ANALYSIS OF DIVERSITY OF HEPATITIS C VIRUS GLYCOPROTEINS E1 AND E2

Natalia Hudson
BSc, MSc Biotechnology

Department of Microbiology and Infectious Diseases
Institute of Infection, Immunity and Inflammation
School of Molecular Medical Sciences
University of Nottingham

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2012

Abstract

Hepatitis C Virus (HCV) exists as a population of sequence variants that evolves during infection adapting to host pressures. The main targets for the immune response are the envelope glycoproteins E1 and E2, which also mediate viral cell entry. The first hypervariable region (HVR1) of E2, previously implicated in the outcome of acute infection, has been a focus of many studies. However more broadly neutralising antibodies tend to target epitopes outside this region, yet evolution of full length E1E2 heterodimer is poorly understood. The HCV transmission and window period as well as seroconversion are the evolutionary events shaping primary infection hence influencing outcome of acute infection. However, due to the asymptomatic character of the early phases of HCV infection, evolutionary data describing this interval is still lacking depth. Defining the genetic and phenotypic characteristics of HCV population of sequence variants that establish infection in a new host would aid vaccine and new therapy design.

This study aimed to identify patterns of HCV envelope glycoprotein evolution upon transmission and during early phases of disease. We studied this in three settings: experimental transmission of immunocompromised mice by known inoculum; occurrence of horizontal transmission in a haemodialysis unit between hypothesised source and index case individuals; and unrelated cases of acutely infected HCV patients. The single genome amplification (SGA) approach was utilised, which allowed us to accurately assign linkage between substitutions and determine the frequency distribution of E1E2 variants in analysed viral populations.

Data from the first experimental setting indicates that a selective sweep occurs upon HCV transmission, with selective amplification of envelope sequence variants that

possess fitness advantage at entry level. Molecular determinants associated with this enhanced infectivity have also been identified. In further part of the project we confirmed a horizontal infection in haemodialysis unit with use of phylogenetic methods and suggested revision of current safety guidelines. Analysis of sequences from the last setting showed that indeed HVR1 might not be a good enough indicator of evolutionary events in the acute phase, as linked substitutions occur also outside this region. Seroconversion is associated with increasing population diversity indicating role of antibodies in driving HCV evolution, which is host specific.

Publications

Brown, R.J.P., **Hudson, N.,** Wilson, G., Ur Rehman, S., Tarr, A.W., Jabbari, S., Hu, K., Borrow, P., Joyce, M., Lewis, L., Zhu, L.F., Law, M., Kneteman, N., Tyrell, D.L., McKeating, J.A. and Ball, J.K. (submitted) Hepatitis C Virus Envelope Glycoprotein Fitness Defines Virus Population Composition Following Transmission to a New Host. *Journal of Virology*.

Brown, R.J.P., Koutsoudakis, G., Urbanowicz, R.A., Mirza, D., Albecka, A., Price, L., **Hudson, N.,** Descamps, V., Backx, M., McClure, C.P., Irving, W.L., Duverlie, G., Dubuisson, J., Perez-del-Pulgar, S., Forns, X., Tarr, A.W. and Ball, J.K. (accepted) Experimental data reveals parallel divergent mechanisms contribute to global hepatitis C virus envelope glycoprotein evolution. *PLoS Pathogens*.

Thomson, P.C., Williams, C., Aitken, C., Ball, J., **Wysocka, N.,** Brown, R. and Rodger, S.R. (2011) A case of hepatitis C virus transmission acquired through sharing a haemodialysis machine. *NDT Plus 4: 32-35***.** First published online October 15, 2010 doi:10.1093/ndtplus/sfq181

Declaration

I hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged. I was not involved in collection of either serum samples or experiments with the xenomice model. Source of initial samples that were used for nucleic acid extractions and further analysis, presented in this thesis, are acknowledged in the text.

Signature:

Date:

Acknowledgments

Firstly I would like to express my gratitude to Professor Jonathan Ball, whose great ideas and hard work resulted in establishing the EIHCV network. I appreciate every discussion, travel, meeting, cultural experience and scientific exchange that I have experienced as part of this successful network. This has allowed me to grow as a researcher and a person. I would like to thank all the partners in EIHCV network for helpful discussions and fruitful meetings with special thanks to Professor Jean Dubuisson for the training I have received in his laboratory. Thanks are also in place to European Framework 7 for providing financial resources to our network.

I would like to acknowledge hard work of Doctor Richard Brown who has assisted difficult troubleshooting throughout the duration of the project, has been a source of motivation during the writing period and has always been a mentor and a friend.

I am grateful to Doctor Alexander Tarr for endless conversations about tissue culture, sharing priceless antibodies and introducing me to the 'Foundation' by Isaac Asimov. Thanks to Doctor Patrick McClure for inviting me to his laboratory practicals, where I could learn from two exceptional teachers. I am thankful to all the members of the Virology Group, for making me feel at home, discussing techniques, sharing reagents and publications as well as organising curry nights and bringing exotic sweets from abroad.

Special thanks go to members of Sue Ryder Care Centre for the Study of Supportive, Palliative and End of Life Care and administrators in the School of Nursing, Midwifery and Physiotherapy of The Nottingham University. Not only had it turned out to be an interesting job supporting driven researchers, but also a source of support that was stimulating my drive. My special gratitude goes to Professor Jane Seymour, Doctor Ruth

Parry and Doctor Kristian Pollock for constant interest in my progress and endless good advice.

My special thanks go to all my great friends who were there for me over last years. I would like to thank my housemates: Hannah, Luke and Dave for introducing me to the ways of life in Britain and never letting me feel like an outsider. I will forever be grateful to Deeman, Victoria, Emily, Magdalena and Avika who patiently helped me through my depression and never allowed me to give up.

I would like to express gratitude to my whole family, including the recently acquired members. Special and biggest thanks go to my parents who have never doubted in my decisions and supported me morally and financially along the way. I am grateful for having been taught that setting high standards and hard work are the way to success.

To my husband- thank you for never-ending patience, motivation and love. I would not have reached the finish line without you!

Natalia

Contents

List of figures

List of tables

Abbreviations

AAT	Alanine Amino Transferase
Alb/uPA	Urokinase plasminogen activator transgene driven by an albumin promoter
AP	Alkaline Phosphatase
APS	Amonium PerSulphate
AspAT	Asparagine Amino Transferase
ATP	Adenosine Tri-Phosphate
BCA	Bicinchoninic acid
BiP	Binding Immunoglobulin Protein
BMI	Body Mass Index
BSA	Bovine Serum Albumin
C	Core protein
CD36	Cluster of Differentiation 36, scavenger receptor
CD81	Cluster of Differentiation 81, tetraspanin
cDNA	Complementary DNA
CLDN	Claudin
CMV	Cytomegalo Virus
DAA	Direct Acting Antivirals
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
ddNTP	Dideoxy Triphosphates
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose Nucleic Acid
DTT	Dithiothreitol
E1	Envelope Glycoprotein 1
E2	Envelope Glycoprotein 2
ECL	Enhanced Chemiluminescence Solution
EDTA	Ethylenediaminetetraacetic acid
EIA	Enzyme ImmunoAssay
ELISA	Enzyme Linked Immunosobent Assay
ER	Endoplasmic Reticulum
FBS	Foetal Bovine Serum
Fc	Fragment Crystallisable region
G418	Geneticin
GNA	Galanthus Nivalis Agglutinin
GST	Glutathione S-Transferase
GTP	Guanosine Tri-Phosphate
HAV	Hepatitis A Virus
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HCVcc	Hepatitis C Virus cultured particles
HCVpp	Hepatitis C Virus pseudoparticles
HDL	High Density Lipoprotein
HEK293T	Human Embryo Kidney cell line
HepG2	Human Hepatocellularcarcinoma cell line
HIV	Human Immunodeficiency Virus
HRP	Horse Radish Peroxydase
Huh7	Human Hepatocellularcarcinoma cell line
HVR	Hyper Variable Region
IFN	Interferon
Ig	Immunoglobulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRES	Internal Ribosome Entry Site
IRF3	Interferon Regulatory Factor 3
IPS	Interferon Promoter Stimulator-1
ISDR	IFN sensitivity-determining region
ISG	Interferon-Stimulated Gene
IVDU	Intravenous Drug Users
JFH-1	Japanese Fulminant Hepatitis-1
LB	Luria Bretani broth

LDL	Low Density Lipoprotein
LEL	Large Extracellular Loop
LPG2	Laboratory of Genetics and physiology 2
L-SIGN	Liver and lymph node Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
LVP	Lipo-Viral Particle
MDA5	Melanoma Differentiation-Associated gene 5
MHC	Major Histocompatibility Complex
MLV	Murine Leukemia Virus
MVA	Modified Virus of Ankara
NANB	Non-A-Non-B
NJ	Neighbour-Joining
NLR	Nucleotide oligomerisation domain-Like Receptors
NS	Non-Structural protein
OCLN	Occludin
Opti-MEM	Reduced serum media
ORF	Open Reading Frame
PAMP	Pathogen- Associated Molecular Patterns
PBS	Phophate Buffered Saline
PBST	Phophate Buffered Saline with 0.05% Tween
PCR	Polymerase Chain Reaction
pE2	Precursor E2 Protein
PEG-IFN	Pegylated Interferone
PEI	Phosphate buffered saline
PKR	Protein Kinase R
PNGase F	Peptide-N-Glycosidase F
PRR	Pattern Recognition Receptor
prM	Precursor Membrane protein
RdRp	RNA-dependent RNA polymerase
Rev	Regulator of expression of viral proteins
RIG-I	Retinoic acid- Inducible Gene-I
RLR	Retinoic acid- inducible gene-I Like Receptor
RRE	Rev-Responsive Element
SCID	Severe Combined Immunodeficiency
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
sE2	Soluble envelope glycoprotein 2
SNP	Short Nucleotide Polymorphism
SR-BI	Scavenger receptor class B type I
SOC	Standard of Care
SVR	Sustained Virological Response
TAE	Tetra acetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
TLR	Toll Like Receptor
TM	Transmembrane domain
UTR	Un-translated region
UVP	UV trans-illuminator
VLDL	Very Low Density Lipoprotein

1. Introduction

1.1. Hepatitis C Virus as a pathogen

1.1.1. Discovery and classification

For many years viral hepatitis was believed to be caused by only two viruses: Hepatitis A Virus (HAV) and Hepatitis B Virus (HBV). Despite establishing robust screening for these agents, post-transfusion hepatitis continued to occur. Such cases were described as Non-A-Non-B (NANB) hepatitis and the unknown agent proved difficult to identify [1, 2]. In 1989, a group of scientists from the Chiron Corporation succeeded in isolating cDNA from the sera of a chimp infected with NANB, and identified this as a unique infectious agent which they called Hepatitis C Virus [3]. This virus was confirmed as the causative agent of NANB hepatitis. Sequence analysis showed that the genome structure was similar to members of the *Flaviviridae* family, but was distinct enough to place HCV into a new genus called *Hepacivirus* [4, 5].

1.1.2. Epidemiology and worldwide distribution

The HCV genome exhibits extensive genetic variation, with at least 6 genotypes and over 70 subtypes having been described [6, 7]. Each genotype differs from other genotypes in 31 to 33% at nucleotide level [8, 9]. HCV genotypes exhibit regional differences in distribution [7]. The high levels of genetic variability apparent within genotypes in sub-Saharan Africa and South-East Asia indicate that HCV is endemic to these regions. Introduction of HCV to non-tropical countries occurred with recent

development of new risk groups, such as blood recipients, vaccinees and drug abusers [10, 11]. HCV is mainly transmitted via contaminated blood and/or blood products. HCV infection occurs following large or repeated percutaneous blood contact (i.e. drug injections, transfusion or transplantation). Viral transmission is less efficient when single exposures to small amounts of blood are concerned [10-15].

The HCV has been divided into 6 genotypes and further subtypes [7] which show specific geographical distribution shown in Figure 1.1. Transmission of specific subtypes, namely 1a, 1b and 3a, by blood transfusion and needle sharing between intravenous drug users (IVDUs) has resulted in the exponential spread of these subtypes [16]. Indeed, these three HCV subtypes represent majority of HCV infections in Europe and the United States. Despite increasing knowledge about HCV molecular evolution and epidemiology, the origins of infection in humans remain unclear. However recently a canine homolog of HCV has been characterised opening new possibilities for understanding worldwide HCV outbreak [17]. More than one million new cases of HCV are reported annually, and reports estimating the number of people infected with HCV varies between 130 to 170 million [9]. These numbers correspond to a 2.2% prevalence of HCV positive persons worldwide. Figure 1.2 presents geographical differences in HCV prevalence. The United Kingdom is amongst countries with the lowest HCV prevalence.[18]. An estimated 250,000 - 600,000 (about 0.5%) individuals in UK carry HCV [19]. The HCV prevalence in Europe varies from 0.4% in Germany, Netherlands and Sweden to 2–3% in some Mediterranean countries [20]. To control a parasitic disease common in Egypt, schistosomiasis, a mass eradication programme was conducted from the 1960s until the 1980s. Unfortunately, non-sterile needles were used throughout the campaign resulting in Egypt having the highest global HCV prevalence of around 20% [21]. Many developing countries still lack any

HCV prevalence data. On average, a quarter of cirrhosis and hepatocellular carcinoma cases worldwide result from HCV infection [9].

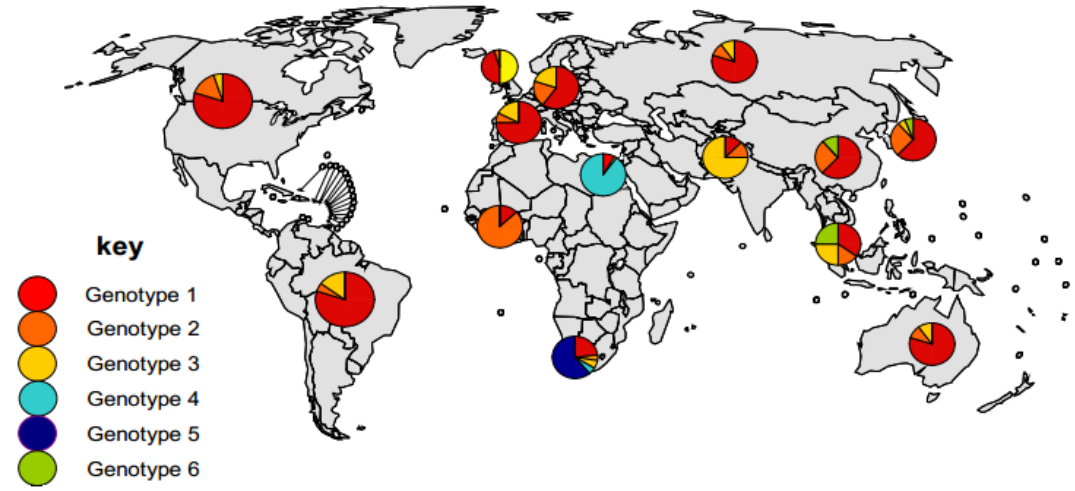


Figure 1.1. Worldwide distribution of HCV genotypes [22].

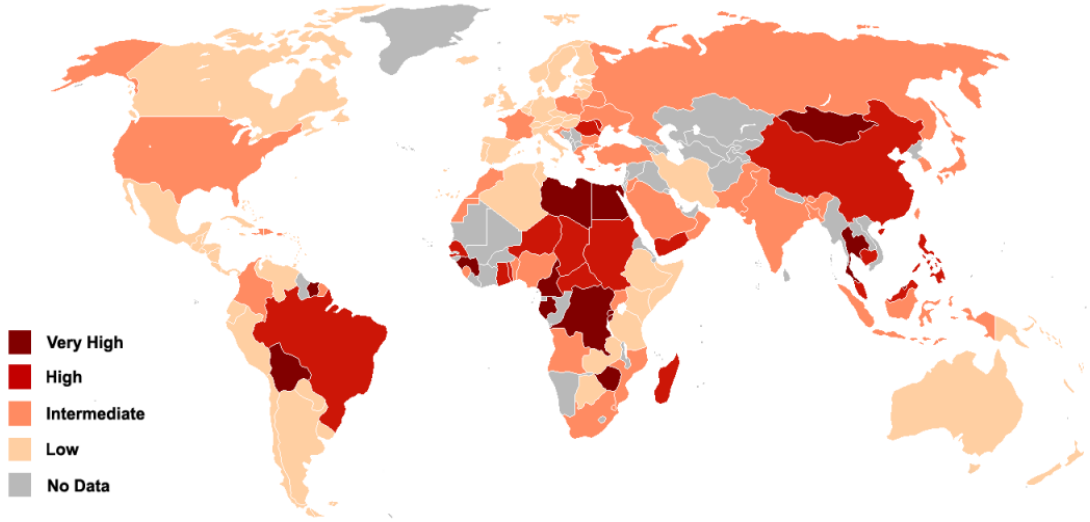


Figure 1.2. Worldwide prevalence of Hepatitis C Virus [23].

1.1.3. Diagnosis

HCV infection is determined by assaying for the presence of anti-HCV antibodies and viral RNA. These detection methods are based on recombinant HCV peptides in the context of enzyme immunoassays (EIA). Third generation assays are now available and

in routine use [24]. . These assays contain a mixture of HCV epitopes from Core, NS3, NS4 and NS5 regions, which detect a broader spectrum of anti-HCV antibodies as compared to previous generation methods and allows estimated specificity up to 99% [24].

Specific anti- HCV antibodies occur upon seroconversion that follows a window period of infection that can last up to six months post infection. The HCV diagnosis by means of antibody testing is thus unreliable for recognition of early phase of infection [25], hence HCV-RNA diagnostics are essential in identifying individuals during the pre-seroconversion window period [4, 5, 26, 27]. The presence of viral RNA and thus diagnosis of HCV replication can be detected and quantified with use of PCR techniques. All commercially available tests implement IU/ml of HCV RNA as a quantitative unit [24, 28]. Automated platforms, which are currently used for HCV RNA detection, are based on real-time PCR technique. Real-time PCR as compared to standard PCR technique is more sensitive and has a broader dynamic range of detection and quantification of viral nucleic acids. As low as 10-15 IU/ml can be detected and quantification can be carried out up to a range of 7-8 $\log_{10}$ IU/ml [28]. Detection and quantification of viral RNA is not only crucial as a diagnostic method but serves as means to monitor the virological response to therapy.

1.1.4. Clinical manifestations

Acute HCV infection is difficult to recognize as symptoms like hyperbilirubinemia are apparent in less than 20% cases [29]. The acute phase of infection is characterized by elevated levels of liver enzymes- alanine aminotransferase (ALT) and asparagine aminotransferase (AspAT) [25]. Raised liver enzymes in sera are due to developing necrosis of hepatocytes and rising cell membrane permeability in the target organ. The

phase of raised AAT can be very short and unnoticed. It is preceded by a point when virus reaches the highest titre. Levels of viremia do not usually correlate with severity of disease or prognosis [30].

The acute phase in 75- 85% of cases develops into chronic hepatitis which is marked by persistence of HCV RNA despite arising of anti- HCV antibodies [25]. The most common histopathologic symptom of HCV is liver fibrosis. Fibrotic changes seem to develop during chronic hepatitis, when continuous destruction of infected cells and mass production of cytokines and growth factors occurs. A prolonged state of developing fibrosis leads to bile duct damage which results in cirrhosis. The final complication in HCV infection is cirrhosis, which most often leads to liver decompensation and the need for liver transplantation [31, 32]. Natural history of HCV infection is summarised in Figure 1.3.

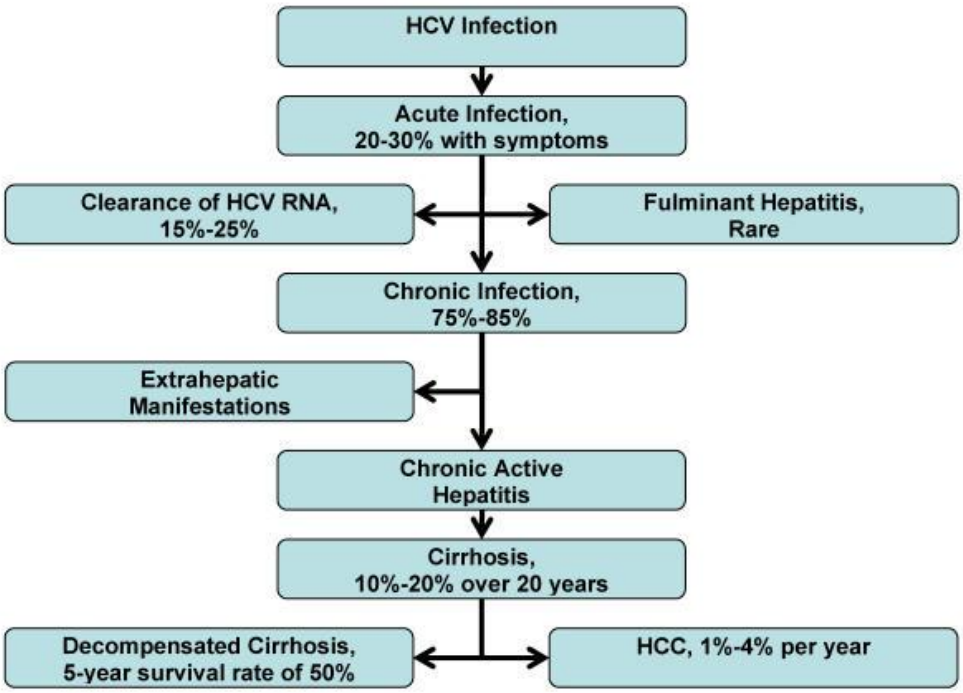


Figure 1.3. Natural history of HCV [32].

1.2. Molecular Insight

1.2.1. Genome and encoded proteins

The HCV genome consists of a single open reading frame (ORF) of around 9000 nucleotides, flanked by 5' and 3' non-coding regions. The 5'UTR (Un-Translated Region), is 341 nt long and possesses an internal ribosome entry site (IRES), which has the ability to start cap-independent translation [33]. The 3'UTR of HCV consists of a poorly conserved stretch of nucleotides followed by polypyrimidine tract and a unique 98 nucleotide sequence. Finally, there is a 3'X tail, which folds into a stem loop structure and is important for the initiation of replication [4, 34]. Model of the HCV RNA structure is presented in Figure 1.4.

The ORF of HCV encodes a single 3000 amino acid polyprotein precursor that is cleaved into at least 10 mature proteins [34, 35]. 5' end of the ORF is translated into structural proteins followed by non-structural proteins at the 3' end. The core protein is located at the N-terminus of the precursor protein and is released by host signal peptidase [5]. This protein is highly basic and is able to form multimeric complexes which eventually form the viral capsid. Core associates with lipid droplets and binds to the 5'UTR of HCV genome, both of which may be essential for virus morphogenesis [36]. It has been shown that due to ribosome frame shifting in the core region, an additional gene product, called the F protein, might be produced, although its function is yet unknown [37].

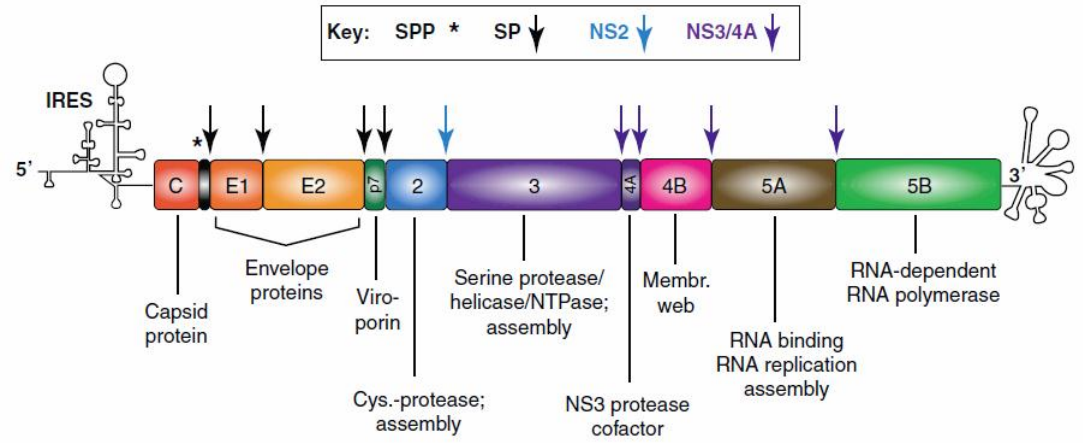


Figure 1.4. HCV genome and polyprotein structure.
Points of protease cleavage are indicated by asterisks. SPP- signal peptide peptidase, SP- signal peptidase [38].

The envelope glycoproteins, E1 and E2, are located downstream of the core protein and are also released via signal peptidase action. These are heavily glycosylated proteins of about 35 and 70 kDa respectively [39], and their functions, including their role in cellular entry and immune escape will be described in detail further.

Encoded next, in the HCV genome, is p7- a small hydrophobic polypeptide that was shown to be essential for the production of infectious virions in vivo [40]. It forms an ion channel in lipid bilayers consistent with its classification into viroporins [41].

Downstream of the structural protein region are the genes encoding non-structural proteins, which are responsible for viral replication and protein release. Two proteinases are coded in the non-structural region: NS2/3 metalloproteinase and NS3 serine proteinase [34]. NS2 is a protein rich in hydrophobic residues and undergoes autocatalytic *cis* cleavage from its precursor protein [42]. NS3 is a protein that contains characteristic motifs for serine proteases (in N-terminal part) and for RNA helicases and nucleoside phosphatases (in C-terminal part) [42, 43]. The protease consists of a catalytic subunit of NS3 activated by NS4A cofactor. The proteolytic activity of this heterodimeric protease is targeted to releasing non-structural proteins from the

polyprotein precursor [42]. The section of NS3 not engaged in polyprotein cleavage forms a helicase domain [43] which is essential in HCV replication.

Located downstream is a hydrophobic NS4B protein for which function was not immediately ascribed. It has been localised on ER membranes and lipid rafts, and further indicated in formation of membranous web, interactions with cellular factors and scaffolding RNA replication [44].

The next protein coded within the ORF, NS5A, is present as two differently phosphorylated proteins: 56kDa and 58kDa [45]. The latter is hyper phosphorylated form of p56. NS5A has been shown to interact with IFN- induced protein kinase (PKR) [46]. The IFN sensitivity-determining region (ISDR) has been mapped in C-terminal part of NS5A which is also responsible for the interaction with protein kinase PKR [47]. The NS5A has been indicated in virus assembly and RNA replication [45].

The sequence of the last protein released from the polyprotein precursor is highly conserved between HCV strains. NS5B is an RNA-dependent RNA polymerase (RdRp), which synthesizes RNA using an RNA template [48]. Crystallisation revealed a 'right hand' structure with palm, thumb and fingers subdomains [49]. The active site of this enzyme is located within the palm, which is fully encircled due to thumb and fingers subdomains creating a tunnel. The latter domains are thought to possess allosteric sites modulating the enzyme interaction with RNA [49, 50].

All HCV proteins have amino acid sequences that localise them to ER membrane, providing evidence that the ER is the site of virus replication and assembly (shown in Figure 1.5).

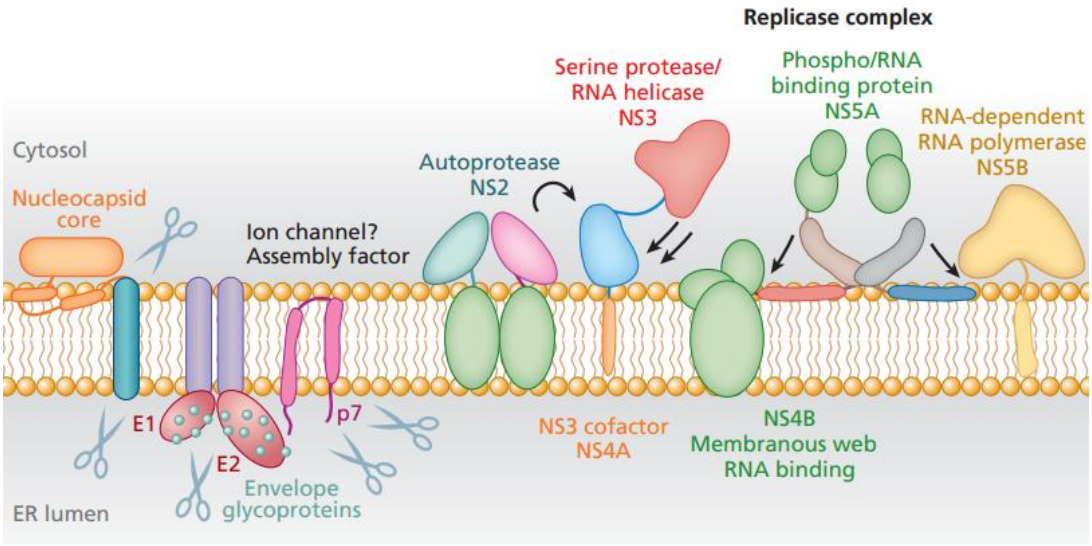


Figure 1.5. ER retention of HCV proteins and their indicated functions [51].

1.2.2. Envelope glycoproteins

HCV viral particle contains E1 and E2 glycoproteins in its bilayer envelope. These surface proteins are essential for virus life cycle as they mediate entry into the host cell and have also been shown to possess epitopes targeted by host immune system.

1.2.2.1. Biogenesis and glycosylation

HCV glycoproteins E1 and E2 are type I transmembrane proteins with N-terminal ectodomains and C-terminal membrane anchors [52]. E1 and E2 are cleaved from the HCV polyprotein as soon as translation of their coding region occurs. Cleavage site between E2 and p7 seems to be recognised less efficiently by signal peptidase giving rise to different pools of E2: fully cleaved E2 and E2p7 [53, 54]. Folding of E1 is a relatively slow process which is dependent on E2 [55, 56]. Cysteine cross-linking that supports the conformation in E2 occurs in a time sufficient for its cleavage from the rest of the precursor protein [57]. Both E1 and E2 possess ER retention signals and undergo maturation in this cellular organelle. The C-termini of E1 and E2 are

responsible for translocation of glycoprotein ectodomains into the ER lumen [55]. HCV glycoproteins associate in ER with chaperones like calreticulin, Binding Immunoglobulin Protein (BiP) and calnexin [58, 59]. Trafficking of E1 and E2 between these chaperone proteins depends on proper processing of N-linked glycans that, among other roles, has an impact on folding of these proteins. Calreticulin and BiP interact preferentially with misfolded aggregates, whereas calnexin associates with properly folded heterodimers [58, 59].

Domains responsible for heterodimerisation of HCV glycoproteins have not been characterized accurately. Initially, that process seemed to depend only on ectodomains [60, 61]. However, experiments on HCV glycoproteins with TM domains deleted showed a tendency for aggregation. Changing of TM domains of E1E2 for TM domains of other proteins disturbed folding of functional heterodimers. Moreover, the N-terminal part of transmembrane E1 domain possesses highly conserved motif GXXXG that enhances interaction of hydrophobic parts of TM proteins [60, 61]. Together these data suggest that important determinants of heterodimerisation reside in the TM domains.

Inconsistent with most Flaviviruses, HCV envelope proteins undergo extensive glycosylation as part of the maturation processes [62]. Mapping of glycosylation sites showed that E1 has up to 6 and E2 - 11 potential N-linked glycosylation sites [63-65] (see Figure 1.6). Glycans present on HCV envelope proteins nearly double their mass. Digested by Peptide-N-Glycosylase F (PNGase F) E1 and E2 are about 37 and 17 kDa, whereas glycosylated proteins migrate at about 65 and 30 kDa, respectively [66]. The characterisation of HCV glycoproteins has shown that glycosylation patterns differ between various truncated and full length proteins indicating their different properties. Absence of E2 or E1, during expression of the other glycoprotein, leads to production of different glycoforms [62-64, 67, 68]. Analyses with conformation- dependent

antibodies show that folding of HCV glycoproteins is a slow process. In addition it has also been shown that although some degree of folding can be observed in E2 expressed alone, both glycoproteins need to be co-expressed to acquire functional properties [57, 59, 69].

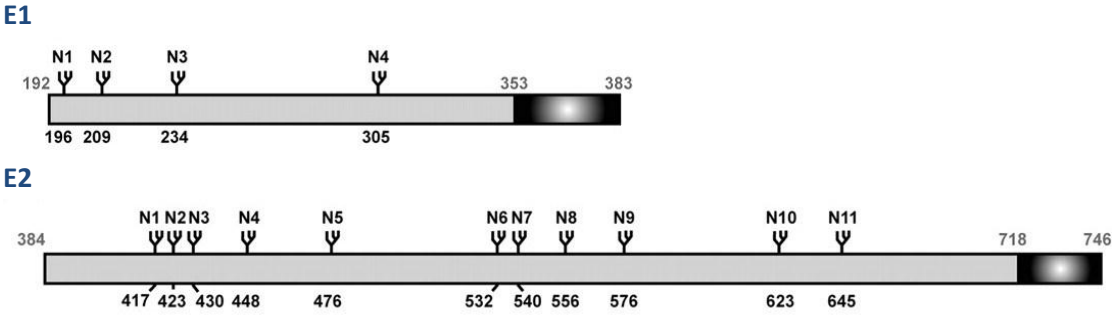


Figure 1.6. Schematic presentation of glycosylation sites on linear E1 and E2 [70].
N indicates N-linked glycan positions numbered according to H strain. Black- trans membrane domains, grey- ectodomains.

Functions of the E1E2 glycans in the HCV life cycle have been studied [65, 71]. Site directed mutagenesis revealed that glycans E1N1, E2N8, E2N10, and E2N11 are essential for correct particle assembly [65]. Glycans E2N1, E2N2, E2N4, E2N6, and E2N11 shield neutralisation sensitive epitopes from immunological targeting [65]. Finally glycans E2N1, E2N2, E2N4, and E2N6 were shown to have a role in E2 interaction with CD81 [65].

A previously mentioned study [66] describes detailed characterization of glycosylation patterns of intracellular, HCVpp- and virion-associated E1 and E2. Intracellular E1 and the majority of E2 glycosylation sites harbour high- mannose- type glycans in both- HCVpp and HCVcc expression systems [66]. Whereas glycans on E1 remain unmodified upon secretion, E2 secreted in the context of HCVcc authentic virions harbour a mixture of high- mannose and complex- type glycans [66]. This observation is not consistent with HCV pseudotypes containing E2 with majority complex- type glycans [66].

All described studies highlight that the HCV glycoproteins maturation depends on a complex and dynamic relationship between cleavage by signal peptidases, close proximity of both E1 and E2, their translocation, folding and glycosylation.

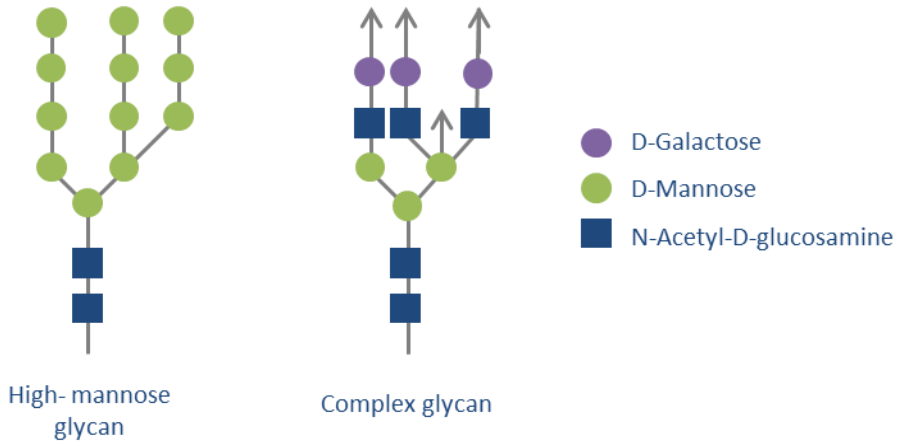


Figure 1.7. Schematic representation of high-mannose and complex N-linked glycan.

1.2.2.2. Structure and function

The current hypothesis assumes that virions acquire lipid envelope from host cell membranes- either external cell membranes or internal membranes of ER and Golgi structures. Immunolocalisation and analysis of E1E2 glycan structure suggests that they are connected to the ER membrane and most probably undergo glycosylation while virus budding [54, 72]. Functional E1E2 heterodimers, present on HCV envelopes, are responsible for virus-cell contact and further membrane fusion [73]. E1, which spans residues 171-383 of the HCV polyprotein, has been suggested as a fusion protein but its exact role in virus entry is not precisely described. E2 glycoprotein, encoded within residues 384- 746 of the HCV polyprotein is much better characterised and is now known to have receptor binding and fusogenic properties as described below. The entry process and its cellular cofactors are further described in paragraph 1.3.1.

Fusion proteins in *Flaviviridae* family have been characterised as class II fusion proteins (reviewed in [74]) and noted to be co-expressed with a second membrane protein. E1 is thought to exist in a noncovalent complex with E2 in a similar manner to flavivirus precursor M protein (prM) and alphavirus precurson E2 protein (pE2) [74]. The usual three domain structure of class II fusion proteins has notable amounts of β-sheet form. Domain II of the protein contains the fusion loop, flanked by domain I at carboxy, and domain III at amino terminus. Such architecture was thought to be true for E2 ectodomain as well [74]. The model of flavivirus fusion protein and its mode of action are presented in Figure 1.8 and described in the figure legend.

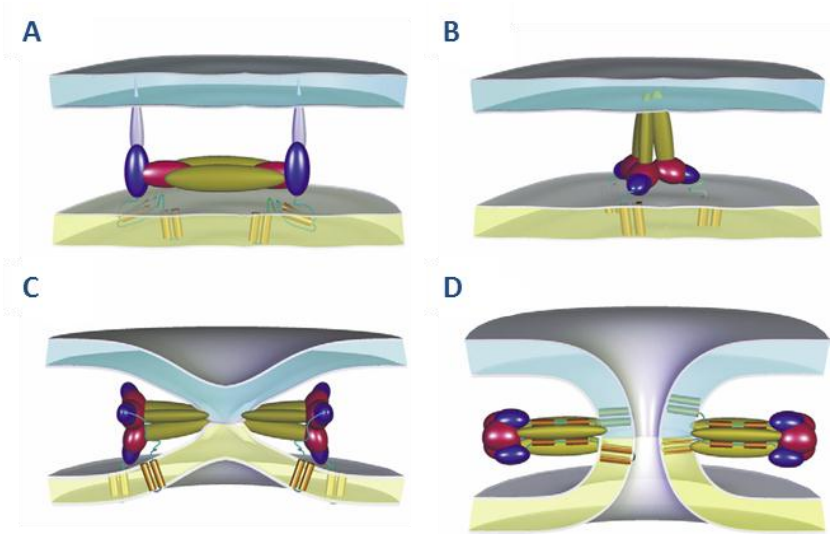


Figure 1.8. Model of flavivirus class II fusion protein triggering membrane fusion [74]. This model has been proposed to be true for HCV E2 as well. Domains are as in figure above: Domain I- red, Domain II- yellow, Domain III- blue. Upper membrane- cell surface, lower membrane- viral envelope. Ectodomain of viral fusion protein binds to cellular receptor and virus is internalised (A). A pH change in the endosomal compartment triggers conformational change, attachment of domain II to host membrane, and trimerisation of ectodomains (B). Further rotation of domain II places the transmembrane domain in close proximity to fusion loop (C), which in effect should lead to membrane fusion (D) [74].

High glycosylation and ER retention of HCV glycoproteins contradict their classification as class II fusion proteins. Due to high level of glycosylation, presence of trasmembrane domains and tendency to misfolding E1 and E2 have been difficult to express at high enough levels to allow crystalisation [57] and deciphering their structure. Krey et al.

[75] designed an effective production system to obtain large amounts of high purity E2 ectodomain that retains functionality and is thought to present a native conformation. In spite of unsuccessful crystallisation efforts, basing on a conserved manner of disulphide bonds across E2 and other structural data the group revealed the tertiary structure presented in Figure 1.9. This model confirms the hypothesised three domain structure in agreement with class II fusion protein model. Domain I consists of eight antiparallel β-sheets (B_0-I_0) and was mapped to possess most of the CD81 binding epitopes. Strands D_0 and E_0 are separated by a long amino acid sequence which in tertiary structure takes form of Domain II. Domain III is at C terminus of the E2 receptor binding domain and is followed by a stem region.

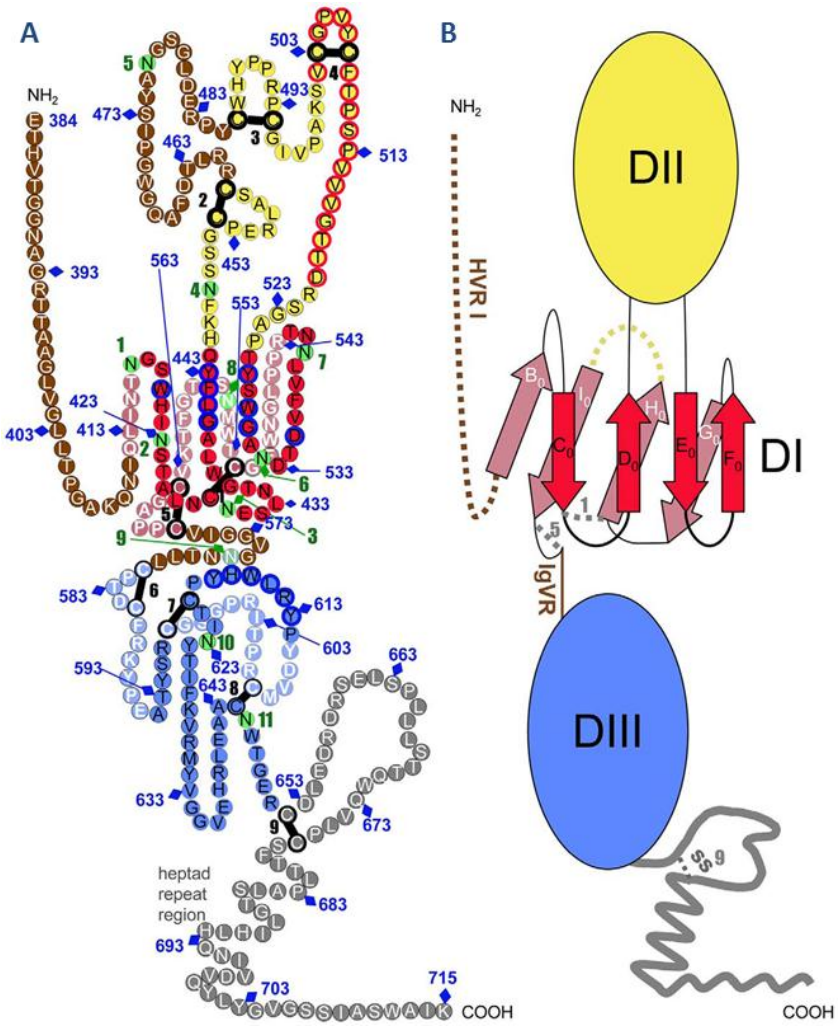


Figure 1.9. Tertiary structure of E2 ectodomain [75].
A- Beads represent sequential amino acids, B- simplistic representation of E2 domains parallel to the structure in panel A. Black bars represent cysteine bridges. Green beads represent glycosylation sites. Beads contoured in blue- amino acids of CD81 binding epitopes, contoured in red- putative fusion loop. Unstructured regions- brown beads.

HCV glycoproteins show high extent of variability with several hot spots termed hyper variable regions (HVR). Hypervariable Region 1 consists of 27 amino acids at N terminus of Domain I. HVR2 was mapped to 15 amino acids of Domain II and intergenotypic variable region (IgVR)- 10 amino acids in a region between domains I and III [76]. Functions of hypervariable regions have not yet been fully ascertained, although HVR1 has been proposed as an immunological decoy [77] and these regions have been shown to have a role in glycoprotein formation and viral infectivity [78]. HVR1 is known to engage in SR-BI receptor [79] and antibody binding. The variability

of HVR1 sequences in acute infection have been shown to be predictive of disease outcome[80]. Other regions across the E2 ectodomain have been indicated in antibody binding and T cell recognition (described later).

1.2.3. Virus life cycle

The HCV entry into host cells (fully described in section 1.3.1) occurs via attachment to set of co-receptors, endocytosis and pH dependent membrane fusion. Capsid released from endosomal body is unpacked and genome RNA is released into the cytosol [5, 81]. The HCV RNA of positive polarity serves as mRNA and its translation is dependent on Internal Ribosome Entry Site (IRES) which directly interacts with 40S ribosomal subunit [33]. Translation occurs on cytoplasmic side of rough Endoplasmic Reticulum (rER). HCV proteins, released from the polyprotein by host and viral proteases, form a membranous web anchored in ER (see Figure 1.5) where HCV genomes are amplified and further encapsidated into new viral particles [81]. The HCV capsids of about 50nm consist of multimeric form of Core protein and viral RNA. They are enveloped in lipid bilayer harbouring glycoproteins E1 and E2 [5, 38]. The mature virus particle has been described to be between 50 and 80 nm in size. Other proteins: p7, NS2 and NS4A have been reported to take part in the assembly process (reviewed in [38]). Viral assembly and release are thought to occur in tight association with lipids. Several host factors related to VLDL metabolism have been indicated in the process which supports the hypothesis of HCV existing as a Lipo-Viral Particle [38, 82]. Schematic of HCV life cycle is presented in Figure 1.10.

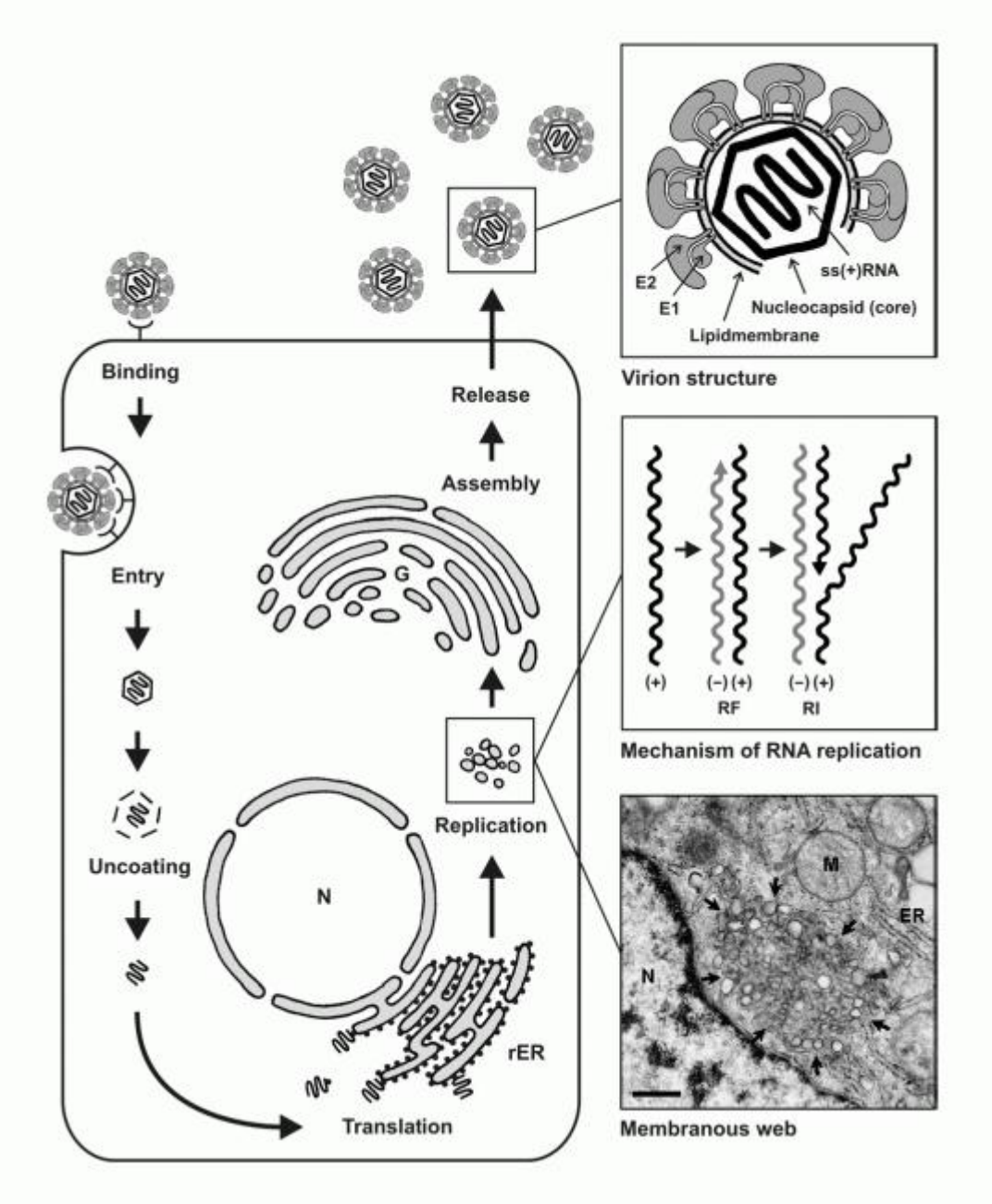


Figure 1.10. Hypothetical Hepatitis C Virus lifecycle [83].
HCV virions bind to the host cell and are internalised via interaction with a set of coreceptors (described further in text). After uncoating of the genome translation occurs on membranes of rough ER and HCV proteins induce formation of membranous web (bottom inset) which is thought to be a scaffold for replication machinery. The middle inset depicts model of HCV genome replication with double stranded replicative form (RF) and replicative intermediate (RI). Progeny virions are assembled on ER membranes in connection with lipid droplets (not shown here and described in further chapters). Model of HCV virion is presented in top right inset.

1.3. Host factors and their interplay with HCV

1.3.1. Cellular receptors and HCV entry

Since the discovery of HCV, several cell surface molecules have been proposed as receptor candidates. The first step in viral entry is attachment to host cell membrane followed by endocytosis and then pH- dependent fusion of membranes. It has been proposed that entry events occur in detailed sequence and that cell surface molecules interact with HCV glycoproteins at different stages of the entry process (presented inFigure 1.11).

1.3.1.1. Molecules involved in attachment

Initial attachment of HCV Viro-Lipo particles, viral particles made of core protein and envelope glycoproteins E1 and E2 in association with lipoproteins, has been reported to engage factors like lectins [52, 84, 85], lipoprotein receptor [86, 87] and glycosaminoglycans [88].

Classical C-type lectins act as adhesion molecules and pathogen recognition receptors that uptake, internalise and lead to elimination of pathogens. Some viruses have been shown to evade degradation and use C-type lectins as entry receptors. Both L-SIGN and DC-SIGN have been shown to capture different forms of HCV glycoproteins and transport them to neighbouring permissive cells [52, 84, 85]. Their ectopic expression does not sustain permissiveness so it has been proposed that they contribute in virus capture and delivery to hepatocytes [52, 84, 85].

The low-density lipoprotein receptor plays a role in lipid uptake and metabolism, and has been indicated in cell attachment for a number of viruses [86]. It preferentially binds to low density lipoproteins (LDL) and very low density lipoproteins (VLDL) which were shown in complex with HCV [38]. Indeed data suggests that attachment of HCV to host cell involves low- density lipoprotein receptor as one of the factors [86, 87, 89].

Glycosaminoglycan [90] chains, although present on surface of many cells, differ in their composition between various cell types [91]. Heparan sulphate has been indicated as primary docking structure for attachment of viruses including a member of the *Flaviviridae* - Dengue virus [92]. Structural analyses of the glycoprotein E2 derived from various HCV strains reported conservatism of positively charged amino acids in HVR1 necessary for interaction with GAGs, and were followed by demonstration of the E2 interaction with heparan sulphate [88, 93].

1.3.1.2. CD81

CD81 was identified as a first putative HCV receptor by using soluble E2 (sE2) as a bait for possible interaction with surface molecules [94]. CD81 belongs to the tetraspanin family. It contains four transmembrane domains spaced with short intracellular loops and two extracellular loops- small and large (SEL and LEL). The latter one contains specific binding regions for E2 [95]

Initial data on the CD81- HCV interaction had to be treated cautiously as it was mostly obtained in studies with soluble E2. This truncated form of E2 has structural differences in comparison to the full-length E1E2 heterodimer [96]. It was proposed that E1 affects structure and properties of E2 which suggest that results obtained in studies with sE2 might not truly reflect the HCV-CD81 interaction [96]. Studies

performed with HCVpp (HCV pseudo particles) and HCVcc (cell culture HCV) showed that anti-CD81 antibodies and recombinant soluble CD81-LEL inhibit infection of hepatoma cell lines [97, 98]. HepG2 cells, which do not express CD81, are not permissive for HCVpp and HCVcc. Ectopic expression of CD81 in the HepG2 line makes it susceptible to HCV infection providing additional proof of CD81 engagement in virus entry to cells [97, 99]. Two out of four disulfides in CD81-LEL are required for interaction with E2 [100]**.** The predominant site of replication and infection by HCV is liver and indeed HCVpp show tropism limited to liver cells expressing human CD81 [101]. A tetraspanin web- in which members of tetraspanin family interact with non tetraspanin molecules, is cell-type specific regarding its composition [102]. Moreover Bitzegeio et al. determined species specific determinants residing in CD81 and using these adapted mouse CD81 for HCV entry [103]

Receptor competition studies suggest that CD81 plays a role after viral attachment [84, 104]. Although tetraspanin members are known to play part in membrane fusion, CD81 does not seem to mediate endocytosis as it undergoes poor trafficking from cell surface to vacuolar web [100].

1.3.1.3. SR-BI

SR-BI is a membrane protein with a large extracellular domain separating two membrane-spanning domains and short terminal internal domains. It is involved in metabolism of lipids acting as a lipoprotein receptor and has been found to interact with HCV [79]. As in the case of CD81, SR-BI was initially characterised by binding with sE2. That interaction has been shown to be selective, as neither closely related human scavenger receptor CD36, nor mouse one was able to bind sE2 [79]. SR-BI binding is dependent on presence of hypervariable region 1 (HVR1) at the N-terminus of sE2 and

can be blocked by anti-HVR1 antibodies [79, 97]. However, interaction of this molecule with the E1E2 heterodimer has not yet been demonstrated, studies with HCVpp are in agreement with SR-BI engagement in HCV entry [97, 105, 106]. Lipoproteins were indicated to mediate HCVpp entry via SR-BI [106], however different regions of SR-B1 are responsible for HCV and HDL recognision [105] indicating that initial lipid mediated binding could be followed by E2-SR-BI interaction. Experiments performed with use of HCVcc showed that SR-BI is engaged in a post attachment step closely linked with CD81 [107].

1.3.1.4. Tight junction proteins

Claudin-1 (CLDN-1) is a tight junction protein that spans plasma membrane four times. It has been recently described as HCV entry coreceptor [108]. Evans et al. showed that this receptor is crucial for HCV infection of hepatoma cell lines and confers susceptibility when ectopically expressed on cell lines non-permissive to HCV [108]. Recent study shows that cell type specific claudin localisation might influence cellular tropism of HCV [109]. Residues within the first extracellular loop of claudin-1 were mapped as critical for HCV entry [110]. Alanine substitutions in these highly conserved motifs impaired claudin-1 cell-to-cell contact, which further supports importance of claudin localisation and/or function in tight junctions for HCV entry process [110]. Additional members of the family: claudin-6 and claudin-9 have been identified as HCV entry coreceptors [111]. Further study localised claudin-1, in association with CD81, at the basolateral surface of hepatoma cells [112] and demonstrated role of these complexes in HCV infection that would be consistent with sinusoidal vessels as route of infection.

SR-BI and CLDN-1 were attributed to have influence on species specificity of HCV. However, until another tight junction protein, occludin (OCLN), has been demonstrated in context of HCV entry process [113], only CD-81 was proved to be responsible for restricting the infection to human and chimpanzee cells [103]. OCLN expression on human cells was demonstrated as a requirement for human cells susceptibility to HCV [114]. Overexpression of OCLN on non-permissive cells renders them susceptible to HCVpp infection and silencing of the receptor in permissive cells hampers HCVpp and HCVcc entry [114]. Similarly to CD81, OCLN has to be of human origin to support efficient entry, and determinants of species specificity are located in its second extracellular loop [115].

1.3.1.5. pH mediated endocytosis

In summary the HCV entry is postulated to be primed by attachment of viral particles in complex with lipoproteins to GAG and/or LDLr. This step is thought to be followed by interaction with CD81 and SR-BI. Late entry steps include binding to tight junction proteins. Then receptor- virus complexes undergo internalisation via clathrin mediated endocytosis [73, 116]. Upon delivery of clathrin-coated vesicles to endosomes, pH change triggers membrane fusion and lipoproteins and particles bound to them are released [117]. Fusion domains have not yet been exactly mapped, however envelope glycoprotein heterodimers are thought to mediate the membrane fusion (described in 1.2.2.2) in a fashion shared by *Flaviviridae* members [74].

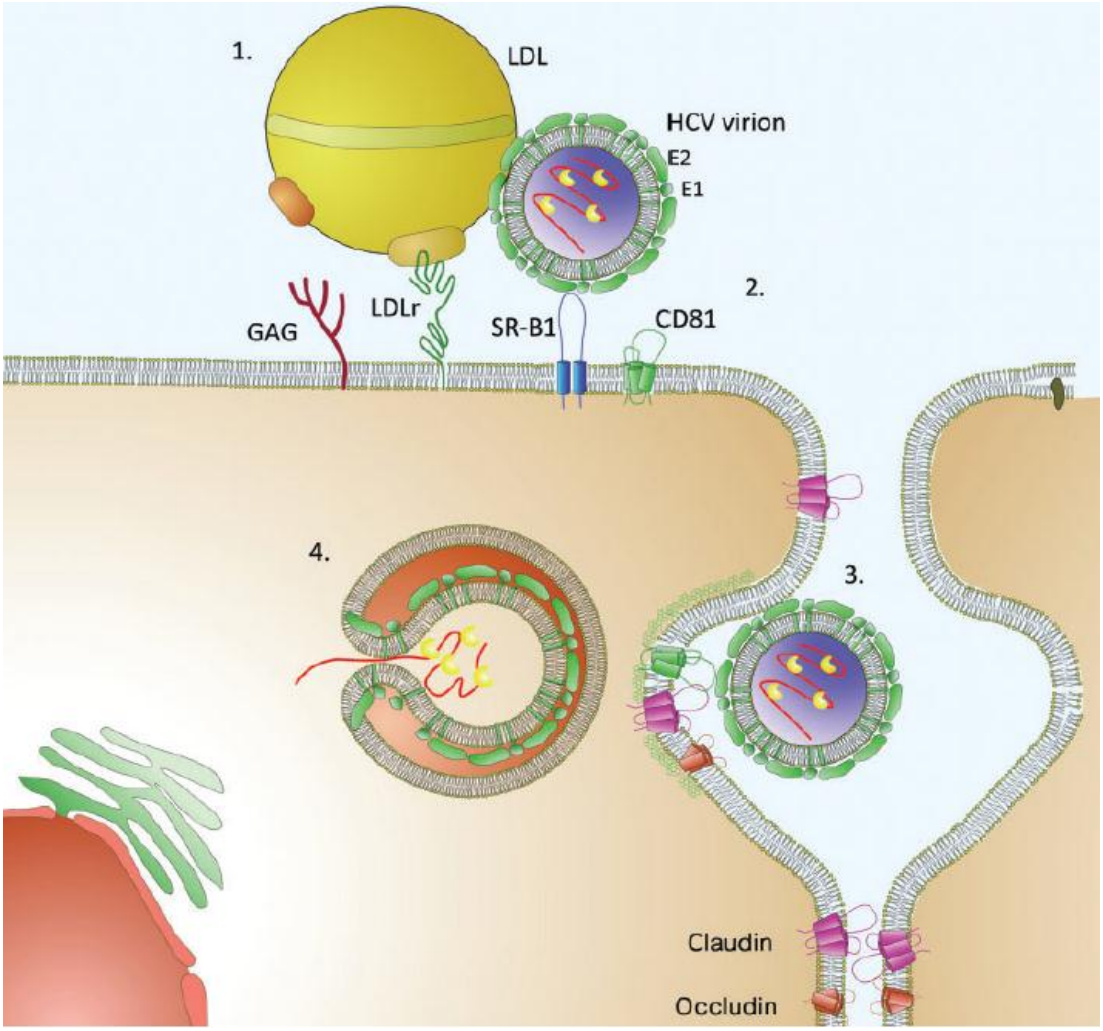


Figure 1.11. Schematic representation of current hypothesis on HCV entry and endocytosis via interplay of envelope glycoproteins and lipoproteins with a set of cellular receptors [118].
Individual steps marked in this figure are described in text (1-attachment, 2- binding to CD81 and SR-B1, 3- interaction with CLDN and OCLN, 4- endocytosis and membrane fusion).

1.3.2. Host factors and immune response

HCV infection triggers complex antiviral responses and the interplay of viral and host factors results in either virus eradication or persistence. In recent years more light has been shed on particular steps of this interplay; however there is much to be discovered before we understand the mutlifactorial host and viral determinants that underlie HCV infection.

1.3.2.1. Innate immune response

The HCV firstly affects signalling pathways leading to production of type I interferon and interferon stimulated genes (ISGs). HCV appears to be able to down regulate these signalling pathways and block their interfering activity [9]. Although much has been learned, the full mechanism by which HCV establishes persistence is not yet known (further described in section 1.5.4).

Innate immune responses are transiently activated by high level of HCV RNA [119]. Initiation occurs through pathogen- associated molecular patterns (PAMP) that are recognised by 3 major classes of pattern recognition receptors (PPRs) [120]. These are: toll like receptors (TLRs), nucleotide oligomerisation domain-like receptors (NLRs) and retinoic acid-inducible gene-I (RIG-I) like receptors (RLRs). Their function is partially regulated by their cellular localisation. RLRs and NLRs are present in cytoplasm of most cells whereas TLRs are located on endosomes and cell surface.

NRLs are thought to be majorly involved in recognition of bacterial PAMPS and their role in antiviral response is not well described [121]. TLRs identified to take part in antiviral detection include TLR 2, 3, 4, 7/8 and 9. Their activation effects in production of proinflammatory cytokines and IFN-α (reviewed in [122]). Main target cells for the HCV- hepatocytes are deficient in TLR signalling [123]. RLRs are expressed in vast majority of cells (including hepatocytes) which strongly suggests their role as primary cytoplasmic detectors of the HCV infection [124]. RLR family consists of RIG-I (retinoic acid- inducible gene- I), MDA5 (melanoma differentiation- associated gene 5) and LPG2 (laboratory of genetics and physiology 2) [124]. These receptors have a RNA helicase domain binding RNA ligands and 2 N-terminal caspase activation and recruitment domains [125]. RIG- I has been shown to detect specific set of ssRNA viruses including flaviviruses [124, 126].

Interferon regulatory factor 3 (IRF3) is a transcription factor expressed in cytoplasm in an inactive form. Transfection of HCV RNA into human hepatoma cell lines triggers activation of IRF3 by C-terminal phosphorylation, followed by its dimerization and transport to nucleus. There transcriptional activation of several ISGs occurs [127]. RIG-I deficient cells do not show such effect upon transfection, but it can be restored with functional RIG- I [128]. This suggests that RIG- I is necessary part of pathway leading to IRF3 activation by HCV and might be the primary intracellular sensor of the virus. Another factor activated by HCV transfection is IFN- β [127]. IFN-β transmits signals to neighbouring cells to induce ISGs as well. These ISGs include some of anti-HCV nature-like protein kinase R (PKR) and ISG56 that prevent translation of viral proteins [129]. RIG- I recognises short dsRNA, 5'triphosphate (ppp) within ssRNA, and viral RNA motifs that are rich in uridine or adenosine [130]. As HCV RNA is not capped it possesses 5'ppp motif. The highly conserved across all HCV genotypes 3'NCR containing poly U region has been recognised as the primary HCV PAMP detected by RIG- I [127, 128]. Different HCV strains have been tested and unsurprisingly genotype 2a strain JFH1 proved to be a weak RIG- I activator, which could in part explain this strain's robustness [130]. It is not yet clear at which point of HCV replication, and in which exact cellular location, the activation of RIG-I happens but it proves to be the key element in innate intracellular anti-HCV response. HCV has been shown to evade interferon response by subverting RIG-I pathway by NS3/4A protease cleavage of specific adaptor proteins [131-133]**.** In the early hours post HCV infection liver cells relocalise IFN-3 to nucleus. NS3/4A cleaves interferone promoter stimulator-1 (IPS-1) and prevents downstream signal pathway of IFN-3 and NF-κB which abrogates production of IFN- and ISGs [133].

1.3.2.2. Adaptive immune response

HCV reaches high titres in serum of infected patients in period as short as 1 week post infection [9]. Compared to the speed of HCV propagation, adaptive immune responses are relatively delayed. Adaptive cell responses are activated relatively late [9, 25, 134]. Lack of clinical symptoms like jaundice (caused by T-cell mediated liver injury) might be related to the delay in immune activation by HCV. Interestingly symptomatic infection correlates with the virus clearance. About 8-14 weeks into acute infection, alanine aminotransferase (ALT) level peaks. High levels of ALT are associated with liver damage. At the same time production of adaptive immune components is unregulated and HCV RNA titres fall [9, 25, 134]. This is followed by seroconversion and detection (in some cases) of anti-HCV antibodies. This event does not always correlate with outcome of infection. Around 20% of patients spontaneously clear the virus[135]. Clearance has been positively correlated with both- strong cell and humoral response, but exact determinants of clearance are not well defined.

It has been reported that spontaneous eradication of HCV is correlated with a strong cellular immune response [136-138]. HCV specific T cells can be detected in blood around 5-9 weeks and in liver 6-12 weeks post infection. As compared to anti HBV cell response, anti HCV is not only delayed but also impaired on the level of proliferation, cytotoxic activity and IFN- γ production [9, 138]. All this might contribute to low rate of viral clearance.

As already mentioned, anti-HCV antibodies arise relatively late in the infection or in radical cases not at all. Moreover they show restricted isotype profile and can disappear after recovery [139]. Antibodies specific to envelope glycoproteins E1 and E2 have been a major focus point [118, 140, 141]. It was shown that these antibodies can neutralise HCV pseudoparticles [101], and also protect chimpanzees from homologous

infection [142]. Unfortunately most of the neutralising antibodies tested so far seem to be strain specific and it can take as long as a year post infection for cross-reactive Abs to arise [143]. However our group and others described widely cross- reactive antibodies [141, 144, 145], which sustains their importance in HCV eradication strategies. The highest levels of neutralising antibodies are present in chronically infected patients, whereas resolved individuals do not maintain anti- HCV antibodies permanently [101].

1.4. Treatment

1.4.1. Current therapy

Current anti-HCV standard of care (SOC) is based on combination of PEG-IFN and ribavirin for a period of between 12 and 48 weeks. Interferons (IFN) were identified in 1957 after it was shown that cells infected by viruses secrete a factor protecting neighbouring cells against infection. Humans produce three types of IFNs: IFN-α, β and ω are classified in type I, IFN-γ is type II, and IFN-λ type III. Binding of IFN to cellular receptor triggers rapid expression of interferon-stimulated genes (ISGs), such as genes coding for MHC particles of class I and II, receptor for Fc of IgG, chemokines, protein kinase R [146, 147].

HCV seems to be easier to eradicate by treatment on early stages of disease. However, infections tend to be detected in chronic phase, usually years after HCV exposure when liver damage cannot be reversed. According to recent estimations, the current therapy is successful in more than 50% cases [147].

Currently, pharmaceutically produced IFNs are recombinant proteins obtained from cell culture systems. IFN with polyethylene glycol (PEG-IFN) shows longer lasting action and better absorption by cells. An additional compound that is used in anti-HCV therapy is the nucleoside analogue ribavirin. Ribavirin triphosphate is used by polymerase as GTP or ATP analogue during transcription. In HCV treatment ribavirin as mono-therapy failed to eradicate virus, however showed to enhance therapeutic effect when combined with IFN. However, many associated side effects and not satisfactory rate of success leaves room for improvement to this therapy regimen [147].

1.4.2. Differential responses to combination therapy

Combination therapy has a significant clinical toxicity, which in many cases prevent completion of the regime. Differences in virological responses have been reported to correlate both with viral and patient related factors. Individuals harbouring genotype 2 or 3 HCV are more likely to clear the virus as compared with genotype 1 infected ones [148, 149]. Also ethnicity and other patient's features have been correlated with treatment outcome. Prognosis seems to be worse for males than women and individuals of African origin as compared with Caucasians. Other negative prognostic factors include high levels of liver fibrosis, age over 50 years, BMI (Body Mass Index) over 25 [150, 151].

Recently published independent genome-wide association studies revealed a genetic polymorphism suggested to be linked with treatment response [152-154]. These studies have been carried out in populations of Japanese [154], Australian [152], and American patients [153]. In the latter individuals of European and African ancestry were included. Several Short Nucleotide Polymorphisms (SNPs) located near the IL28B gene, which encodes interferon-λ-3, strongly associate with differential response to

anti HCV treatment. Suppiah et al. [153] note, that genotype observed to lead to a better response to treatment is more prevalent in European populations, which might explain previous findings of better sustained virological response (SVR) for patients of Caucasian origin. Together these data imply that host genetic make-up has implications for prognosis of anti-HCV treatment response.

1.4.3. Direct acting antivirals and potential vaccine design

The limited success rate and severe side effects of PEG-IFN and ribavirin therapy generated rising interest in direct acting inhibitors of HCV life cycle. In recent years, following development of HCV cell culture system and resolution of tertiary structure of some of HCV proteins, putative targets for direct-acting antiviral agents have been identified. Figure 1.12 depicts HCV life cycle as a chain of sequential steps of which everyone offers a target for direct-acting drugs [155]. Many promising families of drugs have been identified in *in vitro* tests, some reaching clinical phase trials. Two families of molecules have been the biggest focus of antiviral strategies: inhibitors of NS5B RNA dependent RNA polymerase and inhibitors of the NS3/4A protease [155-157].

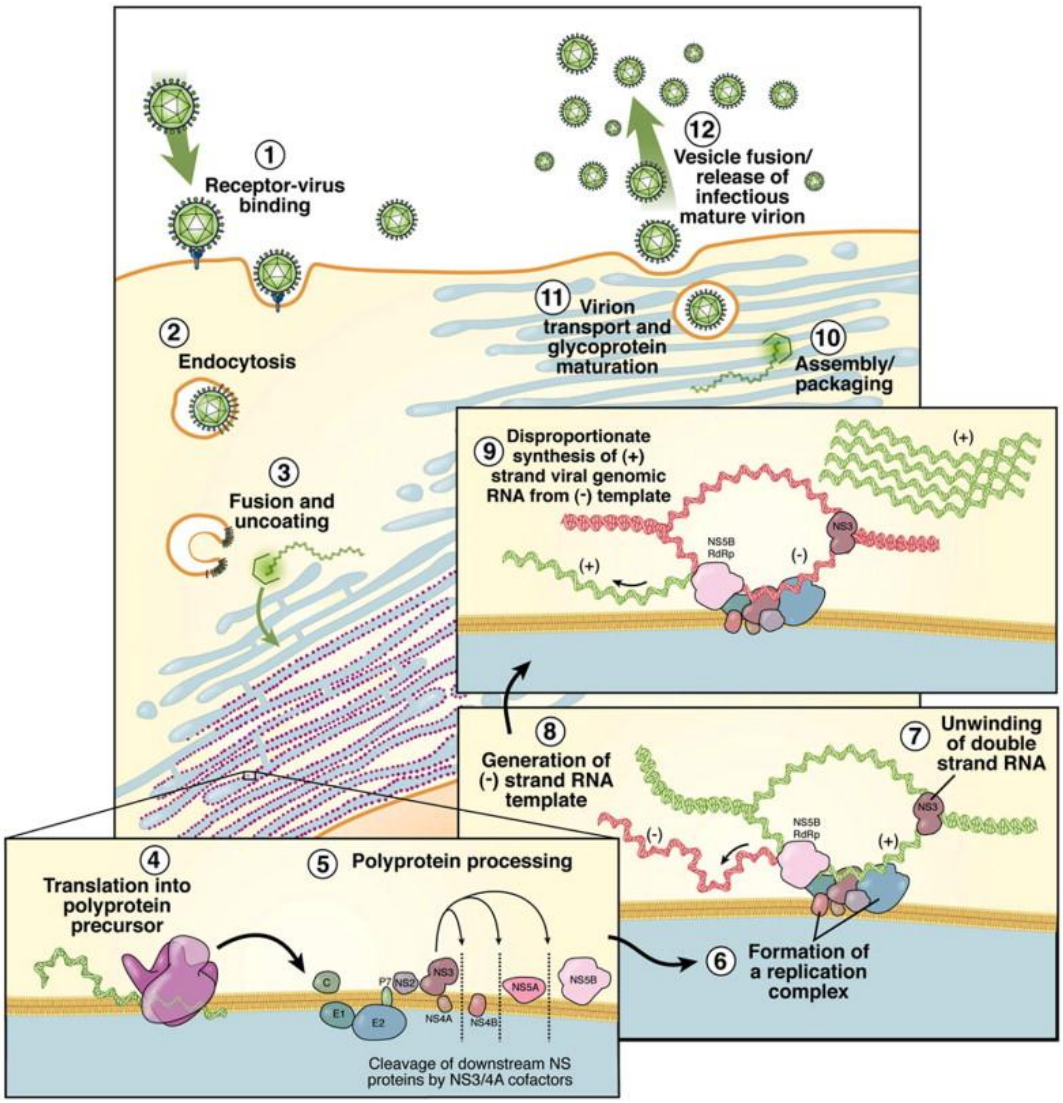


Figure 1.12. HCV life cycle- each steps representing potential target for antiviral therapies [155].

1.4.3.1. Protease inhibitors

Posttranslational processing of the HCV polyprotein engages a number of host peptidases and two viral proteases. The NS3/4A protease is responsible for cleavage at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B sites [42]. Research in developing NS3/4A protease inhibitors has seen the biggest success so far with two drugs: Telaprevir and Boceprevir being approved in Europe and United States in 2011 [157]. Monotherapy, despite showing high antiviral potential, is linked with

development of resistant HCV variants, often ending in viral relapse. However- protease inhibitors administered in combination with PEG-IFN α and ribavirin increase sustained virological response in both: treatment naïve patients and individuals who have previously undergone unsuccessful therapy [157].

Telaprevir is an orally administered linear, ketoamide protease inhibitor. This DAA was initially assessed in monotherapy with patients harbouring genotype 1 HCV. Median viral load decline of 4.4 log10 IU/ml HCV RNA was achieved with a dose of 750 mg every eight hours [158]. Unfortunately viral breakthrough with arising resistant variants was observed in majority of treated individuals [159]. Many subsequent short and long term trials of Telaprevir in combination with PEG-IFN α and ribavirin achieved more pronounced viral decline with reduced resistance profile (reviewed in [157]). Trials with ribavirin removed from treatment were not as successful, highlighting the importance of ribavirin in the drug combination.

Trials with Boceprevir, another ketoamide NS3/4A protease inhibitor, in genotype 1 infected patients that underwent earlier treatment showed HCV RNA reduction up to 2.88 log10 IU/ml [160]. Similarly as for Telaprevir resistance profiles were discovered as well as reduced potency against HCV genotype 2/3 [157].

Newly developed HCV protease inhibitors are currently being tested, with some in phase 1 and 2 clinical trials, all seem to exhibit similar SVR rates as described here Telaprevir and Boceprovir. Newly developed protease inhibition strategies could potentially overcome some of the side effects and, when used with SOC, reduce premature termination of treatment. Ritonavir, a compound enhancing a number of protease inhibitors' exposure is being tested with possibility of improving their pharmacokinetics [157].

1.4.3.2. Polymerase inhibitors

Replication being a complex multifactorial process offers many possible targets for antiviral interference. Most common approaches include nucleoside and non-nucleoside inhibitors of HCV polymerase and cyclophilin B inhibitors. Nucleoside polymerase inhibitors target the enzyme active site. Non-nucleoside molecules interfere with allosteric sites within RdRP. Drugs inhibiting cyclophilin B alter its ability to modulate RNA binding capacity of the viral polymerase [155].

Conservatism within NS5B active site is an obvious advantage in development of nucleoside inhibitors which so far are the only DAA in clinical trials conferring broad cross-genotype activity. Up to date several compounds targeting NS5B have been reported to induce a promising early virological response. Although they do not induce a mutational breakthrough, unfortunately most have severe adverse effects and their development has been put on hold [157].

Among the nucleoside polymerase inhibitors under trial a cytidine analogue-Mericitabine (RG-7128) is most developed and promising. In monotherapy the compound decreased HCV RNA levels by 2.7 $\log_{10}$ and when used in addition to SOC by 5 $\log_{10}$ IU/ml [161]. Observed viral resistance to Mericitabine is marginal. The compound is processed in kidneys without liver metabolism being involved and severe side effects have not been recorded [156].

The structure of HCV polymerase, crucial for the enzyme's function, can be compromised by non-nucleoside inhibitors. Four allosteric sites have been mapped as potential targets- two sites in thumb like domain and two in palm like domain [155, 157]. Benzimidazoles target thumb-1 domain, thumb-2 domain interacts with thiophenes. Palm-1 domain binds benzothiadiazines, palm-2 is targeted by

benzofurans. Drugs representing all above classes have entered clinical trials stage, however, resistant mutations are selected more frequently than for the nucleoside NS5B inhibitors. This is thought to be due to the non-nucleoside inhibitors binding sites being located away from the active site [157].

1.4.3.3. NS5A inhibitors

Recently much focus has been directed to NS5A inhibitors. NS5A having no known enzymatic activity for many years was considered less often for a target of direct acting antivirals. However in 2010 Gao et al. [162] reported development of BMS-790052, a small molecule inhibitor of NS5A. In vitro this molecule showed picomolar potency against a wide range of genotypes [162]. This molecule, also known as Daclatasvir, inhibits HCV RNA replication via interaction with NS5A, a crucial component of HCV replication complex [163]. Further studies showed that small molecules targeting NS5A influence subcellular localisation of this protein relocalising it from ER to lipid droplets [164]. Combination therapies of daclatasvir with NS3 inhibitors [165], or with IFN-α and RBV [166] have shown promising results in early clinical phases.

1.1.3.1. HCV cell entry as a possible antiviral target

Currently developed DAAs focus on non-structural protein targets and intracellular steps of HCV life cycle. From a mechanistic point of view HCV entry offers another attractive target for inhibition, encompassing multiple steps of virus-host interactions. To date several receptors have been defined as factors interacting with HCV envelope glycoproteins in a proposed concerted process of attachment, endocytosis and membrane fusion (described in 1.3.1). In vivo, inhibition of HCV entry process has been demonstrated for neutralising anti-E2 antibodies [141, 144], antibodies targeting

receptors: CD81 and SR-BI [97, 107, 167], cyanovirin-N [168], compound affecting E2 glycosylation and molecules preventing pH dependent endocytosis [73]. Studies performed in chimeric Alb-uPA SCID mice have proved that neutralising antibodies can protect against HCV infection [169], similar results were reported for chimpanzees [170]. Small synthetic molecules targeting the entry process were also found and are at various pre-clinical stages of development [171]. So far only anti-envelope antibodies have reached the clinical trials phase [155].

Entry being the first and crucial step in viral life cycle offers a chance for a prophylactic compound design. However, high variability of E1E2 glycoproteins coupled with asymptomatic progress of acute HCV infection might hamper the design of successful entry inhibitors. This highlights the importance of studies focusing on early stages of infection and HCV glycoprotein evolution under differing selective pressures present in infected host.

1.4.3.4. Vaccines

Viral evasion is a complex process engaging both host and viral factors; hence identification of protective immunity has been elusive. Although HCV in majority causes a chronic infection, neutralising antibodies and robust T cell immune responses have been correlated with viral clearance and thus point out a possibility of designing both preventive and therapeutic vaccines. Major concern in vaccine development is selection for escape mutants. A key issue in designing a HCV vaccine is the choice of target antigens. It is thought that an effective vaccine design will have to elicit both strong and broad cross reactive T cell, as well as neutralising antibody response to clear or prevent HCV infection.

Currently several classes of vaccines are being tested in clinical trials. Recombinant proteins are generated by isolation of the sequence coding for viral antigen and cloning into a heterologous expression system. Proof of concept behind prophylactic value of recombinant proteins was gained with the successful Hepatitis B vaccine [172]. This vaccine consists of HBV surface antigen produced in yeast cells and elicits protective antibody response. Contrary to HBV, HCV envelope glycoproteins are highly variable, especially in antibody binding regions. However, vaccines targeting envelope were designed and Phase I study of a recombinant E1/E2 vaccine with adjuvant MF59 reported [173]. In this study subjects tolerated the vaccine well and showed both neutralising antibody and T-cell proliferation in response to antigen [173]. A core derived vaccine ISCOMATRIX® [174] was reported to elicit specific humoral response in subjects receiving the highest dose. Due to very low level of HCV-specific T cells being detected this vaccine was proposed to be evaluated in a therapeutic approach.

Synthetic peptide vaccines are designed to induce HCV specific immune response by directly presenting vaccine epitopes to the T-cell receptor via HLA molecules. Amongst synthetic peptide vaccines IC41 (Intercell AG, Vienna, Austria) has progressed furthest. This vaccine is composed of five synthetic peptides containing four, highly conserved HCV epitopes that are prevalent in genotypes 1 and 2. When administered in healthy volunteers, IC41 adjuvanted with poly-L-arginine was well tolerated [175]. When given to HCV infected individuals, this vaccine has effectively elicited HCV-specific Th1/Tc1 responses. However this response did not have a significant effect on HCV viral load [176] suggesting that further optimisation is necessary.

An interesting approach to peptide based vaccine was presented by a Japanese group [177]. HCV infected patients were administered a peptide vaccine in stages, with personalised vaccination at the follow up stage, consisting only of these peptides that

were successful in eliciting immune response after first immunisation. Significant decrease of serum HCV RNA level was reported for 8 out of 12 study subjects [177].

Vector based vaccines have been an appealing concept as contained immunogens are not HLA restricted and can be of a broader spectrum than those within peptide vaccines. A highly attenuated poxvirus strain: Modified Virus of Ankara (MVA) lacking several immunomodulatory genes has previously been used as a vaccine platform for HIV [178]. Phase I trial of a MVA vector vaccine TG4040, containing NS3, NS4 and NS5B antigens, showed it was well tolerated [179]. In five out of 15 patients viral load reduction was observed (0.5-1.4 log) and associated with increase in T-cell response [178]. A trial of TG4040 in combination with SOC has been proposed. Recombinant adenoviral strategies delivering HCV NS3 to NS5B proteins are also tested in Phase I trials [180]. Two vectors Ad6 and AdCh3, to which humans are rarely exposed hence would not possess adenoviral specific antibodies, are used.

Another strategy in developing a T cell based vaccine is DNA vaccination. A Swedish company Tripep AB has recently produced a codon optimised, CMV promoter controlled, NS3/4A DNA gene vaccine [181]. This well tolerated vaccine has reduced HCV RNA load in four patients out of 12, which coincided with production of HCV specific T cell responses in three of them. These trials proved that electroporated in vivo DNA-therapeutic vaccination against HCV was a viable concept and further trials should be undertaken.

Although HCV has remarkable ability to persist and lead to chronic infection, a robust cellular and humoral response leading to virus eradication can be produced. A vaccine, to be effective, will have to elicit such broad and cross-reactive immune response. To date only few emerging vaccine constructs have entered clinical trials and their

reported success rate is limited. However, various approaches can be pursued, one being use of the therapeutic vaccines as additives supporting SOC.

1.5. Evolution of HCV

HCV has been classified into 6 genotypes consisting of many subtypes and further characterised by remarkable strain variability [7]. Recent years had shed some light on the genotypes' origins and factors involved in HCV evolution, both on population and intrahost level. However further studies are required to uncover reasons for HCV success in establishing chronic infection and possible ways of clearing the virus.

1.5.1. Origin of HCV genotypes

In the absence of historical data regarding time of Hepatitis C Virus occurrence its evolutionary origin can only be estimated by means of phylogenetic data analysis of sequences collected over last 20 years. Whilst some regions of the HCV genome can be easily used for such analyses, there are parts of genome that show greater and more rapid variability. These are mainly E1 and E2, with typical difference between genotypes of about 50% [11]. Some of the variability observed can be explained by Darwinian selective pressure. Envelope glycoproteins, especially hypervariable regions may be a target for neutralizing antibodies and virus sequence change effects in immune evasion and persistence of infection [182, 183].

Phenotypic differences between genotypes have been described for HCV. The most striking is the differing response to interferon treatment. On average, only 40-50% of chronically infected patients harboring genotype 1 HCV show long term response to combination treatment, as compared to 70-80% in those infected with genotype 2 or 3

[148, 149]. So far the exact mechanism underlying this difference has not been established. Less clear-cut are differences in HCV pathogenicity. There is some evidence to suggest that patients infected with a heterogeneous genotype 1 population are more likely to establish a persistent infection that goes on to develop severe liver disease [184, 185]. However, these issues need clarification via more in-depth longitudinal evolutionary studies coupled with phenotypic characterization.

Divergence of HCV genotypes has proved difficult to establish. Previously reported short times of origin contradict the high prevalence and diversity of HCV across the globe. Simmonds [186] has reviewed some of the divergence times that seem to correlate with historical evidence. Diversity amongst 1b unrelated sequence indicates a divergence time of about 60-70 years [11] which is supported by lack of geographical groupings of subtypes. Time of origin of genotype 2 has been predicted to be 90-150 years [11]. Relatively recent divergence of genotypes like 3a which is common for IDUs is consistent with their rapid spread via shared needles in 1960s [11]. Current phylogenetic analysis of genotype 4a in Egypt is compatible with its mass spread through treatment of schistosomiasis with non- disposable needles in 1950s. [187]. Smith et al. [188] estimate the time of HCV genotypes origin greater than 500 years ago. However recent studies appear to correctly describe the spread of HCV it is worth noting that sequences collected over last two decades might not be sufficient.

1.5.2. HCV intrahost population diversity and quasispecies dispute

HCV exists as a viral intrahost population and evolution of the variants results in virus immune escape and establishment of a persistent viral infection. The high rate of mutation is a direct effect of the low fidelity of the RNA dependent RNA polymerase

[189, 190]. It is known that random nucleotide substitutions occur at a very high rate. In chronic infection the mutation rate is $1.5\text{-}4.0 \times 10^{-3}$ per site per genome per year [188, 191, 192]. Variability is not distributed evenly through the genome, instead it is concentrated in regions like the E1E2 glycoproteins [182]. Although positive selection has been described, much of the observed HCV genetic diversity can be attributed to neutral sequence drift [189, 190, 193].

Sequence drift underlying diversification has been of major interest. The neutral theory of evolution suggests that changes should occur at a constant rate and represent accumulation of neutral changes [194]. Studies revealed that in coding regions of the HCV genome changes occur predominantly at synonymous sites [16, 189]. Some of the variability observed can be explained by Darwinian selective pressure. The relative pressure at a locus can be represented by the ratio of nonsynonymous to synonymous mutations [16]. Envelope glycoproteins, especially hypervariable regions may be a target for neutralizing antibodies and virus sequence change effects in immune evasion and persistence of infection [182, 183]. The nature and rate of substitutions in E1E2, especially in HVR1 during acute infection, has been demonstrated to correlate with disease outcome [80, 195]. Eradication of the virus in acute infection correlates with a stable HVR1 sequence population whilst chronic infection is associated with rapid evolution of the intrahost variants [193]. Although HVR1 is highly variable some sites exhibit extreme negative (conservatory) pressure, suggesting a key biological role for these region. [196].

Previously mentioned Farci et al. reported that diversity of HVR1 sequences was predictive of HCV infection outcome [80]. More diverse HVR1 seqences in acute infection correlated with progression to chronicity and less diverse HVR1 with resolved infection. This was described as evidence of HCV acting as a quasispecies [80]. However, term quasispecies is not a surrogate for genetic diversity; quasispecies

theory predicts that the population as a whole is a target for selection rather than individual variants from within the population [197-199]. To date there has been no experimental data which confirms that HCV behaves as a classical quasispecies. So far the only experimental data supportisng quasispecies theory in viruses was delived by Vignuzzi et al. [200]. Using artificially expanded Polio virus populations a link between viral population complexity with disease pathogenesis in experimentally infected mice was demonstrated. However quasispecies behaviour has yet to be definitively demonstrated in naturally occurring viral populations.

1.5.3. Transmission as an evolutionary event

The transmission of HCV can occur by different routes. Prior to discovery of non-A-non-B hepatitis and establishment of robust blood screening techniques major source of new infections was via transfusion of contaminated blood or blood products [13, 201, 202]. Transmission between sexual partners has been reported [203, 204] but risk of infection was correlated stronger with injecting drug usage than sexual behaviours [204]. Both horizontal and vertical intrafamilial cases of HCV seem to be mostly associated with parenteral route of infection [205]. However transmission has also been reported to occur before birth [206]. Although nosocomial and iatrogenic risks of HCV infection have been reduced in industrialised countries they are still prevalent in countries like Egypt where estimated 500 000 new infections occur yearly [207]. Major risks of HCV infections in countries with well-developed screening procedures are associated with injection drug usage [208, 209] but occupational risk of infections like through needle-stick are still present [210, 211].

Modelling events of transmission and acute HCV infection is challenging due to difficulty of virus detection at the earliest stages (patients are usually asymptomatic).

Most studies of acute HCV infection were lacking samples from donor populations to assess the dynamics of viral transmission. Studies in experimentally infected chimpanzees present many drawbacks and limitations. Research utilising immunocompromised mice with human liver grafts may provide a more robust model to enable study of HCV envelope glycoprotein evolution at transmission.

The majority of previous reports of HCV sequence evolution upon transmission have focused on the hyper variable region 1 (HVR1) at the N-terminus of E2 [90, 193, 212]. A number of groups have reported that identical HVR1 variants occur in donor/recipient pairs during viral transmission events tracked in both humans and chimpanzees [213-217]. Moreover, HVR1 variants can remain stable for several years after transmission in humans [13, 213], during transmission from human to chimpanzee [215, 217] and after further passaging of the virus in chimpanzees [217]. However differences were observed in dynamics of HVR1 evolution between different individual recipients of HCV transmission events [13, 212, 213], reflecting variable fitness requirements provided by different selective environments. The low intrapopulation genetic divergence of viral sequences observed in window phase of HCV infection [218] suggests that selective amplification or selective transmission occurs. HIV-1 transmission studies demonstrated that productive infection can be achieved by a single virus after heterosexual transmission [219, 220]. In contrary, in men who have sex with men multivariant transmission was observed to initiate HIV-1 infection [221]. Up to 16 variants were reported to establish infection in intravenous drug users, amongst which 60% of infections are attributed to multivariant transmission [222]. Transmission data regarding establishment of HCV infection is limited. However, deep sequencing technologies have been used to analyse viral genetic complexity in serial samples from patients undergoing acute HCV infection [223]. Contrastingly, these analyses report that HVR1 is highly variable at the intra-host level in acute infection. They also

demonstrate productive infection is achieved by a single or just a few viral particles, indicating a bottleneck at transmission [224].

The exact mechanism of HCV persistence is not known but the evolutionary adaptation of the virus to host selective pressure seems to play the crucial role. Adaptations in the founder virus population are shaped by equilibrium of beneficial factors and possible fitness costs of substitutions. The envelope glycoproteins have been of major interest as the viral proteins most readily in contact with host cells and the immune system. Their adaptation was suggested to be driven two ways. Substitutions allowing immune escape may be beneficial only if the adapted glycoproteins retain phenotype necessary for receptor binding and cell entry (see schematic model in Figure 1.13). Elucidating the evolutionary mechanisms underlying HCV transmission events and persistence of the infection will have implications for designing antiviral strategies targeting the HCV entry process.

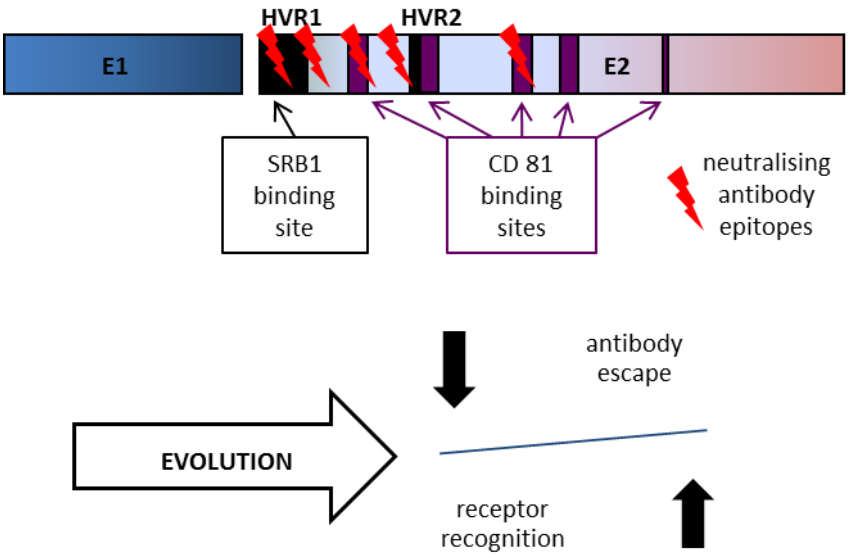


Figure 1.13. Representation of hypothesis of E1E2 glycoproteins evolution being driven by immune pressure and limited by conformational restraints of interactions with entry receptors.

1.5.4. Establishment of chronic HCV infection

The turning point between acute and chronic phase of HCV infection and factors involved are poorly understood. Anti HCV immune response develops with a delay whereas HCV propagates at an estimated rate of 10^{12} virions per day [119]. It is understandable that at this rate, and without proof reading capability, the RdRP will introduce changes which will be deleterious, selectively neutral or advantageous. Evolution of HCV within an individual is driven by a combination of drift and immune escape [189, 217]. Variants most fit to enter into host cells, replicate efficiently and evade eradication, will be selected for and predominate in a population, although selection pressures will change transiently. Outpacing of immune recognition by extreme sequence variability is one of the suspected mechanisms of HCV persistence.

Several studies reported mutations enabling HCV escape from CTL response in chimpanzees and humans (reviewed in [225]), the outcome of infection is also correlated with antibody driven escape substitutions within E2 during seroconversion [80].

Selection pressures exerted by CD4 T cells are thought to be limited as substitutions in epitopes of the major histocompatibility complex (MHC) II are rare [226]. However, selection of escape HCV variants by CD8 T cells is broadly reported [226-228]. Mutational escape in the context of MHC-I epitopes was first proved for RNA viruses in mice infected with murine lymphocytic choriomeningitis [229]. This mechanism was confirmed by observations made in natural HIV-1 infection [230]. Escape mutations were shown to be exerted by CD8 T cell pressure, especially in acute phase of infection [231], and to correlate with development of AIDS [232]. Statistically sufficient proof of similar selection in HCV infection was established first in chimpanzee model [233], where escape mutations were observed in multiple regions and were due to selective

pressure, indicated by relatively high ratio of non-synonymous to synonymous substitutions in these epitopes. Mutational escape from cellular immune responses was preliminary indicated in natural HCV infection by comparison of HCV sequences circulating in host to available prototypical sequences [234]. Further studies with identified donor inoculum constituted a critical mass of evidence of T cell mediated development of HCV escape [226-228] and indicated numerous T cell targets.

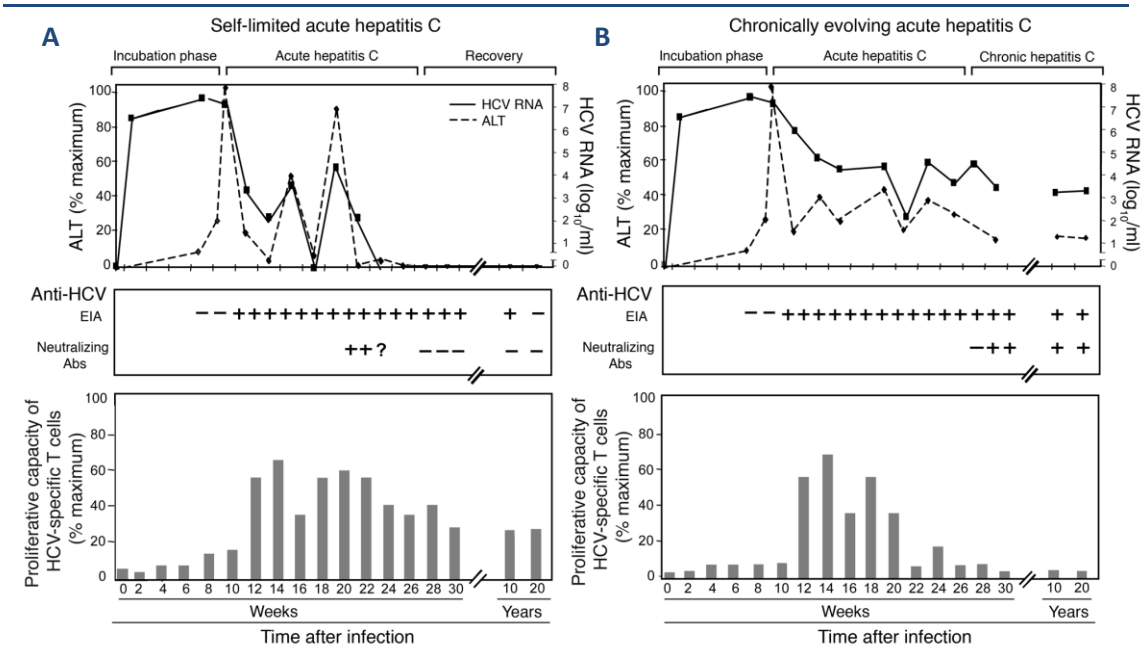


Figure 1.14. Schematic of the course of acute HCV infection followed by recovery (A) or chronic infection (B) [134].
HCV RNA level rises rapidly during incubation phase (first 2-3 months post infection). T cell responses and anti-HCV antibodies become detectable at the same time ALT levels elevate as an effect of liver damage. In a resolving infection (A) clearance of the virus is followed by decrease of anti-HCV antibodies levels that may eventually become undetectable. T cell responses remain detectable and antigenic recall is strong. Chronic infection (B) is characterised by persistent HCV RNA at level lower than during acute phase. Levels of HCV specific antibodies rise and T cell responses wane, with in vitro recall becoming weak or undetectable.

Role of humoral immune response in the outcome of acute HCV infection is controversial. However, HCV specific antibodies have been indicated to drive viral envelope evolution throughout acute and chronic phase [80, 235, 236]. Chronic infection is associated with detectable HCV-RNA levels beyond 6 months [25] despite presence of anti- HCV neutralising antibodies [237, 238], which often are high in titre

and broadly cross reactive [134]. A number of amino acid sites within and around known receptor and antibody binding regions are positively selected [192, 239]. Proof of mutational escape driven by humoral response has been provided in studies with sera sampled sequentially during HCV infection. Given serum sample was not able to neutralise envelopes of viral strains derived from the same sampling point, but had strong neutralising activity towards variants from earlier time points [235, 240, 241]. Neutralising antibodies accelerate HCV glycoprotein evolution during transition from acute to chronic phase [240] and maintain further selection of escape variants throughout the course of infection [241].

Together this data highlights that the knowledge of the sequence of incoming viral variants, or if lacking- sequential samples, might lead to determination of the capacity for, and extent of HCV evolution under immune pressure. This in turn would aid vaccine and DAA development.

A second hypothesis regarding viral persistence is that anti-HCV immune response is too weak to clear total virus from all hepatocytes and prevent persistent infection. HCV antigens being displayed at relatively low level might not stimulate initial response efficiently [136]. T cell responses have been demonstrated to correlate with the outcome of HCV infection [242-244], however production of both HCV specific T cells and antibodies is relatively late [134]. Patients followed into recovery from HCV infection showed readily detectable levels of specific CD4 T cells. Lack or weakening of initially strong response are associated with development of chronic infection or relapse of vireamia despite initial control respectively. Progression of HCV infection to chronic phase associated with continuous antigenic stimulation is thought to lead to exhaustion of HCV- specific T cells. Arising responses are of reduced breadth and impaired function [245].

Some of the HCV proteins have been identified to directly interfere with host immune response. The HCV Core protein has been shown to bind the receptor for the complement component C1q in vitro [246]. In in vivo mice study Core has been shown to downregulate T cell responses [247]. Additionally HCV proteins interfere with IFN activated pathways, which may further explain prevalence of chronic HCV infection. Strains poorly responding to IFN therapy- 1a, 1b as opposed to 2a, 2b and 3a possess within their E2 sequence domain homologous with phosphorylation site of the PKR and eIF2a. They compete for binding thus are able to inhibit kinase activity of PKR [248]. Furthermore PKR is also inhibited by NS5A [47] which has a broader anti-ISG activity. NS5A amino acids 2209 and 2248 have been termed interferon sensitivity- determining region (ISDR) which has been shown to be the PKR-binding domain with its affinity modulated by mutations [46]. NS3/4A serine protease has been indicated to disrupt IFN signaling by degrading RIG-I immune adaptor proteins [249].

Despite the fact that type I IFN can successfully block HCV replication in cell culture [148, 149] the anti- HCV therapy of pegIFN-α and ribavirin has a success rate between 40- 80% of cases with response differing depending on genotype of the HCV infecting patient. So far the exact mechanism underlying this difference has not been established. HCV seems to be able to down regulate IFN signalling pathway probably by preventing activation of a specific anti- HCV ISGs. In a study of biopsies taken from HCV infected patients, non- responders had highly activated ISGs both prior and post IFN treatment, hence it was suggested that HCV interplays with a tight subset of anti-HCV ISGs [250]. Moreover pre-existing high activation of many ISGs might itself have a negative effect on treatment efficacy, which preferably should aim to boost only specific anti-HCV responses. Ericson et al [251] found that consensus IFN, showing increased anti-HCV efficacy as compared to IFN-α, induces different subset of ISGs. Discovery of specific

and anti-HCV potent ISG and learning how they are controlled could possibly lead to enhancing the therapy success rate.

1.6. Laboratory models in HCV study

1.6.1. Animal models

As mentioned above HCV afflicts significant part of worldwide population and is a leading cause of liver transplantation [14], a vaccine being unavailable and current treatment in many cases ineffective and poorly tolerated [151]. The development of vaccine or alternative therapies has been hampered by the lack of suitable animal models of infection; however recent years have seen substantial breakthroughs in several new study models.

1.6.1.1. Chimpanzee, tamarin and tupaia

Resulting from HCV narrow species tropism for many years the common chimpanzee (*Pan troglodytes*) was the only animal available as experimental system. Experiments conducted on these animals enabled identification of basic facts related to the virus. Transmission of established non-A, non-B hepatitis from human to chimpanzee proved that causative agent was enveloped RNA virus, and that measurable RNA levels persisted within this host [252]. This further provided source of infectious serum for the initial molecular cloning of the HCV genome [3, 4]. A close genetic relationship between chimpanzees and humans offered possibility to study viral evolution [253] as well as host responses [254]. Unfortunately HCV infection progression and persistence in chimpanzees differs in some respects to humans. Antibody immunity is either weak or absent in this animal model. Multiple infections of chimpanzees did not result in

their immunity to both homo- and heterologic strains of HCV [139]. Direct immunisation of chimpanzees with recombined HCV antigens triggers production of anti- HCV antibodies which protective effects were disputable [139, 255]. Nowadays they continue to play a role in infectivity study and screening of antiviral molecules and vaccine candidates [256-258].

Unfortunately chimpanzee model has serious drawbacks and limitations. Retrospective analyses showed that hepatitis in chimpanzees does not reflect exact routes of pathogenicity of HCV in humans. While vast majority of infections in people lead to chronicity only 30-50% of chimpanzees develop chronic disease. The liver cirrhosis and fibrosis that are observed in humans does not occur in the chimpanzee liver [256, 259-261]. Such dissimilarities in HCV pathogenesis between humans and chimpanzees, in addition to both moral and financial considerations, have been integral in the development of additional animal models.

1.6.1.2. Surrogate virus model: GBV-B

Possible alternatives to chimpanzee model were identified. The GB virus B (GBV-B) is a flavivirus closely related to HCV [262]. Although persistent infections were reported after intrahepatic inoculation with synthetic GBV-B RNA, in most cases the virus causes acute, resolving disease in tree shrew (*Tupaia belangeri chinensis*) [263], tamarin (*Saguinas Ursula*) [264], marmoset (*Callithrix jacchus*) [265] and owl monkey (*Aotus trivigatus*) [266]. The possibility of using GBV-B as a platform for HCV chimeras might prove its potential for indirect testing of therapeutics [267].

1.6.1.3. Transgenic mouse model

Limitations of chimpanzee model in studies of liver pathology highlighted the need of obtaining small animal models for HCV infection. So far three types of murine models were designed: transgenic mice that express HCV proteins from tissue specific promoters in liver, immuno-compromised mice with human liver grafts and humanised rodents expressing human receptors. Use of the first model has been limited to studies of liver pathology under expression of HCV proteins like induction of hepatocellular carcinoma [268]. The second model seems to be more robust as it enables study of HCV propagation directly in a human tissue. The third model has been introduced recently [269]. These newly established animal models are still not widespread. Generating transgenic animals is technically difficult and very expensive; hence they have not yet been introduced into routine studies.

Additional model- Trimera mice have proved efficient in production of monoclonal antibodies against several pathogens including HCV [270, 271]. This human/mouse radiation chimera is achieved by total body irradiation and further rescue by SCID mouse bone marrow, and engrafted with human lymphocytes. It may serve as a model of infection if transplanted with infected human tissue [271].

Rodent species, even when highly immune deficient are not susceptible to HCV. The lack of permissiveness is most likely multifactorial including block in HCV entry [114]. Transgenic mice expressing human CD81 were not permissive to HCV suggesting that another host receptor is crucial [272]. Mouse cells expressing human CD81 and additionally OCLN were rendered permissive [114]. Although mouse CD81 is more than 90% identical to its human homolog, the 4 amino acid positions identified as crucial for the interaction with HCV envelope are variable in mouse derivative [273]. Nevertheless mCD81 can support HCV entry but at much lower levels than human

CD81 which enabled in vitro selection of E1E2 adaptations. Murine- tropic HCV could in principle efficiently enter mouse cells but there seems to be poor proof of successful assembly and release of infectious progeny (reviewed in [114]). Recently, drawing on the knowledge of CD81 and OCLN bearing species specific determinants for HCV entry, Dorner *et al.* [269] developed a genetically humanised mouse model for HCV infection. This system is the first one to allow HCV entry in immunocompetent rodent opening opportunities for HCV pathogenesis, entry and entry inhibitors *in vivo* [269].

Another approach in obtaining rodent HCV infection models is host adaptation by engrafting them with human primary or transformed cells [274]. This was achieved by transplanting human hepatocytes into murine host with severe immunodeficiency and liver injury. In most often used model transgenic mice overexpress urokinase-type plasminogen activator (uPA) driven by an Albumin promoter [275]. Hepatotoxic effect in homozygous mice leads to severe loss of hepatocytes which can be rescued by xenotransplantation of human hepatocytes. These chimeric mice are susceptible to HCV infection which is characterized by rapid increase of viral loads that are sustained for several weeks [276, 277].

Xenomice model holds the potential to facilitate assessment of antiviral agents [277]. It was previously used to show that HCV infection can be modulated with antibodies [169]. We retrospectively assessed evolution of E1E2 sequence pre and post inoculation of mice.

Although transgenic mice proved most advantageous of small animal models obtaining xenomice chimeras is still very cost and time consuming, limiting their use. The new humanised rodents might bring solution to these problems

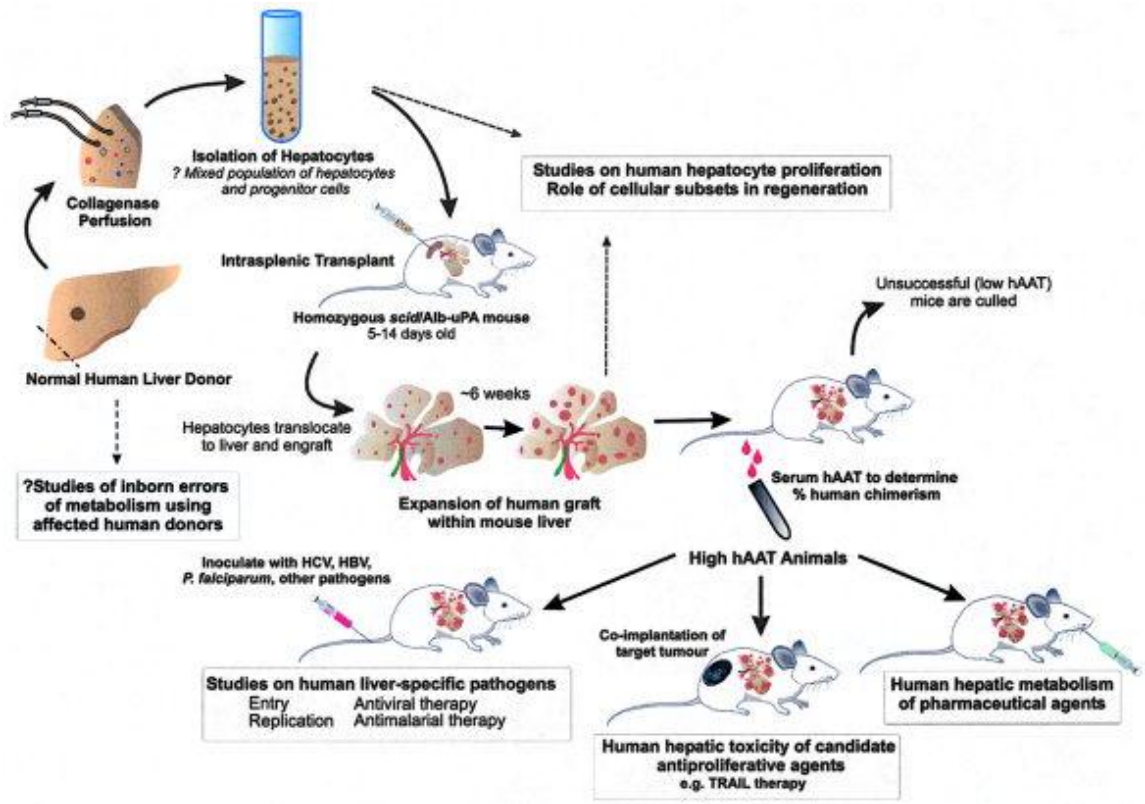


Figure 1.15. Chimeric Alb/uPA SCID mice production scheme [275].

1.6.2. Tissue culture models

For a considerable time the only measures available to study HCV life cycle were based on observation of infected patients' viral populations, experimental infections of chimpanzees and comparison with other *Flaviviridae* members. However, additional data on viral enzymes and structural proteins came from heterologous expression systems in different kinds of cells. Attempts to introduce different approaches in HCV study have succeeded in establishing widely used tissue culture experimental systems [278]. Most of these models allow only limited look on isolated aspects of viral life cycle, replication and pathogenicity.

Transient and stable expression of HCV proteins in different cells created a root for our current knowledge. Various expression systems were applied to identify cellular

localization, folding and structure as well as interactions between viral proteins and their influence on cellular proteins. As mentioned before most cellular receptors for HCV were identified in studies based on recombinant envelope glycoproteins. This approach however, has many limitations as expression environment is far from a natural one, which occurs during liver infection.

1.6.2.1. Subgenomic and genome length replicons

For many years studies of HCV replication were hampered by lack of permissive cell culture systems, hence subgenomic replicon systems were developed [279]. Subgenomic replicons contain sequence encoding antibiotic marker fused in frame in the core sequence which is followed by heterologous IRES to drive translation of described ORF containing non-structural HCV region. Dicistronic RNAs containing and expressing full length ORF have also been developed [280].

Introduction of a replicon system was a big step forward in studies on HCV replication and translation. In 1999, long before establishment of chimeric transgenic mice as robust HCV infection models, Bartenshlager's lab developed first HCV replicon system[279]. HCV replicons either contain genetically engineered whole genomes or shorter sub-genomic fragments. The latter ones comprise minimal non-structural region of HCV genome [281]. The first subgenomic replicon was based on genotype 1b non- structural region with two heterologous elements- neomycin phosphotransferase gene (introducing G418 resistance) and encephalomyocarditis virus IRES. In this system subgenomic RNA consisted of non-structural proteins coding regions as the NS2/3 protease, NS3/4A serine protease, NS3 NTPase- helicase, and NS5B[279]. In vitro transcription of cDNA resulted in RNA that could be used to transfect Huh-7 cells. Upon selection with G418 survivor cell clones were bearing up to 5000 replicons per

cell [281]. Interestingly, efficient replication of these HCV replicons correlates with acquiring so called cell-culture adaptive mutations [280, 282]. It is not clear how exactly these mutations influence viral replication especially that their introduction has not been observed in vivo. These mutations are thought to be responsible for the inability to raise infectious viral particles from full genome replicons [280, 282].

Replicon models played a crucial role in deciphering replication steps of HCV life cycle. However, to study viral entry or assembly and release of the progeny virus, a productive and robust system of viral infection in cell culture must have been established.

1.6.2.2. Viral pseudoparticles

As much as replicon systems shed a light on genuine HCV RNA replication and translation of polyprotein, retrovirus based HCV pseudoparticles (HCVpp) were important in study of virus entry into host cells.

Several surrogate models of HCV entry have been developed. Initially the problem was approached by use of insect cell lines infected with recombinant baculovirus containing the cDNA of HCV structural proteins [283]. Virus- like particles produced in this manner were not infectious and retained in an intracellular compartment. Therefore, it was difficult to evaluate how closely these particles resembled native HCV virions. Another approach was to produce virosomes by incorporating E1E2 glycoproteins into liposomes. Although these virosomes interact with cell surface receptors it has not been shown whether they can induce fusion [284].

The notion that HCV glycoproteins, when over expressed in heterologous systems were retained on the cell surface as well as in original site of replication on the ER, led

scientists to develop the retroviral pseudotype system. Retroviruses are well known to be able to incorporate into their envelope a variety of engineered viral glycoproteins. Moreover, the fact they can easily integrate long heterologous DNA sequences, allows the introduction of genetic markers. Murine leukaemia virus (MLV) or human immunodeficiency virus vectors were used to create particles pseudotyped with HCV E1E2 and containing reporter genes, that allow monitoring of infection [101, 285].

Pseudoparticle production relies on co-transfection of producer cell line with plasmid(s) containing coding sequences of retroviral core, packaging signals, reporter (like GFP or luciferase) and HCV glycoproteins (see Figure 1.16Figure 1.16). Viral capsids assemble in transfected cells and are subsequently transported to cell surface, where they acquire envelope from host cell membrane that contains E1E2 glycoproteins present as an effect of overexpression. HCVpp released into media of producer lines carry reporter gene that is released upon infection of naïve cells. Resulting expression of reporter transcript can be detected and measured, which reflects infection efficacy [101, 286]. Data collected with use of HCV pseudoparticles (HCVpp) proves their potential to mimic first steps of HCV infection. They exhibit preferential tropism to hepatocytes and are neutralised by monoclonal anti-E2 antibodies and sera from HCV-infected patients. Characterisation of E1E2 incorporated on HCVpp with conformation-sensitive antibodies confirms that functional unit in this system is a noncovalent heterodimer [66]. In addition coexpression of both glycoproteins has been shown to be crucial for production of infectious pseudoparticles. So far HCVpp are the only tool to study functional characteristics of HCV envelope glycoproteins derived from various genotypes and strains [97].

HCVpp can be produced in large quantities at a convenient safety level. Furthermore they offer great flexibility in terms of incorporation of marker genes and allow investigation of viral entry independently of replication [287]. The introduction of

pseudoparticle assay proved to be a great advance in HCV entry studies. Previously known putative receptors were confirmed in this assay, candidate receptors were characterized [288]. It has been revealed that none of the putative receptors alone is sufficient to mediate HCVpp infection in non-permissive cells and that set of coreceptors requires both CD81 and SR-BI [97].

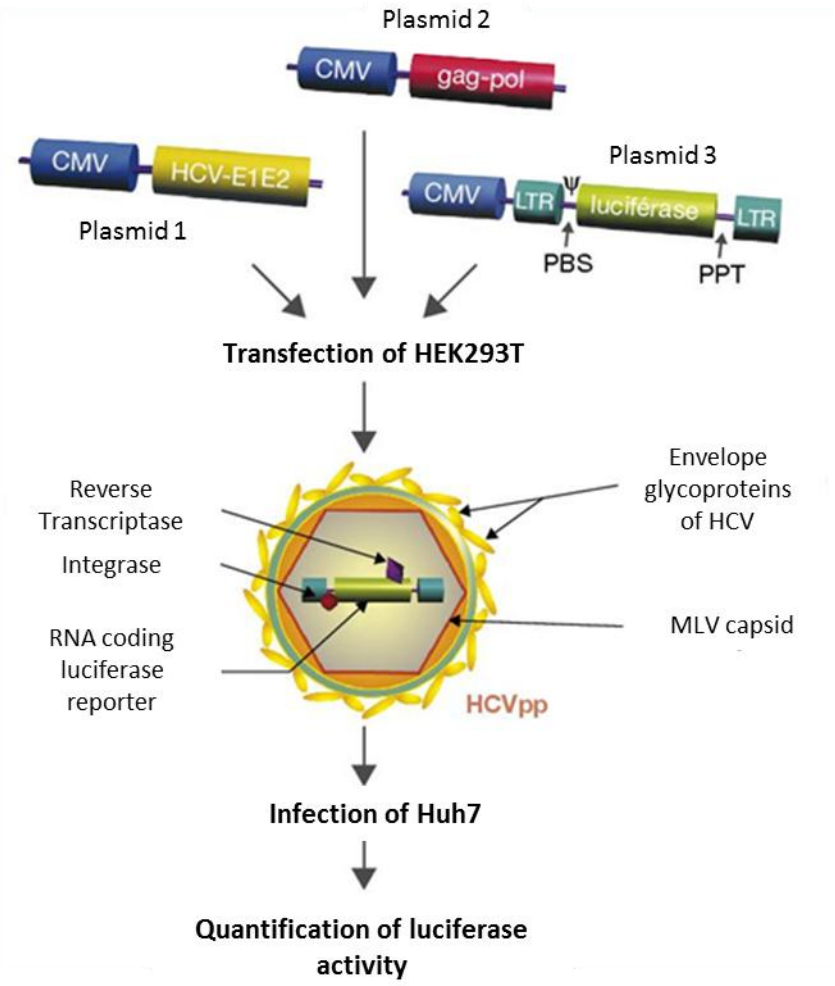


Figure 1.16. HCVpp production process (adapted after [289]).

Although HCVpp are similar to HCV virions they do not mimic the natural infection pathway. Pseudotypes do not interact with lipoproteins in the same way natural HCV virions do, hence lipoprotein mediated entry and role of LDL receptor cannot be dissected in this system.

1.1.3.1. Cell cultured HCV

A breakthrough in HCV research was discovery of a HCV strain that was able to replicate and produce infective particles in cell culture. Many groups tried to establish a system that would sustain HCV replication, most attempts were based on primary human or chimpanzee hepatocytes exposed to HCV contaminated patient sera, or on cultivation of cells from chronically infected patients [290-293]. HCV has been described to infect several types of cells- both primary and transformed, but replication detected is at a limited level [48]. Despite reported infectivity of HCV particles in some of those studies reproducibility of results was poor and none of the systems sustained efficient and long term replication.

Full replication cycle in tissue culture was obtained by electroporation of infectious RNA into a hepatoma cell line [294]. This most recent significant advance in HCV research started with isolation of a genotype 2a HCV strain, termed JFH-1, from a patient with fulminant hepatitis [295]. Initially JFH-1 was introduced as a new subgenomic replicon that possessed efficient replication despite lack of adaptive mutations. It was subsequently found that upon replication, viral particles assemble and are secreted outside the cell and are infectious for Huh-7 cells. It provided opportunity to study many aspects of the host-virus interaction and the whole life cycle including viral entry, assembly and release. Particles generated in this system not only resemble virions present in patient sera in their morphology but also interact with lipoproteins. Thus, lipoprotein-mediated entry pathway and virion release could now be studied [276, 296]. HCVcc infectivity was neutralized by CD81 specific antibodies and immunoglobulins derived from patient sera. Like for HCVpp, HCVcc infection showed specificity to hepatoma cell lines. Chimpanzees experimentally infected with

HCVcc developed disease symptoms identical to those generated upon infection with human derived virus [294, 297, 298].

Original strain JFH-1 remains unique in its ability to infect, replicate and successfully produce viral particles, with underlying reasons still unknown. JFH1 has been recombined to create various chimeras of which only wholly genotype 2a chimeras replicate efficiently in Huh-7 cells (reviewed in [282]). This model possesses a major drawback- only genotype 2a, which is not representative of the genotype 1 associated with highest risk of liver disease and least SOC responsive, can be propagated in such manner [299].

Currently many chimeras of different genotypes with parts of JFH-1 genome are being generated and introduced into the system. However it is still not known which region of JFH-1 genome is responsible for effective replication and infectivity of HCVcc.

1.7. Summary

Hepatitis C Virus is a positive sense enveloped RNA virus and the sole member of *Hepacivirus* genus within *Flaviviridae* family [4, 5]. Globally HCV infects an estimated 170 million people with a high rate of persistence. Only about 15-20% of the population resolve acute HCV infection; the vast majority goes on to develop a chronic infection, which is implicated in liver cirrhosis and hepatocellular carcinoma [4, 12, 300]. Current anti-HCV therapy is based on combination of pegylated Interferon alpha (pegIFN-α) and the nucleoside analogue Ribavirin for a period of between 12 and 48 weeks. Combination therapy is successful in around 40%-80% cases with a resistance profile correlated to genotype of the infecting virus [146, 147, 301], coupled with host SNPs close to the *IL28B* locus which are associated with successful treatment outcomes

[152, 154]. However, combination therapy is not specifically targeted, with the mechanisms of action poorly understood. Although development of small molecule inhibitors specifically targeting HCV proteins has been extensively reported [302-306], none target the entry step of the viral life cycle. Moreover the significant genetic heterogeneity of the virus can result in selection of drug resistant strains, viral breakthrough and treatment failure [307-309].

HCV circulates within an infected host as a swarm of genetically distinct but closely related variants [310]. The diverse nature of HCV infection is a direct effect of the low fidelity of the RNA dependent RNA polymerase (RdRp) coupled with a high replicative rate in vivo. Estimates indicate the viral RdRp makes 1 error per 10-100 kbp copied [311] and that more than 10^{12} virions are produced each day in an infected host [119]. However, genetic variability is not distributed evenly through the viral genome. Instead it is concentrated in specific regions, the most variable of which are the E1E2 genes encoding the envelope glycoproteins [182]. Although positive selection has been demonstrated to contribute to the evolution of E1E2 [183, 192, 195, 239], much of the observed HCV genetic diversity can be also attributed to neutral sequence drift [189, 190]. The E1 and E2 glycoproteins mediate cell entry and membrane fusion [312] as well as possess major determinants for neutralising antibody responses [313]. Some neutralising antibody epitopes overlap or are positioned closely to receptor binding sites [79, 313, 314]. HCV envelope glycoproteins are heavily glycosylated [56]. These potential N-linked glycosylation sites (PNGS) have been shown to possess various functions including masking of antibody epitopes, controlling correct folding and facilitation of entry into permissive cells [71, 140]. Non-synonymous substitutions at N-linked motifs may result in glycan knockouts, resulting in phenotypic changes affecting both receptor- ligand and antibody- epitope affinities. Interactions between the host organism and virus selects for sequence variants which are fittest to facilitate entry and

replication, as well as evading host immune responses [315]. The entry step of the viral lifecycle, facilitated by the envelope glycoproteins, presents a potential target for specific anti-viral drugs and/or monoclonal antibody therapies.

HCV is a blood-borne viral infection. Presently the main risk of infection in well developed countries is associated with IVDU, however nosocomial transmissions still happen [208, 209, 316]. Due to asymptomatic course of acute phase of hepatitis C infection, studying early evolutionary events of transmission and establishment of infection in natural setting have been challenging.

The majority of previous studies of HCV intrahost evolution have suffered from exclusion of important regions outside HVR1. HVR1 variants have been shown to remain stable upon transmission and even up to several years later [213-217]. Differences in HVR1 sequence evolution are patient dependent [13, 212, 213]. Some research indicates that like HIV-1, HCV can establish productive infection with a single or just a few viral particles.

Almost 20 years after discovery of Hepatitis C Virus it still remains one of major worldwide problems. Much knowledge has been gained about different aspects of life cycle, morphology of the virus or its pathogenicity. In face of highly dynamic evolution and the low success rate of current therapies, there is much still to be learned before satisfactory practices for managing HCV infection are developed. The study of viral evolution within patients at distinct disease stages, coupled with phenotypic studies of viral isolates to assess receptor binding and neutralization sensitivity over the course of infection using up-to-date methods, should shed light on HCV response to host-immune pressure over the course of natural infection.

1.8. Aims of the project

In the light of the described knowledge, we aimed to broaden understanding of diversity in HCV glycoproteins E1 and E2 sequences at different stages of infection. We were able to study samples derived from a xenomouse model of transmission and several samples from naturally occurring transmission as well as sequences from acute stage of infection.

In this study we utilised single genome amplification (SGA) technique to recover full-length E1E2 sequences. SGA of full-length E1E2 glycoprotein sequences, followed by direct sequencing minimises in vitro generated artefacts and experimental biases associated with the standard bulk amplification and cloning approach, giving an accurate representation of investigated intrahost population. Our prominent aim was to advocate this method, widely used in evolutionary studies of HIV, as a new gold standard in HCV research.

Comparative analyses of sequences derived from three different settings were performed, in conjunction with a range of phylogenetic tests. As full-length E1E2 sequences were utilised in this study, we also aimed to investigate the phenotypic consequences of E1E2 evolution.

2. Materials and methods

2.1. Source of samples

To investigate genetic and phenotypic variability of HCV envelope glycoproteins, sequences derived from three novel clinical or experimental settings were utilised.

For the first sequence dataset, HCV RNA samples were obtained as described in a previously published study [169]. Eleven human liver-chimeric Alb-uPA/SCID mice were inoculated intrajugularly with 100 µl genotype 1a HCV infected serum KP (2.3×10^6 IU/ml) to simulate a natural exposure to virus. Patient KP acquired HCV via an unknown route and at the time of sampling was in the chronic phase of infection. Experimental infections were monitored over time by tail bleed sampling. HCV RNA in mouse serum was quantified by a real-time 24 TaqMan PCR assay [169]. RNA was recovered from serum aliquots with use of commercially available RNA extraction kit (Qiagen) and resuspended in 20 µl of H_2O. The sequences obtained from these RNA samples were designated the "mouse model" dataset (see Table 2.1 for sample details)

The second sequence dataset was generated from a proposed hospital acquired transmission event where the donor and recipient were suspected [317]. The index case patient had been receiving haemodialysis three times a week at the same unit for 7 years. Blood samples were screened every three months and were HCV negative until September 2009 when routine (unknown) HCV antibody test reached equivocal concentration. Retrospective analysis of stored samples revealed HCV PCR +/antibody - results for a sample taken 3 months earlier, suggesting acute HCV infection. Investigation of possible risk factors leading to index case infection pinpointed treatment received in March/April time for diabetic ulcer at the renal inpatient clinic.

Four haemodialyses were performed of which one was directly after a HCV positive patient. A transmission event between these two haemodialysis patients was suspected and subsequent genotyping identified the infecting HCV genotype as 3a in both cases. The putative date of the transmission event was established to be 26.03.2009. Sequence data was generated from serum samples obtained from these two patients, along with epidemiologically unrelated genotype 3a sequences from various sources including nine Nottingham and eight Glasgow cases. The sequences utilised in this part of the investigation were termed the "transmission" dataset (see Table 2.2). RNA was extracted from this set of samples utilising two different methods described in 2.2.

The third set of sequences was derived directly from patient samples previously enrolled in the Trent Study cohort [318]. Serum samples utilised were collected exclusively from patients in the acute phase of infection of unknown source. Patients were infected with genotype 1a HCV. Window pre-seroconversion period of disease was determined by HCV RNA presence in serum and lack of detectable anti-HCV antibodies in second and third generation enzyme linked immunosorbent assays (unknown manufacturer). The presence of HCV RNA in samples was determined initially by an inhouse RT- PCR assay and, post 1995, by a commercially available reaction (Amplicor; Roche Diagnostics, East Sussex, UK)[318]. HCV glycoprotein sequences generated in this part of project were designated as the "acute" dataset (see Table 2.3).

Table 2.1. Transmission in mouse model samples details.

Patient/ Mouse ID	Population sampling	Serum HCV (IU/ml)	Hepatocyte donor	Time post-inoculation (days)
Donor:				
KP	KP	1×10^6	-	-
Recipients:				
A594	594_1	8.09×10^5	II	32
	594_2	1.01×10^6	II	59
N666	666_1	2.69×10^6	I	14
	666_2	2.49×10^6	I	28
N714	714_1	1.16×10^6	I	14
	714_2	1.54×10^6	I	28
A931	931_1	8.61×10^7	III	21
	931_2	2.94×10^7	III	42
	A931	8.04×10^7	III	28
A594	A594	8.09×10^5	II	32
A583	A583	8.40×10^5	II	32
A596	A596	1.73×10^5	II	32
A585	A585	2.99×10^5	II	32
A587	A587	9.79×10^6	II	32
A902	A902	4.46×10^7	III	28
A909	A909	1.96×10^7	III	28
A965	A965	7.73×10^6	III	28

I = 36 year old female, II = 45 year old female 92% viability polycystic tissue, III = 4 year old male 92% viability normal tissue

Table 2.2. Transmission samples details.

	Sample Label	Patient initials	Sample number	Collection time	Extract or serum sent
Index case samples	IC1	JL	09.626410	04.06.09	serum
	IC2	JL	09.641669	03.09.09	nucleic acid
	IC3	JL	09.643461	12.09.09	nucleic acid
	IC4	JL	09.644581	19.09.09	serum
	IC4*	JL	09.644581*	19.09.09	nucleic acid
Hypothetical source samples	HS1	JM	09.605096	05.02.09	serum
	HS2	JM	09.621248	07.05.09	serum
	HS3	JM	09.628582	17.06.09	serum
Unrelated control patient samples	PG	JMK	09.636795	05.08.09	serum
	PF	Bi	09.625613	03.06.09	serum
	PC	O	09.618182	21.04.09	serum
	PD	B	09.618221	21.04.09	serum
	PA	Q	09.617409	15.04.09	serum
	PE	H	09.623014	15.05.09	serum
	PB	G	09.618166	21.04.09	serum
	PH	M	09.639212	20.08.09	serum
	PI	Be	09.638115	13.08.09	serum
	PT	N/A	N/A	N/A	serum
	PS	N/A	N/A	N/A	serum
	PR	N/A	N/A	N/A	serum
	PP	N/A	N/A	N/A	serum
	PO	N/A	N/A	N/A	serum
	PN	N/A	N/A	N/A	serum
	PM	N/A	N/A	N/A	serum
	PL	N/A	N/A	N/A	serum
	PK	N/A	N/A	N/A	serum
	PJ	N/A	N/A	N/A	serum

IC- index case, HS- hypothetical source, P…- patient- control samples from the unit, N/A- not available

* IC4- 09.644581- both serum and extract sent, only a small volume of serum left

Table 2.3. Acute samples details.

	Patient initials	Sample name	Date of collection	Special notes
Acute 1	N/A	1A20	unknown	N/A
Acute 2	MA	1A30.3	07/10/2003	2 months after first sample collected, month after HCV RNA detected, Anti-HCV negative
		1A30.12	04/02/2005	16 months post infection, RNA+, Anti-HCV+, patient deceased
Acute 3	KB	1A57.2	02/09/2004	5 months after first sample collected, first time HCV RNA detected, no information about Anti-HCV, later sample still Anti-HCV negative
		1A57.14	01/02/2008	42 months post infection, RNA+, Anti-HCV+
Acute 4	RW	1A65.1	05/04/2005	first sample taken, positive RNA, negative Anti-HCV, first Ab+ sample collected 5 months post PCR+
		1A65.10	05/06/2006	14 months post infection
Acute 5		1A83.1	14/12/1996	N/A

N/A- not available

2.2. HCV RNA extraction

Two separate RNA extraction methods were employed dependent on samples. Both are derivatives of Boom's method [319]. From the acute dataset samples, RNA was recovered from 140- 200 µl of serum using a commercially available QIAamp® MinElute® Virus Spin RNA extraction kit (Qiagen). Serum and buffers were equilibrated to room temperature prior to extraction. Serum was mixed with 25 µl QIAGEN protease and 200 µl of buffer AL containing 28 µg/ml of carrier RNA and guanidine hydrochloride. Lysis was performed at 56°C for 15 minutes. Lysate was then incubated at room temperature with 250 µl of ethanol to allow binding of RNA onto the QIAamp MinElute column. RNA was bound to the column upon centrifugation at 6000 × *g* for 1 minute. Three steps of washing were then performed- first with 500 µl of buffer AW1 (containing guanidine hydrochloride), second with 500 µl of buffer AW2 and third with 500 µl of ethanol. After removal of residual contaminants, ethanol carryover was removed by additional centrifugation step. RNA was then eluted with 20- 100 µl of buffer AVE.

Part of the "transmission" RNA (of a very low titre) was extracted with use of the bioMerieux's system- the NucliSens® easyMAG® platform. This system for automated isolation of nucleic acids from clinical samples is based on silica extraction technology as manual method but introduces use of magnetic particles to isolate viral RNA. Serum input volumes were set at 200 µl and eluate volume at 25 µl. 2 ml lysis buffer was added to sample according to manufacturer's pre-set protocol and 550 µl of magnetic silica (prepared in a 1:1 ratio with molecular grade water) introduced. RNA bound to silica was then washed 5 times, further by addition of final buffer and heating eluted from silica beads. Magnetic arrays removed silica beads from the sample and pure RNA eluate was retrieved.

2.3. cDNA synthesis

Viral cDNA was synthesised using commercially available Thermoscript™ RT-PCR System for first strand cDNA synthesis (Invitrogen).

Nine µl of viral RNA were used in a 20 µl reverse transcription reactions. Firstly, RNA was mixed with 10pmol primer OAS 1a (see Table 2.4) and 20pmol dNTPs and incubated at 65ºC for 5 minutes in order to denature template and primer. This step also removes RNA secondary structures that could impede full-length cDNA synthesis.

Denatured RNA was then combined with 8 µl of extension master mix that consisted of 15 U ThermoScript ™ RT, 40 U RNaseOUT™ and 0.1µM DTT in 1× cDNA synthesis buffer. Elongation was performed at 50 ºC for 1 hour or longer. Reaction was terminated at 85 ºC for 5 minutes and DNA:RNA hybrids were digested via incubation with 2 U of RNase H at 37ºC for 20 minutes.

2.4. Amplification of E1E2

Two PCR based methods were used in sampling viral quasispecies. The first method involved generating full length E1E2 glycoprotein sequences using unknown amount of cDNA molecules as a template, followed by direct sequencing of PCR products. This is referred to as “bulk amplification” method as aliquots of non-diluted cDNA samples were used in PCR amplification. It was mainly used to amplify E1E2 fragments from low titre samples or to serve as a comparison to single genome amplification method described below.

In order to avoid *in vitro* generated recombinants, cloning induced errors and also to ensure accurate and representative sampling of quasispecies populations, a single genome amplification (SGA) approach was utilised. The SGA method, followed by direct sequencing, has been previously applied to characterise viral quasispecies populations in HIV-1 [220, 320-322] and HCV infection [323]. In this approach viral cDNAs were serially diluted and aliquots of 2-fold dilutions (1 µl) were used as template in the first round of a nested full-length E1E2 PCR. Two µl of first-round product were subsequently used as template in second-round reactions. For each sample, end-point titration PCRs revealed the dilution at which the concentration of viral cDNA was <1 molecule per µl. Subsequently multiple E1E2 amplicons were obtained for each sample at the end-point dilution (≤3/10 PCR positives). According to Poisson distribution, when no more than 30% of PCR reactions are positive, more than 80% of amplicons will be derived from single molecule templates [321].

These amplicons were sequenced directly and further referred to as "single molecule" derived quasispecies sequences. Chromatograms were checked by eye for presence of double peaks (see Figure 2.1) that would disqualify an amplicon from further analysis due to being derived from multiple templates.

Both bulk and single-molecule methods utilised a nested PCR comprised of two rounds of amplification primed by two sets of primers. A second pair of primers was designed to attach to sequences positioned internally on the template in relation to the first pair of primers. For every PCR carried out negative controls were used to ensure lack of carryover contamination.

Amplification reactions were set up in 20 µl volumes containing 4 pmol of sense and anti-sense outer primers (see Table 2.4), 0.25 mM dNTPs, 0.5 U of Platinum® *Taq* High Fidelity polymerase (Invitrogen), 1× High Fidelity polymerase buffer and 2 mM $MgSO_4$.

One µl of neat or diluted cDNA was used as reaction template. The PCR-cycling parameters were 35 cycles of 94°C for15 seconds, 50°C for 30 seconds and 68°C for 3 minutes. Two micro litres of the first-round product was subsequently used in second-round reactions with inner primers using identical amplification and cycling conditions to the first round but increasing the cycle number to 45.

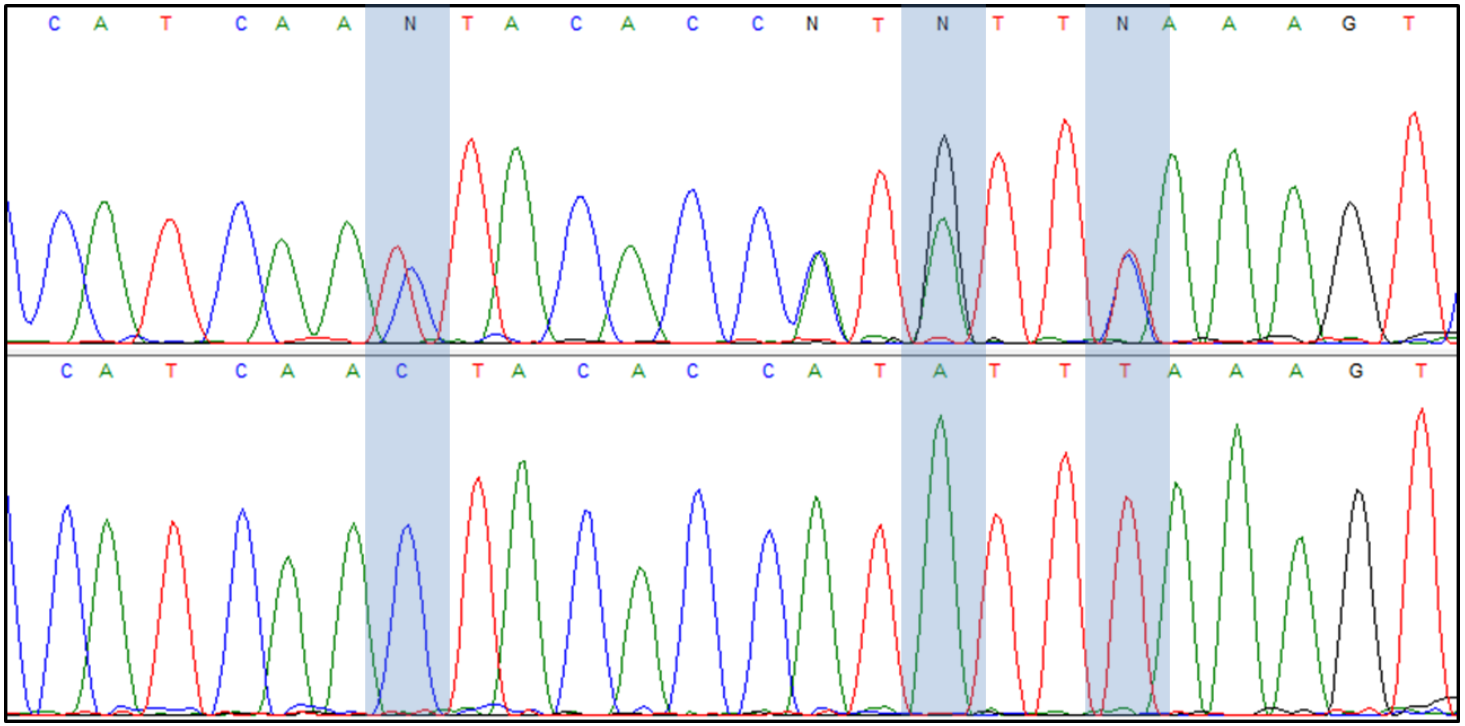


Figure 2.1. Sequence chromatogram representing fragment of a mixed sample compared to sample derived from single molecule.
Highlighted are double peaks in top panel- representing sequence derived from mixed sample. Bottom panel- single molecule amplified sequence, where clean peaks are highlighted.

Table 2.4. List of primers for nested PCR

Primer	PCR round	Direction	Genotype specificity	Sequence	Nucleotide position[1]
OAS 1a	cDNA synthesis, 1st	Antisense	1	5' GGGATGCTGCATTGAGTA 3'	2599 ← 2616
EOS	1st	Sense	Universal	5' GGACGGGGTAAACTATGCAACAGG 3'	818 → 841
1ASGT1a	2nd	Antisense	1	5' TTACGCCTCCGCTTGGGATATGAGTAACATCAT 3'	2550 ← 2582
170gt1	2nd	Sense	1,3,5	5' CACCATGGGTTGCTCTTTCTCTAT 3'	843 → 868
OAS3	cDNA synthesis, 1st	Antisense	3	5' TGCGCTGAGGGCGTTCAG 3'	2620 ← 2601
IASGT3A	2nd	Antisense	3	5' CCAATTATGCTTCTGCCTGTGATATCATCAGC 3'	2555 ← 2586
142-KpnI	Pre-cloning	Antisense	1	5' ATACATACGGTACCTTACGCCTCCGCTTGGGATATGAGTAACATCAT 3'	3'E2+ KpnI
179-HindIII	Pre-cloning	Sense	1	5' CACCATACAAGCTTCATGGGTTGCTCTTTCTCTATC 3'	3' Core+ HindIII

[1] primer position numbered according to H77 genome coordinates; all primers were previously designed by other members of our group

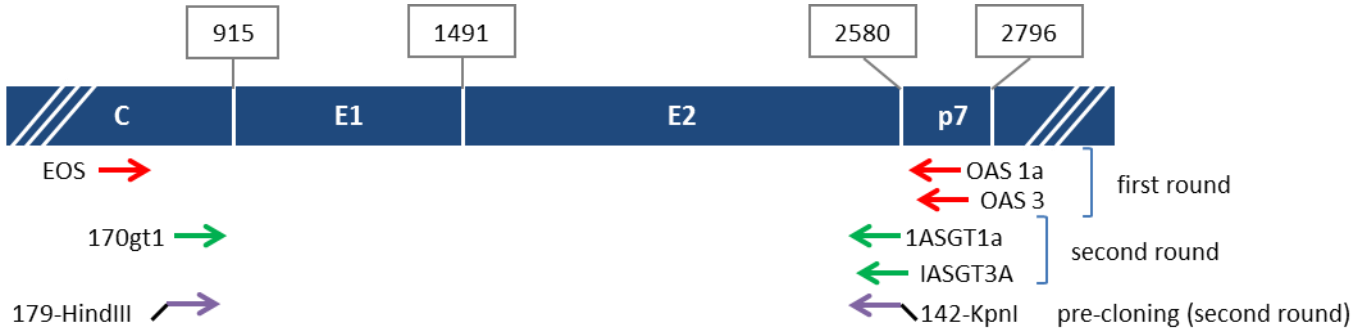


Figure 2.2. Schematic representation of primer locations (according to H77 coordinates) for primers presented in Table 2.4.

2.5. Agarose gel electrophoresis

Analysis of PCR products by eletrophoresis was carried out alongside Gene Ruler DNA ladder mix (MBI Fermentas). Gel was obtained by boiling agarose in 1× tris-acetate ethylene diamine tetra acetic acid (TAE) buffer (Sigma) at 2% weight to volume concentration. Ethidium bromide was added prior to casting the gel (5 µl / 100 ml). Five µl of PCR products were added to 2µl of 6× loading buffer (50% glycerol in water, 0.002% bromophenol blue). Samples were resolved at 6 V/cm for 40-50 minutes. Subsequently DNA was visualized using UV trans-illuminator (UVP). The result of each amplification was then recorded photographically.

2.6. Purification of PCR products

Before subsequent sequencing or cloning PCR products of verified size were purified using QIAquick ® PCR Purification Kit (Qiagen) according to manufacturers' protocol.

One volume of PCR product was mixed with five volumes of Buffer PB, which contains guanidine hydrochloride and isopropanol. The mixture was bound to silica-gel based column by applying a vacuum. To remove unincorporated nucleotides, primers, enzyme and salts, 750 µl ethanol based buffer PE was drawn through the column, again by applying a vacuum. To remove all traces of wash buffer, the column was further spun for one minute at 15800×g. DNA was then eluted with 30-50 µl elution buffer, quantified by spectrophotometry and stored at -20ºC. Measurements were performed with a NanoDrop spectrophotometer (Thermo Scientific) which allows accurate quantification of DNA concentration in 0.5-2µl sample. The peak of light absorption is

at 260 nm for DNA and at 280 nm for protein [324]. NanoDrop software calculates DNA concentration and informs about its purity.

2.7. Sequencing reactions

Amplification products were directly sequenced using BigDye™ Terminator chemistry (Perkin Elmer). This method is based on the dideoxy sidechain termination method developed by Sanger [325] combined with energy transfer dye [326].

Sequencing reaction were set up in 10 µl volumes containing 3.5 µl of Better Buffer (Applied biosystems), 0.5 µl Big Dye™ (Applied biosystems), 3.2 pmol primer (see Table 2.5) and 20-40 ng of PCR product (1.7 kb). The PCR-cycling parameters were 30 cycles of 96°C for 30 seconds, 50°C for 10 seconds and 60°C for 4 minutes. All steps were ramped at 1°C per second.

Amplified ssDNA was diluted with 10 µl of molecular grade water. Precipitation was then performed by addition of 50µl 100% ethanol, 2µl 3M sodium acetate and 2µl 125mM EDTA. Incubation at room temperature was performed for between 1-24 hours in darkness.

Depending on the reaction vessel in which sequencing was performed, centrifugation steps differed. Individual tubes were centrifuged at 18000 × g for 30 minutes at room temperature. Supernatant was then removed by careful pipetting and the pellet was washed twice with 70% ethanol. 96-well plates were centrifuged at 2000 × g for 45 minutes at 4°C and supernatant discarded by centrifuging the plate upside down at 185 × g for 30 seconds. Pellets were washed once with 70% ethanol. Residual alcohol in both cases was removed by evaporation on a thermoblock at 50°C

Sequencing reactions were read on an ABI 3130 analyser. Chromatograms were checked and manually adjusted in FinchTV (Geospiza) and contigs of overlapping sequences were assembled with use of SeqMan® (DNAStar Inc.).

Table 2.5. List of sequencing primers.

Primer	Template	Direction	Sequence	Nucleotide position[1]	Origin
T7	pcDNA3.1	Sense	5' TAATACGACTCACTATAGGG 3'	863 → 882 (upstream of insert)	Invitrogen
BGH	pcDNA3.1	Antisense	5' TAGAAGGCACAGTCGAGG 3'	1111 ← 1128 (downstream of insert)	Invitrogen
E1IS	E1E2	Sense	5' TGGGATATGATGATGAACTGG 3'	1299 → 1319	Previously created in our group
1A-FOR-seq	E1E2, gt 1A	Sense	5' TCGGAGGGGTGGGCAACAA 3'	2053 → 2071	Specifically designed for the project
1A-REV-seq	E1E2, gt 1A	Antisense	5' GCCCGCYAGGACTCCCCA 3'	1398 ← 1415	Specifically designed for the project
3A-FOR-seq	E1E2, gt 3A	Sense	5' TTCACACCATCGCCAGTGG 3'	1866 → 1884	Specifically designed for the project
3A-FOR2-seq	E1E2, gt 3A	Sense	5' CAAACGGTCCAGACCTGTAAC 3'	1236 → 1256	Specifically designed for the project
3A-REV-seq	E1E2, gt 3A	Antisense	5' CATGGAGTAATAGGCTAGGC 3'	1414 ← 1433	Specifically designed for the project

[1] primer position numbered according to H77 genome coordinates apart from primers T7 and BGH- position numbered according to pcDNA3.1 V5/His© TOPO® sequence (Invitrogen)

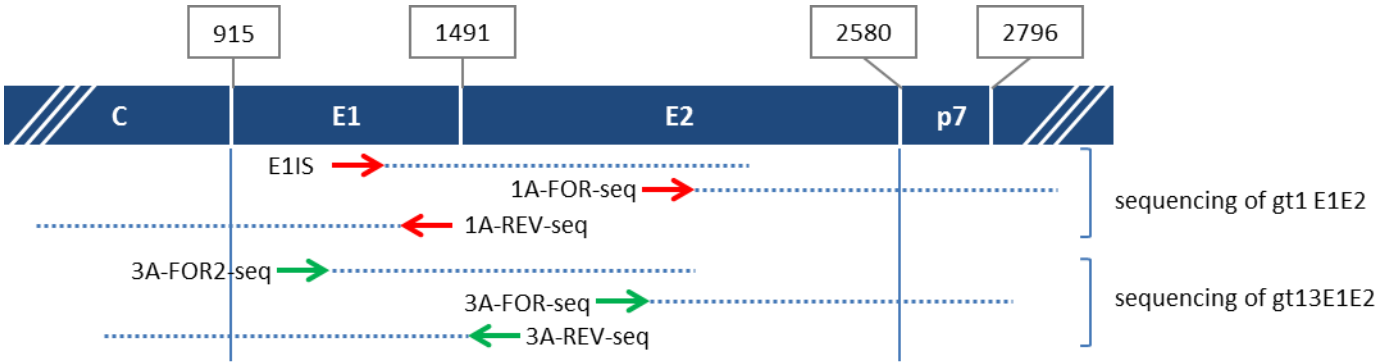


Figure 2.3. Schematic representation of primer locations (according to H77 coordinates) for primers presented in Table 2.5. and approximate overlap of sequencing products further aligned into full length E1E2 contigs.

2.8. Sequence Alignment

Sequence fragments were assembled into an individual E1E2 consensus sequence for each amplicon using SeqMan® (DNAStar Inc.) Nucleotide sequences of full-length E1E2 were then aligned according to their overlying amino acid translation using MEGA4 [327] which implements CLUSTAL W [328] source code, and then manually refined to ensure maintenance of the open reading frame. CLUSTAL W is a progressive multiple alignment method [328] which allows gap penalties to vary in a position- and residue-specific manner.

2.9. Phylogenetic reconstruction

Phylogenetic trees were generated by two separate methods: neighbour-joining (NJ) and maximum likelihood [329]. The NJ method for tree-building from nucleotide sequence data requires a pre-defined distance matrix. A pair of neighbours is defined as a pair of operational taxonomic units (OTUs) connected through an interior node in an unrooted, bifurcating tree. Among the possible pairs of OTUs, the pair that gives the smallest sum of branch lengths is combined, and subsequently regarded as a single OTU. The principle of NJ tree construction is to find operational taxonomic units (OTUs), or terminal nodes, that minimise the total branch length at each stage of successive clustering of OTUs, originating from a star-like tree. [330]. NJ trees were constructed using MEGA4 under a Kimura-2 parameter model [331] of sequence evolution whereby differential weightings are assigned to transitions and transversions [332]. The NJ method combines computational speed with uniqueness of result

Statistical robustness of phylogenetic groupings were estimated using the bootstrap approach via 1000 replications (reviewed in [333]). Bootstrapping is a reliability test which involves generation of pseudoreplicate nucleotide alignments by randomly sampling (with replacement) from the sites in the original alignment. The procedure is repeated 1000 times and a tree is generated for each replicate dataset with scores for identical grouping of sequences given as percentages. Significant bootstrap values at internal tree node (>70%) indicates the monophyly of a group of sequences.

The NJ method is a distance method and derives subsequent trees from pairwise distance matrices. This method is therefore not appropriate to infer ancestral sequences from contemporaneous sequence data. To further investigate sequence evolution within the the generated datasets, a discrete characters method - maximum likelihood _- was applied to enable estimation of ancestral sequences, as implemented in PAUP version 4.0b10 [334] using the best-fit substitution model for the data calculated in Modeltest version 3.7 [335]. ML methods choose amongst competing hypotheses by choosing the one which maximises the likelihood: the hypothesis which is most likely to have generated the observed data, given the chosen model of sequence evolution. ML requires an explicit model of evolution to compute a tree which is estimated directly from the sequence data. Modeltest is a programme that chooses the best fit of 56 models, applying different selection frameworks: Akaike information criterion (AIC), hierarchical likelihood ratio tests (hLRTs), and Bayesian information criterion [336, 335]. Given a model of substitution PAUP computes likelihood of observed nucleotides in each site given all possible combinations of ancestral states. This method allows the construction of high-resolution phylogenies and estimation of ancestral nucleotide state at internal nodes but is computationally demanding.

Substitutions occurring across viral populations were visualised using the HighlighterTool: http://hcv.lanl.gov/content/sequence/HIGHLIGHT/highlighter.html. Consensus sequences were generated using the Consensus Maker tool: http://hcv.lanl.gov/content/sequence/CONSENSUS/consensus.html.

2.10. Cloning of PCR products into pcDNA3.1

2.10.1. Directional TOPO cloning

To perform protein expression HCV E1E2 amplification products, as well as recombinant PCR products, were ligated into a pcDNA3.1 V5/His© TOPO® expression vector (Invitrogen)**.** The vector is supplied in linearised form with single 3' thymidine overhang and is covalently bound topoisomerase I of *Vaccinia* virus. Forward primers for both sets of reactions were designed to contain directional cloning signal- CACC at the 5' end and fragments for cloning were amplified with polymerase creating blunt ends (Phusion HF polymerase, Finnzymes). The TOPO cloning technique eliminates the need for post PCR procedures. Directional TOPO cloning occurs at highest rate when molecular ratio of insert: vector is between 0.5:1 – 2:1. To perform cloning with high efficiency amounts of insert were calculated to closely match a vector/insert ratio of 1:1. Cloning was performed according to manufacturer's instructions but based on half reaction volumes. After 30 minutes of incubation at 22°C, 3-6µl of ligation mix were added to 50 µl of competent TOP10 *E.coli* cells on ice. Transformation and screening steps are described in sections 2.12 and 2.13.

2.10.2. Restriction enzyme cloning

Vector pcDNA3.1 V5/His D-TOPO has been modified in our group by TOPO cloning to obtain circularised version without insert.

For cloning of E1E2 sequences KpnI and HindIII recognition sites were chosen. Primers containing restriction sites were designed (142-KpnI and 179-HindIII see Table 2.4). PCR products of E1E2 amplification from cDNA were used as templates in additional round of amplification (as described before 2nd round nested PCR in 2.4) using primers containing restriction sequences. Products were analysed by gel electrophoresis afterwards column purified and quantitated as previously described.

Vector in total amounts of 1-3 ng was digested in buffer E (Promega). 20 units of HindIII, 40 units of Acc65I and 1 unit of BSA were used in the reaction in a total volume of 50-100 µl. Similarly 1 µg of insert per reaction was digested in buffer E with 10 units of each restriction enzyme and 1 unit of BSA in a total volume of 20 µl. After overnight incubation at 37°C the reaction was heat inactivated at 80°C for 20 minutes. Digested products were purified and quantitated as previously described.

Cloning reactions were assembled at a vector: insert molar ratio of between 1:3-1:4, with a total DNA concentration 50-100 ng in a total volume of 10 µl. Ligation was performed using 20 units of T4 ligase [21] in a buffer provided (50mM Tris-HCL, 10mM $MgCl_2$, 10mM DTT, 1mM ATP, pH 7.5) with additional 1mM ATP. Overnight incubation at 18°C was followed by heat inactivation at 65°C for 15 minutes.

Cloned plasmids were transformed into *E.coli* cells as described in section 2.12. Transformed cells were seeded on agar plates with ampicillin and after overnight

incubation colonies were screened for presence of the right size insert (described in 2.13).

Screening products were analysed via gel electophoresis and colonies containing the right size insert were grown overnight in Luria Broth to amplify desired plasmid. Column purified plasmids were sequenced (as mentioned further in section 2.15).

2.11. Preparation of TOP10 competent cells

Overnight culture of TOP10 *E.coli* cells were refreshed by 1:100 in fresh Luria Broth dilution and incubated in shaker at 37°C until OD600= 0.3- 0.4. Cultures were then centrifuged for 10 min at 4000 g at 4°C. Bacterial pellets were resuspended (gently, keeping cells on ice) in cold 0.1 M CaCl2 (10ml for every 100ml of LB of fresh culture). After 1 hour incubation on ice, cells were spun as previously. The pellet resuspended in cold 0.1 M CaCl2 + 5% glycerol (2ml/ 100LB) and aliquots of 100 µl were frozen down and stored at -80°C.

2.12. Transformation of TOP10 competent cells

Transformation was performed on 50 µl aliquots of TOP 10 competent cells by adding 1-5 µl of plasmid depending on its concentration. The mix was incubated on ice for 30 minutes. Afterwards heat shock was performed at 42°C for 30-45 seconds followed by brief incubation on ice. 250 µl of LB was added and culture was incubated at 37°C with shaking at 225rpm for one hour without antibiotics. To select bacteria that accepted plasmid, culture was seeded on LB agar plates with ampicillin (100 µg/ ml). After overnight incubation at 37°C colonies from the plate were picked for screening.

2.13. Screening of bacterial colonies

To verify that correct insert was cloned into the vector. Single colonies were picked with sterile pipette tip and dipped into PCR master mix. The PCR reactions were set up in 12.5 µl volumes containing 5 pmol primer T7 and BGH (Table 2.5.), 0.2 mM dNTPs, 0.3 U of Hot Start *Taq* polymerase (Quiagen) and 1× polymerase buffer. Thermoblock cycling parameters were: initial denaturation at 95°C for 15 minutes followed by 30 cycles of 94°C for 45 seconds, 50°C for 45 seconds and 72°C for 4 minutes. A final extension at 72°C for 10 minutes was then performed.

Products of screening were analysed on 2% agarose gel. Colonies containing clones with correct size insert were further picked for overnight culture in 5ml of LB with ampicillin to obtain 30-50 µl or 150 ml LB with ampicillin to obtain 1000 µl of plasmid preparations.

2.14. Plasmid purification from 1-5 ml overnight culture- Minipreps

Plasmids were purified on a silica membrane based method using QIAprep® Miniprep kit (QIAGEN). Overnight cultures of target bacteria were centrifuged at 6000 × g for 5 minutes. Pellets were then resuspended in 250 µl of chilled buffer P1 (containing RNase A at 0.1 mg/ml). Then 250 µl of buffer P2 containing sodium hydroxide was added. Lysis was neutralised with 350 µl of buffer N3 containing acetic acid and guanidine hydrochloride. Samples were then centrifuged at 17900 × g for 10 minutes and clarified supernatants were applied onto QIAprep columns. Solutions were drawn through columns under vacuum pressure. Columns were then washed with 0.5 ml of buffer PB

which removes endonucleases. Further washes with 0.75 ml of buffer PE containing ethanol removes salts. Additional centrifugation ensured removal of ethanol carryover. Plasmid DNA was eluted in 30-50 µl of EB buffer and kept in -20˚C for storage.

2.15. Sequencing of clones

To confirm that cloning was successful sequence analysis of the inserted fragment was performed. For full length E1E2 sequencing three primers were used. Sequencing with primer T7 (see Table 2.5) reveals 5' end of insert sequence and a fragment of plasmid sequence flanking this end. Similarly sequences derived from the primer BGH (see Table 2.5.) cover 3' end of insert and downstream fragment of plasmid. To obtain full length contig third primer was applied- E1IS (see Table 2.5).

Sequencing conditions were as described above (chapter 2.7). Amount of plasmid template used was 100-200ng.

2.16. Plasmid purification from 50-150 ml overnight culture Midipreps

In order to obtain sufficient volumes of HCV E1E2 containing plasmids for mammalian cell transfections, larger volumes of bacterial culture were used. Bacteria were harvested from overnight culture by centrifugation at 5000 × g for 10 minutes. Plasmids were purified using GenElute™ HP High Performance Plasmid Midiprep Kit which features a filter syringe for clearing of bacterial lysate and a silica column which binds nucleic acid. Pellet was resuspended in 4 ml of Resuspension RNase solution. Lysis was performed by combining the sample with 4 ml of lysis buffer for 3-5 minutes

and then neutralised with 4 ml of neutralisation solution. A further 3 ml of binding solution was added and the mix was immediately applied to the barrel of filter syringe. Leaving the sample in an upright syringe barrel for 5 minutes allowed separation of cell debris from lysates. After 4 ml Column preparation solution was drawn through columns under vacuum pressure. Samples were filtered through syringes onto columns. A vacuum was applied until lysates passed through columns. Two washes were then performed- first with 4 ml of wash solution one and second with wash solution 2 which contained ethanol. Columns were dried by centrifugation at 6000 × g for 5 minutes to allow removal of ethanol. Plasmid DNA was eluted in 1 ml of elution solution and stored at -20˚C.

2.17. Site directed mutagenesis

Cloned sequences were checked and aligned with original end-point sequence to locate polymerase or cloning induced mutations. Any differences from endpoint amplicon sequences were back mutated via site-directed mutagenesis. Primers for mutagenesis (see Table 2.6) were designed online: http://bioinformatics.org/primerx/cgi-bin/DNA_1.cgi.

Reactions were set up in 50 µl of 1×Phusion HF buffer (Finnzymes) with 3% DMSO, 0.2mM of dNTPs, 0.5 U of Phusion HF polymerase and 125 ng of each mutagenesis primer. Fifty ng of plasmid that required back mutation was used as a template. Cycling contitions were: 98ºC for 2 minutes, then 30 cycles of 98ºC for 10 seconds, 55ºC for 15 seconds and 72ºC for 8 minutes.

Digestion of original template was carried out with 20U of DpnI [21] in 1× concentrated buffer. DpnI is a restriction enzyme that cleaves methylated DNA; hence whilst original

template plasmid undergoes digestion, newly synthesised mutated product remains. After overnight incubation at 37ºC samples were resolved on an agarose gel to confirm presence of the plasmid DNA. Bacteria were transformed and plasmids sequenced as previously described.

Table 2.6. List of mutagenesis primers.

Primer name/ mutated position	Template	Direction	Sequence	Nucleotide position[1]
C1358T_For	lig1A20.1_1_c2	Sense	5' CATTATCCTTGTACTGTCAACTACACCATAT 3'	2190 → 2220
C1358T_Rev		Antisense	5' ATATGGTGTAGTTGACAGTACAAGGATAATG 3'	2190 ← 2220
T943C_For	lig1A30.3_26_c15	Sense	5' CCGACCAACGCCCCTACTGCTGGCACTA 3'	1780 → 1807
T943C_Rev		Antisense	5' TAGTGCCAGCAGTAGGGGCGTTGGTCGG 3'	1780 ← 1807
G1210A_For	lig1A30.3_38_c23	Sense	5' CCCTTGTGCCATCGGAAGGGTGGGCAACAAAAC 3'	2042 → 2074
G1210A_Rev		Antisense	5' GTTTTGTTGCCCACCCTTCCGATGGCACAAGGG 3'	2042 ← 2074
C1183T_For	lig1A57.2_6_c28	Sense	5' GGATTCACCAAGGTGTGCGGAGCGCCTCCTTG 3'	2016 → 2047
C1183T_Rev		Antisense	5' CAAGGAGGCGCTCCGCACACCTTGGTGAATCC 3'	2016 ← 2047

[1] primer position numbered according to H77 genome coordinates; all primers designed by the author

2.18. Mammalian cell cultures

Three different strains of mammalian cells were used for phenotypic analysis of proteins of interest: HEK 293 T, HEK 293 FT [337], Huh7 [338].

HEK 293T cells (Human Embryo Kidney cells) and Huh7 cells (Human Hepatocellular carcinoma cells) were cultured in DMEM (Gibco) with 10% FBS (bioSera), 5 ml of antibiotic/antimycotic solution (Hyclone) and 5 ml of non-essential amino acids (Gibco).

HEK 293FT cells were cultured in DMEM media with additives as above and additional 5ml of geneticin G418 (100mg/ml, Autogen Bioclear) for selection of plasmid bearing SV40 T antigen.

All cells were grown in T-flasks at 37ºC and 0.5% CO_2. Passages were performed after cells reached monolayer confluency of 80-90%. After washing with PBS, cells were trypsinised for 2-5 minutes. Trypsin was neutralised by addition of DMEM with FBS and cells were seeded in fresh media at a dilution 1:12 for HEK cells and 1:6 for Huh7 cells.

Cell stocks were stored in liquid nitrogen in growth medium with 10%FBS and 10% DMSO at around 1×10^7 cells per 1 ml.

Table 2.7. Components of mammalian cell culture media additives.

Components of media additives	Concentration
Antibiotic/antimycotic solution	
Penicillin G	10,000U/mL
Streptomycin	10,000µg/mL
Amphotericin B (Fungizone)	25µg/mL
Non-essential amino acids solution	
Glycine	750 mg/L
L-Alanine	890 mg/L
L-Asparagine	1320 mg/L
L-Aspartic acid	1330 mg/L
L-Glutamic Acid	1470 mg/L
L-Proline	1150 mg/L
L-Serine	1050 mg/L

2.19. Expression of proteins in mammalian cells

For efficient production of HCV E1E2 glycoproteins, 293FT cells were transfected with plasmids bearing target insert sequences. Cells were seeded in 10 cm dishes 24 hours before transfection in amounts allowing 40-60% confluency on the following day. Transfections were performed with PEI transfection reagent (Exgen 500- Fermentas). Twenty-four µl of PEI was diluted in OptiMEM (Gibco) to a total volume of 300 µl. Solution was then mixed with 6 µg of plasmid diluted in total 300 µl of OptiMEM and incubated at room temperature for 45 minutes. For the transfection event, cell media was changed from DMEM to Opti-MEM and subsequently DNA-PEI mix was added. After overnight incubation media was changed back to DMEM. 72h post-transfection, cells were harvested in 1 ml cell lysis buffer (150 mM NaCl, 1% NP-40, 50 mM Tris pH 7.4, 1 mM EDTA).

2.20. Polyacrylamide gel electrophoresis of proteins

Quality of expressed proteins was analysed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [329, 339], followed by visualisation using Western Blotting [340].

Ten % polyacrylamide gels were cast in a slab kit [341]. The resolving gel consisted of 6 ml of deionised water, 5ml of 30% acrylamide: bisacrylamide solution, 3.8 ml of 1.5 M Tris pH 8.8, 75 µl of 20% sodium dodecyl sulphate (SDS), 100 µl of 30% ammonium persulphate (APS) and 22.5 µl of TEMED.

Once polymerised, the resolving gel was overlayed with 5% stacking gel made with 3.4 ml of deionised water, 0.83 ml of 30% acrylamide: bisacrylamide solution, 0.63 ml of 1 M Tris pH 6.8, 25 µl of 20% SDS, 17 µl of 30% APS and 10 µl of TEMED. A comb was inserted between slabs to form sample wells.

The samples for analysis were prepared by mixing equal volumes of protein solution and gel loading buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.002% bromophenol blue in water). Samples were then heated at 95ºC for 5-10 minutes to allow denaturation and coating with SDS, and briefly centrifuged before adding 20-30 µl onto a gel. Proteins were resolved alongside molecular weight markers (Spectra Bright full range, Fermentas) in tris-glycine buffer (25mM Tris, 250 mM glycine, 0.1% SDS, pH 8.5) at 150V for around 90 minutes.

Resolved proteins were either visualised by direct Coomassie staining or transferred onto a nitrocellulose membrane for further Western Blotting analysis.

For staining, gel was transferred into a Coomassie dye (2.5g Brilliant Blue R250, 450 ml water, 450 ml methanol, 100 ml glacial acetic acid) for 1 hour. Destaining was performed by boiling the gel in clean deionised water until protein bands were clearly visible.

2.21. Western Blotting analysis of proteins

After electophoresis, proteins were transferred from the gel onto a sensitive ECL membrane (GE Healthcare) in a semidry blotter (BioRad). Whatman paper and nitrocellulose membrane were soaked in transfer buffer (48 mM Tris, 39 mM glycine, 20% methanol, 1.3 mM SDS) and the transfer apparatus was assembled. The gel was facing the negatively charged plate and the membrane facing the positively charged plate in order to allow protein migration. Transfer was carried out at 1 mA/cm^2 for around 90 minutes. Efficiency of the process was assessed by eye examination of the amount of prestained weight marker that migrated to the membrane.

Following electrotransfer, to prevent interactions between the membrane and the antibody used for detection of the target protein, ECL membrane was soaked in blocking solution (5% milk powder in 0.05% PBS-Tween) for one hour at room temperature or overnight at 4 ºC. Membrane was washed between each stage 3 times for 5 minutes in PBST. For detection of proteins different primary antibodies were used (see Table 2.8). Incubation was carried out for 1 hour at room temperature or overnight at 4ºC. After washing membrane was probed with secondary anti-mouse IgG1 HRP conjugate (DAKO) at 1:1000 dilution in 10 ml of PBST for 1 hour. Following final washing membrane was immersed in ECL detecting reagent (2.5 ml solution 1+ 2.5 ml solution 2, in house method), placed between acetate sheets and exposed onto Kodak BioMax Light film. Solution 1 was made up in deionised water to a total volume

of 100 ml containing 1 ml of 250 mM Luminol in DMSO, 0.44 ml of 90mM Coumaric acid in DMSO and 10 ml of 1 M Tris pH 8.5. One hundred ml of solution 2 consisted of 64 µl of 30% H_2O_2 and 10 ml of 1M Tris pH 8.5 in deionised water.

Table 2.8. List of primary antibodies used in Western Blotting analysis.

Antibody	Protein detected	Concentration	Reference
AP33	HCV E2: soluble, bound in heterodimers, incorporated on pseudoparticles	hybridoma supernatant diluted 1:400 in 10 ml of PBST	[342], kindly provided by Arvind Patel
R187	MLV core	hybridoma supernatant diluted 1:200 in 10 ml of PBST	ATCC CRL1912, kindly provided by Jean Dubuisson

2.22. Generation of HCV pseudotyped particles

Murine leukaemia virus (MLV) - HCV pseudotypes were generated by co-transfection of HEK 293T cells with 3 plasmids: phCMV-5349 containing the MLV polyprotein Gag-Pol, pTG-Luc126 encoding the reporter gene of firefly luciferase and pcDNA3.1 E1E2, a commercially avaiable expression vector with inserts derived from target HCV E1E2 sequences (system first described in (Bartosch, 2003 #14)). Co-transfections were performed in 10 cm dishes with 40-60% confluent 293T cells using PEI transfection reagent (Exgen 500- Fermentas). Twenty four µl of PEI diluted in phyH_2O to a total volume of 300 µl was mixed with 300 µl of plasmid mix (2 µg of each) and incubated in room temperature for 45 minutes. For the transfection event, cell media was changed from DMEM to Opti-MEM and subsequently DNA-PEI mix was added. After overnight incubation media was changed back to DMEM. 72h post-transfection media containing pseudoparticles was collected and filtered through 0.45 µm filter.

2.23. Infectivity assay

Huh 7 cells were seeded in 48 cell clusters, in a concentration 40 000 cells per well in 0.5ml. Growth media was collected and 300 µl media containing pseudoparticles was added along with 200 µl of fresh DMEM. Media and infected cells were collected after 72h. Cell lysis was performed in 50 µl lysis buffer (Promega) and luciferase assay performed according to manufacturer's protocol using 50 µl firefly luciferase substrate (Promega). Relative light units emission (560nm) was read using a luminometer (Fluorostar Optima, BMG Labtech).

2.24. Production and purification of CD81-LEL

In order to perform receptor binding assays of target glycoproteins, large extracellular loop [6] of CD81 in fusion with GST was produced in bacteria. Overnight culture of *E. coli* strain BL21 bearing plasmid with GST-CD81-LEL sequence in LB with ampicillin was refreshed 1:100 and cultured until OD_{600}=0.8. Expression of the protein was then induced by adding IPTG (0.5mM). After further 4 h of incubation cells were centrifuged at 4000 × g for 10 minutes at 4ºC. Pellets from 1000 ml of culture were resuspended in 30 ml of lysis buffer (0.1% Triton, 0.5mM DTT with protease inhibitor cocktail (Roche) in PBS) by vortexing and frozen at -20ºC. Lysates were placed in a water bath set at 25ºC and, after complete thawing, were incubated for a further 25-30 minutes. Freezing- thawing cycle (as above) was repeated and then solutions were centrifuged at 15000 × g for 40 minutes at 4 ºC.

Supernatants were added to 1.5 ml of glutathione agarose slurry obtained by swelling 0.07g of beads (Agarose 4B, Sigma Aldrich) overnight in water. Protein was allowed to bind to the beads overnight; incubation was carried out at 4 ºC with rotation. The

solution was then purified by sequential steps of washing and centrifugation at 750 × g for 2 minutes. Firstly, two wash steps were performed with previously used lysis buffer without protease inhibitors. Second, washing solution containing high salt (20 mM Tris-HCl pH 7.5, 300 mM NaCl) was applied twice. Lastly, two steps of washing were performed with low salt solution (20 mM Tris-HCl pH 8.0). The protein was harvested in solution with high glutathione concentration (0.1 M Tris-HCl pH 8.0, 10 mM glutathione) that out-competed glutathione immobilised on beads.

Quality of purified protein was assessed by SDS-PAGE (described in 2.20) followed by Coomassie staining and Western Blot analysis (see chapter 2.21). Samples were quantified in Pierce BCA protein assay.

2.25. Bicinchoninic acid (BCA) protein assay

Quantitation of total protein in samples was performed in a colorimetric reaction based on bicinchoninic acid [343].

A commercially available kit (Thermo Scientific) was used for BCA quantitation. Ten µl of serial diluted samples were transferred to a microtiter plate in parallel to Bovine Serum Albumin (BSA) standard dilutions. Ninety µl of developing solution (5ml BCA, 100 µl copper (II) sulphate solution) were added. Plates were wrapped in cling film to avoid evaporation and incubated at 37ºC for 1 hour. Absorbance was read in spectrophotometer at 550 nm.

Standard curve for BSA dilutions was generated in Prism and concentrations of protein samples were calculated.

2.26. ELISA

Enzyme Linked ImmunoSorbent Assays [214] [344] were used to characterise antibody binding to target E1E2 glycoproteins.

2.26.1. GNA capture ELISA

96-well Immulon (Thermo) plates were coated with the Galanthus Nivalis Agglutinin (GNA) at 5µg/ml in 50 µl/well carbonate-bicarbonate buffer (schematic design in Figure 2.4). Incubation was carried out at room temperature for 2 hours or overnight at 4ºC. Excess GNA was washed from the wells 3 times with 250 µl of PBST. Afterwards, to prevent interactions between the plate wells and the antibody, a solution of 5% milk powder in 0.05% PBST was added for 1 hour at room temperature or 4ºC overnight. Tested proteins were then captured from cell lystates overnight at 4ºC. Lysates containing proteins were purified prior to use by centrifugation at 20000 × g for 30 minutes at 4ºC. Unbound proteins were washed off as previously described.

Detection of proteins was performed with different antibodies (see Table 2.9) at varying concentrations to obtain titration curves. Secondary antibodies used were conjugated with horse radish peroxidase (HRP) and colour reactions were developed with 100 µl TMB (Sigma) reagent. Spectrophotometric readouts were performed in dynamic conditions at 620nm or after stopping the reaction with 100 µl of 3M H_sSO_4 at 450nm.

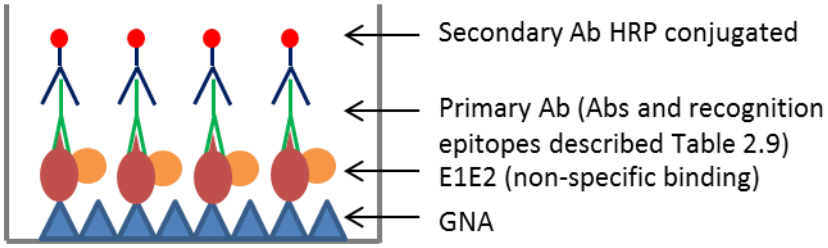


Figure 2.4. Schematic representation of GNA capture ELISA.

Table 2.9. List of primary antibodies used in GNA capture ELISA.

Antibody	Origin	Epitope/ aa residues[1]	CD81 binding competition	HCVpp neutralisation	Reference
AP33	Mouse	412-424	+	+	[342], kindly provided by Arvind Patel
ALP98	Mouse	640-654	-	+	[342], kindly provided by Arvind Patel
1:7	Human	525, 530, 535	+	+	[345] Kindly provided by Matts Persson
ML1	Human	424, 525, 530, 535	+	+	[169] *described as AR3A, Kindly provided by Mansun Law
ML2	Human	Discontinuous	+	+	[169] *described as AR3B, Kindly provided by Mansun Law
DB1	Human	540, 549	-	-	Kindly provided by Mansun Law

[1] amino acid position numbered according to H77 polyprotein coordinates

2.26.2. CD81 capture ELISA

ELISA was also applied to investigate glycoprotein binding to the putative HCV receptor CD81 (schematic design in Figure 2.5). Fifty µl of GST-CD81-LEL (10 µg/ml in carbonate-bicarbonate buffer) was incubated at 4ºC overnight in wells of Immulon microtitre plate. Excess of CD81 was removed and blocking performed as described above. Similarly prepared lysates were captured by CD81 in wells at 4ºC overnight. To reduce background, primary antibody ALP98 was diluted in 5% milk in PBST and hybridised to proteins for 2 hours at room temperature. After washing, probing was

carried out with anti-mouse antibodies conjugated with alkaline phosphatase (AP) (Sigma) at 1:1000 dilution in PBST.

Detection of E1E2 capture by CD81 with antibody ALP98 results in very low signals. To overcome this obstacle AMPAK ELISA detection system (Invitrogen) was used.

Excess of alkaline phosphatase was removed by washing the plate 4 times with 250 µl TBS. Positive control was applied according to manufacturer's protocol. 50 µl of freshly reconstituted AMPAK substrate were carefully sampled to wells. After 15 minutes of incubation at room temperature, reconstituted amplifier was added in the same manner. Reactions were incubated until sufficient colour intensity was observed and then stopped with 50 µl of 0.3M H_2SO_4. Absorbance was read at 495 nm.

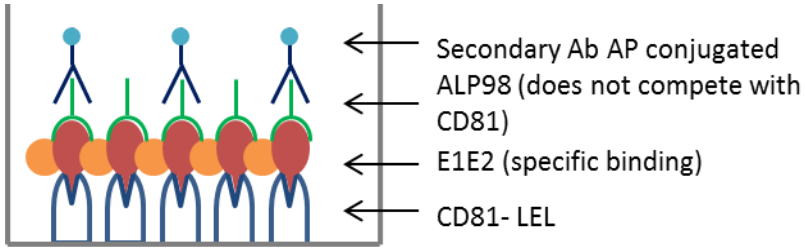


Figure 2.5. Schematic representation of CD81 capture ELISA.

3. Transmission of HCV in the xenomouse model

3.1. Results

In natural infection, identifying the genetic and phenotypic determinants which facilitate HCV transmission is challenging as most cases of acute HCV infection are asymptomatic. Many previous studies of acute HCV infection are missing samples from donor populations, which precludes assessment of the dynamics of viral populations at transmission. Studies performed in experimentally infected chimpanzees have multiple drawbacks and limitations. In this study we utilised immunocompromised mice with human liver grafts to allow assessment of HCV envelope glycoprotein evolution at the transmission step of the viral life cycle, defining viral populations in both the donor inoculum and recipient chimeric mice.

3.1.1. Amplification of full-length E1E2 sequences

A cohort of eleven Alb-uPA/SCID mice transplanted with human hepatocytes from different donors were inoculated with HCV infected serum from a single source. Donor sera (KP) was obtained from an HCV+ patient in the chronic phase of infection, with time since initial infection estimated at more than 5 years. Mice were bled at different time-points post infection (see Table 2.1 in 2) and HCV RNA was extracted. HCV RNA was also extracted from the donor serum.

To accurately assess the changes in E1E2 population dynamics occurring during our *in vivo* HCV transmission events, a single genome amplification (SGA) approach was employed, followed by direct sequencing. Full-length E1E2 genes from sequential

serum samples (2 time points), derived from 4 mice were amplified. Full-length E1E2 sequences were directly sequenced and aligned.

In total we obtained 36 end-point titrated full length E1E2 sequences from KP inoculum and 136 sequences derived from four mice (24 to 44 for each mouse, median 35 and 11 to 22 sequences for each time point, median 18). A total of 10 bulk amplified consensus E1E2 sequences derived from all 10 experimentally infected mice were also obtained for control purposes (see Table 3.1).

Table 3.1. Summary of mice and donor data for sequences amplified by bulk and end-point titration methods.

Patient/ Mouse ID	Population sampling	Serum HCV (IU/ml)	Hepatocyte donor	Time post-inoculation (days)	Number of SGA amplicons
KP	KP	1×10^6	-	-	36
A594	594_1	8.09×10^5	II	32	13 + Bulk
	594_2	1.01×10^6	II	59	17
N666	666_1	2.69×10^6	I	14	22
	666_2	2.49×10^6	I	28	20
N714	714_1	1.16×10^6	I	14	21
	714_2	1.54×10^6	I	28	19
A931	931_1	8.61×10^7	III	21	11
	931_2	2.94×10^7	III	42	13
	A931	8.04×10^7	III	28	NA, Bulk
A583	A583	8.40×10^5	II	32	NA, Bulk
A596	A596	1.73×10^5	II	32	NA, Bulk
A585	A585	2.99×10^5	II	32	NA, Bulk
A587	A587	9.79×10^6	II	32	NA, Bulk
A902	A902	4.46×10^7	III	28	NA, Bulk
A909	A909	1.96×10^7	III	28	NA, Bulk
A965	A965	7.73×10^6	III	28	NA, Bulk

I = 36 year old female, II = 45 year old female 92% viability polycystic tissue, III = 4 year old male 92% viability normal tissue

3.1.2. Assessment of intra and inter population genetic distances and diversity

To determine whether there were significant differences in the genetic diversity apparent in pre and post transmission viral populations, genetic distances were calculated using MEGA 5.0 [346]. Genetic distance was calculated individually for KP inoculum derived sequences and compared to genetic distance from post transmission sequences at two time points (only SGA derived sequences were included in calculations). These analyses show that pairwise genetic distance within sequence populations rises significantly after transmission (see Figure 3.1 and Table 3.2). When post transmission viral populations were analysed according to sampling points, a significant rise in genetic diversity was observed for all comparisons with the exception of 594 time point 2, when compared to KP source. When genetic distance was calculated for total post transmission populations (both time points combined) all apart from mouse 594 showed significant rise as compared to KP (see Figure 3.2 and Table 3.3). Together these data demonstrate an increase in genetic diversity post transmission in Alb-uPA/SCID chimeric mouse model.

Additionally we also calculated the diversity index for pre and post transmission viral populations (Table 3.4 and Figure 3.3 A). Diversity index is a measure of the number of unique sequences in a population. In a similar fashion to increasing intrapopulation genetic distance, we also observe increasing diversity index post transmission i.e. there are more unique sequences identified after the transmission event than prior to it. Next we calculated the ratio of non-synonymous to synonymous substitutions (dN/dS) in all viral populations using MEGA 5.0 (Table 3.4 and Figure 3.3 B). dN/dS ratio is a measure of positive diversifying selection. A dN/dS ratio <1 indicates purifying selection. A dN/dS ratio equal to 1 indicates genetic drift. And dN/dS ratio >1 indicates positive

selection is acting upon our sequence population. Unlike genetic distance and diversity index the dN/dS ratio does not change significantly between pre and post transmission populations. All dN/dS ratios are significantly below 1 indicating purifying selection as the major evolutionary pressure shaping viral populations pre and post transmission. Together these analyses indicate that viral population complexity increases upon experimental transmission in chimeric mouse model, and purifying selection is the dominant evolutionary process shaping envelope glycoprotein evolution [239].

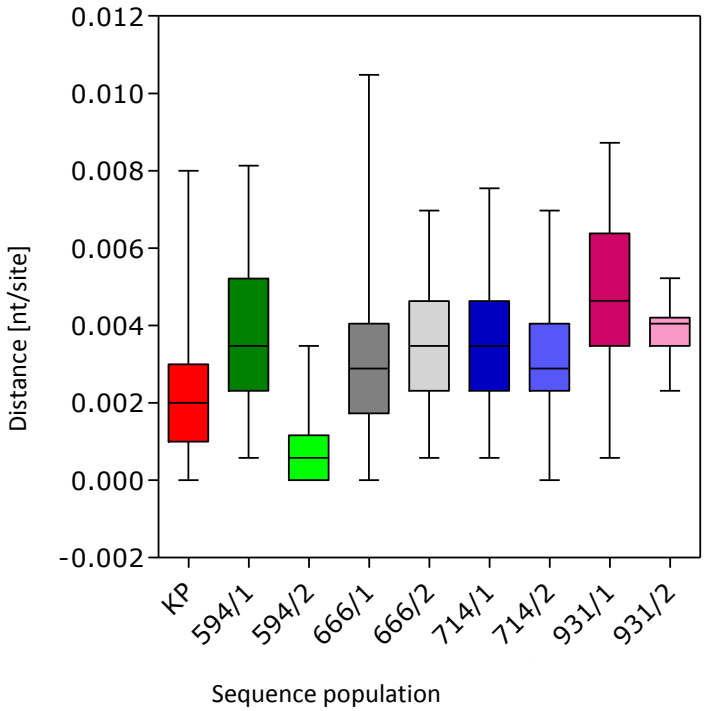


Figure 3.1. Genetic distance (nucleotide substitution per site) values for full length E1E2 sequence populations (derived from SGA) per time-point in each xenomouse host.

Table 3.2. Statistical analysis of differences between pre and post-transmission populations by time-point in each xenomouse host.

	Number of values	Min	Max	Mean	Std. Deviation
KP	625	0	0.008	0.002402	0.001419
594/1	78	0.00058	0.00813	0.003622	0.001592
594/2	136	0	0.00347	0.000943	0.000869
666/1	231	0	0.01048	0.003259	0.00196
666/2	190	0.00058	0.00697	0.003526	0.001336
714/1	210	0.00058	0.00755	0.003554	0.001505
714/2	171	0	0.00697	0.003066	0.001325
931/1	55	0.00058	0.00872	0.004847	0.001921
931/2	78	0.00231	0.00522	0.003813	0.000721

Dunn's Multiple Comparison Test	Difference in rank sum	Significant? P < 0.05?	Summary
KP vs 594/1	-404.0	Yes	***
KP vs 666/1	-246.7	Yes	***
KP vs 714/1	-396.9	Yes	***
KP vs 931/1	-667.3	Yes	***
KP vs 931/2	-551.9	Yes	***

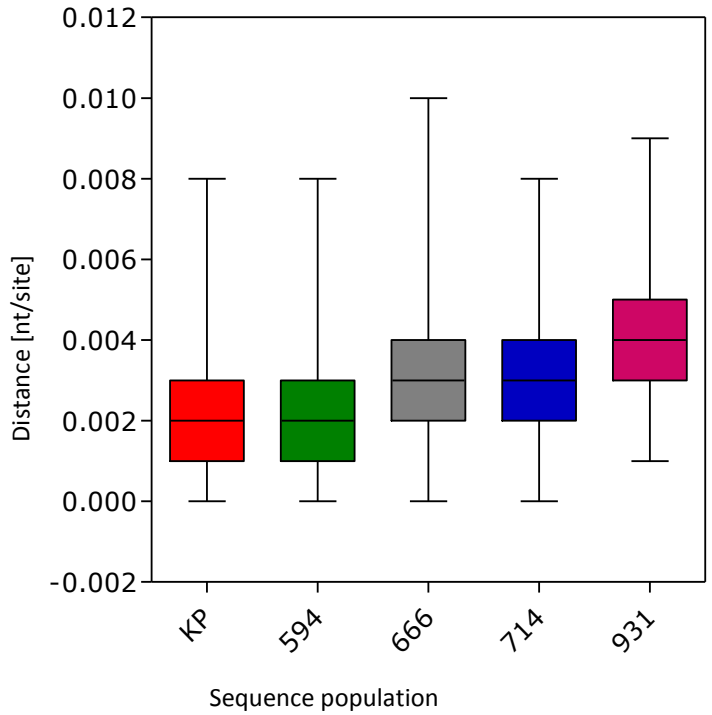


Figure 3.2. Genetic distance (nucleotide substitution per site) values for full length E1E2 sequence populations (derived from SGA) per each xenomouse host (sampling time points combined).

Table 3.3. Statistical analysis of differences between pre and post-transmission populations by each xenomouse host (time points combined).

	Number of values	Min	Max	Mean	Std. Deviation
KP	625	0	0.008	0.002402	0.001419
594	435	0	0.008	0.002147	0.001542
666	861	0	0.01	0.003329	0.001731
714	780	0	0.008	0.003277	0.001483
931	276	0.001	0.009	0.004279	0.001577

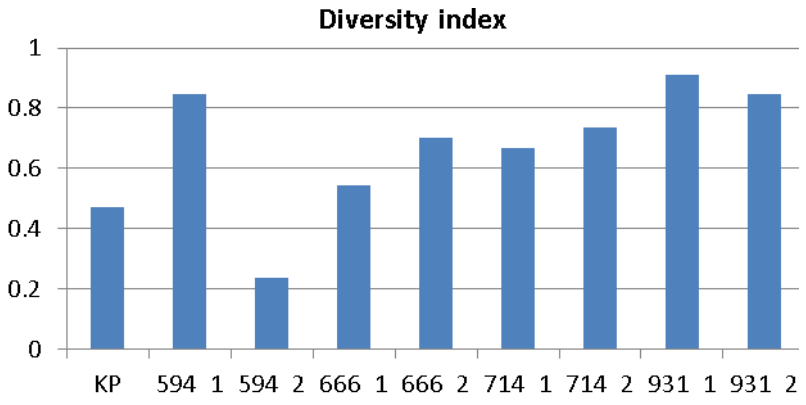
Dunn's Multiple Comparison Test	Difference in rank sum	Significant? P < 0.05?	Summary
KP vs 594	114.2	No	ns
KP vs 666	-499.5	Yes	***
KP vs 714	-520.5	Yes	***
KP vs 931	-990.5	Yes	***

Table 3.4. Diversity index and dN/dS ratio comparison between donor HCV populations circulating in donor and mice hosts.

Patient/ Mouse ID	Sequence population	Number of SGA amplicons	Diversity Index*	Mean dN/dS ratio
KP	KP	36	0.472	0.2177
A594	594_1	13	0.846	0.2420
	594_2	17	0.235	0.06840
N666	666_1	22	0.545	0.2020
	666_2	20	0.700	0.1984
N714	714_1	21	0.667	0.2073
	714_2	19	0.737	0.2640
A931	931_1	11	0.909	0.1793
	931_2	13	0.846	0.2860

* Calculated as number of unique sequences per total number of sequences in given population

A



B

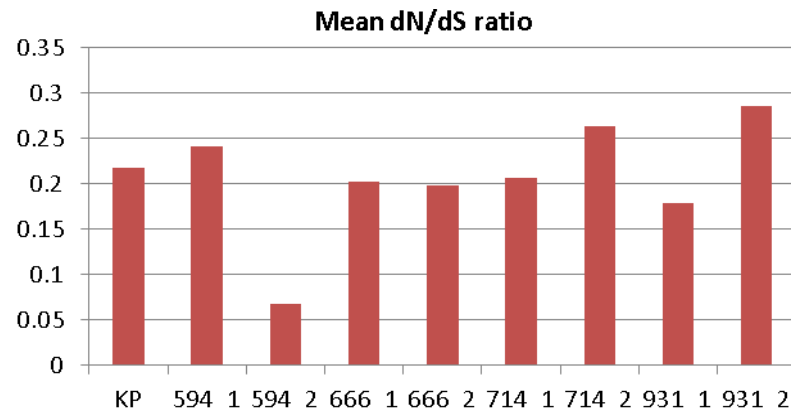


Figure 3.3. Diversity index (A) and dN/dS ratio (B) comparison between HCV populations circulating in donor and mice hosts.

3.1.3. Composite phylogenetic analysis of donor/recipient viral populations

In order to reveal specific patterns of sequence evolution occurring during HCV transmission, a combined alignment containing all 172 SGA derived full length E1E2 sequences from the donor and four recipient mice, was subjected to Maximum Likelihood [330] phylogenetic reconstruction. For control purposes bulk amplified E1E2 sequences derived from inoculum and mice samples were also included. The best fit substitutional model which gave rise to our observed sequence data set was estimated in Model Test v3.7 [335]. The best fit model (HKY+G) was utilised for the combined phylogenetic reconstruction and generated in PAUP under the ML criterion (Figure 3.4).

The composite ML tree in Figure 3.4 depicts phylogenetic relationships between all sequences obtained in the chimeric mouse transmission study. Donor sequences of KP inoculum cluster together in an antecedent clade. Surprisingly, we do not observe monophyletic clustering of post transmission samples on the basis of the host mouse from which they were derived. All post transmission sequences cluster together in a mixed clade. However, interestingly two sequences from the donor inoculum clustered with the mixed recipient clade (KP_7, KP_22).

Star-like phylogenies were constructed for each post transmission sequence population by the Neighbour Joining method (Figure 3.5). The majority of sequences derived from both donor and recipient chimeric mice were unique. However, groups of identical sequences were identified within the composite tree in both the donor and mixed recipient clades (and Figure 3.6). In the mixed post transmission clade identified groups of identical sequences were not necessarily derived from the same chimeric

mouse. Together these data suggest that the post transmission sequences do not cluster on the basis of recipient mouse and are genetically distinct from the donor sequences. While most sequences identified were unique, this data indicates there are genetic determinants, related to composition of incoming sequence population and, associated with transmission and establishment of initial infection in chimeric mouse model. These determinants are further analysed in following subchapters.

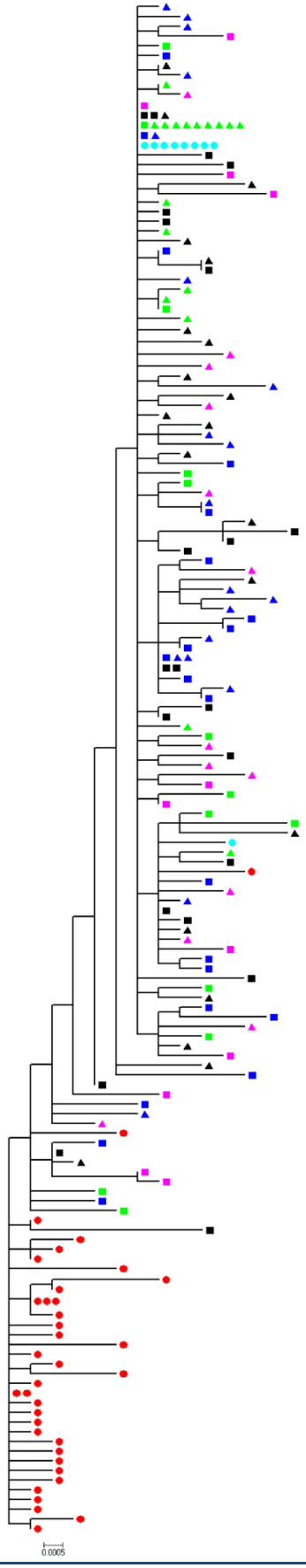
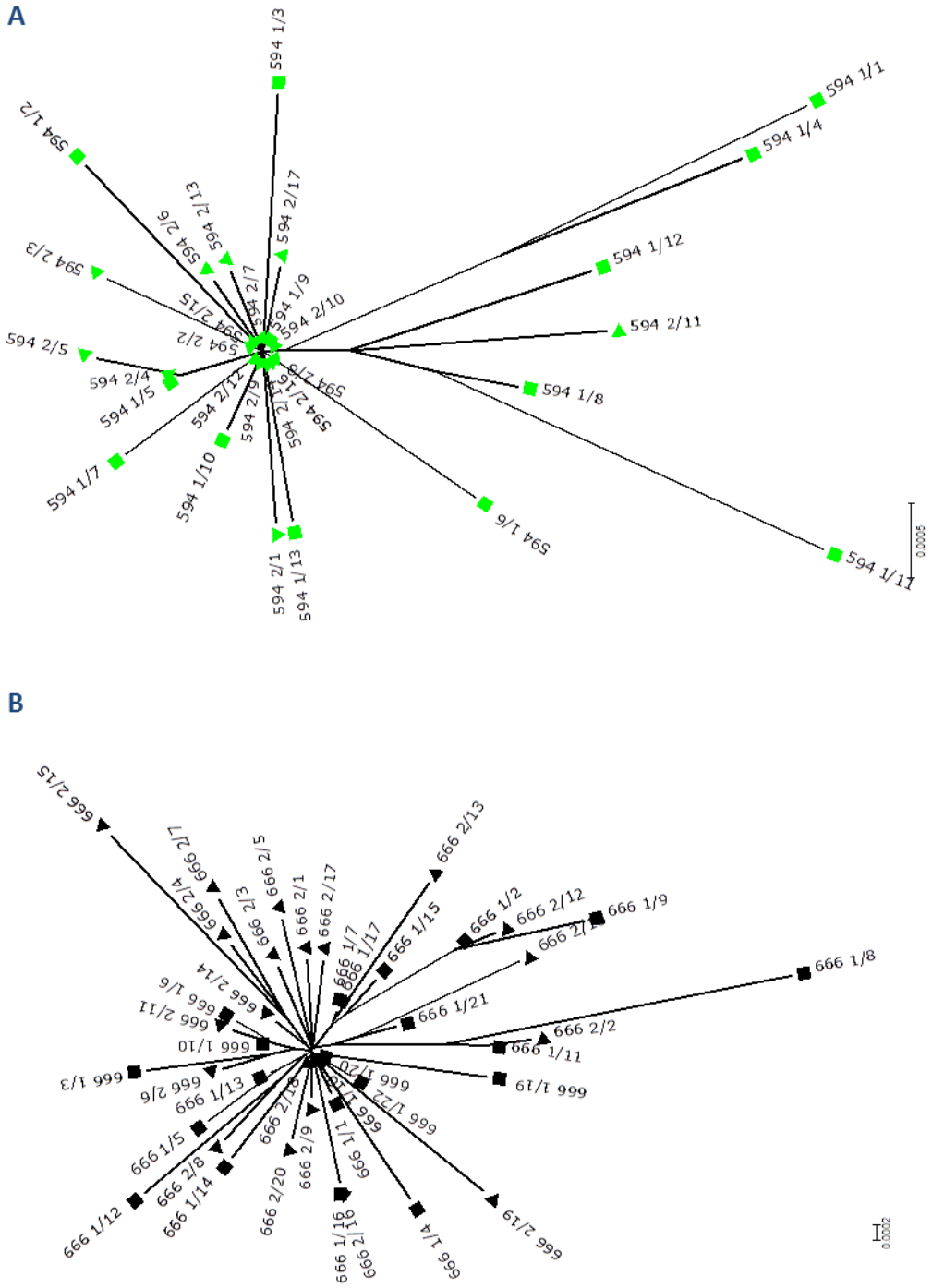


Figure 3.4. Phylogenetic reconstruction of envelope glycoprotein sequence complexity.

Composite mid-point rooted ML (HKY+G) tree of SGA and bulk derived E1E2 sequences from KP inoculum and populations circulating in mice post transmission. Despite lack of contamination sequences do not cluster on the basis of study subject. Key: red circles- KP inoculum, green squares- mouse 594 time point 1, green triangles- 594 tp2, black squares- mouse 666 time point 1, black triangles- 666 tp2, blue squares- mouse 714 time point 1, blue triangles-714 tp2, pink squares- mouse 931 time point 1, pink triangles- 931 tp2, turquoise circles- bulk amplified sequences. The scale bar represents 0.005 nucleotide substitution per site.


A
594 1/3
594 1/1
594 1/2
594 1/4
594 2/13
594 2/17
594 2/6
594 1/12
594 2/3
594 2/7
594 1/9
594 2/15
594 2/10
594 2/11
594 2/2
594 2/5
594 2/4
594 1/8
594 1/5
594 2/12
594 2/9
594 2/16
594 2/8
594 1/7
594 1/10
594 1/6
594 2/1
594 1/13
594 1/11
0.0005
B
666 2/15
666 2/7
666 2/4
666 2/5
666 2/3
666 2/1
666 2/17
666 2/13
666 1/2
666 2/12
666 1/9
666 1/15
666 1/7
666 1/17
666 2/11
666 2/14
666 1/6
666 1/8
666 1/21
666 2/2
666 1/10
666 1/11
666 1/3
666 1/19
666 2/6
666 1/20
666 1/22
666 1/5
666 1/1
666 2/8
666 2/18
666 2/19
666 1/14
666 2/9
666 2/20
666 1/4
666 1/16
666 1/12
0.0002

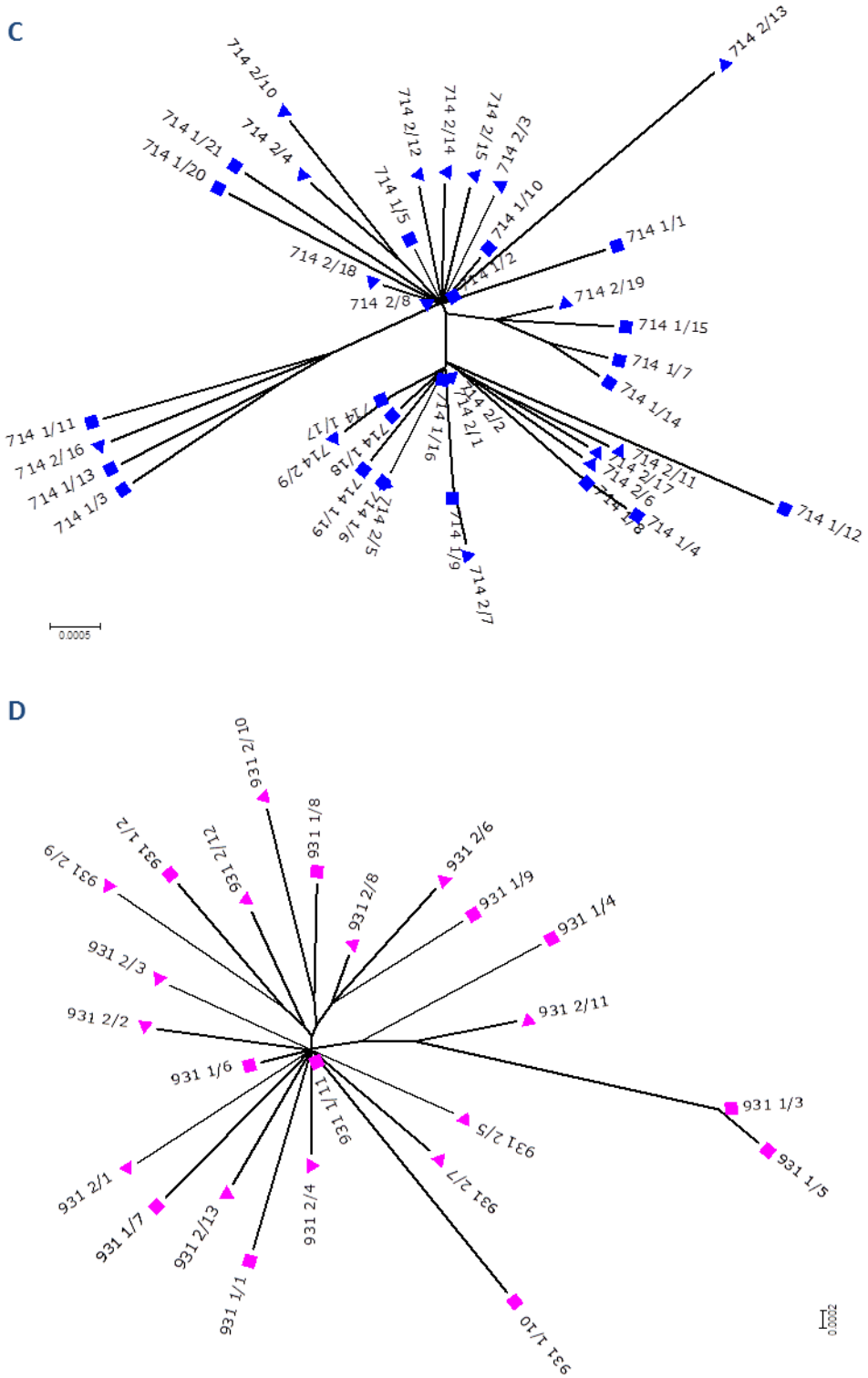


Figure 3.5. Radiation Neighbour Joining trees for each intrahost viral population post-transmission.

Panel A- Mouse 594, panel B- mouse 666, panel C- mouse 714, panel D- mouse 931. Key: Squares- sampling time point 1, triangles- sampling time point 2: green- mouse 594, black- mouse 666, blue- mouse 714, pink- mouse 931.

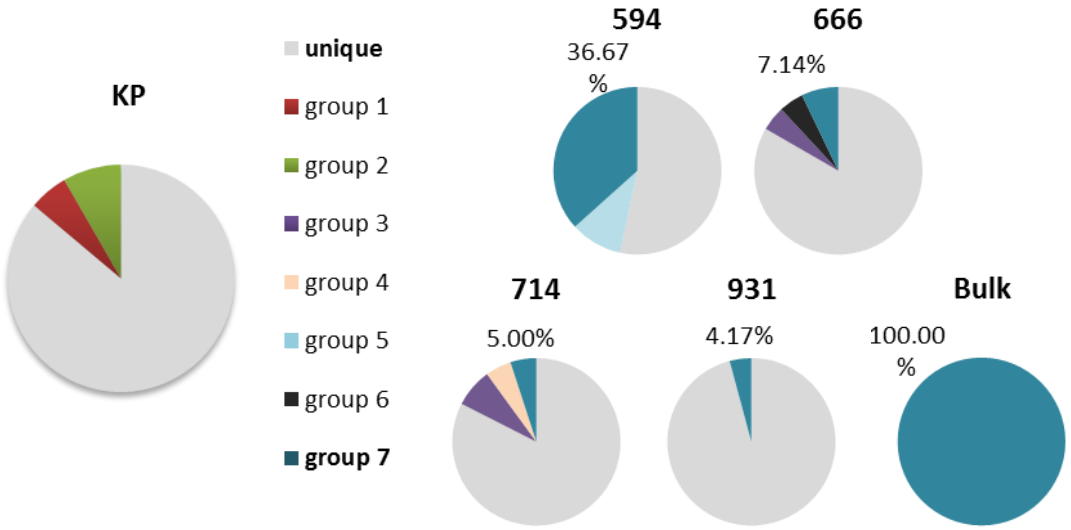


Figure 3.6. Ratio of unique sequences versus groups of identical nucleotide sequences in pre- and post- transmission viral populations.

3.1.4. HVR1 sequence analysis

Previous experimental infections of chimpanzees with HCV indicate the majority HVR1 sequence in the source inoculum is effectively transmitted and becomes the majority circulating variant post transmission. We next wanted to assess whether genetic determinants of effective transmission and establishment of initial infection were located within the HVR1 region. To compare transmission of HCV in uPA/SCID mice to previously reported studies of transmission in chimpanzees and humans [90, 207, 216] we translated HVR1 nucleotide sequences from both donor and recipient uPA/SCID mice into amino acid sequences. Analysis of 172 translated HVR1 amino acid sequences revealed that the major HVR1 variant in the donor population was also the major variant circulating in all inoculated mice (see Figure 3.7).

A

Sequence 1	ETHVTGGATA HGASVLASLL TTGAKQ	**83.14%**
Sequence 2	.....A.	**13.37%**
Sequence 3	....F....	1.16%
Sequence 4	....T...	0.58%
Sequence 5	..Y..... ..... .P....	0.58%
Sequence 6	.... .P....	0.58%
Sequence 7	....F.	0.58%

B

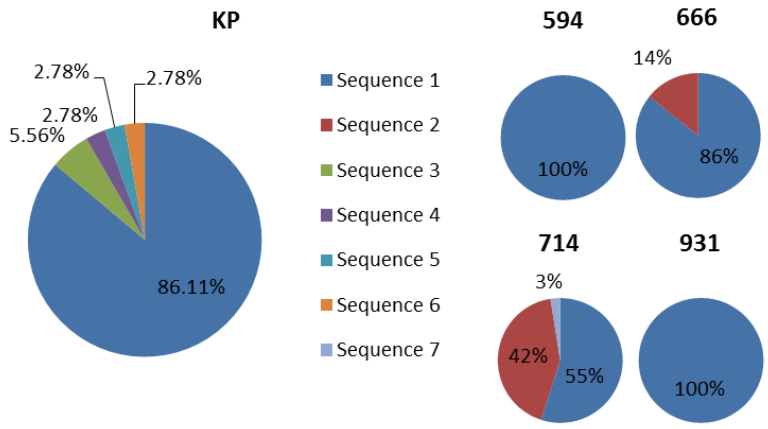


Figure 3.7. Alignment of HVR1 amino acid sequence variants (A) and HVR1 variant frequencies amongst total populations of end-point titrated sequences in the donor and recipient mice (B).

Overall the major HVR1 sequence variant (seq1, Figure 3.7) was present in 83.14% of all analysed sequences. The major HVR1 variant was present in over 86% of donor sequences and constituted 100% of sequences in populations circulating in mouse 594 and 931. None of the minor HVR1 variants present in KP inoculum were detected in any of the inoculated mice. In mouse 666 one new minor variant appeared (seq 2, Figure 3.7). This HVR1 variant also arose in mouse 714 and constituted over 42% of the population. Additionally another minor sequence variant (seq 7) was also detected. Interestingly all new HVR1 variants differed from the majority sequence only by one amino acid. Together these analyses indicate that the HVR1 is relatively stable upon transmission, mirroring the findings of transmission studies in chimpanzees. Importantly these data suggest that determinants of effective transmission and establishment of initial infection reside outside HVR1 and indicate the necessity of using full length E1E2 when studying HCV transmission.

3.1.5. Individual phylogenetic analyses of donor/recipient intrahost populations

Individual donor-recipient transmission events were further analysed using a combination of Highlighter plots coupled with phylogenetic analysis (Figure 3.8). While sequence diversity was limited in all xenomice, mirroring the restricted diversity observed prior to seroconversion in human hosts, the majority of sequences identified were unique. As in the combined analysis, donor sequence populations were largely distinguishable from xenomouse derived E1E2s**.** Highlighter plots and phylogenetic trees revealed mutational patterns associated with defined sequence variants within pre- and post-transmission populations. When compared to the KP consensus master sequence, three non-synonymous substitutions that were present at a low level in the source inoculum predominated in post transmission populations. All three of these

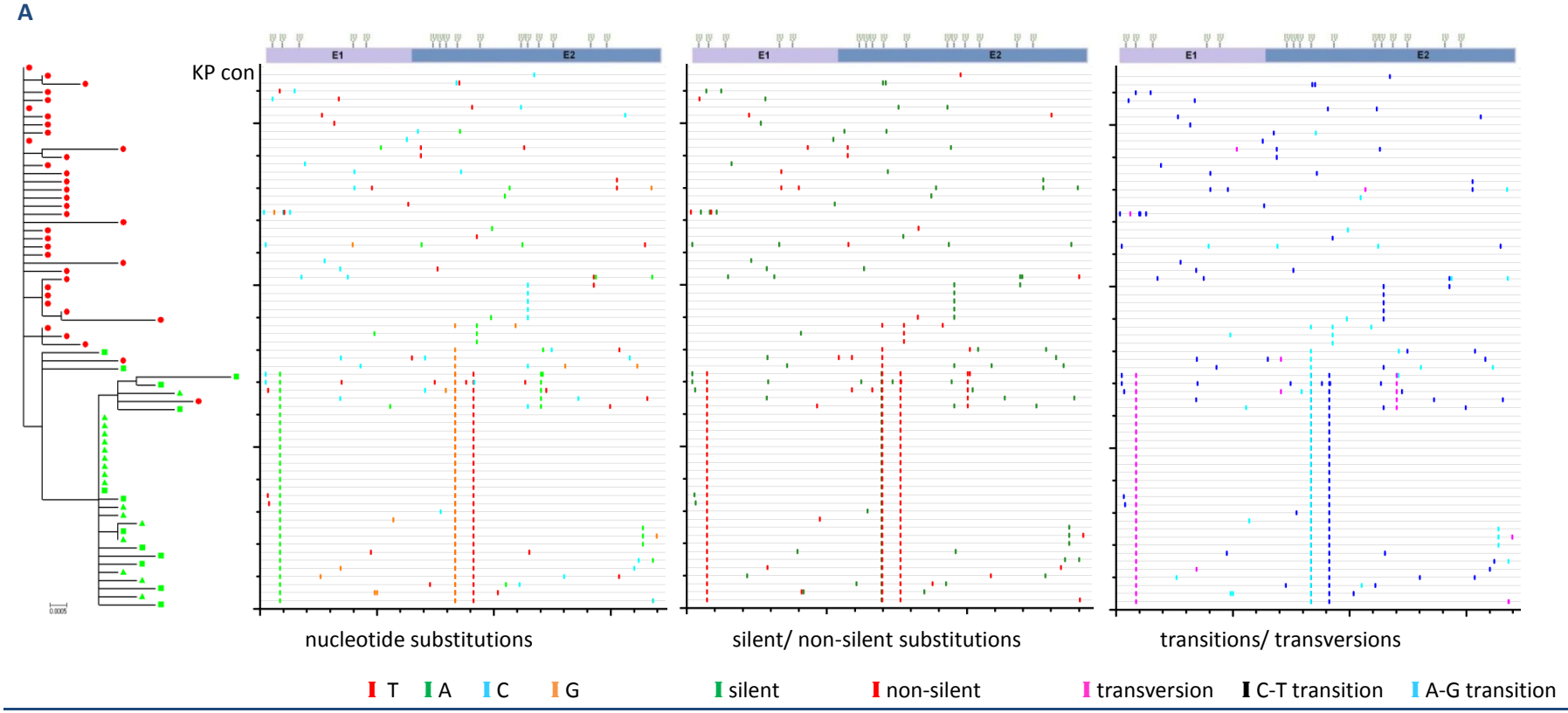
non-synonymous substitutions were located outside HVR1 (see Figure 3.8). The same mutational pattern was observed in each experimental transmission event.

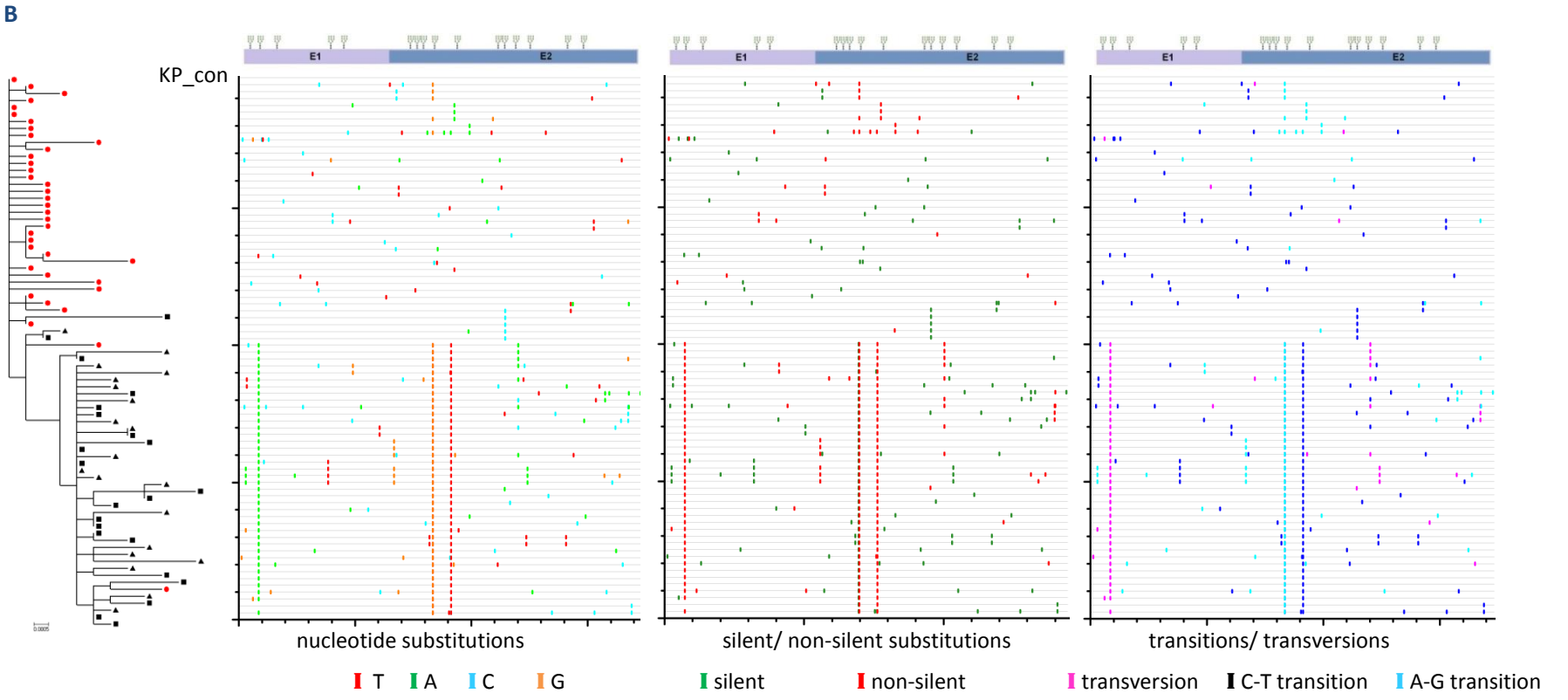
Analysis of encoded amino acid sequence alignments allowed us to identify four key E1E2 residues which are likely to be important in effective transmission and establishment of productive infection in a new host. Substitutions at amino acid residues 198, 448, 474 and more subtle variation at 570 were consistently observed when compared to the donor inoculum, irrespective of xenomouse host. Consensus and MRCA residues at these positions in each xenomice viral population can be seen in Figure 3.9 A. Substitutions at these positions appeared in one of 5 possible combinations SNHV, SDHV TDYD, and SDYV and TDYV/A, the frequencies of which are summarised in Figure 3.9 B.

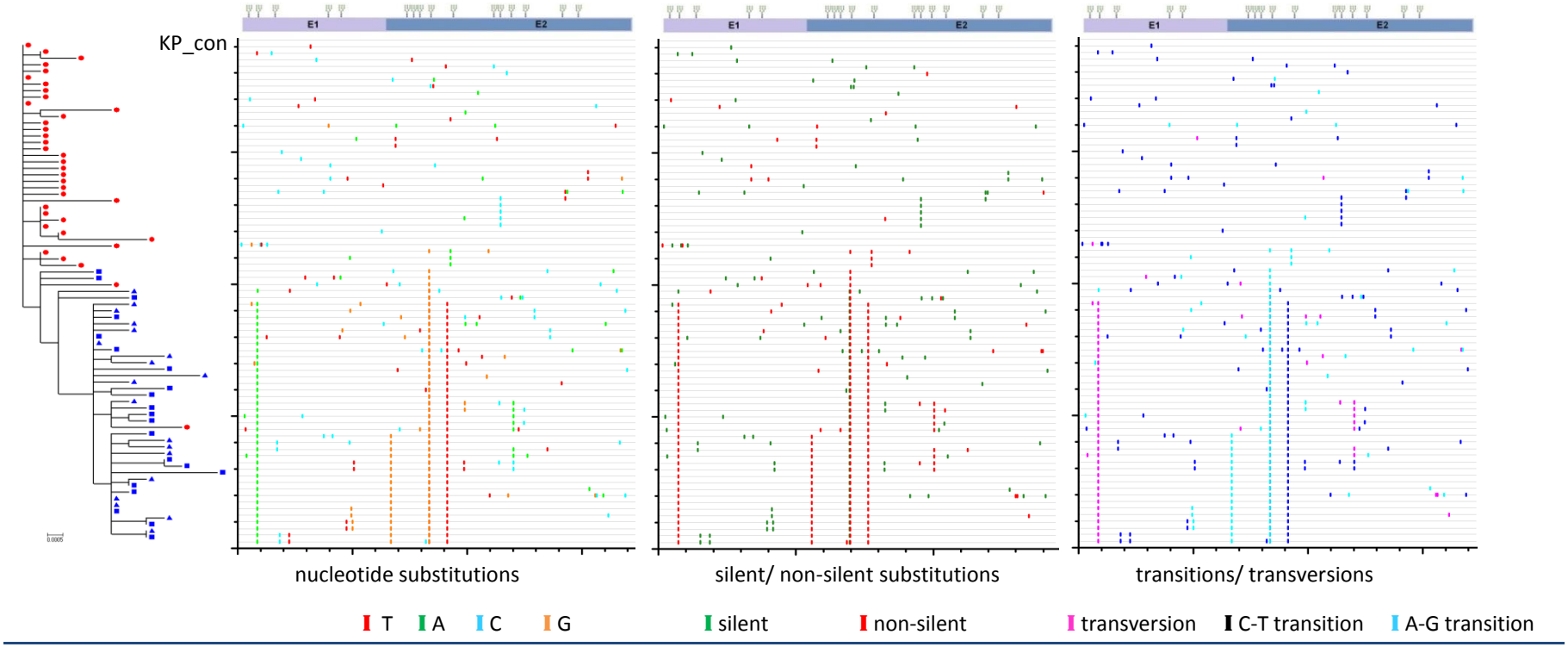
The majority E1E2 variant in the donor population possesses the SNHV discontinuous quartet combination. This combination of residues is encoded in ~90% of all donor E1E2 amplicons. The N448 residue is part of an NXS/T potential N-linked glycosytalion (PNG) motif which is highly conserved in the donor inoculum and in the majority of globally sampled HCV strains, presumably due to functional constraint. Surprisingly the concerned PNG site is absent in all post transmission sequences. The major pre-transmission variant (SNHV), possessing functional glycan site at position 448, is not detectable in any mouse population. The TDYV/A quartet, which is undetectable in the donor population (see Figure 3.9) becomes the dominant post-transmission variant, irrespective of the recipient xenomouse host. Two additional combinations, TDYD and SDHV, appear at low frequency in all recipient xenomice viral populations. Both of these minority combinations possess a PNGS knockout at position 448. Both of the variants are also detected at low frequency in the donor E1E2 population. Additionally a novel SDYV variant, previously undetected in the donor virus population, was also

detected within first sampling time points of recipient mice 666 and 931. This rare variant was absent in later sampling time point of both xenomice.

In conclusion, these data demonstrate that whilst HVR1 sequences remain stable in HCV transmission, key residues elsewhere in E1E2 are important in viral transmission and initiation of infection in a new host. The application of the SGA method allowed accurate assessment of E1E2 sequence variant distribution in donor and recipient plasma. Analysis of full-length E1E2 allowed identification of key residues outside of HVR1 associated with establishment of initial infection and indicate that the major post transmission variant (TDYV/A) was undetectable in the donor population. However, the TDYD variant was detectable in the donor inoculum, and this sequence variant is only one substitution away from the major post transmission variant. In conclusion these findings suggest that a E1E2 key residue variant that is undetectable in the source inoculum is selectively amplified post transmission. It is likely that this key residue combination confers a fitness advantage in the new selective environment.


A
KP con
E1
E2
E1
E2
E1
E2
nucleotide substitutions
silent/ non-silent substitutions
transitions/ transversions
T
A
C
G
silent
non-silent
transversion
C-T transition
A-G transition


B
KP_con
E1
E2
nucleotide substitutions
silent/ non-silent substitutions
transitions/ transversions
T
A
C
G
silent
non-silent
transversion
C-T transition
A-G transition


C
KP_con
E1
E2
nucleotide substitutions
silent/ non-silent substitutions
transitions/ transversions
T
A
C
G
silent
non-silent
transversion
C-T transition
A-G transition

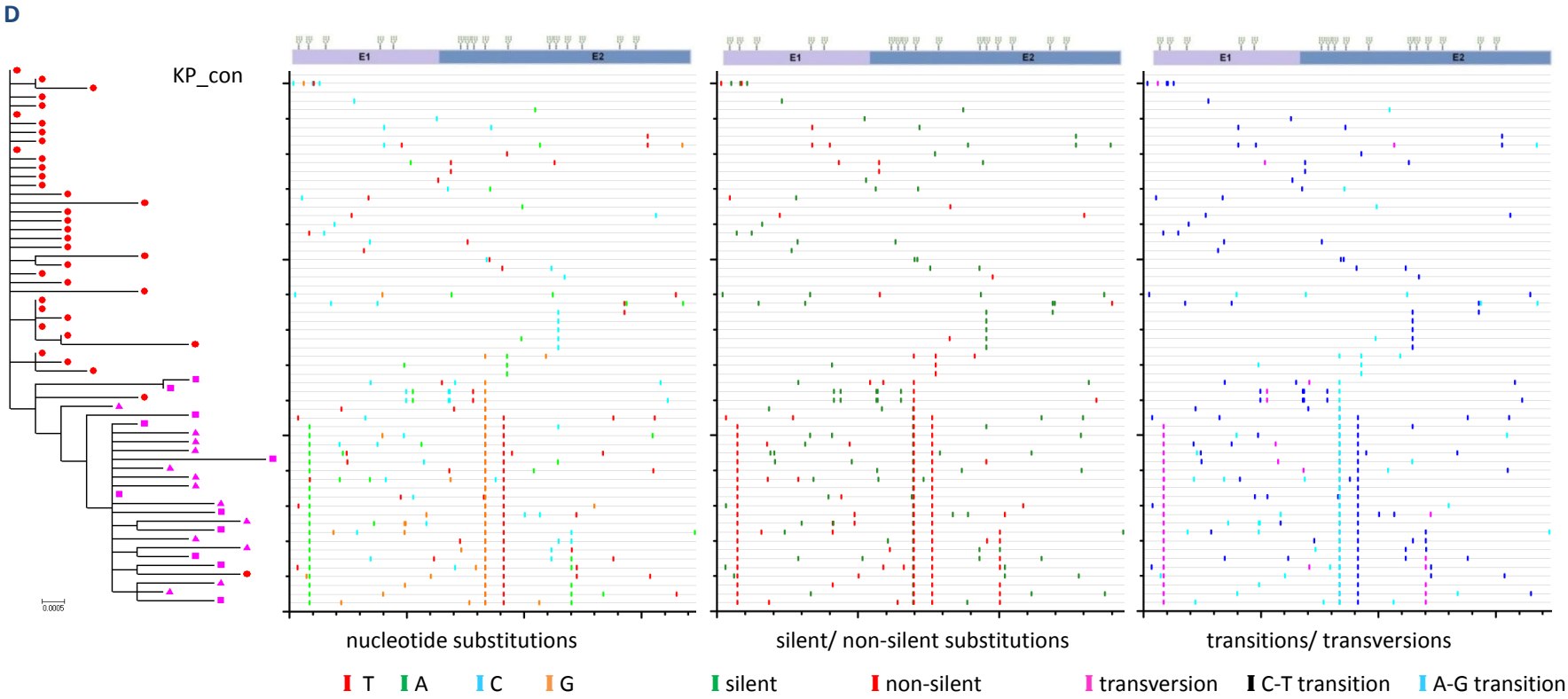


Figure 3.8. Identification of transmitted variants of envelope glycoprotein genes in experimentally infected mice on the basis of phylogenetic reconstructions and highlighter plots.

SGA derived sequences from post transmission samples from mouse 594 (A), 666 (B), 714 (C), 931 (D) were analysed by phylogenetic reconstructions (left panels) and Highlighter plots (right panels) including pre transmission donor population KP. Maximum Likelihood trees are rooted on KP consensus sequence which is a master sequence in corresponding highlighter plots. Highlighter diagrams depict location of substitutions in each E1E2 sequence as compared to chosen master pre transmission sequence (KP consensus).

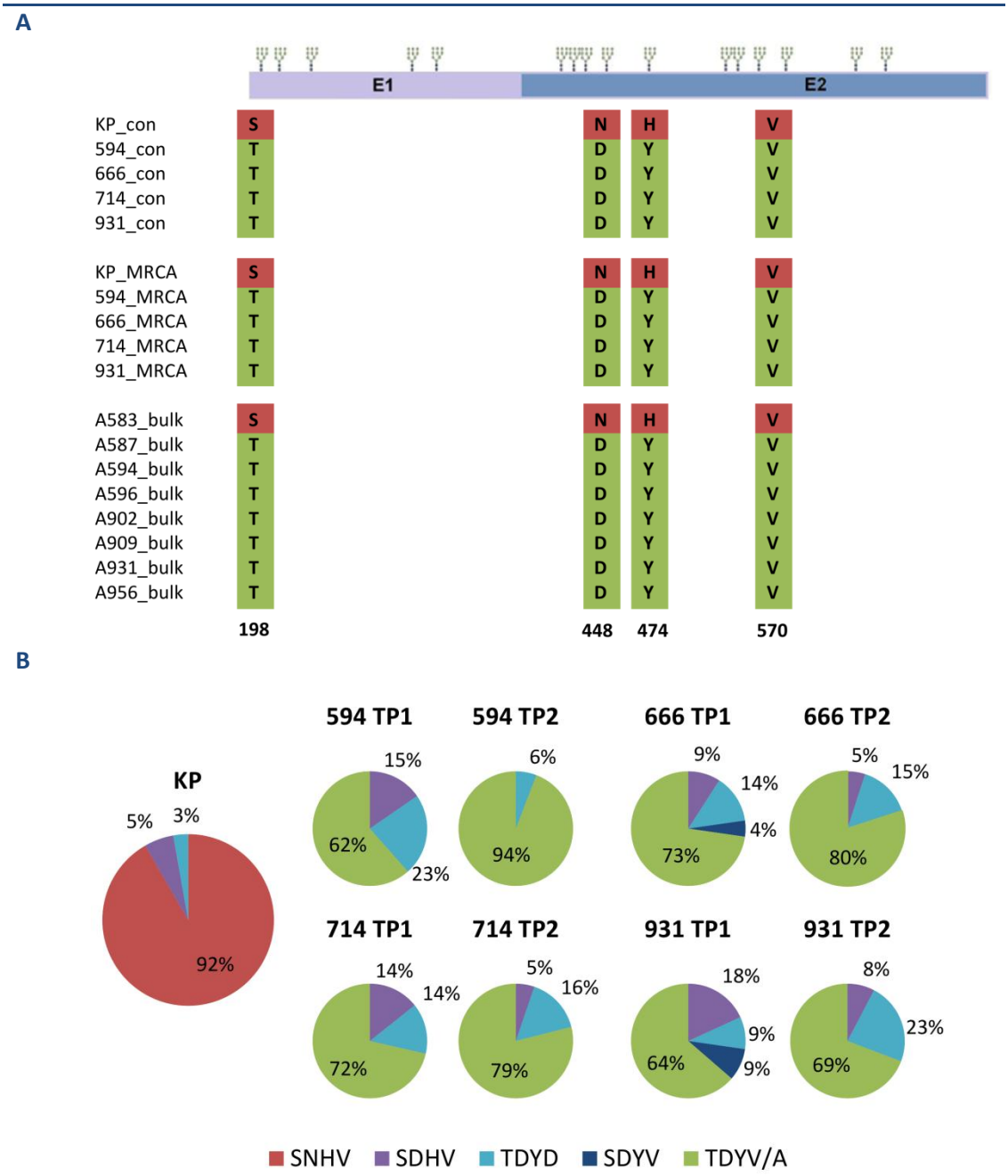


Figure 3.9. Location of key transmission residues and frequency distribution of key residue variants in donor and recipient mice.

(A) Schematic E1E2 diagram depicting the locations of 4 key residues involved in transmission. Position of conserved PNG sites are indicated above the E1E2 protein with absolute polyprotein co-ordinates of four key residues positioned beneath numbered relative to homologous positions in the H77 reference strain polyprotein ORF (accession no. NC_004102). Coloured vertical columns located below key residues indicate donor and recipient consensus (top) and MRCA [164] amino acids derived from SGA amplicon populations, in addition to majority consensus amino acids bulk amplified from recipient xenomice [273]. (B) Frequencies of key residue combinations circulating within donor/recipient hosts. TP1 = time point 1; TP2 = time point 2.

3.1.6. Phenotypic analysis

To test whether different key residue variants possessed different phenotypic properties, and assess the effect of the N448D knockout, three sequence variants (SNHV, SDHV and TDYD) were cloned into a pcDNA3.1 expression vectors. Unfortunately we were unable to clone the majority mouse variant (TDYV).

3.1.6.1. Protein expression and antibody recognition

Variant proteins were expressed individually in 293T cells. Although this method is not quantitative, Western Blotting analysis showed similar amounts of proteins being produced (see Figure 3.10) indicating no difference in translational efficiency and E1E2 expression between the tested variants. Furthermore, variants were compared in plate based assays to assess any differences in binding to a panel of monoclonal antibodies as described in 2.26

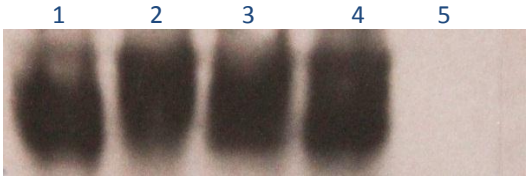
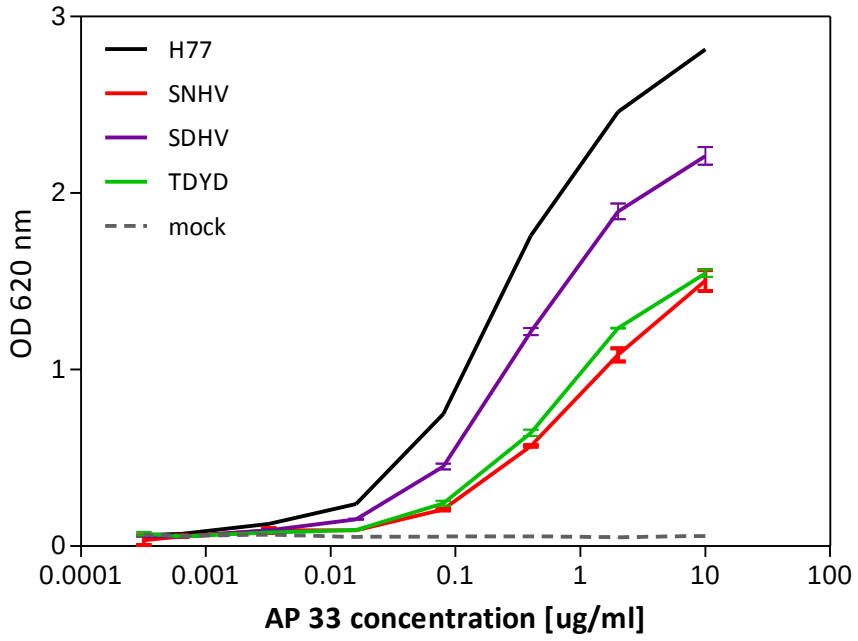


Figure 3.10. Expression of clones derived from representative sequences from variants circulating in experimentally infected mice.
10% SDS-PAGE followed by Western Blotting hybridisation with AP33. Key: 1- H77.20, 2- SNHV, 3- SDHV, 4-TDYD, 5- mock.

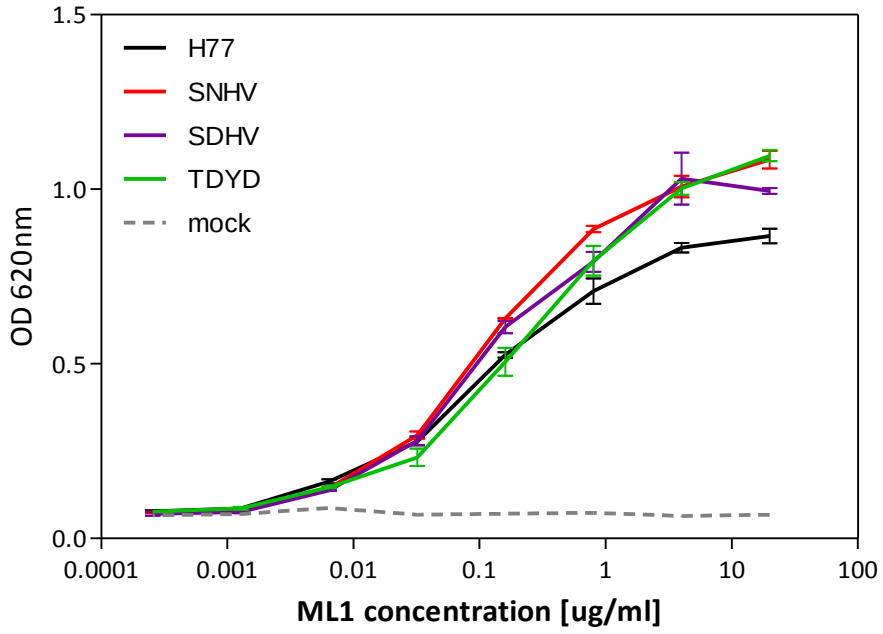
Results of GNA capture ELISAs with 5 recognised anti-E2 antibodies are presented in Figure 3.11. Figure 3.12 presents CD81 capture ELISA with detection by linear anti-E2 antibody ALP98. Probing with antibodies recognising linear (AP33, ALP98) and conformational epitopes (Ab1:7, ML1, ML2, DB1) did not reveal any significant

differences in recognition. All analysed variants of E1E2 glycoproteins appear to be expressed at similar levels and folded correctly when expressed in 293T cells.

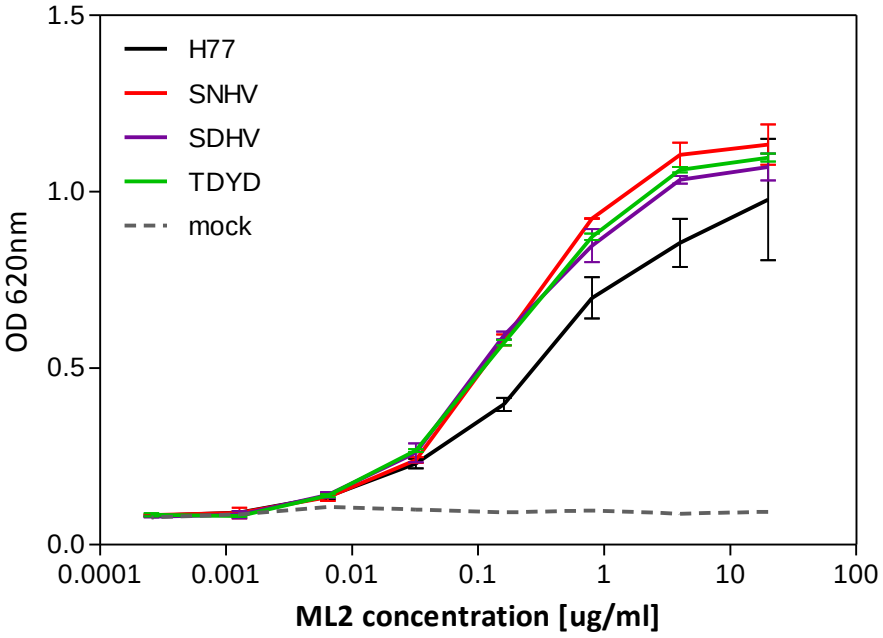
A



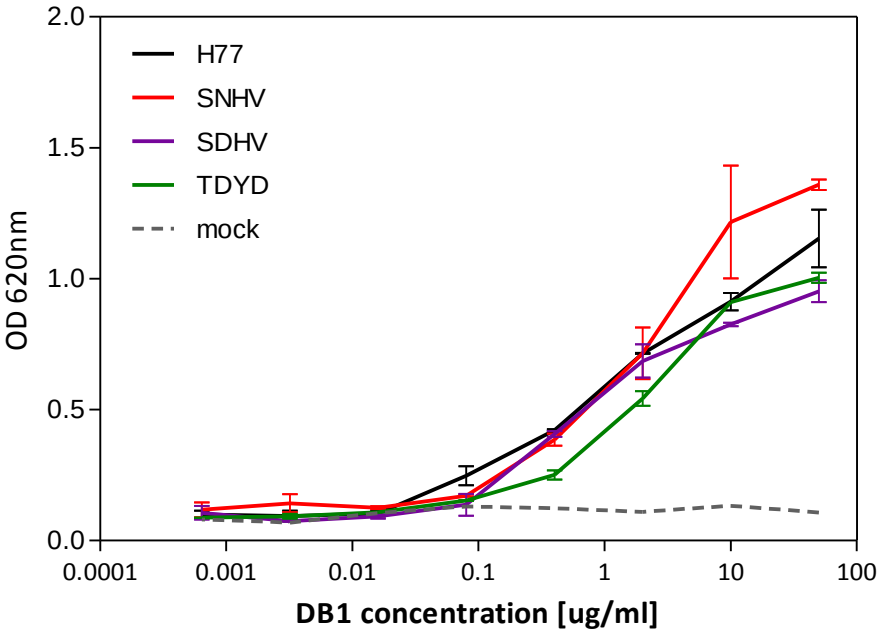
B



C


1.5
1.0
0.5
0.0
OD 620nm
H77
SNHV
SDHV
TDYD
mock
0.0001
0.001
0.01
0.1
1
10
100
ML2 concentration [ug/ml]

D


2.0
1.5
1.0
0.5
0.0
OD 620nm
H77
SNHV
SDHV
TDYD
mock
0.0001
0.001
0.01
0.1
1
10
100
DB1 concentration [ug/ml]

E

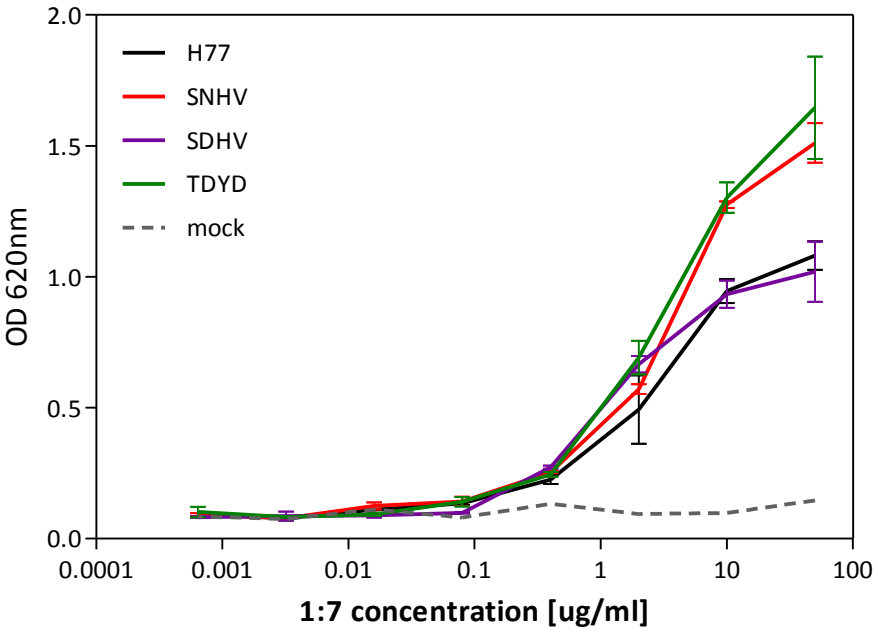


Figure 3.11. Antibody recognition of selected E1E2 in ELISA.
Panels A, B, C, D, and E depict recognition of selected clones by antibodies AP33, ML1, ML2, DB1 and Ab 1:7 respectively. GNA capture ELISA, titration of antibody against set amount of proteins.

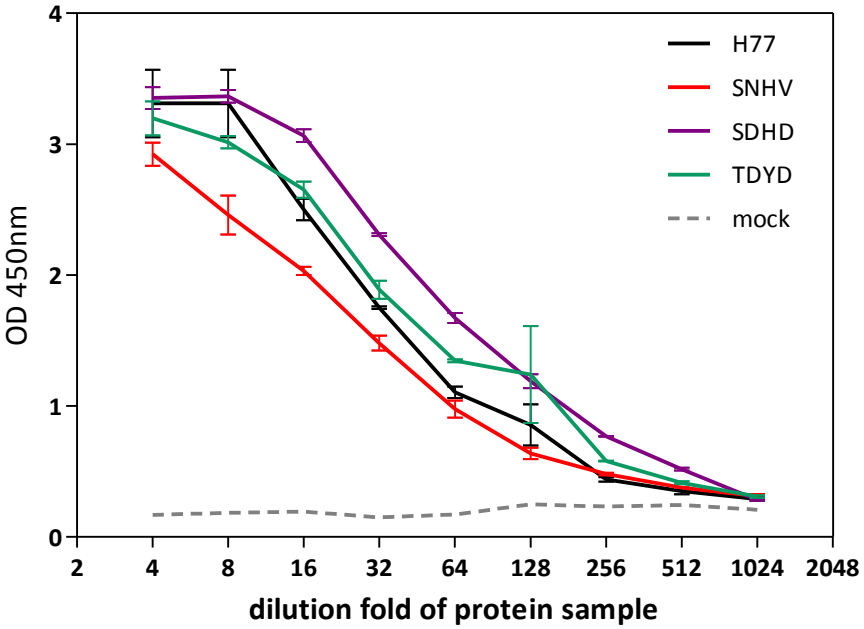


Figure 3.12. E1E2 binding to CD81-LEL in ELISA.
CD81-LEL capture ELISA, titration of antigens against set concentration [1µg/ml] of antibody ALP98.

3.1.6.2. HCV pseudotyping

HCV pseudotype (HCVpp) infection of Huh7 hepatoma cells by key residue E1E2 variants was then performed (for method see sections 2.22 and 2.23). All tested HCVpps had comparatively low infectivity (when compared to the H77 E1E2 positive control) despite being recognised by antibodies in a similar manner to the control strain (Figure 3.14). However, a significant increase in capacity for target cell entry stage was noted for the TDYD and SDHV variants containing the N448D substitution, when compared to dominant donor SNHV variant. Thus we propose that the observed substitutions outside HVR1 may have an impact on the variant's fitness at the entry level and that this may be modulated by the N448D substation which results in a PNG knockout.

Subsequent HCVpp infection of primary hepatocytes and Huh 7.5 cells performed by our collabators at the University of Birmingham (Brown et al. Manuscript submitted) confirmed our preliminary data regarding differential entry fitness of key residue variants and indicate that the N446D substitution is integral in enhanced infectivity.

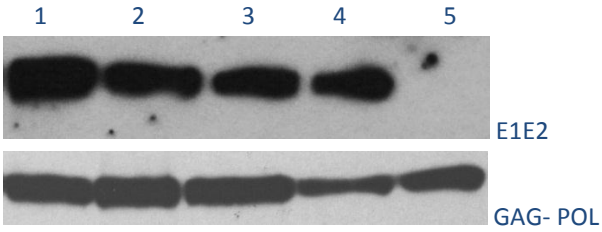


Figure 3.13. Western Blotting analysis of HCV pseudoparticles.
Upper panel- E1E2 detected by AP33 probing, lower panel- Gag-Pol MLV protein detected by Anti-GAG antibody. Key: 1- H77, 2- SNHV, 3- TDYD, 4- SDHV, 5- ΔE1E2.

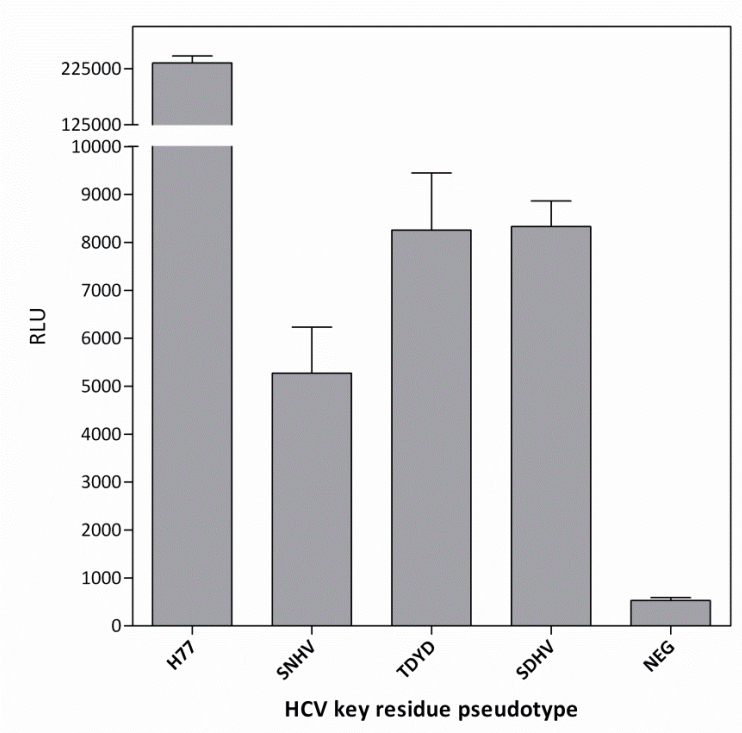


Figure 3.14. Phenotypic analysis of selected E1E2 by HCVpp infectivity assay.

3.2. Discussion

In the first part of this study, the xenomouse model was utilised to investigate the evolutionary mechanisms underlying HCV transmission. The majority of studies focussing on acute HCV infection are unable to assess the dynamics of the preceding transmission event due to lack of known donor samples. This is largely as a result of the asymptomatic nature of acute infection. Additionally, many studies of HCV variants evolution utilise only partial E1E2s, centred on the HVR1 region. Regions important in HCV transmission and establishment of productive infection outside of HVR1 may be missed using only partial envelope sequences. Using partial E1E2 sequences also precludes meaningful subsequent phenotypic investigation. Contrastingly, our investigation is the first to describe HCV full-length E1E2 glycoprotein variants distribution from a defined source at transmission. E1E2 sequences were derived from single genome templates, minimising in vitro generated artefacts, which may skew the

sequence variants distribution and misrepresent evolutionary processes underlying observed sequence diversity. In addition, Alb uPA/SCID xenomice were experimentally infected with a known source inoculum whose sequence population was also interrogated. The full-length nature of the E1E2 sequences retrieved also enabled phenotypic characterisation of donor and recipient glycoproteins at transmission.

To ensure accurate and representative sampling of both donor and recipient quasispecies populations, an SGA approach was utilised. The SGA method, followed by direct sequencing, has been previously applied to characterise viral quasispecies populations in HIV-1 [220, 320-322] and HCV infection [323]. PCR Amplification from single molecule cDNA templates is preferable to the standard bulk amplification and cloning approach when analysing viral quasispecies as it excludes polymerase induced nucleotide misincorporation, amplicon re-sampling, selective amplification of specific isolates, cloning bias and the generation of *in vitro* recombinants via polymerase template switching. In traditional approach of clonal analysis low frequency population variants can be missed or overrepresented, which can be precluded by use of SGA approach. Furthermore, SGA is not affected by the limitations of ultra-deep sequencing technologies. Whilst the apparent nucleotide frequencies at each position within any given sequence population can be assessed using UDS, genomic linkage between substitutions cannot be ascertained via this method.

When the total donor/recipient data set was analyzed in a composite tree, a lack of monophyletic clustering based on recipient xenomouse was observed. However, the majority of donor E1E2 sequences were phylogenetically distinct, residing in an antecedent clade. The branching pattern observed in the composite tree is due to synapomorphic substitutions common to all post transmission populations. These data suggest the different xenomouse hosts constitute highly similar selective environments, despite possessing human liver grafts from different donors. Interestingly, the post

transmission clade contains two isolates derived from the KP source inoculum. Analyses of individual donor/recipient populations revealed that the composition of sequence populations changed dramatically in separate parallel transmission experiments from the composition observed in donor inoculum. Irrespective of recipient xenomouse viral sequences analysed, a minor E1E2 variant circulating within the donor becomes the dominant variant within the newly infected xenomice. It is likely that this minor variant, present in the source incoculum at low frequency, would have been either absent or overrepresented if a clonal analysis of the viral sequence population had been employed. Previous investigations report low genetic diversity shortly after infection, which has been interpreted as founder effect: infection originating from limited diversity inoculum or single virus. The genetic signature of founder effect (establishment of infection from a single strain) is indistinguishable from a selective sweep (a selectively advantageous genetic variant is swept to fixation in the recipient) in contemporaneously sampled populations. However, knowledge of the genetic composition of the donor viral population allows us to assess which of these two competing evolutionary scenarios is most likely to have given rise to the sequence distributions observed in recipient xenomice. By comparison of donor and recipient viral sequence populations, our analyses indicate a selective sweep occurring upon transmission, whereby the frequency of one of the minor variants at undetectable level in donor inoculum becomes the major variant within all the experimentally infected xenomice. This phenotype is advantageous in the immunocompromised mouse, which possesses no adaptive immune system, and is swept to fixation due to increased fitness in a new selective environment. Analysis of transmission events in several mice suggests that this variant arose in the host, preceding the experimental infection, as opposed to developing independently post-transmission in each host.

Our findings are in agreement with previous reports, suggesting the major HVR1 donor variant is established as the major variant in recipients after transmission [212, 213, 215, 217]. In this respect, the xenomouse Alb uPA/SCID model appear to mimic the HVR1 evolutionary stasis observed at transmission between human hosts and experimental transmission from human to chimpanzee. However, expanding the sequence coverage to encompass full-length E1E2 sequences revealed contrasting results to previous transmission studies, that advocated phylogenetic grouping of envelope sequences based on HVR1 sequence analysis [236].

Our data revealed three key amino acid sites (198, 448 and 474) and subtle variation at a fourth position (570) played a role in establishment of a productive infection in a new host in the xenomouse system. The major donor variant SNHV appeared unsuccessful in establishing infection post transmission and was outcompeted by the TDYD variant, which became the dominant component in all recipient xenomice. All post-transmission E1E2s harbour an N→D PNGS knockout, which presumably confers a selective advantage in each xenomouse host. A PNGS knockout at position 448 (corresponding to glycan E2N4) has previously been shown to abrogate H77 E1E2 pseudoparticle infectivity [347]. However, in the JFH-1 HCVcc system PNGS 448, has not affected infectivity although has been demonstrated to render envelopes more susceptible to neutralisation by a panel of anti-E2 monoclonal antibodies, indicating PNGS 448 is critical in shielding neutralisation sensitive epitopes from immunological targeting [71]. This would suggest that TDYV variant is selectively amplified due to the fitness it harbours in the new host in the absence of neutralising antibodies. Unfortunately we were unable to access any of the original KP inoculum to perform neutralisation tests to assess this hypothesis. However, none of the variants demonstrated variability in sensitivity to antibodies whose epitopes overlap or are the in vicinity of the CD81 binding site. This indicates that antibody selective pressure against the TDYV sequence

variant in the donor population would have been directed outside the CD81 binding site.

In our hands pseudotype KP and mouse-derived E1E2 clones showed very low level infectivity in HCVpp when used to infect Huh7 cells. Nevertheless the infectivity conferred by the mouse derived E1E2 was greater than that conferred by the major variant present in the KP inoculum. Further analysis, performed by our collaborators, showed significance of this difference by use of primary hepatocytes and obtaining higher levels of infectivity (Brown et al. submitted). This fact is in agreement with observations made in the liver transplantation setting, where viral variants with increased capacity for cell entry have been shown to lead the re-infection of grafted liver [236].

A number of different donors were used to supply hepatocyte grafts for the xenomice utilised in this investigation. HCV infection of recipient xenomice is dependent on a high percentage of human graft in each uPA/SCID mouse [277]. It has previously been demonstrated that HCV does not replicate in mouse hepatocytes [348]. Thus productive viral replication and infectious particle production observed in recipient xenomice can only be sustained by engrafted human hepatocytes. The observed E1E2 adaptations in recipient xenomouse virons will therefore have been driven to fixation by their ability to enter and replicate in human hepatocytes, indicating the xenomouse system is a useful surrogate model to enable interrogation of HCV quasispecies transmission to a HCV naïve host and establishment of initial infection prior to seroconversion.

4. Tracking a naturally occurring HCV transmission event

4.1. Results

Having assessed the dynamics of HCV experimental transmission in the chimeric mouse model, we then investigated a case of HCV transmission acquired through sharing a haemodialysis machine. Patients undergoing haemodialysis are at increased risk of HCV infection [317]. Horizontal transmission between haemodialysis patients has been well documented [318, 349]. Standard infection control procedures have been shown to drastically reduce nosocomial spread of HCV within dedicated haemodialysis units [350]. Transmission of HCV to multiple patients using the same haemodialysis machine is not well documented. Consequently dedicated hemodialysis machines for HCV infected patients are not routinely used. Using phylogenetic analysis we confirm a case of nosocomial HCV transmission between patients receiving haemodialysis in the same unit, after sharing the same machine. These analyses have implications for health care management guidelines.

4.1.1. Case study background

A 51-year-old HCV negative woman had been receiving regular haemodialysis at the Glasgow unit for 7 years. Routine anti-HBV and anti-HCV antibody testing demonstrated negative results throughout this period. However anti-HCV antibody and corresponding HCV PCR returned positive in September 2009. Retrospective analysis of samples acquired in June 2009 confirmed HCV PCR positivity and antibody negativity indicating acute infection, giving a likely window the time line for the infection event.

Other risk factors for HCV infection were eliminated. All regular unit patients were HCV antibody negative. However, the index patient received four separate dialysis sessions immediately following a known HCV positive patient, at a different inpatient unit after hospital admission for unrelated complication in late March/early April [317].

4.1.2. Amplification of full-length E1E2 sequences

To ascertain whether index case and hypothetical source viruses were related, multiple samples from both patients were received from the renal unit in Glasgow. Subsequently viral RNA was extracted, cDNA reverse transcribed and used for template in PCR amplification. Received samples had very low viral titres and consequently it proved difficult to amplify full-length E1E2 sequences. Serum samples that did not yield sufficient RNA to amplify detectable 5'-NCR fragments or E1E2 after QIAGEN RNA extraction were extracted subsequently using the EasyMAG technique (seeTable 3.5, both methods described in 2.2). Amplification of full length glycoprotein sequences was performed from viral cDNAs and were directly sequenced. From three samples derived from the suspected source, only one yielded a full length E1E2 sequence. For five index case samples only three yielded sequences. Reasons for these apparent low titres could be due to repeated freeze-thawing of the sera prior to arrival to Nottingham, or contamination with RNAses. For the purposes of comparison an additional 19 genotype 3A infected serum samples from Glasgow renal unit were also acquired. Full length E1E2 sequences were derived from six of these samples.

Table 4.1. Summary of extraction and DNA amplification results within low titre transmission samples.
RNA samples that did not yield enough cDNA for 5'NCR or E1E2 amplification were extracted again with alternative method. For second batch of samples only Easy-MAG method was applied.

	Sample Label	QIA- RNA extracted		MAG- RNA extracted	
		5'NCR amplification	E1E2 amplification	5'NCR amplification	E1E2 amplification
Index case samples	IC1	–	–	+ (faint)	–
	IC2	+ (faint)	+	N/A	N/A
	IC3	+ (faint)	+	N/A	N/A
	IC4	+ (faint)	–	N/A	N/A
	IC4 (RNA)	+ (faint)	+	N/A	N/A
Hypothetical source samples	HS1	–	–	–	N/A
	HS2	–	–	+	–
	HS3	+ (faint)	–	+	+
Unrelated control patient samples	PG	–	–	+	–
	PF	–	–	+	+
	PC	+ (faint)	–	–	N/A
	PD	–	–	+	–
	PA	–	–	–	N/A
	PE	+ (faint)	+	N/A	N/A
	PB	+ (v faint)	+	N/A	N/A
	PH	–	–		N/A
	PI	–	–	+	–
	PT	N/A	N/A	+	-
	PS	N/A	N/A	-	N/A
	PR	N/A	N/A	+	-
	PP	N/A	N/A	-	-
	PO	N/A	N/A	+	+
	PN	N/A	N/A	+	+
	PM	N/A	N/A	+	-
	PL	N/A	N/A	+	-
	PK	N/A	N/A	+	+
	PJ	N/A	N/A	+	-

N/A- result not available/ amplification was not performed

4.1.3. Phylogenetic analysis of outbreak and control samples

Nine epidemiologically unrelated Nottingham E1E2 genotype 3A sequences were also introduced to analysis. Composite phylogenetic analysis of source, index and control sequences was conducted. Bootstrapped phylogenetic trees for HCV E1E2s, both nucleic acid and amino acid sequence revealed clustering of suspected epidemiologically linked sequences, indicating that HCV transmission on the renal unit was indeed the source of the outbreak. This analysis suggests infection with closely related strain of virus and high probability of index case infection acquired from the hypothetical source (Figure 3.15). The same relationship is evident for related control sequences. UKN3A2.12, UKN3A2.28 and UKN3A2.4/SP-1/B4 E1E2s cluster together with 100% bootstrap support and are derived from the same patient at different sampling time points. Sequences UKN3A4.6 and UKN3A4.37, which are also both derived from an additional patient also cluster together with significant bootstrap support. Other unrelated control samples do not cluster, or if they do- bootstrap values are below 70% (apart from PN and PO) and internal branches are relatively long.

These analyses confirm a case of nosocomial HCV transmission acquired from sharing a haemodialysis machine, using phylogenetic analysis. No fault was found with the dialysis machine in question. Thus, on the basis of this data, we propose that HCV infected patients should receive dedicated haemodialysis machines which should not be shared with non-HCV infected patients, which is not currently standard practice. Implementation of this strategy would remove the HCV infection risk factor associated with haemodialysis.

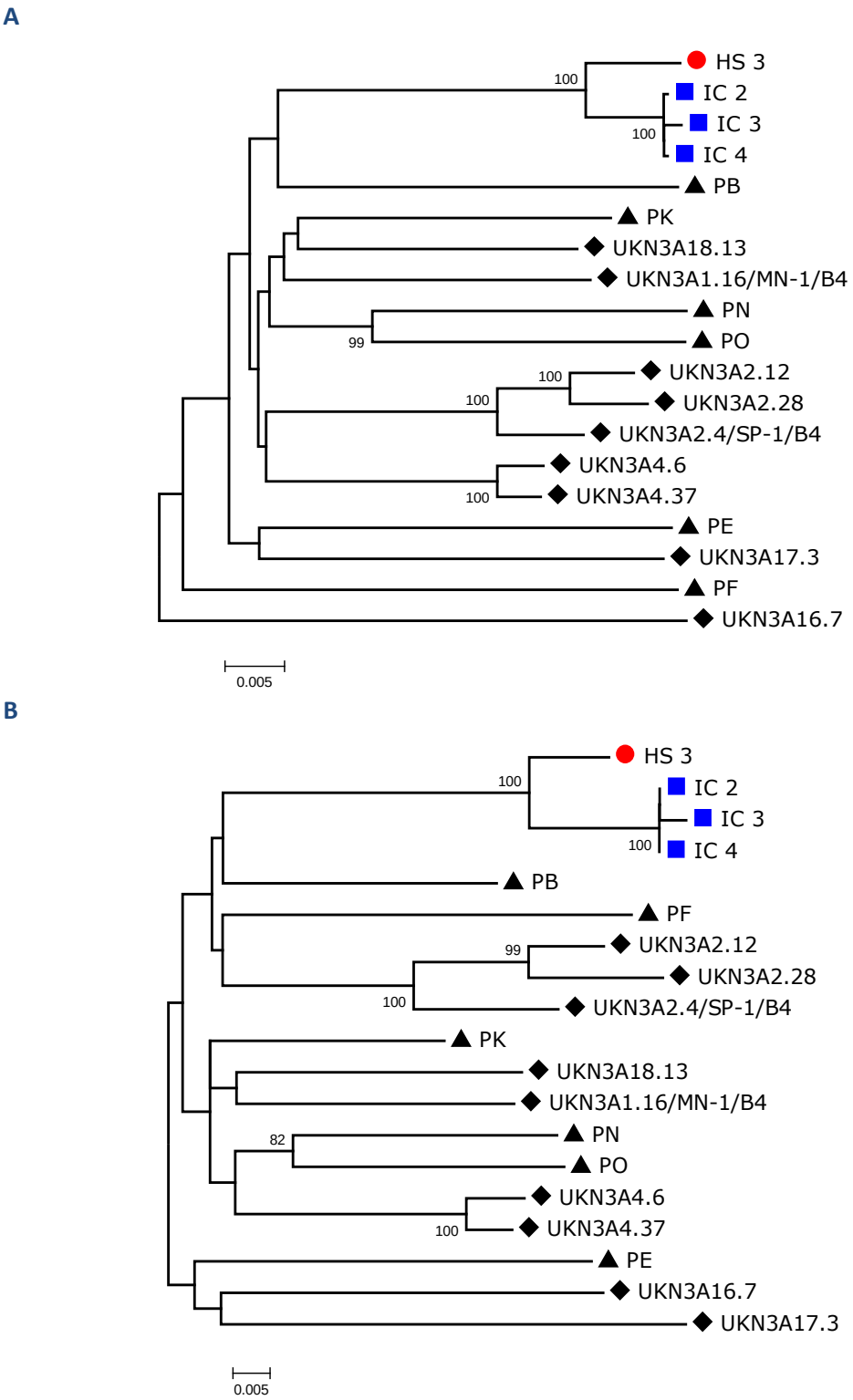


Figure 4.1. Phylogenetic Neighbour-Joining reconstruction of genotype 3A transmission. Clustering of the source and recipient derived sequences is supported by 100% of the bootstrapped replicate trees. Panel A- nucleic acid, Panel B- amino acid sequence phylogeny. Bootstrap values below 70% are hidden. Key: red circle- hypothetical source samples, blue squares- index case samples, black triangles- unrelated control patient samples from Glasgow, black rhombuses- unrelated patient sequences from Nottingham.

4.2. Discussion

In the second part of the study we have demonstrated that the virus strain infecting the index case patient indeed originated from the suspected source, by use of phylogenetic analysis. Molecular techniques have been successfully introduced to evolutionary study of relationships between organisms. The information obtained from nucleic acid sequences has undermined previous location of the human branch on the evolutionary tree, based on fossil data [351]. In case of viral diseases fossils are not available leaving only evolutionary biology as the means of resolving their origins and relationships. Precedence for use of phylogenetic relationship as an evidence of viral transmission has been long observed for HIV-1 outbreaks. The case of a dentist, whose unsafe procedures resulted in multiple patients becoming HIV-1 positive, has been amongst the first ones [352, 353]. Following this case many dental surgeries have undergone screening and guidelines were revised. Molecular proof of relationship between suspected donor and recipient HIV-1 sequences have become a standard evidence for the occurrence of transmission widely used to support criminal cases [354-356]. Similar approach has been employed to investigate HCV outbreak in anti-D immunoglobulin recipient cohort [357].

The investigated infection, as corroborated from the epidemiological data, could have occurred within a very limited time whilst index case individual was hospitalised and used the same haemodialysis unit as the suspected donor. The machine was thoroughly cleaned and prepared according to guidelines, which has not precluded HCV carryover. This supports the notion that ability of HCV to establish productive infection from very small initial inoculum volume should not be neglected. Unfortunately we were not able to obtain SGA sequences for the circulating viral populations to be able to shed light on the number of founder viruses and possible dynamics of viral variants.

Although routine screening of blood products for presence of Hepatitis C Virus was introduced in early 1990s, new nosocomial infections are noted each year. Patients receiving haemodialysis are especially at risk and many direct environmental cross-infections have been reported [350]. Most evidence, correlating HCV contamination to haemodialysis event, include the observation that more haemodialysis patients possess anti-HCV antibodies than peritoneal dialysis patients [358]. Moreover patients dialysed adjacent to HCV infected individuals more often develop anti-HCV antibodies [359], and physical isolation of non-infected patients, from those carrying HCV, correlated with lower contamination rates [360]. Strict appliance of infection control policies seems to have an effect on lowering HCV incidence [359], whereas poor control measures correlate with HCV outbreaks [361].

Guideline bodies have updated standard precautions that have been shown to reduce the risk of cross-infections in haemodialysis units. However use of separate machines for infected individuals was not suggested as a policy to limit HCV transmission, probably due to low number of reports evidencing transmission of HCV to sequential patient via a haemodialysis machine [362, 363].

Current guidelines have been strictly followed in the unit described in this study, however horizontal HCV transmission has occurred. In this unit the population of hepatitis C patients is not isolated from HCV- negative individuals. It seems viable to recommend extending infection control guidance by dedicating separate haemodialysis machines to patients carrying viral diseases.

5. Acute HCV infection

5.1. Results

In the previous two sections we have described experimental HCV transmission in the chimeric mouse model and a hospital acquired patient to patient transmission event. In the UK most HCV infection is acquired through intravenous drug use (IVDU). Consequently identifying source samples of HCV transmission in natural infection is challenging due to the chaotic life styles of drug users. Whilst diagnosis of acute HCV infection is still rare, due to the large size of the Trent Cohort HCV infected sera sample collection (greater than 5000), we were able to retrospectively identify a number of HCV positive samples derived from the acute phase of infection. Strains that establish initial infection are important for prophylactic and therapeutic vaccine development. In this section we describe genetic and phenotypic diversity associated with viral envelopes derived from acute HCV infection, and compare those to envelopes in chronic infection.

5.1.1. Amplification of full-length E1E2 sequences

Amplification of full length E1E2 via SGA was successfully achieved from acute serum samples. In total we obtained 126 end point titrated full length E1E2 sequences from infected patients, 41 acute window sequences from 4 samples, 17 sequences from 2 patients shortly post seroconversion and 68 sequences from chronically infected individuals at two separate time points. Patients' samples are detailed in Table 3.6. All SGA amplicons derived from chronic phase infection were generated by my colleague Mr Shafiq ur Rehman.

Table 5.1. Details of patients' samples.

Patient	Sample	Label/ Time point	Sample date	Number of SGAs
1A20	1A20	ACUTE 1.1	NA	10
1A30	1A30.3	ACUTE 2.1	07/10/2003	9
	1A30.12[1]	ACUTE 2.2	04/02/2005	7
1A57	1A57.2	ACUTE 3.1	02/09/2004	9
	1A57.14[1]	ACUTE 3.2	01/02/2008	10
1A65	1A65.1	ACUTE 4.1	05/04/2005	9
1A83	1A83.8	ACUTE 5.1	NA	4
1A67	1A67.5[2]	CHRONIC 1.1	15/07/1998	9
	1A67.9[2]	CHRONIC 1.2	13/02/2001	8
1A68	1A68.1[2]	CHRONIC 2.1	22/12/1992	7
	1A68.5[2]	CHRONIC 2.2	10/01/1994	9
1A69	1A69.1[2]	CHRONIC 3.1	12/08/1997	6
	1A69.2[2]	CHRONIC 3.2	12/03/2002	10
1A70	1A70.1[2]	CHRONIC 4.1	04/10/1999	10
	1A70.4[2]	CHRONIC 4.2	01/10/2002	9

[1]Indicates post seroconversion samples

[2]Indicates samples obtained and sequenced by Shafiq ur Rehman, further analysis performed by the author

5.1.2. Assessment of genetic distances and diversity in viral populations

To compare and contrast sequence data sets derived from acute and chronic phase infections, genetic distance and diversity analyses were performed. As expected, analyses revealed that sequence genetic distances in pre- seroconversion populations are lower than the ones observed in chronic infection (see Table 3.7 and Figure 3.16).

Table 5.2. Genetic distance (nucleotide substitutions per site) in SGA derived intrahost populations.

Label/ time point	E1E2	E1	HVR1	E2
ACUTE 1	0.00046	0.00031	0.00254	0.00040
ACUTE 2.1	0.00125	0.00069	0.00276	0.00149
ACUTE 2.2	0.00215	0.00089	0	0.00313
ACUTE 3.1	0.00578	0.00628	0.00283	0.00571
ACUTE 3.2	0.00757	0.00610	0.04350	0.00582
ACUTE 4	0.00167	0.00139	0.00281	0.00177
ACUTE 5	0	0	0	0
CHRONIC 1.1	0.00507	0.00502	0.00422	0.00517
CHRONIC 1.2	0.02369	0.01812	0.1281	0.02196
CHRONIC 2.1	0.01314	0.01293	0.01222	0.01377
CHRONIC 2.2	0.00838	0.0108	0.00847	0.00686
CHRONIC 3.1	0.04327	0.01296	0.01754	0.0111
CHRONIC 3.2	0.00978	0.00562	0.03017	0.00959
CHRONIC 4.1	0.0078	0.00783	0.02833	0.00622
CHRONIC 4.2	0.00791	0.00769	0.02756	0.00612

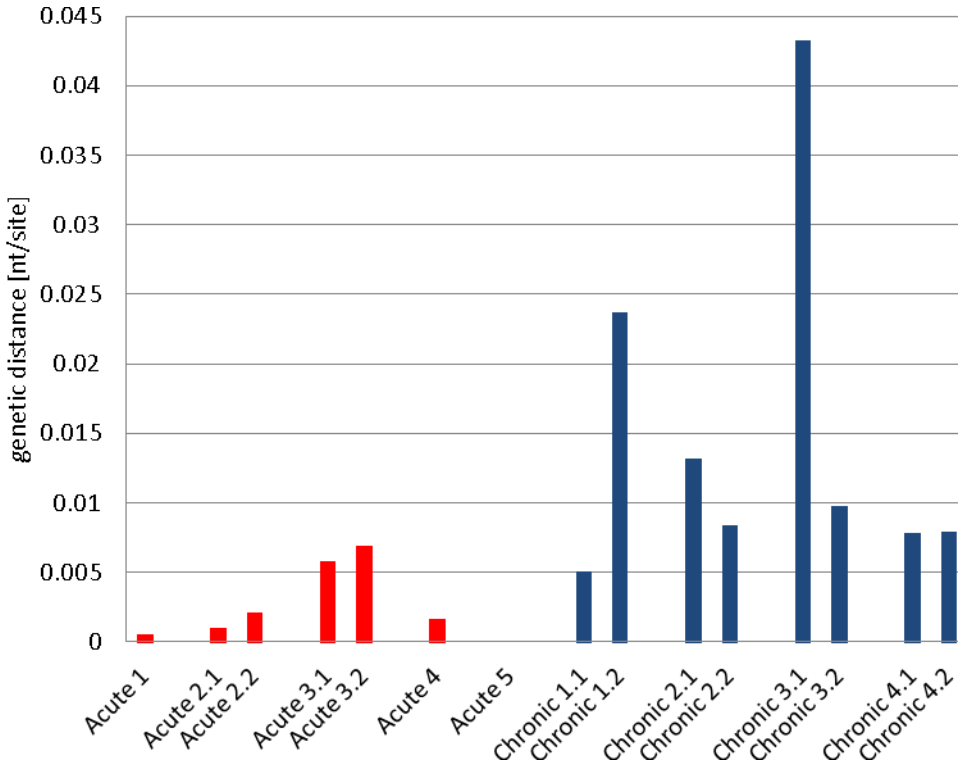


Figure 5.1. Genetic distance (nucleotide substitutions per site) comparison for populations of full length E1E2 sequences from different patients and time points. Graph represents highlighted column in Table 3.7.

Moreover genetic distances of the intrahost populations at different time points vary for all data sets (see Table 3.7 and Figure 3.16) suggesting that populations are genetically dynamic over time. Sequential samples from the acute dataset obtained in this project: Acute 2.1 and 2.2; and Acute 3.1 and 3.2 show an increase of genetic distance with time. For sequential samples of chronic dataset the difference is variable.

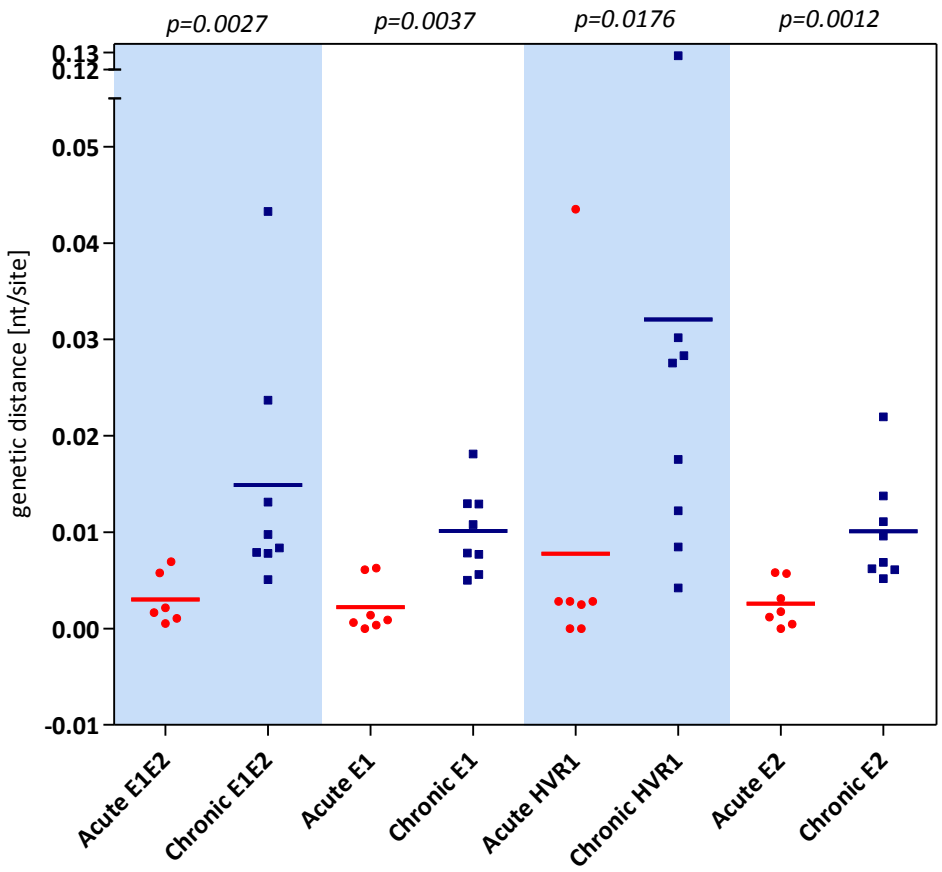


Figure 5.2. Difference in genetic distance between acute and chronic phases of infection. Comparison of full length E1E2 and separate domains' sequences. Red circles- genetic distances for acute populations, blue squares- for chronic populations. Horizontal bars represent mean values. P values associated with acute and chronic comparisons were calculated in GraphPad Prism using Mann-Whitney test.

Figure 3.17 depicts the comparison of genetic distances between acute and chronic data sets. The genetic distance calculations were made on the basis of full length E1E2 sequences as well as fragments spanning sequence coding E1, HVR1 and E2. These data

demonstrate there is a significant difference in genetic distance between sequences from acute and chronic phases of infection irrespective of domain analysed.

As in the chimeric mice transmission experiment, variability of HVR1 variants in acute infection was assessed. All apart from one (Acute 3.2) populations contain either none or single nucleic acid substitutions within all SGA obtained sequences. In Acute 2.1, Acute 3.1 and Acute 4 these substitutions are non-silent leading to population consisting of 2 HVR1 variants- a highly dominant and a minor variant at a 10, 11.1 and 11.1% frequency in respective populations. The HVR1 population in host Acute 3.2 is demonstrably more variable, with nucleic acid substitutions giving rise to 3 amino acid variants (see sequence logo alignments in Figure 3.18). The major variant comprises 50% of sequences, with two minor variants comprising 20 and 30% respectively. The higher observed variability within this population is likely to be directly related to the time of sampling from initial infection. Acute 3.2 sequences were derived from a serum sample collected 4 years after the Acute 3.1 time point, and about 3 years post the seroconversion event (see Table 2.3 in chapter 2). We do not observe such increased diversity in another post seroconversion population derived data set, Acute 2.2. This is likely to be due to reduced sampling time since the presence of anti-HCV antibodies were detected when compared to Acute 3.2 (see Table 2.3). The diversity of HVR1 sequences within assessed acute populations is generally very low, although it seems to increase after seroconversion. The restricted diversity we observe in the HVR1 in acute infection mirrors the limited diversity observed shortly after experimental transmission to chimeric mice and is in agreement with previous reports of acute infection [240, 364, 365]. We observe increased HVR1 diversity in the intrahost population of Acute 3.2 which is likely to be the result of escape from host humoral/ cellular immune responses [77, 235]. Due to the time from initial infection being in excess of 4 years this intrahost population is technically derived from the chronic phase

of infection. However these sequences served as a useful internal experimental control, and demonstrate increasing HVR1 diversity post seroconversion. Together these analyses show restricted HVR1 diversity in acute infection. This restriction is likely due to founder effect upon transmission or a selective sweep post transmission. Our experimental data from the xenomice portion of the study indicates that selective sweeps shape initial viral populations, and this may mirror the post transmission dynamics in natural infection. Due to the asymptomatic nature of acute HCV infection we were unable to identify donor samples to test this hypothesis.

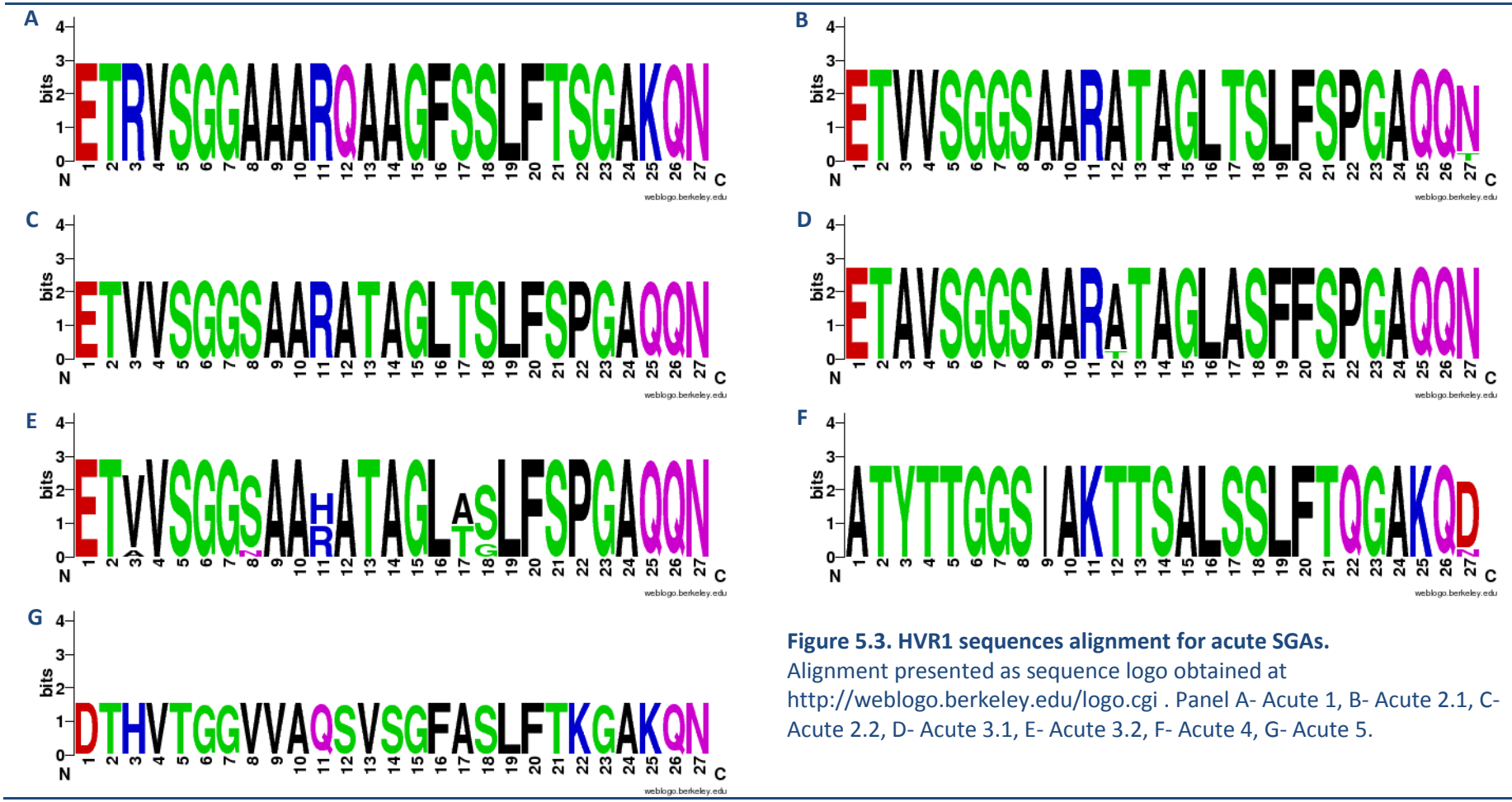


Figure 5.3. HVR1 sequences alignment for acute SGAs.
Alignment presented as sequence logo obtained at http://weblogo.berkeley.edu/logo.cgi . Panel A- Acute 1, B- Acute 2.1, C- Acute 2.2, D- Acute 3.1, E- Acute 3.2, F- Acute 4, G- Acute 5.

5.1.3. Phylogenetic analyses of sequence populations

To further analyze viral populations in acute phases of HCV infection and compare these to populations from chronic phase of infection, phylogenetic analyses of full length E1E2 glycoproteins sequences were performed (see Figure 3.19). Sequences cluster on the basis of sample origin and time point indicating no cross-contamination. Within the acute set samples, where two sequential time points were analyzed, clustering is evident between sequential samples. Pre-seroconversion window sequences are highly similar, which has been detailed in Table 3.7 and is also depicted in the phylogenetic reconstruction in Figure 3.19. Seroconversion seems to have an effect on E1E2 evolution- tree branch length and genetic distance rise as compared to window sequence populations. Further comparison with variants from chronically infected patients depicted in Figure 3.19 B shows that branch lengths and genetic distances within populations depend on time of sampling. The longer post infection and post seroconversion the more diversified the populations.

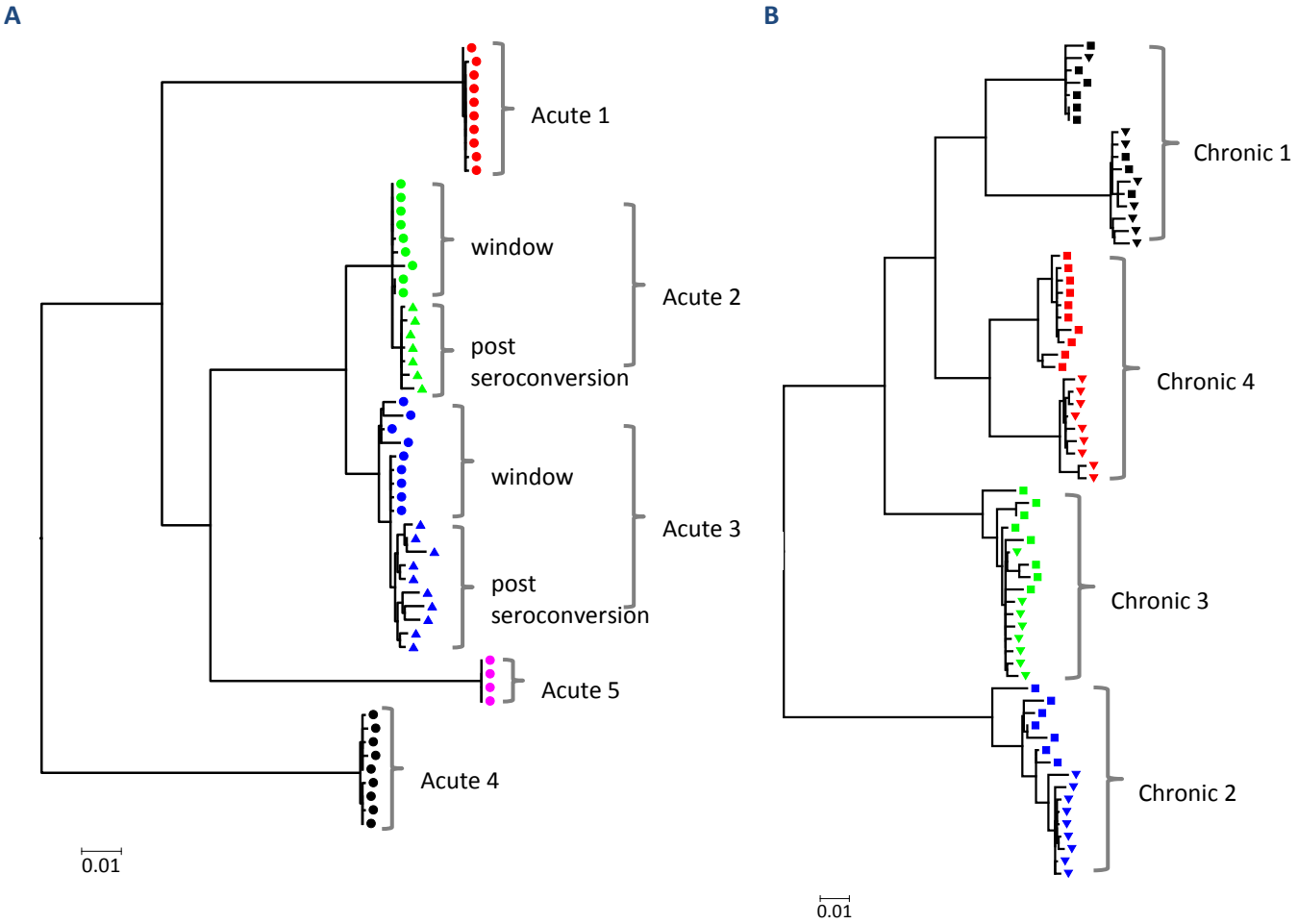


Figure 5.4. Phylogenetic reconstruction of sequences in acutely infected HCV patients. Maximum likelihood composite tree of all acute single molecule amplified sequences. HKY+G model of evolution calculated in modeltest 3.7.

Panel A:
Red circles- UKN1A20.1, green circles- UKN1A30.3, green triangles- UKN1A30.12, blue circles- UKN1A57.2, blue triangles- UKN1A57.14, pink circles- UKN1A83.8, black circles- UKN1A65.1.
Panel B:
Black squares- UKN1A67 time point 1, black triangles- UKN1A67 time point 2, red squares- UKN1A70 time point one, red triangles- UKN1A70 time point 2, green squares- UKN1A68 time point 1, green triangles- UKN1A68 time point 2, blue squares- UKN1A69 time point 1, blue triangles- UKN1A69 time point 2.

Full length E1E2 sequence populations derived from acute infection were compared to their respective most recent common ancestral sequences (MRCA), to elucidate substitutional patterns apparent in each host using highlighter plots (schematic representation in Figure 3.20).

For Acute 1 and Acute 4 E1E2 sequences, there is extremely limited nucleotide diversity apparent. Additionally for Acute 5 there was no nucleotide diversity observed with all sequences being identical. Consequently sequences from this population were omitted from the highlighter analysis. In the sequences derived from patient Acute 1 there is a low level of non-synonymous substitutions with two unlinked amino acid changes apparent. However in patient Acute 4 there is a slightly higher level of amino acid substitutions. Indeed at position 475 in the HVR2, roughly 50% of sequences contain Leucine while all the remaining sequences contain Proline. As these sequences are derived from pre-seroconversion any amino acid substitutions observed are not a result of escape from host humoral responses. Consequently any amino acid substitutions we see in Acute 1 and 4 sequences are likely to be selectively neutral, having negligible effect on viral envelope fitness.

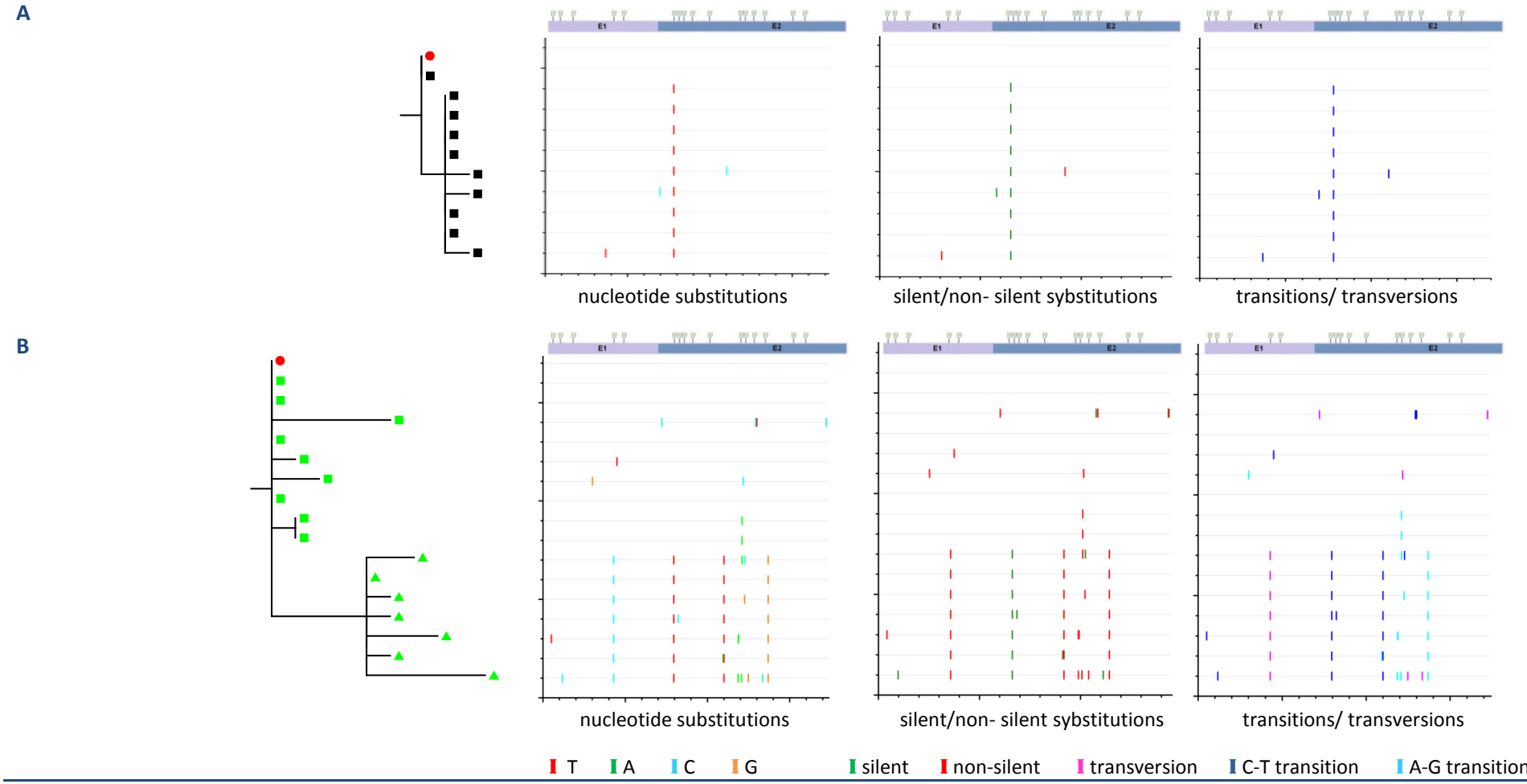
For the samples where sequential time point data is available differences in substitution patterns between pre- and post- seroconversion sampling times are distinguishable (Figure 3.20 B and D). In Acute 2 sequences, whilst multiple unlinked amino acid changes are apparent, non- synonymous substitutions at three positions are associated with differences between pre and post- seroconversion populations: one in E1 and two in E2. At position 312 in E1 we observe an E→D substitution. All pre-seroconversion sequences possess glutamic acid (E) whereas all post- seroconversion possess aspartic acid (D). There is extremely limited structural/ functional information available for the E1 protein so we are unable to ascertain what effect this substitution would have on E1 function, although this change is relatively conservative as both

glutamic and aspartic acid have similar physicochemical properties (polar, negative charge). The next non-synonymous substitution we observe at site 537 (A→V), with all pre-seroconversion possessing an alanine (A) whilst all post sero-conversion sequences possess a valine (V). The recent structural mapping of the E2 protein [75] suggests that this residue lies in the proposed Domain I, and is in close proximity to both an N-linked glycan and residues critical for CD81 binding. Whilst this substitution is relatively conservative it is possible that this change might have an impact on receptor binding or susceptibility to antibody mediated neutralisation. This will need an experimental confirmation. Finally, in patient Acute 2, we observe an I→V substitution at position 627 within E2. This residue is located in proposed Domain III [75] and whilst it is again relatively conservative is located in close proximity to another N-linked glycan site, removal of which is associated with impaired heterodimerisation and folding [65]. Whilst this substitution does not knock- out this glycan site, it is plausible that this change may result in minor conformational shifts which may affect protein folding, susceptibility to host neutralising antibodies, and capacity for cellular entry, although this hypothesis would require experimental confirmation.

Finally we observe multiple substitutions apparent in pre and post- seroconversion viral sequences from Acute 3. The majority of substitutions appear to be synonymous indicating purifying selection is the major force acting on these sequences, however multiple non- synonymous substitutions are also observed. Whilst a proportion of non-synonymus substitutions are unlinked and appear to have a random distribution, four sites in the E2 protein are associated with differences between pre- and post-seroconversion sequences. Substitutions at sites 386 (A→V) and 403 (F→L) are associated with seroconversion and exhibit complete compartmentalisation between time points. Substitutions at sites 395 (R→H) and 401 (A→T) appear to show

diversifying selection post-seroconversion, whilst these residues are completely conserved in the window phase. Whilst all window phase derived sequences possess arginine (R) and alanine (A) at positions 395 and 401 respectively, variability at these residues is observed post-seroconversion. Indeed 50% of sequences retain the ancestral amino acid combination (RA) whilst the other half harbours a new, histidine-threonine (HT), combination. Non- synonymous substitutions at 395 and 401 are linked. Unlike Acute 2, all changes associated with seroconversion are located in the HVR1. This region has been previously reported to act as immunological decoy and thus it is likely that these changes have been driven by host humoral immune targeting [77]. This region is also shown to be essential for SR-BI receptor binding [79] and consequently these mutations may have an impact on capacity for cellular entry.

In summary these analyses demonstrate limited diversity is apparent in acute infection and identifies important molecular determinants associated with seroconversion. However these determinants do not appear to be universal, varying from patient to patient and are located in different functionally important regions.


A
E1
E2
nucleotide substitutions
silent/non- silent sybstitutions
transitions/ transversions
B
E1
E2
nucleotide substitutions
silent/non- silent sybstitutions
transitions/ transversions
T
A
C
G
silent
non-silent
transversion
C-T transition
A-G transition

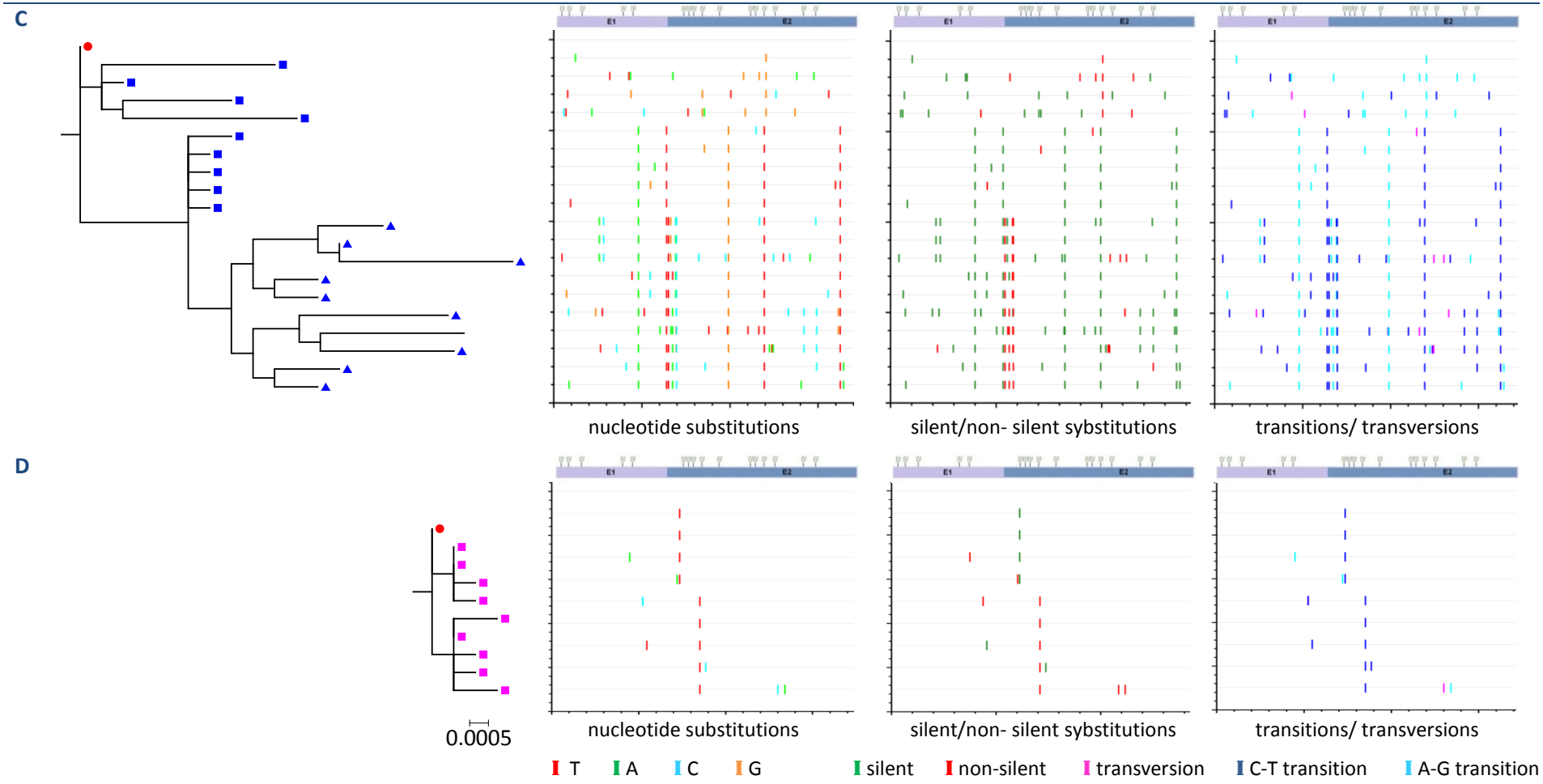


Figure 5.5. Schematic visualisation of distribution of synonymous and non-synonymous substitutions in acute patients.
Comparison made against sequence of MRCA (calculated in PAUP 4.0). Panels A, B, C and D represent sequences from Acute 1, 2, 3 and 4 respectively.

5.1.4. Phenotypic analysis of acute clones

To test whether different sequences within viral populations and between sampling points possessed varying phenotypic properties, E1E2 variants were analysed using the HCV pseudoparticle system. Representative major sequences from acute populations were identified and cloned into pcDNA3.1 background. Unfortunately due to time constraints comparison with minor variants was not possible. Available clones were analysed to confirm presence of a correct sequence or annotate changes (PCR or/and cloning artefacts) and subsequently used in pseudoparticle infectivity assays.

Cloning of SGA products proved to be difficult and error prone. Most of the clones carried at least one amino acid substitution when compared to translated end point titrated sequences (details in Table 5.3). Topoisomerase cloning was initially applied but for this particular set of sequences the reactions had very low efficiency. Subsequently most of the E1E2 clones were obtained by restriction digestion cloning via an in-house optimised method (described in 2.10.2). SGA derived sequences are essentially a consensus of all PCR amplicons in a pool of PCR products and represent the single genome template from which they were derived. However due to inherent error rate of the polymerase a proportion of amplicons will contain artefactual substitutions, which will occur sporadically throughout the length of the sequence. Consequently sporadic PCR induced errors will not be detectable in sequencing as the minimal threshold for detection of minor variants in a chromatograph is approximately 10%. Coupled with the inherent error prone nature of the bacterial replication machinery, which will also introduce sporadic mutations, these are the likely reasons for the high error rate we observe when cloning SGA derived full length E1E2 sequences. This unexpectedly high level of artefactual substitutions associated with

SGA cloning has given us the opportunity to assess the effect on pseudoparticle infectivity.

Table 5.3. Details of clones produced and their sequences.

Sample	Clone label	Sequence differences[1]	Infectivity assessed
restriction cloning			
Acute 1	lig1A20.1_1_c2	V623A	Yes
	lig1A20.1_1_c3	H262R;Y475C	Yes
	lig1A20.1_1_c5	L373V;G524C	Yes
	lig1A20.1_1_c8	L727P	Yes
Acute 2.1	lig1A30.3_26_c9	multiple	Yes
	lig1A30.3_26_c11	C645G;D654G	Yes
	lig1A30.3_26_c14	F292L;C453R	Yes
	lig1A30.3_38_c19	R574G; L736S	Yes
	lig1A30.3_38_c21	N/A	Yes
	lig1A30.3_38_c23	R574G	Yes
	lig1A30.3_C_c37	N411T;T436A;S456C	Yes
Acute 2.2	lig_1A30.12_2_c6	N/A	Yes
Acute 3.1	lig_1A57.2_1_c10	N/A	Yes
	lig_1A57.2_1_c12	N/A	Yes
	lig1A57.2_6_c28	C565R	Yes
Acute 3.2	lig_1A57.14_1.2_1_c20	multiple	No
	lig_1A57.14_1.2_1_c21	N/A	Yes
	lig_1A57.14_1.2_2_c30	N/A	Yes
	lig_1A57.14_1.2_2_c33	N/A	Yes
Acute 4	lig1A65.1_23_c33	L378P	Yes
	lig1A65.1_23_c35	N578S;C735R	Yes
	lig1A65.1_23_c37	no mutations	Yes
	lig1A65.1_23_c40	C185R;P472S	Yes
	lig1A65.1_24_c46	N/A	Yes
Acute 5	lig_6228.8_3_c41	N/A	Yes
	lig_6228.8_3_c42	L727P	Yes
	lig_6228.8_3_c47	N/A	Yes
topo cloning			
Acute 1	topo1A20.1_1	H262R;A320G;H445R	Yes
Acute 2.1	topo1A30.3_26_c7	multiple	Yes
	topo1A30.3_26_c15	P485S	Yes
Acute 3.1	topo1A57.2_1.32.2_c11	L346V	Yes
	topo1A57.2_1.32.2_c14	G279E;G574E	Yes
Acute 3.2	topo1A57.14_1.2.1	multiple	No
Acute 4	topo1A65.1_23_c27	multiple	No
	topo1A65.1_23_c29	S284P	Yes
	topo1A65.1_23_c30	L183P	Yes

N/A- full sequence not available, sequencing reaction not performed for one or more amplicons

[1] Numbering of amino acid position relative to protein start in H77

Amongst 9 clones derived from topoisomerase method all had mutations, only 4 of them singular. The restriction cloning yielded 27 clones, only 17 were fully sequenced owing to time constraints. Only one clone (lig1A65.1_23_c37) proved to possess the correct sequence of the consensus SGA sequence. Despite the presence of mutations a proportion of the clones were utilised in pseudoparticle assays and proved infectious. Furthermore, a number of plasmids possessing one single mutation were reverted by site directed mutagenesis to the SGA consensus (described in section 2.17) and infectivities compared.

Infectivity assays were performed twice in triplicate. The averages of infectivities are presented in Table 5.4 and Figure 3.21 as a percentage of infectivity of H77 positive control, which was taken as 100%. A threshold of 10% infectivity as compared to positive control was set as a cut off for infectious (10% and above) or non- infectious (below 10%) E1E2 clones.

A highly infectious clone lig1A30.3_38_c23 that possessed one mutation when compared to the original end-point titrated consensus sequence was reverted alongside non-infectious clones to compare the effect of reversion. A total of six clones were back mutated and assessed in pseudoparticle infectivity assays (see Table 5.4 and Figure 3.22). A previously infectious clone (lig1A30.3_38_c23) remained functional upon reversion of the artefactual substitution. Infectivities of a further 3 constructs (lig1A20.1_1_c2, topo1A30.3_26_c15, lig_6228.8_3_c42) were brought above the threshold by restoring their sequence to the observed original. Clone topo1A65.1_23_c29 improved infectivity after back mutation, but the overall signal remained below the set threshold.

Table 5.4. Infectivity assay results before and after back-mutation.

Sample	Clone label	infectivity as % H77				
		I	II	Mutation[1]	III	IV
Acute 1	lig1A20.1_1_c2	1.69%	0.17%	C2206T	11.93%	12.35%
	lig1A20.1_1_c3	0.75%	1.35%			
	lig1A20.1_1_c5	0.25%	0.88%			
	lig1A20.1_1_c8	8.13%	12.43%			
	lig1A20.1_29_c1	0.24%	0.56%			
	lig1A20.1_29_c8	0.44%	1.69%			
	topo1A20.1_1	13.31%	18.52%			
Acute 2.1	lig1A30.3_26_c9	1.94%	2.95%			
	lig1A30.3_26_c11	1.19%	0.14%			
	lig1A30.3_26_c14	0.97%	0.08%			
	lig1A30.3_38_c19	5.38%	3.26%			
	lig1A30.3_38_c21	65.45%	75.95%			
	lig1A30.3_38_c23	98.60%	82.76%	G2058A	87.09%	93.62%
	lig1A30.3_1.8_c36	0.65%	0.97%			
	lig1A30.3_2_c29	0.07%	0.81%			
	lig1A30.3_2_c30	0.36%	3.35%			
	lig1A30.3_C_c37	0.92%	0.26%			
	topo1A30.3_26_c7	0.18%	2.09%			
	topo1A30.3_26_c15	3.74%	2.22%	T1791C	112.88%	94.02%
Acute 2.2	lig_1A30.12_2_c6	14.02%	2.67%			
Acute 3.1	lig1A57.2_6_c28	0.74%	0.13%	C2031T	19.42%	22.14%
	lig_1A57.2_1_c10	10.91%	5.04%			
	lig_1A57.2_1_c12	14.86%	1.99%			
	topo1A57.2_1.32.2_c11	25.82%	37.54%			
	topo1A57.2_1.32.2_c14	0.21%	0.67%			
Acute 3.2	lig_1A57.14_1.2_1_c20	13.00%	2.56%			
	lig_1A57.14_1.2_1_c21	16.60%	4.68%			
	lig_1A57.14_1.2_2_c30	98.41%	85.37%			
	lig_1A57.14_1.2_2_c33	19.37%	8.38%			
Acute 4	lig1A65.1_23_c33	0.65%	0.18%			
	lig1A65.1_23_c35	1.00%	3.19%			
	lig1A65.1_23_c37	0.36%	0.39%			
	lig1A65.1_23_c40	0.83%	2.05%			
	lig1A65.1_24_c46	0.91%	0.77%			
	topo1A65.1_23_c29	0.12%	0.11%	T340C	11.60%	9.85%
	topo1A65.1_23_c30	1.30%	0.44%			
Acute 5	lig_6228.8_3_c38	9.89%	3.85%			
	lig_6228.8_3_c41	11.85%	4.66%			
	lig_6228.8_3_c42	11.44%	9.87%	T1670C	13.21%	12.72%
	lig_6228.8_3_c47	11.35%	0.97%			

I, II, II, IV- sequential tests, results presented as an average of triplicate values

[1]Numbering of nucleotide position relative H77 genome

Highlighted in red are values above 10 % threshold

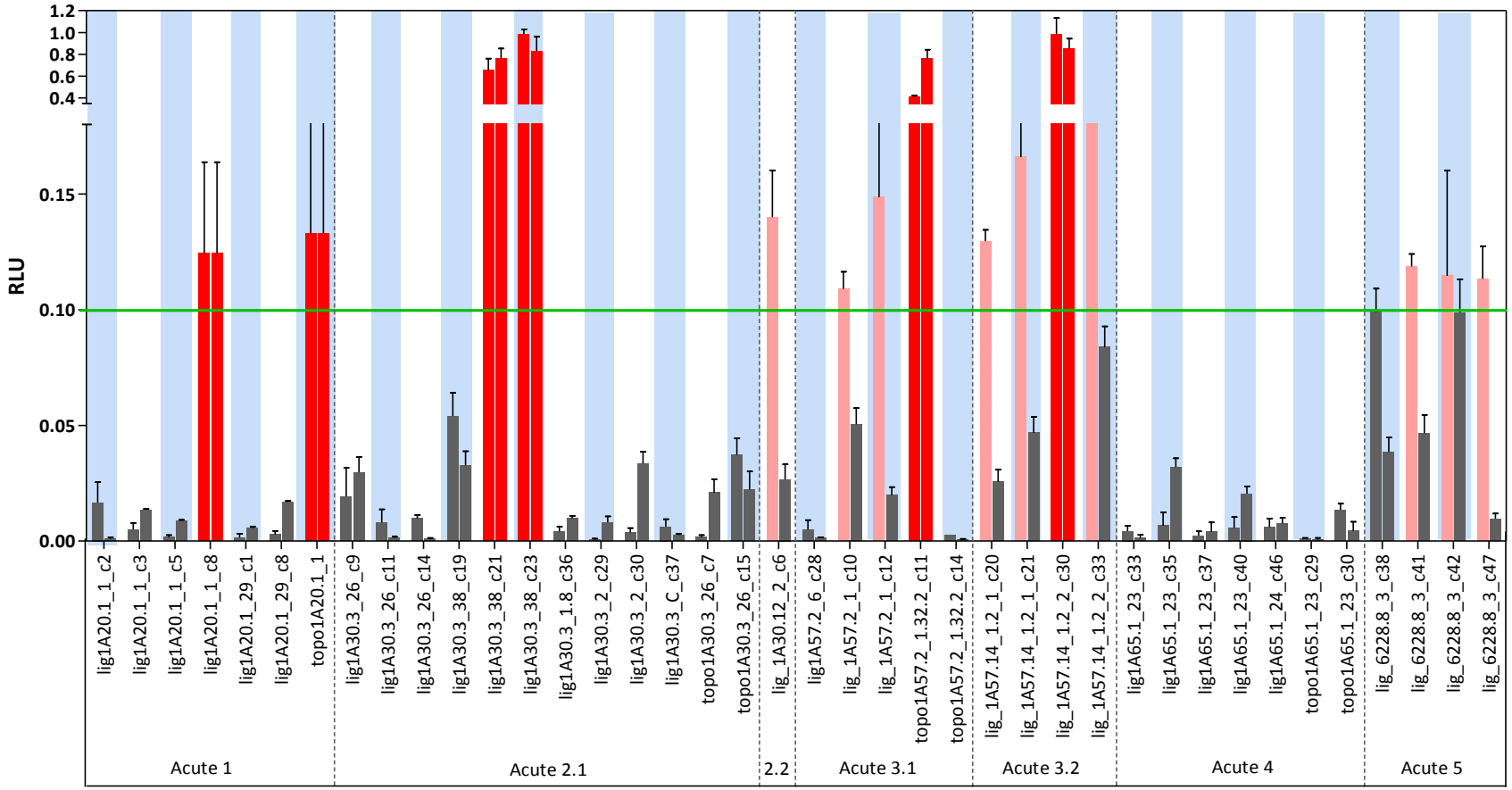


Figure 5.6. Infectivity assay results of unmodified clones.

Key: RLU- relative light units presented as ratio of positive control (positive control=1). Red bars- values above threshold, pink- one of two values above threshold. Green line- infectivity threshold of 10% control (H77).

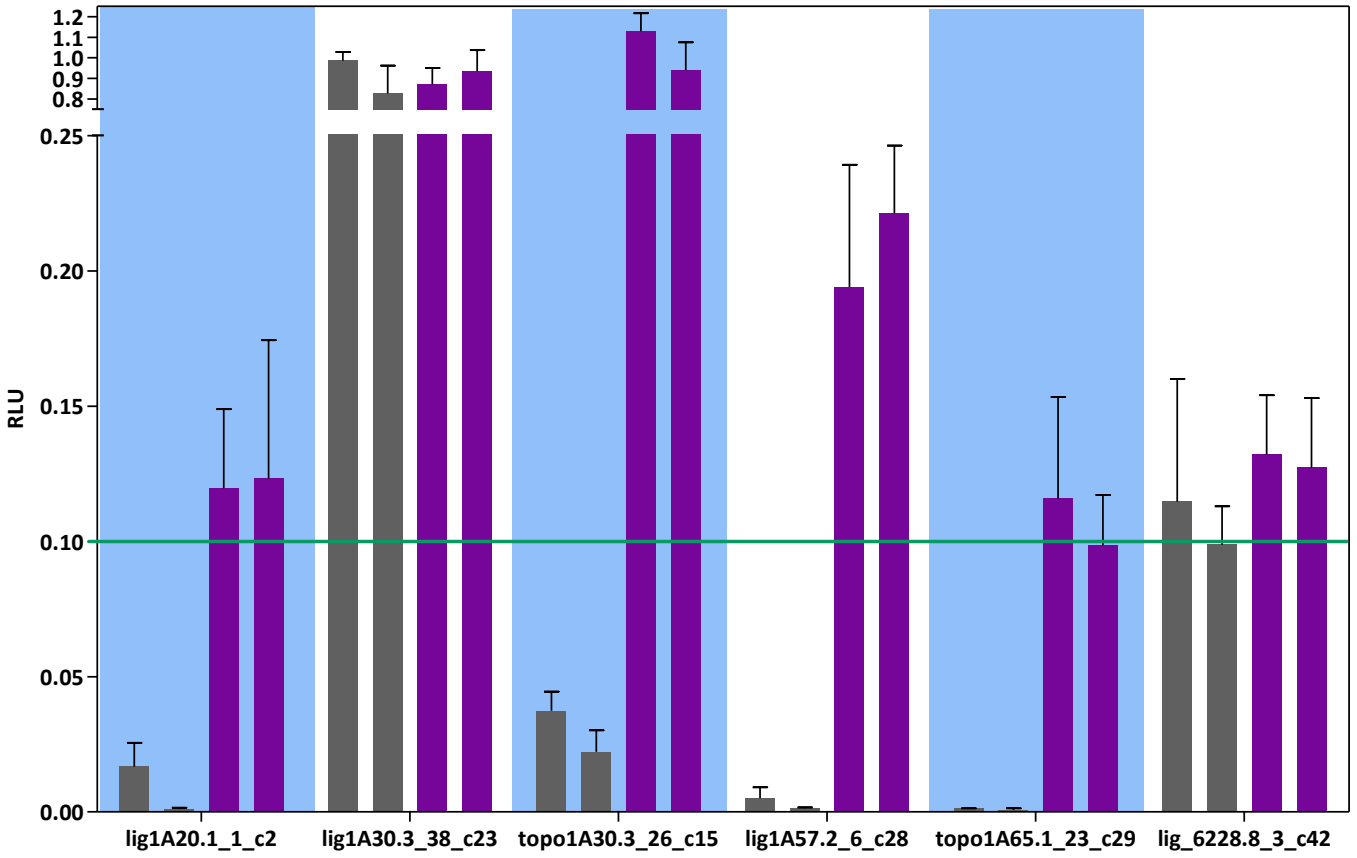


Figure 5.7. Comparison of infectivity of clones prior and post back-mutation.
Key: RLU- relative light units presented as ratio of positive control (positive control=1). Grey bars- two tests before, purple- after reverting mutations. Values as average of three replicas in each test. Green line- infectivity threshold of 10% control (H77).

In conclusion the above data demonstrates the necessity to check the sequences of cloned SGA products prior to any phenotypic analysis. It also reveals the extreme sensitivity of the HCV pseudoparticle system to single point mutations. For example a non- infectious clone topo1A30.3_26_c15 containing a single point mutation resulting in a non- synonymous substitution (P315S) was reverted to SGA consensus. This resulted in a massive increase of infectivity well above the threshold. Unfortunately owing to time constraints and the unexpectedly high level of sporadic mutations, we were unable to further characterise phenotypic properties of E1E2 derived from acute phase of infection.

Finally these data demonstrate the applicability of SGA to the study of intrahost E1E2 viral populations at HCV transmission and in acute and chronic HCV infection. This technique has become the gold standard for analysis of HIV-1 intrahost diversity including analyses of anatomic compartmentalisation [366-368] and identification of number of transmitted variants [220, 369]. This technique has previously not been widely applied to HCV evolutionary studies. However caution must be applied when subsequent cloning of SGA products for phenotypic studies is performed, as this process is error prone: clones must have their sequences confirmed to ensure homology.

5.2. Discussion

The inherent sequence diversity of HCV represents a major challenge for any treatment of this disease. In vivo HCV E1E2 exhibits rapid adaptation to host humoral and cellular responses [16, 134, 228]. Cellular entry is an essential step in the life cycle of any virus and this step may serve as a potential candidate for drug targeting. Consequently, any new prophylactic or therapeutic agent will have to efficiently target structurally or

functionally conserved regions of the viral glycoproteins. Understanding the cellular and viral components of entry, coupled with the evolution of E1E2 in response to host immune targeting, should increase our understanding of the molecular mechanisms involved in the entry process.

The last part of presented project aimed to describe the evolution of HCV envelope genes over several years of infection, including comparison of changes in acute and chronic phases between different patients. We have applied an approach for amplification of full length E1E2 sequences by titration PCR as described here earlier. In brief, our method allows an accurate assessment of the distribution of the viral quasispecies within a specific patient over time. Also, phylogenetic analysis of full length E1E2 sequences aimed to achieve more robust and representative results compared to analyses based only on the HVR1 region.

Reports on tempo of evolutionary change are contrasting with some claiming evolution of HCV is more rapid during acute phase [370], some during chronic phase [218]. Different parts of the HCV genome undergo varying pressure. There has been a substantial interest in studying evolution of E1E2 glycoproteins which facilitate the cell entry and fusion and also undergo selection pressures from the immune system. Accuracy of phylogenetic reconstructions is influenced by the choice of sequence that undergoes analysis and HVR1, being a very short and the most rapidly evolving fragment of the HCV genome, does not allow in depth phylogenetic reconstruction of intrapatient sequence variants. Although HVR1 importance in HCV life cycle as a receptor binding epitope and antibody epitope has been proven, extending reconstructions to full length E1E2 sequences allows inclusion of other structurally and evolutionary important sites. Many receptor binding sites, antibody recognition sites and T-cell targeted epitopes have been characterised for the E1E2 heterodimer, hence analysis of the full length sequence is essential.

In our study the number of substitutions across full length E1E2 clearly increases after seroconversion. HCV E1E2 variants diversity in chronic phase of infection is vastly higher than in acute phase. These findings, suggesting that virus specific antibodies that arise in acute phase during seroconversion and further on drive sequence evolution of the virus, are in accordance with the literature [80, 235, 238]. Furthermore fluctuations of genetic diversity values within population sets at distinct time points of disease might indicate states of dynamic equilibrium between new arising antibody pressure on viral sequences and their adaptation. Moreover phylogenetic analysis shows that the pattern and tempo of substitutions in circulating virus are individual for each host.

The development and implementation of a robust pseudotype cell- entry assay allowed analysis of the E1E2 glycoproteins derived from different genotypes and strains. This method has also been applied to screening for anti-HCV antibodies and neutralising antibodies have been described. Here we aimed to apply the pseudoparticle indectivity assay to study envelopes isolated at distinct disease stages, and further elucidate the relationship between E1E2 evolution, host antibody responses and receptor binding affinity. Unfortunately due to time limitations of the project these aims have been achieved only partially.

Progress of this particular branch of the presented project has been stalled by suprising difficulties related to molecular cloning and high amount of artefact within clonal sequences. However this allowed us to make an observation of a relative sensitivity of the HCV pseudoparticle assay to mutations. This further proves that every precaution needs to be applied when obtaining viral sequences for further analysis of their phenotype. This study was also limited by the number of patient samples obtained and lack of the donor/ source information. Crude analysis of dynamic changes in intrapopulation diversity allows us to hypothesize complex interaction between neutralising antibodies and escape of newly arising variants, further change of antibody

target followed by escape event and transient fixing of a differing variant population. In depth phylogenetic and phenotypic analysis on larger amount of samples, including both infectivity and neutralising assays, would aid understanding of HCV evolution and its molecular determinants in the acute stage of infection.

6. Final conclusion

Over two decades since the discovery of the causative agent of the Non-A-non-B hepatitis, the breadth of the gathered knowledge is still not enough to eradicate the virus from the human population. HCV is a pathogen establishing relatively often a chronic infection. Its high mutation rate and surprisingly late immune response within infected hosts precludes positive outcome of acute infection in more than half of infected individuals. Medical complications related to HCV infection have been an increasing worldwide burden- both moral and financial. Molecular investigation of HCV diversity is of major importance for the design of accurate treatment and development of a preventive vaccine.

In this thesis we attempted to broaden our understanding of HCV evolutionary dynamics in three distinct settings: experimental transmission into immunocompromised mice, naturally occurring horizontal transmission and acute phase of infection.

Overall our findings aid understanding of early post transmission viral glycoprotein populations. Identification of unique features of transmitted E1E2 glycoproteins may potentially be relevant to vaccine and therapeutic design strategies. Beyond this, our study utilises a methodological approach previously applied to evolutionary analyses of HIV-1 intrahost population diversity. Furthermore we confirmed that Alb uPA/SCID mouse model of HCV infection is robust and reliable for HCV transmission studies. These data also suggest universal mechanisms may underly HCV transmission and subsequent establishment of productive infection prior to seroconversion, irrespective of host.

We also confirmed that nosocomial transmission in haemodialysis unit occurred between hypothesised source and index case, highlighting the need for continuous revision of safety guidelines and suggesting possible future change to establishing separate dialysis machines for patients carrying blood borne infections.

Comparison of HCV sequences obtained from human sera at differing sampling time points in the progress of infection provided us with an insight into diversification of viral population at early stages of infection that further sustained the notion that HVR1 alone is not an indicator of evolutionary dynamics of HCV. Forces shaping evolution of HCV during acute infection, especially upon seroconversion are dependent on both the host and viral factors. Knowledge of the initial inoculum and host genetic make-up might in future help predict the course of the disease and aid administering successful treatment.

7. References

1. Feinstone, S.M. and R.H. Purcell, *Non-A, Non-B Hepatitis.* Annual Review of Medicine, 1978. **29**(1): p. 359-366.

2. Purcell, R.H., H. Alter, and J.L. Dienstag, *Non-A, non-B hepatitis.* Yale Journal of Biology and Medicine, 1976. **49**(3): p. 243-250.

3. Choo, Q.L., *Isolation of a cDNA clone derived from a blood borne non-A non-B viral hepatitis genome.* Science, 1989. **244**: p. 359-361.

4. Major, M., B. Rehermann, and S.M. Feinstone, *Hepatitis C Viruses*, in *Fields Virology*, D.M. Knipe and P.M. Howley, Editors. 2001, Lippincott Williams & Wilkins: Philadelphia. p. 1127- 1163.

5. Purcell, R.H., *Hepatits C Virus*. Encyclopedia of Virology, ed. A. Grandorl and R.G. Webster1999, Norwich: Academic Press.

6. Simmonds, P., et al., *A proposed system for the nomenclature of hepatitis C viral genotypes.* Hepatology, 1994. **19**(5): p. 1321-1324.

7. Simmonds, P., et al., *Classification of Hepatitis-C Virus into 6 major genotype and a series of subtypes by phylogenetic analysis of the NS-5 region.* Journal of General Virology, 1993. **74**: p. 2391-2399.

8. Bukh, J., R.H. Miller, and R.H. Purcell, *Biology and genetic heterogeneity of hepatitis C virus.* Clinical and Experimental Rheumatology, 1995. **13**: p. S3-S7.

9. Rehermann, B. and M. Nascimbeni, *Immunology of hepatitis B virus and hepatitis C virus infection.* Nature Reviews. Immunology, 2005. **5**(3): p. 215-229.

10. Cochrane, A., et al., *A genetic analysis of hepatitis Cvirus transmission between injection drug users.* Journal of Infectious Diseases, 2002. **186**(9): p. 1212-1221.

11. Simmonds, P., *Reconstructing the origins of human hepatitis viruses.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2001. **356**(1411): p. 1013-1026.

12. Tong, M.J., et al., *Clinical Outcomes after Transfusion-Associated Hepatitis C.* New England Journal of Medicine, 1995. **332**(22): p. 1463-1466.

13. McAllister, J., et al., *Long-Term Evolution of the Hypervariable Region of Hepatitis C Virus in a Common-Source-Infected Cohort.* Journal of Virology, 1998. **72**(6): p. 4893-4905.

14. Brown, R.S.J., *Hepatitis C and liver transplantation.* Nature, 2005. **436**: p. 973-978.

15. Domínguez-Gil, B. and J.M. Morales, *Transplantation in the patient with hepatitis C.* Transplant International, 2009. **22**(12): p. 1117-1131.

16. Simmonds, P., *Genetic diversity and evolution of hepatitis C virus – 15 years on.* Journal of General Virology, 2004. **85**(11): p. 3173-3188.

17. Kapoor, A., et al., *Characterization of a canine homolog of hepatitis C virus.* Proceedings of the National Academy of Sciences, 2011. **Published online before print May 24, 2011, doi: 10.1073/pnas.1101794108**.

18. Rantala, M. and M.J.v.d. Laar, *Surveillance and epidemiology of hepatitis B and C in Europe - a review.* Eurosurveillance, 2008. **13**(21).

19. *http://www.cks.nhs.uk/hepatitis_c/background_information/prevalence*.

20. Hatzakis, A., et al., *The state of hepatitis B and C in Europe: report from the hepatitis B and C summit conference*.* Journal of Viral Hepatitis, 2011. **18**: p. 1-16.

21. Rao, M., et al., *Further evidence for association of hepatitis C infection with parenteral schistosomiasis treatment in Egypt.* BMC Infectious Diseases, 2002. **2**(1): p. 29.

22. *http://www.who.int/vaccine_research/documents/ViralCancer7.pdf*.

23. *http://commons.wikimedia.org/wiki/Image:HCV_prevalence_1999.png*.

24. Pawlotsky, J.-M., *Use and interpretation of virological tests for hepatitis C.* Hepatology, 2002. **36**(S1): p. S65-S73.

25. Kamal, S.M., *Acute Hepatitis C: A Systematic Review.* American Journal of Gastroenterology, 2008. **103**(5): p. 1283-1297.

26. Moradpour, D., et al., *Molecular virology of hepatitis C.* Praxis, 2002. **91**(51-52): p. 2247-53.

27. Wiegand, J., J. Mossner, and H.L. Tillmann, *Diagnostics and therapy of chronic hepatitis B and C.* Internist, 2008. **49**(7): p. 817-829.

28. Chevaliez, S., *Virological tools to diagnose and monitor hepatitis C virus infection.* Clinical Microbiology and Infection, 2011. **17**(2): p. 116-121.

29. Orland, J.R., T.L. Wright, and S. Cooper, *Acute hepatitis C.* Hepatology, 2001. **33**(2): p. 321-327.

30. Pawlotsky, J.M., *Pathophysiology of hepatitis C virus infection and related liver disease.* Trends in Microbiology, 2004. **12**(2): p. 96-102.

31. Hoofnagle, J.H., *Course and outcome of hepatitis C.* Hepatology, 2002. **36**(5): p. S21-S29.

32. Chen, S.L. and T.R. Morgan, *The Natural History of Hepatitis C Virus (HCV) Infection.* International Journal of Medical Science, 2006. **3**(2): p. 47-52.

33. Tsukiyamakohara, K., et al., *Internal ribosome entry site within hepatitis C virus RNA.* Journal of Virology, 1992. **66**(3): p. 1476-1483.

34. Lindenbach, B.D. and C.M. Rice, *Flaviviridae: The viruses and their replication*. Fields Virology, ed. D.M. Knipe and P.M. Howley2001, Philadelphia: Lippincott Williams & Wilkins. 991- 1042.

35. Grakoui, A., et al., *Expression and identification of hepatitis C virus polyprotein cleavage products.* Journal of Virology, 1993. **67**(3): p. 1385-1395.

36. McLauchlan, J., et al., *Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets.* EMBO Journal 2002. **21**: p. 3980–3988.

37. Choi, J., Z. Xu, and J.-h. Ou, *Triple Decoding of Hepatitis C Virus RNA by Programmed Translational Frameshifting.* Mol. Cell. Biol., 2003. **23**(5): p. 1489-1497.

38. Bartenschlager, R., et al., *Assembly of infectious hepatitis C virus particles.* Trends in Microbiology, 2011. **19**(2): p. 95-103.

39. De Beeck, A.O., et al., *Characterization of functional hepatitis C virus envelope glycoproteins.* Journal of Virology, 2004. **78**(6): p. 2994-3002.

40. Sakai, A., et al., *The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences.* Proceedings of the National Academy of Sciences, 2003. **100**(20): p. 11646-11651.

41. Pavlović, D., et al., *The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives.* Proceedings of the National Academy of Sciences, 2003. **100**(10): p. 6104-6108.

42. Bartenschlager, R., et al., *Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions.* Journal of Virology, 1993. **67**(7): p. 3835-44.

43. Gwack, Y., et al., *Characterization of RNA Binding Activity and RNA Helicase Activity of the Hepatitis C Virus NS3 Protein.* Biochemical and Biophysical Research Communications, 1996. **225**(2): p. 654-659.

44. Sklan, E.H. and J.S. Glenn., *HCV NS4B: From Obscurity to Central Stage*, in *Hepatitis C Viruses: Genomes and Molecular Biology.*, S.L. Tan, Editor 2006, Horizon Bioscience: Norfolk.

45. Dubuisson, J., *Hepatitis C virus proteins.* World Journal of Gastroenterology, 2007. **13**(17): p. 2406-2415.

46. Gale, M.J., *Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein.* Virology, 1997. **230**: p. 217-227.

47. Gale, M.J., et al., *Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation.* Molecular and Cellular Biology, 1998. **18**(9): p. 5208-18.

48. Bartenschlager, R. and V. Lohmann, *Replication of the hepatitis C virus.* Clinical Gastroenterology, 2000. **14**(2): p. 241-54.

49. Bressanelli, S., et al., *Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus.* Proceedings of the National Academy of Sciences, 1999. **96**(23): p. 13034-13039.

50. Lesburg, C.A., et al., *Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site.* Nature Structural and Molecular Biology, 1999. **6**(10): p. 937-943.

51. Rice, C.M., *New Insights Into HCV Replication: Potential Antiviral Target.* Topics in Antiviral Medicine, 2011. **19**(3): p. 117-20.

52. Lozach, P.Y., et al., *DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2.* Journal of Biological Chemistry, 2003. **278**(22): p. 20358-20366.

53. Dubuisson, J. and C.M. Rice, *Hepatitis C virus glycoprotein folding: Disulfide bond formation and association with calnexin.* Journal of Virology, 1996. **70**(2): p. 778-786.

54. Deleersnyder, V., et al., *Formation of native hepatitis C virus glycoprotein complexes.* Journal of Virology, 1997. **71**(1): p. 697-704.

55. Patel, J., A.H. Patel, and J. McLauchlan, *Transmembrane domain of the hepatitis C virus E2 glycoprotein is required for correct folding of the E1 glycoprotein and native complex formation.* Virology, 2001. **279**(1): p. 58-68.

56. Dubuisson, J., et al., *Glycosylation of the hepatitis C virus envelope protein E1 is dependent on the presence of a downstream sequence on the viral polyprotein.* Journal of Biological Chemistry, 2000. **275**(39): p. 30605-30609.

57. Dubuisson, J., *Folding, assembly and subcellular localization of hepatitis C virus glycoproteins.* Hepatitis C Viruses, 2000. **242**: p. 135-148.

58. Choukhi, A., et al., *Involvement of endoplasmic reticulum chaperones in the folding of hepatitis C virus glycoproteins.* Journal of Virology, 1998. **72**(5): p. 3851-3858.

59. Choukhi, A., et al., *Characterization of aggregates of hepatitis C virus glycoproteins.* Journal of General Virology, 1999. **80**: p. 3099-3107.

60. Cocquerel, L., et al., *A retention signal necessary and sufficient for endoplasmic reticulum localization maps to the transmembrane domain of hepatitis C virus glycoprotein E2.* Journal of Virology, 1998. **72**(3): p. 2183-2191.

61. Michalak, J.P., et al., *Characterization of truncated forms of hepatitis C virus glycoproteins.* Journal of General Virology, 1997. **78**: p. 2299-2306.

62. Goffard, A., et al., *Role of N-Linked Glycans in the Functions of Hepatitis C Virus Envelope Glycoproteins.* Journal of Virology, 2005. **79**(13): p. 8400-8409.

63. Goffard, A. and J. Dubuisson, *Glycosylation of hepatitis C virus envelope proteins.* Biochimie, 2003. **85**(3-4): p. 295-301.

64. Meunier, J.C., et al., *Analysis of the glycosylation sites of hepatitis C virus (HCV) glycoprotein E1 and the influence of E1 glycans on the formation of the HCV glycoprotein complex.* Journal of General Virology, 1999. **80**: p. 887-896.

65. Helle, F., et al., *Role of N-Linked Glycans in the Functions of Hepatitis C Virus Envelope Proteins Incorporated into Infectious Virions.* Journal of Virology, 2010. **84**(22): p. 11905-11915.

66. Vieyres, G., et al., *Characterization of the envelope glycoproteins associated with infectious hepatitis C virus.* Journal of Virology, 2010. **84**(19): p. 10159-68.

67. Slater-Handshy, T., et al., *HCV E2 glycoprotein: mutagenesis of N-linked glycosylation sites and its effects on E2 expression and processing.* Virology, 2004. **319**(1): p. 36-48.

68. Nakano, I., et al., *Conformational epitopes detected by cross-reactive antibodies to envelope 2 glycoprotein of the hepatitis C virus.* Journal of Infectious Diseases, 1999. **180**(4): p. 1328-1333.

69. Matsuura, Y., et al., *Processing of E1 and E2 glycoproteins of hepatitis C virus expressed in mammalian and insect cells.* Virology, 1994. **205**(1): p. 141-150.

70. Lavie, M., A. Goffard, and J. Dubuisson, *Assembly of a functional HCV glycoprotein heterodimer.* Current Issues in Molecular Biology, 2007. **9**: p. 71-86.

71. Helle, F., et al., *Role of N-Linked Glycans in the Functions of Hepatitis C Virus Envelope Proteins Incorporated into Infectious Virions.* Journal of Virology, 2007. **84**(22): p. 11905-11915.

72. Duvet, S., et al., *Hepatitis C virus glycoprotein complex localization in the endoplasmic reticulum involves a determinant for retention and not retrieval.* Journal of Biological Chemistry, 1998. **273**(48): p. 32088-32095.

73. Hsu, M., *Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles.* Proceedings of the National Academy of Sciences, 2003. **100**: p. 7271-7276.

74. Voisset, C. and J. Dubuisson, *Functional hepatitis C virus envelope glycoproteins.* Biology of the Cell, 2004. **96**(6): p. 413-420.

75. Krey, T., et al., *The Disulfide Bonds in Glycoprotein E2 of Hepatitis C Virus Reveal the Tertiary Organization of the Molecule.* PLoS Pathogens, 2010. **6**(2).

76. Hijikata, M., et al., *Hypervariable regions in the putative glycoprotein of hepatitis C virus.* Biochemical and Biophysical Research Communications, 1991. **175**(1): p. 220-228.

77. Ray, S.C., et al., *Acute Hepatitis C Virus Structural Gene Sequences as Predictors of Persistent Viremia: Hypervariable Region 1 as a Decoy.* Journal of Virology, 1999. **73**(4): p. 2938-2946.

78. McCaffrey, K., et al., *The variable regions of hepatitis C virus glycoprotein E2 have an essential structural role in glycoprotein assembly and virion infectivity.* Journal of General Virology, 2011. **92**(1): p. 112-121.

79. Scarselli, E., et al., *The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus.* EMBO Journal, 2002. **21**(19): p. 5017-5025.

80. Farci, P., *The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies.* Science, 2000. **288**: p. 339-344.

81. Chevaliez, S. and J.M. Pawlotsky, *HCV Genome and Life Cycle*, in *Hepatitis C Viruses: Genomes and Molecular Biology*, S.L. Tan, Editor 2006, Horizon Bioscience: Norfolk.

82. Joyce, M.A. and D.L.J. Tyrell, *The cell biology of hepatitis C virus.* Microbes and Infection, 2010. **12**(4): p. 263-271.

83. Bartenschlager, R., M. Frese, and T. Pietschmann, *Novel Insights into Hepatitis C Virus Replication and Persistence*, in *Advances in Virus Research*2004, Academic Press. p. 71-180.

84. Cormier, E.G., et al., *L-SIGN (CD209L) and DC-SIGN (0209) mediate transinfection of liver cells by hepatitis C virus.* Proceedings of the National Academy of Sciences, 2004. **101**(39): p. 14067-14072.

85. Gardner, J.P., et al., *L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus.* Proceedings of the National Academy of Sciences, 2003. **100**(8): p. 4498-4503.

86. Agnello, V., et al., *Hepatitis C virus and other Flaviviridae viruses enter cells via low density lipoprotein receptor.* Proceedings of the National Academy of Sciences, 1999. **96**(22): p. 12766-12771.

87. Owen, D.M., et al., *Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor.* Virology, 2009. **394**(1): p. 99-108.

88. Barth, H., et al., *Cellular Binding of Hepatitis C Virus Envelope Glycoprotein E2 Requires Cell Surface Heparan Sulfate.* Journal of Biological Chemistry, 2003. **278**(42): p. 41003-41012.

89. Monazahian, M., et al., *Low density lipoprotein receptor as a candidate receptor for hepatitis C virus.* Journal of Medical Virology, 1999. **57**(3): p. 223-229.

90. Mondelli, M.U., et al., *Hypervariable region 1 of hepatitis C virus: immunological decoy or biologically relevant domain?* Antiviral Research, 2001. **52**(2): p. 153-159.

91. Bernfield, M., et al., *Functions of cell surface heparan sulfate proteoglycans.* Annual Review of Biochemistry, 1999. **68**(1): p. 729-777.

92. Hilgard, P. and R. Stockert, *Heparan sulfate proteoglycans initiate dengue virus infection of hepatocytes.* Hepatology, 2000. **32**(5): p. 1069-1077.

93. Germi, R., et al., *Cellular glycosaminoglycans and low density lipoprotein receptor are involved in hepatitis C virus adsorption.* Journal of Medical Virology, 2002. **68**(2): p. 206-215.

94. Pileri, P., *Binding of hepatitis C virus to CD81.* Science, 1998. **282**: p. 938-941.

95. Drummer, H.E., K.A. Wilson, and P. Poumbourios, *Identification of the Hepatitis C Virus E2 Glycoprotein Binding Site on the Large Extracellular Loop of CD81.* Journal of Virology, 2002. **76**(21): p. 11143-11147.

96. Cocquerel, L., et al., *CD81-dependent binding of hepatitis C virus E1E2 heterodimers.* Journal of Virology, 2003. **77**(19): p. 10677-10683.

97. Bartosch, B., et al., *Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor.* Journal of Biological Chemistry, 2003. **278**(43): p. 41624-41630.

98. Hsu, M., et al., *Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles.* Proceedings of the National Academy of Sciences, 2003. **100**(12): p. 7271-7276.

99. Lavillette, D., et al., *Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus.* Hepatology, 2005. **41**(2): p. 265-274.

100. Petracca, R., et al., *Structure-function analysis of hepatitis C virus envelope-CD81 binding.* Journal of Virology, 2000. **74**(10): p. 4824-4830.

101. Bartosch, B., J. Dubuisson, and F.L. Cosset, *Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes.* Journal of Experimental Medicine, 2003. **197**(5): p. 633-642.

102. Levy, S. and T. Shoham, *Protein-Protein Interactions in the Tetraspanin Web.* Physiology, 2005. **20**(4): p. 218-224.

103. Bitzegeio, J., et al., *Adaptation of Hepatitis C Virus to Mouse CD81 Permits Infection of Mouse Cells in the Absence of Human Entry Factors.* PLoS Pathogens, 2010. **6**(7): p. e1000978.

104. Cormier, E.G., et al., *CD81 is an entry coreceptor for hepatitis C virus.* Proceedings of the National Academy of Sciences, 2004. **101**(19): p. 7270-7274.

105. Catanese, M.T., et al., *Role of SR-BI in HCV entry: kinetics and molecular determinants.* Journal of Virology, 2009.

106. Lavillette, D., et al., *Human Serum Facilitates Hepatitis C Virus Infection, and Neutralizing Responses Inversely Correlate with Viral Replication Kinetics at the Acute Phase of Hepatitis C Virus Infection.* Journal of Virology, 2005. **79**(10): p. 6023-6034.

107. Zeisel, M.B., et al., *Scavenger receptor class B type I is a key host factor for hepatitis C virus infection required for an entry step closely linked to CD81.* Hepatology, 2007. **46**(6): p. 1722-1731.

108. Evans, M.J., et al., *Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry.* Nature, 2007. **446**(7137): p. 801-805.

109. Yang, W., et al., *Correlation of the tight junction-like distribution of Claudin-1 to the cellular tropism of hepatitis C virus.* Journal of Biological Chemistry, 2008. **283**(13): p. 8643-8653.

110. Cukierman, L., et al., *Residues in a Highly Conserved Claudin-1 Motif Are Required for Hepatitis C Virus Entry and Mediate the Formation of Cell-Cell Contacts.* Journal of Virology, 2009. **83**(11): p. 5477-5484.

111. Zheng, A.H., et al., *Claudin-6 and claudin-9 function as additional coreceptors for hepatitis C virus.* Journal of Virology, 2007. **81**(22): p. 12465-12471.

112. Harris, H.J., et al., *Claudin Association with CD81 Defines Hepatitis C Virus Entry.* Journal of Biological Chemistry, 2010. **285**(27): p. 21092-21102.

113. Liu, S., et al., *Tight Junction Proteins Claudin-1 and Occludin Control Hepatitis C Virus Entry and Are Downregulated during Infection To Prevent Superinfection.* Journal of Virology, 2009. **83**(4): p. 2011-2014.

114. Ploss, A., et al., *Human occludin is a hepatitis C virus entry factor required for infection of mouse cells.* Nature, 2009. **457**: p. 882-886.

115. Michta, M.L., et al., *Species-Specific Regions of Occludin Required by Hepatitis C Virus for Cell Entry.* Journal of Virology, 2010. **84**(22): p. 11696-11708.

116. Blanchard, E., et al., *Hepatitis C Virus Entry Depends on Clathrin-Mediated Endocytosis.* Journal of Virology, 2006. **80**(14): p. 6964-6972.

117. Tscherne, D.M., et al., *Time- and Temperature-Dependent Activation of Hepatitis C Virus for Low-pH-Triggered Entry.* Journal of Virology, 2006. **80**(4): p. 1734-1741.

118. Edwards, V.C., et al., *The role of neutralizing antibodies in hepatitis C virus infection.* Journal of General Virology, 2012. **93**(1): p. 1-19.

119. Neumann, A.U., et al., *Hepatitis C Viral Dynamics in Vivo and the Antiviral Efficacy of Interferon- Therapy.* Science, 1998. **282**(5386): p. 103-107.

120. Saito, T. and M.J. Gale, *Principles of intracellular viral recognition.* Current Opinion in Immunology, 2007. **19**(1): p. 17-23.

121. Kanneganti, T.D., et al., *Critical role for Cryopyrin/Nalp3 in activation of caspase-1 in response to viral infection and double-stranded RNA.* Journal of Biological Chemistry, 2006. **281**(48): p. 36560-8.

122. Horner, S.M. and M.J. Gale, *Intracellular Innate Immune Cascades and Interferon Defenses That Control Hepatitis C Virus.* Journal of Interferon & Cytokine Research, 2009. **29**(9): p. 489-498.

123. Seki, E. and D.A. Brenner, *Toll-like receptors and adaptor molecules in liver disease: update.* Hepatology, 2008. **48**(1): p. 322-35.

124. Kato, H., et al., *Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses.* Nature, 2006. **441**(7089): p. 101-5.

125. Keynan, Y., et al., *The role of regulatory T cells in chronic and acute viral infections.* Clinical Infectious Diseases, 2008. **46**(7): p. 1046-1052.

126. Yoneyama, M., et al., *The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses.* Nature Immunology, 2004. **5**(7): p. 730-7.

127. Saito, T., et al., *Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2.* Proceedings of the National Academy of Sciences, 2007. **104**(2): p. 582-7.

128. Sumpter, R.J., et al., *Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I.* Journal of Virology, 2005. **79**(5): p. 2689-99.

129. Wang, C., et al., *Alpha interferon induces distinct translational control programs to suppress hepatitis C virus RNA replication.* Journal of Virology, 2003. **77**(7): p. 3898-912.

130. Uzri, D. and L. Gehrke, *Nucleotide sequences and modifications that determine RIG-I/RNA binding and signaling activities.* Journal of Virology, 2009. **83**(9): p. 4174-84.

131. Qu, L. and S.M. Lemon, *Hepatitis A and Hepatitis C Viruses: Divergent Infection Outcomes Marked by Similarities in Induction and Evasion of Interferon Responses.* Seminars in Liver Disease, 2010. **30**(4): p. 319-332.

132. Li, X.-D., et al., *Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity.* Proceedings of the National Academy of Sciences, 2005. **102**(49): p. 17717-17722.

133. Loo, Y.-M., et al., *Viral and therapeutic control of IFN-β promoter stimulator 1 during hepatitis C virus infection.* Proceedings of the National Academy of Sciences, 2006. **103**(15): p. 6001-6006.

134. Rehermann, B., *Hepatitis C virus versus innate and adaptive immune responses: a tale of coevolution and coexistence.* The Journal of Clinical Investigation, 2009. **119**(7): p. 1745-1754.

135. Pham, T.N.Q., et al., *Hepatitis C Virus Persistence after Spontaneous or Treatment-Induced Resolution of Hepatitis C.* Journal of Virology, 2004. **78**(11): p. 5867-5874.

136. Cooper, S., et al., *Analysis of a successful immune response against hepatitis C virus.* Immunity, 1999. **10**(4): p. 439-49.

137. Lechner, F., et al., *Analysis of successful immune responses in persons infected with hepatitis C virus.* Journal of Experimental Medicine, 2000. **191**(9): p. 1499-512.

138. Thimme, R., et al., *Determinants of viral clearance and persistence during acute hepatitis C virus infection.* Journal of Experimental Medicine, 2001. **194**(10): p. 1395-406.

139. Farci, P., et al., *Lack of protective immunity against reinfection with hepatitis C virus.* Science, 1992. **258**(5079): p. 135-40.

140. Helle, F., et al., *The Neutralizing Activity of Anti-Hepatitis C Virus Antibodies Is Modulated by Specific Glycans on the E2 Envelope Protein.* Journal of Virology, 2007. **81**(15): p. 8101-8111.

141. Tarr, A.W., et al., *Naturally Occurring Antibodies That Recognize Linear Epitopes in the Amino Terminus of the Hepatitis C Virus E2 Protein Confer Noninterfering, Additive Neutralization.* Journal of Virology, 2012. **86**(5): p. 2739-2749.

142. Farci, P., et al., *Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization.* Proceedings of the National Academy of Sciences, 1994. **91**(16): p. 7792-6.

143. Tarr, A.W., et al., *Hepatitis C Patient-Derived Glycoproteins Exhibit Marked Differences in Susceptibility to Serum Neutralizing Antibodies: Genetic Subtype Defines Antigenic but Not Neutralization Serotype.* Journal of Virology, 2011. **85**(9): p. 4246-4257.

144. Owsianka, A., et al., *Monoclonal Antibody AP33 Defines a Broadly Neutralizing Epitope on the Hepatitis C Virus E2 Envelope Glycoprotein.* Journal of Virology, 2005. **79**(17): p. 11095-11104.

145. Perotti, M., et al., *Identification of a Broadly Cross-Reacting and Neutralizing Human Monoclonal Antibody Directed against the Hepatitis C Virus E2 Protein.* Journal of Virology, 2008. **82**(2): p. 1047-1052.

146. Friedman, E., et al., *Factors associated with premature discontinuation of chronic hepatitis C virus treatment using pegylated interferon and ribavirin: A four year experience.* Gastroenterology, 2008. **134**(4): p. A775-A775.

147. Wedemeyer, H., M. Cornberg, and M.P. Manns, *PEG-interferons: significance for the treatment of viral hepatitis B and C.* Deutsche Medizinische Wochenschrift, 2001. **126**: p. S68-S75.

148. Pawlotsky, J.M., *Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C.* Antiviral Research, 2003. **59**(1): p. 1-11.

149. Pawlotsky, J.M., *The nature of interferon-alpha resistance in hepatitis C virus infection.* Current Opinion in Infectious Diseases, 2003. **16**(6): p. 587-592.

150. Liang, T.J. and T. Heller, *Pathogenesis of hepatitis C-associated hepatocellular carcinoma.* Gastroenterology, 2004. **127**: p. S62-S71.

151. Zeuzem, S., et al., *Expert opinion on the treatment of patients with chronic hepatitis C.* Journal of Viral Hepatitis, 2008. **16**(2): p. 75-90.

152. Ge, D., et al., *Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance.* Nature, 2009. **461**(7262): p. 399-401.

153. Suppiah, V., et al., *IL28B is associated with response to chronic hepatitis C interferon-[alpha] and ribavirin therapy.* Nature Genetics, 2009. **41**(10): p. 1100-1104.

154. Tanaka, Y., et al., *Genome-wide association of IL28B with response to pegylated interferon-[alpha] and ribavirin therapy for chronic hepatitis C.* Nature Genetics, 2009. **41**(10): p. 1105-1109.

155. Pawlotsky, J.M., S. Chevaliez, and J.G. McHutchison, *The Hepatitis C Virus Life Cycle as a Target for New Antiviral Therapies.* Gastroenterology, 2007. **132**(5): p. 1979-1998.

156. Soriano, V., et al., *Directly acting antivirals against hepatitis C virus.* Journal of Antimicrobial Chemotherapy, 2011. **66**(8): p. 1673-1686.

157. Sarrazin, C. and J. Vermehren, *New HCV therapies on the horizon.* Clinical Microbiology and Infection, 2011. **17**: p. 122–134.

158. Reesink, H.W., et al., *Rapid Decline of Viral RNA in Hepatitis C Patients Treated With VX-950: A Phase Ib, Placebo-Controlled, Randomized Study.* Gastroenterology, 2006. **131**(4): p. 997-1002.

159. Sarrazin, C., et al., *Dynamic Hepatitis C Virus Genotypic and Phenotypic Changes in Patients Treated With the Protease Inhibitor Telaprevir.* Gastroenterology, 2007. **132**(5): p. 1767-1777.

160. Sarrazin, C., et al., *SCH 503034, a Novel Hepatitis C Virus Protease Inhibitor, Plus Pegylated Interferon α-2b for Genotype 1 Nonresponders.* Gastroenterology, 2007. **132**(4): p. 1270-1278.

161. Le Pogam, S., et al., *RG7128 Alone or in Combination with Pegylated Interferon-α2a and Ribavirin Prevents Hepatitis C Virus (HCV) Replication and Selection of Resistant Variants in HCV-Infected Patients.* Journal of Infectious Diseases, 2010. **202**(10): p. 1510-1519.

162. Gao, M., et al., *Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect.* Nature, 2010. **465**(7294): p. 96-100.

163. Fridell, R.A., et al., *Distinct Functions of NS5A in Hepatitis C Virus RNA Replication Uncovered by Studies with the NS5A Inhibitor BMS-790052.* Journal of Virology, 2011. **85**(14): p. 7312-7320.

164. Targett-Adams, P., et al., *Small Molecules Targeting Hepatitis C Virus-Encoded NS5A Cause Subcellular Redistribution of Their Target: Insights into Compound Modes of Action.* Journal of Virology, 2011. **85**(13): p. 6353-6368.

165. Chayama, K., et al., *Dual therapy with the nonstructural protein 5A inhibitor, daclatasvir, and the nonstructural protein 3 protease inhibitor, asunaprevir, in hepatitis C virus genotype 1b–infected null responders.* Hepatology, 2012. **55**(3): p. 742-748.

166. Corouge, M. and S. Pol, *New treatments for chronic hepatitis C virus infection.* Médecine et Maladies Infectieuses, 2011. **41**(11): p. 579-587.

167. Lavillette, D., et al., *Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus.* Hepatology, 2005. **41**(2): p. 265-274.

168. Helle, F., et al., *Cyanovirin-N Inhibits Hepatitis C Virus Entry by Binding to Envelope Protein Glycans.* Journal of Biological Chemistry, 2006. **281**(35): p. 25177-25183.

169. Law, M., et al., *Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge.* Nature Medicine, 2008. **14**(1): p. 25-27.

170. Farci, P., *Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization.* Proceedings of the National Academy of Sciences, 1994. **91**: p. 7792-7796.

171. Schinazi, R.F., L. Bassit, and C. Gavegnano, *HCV drug discovery aimed at viral eradication.* Journal of Viral Hepatitis, 2010. **17**(2): p. 77-90.

172. McAleer, W.J., et al., *Human hepatitis B vaccine from recombinant yeast.* Nature, 1984. **307**(5947): p. 178-80.

173. Frey, S.E., et al., *Safety and immunogenicity of HCV E1E2 vaccine adjuvanted with MF59 administered to healthy adults.* Vaccine, 2010. **28**(38): p. 6367-6373.

174. Habersetzer, F., T.F. Baumert, and F. Stoll-Keller, *GI-5005, a yeast vector vaccine expressing an NS3-core fusion protein for chronic HCV infection.* Current Opinion in Molecular Therapy, 2009. **11**(4): p. 456-462.

175. Firbas, C., et al., *Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: A randomized, placebo controlled trial for dose optimization in 128 healthy subjects.* Vaccine, 2006. **24**(20): p. 4343-4353.

176. Klade, C.S., et al., *Therapeutic Vaccination of Chronic Hepatitis C Nonresponder Patients With the Peptide Vaccine IC41.* Gastroenterology, 2008. **134**(5): p. 1385-1395.e1.

177. Yutani, S., et al., *Phase I clinical study of a personalized peptide vaccination for patients infected with hepatitis C virus (HCV) 1b who failed to respond to interferon-based therapy.* Vaccine, 2007. **25**(42): p. 7429-7435.

178. Sandström, E., et al., *Broad Immunogenicity of a Multigene, Multiclade HIV-1 DNA Vaccine Boosted with Heterologous HIV-1 Recombinant Modified Vaccinia Virus Ankara.* Journal of Infectious Diseases, 2008. **198**(10): p. 1482-1490.

179. Habersetzer, F., et al., *A Poxvirus Vaccine Is Safe, Induces T-Cell Responses, and Decreases Viral Load in Patients With Chronic Hepatitis C.* Gastroenterology, 2011. **141**(3): p. 890-899.e4.

180. Barnes, E., et al., *Novel Adenovirus-Based Vaccines Induce Broad and Sustained T Cell Responses to HCV in Man.* Science Translational Medicine, 2012. **4**(115): p. 115ra1.

181. Liver, E.A.f.t.S.o.t., *First Evidence For DNA-based Vaccination Against Chronic Hepatitis C*, in *Science Daily*2009, European Association for the Study of the Liver.

182. Kumar, U., et al., *SEQUENCE VARIATION IN THE LARGE ENVELOPE GLYCOPROTEIN (E2/NS1) OF HEPATITIS-C VIRUS DURING CHRONIC INFECTION.* Journal of Infectious Diseases, 1993. **167**(3): p. 726-730.

183. Farci, P. and R.H. Purcell, *Clinical significance of hepatitis C virus genotypes and quasispecies.* Seminars in Liver Disease, 2000. **20**(1): p. 103-126.

184. Yee, T.T., et al., *The natural history of HCV in a cohort of haemophilic patients infected between 1961 and 1985.* Gut, 2000. **47**(6): p. 845-851.

185. Mazzeo, C., et al., *Ten year incidence of HCV infection in northern Italy and frequency of spontaneous viral clearance.* Gut, 2003. **52**(7): p. 1030-1034.

186. Kumar, D., et al., *Response of combination therapy on viral load and disease severity in chronic hepatitis C.* Digestive Diseases and Sciences, 2008. **53**(4): p. 1107-1113.

187. Frank, C., et al., *The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt.* Lancet, 2000. **355**(9207): p. 887-891.

188. Smith, D.B., et al., *The origin of hepatitis C virus genotypes.* Journal of General Virology, 1997. **78**(2): p. 321-8.

189. Rispeter, K., et al., *Hepatitis c virus variability: Sequence analysis of an isolate after 10 years of chronic infection.* Virus Genes, 2000. **21**(3): p. 179-188.

190. Smith, D.B. and P. Simmonds, *Characteristics of nucleotide substitution in the hepatitis C virus genome: Constraints on sequence change in coding regions at both ends of the genome.* Journal of Molecular Evolution, 1997. **45**(3): p. 238-246.

191. Smith, D.B., *Evolution of the hypervariable region of hepatitis C virus.* Journal of Viral Hepatitis, 1999. **6**: p. 41-46.

192. Brown, R.J.P., et al., *Evolutionary dynamics of hepatitis C virus envelope genes during chronic infection.* Journal of General Virology, 2005. **86**(7): p. 1931-1942.

193. Kato, N., et al., *Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection.* Journal of Virology, 1994. **68**(8): p. 4776-4784.

194. Kimura, M., *Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles.* Genetical Research, 1968. **11**(3): p. 247-69.

195. Sheridan, I., et al., *High-Resolution Phylogenetic Analysis of Hepatitis C Virus Adaptation and Its Relationship to Disease Progression.* Journal of Virology, 2004. **78**(7): p. 3447-3454.

196. Sobolev, B.N., et al., *Comparative analysis of amino acid sequences from envelope proteins isolated from different hepatitis C virus variants: possible role of conservative and variable regions.* Journal of Viral Hepatitis, 2000. **7**(5): p. 368-374.

197. Holmes, E.C. and A. Moya, *Is the Quasispecies Concept Relevant to RNA Viruses?* Journal of Virology, 2002. **76**(1): p. 460-462.

198. Holmes, E.C., *The RNA Virus Quasispecies: Fact or Fiction?* Journal of Molecular Biology, 2010. **400**(3): p. 271-273.

199. Smith, D.B., et al., *Virus 'quasispecies': making a mountain out of a molehill?* Journal of General Virology, 1997. **78**: p. 1511–1519.

200. Vignuzzi, M., et al., *Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population.* Nature, 2006. **439**(7074): p. 344-348.

201. Love, A., et al., *Evolution of hepatitis C virus variants following blood transfusion from one infected donor to several recipients: a long-term follow-up.* Journal of General Virology, 2004. **85**(2): p. 441-450.

202. Casino, C., et al., *Variation of hepatitis C virus following serial transmission: multiple mechanisms of diversification of the hypervariable region and evidence*

for convergent genome evolution. Journal of General Virology, 1999. **80**(3): p. 717-725.

203. Carrere-Kremer, S., et al., *Subcellular localization and topology of the p7 polypeptide of hepatitis C virus.* Journal of Virology, 2002. **76**(8): p. 3720-3730.

204. Buchbinder, S.P., et al., *Hepatitis C virus infection in sexually active homosexual men.* Journal of Infection, 1994. **29**(3): p. 263-269.

205. Cavalheiro, N.d.P., et al., *Hepatitis C: sexual or intrafamilial transmission? Epidemiological and phylogenetic analysis of hepatitis C virus in 24 infected couples.* Revista da Sociedade Brasileira de Medicina Tropical, 2009. **42**: p. 239-244.

206. Weiner, A.J., et al., *A unique, predominant hepatitis C virus variant found in an infant born to a mother with multiple variants.* Journal of Virology, 1993. **67**(7): p. 4365-4368.

207. Miller, F.D. and L.J. Abu-Raddad, *Evidence of intense ongoing endemic transmission of hepatitis C virus in Egypt.* Proceedings of the National Academy of Sciences, 2010. **107**(33): p. 14757-14762.

208. Pybus, O.G., et al., *The hepatitis C virus epidemic among injecting drug users.* Infection, Genetics and Evolution, 2005. **5**(2): p. 131-139.

209. Pybus, O.G., et al., *Investigating the endemic transmission of the hepatitis C virus.* International Journal for Parasitology, 2007. **37**(8-9): p. 839-849.

210. Mizuno, Y., et al., *Study of needlestick accidents and hepatitis C virus infection in healthcare workers by molecular evolutionary analysis.* Journal of Hospital Infection, 1997. **35**(2): p. 149-154.

211. Goldberg, D. and E. Anderson, *Hepatitis C: who is at risk and how do we identify them?* Journal of Viral Hepatitis, 2004. **11**: p. 12-18.

212. Manzin, A., et al., *Evolution of Hypervariable Region 1 of Hepatitis C Virus in Primary Infection.* Journal of Virology, 1998. **72**(7): p. 6271-6276.

213. Lin, H.J., et al., *Occurrence of identical hypervariable region 1 sequences of hepatitis C virus in transfusion recipients and their respective blood donors: Divergence over time.* Hepatology, 2001. **34**(2): p. 424-429.

214. Manzin, A., et al., *Dominant Role of Host Selective Pressure in Driving Hepatitis C Virus Evolution in Perinatal Infection.* Journal of Virology, 2000. **74**(9): p. 4327-4334.

215. Sugitani, M. and T. Shikata, *Comparison of amino acid sequences in hypervariable region-1 of hepatitis C virus clones between human inocula and the infected chimpanzee sera.* Virus Research, 1998. **56**(2): p. 177-182.

216. Nainan, O.V., et al., *Selective transmission of hepatitis C virus genotypes and quasispecies in humans and experimentally infected chimpanzees.* Journal of General Virology, 2006. **87**(1): p. 83-91.

217. Ray, S.C., et al., *Hypervariable Region 1 Sequence Stability during Hepatitis C Virus Replication in Chimpanzees.* Journal of Virology, 2000. **74**(7): p. 3058-3066.

218. Chen, S. and Y.-m. WANG, *Multigene tracking of quasispecies in viral persistence and clearance of hepatitis C virus.* World Journal of Gastroenterology, 2005. **11**(19): p. 2874- 2884.

219. Derdeyn, C.A., et al., *Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission.* Science, 2004. **303**(5666): p. 2019-22.

220. Salazar-Gonzalez, J.F., et al., *Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing.* Journal of Virology, 2008. **82**(8): p. 3952-3970.

221. Li, H., et al., *High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men.* PLoS Pathogens, 2011. **6**(5): p. e1000890.

222. Bar, K.J., et al., *Wide variation in the multiplicity of HIV-1 infection among injection drug users.* J Virol, 2010. **84**(12): p. 6241-7.

223. Wang, G.P., et al., *Hepatitis C Virus Transmission Bottlenecks Analyzed by Deep Sequencing.* J. Virol. **84**(12): p. 6218-6228.

224. Wang, G.P., et al., *Hepatitis C Virus Transmission Bottlenecks Analyzed by Deep Sequencing.* Journal of Virology, 2010. **84**(12): p. 6218-6228.

225. Thimme, R., V. Lohmann, and F. Weberc, *A target on the move: Innate and adaptive immune escape strategies of hepatitis C virus.* Antiviral Research, 2006. **69**(3): p. 129-41.

226. Callendret, B., et al., *Transmission of Clonal Hepatitis C Virus Genomes Reveals the Dominant but Transitory Role of CD8+ T Cells in Early Viral Evolution.* Journal of Virology, 2011. **85**(22): p. 11833-11845.

227. Bowen, D.G. and C.M. Walker, *Mutational escape from CD8+ T cell immunity.* The Journal of Experimental Medicine, 2005. **201**(11): p. 1709-1714.

228. Merani, S., et al., *Effect of immune pressure on hepatitis C virus evolution: Insights from a single-source outbreak.* Hepatology, 2011. **53**(2): p. 396-405.

229. Pircher, H., et al., *Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo.* Nature, 1990. **346**(6285): p. 629-633.

230. Phillips, R.E., et al., *Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition.* Nature, 1991. **354**(6353): p. 453-459.

231. Borrow, P., et al., *Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.* Nature Medicine, 1997. **3**: p. 205-211.

232. Goulder, P.J., et al., *Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS.* Nature Medicine, 1997. **3**(212-217).

233. Erickson, A.L., *The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes.* Immunity, 2001. **15**: p. 883-895.

234. Chang, K.M., *Immunological significance of cytotoxic T lymphocyte epitope variants in patients chronically infected by the hepatitis C virus.* Journal of Clinical Investigation, 1997. **100**: p. 2376-2385.

235. Dowd, K.A., et al., *Selection Pressure From Neutralizing Antibodies Drives Sequence Evolution During Acute Infection With Hepatitis C Virus.* Gastroenterology, 2009. **136**(7): p. 2377-2386.

236. Fafi-Kremer, S., et al., *Viral entry and escape from antibody-mediated neutralization influence hepatitis C virus reinfection in liver transplantation.* Journal of Experimental Medicine, 2010. **207**(9): p. 2019-31.

237. Chen, M., *Limited humoral immunity in hepatitis C virus infection.* Gastroenterology, 1999. **116**: p. 135-143.

238. Logvinoff, C., *Neutralizing antibody response during acute and chronic hepatitis C virus infection.* Proceedings of the National Academy of Sciences, 2004. **101**: p. 10149-10154.

239. Brown, R.J.P., et al., *Cross-genotype characterization of genetic diversity and molecular adaptation in hepatitis C virus envelope glycoprotein genes.* Journal of General Virology, 2007. **88**(2): p. 458-469.

240. Liu, L., et al., *Acceleration of Hepatitis C Virus Envelope Evolution in Humans Is Consistent with Progressive Humoral Immune Selection during the Transition from Acute to Chronic Infection.* Journal of Virology, 2010. **84**(10): p. 5067-5077.

241. von Hahn, T., et al., *Hepatitis C Virus Continuously Escapes From Neutralizing Antibody and T-Cell Responses During Chronic Infection In Vivo.* Gastroenterology, 2007. **132**(2): p. 667-678.

242. Gerlach, J.T., *Recurrence of hepatitis C virus after loss of virus-specific CD4+ T-cell response in acute hepatitis C.* Gastroenterology, 1999. **117**: p. 933-941.

243. Thimme, R., *CD8+ T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection.* Journal of Virology, 2003. **77**: p. 68-76.

244. Smyk-Pearson, S., et al., *Spontaneous recovery in acute human hepatitis C virus infection: Functional T-cell thresholds and relative importance of CD4 help.* Journal of Virology, 2008. **82**(4): p. 1827-1837.

245. Wedemeyer, H., *Impaired effector function of hepatitis C virus-specific CD8+ T cells in chronic hepatitis C virus infection.* Journal of Immunology, 2002. **169**: p. 3447-3458.

246. Kittlesen, D.J., et al., *Interaction between complement receptor gC1qR and hepatitis C virus core protein inhibits T-lymphocyte proliferation.* Journal of Clinical Investigation, 2000. **106**: p. 1239-1249.

247. Large, M.K., D.J. Kittlesen, and Y.S. Hahn, *Suppression of host immune response by the core protein of hepatitis C virus: possible implications for hepatitis C virus persistence.* Journal of Immunology, 1999. **162**: p. 931-938.

248. Taylor, D.R., et al., *Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein.* Science, 1999. **285**: p. 107-110.

249. Foy, E., *Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease.* Science, 2003. **300**: p. 1145-1148.

250. Sarasin-Filipowicz, M., et al., *Interferon signaling and treatment outcome in chronic hepatitis C.* Proceedings of the National Academy of Sciences, 2008. **105**(19): p. 7034-9.

251. Erickson, A.K., S. Seiwert, and M.J. Gale, *Antiviral potency analysis and functional comparison of consensus interferon, interferon-alpha2a and pegylated interferon-alpha2b against hepatitis C virus infection.* Antiviral Therapy, 2008. **163**(7): p. 851-862.

252. Hollinger, F.B., et al., *Non-A, non-B hepatitis transmission in chimpanzees: a project of the transfusion-transmitted viruses study group.* Intervirology, 1978. **10**(1): p. 60-8.

253. Okamoto, H., et al., *Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability.* Virology, 1992. **190**(2): p. 894-9.

254. Kato, T., et al., *Interferon-inducible gene expression in chimpanzee liver infected with hepatitis C virus.* Virology, 1992. **190**(2): p. 856-60.

255. Choo, Q.L., et al., *Vaccination of chimpanzees against infection by the hepatitis C virus.* Proceedings of the National Academy of Sciences, 1994. **91**(4): p. 1294-8.

256. Bassett, S.E., et al., *Effects of iron loading on pathogenicity in hepatitis C virus-infected chimpanzees.* Hepatology, 1999. **29**(6): p. 1884-1892.

257. Bukh, J., et al., *Host range studies of GB virus-B hepatitis agent, the closest relative of hepatitis C virus, in New World monkeys and chimpanzees.* Journal of Medical Virology, 2001. **65**(4): p. 694-697.

258. Lanford, R.E., et al., *The chimpanzee model of hepatitis C virus infections.* ILAR Journal, 2001. **42**(2): p. 117-26.

259. Bassett, S.E., et al., *Viral persistence, antibody to E1 and E2, and hypervariable region 1 sequence stability in hepatitis C virus-inoculated chimpanzees.* Journal of Virology, 1999. **73**(2): p. 1118-1126.

260. Thomson, M., et al., *The clearance of hepatitis C virus infection in chimpanzees may not necessarily correlate with the appearance of acquired immunity.* Journal of Virology, 2003. **77**(2): p. 862-870.

261. Bassett, S.E., K.M. Brasky, and R.E. Lanford, *Analysis of hepatitis C virus-inoculated chimpanzees reveals unexpected clinical profiles.* Journal of Virology, 1998. **72**(4): p. 2589-99.

262. Beames, B., et al., *Development of a primary tamarin hepatocyte culture system for GB virus-B: a surrogate model for hepatitis C virus.* Journal of Virology, 2000. **74**(24): p. 11764-72.

263. Xie, Z.C., et al., *Transmission of hepatitis C virus infection to tree shrews.* Virology, 1998. **244**(2): p. 513-20.

264. Karayiannis, P., et al., *Experimental infection of Tamarins with human non-A, non-B hepatitis virus.* Journal of Medical Virology, 1983. **11**(3): p. 251-6.

265. Watanabe, T., et al., *Studies on transmission of human non-A, non-B hepatitis to marmosets.* Journal of Medical Virology, 1987. **22**(2): p. 143-56.

266. Bukh, J. and C.L. Apgar, *Five new or recently discovered (GBV-A) virus species are indigenous to New World monkeys and may constitute a separate genus of the Flaviviridae.* Virology, 1997. **229**(2): p. 429-36.

267. Grakoui, A., H.L. Hanson, and C.M. Rice, *Bad time for Bonzo? Experimental models of hepatitis C virus infection, replication, and pathogenesis.* Hepatology, 2001. **33**(3): p. 489-95.

268. Gale, M. and M.R. Beard, *Molecular clones of hepatitis C virus: applications to animal models.* ILAR Journal, 2001. **42**(2): p. 139-51.

269. Dorner, M., et al., *A genetically humanized mouse model for hepatitis C virus infection.* Nature, 2011. **474**(7350): p. 208-211.

270. Eren, et al., *Human monoclonal antibodies specific to hepatitis B virus generated in a human/mouse radiation chimera: the Trimera system.* Immunology, 1998. **93**(2): p. 154-161.

271. Ilan, E., et al., *The Hepatitis C Virus (HCV)–Trimera Mouse: A Model for Evaluation of Agents against HCV.* Journal of Infectious Diseases, 2002. **185**(2): p. 153-161.

272. Masciopinto, F., et al., *Expression of Human CD81 in Transgenic Mice Does Not Confer Susceptibility to Hepatitis C Virus Infection.* Virology, 2002. **304**(2): p. 187-196.

273. Higginbottom, A., et al., *Identification of amino acid residues in CD81 critical for interaction with hepatitis C virus envelope glycoprotein E2.* Journal of Virology, 2000. **74**(8): p. 3642-9.

274. Legrand, N., et al., *Humanized Mice for Modeling Human Infectious Disease: Challenges, Progress, and Outlook.* Cell Host and Microbe, 2009. **6**(1): p. 5-9.

275. Kneteman, N.M. and D.F. Mercer, *Mice with chimeric human livers: Who says supermodels have to be tall?* Hepatology, 2005. **41**(4): p. 703-706.

276. Lindenbach, B.D., et al., *Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro.* Proceedings of the National Academy of Sciences, 2006. **103**(10): p. 3805-3809.

277. Meuleman, P. and G. Leroux-Roels, *The human liver-uPA-SCID mouse: A model for the evaluation of antiviral compounds against HBV and HCV.* Antiviral Research, 2008. **80**(3): p. 231-238.

278. Brass, V., D. Moradpour, and H.E. Blum, *Molecular virology of hepatitis C virus (HCV): 2006 update.* International Journal of Medical Science, 2006. **3**(2): p. 29-34.

279. Lohmann, V., et al., *Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line.* Science, 1999. **285**: p. 110-113.

280. Ikeda, M., et al., *Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells.* Journal of Virology, 2002. **76**(6): p. 2997-3006.

281. Lohmann, V., et al., *Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line.* Science, 1999. **285**(5424): p. 110-113.

282. Pietschmann, T., et al., *Persistent and transient replication of full-length hepatitis C virus genomes in cell culture.* Journal of Virology, 2002. **76**(8): p. 4008-4021.

283. Baumert, T.F., et al., *Hepatitis C virus structural proteins assemble into viruslike particles in insect cells.* Journal of Virology, 1998. **72**(5): p. 3827-3836.

284. Lambot, M., et al., *Reconstitution of hepatitis C virus envelope glycoproteins into liposomes as a surrogate model to study virus attachment.* Journal of Biological Chemistry, 2002. **277**(23): p. 20625-20630.

285. Op De Beeck, A., et al., *Characterization of functional hepatitis C virus envelope glycoproteins.* Journal of Virology, 2004. **78**(6): p. 2994-3002.

286. Op De Beeck, A., et al., *Characterization of Functional Hepatitis C Virus Envelope Glycoproteins.* J. Virol., 2004. **78**(6): p. 2994-3002.

287. Sandrin, V., et al., *Assembly of functional hepatitis C virus glycoproteins on infectious pseudoparticles occurs intracellularly and requires concomitant incorporation of E1 and E2 glycoproteins.* Journal of General Virology, 2005. **86**(12): p. 3189-3199.

288. Cocquerel, L., C. Voisset, and J. Dubuisson, *Hepatitis C virus entry: potential receptors and their biological functions.* Journal of General Virology, 2006. **87**: p. 1075-1084.

289. *http://www.john-libbey-eurotext.fr/en/revues/medecine/epi/e-docs/00/04/3D/4B/article.md?fichier=images.htm*.

290. Lanford, R., et al., *Demonstration of in vitro infection of chimpanzee hepatocytes with hepatitis C virus using strand-specific RT/PCR.* Virology, 1994. **202**: p. 606-614.

291. Fournier, C., et al., *In vitro infection of adult normal human hepatocytes in primary culture by hepatitis C virus.* Journal of General Virology, 1998. **79**: p. 2367-2374.

292. Dash, S., et al., *Transfection of HepG2 cells with infectious hepatitis C virus genome.* American Journal of Pathology, 1997. **151**(2): p. 363-373.

293. Ikeda, M., et al., *Human hepatocyte clonal cell lines that support persistent replication of hepatitis C virus.* Virus Research, 1998. **56**(2): p. 157-167.

294. Wakita, T., et al., *Production of infectious hepatitis C virus in tissue culture from a cloned viral genome.* Nature Medicine, 2005. **11**(7): p. 791-796.

295. Kato, N., *Molecular virology of hepatitis C virus.* Acta Medica Okayama, 2001. **55**(3): p. 133-159.

296. Diaz, O., et al., *Preferential association of Hepatitis C virus with apolipoprotein B48-containing lipoproteins.* Journal of General Virology, 2006. **87**: p. 2983-2991.

297. Lindenbach, B.D., et al., *Complete replication of hepatitis C virus in cell culture.* Science, 2005. **309**(5734): p. 623-626.

298. Zhong, J., et al., *Robust hepatitis C virus infection in vitro.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(26): p. 9294-9299.

299. Bukh, J. and R.H. Purcell, *A milestone for hepatitis C virus research: a virus generated in cell culture is fully viable in vivo.* Proceedings of the National Academy of Sciences, 2006. **103**(10): p. 3500-1.

300. Villano, A.S., et al., *Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection.* Hepatology, 1999. **29**(3): p. 908-914.

301. Wohnsland, A., W.P. Hofmann, and C. Sarrazin, *Viral Determinants of Resistance to Treatment in Patients with Hepatitis C.* Clinical Microbiology Reviews, 2007. **20**(1): p. 23-38.

302. Hoofnagle, J.H., *A Step Forward in Therapy for Hepatitis C.* New England Journal of Medicine, 2009. **360**(18): p. 1899-1901.

303. De Francesco, R. and A. Carfí, *Advances in the development of new therapeutic agents targeting the NS3-4A serine protease or the NS5B RNA-dependent RNA polymerase of the hepatitis C virus.* Advanced Drug Delivery Reviews, 2007. **59**(12): p. 1242-1262.

304. Lin, K., et al., *VX-950, a Novel Hepatitis C Virus (HCV) NS3-4A Protease Inhibitor, Exhibits Potent Antiviral Activities in HCV Replicon Cells.* Antimicrobial Agents and Chemotherapy, 2006. **50**(5): p. 1813-1822.

305. Jarczak, D., et al., *Hairpin ribozymes in combination with siRNAs against highly conserved hepatitis C virus sequence inhibit RNA replication and protein translation from hepatitis C virus subgenomic replicons.* FEBS Journal, 2005. **272**(22): p. 5910-5922.

306. Hügle, T. and A. Cerny, *Current therapy and new molecular approaches to antiviral treatment and prevention of hepatitis C.* Reviews in Medical Virology, 2003. **13**(6): p. 361-371.

307. Oh, T.S. and C.M. Rice, *Predicting response to hepatitis C therapy.* The Journal of Clinical Investigation, 2009. **119**(1): p. 5-7.

308. McHutchison, J.G., et al., *Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection.* N Engl J Med, 2009. **360**(18): p. 1827-38.

309. Hezode, C., et al., *Telaprevir and peginterferon with or without ribavirin for chronic HCV infection.* N Engl J Med, 2009. **360**(18): p. 1839-50.

310. Bukh, J., R.H. Miller, and R.H. Purcell, *Genetic-Heterogeneity of Hepatitis-C Virus - Quasi-Species and Genotypes.* Seminars in Liver Disease, 1995. **15**(1): p. 41-63.

311. Lindenbach, B.D. and C.M. Rice, *Unravelling hepatitis C virus replication from genome to function.* Nature, 2005. **436**(7053): p. 933-938.

312. Drummer, H.E. and P. Poumbourios, *Hepatitis C Virus Glycoprotein E2 Contains a Membrane-proximal Heptad Repeat Sequence That Is Essential for E1E2 Glycoprotein Heterodimerization and Viral Entry.* Journal of Biological Chemistry, 2004. **279**(29): p. 30066-30072.

313. Keck, Z.Y., et al., *Hepatitis C virus E2 has three immunogenic domains containing conformational epitopes with distinct properties and biological functions.* Journal of Virology, 2004. **78**(17): p. 9224-9232.

314. Pileri, P., et al., *Binding of hepatitis C virus to CD81.* Science, 1998. **282**(5390): p. 938-941.

315. Weiner, A.J., et al., *Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections.* Proceedings of the National Academy of Sciences, 1992. **89**(8): p. 3468–3472.

316. Peter C. Thomson, C.W., Celia Aitken, Jonathan Ball, Natalia Wysocka, Richard Brown and a.R.S. Rodger, *A case of hepatitis C virus transmission acquired through sharing a haemodialysis machine.* Nephrology Dialysis Transplantation Plus, 2010(First published online: October 15, 2010).

317. Thomson, P.C., et al., *A case of hepatitis C virus transmission acquired through sharing a haemodialysis machine.* Nephrology Dialysis Transplantation Plus, 2011. **4**(1): p. 32-35.

318. Mohsen, A.H. and T.H.S. Group, *The epidemiology of hepatitis C in a UK health regional population of 5.12 million.* Gut, 2001. **48**(5): p. 707-713.

319. Boom, R., et al., *Rapid and simple method for purification of nucleic acids.* Journal of Clinical Microbiology, 1990. **28**(3): p. 495-503.

320. Keele, B.F., et al., *Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.* Proceedings of the National Academy of Sciences, 2008. **105**(21): p. 7552-7557.

321. Simmonds, P., et al., *Human Immunodeficiency Virus-Infected Individuals Contain Provirus In Small Numbers Of Peripheral Mononuclear-Cells And At Low Copy Numbers.* Journal of Virology, 1990. **64**(2): p. 864-872.

322. Bar, K.J., et al., *Wide Variation in the Multiplicity of HIV-1 Infection among Injection Drug Users.* Journal of Virology, 2010. **84**(12): p. 6241-6247.

323. Ramachandran, S., et al., *End-point limiting-dilution real-time PCR assay for evaluation of hepatitis C virus quasispecies in serum: Performance under optimal and suboptimal conditions.* Journal of Virological Methods, 2008. **151**(2): p. 217-224.

324. Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*. 3 ed. Vol. 3. 2001: Cold Spring Harbor Laboratory Press.

325. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* Journal of Molecular Biology, 1975. **94**(3): p. 441-446.

326. Rosenblum, B.B., et al., *New dye-labeled terminators for improved DNA sequencing patterns.* Nucleic Acids Research, 1997. **25**(22): p. 4500-4504.

327. Tamura, K., et al., *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0.* Molecular Biology and Evolution, 2007. **24**(8): p. 1596-1599.

328. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.

329. Laemmli, U.K., *Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-685.

330. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Molecular Biology and Evolution, 1987. **4**(4): p. 406-425.

331. Motoo, K., *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences* Journal of Molecular Evolution, 1980. **16**: p. 111-120.

332. Kimura, M., *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.* Journal of Molecular Evolution, 1980. **16**(2): p. 111-20.

333. Efron, B., Halloran, E., Holmes, S. , *Bootstrap confidence levels for phylogenetic trees.* Proceedings of the National Academy of Sciences, 1996. **93**: p. 7085-7090.

334. Swofford, D.L., *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4.04beta*2003: Sinauer Associates, Sunderland, Massachusetts.

335. Posada, D. and K.A. Crandall, *MODELTEST: testing the model of DNA substitution.* Bioinformatics, 1998. **14**(9): p. 817-818.

336. Mee, C.J., et al., *Hepatitis C virus infection reduces hepatocellular polarity in a vascular endothelial growth factor-dependent manner.* Gastroenterology, 2010. **138**(3): p. 1134-42.

337. Graham, F.L., et al., *Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5.* Journal of General Virology, 1977. **36**(1): p. 59-72.

338. Nakabayashi, H., Taketa,T., Miyano,K., Yamane,T., and Sato,J., *Growth of human hepatoma cell lines with differentiated functions in chemically defined medium.* Cancer Research, 1982. **42**: p. 3858-3863.

339. Shapiro, A.L., E. Viñuela, and J. V. Maizel, *Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels.* Biochemical and Biophysical Research Communications, 1967. **28**(5): p. 815-820.

340. Burnette, W.N., *"Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A.* Analytical Biochemistry, 1981. **112**(2): p. 195-203.

341. Masciopinto, F., et al., *Association of hepatitis C virus envelope proteins with exosomes.* European Journal of Immunology, 2004. **34**(10): p. 2834-2842.

342. Clayton, R.F., et al., *Analysis of Antigenicity and Topology of E2 Glycoprotein Present on Recombinant Hepatitis C Virus-Like Particles.* Journal of Virology, 2002. **76**(15): p. 7672-7682.

343. Smith, P.K., et al., *Measurement of protein using bicinchoninic acid.* Analytical Biochemistry, 1985. **150**(1): p. 76-85.

344. Van Weemen, B.K. and A.H.W.M. Schuurs, *Immunoassay using antigen--enzyme conjugates.* FEBS Letters, 1971. **15**(3): p. 232-236.

345. Johansson, D.X., et al., *Human combinatorial libraries yield rare antibodies that broadly neutralize hepatitis C virus.* Proceedings of the National Academy of Sciences, 2007. **104**(41): p. 16269-16274.

346. Tamura, K., et al., *MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.* Molecular Biology and Evolution, 2011. **28** p. 2731-2739.

347. Goffard, A., et al., *Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins.* Journal of Virology, 2005. **79**(13): p. 8400-9.

348. Park, I.-W., et al., *Hepatitis C virus is restricted at both entry and replication in mouse hepatocytes.* Biochemical and Biophysical Research Communications, 2009. **387**(3): p. 489-493.

349. Fissell, R.B., et al., *Patterns of hepatitis C prevalence and seroconversion in hemodialysis units from three continents: The DOPPS.* Kidney International, 2004. **65**(6): p. 2335-2342.

350. Jadoul, M., et al., *The changing epidemiology of hepatitis C virus (HCV) infection in haemodialysis: European multicentre study.* Nephrology Dialysis Transplantation, 2004. **19**(4): p. 904-909.

351. Sarich, V.M. and A.C. Wilson, *Immunological Time Scale for Hominid Evolution.* Science, 1967. **158**(3805): p. 1200-1203.

352. Ciesielski, C., et al., *Transmission of human immunodeficiency virus in a dental practice.* Annals of Internal Medicicne, 1992. **15**(116): p. 798-805.

353. Ou, C.-Y., et al., *Molecular Epidemiology of HIV Transmission in a Dental Practice.* Science, 1992. **256**(5060): p. 1165-1171.

354. Machuca, R., et al., *Molecular Investigation of Transmission of Human Immunodeficiency Virus Type 1 in a Criminal Case.* Clinical and Diagnostic Laboratory Immunology, 2001. **8**(5): p. 884-890.

355. Scaduto, D.I., et al., *Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences.* Proceedings of the National Academy of Sciences, 2010. **107**(50): p. 21242-21247.

356. Metzker, M.L., et al., *Molecular evidence of HIV-1 transmission in a criminal case.* Proceedings of the National Academy of Sciences, 2002. **99**(22): p. 14292-14297.

357. Power, J.P., et al., *Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin.* Lancet, 1995. **345**(8959): p. 1211-1213.

358. Johnson, D.W., et al., *Frequencies of hepatitis B and C infections among haemodialysis and peritoneal dialysis patients in Asia-Pacific countries: analysis of registry data.* Nephrology Dialysis Transplantation, 2009. **24**(5): p. 1598-1603.

359. Jadoul, M., C. Cornu, and C.v.Y. de Strihou, *Universal precautions prevent hepatitis C virus transmission: A 54 month follow-up of the Belgian multicenter study.* Kidney International, 1998. **53**(4): p. 1022-1025.

360. dos Santos, P.J., et al., *Impact of dialysis room and reuse strategies on the incidence of hepatitis C virus infection in haemodialysis units.* Nephrology Dialysis Transplantation, 1996. **11**(10): p. 2017-2022.

361. Okuda, K., et al., *Mode of hepatitis C infection not associated with blood transfusion among chronic hemodialysis patients.* Journal of Hepatology, 1995. **23**(1): p. 28-31.

362. KDIGO, *Guideline 3: Preventing HCV transmission in hemodialysis units.* Kidney International, 2008. **73**(S109): p. S46-S52.

363. Geddes, C., E. Lindley, and N. Duncan, *UK Renal Association Clinical Practice Guidelines—Blood-borne Viruses.*, 2009, UK Renal Association.

364. Major, M.E., et al., *Long-Term Follow-Up of Chimpanzees Inoculated with the First Infectious Clone for Hepatitis C Virus.* Journal of Virology, 1999. **73**(4): p. 3317-3325.

365. Harris, K.A. and C.G. Teo, *Diversity of Hepatitis C Virus Quasispecies Evaluated by Denaturing Gradient Gel Electrophoresi.* Clininal and Diagnostic Laboratory Immunology, 2001. **8**(1): p. 62-73.

366. Brown, R.J.P., et al., *Intercompartmental Recombination of HIV-1 Contributes to env Intrahost Diversity and Modulates Viral Tropism and Sensitivity to Entry Inhibitors.* Journal of Virology, 2011. **85**(12): p. 6024-6037.

367. Anderson, J.A., et al., *HIV-1 Populations in Semen Arise through Multiple Mechanisms.* PLoS Pathog, 2010. **6**(8): p. e1001053.

368. Schnell, G., et al., *HIV-1 Replication in the Central Nervous System Occurs in Two Distinct Cell Types.* PLoS Pathog, 2011. **7**(10): p. e1002286.

369. Li, H., et al., *High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men.* PLoS Pathog, 2010. **6**(5): p. e1000890.

370. Cantaloube, J.-F., et al., *Evolution of hepatitis C virus in blood donors and their respective recipients.* Journal of General Virology, 2003. **84**(2): p. 441-446.