Development of techniques for the isolation and characterisation of human monoclonal antibodies from Hepatitis C virus infected individuals

Victoria C. Edwards BSc, MSc

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

December 2012

Abstract

Infection with hepatitis C virus (HCV) is cleared spontaneously in only 20% of cases with the majority of individuals developing a chronic infection. This discrepancy in disease outcome is incompletely understood but current understanding of the immune response to HCV suggests that rapid induction of a broadly neutralising antibody (nAb) response leads to resolution of acute infection. The majority of nAb identified target the envelope glycoproteins, particularly E2, and most appear to inhibit binding of E2 to the cellular receptor CD81. Antibodies targeting other interactions, such as those with the receptor CLDN or the fusion determinant, are underrepresented in the repertoire of anti-HCV antibodies. However, the antibody discovery process may have been biased by the nature of the assays used. Therefore new assays are needed to enable the discovery and characterisation of antibodies in an unbiased manner. To facilitate this, a novel insect cell display library was developed for mapping antibody-binding epitopes. Cells expressing specific E2 mutants provided the necessary proof-of-principle that loss of antibody binding could be detected in this system before a library expressing randomly mutated E2 was developed. Sorting experiments demonstrated that single cells could be isolated and enriched based on loss of antibody binding. Secondly, a method for characterising the immunoglobulin (Ig) genes of HCV infected patients was developed; Ig genes were isolated from small numbers of B cells and the sequences analysed. Finally, a patient cohort was studied with a view to investigating the evolution of the antibody response during early infection. The unreliable nature of the samples prevented such analysis; however a DNA fingerprinting method of testing the origin and relatedness of serum samples

was developed. This will improve the reliability of future studies. Together these methods provide a work model for the assessment of samples and the isolation and characterisation of antibodies.

Publications

Edwards, V.C., Tarr, A.W., Urbanowicz, R.A. and Ball, J.K. (2012) The role of neutralizing antibodies in hepatitis C virus infection. *J. Gen. Virol*. **93**; 1-19

Edwards V.C., McClure, C.P., Thomson, E., Kamal, S., Irving, W.I. and Ball, J,K. Use of short-tandem repeat (STR) fingerprinting to validate a cohort of samples from acute phase hepatitis C virus infection. Manuscript in preparation.

Acknowledgments

Firstly I would like to thank Professor Jonathan Ball for providing me with the opportunity to carry out this research project. The project has been interesting but challenging and I am very grateful for his guidance in all aspects of the work and the opportunity to develop my scientific, writing and presentational skills. I would also like to thank all members of the Virus Research Group, past and present, for their support, encouragement and practical advice in the lab and for helping to create a pleasant working environment. Particular thanks go to Dr. Richard Urbanowicz for developing the B cell isolation and culturing protocols. Within the School of Molecular Medical Sciences I would like to thank Dr. Adrian Robins for his expertise in setting up and troubleshooting the flow cytometry and cell sorting protocols. I would also like to thank Sue Bainbridge for her practical advice regarding the testing and analysis of STR fingerprint data.

I would like to thank several people who have provided reagents, without which this project would not have been possible; Dr. Thomas Krey (Pasteur Institute, Paris) for the *Drosophila melanogaster* S2 cells and the pMT plasmid, Dr. Jialing Liu, (UCSF, USA) for the pYD1 plasmid, Marlene Schweitzer, (BOKU, Austria) for the *Saccharomyces cerevisiae* EBY100 cells, Dr. Emma Thomson (MRC Virus Research Unit, Glasgow) for the mixed acute serum samples, Dr. Sanaa Kamal (Ain Shams University, Cairo) for the cohort of patient samples from Egypt and Joy Kean (Gartnavel General Hospital, Glasgow) for carrying out antibody avidity testing. I would also like to thank the Trent HCV Cohort from whom B cells were isolated, Nottingham Digestive Diseases Centre/Biomedical Research Unit and the MRC for funding.

Finally I would to thank my family and friends for their practical and emotional support throughout my PhD, particularly my mum and stepdad who have been a huge source of encouragement during the writing-up period. I am very grateful to everyone who has helped make this experience an enjoyable and fulfilling one.

Contents

Abbreviations

5’ RACE	5’ rapid amplification of cDNA ends
aa	amino acid
Ab	antibody
ADCC	antibody-dependent cellular cytotoxicity
AI	avidity index
AID	activation-induced cytidine deaminase
ALT	alanine transaminase
AP	alkaline phosphatase
apo	apolipoprotein
AR	antigenic region
ASC	antibody secreting cell
APS	ammonium persulphate
ATP	adenosine triphosphate
BSA	bovine serum albumin
cDNA	complementary DNA
CDC	complement-dependent cytotoxicity
CDR	complementarity determining region
CH	constant heavy
CHO	Chinese hamster ovary cell line
CL	constant light
CLDN	claudin
CMV	Cytomegalovirus
CSR	class-switch recombination
DC-SIGN	dendritic cell-specific intracellular adhesion molecule-3-grabbing non-integrin
DENV	Dengue virus
DEPC	diethyl pyrocarbonate
DES	*Drosophila* expression system
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
ddNTPs	dideoxynucleotide triphosphates
DTT	dithiothreitol
EBV	Epstein Barr virus
EDTA	ethylene diamine tetraacetic acid
EGFR	epidermal growth factor receptor
EL	extracellular loop
ELISA	enzyme linked immunosorbant assay
EphA2	ephrin receptor A2
Fab	fragment antigen binding
FBS	foetal bovine serum

Fc	fragment crystallisable
FcR	Fc receptor
FI	fluorescence intensity
FR	framework region
GAG	glycosaminoglycan
GNA	*Galanthis nivalis* agglutinin
GFP	green fluorescence protein
HA	haemagglutinin
HAT	hypoxanthine/aminopterin/thymidine
HA_{TM}	haemagglutinin transmembrane (plus cytoplasmic domain)
HBIG	hepatitis B virus immunoglobulin
HCFC1	human host cell factor C1
HCV	hepatitis C virus
HCVcc	HCV cell culture
HCVpp	HCV pseudoparticle
HDL	high-density lipoprotein
HEK 293T	human embryonic kidney 293T cell line
HIV	human immunodeficiency virus
HRP	horseradish peroxidase
HS	heparin sulphate
HVR	hypervariable region
IE	immediate early promoter
IFN	interferon
Ig	immunoglobulin
igVR	intergenotypic variable region
kDa	kilodalton
IL	interleukin
IRF	interferon regulatory factor
LDL/LDLr	low-density lipoprotein/low density lipoprotein receptor
LEL	large extracellular loop
L-SIGN	liver-specific intracellular adhesion molecule-3-grabbing non-integrin
LVP	lipoviral particles
MAb	monoclonal antibody
MAC	membrane attack complex
MB	molecular biology
MMLV	Moloney's murine leukaemia virus
mRNA	messenger RNA
nAb	neutralising antibody
NCR	non-coding region (also referred to as UTR)
NF-κB	nuclear factor kappa B
NHEJ	non-homologous end joining
NOB	neutralisation-of-binding
NPC1L1	Niemann-Pick C1-like 1

NTP	nucleoside triphosphate
OCLN	occludin
PBMCs	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PBS-T	PBS-Tween
PBS-TB	PBS-T containing BSA
PCR	polymerase chain reaction
PEDEL-AA	Programme for estimating diversity in error-prone PCR libraries
pNPP	*p*-Nitrophenyl phosphate
qPCR	quantitative PCR
RAG	recombination activating gene
RIG-I	retinoic acid inducible gene I
RNA	ribonucleic acid
RSS	recombination signal sequence
RT	reverse transcriptase
scFv	single chain Fv
SDS-PAGE	sodium dodecyl-polyacrylamide gel electrophoresis
sE2	soluble E2
SEL	small extracellular loop
SHM	somatic hypermutation
siRNA	small interfering RNA
SR-BI	scavenger receptor class B type I
STR	short tandem repeat
SVR	sustained virological response
TBS-T	Tris buffered saline-Tween 20
TdT	terminal deoxynucleotidyl transferase
TEMED	tetramethylethylene diamine
TLR	Toll-like receptor
UTR	untranslated region (also referred to as NCR)
VDJ	variable-diversity-joining (in terms of Ig genes)
VH	variable heavy
VL	variable light
VLDL	very-low-density lipoprotein
WNV	West Nile virus

1 General Introduction

1.1 Natural History

Hepatitis C virus (HCV) was first identified as the cause of non-A, non-B hepatitis in the late 1980s (Choo *et al.*, 1989; Kuo *et al.*, 1989) and is currently estimated to infect 170 million people worldwide (Ascione *et al.*, 2007) with up to 3 million new infections per year (Anon., 1999). In 20-30% of infections, the virus is cleared spontaneously (Santantonio *et al.*, 2008); however in the majority of cases the virus persists. The reasons for this discrepancy in disease outcome are poorly understood but host genetic polymorphisms have been implicated following the discovery of a protective variant of the IL28B gene. This gene encodes the type III interferon (IFN), IFN-λ3, and the presence of a C allele correlates with spontaneous resolution of infection (Thomas *et al.*, 2009). This variant has also been associated with treatment induced response to HCV (Ge *et al.*, 2009). In those who fail to clear the virus, chronic HCV infection can lead to cirrhosis of the liver, and in some cases hepatocellular carcinoma (HCC). Cirrhosis develops in 20% of chronic infections (Anon., 1999) and approximately 500,000 cases of HCV-associated HCC occur each year (Ascione *et al.*, 2007). Chronic HCV infection is the leading cause of liver transplantation in developed countries (Shepard *et al.*, 2005) but as yet, there is no way to prevent reinfection of the newly grafted liver.

1.2 Transmission

HCV is transmitted primarily through exposure to contaminated blood or blood products, organ transplantations, the use of unsterilized medical equipment and needle-stick injury to healthcare workers. In developed countries, the most common route of transmission is via intravenous drug use and the sharing of drug paraphernalia (Alter, 2007). The role of sexual transmission in the spread of HCV is debatable. There does not appear to be a significant risk of transmission between long-term, monogamous, heterosexual partners of HCV-positive individuals. HCV infection risk increases with the number of sexual partners, although this may be confounded by increased use of injecting drugs in these groups. The greatest association appears to be with human immunodeficiency virus (HIV)-coinfection (Tohme & Holmberg, 2010). Perinatal transmission can occur, although this is rare, and there is no association between breast-feeding and HCV transmission (Alter, 2007).

1.3 Diversity and distribution of HCV

HCV can be classified on the basis of its genome sequence into genotypes and subtypes. Six major genotypes of HCV have been well described (Simmonds *et al.*, 2005); more recently a seventh genotype was identified and has been added to the phylogenetic tree of HCV diversity (Nakano *et al.*, 2012). Genotypes differ by >30% at the nucleotide level. HCV can be further categorised into different subtypes, showing 20-25% sequence divergence (Simmonds *et al.*, 2005). Further diversity is observed between viral variants circulating within a single host. Due to the error-prone nature of the virus-

encoded RNA-dependent RNA polymerase and the high replication rate of the virus (10^{12} virions per day) (Neumann *et al.*, 1998), a highly diverse population of viral variants is generated. This has been described as a 'quasispecies' (Martell *et al.*, 1992), although the ability of HCV to form a true 'quasispecies' in nature has been called into question (Holmes, 2010).

The global distribution of different HCV genotypes varies. Genotypes 1 and 3 are the most common genotypes in Northern Europe and the USA and are associated with intravenous drug use. Genotype 2 is found in the Mediterranean and the Far East. Genotype 4 is dominant in North Africa and the Middle East (Simmonds *et al.*, 2005). Genotype 5 is commonly found in South Africa whilst genotype 6 is prevalent in the Far East and Australia (Simmonds *et al.*, 2005). The prevalence of HCV varies greatly by geographic location, with prevalence estimated to be as low as 0.1% in the United Kingdom and Scandinavia. In comparison, prevalence in Egypt is reported at 15-20% (Alter, 2007).

1.4 Current treatment regime and prospects

The current standard of therapy for HCV infection is a combination of ribavirin plus pegylated IFN-α. This regime is only effective in 40-60% of cases however, and response to therapy varies between the different viral genotypes (genotypes 2 and 3 show better response rates than genotypes 1, 4, 5 and 6). Side effects are also common and can be severe leading to discontinuation of therapy (Poynard *et al.*, 2003). As knowledge of the virus lifecycle has increased, the search for HCV-specific, direct-acting antivirals

(DAAs) has intensified. Two protease inhibitors (targeting the serine protease NS3) improve recovery rates when given in combination with ribavirin and IFN-α. Boceprevir increases sustained virological response (SVR) rates in genotype 1 infected patients from 38% to 75% (Kwo *et al.*, 2010) compared to ribavirin and IFN-α alone. Telaprevir increases SVR rates in genotype 1 infected patients to 67%, compared with 41% of patients receiving standard therapy (McHutchison *et al.*, 2009). Telaprevir also improves SVR rates in patients who have previously failed standard therapy (McHutchison *et al.*, 2010). However, adverse events associated with Boceprevir and Telaprevir have been reported. They also show limited cross-genotype reactivity and a low barrier to resistance leading to the search for second-generation protease inhibitors with improved pharmacokinetics (Sarrazin *et al.*, 2012). The RNA-dependent RNA polymerase NS5B has also been investigated as a target for antiviral therapy leading to the development of nucleoside and non-nucleoside analogue inhibitors. Nucleoside analogues generally show cross-genotype activity, whereas the non-nucleoside inhibitors are more restricted and have a lower barrier to resistance mutations (Sarrazin *et al.*, 2012). Several drugs in both classes are currently under development and initial trials of the nucleoside analogue PSI-7977 in combination with ribavirin plus IFN-α (Lalezari *et al.*, 2011) or with ribavirin only (Sarrazin *et al.*, 2012) have shown promise.

1.5 Genome organisation and protein function

HCV belongs to the *Hepacivirus* genus within the *Flaviviridae* family of viruses. HCV is an enveloped viral particle possessing a single-stranded,

positive sense, RNA genome (Lemon *et al.*, 2007). The genome is 9.6 kb in length and encodes a single polyprotein of 3000 amino acid residues, flanked at the 5' and 3' end by untranslated regions (UTR) (Lemon *et al.*, 2007). The 5'UTR is a highly structured region containing the internal ribosome entry site (IRES) required for polyprotein synthesis, as well as elements necessary for RNA replication (Tellinghuisen *et al.*, 2007). The 3'UTR is also essential for virus replication (Kolykhalov *et al.*, 2000) and may enhance translation from the IRES in human liver cells and initiate negative-strand RNA synthesis (Tellinghuisen *et al.*, 2007). The polyprotein is cleaved co-translationally and post-translationally into 10 products by the cellular proteins, signal peptidase and signal peptide peptidase, as well as the viral proteases, products of the NS2-3 and NS3-4A proteins (Op De Beeck & Dubuisson, 2003; Penin *et al.*, 2004b) (Figure 1-1). The polyprotein encodes three structural proteins; core, and the envelope glycoproteins E1 and E2; the viroporin p7 and six non-structural proteins; NS2, NS3, NS4A, NS4B, NS5A and NS5B.

1.5.1 Core

The core protein forms the viral nucleocapsid housing the RNA genome and is surrounded by a host-derived lipid envelope or membrane. Core is cleaved from E1 by the action of signal peptidase, producing an immature 191 amino acid protein that undergoes further processing by the signal peptide peptidase to yield the mature protein (McLauchlan *et al.*, 2002) predicted to be at least 177 amino acids in length (Kopp *et al.*, 2010). The N-terminal domain (D1) of the mature core protein functions in protein oligomerisation and RNA binding

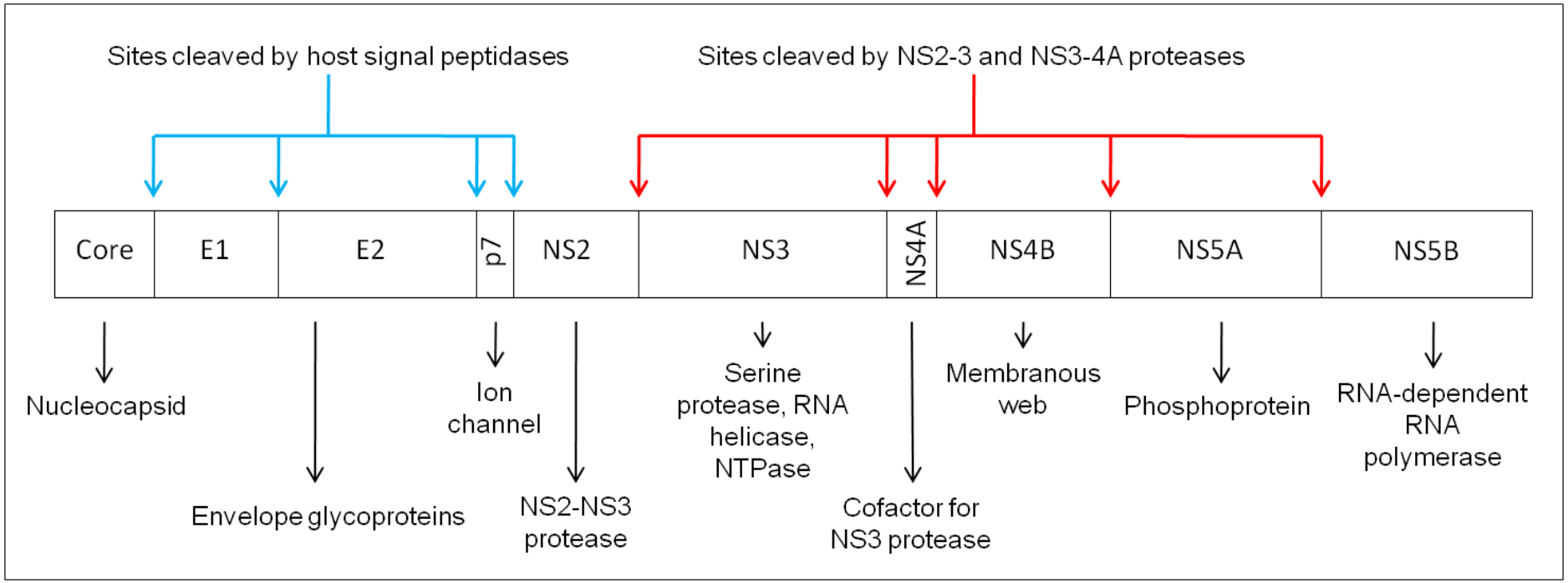

Figure 1-1. Schematic representation of the HCV polyprotein. The structural proteins, core, E1 and E2 and the viroporin p7 are cleaved from the polyprotein by host signal peptidase and signal peptide peptidase. Following translation of the polyprotein, the viral NS2-3 protease and the NS3-4A serine protease cleave the remaining non-structural proteins from the polyprotein.

(Boulant *et al.*, 2005) whilst D2 targets the core protein to lipid droplets and plays an important role in viral particle assembly (Shavinskaya *et al.*, 2007).

1.5.2 E1 and E2

The envelope glycoproteins are produced by cellular signal peptidase cleavage of the polyprotein (Op De Beeck & Dubuisson, 2003) (Figure 1-1). E1 and E2 are highly glycosylated structures displayed on the surface of the viral particle that form non-covalent heterodimers (Dubuisson *et al.*, 1994). The C-terminal transmembrane domains are thought to function in heterodimerisation and contain endoplasmic reticulum (ER) retention signals that may anchor the glycoproteins within the lipid membrane (Op De Beeck *et al.*, 2001). These non-covalent heterodimers have long been considered the functional form of the envelope glycoproteins. Recent studies however, have shown that within the HCV cell culture (HCVcc) system, particles also possess envelope glycoproteins in larger covalent complexes stabilised by disulphide bridges (Vieyres *et al.*, 2010). Further study is needed to determine which of these represents the functional form of the glycoproteins.

E1 and E2 are highly glycosylated; their ectodomains are targeted to the ER lumen where they undergo N-linked and O-linked glycosylation (Op De Beeck *et al.*, 2001; Op De Beeck & Dubuisson, 2003). N-linked glycosylation is carried out by the enzyme oligosaccharyl transferase during protein synthesis. It catalyses the transfer of a $Glc_3Man_9GlcNAc_2$ oligosaccharide from a lipid intermediate to an asparagine residue within the Asn-X-Thr/Ser consensus sequence (Goffard & Dubuisson, 2003). This transfer takes place on the

lumenal face of the ER (Goffard & Dubuisson, 2003; Op De Beeck & Dubuisson, 2003) resulting in the exposure of the glycosylated glycoproteins on the surface of the virion. E1 possesses six potential N-linked glycosylation sites, although not all are glycosylated. E2 has 11 possible N-linked glycosylation sites (Goffard & Dubuisson, 2003; Op De Beeck *et al.*, 2001) (Figure 1-2a), nine of which are conserved across genotypes (Helle *et al.*, 2007). These glycans play an important role in the structure and function of the glycoproteins (Goffard *et al.*, 2005) and some are essential for entry of HCVpp into target cells (Falkowska *et al.*, 2007) suggesting a role for glycans in E2-receptor interactions. Other glycans play a role in modulating immunogenicity; the removal of specific glycans enhances binding to CD81 and increases sensitivity to neutralisation by sera and antibodies (Falkowska *et al.*, 2007; Helle *et al.*, 2007; Helle *et al.*, 2010).

The role of E1 is poorly understood although it has been proposed to function as the fusion determinant, triggering fusion of the viral and cellular membranes during entry (Lavillette *et al.*, 2007). It has also been suggested that E1 functions in virus entry by binding to apolipoprotein B (apoB) and apoE and promoting virus attachment to the low density lipoprotein receptor (LDLr) (Mazumdar *et al.*, 2011).

The N-terminal ectodomain of E2 contains the entry determinants for infection of a host cell. E2 comprises amino acids (aa) 384-748 of the HCV polyprotein, with the transmembrane domain located at aa 718-748. The linear structure of the E2 ectodomain highlights three highly variable regions (Figure 1-2b). Hypervariable region 1 (HVR1), located at the N-terminus, is 26-28 amino acids in length (depending on genotype). This region, between aa 384-

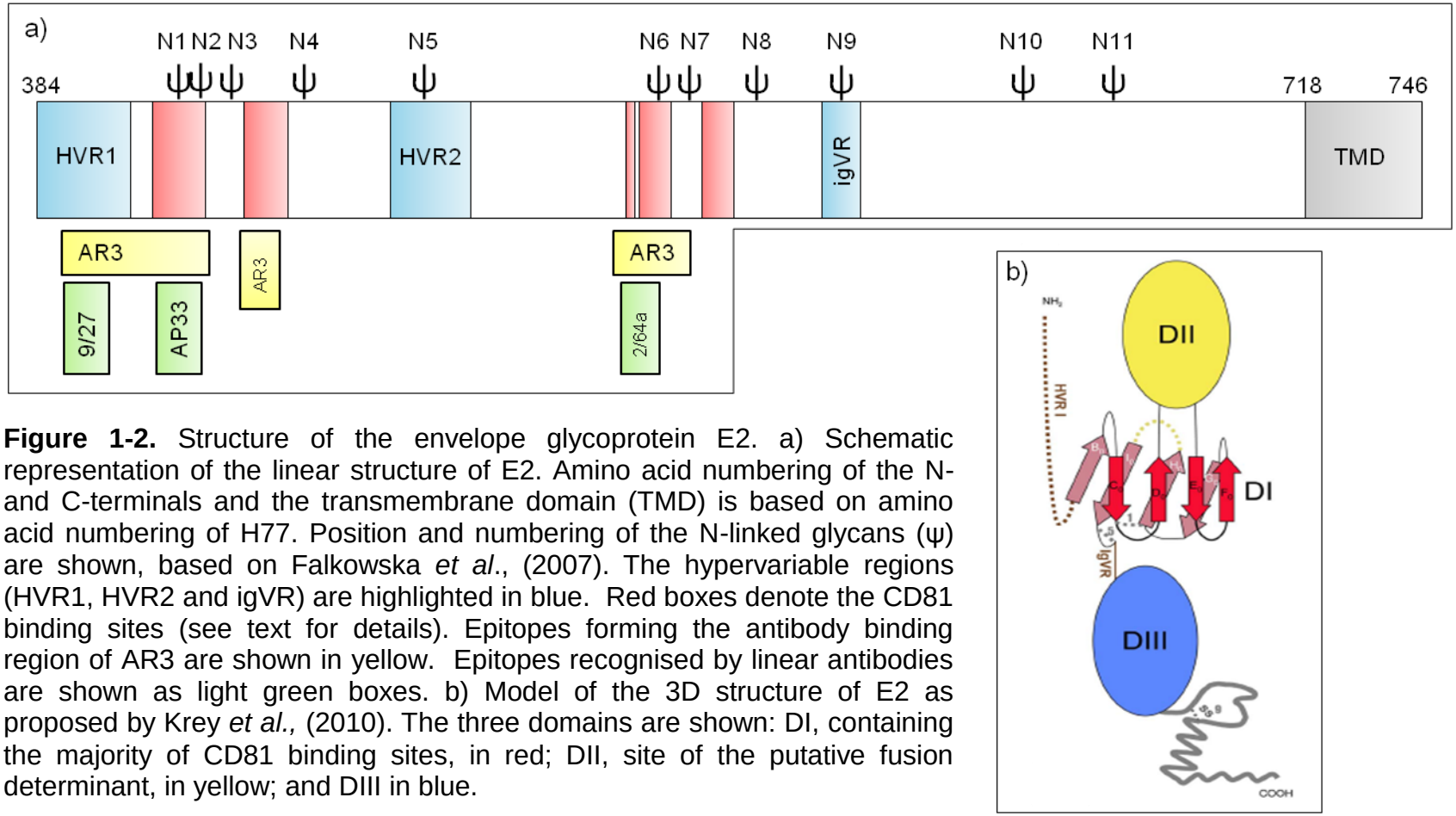

Figure 1-2. Structure of the envelope glycoprotein E2. a) Schematic representation of the linear structure of E2. Amino acid numbering of the N- and C-terminals and the transmembrane domain (TMD) is based on amino acid numbering of H77. Position and numbering of the N-linked glycans (ψ) are shown, based on Falkowska *et al*., (2007). The hypervariable regions (HVR1, HVR2 and igVR) are highlighted in blue. Red boxes denote the CD81 binding sites (see text for details). Epitopes forming the antibody binding region of AR3 are shown in yellow. Epitopes recognised by linear antibodies are shown as light green boxes. b) Model of the 3D structure of E2 as proposed by Krey *et al.,* (2010). The three domains are shown: DI, containing the majority of CD81 binding sites, in red; DII, site of the putative fusion determinant, in yellow; and DIII in blue.

410, plays an important role in entry through its interaction with SR-BI (Bartosch *et al.*, 2005; Bartosch *et al.*, 2003c) as well as in antibody binding (Vieyres *et al.*, 2011) and disease outcome (Farci *et al.*, 2000). HVR2 and the intergenotypic variable region (igVR), both downstream of HVR1, may play a role in heterodimerisation of E1E2 and virus infectivity (Albecka *et al.*, 2011; McCaffrey *et al.*, 2011). Two additional hypervariable regions have been identified in genotype 3a isolates only but the role of these regions, HVR495 (aa 495-501) and HVR575 (aa 575-578), in antibody binding and disease outcome has yet to be established (Humphreys *et al.*, 2009). Several putative sites for CD81 binding (see section 1.7.4), as well the majority of known antibody binding epitopes have also been mapped to E2.

A recently described 3D structural model (Krey *et al.*, 2010) proposes that E2 is a class II fusion protein, analogous to the envelope glycoproteins of the related viruses, alphavirus (E1) and flavivirus (E). These proteins have a three domain architecture consisting of DI, DII and DIII (Figure 1-2b). In HCV E2, DI contains most of the CD81 binding determinants. The majority of glycans also cluster within this domain, consistent with a role of some glycans in modulating HCV entry (Falkowska *et al.*, 2007; Goffard *et al.*, 2005). DI is preceded by an N-terminal extension containing HVRI. DII contains the putative fusion peptide (aa 502-520) (Krey *et al.*, 2010), contradicting studies that attribute the fusion determinant to E1 (Lavillette *et al.*, 2007). However, in class fusion II proteins, the protein analogous to HCV E1 functions as a chaperone, exposing the fusion loop buried within the prefusion form of the protein (Krey *et al.*, 2010). Further study will be needed to determine if E1 does fulfil this role in HCV entry. Domains DI and DIII of E2 are linked by

IgVR which can be extended to enable the necessary movement of DIII during membrane fusion. DIII is followed by the stem region that contains a conserved heptad repeat (aa 675-699) involved in E1E2 heterodimerisation and entry of HCVpp (Drummer & Poumbourios, 2004). A further segment at aa 705-715 has been identified as essential for entry of HCVcc (Albecka *et al.*, 2011). The stem connects to the transmembrane domain responsible for anchoring E2 to the viral membrane (Op De Beeck *et al.*, 2001). The 3D structure of E2 is partly determined by the position of 18 cysteine residues forming 9 putative disulphide bonds (Krey *et al.*, 2010) all of which are essential for infectivity of HCVpp and HCVcc (McCaffrey *et al.*, 2012).

1.5.3 p7

There has been some debate over whether p7 is a structural or non-structural protein. p7 is a 63 amino acid, integral membrane protein containing two transmembrane domains connected by a short cytoplasmic loop (Carrere-Kremer *et al.*, 2002). It is located in the ER membrane where it forms an ion channel that can be blocked by amantadine (Griffin *et al.*, 2003; Wozniak *et al.*, 2010) and iminosugars (Pavlovic *et al.*, 2003). p7 is essential for infectivity *in vivo* (Sakai *et al.*, 2003) and functions in the efficient assembly and release of viral particles (Steinmann *et al.*, 2007) by de-acidifying intracellular vesicles, presumably to protect the immature, acid sensitive viral particles (Wozniak *et al.*, 2010). However, it remains to be seen whether p7 possesses additional, ion channel-independent functions.

1.5.4 NS2

NS2 is a 23 kDa membrane protein that, in combination with the N-terminal portion of NS3, forms a protease required for cleavage of the polyprotein at the NS2-NS3 boundary (Grakoui *et al.*, 1993b). The N-terminus of NS2 contains the transmembrane domain (Santolini *et al.*, 1995) with the cysteine protease activity residing in the C-terminal domain (Lorenz *et al.*, 2006). NS2 is essential for the production of infectious HCV particles (Jirasko *et al.*, 2008; Jones *et al.*, 2007) and is able to form a complex with E1, E2, NS3, and possibly also NS4, that presumably functions to bring structural and non-structural proteins together for particle assembly (Stapleford & Lindenbach, 2011).

1.5.5 NS3 and NS3-4A

NS3 is a multifunctional protein containing both serine protease and NTPase/helicase activity. The serine protease activity requires NS4A as a cofactor (Failla *et al.*, 1994) and is responsible for cleaving the polyprotein at the NS3-NS4A, NS4A-NS4B, NS4B-NS5A and NS5A-NS5B boundaries (Grakoui *et al.*, 1993a). NS3-4A is also able to block innate antiviral responses, such as the RIG-I induced pathway of NF-κB and IFN regulatory factor 3 (IRF3) activation, by cleaving the adaptor protein Cardif /MAVS (Li *et al.*, 2005b; Meylan *et al.*, 2005), thereby reducing production of type-I IFN. It also interferes with the activation of IRF3 and NF-κB via Toll-like receptor (TLR)-3 by cleaving the adaptor protein Trif (Li *et al.*, 2005a). NS3-4A may therefore contribute to the persistence of HCV infection. This dual function makes NS3 a particularly promising target for antiviral therapies, as it raises

the possibility of simultaneously targeting viral replication and enhancing the immune response (Raney *et al.*, 2010). The NTPase (Suzich *et al.*, 1993) and helicase activity (Tai *et al.*, 1996) of NS3 unwinds double-stranded RNA and DNA in a 3' to 5' direction in a reaction coupled to the hydrolysis of NTPs. The precise function of this helicase activity is not completely understood, (Raney *et al.*, 2010) although it is essential for viral replication (Lam & Frick, 2006) and infectivity (Kolykhalov *et al.*, 2000).

1.5.6 NS4B

NS4B is a 27 kDa hydrophobic membrane protein (Yu *et al.*, 2006). The expression of NS4B results in the formation of a membranous web structure. All viral proteins are associated with this structure leading to the suggestion that the membranous web is the site of HCV replication within an infected cell (Egger *et al.*, 2002). Polymerisation and lipid modification, specifically palmitoylation, of NS4B is required for the protein-protein interactions involved in formation of the replication complex to occur (Yu *et al.*, 2006). NS4B has also been postulated to play a role in the production of viral particles (Jones *et al.*, 2009).

1.5.7 NS5A

NS5A is a phosphoprotein that binds NS5B and modulates its RNA-dependent RNA polymerase activity *in vitro* (Shirota *et al.*, 2002). The N-terminal, α-helical region anchors the protein within the ER membrane and provides a platform for formation of the replication complex (Penin *et al.*,

2004a). Adjacent to the α-helix, domain I of NS5A is a zinc-binding domain that is essential for RNA replication, although its precise role is unclear (Tellinghuisen *et al.*, 2004). NS5A is also able to bind single-stranded RNA, suggesting a role in RNA replication via binding to U/G rich regions in the NTRs of the HCV genome (Huang *et al.*, 2005). NS5A exists in a basally phosphorylated (56 kDa) form and a hyperphosphorylated (58 kDa) form. Differential phosphorylation of NS5A modulates RNA replication (Appel *et al.*, 2005), protein-protein interactions (Evans *et al.*, 2004) and infectious particle production (Tellinghuisen *et al.*, 2008) and may therefore function as a regulatory switch during virus replication. NS5A may also modulate the IFN response to HCV by inhibiting PKR, a protein kinase that reduces mRNA translation, with important implications for treatment of the virus (Katze *et al.*, 2002).

1.5.8 NS5B

NS5B is the RNA-dependent RNA polymerase and as such is indispensable for replication (Kolykhalov *et al.*, 2000; Lohmann *et al.*, 1999). It is a tail-anchored protein with the C-terminal 21 amino acids essential for correct insertion into the ER membrane (Moradpour *et al.*, 2004). Cooperative initiation and elongation of RNA requires the formation of NS5B oligomers (Wang *et al.*, 2002) and NS5B interacts with other viral proteins to form the replication complex (Ishido *et al.*, 1998; Shirota *et al.*, 2002). More recently an interaction between NS5B, p7 and NS2 has been postulated to function in virus particle production (Gouklani *et al.*, 2012). The essential role of NS5B in the HCV life-cycle has made it a target for antiviral therapies (see section 1.4).

1.6 Model systems to study HCV

A number of model systems have been developed to study the lifecycle and immune response to HCV. As this thesis is focussed on elucidating the antibody response to HCV, the following section will focus on those systems that have enabled such studies of the antibody response as well studies of the cellular receptors and the envelope glycoproteins. Although the subgenomic replicon system (Lohmann *et al.*, 1999) is an important development that provides insight into the replication of viral RNA, it lacks the structural proteins, precluding any studies of viral assembly or structure, and therefore will not be discussed further.

1.6.1 Chimpanzees as model organisms

Chimpanzees are the only species other than humans that are permissive for HCV infection. Early studies of the antibody response showed that infection of chimpanzees could be prevented by *in vitro* neutralisation of HCV (Farci *et al.*, 1994) and infectivity *in vivo* was also neutralised by the administration of immune sera (Farci *et al.*, 1996). Furthermore, chimpanzees were capable of eliciting an HCV-specific neutralising antibody (nAb) response following immunisation with the envelope glycoproteins, a response that partially protected against experimental challenge with the virus (Choo *et al.*, 1994). However, it has since been noted that the outcome of HCV infection in chimpanzees differs greatly from that in humans. Chimpanzees tend to suffer a milder disease with a higher rate of viral clearance and are less likely to develop liver fibrosis or cirrhosis (Jo *et al.*, 2011). In terms of the immune

response, viral clearance in chimpanzees is associated with a strong cellular immune response (Barth *et al.*, 2011; Grakoui *et al.*, 2003; Shoukry *et al.*, 2003). As will be discussed later, spontaneous resolution in humans is associated with strong nAb responses. Therefore studies of chimpanzees may not give an accurate assessment of the role played by the immune response in clearance of HCV. The ethical and financial constraints of chimpanzee studies also make it a difficult model to work with.

1.6.2 Soluble E2 and neutralisation of binding assays

Recombinant, soluble, full-length or truncated forms of E2 were used to identify the first entry receptors involved in HCV entry. However, the use of soluble E2 (sE2) may not represent the native form of the envelope glycoprotein. Indeed a study of truncated sE2, mammalian expressed E1E2 and VLPs (see section 1.6.3) has highlighted structural, and therefore antigenic, dissimilarities between the three forms of E2 (Owsianka *et al.*, 2001). Following the discovery that E2 binds to the cellular receptor CD81 (Pileri *et al.*, 1998), many early *in vitro* studies of the antibody response probed this particular interaction. The neutralisation-of-binding (NOB) assay measures E2 binding to Molt-4 cells in the presence of nAb (Flint *et al.*, 1999a; Rosa *et al.*, 1996). However, Molt-4 cells, a lymphoma cell line, express CD81 but none of the other HCV entry receptors. Therefore studies using this assay would be limited to identifying antibodies targeting E2-CD81 interactions only. Subsequent studies comparing Molt-4 cells and Huh7 hepatocarcinoma cells have shown that these antibodies inhibit the binding of E2 to Molt-4 cells more

readily than to Huh7 cells (Heo *et al.*, 2004) potentially overstating the neutralising capabilities of some antibodies.

1.6.3 Virus-like particles

Virus-like particles (VLPs) can be generated by the expression of the HCV structural proteins (core-E1-E2) in a baculovirus-insect cell (*Spodoptera frugiperda* Sf9) expression system. The resulting particles are structurally similar to virions isolated from HCV-infected humans and chimpanzees (Baumert *et al.*, 1998) and react with HCV-specific immune sera (Baumert *et al.*, 1999). They are also immunogenic and able to elicit an anti-E2 antibody response in mice (Baumert *et al.*, 1999). However, these VLPs are not secreted from the insect cells but rather are retained within intracellular vesicles (Baumert *et al.*, 1998) making the isolation of VLPs difficult.

1.6.4 HCVpp

The HCV pseudoparticle (HCVpp) system uses HCV envelope glycoproteins displayed on the surface of retroviral or lentiviral core particles (Bartosch *et al.*, 2003b) to mimic the attachment and entry steps of the HCV lifecycle. Infectivity of a human hepatoma cell line is assessed by measurement of a specific reporter incorporated into the pseudoparticle. HCVpp have been used to characterise E1 and E2 (Flint *et al.*, 2004) and to identify nAbs in human and chimpanzee sera (Bartosch *et al.*, 2003a; Yu *et al.*, 2004). More recently, E1E2 isolated from patient sources has been displayed on the surface of the pseudoparticle (Fafi-Kremer *et al.*, 2010; Tarr *et al.*, 2011) and used to assess

the evolution of the antibody response in individual patients (Dowd *et al.*, 2009). However, not all pseudoparticles bearing patient-derived E1E2 are functional (Dowd *et al.*, 2009; Lavillette *et al.*, 2005). One study suggests that less than 50% of pseudoparticles bear E1E2 at the surface and that the number of E1E2 molecules on each particle is low (Bonnafous *et al.*, 2010). It is also important to note that, unlike serum derived particles, HCVpp are not associated with any lipoproteins (Burlone & Budkowska, 2009). Lipoproteins are thought to play an important role in modulating entry and neutralisation of HCV; therefore the HCVpp system may not mimic accurately the *in vivo* situation. Furthermore, glycoprotein maturation and assembly of HCVpp differs from that seen in the HCVcc system (see below); HCVpp are assembled in a post-Golgi compartment following transit of the glycoproteins through the cell secretory pathway independently of other particle components (Sandrin *et al.*, 2005). In contrast, HCVcc assemble in an ER-derived compartment (Miyanari *et al.*, 2007). These differences may lead to variations in morphology and indeed, differential glycosylation of E1E2 has been observed between HCVcc and HCVpp (Vieyres *et al.*, 2010)

1.6.5 HCVcc

For many years, studies of the HCV lifecycle were hampered by the failure of the virus to replicate in cell culture. The discovery of a genotype 2a strain, JFH-1, capable of replicating in cell culture without accumulating adaptive mutations (Kato *et al.*, 2001; Wakita *et al.*, 2005) has greatly improved our knowledge of the virus. This HCV cell culture (HCVcc) system generates infectious viral particles enabling studies of the entire virus lifecycle. Much

effort has gone into developing chimeric infectious clones of this virus, in order to study viruses of different genotypes. This involves replacing the structural proteins (core-E1-E2) plus p7 and NS2 of JFH-1 with the corresponding proteins from genotypes 1-7. Although generation of these chimeras has been successful, they often have poor replication kinetics and acquire cell culture adaptive mutations (Gottwein *et al.*, 2009; Pietschmann *et al.*, 2006). Cell culture adaptations have been reported to increase the sensitivity of some isolates to antibody neutralisation and to alter binding to entry receptors (Dhillon *et al.*, 2010; Grove *et al.*, 2008). Therefore studies of virus neutralisation carried out in the HCVcc system must be interpreted carefully.

1.6.6 Mouse models

Chimpanzees and humans are the only species naturally infected with HCV. Small animal models are available, however these are unable to support the complete replication cycle of HCV. The uPA-SCID mouse model uses immunosuppressed chimeric mice transplanted with human hepatocytes. This renders the mice susceptible to HCV infection (Lindenbach *et al.*, 2006; Mercer *et al.*, 2001), although their immunodeficiency precludes any study of the host adaptive immune response. Despite this, these mice have been used to demonstrate the protective nature of passively transferred anti-HCV antibodies (Law *et al.*, 2008; Vanwolleghem *et al.*, 2008), anti-CD81 antibodies (Meuleman *et al.*, 2008) and anti-SR-BI antibodies (Meuleman *et al.*, 2012). A genetically humanised mouse bearing the human HCV entry factors (CD81, SR-BI, CLDN-1 and OCLN) was recently developed that is able to support

HCV entry but does not permit viral replication. Studies of the complete viral lifecycle are therefore not possible but these mice are immunocompetent and do produce antibody in response to vaccination. Furthermore, this antibody response protects against subsequent viral challenge. Passive transfer of E2-specific and receptor-specific antibodies also protects against infection in these mice (Dorner *et al.*, 2011; Giang *et al.*, 2012).

1.7 Cell entry and receptors

A number of receptors and co-receptors for HCV entry have been identified; however, not all of them bind directly to E2 and instead may be required for the formation of a co-receptor complex at the surface of hepatocytes that mediates HCV entry. A schematic representation of the HCV entry pathway, as it is currently understood, is presented in Figure 1-3.

1.7.1 Lectins

The lectins DC-SIGN (dendritic cell-specific intracellular adhesion molecule-3-grabbing non-integrin) and L-SIGN (liver-specific SIGN) are type II membrane proteins belonging to the C-type lectin family. DC-SIGN and L-SIGN are able to bind sE2 and HCVpp via the mannose containing N-linked glycans (Lozach *et al.*, 2004; Pohlmann *et al.*, 2003). DC-SIGN is expressed on dendritic cells and lymphocytes while L-SIGN is expressed on liver sinusoidal cells. These cells are not thought to internalise the virus and there is evidence to suggest that they act as ‘capture receptors’ which transmit the virus to susceptible cells i.e. hepatocytes (Lozach *et al.*, 2004).

1.7.2 Glycosaminoglycans

Glycosaminoglycans (GAGs) are cell surface polysaccharides found on most cell types and the GAG heparan sulphate (HS) is required by several viruses for attachment to the target cell (Barth *et al.*, 2003). HS binds to E2 with N-terminal epitopes proposed as the sites of interaction (Barth *et al.*, 2003; Barth *et al.*, 2006). E1 is also able to bind HS, although the affinity is much lower than for E2 (Barth *et al.*, 2006). E2-HS binding and HCVpp entry can be inhibited by patient-derived anti-E2 antibodies and HS analogues such as heparin (Barth *et al.*, 2003; Barth *et al.*, 2006). GAGs function at an early step in virus attachment to the target cell, prior to the binding of co-receptors such as CD81 (Barth *et al.*, 2006; Koutsoudakis *et al.*, 2006).

1.7.3 LDL receptor (LDLr)

HCV circulating in patient sera is heterogeneous, displaying a range of buoyant densities caused by association with different lipoproteins (Burlone & Budkowska, 2009). Low-density particles containing apoB and apoE, termed lipoviral particles (LVPs), are the most infectious (Agnello *et al.*, 1999; Andre *et al.*, 2002; Bradley *et al.*, 1991). Lipoproteins appear to play a role in entry as LDLr mediates entry of HCV and other *Flaviviridae* family members via binding of low-density lipoprotein (LDL) or very-low-density lipoprotein (VLDL) but not high-density lipoprotein (HDL) (Agnello *et al.*, 1999). Specifically, the apoB and apoE components of VLDL (or apoB within LDL) mediate attachment to LDLr (Andre *et al.*, 2002) and the reaction is independent of E2 binding to LDLr (Wunschmann *et al.*, 2000). Therefore lipoprotein receptors, such as LDLr, could promote HCV entry indirectly, via

binding to lipoproteins associated with the virus particle rather than by binding to the envelope glycoproteins. However a recent study suggests that LDLr is not involved in productive HCV entry but is essential for replication of the virus via its role in lipid metabolism (Albecka *et al.*, 2012).

1.7.4 CD81

CD81 is a member of the tetraspanin family of transmembrane proteins. It is a ubiquitous protein which normally forms part of the T cell and B cell receptor complex. It has four transmembrane domains, two small extracellular loops (SEL) and one large extracellular loop (LEL). The N and C terminals are located intracellularly (Levy & Shoham, 2005). CD81 was the first receptor identified as being critical for HCV entry (Pileri *et al.*, 1998). CD81 is able to bind sE2 via the LEL and this interaction is inhibited by anti-CD81 monoclonal antibodies (MAb) and soluble CD81 LEL, preventing infectivity of HCVpp and HCVcc (Bartosch *et al.*, 2003b; Wakita *et al.*, 2005). Down-regulation of CD81 expression using small interfering RNA (siRNA) also reduces infectivity of serum-derived HCV (Molina *et al.*, 2008). Human hepatoma cell lines, such as HepG2, which do not naturally express CD81 are rendered permissive to HCV infectivity following ectopic expression of CD81 (Bartosch *et al.*, 2003c; Lavillette *et al.*, 2005). CD81 has been proposed as a species-specific determinant of HCV infectivity (Ploss *et al.*, 2009). CD81 binding sites within E2 have been extensively studied and research indicates that they map to the regions aa 412-423 (Owsianka *et al.*, 2001), aa 432-447 (Clayton *et al.*, 2002), aa 528-535 (Owsianka *et al.*, 2001) and aa 544-551 (Flint *et al.*, 1999a). Additionally, specific residues W420, Y527, W529, G530 and D535

(Owsianka *et al.*, 2006) and the motif ^{436}GWLAGLFY^{443} (Drummer *et al.*, 2006) are critical for E2-CD81 binding (Figure 1-2a).

1.7.5 SR-BI

Human scavenger receptor class B type I (SR-BI) is a glycoprotein found on the surface of several mammalian cells but is particularly abundant on liver cells (Burlone & Budkowska, 2009). SR-BI binds to sE2, a process that requires HVR1 as binding is prevented by the deletion of HVR1 and in the presence of anti-HVR1 antibodies (Bartosch *et al.*, 2003c; Scarselli *et al.*, 2002). SR-BI is thought to function at a post-attachment step and at a similar point in the entry process as CD81 (Zeisel *et al.*, 2007). CD36, a scavenger receptor molecule which is very similar to SR-BI but found on endothelial cells and macrophages, does not bind sE2, indicating a potential role for SR-BI in determining viral tropism (Scarselli *et al.*, 2002). The main function of SR-BI is in lipid metabolism, specifically the selective uptake of cholesteryl ester from HDL (Connelly & Williams, 2003). Serum HCV particles are often associated with lipoproteins, including LDL and VLDL (Andre *et al.*, 2002), raising the possibility that association of lipoproteins with HCV could facilitate infection in an SR-BI-dependent process. It has since been demonstrated that infectivity of HCVpp and HCVcc is enhanced by the addition of HDL, via interaction between SR-BI and HDL (Bartosch *et al.*, 2005; Dreux *et al.*, 2006; Voisset *et al.*, 2005). HCVpp infectivity is also enhanced by the apolipoprotein C1 (Meunier *et al.*, 2005). HDL-enhancement of infection requires HVR1, as infectivity of HCVpp lacking HVR1 cannot be enhanced by HDL (Bartosch *et al.*, 2005). A recent study suggests that the requirement for

HVR1 in infectivity varies between genotypes; HVR1 is essential for infectivity of genotypes 1a, 1b, 2b, 3a and 4a whereas genotypes 2a, 5a and 6a carrying a deletion of HVR1 remain infectious (Prentoe *et al.*, 2011). Therefore regions of contact between E2 and SR-BI may exist outside of HVR1. HCV binding to SR-BI is also mediated via apoB-containing VLDL and may provide an alternative SR-BI entry mechanism that does not require direct HVR1-SR-BI binding (Maillard *et al.*, 2006), although precise details of the role of SR-BI in HCV entry remain to be determined.

1.7.6 Claudin-1

Claudin-1 (CLDN-1) is a tight junction protein required for HCV entry (Evans *et al.*, 2007). It is a four transmembrane protein, with two extracellular loops (EL) and intracellular N- and C-terminals. The entry determinants of CLDN-1 map to EL1 (Cukierman *et al.*, 2009; Evans *et al.*, 2007). There is no evidence of direct contact between HCV and CLDN-1 (Evans *et al.*, 2007; Krieger *et al.*, 2010); however, CD81 and CLDN-1 co-localise within cultured cells and hepatocytes (Harris *et al.*, 2008) and form a CD81-CLDN-1 co-receptor complex necessary for HCV entry (Harris *et al.*, 2010). A recent study suggests that CD81-CLDN-1 endocytosis is enhanced by HCV to promote entry into hepatocytes (Farquhar *et al.*, 2012). The interaction between CD81 and CLDN-1 can be disrupted by anti-CLDN-1 antibodies thereby inhibiting HCV entry (Fofana *et al.*, 2010; Krieger *et al.*, 2010). Claudin-6 and claudin-9 can also function as co-receptors in the entry of HCVcc and HCVpp (Harris *et al.*, 2010; Zheng *et al.*, 2007).

1.7.7 Occludin

The tight-junction protein occludin (OCLN) is also an essential co-receptor for HCV entry (Ploss *et al.*, 2009), functioning at a post-attachment step (Benedicto *et al.*, 2008). OCLN and E2 are thought to interact directly, as demonstrated by co-precipitation of the proteins (Benedicto *et al.*, 2008; Liu *et al.*, 2009). OCLN is a four transmembrane protein possessing two ELs and the entry determinants have been mapped to EL2 (Liu *et al.*, 2010). OCLN, along with CD81, form the minimal species-specific determinants for HCV entry, as the mouse and other non-primate homologues of OCLN are refractory to HCV infectivity (Dorner *et al.*, 2011; Michta *et al.*, 2010; Ploss *et al.*, 2009). These species-specific determinants also map to EL2 (Michta *et al.*, 2010; Ploss *et al.*, 2009).

1.7.8 Additional entry factors

Recently three additional co-factors have been implicated in HCV entry. Epidermal growth factor receptor (EGFR) and ephrin receptor A2 (EphA2) were identified using a siRNA screen of cellular kinase networks. These receptor tyrosine kinases regulate the formation of the CD81-CLDN-1 co-receptor complex, and therefore act at a post-binding step as well as playing a role in fusion. Reduced expression of EGFR and EphA2 reduces infectivity of both HCVcc and HCVpp. Infectivity is also inhibited by the EGFR inhibitor, erlotinib, and the EphA2 inhibitor, dasatinib, as well as antibodies targeting both receptors. Furthermore, erlotinib reduces infectivity in the uPA-SCID mouse model (Lupberger *et al.*, 2011).

Niemann-Pick C1-like 1 (NPC1L1) cholesterol absorption receptor is a transmembrane protein involved in cholesterol homeostasis. Its role in HCV entry was recently demonstrated as siRNA silencing of NPC1L1 reduces HCVcc infectivity. NPC1L1 acts prior to viral fusion and its activity is dependent on the level of cholesterol within the viral particle. Entry is also blocked by an NPC1L1-specific antibody blocking virus binding to the first LEL of NPC1L1. The NPC1L1 inhibitor, ezetimibe, reduces HCV infectivity *in vitro* and delays the development of HCV infection *in vivo* (Sainz *et al.*, 2012).

1.7.9 Proposed mechanism of HCV entry into target cells

The exact role played by each cellular receptor or entry factor is still incompletely understood; however, the experimental evidence has made it possible to piece together an entry model (Figure 1-3).

The viral particles circulate as LVPs in the vascular system (Burlone & Budkowska, 2009). The lectins, L-SIGN and DC-SIGN, are believed to act as 'capture receptors' capable of binding the virus and transmitting it to permissive cells (Lozach *et al.*, 2004). Initial attachment to hepatocytes is mediated by LDLr binding to the lipoprotein components of LVPs (Andre *et al.*, 2002) and GAG binding to the envelope glycoproteins (Barth *et al.*, 2003) in a low affinity reaction which brings the virus into close proximity with the high affinity receptors CD81 and SR-BI at the cell surface. It has been proposed that this triggers cytoskeletal rearrangements which enable translocation of the viral particle to sites containing CLDN-1 and OCLN

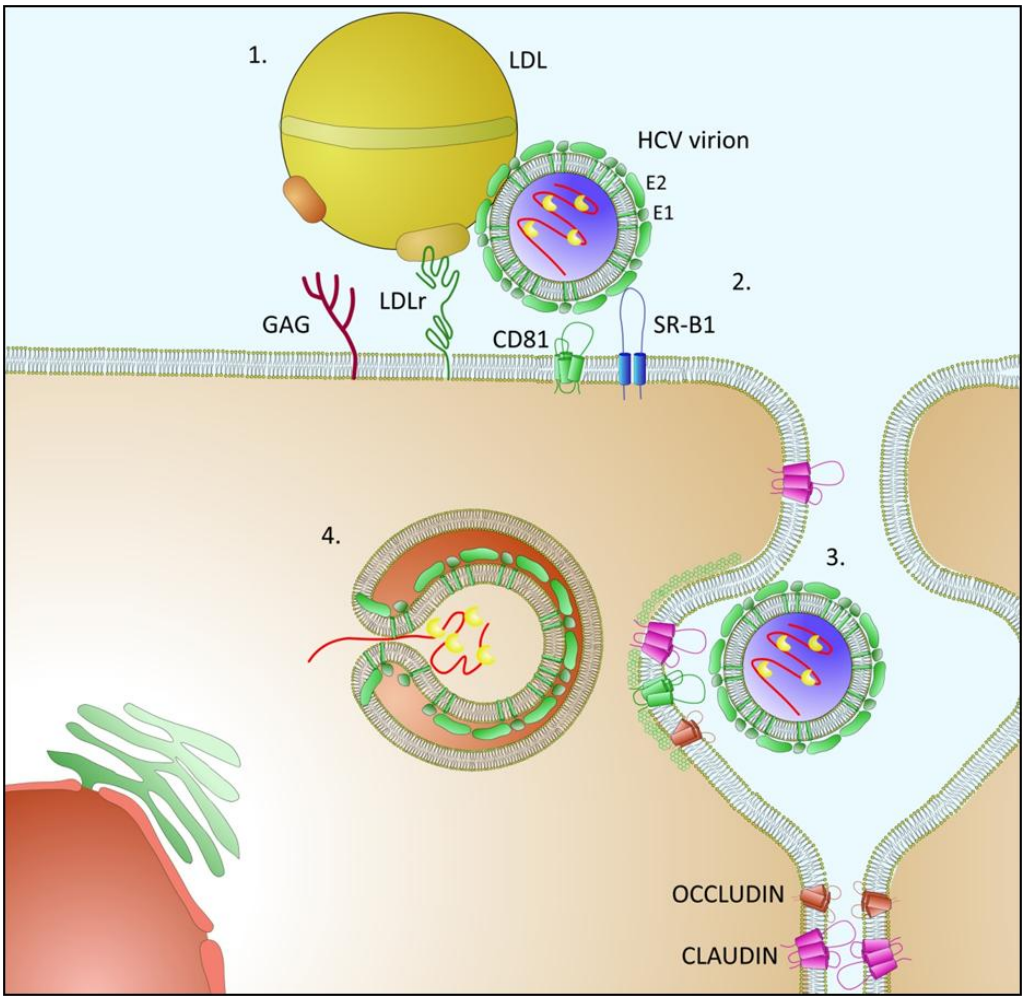

Figure 1-3. Mechanism of HCV entry into hepatocytes. (1) Viral particles circulate as LVPs. The lectins, L-SIGN and DC-SIGN, act as 'capture receptors', binding the virus and transmitting it to permissive cells. Initial attachment is mediated by LDLr and GAG binding in low affinity reactions which bring the virus into close proximity with high affinity receptors. (2) E1E2 binds CD81 and SR-BI at the cell surface triggering cytoskeletal rearrangements leading to translocation of the viral particle to sites containing CLDN-1 and OCLN. (3) The E2-CD81-CLDN-1 co-receptor complex is formed leading to internalisation via clathrin-mediated endocytosis. (4) pH-dependent fusion of the viral envelope with endosomal plasma membranes releases the nucleocapsid into the cytoplasm. Figure is taken from Edwards *et al.,* (2011).

(Brazzoli *et al.*, 2008). An E2-CD81-CLDN-1 co-receptor complex is formed and is necessary for entry to occur (Harris *et al.*, 2010; Harris *et al.*, 2008). The precise function of OCLN has not yet been determined but interaction of the virus with co-receptors leads to internalisation via clathrin-mediated endocytosis (Blanchard *et al.*, 2006; Farquhar *et al.*, 2012). Trafficking via endosomes and pH-dependent fusion of the viral envelope with endosomal plasma membranes is thought to occur (Blanchard *et al.*, 2006; Lavillette *et al.*, 2006; Tscherne *et al.*, 2006), however, further study is needed to elucidate the precise mechanism and to identify the fusion determinant within the envelope glycoproteins.

The entry factors EGFR, EphA2 and NPC1L1 are not included on this diagram as they have only recently been described. However, EGFR and EphA2 are thought to regulate formation of the CD81-CLDN-1 co-receptor complex (Lupberger *et al.*, 2011). The function of NPC1L1 is less clear although its cholesterol-dependent nature may indicate an indirect mode of action, whereby it removes virus-associated cholesterol to reveal glycoprotein binding sites or allow conformational changes necessary for binding (Sainz *et al.*, 2012).

1.8 The immune response to HCV

1.8.1 The cellular immune response

The humoral immune response does not function alone and the role played by the cellular adaptive immune response has been well studied. The importance of $CD8^+$ T cells and $CD4^+$ T cells has been demonstrated by

depletion studies carried out in chimpanzees. Depletion of $CD8^+$ T cells results in the persistence of HCV until $CD8^+$ T cells recover (Shoukry *et al.*, 2003) while depletion of $CD4^+$ T cells causes persistent infection which becomes chronic due to the selection of mutations in MHC class I epitopes (Grakoui *et al.*, 2003).

$CD8^+$ (cytotoxic) T lymphocyte (CTL) responses are directed against epitopes spanning the HCV polyprotein, encompassing both structural and non-structural proteins (Neumann-Haefelin *et al.*, 2008). In acute resolving infection, the $CD8^+$ T cell response peaks within 4 to 10 weeks following exposure to the virus and is thought to contribute to liver pathology as the CTL response coincides temporally with the onset of liver disease (Lechner *et al.*, 2000). $CD8^+$ T cells also preferentially accumulate in the liver compared to the peripheral blood (Neumann-Haefelin *et al.*, 2008; Spangenberg *et al.*, 2005). CTLs produced early during infection exhibit a 'stunned' phenotype and produce low levels of IFN-γ but these are later replaced by a memory phenotype producing high levels of IFN-γ. A strong $CD8^+$ T cell response targeting multiple epitopes is generally associated with viral clearance (Chang *et al.*, 2001; Lechner *et al.*, 2000; Thimme *et al.*, 2002); however, some studies have demonstrated virus persistence despite the presence of a multi-specific $CD8^+$ T cell response (Cox *et al.*, 2005). Therefore the breadth of the $CD8^+$ T cell response is not always a determinant of disease outcome. A number of reasons have been suggested for this apparent immune failure. Dysfunctional $CD8^+$ T cell responses have been demonstrated in some patients with chronic infection. These dysfunctions include impaired cytokine production (IFN-γ and tumour necrosis factor (TNF)-α) (Gruener *et al.*, 2001; Spangenberg *et al.*,

2005; Wedemeyer *et al.*, 2002), reduced lytic activity due to low levels of perforin (Gruener *et al.*, 2001) and low proliferative capacity (Wedemeyer *et al.*, 2002). Mutational escape by the virus may also occur and in fact most $CD8^+$ T cells do not proliferate against the circulating antigen but rather a previous antigen which has been eliminated by viral escape mutations (Neumann-Haefelin *et al.*, 2008).

Despite the ability of the liver to delete $CD8^+$ T cells, virus-specific T cells may persist in the liver for many years and continue to contribute to disease pathology (Neumann-Haefelin *et al.*, 2008; Spangenberg *et al.*, 2005; Wedemeyer *et al.*, 2002). As suggested by depletion studies, $CD8^+$ T cells may be insufficient to clear the virus without $CD4^+$ T cell help (Grakoui *et al.*, 2003). This hypothesis is supported by the findings that $CD4^+$ Th1 cell responses are lower in chronically infected patients (Wedemeyer *et al.*, 2002) and that resolving acute infection is usually characterised by a broad $CD4^+$ T cell response which targets multiple epitopes (Schulze Zur Wiesch *et al.*, 2005). An impaired $CD4^+$ Th1 response during the acute phase is associated with the failure to mount an effective, memory $CD8^+$ T cell response (Urbani *et al.*, 2006). Recently it has been suggested that a $CD4^+$ T cell response develops in all acute HCV infections but that these $CD4^+$ T cells fail to proliferate and rapidly decline in those patients who develop a persistent infection (Schulze Zur Wiesch *et al.*, 2012).

The relative roles played by $CD8^+$ and $CD4^+$ T cells in resolution of infection need to be carefully considered for future vaccination strategies. The importance of both $CD4^+$ and $CD8^+$ T cells in spontaneous resolution of HCV infection is well established although the precise mechanisms for failure are

still being deciphered. It is also important to remember that $CD4^+$ T helper cells are required to stimulate an effective memory B cell response (McHeyzer-Williams *et al.*, 2012); therefore a defective $CD4^+$ T cell response may inhibit the development of the antibody response. As such it may be necessary to simultaneously target both humoral and cellular arms of the immune response to ensure that any vaccine has a protective effect.

1.8.2 The antibody response to HCV infection

1.8.2.1 Evidence of a protective antibody response in HCV

Due to the asymptomatic nature of acute HCV infection it is difficult to identify patients in the early phase of disease. This can make analysis of the role played by antibodies in spontaneous recovery problematic. However, early evidence of a protective effect came from a retrospective study of patients receiving polyclonal immunoglobulin (Ig) against hepatitis B virus antigen (HBIG). Prior to the introduction of screening for HCV infection, patients who received HBIG were less likely to develop HCV than those patients who received screened HBIG. Subsequently, anti-HCV antibodies were found in HCV-negative patients receiving unscreened HBIG indicating that anti-HCV antibodies had been passively transferred from donor to recipient within the HBIG and that these antibodies were protective (Feray *et al.*, 1998). Hypogammaglobulinaemic patients exhibit a rapid disease progression upon infection with HCV, suggesting that antibodies contribute to controlling chronic infection (Bjoro *et al.*, 1994). However, others report spontaneous resolution of acute HCV in hypogammaglobulinaemic patients, suggesting that

cellular immunity may clear infection in some cases (Razvi *et al.*, 2001). Finally, HCV-positive patients with depleted B cell populations following rituximab therapy experience an increase in viral load which returns to normal after cessation of treatment (Ennishi *et al.*, 2008), presumably as the B cell population recovers. Therefore B cells and the antibody response may be involved in maintenance of a steady viral load during chronic infection.

1.8.2.2 Neutralising antibody response in acute vs. chronic disease

Studies of the acute phase nAb response in patient cohorts suggest that nAb do play a critical role in disease outcome. In a cohort infected with a single strain of HCV (genotype 1b strain AD78), studies using HCVpp bearing the autologous AD78-derived E1E2 demonstrated that rapid induction of nAb early during infection was associated with spontaneous recovery. Patients who developed a chronic infection generated no antibodies or a very low titre antibody response in the early phase of the infection. Antibodies isolated from spontaneous resolvers showed broader cross-neutralising activity than those with chronic infection. In contrast, when the nAb response was assessed several years later, nAb were detectable at high titres in chronically infected patients, although these were unable to clear the virus. Neutralising antibodies were largely undetectable at this stage in patients who had resolved infection (Pestka *et al.*, 2007). In a cohort of injecting drug users infected with different strains of HCV a similar pattern was observed. Using patient-specific E1E2 sequences from sequential time points to generate HCVpp, a high titre, autologous nAb response was observed early during infection in all individuals who spontaneously resolved the infection but was absent in those who

developed a chronic infection (Dowd *et al.*, 2009). Despite these striking observations it remains to be established which viral epitopes are targeted by the acute phase antibody response, raising the possibility that the very early antibody response in these patients targets novel viral epitopes. These findings also contrast with many studies carried out in chimpanzees which show poor induction of nAb in animals that spontaneously control the infection (Logvinoff *et al.*, 2004; Meunier *et al.*, 2005). However, this may be due to observed differences in disease progression between chimpanzees and humans (discussed in section 1.6.1) which may make chimpanzees an unreliable model organism.

One of the best studied cases of HCV infection is that of patient H who became infected in 1977 after receiving a blood transfusion and developed a chronic infection. This individual is the source of the 1a strain H77 which is used in most studies of HCV and forms the basis for several assays. A study of the nAb response in patient H over a period of 26 years showed that the early antibody response was targeted toward HVR1 but was not maintained (von Hahn *et al.*, 2007), consistent with the view that HVR1 acts as an immunological decoy (Ray *et al.*, 1999). The induction of a broadly nAb response targeting E1E2 was not observed until the chronic phase (Bartosch *et al.*, 2005; von Hahn *et al.*, 2007).

1.8.2.3 Dynamics of the antibody response during infection

Over the years many studies of anti-HCV antibody responses have relied upon neutralisation of HCVpp bearing H77 E1E2 sequences as a marker of neutralising ability. However several studies have shown that the nAb

response is dynamic and drives the evolution of viral envelope sequences (Dowd *et al.*, 2009; Farci *et al.*, 2000; von Hahn *et al.*, 2007). Both von Hahn *et al.*, (2007) and Dowd *et al.*, (2009) show that sequential serum samples are limited in their ability to neutralise the concurrently circulating viral strains but are able to efficiently neutralise virus strains from earlier time points. Therefore even within one individual, a single viral envelope sequence may not be sufficient to 'pull-out' all nAb. Indeed, the use of heterologous H77-HCVpp to measure nAb is less sensitive than the use of autologous HCVpp. This over-reliance upon H77 may explain why the nAb response appears to lag behind the cellular immune response and why it appears to develop so much later than in other viral infections (Dowd *et al.*, 2009). A recent study also shows that patient-derived E1E2 sequences are more resistant to neutralisation than H77 E1E2 which may have resulted in an over-estimation of the number and potency of nAb. The frequency of patient sera with cross-neutralising activity was also lower than previously reported which was attributed to the limited number of viral genotypes against which many antibodies or sera have been tested (Tarr *et al.*, 2011). This highlights the importance of choosing an appropriate assay method to accurately assess the neutralising potential of a given antibody and possible shortcomings in the way broadly nAbs are defined.

1.8.3 Specific viral targets of neutralising antibodies

1.8.3.1 HVR1 binding sites within E2

HVR1 is a linear epitope located at aa 384-410 of E2 (Figure 1-2) that plays an important role in SR-BI binding (see section 1.7.5). It is immunogenic and anti-HVR1 antibodies have been identified *in vivo* (Kato *et al.*, 1994; Kato *et*

al., 1993; Weiner *et al.*, 1992) although these antibodies tend to be strain specific (Bartosch *et al.*, 2003a; Shimizu *et al.*, 1996; Vieyres *et al.*, 2011) and high concentrations are required for effective neutralisation (Bartosch *et al.*, 2003a). Nevertheless, HVR1 plays an important role in disease outcome.

Antibodies binding to HVR1 are either neutralising or non-neutralising depending on the region of HVR1 to which they bind. The HVR1-specific rat MAb 9/27 inhibits SR-BI binding to E2 and neutralises infectivity of HCVpp bearing genotype 1a-derived envelope glycoproteins (Bartosch *et al.*, 2003a; Hsu *et al.*, 2003). The epitope bound by 9/27 maps to aa 396-407, within the C-terminal region of HVR1. Other nAbs, such as the mouse MAb AP213 and the polyclonal sera R1020 and R140, also bind to the C-terminal region of HVR1 (Vieyres *et al.*, 2011). In contrast, antibodies binding to epitopes within the N-terminal region of HVR1 (aa 384-395) including rat MAbs 7/59, 6/16 and 6/82a, are non-neutralising (Hsu *et al.*, 2003). Therefore, HVR1 possesses two immunogenic domains which differ in their ability to stimulate a nAb response.

1.8.3.2 HVR1 as an immunological decoy

Infectivity of HCVpp and HCVcc is enhanced by HDL in an HVR1 and SR-BI dependent manner (Bartosch *et al.*, 2005; Voisset *et al.*, 2005). This interaction also inhibits neutralisation of HCVpp and HCVcc by antibodies binding to E2 regions outside of HVR1 (Bartosch *et al.*, 2005; Dreux *et al.*, 2006). HVR1 is also able to mask some nAb epitopes within E2 and deletion of HVR1 increases neutralisation by CD81 binding site-specific human MAbs

and patient sera (Bankwitz *et al.*, 2010; Prentoe *et al.*, 2011) presumably by exposing these CD81 binding sites. Therefore HVR1 may protect viral entry determinants within E2 from neutralisation during the early stages of entry (Bankwitz *et al.*, 2010). Thus HVR1 appears to play multiple roles during HCV entry which are somewhat contradictory in nature. This has led some to suggest that HVR1 functions as an immunological decoy.

Studies of viral evolution during the acute phase of infection suggest that resolved infection is associated with stable HVR1 sequences. In contrast persistent infection is characterised by considerable sequence change within HVR1 (Farci *et al.*, 2000; Ray *et al.*, 1999). The immunological decoy hypothesis posits that HVR1 stimulates a strong antibody response that is unable to clear the virus, but which instead drives mutational escape of the virus (Ray *et al.*, 1999). Viruses able to mutate rapidly within HVR1 will escape the nAb response and persist, whereas viruses with constrained evolution of HVR1 do not escape the antibody response and are cleared. The HVR1-focussed acute phase antibody response of patient H and the emergence of escape mutations within HVR1 leading to chronic infection may reflect this situation occurring *in vivo* (von Hahn *et al.*, 2007). However, recently published work contradicts the hypothesis that stable HVR1 sequences correlate with viral clearance. Liu *et al.*, (2012) observed that HVR1 evolved rapidly in patients who spontaneously resolved infection in response to the induction of a strong HVR1-specific nAb response. Sequence change within HVR1 was not observed in patients who progressed to chronic infection due to the lack of a nAb response driving HVR1 evolution (Liu *et al.*, 2012). The failure to identify the target of the antibody response in some spontaneous

resolvers (Dowd *et al.*, 2009; Pestka *et al.*, 2007) also suggests that other broadly neutralising epitopes may contribute to clearance of the virus. Therefore the ability to induce a strong nAb response targeting multiple epitopes, including HVR1, early during infection may be a more critical determinant of disease outcome.

1.8.3.3 The CD81 binding region within E2

A number of CD81 binding sites have been identified within E2 (see section 1.7.4). Among these, residues W420, Y527, W529, G530 and D535 and the ^{436}GWLAGLFY^{443} motif are highly conserved across genotypes. In the 3D structural model of E2, the CD81 binding sites are oriented in close proximity to each other within DI (Krey *et al.*, 2010). Antibodies targeting both linear and conformational epitopes have been isolated and anti-CD81 MAbs are the most numerous identified to date.

A linear epitope immediately downstream of HVR1, aa 412-423, is recognised by the broadly neutralising mouse MAb AP33 (Owsianka *et al.*, 2005) and the rat MAb 3/11 (Flint *et al.*, 1999a)(Figure 1-2a), although residues critical for binding differ between the two MAbs leading to differences in neutralising potency (Tarr *et al.*, 2006). Although the epitope bound by both MAbs is linear, optimal binding is dependent on the overall conformation of the E2 glycoprotein (Tarr *et al.*, 2006), showing the importance of using native E2 when studying antibodies. AP33 shows broad neutralising activity against the major genotypes of HCV (Owsianka *et al.*, 2005) making this region of E2 an interesting target for vaccination. However

antibodies targeting this region of E2 are found in <5% of individuals with resolved or chronic infection (Tarr *et al.*, 2007) suggesting that this region is less immunogenic in humans than in rodents. Two other antibodies targeting this region, HCV1 and 95-2, have been identified in transgenic mice bearing human Ig genes (Broering *et al.*, 2009), further evidence that this domain is preferentially targeted by the murine immune response. Murine MAbs recognising conformational epitopes have been identified (H35 and H48), however these show limited breadth of neutralisation (Cocquerel *et al.*, 1998; Owsianka *et al.*, 2006).

Many human MAbs targeting the CD81 binding site recognise conformational epitopes. E2 has been divided into five antigenic regions (AR); AR1, AR2, AR3 (Law *et al.*, 2008), AR4 and AR5 (Giang *et al.*, 2012) with AR3 consisting of discontinuous epitopes aa 396-424, aa 436-447 and aa 523-540 (Figure 1-2a). A number of neutralising MAbs have been mapped to this region including 1:7 and A8 (Johansson *et al.*, 2007b), AR3A, AR3B, AR3C and AR3D (Law *et al.*, 2008), CBH5 and CBH2 (Hadlock *et al.*, 2000; Owsianka *et al.*, 2008), HC-1 (Keck *et al.*, 2008a) as well as the Fab e137 (Perotti *et al.*, 2008) and Fab e20 (Mancini *et al.*, 2009). All of these antibodies are broadly neutralising and MAbs AR3A and AR3B protect against infection *in vivo* (Law *et al.*, 2008). In contrast antibodies binding to AR1 and AR2 are poorly neutralising (Law *et al.*, 2008). The human MAbs CBH4B, CBH4D and CBH4G, that bind a region distinct from the three described ARs, are also non-neutralising (Keck *et al.*, 2004a). Therefore AR3 may represent an immunogenic, neutralising determinant within E2, which encompasses the CD81 binding site (Law *et al.*, 2008). CD81 binding site-specific MAbs have

also been described that bind to the regions aa 410-446 (overlapping with residues forming AR3) and aa 611-617, indicating that DIII of E2 may also be involved in CD81 binding. This group of MAbs, termed HC-84, were broadly neutralising and did not lead to neutralisation escape mutations (Keck *et al.*, 2012).

Antibodies recognising the CD81 binding site have been described that show no neutralisation or limited neutralisation. The rat MAb 1/39 binds aa 432-443, overlapping the CD81 binding site aa 432-447, but shows only limited neutralisation of HCVpp entry. MAbs 9/75 and 2/64a both recognise the ^{524}APTYSWGA^{531} motif (overlapping aa 528-535) (Figure 1-2a), yet whereas 2/64a shows intermediate neutralisation of entry, 9/75 is non-neutralising (Hsu *et al.*, 2003). Therefore the presence of antibodies targeting CD81 binding does not necessarily correlate with neutralisation of infection and subtle differences in antibody specificity may alter their neutralising potency.

Recently a novel antibody was described with inhibitory activity against both CD81 and SR-BI. The broadly neutralising mouse MAb, H77.39, binds a region of E2 encompassing aa 384-520, with critical binding residues located at N415 and N417 (Sabo *et al.*, 2011) and therefore overlapping the epitope bound by AP33 (Owsianka *et al.*, 2005) and 3/11 (Flint *et al.*, 1999a). Its ability to inhibit CD81 binding to E2 is presumably mediated via binding to, and blocking, the CD81 binding site at aa 412-423. However, it is not yet clear how it might also inhibit SR-BI, although its close proximity to HVR1 may cause steric inhibition of SR-BI binding to E2 (Sabo *et al.*, 2011).

1.8.3.4 The fusion determinant and stem region

The fusion determinant of HCV entry, thought to reside within one of the envelope glycoproteins, has not yet been conclusively identified. However, studies suggest that patient-derived antibodies can block virus entry by inhibiting the fusion process (Haberstroh *et al.*, 2008; Kobayashi *et al.*, 2006). More recently broadly neutralising antibodies targeting the region aa 496-515 (which overlaps the putative fusion determinant at aa 502-520) have been described following vaccination of mice with recombinant E1E2 (Kachko *et al.*, 2011) although their ability to block fusion specifically was not tested. Identifying the function of such antibodies could help to confirm the location of the fusion determinant and provide another target for antibody-based therapies, in a manner analogous to the inhibition of HIV-1 gp41 fusion (Karlsson Hedestam *et al.*, 2008) and influenza virus haemagglutinin (HA) fusion (Sui *et al.*, 2009) by nAbs.

Antibodies targeting other epitopes within E2 have also been described. In contrast to AR3, which encompasses a CD81 binding site within E2, the newly described AR4 and AR5 encompass discontinuous epitopes within the E1E2 complex. Antibodies targeting these AR do not block E2 binding to CD81 yet are able to neutralise infectivity *in vivo*. Binding of MAb AR4A maps to the membrane proximal external region (residue D698) within the stem region, indicating an additional target for the neutralising antibody response beyond those previously described (Giang *et al.*, 2012).

1.8.3.5 Antibodies targeting E1

Antibodies targeting epitopes within E1 are rare in patient sera (Leroux-Roels *et al.*, 1996; Pestka *et al.*, 2007). This may be due to the immunological dominance of E2, the masking of E1 epitopes by E2 (Garrone *et al.*, 2011) or glycan-shielding of immunogenic domains within E1 (Fournillier *et al.*, 2001). More simply, it may be caused by technical difficulties in detecting anti-E1 antibodies, as E1 misfolds unless it is co-translated with E2 (Dubuisson *et al.*, 1994).

Characterised E1-specific MAbs are limited in number. The human MAb H-111 binds the ^{192}YEVRNVSGVYH^{211} epitope near the N-terminus of E1 from genotypes 1a, 1b, 2b and 3a but not genotypes 2a or 4a suggesting partial conservation of this epitope. H-111 binding to VLPs can reduce entry into Molt-4 cells, albeit with a maximum 70% inhibition of entry (Keck *et al.*, 2004b). The human MAbs IGH505 and IGH526 show broad inhibition of entry of HCVpp, as well as neutralising HCVcc infectivity. Both antibodies map to a linear epitope within E1 encompassing aa 313-327 (Meunier *et al.*, 2008) showing that some regions of E1 are immunogenic and elicit a nAb response, although the frequency of such antibodies is low.

Several trials of E1 protein vaccine candidates have demonstrated that anti-E1 antibody responses are induced (Garrone *et al.*, 2011; Leroux-Roels *et al.*, 2004; Nevens *et al.*, 2003), although the efficacy of the anti-E1 response is unclear. In one study, viral load in chronically infected patients was not reduced although improved liver histology was reported (Nevens *et al.*, 2003). A recent study of VLPs expressing E1, used in a prime-boost protocol with soluble E1, reported the induction of an E1-specific antibody response but this

was not neutralising. Neutralisation was only observed when animals were also vaccinated with E1E2 (Garrone *et al.*, 2011). In contrast, induction of anti-E1 antibodies following immunisation with a recombinant form of E1 that was protective against challenge with a heterologous HCV has been reported. This was more effective than an E2-based vaccine (Verstrepen *et al.*, 2011).

The role played by E1 in HCV entry is still unclear. As yet it has not been shown to bind to the entry receptors, CD81, SR-BI, CLDN-1 or OCLN. It has been proposed to function as the fusion determinant (Lavillette *et al.*, 2007) although this is in disagreement with the proposed 3D model of E2 (Krey *et al.*, 2010). It therefore remains to be seen which aspect of E1 structure and function might best be targeted in the search for nAbs.

1.8.3.6 Other targets for the neutralising antibody response

To date, antibodies inhibiting E2 binding to CD81 and SR-BI have been the focus of research. However the discovery of several additional receptors and entry factors has opened up new targets. Although OCLN co-precipitates with E2 (Benedicto *et al.*, 2008; Liu *et al.*, 2009), no virus-specific antibodies targeting this interaction have been identified. Anti-CLDN-1 antibodies have been demonstrated to inhibit the formation of the E2-CD81-CLDN-1 co-receptor complex (Fofana *et al.*, 2010; Krieger *et al.*, 2010) and thereby reduce infectivity of HCVcc and HCVpp. The usefulness of anti-receptor rather than anti-virus antibodies has also been demonstrated with antibodies targeting CD81 (Fafi-Kremer *et al.*, 2010; Meuleman *et al.*, 2008; Molina *et al.*, 2008), SR-BI (Catanese *et al.*, 2010; Meuleman *et al.*, 2012) and, more recently, the

newly described co-factors EGFR (Lupberger *et al.*, 2011) and NPC1L1 (Sainz *et al.*, 2012) which were able to block or inhibit viral entry. Furthermore, combinations of anti-receptor antibodies had an additive effect on neutralisation of infectivity (Krieger *et al.*, 2010) as did a combination of anti-receptor plus anti-HCV antibodies (Fofana *et al.*, 2010). Targeting components of normal human cells may not be appropriate in healthy individuals; however anti-receptor antibodies may provide an important facet of future therapies to control the spread of HCV, particularly among transplant recipients.

1.8.4 Virus mechanisms for evading antibody neutralisation

Neutralising antibodies can clear acute infection, yet broad nAb responses are commonly seen in chronically infected patients. This implies that the virus must be able to subvert or evade the immune response in order to establish a persistent infection. Several mechanisms have been described by which HCV may persist.

1.8.4.1 Mutational escape

The HCV single-stranded RNA genome has a high mutation rate due to the error-prone nature of the RNA-dependent RNA polymerase and the rapid turnover of the virus (Neumann *et al.*, 1998). This results in a diverse population of viral variants within the host, with variable sensitivity to a given antibody. In the presence of nAb, escape variants will have a selective advantage over neutralisation-sensitive variants and will rapidly become the

dominant strain. This situation has been observed *in vivo*, where evolution of the envelope glycoproteins is driven by the nAb response (Dowd *et al.*, 2009; Pestka *et al.*, 2007; von Hahn *et al.*, 2007). In chronic infection, positively selected amino acids are found in E2 and are situated within or in close proximity to receptor binding sites and nAb epitopes (Brown *et al.*, 2005; Brown *et al.*, 2007) and a high number of selected sites within E2 is associated with rapid progression to persistent infection (Sheridan *et al.*, 2004). Therefore mutational escape is a key weapon in HCV's ability to persist despite a strong nAb response.

1.8.4.2 Glycan shielding

The envelope glycoproteins of HCV are highly glycosylated structures and the 11 N-linked glycans of E2 play an important role in its structure, function and immunogenicity (see section 1.5.2). Specific glycans at positions N1, N2, N4, N6 and N11 (Figure 1-2a) modulate HCVpp and HCVcc binding to CD81 and removal of these glycans increases entry inhibition by the CD81-LEL (Falkowska *et al.*, 2007; Helle *et al.*, 2007; Helle *et al.*, 2010). Furthermore, removal of these glycans increases the sensitivity of HCVpp and HCVcc to neutralisation by patient sera and MAbs (Falkowska *et al.*, 2007; Helle *et al.*, 2007; Helle *et al.*, 2010). This has led to the suggestion that HCV possesses a glycan shield, which protects certain epitopes from neutralisation. The finding that some of these glycans also modulate CD81 binding shows that a fine balance has to be achieved by the virus, namely the protection of entry determinants from neutralisation without blocking their function. Both HIV-1 gp120 and influenza HA possess a glycan shield which affords the virus

protection from neutralisation without blocking entry. However, these glycan shields are evolving, continually shifting the position as well the frequency of glycans on the exposed face of the glycoprotein (Abe *et al.*, 2004; Wei *et al.*, 2003). HCV does not exhibit such diversity in its glycosylation state. Indeed, of the 11 potential N-linked glycosylation sites on E2, nine show conservation of >97% across the genotypes (Falkowska *et al.*, 2007) with N7 absent from genotype 3 as well as most genotype 6 sequences and N5 missing in some subtypes (Helle *et al.*, 2007). The majority of the glycans in E2 do not undergo positive selection, with the exception of N5 in genotype 1, N6 in genotype 5 and N9 in genotype 4 (Brown *et al.*, 2007) and some glycans are critical for entry (Falkowska *et al.*, 2007) or infectious particle production (Helle *et al.*, 2010). Therefore HCV is not able to evolve the glycan shield in response to a changing nAb repertoire, but glycans may protect some entry determinants from neutralisation at key steps during entry.

1.8.4.3 Cell-to-cell transmission

The mechanism by which cell-free HCV attaches to the target cell-surface and enters is gradually being deciphered (Figure 1-3 and section 1.7.9). However, it has recently been discovered that HCV is also capable of direct cell-to-cell transmission (Timpe *et al.*, 2008). This requires the entry receptors CD81, SR-BI, CLDN-1 and OCLN and occurs with all viral genotypes (Brimacombe *et al.*, 2011). Cell-to-cell transmission is largely resistant to E2-specific antibody-mediated neutralisation (Timpe *et al.*, 2008) although it is partially inhibited by the rat MAbs 9/27 and 11/20c which target the HVR1 epitope, aa 396-407, and the CD81 binding sites, aa 412-423 and 436-447,

respectively (Brimacombe *et al.*, 2011). However, it is sensitive to anti-SR-BI antibody mediated neutralisation (Meuleman *et al.*, 2012) suggesting an important role for SR-BI in this mechanism of viral spread. Direct cell-to-cell transmission as a mechanism of immune evasion is characteristic of a number of other enveloped viruses, such as the herpesviruses, measles virus, and the retroviruses HIV-1, murine leukaemia virus and human T-cell lymphotrophic virus (Mothes *et al.*, 2010). Such direct cell-to-cell transmission may have implications in terms of antibody-based therapeutics. Once the virus has entered hepatocytes, cell-to-cell transmission may contribute significantly to the establishment of a chronic infection. Therefore only a relatively short window may exist for antibody neutralisation to prevent infection of the liver and this will be an important consideration when treating transplant recipients.

1.8.4.4 Non-neutralising antibody interference

Neutralisation by antibodies may be further inhibited by the presence of non-neutralising antibodies (non-nAb) circulating within patient sera, although this is an area of some disagreement. It has been suggested that non-nAbs, binding distinct epitopes within E2, block antibody binding to neutralising epitopes (Zhang *et al.*, 2007). Two distinct epitopes have been identified within E2; epitope I (aa 412-419), reported to function as the neutralising epitope and epitope II (aa 434-446), reported to bind non-nAbs that block binding to epitope I (Zhang *et al.*, 2007). Furthermore, depletion of epitope II-specific antibodies enabled the detection of nAbs in sera that were previously deemed to be non-neutralising (Zhang *et al.*, 2009). This mechanism of antibody interference was proposed as a reason for the treatment failure of polyclonal Ig

preparations (Davis *et al.*, 2005); however, this is in contrast to the protective effects of polyclonal Ig (HBIG) previously reported (Feray *et al.*, 1998). A recent study found that affinity-purified Igs and MAbs targeting epitope II neutralised HCVpp and HCVcc entry and there was no observed interference between epitope I and epitope II-specific antibodies (Tarr *et al.*, 2012b). Broadly neutralising, conformation sensitive MAbs targeting epitope II have also been described (Keck *et al.*, 2012). Whilst this suggests that epitope II does not constitute a non-neutralising epitope *per se*, different MAbs bind different but overlapping epitopes within this region with varying potency. Therefore, it remains to be seen if individual antibodies are able to interfere with virus neutralisation or if different epitopes elicit a non-neutralising response.

1.8.4.5 Lipid shielding

Lipoproteins form an important component of the HCV particle and play a key role in viral entry via binding to LDLr and SR-BI (as discussed in sections 1.7.3 and 1.7.5). Components of HDL, such as apoCI, can enhance the infectivity of HCVpp and HCVcc (Bartosch *et al.*, 2005; Dreux *et al.*, 2007; Meunier *et al.*, 2005), reduce the sensitivity of HCVpp to nAbs (Dreux *et al.*, 2006) and enhance fusion between viral and cellular membranes (Dreux *et al.*, 2007). This has led to the suggestion that lipid shielding protects the virus from neutralisation and facilitates infectivity, although the precise interactions occurring between the virus, lipoproteins, receptors and antibodies remains to be deciphered.

1.8.5 Role of antibody effector functions in HCV clearance

In addition to mediating viral clearance by neutralisation of entry, antibodies carry out specific effector functions that control infection via non-neutralising mechanisms. Antibody-dependent cellular cytotoxicity (ADCC) is the lysis of virus-infected target cells by macrophages, neutrophils and natural killer (NK) cells following binding to the Fc region of antibody molecules bound on the surface of the target cell. This is mediated by Fc receptors (FcR) expressed by the effector cells (Eales, 2001). Complement-dependent cytotoxicity (CDC) requires binding of the complement protein C1q to the Fc region of an antibody bound to a virus particle or virus-infected cell. This triggers the classical complement cascade leading to the formation of a membrane attack complex (MAC) on infected cells causing lysis of the target (Tarr *et al.*, 2012a). ADCC is an important component of the immune response to HIV that may enhance the activity of nAbs or facilitate inhibition by otherwise non-nAbs (Chung *et al.*, 2008) but its role in HCV infection is less well understood. Acute- and chronic-phase sera can mediate ADCC via binding to E2 expressed on the surface of infected cells (Natterman *et al.*, 2005). Interestingly, in this study, chronic sera showed greater ADCC, compared to the acute phase sera, suggesting that ADCC does not necessarily clear the virus. The broadly neutralising MAbs CBH2 and CBH5 as well as the non-neutralising MAb CBH4 possess CDC activity on target cells (Machida *et al.*, 2008).

Mechanisms by which the virus evades these immune responses have been described. HCV-induced hypermutation of Ig genes has been proposed as a potential mechanism to subvert CDC (Machida *et al.*, 2008) although this would presumably also affect antibody-mediated neutralisation and ADCC.

The core protein of HCV possesses an FcγR-like domain, spanning aa 3-75, that can bind IgG molecules via their Fc region (Maillard *et al.*, 2004). Therefore, core binding to the Fc region of antibodies would inhibit effector functions such as ADCC and CDC. Recently, it has been suggested that HCV particles incorporate the host protein CD59 into their envelope. CD59 is a regulator of complement activation that prevents formation of the MAC and thereby inhibits CDC of target cells or virus particles. Indeed, inhibition of CD59 was found to increase lysis of cell culture and patient-derived virus particles (Amet *et al.*, 2012). It has also been reported that sub-neutralising concentrations of antibodies enhance infectivity through binding of virus-antibody complexes to the FcRII receptor on hepatocytes (Meyer *et al.*, 2011). A greater understanding of antibody-mediated effector functions and their contribution to HCV pathogenesis will inform the design of effective vaccines or therapeutics. It will be particularly important to understand evasion tactics employed by the virus to ensure that treatment is ultimately successful.

1.9 Neutralising antibodies in vaccines or therapeutics

There is currently no vaccine against HCV and preventing reinfection of a newly grafted liver is not yet possible. Therefore identifying protective immune determinants that may be induced by a vaccine is an important goal for HCV researchers. The correlation between strong T cell (Schulze Zur Wiesch *et al.*, 2005; Thimme *et al.*, 2002) and nAb responses (Dowd *et al.*, 2009; Pestka *et al.*, 2007) and the spontaneous resolution of acute infection indicates that the immune system plays an important role in disease outcome.

The protection afforded by polyclonal immunoglubulins (Feray *et al.*, 1998) and resistance to reinfection of IVDUs who spontaneously resolve primary infection (Osburn *et al.*, 2010) support the notion that protection against HCV infection can be generated by vaccination.

Several trials with vaccine candidates have been carried out but have met with mixed results. A formulation of HCV-enriched Ig (Civacir) reduced levels of alanine transaminase (ALT), a marker of liver injury, and caused a transient reduction in viral load but could not prevent reinfection of the liver in transplant recipients (Davis *et al.*, 2005). A neutralising, human anti-E2 MAb HCV-Ab^{XTL}68, reduced HCV RNA levels in patients undergoing liver transplantation but this was not maintained long term and therapeutically useful levels of serum antibody were not achieved (Schiano *et al.*, 2006). A recombinant form of E1E2 developed by Chiron Corp. has proven effective in chimpanzees, inducing a strong T cell and nAb response that completely protected against infection with the homologous virus strain and prevented chronic infection following experimental challenge with a heterologous virus (Houghton & Abrignani, 2005). Preliminary studies of an adjuvanted form of the vaccine, HCV E1E2/MF59.C1, indicate that it is immunogenic, capable of inducing cellular and humoral immune responses in healthy individuals. Furthermore, few adverse events were reported suggesting that the safety profile of the vaccine is good (Frey *et al.*, 2010). Results of phase I clinical trials are awaiting publication. A recombinant form of E1 trialled in patients suffering a chronic, genotype 1 infection was found to induce an E1-specific T cell and antibody response that reduced ALT levels and fibrosis but did not affect viral load (Nevens *et al.*, 2003). The authors suggest that E1-based

therapy may improve liver pathology; however a subsequent study found no improvement in liver histology despite the induction of cellular and humoral immunity (Wedemeyer *et al.*, 2008).

Other vaccine candidates have focussed on inducing cellular immunity against the non-structural and/or core proteins of HCV. These include a synthetic peptide vaccine containing conserved $CD8^+$ and $CD4^+$ T cell epitopes (IC41) from core, NS3 and NS4 that is able to induce a strong cellular immune response. However, trials in chronically infected patients have shown mixed results in terms of reducing viral load (Torresi *et al.*, 2011), highlighting the importance of choosing an appropriate dosing regimen in achieving a favourable outcome. Another novel vaccine candidate expressing a core-NS3 protein, termed GI-5005, induces a strong T cell response that reduces viral load and ALT levels (Torresi *et al.*, 2011). Furthermore, when combined with standard therapy (ribavirin plus IFN-α), GI-5005 improved early virological response rates and SVR rates compared to standard therapy alone (Jacobson *et al.*, 2010). DNA vaccines have also proven successful, specifically a codon-optimised NS3/NS4A gene expressed from a cytomegalovirus (CMV) immediate early (IE) promoter (ChronVac-C). In studies of genotype 1 infected, treatment-naive patients, vaccination reduced viral load in 67% of recipients and was temporally associated with the induction of an HCV-specific T cell response (Sallberg *et al.*, 2009). No adverse events were reported indicating that this vaccination approach is safe although further trials are ongoing.

Given the variable nature of HCV and its ability to evade or subvert both the humoral and cellular arms of the immune response, preventive or therapeutic

vaccines may need to target multiple viral components. Targeting a single viral epitope is likely to lead to the development of escape mutations and therefore resistant strains. The rational design of immunogens to be included in any vaccine is likely to require an approach similar to that used in the treatment of HIV – simultaneous targeting of multiple viral proteins or epitopes. Including host targets (e.g. entry receptors) as well as viral targets may broaden the scope and effectiveness of vaccination. A recent study of the antibody response to HIV suggests that a combination of less potent MAbs, that individually do not neutralise infection, may be more representative of the *in vivo* situation and provide greater protection than a single MAb (Walker *et al.*, 2011). Such an approach would require the identification of multiple antibodies targeting distinct but overlapping targets.

1.10 General aims

Since the discovery of HCV more than 20 years ago, considerable advances have been made in our understanding of this important pathogen. Advances continue to be made in the search for antiviral therapies; however the current standard of therapy is poorly tolerated and unable to resolve all cases of the disease. The burden of HCV-related liver transplantation and the inability to protect the newly grafted organ from reinfection has heightened the search for therapeutic or preventive vaccines. Neutralising antibodies may provide an important component of new vaccines or therapies and improving our knowledge of such antibodies will aid in the design of immunogens.

The majority of nAbs identified to date target CD81 binding sites within E2. However, this may be due to the early bias of experimental methods towards CD81-specific antibodies and it has become clear in recent years that a number of other factors are involved in HCV entry (see section 1.7). The identity of the fusion determinant is also unclear (Krey *et al.*, 2010; Lavillette *et al.*, 2007) but it appears to elicit nAbs capable of blocking infection. The immunogenicity of E1 has also been a subject of debate but anti-E1 antibodies have been reported (Keck *et al.*, 2004b; Meunier *et al.*, 2008) and may contribute to resolution of infection. Any of these may feasibly provide additional targets for nAb and are deserving of thorough investigation. However, studying the full spectrum of the nAb response may be hampered by the assay systems commonly in use, many of which rely upon truncated versions of E2 or use single virus strains. Ideally a method is needed that is not biased towards a specific epitope or receptor binding region within E2 and that faithfully represents the native form of the glycoprotein. Furthermore, identifying nAb that are relevant in viral clearance *in vivo* has proven difficult. Broadly neutralising murine MAbs are limited to targeting the linear epitope aa 412-423, an epitope that is poorly immunogenic in humans (Tarr *et al.*, 2007), whereas broadly neutralising human MAbs target conformational epitopes. Therefore isolating human MAbs is more desirable at it holds the promise of providing broadly nAbs that contribute to virus control in a natural disease setting. A suitable method for screening large numbers of patients for nAbs is needed, in order to isolate those that have a protective effect *in vivo*. The overall aims at the outset were therefore to develop assay systems for the discovery and characterisation of anti-HCV antibodies. Firstly cell-surface

display technology was used to develop an E2 library with which to map MAb-binding residues. Secondly, the Ig genes were PCR amplified and sequenced from the B cells of HCV-infected individuals and finally, a study of the antibody response in a cohort of HCV-infected patients was undertaken. The aim of all three projects was to facilitate studies of nAb and to further our understanding of their role in disease outcome.

2 Design and development of an E2 cell-surface display library

2.1 Introduction

2.1.1 Approaches to peptide and protein mapping

To improve our understanding of the mechanisms of antibody neutralisation it is necessary to identify the location and structure of neutralising epitopes; that is, whether they are linear or conformational in nature. This is turn can tell us the nature of the interaction that is blocked or inhibited by nAbs. Knowledge of the location of several nAb epitopes has already informed the generation of a 3D model of E2 (Krey *et al.*, 2010). A greater understanding could help to direct the design of immunogens, therapeutic antibodies or vaccine components and several approaches have been developed to decipher the nature of antibody-receptor interactions.

Peptide scanning has been used to identify a number of linear epitopes within E2. This method uses MAb binding to a panel of overlapping peptides to determine the specific regions involved in antibody recognition (Broering *et al.*, 2009; Clayton *et al.*, 2002; Owsianka *et al.*, 2001). Linear epitopes have also been mapped using phage display libraries of random peptides (Tarr *et al.*, 2006; Vieyres *et al.*, 2011; Zhang *et al.*, 2007; Zhang *et al.*, 2009). These generally consist of a phage host displaying peptides of between 6-40 amino acids that undergo several rounds of affinity enrichment with a specific MAb or polyclonal serum. After several rounds of biopanning the library is enriched for those peptides binding the target antibody. The phages are propagated in a bacterial host forming plaques from which the DNA sequence of the selected

peptide can be amplified. Alignment of these peptide sequences will identify the amino acids likely to be involved in antibody binding (Smith & Petrenko, 1997). The use of phage display may however, be limited by the use of bacteria, such as *Escherichia coli*, for propagation of the phage host. Prokaryotic cells do not possess an ER or Golgi apparatus and, as a result, are unable to carry out the complex glycosylation and folding of proteins observed in eukaryotic cells (Weerapana & Imperiali, 2006). HCV E2 is a highly glycosylated structure containing up to 11 N-linked glycans that play an important role in its structure, function and immunogenicity (see section 1.5.2). Without complex glycosylation, the peptides expressed in a phage display library may not represent the native glycoprotein, and as such, may not be appropriate for mapping more complex epitopes. However, phage display has been successfully used to produce mimotopes of conformation-sensitive antibody binding epitopes (Prezzi *et al.*, 1996; Zwick *et al.*, 2001), although such peptides often show little or no sequence homology to the native antigen (Larralde *et al.*, 2007).

An alternative approach to mapping conformation-sensitive epitopes is the use of panels of glycoproteins containing single amino acid substitutions to probe antibody-receptor interactions. Typically alanine scanning mutagenesis is used to interrogate antibody binding to a region of particular interest. This method has been successfully used to map the epitopes bound by conformation-sensitive MAbs (Johansson *et al.*, 2007b; Keck *et al.*, 2008a; Law *et al.*, 2008; Owsianka *et al.*, 2008; Perotti *et al.*, 2008) as well as linear epitopes (Tarr *et al.*, 2006). The majority of these studies however, have relied upon preliminary studies, including peptide phage display (Tarr *et al.*, 2006),

NOB assays (Johansson *et al.*, 2007b; Law *et al.*, 2008; Owsianka *et al.*, 2008) or antibody competition experiments (Keck *et al.*, 2008a; Law *et al.*, 2008; Perotti *et al.*, 2008) to identify potential regions of interest for subsequent alanine scanning mutagenesis. This reduces the size of the protein or peptide undergoing mutagenesis but does require some prior knowledge of the likely epitopes. This approach may also miss some rare, genotype-specific variants involved in antibody recognition as it has tended to focus on regions of high conservation within E2 (Perotti *et al.*, 2008).

More recently antibody-binding epitopes have been mapped by generation and sequence analysis of *in vitro* neutralisation escape mutants. Both linear (Gal-Tanamy *et al.*, 2008) and conformation-sensitive (Keck *et al.*, 2011) epitopes have been mapped in this way. This approach has been used to identify small variations in the specific residues involved in recognition by two different antibodies, namely CBH-2 and HC-11, that bind overlapping epitopes within E2 (Keck *et al.*, 2011). Generation of the escape mutants is however, a lengthy and labour-intensive process requiring repeat passage of HCVcc in non-neutralising concentrations of antibody. In some cases no neutralisation escape mutants will be generated. This is possible if all the antibody contact residues are essential for virus fitness, for example residues involved in CD81 binding, making epitope mapping via this method impossible (Keck *et al.*, 2011).

The use of combinatorial polypeptide libraries has been widely used to map antibody- and ligand-binding epitopes within a variety of different proteins. Several different host organisms have been used for the generation of these display libraries, including bacteria, yeast, and insect cells. A number of

bacterial outer membrane, cell envelope and adhesion proteins, from both Gram-negative and Gram-positive organisms, have been modified for the surface display of foreign proteins (Lee *et al.*, 2003). Bacteria can be propagated in large numbers, grow rapidly and are easily handled; however bacterial hosts are unable to support the expression of large proteins (Lee *et al.*, 2003), lack the foldases and chaperones necessary to process disulphide-bonded proteins (Boder & Wittrup, 1997) and cannot glycosylate proteins fully, making them unsuitable for the expression of the E2 glycoprotein.

Yeast cell display library systems have been developed in *Saccharomyces cerevisiae*, the most successful of which exploits the cell surface mating adhesins for the attachment of foreign proteins. The protein folding and secretory machinery of *S. cerevisiae* is very similar to mammalian cells and a eukaryotic expression system has the advantage of processing proteins through the ER, unlike bacterial expression systems (Boder & Wittrup, 1997). This system has been used to map antibody-binding epitopes within the envelope glycoproteins of the flaviviruses West Nile Virus (WNV) (Oliphant *et al.*, 2005) and dengue virus (DENV) (Sukupolvi-Petty *et al.*, 2010; Sukupolvi-Petty *et al.*, 2007). Its use has recently been reported in mapping studies of HCV E2 (Sabo *et al.*, 2011).

An alternative eukaryotic expression system involves the use of insect cells. Typically this has employed expression of proteins on the surface of a baculovirus, such as the *Autographa californica* multiple nucleopolyhedrovirus (*Ac*MNPV). The protein of interest is attached to the *Ac*MNPV major envelope protein gp64. When the virus infects *Spodoptera frugiperda* (Sf9) cells, the gp64-fusion protein becomes localised to the surface of the insect

cells (Grabherr *et al.*, 2001). As a result, the protein of interest can be detected by flow cytometry such that a population of insect cells can be screened based on binding affinity for a test protein. A variety of different proteins have been expressed in this system, including complex mammalian and viral proteins requiring extensive post-translational modification (Grabherr *et al.*, 2001). These include the Japanese encephalitis virus E protein (Xu *et al.*, 2011), influenza A virus HA (Yang *et al.*, 2007) and HIV-1 gp120 (Toellner, 2002). Baculoviruses are able to incorporate large protein-coding sequences into their DNA and grow to high titres but the system is a single round infection cycle leading ultimately to the death of the host insect cell. Therefore it is not possible to develop stable cell lines that can be repeatedly used, making this a labour-intensive process. More recently a secreted protein expression system has been developed in the *Drosophila melanogaster* S2 insect cell line. This has been successfully used to express a number of different proteins from both transiently and stably transfected cells including hepatitis B virus surface antigen (HBsAg) (Deml *et al.*, 1999) and HIV gp120 (Culp *et al.*, 1991). More recently the HCV E2 glycoprotein expressed in *D. melanogaster* S2 cells has been shown to be largely monomeric and correctly folded. This approach also allows for the production of large amounts of protein (Krey *et al.*, 2010).

Expression of correctly folded and glycosylated truncated versions of the E2 ectodomain by transfection of the human embryonic kidney (HEK) 293 cell line (Flint *et al.*, 1999b) and the Chinese hamster ovary (CHO) cell line (Heile *et al.*, 2000) have been described. Such mammalian-derived glycoprotein has been used in studies of protein structure and function. However, due to the inherent difficulty culturing mammalian cells, their growth as monolayers

requiring direct cell-to-cell contact and the expense and technical challenge of scaling-up mammalian tissue culture, these cells would not be the ideal host for a protein display library.

The most important consideration when choosing a protein or peptide mapping method is for the protein to be correctly folded and displayed in its native form. Although E1 and E2 assemble to form non-covalent heterodimers on the surface of the virion (Deleersnyder *et al.*, 1997), E2 can be expressed alone in model systems (Michalak *et al.*, 1997). The majority of the protein generated is monomeric; however disulphide-linked aggregates also form that do not bind conformation-sensitive antibodies (Michalak *et al.*, 1997) or cellular receptors (Flint *et al.*, 2000; Heile *et al.*, 2000). The glycosylation state of E2 is also important as several N-linked glycans are critical for correct protein folding (Goffard *et al.*, 2005). Therefore presentation of the protein in its native form is essential if epitopes bound by naturally occurring antibodies are to be accurately mapped and potential targets for therapeutic antibodies identified. As described, several approaches to epitope mapping are limited in their ability to express E2 in its native form. The most promising host candidates for generation of an E2 protein display library are *S. cerevisiae* yeast cells and *D. melanogaster* S2 insect cells. These will be discussed in more detail in the following sections.

2.1.2 Principles of the yeast cell display system

The yeast cell display system was developed by Boder and Wittrup (1997) for the expression of mammalian proteins on the surface of *S. cerevisiae*

EBY100 cells. This is achieved by exploiting the natural mating machinery of yeast cells. *S. cerevisiae* exists as a diploid cell from which two haploid cells are generated during meiosis. These haploid cells may be one of two mating types – MATa or MATα. Two haploid cells of opposite mating type combine by conjugation to produce a diploid cell (Cross *et al.*, 1988). Cells of different mating type can be identified by the expression of either a-agglutinin or α-agglutinin on the cell surface. These agglutinins are induced in response to pheromones produced by cells of the opposite mating type and bring the cells into direct contact required for conjugation. Therefore MATa cells express a-agglutinin in response to the α-factor pheromone produced by MATα cells whilst MATα cells express α-agglutinin in response to the a-factor produced by MATa cells (Fehrenbacher *et al.*, 1978; Lipke & Kurjan, 1992). The a-agglutinin is composed of two subunits, a core subunit (Aga1p) and a binding subunit (Aga2p) (Figure 2-1a). The core subunit is a 725 amino acid protein rich in serine and threonine residues and is the product of the AGA1 gene (Roy *et al.*, 1991). The binding subunit is a 69 amino acid protein encoded by the AGA2 gene (Cappellaro *et al.*, 1991). Both proteins are highly O-glycosylated but lack any N-linked glycosylation sites (Cappellaro *et al.*, 1991; Roy *et al.*, 1991). The core subunit is attached to β-glucans within the yeast cell wall via its C-terminus in a reaction thought to involve binding of a glycosyl phosphatidyl inositol anchor (Lipke & Kurjan, 1992). The binding subunit is attached to the core subunit via two disulphide bonds (Cappellaro *et al.*, 1994). The binding site within a-agglutinin has been localised to the C-terminal region of Aga2p (Cappellaro *et al.*, 1994; Shen *et al.*, 2001).

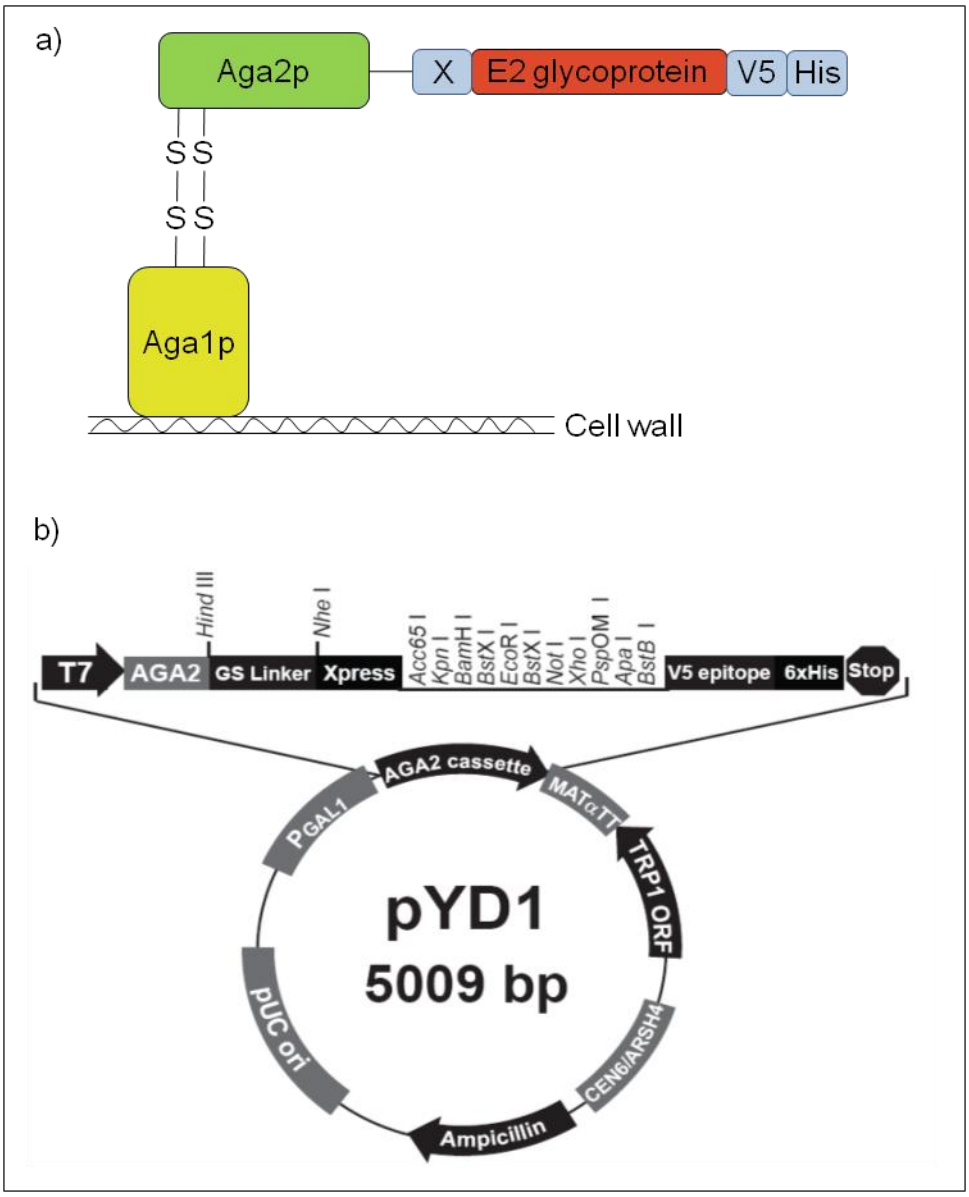

Figure 2-1. Schematic representation of the yeast display system. (a) Display of the E2 glycoprotein on the yeast cell surface via attachment to the Aga2p protein. Aga2p is attached via disulphide bonds to Aga1p which is anchored to the cell wall of *S. cerevisiae*. X, Xpress epitope; V5, V5 epitope; His, poly-histidine tag. (b) The pYD1 plasmid and AGA2 cassette (taken from pYD1 Yeast Display Vector Kit manual, cat. no. V835-01).

The yeast cell display system is based upon fusion of a non-native protein to the C-terminus of the a-agglutinin binding subunit, Aga2p, resulting in expression and display of the protein on the cell surface (Figure 2-1a) (Boder & Wittrup, 1997). The plasmid pYD1 (Invitrogen) (Figure 2-1b) is a 5.0 kb expression vector that contains the AGA2 gene derived from *S. cerevisiae*. The multiple-cloning site allows fusion of the gene-of-interest to the AGA2 gene. AGA2 expression is under the control of a galactose-inducible promoter, P_{GAL1}, allowing for controlled expression of the protein of interest. Aga1p subunit expression is also under the control of the P_{GAL1} promoter, stably integrated into the chromosome of the *S. cerevisiae* EBY100 strain (Boder & Wittrup, 1997). The Aga1p and Aga2p-fusion gene products associate within the secretory pathway and are exported to the cell surface (Boder & Wittrup, 1997). The Xpress epitope, V5 epitope and poly-histidine tag within pYD1 enable detection and purification of the cell surface expressed protein.

2.1.3 Principles of the insect cell display system

The insect cell display system is based upon the *Drosophila* Expression System (DES) developed by Invitrogen. This system uses the *D. melanogaster* S2 cell line, derived from a primary culture of late-stage (20-24 hour old) embryos (Schneider, 1972). These cells are typically spherical in appearance and grow in suspension with a tendency to form clumps of between 10 and 20 cells. Spindle-shaped cells may form at high density (Schneider, 1972). *D. melanogaster* S2 cells display characteristics similar to haemocytes, the insect equivalent of macrophages, including the ability to carry out phagocytosis (Luce-Fedrow *et al.*, 2008). Cells are easily cultured in serum-free media at

28°C without CO_2 and will grow at room temperature (Cherbas & Cherbas, 2000).

The DES utilises the plasmid pMT/BiP/V5-His (Invitrogen) to express secreted proteins in *D. melanogaster* S2 cells. Protein expression is under the control of a metallothionein-inducible promoter, pMT (Angelichio *et al.*, 1991; Bunch *et al.*, 1988), and protein secretion is controlled by the BiP secretion signal (Figure 2-2). Expression can be induced by the addition of the heavy metal ions copper, zinc or cadmium to the growth medium (Bunch *et al.*, 1988). The BiP protein is an Ig-binding chaperone protein which targets proteins to the ER lumen and facilitates protein folding (Kirkpatrick *et al.*, 1995). In this study a modified version of the plasmid pMT/BiP/V5-His was used. The plasmid, termed pMT, has a modified multiple cloning site, lacks the V5 epitope and poly-histidine tag, and contains a Strep-tag (WSHPQFEK) for protein purification (Krey *et al.*, 2010) (Figure 2-2). This plasmid has been successfully used for expression of sE2 from *D. melanogaster* S2 cells (Krey *et al.*, 2010). By cloning the gene of interest in frame with the BiP secretion signal it is possible to generate large amounts of secreted protein.

The DES is designed for secreted protein expression. Therefore it is necessary to adapt this system for cell surface display of proteins. This can be achieved by the addition of a transmembrane domain to the E2 envelope glycoprotein. The native E2 transmembrane domain contains ER retention signals which would prevent translocation of the protein to the cell surface (Cocquerel *et al.*, 1998). To overcome this problem, the native domain can be removed and replaced with the transmembrane and cytoplasmic domain of the influenza A virus HA. The transmembrane and cytoplasmic domains are 27

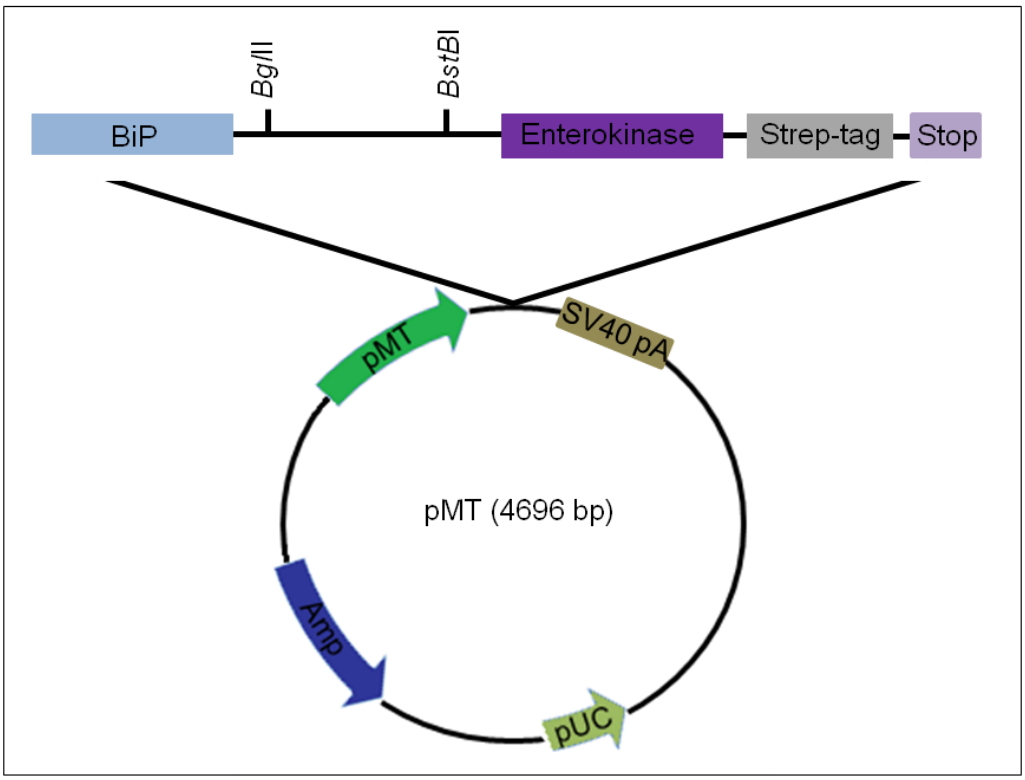

Figure 2-2. Schematic representation of the *Drosophila* expression plasmid, pMT. The expression cassette, containing the BiP secretion signal upstream of the cloning cassette, an enterokinase sequence and the Strep-Tag is under the control of a metallothionein promoter, pMT. The ampicillin resistance cassette, Amp is required for selection and the pUC origin is necessary for replication in *E. coli*. The SV40 late polyadenylation signal, SV40 pA, is necessary for transcription termination and polyadenylation of mRNA (Angelichio *et al.*, 1991). Plasmid map is based on the pMT/BiP/V5-His Vector manual (Invitrogen, Paisley, UK) plus additional information provided by Thomas Krey (personal communication).

and 10 amino acids respectively and are found within the HA2 subunit of the influenza HA (Armstrong *et al.*, 2000). This approach has been used to express a truncated version of E2 on the surface of mammalian cells. The protein is recognised by conformation-dependent antibodies and binds CD81 suggesting that it is representative of the native E2 glycoprotein (Flint *et al.*, 1999b).

Stable *D. melanogaster* S2 cell lines can be generated by cotransfection of the expression plasmid with the selection plasmid, pCoBlast. The plasmid

pCoBlast contains the blasticidin resistance gene, *bsd*, from *Streptomyces griseus* under the control of the *Drosophila* copia promoter (Anon., 2003). *Copia* elements are transposable elements with characteristics similar to some retroviral replication proteins which generate large amounts of transcript (Mount & Rubin, 1985). Stable cell lines contain multi-copy inserts of the gene of interest forming arrays of 500-1000 copies arranged in a head-to-tail fashion. The number of copies inserted can be varied by altering the ratio of expression vector to selection vector (Anon., 2003). Stable cell lines can be maintained over long periods of time by growth in blasticidin-containing media and protein expression selectively induced by addition of copper sulphate. An advantage of using *D. melanogaster* S2 cells is their rapid growth and ease of handling. Cells grow exponentially between 5 $x10^5$ and 1 $x10^7$ cells/ml, with a doubling time of 24 hours. The cells are generally stable over long culture periods and much less susceptible to gross changes in their morphology than mammalian cells (Cherbas & Cherbas, 2000). Cultures of *D. melanogaster* S2 cells can be easily scaled-up for increased protein production.

2.2 Aims of the work presented in this chapter

The major aim of this work was to develop a cell-surface display library system for the expression of randomly mutated HCV E2 protein. This library would then be used to map antibody-binding epitopes within E2 by flow cytometry analysis and sorting of loss-of-binding mutants. To reach this stage it was necessary firstly to demonstrate expression and correct fold of E2 on the surface of the display library host. A suitable random mutagenesis method was

chosen to generate the library. The randomly mutated E2 library was then used to screen a control MAb against known loss-of-binding mutants to demonstrate proof-of-principle that antibody-binding residues could be mapped with this system. The yeast strain, *S. cerevisiae* EBY100, and the insect cell line, *D. melanogaster* S2, were both tested as the host organism for library development.

2.3 Materials and Methods

2.3.1 Construction of a yeast display library in *S. cerevisiae* EBY100

2.3.1.1 Polymerase chain reaction (PCR) amplification of E2

Primers were designed to generate three truncated versions of E2 from HCV strain H77.20 (genotype 1a) (Table 2-1); H77 $E2_{650}$ (amino acids 384-650; primers H77c384_s *Eco*RI and H77c650_as *Xho*I), H77 $E2_{661}$ (amino acids 384-661; primers H77c384_s *Eco*RI and H77c661_as *Xho*I) and H77 $E2_{715}$ (amino acids 384-715; primers H77c384_s *Eco*RI and H77c715_as *Xho*I). Primers contained either *Eco*RI or *Xho*I restriction sites for directional cloning into the vector pYD1.

E2 was amplified from plasmid pCR3.1::H77.20 with appropriate primer pairs. One microlitre of plasmid (diluted 1:1000 in Molecular Biology (MB) grade H_20 (Sigma Aldrich, Gillingham, UK) to give a concentration of approximately 0.3-0.5 ng/µl) was used as template in a total reaction volume of 25 µl containing 200 µM dNTPs (Fermentas St. Leon-Rot, Germany), 5 pM each primer and 0.25 units Phusion High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland). PCR amplification was performed in a 2-step

Table 2-1. Sequence of primers used in this chapter. Primers were synthesised by Eurofins MWG Operon, Germany.

Target	Function of primer	Primer orientation	Primer name	Primer sequence (5'-3')
E2 (H77)	PCR amplification	Sense	H77c384_s *Eco*RI	GATCGAATTCGAAACCCACGTCACCGGG
	PCR amplification	Antisense	H77c650_as *Xho*I	AGCTGAGCTCTTCGCCCCGCGTCCAGTTG
	PCR amplification	Antisense	H77c661_as *Xho*I	AGCTGAGCTCCTCGGACCTGTCCCTGTCT
	PCR amplification	Antisense	H77c715_as *Xho*I	AGCTGAGCTCCTTAATGGCCCAGGACGC
	PCR, fusion PCR, sequencing	Sense	H77c384_s *Bgl*II	GATCAGATCTGAAACCCACGTCACCGGG
	PCR amplification	Antisense	H77_715_as	CTTAATGGCCCAGGACGCGAT
	Sequencing	Internal	S2_IF	CGTATAGGCTTTGGCACTATC
	Sequencing	Internal	S2_IR	TGCCAGCAGTAGGGGCGTTCG
E2 (UKN 2B1.1)	PCR amplification, fusion PCR	Sense	2B1.1_384_s *Bgl*II	GATCAGATCTGAGACTCATACCACCGGC
	PCR amplification	Antisense	2B1.1_715_as	TTTTGTGATAGCCGGAGT
HA TM domain	PCR amplification (H77 compatible)	Sense	HA_H77_linker_s	GCGTCCTGGGCCATTAAGATTCTGGCGATCTACTCA
	PCR amplification (2B1.1 compatible)	Sense	HA_2B1.1_linker_s	ACTCCGGCTATCACAAAAATTCTGGCGATCTACTCA
	PCR amplification, Fusion PCR	Antisense	HA_TM_as *Bst*BI	AGCTTTCGAAGATGCATATTCTGCACTG
pYD1	Plasmid screening	Sense	pYD1_fwd	AGTAACGTTTGTCAGTAATTGC
	Plasmid screening	Antisense	pYD1_rev	GTCGATTTTGTTACATCTACAC
pMT	Plasmid screening	Sense	MT_fwd	CATCTCAGTGCAACTAAA
	Plasmid screening	Antisense	BGH	TAGAAGGCACAGTCGAGG
pGEM-T Easy	Plasmid screening	Sense	M13 fwd	GTAAAACGACGGCCAGT
	Plasmid screening	Antisense	M13 rev	AACAGCTATGACCATG

Restriction sites are underlined. *Eco*RI, GAATTC; *Xho*I, GAGCTC; *Bgl*II, AGATCT; *Bst*BI, TTCGAA

cycling reaction consisting of 98°C for 30s, 35 cycles of 98°C for 10s and 72°C for 30s, followed by 72°C for 60s. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2.

2.3.1.2 Agarose gel electrophoresis analysis of PCR products

Five microlitres of PCR product was mixed with 6x Loading Buffer (50% glycerol in H_20, 0.002% bromophenol blue) and loaded onto a 2% agarose (Bioline, London, UK) gel made with 1x Tris Acetate-ethylene diamine tetraacetic acid (EDTA) buffer (Sigma Aldrich, Gillingham, UK) containing 5 µl ethidium bromide solution (Fisher Scientific, Loughborough, UK). PCR products were loaded alongside 5 µl of Generuler DNA ladder mix (Fermentas, St. Leon-Rot, Germany). Samples were resolved by exposure to a potential difference of 95 V for 0.7 hours and visualised on a UV Transilluminator. Results were recorded photographically.

2.3.1.3 Purification of PCR products

PCR positive samples were purified on a Qiaquick PCR Purification Kit (Qiagen, Crawley, UK) using a vacuum manifold according to the manufacturer's instructions. Briefly, 5 volumes of Buffer PB were added to 1 volume of PCR product, mixed and loaded onto a QIAquick spin column. The column was loaded onto a QIAvac 24 Plus vacuum manifold (Qiagen, Crawley, UK). DNA was bound to the silica membrane by applying the vacuum and allowing the sample to pass through. Sample was washed in 750 µl Buffer PE under vacuum to remove any residual impurities or primers. DNA was eluted from the column in 50 µl Buffer EB by centrifuging at 17,900

x g for 1 minute. DNA concentration was quantified using a Nanodrop ND1000 Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK).

2.3.1.4 Cloning of $E2_{650}$, $E2_{661}$ and $E2_{715}$ into plasmid pYD1

Purified E2 PCR products and pYD1 (Invitrogen, Paisley, UK, a gift from Jialing Liu, UCSF, USA) were digested with 10 units *Eco*RI (New England Biolabs, Hitchin, UK) and 10 units *Xho*I (New England Biolabs, Hitchin, UK) in a 10 µl reaction for 1 hour at 37°C. Restriction enzymes were inactivated by heating to 65°C for 20 minutes. Parental plasmid was removed from the PCR-generated E2 fragments by incubation with 10 units of *Dpn*I (New England Biolabs, Hitchin, UK) in a 10 µl reaction for 3 hours at 37°C. Enzyme was inactivated by heating to 80°C for 20 minutes.

Linearised plasmid was de-phosphorylated by incubation with 5 units of Antarctic phosphatase (New England Biolabs, Hitchin, UK) in a 10 µl reaction for 1 hour at 37°C. Enzyme was inactivated by heating to 65°C for 20 minutes.

E2 fragments were ligated into linearised pYD1 bearing complementary ends at a 3:1 molar ratio of insert to vector in a 20 µl volume reaction containing 1000-2000 cohesive end units of T4 DNA ligase (New England Biolabs, Hitchin, UK) supplemented with 2mM ATP (Fermentas, St. Leon-Rot, Germany). The sample was incubated at room temperature for 1 hour or overnight at 4°C. Ligation products were transformed into *E. coli* TOP10 cells as described in 2.3.1.7.

2.3.1.5 Cloning of $E2_{650}$, $E2_{661}$ and $E2_{715}$ into plasmid pGEM-T Easy

PCR amplification with Phusion High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland) results in blunt-ended PCR products. In order to clone the purified E2 PCR products into pGEM-T Easy (Promega, Southampton, UK), which contains thymine (T) overhangs, it was necessary to add adenosine (A) overhangs to the PCR products. Single A overhangs were added to E2 fragments by incubation with 0.1 µM dNTPs and 0.5 units HotStar Taq DNA Polymerase (Qiagen, Crawley, UK) at 94°C for 15 minutes followed by 72°C for 30 minutes. E2 fragments were *Dpn*I treated as described in 2.3.1.4. E2 fragments were ligated into pGEM-T Easy (Promega, Southampton, UK) at a 3:1 molar ratio of insert to vector in a 10 µl volume reaction containing 3 units T4 DNA ligase (Promega, Southampton, UK). The sample was incubated at room temperature for 1 hour or overnight at 4°C. Ligation products were transformed into *E. coli* TOP10 cells as described in 2.3.1.7.

2.3.1.6 Subcloning of $E2_{650}$, $E2_{661}$ and $E2_{715}$ into pYD1

E2 inserts were cut from pGEM-T Easy (Promega, Southampton, UK) by digestion with 10 units *Eco*RI (New England Biolabs, Hitchin, UK) and 10 units *Xho*I (New England Biolabs, Hitchin, UK) in a 10 µl reaction overnight at 37°C. Samples were run on a 1% low-melting-point agarose (Promega, Southampton, UK) gel as described in 2.3.1.2 with exposure to a potential difference of 40 V for 0.7 hours. Bands corresponding to the digested insert were excised from the gel and the DNA recovered by digestion with Agarase (Fermentas, St. Leon-Rot, Germany). Briefly, agarose was melted by heating

at 70°C for 10 minutes followed by incubation at 42°C for 5 minutes. Two units of agarase per 100 mg of agarose gel was added and incubated at 42°C for a further 30 minutes. Ammonium acetate was added to a concentration of 2.5 M, chilled on ice for 5 minutes and centrifuged at 9,300 x g for 10 minutes. Supernatant was transferred to a clean tube and 2 volumes of ethanol added. Samples were incubated at room temperature for 1 hour to precipitate DNA. DNA was pelleted by centrifugation at 9,300 x g for 15 minutes, supernatant removed and air-dried at room temperature.

Extracted and purified E2 fragments were cloned into pYD1 as described in 2.3.1.4. Plasmids were transformed into *E. coli* TOP10 cells as described in 2.3.1.7.

2.3.1.7 Transformation of plasmids into *E. coli* TOP10 cells

Fifty microlitres of competent *E. coli* TOP10 cells were incubated with 3 µl of ligation product or plasmid on ice for 20 minutes. Cells were transformed by heat-shocking at 42°C for 40s and 125 µl SOC media (Sigma Aldrich, Gillingham, UK) was added. Cells were incubated at 37°C for 1 hour with shaking at 250 rpm before plating onto LB (Sigma Aldrich, Gillingham, UK) agar plates containing 25 µg/ml ampicillin and incubated at 37°C overnight.

2.3.1.8 PCR screening of transformant colonies

Single colonies were picked into a 12.5 µl reaction volume containing 2.5 pM of each plasmid-specific primer, 2.5 mM dNTPs (Fermentas, St. Leon-Rot, Germany) and 0.3 units HotStar Taq DNA Polymerase (Qiagen, Crawley, UK).

Colonies were screened in a 3-step cycling reaction consisting of 95°C for 15 minutes, 40 cycles of 94°C for 30s, 50°C for 30s and 72°C for 1 minute 30s followed by 72°C for 10 minutes. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2. Positive colonies containing the desired insert were inoculated into a 3 ml LB (Sigma Aldrich, Gillingham, UK) culture containing 25 µg/ml ampicillin and incubated at 37°C overnight.

2.3.1.9 Plasmid extraction from cultures of *E. coli* TOP10 cells

Plasmid was extracted from overnight cultures using a Qiaprep Miniprep kit (Qiagen, Crawley, UK) and vacuum manifold according to the manufacturer's instructions. Briefly, bacterial cells were pelleted by centrifugation at 6800 x g for 3 minutes and cells resuspended in 250 µl Buffer P1. Cells were lysed by the addition of 250 µl Buffer P2 and inverted to mix. The lysis reaction was neutralised by the addition of 350 µl of Buffer N3 and inverted to mix. Cell debris was precipitated by centrifugation at 17,900 x g for 10 minutes and the supernatant transferred to a Qiaprep spin column. The column was loaded onto a QIAvac 24 Plus vacuum manifold (Qiagen, Crawley, UK). DNA was bound to the silica membrane by applying the vacuum and allowing the sample to pass through. Sample was washed in 500 µl Buffer PB followed by a wash step with 750 µl Buffer PE under vacuum to remove any residual impurities or primers. DNA was eluted from the column in 50 µl Buffer EB by centrifuging at 17,900 x g for 1 minute. DNA concentration was quantified using a Nanodrop ND1000 Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK). Plasmid was sequenced with plasmid-specific primers as described in 2.3.1.10.

2.3.1.10 DNA sequencing by the dideoxy chain terminator method

DNA sequencing was carried out by the dideoxy chain terminator method (Sanger *et al.*, 1977). For PCR products 200-500 bp in size, 3-10 ng of DNA was sequenced. For PCR products 500-1000 bp in size, 5-20 ng of DNA was sequenced. For double-stranded plasmid, 100-200 ng was sequenced. Template was mixed with 0.5 µl BigDye Terminators V1.1 (Applied Biosystems, Paisley, UK) containing all dNTPs and fluorescently labelled ddNTPs, 3.5 µl BetterBuffer (Microzone, Haywards Heath, UK), and 3.2 pM of sequence-specific primer in a final volume of 10 µl MB H_20 (Sigma Aldrich, Gillingham, UK). Samples were amplified in a thermal cycling reaction of 25 cycles at 96°C for 30s, 50°C for 10 seconds and 60°C for 4 minutes. The single stranded DNA sequencing products were diluted with 10 µl MB H_20 (Sigma Aldrich, Gillingham, UK) and precipitated by addition of 50 µl 100% ethanol, 2 µl sodium acetate and 2 µl 125 mM EDTA. Precipitation reactions were carried out in the dark for 1-24 hours. Samples were centrifuged at 17,900 x g for 30 minutes, washed once with 70% ethanol and centrifuged at 17,900 x g for 15 minutes to pellet the DNA. DNA pellets were air dried at 50°C for 10 minutes. Pellets were resuspended and analysed on an ABI Prism 3130 Fluorescent DNA analyser (Perkin-Elmer, Cambridge, UK). Sequencing reads were edited using Chromas 2.13 (Technelysium, Brisbane, Australia) and contiguous sequences were made from paired or grouped sequences using SeqMan II software (DNASTAR Inc., Madison, USA). Phylogenetic analyses were carried out with MEGA 4 software (Tamura *et al.*, 2007) with sequence alignment to reference sequences.

2.3.1.11 Transformation of plasmids into *S. cerevisiae* EBY100

Competent *S. cerevisiae* EBY100 cells (Invitrogen, Paisley, UK, a gift from Marlene Schweitzer, BOKU, Austria) were prepared according to the *Pichia* EasyComp kit (Invitrogen, Paisley, UK). Briefly, an overnight culture of *S. cerevisiae* EBY100 in YPD broth (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) D-glucose) was diluted to an OD_{600} of 0.3 and grown at 30°C for approximately 4 hours to reach an OD_{600} of 0.6 to 1.0. Cells were pelleted by centrifugation at 500 x g for 5 minutes and resuspended in 10 ml solution I. Cells were immediately pelleted as before and resuspended in 1 ml solution I. Aliquots of competent cells (50 µl) were transformed with 5 µl plasmid according to the *Pichia* EasyComp kit (Invitrogen, Paisley, UK). Briefly, 1 ml solution II was added to the cell-DNA mix and incubated at 30°C for 1 hour with gentle mixing every 15 minutes. Cells were transformed by heat-shock at 42°C for 10 minutes and centrifuged at 3000 x g for 5 minutes. Cells were resuspended in 1 ml solution III and centrifuged at 3000 x g for 5 minutes. Cells were resuspended in 100 µl solution III and plated onto minimal dextrose agar plates (0.67% (w/v) yeast nitrogen base (YNB), 2% (w/v) glucose) containing 0.1% (w/v) leucine and ampicillin (100 µg/ml). Plates were incubated at 30°C for 72-96 hours until colonies appeared.

2.3.2 Expression and characterisation of E2 proteins expressed in *S. cerevisiae* EBY100

2.3.2.1 Induction of protein expression in *S. cerevisiae* EBY100

A single colony of *S. cerevisiae* EBY100 containing the plasmid of interest was inoculated into YNB-CAA glucose media (0.67% (w/v) YNB, 0.5% (w/v)

casamino acids, 2% (w/v) glucose) and grown at room temperature overnight with shaking to an OD_{600} of 2.0 to 5.0. Cells were centrifuged at 5000 x g for 10 minutes and resuspended in YNB-CAA galactose media (0.67% (w/v) YNB, 0.5% (w/v) casamino acids, 2% (w/v) galactose) to an OD_{600} of 0.5 to 1.0. Culture was incubated at room temperature for 48 hours in galactose-containing media to allow sufficient protein expression.

2.3.2.2 SDS-PAGE analysis of E2 protein

An aliquot of cells equivalent to 2 OD_{600} units was removed for analysis, centrifuged at 3000 x g for 5 minutes and washed once with 1x phosphate buffered saline (PBS). Cells were resuspended in 100 µl PBS.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of E2-expressing cells was carried out on a 12% polyacrylamide gel (Laemmli, 1970). The resolving gel was prepared by mixing 2.5 ml H_2O, 3 ml 30% acrylamide:bisacrylamide (37.5:1) (Severn Biotech Ltd., Kidderminster, UK), 1.9 ml 1.5 M Tris (pH 8.8), 37.5 µl 20% (w/v) SDS, 50 µl 30% (w/v) ammonium persulphate (APS) and 11.25 µl tetramethylethylene diamine (TEMED) (Sigma Aldrich, Gillingham, UK). This was loaded onto a minislab gel apparatus (Atto Corporation, Tokyo, Japan) and allowed to set. A 5% stacking gel consisting of 1.7 ml H_2O, 415 µl 30% acrylamide:bisacrylamide (37.5:1) (Severn Biotech Ltd., Kidderminster, UK), 315 µl 1 M Tris (pH 6.8), 12.5 µl 20% (w/v) SDS, 30% (w/v) APS and 5 µl TEMED (Sigma Aldrich, Gillingham, UK) was loaded onto the resolving gel and allowed to set. Fifteen microlitres of cell or protein sample was mixed with an equal volume of 2x Laemmli reducing buffer (100 mM Tris-HCl, 20% (v/v) glycerol, 4% (w/v)

SDS, 0.2% (w/v) bromophenol blue, 2% (v/v) β-mercaptoethanol) and boiled for 5 minutes. Samples were resolved by electrophoresis at 150 V, 400 mA for approximately 90 minutes alongside Spectra Multicolour Broad Range Protein Ladder (Fermentas, St. Leon-Rot, Germany) in 1x running buffer (25 mM Tris, 250 mM Glycine, 0.1 % (w/v) SDS, pH 8.3).

2.3.2.3 Western blot analysis of SDS-PAGE separated samples

Following electrophoretic separation, samples were transferred to a Hybond ECL nitrocellulose membrane (GE Healthcare, Amersham, UK). The polyacrylamide gel was placed on top of the nitrocellulose membrane and then sandwiched between eight layers of Whatman chromatography paper Grade No. 3MM (Fisher Scientific, Loughborough, UK) soaked in Transfer Buffer (38 mM glycine, 47 mM Tris, 0.04% (w/v) SDS, 20% (v/v) methanol, pH 8.3). Proteins were transferred onto the membrane in a Transblot SD semi-dry transfer cell (Biorad, Hemel Hempstead, UK) under a current of 1 mA/cm^2 membrane for 2 hours. Successful transfer could be determined by visualisation of the molecular weight marker on the membrane. Membrane was blocked overnight in blocking buffer (5% milk powder, 1x PBS, 0.05% Tween-20), washed three times in 1x PBS containing 0.05% Tween 20 (PBS-T) and incubated with an appropriate primary antibody (diluted 1:200 in PBS-T) for 1 hour on a rocker. Washing was repeated and the membrane incubated for 1 hour with rabbit anti-mouse or rabbit anti-human IgG conjugated to horseradish peroxidase (HRP) (Dako, Ely, UK) (1:1000 in PBS-T). Membrane was washed six times and soaked in 2 ml ECL Western blotting detection reagents (GE Healthcare, Amersham, UK) for 5 minutes. Proteins present on

the membrane were visualised by chemiluminescent detection of the oxidation of luminol catalysed by HRP. Kodak Biomax light film (Sigma Aldrich, Gillingham, UK) was exposed to the membrane for 2-10 minutes and signal detected by soaking the film sequentially in Kodak GBX developer (Sigma Aldrich, Gillingham, UK), Kodak GBX fixer (Sigma Aldrich, Gillingham, UK) and water.

2.3.2.4 Flow cytometry analysis of EBY100 cells

Cells were counted on a haemocytometer and 1 $x10^6$ cells aliquoted into 12 mm x 75 mm sterile culture tubes with lids (Simport, Beloeil, Canada). Cells were washed once in 500 µl yeast wash buffer (1x PBS, 0.5% bovine serum albumin (BSA), 2 mM EDTA) by centrifuging at 350 x g for 5 minutes. Ten microlitres of primary antibody, diluted to the desired concentration in yeast wash buffer, was added to the cells and incubated on ice for 1 hour. Cells were washed twice in yeast wash buffer as before. Ten microlitres of Alexafluor 488-conjugated goat anti-mouse IgG (Invitrogen, Paisley, UK) (1:200 in yeast wash buffer) was added to the cells and incubated on ice, in the dark for 1 hour. Cells were washed twice in yeast wash buffer as before and resuspended in 500 µl fixing buffer (isoton, 0.5% formaldehyde) and stored at 4°C until analysis on a Beckman Coulter FC500 Flow Cytometer (Beckman Coulter, High Wycombe, UK) with 50,000 events captured. Results were analysed using WEASEL v3.0 for Windows (Walter and Eliza Hall Institute of Medical Research, Parkville, Australia) and presented as dot plots. Wild-type EBY100 cells were included as negative controls in all experiments and provided the baseline from which all values for percentage binding were calculated.

2.3.3 Construction of an insect cell display library in *D. melanogaster* S2 cells

2.3.3.1 PCR amplification of E2 and influenza HA TM domain

E2 comprising amino acids 384-715 ($E2_{715}$) was amplified from plasmid pCR3.1::H77.20 (H77, genotype 1a) with primer pairs H77c384_s *Bgl*II and H77_715_as and from pCR3.1::UKN2B1.1 (2B1.1, genotype 2b) with primer pairs 2B1.1_384_s *Bgl*II and 2B1.1_715_as (Table 2-1). One microlitre of plasmid was used as template in a total reaction volume of 25 µl containing 200 µM dNTPs (Fermentas, St. Leon-Rot, Germany), 5 pM each primer and 0.25 units Phusion High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland). PCR amplification was performed in a 2-step cycling reaction consisting of 98°C for 30s, 35 cycles of 98°C for 10s and 72°C for 30s, followed by 72°C for 60s. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2 and purified as described in 2.3.1.3.

A 111 nucleotide region of the transmembrane and cytoplasmic domain (amino acids 529-565) of influenza HA (H1) (HA_{TM}) was amplified from plasmid HA/A/PR/8/34 (a gift from Nicola Clementi, Universita 'Vita-Salute' San Raffaele, Milan) with primers HA_TM_as *Bst*BI and either HA_H77_linker_s or HA_2B1.1_linker_s (Table 2-1). Reaction mix was set up as described. PCR amplification was performed in a 3-step reaction consisting of 98°C for 30s, 35 cycles of 98°C for 10s, 70°C for 30s and 72°C for 30s, followed by 72°C for 5 minutes. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2 and purified as described in 2.3.1.3.

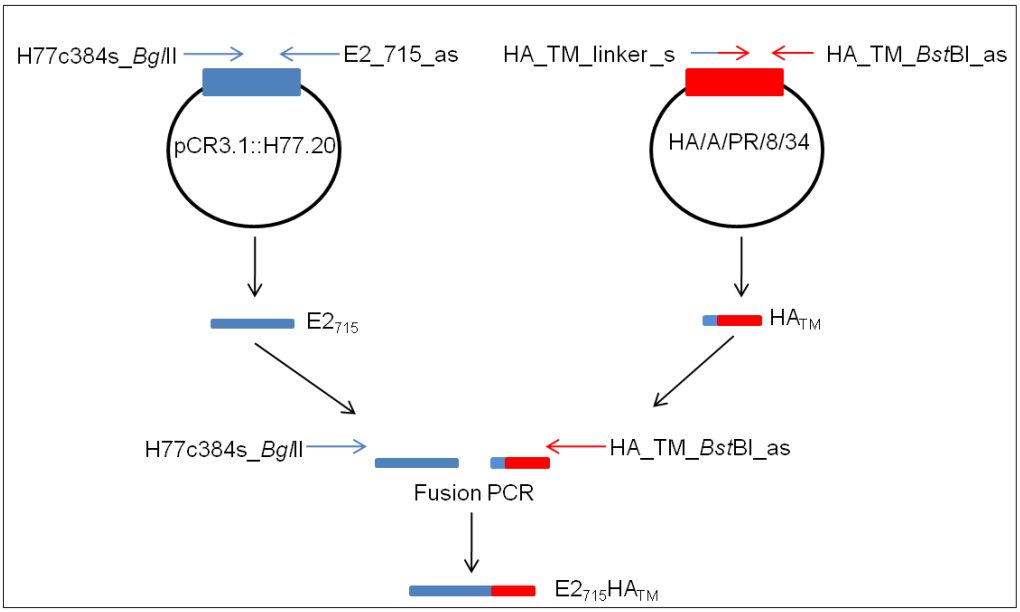

Figure 2-3. Schematic representation of the Fusion PCR cloning strategy. $E2_{715}$ (blue) was PCR amplified from pCR3.1::H77.20. The influenza haemagglutinin transmembrane (HA_{TM}) tag (red) was amplified from HA/A/PR/8/34. The forward primer, HA_TM_linker_s, contains a sequence complementary to the 3' end of $E2_{715}$ (shown in blue) which forms an E2 tag within the HA_{TM} product. This enables the fusion of the HA_{TM} tag to $E2_{715}$ producing the complete $E2_{715}HA_{TM}$ product.

2.3.3.2 Fusion PCR to generate $E2_{715}HA_{TM}$

PCR products generated in 2.3.3.1 were combined in a Fusion PCR reaction (Figure 2-3). H77 $E2_{715}$ and HA_{TM} with an H77-compatible end were combined using primers H77c384_s *Bgl*II and HA_TM_as *Bst*BI. UKN2B1.1 $E2_{715}$ and HA_{TM} with an UKN2B1.1-compatible end were combined using primers 2B1.1_384_s *Bgl*II and HA_TM_as *Bst*BI (Table 2-1). An equal molar concentration of $E2_{715}$ and HA_{TM} PCR products were used as template in a total reaction volume of 50 µl containing 200 µM dNTPs (Fermentas, St. Leon-Rot, Germany), 10 pM each primer and 0.5 units Phusion High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland). PCR amplification was performed in a 3-step reaction consisting of 98°C for 30s, 35 cycles of 98°C for 10s, 70°C for 15s and 72°C for 30s, followed by 72°C for 5 minutes. PCR

products were detected by agarose gel electrophoresis as described in 2.3.1.2 and purified as described in 2.3.1.3.

2.3.3.3 Preparation of the *Drosophila* expression plasmid, pMT

The plasmid pT425, derived from plasmid pMT/BiP/V5-His (Invitrogen, Paisley, UK) (Figure 2-2), was kindly provided by Thomas Krey (Pasteur Institute, Paris). The original plasmid insert was removed by restriction enzyme digestion of 4 µg plasmid with 10 units *Bgl*II (New England Biolabs, Hitchin, UK) and 20 units *Bst*BI (New England Biolabs, Hitchin, UK) in a 30 µl reaction volume containing NEB buffer 2 at 37°C overnight. Digested plasmid was visualised on a 1% agarose gel prepared as described in 2.3.1.2. The band corresponding to the plasmid backbone without the insert was excised and purified using a Qiaquick Gel Extraction kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. Briefly, the excised gel fragment was weighed, mixed with 3 volumes of buffer QG and heated to 50°C for 10 minutes to allow the agarose to dissolve. The solution was loaded onto a Qiaquick column preloaded onto a QIAvac 24 Plus vacuum manifold (Qiagen, Crawley, UK) and DNA bound to the silica membrane by applying the vacuum and allowing the sample to pass through. Sample was washed with 500 µl buffer QG to remove residual agarose followed by 750 µl buffer PE under vacuum to remove any residual impurities. DNA was eluted from the column in 50 µl Buffer EB by centrifuging at 17,900 x g for 1 minute. DNA concentration of the resulting plasmid, pMT, was quantified using a Nanodrop ND1000 Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK).

2.3.3.4 Cloning of $E2_{715}HA_{TM}$ constructs into plasmid pMT

Approximately 1 µg of the constructs H77 $E2_{715}HA_{TM}$ and UKN2B1.1 $E2_{715}HA_{TM}$, generated in 2.3.3.2, were digested with 10 units *Bgl*II (New England Biolabs) and 20 units *Bst*BI (New England Biolabs, Hitchin, UK) in a 20 µl reaction volume containing NEB buffer 2 at 37°C overnight. Digested products were column purified using a Qiaquick PCR Purification kit (Qiagen, Crawley, UK) as described in 2.3.1.3. $E2_{715}HA_{TM}$ fragments were ligated into pMT containing compatible ends at a 3:1 molar ratio of insert to vector in a 20 µl reaction volume containing 400 cohesive end units T4 DNA ligase (New England Biolabs, Hitchin, UK). The reaction was incubated at 16°C for 1 hour and plasmids transformed into *E. coli* TOP10 competent cells as described in 2.3.1.7. Transformant colonies were screened with pMT-specific primers (Table 2-1) as described in 2.3.1.8. Putative positive colonies were picked and grown, plasmid extracted as described in 2.3.1.9 and sequenced with pMT-specific primers as described in 2.3.1.10 to confirm that $E2_{715}HA_{TM}$ was inserted into the plasmid and did not contain any mutations. The resulting plasmids were designated pMT::$E2_{715}HA_{TM}$ (H77) and pMT::$E2_{715}HA_{TM}$ (UKN2B1.1)

2.3.3.5 Site-directed mutagenesis

Site-directed mutagenesis was carried out on plasmid pMT::$E2_{715}HA_{TM}$ (H77) to reverse an asparagine residue at position 391 back to the wild-type serine residue (N391S) and to generate alanine substitutions at amino acids 530 (G530A) and 535 (D535A). These residues are known CD81 binding sites and

Table 2-2. Sequences of the site-directed mutagenesis primers.

Target	Orientation	Primer name	Primer sequence (5'-3')
G530A	Sense	G530A_s	CTACCTACAGCTGGGCTG CAAATGATACGG
G530A	Antisense	G530A_as	CCGTATCATTTGCAGCCCA GCTGTAGGTAG
N391S	Sense	N391S_s	GTCACCGGGGGAAGTGCC GGCCGCAC
N391S	Antisense	N391S_as	GTGCGGCCGGCACTTCCC CCGGTGAC
D535A	Sense	D535A_s	GGTGCAAATGATACGGCT GTCTTCGTCCTTAAC
D535A	Antisense	D535A_as	GTTAAGGACGAAGACAGC CGTATCATTTGCACC

All primers were synthesised commercially by Eurofins MWG Operon, Germany.

form part of the epitope recognised by the antibodies AR3A, A:8 and 1:7 (Johansson *et al.*, 2007b; Law *et al.*, 2008). Mutagenesis primers were designed using PrimerX (http://bioinformatics.org/primerx/cgi-bin/DNA_1.cgi) (Table 2-2). Mutagenesis was carried out on 50 ng of plasmid in a total reaction volume of 50 µl containing 125 ng each plasmid, 200 µM dNTPs (Fermentas, St. Leon-Rot, Germany), 3% dimethyl sulphoxide (DMSO) and 0.5 units Phusion High-Fidelity DNA Polymerase (Finnzymes, Vantaa, Finland). Thermal cycling consisting of 98°C for 2 minutes and 30 cycles of 98°C for 10s, 55°C for 15s and 72°C for 5 minutes was performed. Parental plasmid was removed by overnight digestion with 20 units *Dpn*I (New England Biolabs, Hitchin, UK) at 37°C. Plasmid was transformed into *E. coli* TOP10 competent cells as described in 2.3.1.7 and colonies screened with pMT-specific primers as described in 2.3.1.8. Putative positive colonies were grown and the plasmid extracted as described in 2.3.1.9 and sequenced with pMT-

specific primers as described in 2.3.1.10 to confirm the presence of the desired mutations.

2.3.3.6 Random mutagenesis of pMT::E2$_{715}$HA$_{TM}$

Random mutations within the E2$_{715}$HA$_{TM}$ region of pMT::E2$_{715}$HA$_{TM}$ (H77) were generated using the Genemorph II EZClone Domain Mutagenesis Kit (Agilent Technologies, Stockport, UK) according to the manufacturer's instructions. In the first step, a megaprimer was generated from pMT::E2$_{715}$HA$_{TM}$ (H77). To obtain a medium mutation rate, 300 ng of pMT::E2$_{715}$HA$_{TM}$ (H77) plasmid was used as template in a total reaction volume of 50 µl containing 125 ng/µl insert specific primers (Table 2-1), 200 µM dNTPs and 2.5 units Mutazyme II DNA Polymerase (Agilent Technologies, Stockport, UK). Amplification was carried out with cycling at 95°C for 5 minutes, 30 cycles of 95°C for 30s, 58°C for 30s, and 72°C for 1 minute, followed by 72°C for 10 minutes. The Megaprime reaction product was purified using a Qiaquick PCR Purification Kit (Qiagen, Crawley, UK) as described in 2.3.1.3. The megaprimer was subsequently used to prime amplification of pMT::E2$_{715}$HA$_{TM}$ (H77) in the EZClone reaction. Two hundred and fifty nanograms of megaprimer was used to prime amplification from 50 ng of the template plasmid in a total reaction volume of 50 µl containing 3 µl EZClone solution and 25 µl EZClone enzyme mix (Agilent Technologies, Stockport, UK). Amplification was carried out with cycling at 95°C for 1 minute, 25 cycles of 95°C for 50s, 60°C for 50s and 68°C for 9 minutes and 30s. Parental plasmid was removed by digestion with 10 units

*Dpn*I (Agilent Technologies, Stockport, UK) for 2 hours at 37°C prior to transformation into XL10-Gold Ultracompetent cells.

2.3.3.7 Transformation of plasmids into XL10-Gold Ultracompetent cells

Plasmid was transformed into XL10-Gold Ultracompetent cells (Agilent Technologies, Stockport, UK) according to the manufacturer's instructions. Briefly, 2 µl β-mercaptoethanol mix (Agilent Technologies, Stockport, UK) was added to 45 µl XL10-Gold cells and incubated on ice for 10 minutes with gentle swirling every 2 minutes. To this, 1.5 µl plasmid was added and incubated on ice for a further 10 minutes. Cells were heat-shocked at 42°C for 30s and incubated on ice for 2 minutes. Five hundred microlitres of NZY^+ broth (1% (w/v) casein hydrolysate, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 12.5 mM $MgCl_2$, 12.5 mM $MgSO_4$, 20 mM glucose) preheated to 42°C was added and cells incubated at 37°C for 1 hour with shaking at 250 rpm before plating onto LB (Sigma Aldrich, Gillingham, UK) agar plates containing 25 µg/ml ampicillin and incubated at 37°C overnight. Colonies were screened with pMT-specific primers (Table 2-1) as described in 2.3.1.8. Putative positive colonies were picked and grown, plasmid extracted as described in 2.3.1.9 and sequenced with pMT-specific primers as described in 2.3.1.10 to assess mutation frequency. The Programme for Estimating Diversity in Error-prone PCR libraries (PEDEL-AA) (Patrick *et al.*, 2003) (accessed at http://guinevere.otago.ac.nz/aef/STATS/index.html) was used to determine the size of the library required. The resulting plasmid pool was designated pMT::H77 library.

2.3.3.8 Stable transfection of *D. melanogaster* S2 cells

D. melanogaster S2 cells were stably transfected by co-transfection of the pMT plasmid, containing the gene of interest, and the plasmid pCoBlast (Invitrogen, Paisley, UK), containing the blasticidin resistance gene, at a 20:1 ratio of expression plasmid to selection plasmid. Cells were seeded at a density of 5 x 10^6 cells/well in a 6-well tissue culture plate for suspension cells (Sarstedt, Leicester, UK) in a total volume of 2.5 ml Insect Xpress media (Lonza, Slough, UK). After 24 hours growth at 28°C, 2 µg pMT and 0.1 µg pCoBlast were mixed with 6 µl FuGene 6 transfection reagent (Roche, Burgess Hill, UK), incubated at room temperature for 20 minutes and added dropwise to the cells. Cells were incubated at 28°C for a further 48 hours and stable transfectants selected by the addition of Blasticidin S (25 µg/ml) (Invivogen, Nottingham, UK). Every 2 to 5 days media was removed by centrifuging cells at 100 x g for 5 minutes and replaced with fresh Insect Xpress media (Lonza, Slough, UK) containing 25 µg/ml Blasticidin S (Invivogen, Nottingham, UK) to maintain selection pressure. Clumps of cells became visible after 20 to 25 days of selection indicating that stable transfection was successful. These cultures were expanded into larger volumes and maintained in T25 or T75 tissue culture flasks for suspension cells (Sarstedt, Leicester, UK). Cells were grown in Insect Xpress media (Lonza, Slough, UK) containing 25 µg/ml Blasticidin S (Invivogen, Nottingham, UK) to maintain stable transfection of the gene of interest.

2.3.4 Expression and characterisation of E2 proteins expressed in *D. melanogaster* S2 cells

2.3.4.1 Induction of protein expression in *D. melanogaster* S2 cells

Cells were seeded at a density of 2 x 10^6 cells/ml in a 6-well tissue culture plate for suspension cells (Sarstedt, Leicester, UK) containing 3 ml Insect Xpress media (Lonza, Slough, UK). After 24 hours growth at 28°C protein expression was induced by addition of 700 µM copper sulphate ($CuSO_4$). Cells were grown for a further 5 days at 28°C before protein was assayed.

2.3.4.2 Preparation of *D. melanogaster* S2 cells for SDS-PAGE

Cells were counted and aliquots containing 1 x 10^6 cells washed once in PBS by centrifuging at 100 x g for 5 minutes. For analysis of whole cells, pellets were resuspended in 200 µl PBS. For analysis of cell lysates, pellets were resuspended in 200 µl lysis buffer (150 mM NaCl, 1% Igepal CA-630 (Sigma Aldrich, Gillingham, UK), 50 mM Tris pH 7.4, 1 mM EDTA, 37 ng/ml iodoacetamide), incubated at 37°C for 10 minutes, vortexed briefly and centrifuged as above to pellet cell debris. SDS-PAGE was carried out as described in 2.3.2.2 followed by Western blot analysis of protein samples as described in 2.3.2.3.

2.3.4.3 *Galanthus nivalis* agglutinin (GNA)-capture ELISA analysis of protein expressed in *D. melanogaster* S2 cells

A Maxisorp 96-well plate (Nunc) (Fisher Scientific, Loughborough, UK) was coated with 50µl of GNA (5 µg/ml in 50 mM carbonate/bicarbonate buffer (pH 9.8)) overnight at 4°C in a humidity chamber. Non-specific binding sites were

blocked with 300 µl blocking buffer (5% milk powder, 1x PBS, 0.05% Tween-20) for 2 hours at room temperature. Wells were washed once in PBS-T. *D. melanogaster* S2 cell lysates (prepared as described in 2.3.4.2) were diluted 1:5 in PBS and 50 µl coated onto wells for 2 hours at room temperature alongside the positive control ($sE2_{650}$ from a genotype 1a (H77) isolate expressed in HEK 293T cells). Wells were washed three times in PBS-T. Bound cell lysates were detected by the addition of 50 µl of primary antibody (5 µg/ml diluted in PBS-T) and incubated at room temperature for 1 hour. Wells were washed three times in PBS-T. Bound primary antibody was detected by the addition of 50 µl of alkaline phosphatase (AP)-conjugated secondary antibody (goat anti-mouse or goat anti-human IgG) (Sigma Aldrich, Gillingham, UK) (1:1000 in PBS) and incubated at room temperature for 1 hour. Wells were washed three times in PBS-T and bound antibody identified by the addition of 100 µl *p*-Nitrophenyl Phosphate (pNPP) (Fisher Scientific, Loughborough, UK). Plates were incubated in the dark for approximately 30 minutes to allow a yellow colour to develop indicating a positive reaction. Absorbance at 405 nm was determined using a Fluorostar Optima plate reader (BMG Labtech, Aylesbury, UK). All samples were run in triplicate and the results analysed using Graphpad Prism 4.03 for Windows (Graphpad Software, San Diego, USA). Error bars were calculated as the standard deviation of the replicates and the results presented as percentage of positive control. Significance of differences was determined using a one-way ANOVA with Tukey's multiple comparison test. Values of $P < 0.05$ were deemed significant.

2.3.4.4 Flow cytometry analysis of *D. melanogaster* S2 cells – cell surface staining

Cells were counted and resuspended at 1 x 10^6 cells/ml in Insect Xpress media (Lonza, Slough, UK) in 12 mm x 75 mm sterile culture tubes with lids (Simport, Beloeil, Canada). Cells were washed once in 500 µl media by centrifuging at 300 x g for 5 minutes.

Cells were stained with a single MAb. Cells were incubated with 10 µl of primary antibody, diluted to the desired concentration in media, on ice for 1 hour. Cells were washed twice in media as before. Bound MAb was detected by the addition of 10 µl of an Alexafluor 488-conjugated secondary antibody (goat anti-mouse or goat anti-human IgG) (Invitrogen, Paisley, UK) (1:200 in Insect Xpress media (Lonza, Slough, UK)) and incubated on ice, in the dark for 1 hour. Cells were washed twice in media as before and resuspended in 500 µl media. For flow cytometry analysis without sorting, samples were run immediately on a Beckman Coulter FC500 Flow Cytometer (Beckman Coulter, High Wycombe, UK) with 50,000 events captured. Results were analysed using WEASEL v3.0 for Windows (Walter and Eliza Hall Institute of Medical Research, Parkville, Australia) and presented as histograms or dot plots. Wild-type S2 cells were included as negative controls in all experiments and provided the baseline from which all values for percentage binding were calculated.

2.3.4.5 Flow cytometry analysis of *D. melanogaster* S2 cells – intracellular staining

Cells were counted and resuspended at 1 x 10^6 cells/ml in Insect Xpress media (Lonza, Slough, UK) in 12 mm x 75 mm sterile culture tubes with lids (Simport, Beloeil, Canada). Cells were washed once in 500 µl media by centrifuging at 300 x g for 5 minutes. Cells were fixed by incubating in 1 ml Insect Xpress media containing 4% formaldehyde at room temperature for 5 minutes. Cells were washed once in 1 ml of media as before. Cells were washed twice in 1 ml permeabilisation buffer (Insect Xpress media, 0.4 mg/ml saponin, 50 mM glucose) by centrifuging at 300 x g for 5 minutes. Cells were stained with 10 µl of primary antibody, diluted to the desired concentration in media, and incubated on ice for 1 hour. Cells were washed twice in permeabilisation buffer as before. Bound MAb was detected by the addition of 10 µl of an Alexafluor 488-conjugated secondary antibody (goat anti-mouse or goat anti-human IgG) (Invitrogen, Paisley, UK) (1:200 in Insect Xpress media (Lonza, Slough, UK)) and incubated on ice, in the dark for 1 hour. Cells were washed once in permeabilisation buffer as before, resuspended in 500 µl fixing solution (Insect Xpress media, 0.5% formaldehyde) and stored at 4°C until analysis. Samples were run on a Beckman Coulter FC500 Flow Cytometer (Beckman Coulter, High Wycombe, UK) and analysed as described in section 2.3.4.4.

2.3.4.6 Fluorescence-activated cell sorting of *D. melanogaster* S2 cells

For sorting, *D. melanogaster* S2 cells were stained with a pool of MAbs that had been directly labelled with either the Alexafluor-488 fluorophore or the

Alexafluor-647 fluorophore using a Monoclonal Antibody Labelling Kit (Invitrogen, Paisley, UK). Cells were simultaneously stained with a single Alexafluor 488-conjugated MAb, AR3A, (AR3A-A488) and a pool of Alexafluor 647-conjugated MAbs, AR2A (AR2A-A647) and H53 (H53-A647) (10 µl total volume). Antibodies were used at a concentration of 10 µg/ml each, diluted in Insect Xpress media (Lonza, Slough, UK).

Cells were counted and resuspended at 1 x 10^6 cells/ml in Insect Xpress media (Lonza, Slough, UK) containing Penicillin-Streptomycin (1:100) (Invitrogen, Paisley, UK) in 12 mm x 75 mm sterile culture tubes with lids (Simport, Beloeil, Canada). Cells were washed once in 500 µl media containing Penicillin-Streptomycin by centrifuging at 300 x g for 5 minutes.

Cells were mixed with the pooled MAb mixture and incubated on ice, in the dark for 2 hours. Cells were washed twice as before and resuspended in 500 µl media containing Penicillin-Streptomycin. Samples were analysed and sorted immediately on a Beckman Coulter MoFlo Cell sorter (Beckman Coulter, High Wycombe, UK). All cells within the sample were analysed. Sorting data was analysed as described in section 2.3.4.4.

2.3.4.6.1 Bulk sorting and expansion of *D. melanogaster* S2 cells

According to the manufacturer's protocol *D. melanogaster* S2 cells do not grow well when seeded at a density below 5 x 10^5 cells/ml (or 500 cells/µl). It is also reported that the cells grow better when passaged into conditioned media. Therefore cells were sorted in bulk into a 96-well flat-bottom plate for suspension cells (Sarstedt, Leicester, UK) containing 100 µl of mixed fresh and conditioned Insect Xpress media (Lonza, Slough, UK) containing Penicillin-

Streptomycin (1:100) (Invitrogen, Paisley, UK). Due to the low density of cells sorted, wild-type *D. melanogaster* S2 cells were seeded at a density of 500 cells/µl into each well prior to sorting. This provided a sufficient density of cells onto which stained cells could be sorted, allowing the sorted population to survive and grow. All cells within a tube were run through the Beckman Coulter MoFlo Cell sorter and as many cells falling within the population of interest were sorted into a single well. Cells were incubated at 28°C and the media replenished every other day to maintain a constant volume. When the population reached a sufficient density the cells were expanded into larger growth vessels (48-well, 24-well, 6-well tissue culture plates (Sarstedt, Leicester, UK)). After approximately 3 weeks the cultures had been expanded into a 6-well tissue culture plate. Penicillin-Streptomycin was removed from the growth media at this stage. Blasticidin S (25 µg/ml) (Invivogen, Nottingham, UK) was re-introduced to the growth media to remove the wild-type cells and select only for the sorted cells of interest. After a further 2 weeks selection, the cells were re-sorted as described in section 2.3.4.6.

2.3.4.6.2 Sorting *D. melanogaster* S2 cells into 96-well PCR plate

For a more direct analysis of the library, single cells were sorted into a 96-well PCR plate containing 10 µl 1x Tris-buffered saline (TBS) per well.

2.3.5 PCR amplification from single *D. melanogaster* S2 cells

Following sorting into a 96-well PCR plate, samples were heated to 95°C for 15 minutes to lyse the cells. Amplification of the H77 $E2_{715}HA_{TM}$ DNA

contained within each cell was carried out in a nested PCR reaction with primers MT_fwd and BGH in the first round, and primers H77c384_s *Bgl*II and HA_TM_as *Bst*BI in the second round (Table 2-1). For the first round, 2.5 pM each primer, 200 µM dNTPs (Fermentas St. Leon-Rot, Germany) and 1 unit LongAmp Taq DNA Polymerase (New England Biolabs, Hitchin, UK) was added to each well to a total volume of 15 µl. Amplification was performed in a 3-step cycling reaction consisting of 94°C for 30s, 45 cycles of 94° for 30s, 50°C for 30s and 65°C for 1 minute 15s, followed by 65°C for 10 minutes. One microlitre of first round product was used as template in the second round reaction with the following cycling parameters; 94°C for 30s, 45 cycles of 94° for 30s, 64°C for 30s and 65°C for 1 minute 15s, followed by 65°C for 10 minute. Products were visualised by agarose gel electrophoresis as described in 2.3.1.2. Positive samples were column purified as described in 2.3.1.3 and sequenced with primers H77c384_s *Bgl*II, HA_TM_as *Bst*BI, S2_IF and S2_IR (Table 2-1) as described in 2.3.1.10.

2.4 Results

2.4.1 Development of a yeast display library in *S. cerevisiae* EBY100

2.4.1.1 Cloning E2 into pYD1

To generate an E2 cell-surface display library in the yeast *S. cerevisiae* EBY100, three truncated versions of the E2 glycoprotein were PCR amplified from the genotype 1a strain H77 (Figure 2-4). All constructs began at the first amino acid residue of E2 (E384) and were truncated at residue E650 ($E2_{650}$), E661 ($E2_{661}$) or K715 ($E2_{715}$). $E2_{661}$ is a soluble form of E2 which is widely

used and has been shown to express well in different culture systems (Cocquerel *et al.*, 2006; Flint *et al.*, 1999b).

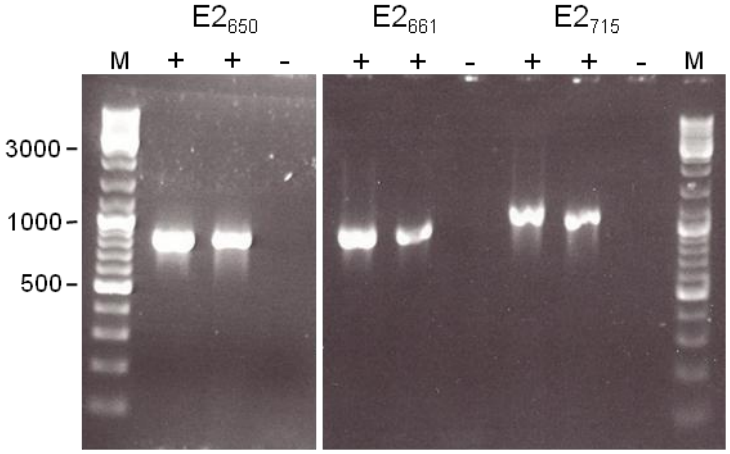

Figure 2-4. Generation of E2 with truncated ectodomains. PCR amplification products $E2_{650}$, $E2_{661}$ and $E2_{715}$. were analysed by agarose gel electrophoresis. + denotes PCR products, - denotes negative control. Position of DNA marker (M) is shown in base pairs.

However, according to the recently described model of E2 (Krey *et al.*, 2010), this truncation results in a 'spare' cysteine residue which is unable to form a disulfide bond. This may result in an unstable or incorrectly folded protein. Truncation at residue E650 removes the 'spare' cysteine and therefore should overcome any problems associated with incorrect folding. $E2_{715}$ lacks only the transmembrane domain (the C-terminal 31 amino acids) and therefore represents the entire ectodomain (Cocquerel *et al.*, 1998). This version of E2 is expressed well and folds correctly in *D. melanogaster* S2 cells (Krey *et al.*, 2010).

Attempts were made to clone E2 PCR products directly into pYD1 (Figure 2-1b) by restriction enzyme cloning. This proved unsuccessful however, despite several attempts. The decision was then made to clone E2 products into pGEM-T Easy (Promega, Southampton, UK) by TA-cloning and subclone E2

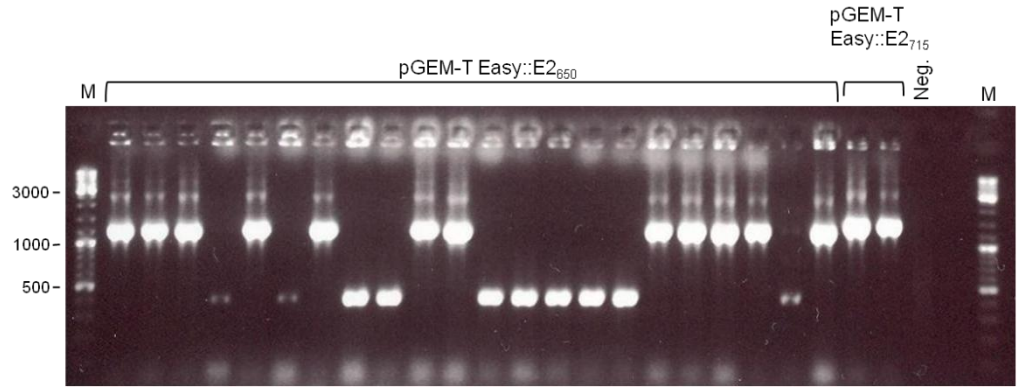

Figure 2-5. Cloning of $E2_{650}$ and $E2_{715}$ constructs into pGEM-T Easy. Colonies were screened by PCR. Bands at approximately 1200 bp represent putative positive colonies containing the desired insert. Bands at approximately 400 bp represent undigested vector. Neg. denotes negative control. Position of DNA marker (M) is shown in base pairs.

into pYD1 by restriction enzyme cloning. Colonies were obtained containing the desired fragment (Figure 2-5) and sequencing of the inserts within these plasmids showed that $E2_{715}$ had been inserted into pYD1. However, the presence of an additional *Eco*RI restriction site within pGEM-T Easy prevented $E2_{715}$ being cut correctly from pGEM-T Easy. This resulted in the insertion of a 14 bp fragment of pGEM-T Easy into pYD1 downstream of $E2_{715}$ and led to the introduction of a stop codon (Figure 2-6 a,c). Alignment of this sequence to a pYD1::E2 sequence generated *in silico* showed that three additional amino acids were present at the end of $E2_{715}$ followed by the stop codon. The $E2_{715}$ was cloned in-frame with Aga2p allowing expression of the fusion protein, however the reading frame of the plasmid was altered knocking out the V5 epitope and the poly-histidine tag (Figure 2-6d). Despite attempting several different cloning strategies, this plasmid, termed pYD1::$E2_{715(stop)}$, was the best construct generated.

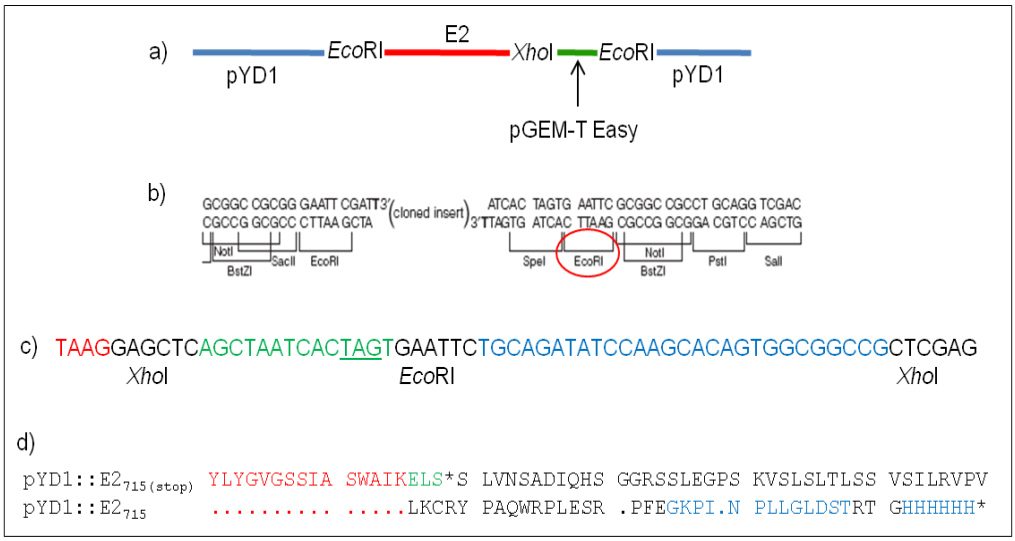

Figure 2-6. Sequence analysis of pYD1::E2 plasmids. (a) Section of pYD1 (blue) containing $E2_{715}$ (red) and a fragment of pGEM-T Easy (green). Restriction sites are shown in black. $E2_{715}$ containing *Xho*I restriction site at the 3' end was cloned into pGEM-T Easy and excised by digestion with *Eco*RI and *Xho*I. However, due to the close proximity of an *Eco*RI site to the TA cloning site (b) there is preferential cleavage of this *Eco*RI site. (c) This results in an *Eco*RI ended product which contains a 14 bp fragment of pGEM-T Easy (green) attached to $E2_{715}$ (red). This is subsequently cloned into pYD1 which has been linearised by a single cut with *Eco*RI. This results in the introduction of a stop codon (underlined) downstream of $E2_{715}$. (d) Amino acid sequence of the pYD1::$E2_{715(stop)}$ aligned to an *in silico* derived pYD1::$E2_{715}$ sequence. E2 is shown in red, the pGEM-T Easy derived sequence is shown in green. The V5 epitope and poly-histidine tag are shown in blue.

2.4.1.2 Expression of E2 on the surface of *S. cerevisiae* EBY100 cells

To test the utility of the yeast display system for expression of HCV E2, plasmid pYD1::$E2_{715(stop)}$ was transformed into *S. cerevisiae* EBY100 and expression of the protein was characterised. *S. cerevisiae* has a tendency to hyper-glycosylate proteins by the addition of extra mannose residues to the $Glc_3Man_9GlcNAc_2$ oligosaccharide (Conde *et al.*, 2004) which can result in mis-folding of the protein. To ascertain if this was the case, expression of pYD1::$E2_{715(stop)}$ was tested by Western blot analysis using an anti-Xpress primary antibody (1:2500) (Invitrogen, Paisley, UK) specific to the Xpress

epitope within pYD1 (Figure 2-1a). Protein expression after 48 hours of induction was tested alongside expression of the Aga2p protein alone (from plasmid pYD1), the human host cell factor C1, HCFC1 (from plasmid pYD1::HCFC1) (a gift from Jialing Liu, UCSF) and untransformed EBY100 cells. Expression of $E2_{715(stop)}$ in *S. cerevisiae* EBY100 cells was observed (Figure 2-7) and the protein did not appear to be hyper-glycosylated, indicated by a single, discrete band.

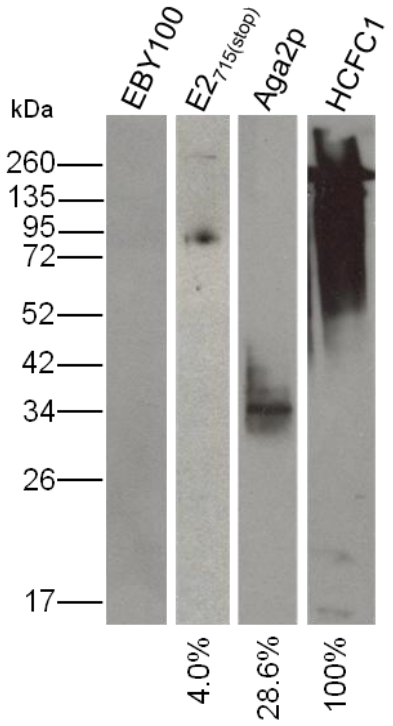

Figure 2-7. $E2_{715(stop)}$ protein expression in *S. cerevisiae* EBY100 cells. After 48 hours induction, an aliquot of whole cells equivalent to 2 OD_{600} units was analysed by Western blot under reducing conditions using Anti-Xpress antibody. Untransformed EBY100 cells served as a negative control. Cells transformed with the pYD1 plasmid alone or pYD1::HCFC1 served as positive controls for Xpress protein expression. Percentage protein expression, relative to HCFC1, is listed below. Position of molecular weight marker is shown in kDa.

The size of the protein, approximately 90 kDa, was larger than would be expected for E2, which is typically observed at 72 kDa. This was due to the presence of the Aga2p protein in the fusion product. The levels of $E2_{715(stop)}$ expression were lower than those seen for the Aga2p protein alone and HCFC1. Aga2p expression was 28.6% of HCFC1 expression, whilst $E2_{715}$ expression was only 4%. The expected size of HCFC1 is 150 kDa, however the large smear on this Western blot suggests that the protein is hyper-

glycosylated in this system. This may unduly impact the relative levels of protein expression. When Aga2p and $E2_{715(stop)}$ expression are compared, $E2_{715(stop)}$ was expression was 14% of Aga2p expression. Increasing the duration of protein induction to 72 hours did not increase expression levels (Figure 2-8) as demonstrated by Western blot analysis of His-Tag expression in pYD1-transfected *S. cerevisiae* EBY100 cells using PentaHis biotin (1:200) (Qiagen, Crawley, UK) and streptavidin-HRP (1:1000).

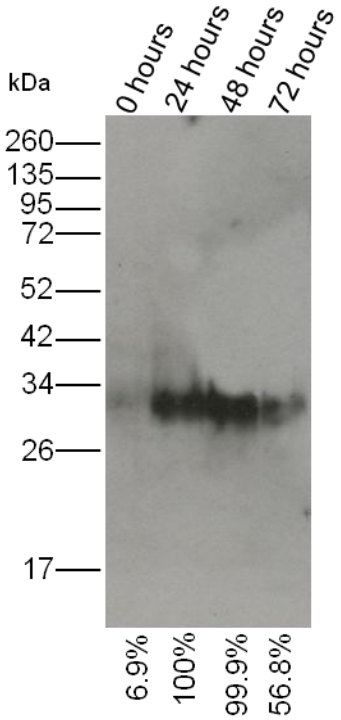

Figure 2-8. Expression of His-tagged protein from pYD1 in *S. cerevisiae* EBY100 over a 72 hour period. An aliquot of whole cells equivalent to 2 OD_{600} units was analysed at each time point by Western blot under reducing conditions with PentaHis Biotin and streptavidin-HRP. Percentage protein expression is listed below. Position of molecular weight marker shown in kDa**.**

Having confirmed that $E2_{715(stop)}$ expressed in *S. cerevisiae* EBY100 cells was not hyper-glycosylated, it was necessary to confirm that protein was expressed on the cell surface and to test if the protein was correctly folded. This was achieved by flow cytometric analysis of cell surface expressed proteins with anti-Xpress antibody (Invitrogen, Paisley, UK) (400 µg/ml) or with the E2-specific antibodies AP33 (400 µg/ml), recognising a linear epitope (Owsianka

et al., 2005), and H53 (400 µg/ml), recognising a conformation-dependent epitope (Cocquerel *et al.*, 1998). Anti-Xpress staining of cells confirmed that Aga2p and $E2_{715(stop)}$ were expressed on the cell surface. Aga2p was expressed at slightly higher levels than $E2_{715(stop)}$, (Figure 2-9) however the expression of $E2_{715}$ was better than that observed by Western blotting.

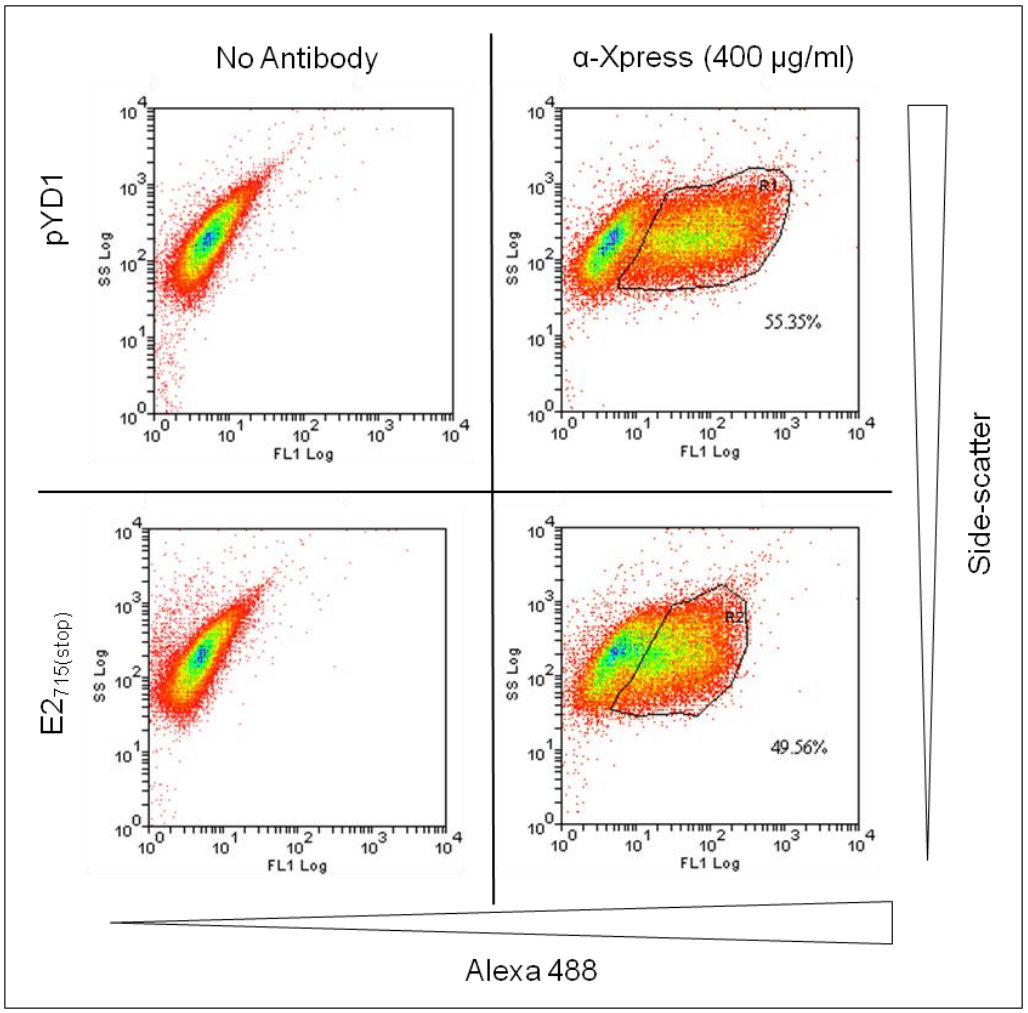

Figure 2-9. Aga2p protein expression on the surface of *S. cerevisiae* EBY100 cells. Cells transformed with pYD1 alone or pYD1::$E2_{715(stop)}$ were induced for 72 hours and protein expression in 1 x 10^6 cells measured by flow cytometry with a MAb specific to the Xpress epitope. Cells stained with α-Xpress are enclosed by a region and the percentage of cells stained is indicated on the right-hand side of the dot-plot. The no antibody control provided a baseline for specific antibody binding.

Cell staining with AP33 and H53 demonstrated that cell-surface E2 was present on cells expressing plasmid pYD1::$E2_{715(stop)}$ (Figure 2-10a). This was specific to $E2_{715(stop)}$-expressing cells as only a small amount of background staining of cells expressing plasmid pYD1 could be seen. Similarly, untransformed EBY100 cells did not bind AP33 or H53. The level of H53 binding to $E2_{715(stop)}$-expressing cells was similar to the level of AP33 binding (16.61% vs. 22.74%) (Figure 2-10a), suggesting that the $E2_{715(stop)}$ construct was correctly folded on the cell surface. However, levels of cell-surface staining with E2-specific MAbs were much lower than staining with anti-Xpress (approx. 20% (Figure 2-10a) vs. approx. 50%. (Figure 2-9). The fluorescence intensity values for surface expressed E2 are also low (Figure 2-10b) with values of 12.7 and 9.8 for AP33 binding and H53 binding respectively. This is compared to values of 7.1 and 8.1 respectively for the pYD1-expressing control. This suggests that the $E2_{715(stop)}$ protein specifically is poorly expressed on the surface of *S. cerevisiae* EBY100 cells. This may be due to the presence of the additional pGEM-T Easy fragment and the stop codon downstream of E2 (Figure 2-6). It is difficult to predict the impact that this might have on protein expression and fold but it would appear to be detrimental. In the absence of a suitable E2 construct expressed in *S. cerevisiae* EBY100 cells and the difficulties encountered with cloning, the decision was made to develop an alternative display library, using the *D. melanogaster* S2 system.

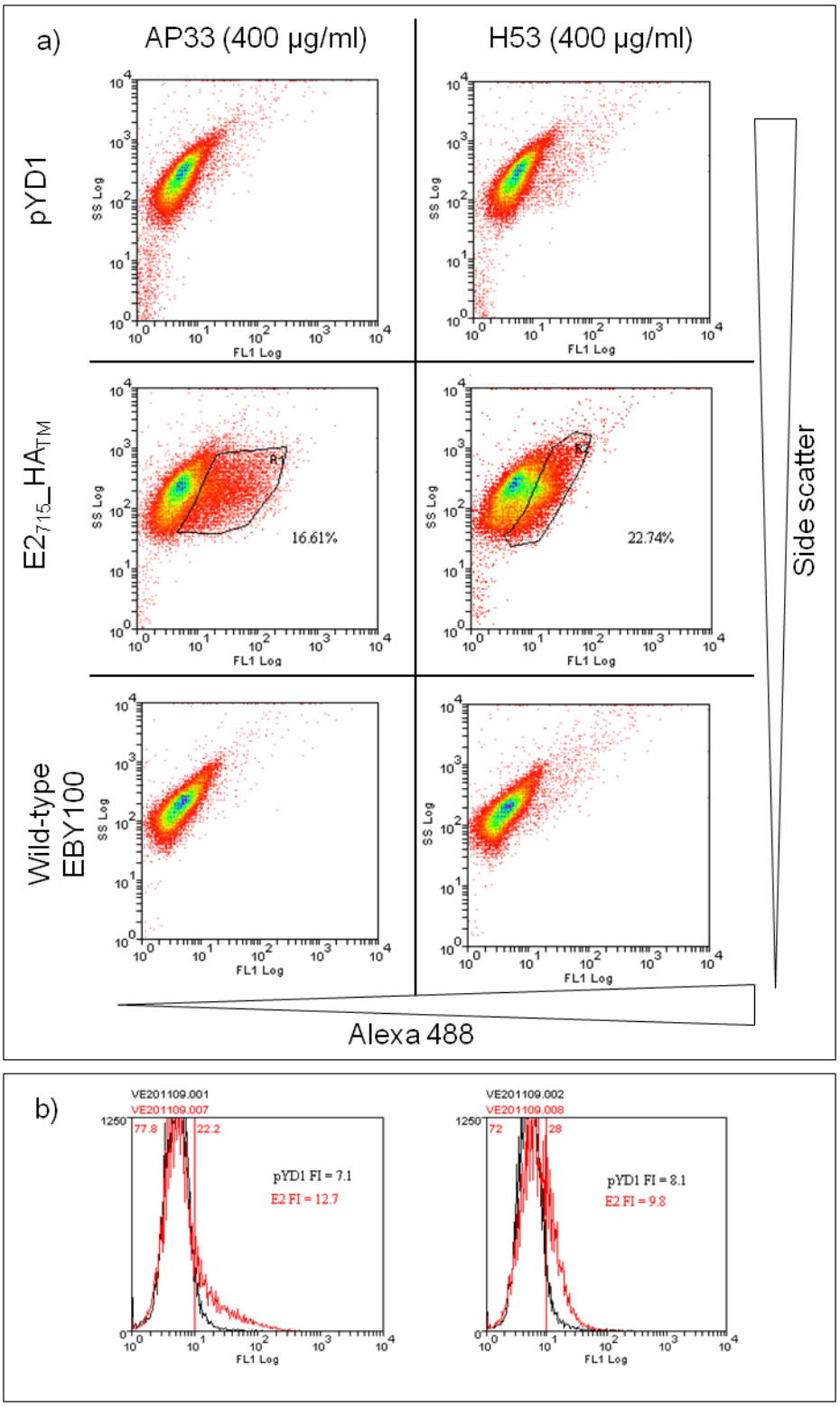

Figure 2-10. $E2_{715(stop)}$ protein expression on the surface of *S. cerevisiae* EBY100 cells. See overpage for full figure legend.

Figure 2-10. $E2_{715(stop)}$ protein expression on the surface of *S. cerevisiae* EBY100 cells. Untransformed cells and cells transformed with pYD1 alone or pYD1::$E2_{715(stop)}$ were induced for 72 hours and protein expression in 1 x 10^6 cells measured by flow cytometry using the E2-specific MAbs AP33 (left panel) and H53 (right panel). a) Cells binding the specific MAb are enclosed within a region and the percentage of cells staining is indicated on the right. Binding was MAb-specific as mock-stained cells did not fluoresce (not shown). b) Histograms of pYD1-expressing cells (black) and $E2_{715(stop)}$-expressing cells (red) staining with AP33 and H53. Mean fluorescence intensity (FI) was calculated using Weasel v3.0 software and given in arbitrary units. Vertical line denotes antibody binding and the percentage of $E2_{715(stop)}$-expressing cells staining positive is shown to the right hand side.

2.4.2 Development of an insect cell display library in *D. melanogaster* S2 cells

2.4.2.1 Generation of HA-tagged E2 and cloning into pMT

To adapt the DES secreted protein expression system for surface display of proteins, the E2 ectodomain alone (amino acids 384-715) was PCR amplified from the genotype 1a strain H77 and the genotype 2b strain UKN2B1.1 (Figure 2-11). This resulted in the removal of the native transmembrane domain containing the ER-retention signals which would prevent trafficking of the protein to the cell surface (Cocquerel *et al.*, 1998). This truncation also removed the endogenous signal sequences from E2 allowing protein translocation to be controlled by the BiP secretion signal within the pMT plasmid. This E2 truncation has previously been successfully expressed in *D. melanogaster* S2 cells (Krey *et al.*, 2010). The influenza A virus HA transmembrane and cytoplasmic domain, HA_{TM}, was PCR amplified from the subtype H1 strain A/PR/8/34 (Figure 2-11). The purified PCR products were subsequently combined in a fusion PCR reaction (Figure 2-3) to generate the products H77 $E2_{715}HA_{TM}$ (Figure 2-11) and UKN2B1.1 $E2_{715}HA_{TM}$. The fusion PCR primers were designed to introduce the restriction sites *Bgl*II and

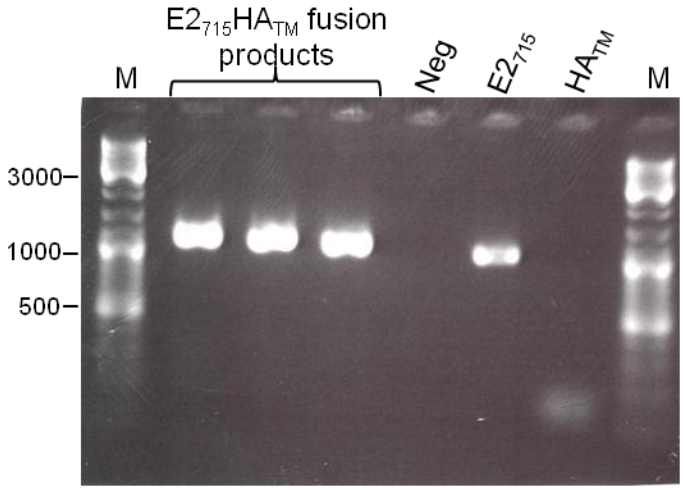

Figure 2-11. Generation of PCR products for Insect cell expression. $E2_{715}$ and HA_{TM} PCR products are shown in columns 6 and 7. Replicates for the Fusion PCR product $E2_{715}HA_{TM}$ are shown in columns 2-4. Neg. denotes negative control. Position of DNA marker (M) is shown in base pairs.

*Bst*BI at the 5' and 3' ends respectively of the PCR products. Restriction enzyme cloning was used to clone H77 $E2_{715}HA_{TM}$ and UKN2B1.1 $E2_{715}HA_{TM}$ into the plasmid pMT bearing compatible ends. The resulting plasmids were designated pMT::$E2_{715}HA_{TM}$ (H77) and pMT::$E2_{715}HA_{TM}$ (UKN2B1.1).

2.4.2.2 Generation of mutant plasmids

Sequencing of the plasmid pMT::$E2_{715}HA_{TM}$ (H77) showed that an erroneous substitution of S391N had been introduced during the PCR amplification step. Therefore site-directed mutagenesis was carried out in order to restore the wild-type sequence and successful mutagenesis was confirmed by sequencing (Figure 2-12). The plasmid pMT::$E2_{715}HA_{TM}$ (H77) was also mutated to introduce alanine substitutions at residues known to be critical in the binding of specific MAbs to E2. These residues, glycine at position 530 and aspartate at

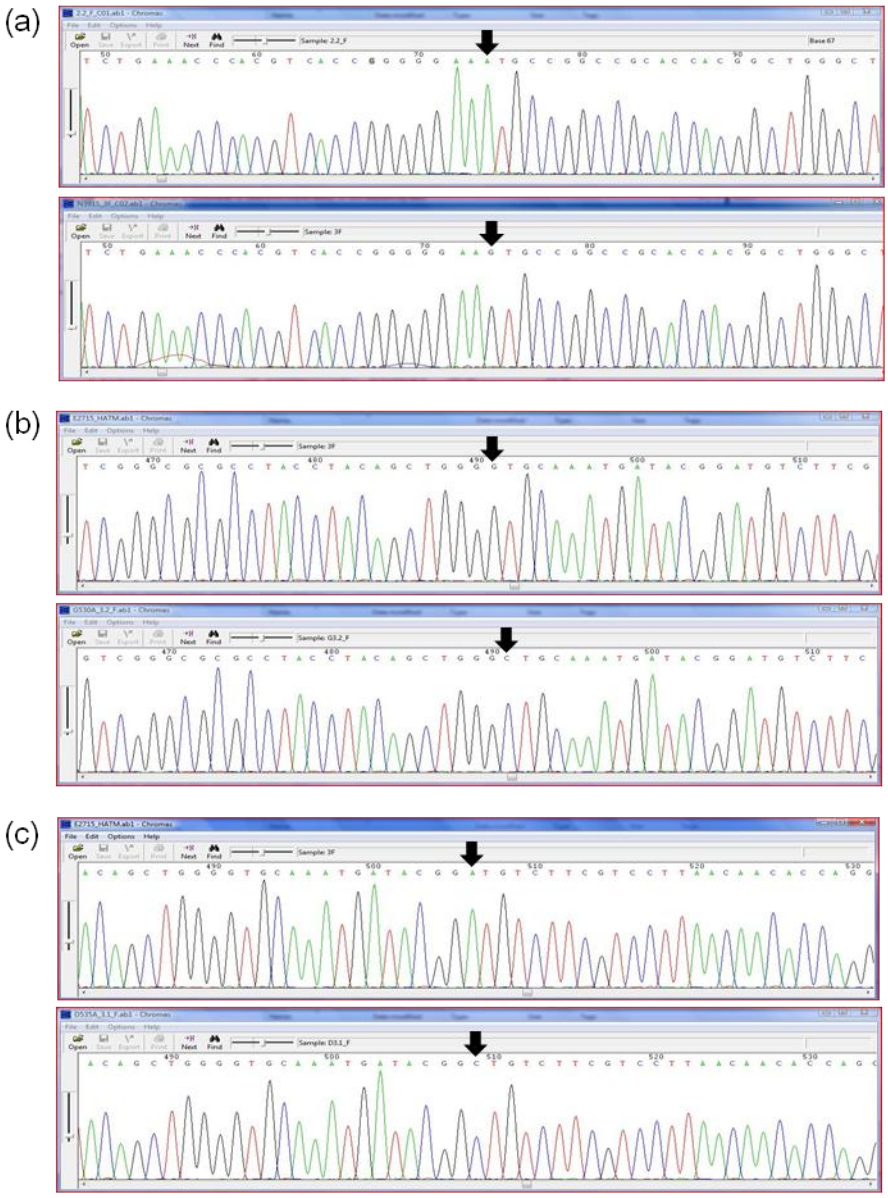

Figure 2-12. Generation of the pMT::E2$_{715}$HA$_{TM}$ mutants D535A and G530A. Sequence chromatograms show (a) replacement of asparagine (N) at residue 391 with serine (S) by mutation of AAT to AGT, restoring the wild-type E2 (H77) sequence; (b) replacement of glycine (G) by alanine (A) at residue 530 by mutation of GGT to GCT. (c) and replacement of aspartate (D) by alanine (A) at residue 535 by mutation of GAT to GCT. Nucleotide substitutions are denoted by a black arrow.

position 535, are involved in the binding of Fab e137 (Perotti *et al.*, 2008), as well as the MAbs AR3A, AR3B and AR3C (Law *et al.*, 2008), 1:7 and A8 (Allander *et al.*, 2000; Johansson *et al.*, 2007b) and CBH-5 (Hadlock *et al.*, 2000; Owsianka *et al.*, 2008). Alanine was introduced at these sites as a search of the Los Alamos HCV sequence database shows that it does not occur naturally at residue 530 or 535. It is also one of the simplest amino acids with respect to molecular structure and does not directly impact protein function. Thus its impact on overall protein fold should be minimal. Substitution of G530 and D535 with alanine was confirmed by sequencing (Figure 2-12) and the resulting plasmids were designated pMT::$E2_{715}$G530A and pMT::$E2_{715}$D535A.

The introduction of random mutations within E2 was achieved using the Genemorph II EZClone Domain Mutagenesis Kit (Agilent Technologies, Stockport, UK). This utilises an error-prone PCR method to amplify the $E2_{715}HA_{TM}$ region. This first round product was then used as a 'megaprimer' to prime amplification of the pMT::$E2_{715}HA_{TM}$ (H77) plasmid in a second reaction. The resulting plasmid should contain mutations only in the $E2_{715}HA_{TM}$ gene and should not be mutated within other functional regions of the plasmid. It may however, contain mutations within the transmembrane domain (HA_{TM}) which could prevent trafficking of the protein to the cell surface or its anchoring within the cell membrane. However the frequency of such mutations should not be so great as to significantly reduce the function of the library.

It was important to generate a library containing, on average, one amino acid substitution per sequence, without introducing deleterious mutations. It was

also important to ensure that the library generated was large enough to be representative of all possible mutations within the E2 ectodomain. The mutation rate achieved with the Genemorph II EZClone Domain Mutagenesis Kit can be altered by adjusting the concentration of template used. Test mutagenesis reactions were carried out with either 300 ng or 500 ng template DNA and several colonies were sequenced. PEDEL-AA online software was used to assess the mutation rate and size of library required (Table 2-3).

Table 2-3. Statistics for determining the size of the E2 display library

Property		
Input DNA (ng)	300	500
Total libary size	1×10^4	1×10^4
No. of variants with no in-frame deletions or stop codons	8701	9478
Mean no. of amino acid substitutions per variant	1.721	0.5267
Unmutated sequences (% of library) (Poisson estimate)	15.57	55.97
No. of distinct full-length proteins in the library (Poisson estimate)	5645	1974
No. of variants that differ from the parent by a single amino acid (Poisson estimate)*	1182	1041

Data generated with the Programme for Estimating Diversity in Error-prone PCR libraries (PEDEL-AA). *E2 is 331 amino acids

This indicated that 300 ng of input DNA was optimal for the generation of a randomly mutated library. With a library size of 1×10^4 cells, each protein should contain an average of 1.7 amino substitutions resulting in 1182 variants that differ from the parent by a single amino acid, based on a Poisson distribution. Given that $E2_{715}$ is 332 amino acids long ($E2_{715}HA_{TM}$ is 369

amino acids), this library should be representative of all single amino acid mutations within E2. PEDEL-AA analysis also indicated that the majority of sequences (8,701 out of 10,000) would not contain stop codons or in-frame deletions. Following transformation of mutated plasmids into *E. coli* XL10 cells, as described in 2.3.3.7, approximately 12,000 colonies were picked and pooled into a single LB culture. This is greater than the 1×10^4 cells required. The resulting pooled plasmid was designated pMT::H77 library.

2.4.2.3 Expression of E2 on the surface of *D. melanogaster* S2 cells

D. melanogaster S2 cells were co-transfected with a pMT plasmid containing the protein of interest and pCoBlast to generate stable cell lines. Expression of the various E2 constructs in the different cell lines was initially confirmed by Western blot analysis of cell lysates using MAb AP33 which recognises a linear epitope within E2 and does not require the residues G530 and D535 for binding (Owsianka *et al.*, 2005; Tarr *et al.*, 2006). All constructs were expressed in *D. melanogaster* S2 cells (Figure 2-13). To confirm localisation of E2 to the cell surface, flow cytometry analysis was carried out on surface stained and intracellularly stained cells expressing $E2_{715}$. Cell staining with AP33 (10 μg/ml) showed that $E2_{715}$ was expressed on the cell surface; however, intracellular staining of permeabilised cells indicated that not all protein was trafficked to the cell surface (Figure 2-14). The level of AP33 binding was only slightly lower for the intracellular staining than the cell surface staining (53.50% compared to 63.43%). This may be due to $E2_{715}$ protein which is in the process of being transported to the cell surface by the protein secretion machinery but which is momentarily retained within the ER.

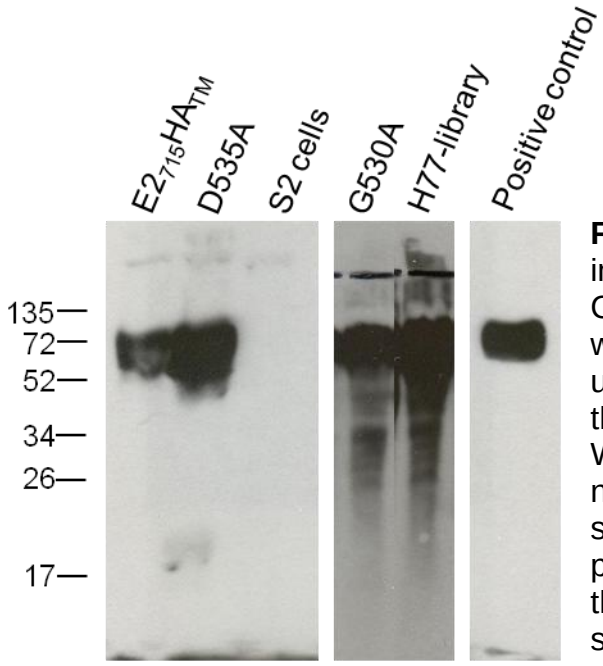

Figure 2-13. $E2_{715}$ expression in *D. melanogaster* S2 cells. Cell lysate from 1 x 10^6 cells was analysed by Western blot under reducing conditions with the E2-specific MAb AP33. Wild-type S2 cells served as a negative control. $sE2_{661}$ served as positive control for protein expression. Position of the molecular weight marker is shown in kDa.

To confirm that all proteins were expressed on the cell surface and in the correct fold or conformation, cells expressing $E2_{715}HA_{TM}$, $E2_{715}$G530A and $E2_{715}$D535A were analysed by flow cytometry with the MAbs AP33 and H53. All stable cell lines expressing E2 could be stained with AP33 and H53 showing that the cell-surface expressed E2 is correctly folded (Figure 2-15). Similar levels of AP33 and H53 binding were observed for each construct (62.41% vs. 58.26% $E2_{715}HA_{TM}$, 47.04% vs. 46.27% $E2_{715}$G530A, 44.86% vs. 38.38% $E2_{715}$D535A) demonstrating that the majority of expressed E2 was in its native conformation. (Cells staining at the top of the dot plot were not included in the analysis due to high side-scatter values indicating that these are probably dead cells). The slightly lower levels of H53 binding to $E2_{715}HA_{TM}$ and $E2_{715}$D535A compared to AP33 binding may be due to the presence of a small amount of incorrectly folded protein. Flow cytometry analysis with AP33 and H53 was repeated a number of times and expression levels of around 50-60% $E2_{715}HA_{TM}$ and around 40-45% $E2_{715}$D535A was typically seen. This may be due to the transfection of some cells with the resistance plasmid

pCoBlast only. This would result in blasticidin-resistant cells that are unable to express E2 protein.

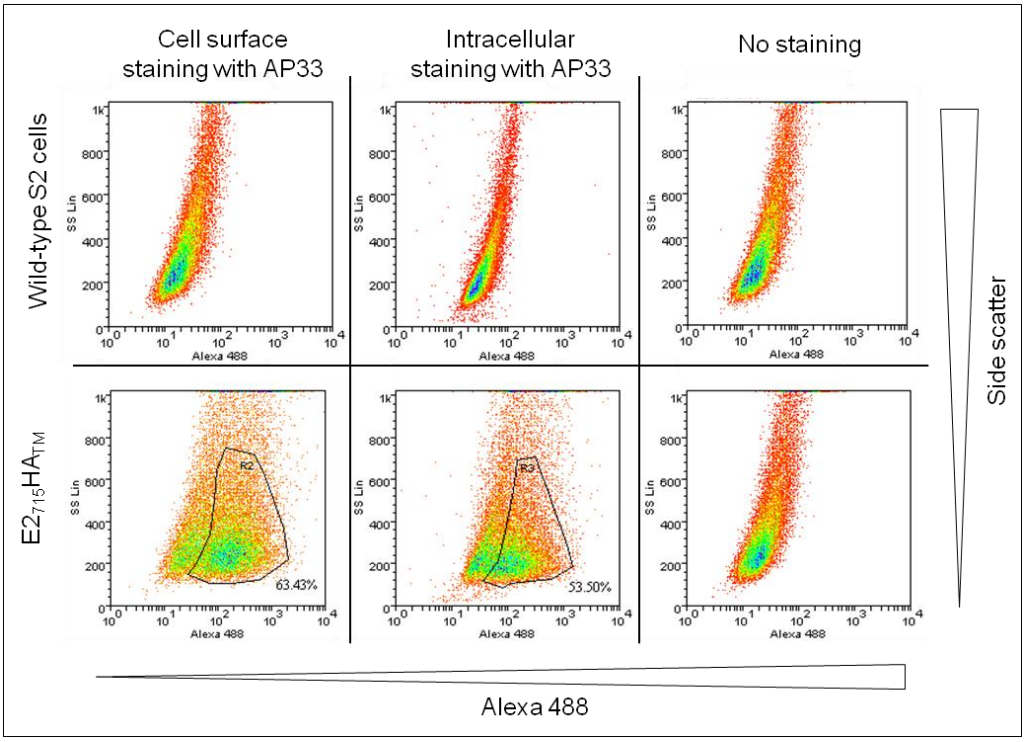

Figure 2-14. Cell surface and intracellular expression of $E2_{715}$ in *D. melanogaster* S2 cells. Cells were induced to express protein for 120 hours and protein expression in 1 x 10^6 cells measured by flow cytometry. Cells were either surface stained (left-hand column) or intracellularly stained (middle column) with the E2-specific MAb AP33 (10 µg/ml). Wild-type cells (top row) were compared to cells expressing $E2_{715}HA_{TM}$. Cells binding the specific MAb are enclosed within a region. Percentage of cells staining with the desired MAb are show in the bottom right-hand corner. A mock staining control was included (right-hand column).

2.4.2.4 Binding of specific MAbs to antibody-binding site mutants of E2

To provide proof-of-principle that the Insect cell display library could be used to map antibody-binding residues, alanine substitutions were introduced into $E2_{715}HA_{TM}$ at known antibody-binding residues (see section 2.4.2.2). Having confirmed expression and correct fold of the wild-type and mutant E2

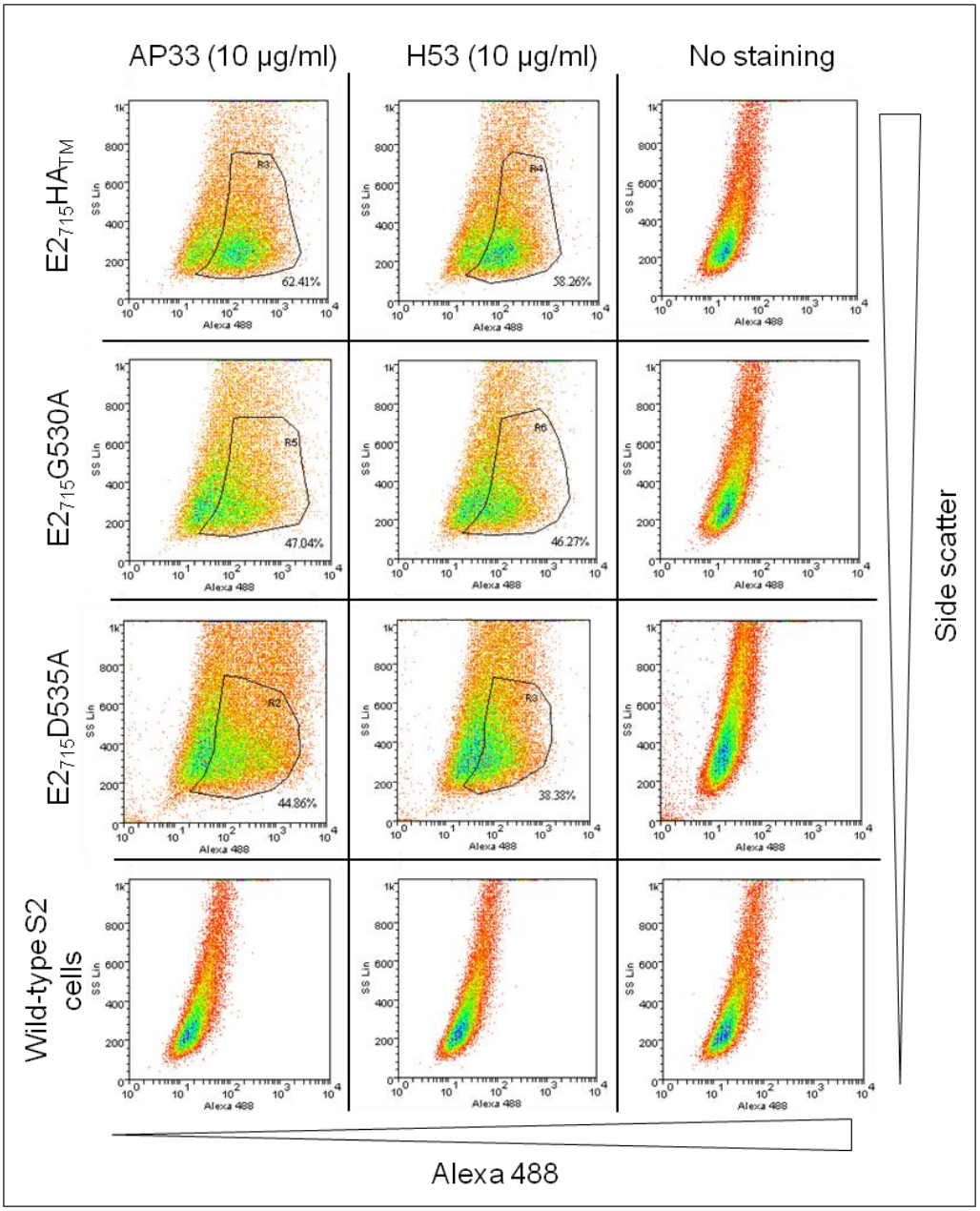

Figure 2-15. Expression of different E2 constructs on the surface of *D. melanogaster* S2 cells. Cells were induced to express protein for 120 hours and protein expression in 1 x 10^6 cells measured by staining with the linear-dependent MAb AP33 (left-hand column) or the conformation-dependent MAb H53 (middle column). Cells binding the specific MAb are enclosed within a region. Percentage of cells staining with the desired MAb are show in the bottom right-hand corner of the dot-plot. A mock staining control was included (right-hand column).

proteins, protein expression was tested with a panel of MAbs whose binding characteristics are well described. The ability of the MAbs to bind the different protein constructs was tested by ELISA (Figure 2-16) and flow cytometry (Figure 2-17). AP33, which recognises a linear epitope within E2 (Owsianka *et al.*, 2005; Tarr *et al.*, 2006), was able to bind to all three protein constructs expressed in *D. melanogaster* S2 cells (Figure 2-16 and 2-17). The increased binding of AP33 to $E2_{715}$G530A and $E2_{715}$D535A relative to $E2_{715}HA_{TM}$ (Figure 2-16) was consistently seen (and found to be significant) and may be due to an increased exposure of the AP33 epitope caused by the introduction of alanine substitutions at residues 530 and 535. Binding of H53 to all protein constructs (Figure 2-16 and 2-17) confirmed previous observations that the surface-expressed proteins are correctly folded. According to the flow cytometry data there was a slight reduction in H53 binding to $E2_{715}$D535A compared to $E2_{715}HA_{TM}$ (mean FI= 159.6 vs. 105.4)(Figure 2-17).

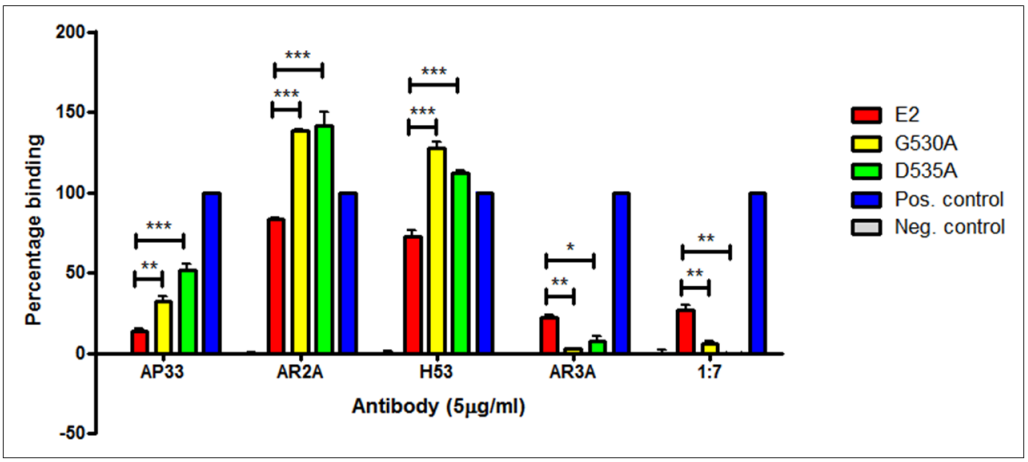

Figure 2-16. MAb reactivity to different E2 constructs expressed in *D. melanogaster* S2 cells. GNA-capture ELISA of cell lysates from 1 x 10^6 cells was performed with a panel of MAbs recognising linear and conformation-dependent epitopes. Results are given as percentage binding relative to a genotype 1a, $sE2_{650}$ protein control. Error bars were calculated from the standard deviation of three assay replicates. One way ANOVA was performed with Tukey's multiple comparison test. Significant differences, P <0.05 are shown; * P = 0.01-0.05, ** P = 0.001-0.01, *** P= <0.001.

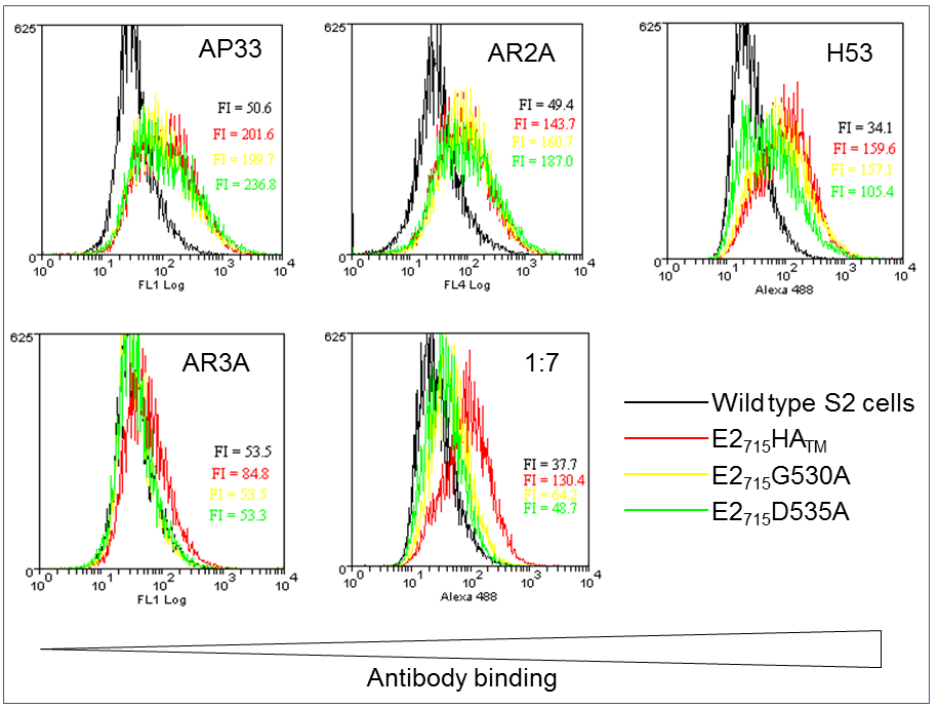

Figure 2-17. MAb binding to *D. melanogaster* S2 cells expressing different E2 constructs. Cells were induced to express protein for 120 hours and 1 x 10^6 cells stained individually with MAbs AP33, AR2A, H53, AR3A or 1:7. Each histogram shows binding of a specific MAb to wild-type S2 cells (black), $E2_{715}HA_{TM}$ (red), $E2_{715}G530A$ (yellow) and $E2_{715}D535A$ (green) expressing cells. Mean FI was calculated for each histogram using Weasel v3.0 software and is given in arbitrary units.

This may indicate a role for D535 in the epitope bound by H53 although this has not been reported previously. Alternatively, this may have been caused by the experimental set up whereby the same number of cells are counted and stained but not all cells will express protein, leading to different levels of antibody binding. AR2A recognises a discontinuous epitope within E2 that does not include the amino acid residues G530 and D535 (Law *et al.*, 2008) and should therefore bind to all protein constructs. As expected, binding of AR2A to all proteins was observed (Figure 2-16 and 2-17) providing further evidence that the proteins were correctly folded. MAbs 1:7 and AR3A recognise a conformation-sensitive epitope which includes residues G530 and D535 (Johansson *et al.*, 2007b; Law *et al.*, 2008). Previous studies have shown

that amino acid substitution at these sites severely impairs or abolishes antibody binding to E2 (Johansson *et al.*, 2007b; Law *et al.*, 2008). As predicted there was a reduction in AR3A and 1:7 binding to $E2_{715}$G530A and $E2_{715}$D535A relative to binding to $E2_{715}HA_{TM}$ (Figure 2-16 and 2-17). The ELISA data showed 22.3% AR3A binding to $E2_{715}HA_{TM}$ (relative to the positive control), reduced to 2.5% binding to $E2_{715}$G530A and 7.4% binding to $E2_{715}$D535A. Using a one-way ANOVA, both differences were significant ($P < 0.05$) (Figure 2-16). MAb 1:7 binding to $E2_{715}HA_{TM}$ was 26.8% of that observed for the positive control. This was reduced to 5.78% for $E2_{715}$G530A and was completely abolished for $E2_{715}$D535A (less than 0%). Both of these observations were found to be significant ($P < 0.05$) (Figure 2-16). The D535A mutation reduced binding of 1:7 to E2 to a greater extent than the G530A mutation. This is in agreement with the observations made by Johansson *et al.* (2007). Although 1:7 binding to $E2_{715}$D535A was completely abolished according to the ELISA data (Figure 2-16), flow cytometric analysis indicated some residual binding to the D535A mutant protein (Figure 2-17) ($E2_{715}$D535A mean FI = 48.7 vs S2 cells mean FI = 37.7). This may have been caused by the different protein preparations used, with cell-surface expressed protein maintaining some residual binding activity not seen in the cell lysates. The use of cell lysates in the ELISA may also explain the fluctuating patterns of MAb binding to different protein preparations. If different cell lysates contained different numbers of protein-expressing cells, the relative amounts of protein would have varied. Unfortunately the protein content of each preparation was not quantified; however it is important to note that the same patterns of antibody binding were observed across at least three repeat experiments using

different cell lysate preparations. According to the flow cytometry analysis of surface expressed protein, binding of AR3A to the mutant constructs $E2_{715}$G530A and $E2_{715}$D535A was abolished. The fluorescence intensity values of 53.5 and 53.3 respectively were similar to the values obtained for wild-type $E2_{715}HA_{TM}$ (Figure 2-17). However, this data does also highlight that AR3A binding to the wild-type protein was poor, with fluorescent intensity values lower than for any other MAb. This may have been due to inefficient detection of the fluorescently-labelled MAb preparation that was used and indeed similar results were observed throughout. This is discussed further in section 2.5. However, taken together this data provided the necessary proof-of-principle that cell-surface display of E2 on *D. melanogaster* S2 cells can be used to map antibody-binding residues within the glycoprotein.

2.4.2.5 Generation of an E2 random-mutation display library in *D. melanogaster* S2 cells

A randomly mutated pMT::H77 library plasmid pool was isolated from *E. coli* XL10 cells and subsequently used to transfect *D. melanogaster* S2 cells. The resulting transfectants were termed the H77 library. Stable transfection and expression of the E2 library was confirmed by Western blot analysis with AP33 (Figure 2-13). Expression of E2 on the cell surface of the random mutant library and correct fold of the protein was determined by flow cytometry using the MAbs AP33, AR3A, AR2A and H53 (10 μg/ml) (Figure 2-18). AP33 binding demonstrated that protein was expressed at the cell surface. AP33 and H53 bound the library-expressing cells at a similar level (34.94% vs. 35.28%) suggesting that the cells expressed correctly folded E2.

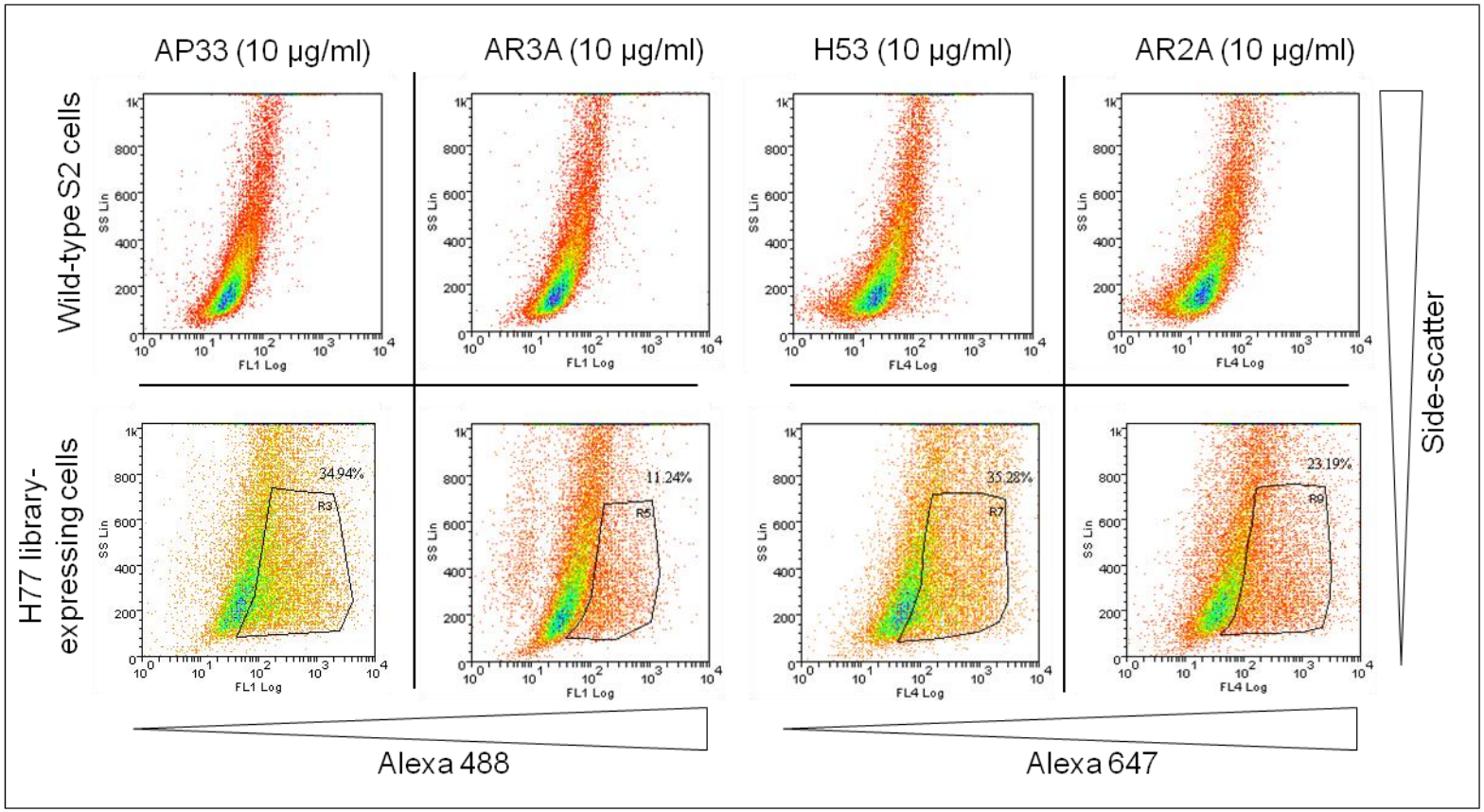

Figure 2-18. Expression of a randomly mutated $E2_{715}$ (H77) library on the surface of *D. melanogaster* S2 cells. Cells were induced to express protein for 120 hours and protein expression in 1 x 10^6 cells measured by staining individual cell aliquots with MAbs AP33, AR3A-A488, H53 or AR2A-A647. Cells binding the specific MAb are enclosed by a region and the percentage of cells given above. Wild-type S2 cells were included as the negative control (top panel).

However, the amount of protein expressed by the H77-library expressing cells appears to be lower than levels of $E2_{715}HA_{TM}$ expression (34.94% vs. 62.41%, Figure 2-18 and 2-15). This may be due to the transfection process which relies upon co-transfection of pMT and pCoBlast and may result in cells transfected with pCoBlast alone. Binding of AR2A and AR3A was observed but at a lower level than AP33 or H53 binding. This may be due to a lower affinity of the MAbs for surface-expressed E2. Alternatively, the process to directly label the MAbs may have been inefficient. This could result in unlabelled MAb binding to the cells that cannot be detected, thereby reducing the fluorescent detection of these cells. However, this provided enough evidence that the H77 library is able to express correctly folded E2 and can bind to all 'test' MAbs.

Protein expression was subsequently tested on both $E2_{715}HA_{TM}$-expressing cells and the H77 library-expressing cells using the directly labelled MAbs AR3A-A488 and pooled AR2A-A647 plus H53-A647 (Figure 2-19). All MAbs bound to the cell confirming the presence of surface expressed protein. The combined level of H53 plus AR2A binding to H77 library-expressing cells (20.7%) was lower than the level of binding observed previously with individual antibodies (35.28%, H53; 23.19%, AR2A, Figure 2-18). Again, this may be due to the presence of inefficiently labelled MAb that cannot be detected. Some non-specific labelling of H53 was also observed suggesting that un-conjugated fluorescent dye was present in this antibody preparation. Reducing the number of washes during the staining protocol did not improve MAb binding, nor did increasing the concentration of antibody used.

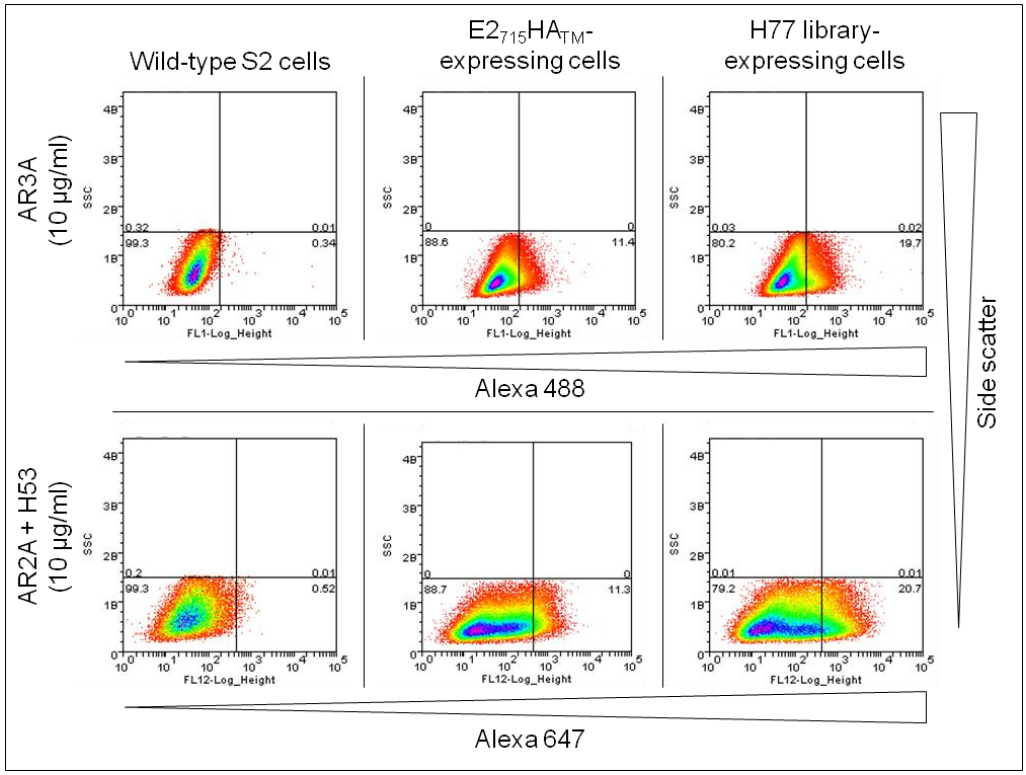

Figure 2-19. Binding of pooled directly labelled MAbs to the surface of *D. melanogaster* S2 cells expressing $E2_{715}HA_{TM}$ or the H77 library. Cells were induced for 120 hours and protein expression in 1 x 10^6 cells measured by flow cytometry. Binding of AR3A-A488 alone is shown in the top row. Binding of pooled H53-A647 and AR2A-A647 are shown in the bottom row. The percentage of cells staining with each antibody is give in the bottom right-hand quadrant. Wild-type S2 cells were included as negative control (left-hand column) and provided the baseline for gating positive cells. Cells were gated on forward scatter vs. side scatter.

2.4.2.6 Maintaining the effective size of the randomly mutated H77 library

To ensure that the diversity of the library was maintained following transfection into *D. melanogaster* S2 cells, stably transfected cells were screened and sequenced. The diversity of the plasmid contained within the *E.coli* XL10 cells was also assessed by sequencing after resurrection of the glycerol stock. The results suggested that the frequency of mutation and the number of distinct sequences within the library were similar to those reported in Table 2-3. Therefore, storage and resurrection of the plasmid pool and

transfection into insect cells did not appear to alter the diversity of $E2_{715}$ within the H77 library.

In order to maintain the diversity of the library, it was important to consider transfection efficiency and protein expression levels. Transfection of *D. melanogaster* S2 cells with a green fluorescence protein (GFP)-containing plasmid indicated that transfection efficiency was approximately 25% (Figure 2-20). Transfection was carried out on 5 x 10^6 cells; therefore between 7.5 x 10^5 and 1.25 x10^6 cells should be transfected with the pMT plasmid. This provides an excess of cells for transfection of a library size of 1 x 10^4.

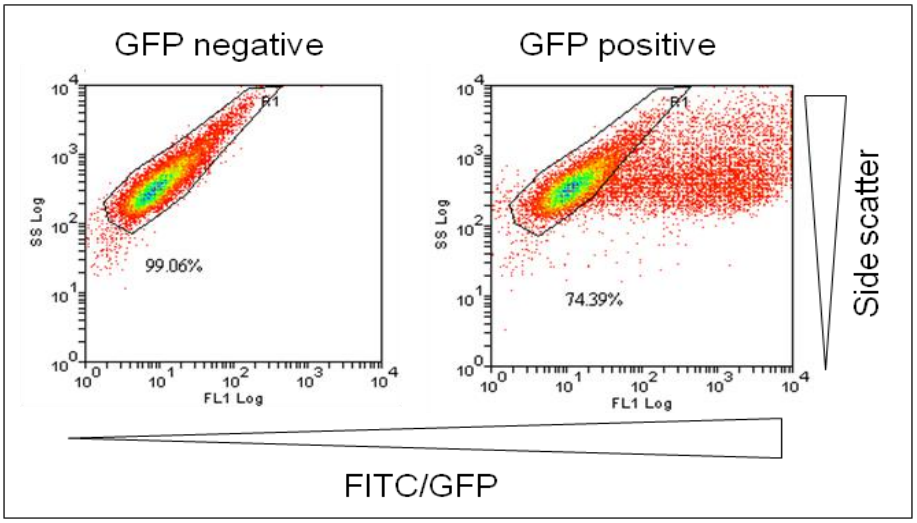

Figure 2-20. Efficiency of plasmid transfection into *D. melanogaster* S2 cells. A pMT::GFP plasmid containing green fluorescence protein was transfected into 5 x 10^6 cells. GFP was measured in the fluorescein isothiocyanate (FITC) channel of a Beckman Coulter FC500 Flow Cytometer with 50,000 events captured. Fluorescence in untransfected, wild-type cells was used to assess background staining. The percentage of cells that do not fluoresce (falling within R1) is given in each dot plot. Approximately 25% of transfected cells fluoresce.

The transfection protocol uses 2 µg plasmid, or 3.95 x 10^{11} copies of plasmid (Staroscik, 2004); therefore each plasmid variant should be represented multiple times over. Finally, it was important to consider levels of protein expression within the library. Based on AP33 and H53 binding to the H77

library, approximately 35% of cells expressed protein (Figure 2-18). Provided therefore that a minimum of 3 x 10^4 cells were labelled and analysed, 10^4 cells would be recovered, sufficient to theoretically cover the entire library, if each amino acid is mutated once. All protein analysis was carried out on 1 x 10^6 cells to provide the greatest possibility of including all variant protein constructs contained within the library. Considering that the effective library size (i.e. sequences that do not contain any stop codons or in-frame deletions) is 8701 (Table2-3), it was felt that sufficient cell numbers were transfected and analysed to allow for assessment of the total library.

2.4.2.7 Sorting of loss-of-binding mutants expressed on *D. melanogaster* S2 cells

The use of *D. melanogaster* S2 cell-expressed E2 as a means of mapping antibody-binding residues was tested in an artificially mixed population of cells and in the H77 library. The artificially mixed population consisted of $E2_{715}HA_{TM}$-expressing cells and $E2_{715}$D535A-expressing cells mixed at a 1:1, 10:1 or 100:1 ratio. Initial staining analysis was carried out to determine the parameters for gating and sorting in order to collect the cells of interest. Analysis of mixed cells at a 1:1 ratio with AR3A showed cells staining in the bottom right-hand quadrant of the dot-plot, while AR2A and H53 stained cells appeared in the top left-hand quadrant (Figure 2-21). When the cells were stained with all MAbs simultaneously a small population was seen in the top left-hand quadrant. This corresponds to those cells which were able to bind the conformation-sensitive MAbs, AR2A and H53, but which failed to bind the test MAb, AR3A. This was therefore deemed to be the population of interest

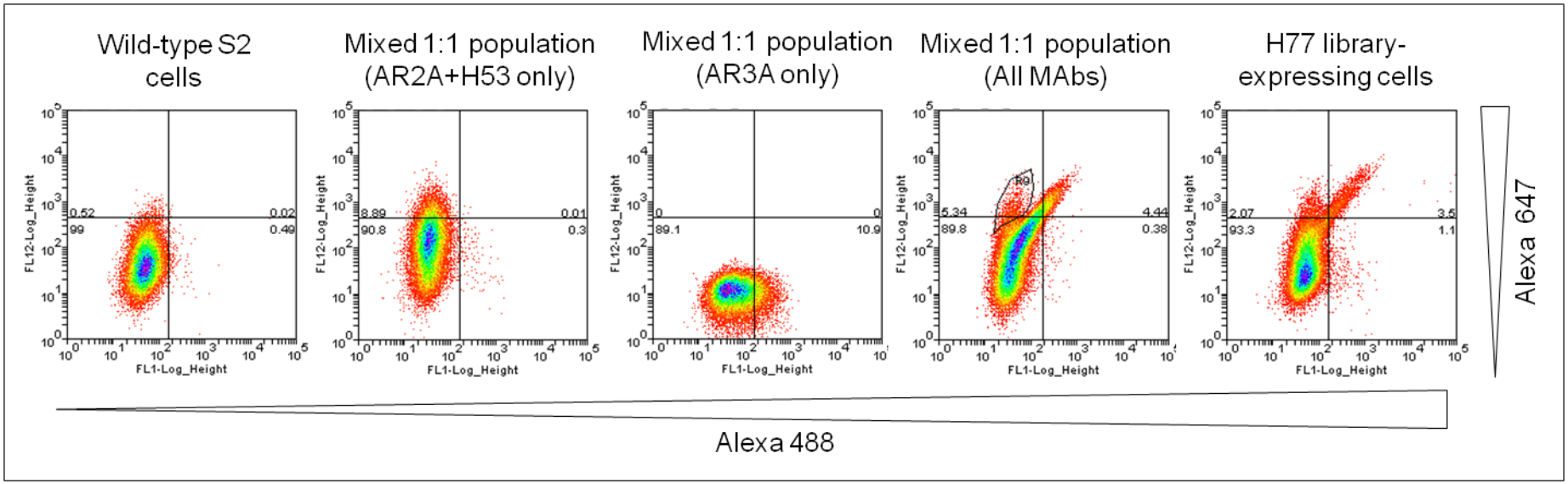

Figure 2-21. Identification of a population of interest within a culture of *D. melanogaster* S2 cells. An artificially mixed population of cells as well as the H77 library were stained with combinations of directly labelled MAbs. Antibody binding was measured in 1 x10^6 cells after 120 hours of protein expression. Wild-type S2 cells were included as negative control (left-hand panel) and provided the baseline for gating positive cells. Bottom left-hand quadrant; percentage of cells that do not bind antibody. Bottom right-hand quadrant; percentage of cells staining with AR3A-A488 only. Top left-hand quadrant; percentage of cells binding AR2A-A647 and H53-A647 only. Top right-hand quadrant; percentage of cells binding all MAbs. Cells were gated on forward scatter vs. side scatter.

for future sorting experiments and is more easily seen when designated as a region (Figure 2-21). The same population of cells in the top left-hand quadrant was also seen when H77 library-expressing cells were tested with pooled MAbs (Figure 2-21). Repeat analysis of the H77 library-expressing cells produced similar results, with the population of interest ranging between 1.00% and 2.50% of the total cell population (Figure 2-22a).

Mixed populations of cells at a 1:1, 10:1 or 100:1 ratio were tested to see if the population of cells gated in the top left-hand quadrant decreased as the number of $E2_{715}$D535A cells decreased. As expected, there was a decrease in the percentage of cells staining positive for H53 and AR2A but negative for AR3A as the ratio of $E2_{715}$D535A cells decreased (Figure 2-22b). Data from two separate experiments showed a similar trend with cells mixed at a 1:1 ratio producing the largest population of gated cells. Fewer cells were observed in the region of interest for cells mixed at a 10:1 ratio with even fewer cells observed for cells at a ratio of 100:1 (Figure 2- 22b). Despite this overall trend, the percentage reduction in cells within the region of interest did not precisely correlate with the ratio of cells in the artificially mixed population. In Figure 2-22b (top row), the population of interest in the 1:1 mixed culture accounts for 2.87% of the total. The population of interest in the 10:1 mixed culture is 0.25% of the total. This 10-fold reduction correlates well with a 10-fold reduction in the number of $E2_{715}$D535A cells; however the population of interest in the 100:1 mixed culture accounts for 0.20% of the total. This is very similar to the percentage of cells observed in the 10:1 mixed culture and is also approximately 10-fold lower than the percentage of cells observed in the 1:1 mixed culture. This could be caused by non-specific staining of cells leading

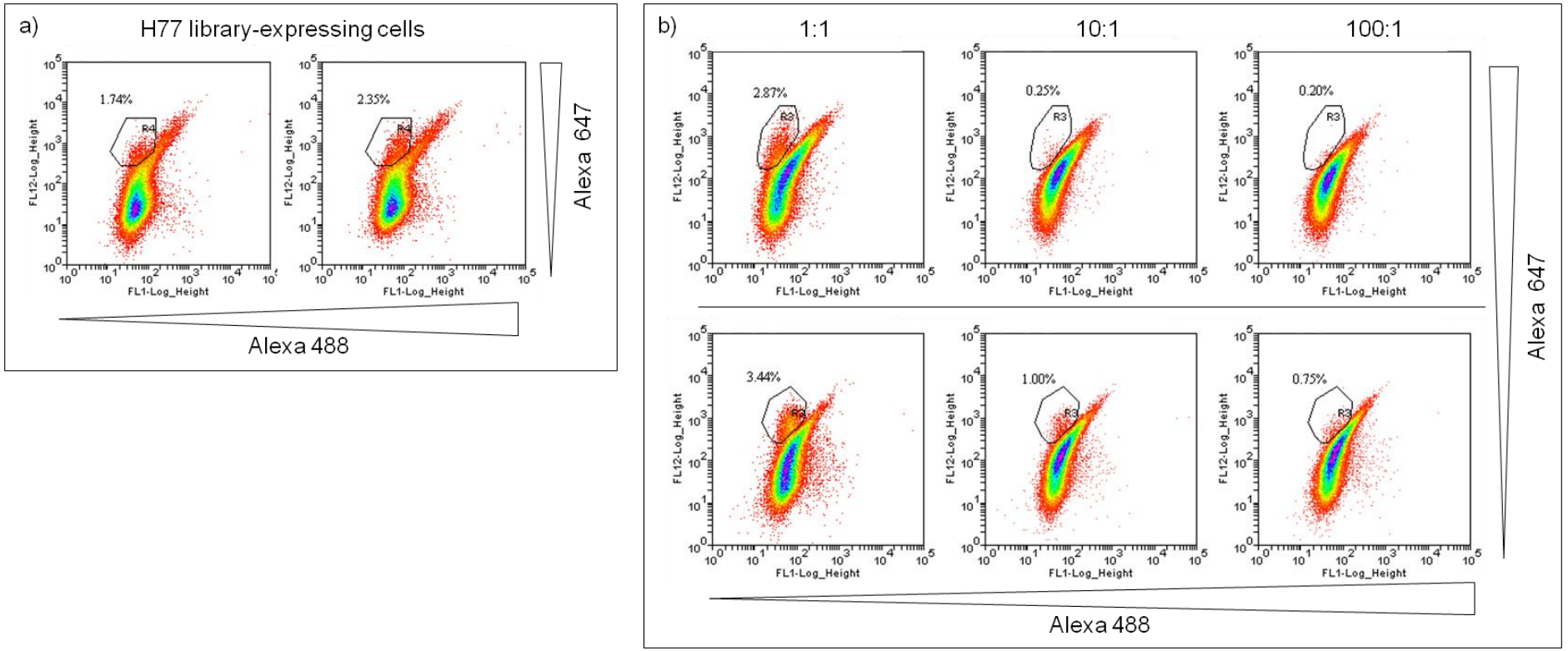

Figure 2-22. Measuring the size of a population of interest within a culture of *D. melanogaster* S2 cells. Cells were stained with pooled directly labelled MAbs. Antibody binding was measured in 1 x10^6 cells after 120 hours of protein expression. Population of interest (AR2A/H53 positive and AR3A negative) is highlighted by a region and percentage of cells in each region given in the top left-hand corner. Cells were gated on forward scatter vs. side scatter. (a) H77 library-expressing cells. Panels correspond to two separate experiments. (b) Artificially mixed populations of $E2_{715}HA_{TM}$-expressing cells and $E2_{715}$D535A-expressing cells at 1:1, 10:1 or 100:1 ratio. Each row corresponds to two separate experiments.

to false positives, thereby increasing the number of cells within the population of interest. Indeed the position of the gate in Figure 2-21 (panel 1) shows 0.52% of wild-type S2 cells staining in the top-left hand quadrant. This value may fluctuate meaning that cells from a 100:1 culture staining in this quadrant are likely to be composed largely of false positive cells. It is also important to bear in the mind the different expression levels observed between $E2_{715}HA_{TM}$ and $E2_{715}$D535A expressing cells; therefore a mixed culture based on cell number will not correlate precisely with the amount of protein present. As approximately 40-45% of $E2_{715}$D535A cells expressed protein (Figure 2-15), at a ratio of 100:1, very few $E2_{715}$D535A-expressing cells would be stained and sorted into the population of interest. A similar pattern was observed in Figure 2-22b (bottom row), with an overall trend towards a decreasing population of interest but discrepancies between the precise figures. The 10:1 mixed culture has a population of interest of 1.00%, a reduction of approximately one third compared to the 1:1 mixed population (3.44%). The 100:1 mixed culture has a population of interest of 0.75%, approximately 4.5-fold lower than the population observed for the 1:1 mixed culture. Again these discrepancies are probably caused by non-specific antibody binding leading to false positives within the region of interest and ambiguities in the actual amount of protein present. This data does still provide evidence that *D. melanogaster* S2 cells expressing E2 on their cell surface can be selected and sorted on the basis of antibody binding.

Once the population of interest had been identified, the cells were sorted in bulk into 96-well tissue culture plates for continued culture and expansion, or individually into 96-well PCR plates for PCR and sequence analysis.

2.4.2.8 Expansion of sorted cells

D. melanogaster S2 cells do not grow well when seeded at a density below 5 x 10^5 cells/ml (or 500 cells/µl). This presents a problem when sorting small populations of cells as the loss-of-binding mutant may represent only a tiny fraction of the total library population. The volume of media into which cells are seeded is also limited by the rate at which the media evaporates in the incubator. Expanding very small numbers of cells is difficult and early experiments were hampered by the low number of cells that were sorted and the inability to keep them alive thereafter. It was therefore necessary to bulk sort all cells into a 96-well plate containing 500 cells/µl of wild-type *D. melanogaster* S2 cells to provide a sufficient cell density for the cells to survive and grow. With this method it was possible to keep sorted cultures of cells alive for several weeks and expand them into larger culture vessels. There were however, issues with contamination, despite the use of Penicillin-Streptomycin in the wash buffer and culture media, and several cultures were lost in the early days post-sorting. Plate-based culturing and microscopic analysis of the bacterial contaminant indicated a *Pseudomonas* spp. which may have been introduced in the sheath fluid contained within the flow cytometer. The sheath fluid was subsequently autoclaved prior to use which largely solved the problem, however observation of strict aseptic technique was essential to ensure survival of the cultures. The sorted cultures were also very sensitive to seeding at a low density and the addition of too much media, even after several days' growth. Only one culture of sorted cells was maintained long enough for re-sorting to be attempted. This was from an artificially mixed population of cells seeded at a 100:1 ratio ($E2_{715}HA_{TM}$:$E2_{715}$D535A) from which the

population of interest had been sorted (Figure 2-22b). These cells were cultured with wild-type *D. melanogaster* S2 cells for 3 weeks and expanded prior to selection with Blasticidin S (25 µg/ml) for 2 weeks. Repeat sorting showed that a higher percentage of cells were now staining within the population of interest, 1.30% vs. 0.20% (Figure 2-23), an increase of 650%, suggesting that the $E2_{715}$D535A-expressing cells had been enriched within the total population. Initial experiments had indicated that cells sorted from a 100:1 ratio mixed culture were likely to contain a large number of false positives (Figure 2-22). However, the ability to isolate a larger population of interest following expansion and re-sorting of cells demonstrates the importance of such iterative enrichment.

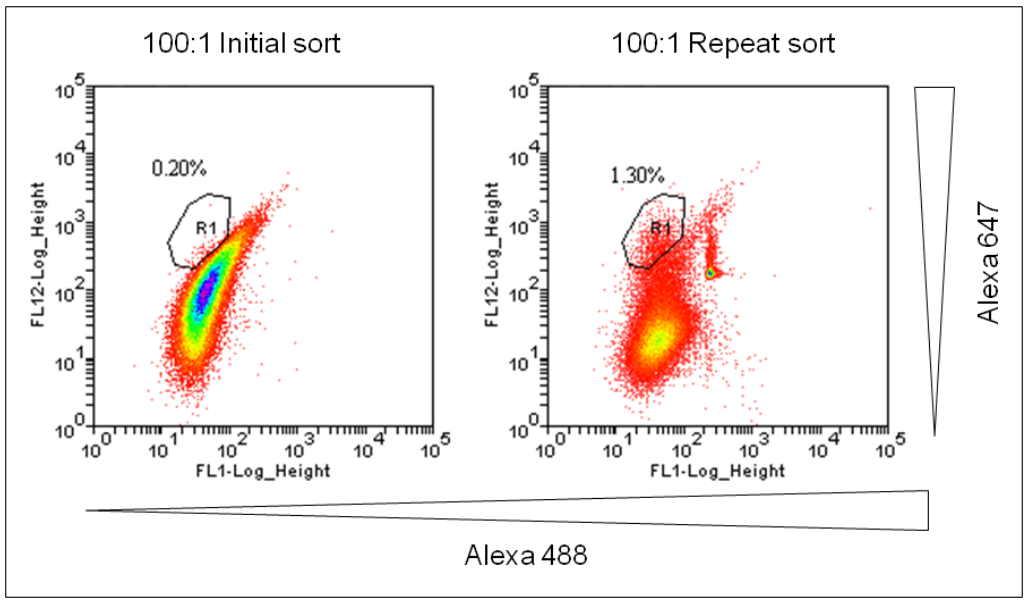

Figure 2-23. Enrichment of a population of interest within a culture of *D. melanogaster* S2 cells by repeat sorting. An artificially mixed population of $E2_{715}HA_{TM}$-expressing cells and $E2_{715}$D535A-expressing cells were sorted for loss-of-binding to MAb AR3A (left-hand plot). The selected population was grown for 3 weeks and selected for a further 2 weeks. Cells were re-sorted for loss-of-binding to MAb AR3A (right-hand plot). Cells were stained with a pool of directly labelled MAbs. In both rounds of sorting, antibody binding was measured in 1 $x10^6$ cells after 120 hours of protein expression. Population of interest is highlighted by a region and percentage of cells in each region given in the top left-hand corner. Cells were gated on forward scatter vs. side scatter.

Although this data does not represent large numbers of cells, it does indicate the potential value of such an approach, particularly if cells undergo several rounds of enrichment. The staining patterns of the two cultures were slightly different; the sorted and enriched population stained less brightly (Figure 2-23) with higher forward scatter and side scatter values (data not shown), indicating that this culture was less healthy than the initial culture. This may be caused by the presence of wild-type *D. melanogaster* S2 cells within the culture which, in the presence of blasticidin, would be dead or dying. Prolonged exposure to the selective agent should remove all wild-type cells and continued culturing of the selected cells should result in a healthier population.

2.4.2.9 Sequence analysis of sorted single S2 cells

Single cells were sorted from the H77 library and from mixed cultures of $E2_{715}HA_{TM}$-expressing cells and $E2_{715}$D535A-expressing cells at a 1:1 ratio. Cells were sorted into 96-well PCR plates and screened (Figure 2-24). Recovery rates of the $E2_{715}$ PCR product ranged from 42% to 70%. Recovery rates less than 100% may be due to failure of the flow cytometer to accurately deposit a single cell into each well, resulting in some empty wells. It is also possible that random mutagenesis of the H77 library generated mutations within the primer binding sites preventing amplification of $E2_{715}$. It is also possible for cells to become transfected with the resistance plasmid pCoBlast only, resulting in blasticidin-resistant cells which do not express protein. These cells may be stained as false positives and deposited into the plate; however subsequent PCR analysis would fail to amplify $E2_{715}$.

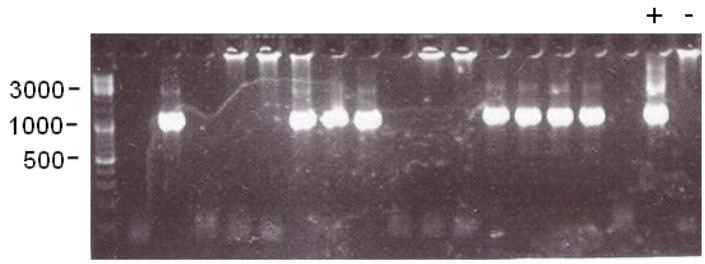

Figure 2-24. Screening single *D. melanogaster* S2 cells for the presence of $E2_{715}$. Representative agarose gel electrophoresis results showing products at approximately 1.1 kb. +, positive control (plasmid pMT::$E2_{715}HA_{TM}$). -, negative control (no template).

PCR positive samples were sequenced to determine which mutations within $E2_{715}$ might be responsible for the loss of antibody binding. Any sequences that appeared to be derived from more than one template were discounted from the analysis. As expected, 100% of samples (out of 10 samples sequenced) sorted from the artificially mixed population contained the D535A mutation. This mutation is known to abrogate AR3A binding; therefore within an artificial population containing wild-type $E2_{715}$ and mutant $E2_{715}$D535A only, cells no longer able to bind AR3A would be expected to contain the D535A mutation. This provided further proof-of-principle that the insect cell display library could be used to map antibody-binding residues.

A total of 48 sequences isolated from the H77 library were analysed. Table 2-4 lists the mutations recorded and their frequency. Most mutations appear only once; however, two mutations appear at high frequency within the samples tested. These are L399I (81.6% of sequences) and C644R (85.7% of sequences). Neither of these residues is critical for AR3A binding, according to previous analysis; however, L399 does fall within the first region forming the discontinuous epitope AR3 (AR3A spans aa396-424, aa436-447 and aa523-540) (Law *et al.*, 2008). In previous studies this residue was not subject to

Table 2-4. Mutations isolated from the H77 display library sorted for loss of reactivity to AR3A. Frequency is shown in parenthesis**.**

Mutation	Frequency (%)
T396Q	1 (2.0)
L399I	40 (81.6)
W420R	1 (2.0)
N423K	1 (2.0)
C429R	1 (2.0)
S473G	1 (2.0)
Y489C	1 (2.0)
G565A	1 (2.0)
I571T	1 (2.0)
G572R	1 (2.0)
R596W	1 (2.0)
C644R	42 (85.7)
P664S	1 (2.0)
S668T	1 (2.0)
H691L	1 (2.0)
D698V	1 (2.0)
Y701F	1 (2.0)
Y720F	1 (2.0)
Y720H	1 (2.0)
S725C	1 (2.0)

alanine-scanning mutagenesis (Law *et al.*, 2008) and it will be necessary to test its role in AR3A binding by carrying out antibody binding studies with E2 containing isoleucine at position 399. Isoleucine and leucine both contain aliphatic side chains and are hydrophobic in nature. Such a conservative

mutation would not be expected to cause massive changes in the overall protein fold but may be sufficient to alter the affinity of the MAb for its epitope. In all sequences containing isoleucine at position 399, mutation was due to a single point mutation from C to A (at nucleotide 46). Isoleucine is a naturally occurring variant at residue 399 in some genotype 2a, 3, 4 and 6 sequences, according to the Los Alamos HCV sequence database. The high frequency of this mutation within isolated sequences may be indicative of a crucial role in AR3A binding, although it may also be due to over-representation of this mutation within the library. If this mutation were introduced early during the mutagenesis process it may have become 'fixed' within the library and its importance in AR3A binding to E2 may be overstated.

The residue C644 does not lie within the AR3 discontinuous epitope nor is it involved with AR3A or CD81 binding. It does however form a disulphide bond with the cysteine residue at position 607 (corresponding to the eighth disulphide bond), according to the recently described E2 structure (Krey *et al.*, 2010). Introduction of arginine at position 644 would abolish the disulphide bond and, as such, this mutation would be expected to alter the overall fold of the protein. However, staining by the conformation sensitive MAbs AR2A and H53 (Figure 2-22a) indicated that this protein was correctly folded. A recent study by McCaffrey *et al*., (2012) showed that although H53 reactivity was abolished by replacement of C644 with an alanine residue in the context of HCVpp, in $sE2_{661}$ H53 binding was maintained, albeit reduced (to 25-75% of wild type). In both cases this mutation prevented E2 binding to CD81-LEL (McCaffrey *et al.*, 2012). This mutation may therefore influence CD81, and consquently also AR3A binding, by altering the overall presentation of the

discontinuous CD81 and AR3 epitope. This would need to be further investigated by testing AR3A binding to E2 containing arginine at position 644. The lower levels of conformational antibody binding to the cysteine mutant in the study by McCaffrey *et al.*, may be caused the use of an alanine replacement. Alanine is a smaller amino acid than the arginine residue observed in our mutants and may result in larger conformational changes to the protein. This study also tested the combined binding of two conformation-sensitive MAbs targeting different epitopes. It is also worth noting that all sequences containing the C644R mutation (except one) also contain one additional mutation within the first AR3 epitope, (Table 2-5) raising the possibility that this mutation does not function directly in AR3A binding and is simply over-represented in the library.

Table 2-5. Combinations of mutations in isolated library sequences

Mutated residues within sequenced $E2_{715}HA_{TM}$	Frequency (%)
L399I, C644R	36 (73.5)
I571T, S668T	1 (2.0)
T396Q, C644R	1 (2.0)
L399I, G565A, C644R	1 (2.0)
L399I, C644R, P664S	1 (2.0)
L399I, G572R, C644R	1 (2.0)
L399I, C644R, L720H	1 (2.0)
W420R, N423K, C429R S473G, Y489C	1 (2.0)
H691L, D698V, Y701F, Y720F, T722S, S725C	1 (2.0)
C644R	1 (2.0)
R596W	1 (2.0)
No mutations	3 (6.1)

The mutations T396Q, W420R and N423K are located within the AR3 discontinuous epitope (Law *et al.*, 2008) and may therefore reduce the affinity of AR3A binding. These mutations only appear once each within the isolated sequences. The mutations W420R and N423K appear within the same sequence (Table 2-5) and it is possible that they have an additive effect on AR3A binding. Mutations within AR3 would be expected to be recovered at a higher frequency as they should have the greatest impact on AR3A binding. The low frequency of such mutations may be due to the relatively small number of sequences that were analysed. The mutation at N423 is unusual as it knocks out the second N-linked glycan of E2. Binding of AR2A and H53 indicated that this protein was correctly folded (Figure 2-22a), an observation supported by previous studies showing that E2 lacking the N-linked glycan at position 423 is immunoprecipitated by H53 (Goffard *et al.*, 2005). This mutation was reported to reduce binding of AR3A to E2, albeit at an intermediate level (Law *et al.*, 2008). It is therefore not surprising that this mutation should have been isolated and demonstrates the utility of this method in identifying residues involved, indirectly at least, in MAb binding.

Residues T396 and L399 both fall within HVR1. Although this encompasses part of AR3, these residues have not been directly implicated in the binding of AR3A to E2 (Law *et al.*, 2008). The position of these residues in HVR1 and their possible role in AR3A binding is potentially interesting. MAb AR3A has been demonstrated to interfere with CD81 binding to E2 (Law *et al.*, 2008), whereas HVR1 has been implicated in binding to SR-BI (Bartosch *et al.*, 2003c; Scarselli *et al.*, 2002). It is tempting to speculate that AR3A binding to these residues may interfere with HVR1-SR-BI interactions. However, further

studies would be needed to test if AR3A really could play a role in inhibiting SR-BI binding to E2.

The cysteine at residue 429 forms a disulphide bond with the cysteine residue at position 552, corresponding to the first disulphide bond (Krey *et al.*, 2010). Replacement of this cysteine with arginine would abolish the disulphide bond. The study by McCaffrey *et al.* (2012) found that mutation of C429 to alanine reduced H53 binding to E2 in the context of HCVpp but had no effect on binding to $sE2_{611}$. This discrepancy may have been caused by the different forms of E2 tested. In both contexts, the C429A mutation abolished binding to CD81-LEL (McCaffrey *et al.*, 2012). This is consistent with the location of disulphide 1 within DI of E2, which also contains the majority of CD81 binding determinants (Krey *et al.*, 2010). This cysteine residue falls between two discontinuous epitopes forming AR3 and within a known CD81 binding region. Therefore it may alter the whole antigenic region preventing or reducing both AR3A and CD81 binding.

The remaining E2 mutations (S473G, Y489C, G565A, I571T, G572R, R596W, P664S, S668T, H691L, D698V and Y701F) have not been extensively studied in the context of AR3A binding. They do not fall within antigenic region 3, as identified by antibody competition studies (Law *et al.*, 2008), nor are they involved in CD81 binding (Clayton *et al.*, 2002; Drummer *et al.*, 2006; Flint *et al.*, 1999a; Owsianka *et al.*, 2001; Owsianka *et al.*, 2006). Indeed residues P664, S668, H691, D698 and Y701 lie within a region of E2 that is often excluded from studies relying upon truncated $sE2_{661}$. As such they present an intriguing set of residues for further investigation. The majority of the substitutions within E2 are non-conservative and as such would be

expected to alter the structure of the protein, thereby reducing AR3A binding. A cluster of mutations around residues 565 to 571 may indicate the position of another region involved in AR3A binding and could be of interest for further study.

The mutations Y720F, Y720H and S725C occur within the HA_{TM} tag. Mutations within the tag are possible due to the nature of the random mutagenesis method. These mutations may alter anchoring of $E2_{715}$ within the cell membrane and reduce levels of expression at the cell surface; however any cells carrying deleterious mutations would most likely be removed from further analysis as they fail to generate surface expressed protein. Such mutations should not interfere directly with the ability of AR3A to bind $E2_{715}$, and all sequences containing mutations within the HA tag were found to contain other mutations within $E2_{715}$ responsible for loss of binding (Table 2-5).

Table 2-5 shows the appearance of multiple mutations within the isolated sequences. The majority of sequences contained one or two mutations and none of the 49 sequences contained a stop codon or in-frame deletion. This is in line with the statistical assessment of the library size and diversity (Table 2-3) which was designed to reduce the frequency of such mutations. Only three sequences were found to contain no mutations. These were clearly among the false positive staining cells and again highlight the importance of iterative enrichment of a population of interest. Six sequences were found to contain three or more mutations. Sequences containing large numbers of mutations are not desirable for mapping studies as the process of identifying the important mutations becomes more difficult. It is likely that only one of the mutations within each of these sequences is responsible for loss-of-binding but each

residue would need to be investigated individually. However, the presence of mutations W420R, N423K and C429R in the same sequence is potentially interesting. The close proximity of these residues suggests they form part of the AR3A epitope and have an additive effect on AR3A binding; indeed W420 and N423 have already been postulated to form part of AR3 (Law *et al.*, 2008). Similarly the presence of H691L, D698V and Y701F in the same sequence may indicate the presence of a novel epitope involved in AR3A binding.

2.4.2.10 Failure of the UKN2B1.1 library

Initially it was hoped that a randomly mutated display library could be generated in a genotype 2b $E2_{715}$ as well as genotype 1a. To this end, the pMT::$E2_{715}HA_{TM}$ (UKN2B1.1) plasmid was successfully generated and stably transfected into *D. melanogaster* S2 cells. Western blot analysis of protein expression with AP33 suggested that protein was produced in the $E2_{715}HA_{TM}$ (UKN2B1.1) transfected cells (Figure 2-25a). However subsequent flow cytometry analysis of intracellular and cell-surface protein expression indicated that protein was not expressed as expected. AR3A bound only around 1% of $E2_{715}HA_{TM}$ (UKN2B1.1) expressing cells regardless of whether staining was at the cell surface or intracellular (Figure 2-25b) suggesting that the protein was incorrectly folded. To exclude the possibility that protein was not translocated to the cell surface, the sequence of the HA_{TM} domain within the $E2_{715}HA_{TM}$ (UKN2B1.1) construct was checked and contained no mutations that might have reduced or abrogated cell-surface expression of E2. There might however have been mutations within the BiP secretion signal contained within pMT, reducing the ability of the cell to translocate protein to the ER. BiP targets

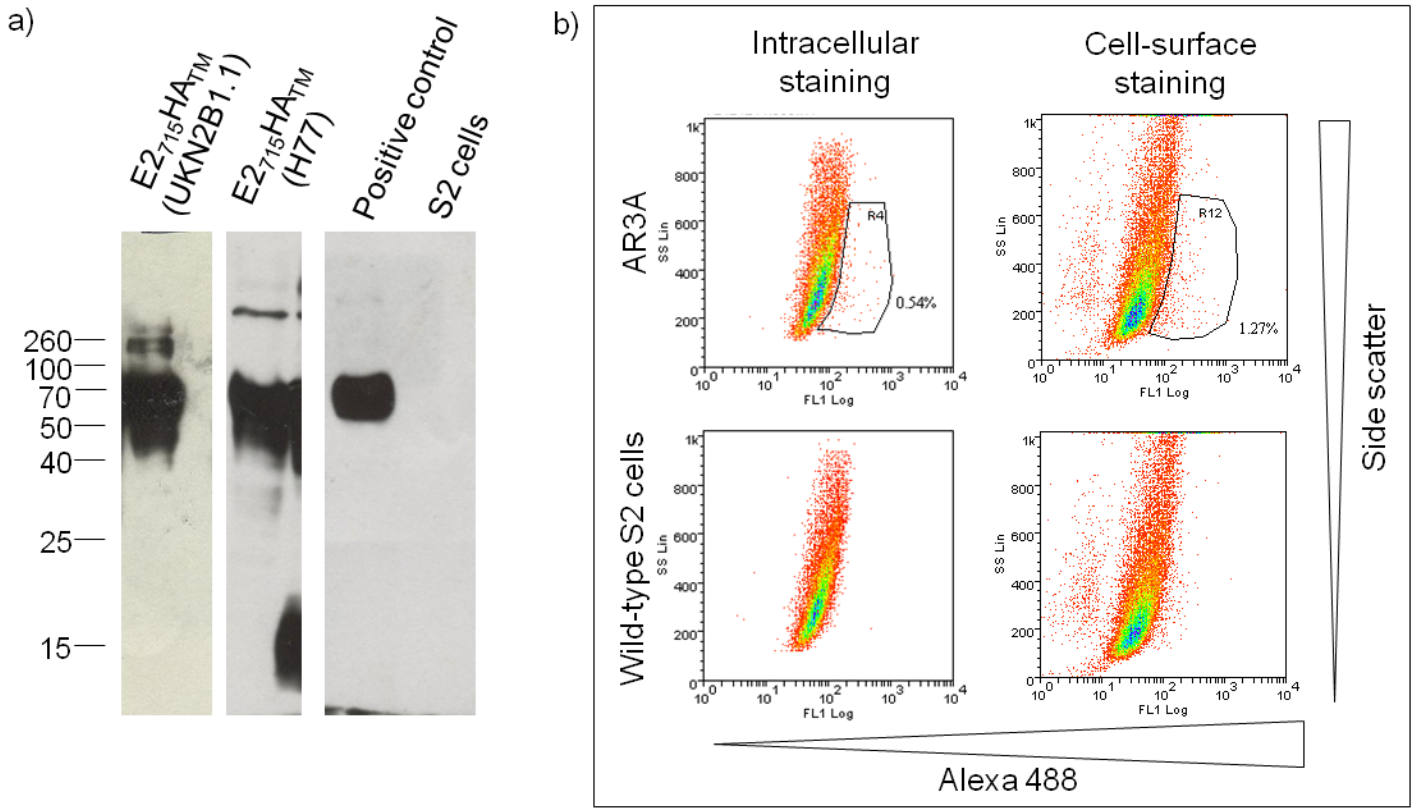

Figure 2-25. Expression of UKN2B1.1 $E2_{715}HA_{TM}$ in *D. melanogaster* S2 cells. a) Cell lysate from 1 x 10^6 cells was analysed by Western blot under reducing conditions using MAb AP33. Wild-type S2 cells served as negative control. $sE2_{661}$ served as positive control. Molecular weight marker is shown in kDa. b) Cell-surface and intracellular protein expression after 120 hours in 1 x 10^6 cells measured by flow cytometry. Cells were intracellularly (left-hand column) or surface stained (right-hand column) with MAb AR3A. Wild-type cells (bottom row) are shown for baseline binding levels. Percentage of cells binding AR3A is shown.

proteins to the ER lumen and facilitates protein folding (Kirkpatrick *et al.*, 1995); therefore any reduction in BiP function may result in incorrectly folded protein. This could explain how protein was produced and detected by AP33 binding a linear epitope, as demonstrated by Western blot analysis, but could not be detected on or within the cell by the conformation-dependent MAb, AR3A. Sequencing of the BiP secretion signal would be needed to confirm this hypothesis. Alternatively, mutation of BiP could result in poor protein expression. However, due to time constraints it was not possible to repeat the flow cytometry with a MAb binding to a linear epitope, such as AP33, that may have resolved the issue of poor expression or incorrect folding.

2.5 Discussion

Initially this project was designed with the Yeast display library system in mind. It was intended that *S. cerevisiae* EBY100 cells would be used as the host for an E2 random mutant library. This system was originally developed for the expression of antibody single chain Fv (scFv) fragments via the cell surface mating machinery of yeast (Boder & Wittrup, 1997). It was found however, that only low levels of E2 expression could be obtained from *S. cerevisiae* EBY100 cells. Expression levels, as assessed by Western blot, were considerably lower than either of the two controls (Figure 2-7). Hyperglycosylation of proteins is a common problem with yeast-expressed proteins; however Western blot analysis of the $E2_{715(stop)}$ protein indicated that it was not hyperglycosylated. *S. cerevisiae* is capable of elongating the core oligosaccharide molecule, $GlcNAc_2$-Man_8, by the addition of up to 200

mannose residues, a phenomenon that occurs in the Golgi apparatus and is postulated to protect yeast glycoproteins excreted into the extracellular environment (Conde *et al.*, 2004). However, hyperglycosylation of E2 could result in incorrect folding of the protein (Goffard *et al.*, 2005) and alter its immunogenicity (Falkowska *et al.*, 2007). The presence of a single band between 72 and 95 kDa (corresponding to $E2_{715(stop)}$ with the Aga2p fusion protein) and the absence of a large protein smear (Figure 2-7) would suggest that monomeric E2 is expressed at the cell surface. $E2_{715(stop)}$ also appeared to be correctly folded as it bound the conformation-sensitive MAb H53 at a similar level to the MAb AP33, which recognises a linear epitope (Figure 2-10). Expression of the Aga2p protein from *S. cerevisiae* EBY100 cells was not enhanced by increasing the duration of protein induction (Figure 2-8) suggesting that the optimal conditions for expression were employed. It is interesting to note that flow cytometry analysis of $E2_{715(stop)}$-expressing cells with the anti-Xpress antibody showed much higher levels of surface staining (49.5%) than staining with the MAbs AP33 or H53 (16.6-22.7%) (compare Figure 2-9 and 2-10). This demonstrates that protein was efficiently expressed in the cells but suggests a problem with E2 expression specifically. The E2 construct expressed in these cells contained a stop codon immediately downstream of the E2 glycoprotein (Figure 2-6). Whilst this abolished the V5 epitope and the poly-histidine tag preventing detection or purification via these tags, it should not affect the coding region of the Aga2p fusion protein or E2. Therefore, it was not expected that expression of E2 would be altered. It is possible however that the addition of ~20 bp derived from pGEM-T Easy and/or the introduction of a stop codon altered the ability of the cells to process

and fold E2, thereby reducing its expression at the cell surface. Western blot analysis of cell lysates to identify intracellular protein aggregates or intracellular staining for flow cytometry may have proved informative in this regard. Alternatively the MATα transcription termination region, responsible for polyadenylation of mRNA, may have been altered. If this were the case however, it would be expected to reduce expression of all protein, not just the $E2_{715(stop)}$ region of the fusion protein.

The Yeast display library system has been successfully used to express and characterise antibody binding to the envelope glycoproteins of WNV (Oliphant *et al.*, 2005) and DENV (Sukupolvi-Petty *et al.*, 2010; Sukupolvi-Petty *et al.*, 2007). The envelope proteins (E) of these viruses are well studied and their crystal structures have been determined. Both the WNV and DENV E proteins are class II fusion proteins, with a three domain architecture (DI, DII and DIII) (Modis *et al.*, 2003; Nybakken *et al.*, 2006), analogous to the structure proposed for E2 (Krey *et al.*, 2010). The DENV E ectodomain (117 or 292 amino acids) used in mapping studies was smaller than the HCV $E2_{715}$ ectodomain (332 amino acids) used in our library. The larger WNV E ectodomain (415 amino acids) was successfully expressed in yeast cells; however the WNV E protein is less highly glycosylated, containing only one N-linked glycan (Nybakken *et al.*, 2006). The DENV E is also less highly glycosylated, containing just two N-linked glycans (Modis *et al.*, 2003), compared to the 11 potential N-linked glycans within HCV E2 (Goffard *et al.*, 2005). The DENV and WNV E proteins each contain just six disulphide bonds (Modis *et al.*, 2003; Nybakken *et al.*, 2006), whereas E2 contains nine (Krey *et al.*, 2010). This smaller size and reduced complexity of the E proteins may

facilitate their expression within the yeast cells. More recently the Yeast display system has been successfully adapted for expression of the HCV E2 protein (Sabo *et al.*, 2011). This was achieved using a more truncated version of the H77 E2 ectodomain, $E2_{660}$, resulting in the removal of one disulphide bond. A smaller E2 construct may therefore be necessary for successful expression within yeast cells. This does however limit the scope of any mapping carried out with such a library. The main advantage of such systems is the ability to interrogate the whole protein simultaneously and to map antibody-binding epitopes in any part of the protein without prior knowledge of potential binding sites. By removing the C-terminal 56 amino acids, any antibodies binding to this region could not be mapped.

This experiment does highlight the importance of checking the quality of any protein construct used in mapping studies. Although the data showed that $E2_{715(stop)}$ expressed at the cell surface was largely correctly folded (Figure 2-10), the overall levels of expression were poor. Use of poorly or incorrectly folded protein could result in the mistaken identification of loss-of-binding mutants. The recent paper by Sabo *et al.* (2011) did not check the fold of yeast-expressed E2 using a conformation-sensitive MAb such as H53. An oligoclonal pool of MAbs was used to test for cell-surface expression of E2; however this does not guarantee that the expressed protein is correctly folded, particularly as the pool of MAbs used was derived from mouse hybridomas and was largely uncharacterised. As discussed in Edwards *et al.* (2012), most neutralising mouse MAbs target linear epitopes within E2 restricting their use as determinants of protein fold. The E2 construct used by Sabo *et al*. was truncated at residue 660 which abolishes the final disulphide bond within E2

and, according to the recent E2 model (Krey *et al.*, 2010), leaves a 'spare' cysteine residue unable to bind with its partner residue. This may make this construct particularly vulnerable to mis-folding. Correct fold of the yeast-expressed E2 would need to be confirmed before the system could be widely adopted for mapping studies.

In attempting to develop a Yeast display library, the limiting factor was found to be the cloning method. Several variations of the cloning method were attempted including the use of different restriction sites, amplification of $E2_{715}$ from different strains or even different genotypes of HCV and the use of subcloning from an intermediate vector. Despite this, cloning continued to be problematic and ultimately resulted in the generation of a sub-standard construct, pYD1::$E2_{715(stop)}$. Poor cloning efficiency is likely to limit the size of the library generated and therefore the decision was taken to abandon this display library and instead to focus on the development of an insect cell display library.

D. melanogaster S2 cells have been used to express a number of different proteins such as Igs (Johansson *et al.*, 2007a; Kirkpatrick *et al.*, 1995) and viral glycoproteins (Ivey-Hoyle & Rosenberg, 1990) including the HCV E2 ectodomain from different viral genotypes (Krey *et al.*, 2010). Therefore the decision was made to generate a display library comprising amino acids 384-715 of E2, enabling us to study the entire ectodomain. $E2_{715}$ was successfully amplified from the H77 (genotype 1a) strain and UKN2B1.1 (genotype 2b) strain of HCV and adapted to cell surface display by fusion to the HA transmembrane region. Generation of an $E2_{715}HA_{TM}$ (UKN2B1.1) construct was not successful (Figure 2-25). Despite detecting protein by Western blot

with AP33, flow cytometry using AR3A could not detect cell-surface or intracellular E2. This suggested that $E2_{715}HA_{TM}$ (UKN2B1.1) was incorrectly folded although it is feasible that mutation to the BiP secretion signal within the plasmid may have reduced protein expression; however due to time constraints this could not be further investigated. In contrast, expression of the $E2_{715}HA_{TM}$ (H77) construct showed that the protein was expressed on the surface of *D. melanogaster* S2 cells and was correctly folded (Figure 2-14 and 2-15). The level of $E2_{715}HA_{TM}$ expression in the *D. melanogaster* S2 cells was considerably higher than that observed in the *S. cerevisiae* EBY100 cells. When comparing AP33 binding, only 16.61% of *S. cerevisiae* EBY100 cells expressing $E2_{715(stop)}$ stained positive, whereas 62.41% of *D. melanogaster* S2 cells expressing $E2_{715}HA_{TM}$ (H77) stained positive (Figure 2-10 vs. 2-15). This strongly indicated that insect cell expression would provide a better platform in which to develop a display library. To test the utility of the system for mapping antibody epitopes, a reverse genetics approach was adopted whereby known antibody-binding residues were replaced with alanine and the resulting mutants subjected to antibody binding studies. The mutated residues, G530A and D535A, are critical for binding of MAb AR3A to E2 (Law *et al.*, 2008). As expected $E2_{715}HA_{TM}$ carrying these mutations were no longer able to bind AR3A (Figure 2-16 and 2-17) demonstrating that antibody epitopes could be mapped by isolating and identifying loss-of-binding mutants. Crucially for mapping studies, $E2_{715}HA_{TM}$ (H77) expressed in *D. melanogaster* S2 cells was correctly folded, as demonstrated by the binding of the conformation-sensitive MAb H53 to all protein constructs (Figure 2-15).

The random mutant library was generated in the $E2_{715}HA_{TM}$ (H77) background and termed the H77 library. The EZClone random mutagenesis method did have the potential to introduce mutations into the HA_{TM} tag, and indeed sequencing analysis showed this to be the case (Table 2-4). It was not considered to be too detrimental to the overall function of the library, however, due to the small size of the tag (37 amino acids) relative to the E2 ectodomain (332 amino acids). Any mutations within the HA_{TM} tag that knocked out cell surface attachment or affected the fold of E2 would be unable to bind the pooled conformation-sensitive MAbs and would simply be removed from any further analysis. Three mutations within the HA_{TM} tag were found in sequences sorted from the H77 library; Y720F, Y720H and Y725H (Table 2-4). These mutations were all found in sequences containing additional mutations within the E2 ectodomain (Table 2-5) which could account for the loss-of-binding phenotype.

The H77 library was not expressed in the *D. melanogaster* S2 cells as strongly as the $E2_{715}HA_{TM}$ (H77) construct. AP33 binding to both cell types showed 62.41% of $E2_{715}HA_{TM}$ (H77)-expressing cells staining positive compared to just 34.94% of H77 library-expressing cells (Figure 2-15 and 2-18). This may be due to differences in the transfection efficiency when generating stable cell lines. This co-transfection method results in arrays of $E2_{715}HA_{TM}$ of between 500 and 1000 repeats inserted into the *Drosophila* genome (Anon., 2003). The exact number of repeats is difficult to control and therefore levels of protein expression may vary. The nature of the co-transfection method may also result in cells becoming transfected with pCoBlast only. These cells are blasticidin resistant but do not express the

protein of interest and therefore have a growth advantage over those cells expressing E2. A greater proportion of pCoBlast-only cells may have been generated during the H77 library transfection resulting in a higher number of cells unable to express E2. This phenomenon has been highlighted previously within the *Drosophila* expression system and may be overcome with the use of a single plasmid containing both the selection marker and the gene of interest. Such a plasmid utilising puromycin resistance has been generated and transfection with this single plasmid results in similar levels of protein expression as co-transfection with two separate plasmids (Iwaki & Castellino, 2008). Slightly reduced antibody binding to the H77 library would also be expected due to the presence of mutations within the antibody-binding epitope. Despite the lower levels of E2 expression, it is important to note that the level of AP33 and H53 binding to the H77 library was similar (Figure 2-18) suggesting that the expressed protein was correctly folded.

The ability to isolate a population of interest based on antibody binding was tested with the H77 library and an artificially mixed population of cells expressing wild-type and mutant (D535A) proteins. The population of interest was characterised as staining positive for the pooled MAbs (H53 and AR2A) but negative for the test MAb (AR3A) and was positioned in the top left-hand quadrant of the dot-plot (Figure 2-21). As expected, the percentage of cells falling within this region decreased as the number of $E2_{715}$D535A mutant cells decreased (Figure 2-22b). This reduction did not correlate precisely with the changing ratio of wild-type to mutant cells, possibly due to false-positive staining of some cells. The overall trend however was encouraging and strongly suggested that a minor population of cells could be sorted on the basis

of antibody binding. Sorting of the H77 library-expressing cells (and the artificially mixed populations) resulted in only a small percentage of cells in the region of interest (Figure 2-22a). This was disappointing but may be improved by small alterations to the experimental protocol. Firstly the MAbs used were directly labelled with Alexa fluorophore conjugates. Comparison of H77 library staining with unlabelled and labelled H53 suggests that MAb labelling is <100% (Figure 2-18 vs. 2-19), resulting in false negatives as cells expressing protein bind unlabelled H53 but do not fluoresce. The labelling process is sensitive to antibody concentration and requires the use of very pure MAb. Labelling may be improved by using a MAb preparation at a higher concentration or labelling the same preparation twice to increase the chance of each MAb molecule carrying at least one fluorescent dye molecule. Improving the labelling of these antibodies could increase the percentage of cells stained and thereby greatly enhance the sensitivity of the assay to isolate loss-of-binding mutants. Secondly, it is clear that stable transfection of *D. melanogaster* S2 cells does not result in 100% protein expression. The percentage of cells expressing protein ranges from ~35% for the H77 library to ~60% for $E2_{715}HA_{TM}$ (H77). Therefore there will always be a considerable percentage of cells which cannot be used for antibody mapping, but which may have a selective growth advantage. It would therefore be extremely advantageous to enrich these stable cell lines for cells which do express E2. This could be done prior to mapping studies by conventional fluorescence activated cell sorting using single MAb staining or by a column-based enrichment method, such as that used by Sabo *et al*. (2010) to enrich yeast cells. For example, cells could be labelled with a pool of human MAbs (i.e.

AR2A, 1:7) and mixed with Anti-IgG microbeads (human) (Miltenyi Biotech, Bisley, UK) before passing through a magnetic column. Cells expressing E2 would be enriched on the column as the remaining, unlabelled cells pass through. By using a pool of conformation sensitive MAbs the population could also be enriched for cells expressing correctly folded E2 only. Alternatively, introduction of a GFP gene into the pMT plasmid would enable cells transfected with pMT, and therefore the gene of interest, to be easily identified. Cells expressing GFP could be sorted by flow cytometry prior to mapping to enrich for E2-expressing cells. Alternatively, cells could be sorted based on GFP expression and antibody binding simultaneously. Due to time constraints and difficulties keeping sorted cells alive, it was not possible to do large scale enrichment of E2-expressing cells. However, the results of preliminary experiments (Figure 2-23) suggest that enrichment for a specific population of cells is possible. Such an enrichment step could greatly improve the library as removing large numbers of non-expressing cells and mis-folded protein would make mapping quicker, easier and reduce the amount of reagents, particularly antibody, required. It was also observed that populations of cells binding antibody did not form a distinct group from non-antibody binders (Figure 2-22); all dot plots contained a continuous populations of cells. This is probably due to individual cells expressing different amounts of protein on the cell surface, possibly due to differences in the size of the repeat arrays inserted into the *Drosophila* genome. It has also been postulated that the position of the repeats within the host cell chromosome may be crucial, as DNA can be integrated at more or less transcriptionally active sites. Alternatively, the percentage of repeats that are transcriptionally active may vary due to

rearrangement of the DNA during chromosomal integration (Deml *et al.*, 1999). As a result a range of antibody binding profiles, from very low to high, may be observed for cells expressing the same protein. This situation is complicated further by the presence of cells transfected with the pCoBlast selection plasmid only. Pre-enrichment of the cultures for protein expression would generate more distinct populations as non-expressers and very low expressers are removed.

One of the biggest problems encountered during the development of the insect cell display library was the maintenance and expansion of sorted cells. *D. melanogaster* S2 cells are very sensitive to growth at low densities and will not survive below a density of 5×10^5 cells per ml. Initially it was hoped that by sorting cells into a small volume of media, the cells would be within the density threshold for survival. Unfortunately due to the low percentage of cells falling within the region of interest, the cells could not be maintained. This might be overcome by staining a larger number of cells. All experiments were carried out with staining and sorting of 1×10^6 cells with approximately 2% of H77 library-expressing cells staining within the population of interest (Figure 2-22a), resulting in a sorted population of around 20,000 cells (200 cells/µl). This number could be increased by sorting a larger number of cells and ultimately, with large enough numbers of cells, it may be possible to remove the feeder culture. This would require a minimum of 50,000 cells to be sorted into a culture vessel containing 100 µl media. Achieving such numbers would be greatly facilitated by pre-enrichment of the H77 library for cells expressing protein.

Maintaining the sorted cultures was also hampered by bacterial contamination. *D. melanogaster* S2 cells are described as being macrophage-like (Luce-Fedrow *et al.*, 2008) and therefore able to remove contaminants from their culture medium. Unfortunately due to the high degree of handling necessary to stain, sort and expand the cells, contamination was common. It was necessary to include Penicillin-Streptomycin in all wash buffers and the post-sorting culture media. Penicillin-Streptomycin was not detrimental to the growth of *D. melanogaster* S2 cells and could be removed from the culture media after a few weeks. All tissue culture handling was carried out in a Class II safety cabinet and the Beckman Coulter MoFlo Cell sorter was also housed within a Class II safety cabinet. Sterilisation of the sheath fluid by autoclaving and cleaning of the cell sorter with bleach and ethanol prior to sorting were all found to reduce the likelihood of contamination. Despite this, several sorted cultures were lost due to bacterial contamination.

An alternative to bulk sorting and enrichment for loss-of-binding mutants involves direct sequencing of the sorted cells. Thanks to advances in flow cytometry technology, it is possible to rapidly and accurately sort large numbers of cells into culture plates or PCR plates, even down to a single cell per well. This overcomes the problem of cell culture contamination or maintenance of cells at low density and allows potential antibody-binding residues to be identified more quickly. It may also facilitate the isolation of rare variants which are only present at a low number within the library. The rate of E2 amplification from single cells was between 42% and 70%. This may be, in part, due to the failure of the cell sorter to deposit a single cell into each well. This could be checked by the inclusion of primers to a

housekeeping gene during the PCR steps. Typically the human hypoxanthine ribosyl transferase (HPRT) gene is used as a control for the presence of template in PCR amplifications. In *D. melanogaster* S2 cells the ribosomal protein 49 (rp49) is commonly used (Luce-Fedrow *et al.*, 2008). Failure to amplify rp49 could be used to detect any empty wells from which E2 would not be recovered. Low PCR recovery may also be caused by cells that do not contain pMT or have not been accurately stained with the MAb pool. This could be improved by pre-enrichment of the H77 library and improved MAb labelling, as described above. As expected, all cells sorted from an artificially mixed population on the basis of lost AR3A reactivity, contained the alanine substitution at D535. Subsequent analysis of the H77 library sorted on the basis of lost AR3A reactivity identified a range of substitutions that may be involved in AR3A binding. Of these, T396Q, L399I, W420R and N423K lie within the first discontinuous epitope forming AR3 and C429R falls between the first and second discontinuous epitopes (Law *et al.*, 2008). Although these residues have not previously been shown to play a direct role in AR3A binding, they lie within the region of E2 recognised by the AR3 group of MAbs. Residues W420 and N423 are found within the CD81 binding site at amino acids 412-423 (Owsianka *et al.*, 2001), consistent with the known function of AR3A as an inhibitor of CD81 binding (Law *et al.*, 2008). Alanine replacement of residues 420 and 423 suggested that they were not involved directly in AR3A binding (Law *et al.*, 2008); however the introduction of larger, charged amino acid residues (arginine and lysine respectively) may cause a localised conformational change that could explain the observed loss of binding. It is also interesting to note that N423 is an N-linked glycosylation

site (Helle *et al.*, 2007). Removal of this glycan has been shown to reduce infectivity in the HCVpp system (Goffard *et al.*, 2005); however in the HCVcc assay system, its removal does not affect infectivity but does enhance neutralisation by polyclonal antibodies and the MAb 3/11 (Helle *et al.*, 2010). MAb 3/11 recognises a linear epitope (Flint *et al.*, 1999a) which is exposed by removal of the glycan at residue 423. Therefore removal of this glycan might be predicted to enhance antibody binding; however, binding of conformation sensitive MAbs, such as AR3A, is more dependent on the fold of the protein which does not appear to be affected by this glycan knock-out (Figure 2-22a) (Goffard *et al.*, 2005).

Two of the mutations result in the replacement of a cysteine residue; C429R and C644R. This would knock out the first and eighth disulphide bond respectively (Krey *et al.*, 2010). This is perhaps surprising given the importance of the disulphide bonds in maintaining protein fold. However E2 truncations up to residue 661 have been expressed and shown to bind H53 (Flint *et al.*, 1999a) demonstrating that the removal of glycans and even the generation of a 'spare' cysteine does not universally generate mis-folded protein but rather a heterogeneous mix of correctly and incorrectly folded E2.

The L399I mutation and the C644R mutation appear at high frequency (>80%) within the recovered sequences. This may be due to over-representation of these sequences within the library, resulting in their recovery at higher than expected frequencies regardless of antibody binding. This may occur if the library is not as diverse as predicted. Attempts were made to assess the diversity of the library and steps were made to ensure that diversity was maintained. This was based on the assumption that at all points the

maximum number of variants would be generated and, more importantly, maintained. There are however, several points at which diversity might have been reduced, if, for example, transfection efficiency was unusually low or resurrection of the glycerol stock was insufficient to represent the full library. This may explain why mutations at G530 and D535 were not isolated during H77 library screening. It was predicted that these variants would appear strongly in AR3A loss-of-binding mutants as they are known to be critical for AR3A binding (Figure 2-16 and 2-17) (Law *et al.*, 2008). It is also possible however, that these residues were mutated to amino acids other than alanine that do not have such a strong impact on AR3A binding. The size of the library was calculated based on mutation of each amino acid at least once; however, a library expressing multiple different amino acids at each residue would give greater mapping capabilities as not all amino acid substitutions will impact antibody binding to the same extent. For example, a conservative mutation of alanine to valine might not change the shape of an epitope as much as a non-conservative mutation to aspartate, resulting in variable antibody binding to epitopes containing aspartate and valine respectively. Therefore a library containing the largest number of variants at a single residue might pick out rare sequences with reduced antibody binding. However, a larger library would be needed to accommodate such variation within the E2 ectodomain, although it was encouraging to note the isolation of sequences containing two different mutations at the same residue, Y720F and Y720H (Table 2-4). Library statistics are listed in Table 2-3. Input DNA of 300 ng was chosen as this would give 1182 variants differing from the parent by a single amino acid. This is theoretically sufficient to allow for amino acid substitution at each

position of the 331 residue protein. However, the statistics also suggest that the library will contain 1790 distinct amino acid substitutions that are present in at least one sequence. This is not sufficient to allow for substitution to every amino acid at each residue. Recalculations using PEDEL-AA indicate that a larger library of 10^7, plus an increased mean mutation frequency from 5 to 16 nucleotide substitutions per daughter sequence would be needed to produce a library containing 6312 distinct amino acid substitutions that are present in at least one sequence. Theoretically this should provide sufficient coverage for every possible substitution to occur at each residue, if 331 residues are mutated from their original amino acid to each of the 19 possible alternatives. However, this is the upper limit of mutations achievable using the Genemorph II EZClone Domain Mutagenesis Kit. Increasing the mutation frequency reduces the number of sequences containing a single substitution, making identification of individual residues involved in antibody binding more time consuming, and increases the number of sequences containing in-frame deletions and stop codons. It may not be possible to achieve all amino acid substitutions at a given residue due to the complexity of the codon mutation that is required (i.e. mutating a single nucleotide vs substituting all nucleotides within the codon). PEDEL-AA uses a generic Poisson model of sequence mutation that assumes that all substitutions are equally likely (Firth & Patrick, 2005) however polymerases favour some mutations over others. The Genemorph II EZClone Domain Mutagenesis Kit is designed to produce substitutions in an unbiased manner; nonetheless, some substitutions may occur more frequently than others. Equally, some substitutions may abrogate protein processing and as such may not be tolerated. Therefore a fine balance is

required in terms of the mutation frequency, such that sufficient diversity is achieved whilst maintaining the effective size of the library.

Finally it is worth considering that the degeneracy of the genetic code may result in some amino acids appearing more frequently than others. Codon usage bias, the unequal use of codons encoding the same amino acid, is a trait shared by *Drosophila* species (Powell & Moriyama, 1997; Vicario *et al.*, 2007). In *D. melanogaster* the trend is towards C/G-ending codons (Vicario *et al.*, 2007) and the most common codons correlate well with the levels of isoaccepting tRNAs (Powell & Moriyama, 1997). Such bias may result in the reduced expression of proteins containing suboptimal codons, further skewing the apparent diversity of the library. This could be overcome by optimising codon usage for expression in *D. melanogaster* S2 cells, a technique that has been used to improve scFv antibody production (Gilmartin *et al.*, 2012) and was also used in the generation of the 3D model of E2 (Thomas Krey, personal communication). This may also improve expression levels of the other $E2_{715}HA_{TM}$ constructs. Codon optimised proteins should be synthesised more rapidly within the cell leading to higher overall expression levels. Lack of codon optimisation may help to explain why a maximum of ~60% of cells from any cell line were able to express protein.

The majority of recovered single cells expressing E2 contained only one or two mutations (Table 2-5). Ideally most sequences should contain a single mutation as this facilitates the process of identifying the critical residues. Interrogating the role played by a specific residue can be achieved through reverse genetics and mapping studies with the MAb of interest. However, this process becomes more complicated and more time-consuming as the number of

residues increases. Therefore the recovery of E2 containing more than two mutations was disappointing, although these were the minority (12%) of sequences. The presence of one or two mutations in 81% of sequences was encouraging and suggested that the mutation frequency was appropriate for this library.

The preliminary data suggested the involvement of some unexpected residues in AR3A binding to E2. Most of the residues identified do not fall within regions of E2 previously shown to be involved in AR3A binding. The potential identification of previously unrecognised residues highlights one of the advantages of using a library expressing full-length E2. The whole protein can be interrogated simultaneously and no prior knowledge of potential antibody-binding epitopes is necessary. A recent study of the CBH-2 epitope initially identified residues G530 and D535 that are also involved in CD81-E2 binding, as critical for antibody binding. However, further alanine mutagenesis of regions not involved in CD81 binding highlighted the importance of residues L640, A642 and A643 (Keck *et al.*, 2008b). The use of a display library encompassing the whole E2 glycoprotein would abrogate the need for lengthy alanine-scanning mutagenesis experiments, which are reliant upon prior knowledge of possible epitopes, and speed up the process of epitope identification. It will be interesting to see if reverse genetics studies of these residues confirm their role in AR3A binding. Collectively this data provides strong evidence that an E2 random mutant display library expressed in *D. melanogaster* S2 insect cells can be used to map antibody residues. A number of steps could be taken to improve the library and ultimately facilitate the use of this library in high-throughput screening of antibody-binding epitopes.

3 Isolation and characterisation of monoclonal antibodies isolated from HCV-infected patients

3.1 Introduction

3.1.1 Antibody isotypes and function

Antibodies or Igs are globular glycoproteins produced by B cells in response to a particular antigen. Five classes, or isotypes, of antibody are found in humans; IgA, IgD, IgE IgM and IgG. Each antibody isotype has different properties with respect to size and function. IgA is predominantly associated with mucosal immunity and can account for up to 50% of the Ig found at mucosal membranes and in secretions. IgE is associated with the allergic immune response and hypersensitivity and accounts for <0.01% of serum Igs. IgD is also found at very low levels in the serum (<0.5%) and is thought to act primarily as a signalling receptor on the surface of B cells. IgM is the main antibody produced during the primary immune response. Monomeric IgM is expressed on the surface of B cells and secreted as a multimer, typically a pentamer. Although usually of low affinity, the pentameric nature of IgM results in high avidity binding to antigen (Schroeder & Cavacini, 2010).

IgG is the most abundant, accounting for ~75% of circulating, serum Igs and has the longest serum half-life. It exists as four subclasses, IgG_1, IgG_2, IgG_3 and IgG_4. IgG_1 is the most abundant and, with the exception of IgG_4, all subclasses are able to carry out complement fixation and opsonise targets (Schroeder & Cavacini, 2010). The different subclasses also differ in their ability to bind cognate receptors via the Ig Fc domain (see section 3.1.2 and Figure 3-1). These receptors, termed FcγR, are expressed on most leukocytes

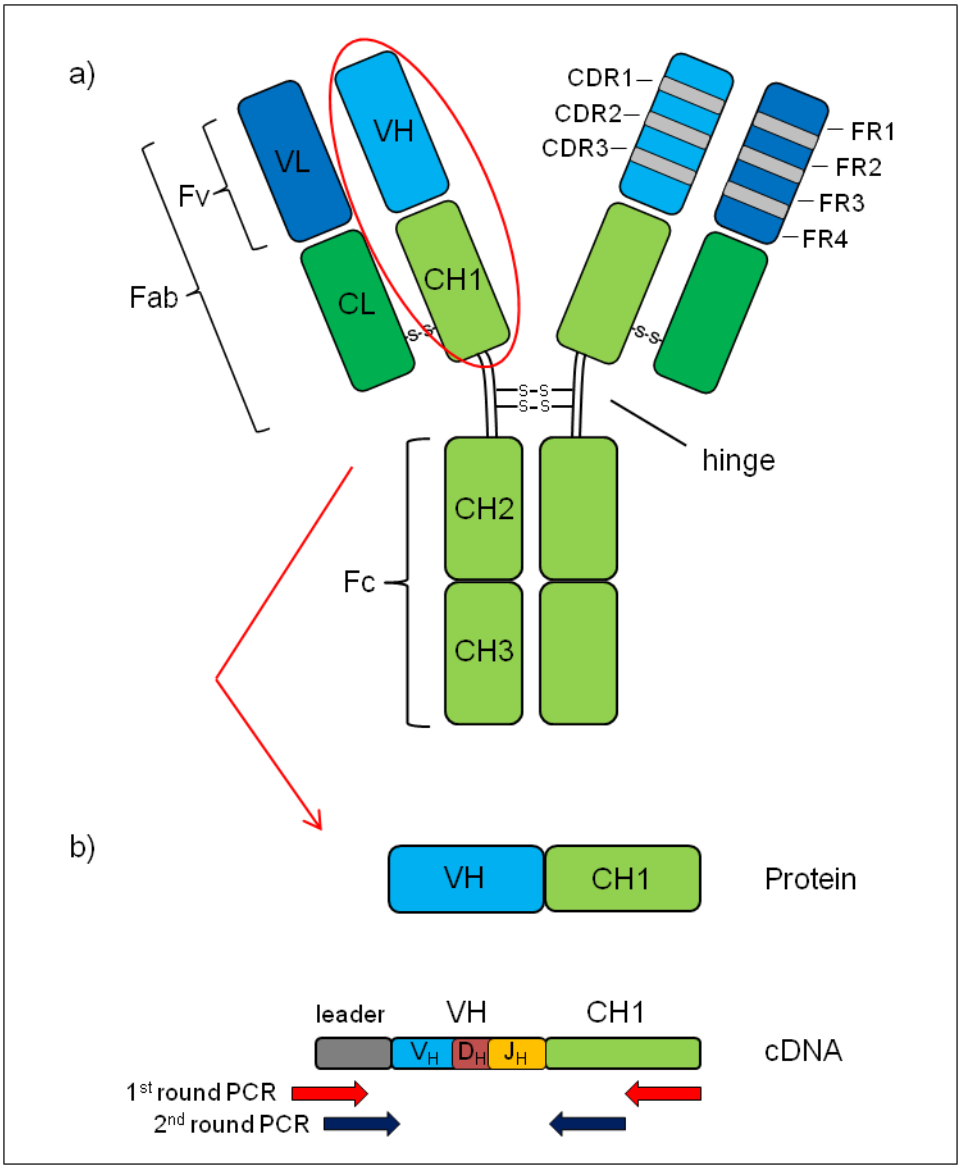

Figure 3-1. Schematic representation of an immunoglobulin molecule. a) The molecule is composed of two heavy chains and two light chains, linked by disulphide bonds (s-s). The constant domains are shown in green and the variable domains shown in blue. CH2 and CH3 make up the fragment crystallisable (Fc) region. The variable domains (VL and VH) plus CL and CH1 make up the antigen-binding fragment (Fab). The Fc portion and the Fab region are connected by a flexible hinge region. The framework regions (FR) and complementarity determining regions (CDR) within Fab are shown. See text for references. b) PCR primer annealing sites within the Ig molecule. The cDNA encoding the heavy and kappa chains of the Fab fragment (red circle = heavy) is the target of the Ig-specific primers. First round sense primer anneals to the leader sequence, the antisense primer to the constant domain. Second round primers anneal internal to the first round primers. cDNA is shown broken down into its separate gene components encoded by the germline genes (listed in Figure 3-2).

and can be divided into three types; FcγRI, FcγRII and FcγRIII (Jiang *et al.*, 2011). IgG_1 and IgG_3 are able to bind all three receptors, IgG_4 binds FcγRII and FcγRIII only, and IgG_2 binds to FcγRII (Schroeder & Cavacini, 2010). This has implications for the ability of the antibody to mediate effector functions, such as ADCC, and is an important consideration for therapeutic antibody or vaccine design (Jiang *et al.*, 2011).

3.1.2 Structure of the IgG molecule

IgG exists as a monomer composed of four polypeptide chains; two light chains and two heavy chains linked by disulphide bonds (Figure 3-1) (Tonegawa, 1983). The type of heavy chain determines the antibody class; therefore IgG has a γ-type chain, IgM has a μ-type chain, IgA has an α-type chain, IgD has a δ-type chain and IgE has an ε-type chain. The light chain within an Ig molecule can be either a kappa (κ)-type or a lambda (λ)-type chain (Eales, 2001). Each heavy chain contains three constant domains (CH1, CH2 and CH3) and one variable (VH) domain (Figure 3-1). The light chain contains a single constant domain (CL) and a single variable (VL) domain. The CH2 and CH3 domains form the fragment crystallisable (Fc) portion of the Ig, which binds FcγR and mediates the antibody effector functions. The variable domain of the heavy and light chain, plus CL and CH1, forms the Fab fragment, responsible for antigen-binding. The Fab and Fc portions of the Ig molecule are separated by a hinge region that enables slight movement of the Fab region, relative to the Fc region, and facilitates antigen binding (Figure 3-1) (Eales, 2001; Schroeder & Cavacini, 2010). The variable domains may also be referred to as the Fv fragment and are further divided into four framework

regions (FR) and three complementarity determining regions (CDR) (Figure 3-1). The CDRs are regions of hypervariability that contribute to the specificity of the antibody for its antigen and enhance the affinity of the interaction (Tonegawa, 1983).

3.1.3 Generation of antibody diversity

3.1.3.1 Recombination of germline genes

Ig light and heavy chains are encoded by germline genes that undergo recombination and somatic hypermutation to produce the great diversity of antibodies generated by a healthy human over their lifetime. The heavy chain locus is located on chromosome 14q32.33 and encodes 39 functional variable (V_H) genes that are divided into 7 families (Matsuda *et al.*, 1998). Located downstream of the V_H genes are 27 D (or diversity) genes and six J (or junction) genes (Matsuda *et al.*, 1998; Schroeder & Cavacini, 2010) (Figure 3-2a). The constant (C) genes, encoding each of the Ig isotypes, are located downstream of the J_H genes. During heavy chain recombination, D_H and J_H genes are the first to recombine, followed by recombination with a single V_H gene to form the complete VDJ segment that comprises the variable heavy (VH) domain (Tonegawa, 1983). The V_H gene encodes CDR1 and CDR2 and the amino terminus of CDR3, with the remainder of CDR3 encoded by the D_H and J_H genes. Combinatorial diversity generated by rearrangement of VDJ genes is estimated at 3.8 $x10^4$ combinations of genes (Schroeder & Cavacini, 2010).

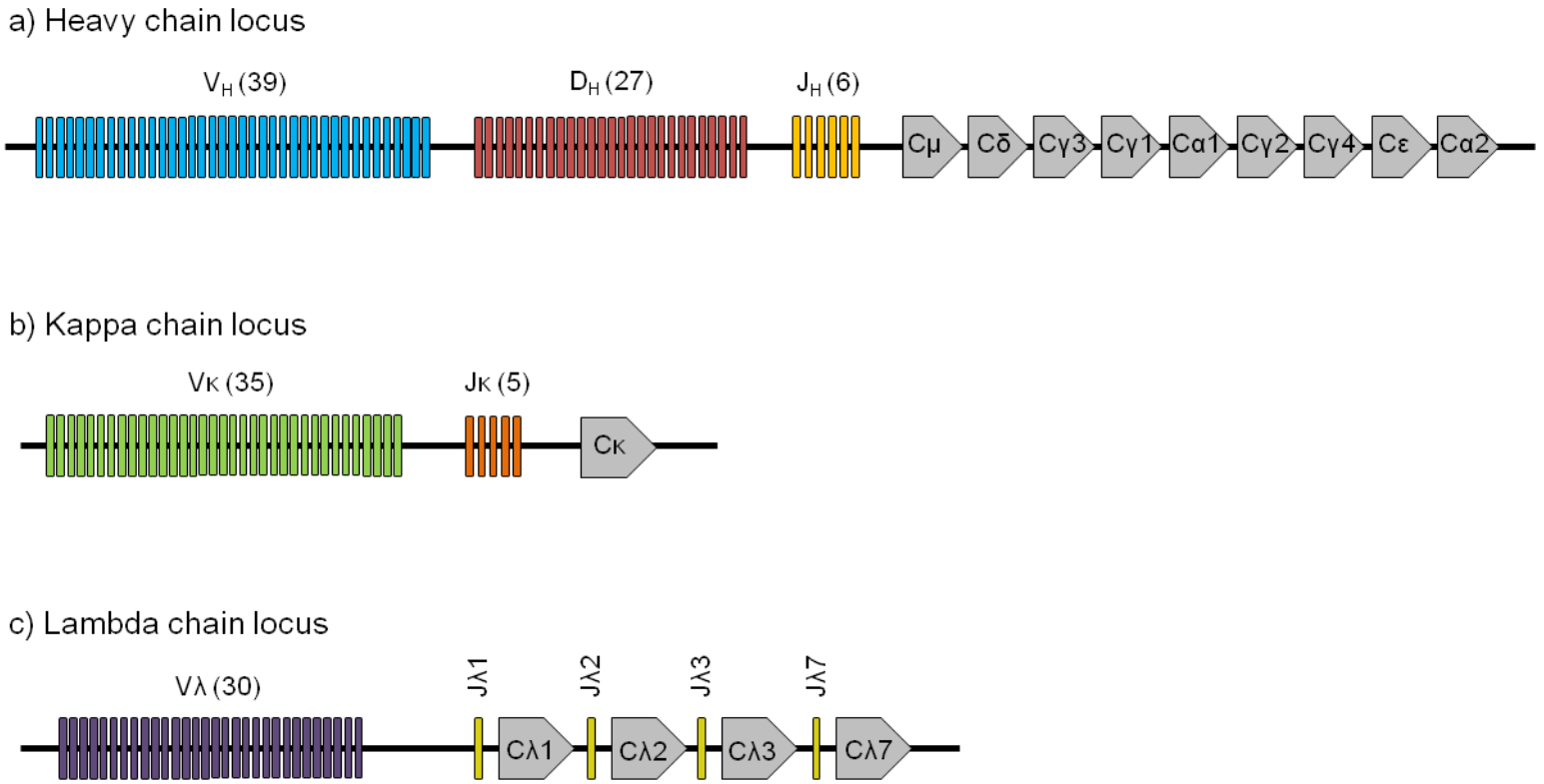

Figure 3-2. Schematic representation of the immunoglobulin gene loci. a) The heavy chain locus contains V, D and J genes that undergo VDJ recombination to form the functional variable domain. Initially recombination with the Cμ gene generates IgM. During B cell maturation, class switch recombination with any of the remaining C genes results in the formation of different isotypes. b) The kappa chain locus contains V and J genes that recombine to form the variable domain. This undergoes recombination with the single Cκ gene to form a functional light chain. C) The lambda chain locus contains Vλ genes that recombine with any of four Jλ genes, each associated with its own Cλ gene. See text for references.

In contrast to the VH domain, the kappa variable domain is composed of V and J genes only. The kappa chain locus is located on chromosome 2p11.2 and encodes an estimated 35 functional Vκ genes divided into 6 families, plus five Jκ genes (Barbie & Lefranc, 1998) (Figure 3-2b). The Vκ and Jκ genes recombine to form the complete VJ segment comprising the variable region, prior to recombination with the single Cκ gene (Tonegawa, 1983). The lambda variable domain is also composed of V and J genes only. The lambda chain locus is located on chromosome 22q11.2 (Kawasaki *et al.*, 1997) and encodes 30 functional Vλ genes divided into 11 families (Pallares *et al.*, 1998) plus four Jλ genes each of which is associated with its own Cλ gene (Schroeder & Cavacini, 2010) (Figure 3-2c). Genetic recombination takes place such that the Jλ-Cλ genes recombine with a Vλ gene (Tonegawa, 1983). In both kappa and lambda variable domains, CDR1 and CDR2 are encoded by the V gene whilst CDR3 is encoded by both the V and J genes (Schroeder & Cavacini, 2010). Approximately 175 kappa gene combinations and 120 lambda gene combinations can be generated by recombination of the germline sequences.

VDJ or VJ recombination is highly controlled to ensure that genes recombine in the correct order and that the same genes do not recombine with themselves. In heavy chain genes D-J recombination occurs first, followed by V-DJ recombination. The process is controlled by recombination signal sequences (RSS) located at the recombination sites between genes. These RSSs are non-coding DNA sequences consisting of a conserved heptamer sequence, a spacer sequence of either 12 or 23-bp, and finally a conserved nonamer sequence. An RSS is found at the 3' end of all V genes and at the 5' end all J genes. D genes are flanked at the 5' and 3' end by an RSS. In heavy chain genes, the V and J

genes have a 23-bp spacer-containing RSS whereas D genes are flanked by a 12-bp spacer. In light chain genes, the Vκ genes contain a 12-bp spacer and Jκ genes contain a 23-bp spacer; Vλ genes contain a 23-bp spacer and Jλ genes contain a 12-bp spacer. Functional recombinations occur only between a sequence flanked by a 12-bp spacer and another sequence flanked by a 23-bp spacer. Therefore, non-functional rearrangements between V-V genes or J-J genes are prevented by this so-called '12/23 rule' (Dudley *et al.*, 2005; Tonegawa, 1983). Recombination is initiated by binding of the recombination-activating gene (RAG)-1 and RAG-2 to a 12-bp RSS, followed by binding to a 23-bp RSS, bringing the two different gene segments into close contact in a synaptic complex. RAG-1 and RAG-2 induce DNA single-strand breaks that result in the formation of a hairpin loop at the end of each coding gene segment. The hairpin loops are resolved by further single-strand nicks carried out by the non-homologous end joining (NHEJ) proteins, Artemis and DNA-PKcs. This results in a 3' DNA overhang that can be modified by the action of exonucleases, DNA polymerases or terminal deoxynucleotidyl transferase (TdT) to add or remove germline nucleotides (Dudley *et al.*, 2005). This introduces mutations into CDR3 and leads to great variation in its length. The introduction of N nucleotides by the action of TdT is thought to increase the diversity of the VH domain to around 10^7 different sequences (Tiller *et al.*, 2008). A variety of host NHEJ proteins finally repair the DNA breaks generating the complete variable region (Dudley *et al.*, 2005).

3.1.3.2 Somatic hypermutation and class switch recombination

Additional genetic diversity is introduced through the process of somatic hypermutation (SHM) of the variable domain upon exposure to antigen. SHM introduces mutations at rate of 10^{-3} mutations per base pair per cell division via the activity of activation-induced cytidine deaminase (AID) (Schroeder & Cavacini, 2010). AID deaminates cytidine residues within single stranded DNA resulting in the conversion of cytidine to uracil. This uracil may be removed by the activity of uracil-DNA glycosylase resulting in an abasic site that is replaced by any nucleotide via the base excision repair mechanism. Alternatively the mismatch repair machinery may introduce additional mutations proximal to the initial C-U mutation. SHM is restricted to the variable region and the constant domain is protected from mutation to ensure antibody function is maintained (reviewed in (Odegard & Schatz, 2006). SHM takes place in B cell germinal centres and can involve several rounds of clonal expansion, exposure to antigen presented on follicular dendritic cells and selection of high-affinity variants (McHeyzer-Williams *et al.*, 2012). B cell germinal centres are also the site of class-switch recombination (CSR), the process by which affinity-matured, antigen-specific variable regions recombine with the constant region genes (Cγ, Cα and Cε) to produce different antibody isotypes (Figure 3-2a). Immature B cells initially express IgM (and also possibly IgD) but may undergo CSR during development to produce the different Ig isotypes. This process also requires the action of AID at specific switch signals located between the constant region genes (Dudley *et al.*, 2005). Maturation of the B cell response in germinal centres results in the production of long-lived plasma cells secreting high-affinity antibody as well the

production of long-lived, high-affinity memory B cells. Germinal centres also play a crucial role in the removal of self-reactive B cells and non-functional Igs (McHeyzer-Williams *et al.*, 2012). IgG^+ memory B cells are thought to reside close to germinal centres and upon re-exposure to antigen, take up and present antigen to $CD4^+$ T helper cells via MHC class II. Antigen-specific memory T helper cell interactions are required for clonal expansion and differentiation of memory B cells into plasma cells (Aiba *et al.*, 2010), the precise mechanisms of which are poorly understood but may involve upregulation of cytokine signals (McHeyzer-Williams *et al.*, 2012).

As a result of the recombination of germline genes, the addition or removal of germline nucleotides and SHM of the variable domain, as well as the recombination of heavy chains with either kappa or lambda-type chains, it has been estimated that the antibody repertoire may exceed 10^{16} different Igs (Schroeder & Cavacini, 2010).

3.1.4 Methods to isolate monoclonal antibodies

The use of polyclonal Igs for prophylactic and therapeutic purposes was developed at the start of the 20^{th} century and has proven effective in the treatment or prevention of respiratory bacterial infections, tetanus, hepatitis B virus and rabies virus (Keller & Stiehm, 2000). However, the widespread use of poly-Igs has been hampered by limited availability, batch variability and the infection risk associated with the use of human blood products. MAbs represent a more consistent and safer option but to date only one MAb therapy has been developed for use against an infectious disease, for the treatment of

respiratory syncytial virus infection (Johnson *et al.*, 1997). Nonetheless, several methods have been established to develop MAb therapies against a variety of different targets, the limitations of which will be discussed below.

3.1.4.1 Murine antibodies

Hybridoma technology was one of the earliest methods developed for MAb production. The technique fuses splenocytes from an immunised mouse donor with myeloma cells to generate a continuous cell line secreting antibody of a defined specificity (Kohler & Milstein, 1975). The first stage in the generation of hybridomas is the immunisation of mice, which may require repeat boosting in order to stimulate a strong immune response. B cells are then removed from the spleen and fused with a myeloma cell line. This is followed by several weeks of selection with hypoxanthine, aminopterin and thymidine (HAT) which selectively removes unfused myeloma cells. Fused hybridomas will grow in the presence of HAT although some loss may occur as unstable hybridomas regress. Successful hybridomas must then be carefully selected for antibody specificity to eliminate non-specific hybridomas and in some cases it is necessary to re-clone hybridomas that have formed by fusion of more than one B cell to the myeloma cell line (Nelson *et al.*, 2000). As such, the process of generating hybridomas is time consuming and can be technically challenging with low fusion rates resulting in few successful hybridomas. Many murine MAbs are immunogenic when administered to patients making them inappropriate for use as therapeutics (Hwang & Foote, 2005). They also show reduced effector function in humans due to differences in the Fc region (Li *et al.*, 2006). This has led to the humanisation of murine antibodies by the

insertion of murine CDRs into a human antibody scaffold (Jones *et al.*, 1986) or by generation of a chimeric MAb containing a murine Fab region and a human Fc region (Boulianne *et al.*, 1984; Morrison *et al.*, 1984). Both approaches reduce the immunogenicity of the MAb (Hwang & Foote, 2005) and have been successfully used to generate a number of effective antibody therapies in clinical use today (reviewed in (Jiang *et al.*, 2011). Such engineering can however, be time-consuming. Transgenic mice bearing human germline Ig genes have been developed, although early methods were unable to recapitulate the entire human VDJ gene diversity (Green *et al.*, 1994; Lonberg *et al.*, 1994). The use of chromosomal transfer techniques has enabled the introduction of the complete IgH and Igκ gene loci into mice lacking endogenous Ig genes; however, low levels of IgA production and unstable Igκ locus transfer were reported (Tomizuka *et al.*, 2000). Those human antibodies generated in transgenic mice do have low immunogenicity (Lonberg, 2005); however production of specific MAbs following immunisation still requires hybridoma generation, which is both time-consuming and challenging. It is also important to consider that, despite the presence of human IgH and Igκ loci, the process of B cell development and maturation takes place in a murine genetic background. This may alter the spectrum of antibodies derived from transgenic mice compared to humans (Lonberg, 2005). Chimeric and humanised MAbs will also retain antigen-binding specificities from mice. The antibody response in mice is unlikely to represent the complexity of the human immune response and therefore methods which directly interrogate human antibody responses to natural infection are likely to be more informative than vaccine induced antibody responses in experimental animals. Indeed within

the field of HCV research it has been noted that mice are limited in their ability to generate conformation-dependent antibodies against the HCV envelope glycoprotein. Conformation-dependent antibodies tend to show greater cross-neutralising activity and therefore present a better therapeutic target (reviewed in Edwards *et al.*, 2012); however, murine anti-HCV MAbs identified to-date predominantly recognise linear epitopes (Clayton *et al.*, 2002; Owsianka *et al.*, 2001; Owsianka *et al.*, 2005; Tarr *et al.*, 2006).

3.1.4.2 Human antibodies

Phage display is widely used for the expression of protein and peptide fragments in *E. coli* (see section 2.1.1). Typically in antibody display, the VH and VL genes are joined by a flexible glycine-serine linker resulting in expression of a scFv on the surface of a filamentous phage (McCafferty *et al.*, 1990). Alternatively Fab fragments may be displayed by anchoring the heavy chain to the phage coat protein and secreting the light chain into the periplasm, where the two chains assemble to form the Fab fragment (Hoogenboom *et al.*, 1991). ScFvs are more commonly used as they are easier to assemble and less likely to be degraded than Fab fragments (Bradbury & Marks, 2004). Antibody display technologies are generally considered to be an improvement over hybridoma technologies as more antibody specificities can be identified from a single immune donor. The enrichment process may also facilitate the isolation of rare antibodies (Hoogenboom, 2005). Phage display can be used for the generation of recombinant libraries from immune donors, non-immune donors and synthetic fragments. Recombinant antibody libraries derived from chronically infected, immune donors have been used successfully to isolate

neutralising anti-HCV antibodies, some of which show broad neutralising activity (Allander *et al.*, 2000; Giang *et al.*, 2012; Johansson *et al.*, 2007b; Law *et al.*, 2008; Perotti *et al.*, 2008). Theoretically, non-immune libraries can be used to identify antibodies binding to any antigen (Carmen & Jermutus, 2002). Non-immune or synthetic libraries undergo *in vitro* affinity maturation via PCR mutagenesis of the CDRs or growth in mutator strains of *E. coli.* However, this can be a time-consuming process requiring iterative rounds of enrichment and selection. Alternatively targeted mutations can be introduced into regions known to enhance affinity, although this may require some prior knowledge of the immune response to the antigen (Hoogenboom, 2005). Recombinant libraries expressing scFv fragments on the surface of *S. cervisiae* cells (Boder & Wittrup, 1997) using flow cytometric sorting of high-affinity binders, rather than biopanning on an immobilised antigen, are also available. Yeast display combined with PCR mutagenesis and *in vitro* recombination has been used for the maturation of antibodies to extremely high affinity (48 fM) (Boder *et al.*, 2000), possibly due to the removal of expression biases associated with expression of mammalian genes in prokaryotic cells. The affinity of the broadly neutralising anti-HCV antibody HC-1 was successfully enhanced by enrichment of scFv mutants displayed on yeast cells (Wang *et al.*, 2011) and a number of therapeutic antibodies have been developed using the recombinant antibody technologies described above (Hoogenboom, 2005).

As discussed previously, hybridomas producing murine antibodies may not accurately mimic the human immune response and can be immunogenic in a therapeutic setting. Recently, human-murine heterohybridomas, generated by the fusion of human splenocytes to modified murine myeloma cell lines

(Dessain *et al.*, 2004) or human myeloma cell lines suitable for the generation of entirely human hybridomas (Karpas *et al.*, 2001) have been developed. These methods are able to generate stable hybridomas (Dessain *et al.*, 2004; Karpas *et al.*, 2001) and the secreted antibodies are likely to be less immunogenic than murine MAbs. However, the frequency of hybridoma generation is still low (Karpas *et al.*, 2001) and obtaining human splenocytes for fusion can be difficult, although hybridoma generation by fusion of peripheral blood mononuclear cells (PBMCs) has been reported (Li *et al.*, 2006). Neutralising anti-HCV MAbs have been isolated from PBMCs using modified hybridoma technology (Hadlock *et al.*, 2000). Antibody producing B cell lines have also been generated by immortalisation of lymphocytes with Epstein-Barr virus (EBV). EBV immortalisation preserves the characteristics of the original cell including surface and secreted Ig, enabling the production of antigen-specific MAb (Steinitz *et al.*, 1977). Recently the technique has been improved by the addition of the polyclonal activator, CpG 2006, a single-stranded DNA oligonucleotide that activates B cells via interaction with TLR-9. This increased the rate of memory B cell immortalisation from 1-2% to 30-100%, and was successfully used to isolate nAbs against the SARS coronavirus (Traggiai *et al.*, 2004). EBV immortalisation is however a technically challenging process with others reporting immortalisation rates of only 40% (Pinna *et al.*, 2009).

Recently methods have been developed to activate memory B cells *in vitro* and characterise the antibodies produced by a single B cell clone. Memory B cells can be activated *in vitro* by antigen-independent mechanisms requiring T helper cell stimulation and cytokine production (Lanzavecchia, 1983;

Lanzavecchia *et al.*, 1983). Memory B cells express a number of TLRs on their surface including TLR-2, -6, -7, -9 and -10 (Lanzavecchia *et al.*, 2006). When stimulated by TLR agonists, including CpG (Bernasconi *et al.*, 2002) and R848 (Pinna *et al.*, 2009), B cells proliferate and differentiate into antibody secreting cells (ASCs). Class-switched memory B cells can also be stimulated by interleukin (IL)-2 (Pinna *et al.*, 2009) and IL-15 (Lanzavecchia *et al.*, 2006). This activation is antigen-independent and does not require direct T-cell interactions. The polyclonal activation of memory B cells is required for plasma cell homeostasis and maintenance of serum antibody levels (Lanzavecchia *et al.*, 2006). More recently, highly proliferating, Ig-secreting B cells have been generated by transduction of germinal centre-specific proteins into memory B cells, followed by stimulation with CD40L and IL-21. This resulted in long-lived antibody producing cells that were able to undergo SHM *in vitro* (Kwakkenbos *et al.*, 2010). This may enable continued affinity maturation, generating improved MAbs without the need for targeted affinity engineering. Studying the memory B cell pool following polyclonal activation abrogates the need for specific antigen boosting, facilitating the identification of antibodies produced against many different antigens. The short-lived nature of most ASCs (10-40 days) (Lanzavecchia *et al.*, 2006) makes identification of B cell specificities problematic, particularly for those diseases that are difficult to diagnose during acute infection, such as HCV. Memory B cells provide a record of all previous antigenic exposure over a patient's lifetime; therefore interrogating the memory B cell pool can shed light on the antibody response to past infections as well as chronic infections.

Single memory B cell isolation has been greatly facilitated by developments in flow cytometry technology. Following isolation, single B cells are stimulated to differentiate into ASCs producing large amounts of antibody and therefore mRNA transcript. The antibodies can then be identified directly by reverse transcription and amplification of the rearranged Ig genes (Wang & Stollar, 2000; Wardemann *et al.*, 2003). Reverse transcription is typically carried out in the same vessel as used for culturing the B cells. This reduces the potential for cross-contamination or loss of RNA. Reverse transcription can be carried out with class-specific primers or universal primers binding to the constant region to isolate any antibody isotype. Amplification of the Ig genes is achieved by nested PCR. In the first round, the forward primers bind the IgH, Igκ or Igλ leader sequence and the reverse primer binds the constant region. In the second round, primers binding within the variable region increase the specificity of the amplified sequences (Figure 3-1b) (Koelsch *et al.*, 2007; Tiller *et al.*, 2008; Wang & Stollar, 2000; Wardemann *et al.*, 2003). Amplification of Ig genes directly from patient B cells has been used for the isolation of high-affinity and broadly reactive nAbs against influenza A virus M2 protein (Grandea *et al.*, 2010) and HA (Corti *et al.*, 2011); isolation of high-affinity, neutralising anti-gp120 HIV antibodies (Scheid *et al.*, 2009; Scheid *et al.*, 2011; Wu *et al.*, 2011); characterisation of the specific antibody response to HIV gp41 (Pietzsch *et al.*, 2010), and to study the antibody response to the seasonal influenza vaccine (Wrammert *et al.*, 2008). The simultaneous amplification of heavy and light chain genes from a single B cell culture overcomes the need for *in vitro* recombination and cloning of multiple sequences, thus avoiding unproductive gene recombination. By looking at

memory B cell populations, it is possible to interrogate naturally occurring human antibody responses, including past exposure to a variety of antigens, without the need for antigen boost. The method is not limited by the number of B cells that can be isolated and rare or infrequent B cells can be identified (Corti *et al.*, 2011). The time taken from single B cell culture to heavy and light chain amplification is shorter than that required for hybridoma production (Nelson *et al.*, 2000) and reported antibody recovery rates are good (Scheid *et al.*, 2009; Scheid *et al.*, 2011; Wrammert *et al.*, 2008). The small sample volumes required both for cell culture and RT-PCR amplification make the process particularly amenable to a high-throughput platform. This technique can identify different antibodies belonging to a single B cell clone, characterised as having high sequence identity, especially within CDR3 and shared V and J genes, in both the light and heavy chains (Scheid *et al.*, 2011). This can be very informative in the search for high affinity variants of naturally occurring antibodies that may correlate with favourable disease outcome. A number of vectors are available for expression cloning of the isolated genes, including mammalian expression vectors (Metheringham *et al.*, 2009; Tiller *et al.*, 2008) and insect cell expression vectors (Johansson *et al.*, 2007a). By introducing restriction sites within the PCR amplification primers it is possible to clone Ig gene products directly into an expression vector, enabling the production of large amounts of MAb. Subsequent phenotypic studies, such as neutralisation, cross-reactivity and epitope binding can be easily correlated to the genotype of the antibody, providing a greater understanding of the antibody response to important pathogens.

3.2 Aims of the work presented in this chapter

Isolation and amplification of the Ig variable (VH and VL) chains from the B cells of HCV-infected individuals represents a powerful technique for studying antibody responses. Several different approaches to B cell isolation and stimulation are available, as well as several primer sets for Ig gene transcript amplification. This study was intended to explore the currently available methods and to develop a protocol for amplification of VH and VL gene transcripts from patients with good nAb responses. All preliminary experiments were carried out on samples collected during the chronic phase of disease. The work to isolate IgG^+ memory B cells and stimulate their differentiation into ASCs under culture conditions was undertaken by Richard Urbanowicz and is covered briefly. Recovering mRNA from B cell cultures, complementary DNA (cDNA) synthesis and PCR amplification of variable regions was undertaken as part of this PhD thesis and the troubleshooting of this method is covered in greater detail. One of the original aims of this work was to clone paired VH and VL chains into an appropriate vector for the expression and characterisation of the antibody. Ultimately this was not achieved, but the progress made towards this goal and the problems encountered are discussed.

3.3 Materials and methods

3.3.1 Isolation and culturing of B cells

The protocols described in section 3.3.1 were developed by Dr. Richard Urbanowicz. All work to isolate B cells from patient samples and culture them for antibody production was carried out by Dr. Richard Urbanowicz.

3.3.1.1 Sample collection

Patients were identified through The Trent HCV Cohort. The Nottingham Local Research Ethics Committee approved the study protocol and written informed consent was obtained from the participants before entering the study.

3.3.1.2 Preparation of PBMCs from whole blood

Approximately 10 ml blood was diluted 2:1 in warm RPMI 1640 media (Invitrogen, Paisley, UK) (supplemented with 10% foetal bovine serum (FBS-GOLD)). Diluted blood was gently layered onto 10 ml warm Histopaque (Sigma Aldrich, Gillingham, UK) and centrifuged at 800 x g for 22 minutes. This separated the blood into four layers; (i) a bottom layer containing red blood cells, (ii) a clear layer of Histopaque, (iii) a thin white layer containing PBMCs and (iv) a top layer of serum. PBMCs were removed using a 2 ml Pastette, taking care to remove as little Histopaque as possible and to avoid contamination of PBMCs with serum. PBMCs were transferred to a 50 ml centrifuge tube and topped up to 50 ml with media. Cells were washed by centrifugation at 300 x g for 10 minutes. Supernatant was removed and cells

resuspended in 5 ml media. A cell count was performed on 10 µl PBMC preparation using a haemocytometer. The PBMC preparation was topped up to 50 ml with media and washed by centrifugation at 300 x g for 5 minutes prior to removal of the supernatant.

3.3.1.3 Separation of IgG^+ memory B cells from total PBMCs

IgG^+ memory B cell isolation was carried out using a MACS IgG^+ Memory B Cell Isolation Kit (Miltenyi Biotech, Bisley, UK) according to the manufacturer's instructions. Briefly PBMCs were resuspended in 400 µl MACS buffer (PBS, pH 7.2; 0.5% BSA; 2 mM EDTA) plus 100 µl MACS Memory B cell Biotin-Antibody Cocktail microbeads (containing antibodies to CD2, CD14, CD16, CD36, CD43 and CD235a) per 10^8 cells. Samples were mixed well and incubated at 4°C for 10 minutes. A further 300 µl MACS buffer was added per 10^8 cells followed by 200 µl MACS Anti-biotin microbeads per 10^8 cells. Samples were mixed well and incubated at 4°C for 15 minutes. Cells were washed by adding 10 ml MACS buffer and centrifuging at 300 x g for 10 minutes. Supernatant was aspirated completely and the pellet resuspended in 1 ml MACS buffer.

An LD column was prepared by placing it in a MACS separator and washing with 2 ml MACS buffer. The cell suspension was applied to the column and washed twice with 1 ml MACS buffer. The flow through containing the total B cell fraction was collected. (The LD column containing non-B cells, i.e. T cells, monocytes, dendritic cells, platelets etc., was removed from the separator

and the B cell negative fraction eluted into a separate tube for generation of macrophage feeder cells (see section 3.3.1.4)).

The total B cell fraction was mixed with 7 ml MACS buffer and centrifuged at 300 x g for 10 minutes. Supernatant was aspirated completely and the pellet resuspended in 50 µl MACS buffer plus 50 µl MACS Anti-IgG microbeads per 10^8 cells. The sample was mixed well and incubated at 4°C for 15 minutes. Labelled cells were washed in 10 ml MACS buffer with centrifuging at 300 x g for 10 minutes. Supernatant was aspirated completely and the pellet resuspended in 500 µl MACS buffer. An MS column was prepared by placing it in a MACS separator and washing with 500 µl MACS buffer. The cell suspension was applied to the column and washed 3 times with 500 µl MACS buffer. The MS column was removed from the MACS separator and placed over a sterile collection tube. IgG^+ memory B cells were eluted from the column in 1 ml MACS buffer. IgG^+ memory B cells were counted using a haemocytometer, centrifuged at 300 x g for 10 minutes and resuspended at the desired concentration for culturing (see section 3.3.1.5).

3.3.1.4 Generation of macrophage feeder cell cultures

The B cell negative fraction isolated in section 3.3.1.3 was resuspended to a total volume of 20 ml in RPMI 1640 media supplemented with 10% FBS-GOLD and 50 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF) (R&D Systems, Abingdon, UK). Cells were placed in a Corning T75 tissue culture flask (Fisher Scientific, Loughborough, UK) and incubated at 37°C with 5% CO_2 for 5-7 days. Supernatant was removed and the adherent

macrophages washed once in cold PBS. Macrophages were removed from the plastic by addition of 5 ml warm cell dissociation buffer (Gibco) (Invitrogen, Paisley, UK) and incubated at 37°C for 5-10 minutes. Macrophages were dislodged by gentle tapping and counted using a haemocytometer. Cells were resuspended in 5 ml media in a sterile polystyrene tube wrapped in Parafilm M (Alpha Laboratories, Eastleigh, UK). Macrophage cells were lightly irradiated (12Gy) using a caesium-137 source and plated at 2500 cells per well in a 384-well flat-bottom tissue culture microplate (BD Biosciences, Oxford, UK).

3.3.1.5 Culturing IgG^+ memory B cells for production of antibody

IgG^+ memory B cells were seeded onto macrophage feeder cells at 1.3, 10 or 100 cells per well in 100 µl total volume of RPMI 1640 media supplemented with 10% FBS-GOLD, R848 (2.5 µg/ml) (Source BioScience, Nottingham, UK) and IL-2 (1000 units/ml) (R&D Systems, Abingdon, UK). Cells were deposited using a Precision Microplate Pipetting System (BioTek UK, Potton, UK). Cultures were incubated at 37°C with 5% CO_2 for 10 days.

3.3.1.6 Screening cell culture supernatants for antibody production

3.3.1.6.1 GNA-capture ELISA

A Maxisorp 96-well plate (Nunc) (Fisher Scientific, Loughborough, UK) was coated, blocked and washed as described in section 2.3.4.3. Cell lysate from HEK 293T cells transfected to express E1E2 (H77 genotype 1a) was diluted 1:5 in 1x PBS and 50 µl coated onto each well for 1 hour at room temperature. Alternatively for detection of total Igs, anti-human IgG was coated onto the

wells. Wells were washed three times in PBS-T. Thirty microlitres of B cell supernatant was added and incubated at room temperature for 2 hours. Wells were washed three times in PBS-T. Bound Ig was detected by the addition of 50 µl AP-conjugated secondary antibody (goat anti-human IgG) (Sigma Aldrich, Gillingham, UK) (1:2000 in PBS) and incubated at room temperature for 1 hour. Wells were washed three times in PBS-T and IgG-containing supernatants identified by the addition of 100 µl pNPP (Fisher Scientific, Loughborough, UK). Absorbance at 405 nm was determined using a Fluorostar Optima plate reader (BMG Labtech, Aylesbury, UK).

3.3.1.6.2 Peptide capture immunoassay

Biotinylated peptides were synthesized using Synphase PA Lanterns (Mimotopes, Melbourne, Australia), using a C-terminal biocytin residue separated from the epitope sequence by a Gly-Ser-Gly spacer sequence. Peptide purity was assessed by MALDI-TOF spectrometry. Peptides were initially dissolved in DMSO then PBS to a final concentration of 1 mg/ml. The peptides corresponded to amino acids 412-423 (QLINTNGSWHIN; Peptide I) and 434-446 (NTGWLAGLFYQHK; Peptide II) of the HCV H77c polyprotein (Tarr *et al.*, 2012b).

Neutravidin (Pierce, Cramlington, UK) was coated on to a Maxisorp 96-well plate (Nunc) (Fisher Scientific, Loughborough, UK) at 5 µg/ml in 50 mM carbonate/bicarbonate buffer (pH9.8) overnight at 4°C in a humidity chamber. Unbound neutravidin was removed by washing with PBS-T, and non-specific binding sites blocked using PBS-T containing 3% bovine serum albumin (PBS-

TB). Peptide solutions (1 µg/ml in PBS-T) were incubated at room temperature for 1 hour. Unbound peptide was removed by washing with PBS-TB and 30 µl of B-cell supernatant bound to the capture peptide for 2 hours at room temperature. Unbound supernatants were removed by washing three times with PBS-T. Bound Ig was detected by the addition of 50 µl AP-conjugated anti-human IgG antibody (Sigma Aldrich, Gillingham, UK), diluted 1:2000 in Tris-buffered saline-0.05% Tween 20 solution (TBS-T). Unbound antibody was removed by washing with TBS-T and IgG-containing supernatants identified by the addition of 100 µl pNPP (Sigma Aldrich, Gillingham, UK). Absorbance at 405 nm was determined using a Fluorostar Optima plate reader (BMG Labtech, Aylesbury, UK).

3.3.2 Isolation and amplification of IgG genes from B cells

3.3.2.1 Extraction of RNA from B cell samples

B cell supernatants testing positive for the GNA-capture ELISA and both peptide capture immunoassays (peptide I and peptide II) were selected for isolation and analysis of the IgG genes expressed. The cells, along with the remaining supernatant (approximately 40 µl), were transferred to a 0.7 ml Eppendorf. The sample was washed twice in 500 µl B cell wash buffer (1x PBS, 0.7 units/µl RNAse OUT (Invitrogen, Paisley, UK)) with centrifugation at 300 x g for 5 minutes. Residual wash buffer was carefully removed by pipetting. Cells were lysed with 10 µl B cell lysis buffer (10 units RNAse OUT (Invitrogen, Paisley, UK), 2.5 µl 5x Thermoscript RT buffer (Invitrogen, Paisley, UK), 0.625 µl 0.1 M dithiothreitol (DTT) (Invitrogen, Paisley, UK),

0.03125 µl Igepal CA-630 (Sigma Aldrich, Gillingham, UK), made up to 10 µl with diethyl pyrocarbonate (DEPC)-treated H_2O (Invitrogen, Paisley, UK)). Samples were transferred to 0.2 ml PCR tubes and immediately snap frozen on dry ice and stored at -80°C. This B cell lysis method is adapted from Wu *et al*. (Wu *et al.*, 2010).

3.3.2.2 IgG-specific cDNA synthesis from RNA samples

RNA samples were thawed slowly on ice. To each sample 75 ng random hexamers (Invitrogen, Paisley, UK), 10 mM dNTPs (Invitrogen, Paisley, UK) and 7.5 units Thermoscript reverse transcriptase (RT) (Invitrogen, Paisley, UK) was added. Samples were incubated on a thermal cycler at 25°C for 10 minutes to allow random hexamer annealing. cDNA synthesis was carried out at 50°C for 1 hour followed by 55° for 1 hour. The reaction was terminated by heating to 85°C for 5 minutes.

3.3.2.3 IgG variable (V) region amplification by nested PCR

IgG V regions were amplified from cDNA using a nested PCR protocol with primer sets specific for the heavy and light chains. Two different primer sets were used in this analysis.

3.3.2.3.1 PCR amplification using primer set A

Primer set A was taken from Wang and Stollar (Wang & Stollar, 2000). Primer names and sequences are listed in Table 3-1 and 3-2. Gamma chain-

Table 3-1. Primer set A, heavy chain primers.

Target	1st or 2nd round	Primer orientation	Primer name	Primer sequence (5'-3')
Heavy chain (γ)	1st	Sense	VHL-1	TCACCATGGACTGCACCTGGA
	1st	Sense	VHL-2	CCATGGACACACTTTGCTCCAC
	1st	Sense	VHL-3	TCACCATGGAGTTTGGGCTGAGC
	1st	Sense	VHL-4	AGAACATGAAACACCTGTGGTTCTT
	1st	Sense	VHL-5	ATGGGGTCAACCGCCATCCT
	1st	Sense	VHL-6	ACAATGTCTGTCTCCTTCCTCAT
	1st	Antisense	CγII	GCCAGGGGGAAGACCGATG
Heavy chain (γ)	2nd	Sense	VH-1	TTGCGGCCGCCAGGTGCAGCTGGTGCAGTC
	2nd	Sense	VH-2	TTGCGGCCGCCAGATCACCTTGAAGGAGTC
	2nd	Sense	VH-3	TTGCGGCCGCGAGGTGCAGCTGGTGGAGTC
	2nd	Sense	VH-4	TTGCGGCCGCCAGGTGCAGCTGCAGGAGTC
	2nd	Sense	VH-5	TTGCGGCCGCGAGGTGCAGCTGGTGCAGTC
	2nd	Sense	VH-6	TTGCGGCCGCCAGGTACAGCTGCAGCAGTC
	2nd	Antisense	CγIII	AGGTCTAGAGACCGATGGGCCCTTGGTGGA

Restriction sites are underlined. *Not*I, GCGGCCGC; *Xba*I, TCTAGA. All primers were synthesised commercially by Eurofins MWG Operon, Germany.

Table 3-2. Primer set A, light chain primers.

Target	1st or 2nd round	Primer orientation	Primer name	Primer sequence (5'-3')
Light chain (κ)	1st	Sense	VκL-1	GCTCAGCTCCTGGGGCTCCTG
	1st	Sense	VκL-2	CTGGGGCTGCTAATGCTCTGG
	1st	Sense	VκL-3	TTCCTCCTGCTACTCTGGCTC
	1st	Sense	VκL-4	CAGACCCAGGTCTTCATTTCT
	1st	Antisense	CγII	GCCAGGGGGAAGACCGATG
Light chain (λ)	1st	Sense	VλL-1	CCTCTCCTCCTCACCCTCCT
	1st	Sense	VλL-2	CTCCTCACTCAGGGCACA
	1st	Sense	VλL-3	ATGGCCTGGATCCCTCTCC
	1st	Antisense	CλII	AGCTCCTCAGAGGAGGGCGG
Light chain (κ)	2nd	Sense	Vκ-1	CATAAGATCTCGACATCCAGATGACCCAGT
	2nd	Sense	Vκ-2	CACCAGATCTCGATATTGTGATGACCCAG
	2nd	Sense	Vκ-3	CACCAGATCTCGAAATTGTGTTGACGCAGTCT
	2nd	Sense	Vκ-4	CACCAGATCTCGACATCGTGATGACCCAGT
	2nd	Antisense	CκIII	TATTCCATGGAAGATGAAGACAGATGGTGC
Light chain (λ)	2nd	Sense	Vλ-1	TATTAGATCTCCAGTCTGTGCTGACTCAGC
	2nd	Sense	Vλ-2	TATTAGATCTCCAGTCTGCCCTGACTCAGC
	2nd	Sense	Vλ-3	CACCAGATCTCTCCTATGAGCTGACTCAGC
	2nd	Antisense	CλIII	CATTCCATGGGGGAACAGAGTGACCG

Restriction sites are underlined. *Bgl*II, AGATCT; *Nco*I, CCATGG. All primers were synthesised commercially by Eurofins MWG Operon, Germany.

specific primers were used to select for Igγ-type heavy chains only. Both kappa and lambda-type light chains were amplified in a single reaction with light-chain specific primers. IgG VH and VL was amplified from 2.5 µl cDNA in a total reaction volume of 15 µl containing 1.5 units LongAmp Taq DNA polymerase (New England Biolabs, Hitchin, UK), 200 µM dNTPs (Fermentas St. Leon-Rot, Germany) 3% DMSO (Finnzymes, Vantaa, Finland) and 5 pM each primer. PCR amplification was performed in a 3-step reaction consisting of 94°C for 30s, 45 cycles of 94°C for 45s, 50°C for 45s and 65°C for 1 minute, followed by 65°C for 10 minutes. The second round PCR was carried out as described using 2.5 µl first round product as template with 55 cycles of amplification. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2 and purified as described in section 2.3.1.3. Purified PCR products were sequenced with the antisense primer specific to the chain type as described in 2.3.1.10.

3.3.2.3.2 PCR amplification using primer set B

Primer set B was generated by combining primers described in Wardemann *et al*. (Wardemann *et al.*, 2003), Koelsch *et al*. (Koelsch *et al.*, 2007) and Tiller *et al*. (Tiller *et al.*, 2008). Primer names and sequences are listed in Table 3-3, 3-4 and 3-5. Gamma chain-specific primers were used to select for Igγ-type heavy chains only. Kappa-type light chains were amplified independently of lambda-type light chains in separate reactions with either kappa chain or lambda chain-specific primers. Ig VH, Vκ and Vλ transcripts were amplified from 1.75 µl cDNA in a total reaction volume of 15 µl containing 1.5 units LongAmp Taq DNA polymerase (New England Biolabs, Hitchin, UK), 200

Table 3-3. Primer set B, heavy chain primers.

Target	1st or 2nd round	Primer orientation	Primer name	Primer sequence (5'-3')
IgH	1st	Sense	5' L-VH1	ACAGGTGCCCACTCCCAGGTGCAG
	1st	Sense	5' L-VH3	AAGGTGTCCAGTGTGARGTGCAG
	1st	Sense	5' L-VH4/6	CCCAGATGGGTCCTGTCCCAGGTGCAG
	1st	Sense	5' L-VH5	CAAGGAGTCTGTTCCGAGGTGCAG
	1st	Antisense	Cγ external	TCTTGTCCACCTTGGTGTTGC
	1st	Antisense	Cγ CH1	GGAAGGTGTGCACGCCGCTGGTC
IgH	2nd	Sense	5' *Age*I VH1/5	CTGCAACCGGTGTACATTCCGAGGTGCAGCTGGTGCAG
	2nd	Sense	5' *Age*I VH3	CTGCAACCGGTGTACATTCTGAGGTGCAGCTGGTGGAG
	2nd	Sense	5' *Age*I VH4	CTGCAACCGGTGTACATTCCCAGGTGCAGCTGCAGGAG
	2nd	Sense	5' *Age*I VH3-23	CTGCAACCGGTGTACATTCTGAGGTGCAGCTGTTGGAG
	2nd	Sense	5' *Age*I VH4-34	CTGCAACCGGTGTACATTCCCAGGTGCAGCTACAGCAGTG
	2nd	Sense	5' *Age*I VH1	CTGCAACCGGTGTACATTCCCAGGTGCAGCTGGTGCAG
	2nd	Sense	5' *Age*I VH1-18	CTGCAACCGGTGTACATTCCCAGGTTCAGCTGGTGCAG
	2nd	Sense	5' *Age*I VH1-24	CTGCAACCGGTGTACATTCCCAGGTCCAGCTGGTACAG
	2nd	Sense	5' *Age*I VH3-33	CTGCAACCGGTGTACATTCTCAGGTGCAGCTGGTGGAG
	2nd	Sense	5' *Age*I VH3-9	CTGCAACCGGTGTACATTCTGAAGTGCAGCTGGTGGAG
	2nd	Sense	5' *Age*I VH4-39	CTGCAACCGGTGTACATTCCCAGCTGCAGCTGCAGGAG
	2nd	Sense	5' *Age*I VH6-1	CTGCAACCGGTGTACATTCCCAGGTACAGCTGCAGCAG
	2nd	Antisense	IgG internal	GTTCGGGGAAGTAGTCCTTGAC
	2nd	Antisense	3' *Sal*I JH1/2	TGCGAAGTCGACGCCTGAGGAGACGGTGACCAG
	2nd	Antisense	3' *Sal*I JH3	TGCGAAGTCGACGCTGAAGAGACGGTGACCATTG
	2nd	Antisense	3' *Sal*I JH4/5	TGCGAAGTCGACGCTGAGGAGACGTGACCAG
	2nd	Antisense	3' *Sal*I JH6	TGCGAAGTCGACGCTGAGGAGACGGTGACCGTG
	2nd	Antisense	Cγ nested	AGGTGCTCTTGGAGGAGGGT

Restriction sites are underlined. *Age*I, ACCGGT; *Sal*I, GTCGAC.Primers were synthesised by Eurofins MWG Operon, Germany.

Table 3-4. Primer set B, kappa chain primers.

Target	1st or 2nd round	Primer orientation	Primer name	Primer sequence (5'-3')
Igκ	1st	Sense	5' L-Vκ1/2	ATGAGGSTCCCYGCTCAGCTGCTGG
	1st	Sense	5' L-Vκ3	CTCTTCCTCCTGCTACTCTGGCTCCCAG
	1st	Sense	5' L-Vκ4	ATTTCTCTGTTGCTCTGGATCTCTG
	1st	Antisense	3' Cκ 543	GTTTCTCGTAGTCTGCTTTGCTCA
Igκ	2nd	Sense	5' Pan Vκ	ATGACCCAGWCTCCABYCWCCCTG
	2nd	Antisense	3' Cκ 494	GTGCTGTCCTTGCTGTCCTGCTC
Igκ	Specific	Sense	5' *AgeI* Vκ 1-5	CTGCAACCGGTGTACATTCTGACATCCAGATGACCCAGTC
	Specific	Sense	5' *AgeI* Vκ 1-9	TTGTGCTGCAACCGGTGTACATTCAGACATCCAGTTGACCCAGTCT
	Specific	Sense	5' *AgeI* Vκ 1D-43	CTGCAACCGGTGTACATTGTGCCATCCGGATGACCCAGTC
	Specific	Sense	5' *AgeI* Vκ 2-24	CTGCAACCGGTGTACATGGGGATATTGTGATGACCCAGAC
	Specific	Sense	5' *AgeI* Vκ 2-28	CTGCAACCGGTGTACATGGGGATATTGTGATGACTCAGTC
	Specific	Sense	5' *AgeI* Vκ 3-11	TTGTGCTGCAACCGGTGTACATTCAGAAATTC
	Specific	Sense	5' *AgeI* Vκ 3-15	CTGCAACCGGTGTACATTCAGAAATAGTGATGACGCAGTC
	Specific	Sense	5' *AgeI* Vκ 3-20	TTGTGCTGCAACCGGTGTACATTCAGAAATTGTGTTGACGCAGTCT
	Specific	Sense	5' *AgeI* Vκ 4-1	CTGCAACCGGTGTACATTCGGACATCGTGATGACCCAGTC
	Specific	Sense	5' *AgeI* Vκ 2-30	CTGCAACCGGTGTACATGGGGATGTTGTGATGACTCAGTC
	Specific	Antisense	3' *BsiW*I Jκ1/2/4	GCCACCGTACGTTTGATYTCCACCTTGGTC
	Specific	Antisense	3' *BsiW*I Jκ3	GCCACCGTACGTTTGATATCCACTTTGGTC
	Specific	Antisense	3' *BsiW*I Jκ5	GCCACCGTACGTTTAATCTCCAGTCGTGTC

Restriction sites are underlined. *Age*I, ACCGGT; *Bsi*WI, CGTACG. Primers were synthesised by Eurofins MWG Operon, Germany.

Table 3-5. Primer set B, lambda chain primers.

Target	1st or 2nd round	Primer orientation	Primer name	Primer sequence (5'-3')
Igλ	1st	Sense	5' L-Vλ1	GGTCCTGGGCCCAGTCTGTGCTG
	1st	Sense	5' L-Vλ2	GGTCCTGGGCCCAGTCTGCCCTG
	1st	Sense	5' L-Vλ3	GCTCTGTGACCTCCTATGAGCTG
	1st	Sense	5' L-Vλ4/5	GGTCTCTCTCSCAGCYTGTGCTG
	1st	Sense	5' L-Vλ6	GTTCTTGGGCCAATTTTATGCTG
	1st	Sense	5' L-Vλ7	GGTCCAATTCYCAGGCTGTGGTG
	1st	Sense	5' L-Vλ8	GAGTGGATTCTCAGACTGTGGTG
	1st	Antisense	3' Cλ	CACCAGTGTGGCCTTGTTGGCTTG
Igλ	2nd	Sense	5' *Age*I Vλ1	CTGCTACCGGTTCCTGGGCCCAGTCTGTGCTGACKCAG
	2nd	Sense	5' *Age*I Vλ2	CTGCTACCGGTTCCTGGGCCCAGTCTGCCCTGACTCAG
	2nd	Sense	5' *Age*I Vλ3	CTGCTACCGGTTCTGTGACCTCCTATGAGCTGACWCAG
	2nd	Sense	5' *Age*I Vλ4/5	CTGCTACCGGTTCTCTCTCSCAGCYTGTGCTGACTCA
	2nd	Sense	5' *Age*I Vλ6	CTGCTACCGGTTCTTGGGCCAATTTTATGCTGACTCAG
	2nd	Sense	5' *Age*I Vλ7/8	CTGCTACCGGTTCCAATTCYCAGRCTGTGGTGACYCAG
	2nd	Sense	Ab sense	GCTTCGTTAGAACGCGGCTAC
	2nd	Antisense	3' XhoI Cλ	CTCCTCACTCGAGGGYGGGAACAGAGTG

Restriction sites are underlined. *Age*I, ACCGGT; *Xho*I, CTCGAG. Primers were synthesised by Eurofins MWG Operon, Germany.

μM dNTPs (Fermentas St. Leon-Rot, Germany) 3% DMSO (Finnzymes, Vantaa, Finland) and 5 pM each primer. Ig VH and Vκ products were PCR amplified in a 3-step reaction consisting of 94°C for 30s, 50 cycles of 94°C for 30s, 57°C for 30s and 65°C for 55s, followed by 65°C for 10 minutes. Ig Vλ products were PCR amplified in a 3-step reaction consisting of 94°C for 30s, 50 cycles of 94°C for 30s, 60°C for 30s and 65°C for 55s, followed by 65°C for 10 minutes. The second round PCR was carried out as described using 1.75 μl first round product as template with extension at 65°C for 45s.

Second round amplification of Vκ genes was initially carried out with the universal primer pair 5'Pan Vκ and 3'Cκ 494. Following sequence analysis of the first Vκ PCR product to identify the gene family, specific amplification of Vκ was carried out with the appropriate primer pair (specific primers are listed in Table 3-4). The reaction was carried out as described above. PCR products were detected by agarose gel electrophoresis as described in section 2.3.1.2 and purified as described in section 2.3.1.3. Purified PCR products were sequenced with the antisense primer or primers specific to the chain-type, as described in 2.3.1.10.

3.3.2.4 Cloning IgG VH and VL PCR products into pGEM-T Easy

Purified PCR products were ligated into pGEM-T Easy (Promega, Southampton, UK) at a 3:1 molar ratio of insert to vector in a 10 μl reaction containing 3 units T4 DNA ligase (Promega, Southampton, UK). The sample was incubated at room temperature for 1 hour or overnight at 4°C. Ligation products were transformed in *E. coli* TOP10 cells as described in 2.3.1.7 with

plating onto LB agar plates containing 25 µg/ml ampicillin (for pGEM-T Easy) or 35 µg/ml zeocin (for pDCORIG). After overnight incubation colonies were screened as described in section 2.3.1.8. Positive colonies containing the desired insert were inoculated into a 3 ml LB (Sigma Aldrich, Gillingham, UK) culture containing 25 µg/ml ampicillin (for pGEM-T Easy) or 35 µg/ml zeocin (for pDCORIG) and incubated at 37°C overnight. Plasmid was extracted from overnight colonies as described in section 2.3.1.9 and sequenced with plasmid-specific primers (Table 3-6) as described in 2.3.1.10.

3.3.3 Characterisation of IgG variable regions

3.3.3.1 Generation of contiguous sequences

Sequencing reads were edited using Chromas 2.13 software (Technelysium, Brisbane, Australia) and contiguous sequences were made using SeqMan II software (DNASTAR Inc., Madison, USA). Sequences were converted into FASTA format for alignment using MEGA 4 software (Tamura *et al.*, 2007).

3.3.3.2 Analysis of VH and VL domains using the Immunogenetics database

The Immunogenetics (IMGT®) database (Lefranc, 2012) was accessed online to characterise the isolated VH and VL domain sequences. The IMGT/V-QUEST (Brochet *et al.*, 2008) sequence alignment software was used to align the isolated sequences to known Ig genes deposited in the database. Highest percentage identity matches to germline sequences for the V, J and D region are given for heavy chain sequences, or to the V and J region for light

Table 3-6. Additional primers used for adaptation of PCR products, sequencing and site-directed mutagenesis.

Target	Function of primer	Primer orientation	Primer name	Primer sequence (5'-3')
pDCORIG huIgG1	Site-directed mutagenesis	Sense	pDC_*Age*I_s	GTAGCAACAGCTACCGGTGTCCACTCCC
	Site-directed mutagenesis	Sense	pDC_*Age*I_KO	CTACTTCCCCGAGCCGGTGACGGTG
	PCR, sequencing	Sense	pDC_H_s	CTCACTATAGGGAGACCCAAGCTTACC
	PCR, sequencing	Antisense	pDC_H_as	AAGGGCGAATTCGCCCTTTAGA
	PCR, sequencing	Sense	pDC_k_s	TAGCTTGGTACCGAGCTCGGATC
	PCR, sequencing	Antisense	pDC_k_as	TCTAGATGCATGCTCGAGTCAACA
Ig variable chain genes	Expression re-amplifcation	Sense	VH_IgExp_s	GATCACCGGTGTCCACTCCGAGGTGCAGCTGGTG
	Expression re-amplification	Antisense	VH_IgExp_as	GATCTGTACACTCCGAMATYSWGATGACC
	Expression re-amplification	Sense	κ_leader	GATCGGATCC**ACCATGGGATGGAGCTGTATCAT CCTCTTCTTGGTAGCAACAGCTACCGGAGTCCA CTCC**GAMATYSWGATGACCCAGWCT
	Expression re-amplification	Antisense	VK_IgExp_as	GATCCGTACGTTTGATCTCCACCTT
pGEM-T Easy	PCR, sequencing	Sense	M13f	GTAAAACGACGGCCAGT
	PCR, sequencing	Antisense	M13r	AACAGCTATGACCATG

Restriction sites are underlined. *Age*I, ACCGGT; *Afe*I, AGCGCT; *Bam*HI, GGATCC; *Bsi*WI, CGTACG. Leader sequence within κ_leader is highlighted in bold. Primers were synthesised by Eurofins MWG Operon, Germany.

chain sequences. Length and amino acid sequence of FRs and CDRs can be obtained and the amino acid sequence of CDR3 for all isolated genes is given. CDR3 is defined as the region from FR3 up to the conserved tryptophan-glycine (WG) motif in VH domains, or the conserved phenylalanine-glycine (FG) motif in VL domains (Lefranc, 2011b). Information regarding gene functionality is also given; productive denotes those gene rearrangements that contain no stop codons, have an in-frame junction and no defects in the genetic regulatory elements. Unproductive rearrangements may result from stop codons, frameshift mutations, out-of-frame junctions or defects in the regulatory elements (Lefranc, 2011c). In the results section, VH refers to the variable heavy region cDNA, encompassing rearranged V_H, D_H and J_H germline genes. VL refers to the variable light region cDNA, encompassing rearranged germline Vκ and Jκ, or Vλ and Jλ genes. In the case of primer set B, those PCR products generated with kappa-chain specific primers are referred to as Vk. Gene rearrangements are annotated according to the IMGT nomenclature whereby IGHV denotes the group i.e. Ig heavy chain variable; IGHV1 specifies the subgroup and contains all genes sharing 75% or greater sequence homology; IGHV1-2 identifies the particular gene and IGHV1-2*01 denotes the allelic variant from which the Ig is derived (Lefranc, 2011a).

3.3.4 Expression of whole IgG in mammalian cells

3.3.4.1 Adaptation of the expression plasmid pDCORIG huIgG$_1$ by site-directed mutagenesis

The expression plasmid pDCORIG huIgG$_1$ was kindly provided by Lindy Durrant (Scancell Ltd, Nottingham, UK) (Figure 3-3). Prior to cloning the VH and Vk domains, it was necessary to introduce appropriate restriction sites into the plasmid. The restriction site *Age*I was introduced into the heavy chain leader sequence and the *Age*I site present in the heavy chain constant region was knocked-out (Figure 3-4a). Site-directed mutagenesis was carried out on all sites simultaneously using the QuikChange Multi Site-directed Mutagenesis kit (Agilent Technologies, Stockport, UK). Site-specific mutagenesis primers were designed using PrimerX (http://bioinformatics.org/primerx/cgi-bin/DNA_1.cgi) (Table 3-6). One hundred nanograms of plasmid was used as

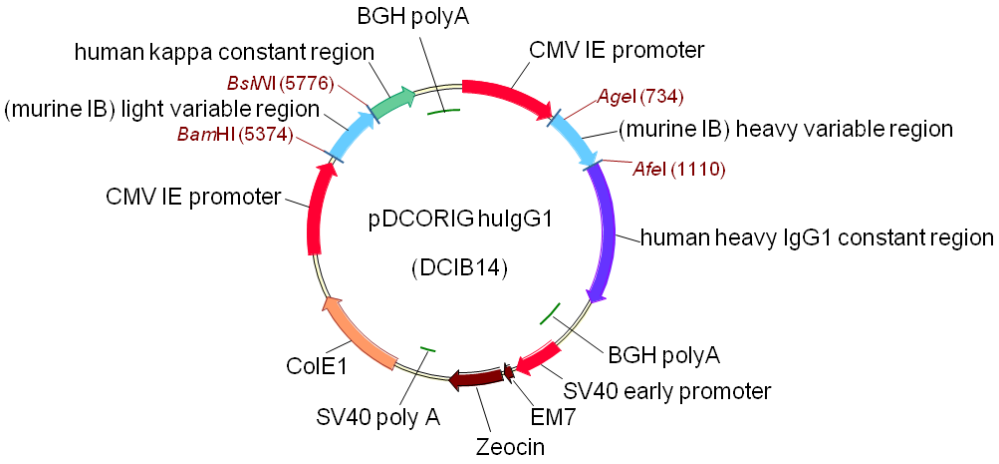

Figure 3-3. Plasmid map of the Ig expression vector, pDCORIG huIgG$_1$ (Scancell Ltd, Nottingham, UK). The light variable and heavy variable regions are shown in light blue. These were replaced by VH and Vκ sequences isolated from patient B cells. Expression cassettes are under the control of a cytomegalovirus immediate early (CMV IE) promoter that directs expression of the variable and constant region genes. The BGH polyadenylation (polyA) signal ensures correct processing of the mRNA transcripts. Zeocin, under the control of the SV40 promoter and polyA signals allows for selection of transfectants in both bacterial and mammalian cells. ColE1 enables replication of the plasmid in *E. coli.* Figure adapted from Metheringham *et al*., (2009).

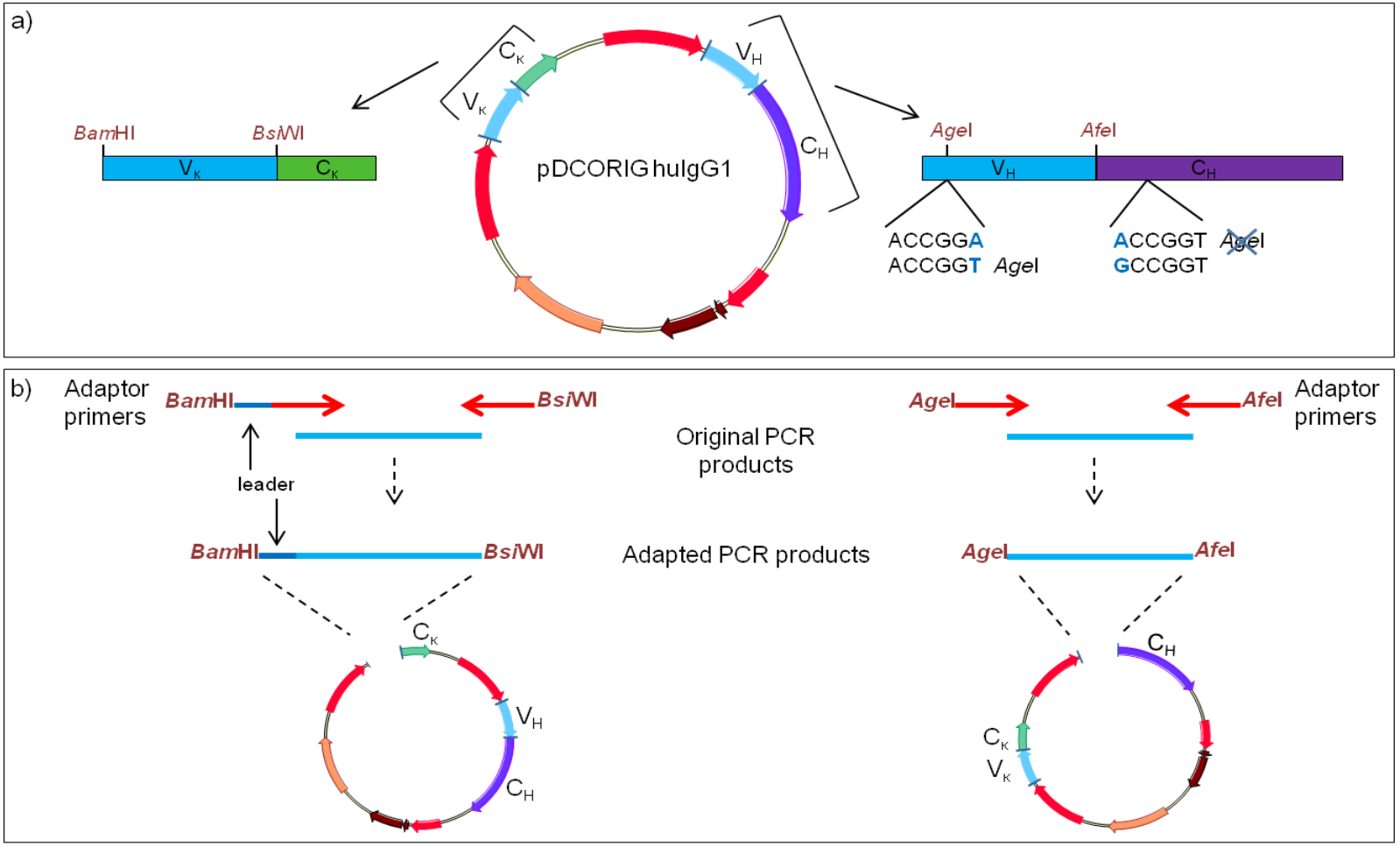

Figure 3-4. Strategy to modify the vector and PCR products for Ig gene cloning. See overpage for full figure legend.

Figure 3-4. Strategy to modify the vector and PCR products for Ig gene cloning. (a) The *Age*I restriction site was introduced into the V_H domain by site directed mutagenesis of A to T (blue text). An *Age*I restriction site was removed from the C_H domain by mutation of A to G (blue text). The light chain cassette was unmodified. (b) The PCR products generated as described in sections 3.3.2.3.1 and 3.3.2.3.2 did not contain compatible restriction sequences. Therefore, V_K PCR products were reamplified with a sense primer containing the Ig leader sequence (dark blue) and the restriction site *Bam*HI. The antisense primer contained *Bsi*WI. V_H PCR products were reamplified with a sense primer containing *Age*I and an antisense primer containing *Afe*I. This allowed restriction enzyme cloning of the adapted PCR products into the vector. Vector map and sequence information was provided by Lindy Durrant (Scancell Ltd, Nottingham, UK).

template in a total reaction volume of 25 µl containing 50 ng of each primer1 µl dNTP mix, 2.5 units QuikChange Multi Enzyme mix and 3% QuikSolution. Mutagenesis was performed in a 3-step cycling reaction consisting of 95°C for 2 minutes, 30 cycles of 95°C for 20s, 55°C for 30s and 65°C for 3 minutes, followed by 65°C for 5 minutes. Parental plasmid was removed by the addition of 10 units *Dpn*I and incubation at 37°C for 1 hour. Plasmid was transformed into TOP10 cells as described in 2.3.1.7. Selected colonies were picked and grown in 3 ml LB culture (Sigma Aldrich, Gillingham, UK) containing 35 µg/ml zeocin and incubated at 37°C overnight. Plasmid was extracted as described in 2.3.1.9 and sequenced with heavy cassette-specific primers, pDC_H_s and pDC_H_as, (Table 3-6) as described in 2.3.1.10 to confirm the presence of the desired mutations.

3.3.4.2 Re-amplification of IgG gene products with universal primers

Universal primers were designed to introduce the restriction sites *Age*I and *Afe*I at the 5' and 3' end respectively of any VH PCR product (Figure 3-4b). A

universal primer was designed to introduce the restriction site *Bsi*WI at the 3' end of any Vk product. Another primer was designed to introduce the Ig leader sequence as well as *Bam*HI at the 5' end of any Vk product (Figure 3-4b). The Ig leader sequence was taken from plasmid pDCORIG huIgG_1. Primers are listed in Table 3-6. Two microlitres of VH or Vk PCR product was used as template in a total reaction volume of 20 µl containing 1 unit LongAmp Taq DNA Polymerase (New England Biolabs, Hitchin, UK), 200 µM dNTPs (Fermentas St. Leon-Rot, Germany), 4 pM each primer (VH_IgExp_s and VH_IgExp_as for amplification of VH products; k_leader and Vk_IgExp_as for amplification of Vκ products) and 3% DMSO (Finnzymes, Vantaa, Finland). PCR amplification was performed in a 2-step reaction consisting of 94°C for 30s, 40 cycles of 94°C for 45s and 65°C for 50s, followed by 65°C for 10 minutes. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2 and purified as described in 2.3.1.3. Products were cloned into pGEM-T Easy (Promega, Southampton, UK) as described in 3.3.2.4 and transformed into *E. coli* TOP10 cells as described in 2.3.1.7.

3.3.4.3 Cloning VH genes into pDCORIG huIgG_1

Purified VH PCR products (containing appropriate restriction sites) and pDCORIG huIgG_1 plasmid were digested with 10 units *Age*I (New England Biolabs, Hitchin, UK) and 10 units *Afe*I (New England Biolabs, Hitchin, UK) in a 20 µl reaction volume for 1 hour at 37°C. Restriction enzymes were inactivated by heating to 65°C for 20 minutes. Plasmid digestion was confirmed by agarose gel electrophoresis as described in 2.3.1.2 and the plasmid column purified as described in 2.3.1.3. VH was ligated into

pDCORIG huIgG$_1$ containing compatible ends at a 3:1 molar ratio of insert to vector in a 20 µl reaction volume containing 400 cohesive end units T4 DNA ligase (New England Biolabs, Hitchin, UK). Reaction was incubated at 16°C for 1 hour and plasmids transformed into *E. coli* TOP10 cells as described in 2.3.1.7. Colonies were screened, as described in section 2.3.1.8, with heavy cassette-specific primers, pDC_H_s and pDC_H_as, (Table 3-6) to confirm the presence of a full-length VH region. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2 and putative positive colonies were inoculated into a 3 ml LB culture (Sigma Aldrich, Gillingham, UK) containing 35 µg/ml zeocin and incubated at 37°C overnight. Plasmid was extracted as described in 2.3.1.9 and sequenced as described in 2.3.1.10 with heavy cassette-specific primers to confirm the introduction of a novel VH region.

3.3.4.4 Cloning Vκ genes into pDCORIG huIgG$_1$

Modified Vk products and plasmid pDCORIG huIgG$_1$ containing a novel VH region were digested with 10 units *Bam*HI (New England Biolabs, Hitchin, UK) in a 20 µl reaction volume for 1 hour at 37°C prior to the addition of 10 units *Bsi*WI (New England Biolabs, Hitchin, UK) and incubation for 1 hour at 55°C. Plasmid digestion was confirmed by agarose gel electrophoresis as described in 2.3.1.2. All digested products were column purified as described in 2.3.1.3. Vk was ligated into pDCORIG huIgG$_1$ containing compatible ends as described in 3.3.4.3. Transformation, screening, growth and plasmid extraction from putative positive colonies and detection of a novel Vk region was carried as described in 3.3.4.3 using kappa specific primers (pDC_k_s and pDC_k_as) (Table 3-6).

3.4 Results

3.4.1 Stimulation of B cell cultures to produce IgG

Initial experiments were carried out with cultures of B cells seeded at either 10 or 100 cells per well. These cultures were stimulated for either 5 days or 10 days with R848, a TLR7 and TLR8 agonist, and IL-2, a B cell growth factor (Pinna *et al.*, 2009). The IgG-specific mRNA was detected by cDNA synthesis and PCR amplification. Both VH and VL gene transcripts could be detected by amplification with primer set A (Figure 3-5 and 3-6) but not with primer set B (data not shown).

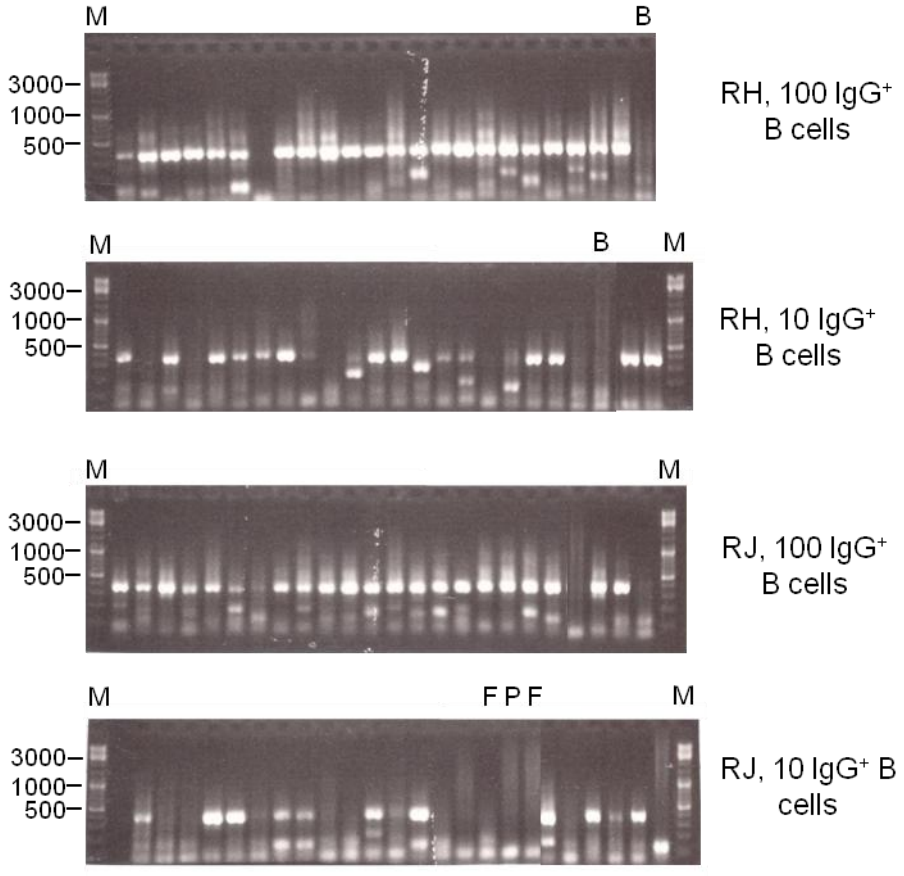

Figure 3-5. Amplification of IgG light chain variable domains. Primer set A was used to amplify VL from cultures of 10 or 100 B cells isolated from two HCV-positive patients, RH or RJ. Positive reactions contain products at ~400 bp. B, B cell negative controls; F, macrophage feeder cell negative controls; P, cDNA negative control. Position of DNA marker, M, is shown in base pairs.

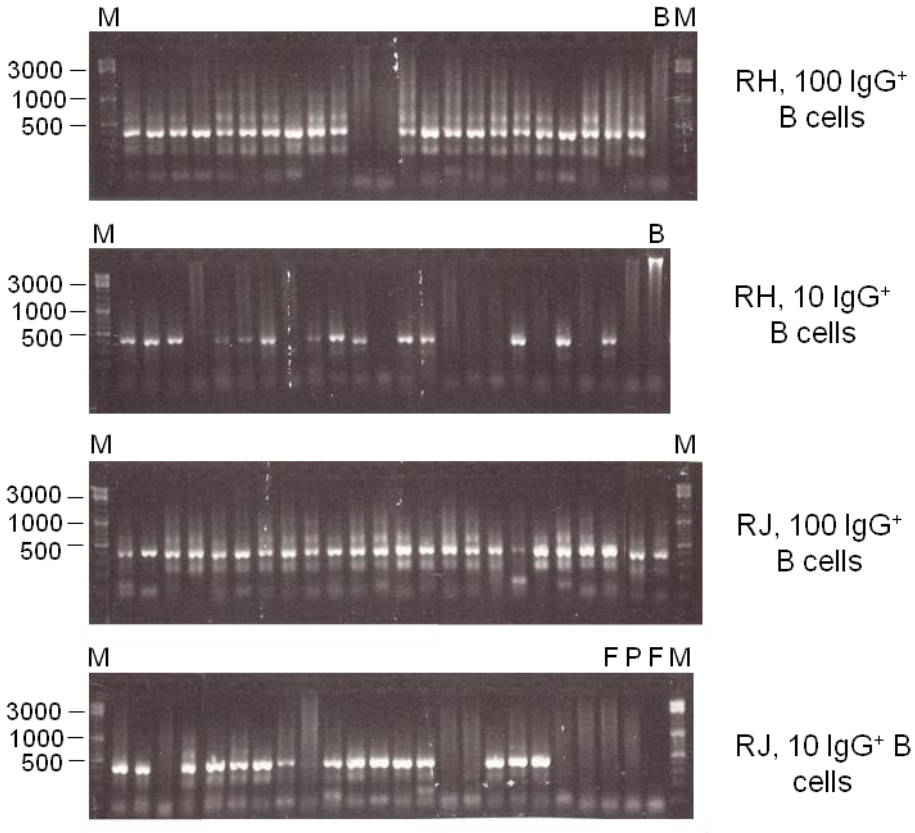

Figure 3-6. Amplification of IgG heavy chain variable domains. Primer set A was used to amplify VH from cultures of 10 or 100 B cells isolated from two HCV-positive patients, RH or RJ. Positive reactions contain products at ~400 bp. B, B cell negative controls; F, macrophage feeder cell negative controls; P, cDNA negative control. Position of DNA marker, M, is shown in base pairs.

VH and VL domains were detected in 5-day old cultures (data not shown) as well as in 10-day old cultures (Figure 3-5 and 3-6). The age of the culture did not impact the number of IgG-PCR positive wells recovered; 74.0% of cultures tested positive for VL and 84.4% tested positive for VH after 5 days, whereas 73.6% of cultures tested positive for VL and 79.0% of cultures tested positive for VH after 10 days. This is comparable to the 70% PCR recovery levels (range 30-80%) reported previously for this primer set (Wang & Stollar, 2000). However, the number of cells seeded per well did influence the recovery of VH and VL transcripts. After 10 days incubation, 93.6% of cultures containing 100 B cells tested positive for VH compared to only 63.6% of cultures

containing 10 B cells (Figure 3-5). Similarly 91.5% of cultures containing 100 B cells tested positive for VL compared to only 54.5% of cultures containing 10 B cells (Figure 3-6). This was expected as not all B cells within the culture will produce IgG. Therefore, there is a greater chance of generating IgG-PCR negative cultures when fewer B cells are seeded. These cultures were generated from B cells isolated from two separate patients, RH and RJ. After 10 days culture, there was very little difference between the patients in terms of IgG production, as quantified by recovery of PCR products; for 100 B cells, 91.6% of RJ-derived cultures tested positive compared to 93.4% of RH-derived cultures; for 10 B cells, 59.5% of RJ-derived cultures tested positive compared to 60.8% of RH-derived cultures. These experiments also showed that the PCR products recovered were specific to the IgG VH and VL domains, as cultures lacking B cells tested negative. Macrophage feeder cell-negative control wells also did not contain IgG PCR products (Figure 3-5 and 3-6). The feeder cells provide additional cytokines that support the proliferation of the IgG^+ memory B cells (Pinna *et al.*, 2009), without which the B cells are not able to produce antibody. This confirmed that the PCR products were specific to IgG VH and VL transcripts expressed in B cells.

Having confirmed that both VH and VL transcripts could be isolated from cultured B cells with at least one primer set, it was necessary to identify those cultures that might be producing HCV-specific Igs. Reactivity of the cell-culture supernatants to HCV E1E2 (genotype 1a) was tested using GNA-capture ELISA. Only 4.1% of 5-day old cultures and 9.8% of 10-day old cultures were HCV E1E2-reactive. This demonstrated that the majority of IgG produced under these culture conditions was not HCV-specific. The B cell

isolation method distinguishes IgG^+ memory B cells from other B cell isotypes (IgM, IgA etc); however, antigen specificity of the IgG response can only be tested after antibody production has been stimulated for several days. Evaluation of subsequent B cell cultures confirmed these initial findings; 20-30% of cultures containing a single cell tested positive by ELISA for IgG production, of which 2-12% was HCV E1E2 (genotype 1a)-specific. For cultures containing 10 B cells, 37-50% of wells tested positive for IgG production of which, 3-11% was HCV E1E2 (genotype 1a)-specific. The values were obtained from multiple experiments with 3 different donors and suggest that the B cell stimulation method did not activate all the B cells, a phenomenon that has been reported in the literature (Pinna *et al.*, 2009). The level of IgG antibody production from single cell cultures was approximately 100 ng/ml (range 74-170 ng/ml) (Figure 3-7) which appears to be lower than the production levels of 200 ng/culture (per cell) reported previously (Pinna *et al.*, 2009). These results prompted us to alter the way in which the cultures were analysed, with assessment of IgG production and HCV E1E2 reactivity preceding VH and VL isolation in all subsequent experiments. It was felt that this would reduce the overall workload by avoiding cDNA synthesis and PCR amplification from cultures which were of no further interest.

3.4.2 Cloning and characterisation of IgG transcripts from patients RH and RJ

Culture wells containing HCV E1E2-specific IgG were matched to those containing VH and VL PCR products. Of thirteen wells testing positive for HCV-specific Igs, 10 were PCR positive for both the VH and VL domain, two

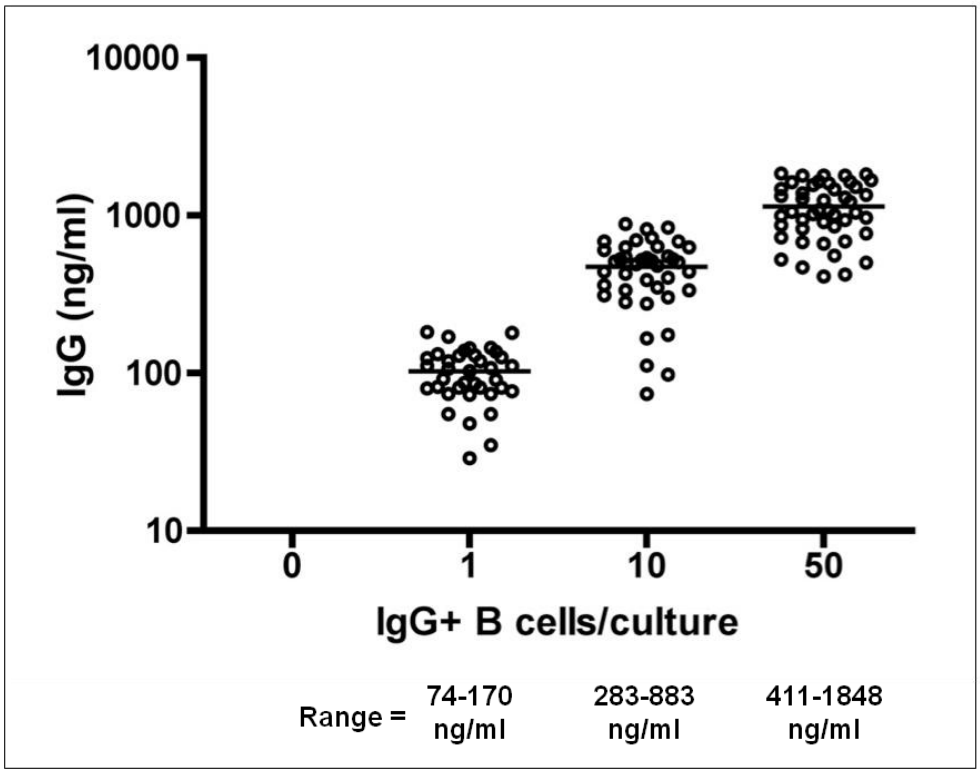

Figure 3-7. IgG production in B cell cultures. Cultures of 1, 10 or 50 cells were plated onto irradiated macrophage feeder cells and stimulated with IL-2 and R848 for 10 days. IgG secreted into the culture supernatant was measured by ELISA. Data was generated and the graph produced in GraphPad Prism 4.03 for Windows (Graphpad Software, San Diego, USA) by Richard Urbanowicz. Data is representative of results for B cell cultures generated from three separate patients.

wells were PCR positive for VH but negative for VL and one well was negative for both PCR products. This may have been caused by insufficient cDNA synthesis which can occur if the cultures become contaminated with RNAse leading to degradation of the IgG-specific mRNA. However, as IgG transcripts were successfully amplified from the majority of cultures it seems unlikely that RNAse contamination would have affected a single sample. It is more likely that the cell pellet was lost during the RNA extraction protocol due to a handling error. Alternatively, mutation within the primer binding sites may have caused the amplification to fail, a problem that has been reported previously (Scheid *et al.*, 2011). This would also explain amplification of VH

but not VL in two of the samples. VH and VL PCR products were purified from all samples that tested positive for both transcripts and were also HCV E1E2 reactive. The PCR products were then cloned into the vector pGEM-T Easy (Promega, Southampton, UK). Multiple colonies containing the VH and VL inserts from different samples were screened. Most colonies were found to contain the product of interest. To ascertain the level of variation between sequences isolated from cultures of either 10 or 100 B cells, several colonies were sequenced across the VH or VL domain. Sequences were then analysed using the IMGT® database (Lefranc, 2012) to determine the VDJ (for the heavy chain) or VJ (for the light chain) gene rearrangement forming the variable domain. Table 3-7 and 3-8 show the results for several sequenced clones from patient RJ. Sequence names correspond to the culture well from which the genes were isolated. Sequences labelled 10.6, 12.2 and 23.2 were derived from cultures of 10 B cells. Sequences labelled 19.2, 19.6 and 21.6 were all derived from cultures of 100 B cells. A single culture well from patient RH tested positive for HCV E1E2-reactive IgG production as well the VH and VL domains, however all sequences generated contained unproductive gene rearrangements. As expected for cultures containing more than one B cell, there was diversity in the sequences isolated from a single culture well; however, some culture wells were more diverse than others. For example sequences isolated from well 12.2 were derived from different V_H-genes (IGHV4-31*03 and IGHV3-53*01) and D_H-genes (IGHD4-4*01 and IGHD5-12*01) with further differences observed between allelic variants (IGHV4-31*03 and IGHV4-31*04), clearly showing the presence of more than one mRNA template. Sequences isolated from well 10.6 showed much less

Table 3-7. Heavy chain sequences isolated from patient RJ.

Sequence name	V-GENE and allele	Identity (%)	J-GENE and allele	Identity (%)	D-GENE and allele	CDR3 (amino acids)
10.6_H1	IGHV1-69*01	88.19	IGHJ6*02	77.42	IGHD3-10*01	GVYYGSTYGVDV
10.6_H2	IGHV1-69*01	88.54	IGHJ6*02	79.03	IGHD3-10*01	GVYYGSTYGVDV
10.6_H3	IGHV1-69*01	88.89	IGHJ6*02	79.03	IGHD3-10*01	GVYYGSTYGVDV
12.2_H2	IGHV4-31*04	91.75	IGHJ4*02	93.75	IGHD4-4*01	AREGTLQNFDY
12.2_H3	IGHV3-53*01	96.14	IGHJ4*02	79.17	IGHD5-12*01	ARDLYGGPGS
12.2_H5	IGHV4-31*03	94.50	IGHJ4*02	95.83	IGHD4-4*01	AREGTLQNFDY
19.2_H1	IGHV3-11*01	94.44	IGHJ4*02	85.42	IGHD3-22*01	ARDSDSSGPSDY
19.2_H2	IGHV3-11*01	94.44	IGHJ4*02	85.42	IGHD3-22*01	ARDSDSSGPSDY
19.2_H3	IGHV3-15*01	97.28	IGHJ4*02	95.83	IGHD3-10*01	TTEGSYFDY
19.2_H4	IGHV3-11*01	94.10	IGHJ4*02	85.42	IGHD3-22*01	ARGSDSSGPSDY
19.2_H5	IGHV3-30*03 or IGHV3-30*18 or IGHV3-30*19	93.06	IGHJ4*02	75.00	IGHD3-22*01	TKDSGYYDSAGYPGEY
19.6_H1	IGHV3-30*01	95.83	IGHJ5*02	94.12	IGHD4-23*01	ASTAEDYGDSWFDP
19.6_H2	IGHV5-51*01	95.83	IGHJ3*02	90.00	IGHD2-2*02	AGPGYCSSTRCSNTFDM
19.6_H3	IGHV3-30*01	95.83	IGHJ5*02	94.12	IGHD4-23*01	ASTTEDYGDSWFDP
19.6_H4	IGHV4-b*01	93.06	IGHJ3*02	94.00	IGHD2-2*01	ARALYCSSPRCIWGKNDAFDI
19.6_H5	IGHV5-51*01	94.44	IGHJ3*02	92.00	IGHD2-2*02	AGPGYCSSTRCSNTFDM
21.6_H1	IGHV4-31*01 or IGHV4-31*03	94.16	IGHJ2*01	98.11	IGHD4-23*01	ARHRRSSWYFDL
21.6_H2	IGHV4-31*01 or IGHV4-31*03	94.85	IGHJ2*01	98.11	IGHD4-23*01	ARHRRSSWYFDL
21.6_H4	IGHV4-31*01 or IGHV4-31*03	95.19	IGHJ2*01	98.11	IGHD4-23*01	ARHRRSSWYFDL
23.2_H1	IGHV1-18*01	96.88	IGHJ6*03	88.71	IGHD2-15*01	ARAVEGAVAAINYYFYMDV
23.2_H2	IGHV1-18*01	97.92	IGHJ6*03	88.71	IGHD2-15*01	ARAVEGAVAAINYYFYMDV
23.2_H5	IGHV1-18*01	96.88	IGHJ6*03	88.71	IGHD2-15*01	ARAVEGAVAAINYYFYMDV

Table 3-8. Light chain sequences isolated from patient RJ.

Sequence name	V-GENE and allele	Identity (%)	J-GENE and allele	Identity (%)	CDR3 (amino acids)
10.6_L1	IGKV1-5*03	91.40	IGKJ1*01	92.11	QHYSDSSGT
10.6_L3	IGKV1-5*03	91.40	IGKJ1*01	89.47	QHYSDSSGT
12.2_L3	IGKV3-20*01	96.81	No rearrangement found, no junction identified		
19.2_L1	IGKV1-39*01 or IGKV1D-39*01	95.34	IGKJ3*01	97.37	QQSYSTPFT
19.6_L2	IGKV3-20*01	97.87	IGKJ1*01	92.11	QQYGSSPGT
19.6_L4	IGKV3-20*01	97.87	IGKJ1*01	92.11	QQYGSSPGT
21.6_L1	IGKV1-39*01 or IGKV1D-39*01	94.98	IGKJ1*01	100.00	QQANSFPRT
21.6_L2	IGKV1-39*01 or IGKV1D-39*01	96.06	IGKJ3*01	96.97	QQSYTIL
21.6_L3	IGKV1-39*01 or IGKV1D-39*01	96.06	IGKJ3*01	96.97	QQSYTIL
23.2_L4	IGKV1-39*01 or IGKV1D-39*01	92.11	IGKJ1*01	97.22	QQSYSTLRT

diversity with all sequences matching the IGHV1-69*01, IGHJ6*02 and IGHD3-10*01 genes. However, percentage identity to the germline gene varied between the isolated sequences, suggesting that more than one cDNA template was present. Sequence variation could be induced during PCR amplification, although the amount of diversity observed is unlikely to be caused by experimental error alone. Amino acid alignment showed variation between the sequences at residue 1 (Q-E), residue 6 (Q-E), residue 44 (Q-L), residue 99 (T-A) and residue 126 (I-V) (Figure 3-8). Although not identical, these sequences were possibly clonal in origin, derived from a common B cell ancestor that had undergone SHM.

These sequences were also aligned to the closest germline sequence IGHV1-69*01 (Figure 3-8). CDRs are highlighted as these are the sites of greatest divergence from the germline sequence. Only alignment to the germline V_H genes is shown here, including the V_H-encoded N-terminus of CDR3. The remainder of CDR3 is encoded by the D (heavy chain only) and J genes. CDR3 shows the greatest length polymorphism and sequence diversity as a result of VDJ gene recombination, N-nucleotide addition and SHM. Mutations

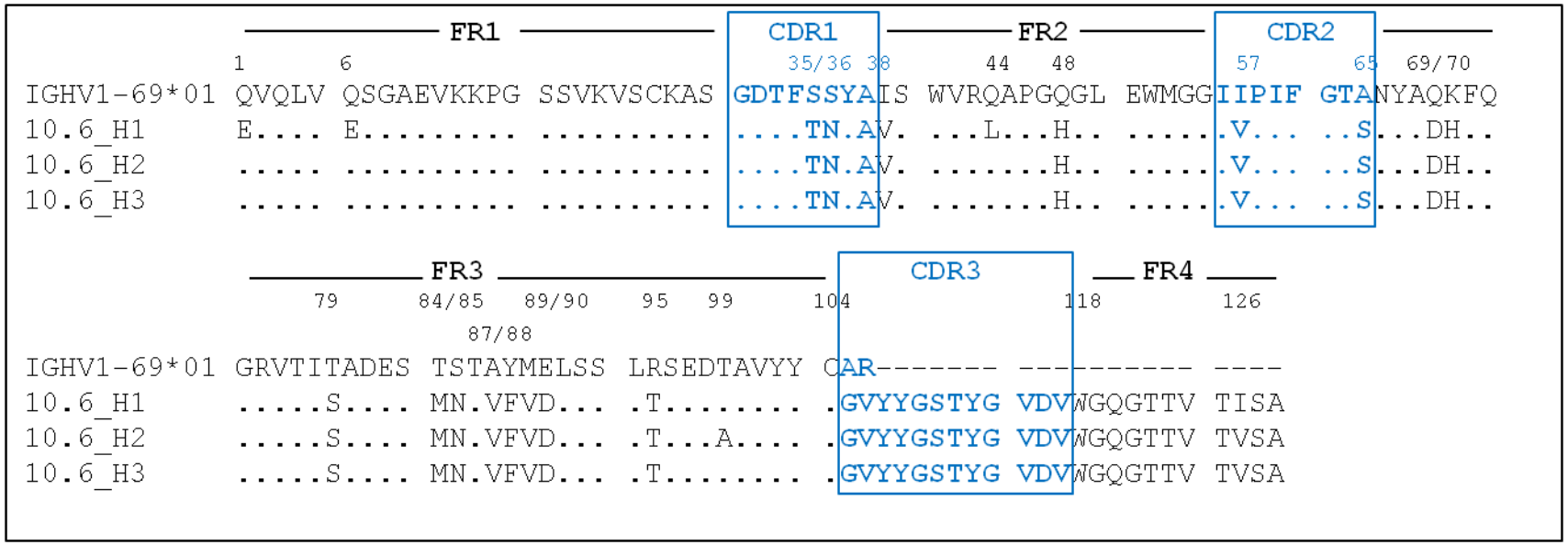

Figure 3-8. Sequence alignment of cloned VH domains and the closest germline relative. VH domain sequences were derived from a 5 day culture of 10 B cells isolated from patient RJ. The diversity between the isolates and the germline gene is a result of affinity maturation. The diversity between each isolate may suggest the presence of more than one cell in the initial culture. FRs are shown and CDRs are highlighted in blue. Dots denote amino acid sequence identity, dashes denote missing amino acids. The C-terminus of CDR3 and FR4 are provided by the J and D gene rearrangements. Residues are numbered according to the IMGT numbering system (Lefranc, 2011b). The conserved cysteine at 104 and conserved tryptophan at 118 are also numbered.

within the CDRs are likely to be those responsible for enhanced binding to the antigen, following affinity maturation; however a large number of mutations were also present in the FRs of these sequences. FRs are responsible for maintaining structural integrity of the IgG molecule and are less likely to tolerate mutations, yet a cluster of mutations within FR3 (residues 84-90) (Figure 3-8) differing from the germline gene was observed. Antibodies have been described that not only tolerate substitutions and insertions within FRs, but also show improved affinity due to these FR mutations (Krause *et al.*, 2011). However, other studies show that substitutions within FRs have little or no impact on antibody binding (Toran *et al.*, 2001). The FR3 mutations in 10.6 are conservative mutations and as such would not be expected to alter the structure of the IgG significantly but it remains to be seen if they improve binding affinity. The overall sequence identity remains high at >88% and all other sequences show germline sequence identity values >91% (Table 3-7).

Diversity was also seen between the light chains, all of which were found to be kappa class chains (Table 3-8). The light chains derived from sample 10.6 however, were very similar, containing identical IGKV germline genes (IGKV1-5*03) and CDR3. However, each isolated sequence matched the IGKJ1*01 gene with different percentage identity and alignment highlighted a single amino acid difference at residue 78 (A-V mutation between clone 10.6_L1 and 10.6_L3)(Figure 3-9). Although these transcripts were derived from cultures containing 10 B cells, it is possible that fewer cells were actually seeded, or functional. The small amount of diversity observed between sequences may have been caused by PCR induced mutations during the amplification and cloning process. It may therefore be possible to isolate the

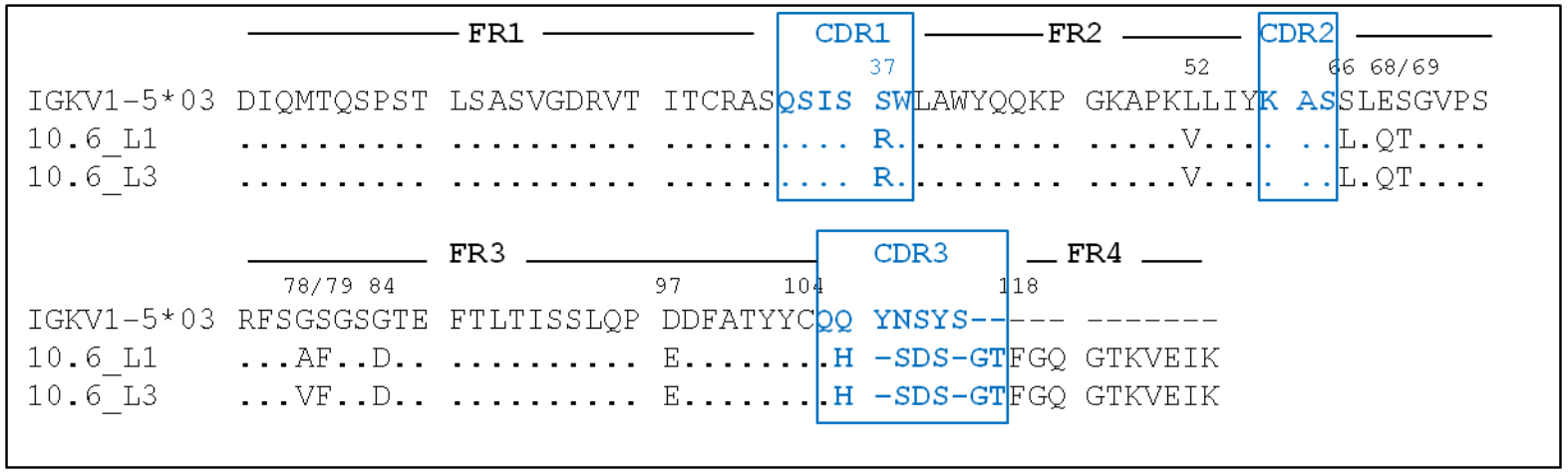

Figure 3-9. Sequence alignment of cloned VL domains and the closest germline relative. VL domain sequences were derived from a 5 day culture of 10 B cells isolated from patient RJ. The diversity between the isolates and the germline gene is a result of affinity maturation. FRs are shown and CDRs are highlighted in blue. Dots denote amino acid sequence identity, dashes denote missing amino acids. The C-terminus of CDR3 and FR4 are provided by the J and D gene rearrangements. Residues are numbered according to the IMGT numbering system (Lefranc, 2011b). The conserved cysteine at 104 and conserved phenylalanine at 118 are also numbered.

different VH and VL domains from cultures such as these and produce antibody from the paired chains. Other transcripts were also found to share common genes and have high levels of sequence identity e.g. transcripts derived from 21.6 and 23.3 (Table 3-7 and 3-8). However, pairing heavy and light chains would be problematic for sequences such as 21.6_L1 and 21.6_L2 that have different Jκ-genes and CDR3 sequences (Table 3-8). Correctly pairing heavy and light chains to produce a functional antibody would be even more problematic for transcripts derived from different germline genes e.g. 19.6. These sequences were isolated from a culture containing 100 B cells. Theoretically there could be 100 different heavy chains and 100 different light chains, resulting in 10,000 possible combinations of heavy and light chains. Generating and testing the functionality of all possible antibodies would be an extremely time-consuming and labour intensive process. Incorrect pairing of chains would result in many of the antibodies being non-functional and the cloning steps required may introduce mutations which abrogate or reduce antibody binding or affinity. Therefore it would be greatly advantageous to isolate VH and VL domains from single cell cultures to avoid the problem of matching heavy and light chains.

3.4.3 Cloning and characterisation of IgG genes from patient SK

In order to obtain cultures of single B cells, fluorescence activated cell sorting was attempted using a Beckman Coulter MoFlo Cell sorter (Beckman Coulter, High Wycombe, UK) to isolate $CD19^+$, $CD27^+$ and IgG^+ memory B cells from the PBMCs of patient SK. Unfortunately, this proved unsuccessful as the sorting was not very efficient and the cells did not produce measurable

amounts of IgG. This was believed to be caused by the failure of the B cells to expand. Therefore, cultures were seeded at approximately 1.3 per well using an automated system that deposits a specific volume of cell suspension into each well. Cultures receiving an average input of 1.3 cells were stimulated for 10 days. HCV-specific IgG-producing wells were identified by ELISA with E1E2 (genotype 1a) and peptide capture immunoassay to peptide I and peptide II. Peptide I corresponds to the E2 region encompassing amino acids 412-423; peptide II corresponds to the E2 region encompassing amino acids 434-446. Although rare in natural infection, antibodies to these epitopes are neutralising and it has recently been shown that antibodies against peptide II can augment neutralisation by antibodies targeting peptide I (Tarr *et al.*, 2012b). Those wells testing positive for HCV-reactive IgG and either peptide I or peptide II reactivity were selected for isolation and characterisation of the VH and VL domains. Eleven wells out of 384 (2.8%) produced HCV-specific IgG with reactivity to peptide I or II. Both primer set A (data not shown) and primer set B (Figure 3-10) were able to amplify VH and VL transcripts from these cultures. Of eleven HCV-specific IgG-positive samples, seven tested positive for the VH domain and all eleven samples tested positive for the VL domain. Primer set B includes kappa-chain and lambda-chain specific primer sets for amplification of the different chain types in separate PCR reactions. Surprisingly, kappa-specific and lambda-specific PCR products were obtained from all eleven samples. Individual B cells produce antibody containing either kappa or lambda-type light chains and mechanisms exist to exclude the expression of both (Dudley *et al.*, 2005). The observation of both chain types was therefore suggestive of more than one B cell in each culture. These

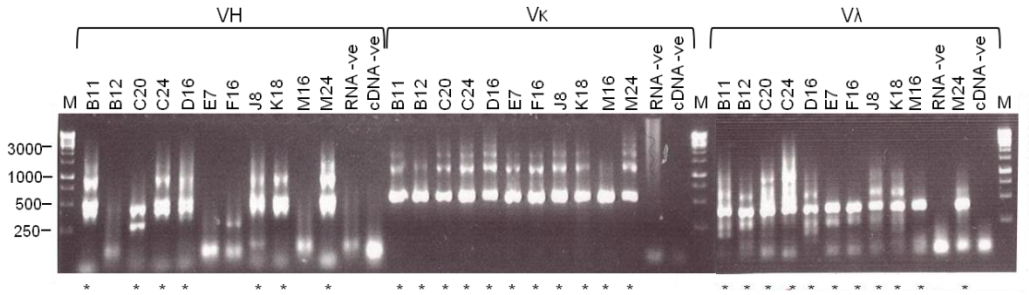

Figure 3-10. Amplification of IgG variable chain domains from B cell cultures. Primer set B was used to amplify VH, Vκ and Vλ from putative single cell cultures isolated from patient SK and cultured for 10 days to produce IgG. Positive reactions containing products at 400-500 bp are highlighted with *. RNA –ve, RNA negative controls; cDNA –ve, cDNA negative controls. Position of DNA marker, M, is shown in base pairs.

cultures were prepared by depositing B cells, diluted to a specific concentration, into a 384-well plate such that each well contained on average 1.3 cells. In reality this would result in a majority of wells containing one B cell but with the potential for a range of cell numbers, with some empty wells and others containing multiple B cells. It is possible that the wells selected for VH and VL analysis were those containing more than one cell that were able to produce sufficient IgG. The B cell culture protocol may have failed to stimulate the production of IgG from single B cells and the low recovery of ELISA/peptide immunoassay positive wells seems to support this hypothesis. The discrepancy between VL recovery (11/11 samples) and VH recovery (7/11 samples) is probably due to mutation of the primer binding site preventing PCR amplification of highly mutated genes.

The VH and VL PCR products isolated from these samples were cloned into the vector pGEM-T Easy and representative colonies from each sample were sequenced to identify their corresponding germline gene (Table 3-9 and 3-10). With the exception of B11, diversity of the VH domain between the sequenced isolates was limited. Transcripts from samples J8, K18 and M24 were derived

Table 3-9. Heavy chain sequences isolated from patient SK.

Sequence name	V-GENE and allele	Identity (%)	J-GENE and allele	Identity (%)	D-GENE and allele	CDR3 (amino acids)
B11_H2	IGHV3-48*03	91.32	IGHJ4*02	93.75	IGHD2-2*01	ARERPAEPYCFDY
B11_H6	IGHV3-23*01	94.44	IGHJ3*02	90.00	IGHD3-22*01	AKDYYYDSGAYSGPGAFNI
C20_H6	IGHV4-39*07	94.50	IGHJ3*02	88.00	IGHD6-19*01	ARDVRYCYSNGCYRNTFDT
J8_H2	Unproductive IgH rearrangement (stop codons)					
J8_H3	IGHV3-7*01 or IGHV3-7*03	91.32	IGHJ5*02	94.12	IGHD3-10*01	ARGFIAQLVRGYINNWFDP
J8_H6	IGHV3-7*01 or IGHV3-7*03	89.93	IGHJ5*02	92.16	IGHD3-10*01	ARGFIAQLVRGYINNWFDP
K18_H3	IGHV1-2*02	87.50	IGHJ4*02	85.42	IGHD2-2*01	ARKYCTSVYCDLRFFDS
K18_H4	IGHV1-2*02	91.67	IGHJ4*02	85.42	IGHD2-2*01	ARKYCTSVYCDLRFFDS
M24_H1	IGHV373*02	92.52	No rearrangement found, no junction identified			
M24_H2	IGHV3-73*02	93.20	IGHJ6*02 or IGHJ6*03	74.19	IGHD4-17*01	SRPPYGDTKADV

Table 3-10. Light chain sequences isolated from patient SK.

Sequence name	V-GENE and allele	Identity (%)	J-GENE and allele	Identity (%)	CDR3 (amino acids)
B11_K3	IGKV1-NL1*01	90.04	IGKJ2*01 or IGKJ2*02	86.84	QQYYSTPPVT
B11_K5	IGKV1-NL1*01	90.04	IGKJ2*01 or IGKJ2*02	86.84	QQYYSTPPVT
C20_L5	IGKV3-20*01	91.13	IGKJ4*01	100	QQYAGSLT
J8_K1	IGKV1-39*01 or IGKV1D-39*01	97.04	IGKJ4*01	91.67	QQSYSTLA
J8_K3	No rearrangement found (stop codons)				
J8_L1	IGKV3-20*01	97.16	IGKJ4*01	94.74	QQYGNSPSLT
J8_L2	IGKV1-12*01 or IGKV1-12*02	92.83	IGKJ3*01	94.74	QQSNSFPFT
K18 L4	IGKV1-9*01	95.34	IGKJ2*01	91.89	QQLDTYPNA
K18 L5	IGKV3-11*01	98.57	IGKJ2*01 or IGKJ2*02	91.43	QQRSNWPPS
M24_K1	IGKV1-39*01	95.57	IGKJ4*01	100	QQSYSTPLT

from the same germline gene and allele for V_H, D_H and J_H genes, with little diversity between sequences. Only one sequence was isolated from sample C20 precluding any analysis of clonal variation. However, upon closer analysis of the sequence alignments, differences between J8_H3 and J8_H6 were observed at the amino acid level (E6Q and S86P) (Figure 3-11). It is possible that these differences were due to PCR-induced errors or mutations introduced during cloning; however it is also likely that these represent descendants of a single B cell that have undergone affinity maturation and are not derived from single B cell cultures. The presence of very different sequences derived from sample B11 was further indication that samples contained more than one B cell. Sample B11_H2 was derived from the rearranged germline genes IGHV3-48*03, IGHJ4*02 and IGHD2-2*01 whereas B11_H6 was derived from IGHV3-23*01, IGHJ3*02 and IGHD3-22*01.

Interestingly when the analysis was extended to the VL products, all were found to contain kappa chain transcripts. No PCR products corresponding to lambda germline genes were observed. However, differing germline gene rearrangements between sequences was observed, providing further evidence of B cell cultures containing more than one cell. In contrast to the situation with the VH transcripts, VL sequences from sample B11 were derived from a single germline gene rearrangement of IGKV1-NL1*01 and IGKJ2*01 (or *02) (Table 3-10). Sequence changes were observed between B11_K5 and B11_K3 (A9F, V49A, C104W), (Figure 3-12) possibly due to affinity maturation of cells derived from the same B cell clone, or they may have been caused by PCR induced errors. The mutation from cysteine to tryptophan at

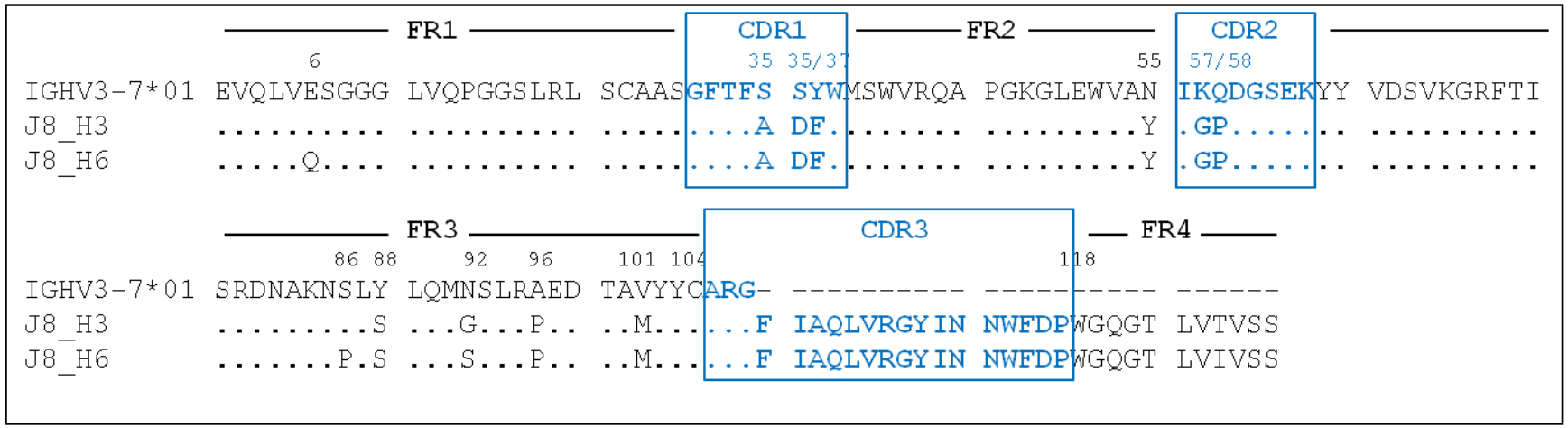

Figure 3-11. Sequence alignment of cloned VH domains derived from a 10 day culture of B cells isolated from patient SK. Cells were seeded at a density of 1.3 cells per well, to give a majority of wells containing a single cell. The diversity between isolated sequences and the germline gene is a result of affinity maturation. The diversity between each isolate may suggest the presence of more than one cell in the initial culture. FRs are shown and CDRs are highlighted in blue. Dots denote amino acid sequence identity, dashes denote missing amino acids. The C-terminus of CDR3 and FR4 are provided by the J and D gene rearrangements. Residues are numbered according to the IMGT numbering system (Lefranc, 2011b). The conserved cysteine at 104 and conserved tryptophan at 118 are also numbered.

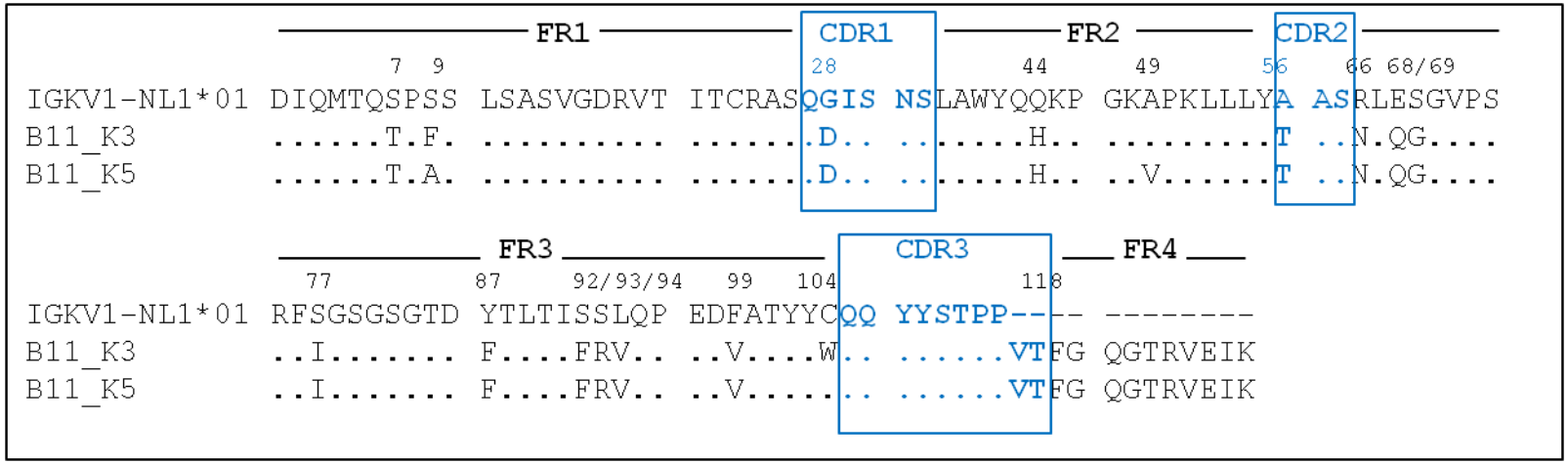

Figure 3-12. Sequence alignment of cloned Vκ domains derived from a 10 day culture of B cells isolated from patient SK. Cells were seeded at a density of 1.3 cells per well, to give a majority of wells containing a single cell. The diversity between isolated sequences and the germline gene is a result of affinity maturation. The diversity between each isolate indicates the presence of more than one cell in the initial culture. FRs are shown and CDRs are highlighted in blue. Dots denote amino acid sequence identity, dashes denote missing amino acids. The C-terminus of CDR3 and FR4 are provided by the J and D gene rearrangements. Residues are numbered according to the IMGT numbering system (Lefranc, 2011b). The conserved cysteine at 104 and conserved phenylalanine at 118 also are numbered.

residue 104 in B11_K3 was particularly interesting. C104 is one of two conserved cysteine residues in the VL domain, along with C23, that form an intra-chain disulfide bond responsible for stabilising the protein fold (Lefranc, 2011d). Although this IGK gene rearrangement was productive, loss of C104 may result in an unstable antibody with reduced function; however this could only be assessed by analysis of the expressed protein, which was beyond the scope of this work. This cysteine residue was present in all other sequences analysed. Other conserved features such as a conserved cysteine at residue 23, a conserved tryptophan at residue 41 and the conserved motif F/W-G-X-G at position 118-121 (Lefranc, 2011b) were also conserved amongst all sequences.

Sequenced transcripts from sample J8 and K18 had different VL domain germline gene rearrangements (Table 3-10) and were clearly not derived from a single B cell. Therefore none of the samples from which VH and VL had been successfully isolated appeared to be derived from a single B cell culture. This observation was supported by direct sequence analysis of PCR products, prior to cloning. Sequence chromatograms from several samples were found to contain mixed peaks (Figure 3-13), indicative of mixed template. The results described led us to believe that the cultures did not contain single B cells and, as such, cloning of paired light and heavy chain genes into an expression vector could be problematic. However, in order to develop the protocol and with the aim of producing antibody, cloning of these sequences into the expression vector was attempted.

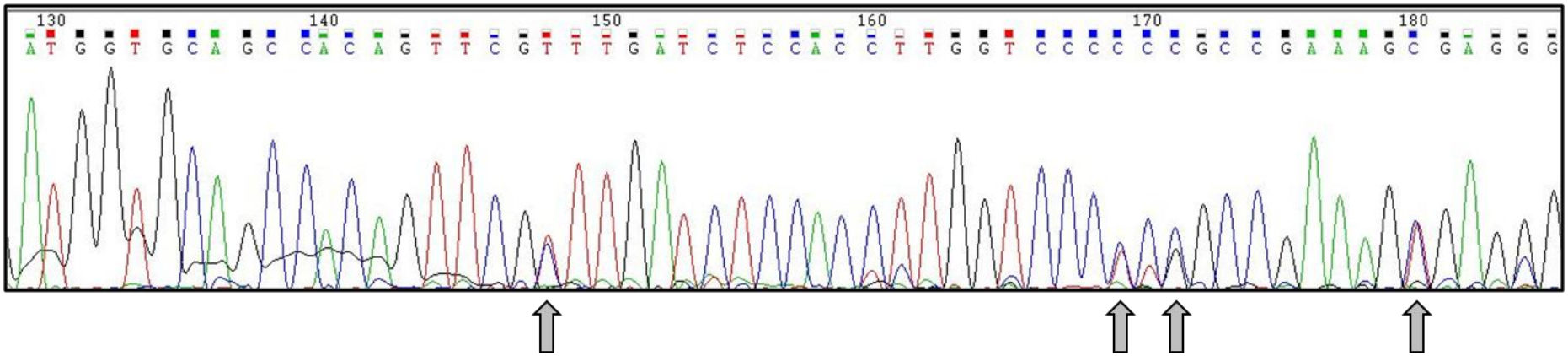

Figure 3-13. Sequence chromatogram of Vκ PCR product. The Vκ domain isolated from patient SK, sample J8, was sequenced directly following PCR amplification without cloning. Mixed peaks, indicative of multiple products within the sequencing reaction mix, were identified in the chromatogram and are highlighted by grey arrows.

3.4.4 Modification of the expression vector

The expression plasmid pDCORIG huIgG_1 was developed by Lindy Durrant and colleagues at Scancell Ltd. for the generation of human IgG_1 antibody containing T cell mimotopes in place of the CDRs (Metheringham *et al.*, 2009). This vector has been used as a DNA vaccine for the induction of high avidity, anti-tumour $CD8^+$ T cell responses (Pudney *et al.*, 2010). pDCORIG huIgG_1 (Figure 3-3) contains a heavy chain cassette, consisting of a murine variable region and a human IgG_1 constant region, under the control of the CMV IE promoter. The light chain cassette contains a murine variable region and a human kappa chain constant region, under the control of a separate CMV IE promoter. pDCORIG huIgG_1 contains restriction sites flanking each of the CDRs, for replacement with T cell mimotopes, as well as restriction sites at the 5' and 3' end of the variable domains, enabling replacement of the entire VH and VL domain. Due to incompatibilities between the expression vector and the primer sets used for PCR amplification, it was not possible to directly clone VH and VL PCR products into pDCORIG huIgG_1. The VH PCR products generated by primer set B contained an *Age*I restriction site at the 5' end and it was thought appropriate to use this site for cloning into pDCORIG. Therefore, *Age*I was introduced by site-direct mutagenesis into the heavy variable region of plasmid pDCORIG (Figure 3-4a). The PCR primer sets used did not amplify the leader sequence of the VH domain, therefore *Age*I was introduced at the 3' terminus of the native leader sequence within the vector. An additional *Age*I site in the constant heavy domain was also knocked out via site-directed mutagenesis.

3.4.5 Adaptation of VH and VL PCR products for expression vector cloning

VH and Vk PCR products generated with different primer sets contained different restriction sites at the 5' and 3' end. Primer set A generated VH domains with *Not*I/*Xba*I ends and Vk domains with *Bgl*II/*Nco*I ends (Table 3-1 and 3-2). Primer set B generated VH domains with *Age*I/*Sal*I ends (Table 3-3) and Vk with *Age*I/*Bsi*WI ends (Table 3-4). The chosen expression vector, pDCORIG, utilised *Age*I/*Afe*I restriction cloning for the VH domain and *Bam*HI/*Bsi*WI cloning for the Vk domain. Due to these incompatibilities, it was necessary to re-amplify the VH and VL products isolated with different primer sets, prior to expression vector cloning. Universal primers were designed to re-amplify VH and introduce *Age*I/*Afe*I restriction sites where necessary. Universal primers were also designed to introduce *Bam*HI/*Bsi*WI into those Vk products lacking the restriction sites (Figure 3-4b and Table 3-6). The sense primer for re-amplification of the kappa chain products also contained a 60 nucleotide sequence corresponding to the Ig leader sequence. Primers were designed such that the PCR products could be cloned into the expression vector without interrupting the reading frame. PCR products amplified from patient SK were modified by re-amplification with the appropriate primer pairs. VH from samples B11, C20 and K18 were successfully re-amplified (Figure 3-14a). Vk re-amplification to introduce the leader sequence was successfully completed for samples B11, C20, M24 and E7 (Figure 3-14b). Re-amplification was attempted on PCR products rather than individual plasmids to maintain any potential diversity between sequences. Sequence analysis of the re-amplified products showed the

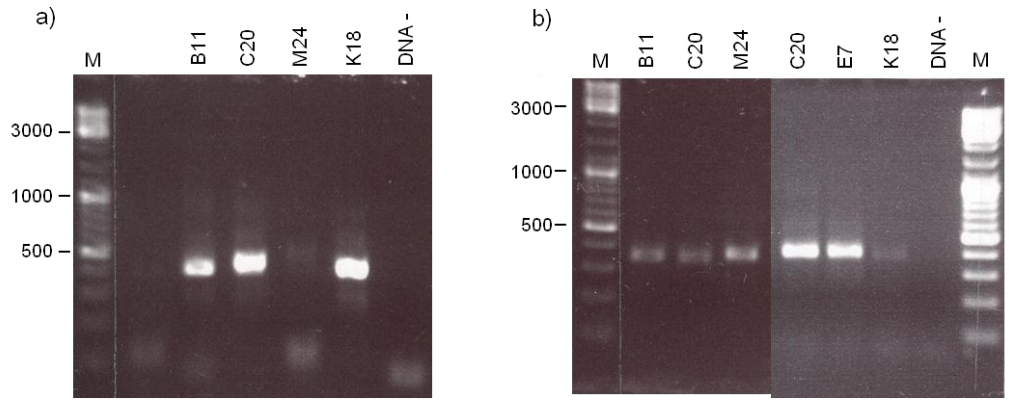

Figure 3-14. Generation of adapted IgG variable chain PCR products. (a) Heavy chains were re-amplified to introduce restriction sites *Age*I and *Afe*I at the 5' and 3' ends respectively. (b) Light chains were re-amplified to introduce the Ig leader sequence plus the restriction site *Bam*HI at the 5' end. Successfully re-amplified products form a band at 350-400 bp. DNA -, no template control. Position of DNA marker, M, is shown in base pairs.

presence of mixed peaks in the chromatogram suggesting that any diversity in the original product had been maintained.

3.4.6 Cloning of VH and Vk gene products into the expression vector

Paired heavy and light chain PCR products had been successfully isolated and adapted for expression vector cloning from two samples, B11 and C20, both derived from patient SK. These products were cloned by restriction digest into the modified pDCORIG huIgG_1 expression vector beginning with VH cloning. VH from sample B11 was successfully inserted into the vector, generating plasmid pDCORIG::B11h2. Sequencing of the insert confirmed the introduction of a novel VH domain with identity to the IGHV3-23*04 germline gene. This matched one of the sequences originally isolated from sample B11, B11_H6 (Table 3-9), showing that the VH domain is maintained throughout

PCR amplification, adaptation and expression vector cloning. It was then hoped that Vk from sample B11 could be cloned into pDCORIG::B11h2 to generate a vector with paired heavy and light chains capable of producing antibody. Unfortunately, despite several attempts, no vector containing the B11 Vk domain was obtained. Vk from sample C20 was successfully inserted into pDCORIG demonstrating that the cloning protocol worked. This plasmid pDCORIG::C20k4, was sequenced and found to contain a novel and functional Vk domain corresponding to the germline gene IGKV1-12*01. This did not correlate with the sequence originally identified in C20 (Table 3-10). However, the original sequence was derived from primer set A, whereas subsequent analyses were carried out on the Vk PCR product amplified with primer set B. It is possible that these primer sets preferentially amplify different genes. It is also important to remember that the products were obtained from cultures probably containing more than one B cell. Only a single sequence from sample C20 was isolated during the initial analysis and this may not have been sufficient to cover the full diversity of sequences within the sample. Due to time constraints it was not possible to continue with the expression vector cloning and, as such, no vector containing VH and Vk domains for antibody expression were generated; however, using the protocol described it should be possible to achieve this with sufficient time.

3.5 Discussion

Methods for isolating antigen-specific IgG^+ memory B cells from infected or recovered individuals can provide important insights into the natural immune

response to infection or vaccination. In order to apply these techniques to studies of HCV, PBMCs were isolated from patients suffering a chronic HCV infection but who were known to generate a strong antibody response. IgG^+ memory B cells were isolated as they provide a record of past exposure to antigen and are able to proliferate upon stimulation leading to the production of ASCs producing high affinity antibody (McHeyzer-Williams *et al.*, 2012). Using a memory B cell stimulation method described by Pinna *et al*. (2009), approximately 30% of single B cell cultures were found to proliferate and produce IgG; however this is only slightly lower than that reported in the literature where 37% of single B cell cultures were activated. Reduced proliferation and expansion is an intrinsic property of some B cells as they enter the memory B cell pool (Pinna *et al.*, 2009) and therefore it may not be possible to recover antibody from all cultured cells. Previous studies cultured IgG^+ memory B cells for 10 days to allow cell proliferation and antibody production (Pinna *et al.*, 2009). We compared B cells cultured for 5 days or 10 days and found no difference between the number of cultures testing positive for IgG production. It was however, noted that stimulating cultures for longer than 10 days reduced the ability to recover IgG-specific cDNA; of eleven cultures producing HCV-specific IgG, all were PCR negative for the VH and VL domains (data not shown). This may have been due to exhaustion of the B cells, such that IgG was still present within the culture medium and detectable by ELISA, but the B cell was no longer secreting IgG and therefore not expressing IgG-specific mRNA.

Analysis of the IgG producing B cell cultures from patient SK (Table 3-9 and 3-10; Figure 3-11 to 3-13) showed that single B cells were not present in these

cultures. This raised the possiblity that the method did not stimulate IgG production from single B cell cultures and that the only IgG-producing cultures yielding detectable IgG were those containing more than one B cell. Seeding more than one B cell is possible with the automated distribution method used but should result in the majority of wells containing a single cell. The low recovery rates (2.8% wells were IgG positive) coupled with the presence of more than one template in the wells from which cDNA was recovered would seem to suggest sub-optimal stimulation of these single cell cultures. Additional factors have been tested for their ability to stimulate memory B cells, but have met with limited success (Pinna *et al.*, 2009). Although CD40L was found to reduce IgG production from memory B cells (Pinna *et al.*, 2009), when combined with IL-21, CD40L was able to stimulate memory B cells expressing B cell lymphoma proteins (Kwakkenbos *et al.*, 2010). Therefore there may be alternative stimulants that could be used to improve the stimulation of single B cell cultures. An alternative explanation is that the ELISA was not sensitive enough to detect the IgG produced by a single cell. However, the detection range for single cell cultures was 74-170 ng/ml IgG (Figure 3-7), within the range of IgG production observed in single cell cultures by Pinna *et al*., (2009). It was estimated that the lower limit of detection for the ELISA was 5 ng/ml IgG (R. Urbanowicz, personal communication) and therefore, it is likely that this assay was sufficiently sensitive to detect the IgG produced by single cells. It seems more likely that there was a failure to sufficiently stimulate the single cells in these cultures.

Of the IgG-secreting B cell cultures recovered, only a small percentage were HCV E1E2-specific. This resulted in the needless culturing of B cells that

would ultimately be discarded. Recent studies have used a pre-culture selection method to isolate HIV gp140-specific (Scheid *et al.*, 2009) or gp120-specific IgG^+ memory B cells (Scheid *et al.*, 2011; Wu *et al.*, 2011). It is feasible that a similar pre-enrichment step using HCV E1E2 selection prior to culturing could reduce the number of non-specific cultures recovered. For patient RJ, only 10 samples (10.7%) tested positive for HCV-specific IgG production and IgG-specific mRNA, while only 1 sample (1.0%) from patient RH tested positive. This is despite culturing multiple B cells in each culture well and probably represents differences in the antibody specificity and quality of memory B cell production between patients. For patient SK, VH and VL domains were recovered from seven HCV-reactive IgG positive samples out of 384 B cell cultures (1.8%). Even if recombinant antibody were generated from all these positive samples, the rate of antibody recovery is low. This is not uncommon; indeed Scheid *et al*. (2011) recently reported the recovery of 576 antibodies from a starting population of 1.5×10^5 IgG^+ memory B cells cultured as single cells, a recovery rate of approximately 0.38%.

When considering the success of the PCR amplification protocol, both methods were capable of recovering VH and VL-specific products. In the initial analysis of samples derived from patients RJ and RH, IgG specific cDNA was amplified with primer set A only. However, in later experiments, including those with samples from patient SK, both primer sets were able to amplify VH and VL. The initial failure of primer set B may be due to the use of HotStar Taq DNA polymerase in the PCR reaction. This enzyme was used in the published literature (Tiller *et al.*, 2008; Wardemann *et al.*, 2003). However, following the failure of the original protocol, LongAmp Taq was

substituted for HotStar Taq to correspond with the protocol employed for primer set A. LongAmp Taq DNA polymerase may be more sensitive and therefore better able to amplify from the low copy number cDNA present in the sample. LongAmp Taq was therefore used in all subsequent experiments. Primer set A has a reported PCR recovery rate of 30-80% (Wang & Stollar, 2000); primer set B is slightly lower at 30-60% recovery rate (Tiller *et al.*, 2008). Recovery rates from cultures of 10 or 100 B cells using primer set A correlate well with those reported previously. Recovery rates using primer set B are more difficult to assess as only those cells known to be producing HCV-specific IgG were screened. Recovery of VL (Vκ or Vλ) was 100% (Figure 3-10). Recovery of VH was 63.6% (7/11) (Figure 3-10), similar to the upper end of reported recovery rates. This analysis is confounded however, by the presence of more than one B cell in most of these cultures and, in particular, the finding that Vκ and Vλ could be amplified from the same culture. Tiller *et al*. (2008) also report the amplification of Igκ and Igλ from a single well but with a frequency of only 5%. Approximately half contained functionally rearranged light chains from both alleles (Tiller *et al.*, 2008; Wardemann *et al.*, 2003). Although it is therefore possible that some of the wells containing both Vκ and Vλ contained only a single B cell, the high number of 'double positive' wells makes it extremely unlikely that this was the case. It is also worth considering the possibility that primer sets were mixed up, contaminated or inadvertently combined. Primer set B is, in fact, a combination of primers taken from three separate papers (Koelsch *et al.*, 2007; Tiller *et al.*, 2008; Wardemann *et al.*, 2003). In an attempt to improve the recovery of VH and VL domains from a variety of different germline genes, some non-specific binding

may have occurred. However, this does not escape the fact that some, if not all, of the samples obtained from patient SK contained more than one B cell, as evidenced by the recovery of transcripts with differing germline gene rearrangements (Table 3-9 and 3-10) and the detection of mixed template in the sequence chromatogram (Figure 3-13).

The failure of some wells to test positive for IgG VH and VL PCR products may be due to the presence of primer site mutations. Improved recovery of highly mutated VH and VL genes has been reported with a primer set that uses sense primers located further upstream within the leader region (Scheid *et al.*, 2011). Neither primer set provides complete coverage of all possible germline genes and alleles. Overall primer set B (Table 3-3 to 3-5) provides greater coverage than primer set A (Table 3-1 and 3-2), as it includes primers to a greater variety of alleles, particularly for the Vλ chains, although primer set B does not contain IGHV2 gene-specific primers. However, a recent study found that IGHV2 was not used in the B cell repertoire of patients suffering from acute resolving or chronic HCV (Racanelli *et al.*, 2011). It is interesting to note that none of the VH domains isolated from patient RJ with primer set A belonged to the IGHV2 gene family (Table 3-7). Therefore the lack of IGHV2-specific primers in set B may not be detrimental; however one cannot exclude the possibility that failure to amplify VH and/or VL chains from all samples was due to a lack of the appropriate gene or allele-specific primer. This problem may be overcome by the use of 5' rapid amplification of cDNA ends (5' RACE). This method amplifies unknown mRNAs by adding a guanine (G) tail to all single-stranded cDNA which is then replicated by the use of an oligo-dC adaptor (Ozawa *et al.*, 2006). 5' RACE was attempted but

was unsuccessful at recovering IgG-specific PCR products. Failure of this method was probably caused by the length of the oligo-dC adaptor (65 nucleotides) and its high melting temperature, making optimisation of the annealing step problematic.

An alternative explanation for reaction failure may be RNAse contamination of the samples. Great care was taken to exclude RNAses from the work environment and RNAse inhibitors were used in the B cell lysis buffer, in accordance with published protocols (Tiller *et al.*, 2008; Wardemann *et al.*, 2003). However, one experiment was lost due to RNAse contamination of the FBS. This prompted us to use certified RNAse- and DNAse-free FBS for all B cell cultures and to introduce a wash step prior to lysis. The B cell wash buffer also contained RNAse inhibitors. These amendments were sufficient to overcome RNAse-induced culture losses and highlight the importance of maintaining an RNAse free environment at all stages of B cell handling and mRNA isolation.

The initial stages of method development were carried out on cultures of either 10 or 100 B cells. Although it was possible to isolate IgG-specific PCR products from these cultures, it was felt that cloning paired light and heavy chain variable domains from such samples would be problematic. Increasing the number of B cells in a culture increases the number of possible heavy and light chain combinations required to make a productive antibody. In order to generate all possible combinations it is necessary to carry out a large amount of cloning. This can be a time-consuming process when one considers that from a culture containing only 10 B cells, there are theoretically 100 different combinations of heavy and light chains. The cloning process may result in the

generation of random reassortments containing novel combining sites that ultimately prove to be unproductive when expressed as whole antibody. It is also important to consider that every time an amplicon is subjected to PCR and cloning or sub-cloning, the chance of introducing artificial errors is increased. These errors may be mistaken for affinity matured antibody binding sites. LongAmp Taq was used in all amplification steps described and, due to its higher fidelity compared to HotStar Taq, should introduce fewer PCR-induced mutations. However, in some of the sequences analysed it is possible that differences were caused by PCR mutation. Direct sequencing analysis of the cDNA isolated from a single cell is the most accurate way to assess the level of affinity maturation. The use of cDNA isolated from cultures containing multiple B cells may also give a misleading picture of the abundance of certain antibody specificities if individual B cells express variable levels of IgG mRNA (Wang & Stollar, 2000). Therefore single B cell cultures present an ideal system in which to interrogate the antibody response to a given pathogen. Unfortunately, in our hands it has not yet been possible to obtain IgG-producing B cell cultures that we are confident contain a single B cell.

Despite only obtaining VH and VL domains from cultures that contained more than one B cell, attempts were made to clone these products into an expression vector. The two different primer sets used and the expression vector, pDCORIG huIgG_1, were all obtained from different sources and, as a result, there were incompatibilities in the restriction sites used. Therefore, it was necessary to adapt some of the PCR products and the expression vector prior to cloning. Figure 3-4 shows the modifications that were made. The restriction sites *Age*I and *Afe*I were chosen for cloning of the VH domains. VH

domains amplified with primer set B contained *Age*I at the 5' end (Table 3-3) and therefore this adaption was not required. However, it was necessary to replace the *Sal*I site at the 3' end with *Afe*I. VH domains amplified with primer set A required adaptation to replace *Not*I and *Xba*I at the 5' and 3' end respectively (Table 3-1). Due to the presence of an *Age*I site at the 5' end of all VL domains amplified with primer set B, it was necessary to clone the heavy chain first to avoid cutting pDCORIG huIgG_1 more than once. The restriction sites *Bam*HI and *Bsi*WI were chosen for cloning of the VL domains. Vκ domains amplified with primer set B contained *Bsi*WI at the 3' end (Table 3-4) and therefore this adaptation was not required. VL products amplified with primer set A contained *Nco*I at the 3' end (Table 3-2). Therefore it was necessary to adapt these products prior to cloning. Attempts were made to adapt the expression vector such that a new restriction site was introduced into the kappa chain cassette, at the 3' end of the leader sequence, thus maintaining the leader sequence needed for expression of the antibody. However, this proved unsuccessful and the decision was taken to further modify the kappa-type light chain PCR products by the introduction of the leader sequence at the 5' end (Figure 3-4b). The leader sequence was taken from the expression vector and contained a *Bam*HI site for direct restriction cloning into the vector. Although cloning adapted VH and VL (Vκ) chains into pDCORIG huIgG_1 was successful, no single vector containing paired heavy and light chains was obtained and this method has a number of drawbacks which would need to be rectified to improve the efficiency and accuracy of antibody cloning and expression. The additional adaptation steps were time consuming and greatly increase the likelihood of introducing PCR-induced errors into the VH and VL

sequences which may reduce the affinity of the resulting antibody. To overcome this problem the restriction sites within the primers could simply be altered to match those within the vector and the PCR products could be cloned directly into pDCORIG huIgG_1. Cloning of the Vκ domain into the expression vector was complicated by the need to introduce the leader sequence into the PCR product itself. It had been hoped that the restriction site *Bsr*GI could be introduced into the kappa chain leader sequence by altering the nucleotide sequence (from AGTCCA to TGTACA), without disrupting the amino acid sequence. Unfortunately this would have resulted in the introduction of another *Age*I site. Therefore kappa chain PCR products were re-amplified to include the leader sequence and cloning was carried out with the original *Bam*HI and *Bsi*WI sites contained within the vector. However, such re-amplifications were not always successful, the process was time-consuming and may introduce PCR-induced errors. Using a method akin to that used for cloning of the heavy chain variable domain would be preferable to avoid additional PCR steps. pDCORIG huIgG_1 was designed as a vector for the production of mouse/human chimeric antibodies engineered to contain T cell mimotopes (Metheringham *et al.*, 2009) (Figure 3-3) and as a result contains a murine leader sequence. It is therefore important to consider the compatibility of the mouse leader sequence for translation of the human variable domain protein. Comparison of the murine sequence within pDCORIG huIgG_1 to human variable domain leader sequences deposited in Genbank shows only 33% nucleotide sequence similarity. This could reduce production of a fully human IgG_1 from this vector and may require replacement of the leader sequences. However, this would require the vector to be more significantly

redesigned, something which was beyond the scope and the time frame of this project. Alternatively the primer sets, which do not currently allow for amplification of the leader sequence, could be modified. The primer set described by Scheid *et al.* (2011), for improved amplification of mutated genes, uses primers located within the leader sequence for both rounds of amplification. Therefore it should be possible to design primer sets capable of amplifying the variable domain including its leader sequence, if necessary.

Attempts at cloning into the expression vector were only partially successful, as paired heavy and light chains from a single culture could not be introduced into the same vector. It was also noted that the recovery of colonies containing the gene of interest was very low. In each case, approximately 20 colonies were screened for the human IgG VH or Vκ insert, yet only a single colony was identified containing the gene of interest. All other colonies contained the original murine variable domain. The cloning protocol would need to be more efficient if clonal diversity is to be maintained and assessed and if different light and heavy chain pairings are to be generated. Although this could be considered a flaw of the cloning method, it may not be a significant issue if single B cell cultures can be obtained and combining functional heavy and light chains from multiple sequences is no longer necessary. However, failure to replace the original murine variable domain is still an issue. Due to the inefficient recovery of DNA from gel extracted vector backbone, the vector was column purified following restriction digest. This would not remove the original murine variable domain and it appears that this was preferentially re-inserted into the vector rather than the human variable domain. Due to the similar size of the murine and human fragments, their ligation into the vector

would be a stochastic event. However, it may be possible to replace the murine variable domain with a stuffer fragment of around 10 bp in length. A small fragment would be more efficiently removed by column purification, increasing the likelihood of inserting the human variable domain during the subsequent ligation step. A plasmid stock containing such stuffer fragments could be used to increase the efficiency of the cloning protocol. In order to improve the speed and accuracy with which Ig variable domains can be cloned from patient B cells and expressed in mammalian cells, several modifications to the primer sets and the expression vector are necessary. These would greatly streamline the process and reduce the potential for introducing errors into the Ig gene sequences.

The expression vector pDCORIG huIgG_1 was chosen as it allows expression of a complete IgG_1 molecule from a mammalian cell culture system. IgG was isolated from patient B cells, rather than IgM, as it is more useful from a therapeutic point-of-view. IgM generally has low affinity for its antigen but due to its pentameric nature, has high avidity. IgM is produced first in response to a pathogen and needs to respond quickly to a variety of antigens. As a result it can be more polyreactive (Schroeder & Cavacini, 2010). Therefore it may bind strongly to the target antigen but does not represent the most affinity matured Ig. In contrast, IgG is produced later during infection by recombination of Cγ with affinity matured V_H genes. The IgG_1 subclass is the most abundant subclass in the serum (Schroeder & Cavacini, 2010) and is able to bind all three FcγR types triggering effector functions such as ADCC and CDC (Jiang *et al.*, 2011). The role played by Fc-mediated effector functions in HCV infection is poorly understood. Both acute and chronic phase sera can

mediate ADCC via binding to E2 expressed on the surface of infected cells (Natterman *et al.*, 2005) and several E2-specific MAbs can induce CDC of E2-expressing cells (Machida *et al.*, 2008). Mechanisms to evade the antiviral effects of ADCC and CDC have also been reported (Machida *et al.*, 2008; Maillard *et al.*, 2004). Therefore expressing antibodies as IgG_1 molecules may provide an opportunity to investigate the Fc-mediated effector functions of anti-HCV antibodies. However, in some cases effector functions are not desirable and in these situations antibodies expressed as IgG_4 molecules are used due to their lower affinity for FcγRI and FcγRII and lack of binding to FcγRIII (Jiang *et al.*, 2011). The expression vector is also available with IgG_4 constant domains (Lindy Durrant, personal communication) for the generation of antibodies with reduced effector functions. The production of whole Ig, rather than Fab fragments or scFv, more accurately replicates the *in vivo* structure and function of the antibody. ScFv may not bind antigen with the same affinity as the whole IgG molecule, which also has a longer half-life in serum and is able to mediate effector functions via the Fc region. The use of mammalian cells for antibody production would also ensure that the Ig molecules are correctly glycosylated. Glycans have been implicated in correct folding and stabilisation of recombinant Igs (Liu *et al.*, 2006) as well as in binding to FcγRs (reviewed in Jiang *et al.*, 2011). The vector pDCORIG contains both the heavy and light chain expression cassettes; therefore transfection of only a single plasmid is required for successful expression. This is in contrast to other protocols that use separate plasmids for the expression of heavy and light chains (Scheid *et al.*, 2011; Tiller *et al.*, 2008; Wardemann *et al.*, 2003). Co-transfection of multiple plasmids may generate

transfectants containing only a single plasmid and therefore expressing only one Ig chain. Stable transfection of pDCORIG into mammalian cells can be selected by the addition of zeocin (Metheringham *et al.*, 2009). Therefore this system should, in the future, provide a suitable system for the continued production of large amounts of whole Ig of known specificity.

Although the VH and VL domains isolated were not from single B cells, is it possible to make any observations or draw any conclusions from the data? Out of 31 functional VH sequences isolated, 14 (45%) were from the IGHV3 gene family. The remainder were from the IGHV1, IGHV4 and IGHV5 gene families (Table 3-7 and 3-9). This pattern of IGHV gene usage was similar to that observed in a study of chronically infected patients (Racanelli *et al.*, 2011), although IGHV4 was more abundant in this study (22% vs. <10% in Racanelli *et al.*). IGHV2 gene usage was not found in either study nor was the use of IGHJ1 genes in memory B cells of chronically infected patients. In fact, the use of IGHJ1*01 was found to be significantly associated with spontaneous resolution of HCV infection (Racanelli *et al.*, 2011) so it is unsurprising that it was not found in either of the chronically infected patients studied here. The IGHD2-2*01 gene and allele was also found more frequently by Racanelli *et al*. in chronically infected patients compared to spontaneous resolvers, an association that they found to be significant. It is perhaps interesting to note that IGHD2-2*01 was used in 13% of the transcripts isolated in this study (Table 3-7 and 3-9).

Several studies of antibody responses to HIV and influenza have been carried out with this method. Most of these studies report that antibodies binding to a defined antigenic epitope are derived from a very few related germline gene

sequences that undergo extensive affinity maturation via SHM (Grandea *et al.*, 2010; Scheid *et al.*, 2009; Scheid *et al.*, 2011). Applying this method to the study of HCV could greatly improve our knowledge of the most highly conserved and immunogenic domains *in vivo*. The method can be adapted to isolate antibodies targeting any part of E2, and could be used to identify antibodies binding to E1. As such, it may help to expose novel antibody-binding epitopes that have yet to be identified. Interrogation of the total antibody response may also help to elucidate the importance of the total antibody pool in reducing viral spread. Indeed, Scheid *et al.* (2009) found that no single antibody targeting HIV gp140 was able to neutralise virus infectivity. However, a combination of several antibodies displayed neutralising activity similar to serum neutralisation levels. They postulate that rather than a single MAb, the total antibody pool may be responsible for reducing virus spread. Therefore studying the total antibody repertoire may represent the *in vivo* situation more accurately and provide a more rational approach to vaccine design. Overcoming the problems encountered during the development of this method could be critical to further our understanding of the role played by nAbs in the outcome of HCV infection.

4 Analysis of acute phase serum samples from a cohort of patients in Egypt

4.1 Introduction

4.1.1 Natural history of HCV infection in Egypt

It is estimated that approximately 2% of the world population is chronically infected with HCV. There is however wide geographic variation in levels of infection with less than 1% of the population in Northern Europe infected with HCV compared to 15-20% of the population in Egypt (Alter, 2007). The high prevalence of HCV in Egypt is largely due to widespread anti-schistosomal treatment programmes established in the 1920s to reduce the burden of *Schistosoma spp*. infection in Egypt. These treatment programmes consisted of a series of intravenous injections with tartar emetic (potassium antimony tartrate) and continued until the 1980s when the advent of oral therapies for the treatment of schistosomiasis rendered this approach obsolete. On average between 6 and 9 anti-schistosomal injections were administered per patient in a treatment regimen carried out over several weeks. Large numbers of patients were treated simultaneously regardless of age or treatment stage. Reusable injection equipment was often used and observations at the time suggest that sterilisation procedures were either omitted or insufficient to fully sterilise equipment (Frank *et al.*, 2000). This resulted in large numbers of people becoming infected with HCV over many years leading to the high prevalence of HCV seen in Egypt today. Infection patterns today correlate with age with a significant bias towards individuals aged 20 years and above (Frank *et al.*, 2000). Genotype 4 is the most common genotype circulating within the

Egyptian population today (Dusheiko *et al.*, 1994; Frank *et al.*, 2000; Ramia & Eid-Fares, 2006), consistent with the theory of a local epidemic derived from a common transmission source (Pybus *et al.*, 2003; Tanaka *et al.*, 2004) i.e. antischistosomal treatment programmes. Thus the Egyptian population provides an interesting study population. Due to the large pool of HCV-positive individuals, new infections, particularly of healthcare workers, are common and provide a good opportunity to identify acutely infected patients. Healthcare workers who obtain a needle-stick injury and close contacts of known HCV-positive individuals are followed closely facilitating the collection of patient samples during the early stages of infection.

The majority of studies into HCV infection and treatment have focussed on genotype 1 and, to a lesser extent, genotypes 2 and 3 as these genotypes predominate in Europe and North America (Simmonds, 2004); genotype 4 has been less well studied. However, the current standard therapy (ribavirin and peglyated IFN-α) has only limited efficacy against genotype 4; response rates are estimated at around 48% (Poynard *et al.*, 2003). Therefore studies of genotype 4 infections are important to increase our understanding of the disease and to aid in the development of better therapies.

4.1.2 Role of the antibody response during the acute phase of infection

Much remains unknown about the antibody response to HCV during the acute phase of infection. This is partly due to the difficulty of identifying patients during the acute phase when the disease is largely asymptomatic (Alter, 2007). However, as discussed in section 1.8.2.2, studies of the acute

phase indicate that resolution of infection is associated with an antibody response which is both high titre and broadly neutralising. Two recent studies demonstrated that spontaneous resolvers exhibit a high titre nAb response during the acute phase which declines slowly over several years. Low titre or absent nAb responses were seen in patients with chronic HCV infection during the acute phase of the infection followed by an increase in nAb titres during the chronic stage of infection (Dowd *et al.*, 2009; Pestka *et al.*, 2007). Studies have also shown that nAb drives envelope sequence evolution leading to the development of escape variants which may allow viraemia to persist. This selective pressure continues throughout chronic infection leading to the development of a diverse viral population within the host (Dowd *et al.*, 2009; von Hahn *et al.*, 2007).

4.1.3 Characteristics of the antibody response and viral sequences during acute infection

Following infection with HCV, antibody is detectable in the serum within 7 to 8 weeks (Pawlotsky, 1999) although it may take up to 3 months to develop. The antibodies produced may target both structural and non-structural viral proteins (Thimme *et al.*, 2008); however only those targeting E1 and E2 have neutralising activity (reviewed in Edwards *et al.*, 2012). Studies have shown that antibodies may persist for several years after the initial infection in patients with a chronic infection and in those who resolve the infection (Dowd *et al.*, 2009; Pestka *et al.*, 2007). IgG class antibodies are typically of low avidity during the acute phase of infection and increase in avidity with time (Gaudy-Graffin *et al.*, 2010; Kanno & Kazuyama, 2002).

HCV genomes appear in the plasma within a few days of infection and peak 6-10 weeks after infection (Bowen & Walker, 2005). Viral sequences are under selective pressure from the cellular and humoral arms of the immune response leading to continuous evolution of the virus. Transmission dynamics during early infection are not well defined; however studies suggest that infection is typically established by one founder virus, although in rare cases more than one founder virus may seed the infection. This is followed by a genetic bottleneck event characterised by a reduction in viral diversity (Bull *et al.*, 2011; Wang *et al.*, 2010). The rate of sequence evolution during the acute phase of infection is limited by strong selective pressures (Bull *et al.*, 2011).

In contrast to this description of a single founder virus, the phenomenon of mixed acute infections has recently been described. In a study by Smith *et al*., (2010), 80% of a study cohort were infected by more than 1 viral subtype during the acute phase of the disease. The dominant variant detected at each time point changed rapidly in some individuals and highly divergent clades within each subtype were also observed. In addition, the viral load within each host was highly dynamic, with changes in viral load of 2 log or greater being described as 'yo-yo' viral load (Smith *et al.*, 2010).

The HCV genome exhibits the greatest sequence diversity in the envelope proteins, E1 and E2, and the NS5A protein. The 5' non-coding region (NCR) and core regions are the most highly conserved regions (Simmonds, 2004). Sequencing of the 5'NCR is commonly used for genotyping of viral isolates although more recently other regions of the genome such as E1 and NS5B have become more popular due to their higher discriminating power (Murphy *et al.*, 2007). Studies of viral sequences from sequential serum samples obtained

from chronically infected individuals show rates of change across the whole genome to be 1.44 x 10^{-3} nucleotide changes per site per year (Simmonds, 2004). The envelope glycoproteins exhibit greater variability; within E1 the rate is 7.1 x 10^{-4} nucleotide changes per site per year (Simmonds, 2004; Smith *et al.*, 1997) and HVR1 within E2 also exhibits considerable sequence diversity. The high viral replication rate (10^{12} virions per day) (Neumann *et al.*, 1998), coupled with the error-prone nature of the RNA polymerase, results in a highly diverse viral population within chronically infected individuals.

Viral load is an important tool for monitoring disease progression and response to treatment. It is typically reported as viral genome copies per ml. Viral load testing can be carried out by amplification of the 5'NCR which, due to its high level of conservation, can be quantified without prior knowledge of the viral genotype (Murphy *et al.*, 2007). Although there is no definitive answer as to what is considered a high or low viral load, values greater than 1 x 10^6 copies per ml are generally considered to be high (Moreno-Otero, 2005). According to the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), which is part of the United States' National Institute of Health (NIH), most patients with chronic hepatitis C have viral loads between 100,000 (10^5) and 10,000,000 (10^7) copies per ml (Bell *et al.*, 2010). Reduction in viral load is used to measure the efficacy of anti-HCV treatment. Any reduction seen in the absence of standard treatment (pegylated IFN-α and ribavirin) would indicate spontaneous resolution of the infection.

4.1.4 Principles of DNA fingerprinting

DNA fingerprinting is a widely used technique in the fields of forensic science and paternity testing. The most commonly used method for DNA fingerprinting employs short tandem repeat (STR) markers (Butler, 2006). STRs (also known as simple sequence repeats or microsatellites) are polymorphic genetic loci containing a repeated nucleotide sequence of between 2 and 7 nucleotides. The number of nucleotides per repeat sequence is consistent for the majority of repeats within a locus; however the number of repeat units at a locus is variable resulting in alleles of many different lengths (Edwards *et al.*, 1991; Klintschar *et al.*, 1999). STRs make up approximately 3% of the human genome and are thought to arise due to slippage of the DNA polymerase during DNA replication (IHGSC, 2001). Over 1 million STR sequences have been identified in the human genome and most are found in the non-coding regions of DNA such as introns and intergenic regions (Ellegren, 2004). Therefore analysis of STRs does not provide any information about an individual's predisposition to disease or any phenotypic characteristics. STRs are inherited DNA motifs and when several STRs are analysed simultaneously, a pattern which is unique to that individual emerges. It is this feature that can be exploited to match unknown samples to an individual. Many different combinations of STR loci can be used to 'fingerprint' a sample. Early commercially available kits employed just three STR loci; however the probability of obtaining a random match with a sample is 1×10^{-3}. That means there is a 1 in 1000 chance that a sample will be wrongly matched to a test sample. Currently available commercial DNA fingerprinting kits use anywhere up to 16 STR loci with random match probabilities reported at between 1.2 x

10^{-18} and 7.2 x 10^{-19} (Butler, 2006). This reduces the chances of an incorrect match to 1 in 1 trillion. Ideally, the STR loci chosen should be located on separate chromosomes and therefore segregate independently during meiosis. This reduces the likelihood of underestimating or overestimating the frequency of a given profile within a population and therefore helps to avoid incorrect matches between samples (Butler, 2006).

4.1.5 Analysis of STR loci

Polymorphic STR loci are amplified by PCR and the size of the product determined. Traditionally this was done by gel electrophoresis of radiolabelled PCR products, however advances in the use of fluorescent dyes and the development of capillary electrophoresis has enabled the process to become automated and therefore quicker and simpler (Lazaruk *et al.*, 1998). Each STR locus is amplified by a pair of primers, one of which is labelled at the 5' end with a fluorescent dye. The resulting fluorescently labelled products are loaded onto a capillary electrophoresis instrument where the negatively charged products migrate towards an anode at different speeds on the basis of their electrophoretic mobility (Frazier *et al.*, 2000). As the products migrate past a detector the fluorescent dye emits light of a specific wavelength which is recorded. A size standard and allelic ladder are run alongside the samples and the information gathered is used to determine the size of the STR loci (Lazaruk *et al.*, 1998). The use of fluorescent dyes with different excitation and emission spectra for each primer set enables several different STRs to be analysed simultaneously. The information gathered is presented in an electropherogram and provides information about the size of the PCR product,

the number of alleles present at a given locus and whether an individual is homozygous or heterozygous at that locus.

Recent advances have been made in the development of reduced size STR loci or 'mini-STR' loci. Typically, amplification of STR loci generates amplicons of between 100 and 450 bp. Mini-STRs are generated from the same loci but with PCR primers designed to bind closer to the repeat region of the STR. This generates smaller PCR products and is particularly useful for the analysis of damaged or degraded DNA samples (Butler *et al.*, 2003; Chung *et al.*, 2004).

4.1.5.1 STR loci used in this study

Three STR loci were chosen for this study. The loci are all widely used in commercial DNA fingerprinting applications and have been well studied and characterised (Butler, 2006; Butler *et al.*, 2003). They have also been studied within an Egyptian population and shown, individually, to have high discriminating power (Ahmed *et al.*, 2001; Klintschar *et al.*, 1999). The 'mini-STR' versions of all three loci were amplified and therefore the STR sizes referred to are the reduced size STRs (Butler *et al.*, 2003).

4.1.5.1.1 THO1

The THO1 locus is found within the tyrosine hydroxylase gene (Genbank ref. D00269) which is involved in the conversion of tyrosine to dopamine. It is located at position 11p15.5 on chromosome 11. This STR is a tetranucleotide TCAT repeat (Butler, 2006). The number of possible alleles ranges from 3 to 14 and the STR size varies from 51 to 98 bp (Butler *et al.*, 2003).

4.1.5.1.2 vWA

The vWA locus is a polymorphism located at position 12p13.3 on chromosome 12 (Genbank ref. M25858). This particular locus is associated with von Willebrand factor, a blood glycoprotein required for clotting. This STR is composed of two tetranucleotide repeats [TCTG][TCTA] (Butler, 2006). The allele range for this STR is 10 to 25 with a size range of 88 to 148 bp (Butler *et al.*, 2003).

4.1.5.1.3 D21S11

The D21S11 STR (Genbank ref. AP000433) is a polymorphism located at position 21q21.1-q21.2 on chromosome 21. It is composed of two tetranucleotide repeats [TCTA][TCTG] (Butler, 2006). The number of alleles for this STR varies from 24 to 38.2 with a possible size range of 153 to 211 bp (Butler *et al.*, 2003).

4.2 Aims of the work presented in this chapter

This work was originally designed to analyse the evolution of viral envelope sequences under pressure from the antibody response during the acute phase of infection in a cohort of patients. However initial analyses of the viral sequences raised questions about the origin of these serum samples. It was therefore deemed necessary to carry out further investigations to confirm that sequential samples were taken from the same patient, as reported. This chapter details those investigations; the analysis of the viral sequences and the antibody

content of the samples as well as the DNA fingerprinting of patient samples and control samples showing.

4.3 Materials and methods

4.3.1 Serum samples

Serum samples were provided by Sanaa Kamal (Ain Shams University, Cairo). Serum was obtained from 34 patients reported to be suffering from an acute HCV infection. Serum was collected at three separate time points during the acute phase of infection (providing a total of 102 samples). Information about sampling time for some patients was also provided by Sanaa Kamal and is given in Table 4-4 (page 271). Samples were labelled with three letters e.g. AWM, to maintain patient confidentiality followed by a single digit e.g. AWM1, AWM2. These numbers correspond to the order in which samples were taken. Two sample tubes were labelled KJO1, KJO2 and KJO3. To distinguish between the two tubes these were relabelled as KJO1a, KJO1b, KJO2a, KJO2b, KJO3a and KJO3b. Each sample contained approximately 2 ml of serum which was aliquoted into 1.5 ml screw cap tubes and stored at -80°C.

Control serum samples were kindly provided by Emma Thomson (MRC Virus Research Unit, Glasgow). These were paired serum samples derived from two patients with reported mixed acute infection. Samples 1611 and 2134 were taken from a single patient, 34, on 27/09/2002 and 21/10/2002 respectively. These will be referred to hereafter as samples 34a and 34b respectively. Samples 5645 and 7087 were taken from a single patient, 31, on

01/09/2003 and 27/10/2003 respectively. These will be referred to hereafter as samples 31a and 31b respectively. Both patients were believed to be suffering mixed acute infection with genotype 1 and genotype 4 virus (Emma Thomson, personal communication).

4.3.2 Isolation and analysis of viral sequences

4.3.2.1 RNA extraction

RNA was extracted from 200 µl of serum using a QIAamp MinElute Virus Spin Kit (Qiagen, Crawley, UK) with carrier RNA according to the manufacturer's instructions. Briefly, 25 µl Qiagen protease, 200 µl serum and 200 µl Buffer AL (containing 28 µg/ml carrier RNA) were mixed by pulse-vortexing and incubated at 56°C for 15 minutes to allow lysis of proteins and inactivation of RNAses to occur. To the lysate, 250 µl of ethanol absolute (200 proof) (Sigma Aldrich, Gillingham, UK) was added and incubated at room temperature for 5 minutes. The lysate was loaded onto a QIAamp MinElute column and centrifuged at 6000 x g for 1 minute to allow adsorption of the nucleic acids onto the silica-gel membrane. Samples were washed with 500 µl Buffer AW1 at 6000 x g for 1 minute, 500 µl Buffer AW2 at 6000 x g for 1 minute and finally 500 µl ethanol absolute (200 proof) (Sigma Aldrich, Gillingham, UK) at 6000 x g for 1 minute to ensure that all contaminants were removed. RNA was eluted from the column in 100 µl AVE Buffer by centrifuging at 20,000 x g for 1 minute and stored at -80°C. Due to the presence of carrier RNA in the samples it was not possible to quantify the eluted RNA.

4.3.2.2 cDNA synthesis with Moloney's Murine Leukaemia Virus (MMLV) RT and random hexamers

Ten microlitres of RNA, 2.5 µl 5x MMLV RT buffer (Fermentas, St. Leon-Rot, Germany) and 25 ng of random hexamers (Invitrogen, Paisley, UK) were heated to 75°C for 5 minutes and snap cooled on ice. To this, 25 µl 5x MMLV RT Buffer, 30 mM dNTPs (Fermentas, St. Leon-Rot, Germany), 200 units MMLV RT (Fermentas, St. Leon-Rot, Germany) and DEPC-treated H_2O to a final volume of 25 µl were added. Samples were incubated at room temperature for 10 minutes followed by heating to 42°C for 1 hour. cDNA was stored in non-stick 0.6 ml tubes (Alpha Laboratories, Eastleigh, UK) at -20°C.

4.3.2.3 PCR amplification of the 5'NCR

The 5'NCR was amplified from cDNA which had been generated with MMLV and random hexamers. One microlitre of cDNA was used as a template in a total reaction volume of 25 µl containing 2.5 mM dNTPs (Fermentas, St. Leon-Rot, Germany), 0.3 units HotStar Taq DNA Polymerase (Qiagen, Crawley, UK) and 5 pM each of primers NCR FWD and NCR REV. The sequence of all primers used for analysis of viral sequences is given in Table 4-1. The 5'NCR was amplified in a 3-step cycling reaction consisting of 95°C for 15 minutes, 60 cycles of 95°C for 30s, 60°C for 30s and 72°C for 60s followed by 72°C for 2 minutes. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2.

Table 4-1. Primers used for analysis of viral isolates

Target	Function of primer	Genotype specificity	Primer orientation	Primer name	Primer sequence (5'-3')
5' NCR	PCR, sequencing	Universal	Sense	NCR FWD	GCGGAACCGGTGAGTACA
	PCR, sequencing	Universal	Antisense	NCR REV	ACTCGCAAGCACCCTATCAG
	Real-time qPCR	Universal	Sense	qHCV NCR F	GCGGAACCGGTGAGTACA
	Real-time qPCR	Universal	Antisense	qHCV Sc	*(fluorophore)*CGGCGCGAAAGGCCTTGTGGTACTGCGCCGX *(quencher+spacer)*ACTCGCAAGCACCCTATCAG
HVR1	Nested PCR 1st round	Universal	Sense	EOS	GGACGGGGTAAACTATGCAACAGG
	Nested PCR 1st round		Antisense	E1O_A	TCATTGCAGTTCAGGGCAGTCCTGTTGATG
	Nested PCR 2nd round, sequencing		Sense	EIIS_MOD	TGGGATATGATGATGAACTGG
	Nested PCR 2nd round, sequencing		Antisense	EIIA	CTGTTGATGTGCCAGCTGCCA
E1E2	cDNA synthesis, nested PCR 1st round	Genotype 4	Antisense	OAS4M	CACCAGCGGCTGAAGCAGCATTGA
	Nested PCR 1st round	Universal	Sense	EOS	GGACGGGGTAAACTATGCAACAGG
	Nested PCR 2nd round,sequencing	Genotypes 1,3,4,5	Sense	170gt1	CACCATGGGTTGCTCTTTCTCTATC
	Nested PCR 2nd round, sequencing	Genotype 4	Antisense	IAS4.31	GACAGTTATGCCTCGACCTGGGATACCATGAATATC
	Sequencing		(Internal) sense	EIIS_MOD	TGGGATATGATGATGAACTGG
	cDNA synthesis, nested PCR 1st round	Genotype 1	Antisense	OAS1a	GGGATGCTGCATTGAGTA
	Nested PCR 2nd round	Genotype 1	Antisense	IASGT1	TTACGCCTCCGCTTGGGATATGAGTAACATCAT

All primers were synthesised commercially by Eurofins MWG Operon, Germany.

4.3.2.4 Purification of the 5'NCR

Five microlitres of each PCR positive sample was purified with 10 units Exonuclease I (Fermentas, St. Leon-Rot, Germany) and 1 unit 'FastAP' Thermosensitive AP (Fermentas, St. Leon-Rot, Germany) by heating at 37°C for 15 minutes. Enzymes were inactivated by heating to 85°C for 15 minutes. Purified PCR product was sequenced with primers NCR FWD and NCR REV as described in 2.3.1.10.

4.3.2.5 Phylogenetic analysis of 5'NCR sequences

Sequencing reads were edited using Chromas 2.13 (Technelysium, Brisbane, Australia) and contiguous sequences were made from paired 5'NCR sequences using SeqMan II software (DNASTAR Inc., Madison, USA). Phylogenetic analyses were carried out with MEGA 4 software (Tamura *et al.*, 2007). Sequences were aligned and compared to reference genotype sequences 1a (Genbank accession number M62321), 1b (M96362, U01214, D10934), 2 (D16433, D13406), 2b (D10077), 3a (D17763, D28917), 3b (D11443), 4b (FJ462435), 4c (FJ462436), 4d (FJ462437, EU392172), 4f (EU392169, EU392174), 4g (FJ462432), 4k (EU392173, EU392171, FJ462438), 4l (FJ839870), 4m (FJ462433), 4n (FJ462441), 4o (FJ462440), 4p (FJ462431), 4q (FJ462434), 4r (FJ462439), 4t (FJ839869), 5 (L29585), 6a (Y12083) and 6b (D84262). The evolutionary relationship was inferred using the Neighbour-Joining method (Saitou & Nei, 1987) and evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004).

4.3.2.6 Nested-PCR amplification of HVR1

HVR1 was amplified from cDNA generated with MMLV RT and random hexamers in a nested-PCR reaction. Two microlitres of cDNA or first round PCR product was used as template in a total reaction volume of 20 µl containing 4 mM dNTPs (Fermentas, St. Leon-Rot, Germany), 1 unit LongAmp Taq DNA Polymerase (New England Biolabs, Hitchin, UK) and 4 pM of each primer. The first round was carried out with primers EOS and E1O_A in a 3-step reaction of 95°C for 30s, 35 cycles of 95°C for 10s, 55°C for 30s and 65°C for 1 minute, followed by 65°C for 10 minutes. The second round was carried out with primers EIIS_MOD and EIIA in a 3-step reaction as described using 45 cycles of amplification. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2.

4.3.2.7 Purification and sequencing of HVR1

PCR positive samples were purified on a Qiaquick PCR Purification Kit (Qiagen, Crawley, UK) as described in section 2.3.1.3. DNA concentration was quantified using a Nanodrop ND1000 Spectrophotometer and the products sequenced with primers EIIS_MOD and EIIA as described in 2.3.1.10.

4.3.2.8 cDNA synthesis with Thermoscript RT and gene-specific primers

To generate full-length E1E2 sequences from the viral RNA, cDNA synthesis was carried out with Thermoscript RT and a gene-specific primer. The genotype 4-specific primer OAS4M was used to analyse Egyptian study samples and control samples. The genotype 1-specific primer OAS1a was also

used to analyse the control samples. In the first step, 8.5 µl RNA, 10 pM primer and 20 mM dNTPs (Invitrogen, Paisley, UK) were heated to 65°C for 5 minutes to allow primer annealing. cDNA was synthesised by adding 5x Thermoscript RT Buffer (Invitrogen, Paisley, UK), 0.1 M DTT (Invitrogen, Paisley, UK), 40 units RNAse OUT (Invitrogen, Paisley, UK), 15 units Thermoscript RT (Invitrogen, Paisley, UK) and DEPC-treated H_2O to a final volume of 20 µl. The reaction mix was heated to 50°C for 1 hour followed by 55°C for 1 hour and finally 85°C for 5 minutes. RNA template was removed by addition of 2 units RNAse H (Invitrogen, Paisley, UK) and heated to 37°C for 20 minutes. cDNA was stored in non-stick 0.6 ml tubes (Alpha Laboratories, Eastleigh, UK) at -20°C.

4.3.2.9 Nested-PCR amplification of full-length E1E2

Two microlitres of cDNA or first round PCR product was used as a template in a total reaction volume of 20 µl containing 4 mM dNTPs (Fermentas, St. Leon-Rot, Germany), 0.5 unit LongAmp Taq DNA Polymerase (New England Biolabs, Hitchin, UK) and 4 pM of each primer. The first round was carried out with primer EOS and either genotype 4-specific primer OAS4M or genotype 1-specific primer OAS1a in a 3-step reaction consisting of 95°C for 1 minute, 35 cycles of 94°C for 15s, 50°C for 30s, and 65°C for 3 minutes followed by 65°C for 10 minutes. The second round was carried out with primer 170gt1 and either genotype 4-specific primer IAS4.31 or genotype 1-specific primer IASGT1 in a 3-step reaction as described using 45 cycles. PCR products were detected by agarose gel electrophoresis as described in 2.3.1.2.

4.3.2.10 Purification and sequencing of full-length E1E2

PCR positive samples were purified as described in 2.3.1.3. DNA concentration was quantified using a Nanodrop ND1000 Spectrophotometer. PCR products were sequenced with primers 170gt1, EIIS_MOD and either IAS4.31 (genotype 4) or IASGT1a (genotype 1) as described in 2.3.1.10.

4.3.2.11 Phylogenetic analysis of full-length E1E2 and HVR1 sequences

Sequencing reads were edited using Chromas 2.13 software (Technelysium, Brisbane, Australia) and contiguous sequences were made from the three E1E2 sequences and paired HVR1 sequences using SeqMan II software (DNASTAR Inc., Madison, USA). Phylogenetic analyses were carried out with MEGA 4 (Tamura *et al.*, 2007). All available HVR1 and E1E2 sequences were translated, aligned and trimmed to the same length resulting in a phylogenetic analysis based on a 270 bp fragment corresponding to E2 HVR1 and the E1 and E2 flanking regions. The evolutionary relationship was inferred using the Neighbour-Joining method (Saitou & Nei, 1987) and evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004). The same process was repeated with the inclusion of HVR1 sequences from reference genotype strains 1a (Genbank accession number M62321, EU155192), 1b (M96362, U01214, D10934), 2b (AY734982, AY734983) 3a (D17763, D28917), 4b (FJ462435), 4c (FJ462436), 4d (EU392172, FJ462437), 4f (EU392169, EU392174), 4g, (FJ462432), 4k (EU392171, EU392173, FJ462438), 4l (FJ839870), 4m (FJ462433), 4n

(FJ462441), 4o (FJ462440), 4p (FJ462431), 4q (FJ462434), 4r (FJ462439), 4t (FJ839869), 5 (NC009826), 6a (Y12083), and 6b (D84262).

4.3.3 Quantitative real-time PCR of the 5'NCR

In order to quantify the viral load of the samples quantitative PCR (qPCR) was carried out. qPCR was carried out with a Scorpion primer, a primer with a tail containing a fluorophore/quencher pair and a probe element attached to the 5' end. As the DNA strand is extended, the probe hybridizes to its target within the nascent DNA causing a structural change in the primer which results in fluorophore unquenching and emission of a fluorescent signal (Whitcombe *et al.*, 1999). The fluorescence levels correlate directly to the amount of specific PCR product. This method has advantages over the use of intercalating agents, such as SYBR green, as it is specific to the target gene and less likely to produce false-positives (Whitcombe *et al.*, 1999). One microlitre of cDNA, generated as described in section 4.3.2.2, was used as template and amplified with 10 pM primer qHCV NCR F and 5 pM of primer qHCV Sc (incorporating the Scorpion probe) in a 15 µl reaction containing 0.25 mM dNTPs (Fermentas, St. Leon-Rot, Germany) and 0.3 units HotStar Taq DNA Polymerase (Qiagen, Crawley, UK). Real-time PCR was carried out on an Mx4000 Multiplex Quantitative PCR System (Agilent Technologies, Stockport, UK) with cycling at 95°C for 15 minutes followed by 50 cycles of 95°C for 30s, 60°C for 30s and 72°C for 30s. Standard controls containing JFH-1 (genotype 2a) cDNA at concentrations ranging from 10^1 to 10^7 cDNA copies per µl were run in duplicate alongside the samples. The results were

given as cDNA copies per µl of input cDNA. This was converted to viral load (i.e. genome copies per ml of serum) by multiplying by 1250.

4.3.4 STR analysis of RNA-extracted serum samples

All STR analyses were carried out on the RNA extracted from patient serum samples. Although the QIAamp MinElute Virus Spin Kit (Qiagen, Crawley, UK) is designed to isolate viral nucleic acids from plasma and serum, it will isolate any nucleic acids contained within a sample. Sufficient human DNA was present within the RNA samples to be amplified by standard PCR techniques. The sequence of all primers used in STR analysis is listed in Table 4-2.

Table 4-2. Sequence of primers used for STR fingerprinting of serum samples

Target	Primer orientation	Primer name	Primer sequence (5'-3')	Modifications
THO1	Sense	THO1_fwd	CCTGTTCCTCCCT TATTTCCC	5' label VIC
	Antisense	THO1_rev	GGGAACACAGACT CCATGGTG	
vWA	Sense	vWA_fwd	AATAATCAGTATG TGACTTGGATTGA	5' label NED
	Antisense	vWA_rev	ATAGGATGGATGG ATAGATGGA	
D21S11	Sense	D21S11_fwd	ATTCCCCAAGTGA ATTGC	5' label PET
	Antisense	D21S11_rev	GGTAGATAGACTG GATAGATAGACGA	

Fluorescently labelled primers were synthesised by Applied Biosystems (Paisley, UK). Antisense primers were synthesised by Eurofins MWG Operon, Germany.

4.3.4.1 PCR amplification of STR loci

Each STR was amplified from serum-extracted RNA in a separate reaction. Five microlitres of RNA was used as template in a total reaction volume of 12.5 µl containing 2.5 mM dNTPs (Fermentas, St. Leon-Rot, Germany), 0.3 units HotStar Taq DNA Polymerase (Qiagen, Crawley, UK) and 2.5 pM each of STR-specific primers. THO1_fwd contains the fluorescent dye VIC at the 5' end. VIC has an emission spectrum of 538-554 nm. vWA_fwd contains the fluorescent dye NED at the 5' end. NED has an emission spectrum of 546-575 nm. D21S11_fwd contains the fluorescent dye PET at the 5' end. PET has an emission spectrum of 558-595 nm. A 3-step reaction of 95°C for 15 minutes, 45 cycles of 94°C for 60s, 57°C for 90s and 72°C for 30s, followed by 60°C for 30s was carried out. THO1 PCR products were diluted 1:5 in MB H_20 (Sigma Aldrich, Gillingham, UK). vWA and D21S11 PCR products were not diluted.

4.3.4.2 Analysis of PCR products on the ABI Prism 3130

Due to different levels of amplification of each STR it was necessary to mix the PCR products to produce final dilution factors of 1:100 (THO1), 1:20 (vWA) and 1:1.1 (D21S11). This was to ensure all PCR products were accurately analysed without producing offscale peaks (too much signal) or very low peaks (too little signal). Therefore, 0.5 µl of THO1 PCR product (already diluted 1:5), 0.5 µl of vWA PCR product and 9 µl of D21S11 PCR product were mixed. 0.5 µl of mixed PCR products was mixed with 12 µl formamide containing the Genescan 500 LIZ Size Standard (Applied Biosystems, Paisley, UK). The size standard contains 16 single-stranded DNA fragments ranging in

size from 35 to 500 nucleotides each labelled with the fluorescent dye LIZ. LIZ has an emission spectrum of 638-655 nm. Samples were run on the ABI Prism 3130 Fluorescent DNA analyser (Perkin-Elmer, Cambridge, UK) and the data analysed using Peakscanner 1.0 software (Applied Biosystems, Paisley, UK). Samples were analysed to determine the allele size (or sizes) of each STR locus. Sequential time-point samples from each patient were then compared on the basis of STR allele size to determine if they were derived from the same patient. Cluster analysis of STR allele sizes was carried out using Genemarker v2.2.0. software (Softgenetics, State College, USA) to generate a distance matrix based on the Euclidean distance between single samples.

4.3.5 Testing antibody content of serum samples

4.3.5.1 Total antibody testing

Antibody testing was carried out in the NHS Microbiology diagnostic facility at The Queen's Medical Centre, Nottingham on a HCV ADVIA Centaur XP Immunoassay System (Siemans, Camberley, UK). This is a third generation assay system with a diagnostic specificity of 99.9% and sensitivity of 100% (Denovel *et al.*, 2004). Samples were scored anti-HCV antibody positive if they had an anti-HCV score >11.0.

4.3.5.2 Antibody avidity testing

Antibody avidity testing was carried out by Joy Kean (Gartnavel General Hospital, Glasgow). Avidity testing uses ELISA to measure the strength of

IgG binding to HCV antigens in the presence or absence of a denaturing reagent such as urea. Results are determined by calculating the ratio of absorbance in the presence of urea to absorbance without urea, expressed as a percentage. Results are reported as an Avidity Index (AI) score. AI values greater than 45 indicate high antibody avidity and are associated with antibodies produced during the chronic stage of infection (Joy Kean, personal communication). An acute phase control sample and a chronic phase control sample were run alongside test samples giving AI values of 9 and 93 respectively.

4.4 Results

4.4.1 Genotyping viral sequences based on the 5'NCR

Initial analyses of the study cohort included genotyping viral isolates from all RNA-positive samples by sequencing of the 5'NCR (Figure 4-1). Although genotyping based on the 5'NCR is not always able to discriminate between genotypes 1 and 6 (Murphy *et al.*, 2007), it was expected that the viral isolates would be predominantly genotype 4 as this is the most common genotype in Egypt (Frank *et al.*, 2000; Ramia & Eid-Fares, 2006). Therefore sequencing of the 5'NCR should be sufficient for this study. Figure 4-1 showed that most of the isolates tested clustered with the genotype 4 reference sequences as expected. However several sequences clustered close to both genotype 3 and genotype 4 sequences. Due to the low discriminatory power of 5'NCR sequencing the genotype of these isolates could not be resolved. There were also some discrepancies in the clustering of genotype 2 sequences which were

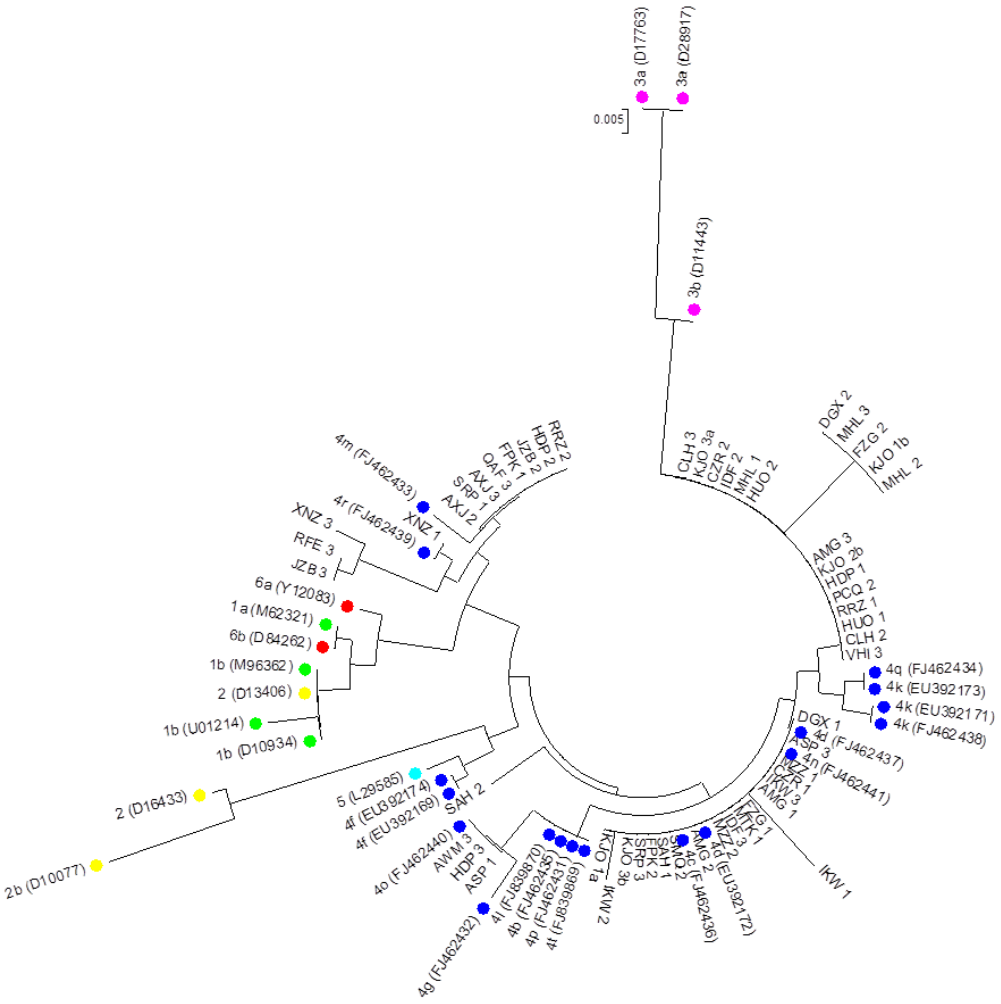

Figure 4-1. Genotyping of viral isolates on the basis of 5'NCR sequencing. Evolutionary history was inferred using the Neighbour-Joining method and evolutionary distances computed using the Maximum Composite Likelihood method. The percentage of replicate trees in which the associated taxa cluster together in the bootstrap test (500 replicates) is given next to the branches. Genotype reference sequences are highlighted with coloured circles: Genotype 1, green; Genotype 2, yellow; Genotype 3, purple; Genotype 4, dark blue; Genotype 5, light blue; Genotype 6, red.

spread throughout the tree. It has been reported that genotyping based on the 5'NCR is limited in its ability to distinguish some subtypes within genotypes 1, 2, 3, 4 and 6 (Murphy *et al.*, 2007). However, subsequent genotyping based on HVR1 alongside several reference sequences, clearly showed that the majority of viral isolates cluster with genotype 4 reference sequences, consistent with the epidemiology of HCV in Egypt. A small number of sequences clustered

with genotype 1 reference sequences. This genotype is also found in Egypt although it is less common (Ramia & Eid-Fares, 2006) (Figure 4-2). The 5'NCR could not be sequenced from samples AXJ1 and FZG3 due to the presence of more than one viral sequence resulting in a mixed chromatogram.

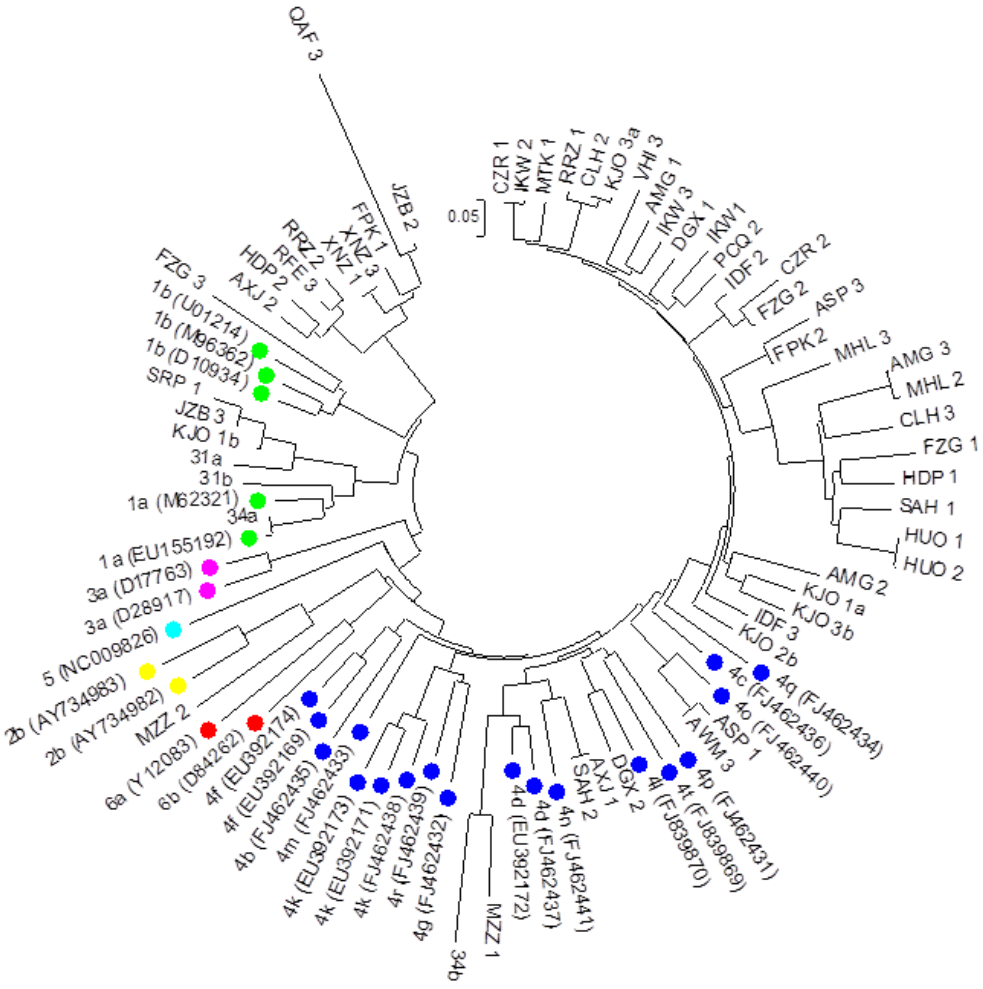

Figure 4-2. Genotyping of viral isolates on the basis of HVR1 sequencing. Evolutionary history was inferred using the Neighbour-Joining method and evolutionary distances computed using the Maximum Composite Likelihood method. The percentage of replicate trees in which the associated taxa cluster together in the bootstrap test (500 replicates) is given next to the branches. Genotype reference sequences are highlighted with coloured circles: Genotype 1, green; Genotype 2, yellow; Genotype 3, purple; Genotype 4, dark blue; Genotype 5, light blue; Genotype 6, red.

It was noted while examining the clustering of samples within Figure 4-1 that several of the sequences taken from the same patient clustered with different

subtypes. For example samples ASP1 and ASP3 showed homology to genotypes 4o and 4d respectively while FPK1 clustered with genotype 4m and FPK2 clustered with genotype 4c. Viral sequences isolated from the same patient during the acute phase of infection should show significant sequence similarity and therefore cluster together in a phylogenetic tree. Although genotyping on the basis of 5'NCR sequences is not ideal, the lack of clustering observed for several patients nevertheless raised the possibility that the triplicate samples were not derived from the same patient. This prompted us to undertake further analysis of viral isolates using a more informative region of the genome. HVR1 and flanking regions, with their greater sequence variability, were chosen to further analyse the relatedness and evolution of viral sequences within each patient.

4.4.2 Phylogenetic analysis of patient-derived viral isolates based on HVR1 sequencing

HVR1 within E2 shows significant sequence variability and evolves under pressure from the host immune response (Farci *et al.*, 2000; Ray *et al.*, 1999). Therefore it has much greater discriminating power in phylogenetic analyses of viral isolates. All viral isolates were aligned on the basis of the HVR1 sequence and the resulting phylogenetic tree (Figure 4-3) confirmed the suspicions raised by the alignment based on the 5'NCR (Figure 4-1). Only two out of the thirty-four patients tested were found to have viral sequences clustering together on the tree. These were samples HUO1 and HUO2 and samples XNZ1 and XNZ3, highlighted in Figure 4-3. HUO3 and XNZ2 contained no detectable viral RNA and therefore no sequencing could be

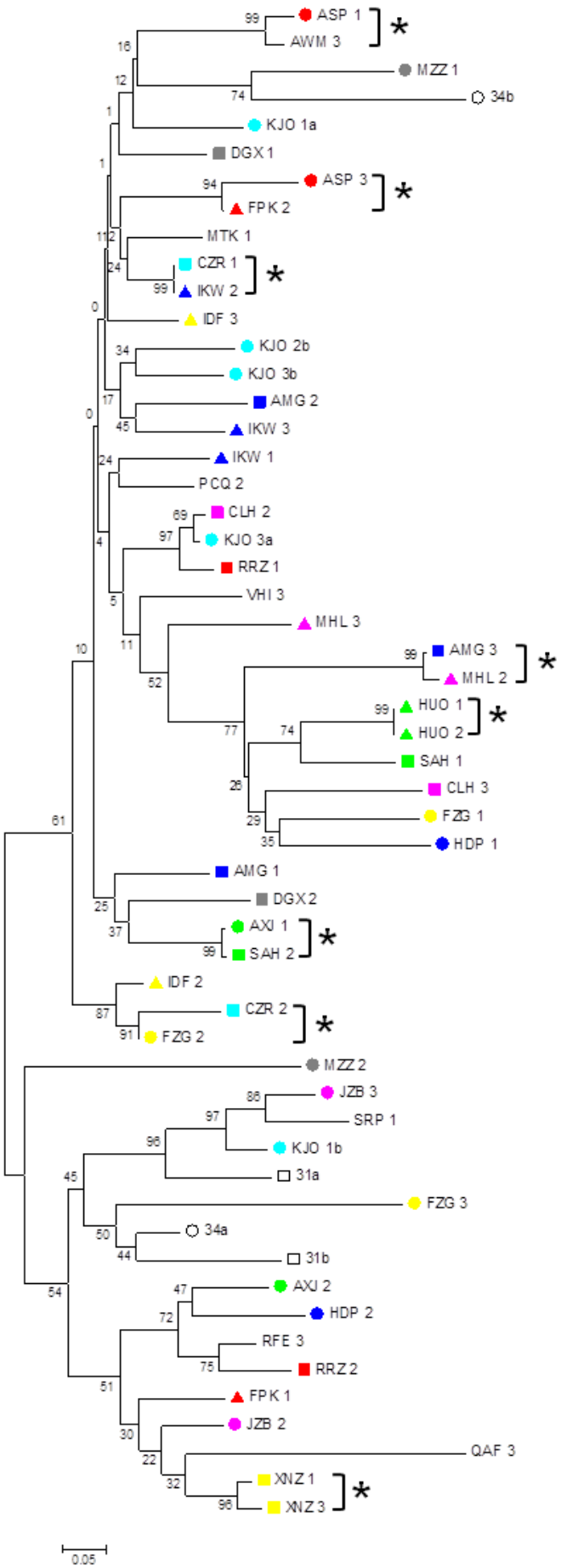

Figure 4-3. Phylogenetic analysis of viral isolates on the basis of HVR1 sequencing. Evolutionary history was inferred using the Neighbour-Joining method and evolutionary distances computed using the Maximum Composite Likelihood method. The percentage of replicate trees in which the associated taxa cluster together in the bootstrap test (500 replicates) is given next to the branches. Patients for whom more than one isolate was sequenced are highlighted with coloured symbols to aid identification of patient clusters. Those patients for whom only one sample contained nucleic acid are left blank. Clusters of interest that are discussed in the text are highlighted with an asterisk (*).

performed on these samples. Based on the hypothesis that infection is typically established by one founder virus (Bull *et al.*, 2011), epidemiologically linked virus strains would be expected to co-cluster within a phylogenetic tree. Samples isolated during the acute or window period of infection should cluster very closely together. All other samples however, were scattered randomly throughout the tree and did not cluster according to the patient ID. It was possible to pick out some virus clusters which could be derived from the same patient; for example, sequences from patients CZR2 and FZG2, AMG3 and MHL2, CZR1 and IKW2, ASP3 and FPK2, ASP1 and AWM3, and AXJ1 and SAH2, all clustered with 90% or greater bootstrap support. These are highlighted in Figure 4-3. It is possible therefore that these sample pairs were isolated from the same patient. Despite these putative groups of patient samples it was not possible to say with complete confidence that they were derived from a single patient. It was also not possible to determine the order in which these samples were taken as there is no reliable information regarding sampling time (see section 4.4.4). Therefore any study of virus evolution under the influence of the host immune response would be impossible and these samples were deemed to be of no further use to this study.

HVR1 could not be sequenced from HDP3, MHL1, SMQ2 and SRP3 either due to failure of the PCR amplification or sequencing reactions.

4.4.3 STR fingerprinting serum samples to identify patient clusters

There are two possible explanations for the phylogenetic observations made above; the first is that the triplicate samples are not derived from the same

patient and therefore are not epidemiologically linked. Alternatively, several of these patients may have developed a mixed acute infection, as suggested by Smith *et al.,* (2010). To distinguish between mixed acute infection and 'mixed-up' or contaminated samples the DNA content of the samples was analysed. To provide a positive control for the analysis, mixed acute samples from two patients, 31 and 34, were provided by Emma Thomson. These paired samples had previously undergone phylogenetic analysis based on HVR1 sequences but had not been subject to DNA fingerprinting analysis. STR fingerprinting of three discrete loci was carried out on the RNA preparations extracted from all samples and the data analysed using Peakscanner 1.0 software. This displays all STR loci as a peak trace; trace data for the positive control 34 alongside data representative of genotype matched and unmatched samples are given in Figure 4-4. Cluster analysis of STR loci sizes was subsequently carried out using Genemarker v2.2.0 software to generate a Euclidean distance tree of the inferred relationships between virus-containing samples and is shown in Figure 4-5.

Samples derived from the same individual should have identical patterns of STR locus size; however this was not the case for the majority of samples and samples reportedly from the same patient were found to have very different STR loci sizes as shown in Figure 4-4. Figure 4-5 also shows that the majority of samples did not cluster according to patient ID and therefore were not epidemiologically linked. The control samples from patients 31 and 34 clustered together according to STR analysis, as expected. HVR1 sequence

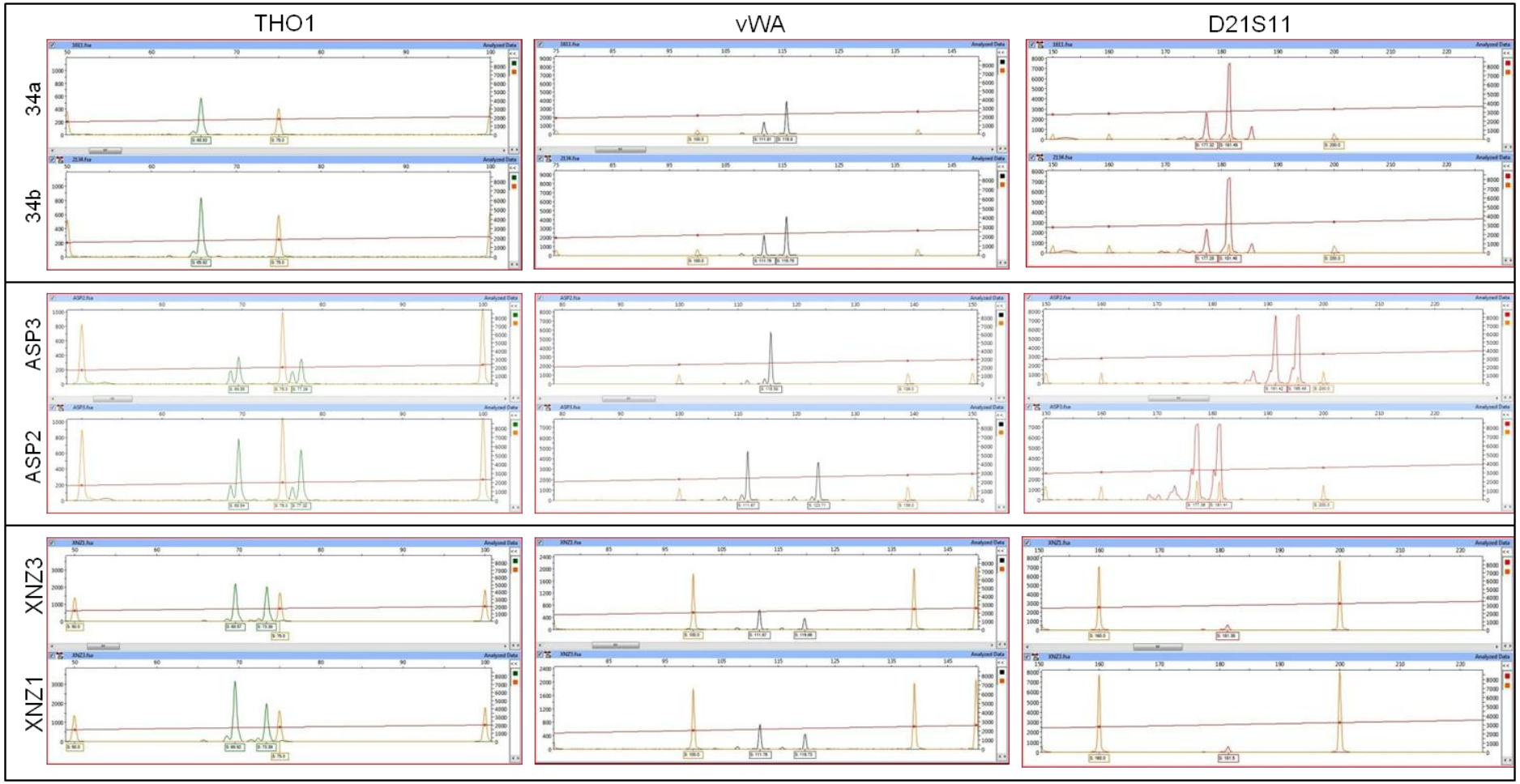

Figure 4-4. Representative trace files for STR locus analysis. See overpage for the full figure legend.

Figure 4-4. Representative trace files for STR locus analysis. Trace files generated in Peakscanner 1.0 software for the positive control samples, 34a and 34b show identical patterns for sequential samples. Paired samples from patient XNZ also possess identical trace patterns indicative of matched genotype. Paired samples from patient ASP have different trace patterns characteristic of an unmatched genotype. Size standards are shown in orange.

analysis was also carried out and confirmed the status of samples 34a and 34b as mixed acute phase samples (Figure 4-5). When included in the genotyping analysis, the samples were found to cluster with genotype 1 and genotype 4 isolates respectively (Figure 4-2). HVR1 sequence analysis of samples 31a and 31b appeared to also show a mixed acute infection (Figure 4-3) although both were of genotype 1a strains (Figure 4-2). This data confirmed the previous findings of the viral isolate analysis, proving that the grouped Egyptian samples were not derived from a single individual. The exceptions to these findings were patient samples HUO1 and HUO2 and samples XNZ1 and XNZ3. It is interesting to note that these samples were the only ones to contain viral isolates which clustered together (Figure 4-3), suggesting that sequencing of viral isolates can provide an indication of the reliability and origin of samples. This data also clearly showed that the majority of Egyptian samples were not taken from patients suffering a mixed acute infection.

Based on the clustering of viral sequences in Figure 4-3 it was possible to pick out some putative patient groups (listed in section 4.4.2.). In some cases the STR fingerprinting data and cluster analysis confirmed these patterns. For example, the samples CZR2 and FZG2, CZR1 and IKW2, AMG3 and MHL2, and ASP3 and FPK2 clustered based on both viral sequence (Figure 4-3) and human genotype (highlighted in Figure 4-5) suggesting that the HCV infection in these paired samples was epidemiologically linked. CZR2 and FZG2 also

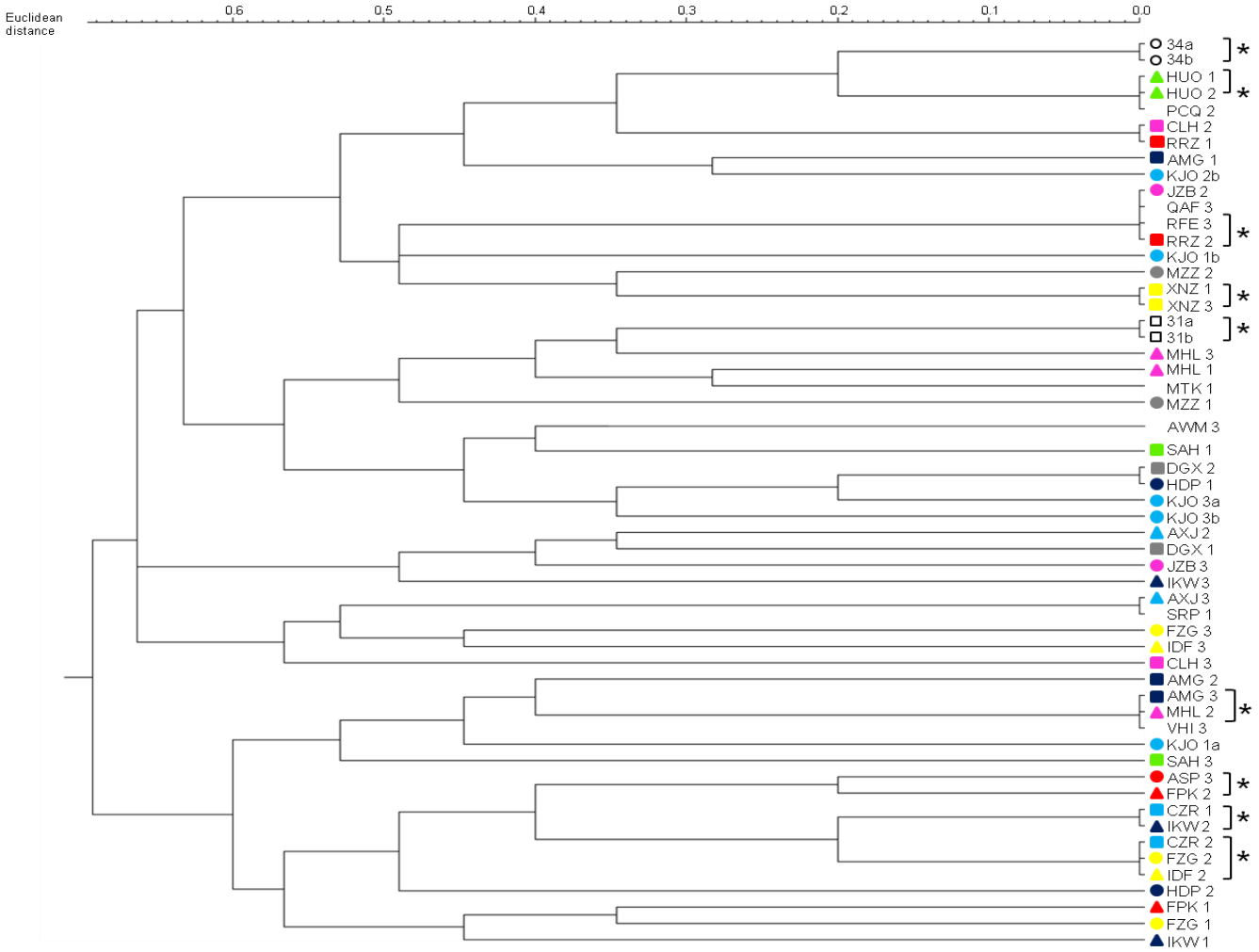

Figure 4-5. Cluster analysis of STR locus sizes. Peak traces were analysed using Peakscanner 1.0 software to determine the size (bp) of the allele(s) at each locus within all samples. Cluster analysis of STR allele sizes was carried out using Genemarker v2.2.2 software to generate a distance matrix based on the Euclidean distance between single samples. Only samples from which viral RNA was extracted are included. Patients for whom more than one isolate was sequenced are highlighted with coloured symbols to aid identification of patient clusters. Symbols correspond to those used in Figure 4-3. Clusters of interest that are discussed in the text are highlighted with an asterisk (*).

clustered with IDF2, an association that was present in Figure 4-3. XNZ1 and XNX3 clustered within Figure 4-5 replicating the pattern seen in Figure 4-3, as did samples HUO1 and HUO 2, as expected for related samples. Interestingly, HUO1 and HUO2 also clustered with PCQ2, although according to HVR1 sequence analysis this isolate was not related. Similarly AMG3 and MHL2 also clustered with VHI3, an association that was not previously observed. The pairing of RRZ2 and RFE3 based on viral sequence similarity (Figure 4-3) was partially replicated by DNA fingerprint analysis (Figure 4-5); however the cluster was extended to include JZB2 and QAF3 also. These slight discrepancies between viral sequence clustering and human genotype analysis are most likely caused by the small number of STR loci analysed. Commercial kits typically use 13 or more loci to provide the statistical power necessary for forensic applications. For the purposes of this study, three loci were sufficient to confirm that samples were not related. However, more loci would be needed to accurately match all samples which are epidemiologically linked.

The sample pairs JZB3 and SRP1 did not cluster based on human genotype despite showing strong viral sequence similarity (86% bootstrap support). Therefore these samples would appear not to be derived from the same patient. The close viral sequence similarity may be due to a common source of infection. STR locus amplification could not be completed on samples AXJ1 and SAH2 therefore comparison of this cluster was not possible. STR locus amplification could also not be completed on sample ASP1, preventing confirmation of the association with sample AWM3. Three novel clusters of samples were observed in the STR fingerprinting analysis (Figure 4-5). The pairings of DGX2 plus HDP1 and AXJ3 plus SRP1 were not observed in

Figure 4-3 and analysis of more STR loci would be needed to confirm any association between these samples. If an association were confirmed, the lack of clustering based on viral sequences may represent mixed acute infections. The clustering of CLH2 and RRZ1 in Figure 4-5 was partially replicated in Figure 4-3, where CLH2 clustered with KJO3a and then formed a larger cluster with RRZ1. Again, further analysis would be needed to confirm the relatedness of samples but these may be derived from a single patient.

STR locus size could not be determined for all samples. This may be due to degradation of the nucleic acid contained within the sample. This may be the case for samples HUO3 and RFE2 which were negative for both viral RNA and human DNA. Alternatively, the primer binding site may contain a mutation which prevents amplification of the target and leads to a phenomenon known as 'allele dropout' (Butler, 2006). This may explain why, in the case of sample KJO2a, the vWA locus could not be amplified. Finally, the presence of PCR inhibitors within the sample may have caused the reaction to fail producing a negative result as was observed for AXJ1 and SAH2.

4.4.4 Analysis of serum samples for antibody and viral load

All Egyptian serum samples were also tested for total antibody content and viral load. Antibody testing was carried out by a commercial third generation immunoassay. Samples with a test score >11.0 were reported as antibody positive (++) in Table 4-3. The remaining samples were non-reactive and therefore reported as antibody negative (-). One sample, HUO3 gave a borderline reading and was therefore reported as weakly positive (+). Viral load was tested by an in-house qPCR method based on amplification of the

Table 4-3. Antibody and viral load status of serum samples.

Sample name	Antibody status	Viral load (copies per ml)
AMG 1	++	5.95×10^6
AMG 2	++	2.41×10^7
AMG 3	++	9.00×10^7
ASP 1	++	6.44×10^6
ASP 2	-	Not detected
ASP 3	++	6.49×10^7
AWM 1	-	Not detected
AWM 2	-	Not detected
AWM 3	++	5.13×10^5
AXJ 1	++	2.10×10^7
AXJ 2	++	1.01×10^7
AXJ 3	++	2.05×10^8
CLH 1	-	Not detected
CLH 2	++	6.20×10^7
CLH 3	++	4.46×10^7
CZR 1	++	2.38×10^7
CZR 2	++	5.21×10^6
CZR 3	++	Not detected
DGX 1	++	2.71×10^7
DGX 2	++	5.03×10^4
DGX 3	++	Not detected
FPK 1	++	6.16×10^6
FPK 2	++	7.49×10^6
FPK 3	-	Not detected
FZG 1	++	1.69×10^7
FZG 2	++	1.30×10^6
FZG 3	++	7.71×10^6
HDP 1	++	2.75×10^6
HDP 2	++	4.13×10^7
HDP 3	++	5.40×10^6
HUO 1	++	1.39×10^7
HUO 2	++	6.34×10^6
HUO 3	+	Not detected
IDF 1	Empty tube	Empty tube
IDF 2	++	2.83×10^7
IDF 3	++	$6.46x\ 10^6$
IEB 1	-	Not detected
IEB 2	-	Not detected
IEB 3	-	Not detected
IKW 1	++	5.09×10^6
IKW 2	++	2.19×10^7
IKW 3	++	1.73×10^7
JZB 1	-	Not detected
JZB 2	++	6.01×10^7
JZB 3	++	6.46×10^6
KJO 1a	++	2.26×10^7
KJO 2a	++	Not detected
KJO 3a	++	7.90×10^7
KJO 1b	++	1.46×10^7
KJO 2b	++	1.22×10^7
KJO 3b	++	5.56×10^7
MHL 1	++	1.45×10^6
MHL 2	++	4.28×10^7
MHL3	++	1.35×10^7

Sample name	Antibody status	Viral load (copies per ml)
MTK 1	++	3.63×10^6
MTK 2	++	Not detected
MTK 3	++	Not detected
MZZ 1	++	3.93×10^6
MZZ 2	++	4.23×10^6
MZZ 3	++	Not detected
NDW 1	++	Not detected
NDW 2	++	Not detected
NDW 3	++	Not detected
OPN 1	++	Not detected
OPN 2	++	Not detected
OPN 3	++	Not detected
PCQ 1	++	Not detected
PCQ 2	++	1.36×10^7
PCQ 3	++	Not detected
QAF 1	++	Not detected
QAF 2	++	Not detected
QAF 3	++	2.43×10^7
RFE 1	++	Not detected
RFE 2	++	Not detected
RFE 3	++	1.90×10^7
RRZ 1	++	2.80×10^7
RRZ 2	++	4.94×10^7
RRZ 3	++	Not detected
SAH 1	++	2.44×10^8
SAH 2	++	9.38×10^7
SAH 3	-	Not detected
SMQ 1	-	Not detected
SMQ 2	++	8.18×10^7
SMQ 3	-	Not detected
SMY 1	-	Not detected
SMY 2	++	Not detected
SMY 3	++	Not detected
SRP 1	++	4.96×10^8
SRP 2	++	Not detected
SRP 3	++	4.15×10^7
TWS 1	++	Not detected
TWS 2	++	Not detected
TWS 3	++	Not detected
UHO 1	++	Not detected
UHO 2	++	Not detected
UHO 3	++	Not detected
VHI 1	++	Not detected
VHI 2	++	Not detected
VHI 3	++	2.34×10^8
XNZ 1	++	1.38×10^8
XNZ 2	-	Not detected
XNZ 3	++	7.80×10^7

5'NCR. Results, as viral genome copies per ml of serum, were calculated and are given in Table 4-3. A large number of samples contained too little viral RNA to be amplified and detected by qPCR (reported as not detected). The lower limit of detection for this method is $<1.25 \times 10^5$ copies per ml of serum, based on the amplification of known size standards. Failure to amplify viral nucleic acid in these samples could also be due to degradation of the sample or the presence of PCR inhibitors rather than as a result of low viral load.

Examination of the data in Table 4-3 raised a few anomalies. The large number of serum samples which were antibody positive is perhaps surprising bearing in mind that these samples were reported to be from the window period of infection. The window phase encompasses the period between exposure to the virus and the appearance of antibodies in the serum whereas the acute phase also includes the early post-seroconversion period. Antibody typically appears within 7-8 weeks of infection although it can take up to 3 months to appear (Pawlotsky, 1999). When compared to the data available regarding sampling time, shown in Table 4-4, it appears that several patients (AMG, ASP, CLH, CZR, FPK and HUO) developed antibodies within one month of exposure to the virus. One patient, CZR, had a period of only 13 days between exposure to the virus and the detection of antibody. Sampling time varies from 2 months to 8 months for those patients for whom data is available. Therefore these samples should fall within the acute phase of disease, however many of the samples contain high viral loads which may suggest a chronic stage of infection.

Four patterns of infection can be seen within the patients tested. (i) In patients CZR, DGX, MTK, MZZ and RRZ viral load became undetectable in

Table 4-4. Information regarding sampling time compared to antibody status and viral load for selected patient samples.

Patient	Exposure Date	Sample no.	Test Date	Antibody status	Viral load
AMG	24/05/2008	AMG 1	28/06/08	++	5.95×10^6
		AMG 2	Unknown	++	2.41×10^7
		AMG 3	Unknown	++	9.00×10^7
ASP	18/04/2008	ASP 1	18/05/08	++	6.44×10^6
		ASP 2	20/07/08	-	Not detected
		ASP 3	15/09/08	++	6.49×10^7
AWM		AWM 1	14/01/08	-	Not detected
		AWM 2	20/05/08	-	Not detected
		AWM 3	25/08/08	++	5.13×10^5
AXJ	12/2007	AXJ 1	07/02/08	++	2.10×10^7
		AXJ 2	09/06/08	++	1.01×10^7
		AXJ 3	12/10/08	++	2.05×10^8
CLH	06/06/2008	CLH 1	09/06/08	-	Not detected
		CLH 2	10/07/08	++	6.20×10^7
		CLH 3	12/08/08	++	4.46×10^7
CZR	12/07/2008	CZR 1	25/07/08	++	2.38×10^7
		CZR 2	31/08/08	++	5.21×10^6
	resolved	CZR 3	02/10/08	++	Not detected
DGX	18/07/2008	DGX 1	05/09/08	++	2.71×10^7
		DGX 2	01/11/08	++	5.03×10^4
		DGX 3	15/12/08	++	Not detected
FPK	04/09/2008	FPK 1	10/10/08	++	6.16×10^6
		FPK 2	14/11/08	++	7.49×10^6
		FPK 3	22/12/08	-	Not detected
FZG	after 03/2008	FZG 1	12/06/08	++	1.69×10^7
		FZG 2	20/07/08	++	1.30×10^6
		FZG 3	02/09/08	++	7.71×10^6
HDP	23/03/2008	HDP 1	10/05/08	++	2.75×10^6
		HDP 2	12/06/08	++	4.13×10^7
		HDP 3	23/07/08	++	5.40×10^6
HUO	14/08/2008	HUO 1	28/09/08	++	1.39×10^7
		HUO 2	20/10/08	++	6.34×10^6
		HUO 3	29/11/08	+	Not detected
IDF		IDF 1	02/08/08	Empty tube	Empty tube
		IDF 2	04/10/08	++	2.83×10^7
		IDF 3	03/12/09	++	6.46×10^6
IKW	14/08/2008	IKW 1	02/10/08	++	5.09×10^6
		IKW 2	05/11/08	++	2.19×10^7
		IKW 3	08/12/08	++	1.73×10^7

sample 2 or 3. As these patients did not receive treatment, this suggests spontaneous resolution of infection. Patients NDW, OPN, TWS, and UHO were all antibody positive yet had undetectable viral load throughout the sampling period. These patients may also have spontaneously cleared

infection. (ii) In patients PCQ, QAF, RFE, SRP and VHI one or more samples were negative for viral RNA followed by a sample with a detectable viral load. This pattern of infection could be caused by reinfection after treatment or after spontaneous resolution of an initial infection. Alternatively, so-called 'yo-yo' viral load, which was recently described (Smith *et al.*, 2010), could account for this pattern. (iii) Several patients were antibody positive and viral RNA positive in all three samples, clearly showing an ongoing infection. However, without accurate sampling times for all samples it is difficult to determine if these patients were suffering from an acute infection or if they were chronically infected with HCV. (iv) Three out of the thirty-four patients showed a pattern typical of early acute phase infection. These are samples AWM, CLH, and JZB. They all became antibody positive and viral RNA positive between samples 1 and 2 or between samples 2 and 3 suggesting that the crucial window period of infection had been captured.

Some more unusual patterns of infection were also identified upon closer examination of individual patients. In patients ASP, FPK, SAH, SMQ, XNZ and possibly also HUO there was a loss of antibody reactivity from one sample to another. In the case of patients ASP and XNZ, antibody disappeared and then reappeared. This is highly unusual as antibody has been shown to persist for several years both in individuals who resolve the infection and in those who develop a chronic infection (Dowd *et al.*, 2009; Pestka *et al.*, 2007). According to the sampling data given, patient ASP was tested four times within a period of five months. It is highly unlikely that antibody was truly lost and regained within that time-frame. It is possible that the samples were mixed up and that samples 2 and 3 should be reversed, however the loss of antibody so soon after

infection is still difficult to explain. Finally it appears that patient IEB was not infected as neither antibody nor viral RNA was detected.

4.4.5 Antibody avidity testing

Avidity testing of the antibodies within the Egyptian serum samples was carried out by Joy Kean to determine whether these samples were from acutely infected patients. IgG class antibodies produced during the acute phase of infection are of low avidity but this increases over time (Gaudy-Graffin *et al.*, 2010; Kanno & Kazuyama, 2002). Therefore measuring the avidity of IgG gives a good indication of the age of the antibody response and therefore the duration of infection. Almost all of the samples tested were found to contain high avidity antibodies with AI values >45 (Table 4-5). This is characteristic of antibodies produced during the chronic stage of infection. Only two samples were found to contain low avidity antibodies below the threshold AI value of 45. These were AMG3 (AI value, 34) and VHI3 (AI value, 35), however the previous samples from these two patients contained high avidity antibodies making these samples very unreliable as a source of acute phase antibodies.

4.5 Discussion

Studies of the acute phase of HCV infection are difficult due to the asymptomatic nature of the acute phase (Alter, 2007). Most patients do not present any outward signs of disease until the virus has established a chronic infection. Therefore the identification of a cohort of patients suffering from

Table 4-5. Avidity of IgG immunoglobulins within patient serum samples. AI values >45 indicate high avidity antibodies.

Sample name	Antibody avidity index (AI)	Sample name	Antibody avidity index (AI)
AMG 1	111	**MHL 1**	115
AMG 2	104	**MHL 2**	47
AMG 3	34	**MHL3**	92
ASP 1	99	**MTK 1**	113
ASP 2	Negative	**MTK 2**	103
ASP 3	102	**MTK 3**	114
AWM 1	Negative	**MZZ 1**	85
AWM 2	Negative	**MZZ 2**	117
AWM 3	103	**MZZ 3**	107
AXJ 1	103	**NDW 1**	104
AXJ 2	102	**NDW 2**	96
AXJ 3	98	**NDW 3**	101
CLH 1	Negative	**OPN 1**	77
CLH 2	99	**OPN 2**	104
CLH 3	94	**OPN 3**	114
CZR 1	107	**PCQ 1**	117
CZR 2	85	**PCQ 2**	108
CZR 3	101	**PCQ 3**	109
DGX 1	96	**QAF 1**	106
DGX 2	98	**QAF 2**	103
DGX 3	97	**QAF 3**	100
FPK 1	84	**RFE 1**	112
FPK 2	87	**RFE 2**	107
FPK 3	Negative	**RFE 3**	104
FZG 1	66	**RRZ 1**	97
FZG 2	80	**RRZ 2**	95
FZG 3	106	**RRZ 3**	98
HDP 1	102	**SAH 1**	101
HDP 2	96	**SAH 2**	98
HDP 3	96	**SAH 3**	Negative
HUO 1	90	**SMQ 1**	Negative
HUO 2	104	**SMQ 2**	79
HUO 3	Negative	**SMQ 3**	Negative
IDF 1	Empty tube	**SMY 1**	Negative
IDF 2	Not tested	**SMY 2**	98
IDF 3	Not tested	**SMY 3**	106
IEB 1	Negative	**SRP 1**	100
IEB 2	Negative	**SRP 2**	100
IEB 3	Negative	**SRP 3**	92
IKW 1	98	**TWS 1**	87
IKW 2	85	**TWS 2**	51
IKW 3	94	**TWS 3**	91
JZB 1	Negative	**UHO 1**	87
JZB 2	105	**UHO 2**	92
JZB 3	106	**UHO 3**	84
KJO 1a	110	**VHI 1**	89
KJO 2a	108	**VHI 2**	83
KJO 3a	111	**VHI 3**	35
KJO 1b	107	**XNZ 1**	99
KJO 2b	116	**XNZ 2**	Negative
KJO 3b	105	**XNZ 3**	93

acute HCV infection would provide an invaluable tool to study the dynamics of the disease. The panel of acute phase serum samples used in this study were obtained from patients with a known exposure to HCV, either through contact with an infected partner or exposure within a healthcare setting. Sampling patients of this type is generally easier than in cohorts of intravenous drug users who are notoriously unreliable at attending follow-up sessions. As such, this cohort contained 34 separate patients and triplicate time-point samples were provided for each patient. It was hoped that a study of the antibody response and virus evolution in this cohort would provide some useful insights into acute HCV infection and the role played by the antibody response in spontaneous resolution of infection. Unfortunately, preliminary studies of these samples raised serious questions about their provenance.

Genotyping analysis of viral isolates was initially carried out on the basis of 5'NCR sequencing. However this was found to have insufficient discriminatory power to accurately assign viral isolates to a specific genotype. This problem was not wholly unexpected as several limitations have been identified with genotyping of the 5'NCR. This method is unable to easily resolve genotype 1 and 6 isolates and to distinguish between some subtypes within genotypes 1, 2, 3, 4 and 6 (Murphy *et al.*, 2007), shortcomings which were observed in this study (Figure 4-1). This led us to carry out genotype analysis on the basis of HVR1 sequencing and, as expected, most isolates clustered with genotype 4 reference sequences (Figure 4-2), the dominant genotype within the Egyptian population, while the remainder clustered with genotype 1.

Phylogenetic trees generated from viral sequences are very useful tools to infer the relatedness of a group of viruses and to reconstruct the epidemiology of an outbreak. This approach has been used to study the origin and epidemiology of HIV, influenza, SARS coronavirus and Ebola virus (Pybus & Rambaut, 2009) and provide important insights into patterns of virus transmission. The method has also been used to assess transmission events between monogamous partners of HCV-infected individuals (de Carvalho-Mello *et al.*, 2010), characterise the risk factors and HCV transmission characteristics within different social networks (Matthews *et al.*, 2011), and to identify the single source of HCV outbreaks within haemodialysis units (Lanini *et al.*, 2010; Thomson *et al.*, 2011). In all of these studies epidemiologically linked viruses, i.e. those with a common source, show significant sequence similarity with high bootstrap values. These studies differ in the region of the virus used for sequencing and there is some disagreement about whether the envelope glycoproteins E1 and E2, which evolve under pressure from the immune system, provide a more accurate phylogeny than more conserved regions of the genome such as NS3 and NS5B (de Carvalho-Mello *et al.*, 2010). As this study was concerned with the evolution of the antibody response during acute infection, phylogenetic analysis of the envelope glycoproteins was deemed most appropriate.

Closer analysis of the clustering patterns in Figure 4-1 (and subsequently Figure 4-2) had shown that, in several cases, isolates from the same patient clustered with different viral subtypes and these unusual patterns were subsequently confirmed by phylogenetic analysis of HVR1 (Figure 4-3). Only two patients out of thirty-four were found to have HVR1 sequences from

sequential samples that clustered together. Studies suggest that HCV infection is typically established by a single founder virus (Bull *et al.*, 2011; Wang *et al.*, 2010); as a result of which, viral sequences isolated from a single patient during the acute phase of infection show significant similarity and cluster closely within a tree. The highly divergent nature of sequences from the same patient therefore suggested multiple different sources of the virus, i.e. these viruses had been isolated from different patients. Other samples were found to cluster very closely on the tree and could be considered to have derived from the same patient. An alternative explanation for the divergent nature of the sequences is the existence of a mixed acute infection, described recently by Smith *et al*. (2010) in a cohort of acutely infected patients. By cloning and sequencing of HVR1 from samples taken during the first three months post-infection they reported coinfection with more than one subtype at individual time points in 70% of study subjects. Moreover, these coinfections were highly dynamic with multiple switches in the dominant viral variant throughout acute infection. They also describe 'yo-yo' viral load values in several of their study cohort, although this did not correlate with viral subtype switching (Smith *et al.*, 2010). However, this study was not supported by any analysis of the human genotype. Therefore it is not possible to rule out that these mixed, highly dynamic acute infections were due to sample contamination or mislabelling of samples.

In order to distinguish between mixed acute infection and the mislabelling of samples within the Egyptian study cohort, STR fingerprinting was carried out. This method is widely used in forensic studies and there are several commercially available kits. The principles behind this technology were

adapted for use in-house using fluorescently-labelled PCR primers to amplify the genes of interest which were then analysed. Due to the presence of human nucleic acids within the RNA samples extracted from serum, no additional DNA isolation or purification was required. Commercial STR fingerprinting kits analyse up to 13 STR loci and contain an internal size standard and an allelic ladder for each locus. The internal size standard contains PCR products of a known size. The allelic ladder contains a copy of every known allele of a given locus. This information is then used to determine the size of sample peaks and correlates this to a specific allele (Butler, 2006). As this analysis did not use a commercial kit no allelic ladder was available for analysis alongside the samples. A size standard containing 16 DNA fragments 35 to 500 nucleotides in length was included in all samples. Therefore the results of STR fingerprinting were given as peak size in base pairs only. As no allelic ladder was available it is not possible to convert this information into allele number. This does not however prevent full analysis of the samples and the identification of those patient samples which do not match. STR fingerprinting analysis (Figure 4-5) did confirm the results of HVR1 sequence analysis (Figure 4-3). It was very clear that most of the samples were isolated from different patients and there was very little clustering according to sample name. The exceptions to this observation were patients HUO and XNZ which clustered by HVR1 sequence and contained identical STR loci (Figure 4-4 and 4-5). The inclusion of positive controls from two patients, 31 and 34, that were found to cluster according to STR locus size provided further confirmation that the Egyptian samples were either mislabelled or contaminated (Figure 4-5). These analyses demonstrate that viral isolate sequencing provides a good

indicator of the origin of a sample and can be used to identify those samples which have a common source. More sample clusters were identified in Figure 4-3 which were confirmed to have a common origin by STR analysis (Figure 4-5). It is however slightly misleading to say that this relationship has been confirmed. As this study used only three STR loci it lacks the discriminating power to provide a positive match between two or more samples. Three loci are sufficient to identify a negative relationship, i.e. two or more samples which do not match. It would be necessary to analyse more STR loci in order to accurately identify those samples which are a positive match, derived from the same individual. Even if more STR loci were analysed it would still not be possible to determine the chronological order in which samples were taken making any study of virus evolution impossible.

DNA fingerprinting of these samples confirmed our suspicion that they were not epidemiologically linked and ruled out the possibility of mixed acute infection. This highlights a major flaw in the recent paper by Smith *et al*. (2010) which failed to analyse human genotype. Many of their observations such as subtype-switching within study subjects and 'yo-yo' viral load dynamics were also observed in our study samples (Table 4-3 and Figure 4-2). Without human genotype analysis it is impossible to confirm the existence of these phenomena within acutely infected patients and the speculation that these are contaminated or mislabelled samples will continue (Irving & Brown, 2010). Mixed acute phase infections do occur, as evidenced by the clustering patterns of samples from patients 31 and 34 (Figure 4-5); however we would urge careful scrutiny of any unusual data and the validation of all samples used, as not all apparently mixed acute infections truly are.

To complete the analysis of this cohort, samples were tested for HCV-specific antibodies and viral load (Table 4-3). Only three patients exhibited a pattern typical of the window phase of infection, during which time antibodies against the virus begin to develop. Doubts were also raised when it was noted that several patients appeared to lose antibody and then regain antibody between sampling. While loss of viral RNA may be caused by spontaneous clearance of infection, with or without subsequent reinfection, antibody is not so readily lost, particularly not within such a short time-frame. Loss of antibody, or its reduction to undetectable levels, has been reported in patients with a resolved acute infection; however this did not occur until, on average, 4 years after infection was resolved (Alter *et al.*, 1989). Antibodies have also been reported to persist for several years in both chronic and resolved infections (Dowd *et al.*, 2009; Pestka *et al.*, 2007). It therefore seems extremely unlikely that six patients should have developed and subsequently lost antibodies within a period of a very few months. Information regarding sampling time was provided in an attempt to clear up the confusion (Table 4-4). However, this only served to raise more problems as it seemed that several patients developed an antibody response and a high viral load within 3 or 4 weeks of exposure to the virus. Typically antibody is detectable within 7-8 weeks but can take up to 3 or 4 months or longer to develop (Alter *et al.*, 1989; Pawlotsky, 1999). Therefore the period of 13 days between exposure to the virus and the appearance of antibodies reported for patient CZR is highly unusual. The antibody avidity data (Table 4-5) indicated that the majority of samples were from the chronic phase of the disease, providing further proof that these samples were not as they were described to be. Antibody avidity

testing has been used to distinguish acute from long-term infection in a number of viral infections including rubella virus (Hedman & Rousseau, 1989), Varicella-Zoster virus (Junker & Tilley, 1994), EBV (Gray, 1995) and CMV (Grangeot-Keros *et al.*, 1997). It has also been used to discriminate between primary HCV infection and the passive transfer of anti-HCV Igs (Ward *et al.*, 1994) and to identify acutely infected patients (Gaudy-Graffin *et al.*, 2010; Kanno & Kazuyama, 2002). The use of such assays to accurately identify acute infection is particularly valuable for HCV due to the largely asymptomatic nature of the disease.

The data presented in this chapter clearly demonstrates that these samples were not isolated during the acute phase of the disease. It is also obvious that the triplicate samples were in fact derived from a number of different individuals and that very few clusters of samples could be identified. Whether this discrepancy is as a result of poor sample storage, mislabelling or a wilful effort to deceive cannot be known. It has however prevented any study of the antibody response during the acute phase of the disease and its impact on the outcome of disease. This study highlights the importance of establishing the provenance and validity of all samples used in studies of viral evolution in order to confirm the accuracy of any conclusions made.

5 General discussion and conclusions

Our current understanding of HCV pathogenesis suggests that a strong immune response can protect against chronic infection. Humoral immunity has been shown to protect against primary infection (Feray *et al.*, 1998) and reinfection (Osburn *et al.*, 2010), and a broad nAb response during acute infection correlates with spontaneous resolution (Dowd *et al.*, 2009; Pestka *et al.*, 2007). The potential for nAbs to prevent infection has made them a promising area of research, yet the development of antibody-based vaccines requires a better understanding of the protective antibody response. The majority of nAbs identified to date target the envelope glycoproteins, in particular E2 and, of these, the most broadly neutralising target the CD81 binding region which forms a cluster within AR3 (reviewed in Edwards *et al.*, 2012). E2-CD81 interactions are necessary for HCV entry into hepatocytes and are clearly an important target for the nAb response, particularly as mutational escape would likely incur fitness costs to the virus. However the scarcity of nAbs binding to other viral epitopes has raised questions about the diversity of the antibody response to HCV. Unfortunately, the assays used to study the HCV-induced antibody response may have introduced bias towards specific regions of E2. As discussed in section 1.6.2, the early use of Molt-4 cells (Flint *et al.*, 1999a; Rosa *et al.*, 1996) would have limited the identification of antibodies to those blocking CD81 binding sites. The choice of E2 construct, whether it is sE2, E1E2 or expressed on pseudoparticles, also has important implications in terms of conformation and antigenicity of the glycoprotein (Owsianka *et al.*, 2001). Most human anti-HCV antibodies target conformational epitopes; however, as discussed in section 2.1.1, methods to

map antibody binding residues in conformation-sensitive epitopes show several limitations. Therefore a system is needed that can rapidly and easily map antibody binding epitopes within any region of the glycoprotein in an unbiased manner. This system would require the expression of the entire glycoprotein ectodomain in a conformation that is representative of its native form.

To this end, a cell surface display library based on expression of the E2 ectodomain fused to the influenza A virus HA_{TM} domain in *D. melanogaster* S2 cells was developed (Figure 2-14 to 2-17). Although it was possible to sort cells that no longer bound the test MAb but maintained E2 conformation (Figure 2-21 and 2-22), the pool of sorted cells was quite small. It was possible however, to enrich for a specific subset of cells by growing and expanding the sorted cells for several weeks prior to re-sorting (Figure 2-23). Analysis of the sequences isolated from single cells sorted for loss-of-binding to MAb AR3A showed that, as expected, some residues within AR3 were mutated. However a number of residues that have not previously been implicated in AR3A binding were also mutated (Table 2-4). Of particular interest, two residues L399I and C644R, were observed in the majority of sequences (discussed in section 2.5). This may reflect over representation of these substitutions in the library; however further study will be needed to investigate what, if any, role these residues have in AR3A binding.

Other limitations of the display library were identified which may be improved upon. These are discussed in detail in section 2.5 but it is apparent that a larger library will be required to ensure that all possible substitutions at all residues are represented. The efficacy of the library could also be improved by increasing the number of cells expressing protein which might be achieved

by pre-enriching the library for cells expressing E2. With these general improvements, the *D. melanogaster* S2 cell library could become a very powerful tool to study antibody epitopes within E2. It is hoped that this system will enable the rapid identification of residues involved in MAb recognition, regardless of the region targeted by the MAb and without any prior knowledge of its function or specificity. Expression of the entire E2 ectodomain should overcome any bias towards a particular region of E2 or a specific receptor binding site. The use of correctly folded glycoprotein and the ability to screen all library cells based on protein conformation will facilitate the identification of conformation-sensitive MAbs, a process which currently is laborious and time-consuming. Any residues identified by display library screening could be investigated further using plate-based binding assays or the HCVpp and HCVcc systems to elucidate their importance in MAb binding.

Although an E2 cell surface display library expressing the entire ectodomain is not biased towards a specific region or epitope within E2, it is limited to antibodies targeting this glycoprotein. E2 appears to be the major target of the antibody response and it has been suggested that E1 is less immunogenic (Fournillier *et al.*, 2001). However, to avoid any bias it might be possible to generate a display library expressing E1. This was not attempted due to the time constraints of the project. Expression of E1 has been hampered in the past by poor expression and the suggestion that E2 co-expression is required for E1 to fold correctly (Michalak *et al.*, 1997). However, an E1-specific human antibody isolated with a soluble form of E1 was also shown to react with E1E2 (Keck *et al.*, 2004b) and broadly nAbs that inhibit infectivity of HCVpp and HCVcc also bind soluble forms of E1 (Meunier *et al.*, 2008). Furthermore,

studies with recombinant E1 vaccine candidates demonstrate the induction of an E1-specific nAb response (Leroux-Roels *et al.*, 2004; Nevens *et al.*, 2003; Verstrepen *et al.*, 2011). Therefore it appears that an antigenic form of E1 can be expressed without E2, although whether the whole protein is correctly folded is not clear. If an E1 display library was successfully generated it could be used to screen patient sera, providing important insight into the prevalence of E1-specific MAbs, as well as for fine mapping of residues bound by E1-specific antibodies.

Attempts were made to generate a display library bearing E2 from a genotype 2b strain. Unfortunately this proved to be unsuccessful and it is not clear whether this was due to poor expression or incorrectly folded protein. The genotype 2b E2 construct used has been expressed as recombinant E1E2 and on the surface of HCVpp and found to fold correctly (Tarr *et al.*, 2011). It is hoped therefore that this construct can be successfully displayed in the insect cell library. The adaptability of the system to express envelope glycoproteins of different genotypes would greatly enhance its utility. Identifying nAbs with broad reactivity is an important consideration as these are most likely to prove therapeutically useful. Testing the breadth of antibody binding with libraries bearing E2 of different genotypes would facilitate the identification of residues or epitopes that are conserved across genotypes.

Ideally, antibody mapping would be done with surface expressed E1E2 to recapitulate the native form of the envelope glycoprotein complex. However, this has proven extremely difficult and is hampered by a lack of understanding of the functional form of the glycoprotein complex (Vieyres *et al.*, 2010). However, if such a library could be generated it may prove useful as a 'pull-

down' method to isolate specific antibodies from patient sera or polyclonal Ig facilitating the isolation of E1E2-specific MAbs in an unbiased manner. Alternatively, experiments with cells expressing E1E2 (or E2) that has been engineered to lack specific immunodominant epitopes may enable the isolation of rarer antibodies. This approach was recently used to screen a library of scFv fragments for those which bound outside of known neutralising epitopes. This led to the identification of a novel antigenic domain within E2 (Keck *et al.*, 2012). By reversing the approach and displaying the glycoprotein on the library platform, it may be possible to screen patient sera for nAbs targeting a specific epitope. Antibody isolation in conjunction with epitope mapping would help to determine the location of additional neutralising determinants or antigenic domains. Such an approach provides an alternative to cell free assays which utilise sE2 to isolate specific antibodies from sera (Sophie Brice, personal communication) and may help to determine the prevalence of specific antibodies within patient sera.

A further problem commonly encountered during studies of the antibody response to HCV is the difficulty of obtaining nAbs. As discussed in section 3.1.4.1 genetic differences mean that the antibody response in mice does not always accurately recapitulate the antibody response seen in humans. Even in mice engineered to express human Ig genes, differences in antigen processing may direct the antibody response towards different epitopes. Isolating nAbs directly from humans may yield more useful antibodies; however antibody isolation and production methods, such as hybridoma generation, are technically challenging and not always successful. Newer methods, such as the immortalisation of B cells, have also met with mixed results (Pinna *et al.*,

2009; Traggiai *et al.*, 2004) and the protocol was not successful in our lab (R. Urbanowicz, personal communication). However, combining B cell activation methods that stimulate the generation of ASCs with Ig gene isolation and sequencing techniques provides a high throughput method to screen the B cell repertoire of a given patient. This method has been widely used in the fields of HIV (Pietzsch *et al.*, 2010; Scheid *et al.*, 2011; Wu *et al.*, 2011) and influenza (Corti *et al.*, 2011; Grandea *et al.*, 2010) research and, using a combination of Ig-specific primers, was successfully adapted to the isolation of Ig genes from HCV-infected patients (Table 3-7 to 3-10). Although single B cell cultures were not successfully analysed during the time available, such cultures have been successfully stimulated to secrete antibody by other groups and therefore it is hoped that, with further troubleshooting (as discussed in section 3.5), single cell cultures of ASCs will be generated. Although isolating paired heavy and light chain sequences from cultures of more than one B cell is technically possible, it would be very time-consuming and the additional PCR steps required would increase the likelihood of introducing experimental artefacts into the sequence. Therefore, in order to develop a robust, high-throughput method for examining B cell populations it would be extremely beneficial to have a reliable protocol for generating single B cell cultures.

The IgG gene isolation and sequencing method is intended ultimately as a means of MAb production. Following the isolation of paired heavy and light chain genes, these would be cloned into a suitable expression vector and expressed in an appropriate host cell. Such expression vector cloning was attempted, however due to incompatibilities between the vector and primer sets used this met with only limited success (see section 3.4.6). The current

protocol described in Chapter 3 is time consuming and increases the likelihood of introducing artefacts into the IgG gene sequences. Removing the additional amplification steps will also be necessary to streamline the IgG isolation and cloning process, making it suitable for routine use in the lab.

With the suggested improvements the isolation of MAbs from patient sera could be efficiently and quickly achieved. The B cell isolation and culturing technique requires only small volumes of blood and has the potential to generate thousands of single B cell cultures from an individual patient. By pre-sorting B cells or by screening the antibody-containing supernatant, it would be possible to select for antibodies of a certain specificity. For example, samples could be targeted for the presence of fusion-specific nAbs or antibodies with cross-genotype reactivity. IgG genes would only need to be isolated from the cultures of interest, thereby reducing the workload and avoiding the unnecessary cloning of non-reactive or non-neutralising antibodies. Paired heavy and light chain genes could be cloned from the appropriate cultures into the expression vector pDCORIG. This is a mammalian expression vector that has been used by other groups to produce large amounts of antibody (Metheringham *et al.*, 2009). Ideally, after cloning variable Ig genes into the vector, it would be transfected into CHO cells, and the antibody secreted into the supernatant. The antibody could then be characterised to determine the strength and breadth of neutralisation as well as the epitope targeted by the antibody. To this end, the insect cell display library could be used for epitope characterisation, whilst HCVpp and HCVcc assays could be employed to determine neutralisation. Only small amounts of antibody would be needed to carry out these preliminary tests. However, by generating stably transfected

cell lines, it would be possible to continually produce antibody allowing for more comprehensive characterisation. As previously suggested, epitope mapping with the insect cell library should provide a non-biased way of determining antibody specificity, particularly if E1 as well as E2 can be displayed in this format. Generating and maintaining stable mammalian cell lines is a fairly straightforward process and, unlike hybridoma generation, there is no risk of cell lines reverting to wild-type, provided that the selective agent is present. Each cell line would be transfected with a single plasmid and would therefore express MAb. Secreted antibody would need to be purified from the cell culture supernatant but this could be achieved using standard protein purification techniques such as affinity chromatography with protein G. Production of larger amounts of antibody might be possible by scaling-up the CHO cell culture although this can be difficult with adherent cell lines. However, CHO cell lines which grow in suspension culture and are more easily transfected and selected have been developed (Freestyle CHO-S cells from Invitrogen). Their use would increase the amount of IgG produced and speed-up the process of MAb generation. Furthermore these cells are adapted to growth in serum-free media, simplifying the purification of antibody from the supernatant (Anon, 2007). Whole antibodies have been successfully expressed in the *D. melanogaster* S2 cell line (Johansson *et al.*, 2007a) which is amenable to large-scale culture in spinner flasks and secretes large amounts of protein (M. Backx, personal communication). However, slight differences in glycosylation of the resulting antibodies have been reported which may require humanisation for optimal antibody activity *in vivo* (Johansson *et al.*, 2007a). No such problems have been reported for antibody produced in CHO cells.

The choice of a vector containing human Ig constant domains is another benefit of this system. In therapeutic applications, human antibodies tend to have a longer half-life, are less-likely to induce an allergic response than murine MAbs or even chimeric MAbs (Hwang & Foote, 2005), and show improved effector functions (Li *et al.*, 2006). Therefore a combination of the IgG gene isolation method and the epitope mapping technique may provide a new system for the production and characterisation of human MAbs produced during natural HCV infection.

The isolation of IgG genes may also provide further insight into the pathogenesis of HCV by allowing close examination of the gene sequences. A recent study attempted to find correlations between HCV disease outcome and the repertoire of germline VH gene segments used (Racanelli *et al.*, 2011). Some of their observations were similar to those made in this study (see section 3.53.5) and it will be interesting to see whether the VH and VL genes commonly used in spontaneous resolvers do produce nAb that helps to clear the virus. Studies carried out on HIV and influenza also suggest that antibodies targeting a conserved epitope are derived from very few related germline genes (Grandea *et al.*, 2010; Scheid *et al.*, 2009; Scheid *et al.*, 2011). If the same is true for HCV, affinity maturation of such germline genes could lead to the generation of MAbs with enhanced affinity or broader reactivity. Understanding how clonally related antibodies undergo affinity maturation may lead to targeted improvements in therapeutic MAbs. Lymphoproliferative disorders, such as mixed cryoglobulinaemia and B-cell non-Hodgkins lymphoma, are frequent manifestations of HCV infection (Poynard *et al.*, 2003) and it has been suggested that HCV induces hypermutation of IgG genes

as a means to evade the immune response (Machida *et al.*, 2008). Studies of the IgG gene repertoire in HCV infected individuals may therefore provide further insight into the interplay between the virus and the immune system. It will be interesting to see what, if any, impact such immune system impairment has on disease outcome. Resistance to such virus-induced mutation might be a hallmark of spontaneous resolvers and identifying the extent to which the virus alters the antibody response throughout infection could have important implications for therapeutic interventions.

A severe impediment to studies of HCV disease dynamics is the difficulty of identifying patients during the acute phase. Therefore, obtaining a cohort of patients suffering from acute infection for which sequential serum samples had been collected provided a great opportunity to study the evolution of nAb responses during acute infection. Unfortunately it quickly became apparent that the samples were not as they appeared and that identifying immune correlates of disease outcome would be impossible. A study of the cohort was carried out nonetheless and provided an object lesson in the importance of checking the origin and reliability of samples. Viral isolates from supposedly related samples did not cluster together on a phylogenetic tree (Figure 4-3), contrary to what was expected for epidemiologically related samples. The development of an in-house DNA fingerprinting technique was critical in confirming that the majority of grouped samples were not related and proved invaluable in distinguishing sample 'mix-up' from genuine mixed-acute infections (Figure 4-5). This approach could be used to assess the reliability of all samples in any future studies requiring blood, serum or PBMCs. This will

improve the reliability of the generated data and the conclusions that are subsequently drawn.

Neutralising antibodies have the potential to clear infection yet their precise role in HCV disease pathogenesis and outcome is incompletely understood. Although the majority of human nAbs identified target the E2-CD81 interaction, several other potential interactions exist for which few, if any, nAbs have been isolated. It is likely that antibodies targeting these interactions do exist and their discovery may provide other means of inhibiting virus entry. Indeed, the recent description of a novel AR within E2 which elicits a broadly neutralising response that does not interfere with CD81 binding (Giang *et al.*, 2012) supports the notion that additional antibody targets remain to be identified. HCV has proven itself very adept at evading immune clearance and lessons learned from HIV research suggest that multiple antibodies targeting several distinct but overlapping epitopes will provide the most effective means of neutralising infectivity. Therefore identifying additional neutralising epitopes will be important for the development of effective therapeutics. Furthermore, the interaction between the virus and the immune response is dynamic, and by studying the evolution of the antibody response during acute and chronic infection, we may identify important determinants of disease outcome. Such studies will be facilitated by new methods of B cell isolation and antibody characterisation, including those developed during this study. Understanding the complexity of the relationship between the virus and the immune response will require varied approaches. The techniques described in this project will hopefully provide the means to study diverse aspects of the

antibody response, techniques which may be used in isolation or combined in a novel work model.

References

Abe, Y., Takashita, E., Sugawara, K., Matsuzaki, Y., Muraki, Y. & Hongo, S. (2004). Effect of the addition of oligosaccharides on the biological activities and antigenicity of influenza A/H3N2 virus hemagglutinin. *J Virol* **78**, 9605-9611.

Agnello, V., Abel, G., Elfahal, M., Knight, G. B. & Zhang, Q.-X. (1999). Hepatitis C Virus and other Flaviviridae viruses enter cells via low density lipoprotein receptor. *P Natl Acad Sci USA* **96**, 12766-12771.

Ahmed, A., Linacre, A. M. T., Mohammed, A. A. A., Vanezis, P. & Goodwin, W. (2001). STR population data for 10 STR loci including the GenePrint PowerPlex 2.1 kit from El-Minia (Central Egypt). *Forensic Sci Int* **117**, 233-234.

Aiba, Y., Kometani, K., Hamadate, M., Moriyama, S., Sakaue-Sawano, A., Tomura, M., Luche, H., Fehling, H. J., Casellas, R. & other authors (2010). Preferential localization of IgG memory B cells adjacent to contracted germinal centers. *P Natl Acad Sci USA* **107**, 12192-12197.

Albecka, A., Belouzard, S., Op de Beeck, A., Descamps, V., Goueslain, L., Bertrand-Michel, J., Terce, F., Duverlie, G., Rouille, Y. & other authors (2012). Role of low-density lipoprotein receptor in the hepatitis C virus life cycle. *Hepatology* **55**, 998-1107.

Albecka, A., Montserret, R., Krey, T., Tarr, A. W., Diesis, E., Ball, J. K., Descamps, V., Duverlie, G., Rey, F. & other authors (2011). Identification of new functional regions in hepatitis C virus envelope glycoprotein E2. *J Virol* **85**, 1777-1792.

Allander, T., Drakenberg, K., Bevene, A., Rosa, D., Abrignani, S., Houghton, M., Widell, A., Grillner, L. & Persson, M. A. (2000). Recombinant human monoclonal antibodies against different conformational epitopes of the E2 envelope glycoprotein of hepatitis C virus that inhibits its interaction with CD81. *J Gen Virol* **81**, 2451-2459.

Alter, H. J., Purcell, R. H., Shih, J. W., Melpolder, J. C., Houghton, M., Choo, Q.-L. & Kuo, G. (1989). Detection of antibodies to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N Engl J Med* **321**, 1494-1500.

Alter, M. J. (2007). Epidemiology of hepatitis C virus infection. *World J Gastroentero* **13**, 2436-2441.

Amet, T., Ghabril, M., Chalasani, N., Byrd, D., Hu, N., Grantham, A., Liu, Z., Qin, X., He, J. J. & other authors (2012). CD59 incorporation protects hepatitis C virus against complement-mediated destruction. *Hepatology* **55**, 354-363.

Andre, P., Komurian-Pradel, F., Deforges, S., Perret, M., Berland, J. L., Sodoyer, M., Pol, S., Brechot, C., Paranhos-Baccala, G. & other authors (2002). Characterization of low- and very-low-density hepatitis C virus RNA-containing particles. *J Virol* **76**, 6919-6928.

Angelichio, M. L., Beck, J. A., Johansen, H. & Ivey-Hoyle, M. (1991). Comparison of several promoters and polyadenylation signals for use in heterologous gene expression in cultured *Drosophila* cells. *Nucleic Acids Res* **19**, 5037-5043.

Anon (2007). FreeStyle CHO-S cells. *Version B 25-0921*. http://tools.invitrogen.com/content/sfs/manuals/FreeStyle_CHO-S_Cells_man.pdf

Anon. (1999). Global surveillance and control of hepatitis C. *J Viral Hepatitis* **6**, 35-47.

Anon. (2003). *Drosophila* expression system. *Version H 25-0191*. http://tools.invitrogen.com/content/sfs/manuals/des_man.pdf

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

Armstrong, R. T., Kushnir, A. S. & White, J. M. (2000). The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. *J Cell Biol* **151**, 425-437.

Ascione, A., Tartaglione, M. T. & Di Costanzo, G. G. (2007). Natural history of chronic hepatitis C virus infection. *Digest Liver Dis* **39** S4-S7.

Bankwitz, D., Steinmann, E., Bitzegeio, J., Ciesek, S., Friesland, M., Herrmann, E., Zeisel, M. B., Baumert, T. F., Keck, Z.-Y. & other authors (2010). Hepatitis C virus hypervariable region 1 modulates receptor interactions, conceals the CD81 binding site and protects conserved neutralizing epitopes. *J Virol* **84**, 5751-5763.

Barbie, V. & Lefranc, M. P. (1998). The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp Clin Immunogenet* **15**, 171-183.

Barth, H., Rybczynska, J., Patient, R., Choi, Y., Sapp, R. K., Baumert, T. F., Krawczynski, K. & Liang, J. T. (2011). Both innate and adaptive immunity mediate protective immunity against hepatitis C virus infection in chimpanzees. *Hepatology* **54**, 1135-1148.

Barth, H., Schafer, C., Adah, M. I., Zhang, F., Linhardt, R. J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T. H. & other authors (2003). Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. *J Biol Chem* **278**, 41003-41012.

Barth, H., Schnober, E. K., Zhang, F., Linhardt, R. J., Depla, E., Boson, B., Cosset, F. L., Patel, A. H., Blum, H. E. & other authors (2006). Viral and cellular determinants of the hepatitis C virus envelope-heparan sulfate interaction. *J Virol* **80**, 10579-10590.

Bartosch, B., Bukh, J., Meunier, J.-C., Granier, C., Engle, R. E., Blackwelder, W. C., Emerson, S. U., Cosset, F.-L. & Purcell, R. H. (2003a). *In vitro* assay for neutralizing antibody to hepatitis C virus: Evidence for broadly conserved neutralization epitopes. *P Natl Acad Sci USA* **100**, 14199-14204.

Bartosch, B., Dubuisson, J. & Cosset, F.-L. (2003b). Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *J Exp Med* **197**, 633-642.

Bartosch, B., Verney, G., Dreux, M., Donot, P., Morice, Y., Penin, F., Pawlotsky, J.-M., Lavillette, D. & Cosset, F.-L. (2005). An interplay between hypervariable region 1 of the hepatitis C virus E2 glycoprotein, the scavenger receptor BI and high-density lipoprotein promotes both enhancement of infection and protection against neutralizing antibodies. *J Virol* **79**, 8217-8229.

Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A. & other authors (2003c). Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J Biol Chem* **278**, 41624-41630.

Baumert, T. F., Ito, S., Wong, D. T. & Liang, T. J. (1998). Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *J Virol* **72**, 3827-3836.

Baumert, T. F., Vergalla, J., Satoi, J., Thomson, M., Lechmann, M., Herion, D., Greenberg, H. B., Ito, S. & Liang, T. J. (1999). Hepatitis C virus-like particles synthesized in insect cells as a potential vaccine candidate. *Gastroenterology* **117**, 1397-1407.

Bell, B., Ghany, M., Hoofnagle, J., Kleiner, D., Liang, J., McHutchison, J. & Seeff, L. (2010). Chronic hepatitis C: Current disease management. *NIH*

Publication No. 10-4230. http://digestive.niddk.nih.gov/ddiseases/pubs/chronichepc

Benedicto, I., Molina-Jimenez, F., Barreiro, O., Maldonado-Rodriguez, A., Prieto, J., Moreno-Otero, R., Aldabe, R., Lopez-Cabrera, M. & Majano, P. L. (2008). Hepatitis C virus envelope components alter localisation of hepatocyte tight junction-associated proteins and promote occludin retention in the endoplasmic reticulum. *Hepatology* **48**, 1044-1053.

Bernasconi, N. L., Traggiai, E. & Lanzavecchia, A. (2002). Maintenance of serological memory by polyclonal activation of human memory B cells. *Science* **298**, 2199-2202.

Bjoro, K., Froland, S., Yun, Z., Samdal, H. & Haaland, T. (1994). Hepatitis C infection in patients with primary hypogammaglobulinemia after treatment with contaminated immune globulin. *N Engl J Med* **331**, 1607-1611.

Blanchard, E., Belouzard, S., Goueslain, L., Wakita, T., Dubuisson, J., Wychowski, C. & Rouille, Y. (2006). Hepatitis C virus entry depends on clathrin-mediated endocytosis. *J Virol* **80**, 6964-6972.

Boder, E. T., Midelfort, K. S. & Wittrup, K. D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *P Natl Acad Sci USA* **97**, 10701-10705.

Boder, E. T. & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553-557.

Bonnafous, P., Perrault, M., Le Bihan, O., Bartosch, B., Lavillette, D., Penin, F., Lambert, O. & Pecheur, E. I. (2010). Characterization of hepatitis C virus pseudoparticles by cryo-transmission electron microscopy using functionalized magnetic nanobeads. *J Gen Virol* **91**, 1919-1930.

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

Boulianne, G. L., Hozumi, N. & Shulman, M. J. (1984). Production of functional chimaeric mouse/human antibody. *Nature* **312**, 643-646.

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

Bradbury, A. R. M. & Marks, J. D. (2004). Antibodies from phage antibody libraries. *J Immunol Methods* **290**, 29-49.

Bradley, D., McCaustland, K., Krawczynski, K., Spelbring, J., Humphrey, C. & Cook, E. H. (1991). Hepatitis C virus: buoyant density of the factor VIII-derived isolate in sucrose. *J Med Virol* **34**, 206-208.

Brazzoli, M., Bianchi, A., Filippini, S., Weiner, A., Zhu, Q., Pizza, M. & Crotta, S. (2008). CD81 is a central regulator of cellular events required for hepatitis C virus infection of human hepatocytes. *J Virol* **82**, 8316-8329.

Brimacombe, C. L., Grove, J., Meredith, L. W., Hu, K., Syder, A. J., Flores, M. V., Timpe, J. M., Krieger, S. E., Baumert, T. F. & other authors (2011). Neutralizing antibody-resistant hepatitis C virus cell-to-cell transmission. *J Virol* **85**, 596-605.

Brochet, X., Lefranc, M. P. & Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* **36**, W503-508.

Broering, T. J., Garrity, K. A., Boatright, N. K., Sloan, S. E., Sandor, F., Thomas Jr, W. D., Szabo, G., Finberg, R. W., Ambrosino, D. M. & other authors (2009). Identification and characterization of broadly neutralizing human monoclonal antibodies directed against the E2 envelope glycoprotein of hepatitis C virus. *J Virol* **83**, 12473-12482.

Brown, R. J. P., Juttla, V. S., Tarr, A. W., Finnis, R., Irving, W. L., Hemsley, S., Flower, D. R., Borrow, P. & Ball, J. K. (2005). Evolutionary dynamics of

hepatitis C virus envelope genes during chronic infection. *J Gen Virol* **86**, 1931-1942.

Brown, R. J. P., Tarr, A. W., McClure, C. P., Juttla, V. S., Taguiri, N., Irving, W. L. & Ball, J. K. (2007). Cross-genotype characterization of genetic diversity and molecular adaptation in hepatitis C virus envelope glycoprotein genes. *J Gen Virol* **88**, 458-469.

Bull, R. A., Luciani, F., McElroy, K., Gaudieri, S., Pham, S. T., Chopra, A., Cameron, B., Maher, L., Dore, G. J. & other authors (2011). Sequential Bottlenecks Drive Viral Evolution in Early Acute Hepatitis C Virus Infection. *PLoS Pathog* **7**, e1002243.

Bunch, T. A., Grinblat, Y. & Goldstein, L. S. B. (1988). Characterisation and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. *Nucleic Acids Res* **16**, 1043-1061.

Burlone, M. E. & Budkowska, A. (2009). Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. *J Gen Virol* **90**, 1055-1070.

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* **51**, 253-265.

Butler, J. M., Shen, Y. & McCord, B. R. (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* **48**, 1054-1064.

Cappellaro, C., Baldermann, C., Rachel, R. & Tanner, W. (1994). Mating type-specific cell - cell recognition of *Saccharomyces cerevisiae*: cell wall attachment and active sites of a- and α-agglutinin. *EMBO J* **13**, 4737-4744.

Cappellaro, C., Hauser, K., Mrsa, V., Watzele, M., Watzele, G., Gruber, C. & Tanner, W. (1991). *Saccharomyces cerevisiae* a and α-agglutinin: characterisation of their molecular interaction. *EMBO J* **10**, 4081-4088.

Carmen, S. & Jermutus, L. (2002). Concepts in antibody phage display. *Brief Funct Genomic Proteomic* **1**, 189-203.

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

Catanese, M. T., Ansuini, H., Graziani, R., Huby, T., Moreau, M., Ball, J. K., Paonessa, G., Rice, C. M., Cortese, R. & other authors (2010). Role of scavenger receptor class B type I in hepatitis C virus entry: kinetics and molecular determinants. *J Virol* **84**, 34-43.

Chang, K. M., Thimme, R., Melpolder, J. J., Oldach, D., Pemberton, J., Moorhead-Loudis, J., McHutchison, J. G., Alter, H. J. & Chisari, F. V. (2001). Differential CD4(+) and CD8(+) T-cell responsiveness in hepatitis C virus infection. *Hepatology* **33**, 267-276.

Cherbas, L. & Cherbas, P. (2000). *Drosophila* protocols. In *Drosophila cell culture and transformation*. Edited by W. Sullivan, M. Ashburner & S. R. Hawley: Cold Spring Harbor Laboratory Press.

Choo, Q. L., Kuo, G., Ralston, R., Weiner, A., D., C., Van Nest, G., Han, J., Berger, K., Thudium, K. & other authors (1994). Vaccination of chimpanzees against infection by the hepatitis C virus. *P Natl Acad Sci USA* **91**, 1294-1298.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**, 359-362.

Chung, A., Rollman, E., Johansson, S., Kent, S. J. & Stratov, I. (2008). The utility of ADCC responses in HIV infection. *Curr HIV Res* **6**, 515-519.

Chung, D. T., Drabek, J., Opel, K. L., Butler, J. M. & McCord, B. R. (2004). A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets. *J Forensic Sci* **49**, 733-740.

Clayton, R. F., Owsianka, A., Aitken, J., Graham, S., Bhella, D. & Patel, A. H. (2002). Analysis of the antigenicity and topology of E2 glycoprotein present on the recombinant hepatitis C virus-like particles. *J Virol* **76**, 7672-7682.

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

Cocquerel, L., Voisset, C. & Dubuisson, J. (2006). Hepatitis C virus entry: potential receptors and their biological function. *J Gen Virol* **87**, 1075-1084.

Conde, R., Cueva, R., Pablo, G., Polaina, J. & Larriba, G. (2004). A search for hyperglycosylation signals in Yeast glycoproteins. *J Biol Chem* **279**, 43789-43798.

Connelly, M. A. & Williams, D. L. (2003). SR-BI and cholesterol uptake into steroidogenic cells. *Trends Endocrinol Metab* **14**, 467-472.

Corti, D., Voss, J., Gamblin, S. J., Codoni, G., Macagno, A., Jarrossay, D., Vachieri, S. G., Pinna, D., Minola, A. & other authors (2011). A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* **333**, 850-856.

Cox, A. L., Mosbruger, T., Lauer, G. M., Pardoll, D., Thomas, D. L. & Ray, S. C. (2005). Comprehensive analyses of CD8+T cell responses during longitudinal study of acute human hepatitis C. *Hepatology* **42**, 104-112.

Cross, F., Hartwell, L. H., Jackson, C. & Konopka, J. B. (1988). Conjugation in *Saccharomyces cerevisiae*. *Ann Rev Cell Biol* **4**, 429-457.

Cukierman, L., Meertens, L., Bertaux, C., Kajumo, F. & Dragic, T. (2009). Residues in a highly conserved claudin-1 motif are required for hepatitis C virus entry and mediate the formation of cell-cell contacts. *J Virol* **83**, 5477-5484.

Culp, J. S., Johansen, H., Hellmig, B., Beck, J., Matthews, T. J., Delers, A. & Rosenberg, M. (1991). Regulated expression allows high level production and secretion of HIV-1 gp120 envelope glycoprotein in *Drosophila* Schneider cells. *Biotechnology* **9**, 173-177.

Davis, G. L., Nelson, D. R., Terrault, N., Pruett, T. L., Schiano, T. D., Fletcher, C. V., Sapan, C. V., Riser, L. N., Li, Y. & other authors (2005). A randomized, open-label study to evaluate the safety and pharmacokinetics of human hepatitis C immune globulin (Civacir) in liver transplant recipients. *Liver Transpl* **11**, 941-949.

de Carvalho-Mello, I. M. V. G., Filho, J. E. M., Gomes-Gouvea, M. S., de Mello Malta, F., de Queiroz, A. T. L., Pinho, J. R. R. & Carrilho, F. J. (2010). Molecular evidence of horizontal transmission of hepatitis C virus within couples. *J Gen Virol* **91**, 691-696.

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

Deml, L., Wolf, H. & Wagner, R. (1999). High level expression of hepatitis B virus surface antigen in stably transfected *Drosophila* Schneider-2 cells. *J Virol Methods* **79**, 191-203.

Denovel, G., van Helden, J., Bauer, R. & Preisel-Simmons, B. (2004). Performance of a new hepatitis C assay on the Bayer ADVIA Centaur Immunoassay System. *Clin Lab* **50**, 75-82.

Dessain, S. K., Adekar, S. P., Stevens, J. B., Carpenter, K. A., Skorski, M. L., Barnoski, B. L., Goldsby, R. A. & Weinberg, R. A. (2004). High efficiency creation of human monoclonal antibody-producing hybridomas. *J Immunol Methods* **291**, 109-122.

Dhillon, S., Witteveldt, J., Gatherer, D., Owsianka, A. M., Zeisel, M. B., Zahid, M. N., Rychlowska, M., Foung, S. K., Baumert, T. F. & other authors (2010). Mutations within a conserved region of the Hepatitis C virus E2

glycoprotein that influence virus-receptor interactions and sensitivity to neutralizing antibodies. *J Virol* **84**, 5494-5507.

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

Dowd, K. A., Netski, D. M., Wang, X. H., Cox, A. L. & Ray, S. C. (2009). Selection pressure from neutralizing antibodies drives sequence evolution during acute infection with hepatitis C virus. *Gastroenterology* **136**, 2377-2386.

Dreux, M., Boson, B., Ricard-Blum, S., Molle, J., Lavillette, D., Bartosch, B., Pecheur, E.-I. & Cosset, F. L. (2007). The exchangeable apolipoprotein ApoC-I promotes membrane fusion of hepatitis C virus. *J Biol Chem* **282**, 32357-32369.

Dreux, M., Pietschmann, T., Granier, C., Voisset, C., Ricard-Blum, S., Mangeot, P.-E., Keck, Z., Foung, S., Vu-Dac, N. & other authors (2006). High density lipoprotein inhibits hepatitis C virus-neutralizing antibodies by stimulating cell entry via activation of the scavenger receptor BI. *J Biol Chem* **27**, 18285-18295.

Drummer, H. E., Boo, I., Maerz, A. L. & Poumbourios, P. (2006). A conserved Gly(436)-Trp-Leu-Ala-Gly-Leu-Phe-Tyr motif in hepatitis C virus glycoprotein E2 is a determinant of CD81 binding and viral entry. *J Virol* **80**, 7844-7853.

Drummer, H. E. & Poumbourios, P. (2004). Hepatitis C virus glycoprotein E2 contains a membrane-proximal heptad repeat sequence that is essential for E1E2 glycoprotein heterodimerization and viral entry. *J Biol Chem* **279**, 30066-30072.

Dubuisson, J., Hsu, H. H., Cheung, R. C., Greenberg, H. B., Russell, D. G. & Rice, C. M. (1994). Formation and intracellular localisation of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses. *J Virol* **68**, 6147-6160.

Dudley, D. D., Chaudhuri, J., Bassing, C. H. & Alt, F. W. (2005). Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol* **86**, 43-112.

Dusheiko, G., Schmilovitz-Weiss, H., Brown, D., McOmish, F., Yap, P.-L., Sherlock, S., McIntyre, N. & Simmonds, P. (1994). Hepatitis C Virus Genotypes: an investigation of type-specific differences in geographic origin and disease. *Hepatology* **19**, 13-18.

Eales, L.-J. (2001). *Immunology for Life Scientists*. Chichester, UK: John Wiley and Sons Ltd.

Edwards, A., Civitello, A., Hammond, H. A. & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* **49**, 746-756.

Edwards, V. C., Tarr, A. W., Urbanowicz, R. A. & Ball, J. K. (2012). The role of neutralizing antibodies in hepatitis C virus infection. *J Gen Virol* **93**, 1-19.

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

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* **5**, 435-445.

Ennishi, D., Terui, Y., Yokoyama, M., Mishima, Y., Takahashi, S., Takeuchi, K., Okamoto, H., Tanimoto, M. & Hatake, K. (2008). Monitoring serum hepatitis C virus (HCV) RNA in patients with HCV-infected CD20-positive B-cell lymphoma undergoing rituximab combination chemotherapy. *Am J Hematol* **83**, 59-62.

Evans, M. J., Rice, C. M. & Goff, S. P. (2004). Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *P Natl Acad Sci USA* **101**, 13038-13043.

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

Fafi-Kremer, S., Fofana, I., Soulier, E., Carolla, P., Meuleman, P., Leroux-Roels, G., Patel, A. H., Cosset, F. L., Pessaux, P. & other authors (2010). Viral entry and escape from antibody-mediated neutralization influence hepatitis C virus reinfection in liver transplantation. *J Exp Med* **207**, 2019-2031.

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

Falkowska, E., Kajumo, F., Garcia, E., Reinus, J. & Dragic, T. (2007). Hepatitis C virus envelope glycoprotein E2 glycans modulate entry, CD81 binding and neutralization. *J Virol* **81**, 8072-8079.

Farci, P., Alter, H. J., Wong, D. C., Miller, R. H., Govindarajan, S., Engle, R., Shapiro, M. & Purcell, R. H. (1994). Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated *in vitro* neutralization. *P Natl Acad Sci USA* **91**, 7792-7796.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J. C., Strazzera, A., Chien, D. Y., Munoz, S. J. & other authors (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339-344.

Farci, P., Shimoda, A., Wong, D., Cabezon, T., De Gioannis, D., Strazzera, A., Shimizu, Y., M., S., Alter, H. J. & other authors (1996). Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 glycoprotein. *P Natl Acad Sci USA* **93**, 15394-15399.

Farquhar, M. J., Hu, K., Harris, H. J., Davis, C., Brimacombe, C. L., Fletcher, S. J., Baumert, T. F., Rappoport, J. Z., Balfe, P. & other authors (2012). Hepatitis C virus induces CD81 and claudin-1 endocytosis. *J Virol* **86**, 4305-4316.

Fehrenbacher, G., Perry, K. & Thorner, J. (1978). Cell-cell recognition in *Saccharomyces cerevisiae*: Regulation of mating-specific adhesion. *J Bacteriol* **134**, 893-901.

Feray, C., Gigou, M., Samuel, D., Ducot, B., Maisonneuve, P., Reynes, M., Bismuth, A. & Bismuth, H. (1998). Incidence of hepatitis C in patients receiving different preparations of hepatitis B immunoglobulins after liver transplantation. *Ann Intern Med* **128**, 810-816.

Firth, A. E. & Patrick, W. M. (2005). Statistics of protein library construction. *Bioinformatics* **21**, 3314-3315.

Flint, M., Dubuisson, J., Maidens, C., Harrop, R., Guile, G. R., Borrow, P. & McKeating, J. A. (2000). Functional characterisation of intracellular and secreted forms of a truncated hepatitis C virus E2 glycoprotein. *J Virol* **74**, 702-709.

Flint, M., Logvinoff, C., Rice, C. M. & McKeating, J. A. (2004). Characterization of infectious retroviral pseudotype particles bearing hepatitis C virus glycoproteins. *J Virol* **78**, 6875-6882.

Flint, M., Maidens, C., Loomis-Price, L. D., Shotton, C., Dubuisson, J., Monk, P., Higginbottom, A., Levy, S. & McKeating, J. (1999a). Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. *J Virol* **73**, 6235-6244.

Flint, M., Thomas, J. M., Maidens, C. M., Shotton, C., Levy, S., Barclay, W. S. & McKeating, J. A. (1999b). Functional analysis of cell-surface expressed Hepatitis C Virus E2 glycoprotein. *J Virol* **73**, 6782-6790.

Fofana, I., Krieger, S. E., Grunert, F., Glauben, S., Xiao, F., Fafi-Kremer, S., Soulier, E., Royer, C., Thumann, C. & other authors (2010). Monoclonal anti-claudin 1 antibodies prevent hepatitis C virus infection of primary human hepatocytes. *Gastroenterology* **139**, 953-964.

Fournillier, A., Wychowski, C., Boucreux, D., Baumert, T. F., Meunier, J. C., Jacobs, D., Muguet, S., Depla, E. & Inchauspe, G. (2001). Induction of hepatitis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites. *J Virol* **75**, 12088-12097.

Frank, C., Mohamed, M. K., Strickland, T. G., Lavanchy, D., Arthur, R. R., Magder, L. S., El Khoby, T., Abdel-Wahab, Y., Ohn, E. S. A. & other authors (2000). The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* **355**, 887-891.

Frazier, R. A., Ames, J. M. & Nursten, H. E. (2000). Background theory and principles of capillary electrophoresis In *Capillary electrophoresis for food analysis*, pp. 1-7: Royal Society of Chemistry.

Frey, S. E., Houghton, M., Coates, S., Abrignani, S., Chien, D., Rosa, D., Pileri, P., Ray, R., Di Bisceglie, A. M. & other authors (2010). Safety and immunogenicity of HCV E1E2 vaccine adjuvanted with MF59 administered to healthy adults. *Vaccine* **28**, 6367-6373.

Gal-Tanamy, M., Keck, Z.-Y., Yi, M., McKeating, J. A., Patel, A. H., Foung, S. K. H. & Lemon, S. M. (2008). In vitro selection of a neutralization-resistant hepatitis C virus escape mutant. *P Natl Acad Sci USA* **105**, 19450-19455.

Garrone, P., Fluckiger, A.-C., Mangeot, P. E., Gauthier, E., Dupeyrot-Lacas, P., Mancip, J., Cangialosi, A., Du Chene, I., LeGrand, R. & other authors (2011). A Prime-boost strategy using virus-like particles pseudotyped for HCV proteins triggers broadly neutralizing antibodies in macaques. *Sci Transl Med* **3**, 94ra71.

Gaudy-Graffin, C., Lesage, G., Kousignian, I., Laperche, S., Girault, A., Dubois, F., Goudeau, A. & Barin, F. (2010). Use of an anti-hepatitis C virus (HCV) IgG avidity assay to identify recent HCV infection. *J Clin Microbiol* **48**, 3281-3287.

Ge, D., Fellay, J., Thompson, A. J., Simon, J. S., Shianna, K. V., Urban, T. J., Heinzen, E. L., Qiu, P., Bertelsen, A. H. & other authors (2009). Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399-401.

Giang, E., Dorner, M., Prentoe, J. C., Dreux, M., Evans, M. J., Bukh, J., Rice, C. M., Ploss, A., Burton, D. R. & other authors (2012). Human broadly neutralizing antibodies to the envelope glycoprotein complex of hepatitis C virus. *Proc Natl Acad Sci U S A* **109**, 6205-6210.

Gilmartin, A. A., Lamp, B., Rumenapf, T., Persson, M. A. A., Rey, F. A. & Krey, T. (2012). High-level secretion of recombinant monomeric murine and human single-chain Fv antibodies from Drosophila S2 cells. *Protein Eng Des Sel* **25**, 59-66.

Goffard, A., Callens, N., Bartosch, B., Wychowski, C., Cosset, F. L., Montpellier, C. & Dubuisson, J. (2005). Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins. *J Virol* **79**, 8400-8409.

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

Gottwein, J. M., Scheel, T. K. H., Jensen, T. B., Lademann, J. B., Prentoe, J. C., Knudsen, M. L., Hoegh, A. M. & Bukh, J. (2009). Development and characterization of hepatitis C virus genotype 1-7 cell culture systems: Role of CD81 and scavenger receptor class B type I and effect of antiviral drugs. *Hepatology* **49**, 364-377.

Gouklani, H., Bull, R. A., Beyer, C., Coulibaly, F., Gowans, E. J., Drummer, H. E., Netter, H. J., White, P. A. & Haqshenas, G. (2012). Hepatitis C virus

nonstructural protein 5B is involved in virus morphogenesis. *J Virol* **86**, 5080-5088.

Grabherr, R., Ernst, W., Oker-Blom, C. & Jones, I. (2001). Development in the use of baculoviruses for the surface display of complex eukaryotic proteins. *Trends Biotechnol* **19**, 231-236.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. (1993a). Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J Virol* **67**, 2832-2843.

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

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

Grandea, A. G., 3rd, Olsen, O. A., Cox, T. C., Renshaw, M., Hammond, P. W., Chan-Hui, P. Y., Mitcham, J. L., Cieplak, W., Stewart, S. M. & other authors (2010). Human antibodies reveal a protective epitope that is highly conserved among human and nonhuman influenza A viruses. *P Natl Acad Sci USA* **107**, 12658-12663.

Grangeot-Keros, L., Mayaux, M. J., Lebon, P., Freymuth, F., Eugene, G., Stricker, R. & Dussaix, E. (1997). Value of cytomegalovirus (CMV) IgG avidity index for the diagnosis of primary CMV infection in pregnant women. *J Infect Dis* **175**, 944-946.

Gray, J. J. (1995). Avidity of EBV VCA-specific IgG antibodies: distinction between recent primary infection, past infection and reactivation. *J Virol Methods* **52**, 95-104.

Green, L. L., Hardy, M. C., Maynard-Currie, C. E., Tsuda, H., Louie, D. M., Mendez, M. J., Abderrahim, H., Noguchi, M., Smith, D. H. & other authors (1994). Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nat Genet* **7**, 13-21.

Griffin, S. D., Beales, L. P., Clarke, D. S., Worsfold, O., Evans, S. D., Jaeger, J., Harris, M. P. & Rowlands, D. J. (2003). The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Lett* **535**, 34-38.

Grove, J., Nielsen, S., Zhong, J., Bassendine, M. F., Drummer, H. E., Balfe, P. & McKeating, J. A. (2008). Identification of a residue in Hepatitis C virus E2 glycoprotein that determines scavenger receptor BI and CD81 receptor dependency and sensitivity to neutralizing antibodies. *J Virol* **82**, 12020-12029.

Gruener, N. H., Lechner, F., Jung, M. C., Diepolder, H., Gerlach, T., Lauer, G., Walker, B., Sullivan, J., Phillips, R. & other authors (2001). Sustained dysfunction of antiviral CD8(+) T lymphocytes after infection with hepatitis C virus. *J Virol* **75**, 5550-5558.

Haberstroh, A., Schnober, E. K., Zeisel, M. B., Carolla, P., Barth, H., Blum, H. E., Cosset, F. L., Koutsoudakis, G., Bartenschlager, R. & other authors (2008). Neutralizing host responses in hepatitis C virus infection target viral entry at postbinding steps and membrane fusion. *Gastroenterology* **135**, 1719-1728.

Hadlock, K. G., Lanford, R. E., Perkins, S., Rowe, J., Yang, Q., Levy, S., Pileri, P., Abrignani, S. & Foung, S. K. H. (2000). Human monoclonal antibodies that inhibit binding of hepatitis C virus E2 protein to CD81 and recognise conserved conformational epitopes. *J Virol* **74**, 10407-10416.

Harris, H. J., Davis, C., Mullins, J. G., Hu, K., Goodall, M., Farquhar, M. J., Mee, C. J., McCaffrey, K., Young, S. & other authors (2010). Claudin

association with CD81 defines hepatitis C virus entry. *J Biol Chem* **285**, 21092-21102.

Harris, H. J., Farquhar, M. J., Mee, C. J., Davis, C., Reynolds, G. M., Jennings, A., Hu, K., Yuan, F., Deng, H. & other authors (2008). CD81 and claudin 1 coreceptor association: role in hepatitis C virus entry. *J Virol* **82**, 5007-5020.

Hedman, K. & Rousseau, S. A. (1989). Measurement of avidity of specific IgG for verification of recent primary rubella. *J Med Virol* **27**, 288-292.

Heile, J. M., Fong, Y.-L., Rosa, D., Berger, K., Saletti, G., Campagnoli, S., Bensi, G., Capo, S., Coates, S. & other authors (2000). Evaluation of hepatitis C virus glycoprotein E2 for vaccine design: an endoplasmic reticulum-retained recombinant protein is superior to secreted recombinant protein and DNA-based vaccine candidates. *J Virol* **74**, 6885-6892.

Helle, F., Goffard, A., Morel, V., Duverlie, G., McKeating, J., Keck, Z.-Y., Foung, S., Penin, F., Dubuisson, J. & other authors (2007). The neutralizing activity of anti-hepatitis C virus antibodies is modulated by specific glycans on the E2 envelope protein. *J Virol* **81**, 8101-8111.

Helle, F., Vieyres, G., Elkrief, L., Popescu, C.-I., Wychowski, C., Descamps, V., Castelain, S., Roingeard, P., Duverlie, G. & other authors (2010). Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. *J Virol* **84**, 11905-11915.

Heo, T.-H., Chang, J.-H., Lee, J.-W., Foung, S. K. H., Dubuisson, J. & Kang, C.-Y. (2004). Incomplete humoral immunity against hepatitis C virus is linked with distinct recognition of putative multiple receptors by E2 envelope glycoprotein. *J Immunol* **173**, 446-455.

Holmes, E. C. (2010). The RNA Virus Quasispecies: Fact or Fiction? *J Mol Biol* **400**, 271-273.

Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol* **23**, 1105-1116.

Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., P., H. & G., W. (1991). Multi-subunit proteins on the surface of filamentous phage:methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res* **19**, 4133-4137.

Houghton, M. & Abrignani, S. (2005). Prospects for a vaccine against the hepatitis C virus. *Nature* **435**, 961-966.

Hsu, M., Zhang, J., Flint, M., Logvinoff, C., Cheng-Mayer, C., Rice, C. M. & McKeating, J. A. (2003). Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *P Natl Acad Sci USA* **100**, 7271-7276.

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

Humphreys, I., Fleming, V., Fabris, P., Parker, J., Schulenberg, B., Brown, A., Demetriou, C., Gaudieri, S., Pfafferott, K. & other authors (2009). Full-length characterization of hepatitis C virus subtype 3a reveals novel hypervariable regions under positive selection during acute infection. *J Virol* **83**, 11456-11466.

Hwang, W. Y. & Foote, J. (2005). Immunogenicity of engineered antibodies. *Methods* **36**, 3-10.

IHGSC (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Irving, W. L. & Brown, R. J. P. (2010). Acute hepatitis C virus infection: A dynamic - and challenging - concept. *J Infect Dis* **202**, 1765-1767.

Ishido, S., Fujita, T. & Hotta, H. (1998). Complex formation of NS5B with NS3 and NS4A proteins of hepatitis C virus. *Biochem Biophys Res Commun* **244**, 35-40.

Ivey-Hoyle, M. & Rosenberg, M. (1990). Rev-dependent expression of Human Immunodeficiency Virus type 1 gp160 in *Drosophila melanogaster* cells. *Mol Cell Biol* **10**, 6152-6159.

Iwaki, T. & Castellino, F. J. (2008). A single plasmid transfection that offers a significant advantage associated with puromycin selection in *Drosophila* Schneider S2 cells expressing heterologous proteins. *Cytotechnology* **57**, 45-49.

Jacobson, I. M., McHutchison, J. G., Boyer, T. D., Schiff, E. R., Everson, G. T., Pockros, P. J., Chasen, R. M., Vierling, J. M., Lawitz, E. J. & other authors (2010). GI-5005 therapeutic vaccine plus Peg-IFN/ribavirin significantly improves virologic response and ALT normalisation at end-of-treatment and improves SVR24 compared to Peg-IFN/ribavirin in genotype chronic HCV patients. *J Hepatol* **52**, A2006.

Jiang, X.-R., Song, A., Bergelson, S., Arroll, T., Parekh, B., May, K., Chung, S., Strouse, R., Mire-Sluis, A. & other authors (2011). Advances in the assessment and control of the effector functions of therapeutic antibodies. *Nat Rev Drug Discovery* **10**, 101-110.

Jirasko, V., Montserret, R., Appel, N., Janvier, A., Eustachi, L., Brohm, C., Steinmann, E., Pietschmann, T., Penin, F. & other authors (2008). Structural and functional characterization of nonstructural protein 2 for its role in hepatitis C virus assembly. *J Biol Chem* **283**, 28546-28562.

Jo, J., Lohmann, V., Bartenschlager, R. & Thimme, R. (2011). Experimental models to study the immunobiology of hepatitis C virus. *J Gen Virol* **92**, 477-493.

Johansson, D. X., Drakenberg, K., Hopmann, K. H., Schmidt, A., Yari, F., Hinkula, J. & Persson, M. A. A. (2007a). Efficient expression of recombinant human monoclonal antibodies in *Drosophila* S2 cells. *J Immunol Methods* **318**, 37-46.

Johansson, D. X., Voisset, C., Tarr, A. W., Aung, M., Ball, J. K., Dubuisson, J. & Persson, M. A. (2007b). Human combinatorial libraries yield rare antibodies that broadly neutralize hepatitis C virus. *P Natl Acad Sci USA* **104**, 16269-16274.

Johnson, S., Oliver, C., Prince, G. A., Hemming, V. G., Pfarr, D. S., Wang, S. C., Dormitzer, M., O'Grady, J., Koenig, S. & other authors (1997). Development of a humanized monoclonal antibody (MEDI-493) with potent in vitro and in vivo activity against respiratory syncytial virus. *J Infect Dis* **176**, 1215-1224.

Jones, C. T., Murray, C. L., Eastman, D. K., Tassello, J. & Rice, C. M. (2007). Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *J Virol* **81**, 8374-8383.

Jones, D. M., Patel, A. H., Targett-Adams, P. & McLauchlan, J. (2009). The hepatitis C virus NS4B protein can trans-complement viral RNA replication and modulates production of infectious virus. *J Virol* **83**, 2163-2177.

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**, 522-525.

Junker, A. K. & Tilley, P. (1994). Varicella-zoster virus antibody avidity and IgG-subclass patterns in children with recurrent chickenpox. *J Med Virol* **43**, 119-124.

Kachko, A., Kochneva, G., Sivolobova, G., Grazhdantseva, A., Lupan, T., Zubkova, I., Wells, F., Merchlinsky, M., Williams, O. & other authors (2011). New neutralizing antibody epitopes in hepatitis C virus envelope glycoproteins are revealed by dissecting peptide recognition profiles. *Vaccine* **30**, 69-77.

Kanno, A. & Kazuyama, Y. (2002). Immunoglobulin G antibody avidity assay for serodiagnosis of hepatitis C virus infection. *J Med Virol* **68**, 229-233.

Karlsson Hedestam, G. B., Fouchier, R. A. M., Phogat, S., Burton, D. R., Sodroski, J. & Wyatt, R. T. (2008). The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat Rev Microbiol* **6**, 143-155.

Karpas, A., Dremucheva, A. & Czepulkowski, B. H. (2001). A human myeloma cell line suitable for the generation of human monoclonal antibodies. *P Natl Acad Sci USA* **98**, 1799-1804.

Kato, N., Ootsuyama, Y., Sekiya, H., Ohkoshi, S., Nakazawa, T., Hijikata, M. & Shimotohno, K. (1994). Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection. *J Virol* **68**, 4776-4784.

Kato, N., Sekiya, H., Ootsuyama, Y., Nakazawa, T., Hijikata, M., Ohkoshi, S. & Shimotohno, K. (1993). Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. *J Virol* **67**, 3923-3930.

Kato, T., Furusaka, A., Miyamoto, M., Date, T., Yasui, K., Hiramoto, J., Nagayama, K., Tanaka, T. & Wakita, T. (2001). Sequence analysis of Hepatitis C virus isolated from a fulminant hepatitis patient. *J Med Virol* **64**, 334-339.

Katze, M. G., He, Y. & Gale, M., Jr. (2002). Viruses and interferon: a fight for supremacy. *Nat Rev Immunol* **2**, 675-687.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J. L., Wang, J. & Shimizu, N. (1997). One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res* **7**, 250-261.

Keck, Z.-Y., Li, T. K., Xia, J. M., Gal-Tanamy, M., Olson, O., Li, S. H., Patel, A. H., Ball, J. K., Lemon, S. M. & other authors (2008a). Definition of a conserved immunodominant domain on hepatitis C virus E2 glycoprotein by neutralizing human monoclonal antibodies. *J Virol* **82**, 6061-6066.

Keck, Z.-Y., Olson, O., Gal-Tanamy, M., Xia, J., Patel, A. H., Dreux, M., Cosset, F. L., Lemon, S. M. & Foung, S. K. H. (2008b). A point mutation leading to hepatitis C virus escape from neutralization by a monoclonal antibody to a conserved conformational epitope. *J Virol* **82**, 6067-6072.

Keck, Z.-Y., Op De Beeck, A., Hadlock, K. G., Xia, J., Li, T.-K., Dubuisson, J. & Foung, S. K. H. (2004a). Hepatitis C virus has three immunogenic domains containing conformational epitopes with distinct properties and biological functions. *J Virol* **78**, 9224-9232.

Keck, Z.-Y., Saha, A., Xia, J., Wang, Y., Lau, P., Krey, T., Rey, F. A. & Foung, S. K. H. (2011). Mapping a region of hepatitis C virus E2 that is responsible for escape from neutralizing antibodies and a core CD81-binding region that does not tolerate neutralization escape mutations. *J Virol* **85**, 10451-10463.

Keck, Z.-Y., Sung, V. M. H., Perkins, S., Rowe, J., Paul, S., Liang, J. T., Lai, M. M. C. & Foung, S. K. H. (2004b). Human monoclonal antibody to hepatitis C virus E1 glycoprotein that blocks virus attachment and viral infectivity. *J Virol* **78**, 7257-7263.

Keck, Z. Y., Xia, J., Wang, Y., Wang, W., Krey, T., Prentoe, J., Carlsen, T., Li, A. Y., Patel, A. H. & other authors (2012). Human monoclonal antibodies to a novel cluster of conformational epitopes on HCV e2 with resistance to neutralization escape in a genotype 2a isolate. *PLoS Pathog* **8**, e1002653.

Keller, M. A. & Stiehm, E. R. (2000). Passive immunity in prevention and treatment of infectious diseases. *Clin Microbiol Rev* **13**, 602-614.

Kirkpatrick, R. B., Ganguly, S., Angelichio, M., Griego, S., Shatzman, A., Silverman, C. & Rosenburg, M. (1995). Heavy chain dimers as well as complete antibodies are efficiently formed and secreted from Drosophila via a BiP-mediated pathway. *J Biol Chem* **270**, 19800-19805.

Klintschar, M., Al-Hammadi, N. & Reichenpfader, B. (1999). Population genetic studies on the tetrameric short tandem repeat loci D3S1358, vWA, FGA,

D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 in Egypt. *Forensic Sci Int* **104**, 23-31.

Kobayashi, M., Bennett, M. C., Bercot, T. & Singh, I. R. (2006). Functional analysis of hepatitis C virus envelope proteins using a cell-cell fusion assay. *J Virol* **80**, 1817-1825.

Koelsch, K., Zheng, N.-Y., Zhang, Q., Duty, A., Helms, C., Mathias, M. D., Jared, M., Smith, K., Capra, J. D. & other authors (2007). Mature B cells class switched to IgD are autoreactive in healthy individuals. *J Clin Invest* **117**, 1558-1565.

Kohler, G. & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497.

Kolykhalov, A. A., Mihalik, K., Feinstone, S. M. & Rice, C. M. (2000). Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *J Virol* **74**, 2046-2051.

Kopp, M., Murray, C. L., Jones, C. T. & Rice, C. M. (2010). Genetic analysis of the carboxy-terminal region of the hepatitis C virus core protein. *J Virol* **84**, 1666-1673.

Koutsoudakis, G., Kaul, A., Steinmann, E., Kallis, S., Lohmann, V., Pietschmann, T. & Bartenschlager, R. (2006). Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses. *J Virol* **80**, 5308-5320.

Krause, J. C., Ekiert, D. C., Tumpey, T. M., Smith, P. B., Wilson, I. A. & Crowe, J. E., Jr. (2011). An insertion mutation that distorts antibody binding site architecture enhances function of a human antibody. *MBio* **2**, e00345-00310.

Krey, T., d'Alayer, J., Kikuti, C. M., Saulnier, A., Damier-Piolle, L., Petitpas, I., Johansson, D. X., Tawar, R. G., Baron, B. & other authors (2010). The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organisation of the molecule. *PLoS Pathog* **6**, e1000762.

Krieger, S. E., Zeisel, M. B., Davis, C., Thumann, C., Harris, H. J., Schnober, E. K., Mee, C., Soulier, E., Royer, C. & other authors (2010). Inhibition of hepatitis C virus infection by anti-Claudin-1 antibodies is mediated by neutralization of E2-CD81-Claudin-1 associations. *Hepatology* **51**, 1144-1157.

Kuo, G., Choo, Q. L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J. & other authors (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362-364.

Kwakkenbos, M. J., Diehl, S. A., Yasuda, E., Bakker, A. Q., van Geelen, C. M., Lukens, M. V., van Bleek, G. M., Widjojoatmodjo, M. N., Bogers, W. M. & other authors (2010). Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. *Nat Med* **16**, 123-128.

Kwo, P. Y., Lawitz, E. J., McCone, J., Schiff, E. R., Vierling, J. M., Pound, D., Davis, M. N., Galati, J. S., Gordon, S. C. & other authors (2010). Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): an open-label, randomised, multicentre phase 2 trial. *Lancet* **376**, 705-716.

Laemmli, U. K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lalezari, J., Lawitz, E., Rodriguez-Torres, M., Sheikh, A., Freilich, B., Nelson, D. R., Hassanein, T., Mader, M., Albanis, E. & other authors (2011). Once daily PSI-7977 plus peginterferon/ribavirin in a Phase 2b trial: rapid virologic suppression in treatment-naive patients with HCV genotype 2/3. *J Hepatol* **54**, S28.

Lam, A. M. & Frick, D. N. (2006). Hepatitis C virus subgenomic replicon requires an active NS3 RNA helicase. *J Virol* **80**, 404-411.

Lanini, S., Abbate, I., Puro, V., Soscia, F., Albertoni, F., Battisti, W., Ruta, A., Capobianchi, M. R. & Ippolito, G. (2010). Molecular epidemiology of a hepatitis C virus epidemic in a haemodialysis unit: outbreak investigation and infection outcome. *BMC Inf Dis* **10**, 257-266.

Lanzavecchia, A. (1983). One out of five peripheral blood B lymphocytes is activated to high-rate Ig production by human alloreactive T cell clones. *Eur J Immunol* **13**, 820-824.

Lanzavecchia, A., Bernasconi, N., Traggiai, E., Ruprecht, C. R., Corti, D. & Sallusto, F. (2006). Understanding and making use of human memory B cells. *Immunol Rev* **211**, 303-309.

Lanzavecchia, A., Parodi, B. & Celada, F. (1983). Activation of human B lymphocytes: frequency of antigen-specific B cells triggered by alloreactive or by antigen-specific T cell clones. *Eur J Immunol* **13**, 733-738.

Larralde, O. G., Martinez, R., Camacho, F., Amin, N., Aguilar, A., Talavera, A., Stott, D. I. & Perez, E. M. (2007). Identification of hepatitis A virus mimotopes by phage display, antigenicity and immunogenicity. *J Virol Methods* **140**, 49-58.

Lavillette, D., Bartosch, B., Nourrisson, D., Verney, G., Cosset, F. L., Penin, F. & Pecheur, E. I. (2006). Hepatitis C virus glycoproteins mediate low pH-dependent membrane fusion with liposomes. *J Biol Chem* **281**, 3909-3917.

Lavillette, D., Pecheur, E.-I., Donot, P., Fresquet, J., Molle, J., Corbau, R., Dreux, M., Penin, F. & Cosset, F. L. (2007). Characterisation of fusion determinants points to the involvement of three discrete regions of both E1 and E2 glycoproteins in the membrane fusion process of hepatitis C virus. *J Virol* **81**, 8752-8765.

Lavillette, D., Tarr, A. W., Voisset, C., Donot, P., Bartosch, B., Bain, C., Patel, A. H., Dubuisson, J., Ball, J. K. & other authors (2005). Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus. *Hepatology* **41**, 265-274.

Law, M., Maruyama, T., Lewis, J., Giang, E., Tarr, A. W., Stamataki, Z., Gastaminza, P., Chisari, F. V., Jones, I. M. & other authors (2008). Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. *Nat Med* **14**, 25-27.

Lazaruk, K., Walsh, P. S., Oaks, F., Gilbert, D., Rosenblum, B. B., Menchen, S., Scheibler, D., Wenz, H. M., Holt, C. & other authors (1998). Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* **19**, 86-93.

Lechner, F., Wong, D. K., Dunbar, P. R., Chapman, R., Chung, R. T., Dohrenwend, P., Robbins, G., Phillips, R., Klenerman, P. & other authors (2000). Analysis of successful immune responses in persons infected with hepatitis C virus. *J Exp Med* **191**, 1499-1512.

Lee, S. Y., Choi, J. H. & Xu, Z. (2003). Microbial cell-surface display. *Trends Biotechnol* **21**, 45-52.

Lefranc, M.-P. (2012). IMGT®, the international ImMunoGeneTics information system ®. http://www.imgt.org

Lefranc, M. P. (2011a). From IMGT-ONTOLOGY CLASSIFICATION Axiom to IMGT standardized gene and allele nomenclature: for immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb Protoc* **2011**, 627-632.

Lefranc, M. P. (2011b). From IMGT-ONTOLOGY DESCRIPTION axiom to IMGT standardized labels: for immunoglobulin (IG) and T cell receptor (TR) sequences and structures. *Cold Spring Harb Protoc* **2011**, 614-626.

Lefranc, M. P. (2011c). From IMGT-ONTOLOGY IDENTIFICATION axiom to IMGT standardized keywords: for immunoglobulins (IG), T cell receptors (TR), and conventional genes. *Cold Spring Harb Protoc* **2011**, 604-613.

Lefranc, M. P. (2011d). IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb Protoc* **2011**, 643-651.

Lemon, S. M., Walker, C., Alter, M. J. & Yi, M. (2007). Hepatitis C Virus. In *Field's Virology*, pp. 1253-1304. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott, Williams and Wilkins

Leroux-Roels, G., Depla, E., Hulstaert, F., Tobback, L., Dincq, S., Desmet, J., Desombere, I. & Maertens, G. (2004). A candidate vaccine based on the hepatitis C E1 protein: tolerability and immunogenicity in healthy volunteers. *Vaccine* **22**, 3080-3086.

Leroux-Roels, G., Esquivel, C. A., DeLeys, R., Stuyver, L., Elewaut, A., Phillipe, J., Desombere, I., Paradijs, J. & Maertens, G. (1996). Lymphoproliferative responses to hepatitis C virus core, E1, E2 and NS3 in patients with chronic hepatitis C infection treated with interferon alfa. *Hepatology* **23**, 8-16.

Levy, S. & Shoham, T. (2005). The tetraspanin web modulates immune-signalling complexes. *Nat Rev Immunol* **5**, 136-148.

Li, J., Sai, T., Berger, M., Chao, Q., Davidson, D., Deshmukh, G., Drozdowski, B., Ebel, W., Harley, S. & other authors (2006). Human antibodies for immunotherapy development generated via a human B cell hybridoma technology. *P Natl Acad Sci USA* **103**, 3557-3562.

Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr. & Lemon, S. M. (2005a). Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *P Natl Acad Sci USA* **102**, 2992-2997.

Li, X. D., Sun, L., Seth, R. B., Pineda, G. & Chen, Z. J. (2005b). Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *P Natl Acad Sci USA* **102**, 17717-17722.

Lindenbach, B. D., Meuleman, P., Ploss, A., Vanwolleghem, T., Syder, A. J., McKeating, J. A., Lanford, R. E., Feinstone, S. M., Major, M. E. & other authors (2006). Cell culture-grown hepatitis C virus is infectious *in vivo* and can be recultured *in vitro*. *P Natl Acad Sci USA* **103**, 3805-3809.

Lipke, P. N. & Kurjan, J. (1992). Sexual agglutination in budding Yeasts: Structure, function and regulation of adhesion glycoproteins. *Microbiol Rev* **56**, 180-194.

Liu, H., Bulseco, G. G. & Sun, J. (2006). Effect of posttranslational modifications on the thermal stability of a recombinant monoclonal antibody. *Immunol Lett* **106**, 144-153.

Liu, L., Fisher, B. E., Thomas, D. L., Cox, A. L. & Ray, S. C. (2012). Spontaneous clearance of primary acute hepatitis C virus infection correlated with high initial viral RNA level and rapid HVR1 evolution. *Hepatology* **55**, 1684-1691.

Liu, S., Kuo, W., Yang, W., Liu, W., Gibson, G. A., Dorko, K., Watkins, S. C., Strom, S. C. & Wang, T. (2010). The second extracellular loop dictates Occludin-mediated HCV entry. *Virology* **407**, 160-170.

Liu, S., Yang, W., Shen, L., Turner, J. R., Coyne, C. B. & Wang, T. (2009). Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. *J Virol* **83**, 2011-2014.

Logvinoff, C., Major, M. E., Oldach, D., Heyward, S., Talal, A., Balfe, P., Feinstone, S. M., Alter, H., Rice, C. M. & other authors (2004). Neutralising antibody response during acute and chronic Hepatitis C Virus infection. *P Natl Acad Sci USA* **101**, 10149-10154.

Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L. & Bartenschlager, R. (1999). Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**, 110-113.

Lonberg, N. (2005). Human antibodies from transgenic animals. *Nat Biotechnol* **23**, 1117-1125.

Lonberg, N., Taylor, L. D., Harding, F. A., Trounstine, M., Higgins, K. M., Schramm, S. R., Kuo, C. C., Mashayekh, R., Wymore, K. & other authors (1994). Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**, 856-859.

Lorenz, I. C., Marcotrigiano, J., Dentzer, T. G. & Rice, C. M. (2006). Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature* **442**, 831-835.

Lozach, P.-Y., Amara, A., Bartosch, B., Virelizier, J.-L., Arenzana-Seisedos, F., Cosset, F.-L. & Altermeyer, R. (2004). C-type lectins L-SIGN and DC-SIGN capture and transmit infectious hepatitis C virus pseudotype particles. *J Biol Chem* **279**, 32035-32045.

Luce-Fedrow, A., Von Ohlen, T., Boyle, D., Ganta, R. R. & Chapes, S. K. (2008). Use of *Drosophila* S2 cells as a model for studying *Ehrlichia chaffeensis* infections. *Appl Environ Microb* **74**, 1886-1891.

Lupberger, J., Zeisel, M. B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C. J., Turek, M. & other authors (2011). EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat Med* **17**, 589-595.

Machida, K., Kondo, Y., Huang, J. Y., Chen, Y.-C., Cheng, K. T.-H., Keck, Z., Foung, S., Dubuisson, J., Sung, V. M. H. & other authors (2008). Hepatitis C virus (HCV)-induced immunoglobulin hypermutation reduces the affinity and neutralizing activities of antibodies against HCV envelope protein. *J Virol* **82**, 6711-6720.

Maillard, P., Huby, T., Andreo, U., Moreau, M., Chapman, J. & Budkowska, A. (2006). The interaction of natural hepatitis C virus with human scavenger receptor SR-BI/Cla1 is mediated by ApoB-containing lipoproteins. *FASEB J* **20**, 735-737.

Maillard, P., Lavergne, J.-P., Siberil, S., Faure, G., Roohvand, F., Petres, S., Teillaud, J. L. & Budkowska, A. (2004). Fcγ receptor-like activity of hepatitis C virus core protein. *J Biol Chem* **279**, 2430-2437.

Mancini, N., Diotti, R. A., Perotti, M., Sautto, G., Clementi, N., Nitti, G., Patel, A. H., Ball, J. K., Clementi, M. & other authors (2009). Hepatitis C virus (HCV) infection may elicit neutralizing antibodies targeting epitopes conserved in all viral genotypes. *PLoS One* **4**, 8254-8260.

Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J. & Gomez, J. (1992). Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* **66**, 3225-3229.

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H., Miyata, T. & Honjo, T. (1998). The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med* **188**, 2151-2162.

Matthews, G. V., Pham, S. T., Hellard, M., Grebely, J., Zhang, L., Oon, A., Marks, P., van Beek, I., Rawlinson, W. & other authors (2011). Patterns and characteristics of hepatitis C transmission clusters among HIV-positive and HIV-negative individuals in the Australian Trial in Acute Hepatitis C. *Clin Infect Dis* **52**, 803-811.

Mazumdar, B., Banerjee, A., Meyer, K. & Ray, R. (2011). Hepatitis C virus E1 envelope glycoprotein interacts with apolipoproteins in facilitating entry into hepatocytes. *Hepatology* **54**, 1149-1156.

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554.

McCaffrey, K., Boo, I., Tewierek, K., Edmunds, M. L., Poumbourios, P. & Drummer, H. E. (2012). Role of conserved cysteine residues in hepatitis C virus glycoprotein e2 folding and function. *J Virol* **86**, 3961-3974.

McCaffrey, K., Gouklani, H., Boo, I., Poumbourios, P. & Drummer, H. E. (2011). The variable regions of hepatitis C virus glycoprotein E2 have an essential structural role in glycoprotein assembly and virion infectivity. *J Gen Virol* **92**, 112-121.

McHeyzer-Williams, M., Okitsu, S., Wang, N. X. & McHeyzer-Williams, L. (2012). Molecular programming of B cell memory. *Nat Rev Immunol* **12**, 24-34.

McHutchison, J. G., Everson, G. T., Gordon, S. C., Jacobson, I. M., Sulkowski, M., Kauffman, R., McNair, L., Alam, J. & Muir, A. J. (2009). Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med* **360**, 1827-1838.

McHutchison, J. G., Manns, M. P., Muir, A. J., Terrault, N. A., Jacobson, I. M., Afdhal, N. H., Heathcote, E. J., Zeuzem, S., Reesink, H. W. & other authors (2010). Telaprevir for previously treated chronic HCV infection. *N Engl J Med* **362**, 1292-1303.

McLauchlan, J., Lemberg, M. K., Hope, G. & Martoglio, B. (2002). Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *EMBO J* **21**, 3980-3988.

Mercer, D. F., Schiller, D. E., Elliot, J. F., Douglas, D. N., Hao, C., Rinfret, A., Addison, W. R., Fischer, K. P., Churchill, T. A. & other authors (2001). Hepatitis C virus replication in mice with chimeric human livers. *Nat Med* **7**, 927-933.

Metheringham, R. L., Pudney, V. A., Gunn, B., Towey, M., Spendlove, I. & Durrant, L. G. (2009). Antibodies designed as effective cancer vaccines. *MAbs* **1**, 71-85.

Meuleman, P., Catanese, M. T., Verhoye, L., Desombere, I., Farhoudi, A., Jones, C. T., Sheahan, T., Grzyb, K., Cortese, R. & other authors (2012). A human monoclonal antibody targeting scavenger receptor class B type I precludes hepatitis C virus infection and viral spread in vitro and in vivo. *Hepatology* **55**, 364-372.

Meuleman, P., Hesselgesser, J., Paulson, M., Vanwolleghem, T., Desombere, I., Reiser, H. & Leroux-Roels, G. (2008). Anti-CD81 antibodies can prevent a hepatitis C virus infection *in vivo*. *Hepatology* **48**, 1761-1768.

Meunier, J. C., Engle, R. E., Faulk, K., Zhao, M., Bartosch, B., Alter, H., Emerson, S. U., Cosset, F. L., Purcell, R. H. & other authors (2005). Evidence for cross-genotype neutralization of hepatitis C virus pseudo-particles and enhancement of infectivity by apolipoprotein C1. *P Natl Acad Sci USA* **102**, 4560-4565.

Meunier, J. C., Russell, R. S., Goossens, V., Priem, S., Walter, H., Depla, E., Union, A., Faulk, K. N., Bukh, J. & other authors (2008). Isolation and characterization of broadly neutralizing human monoclonal antibodies to the E1 glycoprotein of hepatitis C virus. *J Virol* **82**, 966-973.

Meyer, K., Banerjee, A., Frey, S. E., Belshe, R. B. & Ray, R. (2011). A weak neutralizing antibody response to hepatitis C virus envelope glycoprotein enhances virus infection. *PLoS One* **6**, e23699.

Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R. & Tschopp, J. (2005). Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167-1172.

Michalak, J. P., Wychowski, C., Choukhi, A., Meunier, J. C., Ung, S., Rice, C. M. & Dubuisson, J. (1997). Characterization of truncated forms of hepatitis C virus glycoproteins. *J Gen Virol* **78**, 2299-2306.

Michta, M. L., Hopcraft, S. E., Narbus, C. M., Kratovac, Z., Israelow, B., Sourisseau, M. & Evans, M. J. (2010). Species-specific regions of occludin required by hepatitis C virus for cell entry. *J Virol* **84**, 11696-11708.

Miyanari, Y., Atsuzawa, K., Usuda, N., Watashi, K., Hishiki, T., Zayas, M., Bartenschlager, R., Wakita, T., Hijikata, M. & other authors (2007). The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol* **9**, 1089-1097.

Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. *P Natl Acad Sci USA* **100**, 6986-6991.

Molina, S., Castet, V., Pichard-Garcia, L., Wychowski, C., Meurs, E., Pascussi, J. M., Sureau, C., Fabre, J. M., Sacunha, A. & other authors (2008). Serum-derived hepatitis C virus infection of primary human hepatocytes is tetraspanin CD81 dependent. *J Virol* **82**, 569-574.

Moradpour, D., Brass, V., Bieck, E., Friebe, P., Gosert, R., Blum, H. E., Bartenschlager, R., Penin, F. & Lohmann, V. (2004). Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. *J Virol* **78**, 13278-13284.

Moreno-Otero, R. (2005). Therapeutic modalities in hepatitis C: challenges and development. *J Viral Hepatitis* **12**, 10-19.

Morrison, S. L., Johnson, M. J., Herzenberg, L. A. & Oi, V. T. (1984). Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *P Natl Acad Sci USA* **81**, 6851-6855.

Mothes, W., Sherer, N. M., Jin, J. & Zhong, P. (2010). Virus cell-to-cell transmission. *J Virol* **84**, 8360-8368.

Mount, S. M. & Rubin, G. M. (1985). Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol Cell Biol* **5**, 1630-1638.

Murphy, D. G., Willems, B., Deschenes, M., Hilzenrat, N., Mousseau, R. & Sabbah, S. (2007). Use of sequence analysis of the NS5B region for routine genotyping of Hepatitis C virus with reference to C/E1 and 5' untranslated region sequences. *J Clin Microbiol* **45**, 1102-1112.

Nakano, T., Lau, G. M., Sugiyama, M. & Mizokami, M. (2012). An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. *Liver Int* **32**, 339-345.

Natterman, J., Schneiders, A. M., Leifeld, L., Langhans, B., Schulz, M., Inchauspe, G., Matz, B., Brackmann, H. H., Houghton, M. & other authors (2005). Serum antibodies against the hepatitis C virus E2 protein mediate antibody-dependent cellular cytotoxicity. *J Hepatol* **42**, 499-504.

Nelson, P. N., Reynolds, G. M., Waldron, E. E., Ward, E., Giannopoulos, K. & Murray, P. G. (2000). Monoclonal antibodies. *Mol Pathol* **53**, 111-117.

Neumann-Haefelin, C., Timm, J., Spangenberg, H. C., Wischniowski, N., Nazarova, N., Kersting, N., Roggendorf, M., Allen, T. M., Blum, H. E. & other authors (2008). Virological and immunological determinants of intrahepatic virus specific CD8+ T cell failure in chronic hepatitis C virus infection. *Hepatology* **47**, 1824-1836.

Neumann, A. U., Lam, N. P., Dahari, H., Gretch, D. R., Wiley, T. E., Layden, T. J. & Perelson, A. S. (1998). Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon-α therapy. *Science* **282**, 103-107.

Nevens, F., Roskams, T., Van Vlierberghe, H., Horsmans, Y., Sprengers, D., Elewaut, A., Desmet, V., Leroux-Roels, G., Quinaux, E. & other authors

(2003). A pilot study of therapeutic vaccination with envelope protein E1 in 35 patients with chronic hepatitis C. *Hepatology* **38**, 1289-1296.

Nybakken, G. E., Nelson, C. A., Chen, B. R., Diamond, M. S. & Fremont, D. H. (2006). Crystal structure of the West Nile Virus envelope glycoprotein. *J Virol*, 11467-11474.

Odegard, V. H. & Schatz, D. G. (2006). Targeting of somatic hypermutation. *Nat Rev Immunol* **6**, 573-583.

Oliphant, T., Engle, M., Nybakken, G. E., Doane, C., Johnson, S., Huang, L., Gorlatov, S., Mehlhop, E., Marri, A. & other authors (2005). Development of a humanised monoclonal antibody with therapeutic potential against West Nile virus. *Nat Med* **11**, 522-530.

Op De Beeck, A., Cocquerel, L. & Dubuisson, J. (2001). Biogenesis of hepatitis C virus envelope glycoproteins. *J Gen Virol* **82**, 2589-2595.

Op De Beeck, A. & Dubuisson, J. (2003). Topology of hepatitis C virus envelope glycoproteins. *Rev Med Virol* **13**, 233-241.

Osburn, W. O., Fisher, B. E., Dowd, K. A., Urban, G., Liu, L., Ray, S. C., Thomas, D. L. & Cox, A. L. (2010). Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection. *Gastroenterology* **138**, 315-324.

Owsianka, A., Clayton, R. F., Loomis-Price, L. D., McKeating, J. & Patel, A. H. (2001). Functional analysis of hepatitis C virus E2 glycoproteins and virus-like particles reveals structural dissimilarities between different forms of E2. *J Gen Virol* **82**, 1877-1883.

Owsianka, A., Tarr, A. W., Juttla, V. S., Lavillette, D., Bartosch, B., Cosset, F. L., Ball, J. K. & Patel, A. H. (2005). Monoclonal antibody AP33 defines a broadly neutralizing epitope on the hepatitis C virus E2 envelope glycoprotein. *J Virol* **79**, 11095-11104.

Owsianka, A., Tarr, A. W., Keck, Z.-Y., Li, T.-K., Witteveldt, J., Adair, R., Foung, S. K. H., Ball, J. K. & Patel, A. H. (2008). Broadly neutralizing human monoclonal antibodies to the hepatitis C virus E2 glycoprotein. *J Gen Virol* **89**, 653-659.

Owsianka, A. M., Timms, J. M., Tarr, A. W., Brown, R. J. P., Hickling, T. P., Szwejk, A., Bienkowska-Szewczyk, K., Thomson, B. J., Patel, A. H. & other authors (2006). Identification of conserved residues in the E2 envelope glycoprotein of the hepatitis C virus that are critical for CD81 binding. *J Virol* **80**, 8695-8704.

Ozawa, T., Kiski, H. & Muraguchi, A. (2006). Amplification and analysis of cDNA generated from a single cell by 5'-RACE: application to isolation of antibody heavy and light chain variable gene sequences from single B cells. *BioTechniques* **40**, 469-478.

Pallares, N., Frippiat, J. P., Giudicelli, V. & Lefranc, M. P. (1998). The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp Clin Immunogenet* **15**, 8-18.

Patrick, W. M., Firth, A. E. & Blackburn, J. M. (2003). User-friendly algorithms for estimating completeness and diversity in randomized protein-coding libraries. *Protein Eng* **16**, 451-457.

Pavlovic, D., Neville, D. C., Argaud, O., Blumberg, B., Dwek, R. A., Fischer, W. B. & Zitzmann, N. (2003). The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *P Natl Acad Sci USA* **100**, 6104-6108.

Pawlotsky, J.-M. (1999). Diagnostic tests for hepatitis C. *J Hepatol* **31**, 71-79.

Penin, F., Brass, V., Appel, N., Ramboarina, S., Montserret, R., Ficheux, D., Blum, H. E., Bartenschlager, R. & Moradpour, D. (2004a). Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J Biol Chem* **279**, 40835-40843.

Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D. & Pawlotsky, J. M. (2004b). Structural biology of hepatitis C virus. *Hepatology* **39**, 5-19.

Perotti, M., Mancini, N., Diotti, R. A., Tarr, A. W., Ball, J. K., Owsianka, A., Adair, R., Patel, A. H., Clementi, M. & other authors (2008). Identification of a broadly cross-reacting and neutralizing human monoclonal antibody directed against the hepatitis C virus E2 protein. *J Virol* **82**, 1047-1052.

Pestka, J. M., Zeisel, M. B., Blaser, E., Schurmann, P., Bartosch, B., Cosset, F.-L., Patel, A. H., Meisel, H., Baumert, J. & other authors (2007). Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. *P Natl Acad Sci USA* **104**, 6025-6030.

Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M. & other authors (2006). Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *P Natl Acad Sci USA* **103**, 7408-7413.

Pietzsch, J., Scheid, J. F., Mouquet, H., Seaman, M. S., Broder, C. C. & Nussenzweig, M. C. (2010). Anti-gp41 antibodies cloned from HIV-infected patients with broadly neutralizing serologic activity. *J Virol* **84**, 5032-5042.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D. & other authors (1998). Binding of hepatitis C virus to CD81. *Science* **282**, 938-941.

Pinna, D., Corti, D., Jarrosay, D., Sallusto, F. & Lanzavecchia, A. (2009). Clonal dissection of the human memory B-cell repertoire following infection and vaccination. *Eur J Immunol* **39**, 1260-1270.

Ploss, A., Evans, M. J., Gaysinskaya, V. A., Panis, M., You, H. N., de Jong, Y. P. & Rice, C. M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457**, 882-886.

Pohlmann, S., Zhang, J., Baribaud, F., Chen, Z., Leslie, G. J., Lin, G., Granelli-Piperno, A., Doms, R. W., Rice, C. M. & other authors (2003). Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J Virol* **77**, 4070-4080.

Powell, J. R. & Moriyama, E. T. (1997). Evolution of codon usage bias in *Drosophila*. *P Natl Acad Sci USA* **94**, 7784-7790.

Poynard, T., Yuen, M. F., Ratziu, V. & Lai, C. L. (2003). Viral hepatitis C. *Lancet* **362**, 2095-2100.

Prentoe, J., Jensen, T. B., Meuleman, P., Serre, S. B. N., Scheel, T. K. H., Leroux-Roels, G., Gottwein, J. M. & Bukh, J. (2011). Hypervariable region 1 differentially impacts viability of hepatitis C virus strains of genotypes 1 to 6 and impairs virus neutralization. *J Virol* **85**, 2224-2234.

Prezzi, C., Nuzzo, M., Meola, A., Delmastro, P., Galfre, G., Cortese, R., Nicosia, A. & Monaci, P. (1996). Selection of antigenic and immunogenic mimics of hepatitis C virus using sera from patients. *J Immunol* **156**, 4504-4513.

Pudney, V. A., Metheringham, R. L., Gunn, B., Spendlove, I., Ramage, J. M. & Durrant, L. G. (2010). DNA vaccination with T-cell epitopes encoded within Ab molecules induces high-avidity anti-tumor CD8+ T cells. *Eur J Immunol* **40**, 899-910.

Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H. & Rambaut, A. (2003). The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: A Bayesian coalescent approach. *Mol Biol Evol* **20**, 381-387.

Pybus, O. G. & Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious diseases. *Nat Rev Genet* **10**, 540-550.

Racanelli, V., Brunetti, C., De Re, V., Caggiari, L., De Zorzi, M., Leone, P., Perosa, F., Vacca, A. & Dammacco, F. (2011). Antibody V(h) repertoire differences between resolving and chronically evolving hepatitis C virus infections. *PLoS One* **6**, e25606.

Ramia, S. & Eid-Fares, J. (2006). Distribution of hepatitis C virus genotypes in the Middle East. *International Journal of Infectious Diseases* **10**, 272-277.

Raney, K. D., Sharma, S. D., Moustafa, I. M. & Cameron, C. E. (2010). Hepatitis C virus non-structural protein 3 (HCV NS3): a multifunctional antiviral target. *J Biol Chem* **285**, 22725-22731.

Ray, S. C., Wang, Y.-M., Laeyendecker, O., Ticehurst, J. R., Villano, S. A. & Thomas, D. L. (1999). Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. *J Virol* **73**, 2938-2946.

Razvi, S., Schneider, L., Jonas, M. M. & Cunningham-Rundles, C. (2001). Outcome of intravenous immunoglobulin-transmitted hepatitis C virus infection in primary immunodeficiency. *Clin Immunol* **101**, 284-288.

Rosa, D., Campagnoli, S., Moretto, C., Guenzi, E., Cousens, L., Chin, M., Dong, C., Weiner, A. J., Lau, J. Y. N. & other authors (1996). A quantitative test to estimate neutralizing antibodies to the hepatitis C virus: Cytofluorimetric assessment of envelope glycoprotein 2 binding to target cells. *P Natl Acad Sci USA* **93**, 1759-1763.

Roy, A., Lu, C. F., Marykwas, D. L., Lipke, P. N. & Kurjan, J. (1991). The *AGA1* product is involved in cell surface attachment of the *Saccharomyces cerevisiae* cell adhesion glycoprotein a-agglutinin. *Mol Cell Biol* **11**, 4196-4206.

Sabo, M. C., Luca, V. C., Prentoe, J., Hopcraft, S. E., Blight, K. J., Yi, M., Lemon, S. M., Ball, J. K., Bukh, J. & other authors (2011). Neutralizing monoclonal antibodies against hepatitis C virus E2 protein bind discontinuous epitopes and inhibit infection at a postattachment step. *J Virol* **85**, 7005-7019.

Sainz, B., Jr., Barretto, N., Martin, D. N., Hiraga, N., Imamura, M., Hussain, S., Marsh, K. A., Yu, X., Chayama, K. & other authors (2012). Identification of the Niemann-Pick C1-like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor. *Nat Med* **18**, 281-285.

Saitou, N. & Nei, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

Sakai, A., Claire, M. S., Faulk, K., Govindarajan, S., Emerson, S. U., Purcell, R. H. & Bukh, J. (2003). The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. *P Natl Acad Sci USA* **100**, 11646-11651.

Sallberg, M., Frelin, L., Diepolder, H., Jung, M.-C., Mathiesen, I., Fons, M., Hultcrantz, R., Carlsson, T. & Weiland, O. (2009). A first clinical trial of therapeutic vaccination using naked DNA delivered by in vivo electroporation shows antiviral effects in patients with chronic hepatitis C. *J Hepatol* **50**, S18.

Sandrin, V., Boulanger, P., Penin, F., Granier, C., Cosset, F. L. & Bartosch, B. (2005). Assembly of functional hepatitis C virus glycoproteins on infectious pseudoparticles occurs intracellularly and requires concomitant incorporation of E1 and E2 glycoproteins. *J Gen Virol* **86**, 3189-3199.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *P Natl Acad Sci USA* **74**, 5463-5467.

Santantonio, T., Wiegand, J. & Gerlach, J. T. (2008). Acute hepatitis C: Current status and remaining challenges. *J Hepatol* **49**, 625-633.

Santolini, E., Pacini, L., Fipaldini, C., Migliaccio, G. & Monica, N. (1995). The NS2 protein of hepatitis C virus is a transmembrane polypeptide. *J Virol* **69**, 7461-7471.

Sarrazin, C., Hezode, C., Zeuzem, S. & Pawlotsky, J. M. (2012). Antiviral strategies in hepatitis C virus infection. *J Hepatol* **56 Suppl**, S88-S100.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. & other authors (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J* **21**, 5017-5025.

Scheid, J. F., Mouquet, H., Feldhahn, N., Seaman, M. S., Velinzon, K., Pietzsch, J., Ott, R. G., Anthony, R. M., Zebroski, H. & other authors (2009). Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636-640.

Scheid, J. F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T. Y., Pietzsch, J., Fenyo, D., Abadir, A. & other authors (2011). Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633-1637.

Schiano, T. D., Charlton, M., Younossi, Z., Galun, E., Pruett, T., Tur-Kaspa, R., Eren, R., Dagan, S., Graham, N. & other authors (2006). Monoclonal antibody HCV-Ab^{XTL}68 in patients undergoing liver transplantation for HCV: Results of a Phase 2 randomized study. *Liver Transpl* **12**, 1381-1389.

Schneider, I. (1972). Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embryol Exp Morphol* **27**, 353-365.

Schroeder, H. W., Jr. & Cavacini, L. (2010). Structure and function of immunoglobulins. *J Allergy Clin Immun* **125**, S41-52.

Schulze Zur Wiesch, J., Ciuffreda, D., Lewis-Ximenez, L., Kasprowicz, V., Nolan, B. E., Streeck, H., Aneja, J., Reyor, L. L., Allen, T. M. & other authors (2012). Broadly directed virus-specific CD4+ T cell responses are primed during acute hepatitis C infection, but rapidly disappear from human blood with viral persistence. *J Exp Med* **209**, 61-75.

Schulze Zur Wiesch, J., Lauer, G. M., Day, C. L., Kim, A. Y., Ouchi, K., Duncan, J. E., Wurcel, A. G., Timm, J., Jones, A. M. & other authors (2005). Broad repertoire of the CD4(+) Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. *J Immunol* **175**, 3603-3613.

Shavinskaya, A., Boulant, S., Penin, F., McLauchlan, J. & Bartenschlager, R. (2007). The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. *J Biol Chem* **282**, 37158-37169.

Shen, Z.-M., Wang, L., Pike, J., Jue, C. K., Zhao, H., de Nobel, H., Kurjan, J. & Lipke, P. N. (2001). Delineation of functional regions within the subunits of the *Saccharomyces cerevisiae* cell adhesion molecule a-agglutinin. *J Biol Chem* **276**, 15768-15775.

Shepard, C. W., Finelli, L. & Alter, M. J. (2005). Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* **5**, 558-567.

Sheridan, I., Pybus, O. G., Holmes, E. C. & Klenerman, P. (2004). High-resolution phylogenetic analysis of Hepatitis C virus adaptation and its relationship to disease progression. *J Virol* **78**, 3447-3454.

Shimizu, Y. K., Igarashi, H., Kiyohara, T., Cabezon, T., Farci, P., Purcell, R. H. & Yoshikura, H. (1996). A hyperimmune serum against a synthetic peptide corresponding to the hypervariable region 1 of hepatitis C virus can prevent viral infection in cell cultures. *Virology* **223**, 409-412.

Shirota, Y., Luo, H., Qin, W., Kaneko, S., Yamashita, T., Kobayashi, K. & Murakami, S. (2002). Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. *J Biol Chem* **277**, 11149-11155.

Shoukry, N. H., Grakoui, A., Houghton, M., Chien, D. Y., Ghrayeb, J., Reimann, K. A. & Walker, C. M. (2003). Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. *J Exp Med* **197**, 1645-1655.

Simmonds, P. (2004). Genetic diversity and evolution of hepatitis C virus - 15 years on. *J Gen Virol* **85**, 3173-3188.

Simmonds, P., Bukh, J., Combet, C., Deleage, G., Enomoto, N., Feinstone, S., Halfon, P., Inchauspe, G., Kuiken, C. & other authors (2005). Consensus

proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**, 962-973.

Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Peng, L. Y. & Simmonds, P. (1997). The origin of hepatitis C virus genotypes. *J Gen Virol* **78**, 321-328.

Smith, G. P. & Petrenko, V. A. (1997). Phage display. *Chem Rev* **97**, 391-410.

Smith, J. A., Aberle, J. H., Fleming, V. M., Ferenci, P., E.C., T., Karayiannis, P., McLean, A. R., Holzmann, H. & Klenerman, P. (2010). Dynamic coinfection with multiple viral subtypes in acute hepatitis C. *J Infect Dis* **202**, 1770-1779.

Spangenberg, H. C., Viazov, S., Kersting, N., Neumann-Haefelin, C., McKinney, D., Roggendorf, M., von Weizsacker, F., Blum, H. E. & Thimme, R. (2005). Intrahepatic CD8+ T-cell failure during chronic hepatitis C virus infection. *Hepatology* **42**, 828-837.

Stapleford, K. A. & Lindenbach, B. D. (2011). Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *J Virol* **85**, 1706-1717.

Staroscik, A. (2004). Calculator for determining the number of copies of a template. *URI Genomics & Sequencing Center* http://www.uri.edu/research/gsc/resources/cndna.html

Steinitz, M., Klein, G., Koskimies, S. & Makel, O. (1977). EB virus-induced B lymphocyte cell lines producing specific antibody. *Nature* **269**, 420-422.

Steinmann, E., Penin, F., Kallis, S., Patel, A. H., Bartenschlager, R. & Pietschmann, T. (2007). Hepatitis C virus p7 protein is crucial for assembly and release of infectious virions. *PLoS Pathog* **3**, e103.

Sui, J., Hwang, W. C., Perez, S., Wei, G., Aird, D., Chen, L. M., Santelli, E., Stec, B., Cadwell, G. & other authors (2009). Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol* **16**, 265-273.

Sukupolvi-Petty, S., Austin, S. K., Engle, M., Brien, J. D., Dowd, K. A., Williams, K. L., Johnson, S., Rico-Hesse, R., Harris, E. & other authors (2010). Structure and function analysis of therapeutic monoclonal antibodies against Dengue virus type 2. *J Virol* **84**, 9227-9239.

Sukupolvi-Petty, S., Austin, S. K., Purtha, W. E., Oliphant, T., Nybakken, G. E., Schlesinger, J. J., Roehrig, J. T., Gromowski, G. D., Barrett, A. D. & other authors (2007). Type- and subcomplex-specific neutralizing antibodies against domain III of Dengue Virus type 2 envelope protein recognize adjacent epitopes. *J Virol* **81**, 12816-12826.

Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrener, P., Grakoui, A., Rice, C. M., Feinstone, S. M. & Collett, M. S. (1993). Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J Virol* **67**, 6152-6158.

Tai, C. L., Chi, W. K., Chen, D. S. & Hwang, L. H. (1996). The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J Virol* **70**, 8477-8484.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599.

Tamura, K., Nei, M. & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbour-joining method. *P Natl Acad Sci USA* **101**, 11030-11035.

Tanaka, Y., Agha, S., Saudy, N., Kurbanov, F., Orito, E., Kato, T., Abo-Zeid, M., Khalaf, M., Miyakawa, Y. & other authors (2004). Exponential spread of Hepatitis C Virus genotype 4a in Egypt. *J Mol Evol* **58**, 191-195.

Tarr, A. W., Owsianka, A. M., Jayaraj, D., Brown, R. J. P., Hickling, T. P., Irving, W. L., Patel, A. H. & Ball, J. K. (2007). Determination of the human

antibody response to the epitope defined by the hepatitis C virus neutralizing monoclonal antibody AP33. *J Gen Virol* **88**, 2991-3001.

Tarr, A. W., Owsianka, A. M., Timms, J. M., McClure, C. P., Brown, R. J. P., Hickling, T. P., Pietschmann, T., Bartenschlager, R., Patel, A. H. & other authors (2006). Characterization of the hepatitis C virus E2 epitope defined by the broadly neutralizing monoclonal antibody AP33. *Hepatology* **43**, 592-601.

Tarr, A. W., Urbanowicz, R. A. & Ball, J. K. (2012a). The role of humoral innate immunity in hepatitis C virus infection. *Viruses* **4**, 1-27.

Tarr, A. W., Urbanowicz, R. A., Hamed, M. R., Albecka, A., McClure, C. P., Brown, R. J. P., Irving, W. L., Dubuisson, J. & Ball, J. K. (2011). Hepatitis C virus (HCV) patient-derived glycoproteins exhibit marked differences in their susceptibility to serum neutralizing antibodies: genetic subtype defines antigenic but not neutralization serotype. *J Virol* **85**, 4246-4257.

Tarr, A. W., Urbanowicz, R. A., Jayaraj, D., Brown, R. J., McKeating, J. A., Irving, W. L. & Ball, J. K. (2012b). Naturally occurring antibodies that recognize linear epitopes in the amino terminus of the hepatitis C virus e2 protein confer noninterfering, additive neutralization. *J Virol* **86**, 2739-2749.

Tellinghuisen, T. L., Evans, M. J., von Hahn, T., You, S. & Rice, C. M. (2007). Studying hepatitis C virus: making the best of a bad virus. *J Virol* **81**, 8853-8867.

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

Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E. & Rice, C. M. (2004). The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem* **279**, 48576-48587.

Thimme, R., Bukh, J., Spangenberg, H. C., Wieland, S., Pemberton, J., Steiger, C., Govindarajan, S., Purcell, R. H. & Chisari, F. V. (2002). Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. *P Natl Acad Sci USA* **99**, 15661-15668.

Thimme, R., Neumann-Haefelin, C., Boettler, T. & Blum, H. E. (2008). Adaptive immune responses to hepatitis C virus: from viral immunobiology to a vaccine. *Biol Chem* **389**, 457-467.

Thomas, D. L., Thio, C. L., Martin, M. P., Qi, Y., Ge, D., O'Huigin, C., Kidd, J., Kidd, K., Khakoo, S. I. & other authors (2009). Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798-801.

Thomson, P. C., Williams, C., Aitken, C., Ball, J. K., Wysocka, N., Brown, R. J. P. & Rodger, R. S. (2011). A case of hepatitis C virus transmission acquired through sharing a haemodialysis machine. *NDT Plus* **4**, 32-35.

Tiller, T., Meffre, E., Yurasov, S., Tsuiji, M., Nussenzweig, M. C. & Wardemann, H. (2008). Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J Immunol Methods* **329**, 112-124.

Timpe, J. M., Stamataki, Z., Jennings, A., Hu, K., Farquhar, M. J., Harris, H. J., Schwarz, A., Desombere, I., Roels, G. L. & other authors (2008). Hepatitis C virus cell-cell transmission in hepatoma cells in the presence of neutralizing antibodies. *Hepatology* **47**, 17-24.

Toellner, L. (2002). Establishment of a baculovirus surface display library using the major envelope protein gp120 from human immunodeficiency virus 1 as a model. *PhD Thesis* Technischen Universität Berlin

Tohme, R. A. & Holmberg, S. D. (2010). Is sexual contact a major mode of hepatitis C virus transmission? *Hepatology* **52**, 1497-1505.

Tomizuka, K., Shinohara, T., Yoshida, H., Uejima, H., Ohguma, A., Tanaka, S., Sato, K., Oshimura, M. & Ishida, I. (2000). Double trans-chromosomic mice: maintenance of two individual human chromosome fragments containing Ig

heavy and kappa loci and expression of fully human antibodies. *P Natl Acad Sci USA* **97**, 722-727.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**, 575-581.

Toran, J. L., Sanchez-Pulido, L., Kremer, L., del Real, G., Valencia, A. & Martinez, A. C. (2001). Improvement in affinity and HIV-1 neutralization by somatic mutation in the heavy chain first complementarity-determining region of antibodies triggered by HIV-1 infection. *Eur J Immunol* **31**, 128-137.

Torresi, J., Johnson, D. & Wedemeyer, H. (2011). Progress in the development of preventive and therapeutic vaccines for hepatitis C virus. *J Hepatol* **54**, 1273-1285.

Traggiai, E., Becker, S., Subbarao, K., Kolesnikova, L., Uematsu, Y., Gismondo, M. R., Murphy, B. R., Rappuoli, R. & Lanzavecchia, A. (2004). An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. *Nat Med* **10**, 871-875.

Tscherne, D. M., Jones, C. T., Evans, M. J., Lindenbach, B. D., McKeating, J. A. & Rice, C. M. (2006). Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry. *J Virol* **80**, 1734-1741.

Urbani, S., Amadei, B., Fisicaro, P., Tola, D., Orlandini, A., Sacchelli, L., Mori, C., Missale, G. & Ferrari, C. (2006). Outcome of acute hepatitis C is related to virus-specific CD4 function and maturation of antiviral memory CD8 responses. *Hepatology* **44**, 126-139.

Vanwolleghem, T., Bukh, J., Meuleman, P., Desombere, I., Meunier, J. C., Alter, H., Purcell, R. H. & Leroux-Roels, G. (2008). Polyclonal immunoglobulins from a chronic hepatitis C virus patient protect human liver-chimeric mice from infection with a homologous hepatitis C virus strain. *Hepatology* **47**, 1846-1855.

Verstrepen, B. E., Depla, E., Rollier, C. S., Mares, G., Drexhage, J. A. R., Priem, S., Verschoor, E. J., Koopman, G., Granier, C. & other authors (2011). Clearance of genotype 1b hepatitis C virus in chimpanzees in the presence of vaccine-induced E1-neutralizing antibodies. *J Infect Dis* **204**, 837-844.

Vicario, S., Moriyama, E. T. & Powell, J. R. (2007). Codon usage in twelve species of *Drosophila*. *BMC Evol Biol* **7**, 226-242.

Vieyres, G., Dubuisson, J. & Patel, A. H. (2011). Characterization of antibody-mediated neutralization directed against the hypervariable region 1 of hepatitis C virus E2 glycoprotein. *J Gen Virol* **92**, 494-506.

Vieyres, G., Thomas, X., Descamps, V., Duverlie, G., Patel, A. H. & Dubuisson, J. (2010). Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. *J Virol* **84**, 10159-10168.

Voisset, C., Callens, N., Blanchard, E., Op De Beeck, A., Dubuisson, J. & Vu-Dac, N. (2005). High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. *J Biol Chem* **280**, 7793-7799.

von Hahn, T., Yoon, J. C., Alter, H., Rice, C. M., Rehermann, B., Balfe, P. & McKeating, J. A. (2007). Hepatitis C virus continuously escapes from neutralizing antibody and T cell responses during chronic infection *in vivo*. *Gastroenterology* **132**, 667-678.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H.-G. & other authors (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**, 791-796.

Walker, L. M., Huber, M., Doores, K. J., Falkowska, E., Pejchal, R., Julien, J. P., Wang, S. K., Ramos, A., Chan-Hui, P. Y. & other authors (2011). Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466-470.

Wang, G. P., Sherrill-Mix, S. A., Chang, K.-M., Quince, C. & Bushman, F. D. (2010). Hepatitis C virus transmission bottlenecks analyzed by Deep Sequencing. *J Virol* **84**, 6218-6228.

Wang, Q. M., Hockman, M. A., Staschke, K., Johnson, R. B., Case, K. A., Lu, J., Parsons, S., Zhang, F., Rathnachalam, R. & other authors (2002). Oligomerization and cooperative RNA synthesis activity of hepatitis C virus RNA-dependent RNA polymerase. *J Virol* **76**, 3865-3872.

Wang, X. & Stollar, B. D. (2000). Human immunoglobulin variable region gene analysis by single cell RT-PCR. *J Immunol Methods* **244**, 217-255.

Wang, Y., Keck, Z. Y., Saha, A., Xia, J., Conrad, F., Lou, J., Eckart, M., Marks, J. D. & Foung, S. K. (2011). Affinity maturation to improve human monoclonal antibody neutralization potency and breadth against hepatitis C virus. *J Biol Chem* **286**, 44218-44233.

Ward, K. N., Ashworth, K. L., E.J., C. & Teo, C. G. (1994). Measurement of antibody avidity for hepatitis C virus distinguished primary antibody responses from passively acquired antibody. *J Med Virol* **43**, 367-372.

Wardemann, H., Yurasov, S., Schaefer, A., Young, J. W., Meffre, E. & Nussenzweig, M. C. (2003). Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374-1377.

Wedemeyer, H., He, X. S., Nascimbeni, M., Davis, A. R., Greenberg, H. B., Hoofnagle, J. H., Liang, T. J., Alter, H. & Rehermann, B. (2002). Impaired effector function of hepatitis C virus-specific CD8(+) T cells in chronic hepatitis C virus infection. *J Immunol* **169**, 3447-3458.

Wedemeyer, H., Mazur, W., Nevens, F., Horsmans, Y., Adler, M., Blum, H., Inglot, M., Gerken, G., Janczewska, E. & other authors (2008). Factors influencing progression of liver fibrosis in patients with chronic hepatitis C: results of the 3-year T2S-918-HCV study with HCVE1 therapeutic vaccine. *J Hepatol* **48**, S27.

Weerapana, E. & Imperiali, B. (2006). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R-101R.

Wei, X., Decker, J. M., Wang, S., Huxiong, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M. & other authors (2003). Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312.

Weiner, A. J., Geysen, H. M., Christopherson, C., Hall, J. E., Mason, T. J., Saracco, G., Bonino, F., Crawford, K., Marion, C. D. & other authors (1992). Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: Potential role in chronic HCV infections. *P Natl Acad Sci USA* **89**, 3468-3472.

Whitcombe, D., Thekaer, J., Guy, S. P., Brown, T. & Little, S. (1999). Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* **17**, 804-807.

Wozniak, A. L., Griffin, S., Rowlands, D., Harris, M., Yi, M., Lemon, S. M. & Weinman, S. A. (2010). Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production. *PLoS Pathog* **6**, e1001087.

Wrammert, J., Smith, K., Miller, J., Langley, W. A., Kokko, K., Larsen, C., Zheng, N. Y., Mays, I., Garman, L. & other authors (2008). Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**, 667-671.

Wu, X., Yang, Z.-Y., Li, Y., Hogerkorp, C.-M., Schief, W. R., Seaman, M. S., Zhou, T., Schmidt, S. D., Wu, L. & other authors (2010). Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856-861.

Wu, X., Zhou, T., Zhu, J., Zhang, B., Georgiev, I., Wang, C., Chen, X., Longo, N. S., Louder, M. & other authors (2011). Focused evolution of HIV-1

neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593-1602.

Wunschmann, S., Medh, J. D., Klinzmann, D., Schmidt, W. N. & Stapleton, J. T. (2000). Characterization of hepatitis C virus (HCV) and HCV E2 interactions with CD81 and the low-density lipoprotein receptor. *J Virol* **74**, 10055-10062.

Xu, X.-G., Wang, Z.-S., Zhang, Q., Li, Z.-C., Zhao, H.-N., Li, W., Tong, D.-W. & Liu, H.-J. (2011). Baculovirus surface display of E envelope glycoprotein of Japanese encephalitis virus and its immunogenicity of the displayed proteins in mouse and swine models. *Vaccine* **29**, 636-643.

Yang, D.-G., Chung, Y.-C., Lai, Y.-K., Lai, C.-W., Liu, H.-J. & Hu, Y.-C. (2007). Avian influenza virus hemagglutinin display on Baculovirus envelope: cytoplasmic domain affects virus properties and vaccine potential. *Mol Ther* **15**, 989-996.

Yu, G. Y., Lee, K. J., Gao, L. & Lai, M. M. (2006). Palmitoylation and polymerization of hepatitis C virus NS4B protein. *J Virol* **80**, 6013-6023.

Yu, M.-Y. W., Bartosch, B., Zhang, P., Guo, Z.-P., Renzi, P. M., Shen, L.-M., Granier, C., Feinstone, S. M., Cosset, F. L. & other authors (2004). Neutralizing antibodies to hepatitis C virus (HCV) in immune globulins derived from anti-HCV-positive plasma. *P Natl Acad Sci USA* **101**, 7705-7710.

Zeisel, M. B., Koutsoudakis, G., Schnober, E. K., Haberstroh, A., Blum, H. E., Cosset, F. L., Wakita, T., Jaeck, D., Doffoel, M. & other authors (2007). 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* **46**, 1722-1731.

Zhang, P., Wu, C. G., K., M., Virata-Theimer, M. L., Yu, M.-Y. W., Alter, H. J. & Feinstone, S. M. (2007). Hepatitis C virus epitope specific neutralizing antibodies in Igs prepared from human plasma. *P Natl Acad Sci USA* **104**, 8449-8454.

Zhang, P., Zhong, L., Budo Struble, E., Watanabe, H., Kachko, A., Mihalik, K., Virata-Theimer, M. L., Alter, H. J., Feinstone, S. & other authors (2009). Depletion of interfering antibodies in chronic hepatitis C patients and vaccinated chimpanzees reveals broad cross-genotype neutralizing activity. *P Natl Acad Sci USA* **106**, 7537-7541.

Zheng, A., Yuan, F., Li, Y., Zhu, F., Hou, P., Li, J., Song, X., Ding, M. & Deng, H. (2007). Claudin-6 and claudin-9 function as additional coreceptors for hepatitis C virus. *J Virol* **81**, 12465-12471.

Zwick, M. B., Bonnycastle, L. L., Menendez, A., Irving, M. B., Barbas, C. F., 3rd, Parren, P. W., Burton, D. R. & Scott, J. K. (2001). Identification and characterization of a peptide that specifically binds the human, broadly neutralizing anti-human immunodeficiency virus type 1 antibody b12. *J Virol* **75**, 6692-6699.