2012

Progress Towards Integrated Models of HCV Dynamics

Thomas S. Oh

PROGRESS TOWARDS INTEGRATED MODELS OF HCV DYNAMICS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Thomas S. Oh

June 2012

PROGRESS TOWARDS INTEGRATED MODELS OF HCV DYNAMICS

Thomas S. Oh, Ph.D.

The Rockefeller University 2012

Hepatitis C virus is a small, enveloped positive-strand RNA member of the Flaviviridae and the etiological agent of a global epidemic of chronic hepatitis C. One of the salient features of HCV is a complex regulatory scheme involving numerous viral and cellular components and which may be the key to its striking success in initiating and maintaining decades-long chronic disease in infected individuals. Lacking any known latent or integrated form, HCV must persist through ongoing RNA replication, immune evasion, and infection of naïve cells. HCV is a highly-studied virus, and many details of its life cycle have been worked out; however, much remains unknown. One thing lacking is an integrated model of HCV as a single dynamic system. Various viral life cycle events, such as entry, polyprotein translation and processing, RNA replication, virion morphogenesis and secretion have been studied and worked out to some extent. How these activities are coordinated and how they influence each other is less well-understood. We approached this problem from three angles. We developed quantitative mathematical models designed to accurately recapitulate the specific processes of virus production and accumulation and virus entry. These models were applied to gain a more detailed understanding of these systems. We studied the HCV regulatory protein NS5A through genetic and chemical perturbation to try to understand how it might be connecting different life cycle activities. These studies revealed a

surprising role for the host cellular protein cyclophilin A in regulating virus assembly as well as clarifying aspects of NS5A phosphorylation. These findings suggest a complex system of regulation linking polyprotein translation, RNA replication, and virion assembly which will require further work to decipher. We also pursued a bioinformatics approach, in which we analyzed HCV multiple sequence alignments (MSAs) to identify covariant amino acid positions within HCV proteins. Some of the challenges to the study of covariation in HCV MSAs were identified, and we describe a novel method of calculating covariation which addresses some of these issues, particularly the problem of phylogenetic covariance. While an integrated dynamic model of HCV remains far from achieved, these studies are presented as incremental steps towards that goal.

Soli deo gloria

Acknowledgements

I would like to thank my advisor, Charles Rice, for his guidance, support, and the opportunity to work in this incredible lab. His passion for science, commitment to intellectual rigor, and attention to detail have made a lasting impression on how I think about and do science. I hope that my future work will bear the mark of his influence.

I would also like to thank the members of my committee, Tom Muir and Paul Bieniasz, for their input, encouragement, and critique. Thanks also to Hengli Tang for making the time and effort to review my work and take part in the examination process.

The quality of the lab environment is a decisive factor in the experience of a graduate student, and I could not have asked for a better group to learn science in than the Rice Lab. Everyone in the lab has been extremely generous, helpful and easy to get along with, quick to offer reagents, advice, or a non-alcoholic beverage at the faculty club. In particular I would like to thank Matt Evans, Laura McMullan, Shihyun You, John Law, Ivo Lorenz, Joe Marcotrigiano, Catherine Murray, and Donna Tscherne, former lab members who really helped me get started in the lab and taught me so much in my first few years. Chris Jones, another former member, worked closely with me for a significant portion of the NS5A project; his wit and wisdom made that collaboration one of the more enjoyable experiences of my graduate career. I thank some of the more recent (and current) lab members Bill Schneider, Tom Dentzer, Cynthia De La Fuente, Svetlana Marukian, Martina Kopp, Daniel Calarese, Josh Horwitz, Joanna Loureiro, John

Schoggins, Maria-Theresa Catanese, Hyo-Young Chung, Alex Ploss, Linda Andrus, Tim Sheahan, Ype De Jong, and Marcus Dorner for invaluable discussion, reagents, help with experiments, and video game nights. My acknowledgements would not be complete without mentioning the hard-working staff at the Rice Lab: in particular Santa Maria Pecoraro, Ellen Castillo, Julia Sable, Lili Zhang, and of course, Pat Holst, for all the vital work they do behind the scenes.

A few personal thanks. My wonderful wife and best friend Ellen, who may be the one person who has looked forward to my graduation more than I have: thank you for your patient endurance and loving support through this time. A special thanks to my friends and community at Morningside UMC, for your support and friendship. And lastly, my wonderful family, especially my parents, who have always been to me models of diligence, trustworthiness, and stewardship – thank you for the years of unflagging love and wisdom.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Chapter 1 Introduction

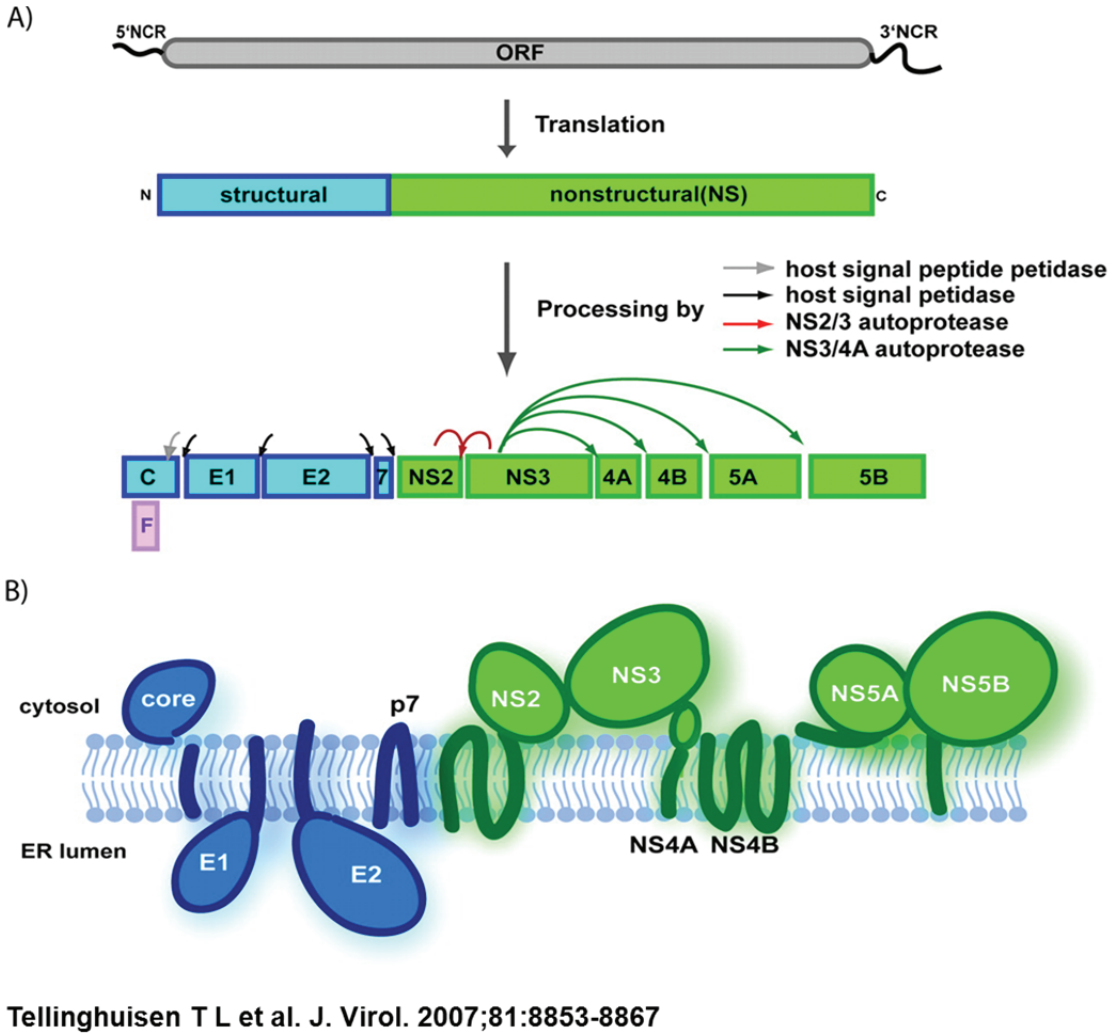
Hepatitis C is a global epidemic, affecting an estimated 160 million people worldwide *(1)*, and is a leading cause of liver disease and hepatocellular carcinoma *(2)*. Prevalence varies widely, with a global average of 2.35% but ranging as high as 14% in countries like Egypt and Cameroon. The acute phase of the disease is often asymptomatic, with an estimated 20% resolving the infection naturally, but the remaining 80% continue on to chronic infection, which is generally life-long without treatment and leads to progressive liver damage. 10-20% of chronically infected individuals will go on to develop complications of liver disease such as cirrhosis within 30 years, and 1-5% will develop liver cancer.

In the United States, HCV-related mortality has increased steadily since its discovery, and recent data indicate HCV has surpassed human immune-deficiency virus (HIV) as a cause of death. It is also the most common chronic blood-borne infection and the main indication for liver transplant *(3)*. The current standard of care consists of pegylated interferon-α and ribavirin, a poorly tolerated regimen with severe adverse effects and success rate of only around 50-60% for genotype 1 HCV patients, the predominant genotype in the U.S. *(4)*. A decade of work on direct-acting antivirals (DAAs) has resulted in the recent approval of two small-molecule inhibitors of HCV *(5)*, with a number of other compounds in the development pipeline. These DAAs, when added on to the standard of care, yield much higher response rates

HCV is the type member of the *Hepacivirus* genus of the Flaviviridae family of enveloped, positive-sense single-stranded RNA viruses. Related genera are *Flavivirus*, which includes the major human pathogens yellow fever virus, West Nile virus, and dengue fever virus, and *Pestivirus*, which includes important agricultural disease-causing agents classical swine fever virus and bovine viral diarrhea virus. Very recently, a second member of the *Hepacivirus* genus was discovered during investigations of a respiratory illness outbreak in dogs. Follow-up study and sequencing uncovered a novel species of the *Hepacivirus* genus, appropriately designated Canine Hepacivirus (CHV) *(6)*. As the closest relative to HCV, comparative studies of CHV and HCV may provide new insights into the biology of both viruses.

The viral genome of HCV comprises a 9.6 kb uncapped RNA molecule bearing a single open reading frame (ORF) spanning most of the genome flanked by shorter 5'- and 3'-untranslated regions (UTRs). The UTRs carry specific RNA secondary structure required for translation and RNA replication, while the ORF codes for the ten viral proteins.

HCV protein translation is initiated from a 5' internal ribosome entry site (IRES) and proceeds through the entire open reading frame. The resulting polyprotein is cleaved by a number of cellular and viral proteases to generate the ten viral proteins, designated C,



Tellinghuisen T L et al. J. Virol. 2007;81:8853-8867

Figure 1-1. HCV genome organization, polyprotein, and topology

(A) HCV genome with single long open reading frame, translated into a single polyprotein, which is post-translationally cleaved by host and viral proteases. (B) Resulting membrane topology of the 10 HCV proteins.

E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. The polyprotein undergoes co-translational translocation across the ER, and the viral proteins are predominantly membrane-anchored or membrane associated. Of these, NS3 through NS5B represent the minimal set of proteins required to initiate and sustain RNA replication along with the 5'- and 3'-UTRs.

1.1 Cell culture systems

Although the HCV genome was originally cloned in 1989, allowing the first studies of HCV as a genetic entity, any research on the viral life cycle itself was severely hampered by a lack of model systems. Molecular clones were infectious in chimpanzees, but no other animal model existed. Replication in cell culture was also was also unsuccessful for over ten years, until the development of subgenomic replicons systems and the discovery of numerous adaptive mutations which increased replication efficiency by orders of magnitude. These subgenomic replicons typically consist of the 5'-UTR with its IRES driving expression of an antibiotic resistance gene, followed by an encephalomyocarditis virus (EMCV) IRES, the subgenome encoding NS3 through NS5B, and the 3'-UTR. These constructs are transcribed in vitro to produce subgenomic RNA, which is transfected into cells and maintained under drug selection. Even the choice of cell line is rather restricted, with efficient replication seen in primary liver cells and a small number of human hepatoma cell lines, in particular Huh7 and its derivatives. Transfection and drug selection results in colonies of drug-resistant cells actively

replicating HCV subgenomic RNA; acquisition of adaptive mutations dramatically increases the frequency of colony formation.

An obvious limitation of these subgenomic replicons is that, not expressing any of the structural proteins, they are not capable of producing new infectious virions, and therefore do not allow investigation of much of the virus life cycle, including assembly, release, and entry. This limitation was overcome with the discovery of JFH-1, a genotype 2a patient isolate cloned from an individual with fulminant hepatitis in 2005. The JFH-1 clone supported both RNA replication and infectious virus production, allowing study of the complete viral life cycle for the first time. This breakthrough was quickly followed up by the report of intragenotypic chimeras of the J6 and JFH-1 strains. These chimeric constructs, comprising the N-terminal region from J6 and the C-terminal region from JFH-1 with a breakpoint either between NS2 and NS3 (J6/JFH) or within NS2 (Jc1), were notable for producing significantly more virus than the wild type JFH1 virus. These two lab strains, along with the Huh7 and Huh7.5 cell lines, have become standard models for the study of the viral life cycle in cell culture.

1.2 The virus particle and entry

The virion is a roughly spherical particle approximately 60nm in diameter displaying the two viral glycoproteins, E1 and E2, on a lipid bilayer envelope. The envelope surrounds the viral nucleocapsid, a multimeric shell of HCV Core protein encapsulating the viral genomic RNA. Virus particles are heterogeneous in composition, size, and

density, and are normally associated with host lipoproteins in larger assemblies known as lipo-viral particles (LVPs). Apolipoproteins B and E (ApoB, ApoE) co-purify with infectivity, suggesting an LVP with some resemblance to VLDL *(7)*. That the maturation and secretion of viral particles utilizes many components of the VLDL assembly pathway further reinforces this notion *(8)*.

HCV entry is a multi-step process involving numerous cellular proteins. Four major entry factors have been identified – SRB1 *(9, 10)*, CD81*(11)*, CLDN1 *(12)*, and OCLDN *(13)*, the ablation of any one of which is sufficient to severely impair entry of HCV and HCV pseudo-typed lentivirus. Likewise, the expression of all four factors renders non-permissive cells permissive for HCV entry. SRB1 and CD81 both exhibit affinity for the HCV E2 glycoprotein and are believed to mediate the first steps, with initial engagement and binding to SRB1 followed by a CD81-mediated priming of the particle for fusion *(14)*. The two tight junction proteins, CLDN1 and OCLDN, are involved in a post-binding step downstream of both SRB1 and CD81 *(12)*. Fusion of the viral envelope is a pH-dependent process *(15)*, implying that fusion takes place following endocytosis and acidification of the endosome.

1.3 Translation and polyprotein processing

After uncoating of the particle, the genomic RNA is translated to generate the HCV polyprotein. The viral RNA is not capped, and translation is initiated from the HCV IRES, which recruits the 40S ribosomal subunit and eukaryotic initiation factor 3 (EIF3), which

is followed by binding of the ternary complex and completion of the 80S ribosome complex. The AUG start codon is positioned in the ribosomal P-site upon recruitment, and no ribosomal scanning is required *(16)*.

As a consequence of HCV's expression as a single polyprotein, numerous cleavage events are required to generate the ten viral proteins, and notably, all viral proteins are produced in equimolar quantities - the non-structural proteins are presumably produced in vast excess. The structural proteins and p7, a small putative ion-channel required for infectious virion production, are cleaved from each other by host signal peptide peptidases and host signal peptidases. Processing of the remainder of the polyprotein is mediated by two virally encoded proteases domains, found in NS2 and NS3 respectively. The structure of the protease domain of NS2 shows a homodimeric configuration in which two copies of the NS2 protein form two composite active sites, with residues from both copies cooperating at each site to catalyze an autoproteolysis which serves the purpose of separating NS3 from the C-terminus of each of the two NS2 molecules *(17)*. This cleavage requires the N-terminal 180 residues of NS3, and no other substrate for the NS2 protease is known apart from the NS2-NS3 junction. The consequent liberation of the N-terminus of NS3 is essential for RNA replication, with NS2 cleavage mutants unable to replicate *(18)*, but the NS2 protein itself is not required for RNA replication, as sub-genomic replicons expressing only NS3-NS5B replicate efficiently *(19)*. Constructs containing an EMCV IRES between NS2 and NS3 also replicate

well, further indicating that, unlike some other members of the Flaviviridae, the uncleaved intermediate does not perform an essential function *(20)*.

The remaining non-structural proteins are cleaved by the N-terminal serine protease domain of NS3, along with the NS4A co-factor, which forms a stable complex with NS3 and is required for efficient processing. The NS3/4A protease also plays an important role in evading the innate immune system of the cell by cleaving a component of the RIG-I signaling pathway, the adapter molecule IPS-1, and thereby disrupting induction of interferon and the cellular antiviral response *(21, 22)*. The crucial role of the NS3/4A protease makes it a very attractive drug target, and numerous small-molecule inhibitors have been identified and are in various stages of development and clinical use *(23–25)*.

1.4 RNA replication

Expression of the integral membrane protein NS4B is sufficient to induce the membrane convolutions, dubbed the membranous web, to which the RNA replication complexes localize *(26)*. These membrane reorganizations form protected structures which enclose the viral RNA replication machinery and may function to conceal viral components from cellular pathogen-associated molecular pattern (PAMP) sensors or control the local biochemical environment.

Once the replication complex is formed, RNA replication depends on the cooperation of the NS3 RNA helicase, the RNA-dependent RNA polymerase NS5B, and NS5A, a phosphoprotein with no known enzymatic activity. For full RNA replication

activity, these proteins must be expressed from the RNA to be replicated; most of the protein functions cannot be *trans*-complemented *(27)*.

The extensive biochemical and structural studies of NS5B as well as its clear function make it the best understood component of the replicase as well as an important drug target *(28, 29)*. As the viral RNA-dependent RNA polymerase, NS5B is obviously essential for HCV replication *(30, 31)*, with mutations in the catalytic GDD motif rendering the virus completely unable to replicate *(32)*. Replication of RNA requires RNA elements in the 5'-UTR *(33, 34)* and 3'-UTR *(31)*, as well as, interestingly, RNA elements within the coding sequence of NS5B *(35)*. Like the NS3/4A protease, multiple inhibitors of the NS5B polymerase are in development *(28)*. Inhibitors include both chain-terminating nucleoside analogues and non-nucleoside allosteric inhibitors directly binding the NS5B molecule.

NS3 is also required for HCV replication, not only for its proteolytic activity in cleaving the other replicase components from each other, but also for a distinct helicase activity *(36)*. The HCV double-stranded RNA is the presumed substrate for this helicase, however its exact function in the RNA replication cycle is unknown. The helicase activity is also a potential drug target, and various inhibitory compounds have been reported *(37–39)*.

1.5 Assembly, Maturation, Secretion

Compared to the earlier events, much less is known about the final stages of the viral life cycle – assembly and secretion of infectious particles. In addition to the structural proteins core, E1, and E2, p7 and NS2 are all dispensable for RNA replication but required for virus particle production *(20)*. A third non-structural protein, NS5A, also plays an active role in virus assembly *(40)* in addition to its essential role in RNA replication *(41)*. Cytosolic lipid droplets (LDs) are believed to be the site of virus assembly, with core protein associating with lipid droplets *(42)* and colocalizing with NS5A, suggesting a rendezvous of HCV RNA and core for packaging into capsids *(43)*. NS5A, an RNA-binding protein and a component of the replicase complex, is thought to shuttle HCV RNA from replication complexes to the sites of assembly and has been recorded making rapid, microtubule-dependent movements over long-distances *(44)*.

NS2 is believed to play a central role in virus assembly, with interactions with structural and non-structural proteins *(45)* suggesting a possible role as a scaffold for virus assembly *(46)*. Mutagenesis studies of NS2 have identified numerous mutations which specifically alter virus production and not RNA replication in both the N-terminal transmembrane domain *(47)* and the protease domain *(48)*, although the protease activity itself is not required *(49)*. Chimeras of J6 and JFH show large differences in virus production which map to NS2 *(50)*, and serial passaging in tissue culture has generated higher titer strains with important adaptive mutations in NS2 *(51, 52)*.

The small, α-helical, trans-membrane protein p7 is also required for virus assembly *(20)*. The protein oligomerizes in vitro to form ion channels which can be blocked by the influenze A M2 inhibitor, Amantadine *(53)*. This ion-channel activity has been shown to de-acidify intracellular vesicles, protecting the nascent virion as it passes through an acid-sensitive phase prior to maturation *(54)*to the final acid-resistant form *(15)*. P7 is also reported to possess a second, ion channel-independent function of regulating NS2 localization and interaction with NS5A *(55)*.

A major hallmark of HCV particle assembly and secretion is its close association with cellular lipid metabolism, especially very low-density lipoprotein (VLDL) synthesis. Several components of the VLDL pathway, including microsomal triglyceride transport protein (MTP) *(8)*, apolipoprotein B (apoB) *(56)*, and apolipoprotein E (ApoE) *(57)* have been shown to participate in HCV assembly and maturation. Both ApoB and ApoE can be found incorporated into LVPs, while MTP is believed to be involved in formation and lipidation of the infectious particle. ApoE has also been shown to interact with NS5A, and that interaction has been correlated with efficient assembly of viral particles *(58)*.

1.6 NS5A

The penultimate member of the HCV polyprotein, NS5A is a putatively three domain phosphoprotein of approximately 460 amino acids (actual length varying somewhat between genotypes and isolates). The three predicted domains are separated by low-complexity sequences (LCS-I and LCS-II), however only domain I has been found to

exhibit any clear structure, with two distinct homodimeric forms having been crystallized. Domain I consists of the first 213 amino acids, which contain an N-terminal amphipathic membrane anchor, four essential cysteine residues which coordinate a single structural zinc-ion, and a large basic groove which is formed at the homodimeric interface of two NS5A molecules and is proposed as an RNA binding surface. NS5A indeed has been shown to have RNA binding affinity, with the ability to bind the 3'- ends of both the plus and minus strands of HCV RNA *(59, 60)*.

The remainder of the protein, domains II and III, are believed to be generally disordered, and NMR studies support their classification as natively unfolded domains. Their lack of distinct structure does not imply a lack of function however, and both domains II and III show important and specific functions in the viral life cycle. At the same time, both domains display a degree of genetic flexibility generally absent from domain I and from the HCV genome at large, with a higher degree of polymorphism across isolates and tolerating various deletions and insertions. Deletion mapping of domains II and III show that for the most part small deletions have little effect on RNA replication. Similarly, a 47-amino acid deletion in domain II has been isolated as an adaptive mutation in subgenomic replicons. However, a 56 amino acid stretch in the second half of domain II is absolutely required for RNA replication and even small deletions in this region render subgenomic replicons completely dead. A total deletion of domain III (except for a short C-terminal stretch required for 5A-5B cleavage) also replicates RNA efficiently; however production of infectious virions is abolished,

pointing to a role of domain III in assembly and release. Similarly, sites in domain III have been shown to accept large heterologous insertions like GFP with little effect on RNA replication, at the cost of severely impaired virus production. More recently, inter-genotypic exchanges of domain III have also been shown to affect virus production in a consistent way *(43)*, suggesting that certain domain III sequences are generally more favorable for virus production than others.

These data point to roles of NS5A in both RNA replication and virus assembly. In fact NS5A has long been proposed as a regulatory switch, balancing these two competing activities. This notion is further encouraged by its ability to bind HCV RNA, imaging studies which have found NS5A localizing with core at lipid droplets *(41)*, the proposed site of virus assembly, and the long-standing observation of NS5A phosphorylation.

Direct inhibitors of NS5A have been developed *(61)*, in particular the compound BMS-790052 *(62)*. This NS5A inhibitor has shown astonishing efficacy, with effective concentrations in the pico-molar range and highly rapid reductions in patient viral load. The mechanism of these inhibitors is unknown, all the more so as NS5A has no known enzymatic function. Resistance has been mapped to domain I, both in cell culture and in HCV patients. Treatment of HCV infected cells with BMS-790052 results in changes in localization *(55)*, polyprotein processing *(63)*, and phosphorylation status of NS5A, and evidence suggests that the inhibitor targets multiple functions of NS5A *(64)*.

1.6.1 NS5A phophorylation

In cells, NS5A exists in two major isoforms differing by their phosphorylation state: the p56 basally phosphorylated form, also referred to as the hypophosphorylated form, and the p58 hyperphosphorylated form. Under phosphatase treatment, the two isoforms collapse down to a single species running slightly faster than the p56 form. Neither the process of phosphorylation nor the functional significance of the two phosphorylation states have yet been fully elucidated, and experimental results have not yet coalesced into a unified picture. The hyperphosphorylated form can be reduced in abundance or eliminated by a variety of means – mutation of serine 232 in domain I (often referred to as S2204, by polyprotein numbering), mutation of serine 457 in domain III, chemical inhibition of casein kinase I-alpha (CKIα), chemical inhibition of casein kinase II *(40)*, and antiviral compounds such as BMS-790052 *(64)*, which target NS5A. The wide variety of ways to reduce p58, and the fact that they affect p58 abundance as an integral species (rather than changing its degree of separation from p56) suggests a cascade of phosphorylation with multiple prerequisite conditions, the disruption of any one of which interrupts the entire process.

The effects of disrupting p58 are as varied as the means. The S2204I mutation in domain I is an adaptive mutation in the genotype 1b Con1, reduces the hyperphosphorylated form and enhances RNA replication 80-fold *(65)*. The same mutation in JFH1, where it also inhibits the formation of p58, severely impairs RNA replication *(64)*. On the other hand, mutating serine 457 to an alanine has minimal

effect on RNA replication but reduces infectious virus production over a thousand-fold, a defect that does not occur when serine 457 is mutated to the phospho-serine mimic aspartic acid *(40)*. Hyperphosphorylation has also been linked to a loss of interaction with the host protein hVAP-A, a required factor for RNA replication, leading to the proposal that hyperphosphorylation negatively regulates replication.

Advancing the theme of a replication/assembly tradeoff, multiple replication-enhancing mutations (REMs) in full length genomes have been demonstrated to reduce virus assembly efficiency *(66)*. Many, if not all, of these REMs also reduce NS5A hyperphosphorylation, reinforcing the link between NS5A phosphorylation status and virus replication/assembly regulation.

These correlations between alterations in the phosphorylation states of NS5A and differences in RNA replication and virus assembly capacity, while not easily understood, provide perhaps the most convincing evidence that NS5A acts as a regulator of the different life cycle activities of HCV.

1.7 Cyclophilin A

The topic of host factors required for the viral life cycle has been one of extensive research and discussion. Numerous siRNA screens, viral protein pull-downs, yeast two-hybrid screens, and small-molecule inihibitor screens have identified a vast array of candidates, and validation of new hits by biochemical and functional assays are published with regularity. These hits include replication factors human vesicle-

associated membrane protein-associated protein A (hVAP-A) *(67)* and phosphatidylinositol 4-kinase (PI4KA) *(68)*, assembly factor annexin A2 (ANXA2), and interferon signaling component PKR.

One of the more extensively studied host replication factors for HCV has been cyclophilin A (CypA, PPIA). This highly abundant, cytosolic peptidyl-prolyl isomerase is an important target of the immunosuppressive drug cyclosporine A (CsA), which when bound to CypA inhibits calcineurin, an activator of IL-2 and the T-cell immune response. However, its role in HCV replication is independent of its function as an immune activator, as demonstrated by analogues of CsA lacking immunosuppressive activity *(69)*. The critical interaction of HCV and CypA is evidenced through direct binding, chemical inhibition, genetic knockdown, and perhaps most convincingly, by the ongoing clinical trials of Alisporivir, a non-immunosuppressive CsA analogue, currently in Phase III and being proposed as the linchpin of what could be the first interferon-free HCV treatment regimen. However, despite these strong findings, the CypA-HCV story has proven complex and is far from complete. At the outset, studies conflicted as to the identity of the required host cyclophilin, with both CypA and CypB identified as the essential replication factor *(70)(71)*. Since then CypA has emerged as the general consensus within the field, although studies continue to explore the role of CypB, focusing specifically on its interaction with NS5B*(72)*. In contrast, CypA's role in HCV replication has been traced to no less than three separate viral proteins, NS2, NS5A, and NS5B. CypA is believed to play a role in the proper folding and activity of the NS2

protease, the crystal structure of which reveals a rare cis-proline which may be the substrate of CypA *(73)*. Experiments separating NS2 from NS3 with an IRES or comparing NS2-5B replicons with NS3-5B replicons show a pronounced decrease in CsA sensitivity when NS2 is not expressed contiguously with NS3.

CsA resistance mutations have been traced to NS5B *(74, 75)*, making it another candidate for interaction partner and drug target. Supporting this hypothesis, NS5B has been shown to associate with CypA via its enzymatic pocket *(76)*, and CsA has been shown to block incorporation of NS5B into replication complexes *(77)*. There is some conflicting data, however, as CypA-independent recruitment of NS5B to replication complexes has also been demonstrated *(72)*.

A consensus has emerged, however, that the primary target of CsA and the primary viral protein partner in the HCV-CypA interaction is NS5A. The interaction of CsA, CypA, and NS5A is the most-studied; nevertheless, findings have been highly diverse, with varying reports as to the locus of interaction, the substrate for CypA's isomerase activity, and mechanism of action. Both domain II *(78, 79)* and domain III *(80)* of NS5A have been shown to interact with CypA, and both domains have been reported to be substrates of the peptidyl-prolyl isomerase. Two proline residues have been identified in domain II which are critical for CypA interaction and for replication *(81)*, but additional prolines scattered throughout domain II are believed to be substrates and contribute to binding in a distributed fashion *(82)*. What effect the CypA interaction has on NS5A function is not yet clear. One possibility is that CypA interaction with NS5A alters the

kinetics of the NS5A-NS5B cleavage event, which may have ramification on both RNA replication and virus replication *(83)*. Another recent report has shown CypA to modulate the RNA binding properties of NS5A domain II *(79)*; the importance of this effect is unclear given that domain I is believed to be the major RNA-binding domain.

1.8 Objective

The objective of this present study is to further our understanding of the HCV life cycle as a system. The long-term goal would be to integrate functional knowledge of HCV molecular biology with quantitative measurements from controlled experimental systems to deepen our understanding of HCV as a system and to develop more comprehensive conceptual and numerical models. The work presented here describes three avenues of approach towards that goal. In chapter 3 we develop two simple mathematical models which we then refine with experimental data to arrive at quantitative predictive models. We then explore some of the implications of these models as well as some potential applications. In chapter 4 we focus on a specific viral protein, NS5A, and probe its role in regulating the various components of the HCV life cycle. In chapter 5 we explore a completely different, bioinformatic approach to HCV biology: covariation analysis, a statistical study of mutational patterns in HCV sequence alignments. These three approaches do not converge in this work; however, these approaches are potentially convergent and we also outline future directions towards this long-term goal.

Chapter 2. Materials and Methods

Chapter 2 Materials and Methods

2.1 Cell culture

Continuous cell lines were cultured at 37°C in 5% CO_2.

2.1.1 Huh7.5 (Human hepatoma)

Cells were propagated in Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen) supplemented with 10mM non-essential amino acids and 10% fetal bovine serum (FBS, Omega Scientific) *(30)*.

2.1.2 Huh7.5(CD81-)

Huh7.5 cells transduced with a shRNA against the HCV entry factor CD81 are described in *(84)*. Cells were propagated as described for the parental Huh7.5 cell line with the addition of 6 µg/ml Blasticidin-S (Invitrogen).

2.2 Plasmid Constructs

2.2.1 Cloning primers

DNA oligonucleotides were ordered and custom synthesized (IDT) for use in generating the plasmid constructs used in this work. The primers and sequences are listed in Table 2-1.

Table 2-1. Cloning primers

Primer Number	Sequence
RU-O-5933	5’-TTGCCTTTGCTTCCAGAGG-3’
RU-O-6769	5’-GAGAGGGTACGCGTTTAGAGAAGACTCCACCCCTTGG-3’
RU-O-12685	5’-CATAGTGCGCTCAGCCGTGAGCAAAGGCGAAGAGCTG-3’
RU-O-12686	5’-CGTCTACTAGATCTCTTATAGAGCTCGTTCATGCCCTCGG-3’
RU-O-12709	5’-GATTCAGGAGTGGGTACCACCCATGC-3’
RU-O-14251	5’-CAGCTCTTCGCCTTTGCTCAC-3’
RU-O-14775	5’-GAGGCTCAGCCGTCAACCCCACCGTGTTCTTC-3’
RU-O-14776	5’-GCCAGATCTTTCGAGTTGTCCACAGTCAGC-3’
RU-O-15105	5’-ATGCACTCCGATATGAGCAGGTTGGCATCGACC-3’
RU-O-15106	5’-CAACCTGCTCATATCGGAGTGCATGCTCCCCAG-3’
RU-O-15869	5’-GGACGTCGGACCGCCGGATTCGG-3’
RU-O-15870	5’-GGATGGCAAGCAGGTGGTGTTTGG-3’
RU-O-15871	5’-CCAAACACCACCTGCTTGCCATCC-3’
RU-O-16248	5’-GCTGTCGGGTCCCAGCTTCCCTGT-3’
RU-O-16249	5’-CGGCCTGAGAACAACCCGCCGCTCGTGGAA-3’
RU-O-16250	5’-CGGGTTGTTCTCAGGCCGTGCCCAAGCCGG-3’

2.2.2 Jc1FLAG2(p7nsGluc2A)

This reporter genome based on the Jc1 chimeric genotype 2A virus has been previously reported *(85)*.

2.2.3 Jc1(NS2-IRES-nsGluc2AUbi).

The previously published reporter construct J6/JFH(NS2-IRES-nsGluc2AUbi) *(20)* was modified to contain the Jc1 breakpoint in NS2 *(50)*. The primer pair RU-O-12709/RU-O-6769 was used to amplify a fragment of the Jc1 NS2 sequence with an N-terminal NotI site and a C-terminal MluI site corresponding to sites in the J6/JFH(NS2-IRES-nsGLuc2AUbi). A second fragment was generated by digesting Jc1 with BsaBI and NotI. These two fragments were ligated into the J6/JFH(NS2-IRES-nsGluc2AUbi) vector digested with BsaBI and MluI.

2.2.4 Jc1(378-Ypet)

A multiple cloning site encoding BlpI and BglII site surrounded by flexible linkers (GGGS) was engineered into APP23 at position 378 (NS5A protein numbering, domain III), resulting in the construct Jc1 378-Link. Ypet was amplified from Bi-Ypet-Jc1-FLAG2 *(86)* using primers RU-O-12685 and RU-O-12686 to generate a Ypet amplicon with flanking BlpI and BglII sites, which was then cloned into the Jc1-378-Link plasmid.

2.2.5 Jc1(Δ34-378-Ypet)

Virus production-competent revertants were generated by passaging supernatants from Jc1(378-Ypet). These clones were reverse transcribed and sequenced. A deletion found in NS5A was engineered back into the Jc1(378-Ypet) background, yielding the Jc1(378-Ypet-1) construct. The deletion was reengineered by PCR amplification of the revertant cDNA using the primer pair RU-O-5933/RU-O-14251, which was digested with BspEI and BlpI. A plasmid prep of J6/JFH made in DAM- E. Coli was digested with SpeI and BspEI, and the 2,164bp fragment was purified. This fragment along with the PCR amplicon was ligated into the SpeI-BlpI vector of Jc1(378-Ypet) in a 3-fragment ligation.

2.2.6 Jc1-H34L, Jc1(378-Ypet-H34L)

The NS5B adaptive mutation H34L has been previously published *(87)*. The N-terminal region of NS5B containing the mutation was cloned into Jc1 and 378-Ypet background by the unique restriction sites RsrII and HindIII.

2.2.7 Jc1-Δ34

The -1 deletion from Jc1(Δ34-378-Ypet) was engineered back into the original Jc1 using overlapping PCR mutagenesis to make a construct with the deletion but without the Ypet fusion in NS5A. NS5A was amplified from Jc1 with the primer pairs RU-O-5933/RU-O-15105 and RU-O-15106/RU-O-15869. The two resulting overlapping amplicons were assembled using the outer primers and digested with BspEI and RsrII.

The same DAM- SpeI-BspEI fragment described above in the Jc1(Δ34-378-Ypet) cloning was used along with the PCR assembled fragment in a 3-fragment ligation into the SpeI-RsrII vector of Jc1.

2.2.8 Jc1(378-CypA) and Jc1(Δ34-378-CypA)

Cyclophilin A was PCR amplified from a commercially available PCMV-SPORT6-PPIA plasmid (Open Biosystems) using primers RU-O-14775 and RU-O-14776. These primers carry 5' overhangs with BlpI and BglII sites respectively, matching the cloning sites flanking the NS5A 378 locus and allowing the CypA gene to be swapped for the Ypet in both the Jc1(378-Ypet) and Jc1(Δ34-378-Ypet).

2.2.9 Jc1(378-mutCypA) and Jc1(Δ34-378-mutCypA)

Residue 126 of CypA was mutated from a histidine to a glutamine (H126Q) by overlapping PCR mutagenesis, resulting in a catalytically inactive mutant of CypA. One fragment was generated by amplifying CypA with RU-O-14775 and RU-O-15871, and the second fragment was made using RU-O-15870 and RU-O-14776, with the active site mutation encoded in the two inner primers. The two overlapping fragments were amplified in an assembly PCR reaction using the outer primers to generate the full length mutant CypA. BlpI and BglII sites in the outer primers allowed the mutant CypA to be swapped into the Jc1(378-Ypet) and Jc1(Δ34-378-Ypet) backbone.

2.2.10 Serine 457 mutants – Jc1(AEED), Jc1(DEED), Jc1(378-Ypet-AEED), Jc1(378-Ypet-DEED), Jc1(Δ34-378-Ypet-AEED), and Jc1(Δ34-378-Ypet-AEED)

Mutations of serine 457 of NS5A to alanine and aspartic acid have been previously generated in the J6/JFH background *(40)*. The C-terminal region of NS5A containing the mutations was transferred into Jc1, 378-Ypet, and Δ34-378-Ypet by restriction digest cloning using the unique RsrII and HindIII sites in NS5A and NS5B respectively.

2.2.11 Jc1(DEYN)

The CsA resistance mutations D316E and Y317N were generated in Jc1 by PCR mutagenesis. Overlapping fragments were made from the primer pairs RU-O-16248/RU-O-16250 and RU-O-16249/RU-O-15869. These fragments were joined by assembly PCR and cloned into Jc1 by SanDI and RsrII.

2.2.12 Cloning into the Jc1(NS2-IRES-nsGluc2AUbi)

378-Ypet and Δ34-378, the corresponding CypA fusions and mutCypA fusions, the DEYN mutant and the serine 457 mutants were all cloned into the Jc1(NS2-IRES-nsGluc2AUbi) background using the unique SanDI and XbaI sites.

2.3.1 HCV RNA transcription

HCV template plasmids were linearized by XbaI restriction digest for 3h at 37°C and purified on MinElute columns (Qiagen). RNA was transcribed from 1 µg of linearized template in a 20 µL T7 RiboMax Express (Promega) in vitro transcription reaction at 37°C for 90 min, followed by a 15 min digestion with 1 µL of the supplied RQ1 DNAse. RNA was purified from the reaction using an RNeasy column (Qiagen), quantified by 260 nm absorpance, and visualized by 1% agarose gel electrophoresis to verify integrity.

2.3.2 HCV RNA transfection

Huh 7.5 cells (or derivative cell lines) were trypsinized, washed twice with ice-cold RNAse-free PBS, and counted. The cells were finally resuspended in PBS at a concentration of 1×10^7 cells/ml. 400 µL of cell suspension was mixed with 5 µg of RNA, transferred to a 2 mm gap electroporation cuvette (BTX), and given 5 pulses of 860 V, 99 µsec using an ElectroSquare Porator ECM 830 (BTX). The cells were left for 10 min before being resuspended in complete DMEM with 10mM HEPES and subsequently distributed into multi-well plates or dishes as needed.

2.4.1 *Gaussia* luciferase assay

To measure intracellular guassia, cells were lysed with 1x Renilla Lysis Buffer (Promega). To measure guassia secreted into the extracellular media, cell supernatants were collected and immediately diluted 1:1 in 2x Renilla Lysis Buffer. 10 µL of each sample were transferred to an opaque white 96-well plate and assayed for luciferase activity with the Renilla Luciferase Assay Kit (Promega) on a FluoSTAR Omega plate reader (BMG). 50 µL of Renilla substrate was injected into each well and luminescence was counted for 5 sec.

2.4.2 Cell-associated infectivity

Following electroporation, cells were plated into poly-lysine coated 96-well plates and cultured. At the endpoint time, cell media was removed and each well was filled with 200 µL PBS, aspirated, and refilled with 50 µL complete DMEM. The entire plate was subjected to three freeze-thaw cycles at -80°C (>30 min) and room temperature (10 min). To assay infectivity, the entire sample was pipetted up and down to mix and put onto naïve Huh7.5 cells.

2.4.3 Luciferase-based infectivity assay

Naïve Huh7.5 cells were plated in 96-well plates at a density of 10^4 cells per well. The following day, infectious samples of guassia-expressing HCV were added to each well

and incubated to allow infection and expression of guassia. After 48 h, cells were lysed and assayed according to the *Gaussia* luciferase assay described above.

2.4.4 HCV titer by limiting dilution

Naïve Huh7.5 cells were plated in 96-well plates, 5×10^3 cells per well. The following day, HCV-containing supernatants were serially diluted and 50 µL of each dilution were added to each of 8 replicate wells. The infections were allowed to proceed for 72 h before cells were fixed and stained for HCV NS5A antigen by immunohistochemistry.

2.4.5 HCV NS5A immunohistochemistry

Cells to be stained were fixed in 100% methanol at -20°C for 30 min and then washed (each wash step is 2 times in PBS, once in PBS-T). Cells were blocked in 1% (w/v) BSA and 0.2% milk in PBS-T (PBS with 0.1% Tween-20) for 30 min and then incubated in 3% hydrogen peroxide. Cells were incubated for 1 h in a 1:25,000 dilution of 9E10 mouse anti-NS5A monoclonal antibody *(88)* in PBS-T and then washed. An HRP-conjugated anti-mouse antibody (Jackson, 1:200 in PBS-T) was added to the cells and incubated for 30 min. The cells were washed once more and developed with DAB peroxidase substrate (Vector) for 5 min before stopping the reaction with a PBS rinse.

2.4.6 HCV RNA qRT-PCR

HCV-infected cells in 96-well plates were lysed in 30 µL of SideStep Lysis and Stabilization Buffer (Agilent) and frozen at -80°C. Thawed lysates were diluted 1:20 in

nuclease-free water and 5 µL of diluted sample was added to make a 25 µL MultiCode-RTX (EraGen) qRT-PCR reaction. Amplification and quantification was performed on a LightCycler 480 (Roche) real-time PCR system.

2.4.7 Cell viability assay

Cell viability was assessed using the CellTiter-Glo assay (Promega). Cells were cultured in a 96-well format. The well volume adjusted to 50 µL of complete DMEM and 50 µL of CellTiter-Glo reagent was added to each well. Plates were rocked for 30 min to mix and assayed for luminescence on a FluoSTAR Omega plate reader (BMG) with an integration time of 0.5 s.

2.4.8 HCV NS5A Western Blot

Supernatants were removed from cells, and cells were lysed in SDS sample buffer. Lysates were resolved by PAGE on SDS-10% polyacrylamide denaturing gels and transferred to PVDF membranes. Membranes were blocked in 5% milk/PBS-T for 1 h. The 9E10 anti-NS5A mAB *(88)* was allowed to bind for 1h at RT in 5% milk/PBS-T (1:1000 dilution). Membranes were washed 5 times in PBS-T and incubated for 30 min at RT with an anti-mouse-HRP secondary antibody (Jackson) in 5% milk/PBS-T (1:5000 dilution). After an additional 5 rinses with PBS-T, blots were developed using the SuperSignal West Pico chemiluminescent substrate (Pierce).

2.5 Inhibitors

Cyclosporine A (CsA) was purchased from Sigma (St. Louis, MO). 2'-C-methyl-adenosine (2'CMA) was obtained from Merck (Whitehouse Station, NJ). BMS-790052 and danoprevir were purchased from Selleck Chemicals (Houston, TX). H479 was kindly provided by Raffaele De Francesco (IRBM).

2.6 Experiments

2.6.1 Inhibitor Dose Response

Following electroporation, HCV-transfected Huh7.5 cells were transferred to 96-well plates at a density of 10^4 cells per well in a well volume of 100 µL. Inhibitors were diluted in complete media to 3x the final desired concentration. 50 µL of 3x inhibitor was added to each well for a final volume of 150 µL.

2.6.2 Inhibitor kinetic response

Following electroporation, HCV-transfected Huh7.5 cells were seeded into multiple 96-well plates. Infection was allowed to proceed for 24 or 48 h before the addition of chemical inhibitors. Entire plates were harvested at each selected time point following drug addition and assayed for RNA replication, luciferase expression, and/or intracellular infectivity.

2.6.3 Infection volume curve

Huh7.5 cells were seeded into poly-L-lysine treated 24-well plates and incubated for 24 h to allow attachment. Wells were then aspirated and then infected with either Jc1FLAG2(p7nsGLuc2AUbi) or TOTO1101(Fluc) *(89)* in varying volumes of complete DMEM ranging from 200 µL to 2 ml. Infections were allowed to proceed at 37°C for 8 h and 4h for HCV and Sindbis virus infections, respectively. Wells were then washed once with 2 ml complete media and filled with 400 µL complete media. Plates were harvested either 48 h (HCV) or 4 h (Sindbis) subsequently and assayed for luciferase activity to determine the level of infection.

2.6.4 Virus production kinetic

Huh7.5 cells were electroporated with Jc1FLAG2(p7nsGLuc2AUbi) and plated in poly-L-lysine coated 24-well plates. 24 h after electroporation, wells were washed twice with PBS and refilled with complete DMEM. At fixed time intervals following the wash, cell supernatant was sampled and split for the secreted *Gaussia* luciferase assay and luciferase-based infectivity assay.

2.7 Numerical models

Several rate equation models were developed to describe various aspects of the HCV life cycle. These models were represented as systems of ordinary differential equations

(ODEs), with an independent variable of time and each dependent variable representing a state value of the system (i.e. number of virions, infected cells, RNA molecules, etc.)

2.7.1 Virus release and stability model

To model the release and stability of virus into the cell supernatant, a single ODE was postulated:

$$\frac{dN_V}{dt} = R_P(t) - \frac{N_V}{\tau_V}$$

Here N_V represents the number of infectious virus particles, $R_P(t)$ is the rate of production, and τ_V is the lifetime of the virus in the supernatant. The rate of production is time dependent and can be fit using a polynomial regression to an experimentally determined rate protein expression.

By way of comparison, a second model was proposed which considered the possibility of a secreted factor which might actively degrade virus particles. The model is accordingly modified:

$$\frac{dN_V}{dt} = R_P(t) - \frac{N_V}{\tau_V} - N_V N_F k$$

$$\frac{dN_F}{dt} = R_{PF} - \frac{N_F}{\tau_F}$$

N_F represents the secreted degrading factor with lifetime τ_F [s] and degradation rate k [s^{-1}], which is secreted a rate R_{PF}.

2.7.2 Virus entry model

The binding and entry process was modeled with a system of three equations, simulating the kinetics of entry while virus in the supernatant continues to degrade.

$$\frac{dN_{CU}}{dt} = -\frac{RN_{CU}N_V}{Vol}$$

$$\frac{dN_{CI}}{dt} = \frac{RN_{CU}N_V}{Vol}$$

$$\frac{dN_V}{dt} = -\frac{RN_{CU}N_V}{Vol} - \frac{N_V}{\tau_V}$$

where N_{CU} = number of uninfected cells, R [ml*s^{-1}]= entry rate, N_V = number of virions, Vol = infection volume (ml), N_{CI} = number of infected cells, and τ_V = virus decay lifetime.

2.8 Numerical solutions for systems of ordinary differential equations

ODE rate equation models were solved using a fifth-order Runge-Kutta-Fehlberg algorithm *(90)* implemented in Visual C++ 6.0 (Microsoft). Rate equations were also custom coded in C++ and data was output to text files for plotting.

2.9 Covariation

2.9.1 HCV sequence alignments

HCV protein multiple sequence alignments (MSAs) were downloaded from euHCVdb *(91)*. For the covariance analyses, only non-gapped sequences of equal length were

considered. For each protein, the most common length was selected and all sequences of different length were discarded from the set. Accession and genotype information was also retrieved from the database. Where appropriate, sub-alignments were generated by filtering by genotype annotation.

2.9.2 Implementation of covariance algorithms

Covariance analyses were done using custom software written in Visual C++ 6.0. MSAs were imported from text files, and all algorithms were implemented in C++. Details for each algorithm are provided in within the chapter text.

Chapter 3. HCV Kinetics and Numerical Models

Chapter 3 HCV Kinetics and Numerical Models

Numerical models have been applied to understanding various aspects of HCV, including in vitro RNA replication *(92)*, epidemiology *(93, 94)*, and most of all, in vivo response to clinical therapy *(95–97)*. These models often provide valuable insight precisely because of their quantitative nature, revealing through rate constants and half-lives subtleties of the underlying biology which could be missed in a more qualitative study. A most striking example comes from the field of HIV, in which a set of modeling studies in the mid-1990s provided the theoretical principles for the development of highly active anti-retro viral therapy (HAART), also known as combination therapy *(98–100)*. Two models, representing HCV production and HCV entry, are presented here along with corresponding experimental data and analysis.

3.1 Virus production and decay

3.1.1 A simple model describes virus production and decay

As an initial foray into the effort of matching experimental results to numerical models, the process of HCV release into the supernatant and subsequent decay was chosen. A simple one-equation model was initially proposed to describe the system.

$$\frac{dN_V}{dt} = R_P - \frac{N_V}{\tau_V}$$

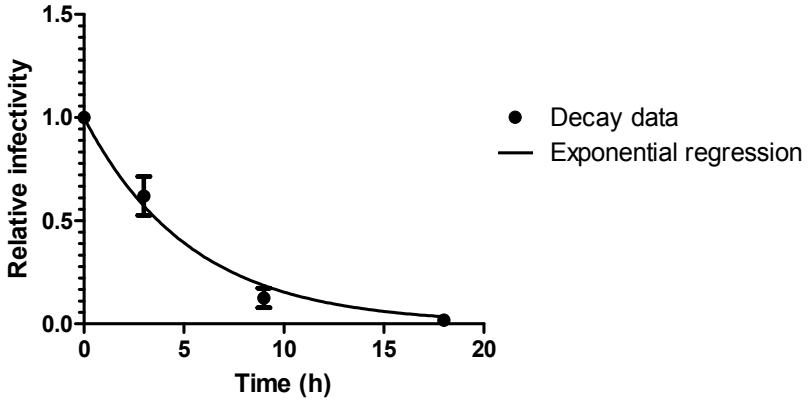


Figure 3-1. Thermal decay of HCV at 37°C

Relative infectivity as a function of incubation time normalized to non-incubated virus shown with exponential regression. Means and SEM of at least two independent experiments are shown.

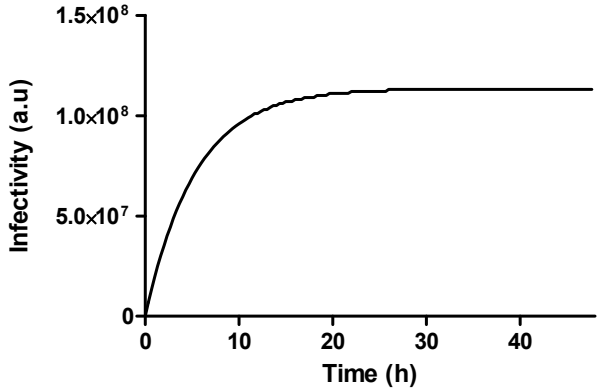
Table 3-1. Production model parameters

Parameter	Description	Value	Units
R_P	Rate of production	arb	virions/s
τ_V	Virus lifetime	5.35	h
$R_P(t)$	Time-dependent rate of production	685.8t + 29666	virions/s

In this model, virus is produced at a rate R_P and decays with an average rate τ_V. The instability of the infectious virus particle can be experimentally determined by measuring the decrease of infectivity in a virus sample over the course of incubation at 37°C. This data is shown in Figure 3-1, with an exponential regression showing a half-life of 4.0 h. This experimentally determined value can be plugged into the model above and then solved numerically to predict the accumulation of infectivity in the supernatants of infected cells. The timescale is assumed to be short enough that the effects of secondary infection can be ignored. The predicted accumulation as a function of time is shown in Figure 3-2.

The actual shape of the infectivity accumulation curve can be experimentally determined by washing virus-producing cells to remove all preexisting virus, and taking samples of the supernatants at selected time intervals following the wash. The resulting data shows some resemblance to the predicted curve in its basic shape and the plateau; however, the experimental curve continues to rise rather than reach a steady state. This discrepancy could be explained by a time-varying rate of virus production, which might be caused by an increase in overall replication levels either on a per cell basis, or as a result of ongoing spread. Ongoing spread was ruled out by the use of Huh7.5 cells deficient in the HCV entry factor CD81. In order to assess a possible increase in per cell viral replication, a reporter gene, secreted *Gaussia* luciferase, was monitored simultaneously with the release of infectious virus. Because the secreted luciferase is

A



B

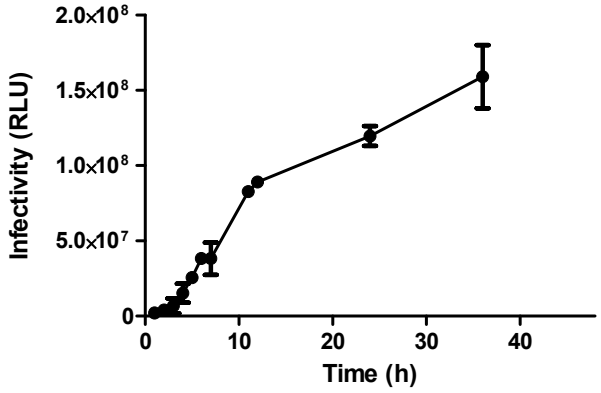


Figure 3-2. Production model and experimental data

(A) Predicted HCV titers from model assuming constant production, shown as a function of hours post-wash. (B) Actual HCV infectivity measured by *Gaussia* luciferase infectivity assay. Cells were electroporated and then washed after 24 h. Means and SEM of at least two independent experiments are shown.

highly stable in the cell supernatant (data not shown), an unvarying level of replication would result in a linear increase of luciferase over time. Instead, luciferase accumulation accelerates over time (Figure 3-3). A polynomial regression was applied to the luciferase data, and R_P, the rate of virion production in the model, was changed to a time-dependent variable proportional to the first derivative of luciferase accumulation. Applying this correction resulted in a markedly improved correlation between the numerical model and experimental results.

The remaining discrepancy can be described as an initial delay after washing. The reason for this discrepancy is not immediately apparent, and may be an effect of the washing step, which subjects the cells to changes in temperature, CO_2 levels, and composition of the extracellular media. The quadratic fit to the luciferase also shows a delay after washing, although this delay of 1 h is smaller than the delay observed in virus production. If a 3 h delay is allowed in the model, the agreement between the model and the experimental data becomes quite high, with an R-squared value of 0.98.

3.1.2 Applications of the virus production model

One practical application of the numerical model is in optimizing the production of virus stocks, specifically in timing the harvest of supernatants. It is possible to simulate an arbitrary set of incubations and harvests and predict the virus yield. Figure 3-4 shows the effect of harvesting frequency on virus yield. At one extreme, a single harvest

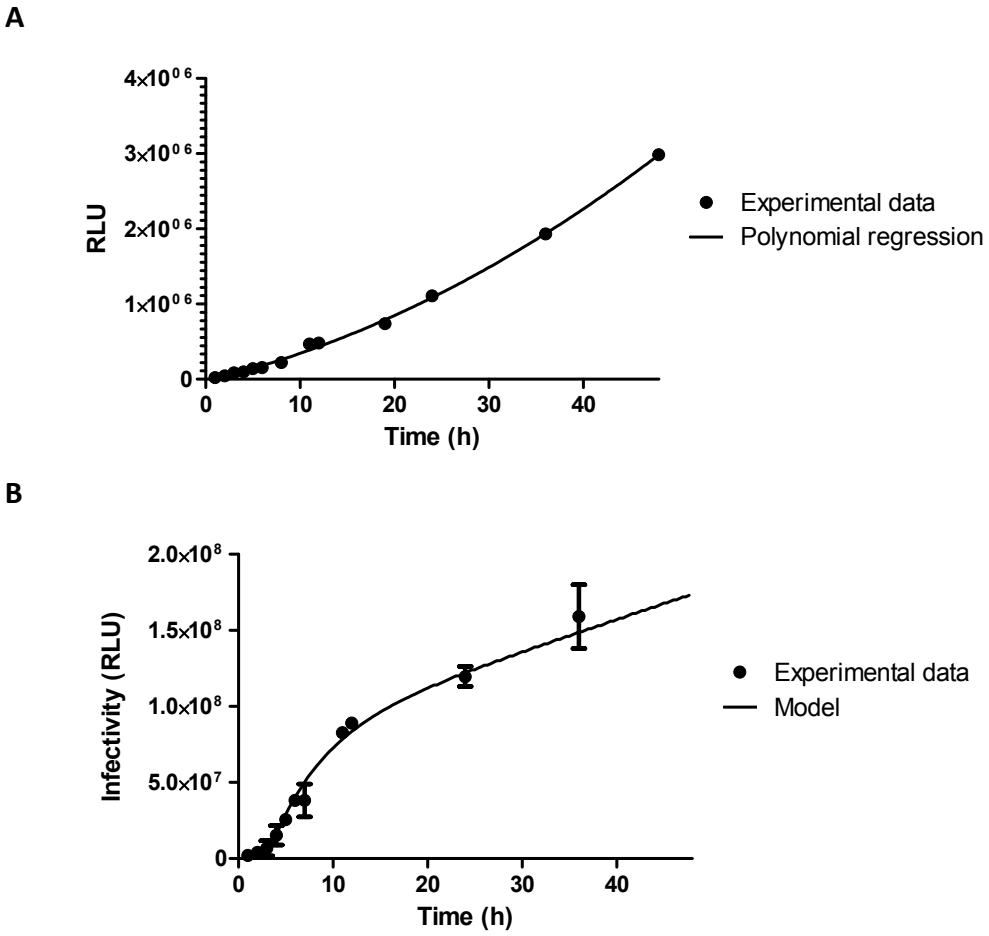


Figure 3-3. Luciferase expression corrected model

(A) Luciferase expression from HCV producer cells as a function of time post-wash, overlaid with a second-order polynomial regression. (B) Overlay of actual HCV production data with output from luciferase-corrected numerical model (R^2=0.98). A three hour delay has been added to the model. Means and SEM from two independent experiments are shown.

following a 48 h incubation yields the highest concentration (by minimizing the total volume of media used), but a lower total yield of virus. If the frequency of harvests is increased, loss due to thermal decay can be reduced by moving harvested supernatants to 4° C, where the virus is more stable. In this way, more total virus is obtained, though at a lower concentration. Based on the model, one might choose to harvest every 8 h, which yields four times more total virus than harvesting once after 48 h while maintaining a final concentration which is only reduced by 40%.

The close agreement between the numerical model and the experimental data (apart from the initial delay in virus release after wash) also supports the completeness of the model and argues against additional factors affecting virus release and decay. It appears that virus production is generally proportional to viral protein expression (as evidenced by luciferase release), and that the decay rate is an essentially intrinsic matter of virus particle stability, without a need to invoke any additional mechanisms, such as a separate secreted virolytic cellular factor. Enzymes such as hepatic lipase *(101)* and lipoprotein lipase *(102)* have been proposed as agents of virus degradation; the completeness of the model described here argues either that such factors are not major mechanisms of virus decay , or that such factors are co-packaged with the virus, making their activity intrinsic to the virus particle.

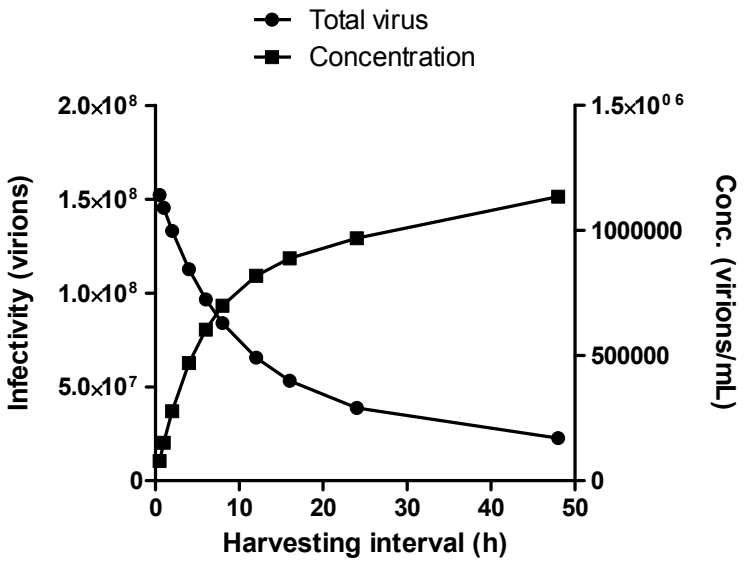


Figure 3-4. Effect of harvest interval on virus stocks

Different harvest intervals over a 48 h period simulated with the HCV production model, going from frequent harvests on the left to a single harvest after 48 h as the final point on the right. Absolute scale of simulated virus number is arbitrary but consistent with concentration on the secondary Y-axis.

3.2.1 Description of binding and entry model

The process of virus uptake involves diffusion of the particle to the cell, attachment to cell surface receptors and internalization, accomplished by cellular endocytosis machinery. Alternatively, the entire process can be simply conceived of as a single binding-entry event. This process can be described by the following model:

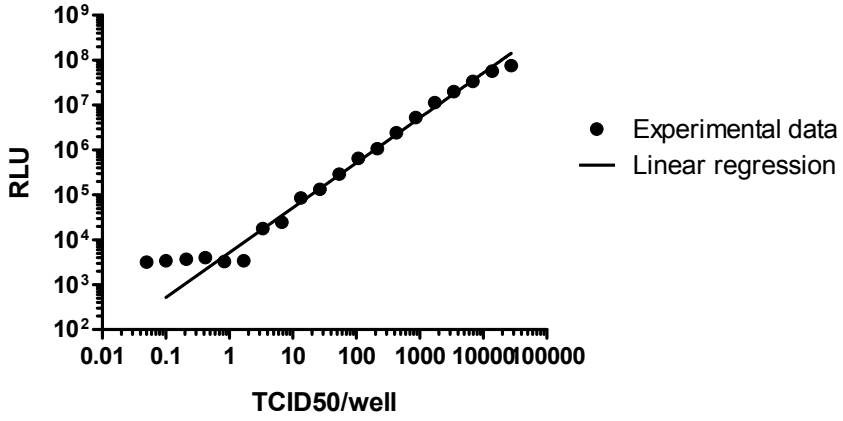
$$\frac{dN_{CU}}{dt} = -\frac{k_I N_{CU} N_V}{Vol}$$

$$\frac{dN_{CI}}{dt} = \frac{k_I N_{CU} N_V}{Vol}$$

$$\frac{dN_V}{dt} = -\frac{k_I N_{CU} N_V}{Vol} - \frac{N_V}{\tau_V}$$

Three state variables – the number of uninfected cells N_{CU}, the number of infected cells N_{CI}, and the number of virus particles N_V – describe the system, augmented by a number of constant parameters: the virus decay lifetime (determined experimentally in the previous section), the infection volume, and a rate constant k_I, with dimensions of ml/(cell × virus × s), which is specific for a particular well geometry and describes the overall rate at which the binding/entry process proceeds. According to this model, the rate of cells becoming infected is proportional to the probability of a virus-cell

A



B

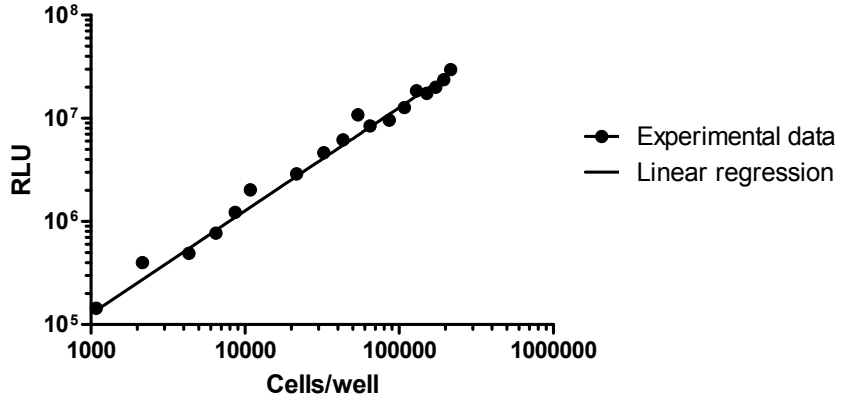


Figure 3-5. Effect of virus dose and cell number on infection

(A) Effect of varying virus dose on total infection showing a wide linear range down to 3 TCID50/well, with linear regression. (B) Effect of varying cell number of total infection, with linear regression. Range spans approximately 0.5% to 80% confluence in one well of a 24-well plate.

interaction (number of virions times number of cells divided by interaction volume). The infection is further limited by the previously quantified decay of virus in the supernatant.

3.2.2 Fitting model to experimental data shown a low rate constant for entry

To generate real-world data to test against the model, several experiments were carried out in which a single variable was varied across a range while the remaining values were held constant. The dependence of infection rate on cell number, virus number, and infection volume were thus separately tested and plotted, as show in Figure 3-5 and Figure 3-6. Infection efficiency was found to be proportional to cell number and virus dose across a wide range, as expected from the nature of the model and providing some confirmation that the basic form of the infection model was correct.

The dependence of infection on the incubation volume was more interesting, showing a rather strong nonlinear dependence on volume. This shape of this curve can be reproduced with excellent agreement to the experimental data, with an R^2 value of 0.98, by setting the infection rate constant k_I equal to 8×10^{-11} (Figure 3-6).

As a point of comparison, the same experiment was repeated with Sindbis virus. The best fit to the experimental data was achieved with the value of k_I set to 3×10^{-9}, a difference in infection rate constant of nearly 40-fold. While measurements for this constant have not been previously reported, the comparison is consistent with the report of a relatively high mean time to infection for HCV *(103)*.

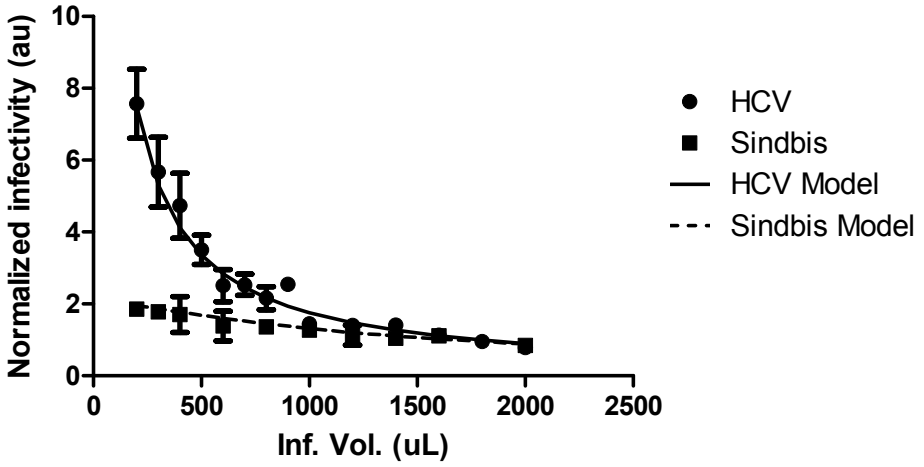


Figure 3-6. Constant dose infection volume curve

Effect of varying infection volume for a fixed virus dose. HCV and Sindbis virus are compared, overlaid with numerical model results. Means and SEM of at least two independent experiments are shown.

Table 3-2. Entry model parameters

Parameter	Description	Value	Units
k_I	Entry rate constant	8×10^{-11} (HCV), 3×10^{-9} (sindis)	ml/(cell × virus × s)
τ_V	Virus lifetime	5.35	h
Vol	Infection volume	0.2 - 2	ml
$N_{CU}0$	Initial number of cells	1,000 - 200,000	cells
N_V0	Initial number of virions	10-10,000	virions

3.2.3 Consequence of low entry rate constant on measures of infectivity

The low value of k_I in an HCV infection has a number of consequences on how infectivity is measured. One result of the slow infection of HCV is that virus titers tend to be underestimated compared to the actual number of infectious particles present in the sample. The dependence of effective infectivity on k_I is shown in Figure 3-7. At higher values of k_I (faster infection), the number of infected cells accurately reflects the infectious content of the inoculum. With decreasing k_I, however, more of the initially infectious virus is lost to thermal decay before the virus is able to enter. For Sindbis virus in a typical infection volume, this represents a relatively minor loss of 20%. For HCV on the other hand, 90% of the infectivity is lost. This figure should also apply to a standard virus titering assay; determinations of virus titer therefore typically underestimate the number of infectious particles in a given virus stock by 10-fold. It is also advisable that the infection volume for the virus titering be standardized.

By extension, measures of specific infectivity could potentially be underestimated as well, as infectivity measurements are subject to this problem while RNA and core protein measurements may not be subject to the same thermal decay. It is interesting to consider a ten-fold revision of previously reported values of specific infectivity.

Because of the low infection rate constant for HCV and the resulting sensitivity to infection volume, the concept of virus dose is much less applicable to HCV as compared

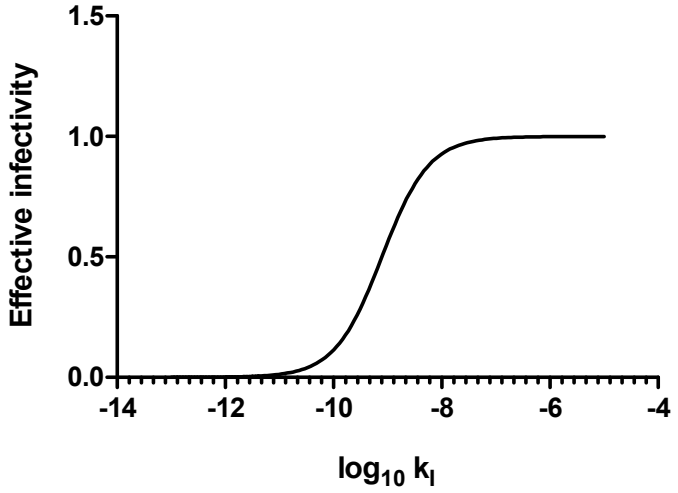


Figure 3-7. Effective infectivity as a function of k_I

Fraction of total infectivity in inoculum which can productively infect cells as a function of infection rate constant, k_I, based on numerical model. Effective infectivity is a fraction of actual infectivity, non-infectious particles are not considered.

with a virus like Sindbis. At a fixed dose, the observed infectivity drops by over 70% when the volume is changed from 200 µl to 800 µl, whereas in a Sindbis virus infection, only 20% is lost over the same range of volumes.

On the other hand, given a fixed concentration of virus, rather than a fixed dose, observed HCV infectivity is almost invariant over the range of volumes, changing by only 10%. Figure 3-8 shows the effect of changing infection volume at a fixed concentration of virus, with some preliminary experimental data which supports the model. For HCV therefore, the virus concentration is actually a much more meaningful measure of infectivity than virus dose, as the number of cells infected in a well is nearly constant for a fixed concentration over a range of volumes. Conversely, Sindbis infectivity varies significantly when the virus concentration is held constant but the volume is changed.

Another commonly used descriptor for cell culture viral infection is the multiplicity of infection (MOI), defined as the ratio of infectious virions to target cells. The MOI is often used to calculate the percentage of cells which will be infected and the distribution of uninfected, singly-infected, and multiply-infected cells. Based on the numerical model, using the MOI in this way also suffers from certain fundamental difficulties. The first is the same as already discussed – the fact that effective infectivity is dependent on the infection volume. The second problem is the way in which cell number affects infection frequency, even when volume is held constant. One of the assumptions underlying the use of MOI is that cell number does not affect the number

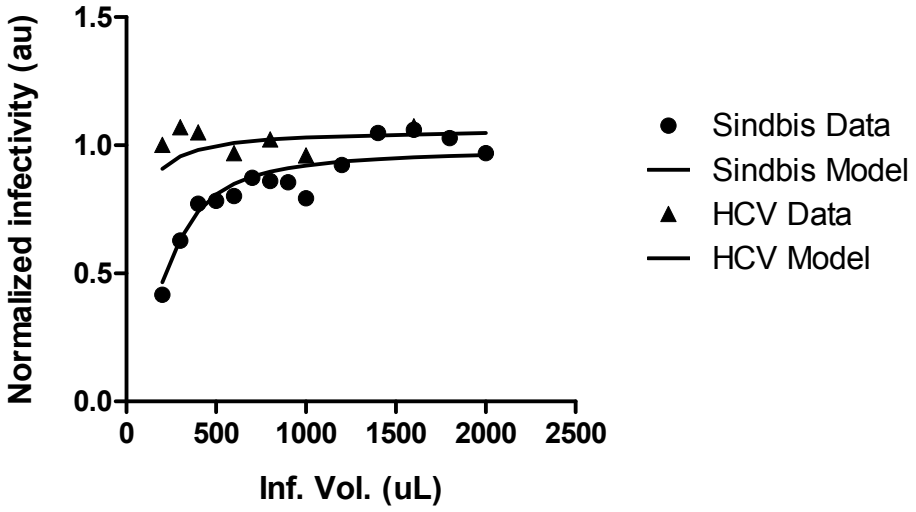


Figure 3-8. Constant concentration infection volume curve

Effect of varying infection volume for a fixed virus concentration. HCV and Sindbis virus are compared, overlaid with predicted curves from the numerical model. Means of triplicate wells from one experiment are shown.

of infection events. So for example, doubling the number of cells, which halves the MOI, does not affect the total number of infections but spreads the infection events over a greater number of cells. The dependence of infection events on cell number is shown for various values of k_I. For very large values of k_I, the assumption of cell number independence is in fact valid over a wide range, and the concept of MOI is applicable. For Sindbis, the assumption is approximately true except at the low cell numbers, which correspond to a confluence of less than 10% (Figure 3-9).

At low values of k_I, as with HCV, this assumption absolutely does not hold except at unattainably large cell numbers, well in excess of the maximum capacity of the cell culture vessel. Rather than being unaffected by cell number, the number of infection events is nearly proportional to cell number – doubling the cell number roughly doubles the number of infections, meaning the number of infections per cell is nearly the same, even though the MOI is cut in half.

The applicability of MOI in the context of HCV is therefore quite limited. In order to use MOI as relative measure, infection volume, cell number, and well format must be constant over the compared conditions, in which case there is no advantage to using MOI over either virus concentration or dose. For MOI to be used as an absolute measure – for example, to predict the percentage of cells infected with a specific dose or the number of doubly-infected cells according to the Poisson distribution – infection

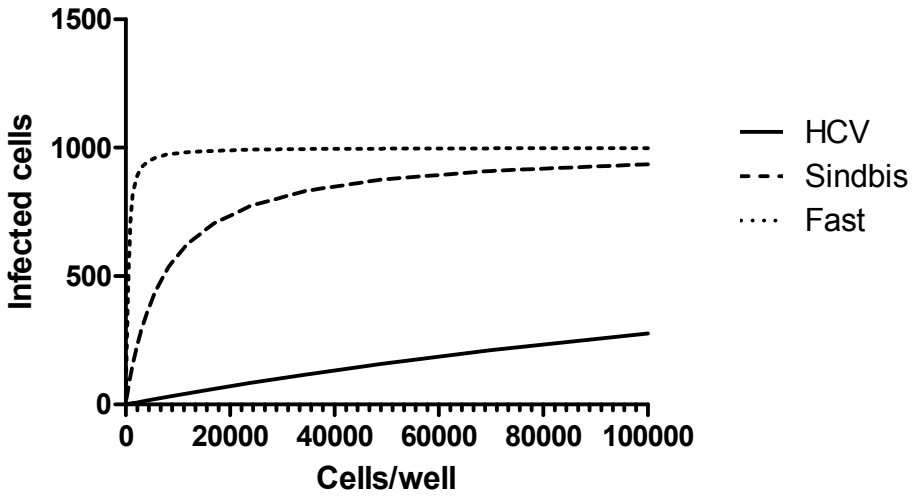


Figure 3-9. Effect of cell number for three values of K_I

Simulated infections for 1000 infectious particles as cell number is varied. Each curve represents a different value of k_I, with HCV $K_I = 8\times10^{-11}$, Sindbis $k_I = 3\times10^{-9}$, and a "fast" infection with $k_I = 10^{-7}$.

volume, cell number, and well format must be identical to the conditions under which the virus titers were determined. Alternatively, the mathematical model could potentially be used to correct for differences in cell number, infection volume, etc.

3.3 Discussion

In these studies, mathematical models were developed to describe different aspects of the HCV cell culture system. Through comparison with the appropriate experimental data, the models were refined and verified. Simulations based on these models provide additional insight into the behavior of HCV and allow a more detailed and quantitative understanding of the experimental systems in use. While these models are relatively simple, they have a particular value in that they conform quantitatively to experimental data. Not only does this increase confidence that the model is sound, it also allows quantitative determinations to be made, such as the relative concentration of virus stocks or the fraction of infectious virions which are lost to thermal degradation. One practical application of such detailed information is to optimize protocols for virus stock preparation. The development of these models also imposes a certain level of rigor on how experimental parameters are defined and can reveal issues with how those parameters are used, such as the notion of infectivity and how the idiosyncrasies of the HCV infection process create issues with the traditional use of concepts like titers, virus dose, and MOI.

These implications and applications of the numerical models derive from a completed set of simulations leading to certain conclusions. It is also interesting to consider a more interactive modality of use, in which these models are encapsulated in a more accessible user interface and become tools for design and analysis of experiments. For example, given that MOI in its traditional sense is not applicable to HCV, there would be some practical value in developing a software tool to determine the equivalent MOI for an arbitrarily chosen set of experimental parameters (virus dose, cell number, well format, infection volume, infection time, etc.), or to calculate the virus dose required to achieve a desired MOI for a set of specified experimental conditions. This is, in fact, the only general way to infect cells at a predetermined multiplicity. At a broader level, it is worth considering how an increased integration of mathematical models and other software tools might prove beneficial to the practice of experimental biology.

In terms of the specific biological findings, the impact of the entry rate constant k_I on virus titers offers some areas for further study. Having seen differences in k_I between HCV and Sindbis, it remains to be seen whether k_I will vary between different strains of HCV and whether this can account for differences in infectious virus production or specific infectivity. This possibility is already being explored as a factor in the elevated infectivity of J6JFH-Clone2, a highly infectious cell culture adapted strain with multiple mutations. In a related vein, determining the correlates of k_I would be an interesting study as well. There are many possible reasons HCV entry proceeds so slowly, whether

physical properties such as buoyant density or biochemical determinants like glycoprotein sequence or target cell properties like receptor concentration. Some of these factors have been modeled in great detail for receptor mediated viral entry in general and include more complex scenarios such as sequential engagement of receptors, which would be highly relevant for HCV entry *(104)*. Other studies have focused on one particular aspect of entry, such as the number of E2:CD81 complexes required for entry *(103)*. Understanding the factors which determine k_I may also open up new targets for drug development, by revealing exploitable bottlenecks in the HCV entry process as well as suggest ways to improve the efficiency of current cell culture systems.

Chapter 4. NS5A and elements of HCV regulation

Chapter 4 NS5A and elements of HCV regulation

As discussed in the introduction, NS5A has been long seen as a key regulator of the various HCV life cycle activities; however, the mechanisms by which it accomplishes this task remain largely undefined. Phosphorylation is certainly involved, and domains II and III have been shown to profoundly affect both RNA replication and virus production. We performed a series of experiments which shed additional light on the regulatory role of NS5A, particularly in the balance of RNA replication and virus production.

4.1 The 378 insertion site and the related domain II deletion

As previously reported, insertions into NS5A domain III are well tolerated in replicons, which maintain efficient RNA replication with GFP inserted in-frame near the C-terminus of NS5A *(105)*. The effect on virus production in a full-length construct, however, is profound, with 70-fold reductions in virus titer reported *(106)*. We inserted the fluorescent protein Ypet *(107)* between amino acids 378 and 379 in NS5A (to be referred to as the 378 insertion site) in a Jc1 background to produce Jc1(378-Ypet), also referred to as 378-Ypet. Consistent with previous reports, RNA replication was robust, but infectivity was severely impaired (Figure 4-1). Serial passaging of cell supernatants following electroporation generated revertants with restored infectivity. Sequencing of these mutants revealed deletions in domain II of NS5A. A 34 amino acid deletion, dubbed Δ34, was reengineered back into the 378-Ypet background to create the Δ34-

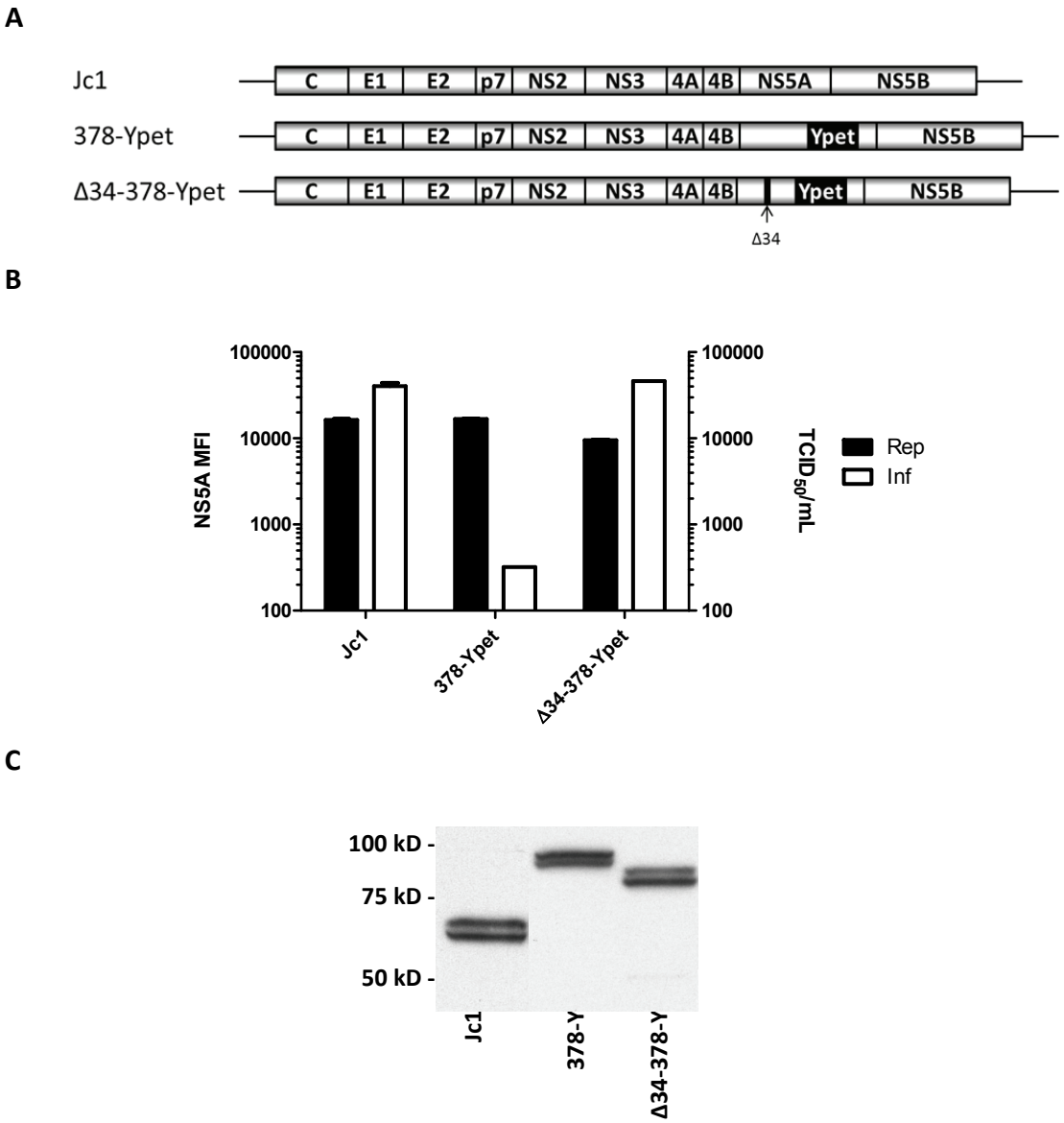


Figure 4-1. 378 Ypet insert and compensatory deletion

(A) Genome layouts for Jc1, 378-Ypet, and Δ34-378-Ypet. (B) Comparison of replication and virus production for the three constructs. Replication is assessed by FACS MFI for a fluorophore-conjugated antibody against NS5A. Infectivity (secondary Y-axis) is measured by limiting dilution assay. (C) Western blot against NS5A comparing the three constructs. Means and SEM of duplicate electroporations are shown.

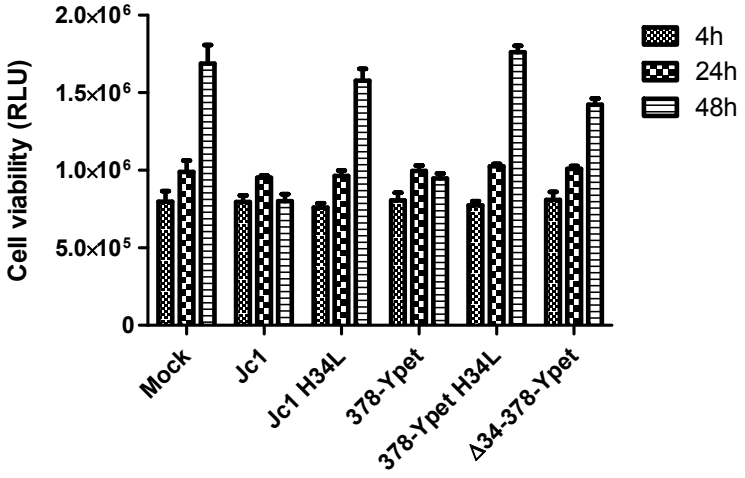
378-Ypet construct. This construct displayed close to wild-type levels of virus assembly. Interestingly, there appears to be a small reduction in replication relative to the 378-Ypet when the deletion is added. The increase in assembly and concomitant decrease in replication, confirmed by NS5A MFI in FACS analysis, suggested that these observations were linked to changes in how NS5A was regulating the relative balance of assembly and replication. The fact that the reversion came in the form of a somewhat remote deletion also seemed to hint that the decreased titers in the 378-Ypet virus were not due to the loss of a NS5A assembly function, but rather a dysregulation.

The phosphorylation status of 378-Ypet and Δ34-378-Ypet were compared by Western blot; the difference was not dramatic, but 378-Ypet showed a small bias towards the larger p58 form, while Δ34-378-Ypet had a more pronounced p56 band. This is in contrast both to the serine 457 studies *(40)* and to studies involving replication enhancing adaptive mutations, which both showed a correlation between p58 levels and virus production *(66)*.

4.1.1 Decrease in virus production not caused by increased cytotoxicity

In the initial characterization of the virus, we observed higher degree of cell death in the 378-Ypet virus compared with the Δ34-378-Ypet, which raised the question as to whether the lower titers from 378-Ypet might be a kind of cytopathic effect. In order to test this possibility, we engineered an NS5B mutation, H34L, into the Jc1 and 378-Ypet

A



B

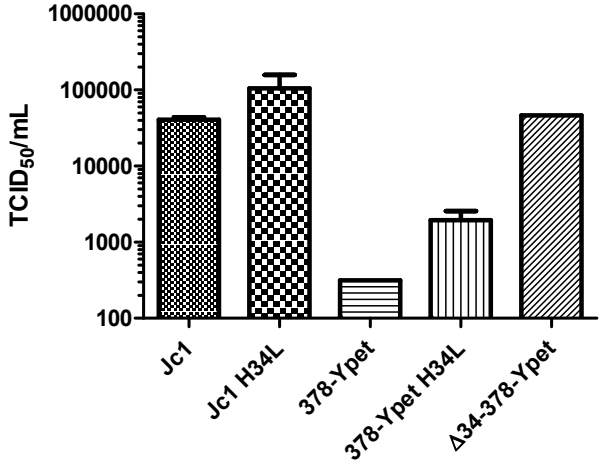


Figure 4-2. H34L mutant and cell viability

(A) Cell viability of HCV RNA-transfected cells as a function of time post-electroporation, with Jc1 and 378-Ypet both showing impaired proliferation at 48 h PE, both rescued by the NS5B H34L mutation. (B) Viral titers at 48 h PE, measured by limiting dilution assay. Means and SEM of duplicate electroporations are shown.

backgrounds. This mutation, originally isolated as an adaptive mutation in JFH1 replicons *(87)*, has been shown to attenuate HCV replication to some extent *(88)*, resulting in reduced cytopathic effect. With the adaptive mutation, increased cell viability was observed in both the Jc1 and 378-Ypet following transfection (Figure 4-2). Virus production was also improved in both cases; however the magnitude of the differences was not enough to explain the phenotypes. In addition, the 378-Ypet and Δ34-378-Ypet modifications were also later engineered into a bicistronic reporter construct, which also displays a moderated replication phenotype and good cell viability. In this background, the differences in virus production were the same, confirming that these differences are not a result of differences in cytotoxicity.

4.1.2 378 Ypet insertion and Δ34 deletion affect an early step of production

To determine whether the lower virus titer with the 378 Ypet insertion was due to a block in the maturation/secretion pathway or a defect in de novo virus assembly, we compared the extracellular titers with intracellular virus released by freeze/thaw cycles of the transfected cells. The differences in extracellular infectivity of the three constructs were closely reflected in the intracellular infectivity levels, showing that the Ypet insertion affects an early stage of virus assembly (Figure 4-3). Given the RNA binding function of NS5A, this early step could be RNA transport and packaging at the assembly site.

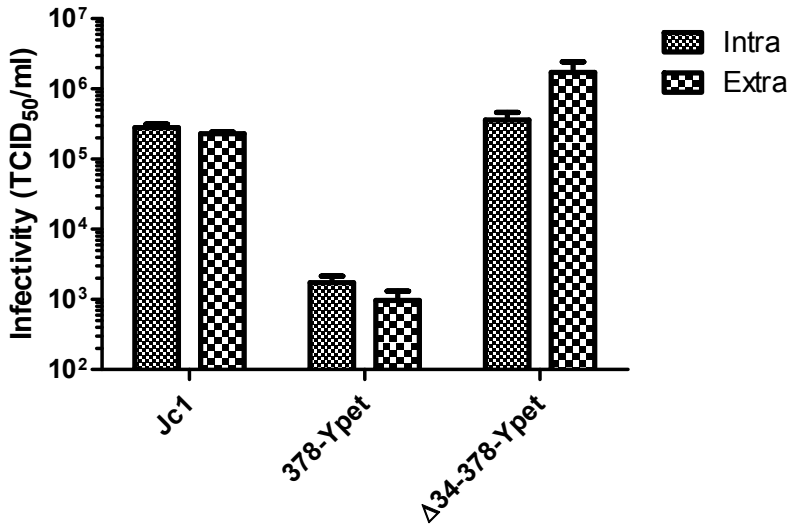


Figure 4-3. Intra- and Extracellular infectivity

Measurements of intracellular and extracellular infectivity, determined by limiting dilution assay, with intracellular infectivity obtained by multiple freeze/thaw cycles of washed cells. Means and SEM of duplicate electroporations are shown.

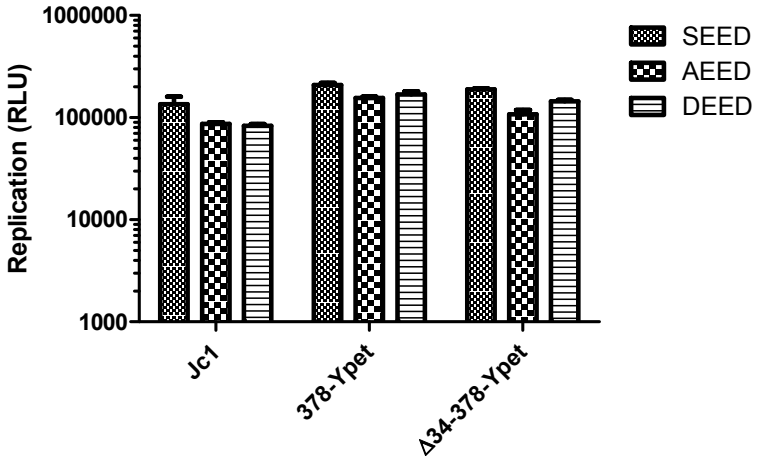
4.1.3 Independence of 378 and Δ34 phenotypes from serine 457

The loss of virus production and compensating recovery is reminiscent of the study on NS5A serine 457 *(40)*, which showed decreases in virus production when mutated which are comparable to the 378-Ypet and a nearly complete recovery of virus production when mutated to a phosphor-mimetic aspartic acid. To test whether serine 457 might play a role the phenotypes of 378-Ypet and Δ34-378-Ypet, the S457A and S457D mutations were engineered into the 378-Ypet and Δ34-378-Ypet genomes. As shown in Figure 4-4, the two sets of perturbations appear to act completely independently. Regardless of whether they are in the Jc1, 378-Ypet, or Δ34-378-Ypet background, the serine 457 mutations have the same effects on virus production.

4.1.4 The effect of the domain II deletion is distributed across its span

We constructed a collection of mutants in domain II mutants to further dissect interplay of the domain II deletion and the 378 Ypet insertion, as outlined in Table 4-1. They include a set of sub-deletions of the Δ34 deletion along with proline to alanine substitutions in that region. These prolines have been proposed as potential binding and/or isomerization targets of cellular proline isomerase cyclophilin A (CypA) *(82)*. Also included was a deletion of a reported CypA binding motif *(82)* and mutations of the prolines contained therein.

A



B

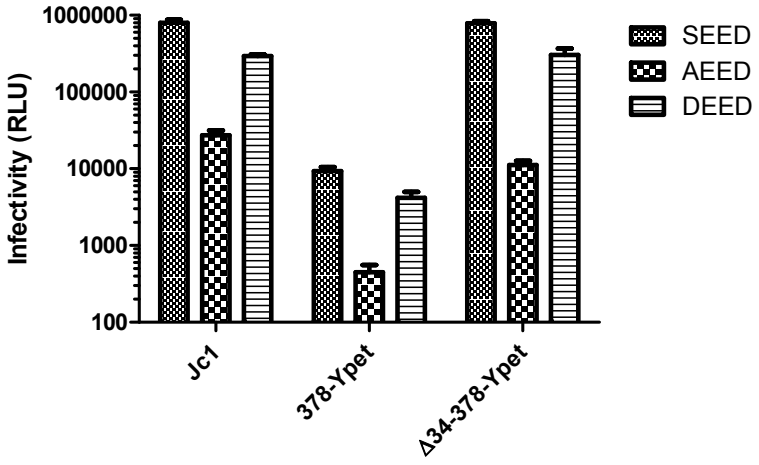


Figure 4-4. Effect of serine 457 mutations

(A) Effect of serine 457 mutations on replication in the backgrounds of 3 different constructs, with SEED being the native motif and AEED and DEED the respective substitutions. (B) The effect of the same mutations on virus production, showing identical effects in all three backgrounds. Means and SEM of duplicate electroporations are shown.

Growth curves and the relative replication and viral titers are shown in Figure 4-5. Replication, with exceptions, was not highly affected by these mutations. The exceptions are both alterations of the CypA binding site ^{310}PAWARP^{315}, which result in a severe decrease in replication. The large boost in infectious virus output from the Δ34 deletion could not be mapped to any particular subdivision; the effect seemed to be distributed across the span of the deletion. The second half of the deletion showed a stronger contribution, an effect that continued beyond the C-terminus of the Δ34 deletion out to amino acid 306.

The effect of the five combined proline to alanine substitutions within the deletion region showed just a two-fold decrease in replication and a minimal increase in virus production. This small effect may, however, be the result of two or more opposing effects. The first two proline mutations together decreased the yield of infectious virions below the limit of detection; however, due to the low virus production of 378-Ypet, titers below a 7-fold decrease could not be measured. The other three proline residues together increased virus production 4-fold.

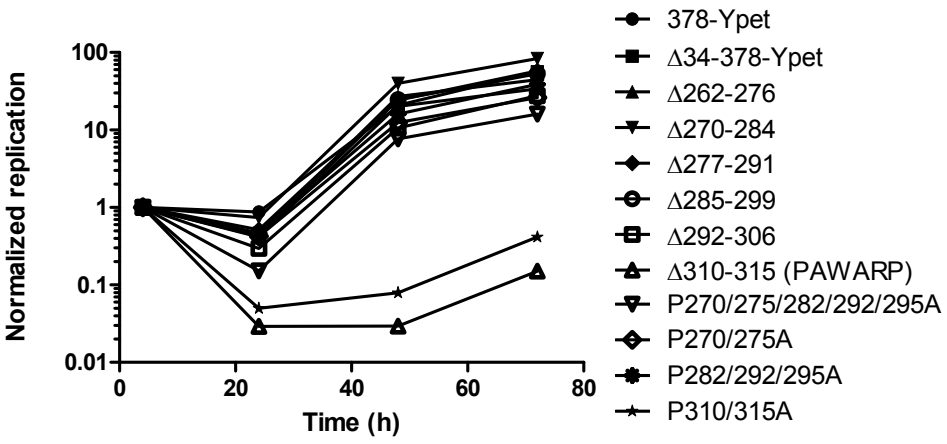
4.2 CsA differentially effects RNA replication and virus assembly

Because of the well-documented importance of CypA for HCV the HCV life cycle and because CypA has been shown to interact with both domains II and III of NS5A *(78, 80)*, this interaction was seen as potentially relevant. We therefore decided to test the effect

Table 4-1. Domain II mutagenesis constructs

Construct	Sequence
378-Ypet	LMEGGVAQTEPESRVPVLDFLEPMAEEESDLEPSIPSECMLPRSGFPRALPAWARPDYNPP
Δ262-295 (Δ34)	LM----------SECMLPRSGFPRALPAWARPDYNPP
Δ262-276	LM----------LDFLEPMAEEESDLEPSIPSECMLPRSGFPRALPAWARPDYNPP
Δ270-284	LMEGGVAQTE----------EEESDLEPSIPSECMLPRSGFPRALPAWARPDYNPP
Δ277-291	LMEGGVAQTEPESRVPV----------PSIPSECMLPRSGFPRALPAWARPDYNPP
Δ285-291	LMEGGVAQTEPESRVPVLDFLEPMA----------LPRSGFPRALPAWARPDYNPP
Δ292-306	LMEGGVAQTEPESRVPVLDFLEPMAEEESDLE----------RALPAWARPDYNPP
Δ310-315	LMEGGVAQTEPESRVPVLDFLEPMAEEESDLEPSIPSECMLPRSGFPRAL-----DYNPP
P270/275/282/292/295A	LMEGGVAQTE**A**ESRV**A**VLDFLE**A**MAEEESDLE**A**SIPSECML**A**RSGFPRALPAWARPDYNPP
P270/275A	LMEGGVAQTE**A**ESRV**A**VLDFLEPMAEEESDLEPSIPSECMLPRSGFPRALPAWARPDYNPP
P282/292/295A	LMEGGVAQTEPESRVPVLDFLE**A**MAEEESDLE**A**SIASECML**A**RSGFPRALPAWARPDYNPP
P310/315A	LMEGGVAQTEPESRVPVLDFLEPMAEEESDLEPSIPSECMLPRSGFPRAL**A**AWAR**A**DYNPP

A



B

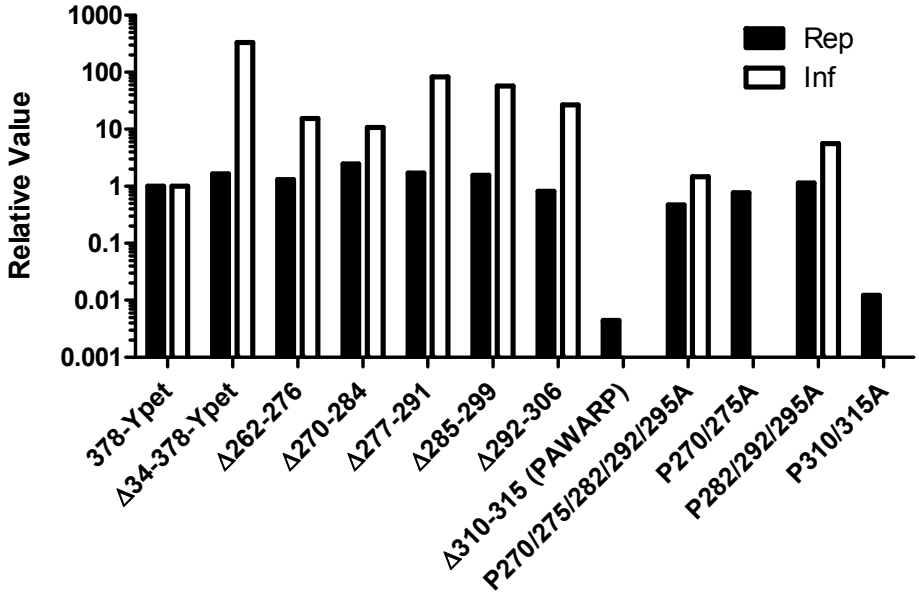


Figure 4-5. Domain II mutagenesis

(A) Growth kinetics of the 12 different constructs, with replication determined by luciferase expression. (B) 48 h replication and viral titers normalized to the parental 378-Ypet. Infectivity was undetectable for two constructs – at least 7-fold reduced.

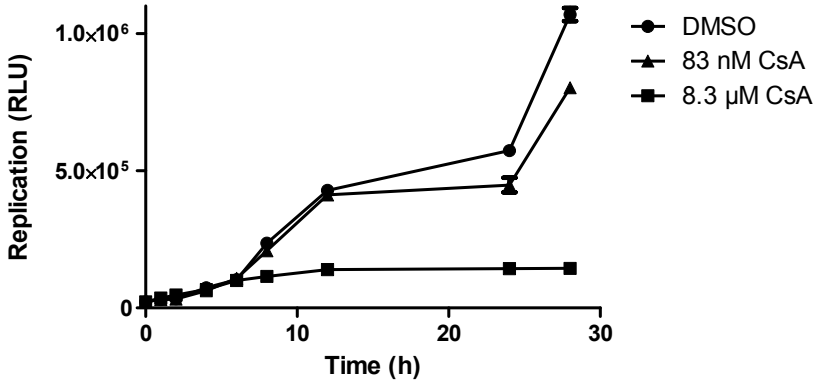
of cyclosporine A (CsA), an inhibitor of CypA, on virus production. CsA is a known inhibitor of HCV RNA replication; in order to look for a possible second effect on virus assembly, we delayed administration of the drug until 24 h after transfection, to allow some time for replication to be established. Two concentrations were tested, 8.3 µM and 83 nM[1], and both replication and virus assembly were then tracked periodically following drug addition.

Because CsA has been shown to inhibit an interaction between CypA and NS2 as well, we chose to perform these experiments using a bicistronic construct containing an IRES between NS2 and NS3, as well as a *Gaussia* luciferase reporter, diagrammed in Figure 4-6a. Separating NS2 and NS3 with an IRES has been shown to modulate the CsA sensitivity, leading to the hypothesis that CypA may play a role in NS2-NS3 cleavage *(73)*. By using a construct in which NS2 and NS3 are already separated, we hoped to bypass any effects of CsA on NS2. Unless otherwise specified, all of the remaining experiments were done in this bicistronic background. Replication was represented by luciferase activity in cell lysates, while intracellular virus released by freeze/thaw cycles was passed on to naïve cells to assay infectivity. The use of intracellular virus allowed CsA in the supernatants to be diluted away, so it would not interfere with the infectivity assay.

[1] CsA concentrations were initially chosen in units of µg/ml, and then converted to molarity at the time of writing for consistency. 8.3 µM = 10 µg/ml, 2.5 µM = 3 µg/ml.

A

B



C

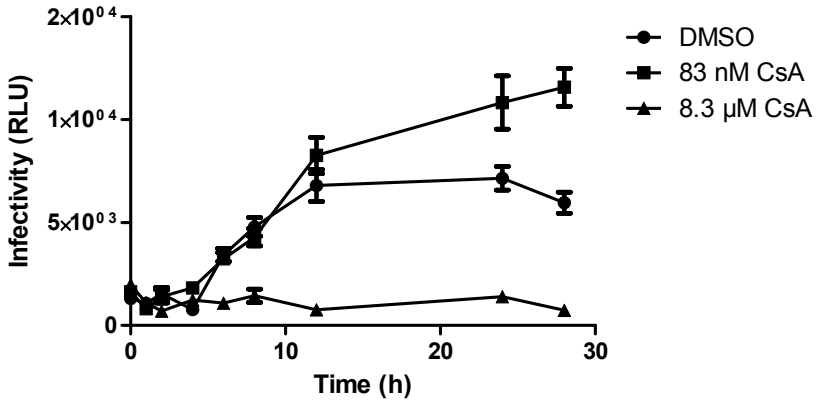


Figure 4-6. Differential effect of CsA on replication and infectivity

(A) The bicistronic reporter virus with NS2 and NS3 separated and expressing *Gaussia* luciferase. (B) Kinetic response to CsA as a function of time post-addition for three different concentrations. (C) Intracellular infectivity as a function of time post-addition of CsA. Note that infectivity with 83 nM CsA is higher than when untreated. Means and SEM of replicate wells are shown.

The resulting drug response time course yielded several interesting observations. As expected, replication was significantly inhibited at the high concentration of CsA and less so at the lower concentration. Surprisingly, however, even at the high concentration, 40-fold higher than the EC_{50}, replication levels did not display any net decrease at all after administration – only further increase in replication was halted.

Another interesting observation, which is followed up on in subsequent experiments, is that though the low concentration of CsA is slightly inhibitory for replication, virus yield is modestly enhanced. This simultaneous inhibition of replication and stimulation of assembly again suggests a shift in some regulatory balance between these life cycle activities, quite possibly through NS5A. CsA treatment did not rescue the virus assembly defect of 378-Ypet, however, which might have been expected from the proline mutagenesis results, but showed the same modest enhancement of assembly coupled with moderate inhibition of replication as seen with the Jc1 genome at the low concentration of CsA (data not shown).

4.3 Various compounds show diverse effects on HCV

In order to gain a clearer picture of the effects of CsA and place it in a broader context for comparison, dose-response curves were measured for a set of compounds with potential activity against HCV, including CsA, 2'CMA, danoprevir, BMS-790052, and H479 (See Table 4-2). The activity of the compounds was assessed on HCV

Table 4-2. List of inhibitors

Inhibitor	Target	Mechanism
Cyclosporine A (CsA)	Cellular cyclophilins	Binding of catalytic pocket
H479	Cellular casein kinase I (CKI)	ATP-competitive imidazole inhibitor
ITMN-191 (Danoprevir)	HCV NS3-4A protease	Peptiomimetic inhibitor
BMS-790052	HCV NS5A domain I	Unknown
2'-C-methyl-adenosine (2'CMA)	HCV NS5B	Chain-terminating nucleoside analogue

electroporated cells at two different timing intervals, based on initial experiments – 4 h PE administration and 48 h PE harvest (4-48 h) and 48 h PE administration with 72 h PE harvest (48-72 h). The dose-response curves are shown in Figure 4-7, and show a number of notable features. As noted, CsA at low doses has a stimulatory effect on infectivity at both timings. Interestingly, 2'CMA has a similar effect at the 4-48 h schedule, an effect which is completely absent in 48-72 h, unlike CsA. BMS-790052 and danoprevir are both monotonically inhibitory, but share a common effect at the later time interval in that both inhibit virus production more strongly than replication.

H479 stands out as a potential counterpart to CsA – this CKIα kinase inhibitor appears to stimulate replication and impair virus production, particularly when added after the establishment of replication. This effect, along with its previously documented effect on NS5A, made it a particularly interesting compound for further study of HCV regulation through the NS5A protein.

The Hill slope for compounds that produced monotonic sigmoidal inhibition curves also varied in interesting ways. At the early interval, BMS-790052 produced a steep inhibition with a Hill slope of approximately 2 for both replication and virus production. At the later time, the effect on infectivity remained steep, but the effect on replication showed a shallower Hill slope of close to 1. Although the steepness of a Hill slope does not lend itself to any conclusive interpretation, the suggestion which has been made that BMS-790052 may inhibit two distinct activities of NS5A *(64)* certainly comes to

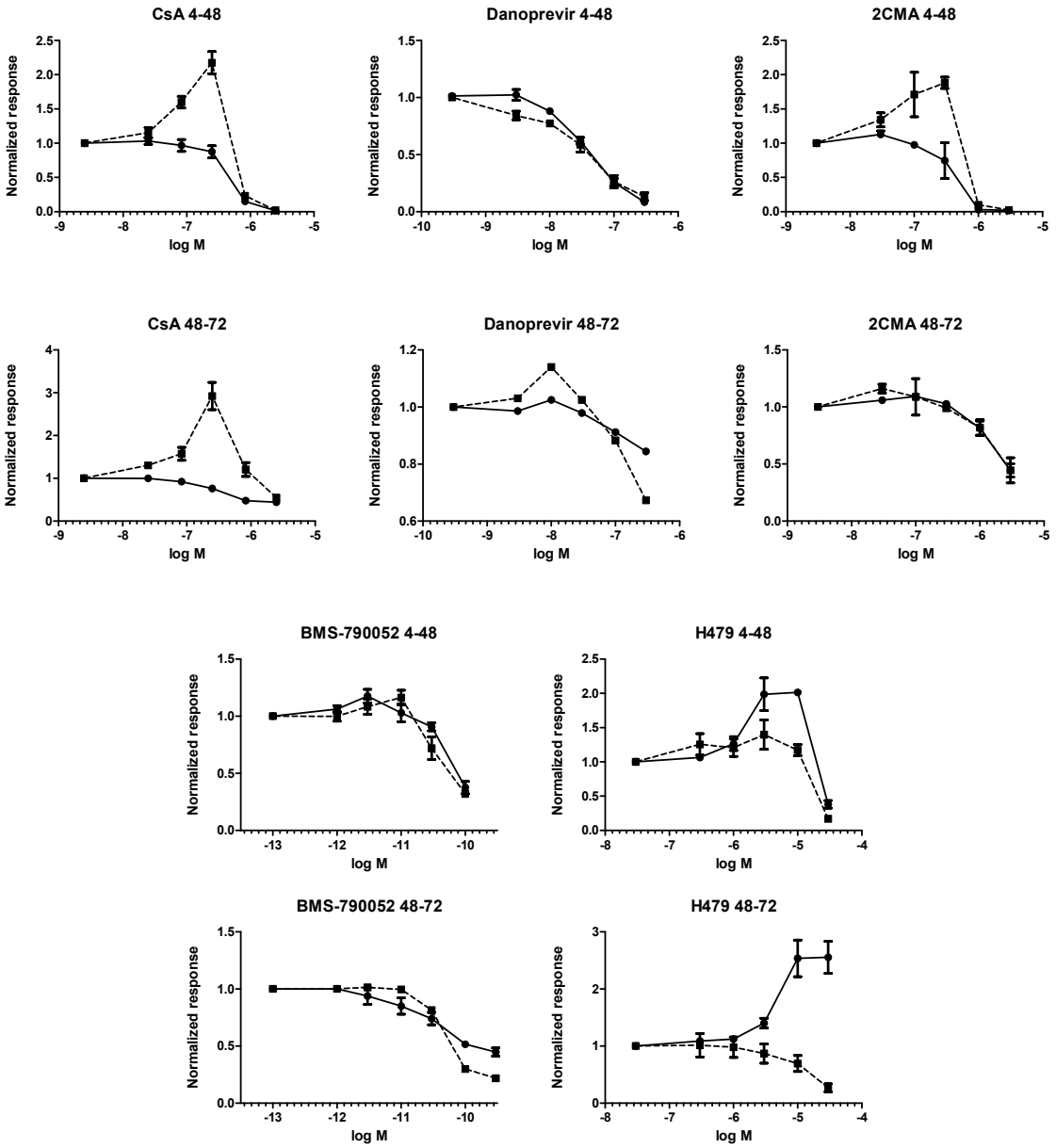


Figure 4-7. Dose-response curves for various compounds

Effect of various compounds on replication and infectivity over a range of doses. In each plot the solid curve represents replication and the dotted curve represents infectivity. Timing intervals are given by time of addition and time of harvest (ex. 4-48 means 4 h addition and 48 h harvest). Means and SEM of at least two independent experiments are shown.

mind. Perhaps at the early time points both activities are highly critical for both replication and assembly, leading to steeper inhibition curves for BMS-790052, whereas at later times one of the two activities becomes less crucial for replication, though still required for assembly.

These experiments also showed a dependence of the dose-response curve on timing. The curves shift laterally (in concentration), and the level of maximum inhibition also varies depending on time of addition. CsA, for example, shows a maximum inhibition of 99% at the earlier interval, but levels off at 58% inhibition at the same concentration if administered and harvested later. This result is consistent with the earlier observation that CsA does not efficiently inhibit pre-established replication.

Overall, the dose-response curves reveal a complexity of behavior beyond what can be captured in a single metric like EC50, complexity which reflects details about the HCV life cycle and the mechanism by which these inhibitors interact with HCV. Some of this complexity is explored further in the following experiments; however, much of it is left as promising avenues for future work.

4.3.1 CsA-mediated inhibition is phenotypically distinct from 2'CMA inhibition

The differences in inhibitory properties of CsA and 2'CMA was examined in further detail by measuring the kinetic response of replication to both drugs. Cells were transfected with HCV RNA and then treated with the inhibitors at a high concentration

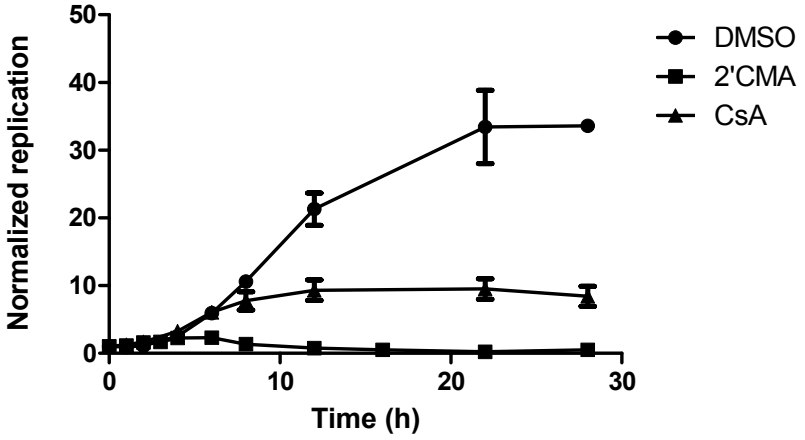


Figure 4-8. Kinetic response to CsA and 2'CMA compared

Replication as a function of time after drug addition, normalized to replication at the time of addition. HCV-transfected cells were treated with 10 µM 2'CMA, 8.3 µM CsA, or DMSO at 24 h PE. Replication was monitored by luciferase expression. Means and SEM of two independent experiments are shown.

(~30 times the IC_{50}) 24 h later. Replication was monitored over the course of 28 h post-addition to generate the time course shown in Figure 4-8. The curves show a clear difference in the mode of inhibition of the two compounds. Whereas CsA only arrests further increase in replication levels, 2'CMA effectively eliminates ongoing replication. While a conclusive explanation of this would require further experimentation, one possible explanation would be that CsA blocks formation of new replication complexes, but lacks activity against ongoing replication within those that were formed before addition of the drug. This is in contrast to treatment with 2'CMA, which directly interferes with the enzymatic activity of the NS5B RNA polymerase.

4.4 Factors affecting the biphasic response to CsA

4.4.1 Assembly enhancement is sensitive to timing of drug addition and harvest

The effect of CsA on HCV virus production is biphasic, with low doses increasing infectious virion yield while high doses decrease it. The effect appears to depend on timing, with a stronger enhancement for 48-72 h treatment than 4-48 h. To further study the role of timing, several combinations of drug addition times and harvest times were tested, as shown in Figure 4-9a. In the panels which show the stimulation the magnitude of the effect follows a trend of increased effect with increased time of treatment and later harvest.

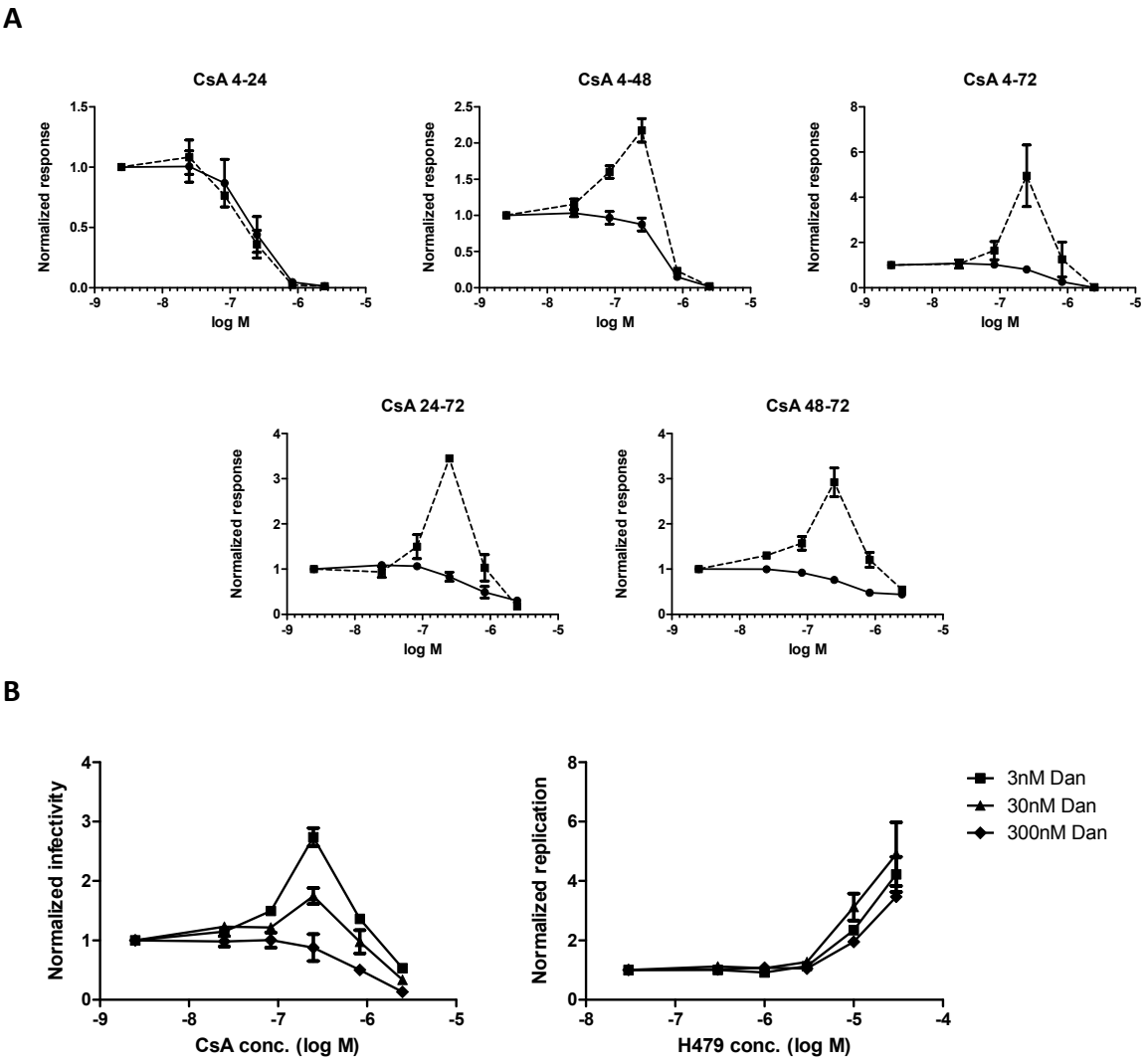


Figure 4-9. Effect of timing and replication levels

(A) The effect of drug addition time and harvest time on CsA dose-response curves. Solid lines = replication, dotted line = infectivity. (B) Pre-treatment with the protease inhibitor danoprevir. Transfected cells were pre-treated from 4 h with danoprevir and then CsA or H479 were added for 48-72 h. Infectivity curves are shown for CsA, replication curves for H479. Legend applies to both plots. Means and SEM of at least two independent experiments are shown.

The most immediately apparent effect however is the fact that there is no stimulation of virus assembly at the 24 h harvest; in fact, the same dose which gives the most stimulation at other times inhibits infectivity by over 50% at 24 h, which is very much in line with the effect of CsA on replication directly. Given the dependence of assembly on RNA replication, the simplest explanation is that at 24h CsA has no direct effect on virus production, with the stimulatory effect asserting only at later time points. This could be due to a number of reasons. One possibility is that cells simply need to be exposed to CsA for a longer duration for the effect to be manifested. However, the clear effect of the 48-72 h timing argues against this, as the treatment duration is essentially the same as the 4-24 h timing. It seems more likely then that the stimulatory effect only manifests after certain conditions have been met as the HCV replication cycle proceeds. If CypA is indeed negatively regulating virus assembly, it may be that this negative feedback does not come into play until, for example, the overall replication levels have crossed some threshold.

4.4.2 Assembly enhancement is dependent on overall replication level

To test this possibility, HCV electroporated cells were pre-treated with varying amounts of danoprevir for the first 48 h before adding CsA, with the intention of moderating replication to a lower level and gauging the impact on CsA-mediated stimulation of infectivity. The results, shown in Figure 4-9b, support the hypothesis that regulation of virus production by CypA only takes effect as the replication levels exceed some threshold value. In contrast, the stimulation by H479 is not affected by pre-

treatment of danoprevir in the same way, and maintains the same relative enhancement even as the overall levels of replication are brought down by higher doses of the protease inhibitor.

4.4.3 Stimulation of intracellular infectivity is reflected in extracellular titers

To avoid the carryover of inhibitors in the dose-response experiments interfering with the downstream measurements of infectivity, intracellular infectivity was measured to generate the response curves for infectivity rather than assaying the supernatants directly. Therefore, it would be possible that an observed increase in infectivity is a consequence of blocks in late stages of virus maturation or secretion leading to an accumulation of intracellular virus rather than an increase in assembly. Indeed it has been reported that inhibition of cyclophilins by NIM811, a non-immunosuppressive CsA analog, alters lipid trafficking and blocks secretion of HCV *(108)*. However, this effect was only observed as a decrease in extracellular infectivity and not accompanied by any increase in intracellular infectivity. To clarify the nature of the infectivity enhancement, supernatants were titered by limiting dilution assay, which would dilute away any carryover of inhibitor. The extracellular infectivity response curve, shown in Figure 4-10, closely resembles the intracellular infectivity response curve, indicating that the measured enhancement reflects an increase in assembly rather than a block in secretion.

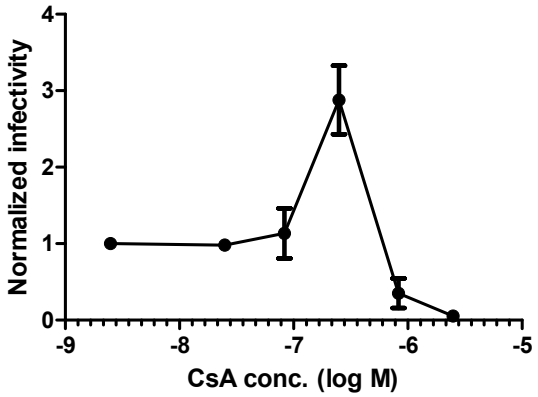


Figure 4-10. Extracellular infectivity dose-response curve

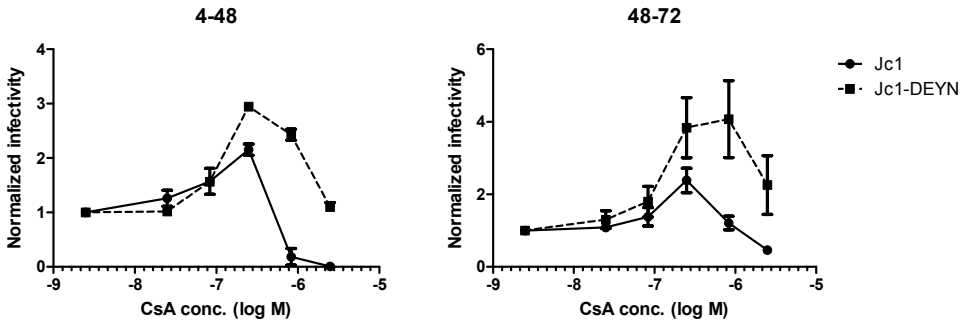
Effect of CsA treatment on extracellular infectivity, measured by limiting dilution assay. Transfected cells were treated at 4 h PE and supernatants were harvested at 72 h PE for titering. Means and SEM of two independent experiments are shown.

4.4.4 CsA resistance mutations in NS5A alter dose response

Mutations D316E and Y317N of NS5A conferring resistance to CsA have been published *(109)*. These mutations were engineered into the Jc1(NS2-IRES-nsGluc2AUbi) and compared with the unaltered constructs for their CsA dose response curve. Like the wild-type sequences, the DEYN mutant constructs showed a biphasic response curve in virus production, with stimulation at low doses and inhibition at higher doses (Figure 4-11). The magnitude of the effect was larger both in 4-48 h and 48-72 h treatments. The shape of the curve also suggested a shift in effective concentration.

A more detailed examination of the peak indeed shows a small but significant shift in both the concentration of maximum effect, from 250 nM to 500 nM, and the EC50 of the first phase, which shifts from 150 nM to 340 nM. The shift of peak could have been attributable to an increased resistance to the anti-replication effect of CsA allowing a stronger effect at the higher concentrations – in effect, a non-uniform vertical shift resulting in an apparent horizontal shift. However, the fact that the stimulatory phase of the curve begins its upward rise at a higher concentration supports the notion that the curve has really shifted horizontally. It seems probable then that the altered sensitivity of virus production to CsA is due to the fact that DEYN mutations directly perturb the regulation of HCV production rather than a side effect of improved HCV replication under CsA treatment.

A



B

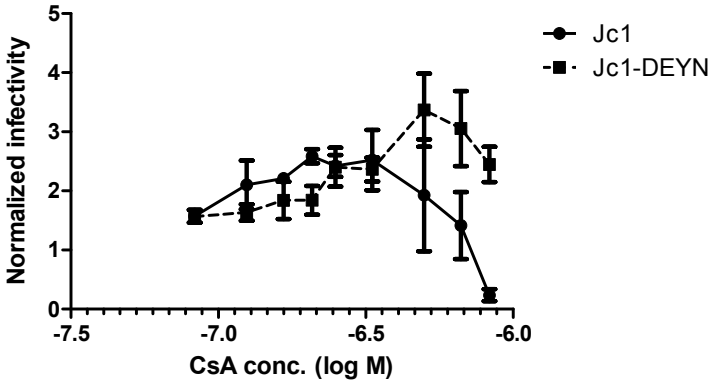


Figure 4-11. Dose-response for DEYN mutation

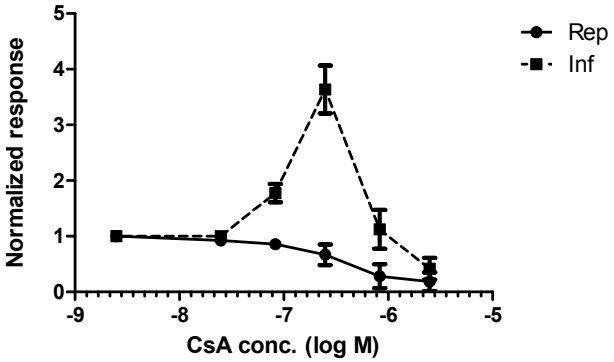
(A) Infectivity dose-response curves to CsA for Jc1 and the CsA resistance mutant D316E Y317N (DEYN). Two different drug timing intervals are shown. In both plots the dotted curve shows the DEYN construct. (B) Detailed infectivity dose-response to CsA mapping out the peak for the 4-48 h CsA treatment. Means and SEM of two independent experiments are shown.

4.4.5 CsA-mediated stimulation of assembly with different NS2 configurations

In order to bypass the effect of CsA on NS2, dose response curves were measured in a bicistronic construct expressing NS3-NS5B from an EMCV IRES following NS2. To confirm that the enhancement of infectivity from CsA treatment was not dependent on this abnormal configuration of NS2 and NS3, a monocistronic Jc1 reporter virus, Jc1(p7nsGLuc2A), was also tested. This construct preserves the natural NS2/NS3 junction. The dose-response curve shows that the stimulation is the same regardless whether NS2 and NS3 are expressed together or separately (Figure 4-12a).

We also considered the possibility that absolute levels of virus assembly might affect the nature of CsA-mediated assembly enhancement. To test this, dose response curves were generated from three different cell culture strains: JFH1, J6/JFH1, and Jc1 (Figure 4-12b). These three strains replicate with similar efficiency but differ in virus production ability, with JFH1 having the lowest and Jc1 the highest *(50, 88, 110)*. J6/JFH showed a very similar response profile to Jc1, perhaps shifted slightly to the right. JFH1, on the other hand, showed a very different profile. The response in infectivity was not biphasic in the case in the case of JFH1, but monotonically increasing with CsA concentration. J6/JFH and Jc1 are both chimeras of the J6 isolate from the N-terminus and the JFH1 isolate from the C-terminus with a breakpoint between NS2 and NS3 for J6/JFH and within NS2 for Jc1. Therefore the two constructs are identical in all genes except for part of NS2. By contrast, the JFH1 strain is a single isolate, which therefore differs from both

A



B

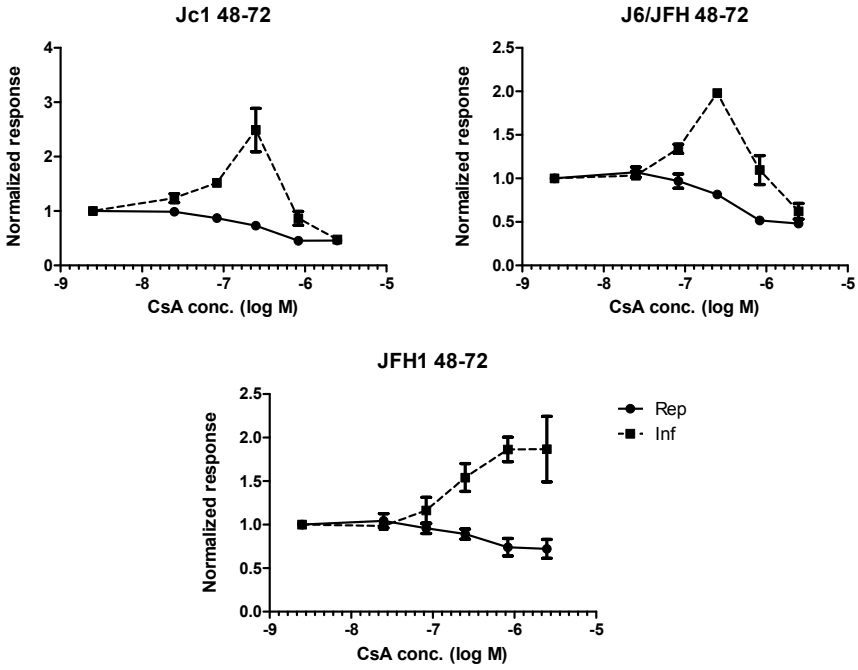


Figure 4-12. Alternate NS2 configurations

(A) Dose-response curve for a monocistronic virus construct. The 48-72 h CsA dose-response for a native NS2-NS3 configuration shows the same enhancement of infectivity. (B) Dose-response curves for different genotype 2A strains, all showing infectivity enhancements from CsA.

chimeric strains in all the structural proteins, p7, and NS2. There are therefore many possible causes for the observed difference in the JFH1 response profile, including direct causes such as differences in NS2, which has been implicated in coordinating virus assembly *(45, 46)* and is a target of CypA, or in any of the structural proteins, as well as indirect causes such as the large differences in overall virus production possibly stemming from differences in regulatory setpoints or feedback loops. However, the overall magnitude of the effect was roughly the same across the three strains; significant differences in virus production did not dramatically enhance or abolish the stimulatory effect of CsA on virus assembly.

4.5 CypA fused to NS5A negatively modulates replication and assembly

In order to test the effect of CypA directly on NS5A, the CypA gene was cloned directly into the 378 insertion site in NS5A. Clearly, whether such a fusion would be functional at all is impossible to know beforehand. However, the small, globular nature of CypA, not entirely unlike Ypet in shape and size, and the previously determined tolerability of fusion inserts into the 378 site provided enough encouragement to construct and test this configuration. The structures of GFP and CypA are compared in Figure 4-13a. Notably, the N- and C-termini of the two proteins are similarly configured.

In addition to the wild-type CypA gene, an active site mutant, H126Q, was also engineered. These two genes were inserted into the 378 locus in the context of both the

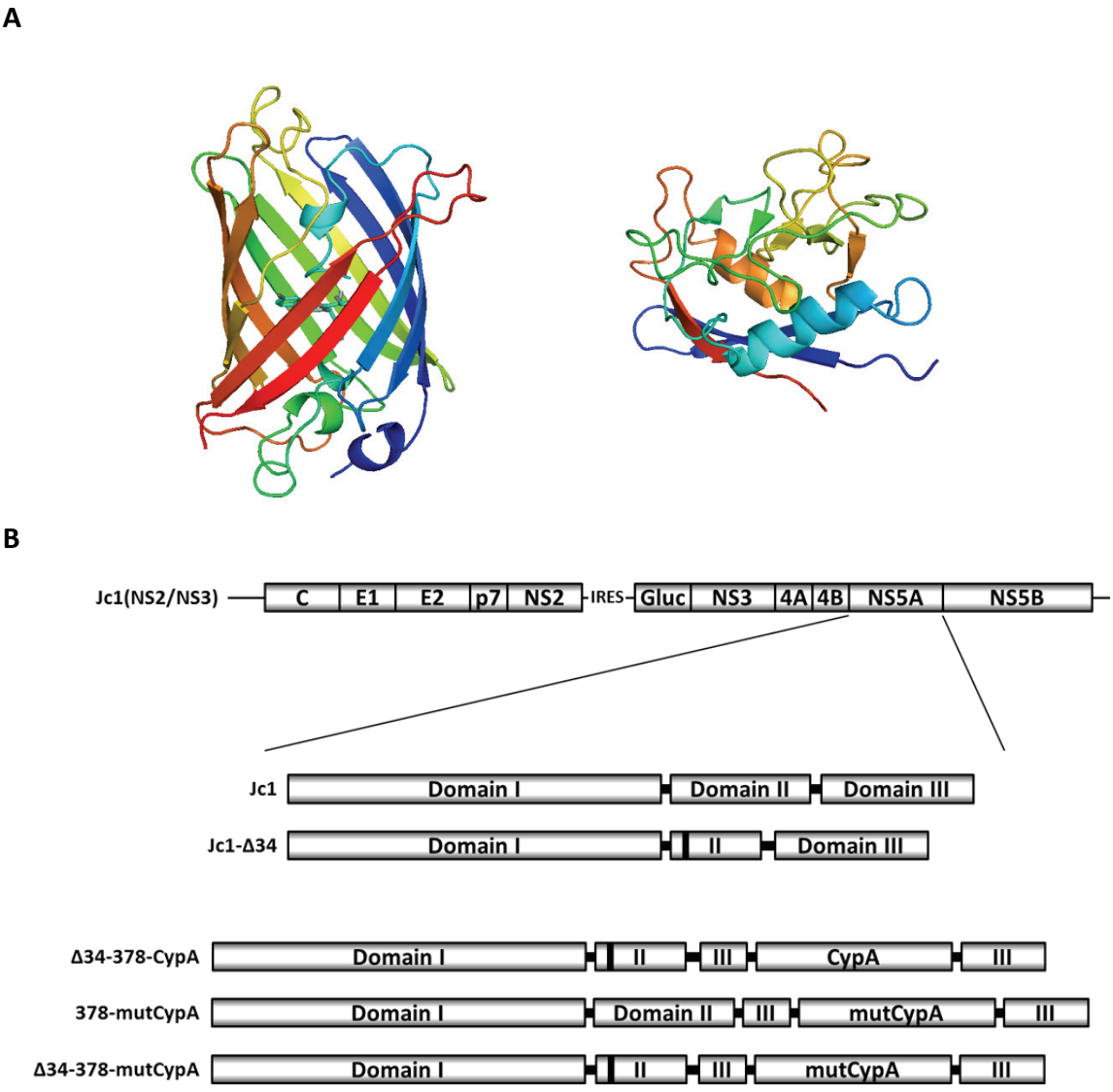


Figure 4-13. NS5A 378-CypA fusion constructs

(A) Crystal structure of Ypet on the left and CypA on the right, showing similar size and globular configuration, as well as proximal N- and C- termini. (B) Engineered NS5A constructs in the bicistronic reporter background showing the three domains of NS5A with either the Δ34 deletion in domain II, the CypA or mutCypA insert in domain III, or both.

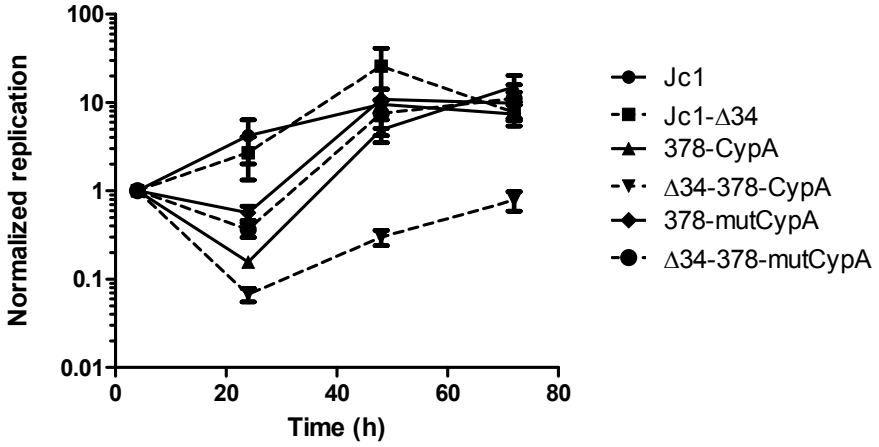
full-length NS5A and the domain II compensatory deletion, yielding a total of four constructs. As described earlier, the domain II deletion restores lost virus assembly function in the context of the Ypet insertion. To make a complete set, Jc1-Δ34 was also included (domain II deletion without any insert). All six constructs are shown in Figure 4-13b.

4.5.1 NS5A-CypA active site-dependent impairment of replication and production

The CypA fusion constructs show several interesting phenotypes. The growth curves following electroporation are shown in Figure 4-14a, normalized to the initial 4 h translation. By 24 h, differences between the constructs become apparent. Jc1 replicates the most efficiently from the outset, with Jc1-Δ34 virtually identical. All the constructs containing insertions are impaired at this point, with the differences between them systematic: the constructs with the mutated CypA replicate better, and the constructs without the domain II deletion replicate better. Δ34- 378-CypA, with both the active CypA and the domain II deletion, is the most impaired.

As replication hits the plateau around the 48 h mark, the relative ranking of replication efficiency becomes shuffled, with most of the constructs clustering around the wild-type Jc1 level of replication. The exception is Δ34-378-CypA, which continues to show a marked impairment of more than 1 log compared to all the other constructs including Δ34-378-mutCypA, which is, apart from the single-residue active site mutation of CypA, an identical genome.

A



B

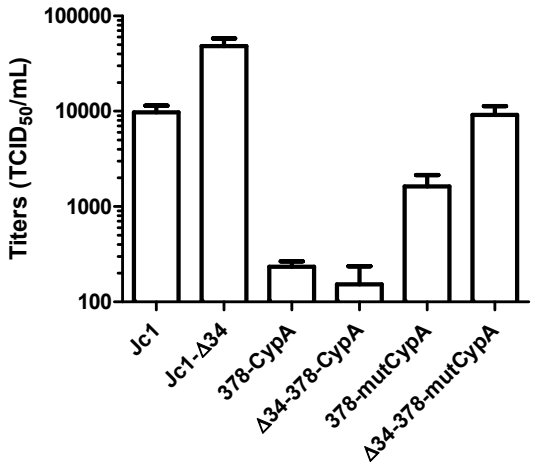


Figure 4-14. Phenotypes of 378-CypA constucts

(A) Growth curves for the six constructs following electroporation. Replication is normalized to the 4 h PE luciferase expression level. Constructs with the deletion are shown with dashed lines for readability. (B) Viral titers for the six constructs measured by limiting dilution assay 48 h PE. Means and SEM of at least two independent experiments are shown.

The virus production capacity of these constructs shows greater variation – 48 h titers are shown in Figure 4-14b. As previously noted, the domain II deletion without any insertion (Jc1-Δ34) has improved titers. Both CypA fusion constructs show major reductions in titers – close to 2 logs. In the case of Δ34-378-CypA, some of this reduction in titers can be attributed to a reduction in overall replicative potential. In both cases, mutating the active site improves titers, with a 1 log gain for 378-CypA and a 2 log gain for Δ34-378-CypA. In order to visualize the combined effects on replication and virus production, the 48 h luciferase expression and 48 h viral titers are shown side-by-side in Figure 4-15.

A simple, though admittedly naïve, measure of the relative effect on infectivity of the domain II deletion and the CypA active site mutation can be calculated by assuming infectivity scales linearly with replication. The marginal impact of a perturbation on infectivity can be estimated by normalizing to the effect of that perturbation on replication. The results of these calculations are shown in Table 4-3. The effect of the domain II deletion on replication varies widely, ranging from a 2-fold increase to a 15-fold decrease. Its marginal effect on infectivity, on the other hand, was fairly consistent, ranging from a 3-fold to 11-fold increase. The effect of mutating the CypA active site was always positive for replication, with a much larger effect on the highly impaired Δ34-378-CypA. Likewise, the CypA mutation had a consistently positive marginal effect

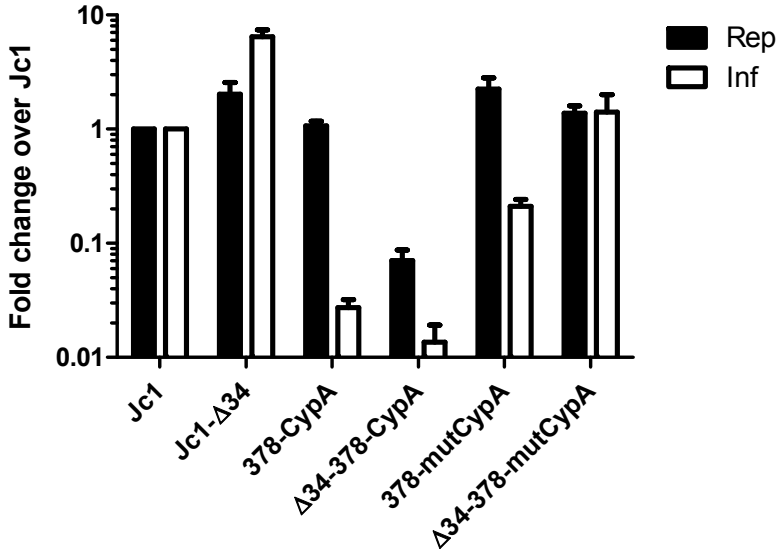


Figure 4-15. Relative replication and titers

48 h PE luciferase-based replication values and viral titers normalized to the Jc1 for comparison. Means and SEM of at least two independent experiments are shown.

Table 4-3. Marginal effects of genetic perturbations

Reference construct	Modification	Relative effect of modification	
		On replication	On infectivity
Jc1	Δ34 deletion	2.0	3.2
378-CypA	Δ34 deletion	0.066	7.6
378-mutCypA	Δ34 deletion	0.61	11
378-CypA	CypA active site mutation	2.1	3.7
Δ34-378-CypA	CypA active site mutation	19	5.3
Jc1	CypA insertion	1.1	0.025
Jc1-Δ34	CypA insertion	0.035	0.061
Jc1	mutCypA insertion	2.2	0.094
Jc1-Δ34	mutCypA insertion	0.68	0.32

on virus production. CypA insertion has contrasting effects for replication depending on whether the deletion was present, but impaired virus production regardless of the background.

4.5.2 Effect of CsA on NS5A-CypA fusion constructs

As the effects of the CypA fusion depend on the active site, and knowing that CsA binds to the active site of CypA and competes out NS5A in a CypA binding assay *(82)*, it is natural to ask whether CsA can reverse the negative effects of the CypA fusion on replication and virus assembly. The various constructs were electroporated into Huh7.5 cells and treated with a series of CsA concentrations at 4 h post-electroporation. At 48 h, the replication and infectivity levels were assayed to determine the effect of CsA treatment. The results are shown in Figure 4-16.

The typical 4-48 h effect on Jc1 was seen, with decreasing replication and a biphasic stimulation/inhibition of virus assembly. With CypA inserted at the 378 position, the resulting virus becomes highly resistant to CsA, even exhibiting a minor increase in replication with increasing CsA, up to a concentration of 2.5 μM. The effect of CsA on virus assembly is modest yet complex, with both a dip at 80 nM and a small peak at 800 nM CsA.

The 378-mutCypA construct, on the other hand, is actually more sensitive to CsA than the wild-type, with a replication IC50 of 208 nM, compared to 499 nM for Jc1.

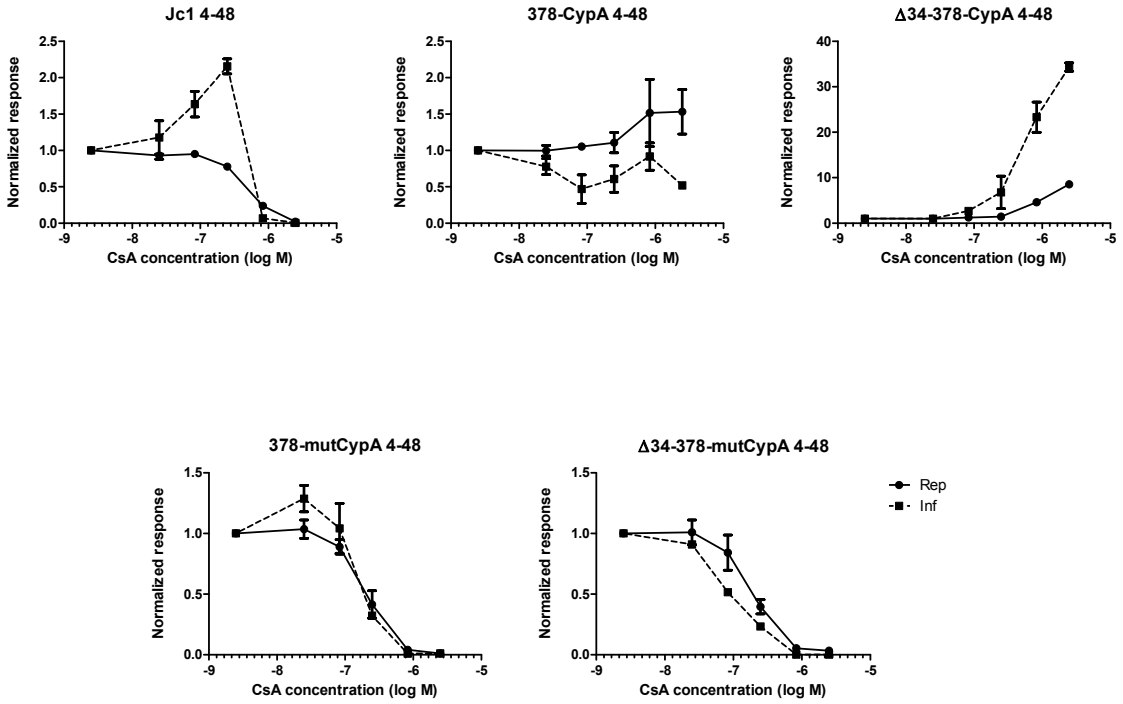


Figure 4-16. Effect of CsA on 378 CypA constucts

Replication and infectivity dose-response curves to CsA for the NS5A-CypA fusion constructs. Means and SEM of at least two independent experiments are shown.

Interestingly, the CsA induced stimulation of assembly appears to be largely erased, with a small peak creeping up around 25 nM instead of the usual 250 nM. The Δ34-378-mutCypA shows no stimulation of assembly at all.

The most dramatic response was seen for Δ34-378-CypA – the construct harboring both the CypA insertion and the domain II deletion. Replication increased with CsA concentration, nearly 10-fold at the maximum concentration 2.5 µM. The effect on virus production was even more pronounced, with an increase of approximately 35-fold at the highest dose of CsA. The effect of CsA on this construct appears to be analogous to the effect of mutating the active site of CypA. Extracellular infectivity also exhibited an approximately 30-fold increase with 2.5 µM CsA (data not shown).

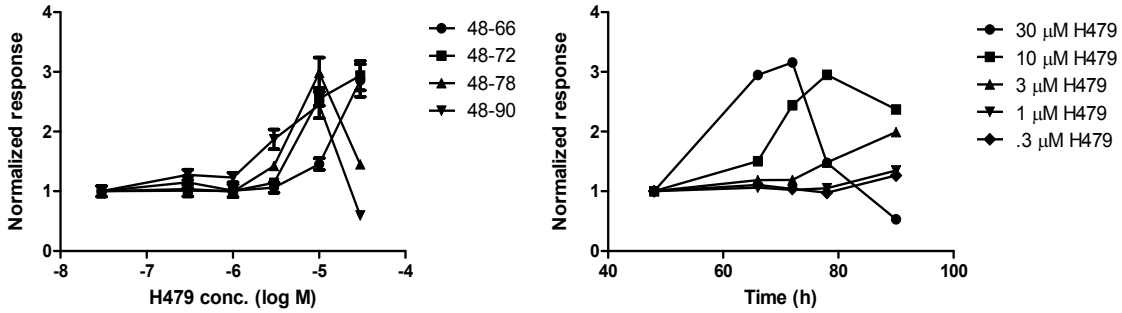
4.6 H479 increases translation at the expense of replication

As noted earlier, when measuring the dose response curves of various compounds, the kinase inhibitor H479 showed an interesting effect on HCV, apparently increasing replication and reducing virus production. H479 is an inhibitor of the alpha isoform of casein kinase I (CKI-α), of which NS5A has been shown to be a direct substrate *(111)*. Previous studies on H479 treatment in HCV replicons have shown an efficient reduction of the p58 hyperphosphorylated form of NS5A. Interestingly, this inhibition of p58 was also shown to potentiate replication in the poorly replicating Con1 unadapted replicons; however, in Con1 strains harboring an NS5A adaptive mutation which decreases p58, H479 treatment was inhibitory *(65)*. Reduction in p58 by adaptive mutations has also

been linked to decreased virus production in Con1 full-length genomes *(66)*. Taken together these data suggest that H479 would favor RNA replication and reduce virus assembly, in agreement with response we measured with H479. However, this prediction overlooks other data. The same p58-reducing adaptive mutation, when put into a genotype 2A genome (the genotype of Jc1), is highly maladaptive, completely abolishing replication *(64)*.

We observed that at the highest concentration of H479, the effect on replication was in fact ambiguous (Figure 4-7). For 4-48 h treatment, H479 was quite inhibitory at 30 µM. For 48-72 h treatment, luciferase activity was generally increased at 30 µM over DMSO alone, but it could be either higher or lower than at 10 µM H479. We reasoned that the ambiguity in the final data point might be due to variations in timing. Therefore dose response curves to H479 were measured for a set of different treatment durations, with the compound added at 48 h and harvested at 66, 72, 78, and 90 h post-transfection (Figure 4-17a). The shape of the curves confirmed the time-dependence of the H479 stimulation. At the 66 h harvest, a simple dose-dependent stimulation is seen, an effect which increases uniformly by 72 h. However, by 78 h the shape of the curve changes, with the highest dose now starting to decrease, a trend which continues to the 90 h harvest where the highest does is actually inhibitory, and the second highest dose has also started to come down from its peak. Transforming this same data into a time course with time as the independent variable further clarifies the situation. High doses

A



B

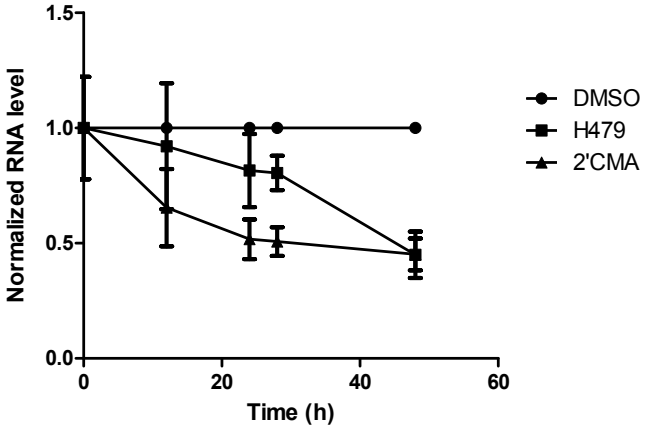


Figure 4-17. H479 effect over time

(A) Effect of harvest time on dose-response curves. 48 h PE treatment with H479 was followed by harvest at four different time points, producing four different dose-response curves. In the second panel, the same data has been re-plotted in the time domain, showing the kinetic response of different doses of H479 following addition at 48 h PE. (B) RNA response kinetic shown as a function of time post-drug addition, normalized to the DMSO treated cells. Means and SEMs for triplicate samples are shown.

of H479 induce a transient increase in luciferase expression which cannot be sustained. Genotype 2a virus has already been shown to be sensitive to reductions in p58 NS5A; H479 may be reducing the level of p58 below an optimum level.

One possible explanation for this transient increase is that the p58 form favors translation of RNA over replication. To test this, we measured HCV RNA in cells along with luciferase activity over time after H479 addition (Figure 4-17b). The results confirmed the hypothesis. Although H479 produced a transient increase in luciferase expression, it simultaneously led to a decline in RNA levels. Neither the increase in luciferase nor the decline in RNA was immediate, which probably reflects the time it takes for the existing p58 to decay away.

4.7 H479 is dominant over CsA enhancement of virus production

Having observed the myriad effects of H479 and CsA on HCV regulation, we tested the effect of simultaneous administration of H479 and CsA for the 48-72 h treatment interval. As seen in Figure 4-18, H479 seems to cancel the stimulatory effect of CsA on virus production. The converse was not the case, and the inhibitory effect of CsA on replication was additive with the higher luciferase expression caused by H479. At each of the three tested concentrations, the response to simultaneous addition of CsA and H479 closely resembled the response to H479 alone.

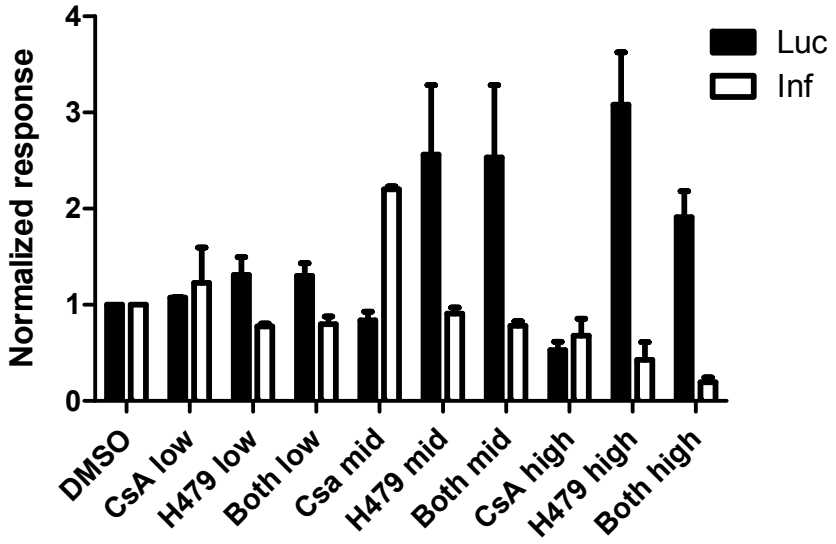


Figure 4-18. Simultaneous administration of CsA and H479

Response to three different doses of CsA and H479, given separately or together for the 48-72 h interval. Luciferase expression and intracellular infectivity is shown normalized to the DMSO control. For CsA, low = 83nM, mid = 250nM, high = 830nM. For H479, low = 3µM, mid = 10µM, high = 30µM. Means and SEM of two independent experiments are shown.

The interaction between the two compounds suggests that the stimulation of infectivity from CsA is mediated through p58, and possibly by newly formed p58 only. Additional experimentation, such as kinetic measurements or variations on the timing of drug addition would provide greater detail into the mechanism of this interplay; however, the association of virus assembly with hyperphosphorylation is consistent with multiple reports making the same correlation *(40, 66, 112)*.

4.8 Discussion

Employing a combination of genetic and chemical manipulations, we sought to further our understanding of NS5A and its role in regulating the different elements of the HCV life cycle. The results suggest an intricate system centered on NS5A and involving numerous interactions with host factors to coordinate protein translation, RNA replication, and virion assembly. The NS5A/CypA interaction, and its effects on RNA replication and virus assembly, was the primary subject of this study. The requirement of this interaction for RNA replication is already well-documented; here we demonstrate negative regulation via CypA of virus assembly and RNA replication as well as a separate positive requirement for virus assembly. This bi-directionality is something of a recurrent theme for NS5A, where a particular perturbation or interaction can have either a positive or negative effect, depending on circumstances. Inhibition of CKI-α, is another example, which enhances replication in the genotype 1 context, but inhibits it in the 2a context. We have shown that inhibition by H479 can transiently increase protein expression, only to ultimately diminish it.

One possible explanation for the bi-directional nature NS5A perturbation is quite apparent in light of the widely-held view that the role of NS5A is best described as regulatory in nature. As a regulator of the virus life cycle, NS5A is not merely contributing to one activity or another, but setting a compromise between several different and, to some extent, competing activities.

4.8.1 CypA, CsA and NS5A

We have shown inhibition of CypA by CsA treatment to have a stimulatory effect on virus production at low concentrations. Several lines of evidence support the case that this is a specific effect on HCV regulation, rather than a general effect (for example, creating cellular environment more conducive to virus assembly) and that this effect is mediated through NS5A. Temporally, the effect depends more on the progression of HCV replication than on treatment duration, arguing for a viral, rather than cellular, pathway. The effect depends on overall HCV replication level, also arguing the same.

The relationship between timing, replication levels, and CsA-mediated assembly enhancement might be understood by supposing CypA to have a restraining effect on virus assembly, either directly or as an indirect consequence of its role in facilitating RNA replication. At early stages of the viral life cycle, virus production may be constrained by a more basic limitation, such as availability of RNA or structural proteins. Later in the infection cycle, as RNA levels rise, the regulation of virus production through

CypA may become the limiting factor. Inhibition of CypA via CsA would then release that limitation and allow virus production to increase.

Incidentally, the timing of CsA-mediated assembly enhancement coincides with the plateau of replication levels which follows the initial exponential growth phase, with the transition in phase occurring between 24 and 48 h, as typically seen in a growth curve following RNA electroporation, such as Figure 4-5a. It is tempting to hypothesize that the limiting of replication and the limiting of virus assembly, which are correlated in time, are also mechanistically linked.

The shift in peak enhancement concentration caused by the DEYN mutations in NS5A makes a strong case for it being the primary partner of CsA in the enhancement of infectivity by CsA. While the DEYN mutations do not rescue the direct interaction between NS5A and CypA that is inhibited by CsA *(113)*, they do confer resistance to the drug. The implication is that the mutations render NS5A less dependent on CypA, perhaps altering its conformation to one better approximating the CypA-NS5A complex or the CypA-isomerized NS5A. The direction of the peak shift and the increase in magnitude are both consistent with this model. If the CypA-NS5A complex has a restraining effect on virus assembly, the mutated NS5A might restrain it more, which would result in both a higher concentration of CsA required to release the restraint, and a stronger stimulation of assembly upon that release.

By fusing CypA to NS5A, the effect of this particular interaction could be studied more specifically, albeit in a more artificial system. These constructs provided additional corroborating data. Consistent with previous experiments, these fusion constructs show that the interaction between CypA and NS5A results in a moderate negative effect on virus production, an effect which requires the active site of CypA. The difference in virus production which maps to the active site of CypA is 4- to 5-fold when corrected for differences in replication, roughly on par with the magnitude of the enhancement of virus production observed at low doses of CsA. Unlike the previous experiments, the CypA-NS5A fusion experiments also show a strong negative effect on replication. This result is notable as it stands in contrast to the conventional role of CypA supporting replication. The effect is not seen in the full-length NS5A, but in the context of the Δ34 domain II deletion, there is a nearly 30-fold reduction in replication from inserting CypA. The reason for this difference, presumably due to an interaction between CypA and domain II, remains unknown. In either case mutation of the CypA active site results in an increase in replication; in the case of the domain II deletion, the rescue is nearly 20-fold.

The effect of Csa on the fusion constructs is not consistently equivalent to mutation of the active site; while CsA treatment mimics the CypA active site mutation in the context of the domain II deletion (Δ34-378-CypA), resulting in significant increases in both replication and virus production, CsA treatment of 378-CypA without the deletion does not exhibit an increase in virus production comparable to that achieved by mutating the CypA active site. One possibility is that the interaction between NS5A and

CypA is stronger with the full length domain II, and therefore a higher concentration of CsA is required to have an effect; this notion is supported by an NMR study which indicates that CypA interacts with NS5A at numerous points distributed across domain II *(82)*. Compensating for this with an increased CsA dose is complicated by the fact that CsA targets the endogenous cellular CypA as well as the engineered NS5A-CypA and because even the CypA in NS5A may be playing multiple roles with different CsA sensitivities. For example, the strong inhibition of assembly usually seen at the higher doses of CsA could be an independent effect now muting the anticipated rescue of infectivity.

The dose-response curves of the mutant CypA constructs are also noteworthy for their deviation from the unmodified Jc1 and for the consistency of the phenotypes with the working model. Both mutant CypA constructs exhibit higher sensitivity to inhibition by CsA, which can be interpreted as an increased requirement for CypA or a decreased efficiency of interaction. This may be due to some residual interaction from the mutant CypA interfering with the authentic CypA-NS5A interaction as a weak dominant negative element. Based on the working model, we would predict that the infectivity peak in the dose-response would shift to the lower concentrations and become smaller in magnitude, reasoning analogously to the DEYN mutation. This effect can be seen in seen in the 378-mutCypA response. If the domain II deletion further reduces CypA binding, the peak might reasonably be imagined to become undetectable altogether.

In general, attempts to formulate an overarching model to explain the full range of observed phenotypes are hampered by the complexity of the system. The underlying range of possible mechanisms is correspondingly complex. CypA is believed to interact with a number of prolines in domain II, not to produce any single structure, but to effect changes in the overall conformation of what is considered an intrinsically disordered protein (IDP) domain *(81)*. These conformational changes in turn could influence a number of other interactions or activities, including NS5A/5B cleavage kinetics *(83)*, RNA binding *(79)*, or localization.

4.8.2 H479

The interaction between H479 and CsA under co-treatment provides additional evidence that the infectivity enhancement of CsA is mediated through NS5A. The mechanism behind this interaction is unclear; however, the p58 isoform has been linked to virus assembly in several studies, and our results, which show H479 consistently reducing virus assembly, conform to this hypothesis. In addition, preliminary kinetic studies (not shown), show a delay of about 10 h after H479 addition before increases in luciferase and decreases in infectivity begin to manifest. Remarkably, a nearly identical delay is seen before enhancement of infectivity begins under CsA treatment. In contrast, this delay is not seen under treatment with the NS5A inhibitor BMS-790052, which begins to inhibit both activities immediately upon addition. The most probable reason for the delay in the case of H479 is fairly apparent: H479 has no effect on existing p58, which has already been phosphorylated. H479, therefore, would have no effect until

some portion of the extant p58 has turned over. The observation that CsA exhibits the same delay suggest that 1) the effect of CsA is to prevent isomerization of NS5A by CypA rather than inhibit the (reversible) binding of CypA to NS5A, and 2) that the target of CypA, at least in the context of regulation of infectivity, is also p58. Again, this is consistent with the association of p58 with virus assembly.

One distinctive conclusion drawn from our experiments with H479 is that RNA replication is associated with p58 rather than p56, as generally believed. Instead, we hypothesize that p56 is more specifically linked with translation, not RNA replication. As both translation and RNA replication are required for efficient RNA replication, this model is not necessarily in conflict with previous data, and potentially resolves certain ambiguities. In the case of genotype 1 replication activation, one need only hypothesize that the genotype 1 NS5A heavily favors the p58 replication activity over the p56 translation activity. By locking up what little RNA there is into replication complexes rather than translating more of it to generate more proteins, replication can be made very inefficient, to the point where it is not self-sustaining. Inhibiting p58 by H479 or adaptive mutations would then re-allocate RNA for translation and dramatically increase replication efficiency. A similar hypothesis has been recently been put forward to explain the inability of some HCV strains to replicate, based on a theoretical analysis *(114)*. Work by Linda Andrus has confirmed that H479 is inhibitory for Jc1 and also suggests that Con1 genomes may have ten-fold lower baseline translation efficiency than Jc1 (person communication), which is consistent with our model.

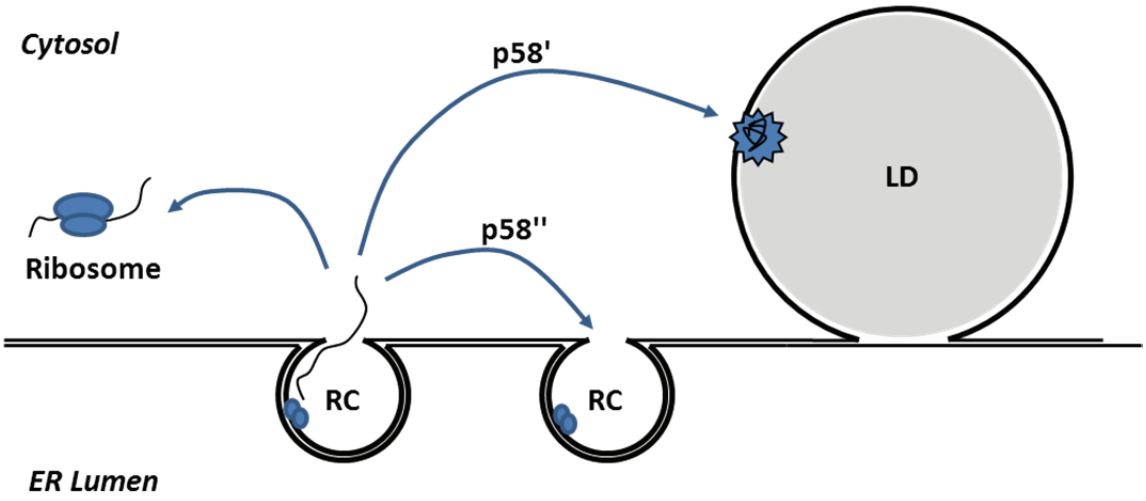


Figure 4-19. Conceptual model of NS5A regulation

Model of HCV regulation by NS5A. Nascent RNA formed within the replication complex (RC) has numerous possible destinations. RNA untended by NS5A goes to ribosomes for translation. CypA-isomerized p58 targets RNA to lipid droplets (LD) for packaging into virions. Doubly isomerized p58 targets RNA to RCs for replication.

4.8.3 Conceptual model

The task of regulating RNA replication and virus assembly can be compared in some respects to regulation of cell division, especially when one considers the problem of cell size sensing. Unicellular organisms generally maintain a fairly uniform maximum size over many generations; in order to do this there must be a kind of molecular measuring stick to sense cell size *(115)*. Similarly, for HCV to regulate its own replication, it must have some way of detecting and responding to its current state. CypA is remarkably well suited to this purpose; its ubiquity and abundant, stable expression make it a reliable standard for comparison, the same qualities which make it an excellent housekeeping gene *(116)*. The coupling of CypA with NS5A is particularly attractive in light of what appears to be NS5A's modus operandi: a multi-functional protein, with intrinsically disordered, proline-rich domains, the activity of which is modulated through overall conformational changes.

Because of the large number of possible relationships between conformation and function and the dearth of data defining those relationships, any specific model of this paradigm would be by necessity highly speculative. However, as an illustrative example, we propose one possible model, shown in Figure 4-19. Here the primary function of NS5A is assumed to be RNA allocation, and various modifications to NS5A modulate to which of three compartments NS5A will direct the HCV RNA. Without targeting by NS5A, RNA moves from the sites of replication to ribosomes for polyprotein translation. An altered conformation of hyperphosphorylated NS5A, dubbed p58', brings RNA to lipid

droplets for packaging. A third conformation, p58'', feeds nascent RNA into new replication complexes for further amplification. No function is assigned in this model to p56 or to the native conformation of p58. Assuming that p58' is the product of CypA *cis-trans* isomerization and p58'' requires further isomerization, a number of observations can be explained.

High doses of CsA would be expected to prevent formation of both p58' and p58''. Under these conditions, assembly of virus particles would be inhibited, as would any expansion of replication complexes. Existing RCs could continue replicating, in keeping with Figure 4-8. Intermediate doses, on the other hand, might increase p58' (and increase virus assembly) while decreasing p58'', particularly if the second CypA interaction was of a lower affinity. On the other hand, in the case of excess interaction with CypA, as might be in the NS5A-CypA fusion, more p58' may be converted to p58'', impairing virus assembly.

The time- and replication-dependence of CsA enhancement of infectivity can be understood in light of the fact that while CypA abundance in the cell is quite stable, the abundance of NS5A changes dramatically. Therefore the relative abundance of CypA, and the resulting efficiency of p58' and p58'' formation would depend on time and replication level. In particular, at early time points or low replication levels, the relative abundance of CypA would be quite high, meaning p58' may exist in excess. In this case increasing p58' levels by low-dose CsA treatment would not result in any increase in virus assembly. Another attractive feature of this model is that at early times with very

high CypA abundance, p58'' production would be at its highest, while at later times p58' might become more emphasized. This would effect an early focus on replication which would gradually shift towards virus assembly as NS5A levels in the cell increased. At still higher levels, both activities might be moderated.

The DEYN mutations, mutCypA insertion and H479 dose-response have already been discussed and fit easily into this model. The model is also consistent with published work on the NS5A inhibitor BMS-790052, which has been shown to block p58 formation, alter localization, and putatively interferes with two distinct activities of NS5A *(63, 64)*.

The 378-Ypet insertion and compensatory deletion can be explained if the insertion into domain III is seen as somehow altering the function of p58', perhaps causing p58' to function more like p58''. The anticipated result would be to reduce virus production and increase replication. The deletion then, might do the opposite, and cause p58'' to function more like p58'. Further work is required to further define these relationships; however, the model illustrates the potential complexity of the system and suggests lines of future exploration.

4.8.4 Viral regulation in context

One significant question raised by this model is how the regulation of HCV proposed here compares with the regulatory mechanisms of other viruses. Interestingly, not all members of the Flaviviridae express a separate NS5A protein – the classical flaviviruses have a single NS5 protein which functions as the viral RdRp and is also phosphorylated

(117). The N-terminal portion of NS5 in this case catalyzes the formation of the 5'-RNA cap, explicitly linking NS5 with polyprotein translation *(118)*. Both the protein and its activity differ significantly from HCV NS5A; however, it seems possible that regulatory loops, albeit of a different nature, are also in operation here. On the other hand, these differences in regulation may be related to the very different survival strategies and pathogenesis of these two virus genera. The flaviviruses tend to produce an acute, vigorous infection *(119)*, while HCV produces a relatively mild and often asymptomatic acute phase followed by a protracted chronic infection which may go unnoticed for many years. The long-term evasion of host immunity may be the driving factor in the development of what seems to be a complex, highly controlled system of viral regulation. This also raises the concern that the atypical, highly productive cell culture strains of HCV may not be appropriate subjects of study for truly understanding the regulatory systems of the virus. Development of more authentic model systems may prove necessary in this case.

The complex regulation of a chronically infecting virus is reminiscent another small RNA virus, HIV. HIV has a number of proteins involved in regulation of the viral life cycle, the most-studied being Tat and Rev. Tat, the trans-activator of transcription, interacts with the long terminal repeat (LTR) promoter of HIV *(120)*. In the absence of Tat, transcription of viral mRNAs is highly inefficient, resulting in stalling of the cellular polymerase Pol II and the production of abortive transcripts *(121)*. Expression of tat potentiates cellular transcription several hundred-fold *(122)* - a positive feedback loop

yielding an explosive increase in viral gene expression. Interestingly, the *tat* gene resides near the 3' end of the HIV genome. In light of the poor processivity of transcription in the absence of Tat, it would seem reasonable for Tat to be situated closer to the LTR for efficient activation; its distal placement suggests a need for delaying or otherwise regulating the HIV in a way other than simply maximizing raw transcription efficiency. This regulation could, for example, represent a switch from a latent to active state. The Rev protein plays a role in export of the viral RNA *(123)*. Because HIV is normally spliced by the cellular splicing machinery, exporting a complete genomic RNA for packaging into new virions requires special measures. Rev binds HIV RNA in the nucleus before it is spliced and by oligomerizing onto the RNA, protects the splice junctions from recognition *(124)*. The regulatory rationale is easily guessed at: at early times, HIV RNA is spliced to produce all the different protein products including virion components, immune evasion factors, and Rev itself. As Rev accumulates, genomic RNA is exported for packaging, and the virus shifts from a protein translation-focused early phase, to a virion assembly-focused late phase *(125)*. A second feedback loop occurs as Rev levels rise and spliced HIV mRNA levels fall to the point where Rev expression, which itself requires splicing, decreases. This in turn increases splicing and restores Rev levels *(126)*. The complexity of HIV regulation further underscores the importance of careful coordination of the viral life cycle activities for fitness and surivival.

4.8.5 Outlook

The regulation of HCV through NS5A is strikingly complex, with numerous modifications affecting NS5A in myriad ways. CKII phosphorylation has not even been considered in this study, nor have we considered NS5A effects on innate immunity, modulation of NS5B function, or the many adaptive mutations which have been reported to enhance RNA replication or virus assembly. Adaptive mutations in particular raise an interesting point. Naturally circulating HCV strains have been notoriously difficult to culture, and almost always require adaptive mutations, which, surprisingly, have appeared in almost every part of the HCV genome. Superficially it might appear that wild-type HCV is highly suboptimal and inefficient, when ten- or hundred-fold improvements in replication or virus production can be achieved with single residue mutations. Alternatively, HCV could be seen as operating under a set of strict controls which allow it to evade both innate and adaptive immunity and establish spectacular success as a chronic infection. From this perspective, adaptive mutations might simply be a breaking open of normally closed control loops and allowing viral processes to proceed out of control. If this is true, it further complicates the study of HCV regulation. Any “good” cell culture system would almost by definition have lesions in its regulatory systems. Therefore we must be cautious in interpreting experimental data. One possible approach is to study the same regulatory phenomena (ex. NS5A phosphorylation) in multiple different cell culture systems (Con1, JFH1, Huh7.5, primary cells), which are

hopefully broken in different ways. Comparisons between more and less adapted strains (JFH1 vs. J6/JFH vs. J6/JFH-Clone 2) might also yield valuable insights.

This important caveat notwithstanding, there are several opportunities for extending the current work. Further work is required with H479, particularly as it is somewhat at odds with existing hypotheses. One testable prediction from our model is that Con1 would have a low protein-to-RNA ratio, which would increase with H479 treatment. Measurements of protein-to-RNA ratios would also be helpful in understanding what the CypA fusion constructs are doing to the relative balance of RNA replication and protein translation.

The difference in infectivity enhancement curves between Jc1 and JFH1 is also very interesting. It seems quite possible that this difference would map to the N-terminus of NS2, which amounts to five amino acid changes between the two strains, one of which is a proline to serine substitution – the least conservative of the five changes. If the difference were to map to this residue, it would be a candidate for yet another CsA target in HCV.

One system we have not capitalized on is the CypA knockdown Huh7.5 cells. We have gone as far as to test that the CypA fusion constructs replicate in these cells (data not shown); however, testing the dose response curves to CsA for these constructs as well as for the DEYN mutants would be valuable in testing our working model. We

would predict that there would be little or no infectivity enhancement in the DEYN mutant under CsA treatment in the CypA knockdown cells.

One other major avenue for future work is the inhibitor response experiments. We believe that measuring the response of HCV to a panel of inhibitors shows promise as a powerful approach to unraveling the details of the HCV life cycle as a system. This, however, is left for the final discussion in chapter 6.

Chapter 5. Covariation in HCV sequence alignments

Chapter 5 Covariation in HCV sequence alignments

5.1 Introduction

The statistical analysis of multiple sequence alignments (MSAs) has long been pursued as a means of inferring underlying biophysical facts about a protein from its genetic diversity *(127, 128)*. The degree of amino acid conservation at particular positions, for example, is seen as evidence of structural or functional significance of those residues ultimately resulting in a fitness advantage for the organism and the consequent stability of that residue across evolutionary history. Like conservation, covariation, also known as co-evolution or correlated mutation, is a statistical property of a MSA. Unlike conservation, which measures the variation at a single position in the protein (a single column in the MSA), covariation is a pairwise property describing the correlation of variation between two amino acid positions (two distinct columns of the MSA). If two positions vary in a completely independent fashion, those sites are said not to covary. On the other hand, if a mutation at one site is frequently accompanied by a mutation at the second site, those two sites are said to covary.

Strong covariation between two sites may be indicative of intra-molecular structural contact or coordination between the sites for enzymatic function, ligand binding, correct folding, etc. Hence, correlated mutation analysis has been proposed as a method to predict residue contacts and generate information about protein structure and function *(127)*. While attractive in principle, practical realization of this concept has

proven exceedingly difficult, for reasons to be explained below, and accurate prediction of residue contacts directly from analysis of MSAs remains an open problem.

Numerous methods for estimating correlation between two random variables exist; of these, a smaller subset of methods is suitable for calculating the correlation between discrete, non-numerical variables, such as the letters in a protein sequence. These methods including Observed Minus Expected Squared (OMES) *(129)*, Statistical Coupling Analysis (SCA) *(130)*, and Mutual Information (MI) *(128)*.

5.1.1 Mutual Information

Mutual information (MI) is a technique originally developed in the field of information theory which quantifies the mutual dependence of two random variables (in this application the identity of two amino acid residues). MI is calculated by comparing the joint probability distribution between two variables X and Y with the product of the two separate probability distributions, given by

$$MI(X,Y) = \sum_{y \in Y} \sum_{x \in X} P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

The summations are over all possible values of x and y, in our case, the twenty possible amino acids. P(x) is the probability of a particular amino acid appearing at position X, and P(x,y) is the probability of encountering a specific pair of amino acids at

positions X and Y simultaneously in a one sequence. If the amino acid positions vary completely independently, the frequency (or probability[2]) of a particular pair of amino acids appearing in the same sequence will be equal to the product of the probabilities of each of the two amino acids appearing at each of the two sites respectively. Any disparity between the product of the probabilities and the actual distribution is indicative that the identity of the amino acid at one site is affected by what amino acid is at the second site and vice versa (there is no directionality to mutual information). For example, consider a MSA in which histidine occurs at position 1 in 10% of sequences, and a valine occurs at position 2 in 10% of the sequences. If the two positions are completely unlinked, the chance of a sequence containing both a histidine at position 1 and a valine at position 2 is 1% - the product of the two probabilities. If in fact, that particular pairing occurs in 10% of sequences, which means the two amino acids always appear together, position 1 and 2 are considered very strongly linked.

Another way to understand MI is to ask, given two sites of unknown amino acid identity in a sequence, if the identity of one is made known, how much is the uncertainty of the second site diminished? If two sites are perfectly linked (ex. an histidine at position 1 is always accompanied by a valine at position 2, but position 2 is always a threonine when site 1 is a glycine), then revealing the identity of one amino

[2] In the literature, the term "probability" is typically used instead of "frequency". For the purposes of this study, the terms are interchangeable – the observed frequency of an amino acid at a site in the protein is implicitly an estimation of the underlying probability of that amino acid appearing at that site.

acid will completely eliminate any uncertainty about the second amino acid – there is information which is *mutual* between the two sites.

5.1.2 Sources of covariation

Any analysis attempting to infer protein structure or function from sequence covariation depends on the degree to which observed covariation is actually a consequence of functional and structural linkage. In fact, observed covariation may have multiple sources, of which true interaction between amino acids is only one.

In addition to direct interactions between residues, covariation is also caused by indirect interactions. Allosteric effects are an obvious example, but indirect correlations need not stem from dramatic conformational changes and may result from subtle distributed effects or a series of interconnected direct interactions. Efforts to distinguish direct interactions from indirect interactions by more sophisticated mathematical techniques have met with some success; however the approach has been limited to proteins existing in multiple copies per genome *(131)*.

A third potential source of covariation in the amino acid MSAs is the presence of RNA secondary structure. The one-to-one base pairing of RNA secondary structure generally produces strong covariation signal from the nucleotide sequence which may result in apparent covariation between two amino acids which in fact have no interaction at all. Generally speaking, RNA secondary is more often located in untranslated regions of an RNA molecule, although for RNA viruses, which tend to pack

information more densely, RNA secondary structure often does overlap with protein coding sequences.

One feature the three sources of covariation described thus far share is that they all are rooted in biophysical interactions and in that sense are all “real”. The fourth and fifth major sources of covariation, random covariation and phylogenetic covariation, differ in that the resulting correlations are largely incidental. These correlations can easily drown out the other “true” sources of covariation and present a significant challenge to extracting useful information from any analysis of covariation.

Random covariation can result from sampling error, where there are not enough sequences in the MSA for the signal to outweigh the background noise *(132)*. A related type of random MI occurs when there are large differences in conservation at different positions in the MSA *(133)*. Highly variable positions will tend to score higher in measurements of covariation than highly conserved positions. This makes it more difficult to detect covariation between more conserved positions, when these covariances are often the most interesting.

5.1.3 Phylogenetic covariation

The problem of phylogeny is a general one affecting any attempt to determine the correlation between two random variables across a genetically diverse sample set – not only between two amino acids, but between any pair of variables at all, for example, attempting to correlate the brain size of animals with their body mass. The problem is

described in detail by Felsenstein *(134)*, in which a set of data (protein sequences, physical attributes, etc.) taken from numerous species might appear to show interesting correlations when in fact they are artifacts of phylogenetic relationships between the sampled species. How this occurs is best illustrated with a specific example.

In the case of a covariation analysis, a large number of related HCV isolates are sequenced, and the sequences for the protein of interest are aligned and compared. If a particular pair of mutations occurs exclusively in a concomitant fashion – every sequence has either both mutations or neither – the conclusion that the two sites are meaningfully linked may appear highly compelling. An attractive interpretation would be that either mutation must be highly deleterious and selected against, and only in the presence of a second, compensatory mutation can the protein maintain its necessary function and be maintained in the gene pool. There is, however, another possibility: that all the sequences containing the double mutations share a common ancestor. The two mutations, which may be essentially neutral for protein function individually or in tandem, happened to exist in that ancestral sequence, and having no fitness cost, these two mutations were maintained in all of its progeny. As a result, the two mutations repeatedly appear together, not due to any functional relationship between the two sites, but by virtue of a shared lineage. Various methods have been developed to correct for phylogenetic covariation, which are discussed further below.

5.2 Calculating mutual information in HCV p7 alignments

In order to calculate MI from HCV sequence alignments, the MI algorithm as described by Korber *(128)* was implemented in C++. As a quick check for functionality, the MI algorithm was executed on an alignment of 874 protein sequences of the V3 loop region of HIV-1 Env, similar to the original analysis by Korber. While it was not possible exactly reproduce the results (which would require using the same exact alignment used in the earlier study), there was good overlap with the original covariation results, with 6 of the 7 originally identified hits showing up in the top 20 hits of the new analysis. The overlap is shown in Figure 5-1, where the horizontal and vertical axes are residue number, with a spot marking a covarying pair. The size of the spot indicates the strength of the correlation.

We next chose to test the MI program on a set of HCV p7 sequences. Being the shortest of the HCV proteins, p7 gave the simplest test case with least computational cost. Sequences were downloaded from the euHCVdb *(91)* and culled by hand to remove sequences less than full 63 amino acid length of p7 or containing any gaps. From the remaining sequences we selected two sets according to genotype: one containing 455 sequences of genotype 1a only, and one containing 538 genotype 1b sequences. MI was calculated from these two alignments, with the results shown in Figure 5-2. Certainly very few observations can be made directly from these visual representations.

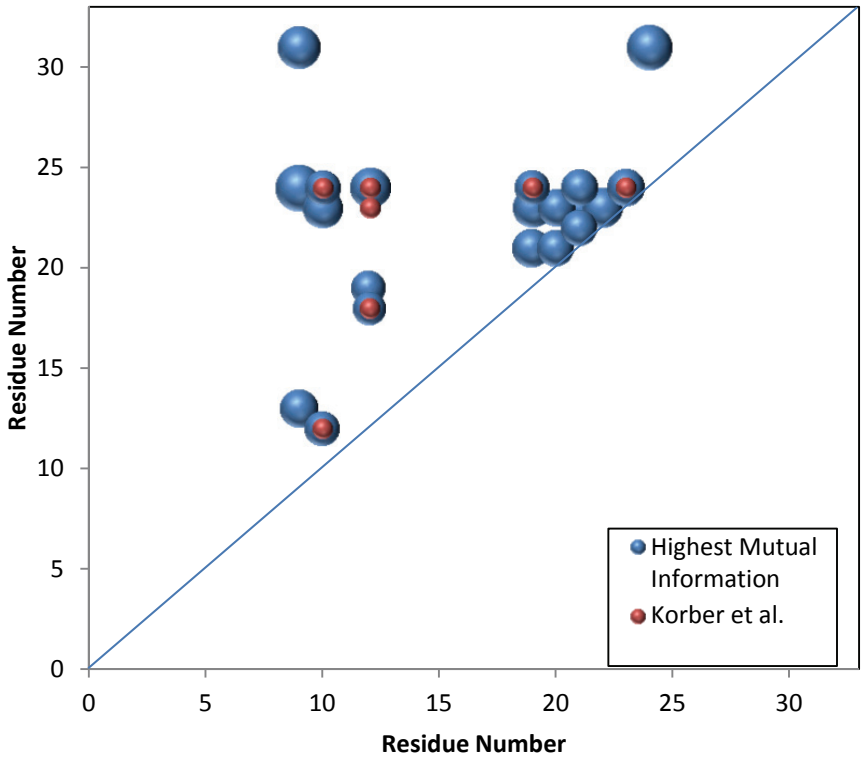


Figure 5-1. HIV Env V3 Mutual information

Top hits from MI calculation, compared with previously published results. MI was calculated for an alignment of 874 sequences of the V3 loop of HIV-1 B Env protein.

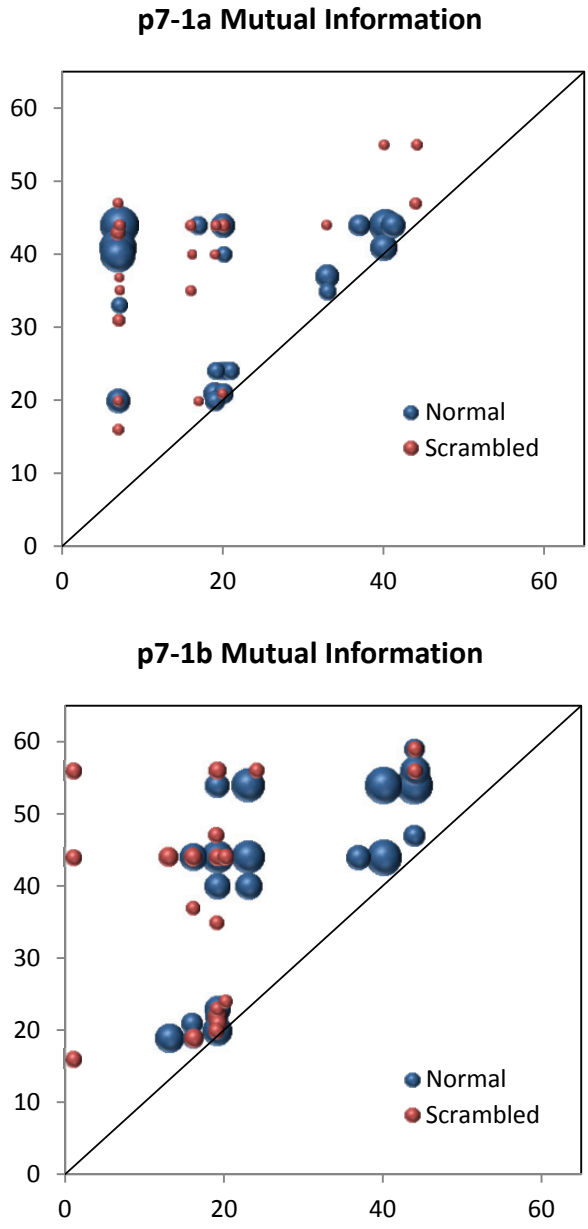


Figure 5-2. p7-1a and p7-1b MI

Top hits from MI analysis for p7-1a and p7-1b MSAs. p7-1a MSA contains 455 sequences. p7-1b MSA contains 538 sequences. The MSAs were also scrambled and MI was calculated from the scrambled MSAs. Both axes show residue number. A spot represents a detected covariation between the residue pair given by the coordinates.

There appears to be significant agreement between the two covariation sets; however, whether the identified covariations are meaningful hits or artifacts of phylogeny or random noise is unknown.

In an attempt to differentiate this covariation data from noise, scrambled alignments were generated by repeatedly exchanging letters from one sequence with the letter from the same position from a randomly chosen sequence. By exchanging letters only within columns, the distribution of amino acids within any one column in the sequence alignment is preserved; however, correlations between columns should be completely abolished. These scrambled alignments were then analyzed for MI in the same way. The results suggest that random variation does account for part of the correlation observed. Many of the same pairs identified in the MI analysis also appear with the scrambled alignment. This problem appears to be a manifestation of random covariation stemming from differences in the degree of conservation. Examining the sequence conservation plot, it is apparent that the high scoring pairs from the scrambled MSA are all pairs of less-conserved residues.

The problem appears to be more pronounced with the 1b sequence set. In the 1a results, the MI values from the scrambled MSA are smaller, and many of the points are unique to either the normal or scrambled dataset. The MI scores from the 1a MSA are also higher than the 1b scores overall, which further indicates that the 1a set contains more useful covariance information.

5.2.1 Evaluating MI results by detection of alpha-helix stacking

To determine whether the detected covariation has any basis in the biophysical properties of the protein, it is necessary to compare against some known biophysical data about the protein. If a crystal structure is known, the mutual information between amino acids can be compared to their physical proximity to see whether higher scores tend to correlate with smaller inter-residue distances. For p7, the crystal structure is not known; however, NMR studies and structure predictions indicate that the majority of the protein is in the form of two trans-membrane alpha-helices *(135, 136)*. The knowledge of the largely alpha-helical nature of p7 provides enough a basis for making an evaluation of the covariation results by whether they can detect evidence of stabilizing interactions between stacked residues in the alpha helix. This type of interaction has been shown to produce linkage between residues 3 to 4 residues apart, which matches the typical alpha helix periodicity of 3.6 residues per turn *(137, 138)*.

If the covariation detected by MI analysis is due to intramolecular interactions within p7, one would then expect a propensity towards finding high scores in covariant pairs 3 to 4 residues apart. In fact, for the genotype 1a analysis, 5 of the top 20 MI scores are from pairs 3 or 4 apart. This result is bolstered by the fact that from the scrambled 1a alignment, only 2 of the top 20 hits meet this criterion, which is the value one would expect if ranking the pairs randomly.

The results from the 1b sequences are less clear overall. Like the 1a MI results, 5 of the top 20 pairs are 3 or 4 residues apart. However, when the sequences are scrambled, there are still 4 such pairs in the top 20. The higher degree of overlap between the unscrambled and scrambled 1b MSA has already been noted, so the result is not particularly surprising. That the scrambled data produces 4 helix-stacking hits in the top 20 is more surprising. The reason is that a number of the variable positions in the 1b MSA happen to be 3 or 4 residues apart, which gives a preponderance of random MI hits with that spacing, independent of any correlated mutations between those columns.

5.2.2 Combining the 1a and 1b MSAs brings phylogeny to the forefront

Some of the difficulty in detecting meaningful covariation can be attributed to a lack of diversity; with many of the positions in the MSA quite highly conserved, MI signal tends to be weak and subject to greater sampling error due to the small number of mutations. One way to increase the diversity of the p7 MSA is to combine the 1a and 1b sets. The two sets were appended to generate a 1ab MSA, and once again tested for MI. The results are very different from the covariance sets previously shown (Figure 5-3). The magnitudes of the MI scores are three times higher those of the 1a set, and much more uniform. When helix-stacking performance was tested it was seen to be no better than a random ranking, implying that the detected MI is not a result of true interactions.

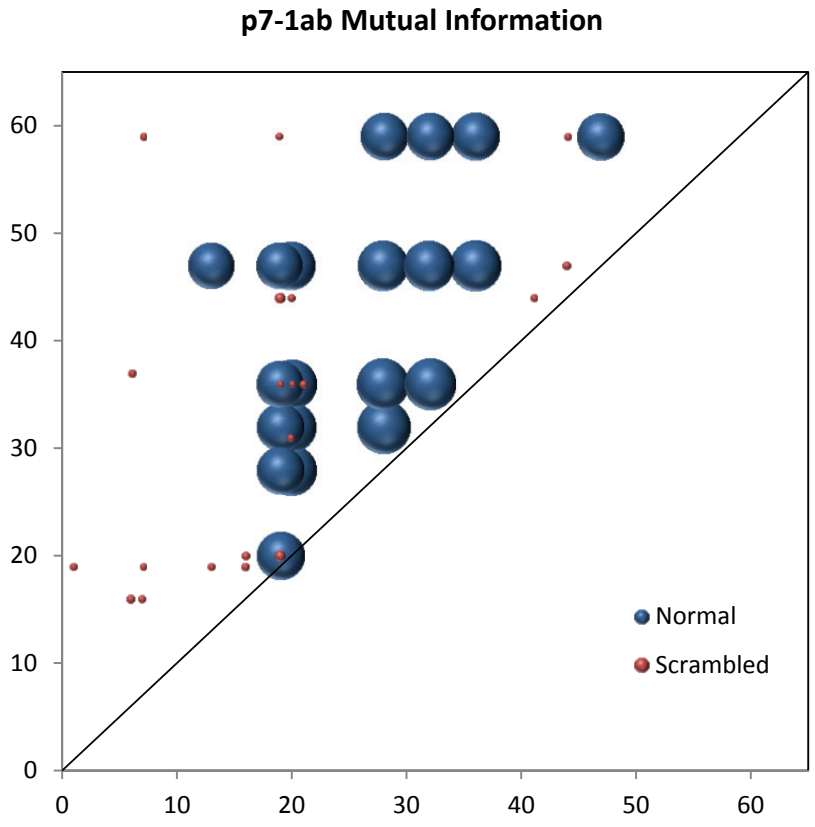


Figure 5-3. MI for combined 1a and 1b p7 MSA

MI for the normal and scrambled combined MSA of p7-1a and p7-1b. MSA contains 993 sequences total.

For comparison, the scrambled MSA was also analyzed; unlike the previous results, the scrambled MSA here produces much smaller scores which are almost entirely non-overlapping, suggesting a qualitatively different source of the false MI than observed thus far. Upon a closer examination of the results, it becomes clear that the high scoring pairs are various combinations of a few residues, and that those residues are the ones which are specific to genotype 1a or 1b. In other words, the observed covariation is now dominated by phylogenetic influence.

This effect is most easily understood by considering an MSA containing multiple copies of just two sequences. Any residue conserved between the two will have no variation, and naturally, no covariation. Any residue which differs between the two will covary perfectly with every other residue which differs between the two, producing an extremely strong covariation signal. Looking at a phylogenetic tree of the 993 sequences (Figure 5-4), it is clear that this is more or less the situation we are encountering – multiple copies of essentially only two really distinct sequences.

5.3 MI with average product correction

Mutual information with average product correction (MI-APC or MIp) is a mathematical technique developed by Dunn et al. *(139)* for the purpose of addressing the issues of random MI and phylogeny, two main problems encountered so far in this

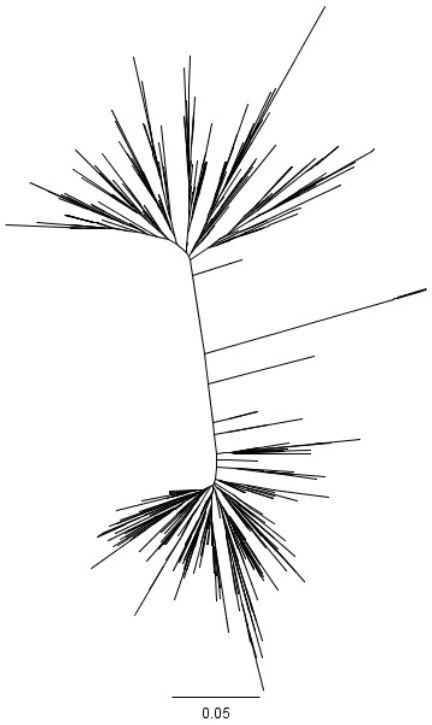


Figure 5-4. Phylogenetic tree for the p7-1ab MSA

Phylogenetic tree showing the clustering of the 993 1a and 1b sequences. Tree was generated by neighbor-joining method.

analysis. The idea is to estimate the contribution of those two sources of error and subtract that contribution from the standard MI calculation to give a corrected measure of covariation for each pair. The correction, APC, is given by

$$APC(i,j) = \frac{MI(i,\bar{x})MI(j,\bar{x})}{\overline{MI}}$$

where

$$MI(i,\bar{x}) = \frac{1}{L-1}\sum_{x\neq i}^{L} MI(i,x)$$

$$\overline{MI} = \frac{2}{L(L-1)}\sum_{x=1}^{L-1}\sum_{y=x+1}^{L} MI$$

The mean mutual information,$MI(i,\bar{x})$, in the numerator of APC represents the average MI of a specific column with every other column in the MSA, where L is the length of the sequences in the MSA. The summation is therefore over all the columns. The denominator is the mean MI of the entire set – every column with every other column. MIp takes the form of MIp = MI – APC. This corrected figure has been shown to have improved performance over MI in contact prediction and detection of alpha-helix stacking *(140)*.

5.3.1 MIp of p7-1a and p7-1b MSAs and the receiver operating characteristic

We first applied the average product correction to the single subtype MSAs of p7, to test its effectiveness in reducing random MI. The MIp hits for the p7-1a and p7-1b MSAs and their scrambled counterparts are shown in Figure 5-5. When compared with the MI results, there is noticeably less overlap between the hits for the scrambled sets and the normal sets. This is especially true for the 1b, which displayed a greater degree of random MI in the earlier analysis. The spurious detection of helix stacking in the p7-1b MSA is also eliminated, which is expected given that those hits were attributed to random MI. The MIp hits for p7-1a tend to cluster near the diagonal, meaning that more of the detected covariation was between residues that were closer to each other in the sequence, and subjectively, the MIp results have a more natural, less grid-like appearance.

There was no obvious improvement over MI in the detection of helix-stacking by MIp as determined by the number of pairs in the top 20 scores with a spacing of 3 or 4 residues. In order to compare these algorithms more rigorously, receiver operating characteristic (ROC) curves were generated for both algorithms acting on the p7-1a MSA and scrambled p7-1a MSA. The ROC is a graphical depiction of the performance of a binary classifier, which plots the true positive rate of that classifier against the false positive rate as the discrimination threshold is varied over its entire range. A good classifier will achieve a high true positive rate while maintaining a low false positive rate,

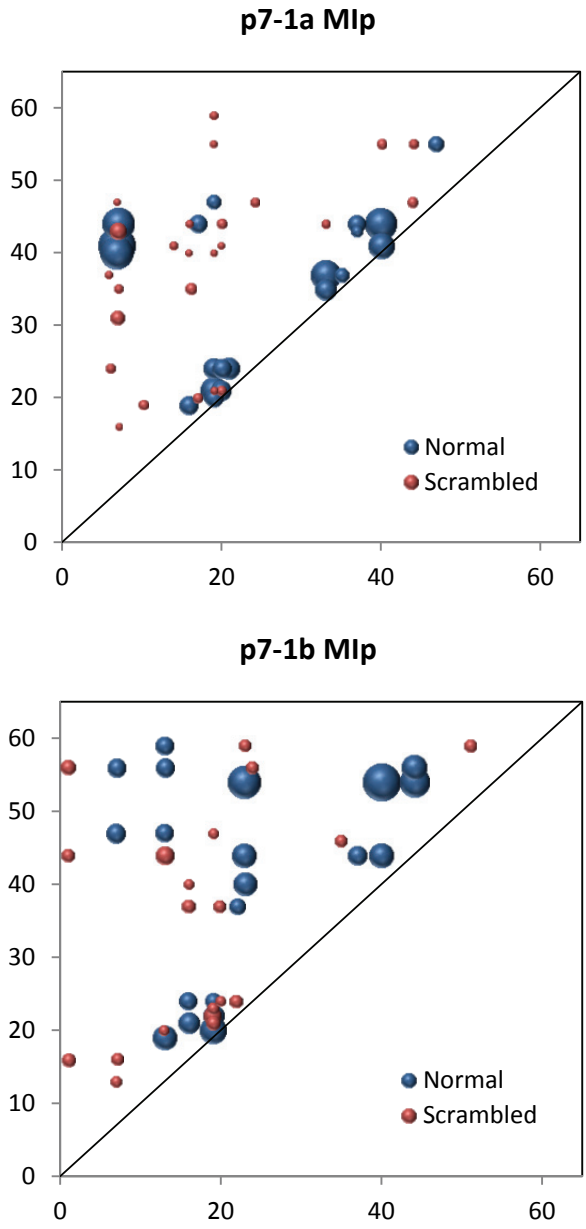
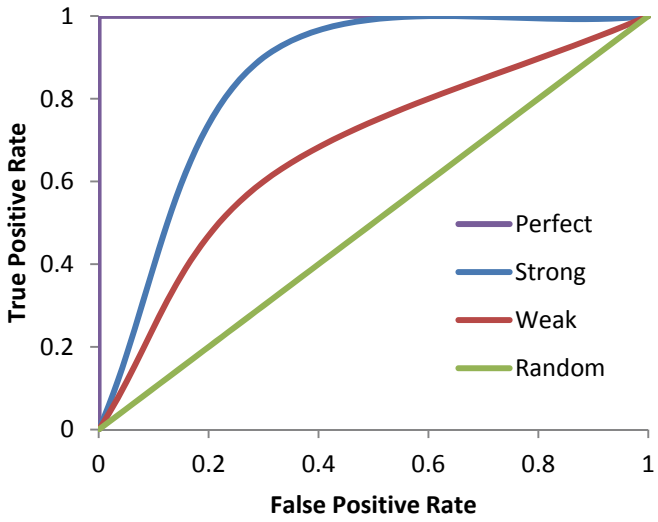


Figure 5-5. MIp for p7-1a and p7-1b

Top MIp hits for p7-1a and p7-1ab MSAs. Results for the scrambled MSAs are also shown for comparison.

A



B

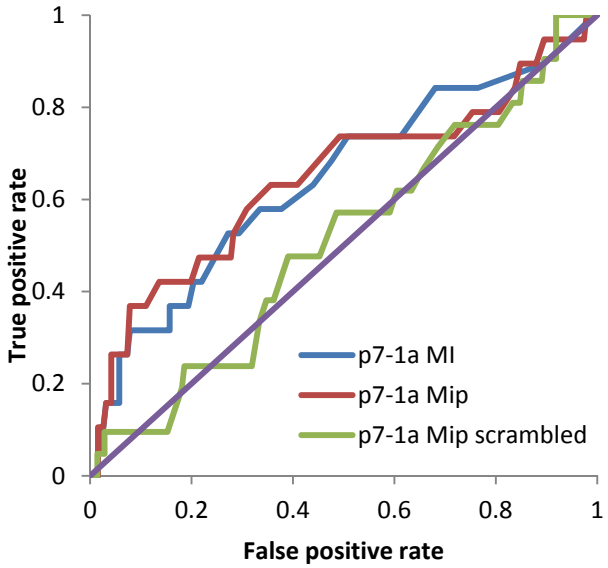


Figure 5-6. Receiver operating characteristic (ROC) curves

(A) A set of hypothetical ROC curves illustrating the difference between strong and weak classifiers. (B) The ROC curves for helix-stacking detection, evaluating performance of MI and MIp on p7-1a MSAs.

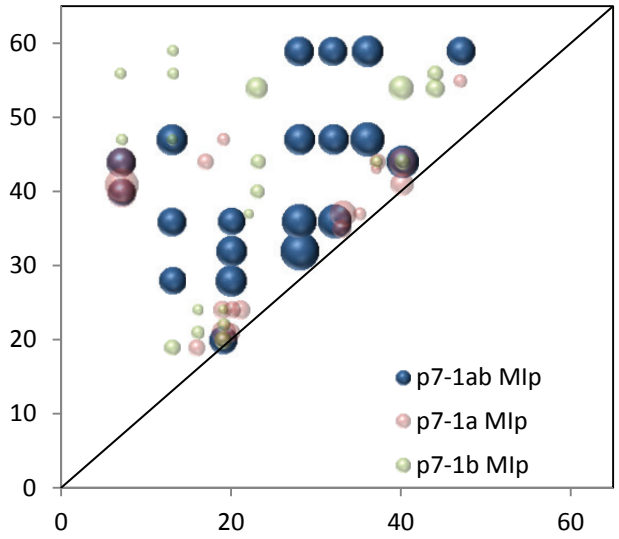
whereas a random classifier will produce many false positives and achieve a true positive rate roughly equal to its false positive rate, producing a line from the origin of slope 1. The ROC curves for p7-1a show MI and MIp nearly identical in performance, with a very small advantage to MIp Figure 5-6. On the scrambled alignment, MIp performs no better than a random classifier, as expected.

5.3.2 MIp for the combined p7-1a and 1b MSA

MIp was calculated for the p7-1a and 1b combined MSA. The results indicate that MIp is not efficiently eliminating the effect of phylogeny, as evidenced by the large overlap (14 of 20) with the uncorrected MI results (Figure 5-3), where the MI results are almost entirely composed of phylogenetic covariation, and little overlap with either the p7-1a results (4 of 20) or the p7-1b results (2 of 20), evident in Figure 5-7a. One would expect that many of the covariant pairs in the two separate sets would still score highly (as they are still covariant) in the combined set. The ROC curves for helix-stacking detection showed a small improvement over MI near the origin (which is the most important region of the curve), but well short of the results from the p7-1a MSA. Generally the latter part of the ROC curve can be ignored, as by that point the false positive rate is very high.

The ineffectiveness of the average product correction to sufficiently diminish the effects of phylogenetic covariation reflects the particular challenges of using covariation

A



B

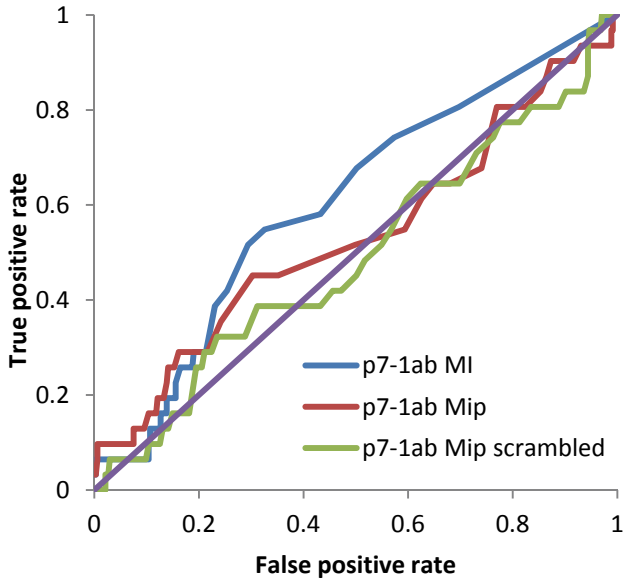


Figure 5-7. MIp for p7-1ab MSA

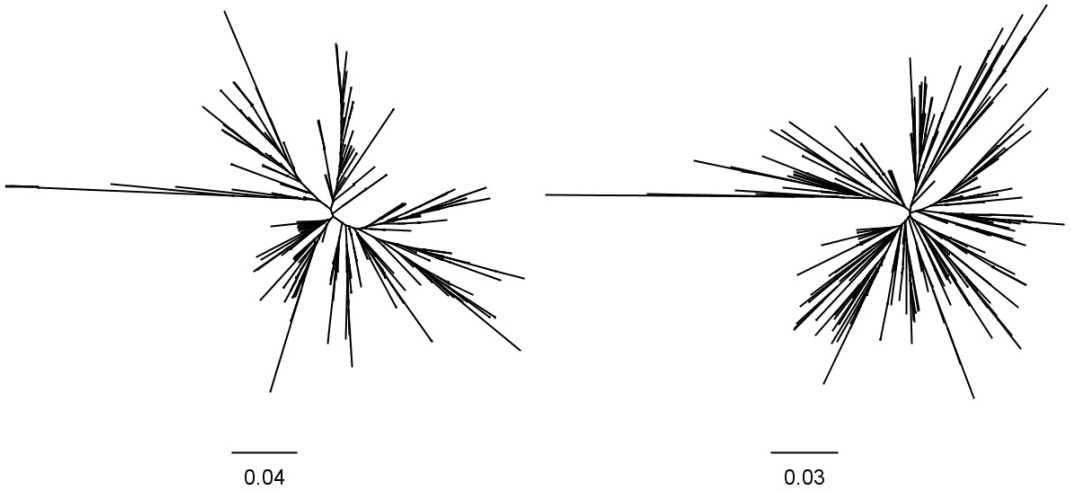
(A) Comparison of hits for combined p7-1ab MIp with MIp hits for the two MSAs separately. (B) ROC curves comparing performance of the MI with MIp for p7-1ab.

analysis to study HCV. When compared with trans-species sequence alignments which are commonly used in such analyses, the HCV MSAs suffer from at least two major limitations: lack of diversity and non-uniform distributions. Figure 5-8 compares phylogenetic trees from the HCV p7 MSAs with a sample MSA used in the study originally describing the average product correction *(139)*, a set of 130 sequences of the methionine aminopeptidase (MetAP) protein from many different bacterial species. The trees illustrate the difficulties well. The lack of diversity is apparent from the relative scale – the MetAP tree is ten times larger than the p7-1a or p7-1b MSA – and is problematic because of the lower overall signal levels and higher susceptibility to random variation. One way to counteract the problem of random variation is to have more sequences within the distribution to average it out. To some extent this strategy is already being exploited with the two p7 MSAs, which contain 455 and 538 sequences, respectively, as compared with 130 sequences in the MetAP MSA. Another way to increase diversity would be to combine the 1a and 1b MSAs; this however creates the second major problem, which is the highly non-uniform distribution of the MSA in sequence space, resulting in overwhelming phylogenetic covariation.

Thus, there is a kind of dilemma: it is desirable to improve MSA diversity, but increased diversity comes at the heavy cost of phylogenetic effects. One common approach to this problem is to cull the sequence set to try to reduce the skewing; this by necessity reduces the number of sequences and increases susceptibility to random

A) p7-1a

B) p7-1b



C) MetAP

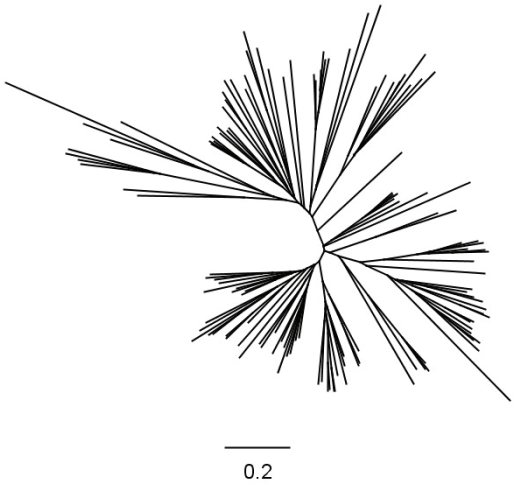


Figure 5-8. Phylogenetic trees

(A) p7-1a tree. (B) p7-1b tree. (C) MetAP tree. All trees generated using neighbor-joining method. Note different scale bars for each tree.

variation. In any case, it can be seen that no amount of culling will produce a star-like phylogenetic tree. Another option is to try to fill in the missing areas; the tree for all p7 sequences downloaded from the euHCVdb is shown.

5.4 Development of a novel method for estimating covariation

In light of the particular challenges involved in assessing HCV covariation, we chose to design an algorithm to cope with these difficulties. The basic approach is, for each sequence, to find the most closely related sequences, and by comparison with those relatives, to assign each letter of the sequence an interest score (IS), which quantifies how "interesting" that amino acid is. After assigning every letter of every sequence an IS, this set of scores can then be analyzed to find correlations between occurrences of interesting residues.

5.4.1 Calculating interest scores

An amino acid in a sequence is considered interesting if, based on its relatives, one would have predicted something else. Therefore, if all the related sequences have a glycine at position 24, finding a glycine at the position would be uninteresting, whereas finding a lysine would be very interesting. There would also be intermediate cases, where a minority amino acid would be somewhat more interesting than the majority amino acid. If the position is highly variable among the related sequences, then whatever amino acid is found there in that sequence, it would not be as interesting as if the position was generally conserved among the neighboring sequences. Further

sophistication can be achieved by taking into account the likelihood of a particular amino acid substitution. Finding a glutamic acid when one would have predicted an aspartic acid would not be as surprising as finding a tryptophan. We chose to calculate the interest score according to the following equation:

$$I(i,n) = \frac{\sum_{n'=0}^{N} 10^{SSF \times S(n,n') + M[s_{ni}][s_{ni}]}}{\sum_{n'=0}^{N} 10^{SSF \times S(n,n') + M[s_{ni}][s_{n'i}]}}$$

The interest score is calculated for a position i in sequence n, and is a ratio between two values. The denominator, which is the main determinant of the score, is the sum of the substitution rates of the amino acid at position (n,i) with the amino acid at that position for every other sequence in the MSA, weighted by the similarity of the that sequence to sequence n. The numerator is the normalizing factor, and is also a sum of substitution rates weighted the same way but the rates are for substituting the amino acid at (n,i) with the same amino acid. Therefore the minimum possible value of I(n,i) is 1, which is the case where position i is completely conserved in all sequences. The largest value of I(n,i) would occur if every sequence differed from sequence n at position i. The matrix M is a substitution matrix, for example, PAM20, which gives the log-odds of seeing a particular substitution. The value S(n,n') is a measure of similarity between two sequences, and is given by sum of the amino acid substitution rates between the two sequences for each position. SSF, the similarity scaling factor, determines how heavily to weigh sequence similarity. It can be thought of as determining how near or far to search

when comparing with related sequences. A large SSF would only consider very closely related sequences, with more distant sequences given very little weight.

5.4.2 Testing interest score calculations

An artificial test MSA was generated by taking one sequence of p7 each from the 1a MSA and the 1b MSA and replicating them multiple times. Then two amino acids were exchanged in one pair, effectively creating an MSA with just two sequences with two "interesting" residues each. A subset of the MSA is shown in Figure 5-9a. Interest scores were then calculated according to the above algorithm for this test case MSA. For all but the two mutated sequences, the interest score was approximately 1 for all residues. The scores for the two mutated sequences are shown in Figure 5-9b. The scores for most of the residues is 1, except for the two mutated residues, which are identified as highly interesting.

5.4.3 Computing covariation from interest scores

After calculating the interest score for every residue in the alignment, the next step is to search the set of scores to find correlated pairs of columns. This could be accomplished a number of different ways; we chose to use a simple correlation function to calculate the correlation between two columns of interest scores. We dubbed the final computed value the interest-score correlation (iCor).

A

```
ALENLVVLNAASLAGVHGLSSFLVFFCAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
ALENLVVLNAASLAGVHGLSSFLVFFCAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
ALENLVVLNAASLAGVHGLSSFLVFFCAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
ALENLVVLNAASLAGVHGLSSFLVFFCAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
ALENLVVLNAASLAGVHGLMSFLVFFCAAWYIKGRLAPGAAYAFYGVWPFLLLLLALPPRAYA
ALEKLIILHSASAASANGPSWFFIFFIAAWYLKGRVVPVATYSVLGLWSLLLLVLALPQQAYA
ALEKLIILHSASAASANGPMWFFIFFIAAWYLKGRVVPVATYSVLGLWSFLLLVLALPQQAYA
ALEKLIILHSASAASANGPMWFFIFFIAAWYLKGRVVPVATYSVLGLWSFLLLVLALPQQAYA
ALEKLIILHSASAASANGPMWFFIFFIAAWYLKGRVVPVATYSVLGLWSFLLLVLALPQQAYA
ALEKLIILHSASAASANGPMWFFIFFIAAWYLKGRVVPVATYSVLGLWSFLLLVLALPQQAYA
```

B

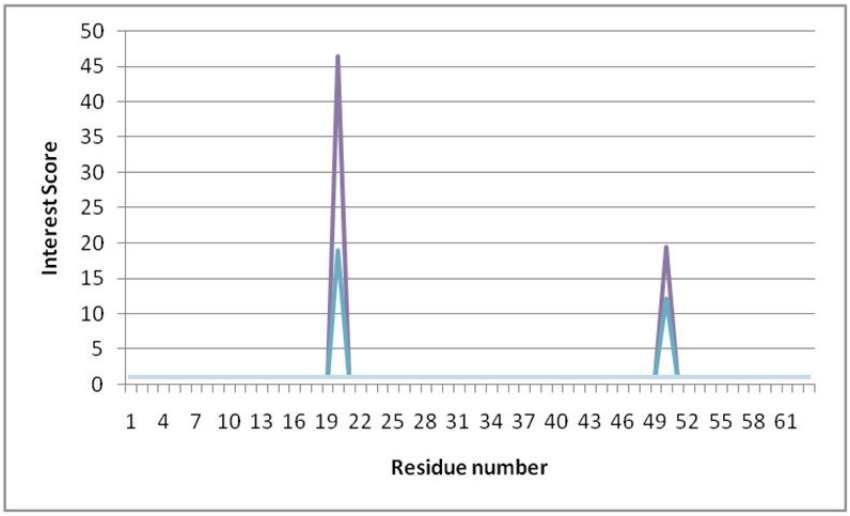


Figure 5-9. Test MSA interest score

(A) Subset of the artificial test MSA. Actual MSA contains 77 sequences, which are all copies of the blue sequences and red sequences. Two sequences contain 2 mutations each, forming a covariant pair. (B) Interest scores for every residue in every sequence. All of the unmutated sequences have scores of 1 across the sequence. The two mutated sequence, shown in violet and teal, show high interest scores corresponding to the two mutations.

$$iCor(i,j) = \frac{1}{N-1}\sum_{n=1}^{N}(I(i,n) - \bar{I}(i))(I(j,n) - \bar{I}(j))$$

Here $I(i,n)$ is the interest score for residue i in sequence n. $\bar{I}(i)$ is the average interest score over all N sequences for column i. The result of this calculation for the test MSA described above is shown in Figure 5-10 and compared with MI and MIp. Because MI has no correction for phylogenetic MI, it detects a large covariation between every pair of residues that varies between the two sequences used to generate the test MSA. The one actual covariation is buried under the phylogenetic background. MIp is much more successful at reducing this background, and the single covariant pair is apparent. However, the iCor algorithm is more effective still at removing the effect of phylogeny, as seen by the lower background.

5.5 iCor results for p7

Interest scores were computed for the p7-1ab MSA. The scores for one representative sequence is shown in Figure 5-11a. In this particular sequence, 3 residues were scored as "interesting": T16, P21, and I44. The higher scores for these three residues indicate that they differ at those points from their related sequences. The array of interest scores was then analyzed for correlation to generate the iCor values for covariation. The results are shown in Figure 5-11b. One salient feature is the extensive overlap with the separate p7-1a and p7-1b results (only a few of the blue dots are not

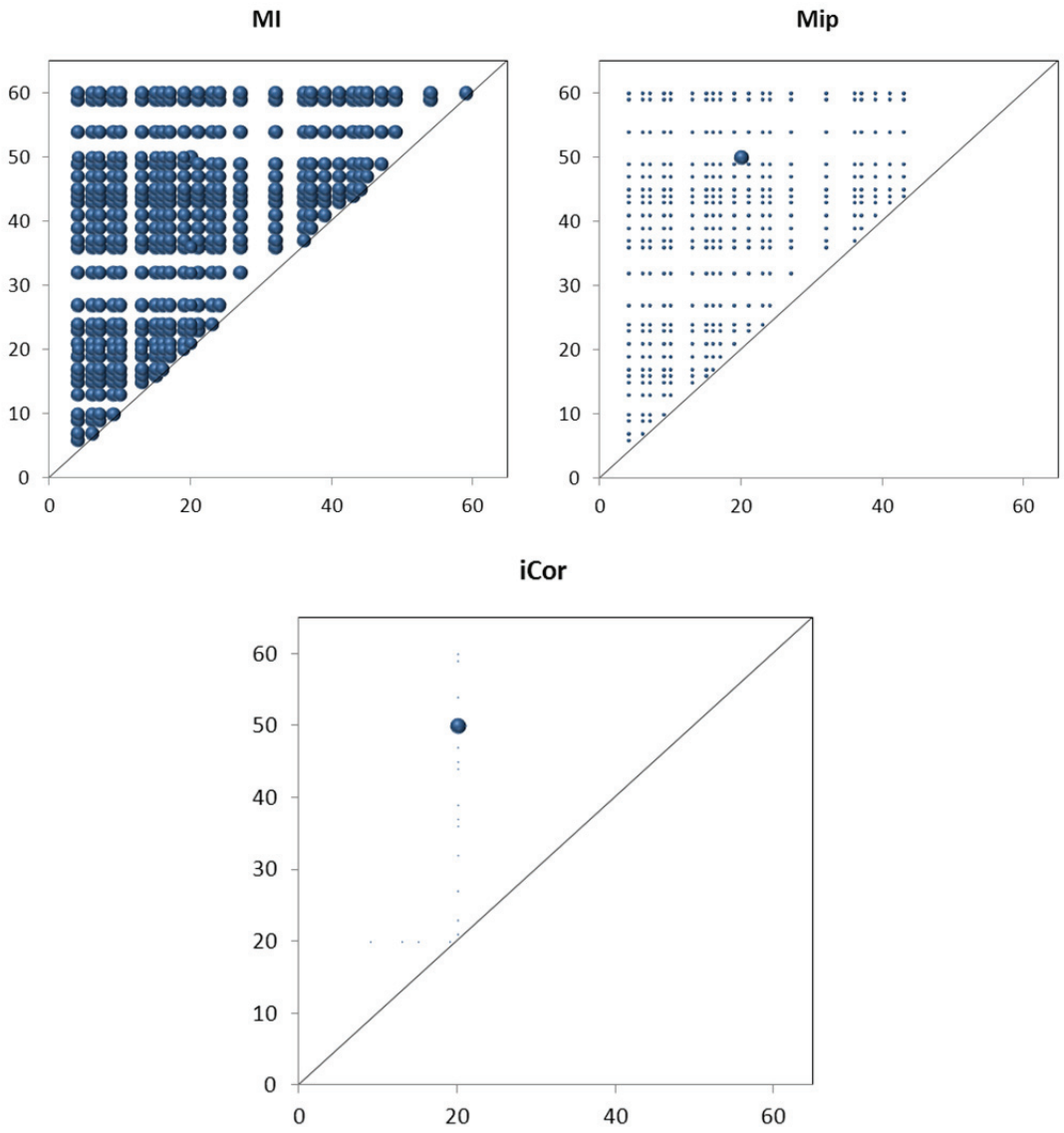


Figure 5-10. Test MSA results

Comparison of three different algorithms for estimating covariance in the test MSA shown in the previous figure. MI and MIp are two published methods, iCor is a novel method.

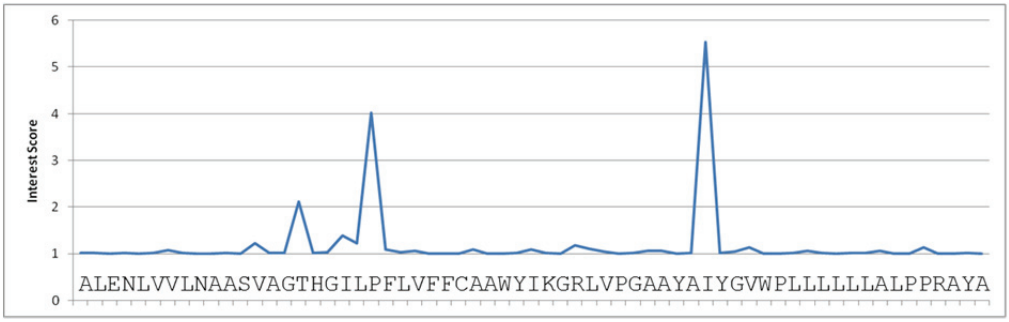
occluded by red or green). This is in contrast to the results from MIp (Figure 5-7a). These results indicate that iCor is able to reduce phylogenetic covariance to the point where underlying covariance signals are detectable.

ROC curves for helix-stacking detection were also computed to test the performance of the iCor algorithm and are shown in Figure 5-12. The results confirm the improvement over MIp for the p7-1ab MSA. Comparing the left-hand-most region of the curve, the performance of iCor on the mixed p7-1ab MSA is actually comparable to that of MIp on the single-genotype p7-1a MSA, which is the best-case scenario thus far observed.

5.6 Discussion

In this study we identified and attempted to solve some of the specific challenges facing covariation analysis of HCV MSAs. In particular, we found that lack of diversity and highly non-uniform phylogenetic distribution to be major concerns when working with HCV MSAs. We attempted to address the problem of phylogeny first by implementing a well-known existing technique, MIp, but found it insufficient to remove the very large phylogenetic component of covariance found in the HCV MSAs. Instead, we developed a novel algorithm for computing covariance by identifying “interesting” residues and searching for correlations in the occurrence of these mutations. This algorithm, iCor was compared with MI and MIp and found to be better suited to dealing with the phylogenetic covariance.

A



B

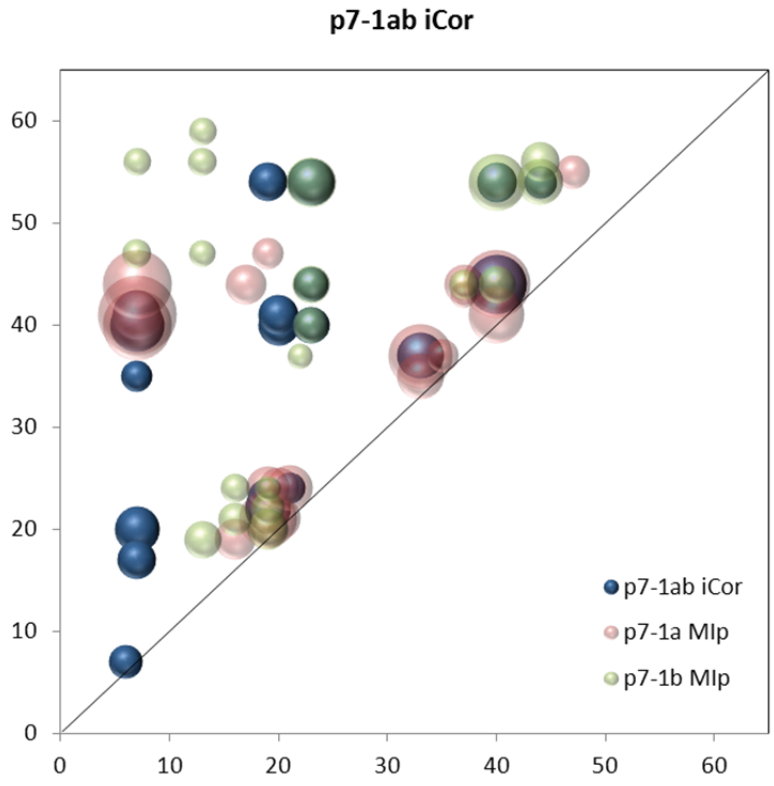


Figure 5-11. iCor results for p7-1ab

(A) Interest scores for a single selected sequence in the p7-1ab MSA. (B) Top hits for the p7-1ab MSA using the iCor method. Covariance scores for p7-1a and p7-1b analyzed separately are overlaid for comparison.

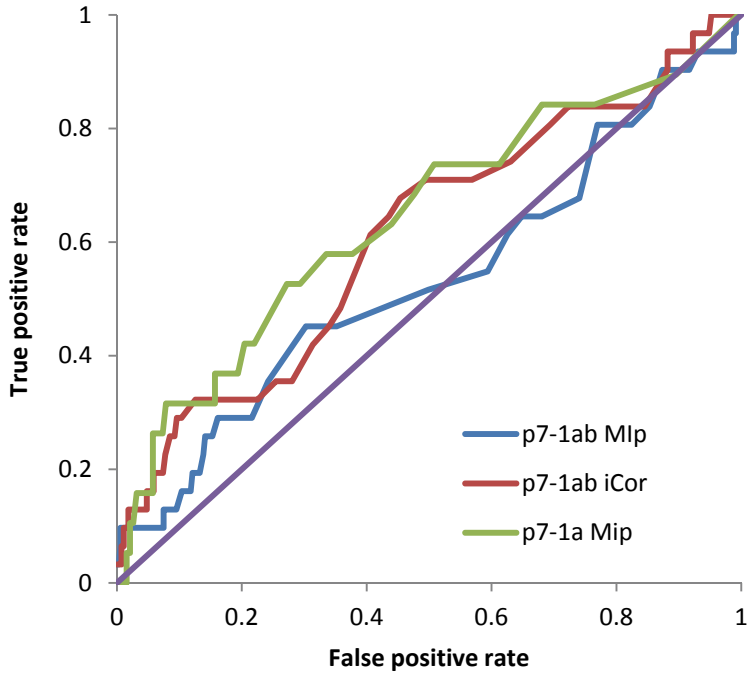


Figure 5-12. iCor helix-stacking ROC curve

ROC curves of iCor and MIp performance in detection of helix-stacking in the p7-1ab MSA. For comparison, the MIp results for p7-1a are also shown.

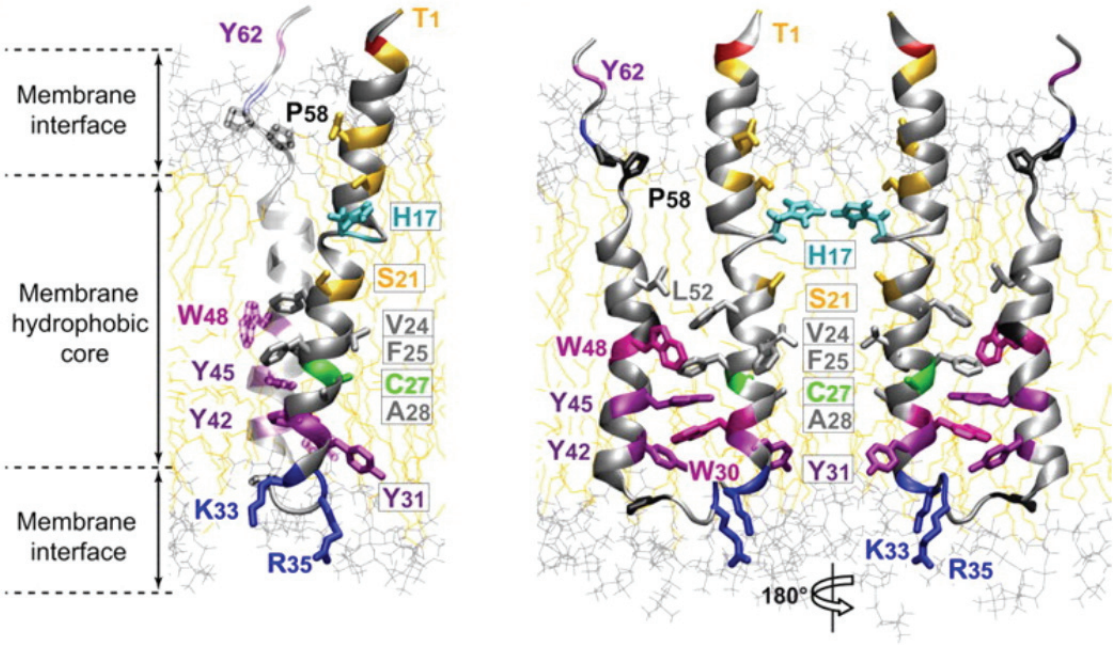


Figure 5-13. Structural view of p7

Illustration showing a tentative structure of the p7 monomer on the left and the oligomeric subunit of the ion channel on the right. From Montserret et al., JBC. 2010; 287: 31446-31461. *(141)*

The next step in evaluating these algorithms and also evaluating the covariance approach in general would be to map covariant pairs onto the crystal structure of a solved protein, such as NS3 or NS5B. A recently published tentative structure for p7 is shown in Figure 5-13 *(141)*. When compared with the iCor results, fully half of the top 20 hits appear to be physically interacting pairs. A larger protein like NS3 would be a more rigorous test; as the volume of the protein increases, the probability of any two randomly selected residues interacting decreases proportionally. In general, efforts to predict pairwise physical interactions by covariance analysis have been largely unsuccessful, and preliminary results with NS5B and NS3 do not show any major breakthroughs in this regard. The question of how to utilize covariance data, even assuming truly covariant pairs can be reliably identified, remains open. Apart from RNA covariance *(142, 143)*, which tends to produce better results because the interactions are explicitly pairwise and there are fewer letters, covariance analyses in HCV have focused on general trends and network effects rather than specific interactions *(144, 145)*. The correct interpretation of this type of observation is not altogether clear. For these reasons, the development of the iCor algorithm represents an incremental step towards a destination which is still being defined.

Chapter 6. Discussion

Chapter 6 Discussion

We have followed three separate avenues towards gaining a more comprehensive understanding of HCV dynamics. We developed numerical models to accurately recapitulate narrowly defined aspects of HCV biology, virus accumulation and virus entry, and applied these models to gaining a more detailed understanding of these life cycle events. We also studied the regulation of HCV through NS5A by chemical and genetic perturbations. Finally, we attempted to apply covariation analysis to the study of HCV, laying some fundamental groundwork and developing novel techniques to advance this area of study.

At the outset of these studies, we had hoped to begin integrating the three approaches developed in chapters 3-5. At this point, they remain separate; however, paths toward convergence can be seen. The most obvious point of convergence is between HCV regulation and numerical models. This particular coupling was seen as the most accessible and had been an aim of the study from early on; however we found that the complexity of NS5A's regulation of HCV put a numerical model outside our reach. What is needed to move forward is rich, quantitative data, and our work suggests several areas for progress in this regard.

We feel that inhibitor studies have tremendous potential which we have only begun to tap. The dose-response curves of HCV to these compounds demonstrate that there is much more information to be gleaned than simply whether a compound is inhibitory

and at what concentration. From the standpoint of drug development it might be beneficial to rate efficacy by a single value in a single consistent assay done the same way every time, for efficient comparison of large libraries of compounds. From the standpoint of unraveling HCV biology on the other hand, we see value in doing many variations of the assay with many different conditions. A dose-response at early times vs. late times or long and short duration of treatment provide different pieces of information which can be integrated, as seen with CsA treatment in particular. Similarly, differences between dose-response curves measuring luciferase expression vs. measuring RNA directly reveal underlying details not only about the mechanism of inhibition, but about HCV itself, as is the case with H479. Even further insight can be gleaned by combining inhibitors in various ways, such as pre-treating cells with different doses of danoprevir before adding a second drug, or simultaneous addition of CsA and H479, which suggested the functional target of CsA might be p58 NS5A.

One type of experiment which should be very powerful is the inhibitor response kinetic. It did not receive much treatment in this work, due to technical issues, although we did show some smaller-scale kinetic experiments for CsA, H479, and 2'CMA. Comparing the kinetic response to a panel of compounds, which would include not only the specific HCV inhibitors already discussed but also broader cellular inhibitors, would provide invaluable information as to temporal relationships between different events. Other inhibitors would include cyclohexamide, brefeldin A, the MTP inhibitor BMS-200150, and kinase inhibitors like TBCA. Preliminary experiments have shown, for

example, that danoprevir treatment results in a more rapid decline of intracellular infectivity than cyclohexamide, which might be taken as evidence for the existence of a pool of uncleaved polyprotein. It would also imply that a constant, fresh supply of at least one of the NS3/4A-processed non-structural proteins is required for ongoing viral assembly. The potential value of these experiments is only increased when we consider combinations of drugs.

The challenges to this approach are largely technical – achieving a high degree of uniformity and repeatability across time points and managing the large number of samples resulting from the multiplication of the number of time points, the number of compounds, the number of concentrations of each compound, the number of assays, and the number of replicates. An additional challenge is presented in synchronizing the experiment temporally for this large number of samples. For example, taking an hour to harvest a set of plates for one time point would not be acceptable if the time points are one hour apart. Obviously a significant amount of workflow optimization is required; automation might also be appropriate.

The introduction of new assays is another exciting area of development to consider with the inhibitor studies. So far, we have measured total HCV RNA, protein expression by luciferase, infectivity, and phosphorylation as they respond to drug treatment. Localization is a more difficult attribute to quantify, but with automated microscopy and image analysis software it is becoming more common to do so. Two brand-new assays are worth mentioning as well, both developed in cooperation with our lab. A PCR-based

strand-specific assay is under development (Joseph Luna, personal communication), which would provide a more detailed characterization of than total HCV RNA. Positive-strand RNA alone is the substrate for polyprotein translation and for virion packaging, whereas both strands are part of the RNA replication process; being able to separately quantify the strands and their response to various inhibitors would shed additional light on all these processes. The second assay is a fluorescent in situ hybridization (FISH) type of assay developed by Salman Khetani (personal communication) for determining HCV localization within the cell. Especially with the model of NS5A primarily regulating RNA targeting, being able to image RNA localization and how it response to, for example, H479 treatment, would be highly informative.

Armed with such a formidable collection of quantitative data, we would be in a much better position to consider more comprehensive numerical models describing the HCV system. The approach would be similar to that taken in chapter 3 but expanded in scale and sophistication, and similarly the aim would be to develop a quantitative description of a specific experimental system which could then be used to make testable predictions and verified. In one sense, the mathematical model not an end in itself, but rather a means of rigorously confirming our understanding of the underlying system.

Where covariation analysis might fit into the broader picture is more difficult to answer. The hope would be to identify important intra- and inter- protein interactions, which might provide insight into some of the molecular mechanics of HCV; in practice, the ability to discover these connections is yet to be realized. One obvious thing to try is

to search an NS5A MSA for covariation. In fact, this analysis was done; the issue is how to utilize the results. No intuitive interpretation presents itself, just a list of putatively covariant pairs. The hits did not clearly link any one region or domain with any other, nor did they cluster together in hot spots. Either a more sophisticated method of interpretation is required, such as the protein sector interpretation *(146)*, or better algorithms to improve the signal-to-noise ratio, or more sequence data. With the rapid advance of sequencing technology, the latter is almost guaranteed. The covariation approach therefore remains something of a dark horse which is of doubtful utility at the moment but may yet yield payoffs in the long-term.

References

1. D. Lavanchy, Evolving epidemiology of hepatitis C virus., *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **17**, 107-15 (2011).

2. S. Munir et al., Hepatitis C treatment: current and future perspectives., *Virology journal* **7**, 296 (2010).

3. C. W. Shepard, L. Finelli, M. J. Alter, Global epidemiology of hepatitis C virus infection., *The Lancet infectious diseases* **5**, 558-67 (2005).

4. M. Gambarin-Gelwan, I. M. Jacobson, Optimal dose of peginterferon and ribavirin for treatment of chronic hepatitis C., *Journal of viral hepatitis* **15**, 623-33 (2008).

5. A. A. Butt, F. Kanwal, Boceprevir and telaprevir in the management of hepatitis C virus-infected patients., *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **54**, 96-104 (2012).

6. A. Kapoor et al., Characterization of a canine homolog of hepatitis C virus., *Proceedings of the National Academy of Sciences of the United States of America* **108**, 11608-13 (2011).

7. S. U. Nielsen et al., Association between hepatitis C virus and very-low-density lipoprotein (VLDL)/LDL analyzed in iodixanol density gradients., *Journal of Virology* **80**, 2418-2428 (2006).

8. P. Gastaminza et al., Cellular determinants of hepatitis C virus assembly, maturation, degradation, and secretion., *Journal of virology* **82**, 2120-9 (2008).

9. E. Scarselli et al., The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus., *The EMBO journal* **21**, 5017-25 (2002).

10. B. Bartosch et al., Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor., *The Journal of biological chemistry* **278**, 41624-30 (2003).

11. P. Pileri et al., Binding of hepatitis C virus to CD81., *Science (New York, N.Y.)* **282**, 938-41 (1998).

12. M. J. Evans et al., Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry., *Nature* **446**, 801-5 (2007).

13. A. Ploss et al., Human occludin is a hepatitis C virus entry factor required for infection of mouse cells., *Nature* **457**, 882-6 (2009).

14. N. R. Sharma et al., Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion., *The Journal of biological chemistry* **286**, 30361-76 (2011).

15. D. M. Tscherne et al., Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry., *Journal of virology* **80**, 1734-41 (2006).

16. G. A. Otto, J. D. Puglisi, The pathway of HCV IRES-mediated translation initiation., *Cell* **119**, 369-80 (2004).

17. I. C. Lorenz, J. Marcotrigiano, T. G. Dentzer, C. M. Rice, Structure of the catalytic domain of the hepatitis C virus NS2-3 protease., *Nature* **442**, 831-5 (2006).

18. S. Welbourn et al., Hepatitis C virus NS2/3 processing is required for NS3 stability and viral RNA replication., *The Journal of biological chemistry* **280**, 29604-11 (2005).

19. V. Lohmann et al., Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line., *Science (New York, N.Y.)* **285**, 110-3 (1999).

20. C. T. Jones, C. L. Murray, D. K. Eastman, J. Tassello, C. M. Rice, Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus., *Journal of virology* **81**, 8374-83 (2007).

21. J. D. Evans, C. Seeger, Cardif: a protein central to innate immunity is inactivated by the HCV NS3 serine protease., *Hepatology (Baltimore, Md.)* **43**, 615-7 (2006).

22. G. Cheng, J. Zhong, F. V. Chisari, Inhibition of dsRNA-induced signaling in hepatitis C virus-infected cells by NS3 protease-dependent and -independent mechanisms., *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8499-504 (2006).

23. S. Ciesek, T. von Hahn, M. P. Manns, Second-wave protease inhibitors: choosing an heir., *Clinics in liver disease* **15**, 597-609 (2011).

24. C. Lin, A. D. Kwong, R. B. Perni, Discovery and development of VX-950, a novel, covalent, and reversible inhibitor of hepatitis C virus NS3.4A serine protease., *Infectious disorders drug targets* **6**, 3-16 (2006).

25. D. Lamarre et al., An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus., *Nature* **426**, 186-9 (2003).

26. D. Egger et al., Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex., *Journal of virology* **76**, 5974-84 (2002).

27. N. Appel, U. Herian, R. Bartenschlager, Efficient rescue of hepatitis C virus RNA replication by trans-complementation with nonstructural protein 5A., *Journal of virology* **79**, 896-909 (2005).

28. F. E. Membreno, E. J. Lawitz, The HCV NS5B nucleoside and non-nucleoside inhibitors., *Clinics in liver disease* **15**, 611-26 (2011).

29. R. R. Deore, J.-W. Chern, NS5B RNA dependent RNA polymerase inhibitors: the promising approach to treat hepatitis C virus infections., *Current medicinal chemistry* **17**, 3806-26 (2010).

30. K. J. Blight, A. A. Kolykhalov, C. M. Rice, Efficient initiation of HCV RNA replication in cell culture., *Science (New York, N.Y.)* **290**, 1972-4 (2000).

31. A. A. Kolykhalov, K. Mihalik, S. M. Feinstone, C. M. Rice, Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo., *Journal of virology* **74**, 2046-51 (2000).

32. T. Yamashita et al., RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus NS5B protein truncated at the C-terminal region., *The Journal of biological chemistry* **273**, 15479-86 (1998).

33. P. Friebe, V. Lohmann, N. Krieger, R. Bartenschlager, Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication., *Journal of virology* **75**, 12047-57 (2001).

34. Y. K. Kim, C. S. Kim, S. H. Lee, S. K. Jang, Domains I and II in the 5' nontranslated region of the HCV genome are required for RNA replication., *Biochemical and biophysical research communications* **290**, 105-12 (2002).

35. P. Friebe, J. Boudet, J.-P. Simorre, R. Bartenschlager, Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication., *Journal of virology* **79**, 380-92 (2005).

36. A. M. I. Lam, D. N. Frick, Hepatitis C virus subgenomic replicon requires an active NS3 RNA helicase., *Journal of virology* **80**, 404-11 (2006).

37. G. Manfroni et al., Inhibition of Subgenomic Hepatitis C Virus RNA Replication by Acridone Derivatives: Identification of an NS3 Helicase Inhibitor, *Journal of Medicinal Chemistry* **52**, 3354-3365 (2009).

38. D. N. Frick, The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target., *Current issues in molecular biology* **9**, 1-20 (2007).

39. K. Li et al., Optimization of Potent Hepatitis C Virus NS3 Helicase Inhibitors Isolated from the Yellow Dyes Thioflavine S and Primuline, *Journal of Medicinal Chemistry* **55**, 3319-3330 (2012).

40. T. L. Tellinghuisen, K. L. Foss, J. Treadaway, Regulation of hepatitis C virion production via phosphorylation of the NS5A protein., *PLoS pathogens* **4**, e1000032 (2008).

41. N. Appel et al., Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly., *PLoS pathogens* **4**, e1000035 (2008).

42. A. Shavinskaya, S. Boulant, F. Penin, J. McLauchlan, R. Bartenschlager, The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly., *The Journal of biological chemistry* **282**, 37158-69 (2007).

43. S. Kim, C. Welsch, M. Yi, S. M. Lemon, Regulation of the production of infectious genotype 1a hepatitis C virus by NS5A domain III., *Journal of virology* **85**, 6645-56 (2011).

44. B. Wölk, B. Büchele, D. Moradpour, C. M. Rice, A dynamic view of hepatitis C virus replication complexes., *Journal of virology* **82**, 10519-31 (2008).

45. K. A. Stapleford, B. D. Lindenbach, Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes., *Journal of virology* **85**, 1706-17 (2011).

46. Y. Ma et al., Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins., *Journal of virology* **85**, 86-97 (2011).

47. V. Jirasko et al., Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly., *PLoS pathogens* **6**, e1001233 (2010).

48. T. G. Dentzer, I. C. Lorenz, M. J. Evans, C. M. Rice, Determinants of the hepatitis C virus nonstructural protein 2 protease domain required for production of infectious virus., *Journal of virology* **83**, 12702-13 (2009).

49. V. Jirasko et al., Structural and functional characterization of nonstructural protein 2 for its role in hepatitis C virus assembly., *The Journal of biological chemistry* **283**, 28546-62 (2008).

50. T. Pietschmann et al., Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras., *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7408-13 (2006).

51. M. V. Pokrovskii et al., Novel mutations in a tissue culture-adapted hepatitis C virus strain improve infectious-virus stability and markedly enhance infection kinetics., *Journal of virology* **85**, 3978-85 (2011).

52. Y. Bungyoku et al., Efficient production of infectious hepatitis C virus with adaptive mutations in cultured hepatoma cells., *The Journal of general virology* **90**, 1681-91 (2009).

53. S. D. C. Griffin et al., The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine., *FEBS letters* **535**, 34-8 (2003).

54. A. L. Wozniak et al., Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production., *PLoS pathogens* **6**, e1001087 (2010).

55. P. Tedbury et al., The subcellular localization of the hepatitis C virus non-structural protein NS2 is regulated by an ion channel-independent function of the p7 protein., *The Journal of general virology* **92**, 819-30 (2011).

56. H. Huang et al., Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins., *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5848-53 (2007).

57. K.-S. Chang, J. Jiang, Z. Cai, G. Luo, Human apolipoprotein e is required for infectivity and production of hepatitis C virus in cell culture., *Journal of virology* **81**, 13783-93 (2007).

58. W. J. A. Benga et al., Apolipoprotein E interacts with hepatitis C virus nonstructural protein 5A and determines assembly of infectious particles., *Hepatology (Baltimore, Md.)* **51**, 43-53 (2010).

59. J. Hwang et al., Hepatitis C virus nonstructural protein 5A: biochemical characterization of a novel structural class of RNA-binding proteins., *Journal of virology* **84**, 12480-91 (2010).

60. L. Huang et al., Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein., *The Journal of biological chemistry* **280**, 36417-28 (2005).

61. J. A. Lemm et al., Identification of hepatitis C virus NS5A inhibitors., *Journal of virology* **84**, 482-91 (2010).

62. M. Gao et al., Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect., *Nature* **465**, 96-100 (2010).

63. D. Qiu et al., The effects of NS5A inhibitors on NS5A phosphorylation, polyprotein processing and localization, *The Journal of general virology* , 2502-2511 (2011).

64. R. A. Fridell et al., Distinct Functions of NS5A in Hepatitis C Virus RNA Replication Uncovered by Studies with the NS5A Inhibitor BMS-790052., *Journal of virology* **85**, 7312-20 (2011).

65. P. Neddermann et al., Reduction of hepatitis C virus NS5A hyperphosphorylation by selective inhibition of cellular kinases activates viral RNA replication in cell culture, *Journal of virology* **78**, 13306 (2004).

66. T. Pietschmann et al., C. M. Rice, Ed. Production of Infectious Genotype 1b Virus Particles in Cell Culture and Impairment by Replication Enhancing Mutations, *PLoS Pathogens* **5**, 14 (2009).

67. H. Tu et al., Hepatitis C virus RNA polymerase and NS5A complex with a SNARE-like protein., *Virology* **263**, 30-41 (1999).

68. A. W. Tai et al., A functional genomic screen identifies cellular cofactors of hepatitis C virus replication., *Cell host & microbe* **5**, 298-307 (2009).

69. X. Puyang et al., Mechanism of resistance of hepatitis C virus replicons to structurally distinct cyclophilin inhibitors., *Antimicrobial agents and chemotherapy* **54**, 1981-7 (2010).

70. K. Morohashi et al., Cyclosporin A associated helicase-like protein facilitates the association of hepatitis C virus RNA polymerase with its cellular cyclophilin B., *PloS one* **6**, e18285 (2011).

71. F. Yang et al., Cyclophilin A is an essential cofactor for hepatitis C virus infection and the principal mediator of cyclosporine resistance in vitro., *Journal of virology* **82**, 5269-78 (2008).

72. U. Chatterji, M. D. Bobardt, P. Lim, P. a Gallay, Cyclophilin A-independent recruitment of NS5A and NS5B into hepatitis C virus replication complexes., *The Journal of general virology* **91**, 1189-93 (2010).

73. S. Ciesek et al., Cyclosporine A inhibits hepatitis C virus nonstructural protein 2 through cyclophilin A., *Hepatology (Baltimore, Md.)* **50**, 1638-45 (2009).

74. J. M. Robida, H. B. Nelson, Z. Liu, H. Tang, Characterization of hepatitis C virus subgenomic replicon resistance to cyclosporine in vitro., *Journal of virology* **81**, 5829-40 (2007).

75. Z. Liu et al., Mutations in the hepatitis C virus polymerase that increase RNA binding can confer resistance to cyclosporine A., *Hepatology (Baltimore, Md.)* **50**, 25-33 (2009).

76. U. Chatterji et al., The isomerase active site of cyclophilin A is critical for hepatitis C virus replication., *The Journal of biological chemistry* **284**, 16998-7005 (2009).

77. Z. Liu, F. Yang, J. M. Robotham, H. Tang, Critical role of cyclophilin A and its prolyl-peptidyl isomerase activity in the structure and function of the hepatitis C virus replication complex., *Journal of virology* **83**, 6554-65 (2009).

78. L. Coelmont et al., DEB025 (Alisporivir) inhibits hepatitis C virus replication by preventing a cyclophilin A induced cis-trans isomerisation in domain II of NS5A., *PloS one* **5**, e13687 (2010).

79. T. L. Foster, P. Gallay, N. J. Stonehouse, M. Harris, Cyclophilin A interacts with domain II of hepatitis C virus NS5A and stimulates RNA binding in an isomerase-dependent manner., *Journal of virology* **85**, 7460-4 (2011).

80. D. Verdegem et al., Domain 3 of NS5A protein from the hepatitis C virus has intrinsic alpha-helical propensity and is a substrate of cyclophilin A., *The Journal of biological chemistry* **286**, 20441-54 (2011).

81. H. Grisé, S. Frausto, T. Logan, H. Tang, A Conserved Tandem Cyclophilin-Binding Site in Hepatitis C Virus Nonstructural Protein 5A Regulates Alisporivir Susceptibility., *Journal of virology* (2012), doi:10.1128/JVI.06641-11.

82. X. Hanoulle et al., Hepatitis C virus NS5A protein is a substrate for the peptidyl-prolyl cis/trans isomerase activity of cyclophilins A and B., *The Journal of biological chemistry* **284**, 13589-601 (2009).

83. A. Kaul et al., Essential role of cyclophilin A for hepatitis C virus replication and virus production and possible link to polyprotein cleavage kinetics., *PLoS pathogens* **5**, e1000546 (2009).

84. J. Witteveldt et al., CD81 is dispensable for hepatitis C virus cell-to-cell transmission in hepatoma cells., *The Journal of general virology* **90**, 48-58 (2009).

85. S. Marukian et al., Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells., *Hepatology (Baltimore, Md.)* **48**, 1843-50 (2008).

86. C. T. Jones et al., Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system., *Nature biotechnology* **28**, 167-71 (2010).

87. T. Kato et al., Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon., *Gastroenterology* **125**, 1808-17 (2003).

88. B. D. Lindenbach et al., Complete replication of hepatitis C virus in cell culture., *Science (New York, N.Y.)* **309**, 623-6 (2005).

89. M. J. Bick et al., Expression of the zinc-finger antiviral protein inhibits alphavirus replication., *Journal of virology* **77**, 11555-62 (2003).

90. R. J. Schilling, S. L. Harris, *Applied numerical methods for engineers using MATLAB and C* (Brooks/Cole, 2000; http://books.google.com/books?id=dSF2QgAACAAJ&pgis=1), p. 715.

91. C. Combet et al., euHCVdb: the European hepatitis C virus database., *Nucleic acids research* **35**, D363-6 (2007).

92. H. Dahari, R. M. Ribeiro, C. M. Rice, A. S. Perelson, Mathematical modeling of subgenomic hepatitis C virus replication in Huh-7 cells., *Journal of virology* **81**, 750-60 (2007).

93. N. Durier, C. Nguyen, L. J. White, Treatment of hepatitis C as prevention: a modeling case study in Vietnam., *PloS one* **7**, e34548 (2012).

94. S. Corson, D. Greenhalgh, S. Hutchinson, Mathematically modelling the spread of hepatitis C in injecting drug users., *Mathematical medicine and biology : a journal of the IMA* (2011), doi:10.1093/imammb/dqr011.

95. J. Guedj, L. Rong, H. Dahari, A. S. Perelson, A perspective on modelling hepatitis C virus infection., *Journal of viral hepatitis* **17**, 825-33 (2010).

96. H. Dahari, B. Sainz, A. S. Perelson, S. L. Uprichard, Modeling subgenomic hepatitis C virus RNA kinetics during treatment with alpha interferon., *Journal of virology* **83**, 6383-90 (2009).

97. H. Dahari, E. Shudo, R. M. Ribeiro, A. S. Perelson, Mathematical modeling of HCV infection and treatment., *Methods in molecular biology (Clifton, N.J.)* **510**, 439-53 (2009).

98. D. D. Ho et al., Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection., *Nature* **373**, 123-6 (1995).

99. A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, D. D. Ho, HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time., *Science (New York, N.Y.)* **271**, 1582-6 (1996).

100. A. S. Perelson et al., Decay characteristics of HIV-1-infected compartments during combination therapy., *Nature* **387**, 188-91 (1997).

101. Y. Shimizu et al., Lipoprotein lipase and hepatic triglyceride lipase reduce the infectivity of hepatitis C virus (HCV) through their catalytic activities on HCV-associated lipoproteins., *Virology* **407**, 152-9 (2010).

102. U. Andréo et al., Lipoprotein lipase mediates hepatitis C virus (HCV) cell entry and inhibits HCV infection., *Cellular microbiology* **9**, 2445-56 (2007).

103. P. Padmanabhan, N. M. Dixit, Mathematical model of viral kinetics in vitro estimates the number of E2-CD81 complexes necessary for hepatitis C virus entry., *PLoS computational biology* **7**, e1002307 (2011).

104. M. M. Gibbons, T. Chou, M. R. D'Orsogna, Diffusion-dependent mechanisms of receptor engagement and viral entry., *The journal of physical chemistry. B* **114**, 15403-12 (2010).

105. D. Moradpour et al., Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes., *Journal of virology* **78**, 7400-9 (2004).

106. J. M. Gottwein et al., Development and application of hepatitis C reporter viruses with genotype 1 to 7 core-nonstructural protein 2 (NS2) expressing fluorescent proteins or luciferase in modified JFH1 NS5A., *Journal of virology* **85**, 8913-28 (2011).

107. A. W. Nguyen, P. S. Daugherty, Evolutionary optimization of fluorescent proteins for intracellular FRET., *Nature biotechnology* **23**, 355-60 (2005).

108. L. J. Anderson, K. Lin, T. Compton, B. Wiedmann, Inhibition of cyclophilins alters lipid trafficking and blocks hepatitis C virus secretion., *Virology journal* **8**, 329 (2011).

109. F. Yang et al., A major determinant of cyclophilin dependence and cyclosporine susceptibility of hepatitis C virus identified by a genetic approach., *PLoS pathogens* **6**, e1001118 (2010).

110. T. Wakita et al., Production of infectious hepatitis C virus in tissue culture from a cloned viral genome., *Nature medicine* **11**, 791-6 (2005).

111. M. Quintavalle et al., Hepatitis C virus NS5A is a direct substrate of casein kinase I-alpha, a cellular kinase identified by inhibitor affinity chromatography using specific NS5A hyperphosphorylation inhibitors., *The Journal of biological chemistry* **282**, 5536-44 (2007).

112. N. Appel, T. Pietschmann, R. Bartenschlager, Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain, *Journal of virology* **79**, 3187-3194 (2005).

113. U. Chatterji et al., HCV resistance to cyclosporin A does not correlate with a resistance of the NS5A-cyclophilin A interaction to cyclophilin inhibitors., *Journal of hepatology* **53**, 50-6 (2010).

114. J. Nakabayashi, A compartmentalization model of hepatitis C virus replication: An appropriate distribution of HCV RNA for the effective replication., *Journal of theoretical biology* **300**, 110-7 (2012).

115. F. Tostevin, Precision of sensing cell length via concentration gradients., *Biophysical journal* **100**, 294-303 (2011).

116. J.-Q. He et al., Selection of housekeeping genes for real-time PCR in atopic human bronchial epithelial cells., *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **32**, 755-62 (2008).

117. K. E. Reed, A. E. Gorbalenya, C. M. Rice, The NS5A/NS5 proteins of viruses from three genera of the family flaviviridae are phosphorylated by associated serine/threonine kinases., *Journal of virology* **72**, 6199-206 (1998).

118. M.-P. Egloff, D. Benarroch, B. Selisko, J.-L. Romette, B. Canard, An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization., *The EMBO journal* **21**, 2757-68 (2002).

119. E. A. Gould, T. Solomon, Pathogenic flaviviruses., *Lancet* **371**, 500-9 (2008).

120. M. A. Muesing, D. H. Smith, D. J. Capon, Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein., *Cell* **48**, 691-701 (1987).

121. S. Y. Kao, A. F. Calman, P. A. Luciw, B. M. Peterlin, Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product., *Nature* **330**, 489-93.

122. J. Sodroski et al., Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat., *Science (New York, N.Y.)* **227**, 171-3 (1985).

123. M. H. Malim, J. Hauber, S. Y. Le, J. V. Maizel, B. R. Cullen, The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA., *Nature* **338**, 254-7 (1989).

124. H. S. Olsen, A. W. Cochrane, P. J. Dillon, C. M. Nalin, C. A. Rosen, Interaction of the human immunodeficiency virus type 1 Rev protein with a structured region in env mRNA is dependent on multimer formation mediated through a basic stretch of amino acids., *Genes & development* **4**, 1357-64 (1990).

125. S. Y. Kim, R. Byrn, J. Groopman, D. Baltimore, Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression., *Journal of virology* **63**, 3708-13 (1989).

126. B. K. Felber, C. M. Drysdale, G. N. Pavlakis, Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein., *Journal of virology* **64**, 3734-41 (1990).

127. U. Göbel, C. Sander, R. Schneider, A. Valencia, Correlated mutations and residue contacts in proteins, *Proteins* **18**, 309–317 (1994).

128. B. T. Korber, R. M. Farber, D. H. Wolpert, a S. Lapedes, Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis., *Proceedings of the National Academy of Sciences of the United States of America* **90**, 7176-80 (1993).

129. S. M. Larson, a a Di Nardo, a R. Davidson, Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions., *Journal of molecular biology* **303**, 433-46 (2000).

130. S. W. Lockless, R. Ranganathan, Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families, *Science* **286**, 295-299 (1999).

131. M. Weigt, R. a White, H. Szurmant, J. a Hoch, T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing., *Proceedings of the National Academy of Sciences of the United States of America* **106**, 67-72 (2009).

132. L. C. Martin, G. B. Gloor, S. D. Dunn, L. M. Wahl, Using information theory to search for co-evolving residues in proteins., *Bioinformatics (Oxford, England)* **21**, 4116-24 (2005).

133. A. A. Fodor, R. W. Aldrich, Influence of conservation on calculations of amino acid covariance in multiple sequence alignments., *Proteins* **56**, 211-21 (2004).

134. J. Felsenstein, Phylogenies and the comparative method, *American Naturalist* **125**, 1-15 (1985).

135. N. Saint, R. Montserret, C. Chipot, F. Penin, Structural and functional analysis of the HCV p7 protein., *Methods in molecular biology (Clifton, N.J.)* **510**, 125-43 (2009).

136. P. Luik et al., The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy., *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12712-6 (2009).

137. T. M. Klingler, D. L. Brutlag, Discovering structural correlations in alpha-helices., *Protein science : a publication of the Protein Society* **3**, 1847-57 (1994).

138. M. Meier, P. Burkhard, Statistical analysis of intrahelical ionic interactions in alpha-helices and coiled coils., *Journal of structural biology* **155**, 116-29 (2006).

139. S. D. Dunn, L. M. Wahl, G. B. Gloor, Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction., *Bioinformatics (Oxford, England)* **24**, 333-40 (2008).

140. J. G. Caporaso et al., Detecting coevolution without phylogenetic trees? Tree-ignorant metrics of coevolution perform as well as tree-aware metrics., *BMC evolutionary biology* **8**, 327 (2008).

141. R. Montserret et al., NMR structure and ion channel activity of the p7 protein from hepatitis C virus., *The Journal of biological chemistry* **285**, 31446-61 (2010).

142. H.-Y. Sun et al., Novel nucleotide and amino acid covariation between the 5'UTR and the NS2/NS3 proteins of hepatitis C virus: bioinformatic and functional analyses., *PloS one* **6**, e25530 (2011).

143. A. Tuplin, Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods, *Journal of General Virology* **85**, 3037-3047 (2004).

144. M. J. Donlin, B. Szeto, D. W. Gohara, R. Aurora, J. E. Tavis, Genome-Wide Networks of Amino Acid Covariances Are Common among Viruses., *Journal of virology* **86**, 3050-63 (2012).

145. R. Aurora, M. J. Donlin, N. A. Cannon, J. E. Tavis, Genome-wide hepatitis C virus amino acid covariance networks can predict response to antiviral therapy in humans., *The Journal of clinical investigation* **119**, 225-36 (2009).

146. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Protein sectors: evolutionary units of three-dimensional structure., *Cell* **138**, 774-86 (2009).