Characterisation of the Membrane and Protein Interactions of the Hepatitis C Virus NS2

Barnabas James King (BSc (Hons))

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
Institute of Molecular and Cellular Biology
September 2011

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Go Team!

Acknowledgements

I would like to thank my supervisors Mark and Steve; for putting up with me and showing such patience in the face of my many mistakes and misunderstandings, and my previous supervisor, Geoff, whos wonderful idea this all was.

Huge thanks to Dr Tedbury who, despite moving over 3,500 miles away, has always been encouraging and ready to discuss all things NS2-related. Thanks also to Dr Tedbury and Dr (Dr) Corless for reading through my thesis to make sure every sentence contains a verb and keeping the tangents to a minimum.

To all those in 8.61, past and present, who made my time in Leeds so enjoyable and who got me through the late nights, the 'cloning-wars' and the n^{th} repeats – thank you! In particular I would like to thank Jamel, Andy, Sarah G, Anna, Corine, Mair, Tosh, Carsten, Phil, Lynsey and Loo-Loo for making me feel so welcome and helping me to settle in.

Over the last four years in Leeds I have lived with a lot of different people. I hope they all know how much I care for them all and appreciated the company and distraction from science. The fact that I left my room and ate something for the last month is entirely due to Jonny, Flis and Tim. Thanks are not enough.

Finally I would like to thank my family who have always been supportive, encouraging and put up with me through the highs and protracted lows.

There are many more people to thank for many different reasons, but hopefully I will be able to do that in person.

Abstract

Hepatitis C virus (HCV) readily establishes a chronic infection and it is currently thought to infect 2-3% of the global population. Treatment for HCV places a severe burden on patients, leading to a premature treatment termination in approximately 20% of individuals, and has poor efficacy against the predominant genotype. A greater understanding of the virus lifecycle and mechanisms of persistence will provide valuable information in the continuing development of direct-acting antiviral compounds.

HCV encodes 10 proteins which are translated as a single polyprotein. Non-structural protein 2 (NS2) is a small, hydrophobic, *trans*-membrane protein, yet the precise number and position of its *trans*-membrane domains remain unclear. NS2 is required for virion morphogenesis but it is not required for replication of the viral genome and it does not form part of the mature virion. NS2 has been shown to interact with other viral proteins, potentially via intra-membrane contacts. Determining the topology of NS2 is therefore vital to our understanding of how NS2 interacts with the other viral proteins and functions within the virus lifecycle.

The interactions between NS2 and membranes and the viral glycoprotein E2 were investigated by truncation analysis and fusion with reporter proteins. Glycosylation analysis demonstrated that the N- and C-termini of NS2 are oriented to the luminal and cytosolic faces of the ER membrane, respectively. Truncation of NS2 at residue 70 oriented a C-terminal reporter fusion to the ER lumen consistent with the formation of a luminal loop. This is the first biochemical evidence that NS2 assumes a 3 TMD topology. Co-immunoprecipitation studies with E2 and eGFP-tagged truncation of NS2 revealed that NS2 forms multiple interactions with E2 and that these interactions are dependent upon NS2 targeting to membranes but likely independent of NS2 topology. A model of NS2 topology is presented.

Table of Contents

List of Figures

List of Tables

Abbreviations

3D	3 dimensional
A	Adenosine
aa	Amino acid
ADRP	Adipose differentiation-related protein
AH	Amphipathic helix
AMP	Adenosine monophosphate
AMPK	AMP-activated protein kinase
Apo	Apolipoprotein
APS	Ammonium persulphate
ARFP	Alternate reading frame protein
Asn	Asparagine
ATP	Adenosine triphosphate
β-gal	β-galactosidase
β-lac	β-lactamase
B#	NS2- β-gal fusion where # is a number from 1 to 10
BN	Blue-native
BSA	Bovine serum albumin
BVDV	Bovine viral diarrhoea virus
C	Core protein
C	Cysteine
C	Cytosine
CD4+	Cluster of differentiation 4 positive
CD81	Cluster of differentiation 81
cDNA	Complementary DNA
CFP	Cerulean fluorescent protein
CFTR	Cystic fibrosis transmembrane conductance regulator
CID/ml	Chimpanzee infectious dose per ml
CIDE	Cell death-inducing DFF 45-like effecter protein
CKII	Casein kinase II
CLDN-I	Claudin I
CMM	Canine microsomal membranes
CoA	Acetyl co-enzyme A
Co-IP	Co-immunoprecipitation

COS7	CV-I in origin (African Green Monkey kidney cell line), replication defective SV40-containing cell line 7
CSFV	Classical swine fever virus
C-terminus	Carboxyl-terminus
CyP	Cyclophilin
D	Aspartic acid
DAA	Direct-acting antivirals
DEPC	Diethyl pyrocarbonate
DFF	DNA fragmentation factor
DM	n-decyl-β-D-maltopyranoside
DMAT	2-Dimethylamino-4, 5, 6, 7-tetrabromo-1H-benzimidazole
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNJ	Deoxynojirimycin
ds	Double stranded
DTT	Dithiothreitol
E	Envelope
E	Glutamic acid
E. coli	Escherichia coli
E5	Early protein 5
EBV	Epstein-Barr virus
ECL	Enhanced chemiluminescence
EDTA	Ethylenediamine tetraacetic acid
eGFP	Enhanced green fluorescence protein
EGFR	Epidermal growth factor receptor
EHMs	Extra-hepatic manifestations
eIF3	eukaryote initiation factor 3
ELISA	Enzyme-linked immuno-sorbent assay
EMCV	Encephalomyocarditis virus
Endo H	Endoglycosidase H
Eps15	EGFR protein tyrosine kinase substrate 15
ER	Endoplasmic reticulum

ERAD	Endoplasmic reticulum-associated degradation
F protein	Frameshift protein
FAS	Fatty acid synthase
FFP	Fluorescence protease protection
FL	Full-length
FLAG®	FLAG oligo-peptide N-DYKDDDDK-C
FMDV	Foot and mouse disease virus
FRET-FLIM	Fluorescence resonance energy transfer with fluorescence lifetime imaging microscopy
fur	Furin
G	Glycine
g	Gravitational force
G	Guanidine
GAG	Glucosaminoglycans
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GBV-B	GB virus B
GBV-C	GB virus C, synonymous with hepatitis G virus
GBV-D	GB virus D
GFP	Green fluorescence protein
GlcNAc	*N*-acetylglucosamine
Gly	Glycine
H	Histidine
H77	Hutchinson HCV isolate 77, genotype 1a
HAART	Highly active anti-retroviral therapy
HAV	Hepatitis A virus
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
hCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HDL	High density lipoprotein
HDV	Hepatitis D virus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HepG2	Hepatocellular carcinoma cell line G2
HEV	Hepatitis E virus

HGV	Hepatitis G virus, synonymous with GB virus C
HIV	Human immunodeficiency virus
HMM	Hidden Markov models
HPV	Human papilloma virus
HRP	Horse radish peroxidase
Huh7	Human hepatocellular carcinoma cell line 7
I	Isoleucine
IBV	Infectious bronchitis virus
ICAM-I	Intercellular adhesion molecule 1
IF	Immunofluorescence
IFN-α	Interferon-α
IgM	Immunoglobulin M
IL	Interleukin
IP	Immunoprecipitate
IRES	Internal ribosomal entry site
IU	International units
J6	Japanese HCV isolate 6, genotype 2a
JFH-1	Japanese fulminant HCV isolate 1, genotype 2a
Jiv	J-domain protein interacting with viral protein
kDa	Kilodaltons
L	Leucine
LB	Luria-Bertani
LCS	Low complexity sequence
LD	Lipid droplets
LDL	Low density lipoprotein
LDL-R	Low density lipoprotein receptor
LEL	Large, extracellular loop
Lys	Lysine
M	Membrane fraction
MAVS	Mitochondrial antiviral signalling protein
MCS	Multiple cloning site
miR-122	Micro RNA 122
MM	Micosomal membranes
mRNA	Messenger RNA

mtHSP70	Mitochondrial heat-shock protein 70
MTP	Microsomal transfer protein
N	Asparagine
NANB/NANBH	Non-A, non-B hepatitis
NF-κB	Nuclear factor κB
NI	Nucleoside-analogue inhibitors
NMR	Nuclear magnetic resonance
NN-DNJ	N-nonyldeoxynojirimycin
NN-DNJ	N-nonyldeoxynojirimycin
NNI	Non-nucleoside-analogue inhibitors
NPT	Neomycin phosphotransferase II
NS#	Non-structural protein
nsAP	Non-secreted alkaline phosphatase
nt	Nucleotide
N-terminus	Amino-terminus
OCLN	Occluin
ONPG	*O*-nitrophenyl-β-D-galactopyranoside
ORF	Open reading frame
OST	Oligosaccharyltransferase
P	Pellet fraction
P	Proline
PACE	Paired basic amino acid cleaving enzyme
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBST	PBS containing0.1% v/v Tween20
PCR	Polymerase chain reaction
PDI	Protein disulphide isomerase
PhoA	Alkaline phosphatase
PIPES	Piperazine-N,N'-bis (2-ethanesulfonic acid)
PLAP	Placental alkaline phosphatase
PLB	Passive lysis buffer
PNGase F	Peptide *N*-glycosidase F
poly (U/UC)	Polypyrimidine tract
PTB	Polypyrimidine tract-binding protein

PVDF	Polyvinylidene fluoride
Q	Glutamine
QC	QuikChange
R.E.	Reporter enzyme
RdRp	RNA-dependent RNA polymerase
RLU	Relative light units
RNA	Ribonucleic acid
RNase	Ribonuclease
RT-PCR	Reverse-transcriptase PCR
RVR	Rapid virological response
S	Sedimentation (unit of sedimentation)
S	Serine
S	Soluble fraction
S#	NS2-SEAP fusion nomenclature where # is a number from 1 to 13
SCAM™	Substituted cysteine accessibility method for *trans*-membrane topology
SDS PAGE	Sodium dodecyl sulphate
SEAP	Secreted alkaline phosphatase
Ser	Serine
SGR	Subgenomic replicon
SH#	Src-homology domain where # is the number 2 or 3
SLO	Streptolysin-*O*
SP	Signal peptide
SPP	Signal peptide peptidase
SPp7	Signal peptide – p7
SR-BI	Scavenger receptor class B type I
SREBP-1c	Sterol regulatory element-binding protein 1c
ss	single stranded
STAT-C	Specifically targeted antiviral therapies for hepatitis C
SV40	Simian virus 40
SVR	Sustained virological response
T	Threonine
T	Thymine
TAP 1/2	Antigen peptide transporter 1/2
TBS	Tris buffered saline

T-Cell	Thyroid cell
TEMED	N,N,N',N'-tetramethylethylenediamine
TEV	Tobacco etch virus
TFE	Tetrafluoroethylene
TGN	*Trans*-Golgi network
TIR	Toll/IL-1 receptor
TMA	Transcription-mediated amplification
TMD	*Trans*-membrane domain
TMS	*Trans*-membrane segment
TNF-α	Tumour necrosis factor-α
tNS2	Truncated NS2
TNT	Transcription and translation
TRIFF	TIR-domain-containing adapter-inducing interferon-β
U	Units
U	Uridine
UC	Uridine-Cytosine dimer
UK	United Kingdom
Un	Untreated
UPR	Unfolded protein response
US	United States of America
UTR	Untranslated region
V	Valine
v/v	Volume per volume
VLDL	Very low density lipoprotein
VLP	Virus-like particule
vNS2	Virus NS2
VSK3	Vesicle surface kinase 3
w/v	Weight per volume
WB	Western blot
WHO	World Health Organisation
wt	Wild type
X-gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YFP	Venus yellow fluorescent protein
YFV	Yellow fever virus

CHAPTER 1: INTRODUCTION

1.1 Viral hepatitis

Viral hepatitis is a broad term referring to inflammation of the liver caused by infection with one of five human viruses: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV) and hepatitis E virus (HEV). A sixth virus known as hepatitis G virus (HGV) is synonymous with the name GB virus C (GBV-C) (Linnen *et al.*, 1996). GBV-C is widespread throughout the population but it is not directly associated with hepatitis or any other disease pathology (Mohr *et al.*, 2011; Stapleton, 2003). Although other viruses have been associated with hepatitis in humans, including yellow fever virus (YFV), human cytomegalovirus (CMV) and Epstein-Barr virus (EBV); they are not commonly associated with the term viral hepatitis. It is thought that approximately 20% of viral hepatitis is not caused by hepatitis A-E viruses (Leary *et al.*, 1996). It may be that newly identified viruses such as GB virus D (GBV-D), which has currently only been recovered from bats, are responsible for some of these infections (Epstein *et al.*, 2010).

HAV is the prototype member of the *Hepatovirus* genus within the *Picornaviridae* family. HAV is spread by the faecal-oral route and typically causes an acute, self-limiting disease. HAV causes roughly 1.5 million cases of clinical hepatitis annually. The majority of cases occur in areas of high endemicity where people are infected in early childhood and infection is often asymptomatic. There are several inactivated-virus vaccines available for HAV but their use is recommended for those in high risk groups, such as those travelling from developed countries to areas where the virus is endemic (WHO, 2011a).

HBV is a member of the *Orthohepadnavirus* genus, within the *Hepadnaviridae* family. HBV causes both acute and chronic diseases and is spread parenterally. The symptoms of acute infection tend to be mild but chronic infection can cause severe liver disease and hepatocellular carcinoma (HCC). HBV particles were identified in 1970 (Dane *et al.*, 1970) following the description of a viral antigen in the blood of an Australian aborigine (Blumberg *et al.*, 1965) which led to the development of highly effective blood screening programs. Despite the development of a vaccine for hepatitis B in the early 1980s, over 350 million people world-wide are thought to be chronically infected with HBV (WHO, 2011b).

HDV requires HBV surface antigen for infectivity and as such infection only occurs as a co-infection with HBV (Rizzetto *et al.*, 1980). Consequently, transmission and host specificity of HDV are assumed to be the same as HBV. As HDV is unrelated to HBV it is not considered a

defective-interfering particle, rather a satellite virus of HBV. HDV co-infection increases the severity and progression of hepatitis B disease (Gerin *et al.*, 2001).

A non-HAV, non-HBV agent causing acute, water-borne hepatitis was first suspected in the 1980s but it was not until 1991 that the first full-length HEV genome was cloned (Tam *et al.*, 1991). HEV has a higher rate of mortality than HAV, especially among pregnant women (Purcell & Emerson, 2008). Pigs and other animals act as animal reservoirs for HEV and zoonosis is thought to be responsible for some of sporadic HEV infections of unknown source in developed countries (Pavio *et al.*, 2010).

One of the closest related viruses to HCV is GB virus B (GBV-B). This virus has been shown to cause an acute, self-limiting hepatitis in tamarins (Karayiannis *et al.*, 1989; Schaluder *et al.*, 1995) prompting its use as an animal model for HCV. The pathogenicity in humans is unknown as is the natural host of GBV-B although a host range study concluded that it is likely to be a virus of New World monkeys (Bukh *et al.*, 2001).

1.2 Hepatitis C Virus

1.2.1 Discovery of HCV

After the generation of diagnostic tests for HAV and HBV in 1975 it became apparent that the vast majority of transfusion-associated hepatitis was not caused by either of these agents. Although associated with transfusions, this agent was prevalent in the general population and in particular among haemophiliacs and intravenous drug users. Initially termed non-A, non-B hepatitis (NANB or NANBH) the disease was thought to be caused by more than one agent (Bradley *et al.*, 1983; Shimizu *et al.*, 1979). Liver biopsies from experimentally infected chimpanzees revealed the presence of membranous tubules. This 'tubule-forming' NANBH agent was thought to be a small, enveloped virus due to its sensitivity to organic solvents (Bradley *et al.*, 1983) and ability to pass through an 80 nm filter (Bradley *et al.*, 1985).

Due to the low titre of this agent, 10^2-10^3 chimpanzee infectious doses per ml (CID/ml), the pooling of plasma from several chronically infected chimpanzees enabled the generation of a cDNA library using random primers. The cDNA molecules were cloned into an expression vector and screened for the expression of pathogen-specific peptides using sera from infected individuals. This technique identified 1 clone that was used as a specific hybridisation probe that hybridised with an approximately 10,000 nucleotide (nt) RNA molecule. Sequencing of this molecule revealed that it generated a single polyprotein of approximately 3,000 amino acids (aa) with homology to the non-structural proteins of members of the *Flaviviridae* family. This

new virus was renamed hepatitis C virus and speculatively named as a member of the *Flaviviridae* (Choo *et al.*, 1989; Choo *et al.*, 1990).

1.2.2 Epidemiology and transmission of HCV

Hepatitis C virus (HCV) can cause either an acute or chronic infection. Estimates of the number of people chronically infected with HCV have varied from 100-170 million people, or 2-3% of the global population (Lavanchy, 2009; Perz *et al.*, 2004). It is thought that a further 4 million new incidences of HCV infection are diagnosed every year (Mast *et al.*, 1999); however this figure may be out of date. HCV prevalence is determined from sero-conversion data but accurate quantification is hampered due to the often asymptomatic nature of acute HCV infections, poor screening and surveillance in many countries and sero-reversion, where chronically infected individuals can lose HCV-specific antibodies over time (Lavanchy, 2009). The global distribution of HCV varies greatly with areas of northern Africa, South America and the Far East showing significantly higher prevalence than Europe and North America (WHO, 2002) (Figure 1.1).

The most common route of HCV infection is parenteral, either by injection drug use or transfusion of infected blood products. The latter route is on the decline in most countries since screening of blood products for HCV became routine. Other rare routes have been identified such as sexual transmission; religious rites and body piercings; although often the route of transmission is not identified. While breast feeding does not appear to be a route of transmission; the spread of HCV perinatally has been demonstrated (Ohto *et al.*, 1994) and there is no conclusive evidence that caesarean section avoids infection of the neonate. There are seven genotypes of HCV that also vary in their global distribution but also the route of transmission. Sub-types 1b and 2a are associated with blood transfusion while 1a and 3a are associated with intravenous drug use.

1.2.3 Acute HCV infection

It is thought that the majority of acute HCV infections go undetected and the natural course of acute infection is highly variable. Acute infection is characterised by an increase in serum levels of liver-produced aminotransferases to 20-30 times in excess of the normal range between 6 and 12 weeks after exposure. HCV genomic material – ribonucleic acid (RNA) - can be detected by polymerase chain reaction (PCR) several days after infection but it can take weeks (Hoofnagle, 1997). Sero-positivity is often achieved 8 weeks after initial exposure but it may occur after several months or not at all. Symptoms of acute HCV infection include malaise, nausea, vomiting and jaundice.

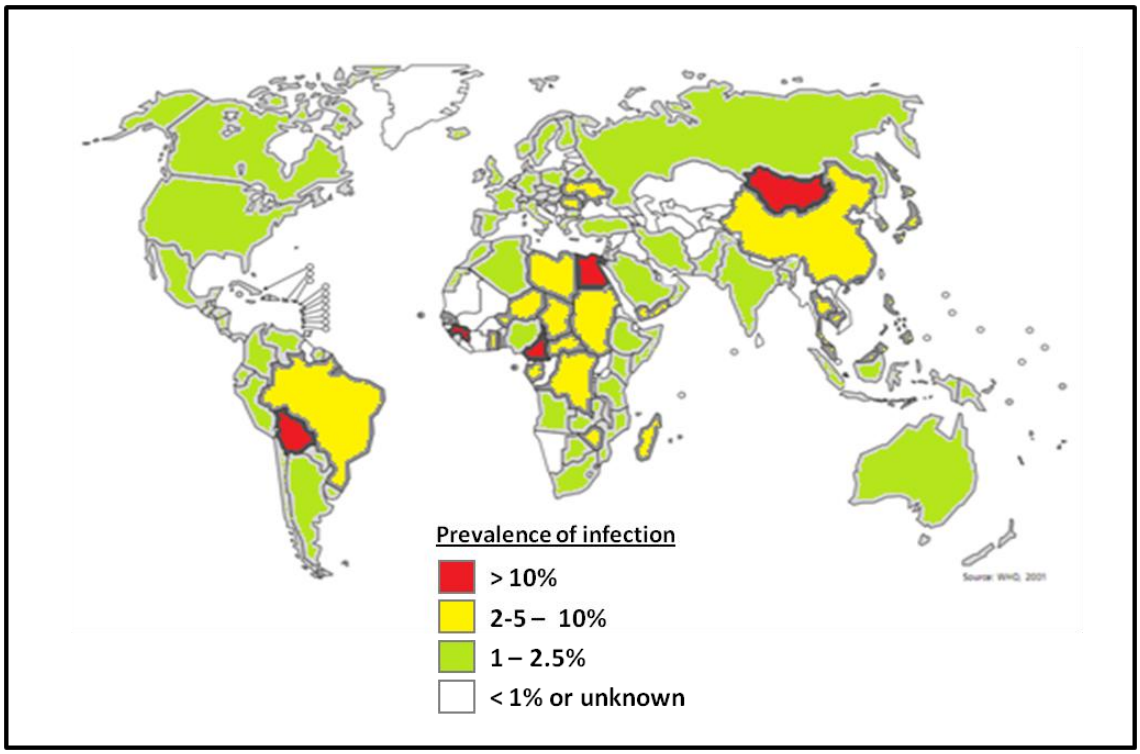


Figure 1.1 Global prevalence of hepatitis C virus.

HCV is most prevalent in less developed regions: South America, central and northern Africa and South East Asia. Reproduced from (WHO, 2002).

1.2.4 Chronic HCV infection

A key clinical feature of HCV is its ability to establish a persistent chronic infection. Clearance of the virus is thought to be achieved in as few as 20% of individuals and spontaneous clearance of chronic HCV is thought to be very rare (Yokosuka *et al.*, 1999). Chronic hepatitis C is defined as infection with HCV for more than 6 months as determined by detection of viral RNA. In the absence of liver cirrhosis most chronically infected individuals are asymptomatic or experience only mild symptoms such as fatigue (Merican *et al.*, 1993). Chronic hepatitis C can lead to liver damage but disease progression usually requires decades and the causes are not clear. Disease complications are almost entirely restricted to those with cirrhosis, as is the development of hepatocellular carcinoma (HCC). The risk of developing cirrhosis within 20 years of infection is estimated at 10-20% (Seeff, 2002). Numerous factors are known to influence disease progression such as age, gender, ethnicity, HBV co-infection and alcohol abuse. Viral load and genotype do not appear to be significant in the natural course of infection but are major factors in treatment outcome. However, a recent analysis of patient histories and disease progression identified infection with genotype 3 as an independent risk factor for the development of HCC (Nkontchou *et al.*, 2011).

1.2.5 Diagnosis

HCV infection is diagnosed either by serologic assay or nucleic acid testing. Serological diagnosis is determined by enzyme-linked immuno-sorbent assays (ELISA) that detect the presence of anti-HCV antibodies using regions of NS5A and NS3 as the bait. Detection of anti-HCV immunoglobulin M (IgM) may precede detection by other methods, however this is uncommon and it does not allow disease characterisation as an acute, chronic or cleared infection. False-positives are more common in patients with rheumatoid conditions and patients from low-prevalence groups such as blood or organ donors. False-negatives can occur more frequently in immuno-compromised patients. Core antigen-based ELISAs exist that are as effective and accurate as RNA-based detection methods at distinguishing between acute or chronic infections but their significantly reduced sensitivity compared to RNA-based detection methods preclude the use of core ELISAs as diagnostic tools (Morota *et al.*, 2009). The most powerful qualitative technique for detecting HCV infection is transcription-mediated amplification (TMA)-based RNA detection. TMA uses a combination of two primers and two enzymes (Figure 1.2). From the positive sense RNA genome of HCV a region is amplified as a double stranded (ds) DNA molecule by reverse transcriptase. The primer used to generate the positive strand of this DNA duplex contains a promoter region for RNA polymerase which

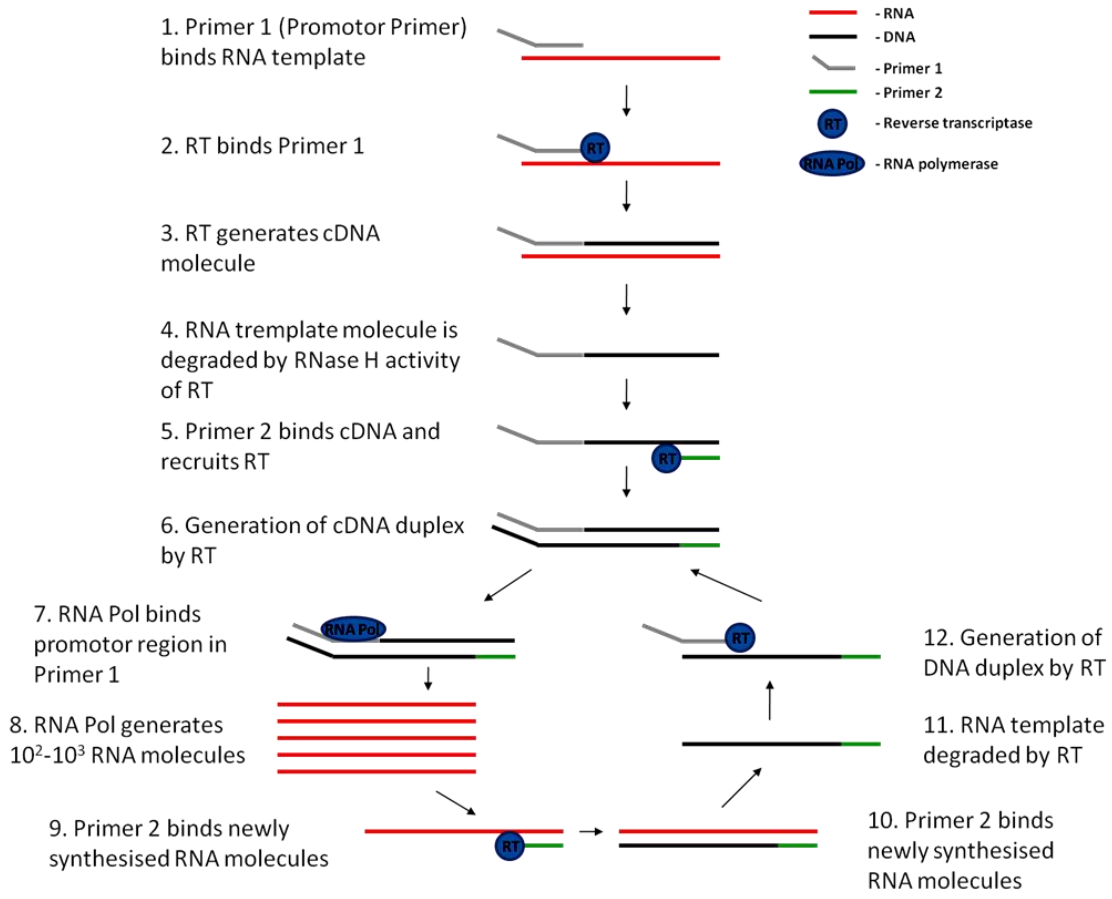


Figure 1.2 Schematic diagram of transcription-mediated amplification (TMA) polymerase chain reaction (PCR) used as a clinical test for HCV infection.

The diagnostic kit comprises two enzymes; reverse transcriptase (RT) and DNA-dependent RNA-polymerase (RNA Pol), and two primers; the 5' overhang of Primer 1 contains an RNA Pol promoter region.

generates further mRNA copies of the amplified region and the process is repeated. Attempts to standardise treatment of HCV led to treatment protocols based upon the monitoring of viral load. Viral load is measured in international units (IU), which equate to 1 IU = ~2.5 viral genome copies/ml, and is monitored using highly sensitive quantitative real-time reverse-transcriptase polymerase chain reaction (real-time RT-PCR) assays. Cross-centre and genotype variation is a major issue in diagnosis and treatment of HCV, despite calibration of assays to a World Health Organisation (WHO) standard (Chevaliez *et al.*, 2007).

1.2.6 Disease pathology

HCV is not thought to be cytopathic as virus load has no bearing on the severity of liver disease or prognosis. Therefore, similar to HBV, the disease pathophysiology of HCV infection, particularly cirrhosis, is thought to be resultant of the host immune-response or anti-viral therapy. The only cytotoxic effect attributed to HCV is steotosis (accumulation of lipid in hepatocytes) and while patients infected with all genotypes can develop steotosis it is only thought be virus-induced in genotype 3 infections (Kumar *et al.*, 2002). The particular mechanism by which genotype 3 viruses induce steotosis is not known, nor is it clear whether the virus benefits from this pathology or not.

Extra-hepatic manifestations (EHMs), such as rheumatic, dermatologic, endocrine and haematologic disorders or chronic fatigue, are often the only clinical signs of HCV infection. Hepatitis as a result of the immune response to viral infection leads to fibrosis. In the case of chronic infection fibrosis in turn leads to cirrhosis of the liver. While fibrosis is reversible, cirrhosis is rarely reversible even after viral clearance. Cirrhosis does not confer liver failure, rather it can be classified as compensated, when liver function is retained, or decompensated, when liver function is impaired. The development of HCC as a result of liver cirrhosis is a rare but significant complication associated with HCV infection (Saito *et al.*, 1990). HCV is the leading indication for liver transplantation in the UK and US and transplant is the only effective treatment once liver decompensation or HCC have developed (Hoofnagle, 1997).

1.2.7 Immune response to HCV

Effective clearance of HCV requires rapid responses from both the innate and adaptive arms of the immune response. Clearance of the virus does not confer immunity upon re-exposure but does reduce the risk of viral persistence upon re-infection (Bowen & Walker, 2005). The role of antibodies in viral clearance is unclear as neutralising antibodies have been identified yet sero-conversion does not prevent re-infection (Farci *et al.*, 1992; Lai *et al.*, 1994) and viral clearance has been observed without sero-conversion (Post *et al.*, 2004). Viraemia is used as a marker

for HCV infection and levels usually peak 6-10 weeks after infection, regardless of outcome (Abe *et al.*, 1992). Clearance is associated with rapid and persistent activation of CD4+ (helper) and CD8+ (cytotoxic) T-cells (Cooper *et al.*, 1999; Grakoui *et al.*, 2003). HCV-specific T-cells and expression of IFN-gamma within the liver of experimentally infected chimpanzees coincided with a reduction in HCV RNA levels (Cooper *et al.*, 1999; Major *et al.*, 2004). The innate immune response to viral infection is based upon interferons which are cytokines produced by cells in response to recognition of intracellular pathogens. HCV replicates via a ds RNA intermediate which is a stimulus for interferon production via the toll-like receptor (TLR) 3. It is hypothesised that upon infection of a cell HCV undergoes initial rounds of translation from its positive stranded RNA genome to produce sufficient protein prior to initiation of viral replication. This enables HCV to disrupt the interferon signalling pathway before it generates immunogenic ds RNA and thus avoid detection by the innate immune response (Li *et al.*, 2005a). Levels of IFN-β (Yang *et al.*, 2011) and IFN-λ (Marukian *et al.*, 2011) have been shown to be stimulated during HCV infection in model systems. However, it is not clear how these correlate with clearance of the virus and to what extent they are representative of natural infections in humans.

1.2.8 Treatment

Current treatment for Hepatitis C is combination anti-viral therapy using interferon-α (IFN-α) 2 and Ribavirin. There is no consistent difference in efficacy between IFN-α 2a or 2b (McHutchison *et al.*, 2007), although IFN-α 2b is cleared by the kidneys so is inappropriate for patients suffering renal complications. Achievement of a sustained virological response (SVR) is the goal of therapy as 99.9% of those attaining SVR remain negative for HCV RNA 5 years after treatment end (Swain *et al.*, 2007). The addition of polyethylene glycol to IFN-α 2a increased the half-life of the drug, permitting less frequent administration, and improved the rate of SVR (Wedemeyer *et al.*, 2002). 80/80/80 treatment adherence is a major factor in achieving SVR and non-adherence is an indicator for discontinuation (McHutchison *et al.*, 2002). Adherent patients are those who receive 80% of IFN-α, 80% of Ribavirin and are treated for more than 80% of the prescribed treatment period.

Treatment efficacy varies depending upon the virus genotype. This is reflected in the therapy durations which are 24 weeks for genotype 2 and 3 infections and typically 48 weeks for genotype 1 infections. The rate of SVR in genotype 2 and 3 infections is around 80% (Hadziyannis *et al.*, 2004) and treatment can be reduced to 12 or 16 weeks if a rapid virological response (RVR) is achieved. RVR is defined as the loss of detection of HCV RNA 4 weeks after

treatment initiation and correlates strongly with SVR. If HCV RNA has not decreased by 2 logs by week 12 treatment is discontinued.

The rate of SVR in genotype 1 infections is around 45% (Hadziyannis *et al.*, 2004) and efficacy is evaluated at weeks 4, 12 and 24. Undetectable HCV RNA at 4 weeks can justify a 24 week treatment program whereas a poor response at 4 weeks is an indicator for discontinuation. If a patient is negative for HCV RNA at 12 weeks a 48 week regimen is taken and if the patient becomes negative for HCV RNA between week 12 and 24 treatment can be extended to 72 weeks.

The production of interferon stimulates antiviral processes within the cell as well as shutting down protein synthesis and cellular proliferation and inducing a similar environment in neighbouring cells reducing the ability of the virus to spread. The mechanism of action of Ribavirin is unclear although it is has been speculated to be a competitive inhibitor of RNA replication (Crotty *et al.*, 2000) as it is a nucleoside analogue (Oxford, 1975), or as a modulator of cellular adenosine monophosphate (AMP)-activated protein kinase (AMPK) (Mankouri *et al.*, 2010). Ribavirin monotherapy is ineffective at reducing viral load.

Numerous side-effects are associated with combination therapy which lead to termination of treatment in up to ~20% of patients and interferon is contraindicated in a significant proportion of individuals, such as people with auto-immune diseases, as IFN-α is a potent stimulator of the immune system and has numerous side effects. The development of 'direct-acting antiviral agents' (DAA) or 'specifically targeted antiviral therapies for hepatitis C' (STAT-C) have predominantly focused on the viral polymerase and protease with an aim to producing better tolerated treatment with less side effects and thus increase treatment adherence. Clinical trials are underway examining the use of the protease inhibitors Boceprivir and Telaprevir, the nucleoside-analogue inhibitor Mericitabine and an inhibitor of the primary viral protein involved in host-cell modulation, non-structural protein 5A (NS5A). Work is also being carried out to develop novel interferons with less side-effects (Muir *et al.*, 2010) and inhibitors of cellular proteins that facilitate HCV replication (Flisiak *et al.*, 2009). Furthermore, combination therapy lacking IFN is also under investigation and a recent study examining the efficacy of triple therapy using a protease inhibitor, a polymerase inhibitor and Ribavirin in genotype 1-infected individuals showed high levels of SVR and none of the patients discontinued therapy prematurely (Zeuzem *et al.*, 2011).

1.2.9 HCV/human immunodeficiency virus (HIV) co-infection

Co-infection prevalence is linked to blood-borne transmission of human immunodeficiency virus (HIV) and in 2006 it was estimated that 4-5 million of the 33.4 million people infected worldwide with HIV were co-infected with HCV (Alter, 2006). Levels of HCV viraemia increase 8 times faster due to the lack of CD4+ T-cells however it has been suggested that up to 20-30% of newly infected people are still able to spontaneously resolve HCV infections despite underlying HIV infection (Vogel & Rockstroh, 2010). The chance of developing cirrhosis increases in co-infection, from 2-6% to 15-25% (Soto *et al.*, 1997), as does that of developing HCC (Giordano *et al.*, 2004). Highly active anti-retroviral therapy (HAART) seems to have positive effect on treatment of HCV however this is thought to be due to immune system recovery rather than direct anti-viral effects. Liver transplantation requires undetectable levels of HIV DNA or the assurance of effective control post-transplant.

1.2.10 HBV/HCV co-infection

Due to the lack of large scale population data the exact number of HBV/HCV co-infected individuals in unknown but as they share the same routes of transmission, co-infection is most common in populations where HBV infection is endemic. The viral interactions are complex and inconsistent but the basic trend appears to be that HCV is suppressive of HBV (Liaw, 2001). HCV super-infection is most common in HBV endemic populations but HBV super-infection remains largely unstudied. The incidences of cirrhosis and decompensated liver disease are more likely in co-infection (Mohamed Ael *et al.*, 1997) and HCC is twice as likely in co-infected patients as HCV mono-infected individuals (Chiaramonte *et al.*, 1999). There is currently no standardised treatment for HCV/HBV co-infection.

1.3 Molecular virology of HCV

1.3.1 Classification

HCV is a member of the *Flaviviridae* family which includes the *Flavivirus*, *Hepacivirus* and *Pestivirus* genera. A recent report called for the formation of a forth genus – *Pegivirus* – to include GBV-A, C and D (Stapleton *et al.*, 2011). HCV is the prototype member of the *Hepacivirus* genus. It shares greatest sequence homology with GB virus B (GBV-B), tentatively classed in the genus *Hepacivus*, bovine viral diarrhoea virus (BVDV), a *Pestivirus*, and the as yet unclassified canine hepacivirus (CHV) (Kapoor *et al.*, 2011). The *Flaviviruses* are vector borne pathogens.

1.3.2 Genetic heterogeneity

There are six recognised genotypes of HCV (1-6), with proposals for the formation of a seventh genotype in progress (Murphy *et al.*, 2007), which vary by ~25-35% at the nucleotide level, and numerous subtypes (a, b, c, etc) which vary by ~15-25%. The high level of sequence divergence in HCV is due to the high levels of RNA replication and the low fidelity of the virus encoded polymerase which has been estimated to be 10^{-4} mutations per nucleotide per year (Bartenschlager & Lohmann, 2000).

Within an infected individual subtype quasi-species, which differ by 1-5% at the nucleotide level (Davis, 1999; Honda *et al.*, 1994), often occur with the predominant quasi-species fluctuating over time. This 'sequence drift' is thought to aid in viral escape as the major areas of non-synonymous mutations occur in the ecto-domains of the glycoproteins. Patients infected with multiple subtypes also see shifts in the dominant subtype.

The global distribution of the genotypes is uneven (Figure 1.3), with genotype 1 the dominant global genotype worldwide. Genotypes 2 and 3 are dominant in eastern Africa and the Asian sub-continent respectively. Genotype 4 is localised to northern Africa, particularly Egypt, where it is the dominant genotype, and genotype 5 appears largely confined to and dominant in southern Africa. Genotype 6 is mainly restricted to the Far East (WHO, 2009).

1.3.3 Genome organisation

The genome of HCV is a single positive stranded RNA molecule acting as the transcription template for the viral proteome (Choo *et al.*, 1989). The genome comprises the protein coding region which is flanked by 5' and 3' non-translated regions (NTR) containing conserved RNA secondary structures (Figure 1.4A). The virus encodes 10 proteins (Grakoui *et al.*, 1993a; Lin *et al.*, 1994) expressed as a 3010 aa polyprotein from a single open reading frame (ORF) (Choo *et al.*, 1990). The N-terminal region of the polyprotein contains the structural proteins: core and the glycoproteins E1 and E2. The non-structural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A and NS5B make up the C-terminal region of the polyprotein. The NS proteins are involved in replication of the viral genome and are assumed to play varying roles in the assembly and secretion of virus particles. The p7 protein is encoded between the structural and non-structural proteins and its presence or absence within virions has yet to be confirmed. Translation is carried out by a cap-independent mechanism whereby the 40S ribosomal subunit is recruited, along with other cellular proteins, to the highly ordered RNA structures within the 5' UTR which forms an internal ribosomal entry site (IRES) (Fukushi *et al.*, 1994; Tsukiyama-Kohara *et al.*, 1992). The structural proteins are released by host signal peptidases

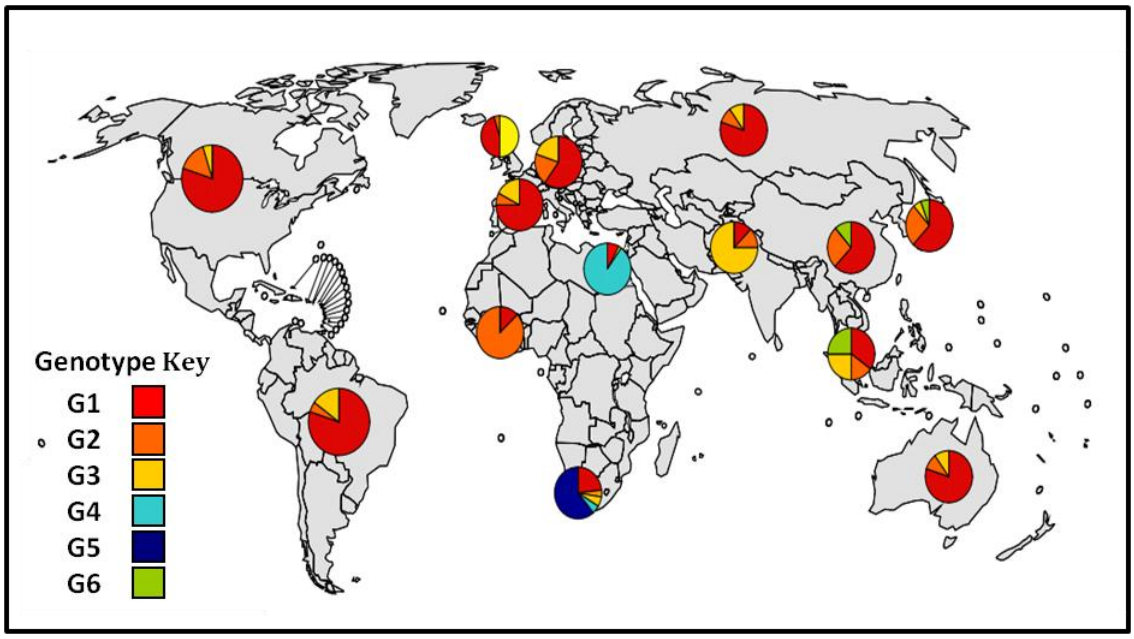


Figure 1.3 Global distribution of HCV genotypes.

Genotype 1 is the most prevalent genotype. Genotypes 2 and 3 are found to varying levels across the globe, but Genotypes 4, 5 and 6 are largely confined to northern Africa (particularly Egypt), South Africa and the Far East respectively. Reproduced from the Global Distribution of HCV Genotypes (WHO, 2009).

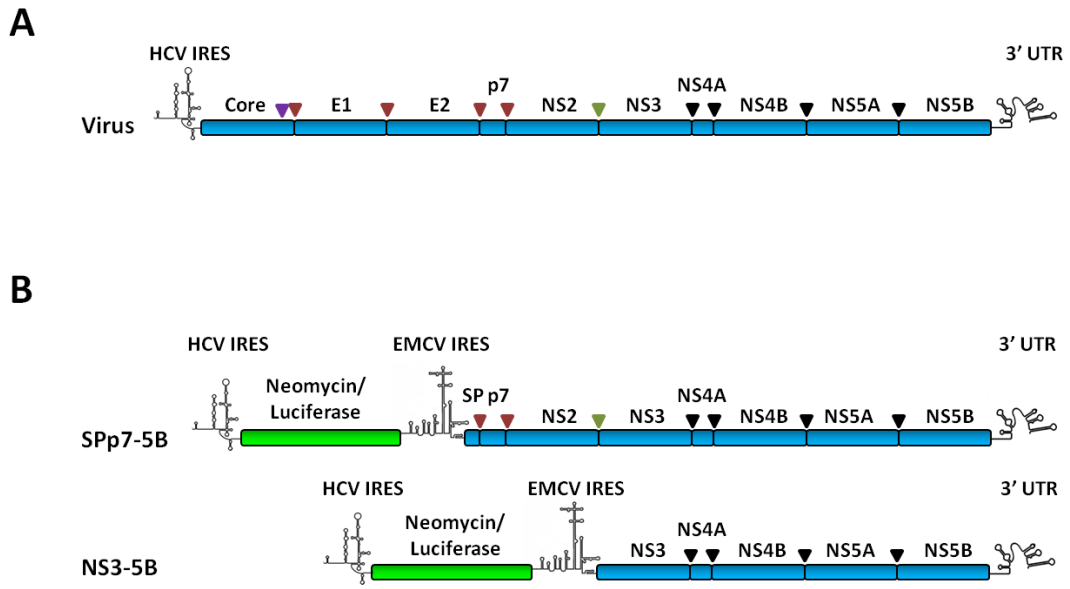


Figure 1.4 Genome organisation of HCV.

(A) Full length HCV genome depicting 5'UTR (HCV IRES), 3' UTR and core-NS5B coding region. (B) Structure of autonomously replicating sub-genomic RNA molecules. The HCV 5'UTR (HCV IRES) and 3' UTR are maintained. Translation of the reporter gene (neomycin or luciferase) is mediated by the HCV IRES. Translation of the truncated HCV genome is mediated by an encephalomyocarditis virus (EMCV) IRES. The EMCV IRES was used in favour of using a duplicate HCV IRES as translation driven by this IRES is more efficient. Arrow heads denote cleavage of the polyprotein by signal peptide peptidase (purple), signal peptidase (red), the NS2/3 cysteine protease (green) and the NS3/4A serine protease (black). SP represents the p7-cognate signal peptide encoded by the carboxyl (C)-terminal 23 residues of E2.

(Grakoui *et al.*, 1993a; Hijikata *et al.*, 1991) and the non-structural proteins by two viral proteases: a cysteine auto-protease; NS2 (Grakoui, 1993b; Lorenz *et al.*, 2006b; Schregel *et al.*, 2009) that catalyses cleavage of the NS2/NS3 junction, and a serine protease; NS3/4A (Tomei *et al.*, 1993) that cleaves to release the other NS proteins.

1.3.4 Virus Lifecycle

1.3.4.1 Virus Entry

The hepatropic nature of HCV was established early on in HCV history and more recently HCV has been identified in brain macrophages from infected individuals (Wilkinson *et al.*, 2009). Virus entry is mediated by the E1 and E2 glycoprotein hetero dimer (Figure 1.5). The cellular entry factors required for entry are cluster of differentiation 81 (CD81) (Pileri *et al.*, 1998), scavenger receptor B type I (SR-BI) (Scarselli *et al.*, 2002) and two tight junction-associated ligands, claudin-I (CLDN-I) (Evans *et al.*, 2007) and occludin (OCLN) (Ploss *et al.*, 2009). The virus is thought to encounter the baso-lateral membrane of hepatocytes in the liver where CD81 and CLDN-I are localised (Reynolds *et al.*, 2008). Polarisation of cell mono-layers in tissue culture systems is not required for entry but other factors such as glucosaminoglycans (GAG), the low density lipoprotein (LDL) receptor (LDL-R) and C-type lectins are important. Entry can be inhibited by mannan binding lectin (MBL), a component of the innate immune response (Brown *et al.*, 2010).

CD81 is a tetraspanin with a large, extracellular loop (LEL) that is bound by E2. CD81 is expressed on the surface of a wide variety of cell types and is associated with activation of the immune response by co-stimulation of T-cells. Antibodies specific for CD81 impair virus entry, although CD81 may be involved after initial virus binding (Koutsoudakis *et al.*, 2006). SR-BI is highly expressed on liver cells and is responsible for uptake and efflux of cholesterol from cells to high density lipoprotein (HDL) and LDL particles. The SR-BI ligand HDL also appears to play a role in HCV entry post-binding (Voisset *et al.*, 2005). The role of LDL-R is to internalise circulating LDL as part of the cholesterol homeostasis pathway. LDL-R complexes are internalised by clathrin-mediated endocytosis and traffic via endosomes (Sabahi, 2009) in a similar mechanism to that predicted for HCV (Thorley *et al.*, 2010). Antibodies against LDL-R inhibit HCV entry as do exogenous LDL and peptide inhibitors of LDL-R. GAG chains are important for the binding of numerous viruses. HCV E2 has been shown to bind the GAG heparan sulphate and pre-incubation of cells with a heparan sulphate homologue, heparin, or a heparinase, which cleaves heparin sulphate, impairs entry of HCV (Morikawa *et al.*, 2007).

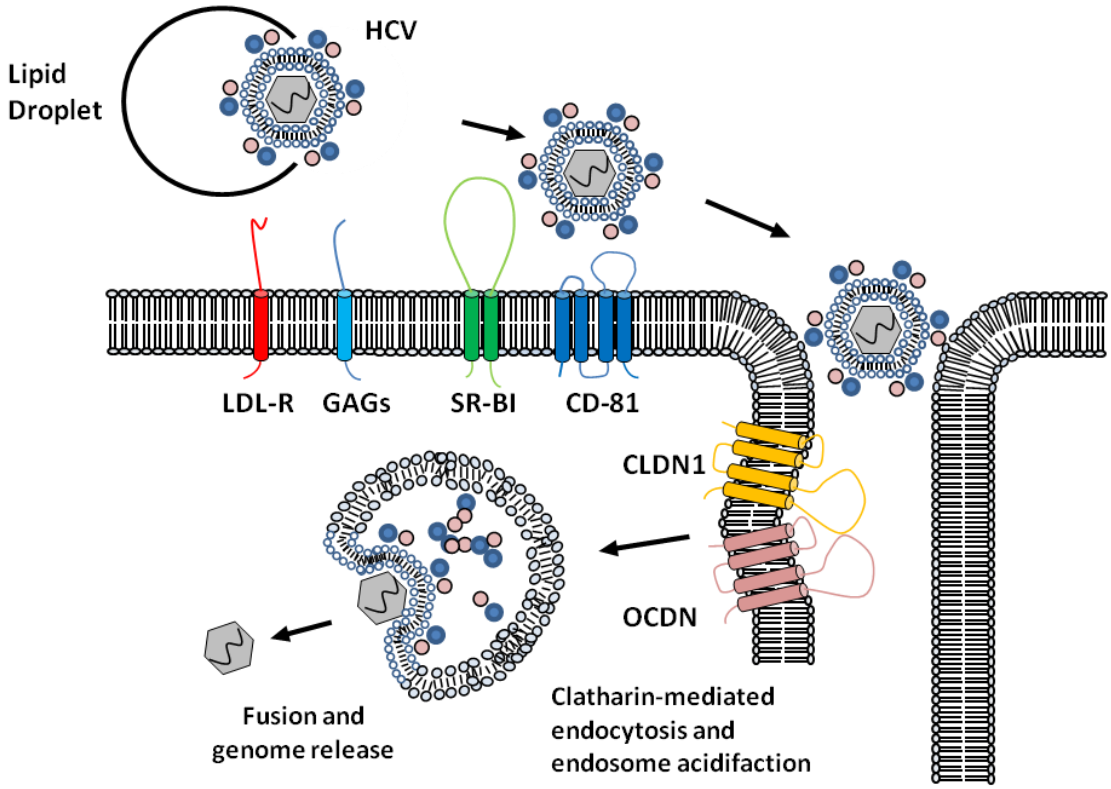


Figure 1.5 HCV particle entry is associated with lipoproteins.

Lipid droplet-associated HCV particles attach to cells via the LDL receptor and GAG molecules. The E2 glycoprotein then specifically binds to SR-B1 CD81. The virion is laterally trafficked to the tight-junctions where interactions with CLDN-1 and OCLN facilitate clathrin-mediated endocytosis. Endosome acidification induces fusion of the virion envelop resulting in genome release. Reproduced from (Eyre *et al.*, 2009).

Heparan sulphate appears to be involved in initial binding as incubation of cells with heparin after addition of virus did not significantly inhibit HCV entry (Koutsoudakis *et al.*, 2006).

The current model for HCV entry is outlined in Figure 1.5. Initial cell binding is thought to be mediated by interactions with surface GAGs, on the baso-lateral membrane of hepatocytes, followed by secondary interactions with SR-BI and CD81. The cell-associated particle is thought to traverse to tight junctions and associate with CLDN1 and OCLN prior to internalisation by clatharin-mediated endocytosis. Internal trafficking is thought to arrest at early endosomes where acidification results in fusion of the virus membrane with the endosome membrane allowing release of the viral RNA into the cytoplasm.

1.3.4.2 Genome Replication

After entry and uncoating the HCV genome is released into the cytoplasm and undergoes initial rounds of translation. The HCV proteins are expressed as a single polyprotein: core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B (Grakoui *et al.*, 1993a). An additional initiation site in the +1 reading frame, within the core coding region, produces a species known as frame-shift (F) protein, or alternate reading frame protein (ARFP). Initial translation of the viral polyprotein is thought to allow the accumulation of viral proteins to sufficiently disable the innate immune response and the formation of a membranous web structure called the replication complex (Egger *et al.*, 2002; Gosert *et al.*, 2003). Replication of the HCV genome is carried out by the virus-encoded RNA-dependent RNA-polymerase (RdRp) NS5B, via a dsRNA intermediate. Newly synthesised genomes are either packaged into virions or used as the template for protein synthesis or further rounds of replication. RNA levels, protein levels and infectivity do not directly correlate leading to the hypothesis that each process could be differentially prioritised throughout the life cycle, although the mechanisms involved remain unclear.

RNA structure is important to the pathogenesis of the virus and HCV RNA structure has been shown to be maintained as a survival factor. Slow growing, persistent RNA viruses like HCV and GBV-C correlate with high levels of genomic structure whereas acute viruses like BVDV have lower levels of RNA secondary structure. RNA secondary structure is thought to prevent the binding of host RNAs and thus allow the genomes to remain autonomous and less detectable to ds RNA sensing antiviral pathways (Davis *et al.*, 2008).

1.3.4.3 Particle Assembly and Egress

HCV has poor assembly efficiency in tissue culture and as a result it has been difficult to discern the mechanism by which infectious particles are produced. Core association with lipid

droplets is a key early step (Boulant *et al.*, 2006) but the site of assembly has not been identified. Core is able to bind RNA (Boulant *et al.*, 2005), but whether it dissociates from lipid droplets to begin packaging of the genome at sites of replication or whether NS5A, which also contains RNA binding domains (Huang *et al.*, 2005), recruits the genome to LDs is also unclear (Bartenschlager *et al.*, 2011). Numerous studies have provided evidence that HCV utilises the very low density lipoprotein (VLDL) biosynthesis pathway, however the evidence is conflicting as to the involvement of apolipoprotein (Apo) B, Apo E and microsomal transfer protein (MTP). Recruitment of the virus envelope likely occurs with the addition of E1/E2 hetero-dimers at the ER or an ER-derived membrane. Other viral proteins are vital for the production of infectious virus particles. NS2 and p7 were shown to act at an early stage in virus assembly (Jones *et al.*, 2007) possibly through the co-ordination of the replication complex with the glycoproteins (Jirasko *et al.*, 2010; Ma *et al.*, 2011; Stapleford & Lindenbach, 2011). The co-localisation of E2, NS2, NS3 and NS5A to cytosolic lipid droplets at 72 hours post-transfection with full-length virus suggested that assembly may be occurring at these sites (Jirasko *et al.*, 2010). The mechanism of virus egress from these sites is therefore unlikely be that of VLDL, which is synthesised at the ER membrane and secreted from the cell via the Golgi (Bamberger & Lane, 1990), rather it may utilise the endosomal trafficking pathway, aspects of which have been shown to be essential for virus release (Corless *et al.*, 2009).

1.4 Systems for studying HCV

1.4.1 Chimpanzees

Chimpanzees are the only animal other than humans that can be infected with HCV and their use enabled the demonstration of an infectious agent as the cause NANBH (Alter *et al.*, 1978; Tabor *et al.*, 1978). Indeed, ultra-structural studies on livers from infected chimpanzees led to the suggestion that two agents were responsible for NANBH (Shimizu *et al.*, 1979). The use of chimpanzees were also instrumental in the isolation and cloning of HCV (Choo *et al.*, 1989) and confirmation of the *in vivo* infectivity of culture derived virus produced by the JFH-1 isolate (Wakita *et al.*, 2005). Chimpanzees infected with the JFH-1 isolate were positive for viral RNA 3 days post-infection, however they presented no signs of hepatitis and viraemia was limited; approximately 10^3 copies/ml (Kato *et al.*, 2008; Wakita *et al.*, 2005).

There are differences in the disease course in chimpanzees and humans as disease is generally milder in chimpanzees (Boonstra *et al.*, 2009) and one longitudinal study found that the rate of chronic infection was much lower in chimpanzees (33%) than in humans (~80%) (Bassett *et al.*, 1998). As GBV-B and HCV share a relatively high degree of sequence homology it is hoped that

by studying both infections in animal systems that it will become apparent why one causes an acute infection and the other is so effective at achieving a persistent infection (Beames *et al.*, 2001). Despite the exploration of GBV-B infection of tamarins as a surrogate for HCV, chimpanzees are currently used for testing of anti-HCV compounds and potential vaccines (Carroll *et al.*, 2009; Chen *et al.*, 2007; Verstrepen *et al.*, 2011) and examination of the immune response and viral persistence (Kato *et al.*, 2008). Draw-backs of the use of chimpanzees include experimental cost, limited sample numbers, absence of disease-induced liver fibrosis and ethical concerns.

1.4.2 Non-primate animal models

As HCV does not infect any non-primates, attempts have been made to genetically engineer or humanise a strain of mice permissive to HCV infection. Progress has been made recently in the development of two such models (Dorner *et al.*, 2011; Washburn *et al.*, 2011). Earlier mouse models relied upon the engraftment of human hepatocytes into animals with severely depleted immune systems in order to achieve infection but as a result no liver damage was observed (Bissig *et al.*, 2010; Mercer *et al.*, 2001). It was shown that expression of human orthologues of CD81 and occludin in murine cells (Ploss *et al.*, 2009) and mice (Dorner *et al.*, 2011) was sufficient to render them permissive to HCV infection. However, viral proteins could not be detected in HCV infected mice expressing human CD81 and occludin, suggesting that virus replication, and possibly also translation, is inefficient in mouse hepatocytes (Dorner *et al.*, 2011). Mice engrafted with human liver and immune cells were permissive to HCV infection (Washburn *et al.*, 2011). These mice generated an HCV-specific immune response of preferentially expanded human T-cells; they developed hepatitis and fibrosis in response to HCV-derived peptide and could be re-stimulated upon re-challenge suggesting the establishment of memory T-cells (Washburn *et al.*, 2011). In combination these models could dramatically increase our understanding of virus entry, replication, host tropism and immune response/evasion.

1.4.3 Replicons

The development of the HCV sub-genomic replicon (SGR) in 1999 enabled dissection of HCV replication and the roles of the non-structural proteins in replication and host cell modulation (Lohmann *et al.*, 1999) (Figure 1.4B). The original SGR was a bi-cistronic RNA molecule and was first developed from the genotype 1b isolate Con1. Expression of a 5' neomycin phosphotransferase (NPT) gene is controlled by the viral IRES and translation the non-structural proteins NS3-5B is mediated by an EMCV IRES (Lohmann *et al.*, 1999). The SGR originally included NS2 but it was found to be detrimental to replication (Lohmann *et al.*,

1999). Culture adaptations in the original SGR were noted and upon introduction into the wild-type SGR they increased replication by 10^5 (Blight *et al.*, 2000; Krieger *et al.*, 2001; Lohmann *et al.*, 1999). The most commonly used laboratory adapted SGR is termed FK5.1 which contains 2 mutations in NS3, one in NS4B and four in NS5A (Krieger *et al.*, 2001). Replication of the SGRs was initially restricted to the human hepatocellular carcinoma cell line Huh7. Treatment of Huh7 cells stably harbouring a genotype 1b replicon with IFN-α led to the removal of HCV RNA; however these cured cells, Huh7.5 cells, were then more permissive to replication upon re-introduction of HCV RNA (Blight *et al.*, 2002). Full-length replicons were generated but they failed to produce infectious virus and replicated less efficiently (Blight *et al.*, 2002; Pietschmann *et al.*, 2002). Replacement of the resistance gene with a Firefly luciferase reporter gene created a system for analysis of transient replication (Lohmann *et al.*, 2001) enabling the high-throughput analysis of anti-viral compounds. Other replicon systems include NS3-5B, NS2-5B and SPp7-5B replicons based on the genotype 1b isolate Con1, corresponding replicons made using the genotype 2a isolate JFH-1 (Tedbury *et al.*, 2011), which doesn't require culture adaptations (Kato *et al.*, 2003), and a mono-cistronic NS3-5B replicon employing a green fluorescence protein-tagged NS5A (Schaller *et al.*, 2007).

1.4.4 Infectious HCV culture systems

A major step forward in the study of HCV came with the identification of the fully infectious JFH-1 isolate which is capable of producing infectious virus in cell culture and chimpanzees (Wakita *et al.*, 2005; Zhong *et al.*, 2005). Further analysis revealed that chimeric viruses based on the JFH-1 isolate were capable of causing infection both in cell culture and *in vivo* (Lindenbach *et al*., 2005; Lindenbach *et al*., 2006; Pietschmann *et al*., 2006). Infectivity of cell culture-derived HCV can be neutralised with E2 and CD81-specific antibodies as well as patient sera (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005; Yi *et al.*, 2006; Zhong *et al.*, 2005).

Although the majority of infectious clones have been generated in the JFH-1 background, a fully infectious genotype 1a clone, derived from the Hutchinson isolate H77, has been developed although it is significantly less infectious than JFH-1 based viruses (Yi *et al.*, 2006). Chimeric viruses involving the structural proteins from non-genotype 2a isolates can require compensatory mutations for efficient infectivity (Yi *et al.*, 2007). Chimeric viruses, particularly those including the structural proteins from the most prevalent genotype 1a and 1b isolates, could play a central role in the development of cross-genotype therapies targeting virus entry and uncoating (Pietschmann *et al.*, 2006). The unique ability of the JFH-1 isolate to replicate in cell culture has recently been proposed to centre on the replication efficiency of the viral polymerase (Simister *et al.*, 2009). Comparative analysis of the polymerase encoded by the

JFH-1 isolate and the closely related J6 isolate demonstrated that the JFH-1 polymerase initiates replication more efficiently resulting in increased replication *in vitro*. Mutations within the polymerase were shown to result in a more closed conformation which may improve de novo initiation kinetics (Simister *et al.*, 2009). Whether the increased replication efficiency of the polymerase is solely responsible for enabling detectable replication of JFH-1 in cell culture is not clear.

1.5 Untranslated regions (UTRs)

1.5.1 5' UTR

The HCV 5' UTR contains a highly conserved internal ribosomal entry site (IRES) (Tsukiyama-Kohara *et al.*, 1992). The 5' UTR is 341 nucleotides in length and the IRES encompasses nucleotides 40-341 of the genome (Honda *et al.*, 1999) (Figure 1.6A). The 5' UTR forms 4 stem loop structures (I, II, III, IV), the first of which, along with stem loop II, is essential for replication (Kim *et al.*, 2002). A predicted structure for loop II was proposed from sequence comparison with the GBV-B 5' UTR structure (Brown *et al.*, 1992; Honda *et al.*, 1996) and was refined to one containing more base-pairing following application of folding constraints based on the GBV-B structure (Honda *et al.*, 1999). HCV RNA replication is enhanced by a liver specific micro RNA 122 (miR-122) which has two seed-matches between nts 22 and 43 of the HCV 5' UTR (Jopling *et al.*, 2005), both of which are required for optimal HCV replication (Jopling *et al.*, 2008). The tissue specificity of miR-122 could be a crucial determinant in the tissue and host tropism of HCV. A recent report where miR-122 was exogenously expressed in the non-permissive HepG2 cell-line rendered cells permissive to HCV replication (Narbus *et al.*, 2011).

Loops II, III and IV form the IRES and, under physiological conditions, form two recognition domains for eukaryote initiation factor 3 (eIF3) and the 40S ribosomal subunit (Kieft *et al.*, 1999) to facilitate cap-independent translation of the HCV polyprotein. The first ~40 nts of the 5' UTR, which contain stem loop I, are not required for initiation of translation (Fukushi *et al.*, 1994) but translation does benefit from the 5' 30 nts of the core coding sequence (Reynolds *et al.*, 1995).

A long-range, protein independent interaction between the 5' and 3' UTRs has been implicated as the switching mechanism between RNA replication and polyprotein translation (Romero-Lopez & Berzal-Herranz, 2009).

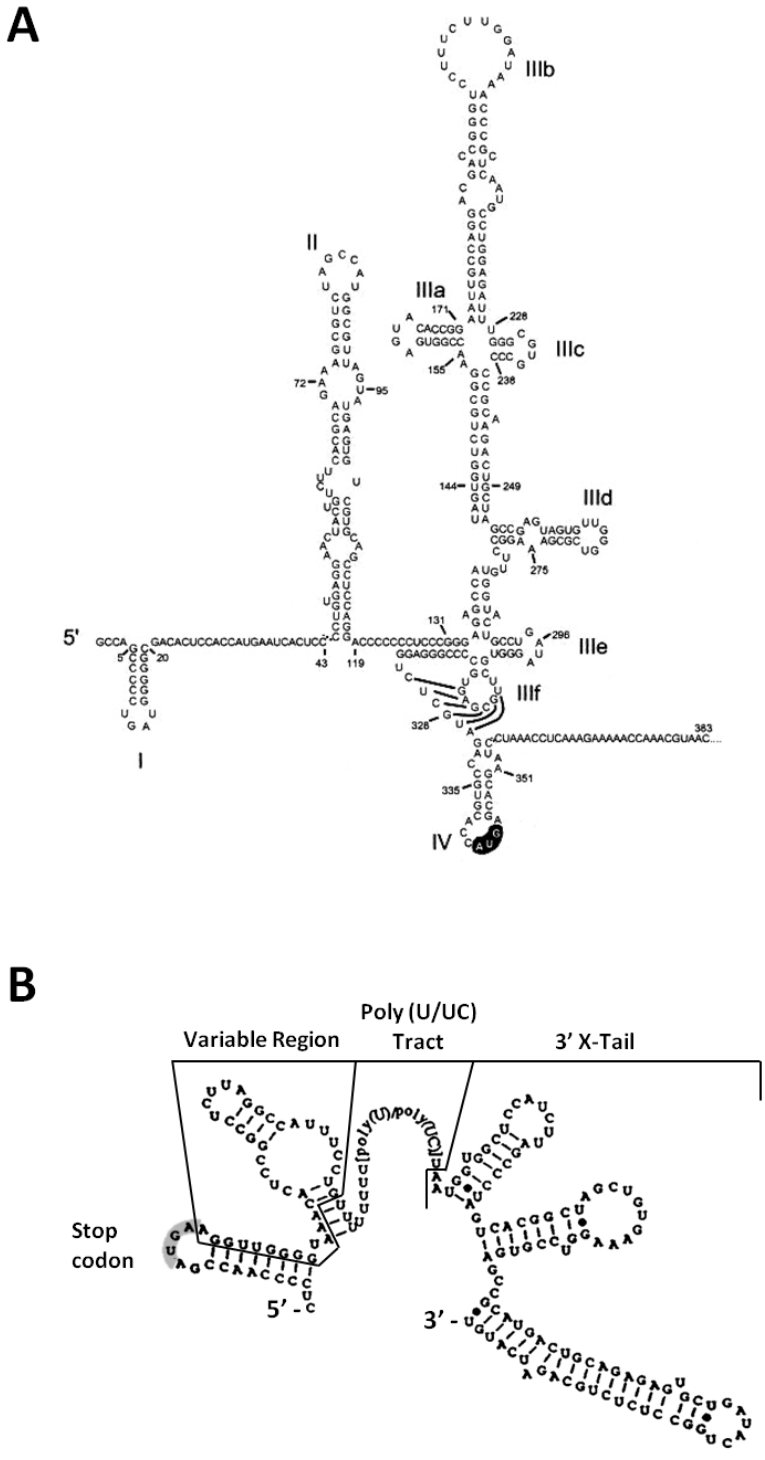


Figure 1.6 The predicted structures of the HCV UTRs.

(A) The 5'UTR forms 4 stem-loop structures. Loops I and II are required for RNA replication, loops II, III and IV are required for recruitment of the 35S ribosome and eIF3. The start codon for translation of the primary HCV ORF is highlighted in black. Reproduced from (Honda *et al.*, 1996). (B) Structure of the 3' UTR. The three major structural components are marked and the stop codon is highlighted. Reproduced from (Kolykhalov *et al.*, 1996).

Two chimeric viruses encoding the core-NS2 and 5' UTR-NS2 from the J6 isolate in the JFH-1 background had comparable replication and translation, but the J6 5' UTR-NS2 virus produced 10-fold more infectious virus (Mateu *et al.*, 2008). The JFH-1 and J6 5' UTRs differ by three nts and it was found that increased infectivity required the mutation of all three bases (Mateu *et al.*, 2008). This was suggestive that the 5' UTR plays a role in particle assembly, as well as replication and translation, making it a viable target for antiviral compounds (Guerniou *et al.*, 2007).

1.5.2 3' UTR

The 3' UTR extends for 236 nts following the translation termination site at the 3' end of the NS5B coding sequence (Figure 1.6B). It comprises a variable region of ~40 nts; a polyuridine and polypyrimidine [poly (U/UC)] tract of variable length; and a highly conserved region called the 3' X-tail which is 98 nts in length (Kolykhalov *et al.*, 1996; Tanaka *et al.*, 1995; Tanaka *et al.*, 1996). The X-tail forms three stem loops of conserved structure (Blight & Rice, 1997; Ito & Lai, 1997) and has been shown to enhance translation in *cis* but not in *trans* (Ito *et al.*, 1998). Numerous proteins, such as polypyrimidine tract-binding protein (PTB) which can bind both the 5' and 3' UTR, have been shown to interact with the 3'UTR of HCV and enhance translation (Ito & Lai, 1999; Spangberg *et al.*, 1999; Tsuchihara *et al.*, 1997). However, a contradictory report demonstrated that replication was independent of PTB and that deletion of the X-tail or the poly (U/UC) tract enhanced replication (Murakami *et al.*, 2001).

The 3' terminus is thought to end in a dsRNA stem loop (Tanaka *et al.*, 1996), but one cDNA clone was found to extend for a further two uridine (U) bases forming a single stranded terminus (Yamada *et al.*, 1996).

1.6 Structural proteins

1.6.1 Core

The nucleocapsid or core protein is the amino (N)-terminal protein on the nascent polyprotein (Figure 1.7). An internal signal sequence encoded between core and E1 is responsible for directing translation of the developing polyprotein to the ER membrane. Cleavage at the N-terminus of E1 by signal peptidase releases an immature core protein of 191 aa and subsequent intra-membrane cleavage by signal peptide peptidase (SPP) releases the mature core protein of 179 aa (McLauchlan *et al.*, 2002). Production of the mature core and its association with lipid droplets are essential for virus particle assembly (Boulant *et al.*, 2007; Targett-Adams *et al.*, 2008). Core remains membrane associated by virtue of 2 amphipathic

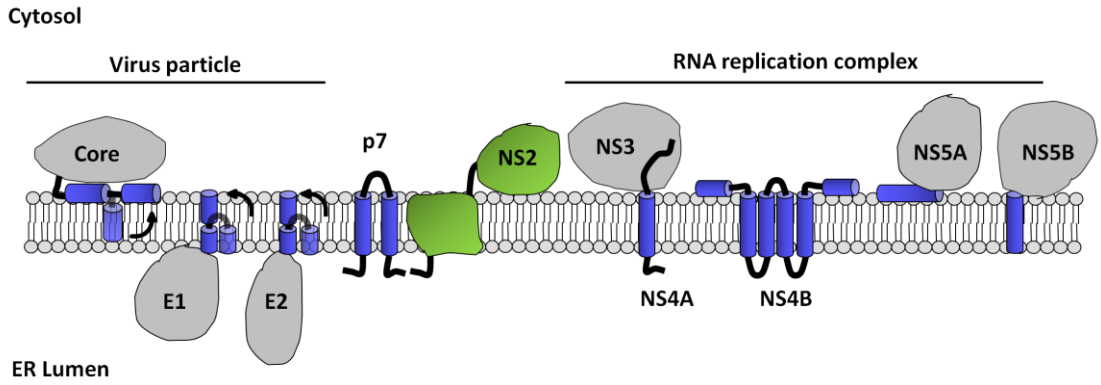


Figure 1.7 Cartoon depicting the membrane interactions of the HCV proteins about the ER membrane.

NS2 is shown in green and the membrane-spanning domain is undefined. Following release from E2 by signal peptidase, the C-terminal domain of core is further processed within the membrane to enable flipping of the C-terminus to the cytosol. The *trans*-membrane domains (TMDs) of E1 and E2 initially form short hairpin structures to act as signal peptides for E2 and p7 respectively. Following cleavage by signal peptidase the E1 and E2 TMDs re-orient their C-termini to the cytosol to generate single pass, tail-anchored membrane topologies. NS5B also forms a tail-anchored protein topology. Core and NS5A remain membrane-associated by virtue of their amphipathic helices. The N- and C-terminal amphipathic helices of NS4B are shown.

α-helices (AH) between residues 119-136 and 148-164 (Boulant *et al.*, 2006). Core is also able to target to the outer membrane of mitochondria via a 10 aa stretch within the C-terminal AH (Schwer *et al.*, 2004). Biochemical analysis suggested that this 10 aa sequence in isolation acted like an AH as GFP expressed as a fusion to this domain could be extracted from purified mitochondria with alkaline carbonate along with the amphitropic mitochondrial heat-shock protein 70 (mtHSP70) (Schwer *et al.*, 2004).

Core has two distinct domains; an N-terminal RNA-binding domain (Boulant *et al.*, 2005; Santolini *et al.*, 1994) and a C-terminal domain responsible for targeting core to lipid droplets (Boulant *et al.*, 2006; Hope & McLauchlan, 2000). Core protein is thought to homo-oligomerise to form the virus nucleocapsid encapsulating the viral genome. It has been proposed that the nucleocapsid is enveloped in a host endoplasmic reticulum (ER)-derived lipid bi-layer, the surface of which is believed to be punctuated by the virally encoded glycoproteins E1 and E2 (Moradpour *et al.*, 2007) although the extent to which they are exposed is the subject of controversy. Core is localised predominantly to the cytosolic face of the ER and it remains ER-associated if cleavage by SPP is prevented (McLauchlan *et al.*, 2002). Core localises to lipid droplets (Hope & McLauchlan, 2000; Santolini *et al.*, 1994) which are increased in abundance in HCV-infected cells. This is thought to be caused by core protein as over-expression studies found that core induced lipid droplet accumulation in tissue culture (Abid *et al.*, 2005). Genotype 3a core had the greatest effect prompting speculation that core could be responsible for the observed virus-induced steotosis in genotype 3a infections (Abid *et al.*, 2005). Core is thought to alter the cellular localisation of lipid droplets to a perinuclear distribution as it disrupts the association of adipocyte differentiation-related protein (ADRP), the major protein content of lipids, with lipid droplets (Boulant *et al.*, 2008). More recently the host proteins hnRNPH1, NF45, and C14orf166 have been identified as interacting partners of core protein (Lee *et al.*, 2011) and core has also been shown to form an interaction with both UTRs in cells (Yu *et al.*, 2009).

1.6.2 Frameshift protein (F)

Translation of the HCV polyprotein occurs from a single ORF, although it had previously been noted that mutation of the primary AUG resulted in translation initiation from a downstream AUG (Reynolds *et al.*, 1995). Subsequently, several reports noted the translation of a second ORF within the core coding region in the +1/-2 frame (Walewski *et al.*, 2001; Xu *et al.*, 2001). Expression of frame-shift (F) protein or alternate reading frame protein (ARFP) is independent of cellular and viral proteases and the gene product varies in length between 126 aa and 162

aa depending upon the genotype. This variation has led to speculation of the authenticity of this gene product during natural infection.

Antibodies to F protein have been detected in sera from chronically (Shesheer Kumar *et al.*, 2008; Walewski *et al.*, 2001) and acutely infected HCV patients (Morice *et al.*, 2009). A cell-mediated immune response to F protein was not observed in a selection of acutely infected individuals (Drouin *et al.*, 2009) but F protein-specific CD4+ T-cell responses were detected in transgenic mice (Gao *et al.*, 2010). F protein appears to be unnecessary for virus infection in cell culture but a recent study found high levels antibodies to F protein in patients with HCC suggesting a role of F protein in HCV-related pathogenesis (Dalagiorgou *et al.*, 2011).

1.6.3 E1 and E2

E1 and E2 are the viral envelop glycoproteins each consisting of a single C-terminal TMD, that anchors them to the ER membrane (Cocquerel *et al.*, 1999; Cocquerel *et al.*, 1998; Selby *et al.*, 1994), and an N-terminal ectodomain which is oriented to the ER lumen (Cocquerel *et al.*, 2000; Santolini *et al.*, 1994). E1 and E2 contain 6 and 11 predicted *N*-linked glycosylation sites and have observable masses of 31 kDa and 70 kDa by electrophoresis, respectively (Grakoui *et al.*, 1993a; Stapleford & Lindenbach, 2011). The correct folding of both E1 and E2 requires the formation of numerous di-sulphide bonds and the ER chaperones calreticulin and calnexin (Choukhi *et al.*, 1998; Dubuisson & Rice, 1996). The folded glycoproteins form non-covalent hetero-dimers via interactions of their TMDs (Deleersnyder *et al.*, 1997; Op De Beeck *et al.*, 2000; Selby *et al.*, 1994).

The TMDs comprise two hydrophobic sections separated by at least one charged residue (Ciczora *et al.*, 2007; Cocquerel *et al.*, 2002; Cocquerel *et al.*, 2000) and dimerisation is severely reduced following mutation of these residues (Ciczora *et al.*, 2007; Stapleford & Lindenbach, 2011). These charged residues are also essential for correct polyprotein processing as the C-terminal portion of each TMD forms a signal sequence for the downstream protein and enables cleavage by signal peptidase (Cocquerel *et al.*, 2002) (Figure 1.6). Both TMDs are understood to form intra-membrane hairpins, with the charged residues acting as the pivot point, with the C-terminus oriented to the ER lumen prior to signal peptidase cleavage. Once released from the downstream protein the signal sequence then inverts to form a single α–helical membrane-spanning domain (Op De Beeck *et al.*, 2000; Penin *et al.*, 2004).

The glycoproteins are thought to be exposed on the surface of the virus particle (Wakita *et al.*, 2005) to facilitate binding and fusion of the virus to cell surface membranes. Experiments using the soluble ectodomain of E2 have shown that it interacts with CD81 (Pileri *et al.*, 1998), SR-BI

(Scarselli *et al.*, 2002) and claudin I (Evans *et al.*, 2007) while E1 coordinates the interaction with the LDL-R (Mazumdar *et al.*, 2011). E1 has also been shown to interact with core, presumably within the ER membrane (Lo *et al.*, 1996), but it has not been determined whether this interaction is required for particle assembly and maintained within mature virions. Antibodies targeting the E2 glycoprotein or the cellular entry factors prevent virus entry.

1.6.4 p7

The p7 protein is an integral membrane protein (Carrere-Kremer *et al.*, 2002) and comprises two TMDs and a conserved charged basic-loop, with both termini oriented to the ER lumen (Carrere-Kremer *et al.*, 2002; Lin *et al.*, 1994) (Figure 1.7). A dual topology of p7 has been proposed whereby the C-terminal helix does not span the membrane, instead it lies parallel to the membrane on the cytosolic face (Isherwood & Patel, 2005). The C-terminal helix of p7 is capable of acting like a signal peptide (SP) and so appears to form the SP sequence for NS2 (Carrere-Kremer *et al.*, 2002). p7 is released from the viral polyprotein by cleavage at both termini by signal peptidase and mutation of the C-terminal alanine to arginine abrogates p7-NS2 cleavage (Cocquerel *et al.*, 2002; Mizushima, 1994).

p7 is 63 residues in length and it has been shown to homo-oligomerise to form a hexameric or heptomeric ion channel, leading to its classification as a viroporin (Clarke *et al.*, 2006; Luik *et al.*, 2009). The inclusion of p7 within virus particles has not been established but it is vital for infection *in vitro* (Jones *et al.*, 2007) and in chimpanzees (Sakai et al., 2003). The expression of core, E1, E2 and p7 in *cis* in insect cells led to the formation of virus-like particles (VLPs) with uncleaved E2-p7 detectable on the surface of these particles (Isherwood & Patel, 2005). However, cleavage of the E2-p7 junction is severely impaired in insect cells compared to mammalian cells (Isherwood & Patel, 2005).

p7 ion channel activity can be chemically inhibited by amantadine but this is genotype dependent (Griffin *et al.*, 2008; StGelais *et al.*, 2007). Ion-channel activity can also be inhibited in a genotype-independent manner by rimantadine and iminosugars such as N-nonyldeoxynojirimycin (NN-DNJ) (Gottwein *et al.*, 2011; Griffin *et al.*, 2008; Pavlovic *et al.*, 2003; StGelais *et al.*, 2007; Wozniak *et al.*, 2010).

Compounds that inhibit ion channel activity also block virion release, implicating ion-channel function in virion morphogenesis (Griffin, 2010; Griffin *et al.*, 2008; Luik *et al.*, 2009). Mutation of the two basic residues in the cytoplasmic loop, termed the p7 basic-loop, block ion-channel function (Griffin *et al.*, 2004; StGelais *et al.*, 2007) and severely reduce infectivity (Jones *et al.*, 2007; Ma *et al.*, 2011; Steinmann *et al.*, 2007). Deletion of p7 from the full-length JFH-1 virus

(Jones *et al.*, 2007; Steinmann *et al.*, 2007) or chimeric viruses rendered them uninfectious (Popescu *et al.*, 2011). p7 basic-loop mutants could be rescued by the vATPase inhibitor bafilomycin A or exogenous expression of influenza M2 but this wasn't the case for p7 deletion viruses, further implicating a channel-dependent function of p7 in the virus life cycle (Brohm *et al.*, 2009; Wozniak *et al.*, 2010). The inclusion of p7 within the virus particle remains an area of investigation.

Analysis of p7 function using liposomes revealed that the basic-loop mutant was less efficient at inserting into membranes (StGelais *et al.*, 2007). The additional mutant histidine (H) 17 alanine (A) and glycine (G) 39A both reduced ion-channel function but did not alter membrane insertion. The effect of G39A on ion-channel function was more profound than the H17A mutant and the G39A. The H17A mutation could be compensated for by increased concentrations of p7 but the G39A mutant could not. This was suggestive of a more significant effect on channel structure (StGelais *et al.*, 2007).

1.7 Non-structural proteins

1.7.1 NS2

NS2 is discussed in detail in section 1.8.

1.7.2 NS3

NS3 is a multi-functional protein and is the only viral protein whose maturation is independent of membranes (Errington *et al.*, 1999; Svitkin *et al.*, 2005). The N-terminal 189 residues of NS3 form a chymotrypsin-like serine protease. Histidine 57, aspartic acid (D) 81 and serine (S) 139 form the catalytic triad. A zinc ion is co-ordinated by 3 cysteine (C) residues and a histidine (C97, C99, C145 and H149) which is thought to be of structural, not catalytic, importance (De Francesco *et al.*, 1996; Stempniak *et al.*, 1997). The NS3 protease *cis*-cleaves at the NS3/4A junction and, in association with NS4A, *trans*-cleaves the subsequent downstream viral protein junctions (Eckart *et al.*, 1993; Failla *et al.*, 1994; Grakoui *et al.*, 1993a; Hijikata *et al.*, 1993a; Hijikata *et al.*, 1993b; Tomei *et al.*, 1993). As mentioned earlier, there are numerous antiviral compounds targeting NS3/4A protease activity currently in clinical trials. The serine protease domain is also required for efficient cleavage of the NS2/3 junction (Thibeault *et al.*, 2001) and zinc co-ordination by NS3 is thought to play a structural role here as well (Tedbury & Harris, 2007). NS2/3 cleavage is discussed in greater detail in section 1.8.2.

The C-terminal two thirds of NS3 possess helicase and NTPase activities (Kim *et al.*, 1995) but there is no dependence between the helicase and protease activities (Gallinari *et al.*, 1998).

Helicase activity progresses in a 3' to 5' direction (Tai *et al.*, 1996) and it has been shown to unwind dsRNA and RNA-DNA hetero-duplexes. Helicase activity is dependent upon dimerisation of NS3 (Serebrov & Pyle, 2004) and activity is ATP and divalent cation-dependent (Frick *et al.*, 2007; Suzich *et al.*, 1993; Wardell *et al.*, 1999). NS3 helicase and ATPase activities are unaffected by the presence of uncleaved NS2 (Welbourn *et al.*, 2005).

NS4A forms a β-strand in complex within the NS3 β-barrel which is essential to induce a conformational change in NS3 (Kim *et al.*, 1995; Yan *et al.*, 1998). The NS3/4A protease has also been shown to inhibit antiviral response pathways by specifically cleaving the cellular proteins mitochondrial antiviral signalling protein (MAVS) (Li *et al.*, 2005b) and Toll/IL-1 receptor (TIR)-domain-containing adapter-inducing interferon-β (TRIF) (Li *et al.*, 2005b).

The enhancing mutation of glutamine (Q) 221 to leucine (L) (Q221L) within the helicase domain was originally described in the H77S/JFH-1 chimeric virus and confers high levels of infectivity to the wt chimeric virus (Ma *et al.*, 2008; Yi *et al.*, 2006) as well as increasing infectivity of the wt JFH-1 virus (Ma *et al.*, 2008). It was subsequently found to rescue deleterious mutations within NS2 in other chimeric viruses (Jirasko *et al.*, 2010; Phan *et al.*, 2009). Introduction of the Q221L mutation into the Jc1 chimeric virus, which achieves improved titres over full-length JFH-1, did not produce and additive effect on virus titres (Phan *et al.*, 2009). The role of this residue is unlikely to be a direct interaction with specific regions of NS2 as a large number of widely distributed NS2 mutations were suppressed by this NS3 mutation. The mechanism by which this residue confers enhanced infectivity is unclear as two other mutations in NS3 (Isoleucine (I) 286 valine (V) and I399V) were able to functionally replace the Q221L mutation in the H77S/JFH-1 virus (Ma *et al.*, 2008). Comprehensive analysis of the NS3 enzyme functions found little difference between wt and the Q221L mutant except that the mutant showed a decrease in functional RNA binding and a minor decrease in RNA unwinding (Phan *et al.*, 2009). Unfortunately, the phenotypes of the I286V and I399V mutants were not examined in this study.

1.7.3 NS4A

NS4A is an 8 kDa (Grakoui *et al.*, 1993a) integral membrane protein that interacts non-covalently with NS3 to anchor it to membranes via its C-terminal TMD (Brass *et al.*, 2008; Hijikata *et al.*, 1993b; Kim *et al.*, 1999; Wolk *et al.*, 2000) (Figure 1.7). NS4A associates with NS3 forming an integral β-strand as part of the N-terminal protease domain of NS3 and acts a co-factor for the NS3 serine protease (Failla *et al.*, 1994; Kim *et al.*, 1999; Kim *et al.*, 1996). Truncation analysis identified residues 21-35 of NS4A as the minimum required to stimulate

the release of NS4B, NS5A and NS5B *in vitro* and in tissue culture (Lin *et al.*, 1995; Tomei *et al.*, 1996). Binding of NS4A was initially thought be unnecessary for NS3 helicase activity (Gallinari *et al.*, 1999), however a more recent report suggested that it may also be a co-factor for helicase activity as it enhanced the RNA binding capacity of the helicase in the presence of ATP (Beran *et al.*, 2009). Mutation analysis implicated a conserved stretch of acidic residues in the C-terminal domain of NS4A as the regulatory domain of the helicase (Shiryaev *et al.*, 2010).

Whether NS4A has any functions independent of NS3 is not known. If NS4A acts solely in *trans* with NS3 it may be that one of the functions of NS3 requires it to dissociate from the membrane. As part of the NS3/4A protease; NS4A is linked with mitochondrial dysfunction and over expression studies suggested it functions as a transcription /translation regulator (Florese *et al.*, 2002; Kadoya *et al.*, 2005), but these require validation in more physiological systems.

1.7.4 NS4B

NS4B is a 27 kDa integral membrane protein capable of targeting independently to ER membranes, indicating the presence of an internal signal sequence (Grakoui *et al.*, 1993a; Hugle *et al.*, 2001). NS4B has been shown to alter membrane formation and has therefore been implicated as the architect of the 'membranous web' associated with the HCV replication complex and is essential for replication (Egger *et al.*, 2002). Both termini are processed by the NS3/4A protease and it is predicted to form a 4 TMD protein with both termini oriented to the cytosol (Lundin *et al.*, 2006; Lundin *et al.*, 2003) (Figure 1.7). As the result of an assumed post-translational event, a minor proportion of the NS4B population has been proposed to assume a different topology whereby the N- terminus is oriented to the ER lumen, giving it a topology common to the NS4B protein of other viruses of the Flaviviridae (Lundin *et al.*, 2003). However, this phenomenon is of limited occurrence when NS4B is co-expressed with NS5A (Lundin *et al.*, 2006). The C-terminal domain of NS4B has been reported to modulate NS3/4A activity, NS5A hyper-phosphorylation and genome replication (Lindenbach *et al.*, 2007).

In addition to the four characterised TMDs of NS4B, two N-terminal amphipathic helices have been described between residues 6-29 (Elazar *et al.*, 2004) and 42-66 (Gouttenoire *et al.*, 2009a). Residues 42-66 displayed α-helical structure in membrane mimetic detergents by solution nuclear magnetic resonance (NMR) and was speculated to be capable of traversing the membrane upon NS4B oligomerisation (Gouttenoire *et al.*, 2009a). Both domains were sufficient to target GFP to membranes but targeting was abrogated following mutagenesis (Elazar *et al.*, 2004; Gouttenoire *et al.*, 2009a). Furthermore, residues 227-254 within the C-terminal domain of NS4B have also been solved by NMR and proposed to form an amphipathic

helix (Gouttenoire *et al.*, 2009b). Disruption of the membrane targeting of any of these three domains prevented replication complex formation (Elazar *et al.*, 2004; Gouttenoire *et al.*, 2009a; Gouttenoire *et al.*, 2009b). Residues 40-69 and 227-254 fused to GFP were fully or partially extracted from membranes, respectively, following alkaline carbonate treatment implying that they were superficially associated with membranes (Gouttenoire *et al.*, 2009a; Gouttenoire *et al.*, 2009b).

1.7.5 NS5A

NS5A is essential for replication of the viral genome and is a component of the replication complex (Macdonald *et al.*, 2005). It exists in two phosphorylation states: the basally-phosphorylated p56 (56 kDa) and the hyper-phosphorylated p58 (58 kDa) forms (Huang *et al.*, 2005). Limited proteolysis by tryptic digest confirmed computational prediction of a three domain organisation of NS5A separated by two low complexity sequences (LCS) (Tellinghuisen *et al.*, 2004). NS5A associates with membranes via a membrane-targeting sequence within domain 1 (Figure 1.7). Expression of the N-terminal 30 residues of NS5A as a fusion to eGFP altered the sub-cellular localisation of eGFP to a pattern comparable to NS5A (Brass *et al.*, 2002; Elazar *et al.*, 2003). These residues are well conserved, particularly for hydropathy, and helical wheel projection analysis found that aa 4-26 resemble an amphipathic helix (Brass *et al.*, 2002; Elazar *et al.*, 2003). The solution NMR structure of this domain showed it to be α-helical in nature with a flexible helical loop in the middle (Penin *et al.*, 2004). Further NMR characterisation suggested that it formed an in-plane amphipathic helix that embedded deeply into the membrane bi-layer (Penin *et al.*, 2004). The ability of NS5A to associate with membranes is vital for its role in replication (Elazar *et al.*, 2003).

A similar domain was also found in the NS5A protein of other related viruses (Brass *et al.*, 2007; Sapay *et al.*, 2006) and an 18-mer synthetic peptide of the NS5A AH that was potently virocidal against HCV and numerous other enveloped, ss RNA viruses was hypothesised to function by disrupting the virion envelops (Cheng *et al.*, 2008). Whether NS5A functions *in vivo* to limit the viral load via it's AH and thus aid in immune-evasion is not known.

Co-ordination of a zinc ion by a zinc binding finger within domain I is essential for replication (Tellinghuisen *et al.*, 2004). Structural characterisation of domain I showed that it may form two distinct dimer conformations that may play unique roles within the life cycle (Love *et al.*, 2009; Tellinghuisen *et al.*, 2005). Domains II and III appear to be mostly unstructured (Hanoulle *et al.*, 2010; Hanoulle *et al.*, 2009; Liang *et al.*, 2007; Verdegem *et al.*, 2011) or flexible which is thought to enable the formation of numerous protein:protein interactions. NS5A is thought to

modulate replication by interacting with HCV RNA (Huang *et al.*, 2005) and NS5B (Shirota *et al.*, 2002). NS5A RNA-binding involves all three protein domains and it preferentially binds the poly (U/UC) tract of the HCV 3' UTR, whereas NS5B preferentially binds the 3' UTR X-tail region (Foster *et al.*, 2010a).

NS5A is a proline-rich protein that has been shown to interact with numerous cellular proteins in order to modulate the cellular environment and the virus life cycle (Macdonald *et al.*, 2004). NS5A contains two poly-proline motifs within the low-complexity sequence separating domains 2 and 3 (LCSII) which bind cellular proteins containing Src-homology 3 (SH3) domains. Domain 2 contains the interferon sensitivity determining region (ISDR) which was identified by comparative analysis of samples from sensitive and insensitive patients and confirmed by mutagenesis (Enomoto *et al.*, 1995; Enomoto *et al.*, 1996). However, other studies have questioned the correlation between this domain and indeed NS5A in determining virus sensitivity to IFN-α (Aizaki *et al.*, 2000; Brillet *et al.*, 2007).

While domain III of NS5A is not required for RNA replication (Appel *et al.*, 2008; Tellinghuisen *et al.*, 2008b), deletion mapping (Appel *et al.*, 2008; Hughes *et al.*, 2009b) and domain swapping (Kim *et al.*, 2011) identified that it plays an essential role in virion morphogenesis. Subsequently specific residues within domain III and LCSII have been shown to regulate the assembly function of NS5A (Hughes *et al.*, 2009a; Masaki *et al.*, 2008; Tellinghuisen *et al.*, 2008a).

1.7.6 NS5B

NS5B is the C-terminal protein of the HCV polyprotein and is a 65 kDa RNA dependant RNA polymerase (RdRp) (Behrens *et al.*, 1996; Grakoui *et al.*, 1993a). NS5B is a so-called 'tail-anchored' protein as it is targeted to the ER or ER-derived membranes via the hydrophobic C-terminal 21 residues which form a TMD (Ivashkina *et al.*, 2002; Schmidt-Mende *et al.*, 2001) (Figure 1.7). NS5B is capable of post-translationally associating with membranes and inserts to orientate its N-terminal enzyme domain to the cytosol (Schmidt-Mende *et al.*, 2001). Membrane association is essential for enzyme activity (Lee *et al.*, 2004; Moradpour *et al.*, 2004).

The C-terminal *trans*-membrane domain can be replaced with that of poliovirus 3A protein (Lee *et al.*, 2006) and the TMD from the GBV-B polymerase (Brass *et al.*, 2010). Maintenance of the 3A anchor region-containing replicon under prolonged selection pressure recovered replication but no mutations were found in the 3A domain, suggesting that its role is purely that of a membrane anchor (Lee *et al.*, 2006). Subsequently, a more detailed structural analysis

of chimeric replicons identified a conserved structural feature present within the 3A anchor domain and the NS5B TMD from HCV and GBV-B that conferred viability, possibly through enabling intra-membrane interactions (Brass *et al.*, 2010).

Solution of the X-ray crystal structure of NS5B showed that it forms a 'right-hand' structure with thumb, finger and palm domains (Lesburg *et al.*, 1999). The three domains completely encircle the active site, in contrast to other similar polymerases which form an open-circle, or 'U' shape. The global structure was found to be relatively inflexible and is believed to remain largely unchanged throughout the catalytic cycle (Lesburg *et al.*, 1999).

The RdRp enzyme contains a highly conserved GDD motif that is flanked by hydrophobic sequences. This motif is a common feature of the RdRp of RNA viruses (Kamer & Argos, 1984; Poch *et al.*, 1989) and mutation of this motif completely blocks genome replication (Yamashita *et al.*, 1998).

Using the positive sense genome as the template, the RdRp binds the 3' X-tail within the 3' UTR (Foster *et al.*, 2010a; Zhong *et al.*, 2000) and initiates synthesis of a negative-strand copy in a primer-independent manner (Behrens *et al.*, 1996). NS5B is capable of synthesising RNA duplexes by *de novo* and 'copy–back' mechanisms (Ferrari *et al.*, 1999); although de novo synthesis is thought to be the mechanism employed *in vivo* (Zhong *et al.*, 2000). The replication intermediate is then used by the RdRp to replicate the genome. RdRp activity was found to require the presence of Mn^{2+} ions, preferentially, or Mg^{2+} ions *in vitro* (Behrens *et al.*, 1996; Ferrari *et al.*, 1999). Glutamic acid 18 and histidine 502 were found to be essential for catalytic activity (Qin *et al.*, 2001), and subsequently also for oligomerisation, implying that RdRp activity may be dependent upon multimerisation (Qin *et al.*, 2002).

The RdRp does not contain any proof-reading activity which leads to the high sequence variability of the major genotypes and of the quasi-species within infected individuals and is central to the virus' ability to escape the immune response and become resistant to specific antivirals (Lindenbach & Rice, 2003). Two forms of drugs targeting NS5B RdRp function are currently under development: nucleoside-analogue inhibitors (NIs) and non-nucleoside-analogue inhibitors (NNIs). Because NIs target the highly conserved active site they confer better genotype cross-reactivity and a greater barrier to resistance.

NS5B forms part of the replication complex and is able to function without any co-factors *in vitro*. NS3, NS4B and NS5A have been shown to modulate activity of NS5B (Behrens *et al.*, 1996; Piccininni *et al.*, 2002; Shirota *et al.*, 2002) and the interaction between NS5A and NS5B

is essential for genome replication (Shimakami *et al.*, 2004). The cellular protein c-Src has been shown to interact with NS5A and NS5B via its SH2 and SH3 domains, respectively, thus mediating the NS5A:NS5B interaction and promoting replication (Pfannkuche *et al.*, 2011). The cellular protein cyclophilin B (CyPB) was shown to stimulate NS5B RNA binding activity (Watashi *et al.*, 2005), although cyclophilin A (CyPA) was later shown to be the only cyclophilin with a pivotal role in replication (Yang *et al.*, 2008). Other cellular interacting partners of NS5B include two helicases, both of which enhance replication (Goh *et al.*, 2004; Kyono *et al.*, 2002).

1.8 NS2

1.8.1 Topology and structure

NS2 is a 217 amino acid *trans*-membrane protein comprising residues 814 to 1030 of the JFH-1 polyprotein (Grakoui, 1993b; Mizushima *et al.*, 1994; Selby *et al.*, 1994). Release of NS2 from the polyprotein requires cleavage at the N-terminus by signal peptidase (Mizushima *et al.*, 1994), and processing at the C-terminus by the NS2/3 cysteine autoprotease (Grakoui, 1993b). NS2 has two domains: an N-terminal *trans*-membrane domain comprising residues 1-92 and a soluble C-terminal catalytic domain comprising residues 93-217, although the exact definitions of these domains vary. NS2 localises to the ER membrane (Santolini *et al.*, 1995), and can co-translationally target to membranes independently from the other viral proteins (Yamaga & Ou, 2002). Truncation analysis demonstrated that NS2 contains two distinct internal membrane-targeting sequences, between residues 30-74 and 119-151 (Yamaga & Ou, 2002).

The catalytic domain was originally defined as the minimum sequence required for cleavage of the NS2/3 junction (Thibeault *et al.*, 2001) and more recently the X-ray crystal structure of the catalytic domain in a post-cleavage state was solved (Lorenz *et al.*, 2006b) (Figure 1.8).

Mutational analysis had shown that histidine 143, glutamic acid 163 and cysteine 184 were involved in catalysis of NS2/3 cleavage (Grakoui, 1993b; Hijikata *et al.*, 1993a; Welbourn *et al.*, 2005). Elucidation of the 3D structure of the NS2 catalytic domain (aa 94-217), post-cleavage, by X-ray crystallography confirmed that these three residues form the enzyme catalytic triad (Figure 1.8). Interestingly, the NS2 catalytic domain, post-cleavage, was shown to form a chimeric active-site dimer, comprising H143 and E163 residues from one monomer and the C184 residue from another monomer (Lorenz *et al.*, 2006b). Dimerisation of the mature full-length protein has not been directly addressed, although Stapleford *et al.,* reported an NS2-specific protein species of approximately 40 kDa following blue-native polyacrylamide gel electrophoresis (BN-PAGE) analysis of NS2-containing complexes which they speculated to be

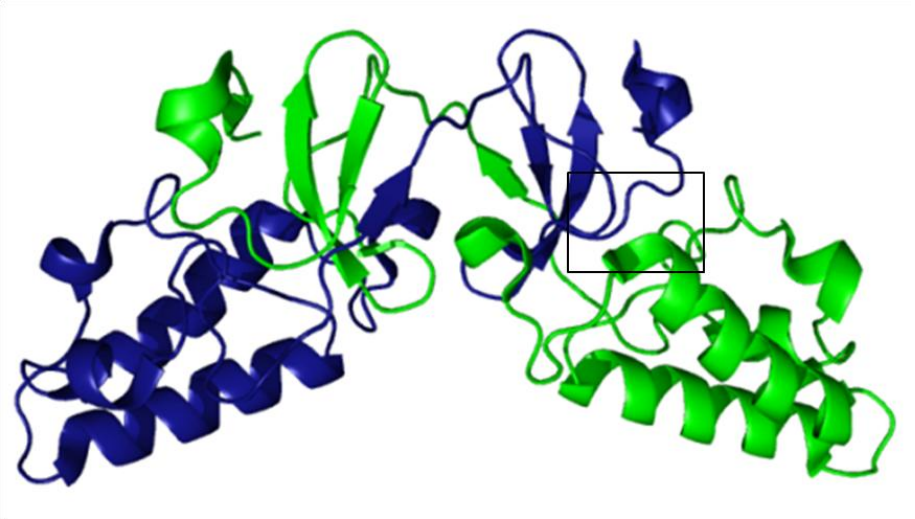


Figure 1.8 3D crystal structure of the NS2 catalytic domain (aa 94-217) in the post-cleavage state.

The domain crystallised as an active-site dimer and the two monomers are shown in blue and green (Lorenz *et al.*, 2006b). The active-site is highlighted. Two of the catalytic residues, H143 and E163, are provided by one monomer, the third catalytic residue, C184, and the substrate sequence is provided by the other monomer. Post-cleavage, the C-terminal leucine, L217, is coordinated into the activate site.

NS2 dimers (Stapleford & Lindenbach, 2011). Using site directed mutagenesis to disrupt the putative dimer interface, Dentzer and colleagues (Dentzer *et al.*, 2009) concluded that dimerisation was essential for infectivity; however they did not confirm that the NS2 mutants disrupted dimer formation.

The catalytic domain can be further divided into two sub-domains; the N-terminal domain that contains two α-helixes and the C-terminal domain that consists of four anti-parallel β sheets. The two sub-domains are connected by an extended linker (Lorenz *et al.*, 2006b). In relation to the active site residues; H143 and E163 are in the N-terminal sub-domain and C184 is in the C-terminal sub-domain. Another notable feature of the crystal structure was that proline 164 assumed an unusual *cis* conformation which heavily contributed to the active site architecture (Lorenz *et al.*, 2006b). Mutation of this residue in a mono-cistronic J6/H77/JFH virus blocked replication and mutation in the bi-cistronic virus reduced infectivity by 10 fold but did not alter replication (Phan *et al.*, 2009). This suggests that correct active site architecture is required for efficient virus assembly, as well as NS2/3 cleavage.

Glycosylation analysis of *in vitro* expressed NS2 was combined with hydropathy scoring to generate a 4 TMD topology prediction for NS2 (Yamaga & Ou, 2002) (Figure 1.9A). In this model the N-terminal ~140 aa were described as forming the membrane-spanning domain and both termini were predicted to be orientated to the ER lumen (Yamaga & Ou, 2002), in agreement with previous findings (Santolini *et al.*, 1995). In this model the fourth TMD was predicted to be within the catalytic domain (aa 119-147), yet the crystal structure of the catalytic domain provided no evidence of a TMD (Lorenz *et al.*, 2006b).

By analysing the outputs of 6 topology prediction programs, and taking into account the localisation of the cleavage events that release the mature protein and the 3D structure of the catalytic domain, Jirasko *et al.,* (Jirasko *et al.*, 2008) produced a 3 TMD model that omitted the fourth putative TMD defined by Yamaga and Ou (Yamaga & Ou, 2002) (Figure 1.9B).

Subsequently the 3D structure of the N-terminal 27 residues of NS2 from the Con1 isolate were solved by solution NMR and were found to assume an α-helical conformation in 50% tetrafluoroethylene (TFE) (Jirasko *et al.*, 2008). The structure of this domain was incorporated into a refined model of NS2 topology based on the output from a single topology prediction program (Phan *et al.*, 2009) (Figure 1.9C). This model estimated the 3 TMDs to be positioned between aa 4-24, aa 30-50 and aa 63-84. More recently 3D structures of peptides encoding residues 27-59 and 60-99 from the Con1 isolate have also been solved by solution NMR and

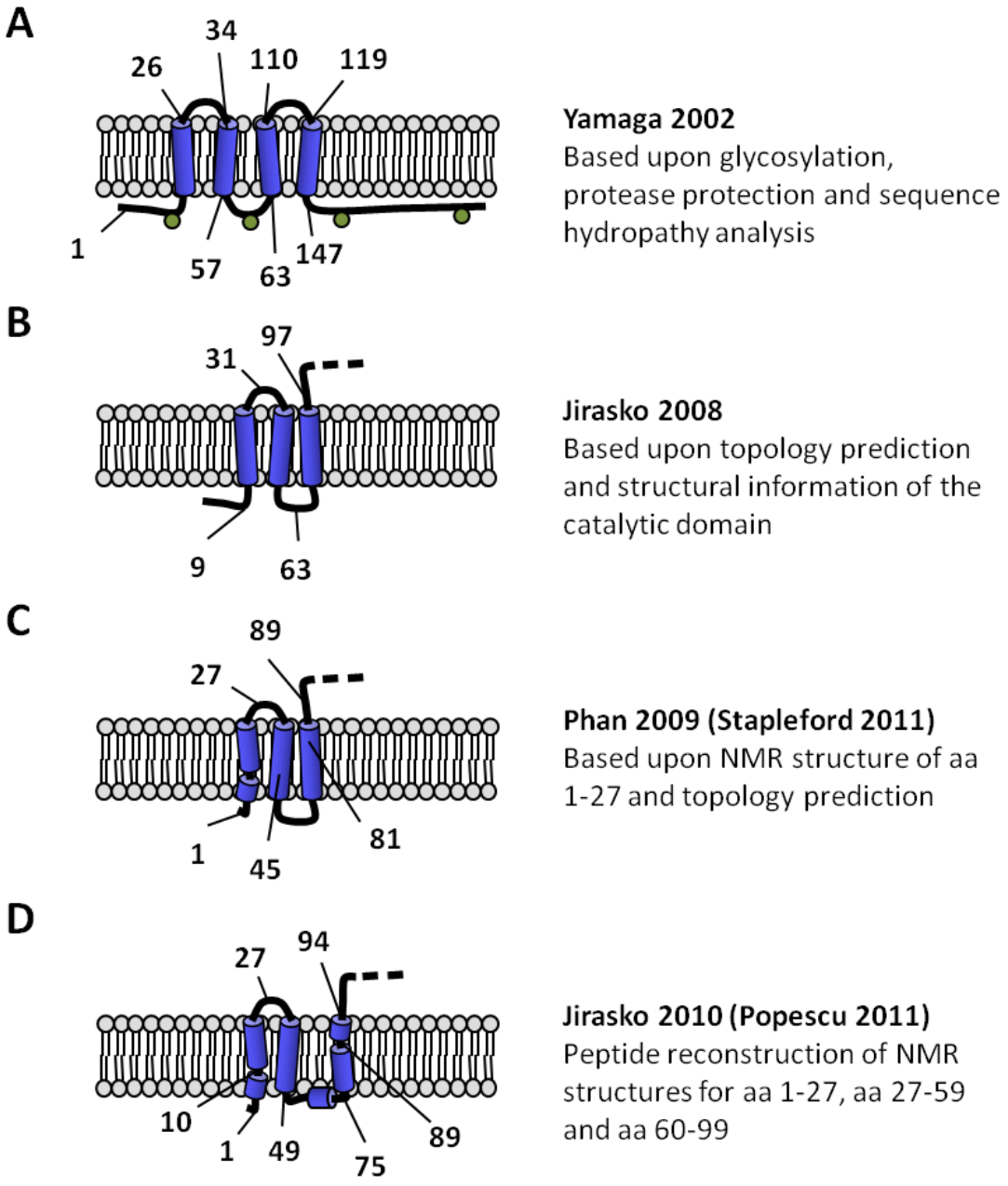


Figure 1.9 Topology models for NS2.

Model A represents the complete NS2 protein; the catalytic domain is not shown for models B-D. Sites of artificial glycan acceptor sites that were glycosylated are shown (green circles). Key residues are shown. No information was given regarding the position of the luminal loop in model C. All topology models predicted the N-terminus to be luminal oriented.

shown to form stretches of α-helical structure in 50% TFE (Jirasko *et al.*, 2010). This has led to the proposal of a refined model of NS2 topology, generated from domain reconstruction of the three peptide sequences (Jirasko *et al.*, 2010; Popescu *et al.*, 2011) (Figure 1.9D). In this latest model the three TMDs are shown to span residues 8-24, 29-48 and 74-97. A three TMD topology is in agreement with recent observations of NS2 topology in virus expressing cells following partial permeabilisation of the plasma membrane. Under these conditions the catalytic domain was detected in the cytosol and sensitive to protease treatment (Ma *et al.*, 2011).

Residue 10 of NS2 is a highly conserved glycine and in the NMR structure of the aa 1-27 peptide it was shown to disrupt the α-helical structure of the putative first TMD. Mutation of G10 to alanine was lethal in the Con1/C3 virus (Figure 1.10), leading to the proposal that this residue and the role it performs as a flexible joint in the first putative TMD of NS2 were vital for virus assembly (Jirasko *et al.*, 2008). However, mutation of this residue in the Jc1 virus and JFH-1 virus (Figure 1.10) resulted in an attenuated phenotype with extracellular infectivity reduced by 1 log (Phan *et al.*, 2009). While different viruses may be the root cause of the differences it is also important to note that while Jirasko and colleagues observed that p7-NS2 cleavage was impaired in this mutant, Phan *et al.,* employed foot and mouse disease virus (FMDV) 2A autoprotease to artificially mediate p7-NS2 cleavage (Phan *et al.*, 2009). So the more impaired phenotype observed by Jirasko *et al* may have been a direct result of impaired p7/NS2 cleavage (Jirasko *et al.*, 2008), at least in the Jc1 virus.

Conservation of specific residues and small amino acids along one face of the helix formed by aa 1-27 led to the speculation that the first putative TMD may form stabilising inter or intra-molecular interactions with other helices (Jirasko *et al.*, 2008). The residues of this helix are well conserved within genotypes but show significant variability between genotypes suggesting co-adaptation of this domain with corresponding domains of NS2 or other viral proteins.

The second peptide (aa 27-59) formed a single, unbroken α-helical structure between aa 27-49, with polar residues populating one face of the helix (Jirasko *et al.*, 2010). This structure was reminiscent of an amphipathic helix and it was noted that in order for this domain to stably span the membrane it would require neutralisation of the charged residues, potentially by interacting with another membrane-spanning domain (Jirasko *et al.*, 2010).

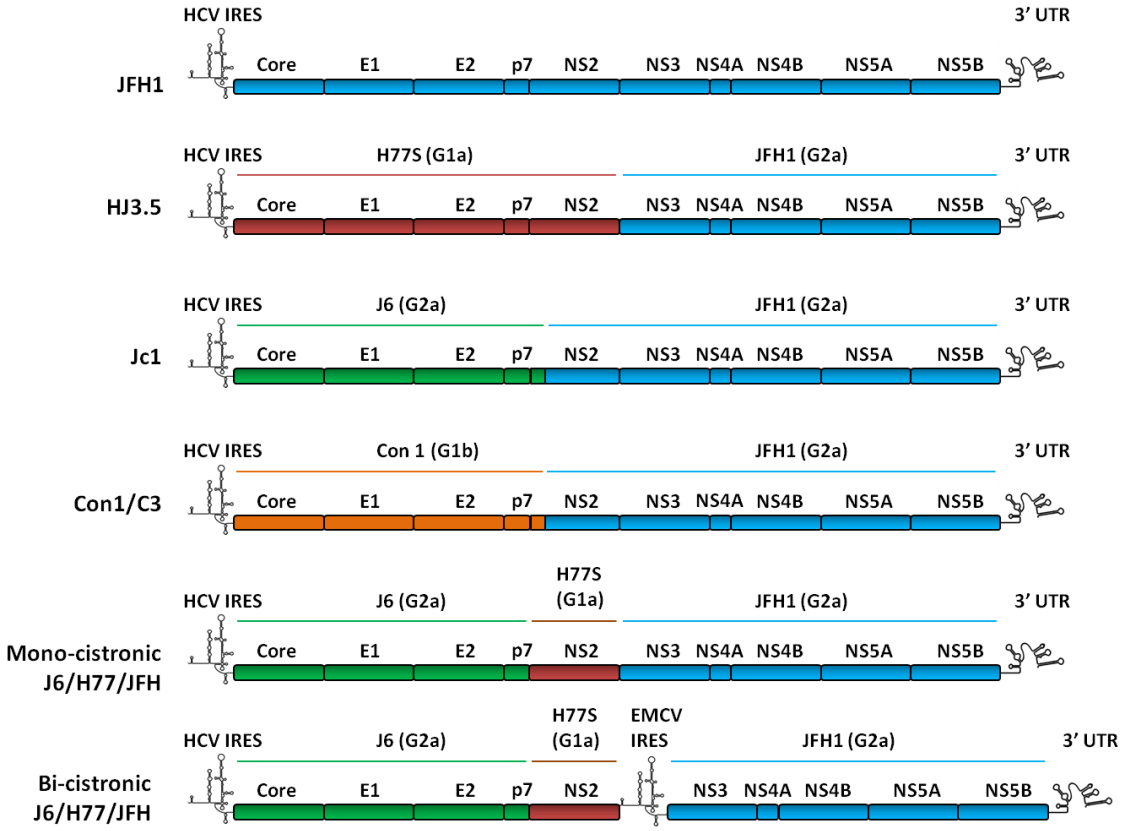


Figure 1.10 Schematic representation of the genome organisation of some of the viruses used to study NS2.

Genome segments provided by each isolate are coloured. The genotypes of donor isolates are noted. The 5' and 3' UTRs were JFH-1 in origin. The point of recombination in the Jc1 and Con1/C3 viruses is after residue 33 of NS2.

The third peptide (aa 60-99) formed three α-helical structures; designated *trans*-membrane segment (TMS) 3.1 (aa 61-70), TMS3.2 (aa 73-86) and TMS3.3 (aa 89-95). TMS3.1 was predicted to orientate in-plane with membrane bi-layers and at right angles to TMS3.2 and TMS3.3 because of a proline at position 73. A glycine and proline at residues 88 and 89 were responsible for a loss of secondary structure between TMS3.2 and TMS3.3.

1.8.2 NS2/3 processing

NS2/3 processing is essential for infection in chimpanzees (Kolykhalov *et al.*, 2000) and in cell culture using wild type (wt) virus (Jones *et al.*, 2007). The insertion of a second IRES between NS2 and NS3 clearly demonstrated that NS2/3 cleavage was only necessary to correctly release NS3 in cell culture as mutation of the active residue C184 to alanine in the bi-cistronic virus had no effect on RNA replication or infectivity of the virus at 72 hours post-transfection, although it displayed delayed kinetics (Jones *et al.*, 2007). This also demonstrated that while uncleaved NS2/3 precursor is essential for infectious particle production in BVDV and other *Pestiviruses*, this was not the case for HCV (Jones *et al.*, 2007). As no other target of the NS2/3 protease has been identified and catalysis of cleavage of the NS2/3 junction appears to be required solely to correctly release NS3; the NS2/3 protease has been described as a positive strand RNA virus accessory protease (Thibeault *et al.*, 2001).

Proteolytic cleavage of NS2/3 precursor molecules can occur in the absence of microsomal membranes (Santolini *et al.*, 1995; Thibeault *et al.*, 2001) and independently of cellular co-factors (Hijikata *et al.*, 1993a; Thibeault *et al.*, 2001); however it does require membrane mimetic detergents (Pieroni *et al.*, 1997; Thibeault *et al.*, 2001) highlighting the necessity of a hydrophobic environment to facilitate folding and/or functionality of the protease. The NS2/3 precursor, identified by truncation analysis as the minimal region required for efficient auto-cleavage, encompasses the C-terminal domain of NS2 (aa 95-217) and the serine protease domain of NS3 (aa 1-181) (Thibeault *et al.*, 2001). Although this construct expressed well in *E. Coli* and was cleavage-competent it was insoluble so required denaturing, purifying and re-folding before it could be analysed (Thibeault *et al.*, 2001).

The inhibition of NS2/3 processing by EDTA and 1, 10 phenanthroline (Grakoui, 1993b; Hijikata *et al.*, 1993a; Thibeault *et al.*, 2001), but not by specific NS3 inhibitors (Thibeault *et al.*, 2001), and stimulation of processing by $ZnCl_2$ implied that NS2/3 cleavage was carried out by a zinc metalloprotease and independent of NS3 protease activity (Hijikata *et al.*, 1993a). It was later found that the protease domain of NS3 contains three cysteine residues and a histidine residue which co-ordinate a zinc ion and that loss of zinc-co-ordination following mutation of these

residues inhibited auto-cleavage of the NS2/3 pre-cursor (Tedbury & Harris, 2007). NS3 serine protease activity also requires zinc coordination but NS2/3 auto-cleavage activity is more sensitive to zinc chelation than NS3 protease activity (Tedbury & Harris, 2007).

NS2/3 processing was efficiently inhibited by expression of the C-terminal 10 aa of NS2 as a peptide (Thibeault *et al.*, 2001). A mechanism of competitive inhibition is consistent with the X-ray crystal structure of the NS2 catalytic domain, post-cleavage, in which the C-terminus of NS2 was co-ordinated within the active site (Lorenz *et al.*, 2006b). Peptides of NS4A have also been shown to inhibit *in vitro* translated and bacterially expressed NS2/3 while enhancing NS3 protease activity (Darke *et al.*, 1999; Thibeault *et al.*, 2001).

Recent work demonstrating that inefficient cleavage was possible with as few as 2 aa of NS3 provided the first evidence that NS2 is an independent protease (Schregel *et al.*, 2009). The role of NS3 is therefore likely one of a structurally important co-factor involving co-ordination of a zinc ion (Dentzer *et al.*, 2009; Schregel *et al.*, 2009) as protein re-folding of the NS2/3 precursor was blocked upon the addition of EDTA to the re-folding buffer (Foster *et al.*, 2010b).

Mutations to residues proximal to the cleavage site were well tolerated by NS2/3 pre-cursor expressed *in vitro*. Mutation of the C-terminal residue of NS2, L217, to alanine, arginine and phenylalanine had no effect on processing *in vitro*. L217E caused a moderate inhibition of cleavage and mutation to proline prevented cleavage entirely (Hirowatari *et al.*, 1993; Reed *et al.*, 1995). When similar mutations were analysed in the context of a fully infectious bi-cistronic virus, with an IRES separating NS2 and NS3 (Figure 1.10), they were found to be largely ineffectual towards replication but severely inhibited infectious virus production (Dentzer *et al.*, 2009). Even the relatively subtle mutation of L217I produced a 10-fold decrease in virus infectivity.

Similar to the findings of Jones *et al.,* (Jones *et al.*, 2007), mutation of the active-site residue C184 to alanine in a bi-cistronic Jc1 virus, but encoding NS2 from the H77 isolate (Figure 1.10), had no effect on replication, intra- and extracellular core and intra- and extracellular infectivity (Dentzer *et al.*, 2009). However, mutation of H143, also part of the catalytic triad, to alanine in a similar bi-cistronic virus caused a minor decrease in released core but a 1.5-2 log decrease in intra- and extracellular infectivity (Jirasko *et al.*, 2008). Examination of the 3D structure of the catalytic domain suggested that H143 interacts with the two C-terminal leucines and serine 194 from the other monomer and forms a salt bridge with E163 of the same molecule (Lorenz

et al., 2006b). It was speculated that mutation of H143 to alanine would disrupt these interactions, whereas the C184A mutation was expected to have little impact (Jirasko *et al.*, 2008). Analysis of these two mutations in the NS2/3 precursor showed that while both mutations impaired protein re-folding; H143A was significantly less folded than C184A and lacked the α-helical secondary structure characteristics observed in the wt and C184A precursors (Foster *et al.*, 2010b). Together this demonstrates that, while catalytic activity of the NS2 protease is not required, correct structural conformation of the catalytic domain and the C-terminal leucine, governed in part by interactions formed by the active-site residue H143, are required for infectivity.

As mentioned previously, the post-cleavage NS2 catalytic domain crystallised as an active-site dimer. The potential for *trans* cleavage of the NS2/3 junction was initially identified by co-expression of cleavage-deficient (C184A and H143A) truncated NS2/3 precursor molecules with full-length polyprotein and monitoring for the formation of truncated NS2 (Grakoui, 1993b). This was explored in more detail by characterising the cleavage in *trans* of a precursor molecule lacking the P1 and P1' residues of the NS2/3 cleavage site (ΔP1-P1') (Reed *et al.*, 1995). In isolation the ΔP1-P1' precursor was not able to catalyse auto-cleavage; however co-expression with a C184A mutant precursor resulted in cleavage, albeit highly attenuated by comparison to wt precursor molecules. Interestingly, co-expression of ΔP1-P1' with a H143A mutant precursor produced no cleavage products (Reed *et al.*, 1995). These experiments were reproduced by Lorenz *et al.,* and demonstrate the potential for NS2/3 cleavage in *trans* (Lorenz *et al.*, 2006b). These experiments corroborate the crystal structure data showing that the active site cysteine residue and the prospective C-terminus are co-ordinated by one molecule, while the other two active site residues, H143 and E163, are provided by the other molecule (Lorenz *et al.*, 2006b; Reed *et al.*, 1995).

Controversy still surrounds the notion of bi-molecular cleavage at the NS2/NS3 junction as if it is an obligatory factor for NS2/3 cleavage it would require the coordination of at least one nascent polyprotein at the membrane surface and therefore could impose a rate limiting step on protein expression and replication. It is interesting to note therefore that separation of NS2 and NS3 with a second IRES, thereby negating this potential rate-limiting step, did not result in an increase in virus replication kinetics (Jones *et al.*, 2007). Furthermore, the highly attenuated cleavage observed in *trans* suggests that this mechanism is very inefficient. It is possible that the conformation shown in the catalytic domain structure produced by Lorenz *et al.,* (Lorenz *et*

al., 2006b) represents a less abundent, but more stable, conformation that was more amenable to crystallisation techniques.

1.8.3 tNS2

Several groups have reported the generation of a truncated form of NS2, termed tNS2 (Jirasko *et al.*, 2008), in experiments involving full-length virus. tNS2 appears to be the results of an N-terminal cleavage event as it has been observed with anti-NS2 antiserum (Boson *et al.*, 2011; Jirasko *et al.*, 2010; Stapleford & Lindenbach, 2011) but not N-terminal tag epitopes (Jirasko *et al.*, 2010; Stapleford & Lindenbach, 2011). tNS2 was detectable in whole-cell lysates but not in high molecular weight, NS2-containing complexes (described in detail in section 1.8.6) purified from the same samples (Stapleford & Lindenbach, 2011). Ma *et al.,* (Ma *et al.*, 2011) observed tNS2 only when an IRES was inserted between p7 and NS2 leading the authors to suggest that is an aberrant initiation product.

The NS2A of YFV plays an essential role in virus assembly which is mediated by a cleavage event carried out by the viral protease (Nestorowicz *et al.*, 1994). Stapleford *et al.,* (Stapleford & Lindenbach, 2011) speculated that the generation of tNS2 may operate as a means to prevent premature particle assembly or a mechanism to dissociate the NS2 complexes prior to particle formation. It is unlikely to be as straight forward as this however, as point mutations that altered the amount of tNS2 did not correlate with observed decreases in infectivity (Jirasko *et al.*, 2010).

The role and indeed the relevance of tNS2 in the virus cycle remains to be confirmed. Truncated forms of NS2 have not been reported using *in vitro* translation systems (Hijikata *et al.*, 1993a; Santolini *et al.*, 1995; Yamaga & Ou, 2002) as well as some cellular systems (Tedbury *et al.*, 2011; Yi *et al.*, 2009). This could be for several reasons; the absence of the specific protease from the systems studied, experimental constraints such as protein resolution and detection or that this cleavage event is not ubiquitous among HCV isolates.

1.8.4 S168

NS2 is a short lived protein with a reported half-life of <30 minutes for DNA-expressed (Franck *et al.*, 2005) and between 3-6 hours for RNA-expressed protein (Welbourn *et al.*, 2009). Serine 168 was originally described as being critical for the degradation of NS2 in the genotype 1a isolate HCV-H via the proteasome as levels of genotype 1a NS2 could be rescued following treatment with the proteasome inhibitors MG132 and clasto-lactocystin-β-lactone (Franck *et al.*, 2005; Yi *et al.*, 2009). Sequence analysis revealed that serine 168 was part of a consensus casein kinase II (CKII) phosphorylation site S/TXXE, where X is any residue. Mutation of S168 to

alanine in an over-expression system stabilised NS2 to the same extent as MG132 treatment. Furthermore treatment of cells expressing a full-length replicon with either the CKII inhibitor curcumin or MG132 both resulted in NS2 stabilisation (Franck *et al.*, 2005). Similarly NS2/3 precursor levels could be stabilised using MG132 in NS2/3 cleavage-deficient replicons based on a genotype 1b isolate, Con 1, suggesting proteasome-mediated degradation (Welbourn *et al.*, 2005; Welbourn *et al.*, 2009). However, mature Con1 NS2 was insensitive to MG132 treatment (Welbourn *et al.*, 2009). Furthermore, mutation of the 6 lysine residues within Con1 NS2 to arginine had no effect on mature protein stability, suggesting that degradation of NS2 and NS2/3 may be different and may vary between genotypes (Welbourn *et al.*, 2009).

Tedbury *et al.,* generated two replicons encoding NS2 through to NS5B (NS2-5B) and signal-peptide p7 through to NS5B (SPp7-5B) (Tedbury *et al.*, 2011) (Figure 1.4B). Using these constructs it was found that NS2 partitioned to a detergent-insoluble fraction in the presence of p7 and co-localised with NS3 and NS5A. While the S168A mutation did not alter NS2 partitioning, it did disrupt co-localisation with NS5A (Tedbury *et al.*, 2011). NS2 expressed from SPp7-5B replicons was able to bind NS5A and co-localise to discrete punctae with NS3 and NS5A, both of which were blocked by the S168A mutation (Tedbury *et al.*, 2011). Although this data implicated S168 as vital for the NS2:NS5A interaction, it was inconsistent with full-length virus experiments where the S168A/G mutations were lethal but did not disrupt NS2 subcellular localisation (Ma *et al.*, 2011). The mechanism of action of the S168 mutations is further confused by the speculation that their lethal phenotype in full-length virus was a result of a disrupting co-localisation between NS5A and E2 (Ma *et al.*, 2011).

Interestingly the S168A mutation resulted in attenuation of the HJ3-5 virus but the NS2:NS3 interaction was unaffected (Ma *et al.*, 2011). Rescue experiments found that second site mutations in NS2 (L174V) and NS5A (V464L) could recover infectivity from the HJ3-5 S168G virus suggesting that the C-terminal catalytic domain of NS2 interacts with NS5A (Yi *et al.*, 2009). The X-ray crystal structure of the NS2 catalytic domain shows S168 and V174 to be positioned on the same solvent-exposed surface (Lorenz *et al.*, 2006b). By supplying NS2 in *trans*, it was shown that wt NS2 could *trans*-complement mutations to S168, in a genotype-specific, or possibly strain-specific, manner (Yi *et al.*, 2009).

Mutation of S168 to alanine or glycine in the HJ3-5 virus (Yi *et al.*, 2009) or to the phospho-mimetic residues aspartic acid or glutamic acid in the Jc1 virus severely attenuated infectious virus formation and extracellular RNA levels (Phan *et al.*, 2009). CKII phosphorylation of S457 of NS5A is essential for virus infectivity, but mutation to the phospho-mimicking aspartic acid

enables infectivity while conferring resistance to CKII inhibitors (Tellinghuisen *et al.*, 2008a). Treatment of infectious virus clones, containing the S457D mutation in NS5A, with the CKII inhibitor DMAT (2-Dimethylamino-4, 5, 6, 7-tetrabromo-1H-benzimidazole) had no effect on intra- or extracellular infectivity (Tellinghuisen *et al.*, 2008a; Yi *et al.*, 2009), implying that the role of S168 is phospho-independent.

Analysis of released core demonstrated that infectious particles produced by the HJ3-5 virus were 'fast sedimenting' while the majority of core released from the uninfectious progenitor virus, HJ3 (which lacks culture adapted mutations that confer viability), was 'slow sedimenting' (Yi *et al.*, 2009). Surprisingly, the majority of extracellular core produced by the uninfectious HJ3-5 (S168G) mutant virus was also 'fast-sedimenting' (Yi *et al.*, 2009). Consistent with the findings using the Jc1 virus (Jirasko *et al.*, 2008), the levels of intracellular core for HJ3-5 (S168G) virus were comparable to the wt HJ3-5 virus, but extracellular core was dramatically decreased (Yi *et al.*, 2009). This fits with the hypothesis that S168 is involved in a maturation step is required to confer infectivity and facilitate virus particle release, which, if not completed, results in degradation of the uninfectious particles, via the proteasome (Gastaminza *et al.*, 2008).

Together these data advocate the hypothesis that S168 plays a role in virus morphogenesis prior to release, and that final maturation of virions likely requires the formation of protein:protein interactions between NS2 and NS5A.

S168 is highly conserved in all genotypes except genotype 4a where it is a threonine. Interestingly, NS2 T168S was identified as a culture adaptive mutation in a genotype 4a/2a chimeric virus encoding core-NS2 from the genotype 4a sequence (Scheel *et al.*, 2008). It is possible that the interacting domain of NS5A in genotype 4a viruses may vary from other genotypes such that it favours a threonine at this site.

1.8.5 Interactions between NS2 and other HCV proteins

Numerous reports have demonstrated physical interactions between NS2 and other viral proteins (Figure 1.11). Stapleford and Lindenbach generated several clones of the Jc1 virus with inserted biotin-acceptor sites and expressed them in a modified Huh7.5 cell line expressing a codon optimised, secreted form of the biotin ligase BirA – Huh7.5 (BirA) (Stapleford & Lindenbach, 2011). Using this system it was found that biotin-tagged E2 in the context of full-length virus was able to interact with NS2 and NS3 while biotin-tagged NS2 interacted with E2 and NS3 (Stapleford & Lindenbach, 2011). Mutagenesis whereby an alanine was inserted following aa 16, aa 41 and aa 82 in order disrupt the helical face of the 3 putative

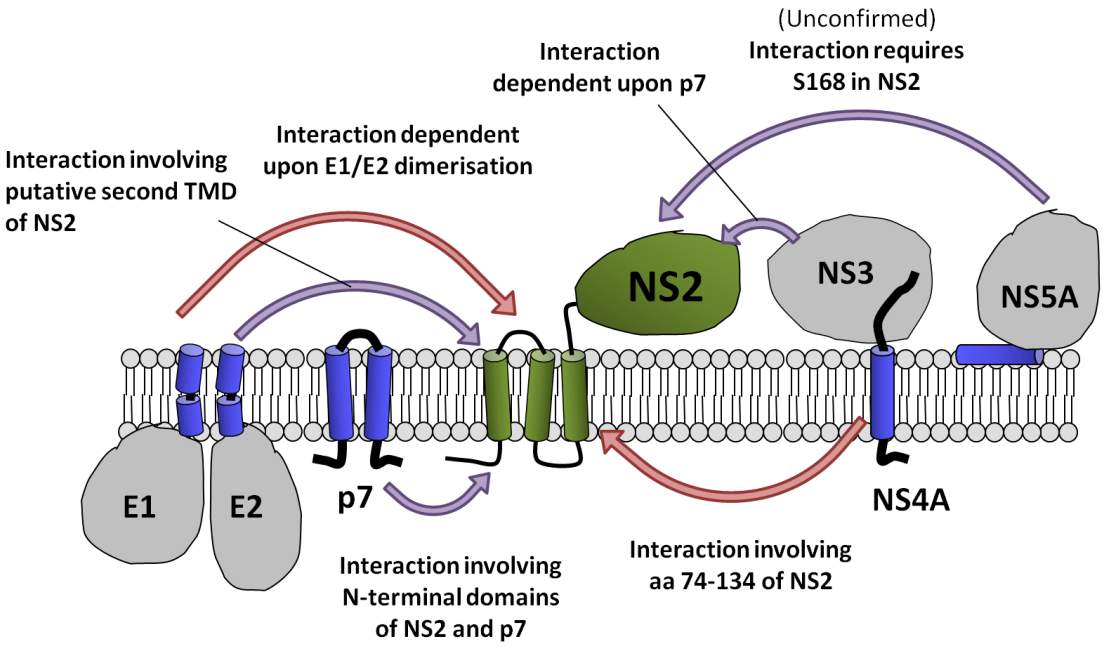


Figure 1.11 Cartoon of the interactions between NS2 and the other viral proteins.

NS2 (green) forms interactions with the other viral proteins. Arrows denoted whether the interaction was shown to be disrupted by the p7 basic loop mutation (purple) or not (pink).

TMDs (defined by Jirasko *et al*., 2010) was combined with deletion mapping to suggest that the E2 interacting domain of NS2 was between aa 32-62 i.e. within the putative second TMD (Popescu *et al.*, 2011).

Interactions between NS2 and NS3 have been reported in a wide range of other systems (Dimitrova *et al.*, 2003; Jirasko *et al.*, 2008; Kiiver *et al.*, 2006). Several mutations within the NS2 *trans*-membrane that reduced infectivity also impaired the NS2:NS3 interaction. The interaction could be partially recovered by compensatory mutations in other proteins that also recovered infectivity (Jirasko *et al.*, 2010). A loss of the NS2:NS3 interaction was noted in the Δp7 virus but not the ΔE1/E2 virus, suggesting the NS2:NS3 interaction is dependent upon p7 and perhaps independent of the glycoproteins (Stapleford & Lindenbach, 2011). Interestingly the interaction was also reduced when NS2 and NS3 were separated by a second IRES (Stapleford & Lindenbach, 2011). NS2 and p7 have been shown to interact directly by immunoprecipitation studies (Ma *et al.*, 2011; Popescu *et al.*, 2011) and FRET-FLIM (fluorescence resonance energy transfer with fluorescence lifetime imaging microscopy) (Popescu *et al.*, 2011). The interaction appears to be genotype-independent and it is the *trans*-membrane domain of NS2, not the catalytic domain, that interacts with p7 (Boson *et al.*, 2011; Popescu *et al.*, 2011). Replacement of the coding sequence for aa 1-27 of NS2 in the JFH-1 virus with that of the Con 1 sequence reduced infectivity and impaired the ability of NS2 to co-immunoprecipitate E2, NS3 and NS5A (Jirasko *et al.*, 2010). However, addition of the E3D rescue-mutation within p7 recovered infectivity and the NS2 interactions providing further evidence that NS2 and p7 interact via their respective N-terminal domains. This also suggested that the NS2:p7 interaction is important for the interaction between NS2 and the other virus proteins (Jirasko *et al.*, 2010). However, the role of p7 in determining the interactions between NS2 and other virus proteins is inconsistent as experiments using full-length JFH-1 virus found the NS2:E2 interaction to be independent of p7 (Popescu *et al.*, 2011). Mutation of the p7 basic loop dramatically disrupts normal protein function as it blocks p7 membrane insertion, abolishes p7 ion-channel function (StGelais *et al.*, 2007) and increases the degradation of NS2 and p7 (Tedbury *et al.*, 2011). Introduction of the p7 basic loop mutation into full length HJ3-5 virus was lethal and abrogated the NS2:E2, NS2:NS3 and NS2:NS5A interactions (Ma *et al.*, 2011), suggesting p7 maybe stabilise or facilitate the NS2:protein interactions.

NS2 has been shown to interact with E1 from virus expressing cells (Ma *et al.*, 2011; Popescu *et al.*, 2011). Long-term culturing of a triple chimeric virus, where the NS2 coding sequence was transposed with that of the H77 isolate in the Jc1 background (J6/H77/JFH – Figure 1.10),

produced an adaptive mutation in E1 (A269T), suggesting direct interaction between E1 and NS2 (Dentzer *et al.*, 2009). However, disruption of E1:E2 hetero-dimerisation (Ciczora *et al.*, 2007) abolished the E1:NS2 interaction which is consistent with E1 interacting with NS2 via its association with E2 (Stapleford & Lindenbach, 2011). An interaction between NS2 and NS4A has been reported and truncation mapping revealed that aa 74-134 of NS2 were sufficient to immunoprecipitate NS4A (Flajolet *et al.*, 2000).

NS2 has also been shown to interact with NS5A, although the interaction is weak compared to other NS2 interactions (Jirasko *et al.*, 2010). Interestingly, it was reported that NS2 preferentially binds the p56 form of NS5A (Jirasko *et al.*, 2010). Confirmation of the NS2:NS5A interaction is lacking though as one study found that NS2 was able to IP NS5A but that this is not reciprocal (Ma *et al.*, 2011), while another group observed no physical interaction between the two proteins (Stapleford & Lindenbach, 2011). This may be a reflection of a weak interaction between NS2 and NS5A or that the tagging methods used produced steric clashes (Stapleford & Lindenbach, 2011). The latter may be less likely as NS2 interactions with E2 and NS3 were consistent across studies using a variety of tagging methods. It has been reported that the NS2:NS5A interaction is dependent upon p7 (Ma *et al.*, 2011; Tedbury *et al.*, 2011).

Despite the inference of a core:NS2 interaction from an NS2-based rescue mutant of a non-infectious core mutation (Murray *et al.*, 2007), no physical interaction has been observed between core and NS2 by numerous co-immunoprecipitation studies.

1.8.6 High molecular weight NS2-containing complexes

High molecular weight complexes were obtained from virus-infected cells and, following blue-native polyacrylamide gel electrophoresis (BN-PAGE), they were found to contain a proportion of E1, E2, NS2 and NS3 (Stapleford & Lindenbach, 2011). The formation of high molecular weight complexes containing both E1 and E2 have been reported previously, although the presence of NS2 was not addressed (Vieyres *et al.*, 2010). The glycoproteins present within the NS2-containing complexes were sensitive to Endo H glycosidase suggesting that they contained unmodified, high-mannose glycans indicative of retention in a pre-Golgi compartment (Stapleford & Lindenbach, 2011). This was consistent with the identification of ER-retention signals within the glycoproteins (Cocquerel *et al.*, 1999; Cocquerel *et al.*, 1998). The presence of p7 and NS4A within high molecular weight NS2-containing complexes was not determined. E1 has been shown to interact with core (Lo *et al.*, 1996) but core was absent from NS2-containing complexes, as was NS5A (Stapleford & Lindenbach, 2011). As not all of the protein species were incorporated into these complexes the core:E1 interaction may well involve a

different sub-population of E1, such as E1 that is not dimerised with E2, or perhaps the core:E1 interaction occurs at a later stage in virion assembly following dissociation from the NS2-containing complexes.

Formation of NS2-containing complexes was impaired by the insertion of a second IRES between E2/p7, p7/NS2 and NS2/NS3 and, although replication was not affected by the introduction of a second IRES, $TCID_{50}$ was reduced by 2 logs (Stapleford & Lindenbach, 2011). IRES insertion between p7 and NS2 disrupted the protein:protein interaction of these two proteins, similarly a reduction in the NS2:NS3 interaction was observed when NS2 and NS3 were separated by an IRES (Stapleford & Lindenbach, 2011). Mutation of residues key to E1/E2 hetero-dimerisation demonstrated that E1 is incorporated into the complexes via interaction with E2 (Stapleford & Lindenbach, 2011). It was reported that these complexes did not form in the uninfectious Δp7 virus suggesting p7 plays a key role in complex formation, however no data was presented to substantiate this (Stapleford & Lindenbach, 2011).

Analysis of previously characterised NS2 mutations that caused a reduction in virus infectivity (Phan *et al.*, 2009) revealed that K27A, W35A, Y39A and E45K severely reduced or abrogated the NS2:NS3 interaction, while the mutations K81A and P89A disrupted the interactions between NS2 and the glycoproteins (Stapleford & Lindenbach, 2011). The Q221L mutation in NS3 and the A78T mutation in E1, which rescued the NS2 K27A and the NS2 K81A mutations respectively, failed to restore the impaired protein interactions, indeed the A78T mutation in E1 appeared to further impair the ability of NS2 to interact with the glycoproteins (Stapleford & Lindenbach, 2011). As this implied that not all of the NS2 interactions were essential for virion assembly it was hypothesised that NS2 may ordinarily alter the function of the glycoproteins and protease complexes, but that these compensatory mutations bring about these alterations in an NS2-independent manner (Stapleford & Lindenbach, 2011).

1.8.7 Role in Replication

NS2 is not required for replication of the viral genome and its inclusion in SGRs had a negative effect on HCV RNA replication in comparison to NS3-5B replicons (Lohmann *et al.*, 1999; Tedbury *et al.*, 2011). Mutation of 27 highly conserved NS2 residues in a mono-cistronic Jc1 virus (Phan *et al.*, 2009) and 14 residues in both a mono and a bi-cistronic J6/H77/JFH virus (Dentzer *et al.*, 2009) found that replication was only inhibited when NS2/3 cleavage was impaired, providing further evidence that that NS2 is unlikely to play a role in genome replication. The apparent negative regulation of RNA replication by NS2 observed in the

replicon systems may be resultant of NS2 interacting with the other non-structural proteins and directing them to prospective sites of virus assembly.

1.8.8 Role in Assembly

NS2 has been shown to play a vital role in virus assembly (Jones *et al.*, 2007) and was hypothesised to form a link between replication complexes and virus factories (Dimitrova *et al.*, 2003; Pietschmann *et al.*, 2006). Naturally occurring inter-genotypic chimeric viruses have been isolated from patients where recombination has taken place within NS2 (Kalinina *et al.*, 2002; Noppornpanth *et al.*, 2006). Numerous groups have generated viable chimera in the lab, based upon the JFH-1 isolate (Lindenbach *et al.*, 2005; Lindenbach *et al.*, 2006; Mateu *et al.*, 2008; Pietschmann *et al.*, 2006; Scheel *et al.*, 2008; Yi *et al.*, 2007). A study generating and characterising chimeric viruses, where the JFH-1 structural proteins were replaced with those of another isolate, found that all but one virus (a genotype 3a/2a chimera) produced highest titres when the point of recombination was after residue 33 in NS2 (Pietschmann *et al.*, 2006). These phenomena have led some to speculate that the N-terminal domain of NS2 forms interactions with the structural proteins, while the catalytic domain may form interactions with the non-structural proteins.

The exact mechanism by which NS2 contributes to infectivity is poorly understood. It has been demonstrated that both the *trans*-membrane and catalytic domains are essential, although catalytic activity is dispensable (Jirasko *et al.*, 2008; Jones *et al.*, 2007). Indeed, many of the conserved residues within NS2 appear to have roles in efficient virus production independent of NS2 protease activity, such as S168. Yet, viruses encoding lethal NS2 deletions (aa 10-62, aa32-96 and aa 97-217) can be *trans*-complemented by SGRs encoding wt NS2 (Jirasko *et al.*, 2008).

As mentioned in section 1.8.5; the role of NS2 protein interactions remains unclear. The delayed virus kinetics observed when NS2 and NS3 were separated by an IRES (Jones *et al.*, 2007; Stapleford & Lindenbach, 2011) could be resultant of impaired interactions between the two proteins (Stapleford & Lindenbach, 2011). No compensatory mutation is known of that rescues an attenuated NS2 mutant with an impaired NS2:NS3 interaction, without concomitantly partially recovering the NS2:NS3 interaction. Therefore this interaction may be pivotal to infectious virus production.

NS2 expressed in isolation has a diffuse cellular distribution and co-localises with ER markers (Franck *et al.*, 2005; Kim *et al.*, 1995; Yamaga & Ou, 2002; Yang *et al.*, 2006). In virus-expressing cells NS2 and E2 co-localise early on with diffuse, cytosolic distributions (Jirasko *et al.*, 2010; Yi

et al., 2009) with the formation of more punctate structures between 48h and 72h (Ma *et al.*, 2011; Popescu *et al.*, 2011; Tedbury *et al.*, 2011). A similar pattern, but with a lower degree of co-localisation, was also observed between NS2 and NS3 (Jirasko *et al.*, 2010). NS2 in these punctate structures co-localises with lipid droplets (LD) and NS5A, although NS5A associates with LD independently of NS2 (Jirasko *et al.*, 2010; Ma *et al.*, 2011; Popescu *et al.*, 2011). No co-localisation has been observed between core and NS2 around LD (Jirasko *et al.*, 2010; Popescu *et al.*, 2011; Yi *et al.*, 2009). As core has been shown to accumulate on lipid droplets (Barba *et al.*, 1997) it was hypothesised that NS2 facilitates assembly by conveying the glycoproteins and aspects of the replication complex to sites of virion assembly on lipids droplets.

Exchanging of the putative first TMD of NS2 within the full length JFH-1 isolate with that of another isolate reduced infectivity and resulted in the recruitment of E2 and NS3 to LD being blocked and severely impaired, respectively (Jirasko *et al.*, 2010). The ability of NS5A to co-localise with LD was unaffected. Introduction of rescue mutations restored the association of E2 and NS3 with LD, suggesting that NS2 mediates the association of these proteins with LDs (Jirasko *et al.*, 2010). Furthermore, deletion of the TMD of E2, deletion of p7 and mutation of the p7 basic-loop all abolished the localisation of NS2 and E2 to punctate structures, suggesting a complex interaction between E2, p7 and NS2 (Ma *et al.*, 2011; Popescu *et al.*, 2011).

It has been proposed that while nucleocapsid assembly may take place on lipid droplets in the absence of NS2 (Miyanari *et al.*, 2007), addition of the glycoproteins and particle maturation may occur at the ER or ER-derived sites in proximity to NS2 (Jirasko *et al.*, 2010; Stapleford & Lindenbach, 2011). Assembly defective mutants had markedly less NS2 associated with lipid droplets, but this association was partially recovered by rescue mutations (Jirasko *et al.*, 2010).

A recent report observed that core is predominantly localised to lipid droplets in the JFH-1 virus, but it co-localises to the ER with the virus glycoproteins and calnexin in the more infectious Jc1 virus (Boson *et al.*, 2011). It was discovered that this differential localisation influences the infectivity of the HCV as long-term culturing of the wt JFH-1 virus identified higher titre mutants in which core was re-distributed to ER membranes. Co-expression of core protein from each isolate with core-NS2 from the opposing isolate inverted the core localisation phenotypes, and further co-expression studies identified NS2 and p7 to be responsible for core localisation. This role was shown to be dependent upon interactions between the C-terminal helix of p7 and aa 5-43 of NS2. Mutation of the p7 basic-loop blocked

the influence of p7 and NS2 on core re-distribution (Boson *et al.*, 2011). This evidence favours the hypothesis that while lipid droplets may be a reservoir for core and other virus proteins, assembly takes place at the ER (McLauchlan, 2009). The mechanism by which NS2 and p7 are able to modulate core localisation is confused by the reports that NS2 and p7 do not co-localise with or immunoprecipitate core and that core was not detected in high molecular weight NS2-containing complexes (Jirasko *et al.*, 2010; Ma *et al.*, 2011; Popescu *et al.*, 2011; Stapleford & Lindenbach, 2011).

1.8.9 Cellular interactions

NS2 appears to impose a generally suppressive effect on promoter-controlled gene expression (Kaukinen *et al.*, 2006). Reporter gene expression controlled by hCMV, HBV and simian virus – 40 (SV-40) promoters, as well as the cellular tumour necrosis factor-α (TNF-α), ferrochelatase and nuclear factor κB (NF-κB) regions promoter/enhancer sequences, was inhibited by NS2 expression. Over-expression of core, E1 and p7 did not inhibit reporter expression (Dumoulin *et al.*, 2003). General inhibition of cellular promoters and suppression of gene expression controlled by the NF-κB promoter by NS2 compared to empty vector has also been described by Yang *et al*., (Yang *et al.*, 2006). Yang and colleagues also found that NS2 expression reduced cell proliferation and this was attributed to NS2-mediated down regulation of cyclin A (Yang *et al.*, 2006). However, conflicting data describing NS2-mediated stimulation of interleukin - 8 (IL-8) production, through activation of the NF-κB promoter (Oem *et al.*, 2008a), and fatty acid synthase (FAS) gene expression, as a result of up regulation of the transcription factor sterol regulatory element-binding protein 1c (SREBP-1c) by NS2 (Oem *et al.*, 2008b), have been reported. As all of these studies were conducted using over-expressed NS2 in isolation from the other viral proteins the validity of these observations in the context of a natural infection remains to be demonstrated.

Residues 99-139 of NS2 were shown to interact with the cellular pro-apoptotic protein cell death-inducing DNA fragmentation factor (DFF) 45-like effector (CIDE)-B using a yeast two-hybrid system (Erdtmann *et al.*, 2003). NS2 co-localised with CIDE-B by confocal immunofluorescence (IF) microscopy and it was shown that NS2 interacts specifically with the liver-associated CIDE-B protein and not the non-hepatic homologue CIDE-A. NS2 was shown to prevent CIDE-B-induced cytochrome *c* release from mitochondria, and thus inhibit apoptosis (Erdtmann *et al.*, 2003). Unfortunately, an attempt to verify this interaction in virus-infected cells by immunoprecipitation of tagged-NS2 was unsuccessful (Stapleford & Lindenbach, 2011).

1.9 NS2 in related viruses

HCV NS2 does not appear to have a homologue within the *Flaviviruses*. The sequence upstream of NS3 in the *Flaviviruses* encodes two proteins termed NS2A and NS2B. Catalysis of the NS2A/NS2B and NS2B/NS3 junctions is carried out by the NS3 protease. NS2B is thought to span the membrane twice and acts as the protease co-factor (Falgout *et al.*, 1991). NS2A is thought to span the membrane three times and is a vital component of the replicase (Khromykh *et al.*, 2000). NS2A has been shown to undergo an additional cryptic cleavage event, mediated by NS3, to produce the NS2Aα protein (Nestorowicz *et al.*, 1994).

The *Pestiviruses,* of which BVDV is the prototype species, are more closely related to the *Hepaciviruses* and BVDV NS2 shares common features with HCV NS2. The N-terminus of BVDV NS2 is cleaved by signal peptidase and NS2 catalyses auto-cleavage of the NS2/3 junction (Lackner *et al.*, 2004); however this event requires the coordination of the cellular factor J-domain protein interacting with viral protein (Jiv) at two distinct sites (Lackner *et al.*, 2005; Lackner *et al.*, 2006; Rinck *et al.*, 2001). Efficient cleavage of the NS2/3 junction correlates with replication and is essential for the cytopathogenicity of BVDV (Rinck *et al.*, 2001). The dependence of NS2/3 cleavage on cellular Jiv is pivotal to the establishment of a persistent infection by BVDV (Lackner *et al.*, 2005).

In contrast to HCV the NS2 protein in *Pestiviruses* contains the zinc coordination motif rather than NS3 (Lackner *et al.*, 2004). A further difference between the genera is that although *Pestivirus* NS2 functions in virus assembly, it does so only as part of an uncleaved NS2/3 precursor which is essential for infectious virus production (Agapov *et al.*, 2004).

In common with HCV NS2, expression of classical swine fever virus (CSFV) NS2 in isolation was found to induce ER-stress, activate transcription via NF-κB and induce an S-phase cell-cycle arrest by inducing cyclin A degradation (Tang *et al.*, 2010). A recent report proposed that CSFV NS2 contains four TMDs and contains two signal sequences using a combination of computational prediction and deletion mapping (Guo *et al.*, 2011).

1.10 Mechanisms for protein targeting to membranes

Nascent mammalian membrane proteins destined for cellular organelles or the cell surface are translated at the ER membrane. The growing peptide encodes an N-terminal signal peptide that is recognised by the signal recognition particle (SRP) that temporarily stalls translation and targets the ribosome to the ER membrane. This complex then joins with the translocon to insert the signal peptide into the membrane and reinitiate translation of the peptide into the

translocon, into the ER lumen or the membrane bi-layer (Grudnik *et al.*, 2009; Saraogi & Shan, 1111). The translocon in mammalian cells consists of three or four copies of the sec61 heterotrimer, which form an aqueous pore within the ER membrane (Hanein *et al.*, 1996), and the ATPase and pore-gating protein BiP (Hamman *et al.*, 1998).

Targeting of nascent polypeptides to membranes independently of a signal peptide has been reported in *E. coli* (Bornemann *et al.*, 2008), although this was speculated to be a compensatory mechanism for the lack of elongation arrest function which is provided by the Alu domain of the mammalian SRP (Saraogi & Shan, 2011).

The manner in which *trans*-membrane proteins are inserted into the ER membrane is thought to occur by one of two routes. The '*en masse*' hypothesis asserts that all of the TMD of a nascent polytopic membrane protein are translated and held in the lumen of the translocon and released into the membrane as a mature protein (Borel & Simon, 1996). The spatial constraints of the translocon pore has led some to exclude the *en masse* model in the case of large polytopic proteins; in favour of another model that where TMDs exit the translocon 'sequentially' upon formation or perhaps upon reaching a limit of the translocon pore size (Mothes *et al.*, 1997).

1.11 Systems for studying protein topology

Determining the topology of *trans*-membrane proteins is fundamental to our understanding of their function. Attempts to resolve the structure of *trans*-membrane proteins are hampered due to their inherent hydrophobic nature, making it difficult to achieve the solubility that is required for crystallisation. Consequently, numerous different biochemical techniques have been developed to study the topology of membrane proteins. A selection of cell-based and in vitro topology investigation techniques are described below.

1.11.1 Topology prediction

Numerous computational programs are available and produce varying forms of topology prediction based upon database information, relative hydropathy values assigned to each amino acid (Kyte & Doolittle, 1982)(Appendix 2), sequence alignment with topologically defined homologues and more global traits like the 'positive inside rule' (Hartmann *et al.*, 1989; Heijne, 1986) whereby positively charged residues are four times more likely to reside on the cytosolic face of a membrane. More complex modelling algorithms have been generated using hidden Markov models (HMM) with Bayesian networks and statistically averaging the outputs of numerous individual programs. These programs do not take into

account intra-protein interactions, such as salt bridges, or lipid-protein interactions, such as dynamic membrane domains or amphipathic helices (Ott & Lingappa, 2002). Prediction of prokaryote membrane proteins is most accurate due to the number of proteins on which to base, confirm and refine predictions. Eukaryote proteins have an added diversity in that the lipid composition of membranes of different cellular compartments varies, affecting the pKa of local protein side-chains (van Klompenburg *et al.*, 1997), and this can significantly influence the topology and targeting of *trans*-membranes proteins (Bogdanov *et al.*, 2005).

The membrane-spanning domains of viral proteins provide an additional challenge to topology prediction as they may not follow exactly the same rules as host membrane proteins such as helix length. They may also be intentionally different in order to manipulate the cellular environment such as in generating the 'membranous web' characteristic of many positive-sense RNA virus families (den Boon & Ahlquist, 2010). As result, these programs are used as a basis for further biochemical analysis.

1.11.2 Fusion with reporter proteins

Reporter proteins such as alkaline phosphatase (PhoA), β-galactosidase (β-gal) and β lactamase (β-lac) are commonly used for the determination of integral membrane-protein topology (Bogdanov *et al.*, 2005). Reporter molecules are expressed as fusions to the C-terminus of a series of C-terminal truncations of the protein of interest. C-terminal truncations are typically used as the orientation of the N-terminal TMD was proposed to directly influence the orientation of subsequence TMDs within the protein (Hartmann *et al.*, 1989). Enzyme activity can be monitored by colourimetric or chemiluminescent assay.

Dual reporter systems involve fusions of two reporters with complementary activity such that they are active on opposing sides of the target membrane. Alkaline phosphatase requires oxidation in the periplasm in order to become active, while β-gal appears to become trapped in the periplasmic membrane upon translocation as a fusion to a transmembrane domain (Georgiou *et al.*, 1988). More recently PhoA and β-gal have been used as combined fusion where the PhoA gene is in frame with the β-galactosidase α-subunit of (LacZα). It this system the LacZα fragment is not toxic to bacteria but is still inactive if it is localised to the periplasm as its activity requires complementation with Ꞷ-fragment localised exclusively to the cytosol (Alexeyev & Winkler, 1999; Alexeyev & Winkler, 2002; Korres & Verma, 2004).

Antibiotics can be used as reporter fusion partners. The gene product of *bla* - β lactamase, is active in the periplasm but not in the cytoplasm. Conversely the product of the *cat* gene, chloramphenicol acetyltransferase, confers resistance when it is expressed in the cytoplasm

where there is a source of acetyl co-enzyme A (CoA) which is required for inactivation of the antibiotic.

The main limitation of C-terminal reporter fusions is that they do not accommodate the possibility of long range or inter-helix interactions which may be required for correct topology. The 'sandwich' reporter system whereby the reporter is inserted into the full length protein has been found to give a more accurate topology (Ehrmann *et al.*, 1990). This approach still suffers limitations as insertion of a non-native protein within the protein sequence may alter protein folding (of both the protein of study and the reporter which could lead to false results) and membrane insertion.

1.11.3 Glycosylation site insertion

Introducing glycosylation sites is a commonly used method whereby the coding sequence is altered to introduce the conserved glycosylation sequence Asn-X-Ser/Thr, where X is any residue except proline, or a peptide loop domain containing a known glycan acceptor site. *N*-linked glycosylation is highly dependent upon the size of the luminal loop as oligosaccharyltransferase (OST) requires 12 residues upstream and 14 residues downstream between the Asn-acceptor residue and the reticular membrane in order to catalyse the reaction (Nilsson & von Heijne, 1993). If the asparagine (Asn) residue is suitably exposed within the lumen of the ER it will be *N*-glycosylated by the addition of a polysaccharide ($Glc_3Man_9GlcNAc_2$) to its amine group by OST. The addition of this oligosaccharide results in a molecular weight increase that is detectable by SDS PAGE. The two *N*-acetylglucosamine molecules (GlcNAc) form the chitobiose core and are fused directly to the Asn-acceptor residue. The glycan moiety of a protein trafficked via the Golgi can undergo maturation where the high-mannose chain is removed and replaced with complex or hybrid oligosaccharide chains. Endoglycosidase H (Endo H) only cleaves within the chitobiose core of high mannose and some hybrid oligosaccharides (Maley *et al.*, 1989). Peptide: *N*-Glycosidase F (PNGase F) cleaves nearly all types of *N*-linked glycan chains except those containing core α1-3 fucose (Tretter *et al.*, 1991) and cleaves between the Asn-acceptor residue and the chitobiose core (Maley *et al.*, 1989). As glycosylation occurs exclusively in the ER, this can be used to determine the orientation of a protein domain about the ER membrane.

1.11.4 Partial membrane permeabilisation

Partial or selective membrane permeabilisation can be achieved using a number of compounds. Digitonin is a detergent produced by *Digitalis purpurea* while streptolysin-*O* (SLO), a toxin obtained from *Streptococci* and *Escherishia coli* (*E. coli*). Both digitonin and SLO are

selective for cholesterol-rich membranes and produce pores of variable sizes. At the right concentration these toxins are specific for the plasma membrane (Bhakdi *et al.*, 1993; Dunn & Holz, 1983) so optimisation is required to achieve pores large enough for antibody infiltration into cells without compromising cellular integrity or intracellular compartments.

Other methods of partial permeabilisation of the plasma membrane include dilute concentrations of detergents such as Tween20 or saponin, repeated cycles of freeze thaw action or other toxins such *E. coli* haemolysin and *Staphylococcus aureus* α-toxin.

Digitonin and SLO treatment has recently been combined with confocal IF microscopy to examine the topology of human papilloma virus (HPV) early protein 5 (E5) which localises to ER membranes and shown to form a three TMD topology using this method (Krawczyk *et al.*, 2010).

1.11.5 Protease protection studies

Protease protection studies can be performed using trypsin or proteinase K treatment of labelled protein expressed *in vitro* in the presence of microsomal membranes (MM) and with or without membrane solubilisation. Any remaining protein can be visualised by SDS PAGE gel. This system can be used in combination with N and C-terminal truncations to identify the protected domain. Protease protection was used in a yeast expression system in combination with glycosidase treatment and reporter fusion activity to characterise the unstable nature of TMD 6 of the cystic fibrosis transmembrane conductance regulator (CFTR) (Tector & Hartl, 1999).

Protease protection method has twice been applied to the study of NS2 topology *in vitro* and on both occasions the majority of the protein, including the C-terminus, was determined to be oriented to the ER lumen (Santolini *et al.*, 1995; Yamaga & Ou, 2002). This method can also be applied, in combination with selective permeabilisation of the plasma membrane, to cell-expressed proteins resident on intracellular membranes. Interestingly it has recently been used to investigate the topology of NS2 expressed from the viral polyprotein and it was found that in this system the majority of the protein, including the C-terminus, was localised to the cytosol (Ma *et al.*, 2011).

1.11.6 Fluorescence protease protection (FFP)

FFP is in combination with partial permeabilisation and protease digestion; GFP, or fluorophore derivatives, are expressed as terminal fusions to the protein of interest and viewed by microscopy (Lorenz *et al.*, 2006a). Cell plasma membranes are selectively

permeabilised with digitonin following fixation and then trypsin is added. Cells are viewed with and without trypsin treatment and if fluorescence remains then the fluorophore is inferred to be luminal oriented. This can be confirmed by immunofluorescence labelling of control proteins and fully permeabilising cells with detergent, followed by addition of trypsin, which should abolish any fluorescence. FFP has been used in the topology investigation of vesicle surface kinase 3 (VSK3) (Fang *et al.*, 2007).

1.11.7 Antibody Epitope insertion

Epitope insertion involves the systematic insertion of an antibody recognition sequence throughout the protein avoiding potential disruption caused by C-terminal deletions and reducing the potential impact caused by insertion of large reporter enzymes. In a polytopic membrane protein the orientation of the epitope about the membrane will determine its accessibility to antibody detection following selectively permeabilising cells for IF microscopy. Using a small epitope that is unlikely to disrupt topology is best. Similar to FFP this technique requires the selective permeabilisation of the plasma membrane followed by indirect IF detection of the epitope tag which can be blocked by proximity of the epitope to the membrane bi-layer or by protein structure.

1.11.8 Specific cleavage motif insertion

A consensus cleavage motif for a specific proteinase, such as tobacco etch virus (TEV) protease, is systematically inserted throughout the protein. This system is typically used *in vitro* where the protein is translated in the presence of MM and the enzyme is added either with or without membrane permeabilisation in order that the orientation of the cleavage site can be determined. This method has met with limited success as access to the cleavage site can suffer the same impedances as epitope insertion analysis (Henderson *et al.*, 2004).

1.11.9 *Trans*-Membrane Domain Trapping

Sugano *et al*., (Sugano *et al.*, 1998) described a method of screening cDNAs for the presence of *trans*-membrane domains that they termed '*trans*-membrane domain trapping'. This method involves the expression of cDNA from a single ORF in frame with a signal peptide and the IL-2 receptor alpha-chain as a detection epitope. If the cDNA encodes a TMD it presents the IL-2 domain to the extracellular face of the plasma membrane, if not the epitope domain is secreted into the extracellular media or retained within the cell. Surface expression of the IL-2 epitope is determined by indirect immunofluorescence microscopy.

1.11.10 SCAM™

Another technique is substituted cysteine accessibility method for *trans*-membrane topology (SCAM™), which involves replacing single amino acids with cysteine residues (Bogdanov *et al.*, 2005). Maleimides react with the thiol groups to form a stable thioester bond that is non-reducible by β-mercaptoethanol or dithiothreitol (DTT). Identification of modified cysteine containing proteins can be done by various methods based on the principle of fusion of the maleimide with a radiolabel, a biotin-tag, a fluorophore or as an appreciable size increase following SDS PAGE analysis. Cytoplasmic cysteine residues can be determined by partially-permeabilising the cell membrane and introducing a membrane impermeable thiol-specific reagent. Direct identification of luminal regions can be achieved by quenching thiol reactive groups (with a thiol reagent that is transparent in the detection phase) in partially permeabilised cells followed by complete permeabilisation and addition of a detectable thiol reagent.

1.12 Aims

At the undertaking of this work little was known of the membrane topology of NS2 in cells and that which had been reported was contradictory to known localisation of the cleavage events that release mature NS2 from the HCV polyprotein. It had recently been demonstrated that deletion of NS2 from the virus prevented infectious virion production. However, the manner in which NS2 facilitates virus particle assembly, and the significance of NS2 membrane topology has in this role, was similarly unknown.

In this study a series of C-terminal truncations of NS2 were generated to investigate the membrane topology of NS2. In response to these findings, and those reported within the field, a further set of truncations were generated to investigate the membrane targeting, subcellular localisation and protein interactions of NS2, specifically the NS2:E2 interaction and NS2 oligomerisation. These studies were carried out in a mammalian cell expression system to validate and expand upon previously published *in vitro* data regarding NS2 topology and to further explore the role of the NS2 *trans*-membrane domain in forming protein:protein interactions with the other viral proteins.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Bacterial strains and culture conditions

Escherichia coli (E. coli) DH5α genotype: F- ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^- (Invitrogen) were used for cloning.

Bacteria were cultured in Luria-Bertani (LB) liquid medium (1% w/v bacto-tryptone, 0.5% w/v Bacto-yeast extract, 171 mM NaCl, pH7.0) or isolated as single clones on solidified LB containing 1.5% w/v bactoagar, at 37°C with sufficient aeration. Cultures were supplemented with 100 μg/ml ampicillin or 50 μg/ml kanamycin where appropriate.

Chemically competent bacteria were produced by the Inoue method (Sambrook, 2001)(section 2.2.1.12).

2.1.2 Mammalian cell lines and culture conditions

Protein expression was carried out in Huh7 cells, a human hepatocellular carcinoma cell line susceptible to propagation of HCV subgenomic replicons and full-length virus; and COS7 cells, a simian cell line endogenously expressing the SV40 large T antigen which increases plasmid DNA copy number by episomal plasmid replication.

Huh7.5 cells, a derivative of the Huh7 cell line shown to support higher levels of replication of replicons (Lohmann *et al.*, 1999), stably harbouring the SGR NS3-5B and SPp7-B (Tedbury *et al.*, 2011) were maintained in 250 ng/ml G418.

Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma) supplemented with 10% v/v foetal bovine serum (FBS), 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine and incubated at 37°C in a humidified incubator with 5% CO_2. Culture medium for Huh7 cells was additionally supplemented with 1% v/v non-essential amino acids (Cambrex). Huh7 cells used to culture virus were supplemented with 10 mM HEPES.

Cells were passaged by washing in phosphate buffered saline (PBS) pH 6.8 prior to incubating with trypsin-EDTA solution pH 8 (Sigma Aldrich) to disassociate cells from growth surface. Cells were mechanically resuspended in culture medium, quantified using a haemocytometer and seeded into fresh culture medium at the desired density.

2.1.3 Virus sequences

HCV sequences:

EUHK2 - genotype 6a, accession number: Y12083

EUH1480 - genotype 5a, accession number: Y13184

ED43 - genotype 4a, accession number: GU814265

NZL1 - genotype 3a, accession number: D17763

JFH-1 - genotype 2a, accession number: AB047639

HC-J6 - genotype 2a, accession number: D00944

J4L6S - genotype 1b, accession number: AF054247

H77 - genotype 1a, accession number: AF011753

Con1- genotype 1b, accession number: AJ238799

QC69 - genotype 7a, accession number: EF108306

Canine hepacivirus (CHV), accession number: JF744991

GB virus B (GBV-B), accession number: AF179612

2.1.4 Plasmids

pSEAP2-control (Clontech) was used as the source of the secreted alkaline phosphatase (SEAP) gene and as a control in SEAP expression assays in mammalian cells. It carries an ampicillin resistance gene.

pEBBLacZ and pcDNA3.1LacZ were kind gifts from Dr T. Tuthill and Dr A. MacDonald and were used as controls in β-galactosidase (β-Gal) expression assays. pEBBLacZ was used as the source of the LacZ gene. They both carry an ampicillin and a neomycin resistance gene.

pJFH-1 was kindly provided by Dr T. Wakita. It carries an ampicillin resistance gene and was the source of HCV genotype 2a sequence segments. Culturing of bacteria containing pJFH-1 was carried out at 30°C.

pCRBlunt (Invitrogen) was used as a cloning intermediate for polymerase chain reaction (PCR) products. It carries a kanamycin resistance gene.

pcDNA3.1(+) (Invitrogen) was used for protein expression in mammalian cells. It carries an ampicillin and a neomycin resistance gene.

peGFP-N1 and peGFP-C2 (Clontech) were used for the expression of GFP fusion proteins. Both encode resistance to kanamycin and neomycin.

The following previously characterised plasmids were used: pSG5-NS5A-Flag (GB virus - B) (Mankouri *et al.*, 2008b), peGFPC1Eps15, pSG5-NS5A (JFH-1) and pSG5-NS5A-eGFP (JFH-1) (Mankouri *et al.*, 2008a), pFBM(JFH-1)core-p7 (Adair *et al.*, 2009), peGFPC1(IBV)N-protein (Emmott *et al.*, 2008)

pFBM(Con1)core-p7 was a kind gift from Dr C. McCormick and peGFPC1(HCV)NS4B was a kind gift from Dr S. Gretton.

2.1.5 Antibodies

Anti-NS2 serum (Rabbit) (A kind gift from Dr A. Pause) was used at 1:5,000

Anti-NS5A serum (Sheep) was used at 1:5,000 (Macdonald *et al.*, 2003)

Anti-NS3 serum (Sheep) was used at 1:10,000 (Aoubala *et al.*, 2001)

Anti-E2 AP33 clone (Mouse) (A kind gift from Dr A. Patel) was used at 1:2,000

Anti-placental alkaline phosphatase (PLAP) (SP15 - Abcam) was used at 1:500

Anti-β-galactosidase (Rabbit) (AB1211 - Abcam) was used at 1:2,000

Anti-GAPDH antibody (Mouse) (AB8245 - Abcam) was used at 1:30,000

Anti-ICAM-I 208 clone (Rabbit) (A kind gift from Dr D. Rowlands) was used at 1:1,000

Anti-FLAG M2 clone (Mouse) (F1804 – Sigma-Aldrich) was used at 1:5,000 for Western blotting or 1:1,000 for immunofluorescence microscopy.

Anti-neomycin phosphotransferase II antibody (06-747 - UpState) was used at 1:5,000

Anti-core serum (Sheep) was used 1:5,000

Anti-GFP raised against full-length GFP (SC-8334 - Santa Cruz) was used 1:5,000

Horseradish peroxidase (HRP) conjugated secondary antibodies; goat anti-rabbit, goat anti-mouse and donkey anti-sheep (Sigma-Aldrich) were used at 1:10,000.

Alexafluor conjugated secondary antibodies (Invitrogen) were used at 4 µg/ml for confocal IF microscopy.

All antibodies were diluted 0.1% v/v Tween20, 5% w/v milk powder (Marvel) in phosphate buffered saline (PBS) for Western blotting or 1% w/v bovine serum albumin (BSA) in PBS for immunofluorescence staining.

2.1.6 Chemicals

Phenol, phenol:chloroform, phenol:chloroform:isoamyl alcohol (25:24:1), MOPS, G418 and dithiothreitol (DTT) were obtained from BDH.

Sodium dodecyl sulphate (SDS), magnesium chloride ($MgCl_2$) and ethylenediamine tetraacetic acid (EDTA) were obtained from VWR.

Formalydehyde, Triton X-100, Z-Leu-Leu-Leu-al (MG132), ONPG, and protease inhibitors (leupeptin, pepstatin, aprotinin and AEBSF (perfabloc)) were obtained from Sigma.

Sodium chloride (NaCl), methanol, ethanol, tris (hydroxymethyl) aminomethane (Tris) and bovine serum albumin (BSA) were obtained from Fisher Thermoscientific.

N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), yeast extract and tryptone were obtained from Melford Laboratories.

PEI was obtained from Polysciences, CDPStar was from Tropix and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) was from AnaSpec.

2.2 Methods

2.2.1 DNA manipulation

2.2.1.1 Purification of DNA

For plasmid DNA analysis small scale DNA purifications from 1 ml cultures were prepared by alkaline lysis. Bacteria from a single colony were inoculated into 1 ml of liquid LB medium containing appropriate antibiotic and cultured over night at 37°C. Bacteria were pelleted by centrifugation and resuspended in 50 µl liquid LB medium followed by the addition of 300 µl TENS lysis solution (10 mM Tris, 1 mM EDTA, 100 mM sodium hydroxide, 0.5% sodium dodecyl sulphate (SDS)) and thoroughly mixed. Lysates were neutralised with 150 µl 3 M sodium acetate pH 5.3 and cellular debris was removed by centrifugation followed by transferral of the supernatants to a fresh tube containing 900 µl ethanol (ice-cold to aid precipitation). Samples were mixed thoroughly and precipitates were collected by centrifugation, resuspension in 500 µl 70% ethanol v/v and repeat centrifugation. The precipitated DNA was finally resuspended in 50 µl of 10 µg/ml RNaseA. All centrifugation steps were carried out at 16,100 x *g* for 5 minutes.

For large scale purifications of 400 ml cultures the Qiagen Maxiprep kit was used as per the manufacturer's instructions. For high purity small scale DNA preparations the Qiagen Miniprep kit was used.

2.2.1.2 Quantification of nucleic acids

Quantitation of nucleic acid preparations was performed by comparison of nucleic acid samples with markers of known concentration (Millenium Markers – Millipore; for RNA

samples and Hyperladder I – Bioline; for DNA samples) by agarose gel electrophoresis or by measuring the absorbance at 260/ 280 nm using a NanoDrop 1000 spectrophotometer (Thermo Scientific) according to the manufacturer's instructions.

2.2.1.3 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify regions of the HCV JFH-1 sequence or the SEAP and β-gal genes using the high-fidelity polymerase pfu Turbo (Stratagene). PCR reaction mixtures comprised 1.25 mM deoxynucleotide triphosphates (dNTPs), 10-20 ng template vector DNA, 50 nM of each primer, 1.25 U Pfu Turbo DNA polymerase in a reaction volume of 50 µl Pfu Buffer (Stratagene), 5% v/v dimethyl sulphoxide (DMSO). Reactions were initially denatured at 95°C for 30 seconds then subjected to 30 seconds of denaturing at 95°C, 30 seconds of annealing at 65°C and extension at 70°C for an appropriate length of time. This was repeated for 20 cycles. An additional extension step of 10 minutes was used immediately prior to termination of amplification reactions. Gradient PCRs were used for PCR optimisation; the annealing temperature was varied from 55 – 70°C.

2.2.1.4 QuikChange Mutagenesis

Single amino acid mutations were introduced by QuikChange mutagenesis. Complimentary primers were designed with 17-20 nucleotides of homologous sequence flanking non-homologous nucleotides. PCR was carried out as described in section 2.2.1.3. Following PCR reactions, parental DNA was degraded by Dpn I treatment. 20 U units of Dpn I and 10% v/v Buffer 4 (NEB) to PCR volume were added and samples incubated at 37°C for 1 hour. PCR reactions were then transformed into competent bacteria as described in section 2.2.1.13.

2.2.1.5 Oligo-Cloning

When oligo cloning was used complimentary oligonucleotides were synthesised with 5' phosphate modifications and with appropriate over-hangs corresponding to the restriction endonuclease used in the cloning strategy. An 50 nM of each oligo was added to annealing reaction buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl and 1 mM EDTA and incubated at 95°C for 10 minutes and allowed to equilibrate to room temperature over 1 hour. Annealed oligos were then incubated on ice prior to ligation (section 2.2.1.11) into pre-digested vector, or stored at -20°C.

2.2.1.6 Removal of protein from DNA solutions

1 volume of dH_2O and 2 volumes of phenol:chloroform:isoamylalcohol (25:24:1) (equilibrated with Tris pH 8.0) were added to DNA-protein mixtures and vortexed thoroughly. Samples were

then centrifuged at 16,100 x *g* for 5 minutes and the top, aqueous phase was collected as the DNA containing fraction.

2.2.1.7 Restriction digestion of DNA

Restriction digests were carried out in volumes of 20 µl containing the recommended buffer and 10-20 enzyme units according to the manufacturer's guidelines. Digests were carried out for a minimum of 1 hour at the optimal temperature. For double digests using enzymes with differing optimal temperatures, the digest mix was incubated with the more heat stable enzyme for 1 hour at its optimal temperature prior to addition of the second enzyme and overnight incubation at the lower temperature.

2.2.1.8 Agarose gel electrophoresis

Gels were prepared by dissolving 0.8-1.0% w/v agarose in 0.5 x TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA) or 1 x TAE (40 mM Tris, 0.11% acetic acid, 1 mM EDTA) with the addition of 10 ng/ml ethidium bromide. Gels were submerged in gel running tanks (Bio-Rad) containing either TBE or TAE and supplemented with a further 0.5 µg/ml ethidium bromide. Samples were equilibrated using TBE loading buffer (TBE buffer in 5% v/v glycerol containing Orange G dye) or TAE loading buffer (TAE buffer in 5% v/v glycerol containing bromophenol blue dye) and separated though the gel. The size of DNA fragments was determined by comparison with Hyperladder I (Bioline) markers. DNA bands were visualised using Gene Genius Bio-imaging System.

2.2.1.9 Extraction of DNA from agarose gels

DNA bands were excised from agarose gels and frozen at -20°C for 20 minutes. The gel slices were mashed with a pipette tip and twice the volume of phenol, equilibrated with Tris pH 8.0, was added. The mixture was vortexed and incubated at -80°C for 30 minutes. Thawed samples were added to ¼ volume v/v dH_2O and vortexed thoroughly. Phases were separated by centrifugation at 16,100 x *g* for 5 minutes. The aqueous phase was collected and ethanol precipitated (section 2.2.1.10). Final DNA pellets were resuspended in an appropriate volume of dH_2O.

2.2.1.10 Ethanol precipitation of DNA

Protein-free DNA suspensions were precipitated by the addition of 2 volumes of ice-cold 100% ethanol v/v and sodium acetate pH 5.2 to a final concentration of 100 mM. Samples were vortexed briefly and incubated at -20°C for 20 minutes. DNA was precipitated by centrifugation at 16,100 x *g* for 20 minutes. DNA pellets were washed in 70% ethanol v/v and resuspended in appropriate volumes of dH_2O.

2.2.1.11 DNA ligation

Ligation reactions were carried out in 20 µl volumes of T4 Ligase buffer (Invitrogen) containing 1 Unit T4 Ligase with a 1:4 molar ratio of vector to insert. Ligations were incubated for 30 minutes on the bench with subsequent overnight incubation at 4°C. 1 µl of ligation mix was transformed (section 2.2.1.13) into 50 µl aliquot of chemically competent DH5α *E. coli* (section 2.2.1.12) following 30 minute and overnight incubations steps, and grown on solid LB-agar supplemented with the appropriate antibiotic over night at 37°C.

2.2.1.12 Preparation of chemically competent bacteria - Inoue Method

Competent bacteria were produced using the Inoue method as detailed in (Sambrook, 2001). A single bacterial culture was grown in 25 ml LB medium for 6-8 h at 37°C with aeration. This starter culture was used to inoculate 250 ml LB medium and grown at 18°C with aeration until an OD_{600} of 0.55 was reached, taking approximately 16 hours. Bacterial cells were incubated in an ice water bath for 10 minutes and then collected by centrifugation at 2,500 x *g* for 10 minutes at 4°C. The supernatant was removed and the bacterial pellet resuspended in 80 ml of ice cold Inoue transformation buffer (55 mM $MnCl_2$ 15 mM CsCl, 250 mM KCl, 10 mM piperazine-N,N'-bis(2-ethanesulfonic acid (PIPES)-NaOH (pH 6.7)). The bacteria were again pelleted by centrifugation and resuspended in 20 ml cold Inoue transformation buffer by swirling. 1.5 ml of dimethyl sulfoxide (DMSO) (Sigma Aldrich) was added and mixed by swirling. The bacteria culture was stored on ice for 10 minutes and the bacteria aliquoted into pre-cooled microfuge tubes and immediately snap frozen in liquid nitrogen, and stored at -80°C.

2.2.1.13 Transformation of Chemically Competent E. coli

Transformation of chemically competent bacteria was carried out according to the Inoue method described by (Sambrook, 2001). Briefly, 100 ng of purified plasmid DNA or 1 µl of ligation reaction mix was added to 50 µl of competent bacteria and incubated on ice for 10 minutes prior to heat-shocking at 42°C for 90 seconds. Bacteria were allowed to recover for 1 hour in 500 µl LB medium before spreading out onto LB-agar plates supplemented with the appropriate antibiotic and culturing overnight.

2.2.1.14 Colony Screening

Large scale screening of transformed bacteria colonies was used during plasmid generation. Colonies were inoculated using a Gilson tip into 10µl dH_2O in a 0.1ml PCR tube and incubated on the bench for 10 minutes, after which the tip was transferred to a 1ml culture containing the appropriate antibiotic and incubated overnight at 37°C. 10 µl of dH_2O was used as the

template for a PCR reaction which was carried as described in (section 2.2.1.3). DNA from 1ml cultures of colonies screening positive by PCR was purified as described in section 2.2.1.1.

2.2.1.15 DNA sequencing

All plasmids and constructs were verified by analytical restriction digest and sequenced by ABI 3130xl capillary sequencer analysis prior to subsequent use with appropriate primers.

2.2.2 Cloning of DNA constructs

A list of oligonucleotides used for cloning along with their sequences and usage are provided in Appendix 1.

2.2.2.1 Generation of NS2-SEAP/β-gal reporter fusions

The SEAP gene lacking its signal peptide was amplified from the pSEAP2-Control vector (Clontech) with 5' XhoI and 3' ApaI restriction sites and a 5' kozak sequence (GCGGCCATG). The amplified region was ligated into pcDNA3.1(+) to generate the nsAP control construct. The SEAP gene lacking its signal peptide was amplified from the pSEAP2-Control vector (Clontech) with 5' XhoI and 3' ApaI restriction sites and a 5' furin cleavage site (R-X-(R/K)-R) flanked by 5' and 3' flexible linkers (GGCGGAAGCAGGGGTCGACGCGGCGGAAGC). The amplified region was ligated into pcDNA3.1(+) to generate the intermediate vector pcDNA-furSEAP. The LacZ gene was amplified from the pEBBLacZ vector (a kind gift from Dr T. Tuthill) with 5' XhoI and 3' ApaI restriction sites and 5' furin cleavage site and flanker sequences as described above. The amplified region was ligated into pcDNA3.1(+) to generate the intermediate vector pcDNA-furLacZ. The signal peptide encoded by the pSEAP2-Control vector was synthesised as an oligonucleotide (composed of a 5' phosphate group, pre-digested 5' HindIII and 3' XhoI restriction sites and a 5' Kozak sequence) and ligated into the pcDNA-furLacZ vector to generate the SP-β-gal control vector.

NS2 C-terminal truncations were amplified from the pJFH-1 plasmid (a kind gift from Dr T. Wakita). All HCV sequences were amplified using the same forward primer which contained a HindIII restriction site, a Kozak sequence and sequence complimentary to the C-terminal 23 amino acids of E2 which form a signal peptide for p7. Reverse primers included a XhoI restriction site and terminated HCV sequence after amino acid 813, 840, 853, 965, 883, 900, 908, 923, 933 or 1030 of the JFH-1 polyprotein. Amplified regions of NS2 were ligated into the pcDNA-furSEAP and pcDNA-furLacZ vectors to generate the S1-10 and B1-10 constructs. The Δ63 construct was generated using the same forward primer and a reverse primer that introduced a XhoI restriction site and terminated HCV sequence at 812 (residue 62 of p7). This HCV sequence was ligated into pcDNA3.1(+) upstream of the SEAP gene (with 5' XhoI and 3'

ApaI restriction sites but lacking; a kozak sequence, furin cleavage linker and signal peptide) to generate the Δ63 construct in which SPp7 was encoded as a direct fusion to a non-secreting form of SEAP.

The S11, S12 and S13 constructs were generated by Quikchange mutagenesis. Signal peptide cleavage site was disrupted by mutation of the C-terminal amino acid of p7 from alanine to proline (5' CGGCAGGCTTATCCCCTCGAGGGCG). The furin cleavage site was disrupted by mutation of the consensus sequence (R-R-G-R) to A-R-G-R (5' GAGGGCGGAAGCGCGG GTCGACGCG).

2.2.2.2 Generation of NS2-eGFP constructs

The peGFP-N1.QC vector was generated by Quikchange mutagenesis (5' CACCGGTCGCCACCGCGGTG AGCAAGGGCG) whereby the vector-encoded eGFP translation initiation codon was mutated resulting in the generation of a unique SacII restriction site.

Domains of NS2 were amplified from the pJFH-1 vector with 5' EcoRI and 3' BamHI restriction sites and 5' Kozak sequences and ligated into the peGFP-N1.QC vector in-frame (by virtue of two additional 3' cytosine nucleotides upstream of the BamHI restriction site) with the eGFP gene with a vector-encoded linker of GDPPVAT.

2.2.2.3 Generation of NS2-FLAG constructs

NS2-FLAG constructs were generated by ligating in a synthetic, 5' phosphorylated, pre-digested, FLAG-tag oligo (p-GATCC**A**GGCGGAAGCGGCGGAAGCGACTACAAAGACGATGACGACAA G**TAA**CCGC). The oligo was composed of a 5' BamHI restriction site, an in-frame adenine nucleotide, a collagen repeat linker, the FLAG coding sequence, a STOP codon and a 3' SacII restriction site. The FLAG oligo was ligated into each of the NS2-eGFP fusion constructs to generate the corresponding NS2-FLAG constructs.

2.2.2.4 Generation of eGFP-NS2 constructs

Domains of NS2 were amplified from the pJFH-1 vector with 5' XhoI and 3' EcoRI restriction sites and 5' cytodine nucleotide (to ensure in-frame translation with the eGFP gene) and ligated into the peGFP-C2 vector in-frame with the eGFP gene with a vector-encoded linker of SGRTQISS.

The peGFP-C2-STOP vector was generated by QuikChange mutagenesis of the wt vector to encode a STOP codon as the 9th codon after the eGFP sequence (5' CGGACTCAGATCTCGAGC TAAAGCTTCGAAT TCTGCAGTC) corresponding to the first codon of the inserted NS2 sequence.

The NS2 deletion constructs Δ25-50, Δ25-92 and Δ51-92 were generated by QuikChange deletion mutagenesis.

2.2.3 RNA manipulations

All procedures involving RNA were carried out in RNA safe conditions, by treatment of surfaces and equipment with RNase Zap® (Ambion).

2.2.3.1 RNase-free water and phosphate buffered saline (PBS)

0.1% v/v diethyl pyrocarbonate (DEPC) was added to dH_2O or PBS, mixed thoroughly and incubated at 37°C overnight with shaking to inactivate any protein present. DEPC was inactivated by autoclaving.

2.2.3.2 Preparation of DNA template

Plasmids encoding the full-length and SGRs of the JFH-1 isolate were linearised by XbaI digestion overnight at 37°C. XbaI was heat-inactivated at 65°C for 20 minutes after which samples were cooled to RT then incubated for 10 minutes on ice. Nucleotide overhangs produced by linearization were removed by treatment with 1 µl mungbean nuclease (NEB) for 45 min at 30°C. Mungbean nuclease was inactivated with the addition of 0.1 % w/v SDS and the DNA purified by phenol/chloroform extraction and ethanol precipitation before being resuspended in RNase free water.

2.2.3.3 In vitro transcription of RNA

Transcriptions were performed using a T7 RiboMAX™ Express Large Scale RNA Production System kit (Promega), according to the manufacturer's instructions. Reaction mixes comprised 1x RiboMAX™ Express T7 buffer, 1 µg of linearised template DNA, 2 µl of T7 express enzyme mix and made to a volume of 20 µl with nuclease free water. Reactions were mixed gently and incubated at 30°C for 30 minutes. The DNA template was degraded by the addition of 1 U of RQ 1 RNase-free DNase and incubation for 15 minutes at 37°C. Protein was removed by the addition of an equal volume of phenol:chloroform:isoamylalcohol (125:24:1) pH 4.0 (Sigma) and vortexed for 1 minute followed by centrifugation at 16,100 x *g* for 2 minutes. The aqueous phase was transferred to a fresh tube and extracted with an equal volume of chloroform, vortexed for 1 minute and centrifuged for 2 minutes at 16,100 x *g*. The aqueous phase was again transferred to a fresh tube and RNA precipitated by the addition of 1/10th volume v/v of 3 M sodium acetate pH 5.2 and an equal volume of isopropanol. Samples were vortexed thoroughly and incubated on ice for 5 minutes followed by pelleting of precipitated RNA by centrifugation at 16,100 x *g* for 10 minutes. RNA pellets were washed in 70 % v/v ethanol and

pelleted again. RNA pellets were air dried and finally recovered in 20 µl RNase-free water (section 2.2.3.1).

2.2.3.4 RNA agarose gel electrophoresis

RNA transcripts were analysed by denaturing 3-(N-morpholino) propanesulfonic acid (MOPS) formaldehyde gel electrophoresis. 1% w/v agarose (Sigma-Aldrich) was heated until molten in MOPS buffer (40 mM MOPS, 10 mM sodium acetate, 1 mM EDTA in RNase-free H_2O (section 2.2.3.1). Upon cooling, 6.5% v/v formaldehyde was added and the mixture was poured into the casting stand and the comb added. Once set, running buffer (MOPS buffer) was added to the tank and RNA samples prepared by the addition of 10 µl loading buffer (56% v/v formamide, 10% v/v formaldehyde, 6% glycerol, a trace amount of bromophenol blue and 10 µg/ml ethidium bromide in MOPS buffer) to 1 µl of RNA transcript or 1 µl of an RNA ladder (Ambion) and headed at 65°C for 10 minutes. Gels were run at 0.1 V/mm^2 until the dye front had migrated two thirds down the gel, RNA was visualised using the UV transilluminator (Syngene) bio-imaging system and software.

2.2.4 Protein biochemistry methods

2.2.4.1 TNT coupled reactions

Coupled transcription and translation (TNT) reactions were conducted as follows. 250 ng of purified DNA was added to 8.75 µl TNT™ Quick Master Mix (Promega) supplemented with 10 µCi Trans Label (70% w/v methionine, 15% w/v cysteine, [35S]) (MP Biomedicals) and to a reaction volume of 12.5 µl. Reactions were allowed to proceed for 1 hour at 30°C before termination by the addition of 1 U of RNaseA. Samples were added to 9 volumes v/v of 2 x Laemmli buffer (10% v/v glycerol, 4% w/v SDS, 20 mM DTT, 0.01% w/v bromophenol blue and 125 mM Tris HCl pH 6.8), denatured at 65°C for 5 minutes and analysed by SDS PAGE (section 2.2.4.3). Gels were fixed in gel fixing solution (40% v/v methanol, 10% v/v glacial acetic acid) for 10 minutes followed by incubation in Amplify solution (Amersham Biosciences) for 10 minutes. Gels were placed in a gel-drying dock (Model 583 Gel Drier - BioRad) for 1 hour on top of two sheets of filter paper (Fisher Scientific). Gels were exposed to film for an appropriate length of time at -80C and developed using an automated developer (SRX-101A Autorad - Konica).

2.2.4.2 Acetone precipitation of protein

Concentration of protein suspensions was achieved by acetone precipitation. 4 volumes v/v acetone were added to samples, vortexed and incubated overnight at -20°C. Precipitated protein samples were pelleted by centrifugation at 4,000 x *g* for 10 minutes at 4°C. Pellets

were washed in 70% v/v ethanol and finally resuspended in the desired volume of appropriate buffer.

2.2.4.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples in Laemmli buffer were boiled for 5 minutes prior to separation by SDS PAGE. Gels were made with 15% resolving gel (50% 30:1 acrylamide:bis-acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% w/v sodium dodecyl sulphate (SDS), 0.1% w/v ammonium persulphate (APS), 0.01% v/v N,N,N',N'-tetramethylethylenediamine (TEMED)) and 5% stacking gel (16.7% v/v 30:1 acrylamide:bis-acrylamide, 187.5 mM Tris-HCl pH 6.6, 0.1% w/v SDS, 0.1% w/v APS, 0.01% TEMED) in Mini Protean 3 chambers (BioRad). Samples were run alongside unstained (Precision Plus - BioRad) or pre-stained markers (SeeBlue - Invitrogen) molecular weight markers to allow estimation of protein size.

2.2.4.4 Western blotting

Specific protein identification was carried out by Western blotting. Gels were transferred to polyvinylidene fluoride (PVDF) membrane (Millipore) using a semi-dry trans-blotter (BioRad) in transfer buffer (20% v/v methanol, 25 mM Tris-HCl pH 8.3, 192 mM glycine). Membranes were activated prior to transfer by washing briefly in methanol. Transfers were run for 45 minutes at 15 mA. Transferred membranes were then washed in PBS 0.1% v/v Tween20 (PBST) and blocked for >1 hour in 10% w/v dried, skimmed milk powder (Marvel) in PBST at room temperature on a rotating platform. Blocking solution was removed by washing 3 times for 5 minutes in PBST before incubation over night in primary antibody at 4°C on a rotating platform. For small volumes of antibody solution, membranes were sealed in close-fitting plastic wallets. Membranes were washed 3 times in PBST prior to incubation in appropriate HRP-conjugated secondary antibody for 1 hour. Secondary antibody was removed by washing 3 times in PBST and once in dH_2O. The enhanced chemiluminescence (ECL) system (Amersham) was used according to the manufacturer's instructions to visualise antibody reactive bands. Briefly, membranes were soaked in 1ml of ECL (1:1 ratio of Solution 1: Solution 2) for 1 minute before removal of the ECL solution and overlaying of the membrane with photographic film for varying exposures. Films were developed using a Konica SRX101 Processor.

2.2.4.5 Densitometry Analysis

Autoradiographs were scanned in and analyses using ImageJ software. Images were colour-inverted and a fixed area was boxed for analysis. Protein species were analysed for density and the mean density taken for a fixed area. A blank region of the image adjacent to the area

under analysis was taken as the background density and subtracted from the mean density values of the sample areas.

2.2.5 Tissue culture techniques

All samples were examined with a Zeiss Axioplan 2 fluorescence microscope, using a 63x oil immersion objective. Images were captured using an AxioCam Mrm CCD camera with Axiovision software.

2.2.5.1 Routine passaging of mammalian cells

Cell lines were cultured in DMEM (Sigma), supplemented with 10% v/v FBS (Sigma-Aldrich), 1 U/ml penicillin (Sigma-Aldrich), 100 µg/ml streptomycin (Sigma-Aldrich) in a humidified 5% CO_2 incubator at 37°C. Huh7 cells were supplemented with 1% v/v non-essential amino acids and Huh7 cells for use in virus experiments were further supplemented with 20 mM HEPES (GIBCO). To passage the cells the media was removed and the cell monolayer washed with PBS. Cells were incubated in 0.5 mg/ml trypsin and 2 mM EDTA in PBS (Sigma) for approximately 5 minutes at 37°C to facilitate cell detachment. Trypsin was then inactivated by the addition of complete media. Cells were counted using a haemocytometer if required and seeded at the desired density.

Huh7 cells harbouring stably replicating SGRs were cultured in the same manner, however the selection of replicon harbouring cells was achieved by supplementing culture media with 250 µg/ml G418 (Sigma-Aldrich)

2.2.5.2 Electroporation of mammalian cells

Sub-confluent Huh7 cell monolayers were detached from the growth surface using trypsin/EDTA solution, as described in the previous section, resuspended in complete media and counted. The desired number of cells was pelleted by centrifugation at 1,000 x *g* for 5 minutes at room temperature, the supernatant was discarded and the cells washed twice in 4°C RNase-free PBS. Cells were pelleted at 1,000 x *g* for 5 minutes at room temperature and resuspended at 1 x 10^7 cells/ml for SGR electroporations and at 2 x 10^7 cells/ml for full-length virus electroporations. 4 x 10^6 cells were added to a pre-chilled sterile 4 mm electroporation cuvette (Geneflow) and 2 µg of subgenomic RNA or 10 µg full-length virus RNA added directly to the cells and mixed. Electroporation was carried out using a Bio-Rad Gene Pulser at 270 Volts (V) and 0.950 Farads (F). Cells were resuspeneded in complete media using a fine tipped Pasteur pipette, and seeded at the desired density. Any equipment that came into contact with cells electroporated with full-length infectious virus was treated with 5 % Virkon before disposal.

2.2.5.3 Transfection of Huh7 cells

Cells were seeded in 12 well plates at 10^5 cells per well as with passaging and grown to ~80% confluency. 1-4 hours prior to transfection culture media was replaced with serum free media (as culture medium except lacking FBS). Transfection mixture (2 µg/ml DNA, 10 µg/ml polyethylenimine (PEI), in Optimem (GIBCO)) was prepared by thorough mixing of the DNA and PEI with the Optimem and the mixture was allowed to equilibrate at room temperature for 15-45 minutes. 250 µl of transfection mixture was inoculated per well and plates were incubated at 37°C for 48 hours.

2.2.5.4 Immunofluorescence microscopy

Cells were seeded onto coverslips in a 12-well tissue culture plate at a low density. Samples were prepared by washing coverslips three times in PBS before fixing in 3-4% paraformaldehyde (PFA) for 10 minutes on ice prior to a final PBS wash. If necessary, cells were permeabilised in 0.1% v/v Triton X-100/PBS for 10 minutes followed by three washes in PBS. Cells were blocked prior to antibody incubation in 1% w/v bovine serum albumin (BSA)/PBS for 30 minutes. Antibodies were diluted in 1% BSA/PBS and incubated at room temperature for 1 hour followed by removal of unbound antibody with three washes in PBS. Coverslips were mounted onto slides using Vector Shield (Vector Laboratories Inc.) and sealed with nail varnish (Boots).

Ice-cold PBS was used throughout the procedure; all other reagents were pre-chilled except Vector Shield.

2.2.5.5 X-Gal staining

Culture medium was discarded and cell monolayers were washed twice in PBS. Cells were fixed with 3% para-formaldehyde in PBS for 5min at 4^oC. Cells were washed twice in PBS then overlaid with 250 µl reaction mix (1 mg/ml 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-Gal - AnaSpec) dissolved in DMF, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 2 mM $MgCl_2$) and incubated overnight at 37°C. Reaction was stopped by washing in PBS and X-Gal staining was observed and imaged using the 10 x objective lens on an eclipse TS100 microscope fitted with a Digital Sight DS-L1 camera (Nikon).

2.2.6 Cell biology techniques

2.2.6.1 Lysis of mammalian cells

Cells were washed three times in PBS and lysed by the addition of Glasgow Lysis Buffer (GLB – 1% Triton X-100, 120 mM KCl, 30 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 10 mM PIPES-NaOH, pH 7.2) and incubated on ice for 15 minutes.

2.2.6.2 Isolation of post-nuclear cell suspension

6.0×10^5 Huh7 cells were seeded into 10 cm dishes 16 hours prior to transfection. 48 hours after transfection cells were washed three times in ice-cold PBS and scraped into hypotonic buffer (10 mM piperazine-N,N'-bis[2-ethanesulfonic acid] (PIPES)/NaOH pH 7.2, 0.5 mM $MgCl_2$) supplemented with protease inhibitors (Leupeptin 1 µg/ml, Pepstatin 1 µg/ml, Aprotinin 2 µg/ml, AEBSF (Perfabloc) 0.2 mM). Cells were incubated on ice for 15 minutes to induce swelling then mechanically lysed on ice by 100 strokes of a Dounce homogeniser grade B. One quarter volume v/v of salt-balancing buffer (10 mM PIPES/NaOH pH 7.2, 600 mM KCl, 150 mM NaCl, 22.5 mM $MgCl_2$) supplemented with protease inhibitors was added prior to pelleting of nuclei and intact cells by centrifugation at 3,000 x *g* for 10 minutes at 4°C. Supernatants were transferred to a fresh tube.

2.2.6.3 Isolation of membrane-associated proteins

Post-nuclear supernatants were transferred to ultraspin centrifuge tubes and clarified at 100,000 x *g* for 1 hour. Supernatants were collected and acetone precipitated. Pellets were resuspended in isotonic buffer (4 volumes v/v of isotonic buffer plus 1 volume v/v of salt-balancing buffer) supplemented with protease inhibitors and clarified again. Pellets were collected into 2x Laemmli buffer (4% w/v SDS, 20% v/v glycerol, 20 mM (dithiothreitol) DTT, 125mM Tris-HCl pH 6.8, 0.01% w/v bromophenol blue).

2.2.6.4 Dissociation of membrane proteins

Post-nuclear fractions were separated into four ultraspin centrifuge tubes and clarified at 100,000 x *g* for 10 minutes at 4°C. Supernatants were collected and acetone precipitated. Pellets were thoroughly resuspended in either isotonic buffer (1 M NaCl, 100 mM Na_2CO_3 pH 11, 4 M urea) or 1% Triton X-100 in isotonic buffer and incubated at room temperature for 30 minutes. All buffers were supplemented with protease inhibitors. Samples were clarified at 100,000 x *g* for 10 minutes at 4°C. Supernatants were collected and acetone precipitated. Pellets were thoroughly resuspended in the isotonic buffer used previously and re-clarified. Pellets were finally recovered in 2x Laemmli buffer.

2.2.6.5 Immunoprecipitation of exogenous protein

Culture media was removed from cell monolayers and cells were washed three times in ice-cold PBS. All residual PBS was removed by aspiration. Cells were harvest by scraping into 200 µl lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) supplemented with protease inhibitors (Leupeptin 1 µg/ml, Pepstatin 1 µg/ml, Aprotinin 2 µg/ml, AEBSF (Perfabloc) 0.2 mM) and incubated on ice for 30 minutes. $1/10^{th}$ volume was removed as an input sample and added to 2x Laemmli buffer. The remaining sample was made up to 1 ml with Tris buffered solution (TBS) (50 mM Tris HCl pH 7.4, 150 mM NaCl) and clarified at 20,100 x *g* at 4°C for 10 minutes.

10 µl of GFP-Trap_A beads (Chromotek) were used per sample. Beads were washed three times in 500µl ice-cold TBS buffer with centrifugation at 2,700 x *g* at 4°C for 2 minutes and aliquoted into fresh microfuge tubes. Clarified lysates were added to pre-washed beads and incubated overnight at 4°C on a blood mixer. Bound protein was collected by centrifugation at 2,700 x *g* at 4°C for 2 minutes. Supernatants were aspirated off and pellets were washed 3 times in ice-cold TBS buffer. Immunoprecipitated proteins were finally recovered by the addition of 20 µl 2 x Laemmli buffer and incubating at 95°C for three minutes. Beads and insoluble material were removed by centrifugation at 2,700 x *g* 2 minutes at room temperature prior to analysis by SDS PAGE.

2.2.6.6 Acetone precipitation

Samples were added to 4 volumes v/v of acetone, mixed thoroughly and incubated overnight at -20°C. Precipitates were pelleted by centrifugation at 20,100 x *g* for 10 minutes at 4°C. Pellets were resuspended in 70% v/v ethanol and centrifuged again. Pellets were finally recovered in 2 x Laemmli buffer.

2.2.6.7 SEAP Assay

For analysis of intracellular SEAP activity, cells were lysed in an appropriate volume of 1x Passive Lysis Buffer (Promega). Endogenous non-specific alkaline phosphatases were inactivated by incubating samples at 65°C for 10 minutes. 20 µl of transfection cell lysate or culture medium in phosphate buffer (10 mM NaCl, 1 mM $MgCl_2$, 10 mM Tris-HCl pH 9.5) were transferred to separate wells of a 96 well plate. The chemiluminescent substrate CDP Star - Ready to use (Sigma) was added to a final concentration of 0.125 mM. Light emission was determined in relative light units (RLU) using a FLUOstar Optima luminometer (BMG Labtech).

2.2.6.8 ONPG assay

90 µl of transfection cell lysate or culture medium were transferred to separate wells of a 96 well plate containing 10 x LacZ buffer (500 mM NaCl, 100 mM $MgCl_2$, 100 mM β-mercaptoethanol). An equal volume v/v of substrate (10mM ortho-nitrophenyl-β-D-galactopyranoside (ONPG - Sigma) in 100 mM sodium phosphate buffer (22 mM sodium dihydrogen orthophosphate, 22 mM disodium hydrogen orthophosphate, pH 7.5) was added to each well. The reaction was stopped with a final concentration of 250 mM Na_2CO_3 after 45 minutes and the absorbance was read at 400 and 450 nm.

2.2.6.9 Endoglycosidase H (Endo H) treatment

$5x10^4$ cells were seeded into 6 well culture dishes 16 hours prior to transfection and samples were harvested 48 hours post-transfection. Lysates were prepared by harvesting into 100 µl PLB; supernatants were prepared by acetone precipitation of 1 ml culture medium and resuspension in 100 µl PLB. Samples were glycosidase treated according to the manufacturer's (New England Biolabs (NEB)) instructions. Briefly, samples were made up to 17 µl by the addition of 10 x denaturing buffering and denatured at 95°C for 10 minutes. Samples were made up to 20 µl with 10 x G5 reaction buffer and 500 U of Endo H and incubated for 1 hour at 37°C. Samples were finally recovered in 5x Laemmli buffer and analysed by SDS PAGE.

2.2.6.10 Peptide: N-Glycosidase F (PNGase F) treatment

$5x10^4$ cells were seeded into 6 well culture dishes 16 hours prior to transfection and samples were harvested 48 hours post-transfection. Lysates were prepared by harvesting into 100 µl PLB; supernatants were prepared by acetone precipitation of 1 ml culture medium and resuspension in 100 µl PLB. Samples were glycosidase treated according to the manufacturer's (NEB) instructions. Briefly, samples were made up to 15 µl by the addition of 10x denaturing buffering and denatured at 95°C for 10 minutes. Samples were made up to 20 µl with 10 x G7 reaction buffer, a final concentration of 1% v/v nonident P40 (NP40) and 500 U of PNGase F and incubated for 1 hour at 37°C. Samples were finally recovered in 5x Laemmli buffer and analysed by SDS PAGE.

2.2.7 *In silico* protein sequence analysis

Alignment of the NS2 amino acid sequence from each HCV sub-genotype was carried out using ClustalW2 (Goujon *et al.*, 2010; Larkin *et al.*, 2007). Hydropathy analysis was carried out using the AlignMe program (Khafizov *et al.*, 2010) using a window size of 19 and Kyte-Doolittle scoring. Membrane topology predictions were generated using 6 prediction programs: Sosui (Hirokawa *et al.*, 1998), HMMTOP (Tusnady & Simon, 1998), TMpred (Hofmann & Stoffel,

1993), ConPred II (Arai *et al.*, 2004), PHDhtm (Rost *et al.*, 1995) and TMHMM (Krogh *et al.*, 2001).

Sosui uses four parameters of primary sequence analysis: the Kyte-Doolittle hydropathy scale (Appendix 2), an amphiphilicity index (used to determine potential helix capping residues), amino acid charges and the length of each hydrophobic domain.

TMpred and PHDhtm make predictions by comparison of primary sequence to those of known trans-membrane protein. The TMpred algorithm uses statistical analysis of the TMbase database which contains characterised topology information of transmembrane proteins. PHDhtm topology predictions are based upon homologous protein sequences. Firstly a BLAST search is performed of the entered sequence against the SWISSPROT database. Secondly a prediction is made as to whether the entered sequence encodes a transmembrane protein by analysing the highest-scoring putative TMD. Finally dynamic algorithms are applied to achieve the most probable number, position and orientation of TMDs (Melen *et al.*, 2003).

HMMTOP and TMHMM use hidden Markov models with 5 and 7 state types respectively (such as helix core composition, helix caps, short cytosolic loops, short and long non-cytosolic loops) to predict over all topology. Rather than just taking into account absolute prediction values for each residue and segment in turn, these programs generate the most probable model based on overall sequence (Melen *et al.*, 2003).

ConPred II compiles and averages data from numerous different prediction programs regarding i) the total number of predicted TMDs and ii) the predicted position of TMDs to achieve an overall predicted topology. Among others, this software uses outputs from SOSUI, TMpred, TMHMM and HMMTOP.

CHAPTER 3: NS2 TOPOLOGY ANALYSIS

3.1 Introduction

Early studies determined that NS2 is a *trans*-membrane protein which associates with ER-derived membranes (Santolini *et al.*, 1995). Although the structure of the C-terminal sub-domain of NS2 (aa93-217), termed the catalytic domain, has been solved by X-ray crystallography (Lorenz *et al.*, 2006b), the precise membrane topology of the mature NS2 protein remains unclear, specifically with regard to the number and position of the *trans*-membrane domains (TMDs) within the N terminal region.

The 92 amino acids comprising the N-terminal portion are collectively referred to as the *trans*-membrane domain. Analyses of this domain, by topology prediction and dissection of the cleavage events which culminate in the release of mature NS2, provided strong evidence that aa 1-92 form three distinct TMDs (Phan *et al.*, 2009; Pietschmann *et al.*, 2006). Cleavage of the N-terminus of NS2 from p7 occurs in the ER lumen (Carrere-Kremer *et al.*, 2002), whilst cleavage of the C-terminus of NS2 from NS3 is thought to occur in the cytosol as NS3 has been shown to localise to the cytosol and nuclei of infected cells (Errington *et al.*, 1999) and recruits NS4A to cleave the downstream viral protein junctions (Failla *et al.*, 1994; Grakoui *et al.*, 1993a; Tomei *et al.*, 1993). Orientation of the N- and C-termini to opposing faces of a lipid bi-layer to allow the release of NS3 into the cytosol could theoretically be achieved by a single TMD topology. Similarly, a single TMD would be sufficient if NS2 function was solely dependent on anchorage to membranes. However, NS2 has known interactions with E1, E2 and p7 which are hypothesised to occur via intra-membrane contacts (Jirasko *et al.*, 2010; Popescu *et al.*, 2011; Stapleford & Lindenbach, 2011). If this hypothesis is correct, a triple TMD topology may create a luminal loop capable of interacting with the glycoprotein ectodomains as well as providing more opportunities for intra-membrane protein:protein interactions with the TMDs of E1, E2 and p7. In contrast to these findings, there is also evidence to support a four TMD composition of NS2 (Yamaga & Ou, 2002).

Work investigating the topology and membrane integration of NS2 from two different genotype 1 strains (HCV-BK and HCV-J) expressed *in vitro,* in the presence of canine microsomal membranes (CMM), showed NS2 membrane association occurred in a p7-independent, NS3-dependent manner (Santolini *et al.*, 1995). The group also described a 17 kDa fragment which localised to a protease-protected environment when NS2 was expressed in *cis* with NS3 or as part of a polypeptide spanning the C-terminal domain of E2 through to NS3. Although these experiments were primarily conducted in cell-free environments supplemented with CMM, validation was provided using a cell-based system, suggesting that

the majority of NS2 resides within the membrane or lumen of vesicles or organelles (Santolini *et al.*, 1995). Subsequently, Yamaga and Ou (Yamaga & Ou, 2002) found that *in vitro* expressed NS2 from a different genotype 1 strain (HCV-1) also associated with membranes, even when expressed separately from other HCV proteins and that the full-length protein behaved as an integral membrane protein. *In vitro* analysis of an NS2 deletion series identified two internal signal sequences within peptides aa 30-74 and aa 119-151, whilst the introduction of synthetic glycosylation acceptor sites led to the conclusion that residues 3, 50, 179 and 209 were oriented to the lumen of the ER (Yamaga & Ou, 2002). From these observations, and a Kyte-Doolittle hydropathy plot, a four TMD topology for NS2 was consequently proposed, with both termini oriented to the ER lumen and residues (4-26), (34-57), (63-110) and (119-147) containing membrane spanning domains (Yamaga & Ou, 2002). The 17 kDa protease protected peptide observed by Santollini *et al.,* (Santolini *et al.*, 1995) was proposed to correspond to the C-terminal domain of NS2 which would correlate with the observations of Yamaga and Ou (Yamaga & Ou, 2002) who predicted that the largest fragment of NS2 not exposed to exogenous protease would be the residues aa119-217. This work has been questioned by the solution of the catalytic domain by X-ray crystallography, which has cast doubt on the hypothesis of a four TMD topology and the orientation of the C-terminus to the ER lumen (Lorenz *et al.*, 2006b), as discussed in section 1.8.1. In addition, a 17kDa peptide of NS2, based on the HCV-H sequence, would require the protection of either the first 152 residues of NS2 or the C-terminal 157 residues (calculated) i.e. residue 61 onwards, making a four TMD topology less likely. More recently, several studies have proposed models predicting NS2 to form three TMDs based upon primary sequence analysis (section 1.8.1), however little has work been carried out to validate these models.

The role of p7 in conferring NS2 topology, as well as the sub-cellular trafficking of these two proteins, is poorly understood. The signal peptide capability of the C-terminal helix of p7 suggests that it may function as such in the HCV polyprotein to facilitate NS2 topology (Carrere-Kremer *et al.*, 2002). NS2 expressed alone has a diffuse, ER-like distribution, whereas virus-expressed NS2 co-localises with NS5A to lipid droplets by a mechanism that is dependent upon the expression of functional p7 in *cis* (Ma *et al.*, 2011; Tedbury *et al.*, 2011). It was also demonstrated that NS2 partitioning to a triton-resistant compartment was dependent upon addition of signal peptide-p7 to NS2-5B replicons (Tedbury *et al.*, 2011). A sub-population of E2-p7 precursor has been shown to localise to the surface of cells (Isherwood & Patel, 2005) and p7 is able to alter the pH of intracellular vesicles (Wozniak *et al.*, 2010), yet the precise cellular trafficking of p7 has not been described.

A complementary reporter enzyme system utilising alkaline phosphatase (PhoA) and β-galactosidase (β-gal) has been widely used in combination with the bacterial plasma membrane to provide a complementary activity profile for truncation fusions of bacterial polytopic plasma membrane proteins. This combination of reporter was used in the initial identification of many of the *trans*-membrane domains of the mammalian plasma membrane proteins TAP1 and TAP2 (Vos *et al.*, 1999).

Placental alkaline phosphatase (PLAP) is expressed on the surface of mammalian cells and is anchored to the membrane by a glycosylphosphatidylinositol (GPI) anchor (Micanovic *et al.*, 1988). Secreted alkaline phosphatase (SEAP) is a modified form of PLAP in which the C-terminal 8 amino acids responsible for covalent addition of the GPI anchor have been deleted (Berger *et al.*, 1988). This results in trafficking of the enzyme to the cell surface and secretion from the cell. SEAP is a heat-stable enzyme and several methods are available for the detection of its enzyme activity. Enzyme activity can be determined using substrates that are converted to colourimetric or chemi-luminescent products and specific SEAP activity can be obtained following heat inactivation of endogenous phosphatases. SEAP can therefore replace PhoA in a dual reporter-fusion system. SEAP possesses one confirmed *N*-glycosylation site at N249 (Endo *et al.*, 1988; Le Du *et al.*, 2001) (Appendix 3) and SEAP maturation has been shown to be ER-dependent as non-cytotoxic levels of reagents that induce ER-dysfunction (Hiramatsu *et al.*, 2006) and deoxynojirimycin (DNJ), an inhibitor of the ER chaperone calnexin, impair SEAP enzyme activity (Norton *et al.*, 2005).

β-gal catalyses hydrolysis of β-galactosides into monosaccharides. Both prokaryotic and eukaryotic β-gal is active in the cytosol and inactive in the oxidising environment of a bacterial periplasm. The activity of β-gal when anchored to the luminal face of the ER membrane is unknown. If β-gal is inactive when orientated to the ER lumen then it will complement SEAP as part of a dual reporter system for the topological analysis of ER-resident *trans*-membrane proteins.

The novel application of a SEAP/ β-gal dual reporter system to assess the topology of an ER transmembrane protein was investigated by expression of a series of C-terminal truncations of NS2, in hepatocellular carcinoma (Huh7) cells. The sub-cellular trafficking of NS2 and p7 was also investigated by introduction of a consensus cleavage motif for the *trans*-Golgi network (TGN) -localised furin protease between the HCV protein sequence and the reporter fusion. The truncation fusions were designed such that release of the reporter fusion from the HCV

protein sequence would only occur if the reporter was oriented to the lumen and the HCV protein sequence enabled trafficking to the TGN.

3.2 Results

3.2.1 Sequence based analysis of NS2

The NS2 coding sequence from 9 different HCV isolates encompassing all 7 genotypes were analysed for sequence homology. The NS2 amino acid sequences were aligned using the ClustalW2 program (Goujon *et al.*, 2010; Larkin *et al.*, 2007). As the crystal structure of the C-terminal catalytic domain (aa 93-217) revealed no membrane spanning domains (Lorenz *et al.*, 2006b), the attention of this study focused on the N-terminal 100 aa (Figure 3.1). The most striking feature of these sequences was their relative hydrophobicity and in particular the conserved stretches of hydrophobic amino acids (red) from aa 10-24, aa 29-37, aa 51-57, aa 64-77 and aa 82-92. All domains were too small to form a typical α-helical TMD (>16 residues) in isolation, suggesting that membrane-spanning domains of NS2 may contain polar residues which could be involved in intra-membrane interactions.

Residues 51-62 and 76-81 contained highly conserved or synonymous residues, indicating that they may form interaction motifs or key structural features, with regions proximal to the N-terminus and residue 27 also showing high degrees of sequence conservation. In addition four conserved prolines, with disrupt α-helical structures, noted at positions 24, 53, 73 and 89 could represent TMD boundaries.

HCV sequences were also aligned with the corresponding coding sequence from the two closest genetic relatives to HCV: canine hepatitis virus (CHV) and GB virus B (GBV-B). Although GBV-B NS2 was 9 amino acids shorter than HCV NS2, the catalytic domain showed a high degree of homology; with H134, E163 and C184 which form the catalytic triad being conserved across viruses (Kapoor *et al.*, 2011). In contrast, despite showing a similar degree of hydrophobicity, the N-terminal domain of GBV-B NS2 has markedly less sequence homology with that of HCV. The NS2 coding sequence of the recently characterised CHV sequence (Kapoor *et al.*, 2011) was the same length as HCV NS2 and sequence alignment highlighted a greater level of sequence homology between these viruses than between HCV and GBV-B. A similar pattern of hydrophobic and hydrophilic sections, as well as two small clusters of conserved residues around aa 26 and 35, were evident within the *trans*-membrane domain of NS2 from HCV and CHV. It should be noted that to date only one isolate of CHV has been submitted to the European HCV database (http://euhcvdb.ibcp.fr/euhcvdb/) so this isolate may not be representative.

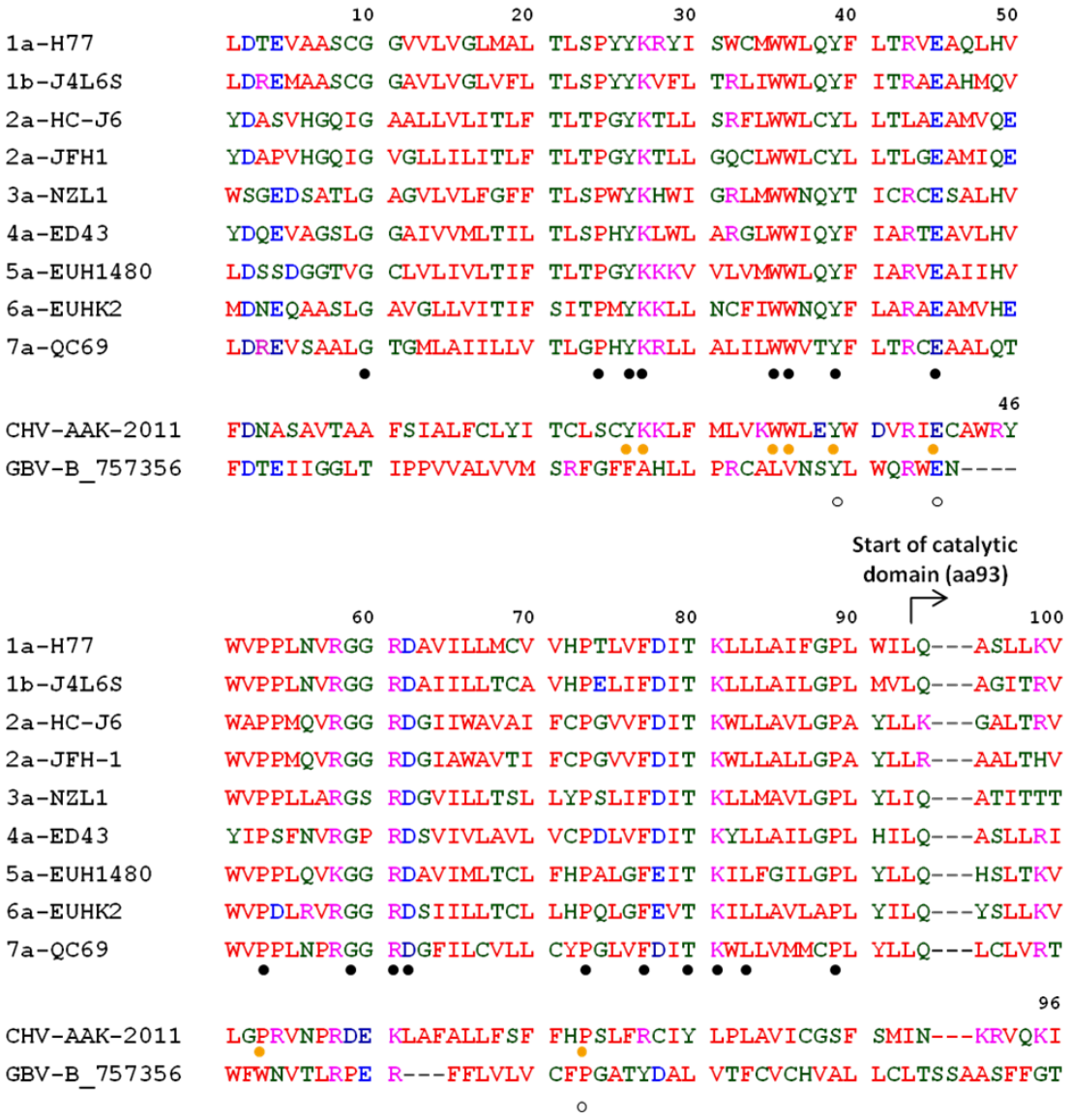


Figure 3.1 Alignment of the NS2 amino acid sequence from a group of representative genotype isolates.

The NS2 coding sequences of 9 HCV isolates from different genotypes were aligned for sequence conservation. Residues are coloured to denote biochemical properties: hydrophobic (red), polar (green), acidic (blue) and basic (magenta). Sequence alignment from the closest genetic relatives to HCV – canine hepacivirus (CHV) and GB virus B (GBV-B) are also shown. Circles below residues denote: conserved residue throughout HCV sequences (black), residues conserved between HCV and CHV (orange) and residues conserved between all three viruses (white). Amino acid numbering is shown above the alignments; isolate genotype and reference name are shown to the left.

The protein sequence for NS2 was also analysed for hydropathy (Figure 3.2) using Kyte-Doolittle scoring (Appendix 2). From this analysis it is evident that the N-terminal ~30 residues form a hydrophobic domain consistent a TMD. A second hydrophobic domain is formed by residues ~31-50, although this is not sufficiently hydrophobic to be predicted as forming a TMD by this method as it does not reach a peak hydrophobicity of 1.8. This could be indicative of an amphipathic helix i.e. one that lies parallel to the membrane with a non-polar face that interacts with the membrane and a polar face that is solvent exposed. It is also possible that this domain forms a predominantly hydrophobic α-helix with polar residues that are involved in forming intra-membrane interactions. These possibilities have been previously noted for the corresponding domain of NS2 from the Con1 isolate (Jirasko *et al.*, 2010). A third extended domain of hydrophobicity was calculated from approximately residue 70 through to residue 110, within the catalytic domain, and forming two distinct shoulders. This domain is long enough to form two TMDs or, as it extends into the catalytic domain, a single TMD the leads into a hydrophobic, membrane-proximal domain of the catalytic domain.

The amino acid (aa) sequence of JFH-1 NS2 was used as the reference for sequence analysis. The JFH-1 isolate is able to undergo a full round of infectivity in cell culture allowing for the potential to directly translate observations from protein expression studies into the full-length virus system and so was the focus of this study. The JFH-1 NS2 amino acid sequence was analysed using 6 topology prediction programs. The outputs from these programs are summarised in Table 3.1. The programs use varying methods to prdict protein topology. SOSUI uses a set of 4 fixed scales to analyse primary sequence. TMpred and PHDhtm are closely based upon comparison with known and homologous sequences from characterised *trans*-membrane proteins. HMMTOP and TMHMM use hidden Markov models taking into account several factors based on statistical bias determined by analysis of the primary peptide sequences of characterised *trans*-membrane proteins. ConPred II analyses outputs from 5 different programs for eukaryote protein topology, combines the 4 most similar and dynamically refines a model of the highest probability. More detail is provided in Methods section 2.2.7.

Every program used predicted NS2 as a polytopic membrane protein comprising three or four TMDs with the N-terminus oriented 'Out', that is, on the extracellular or luminal face of a lipid bi-layer. Consistent with the previous predictions (detailed in section 1.8.1, Figure 1.9) (Jirasko *et al.*, 2008; Phan *et al.*, 2009; Pietschmann *et al.*, 2006), all but one of the programs predicted a membrane-spanning domain within aa 5-33. Two less consistently identified sequences (aa

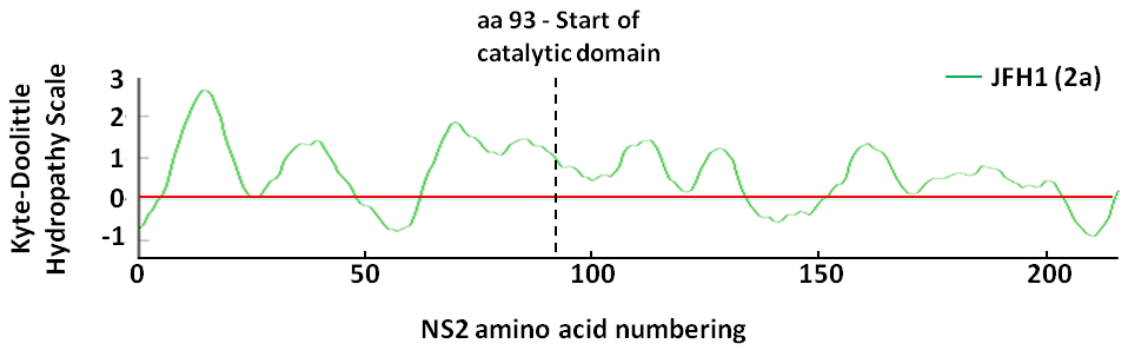


Figure 3.2 Hydropathy analysis of JFH-1 NS2 protein sequence.

Analysis was carried out using the AlignMe software using Kyte-Doolittle scoring (Appendix 2) and a window size of 19. Stretches of sequence scoring higher than 1.8 are consistent with membrane spanning domains.

	TMD 1	TMD 2	TMD 3	TMD 4
Sosui	7-29	31-53	64-86	96-118
HMMTOP	15-39		69-93	98-116
TMpred *Model I*	5-24	28-48	62-80	
TMpred *Model II*	7-26	29-46	60-77	82-98
ConPred II	7-27	29-49	65-85	
PHDhtm	5-22	29-48	63-86	
TMHMM	9-33		64-88	119-140
Consensus	**7-27**	**29-49**	**63-86**	

Table 3.1 Defining a computer model for the membrane spanning domains of NS2 using topology prediction software.

7 topology predictions for JFH-1 NS2 were obtained using 6 programs shown on the left. Comparable domains predicted to a section of sequence are shaded. An inferred consensus computer model for NS2 topology is shown (Consensus). All programs used predicted the N-terminus to be orientated to the luminal/ extracellular face of a membrane bi-layer.

28-53 and aa 60-93) were also predicted to contain membrane-spanning domains. From these data a consensus computer model (hereafter referred to as the 'computer model') for NS2 topology was inferred, which defined three putative TMDs between positions 7-27, 29-49 and 63-86. The computer model was roughly determined by-eye as amino acid sequences predicted to form membrane spanning domains by the majority of the programs. The computer model placed the N-terminus on the luminal side of the ER membrane with the C-terminus oriented to the cytosol (Table 3.1). The predicted orientation of the termini to respective sides of the ER membrane was consistent with the known cleavage events that release mature NS2.

The computer model predicted NS2 to form a tight, cytosolic hairpin loop between the first and second TMDs. The computer model also defined a short luminal loop of approximately 15 aa between the second and third predicted TMDs. Half of the programs utilised predicted a TMD within the N-terminal portion of the catalytic domain, however the position and length of the domain was inconsistent. Therefore the computer model did not include the prediction of a TMD within the catalytic domain.

The prediction of a TMD within the catalytic domain by some programs was consistent with a previous report (Yamaga & Ou, 2002), inspection of the X-ray crystal structure of the catalytic domain suggest it is likely due to misinterpretation of extended hydrophobic stretches shown to form ordered structural domains (Lorenz *et al.*, 2006b). Together, the computer model and previous experimental observations strongly suggested that NS2 contained 3 TMDs (Figure 3.3A).

3.2.2 Development of an NS2 dual reporter-fusion system

Topology analysis was undertaken using a series of C-terminal truncations of NS2 from the JFH-1 isolate expressed as fusions to a reporter enzyme. It was hypothesised that progressive C-terminal truncations through a polytopic membrane protein would result in changes to the orientation of the reporter about the ER membrane as single TMDs were removed from the sequence (Figure 3.3B). The activity of a C-terminally fused reporter molecule should enable the identification of NS2 truncations that orientate their C-terminus to the ER lumen by assessing the enzyme activity of the reporter fusion. A complementary dual reporter system was designed using SEAP and β-gal. It was anticipated that SEAP would be selectively active if orientated to the ER lumen and β-gal would be selectively active in the cytosol.

C-terminal truncations of NS2 were generated such that they would terminate NS2 sequence either within or between predicted TMDs defined by the computer model (Table 3.1). Reporter

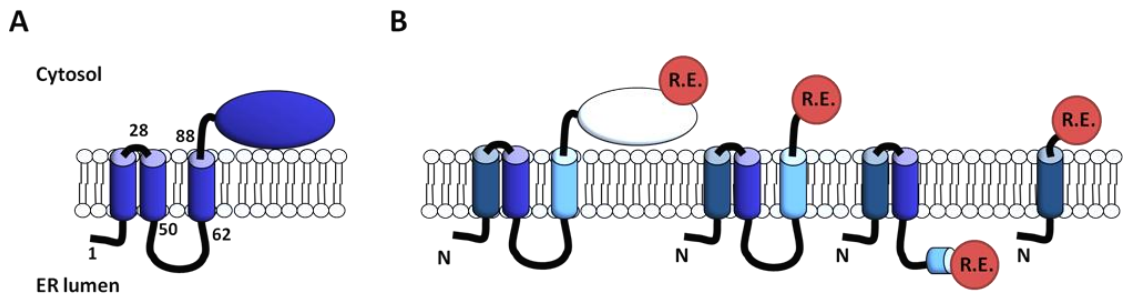


Figure 3.3 Model of predicted NS2 topology and principle of C-terminal truncation mapping. (**A**) A cartoon of NS2 modelled onto a lipid bi-layer. TMDs are shown as cylinders and key residues are noted. Residues marking the TMD boundaries are noted (green boxes). (**B**) Topology mapping of C-terminal truncations of a polytopic protein fused to a reporter enzyme (R.E.) result in re-orientation of the C-terminus about the membrane. The N-terminus (N) remains oriented to one side of the membrane.

fusions to SEAP (S) or β-gal (B) were numbered 1-10 in order of the magnitude of the truncation being made: S1 contained full-length NS2, S9 contained the shortest section of NS2 sequence (the N-terminal 27 residues) and S10 contained no NS2 sequence. A schematic outlining the NS2 truncation SEAP-fusions is shown in Figure 3.4.

Taking into account the ambiguous role of p7 in NS2 topology and subcellular trafficking; the C-terminal NS2 truncation-SEAP fusions were encoded in *cis* with p7 and the E2-derived cognate signal peptide, collectively termed signal peptide-p7 or SPp7 (Tedbury *et al.*, 2011).

To determine whether p7 and NS2 traffic via the Golgi, viral protein sequences were separated from reporters by a cleavage motif specific for furin. To more thoroughly characterise the trafficking of p7 within the cell, the C-terminal residue of p7 (p7-A63) was mutated to proline in the S10 background and the resulting mutant termed S11 (Figure 3.4). In line with this another mutant of S10 was designed containing a mutation to the consensus furin cleavage site, whereby the first arginine of the motif R-X-(R/K)-R was mutated to alanine and the resulting construct termed S12. A double mutant was also generated - S13. As it was feasible that mutations predicted *in silico* to abrogate peptidase cleavage may not function accordingly in tissue culture an additional construct was generated whereby SEAP was encoded immediately downstream of SPp7 without an intervening furin cleavage motif site and where the C-terminal amino acid of p7 was also deleted and termed Δ63. The p7 trafficking mutants were not generated in the β-gal background as the reporter was predicted to be inactive in these constructs.

3.2.3 Characterisation of SEAP control reporters

To verify the use of SEAP as a reporter enzyme it was necessary to demonstrate that wild type (wt) SEAP, expressed from the pSEAP control vector, was active and secreted from cells. Conversely it was necessary to confirm that if SEAP is orientated to the cytosol rather than the ER lumen the opposite would be observed. To this end a non-secreted form of SEAP was generated by cloning the sequence encoding SEAP but lacking the N-terminal signal peptide sequence into pcDNA3.1(+). This construct was termed 'non-secreted alkaline phosphatase' (nsAP) (Figure 3.5).

COS7 cells were transfected with pcDNA3.1(+) plasmid (vector) alone, pSEAP (wt SEAP) or nsAP and the reporter phenotypes assessed. The vector pcDNA3.1(+) was used as the backbone for all novel constructs described in this chapter and is hereafter refered to as 'vector' or 'vector alone' and was used as a negative control. Enzyme activity was determined by assessing cell lysates and culture supernatants for SEAP activity by chemi-luminescence assay (Figure 3.6A).

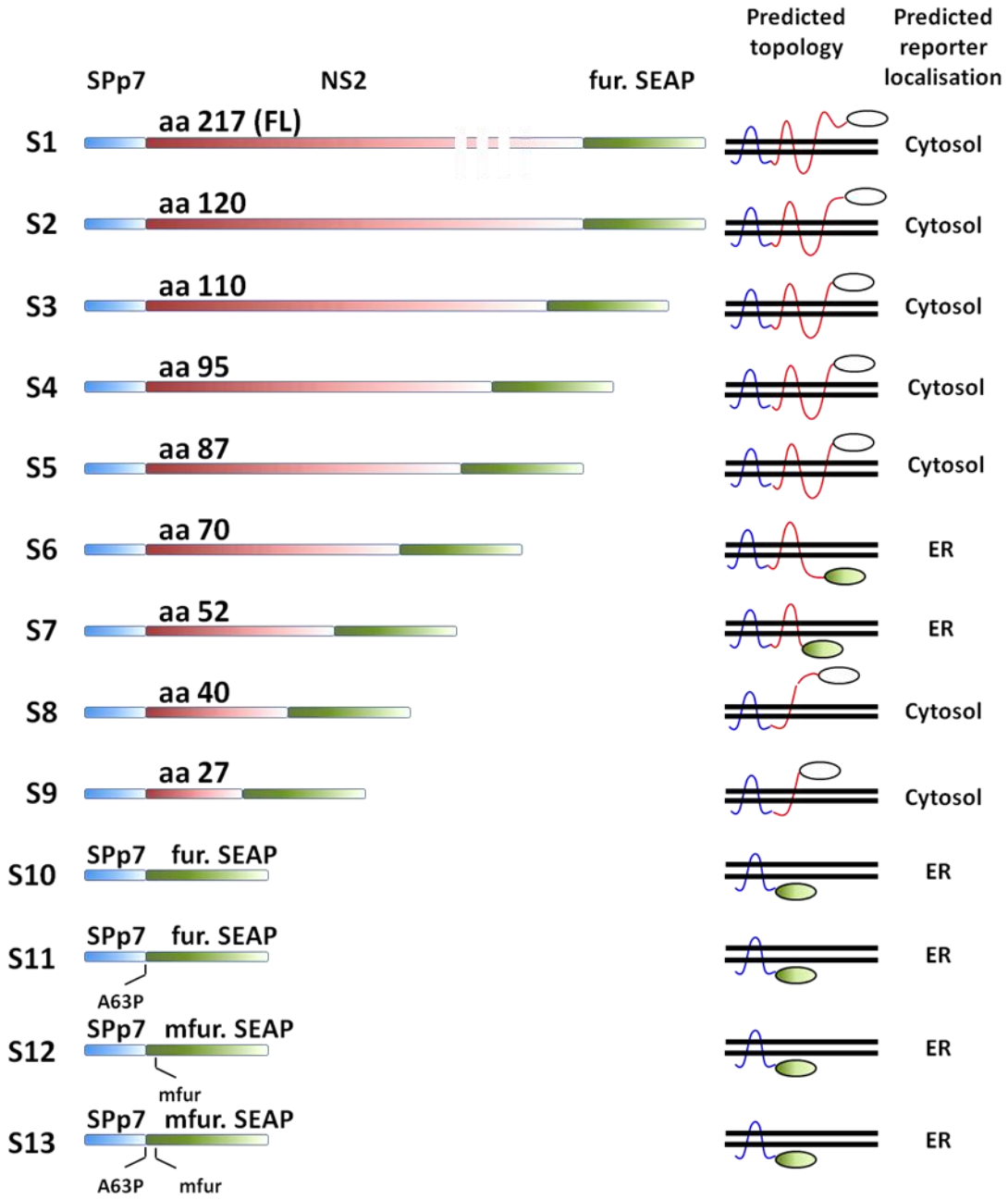


Figure 3.4 Schematic of NS2-SEAP fusion truncation series.

Schematic representation of the expression constructs are shown alongside predicted membrane topology and reporter orientation about the membrane. Constructs comprise p7 along with the E2-derived cognate signal peptide (SPp7) (blue), NS2 sequence (red) and secreted alkaline phosphatise (SEAP) reporter fusion (green). Terminating residue of NS2 sequence is noted above each construct. FL refers to full length NS2. Nomenclature of constructs is noted to the left of each fusion construct. The predicted topology the SEAP reporter is represented either by an empty ellipsoid (inactive) or a green ellipsoid (active) denoting the predicted enzyme activity of the reporter expressed from the fusions. Signal peptidase and furin cleavage mutations (A63P and mfur respectively) are indicated.

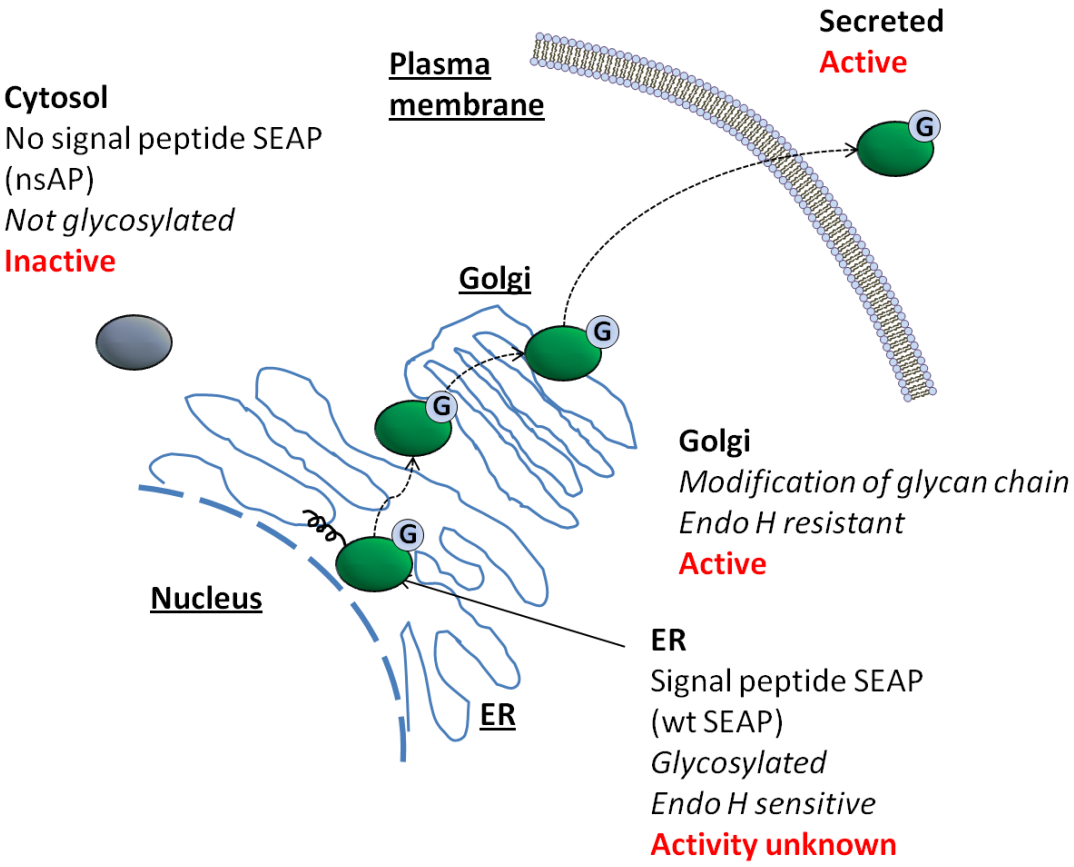
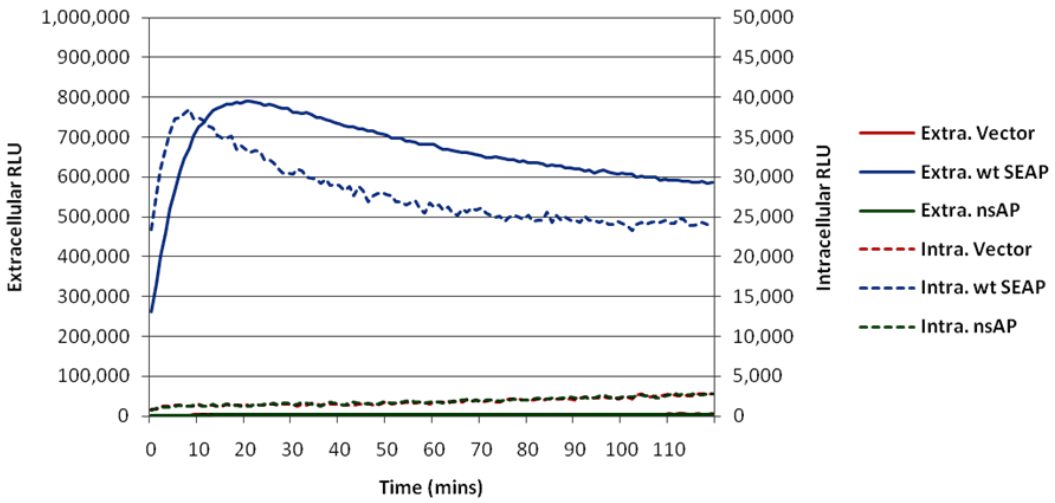


Figure 3.5 Schematic diagram showing the maturation of SEAP.

Wt SEAP is translocated to the lumen of the ER by recognition of the signal peptide by the signal recognition particle (SRP). Signal peptidase releases the enzyme domain of SEAP from the membrane and oligo-saccharyltransferase (OST) glycosylates SEAP at asparagine 249. Wt SEAP is trafficked to the *trans*-Golgi network (TGN) where the glycan group is modified prior to secretion of the mature protein. Modification of the glycan moiety to a more complex or hybrid saccharide chain renders it insensitive to endoglycosidase H (Endo H) treatment. SEAP lacking an N-terminal signal peptide (nsAP) is not targeted to membranes and is therefore not post-translationally modified or functionally active.

A


Extracellular RLU
Intracellular RLU
Time (mins)
Extra. Vector
Extra. wt SEAP
Extra. nsAP
Intra. Vector
Intra. wt SEAP
Intra. nsAP

B

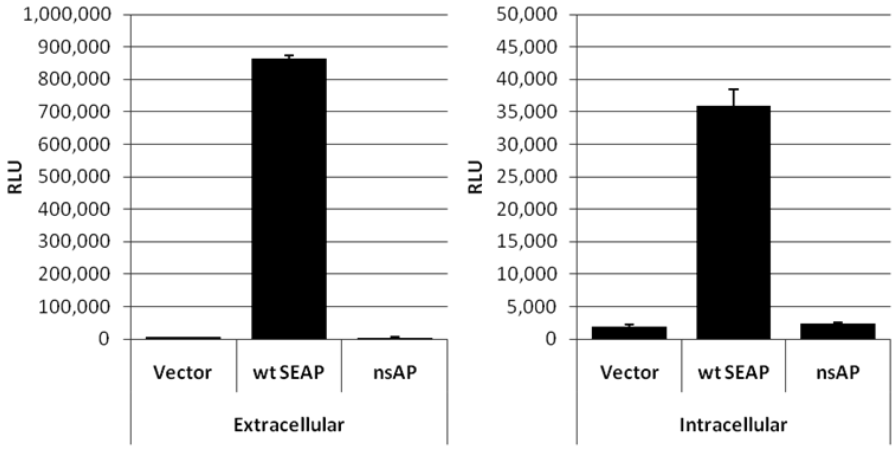

RLU
Vector
wt SEAP
nsAP
Extracellular
Intracellular

Figure 3.6 Enzyme activity of SEAP control reporters.

SEAP activity was determined by chemi-luminescence assay using the CDPStar™ substrate. COS7 cells were transfected with 1µg pcDNA3.1(+) (vector), pSEAP (wt SEAP) or pcDNA3.1-nsAP (nsAP). Cell lysates and culture supernatants were collected 48h post-transfection. Endogenous alkaline phosphatases were inactivated by heating of samples at 65°C for 10 min. (**A**) Chemi-luminescence signal was monitored for 2 hr after addition of substrate. Extra- and intracellular enzyme activities are plotted on opposing axes. Peak enzyme activity was achieved approx. 10-20 mins after substrate addition. (**B**) Peak extra- and intracellular SEAP activity. Measurement of chemi-luminescence was initiated 10 min (T=10) after addition of the substrate (T=0) whereby readings were taken every minute for 10 min (T=20). Values represent mean of values for a sample taken between T=10 and T=20. Where possible, each sample was assessed in triplicate. The numerical means are shown and positive standard deviations were calculated and included for each sample. 20µl of culture media or 5% cell lysate were used. Enzyme activity is shown in relative light units (RLU).

Cells expressing wt SEAP had high levels of extracellular SEAP activity with intracellular activity approximately 20-fold lower. Cells transfected with empty vector or nsAP showed neither extra- nor intracellular activity. Monitoring of reactions for 2 hours after addition the substrate revealed that the intracellular chemi-luminescence signal peaked 8 minutes after substrate addition and at 19 minutes after substrate addition for extracellular SEAP (Figure 3.6A). To expedite data collection measurement of chemi-luminescence was initiated 10 minutes (T=10) after addition of the substrate (T=0) whereby readings were taken every minute thereafter for 10 minutes (T=20). Values therefore represented the mean of 11 readings per sample taken from T=10 to T=20 (Figure 3.6B).

3.2.4 Expression of SEAP fusions in mammalian cells

NS2-SEAP fusions were initially expressed in the hepatocellular carcinoma cell line Huh7. However, due to poor expression levels the truncations were subsequently expressed in COS-7 cells. COS-7 cells contain the SV40 large T antigen which enables episomal replication of transfected plasmids that contain the SV40 origin of DNA replication within mammalian cells, leading to enhanced expression of recombinant proteins. Analysis of equal amounts of cell lysate by SDS PAGE and immunoblotting for SEAP revealed that all fusion proteins were expressed, but at varying levels (Figure 3.7). This was confirmed by probing for the vector-encoded neomycin phosphotransferase (NPT) as a measure of transfection efficiency, and the cellular protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a loading control. NPT levels were relatively constant indicating that the discrepancy did not arise as a consequence of unequal transfection rates and GAPDH levels demonstrated that protein loading was consistent and that none of the constructs were noticeably cytotoxic. The vector encoding wt SEAP did not encode the NPT resistance gene. Fusion peptides that included NS2 sequence (S1-S9) were detectable at significantly lower levels than those where NS2 was not present (Δ63 and S10-S13). Full-length NS2 has a molecular mass of 23 kDa and the SEAP-reactive species seen for S1 migrated at a slower rate equivalent to an increase in size of approximately 20 kDa over wt SEAP. The altered migration of the NS2-containing fusions in comparison to wt SEAP was consistent with the respective NS2 sequences remaining fused to the reporter. Lysates were not clarified prior to analysis by SDS PAGE because up to 50% of NS2 has been shown to associate with detergent insoluble lipid-rafts when expressed in a replicon also containing SPp7 (Tedbury *et al.*, 2011).

3.2.5 Expression of NS2-SEAP fusions *in vitro*

Since the consistent expression of NPT had already excluded transfection efficiency as the cause of low fusion-protein expression levels seen in some of the constructs, it was important

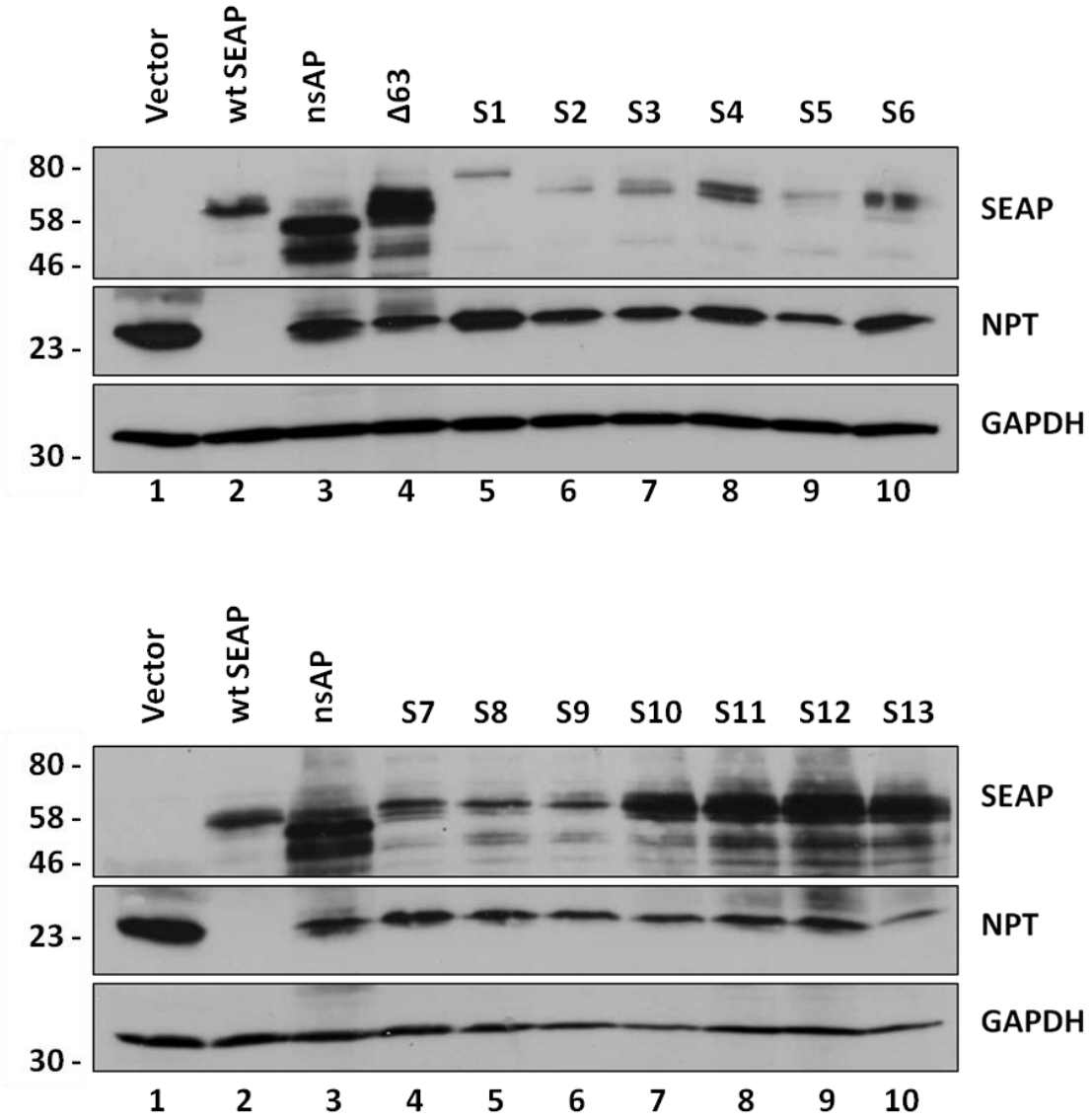


Figure 3.7 Expression of NS2-SEAP fusions in COS7 cells.

COS7 cells were transfected with NS2-SEAP fusion series and cells harvested 48h post-transfection. Fusion protein expression levels were determined by probing for SEAP following separation by SDS PAGE. Transfection efficiency was gauged from levels of the vector-encoded neomycin phosphotransferase II resistance gene (NPT). The cellular protein glyceraldehydes-3-phosphate dehydrogenase (GAPDH) was used as loading control. Vector alone, wt SEAP and nsAP were separated on both gels for comparison. The vector encoding wt SEAP did not encode NPT. Molecular weight markers (kDa) are shown to the left.

to determine whether there were unequal levels of transcription or translation across the panel of constructs. To investigate this; equal amounts of each plasmid were added to the coupled transcription and translation (TNT) reactions supplemented Trans-^{35}S-label containing with ^{35}S methionine and ^{35}S cysteine. Transcription from these constructs *in vitro* was facilitated by the T7 promoter encoded by the vector. TNT reactions proceeded for 90 minutes at 30°C before termination through the addition of ribonuclease A (RNase A). The levels of radiolabelled amino acids incorporated-protein were determined by separation of samples by SDS PAGE followed by autoradiography (Figure 3.8). The reporter fusions migrated corresponding to the varying size of the NS2 sequence encoded by each construct. In this context wt SEAP and nsAP would be expected to run at similar sizes as neither reporter should be post-translationally modified, although this could not be demonstrated. The vector encoding wt SEAP did not contain a T7 promoter site and so could not be used in this system. SPp7 should have been present as part of the precursor polyprotein as no signal peptidase or furin were present in this system. A slightly retarded migration of the p7-SEAP fusions (S10-S13) in comparison to the nsAP control was evidence of this (Figure 3.8).

Trans-^{35}S-label includes radio-labelled methionine and cysteine which are incorporated in growing peptide chains and enable their detection by autoradiography. All of the fusions appeared to express to similar levels, with the possible exceptions of S2 and S5 which gave less intense signals which could not be explained by the different number of methionines and cysteines encoded by the constructs (Figure 3.8). From this it was concluded that the low expression levels in COS7 cells of the NS2-containing fusions in comparison to the p7-SEAP fusions was not due to poor transcription or translation from the plasmids. Rather it was likely that observed phenotypes were a result of differential stability of the resultant proteins, either caused by specific targeting motifs present within NS2 or due to recognition of incorrectly folded proteins by cellular quality control mechanisms.

3.2.6 Enzymatic activity of NS2-SEAP fusions

Cells expressing the NS2-SEAP fusion were assayed for extra- and intracellular SEAP activity (Figure 3.9). The fusion peptides examining p7 trafficking (Δ63, S11, S12 and S13) are discussed in section 3.2.11. Wt SEAP showed high levels of SEAP activity both extra- and intracellularly. Cells transfected with the non-secreted form of SEAP, nsAP, or vector alone had undetectable enzyme activity. S10 i.e. SPp7-SEAP, displayed extracellular levels of reporter activity comparable to wt SEAP and intracellular levels slightly reduced in comparison with wt SEAP. S9 was the only NS2-containing fusion that repeatedly showed significant extra- and intracellular enzymatic activity above that of background i.e. nsAP or vector alone. All NS2-containing

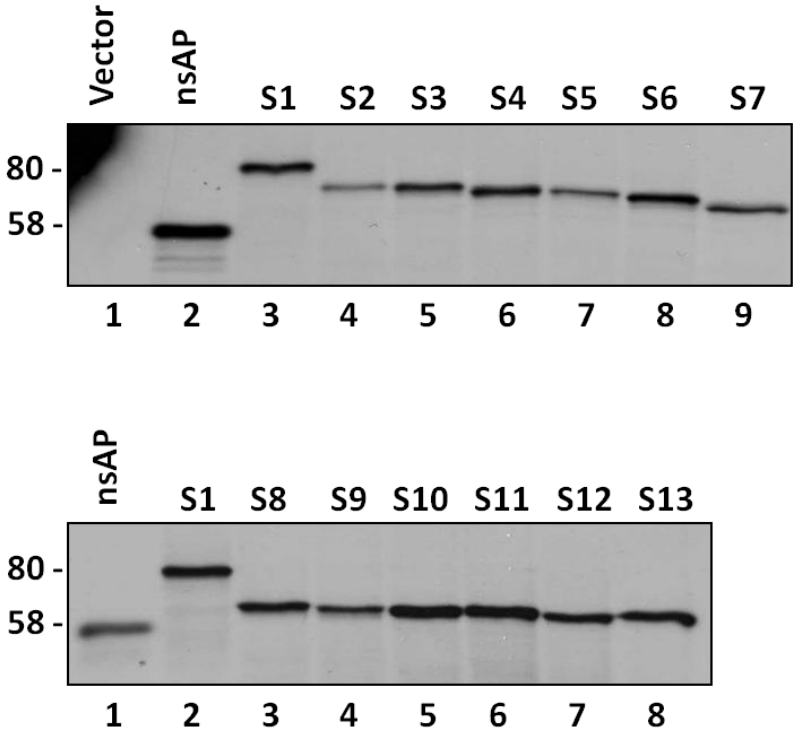


Figure 3.8 *In vitro* translation of NS2-SEAP fusions.

NS2-SEAP fusions were expressed *in vitro* using a coupled transcription and translation system in the presence of ^{35}S radiolabelled amino acids. Reactions were incubated for 90 min at 30°C. Samples were separated by SDS PAGE and reaction products detected by autoradiography. 250ng of plasmid DNA was used per reaction. Molecular weight markers (kDa) are shown to the left. nsAP and S1 were loaded on both gels for size comparison.

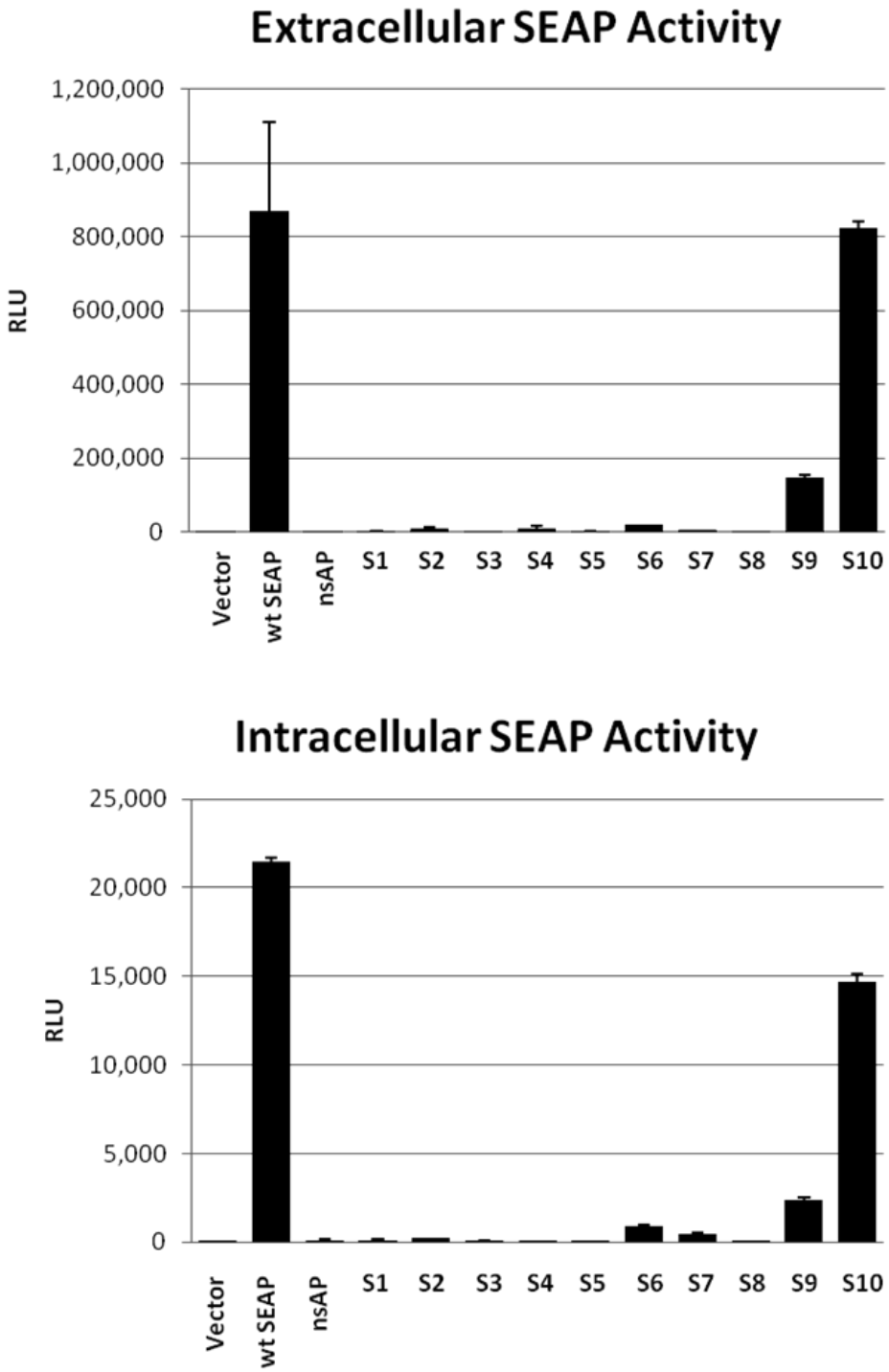


Figure 3.9 Reporter activity of NS2-SEAP fusions.

COS7 cells were transfected with NS2-SEAP fusion series and extra- and intracellular reporter enzyme activity determined by chemiluminescent assay using CDPStar substrate. Each sample was assayed in triplicate, numerical means were plotted and standard deviations shown. Assay was carried out ≥3 times. A representative data set is shown. Enzyme activity is shown in relative light units (RLU).

fusions (S1-S9) were expressed at comparable levels, although expression was significantly reduced compared with S10, nsAP and wt SEAP. In light of these findings, the relative inactivity of the reporters within the NS2-containing fusions S1-S8 compared with S9 cannot be attributed solely to sub-optimal levels of expression.

Unfortunately, no information regarding NS2 topology could be inferred directly from the reporter activity. However, as it is not known at what stage during maturation SEAP becomes active it is possible that NS2 sequences oriented the reporter to the ER lumen but undefined trafficking signals prevented trafficking of the fusion proteins through the secretion pathway and thus prevented activation of SEAP.

SEAP is a characterised glycoprotein and initial addition of a core oligosaccharide ($Glc_3Man_9GlcNAc_2$), to an acceptor asparagine on a nascent protein by oligosaccharyltransferase (OST), occurs in the ER lumen. Upon trafficking to the Golgi, glycan chains are modified to produce high-mannose, complex or hybrid oligosaccharide chains. Endoglycosidase H (Endo H) removes only high mannose and some hybrid oligosaccharides whereas Peptide: *N*-Glycosidase F (PNGase F) cleaves nearly all types of *N*-linked glycan chains except those containing core α1-3 fucose (Tretter *et al.*, 1991). If an NS2 truncation orients the C-terminus to the ER lumen but prevents trafficking through the secretory pathway, the reporter should be glycosylated and thus sensitive to Endo H. Therefore the glycosylation status of wt SEAP and nsAP was assessed for the potential investigation of topology using glycosidase treatment. To exclude the possibility that any NS2 truncation was enabling modification of the SEAP glycan chain to an Endo H resistant form but were still sequestered in an inactive form; PNGase F sensitivity was also assessed.

3.2.7 Validation of wt SEAP as a glycosidase reporter

The activity of Endo H and PNGase F were confirmed using the HCV envelope glycoprotein E2. E2 is well characterised and contains 11 potential *N*-linked glycan acceptors (Goffard & Dubuisson, 2003). Lysates from Huh7 cells electroporated with full-length JFH-1 RNA were treated with both glycosidases (Figure 3.10A). A protein species at ~58 kDa was detected in untreated samples (E2) using the anti-E2 monoclonal antibody. Upon treatment with Endo H and PNGase F this species was lost and replaced by two protein species equivalent to ~36 (E2*) and ~34 kDa (E2†), respectively. This pattern corresponds to the cleavage of the 11 high-mannose *N*-linked glycans within the core of the chains and the complete removal of the glycans respectively (Stapleford & Lindenbach, 2011). The 30 kDa protein species was thought to be a non-specific host protein.

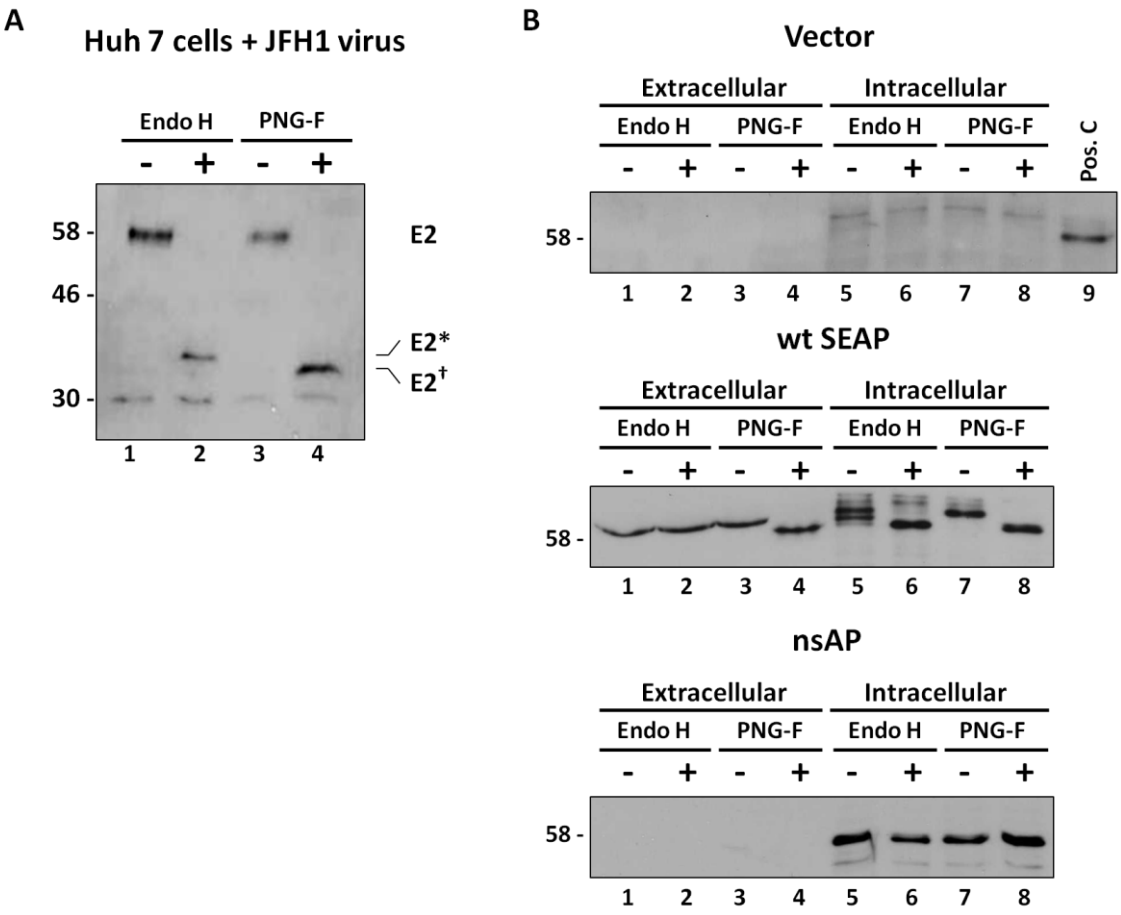


Figure 3.10 Functional confirmation of glycosidases and glycosylation state of SEAP expressed from reporter control constructs.

(**A**) Functional activity of Endo H and PNGase F was confirmed using HCV E2 glycoprotein. Huh7 cells were harvested 72 hr post-electroporation with full length JFH-1 virus. Lysates were denatured for 10 min at 95°C prior to incubation with indicated enzyme for 3 hr at 37°C. Samples were separated by SDS PAGE and the glycosylation sensitivity of E2 determined by immunoblotting with anti-E2 antibody (AP33 clone). E2 was completely de-glycosylated with PNGase F (E2†) and partially de-glycosylated with Endo H (E2*). (**B**) Glycosidase sensitivity of extra- and intracellular SEAP expressed from control vectors. COS7 cells were harvested 48 hours post-transfection. Extracellular protein was obtained from acetone-precipitated culture media resuspended in lysis buffer. SEAP was detected using anti-PLAP antiserum (Abcam). Lysate from untreated nsAP-transfected cells was used as a comparator for vector alone (Pos. C). Molecular weight markers (kDa) are shown to the left of SDS-PAGE gels.

Next, the intra- and extracellular glycosylation states of the control reporter enzymes were determined. SEAP possesses two putative *N*-linked glycan acceptor sites, although one is thought to be glycosylated in mammalian cells (Le Du *et al.*, 2001) (Appendix 3). To allow detection of proteins secreted into the culture media, one quarter volume of culture supernatant was acetone precipitated and resuspended in lysis buffer in an equivalent volume to that used for harvesting the cellular fraction. Samples were treated with Endo H and PNGase F and analysed by SDS PAGE and western blotting for the SEAP reporter using an anti-PLAP antibody (Figure 3.10B). A SEAP-specific species was observed in extracellular and intracellular wt SEAP samples but only in intracellular nsAP samples. No SEAP-specific species was detected in vector control samples as compared with an untreated intracellular nsAP control sample (Pos C.). A slower migrating non-specific protein species was observed in intracellular empty vector samples on higher exposures. Intracellular wt SEAP was susceptible to both glycosidases whereas extracellular wt SEAP was only deglycosylated by PNGase F. Protein expressed from the nsAP control vector was insensitive to both glycosidases.

The glycosylation and secretion patterns for wt SEAP and nsAP correlated to the enzyme activity data (Figure 3.9) and were consistent with a previous investigation which demonstrated that secreted SEAP was Endo H resistant but PNGase F sensitive (Davis, 1993). Together these observations confirmed that the enzyme produced by the nsAP construct was not oriented to the ER lumen and consequently not glycosylated (as demonstrated by insensitivity to glycosidases and a similar migration to glycosidase-treated wt SEAP).

To validate the application of glycosidase treatment of SEAP to an NS2-fusion system it was first necessary to determine the sensitivity of NS2 expressed alone to Endo H and PNGase F. C-terminally FLAG-tagged NS2 (JFH-1) was expressed in COS7 cells and lysates analysed by SDS PAGE and western blotting, and probed using antibodies raised against JFH-1 NS2 and the FLAG epitope (Figure 3.11A). Four protein species were detected using the anti-NS2 antibody (Figure 3.11A left panel) but only two prominent protein species were detected when the same membrane was re-probed for the FLAG epitope (Figure 3.11A right panel), with the larger species in each case corresponding to full-length NS2.

Glycosylation of p7 has not been reported in any context and in the proposed reporter-fusion system the p7/NS2 junction would be maintained for all NS2-SEAP fusions allowing cleavage by signal peptidase, thus p7 should not form part of the mature NS2-SEAP fusions. Taking these into account, and combined with the technical difficulties of reliably detecting p7, the glycosidase sensitivity of p7 was not determined.

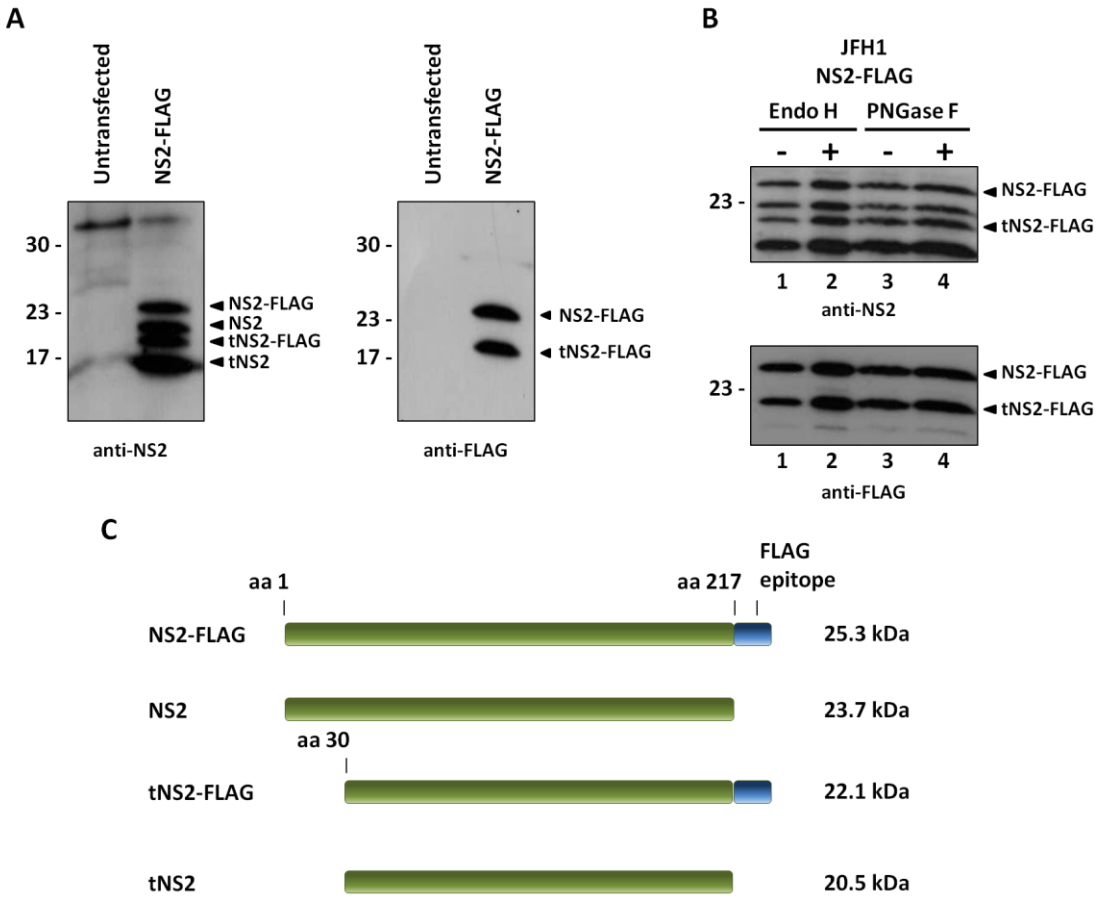


Figure 3.11 Glycosidase treatment of full-length NS2.

(**A**) Expression of C-terminally FLAG-tagged NS2 (JFH-1). Lysates from untransfected COS7 cells or cells transfected with NS2-FLAG were separated by SDS PAGE and NS2 was detected by immunoblotting using anti-NS2 polyclonal antisera (left panel) or M2 anti-FLAG monoclonal antibody (right panel). (**B**) Protein species were labelled to the right. Lysates from COS7 cells transfected with NS2-FLAG expression construct were treated with Endo H and PNGase F. Samples were immunoblotted using rabbit anti-NS2 antibody (top panel) then re-probed with mouse anti-FLAG antibody (bottom panel). Samples were collected 48 hours post-transfection. Molecular weight markers (kDa) are shown to the left. (**C**) Schematic representation of possible NS2-FLAG degradation products. NS2 sequence (green) with FLAG epitope (blue). Speculative N-terminus of tNS2 was placed at aa 30. Predicted masses for each fusion peptide are shown to the right.

The migrations of the four NS2-specific protein species were unaltered upon treatment with either glycosidase, indicating that NS2 was not *N*-glycosylated (Figure 3.11B). Thus glycosidase sensitivity of an NS2-SEAP fusion could be deduced to be the result of a SEAP-specific glycosylation event and therefore provide a means of determining the C-terminus orientation of the NS2 sequence within the fusion.

Several groups have reported a truncated form of NS2 (tNS2) detectable using polyclonal NS2 antiserum which migrates at ~18kDa. Corresponding truncations were observed when expression of tagged NS2 confirmed using antibodies directed against C-terminal tag epitope but not against N-terminal tags. The nature of tNS2 has not been determined but it is likely to be the product of a specific cleavage event in the vicinity of aa 35 of NS2. Furthermore, as NS2 is responsible for catalysing the cleavage of the NS2/3 junction it is possible that NS2 catalyses the cleavage of non-native sequences i.e. a non-NS3 sequence, from the NS2 C-terminus. It was therefore proposed that the four protein species corresponded to full-length NS2-FLAG, full-length NS2, tNS2-FLAG and tNS2 (Figure 3.11C).

Alternatively, as NS2-SEAP showed no evidence of auto-cleavage (Figure 3.7 lane 5) it is possible that the fusions were subjected to non-specific cleavage by cellular proteases within the linker sequence, causing the pattern described in Figure 3.11C. This appears to be a common phenomenon encountered with reporter fusions and has recently be noted by Popescu *et al.,* for YFP (Venus yellow fluorescent protein)-tagged NS2 and CFP (Cerulean fluorescent protein)-tagged p7 (Popescu *et al.*, 2011).

3.2.8 Glycosidase sensitivity of NS2-SEAP fusion proteins

To investigate whether any NS2 truncations orientated the reporter to the ER lumen, the sensitivity of the panel of fusion proteins to glycosidases were determined. Lysates from cells transfected with SEAP reporters were treated with Endo H and PNGase F (Figure 3.12). As shown previously (Figure 3.10) wt SEAP was sensitive to both glycosidases as characterised by a reduction in size of ~2 kDa by SDS PAGE corresponding to the removal of a glycan chain, whereas nsAP results showed no evidence of glycosylation. All p7-SEAP fusions (S10-S13 and Δ63) were sensitive to both glycosidases. Of the NS2 containing fusions only S6 and S9 appeared to facilitate glycosylation of the reporter. Two distinct protein species were detected in the untreated samples for S6 which migrated to just under and just over 58 kDa respectively (Figure 3.12). Upon treatment with either glycosidase the protein remained unchanged in size, although somewhat reduced in intensity. This protein was thought to represent a non-specific protein as it was also detected in vector only transfected cells. The smaller protein species was

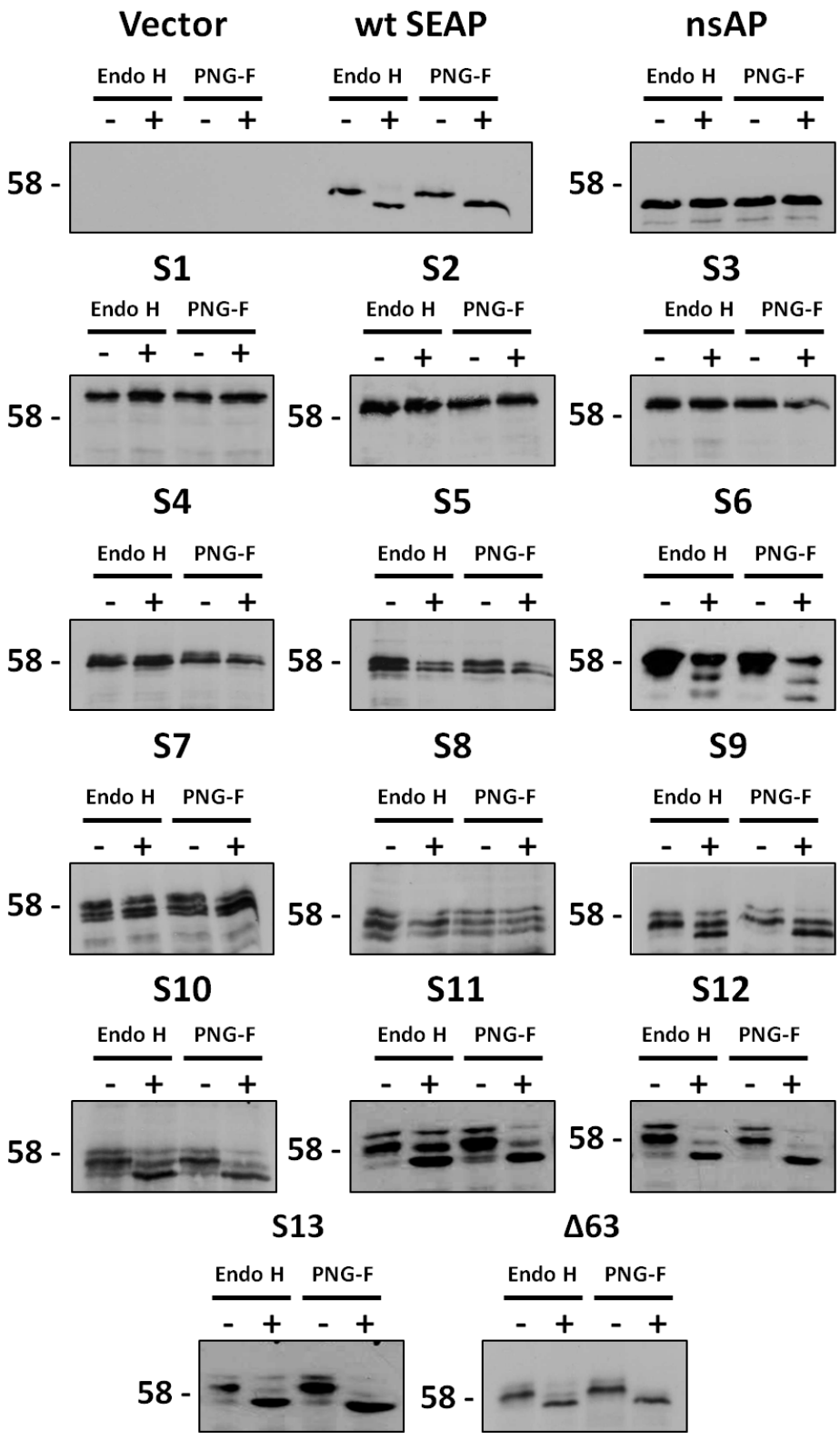


Figure 3.12 Glycosidase sensitivity of NS2-SEAP fusions.

Lysates from COS7 cells expressing the NS2-SEAP fusions were treated with Endo H and PNGase F. Cells were harvested 48h post-transfection. Molecular weight markers (kDa) are shown to the left.

noticeably shifted down indicating the complete removal of a high-mannose glycan chain from all fusion peptides. A third protein appeared upon glycosidase treatment which was smaller still; however, the nature of this protein was not clear. While wt SEAP and Δ63 were efficiently de-glycosylated by both enzymes, S9-S13 showed varying degrees of sensitivity to Endo H and PNGase F. All untreated samples contained a single predominant protein species. Two species of varying prominence were detected in treated samples: one of a similar weight as the untreated sample and one slightly smaller, equivalent to a reduction in molecular weight of ~2 kDa. Interestingly, a further species was detected at a low level of expression in the untreated samples that corresponded to the prominent de-glycosylated form of the reporters. This may represent an unglycosylated population of the fusion proteins, possibly caused by inefficient glycosylation of the cryptic glycan acceptor site.

Collectively the data suggests that NS2 forms a three TMD topology with the N- and C-termini oriented to the lumen and cytosol respectively, and a luminal loop in the vicinity of residue 70. The glycosylation of S9 suggests that either the first 27 residues of NS2 cannot function as a TMD in this system, or that the residues do not completely represent the first TMD of NS2.

3.2.9 Analysis of reporter activity in NS2-β-galactosidase fusions

To determine whether the use of β-gal as part of a dual reporter system could be extended for the study of ER *trans*-membrane proteins by displaying localisation dependent enzyme activity; a control reporter construct was also generated where the signal peptide encoded

within the SEAP reporter vector was generated as an oligo-nucleotide and inserted upstream of the wt β-gal coding sequence. This construct, termed 'SP β-gal', was therefore predicted to orientate the enzyme into the ER lumen. The same NS2 truncations used in the NS2-SEAP fusions (Figure 3.4) were generated as C-terminal fusions to β-gal and termed B1 – B10.

NS2 β-gal fusions were expressed in COS7 cells and assessed for fusion protein expression levels and reporter activity by colorimetric O-nitrophenyl-β-D-galactopyranoside (ONPG) solution-based assay and 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) staining of fixed cells (Figure 3.13). All NS2-β-gal fusions showed reporter activity to varying degrees by both enzyme functionality methods. The specificity of the assays for β-gal activity was confirmed by a completely negative response from cells transfected with vector alone. Enzyme responses in the ONPG and X-gal staining assays appeared to mirror fusion-protein expression levels, which matched the previously described SEAP fusions in that all NS2-sequence containing fusions were detected at lower levels than controls and the B10/S10 fusion. Unfortunately, as the SP β-gal control reporter displayed positive enzyme activity in both

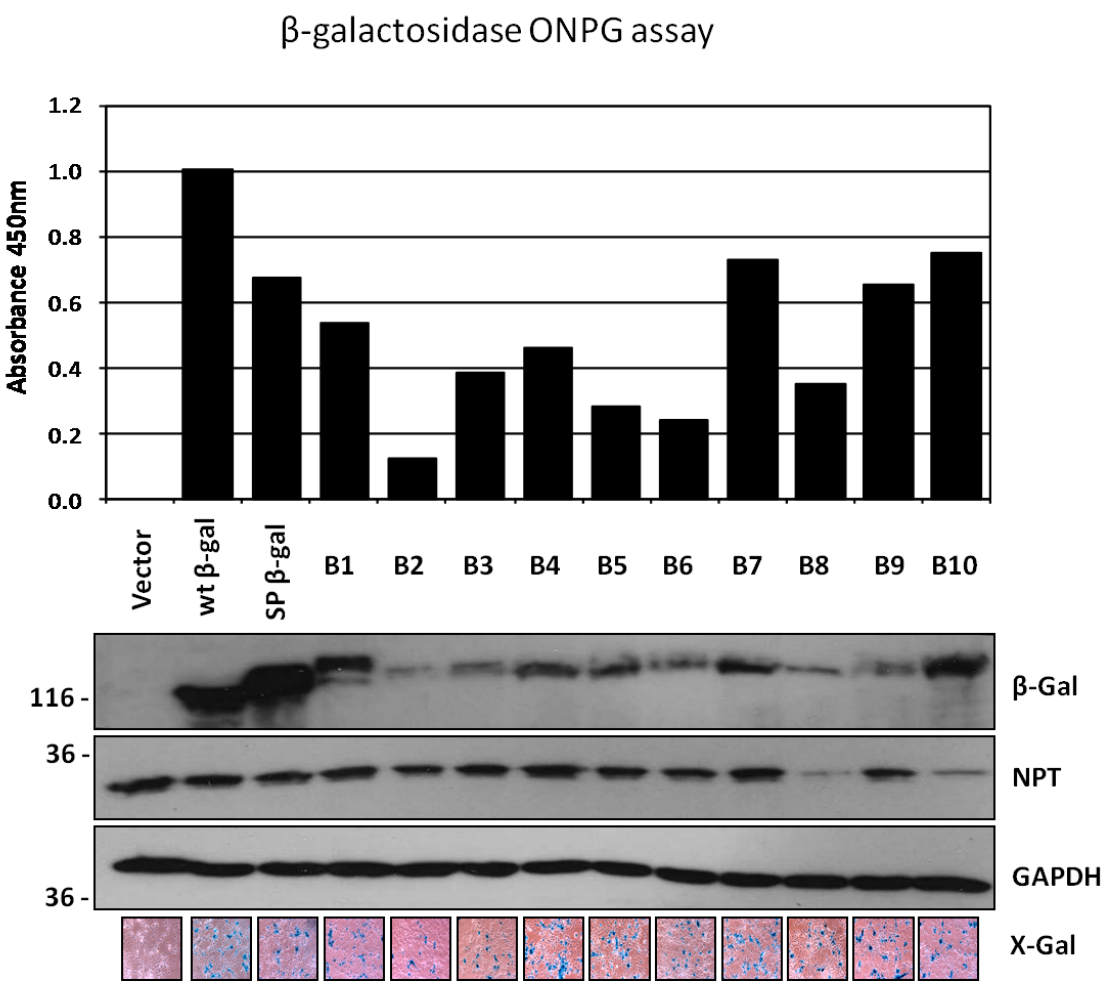


Figure 3.13 Expression and reporter activity of NS2-β-gal fusions.

COS7 cells were transfected in duplicate with NS2-β-gal fusion series. 48h post-transfection cells were either; lysed and used to determine reporter activity by ONPG colorimetric assay (top graph) and fusion expression by immunoblotting (middle three panels) or; fixed and incubated with X-gal reagent (bottom row of panels). Lysates were incubated for 10 min in ONPG reaction mixture at 37°C. Absorbance was read at 420nm. Fusion protein expression levels were determined by probing for β-galactosidase. Transfection efficiency was gauges from levels of the vector-encoded neomycin phosphotransferase II resistance gene (NPT). The cellular protein glyceraldehydes-3-phosphate dehydrogenase (GAPDH) was used as a loading control. Molecular weight markers (kDa) are shown to the left. Fixed cells were overlaid with X-gal reagent and incubated for 16h at 37°C. Cells were viewed microscopically and imaged using a Digital Sight DS-L1 (Nikon). Cells positive for β-galactosidase activity stained blue. Representative wide-field images are shown.

formats no further information could be discerned regarding fusion protein topology from analysis of enzymatic function.

Low resolution separation of cell lysates by SDS PAGE indicated that SP β-gal ran slower than wt β-gal. This could represent irregular folding in the ER lumen or a non-native post-translational modification, such as the addition of a glycan chain to a cryptic glycosylation site. As enzyme function was maintained for it is unlikely that folding differed significantly from wt. B1 migrated as a doublet with the larger protein appearing to correspond to the full-length NS2 β-gal fusion, and the smaller possibly representing a cleavage event product near the NS2/β-gal junction, as it migrated faster than the other fusion peptides and at a similar rate to SP β-gal.

3.2.10 Glycosidase sensitivity of NS2-β-galactosidase fusions

β-gal is a cytosolic protein and so is not glycosylated under normal conditions. Analysis of the β-gal sequence identified 4 possible *N*-linked glycan acceptor sites (Appendix 3); two of which were predicted with a high probablility to be glycosylated of oriented to the ER lumen. To investigate whether this apparent gain in molecular weight was due to the glycosylation of a cryptic glycan acceptor site, lysates from COS7 cells transfected with the NS2-β-gal fusions were treated with Endo H and PNGase F and analysed by SDS PAGE (Figure 3.14). Samples were run for longer than in the initial expression profile (Figure 3.13) to achieve better separation. As expected wt β-gal migrated as a single protein that was unaltered by treatment with either glycosidase. Surprisingly, the SP β-gal control migrated as two distinct species, and the more intense larger species was lost after treatment with both enzymes. Concomitantly, the smaller protein became more intense. Consistent the SEAP fusion data; the only NS2-β-gal fusion that displayed sensitivity to glycosidase treatment was B6. Unlike the corresponding SEAP fusion, B9 did not show signs of having been glycosylated. B10, which encodes p7 as a fusion to β-gal with no intervening NS2 sequence, also showed sensitivity to glycosidase treatment. As with the SP β-gal control, untreated B6 and B10 appeared to migrate as doublets but were reduced to single protein species with the loss of the higher molecular weight protein after treatment with either enzyme. This pattern was indicative of only a sub-population having been glycosylated and was consistent with the SEAP fusions, in that truncation of NS2 at residue 70 resulted in orientation of the artificial NS2 C-terminus to the ER lumen.

3.2.11 Analysis of p7 trafficking

As described in section 3.2.2; a series of S10-derived mutants were generated to enable the examination of p7 trafficking. S11 contained a mutation to the signal peptidase cleavage site;

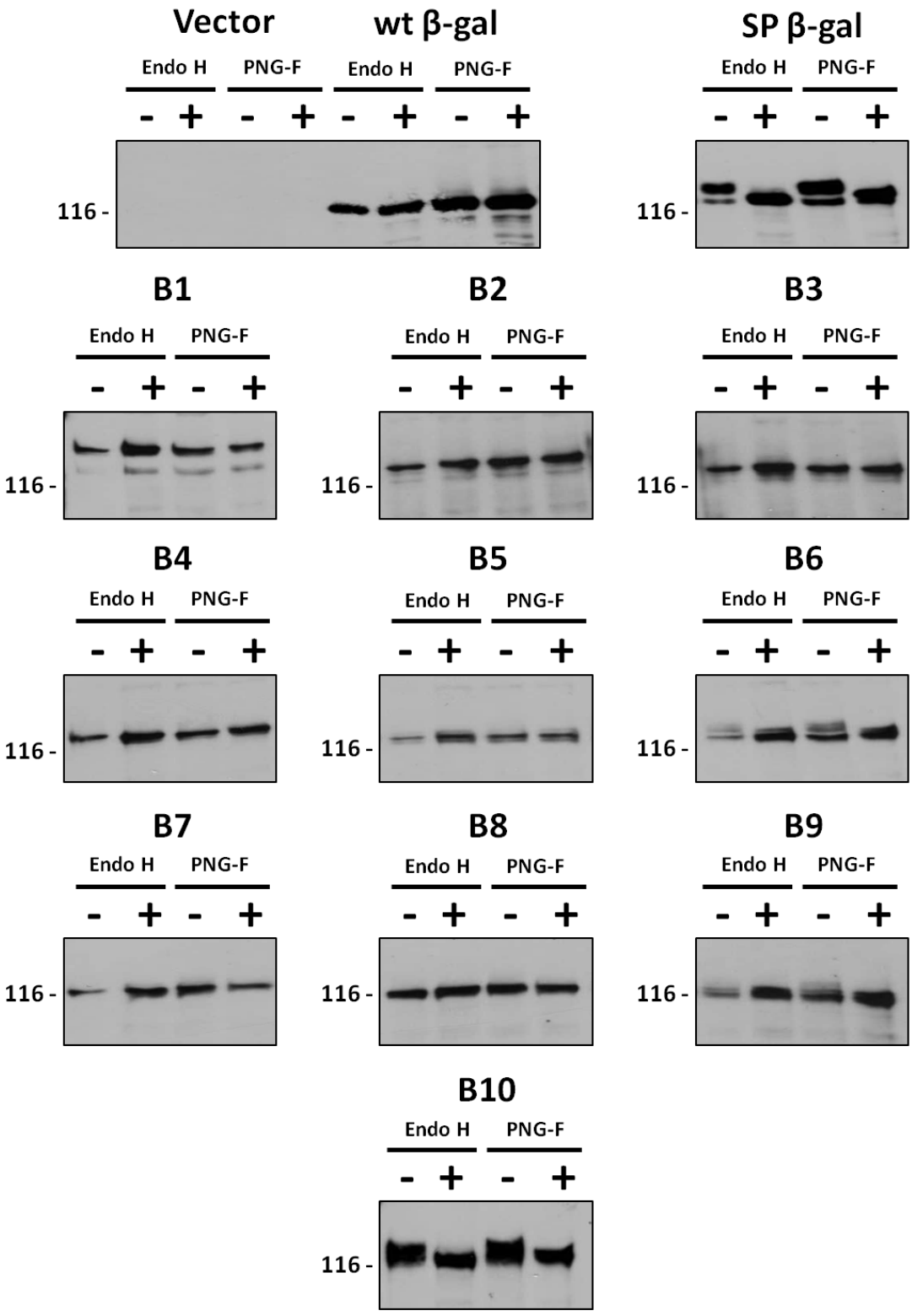


Figure 3.14 Glycosidase sensitivity of NS2-β-gal fusions.

Lysates from COS7 cells expressing the NS2-β-gal fusions were treated with Endo H and PNGase F. Cells were harvested 48h post-transfection. Molecular weight markers (kDa) are shown to the left.

S12 contained a mutation to the furin cleavage site, S13 contained both mutations and Δ63 encoded p7 as a direct fusion to SEAP without the intervening furin cleavage linker and without the C-terminal residue of p7 (alanine 63), which was expected to abrogate signal peptidase cleavage at this site.

Initially it was found that the p7-trafficking mutant fusions expressed to similar, high levels (Figure 3.7) and that the reporters were all active and sensitive to both Endo H and PNGase F suggesting that p7 does indeed orient the C-terminus to the ER lumen (Figure 3.12). To investigate whether any of the constructs showed altered reporter activity they were transfected into cells and the supernatants and lysates assessed for SEAP activity and protein expression, respectively. Surprisingly, S10-S13 each showed extracellular SEAP activity levels comparable to wt SEAP, and Δ63 showed SEAP activity nearly double that of wt SEAP (Figure 3.15). Western blotting of cell lysates following SDS PAGE separation showed all fusions expressed at similarly high levels, and significantly higher than wt SEAP (Figure 3.15).

To further examine whether the increased SEAP activity of the Δ63 peptide over wt SEAP was a result of higher transfection efficiency; the extracellular SEAP activities of each of the trafficking mutants and the nsAP control vector were normalised to respective NPT expression levels and represented graphically (Figure 3.16). When secreted SEAP activity was shown as a function of NPT expression it became apparent that all of the p7-SEAP fusion peptides secreted active SEAP to comparable levels. Δ63 and S13 were engineered to possess neither functional signal peptidase or furin cleavage sites yet were still able to secrete active reporter. There are several possible explanations for this observation; i) these fusions were subject to an additional cleavage event by a cellular protease resulting in secretion of the active reporter; ii) p7 was unable to anchor to membranes and was being secreted as an uncleaved fusion to SEAP; iii) extracellular activity noted for all p7-fusions was the result a population of p7 released from the cell in exosomes.

The amino acid sequence for the p7-SEAP fusions was analysed for the presence of non-SP, non-furin cleavage motif. This revealed a consensus motif for thrombin cleavage within the C-terminal domain of p7, but because thrombin is activated from pro-thrombin in the extracellular matrix by the pro-thrombinase complex it was not considered to be the cause of intracellular p7-SEAP cleavage. Core protein of HCV is released from E1 by signal peptidase in the ER lumen and the C-terminal domain of core is further processed within the membrane by signal peptide peptidase (SPP) (McLauchlan *et al.*, 2002). If p7 is cleaved within the C-terminal

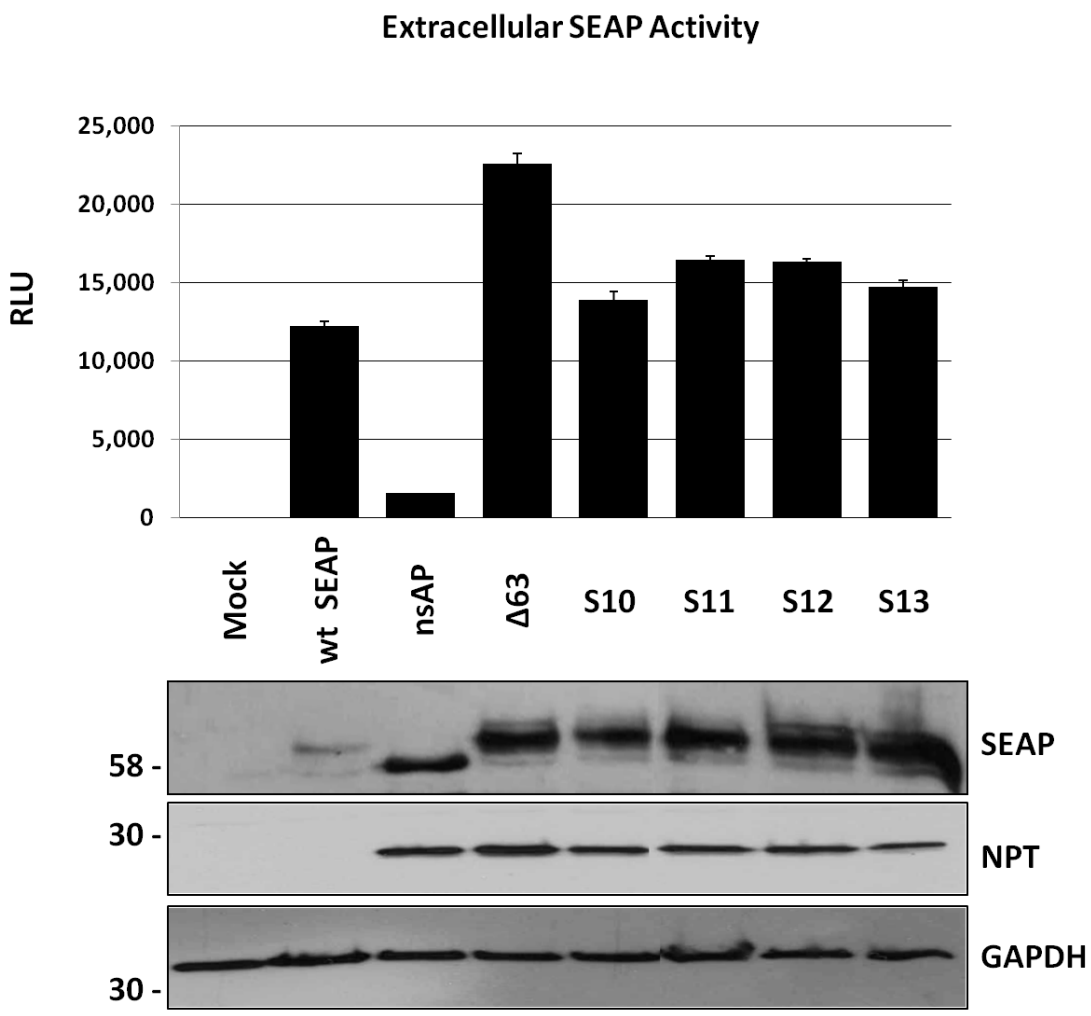


Figure 3.15 Analysis of p7 trafficking mutants.

Constructs encoding SPp7 fused to SEAP without any intervening NS2 sequence were expressed in COS7 cells and analysed for the levels of active extracellular reporter (top graph). Reporter activity is shown in relative light units (RLU). Lysates were assessed for levels of SEAP, vector-encoded NPT as a marker of transfection efficiency and endogenous GAPDH as a loading control (bottom panels). Molecular weight markers (kDa) are show to the left.

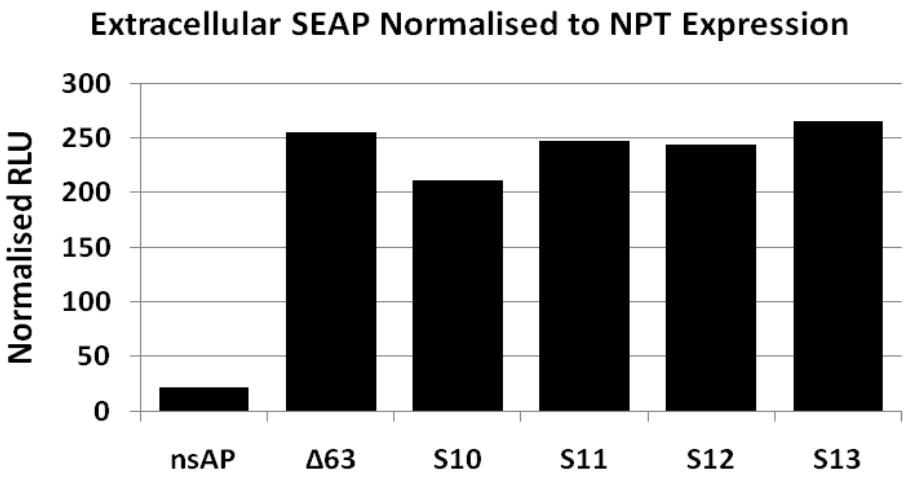


Figure 3.16 Normalised activity of secreted reporter enzymes expressed from p7 trafficking mutants.

Extracellular SEAP activity values of p7-SEAP fusion shown in Figure 3.15 were normalised to their corresponding, background-subtracted, NPT expression levels and re-plotted. Reporter activity was defined in normalised relative light units (RLU).

helix it would be a credible target for SPP, unfortunately there is no known cleavage consensus sequence for SPP.

As p7 has a mass of 7 kDa uncleaved p7-SEAP fusion would be expected to have a retarded migration by SDS PAGE analysis compared with wt SEAP, yet no such difference was apparent between any of the p7 trafficking mutants and wt SEAP (Figure 3.7). The presence or absence of uncleaved p7 in lysates, on the cell surface or in precipitated culture supernatants could not be directly addressed due to poor detection of genotype 2a p7 under these conditions.

3.3 Discussion

A model of NS2 topology involving an odd number of TMDs has been proposed whereby the N- and C-termini are oriented to the ER lumen and cytoplasm respectively (Jirasko *et al.*, 2008; Phan *et al.*, 2009; Pietschmann *et al.*, 2006) (Figure 3.17 **- Model B**). However this model conflicts with experimental data suggesting a four TMD topology with both termini oriented to the ER lumen (Yamaga & Ou, 2002) (Figure 3.17 **- Model A**).

3.3.1 Membrane topology of NS2

Comparison of the outputs from 6 different topology prediction programs was used to generate a consensus computer model (Figure 3.3A, Figure 3.17 **– Model C**) for NS2 topology which defined three putative TMDs between positions 7-27, 29-49 and 63-86 (Table 3.1). The similarity between the computer model (Figure 3.17 **– Model C**) and the model published by Jirasko *et al.,* (Jirasko *et al.*, 2008) (Figure 3.17 **– Model B**) are unsurprising as 5 of the 6 programs used were common to the two studies. The computer model (Figure 3.17 **– Model C**) predicted that the fusion of a reporter to aa 52 and aa 70 would result in luminal orientation of the reporter as they were modelled to terminate NS2 sequence after the second TMD and within the third TMD, respectively (Table 3.1).

Control reporter enzymes where SEAP was expressed in the ER lumen (wt SEAP) or the cytosol (nsAP) were active and inactive respectively (Figure 3.6). These localisations were further confirmed by glycosidase treatment; intracellular wt SEAP was functionally active and sensitive to both Endo H and PNGase F, whereas extracellular wt SEAP was active but insensitive to Endo H treatment. nsAP was insensitive to this treatment. This pattern of glycosidase sensitivity demonstrates that secretion of SEAP following final glycan modification in the Golgi occurs rapidly and efficiently and that activity of SEAP is dependent on luminal localisation (Figure 3.10B).

The corresponding β-gal controls produced active enzyme irrespective of predicted reporter orientation about the ER membrane (Figure 3.13). As a result, topological information of the NS2-truncations could not be gained from the enzyme activity of the β-gal series. However, the apparent *N*-glycosylation of a cryptic asparagine acceptor residue when β-gal was translocated to the ER lumen (SP β-gal; Figure 3.14) enabled comparative characterisation of the glycosylation states of NS2 fusions to both reporters.

The extracellular and intracellular reporter activities of the NS2-SEAP fusions were determined by chemi-luminescent assay (Figure 3.9). Of the NS2-containing fusions only fusion of SEAP to

the first 27 amino acids of NS2 (S9) resulted in the secretion of active reporter. The relatively low extracellular reporter activity of S9 (nearly 5 fold down on wt SEAP) could have been due to instability or inefficient secretion of this fusion compared to wt SEAP and the p7-SEAP fusions S10 – S13 (Figure 3.7). The inability of a fusion to secrete the reporter would have led to accumulation of intracellular active SEAP; however intracellular reporter activity followed a similar trend with only S9 showing levels markedly above background (Figure 3.9). S6 displayed a small but reproducible level of SEAP activity albeit only 3 times that of background. These data suggest that only the first 27 residues of NS2 orient the reporter to the ER lumen where it subsequently becomes activated and secreted from the cell.

Consistent with the enzyme activity assays intracellular S9 and S10 were sensitive to both glycosidases (Figure 3.12). Interestingly, S6 also displayed sensitivity to both enzymes. A similar pattern of apparent glycosylation was noted for the NS2-β-gal fusions (Figure 3.14) with SP β-gal, B6 and B10 all displaying sensitivity to both enzymes, although corresponding glycosylation of the reporter encoded by B9 was not detected.

The computer model (Figure 3.17 **– Model C**) for NS2 predicted that a reporter fused to the C-terminus of aa 52 would be orientated to the ER lumen. Glycosylation analysis of the SEAP and β-gal fusions suggested that this was not case (Figure 3.12 and 3.14). This could be due to incorrect assignment of the second TMD within the computer model or that fusion of the reporter immediately proximal to the C-terminus of the second TMD impeded the ability of residues 1-52 to stably form a second TMD. The glycosidase sensitivity of S6 and B6 (aa 1-70) was concurrent with the predicted phenotype in that it appeared to orientate the reporter to the ER lumen. The computer model positioned residue 70 within the N-terminal half of a 24 aa TMD, spanning residues 63-86. Truncations either side of S6 (S5 – aa 87, S7 – aa 52) were not glycosylated implying that their respective C-termini were oriented to the cytosol. Taken together, the glycosylation data for S/B5, S/B6 and S/B7 describes NS2 topology in that residues 1-52 are sufficient to form only one stable TMD, aa 1-70 form two TMDs and that residues 1-87 are capable of forming 3 TMDs. This provided firm evidence for the presence of a luminal loop in NS2 between residues 52-87.

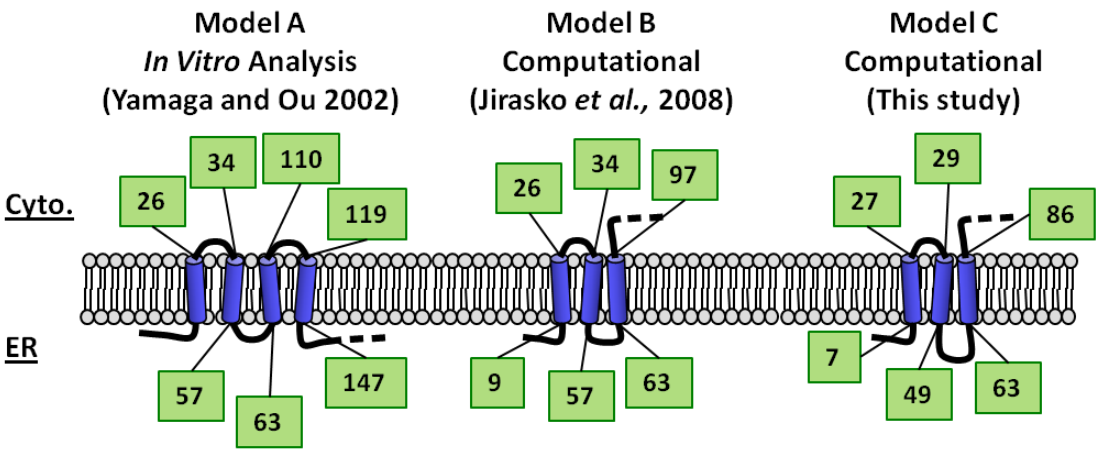
The glycosylation profile of S9 was inconsistent with the computer model and previous experimental data identifying this sequence contains the first TMD of NS2 (Jirasko *et al.*, 2008; Pietschmann *et al.*, 2006; Yamaga & Ou, 2002), in that fusion of a reporter to the C-terminus of the first 27 aa of NS2 appeared to orientate NS2 to the ER lumen. Due to the poor expression and low SEAP activity of the S9 fusion it was not possible to discern whether all fusion

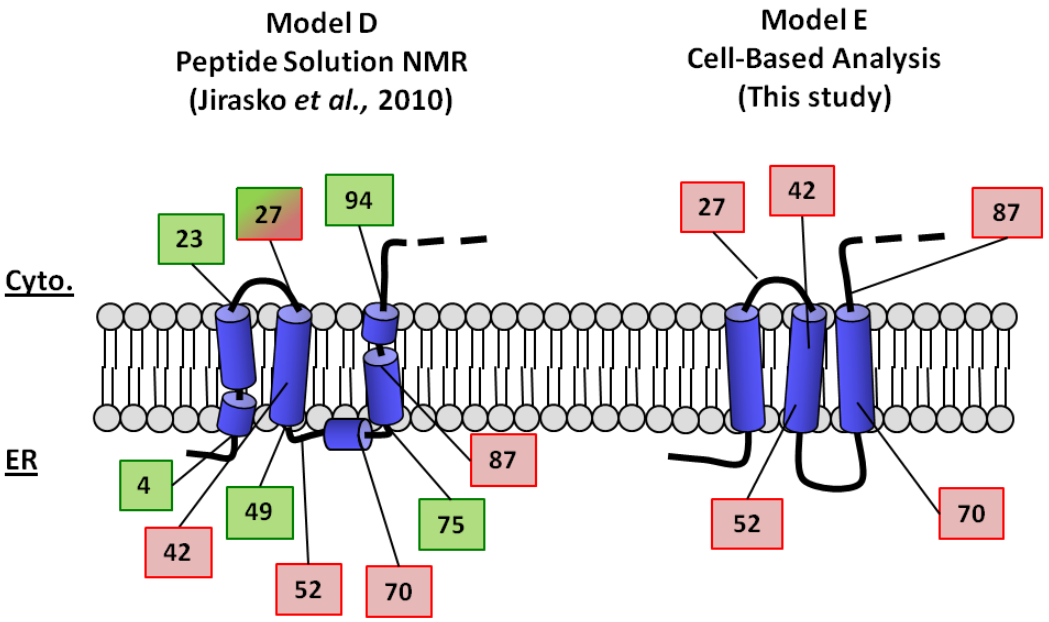
molecules or only a sub-population oriented their reporter to the lumen. Attempts to determine whether aa 1-27 could be recovered in the extracellular media were unsuccessful (data not shown). Furthermore, limited resolution of the S9 and S10 fusions by SDS PAGE analysis meant that it was not possible to investigate the presence of uncleaved S9 fusion molecules (Figure 3.7). The disparity between S9 and B9, whereby S9 was sensitive but B9 insensitive to glycosidase treatment, was likely caused by disruption to the correct topology by one of the reporters and thus prevented definitive characterisation of the topology of aa 1-27.

As the reporters encoded by S6 and B6 were oriented to the ER lumen, yet S6 displayed little to no SEAP activity, either intra or extracellularly, it is likely that the N-terminal 70 residues of NS2 were capable of retaining the fusion protein in a pre-TGN compartment as trafficking to the TGN would have resulted in cleavage by furin. The lack of extracellular reporter activity also suggests that NS2 is not secreted from the cell (Figure 3.9).

Several membrane proteins including HCV NS4B and p7 have been speculated to assume multiple topologies (Isherwood & Patel, 2005; Lundin *et al.*, 2003) with their C-termini able to orientate to both sides of the ER membrane. Two distinct topologies for NS2 have been predicted in which the C-terminus is oriented to opposing sides of the ER membrane, while the N-terminus is always predicted as luminal (Figure 3.17). Numerous inconsistencies have been noted between the studies carried out *in vitro* and those examining NS2 as part of the viral polyprotein. The studies that led to the generation of the 4 TMD model (Figure 3.17 **– Model A**) were principally conducted *in vitro* and it was reported that the C-terminus of NS2 is oriented to the ER lumen (Yamaga & Ou, 2002). More recently, experiments employing partial membrane permeabilisation in combined with confocal IF microscopy or protease treatment showed that the C-terminus of NS2 in virus-expressing cells to be oriented to the cytosol (Ma *et al.*, 2011). In the present study neither S1 nor B1, encoding the full-length NS2 protein, displayed any glycosidase sensitivity. This study, combined with the work of Ma and colleagues (Ma *et al.*, 2011), demonstrates that NS2 likely forms a single topology in cells with the C-terminus oriented to the cytosol. It is possible that the observed difference in the localisation of the C-terminus of NS2 between cell-based and *in vitro* studies could be due to a requirement for a cellular chaperone to achieve correct topology.

Together, the data presented here enabled the generation of an experimentally defined model topology of NS2 topology (Figure 3.17 **– Model E**). As discussed, NS2 C-terminal truncations displayed phenotypes consistent with a three TMD topology. The difference in glycosidase sensitivity between S9 and B9 meant that no clear conclusions could be made regarding its


Model A
In Vitro Analysis
(Yamaga and Ou 2002)
Model B
Computational
(Jirasko et al., 2008)
Model C
Computational
(This study)
Cyto.
ER
34
110
26
119
57
63
147
26
34
97
9
57
63
29
27
86
7
49
63


Model D
Peptide Solution NMR
(Jirasko et al., 2010)
Model E
Cell-Based Analysis
(This study)
Cyto.
ER
23
27
94
4
42
49
52
70
75
87
27
42
87
52
70

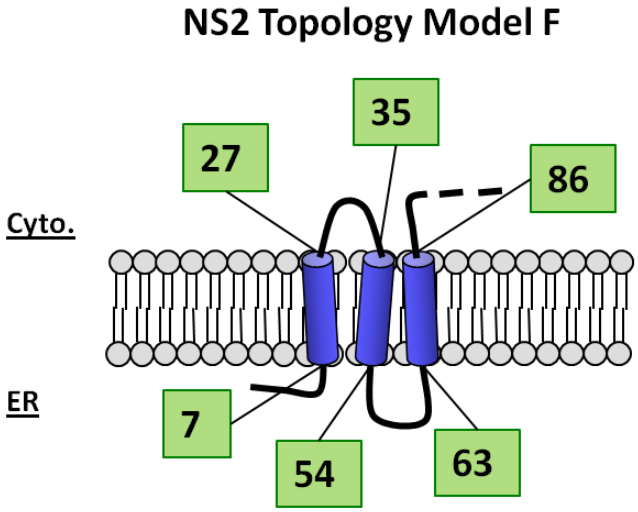

NS2 Topology Model F
Cyto.
ER
35
27
86
7
54
63

Figure 3.17 Models of NS2 Topology.

Model A - 4 TMD topology defined based upon glycosylation, protease protection and primary sequence analysis (Yamaga & Ou, 2002). **Model B -** 3 TMD topology based on predictions from 6 different programs (Jirasko *et al.*, 2008). **Model C -** Consensus computer model defined in Table 3.1. **Model D -** 3 TMD topology composition following 3D structure determination of three distinct peptides (aa 1-27, aa 27-59 and aa 60-99) by solution NMR (Jirasko *et al.*, 2008; Jirasko *et al.*, 2010). **Model E -** Generated from the experimental data described in this chapter. **Model F -** Predicted position of TMDs based upon experimental data and computation Model C. Residues marking the predicted boundaries of TMDs (green boxes), the position of C-terminal truncation points (red boxes) or both (green and red box) are noted. The means by which the models were derived are noted. NS2 sequence within fusions S6 and B6 was terminated at residue 70. The N-terminus in each model is localised to the ER lumen. The catalytic domains are not shown.

orientation about the membrane. In this model, aa 27 it is shown as positioned on the cytosolic face of the ER membrane in accordance with the computer model (Figure 3.17 **– Model C**).

Since the commencement of this work solution NMR structures for three peptides representing the N-terminal 99 residues of NS2 from the Con-1 isolate have been published (Jirasko *et al.*, 2008; Jirasko *et al.*, 2010). Residues 1-27 synthesised as a peptide were shown to adopt an overall α-helical conformation in 50% TFE and the authors proposed that residues 3-22 form the first TMD of NS2. As with the first TMD, peptides spanning the predicted second (aa 27-59) and third (aa 60-99) TMDs were expressed as peptides and shown to assume α-helical conformations in 50% TFE. Analysis of the second putative TMD revealed that it formed a single helix spanning residues 31-52 and possessed characteristics of an amphipathic helix i.e. one that interacts with the membrane interface rather than penetrating the hydrophobic core of the membrane. The authors noted that charged and polar residues along the spine of the helix would need to be neutralised by residues in a complementary domain to enable a stable, membrane spanning conformation (Jirasko *et al.*, 2010). It is therefore possible that in the full length protein aa 52 is oriented to the ER lumen and that polar trans-membrane residues are not stabilised when NS2 sequence is terminated at aa 52, but are when sequence is terminated at aa 70.

The third peptide formed three small helical domains with the N-terminal domain predicted to lie parallel to the membrane with the other domains forming a third putative TMD between residues 74-95 (Jirasko *et al.*, 2010). A reconstruction of the three separate peptide structures is shown in Figure 3.17 **– Model E**. The positions of the first and second putative TMDs defined in **Model E** map closely to those modelled based upon the current work (Figure 3.17 **- Model D**). The most significant difference between **Model D** and **Model E** is the position of the C-terminus of the putative third TMD. Jirasko *et al.* modelled this TMD as terminating at residue 95 (Jirasko *et al.*, 2010), whereas truncation analysis presented in the current work suggested that aa 87 was oriented to the cytosol. Another difference between the two models is the positioning of the C-terminal boundary of the second TMD. From the NMR structure data it was calculated that α-helical structure ended at residue 49 (Jirasko *et al.*, 2010). However, in the present study aa 52 was not oriented to the ER lumen.

By combining the experimentally determined **Model D** with the computer model (Figure 3.17 **- Model C**); a more detailed model for NS2 topology was defined (Figure 3.17 **- Model F**). Model F defines three TMDs between residues 7-27, 35-54 and 63-86 and provides a more detailed,

experimentally developed, model for future investigations to more precisely map NS2 topology.

3.3.2 P7 topology and trafficking

The N- and C-termini of p7 are thought to localise to the ER lumen as cleavage of p7 from adjacent proteins (E2 and NS2) is facilitated by signal peptidase in the ER lumen(Carrere-Kremer *et al.*, 2002). Purified p7 spontaneously inserts into lipid bi-layers and was proposed to form two membrane-spanning domains (Griffin *et al.*, 2005; Pavlovic *et al.*, 2003; StGelais *et al.*, 2007) and p7 expressed without a signal peptide is functionally active (Griffin *et al.*, 2005; Jones *et al.*, 2007; Wozniak *et al.*, 2010). The p7 monomer has been solved by solution NMR and shown to assume a hair-pin structure in methanol with two α-helical domains separated by a short, basic-charged loop (Foster *et al.*, 2011; Montserret *et al.*, 2010).

S10 encoded a polypeptide of SPp7 and SEAP separated by a furin cleavage linker. Cells transfected with S10 and S10-derived mutants secreted high levels of active SEAP (Figure 3.7). Treatment of cell lysates with Endo H and PNGase F caused a decrease in the molecular weight of the SEAP reporter, such that its migration was equivalent to nsAP; consistent with it having been glycosylated (Figure 3.12). Together these observations confirm previous work characterising the orientation of the C-terminus of p7 to the ER lumen and therefore that p7 likely forms two TMDs in a hairpin-like topology.

Analysis of the extracellular reporter activity expressed from the p7-SEAP trafficking mutants revealed that they all secreted high levels of active reporter (Figure 3.15). The amount of active reporter released from S10-S13 was comparable to the wt SEAP control vector and the Δ63 showed significantly more secreted active reporter, albeit to varying levels. Moreover, when extracellular reporter activities were normalised to levels of the vector-encoded resistance gene NPT, there was no discernible difference in levels of active reporter secretion caused by mutation of the two potential cleavage sites (Figure 3.16).

While it is possible that the fusion proteins were undergoing an additional cleavage event, possibly mediated by SPP, it is interesting to note that intracellular SEAP protein levels of the p7 mutants were far in excess of the wt SEAP control (Figure 3.15), yet extracellular reporter activities were comparable between the wt SEAP control and the p7 mutants (Figure 3.15). This implies that either only a population of the reporter was secreted from cells transfected with the p7 mutants, or that the specific activity of the fusion reporter compared to the wt SEAP was much lower. While both hypotheses are testable by chemical inhibition, time constraints prevented further investigation, it was not possible to distinguish between

whether p7 is retained in the ER where the fusion protein was cleaved and secreted, or that a population of p7 is secreted from the cell in exosomes.

3.4 Conclusions

The topology analysis of a series of NS2 C-terminal truncations has provided evidence that the N- and C-termini of NS2 are oriented to the ER lumen and the cytosol respectively. Furthermore, termination of the NS2 sequence at aa 70 resulted in glycosylation of the C-terminal reporter, while truncation at aa 52 and aa 87 did not result in reporter glycosylation. Taken together, these results provide conclusive evidence that NS2 forms three TMDs.

The confirmation that NS2 forms a luminal loop region enabled the generation of a more defined model of NS2 topology providing a firm basis for precise mapping of the TMDs of NS2.

CHAPTER 4: MEMBRANE ASSOCIATION STUDIES

4.1 Introduction

The extent to which a peptide domain interacts with membranes i.e. whether it forms a TMD or an peripheral association via an amphipathic helix, can be determined by chemically treating isolated cellular membranes with high molarity NaCl, urea or alkaline carbonate to selectively remove peripherally associated proteins or detergents like Triton X-100, to solubilise lipids and integral membrane proteins (Fujiki *et al.*, 1982; Matsunaga *et al.*, 2002). NS2 is an integral membrane protein as its association with microsomal membranes cannot be disrupted post-translationally by washing with 4.5M urea (Santolini *et al.*, 1995), high molar salt or alkaline carbonate (Yamaga & Ou, 2002). NS2 is capable of independently inserting into membranes and deletion mapping indicated the presence of two internal membrane-targeting sequences within residues 30-74 and 119-151 (Yamaga & Ou, 2002). The presence of these sequences, and whether they confer an integral or peripheral membrane association, has not been explored in a cell-based system. Furthermore, although expression of NS2 alone has been known to co-localise with ER markers by confocal IF microscopy (Franck *et al.*, 2005; Kim *et al.*, 1995; Yamaga & Ou, 2002; Yang *et al.*, 2006), the domains responsible for this sub-cellular targeting have not been determined.

Prior to the completion of the studies described in Chapter 3 a solution NMR structure of the N-terminal 27 aa of NS2 was published (Jirasko *et al.*, 2008). This structure described the putative first TMD of NS2 as spanning residues 4-23. Based upon this evidence the computer model of NS2 topology described in the previous chapter (Table 3.1) was refined. Using this refined model a series of NS2 truncations were designed to functionally dissect NS2 membrane-targeting, subcellular distribution and membrane affinity in the absence of other viral protein sequence.

4.2 Results

Concomitantly with the C-terminal truncation analysis described in Chapter 3; the computer model (Figure 4.1A) was altered to take into account the publication of the solution NMR structure for aa 4-23 (Jirasko *et al.*, 2008) which was proposed to form the first TMD of NS2. The refined computer model (Figure 4.1B) described three TMDs spanning residues 4-23 (rather than residues 7-27 as defined in the original computer model - Figure 3.3A), 29-49 and 63-86. A novel series of NS2 truncation and deletion constructs were generated as N- and C-terminal fusions to eGFP and as C-terminal FLAG-tagged fusions i.e. eGFP-NS2, NS2-eGFP and NS2-FLAG. NS2 truncations were defined based upon the description of three distinct sections within the *trans*-membrane: aa 1-24 encompassing the first putative TMD; aa 25-50 encompassing the second putative TMD; and aa 51-92 comprising the remainder of the *trans*-membrane domain (Figure 4.2).

These NS2 fusion constructs were generated, without any additional HCV protein sequence, in order to further characterise the role of the *trans*-membrane domain with regards to NS2 membrane-targeting, topology and subcellular distribution. To facilitate direct comparison with the work presented in the previous chapter and the potential for translation of findings into the fully infectious virus system, the constructs were generated using JFH-1 (genotype 2a) coding sequence.

4.2.1 Generation and expression of eGFP and FLAG-tagged NS2 fusions

The NS2 constructs were cloned into the peGFP-N1 vector to generate the NS2-eGFP series. However, when lysates from transfected cells were probed with antibodies to GFP and the NS2 catalytic domain it was found that the GFP and the NS2 proteins were detectable as separate proteins as well as the expected fusion proteins (Appendix 4). Detection of GFP alone suggested that the fusion proteins were either being cleaved or that eGFP was being translated from a second mRNA transcript. To eliminate the chance of secondary initiation of eGFP alone, the vector- encoded initiation site, which enables the use of this plasmid as an eGFP expression control vector, was mutated from ATG to GCG by QuikChange mutagenesis generating a mutant form of the vector and termed peGFP-N1.QC (QC).

To confirm the mutation at the protein level the parental (wt) and mutant (QC) peGFP-N1 vectors were transfected into Huh7 cells and analysed for exogenous protein expression and direct eGFP fluorescence 48 hours post-transfection. Confocal IF microscopy of cells transfected with wt and QC vectors and stained with anti GFP antibody demonstrated that the

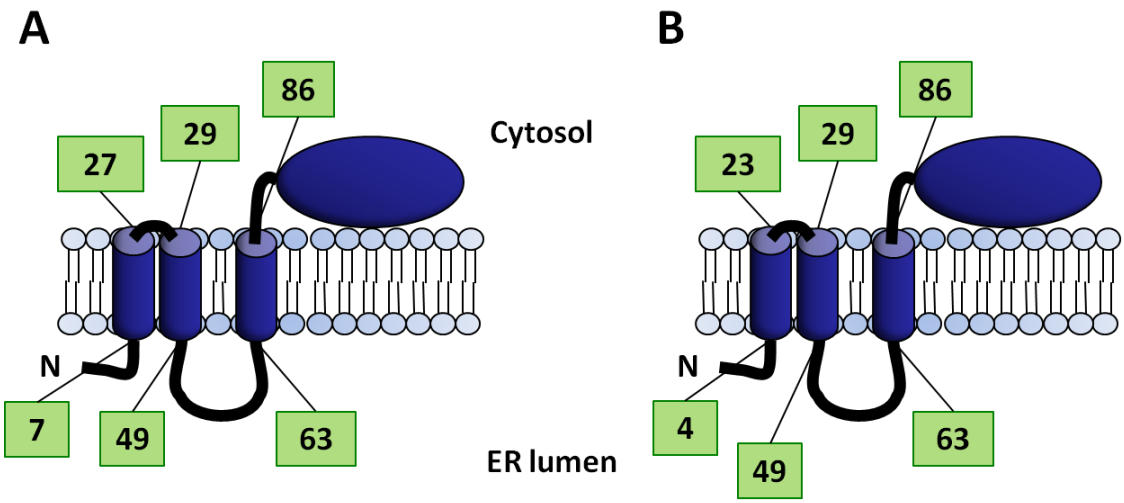


Figure 4.1 Refined computer model of NS2 membrane topology

(**A**) The original computer model described in Figure 3.3. (**B**) A refined computer taking into account the description of the first putative TMD of NS2 by solution NMR to aa 4-23 (Jirasko *et al.*, 2008). Residues marking the predicted TMD boundaries are noted (green boxes). Note the altered prediction of the first TMD from aa 7-27 to aa 4-23.

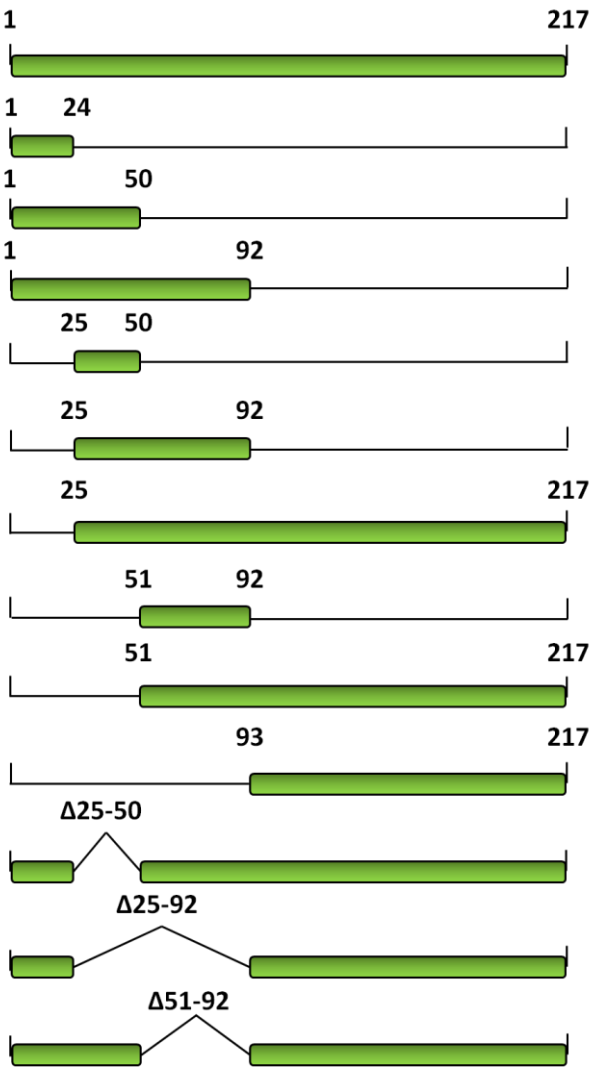


Figure 4.2 Cartoon of NS2 peptides generated as fusions to eGFP or with a C-terminal FLAG-tag.

A series of N-terminal and C-terminal truncations and deletion peptides were generated. The Residues encompassed by each peptide are noted at either end except for the three deletion peptides where the residues deleted are noted.

QC vector produced a protein that lacked direct fluorescence but reacted with the anti-GFP antibody (Figure 4.3A). Analysis of the wt and QC vector by western blot revealed that the mutation did not prevent the expression of a GFP-specific species migrating at the rate as mature eGFP (Figure 4.3B). A second GFP-specific protein species migrating close to the 17 kDa marker was also detected in cells transfected with the QC form of the vector. The next ATG in-frame with the eGFP ORF is approximately 240 nt downstream. Translation initiated from this ATG would produce a severely truncated eGFP molecule with an estimated mass of 18. 5 kDa (wt eGFP is 26.9 kDa). This could explain the 17 kDa protein species noted in peGFP-N1.QC expressing cells. As a protein species of comparable size to mature eGFP was detected in cells expressing the mutant vector it is possible that translation in-frame with the eGFP ORF was being initiated from a non-ATG codon proximal to the native eGFP initiation codon as no examples of GCG initiated translation could be found. It has been shown that the N-terminal residues of GFP are vital for its direct fluorescence, to the extent that deletion of aa 2-8 abrogated the fluorescent activity of the protein (Dopf & Horiagon, 1996) it reasonable to assume any non-native N-terminal region of eGFP would not possess fluorescence activity. The NS2-eGFP constructs generated in the peGFP-N1.QC vector were carried forward as aberrant initiation products could be differentiated from fusion molecules by SDS PAGE and the lack of fluorescence from these aberrant products would enable confocal fluorescence studies of the fusions.

The majority of N-terminal fusions of NS2 to eGFP expressed to low levels or were undetectable in whole cell lysates from multiple, repeated transfections. Therefore, lysates from transfected cells were immunoprecipitated prior to SDS PAGE analysis using GFP-Trap beads to confirm expression (Figure 4.3C). However, even under these circumstances (51-217) eGFP expression was undetected. Despite mutation of the vector encoded eGFP start site; all of the NS2-eGFP fusions, with the exception of (1-92) eGFP, expressed a protein which reacted with the GFP antibody and migrated at a similar rate to eGFP alone. As discussed in section 3.2.7, it is possible that those fusions that contained the catalytic domain may have been capable of cleaving the junction between the NS2 sequence and C-terminal reporter. However, this GFP-specific species was present in all other fusions it is more likely that it is the result of cleavage of the fusions by a cellular protease.

A linker comprising the FLAG epitope followed by a stop codon (FLAG-STOP) was cloned into the NS2-eGFP following the NS2 sequence to generate the NS2-FLAG constructs. Expression of the NS2-FLAG fusions in Huh7 cells produced a further pattern of protein expression/stabilities

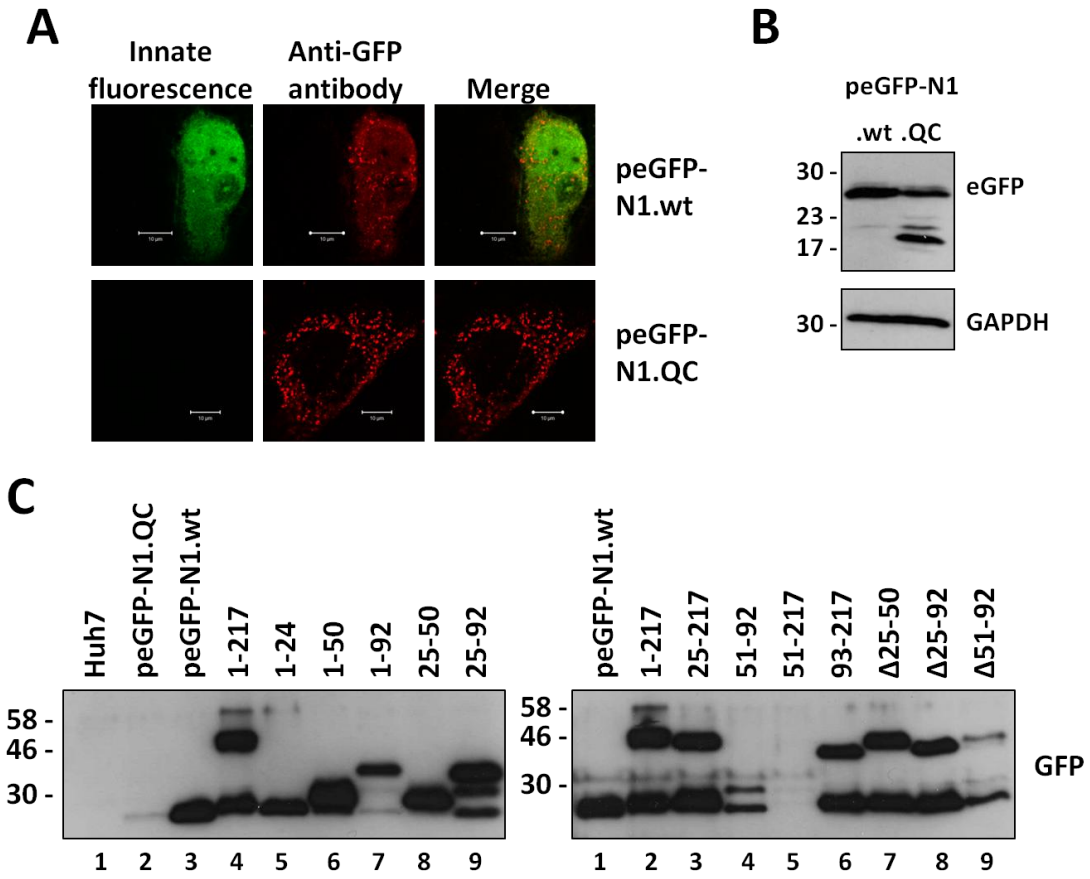


Figure 4.3 Expression of NS2-eGFP fusions

(**A**) Huh7 cells transfected with the wild type peGFP-N1 vector (.wt) or the QuikChange mutant (.QC) were assessed by confocal IF microscopy and (**B**) immuno-blotting. For IF cells were fixed and permeabilised 48 hours post-transfection and eGFP expression was determined directly by observing direct fluorescence (green) or by indirectly using an anti-GFP antibody (red). Representative images are shown. Scale bar - 10nm. For immuno-blotting cells were harvested 48 h post-transfection, analysed by SDS PAGE and transferred to PVDF membranes. Membranes were probed with a polyclonal GFP antibody and a GAPDH antibody. (**C**) Huh7 cells were transfected with the NS2-eGFP fusions. 48 hours post-transfection cells were lysed in 1% triton and immunoprecipitated using the GFP-Trap system prior to analysis by SDS PAGE and subsequent immunoblot. Molecular weight markers are shown to the left of each immunoblot panel. Molecular weight markers are shown to left and lane numbers are given below.

(Figure 4.4A). All peptides could be detected to some degree except (1-24) FLAG, which was never detected following multiple, repeated transfections. Similarly consistent was the detection of a second protein species (tNS2-FLAG – described in section 3.2.7) using the anti-FLAG antibody in lysates of cells transfected with the full-length NS2-FLAG fusion which migrated at the predicted molecular weight of (25-217) FLAG. Both (51-217) FLAG and (25-217) FLAG migrated as single protein species suggesting that they are not susceptible to cleavage. This could be due to lacking the cleavage site as a result of the truncation or different folding of the cleavage site.

C-terminal fusions of NS2 to eGFP were generated in the peGFP-C2 vector and typically expressed to detectable levels, with the exception of (25-92) and (Δ51-92) (Figure 4.4B). Thethree smallest eGFP- NS2 fusions (1-24), (25-50) and (51-92) were expressed to markedly higher levels than the other fusions.

4.2.2 Membrane-targeting of NS2 truncation fusions by membrane fractionation

In vitro studies of NS2 (described in Chapter 3) identified the presence of two distinct membrane-targeting domains within residues 30-74 and 119-151 (Yamaga & Ou, 2002). To confirm these findings in a cell-based system the membrane targeting of NS2 truncations were assessed by membrane fractionation. By analysing all three fusion sets any phenotype observed for a given peptide of NS2 was more likely to be a result of the specific NS2 sequence involved. Huh7 cells were harvested and mechanically lysed in hypotonic buffer 48 hours post-transfection. Post-nuclear supernatants were subjected to centrifugation at 100,000 x *g* for 1 hour (Figure 4.5). Membrane-associated proteins were pelleted under these conditions (**M**) whereas soluble proteins remained in the supernatant (**S**). The endogenous integral membrane protein intercellular adhesion molecule I (ICAM-I) and the soluble protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from mock transfected Huh7 cells were found exclusively in the membrane and soluble fractions respectively (Figure 4.5 right hand side western blot panels). The fractionation of GAPDH exclusively to the soluble fraction demonstrates that this method of homogenisation did not result in the formation of microsomes that could engulf soluble cytosolic protein leading to mis-interpretation of their membrane association.

4.2.2.1 Membrane-targeting: NS2-eGFP fusions

Full-length NS2-eGFP associated entirely with membranes. However, the truncated form of the fusion, tNS2-eGFP, was predominantly soluble (Figure 4.6). When the *trans*-membrane domain (aa 1-92) and the catalytic domain (aa 93-217) were analysed separately the *trans*-membrane

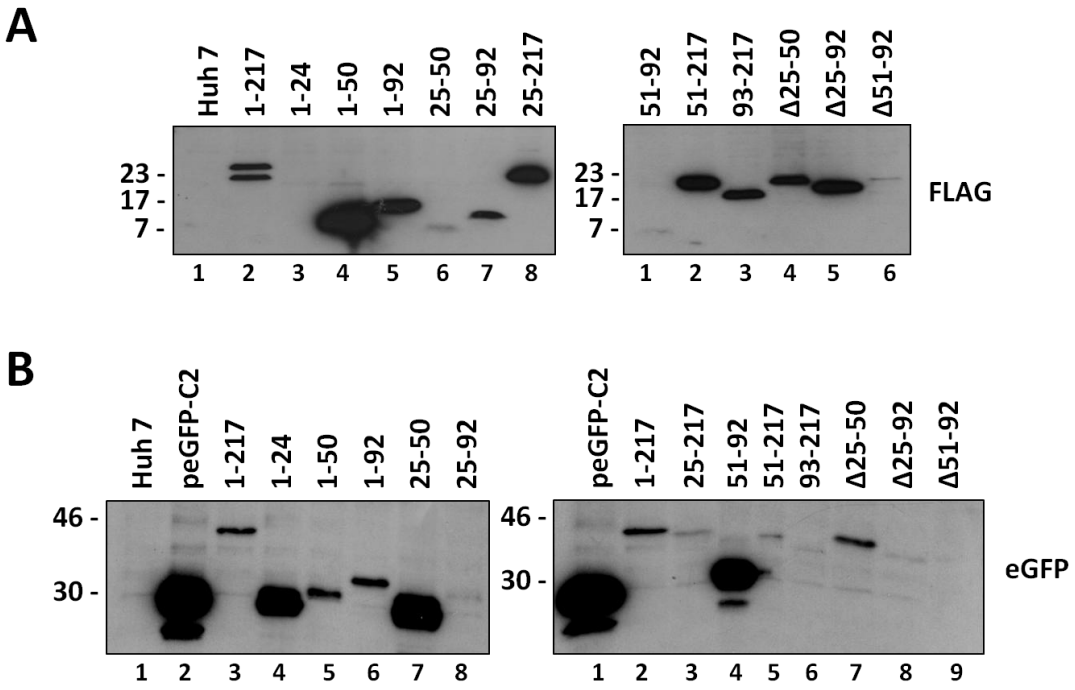


Figure 4.4 Expression of NS2 peptides with C-terminal eGFP or FLAG fusions

(**A**) Huh7 cells were transfected with either NS2-FLAG truncation panel or (**B**) the eGFP-NS2 panel. 48 hours post-transfection lysates were collected and fusion peptide expression was detected by probing for the appropriate fusion reporter following SDS PAGE analysis. Molecular weight markers are shown to the left, the epitope probed for is noted to the right and lane numbers are provided below each panel.

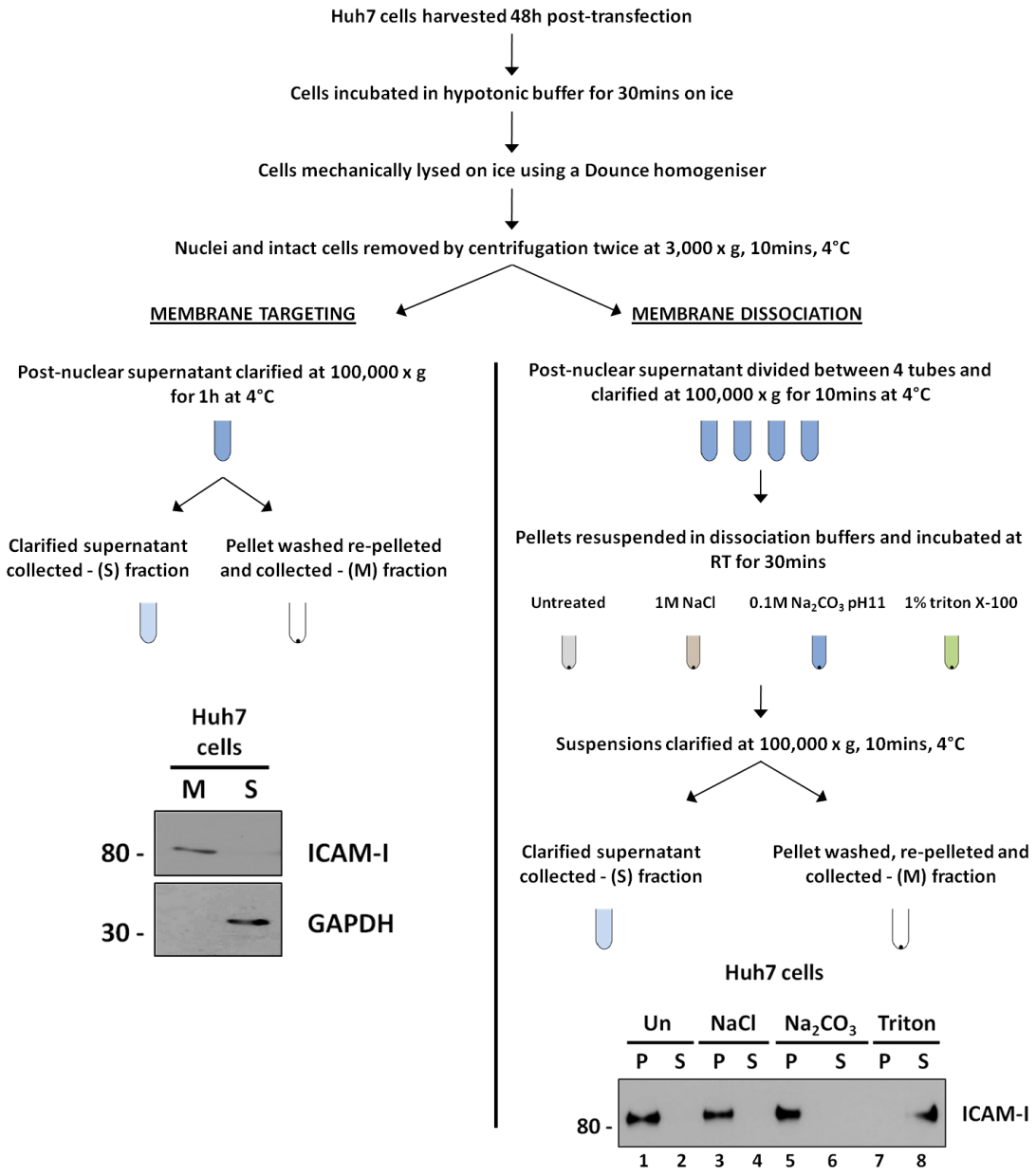


Figure 4.5 Flow diagram outlining the membrane targeting and dissociation procedures.

The membrane targeting (Left) and membrane dissociation (Right) experiments are described in brief. Control blots for each procedure using mock transfected Huh7 cells are shown below each procedure. Membrane targeting experiments: (M) membrane fraction (S) soluble fraction. Membrane Dissociation experiments: (P) pelleted fraction (S) soluble fraction. Colouring is used as a visual aid. Molecular weight markers are shown to the left of each panel and lane numbers are given below.

domain was not expressed to sufficient levels to permit detection, whereas as majority of the catalytic domain was detected in the soluble fraction. Dissection of the *trans*-membrane domain showed that peptides (1-50) and (25-92) were predominantly or entirely soluble. This pattern was mirrored when the individual putative TMDs were expressed individually: the majority of (1-24) was not entirely membrane associated although a modest portion pelleted with membranes, (25-50) was entirely soluble and (51-92) was mostly soluble. Interpretation of the latter's fractionation profile was hampered by a non-specific protein species at similar size to the fusion protein. Due to a lack of detectable protein expression in these experiments, it was not possible to determine the effect of aa 51-92 on membrane targeting of the catalytic domains as (51-217) and (Δ51-92) fusion peptides were. Interestingly (25-217) displayed the same phenotype as the full-length fusion (1-217) whereas (Δ25-50) showed disrupted membrane targeting comparable to the catalytic domain alone (93-217), yet the more severe deletion peptide (Δ25-92) was recovered in the membrane fraction (Figure 4.6).

The unexpected phenotypes of (Δ25-50) and (Δ25-92) could be explained by their predicted topologies. As residues 25-50 and 51-92 were predicted to encompass the second and third TMDs, it is possible that the deletion of a single TMD from (Δ25-50) eGFP resulted in incorrect orientation of the catalytic domain to the ER lumen, which in turn acted to destabilise the membrane interactions. Whereas the deletion of two TMDs from the (Δ25-92) eGFP construct produced the correct orientation of the C-terminus about the ER membrane.

The ability of (1-24) and (25-217) to partially target to membranes indicated the presence of two distinct membrane-targeting sequences within NS2 (Figure 4.6).

4.2.2.2 Membrane-targeting: eGFP-NS2 fusions

Full-length eGFP-NS2 was detected as a single protein species that associated with membranes (Figure 4.7). In contrast with the NS2-eGFP fusions; eGFP (1-92) was detectable and found in the membrane fraction. Fractionation of eGFP (93-217) was inconsistent. While other fusions gave reproducible phenotypes; of the four times eGFP (93-217) was examined it was recovered in either one or the other fractions or not detected at all. It was not clear why this inconsistency was observed, however because of this it was omitted from the analysis. Both (25-217) and (51-217) were able to associate completely with membranes. Further truncation analysis of the *trans*-membrane domain revealed that (1-24) and (25-50) may be involved in membrane targeting as individually they both evenly distributed between the soluble and membrane fractions while as a single polypeptide (1-50) they were fully able to target to membranes. (51-92) was also evenly distributed when expressed in isolation but was not able

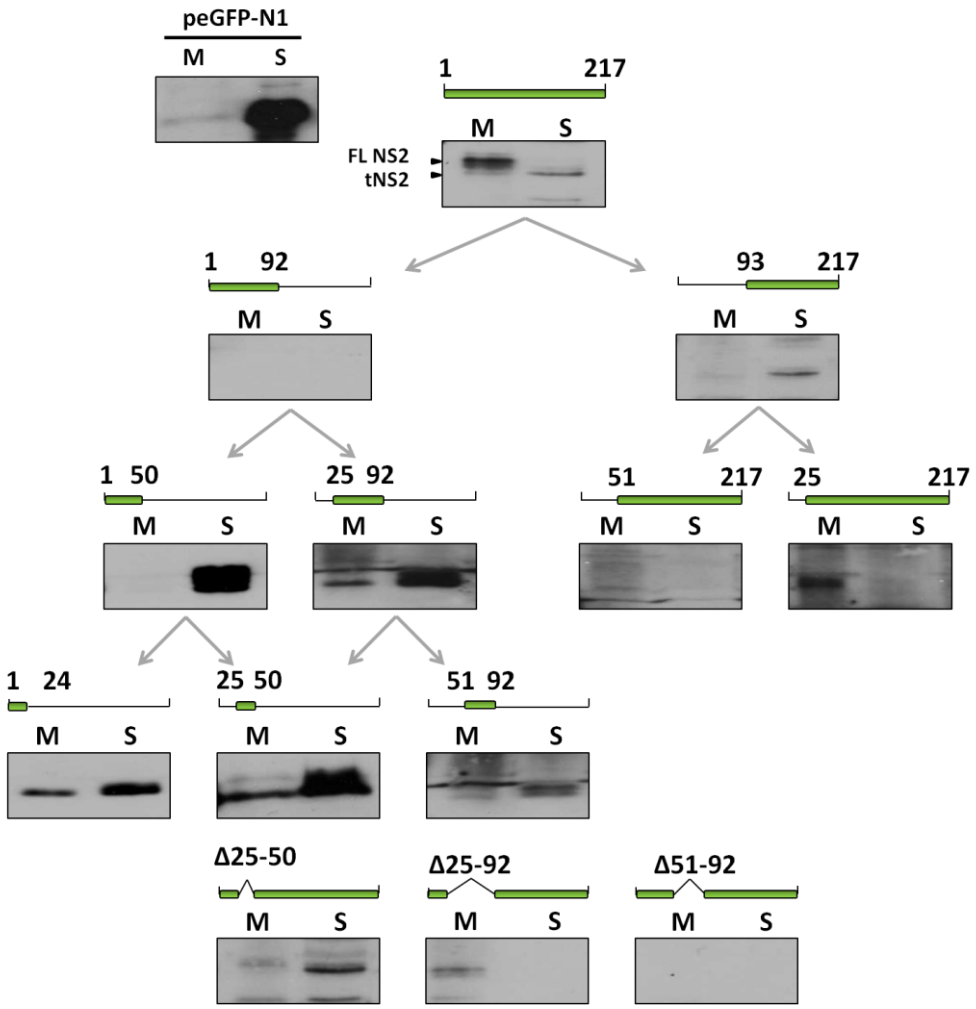


Figure 4.6 Membrane targeting of the NS2-eGFP fusions.

Truncations of NS2 were expressed in Huh7 cells as C-terminal fusions to eGFP and analysed for their abilities to target to membranes. Those fusions able to target to membranes were found in the pellet fraction (M) following centrifugation of lysates at 100,000 x *g*. Fusions unable to target to membranes were recovered in the supernatant (S). Fusions were detected by immuno-blotting for the eGFP fusion. The reporter protein expressed in isolation is shown in the top left panel and full-length NS2-eGFP is shown in the middle at the top. Two distinct protein species were seen for the full-length fusion: full-length NS2 (FL NS2) and an N-terminal truncation (tNS2) (black arrow heads). Two protein species were detected for (25-50); (25-50) eGFP (white arrow head) and 'free' eGFP (black arrow head). The NS2 sequence analysed in each blot panel is shown above.

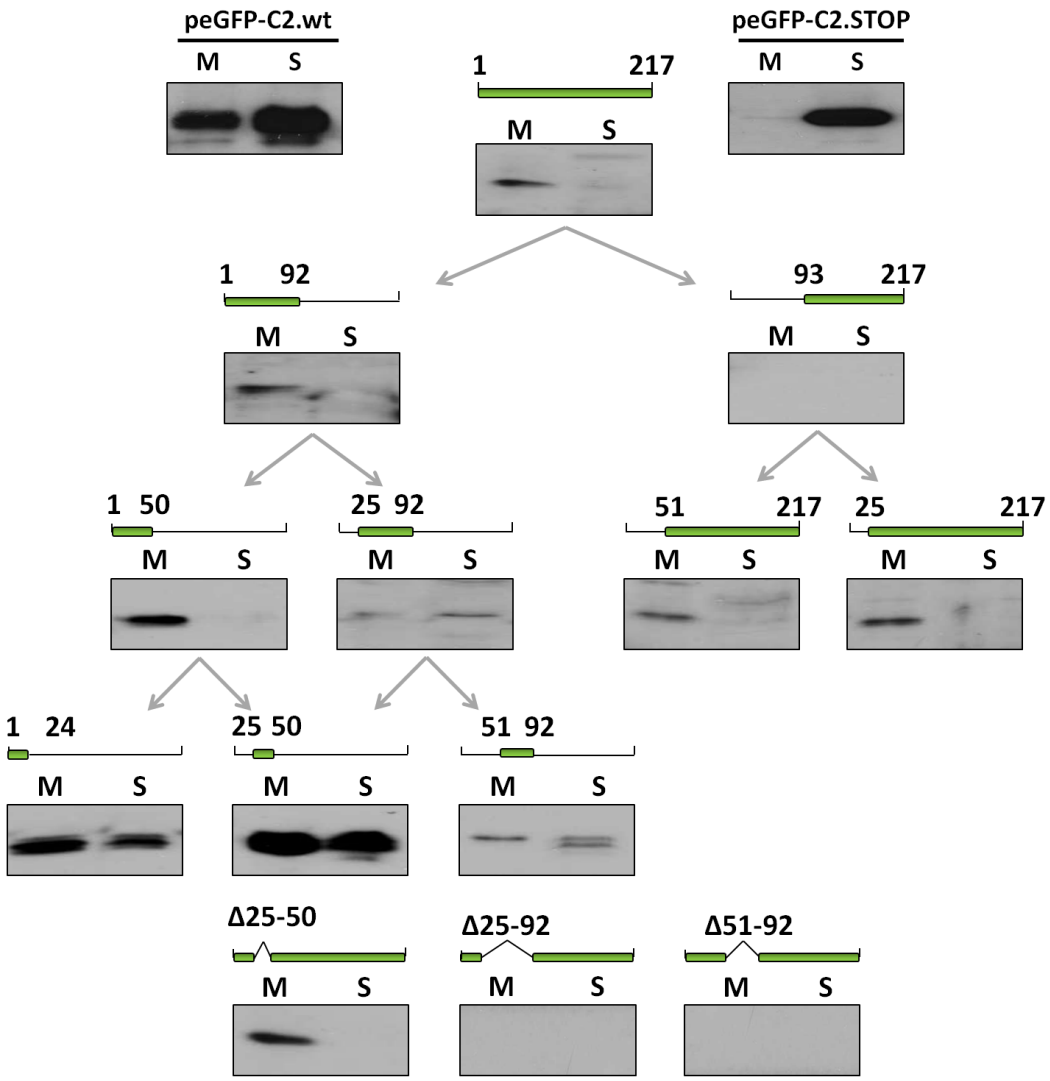


Figure 4.7 Membrane targeting of the eGFP-NS2 fusions.

Truncations of NS2 were expressed in Huh7 cells as N-terminal fusions to eGFP and analysed for their abilities to target to membranes. Those fusions able to target to membranes were found in the pellet fraction (M) following centrifugation of lysates at 100,000 x *g*. Fusions unable to target to membranes were recovered in the supernatant (S). Fusions were detected by immuno-blotting for the eGFP fusion. The membrane targeting of the reporter protein expressed from the wild type vector (peGFP.C2.wt) is shown in the top left panel. The phenotype of the reporter protein following introduction of a STOP codon at the site of NS2 sequence insertion is shown in the top right panel. Full-length NS2-eGFP is shown in the middle at the top. The NS2 sequence analysed in each blot panel is shown above.

to influence the targeting of (25-50) to membranes as (25-92) was similarly evenly distributed. A second slightly faster migrating protein species was observed in the soluble fraction of (51-92); this second protein species was not consistently observed but probably corresponded to 'free' eGFP. The only deletion peptide to be detected from this panel was (Δ25-50) which fully associated with membranes (Figure 4.7). This was in stark contrast to the equivalent C-terminally eGFP-tagged peptide ((Δ25-50 eGFP) - Figure 4.6), which was predominantly soluble.

The ability of eGFP (1-50) and eGFP (51-217) to both completely target to membranes was evidence of two distinct membrane-targeting sequences within NS2. Evidence in support of three membrane-targeting sequences was given by the partial targeting of (1-24), (25-50) and (51-92) which were recovered to a similar degree in both fractions.

4.2.2.3 Membrane-targeting: NS2-FLAG fusions

The NS2-FLAG fusions were similarly analysed for their abilities to target to membrane (Figure 4.8). As noted previously (Figure 4.4A) two FLAG-specific species were detected in lysates from cells transfected with the full-length NS2-FLAG peptides. The top protein species, corresponding to full-length NS2-FLAG, pelleted in the membrane fraction while the lower species, corresponding to the truncated form (tNS2), was more evenly distributed between the two phases (Figure 4.8). As with the previous fusion sets; specific truncations or deletions of NS2 with C-terminal FLAG-tags were assessed for their ability to associate with membranes.

The *trans*-membrane domain (1-92) was found exclusively in the membrane fraction while the catalytic domain (93-217), although mostly detected in the membrane fraction, was also present in the soluble fraction. This pattern was mirrored by the N-terminal truncations (51-217) and (25-217) and the deletion peptides (Δ25-50) and (Δ25-92). Of those constructs encoding the catalytic domain only (Δ51-92) showed significantly impaired membrane targeting. Examination of the *trans*-membrane domain revealed that (1-50) and (25-50) were entirely membrane associated whereas (25-92) and (51-92) were evenly distributed between the two fractions. Despite being able to independently target to membranes (25-50) was not able to influence the targeting of (51-92), as seen by the similar targeting of (25-92) and (51-92). Consistent with the initial expression blots, (1-24) FLAG was not detectable in the membrane-targeting experiments (Figure 4.8).

The detection of both (1-50) FLAG and (51-217) FLAG in the membrane fraction demonstrated that each of these domains contains a distinct membrane-targeting sequence (Figure 4.8). Interestingly, dissection of these peptides revealed the potential for three membrane-

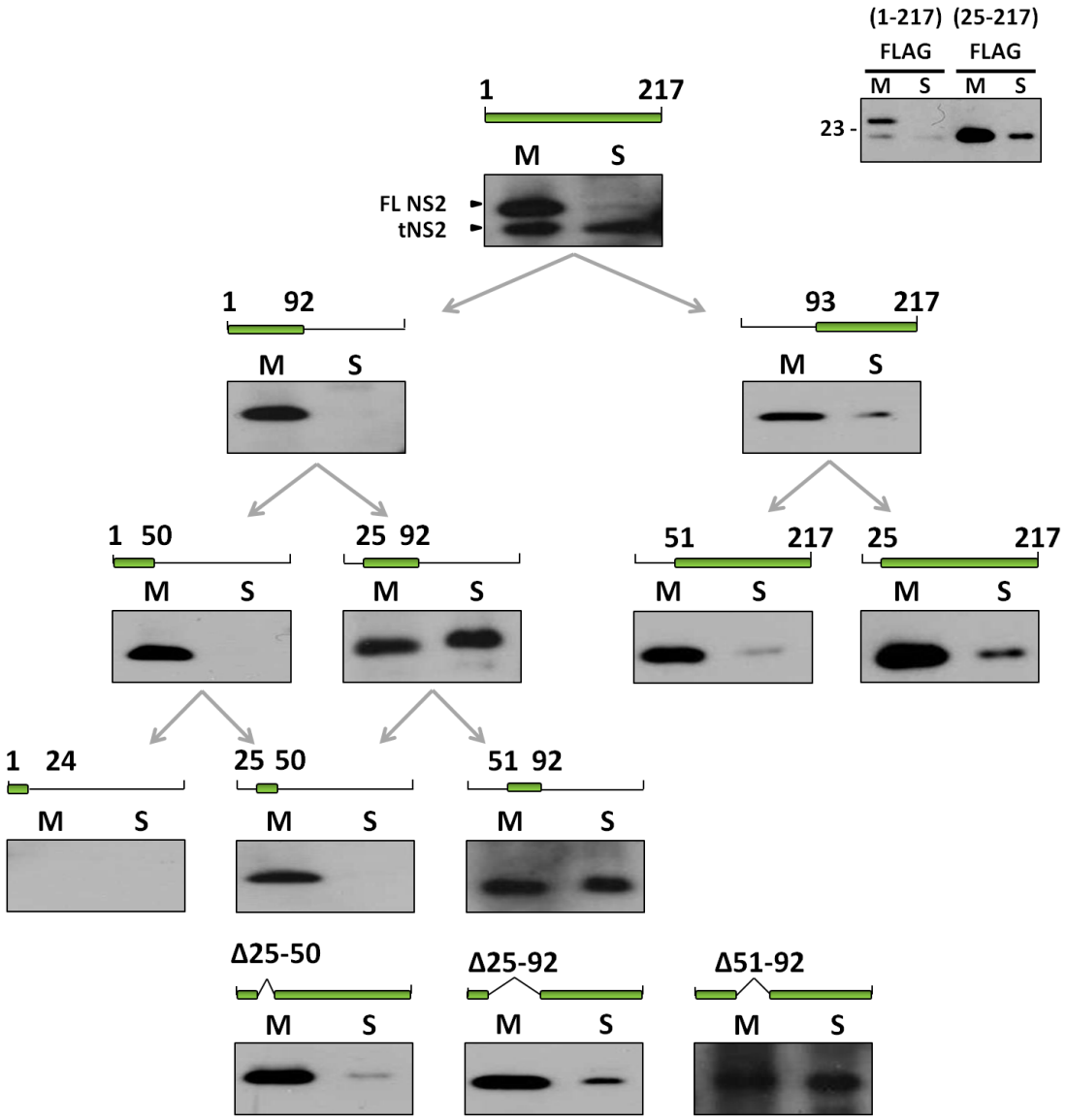


Figure 4.8 Membrane targeting of the NS2-FLAG panel.

Truncations of NS2 were expressed in Huh7 cells with C-terminal FLAG-tags and analysed for their abilities to target to membranes. Those peptides able to target to membranes were found in the pellet fraction (M) following centrifugation of lysates at 100,000 x *g*. Peptides unable to target to membranes were recovered in the supernatant (S). Peptides were detected by immuno-blotting for the FLAG epitope. Two distinct protein species were seen for the full-length fusion: full-length NS2 (FL NS2) and an N-terminal truncation (tNS2) (black arrow heads). The NS2 sequence analysed in each blot panel is shown above.

targeting sequences within NS2 as (25-50) FLAG and (93-217) FLAG could target efficiently to membranes and a significant proportion of (51-92) FLAG was also recovered in the membrane fraction.

4.2.3 Analysis of subcellular localisation by confocal IF microscopy

Membrane targeting analysis suggested the presence of multiple signal sequences within NS2. In order to discern whether these sequences were also responsible for the sub-cellular localisation of NS2 Huh7 cells expressing the NS2 fusion sets were analysed 48h post-transfection by confocal IF microscopy.

4.2.3.1 Subcellular localisation: NS2-eGFP fusions

The subcellular localisation of the NS2-eGFP fusions was assessed (Figure 4.9). Some fusions showed no direct eGFP fluorescence. Fusion expression from these constructs was determined indirectly by staining cells with an anti-GFP antibody and imaged in red. Those with direct eGFP fluorescence were imaged in green. GFP expressed from the wild type vector showed typical staining (Figure 4.9 **-** top right frame) whilst a GFP-antibody-reactive species was expressed from the mutant vector that lacked direct fluorescence and had a different sub-cellular distribution to the wt protein (Figure 4.9 **-** top left frame). Full-length NS2-eGFP (1-217) showed a diffuse, non-nuclear fluorescence pattern, reminiscent or ER staining. When the two domains of NS2 were assessed in isolation (1-92) gave a similar pattern of eGFP fluorescence to the full-length NS2 fusion. This was in contrast to (93-217) which showed a dramatically disrupted staining pattern. The less severe N-terminal truncation (51-217) also showed disrupted staining and was only detectable using the anti-GFP antibody. Subcellular fluorescence similar to the full-length fusion was achieved when only the first TMD of NS2 was lacking in the case of the (25-217) fusion (Figure 4.9). Of the deletion mutants (Δ25-50) showed staining comparable to full-length NS2 while (Δ25-92) and (Δ51-92) showed a mixed phenotype of significant diffuse distribution combined with bright punctae, but not the extent of aggregation seen for (93-217). Dissection of the *trans*-membrane domain showed that each of the three sub-domains expressed in isolation disrupted the direct fluorescence of the eGFP reporter molecule. Expression of two of the putative sub-domains in tandem, either as (1-50) or (25-92) resulted in either wild-type staining or the formation of aggregates similar to those seen for (93-217) respectively (Figure 4.9).

No interpretation was made of those fusions lacking direct eGFP fluorescence as this was possibly due to mis-folding of the reporter protein which could result in fusion peptides being

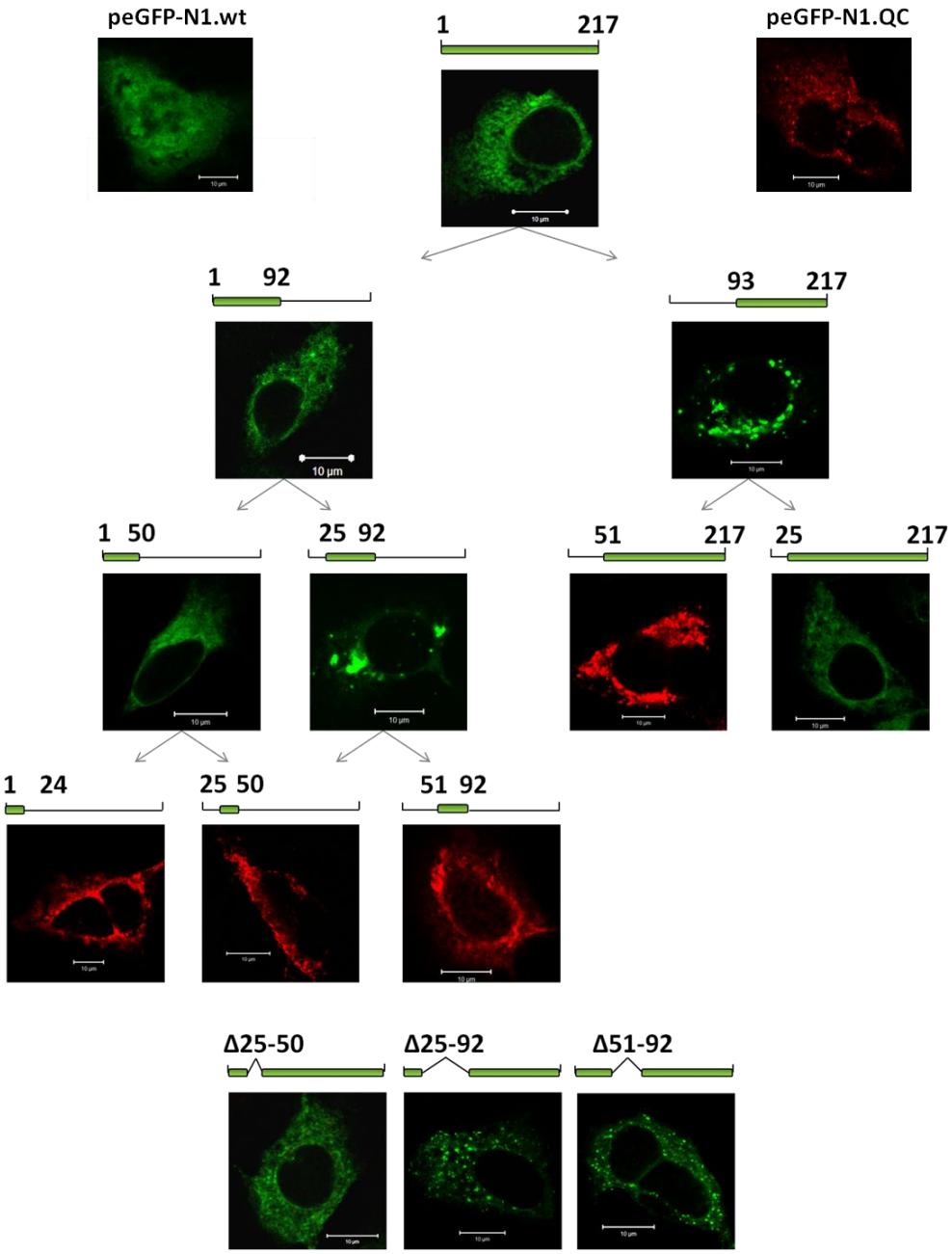


Figure 4.9 Subcellular localisation of the NS2-eGFP fusions.

The NS2-eGFP fusions were expressed in Huh7 cells and fixed 48 hours post-transfection and visualised for eGFP fluorescence (green). Expression of those fusions that displayed no direct reporter fluorescence was confirmed by permeabilisation of fixed cells and detection of the fusions by indirect IF using an anti-GFP antibody (red). eGFP expressed from the wild type peGFP-N1 (wt) vector is shown in the top left panel. The vector into which the NS2 truncations were cloned (peGFP-N1.QC) which contains a mutation within the vector encoded eGFP initiation codon is shown in the top right. Cells transfected with this mutant empty vector displayed no direct eGFP fluorescence. Full-length NS2-eGFP is shown in the middle at the top. Representative images are shown.

targeted to protein recycling pathways. Thus any apparent subcellular localisation of these fusions may not have been determined solely by the NS2 sequence involved.

Analysis of those constructs that displayed direct eGFP fluorescence suggested the *trans*-membrane domain as the domain responsible for NS2 sub-cellular localisation as (1-92) eGFP displayed a wt pattern, whereas the localisation of (93-217) eGFP was disrupted compared with the full-length fusion (Figure 4.9). Furthermore, the putative first and second TMDs may be responsible for NS2 localisation as (1-50) eGFP also displayed localisation comparable to wt.

4.2.3.2 Subcellular localisation: eGFP-NS2 fusions

eGFP expressed from the wt and mutant STOP peGFP-C2 vectors were diffuse and displayed characteristic nuclear staining. As described earlier, the mutant STOP vector encoded an additional STOP codon, preventing translation of the multiple cloning site (MCS) as a C-terminal extension to the eGFP molecules.

The non-nuclear, diffuse staining observed for full-length NS2-eGFP was corroborated by the fusion of eGFP to the N-terminus of NS2 which produced a similar staining pattern (Figure 4.10 (1-217)). Removal of the catalytic domain (1-92) did not alter the subcellular distribution of the eGFP reporter. However, fusion of eGFP to the catalytic domain alone resulted in the accumulation of eGFP in small cytoplasmic foci. Surprisingly (51-217) and (25-217) both displayed a more exaggerated disruption to full-length NS2 staining than (93-217), as did (Δ51-92). Further investigation of the *trans*-membrane domain revealed that (1-50) was able to target eGFP in a similar pattern to full-length NS2 while (25-92) showed a large degree of fusion peptide aggregation in addition to the more typical diffuse staining. None of the three sub-domains in isolation were able to produce a fluorescence pattern similar to full-length NS2: (1-24) had a pattern more similar to wt eGFP, in that it was present throughout the cell, but also appeared to be aggregating in nuclear or peri-nuclear compartments; (25-50) produced a distinctive punctuate distribution somewhat similar to (Δ51-92); while (51-92) was diffuse throughout the cytoplasm as well as in large aggregates, akin to (25-217). The staining pattern observed for (Δ25-92) was comparable to (93-217) but the less severe deletion (Δ25-50) was like full-length NS2 (Figure 4.10).

Consistent with the NS2-eGFP fusion (Figure 4.9), this data set also indicate that the *trans*-membrane domain mediates NS2 localisation as eGFP (1-92) sub-cellular distribution was indistinguishable from full-length eGFP NS2 (Figure 4.10). eGFP fused to the catalytic domain, eGFP (93-217), showed a partially disrupted pattern of fluorescence compared to the full-

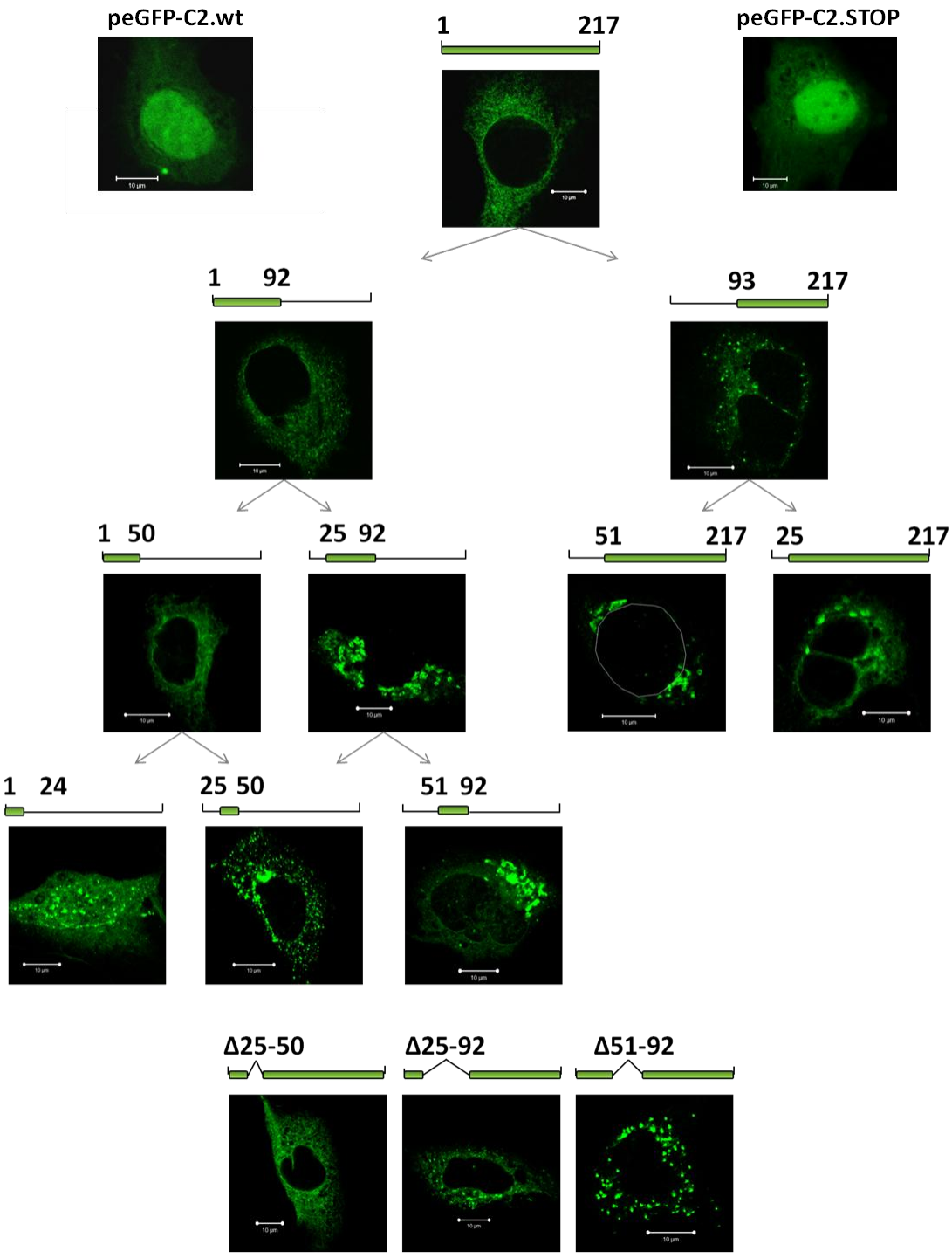


Figure 4.10 Subcellular localisation of the eGFP-NS2 fusions.

The eGFP-NS2 fusions were expressed in Huh7 cells and fixed 48 hours post-transfection and visualised for eGFP fluorescence (green). eGFP expressed from the wild type peGP-C2 (wt) vector is shown in the top left panel. The vector containing a stop site corresponding to the site of NS2 sequence insertion (peGFP-C2.STOP) is shown in the top right. Full-length NS2-eGFP is shown in the middle at the top. Representative images are shown.

length fusion. While (1-50) eGFP also displayed localisation comparable to wt, it maybe that the first putative TMD (aa 1-24) is more important for achieving correct localisation than the second TMD (aa 25-50) as eGFP (25-92) and (25-217) showed disrupted localisation but eGFP (Δ25-50) was comparable to wt (Figure 4.10).

4.2.3.3 Subcellular localisation: NS2-FLAG fusions

The NS2-FLAG fusions were generated in the peGFP-N1.QC vector background which displayed no direct fluorescence (Figure 4.3B). Indirect detection of the FLAG epitope enabled the analysis of the NS2-FLAG fusions. The ability of NS2 to determine its own sub-cellular distribution in spite of fusion to other proteins was confirmed as (1-217) FLAG had a non-nuclear diffuse staining pattern (Figure 4.11), although the staining appeared to be more discrete when directly compared to NS2-eGFP (Figure 4.9) and eGFP-NS2 (Figure 4.10). This discrepancy may be due to the respective fluorophores and techniques used to identify the three fusion sets.

Analysis of the NS2-FLAG fusion revealed that the only FLAG-tagged peptides to show a similar staining pattern to full-length NS2-FLAG were (25-217) and (Δ25-50) (Figure 4.11). (Δ25-92) FLAG was similarly diffuse but lacked the defined perinuclear staining observed in the full-length protein. The only consistent feature of these three fusions was the catalytic domain, which expressed on its own, (93-217 FLAG, showed a disrupted phenotype. Surprisingly (Δ51-92) showed a more disrupted phenotype than the more extensive deletion peptide (Δ25-92) (Figure 4.11).

It was not clear from the NS2-FLAG fusions which domains of NS2 were necessary to confer wild-type NS2 sub-cellular localisation, in the absence of the other viral proteins.

4.2.4 Affinity of the membrane interactions of NS2 truncation fusions

To investigate whether any of the fusions were forming integral or peripheral membrane associations; the membrane affinity of the fusion peptides was determined following treatment of the membrane fraction with isotonic buffer, 1 M NaCl, 100 mM Na_2CO_3 pH 11.5 and 1% Triton X-100 (Figure 4.5). When this procedure was applied to mock-transfected Huh7 cells the endogenous protein intercellular adhesion molecule 1 (ICAM-I) was found in the pellet fraction following treatment of the membranes with isotonic buffer, high salt and alkaline carbonate. ICAM-I was fully recovered in the soluble fraction only following treatment with detergent (Figure 4.5).

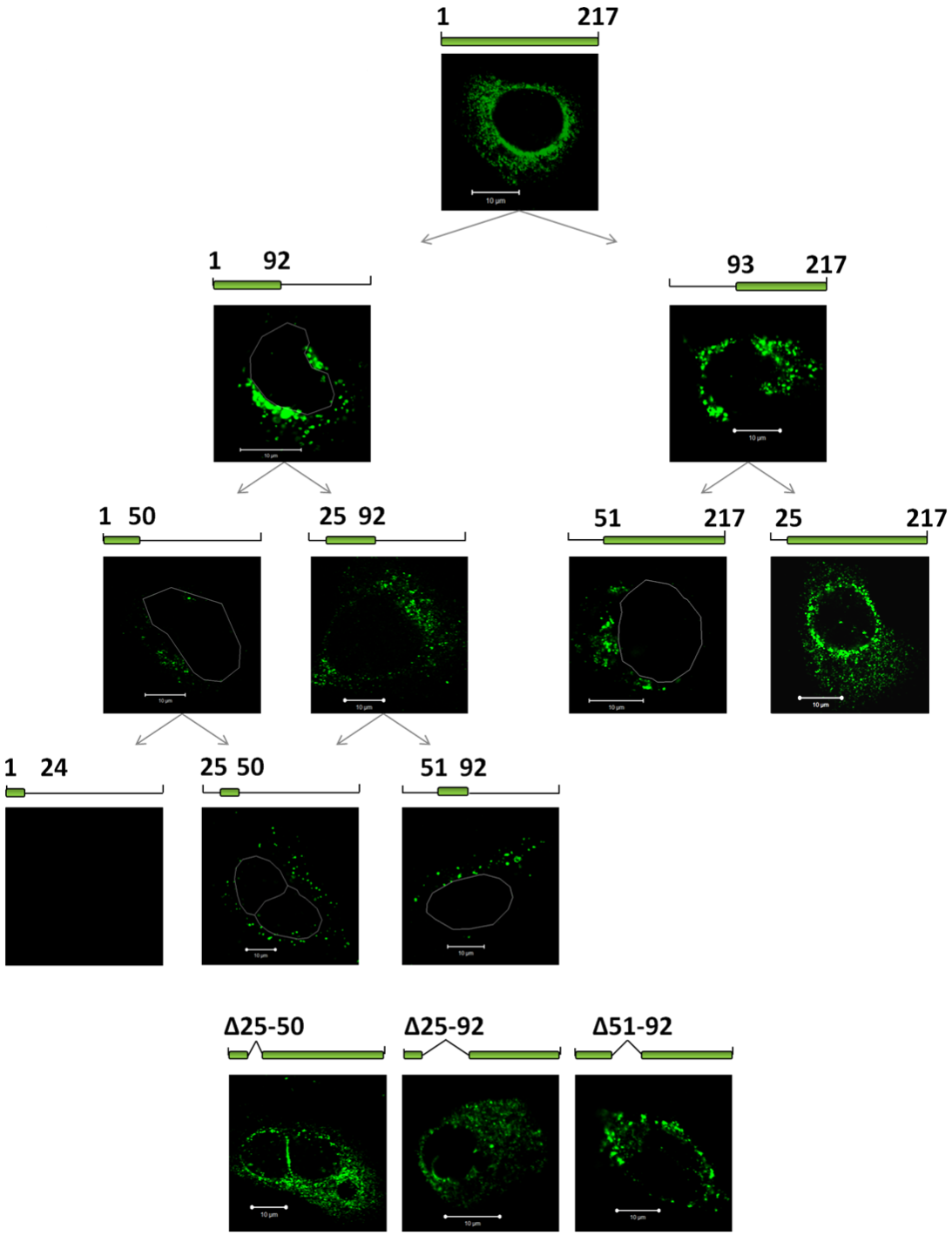


Figure 4.11 Subcellular localisation of the NS2-FLAG fusions.

The NS2-FLAG peptides were expressed in Huh7 cells and fixed 48 hours post-transfection. Cells were permeabilised and tagged NS2 peptides were visualised following indirect immunofluorescence (IF) labelling of the FLAG epitope (green). Where necessary, nuclei are outlined (white rings). Full-length NS2-eGFP is shown in the middle at the top. Representative images are shown.

The eGFP control expression vectors were assessed by this procedure and GFP expressed from the C2.wt and C2.STOP vectors was found predominantly in the supernatant of the initial clarified membrane fractionation such that they could not be further removed from membranes by this technique (Figure 4.12A and B). The protein produced by the peGFP-N1.wt vector was almost entirely soluble however a trace amount did pellet with the initial membrane fraction (Figure 4.12C). This population could be partially removed from membranes was not removed by washing in 1 M NaCl, which strips proteins from membranes, but was extracted by alkaline carbonate, which strips proteins from membranes and linearises membrane vesicles suggesting that a small population of eGFP expressed from this vector was being packaged into vesicles.

These controls demonstrate that the eGFP reporter has a very low affinity for membranes allowing the examination of NS2 sequences expressed as fusions to eGFP. Therefore this method was applied to the eGFP and FLAG-tagged NS2 peptides. Whenever possible the FLAG-tagged versions were shown for simplicity as they represented the least altered versions of the NS2 sequence. Full-length NS2 fusions could only be removed from the membrane fraction with detergent (Figure 4.13A and B) consistent with previous data describing NS2 as an integral membrane protein (Santolini *et al.*, 1995). Interestingly the truncation form of NS2, tNS2, observed in the full-length NS2-FLAG transfected cells showed a partial association with membranes in the membrane-targeting experiments (Figure 4.8), yet that which was associated with membranes was also only solubilised after detergent treatment (Figure 4.13B). The two N-terminal truncations of NS2 within the *trans*-membrane domain (25-217) and (51-217) also appeared to insert into membranes as neither could be stripped from membranes using high salt or alkaline carbonate washes (Figure 4.13C and D respectively). Analysis of the complete *trans*-membrane domain (aa 1-92) showed that it was able to act like an integral membrane polypeptide (Figure 4.14A) as it was only recovered in the soluble fraction after treatment with detergent. As the three catalytic domain fusions showed different propensities towards membranes (Figure 4.6, 4.7 and 4.8 – (93-217)) all versions were tested for their membrane affinity. In the membrane-targeting experiment the catalytic domain expressed as either eGFP (93-217) or (93-217) FLAG targeted predominantly to the membrane fraction (Figure 4.6 and 4.8). When the membrane-associated molecules were treated to disrupt the membrane interactions; washing in isotonic buffer and high molarity salt removed little or no fusion peptides but they were partially solubilised by alkaline carbonate treatment (Figure 4.14B and C). This behaviour is most consistent with that of a soluble protein that has become trapped in microsomes as alkaline carbonate linearises membrane bi-layers, as well as

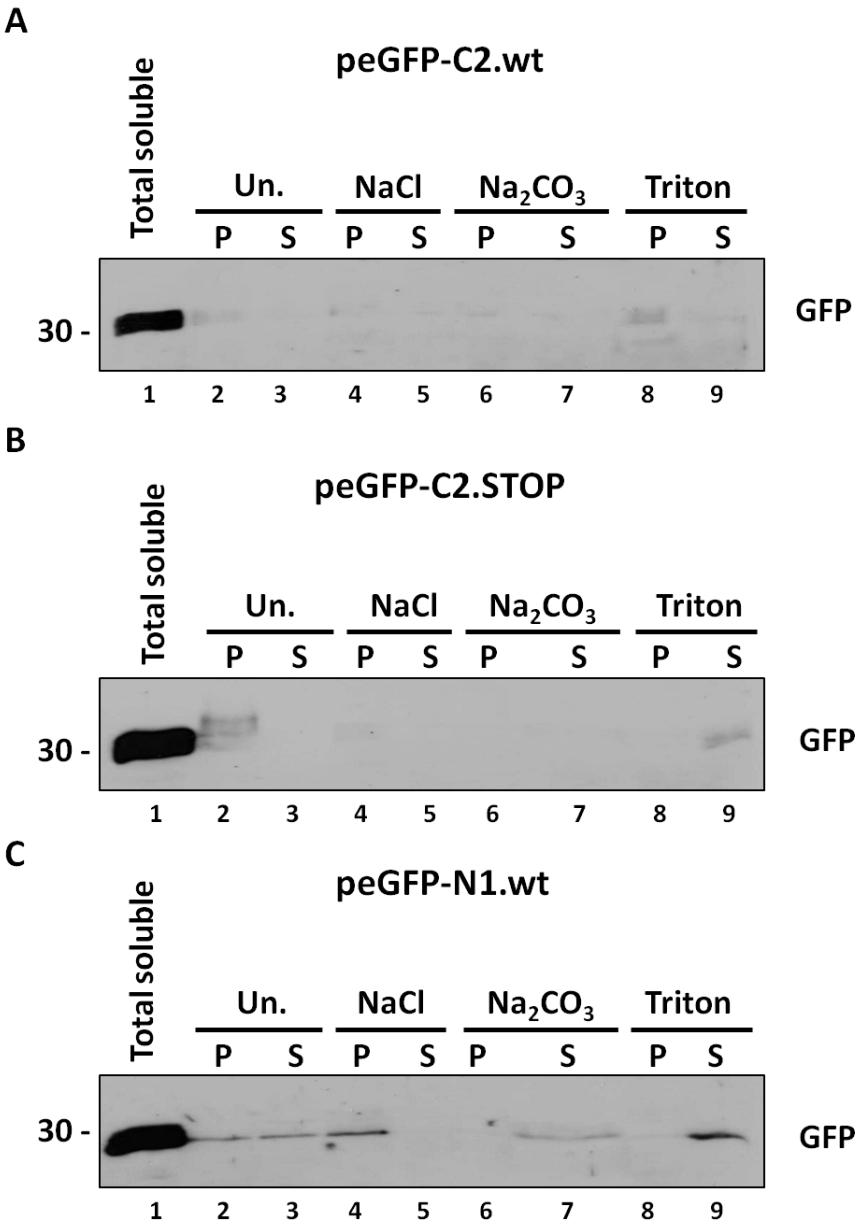


Figure 4.12 Membrane dissociation of eGFP expression control plasmids.

Huh7 cells were transfected with (**A**) peGFP-C2.wt, (B) peGFP-C2.STOP and (**C**) peGFP-N1.wt and the membrane affinity of the expressed eGFP molecules was determined by treatment with chaotropic agents and detergent. Proteins associated with the membrane fraction following centrifugation of the post-nuclear supernatant at 100,000 x *g* for 10 minutes were resuspended in either isotonic buffer (Un.), 1 M NaCl (NaCl), 100 mM Na_2CO_3 pH 11.5 (Na_2CO_3) or 1% triton (Triton) and incubated at 25°C for 30 minutes. Fusions were either recovered in the pellet (P) or soluble (S) fraction following centrifugation of treated membranes at 100,000 x *g*. Molecular weight markers are shown to the left, the fusion tag probed for by immuno-blot is noted to the right and lane numbers are provided below each panel.

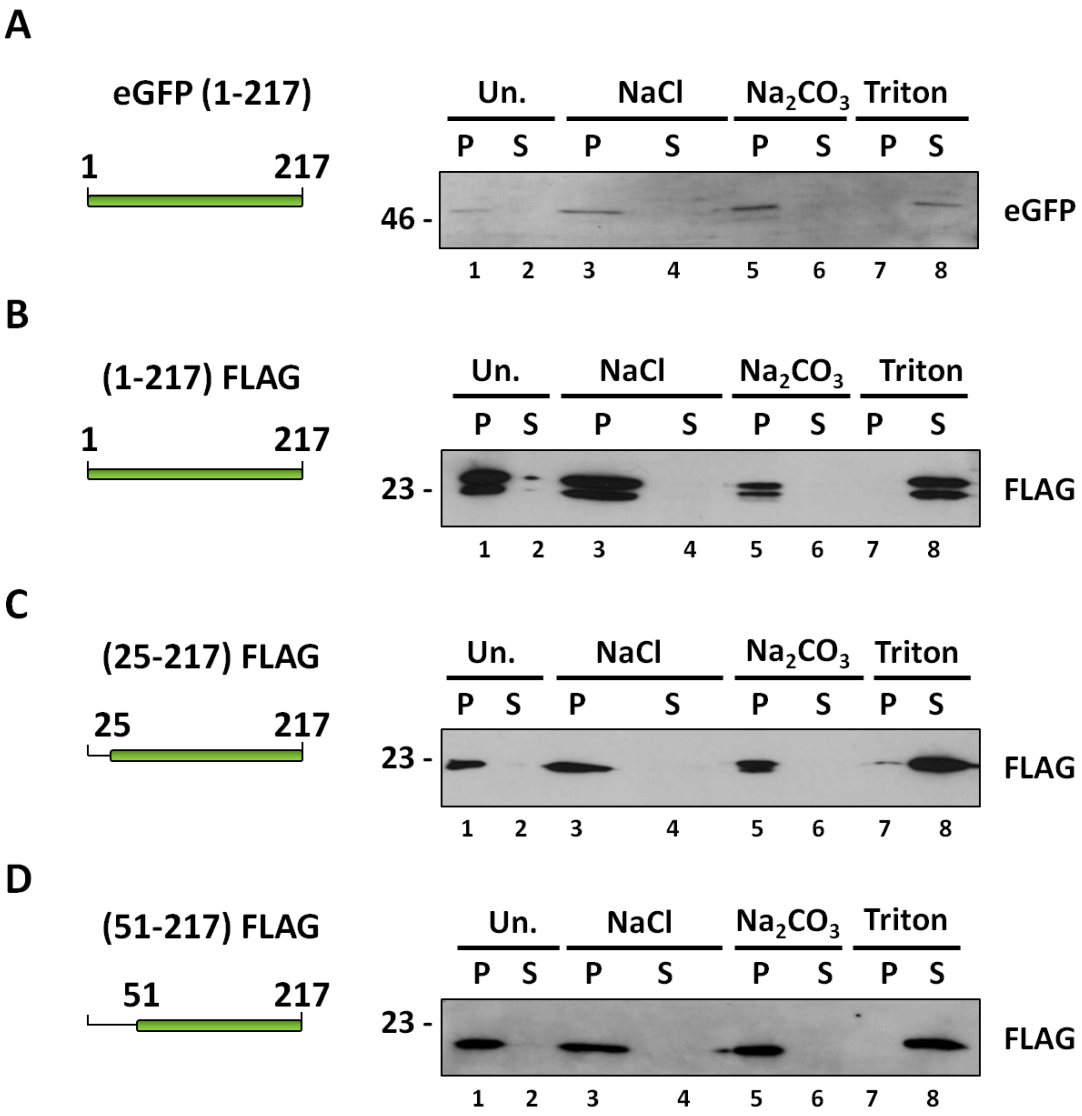


Figure 4.13 Membrane dissociation of full-length NS2 and two N-terminal truncations.

Full length NS2 with either (**A**) an N-terminal eGFP fusion, (**B**) a C-terminal FLAG-tag and (**C**) the FLAG-tagged N-terminal truncations (25-217) and (**D**) (51-217) were assessed for the affinity of their interactions with membranes. Proteins associated with the membrane fraction following centrifugation at 100,000 x *g* for 10 minutes were resuspended in either isotonic buffer (Un.), 1 M NaCl (NaCl), 100 mM Na_2CO_3 pH 11.5 (Na_2CO_3) or 1% triton (Triton) and incubated at 25°C for 30 minutes. Fusions were either recovered in the pellet (P) or soluble (S) fraction following centrifugation of treated membranes at 100,000 x *g*. Molecular weight markers are shown to the left, the fusion tag probed for by immunoblot is noted to the right and the lane numbers are provided below each panel.

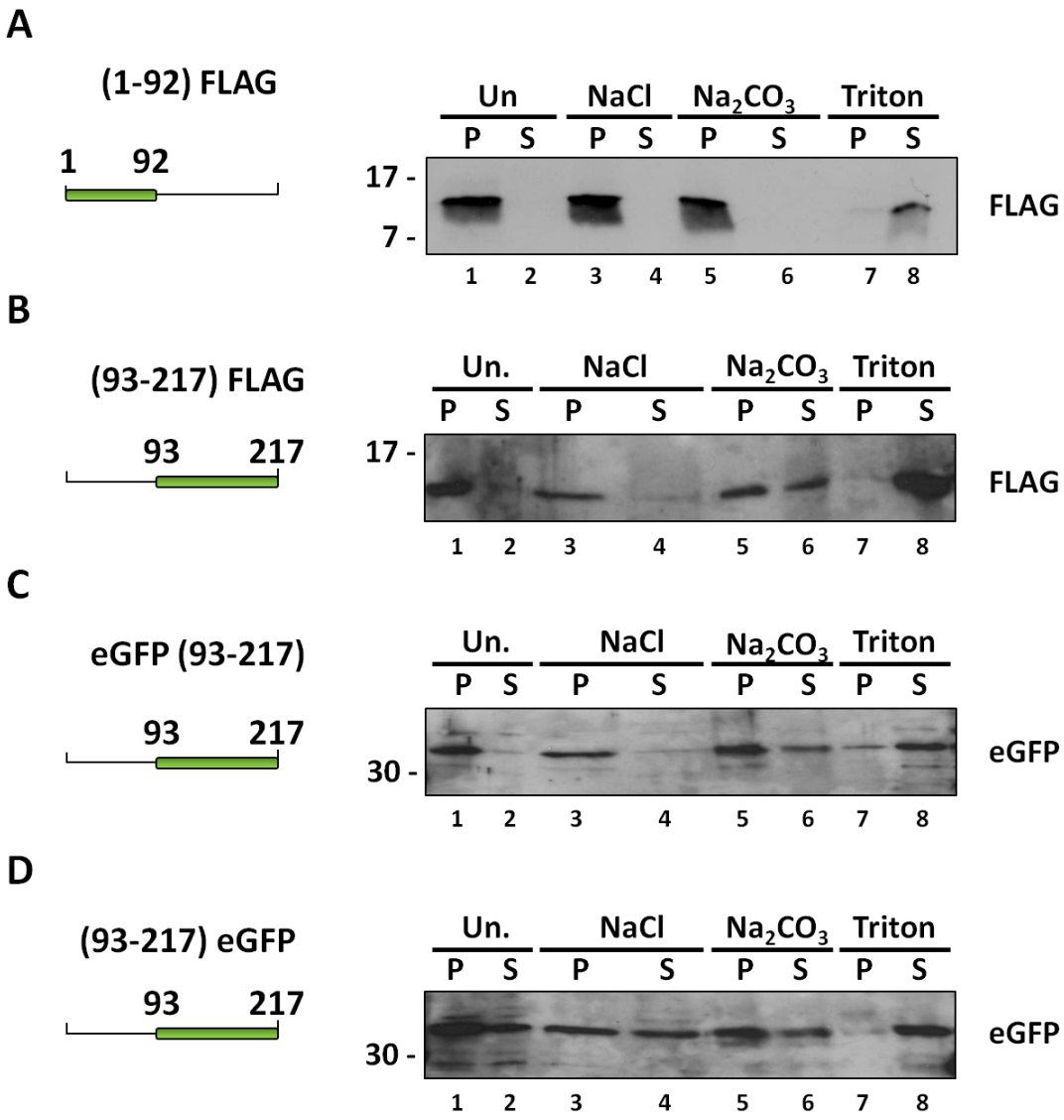


Figure 4.14 Membrane dissociation of the trans-membrane and catalytic domains of NS2.

(**A**) The FLAG-tagged NS2 *trans*-membrane domain and (**B**) the catalytic domain expressed with a C-terminal FLAG-tag, (**C**) an N-terminal eGFP fusion or (**D**) a C-terminal eGFP fusion were assessed for the affinity of their interactions with membranes. Proteins associated with the membrane fraction following centrifugation at 100,000 x *g* for 10 mins were resuspended in either isotonic buffer (Un.), 1 M NaCl (NaCl), 100 mM Na_2CO_3 pH 11.5 (Na_2CO_3) or 1% triton (Triton) and incubated at 25°C for 30 minutes. Fusions were either recovered in the pellet (P) or soluble (S) fraction following centrifugation of treated membranes at 100,000 x *g*. Molecular weight markers are shown to the left and the fusion tag probed for by immunoblot is noted to the right of each panel.

stripping them of protein content. As a similar profile was not observed for the soluble protein control GAPDH (Figure 4.5) this indicates that (93-217) FLAG and eGFP (93-217) may be packaged to vesicles. The incomplete solubilisation after triton treatment seen for eGFP (93-217) was likely the result of experimental error as all peptides were consistently solubilised by detergent treatment. Only a small amount of the (93-217) eGFP fusion was able to target to membranes (Figure 4.7) and this population was readily disrupted by further washing of the membrane pellet in isotonic buffer alone as well as high molarity salt and alkaline carbonate (Figure 4.14D). The fractionations shown in Figure 4.14B, C and D were carried out at the same time. Attempts to repeat these findings, particularly using (93-217) FLAG, were highly inconsistent. In some instances they did not associate with membranes at all, in others they were so poorly expressed as to prevent effective phenotyping and in other instances they were resolutely associated with membranes such that they could only be solubilised following detergent treatment.

While the reason for the inconsistencies was not clear, it is possible that the catalytic domain forms weak or unstable interactions with membranes such that variations in pellet resuspension between experiments or fluctuations in ambient laboratory temperature were responsible for the lack of reproducibility as membrane dissociations incubations were carried at room temperature, rather than on ice, to enhance membrane fluidity and lipid solubility.

Isolated sub-domains of the *trans*-membrane domain were analysed for the affinity of their membrane interactions (Figure 4.15). As the structure of the putative first TMD had been solved by solution NMR (Jirasko *et al.*, 2008), dissociation experiments focused on the putative second and third TMDs encoded by aa 25-50 and 51-92, respectively. Analysis of both putative TMDs together (25-92) showed that, like the complete *trans*-membrane domain, it was able to form a strong interaction with membranes as it was only found in the soluble fraction after triton treatment (Figure 4.15A). In the membrane-targeting experiments eGFP (51-92) and (51-92) FLAG were recovered to equal amounts in both fractions. Despite numerous attempts (51-92) FLAG could not be detected to sufficient levels under these conditions. The membrane-targeted population of the eGFP (51-92) fusion could be detected and displayed a dissociation profile more consistent with that of an integral membrane protein than an amphitropic protein (Figure 4.15B). Residues 25-50 showed a varying affinity for membranes. eGFP (25-50) showed even distribution between the membrane and soluble fractions in the membrane-targeting experiments and the membrane-associated portion could be partially disrupted by washing with high molarity salt or alkaline carbonate (Figure 4.15C). This response following incubation

with agents that exhibit chaotropic properties is characteristic of a peripherally-associated protein. However, when (25-50) was examined with either a C-terminal eGFP fusion or FLAG-tag the membrane targeted population was insensitive to all but detergent treatment (Figure 4.15D and E, respectively) suggesting that it is capable of forming a TMD.

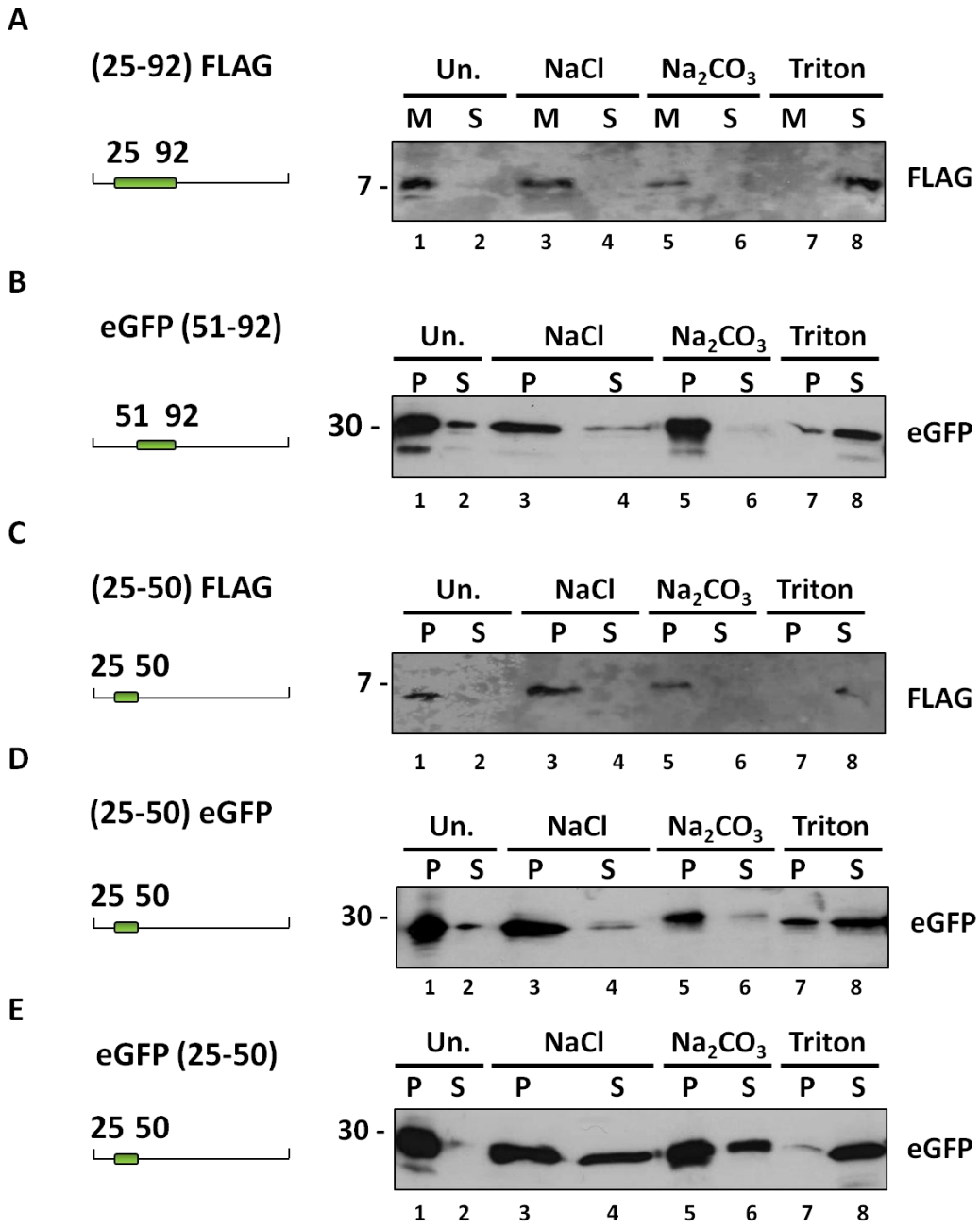


Figure 4.15 Membrane dissociation of the putative membrane spanning domains of NS2.
Isolated sub-domains of the NS2 *trans*-membrane domain were assessed for the affinity of their interactions with membranes. Proteins associated with the membrane fraction following centrifugation at 100,000 x *g* for 10 mins were resuspended in either isotonic buffer (Un.), 1M NaCl (NaCl), 100mM Na_2CO_3 pH 11.5 (Na_2CO_3) or 1% triton (Triton) and incubated at 25°C for 30 minutes. Fusions were either recovered in the pellet (P) or soluble (S) fraction following centrifugation of treated membranes at 100,000 x *g*. Molecular weight markers are shown to the left, the fusion tag probed for by immunoblot is noted to the right and lanes number are provided below each panel.

4.3 Discussion

Building upon the topology analysis of the previous chapter, truncations of NS2 were assessed for their ability to interact with membranes, the affinity of those interactions and whether the ability to target to membranes determined NS2 subcellular distribution in the absence of the other viral proteins.

4.3.1 NS2 contains multiple membrane-targeting sequences

Examination of the membrane-targeting of the NS2 fusions revealed the presence of multiple membrane-targeting sequences. As both eGFP (1-50) and eGFP (51-217) were recovered in the membrane fraction this was evidence of two distinct membrane-targeting sequences within NS2. Further investigation revealed that (1-24) and (25-50) were able to partially target to membranes suggesting that the targeting of residues 1-50 may be a result of two inefficient membrane-targeting sequences working synergistically. This hypothesis would support interactions between the predicted first and second TMD of NS2.The NS2-eGFP fusions indicated the presence of two membrane-targeting sequences located within residues 1-24 and 25-217, although targeting was inefficient.

The NS2 FLAG constructs displayed a higher propensity for membrane-targeting than the eGFP fusion, with both the *trans*-membrane domain (1-92) and the catalytic domain (93-217) targeting to membranes. This could be because a FLAG-tag (8 aa) represented a less disruptive feature to NS2 topology formation and membrane interactions than an eGFP molecule (238 aa). (1-50) FLAG and (51-217) FLAG both targeted exclusively to the membrane fraction (Figure 4.8) demonstrating that they each contain a distinct membrane-targeting sequence. Interestingly, dissection of these peptides revealed the potential for three membrane-targeting sequences within NS2 as (25-50) FLAG and (93-217) FLAG were targeted efficiently to membranes and a significant proportion of (51-92) FLAG was also recovered in the membrane fraction. This demonstrated the sequence spanning the second predicted TMD (aa 25-50) could efficiently target to membranes. However, the apparent membrane-targeting sequence located within aa 25-50 was not able to enhance the membrane targeting capacity of the sequence spanning the third predicted TMD (51-92) as both (51-92) and (25-92) FLAG were recovered in both the membrane and soluble fractions.

Solution of the catalytic domain by X-ray crystallography showed a tightly packed dimer with no apparent *trans*-membrane helices (Lorenz *et al.*, 2006b) (Figure 1.8). However the dimer was crystallised with two molecules of n-decyl-β-D-maltopyranoside (DM), a lipid-like non-ionic detergent. These detergent molecules co-crystallised on what is believed to be the

membrane-proximal face of the crystal dimer (Lorenz *et al.*, 2006b). The presence of these molecules and the requirement for detergent in NS2/3 cleavage buffers (Thibeault *et al.*, 2001) could indicate that the catalytic domain requires a stabilising hydrophobic environment which is ensured by the C-terminal membrane-targeting sequence. In agreement with this the FLAG-tagged catalytic domain, (93-217) FLAG, was able to target to membranes. On the surface, this appeared to be contradicted by expressed as a fusion to eGFP which was either soluble, (93-217) eGFP, or inconsistent, eGFP (93-217). In the absence of the *trans*-membrane domain the catalytic domain may only form a weak associate with membranes. Expression of the catalytic domain as a fusion to a large, soluble eGFP molecule may be sufficient to destabilise the interaction between the catalytic domain and membranes.

Overall these data suggest that residues 25-50, which are predicted to encompass the second TMD of NS2, contain a membrane-targeting sequence that may interact with upstream sequences in the putative first TMD. A second membrane-targeting sequence may be located within residues 51-92 (predicted to encompass the third TMD) as this domain expressed as a FLAG-tagged peptide or as an N-terminal fusion to eGFP i.e. eGFP (51-92), was able to partially target to membranes. The FLAG-tagged truncations also indicated the presence of a membrane-targeting sequence within the catalytic domain (aa 93-217). These data are consistent with the observations of Yamaga and Ou who reported the presence of two membrane-targeting sequences within between residues 30-74 and 119-151 (Yamaga & Ou, 2002). As no truncations were made into the catalytic domain no refinement can be made regarding the proposed location of the C-terminal membrane-targeting sequence (aa 119-151). However, the data presented here suggests that the N-terminal membrane-targeting sequence located between residues 30-74 may in fact represent two distinct targeting domains between residues 30-50 and 51-74.

4.3.2 Correlation between membrane-targeting and subcellular localisation

NS2 has been shown to target to membranes (Santolini *et al.*, 1995; Yamaga & Ou, 2002) and to co-localise with markers of the ER (Franck *et al.*, 2005; Kim *et al.*, 1995; Yamaga & Ou, 2002; Yang *et al.*, 2006). Consistent with this all three tagged versions of full-length NS2 (wt) associated with membranes and produced a diffuse, peri-nuclear staining pattern by confocal IF microscopy. However, NS2-FLAG appeared to localise to more discrete structures (Figure 4.11) compared to the two full-length eGFP-tagged versions (Figure 4.9 and 4.10), possibly as a result of the detection methods used.

One of the questions addressed by this work was whether the membrane-association of NS2 determined its subcellular distribution. To aid in the analysis of the NS2 truncations; a table was generated summarising their membrane-targeting abilities and whether they achieved a sub-cellular localisation similar to their respective full-length NS2 fusions (Table 4.1). From this summary it was evident that membrane-targeting of NS2 truncation was not sufficient to confer correct localisation as numerous fusions were recovered in the membrane fraction and yet were unable to achieve a cellular distribution similar to wt. Furthermore, it was apparent that the two mechanisms may be entirely distinct as two truncations, (1-50) eGFP and (Δ25-50) eGFP, achieved wt localisation despite being unable to target to membranes (Table 4.1).

The *trans*-membrane domain, in particular residues 1-50, seems to be involved in both processes as the (1-92) and (1-50) fusions showed a general ability to target to membranes and assume a wt localisation (Table 4.1). However, it appears as though neither of the sequences predicted to contain the first and second TMDs (aa 1-24 and 25-50 respectively) are absolutely required for membrane-targeting or subcellular localisation. All of the (25-217) fusions were able to target to membranes and all had subcellular distributions similar to wt protein. Dissection of the *trans*-membrane domain into three peptides (1-24, 25-50 and 51-92), each predicted to encompass one of the three TMDs, suggested a co-operative effect of the three domains as individually they displayed aberrant sub-cellular localisations and a reduced ability to target to membranes (Table 4.1).

4.3.3 Domains responsible for subcellular distribution

As membrane association did not appear to dictate the subcellular distribution of NS2, the fusions were scrutinised with an aim to identifying sequences responsible for producing NS2 subcellular distribution. However no single domain appeared to be solely responsible for achieving wt subcellular distribution.

Analysis of the FLAG-tagged peptides implicated sequences within the *trans*-membrane domain and the catalytic domain that combined to achieve wt staining as only the catalytic domain was common to the three fusions which displayed similar distribution to full-length NS2, yet the catalytic domain on its own, (93-217) FLAG, displayed an aberrant phenotype.

A greater degree of consistency was seen between the two sets of eGFP fusions (Figure 4.9 and 4.10). In both cases (1-92) and (1-50) redistributed the reporter in similar pattern to that of the full-length fusions. The catalytic domains in isolation were unable to achieve a correct subcellular distribution, but could be partially or fully recovered by the addition of all but the putative first TMD in the form of (25-217) eGFP and eGFP (25-217), respectively. Deletion of

NS2 region	Tag	M/S	WT IF
1-217	N1	M	Y
	C2	M	Y
	FLAG	M	Y
25-217	N1	M	Y
	C2	M	Y/N
	FLAG	M	Y
51-217	N1	-	-
	C2	M	N
	FLAG	M	N
93-217	N1	S	N
	C2	-	Y/N
	FLAG	M	N
1-92	N1	-	Y
	C2	M	Y
	FLAG	M	N
1-50	N1	S	Y
	C2	M	Y
	FLAG	M	N
25-92	N1	Mostly s	Y/N
	C2	Mostly s	N
	FLAG	M/S	N
1-24	N1	M/S	-
	C2	M/S	N
	FLAG	-	-
25-50	N1	S	-
	C2	M/S	N
	FLAG	M/S	N
51-92	N1	S	-
	C2	M/S	N
	FLAG	M/S	N
Δ25-50	N1	S	Y
	C2	M	Y
	FLAG	M	Y
Δ25-92	N1	M	Y/N
	C2	-	Y/N
	FLAG	M	N
Δ51-92	N1	-	Y/N
	C2	-	N
	FLAG	M/S	N

Table 4.1 Summary of the membrane-targeting and subcellular localisation of the eGFP and FLAG-tagged NS2 fusions.

The phenotypes observed for the membrane-targeting (M/S) i.e. recovered in the membrane (M), soluble (S) or both fractions (M/S), of each fusion and their subcellular distributions were similar to that of the full-length proteins (WT IF) i.e. yes (Y), no (N) or mixed (Y/N), as determined by confocal IF are summarised. Phenotypes were either comparable to (green) or different from (red) the full-length fusions. Peptides that were recovered in both the membrane and soluble fractions or displayed subcellular localisation similar to wt but with limited additional punctate structures are shown in yellow. (-) denotes peptides where no phenotype could be assigned.

the second sub-domain (Δ25-50) had no effect on the staining pattern for either eGFP fusion or the corresponding FLAG fusion (Figure 4.11). Deletion of the both second and third sub-domains (Δ25-92) from the eGFP fusions resulted in a partially disrupted phenotype yet surprisingly this was more significantly disrupted by deletion of the third putative TMD, (Δ51-92). A hypothesis suggesting that interactions between the *trans*-membrane domain and the catalytic domain combine to achieve wt distribution was further added to by the wt distribution of (25-217) eGFP yet when it was expressed as two constituent domain (25-92) and (93-217) they aggregated in large punctae.

In conclusion, truncation analysis of NS2 indicated that the catalytic domain is not required for subcellular localisation and that regions of the trans-membrane domain may interact to induce NS2 subcellular localisation.

4.3.4 Membrane affinity

The membrane targeting experiments highlighted that multiple domains of NS2 were capable of interacting with membrane. To determine whether truncations of NS2 that targeted to membranes were forming integral membrane domains the membrane-targeted fractions were treated with high molarity salt (1 M NaCl) and alkaline carbonate (100 mM Na_2CO_3, pH 11.5) which remove peripherally associated proteins from membranes (Fujiki *et al.*, 1982; Matsunaga *et al.*, 2002).

Treatment of full-length NS2 fusions with high molarity salt and alkaline carbonate confirmed previous reports that NS2 is an integral membrane protein as all three tagged versions were resistant to both agents (Figure 4.13A and B) (Santolini *et al.*, 1995; Yamaga & Ou, 2002).

Deletion of the first or the first and second putative TMDs did not prevent NS2 from inserting into membranes as neither (25-217) nor (51-217) FLAG were displaced with high molarity salt or alkaline carbonate (Figure 4.13C and D). Similarly, the complete *trans*-membrane domain (1-92) with a C-terminal FLAG was efficiently targeted to and inserted into membranes.

The membrane-targeting experiments highlighted that the phenotype of the catalytic domain (93-217) varied dramatically depending upon the fusion situation. A similar inconsistency was noted in the membrane affinity of those molecules able to target to membranes (Figure 4.14B, C and D). As discussed earlier, the catalytic domain may form a weak association with membranes that could be disrupted by the fusion of this domain to an eGFP molecule, whereas epitope tagging of this domain with the FLAG peptide may be less disruptive. The inconsistency of membrane-targeting and membrane affinity of the catalytic domain is highly

suggestive that this domain does not integrally insert into the membrane. In support of this conclusion; the integral membrane protein ICAM-I was consistently recovered in the membrane fraction and only extracted following addition of detergent, as were full-length NS2 (1-217) and the NS2 *trans*-membrane domain alone (1-92). In agreement with the X-ray crystal structure of the catalytic domain it was concluded that residues 93-217 likely form a soluble domain that interacts superficially with membranes via an internal membrane-targeting sequence.

In the previous chapter a model was presented for the topology of NS2 based upon the glycosylation data (Figure 3.17 **– Model F**). The glycosylation data showed that NS2 achieves a three TMD topology but little could be concluded as to the precise position of the TMDs suffice to say they are all contained within the N-terminal 86 residues.

It has previously been shown that aa 1-27 of NS2 from the Con1 isolate, expressed with a C-terminal eGFP molecule, displayed a comparable subcellular localisation to that of full-length NS2 from the same isolate (Jirasko *et al.*, 2010). The fusions involving only residues 1-24 of NS2 described in this chapter were either not detected by confocal IF microscopy or showed highly altered distribution compared with full-length NS2 fusions. The subtle difference in peptide length or the different genetic origins of the sequences examined may be responsible for these inconsistencies. Analysis of S9 and B9 in the previous chapter, which included aa 1-27 of NS2, suggested that this sequence may not encompass the whole of the first TMD of NS2. Two of the fusions examined in the present chapter including aa 1-24 of NS2 showed a partial membrane association (Figure 4.6 and 4.8). The N-terminal domain of NS2 may form the first TMD but in the systems employed in this study it was not able to do so very efficiently. Poor efficiency of this integration could be inferred from the low levels of active SEAP detected for the S9 fusion and minimal sensitivity of B9 to glycosidase treatment. The NMR structure of this domain led the authors to speculate that this domain could form a *trans*-membrane helix provided hydrophilic residues within the helix were neutralised by complementary interactions with another intra-membrane domain (Jirasko *et al.*, 2010).

Analysis of (25-50) and (51-92) in isolation showed that they did not target efficiently to membranes (Table 4.1). More interesting though was the affinity of the interactions of those molecules that were able to target to membranes. Both of these domains, once targeted to membranes, appeared to form integral associations as only eGFP (25-50) was displaced from the membranes following treatment with chaotropic agents (Figure 4.15). Further evidence for aa 51-92 containing a TMD was provided by the membrane dissociation response of (51-217)

FLAG, which responded as an integral membrane protein while (93-217) did not. These data are consistent with aa 25-50 and 51-92 each containing a TMD, thus supporting the experimentally determine model proposed in the previous chapter (Figure 3.17 – **Model F**).

4.3.5 Truncated NS2 (tNS2)

A truncated form of NS2 (tNS2) was observed in lysates of cells transfected with the full-length NS2-FLAG and NS2-eGFP constructs. Termed tNS2 this protein has been detected with antibodies raised against the C-terminal catalytic domain of NS2 and antibodies targeting C-terminal tag epitopes, but not with corresponding N-terminal tagged NS2 (Popescu *et al.*, 2011; Stapleford & Lindenbach, 2011). In this study tNS2 appeared to be generated from a specific cleavage event occurring in the vicinity of residue 25 as side by side comparison of tNS2 with (25-217) FLAG showed them to migrate at similar rates by SDS PAGE analysis (Figure 4.8 top right panel). Membrane-targeting analysis revealed that although only a proportion of tNS2-eGFP and tNS2-FLAG were able to target to membranes (Figure 4.6 and Figure 4.8). The population of tNS2-FLAG that did associate with membranes appeared to be anchored to the membrane as it too was only solubilised by detergent treatment (Figure 4.13B).

tNS2 has been observed in experiments using full-length JFH-1 virus (Welbourn *et al.*, 2009) and chimeric viruses (Boson *et al.*, 2011; Jirasko *et al.*, 2008; Jirasko *et al.*, 2010; Ma *et al.*, 2011; Stapleford & Lindenbach, 2011). However, tNS2 has not been reported in all virus systems (Tedbury *et al.*, 2011; Yi *et al.*, 2009) and was not noted in initial work on NS2 carried out *in vitro* (Hijikata *et al.*, 1993a; Hijikata *et al.*, 1993b; Santolini *et al.*, 1995; Yamaga & Ou, 2002). The role, if any, of tNS2 within the virus life-cycle has not been explored.

NS2A of yellow fever virus (YFV) has been shown to undergo an additional internal cleavage event mediated by the viral protease (Nestorowicz *et al.*, 1994). As tNS2 was observed here in the absence of the other viral proteins this does not seem to be the case in HCV. By comparison with the (25-217) FLAG; the tNS2 cleavage event was predicted to occur in the vicinity of residue 25. This residue is predicted to reside in the cytosol between the putative first and second TMDs making it a potential target for a cytosolic protease. Alternatively, as the loop region between TMD 1 and TMD2 is predicted to be very small, the cleavage point that produces tNS2 may well be positioned within the membrane making it potential target for signal peptide peptidase, which has been shown to cleave HCV core within the membrane (McLauchlan *et al.*, 2002).

Identification of the cleavage site and subsequent mutation within the full length virus would help to establish the significance of tNS2 or whether it is an experimental or genotype artefact.

4.4 Conclusions

Full length NS2 associates with membranes regardless of terminal fusion reporter. In addition to this NS2 contains specific subcellular targeting motifs, which appear unaffected by N- and C-terminal reporters/epitope tags. Building on previous findings by Yamaga and Ou (Yamaga & Ou, 2002), evidence provided here describes the presence of three internal membrane-targeting sequences within NS2 between residues 30-50, 51-74 and 119-151.

The ability of NS2 to associate with membranes is not sufficient for its native sub-cellular localisation, nor does native sub-cellular distribution confer membrane association. This suggests that different domains are required for the two processes.

Finally, FLAG-tagged fusions provided evidence that aa 25-50 and 51-92 encode the second and third TMDs of NS2, consistent with topology **Model F** proposed in the previous chapter (Figure 3.17).

CHAPTER 5: CHARACTERISATION OF NS2 PROTEIN INTERACTIONS

5.1 Introduction

Since the identification of the JFH-1 isolate, which is capable of undergoing a full round of infectivity in cell culture (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005; Zhong *et al.*, 2005), the roles of NS2 and p7 have been the focus of several studies. The generation of bi-cistronic JFH-1 viruses allowed the systematic deletion of p7 and NS2 identifying that both had essential roles in the morphogenesis of infectious virus particles (Jones *et al.*, 2007; Steinmann *et al.*, 2007). Since then there have been numerous studies investigating the interactions between NS2 and the other viral proteins.

Co-immunoprecipitations (co-IP) using epitope tagged NS2 in full-length virus experiments identified interactions between NS2 and E1, E2, p7, NS3 and NS5A (Jirasko *et al.*, 2010; Ma *et al.*, 2011; Stapleford & Lindenbach, 2011; Tedbury *et al.*, 2011). Since NS2 has previously been shown to interact with both structural and non-structural proteins this has strengthened the hypothesis that NS2 may function in virion assembly by uniting nascent particles with RNA replication complexes (Dimitrova *et al.*, 2003; Pietschmann *et al.*, 2006). A schematic of the viral proteins along with two models of NS2 membrane topology are shown in Figure 5.1**.**

The interaction between NS2 and E2 has been implicated as essential for the production of infectious virions as mutations that block this interaction showed impaired or attenuated infectivity (Jirasko *et al.*, 2010; Popescu *et al.*, 2011; Stapleford & Lindenbach, 2011). Mutation analysis found that K81A had a slightly reduced E2 binding (Phan *et al.*, 2009; Stapleford & Lindenbach, 2011) and that mutation of D62A in NS2, which reduced infectivity 100 fold, could be recovered by the additional mutation of I360T in E2 (Phan *et al.*, 2009). Mutation of F77 to alanine impaired the NS2:E2 interaction but this could be recovered by a compensatory mutation with the TMD of E2 (V341A) (Jirasko *et al.*, 2010). Popescu *et al*., (Popescu *et al.*, 2011) found that by introducing alanines after residues 16 and 41, in order to alter the rotation of the α-helix of putative TMDs by 130°, that A41 insertion blocked the NS2:E2 interaction but was unaffected by the A16 insertion. It was also reported that deletion of residues 10-62 of NS2 blocked the NS2:E2 interaction (Popescu *et al.*, 2011). From this it was proposed that the NS2 *trans*-membrane domain is important for mediating an intra-membrane interaction between NS2 and E2 (Popescu *et al.*, 2011; Stapleford & Lindenbach, 2011).

Transposing the JFH-1 coding sequence for aa 1-27 with that of the Con 1 isolate blocked infectivity, but long-term passage in cell-culture produced a compensatory mutation of E3D in p7 which rescued infectivity, suggesting that p7 plays a regulatory role in the NS2:E2

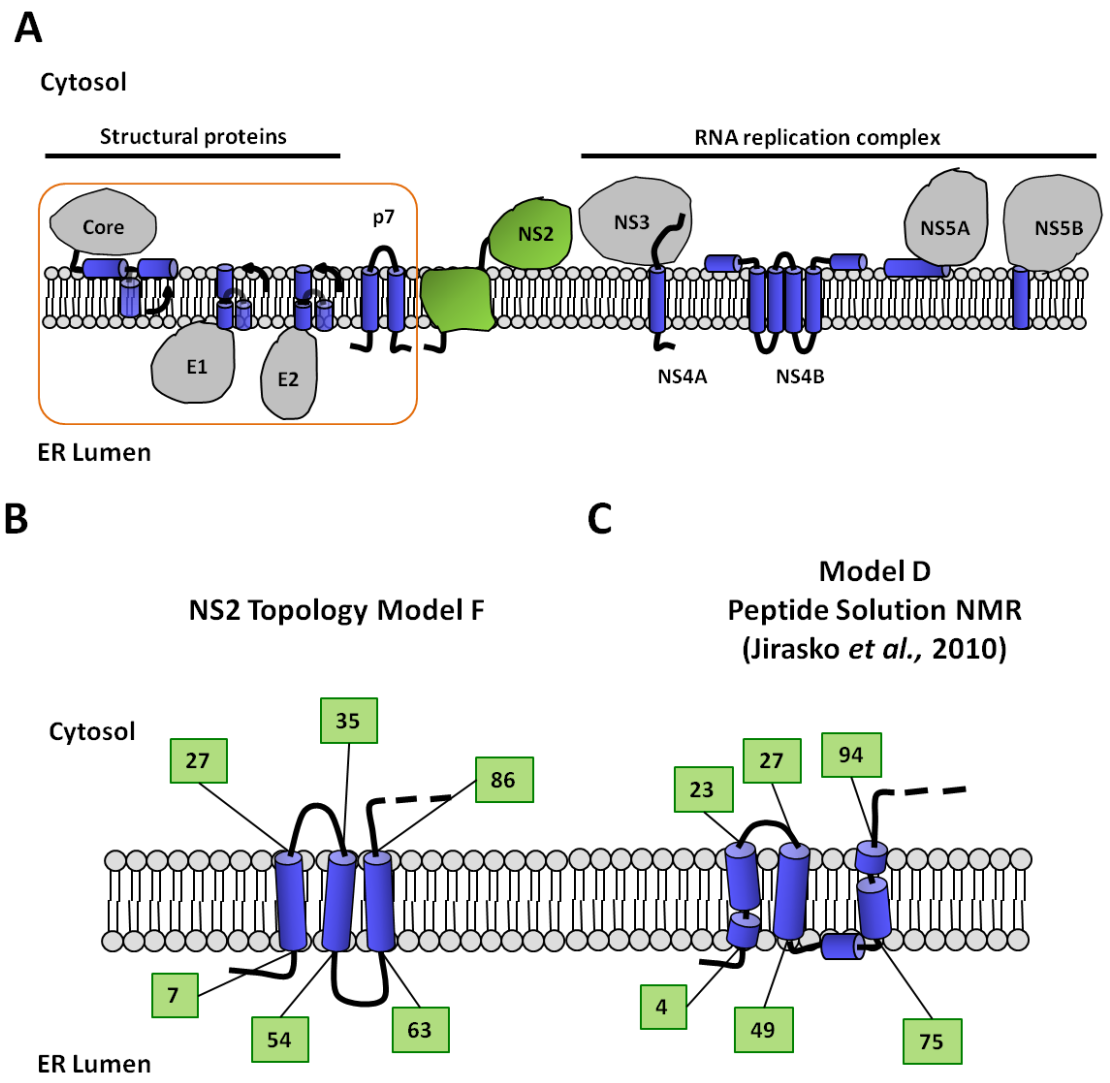


Figure 5.1 Membrane topology of HCV proteins.

(**A**) Schematic representation of the membrane interactions of the HCV proteins. The region encoding core to p7 (C-p7) (orange rectangle) was translated as a polypeptide in studies investigating the E2/NS2 interaction. NS2 is shown with three putative TMDs (green). TMDs of other viral proteins and the amphipathic helix of NS5A are shown as cylinders (blue) and the membrane is shown as a lipid bi-layer. (**B**) **Model F** of NS2 topology inferred from the reporter glycosylation and computational analysis (Chapter 3). (**C**) **Model D** of NS2 topology based on three solution NMR structures for peptides encompassed the 1st (aa 4-23), 2nd (aa 28-49) and 3rd (aa 73-93) putative TMDs (Jirasko *et al.*, 2008; Jirasko *et al.*, 2010). The catalytic domain is not shown. Residues marking the boundaries of predicted TMDs are noted (green boxes). JFH-1 coding sequence was used unless otherwise stated.

interaction (Jirasko *et al.*, 2010). In addition, it was reported that mutation of the p7 basic-loop or deletion of p7 blocked the NS2:E2 interaction (Ma *et al.*, 2011). However the role of p7 in the interaction of NS2 and E2 (NS2:E2) is contentious as another group found that neither mutation of the basic-loop or deletion of p7 was deleterious to the interaction, although deletion of p7 resulted in an increased mobility of the E2-specifc protein species detected compared with the wild type virus, suggesting that NS2 was ordinarily interacting with E2-p7 (Popescu *et al.*, 2011). Furthermore, a third group reported that while deletion of p7 dramatically impaired the NS2:E2 interaction it did not completely ablate it (Stapleford & Lindenbach, 2011).

In addition to forming interactions with other viral proteins, NS2 is proposed to form dimers. Bimolecular cleavage at the NS2/3 junction was first observed by Grakoui *et al*., (Grakoui, 1993b; Reed *et al.*, 1995). Subsequent solution of the catalytic domain structure by X-ray crystallography combined with biochemical validation provided evidence that the NS2 catalytic domain forms active-site dimers (Lorenz *et al.*, 2006b). Specifically, the catalytic residues histidine 143 and glutamate 163 are contributed by one monomer while cysteine 184, which facilitates cleavage by nucleophilic attack of the peptide bond (Lorenz *et al.*, 2006b), was coordinated by the other monomer. Co-expression of two separate precursors with distinct active mutations confirmed the potential for bi-molecular cleavage through dimerisation of the active site (Lorenz *et al.*, 2006b).

Further evidence of dimerisation of the catalytic domain was provided using analytical ultracentrifugation, protein cross-linking in solution and co-immunoprecipitations from cells co-transfected with two differently tagged forms of NS2/3 pre-cursor (Lorenz *et al.*, 2006b). Targeted mutagenesis of the dimer interface suggested that dimerisation of the catalytic domain was important for the role of NS2 in virus assembly, but not NS2/3 cleavage (Dentzer *et al.*, 2009). Whether or not the *trans*-membrane domain is involved in NS2 dimerisation has also not been investigated.

Taking advantage of the NS2 fusions described in the previous chapter; the NS2 intermolecular interactions were explored using the GFP-Trap^{TM} system of co-IP with the aim of identifying the specific domains of NS2 involved and whether these interactions correlated to NS2 membrane-targeting. It was hoped that identification of specific domains or functions of NS2 required for protein interactions would facilitate consolidation and interpretation of previous directed mutagenesis and rescue mutation studies.

5.2 Results

The tagged-NS2 constructs described in Chapter 4 provided an ideal opportunity to explore which regions of NS2 are involved in forming protein: protein interactions. As the fusion peptides facilitate more detailed characterisation of interactions involving the *trans*-membrane domain of NS2; E1, E2 and p7 were considered for initial study. The reliability and availability of detection reagents singled out E2 as the primary candidate for interaction studies.

5.2.1 Co-immunoprecipitation of E2 with NS2

A previous study investigated the interaction between NS2 and other viral proteins by performing immunoprecipitations (IPs) of tagged-NS2 in the context of fully infectious virus isolates and provided evidence that p7 was essential for facilitating the NS2/E2 interaction (Ma *et al.*, 2011). Taking this into account it was decided to express E2 *in cis* with p7 by taking advantage of a previously characterised construct in which the JFH-1 core-p7 (C-p7) coding region is expressed as a single polyprotein (Adair *et al.*, 2009).

Huh7 cells were singly transfected with a range of control constructs to confirm the specificity of the GFP-Trap™ system. Cells were harvested 48h post-transfection into lysis buffer containing 1% Triton. 5% of crude lysates were retained as inputs for comparison by immunoblot analysis. Clarified lysates were incubated with GFP-Trap™ beads overnight and immunoprecipitated proteins were recovered by addition of 2 x Laemmli buffer. Single transfections showed that E2 did not bind to the GFP-Trap™ beads as E2 was detectable in the input but not the IP fraction (Figure 5.2A **-** lanes 1 and 2). tNS2-eGFP was not detected under these conditions due to limited resolution of samples to facilitate probing of a single membrane for multiple protein species. Two protein species were detected using the E2 antibody, one of approximately 65 kDa; presumed to be mature E2, and a second high molecular weight species that was unable to enter the resolving gel matrix. A faint non-specific protein species of approximately 46 kDa was evident in the IP fractions of all samples. The majority of the eGFP fusions were expressed to levels below the detection limit for the input samples but all were concentrated following IP with the GFP-Trap™ beads.

Co-transfection of the C-p7 expression construct with either eGFP alone or eGFP-tagged NS2 confirmed previous reports that NS2 is able to immunoprecipitate E2 (Figure 5.2B) (Jirasko *et al.*, 2010; Ma *et al.*, 2011; Popescu *et al.*, 2011; Stapleford & Lindenbach, 2011). NS2 as an N and C-terminal fusion to eGFP was able to co-IP E2. eGFP alone, despite being

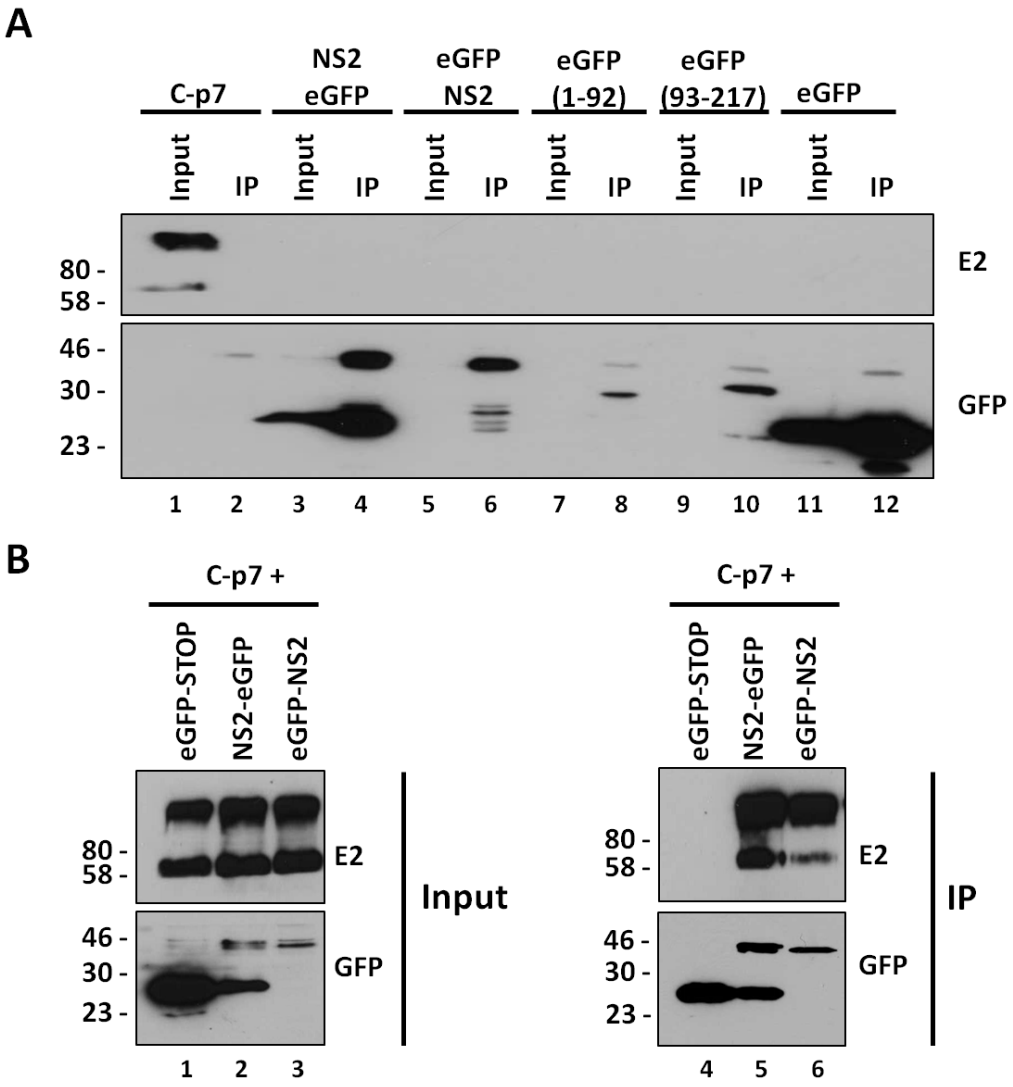


Figure 5.2 eGFP-tagged NS2 interacts with E2 expressed from the C-p7 construct.

(**A**) Huh7 cells were singly transfected with C-p7, eGFP alone and eGFP-fusions to test for non-specific interactions. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. Membranes were probed with antibodies specific for E2 (AP33, mouse monoclonal) and GFP (rabbit polyclonal). (**B**) Huh7 cells were co-transfected with C-p7 and eGFP-tagged NS2 or eGFP alone. Immunoprecipitations were carried out as above.

immunoprecipitated by the GFP-Trap™ beads, did not co-IP E2. Although side-by-side comparisons suggested that NS2-eGFP was more efficient at binding E2, the eGFP-NS2 fusion was used in subsequent experiments as a precaution because NS2-eGFP appeared to produce cleaved fusion molecules as well as tNS2-eGFP (Figure 5.2B and Appendix 4) and it was not known what effect these may have on poorly expressing fusions or weak interacting domains. Two protein species were detected again using the E2 antibody. As the higher molecular weight band may represent an insoluble form of E2, 4 M urea was added to samples to increase denaturation. This dramatically reduced the amount of the high molecular weight species (data not shown) and so 4 M urea was added to protein sample buffer thereafter.

5.2.2 Co-IP of E2 with other GFP-tagged proteins

To verify that the eGFP tag was not in some way responsible for the observed NS2:E2 interaction cells were co-transfected with C-p7 (JFH-1) and a range of other GFP-tagged proteins (Figure 5.3A). GFP-tagged N protein from infectious bronchitis virus (IBV) (Emmott *et al.*, 2008) and the cellular protein epidermal growth factor receptor (EGFR) Protein tyrosine kinase Substrate #15 (Eps15) (Mankouri *et al.*, 2008a), which is involved in the endocytosis of EGFR, showed that GFP expressed as a fusion to other proteins did not interact with E2. A non-GFP tagged protein, NS5A, was not detected and did not precipitate with E2 further confirming the specificity of the GFP-Trap™ system. On high exposures a small, but reproducible, amount of E2 was seen to be co-immunoprecipitated by NS4B (Figure 5.3A and B). On review of the literature; a physical interaction between E2 and NS4B expressed from an NS3-5B polyprotein has been suggested (Selby *et al.*, 1994), however, due to a lack of NS4B-specific reagents, this interaction has not been explored further.

5.2.3 Investigation of the cross-genotype interaction between NS2 and E2

Studies of chimeric JFH-1 viruses found that the highest extracellular virus titres were achieved if the genetic cross-over point was within the *trans*-membrane domain of NS2 (Pietschmann *et al.*, 2006). It was hypothesised that the increase in viral titres was due to increased affinity between N-terminal domains of NS2 and the structural proteins and p7 derived from the same sequence. As NS2 is able to interact with NS3 (Ma *et al.*, 2011) and p7 (Popescu *et al.*, 2011) from different genotypes the question was asked whether the increase in viral titres observed by Pietschmann *et al.,* (Pietschmann *et al.*, 2006) was due to a genetic determinant of the NS2:E2 interaction. To investigate this, cells were co-transfected with GFP-NS2 and a vector encoding C-p7 from the JFH-1 or the Con1 (genotype 1b) isolate (Figure 5.4A). E2 expression from the two C-p7 constructs was comparable (Figure 5.4A). Interestingly, the levels of E2 co-immunoprecipitated by eGFP-NS2 were also comparable.

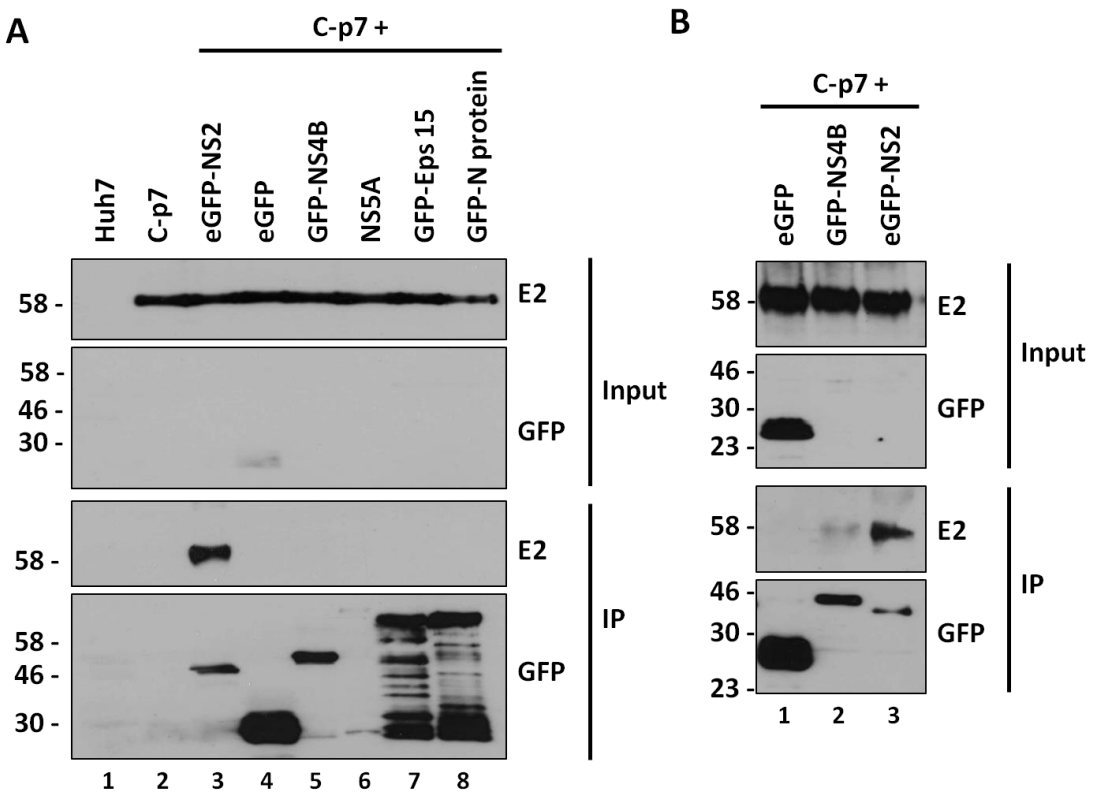


Figure 5.3 Verification of the specificity of the E2/NS2 interaction.

(**A**) Huh7 cells were co-transfected with C-p7 and control plasmids. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. Membranes were probed with antibodies specific for E2 (AP33, mouse monoclonal) and GFP (rabbit polyclonal). (**B**) IPs carried out on lysates from Huh7 cells co-transfected with C-p7 and either eGFP, NS4B-eGFP or eGFP-NS2. Molecular weight markers are shown to the right and lane numbers are noted below.

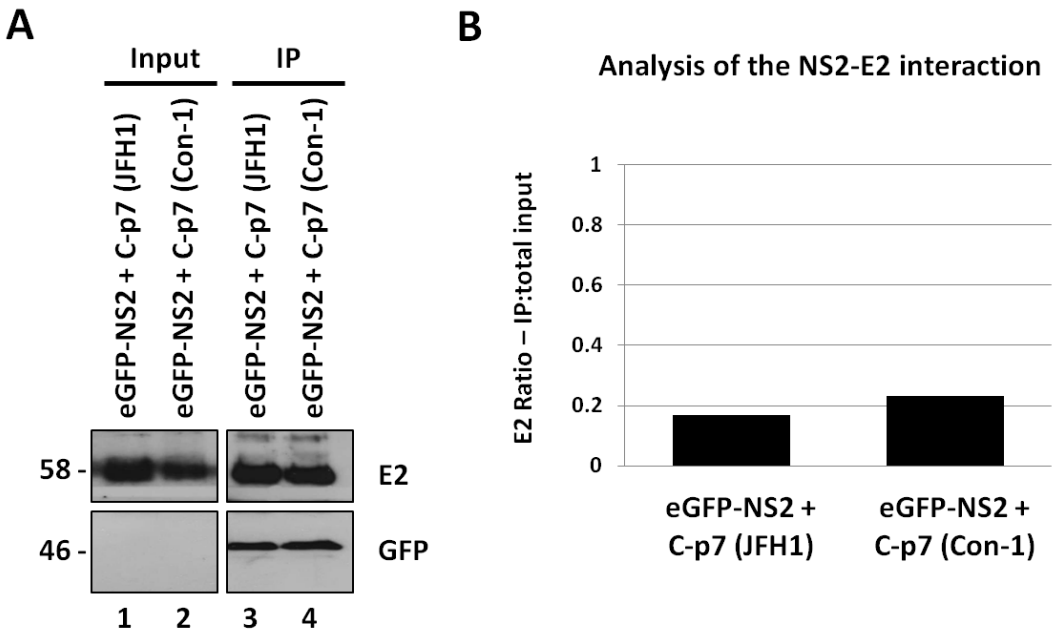


Figure 5.4 The NS2:E2 interaction is genotype-independent.

(**A**) Huh7 cells were co-transfected with eGFP-NS2 and C-p7 from the JFH-1 (genotype 2a) or Con1 (genotype 1b) isolates. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. Membranes were probed with antibodies specific for E2 (AP33, mouse monoclonal) and GFP (rabbit polyclonal). Molecular weight markers are shown to the right. Lane numbers are noted below. (**B**) Quantification of the interaction between NS2 and E2 from the JFH-1 and Con1 isolates by densitometry. Immunoblot images were colour-inverted and the background density was subtracted from average densities of the input and IP bands for each E2 species. Normalised input values were multiplied by 20 to represent total input. IP efficiency was plotted as a ratio of IP: total input E2.

The comparative affinity of NS2 for E2 was calculated using densitometry analysis of the respective input and IP fractions for each of the E2 species on the same exposure. As the IP fractions represented the total IP fraction whereas the input fractions represented only 5% of the total input sample the normalised input value was multiplied by 20. The efficiency of the NS2/E2 interaction was then expressed as a ratio of IP: input (Figure 5.4B). A ratio value of 1 would have indicated that all the expressed E2 was co-immunoprecipitated with NS2. This showed the amount of E2 immunoprecipitated to be approximately 20% of the total input E2 irrespective of the genetic origin of E2 (Figure 5.4B).

5.2.4 Analysis of the E2 binding domain of NS2

Following confirmation of the NS2:E2 interaction using GFP-tagged NS2 the question was asked whether any sub-domains of NS2 were dispensable for the interaction. Huh7 cells were co-transfected with the eGFP-NS2 fusion constructs and the JFH-1 C-p7 vector (the JFH-1 vector was used hereafter and all NS2 truncation fusions were generated from the JFH-1 sequence) and IPs performed on lysates harvested 48 hours post-transfection (Figure 5.5). Control samples were transfected in duplicate and the lysates pooled and loaded on both gels to allow direct comparison. Input and IP fractions were probed for E2, core and GFP. Core and E2 input blots confirmed relatively consistent transfection of the C-p7 plasmid (Figure 5.5).

As seen in the previous chapter, the eGFP-NS2 fusions expressed to varying levels (Figure 4.5B), but they were all detected in the IP fractions (Figure 5.5). Full-length NS2 (1-217) was able to efficiently interact with E2, but E2 did not co-IP with all of the NS2 fusions. Levels of E2 comparable to the full-length NS2 fusion were immunoprecipitated by the (25-217) and (Δ25-50) fusions. (1-50) and (1-92) also interacted with E2 but to a lesser extent. Higher exposures showed that trace amounts of E2 were immunoprecipitated by (25-92), (51-92), (51-217) and (Δ25-92) (data not shown). No interaction was detected between E2 and (1-24), (25-50), (93-217) or (Δ51-92). None of the fusions immunoprecipitated core protein. All the core blots shown were carried out in parallel and exposed on the same film (Figure 5.5).

The apparent low levels of input E2 from cells co-transfected with (25-217) and C-p7 (Figure 5.5 - lane 10) could represent an artefact of the transfer procedure as the corresponding core immuno-blots showed protein levels comparable with all other samples. Furthermore levels of E2 comparable to that of the full-length eGFP fusion were immunoprecipitated by (25-217). A similarly low level of E2 Input was noted for (Δ51-92) (Figure 5.5 **-** lane 20) but upon repetition this was not the case (Figure 5.6A **-** lane 20). In both transfections there was no evidence of (Δ51-92) interacting with E2.

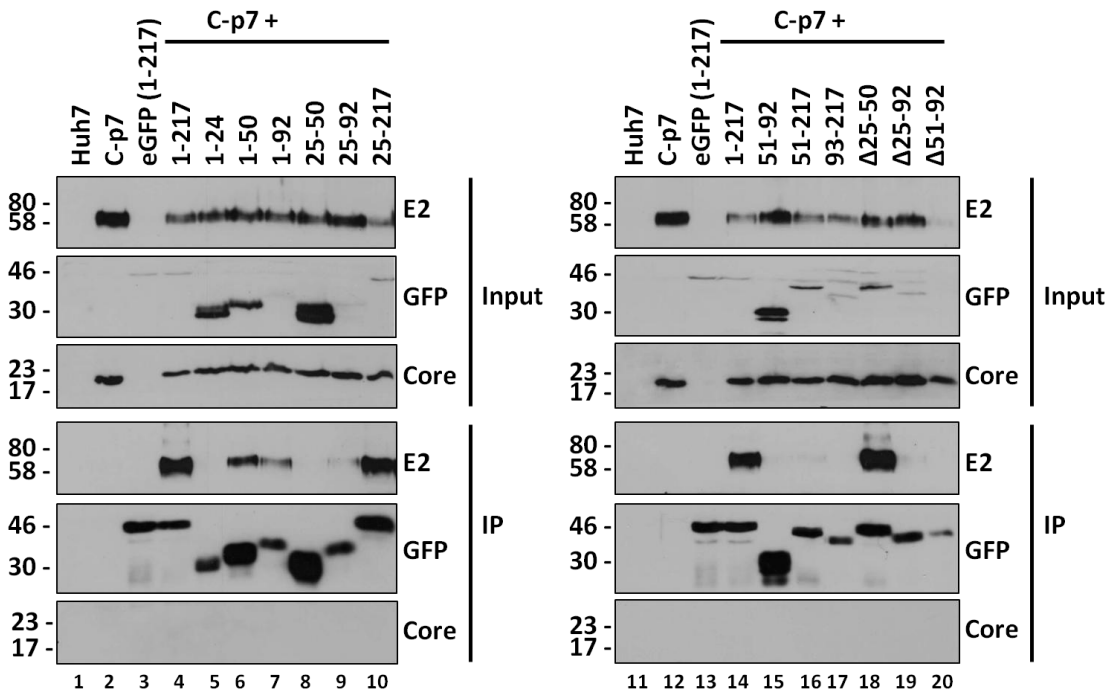


Figure 5.5 Characterisation of the E2-binding domain of NS2.

Huh7 cells were co-transfected with C-p7 and the eGFP-NS2 fusion set. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. Membranes were probed with antibodies specific for E2 (AP33, mouse monoclonal), core protein (sheep polyclonal) and GFP (rabbit polyclonal). Fusions are numbered according to the NS2 residues they encode. The first four samples, lanes 1-4, were performed in duplicate and the lysates pooled and electrophoresed on both gels to allow direct comparison. Molecular weight markers are shown to the right and lane numbers are noted below.

To confirm the phenotypes of these the eGFP-NS2 fusions the experiment was repeated using the eGFP-NS2 fusions (Figure 5.6A) and the NS2-eGFP fusions (Figure 5.6B). The eGFP-NS2 fusions produced the same pattern as before (Figure 5.5) except that on higher exposures (25-50) appeared to interact with E2 and the weak interaction previously observed between eGFP (51-92) and E2 was not reproduced (data not shown).

Analysis of the NS2-eGFP fusions confirmed the phenotypes of the (1-92) (25-217) and (Δ25-50) truncations (Figure 5.6B). The most notable discrepancy between the two sets of fusions was that eGFP (1-50) efficiently immunoprecipitated E2 while (1-50) eGFP showed no signs of an interaction. This suggests that the fusion molecule was disrupting the NS2:E2 interaction in the latter instance either by steric inhibition or by altering the targeting or structure of the NS2 domain.

Higher exposures only identified a weak interaction between eGFP (25-92) and E2. The phenotypes described for eGFP (51-92) and eGFP (51-217) could not be confirmed as only a trace amount of (51-92) eGFP was visible on over-exposed films and no expression was detected for (51-217) eGFP. Poor expression was repeatedly observed for these fusions, particularly (51-217) eGFP; the expression of which was only ever detected by confocal IF microscopy (Figure 4.9).

The apparent cleaved fusion products detected in cells transfected with the NS2-eGFP constructs (Appendix 4) did not appear to have a detrimental effect on those fusions displaying a high affinity for E2 as (1-217), (25-217) and (Δ25-50) all efficiently immunoprecipitated E2. However, the reduced number of weak interactions observed on higher exposures compared with the eGFP-NS2 fusions could have been the result of interference by cleaved NS2 species competing for E2.

From comparison of the two GFP-tagged NS2 truncation sets it is clear that certain regions of NS2 can interact with E2, however no single domain appeared essential for enabling the NS2:E2 interaction. E2 was efficiently co-immunoprecipitated by (1-92), (25-217) and (Δ25-50) from both sets and eGFP (1-50). This suggests that NS2 contains multiple E2-binding domains. It has also been reported that E1, E2, p7 and NS2 form a complex (Stapleford & Lindenbach, 2011) which may mean that NS2 has indirect associations with E2 via p7 or E1. Unfortunately E1 and p7 expression could not be detected in this system preventing exploration of this hypothesis.

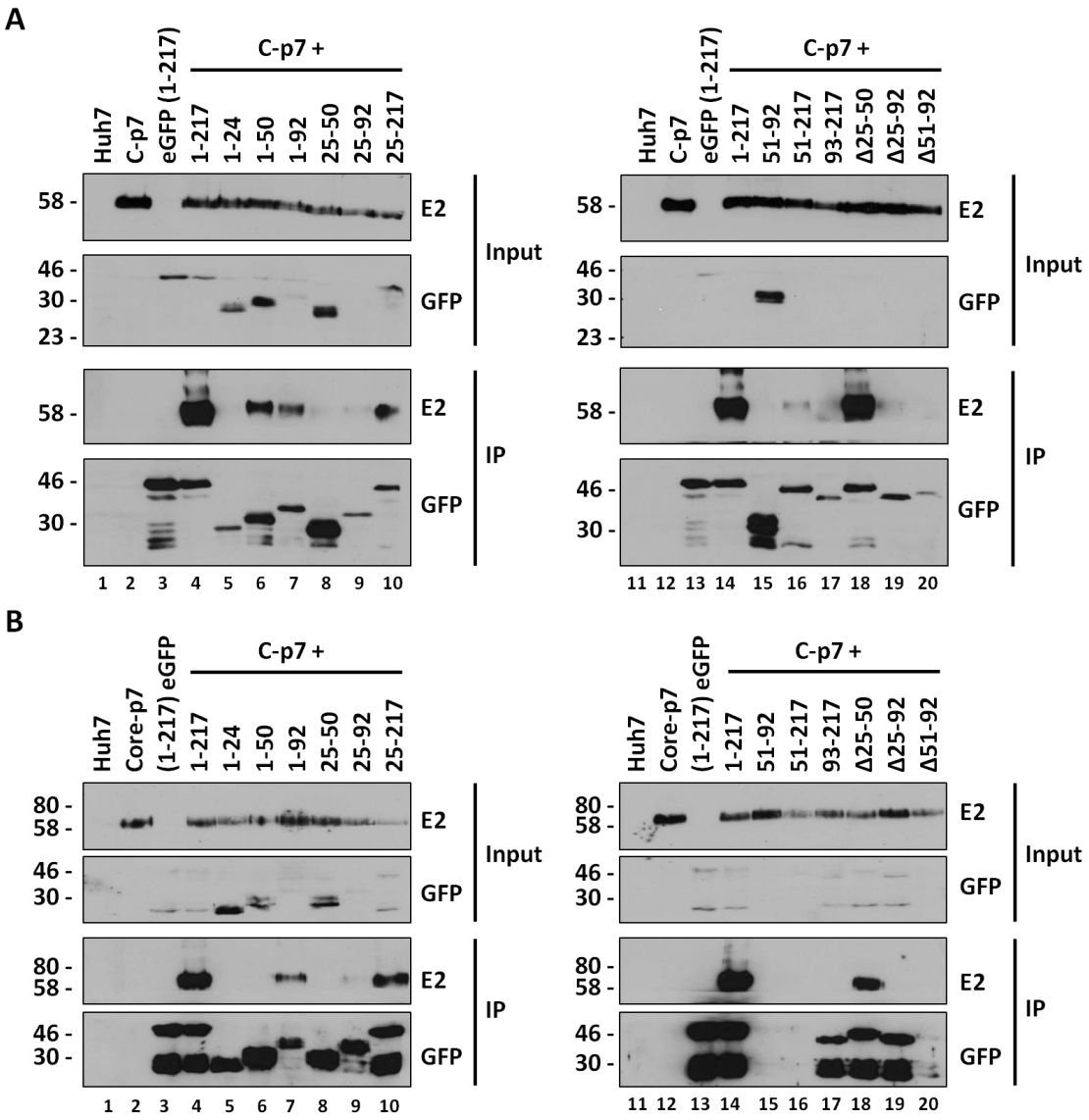


Figure 5.6 Further characterisation of the E2-binding domain of NS2.

Huh7 cells were co-transfected with C-p7 and either the eGFP-NS2 fusions (**A**) or the NS2-eGFP fusions (**B**). Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into 4 volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted, washed three times in TBS and heated at 95°C in 2 x protein sample buffer for 3 minutes to remove bound proteins (IP). Total IP fraction and 5% of total input were analysed by SDS PAGE, transferred to PVDF membranes and probed for E2 and GFP. Fusions are numbered according to the NS2 residues they encode. The first four samples, lanes 1-4, were in duplicate and the lysates pooled and electrophoresed on both gels to allow direct comparison. Molecular weight markers are shown to the right and lane numbers are noted.

5.2.5 Interactions between NS2 and the non-structural proteins

As discussed previously, NS2 has been shown to interact with NS3, NS4A and NS5A (Flajolet *et al.*, 2000; Jirasko *et al.*, 2010; Ma *et al.*, 2011; Stapleford & Lindenbach, 2011; Tedbury *et al.*, 2011). With the aim of identifying which regions of NS2 were capable of interacting with individual non-structural proteins, and whether p7 was required for these interactions, Huh7 cells harbouring the autonomously self-replicating sub-genomic RNA, or sub-genomic replicon (SGR), encoding the neomycin resistance gene and the NS3-5B or SPp7-NS5B proteins from the JFH-1 isolate were used (Tedbury *et al.*, 2011). The non-structural proteins are thought to form replication complexes that may exclude elements of the innate immune response (Moradpour *et al.*, 2007), and possibly other viral proteins. As these would be pre-formed in cells stably harbouring HCV replicons and may prevent exogenously expressed eGFP-NS2 from interacting with the other viral proteins, RNA encoding transient NS3-5B replicon with a luciferase reporter gene in place of the antibiotic resistance gene was generated (Tedbury *et al.*, 2011). The NS2:NS5A interaction appears to be independent of the structural proteins (Tedbury *et al.*, 2011) whereas the NS2:NS3 interaction seems to be destabilised by disruption to the structural protein suggesting that the NS2:NS3 interaction may require the presence of the structural proteins. To enable the investigation of this possibility full-length JFH-1 virus RNA (Wakita *et al.*, 2005) was also generated. Schematic diagrams of the two SGRs compared to the full-length virus genome are shown in Figure 5.7A**.**

Huh7 cells were electroporated in duplicate with full-length JFH-1 virus RNA (Virus) or JFH-1 NS3-5B transient replicon RNA. 24 hours post-electroporation cells were either transfected with eGFP alone or eGFP-NS2 and harvested 72 hours post-electroporation. The two stable replicon harbouring cell-lines described above were also transfected in parallel to enable direct comparison. Input levels of core, E2 and NS5A confirmed that cells had been electroporated with the virus and replicon RNAs (Figure 5.7B). Levels of NS3 and NS5A in the stable replicon cells were inconsistent and notably higher in the mock transfected cells. The transfection efficiency of all six cell populations was comparable. E2 from virus-electroporated cells co-immunoprecipitated with eGFP-NS2 and not with eGFP alone (Figure 5.7B - lanes 4 and 14), although considerably less efficiently than when the two were expressed from DNA constructs (Figure 5.2B). At first it appeared as though a trace amount of NS5A might have co-immunoprecipitated with eGFP-NS2 in virus-electroporated cells (Figure 5.7B - lane 4 and 14), but on closer inspection trace amounts of this protein species were also detected in the GFP alone transfected cells. None of the other HCV proteins probed for were detected in IP fractions in both the transient and stable expression systems (Figure 5.7B).

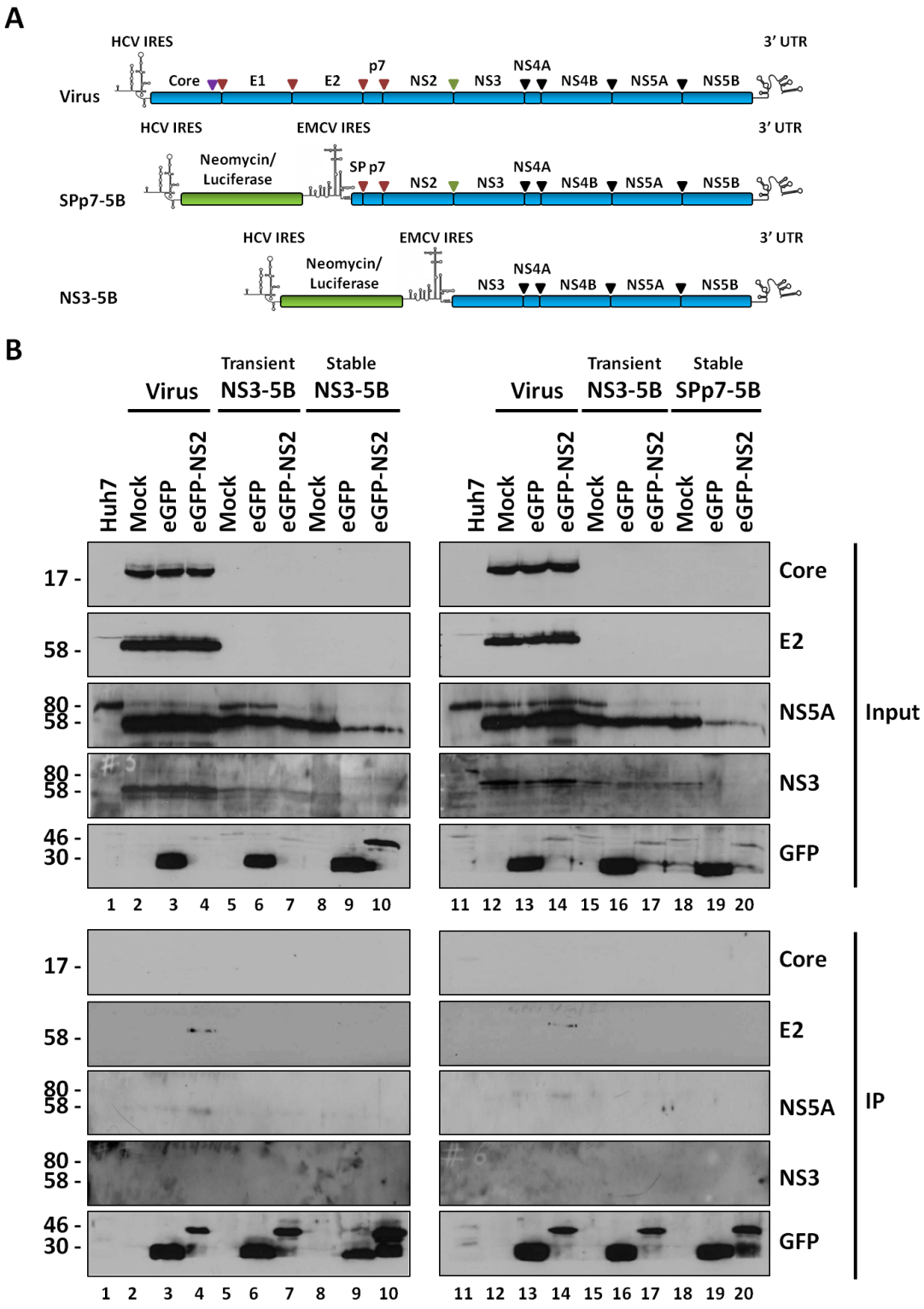

A
HCV IRES
3' UTR
Virus
Core
E1
E2
p7
NS2
NS3
NS4A
NS4B
NS5A
NS5B
HCV IRES
EMCV IRES
3' UTR
SPp7-5B
Neomycin/
Luciferase
SP p7
NS2
NS3
NS4A
NS4B
NS5A
NS5B
HCV IRES
EMCV IRES
3' UTR
NS3-5B
Neomycin/
Luciferase
NS3
NS4A
NS4B
NS5A
NS5B
B
Virus
Transient
NS3-5B
Stable
NS3-5B
Virus
Transient
NS3-5B
Stable
SPp7-5B
Huh7
Mock
eGFP
eGFP-NS2
Mock
eGFP
eGFP-NS2
Mock
eGFP
eGFP-NS2
Huh7
Mock
eGFP
eGFP-NS2
Mock
eGFP
eGFP-NS2
Mock
eGFP
eGFP-NS2
17 -
58 -
80 -
58 -
80 -
58 -
46 -
30 -
Core
E2
NS5A
NS3
GFP
Input
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
17 -
58 -
80 -
58 -
80 -
58 -
46 -
30 -
Core
E2
NS5A
NS3
GFP
IP
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20

Figure 5.7 Investigation of the interactions between NS2 and other viral proteins.

(**A**) Schematic diagram of the genome organisation of full-length HCV and two sub-genomic replicons. The JFH-1 sequence encoding core-p7 (NS3-5B) or core-E2 (SPp7-5B) was replaced with a reporter gene enabling the study of replication in stable (neomycin) or transient (luciferase) replication systems (Tedbury *et al.*, 2011). Native JFH-1 5' and 3' UTRs were maintained to enable replication. Translation of HCV proteins from the replicons was controlled by an encephalomyocarditis virus (EMCV) IRES. Polyprotein processing events are marked by arrow heads: signal peptidase (red), signal peptide-peptidase (purple), NS2 (green), NS3/4A (Black). (**B**) The ability of NS2 expressed in *trans* to interact with other viral proteins expressed from different environments was assessed. Huh7 cells were electroporated with full-length JFH-1 virus RNA (virus) or RNA encoding the non-structural protein NS3-5B (Transient NS3-5B). 24 hours post-electroporation cells were transfected with either eGFP alone or eGFP-NS2. Huh7 cell populations containing stably replicating sub-genomic RNA molecules encoding either NS3-5B or SPp7-5B were transfected in parallel. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. Molecular weight markers are shown to the right and lane numbers are noted below. Membranes were immunoblotted for the proteins indicated to left.

The protein species detected with the NS3 (~70 kDa)-specific antibody in the input fractions migrated faster than predicted (~58kDa) (Figure 5.7B). Due to the nature of the experiment only one gel could be produced for each of the input and IP samples. Consequently the top portions of the membranes were probed, then stripped and re-probed twice to allow the systematic detection of E2, NS5A and NS3 respectively. NS3 and NS5A antiserum were both generated in sheep; therefore it seemed likely that the protein species observed for NS3 were due to the re-detection of residual NS5A antibody that had not been completely removed. In order to verify that eGFP-NS2 was indeed less efficient at interacting with E2 expressed from the full-length virus RNA compared with E2 expressed from the C-p7 plasmid DNA; virus RNA electroporated cells were transfected as before and immunoprecipitated in parallel with cells co-transfected with eGFP-NS2 and the C-p7 vectors (Figure 5.8A).

As input levels of core, E2 and GAPDH were higher in eGFP-NS2 transfected cells it was possible that over-expression of NS2 in the context of the other viral proteins increased cell proliferation (Figure 5.8A). However, as this increase in protein levels was not seen previously (Figure 5.7B), it was suspected that virus electroporated cells were unevenly seeded out prior to transfection.

Levels of eGFP-NS2 immunoprecipitated from the virus-electroporated cells and the C-p7-transfected cells were comparable, yet the amount of E2 they immunoprecipitated was not. eGFP-NS2 was able to immunoprecipitate considerably more E2 when it was expressed from the C-p7 vector despite the input levels of E2 being significantly higher in cells expressing the full virus polyprotein. A similar number of cells were harvested for the eGFP-NS2 transfected samples as indicated by the input levels of GAPDH (Figure 5.8 **-** lane 3 and 4).

Densitometry was performed to quantify the efficiency of the E2/NS2 co-IPs in a similar manner to that described in Figure 5.4B. Due to the continuous nature of the E2 input band it was not possible to examine the total density of the bands, instead the average density of a vertical line taken through the centre of the E2 bands was used. Similar to the analysis shown in Figure 5.4B, NS2 co-transfected with the C-p7 construct was able to precipitate approximately 15% of the total E2 (Figure 5.8B). Although crude, this analysis highlighted that NS2 was 10-fold more efficient at interacting with E2 expressed from the C-p7 DNA construct than when E2 was expressed from an RNA molecule in the context of the complete virus polyprotein.

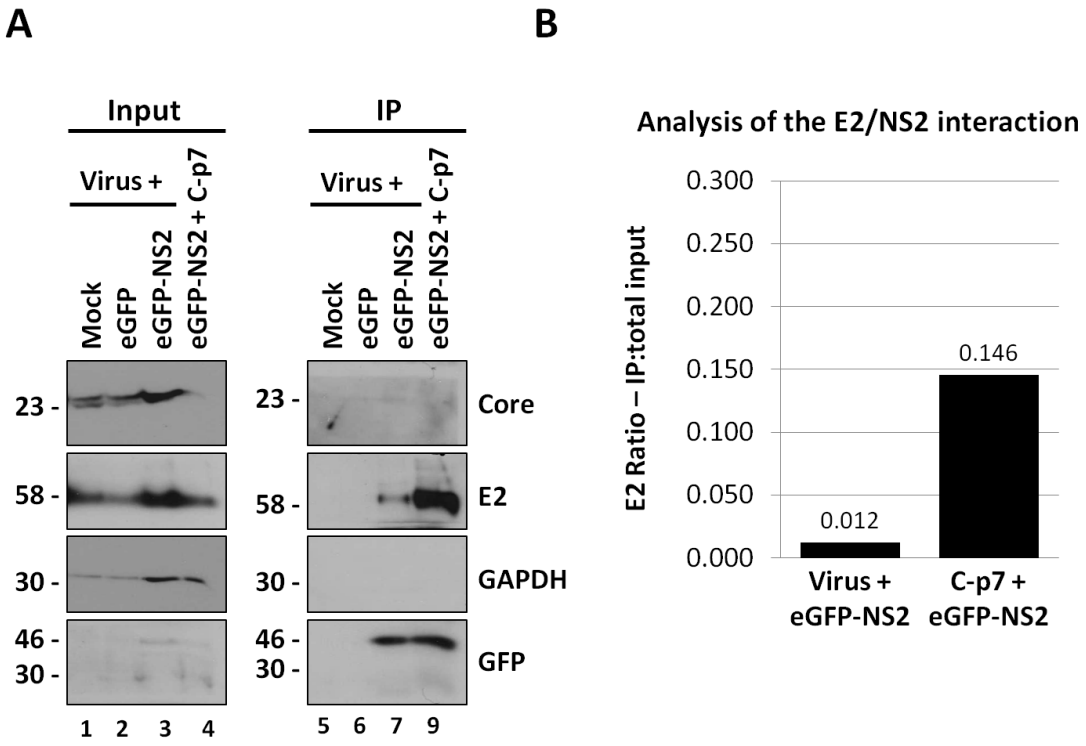


Figure 5.8 The eGFP-NS2/E2 interaction is less efficient in virus expressing cells.

(**A**) Huh7 cells were electroporated with full-length JFH-1 virus RNA (Virus). 24 hours post-electroporation cells were transfected with either eGFP alone or eGFP-NS2. In parallel Huh7 cells were co-transfected with eGFP-NS2 and C-p7. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. Molecular weight markers are shown to the right and lane numbers are noted below. Membranes were immunoblotted for the proteins indicated to left. (**B**) Quantification of the interaction between NS2 and E2 expressed from virus and the C-p7 construct by densitometry. Immunoblots were inverted and a vertical line taken through the centre of the E2 bands, the average densities were recorded and the background subtracted. Normalised input values were multiplied by 20 to represent total input. IP efficiency was plotted as a ratio of IP: input E2. The numerical values are noted.

5.2.6 NS2 oligomerisation

To confirm previous observations that NS2 is able to dimerise (Dentzer *et al.*, 2009; Lorenz *et al.*, 2006b), Huh7 cells were co-transfected with full-length eGFP-tagged NS2 and full-length NS2-FLAG. IPs were carried out using GFP-Trap™ beads and the oligomerisation of NS2 was established by probing for the FLAG epitope (Figure 5.9A). Both N- and C-terminally eGFP-tagged NS2 were able to IP NS2-FLAG. The truncated form of NS2, tNS2, with a C-terminal FLAG-tag was detected in the IP fraction from eGFP-NS2 expressing cells but not NS2-eGFP. The apparent difference could be due to levels of tNS2-FLAG detected in the input fractions. The significance of the different expression levels of tNS2-FLAG was unclear and not explored further. This experiment also demonstrated that NS2-FLAG did not bind non-specifically to the beads and that the protein species detected with the FLAG antibody were specific to FLAG-tagged NS2 as they were not observed in samples from cells singly transfected with GFP-tagged NS2 (Figure 5.9A). Because all gels were run under reducing conditions and using ionic detergents it was not possible to distinguish between NS2 dimerisation and higher order oligomerisation. Therefore the interaction hereafter is referred to as NS2 oligomerisation.

Previous studies have shown that the catalytic domain of NS2 was capable of oligomerising independently of the *trans*-membrane domain (Lorenz *et al.*, 2006b). In order to establish whether the *trans*-membrane of NS2 is also capable of forming intermolecular interactions, Huh7 cells were co-transfected with combinations of full-length NS2, the *trans*-membrane domain alone and the catalytic domain alone (Figure 5.9B). As both conformations of eGFP fusions to NS2 were shown to interact with NS2-FLAG the experiment was simplified by the omission of the NS2-eGFP fusions. Furthermore, in an attempt to improve the detection of co-immunoprecipitated FLAG-tagged NS2 co-transfections were carried out with 0.25µg of GFP-fusion DNA and 0.75µg of FLAG-fusion DNA.

NS2-FLAG was again immunoprecipitated by eGFP-NS2 demonstrating the reproducibility of the interaction (Figure 5.9B), but it was not immunoprecipitated by eGFP alone or NS4B-eGFP. Surprisingly, immunoprecipitation of the FLAG-tagged *trans*-membrane domain was observed by the corresponding GFP-tagged domain suggesting that these two domains oligomerise. It was interesting to observe that the GFP-tagged *trans*-membrane domain was not able to IP the full-length NS2-FLAG fusion. Contrary to previous reports, no interaction was noted between the catalytic domain i.e. eGFP (93-217), and either full-length NS2-FLAG or the FLAG-tagged catalytic domain, (93-217) FLAG (Figure 5.9B - lane 7 and 8).

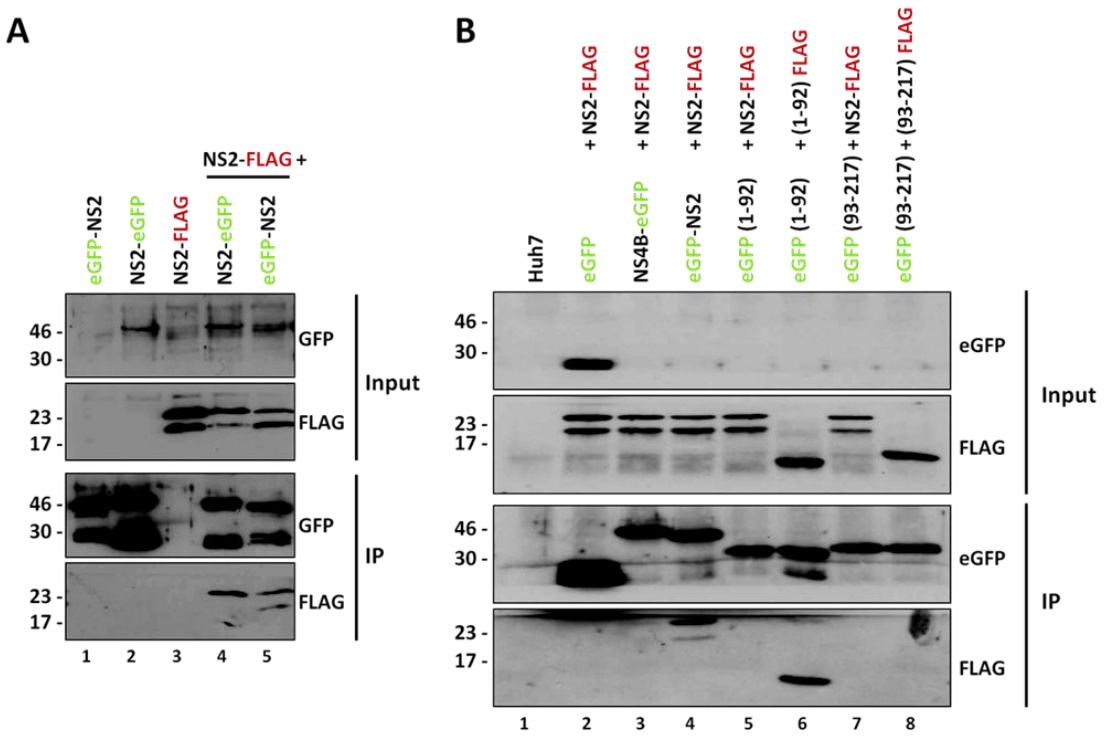


Figure 5.9 NS2 oligomerisation.

(**A**) To investigate whether NS2 was able to oligomerise, immunoprecipitations were carried on lysates from Huh7 cells transfected with full-length NS2 FLAG, eGFP-NS2 and NS2-eGFP. (**B**) To further investigate NS2 oligomerisation combinations of the tagged full-length NS2, the *trans*-membrane domain alone and the catalytic domain alone were co-transfected into cells and immunoprecipitations performed. Cells were harvested 48 hours post-transfection and lysed on ice for 30 minutes in TBS containing 1% Triton X-100. 5% of total input (Input) was kept aside and added to 1 volume of 2x protein sample buffer. Remaining lysates were diluted into four volumes of TBS, clarified at 20,100 x *g* for 10 minutes at 4°C and the resulting supernatants incubated O/N at 4°C with GFP-Trap™ beads. Beads were pelleted and washed three times in TBS. Immunoprecipitated proteins (IP) were harvested by heating the beads at 95°C in 2 x protein sample buffer for 3 minutes. Total IP fraction and 5% of total input were analysed by SDS PAGE and transferred to PVDF membranes. For clarity eGFP (green) and FLAG (red) tags in sample names have been coloured. Molecular weight markers are shown to the right and lane numbers are noted below. Membranes were immunoblotted for the epitope tags indicated to left.

An additional protein species detectable with the GFP antibody in the IP fraction of cells transfected with eGFP (1-92) + (1-92) FLAG was presumed to be 'free' eGFP, although it was not clear why this was restricted to this sample and not detected in the eGFP (1-92) + (1-217) FLAG sample. It could be that stable dimerisation of the NS2 *trans*-membrane domain altered the stability of the eGFP (1-92) fusion making it more susceptible to cleavage by cells proteases.

Together these data suggest that NS2 does indeed oligomerise and that the *trans*-membrane domain may play a significant role in forming intermolecular interactions.

5.3 Discussion

NS2 has been reported to interact with the majority of the other HCV proteins and also to dimerise within the catalytic domain. The X-ray crystal structure of the catalytic domain identified the potential dimer interface and allowed the targeted mutation of conserved residues about this face (Dentzer *et al.*, 2009), but it did not elucidate the role of the *trans*-membrane domain in dimerisation. Most studies looking at the interaction between NS2 and other viral proteins have identified residues key to maintaining the interactions. Immunoprecipitation studies were undertaken to identify the E2 binding domain of NS2 and further characterise NS2 dimerisation.

5.3.1 E2 interacts with NS2 expressed in *trans*

The observation that NS2 and E2 interact in the context of the full-length virus was confirmed here as GFP-tagged NS2 expressed in cells containing the full-length replicating virus genome was able to specifically co-IP E2 (Figure 5.7 and 5.8). This interaction was also confirmed when E2 and NS2 were expressed in *trans* from DNA vectors (Figure 5.2B). However, despite the published observation of NS2 interacting with NS3 and NS5A (Jirasko *et al.*, 2010; Ma *et al.*, 2011; Stapleford & Lindenbach, 2011), these interactions could not be detected in cells containing transiently replicating full-length virus RNA or cells containing stably or transiently replicating sub-genomic sequence of HCV RNA (Figure 5.6B and 5.7A). Indeed eGFP-NS2 was only able to co-IP a small population of E2 from virus containing cells compared with E2 expressed from the C-p7 construct (Figure 5.7B).

Experiments demonstrated that NS2, E2 and numerous other viral proteins aggregate in high molecular weight complexes (Stapleford & Lindenbach, 2011) which may form shortly after translation and could exclude other proteins or further sequester E2 to an exclusive environment. Furthermore, as NS2 has been shown to interact with numerous other HCV proteins, it is reasonable to speculate that the eGFP-NS2 molecules could have been interacting with E2 and other non-structural proteins that were not included in the co-transfection system. NS2 has been proposed to localise proximally to proposed sites of genome replication (Tedbury *et al.*, 2011) and replication complexes extracted from NS3-5B replicon cells have been found to be protease and nuclease resistant (Aizaki *et al.*, 2000; Yang *et al.*, 2004). It has also been shown that when NS2 is expressed in *cis* with p7 it localises to a detergent-insoluble compartment (Tedbury *et al.*, 2011). It is therefore possible that NS2 expressed in *trans* was unable to access these replication complexes, or that E2 was localised

to detergent-insoluble complexes by NS2 expressed as part of the viral polyprotein (vNS2) which would have been removed from lysates prior to incubation with GFP-Trap™ beads.

Proximity following translation could be vital to the establishment of the E2:NS2 interaction as the introduction of a second IRES between E2 and p7 resulted in a reduction in the amount of E2 immunoprecipitated by NS2 (Stapleford & Lindenbach, 2011). Likewise, introduction of a second IRES between NS2 and NS3 impaired the NS2:NS3 interaction by co-IP (Stapleford & Lindenbach, 2011). Both of these bi-cistronic viruses had reduced viral titres by 1.5-2 log10. This suggests that the spatial separation afforded by translation in *trans* could have been detrimental for the ability of NS2 to interact with other viral proteins and thereby impede its role in assembly. Therefore, one explanation for the apparent reduced affinity of eGFP-NS2 for E2 in the context of the fully infectious virus system is that vNS2 was competing for E2 and, as this species would have been expressed proximally to the E2 molecules, vNS2 could have sequestered the majority of the E2 population.

5.3.2 The E2-binding domain of NS2

It was hypothesised that the *trans*-membrane domain of NS2 interacts with E2, as early truncation analysis of NS2 and E2 found that the N-terminal 96 residues of NS2 co-immunoprecipitated with E2 using an antibody specific for E2 (Selby *et al.*, 1994). As the catalytic domain is predicted to orientate to the cytosol and contain no membrane-spanning domains it was believed that it would not interact with E2 (Lorenz *et al.*, 2006b). Consistent with these hypotheses; the *trans*-membrane domain (1-92) was able to interact with E2 and the catalytic domain (93-217) was not, irrespective of the termini to which the eGFP reporter was fused (Figure 5.4 and 5.5). Compensatory mutation analysis of NS2 defective viruses suggested that the putative third TMD (aa 73-93) of NS2 interacts with E1 and E2 (Jirasko *et al.*, 2010). Furthermore, Popescu *et al*., proposed the main E2 interacting domain to be the second putative TMD of NS2 (aa 31-49) (Popescu *et al.*, 2011). However, data provided in this chapter demonstrated that neither of these domains was capable of interacting with E2 in isolation as the truncations (25-50) and (51-92), which encompasses the predicted second and third TMDs, did not IP E2, regardless of the position of the eGFP fusion molecule (Figure 5.5).

Analysis of those fusions that were able to co-IP E2 found that they did not encode a common domain although they all encoded two of the putative TMDs in all combinations (Figure 5.10). As E2 and NS2 are *trans*-membrane proteins it was feasible that their interaction may be dependent upon membrane association. A table was generated summarising the membrane targeting and E2 interaction phenotypes of the two eGFP fusion sets (Table 5.1). A positive

correlation was noted between the ability of a protein to target to membranes and to interact with E2. Only (Δ25-50) eGFP and (Δ25-92) eGFP did not fit this trend as (Δ25-50) eGFP was able to target to membranes but not co-IP while (Δ25-92) eGFP was unable to target to membranes but was able to co-IP E2 (Table 5.1). This strongly suggests a link between NS2 membrane association and E2 interaction. It could be interesting to investigate the NS2:E2 interaction *in vitro,* expressed in the presence and absence of purified membranes, to determine the dependency of the interaction on membranes.

The major difference between the co-IP-profiles of the two NS2 GFP-fusion sets was that (1-50) eGFP did not co-IP E2 whereas eGFP (1-50) efficiently bound E2 (Figure 5.6). Comparison with the membrane targeting data highlighted that these fusions presented opposing membrane-targeting abilities with (1-50) eGFP also unable to target to membranes while eGFP (1-50) was detected exclusively in the membrane fraction (Table 5.1; Figure 4.6 and 4.8). This is consistent with the hypothesis that the NS2:E2 interaction requires NS2 to be membrane-associated.

The ability of (Δ25-92) eGFP to interact with membranes but not E2 suggests that, despite eGFP (Δ25-92) showing a weak interaction with E2, the eGFP reporter forming part of the (Δ25-92) eGFP fusion disrupted the NS2:E2 interaction. The phenotype of (Δ25-50) eGFP is more intriguing as it was unable to target to membranes when expressed in isolation but still efficiently interacted with E2 (Table 5.1). It is possible that this fusion, when co-expressed with core-p7 was able form protein interactions with p7 that enabled the truncated NS2 fusion to associate with the membrane and thus interact with E2. Re-examination of the membrane-targeting of this (Δ25-50) eGFP from co-transfected cells would help to clarify this.

5.3.3 Requirements of the E2/NS2 interaction

Whether or not the E2:NS2 interaction was dependent upon core, E1 or p7 could not be addressed using systems employed herein and would have required the generation of additional expression constructs. Despite this it is unlikely that the E2:NS2 interaction requires the presence of core as it did not interact with any of the eGFP-NS2 fusions when it was expressed from a DNA construct (Figure 5.4) or from viral RNA (Figure 5.6B). Stapleford and Lindenbach (Stapleford & Lindenbach, 2011) demonstrated that E1 and E2 could be co-immunoprecipitated by tagged NS2, yet mutations that disrupted the E1:NS2 interaction did not alter the E2:NS2 interaction. Therefore it is unlikely that E1 is necessary for stabilisation of the E2:NS2 interaction.

NS2 region	Terminus of eGFP tag	Target to Membranes	Co-IP E2
1-217	C	Y	Y
	N	Y	Y
25-217	C	Y	Y
	N	Y	Y
51-217	C	-	-
	N	Y	Y/N
93-217	C	N	N
	N	-	N
1-92	C	-	Y
	N	Y	Y
1-50	C	N	N
	N	Y	Y
25-92	C	Mostly N	Y/N
	N	Mostly N	Y/N
1-24	C	Y/N	N
	N	Y/N	N
25-50	C	N	N
	N	Y/N	N
51-92	C	N	-
	N	Y/N	N
Δ25-50	C	N	Y
	N	Y	Y
Δ25-92	C	Y	N
	N	-	Y/N
Δ51-92	C	-	N
	N	-	N

Table 5.1 Summary of the membrane targeting and E2 interaction of the eGFP fusions.

The fusions were either able to (Yes – Y, green) or unable (No – N, red) to target to membranes and interact with E2. Fusions that were recovered in both the membrane and soluble fractions, or displayed only a weak interaction with E2, are shown in yellow (Y/N). As the weak E2 interactions between eGFP (25-50) and eGFP (51-92) were not reproducible they are not included. (-) denotes peptides where no phenotype could be assigned. The position of the eGFP tag with respect to the NS2 sequence is noted. eGFP-NS2 fusions, N; NS2-eGFP fusions, C.

As highlighted in the introduction to this section, the role of p7 in mediating NS2:protein interactions is the subject of current debate. When Jirasko *et al.,* (Jirasko *et al.*, 2010) swapped the sequence encoding the first putative TMD from JFH-1 with that of the Con1 coding sequence, in the JFH-1 virus, the resulting virus showed impaired intra- and extracellular infectivity and reduced interactions with E2, p7, NS3 and NS5A. However, repeated passage in cell culture led to the selection and identification of the pseudo-revertant mutation E3D in p7. This pseudo-revertant mutation restored NS2 interactions with E2, NS3 and NS5A (p7 was undetectable due to mutation of the antibody recognition epitope) and significantly improved infectivity (Jirasko *et al.*, 2010). This implied that the NS2:E2 interaction may be dependent upon the pre-formation of a functional p7-NS2 interaction via the respective N-terminal domains. It had been hoped to address the point of p7 involvement in NS2:NS3 and NS2:NS5A interaction by comparing two replicon systems with and without p7, however no NS2 interactions were detected using these systems (Figure 5.7B)

The topology of NS2 does not appear to play a role in determining the NS2:E2 interaction as the (Δ25-50) and (25-217) truncations could be expected to assume significantly altered topologies compared to the full-length protein (Figure 5.10A). Based upon the SEAP fusion S7 (aa 1-52), which localised the C-terminus to the cytosol and therefore inferred to form a single TMD, eGFP (1-50) may form a single TMD topology or an incomplete hairpin topology (Figure 5.10A). Any fusions of domains (1-24) and (25-50) did not target efficiently to membranes, whereas eGFP (1-50) did (Table 5.1). It is therefore possible that hydrophilic residues within the first and second putative TMDs (Figure 5.10B) require the formation of stabilising salt bridges to facilitate membrane insertion. Whether these interactions occur between the first and second TMDs of the same molecule or other molecules is unclear although it is clear that the *trans*-membrane domain is cpapable of forming inter-molecular interactions (Figure 5.9B). As (1-50) eGFP did not associate with membranes this adds to the evidence that residue 50 is within the membrane. It is possible that fusion of eGFP to this residue prevented insertion of the partial second TMD which inturn prevented the formation of stabilising interactions with the putative first TMD resulting in the soluble phenotype observed for this construct [(1-50) eGFP].

While the topology of (Δ25-50) and (25-217) can only be speculated upon, the deletions were predicted to remove a single TMD from each of them resulting in a opposed localisation for the catalytic domain for these fusions compared to the full-length protein. Although the NS2 truncations capable of targeting to membranes appeared to be those most able to interact

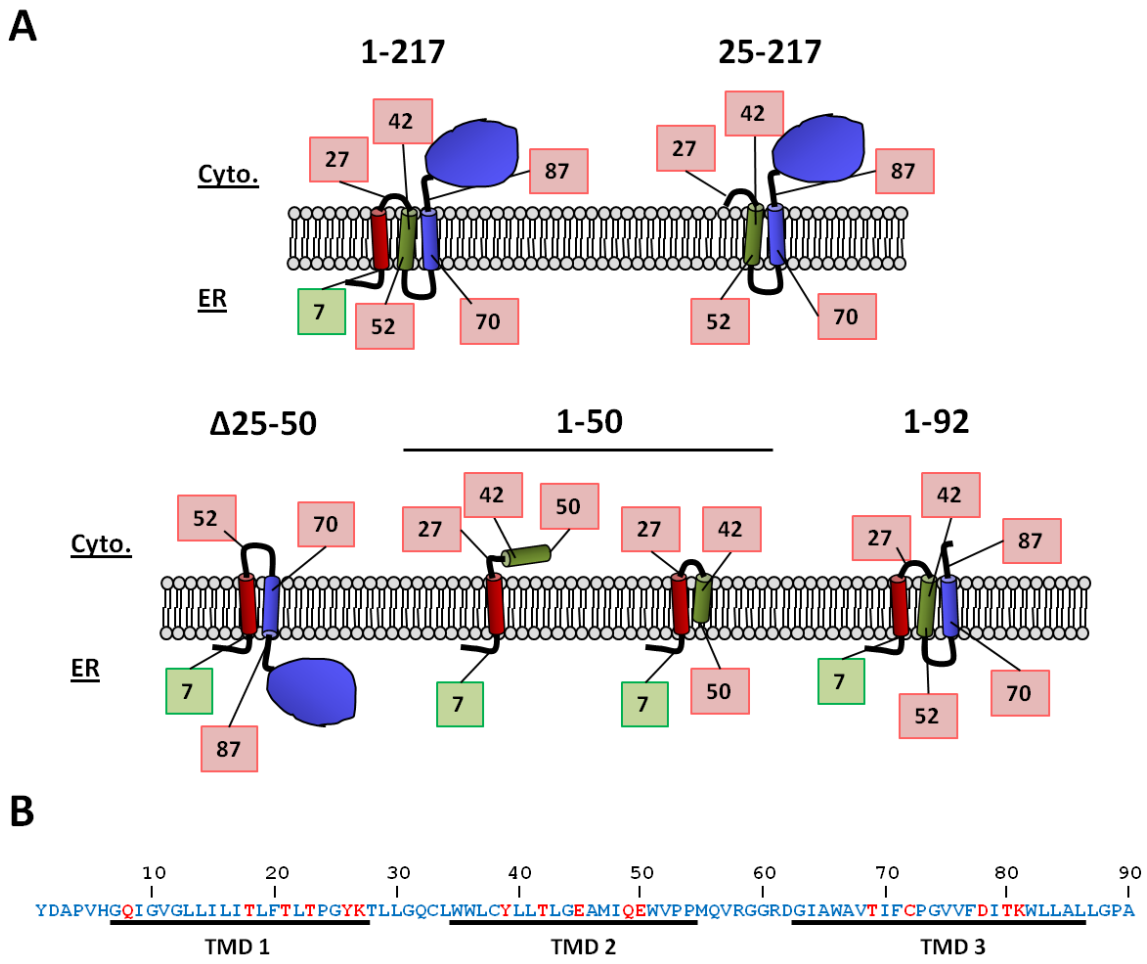


Figure 5.10 Predicted membrane topologies of the truncations able to efficiently interact with E2.

(**A**) Models based upon the truncation analysis data and the consensus topology prediction presented in Chapter 3. The positions of the C-terminal truncations used in Chapter 3 (red boxes) and the predicted start of the first TMD (green boxes) are noted. The putative TMDs are coloured for comparison; first TMD (red), second TMD (green) and third (blue). (**B**) JFH-1 NS2 amino acid sequence. Hydrophilic residues (red) within the predicted TMDs (underlined) are highlighted.

with E2, membranes were not required for the perpetuation of the interaction as the lysis conditions of 1% Triton X-100 should solubilise the majority of membranes, suggesting the formation of a stable protein-protein interaction rather than independent co-localisation to a membrane compartment.

5.3.4 Oligomerisation of the NS2 *trans*-membrane domain

Bimolecular cleavage of NS2/3 precursors led to the hypothesis that NS2 molecules may interact to facilitate autocleavage (Grakoui, 1993b; Lorenz *et al.*, 2006b; Reed *et al.*, 1995).

Evidence of NS2 dimerisation has been presented by blue native (BN)-PAGE analysis of cell lysates (Lorenz *et al.*, 2006b; Stapleford & Lindenbach, 2011) and crystallisation of the isolated catalytic domain as a dimer (Lorenz *et al.*, 2006b). Cleavage of the NS2/3 junction is essential for genome replication (Jones *et al.*, 2007). Mutation of three residues predicted to be at the dimer interface, based on the crystal structure, did not affect replication but severely reduced infectious virus production, although biochemical confirmation of a concomitant disruption of NS2 dimerisation was not provided (Dentzer *et al.*, 2009). This suggested that dimerisation could be more important for assembly than NS2/3 cleavage and may be consistent with observations from mixing experiments using two differently cleavage-defective precursors that showed bi-molecular cleavage to be very inefficient (Grakoui, 1993b; Lorenz *et al.*, 2006b; Reed *et al.*, 1995).

Co-IP of FLAG-tagged full-length NS2 by eGFP-NS2 (Figure 5.9) was consistent with the previous biochemical evidence indicative of NS2 dimerisation (Lorenz *et al.*, 2006b). However, observations that the catalytic domain was capable of dimerisation were not reproduced here as eGFP (93-217) was not able to co-IP either NS2-FLAG or (93-217) FLAG (Figure 5.9B). Conversely, the *trans*-membrane domain was able to co-IP the corresponding FLAG-tagged catalytic domain fusion (Figure 5.9B). The reason for this apparent discrepancy was not immediately clear as both this study and Lorenz *et al.,* (Lorenz *et al.*, 2006b) used co-IP of differentially tagged-NS2 expressed in cells to assess oligomerisation. Although the crystal structure of the catalytic domain included residues 93-217 of NS2, subsequent co-IP experiments utilised differentially tagged NS2/3 consisted of residues 19–217 of NS2 and 1–181 of NS3 (Lorenz *et al.*, 2006b). Therefore it is possible that the oligomerisation observed for the catalytic domain was enhanced by the homogeneity and concentration of the purified NS2 catalytic domain and that NS2 oligomerisation is primarily coordinated by the N-terminal residues (19-92). The co-IP of tNS2-FLAG by eGFP-NS2 (Figure 5.9) is consistent with this hypothesis as the truncation point of this peptide appears to be in the vicinity of residue 25

(Figure 4.8). The NS2 truncation sets described in Chapter 4 represent the ideal tools to investigate this more thoroughly.

5.4 Conclusions

Co-immunoprecipitation studies presented here reveal that NS2 oligomerises via its *trans*-membrane domain and this region is also responsible for interacting with E2. Consistent with previous findings regarding the NS2:NS3 and NS2:p7 interactions; the NS2:E2 interaction also appears to be genotype independent.

Truncation analysis suggested that NS2 forms multiple interactions with E2 via the putative second and third TMDs, and to a lesser extent the putative first TMD, but that the catalytic domain is dispensable. By comparing with membrane-targeting data presented in the previous chapter, it is proposed that the NS2:E2 interaction is dependent upon membrane targeting of NS2 and the two proteins are likely to form intra-membrane contacts. The ultimate membrane topology of NS2 does not appear to be pivotal for the NS2:E2 interaction.

CHAPTER 6: FINAL DISCUSSION AND FUTURE PERSPECTIVES

The data presented here demonstrate that NS2 assuming a three TMD topology. C-terminal truncation analysis demonstrated that aa 1-52 are capable of forming one TMD, aa 1-70 form two TMDs and aa 1-87 form three TMDs (Figure 3.17 **Model E**). Analysing protein topology using C-terminal truncations introduces the possibility of disruption to long-range intra-molecular interactions which are required for achieving the native topology. Therefore it is possible that in full-length NS2 aa 1-52 do form two TMDs but that stabilising interactions with upstream residues (aa 53-70) are required in order to achieve this topology. However, the ability of residues 25-50, expressed with no additional NS2 sequence, to partially target to and insert into membranes (Figure 4.15) is suggestive that an absence of long-range interactions was not responsible for the phenotype of the S7 fusion (aa 1-52 of NS2 fused to SEAP). Taken together the truncation and membrane dissociation data suggest that residues 1-52 contain all of the first TMD and a significant portion of the second TMD of NS2, but not all of it. This led to the description of a refined model, **Model F**, of NS2 topology shown in (Figure 6.1). Although this model has been designed in part by experimental observations, the exact positions of the TMDs remain to be precisely mapped. As all the experiments carried out here were done using JFH-1 protein sequence this model represents a platform for confirmation of these findings within the full-length system and a more detailed mapping of NS2 topology.

The extent to which NS2 topology is required for the role of NS2 in virus production is not clear. C-terminal truncation analysis of NS2 is not a valid approach to explore this in full-length virus as deletion of the protease domain (aa 94-217) in the context of a bi-cistronic virus, in which NS2 and NS3 were separated by a second IRES, was lethal to the virus (Jirasko *et al.*, 2008; Jones *et al.*, 2007). Indeed, more recently it has been shown that deletions of aa 10- 62 or aa 32-96 in a bi-cistronic virus were also lethal to virus production. Investigation of NS2 topology by means of epitope or reporter-enzyme insertion in a bi-cistronic JFH-1 virus would also provide a means of validating that protein function was altered by introduction of the epitope/reporter.

Experiments presented in Chapter 4 verified *in vitro* data describing the presence of multiple membrane-targeting within NS2, including one within the protease domain (Yamaga & Ou, 2002). However, it is possible that some of the truncations recovered in the membrane fraction during the membrane-targeting experiments were forming aggregates rather than direct associations with membranes. Although it was not feasible to do so initially with such a large panel of fusions, re-examination of those constructs targeting to membranes by membrane flotation analysis would circumvent this problem. Co-immunoprecipitation studies

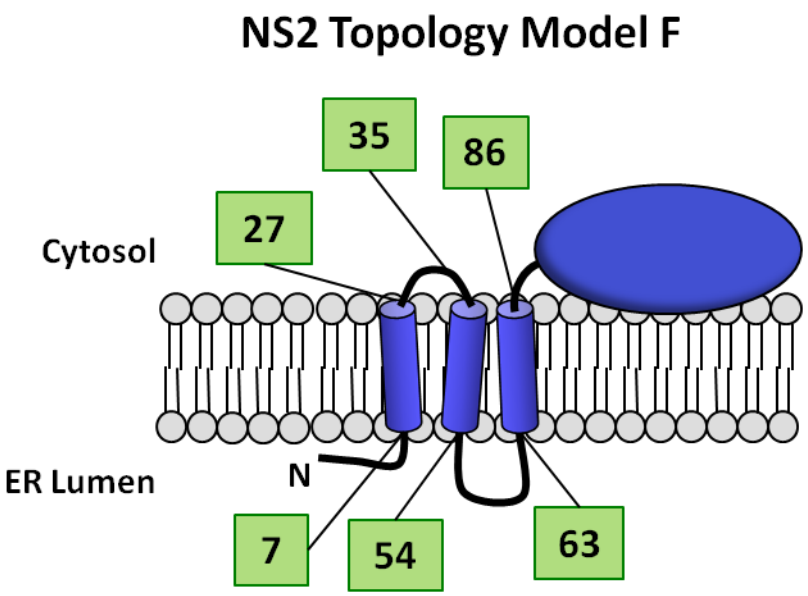


Figure 6.1 NS2 topology model.

NS2 is shown modelled on to a lipid bi-layer. The N terminus is shown oriented to the ER lumen. The catalytic domain is shown oriented to the cytosol and in peripheral association with the membrane in accordance with membrane-targeting data. The predicted boundaries of the TMDs are noted (green boxes).

showed that NS2 and E2 expressed from separate DNA molecules were able to stably interact. As membrane association of NS2 was implicated in the NS2:E2 interaction, the membrane-targeting of those fusions able to interact with E2 by membrane flotation should be confirmed. Co-expression of eGFP-NS2 and core-p7 *in vitro* in the presence and absence of microsomal membranes could also help to clarify the requirement of membranes for formation of the NS2:E2 interaction.

Although time constraints and an inability to detect p7 in these systems prevented the investigation of the role of p7 in mediating the NS2:E2 interaction, repetition of the experiments described in chapter 5 with a core-E2, rather than core-p7, expression construct would provide a simple means of testing this. Mutation of the p7 basic loop was found to abrogate the NS2:E2 interaction leading to speculation that p7 ion channel function is specifically required for the interaction. However, as mutation of the basic loop may also alter p7 topology (Ma *et al.*, 2011) and its ability to insert into membranes (StGelais *et al.*, 2007); treatment of cells co-transfected with eGFP-NS2 and core-p7 with the drug rimantidine, which blocks p7 ion channel function (Griffin, 2010; Griffin *et al.*, 2008; Luik *et al.*, 2009), presents a specific means of investigating p7 ion channel function in relation to the NS2:E2 interaction.

It is not clear whether the formation of physical interactions between NS2 and other viral proteins are required for the role of NS2 in virion morphogenesis. If the NS2:E2 interaction is found to be required for this function, the co-IP system described here provides a tool for the testing of compounds that disrupt the NS2:E2 interaction. As genotype 2a NS2 was able to efficiently interact with E2 from both genotype 2a and 1b, it is reasonable to assume that compounds that disrupt the NS2:E2 interaction may be cross-genotypic. Chemical disruption of NS2 dimerisation would also help to elucidate the role of dimerisation in cleavage of the NS2/3 junction and virion morphogenesis.

CHAPTER 7: REFERENCES

Abe, K., Inchauspe, G., Shikata, T. & Prince, A. M. (1992). Three different patterns of hepatitis C virus infection in chimpanzees. *Hepatology* **15**, 690-695.

Abid, K., Pazienza, V., de Gottardi, A., Rubbia-Brandt, L., Conne, B., Pugnale, P., Rossi, C., Mangia, A. & Negro, F. (2005). An in vitro model of hepatitis C virus genotype 3a-associated triglycerides accumulation. *J Hepatol* **42**, 744-751.

Adair, R., Patel, A. H., Corless, L., Griffin, S., Rowlands, D. J. & McCormick, C. J. (2009). Expression of hepatitis C virus (HCV) structural proteins in trans facilitates encapsidation and transmission of HCV subgenomic RNA. *J Gen Virol* **90**, 833-842.

Agapov, E. V., Murray, C. L., Frolov, I., Qu, L., Myers, T. M. & Rice, C. M. (2004). Uncleaved NS2-3 is required for production of infectious bovine viral diarrhea virus. *J Virol* **78**, 2414-2425.

Aizaki, H., Saito, S., Ogino, T., Miyajima, N., Harada, T., Matsuura, Y., Miyamura, T. & Kohase, M. (2000). Suppression of interferon-induced antiviral activity in cells expressing hepatitis C virus proteins. *J Interferon Cytokine Res* **20**, 1111-1120.

Alexeyev, M. F. & Winkler, H. H. (1999). Membrane topology of the Rickettsia prowazekii ATP/ADP translocase revealed by novel dual pho-lac reporters. *Journal of Molecular Biology* **285**, 1503.

Alexeyev, M. F. & Winkler, H. H. (2002). Transposable dual reporters for studying the structure-function relationships in membrane proteins: permissive sites in R. prowazekii ATP/ADP translocase. *Biochemistry* **41**, 406-414.

Alter, H. J., Purcell, R. H., Holland, P. V. & Popper, H. (1978). Transmissible agent in non-A, non-B hepatitis. *Lancet* **1**, 459-463.

Alter, M. J. (2006). Epidemiology of viral hepatitis and HIV co-infection. *J Hepatol* **44**, S6-9.

Aoubala, M., Holt, J., Clegg, R. A., Rowlands, D. J. & 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**, 1637-1646.

Appel, N., Zayas, M., Miller, S., Krijnse-Locker, J., Schaller, T., Friebe, P., Kallis, S., Engel, U. & Bartenschlager, R. (2008). Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog* **4**, e1000035.

Arai, M., Mitsuke, H., Ikeda, M., Xia, J., Kikuchi, T., Satake, M. & Shimizu, T. (2004). ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability.

Bamberger, M. J. & Lane, M. D. (1990). Possible role of the Golgi apparatus in the assembly of very low density lipoprotein. *Proc Natl Acad Sci U S A* **87**, 2390-2394.

Barba, G., Harper, F., Harada, T., Kohara, M., Goulinet, S., Matsuura, Y., Eder, G., Schaff, Z., Chapman, M. J., Miyamura, T. & Brechot, C. (1997). Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. *Proc Natl Acad Sci U S A* **94**, 1200-1205.

Bartenschlager, R. & Lohmann, V. (2000). Replication of hepatitis C virus. *J Gen Virol* **81**, 1631-1648.

Bartenschlager, R., Penin, F., Lohmann, V. & Andre, P. (2011). Assembly of infectious hepatitis C virus particles. *Trends Microbiol* **19**, 95-103.

Bassett, S. E., Brasky, K. M. & Lanford, R. E. (1998). Analysis of hepatitis C virus-inoculated chimpanzees reveals unexpected clinical profiles. *J Virol* **72**, 2589-2599.

Beames, B., Chavez, D. & Lanford, R. E. (2001). GB virus B as a model for hepatitis C virus. *Ilar J* **42**, 152-160.

Behrens, S. E., Tomei, L. & De Francesco, R. (1996). Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J* **15**, 12-22.

Beran, R. K., Lindenbach, B. D. & Pyle, A. M. (2009). The NS4A protein of hepatitis C virus promotes RNA-coupled ATP hydrolysis by the NS3 helicase. *J Virol* **83**, 3268-3275.

Berger, J., Hauber, J., Hauber, R., Geiger, R. & Cullen, B. R. (1988). Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. *Gene* **66**, 1-10.

Bhakdi, S., Weller, U., Walev, I., Martin, E., Jonas, D. & Palmer, M. (1993). A guide to the use of pore-forming toxins for controlled permeabilization of cell membranes. *Med Microbiol Immunol* **182**, 167-175.

Bissig, K. D., Wieland, S. F., Tran, P., Isogawa, M., Le, T. T., Chisari, F. V. & Verma, I. M. (2010). Human liver chimeric mice provide a model for hepatitis B and C virus infection and treatment. *J Clin Invest* **120**, 924-930.

Blight, K. J., Kolykhalov, A. A. & Rice, C. M. (2000). Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972-1974.

Blight, K. J., McKeating, J. A. & Rice, C. M. (2002). Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol* **76**, 13001-13014.

Blight, K. J. & Rice, C. M. (1997). Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *J Virol* **71**, 7345-7352.

Blumberg, B. S., Alter, H. J. & Visnich, S. (1965). A "New" Antigen in Leukemia Sera. *Jama* **191**, 541-546.

Bogdanov, M., Zhang, W., Xie, J. & Dowhan, W. (2005). Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAMTM): Application to lipid-specific membrane protein topogenesis. *Methods* **36**, 148.

Boonstra, A., van der Laan, L. J., Vanwolleghem, T. & Janssen, H. L. (2009). Experimental models for hepatitis C viral infection. *Hepatology* **50**, 1646-1655.

Borel, A. C. & Simon, S. M. (1996). Biogenesis of polytopic membrane proteins: membrane segments assemble within translocation channels prior to membrane integration. *Cell* **85**, 379-389.

Bornemann, T., Jockel, J., Rodnina, M. V. & Wintermeyer, W. (2008). Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nat Struct Mol Biol* **15**, 494-499.

Boson, B., Granio, O., Bartenschlager, R. & Cosset, F. L. (2011). A concerted action of hepatitis C virus p7 and nonstructural protein 2 regulates core localization at the endoplasmic reticulum and virus assembly. *PLoS Pathog* **7**, e1002144.

Boulant, S., Douglas, M. W., Moody, L., Budkowska, A., Targett-Adams, P. & McLauchlan, J. (2008). Hepatitis C virus core protein induces lipid droplet redistribution in a microtubule- and dynein-dependent manner. *Traffic* **9**, 1268-1282.

Boulant, S., Montserret, R., Hope, R. G., Ratinier, M., Targett-Adams, P., Lavergne, J. P., Penin, F. & McLauchlan, J. (2006). Structural determinants that target the hepatitis C virus core protein to lipid droplets. *J Biol Chem* **281**, 22236-22247. Epub 22006 May 22216.

Boulant, S., Targett-Adams, P. & McLauchlan, J. (2007). Disrupting the association of hepatitis C virus core protein with lipid droplets correlates with a loss in production of infectious virus. *J Gen Virol* **88**, 2204-2213.

Boulant, S., Vanbelle, C., Ebel, C., Penin, F. & Lavergne, J. P. (2005). Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *J Virol* **79**, 11353-11365.

Bowen, D. G. & Walker, C. M. (2005). Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature* **436**, 946-952.

Bradley, D. W., Maynard, J. E., Popper, H., Cook, E. H., Ebert, J. W., McCaustland, K. A., Schable, C. A. & Fields, H. A. (1983). Posttransfusion non-A, non-B hepatitis: physicochemical properties of two distinct agents. *J Infect Dis* **148**, 254-265.

Bradley, D. W., McCaustland, K. A., Cook, E. H., Schable, C. A., Ebert, J. W. & Maynard, J. E. (1985). Posttransfusion non-A, non-B hepatitis in chimpanzees. Physicochemical

evidence that the tubule-forming agent is a small, enveloped virus. *Gastroenterology* **88**, 773-779.

Brass, V., Berke, J. M., Montserret, R., Blum, H. E., Penin, F. & Moradpour, D. (2008). Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. *Proc Natl Acad Sci U S A* **105**, 14545-14550.

Brass, V., Bieck, E., Montserret, R., Wolk, B., Hellings, J. A., Blum, H. E., Penin, F. & Moradpour, D. (2002). An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J Biol Chem* **277**, 8130-8139. Epub 2001 Dec 8114.

Brass, V., Gouttenoire, J., Wahl, A., Pal, Z., Blum, H. E., Penin, F. & Moradpour, D. (2010). Hepatitis C virus RNA replication requires a conserved structural motif within the transmembrane domain of the NS5B RNA-dependent RNA polymerase. *J Virol* **84**, 11580-11584.

Brass, V., Pal, Z., Sapay, N., Deleage, G., Blum, H. E., Penin, F. & Moradpour, D. (2007). Conserved determinants for membrane association of nonstructural protein 5A from hepatitis C virus and related viruses. *J Virol* **81**, 2745-2757.

Brillet, R., Penin, F., Hezode, C., Chouteau, P., Dhumeaux, D. & Pawlotsky, J. M. (2007). The nonstructural 5A protein of hepatitis C virus genotype 1b does not contain an interferon sensitivity-determining region. *J Infect Dis* **195**, 432-441.

Brohm, C., Steinmann, E., Friesland, M., Lorenz, I. C., Patel, A., Penin, F., Bartenschlager, R. & Pietschmann, T. (2009). Characterization of determinants important for hepatitis C virus p7 function in morphogenesis by using trans-complementation. *J Virol* **83**, 11682-11693.

Brown, E. A., Zhang, H., Ping, L. H. & Lemon, S. M. (1992). Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* **20**, 5041-5045.

Brown, K. S., Keogh, M. J., Owsianka, A. M., Adair, R., Patel, A. H., Arnold, J. N., Ball, J. K., Sim, R. B., Tarr, A. W. & Hickling, T. P. (2010). Specific interaction of hepatitis C virus glycoproteins with mannan binding lectin inhibits virus entry. *Protein Cell* **1**, 664-674.

Bukh, J., Apgar, C. L., Govindarajan, S. & Purcell, R. H. (2001). Host range studies of GB virus-B hepatitis agent, the closest relative of hepatitis C virus, in New World monkeys and chimpanzees. *J Med Virol* **65**, 694-697.

Carrere-Kremer, S., Montpellier-Pala, C., Cocquerel, L., Wychowski, C., Penin, F. & Dubuisson, J. (2002). Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. *J Virol* **76**, 3720-3730.

Carroll, S. S., Ludmerer, S., Handt, L., Koeplinger, K., Zhang, N. R., Graham, D., Davies, M. E., MacCoss, M., Hazuda, D. & Olsen, D. B. (2009). Robust antiviral efficacy upon administration of a nucleoside analog to hepatitis C virus-infected chimpanzees. *Antimicrob Agents Chemother* **53**, 926-934.

Chen, C. M., He, Y., Lu, L., Lim, H. B., Tripathi, R. L., Middleton, T., Hernandez, L. E., Beno, D. W., Long, M. A., Kati, W. M., Bosse, T. D., Larson, D. P., Wagner, R., Lanford, R. E., Kohlbrenner, W. E., Kempf, D. J., Pilot-Matias, T. J. & Molla, A. (2007). Activity of a potent hepatitis C virus polymerase inhibitor in the chimpanzee model. *Antimicrob Agents Chemother* **51**, 4290-4296.

Cheng, G., Montero, A., Gastaminza, P., Whitten-Bauer, C., Wieland, S. F., Isogawa, M., Fredericksen, B., Selvarajah, S., Gallay, P. A., Ghadiri, M. R. & Chisari, F. V. (2008). A virocidal amphipathic {alpha}-helical peptide that inhibits hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* **105**, 3088-3093.

Chevaliez, S., Bouvier-Alias, M., Brillet, R. & Pawlotsky, J. M. (2007). Overestimation and underestimation of hepatitis C virus RNA levels in a widely used real-time polymerase chain reaction-based method. *Hepatology* **46**, 22-31.

Chiaramonte, M., Stroffolini, T., Vian, A., Stazi, M. A., Floreani, A., Lorenzoni, U., Lobello, S., Farinati, F. & Naccarato, R. (1999). Rate of incidence of hepatocellular carcinoma in patients with compensated viral cirrhosis. *Cancer* **85**, 2132-2137.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (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. & Bradley, D. W. (1990). Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. *Br Med Bull* **46**, 423-441.

Choukhi, A., Ung, S., Wychowski, C. & Dubuisson, J. (1998). Involvement of endoplasmic reticulum chaperones in the folding of hepatitis C virus glycoproteins. *J Virol* **72**, 3851-3858.

Ciczora, Y., Callens, N., Penin, F., Pecheur, E. I. & Dubuisson, J. (2007). Transmembrane domains of hepatitis C virus envelope glycoproteins: residues involved in E1E2 heterodimerization and involvement of these domains in virus entry. *J Virol* **81**, 2372-2381.

Clarke, D., Griffin, S., Beales, L., Gelais, C. S., Burgess, S., Harris, M. & Rowlands, D. (2006). Evidence for the formation of a heptameric ion channel complex by the hepatitis C virus p7 protein in vitro. *J Biol Chem* **281**, 37057-37068. Epub 32006 Oct 37010.

Cocquerel, L., Duvet, S., Meunier, J. C., Pillez, A., Cacan, R., Wychowski, C. & Dubuisson, J. (1999). The transmembrane domain of hepatitis C virus glycoprotein E1 is a signal for static retention in the endoplasmic reticulum. *J Virol* **73**, 2641-2649.

Cocquerel, L., Meunier, J. C., Pillez, A., Wychowski, C. & Dubuisson, J. (1998). A retention signal necessary and sufficient for endoplasmic reticulum localization maps to the transmembrane domain of hepatitis C virus glycoprotein E2. *J Virol* **72**, 2183-2191.

Cocquerel, L., Op de Beeck, A., Lambot, M., Roussel, J., Delgrange, D., Pillez, A., Wychowski, C., Penin, F. & Dubuisson, J. (2002). Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. *Embo J* **21**, 2893-2902.

Cocquerel, L., Wychowski, C., Minner, F., Penin, F. & Dubuisson, J. (2000). Charged residues in the transmembrane domains of hepatitis C virus glycoproteins play a major role in the processing, subcellular localization, and assembly of these envelope proteins. *J Virol* **74**, 3623-3633.

Cooper, S., Erickson, A. L., Adams, E. J., Kansopon, J., Weiner, A. J., Chien, D. Y., Houghton, M., Parham, P. & Walker, C. M. (1999). Analysis of a successful immune response against hepatitis C virus. *Immunity* **10**, 439-449.

Corless, L., Crump, C. M., Griffin, S. D. & Harris, M. (2009). Vps4 and the ESCRT-III complex are required for the release of infectious hepatitis C virus particles. *J Gen Virol* **91**, 362-372.

Crotty, S., Maag, D., Arnold, J. J., Zhong, W., Lau, J. Y., Hong, Z., Andino, R. & Cameron, C. E. (2000). The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat Med* **6**, 1375-1379.

Dalagiorgou, G., Vassilaki, N., Foka, P., Boumlic, A., Kakkanas, A., Kochlios, E., Khalili, S., Aslanoglou, E., Veletza, S., Orfanoudakis, G., Vassilopoulos, D., Hadziyannis, S. J., Koskinas, J. & Mavromara, P. (2011). High levels of HCV core+1 antibodies in HCV patients with hepatocellular carcinoma. *J Gen Virol* **92**, 1343-1351.

Dane, D. S., Cameron, C. H. & Briggs, M. (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet* **1**, 695-698.

Darke, P. L., Jacobs, A. R., Waxman, L. & Kuo, L. C. (1999). Inhibition of hepatitis C virus NS2/3 processing by NS4A peptides. Implications for control of viral processing. *J Biol Chem* **274**, 34511-34514.

Davis, G. L. (1999). Hepatitis C virus genotypes and quasispecies. *Am J Med* **107**, 21S-26S.

Davis, M., Sagan, S. M., Pezacki, J. P., Evans, D. J. & Simmonds, P. (2008). Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. *J Virol* **82**, 11824-11836.

Davis, T. R. W., T.J. McKenna, K.A. Granados, R.R. Shuler, M.L. Wood, H.A. (1993). Comparative recombinant protein production of eight insect cell lines. *In Vitro Cell Dev Biol Anim* **29A**, 388-390.

De Francesco, R., Urbani, A., Nardi, M. C., Tomei, L., Steinkuhler, C. & Tramontano, A. (1996). A zinc binding site in viral serine proteinases. *Biochemistry* **35**, 13282-13287.

Deleersnyder, V., Pillez, A., Wychowski, C., Blight, K., Xu, J., Hahn, Y. S., Rice, C. M. & Dubuisson, J. (1997). Formation of native hepatitis C virus glycoprotein complexes. *J Virol* **71**, 697-704.

den Boon, J. A. & Ahlquist, P. (2010). Organelle-like membrane compartmentalization of positive-strand RNA virus replication factories. *Annu Rev Microbiol* **64**, 241-256.

Dentzer, T. G., Lorenz, I. C., Evans, M. J. & Rice, C. M. (2009). Determinants of the hepatitis C virus nonstructural protein 2 protease domain required for production of infectious virus. *J Virol* **83**, 12702-12713.

Dimitrova, M., Imbert, I., Kieny, M. P. & Schuster, C. (2003). Protein-protein interactions between hepatitis C virus nonstructural proteins. *J Virol* **77**, 5401-5414.

Dopf, J. & Horiagon, T. M. (1996). Deletion mapping of the Aequorea victoria green fluorescent protein. *Gene* **173**, 39-44.

Dorner, M., Horwitz, J. A., Robbins, J. B., Barry, W. T., Feng, Q., Mu, K., Jones, C. T., Schoggins, J. W., Catanese, M. T., Burton, D. R., Law, M., Rice, C. M. & Ploss, A. (2011). A genetically humanized mouse model for hepatitis C virus infection. *Nature* **474**, 208-211.

Drouin, C., Lamarche, S., Bruneau, J., Soudeyns, H. & Shoukry, N. H. (2009). Cell-mediated immune responses directed against hepatitis C virus (HCV) alternate reading frame protein (ARFP) are undetectable during acute infection. *J Clin Virol* **47**, 102-103.

Dubuisson, J. & Rice, C. M. (1996). Hepatitis C virus glycoprotein folding: disulfide bond formation and association with calnexin. *J Virol* **70**, 778-786.

Dumoulin, F. L., von dem Bussche, A., Li, J., Khamzina, L., Wands, J. R., Sauerbruch, T. & Spengler, U. (2003). Hepatitis C Virus NS2 Protein Inhibits Gene Expression from Different Cellular and Viral Promoters in Hepatic and Nonhepatic Cell Lines. *Virology* **305**, 260.

Dunn, L. A. & Holz, R. W. (1983). Catecholamine secretion from digitonin-treated adrenal medullary chromaffin cells. *J Biol Chem* **258**, 4989-4993.

Eckart, M. R., Selby, M., Masiarz, F., Lee, C., Berger, K., Crawford, K., Kuo, C., Kuo, G., Houghton, M. & Choo, Q. L. (1993). The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. *Biochem Biophys Res Commun* **192**, 399-406.

Egger, D., Wolk, B., Gosert, R., Bianchi, L., Blum, H. E., Moradpour, D. & Bienz, K. (2002). Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol* **76**, 5974-5984.

Ehrmann, M., Boyd, D. & Beckwith, J. (1990). Genetic analysis of membrane protein topology by a sandwich gene fusion approach. *Proc Natl Acad Sci U S A* **87**, 7574-7578.

Elazar, M., Cheong, K. H., Liu, P., Greenberg, H. B., Rice, C. M. & Glenn, J. S. (2003). Amphipathic helix-dependent localization of NS5A mediates hepatitis C virus RNA replication. *J Virol* **77**, 6055-6061.

Elazar, M., Liu, P., Rice, C. M. & Glenn, J. S. (2004). An N-terminal amphipathic helix in hepatitis C virus (HCV) NS4B mediates membrane association, correct localization of replication complex proteins, and HCV RNA replication. *J Virol* **78**, 11393-11400.

Emmott, E., Dove, B. K., Howell, G., Chappell, L. A., Reed, M. L., Boyne, J. R., You, J. H., Brooks, G., Whitehouse, A. & Hiscox, J. A. (2008). Viral nucleolar localisation signals determine dynamic trafficking within the nucleolus. *Virology* **380**, 191-202.

Endo, T., Ohbayashi, H., Hayashi, Y., Ikehara, Y., Kochibe, N. & Kobata, A. (1988). Structural study on the carbohydrate moiety of human placental alkaline phosphatase. *J Biochem* **103**, 182-187.

Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C., Izumi, N., Marumo, F. & Sato, C. (1995). Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* **96**, 224-230.

Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C., Ogura, Y., Izumi, N., Marumo, F. & Sato, C. (1996). Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* **334**, 77-81.

Epstein, J. H., Quan, P. L., Briese, T., Street, C., Jabado, O., Conlan, S., Ali Khan, S., Verdugo, D., Hossain, M. J., Hutchison, S. K., Egholm, M., Luby, S. P., Daszak, P. & Lipkin, W. I. (2010). Identification of GBV-D, a novel GB-like flavivirus from old world frugivorous bats (Pteropus giganteus) in Bangladesh. *PLoS Pathog* **6**, e1000972.

Erdtmann, L., Franck, N., Lerat, H., Le Seyec, J., Gilot, D., Cannie, I., Gripon, P., Hibner, U. & Guguen-Guillouzo, C. (2003). The Hepatitis C Virus NS2 Protein Is an Inhibitor of CIDE-B-induced Apoptosis. *J Biol Chem* **278**, 18256-18264.

Errington, W., Wardell, A. D., McDonald, S., Goldin, R. D. & McGarvey, M. J. (1999). Subcellular localisation of NS3 in HCV-infected hepatocytes. *J Med Virol* **59**, 456-462.

Evans, M. J., von Hahn, T., Tscherne, D. M., Syder, A. J., Panis, M., Wolk, B., Hatziioannou, T., McKeating, J. A., Bieniasz, P. D. & Rice, C. M. (2007). Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* **446**, 801-805. Epub 2007 Feb 2025.

Eyre, N. S., Baumert, T. F. & Beard, M. R. (2009). Closing the gap: the tight junction protein occludin and hepatitis C virus entry. *Hepatology* **49**, 1770-1772.

Failla, C., Tomei, L. & De Francesco, R. (1994). Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J Virol* **68**, 3753-3760.

Falgout, B., Pethel, M., Zhang, Y. M. & Lai, C. J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J Virol* **65**, 2467-2475.

Fang, J., Brzostowski, J. A., Ou, S., Isik, N., Nair, V. & Jin, T. (2007). A vesicle surface tyrosine kinase regulates phagosome maturation. *J Cell Biol* **178**, 411-423.

Farci, P., Alter, H. J., Govindarajan, S., Wong, D. C., Engle, R., Lesniewski, R. R., Mushahwar, I. K., Desai, S. M., Miller, R. H., Ogata, N. & et al. (1992). Lack of protective immunity against reinfection with hepatitis C virus. *Science* **258**, 135-140.

Ferrari, E., Wright-Minogue, J., Fang, J. W., Baroudy, B. M., Lau, J. Y. & Hong, Z. (1999). Characterization of soluble hepatitis C virus RNA-dependent RNA polymerase expressed in Escherichia coli. *J Virol* **73**, 1649-1654.

Flajolet, M., Rotondo, G., Daviet, L., Bergametti, F., Inchauspe, G., Tiollais, P., Transy, C. & Legrain, P. (2000). A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* **242**, 369-379.

Flisiak, R., Feinman, S. V., Jablkowski, M., Horban, A., Kryczka, W., Pawlowska, M., Heathcote, J. E., Mazzella, G., Vandelli, C., Nicolas-Metral, V., Grosgurin, P., Liz, J. S., Scalfaro, P., Porchet, H. & Crabbe, R. (2009). The cyclophilin inhibitor Debio 025 combined with PEG IFNalpha2a significantly reduces viral load in treatment-naive hepatitis C patients. *Hepatology* **49**, 1460-1468.

Florese, R. H., Nagano-Fujii, M., Iwanaga, Y., Hidajat, R. & Hotta, H. (2002). Inhibition of protein synthesis by the nonstructural proteins NS4A and NS4B of hepatitis C virus. *Virus Res* **90**, 119-131.

Foster, T. L., Belyaeva, T., Stonehouse, N. J., Pearson, A. R. & Harris, M. (2010a). All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. *J Virol* **84**, 9267-9277.

Foster, T. L., Kalverda, A., Thompson, G., Thompson, J., Pearson, A., Foster, R., Homans, S., Harris, M. & Griffin, S. D. C. (2011). Solution structure of monomeric hepatitis C virus (HCV) p7 in a native, drug binding,

hairpin conformation. In *18th International Symposium on Hepatitis C Virus and Related Viruses*. Seattle, WA, USA: Gilead.

Foster, T. L., Tedbury, P. R., Pearson, A. R. & Harris, M. (2010b). A comparative analysis of the fluorescence properties of the wild-type and active site mutants of the hepatitis C virus autoprotease NS2-3. *Biochim Biophys Acta* **1804**, 212-222.

Franck, N., Le Seyec, J., Guguen-Guillouzo, C. & Erdtmann, L. (2005). Hepatitis C Virus NS2 Protein Is Phosphorylated by the Protein Kinase CK2 and Targeted for Degradation to the Proteasome. *J Virol* **79**, 2700-2708.

Frick, D. N., Banik, S. & Rypma, R. S. (2007). Role of divalent metal cations in ATP hydrolysis catalyzed by the hepatitis C virus NS3 helicase: magnesium provides a bridge for ATP to fuel unwinding. *J Mol Biol* **365**, 1017-1032.

Fujiki, Y., Hubbard, A. L., Fowler, S. & Lazarow, P. B. (1982). Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J Cell Biol* **93**, 97-102.

Fukushi, S., Katayama, K., Kurihara, C., Ishiyama, N., Hoshino, F. B., Ando, T. & Oya, A. (1994). Complete 5' Noncoding Region Is Necessary for the Efficient Internal Initiation of Hepatitis C Virus RNA. *Biochemical and Biophysical Research Communications* **199**, 425.

Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkuhler, C. & De Francesco, R. (1998). Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *J Virol* **72**, 6758-6769.

Gallinari, P., Paolini, C., Brennan, D., Nardi, C., Steinkuhler, C. & De Francesco, R. (1999). Modulation of hepatitis C virus NS3 protease and helicase activities through the interaction with NS4A. *Biochemistry* **38**, 5620-5632.

Gao, D. Y., Jin, G. D., Yao, B. L., Zhang, D. H., Gu, L. L., Lu, Z. M., Gong, Q., Lone, Y. C., Deng, Q. & Zhang, X. X. (2010). Characterization of the specific CD4+ T cell response against the F protein during chronic hepatitis C virus infection. *PLoS One* **5**, e14237.

Gastaminza, P., Cheng, G., Wieland, S., Zhong, J., Liao, W. & Chisari, F. V. (2008). Cellular determinants of hepatitis C virus assembly, maturation, degradation, and secretion. *J Virol* **82**, 2120-2129.

Georgiou, C. D., Dueweke, T. J. & Gennis, R. B. (1988). Beta-galactosidase gene fusions as probes for the cytoplasmic regions of subunits I and II of the membrane-bound cytochrome d terminal oxidase from Escherichia coli. *J Biol Chem* **263**, 13130-13137.

Gerin, J. L., Casey, J. L. & Purcell, R. H. (2001). *Hepatitis D virus*. Philidelphia: Lippincott Williams and Wilkins.

Giordano, T. P., Kramer, J. R., Souchek, J., Richardson, P. & El-Serag, H. B. (2004). Cirrhosis and hepatocellular carcinoma in HIV-infected veterans with and without the hepatitis C virus: a cohort study, 1992-2001. *Arch Intern Med* **164**, 2349-2354.

Goffard, A. & Dubuisson, J. (2003). Glycosylation of hepatitis C virus envelope proteins. *Biochimie* **85**, 295-301.

Goh, P. Y., Tan, Y. J., Lim, S. P., Tan, Y. H., Lim, S. G., Fuller-Pace, F. & Hong, W. (2004). Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV)

NS5B protein and the potential role of p68 in HCV RNA replication. *J Virol* **78**, 5288-5298.

Gosert, R., Egger, D., Lohmann, V., Bartenschlager, R., Blum, H. E., Bienz, K. & Moradpour, D. (2003). Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J Virol* **77**, 5487-5492.

Gottwein, J. M., Jensen, T. B., Mathiesen, C. K., Meuleman, P., Serre, S. B., Lademann, J. B., Ghanem, L., Scheel, T. K., Leroux-Roels, G. & Bukh, J. (2011). Development and Application of Hepatitis C Reporter Viruses with Genotype 1 to 7 Core-Nonstructural Protein 2 (NS2) Expressing Fluorescent Proteins or Luciferase in Modified JFH1 NS5A. *J Virol* **85**, 8913-8928.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**, W695-699.

Gouttenoire, J., Castet, V., Montserret, R., Arora, N., Raussens, V., Ruysschaert, J. M., Diesis, E., Blum, H. E., Penin, F. & Moradpour, D. (2009a). Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B. *J Virol* **83**, 6257-6268.

Gouttenoire, J., Montserret, R., Kennel, A., Penin, F. & Moradpour, D. (2009b). An amphipathic alpha-helix at the C terminus of hepatitis C virus nonstructural protein 4B mediates membrane association. *J Virol* **83**, 11378-11384.

Grakoui, A., Shoukry, N. H., Woollard, D. J., Han, J. H., Hanson, H. L., Ghrayeb, J., Murthy, K. K., Rice, C. M. & Walker, C. M. (2003). HCV persistence and immune evasion in the absence of memory T cell help. *Science* **302**, 659-662.

Grakoui, A., Wychowski, C., Lin, C., Feinstone, S. M. & Rice, C. M. (1993a). Expression and identification of hepatitis C virus polyprotein cleavage products. *J Virol* **67**, 1385-1395.

Grakoui, A. M., D. W. Wychowski, C. Feinstone, S. M. Rice, C. M. (1993b). A second hepatitis C virus-encoded proteinase. *Proc Natl Acad Sci U S A* **90**, 10583-10587.

Griffin, S. (2010). Inhibition of HCV p7 as a therapeutic target. *Curr Opin Investig Drugs* **11**, 175-181.

Griffin, S., Clarke, D., McCormick, C., Rowlands, D. & Harris, M. (2005). Signal peptide cleavage and internal targeting signals direct the hepatitis C virus p7 protein to distinct intracellular membranes. *J Virol* **79**, 15525-15536.

Griffin, S., Stgelais, C., Owsianka, A. M., Patel, A. H., Rowlands, D. & Harris, M. (2008). Genotype-dependent sensitivity of hepatitis C virus to inhibitors of the p7 ion channel. *Hepatology* **48**, 1779-1790.

Griffin, S. D., Harvey, R., Clarke, D. S., Barclay, W. S., Harris, M. & Rowlands, D. J. (2004). A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria. *J Gen Virol* **85**, 451-461.

Grudnik, P., Bange, G. & Sinning, I. (2009). Protein targeting by the signal recognition particle. *Biol Chem* **390**, 775-782.

Guerniou, V., Gillet, R., Berree, F., Carboni, B. & Felden, B. (2007). Targeted inhibition of the hepatitis C internal ribosomal entry site genomic RNA with oligonucleotide conjugates. *Nucleic Acids Res* **35**, 6778-6787.

Guo, K. K., Tang, Q. H., Zhang, Y. M., Kang, K. & He, L. (2011). Identification of two internal signal peptide sequences: critical for classical swine fever virus non-structural protein 2 to trans-localize to the endoplasmic reticulum. *Virol J* **8**, 236.

Hadziyannis, S. J., Sette, H., Jr., Morgan, T. R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H., Jr., Bernstein, D., Rizzetto, M., Zeuzem, S., Pockros, P. J., Lin, A. & Ackrill, A. M. (2004). Peginterferon-alpha2a and ribavirin combination therapy in

chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* **140**, 346-355.

Hamman, B. D., Hendershot, L. M. & Johnson, A. E. (1998). BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation. *Cell* **92**, 747-758.

Hanein, D., Matlack, K. E., Jungnickel, B., Plath, K., Kalies, K. U., Miller, K. R., Rapoport, T. A. & Akey, C. W. (1996). Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **87**, 721-732.

Hanoulle, X., Badillo, A., Verdegem, D., Penin, F. & Lippens, G. (2010). The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept Lett* **17**, 1012-1018.

Hanoulle, X., Verdegem, D., Badillo, A., Wieruszeski, J. M., Penin, F. & Lippens, G. (2009). Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem Biophys Res Commun* **381**, 634-638.

Hartmann, E., Rapoport, T. A. & Lodish, H. F. (1989). Predicting the orientation of eukaryotic membrane-spanning proteins. *Proc Natl Acad Sci U S A* **86**, 5786-5790.

Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *Embo J* **5**, 3021-3027.

Henderson, N. S., So, S. S., Martin, C., Kulkarni, R. & Thanassi, D. G. (2004). Topology of the outer membrane usher PapC determined by site-directed fluorescence labeling. *J Biol Chem* **279**, 53747-53754.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M. & Shimotohno, K. (1991). Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. *Proc Natl Acad Sci U S A* **88**, 5547-5551.

Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K. & Shimotohno, K. (1993a). Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J Virol* **67**, 4665-4675.

Hijikata, M., Mizushima, H., Tanji, Y., Komoda, Y., Hirowatari, Y., Akagi, T., Kato, N., Kimura, K. & Shimotohno, K. (1993b). Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. *Proc Natl Acad Sci U S A* **90**, 10773-10777.

Hiramatsu, N., Kasai, A., Hayakawa, K., Yao, J. & Kitamura, M. (2006). Real-time detection and continuous monitoring of ER stress in vitro and in vivo by ES-TRAP: evidence for systemic, transient ER stress during endotoxemia. *Nucleic Acids Res* **34**, e93.

Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378-379.

Hirowatari, Y., Hijikata, M., Tanji, Y., Nyunoya, H., Mizushima, H., Kimura, K., Tanaka, T., Kato, N. & Shimotohno, K. (1993). Two proteinase activities in HCV polypeptide expressed in insect cells using baculovirus vector. *Arch Virol* **133**, 349-356.

Hofmann, K. & Stoffel, W. (1993). TMPred, p. 166.

Honda, M., Beard, M. R., Ping, L. H. & Lemon, S. M. (1999). A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J Virol* **73**, 1165-1174.

Honda, M., Brown, E. A. & Lemon, S. M. (1996). Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *Rna* **2**, 955-968.

Honda, M., Kaneko, S., Sakai, A., Unoura, M., Murakami, S. & Kobayashi, K. (1994). Degree of diversity of hepatitis C virus quasispecies and progression of liver disease. *Hepatology* **20**, 1144-1151.

Hoofnagle, J. H. (1997). Hepatitis C: The clinical spectrum of disease. *Hepatology* **26**, 15S-20S.

Hope, R. G. & McLauchlan, J. (2000). Sequence motifs required for lipid droplet association and protein stability are unique to the hepatitis C virus core protein. *J Gen Virol* **81**, 1913-1925.

Huang, L., Hwang, J., Sharma, S. D., Hargittai, M. R., Chen, Y., Arnold, J. J., Raney, K. D. & Cameron, C. E. (2005). Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein. *J Biol Chem* **280**, 36417-36428.

Hughes, M., Gretton, S., Shelton, H., Brown, D. D., McCormick, C. J., Angus, A. G., Patel, A. H., Griffin, S. & Harris, M. (2009a). A conserved proline between domains II and III of hepatitis C virus NS5A influences both RNA replication and virus assembly. *J Virol* **83**, 10788-10796.

Hughes, M., Griffin, S. & Harris, M. (2009b). Domain III of NS5A contributes to both RNA replication and assembly of hepatitis C virus particles. *J Gen Virol* **90**, 1329-1334.

Hugle, T., Fehrmann, F., Bieck, E., Kohara, M., Krausslich, H. G., Rice, C. M., Blum, H. E. & Moradpour, D. (2001). The hepatitis C virus nonstructural protein 4B is an integral endoplasmic reticulum membrane protein. *Virology* **284**, 70-81.

Isherwood, B. J. & Patel, A. H. (2005). Analysis of the processing and transmembrane topology of the E2p7 protein of hepatitis C virus. *J Gen Virol* **86**, 667-676.

Ito, T. & Lai, M. M. (1997). Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *J Virol* **71**, 8698-8706.

Ito, T. & Lai, M. M. (1999). An internal polypyrimidine-tract-binding protein-binding site in the hepatitis C virus RNA attenuates translation, which is relieved by the 3'-untranslated sequence. *Virology* **254**, 288-296.

Ito, T., Tahara, S. M. & Lai, M. M. (1998). The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. *J Virol* **72**, 8789-8796.

Ivashkina, N., Wolk, B., Lohmann, V., Bartenschlager, R., Blum, H. E., Penin, F. & Moradpour, D. (2002). The hepatitis C virus RNA-dependent RNA polymerase membrane insertion sequence is a transmembrane segment. *J Virol* **76**, 13088-13093.

Jirasko, V., Montserret, R., Appel, N., Janvier, A., Eustachi, L., Brohm, C., Steinmann, E., Pietschmann, T., Penin, F. & Bartenschlager, R. (2008). Structural and functional characterization of nonstructural protein 2 for its role in hepatitis C virus assembly. *J Biol Chem* **283**, 28546-28562. Epub 22008 Jul 28521.

Jirasko, V., Montserret, R., Lee, J. Y., Gouttenoire, J., Moradpour, D., Penin, F. & Bartenschlager, R. (2010). Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly. *PLoS Pathog*.

Jones, C. T., Murray, C. L., Eastman, D. K., Tassello, J. & Rice, C. M. (2007). Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *J Virol* **81**, 8374-8383. Epub 2007 May 8330.

Jopling, C. L., Schutz, S. & Sarnow, P. (2008). Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome. *Cell Host Microbe* **4**, 77-85.

Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M. & Sarnow, P. (2005). Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**, 1577-1581.

Kadoya, H., Nagano-Fujii, M., Deng, L., Nakazono, N. & Hotta, H. (2005). Nonstructural proteins 4A and 4B of hepatitis C virus transactivate the interleukin 8 promoter. *Microbiol Immunol* **49**, 265-273.

Kalinina, O., Norder, H., Mukomolov, S. & Magnius, L. O. (2002). A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* **76**, 4034-4043.

Kamer, G. & Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res* **12**, 7269-7282.

Kapoor, A., Simmonds, P., Gerold, G., Qaisar, N., Jain, K., Henriquez, J. A., Firth, C., Hirschberg, D. L., Rice, C. M., Shields, S. & Lipkin, W. I. (2011). Characterization of a canine homolog of hepatitis C virus. *Proc Natl Acad Sci U S A* **108**, 11608-11613.

Kato, T., Choi, Y., Elmowalid, G., Sapp, R. K., Barth, H., Furusaka, A., Mishiro, S., Wakita, T., Krawczynski, K. & Liang, T. J. (2008). Hepatitis C virus JFH-1 strain infection in chimpanzees is associated with low pathogenicity and emergence of an adaptive mutation. *Hepatology* **48**, 732-740.

Kato, T., Date, T., Miyamoto, M., Furusaka, A., Tokushige, K., Mizokami, M. & Wakita, T. (2003). Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* **125**, 1808-1817.

Kaukinen, P., Sillanpaa, M., Kotenko, S., Lin, R., Hiscott, J., Melen, K. & Julkunen, I. (2006). Hepatitis C virus NS2 and NS3/4A proteins are potent inhibitors of host cell cytokine/chemokine gene expression. *Virol J* **3**, 66.

Khafizov, K., Staritzbichler, R., Stamm, M. & Forrest, L. R. (2010). A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe. *Biochemistry* **49**, 10702-10713.

Khromykh, A. A., Sedlak, P. L. & Westaway, E. G. (2000). cis- and trans-acting elements in flavivirus RNA replication. *J Virol* **74**, 3253-3263.

Kieft, J. S., Zhou, K., Jubin, R., Murray, M. G., Lau, J. Y. & Doudna, J. A. (1999). The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. *J Mol Biol* **292**, 513-529.

Kiiver, K., Merits, A., Ustav, M. & Zusinaite, E. (2006). Complex formation between hepatitis C virus NS2 and NS3 proteins. *Virus Research* **117**, 264.

Kim, D. W., Gwack, Y., Han, J. H. & Choe, J. (1995). C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem Biophys Res Commun* **215**, 160-166.

Kim, J. E., Song, W. K., Chung, K. M., Back, S. H. & Jang, S. K. (1999). Subcellular localization of hepatitis C viral proteins in mammalian cells. *Arch Virol* **144**, 329-343.

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. & Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **87**, 343-355.

Kim, S., Welsch, C., Yi, M. & Lemon, S. M. (2011). Regulation of the production of infectious genotype 1a hepatitis C virus by NS5A domain III. *J Virol* **85**, 6645-6656.

Kim, Y. K., Kim, C. S., Lee, S. H. & Jang, S. K. (2002). Domains I and II in the 5' nontranslated region of the HCV genome are required for RNA replication. *Biochem Biophys Res Commun* **290**, 105-112.

Kolykhalov, A. A., Feinstone, S. M. & Rice, C. M. (1996). Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol* **70**, 3363-3371.

Kolykhalov, A. A., Mihalik, K., Feinstone, S. M. & Rice, C. M. (2000). Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *J Virol* **74**, 2046-2051.

Korres, H. & Verma, N. K. (2004). Topological analysis of glucosyltransferase GtrV of Shigella flexneri by a dual reporter system and identification of a unique reentrant loop. *J Biol Chem* **279**, 22469-22476.

Koutsoudakis, G., Kaul, A., Steinmann, E., Kallis, S., Lohmann, V., Pietschmann, T. & Bartenschlager, R. (2006). Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses. *J Virol* **80**, 5308-5320.

Krawczyk, E., Suprynowicz, F. A., Sudarshan, S. R. & Schlegel, R. (2010). Membrane orientation of the human papillomavirus type 16 E5 oncoprotein. *J Virol* **84**, 1696-1703.

Krieger, N., Lohmann, V. & Bartenschlager, R. (2001). Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. *J Virol* **75**, 4614-4624.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.

Kumar, D., Farrell, G. C., Fung, C. & George, J. (2002). Hepatitis C virus genotype 3 is cytopathic to hepatocytes: Reversal of hepatic steatosis after sustained therapeutic response. *Hepatology* **36**, 1266-1272.

Kyono, K., Miyashiro, M. & Taguchi, I. (2002). Human eukaryotic initiation factor 4AII associates with hepatitis C virus NS5B protein in vitro. *Biochem Biophys Res Commun* **292**, 659-666.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132.

Lackner, T., Muller, A., Konig, M., Thiel, H. J. & Tautz, N. (2005). Persistence of bovine viral diarrhea virus is determined by a cellular cofactor of a viral autoprotease. *J Virol* **79**, 9746-9755.

Lackner, T., Muller, A., Pankraz, A., Becher, P., Thiel, H. J., Gorbalenya, A. E. & Tautz, N. (2004). Temporal modulation of an autoprotease is crucial for replication and pathogenicity of an RNA virus. *J Virol* **78**, 10765-10775.

Lackner, T., Thiel, H. J. & Tautz, N. (2006). Dissection of a viral autoprotease elucidates a function of a cellular chaperone in proteolysis. *Proc Natl Acad Sci U S A* **103**, 1510-1515.

Lai, M. E., Mazzoleni, A. P., Argiolu, F., De Virgilis, S., Balestrieri, A., Purcell, R. H., Cao, A. & Farci, P. (1994). Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. *Lancet* **343**, 388-390.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Lavanchy, D. (2009). The global burden of hepatitis C. *Liver Int* **29 Suppl 1**, 74-81.

Le Du, M. H., Stigbrand, T., Taussig, M. J., Menez, A. & Stura, E. A. (2001). Crystal structure of alkaline phosphatase from human placenta at 1.8 A resolution. Implication for a substrate specificity. *J Biol Chem* **276**, 9158-9165.

Leary, T. P., Muerhoff, A. S., Simons, J. N., Pilot-Matias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996). Sequence and genomic organization of GBV-C: a novel member of the flaviviridae associated with human non-A-E hepatitis. *J Med Virol* **48**, 60-67.

Lee, H., Liu, Y., Mejia, E., Paul, A. V. & Wimmer, E. (2006). The C-terminal hydrophobic domain of hepatitis C virus RNA polymerase NS5B can be replaced with a heterologous domain of poliovirus protein 3A. *J Virol* **80**, 11343-11354.

Lee, J. W., Liao, P. C., Young, K. C., Chang, C. L., Chen, S. S., Chang, T. T., Lai, M. D. & Wang, S. W. (2011). Identification of hnRNPH1, NF45, and C14orf166 as Novel Host Interacting Partners of the Mature Hepatitis C Virus Core Protein. *J Proteome Res* **2011**, 24.

Lee, K. J., Choi, J., Ou, J. H. & Lai, M. M. (2004). The C-terminal transmembrane domain of hepatitis C virus (HCV) RNA polymerase is essential for HCV replication in vivo. *J Virol* **78**, 3797-3802.

Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. F. & Weber, P. C. (1999). Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol* **6**, 937-943.

Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr. & Lemon, S. M. (2005a). Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci U S A* **102**, 2992-2997. Epub 2005 Feb 2914.

Li, X. D., Sun, L., Seth, R. B., Pineda, G. & Chen, Z. J. (2005b). Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc Natl Acad Sci U S A* **102**, 17717-17722. Epub 12005 Nov 17721.

Liang, Y., Ye, H., Kang, C. B. & Yoon, H. S. (2007). Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry* **46**, 11550-11558.

Liaw, Y. F. (2001). Concurrent hepatitis B and C virus infection: Is hepatitis C virus stronger? *J Gastroenterol Hepatol* **16**, 597-598.

Lin, C., Lindenbach, B. D., Pragai, B. M., McCourt, D. W. & Rice, C. M. (1994). Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J Virol* **68**, 5063-5073.

Lin, C., Thomson, J. A. & Rice, C. M. (1995). A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro. *J Virol* **69**, 4373-4380.

Lindenbach, B. D., Evans, M. J., Syder, A. J., Wolk, B., Tellinghuisen, T. L., Liu, C. C., Maruyama, T., Hynes, R. O., Burton, D. R., McKeating, J. A. & Rice, C. M. (2005). Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623-626. Epub 2005 Jun 2009.

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. & 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**, 3805-3809. Epub 2006 Feb 3816.

Lindenbach, B. D., Pragai, B. M., Montserret, R., Beran, R. K., Pyle, A. M., Penin, F. & Rice, C. M. (2007). The C terminus of hepatitis C virus NS4A encodes an electrostatic switch that regulates NS5A hyperphosphorylation and viral replication. *J Virol* **81**, 8905-8918. Epub 2007 Jun 8920.

Lindenbach, B. D. & Rice, C. M. (2003). Evasive maneuvers by hepatitis C virus. *Hepatology* **38**, 769-771.

Lo, S. Y., Selby, M. J. & Ou, J. H. (1996). Interaction between hepatitis C virus core protein and E1 envelope protein. *J Virol* **70**, 5177-5182.

Lohmann, V., Korner, F., Dobierzewska, A. & Bartenschlager, R. (2001). Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *J Virol* **75**, 1437-1449.

Lohmann, V., Korner, F., Koch, J. O., Herian, U., Theilmann, L. & Bartenschlager, R. (1999). Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line. *Science* **285**, 110-113.

Lorenz, H., Hailey, D. W. & Lippincott-Schwartz, J. (2006a). Fluorescence protease protection of GFP chimeras to reveal protein topology and subcellular localization. *Nat Methods* **3**, 205-210.

Lorenz, I. C., Marcotrigiano, J., Dentzer, T. G. & Rice, C. M. (2006b). Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature* **442**, 831-835. Epub 2006 Jul 2023.

Love, R. A., Brodsky, O., Hickey, M. J., Wells, P. A. & Cronin, C. N. (2009). Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J Virol* **83**, 4395-4403.

Luik, P., Chew, C., Aittoniemi, J., Chang, J., Wentworth, P., Jr., Dwek, R. A., Biggin, P. C., Venien-Bryan, C. & Zitzmann, N. (2009). The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. *Proc Natl Acad Sci U S A* **106**, 12712-12716.

Lundin, M., Lindstrom, H., Gronwall, C. & Persson, M. A. (2006). Dual topology of the processed hepatitis C virus protein NS4B is influenced by the NS5A protein. *J Gen Virol* **87**, 3263-3272.

Lundin, M., Monne, M., Widell, A., Von Heijne, G. & Persson, M. A. (2003). Topology of the membrane-associated hepatitis C virus protein NS4B. *J Virol* **77**, 5428-5438.

Ma, Y., Anantpadma, M., Timpe, J. M., Shanmugam, S., Singh, S. M., Lemon, S. M. & Yi, M. (2011). Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins. *J Virol* **85**, 86-97.

Ma, Y., Yates, J., Liang, Y., Lemon, S. M. & Yi, M. (2008). NS3 helicase domains involved in infectious intracellular hepatitis C virus particle assembly. *J Virol* **82**, 7624-7639.

Macdonald, A., Chan, J. K. & Harris, M. (2005). Perturbation of epidermal growth factor receptor complex formation and Ras signalling in cells harbouring the hepatitis C virus subgenomic replicon. *J Gen Virol* **86**, 1027-1033.

Macdonald, A., Crowder, K., Street, A., McCormick, C. & Harris, M. (2004). The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *J Gen Virol* **85**, 721-729.

Macdonald, A., Crowder, K., Street, A., McCormick, C., Saksela, K. & Harris, M. (2003). The hepatitis C virus non-structural NS5A protein inhibits activating protein-1 function by perturbing ras-ERK pathway signaling. *J Biol Chem* **278**, 17775-17784.

Major, M. E., Dahari, H., Mihalik, K., Puig, M., Rice, C. M., Neumann, A. U. & Feinstone, S. M. (2004). Hepatitis C virus kinetics and host responses associated with disease and outcome of infection in chimpanzees. *Hepatology* **39**, 1709-1720.

Maley, F., Trimble, R. B., Tarentino, A. L. & Plummer, T. H., Jr. (1989). Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal Biochem* **180**, 195-204.

Mankouri, J., Griffin, S. & Harris, M. (2008a). The hepatitis C virus non-structural protein NS5A alters the trafficking profile of the epidermal growth factor receptor. *Traffic* **9**, 1497-1509.

Mankouri, J., Milward, A., Pryde, K. R., Warter, L., Martin, A. & Harris, M. (2008b). A comparative cell biological analysis reveals only limited functional homology between the NS5A proteins of hepatitis C virus and GB virus B. *J Gen Virol* **89**, 1911-1920.

Mankouri, J., Tedbury, P. R., Gretton, S., Hughes, M. E., Griffin, S. D., Dallas, M. L., Green, K. A., Hardie, D. G., Peers, C. & Harris, M. (2010). Enhanced hepatitis C virus genome replication and lipid accumulation mediated by inhibition of AMP-activated protein kinase. *Proc Natl Acad Sci U S A* **107**, 11549-11554.

Marukian, S., Andrus, L., Sheahan, T. P., Jones, C. T., Charles, E. D., Ploss, A., Rice, C. M. & Dustin, L. B. (2011). Hepatitis C virus induces interferon-lambda and interferon-stimulated genes in primary liver cultures. *Hepatology* **2011**, 24580.

Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T. & Suzuki, T. (2008). Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles. *J Virol* **82**, 7964-7976.

Mast, E. E., Alter, M. J. & Margolis, H. S. (1999). Strategies to prevent and control hepatitis B and C virus infections: a global perspective. *Vaccine* **17**, 1730.

Mateu, G., Donis, R. O., Wakita, T., Bukh, J. & Grakoui, A. (2008). Intragenotypic JFH1 based recombinant hepatitis C virus produces high levels of infectious particles but causes increased cell death. *Virology* **376**, 397-407.

Matsunaga, J., Young, T. A., Barnett, J. K., Barnett, D., Bolin, C. A. & Haake, D. A. (2002). Novel 45-kilodalton leptospiral protein that is processed to a 31-kilodalton growth-phase-regulated peripheral membrane protein. *Infect Immun* **70**, 323-334.

Mazumdar, B., Banerjee, A., Meyer, K. & Ray, R. (2011). Hepatitis C virus E1 envelope glycoprotein interacts with apolipoproteins in facilitating entry into hepatocytes. *Hepatology* **2011**, 24523.

McHutchison, J. G., Dusheiko, G., Shiffman, M. L., Rodriguez-Torres, M., Sigal, S., Bourliere, M., Berg, T., Gordon, S. C., Campbell, F. M., Theodore, D., Blackman, N., Jenkins, J. & Afdhal, N. H. (2007). Eltrombopag for thrombocytopenia in patients with cirrhosis associated with hepatitis C. *N Engl J Med* **357**, 2227-2236.

McHutchison, J. G., Manns, M., Patel, K., Poynard, T., Lindsay, K. L., Trepo, C., Dienstag, J., Lee, W. M., Mak, C., Garaud, J. J. & Albrecht, J. K. (2002). Adherence to combination therapy enhances sustained response in genotype-1-infected patients with chronic hepatitis C. *Gastroenterology* **123**, 1061-1069.

McLauchlan, J. (2009). Hepatitis C virus: viral proteins on the move. *Biochem Soc Trans* **37**, 986-990.

McLauchlan, J., Lemberg, M. K., Hope, G. & Martoglio, B. (2002). Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *EMBO J* **21**, 3980-3988.

Melen, K., Krogh, A. & von Heijne, G. (2003). Reliability measures for membrane protein topology prediction algorithms. *J Mol Biol* **327**, 735-744.

Mercer, D. F., Schiller, D. E., Elliott, J. F., Douglas, D. N., Hao, C., Rinfret, A., Addison, W. R., Fischer, K. P., Churchill, T. A., Lakey, J. R., Tyrrell, D. L. & Kneteman, N. M. (2001). Hepatitis C virus replication in mice with chimeric human livers. *Nat Med* **7**, 927-933.

Merican, I., Sherlock, S., McIntyre, N. & Dusheiko, G. M. (1993). Clinical, biochemical and histological features in 102 patients with chronic hepatitis C virus infection. *Q J Med* **86**, 119-125.

Micanovic, R., Bailey, C. A., Brink, L., Gerber, L., Pan, Y. C., Hulmes, J. D. & Udenfriend, S. (1988). Aspartic acid-484 of nascent placental alkaline phosphatase condenses with a phosphatidylinositol glycan to become the carboxyl terminus of the mature enzyme. *Proc Natl Acad Sci U S A* **85**, 1398-1402.

Miyanari, Y., Atsuzawa, K., Usuda, N., Watashi, K., Hishiki, T., Zayas, M., Bartenschlager, R., Wakita, T., Hijikata, M. & Shimotohno, K. (2007). The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol* **9**, 1089-1097.

Mizushima, H., Hijikata, M., Asabe, S., Hirota, M., Kimura, K. & Shimotohno, K. (1994). Two hepatitis C virus glycoprotein E2 products with different C termini. *J Virol* **68**, 6215-6222.

Mizushima, H. H., M. Asabe, S. Hirota, M. Kimura, K. Shimotohno, K. (1994). Two hepatitis C virus glycoprotein E2 products with different C termini. *J Virol* **68**, 6215-6222.

Mohamed Ael, S., al Karawi, M. A. & Mesa, G. A. (1997). Dual infection with hepatitis C and B viruses: clinical and histological study in Saudi patients. *Hepatogastroenterology* **44**, 1404-1406.

Montserret, R., Saint, N., Vanbelle, C., Salvay, A. G., Simorre, J. P., Ebel, C., Sapay, N., Renisio, J. G., Bockmann, A., Steinmann, E., Pietschmann, T., Dubuisson, J., Chipot, C. & Penin, F. (2010). NMR structure and ion channel activity of the p7 protein from hepatitis C virus. *J Biol Chem* **285**, 31446-31461.

Moradpour, D., Brass, V., Bieck, E., Friebe, P., Gosert, R., Blum, H. E., Bartenschlager, R., Penin, F. & Lohmann, V. (2004). Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. *J Virol* **78**, 13278-13284.

Moradpour, D., Penin, F. & Rice, C. M. (2007). Replication of hepatitis C virus. *Nat Rev Microbiol* **5**, 453-463. Epub 2007 May 2008.

Morice, Y., Ratinier, M., Miladi, A., Chevaliez, S., Germanidis, G., Wedemeyer, H., Laperche, S., Lavergne, J. P. & Pawlotsky, J. M. (2009). Seroconversion to hepatitis C virus alternate reading frame protein during acute infection. *Hepatology* **49**, 1449-1459.

Morikawa, K., Zhao, Z., Date, T., Miyamoto, M., Murayama, A., Akazawa, D., Tanabe, J., Sone, S. & Wakita, T. (2007). The roles of CD81 and glycosaminoglycans in the adsorption and uptake of infectious HCV particles. *J Med Virol* **79**, 714-723.

Morota, K., Fujinami, R., Kinukawa, H., Machida, T., Ohno, K., Saegusa, H. & Takeda, K. (2009). A new sensitive and automated chemiluminescent microparticle immunoassay for quantitative determination of hepatitis C virus core antigen. *J Virol Methods* **157**, 8-14.

Mothes, W., Heinrich, S. U., Graf, R., Nilsson, I., von Heijne, G., Brunner, J. & Rapoport, T. A. (1997). Molecular mechanism of membrane protein integration into the endoplasmic reticulum. *Cell* **89**, 523-533.

Muir, A. J., Shiffman, M. L., Zaman, A., Yoffe, B., de la Torre, A., Flamm, S., Gordon, S. C., Marotta, P., Vierling, J. M., Lopez-Talavera, J. C., Byrnes-Blake, K., Fontana, D., Freeman, J., Gray, T., Hausman, D., Hunder, N. N. & Lawitz, E. (2010). Phase 1b study of pegylated interferon lambda 1 with or without ribavirin in patients with chronic genotype 1 hepatitis C virus infection. *Hepatology* **52**, 822-832.

Murakami, K., Abe, M., Kageyama, T., Kamoshita, N. & Nomoto, A. (2001). Down-regulation of translation driven by hepatitis C virus internal ribosomal entry site by the 3' untranslated region of RNA. *Arch Virol* **146**, 729-741.

Murphy, D., Chamberland, J., Dandavino, R. & Sablon, E. (2007). A NEW GENOTYPE OF HEPATITIS C VIRUS ORIGINATING FROM CENTRAL AFRICA. *Heptology* **46**, 623A.

Murray, C. L., Jones, C. T., Tassello, J. & Rice, C. M. (2007). Alanine scanning of the hepatitis C virus core protein reveals numerous residues essential for production of infectious virus. *J Virol* **81**, 10220-10231.

Narbus, C. M., Israelow, B., Sourisseau, M., Michta, M. L., Hopcraft, S. E., Zeiner, G. M. & Evans, M. J. (2011). HepG2 cells expressing miR-122 support the entire hepatitis C virus life cycle. *J Virol* **2011**, 14.

Nestorowicz, A., Chambers, T. J. & Rice, C. M. (1994). Mutagenesis of the yellow fever virus NS2A/2B cleavage site: effects on proteolytic processing, viral replication, and evidence for alternative processing of the NS2A protein. *Virology* **199**, 114-123.

Nilsson, I. M. & von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J Biol Chem* **268**, 5798-5801.

Nkontchou, G., Ziol, M., Aout, M., Lhabadie, M., Baazia, Y., Mahmoudi, A., Roulot, D., Ganne-Carrie, N., Grando-Lemaire, V., Trinchet, J. C., Gordien, E., Vicaut, E., Baghad, I. & Beaugrand, M. (2011). HCV genotype 3 is associated with a higher hepatocellular carcinoma incidence in patients with ongoing viral C cirrhosis. *J Viral Hepat* **18**, e516-522.

Noppornpanth, S., Lien, T. X., Poovorawan, Y., Smits, S. L., Osterhaus, A. D. & Haagmans, B. L. (2006). Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J Virol* **80**, 7569-7577.

Norton, P. A., Conyers, B., Gong, Q., Steel, L. F., Block, T. M. & Mehta, A. S. (2005). Assays for glucosidase inhibitors with potential antiviral activities: secreted alkaline phosphatase as a surrogate marker. *Journal of Virological Methods* **124**, 167-172.

Oem, J. K., Jackel-Cram, C., Li, Y. P., Kang, H. N., Zhou, Y., Babiuk, L. A. & Liu, Q. (2008a). Hepatitis C virus non-structural protein-2 activates CXCL-8 transcription through NF-kappaB. *Arch Virol* **153**, 293-301. Epub 2007 Dec 2013.

Oem, J. K., Jackel-Cram, C., Li, Y. P., Zhou, Y., Zhong, J., Shimano, H., Babiuk, L. A. & Liu, Q. (2008b). Activation of sterol regulatory element-binding protein 1c and fatty acid synthase transcription by hepatitis C virus non-structural protein 2. *J Gen Virol* **89**, 1225-1230.

Ohto, H., Terazawa, S., Sasaki, N., Sasaki, N., Hino, K., Ishiwata, C., Kako, M., Ujiie, N., Endo, C., Matsui, A. & et al. (1994). Transmission of hepatitis C virus from mothers to infants. The Vertical Transmission of Hepatitis C Virus Collaborative Study Group. *N Engl J Med* **330**, 744-750.

Op De Beeck, A., Montserret, R., Duvet, S., Cocquerel, L., Cacan, R., Barberot, B., Le Maire, M., Penin, F. & Dubuisson, J. (2000). The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J Biol Chem* **275**, 31428-31437.

Ott, C. M. & Lingappa, V. R. (2002). Integral membrane protein biosynthesis: why topology is hard to predict. *J Cell Sci* **115**, 2003-2009.

Oxford, J. S. (1975). Inhibition of the replication of influenza A and B viruses by a nucleoside analogue (ribavirin). *J Gen Virol* **28**, 409-414.

Pavio, N., Meng, X. J. & Renou, C. (2010). Zoonotic hepatitis E: animal reservoirs and emerging risks. *Vet Res* **41**, 46.

Pavlovic, D., Neville, D. C., Argaud, O., Blumberg, B., Dwek, R. A., Fischer, W. B. & Zitzmann, N. (2003). The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc Natl Acad Sci U S A* **100**, 6104-6108.

Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D. & Pawlotsky, J. M. (2004). Structural biology of hepatitis C virus. *Hepatology* **39**, 5-19.

Perz, J. F., Farrington, L. A., Pecoraro, C., Hutin, Y. J. F. & Armstrong, G. L. (2004). Estimated global prevalence of hepatitis C virus infection. In *42nd Annual Meeting of the Infectious Diseases Society of America*. Boston, MA, USA.

Pfannkuche, A., Buther, K., Karthe, J., Poenisch, M., Bartenschlager, R., Trilling, M., Hengel, H., Willbold, D., Haussinger, D. & Bode, J. G. (2011). c-Src is required for complex formation between the hepatitis C virus-encoded proteins NS5A and NS5B: a prerequisite for replication. *Hepatology* **53**, 1127-1136.

Phan, T., Beran, R. K., Peters, C., Lorenz, I. C. & Lindenbach, B. D. (2009). Hepatitis C virus NS2 protein contributes to virus particle assembly via opposing epistatic interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *J Virol* **83**, 8379-8395. Epub 2009 Jun 8310.

Piccininni, S., Varaklioti, A., Nardelli, M., Dave, B., Raney, K. D. & McCarthy, J. E. (2002). Modulation of the hepatitis C virus RNA-dependent RNA polymerase activity by the non-structural (NS) 3 helicase and the NS4B membrane protein. *J Biol Chem* **277**, 45670-45679.

Pieroni, L., Santolini, E., Fipaldini, C., Pacini, L., Migliaccio, G. & La Monica, N. (1997). In vitro study of the NS2-3 protease of hepatitis C virus. *J Virol* **71**, 6373-6380.

Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M., Cosset, F. L. & Bartenschlager, R. (2006). Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc Natl Acad Sci U S A* **103**, 7408-7413. Epub 2006 May 7401.

Pietschmann, T., Lohmann, V., Kaul, A., Krieger, N., Rinck, G., Rutter, G., Strand, D. & Bartenschlager, R. (2002). Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. *J Virol* **76**, 4008-4021.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G. & Abrignani, S. (1998). Binding of hepatitis C virus to CD81. *Science* **282**, 938-941.

Ploss, A., Evans, M. J., Gaysinskaya, V. A., Panis, M., You, H., de Jong, Y. P. & Rice, C. M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457**, 882-886.

Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *Embo J* **8**, 3867-3874.

Popescu, C. I., Callens, N., Trinel, D., Roingeard, P., Moradpour, D., Descamps, V., Duverlie, G., Penin, F., Heliot, L., Rouille, Y. & Dubuisson, J. (2011). NS2 protein of hepatitis C virus interacts with structural and non-structural proteins towards virus assembly. *PLoS Pathog* **7**, e1001278.

Post, J. J., Pan, Y., Freeman, A. J., Harvey, C. E., White, P. A., Palladinetti, P., Haber, P. S., Marinos, G., Levy, M. H., Kaldor, J. M., Dolan, K. A., Ffrench, R. A., Lloyd, A. R. & Rawlinson, W. D. (2004). Clearance of hepatitis C viremia associated with cellular immunity in the absence of seroconversion in the hepatitis C incidence and transmission in prisons study cohort. *J Infect Dis* **189**, 1846-1855.

Purcell, R. H. & Emerson, S. U. (2008). Hepatitis E: an emerging awareness of an old disease. *J Hepatol* **48**, 494-503.

Qin, W., Luo, H., Nomura, T., Hayashi, N., Yamashita, T. & Murakami, S. (2002). Oligomeric interaction of hepatitis C virus NS5B is critical for catalytic activity of RNA-dependent RNA polymerase. *J Biol Chem* **277**, 2132-2137.

Qin, W., Yamashita, T., Shirota, Y., Lin, Y., Wei, W. & Murakami, S. (2001). Mutational analysis of the structure and functions of hepatitis C virus RNA-dependent RNA polymerase. *Hepatology* **33**, 728-737.

Reed, K. E., Grakoui, A. & Rice, C. M. (1995). Hepatitis C virus-encoded NS2-3 protease: cleavage-site mutagenesis and requirements for bimolecular cleavage. *J Virol* **69**, 4127-4136.

Reynolds, G. M., Harris, H. J., Jennings, A., Hu, K., Grove, J., Lalor, P. F., Adams, D. H., Balfe, P., Hubscher, S. G. & McKeating, J. A. (2008). Hepatitis C virus receptor expression in normal and diseased liver tissue. *Hepatology* **47**, 418-427.

Reynolds, J. E., Kaminski, A., Kettinen, H. J., Grace, K., Clarke, B. E., Carroll, A. R., Rowlands, D. J. & Jackson, R. J. (1995). Unique features of internal initiation of hepatitis C virus RNA translation. *Embo J* **14**, 6010-6020.

Rinck, G., Birghan, C., Harada, T., Meyers, G., Thiel, H. J. & Tautz, N. (2001). A cellular J-domain protein modulates polyprotein processing and cytopathogenicity of a pestivirus. *J Virol* **75**, 9470-9482.

Rizzetto, M., Hoyer, B., Canese, M. G., Shih, J. W., Purcell, R. H. & Gerin, J. L. (1980). delta Agent: association of delta antigen with hepatitis B surface antigen and RNA in serum of delta-infected chimpanzees. *Proc Natl Acad Sci U S A* **77**, 6124-6128.

Romero-Lopez, C. & Berzal-Herranz, A. (2009). A long-range RNA-RNA interaction between the 5' and 3' ends of the HCV genome. *Rna* **15**, 1740-1752.

Rost, B., Casadio, R., Fariselli, P. & Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. *Protein Science*.

Sabahi, A. (2009). Hepatitis C Virus entry: the early steps in the viral replication cycle. *Virol J* **6**, 117.

Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y. & et al. (1990). Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc Natl Acad Sci U S A* **87**, 6547-6549.

Sakai, A., Claire, M. S., Faulk, K., Govindarajan, S., Emerson, S. U., Purcell, R. H. & Bukh, J. (2003). The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. *Proc Natl Acad Sci U S A* **100**, 11646-11651. Epub 12003 Sep 11622.

Sambrook, J. R. (2001). *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor Laboratory Press.

Santolini, E., Migliaccio, G. & La Monica, N. (1994). Biosynthesis and biochemical properties of the hepatitis C virus core protein. *J Virol* **68**, 3631-3641.

Santolini, E., Pacini, L., Fipaldini, C., Migliaccio, G. & Monica, N. (1995). The NS2 protein of hepatitis C virus is a transmembrane polypeptide. *J Virol* **69**, 7461-7471.

Sapay, N., Montserret, R., Chipot, C., Brass, V., Moradpour, D., Deleage, G. & Penin, F. (2006). NMR structure and molecular dynamics of the in-plane membrane anchor of nonstructural protein 5A from bovine viral diarrhea virus. *Biochemistry* **45**, 2221-2233.

Saraogi, I. & Shan, S. O. (1111). Molecular mechanism of co-translational protein targeting by the signal recognition particle. *Traffic* **12**, 535-542.

Saraogi, I. & Shan, S. O. (2011). Molecular mechanism of co-translational protein targeting by the signal recognition particle. *Traffic* **12**, 535-542.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. & Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J* **21**, 5017-5025.

Schaller, T., Appel, N., Koutsoudakis, G., Kallis, S., Lohmann, V., Pietschmann, T. & Bartenschlager, R. (2007). Analysis of hepatitis C virus superinfection exclusion by using novel fluorochrome gene-tagged viral genomes. *J Virol* **81**, 4591-4603.

Scheel, T. K., Gottwein, J. M., Jensen, T. B., Prentoe, J. C., Hoegh, A. M., Alter, H. J., Eugen-Olsen, J. & Bukh, J. (2008). Development of JFH1-based cell culture systems for hepatitis C virus genotype 4a and evidence for cross-genotype neutralization. *Proc Natl Acad Sci U S A* **105**, 997-1002.

Schmidt-Mende, J., Bieck, E., Hugle, T., Penin, F., Rice, C. M., Blum, H. E. & Moradpour, D. (2001). Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase. *J Biol Chem* **276**, 44052-44063.

Schregel, V., Jacobi, S., Penin, F. & Tautz, N. (2009). Hepatitis C virus NS2 is a protease stimulated by cofactor domains in NS3. *Proc Natl Acad Sci U S A* **106**, 5342-5347. Epub 2009 Mar 5312.

Schwer, B., Ren, S., Pietschmann, T., Kartenbeck, J., Kaehlcke, K., Bartenschlager, R., Yen, T. S. & Ott, M. (2004). Targeting of hepatitis C virus core protein to mitochondria through a novel C-terminal localization motif. *J Virol* **78**, 7958-7968.

Seeff, L. B. (2002). Natural history of chronic hepatitis C. *Hepatology* **36**, S35-46.

Selby, M. J., Glazer, E., Masiarz, F. & Houghton, M. (1994). Complex processing and protein:protein interactions in the E2:NS2 region of HCV. *Virology* **204**, 114-122.

Serebrov, V. & Pyle, A. M. (2004). Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature* **430**, 476-480.

Shesheer Kumar, M., Venkateswara Rao, K., Mohammed Habeebullah, C. & Dashavantha Reddy, V. (2008). Expression of alternate reading frame protein (F1) of hepatitis C virus in Escherichia coli and detection of antibodies for F1 in Indian patients. *Infect Genet Evol* **8**, 374-377.

Shimakami, T., Hijikata, M., Luo, H., Ma, Y. Y., Kaneko, S., Shimotohno, K. & Murakami, S. (2004). Effect of interaction between hepatitis C virus NS5A and NS5B on hepatitis C virus RNA replication with the hepatitis C virus replicon. *J Virol* **78**, 2738-2748.

Shimizu, Y. K., Feinstone, S. M., Purcell, R. H., Alter, H. J. & London, W. T. (1979). Non-A, non-B hepatitis: ultrastructural evidence for two agents in experimentally infected chimpanzees. *Science* **205**, 197-200.

Shirota, Y., Luo, H., Qin, W., Kaneko, S., Yamashita, T., Kobayashi, K. & Murakami, S. (2002). Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. *J Biol Chem* **277**, 11149-11155.

Shiryaev, S. A., Chernov, A. V., Shiryaeva, T. N., Aleshin, A. E. & Strongin, A. Y. (2010). The acidic sequence of the NS4A cofactor regulates ATP hydrolysis by the HCV NS3 helicase. *Arch Virol* **156**, 313-318.

Simister, P., Schmitt, M., Geitmann, M., Wicht, O., Danielson, U. H., Klein, R., Bressanelli, S. & Lohmann, V. (2009). Structural and functional analysis of hepatitis C virus strain JFH1 polymerase. *J Virol* **83**, 11926-11939.

Soto, B., Sanchez-Quijano, A., Rodrigo, L., del Olmo, J. A., Garcia-Bengoechea, M., Hernandez-Quero, J., Rey, C., Abad, M. A., Rodriguez, M., Sales Gilabert, M., Gonzalez, F., Miron, P., Caruz, A., Relimpio, F., Torronteras, R., Leal, M. & Lissen, E. (1997). Human immunodeficiency virus infection modifies the natural history of chronic parenterally-acquired hepatitis C with an unusually rapid progression to cirrhosis. *J Hepatol* **26**, 1-5.

Spangberg, K., Goobar-Larsson, L., Wahren-Herlenius, M. & Schwartz, S. (1999). The La protein from human liver cells interacts specifically with the U-rich region in the hepatitis C virus 3' untranslated region. *J Hum Virol* **2**, 296-307.

Stapleford, K. & Lindenbach, B. D. (2011). Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *J Virol* **85**, 1706-1717.

Stapleton, J. T., Foung, S., Muerhoff, A. S., Bukh, J. & Simmonds, P. (2011). The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *J Gen Virol* **92**, 233-246.

Steinmann, E., Penin, F., Kallis, S., Patel, A. H., Bartenschlager, R. & Pietschmann, T. (2007). Hepatitis C virus p7 protein is crucial for assembly and release of infectious virions. *PLoS Pathog* **3**, e103.

Stempniak, M., Hostomska, Z., Nodes, B. R. & Hostomsky, Z. (1997). The NS3 proteinase domain of hepatitis C virus is a zinc-containing enzyme. *J Virol* **71**, 2881-2886.

StGelais, C., Tuthill, T. J., Clarke, D. S., Rowlands, D. J., Harris, M. & Griffin, S. (2007). Inhibition of hepatitis C virus p7 membrane channels in a liposome-based assay system. *Antiviral Res* **76**, 48-58. Epub 2007 Jun 2004.

Sugano, S., Yoshitomo-Nakagawa, K., Yu, Y. S., Mizushima-Sugano, J. & Yoshida, K. (1998). Transmembrane-domain trapping: a novel method for isolation of cDNAs encoding putative membrane proteins. *DNA Res* **5**, 187-193.

Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrener, P., Grakoui, A., Rice, C. M., Feinstone, S. M. & Collett, M. S. (1993). Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J Virol* **67**, 6152-6158.

Svitkin, Y. V., Pause, A., Lopez-Lastra, M., Perreault, S. & Sonenberg, N. (2005). Complete translation of the hepatitis C virus genome in vitro: membranes play a critical role in the maturation of all virus proteins except for NS3. *J Virol* **79**, 6868-6881.

Swain, M. G., Lai, M.-Y., Shiftman, M. L., Cooksley, W. G. E., Abergel, A., Lin, A., Connell, E. & Diag, M. (2007). Durable sustained virological response after treatment with peginterferon alpha-2a (PEGASYSR) alone or in combination with ribavirin (COPEGUSR): 5-year follow-up and the criteria of a cure. *Journal of Hepatology* **48**, S23.

Tabor, E., Gerety, R. J., Drucker, J. A., Seeff, L. B., Hoofnagle, J. H., Jackson, D. R., April, M., Barker, L. F. & Pineda-Tamondong, G. (1978). Transmission of non-A, non-B hepatitis from man to chimpanzee. *Lancet* **1**, 463-466.

Tai, C. L., Chi, W. K., Chen, D. S. & Hwang, L. H. (1996). The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J Virol* **70**, 8477-8484.

Tam, A. W., Smith, M. M., Guerra, M. E., Huang, C. C., Bradley, D. W., Fry, K. E. & Reyes, G. R. (1991). Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* **185**, 120-131.

Tanaka, T., Kato, N., Cho, M. J. & Shimotohno, K. (1995). A novel sequence found at the 3' terminus of hepatitis C virus genome. *Biochem Biophys Res Commun* **215**, 744-749.

Tanaka, T., Kato, N., Cho, M. J., Sugiyama, K. & Shimotohno, K. (1996). Structure of the 3' terminus of the hepatitis C virus genome. *J Virol* **70**, 3307-3312.

Tang, Q. H., Zhang, Y. M., Fan, L., Tong, G., He, L. & Dai, C. (2010). Classic swine fever virus NS2 protein leads to the induction of cell cycle arrest at S-phase and endoplasmic reticulum stress. *Virol J* **7**, 4.

Targett-Adams, P., Hope, G., Boulant, S. & McLauchlan, J. (2008). Maturation of hepatitis C virus core protein by signal peptide peptidase is required for virus production. *J Biol Chem* **283**, 16850-16859.

Tector, M. & Hartl, F. U. (1999). An unstable transmembrane segment in the cystic fibrosis transmembrane conductance regulator. *Embo J* **18**, 6290-6298.

Tedbury, P., Welbourn, S., Pause, A., King, B., Griffin, S. & Harris, M. (2011). The subcellular localization of the hepatitis C virus non-structural protein NS2 is regulated by an ion channel-independent function of the p7 protein. *J Gen Virol* **92**, 819-830.

Tedbury, P. R. & Harris, M. (2007). Characterisation of the role of zinc in the hepatitis C virus NS2/3 auto-cleavage and NS3 protease activities. *J Mol Biol* **366**, 1652-1660. Epub 2006 Dec 1628.

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

Tellinghuisen, T. L., Foss, K. L., Treadaway, J. C. & Rice, C. M. (2008b). Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J Virol* **82**, 1073-1083.

Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E. & Rice, C. M. (2004). The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem* **279**, 48576-48587. Epub 42004 Aug 48531.

Tellinghuisen, T. L., Marcotrigiano, J. & Rice, C. M. (2005). Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature* **435**, 374-379.

Thibeault, D., Maurice, R., Pilote, L., Lamarre, D. & Pause, A. (2001). In vitro characterization of a purified NS2/3 protease variant of hepatitis C virus. *J Biol Chem* **276**, 46678-46684. Epub 42001 Oct 46678.

Thorley, J. A., McKeating, J. A. & Rappoport, J. Z. (2010). Mechanisms of viral entry: sneaking in the front door. *Protoplasma* **244**, 15-24.

Tomei, L., Failla, C., Santolini, E., De Francesco, R. & La Monica, N. (1993). NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J Virol* **67**, 4017-4026.

Tomei, L., Failla, C., Vitale, R. L., Bianchi, E. & De Francesco, R. (1996). A central hydrophobic domain of the hepatitis C virus NS4A protein is necessary and sufficient for the activation of the NS3 protease. *J Gen Virol* **77 (Pt 5)**, 1065-1070.

Tretter, V., Altmann, F. & Marz, L. (1991). Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem* **199**, 647-652.

Tsuchihara, K., Tanaka, T., Hijikata, M., Kuge, S., Toyoda, H., Nomoto, A., Yamamoto, N. & Shimotohno, K. (1997). Specific interaction of polypyrimidine tract-binding protein with the extreme 3'-terminal structure of the hepatitis C virus genome, the 3'X. *J Virol* **71**, 6720-6726.

Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. & Nomoto, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. *J Virol* **66**, 1476-1483.

Tusnady, G. E. & Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* **283**, 489-506.

van Klompenburg, W., Nilsson, I., von Heijne, G. & de Kruijff, B. (1997). Anionic phospholipids are determinants of membrane protein topology. *Embo J* **16**, 4261-4266.

Verdegem, D., Badillo, A., Wieruszeski, J. M., Landrieu, I., Leroy, A., Bartenschlager, R., Penin, F., Lippens, G. & Hanoulle, X. (2011). Domain 3 of NS5A protein from the

hepatitis C virus has intrinsic alpha-helical propensity and is a substrate of cyclophilin A. *J Biol Chem* **286**, 20441-20454.

Verstrepen, B. E., Depla, E., Rollier, C. S., Mares, G., Drexhage, J. A., Priem, S., Verschoor, E. J., Koopman, G., Granier, C., Dreux, M., Cosset, F. L., Maertens, G. & Heeney, J. L. (2011). Clearance of genotype 1b hepatitis C virus in chimpanzees in the presence of vaccine-induced e1-neutralizing antibodies. *J Infect Dis* **204**, 837-844.

Vieyres, G., Thomas, X., Descamps, V., Duverlie, G., Patel, A. H. & Dubuisson, J. (2010). Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. *J Virol* **84**, 10159-10168.

Vogel, M. & Rockstroh, J. K. (2010). Treatment of acute hepatitis C in HIV infection. *J Antimicrob Chemother* **65**, 4-9.

Voisset, C., Callens, N., Blanchard, E., Op De Beeck, A., Dubuisson, J. & Vu-Dac, N. (2005). High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. *J Biol Chem* **280**, 7793-7799.

Vos, J., Spee, P., Momburg, F. & Neefjes, J. (1999). Membrane topology and dimerization of the two subunits of the transporter associated with antigen processing reveal a three-domain structure. *J Immunol* **163**, 6679-6685.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H. G., Mizokami, M., Bartenschlager, R. & Liang, T. J. (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**, 791-796. Epub 2005 Jun 2012.

Walewski, J. L., Keller, T. R., Stump, D. D. & Branch, A. D. (2001). Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**, 710-721.

Wardell, A. D., Errington, W., Ciaramella, G., Merson, J. & McGarvey, M. J. (1999). Characterization and mutational analysis of the helicase and NTPase activities of hepatitis C virus full-length NS3 protein. *J Gen Virol* **80 (Pt 3)**, 701-709.

Washburn, M. L., Bility, M. T., Zhang, L., Kovalev, G. I., Buntzman, A., Frelinger, J. A., Barry, W., Ploss, A., Rice, C. M. & Su, L. (2011). A humanized mouse model to study hepatitis C virus infection, immune response, and liver disease. *Gastroenterology* **140**, 1334-1344.

Watashi, K., Ishii, N., Hijikata, M., Inoue, D., Murata, T., Miyanari, Y. & Shimotohno, K. (2005). Cyclophilin B is a functional regulator of hepatitis C virus RNA polymerase. *Mol Cell* **19**, 111-122.

Wedemeyer, H., Wiegand, J., Cornberg, M. & Manns, M. P. (2002). Polyethylene glycol-interferon: current status in hepatitis C virus therapy. *J Gastroenterol Hepatol* **17 Suppl 3**, S344-350.

Welbourn, S., Green, R., Gamache, I., Dandache, S., Lohmann, V., Bartenschlager, R., Meerovitch, K. & Pause, A. (2005). Hepatitis C virus NS2/3 processing is required for NS3 stability and viral RNA replication. *J Biol Chem* **280**, 29604-29611.

Welbourn, S., Jirasko, V., Breton, V., Reiss, S., Penin, F., Bartenschlager, R. & Pause, A. (2009). Investigation of a role for lysine residues in non-structural proteins 2 and 2/3 of the hepatitis C virus for their degradation and virus assembly. *J Gen Virol* **90**, 1071-1080.

WHO (2002). Global distribution of hepatitis A, B and C, 2001.

WHO (2009). Global distribution of HCV genotypes.

WHO (2011a). Hepatitis A.

WHO (2011b). Hepatitis B.

Wilkinson, J., Radkowski, M. & Laskus, T. (2009). Hepatitis C virus neuroinvasion: identification of infected cells. *J Virol* **83**, 1312-1319.

Wolk, B., Sansonno, D., Krausslich, H. G., Dammacco, F., Rice, C. M., Blum, H. E. & 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**, 2293-2304.

Wozniak, A. L., Griffin, S., Rowlands, D., Harris, M., Yi, M., Lemon, S. M. & Weinman, S. A. (2010). Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production. *PLoS Pathog* **6**, e1001087.

Xu, Z., Choi, J., Yen, T. S., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M. J. & Ou, J. (2001). Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *Embo J* **20**, 3840-3848.

Yamada, N., Tanihara, K., Takada, A., Yorihuzi, T., Tsutsumi, M., Shimomura, H., Tsuji, T. & Date, T. (1996). Genetic organization and diversity of the 3' noncoding region of the hepatitis C virus genome. *Virology* **223**, 255-261.

Yamaga, A. K. & Ou, J. H. (2002). Membrane topology of the hepatitis C virus NS2 protein. *J Biol Chem* **277**, 33228-33234. Epub 32002 Jun 33224.

Yamashita, T., Kaneko, S., Shirota, Y., Qin, W., Nomura, T., Kobayashi, K. & Murakami, S. (1998). RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus NS5B protein truncated at the C-terminal region. *J Biol Chem* **273**, 15479-15486.

Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J. L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C. & Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 A resolution structure in a hexagonal crystal form. *Protein Sci* **7**, 837-847.

Yang, D., Liu, N., Zuo, C., Lei, S., Wu, X., Zhou, F., Liu, C. & Zhu, H. (2011). Innate host response in primary human hepatocytes with hepatitis C virus infection. *PLoS One* **6**, e27552.

Yang, F., Robotham, J. M., Nelson, H. B., Irsigler, A., Kenworthy, R. & Tang, H. (2008). Cyclophilin A is an essential cofactor for hepatitis C virus infection and the principal mediator of cyclosporine resistance in vitro. *J Virol* **82**, 5269-5278.

Yang, G., Pevear, D. C., Collett, M. S., Chunduru, S., Young, D. C., Benetatos, C. & Jordan, R. (2004). Newly synthesized hepatitis C virus replicon RNA is protected from nuclease activity by a protease-sensitive factor(s). *J Virol* **78**, 10202-10205.

Yang, X. J., Liu, J., Ye, L., Liao, Q. J., Wu, J. G., Gao, J. R., She, Y. L., Wu, Z. H. & Ye, L. B. (2006). HCV NS2 protein inhibits cell proliferation and induces cell cycle arrest in the S-phase in mammalian cells through down-regulation of cyclin A expression. *Virus Res* **121**, 134-143.

Yi, M., Ma, Y., Yates, J. & Lemon, S. M. (2007). Compensatory mutations in E1, p7, NS2, and NS3 enhance yields of cell culture-infectious intergenotypic chimeric hepatitis C virus. *J Virol* **81**, 629-638.

Yi, M., Ma, Y., Yates, J. & Lemon, S. M. (2009). Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation. *PLoS Pathog* **5**, e1000403.

Yi, M., Villanueva, R. A., Thomas, D. L., Wakita, T. & 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**, 2310-2315.

Yokosuka, O., Kojima, H., Imazeki, F., Tagawa, M., Saisho, H., Tamatsukuri, S. & Omata, M. (1999). Spontaneous negativation of serum hepatitis C virus RNA is a rare event in type C chronic liver diseases: analysis of HCV RNA in 320 patients who were followed for more than 3 years. *J Hepatol* **31**, 394-399.

Yu, K. L., Jang, S. I. & You, J. C. (2009). Identification of in vivo interaction between Hepatitis C Virus core protein and 5' and 3' UTR RNA. *Virus Res* **145**, 285-292.

Zeuzem, S., Asselah, T., Angus, P., Zarski, J. P., Larrey, D., Mullhaupt, B., Gane, E., Schuchmann, M., Lohse, A., Pol, S., Bronowicki, J. P., Roberts, S., Arasteh, K., Zoulim,

F., Heim, M., Stern, J. O., Kukolj, G., Nehmiz, G., Haefner, C. & Boecher, W. O. (2011). Efficacy of the Protease Inhibitor BI 201335, Polymerase Inhibitor BI 207127, and Ribavirin in Patients with Chronic HCV infection. *Gastroenterology*, 14.

Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D. R., Wieland, S. F., Uprichard, S. L., Wakita, T. & Chisari, F. V. (2005). Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* **102**, 9294-9299. Epub 2005 Jun 9296.

Zhong, W., Uss, A. S., Ferrari, E., Lau, J. Y. & Hong, Z. (2000). De novo initiation of RNA synthesis by hepatitis C virus nonstructural protein 5B polymerase. *J Virol* **74**, 2017-2022.

CHAPTER 8: APPENDICES

Appendix 1. Oligonucleotides used for cloning.

Primer	Sequence 5'-3'	Purpose
Wt SEAP XhoI f	NNNNCTCGAGGCGGCCATGGCCATCATCCCAGTTGAGGAGGAG AACC	SEAP reporter controls
Wt SEAP ApaI r	NNNNGGGCCCTTATGTCTGCTCGAAGCGGCCGGC	SEAP reporter controls
Wt LacZ XhoI f	NNNNCTCGAGGCGGCCATGGCCACCATGATTACGGATTCACTGG	β-gal reporter controls
Wt LacZ ApaI r	NNNNGGGCCCTTATTTTTGACACCAGACCAACTGG	β-gal reporter controls
SPp7 HindIII f	NNNNAAGCTTGCGGCCATGGCCAGAGTCTGCGCCTGCTTGTGG	C-terminal truncations
Furin-LacZ XhoI f	NNNNCTCGAGGGCGGAAGCAGGGG**T**CGACGCGGCGGAAGCAC CATGATTACGGATTCACTGG	β-gal reporter fusions
Furin-SEAP XhoI f	NNNNCTCGAGGGCGGAAGCAGGGG**T**CGACGCGGCGGAAGCAT CATCCCAGTTGAGGAGGAGAACC	SEAP reporter fusions
SEAP (no linker) XhoI f	NNNNCTCGAGATCATCCCAGTTGAGGAGGAGAACC	SEAP control fusions
NS2 aa 217 XhoI r	NNNNCTCGAGAAGGAGCTTCCACCCCTTGG	C-terminal truncations
NS2 aa 120 XhoI r	NNNNCTCGAGCGCGAGCTGCTTCACCAAAGC	C-terminal truncations
NS2 aa 110 XhoI r	NNNNCTCGAGTATCAGAGCGTGAGCTCTGACG	C-terminal truncations
NS2 aa 95 XhoI r	NNNNCTCGAGGGCCCTTAAGAGGTAAGCAGG	C-terminal truncations
NS2 aa 87 XhoI r	NNNNCTCGAGAAGCAACGCCAAAAGCCATTTGG	C-terminal truncations
NS2 aa 70 XhoI r	NNNNCTCGAGTATAGTGACGGCCCACGCGATGC	C-terminal truncations
NS2 aa 52 XhoI r	NNNNCTCGAGTACCCACTCCTGAATCATGGC	C-terminal truncations
NS2 aa 40 XhoI r	NNNNCTCGAGGAGATAGCACAACCACCACAGACACTGG	C-terminal truncations
NS2 aa 27 XhoI r	NNNNCTCGAGCTTATACCCCGGGGTGAGTGTGAAGAGG	C-terminal truncations
P7 XhoI r	NNNNCTCGAGGGCATAAGCCTGCCGGGGCAGTGCC	C-terminal truncations
QC furin mutation f	GAGGGCGGAAGCGCGGGTCGACGCG	Generation of S12 and S13
QC furin mutation r	CGCGTCGACCCGCGCTTCCGCCCTC	Generation of S12 and S13
QC signal peptidase mutation f	CGGCAGGCTTATCCCCTCGAGGGCG	Generation of S11 and S13

Primer	Sequence 5'-3'	Purpose
QC signal peptidase mutation r	CGCCCTCGAGGGGATAAGCCTGCCG	Generation of S11 and S13
P7 aa 62 XhoI r	NNNNCTCGAGATAAGCCTGCCGGGGCAGTGCC	Generation of delta 63 (S10 derivative)
Signal peptide oligo f	AGCTTGCGGCCATGGCCCTGCTGCTGCTGCTGCTGCTGGGCCTGAGGCTACAGCTCTCCCTGGGCC	Generation of SP-β-gal
Signal peptide oligo r	TCGAGGCCCAGGGAGAGCTGTAGCCTCAGGCCCAGCAGCAGCAGCAGCAGGGCCATGGCCGCA	Generation of SP-β-gal
BK40(XhoI-mfur)F Forward: 5'	CTCGAGGGCGGAAGCGCGGGTCGACGCG	
NS2 aa 1 EcoRI f	GAATTCGCGGCCATGTATGACGCACCTGTGCACGG	NS2-eGFP fusions
NS2 aa 25 EcoRI f	GAATTCGCGGCCATGGGGTATAAGACCCTCCTCG	NS2-eGFP fusions
NS2 aa 51 EcoRI f	GAATTCGCGGCCATGTGGGTACCACCCATGCAGG	NS2-eGFP fusions
NS2 aa 93 EcoRI f	GAATTCGCGGCCATGTTAAGGGCCGCTTTGACAC	NS2-eGFP fusions
NS2 aa 24 BamHI r	GGATCCCGGGGTGAGTGTGAAGAGG	NS2-eGFP fusions
NS2 aa 50 BamHI r	GGATCCCTCCTGAATCATGGCTTCC	NS2-eGFP fusions
NS2 aa 92 BamHI r	GGATCCCCGAGGTAAGCAGGCCCAAGC	NS2-eGFP fusions
NS2 aa 217 BamHI r	GGATCCAAGGAGCTTCCACCCCTTGG	NS2-eGFP fusions
SEAP (no linker) XhoI f	NNNNCTCGAGATCATCCCAGTTGAGGAGGAGAACC	SEAP control fusions
NS2 aa 1 XhoI f	NNNNCTCGAGCTATGACGCACCTGTGCACG	eGFP-NS2 fusions
NS2 aa 25 XhoI f	NNNNCTCGAGCGGGTATAAGACCCTCCTCG	eGFP-NS2 fusions
NS2 aa 51 XhoI f	NNNNCTCGAGCTGGGTACCACCCATGCAGG	eGFP-NS2 fusions
NS2 aa 93 XhoI f	NNNNCTCGAGCTTAAGGGCCGCTTTGACAC	eGFP-NS2 fusions
NS2 aa 24 EcoRI r	NNNNGAATTCTTACGGGGTGAGTGTGAAGAGG	eGFP-NS2 fusions
NS2 aa 50 EcoRI r	NNNNGAATTCTTACTCCTGAATCATGGCTTCC	eGFP-NS2 fusions
NS2 aa 92 EcoRI r	NNNNGAATTCTTAGAGGTAAGCAGGCCCAAGC	eGFP-NS2 fusions

Primer	Sequence 5'-3'	Purpose
NS2 aa 217 EcoRI r	NNNNGAATTCTTAAAGGAGCTTCCACCCCTTGG	eGFP-NS2 fusions
Flag-tag oligo f	P-GATCCAGGCGGAAGCGGCGGAAGCGACTACAAAGACG ATGACGACAAGTAACCGC	NS2-FLAG fusions
Flag-tag oligo r	P-GGTTACTTGTCGTCATCGTCTTTGTAGTCGCTTCCGCCGC TTCCGCCTG	NS2-FLAG fusions
NS2 delta aa 25-50 f	ACCCTCTTCACACTCACCCCGTGGGTACCACCCATGCAGG	NS2 deletion fusions
NS2 delta aa 25-50 r	CCTGCATGGGTGGTACCCACGGGGTGAGTGTGAAGAGGGT	NS2 deletion fusions
NS2 delta aa 25-92 f	ACCCTCTTCACACTCACCCCGTTAAGGGCCGCTTTGACAC	NS2 deletion fusions
NS2 delta aa 25-92 r	GTGTCAAAGCGGCCCTTAACGGGGTGAGTGTGAAGAGGGT	NS2 deletion fusions
NS2 delta aa 51-92 f	GGGAAGCCATGATTCAGGAGTTAAGGGCCGCTTTGACAC	NS2 deletion fusions
NS2 delta aa 51-92 r	GTGTCAAAGCGGCCCTTAACTCCTGAATCATGGCTTCCC	NS2 deletion fusions
QC peGFP-N1 ATG f	CACCGGTCGCCACCGCGGTGAGCAAGGGCG	peGFP-N1 QC control mutant
QC peGFP-N1 ATG r	CGCCCTTGCTCACCGCGGTGGCGACCGGTG	peGFP-N1 QC control mutant
QC peGFP-C2 stop f	CGGACTCAGATCTCGAGCTAAAGCTTCGAATTCTGCAGTC	peGFP-C2 STOP control mutant
QC peGFP-C2 stop r	GACTGCAGAATTCGAAGCTTTAGCTCGAGATCTGAGTCCG	peGFP-C2 STOP control mutant
SEAP r	NNNNTGGCGGCTGTCTGTGCAGGCTGC	Sequencing Primer
Lac Z f1	NNNNGCCACTCGCTTTAATGATG	Sequencing Primer
Lac Z f2	NNNNGGCCACCGATATTATTTGC	Sequencing Primer

Appendix 2. Amino acid nomenclature and hydropathy scoring.

Amino acid	3-letter code	1-letter code	Kyte-Doolittle score
Alanine	Ala	A	1.8
Arginine	Arg	R	-4.5
Asparagine	Asn	N	-3.5
Aspartic Acid	Asp	D	-3.5
Cysteine	Cys	C	2.5
Glutamic Acid	Glu	E	-3.5
Glutamine	Gln	Q	-3.5
Glycine	Gly	G	-0.4
Histidine	His	H	-3.2
Isoleucine	Ile	I	4.5
Leucine	Leu	L	3.8
Lysine	Lys	K	-3.9
Methionine	Met	M	1.9
Phenylalanine	Phe	F	2.8
Proline	Pro	P	-1.6
Serine	Ser	S	-0.8
Threonine	Thr	T	-0.7
Tryptophan	Trp	W	-0.9
Tyrosine	Tyr	Y	-1.3
Valine	Val	V	4.2

Amino acid names and three and single letter abbreviations. Hydropathy scores for each residue are those assigned by Kyte-Doolittle (Kyte & Doolittle, 1982).

Appendix 3. Prediction of N-linked glycosylation sites (N-X-S/T).

Sequences were analysed using the NetNGlyc 1.0 Server.

β-Galactosidase amino acid sequence:

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWRFAWF
PAPEAVPESWLECDLPEADTVVVPSNWQMHGYDAPIYTN[103]VTYPITVNPPFVPTEN[119]PTGCY
SLTFNVDESWLQEGQTRIIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRAGENRLAVM
VLRWSDGSYLEDQDMWRMSGIFRDVSLLHKPTTQISDFHVATRFNDDFSRAVLEAEVQMCGELR
DYLRVTVSLWQGETQVASGTAPFGGEIIDERGGYADRVTLRLNVENPKLWSAEIPNLYRAVVEL
HTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLIRGVNRHEHHPLHGQVMDEQTMVQDILLM
KQNNFNAVRCSHYPNHPLWYTLCDRYGLYVVDEANIETHGMVPMNRLTDDPRWLPAMSERVTRM
VQRDRNHPSVIIWSLGN[461]ESGHGANHDALYRWIKSVDPSRPVQYEGGGADTTATDIICPMYAR
VDEDQPFPAVPKWSIKKWLSLPGETRPLILCEYAHAMGNSLGGFAKYWQAFRQYPRLQGGFVWD
WVDQSLIKYDENGNPWSAYGGDFGDTPNDRQFCMNGLVFADRTPHPALTEAKHQQQFFQFRLSG
QTIEVTSEYLFRHSDNELLHWMVALDGKPLASGEVPLDVAPQGKQLIELPELPQPESAGQLWLT
VRVVQPNATAWSEAGHISAWQQWRLAEN[726]LSVTLPAASHAIPHLTTSEMDFCIELGNKRWQFN
RQSGFLSQMWIGDKKQLLTPLRDQFTRAPLDNDIGVSEATRIDPNAWVERWKAAGHYQAEAALL
QCTADTLADAVLITTAHAWQHQGKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQ
LAQVAERVNWLGLGPQENYPDRLTAACFDRWDLPLSDMYTPYVFPSENGLRCGTRELNYGPHQW
RGDFQFNISRYSQQQLMETSHRHLLHAEEGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYH
YQLVWCQK

Probablility of N-glycosylation:

N103 - 0.8114, N119 - 0.6079, N461 - 0.5674, N726 - 0.7344

SEAP amino acid sequence:

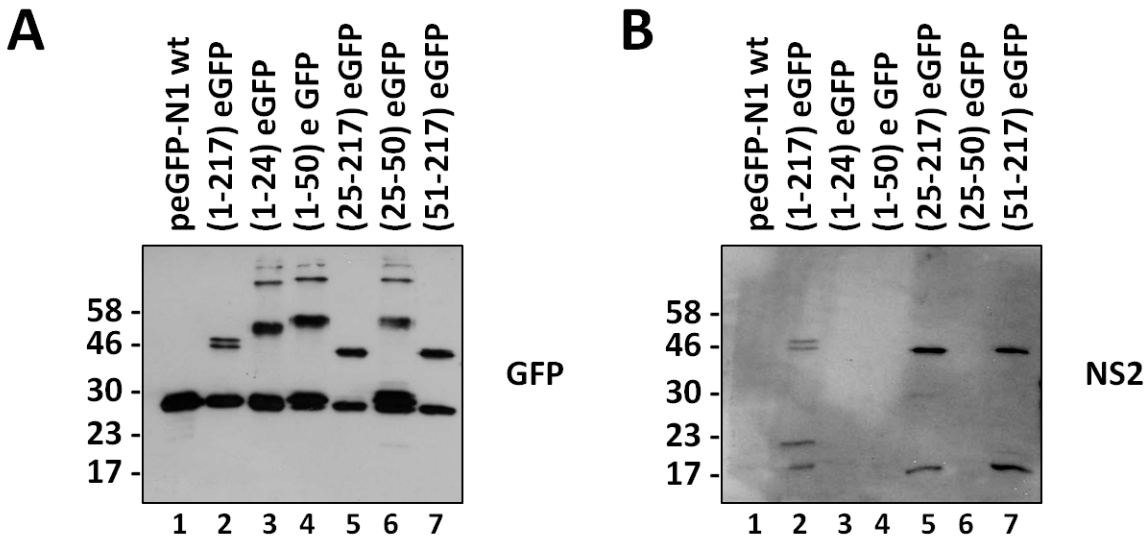
IIPVEEENPDFWNREAAEALGAAKKLQPAQTAAKNLIIFLGDGMGVSTVTAARILKGQKKDKLG
PEIPLAMDRFPYVALSKTYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFNQCN[122]TTRG
NEVISVMNRAKKAGKSVGVVTTTRVQHASPAGTYAHTVNRNWYSDADVPASARQEGCQDIATQL
ISNMDIDVILGGGRKYMFRMGTPDPEYPDDYSQGGTRLDGKNLVQEWLAKRQGARYVWN[249]RTE
LMQASLDPSVTHLMGLFEPGDMKYEIHRDSTLDPSLMEMTEAALRLLSRNPRGFFLFVEGGRID
HGHHESRAYRALTETIMFDDAIERAGQLTSEEDTLSLVTADHSHVFSFGGYPLRGSSIFGLAPG
KARDRKAYTVLLYGNGPGYVLKDGARPDVTESESGSPEYRQQSAVPLDEETHAGEDVAVFARGP
QAHLVHGVQEQTFIAHVMAFAACLEPYTACDLAPPAGTTDAAHPG

Probablility of N-glycosylation:

N122 - 0.5276, N249 - 0.5333

Only N249 is thought to be N-glycosylated in humans (Endo *et al.*, 1988; Le Du *et al.*, 2001).

Appendix 4. NS2-eGFP truncations produce 'cleaved' fusion products.



COS7 cells were transfected with a selection of NS2-eGFP truncations (lacking the .QC mutation of the eGFP initiation site). Lysates were collected 48 h post-transfection and analysed by SDS PAGE and immuno-blotting. Membranes were probed with anti-GFP antibody (**A**) and antiserum raised against the NS2 catalytic domain (aa 94-217) (**B**). Oligomerisation of (1-24), (25-50) and (1-50) was not seen in any subsequent experiments. The observation here was thought to be caused by non-specific aggregation in the absence of sufficient reducing agent. Molecular weight markers are shown to the left and lanes are marked below.