

Hepatitis C Virus Protein NS3 : Interactions with the Liver Host Cell Proteome

A thesis submitted to the Faculty of
Graduate studies in partial fulfillment of the
requirements for the degree of
Master of Science

© Candace D. Rypien
Department of Medical Microbiology and Infectious Disease
University of Manitoba
2006

**THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION**

Hepatitis C Virus Protein NS3 : Interactions with the Liver Host Cell Proteome

by

Candace D. Rypien

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree**

Of

MASTER OF SCIENCE

Candace D. Rypien © 2006

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Acknowledgments

To my supervisor Mike. Thank you for your support, for reading innumerable rough drafts of my work, and for giving me the strength to believe in myself, when no one else does. You truly are one of the most intelligent people that I know.

To my advisory committee: Drs Kevin Coombs and Mike Mowat. My most sincere thanks for all the time and effort you put into my thesis. I am grateful for your guidance, interest and most importantly, for reading and giving comments on all 104 pages. Sorry about that.

To my lab mates: Julie – It has been three long years, of snapping turtles, misery, chocolate, tattoos and maybe a bit of science. Take care. Amanda – I would like to thank you for my Spanish lessons and comic relief. The lab is not the same without you. Edmund – You may have only been here a year, but I really appreciate the good talks, and support that you provided me, despite being so very young and a monk.

To my peers at Health Canada. A huge thank you to the members of my extended lab, I am not sure who made more of an impact on each other. Most importantly, this research would not have been possible without the support of DNA core at the NML for sequencing my many constructs and making all my primers.

To my peers in graduate studies, specifically Charlene (thank you for all the support I asked for, and all the support that I didn't), Bevan (I will book shop with you anytime), and Amy (good luck in your life and finding happiness).

To the many volunteer activities in my life. I owe a lot to my athletes at Special Olympics, my sparks, brownies and pathfinders, my fellow volunteers at the red cross and students union; you kept me sane (well, some of the time), or at least rooted in the real world.

To my friends. No matter where you are right now, you have all been there for me when I have needed you, and sometime when I didn't think I did. In Winnipeg. Chris, Carol, Joanna, Sarah, Marika and Alex. I don't think I would have made it this far, this sane, without you. Cynthia – I really would not be here without all your support and the weekly phone-calls. To my genetics girls, and Scona-ites, we have all taken such different directions in our lives, I am always immensely impressed that the farther we get apart, the more we seem to have in common.

To my family. Trisororia, I look forward to many trips and years to come. Krystal, you are so intelligent and strong. You inspire me to be myself. Michelle, your drive to do what you believe in is awe-inspiring. You help give me the strength to be me. Thank you both. To both sets of my grandparents, thank you for instilling me with goals and drive. To the Rypiens and Lochers, thanks for the support and a wee bit of sarcasms that has kept me warm these three long winters.

To my mother and father. I count myself lucky to have both of you not only as my parents, but as friends. I cannot count the number of phone calls that you have received at every time of day. Thank you for your advice, support and unending love. I really would not be here without you, literally and figuratively.

Finally, to my Aunt Helena, who inspired my research from the start. I only wish I could live my life with half as much inner strength and positive outlook as you do. You are truly an inspiration.

I would like to thank the National Research Council (NSERC), the Alberta Provincial government Heritage Foundation, and the University of Manitoba for personal financial support funding this project.

Table of Contents :

List of Figures.....	iv
List of Tables.....	vi
Abstract.....	vii
I) Introduction	1
1.0. Hepatitis C Virus.....	1
1.1. HCV Replication Models.....	4
1.1.1. Replicons.....	4
1.1.2. Cell Culture Replication of NS3	6
2.0 Structure and Function of NS3	7
2.1. NS3 protease complex Structure and Function.....	7
2.2. Helicase Structure and Function of NS3.....	8
2.3. NS3 Localization	9
3.0 NS3 and the Host Cell.....	10
3.1. Functional Interactions.....	10
3.1.1 NS3 and the host cell immune response	10
3.1.2. Potential Transformation Capability of NS3	12
3.1.2.1. NS3 truncation <i>in vitro</i> and <i>in vivo</i>	13
3.1.2.2. NS3 and p21 ^{waf1}	14
3.1.3. Oxidative Stress	14
3.2. Physical Interactions with NS3	15
3.2.1. NS3 interactions with host cell proteome	15
3.2.1.1 NS3 and Cell Cycle Proteins.....	15
3.2.1.2. NS3 Host Immune.....	15
3.2.1.3. NS3 and Other Proteins	16
3.2.2 NS3 and other HCV viral proteins.....	17
4.0. Studying Protein-Protein Interactions.....	19
5.0. Aims of Research.....	22
II) Materials and Methods.....	23
1.0 General Techniques	23
1.1 Polymerase Chain Reaction (PCR).....	23
1.2 DNA Purification.....	23
1.3. DNA Ligations.....	25
1.4. Preparation of Electro-competent <i>Escherichia coli</i> Cells.....	26
1.5. Electrocompetent Bacterial Cell Transformation	26
1.6 Chemically Competent Cell Transformation.....	27

1.7. Sequencing.....	27
1.8 Sodium Dodecyl Sulphate -Polyacrylamide Gel Electrophoreses (SDS-PAGE)...	28
1.9 Coomassie Staining.....	28
1.10 Silver staining	28
1.11 Western Transfer.....	29
1.12 Western Blot Detection.....	29
2.0 Yeast-2-Hybrid (Y2H) Analysis.....	30
2.1 NS3 Construct Cloning:.....	30
2.2 Yeast Chemically Competent Transformation.....	31
2.3 Yeast Protein Expression	31
2.4 Yeast Protein Extraction	32
2.5 Yeast-2-Hybrid Screen.....	32
2.6. Yeast Colony PCR.....	33
2.7 Bioinformatic Analysis	33
2.8 Yeast Plasmid DNA Extraction	34
3.0 NS3 Recombinant Protein Expression.....	34
3.1 NS3 cloning	34
3.2. Expression of NS3	35
3.3 Purification of NS3p from HCV strain 1a	36
3.4 Large Scale Expression and Purification of NS3 from HCV strain 1b.....	36
3.5 Identification of NS3 using mass spectrometry	37
4.0 Confirmation of Protein-Protein Interactions	38
4.1 Protein Interactions in Yeast.....	38
4.2 Protein Expression in Reticulate Lysates.....	39
4.3 Affinity co-precipitations (HIS Pull Down)	39
III) Results	41
1.0. Yeast 2 Hybrid Screen (Y2H).....	41
1.1. Expression of HCV NS3 in AH109 yeast cells	41
1.2. NS3 Yeast 2 Hybrid Screen	44
1.3. Protein-protein interactions between NS3 and individual prey of interest in AH109 yeast.....	53
2.0. Expression and Purification of NS3.....	57
2.1. Expression of NS3 1a in bacteria.....	57
2.2. Expression of NS3 1b in bacteria.....	66
2.3. Characterization of NS3/4a 1b by mass spectrometry.....	74
3.0. HIS pull-down experiments	80
IV) Discussions	85

1.0. Yeast-2-Hybrid Screen.....	85
1.1 Yeast-2-Hybrid Screen pulls out many non-relevant interactions.....	85
1.2. Confirmation of Y2H protein-protein interactions	87
2.0 Purification of NS3/4A	90
3.0. Co-Precipitation of NS3 with Prey Protein.....	92
4.0 Future work.....	94
5.0. Summary.....	95
V) References.....	96
VI) Appendix 1	104

List of Figures :

I) Introduction :

Figure 1. Hepatitis C virus genome structure 2

Figure 2: Generalized *Flaviviridae* Lifecycle..... 5

Figure 3: Yeast-2-Hybrid (Y2H) Screen..... 21

III) Results :

Figure 4: Expression of NS3 constructs in yeast cells..... 42

Figure 5: Physical Interactions of DBD- and AD- proteins in yeast. 45

Figure 6: Yeast 2 Hybrid (Y2H) screening procedure..... 46

Figure 7: Y2H direct colony PCR amplification. 49

Figure 8: Specificity of Y2H interactions for selected prey protein..... 56

Figure 9: Induction and growth rate of pGEX-2T-NS3 in Escherichia coli BL21(DE3) bacteria cells..... 59

Figure 10: Expression of pHAT-NS3p 1a in Escherichia coli BL21(DE3) cells 61

Figure 11: Expression of HAT-NS3p in the presence of protein chaperone proteins, or human codon biased tRNA bacterial strains..... 64

Figure 12: Expression and purification of pHAT-NS3p 1a in a pG-Tf2 bacterial background..... 67

Figure 13: Expression of pHAT-NS3ps 1a..... 69

Figure 14. Protein alignment HCV NS3 strain 1a and HCV NS3 strain 1b. 71

Figure 15: Expression and purification of pHAT-NS3ps 1b 72

Figure 16: Sequence of the HIS-NS3/4Apep-1b region at the N-terminus. 75

Figure 17: Purification of full-length HIS-NS3/4Apep-1b by Ni chelate and size exclusion chromatography. 76

Figure 18: Detection of HIS-NS3/4Apep-1b with anti-HIS Antibodies..... 77

Figure 19: Identification of HIS-NS3/4Apep-1b by mass spectrometry. 78

Figure 20: Ability of purified HIS-NS3/4Apep-1b to bind radiolabeled prey proteins... 81

Figure 21: Co-precipitation of ³⁵S methionine labeled prey proteins with HIS-NS3/4Apep-1b 83

List of Tables :

II) Materials and Methods :

Table 1: Comprehensive list of primers used in this study..... 24

III) Results :

Table 2: Summary of yeast 2 hybrid prey not further evaluated for interaction with HCV-NS3 based on sequence information..... 51

Table 3: Unique in-frame sequences found in NS3 yeast 2 hybrid screen 52

Table 4: Summary of Y2H specificity interactions between prey constructs and DBD-NS3, DBD-NS5A or DBD-lamin. 55

Abstract

Hepatitis C virus (HCV) infects over 200 million people worldwide with most of these cases leading to chronic infection that can manifest into liver cirrhosis or hepatocellular carcinoma. With limited treatment options with poor efficacy and harsh side effects and no viable vaccine, understanding which host cellular proteins each viral protein interact with could lead to novel therapeutic regimens. One HCV-encoded protein, non-structural 3 (NS3), encodes a multi-functional protein containing both a chymotrypsin-like serine proteinase and a NTPase/helicase activity. Besides playing critical roles in virus polyprotein processing and genome replication, it has also been shown to play roles in cellular transformation, immune response and oxidative stress. The mechanism by which these cellular processes are effected, however, are still unknown. In order to better understand the role that NS3 plays in cellular events, an yeast two-hybrid (Y2H) screen was performed in order to identify novel NS3-host cell protein interactions.

As a control in the Y2H screen, NS3 was analyzed for its ability to interact with itself. It was determined through Y2H and co-precipitation experiments that HCV NS3 forms a homodimer, an interaction mediated through its N-terminally located protease domain. This finding is in contrast to a previously published report suggesting dimerization through the helicase domain of NS3.

For the Y2H screen, 3.9×10^5 yeast colonies were screened giving a total of 425 colonies (0.11%) which showed growth suggesting a positive interaction between NS3 and a host cell protein. A high-throughput yeast colony PCR amplification/sequencing protocol provided sequence information for 98% of these isolates which allowed for almost half to be eliminated as false positive interaction sequences. In total, 75 unique sequences were identified and 23 were selected for further study. The first method of examination was to determine if the prey protein interacted specifically with NS3 and 5 of 23 interaction sequences were shown to absolutely require the presence of NS3 within the Y2H system. In order to confirm these findings, expression and purification of a recombinant hexahistidine-tagged NS3 (HCV strain 1b) protein was achieved and used as bait for *in vitro* co-precipitation experiments. Human haptoglobin, one of the proteins found to specifically bind NS3 in the Y2H system, was also found to interact with the virus protein in a co-precipitation analysis. This liver secreted protein has roles in immune response and oxidative stress and alterations to its abundance has been noted in HCV infected patients. It is possible that the interaction with NS3 may be in part responsible for these observations.

D) Introduction

1.0. Hepatitis C Virus

Hepatitis C Virus (HCV) chronically infects over 200 million people worldwide (Shepard *et al* 2005) often leading to a disease state that can culminate in liver cirrhosis, liver fibrosis, digestive tract hemorrhaging, or liver carcinoma. In Canada, it is estimated that HCV infects approximately 240 000 people, 70% of whom are undiagnosed (Remis, R.S. *et al*, 2004). Current treatment of HCV is a combination of PEGylated interferon- α and ribavirin, but overall the antiviral therapy is expensive, associated with toxicity and is only successful in approximately 40% of infected individuals (Shepard *et al* 2005). There is no preventative method of controlling HCV infection.

HCV is in the viral family of *Flaviviridae* and the only member of the genus *Hepacivirus* (Poynard *et al* 2003). *Flaviviridae* have an enveloped nucleocapsid which contains a positive-polarity RNA genome. There are 6 genotypes of HCV and numerous subtypes that differ in their nucleotide sequence by 31-34% (Pawlotsky *et al* 2003). Strain specificity has a great effect on pathogenicity and treatment efficacy. The most common genotype in North America and Western Europe is genotype group 1, which accounts for 60% of HCV globally. Genotype 1a predominates in Northern Europe and North America, and genotype 1b predominates in Southern and Eastern Europe and Japan (Shepard *et al* 2005). Combination therapy is often ineffective against these strains making preventative measures and alternative therapies necessary (Pawlotsky *et al* 2003). Genotype 3 HCV is endemic in South-East Asia, whereas genotype 4 is found primarily in the Middle-East, Central Africa and Egypt. Genotype 5 is found primarily in South Africa and genotype 6 is found in Asia (Mondelli *et al* 1999).

HCV has a 9.6 kb genome which encodes an approximately 3000 amino acid polyprotein (3008-3037aa) (Poynard *et al* 2003). This protein is cleaved by viral and cellular proteases to yield at least 10 unique polypeptides (Figure 1). The genome is composed of three structural genes: core (C), and the viral glycoproteins (E1 and E2). These genes are followed by 6 nonstructural genes: NS2, NS3, NS4A, NS4B, NS5A and NS5B. The tenth protein, p7, is found after E1 and E2 but it is unclear whether this

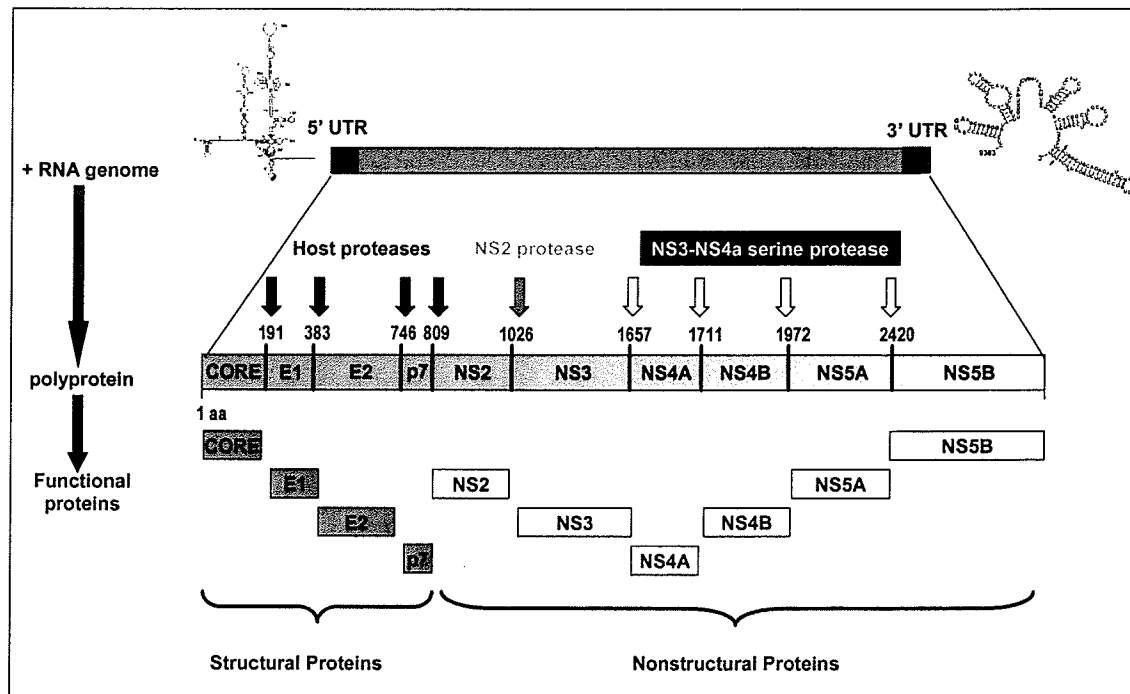


Figure 1 : Hepatitis C genome.

Figure 1. Hepatitis C virus genome structure

The HCV genome is a (+) sense RNA genome. The 5' untranslated region (UTR) terminus contains an IRES structure and the 3' UTR contains a highly conserved folding sequence/structure known as an X-box. Once inside the cell's cytoplasm, the RNA genome is directly translated into a long polyprotein using the host cell machinery. This polyprotein is cleaved co- and post translationally by host cell and viral (NS2 and NS3/4A) proteases. There are four structural genes located in the 5' portion of the polyprotein: core, E1, E2 and p7, and 6 non-structural genes located in the 3' portion of the polyprotein: NS2, NS3, NS4A, NS4B, NS5A and NS5B. (Figure is modified from Lindenbach, B. D. *et al* 2001).

protein has a structural role or not, although there is some evidence for a role as a cationic channel (Griffin *et al* 2003). Recently, an eleventh protein, termed the F protein, has been found. The F protein is produced by a ribosomal frame-shift (Xu *et al* 2003), but the role that it plays in the viral life cycle remains unknown. The 5' untranslated region (UTR, 340 nucleotides) consists of an internal ribosome entry sequence (IRES) which is involved in cap-independent initiation of translation. The 3' UTR contains an X box structure whose highly conserved nature suggests an important role in viral replication, although the precise role remains unknown.

The N-terminal portion of the polyprotein consists of structural proteins. E1 and E2 are highly glycosylated type I transmembrane/class II fusion proteins which contain an N-terminal ectodomain and a C-terminal hydrophobic membrane domain (Dubuisson *et al* 2002). These proteins form stable non-covalent heterodimers. E2 is involved in the attachment of the virus to the host cell, and E1 is involved in the fusion process between the virus and host cell. The C or core protein is an RNA binding phosphoprotein and is the main constituent of the nucleocapsid. The C protein has been implicated in many processes besides packaging of the viral genome. These include altered signal transduction cascades, effecting cellular transformation, inhibition and activation of apoptosis among others (Watashi *et al* 2003). The p7 protein forms an ion channel and is thought to be a viroporin (Carrere-Kremer *et al* 2002) involved in enhancing membrane permeability in the later stages of virus assembly.

There are 6 nonstructural proteins of HCV which are involved in replication and viral pathogenesis, NS2, NS3, NS4A, NS4B, NS5A and NS5B. NS2 is a protease which self-cleaves the polyprotein at the NS2-NS3 junction (Penin *et al* 2004). NS3 is a proteinase which cleaves the polyprotein at all the sites between NS3 and NS5B in conjunction with its cofactor NS4A (Fig. 1). NS3 also has helicase and ATPase activities which allow it to unwind RNA (and DNA) secondary structures. NS4B is a membrane protein involved in the formation of an ER membranous web which helps to bring HCV viral proteins and host cell machinery together during the replication process (Penin *et al* 2004). NS5A is a phosphoprotein whose precise role in replication remains undefined. There is some evidence to suggest a role in resistance to interferon (IFNs) (Foy *et al*

2003). Finally, NS5B is an RNA dependant RNA polymerase required for the replication of the HCV RNA genome.

1.1. HCV Replication Models

Until recently, there was no method to replicate HCV efficiently in human cell culture leading to a poor understanding of the HCV replication cycle (Zhong *et al* 2005). Therefore a generic *Flaviviridae* replication cycle was proposed based primarily on knowledge of other *Flaviviridae* (Fig. 2). The virus is uptaken via receptor mediated endocytosis (Fig. 2.1) and then uses host cell machinery to translate the (+) sense RNA genome into a large polyprotein which is cleaved by host and viral proteases (Fig 2.2). The viral protein NS5A then reverse transcribes the (+) RNA genome making a (-) RNA template which is used to synthesize full length RNA genomes for packaging (Fig. 2.3). The viral genomes and proteins are then packaged together creating virus progeny which gain a lipid membrane (Fig. 2.4) before they exit the cell via exocytosis (Fig. 2.5)

1.1.1. Replicons

The lack of an efficient and reliable cell culture system or small animal model that supports the replication of all HCV genotypes has hampered research regarding the fundamental biology of the virus (reviewed in Pietschmann *et al* 2003). Consequently, until very recently, the main method of studying virus replication has been through replicon systems (Lohmann *et al* 2003). In this system, the structural genes (C-p7) and NS2 from HCV stain 1b were replaced with a neomycin resistance cassette, followed by the encephalomyocarditis virus (EMCV) IRES sequence. This leads to the creation of a bicistronic construct, whereby the neomycin resistance cassette is expressed by the HCV IRES, and the HCV nonstructural genes are expressed from the EMCV IRES. Although this subgenomic replicon recapitulates events in the virus life cycle it cannot provide information on entry and virus assembly. Replicons containing the entire strain 1b coding sequence have been generated but these do not produce infectious particles.

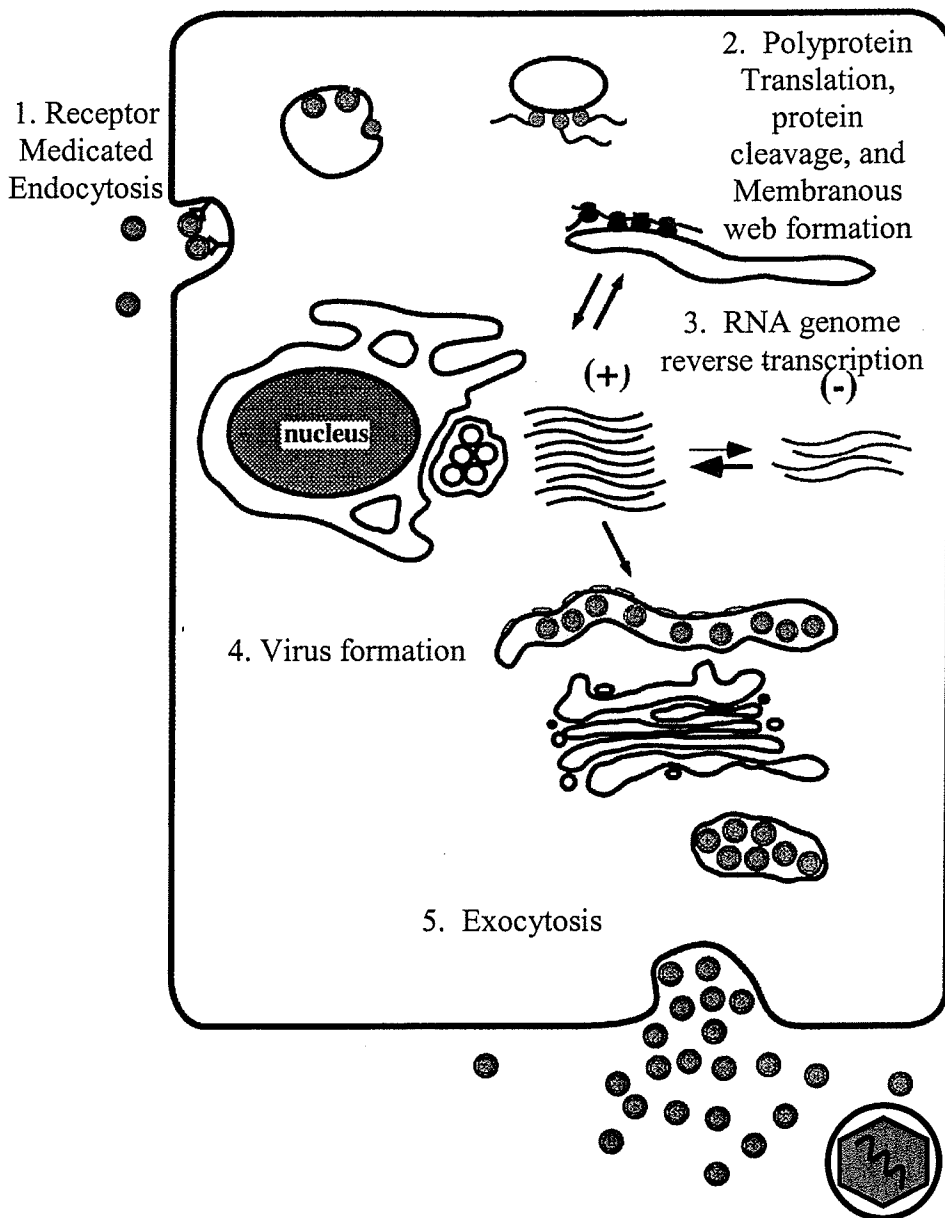


Figure 2: Generalized *Flaviviridae* Lifecycle

The *Flaviviridae* life cycle is similar among all members of the class. It involves the (1) uptake of the virus via receptor mediated endocytosis . (2) Host cell machinery is used to directly translate the viral (+) sense RNA genome into a polyprotein which is co- or post-translationally processed via viral and host proteases. These viral proteins assemble a replication complex on membranous web regions. (3) The (+) RNA genome is reverse transcribed using the viral RNA dependant RNA polymerase (NS5B) to produce negative stranded RNA molecules which are in-turn used as template to synthesize (+) RNA genomes for packaging . (4) RNA is then packaged into a nucleocapsid and (5) leave the cell via exocytosis. (adapted from Lindenbach, B. D. *et al* 2001)

This model system has led to the discovery of numerous cell culture adapted mutations in the replicon which allow for enhanced ability to initiate replication in Huh7 human cell lines. The majority of these sites are found in NS5A and NS3 (Krieger *et al* 2001) suggesting the importance that these proteins play in regulating HCV replication and possibly interacting with the host cell proteome.

1.1.2. Cell Culture Replication of NS3

High levels of RNA replication are attainable in human Huh7 cells, especially as the replicons mutate, or acquire cell culture adaptive mutations. These mutations are found in a cluster at the center of NS5A, a second cluster is found at the C terminus of NS3, and a third group is found in NS4B (Lohmann *et al* 2003). For NS3, specific alterations were found at the amino acid sites 176 and 254 in the protease and helicase domains respectively (Krieger *et al* 2001). A further cell cultured adaptation was found to occur at residue 470 in the helicase domain of NS3 (Grobler *et al* 2003). This mutation, when in the presence of the NS5A-S232I cell cultured adaptation, is necessary and sufficient to confer replication capability to previously replication-incompetent replicons. These mutations specifically allow replicons made from genotype 1b HCV-BK and genotype 1a HCV-H77 (Grobler *et al* 2003) to replicate in cell culture. Interestingly, NS3 cell cultured adaptations alone do not significantly increase replication fitness in cell culture, however, when combined with the change in NS5A, this leads to more efficient replication than either the NS3 or NS5A adaptation on their own.

Recently, HCV JFH-1 (genotype 2a) was demonstrated to replicate (up to 4.6×10^4 focus-forming units/mL, ffu/mL) in Huh7 derived cell lines (Zhong *et al* 2005). These virus particles were found not only to be infectious in human cell lines (Lindenbach *et al* 2006), but also in chimpanzee studies (Wakita *et al* 2005). Although this discovery greatly aids in our ability to study HCV, genotype 2a is a relatively rare genotype. A more common genotype, HCV H77-S (genotype 1a Hutchinson strain), has also been “induced” to replicate and produce infectious virus in highly permissive Huh7.5 cells when the cell culture adapted mutations discovered in the subgenomic replicon system were introduced into them (Yi *et al* 2006). This 1a strain virus did not

replicate as efficiently as the 2a strain, demonstrated by the H77-S (1a) virus exhibiting a specific infectivity of 5.4×10^4 RNA copies/ffu, whereas the JFH-1 (2a) virus specific infectivity was 1.4×10^2 RNA copies/ffu.

Although a cell culture model system has been discovered, a small animal model for HCV does not exist. The chimpanzee studies have provided invaluable information including the first transmission of the virus, proving the transmissible nature of the disease (Alter *et al* 1989), and the generation of high titres of serum needed for the first HCV genome clone (Choo *et al* 1989). The clinical courses of the disease in chimpanzees and humans are similar, allowing in part the study of disease progression but research is limited by the ethics, costs and variability associated with working with chimpanzees.

2.0 Structure and Function of NS3

HCV NS3 is an approximately 67kDa multifunctional protein playing roles not only in viral polyprotein processing and replication, but also in host cell function modulation. The N-terminal one third of the protein is made up of a serine-protease which is responsible for the cleavage between most of the non-structural proteins (NS3/4A, NS4A/4B, NS4B/5A and NS5A/5B (Bartenschlager *et al* 1993, Reed *et al* 1995). The serine protease activity is enhanced by the cofactor NS4A which forms a complex with NS3 (Lin *et al* 1995). The helicase functions in the replication of the viral genome. Crystallization studies have confirmed a physical interaction between NS3 and its cofactor NS4A (Kim *et al* 1996). Furthermore they emphasize the importance a zinc ion plays in the protease domain. Each domain relies on the other for complete activity (Gu *et al* 2005); however, the protease can function on its own with somewhat decreased activity.

2.1. NS3 proteinase complex Structure and Function

The protease function of NS3 is well characterized and ample research has been devoted to the study of this proteinase as a drug target. NS3 is a chymotrypsin-like serine

proteinase which requires NS4A for full activity. The catalytic triad consists of the amino acids His-57, Asp-81 and Ser-139 found in the amino-terminal portion of NS3 (Miller *et al* 1990). NS4A complexes with NS3 in a 1:1 ratio forming a tight interaction which occurs between the 22 amino-terminal residues of NS3 with a 12-residue sequence in the center of NS4A (Bartenschlager *et al* 1995). The presence of NS4A induces a conformational change in NS3 as seen by comparative crystallography studies (Kim *et al* 1996, Love *et al* 1996, Yan *et al* 1998). When NS3 is crystallized alone, the 28 amino-terminal residues are flexible and extend away from the protein. However, when NS3 is complexed with NS4A, these amino acid residues fold into a β -sheet and α -helix and NS4A forms a β -sheet which intercalates into NS3 (Love *et al* 1996, Yan *et al* 1998). For full catalytic activity, NS3/4A also requires the presence of a Zn^{2+} atom. This metal ion is complexed with three cysteine residues (Cys 97, 99, 105), a histidine residue (His149) and a water molecule and plays an important structural role in NS3. Mutations to those residues also render the protein insoluble during recombinant protein production (De Francesco *et al* 1996).

NS3/4A cleaves all the sites between the non-structural proteins NS3-NS5B with different cut site preferences. The first cleavage event occurs in *cis* between NS3 and NS4A, while the remaining cleavage events are mediated in *trans*. The most rapid cleavage event for the remaining polyprotein is between NS5A and NS5B, liberating the NS5B RNA polymerase (Bartenschlager *et al* 1994). The crucial residue determinants in the cleavage site include an acidic residue [Aspartic acid (D) or Glutamic acid (E)] at position P6, a cysteine or threonine at site P1 during *trans* cleavage events, and a small chain amino acid at position P1' (Grakoui *et al* 1993) giving a consensus sequence of D/E-X-X-X-X-C/T S/A-X-X-X. The very general nature of this recognition sequence suggests that other factors including tertiary structure between proteinases and substrate are heavily involved.

2.2. Helicase Structure and Function of NS3

The NS3 helicase can unwind dsRNA, dsDNA and RNA/DNA heteroduplexes using energy generated from an ATPase also found in NS3 (Kadare *et al* 1997). It is not

readily apparent why this helicase can unwind DNA/DNA helicases as HCV has an RNA genome. The HCV helicase belongs to the super family 2 of helicases characterized by a conserved seven amino acid motif (Kadare *et al* 1997). The minimal functional domain is approximately 400 amino acids between residues 1209 and 1608 of the HCV polyprotein (Kanai *et al* 1995). Although ATP is the preferred nucleotide molecule for the ATPase activity, NS3 can hydrolyze all ribonucleotides and deoxynucleotides (Jin *et al* 1995). As well, activity requires a divalent metal ion such as Mg^{2+} or Mn^{2+} . For full activity, it also requires the cofactor NS4A (Pang *et al* 2002). Not only does the helicase domain interact with NS4A, but there is some evidence for dimerization of isolated helicase domains (Khu *et al* 2001). In single-cycle conditions where rebinding of substrate is stopped by excess competing oligonucleotides, full length NS3 was an efficient DNA helicase, but a poor RNA helicase. When NS4A was added, RNA helicase activity was increased substantially. This suggests that other than a structural role, NS4A might function with the NS3 helicase to load the RNA into the binding site. Although structure may suggest that the protease and helicase domains function independently, this is clearly not the case since both NS4A and an intact protease domain are essential for full helicase activity (Frick *et al* 2003).

2.3. NS3 Localization

Although the HCV replication cycle is believed to occur exclusively in the cytoplasm, there are several reports in the literature of viral proteins being found within cellular organelles including the nucleus. NS3 has been found to localize in 3 distinct patterns in HeLa cells (Deng *et al* 2006). Both the protease domain and the full length NS3 from patient isolates were examined and NS3 was found either concentrated in both the nucleus and cytoplasm, present diffusely in the cytoplasm or a combination of both patterns. The propensity for NS3 to localize to the nucleus in dot-like patterns was augmented by the ability of the NS3 sequence to bind p53, and was disrupted by the presence of NS4A. Co-expression of NS3 and NS4A was found to direct NS3 to the endoplasmic reticulum (ER) (Wolk *et al* 2000). Site directed mutagenesis found that targeting of NS3 to the ER required the hydrophobic amino terminus of NS4A (Wolk *et*

et al 2000) and confirms that NS4A likely plays a role in sequestering NS3 to the membranous web during viral infections. NS3 has also been found to localize to the mitochondria in chronic hepatitis C patients (Kasprzak *et al* 2005), and in HCV replicon expressing Huh7 cells (Normura-Takigawa *et al* 2006). The role that NS3 plays in these locations remains to be defined.

3.0 NS3 and the Host Cell

3.1. Functional Interactions

3.1.1 NS3 and the host cell immune response

Successful clearance of many viral infections requires both a strong humoral and cellular response. In HCV infections, NS3 plays a role in inducing and modulating the host immune response. In acute hepatitis, clearance of the virus is associated with a strong T helper cell response against HCV core, NS3 and NS4 antigens (Diepolder *et al* 1995). Specifically, the amino acids 1251-1259 in NS3 elicit a strong antibody response (Diepolder *et al* 1995). NS3 and core protein have been found to activate monocyte and myeloid dendritic cells inducing production of anti-inflammatory cytokine IL-10 (Dolganiuc *et al* 2003). During viral infections however, many patients fail to resolve acute infections and develop chronic hepatitis. In both chronically infected patients (Cerny *et al* 1995) and chimpanzees (Erickson *et al* 1993), a limited cytotoxic T-cell response exists which is insufficient to clear the virus, but strong enough to stimulate liver inflammation.

NS3 also appears to play a role in hiding the virus from immune detection possibly through functional interactions with caspases and the IFN pathways. Apoptosis is thought to be one of the causes of pathogenesis in HCV disease progression. It has been found that the NS2/NS3 precursor polyprotein induces activation of caspases and thereby apoptosis (Prikhod'ko *et al* 2004). This effect could be inhibited by a caspase-8 inhibitor, and studies found that NS3 co-immunoprecipitated with caspase-8. Cell lines HepG2 and Vero stably expressing NS3 showed increased Fas-induced cell death and

mutagenesis experiments demonstrated that apoptosis was independent of the NS3 protease or helicase activity (Prihod'ko *et al* 2004).

Interferon regulatory factors (IRFs) are transcription factors that initiate an antiviral state in the host cells. These states are induced by a variety of factors, including RNA intermediates and viral products, which activate viral-activated kinases (VAK) (Smith *et al* 2001). VAK, in-turn, phosphorylates and activates IRF-3, a latent cytoplasmic transcription factor. The phosphorylated IRF-3 enters the nucleus and induces the upregulation of interferon stimulated genes (ISGs) including double stranded RNA dependent protein kinase R and 2,5-oligoadenylate synthase (OAS) (Reich *et al* 2002). In chimpanzee studies, many of these ISGs have been found to be upregulated upon infection with HCV (Bigger *et al* 2001). Foy *et al* (2003) found that HCV NS3/4A blocks phosphorylation, and therefore activation, of IRF-3 but IRF-3 phosphorylation activity was restored upon disruption of the protease function using protease inhibitors or mutational knockout (Foy *et al* 2003). Conversely, constitutively expressed IRF-3 mutants reduced HCV RNA replication (Foy *et al* 2003).

The mechanism by which NS3 was able to disrupt IRF3 activity has recently been ascribed to NS3/4A's ability to disrupt retinoic acid-induced gene 1 (Rig-I) (Foy *et al* 2005). Rig-I is a cytosolic DexD/H box RNA helicase which functions independently of toll-like-receptor 3 (TLR3) to signal IFN- β production via the activation of IRF-3 and NF- κ B. Rig-I is induced by pathogen-associated molecular patterns (PAMP) found in the 5' and 3' non-translated regions of the HCV genomes (Sumpter *et al* 2005). The disruption of Rig-I by NS3 could be inhibited by the NS3 protease inhibitor SCH6 NS3/4A (Foy *et al* 2005). NS3, therefore, disrupts Rig-I, which no longer activates IRF-3 and NF- κ B, thereby blocking IFN- β production, which limits the host cells' ability to control HCV replication.

HCV replication produces double stranded RNA (dsRNA) which induces the helicases Rig-I and Mda5. These helicases both detect dsRNA via their CARD domains which relay a signal to IRF-3 and NF- κ B activating them. The CARD domain is a member of the death-fold super family and contains 6 helix bundles with a Greek-key topology and a death domain. A newly discovered CARD-motif containing protein, known variously as mitochondrial antiviral signaling protein (MAVS) (Seth *et al* 2005),

virus-induced signaling adaptor (VISA) (Xu *et al* 2005), interferon- β promoter stimulator 1 (IPS-1) (Kawai *et al* 2005), or Cardif (Meylan *et al* 2005), is now known to play a role in this signaling event. It is thought that Cardif interacts with Rig-I to recruit IKK α , IKK β , and IKK ϵ kinases, which activate IRF-3 and NF- κ B directly. Cardif was found to activate IFN- β reporter genes, in a dose dependant method, when over-expressed in 293T cells (Meylan *et al* 2005). It was also found to co-immunoprecipitate with IKK α , IKK β and IKK ϵ (Meylan *et al* 2005). Most significantly from an HCV perspective, Cardif is also cleaved and thus inactivated by NS3/4A (Meylan *et al* 2005). This demonstrates a method by which NS3/4A protease activity can disrupt the activation of IFN- β via the Rig-I pathway.

Besides through the Rig-I pathway, the production of IFN- β can also be induced by engagement of TLR3 which itself is linked to the activation of IRF-3 and NF- κ B by the adaptor protein TRIF. TRIF has also been shown to be a cleavage target for NS3/4A (Li *et al* 2005).

3.1.2. Potential Transformation Capability of NS3

Clinically, HCV infections have been found to lead to hepatocellular carcinoma in a significant number of cases thus strongly suggesting an association between chronic HCV and oncogenesis. NS3 itself has been found to have transforming potential in a variety of cultured cell lines, with the amino-terminal portion of NS3 being the minimal requirement. This transformation capability has been demonstrated in NIH 3T3 cells (Sakamuro *et al* 1995), rat fibroblasts (Zemel *et al* 2001), and the human liver cell line QSG7701 (He *et al* 2003). Upon transformation of NIH 3T3 cells with the protease domain of NS3, cells exhibited rapid proliferation, loss of contact inhibition and loss of anchorage dependence (Sakamuro *et al* 1995). Furthermore, when injected into nude mice tumor formation occurred (Sakamuro *et al* 1995). These results were confirmed by Zemel *et al* 2001 who demonstrated the transformation potential of the protease domain of NS3 in non-tumorigenic rat fibroblast cells (RF). The transformation capability of NS3 protease was eliminated by mutating the active site of the protease, or by adding protease inhibitors (Zemel *et al* 2001). These experiments provide substantial proof that

the active serine protease is cleaving an as yet undefined target leading to increased probability of transformation.

3.1.2.1. NS3 truncation *in vitro* and *in vivo*

Most of the previous studies described above utilize the N terminal portion of NS3 and therefore, the transformation capability of the protease domain of NS3 would be more applicable if the truncated protease domain was seen during authentic viral infections. In fact, NS3 cleavage products have been seen in various mammalian cells (Shoji *et al* 1999, Yang *et al* 2000) and human patient samples (Nielsen *et al* 2004). There are contradictory reports regarding the cleavage site and the origin of the protease (cellular or viral) responsible for the cleavage. Shoji *et al* 1999 suggested that the cleavage occurs due to a host cell protease, whereas it was also reported that NS3/NS4A is responsible for self-cleavage (Yang *et al* 2000). There are also conflicting opinions as to the correct internal cleavage site in NS3 (QRR/GRTGR, or HLIFCH/S, or DVSIPT/S) (Shoji *et al* 1999, Yang *et al* 2000). Despite these differences, both studies used the genotype HCV 1b and examined the putative cleavage site using wild type and mutated versions and both observed loss of the cleavage event in the presence of the mutation. The major difference between these studies is the presence (Yang *et al* 2000) or absence (Shoji *et al* 1999) of the protease cofactor NS4A. Further studies should determine whether a combination of host cell and viral factors are involved in the cleavage event, and if there is more than one mechanism of cleavage. This leaves the question of what function this cleavage performs, whether to the benefit or detriment of the infection. It should be noted that the internally cleaved N terminal domain of NS3 has been found to have a higher oncogenic potential than the full length NS3 (Yang *et al* 2000). As well, a similar cleavage in NS3 is described in other Flaviviruses, such as Dengue virus (Arias *et al* 1993).

3.1.2.2. NS3 and p21^{waf1}

It has been found that NS3, specifically the protease domain, can repress p21^{waf1} transcription (Kwun *et al* 2001). The protein p21^{waf1} is a negative effector involved in controlling the cell cycle via negatively regulating cyclin dependant kinases. p21^{waf1} transcription is normally controlled by the binding of p53 to it's promoter region and the repression of p21^{waf1} was lost when p53 binding sites were removed from that region (Kwun *et al* 2001). Both transcription and protein stability of NS3 in these experiments were not affected. Interestingly, NS3 has been found to localize to the nucleus (Muramatsu *et al* 1997) in HeLa cells. The localization of NS3 to the nucleus was augmented by p53, and inhibited by NS4A (Muramatsu *et al* 1997). It is possible that NS3 is normally sequestered to the membranous web and that somehow p53 inhibits this interaction allowing NS3 to enter the nucleus and effect the regulation of host cell genes such as p21^{waf1}. In patient samples, there is a correlation between NS3 and p21^{waf1} loss which suggests that this interaction is observed as in early phases of hepatocellular carcinoma (Bahnassi *et al* 2005).

3.1.3. Oxidative Stress

HCV infections have been associated with host cell oxidative stress including increased reactive oxygen and nitrogen species and decreased antioxidant defenses (reviewed in Choi *et al* 2004). This may be explained by HCV protein NS3's ability to activate Nox 2 (Bureau *et al* 2001). Nox 2, or NAD(P)H oxidase is a phagosomal protein involved in generation of reactive oxygen species (ROS). NS3 was found to induce a rapid increase in intracellular calcium required for ROS production (Bureau *et al* 2001). The activation of Nox2 by NS3 induces a release of oxygen radicals from phagocytes which in turn induce apoptosis in CD3+/56 T cells, CD3-/56+ natural killer cells and CD3+/56+ natural killer T cells (Thoren *et al* 2004). It was also found that Nox2 inhibitors, histamine and diphenylene iodonium, suppress the NS3 oxygen radical production thus suppressing immune cell apoptosis (Thoren *et al* 2004).

3.2. Physical Interactions with NS3

3.2.1. NS3 interactions with host cell proteome

NS3 has been found to interact with a variety of proteins in the host cell proteome. In most cases, it still remains unclear what role these interactions play in the replication and pathogenicity of HCV (Reviewed by Tellinghuisen *et al* 2002).

3.2.1.1 NS3 and Cell Cycle Proteins

NS3 was found to interact with the tumour suppressor protein, p53, as evidenced by co-precipitation experiments (Ishido *et al* 1998). This interaction occurs both in the presence and absence of the cofactor NS4A and deletion analysis found that the N terminus of NS3 interacts with the C terminus of p53 (Ishido *et al* 1998). It is thought that NS3 interacts with p53 in order to inhibit apoptosis of the infected host cell, a finding further supported by the fact that NS3 can inhibit actinomycin D-induced apoptosis (Fujita *et al* 1996).

3.2.1.2. NS3 Host Immune Proteins

NS3 has also been found to form a physical interaction with several host cell proteins involved in the immune response and these interactions may help define how HCV and specifically NS3 is involved in immune response.

Low-molecular-mass-protein 7 (LMP7) is a component of the host cell immunoproteasome responsible for digesting proteins for presentation on MHC class I cells. This process allows the host immune system to view potential viral infections of cells as their proteins are digested and presented on the outside of the infected host cells. LMP7 has been found to co-immunoprecipitate with NS3, specifically through the NS3 protease domain (Khu *et al* 2004). Furthermore, the LMP7-immunopeptidase activity of Huh7 cell lines is reduced in stably expressing cell lines containing the HCV subgenomic

replicon thus demonstrating interference with presentation of HCV proteins to the host cell immune system (Khu *et al* 2004).

Transforming growth factor- β (TGF- β) is a pleotropic cytokine that affects cell growth, cell death, differentiation and morphogenesis. The TGF- β family activates the downstream Smad family of cytoplasmic effectors. Once activated, these proteins translocate to the nucleus and act as transcriptional activators. The full length NS3 was found to suppress Smad3 mediated transcription activation (Cheng *et al* 2004). A possible mechanism may be the binding of Smad3 by NS3 as the two were found to physically interact via GST pull down assays (Cheng *et al* 2004). In addition, the expression of NS3 caused cells to resist TGF- β mediated apoptosis (Cheng *et al* 2004). In this way, NS3 may play a role in TGF- β induced immune responses.

3.2.1.3. NS3 and Other Proteins

Human protein kinases have been found to interact with NS3. The helicase domain has been shown to interact with human protein kinase A (PKA) (Aoubala *et al* 2001, Borowski *et al* 1999-3) and thereby affect its ability to translocate across the nuclear membrane. This means that NS3 can potentially interfere with phosphorylation signal cascades in the host cell. Furthermore, NS3 itself serves as a substrate for phosphorylation by the human protein kinase C (PKC) (Borowski *et al* 1999-1). The impact of this interaction on the host cell is unknown; however, it does function to sequester PKC away from other substrates thus interfering with PKC signaling.

Sm-D1 is a component of the small nuclear ribonucleoprotein complex (snRNP) and is associated with the autoimmune disease systemic lupus erythematosus. An interaction between NS3 and Sm-D1 was found using yeast 2 hybrid screening and the two proteins were co-localized in the nucleus of cells over-expressing Sm-D1 (Iwai *et al* 2003). It is possible that an interaction between NS3 and Sm-D1 could induce stress in the cell increasing the probability of developing a transformed state (Iwai *et al* 2003). This interaction offers another possible mechanism contributing to the development of hepatocarcinoma after chronic HCV infection.

The NS3 helicase domain interacts with the core histones H2B and H4. Histones are responsible for DNA packaging and organization within the cell and play an important role in the regulation of gene expression. If the level of the core histones (H2B and H4) is reduced, by binding to NS3, it could up regulate host genome expression as host cell promoters are no longer occluded from transcription machinery by histones. This interaction may also explain how NS3 fragments can be found in the host cell nucleus (Muramatsu *et al* 1997). The histone binding domain of NS3 occurs in the helicase domain of NS3 (Borowki *et al* 1999-2). Importantly, NS3 has been found to interfere with the binding of histones to DNA (Borowski *et al* 1999-2).

Using a yeast-2-hybrid system with the amino-terminal protease portion of NS3 as bait, NS3 was found to interact with a protein involved in the intracellular transport and secretory pathways, ELKS- δ (Hidajat *et al* 2005). This interaction was confirmed with co-immunoprecipitation, GST-pulldown and immunoelectron microscopy. NS3 interacts both with the full length ELKS- δ and the splice variant ELKS- α . SEAP (secreted alkaline phosphatase) was used as a marker of secretion and both full-length NS3 and the protease domain were found to enhance secretion from HeLa and Huh 7 cells. Increased secretion was linked with a stronger interaction between NS3 and ELK- δ . This suggests that NS3 may play a role in modulating host cell intracellular transportation and secretion which may help facilitate intracellular transport of viral components aiding in virion formation. It has also been found that another HCV protein, NS5A, also interacts with a protein involved in intracellular transportation, hVAP-A/B. It is unknown if the interaction between NS3 and ELK- δ and the interaction between NS5A and hVAP-A/B are related or if these interactions are independent.

3.2.2 NS3 and other HCV viral proteins

The replication of many (+) stranded RNA viruses, such as Poliovirus and Flaviviruses, occurs on to the endoplasmic reticulum in membrane associated replication complexes and it is hypothesized that this occurs with HCV as well. These membranous web systems aid in virus replication and viral particle construction by sequestering a high local concentration of viral components. If this is true of HCV, the nonstructural and

structural components of HCV should somehow interact and stay clustered around the ER membrane. Although the protease and helicase domains have clear roles in replication, it is currently inconclusive how the replication complex forms and how the viral proteins, including NS3, interact.

Protein-protein interactions between HCV proteins have been studied using GST-pull-down experiments, co-immunoprecipitation and yeast-2-hybrid screens, although often results differed depending on the assay being used. GST pull-down experiments found that NS3 bound to NS2, NS4B, and NS5A whereas Y2H screening found that NS3 interacted only with the NS2 protein (Dimitrova *et al* 2003). *In vitro* expressed co-immunoprecipitation found that NS3 interacted with NS2, NS4A, NS4B, NS5A and NS5B whereas *in vivo* co-immunoprecipitation of the HCV non-structural genes expressed in Huh7 cells found NS3 co-precipitated with NS2, NS4B, NS5A and NS5B, and not NS4A (Dimitrova *et al* 2003). It is possible that in the co-immunoprecipitation experiments, the *in vivo* co-immunoprecipitation of NS4A may not be through direct physical interaction with NS3, but rather the proteins are interacting and precipitating in a large protein complex, although this conflicts with previous research that states that NS4A integrates into NS3 as its cofactor (Bartenschlager *et al* 1999, De Francesco *et al* 2000, Wolk *et al* 2000). It should be noted that in the GST pull-down as well as in the Y2H screen and *in vivo* co-immunoprecipitation, NS3 did not show an interaction with NS4A (De Francesco *et al* 2000, Bartenschlager *et al* 1999, Wolk *et al* 2000). It is possible that for an interaction to occur in these screens, NS4A must initially be part of the same polyprotein (translated *in cis*). It is also possible that the interaction requires another cofactor or protein folding chaperone to occur. NS3 also was not found to interact with itself despite previous research that suggesting that the NS3 helicase domain may form homo-dimers (Khu *et al* 2001). This lack of interaction may be due to improper folding, strain specificity, binding conditions or occlusion of the binding site by the GST or Y2H protein tags. Other work examining the interaction between all HCV proteins via Y2H screening, using random HCV genomic libraries to increase the likelihood of proper folding and complete coverage, found that NS3 interacted with NS4A, and E2 (Flajolet *et al* 2000). This confirms the already described interaction between NS3 and its cofactor NS4A, but it remains unclear what the interaction between

NS3 and the glycoprotein E2 suggests about HCV protein interactions. As with all *in vitro* protein-protein interaction experiments, it will be essential to validate these observations in a functional sense.

4.0. Studying Protein-Protein Interactions

Viruses are the ultimate parasites, requiring a strict dependence on their host cells. This can be affected by both direct physical and functional interactions. There are numerous methods available to study these interactions including physical methods such as protein affinity chromatography, co-immunoprecipitation, and/or library based methods such as a yeast-2 hybrid screen (Phizicky, E. *et al*, 1995, Ausubel, F.M. 1998).

In protein affinity chromatography, a protein is covalently bound to a matrix, such as sepharose beads. This matrix-protein combination is then used to screen cell lysates or other protein extracts for binding partners. The extract flows through the matrix and is washed with a low salt solution to decrease non-specific binding. The proteins which physically bind to the protein of interest are then eluted using a high-salt solution, cofactors, or a detergent such as SDS. Although a relatively straightforward procedure, the main disadvantage to this method is the requirement that the screening protein in question be present in a purified and soluble form. Protein interacting partners can be identified via mass spectrometry or antibody confirmation. This limits the identifiable protein binding partners to those that are in a high enough quantity to be identified by mass spectrometry or which are expected to bind and have an antibody which specifically interacts with them. Further, identification does not prove a direct interaction, as the identified protein may be part of an interacting complex. The advantage, however, is that any interaction discovered is very likely to be a “real” interaction, as the experiment uses cell lysates to examine interactions.

Co-immunoprecipitation is a classic method used to study protein-protein interactions. Cell lysates are prepared and an antibody to the protein of interest is added to precipitate the protein of interest as an immune complex. The protein of interest and any proteins bound to it, are washed, eluted and analyzed. The hurdles for this type of experiment include having a strong, monoclonal, antibody that binds the antigen of

interest. Significantly, the antibody-antigen binding site must not interfere with or compete for the interaction site between the two proteins. One advantage of this method is that the proteins being studied are taken directly from cell culture and therefore should represent biologically relevant proteins/interactions. A disadvantage is that, similar to affinity chromatography, it does not prove a direct interaction between the protein of interest and the prey, as interacting complexes can also be purified.

The Yeast-2 Hybrid (Y2H) screening method allows for the study of protein-protein interactions in an *in vivo* environment. First used to study protein-protein interactions in *Saccharomyces cerevisiae* (Fields *et al* 1989), this method has been utilized to identify protein interactions from a variety of species. A gene encoding a protein of interest (prey) is cloned into a yeast construct which expresses the protein as a fusion partner coupled to a specific DNA Binding Domain protein (Figure 3). This vector is transformed into yeast cells along with a cDNA library. Following transformation into appropriate yeast cells, the presence of two plasmids is confirmed by the expression of the auxotrophic markers tryptophan and leucine (one from each of the bait and prey plasmids). If there is an interaction between the prey and the bait, it brings the activation domain (AD) and DNA binding domain (DBD) into close proximity thus reforming a functional transcription factor, which up-regulates the auxotrophic markers histidine and adenine. Positive interactions are screened as growth of yeast colonies on plates that require enzymes to convert or produce histidine, adenine, tryptophan and leucine for growth to occur. False positive interactions are decreased by using 2 different selection agents (tryptophan, and leucine). Proteins with particular physical attributes, such as helix-loop-helix motifs, or large hydrophobic regions, are usually considered false positives as well, due to their nature to interact ubiquitously with other proteins (Staudinger *et al* 1993). Furthermore, proteins identified must undergo screening against a control protein to make sure that the interaction is real, and is not directed against the tags used in the Y2H screen (DBD or AD). The advantage of this screening method is that it allows for rare and weak protein interactions to be discovered. The main disadvantages is that the system is not in a mammalian cell background and that random insertion of the cDNA library into the second vector can lead to a large number of out-of-

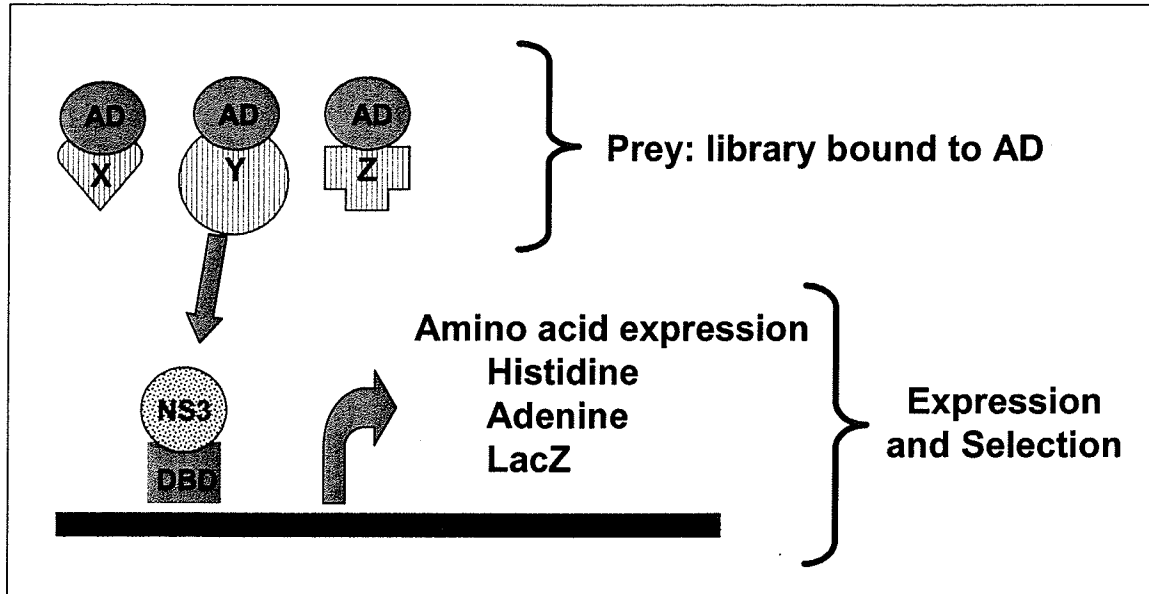


Figure 3: Yeast-2-Hybrid (Y2H) Screen

The Y2H screen is an *in vivo* procedure that allows screening of large “prey” protein libraries for interactions with a bait protein of interest. In this study, HCV-NS3 is fused downstream of DNA binding domain (DBD). Components from a random cDNA library are fused downstream from an activation domain sequence. These constructs are co-expressed in a single yeast cell and if there is a physical interaction between the bait protein and the prey then the activation domain and DBD domain are brought together. When in close proximity, the two domains allow for the transcription and translation of markers, including the expression of the amino acids adenine and histidine, and the protein marker lac Z. Transcription of these markers allow for growth of the yeast colony on selective media.

frame insertions which limits the efficacy of the library and increases the number of false positive identifications observed.

5.0. Aims of Research

NS3 is a multifunctional protein. Its role in viral replication is clear given the essential nature of its protease, and helicase domains. Furthermore, the protein has been found to interact in a multitude of ways with the host cell proteome. However, that picture remains incomplete, particularly with respect to its ability to contribute to cellular carcinoma. Identification of cellular proteins that interact with HCV NS3 should provide a better understanding of the mechanisms by which HCV replicates, induces cellular deregulation/pathogenesis and evades host immune clearance. Because of its proven record to identify protein-protein interactions, a high throughput yeast two hybrid (Y2H) based screening approach was developed to search for novel interactions between NS3 and the host cell proteome.

II) Materials and Methods

1.0 General Techniques

1.1 Polymerase Chain Reaction (PCR)

Polymerase chain reactions (PCR) were carried out using Qiagen Taq DNA PCR kits. 50 μ L reactions were prepared using 25 μ L 2x master mix (Qiagen) combined with 0.1 μ mol primers (Table 1) and 0.01-1 μ g DNA template.

Some PCR reactions for the yeast-2-hybrid screen were carried out using Roche Expand long template PCR system, or combitaq (Roche Biosciences). 50 μ L reactions were prepared using 500 μ M dNTP, 5 μ L 10x PCR buffer IV (Roche), 0.75 μ L combitaq polymerase (Roche), and 1.75mM-500ng DNA. Further PCR reaction parameters are described in section 2.6.

Amplicons for the reticulate lysate experiments were performed using BioRad iproof polymerase system. 34.5 μ L of sterile water was mixed with 10 μ L 5x iproof buffer (BioRad), 1 μ L 10mM dNTP, 0.2 μ L MgCl solution (BioRad), 0.1 μ mol primers and 0.01-1 μ g DNA template.

PCR reactions were carried out using an Eppendorf thermocycler (Eppendorf EP gradient S thermocycler) with a 5 minute denaturing step (96°C), followed by 30-40 cycles of one minute denaturation (96°C), 30-90 seconds annealing (50-65°) and 1 to 2 minutes per kilobase elongation (72°C) followed by a 10 minute polishing step at 72°C. Reactions were left at 4°C overnight or at -20°C for long-term storage.

1.2 DNA Purification

Plasmid DNA was purified using Qiagen miniprep kits from 3mL overnight cultures of bacteria grown in Luria Bertani (LB) media (0.01g/mL tryptone, 0.005g/mL yeast extract, 0.01g/mL NaCl) at 37°C with vigorous agitation (250 rpm). Bacteria were pelleted at 14000xg in a bench top centrifuge (Eppendorf 55417R) at room temperature

No.	Sequence
232F	AAACAAATTCTCAAGCGCTTTCACAACCA
233R	TGAACTTGCGGGGTTTTTCAGTATCTACGA
312F	TGTTGGATCCATGGCGCCCATCACGGCGTA
313R	TGATGAATTCTACGTGACGACCTCCAGGTC
341R	TGATGGATCCTACATGGTTGTCCCTAGGTTTC
341R	TGATGGATCCTACATGGTTGTCCCTAGGTTTC
342F	TGTTGAATTCATGGCGGTGTGCACCCGTGGAG
360F	ATGTTTAATACCACTACAATGGATG
365FN	GACTGTCGACGCGCCCATCACGGCG
366R	GACTGAATTCCTAGGTCGGGATGACAGACA
385F	GACTGTCGACGCGCCTATTACGGCC
386R	GACTGAATTCTACATAGTGGTTTCCATAGAC
405F	GATATACATATGGGTTCTGTTGTTATTGTTGGT AGAATTATTTATCTGGTAGTGGTAGTATCAGG GCCTACTCCCAA
407R	TGATGCGGCCGCGGTTTTTCAGTATCTACGAT TCATA
408R	AACAGAATTCTACTACGTGACGACCTCCAGGT
415F	TGTTAAGCTTGTAATACGACTCACTATAGGGCG AGCCGCCACCATGGCTTACCCATACGATGTT
Sequencing Primers	Sequence
interior NS3 1a F	ACTGTGTCCCATCCTAAC
interior NS3 1a R	GTTAGGATGGGACACAGT
NS3 1a F	TGTTGAATTCATGGCGCCCATCACGGCGTA
NS3 1a R	TGATGGATCCTACGTGACGACCTCCAGGTC
NS3 1b F	CATCACGTA CTCCACCTA
NS3 1b R	GCTTCGCGGCGAGCTC
pGADF	AGATGGTGCACGATGCACAG
pGBKTR	GTCAC TTTAAAATTTGTAT
pGEX2TF	GGGCTGGCAAGCCACGTTTGGTG
pGEX2TR	CCGGGAGCTGCATGTGTCAGAGG
T3 promoter	CAATTAACCCTCACTAAAG
T7 promoter	GTAATACGACTCACTATAG

Table 1: Comprehensive list of primers used in this study.

for six minutes. The supernatant was discarded and the pellet was resuspended in 250 μ L buffer P1 (50mM Tris-Cl pH 8.0, 10mM EDTA, 100 μ g/mL RNase A), lysed with the addition of 250 μ L P2 buffer (200mM NaOH, 1% SDS) and then pH neutralized with the addition of 350 μ L of N3 buffer (3.0M potassium acetate, pH5.5). The lysed bacteria were pelleted at 14000xg for 20 minutes at 4°C in a bench top centrifuge and the supernatant was passed through a Qiagen spin column. Elutions were carried out in 35 μ L of EB buffer (25 mM Tris-HCl , pH 8.0)

DNA was purified from agarose gels using Qiagen Qiaex II kits. DNA was electrophoresed on 0.7-1.0% agarose gels in TBE (107.8g/L Trizma base, 55.03g/L Boric Acid, 7.44g/L EDTA) and the appropriate bands were cut out and heated to 50°C in 300-500 μ L of buffer QX1 (3M NaI, 4M NaClO₄, 5mM Tris/HCl pH 7.5, 0.1% Na₂SO₃) for 10 minutes with 10 μ L of silica gel-bead slurry (Qiaex II resin). Beads were spun down at 14000xg for 1 minute and washed successively with 500 μ L QX1 buffer twice and 750 μ L PB buffer once. Samples were eluted from the silica bead resin with 12.5 μ L elution buffer.

DNA ligations were purified with BioRad Quantum Prep PCR Kleen Spin columns or with Qiagen MinElute PCR purification kits. In the BioRad Quantum Prep PCR Kleen Spin columns, spin columns were pre-cleared of storage buffer by spinning at 425xg for 1 minute followed by loading of the entire PCR reaction onto the column. The columns were spun for 425xg for one minute on a bench top centrifuge and the flow-through was collected. Qiagen PCR purification kits involved diluting the entire PCR reaction in 250 μ L P1 buffer. Samples were loaded on Qiagen PCR purification spin columns and spun at 14000xg for one minute. Bound DNA was washed with 500 μ L buffer PB and eluted in 10-35 μ L of elution buffer (EB).

1.3. DNA Ligations

Both insert and vector were digested for two hours to overnight then electrophoresed on an agarose gel and purified as outlined in section 1.2. The vector back-bone was cleaned using BioRad Quantum Prep PCR Kleen Spin columns (section 1.2). Insert and vector backbone were mixed in a 2:1-8:1 ratio and incubated with 1 μ L

T4 DNA ligase (New England Biolabs) and 1-2 μ L 10x DNA ligase buffer (500 mM Tris-HCl, 100 mM MgCl₂, 10 mM ATP, 100 mM Dithiothreitol, 250 μ g/ml BSA, pH 7.5 at 25°C) in a total volume of 10-20 μ L. The reaction was left at room temperature for 4 hours or at 16°C overnight. Ligations were purified using Qiagen PCR spin columns (section 1.2).

1.4. Preparation of Electro-competent *Escherichia coli* Cells

E. coli strain Top 10 (Invitrogen) was streaked onto LB agar plates (35g LB Agar Lennox/L Disco) from glycerol stocks and grown overnight at 37°C. A single colony was used to inoculate a 5mL culture incubated overnight at 37°C at 250rpm. This culture was then used to inoculate 500mL of LB broth and was grown for approximately 3-4 hours or until an OD₆₀₀ of 0.6-0.7 was reached. Cultures were centrifuged in a Sorvall centrifuge using rotor JA 25.5 for 15 minutes at 6000xg and 4°C in 50mL sterile culture tubes. The cell pellet was resuspended in 500mL ice-cold distilled water followed by centrifugation at 6000xg for 15 min at 4°C. This process was repeated 3 times. After the final spin, supernatant was drained immediately and the pellet was resuspended in approximately 1-2mL of equal amounts of ice-cold water and 60% glycerol. Cells were frozen in 100 μ L aliquots in a dry ice/ethanol bath before storage at -80°C.

1.5. Electrocompetent Bacterial Cell Transformation

Frozen electrocompetent *E. coli* were thawed on ice. Ligations were mixed with the cells and left on ice for one hour. Different amounts of DNA were used for transformation depending on the state of the DNA: 0.1-1 μ L for intact plasmid DNA isolated from bacterial cells, 3-5 μ L of purified ligation mixture or 5-20 μ L of construct if the plasmid DNA was purified from yeast cells. Cells were placed in a 2mm ice-cold electroporation cuvette and electroporated at 2.5kV s (25 μ F capacitance, 200 ohm resistance) in a BioRad Gene Pulser II electroporator. Cells were immediately mixed with 0.75mL of LB or SOC media (2% tryptone D, 0.5% difco yeast extract, 0.058% NaCl, 0.019%KCl, 0.036% dextrose, 0.02% MgCl₂-6H₂O, 0.012% MgSO₄) and

incubated with gentle rocking at 37°C for 45 minutes to one hour. The mixture was then plated on pre-warmed selective LB-agar media with 10 μ L, 100 μ L or 750 μ l of the culture. Plates were left at room temperature until the liquid media absorbed into the agar and were then inverted and placed at 37°C overnight.

1.6 Chemically Competent Cell Transformation

Commercially purchased ultra-competent gold XL-10 cells (Stratagene) were thawed on ice. The cells were aliquoted into tubes with 80 μ L of cells per reaction. The cells were gently mixed with 1 μ L B-mercaptoethanol (Stratagene) and 5-20 μ L of DNA purified from yeast. The cells were left on ice for 1 hour before shocking at 42°C for 45 seconds. The cells were then placed on ice for 5 min and then mixed with 0.75mL 37°C SOC media and left shaking for 1 hour at 37°C. The samples were equally split and poured onto two pre-warmed (37°C) agar plates and left to absorb into the agar. The plates were then left at 37°C inverted overnight.

1.7. Sequencing

All cycle sequencing reactions were performed by the DNA Core Facility (National Microbiology Laboratory, Winnipeg, Canada). PCR reactions were carried out in MJ Research thermocyclers using 1mM sequencing primer (Table 1) and variable template concentrations depending on the source of template DNA (typically 100ng of target) along with Applied Biosystems BigDye Terminator Fluorescent NTPs (ver 3.1). PCR reactions were cleaned with Agencourt Biosciences' CleanSeq magnetic bead technology and reactions were loaded onto ABI 3730XL 48 capillary DNA sequencer using 50cm capillary arrays with POP7 polymer for long sequences or on ABI 3130XL 16 capillary DNA sequencers with 36cm capillary arrays with POP7 polymer for short sequences.

1.8 Sodium Dodecyl Sulphate -Polyacrylamide Gel Electrophoreses (SDS-PAGE)

12.5% SDS-PAGE polyacrylamide gels were hand-cast in our lab. The minigels (7.5cm x 8.2cm) contained 8% acrylamide in the stacking gel and 12.5% acrylamide in the separating gel. Samples were mixed with SDS PAGE gel loading buffer (6x stock = 3% glycerol, 0.6g/mL bromophenol blue, 0.6% SDS, 10mM Tris pH 6.8, 20 μ l/mL B-mercaptoethanol) and boiled for five minutes at 95°C. The gels were then electrophoresed at 25mAmps/gel for 1.40 hours. The gels were then stained with Coomassie blue R250, Silver or processed for western blotting (all described in Sections 1.10-1.12 below).

Commercially purchased Bis-Tris 10% gels (Invitrogen) were also used. Gels were run using 1x MES buffer (Invitrogen) for one hour at 140 volts prior to processing.

1.9 Coomassie Staining

SDS-PAGE gels (section 1.7) were stained for 1h to overnight using Coomassie Blue stain (40% methanol, 10% acetic acid, 2g/L Coomassie brilliant blue R250). Destaining was performed until background was clear with express destain solution (40% methanol, 10% acetic acid) with Kimwipes (Kimberly Clark) to soak up excess Coomassie stain.

1.10 Silver staining

Silver staining of protein gels was accomplished using the Amersham Plus One silver stain kit (GE HealthCare). Gels were rinsed briefly in 50mL ddH₂O before being fixed in a solution of 20% acetic acid and 80% methanol overnight. The next day, the gel was sensitized to the silver stain reaction in a solution of 5% sodium thiosulfate, 75% methanol and 17g/100mL sodium acetate for 30 minutes at room temperature with gentle rocking. The gel was then washed three times in 250mL of ddH₂O for five minutes. A 2.5% silver stain solution was left on the gel for 20 minutes and then the gel was washed two times in 250mL of ddH₂O for one minute. Development of the gel was in a solution

of 6.25g sodium carbonate and 100 μ L formaldehyde in 200mL water. The reaction was stopped by incubating the gel for 10 min in a solution of 3.56g/100mL of EDTA in 250mL water. Finally, gels were washed three times for five minutes with water.

1.11 Western Transfer

SDS-PAGE protein gels were transferred to PVDF membrane using the BioRad semi-dry electrophoretic transfer cell (BioRad trans-blot SP semi-dry transfer cell). The gels were first equilibrated for 30 min in Towbin buffer (192mM glycine, 25mM Tris pH8.2, 20% methanol). PVDF transfer membrane was rinsed in methanol and then Towbin buffer, along with the blotting paper for 10 minutes prior to use. Five sheets of 1mm Whatman paper, followed by the membrane, gel and additional filter paper were stacked on the anode plate. Gels were transferred at 25 volts for 22 min for one minigel or 25 volts for 30 minutes for two gels. Transferred membranes were blocked in 5% skim milk powder in TBST [TBS (60.55g/L Trizma base, 80.06g/L NaCl, 2.01g/L KCl), pH8.0 + 0.05% Tween20] (5% SM-TBST) overnight at 4°C. Blocking solution was removed and replaced with the appropriate dilution of primary antibody in 5% SM-TBST for 1 hour at room temperature or overnight at 4°C with gentle rocking. Membranes were washed once for 15 minutes and twice for 5 minutes in TBST before soaking in the appropriate dilution of secondary antibody in 5% skim milk in TBST for 45 minutes at room temperature with rocking. Membranes were washed in TBST for 15 minutes followed by 4 times for 5 minutes in TBST before visualization.

1.12 Western Blot Detection

Visualization was carried out using the Visualizer (Upstate) or ECL-Plus kits (GE Healthcare) following the manufacturers' protocols. For Visualizer, buffers were brought to room temperature before 2mL buffer A was mixed with 1mL buffer B. The solution was poured onto the side of the membrane that proteins were transferred to and left for three minutes. The excess solution was drained off and the membrane was wrapped in plastic wrap before being exposed to X-ray film for 10 seconds to 30 minutes. X-ray

films were developed using a Feline 14 X-ray film processor (Fisher) and membranes were placed at -20°C for long term storage. With the ECL-Plus protocol, buffers were also brought to room temperature before mixing 7mL of solution A and 175 μL of solution B. The solution was poured onto the side of the membrane the proteins were transferred to and left for five minutes before draining excess reagent off. The membrane was wrapped in plastic wrap and exposed to X-ray film as outlined above.

2.0 Yeast-2-Hybrid (Y2H) Analysis

2.1 NS3 Construct Cloning:

The HCV infectious clone H77C (genotype 1a, NCBI accession number AF011751) was a kind gift from Dr. Jens Bukh (NIH, Maryland, USA) and served as the template from which the full length NS3 coding region was derived (Yanagi et al., 1997). The nucleotides 3420-5312, (aa 1140-1771) served as the template for PCR amplification to create the clones pGBK-T7-NS3 1a and pGAD-T7-NS3 1a. The resultant amplicon was cloned into bait vector pGBK-T7-DBD (DNA Binding Domain)(ClonTech) or prey vector pGAD-T7-AD (Activation Domain)(ClonTech). Initially, the NS3 coding region was amplified by PCR (as per section 1.1) using primers 222F then digested with *Bam HI* and *EcoRI* prior to being ligated into similarly digested pBSKII+ vector (Stratagene). Resulting clones were digested with restriction enzymes *BamHI* and *EcoRI* and gel purified (section 1.2). Vectors were digested with the same restriction enzymes and spin column purified (Biorad PCR Kleen spin columns, section 1.2) prior to use in a ligation reaction (section 1.3). Following purification, transformation and selection, colonies were grown overnight and plasmid DNA purified as described in section 1.2. Purified plasmid DNA samples were digested with *BamHI* and *EcoRI* to screen for inserts then sequences were confirmed using the sequencing primers (T7 promoter and T3 promoter, interior NS3 1a F and interior NS3 1a R, section 1.7). NS3 protease (NS3-P aa 1140-1348) and NS3 helicase (NS3-H aa 1261-1771) domains were also cloned into the bait and prey vectors. The protease domain was amplified using primers 222F and 341R then the amplicon was digested with *EcoRI* and *BamHI* and ligated into a likewise digested

pGAD-T7 or pGBK-T7 vector. The helicase domain was cloned via amplification of pGBK-T7-NS3 1a with the primers 342F and 223R then digested with *EcoRI* and *BamHI* prior to insertion into either the pGAD-T7 or pGBK-T7 template previously digested with *EcoRI* and *BamHI*. Ligations were purified and transformed into *E. coli* Top10 cells (section 1.5) and screened for inserts using digestion before being sequenced (section 1.7).

2.2 Yeast Chemically Competent Transformation:

Yeast (AH109) was transformed with bait and prey vectors in a modified yeast transformation protocol (Gietz *et al* 2002). Plate grown yeast was resuspended in 1mL of water (approximately 10 μ L yeast pellet per transformation) then pelleted by spinning at maximum speed in a benchtop centrifuge for 10 seconds. Pellets were resuspended in 100mM LiAc and left at 30°C for five minutes. The supernatant was removed and the pellet was layered with 240 μ L 50% PEG, 50 μ L Salmon sperm carrier DNA (boiled for 5 minutes), 36 μ L LiAc, 20 μ L water and 5 μ L template DNA. Layered samples were vortexed for one minute and left at 42°C for 20-40 minutes. Samples were then spun at top speed for 10 seconds and the supernatant removed. Samples were resuspended in 150 μ L water before plating on 2 drop-out (2DO media) (-Trp, -Leu 13.35g SD media, 0.35g -Trp/-Leu selective reagent, 18g agarose/500mL water then autoclaved for 15 minutes). Plates were left for 4 days at 30°C. Colonies were re-streaked on 4DO plates (-Leu, -Trp, -Ade, -His selective reagent).

2.3 Yeast Protein Expression

After yeast were transformed with appropriate plasmid DNA, they were grown overnight in appropriate selective liquid media usually 1DO media (13.35g SD media and 0.35g -Trp selective reagent/500mL water then autoclaved for 15 minutes) at 30°C. Overnight cultures were vortexed for 30 seconds on medium speed to resuspend yeast cells that had settled overnight. 50 μ l of resultant yeast culture was used to inoculate 1mL of 1DO liquid media and grown for approximately 6 hours at 30°C or until an OD₆₀₀ of

0.6-0.8. The cultures were spun down at 10 000 rpm for 5 minutes at 4°C. The pellet was resuspended in 500 μ L ice-cold water and spun down again. The pellets were frozen overnight at -20°C.

2.4 Yeast Protein Extraction:

Frozen yeast pellets were rapidly resuspended in 150 μ L 60°C cracking buffer (8M Urea, 5% SDS, 40mM Tris pH 6.8, 0.1 MM EDTA, 0.4mg/ml bromophenol blue, 0.01% B-mercaptoethanol, Roche EDTA-free protease inhibitor tablets) to which 50 μ l of a glass bead (50:50 slurry) were added. Samples were heated to 70°C for 10 minutes before vortexing for one minute. The solution was then spun for five minutes at 14000xg and boiled for a further five minutes before being spun for 1 minute at top speed to pellet the glass beads. The sample was then loaded and run on a 12.5% SDS polyacrylamide gel (section 1.8). Proteins were transferred to PVDF (section 1.11) and probed with anti-DBD (1/100, Santa Cruz) and secondary anti-mouse (1/10000, GE Healthcare) antibodies to detect the Y2H DBD- Fusion proteins. Western blots were visualized with ECL-Plus (section 1.12).

2.5 Yeast-2-Hybrid Screen:

To perform the Y2H screen, yeast (strain AH109) containing the bait construct (pGBK-T7-NS3 1a) was mated with yeast containing a human fetal liver cDNA library in vector pACT2 (CloneTech). PGBK-T7-NS3 1a AH109 was grown in 50mL culture of - Trp 1DO media overnight at 30°C and 250 rpm. When the OD₆₀₀ reached approximately 0.8 the yeast cells were spun down at 1000xg for five minutes in a bench-top centrifuge. The pellet was resuspended in 5mL of residual media and the cells were counted. The pre-transformed library was thawed in a water bath and vortexed briefly before being mixed with the yeast cells. A 10 μ l aliquot of the library was saved for a control. The 10 μ l were mixed with 990 μ l yeast and were plated at a dilution of 1/100, 1/1000 and 1/10000. The yeast cells and remaining library were placed in a two liter flask with 47mL of 2xYPDAK yeast media (1% yeast extract O, 2% peptone D, 2% dextrose,

0.01% kanamycin, 0.02% adenine) and incubated for 24 hours at 42 rpm and 30°C to allow for mating to occur. The yeast cells were visually screened for zygote formation and were left for four more hours before being spun down at 1000xg for 10 minutes. The flask was rinsed twice with 50mL 2xYPDAK media and the supernatant was pelleted and combined. The final pellet was resuspended in 10mL 4DO media and was plated on quadruple dropout media plates (4DO, -Ade, -His, -Leu, -Trp) for 14 days at 30°C. Yeast was also spread at dilutions of 1/10, 1/100, and 1/1000 on -Leu and -Trp and 2DO plates as quantitation controls.

2.6. Yeast Colony PCR:

Yeast colonies were picked and dissolved in 100 μ L lyticase solution (1.2M sorbitol, 0.1 M NaPO₄ pH7.4, 0.2mg lyticase). The primers 232F and 233R were used in PCR reactions (Taq polymerase and Roche Combitaq section 1.1) to amplify the prey inserts. The reactions were carried out using 3 μ L lyticase treated yeast as the template DNA. The PCR reaction had a three minute initial denaturing step followed by 35 cycles of 20 seconds at 94°C (denature) and three minutes at 68°C for annealing and elongation. This was followed by an eight minute polishing extension step at 72°C before storage at 4°C. Dilutions (1/10 and 1/100) were sent for sequencing (section 1.7) with sequencing primer 360F.

2.7 Bioinformatic Analysis:

MegaBLAST analysis techniques were used (NCBI) to ascertain the identity of the sequences present in the amplified prey vectors. The sequences were downloaded from NCBI (<http://www.ncbi.gov>) and SeqMan (Lasergene) or Vector NTI 10 (ver. 10, Invitrogen) was used to evaluate the similarity between the official NCBI sequence and the sequence obtained from the amplicons. Software was also used to study multiple alignments (Clustal X, ver 1.81) and analyze the sequences for appropriate placement of the open reading frame (Vector NTI, ver 10).

2.8 Yeast Plasmid DNA Extraction:

Frozen (-80°C) lyticase treated yeast were mixed with 100µL buffer P1 (Qiagen). The standard Qiagen miniprep protocol was followed as described in section 1.2 except plasmids were eluted in 20µL EB buffer. A subsample of this mixture (5-20µL) was transformed into ultra-competent XL10-gold cells (section 1.6). Bacterial colonies were grown and plasmids extracted using Qiagen miniprep DNA plasmid purification. The presence of an appropriate insert was evaluated by PCR amplification with primers 232F and 233R (section 1.1) followed by direct sequencing (section 1.7) with sequencing primer 360F.

3.0 NS3 Recombinant Protein Expression

3.1 NS3 cloning:

NS3 strain 1a (H77C) was sub-cloned from pBSKII+-NS3 1a into the expression vectors pGEX-2T (GE Healthcare; primer 312F, 313R) using restriction enzymes *Bam*HI and *Eco*RI to create the clone pGEX-2T-NS3fl (full length). Two shortened versions of NS3 strain 1a were created; NS3p (aa 1 – 200) and NS3ps (aa 1 – 150). HAT-NS3p was created using the primers 365FN and 366R, and was subcloned into a pHAT10 vector which makes use of an N-terminal Ni-chelate binding domain (Clonetech) using the restriction enzymes *Eco*RI and *Sall*. The shorter version, clone pHAT-NS3ps 1a, was created using the primers 365FN and 341R, and was subcloned into pHAT10 using the restriction enzymes *Eco*RI and *Sall*.

Constructs were also created using NS3 from strain 1b (HCV strain1b-plasmid pHCVrep1bBB7, a kind gift from Dr. Charles Rice, Rockefeller University, New York, USA). A short version of NS3 (aa 1-150) from strain 1b was cloned into pHAT10 with the primers 385F and 386R and the restriction enzymes *Eco*RI and *Sall* to create the clone pHAT-NS3ps 1b. To create the clone HIS-NS3/4Apep-1bp, a primer (405F) was designed which fused a NS4A strain 1b peptide to the N terminal region of NS3 strain 1b. This primer, along with the primer 386R was used to amplify NS3ps from the HCV strain

1b replicon. This amplified NS3ps, with an N-terminal 4A peptide, was then cloned into pET29b containing NS3p strain 1b using the restriction enzymes *EcoRI* and *NdeI*. The full length NS3 from strain 1b (pHCVrep1bBB7) was also cloned into pET29b using the primers 405F and 408R, which fused the 4A peptide to the full length NS3 strain 1b, using the restriction enzymes *EcoRI* and *NdeI* to create a full length NS3/4A construct. All of the above cloning was performed using the techniques outlined in section 1.0.

3.2. Expression of NS3:

Expression constructs in bacterial strain BL21(DE3) were grown overnight and seeded 1/100 into 5-250 mL fresh LB media. Cultures were grown to an OD₆₀₀ of 0.6-0.8 and in some instances were supplemented just prior to induction with ZnSO₄·7H₂O to 100µM final concentration. Cultures were grown to an OD₆₀₀ of 0.6-0.8 then half the culture was induced with a final concentration of 0.6mM IPTG for 3.0 hours at 22°C. Cells were centrifuged prior to storage at -20°C. Cultures were lysed using 500µl-50ml lysis buffer (PBS, 10% glycerol, 0.5M NaCl, 3mM DTT, 0.5% CHAPS, Roche complete protease inhibitors-EDTA) and sonicated in an ice slurry water bath with a Branson sonicator (4 times for 4 seconds at level 8) before analysis by SDS PAGE (section 1.8).

In some cases, HAT-NS3p and HAT-NS3ps strain 1a were expressed in bacteria cell lines over-expressing chaperones (Takara) to aid in proper folding, or to utilize human codon biased tRNAs to aid in increased translation (Rosetta cell lines, Novagene). The bacterial cell lines included pG-KJE8 (groEL, dnaK, dnaJ, grpE), pGro7 (groES, groEL), pKJE7 (dnaK, dnaJ, grpE), pG-Tf2 (groES, groEL, tig), pTf16 (tig), and rosetta (Novagene, expressing tRNAs with anticodons: AGG, AGA, AUA, CUA, CCC, GGA on chloramphenicol resistant plasmids) were all cloned into BL21(DE3) pLysS *E. coli*. Overnight cultures were used to inoculate 5mL cultures. Samples were grown to an OD₆₀₀ of 0.7 then induced with 1mM IPTG and 3mg/mL arabinose, and/or tetracycline (0.0075µg/mL) depending on the genetic background. Cultures were induced at 22°C for 2.5 hours and then spun down at 6000 x g for 15 min.

3.3 Purification of NS3p from HCV strain 1a

A pellet from a 250 ml culture was lysed in 50mL of lysis buffer (PBS, 10% glycerol, 0.5M NaCl, 3mM DTT, 0.5% CHAPS, protease inhibitors) then sonicated as described above. The sample was spun at 14000xg for 20 min and the soluble fraction was filtered through a 0.45 micron filter. It was then loaded onto a 1mL HIS-FF HiTrap column (GE Healthcare) on an FPLC (GE Healthcare) using a PBS based buffer A (PBS, 500mM NaCl, 20mM imidazole) and buffer B for elution (buffer A + 500mM imidazole). The column was run with a 2 step gradient elution (20% buffer B gradient over 10 column volumes, and 100% buffer for 10 column volumes) and an average flow rate of 1mL/minute. This was followed by purification on a 1mL anion exchange HiTrap QFF column (GE Healthcare; buffer A = 25mM Tris pH8.0, buffer B = buffer A + 1M NaCl). This column was run with a 20% buffer B gradient over 10 column volumes, followed by 20% buffer B for 5 column volumes and 100% buffer B for 5 column volumes with a 1mL/minute average flow rate.

3.4 Large Scale Expression and Purification of NS3 from HCV strain 1b:

The shortened version of NS3 from strain 1b (NS3ps1b; aa 1 –150) was grown on agar plates overnight and then in a 5mL overnight culture. This was used to inoculate a 500mL culture, which was grown to an OD₆₀₀ of 0.6, then 0.1mM ZnSO₄ was added. Thirty minutes later the culture was induced with 0.6mM IPTG and grown for 3 hours at 25°C. Bacterial cultures were pelleted at 6000 x g and lysed (50mM NaPO₄ pH 7.4, 10% glycerol, 0.5M NaCl, 3mM DTT, 0.5% CHAPS, protease inhibitors). Proteins were purified on a 1mL SPFF cation exchange column (GE Healthcare; buffer A= 50mM NaHPO₄ pH 6.5, 0.2mM EDTA, 10% glycerol, 0.2mM DTT, buffer B =buffer A + 1M NaCl) with a gradient elution increasing from 0-100% buffer B over 20 column volumes and a flow rate of 1mL/minute. This was followed by a Heparin column or a HIS column. The 1mL Heparin FF column (GE Healthcare; buffer A = 10mM NaPO₄ pH 7.0, 10% glycerol, 0.5mM DTT, buffer B = buffer A + 1M NaCl) was run with a gradient elution (0-100% over 20 column volumes) and flow rate of 1mL/minute, whereas the

HIS-FF HiTrapp column (GE Healthcare; 20mM Tris pH8.0, 10% glycerol, 0.5M NaCl, 10mM imidazole, buffer B = buffer A + 250mM imidazole) was run with a step elution increasing from 0% buffer B for 5 column volumes, to 100% buffer B for 10 column volumes and 1mL/minute flow rate.

The full length HIS-NS3/ 4Apep-1b clone was grown in 1L LB broth from an overnight colony and induced with 0.1mM IPTG at 22°C for 2.5 hours. The culture was spun down and stored overnight at -20°C. The bacterial pellet was resuspended in 50mL lysis buffer and sonicated (8 times for 8 seconds at level 8). The sample was then incubated for 20 min on ice and then spun for 35 min at 14000 x g. The soluble fraction was filtered using 0.45 micron syringe filters (Corning) before loading on a 50mL super loop on an FPLC. The sample was purified using a 1mL HIS-FF column (GE Healthcare, buffer A = 20mM NaPO₄ pH 7.0, 10% glycerol, 0.5M NaCl, 20mM imidazole, buffer B = buffer A + 500mM imidazole) with a gradient elution (0-100% over 20 column volumes) and 1mL/minute flow rate and then loaded onto a 50mL S200-HR gel filtration column (GE Healthcare, buffer A= 50mM NaHPO₄ pH 6.55, 0.2mM EDTA, 10% glycerol, 0.2mM DTT, buffer B =buffer A + 1M NaCl) with a gradient elution (0-100% buffer B over 20 column volumes). The samples were pooled and concentrated before storage in gel filtration elution buffer and 50% glycerol.

3.5 Identification of NS3 using mass spectrometry

Following electrophoresis on 12.5% SDS PAGE gels, proteins were stained with either Coomassie Blue (section 1.9) or silver (section 1.10). For silver stained gels, bands were cut out using a clean scalpel in a fume hood before agitation with 30mM potassium ferricyanide and 100mM sodium thiosulfate 1:1 solution for 20 minutes to destain the slices. The reaction was stopped by washing three times with sterile water. For either staining type sample, slices were left overnight in water with agitation at room temperature. Gel slices were dehydrated by removing the water and replacing it with 200μL of 25mM NH₄HCO₃/50% acetonitrile and shaking for 15 minutes at 900 rpm. This was repeated a total of five times, and finally with 100% acetonitrile once before being dried in a speed vacuum centrifuge for approximately 15 minutes with no heat

(Savant). The gel slices were then reduced and alkylated via the addition of 10mM DTT for 30 min at 50°C, then 55mM iodoacetamine for 45 min at room temperature in the dark. The samples were washed with 25mM NH₄HCO₃/50% acetonitrile followed by the samples being dried in a speed vacuum centrifuge. The samples were trypsinized via the addition of 30μL ice-cold 20μg/mL trypsin in 40mM NH₄HCO₃/9% acetonitrile for 15 minutes on ice followed by removing excess liquid then incubation at 37°C overnight. The peptides were extracted via washing with 0.1% formic acid, and 50%acetonitrile/0.1% formic acid. The supernatant was concentrated and desalted using C18 resin Zip tips (Millipore) according to the manufacturers protocols with elution in 50% acetonitrile/0.1% formic acid. 1.5μL of sample was placed into Proxeon 0.1um discrete spray tips and nanospray was utilized to collect peptide fingerprint (MS) data as well as peptide sequence data (MS/MS) in information data acquisition mode (repetitions of 1 survey scan followed by collision analysis of 3 most abundant peaks) from a QSTAR XL mass spectrometer (Applied Biosystems). Data were analyzed using Mascot Software (Matrix Science; www.Matrixscience.com). The NCBI non-redundant protein data base was searched with criteria set to: all species, any molecular weight and pI, with a possibility of one missed cleavage and standard modifications of carbamidomethylation of cysteine and oxidation of methionine.

4.0 Confirmation of Protein-Protein Interactions

4.1 Protein Interactions in Yeast:

Purified prey plasmids (pACT2) extracted from AH109 yeast were transformed into AH109 chemically competent yeast cells containing pGBK-T7-NS3, pGBK-T7-NS5A or pGBK-T7-lamin (negative control)(Clontech) and plated on 2DO selective media. Colonies were then re-streaked on 4DO media to test for positive interactions indicated by colony growth.

4.2 Protein Expression in Reticulate Lysates

Genes of interest were amplified with the primers 415F and 407R. The forward primer added a T7 promoter region including a Kozac sequence (GCCGCCACC) and AUG start site to the front end of the amplicon. The amplified genes, or amplicons, were then used as a template in a TNT T7 PCR Quick Reticulate Lysate (Promega) to express the prey proteins. An aliquot of 40 μ L of TNT T7 PCR reticulate lysate master mix was mixed with 7 μ L of PCR amplicon, 3 μ L of water and 1 μ L of 35-S radioactive methionine. 1.5 μ L of sample was evaluated by 10% SDS PAGE (section 1.8) followed by Coomassie Blue visualization (section 1.9). The gels were dried on 1mm Whatman paper in a gel dryer for 1.5 hours at 50°C and exposed to a blank phospho-screen for 2 hours before image capture on a Typhoon 9400 imager (GE Healthcare). Scans were performed at 200 microns resolution.

4.3 Affinity co-precipitations (HIS Pull Down):

Pull-down experiments were conducted with Ni bound beads (Amersham) washed twice in 1mL of wash buffer (0.1M Tris pH 7.4, 0.01% triton X-100, 10% glycerol, 20mM imidazole). 30 μ l of ³⁵S prey protein was pre-cleared by mixing with 20 μ L HIS beads for four hours at 4°C with agitation to remove proteins that bound HIS beads non-specifically. The HIS-NS3/4Apep-1b recombinant protein was pre-bound to the Ni beads at a molar excess of protein to binding sites on the beads. The HIS-NS3/4Apep-1b and beads were brought up to a total volume of 500 μ L using the wash buffer and left at 4°C with gentle rocking for four hours. The pre-bound HIS-NS3/4Apep-1b resin was washed twice with 1mL of wash buffer. The pre-bound HIS-NS3/4Apep-1b beads, or control beads with no HIS-NS3/4Apep-1b, were mixed with 250 μ L of the pre-cleared reticulate lysate and an additional 250 μ L of wash buffer and incubated at 4°C with agitation overnight to allow protein-protein interactions to occur. Samples were spun down and supernatants removed. Resins were washed 3x 0.75mL in wash buffer. The final pelleted resins were mixed with 20 μ L 2xSDS PAGE protein loading buffer and boiled for 5 minutes at 95°C before loading on 10% SDS PAGE acrylamide gels (Invitrogen,

section 1.8). Gels were dried for 3.0 hours at 50°C and exposed on a blanked phospho-screen for 12-48 hours. Scanning on the Typhoon was as described above.

III) Results:

1.0. Yeast 2 Hybrid Screen (Y2H)

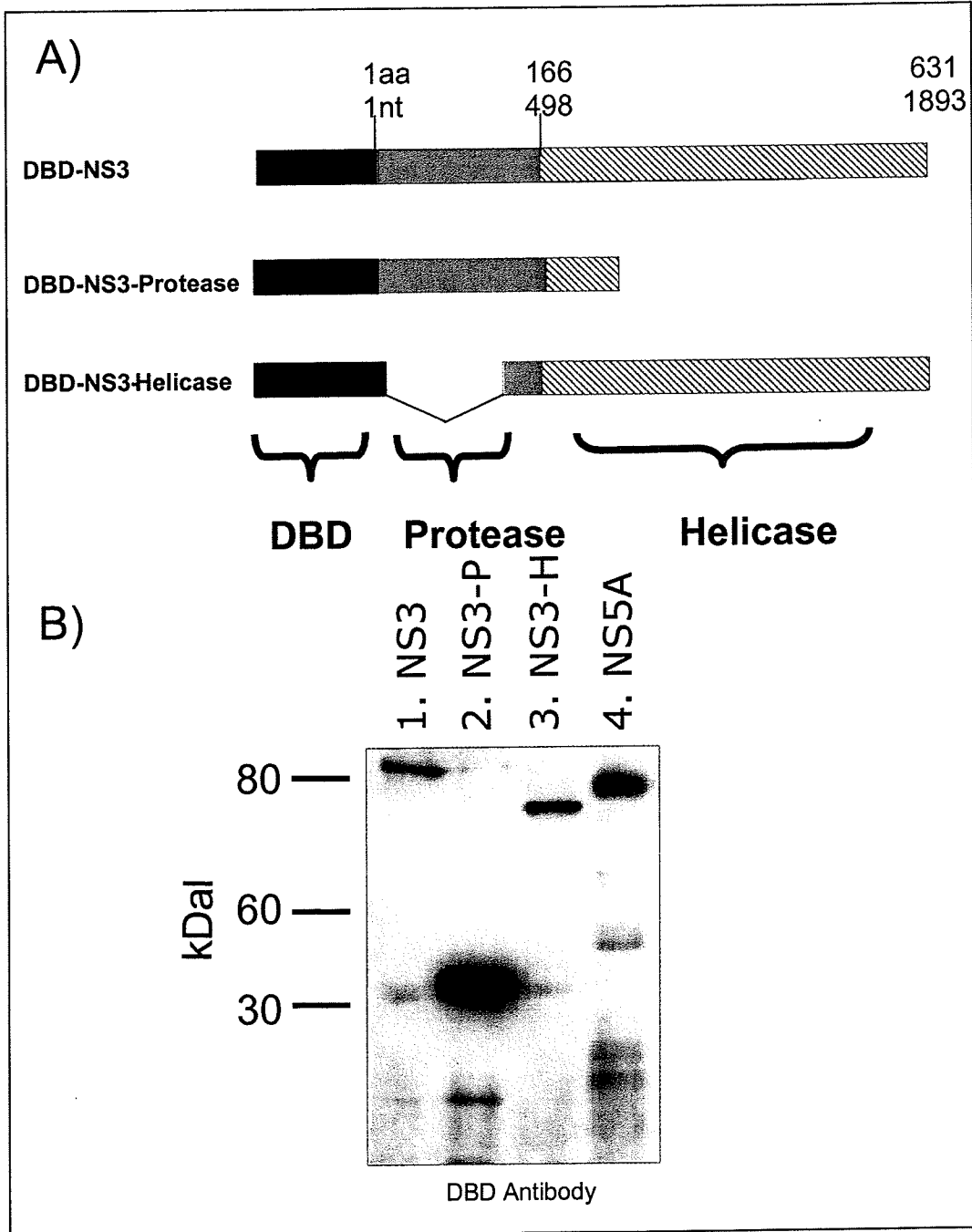
1.1. Expression of HCV NS3 in AH109 yeast cells

Hepatitis C virus non-structural 3 (NS3) protein encodes an approximately 67 kDa polypeptide which was used as “bait” in a yeast two hybrid (Y2H) genetic screen to identify host cellular physical binding partners. In this screen, when two introduced fusion proteins interact they reform a transcription factor essential for yeast viability on appropriate selection media. This required that the open reading frame encoding NS3 strain 1a be amplified and placed into the yeast expression vector pGBK-T7 (Fig 4A). The pGBKT7 vector encodes a DNA binding domain (DBD) to which NS3 was fused. This fusion increases the molecular weight of expressed NS3 by 14.5 kDa. In addition to the full length NS3 construct, regions corresponding to the NS3 protease (P) or helicase (H) domains were also placed into the yeast vector (Fig 4A). Prior to the full Y2H screen, the expression of the DBD-NS3 fusion constructs within yeast cells had to be evaluated to ensure that the DBD-NS3 target “bait” would be present during the genetic screen. Yeast cells were individually transformed with the various constructs, proteins were extracted, and the presence of the fusion proteins determined by Western blot with a primary antibody directed against the protein tag DNA binding domain. Expression was successfully demonstrated for three NS3 constructs at their appropriate molecular weights: DBD-NS3 (83.9 kDa), -NS3P (32.8 kDa), and -NS3H (65.5 kDa) (Fig. 4B). In addition, a control plasmid encoding the HCV NS5A protein (70.5 kDa) fused to the DBD was also included in this Western analysis. All constructs expressed at sufficient levels for detection and in non-degraded forms, thus allowing for their use in an Y2H screen.

A second criteria for a successful Y2H screen is that the DBD-NS3 fusion must not allow growth in the absence of an interacting binding partner. In other words, the

Figure 4: Expression of NS3 constructs in yeast cells.

- (A) Three different constructs were created for use in the Y2H screen; NS3, NS3P (protease) and NS3H (helicase). Each sequence was fused to the DNA binding domain (DBD) present in the yeast expression vector pGBKT7.
- (B) Protein extracts from yeast expressing various DBD-sequences were analyzed by Western blot analysis and visualized using a primary antibody to the DBD. Extracts were loaded as: (1) DBD-NS3, (2) DBD-NS3P (3) DBD-NS3H (4) DBD-NS5A. Molecular weights are indicated alongside the gel in kDa. The predicted Molecular Weights are: DBD-NS3 (83.9kDa), DBD-NS3P (32.8kDa), DBD-NS3H (65.5kDa) and DBD-NS5A (70.5 kDa).



DBD-NS3 protein must not act as a functional transcription factor by itself. It was previously determined that NS3 does not interact with a lamin peptide (AD-Lamin C, Clontech), which could be used as a functional negative control (Carpenter, unpublished). In addition, other functional aspects of NS3 were studied, including the independent binding of the helicase and protease domains of this protein with each other. For this, several constructs were cloned in-frame behind the activation domain (AD) of “prey” vector pGAD-T7 and “bait vectors pGBK-T7 including: NS3 (full length), NS3P, NS3H, NS5A, and another negative control interaction protein, growth receptor bound protein 2 (Grb2). Yeast were co-transformed with both a bait construct (DBD) and a prey construct (AD) and selected on 2 drop-out (2DO) media to ensure that both plasmids were present in the yeast cell. Colonies were re-streaked on 4 drop-out (4DO) plates to assess the physical interactions between the bait and prey proteins (Fig. 5). Yeast containing HCV DBD-NS5A co-expressing AD-Grb2 showed vigorous growth on selection media as has been previously demonstrated for the Y2H (Tan *et al* 1999) and served as a positive control for the 4DO selection plates (Fig 5-1). DBD-NS5A also interacted with AD-NS3 (Fig. 5-2) and AD-NS3P (Figure 5-9). NS3 was found to interact with the NS3P domain (Fig. 5-10), but not with itself (Fig. 5-3). The helicase domain of NS3 did not interact with itself (Fig. 5-7). The protease domain of NS3 was found to interact with NS3P (Fig. 5-6), NS3H (Fig. 5-4 and 5-5) and NS3 (Fig. 5-10), but did not interact with Grb2 (Fig. 5-8). Therefore, specific and functional interactions for HCV NS3 were evident in the Y2H protein-protein interaction analysis.

1.2. NS3 Yeast 2 Hybrid Screen

Once the expression and ability of pGBK-T7-NS3 to interact appropriately within an Y2H screen was demonstrated, the screen could be initiated. When performing a full Y2H screen, a cDNA library is cloned into the prey vector and introduced into yeast containing a DBD domain fusion protein (in this case, DBD-NS3) (Fig. 6). The yeast is grown on 4DO selective media and normally, the largest yeast colonies that grow are individually grown in liquid culture, then the “prey” plasmid DNA is extracted and

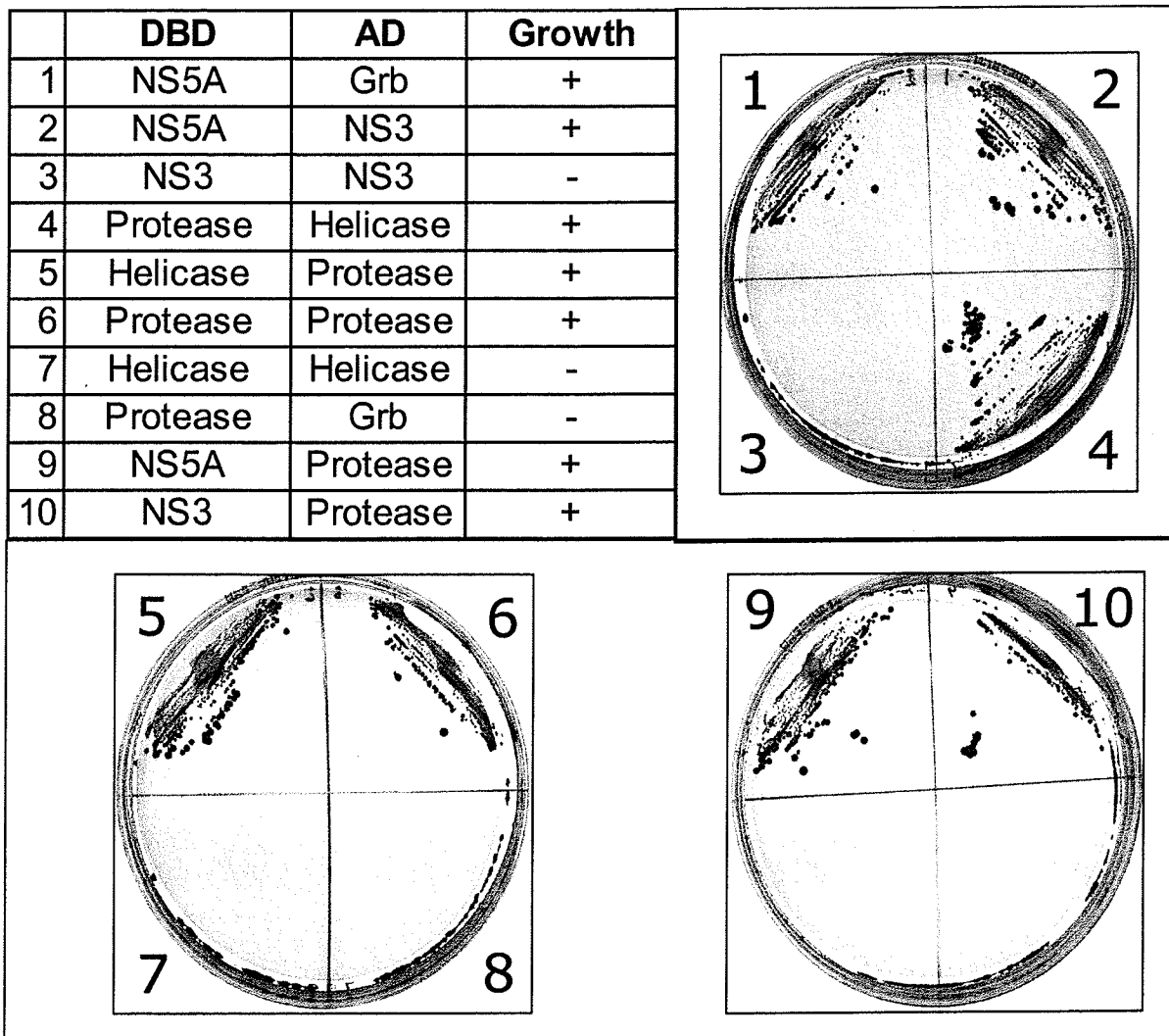
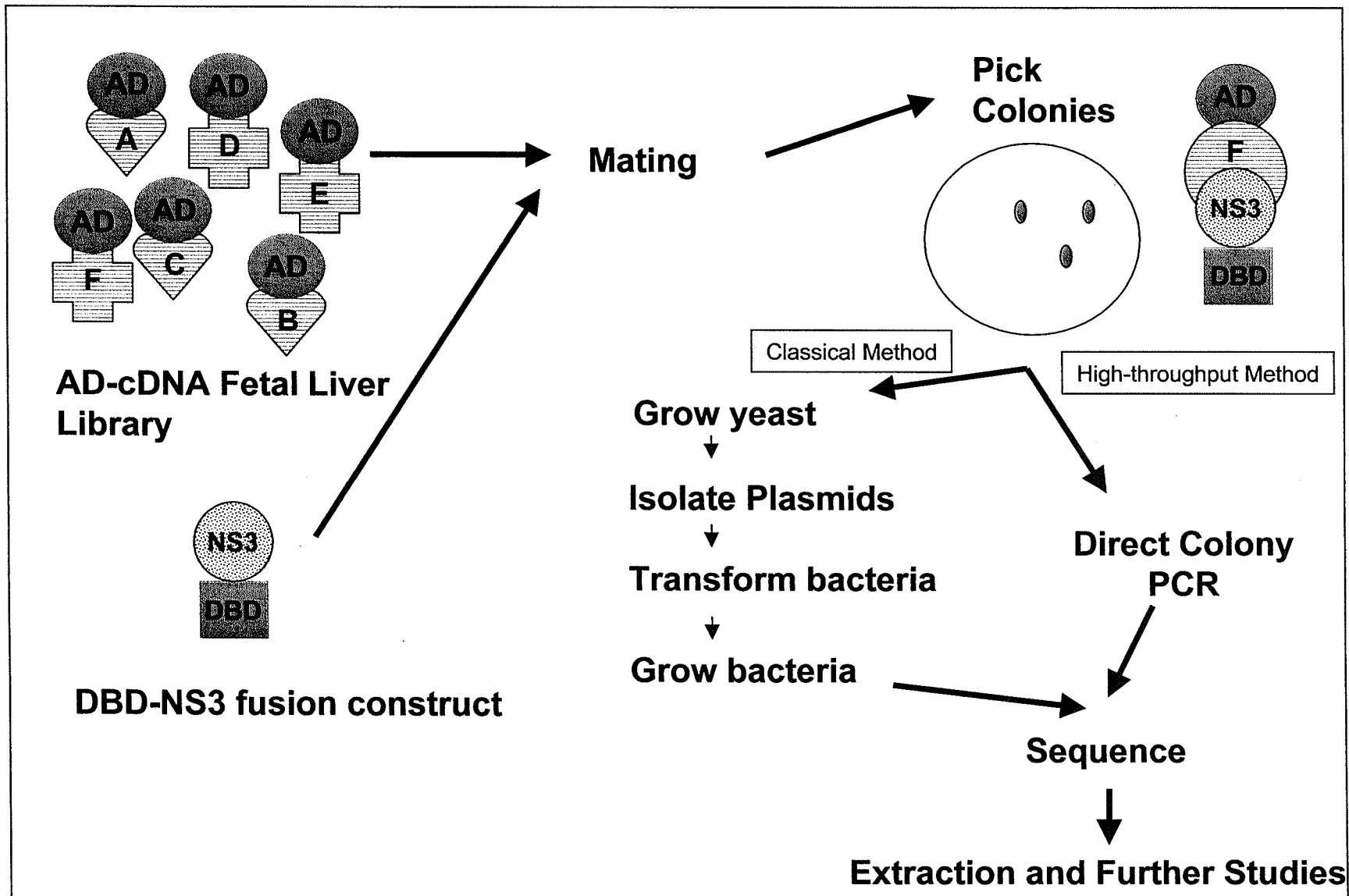


Figure 5: Physical Interactions of DBD- and AD- proteins in yeast.

Yeast were co-transformed with a DBD expression vector :DBD-NS5A (1,2,9), DBD-NS3 (3, 10), DBD-NS3-Protease (4,6,8) or DBD-NS3-Helicase (5,7) and one AD expression vector: AD-Grb2 (1,8), AD-NS3 (2,3), AD-Helicase (4,7) or AD-Protease (5,6,9,10). Double transformants were grown on 4 drop out (4DO) selective media. Growth of colonies shown here on Petri plates suggest a physical interaction between the proteins expressed from the DBD and AD domain vectors.

Figure 6: Yeast 2 Hybrid (Y2H) screening procedure.

Yeast expressing a DBD-NS3 fusion construct are mated with yeast expressing an AD-cDNA fetal liver library (letters A-F). Only if the NS3 and prey protein elements (letters A-F) interact is a transcription factor restored leading to expression of genes essential for yeast colony growth on selective media. In the “classical” approach to screening, yeast colonies are grown in liquid broth, the AD containing plasmid DNA is isolated then used to transform bacteria. These bacteria are subsequently grown, and the plasmid re-isolated in preparation for sequence analysis. Here a high-throughput methodology was developed in which direct yeast colony PCR replaces the majority of the steps for plasmid isolation and identification in the classical method



used to transform bacteria. This plasmid DNA is then amplified, extracted and sequenced. Because this is a very time consuming approach especially when a large number of colonies must be dealt with (as in the Y2H approach), only the largest most robust growing colonies are evaluated. In our study, we wished to examine all of the prey constructs which resulted in a growth phenotype for several reasons: 1) size of a growth positive colony may not be indicative of a true/relevant interaction with NS3, 2) rare cDNAs may be underrepresented in the fetal liver library we intended to use in our screen and therefore these species may be overlooked and 3) to understand the nature of “sticky proteins” that allow colony growth independent of the bait sequence and thereby disregard these false positives in future screens. Therefore, a high-throughput method was developed which focused on direct PCR and sequencing on yeast cells containing the prey vector template thus eliminating most of the time consuming steps of the classical method (Fig. 6). Based on the results of the sequence analysis, only prey plasmids of interest were extracted for further studies.

AH109 yeast cells containing pGBK-T7-NS3 were grown in liquid culture before being mated with Y187 yeast transformed with a fetal liver cDNA library cloned into the AD vector pACT2 (Clontech). The viability of the pGBKT7-NS3 culture was 5.9×10^7 cfu/mL while that of the pACT2 liver library was 1.63×10^7 cfu/mL (colony forming units/mL). Following mating overnight, the viability of yeast diploids was determined to be 3.2×10^5 cfu/mL. A total of 3.9×10^5 diploid colonies were plated on 4DO selection media and screened for colony growth. 425 colonies grew on 4DO media indicating that only 0.11% of all yeast containing both a DBD-NS3 and an AD-liver cDNA presented a growth phenotype. These colonies were re-streaked onto fresh 4DO media to confirm positive growth and increase the colony size prior to harvesting in lyticase solution.

Lyticase treated yeast were used directly as a template for PCR reactions in order to generate targets which were later sequenced. Using long DNA PCR primers, a total of 295 samples were amplified using unmodified Taq DNA polymerase (Qiagen). Of these PCR reactions, 237 were sequence competent thus generating information for 56% of the growth positive yeast (Fig. 7A). Optimizing the PCR reaction (Taq opt) by changing the template concentration, number of PCR cycles and/or annealing temperature allowed an

A)

	Amplify	No Amplify	Sequence	No Seq.
Taq	295	130	237	58
Taq (opt)	126 (68/58)	62	11	53
Taq Total	363	62	248	115
Combitaq	168 (53/115)	9	159	9
Total	416	9	407	9

B)

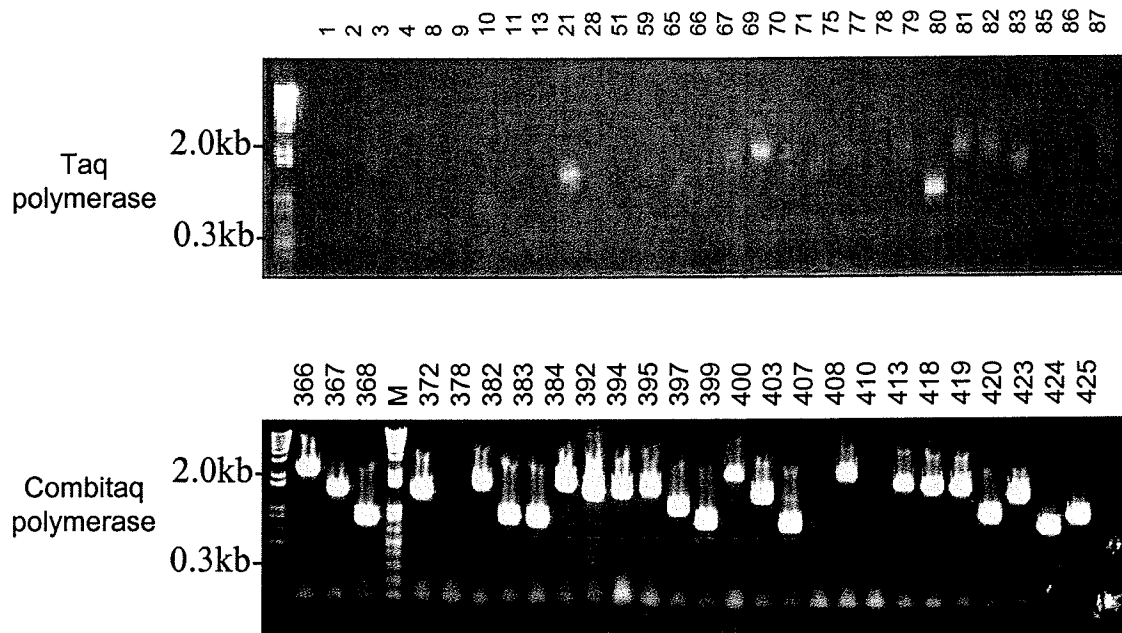


Figure 7: Y2H direct colony PCR amplification.

- (A) Direct yeast colony amplification was attempted for all 425 isolates obtained from the NS3 Y2H screen. Amplification was attempted with Taq polymerase alone, after optimization of the Taq polymerase amplification (Taq opt) or after using a combitac mixture (Taq and a proof-reading polymerase). Columns indicate successful (Amplify) or unsuccessful (No Amplify) PCR amplification or successful (Sequence) or unsuccessful (No Sequence) sequencing reactions. The amplified column contains the data in the form [a (b/c)]. It should be noted that the amplified column is the sum (a) of both primary amplifications (b), and amplicons that were reamplified (c) to obtain better sequencing results.
- (B) Representative agarose gel electrophoresis/ethidium bromide detection of direct yeast colony PCR amplifications performed with Taq polymerase (upper panel) or Combitaq (lower panel) PCR. Yeast colony numbers are indicated above the lanes. A DNA ladder (left) indicates size of products in kb.

additional 68 samples to be amplified and, of these, 11 were successfully sequenced (Fig. 7A). Finally, switching to a polymerase mixture containing proof-reading and non-proof reading polymerases (Roche Combitaq) allowed for an additional 53 samples to be amplified and an additional 159 to be sequenced (Fig. 7A). In total 416 of the 425 positive interactions were amplified (98%) and 407 amplicons provided some sequence information (Appendix 1).

The 407 “prey” sequences were initially analyzed in order to eliminate irrelevant prey targets. These included sequences that were out of frame with the activation domain or did not fall within the open reading frame of a known gene (Table 2A). In total, 125 prey sequences were out of frame behind the AD and therefore represent irrelevant protein interactions. An additional 10 were classified as “unclarified” human genome open reading frames. These were sequences that matched to a human DNA sequence, but whose sequence has not yet been defined as an open reading frame. In many cases these sequences were present in Bacmid cloning vehicles from large scale sequencing projects. There were 9 prey that upon subsequent analysis and PCR amplification confirmation would not provide any sequence information at all and 28 prey which gave poor sequence data. The 9 prey which did not provide sequence is most likely due to an inability for primers to bind due to loss of the primer binding site or to mutations to the binding sites. Recombination events in yeast is also possible. The 28 prey which gave poor sequence is thought to be due to the presence of more than one prey vector occluding clean DNA sequence. Finally, 24 genes were found to have unusual insertions. Typically these sequences represented a fusion of 2 apparently non-related sequences following the activation domain (Table 2B). In most cases (20 of 24), despite the fusion, at least one of the prey elements was still in frame with the AD. Although not dealt with further, subsequent analysis may deem that these prey constructs do represent real interactions. To evaluate this, each in-frame gene must be re-cloned separately into the AD vector and re-screened. In total, this allowed 205 “prey” sequences to be eliminated from further characterization as an NS3 interaction partner (Table 2A).

Of the remaining 220 prey, 75 unique sequences were found to be in-frame (Table 3). In order to more clearly focus our analysis of the putative interacting partners, a focus was placed on proteins involved in cellular pathways affected by hepatitis C virus

A)

Unusual prey constructs	Number
Frame shift mutations	125
Unclarified ORF	10
Not amplified	9
Poor sequence	28
No sequence	9
Fusion proteins	24
Total	205

B)

Colony #	gene	In frame
3	apolipoprotein L	Q starts at 105
51	aspartoacylase (aminocyclase) 3	Q starts at 12
62	chromosome 19 clone	Q starts at 14
66	chromosome 22	Q starts at 111
76	UDP-glucose dehydrogenase (UGDH)	Q starts at 197
77	elongation factor 1-alpha (EF1A)	Q starts at 51
85	Jak2 kinase	Q starts at 16
88	seq	Q starts at 191
92	complement factor B mRNA	Q starts at 33
103	FLJ32065 protein	Q starts at 13
164	chromosome 3p21.3	Q starts at 13
179	alpha-1-antitrypsin	Q starts at 292
212	ND1 (mitochondria)	Q starts at 110
241	ND1 (Mitochondrial)	Q starts at 98
243	HBeAg binding protein	Q starts at 172
265	retinol binding protein 4	Q starts at 15
275	chromosome 22	Q starts at 46
292	aldolase B, fructose bisphosphate	Q starts at 2
345	sirtuin (silent mating type information regulation 2 homolog	Q starts at 190
365	complement C1q-C	Q starts at 15
2	aldolase B	Q starts at 9 - N
4	HERPUD1 gene for stress protein	Q starts at 394 - N
56	HERPUD1 gene for stress protein	Q starts at 333 - N
144	FLJ10856/transmembrane protein 30A	Q starts at 115 - N

Table 2: Summary of yeast 2 hybrid prey not further evaluated for interaction with HCV-NS3 based on sequence information.

- (A) Following PCR amplification and sequencing, prey sequences were classified into several categories. None of the sequences was further evaluated as an NS3 binding partner (see text for details).
- (B) Fusion proteins formed by two different mRNA species being present following the activation domain (AD) sequence. The relevant colony number and major genes identified are listed. For the in frame column, Q indicates the position at which the major gene begins (for example in colony 3, there is an insert of 104 bp prior to the appearance of the apolipoprotein gene sequence). In frame columns with a -N do not have the major gene in frame to the AD.

Colony #	Gene
18	insulin-like growth factor binding protein 3
25	haptoglobin
27	endothelial cell apoptosis protein E-CE1
36	fibrinogen, gamma polypeptide
41	cyclin G1 binding protein 1
48	inhibin beta E (Activin beta E)
53	alpha 2 HS glycoprotein
54	fibrinogen beta-chain mRNA
55	complement C4A
63	complement C8
64	RNA Polymerase 1 polypeptide C
70	plasminogen
71	myosin light chain kinase
75	Heat shock 60kDa Protein
77	eukaryotic elongation translation factor 1 alpha
89	MHC II complement
90	ficollin-1 mRNA
96	apolipoprotein H (beta 2 glycoprotein)
98	elongation factor e2
110	SOUL protein (SOUL)/Heme binding protein 2/placental protein 23
114	branched chain alpha-keto acid, dehydrogenase E1-beta subunit
116	enoyl Coenzyme A hydratase, short chain, 1, mitochondrial
121	complement factor B mRNA
123	Vitronectin
124	Macaca fascicularis brain/HepBeAg binding protein 2
126	aldehyde dehydrogenase 2 family
129	sphingolipid activator proteins 1/prosaposin
133	coagulation factor XII
138	corticotropin releasing hormone binding protein
142	Cox 2 (Mitochondria)
157	complement factor H-related protein 4
158	decorin B
165	factor XIIb
181	Chrom 17/espin 2 (EPN2) transcript variant
188	procollagen C proteinase (bone morphogenetic protein BMP1)
190	aminocyclase-1 (ACY1)
200	NADH ubiquinone oxidoreductase
201	Threonine synthase-like-1
203	Complement component 6
208	aldolase B mRNA
214	ATPase Na K transporting, beta 1 polypeptide
229	Inter alpha trypsin inhibitor
230	tyrosine phosphate PTPase
234	beta actin
238	clathrin assembly protein complex 2
242	beta ureidopropionase
249	L-arginine:glycine amidinotransferase
250	ribonuclease T2 (ribonuclease B precursor)
262	catalase
266	dipeptidyl-peptidase 7
270	Galactose Mutarotase (aldolase 1 epimerase)
285	transferrin
293	lazarotene-induced gene 2 (TIG2)
298	dynactin B
299	choline dehydrogenase
309	Complement Factor H
331	Zinc finger protein 398
337	cysteine proteinase inhibitor precursor cystatin C
340	phenylalanine hydroxylase
342	dihydroipoamide dehydrogenase
353	HBV pre-s2 binding protein 1
354	putative DNA polymerase delta p38 subunit
360	sirtuin (silent mating type information regulation 2 homolog)
361	SAD1 unc-84 domain protein
363	Zn alpha 2 glycoprotein
369	lactamase, beta (LACTB), nuclear gene
370	cytoskeleton gamma actin
373	F box protein FBG2
387	actin beta chain
392	glyceraldehyde 3 phosphate dehydrogenase
409	adenosine kinase
410	alpha-1-B glycoprotein

Table 3: Unique in-frame sequences found in NS3 yeast 2 hybrid screen

Unique colony numbers (column 1) and the gene present within the AD-domain plasmid (column 2) are shown. The columns which are grey are columns which were selected for further examination.

replication. In most cases, the mechanism by which the virus affects these pathways remains unknown and therefore identification of the specific proteins involved would be an important advancement. Using these criteria, 23 proteins were selected for further study. The proteins were involved in pathways known to be affected by HCV infection such as oxidative stress (cox 2, NADH ubiquinone oxidoreductase), cell cycle control (cyclin G1 binding protein, elongation factor 1-alpha, elongation factor ef2, endothelial cell apoptosis protein E-CE1, vitronectin), immune regulation (inhibin beta E), or energy metabolism (adenosine kinsase, aldehyde dehydrogenase 2, beta ureidopropionase, catalase). Secondly, proteins involved in protein metabolisms (procollagen C proteinase, inter-alpha trypsin inhibitor, plasminogen) and in signal transduction (corticotropin releasing hormone binding protein, flotillin) were chosen for further evaluation as they could explain the ramifications of HCV infection on liver cells. Finally, proteins which were present multiple times in the Y2H screen (aldolase B fructose biphosphate, haptoglobin, apolipoprotein H, alpha 2-HS glycoprotein, insulin-like growth factor binding protein, transthyretin) were also further examined as redundancy in the screen increases the possibility that the interaction is real.

Typically, in an Y2H screen, it is difficult to tell whether all of the colonies evaluated are independent individual colonies, or if some may be daughter colonies. For example, several of the prey targets in this screen appeared multiple times; for example, metallothionein was seen 52 times (Appendix 1). In our study, however, the hypothesis that these may represent daughter colonies from a progenitor unique diploid mating is not supported by the fact that so many single unique sequences are present within the screen.

1.3. Protein-protein interactions between NS3 and individual prey of interest in AH109 yeast

In order to evaluate the specificity of the NS3/prey interactions, prey plasmids were introduced into yeast expressing different DBD-proteins and assessed for protein-protein interaction (colony growth). First, prey plasmids were isolated from yeast cells then transferred into bacteria. Plasmid DNA was purified and the prey insert sequence verified. The plasmid DNA was then retransformed into AH109 yeast cells expressing

DBD-NS3 (pGBK-T7-NS3, confirmation of interaction), DBD-lamin C (pGBK-T7-lamin C, negative control) or DBD-NS5A (pGBK-T7-NS5A) and was assessed for growth (indicating a potential interaction between the proteins). These experiments were carried out in duplicate.

The interaction with DBD-NS3 was used to confirm the original results of the Y2H screen. In every case growth of the yeast colonies was seen, thereby confirming that growth was specific for the AD-protein partner being evaluated (Table 4). The DBD-lamin control protein was used to screen out non-relevant or “sticky proteins”. This control thereby eliminated AD-polypeptides which 1) bound to the DBD rather than the NS3 sequence or 2) bound protein sequences irrespective of specificity or 3) can activate the yeast auxotrophic promoters and allow colony growth in the absence of an interaction with the DBD-NS3 sequence. Finally, AD-constructs were also screened against another HCV viral protein, HCV NS5A which is distinct from NS3 or lamin C in sequence and function.

Interaction types fell into three categories, examples of which are shown in Fig. 8. These classifications include AD-proteins that interacted with DBD-NS3 only (Fig. 8A), proteins that interacted with DBD-NS3 and DBD-NS5A (Fig. 8B), and proteins that interacted with DBD-NS3, DBD-NS5A and DBD-lamin C (Fig. 8C). Based on these results, nine AD-protein sequences were excluded from further analysis primarily because they bound lamin C (Table 3). Interestingly, nine other proteins interacted with both DBD-NS3 and DBD-NS5A but not with DBD-lamin C. It is unclear whether this is because the proteins are “sticky” and the DBD-NS5A prey represents a larger target than DBD-lamin C or whether some of these interactions may represent legitimate interactions. In fact, the binding of more than one hepatitis C virus protein to a single cellular protein is not without precedence. For example, the human proteins hVap-A and hVap-B interact with both NS5A and NS5B simultaneously (Tu *et al* 1999). For this reason, a positive interaction of one prey protein with both NS5A and NS3 does not necessitate that the interaction is false. Also, NS5A and NS3 are both believed to be present in the virus replication complex, and as both may physically close together, both may indeed interact with the same host cellular protein.

Protein Name	Colony # studied	# of times interaction found	Interaction with NS3	Interaction with NS5A	Interaction with Lamin
Adenosine Kinase	409	2	YES	NO	NO
Cox 2	142	9	YES	NO	NO
elongation factor ef2	98	1	YES	NO	NO
Haptoglobin	25	11	YES	NO	NO
transthyretin (pre albumin)	285	2	YES	NO	NO
aldehyde dehydrogenase 2	125	2	YES	YES	NO
Aldolase B fructose biphosphate (ALDOB)	208	20	YES	YES	NO
catalase	262	1	YES	YES	NO
cyclin G1 binding protein	41	1	YES	YES	NO
insuline-like growth factor binding protein	18	3	YES	YES	NO
Inter alpha trypsin inhibitor	229	4	YES	YES	NO
NADH ubiquinone oxidoreductase	200	1	YES	YES	NO
plasminogen	70	3	YES	YES	NO
Vitronectin	123	1	YES	YES	NO
alpha 2 -HS glycoprotein	53	5	YES	NO	YES
endothelial cell apoptosis protein E-CE1	27	1	YES	NO	YES
apolipoprotein H, beta 2 glycoprotein 1	96	23	YES	YES	YES
beta ureidopropionase	242	2	YES	YES	YES
corticotropin releasing hormone binding protein	136	2	YES	YES	YES
elongation factor 1-alpha	77	1	YES	YES	YES
flotillin	90	3	YES	YES	YES
inhibin beta E	46	1	YES	YES	YES
procollagen C proteinase	186	4	YES	YES	YES

Table 4: Summary of Y2H specificity interactions between prey constructs and DBD-NS3, DBD-NS5A or DBD-lamin.

Prey constructs were transformed into yeast AH109 cells expressing one of three DNA binding domain fusion proteins (NS3, NS5A or lamin). The sequences found in the screen (protein name), the colony used in the study and the number of times the interaction was found are indicated along with the protein-protein interaction status (as indicated by colony growth). There are three different types of proteins, the white proteins interacted only with NS3, the light grey proteins interacted with NS3 and NS5a and the dark grey proteins interacted with NS3 and lamin (and usually NS5A as well).

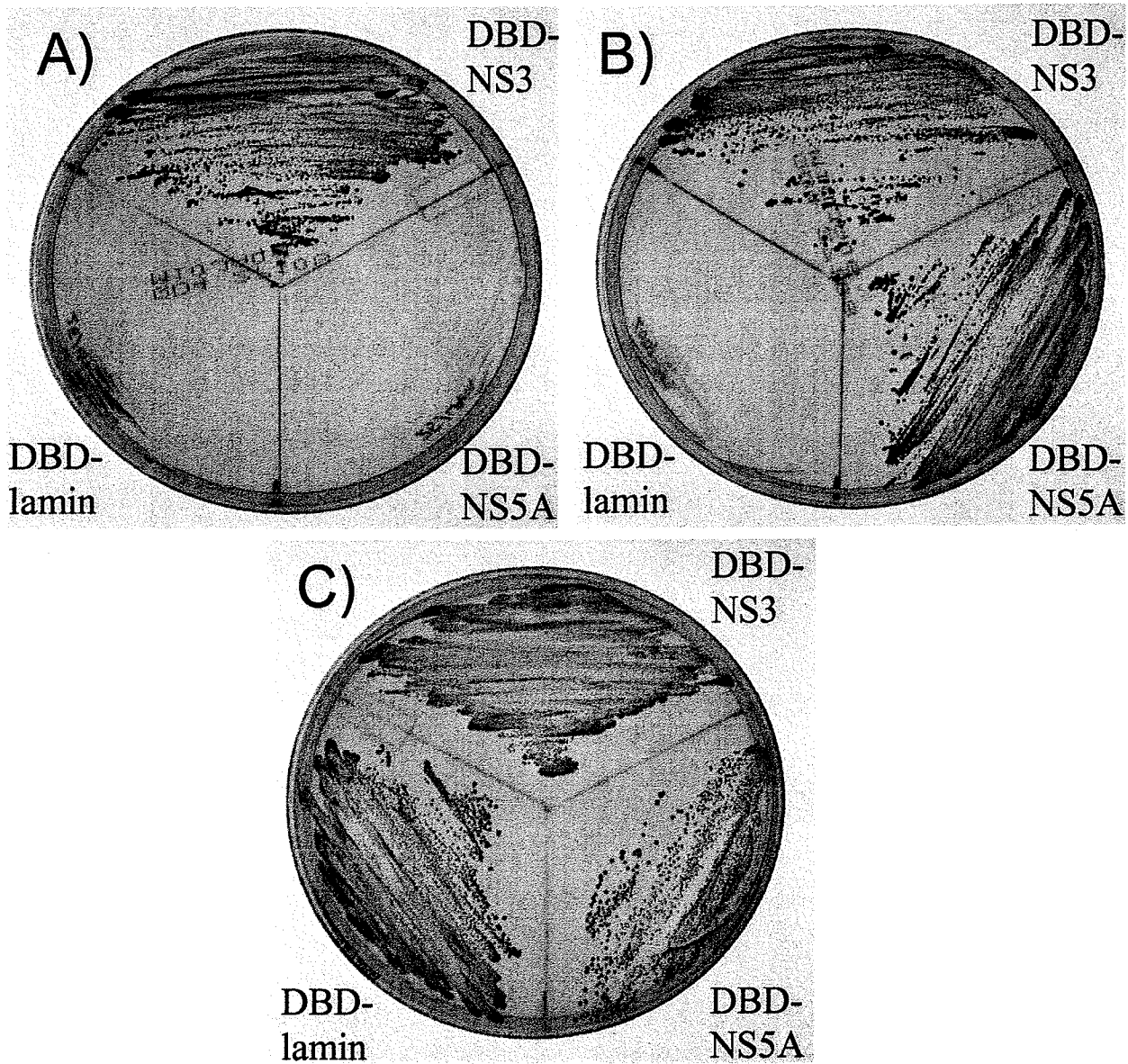


Figure 8: Specificity of Y2H interactions for selected prey protein

To further analyze the specificity of the prey sequences found in the NS3 Y2H screen, the AD- containing plasmids were transformed into yeast expressing either DBD-NS3, DBD-NS5A or DBD-lamin C. The yeast transformants were plated on selective media to test for a positive physical interaction (indicated by growth). Interactions types fell into three different categories.

- A) AD-haptoglobin (colony 25) interacted only with DBD-NS3
- B) AD-catalase (colony 262) interacted with both DBD-NS3 and DBD-NS5A
- C) AD-flotilin (colony 90) interacted with DBD-NS3, DBD-NS5A and DBD-lamin C.

Table 4 also examines the number of times that the interacting protein partner was independently detected in the Y2H screen. When a protein is independently identified multiple times, this could either be due to a strong interaction between NS3 and the binding partner, or the fact that the binding partner exists at high levels in the liver library and therefore has a greater chance of coming into contact with NS3. These possibilities were not further evaluated at this point.

Of the 23 AD-prey proteins tested, only 5 demonstrated a positive interaction with DBD-NS3 only. These proteins included adenosine kinase, cox2, elongation factor ef2, haptoglobin and transthyretin. These targets represent the most specific interactions within the Y2H NS3 screen found to date and represent the highest potential for having a legitimate interaction with HCV NS3. Although the Y2H analysis allowed identification of putative binding partners of NS3, an alternative approach was required to confirm these interactions. Therefore, examination by an *in vitro* physical interaction (co-precipitation) was necessary. However before this could be performed, it was necessary to generate a recombinant NS3 protein to serve as a target for the prey sequences identified above.

2.0. Expression and Purification of NS3

2.1. Expression of NS3 1a in bacteria

To develop an *in vitro* “pull-down” assay, we wished to express and purify recombinant NS3 to serve as a binding target. ³⁵S-methionine radio-labeled proteins representing the prey proteins of interest would then be added to an NS3 interaction reaction. This strategy was chosen for several reasons. Firstly, antibodies are not available for each prey protein to be examined, and secondly, NS3 on its own is known to induce an apoptotic response in many of the cell lines commonly used for HCV research (Prikhod'ko *et al* 2004) and therefore could limit our ability to detect interactions occurring in cells. Finally, we wished to develop a recombinant protein that could be

used for the development of anti-sera against NS3, as commercially available sources have proven to be expensive and generally of poor quality (data not shown).

Initially, NS3 expression was attempted using full length NS3 strain 1a fused to glutathione-S-transferase (GST) in the expression vector pGEX-2T. The GST fusion tag eases protein purification and as anti-sera against NS3 is of limited quality, it would allow unambiguous detection of the recombinant protein. However, following bacterial transformation and induction GST-NS3 expression in the bacteria multiple times, NS3 was not readily visible by Coomassie stained SDS-PAGE analysis of harvested bacterial protein extracts (Fig. 9A, compare lanes 1 and 2). Western blot analysis with a GST specific antibody showed no GST-NS3 in the induced bacterial sample (Fig 9B, lane 5) even though induction performed with a control protein (GST-Grb2) was effective (Fig. 9B, lane 4). In fact, even without induction the GST-Grb2 bacteria showed a small amount of protein production (Fig 9B, lane 3).

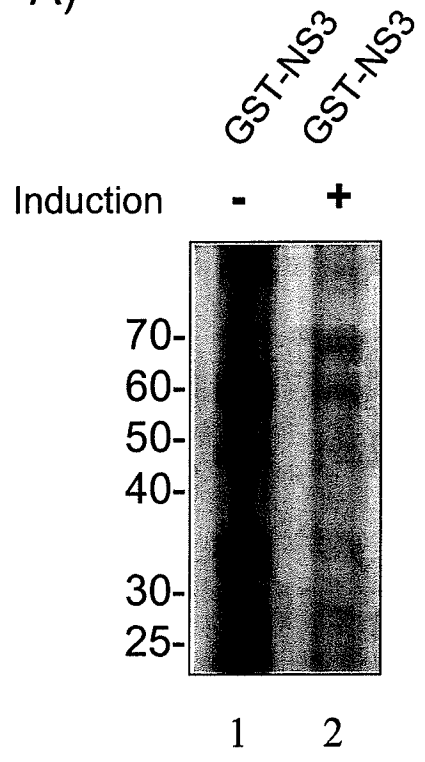
Considering that NS3 has a functional protease activity, toxicity to the bacteria may not be unanticipated. Therefore, effect on bacterial growth was assessed by comparing the growth rates of bacterial cells in the presence or absence of NS3 (Fig 9C). These experiments were done in triplicate and it was found that even without induction, bacteria harboring pGEX2T-NS3 grew slower than either an empty vector (pGET-2T) or pGEX-2T containing a control protein, Grb2. Over time, the bacteria harboring pGEX-2T-NS3fl appear to recover, and this may be due to an accumulation of mutations in NS3 or plasmid loss alleviating the growth suppression on the bacteria.

To increase the expression levels of NS3 in bacteria, other groups have successfully cloned and expressed the protease domain of NS3 alone (Vishnuvardhan *et al* 1997). Although this would limit the interaction studies we wished to perform with NS3, it would still allow confirmation of protease domain interactions and be a useable reagent for antibody production. Therefore, the amino-terminal region of NS3 (aa 1-200) was cloned into the pHAT bacterial expression vector to generate the construct HAT-NS3p. Here, the GST tag was replaced with a HAT tag in order to reduce the size of the affinity tag and also, since the HAT tag is a modified Ni-binding tag would allow purification under reducing conditions in case the protein was insoluble (Fig. 10A).

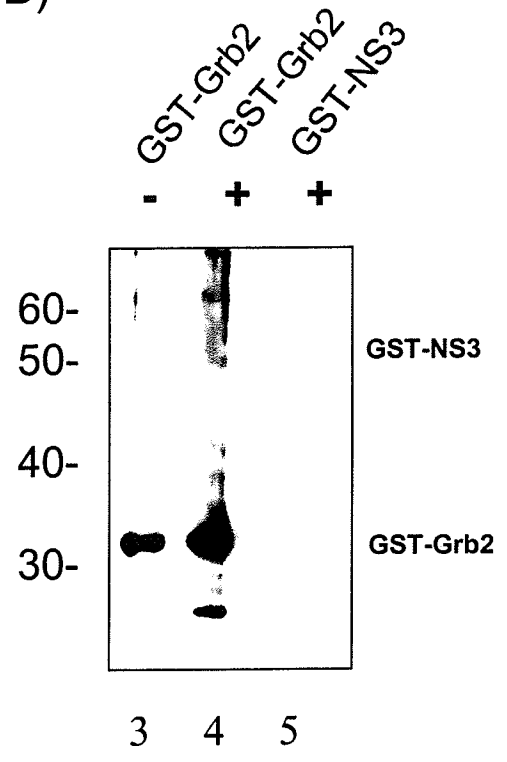
Figure 9: Induction and growth rate of pGEX-2T-NS3 in Escherichia coli BL21(DE3) bacteria cells.

- (A) Bacterial protein extracts were prepared from strains harboring GST-NS3 or GST-Grb2 without (lanes 1 and 3) or with induction of the recombinant protein by IPTG (lanes 2,4 and 5). Coomassie blue stained gel (lanes 1 and 2).
- (B) Western blot with antibody to GST (lanes 3-5). NS3 is not visibly expressed in lane 3, whereas the positive control GST-Grb2 expressed well.
- (C) Bacterial cultures containing various GST-protein constructs were grown without induction in liquid culture. OD₆₀₀ readings were taken every 30 minutes. GST-NS3 (blue) shows a decreased growth rate compared to an empty vector (yellow) and the positive control Grb2 (pink).

A)



B)



C)

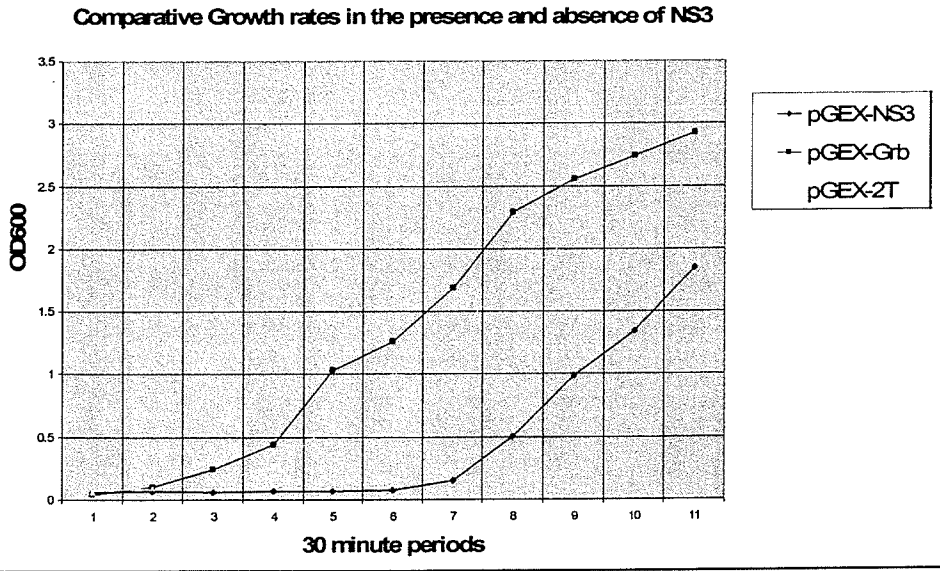
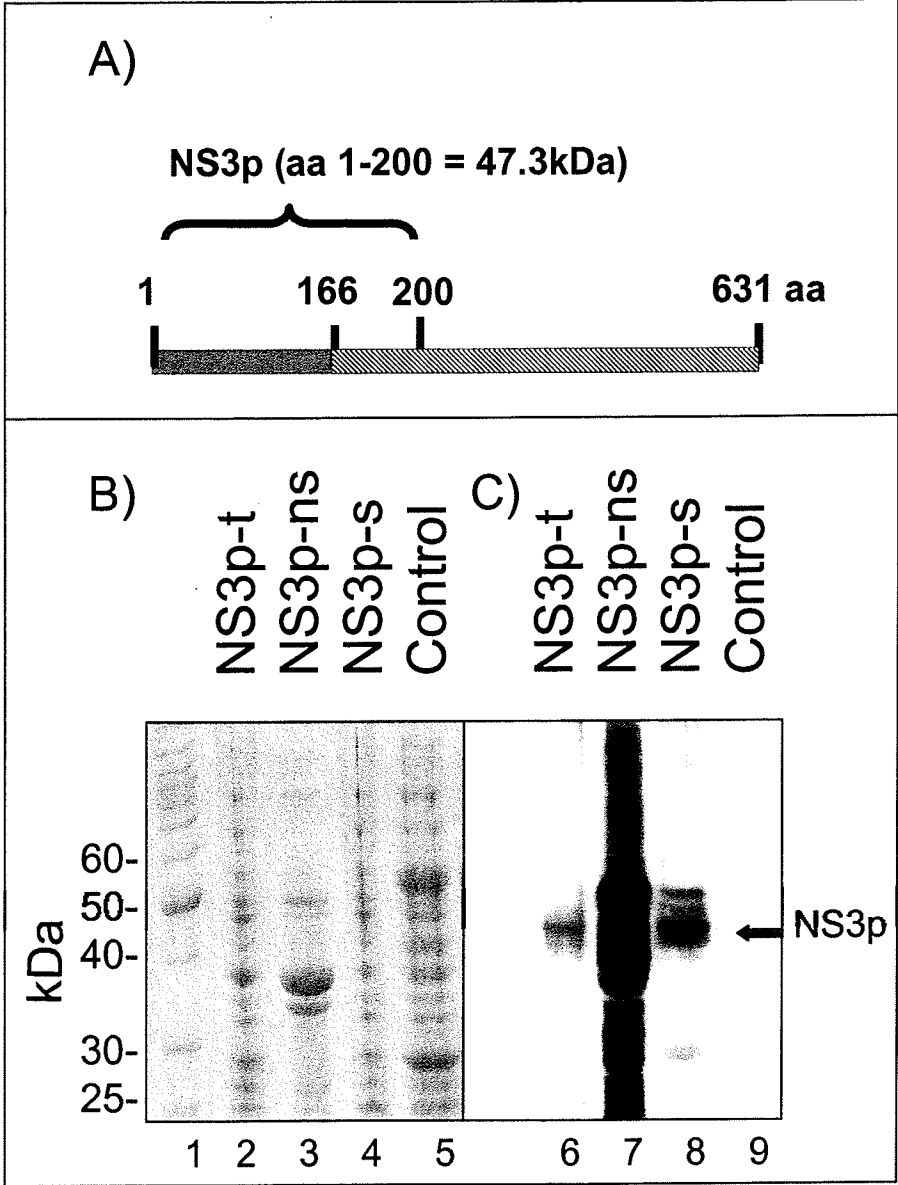


Figure 10: Expression of pHAT-NS3p 1a in Escherichia coli BL21(DE3) cells

- (A) Cartoon depicting the region of NS3 used to generate NS3p (NS3 protease domain) (aa 1-200). Gray shaded area (aa 1-166) corresponds to the minimal protease domain within NS3.
- (B) Coomassie blue stained SDS-PAGE depicting (lane 1) Molecular weight markers, (2) total extract (NS3p-t), (3) insoluble fraction (NS3p-ins) and (4) soluble fraction (NS3p-s) from a bacterial culture induction of pHAT-NS3p, (5) control bacterial culture containing only the pHAT expression vector. Anticipated position of recombinant NS3p is indicated.
- (C) Western blot of samples from part (B) probed with anti-Hat tag antibody. (lane 6) total extract (NS3p-t), (7) insoluble fraction (NS3p-ins) and (8) soluble fraction (NS3p-s) from a bacterial culture induction of pHAT-NS3p, (9) control bacterial culture containing only the pHAT expression vector.



Following multiple inductions and analyses of bacterial samples by Coomassie-stained SDS-PAGE, no recombinant HAT-NS3p was clearly evident compared with a control lane (Fig 10B, compare lanes 2 and 5). However HAT-NS3p was visible on a Western transfer probed with anti-HAT antisera (Fig. 10B, lanes 6-9). The Western blot, however, demonstrated two problems. Firstly, a large proportion of the NS3p protein was insoluble (Fig 10B, lane 7), and secondly, that the protein was degrading or was being incompletely translated (lanes 6-8).

When proteins are misfolded in bacterial expression systems, they often form insoluble inclusion bodies which can often protect the full length protein species. Alternatively, at lower expression levels, misfolded proteins can be targeted for degradation. To increase the amount of soluble Hat-NS3p protein being produced and hopefully decrease degradation due to protein misfolding, pHAT-NS3p was expressed in bacterial cell lines containing various plasmid vectors which express a variety of chaperone proteins (Fig. 11A). Chaperone proteins are known to help fold proteins properly, thus aiding in both the problems of insolubility and degradation/incomplete translation (Nishihara *et al* 1998, Nishihara *et al* 2000). Soluble extracts from bacteria expressing pHAT-NS3p were prepared and electrophoresed on a 12.5% SDS-PAGE gel and stained with Coomassie blue stain (Fig 11B). The Coomassie stained gel indicates that the extracts were equally loaded and that expression of the chaperone proteins was evident. For example the over-expressed chaperone proteins are easily identifiable on the Coomassie blue gel, especially GroEL (Fig 11B, lane 6, 63kDa induced species) and tig (lane 7, 55kDa induced species). Western analysis with anti-Hat anti-sera demonstrated that extracts prepared from bacteria over-expressing GroES, GroEL and tig allowed for highest expression of soluble Hat-NS3p (Fig. 11A lane 7) in comparison to the expression of Hat-NS3 in a bacterial strain without chaperone proteins (Fig. 11A, lane 2). Increased length of exposure of this blot revealed HAT-NS3p weakly in all lanes. Hat-NS3p 1a was also grown in a bacterial background which more accurately reflects the codon bias seen in human cells (Rosetta *E. coli*, Novagen). This genetic background also increased the amount of soluble HAT-NS3p (Fig. 11, lane 8). Duplicate experiments gave the same results. Of all the bacterial expression lines, the background expressing GroES/EL and tig gave the best results and further studies with pHAT-NS3p were

Figure 11: Expression of HAT-NS3p in the presence of protein chaperone proteins, or human codon biased tRNA bacterial strains.

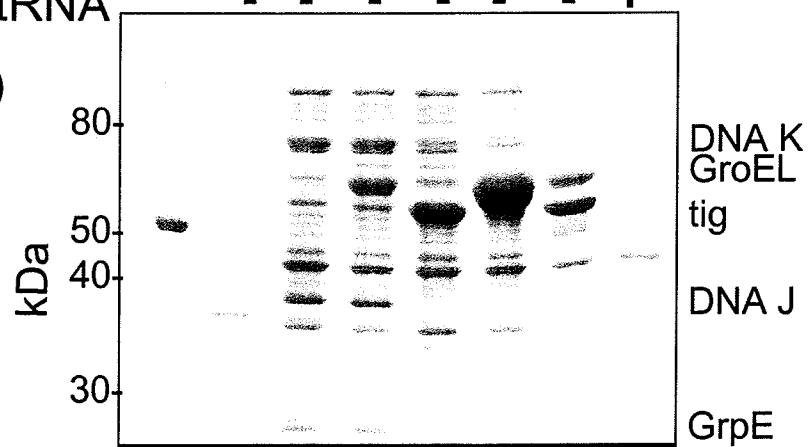
- (A) List of bacterial expression vectors used to express specific chaperone proteins.
- (B) Western analysis of the protein extracts from section(C) using anti-sera detected against HAT tag (to detect HAT0NS3p). Lanes are as for (B). The expected MW for full-length HAT-NS3p is indicated.
- (C) Soluble protein extracts were prepared from bacterial strains expressing HAT-NS3p and chaperone protein/codon biased tRNAs prior to SDS-PAGE and detection with Coomassie Brilliant Blue. Lanes: (1) MWM. Extracts prepared from the various bacterial chaperone strains: (2) no chaperones, (3) DNA J/K+grpE, (4) groES/EL+ DNA J/K+grpE, (5) tig + grpE (6) gro ES, (7) gro ES + tig, (8) humanized codon biased tRNAs (strain Rosetta).

A)

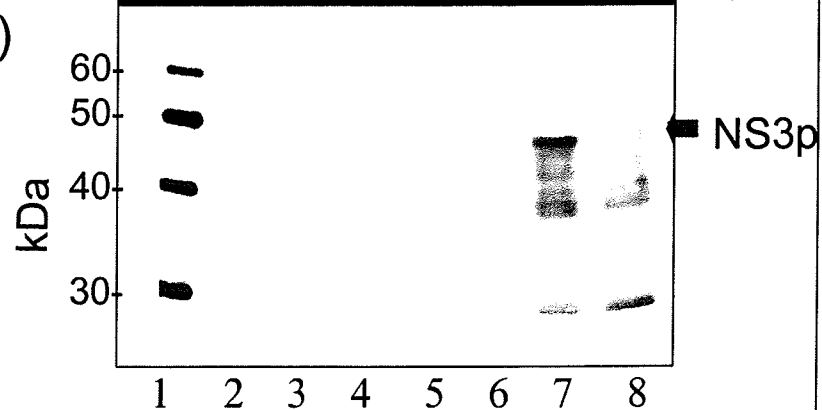
Vector	Chaperones
pG-Tf2	groES, groEL, tig
pTf16	tig
pKJE7	dnaK, dnaJ, grpE
pG-KE7	dnaK, dnaJ, grpE, groES, groEL
pGro7	groES, groEL

groES/EL	-	-	+	-	+	+	-
tig	-	-	-	+	-	+	-
Dna K/J	-	+	+	-	-	-	-
grpE	-	+	+	+	-	-	-
CtRNA	-	-	-	-	-	-	+

B)



C)



conducted in a pG-Tf2 background expressing these chaperones. Although increased amounts of HAT-NS3p were produced, the protein was still largely in a degraded form. Therefore attempts were made to purify the protein to ultimately obtain a single protein species. The expressed pHAT-NS3p, in *E. coli* expressing pG-Tf2, was purified on a nickel chelate column by fast protein liquid chromatography (Fig. 12A). During the purification, protein (presumed to be NS3p) eluted over two fractions (A2 and A3) and little was lost in the flow-through. The concentrated protein eluted from the nickel chelated column, however, was still not clearly evident by Coomassie blue gel (Fig. 12B). It was, however, visible on a Western blot probed with anti-HAT anti-sera (Fig. 12C). The Western blot, however, still demonstrated the presence of degradation products, or incomplete translation indicating that further purification and larger cultures sizes would be required to isolate HAT-NS3p.

Due to the high levels of degradation seen with expressed HAT-NS3p strain 1a in backgrounds expressing a variety of chaperone proteins or rare tRNAs, a second truncated product was generated, HAT-NS3ps (aa 1- 150, Fig. 13A), which stands for NS3 protease domain short version. This construct had detectable levels of soluble expression, as seen on a western blot probed for anti-HA (Fig. 13B, lanes 2 and 3) in multiple experiments, but still had ample soluble degradation products or incomplete translation.

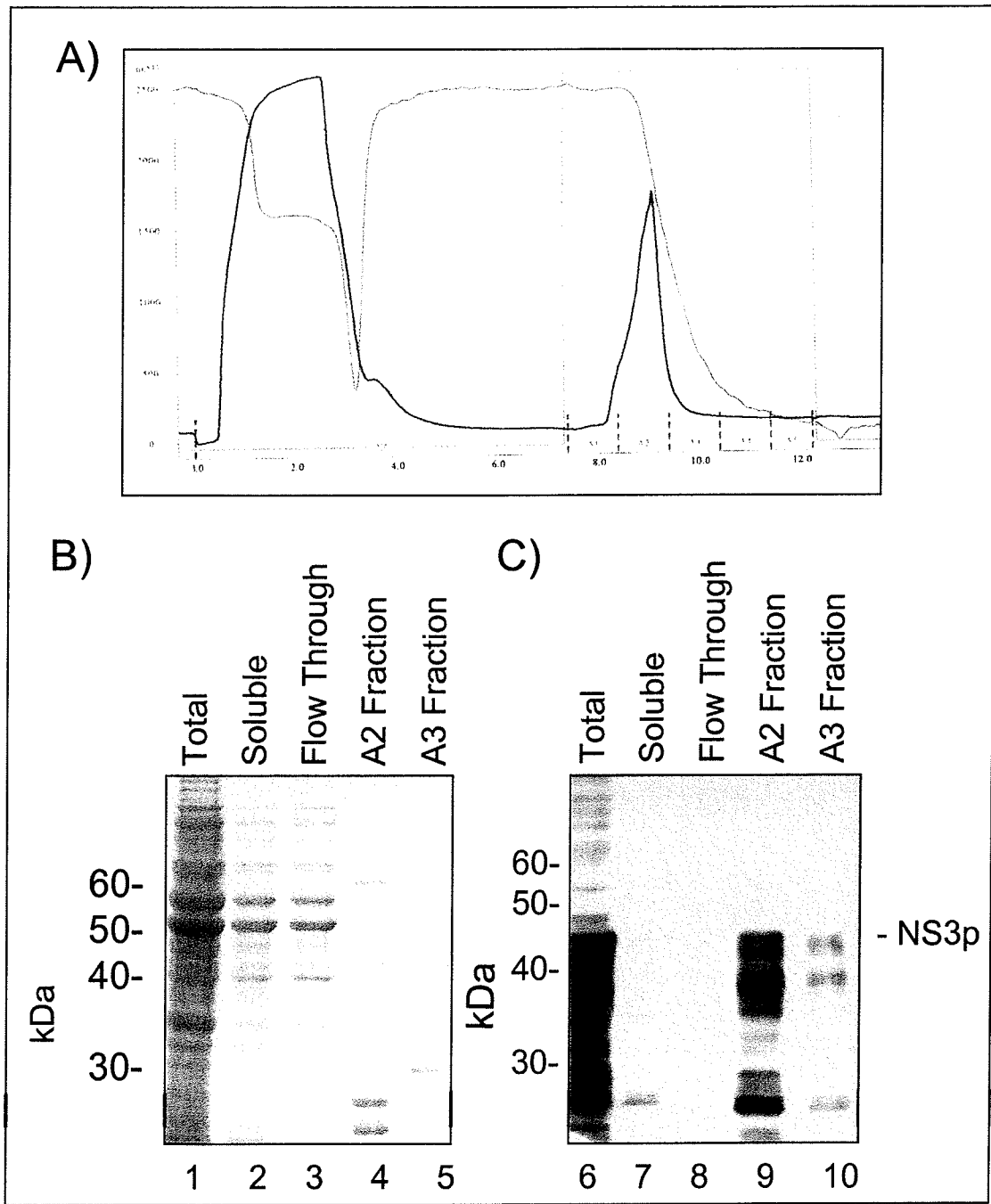
The presence of large quantities of degradation products or incomplete translation in either HAT-NS3p, or HAT-NS3ps, regardless of the presence of chaperone or tRNA molecules, suggested that these products would not easily be developed for complete purification and therefore use of the HCV strain 1a NS3p protein was abandoned.

2.2. Expression of NS3 1b in bacteria

Although strain 1a remains the prototype hepatitis C virus and is the strain we used for the Y2H screen, other investigators have successfully used different strains of HCV as a source for protein expression (Poliakov *et al* 2002). According to the Clustal algorithm, difference at the amino acid level between NS3 strain 1a and NS3 strain 1b is less than

Figure 12: Expression and purification of pHAT-NS3p 1a in a pG-Tf2 bacterial background.

- (A) A bacterial culture containing pHAT-NS3p and pGTf2 was grown and induced and the soluble extract purified on an FPLC using Ni chelate resin. Blue line indicates absorbance (A_{280}), brown line indicates relative conductivity, green line indicates elution gradient profile with imidazole.
- (B) Fractions from the FPLC run were evaluated by SDS-PAGE and visualized using Coomassie blue stain. Lanes: (1) Total extract, (2) Soluble extract, (3) Flow through Ni resin, (4) Ni-resin A2 fraction (5) Ni-resin A3 fraction.
- (C) Fractions from the FPLC run were evaluated by Western blot with anti-HAT antiserum. Lanes: (6) Total extract, (7) Soluble extract, (8) Flow through Ni resin, (9) Ni-resin A2 fraction (10) Ni-resin A3 fraction.



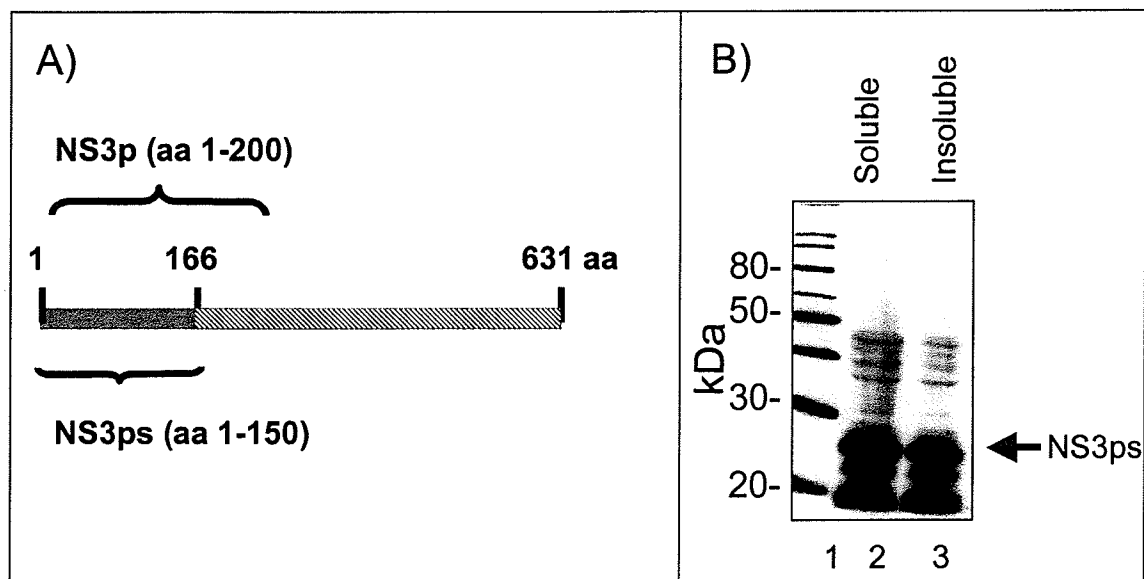


Figure 13: Expression of pHAT-NS3ps 1a

- (A) A cartoon depicting the coding region of NS3 contained in pHAT-NS3ps 1a and a comparison to pHAT-NS3p. Amino acids 1-166 represent the minimal proteinase domain.
- (B) Extracts from bacterial cultures expressing pHAT-NS3ps were evaluated by Western blot analysis probed with anti-HAT antibodies. Lanes: (1) MW markers, (2) soluble fraction (3) insoluble fraction. Position of full-length NS3ps is indicated with an arrow.

7% (45/632 amino acids) (Fig. 14). Of these differences, 34 represent highly similar substitutions, 4 are similar and only 7 are biophysically significantly different. Due to the similarity between NS3 strain 1a and NS3 strain 1b, it was felt that the purified NS3 strain 1b would still represent a valid binding partner for “pull-down” studies if it could be purified.

Initially, a segment of NS3 strain 1b homologous to the HAT-NS3ps clone was developed and termed HIS-NS3ps-1b. As seen with the prior NS3 expression studies, over-expression of HIS-NS3ps-1b in bacteria was not detectable by Coomassie blue staining (data not shown); however, upon Western analysis, the protein migrated at a position consistent with the size expected for HIS-NS3ps-1b. (22.9kDa, Fig.15B, lanes 1). The protein also appeared to be at least 50% soluble (Fig 15B, compare lanes 2 and 3). To determine if the protein was competent for purification, soluble extract was applied to a Ni-affinity column and purified by FPLC where it eluted over the course of several fractions. Several other proteins were seen in the HIS purification, as visualized by Coomassie Blue stained gel (Fig. 15C) requiring that further purification steps would be necessary and a western blot probing with HIS antibodies demonstrated the presence of the NS3ps protein (Figure 15D lanes 7-9). Importantly, however, there was little degradation of HIS-NS3ps-1b as seen by Western analysis probed with anti-HIS antibodies (Fig. 15D).

Prompted by the successful expression and partial purification the NS3ps strain 1b protease domain with minimal degradation, the possibility of expressing a full-length form of the NS3-1b protein was attempted. Data from Fig. 15B suggests that approximately 50% of the expressed protein remained insoluble even in the presence of detergent (CHAPS). Previous groups have suggested that incorporation of a portion of the natural viral binding partner NS4A to NS3 at the amino-terminus of NS3 could further stabilize the protein and increase the solubility of the expressed protein (Taremi *et al* 1998). Therefore a primer designed to include the first 22 residues of NS4A was synthesized and used to attach this region to the N-terminus of the NS3-1b coding

```

1a      MAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYHGA
1b      MAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYHGA
*****;*****;*****;*****;*****;*****;*****

1a      GTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRLTPCTCGSSDLYLVTRHADVIPVRRR
1b      GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRR
*::*:*.****;***** ** *;*****

1a      GDSRGSLLSPRPI SYLKGSSGGPLLCPAGHAVGLFRAAVCTRGVAKAVDFIPVENLGTMM
1b      GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGI FRAAVCTRGVAKAVDFVPVESMETMM
*****;*****;****;*****;***; ***

1a      RSPVFTDNSSPPAVPQSFQVAHLHAPTGS GKSTKVPAAYAAQGYKVLVLNPSVAATLGFG
1b      RSPVFTDNSSPPAVPQTFQVAHLHAPTGS GKSTKVPAAYAAQGYKVLVLNPSVAATLGFG
*****;*****

1a      AYMSKAHGVDPNIRTGVRTITTGSPITYSTYGKFLADGGC SGGAYDIIICDECHSTDATS
1b      AYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLADGGC SGGAYDIIICDECHSTDSTT
*****;*****;*****;*****;*****;*****;*:

1a      ILGIGTVLDQAETAGARLVVLATATPPGSVTVSHPNIEEVALSTTGEIPFYGKAIPLEVI
1b      ILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFYGKAIPLETI
*****;*****;*****;*:

1a      KGGRHLIFCHSKKKCDELA AKLVALGINAVAYYRGLDVSVIPTSGDVVVVSTDALMTGFT
1b      KGGRHLIFCHSKKKCDELA AKLSGLGLNAVAYYRGLDVSVIPTSGDVI VVATDALMTGFT
***** .**;*****;**;*****

1a      GDFDSVIDCNTCVTQTVD FSLDPTFTIETTLPQDAVSRTQRRGRTGRGKPGIYRFVAPG
1b      GDFDSVIDCNTCVTQTVD FSLDPTFTIETTVPQDAVSRQRRGRTGRGRMGIIYRFVTPG
*****;*****;*****;*****;**

1a      ERPSGMFDSSVLC ECYDAGCAWYELTPAETTVRLRAYMNT PGLPVCQDHLEFWEGVFTGL
1b      ERPSGMFDSSVLC ECYDAGCAWYELTPAETSVRLRAYLNT PGLPVCQDHLEFWESVFTGL
*****;*****;*****;*****

1a      THIDAHFLSQTKQSGENFPYLVA YQATVCARAQAPPPSWDQMWKCLIRLKPTLHGPTPLL
1b      THIDAHFLSQTKQAGDNFPYLVA YQATVCARAQAPPPSWDQMWKCLIRLKPTLHGPTPLL
*****;*:*****

1a      YRLGAVQNEVTLTHPITKYIMTCMSADLEVVT
1b      YRLGAVQNEVTTTHPITKYIMACMSADLEVVT
***** *****;*****

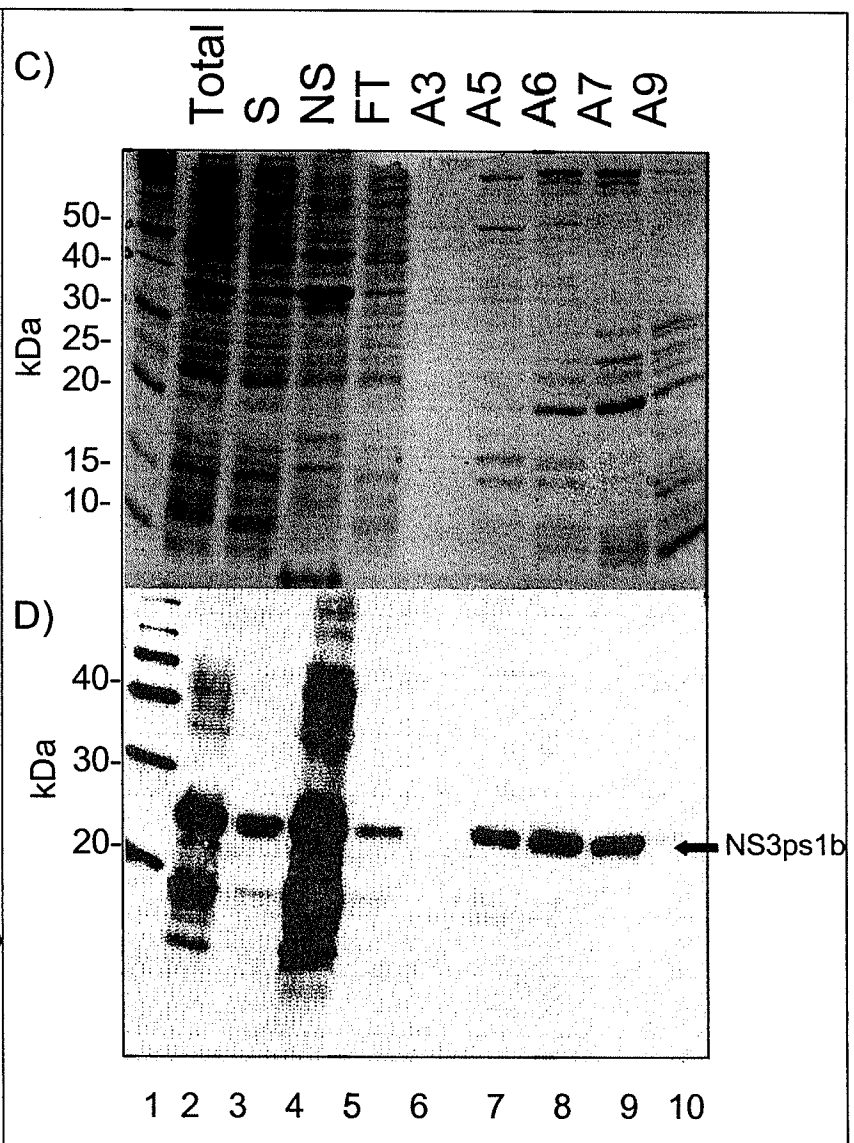
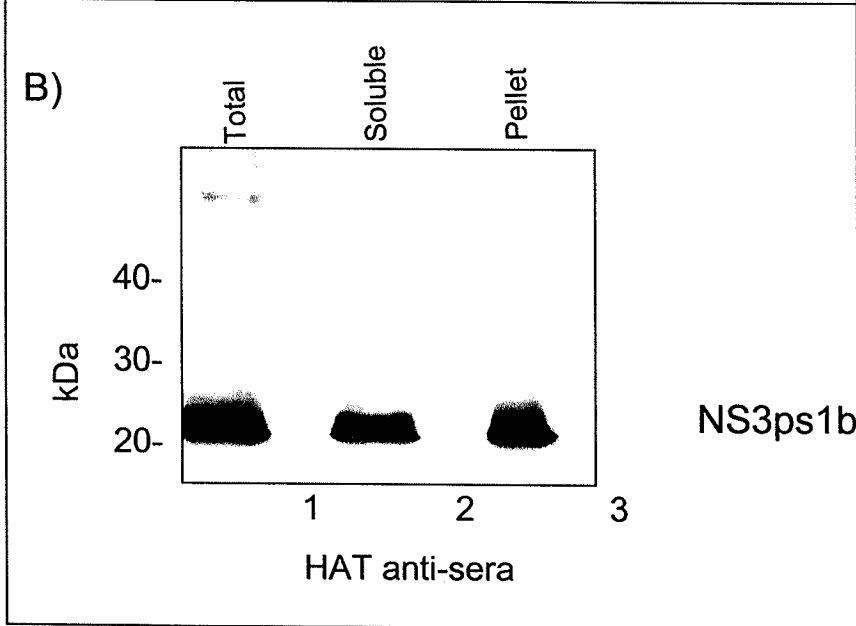
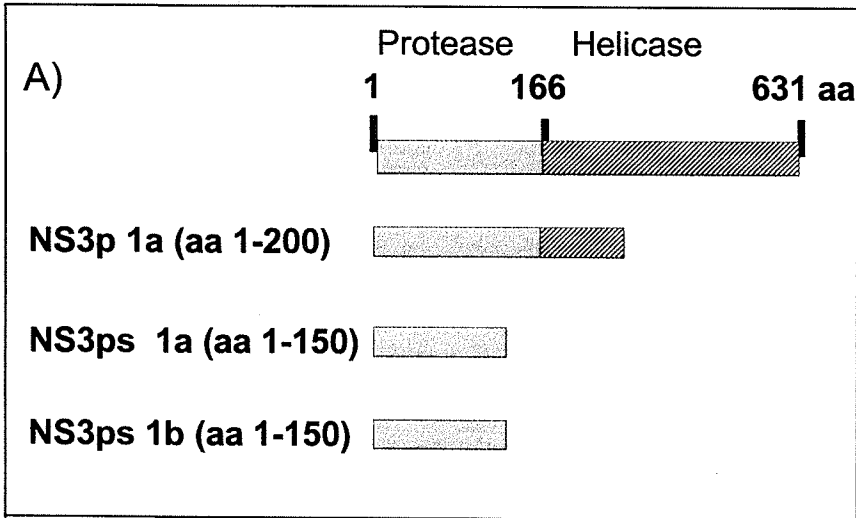
```

Figure 14. Protein alignment HCV NS3 strain 1a and HCV NS3 strain 1b.

Protein sequences for the HCV strains 1a and HCV strain 1b were aligned using Clustal X (1.81) (*) Identical amino acids, (:) highly similar amino acids, (.) similar amino acids and () amino acids with dissimilar physical attributes.

Figure 15: Expression and purification of pHAT-NS3ps 1b

- (A) A cartoon depicting the various sections of the NS3 protein expressed in bacteria.
- (B) Western blot analysis of bacterial expressed HIS-NS3ps-1b protein detected using anti-HIS Ab. Lanes: (1) Total extract, (2) soluble extract, (3) Insoluble extract.
- (C) Western blot analysis of extracts obtained from NS3ps-1b protein purification on Ni-resin. Lanes: (1) MWM, (2) Total extract prior to binding column, (3) soluble extract (4) insoluble extract (5) flow through during column purification, (6) fraction A3, (7) fraction A5, (8) fraction A7, (9) fraction A9.
- (D) Coomassie blue stained SDS PAGE gel analysis of purification of NS3ps-1b by nickel chelate column chromatography. Lanes are as per section (C) above.

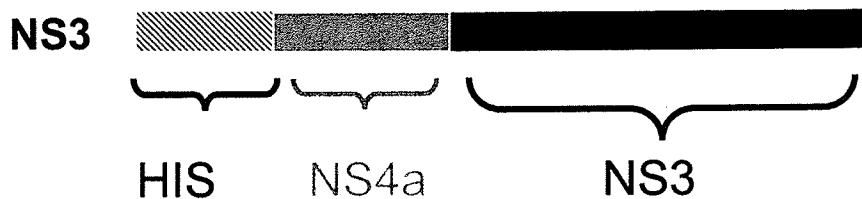


sequence (Fig.16). The 4A coding region was placed downstream of the His6 affinity tag. Clone HIS-NS3/4Apep-1b was expressed in bacteria and a protein at the predicted molecular weight for NS3/4Apep-1b (~71 kDa) was seen by Coomassie blue stained SDS-PAGE (Fig. 17A, compare lanes 1 and 2). Extracts were prepared and loaded onto nickel chelate columns for purification. The protein was most concentrated in fractions A11-B1 (Fig. 17B, lanes 2-4) and these were pooled, concentrated and applied to a 50 ml gel filtration column to obtain a further purified protein. Interestingly, two major peaks were seen during the chromatography run (not shown). Electrophoretic analysis of the two fractions suggested that both contained the NS3/4A-1b protein (Fig 17C, lanes 1-4). The reason for the two elution peaks is currently unknown.

To confirm that the protein purified was indeed HIS-NS3/4A-1b, anti HIS anti-sera was used for Western analysis. (Figure 18). The HIS anti-sera, however, was unable to detect this protein species despite being able to detect a control protein containing the identical HIS epitope (control protein, Carpenter, unpublished data). The non-specific binding in the control lane is due to the HIS6 antiserum binding non-specifically to other proteins expressed in the total bacterial lysate. These proteins are not in the NS3/4A1b lanes as these proteins have been purified. Due to the HIS and commercial NS3 antibodies working poorly to identify the purified protein, an alternate means was taken to prove that the purified protein was NS3.

2.3. Characterization of NS3/4a 1b by mass spectrometry

Purified HIS-NS3/4Apep-1b were subjected to SDS-PAGE then bands corresponding to the purified protein species were cut out and the samples were trypsin digested and evaluated by MS (mass spectrometry) and MS/MS (tandem MS). Fig. 19A shows the spectrum from a discrete nanospray (hollow-bore glass needle) injection of the digested protein into a Q-TOF mass spectrometer. Ion species corresponding to a theoretical trypsin digestion of the HIS-NS3/4Apep-1b protein are indicated with arrows. These fragments gave 20% coverage of the protein. In order to unequivocally identify this protein species, 2+ and 3+ ions from the nanospray analysis were subjected to energy



5051	CTTTAAGAAG	GAGATATACC	ATGGGCAGCA	GCCATCATCA	TCATCATCAC	
			M G S S	H H H	H H H	His6
5101	AGCAGCGGCC	TGGTGCCGCG	CGGCAGCCAT	ATGGGTTCTG	TTGTTATTGT	
	S S G L	V P R	G S H	M G S V	V I V	4A
5151	TGGTAGAATT	ATTTTATCTG	GTAGTGGTAG	TATTACGGCC	TACTCCCAAC	
	G R I	I L S G	S G S	I T A	Y S Q Q	NS3
5201	AGACGCGAGG	CCTACTTGGC	TGCATCATCA	CTAGCCTCAC	AGGCCGGGAC	...
	T R G	L L G	C I I T	S L T	G R D	...

Figure 16: Sequence of the HIS-NS3/4Apep-1b region at the N-terminus.

NS3 has been found to be more soluble upon the addition of a NS4A peptide fused to the amino-terminal portion of the protein. A 6-Histidine tag (red) was added to the front of the fusion protein, followed by a portion of the 4A peptide (blue) (SSGLVPRGSHMGSVVIVGRIL). This sequence was attached to the amino terminal portion of NS3 via a linker peptide (pink) (SGSGS). Note that the amino-terminal sequence (MGSS) is pET29 vector derived.

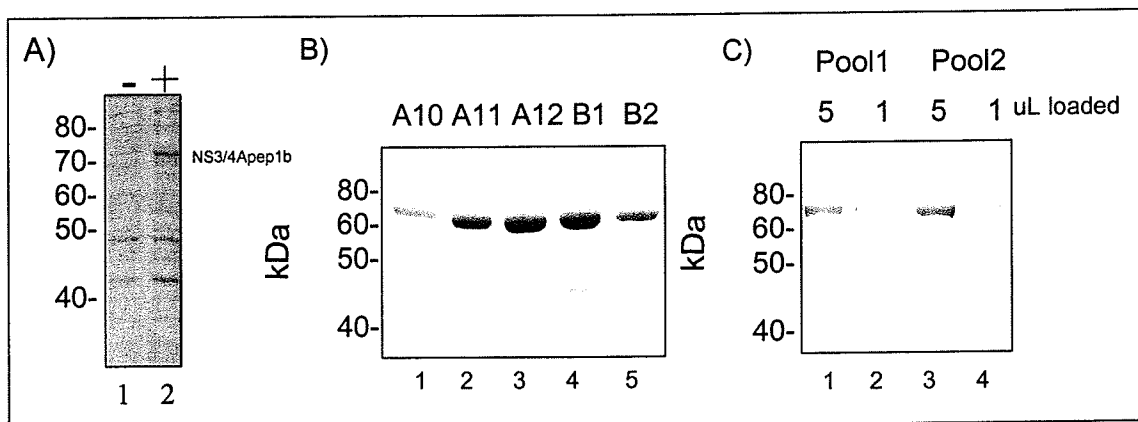


Figure 17: Purification of full-length HIS-NS3/4Apep-1b by Ni chelate and size exclusion chromatography.

- A) Bacterial extracts containing pHAT-NS3/4Apep-1b without (1) or with (2) IPTG induction were electrophoresed by SDS-PAGE and detected with Coomassie blue stain.
- B) The induced sample of pHAT-NS3fl/4A 1b was loaded on a 1mL nickel chelate column for purification via FPLC. A protein consistent with the size expected for NS3/4Apep-1b eluted at approximately 100mM imimidazole in fractions A10-B2.
- C) Fractions A10-B2 were pooled and concentrated before loading on a 50mL gel filtration column. Two peaks were detected (data not shown). Proteins present in the two pools were evaluated by SDS-PAGE and Coomassie blue gel staining. Samples were evaluated at two different concentrations (1 μ L and 5 μ L).

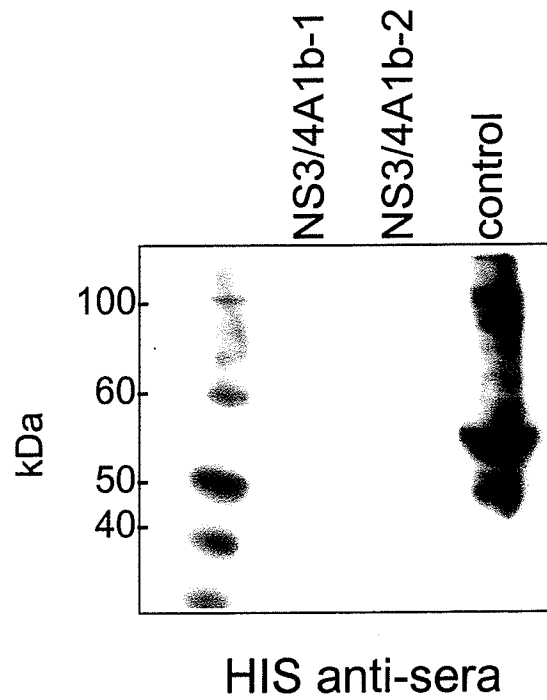


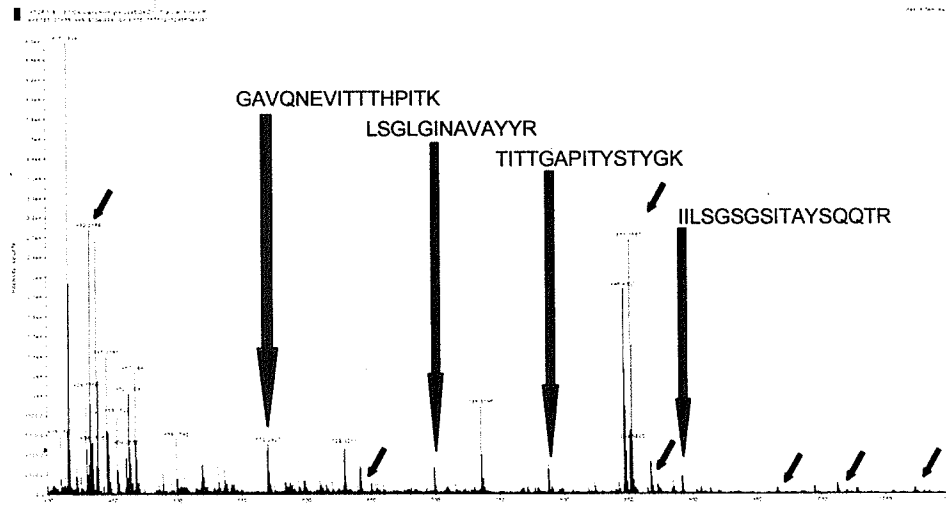
Figure 18: Detection of HIS-NS3/4Apep-1b with anti-HIS Antibodies.

- (A) Purified His-NS3/4Apep-1b was subjected to Western analysis and the blot probed with antiHis6 antiserum. The control is a bacterial extract expressing a control protein (50.8kDa) with a HIS tag identical to that found on NS3/4Apep-1b attached to a portion of the HCV1a NS5A/NS5B fusion junction site (Carpenter, unpublished).

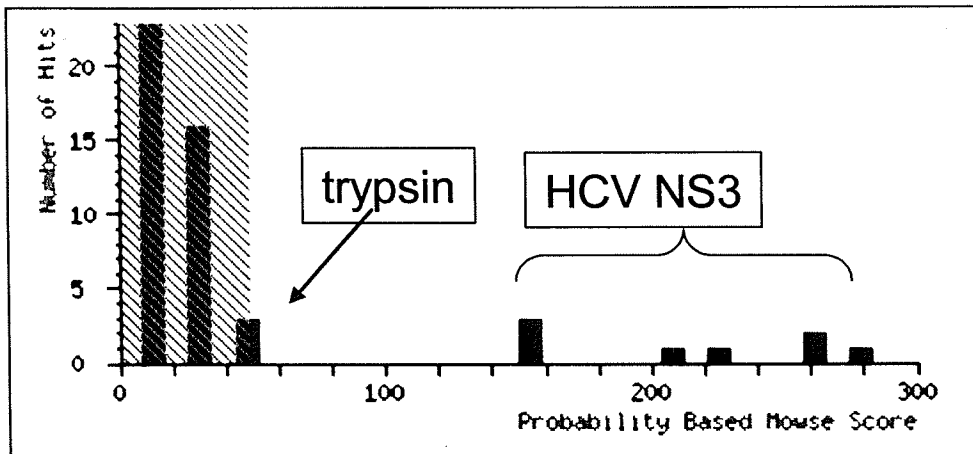
Figure 19: Identification of HIS-NS3/4Apep-1b by mass spectrometry.

- (A) HIS-NS3/4Apep-1b protein was trypsin digested and subjected to discrete nanospray injection into a Q-TOF mass spectrometer. The MS spectrum is shown with X-axis representing mass/charge (m/z) and the y-axis representing intensity. Arrows indicate ion species predicted from a theoretical digest of HIS-NS3/4Apep-1b.
- (B) Information dependant acquisition of the sample used in part (A) was performed in order to gain sequence data via MS/MS fragmentation analysis. Data was collected and evaluated by Mascot. X-axis represents Mowse score, y-axis represents number of peptide sequences matched to a given protein. Mascot graph indicates scores and proteins identified with confidence based on a probability Mowse score (higher Mowse scores represent more statistically significant identifications).
- (C) Amino acid sequence of HIS-NS3/4Apep-1b detailing the amino acid sections identified during mass spectrum analysis (). Peptides for which partial sequence was obtained are in green underlined.

A)



B)



C)

NS3/4Apep sequence:

MGSSHHHHHSSGLVPRGSHMGSVV
 IVGRILSGSITAYSQOTRGLLGCIITSL
 TGRDRNQVEGEVQVVSTATQSFLATC
 VNGVCWTVYHGAGSKTLGPKGPIT
 QMYTNVDQDLVGWQAPPGARSLTPC
 TCGSSDLYLVTRHADVIPVRRRGDSR
 GSLLSPRPVSYLKGSSGGPLLCPGSHA
 VGIFRAAVCTRGVAKAVDFVPVESME
 TTMRSVFTDNSSPPAVPQTFQVAHL
 HAPTGS GKSTKVPAAYAAQGYKVLV
 LNPSVAATLGF GAYMSKAHGIDPNIR
 TGVRTITGAPITYSTYKFLADGGCS
 GGAYDIHCDECHSTDSTILGIGTVLD
 QAETAGARLVVLATATPPGSVTVPHF
 NIEVALSSTGEIPFYGKAIPETIKGG
 RHLIFCHSKKKCEDELA AKLSGLGLNA
 VAYYRGLDVSVIPTSGDVIVVATDAL
 MTGFTGDFDSVIDCNTCVTQTVD FSL
 DPTFTIETTTVPQDAVSR SRQRRTG
 RGRMGIYRFVTPGERPSGMFDSSVLC
 ECYDAGCAWYELTPAETSVRLRAYL
 NTPGLPVCQDHLEFWESVFTGLTHID
 AHFLSQT KQAGDNFPYLVA YQATVC
 ARAQAPP SWDOMWKCLIRLKPTLH
 GPTPLLYRIGAVQNEVTTTHPITKYIM
 ACMSADLEVVT

dependent fragmentation resulting in the partial sequence for 4 peptides (Fig 19A, sequences above arrows). When these spectra were submitted to the MASCOT search engine (www.matrixscience.com), an unequivocal identification to HCV-NS3 was made. Mowse scores greater than 45 are considered significant ($p < 0.05$). Here only peptides corresponding to NS3 (or the digestion agent trypsin) were identified with high probability based on MOWSE scores. Fig. 19C shows the peptides detected by MS nanospray (underlined) with the peptides determined by tandem MS analysis for which partial sequence was obtained (lowercase underlined). With soluble full-length NS3 in hand, we next wished to look at the proteins identified in the Y2H screen and evaluate the putative interactions in an in vitro NS3 system.

3.0. Co-precipitation (HIS pull-down) experiments

It was decided that targets identified by the Y2H screen would be translated in rabbit reticulate lysates in the presence of ^{35}S -methionine to label the product. This strategy was chosen since antibodies to many of the prey targets are unavailable and the radioactivity system allows simultaneous testing of many samples and is thus amenable to high-throughput analysis. All co-precipitation experiments were done in duplicate. To test this system, 50 μL reticulate lysates expressing the ^{35}S labeled prey protein (NS3, NS3p, NS3H, NS5A and Grb2) were divided into two tubes. Purified HIS-NS3/4Apep-1b was added to one of the tubes and, following a two hour incubation at 4°C, Ni-beads were added to each tube. All of the reticulate lysate samples expressed their prey protein of interest (Figure 20A, lanes 1-5)). HIS-NS3/4Apep-1b was found to co-precipitate NS3-1a (Fig 20B, lane 6) and NS3p 1a (lane 8). NS3/4A-1b did not co-precipitate NS3H, NS5A or Grb2 (Fig. 20B lanes 10, 11, 13). Importantly, without the addition of HIS-NS3/4Apep-1b, no labeled proteins were precipitated (lanes 7, 9, 12, 14). This demonstrated that NS3/4Afl 1b can co-precipitate a binding partner seen in the Y2H system (NS3p) but does not co-precipitate non-specific proteins (Grb2). Interestingly, it also demonstrated that full-length NS3 can interact with itself in a co-precipitation experiment, a result that differs from that seen in the Y2H analysis.

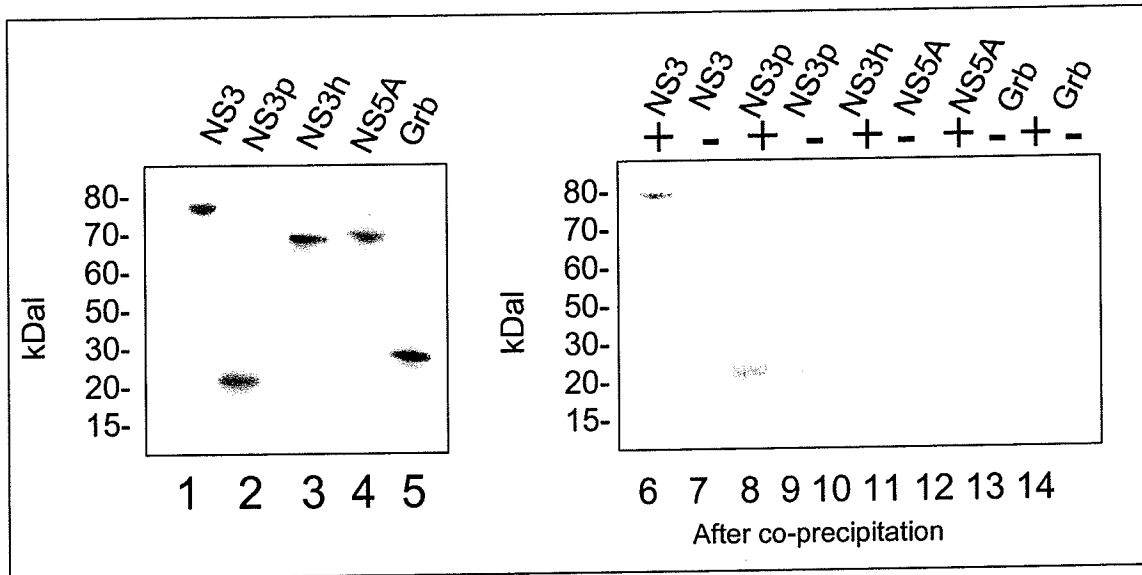


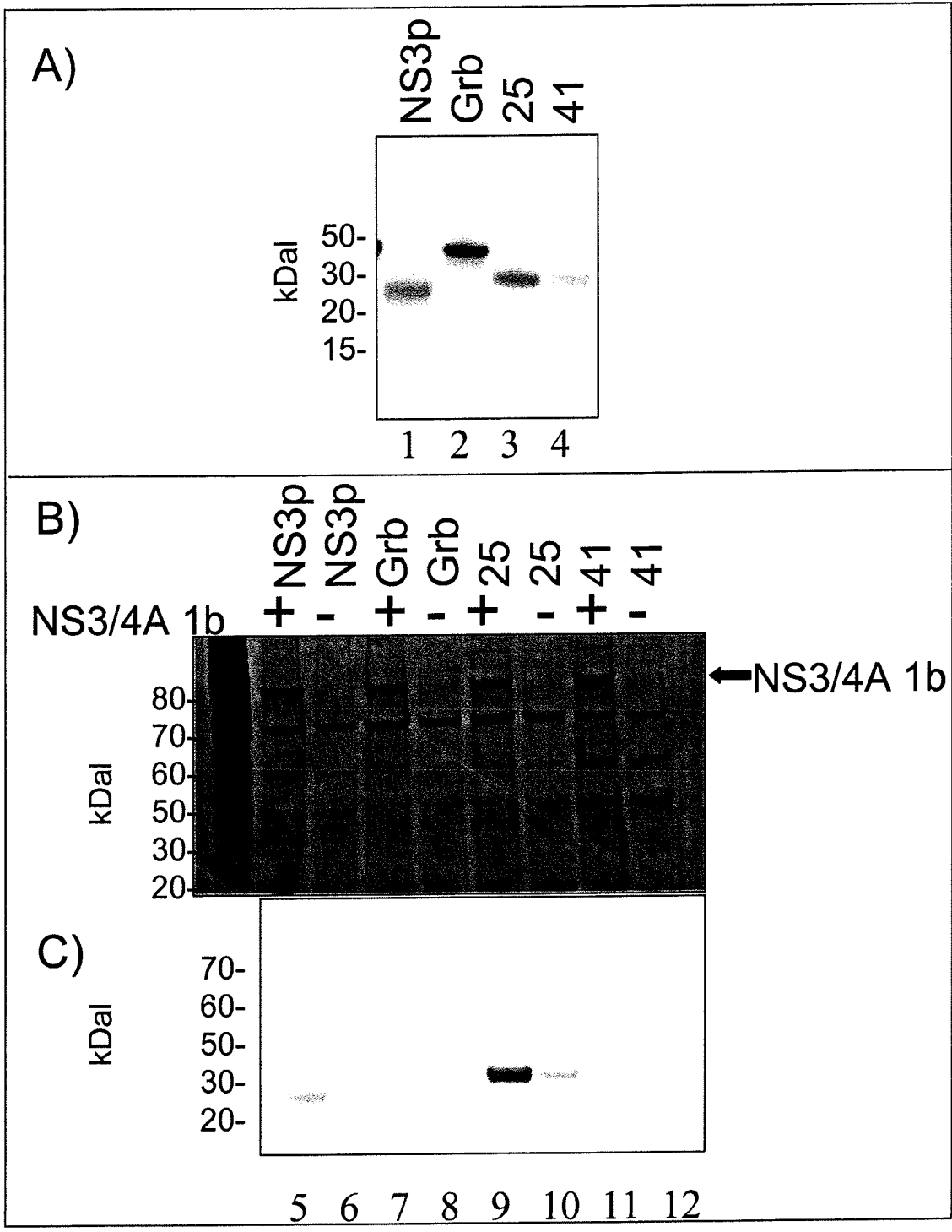
Figure 20: Ability of purified HIS-NS3/4Apep-1b to bind radiolabeled prey proteins

- (A) In vitro translated ³⁵S methionine labeled proteins were subjected to SDS-PAGE and autoradiography. Lane: (1) NS3 1a, (2) NS3 protease domain (NS3p), (3) NS3 helicase domain (NS3H), (4) NS5A, (5) Grb2.
- (B) Autoradiograph detailing the interaction of NS3/4Apep-1b with various target proteins. Co-precipitations included the indicated labeled protein. Lanes: (6, 7) NS3, (8, 9) NS3P, (10, 11) NS3H, (12, 13) NS5A, (14, 15) Grb2. Each co-precipitation contains HIS-NS3/4Apep-1b (+) or no protein added (-).

Finally, 2 prey protein identified in the Y2H system (colony 25=haptoglobin, colony 41=cyclin G1 binding protein) were expressed in rabbit reticulate lysate and labeled with ^{35}S methionine (Fig. 21A). Along with control samples (NS3p and Grb2), samples were subjected to co-precipitation experiments as described above then subjected to SDS-PAGE and Coomassie staining (Fig 21B). Here the NS3/4Apep-1b protein can be clearly visualized in the appropriate lanes (+). The gels were then subjected to autoradiography (Fig 21C). Here, sample 25 (haptoglobin) co-precipitated strongly with NS3/4Apep 1b whereas only a small amount was present in the lanes where the NS3 protein was absent (Figure 21 lane 10). The large difference in intensity strongly suggests that the interaction is specific for the NS3 protein. For sample 41 (cyclin G1 binding protein), no co-precipitation was detected (Figure 21C, lanes 11 and 12).

Figure 21: Co-precipitation of ^{35}S methionine labeled prey proteins with HIS-NS3/4Apep-1b

- A) Expression of ^{35}S methionine labeled constructs in rabbit reticulate lysates. The positive control (NS3p), negative control (Grb2) and two experimental prey (25 and 41) were selected for study
- B) Co-precipitation experiments were performed and run on a SDS PAGE polyacrylamide gel and stained with Coomassie blue. Lanes are evenly loaded and the presence of NS3/4Apep 1b is visible in lanes 1, 3, 5, and 7.
- C) Western transfer probed with anti-HIS anti-sera demonstrates that the positive control, NS3P (lane 5) and colony number 25 (lane 9) successful co-precipitated with NS3/4Apep 1b.



IV) Discussion

The hepatitis C virus genome encodes 10-11 proteins and as a consequence of this limited coding capacity, each of the viral proteins tends to serve multiple functions during the viral life cycle. In many cases, a physical interaction has been identified such as between NS3 and p53 (Ishito et al, 1998) or LMP7 (Khu et al.,2004). In other cases, a functional consequence such as cellular transformation (Zemel et al., 2001) has been determined. The viral nonstructural protein 3 (NS3) is a particularly interesting target due to known and perceived roles in immune modulation and cellular transformation. In order to gain further insight into functions that NS3 performs, protein-protein interaction screen was performed using the yeast two hybrid system. This system has several advantages in that even weak or rare interactions can be discerned but there are also a number of drawbacks to this technique, the most important one being the large number of false positives which can be seen.

1.0. Yeast-2-Hybrid Screen

1.1 Yeast-2-Hybrid Screen pulls out many non-relevant interactions

For most published Y2H screens, only a few clones are studied. This study represents the first comprehensive study for which all prey sequences were determined for a single Y2H screen. This allows for a better understanding of the Y2H as a tool in a way which is not normally done due to relatively laborious nature of looking at individual prey targets. Therefore, a relatively high throughput method was developed in which sequence information from 98% of the prey targets was obtained using a combination of direct lyticase treatment of yeast colonies combined with direct PCR-sequencing. Analysis revealed that besides in-frame protein targets, there were several different categories that could be termed “inappropriate binding”. These include proteins that are out of frame with an authentic protein sequence, un-clarified open reading frames (ORF), or fusion proteins and will be discussed in greater detail below.

Gene sequences with frame shift mutations produce non-natural proteins. The fact that these are being pulled out of an Y2H screen occurs because they produce “sticky proteins” or proteins which alleviate the auxotrophic pressure on the yeast cells. How proteins perform these two functions will be addressed later in regards to other non-specific interacting proteins. Theoretically, 2/3 of all prey should be out-of-frame due to the method that a library is generated. However, our data found that 33.9% of prey were out-of-frame (Table 1). This discrepancy is most likely because most frame shifts would generate short reading frames due to premature stop codons. In addition, out-of-frame sequences do not represent natural proteins in the host cell proteome and therefore viral proteins would not evolve to interact with them. Considering the likelihood of these out-of-frame sequences interacting with DBD-NS3 or of their ability to alleviate auxotrophic pressure, it is surprising that so many were found in this screen. Of the 125 out-of-frame sequences, 52 were out-of-frame sequences of metallothioneine (Appendix 1) potentially skewing the number of out-of-frame sequences that would typically be seen in a yeast-2-hybrid screen. Although these proteins are not natural, their interaction with NS3 may be authentic and thus future work could examine these proteins for motifs or other explanations as to why they were positive in the screen.

There were 24 fusion proteins present in the Y2H screen (Table 1). These samples occurred when two different mRNA species, or cDNA species, were spliced together during the generation of the liver library. Although rare, these proteins are also of interest as, despite being formed from two proteins, they still form a protein with the capacity to positively interact with NS3. These cases also require further analysis to determine whether either protein on its own has a natural interaction with NS3, or if the interaction is due to the formation of a hybrid protein.

There were a few cases where the prey which interacted with DBD-NS3 was found to be from an un-clarified open reading frame. Most of these prey were originally sequenced from large scale Bacmid sequencing projects, but these regions have not yet been fully annotated. The completion and clarification of the human genome annotations will determine if these are worthwhile targets. Alternatively, more advanced bioinformatic analysis may provide additional direction.

1.2. Confirmation of Y2H protein-protein interactions

Before the yeast-2-hybrid experiment could be carried out, the ability of DBD-NS3 to interact with other proteins had to be evaluated. DBD-NS3 was transformed into yeast cells containing a variety of AD proteins, including NS3fl, NS3P, NS3H, NS5A and Grb2. NS3 has previously been found to interact with NS5A by GST pull-down and co-immunoprecipitation (Dimitrova *et al* 2003). In this Y2H screen, NS3 did interact with NS5A (Figure 5-2). Studies have also suggested that NS3 interacts as a dimer through its helicase domain (Khu *et al* 2001). In this Y2H screen, NS3 was not found to interact with NS3 (Figure 5-3), and the helicase domain did not interact with itself (Figure 5-7). The difference in results regarding the ability for NS3H to form homodimers may be strain related, as this work used strain 1a and previous studies used strain 1b (Khu *et al* 2001), as both experiments used Y2H screens as a method to test protein interactions. In the Y2H portion of this study, novel interactions were observed between NS3P and NS3. In addition, interactions were seen between NS3P and NS3P, NS3H and NS5A (Figure 5) and suggest that NS3 may interact as a homodimer through its protease domain. These results were partially confirmed in the co-precipitation studies as NS3/4Apep 1b co-precipitated both NS3 and NS3p. The idea that the interaction seen between the protease domains of NS3 is strain related seems unlikely as HIS-NS3/4Apep-1b was able to co-precipitate NS3-1a (Figure 20 lane 6). This further supports the suggestion that the homodimerization of NS3 is due to an interaction in the protease domain, not the helicase domain.

The remaining Y2H sequences which were shown to be in-frame and map to a human open reading frame were further examined to determine if the interactions were specific for the presence of NS3 in the DBD fusion. Confirmation of specificity is important since it is not possible to differentiate between an interaction which is “real” or which is due to a protein with a high propensity to bind any protein, a so-called “sticky” protein. Such interactions could be due to non-specific hydrophobic regions or highly charged regions. It is also not impossible that prey sequences uncovered in this screen are also binding the fusion partner (DBD) rather than NS3. After prioritizing the list of 78

prey, 23 proteins were selected for further examination and tested for their ability to interact specifically with DBD-NS3, versus DBD-NS5A or DBD-lamin.

Eight proteins were found to interact with both NS3 and lamin. Six of these proteins interacted with NS3, NS5A and lamin. It is unlikely that these proteins naturally interact with these three diverse proteins, and therefore likely represent either an interaction with the DBD or production of a protein which alleviates auxotrophic pressure on the cells. These proteins included apolipoprotein H, beta ureidopropionase, corticotropin releasing hormone binding protein, elongation factor 1-alpha, flotillin, and inhibin beta E. These proteins were not further examined in this study for interactions with NS3, although future work could examine them for common binding motifs that will allow rapid identification of likely false positives in subsequent screens. In addition, there were two proteins which were found to interact with both NS3 and lamin, but not with NS5A. These proteins may still represent "sticky" proteins which interact with both NS3 and lamin, but most likely do not interact with the DBD or alleviate auxotrophic pressure as they do not interact with NS5A. The fact that they do interact with proteins as different in function and sequence as NS3 and lamin, however, suggests that the interaction is not likely biologically relevant. It is worth noting that it is virtually impossible to find a true negative control in an Y2H screen, as any protein that is used in the screen has a possibility of interacting with both the protein of interest and the negative control. Even when used from different species, the screen does not assess biological relevance, instead it identifies physical interaction only. This is one of the challenges of working with Y2H screening as a technology.

Some of these sticky protein interactions may have occurred as a result of the cDNA library that was used in the Y2H screen, i.e. a liver library. Liver cells, due to their function in the body, express higher than normal levels of secreted proteins. Secreted proteins, such as glycosylated proteins, are often highly charged and can have a high propensity to interact with other proteins. These could explain NS3 interacting with alpha 2-HS glycoprotein which is a cell membrane protein and would rarely come in contact with NS3 normally. NS3 also has a large hydrophobic region at the N terminus of the protein. These regions could interact with any protein with a corresponding

hydrophobic patch. Further work could examine the “sticky” and non-specific binding interactions with NS3 for possible hydrophobic spans.

Nine prey samples interacted with both NS3 and NS5A, but not with lamin including aldehyde dehydrogenase 2, aldolase B, catalase, cyclin G1 binding protein, insuline-like growth factor, inter-alpha trypsin inhibitor, NADH ubiquitinone oxidoreductase, plasminogen, and vitronectin. As NS3 and NS5A are thought to be localized to the same region of the cell (membranous webs) and there is previous work demonstrating viral proteins interacting with the same host cell proteins (NS5A and NS5B both interact with hVap-A and -B) (Gao *et al* 2004, Hammamoto *et al* 2005) there is a possibility that these interactions are real. Although still of interest, these proteins are less likely to be truly specific. Cyclin G1 binding protein was selected for further studies using co-precipitation and did not precipitate with NS3/4Apep 1b. It is unclear at present whether this inability to co-precipitate is due to the transition of working with NS3 strain 1a to NS3 strain 1b, or if the interaction is functional only in the Y2H system.

Finally, five proteins were found to specifically interact with NS3 in the Y2H including adenosine kinase, cox 2, elongation factor ef2, haptoglobin and transthyretin. These proteins are most likely to have a specific interaction with NS3 and therefore have the most potential to be biologically relevant. NS3 interacted with two transport proteins which are localized extracellularly, Haptoglobin and transthyretin. Although NS3 is not found extracellularly, there is a probability that NS3 could interact with this protein before they are released. Haptoglobin binds free plasma hemoglobin to prevent iron loss and is also thought to be involved in immunomodulation as it interacts with CD1, integrin beta 2 and CD 163 (El-Ghmati *et al* 2002). Interaction with this protein may allow NS3 to modulate immune response. Transthyretin is involved in transporting the thyroid hormone thyroxine (T4). Thyroxine functions to increase basal metabolic rate, and protein synthesis. NS3 binding this protein could be involved in HCV hijacking of the host cell machinery.

Adenosine Kinase is a nucleoside kinase responsible for phosphorylating nucleoside analogues. It is primarily found in the cytoplasm, and therefore could interact with NS3 during nucleic acid metabolism. NS3 has previously been found to interact with kinases, such as human protein kinase A (PKA) (Aoubala *et al* 2001, Borowski *et al*

1999-3), and NS3 serves as a substrate for phosphorylation by the human protein kinase C (PKC) (Borowski *et al* 1999-1).

Cox2, or Cytochrome C Oxidase subunit IV isoform 2, is an oxidase found in the mitochondrial inner membrane and is the terminal enzyme in the electron transport chain (Huttemann *et al* 2001). NS3 has been localized to the mitochondria in chronic hepatitis C patients (Kasprzak *et al* 2005), and in HCV replicon-expressing Huh7 cells (Normura-Takigawa *et al* 2006). Cells expressing NS3/4A have been shown to undergo actinomycin D-induced mitochondria-mediated apoptosis (Normura-Takigawa *et al* 2006). This may be another method whereby NS3 induces apoptosis in human cell culture, as it is also associated with caspase-8 (Prihod'ko *et al* 2004). Interaction with this protein indicates NS3 may play a larger role in inducing hepatocyte cell death in chronic HCV patients, exacerbating liver cirrhosis.

Finally, elongation factor 2 is a translation regulatory protein localized in the cytoplasm. NS3 has been found to induce apoptosis in human cell culture, and this is one potential method by which apoptosis could be induced.

2.0 Purification of NS3/4A

To further evaluate the hypothetical interactions between NS3 and other proteins, co-precipitation was performed. This required a soluble, purified version of NS3. The main issues involved in its purification including translation levels, degradation, and solubility.

NS3 was produced at low levels in bacteria due to its toxicity to bacteria, or codon bias in the coding region. Toxicity, as defined by decreased growth of bacteria, was evidenced by the growth curve of NS3 (Figure 9C). However, somewhat higher levels of protein production could be seen when a truncated form of NSP (NS3P 1a) was used (Figure 10). Unfortunately, the higher levels of expression seen when expressing NS3P 1a were accompanied by the formation of truncation products, either due to incomplete translation, or degradation (Figure 10).

Degradation can occur when bacterial proteins are misfolded. Incomplete translation could be due to the difference in codon bias seen between human cells and

bacterial cells. Both issues were addressed by expressing NS3P 1a in bacterial background expressing either chaperone proteins, or plasmids that adjusted the anticodon tRNA levels, accounting for the different codon bias in a human gene expressed in a bacterial environment (Figure 11). In all cases, the expression of NS3P 1a still produced large quantities of non-soluble proteins (data not shown) and truncated protein byproducts, even after nickel chelated column purification (Figure 12). To attempt to improve both solubility and decrease truncated byproducts, an even shorter NS3 clone, NS3ps 1a, was created. The further truncation of NS3 decreased its use in screening of potential interacting partners, as it would be limited to interactions with the protease domain, but truncation was the only viable method to obtain the minimal expression required. Unfortunately, further truncation of NS3 still produced non-soluble protein with truncated byproducts (Figure 13).

The main alteration which aided in decreasing degradation was the use of the 1b strain instead of the 1a strain. Although initial Y2H screens used the 1a strain, this strain proved difficult to express and purify as a recombinant protein in bacteria. Switching to the 1b strain decreased the degradation of NS3 dramatically (Figure 13 and Figure 14). The difference in degradation or incomplete translation between the species is surprising considering the minimal differences between the strains at the amino acid level (Figure 14). Nonetheless the expression of NS3ps1b was far more stable than strain 1a.

Although NS3ps 1b produces a protein that did not produce as many degradation products, purification was still problematic as almost half of the produced protein was insoluble (Figure 14). There were several methods utilized to help with the solubility of NS3. The main method which aided in solubility was the addition of a portion of the NS4A cofactor to the front end of the protein being expressed (Figure 16). Previous research has demonstrated the success in solubilizing and expressing NS3 with the addition of a NS4A peptide linked to the N terminal portion of NS3 (Taremi *et al* 1998). It is possible that the addition of the 4A peptide blocked the hydrophobic patch at the N-terminus of NS3 thus likely decreasing aggregation of the protein. It is also possible that the interaction with 4A altered the conformation of NS3 increasing its solubility.

Despite not being visible on a Coomassie blue stained gel (Figure 17A), NS3/4Apep-1b was purified at sufficient levels with high purity using a two-step

purification on a nickel chelate column followed by a 50mL gel filtration column. NS3/4A eluted over several fractions off the nickel chelate column. This could demonstrate a weak binding affinity of NS3/4A1b to the column, or multiple binding affinities (perhaps due to dimerization of NS3). It is also possible that the HIS antigen of the NS3/4A-His protein might be slightly occluded. The second step in purification was a 50mL HR-200 sephadex gel filtration column. On this column, the protein eluted off over two peaks. These two elution peaks could be two different forms of the protein eluting, such as a dimer formation. Analytical gel filtration column chromatography will be required to determine if this is indeed the case.

Curiously, even though NS3/4A was expressed with a HIS tag which bound to a nickel chelate column, it could not be detected when analyzed by Western blot using His anti-sera for detection. This was not due to a problem with the HIS tag or with the anti-sera used since the His antibody worked when run against an in-house control (Figure 18). It is possible that the His tag, which is bound on the amino terminus of the protein, gets folded inside the NS3 protein, and despite denaturing conditions, is occluded when run on SDS-PAGE acrylamide gels and transferred to a membrane. This, however, further complicated analysis as there was no definitive method to identify the purified protein as neither the HIS-tag antibody nor the antibodies against NS3 in our collection could be used for identification. We therefore used mass spectrometry to confirm the identity of the protein (Figure 19).

3.0. Co-Precipitation of NS3 with Prey Protein

Prey were successfully amplified and labeled with ^{35}S methionine. Both of the positive controls, NS3fl 1a and NS3p 1a, were pulled down by NS3/4Apep 1b (Figure 20). The negative control (Grb2), however, was not pulled down in the presence of NS3/4Apep 1b. Interestingly, NS5A 1a was not pulled down by the protein (Figure 20). This is curious as the NS3 strain 1a was previously shown to interact physically with NS5A strain 1a in a yeast-2-hybrid screen and co-precipitation (Dimitrova *et al* 2003). It is possible that the strain specificity allows for NS5A 1a to interact with NS3 strain 1a but not 1b. The difference in binding may also be accounted for by the presence of the

NS4A peptide which is absent in the Y2H screen but present in the co-precipitation experiments which may alter the binding capabilities of the protein.

Two sample proteins, haptoglobin (25) and cyclin G1 binding protein (41), were expressed in reticulocyte lysates (Figure 21). According to the earlier Y2H screen, haptoglobin was known to interact only with NS3, whereas cyclin G1 binding protein interacted with both NS3 and NS5A. Pull-down studies found that cyclin G1 binding protein was not pulled down by NS3/4A 1b. This suggests that either NS3/4Apep-1b does not truly interact with cyclin G1 binding protein, as it may not be a biologically relevant interaction, or it is an interacting partner of NS3 strain 1a, and did not pull down in our hands as the pull-down experiment was performed using an NS3 strain 1b protein. Haptoglobin, however, did show increased interaction in the presence of NS3/4Apep 1b in comparison to a negative control thus further supporting a real interaction between NS3 and haptoglobin.

Haptoglobin (Hp) is a protein found in blood plasma which binds free hemoglobin and is released from erythrocytes upon cell death. In its simplest form, Hp consists of two alpha helices and two beta sheets connected by disulfide bridges (Langois *et al* 1996). The binding observed between NS3 and haptoglobin, as evidenced by both an Y2H screen and co-precipitation, could be the result of protein interactions between cysteines. Haptoglobin is primarily produced in hepatocytes (Langois *et al* 1996), allowing HCV to come into contact with the protein before it leaves the cell. Hp decreases the loss of iron and hemoglobin from the body and decreased levels of haptoglobin are associated with hemolytic anemia. Chronic hepatitis C infection is often associated with anemia, although this is usually considered a side effect of ribavirin treatment (Borgia *et al* 2003). However, hemolytic anemia has been reported in a treatment-naïve patient (Srinivasan *et al* 2001) as diagnosed by reduced hemoglobin in the absence of blood loss, elevated serum lactate dehydrogenase, and low haptoglobin. It is tempting to speculate that the anemia could be the result of the NS3 protein binding to haptoglobin within the cell thus reducing the amount of free hp available to be secreted. Furthermore, haptoglobin polymorphism has been linked to HCV infection (Louagie *et al* 1996) suggesting a correlation between the two proteins. HCV infected patients were observed to have an increased phenotype of Hp 1-1, and a decrease in Hp 2-2 and there was no

association between Hp phenotype and response to therapy or HCV strain (Louagie *et al* 1996). Similarly, a large scale proteomics project found that haptoglobin levels were significantly changed in Hepatitis B virus-infected sera (He *et al* 2003) suggesting a greater role of haptoglobin in liver infections. This might be due to haptoglobin's association with the immune response. The carbohydrate moiety of Hp has been found to bind with CD22 (Hanasaki *et al* 1995), which is involved in communication between B lymphocytes and other immune cells. Haptoglobin has also been shown to have an inhibitory effect on prostoglandin synthesis, a protein with known involvement in anti-inflammatory action (Lange *et al* 1992). If NS3 bound to and reduced haptoglobin, it would follow that there would be an increase in prostaglandin synthesis, reducing activation of T cells and inflammation. Furthermore, haptoglobin may play a role in regulating oxidative stress, another function HCV is known to have a role in. Free hemoglobin promotes the generation of hydroxyl radicals as a fenton reagent (Sadrzadeh *et al* 1984). This is mitigated by the binding of hemoglobin to haptoglobin (Javid *et al* 1965). If haptoglobin was bound by NS3, free hemoglobin could promote the increased oxidative stress seen during HCV infection. The ability of NS3 to bind to and sequester haptoglobin, and the role haptoglobin plays in host immune response and oxidative stress could help to explain the increased oxidative stress and immune modulation seen during an HCV infection.

4.0 Future work

Future work could investigate the ability of all the 78 unique in-frame yeast to interact with NS3, NS5A and lamin. Furthermore, the non-specific interacting proteins could be studied for the presence of common motifs which cause non-relevant interactions with target proteins. Interactions specific to NS3 can be determined using co-precipitation experiments and subsequently analyzed in several ways. First, because the fusion proteins identified in the Y2H are not typically complete open reading frames, the ability for the full-length prey protein to interact with NS3 must be ascertained. If an interaction occurs between the two full-length proteins, then the interaction must be evaluated *ex vivo*. The easiest way to evaluate this is to express NS3/4A and the prey of

interest in a human cell background. This would test the proteins ability to interact in a human cell system. If the interaction is still observed, then immunofluorescence studies should be performed to demonstrate that the proteins are expressed in similar areas in the cell. Finally, experimentation attempting to evaluate the functional role that this interaction plays in infection, replication and disease progression.

5.0. Summary

NS3 is a complicated protein, which plays a vital role in viral replication and disease progression. It has been found to be involved in numerous cellular pathways including immune regulation and oxidative stress. The protein interaction screen performed here has identified a number of proteins which appear to specifically and physically interact with NS3. The ability for NS3 to bind with these targets should further our knowledge of how HCV causes disease and possibly allow better predictors of treatment response. By better understanding the interactions NS3 has with the host cell proteome, we can elucidate its role in viral replication and disease progression, allowing more effective treatment alternatives to alleviate or treat this disease.

V) References

- Alter, H. J., Purcell R.H., Shih J.W., Melpolder J.C., Houghton M., Choo Q.L., Kuo G. (1989). Detection of antibody to hepatitis C virus in prospectively followed transfusions with acute and chronic non-A, non-N hepatitis. *NEJM* **321**(22), 1494-500.
- Aoubala, M., Holt, J., Clegg, R. A., Rowlands, D. J., and Harris, M. (2001). The inhibition of cAMP-dependent protein kinase by full-length hepatitis C virus NS3/4A complex is due to ATP hydrolysis. *J Gen Virol* **82**(Pt 7), 1637-46.
- Arias, C. F., F. Preugschat, J.H. Strauss (1993). Dengue 2 virus NS2B and NS3 form a stable complex that can cleave within the helicase domain. *Virology* **193**(2), 888-99.
- Ausubel, F.M. (1998) Current Protocols in Molecular Biology. John Wiley and Sons, Inc, NY.
- Bahnassi, A. A., Zekri, A. R., El-Houssini, S., Mokhtar, N. M., Abdel-Aziz, A. O., Sherif, G. M., El-Mishad, A. M., and Khaled, H. M. (2005). Hepatitis C virus-NS3P in relation to p53, p21waf, mdm2, p21-ras and c-erbB2 in hepatocarcinogenesis. *J Gastroenterol Hepatol* **20**(11), 1731-40.
- Bartenschlager, R. (1999). The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy. *J Viral Hepat* **6**(3), 165-81.
- Bartenschlager, R., and Pietschmann, T. (2005). Efficient hepatitis C virus cell culture system: what a difference the host cell makes. *Proc Natl Acad Sci U S A* **102**(28), 9739-40.
- Bartenschlager, R., L. Ahlborn-Laake, K. Yasargil, J. Mous, and H. Jacobsen (1994). Kinetic and structural analysis of hepatitis C virus polyprotein processing. *J Virol* **68**, 5045-55.
- Bartenschlager, R., V. Lohmann, T. Willkinson, and J.O. Koch (1995). Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *J Virol* **69**, 7519-7528.
- Bigger, C. B., K.M. Brasky, R.E. Lanford (2001). DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J Virol* **75**(15), 7059-66.
- Borgia, G., Reynaud, L., Gentile, I., Borrelli, F., Cerini, R., Ciampi, R., and Piazza, M. (2003). Pernicious anemia during IFN-alpha treatment for chronic hepatitis C. *J Interferon Cytokine Res* **23**(1), 11-2.
- Borowski, P., Heiland, M., Feucht, H., and Laufs, R. (1999). Characterisation of non-structural protein 3 of hepatitis C virus as modulator of protein phosphorylation mediated by PKA and PKC: evidences for action on the level of substrate and enzyme. *Arch Virol* **144**(4), 687-701.
- Borowski, P., Kuhl, R., Laufs, R., Schulze zur Wiesch, J., and Heiland, M. (1999). Identification and characterization of a histone binding site of the non-structural protein 3 of hepatitis C virus. *J Clin Virol* **13**(1-2), 61-9.
- Borowski, P., Schulze zur Wiesch, J., Resch, K., Feucht, H., Laufs, R., and Schmitz, H. (1999). Protein kinase C recognizes the protein kinase A-binding motif of nonstructural protein 3 of hepatitis C virus. *J Biol Chem* **274**(43), 30722-8.

- Bureau, C., Bernad, J., Chaouche, N., Orfila, C., Beraud, M., Gonindard, C., Alric, L., Vinel, J. P., and Pipy, B. (2001). Nonstructural 3 protein of hepatitis C virus triggers an oxidative burst in human monocytes via activation of NADPH oxidase. *J Biol Chem* **276**(25), 23077-83.
- Carrere-Kremer, S., Montpellier-Pala, C., Cocquerel, L., Wychowski, C., Penin, F., and Dubuisson, J. (2002). Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. *J Virol* **76**(8), 3720-30.
- Cerny, A., J.G. McHutchison, C. Pasquinelli, M.E. Brown, M.A. Brothers, B. Grabscheid, P. Fowler, M. Houghton, F.V. Chisari (1995). Cytotoxic T lymphocyte response to hepatitis C virus-derived peptides containing the HLA A2.1 binding motif. *J Clin Invest* **95**(2), 521-30.
- Cheng, P. L., Chang, M. H., Chao, C. H., and Lee, Y. H. (2004). Hepatitis C viral proteins interact with Smad3 and differentially regulate TGF-beta/Smad3-mediated transcriptional activation. *Oncogene* **23**(47), 7821-38.
- Choi, J., Lee, K. J., Zheng, Y., Yamaga, A. K., Lai, M. M., and Ou, J. H. (2004). Reactive oxygen species suppress hepatitis C virus RNA replication in human hepatoma cells. *Hepatology* **39**(1), 81-9.
- Choo, Q. L., G. Kuo, A.J. Weiner, L.R. Overby, D.W. Bradley, and M. Houghton (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**, 359-362.
- Choo, Q. L., Weiner, A. J., Overby, L. R., Kuo, G., Houghton, M., and Bradley, D. W. (1990). Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. *Br Med Bull* **46**(2), 423-41.
- De Francesco, R., A. Urbani, M. Nardi, L. Tomei, C. Steinkuhler, and A. Tramontano (1996). A zinc binding site in viral serine proteinases. *Biochemistry* **35**, 13282-87.
- De Francesco, R., and Steinkuhler, C. (2000). Structure and function of the hepatitis C virus NS3-NS4A serine proteinase. *Curr Top Microbiol Immunol* **242**, 149-69.
- Deng, L., M. Nagano-Fujii, M. Tanaka, Y. Nomura-Takigawa, M. Ikeda, N. Kato, K. Sada, H. Hotta (2006). NS3 protein of Hepatitis C virus associates with the tumour suppressor p53 and inhibits its function in an NS3 sequence-dependant manner. *J Gen Virol* **87**(6), 1703-13.
- Diepolder, H. M., Zachoval, R., Hoffmann, R. M., Wierenga, E. A., Santantonio, T., Jung, M. C., Eichenlaub, D., and Pape, G. R. (1995). Possible mechanism involving T-lymphocyte response to non-structural protein 3 in viral clearance in acute hepatitis C virus infection. *Lancet* **346**(8981), 1006-7.
- Dimitrova, M., Imbert, I., Kieny, M. P., and Schuster, C. (2003). Protein-protein interactions between hepatitis C virus nonstructural proteins. *J Virol* **77**(9), 5401-14.
- Dolganiuc, A., Kodys, K., Kopasz, A., Marshall, C., Do, T., Romics, L., Jr., Mandrekar, P., Zapp, M., and Szabo, G. (2003). Hepatitis C virus core and nonstructural protein 3 proteins induce pro- and anti-inflammatory cytokines and inhibit dendritic cell differentiation. *J Immunol* **170**(11), 5615-24.
- El Ghmati, S. M., E.M. Van Hoeyveld, J. A. G. Van Strijp, J. L. Ceuppens, E. A. M. Stevens (1996). Identification of Haptoglobin as an alternative ligand for CD11b/CD18. *Am Association of Immunologists* **156**, 2542-52.

- Erickson, A. L., M. Houghton, Q.L. Choo, A.J. Weiner, R. Ralston, E. Muchmore, C.M. Walker (1993). Hepatitis C virus-specific CTL responses in the liver of chimpanzees with acute and chronic hepatitis. *J Immunol* **151**(8), 4189-99.
- Flajolet, M., Rotondo, G., Daviet, L., Bergametti, F., Inchauspe, G., Tiollais, P., Transy, C., and Legrain, P. (2000). A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* **242**(1-2), 369-79.
- Foy, E., Li, K., Sumpter, R., Jr., Loo, Y. M., Johnson, C. L., Wang, C., Fish, P. M., Yoneyama, M., Fujita, T., Lemon, S. M., and Gale, M., Jr. (2005). Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling. *Proc Natl Acad Sci U S A* **102**(8), 2986-91.
- Foy, E., Li, K., Wang, C., Sumpter, R., Jr., Ikeda, M., Lemon, S. M., and Gale, M., Jr. (2003). Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science* **300**(5622), 1145-8.
- Frick, D. N. (2003). Helicases as antiviral drug targets. *Drug News Perspect* **16**(6), 355-62.
- Fujita, T., Ishido, S., Muramatsu, S., Itoh, M., and Hotta, H. (1996). Suppression of actinomycin D-induced apoptosis by the NS3 protein of hepatitis C virus. *Biochem Biophys Res Commun* **229**(3), 825-31.
- Gao, L., Aizaki, H., He, J. W., and Lai, M. M. (2004). Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J Virol* **78**(7), 3480-8.
- Gietz, R. D., and R.A. Woods (2002). Transformation of yeast by the LiAc/SS carrier DNA/PEG Method. *Methods in Enzymology* **350**, 87-96.
- Grakoui, A., D.W. McCourt, C. Wychowski, S.M Feinstone, and C.M. Rice (1993). Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J Virol* **67**(5), 2832-43.
- Griffin, S. D., Beales, L. P., Clarke, D. S., Worsfold, O., Evans, S. D., Jaeger, J., Harris, M. P., and Rowlands, D. J. (2003). The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Lett* **535**(1-3), 34-8.
- Grobler, J. A., Markel, E. J., Fay, J. F., Graham, D. J., Simcoe, A. L., Ludmerer, S. W., Murray, E. M., Migliaccio, G., and Flores, O. A. (2003). Identification of a key determinant of hepatitis C virus cell culture adaptation in domain II of NS3 helicase. *J Biol Chem* **278**(19), 16741-6.
- Gu, B., Pruss, C. M., Gates, A. T., and Khandekar, S. S. (2005). The RNA-unwinding activity of hepatitis C virus non-structural protein 3 (NS3) is positively modulated by its protease domain. *Protein Pept Lett* **12**(4), 315-21.
- Hamamoto, I., Nishimura, Y., Okamoto, T., Aizaki, H., Liu, M., Mori, Y., Abe, T., Suzuki, T., Lai, M. M., Miyamura, T., Moriishi, K., and Matsuura, Y. (2005). Human VAP-B is involved in hepatitis C virus replication through interaction with NS5A and NS5B. *J Virol* **79**(21), 13473-82.
- Hanasaki, K., L.D. Powell, A. Varki (1995). Binding of human plasma sialoglycoproteins by the B cell-specific lectin CD22. Selective recognition of immunoglobulin M and haptoglobin. *J Biol Chem* **270**(13), 7543-50.
- He, Q. Q., Cheng, R. X., Sun, Y., Feng, D. Y., Chen, Z. C., and Zheng, H. (2003). Hepatocyte transformation and tumor development induced by hepatitis C virus

- NS3 c-terminal deleted protein. *World J Gastroenterol* **9**(3), 474-8.
- He, Q. Y., Lau, G. K., Zhou, Y., Yuen, S. T., Lin, M. C., Kung, H. F., and Chiu, J. F. (2003). Serum biomarkers of hepatitis B virus infected liver inflammation: a proteomic study. *Proteomics* **3**(5), 666-74.
- Hidajat, R., Nagano-Fujii, M., Deng, L., Tanaka, M., Takigawa, Y., Kitazawa, S., and Hotta, H. (2005). Hepatitis C virus NS3 protein interacts with ELKS- $\{\delta\}$ and ELKS- $\{\alpha\}$, members of a novel protein family involved in intracellular transport and secretory pathways. *J Gen Virol* **86**(Pt 8), 2197-208.
- Huttemann, M., B. Kadenbach, L.I. Grossman (2001). Mammalian subunit IV isoforms of cytochrome c oxidase. *Gene* **267**(1), 111-23.
- Ishido, S., and Hotta, H. (1998). Complex formation of the nonstructural protein 3 of hepatitis C virus with the p53 tumor suppressor. *FEBS Lett* **438**(3), 258-62.
- Iwai, A., Hasumura, Y., Nojima, T., and Takegami, T. (2003). Hepatitis C virus nonstructural protein NS3 binds to Sm-D1, a small nuclear ribonucleoprotein associated with autoimmune disease. *Microbiol Immunol* **47**(8), 601-11.
- Javid, J. (1965). The effect of Haptoglobin polymer size on hemoglobin binding capacity. *Vox Sang* **10**, 320-5.
- Jin, L., and D.L. Peterson (1995). Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA helicase. *Arch Biochem Biophys* **323**, 47-53.
- Kadare, G., and A.L. Haenni (1997). Virus-encoded RNA helicases. *J Virol* **71**, 2588-2590.
- Kanai, A., K. Tanabe, and M. Kohara (1995). Poly(U) binding activity of the hepatitis C virus NS3 protein, a putative RNA helicase. *FEBS Lett* **376**, 221-224.
- Kasprzak, A., Seidel, J., Biczysko, W., Wysocki, J., Spachacz, R., and Zabel, M. (2005). Intracellular localization of NS3 and C proteins in chronic hepatitis C. *Liver Int* **25**(4), 896-903.
- Kawai, T., K. Takahashi, S. Sato, C. Coban, H. Kumar, H. Kato, K.J. Ishii, O. Takeuchi, S. Akira (2005). IPS-1, an adaptor triggering RIG-I and Mda5-mediated type I interferon induction. *Nat Immunol* **6**(10), 981-8.
- Khu, Y. L., Koh, E., Lim, S. P., Tan, Y. H., Brenner, S., Lim, S. G., Hong, W. J., and Goh, P. Y. (2001). Mutations that affect dimer formation and helicase activity of the hepatitis C virus helicase. *J Virol* **75**(1), 205-14.
- Khu, Y. L., Tan, Y. J., Lim, S. G., Hong, W., and Goh, P. Y. (2004). Hepatitis C virus non-structural protein NS3 interacts with LMP7, a component of the immunoproteasome, and affects its proteasome activity. *Biochem J* **384**(Pt 2), 401-9.
- Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **87**(2), 343-55.
- Krieger, N., Lohmann, V., and Bartenschlager, R. (2001). Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. *J Virol* **75**(10), 4614-24.
- Kwun, H. J., Jung, E. Y., Ahn, J. Y., Lee, M. N., and Jang, K. L. (2001). p53-dependent transcriptional repression of p21(waf1) by hepatitis C virus NS3. *J Gen Virol*

- 82(Pt 9), 2235-41.
- Lange, V. (1992). Haptoglobin polymorphism - not only a genetic marker. *Anthropol Anz* **50**(4), 281-302.
- Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr., and Lemon, S. M. (2005). Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci U S A* **102**(8), 2992-7.
- Lindenbach, B. D., C.M. Rice (2001). "Flaviviridae: The Viruses and Their Replication." 4 ed. Fields Virology (D. M. Knipe, P.M. Howley, Ed.), 1. 2 vols. Lippincott Williams and Wilkins, Philadelphia.
- Lindenbach, B. D., Meuleman, P., Ploss, A., Vanwolleghem, T., Syder, A. J., McKeating, J. A., Lanford, R. E., Feinstone, S. M., Major, M. E., Leroux-Roels, G., and Rice, C. M. (2006). Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro. *Proc Natl Acad Sci U S A* **103**(10), 3805-9.
- Lohmann, V., Hoffmann, S., Herian, U., Penin, F., and Bartenschlager, R. (2003). Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J Virol* **77**(5), 3007-19.
- Louagie, H. K., J.T. Brouwer, J.R. Delanghe, M.L. DeBuyzere, G.G. Leroux-Roels (1996). Haptoglobin polymorphism and chronic hepatitis C. *J Hepatol* **25**(1), 10-4.
- Love, R. A., H.E.Parge, J.A. Wickersham, Z. Hostomsky, N. Habuka, E.W. Moomaw, T. Adachi, and Z. Hostomska (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **87**, 331-342.
- Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., and Tschoopp, J. (2005). Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**(7062), 1167-72.
- Miller, R., and R.H. Purcell (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc Natl Acad Sci U S A* **87**, 2057-2061.
- Mondelli, M. U., Silini E. (1999). Clinical Significance of hepatitis C virus genotypes. *J Hepatol* **31**, 65-70.
- Muramatsu, S., Ishido, S., Fujita, T., Itoh, M., and Hotta, H. (1997). Nuclear localization of the NS3 protein of hepatitis C virus and factors affecting the localization. *J Virol* **71**(7), 4954-61.
- Nielsen, S. U., M.F. Bassendine, A.D. Burt, D.J. Bevitt, G.L. Toms (2004). Characterization of the genome and structural proteins of hepatitis C virus resolved from infected human liver. *J Gen Virol* **85**(6), 1497-507.
- Nishihara, K., M. Kanemori, M. Katigawa, H. Yanagi, T. Yura (1998). Chaperone coexpression plasmids: differential and synergistic roles of Dna-K-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol* **64**(5)(1694-9).
- Nishihara, K., M. Kanemori, H. Yanagi, T. Yura (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl Environ Microbiol* **66**(3), 884-9.
- Normura-Takigawa, Y., M. Nagano-Fujii, L. Deng, S. Kitazawa, S. Ishido, K. Sada, H.

- Hotta (2006). Non-structural protein 4A of Hepatitis C virus accumulates on mitochondria and renders the cells prone to undergoing mitochondria-mediated apoptosis. *J Gen Virol* **87**, 1935-45.
- Pang, P. S., Jankowsky, E., Planet, P. J., and Pyle, A. M. (2002). The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *Embo J* **21**(5), 1168-76.
- Pawlotsky, J. M. (2003). Hepatitis C virus genetic variability: pathogenic and clinical implications. *Clin Liver Dis* **7**(1), 45-66.
- Penin, F., Brass, V., Appel, N., Ramboarina, S., Montserret, R., Ficheux, D., Blum, H. E., Bartenschlager, R., and Moradpour, D. (2004). Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J Biol Chem* **279**(39), 40835-43.
- Phizicky, E.M., Fields, S. (1995) Protein-protein interactions : methods for detection and analysis. *Microbiol Rev.* **59**(1), 94-123.
- Pietschmann, T., and Bartenschlager, R. (2003). Tissue culture and animal models for hepatitis C virus. *Clin Liver Dis* **7**(1), 23-43.
- Poliakov, A., Hubatsch, I., Shuman, C. F., Stenberg, G., and Danielson, U. H. (2002). Expression and purification of recombinant full-length NS3 protease-helicase from a new variant of Hepatitis C virus. *Protein Expr Purif* **25**(3), 363-71.
- Poynard, T., Yuen, M. F., Ratziu, V., and Lai, C. L. (2003). Viral hepatitis C. *Lancet* **362**(9401), 2095-100.
- Prikhod'ko, E. A., Prikhod'ko, G. G., Siegel, R. M., Thompson, P., Major, M. E., and Cohen, J. I. (2004). The NS3 protein of hepatitis C virus induces caspase-8-mediated apoptosis independent of its protease or helicase activities. *Virology* **329**(1), 53-67.
- Reed, K. E., A. Grakoui, and C.M. Rice (1995). Hepatitis C virus-encoded NS2-3 protease: Cleavage-site mutagenesis and requirements for bimolar cleavage. *J Virol* **69**, 4127-4136.
- Reich, N. C. (2002). Nuclear/cytoplasmic localization of IRFs in response to viral infection or interferon stimulation. *J Interferon Cytokine Res* **22**(1), 103-9.
- Remis, R.S. and the Hepatitis C Writing Group for Blood-born Pathogens Division, LCDC, Health Canada. Estimating the number of blood transfusion recipients infected by hepatitis C virus in Canada, 1960-85 and 1990-92. June 22, 1998.
- Sadrzadeh, S. M., E. Graf, S.S. Panter, P.E. Hallaway, J.W. Eaton (1984). Hemoglobin. A biological fenton reagent. *J Biol Chem* **259**(23), 14354-6.
- Sakamuro, D., Furukawa, T., and Takegami, T. (1995). Hepatitis C virus nonstructural protein NS3 transforms NIH 3T3 cells. *J Virol* **69**(6), 3893-6.
- Seth, R. B., L. Sun, C.K. Ea, Z.J. Chen (2005). Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF3. *Cell* **122**(5), 669-82.
- Shepard, C., L. Finelli and MJ Alter (2005). Global Epidemiology of Hepatitis C Virus infections. *Lancet Infectious Disease* **5**, 558-67.
- Shoji, I., Suzuki, T., Sato, M., Aizaki, H., Chiba, T., Matsuura, Y., and Miyamura, T. (1999). Internal processing of hepatitis C virus NS3 protein. *Virology* **254**(2), 315-23.
- Smith, E. J., I. Marie, A. Prakash, A. Garcia-Sastre, D.E. Levy (2001). IRF3 and IRF7

- phosphorylation in virus-infected cells does not require double-stranded RNA-dependent protein kinase R or Ikappa B kinase but is blocked by Vaccinia virus E3L protein. *J Biol Chem* **276**(12), 8951-7.
- Srinivasan, R. (2001). Autoimmune hemolytic anemia in treatment-naive chronic hepatitis C infection. *J Clin Gastroenterol* **32**(3), 245-7.
- Sumpter, R., Jr., Loo, Y. M., Foy, E., Li, K., Yoneyama, M., Fujita, T., Lemon, S. M., and Gale, M., Jr. (2005). Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J Virol* **79**(5), 2689-99.
- Tan, S. L., H. Nakao, Y. He, S. Vijaysri, P. Neddermann, B.L. Jacobs, B.J. Mayer, M.G. Katze (1999). NS5A, a non-structural protein of hepatitis C virus binds growth factors receptor bound protein 2 adaptor protein in a Src homology 3 domain/ligand dependent manner and perturbs mitogenic signaling. *Proc Natl Acad Sci U S A* **96**, 5533-8.
- Taremi, S. S., Beyer, B., Maher, M., Yao, N., Prosis, W., Weber, P. C., and Malcolm, B. A. (1998). Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease. *Protein Sci* **7**(10), 2143-9.
- Tellinghuisen, T. L., C.M. Rice (2002). Interactions between hepatitis C virus proteins and host cell factors. *Curr Opin Microbiol* **5**(4), 419-27.
- Thoren, F., Romero, A., Lindh, M., Dahlgren, C., and Hellstrand, K. (2004). A hepatitis C virus-encoded, nonstructural protein (NS3) triggers dysfunction and apoptosis in lymphocytes: role of NADPH oxidase-derived oxygen radicals. *J Leukoc Biol* **76**(6), 1180-6.
- Tu, H., D.R. Taylor, T. Yang, A.K. Mircheff, Y. Wen, A.E. Gorbalenya, S.B. Hwang, and M.M. Lai. (1999). Hepatitis C Virus RNA polymerase and NS5A complex with a SNARE-like-protein. *Virology* **263**(1), 30-41.
- Vishnuvardhan, D., Kakiuchi, N., Urvil, P. T., Shimotohno, K., Kumar, P. K., and Nishikawa, S. (1997). Expression of highly active recombinant NS3 protease domain of hepatitis C virus in E. coli. *FEBS Lett* **402**(2-3), 209-12.
- Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H. G., Mizokami, M., Bartenschlager, R., and Liang, T. J. (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**(7), 791-6.
- Watashi, K., and Shimotohno, K. (2003). The roles of hepatitis C virus proteins in modulation of cellular functions: a novel action mechanism of the HCV core protein on gene regulation by nuclear hormone receptors. *Cancer Sci* **94**(11), 937-43.
- Wolk, B., Sansonno, D., Krausslich, H. G., Dammacco, F., Rice, C. M., Blum, H. E., and Moradpour, D. (2000). Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3-NS4A complex expressed in tetracycline-regulated cell lines. *J Virol* **74**(5), 2293-304.
- Xu, L. G., Y.Y. Wang, K.J. Han, L.Y. Li, Z. Zhai, H.B. Shu (2005). VISA is an adaptor protein required for virus-triggered IFN-beta signaling. *Mol Cell* **19**(6), 727-40.
- Xu, Z., Choi, J., Lu, W., and Ou, J. H. (2003). Hepatitis C virus f protein is a short-lived protein associated with the endoplasmic reticulum. *J Virol* **77**(2), 1578-83.

- Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J. L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C., and Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci* **7**(4), 837-47.
- Yang, S. H., Lee, C. G., Song, M. K., and Sung, Y. C. (2000). Internal cleavage of hepatitis C virus NS3 protein is dependent on the activity of NS3A protease. *Virology* **268**(1), 132-40.
- Yi, M., Villanueva, R. A., Thomas, D. L., Wakita, T., and Lemon, S. M. (2006). Production of infectious genotype 1a hepatitis C virus (Hutchinson strain) in cultured human hepatoma cells. *Proc Natl Acad Sci U S A* **103**(7), 2310-5.
- Zemel, R., Gerechet, S., Greif, H., Bachmatove, L., Birk, Y., Golan-Goldhirsh, A., Kunin, M., Berdichevsky, Y., Benhar, I., and Tur-Kaspa, R. (2001). Cell transformation induced by hepatitis C virus NS3 serine protease. *J Viral Hepat* **8**(2), 96-102.
- Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D. R., Wieland, S. F., Uprichard, S. L., Wakita, T., and Chisari, F. V. (2005). Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* **102**(26), 9294-9.

VI) Appendix 1

Colony #	In frame	Interaction Region	Gene Start	gene
88	seq issue			?
7	seq issue			?
289	seq issue			?
297	seq issue			?
23	unclarified ORF			?
101	unclarified ORF			?
150	unclarified ORF			?
172	unclarified ORF			?
97	unclarified ORF			?
318	N			10 formyltetrahydrofolate dehydrogenase
335	seq issue			actin beta chain
387	Y			actin beta chain
35	N			actin, gamma 1
409	Y	911	188-1225	adenosine kinase
267	Y			adenosine kinase
402	Y			aldehyde dehydrogenase 2 family
125	Y	1621	442-1995	aldehyde dehydrogenase 2 family
371	N			aldehyde oxidase
383	N			aldehyde oxidase
400	N			aldehyde oxidase
184	N			aldehyde oxidase
315	N			aldehyde oxidase
401	N			aldehyde oxidase
259	N			aldehyde oxidase
286	N			aldehyde oxidase
292	fusion	685	126-1220	aldolase B
1	Y			aldolase B
6	Y			aldolase B
10	Y			Aldolase B
59	Y			Aldolase B
104	Y			Aldolase B
162	Y			Aldolase B
355	Y			Aldolase B
384	Y			Aldolase B
394	Y			Aldolase B
2	Y			aldolase B
81	Y	453	126-1220	aldolase B
208	Y			aldolase B
227	Y			aldolase B
320	Y			aldolase B
324	Y			aldolase B
338	Y			aldolase B
291	Y			aldolase B
17	Y			aldolase B
91	Y	444	126-1220	aldolase B
367	Y			alpha 2 HS glycoprotein
221	Y			alpha 2 HS glycoprotein
341	Y			alpha 2 HS glycoprotein
393	Y			alpha 2 HS glycoprotein

248	Y	340	1-1104	alpha 2 HS glycoprotein
418	Y			alpha enolase
155	Y	1086	489-1514	alpha enolase
321	Y	1086	489-1514	alpha enolase
179	fusion	1		alpha-1-antitrypsin
269	seq issue			alpha-1-antitrypsin
276	seq issue			alpha-1-antitrypsin
8	Y			alpha-1-B glycoprotein
145	Y			alpha-1-B glycoprotein
9	Y			alpha-1-B glycoprotein
40	Y			alpha-1-B glycoprotein
198	Y			alpha-1-B glycoprotein
257	Y			alpha-1-B glycoprotein
410	Y	1156	55-1542	alpha-1-B glycoprotein
12	seq issue			alpha-2-HS glycoprotein
53	unclarified ORF			alpha-2-HS glycoprotein
252	Y	71aa	1-367aa	alpha-2-HS glycoprotein
190	Y	657	69-1295	aminoacylase-1 (ACY1)
68	N			Angiogenin
160	N			angiogenin gene
421	N			angiogenin gene
149	seq issue			apolipoprotein H
82	Y			Apolipoprotein H
95	Y			Apolipoprotein H
96	Y			Apolipoprotein H
127	Y			apolipoprotein H
148	Y			apolipoprotein H
154	Y			Apolipoprotein H
295	Y			Apolipoprotein H
419	Y			Apolipoprotein H
128	Y	11	32-1069	apolipoprotein H
414	Y	11	32-1069	apolipoprotein H
135	Y	20	32-1069	apolipoprotein H
78	Y	11	32-1069	apolipoprotein H
372	Y	20	32-1069	apolipoprotein H
282	Y			apolipoprotein H
239	Y			apolipoprotein H
218	Y			apolipoprotein H
224	Y			apolipoprotein H
339	Y			apolipoprotein H
280	Y			apolipoprotein H
102	Y	11	32-1069	apolipoprotein H
141	Y			apolipoprotein H
3	fusion			apolipoprotein L
51	fusion			aspartoacylase (aminocyclase) 3
214	Y	564	97-998	ATPase Na K transporting, beta 1 polypeptide
117	N			ATPase type 13A3
246	N			Autism related protein 1
24	seq issue			B beta polypeptide
417	unclarified ORF			basic transcription factor 3 (BTF3) pseudogene

234	Y			beta actin
319	Y			beta actin
242	Y			beta ureidopropionase
177	Y	587	122-1276	beta ureidopropionase
264	N			BMP binding endothelial regulator (BMPER)
114	Y	334	1-1178	branched chain alpha-keto acid, dehydrogenase (BCKDHB) E1-beta subunit
314	N			carbaryl phosphate synthetase
348	N			carbaryl phosphate synthetase
262	Y	795	84-1667	catalase
413	Y	710	113-1132	cathepsin B proteinase
99	N			CD14
192	N			CD14
44	N			cell surface protein TAPA-1 mRNA (CD-81)
204	N			CG9886-like mRNA
299	Y	1265	200-1984	choline dehydrogenase
245	N			chornicon homolog 4 (Drosophila)
115	unclarified ORF			Chrom 14/alpha antitrypsin flanking region?
181	Y	protein analysis 108-170aa		Chrom 17/esp2 2 (EPN2) transcript variant
237	unclarified ORF			Chrom 4?
216	seq issue			Chromosome 1
226	seq issue			Chromosome 10
134	seq issue			chromosome 17
312	seq issue			chromosome 19
62	fusion			chromosome 19 clone
275	fusion			chromosome 2
66	fusion			chromosome 22
219	seq issue			Chromosome 3
225	seq issue			Chromosome 3
164	fusion			chromosome 3p21.3
131	unclarified ORF	-	-	chromosome 5
283	seq issue			chromosome 8
83	Y	11	32-1069	Chromosome 9
407	Y			coagulation factor XII
133	Y	917	50-1897	coagulation factor XII
313	Y	65	50-1897	coagulation factor XII
365	fusion	15	102-839	complement C1q-C
55	Y	2890	52-5286	complement C4A
73	Y			complement C8
140	Y			complement C8
175	Y			complement C8
178	Y			complement C8
294	Y			complement C8
63	Y	1142	68-1843	complement C8
202	Y			complement C8
349	Y	1169	68-1843	complement C8
130	Y			complement C8 alpha subunit

268	Y	1314	138-1892	complement C8 alpha subunit
170	Y			complement C8 beta subunit
271	Y			complement C8 beta subunit
203	Y	1728	156-2960	Complement component 6
308	Y	2286	156-2960	Complement component 6
107	Y	4690	86-5318	complement component C4A mRNA
157	Y	658 (starts at +4)	46-1038	complement factor H
374	Y			complement factor B
92	fusion			complement factor B mRNA
121	Y	2138	227-2521	complement factor B mRNA
376	Y			complement factor B mRNA
396	Y			complement factor B mRNA
405	Y			complement factor B mRNA
309	Y	187	46-1038	Complement H
137	Y			corticotropin releasing factor
136	Y	782	281-1249	corticotropin releasing hormone binding protein
52	N			Cox 1
284	Y	7832	7337-8020	Cox 2 (Mitochondria)
235	Y	7871	7337-8020	Cox 2 (mitochondrial)
244	N			Cox1 (Mitochondrial)
142	Y	7553	7337-8020	Cox-2
422	Y	7907	7586-8269	Cox2 (Mitochondria (cox2))
322	Y	7697	7337-8020	Cox2 (Mitochondria)
379	Y	7697	7337-8020	Cox2 (Mitochondria)
197	Y	7826	7337-8020	Cox2 (mitochondrial)
375	Y	7871	7337-8020	Cox2 (mitochondrion)
228	Y	7658	7337-8020	Cox2 (mitochondrion)
423	N			Cox3 (mitochondrial protein)
41	Y	484	493-957	cyclin G1 binding protein 1 AF093571
337	Y	142	76-441	cysteine proteinase inhibitor precursor cystatin C
288	Y			cytoskeletal gamma actin
370	Y	768	60-1187	cytoskeletal gamma actin
420	Y	119	26-382	D-dopachrome tautomerase
329	seq issue			DEAD (Asp-Glu-Ala-Asp) box polypeptide 50
158	Y	538	1-1080	decorin B
342	Y	868	22-1584	dihydrolipoamide dehydrogenase
266	Y	984	6-1484	dipeptidyl-peptidase 7
21	-			DOES NOT AMPLIFY
79	-			DOES NOT AMPLIFY
112	-			DOES NOT AMPLIFY

132	-			DOES NOT AMPLIFY
183	-			DOES NOT AMPLIFY
199	-			DOES NOT AMPLIFY
251	-			DOES NOT AMPLIFY
258	-			DOES NOT AMPLIFY
332	-			DOES NOT AMPLIFY
298	Y	298	32-604	dynactin 6
272	Y		29	Dynactin 6 (WS-3 protein)
352	Y	1357	469-1857	elongation factor 1 alpha
77	fusion			elongation factor 1-alpha (EF1A)
98	Y	1693	1-2577	elongation factor ef2
27	Y	352	172-861	endothelial cell apoptosis protein E-CE1
116	Y	3	15-887	enoyl Coenzyme A hydratase
				eukaryotic elongation translation factor 1 alpha
305	Y	1219	469-1857	alpha
373	Y	559	136-1017	F box protein FBG2 AF233223
165	Y	1353	36-2021	factor XIIIb
351	Y			factor XIIIb
215	Y	1338	36-2021	factor XIIIb
31	Y			fibrinogen beta chain
28	Y			fibrinogen beta chain
69	Y			fibrinogen beta chain
156	Y			fibrinogen beta chain
220	Y			fibrinogen beta chain
240	Y			fibrinogen beta chain
316	Y			fibrinogen beta chain
323	Y			fibrinogen beta chain
343	Y			fibrinogen beta chain
344	Y			fibrinogen beta chain
346	Y			fibrinogen beta chain
357	Y			fibrinogen beta chain
377	Y			fibrinogen beta chain
408	Y			fibrinogen beta chain
277	Y			fibrinogen beta chain
33	Y			fibrinogen beta chain
54	Y	524	9-1484	fibrinogen beta chain
36	Y	640	1-1314	fibrinogen, gamma polypeptide
171	N			Flavin containing monooxygenase
169	N			Flavin containing monooxygenase
168	seq issue			Flavin containing monooxygenase
144	fusion - N			FLJ10856/transmembrane protein 30A
103	fusion			FLJ32065 protein
159	N			FLJ33074 fis (WD repeat domain 23)
247	Y			flotillin-1 mRNA
307	Y			flotillin-1 mRNA
90	Y	41	23-1306	flotillin-1 mRNA
				Galactose Mutarotase (aldolase 1 epimerase)
270	Y			
14	N			Glutathione peroxidase
303	N			Glutathione peroxidase
				glyceraldehyde 3 phosphate dehydrogenase
392	Y	269	53-1060	

16	unclarified ORF			Green Homo sapiens STS cDNA
281	N			H.sapiens SPFH domain family member 2 (SPFH2)
222	N			H.sapiens Zinc finger protein
25	Y			haptoglobin
118	Y			haptoglobin
139	Y			haptoglobin
143	Y			haptoglobin
146	Y			haptoglobin
34	Y	363	99-944	haptoglobin
196	Y			haptoglobin
163	Y			haptoglobin
191	Y			haptoglobin
29	Y			haptoglobin alpha 2
300	Y	211	328-1014	haptoglobinalpha 1S (X00637)
243	fusion	208	1-372	HBeAg binding protein
353	Y	397	430-771	HBV pre-s2 binding protein 1 (SBP1)
75	Y	562	46-1767	heat shock 60kDal protein
254	Y	586	46-1767	heat shock 60kDal protein
56	fusion - N			HERPUD1 gene for stress
4	fusion - N			HERPUD1 gene for stress
57	N			HERPUD1 gene for stress
58	N			HERV-H LTR-associating 3
60	N			HERV-H LTR-associating 3 similar to
61	N			HERV-H LTR-associating 3 similar to
296	fusion			histidine rich glycoprotein
173	unclarified ORF			IgM heavy chain constant/?
46	Y	671	209-1261	inhibin beta E (Activin beta E)
260	seq issue			insulin like growth factor II
18	Y	393	117-992	insulin-like growth factor binding protein 3
328	Y	414	117-992	insulin-like growth factor binding protein 3
231	Y			inter-alpha-trypsin inhibitor
229	Y	2146	25-2760	inter-alpha-trypsin inhibitor
287	Y			inter-alpha-trypsin inhibitor
86	Y			inter-alpha-trypsin inhibitor
72	N			ISG12 protein/Interferon alpha inducible protein 27
85	fusion			Jak2 kinase
369	Y	1342	40-1683	lactamase, beta (LACTB)
249	Y	704	71-1342	L-arginine:glycine amidinotransferase
39	N			leukemia virus receptor 1 (GLVR1)
385	N			lymphocyte antigen 96
113	N			lysophospholipase (LPL1)
124	Y	97	1-351	Macaca fascicularis brain/HepBeAg binding protein 2
93	N			mahogunin ring finger 1 (MGRN1)
152	N			mannose-specific lectin U09716
105	N			metallothionein
87	N			metallothionein
45	N			metallothionein 2 A

336	N			metallothionein 2 A
26	N			metallothionein 2A
42	N			metallothionein 2A
48	N			metallothionein 2A
65	N			metallothionein 2A
67	N			metallothionein 2A
100	N			metallothionein 2A
108	N			metallothionein 2A
109	N			metallothionein 2A
120	N			metallothionein 2A
147	N			metallothionein 2A
185	N			metallothionein 2A
188	N			metallothionein 2A
206	N			metallothionein 2A
209	N			metallothionein 2A
210	N			metallothionein 2A
211	N			metallothionein 2A
213	N			metallothionein 2A
223	N			metallothionein 2A
233	N			metallothionein 2A
236	N			metallothionein 2A
256	N			metallothionein 2A
273	N			metallothionein 2A
274	N			metallothionein 2A
290	N			metallothionein 2A
310	N			metallothionein 2A
333	N			metallothionein 2A
334	N			metallothionein 2A
356	N			metallothionein 2A
358	N			metallothionein 2A
359	N			metallothionein 2A
364	N			metallothionein 2A
366	N			metallothionein 2A
378	N			metallothionein 2A
381	N			metallothionein 2A
382	N			metallothionein 2A
386	N			metallothionein 2A
388	N			metallothionein 2A
390	N			metallothionein 2A
391	N			metallothionein 2A
397	N			metallothionein 2A
398	N			metallothionein 2A
403	N			metallothionein 2A
404	N			metallothionein 2A
415	N			metallothionein 2A
424	N			metallothionein 2A
166	N			metallothionein 2A
5	N			metallothionein 2A
182	N			metallothionein 2A
189	N			metallothionein 2A
89	Y	1728	156-2960	MHC II complement C6 J0524
411	N			Mitochondria

412	N			Mitochondria
176	N			Mitochondria
11	seq issue			mitochondrial matrix protein P1
71	Y	1071	3-5540	myosin light chain kinase
207	N			Na K ATPase beta 1 U16799
200	Y	181	184-418	NADH ubiquinone oxidoreductase
212	fusion	3342	3059-4015	ND1 (mitochondria)
241	fusion	3342	3059-4015	ND1 (Mitochondrial)
126	N			ND4L (mitochondria)
263	N			Niemann-Pick disease (Epididymal secretory protein)
326	N			non-metastatic cells 1, protein (NM23A)
195	N			Not human
395	seq issue			PAC clone
20	N			PEST-containing transporter(XPCT) gene (X-linked)
30	Y			phenylalanine hydroxylase
340	Y	1514	473-1831	phenylalanine hydroxylase L47726
253	N			phosphatase methylesterase-1
122	N			phosphoglucomutase 1
151	Y			plasminogen
416	Y			plasminogen
70	Y	1963	55-2487	plasminogen
186	Y	2083	1-2961	procollagen C proteinase
193	Y	2083	1-2961	procollagen C proteinase
306	Y	BY SEQUENCE ANALYSIS		procollagen C proteinase
37	Y			procollagen C proteinase
399	Y			procollagen C-proteinase
302	N			procollagen lysine 1,2 oxoglutarate 5-dioxygenase 1
205	N			procollagen lysine 1,2 oxoglutarate 5-dioxygenase 1
301	N			procollagen-lysine
138	N			Procollagen-lysine/1,2-oxoglutarate 5-dioxygenase
354	Y	493	1-1107	putative DNA polymerase delta p38 subunit
84	N			RAB14 member of RAS oncogene family
232	seq issue			reseq
22	seq issue			reseq
32	seq issue			reseq
255	seq issue			reseq
261	seq issue			reseq
368	seq issue			reseq
94	seq issues			reseq
161	seq issues			reseq
119	seq issues			reseq
180	seq issues			retinoic acid induced RIG-E precursor
43	unclarified ORF			retinoic acid inducer RIG-E precursor

265	fusion		291	retinol binding protein 4
304	seq issue			ribonuclease 6 precursor
250	Y	585	408-1178	ribonuclease T2 (ribonuclease 6 precursor)
64	Y	46	58-1098	RNA Polymerase 1 polypeptide C
74	Y			RNA Polymerase 1 polypeptide C subunit
123	Y	543	150-1583	S protein/Vitronectin
361	Y	2252	218-2371	SAD1 unc-84 domain protein
311	seq issue			sequence 191 from patent WO0129221
194	N			sequence 4 from patent Wo0138579
38	Y			Sequence 62 from patent
167	Y			Sequence 62 from patent W00073328
174	Y			sequence 62 from patient W00073328
347	Y			sequence 62 from patient W00073328
278	N			SH3 and multiple ankaryin repeat (SHANK3)
279	N			SH3 and multiple ankaryin repeat (SHANK3)
350	seq issue			signal recognition particle receptor ('docking protein')
238	Y	5005	1-5757	similar to adapter related protein complex 2
345	fusion			sirtuin (silent mating type information regulation 2) homolog
360	Y	759	297-1229	sirtuin (silent mating type information regulation 2) homolog
406	N			solute carrier family 20 (phosphate transporter)
47	N			Solute carrier family 20 (phosphate transporter) member 1
330	fusion			solute carrier family 39 (zinc transporter)
110	Y	369	180-795	SOUL protein (SOUL)/Heme binding protein 2
187	Y			sphingolipid activator proteins 1
129	Y	449	53-775	sphingolipid activator proteins 1
425	Y	572	152-994	succinate-ubiquinone oxidoreductase iron sulfate (lp)
327	Y			tararotene induced gene 2, retinoic acid receptor
293	Y	163	61-552	tazarotene-induced gene 2 (TIG2)
201	Y	491	383-1051	Threonine synthase-like-1
15	Y	491	383-1051	Threonine synthase-like-1
362	N			transcription factor 7-like 1
389	seq issue			transcription factor IIB
285	Y			transthyretin
325	Y	325	27-470	transthyretin
153	N			tryptophan 2,3-dioxygenase
106	N			tryptophan 2,3-dioxygenase
49	Y			tyrosine phosphate PTPase
230	Y	103	195-1256	tyrosine phosphate PTPase
13	seq issue			ubiquitouslyexpressed transcript

76	fusion			UDP-glucose dehydrogenase (UGDH)
19	N			X-linked PEST-containing transporter
111	N			zinc finger
80	N			Zinc finger protein
380	N			Zinc finger protein
50	N			Zinc finger protein
331	Y	1122	18-1946	Zinc finger protein 398
363	Y	279	12-908	Zn alpha 2 glycoprotein
217	Y			Zn alpha 2 glycoprotein M76707
317	Y			Zn Alpha 2 glycoprotein M76707