From THE DEPARTMENT OF MEDICINE
Karolinska Institutet, Stockholm, Sweden

TOPOLOGY AND MEMBRANE REARRANGEMENTS OF THE HEPATITIS C VIRUS PROTEIN NS4B

Marika Lundin

Stockholm 2006
ABSTRACT

Hepatitis C virus (HCV) is a major cause of chronic liver disease and viral hepatitis around the world. Approximately 180 million people or 3% of the world’s population are infected with the virus. The infection only rarely give rise to any acute symptoms, but more than 70% of those infected will develop chronic hepatitis with the risk of developing liver cirrhosis and liver cancer. The only treatment available is a combination therapy of pegylated interferon and the broad spectrum antiviral agent ribavirin. All patients cannot be subjected to this treatment however, and of those treated only about 50% (with variations for different genotypes) will clear the virus. A better understanding of HCV and its life cycle is therefore needed to achieve a more efficient and well tolerated therapy against HCV infection.

In order to improve the understanding of HCV, we have focused on one of the least known of the viral proteins - the non structural protein 4B (NS4B). NS4B is a relatively small protein (27 KDa) of mostly unknown functions. It is believed however to play an important, but yet unknown part in the viral replication process. By employing fluorescence microscopy we found that NS4B located in the endoplasmic reticulum (ER) membrane, and that it also had the ability to rearrange the membranes into dense aggregates in the cytoplasm. These new structures are called membrane associated foci (MAF) and are believed to be the locale for viral replication. They are assumed to be the same structures, or closely related, as the structure called “membraneous web” found with electron microscopy.

In our studies we also found that NS4B is an integral ER membrane protein. By computer predictions and glycosylation mapping we generated a model where the protein initially has four transmembrane segments with the C- and the N-termini in the cytoplasm. After translation and processing from the polyprotein however, the N-terminus of NS4B translocates to the ER lumen, giving the protein a fifth transmembrane segment. In the presence of another viral protein, NS5A, a lower degree of translocation was observed, indicating that NS5A may influence the dual topology of NS4B. Immunoprecipitation showed us that NS4B could homo-oligomerize, and this ability seemed to correlate with induction of MAF structures. Furthermore, we obtained experimental data suggesting that the translocation of the N-terminus may be involved in this process too, in particular an amphipatic helix around aa 40-69.

Finally, based on our topology model we created 14 point mutations in the NS4B protein in the subgenomic replicon to assess what portions of NS4B that are important for replication. All mutants except one had negative effects on the replicon establishment efficiency, even changes located to the luminal loops of NS4B (replication occurs on the cytoplasmic side). Neither the polyprotein efficiency nor the ability to induce MAF were affected, supporting a role of NS4B in HCV replication besides providing the necessary environment for viral replication.
LIST OF PUBLICATIONS

This thesis is based on the following papers, which are referred to in the text by their roman numerals:


* Both authors have contributed equally
# CONTENTS

1 List of abbreviations ................................................................................. 6

2 Aims of the project ..................................................................................... 7

3 Introduction .................................................................................................. 9
   3.1 Hepatitis C virus ................................................................................... 9
      3.1.1 Prevalence and treatment of HCV infection ................................ 9
      3.1.2 The HCV genome organisation ................................................. 10
      3.1.3 The viral proteins ....................................................................... 11
      3.1.4 Model systems .......................................................................... 13
      3.1.5 The HCV life cycle ..................................................................... 15
      3.1.6 NS4B’s potential role in the viral life cycle and replication ....... 17
   3.2 Biological membranes and membrane proteins .............................. 19
      3.2.1 The eukaryotic cell .................................................................... 19
      3.2.2 Viruses - membrane hijackers .................................................. 19
      3.2.3 The ER membrane and translocating proteins ....................... 20
      3.2.4 Topology of membrane proteins .............................................. 23

4 Comments on material and methods .................................................... 26
   4.1 Studies of NS4B topology ................................................................. 26
      4.1.1 Topology prediction ................................................................... 26
      4.1.2 Experimentally verifying the predicted topology ....................... 27
   4.2 In vitro translation .............................................................................. 28
      4.2.1 Kinetics of glycosylation ......................................................... 28
   4.3 Choice of cell lines ............................................................................ 29
   4.4 The subgenomic replicon .................................................................. 30
      4.4.1 Choice of mutations .................................................................. 30
      4.4.2 Replicon establishment efficiency .......................................... 31
   4.5 NS4B oligomerization ....................................................................... 31
      4.5.1 The oligomerization assay ....................................................... 33
      4.5.2 Crosslinking .............................................................................. 33

5 Results and discussion ............................................................................. 34
   5.1 Localization and topology of NS4B (paper I-II) ............................... 34
      5.1.1 Localization ............................................................................... 34
      5.1.2 Topology ................................................................................... 35
   5.2 Mutational analysis of the different domains of NS4B and their role in replication (paper III) ............................................................ 38
   5.3 Correlation between homo-oligomerization, N-terminal translocation and MAF induction ................................................................. 40

6 Concluding remarks .................................................................................. 43

7 Acknowledgements ................................................................................... 45

8 References ................................................................................................. 47
1 LIST OF ABBREVIATIONS

Aa  amino acid
AH Amphipathic helix
BVDV bovine viral diarrhea virus
DSS Disucinimidyl suberate
EGFP Enhanced green fluorescent protein
EM Electron microscopy
ER Endoplasmic reticulum
IFN-α Interferon-α
IP Immunoprecipitation
IRES Internal ribosome entry site
HAV hepatitis A virus
HBV hepatitis B virus
HCV hepatitis C virus
HCVcc Hepatitis C virus grown in cell culture
HCVpp HCV pseudoparticles
IRES Internal ribosome entry site
MAF Membrane associated foci
MW Membranous web
NANBH non-A, non-B hepatitis
NMR Nuclear magnetic resonance spectroscopy
NS non-structural
NST asparagine-serine-threonine
ORF Open reading frame
OST Oligosaccharyl transferase
PDI protein disulphide isomerise
RC Replication complex
RdRp RNA-dependent RNA polymerase
rSA reverse signal anchor
SS Signal sequence
SA Signal anchor
ST Stop transfer
SRP Signal recognition particle
TM Transmembrane
UTR Untranslated region
2. AIMS OF THE PROJECT:

The overall aim of our studies was to characterize the HCV protein NS4B. Almost nothing was known about this protein as we started our investigations, least of all its function. We hoped that the gathering of knowledge about this protein would help us acquire a better understanding of it. The specific aims that emerged during our studies were as follows:

Investigate the cellular location of NS4B in detail.

Determine a model of the two dimensional topology of NS4B in the ER membrane

Investigate what parts of the NS4B that are important for its role in viral replication

Investigate possible connections between the membrane altering capacity of NS4B, its dual topology, and its ability to oligomerize.
3. INTRODUCTION

3.1 Hepatitis C Virus

In the mid 1970s serological tests indicated that the world’s supply of blood was contaminated with an unknown infectious agent that caused chronic post-transfusion hepatitis. At that time tests had recently been developed that could detect hepatitis A and B viruses (HAV and HBV). This had made it possible to identify and differentiate between patients with hepatitis caused by these two viral agents. However, there were still a large number of patients negative for both HAV and HBV that developed severe inflammatory liver disease. The term “non-A, non-B hepatitis” (NANBH) was coined and it was found to be a transfusion associated hepatitis (48).

In the early 1980s chloroform sensitivity tests and filtration studies revealed the NANBH agent to be a small enveloped virus, and in 1989 Choo and colleagues reported the cloning of the NANBH agent (28, 61). It turned out to be a positive stranded RNA virus and was given the name hepatitis C virus (HCV).

Of course the cloning of the HCV was a big step in the understanding of this viral agent. The research of HCV has however been hampered for a long time by the lack of suitable animal models and cell culture systems. Chimpanzee is still the only animal that can be naturally infected with HCV and this causes both ethical and economical concerns. No reliable or reproducible cell culture system has been available until very recently. Therefore, the non clinical research on HCV during the first 16 years since its discovery has relied in large parts on alternative methods; comparisons with other viruses, recombinant expression of proteins, viral like particles etc. But despite of this, a large pool of data on HCV is available today. It will be very interesting to see in the near future how these data stands in the light of culture grown virus.

3.1.1 Prevalence and treatment of HCV infection

HCV is the major cause of chronic liver disease and viral hepatitis all around the world and it has been compared to a “viral time bomb”. According to WHO, 180 million people (approximately 3% of the world’s population) are infected with HCV. Patients are usually asymptomatic during acute infection, but 70% of those infected will develop chronic hepatitis with the risk of developing liver cirrhosis and liver cancer. In fact, HCV is responsible for 50-70% of all cases of liver cancer and two thirds of all liver transplants in the developed world. Intriguingly, HCV is one of the rare RNA viruses known so far that manages to establish a chronic infection in the host. It infects mainly the liver, but there is some evidence also for the existence of extra-hepatic reservoirs.(183)

It has been estimated that three to four million persons are newly infected with HCV each year. Since the virus is mainly transmitted by blood, the introduction of anti-HCV screening tests for blood donors has significantly reduced the number of newly infected persons. Currently, intravenous drug users are the main risk group in the developed world, and there are also infrequent sexual, vertical and nosocomial transmissions. However, in many less developed countries the screening of blood donors is not a rule, and HCV is still a major cause of transfusion-associated hepatitis.

Unfortunately, there is currently no therapy against HCV that is sufficiently efficient to allow complete eradication of the virus. Even before the infectious agent HCV was identified,
Interferon-α (IFN-α) was shown to improve the health of patients with non-A, non-B hepatitis (69). This treatment only had limited success but IFN is still part of the therapy given to patients infected with HCV. The treatment has been improved however step by step over the years. The current therapy of choice is a combination of pegylated interferon and the broad-spectrum antiviral agent ribavirin. This combination gives sustained response rates of about 50 % (with variations for different genotypes)(111). Besides not being efficient enough, this therapy is often associated with severe side effects. Furthermore, because of the adverse effects of the treatment, some patients can not be given treatment at all. The response of the patients to the therapy varies greatly depending on both viral and host characteristics; especially the viral genotype, with genotype 1b being the most difficult to eliminate. A better understanding of the viral life cycle is therefore needed to achieve a more-effective and better-tolerated therapy against HCV.

The need for a more efficient and safe therapy against HCV infection has of course initiated major research efforts. These have aimed at both inhibiting specific steps in the viral life cycle and to find ways to modulate the host’s immune response. The main viral targets have been the NS3-4A protease as well as the NS5B polymerase, and some of the newly discovered drugs are presently in different stages of clinical trials. There is however one big issue that any successful drug against HCV has to overcome; the emergence of drug-resistant viruses. HCV easily mutates and has a fast turnover rate – a combination that may quickly result in escape mutants. A way to overcome this is to treat patients with combinations of several antiviral agents, each with a different target. The ultimate goal however is to be able to leave the IFN-based treatments in favour for something else with less severe side effects and possibly shorter treatment periods. (See review in reference (32).)

There are also attempts to make efficient vaccines against HCV – both prophylactic and therapeutic. The lack of an in vitro culture system has made it difficult to develop inactivated or live attenuated vaccines and also to correctly assess the response raised by putative vaccines. Furthermore, the high genetic diversity of the HCV genome has posed yet another problem – the difficulty of making good, cross reactive reagents. Most vaccine candidates are based on HCV envelope proteins, but there are also some using non structural genes.

3.1.2 The HCV genome organisation

HCV has been classified in the genus Hepacivirus in the Flaviviridae family. Other genera of this family are Pestiviruses (e.g. Bovine viral diarrhea virus (BVDV)) and Flaviviruses (e.g. Yellow fever virus, TBE, Dengue virus). Several features of the viruses in these genera are very similar; i.e. their genome organisation, presumed replication strategy and virion morphology. But their biological properties such as tropism and virulence can differ considerably.

The HCV particle is relatively small, only approximately 50 nm. It is enveloped and contains a positive strand RNA genome of 9.6 kb. The genome codes for a single open reading frame (ORF) of approximately 3010 amino acids (see Fig 1). The open reading frame is flanked by the 5’ and the 3’ UTRs (untranslated regions). Both are highly structured RNA elements that are able to bind different proteins important for replication. The 5’UTR and the fist part of the ORF contain the internal ribosome entry site (IRES) that direct translation in a cap-independent manner. The 3’UTR consists of three different parts; a variable part of ~40 nucleotides, a poly(U)/(UC) stretch of different lengths, and a highly conserved 98-base sequence designated the x-tail and representing the 3’terminus of the genome.(110.)
As for most RNA viruses, the mutation rate of HCV is high which result in a heterogeneous virus population. Furthermore, it also has a high replication rate ($\sim 10^{12}$ virions per day in humans (133)) which quickly gives rise to numerous viral quasispecies in each infected individual. The high variation of the virus is used to produce escape mutants that can sneak past the immune response of the host. HCV has been divided into six different genotypes (1 to 6) and numerous subtypes (a, b, c etc) that all show distinct differences both in geographical distribution, disease progression and also response to therapy. If comparing the complete genome of HCV from different genotypes, they may differ with as much as 30-35% on a nucleotide level (subtypes vary 20-25%). Different parts of the viral genome have been shown to be more variable than other; E2 and NS5A both have regions of very high variability whereas the 5’UTR, core and NS3 are more conserved (164, 165).

3.1.3 The viral proteins

Even though HCV was discovered seventeen years ago, much is still unknown about its basic characteristics. Several of the proteins have still not been ascribed a function; receptor requirements and exact ways of replication and assembly are still unclear. Most of the knowledge about the HCV proteins has been acquired by comparisons with related viruses, recombinant expression of proteins and expression of virus like particles.

The genome of HCV codes for a polyprotein of about 3000 amino acids, which is proteolytically processed by cellular and viral proteases. This occur both co- and posttranslationally at the ER membrane (Fig 1, 2):

The structural proteins Core, E1 and E2 are located in the N-terminal part of the polyprotein (see Fig 1). Core forms the capsid around the viral RNA, and E1 and E2 are the highly glycosylated envelope proteins. E1 and E2 form hetero-complexes by non-covalent interactions and E2 is needed for the correct folding of E1 (33, 118, 152). The small protein p7 is located between the structural and the non-structural (NS) proteins in the polyprotein, but to which region it belongs to is still a disputed question. By forming hexamers in the membrane, it forms ion channels that can be blocked by several different substances (60, 140). The exact function of this channel is still unknown, but the corresponding protein of the related BVDV seem to be important for the viral assembly (63).
The NS proteins are all assumed to play different roles in the replication process. The NS2 protein however is dispensable for replication in a subgenomic replicon system (see section 3.1.4), and its primary function in the viral life cycle is still unclear (Lohmann-99, Blight-2000). Currently, the only known activity of it is that it works as an autoprotease in concert with NS3. Together they digest the junction between the individual proteins and thereby releasing NS2 from the polyprotein (57, 67). Still, NS2 has been shown to interact with all other NS proteins (36), which surely indicates a function close to the replication complex even if not apparent in the subgenomic replicon system. A recent article mapped the optimal crossover site for the making of chimeric infectious virus in NS2. They showed that the first transmembrane domain of NS2 needed to be of the same isolate as the structural proteins in order to have an efficient virus assembly. Similarly, it was advantageous for viral production if the second transmembrane domain and the regions downstream of it were of the same isolate as the other non structural proteins (143). These results indicate that NS2 may be an important link between viral assembly and the replication complex.

The NS3 protein displays a serine-type protease activity in its N-terminal domain and an NTPase and helicase activity in the C-terminal part (11, 51, 56, 67, 83, 169). The serine protease activity has been very well characterised and is one of the main targets in the development of antiviral drugs. To reach optimal enzymatic activity NS3 needs to bind to its cofactor; NS4A. These two proteins form a heterodimer where the N-terminal residues of NS3 bind tightly to the centre of NS4A (10, 47, 187). Binding of NS3 to NS4A also targets NS3 to the intracellular membranes where it can interact with the other proteins in the replication machinery all that are membrane associated (see Fig. 2). Being close to the membrane is probably not important for the protease activity as such; also small peptides spanning the NS3-binding region of NS4A is sufficient for robust NS3 activity (10, 47, 100). The NS3/NS4A complex is responsible for processing the junctions between the rest of the NS proteins, releasing NS3, NS4A, NS4B, NS5A and NS5B as separate proteins from the polypeptide (see Fig. 1) (56) (11, 39, 172). The helicase activity of NS3 is located in the C-terminal part of the protein. The role of the helicase is not clear, but it probably unwinds structured RNA to make it available for translation or replication. This unwinding requires hydrolysis of nucleoside triphophates (NTPs) and all helicases found so far possess inherent NTPase as does the HCV NS3 (reviewed in (76)).
The NS4B protein is essential for HCV replication, but its exact mode of action is currently unknown (6). The only function that has been established so far is the ability of NS4B to rearrange the intracellular membranes into membranous vesicles or invaginations (42, 55, 59, 107). These may serve as scaffolds for the HCV replication complex or a way to protect the viral RNA and proteins from the host cell defence. Since NS4B is of special interest in this thesis, it will be discussed in more detail in a separate section below (see 3.1.6).

NS5A has unknown functions but is indispensable for replication (6). Furthermore, single amino acid changes in NS5A can highly improve the replication efficiency of a subgenomic replicon (18, 104). Intriguingly, this protein still allows insertions of the full EGFP protein at two different sites (125). The structure of one part of this protein (domain I) was recently solved, and it turns out to form homodimers and to have a potential ability to bind RNA (171). NS5A is a phosphoprotein with two different phosphorylation states, p56 and p58 indicating the corresponding molecular weights. The functions of them are not entirely clear, but several lines of evidence indicate that the phosphorylation status of NS5A may play a regulatory role in replication. First, treatment of cells containing subgenomic replicon with inhibitors of the cellular kinase that is responsible for the hyperphosphorylation of NS5A (giving the p58 product) leads to an increase of HCV RNA replication (132). Second, mutations in NS5A (and also in NS4B) that reduced the hyperphosphorylation of 5A lead to a more efficient viral RNA replication (7, 46). It was also shown that there was an inverse correlation between NS5A phosphorylation and the ability of NS5A to bind a cellular protein called hVAP-A. The binding of NS5A to hVAP-A has previously been shown to be important for an efficient replication (46, 52, 174).

NS5B is the RNA dependent RNA polymerase (RdRps) and located at C-terminal end of the polyprotein (16). Similarly to all RdRps it can be subdivided into palm, fingers and thumb domains, but unlike other RdRps, it fully “closes” the active site by a tight contact between the thumb and the fingers (2, 21, 96). It also contains four different highly conserved amino acid motifs (conserved among RdRps), one of which is the well known GDD motif (29, 146). NS5B belongs to a class of integral membrane proteins termed tail-anchored that use a hydrophobic C-terminal sequence to anchor the protein to the ER membrane via a posttranslational mechanism (159). This places the large ectodomain of NS5B on the cytosolic side, but still attached to the ER membrane. The attachment to the ER membrane is not important for the activity of the RdRp in vitro, but it seems to be needed in the context of the subgenomic replicon (124, 186). The membrane anchoring of NS5B probably enhances the possibility for it to interact and form functional and efficient replication complexes in the complete HCV context.

### 3.1.4 Model systems

As previously alluded to, HCV research has long been limited by the lack of efficient culture systems and small animal models. Below follows a brief description of some of the systems that have been used to acquire our current knowledge about HCV.

#### Animal models

Valuable data has been collected by infecting chimpanzees, but the use of these animals is not optimal for ethical reasons. A small animal model such as mouse would be desirable. There are in fact at least two different mouse models available, but both require quite extensive and difficult procedures on the mice in order to make them susceptible for infection. (73, 117). Both models also require transplantation of human liver tissue, making access of such a limitation for these
techniques. Furthermore, infection of these mice does not mimic a human infection as well as infection in chimpanzee does, making the use of them further limited.

Cell culture systems including replicons.

The efforts to infect already established cell lines or primary hepatocytes with serum-derived HCV has yielded only low-level replication - levels that were not sufficient to allow studies of the many processes of the HCV life cycle. In the absence of an efficient cell culture system that would cover the entire life cycle of HCV, other models were developed that made it possible to study parts of it. Data concerning particle assembly, host cell binding and infection has been collected by production of virus like particles in insect and mammalian cells, pseudotyped vesicular stomatitis virus or putative virions (14, 17, 65, 92). But it was not until the development of an infectious retroviral pseudotype virus system that a really reliable surrogate model for attachment and possibly entry was established (12, 71).

When it comes to studying the replication of HCV a big roadblock was removed with the development of subgenomic replicon system; a self-replicating RNA in the human hepatoma cell line Huh-7 (106). This system allowed researchers for the first time to study efficient HCV RNA replication in cell culture and also to do reverse genetics and study the effects on replication. The prototype replicon was a bicistronic RNA encoding a neomycin resistance gene under the control of the HCV IRES, followed by a second IRES from Encephalomyocarditis virus that controlled the genes for NS3-NS5B. See Fig 3. Upon transfection of the replicon RNA into Huh-7 cells cell clones expressing the HCV replication complex (NS3-NS5B) were able to efficiently replicate their RNA in high amounts and thereby conferring resistance against the G418 selection agent. Replication efficiency could therefore be scored by cell survival and quantification of new viral RNA. The efficiency of replicons was improved by the discovery of certain amino acid substitutions, the so called cell culture-adaptive mutations that increased the replication by several orders of magnitude (18, 104, 105). This advance made it possible to develop transient replication assays that made use of more direct and faster readouts than cell survival. Besides the many answers about HCV replication that the replicon system has provided, it has also become an important tool for drug discovery and studies of antiviral resistance. Attempts were made to add the structural genes to the subgenomic replicons, but no clear evidence for virus production could be found despite efficient replication of the full length RNA (144).

![Fig 3. The subgenomic replicon](image)

Production of viral particles in cell culture

A major step for HCV research was taken during 2005 when three groups simultaneously reported that they had been able to grow infectious HCV in cell culture (101, 177, 191). All three groups had used a highly efficient HCV clone of genotype 2a, that had been cloned from a patient with fulminating hepatitis C. This magic clone was called JFH-1 and had previously been shown to efficiently replicate as a replicon in a panel of different cell lines; both in hepatic and nonhepatic cell lines. Furthermore, it did so without the requirement for adaptive mutations (80). Why the
JFH-1 managed to produce infectious virus in cell culture but no other of all clones tested is an important but still unresolved question. But whatever the reason for it, it surely represents an important milestone in HCV research, since it will finally be possible to apply classical virological techniques to this elusive virus.

Recent reports have also described further advancements in the use of the infectious HCV grown in cell culture. The addition of a luciferase reporter gene to the virus contributed with a fast and convenient readout for infection (87). Furthermore, by constructing chimeric virus with the NS region from the JFH-1 clone and the structural region from other genotypes (1a, 1b, 2a and 3a) the use of the infectious virus system was has been further expanded (143).

3.1.5 The HCV life cycle

The life cycle of viruses can briefly be divided into three different steps; 1) binding to host cells and cell entry, 2) translation and replication, 3) virus assembly and release. Since the HCV field has long been hampered by the lack of efficient culture systems, the current understanding of HCV life cycle is largely based on studies of different surrogate systems as well as comparisons with other viruses. A schematic illustration of it can be seen in Fig 4.

Binding of host cells and cell entry.

Generally, virus entry into host cells is a complex process that may involve many different interactions; docking to the target cell, fusion of membranes and subsequent internalisation of the virus. In the case of HCV, the search for candidate receptors has been made with HCV pseudo virus, virus-like particles, or recombinant expression of E1 and E2. The main receptor candidates today are CD81 and the human scavenger receptor class B type I (SR-BI), but binding has also been reported to two lectins called L-SIGN and DC-SIGN as well as to the LDL receptor (1, 53, 145, 147, 158). The expression of CD81 and SR-BI in non-liver cell lines does however not allow entry of HCV pseudo virus into cells (13, 71). Furthermore, these receptor candidates are both expressed on many different cell types, in contrast to the rather restricted tropism of the HCV. Clearly, additional cofactors need to be identified to explain HCV’s specific interaction with liver cells. The internalisation of HCV into target cells has been suggested to occur by receptor-mediated endocytosis which includes a low pH compartment that triggers membrane fusion (13, 71). The use of HCVcc has confirmed the requirement of CD81 as a main receptor, and also the pH-dependent virus entry (87, 101, 177, 191).

Translation and processing

Once the genomic RNA has been released into the cytoplasm it is ready to take the role of mRNA, and initiate translation of viral proteins. The IRES binds directly the 40S ribosomal subunit without the need for pre-initiation factors that are required for cap-dependent translation. Subsequent binding of eukaryotic initiation factor (eIF) 3, the complex Met-tRNA-eIF2-GTP and the 60S ribosomal subunit slowly makes up the translationally active ribosome (138). Translation of the HCV genome results in a large polyprotein that is integrated in the ER membrane. By co- and posttranslational processing it is cleaved into at least ten different proteins (see Fig. 1). The N-terminal part of the polyprotein is digested by host cell signal peptidase located in the ER lumen (C/E1, E1/E2, E2/p7, p7/NS2). NS2-3 autoprotease cuts the NS2/NS3 junction and the NS3/4A protease releases the rest of the NS proteins (NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B).
Replication

Little is known of the details of HCV replication, but it is generally believed that most NS proteins are involved in the process. This assumption is based on comparisons with other members in the Flaviviridae family, and the fact that all NS proteins are localized together with newly synthesised viral RNA on membranes derived from the ER (44, 55, 162). Most NS proteins also interact with several other NS proteins, and some of them interact with all (36). Furthermore, adaptive mutations in NS3, NS4A, NS4B, NS5A and NS5B have been found in subgenomic replicons and this strongly indicates that all these proteins have important roles in the replication process. The main players in the replication complex (RC) however are most likely the NS3 helicase and the NS5B RdRp while the roles of the other NS proteins in the replication process remains to be elucidated. It is interesting to note that adaptive mutations in NS4B, NS5A or NS5B are able to strongly enhance replication on their own, but can not to be combined with each other (104). In contrast, the adaptive changes in NS3 were comparatively weak, but when combined with one of those in NS4B, NS5A or NS5B they gave a cooperative effect (88, 104).

The making of new HCV RNA genomes goes via a negative stranded RNA intermediate that is produced by the replication complex. By studying replicon containing cells, it was estimated that each replication complex contains one negative strand RNA and two to ten RNAs of positive polarity (151). It also contains several hundred NS protein copies where the majority may be needed as structural components. Furthermore, only a small fraction of all NS proteins seem to involved in the actual replication process, since there seem to be an excess of NS proteins also outside the RC (121, 151). If the vast amount of NS proteins outside the RC are simply by-products or if they have some kind of function is currently unknown.

All positive-strand RNA viruses studied so far form membrane associated replication complexes and HCV is no exception (reviewed in (3, 109)). By using host cell membranes as platforms for the replication, the replication factors are concentrated for efficient assembly into replication complexes. Furthermore, all positive-strand RNA viruses also rearrange intracellular
membranes into vesicles or invaginations. This phenomenon will be further discussed in a separate section below (see 3.2.2).

There seem to be a close functional association between the HCV replication complex and host cell membranes. Several reports have indicated that the RC of HCV is located on membranes that have similar properties to lipid rafts and also that the actual replication may need a specific lipid environment (4, 77, 162). The viral replication is stimulated by the presence of saturated or monounsaturated fatty acids, but inhibited by polyunsaturated fatty acids or inhibitors of fatty acid synthesis (77). The rationale behind this is not clear, but maybe the right composition of lipids helps maintaining a membrane structure that is optimal for viral replication. In addition lipid modification of proteins has also shown to be important for HCV replication. Inhibition of geranylgeranylation in cells leads to disassembly of the replicase complex and abolishment of HCV replication (77, 188). One explanation behind this may be the recently reported interaction between the geranylgeranylated protein FBL-2 and NS5A (181). Palmitoylation of NS4B has also been reported (189).

Virus assembly and release

It is still unclear how the different parts of the HCV particle are assembled. But there must be a precise process that manages to pack the right RNA together with the core proteins and then envelope the nucleocapsid with a lipid membrane having E1 and E2 correctly assembled into heterodimers. The core proteins has been shown to be able to self-assemble into nucleocapsid like particles in the presence of structured RNA (90). How the virus achieves packaging specificity of the right RNA has not been elucidated yet. Neither is it known how the newly synthesised genomic RNA finds the core particle in the cells or how the envelopment of the nucleocapsid is achieved. In analogy to other members in the Flaviviridae family, virus formation is however thought to occur by budding of E1- and E2- containing membranes into the lumen of the ER. This kind of budding has also been seen with virus like particles that has been generated by heterologous expression of Core, E1 and E2 in cells (17). Both E1 and E2 contain ER retention signals and have been shown to be localized to the ER (37, 38, 156). Core on the other hand seems to localize to ER membranes, but also to lipid droplets (8, 126) (70, 156). A recent report by Rouillé et al investigated the localization of the structural proteins and NS3 in cells infected by the HCVcc. Interestingly, there were less colocalization of the different HCV proteins than previously thought: Core and E2 did not colocalize anywhere in the cell, and both core and E2 partly colocalized with NS3 but not in the membranous web. This is in contrast to earlier reports that used heterologous expression of HCV proteins and found that both core and envelope proteins colocalized at certain sites in the cells and structural and NS protein colocalized in the membranous web (42, 126). In the study by Rouillé et al however, no structural proteins were found in the MW. Overall the results in the HCVcc system suggests that generation of HCV particles is a multi-step process that take place at different sites in the cell. How the release of HCV progeny from cells is achieved is still an open question. It has long been assumed however that HCV exits through the secretory pathway, but this has not been shown yet due to the lack of suitable model systems. Hopefully the HCVcc system will now show us whether this assumption is true within the near future.

3.1.6 NS4B’s potential role in the viral life cycle and replication

Surprisingly little was known about NS4B during the first decade after the discovery of HCV; basically only its size (27kDa) and sequence. The size and hydrophobicity profile was also very
Marika Lundin

similar to the NS4Bs in other members of the Flaviviridae family, even though there is no amino acid sequence conservation between the different genera (27, 29) The functions of it were – and to large parts still are – unknown. NS4B has long been assumed to be a part of the replication complex, even though proof of it was long in coming. And even today indications to support this notion are only indirect proofs. For example, a single amino acid mutation in the HCV NS4B strongly enhances the replication in a subgenomic replicon system, but the same mutation is incompatible with a substitution in NS5B that normally results in a more efficient replication (105). This clearly suggests that NS4B at some point in the replication cycle has a function that is tightly coupled to that of the viral RNA polymerase. Furthermore, several other single amino acid mutations in NS4B have been shown both to be able to inhibit as well as to enhance replication (61, 102, 104). Studies of the Kunijin virus and BVDV (both belonging to Flaviviridae family) also support the notion that NS4B may indeed be an important part of the replication machinery; both deletion of NS4B as well as single amino acid mutations had negative effects on replication (58, 81, 98). Still, no exact function of NS4B in viral replication in any of the members of the Flaviviridae family has yet been found. A recent study reported however that Dengue virus NS4B has the ability to bind the NS3 helicase domain, and may play a role for the dissociation of RNA (176).

In 2002 the first report that clearly showed a function for NS4B was published. Egger and colleagues showed convincingly that NS4B had the ability to induce membranous vesicles or invaginations in the host cell, a so called membranous web where the RNA replication was shown to occur (42, 55). These structures may serve as scaffolds and protection for the HCV replication complex. Similar intracellular membrane rearrangements have been found in all positive strand RNA viruses investigated so far and will be discussed in more detail in a separate section below (see 3.2.2.). Egger et al. used electron microscopy to detect the membranous web, but several other groups have also detected intracellular structural changes by using immunofluorescence microscopy (45, 59, 84, 107) (Fig 5). These structures have been called foci or membrane associated foci (MAF). Since appearance of foci seem to correlate with the presence of a membranous web, it is probable that these two structures are related. However, NS4B most likely has other functions in the replication complex than the induction of the membranous web – as we show in paper III, the W1963S mutant still has the ability to induce membrane rearrangements, even though it is completely incapable of replication (102).

In EM: Membranous web

In fluorescence microscopy: membrane associated foci (MAF)

Egger et al. 2002 J. Virol. 76: 5974-84

Lundin et al. 2003 J. Virol. 77: 5428-38

Fig 5. Structural changes in the intracellular architecture caused by NS4B
According to the reports about HCV NS4B that has been published during the last five or six years, it seems this protein is a versatile player. Single amino acid mutations have been shown to alter tropism or interferon resistance of a subgenomic replicon (130, 192). And interestingly, NS4B in Dengue virus, Yellow Fever virus and West Nile virus have also been shown to influence interferon resistance, so this may be a common trait among the NS4Bs from different genera in the Flaviviridae family (127, 128). The HCV NS4B has also been reported to influence the hyperphosphorylation of NS5A (7, 46, 85, 131), inhibit protein expression (49, 78), affect the anterograde traffic from the ER to golgi (86), transform NIH3T3 cells (139), modulate the unfolded protein response (190), activating NF-κB associated signalling (79) and binding to CREB-RP/ATF6β which is a ER stress-associated transcription factor (173). NS4B has also been reported to have a GTP binding motif (43).

3.2 Biological membranes and membrane proteins

3.2.1 The Eucaryotic cell

The eukaryotic cell needs to host many different functions and perform many complex tasks. It has to deal with DNA replication, transcription, translation, sorting of proteins, sending and receiving different signals, defending itself against micro organisms, and much more. In order to execute all these specialized functions, the cell is divided into different subcellular compartments, each separated from each other by lipid bilayers. These provide barriers against the surrounding environment but also contain transport systems to allow selected molecules to pass from one compartment to another. Furthermore, many biological processes take place in and across the membranes, and the different membranes of the cell also communicate with each other either by direct association or through vesicular transport. The nuclear membrane is continuous with the ER which forms an intricate network of channels and subcompartments within the cell. The smooth ER is involved in synthesis and metabolism of lipids. The rough ER is covered by ribosomes that synthesize secretory and integral proteins and is the first compartment of the secretory pathway. In this pathway secretory proteins are loaded into vesicles that travel through the Golgi complex out to the plasma membrane and are exported out of the cell.

3.2.2 Viruses – membrane hijackers

As indicated earlier, viruses may take advantage of host cell membranes in several ways. At uptake via the endocytic pathway the viral envelope can fuse with the cell’s membrane and thereby release its nucleocapsid. The secretory pathway can be used for release of the viral progeny via budding, instead of cell lysis which is the other option. And all positive-strand RNA viruses, whether they infect mammalian, insect or plant cells, assemble their replication complexes on intracellular membranes of the host cell. Furthermore, they also form vesicles or invaginations of the membranes to facilitate and protect the virus replication (reviewed in (3) and (109)). This rearrangement of membranes serves several purposes. First, it provides a confined surface area for RC to be concentrated and assembled onto. Second, it creates secluded areas where the RC and the newly made viral genomes are protected against host cell defenses. Third, the viral genomes are also sequestered from competing host RNA templates. Different viruses use
different membranes as origins for these new structures; they can use the ER (141, 155), golgi (94), mitochondrial membranes (119) or the endosomal/lysosomal compartment (50).

The HCV polyprotein also causes rearrangements of the intracellular membranes. This has been seen both in liver tissue from infected chimpanzees, infected cells in vitro, cells harboring subgenomic replicons and also in cells expressing HCV proteins under a tetracycline regulated promoter (42, 55, 142, 156). Egger et al. first saw these distinct membrane alterations – tightly packed vesicles in a membranous matrix- using electron microscopy and termed it a “membranous web”. The protein responsible for this effect is the NS4B, and the new structures most likely originate from the ER. Replicating RNA as well as the other NS proteins has been found to localize to these structures, suggesting that it indeed is the site of replication (55). The mechanism behind this membrane altering effect is currently unknown, and is ill defined also for most other viruses. But considering the difficulty to transcomplement proteins in the HCV RC it seems likely that the rearrangement of membranes need translation in cis (6). The translation of HCV proteins may in turn bend the membrane and create the secluded vesicles to which no proteins expressed in trans have access.

3.2.3 The ER membrane and translocating proteins

Much of what is discussed in this thesis takes place on ER membranes or membranes originating from them. The HCV replication takes place on them, NS4B localizes to them as well as rearranges them. We have also used them as tools in the glycosylation studies. Therefore I think it is appropriate to introduce this part of the cell a bit closer and also discuss how the proteins in these membranes are integrated and how this integration can be used for topology predictions.

Biological membranes

The most important building blocks in all biological membranes are the phospholipids (phosphatidyicholine, phosphatidylserine, phosphatidylethanolamine and sphingomyelin). They consist of hydrophobic fatty acid tails in one end that are linked to hydrophilic phosphate-containing head groups in the other. Because of their amphipathic nature, they spontaneously form bilayers in aqueous solutions (see Fig 6) with the hydrophobic tails buried in the interior of the membrane and the polar head groups in contact with water. The membranes of mammalian cells consist not only of phospholipids, but also of glycolipids, cholesterol, and proteins. The proportion of each varies between the different cell membranes. The lipid bilayers behave as two-dimensional fluids (“The fluid mosaic model” (167)). This is an important property because it allows individual molecules – both lipids and proteins – to rotate and to move freely in all lateral directions. The fluidity of the membrane is critical for the lateral diffusion and it is in turn is affected by both the temperature and the lipid composition. The lateral diffusion in membranes can result in clusters of certain lipids and proteins held together in so called lipid rafts. These can form micro domains
within the lipid bilayer that may provide a certain environment for specific biological processes. (Reviewed in (30) and (166)).

The translocation machinery

All biological membranes and subcellular compartments contain proteins. Therefore there must be transport mechanisms to allow proteins to cross or integrate into the lipid bilayers. All proteins destined to the ER, golgi, plasma membrane or secretion need first to pass through the ER membrane. Protein transport across this membrane can occur both co- and posttranslationally, but the former predominates in mammalian cells.

Proteins that are bound for integration in the ER membrane or for the secretory pathway are first recognized by the signal recognition particle (SRP) as their first hydrophobic sequence emerge from the ribosome (Fig 7). SRP binds to the signal peptide or the first TM segment of the protein which causes a translational arrest. SRP subsequently binds to the SRP receptor at the ER membrane which leads to the ribosome docking to the translocation machinery; the Sec61-complex. Once targeted, the newly made polypeptide chain is transferred from SRP to the translocation channel in a GTP-dependent step which results in a continuation of translation. (Reviewed in (40, 108, 178)).

The Sec61 complex - the eukaryotic translocon – is a heterotrimeric complex consisting of Sec61α, Sec61β and Sec61γ that forms a cylindrical structure in the ER membrane (62). Other proteins associated with the translocon are the translocating associated membrane protein (TRAM; required for efficient translocation of some proteins), oligosaccharyl transferase (OST; adds N-linked carbohydrates to proteins), the signal peptidase (SPase; cleaves of signal sequences), the translocating chain associated protein (TRAP; enhances translocation of some proteins) (31, 34, 54, 116). Not all of these proteins are needed for translocation to occur; reconstitution studies of purified components have shown that the only parts needed to catalyze protein translocation across the ER membrane are the SRP receptor, the Sec61complex and TRAM (for some proteins) (54).

When the ribosome docks at the cytosolic side of the Sec61 complex, translation starts again and the nascent protein is passed through the translocon. At this moment the channel must be open.

Fig 7. Schematic representation of the translocon and some of the proteins associated with it. SRP recognizes the signal sequence of the nascent peptide and targets the complex to the ER translocon. Translocation of the protein can then continue in a cotranslational manner.
across the membrane, but still prevent the passage of small molecules from one side of the membrane to the other. Two models have been proposed to explain this: The first model proposes that the ribosome itself seals the channel at the cytosolic side of the membrane. Thereby the luminal side of the channel can open and release the newly made protein without risking leakage of ER components. Next, at detachment of the ribosome from the Sec61 complex the channel is instead closed by the chaperon BiP on the luminal side. The second model is based on more recent structural studies of the archeal translocon suggesting that the channel itself is shaped as an hourglass (179). The narrowest point of the channel is located in the middle of the membrane and lined with hydrophobic residues. These act as a gasket around the nascent polypeptide and thereby prevent membrane permeation of small molecules during translation. Furthermore, at the luminal side of the channel there a short helix extending from the α-subunit that acts as a “plug” to seal the translocon when in inactive or the TM helix has moved laterally out of the channel. (Reviewed in (153) and (5).)

Besides being able to open itself as a channel through the membrane, the translocon must also be able to allow lateral integration into the ER membrane. X-ray structure showed that the α-subunit of the translocon is formed like a “clam shell” hinged by a loop at the back of the molecule (179). The mouth of the clam shell most probably represents the lateral gate through which the hydrophobic segments of nascent proteins can leave the channel. The exact mechanism behind this is currently not clear and for proteins containing multiple TM segments; it could either occur sequentially or collectively.

**Topogenic sequences and protein insertion into the ER membrane**

Secreted and integral ER proteins all have one or more signal sequences that target the polypeptide to the ER membrane and also influence the transmembrane orientation of the protein. These sequences are classified differently depending on where they are located in the protein how they orient themselves in the membrane:

A signal sequence (SS) is always located in the absolute N-terminus of the protein. In this way the SRP recognizes it as soon as it emerges from the ribosome and targets the mRNA-ribosome-protein quickly to the membrane where the SS is inserted in the membrane with the N-terminus facing the cytosol (see Fig 8). The signal sequence consists of short stretch of hydrophobic residues (6-15 aa residues.) that are flanked by one or more positively charged residues at the N-

![Hydrophobicity](image)

**Fig 8.** Topogenic sequences. The targeting signal (white) of ER membrane proteins can be a signal sequence (SS) which will be cleaved off, signal anchor (SA), reverse signal anchor (rSA) or a C-tail. The TM segment following the targeting signal in polytopic membrane proteins are either SA or stop-transfer (dotted line) depending on their orientation. Adapted from Goder and Spiess, FEBS letters 2001, 504:87-93.
terminus and a recognition site for the signal peptidase in the C-terminus (112). In most cases the SS is rapidly cleaved off during translocation and subsequently degraded.

Also internal TM segments in a membrane protein can act as targeting signals. If it is inserted in an N_cyt-C_lum orientation it is called “signal anchor” sequence (see Fig 8), and if in the opposite direction (N_lum-C_cyt) “reverse signal anchor”. After initiation of translocation of a polytopic membrane protein, the translocation across the ER membrane continues until the translocon comes into contact with the next hydrophobic sequence, the “stop-transfer” sequence, which halts the translocation and integrates the stop-transfer sequence into the membrane. Another rare, but still important topogenic signal is found in the so-called tail-anchored proteins (to which the HCV protein NS5B belongs). They have a single TM helix in the C-terminus and therefore require posttranslational insertion into the ER membrane (91).

What determines the orientation of the TM segments? There seem to be several different factors that influence the orientation, but three seem to be especially important: First, “the positive inside rule” which compares the flanking regions of a TM segment. The one that contains more positive charges tend to have an orientation towards the cytosol (184, 185). Second, the folding of the domain preceding the TM helix: only regions that are not stably folded into a tertiary structure can be translocated (35). Third, the length and hydrophobicity of the TM helix is also important: longer sequences favour a N_lum-C_cyt orientation (161). These three factors together influence the orientation of TM segments. For polytopic proteins, they seem to be particularly important for the first hydrophobic segment. The orientations of the subsequent TM helix often alternate correspondingly. There are however many examples that suggesting that the process of integration in the membrane is even more complex; Internal TM segments from polytopic proteins have been shown to have preferred orientations when expressed on their own (103). There are also examples that show that both upstream and downstream TM segments affect the orientation of a specific TM helix within a multispanning protein (26, 103, 122, 135). It all suggests a rather complex process of topology determination in the ER translocon.

Posttranslational modifications of proteins

Secretory and integral proteins translocated into the ER lumen may achieve one or several different modifications. Examples of such are signal sequence cleavage (briefly discussed above), N-linked glycosylation and disulphide bond formation.

At N-linked glycosylation the enzyme oligosaccharyl transferase (OST) adds an oligosaccharide substrate to a consensus site in the nascent polypeptide chain. The OST is associated with the translocase complex on the luminal side of the ER membrane and adds a glycan antenna of 14 sugar residues to the protein as it emerges through the translocon channel. The recognition site on the protein for N-linked glycosylation is the sequon asn-X-ser/thr, and the oligosaccharide is attached to the asparagine in this sequence. X can be any amino acid except proline, and proline residues in the flanking region of the recognition site also impair glycosylation (115). Four of the fourteen sugar residues are removed from the oligosaccharide unit while still in the ER, and if transported further into the Golgi apparatus it will be modified even more. (30).

The formation of disulphide bonds between cystein residues is performed by protein disulphide isomerase (PDI). This process is important for stabilizing folded structures of many proteins both within the same peptide and between peptides to create quaternary structures. The cystein residues that are to be linked by disulphide bridges may be far apart in the primary structure, so disulphide formation between them is intimately linked with the folding of the polypeptide chain. (30).
3.2.4 Topology of membrane proteins

Membrane proteins are estimated to make up approximately one third of the proteome, and they are often responsible for critical cellular mechanisms. But despite of this they are generally ill characterized. Only a handful of membrane proteins have had their structures solved by high-resolution techniques, as compared to several thousands of globular proteins. The reason for this is simple. Membrane proteins are due to their hydrophobic nature hard to analyze and characterize biochemically due to their hydrophobic nature and limited solubility; they are difficult to over-express, to purify and to handle. Methods available for structure determination of proteins are X-ray crystallography, NMR, electron microscopy, biophysical methods and molecular biology techniques. For almost all of them, they are more difficult to apply to membrane proteins than to soluble ones. For instance, membrane proteins seldom yield good crystals (needed for both X-ray crystallography and electron microscopy) and NMR is only applicable on small membrane proteins at high concentration.

Predicting the topology of membrane proteins

Membrane proteins are hard to work with practically, but by using computer modeling it is possible to get a theoretical model of the protein of interest. The advantage with membrane proteins is that even though they come in a variety of sizes and shapes they still have to follow some basic structural principles due to the lipid environment into which they are embedded. These structural principles are used in the prediction programs and will be briefly described below.

The membrane spanning portions of membrane proteins can be of helix bundle or beta-barrel type. Beta-barrel membrane proteins are however uncommon and will not be dealt with here. Helix bundle proteins contain one or more alpha helices, each of which is a stretch of largely hydrophobic side chains. The overall length of the TM helix can vary between 15 and 40 amino acid residues with an average of 26 (20). The different lengths of the helices can be explained by different tilt angles of the helices in the membrane.

The TM alpha helices are surrounded by the lipid bilayer and this is of course reflected in their amino acid residue composition. The residues exposed to the lipid hydrophobic core of the membrane are generally hydrophobic (e.g. Ile, Leu, Val, Ala, Phe, Gly). And since the TM helices of a polytopic protein can protect each other from the hydrophobic environment, they are usually less hydrophobic than TM helices from single spanning membrane proteins. To appreciate what amino acid residues that is more likely to be used in a TM helix, different hydrophobicity scales have been developed (see Fig 9). The order of the amino acids residues in these rankings differ somewhat however, depending on how the hydrophobicity was estimated (different biophysical or statistical methods have been used). A "biological" hydrophobicity scale was recently developed by challenging the Sec61 complex with a set of designed TM helices. It corresponded quite well with both biophysical and statistical hydrophobicity scales, but also emphasized the positional preferences for some residues in the TM helices (66). Aromatic residues (Tyr and Trp) are often positioned within the lipid head group region (82).

| I | V | L | F | C | M | A | G | T | W | S | Y | P | H | E | Q | D | N | K | R |

Fig 9. The Kyte-Doolittle scale is one example of the hydrophobicity scales available (J. Mol. Biol. 157:105-32. The amino acids are given with their one-letter abbreviations.)
In contrast to the TM helices described above, the connecting loops between them as well as the N- and C-termini have to be rather hydrophilic and can be of varying lengths and shapes. They are however also important for the decisions taken by the translocon in the process of integrating the protein in the membrane. As was described above, both folding and charge distributions in these parts are of importance for the final topology (see Fig 8).

When using the word “topology” of membrane proteins, it is the two dimensional structure that is referred to. In other words; what parts of the protein is integrated in the lipid bilayer and in what orientation? By using the basic architectural principles of membrane proteins described above prediction programs such as TMHMM, HMMTOP, PHD, TopPred or MEMSAT can offer initial theoretical suggestions of the 2D topology of the protein of interest. Preferably, the several different programs should be used for the analysis.

The predicted topology also needs to be experimentally verified. There are several methods that may be considered, such as chemical modification of cysteins, N-linked glycosylation, protease treatment, addition of reporter molecules or epitopes.
4. COMMENTS ON MATERIAL AND METHODS

This section is a general commentary on some of the different methods we have used in our studies. More detailed descriptions of the separate methods can be found in the material and method sections of papers I-IV.

4.1. Studies of NS4B topology (paper I, II, IV)

4.1.1 Topology prediction

A good topology model of a membrane protein is a necessary starting point for most experimental structure-function studies. The computer programs applied for this usually tries to discriminate between membrane proteins and non-membrane proteins, to predict locations of the TM helices and to determine the orientation of them in the membrane. The predictions are usually based on the hydrophobicity of the TM spanning segments and the “positive inside rule” described in section 3.2.3 (“topogenic sequences…”). It is important to remember that the predictions are only a first approximation and are often imprecise and may also be incorrect. Predictions are complicated by the facts that some proteins have dual topologies (64, 97), that it may be hard to discriminate between one long TM helix and a helical hairpin, difficulties in recognizing half-TM helices and that the “rules” followed by the prediction programs are far from universal.

Some examples of the prediction programs used in our studies are MEMSAT (75), TMHMM (89), and HMMTOP (175), all three used for the initial prediction of NS4B (paper I). The MEMSAT program differentiates between five structural states (inside loop, inside helix end, helix middle, outside helix end and outside loop). Each state is then compared with a statistical table of the 20 different amino acids and the frequencies with which they occupy the different states. A dynamic programming algorithm is used for making the subsequent prediction. Similarly, HMMTOP and TMHMM assess five or seven defined regions of an integral membrane protein. Instead of dynamic programming, they use algorithms called hidden Markov models (HMM) to make the predicted model. At the bottom of all three programs are tables and statistics coming from experimentally determined topologies of membrane proteins. There may however be a certain bias in the test sets due to the facts that relatively few eukaryotic membrane proteins have known structures, and also that certain types of membrane proteins are easier to determine the structure for than others.

In our first analysis of NS4B using the above mentioned modelling programs, the N-terminal half of the protein was more difficult to predict than the C-terminal half. This was reflected both by the fact that the MEMSAT program disagreed with the other two programs about the placement of the first TM segment and that the TMHMM program had lower probability scores in that region of the protein (see Fig 10 and 11). However, the TMHMM program has been further improved by allowing...
the input of experimental data (114). By using this improved predictor together with our results from the glycosylation mapping, the new prediction got high probability scores in all regions of the protein (paper I, Fig 8).

### 4.1.2 Experimentally verifying the predicted topology

Even the best predicted topology model needs to be experimentally verified. There are many different methods that may be used, and all are based on the fact that the membrane separates two different compartments that have different properties and different accessibility to various agents. Briefly, available methods can be divided into two groups; the first is based on gene-fusion, where the protein of interest is attached to different reporter molecules (e.g. alkaline phosphatase, lacZ, β-lactamase) that will react differently depending on which side of the membrane it has been located to. The second group relies on smaller insertions/mutations within the protein of interest; N-linked glycosylation (further discussed below), cystein modification, protease treatment, or epitope insertion.

#### N-linked glycosylation

The glycosylation mapping technique makes use of the natural system of N-linked glycosylation present in all eukaryotes. The enzyme OST is located in the ER membrane close to the translocon where it adds a glycan antenna to translocating proteins as it emerges from the translocon (see Fig 12). The OST only acts on the luminal side of the ER membrane.

Therefore, adding the consensus motif asn-Xaa-ser/thr at various positions in a protein and subsequently assess if they are glycosylated or not, will indicate what portion are located to the lumen (glycosylated) or those that are not (not glycosylated). Glycosylation is indicated as a 2KDa shift in SDS-PAGE. However, the OST has a sharp threshold for glycosylation and need a distance from the membrane to the acceptor asparagine of 12-14 amino acid residues (136). This feature can be used to map borders of TM helices (paper IV), but it can also cause problems when investigating membrane proteins with very short loop regions (paper I). An advantage with glycosylation mapping is that it can easily be used in *in vitro* systems. A disadvantage is that endogenous glycosylation sites of the proteins should be removed to facilitate analysis. Furthermore, the insertion of the new glycosylation sites may destroy the activity of the protein - not necessarily due to the amino acid changes, but also by the addition of the glycan antenna that may disturb the folding of the protein and/or hinder the accessibility of functional residues. It is

<table>
<thead>
<tr>
<th>NS4B sequence</th>
<th>TMBMM</th>
<th>HAMTOM</th>
<th>MEMSAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQHLPYIQGQ MMILAQPKDK ALGILGTAIR KAEVTLPWQ TNIQQLPVWPAKHMWNPISG QCYLESLSL RTGQPAASLM AIFTAAATSPL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E92</td>
<td>TTGQLTLPL LNQAHAAQLAQ PQAATAFPAG ALHAFAGAAGS VGOLGTVLDI LAGYGAGVAGAALPAKMSG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E135</td>
<td>LAYGAGVAGAALPAKMSG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E161</td>
<td>ENPSTEDLN LLPALSPGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E216</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig 11. Overview of the results from different programs used to predict the topology of NS4B. The amino acid sequence of NS4B (genotype 1a) is listed at the top, and the sites of EC4 insertion in the different constructs are indicated above it. The gray bars show the location of the predicted TM segments.
also important to remember that negative glycosylations may have several different explanations. Except the obvious one – being on the cytosolic side of the ER membrane – the lack of glycosylation may also be due to a location in the membrane or steric hindrance on the lumenal side of the membrane. (Reviewed in (180).)

We used N-linked glycosylation in paper I, II and IV. The reasons for choosing this method were several. First, NS4B does not contain any endogenous glycosylation sites that needed to be removed. Second, it is a convenient method as it can be used in in vitro systems, and modifications of the gene can easily be made by a single PCR step instead of cloning. Third, as the OST has a sharp threshold for glycosylation, we could use it as a “molecular ruler” to find the borders of TMX (paper IV).

4.2 In vitro translation (paper I-IV)

The cell-free translation systems consist of extracts from rabbit reticulocytes, wheat germ or Escherichia coli containing all the macromolecular components required for translation of exogenous RNA. It is a convenient system in many ways; First, the cloning of mutant constructs is in many cases not required, since PCR products may be used as templates for the in vitro transcription. Second, the results can be achieved much faster than when expressing the genes of interest in cells, since transcription, translation, and gel run can all be done in the same day. Third, in vitro expression may be the solution if the over-expressed protein is toxic to the host cell. The system used in our studies, rabbit reticulocyte lysate with dog pancreas microsomes, is generally considered to closely mimic the situation in a living mammalian cell. Microsomes are the crude extracts of ER membranes, thereby providing the environment needed for membrane or secretory proteins. There are however some drawbacks with this system that have to be considered. For example, the microsomes used in glycosylation studies often have a slight batch variation that may cause the glycosylation efficiency to vary between experiments. The glycosylation of proteins is also less efficient than when expressed in cells.

4.2.1 Kinetics of glycosylation (paper I):

When testing the time of glycosylation experimentally, the translation of proteins is first synchronized. This is done by addition of aurintricarboxylic acid (ATA) which at low concentrations inhibits the interaction between the mRNA and the ribosome. Thereby any new initiation of translation is prevented, and it is possible to follow the synthesis and glycosylation of peptide chains that had been initiated before the addition of ATA. Detergent is then added to aliquots of the translation mixture at different time points. This disrupts the microsomal membranes and thereby prevents any further glycosylation of proteins in that aliquot. All aliquots are however incubated at 30°C for a full hour, allowing all proteins to be fully translated. When running the aliquots on SDS PAGE, the appearance of a band

Fig 12. Glycosylation of a protein in the ER lumen. The number of amino acids required to cover the distance of the ribosome, the lipid bilayer and to reach the active site in the OST is indicated.
corresponding to the glycosylated version of the protein shows at which time point the acceptor asparagine reached the oligosaccharyltransferase.

The result of the experimentally determined time of glycosylation can then be compared to the theoretically estimated time point: The protein is translated at a speed of approx 0.4 aa/s in the in vitro system. When calculating the time point for a certain site to be glycosylated, the extra distance needed for the acceptor asparagine to reach the OST also has to be considered. This distance is approximately 70 amino acid residues which add almost three extra minutes (see Fig 12).

In the case of NS4B, we tested the time of glycosylation for two different sites: the second luminal loop (E161) and the N-terminus (E33). The theoretical times of glycosylation of these sites were estimated as follows:

\[
\text{E161: } \frac{(161^a + 15^b + 70^c)}{0.4} = 615 \text{ seconds} = 10.25 \text{ minutes} \\
\text{E33: } \frac{(33^a + 15^b + 70^c)}{0.4} = 295 \text{ seconds} = 4.9 \text{ minutes}
\]

\(a\) Number of amino acid residues in NS4B preceding the insertion site of the EC4 loop
\(b\) Number of amino acid residues in the EC4 loop preceding the acceptor Asparagine
\(c\) Number of amino acid residues needed to be translated downstream of the glycosylation motif before the acceptor asparagine reaches the OST.

The experimentally defined time points for glycosylation of these sites were 9 minutes for E161 which corresponded quite well with the theoretical value of 10 minutes. It should be noted that translation speed of 0.4 aa/s is a rather low figure – other estimates have stated translation speeds up to 1 aa/s. The experimentally defined time point for the glycosylation of E33 however was 12 minutes, which was far later than the expected 5 minutes. This showed that the N-terminus was indeed glycosylated after the E161 site, even though the site of E33 was located almost 130 residues upstream of the second luminal loop. This was the first indication that the N-terminus of NS4B indeed translocated across the ER membrane in a posttranslational process.

4.3 Choice of cell lines (paper I, II, III, IV)

In our four studies several different cell-lines were used. For fluorescence microscope studies we mainly used Hep3B (paper I) or Huh 7 (paper I, II, III, IV). Both are human hepatoma cell lines that exhibit large cytoplasms. This made them especially suitable for studying cellular patterns and MAF formations.

Huh 7 is the only cell line reported to support replication of the unadapted pHCVrep1b.BB7 replicon (paper III). The replication of the replicon could however be further enhanced by the use of a “cured” Huh 7 cell line. This was achieved by treating Huh 7 cells that had been supporting replicon replication for nine months and with 1000 IU/ml IFN-\(\gamma\) to the media for 5 weeks. Successful treatment (loss of replicon replication) was verified by the loss of resistance to neomycin, loss of expression of NS3 and NS5 detectable by Western blotting and loss of replicon RNA detectable by RT-PCR.

For the polyprotein study (paper II) and the oligomerization study (paper IV), we used Cos-7 cells. Especially in the polyprotein study, it would have been desirable to use a human hepatoma cell line instead (Cos-7 are derived from Africa green monkey kidney cells). Both Huh 7 and Hep 3B, as well as Hela cells (human cervical cells) were very inefficient in expressing the long polyprotein. Instead we turned to Cos-7 cells that are known for expressing heterologous proteins
at high levels. In the oligomerization study we kept working with Cos-7 cells since the main goal was a high protein expression. The NS4B-NS4B interaction was however also verified in Huh 7 cells.

4.4 The subgenomic replicon (paper III)

The subgenomic replicon system has already been described in section 3.1.4. The replicon used in our study (paper III) was the pHCVrep1b.BB7 that contain an adaptive mutation in the NS5A gene, S2204I (18).

The replicon system is very useful when studying viral replication, as it makes it possible to study the replication process in isolation – apart from viral infection, assembly etc. It should be noted however, that it does not necessarily reflect the efficiency of viral proliferation in humans or chimpanzees (24).

4.4.1 Choice of mutations:

The amino acid changes made to the replicon in this study were made with somewhat different considerations that will briefly be mentioned here (see Fig 13).

The N- and C-terminal mutations:

These mutations were originally introduced in the same replicon, in order to introduce three unique restriction endonuclease sites. The main goals were to 1) not disturb the polyprotein processing and 2) make mutations in areas that were highly variable. However, one mutation (W1963S) had to be made at a conserved site. When the replicon containing all these six mutations turned out not to be able to replicate at all, we wanted to elucidate which of the mutations that caused this effect. Therefore, the exact same amino acid changes were introduced separately into different replicons.

The transmembrane loop regions:

Both amino acids are located in the putative first lumenal loop. This loop was predicted to be very short, and we did not know for certain what residues in this area that could constitute the short bend between TM1 and TM2. This was however a likely site, since prolines have a high turn propensity (123). We would preferably have mutated conserved amino acid residues, but none were available in this region. Since we wanted them replaced with a small neutral amino acid residue instead; they were both mutated to alanines.

Both amino acids located in the first cytosolic loop. G1871 was a conserved amino acid residue and therefore was a natural choice for mutation. Even if not conserved, K1846 had shown to be an interesting site since a K1846T change had proven to be a highly efficient adaptive mutation (104). Both were mutated to alanines since we wanted small neutral amino acid residues.
These four amino acid residues are all located in the second luminal loop and all four are fully conserved between genotypes. M1869, G1871 and P1874 were all exchanged to alanines. We considered this loop long enough to enable protein-protein interactions. Therefore the negatively charged E1877 was exchanged to a positively charged arginine to maximize the effect of the mutation.

![Locations of the 14 point mutations strategically placed in NS4B. The TM domains are shaded in gray, and the amino acid residue changes are indicated.](image)

### 4.4.2 Replicon establishment efficiency

To assess the effect of the mutations in the replicons, we measured their efficiency to establish continuous replication in Huh7 or T#9 cells (= replicon establishment efficiency). This was done by counting the resulting colonies after 10-16 days of growth in selection media.

Huh7 cells, and cell lines derived from them are very unstable by nature (104, 129). This causes a problem when the results are of quantitative nature rather than qualitative; there was a significant difference in absolute numbers of the surviving colonies from the same mutant in different transfection rounds. Yet the ranking order between the different mutants however showed a stable pattern. For this reason we chose to present our results as the relative number of colonies instead of the absolute numbers. The original replicon (S2204I) and the pol(-) replicon (negative control) were always included in each transfection round and the results from them were set as 100% and 0% respectively. To further minimize the effect of the unstable nature of Huh7 cells we also performed the following measures:

- The mutant replicons were transfected both into both Huh7 cells and T#9 cells, to see assure that the ranking order was the same in both cell lines. Moreover, as the T#9 cells generally resulted in more colonies, these cells allowed better dilution series.
- Both cell lines were stored in liquid nitrogen prior to use and discontinued after a maximum of 20 passages for Huh7 and 16 passages for T#9.
- In the literature, replicon establishment efficiency is usually the average from 3-5 transfection rounds. We did 12 and 11 transfection rounds for the N-and C-terminal mutants and inter-transmembrane loop mutants, respectively to obtain reliable results.

### 4.5 NS4B oligomerization (paper IV)

#### 4.5.1 The oligomerization assay

In the oligomerization study (paper IV), we chose to assay for interactions between NS4B molecules that could be detected without crosslinking. This was done by co-expressing two different constructs of NS4B, each having a different epitope tag. The tag of one of the constructs was used for immunoprecipitation and the tag on the other was used for detection (Fig. 14). In this
way we could detect that we had at least a dimerization of the protein. All oligomerization assays reported in paper IV were done in similar manner; the different mutant- and deletion constructs of NS4B were expressed in fusion with EGFP (in the C-terminus) which were then co-expressed with a full length NS4B fused to an Xpress tag (in the N-terminus). To assure that the tags themselves did not interact with NS4B, we did different versions of the oligomerization assay using differently tagged proteins as well as wild type proteins without tags.

The anti-Xpress antibody worked very well in the immunoprecipitation, giving an efficient pull down without much background. One concern with this construct was that it not only added the Xpress-tag to the protein, but also a histidine-tag and a cleavage site for enterokinase. This adds approximately 30 extra amino acid residues to the N-terminus of the NS4B. Considering the length of this tag and also that many of the added amino acid residues were positively charged, we deemed it necessary to make sure that the N-terminus could still translocate when in fusion with this tag. This was done by adding a glycosylation motif (NST) to the N-terminus of the X-NS4B construct and expressing it in vitro in the absence and presence of microsomal membranes. The result clearly showed that NS4B still retained the capacity to translocate the N-terminal tail in the presence of the added epitope (see Fig 15).

Admittedly, it would have been optimal to have the other NS4B in the assay without any tag at all, minimizing the risk of crossreactions, steric hindrance etc. The antibodies against NS4B we had available were all directed against different domains of the protein. Because of this we would have had to use different antibodies for the detection of the different deletion-mutants of NS4B. This would in turn have made it more difficult to compare the results from the different assays. We preferred to be able to use the same antibody for detection of all constructs and to begin with we tried to use smaller epitopes (8 amino acid residues) for the detection. However, a large portion of the
deletion mutants of NS4B seemed then to be rapidly degraded when in fusion with the smaller epitope, and the addition of the larger protein EGFP solved this problem.

4.5.2 Crosslinking

From the oligomerization assay described above, it was not possible to find out the size of the oligomerized complexes, only that they were at least made up by dimers. To have a rough estimation, we added crosslinker to cells expressing X-NS4B and analyzed the lysates by Western blotting. The crosslinker we choose was Disuccinimidyl suberate (DSS), which is an uncleavable, membrane permeable crosslinker that crosslinkes primary amine-groups within a distance of 11.4 Å. Our results indicated that besides dimers there may also be trimers and even larger complexes. However, since we have not analyzed the “larger complexes” in detail, we can not know if those consist only of NS4Bs or perhaps also include other proteins.
5. RESULTS AND DISCUSSION

5.1 Localization and topology of NS4B (paper I and II)

The aim of paper I was to characterize the cellular location and molecular topology of NS4B in detail. At the time when this study was initiated some early studies of its cellular location had been made, but nothing was known about its structure or topology (72, 84, 160). Paper II was a continuation of the topology study of paper I.

5.1.1 Localization

The cellular localization of NS4B was investigated by fusing the protein to EGFP in the C-terminal end of the protein. The fusion protein was then expressed in various different cell lines, which were also immuno-stained against different cell organelles. The resulting cellular pattern was investigated using a fluorescence microscope. We found that NS4B, when recombinantly expressed, localized mainly to the ER, and possibly also partly to the golgi (Fig 16). An unexpected finding was that apart from the typical reticular ER pattern, NS4B could also induce new structures in cells. These were visible as aggregates in the cytoplasm and they tended to grow bigger with time (see Fig. 16). By immuno staining the cells we concluded that also these structures were positive for the ER marker Calnexin, indicating that these structures were rearranged ER membranes.

Rearrangement of membranes seemed to be an intrinsic property of NS4B, although we could not at this point rule out the possibility that this characteristic was restricted to cells overexpressing the protein or that expression of the full polyprotein would have given different results. Similar structures seen in light microscopy has however later been reported in cells expressing subgenomic replicons (44, 55). This would indicate that NS4B indeed has the ability to induce these cytoplasmic aggregates, that we now call membrane associated foci (MAF). Soon before our results were published, Egger and colleagues reported a study where they had seen NS4B-induced structures with electron microscopy (in contrast to ours that used light microscopy) (42). They found that when expressing the complete polyprotein in mammalian cells, it induced a tight structure consisting of small vesicles embedded in a membranous matrix. This structure was called a “membranous web” (MW) and it was found to be induced by NS4B. They also showed that all other HCV encoded proteins located to these structures. Similar structures have previously been seen in HCV infected livers of chimpanzees, indicating relevance during a true

![NS4B-EGFP, Calnexin, Merger](image)

Fig 16. NS4B-EGFP expressed in cells that have also been labelled with the ER marker Calnexin. MAF structures induced by NS4B are clearly visible in the cytoplasm, seen as brightly fluorescing dots. These were also stained by the anti-calnexin antibody, showing colocalization of NS4B and ER membrane.
viral infection (142). The relationship between MAFs (seen in light microscopy) and MW (seen in electron microscopy) is not yet known (see Fig 5 for comparison). To date, we cannot know if the structures are related, but it seems likely for several reasons. First, both MAFs and MW are induced by NS4B alone. Second, if they were not the same, the MAF structures should have been seen in the EM study as separate structures. Third, the MW was labelled with antibodies recognizing all other NS proteins and no label was observed on any other membrane alterations (42). This was true also for the MAFs (44, 55).

All RNA viruses known to date rearrange intracellular membranes, a process assumed necessary for viral replication complex (see section 2.2.2). HCV is no exception and as paper I show, NS4B is the protein responsible for this mechanism. NS4B cannot be complemented in trans to rescue replication (6). Maybe the explanation behind this is the same as for poliovirus; viral replication needs membrane modification, viral translation and RNA synthesis working together in a coupled process (41). Also in other members in the Flaviviridae family have NS4B been shown to be required in cis configuration (58, 81). However, NS4B does not induce membrane rearrangements in Dengue virus or Kunjin virus (120, 127, 182), indicating that this function may be specific for HCV NS4B.

5.1.2 Topology

We showed in paper I, in agreement with others, that NS4B is an integral membrane protein (68, 72, 170). Therefore, we next set out to determine its topology in the ER membrane. Besides a general interest for structure of this protein, we considered that this type of information could be useful for future studies on the function of NS4B.

To create a preliminary model of NS4B in the ER membrane, the amino acid sequence of the protein was analyzed by three different prediction programs; TMHMM, HMMTOP and MEMSAT ((75, 89, 175). All programs predicted that the protein had four TM segments with the N- and the C-termini in the cytosol. This corresponded well with current knowledge of HCV polyprotein processing; the NS3/NS4A protease complex that cleaves NS4B from neighbouring proteins is located on the cytosolic side of the ER membrane. The predictions also indicated that the loops between the TM segments could be expected to be very short, most likely just bends. The prediction of the N-terminal half of the NS4B was however more uncertain than the C-terminal half. All three programs agreed in the placement of the third and the fourth TM segments and they also clearly separated them from each other by a loop. MEMSAT however, indicated another location of the first segment than did TMHMM and HMMTOP (Fig 11). The difficulties in predicting the topology in the N-terminal half of NS4B was also reflected by the fact that TMHMM gave lower probability scores in this region (Fig 10).

To test the predictions experimentally, we added glycosylation sites at seven different locations in the protein. Since the loops between the TM segments were predicted to be very short, an insertion of 30 amino acid residues with a glycosylation motif in the middle had to be added to the sites located in the transmembrane region (the EC4 loop(148)). This was necessary as a distance of 12-14 amino acid residues between the acceptor asparagine to the membrane is required for a glycosylation to occur (136). The same insertion was also made in the N- and C-terminal tails to make results easy to compare.

Our experimental data, based on the seven glycosylation mutants, were only partially consistent with the predicted topology. The second luminal loop (construct E161) was glycosylated as expected, indicating luminal localization. In paper I however, we failed to experimentally show the presence of the first two TM segments and the luminal loop between
them. We still believed that NS4B had four TM segments, and we also believed that the first lumenal loop was located around residue 112 in the NS4B sequence. An indication of this was that the construct E112, with the EC4 loop inserted after the 112\textsuperscript{th} amino acid residue, sometimes gave a weak glycosylation. The results were however inconsistent and we chose to report this site as negative. If the luminal loop was indeed located around this site, we reasoned that the lack of glycosylation at this site may be due to one of two reasons. Either the insertion of the long EC4 loop could easily have disrupted critical interactions between the two helices TM1 and TM2 that may be required for correct membrane insertion. Or, the EC4 loop had been inserted somewhat askew and not at the very top of the bend between TM1 and TM2. In that case, the distance from the ER membrane would be slightly too short for an efficient glycosylation. The latter case could easily be tested by elongating the EC4 loop and thereby increasing the distance to the membrane. This was done in paper II, and the addition of three amino acid residues to each end of the EC4 loop at this site resulted in a more efficient and also consistent glycosylation. This clearly indicated that the first luminal loop was indeed located in the vicinity of the 112\textsuperscript{th} amino acid residue of the NS4B sequence. It would have been interesting to move the EC4 loop to different positions around the 112\textsuperscript{th} residue, and thereby find the “top” of the first luminal loop.

An unexpected finding in paper I was that the N-terminal end of NS4B also had a luminal position (construct E33). It was unexpected, since both the N- and C-termini of NS4B are cleaved from the polyprotein by the cytoplasmic NS3/NS4A protease. In paper I, we had only expressed NS4B by itself in vitro, and could therefore not exclude the possibility that it may behave differently if expressed in cells and in presence of the other NS proteins. But as was shown in the kinetics study in paper I, the translocation obviously occurred as a posttranslational event. This strongly suggested that this was a general mechanism of NS4B and not the effect of a misfolding of the protein (Fig 17).

Paper II continued to study the translocation of the N-terminal end of NS4B in detail. We found that it still occurred when the protein was expressed in cells, and also in a polyprotein
context. When expressing NS4B from a polyprotein encoding NS3-NS4A-NS4B, the translocation of the N-terminus was still efficient, but when NS5A was part of the polyprotein the translocation was significantly inhibited, suggesting that the presence of NS5A somehow influences this process (Fig 18). As NS5A is always present in the full polyprotein, an obvious objection is that the translocation of the N-terminus of NS4B may never actually occur in an HCV infection. There are several lines of evidence speaking against this though. First, in the presence of NS5A approximately 15% of the proteins still translocated their N-termini (as compared to 60% in its absence). Second, we also showed in paper II that the N-terminal translocation is a feature conserved between genotypes. Admittedly, this part of the study was made in vitro and also on NS4B expressed alone, and not in a polyprotein context. It is however difficult to imagine such a feature being conserved without having a function in the viral life cycle.

Thus, that the N-terminus of NS4B had a luminal location also after having been processed from NS4A (paper II), strongly supported the conclusion from the kinetic assay in paper I – that the relocation of the N-terminus was the result from a posttranslational process (Fig 19). The mechanism behind this event may be the same as for posttranslational translocation of whole proteins, which occur in all mammalian cells. In yeast it is a common and quite well characterized process, but it may be more unusual in mammalian cells (154). Another possibility is that the translocation of the N-terminus uses the same pathway as the tail anchored proteins do (19). Especially viruses with their condensed genomes may benefit from a posttranslational translocation of a protein already inserted in the membrane; in this way the viral protein can achieve a dual topology and thereby also two different functions. (see for example the Hepatitis B virus L protein (22, 23, 93, 95, 134, 137, 149). Also the HCV protein p7 has been reported to have an alternative topology when in the precursor protein E2-p7 (74). Currently, we cannot know if both topologies of NS4B (4 TM segments vs. 5 TM segments) have functions in the viral life cycle. But when destroying the amphipathic helix reported to be located in the N-terminal end of NS4B, we simultaneously abolished the ability of the protein to create MAF structures as well as to translocate its N-terminus to the ER lumen (paper II and (45)). This indicated an association between the 5 TM helix topology and the ability to rearrange membranes. This was further explored in paper IV (see below).

The NS4B proteins of other members in the Flaviridae family are as ill defined as for HCV. The sizes and hydrophobicity plots of the corresponding proteins are however similar for viruses in all genera in the family. Interestingly, the NS4B proteins of Yellow fever virus and Dengue virus (both flaviviruses), have been reported to have their N-terminal tails in the lumen of the ER. In these cases however, translocation of the N-terminal tail is initiated by cleavable signal sequence and not posttranslational translocation as for the corresponding HCV protein (25, 99, 120). In the case of Dengue virus, it was further shown that the 2K signal sequence preceding NS4B was not required for integration in the membrane, indicating that the role of the signal
sequence is primarily to help the protein achieve the right topology with the N-terminus in the lumen (120). A shared topology achieved in two different ways may suggest a common function for NS4B in the Flaviviridae family.

The HCV protein NS4B localizes to the ER membranes in mammalian cells. It also has the ability to create new structures in the cytoplasm by rearranging the ER membranes. The new structures, called membrane associated foci, are visible in light microscopy as dense aggregates in the cytoplasm. NS4B is integrated in the ER membrane and has initially four transmembrane segments and the N- and the C-termini in the cytoplasm. Via a posttranslational mechanism the N-terminus of the protein is relocated to the ER lumen giving the protein a fifth TM segment. NS4B of all major genotypes have the ability to translocate the N-terminus. When NS4B is expressed in a polyprotein context, the N-terminus still translocates, but less efficient when NS5A is expressed from the same polyprotein. This indicates that NS5A somehow influences the translocation of the N-terminus of NS4B.

5.2 Mutational analysis of the different domains of NS4B and their role in replication (paper III)

The initial idea behind this study was to remove NS4B from the subgenomic replicon to assess its importance for replication. We also wanted the possibility to exchange the NS4B sequence with a reporter gene (e.g. EGFP), NS4B of other genotypes or NS4B carrying designed mutations. In order to be able to easily remove or insert sequences at the site of NS4B, we introduced three different restriction sites in the replicon; KspI, PacI and PvuI. This resulted in six amino acid changes in the N- and C-terminal ends of NS4B; L1715R, P1716G, Y1717L, E1719K, W1963S, and N1965D. The mutations were designed to be in the proximal ends of NS4B, but well distanced from the sites needed for poly protein processing (P6, P4, P1 and P1', see (9, 56, 168). We also tried to introduce the mutations in areas that were highly variable. One of the mutations (W1963S) was however at a conserved site. The new replicon containing the restriction sites was not able to replicate. It was therefore obvious that one or more of these mutations disturbed some essential function(s) of NS4B in the replication process. To elucidate which of these mutations that caused this effect, the same mutations were introduced into six separate replicons (Fig 13 and 20).

Five of the six mutations resulted in replicons that produced fewer colonies than the original, and one of those totally abolished replication (W1963S). One of the mutations on the other hand gave higher survival efficiency than the original (L1715R). According to literature the mutations should not interfere with polypeptide processing, but we decided to test this possibility experimentally (9, 56, 168). When testing this by an in vitro assay we found no significant difference in the polypeptide processing between the different mutants.

The L1715R mutation in the N-terminal end of NS4B changed a non-polar residue to a positively charged one. Even though this resulted in a more efficient replicon establishment, the same mutation can only be found in rare clinical isolates. Mostly, the L1715 residue is exchanged to similar amino acids such as isoleucine or alanine. This may imply that the site is important for a function in the viral life cycle not connected with replication. Interestingly, the N-terminal region of NS4B has previously shown to influence cell tropism (192) and interferon resistance, both in HCV and Dengue virus (127, 130). Furthermore, a single mutation in the N-terminal region of NS4B of BVDV altered the cytopathy of the virus, independently of viral RNA or protein levels.
in the cell (150). All this imply an important role for this region of the NS4B – not necessarily coupled to the replication process.

The tryptophane at site 1963 on the other hand is conserved, and the exchange to serine (W1963S) fully abolished replication. Tryptophane has been reported to be over represented at interaction sites with other proteins (163). In paper IV we showed that NS4B has the ability to homo-oligomerize and also that the C-terminal end of the protein was required for this interaction. A possibility is that the W1963 plays a role for this interaction, but this has not been tested. Furthermore, the W1963S could still create MAF structures; an ability we believe is connected to oligomerization of the protein (paper IV). Interestingly, no surviving colonies for the W1963S mutant could be found, implying a function of NS4B that could not be compensated for by mutations in other parts of the polyprotein or interacting cellular proteins. This in turn suggests that this site is involved in some other function than protein binding, maybe by having some intrinsic function or binding of RNA?

To complement the above results we also introduced mutations in the predicted lumenal and cytoplasmic loops between the TM segments. For these mutations we used another approach than above, and choose to mutate conserved amino acid residues where such were available (Fig 20).

One concern with the mutations in the TM loop regions was the risk of protein misfolding. As the loops were predicted to be very short, each amino acid residue in them may be important for correct folding, especially in the short loops between TM1-TM2 and TM2-TM3. Correct folding is almost certainly a prerequisite for a protein to function correctly. Therefore, all 14 mutants were tested for the only known function of NS4B - the ability to rearrange intracellular membranes into

![Diagram](image_url)

**Fig 20.** Location and results of the 14 replicon mutants. The replicon establishment efficiencies are listed to the left in the picture, and also displayed using a colour gradient in the topology picture. Mutations of conserved amino acid residues are depicted with blue circles, and the mutation that abolished replication depicted as a cross. The table below shows a list of amino acid residues that have been found in clinical isolates at the non-conserved sites. The list is in the order of most frequent to the least common. Bold red lettering indicate amino acid residues found in genotype 1.
MAF structures. We found that all mutants retained this ability, at least as far as could be judged by light microscopy, and therefore assumed that they had correct folding. Furthermore, this also implied that NS4B have additional roles in viral replication than merely providing the necessary environment.

Replicons having the mutations on the lumenal side of the ER membrane were also affected negatively, and in some cases they almost abolished replication. The replication complex (RC) is located on the cytosolic side of the ER membrane. The only parts of the RC that could possibly interact with the lumenal loops of NS4B are the small parts of NS4A or NS5B that anchors them to the membrane (see Fig 2). NS4B also has the ability to interact with itself, but as we show in paper IV, the TM segments or the loops between them are not involved in this oligomerization. Hence, the most likely interacting partners for NS4B in the ER lumen are cellular proteins or other smaller substrates. Unfortunately, there is scarcely any information about interacting cellular proteins to NS4B, especially not for those that bind to the hydrophobic mid-domain of the protein.

In Paper III we tried to obtain an overview of the role of the different parts of NS4B in the replication complex by introducing single point mutations at various positions in NS4B. All mutations except one affected replicon establishment efficiency negatively. The effects could neither be ascribed to incorrect processing nor inability of the mutants to rearrange intracellular membranes, implying a role for NS4B in replication besides just providing the necessary environment for viral RNA synthesis. Furthermore, the amino acid change that improved efficiency (L1715R) only occurs very rarely in clinical isolates. Possibly, this mutation may negatively affect some other process(es) in the viral life cycle despite improving replication, in turn suggesting that NS4B has functions besides those in the replication process. Even though NS4B is not well conserved between genotypes, our data show that the molecular integrity of NS4B is vital for the HCV replication process.

5.3 Correlation between homo-oligomerization, N-terminal translocation and MAF induction (paper IV)

In paper IV we were interested in finding the relationship between the N-terminal translocation of NS4B, and MAF formation – if any. As reported by us and others in earlier studies, NS4B with an abolished amphipathic helix (AH) in the N-terminus lacked the ability both to create MAFs, to translocate its N-terminus into the ER lumen, and also to replicate (paper II and (45)). This suggested that these functions were interlinked, and we were interested in finding the relationship between the first two. The most probable mechanism of the membrane altering capacity of NS4B involves either a particular intrinsic property of NS4B or some protein-protein interaction. As has been shown in several studies, the other NS proteins of HCV are not required to induce membrane rearrangements (paper I, II, III (42, 59, 84)). Therefore, we also wanted to investigate if NS4B possibly could homo-oligomerize, and in that case how this function is coupled to induction of MAFs and N-terminal translocation.

Initially, we found that NS4B indeed had the ability to homo-oligomerize, and that both the N- and the C-terminal domains were necessary for this interaction. Possibly this involved a disulphide bond formation in the C-terminus. The N-terminal domain was obviously important too, but we had indications of a weaker interaction in this area than in the C-terminal part. We found that it did not matter on what side of the ER membrane the N-terminus of NS4B resided on
proteins of both topologies were able to interact. Homo-oligomerization of NS4B was also recently reported both for HCV and Dengue virus (176, 189).

As the N-terminal part of NS4B had shown several interesting features mentioned above, we wanted to identify what regions that were responsible for these different functions. According to our topology model, current literature, and other reports we divided the N-terminal domain into three different regions and deleted them accordingly (see Fig 21); AH1, AH2 and TMX. AH1 was the amphihelix reported by Elazar and colleagues (amino acid residue 6-29, (45)), and AH2 was a second amphipathic helix that had been proposed by Castet et al at the 11th International Symposium on HCV and Related Viruses (amino acid residue 40-69). TMX was found by our own computer predictions and glycosylation scan (located approximately between amino acid residue 60-84, paper IV).

All three deletion mutants were tested for their abilities to homo-oligomerize, translocate their N-termini and to induce MAF. The results are summarized in Fig. 21. We found that the homo-oligomerization and the MAF induction seemed to go hand in hand, at least in these rather crude deletion mutants. The deletion mutant lacking AH1 (NS4B(30-261)) was able to perform all three processes. This was intriguing, since a full length NS4B with a disrupted AH was not able to perform any of them (AHmut). The one lacking both AH1 and AH2 (NS4B(60-261)) had lost its ability to homo-oligomerize and to form the typical MAF structures. And the one lacking also TMX (NS4B(92-261)) was unable to perform any of the three processes. We concluded that the AH1 is not needed to perform any of the processes we had been looking at. AH2 on the other hand seem to be important both for homo-oligomerization and MAF formation, but not for the N-terminal translocation. Interestingly, AH2 cannot perform any of these processes in the presence of a mutated AH1.

Our results suggested that a mutated AH1 may inhibit the N-terminal translocation. By doing this, the distorted AH1 is locked on the cytoplasmic side of the ER membrane as is AH2 and TMX. By removing the mutated AH1 (NS4B(30-261)), or by forcing it to the lumenal side of the ER membrane (ssAHmut), any inhibitory effects performed by it on the cytoplasmic side were abolished. As was also shown, at least oligomerization can occur with the N-terminus of NS4B still on the cytosolic side of the membrane, suggesting a deleterious interaction between the mutated AH1 and AH2 on this side of the membrane. Obviously however this interaction did not cause any negative effects if the two AHs were located on luminal side of the membrane. Instead,
AH2 is fully functional in the ER lumen and able both to interact with other NS4B molecules and to induce MAF formation.

The conclusions from this study are based on the results from rather crude deletion mutants. It would be interesting to have a more detailed picture of the processes described. First, by disrupting the AH1 with less pronounced mutations might have given different results. Also AH2 should be destroyed by mutations instead of being deleted. Second, it would be possible to quantify the oligomerization data and also to employ electron microscopy for studying the membrane rearrangements. By doing this it would be possible to obtain a qualitative comparison between the different mutants. Our use of light microscopy served our purpose – to distinguish between mutants that were able or not able to induce structures in cells looking like MAF. But we cannot actually be sure that these structures are exactly similar. Third, it would be very interesting to “lock” the N-terminus of NS4B on the cytosolic side of the ER membrane without having to mutate the actual protein sequence (as was the case with AHmut). Only by doing this can we know for sure about the importance of the N-terminal translocation.

In this study we found that NS4B has the ability to homo-oligomerize and that this may be connected with its ability to form MAF. The C-terminal domain is required for this to occur, as is an amphipathic helix located in the N-terminal part (located approximately between a. a. residues 40-69). The relocation of the N-terminus to the ER lumen may be a prerequisite for these processes to occur. We also determined that TMX, the TM segment created by the N-terminal translocation, is located downstream of amino acid residue 60 in the protein.
6. CONCLUDING REMARKS

When we started our studies of NS4B, almost nothing was known about the characteristics of this protein. What are the first steps towards the understanding of an almost unknown protein? We considered location in cells and finding putative interaction partners to be important goals in this characterization.

When studying the location of the NS4B in cells, we found that it induced new membrane structures in the cells, visible as dense aggregates in the cytoplasm (MAF) and the result of rearranged ER membranes. While conducting this study this was also seen and reported by other groups (42, 84). In electron microscopy the changes were visible as small vesicles enclosed in a membraneous matrix (42). HCV had been known to induce structural changes in infected cells (142), just as other positive stranded RNA viruses in order to create a favorable environment for viral replication. That NS4B was the protein behind this effect was therefore an important finding. The underlying mechanism behind this event is unknown however, and we can only speculate. One possibility is that the protein itself has a conical shape, and as an integral protein in the ER membrane it could therefore induce a positive curvature of it. This could in turn create vesicles, especially as the protein has the ability to oligomerize. Another possibility is that NS4B induces the concentration of certain lipids in the ER membrane, which may promote the deformation of membranes (15). Indeed, the replication complex of HCV, and specifically NS4B, have been found associated to areas of the ER membranes having specific lipid compositions (lipid rafts or detergent resistant membranes, (4, 52, 162)). In paper IV we found that the second amphipathic helix in the N-terminus of NS4B played a role for the induction of MAFs. Yet another suggestion is therefore that this AH structure is inserted as a wedge from the luminal side of the ER membrane and thereby induces a positive membrane curvature that can enclose and protect the replication complex.

Our second original goal was to find interaction partners to NS4B. There was however no good interaction assays for integral ER proteins available. Therefore we decided first to map the topology of the protein in the membrane to identify interesting domains that could be assayed for interactions. During the topology mapping we got the unexpected but intriguing result of the posttranslational relocation of the N-terminus into the ER lumen. This raised the interesting possibility that NS4B exist in two different topologies. There could be several different reasons for this. First, it may be a way for the virus to use the same protein for several different functions, as has been reported for other proteins with dual topologies (64, 97). Second, the translocation may be some kind of regulatory event, marking the transition between two different phases in the viral life cycle. Or third, NS4B of both topologies may interact with each other, and may therefore exist at the same time and needed for the same function. It is easy to imagine that the N-tail needs to be kept in a relatively unfolded state in order to be able to translocate across the ER membrane. It could therefore be interesting to search for interactions between the N-terminal part of the protein and cytosolic chaperones.

So what about our intentions of finding interaction partners? In collaboration with other groups we tried to screen for interaction partners to the N-terminal and C-terminal domains using yeast-two-hybrid assay. Both domains were predicted to be highly structured and therefore likely to be able to bind to other proteins. From the putative interactants found, we found the p11 protein to be highly interesting, as it has been found to be involved in exocytosis. Unfortunately, we were unable to confirm this interaction. But the same interaction was also reported in an oral presentation by Charles M. Rice on the 12th International HCV symposium (October 2005).
A curious finding was that NS4B was so extremely sensitive to mutations. Even single amino acid mutations in non conserved regions affected replication both positively and negatively. This despite the fact that the NS4B amino acid sequence is quite variable between genotypes. Clearly it posed a problem when wanting to study NS4B in a larger context, as designed mutations or inserted epitopes almost certainly would disrupt replication. Consequently, one may object to any conclusions drawn from mutated NS4Bs, since they probably will not be functional in an infectious virus. We believe however, that studying NS4B in isolation can still provide us with important information about the protein per se. Subsequently, particularly interesting mutants may be introduced into the recently developed HCVcc propagation system and assessed for virological relevance and fitness.

So far, we have only studied the soluble parts of the protein. It would be interesting also to look closer at the TM segments of NS4B. They are rich in glycine residues, and this small amino acid creates “holes” in the alpha helical structure. Glycine are therefore common at interaction sites in TM segments, as these holes can be filled with the side chains from larger hydrophobic amino acid residues in neighbouring segments. Notably, the GxxxG motif can be found at two different places in the TM domain of NS4B (in TM2 and TM3). As this motif is frequently found at interaction sites between TM helices (157, 161), this does indicate that the TM segments of NS4B are indeed involved in protein-protein interactions. Either within the protein itself or with other proteins in the ER membrane. This is further supported by the fact that these two motifs are fully conserved between genotypes.

Unfortunately we do not know the full function of NS4B yet. We know that it rearranges the intracellular membranes to create MAF – structures that are needed for viral replication. Obviously however, NS4B have also other, unknown functions. About these we can only speculate. Most probably it plays an important role for the viral replication process. The results from our 14 replicon mutants clearly suggest this, as does the fact that NS4B can interact with other proteins in the replication complex, and the adaptive mutations found in NS4B (36, 52, 61, 104). We can however only guess about its role in this process. An interesting finding was that the N-terminus of NS4B, when expressed on its own, located to lipid droplets (paper IV). The core protein of HCV locates partly to the ER membrane and partly to lipid droplets (8, 113), and lipid droplets bud off from the ER membrane. If the N-terminus of NS4B has the ability to bind to areas in the ER membrane that will subsequently bud off to form lipid droplets (areas where core may assemble), there is a possibility that NS4B forms the connection between the replication complex and the core proteins that will subsequently form nucleocapsids. Perhaps NS4B forms a channel in the membrane surrounding the RC, transporting newly synthesized RNA to the site of assembly? Partial colocalization has indeed been found between core and part of the replication complex (NS3) in cells infected with HCVcc (156).

We cannot tell the full story of NS4B yet – much is still left to be discovered. Our characterization of NS4B has however provided an improved understanding of the protein that in turn will facilitate future studies. Recently a new chapter in HCV research commenced, when the possibility to grow virus in culture emerged. It is now possible to study the proteins in its real context, and hopefully this will provide more answers also concerning the role of NS4B in the in the viral life cycle.
7. ACKNOWLEDGEMENTS

Many people have contributed to this thesis directly or indirectly. I would like to express my sincere gratitude especially to the following persons

My supervisor Mats Persson for your open door, generosity and enthusiasm. You have always believed in me and let me follow my own ideas – both when good and not so good…

My co-supervisor Anders Widell for providing the clinical aspect of HCV, and for knowing all the good (no matter how small) restaurants worldwide.

Gunnar von Heijne for valuable discussions and inviting me to work in your lab. This provided me with a solid ground to stand on when venturing into the area of hydrophobic proteins…

All members of the MAAP group:

Hannah, for being a good friend and an inexhaustible source of odd knowledge picked up from the Discovery channel. But most of all, for joining me in the struggle to understand NS4B and for discussing models, ideas, and suggestions about this protein at any time – either by SMS 3.30 in the morning or during our numerous coffee breaks.

Aster, for your warm personality and for taking us to “Starlight”.

Mikaela, for being the invaluable organizer and for making the best cakes…

Fayezeh, for knowing everything that happens in CMM, and in hard competition with Mikaela – for making the best cakes.

Daniel, for teaching me to climb higher (literally speaking).

Johan, for tipping me off about Åland – that became my favourite place for kayaking.

Katarina, for lunch-skating on Brunnsviken and for taking care of me when I first came to the lab.

Also thanks to Daniela, Lubna, Alexej, Kathrin, Ann-Katrin, Anila, Tobias, Oguz, Ingrid, Karin and Marko for the time you spent with the MAAP group, and contributing to the friendly atmosphere in the lab.

Special thanks to my former students Emma, Caroline and Sandra - for asking all those questions that made me reassess what I thought I knew…

All friends at CMM:01 - former and present members of the groups of Georgy Bakalkin, Tomas Ekström, Catharina Larsson and Gunnar Nordstedt and Lars Terenius. There have been many enjoyable lunch and coffee breaks with different combinations of people over the years and everyone has contributed to interesting discussions – with topics spanning anything between science (occasionally) and reality shows…

Delphi Post and Catarina Zanninni for taking care of everything in CMM.

Kåre Hultén for bringing good coffee to CMM:01

Bob Harris, for teaching me the microscope at the very beginning of this project. Thank you also for linguistic advice.
Magnus Monné, IngMarie Nilsson and Marika Hermansson for knowing so much about membrane proteins, for making me feel at home the GvH lab, and also helping me with practical issues.

All the organizers, students and supervisors at Uppsala Graduate School in Biomedicine (UGSBR)

All dear friends outside work - You know who you are. You have all contributed indirectly to this thesis by always being there and providing a space that I can be sure is 100% free of NS4B or anything remotely connected to it.

Both my wonderful families Lundströms & Lundins. Special thanks to Mamma Margareta, Pappa Ingmar, Johanna, Kerstin and Enrico for all help with Nora and Jakob during the writing of this thesis. Extra thanks to Johanna for valuable comments on the manuscript.

Nora and Jakob for your existence and for giving me perspective.

Douglas, for love, support and understanding. We have been joking about this worn phrase, but there is no other way to put it. Your support during my thesis writing has been immense, your understanding for me never being home is impressive and your love despite of all this is beyond belief. Also for all Seinfeld quotations and the great laughs following them. Love you.

This work has been supported by the Swedish Strategic Research Foundation (I&V programme), the Swedish Research Council, the Swedish Cancer Society, the Torsten and Ragnar Söderberg Foundations, and the Department of Medicine, Karolinska Institutet.

... and thanks to everyone that reads more in this thesis than pages 45-46. You made the effort worthwhile.
8. References


