Structural characterization of viral envelope glycoproteins

Ieva Vasiliauskaite

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant

Institut Pasteur, Departement de Virologie, Unité de Virologie Structurale

Structural Characterization of Viral Envelope Glycoproteins

Ieva Vasiliauskaite

Thèse de doctorat de Virologie Structurale

Dirigée par Felix Rey

Présentée et soutenue publiquement le 14 Novembre 2014

Devant un jury composé de:

Rapporteur: Prof. Jean Dubuisson

Rapporteur: Prof. Yves Gaudin

Examinateur: Prof. Germain Trugnan

Examinateur: Prof. Andreas Herrmann

Examinateur: Dr. Berend-Jan Bosch

Directeur de thèse: Prof. Felix Rey

Co-directeur: Dr. Thomas Krey

Acknowledgements

I would like to express my sincere gratitude to a number of people (co-workers, colleagues, friends and family) who helped me and supported me during my doctoral studies.

First I would like to thank my thesis director Prof. Felix Rey for providing me with the opportunity to work in his laboratory, his guidance and advice throughout my thesis.

I also want to express my immense gratitude to my supervisor Dr. Thomas Krey for his extensive supervision, teaching me various techniques, never ending ideas, many stimulating discussions and for open door at any time. Your contribution to my PhD and my development as a researcher has been invaluable. Thank you for all your help on every aspect of the projects and teaching me how to work and think as a scientist.

I would like to thank my thesis reviewers Prof. Yves Gaudin and Prof. Jean Dubuisson for critical reading of the manuscript and all of their helpful suggestions and Prof. Andreas Herrmann, Prof. Germain Trugnan and Dr. Berend Jan Bosch for accepting to serve on my defense committee. Also thanks to my tutor Nicolas Wolff for his support during the past four years.

I am very grateful for all my collaborators for interesting and fruitful interactions: Dr. Annette Martin and Dr. Caroline Marnata (Institut Pasteur) for GBV-B infection inhibition assays, Dr. Patrick England (Plate-Forme de Biophysique des Macromolécules et de leurs Interactions, Institut Pasteur) for introducing me to SPR techniques, Prof. Peter Rottier, Dr. Berend Jan Bosch, Dr. Qiushi Wang (Utrecht University) and the former post-doc of our lab Dr. Scott Jeffers for all collaboration on baculovirus F project, Prof. Arvind Patel and Ania Owsianka (MRC - University of Glasgow Centre for Virus Research) for all contribution to DAO5 project, Steven K. H. Foung for providing a number of antibodies and Christine Girard-Blanc for her initial work on GBV-B E2 and her help with protein expression. Special thank to Dr. Pierre Legrand for his immense help with experimental phasing and all invaluable lessons in protein crystallography.

I am very thankful to all current and former group members for creating a nice working environment and keeping our lab running smoothly as well as sharing their expertise with me. I extend my gratitude to Danielle Senlecques for excellent administrative work and assistance.

I also would like to acknowledge Ahmed Haouz and Patrick Weber from the crystallization platform for help in crystallization and staff of the synchrotron beamlines PX-I

at the Swiss Light Source and Proxima-1 at Synchrotron Soleil for help during data collection.

This work was supported by two grants: Marie Curie Initial Training Network “Virus Entry” and ANRS grant for Thomas Krey. “Virus Entry” network was a very useful training program and also provided me the opportunity to present my work at many different meetings. I want to thank to all the fellows within “Virus Entry” network for all nice moments together.

Finally, I cannot express how much I appreciate emotional and practical support of my closest family and friends. Thanks to my parents and grandparents, my brother and his family for their love and belief in me that always give me so much strength in everything I do. I would like to end with my immense gratitude for Bryan for his support and patience during this stressful period. Thank you so much for helping me through hard times and filling my life with love. No words can express what your love and support have meant to me.

Table of Contents

List of abbreviations

3D	three-dimensional
6HB	six helix bundle
aa	amino acid
Ab	antibody
ALT	alanine aminotransferase
apo	apolipoprotein
ATP	adenosine-5'-triphosphate
BVDV	Bovine viral diarrhea virus
cDNA	complementary DNA
CDR	complementarity-determining region
CLDN	claudin
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
DTT	dithiothreitol
DUF	domain of unknown function
EDTA	ethylenediaminetetraacetic acid, $C_{10}H_{16}N_2O_8$
EGFR	epidermal growth factor receptor
EK	enterokinase
ELISA	enzyme linked immunosorbant assay
EM	electron microscopy
EndoD	Endo-β-N-acetylglucosaminidase D
EndoH	Endo-β-N-acetylglucosaminidase H
EPHA2	ephrin type A receptor 2
ER	endoplasmic reticulum
Fab	Fragment antigen binding
g	acceleration of gravity at the Earth's surface, g = 9.81 m/s^2

GAG	glycosaminoglycan
GBV-B	GB virus B
HC	heavy chain
HCV	hepatitis c virus
HCVcc	HCV cell culture
HCVpp	HCV pseudo-particle
HDL	high-density lipoproteins
HEK	human embryonic kidney (Agnello *et al.*)
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid, $C_8H_{18}N_2O_4S$
HIV	human immunodeficiency virus
HN	hemagglutinin-neuraminidase
hPIV3	human parainfluenza virus 3
HR	heptad repeat region
HRP	horseradish peroxidase
HVR	hypervariable region
Ig	immunoglobulin
igVR	intergenotypic variable region
JFH-1	Japanese fulminant hepatitis-1
kD	kiloDalton, 1 Da = $1.660538921(73)\times10^{-27}$ kg
LB	lysogeny broth
L/DC-SIGN	liver/dendritic cell specific intracellular adhesion molecule 3-grabbing nonintegrins
LC	light chain
LDL	low-density lipoproteins
LDL-R	low-density-lipoprotein receptor
LEL	large extracellular loop
L-SIGN	liver-specific intracellular adhesion molecule-3-grabbing non- integrin
mAb	monoclonal antibody
MAD	multi-wavelength anomalous diffraction
MALDI-TOF_MS	matrix assisted laser desorption ionisation time-of-flight mass

	spectrometry
MALLS	multi-angle laser light scattering
MES	2-(N-morpholino)ethanesulfonic acid, $C_6H_{13}NO_4S$
MLV	murine leukemia virus
MPD	2-Methyl-2,4-pentanediol
nAb	neutralizing antibody
NCS	non-crystallographic symmetry
NDV	Newcastle disease virus
NMR	nuclear magnetic resonance
NPC1L1	Niemann-Pick C1-like 1
NS	non-structural
nt	nucleotide
^oC	degrees Celsius
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PIV5	parainfluenza virus 5
PNGase	peptide-*N*-Glycosidase F
OCLN	occludin
RBV	ribavirin
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RSV	Respiratory syncytial virus
SAD	single-wavelength anomalous diffraction
scFv	single chain antibody fragment
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sE2ΔHVR1	soluble E2 harboring HVR1 deletion
SEC	size-exclusion chromatography

SeFe	Spodoptera exigua multicapsid nucleopolyhedrovirus F protein ectodomain
SeFet	trimer of Spodoptera exigua multicapsid nucleopolyhedrovirus F protein ectodomain
SeMet	selenomethionine
SeMNPV	Spodoptera exigua multicapsid nucleopolyhedrovirus
SFV	Semliki Forest virus
SPR	surface plasmon resonance
SR-BI	scavenger receptor class B type I
S-SAD	sulphur SAD
SV5	simian virus 5
TB	terrific Broth
TBEV	tick-borne encephalitis virus
TfR1	transferrin receptor
TLS	Translation/Libration/Screw
Tris	tris(hydroxymethyl)aminomethane, $(HOCH_2)_3CNH_2$
UTR	untranslated region
VH	Variable heavy
VL	Variable light
VLDL	very-low-density lipoprotein
VLP	virus like particle
VSV	Vesicular stomatitis virus

A	alanine
V	valine
L	leucine
I	isoleucine
F	phenylalanine
M	methionine
T	threonine
K	Lysine
L	leucine
D	aspartate
N	asparagine
E	glutamate
G	glycine
H	histidine
Y	tyrosine
S	serine
C	cysteine
Q	glutamine
W	tryptophan
P	proline

General Introduction

Virus Entry

The role of viral envelope proteins in virus entry

Viruses are small infectious agents that depend on cellular machinery for their own replication. The delivery of the viral genome into the cell is, therefore, an essential step in the virus life cycle. In order to release their genome into the cytoplasm of the host cell, viruses have to go through barriers such as the plasma membrane imposed by the cell. Virus entry is composed of three main steps: (1) attachment, (2) penetration and (3) uncoating. The genome of enveloped viruses is protected by the lipid bilayer membrane surrounding these viruses. Therefore, the penetration step for enveloped viruses requires fusing two lipid bilayers, i.e. the viral envelope and the host cell membrane.

The viral envelope is derived from a host cell membrane during budding, and it contains the viral transmembrane proteins. The viral envelope proteins mediate the steps of attachment and penetration mentioned above and are indispensable for virus infection. The virus envelope usually contains several different proteins, nevertheless, in many viruses the same envelope protein can function both as a receptor-binding protein and a fusion protein (for example influenza hemagglutinin and flavivirus E protein) (Anderson *et al.*, 1992; Chen *et al.*, 1997; Rey *et al.*, 1995; Skehel & Wiley, 2000). However, viruses often use two distinct and specialized envelope proteins that associate in the viral envelope forming envelope spikes (for instance alphavirus E1 and E2, paramyxovirus F and HN) (Kielian *et al.*, 2010) (Chang & Dutch, 2012). The receptor-binding proteins are much more diverse than the fusion proteins which are categorized into several classes based on their structural features (White *et al.*, 2008) (Schneider-Schaulies, 2000).

Virus interaction with cell surface molecules

Virus entry begins when envelope proteins bind to a host cell through specific receptors and/or attachment factors on the cell surface (reviewed in (Grove & Marsh, 2011;

Mercer *et al.*, 2010). Virus binding to attachment factors is a relatively unspecific process and leads to an accumulation of viral particles at the cell surface . Many viruses anchor onto the cells via interactions with heparan sulfate or other carbohydrate structures that contain negatively charged moieties (Jolly & Sattentau, 2013). In contrast, the interaction of the virus with specific receptors results in an active entry process. The latter interactions can be sufficient to trigger the fusion event directly at the cell surface. They may also lead to the internalization of the virus into specific endocytic compartments where exposure to low pH, cleavage by cellular proteases, and/or other cellular factors lead to the release of the genome into the cytoplasm (reviewed in (Mercer *et al.*, 2010; Smith & Helenius, 2004)).

The interaction between the viral envelope proteins and cell surface receptors frequently determines the host tropism and/or the susceptible cell type for virus infection (reviewed in (Schneider-Schaulies, 2000). A virus can interact with one or several receptors/entry factors, a process that may also vary during the course of an infection. In addition, the evolutionary pressure for productive infection and entry may result in the emergence of new virus variants with altered infectivity.

Principles of viral membrane fusion

In order to translocate its genome into the cytoplasm of a cell, an enveloped virus must fuse its membrane to a cellular membrane. The fusion of two lipid bilayers, which is the crucial step of enveloped virus penetration to the target cell, proceeds through several stages (**Figure 1**) (reviewed in (Chernomordik & Kozlov, 2008; Harrison, 2008; White *et al.*, 2008). First, it requires two membranes to approach one another overcoming the repulsive forces generated by the ordered membrane-surface water molecules. Subsequently, a local distortion of the individual bilayers is required to eventually lead to their merging into a single membrane.

These processes have high kinetic barriers (Chernomordik & Kozlov, 2003; 2008). Viral fusion proteins overcome these kinetic barriers by providing the free energy liberated during a conformational change of the protein from a metastable pre-fusion form to a more stable post-fusion form. Independent of the triggering event, all viral fusion proteins undergo structural rearrangements that lead to the exposure of a distinct hydrophobic patch (a fusion peptide or a fusion loop). Insertion of this hydrophobic peptide into a target membrane results in the formation of an extended pre-hairpin intermediate.

The pre-hairpin intermediate of the fusion protein bridges the viral and target membranes by having its C-terminal transmembrane anchor embedded in the viral membrane and the fusion peptide inserted into the target membrane. When the pre-hairpin intermediate collapses into a stable "hairpin" conformation it brings these two membrane inserted elements together resulting in the distortion of the two lipid bilayers (**Figure 1**). This process is followed by lipid mixing of the two proximal leaflets (a step called hemifusion) and the opening of the fusion pore. Pore expansion requires the coordinated action of several fusion proteins at, and outside, the contact sites (Danieli *et al.*, 1996; Roche & Gaudin, 2002). Despite varied oligomeric forms of pre-fusion forms, all post-fusion structures of the viral fusion protein described to date display a trimeric hairpin conformation.

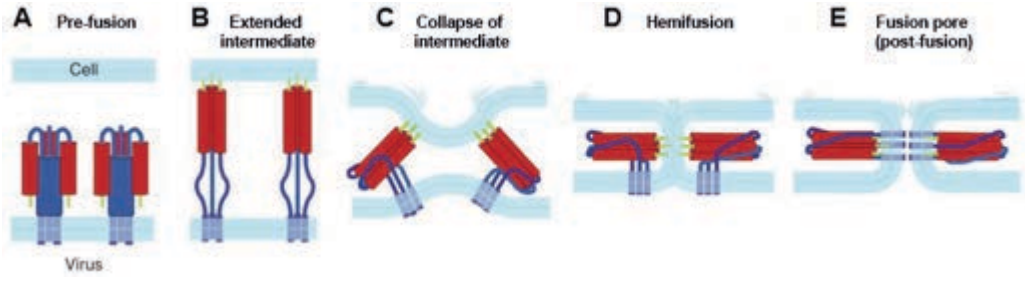


Figure 1. Membrane fusion induced by viral fusion proteins. (A) The viral glycoproteins are present in a viral membrane in a metastable pre-fusion form. (B) Certain triggers such as low pH induce conformational changes in the viral fusion proteins resulting in the insertion of the fusion peptides (light green) into the target membrane and the formation of a trimeric extended intermediate. (C) The extended intermediate collapses when the C-terminal region of the protein (blue) folds back along the N-terminal portion of the protein (red) pulling viral and cellular membranes closer to each other. (D) The proximal leaflets of the opposing membranes start to merge leading to hemifusion. (E) The fusion protein folds into the final hairpin-like, post-fusion conformation with the fusion peptides and transmembrane domains at the same end of the trimer which promotes opening of the fusion pore. Steps C to E may require the concerted action of several trimmers. Adapted from (Harrison, 2008).

Mechanisms to trigger viral membrane fusion

The activity of the fusion proteins must be controlled to ensure that the fusion reaction occurs at the location and time optimal for virus infection. This regulation is achieved by keeping the fusion machinery inactive until a specific switch induces the fusion cascade (reviewed in (White *et al.*, 2008).

During virus biogenesis and maturation the fusion protein adopts the pre-fusion state. Some viral fusion proteins (e.g. the fusion protein of influenza virus) are synthesized as an inactive precursor and must be proteolytically cleaved to generate the metastable pre-fusion form (McCune *et al.*, 1988; Wiley & Skehel, 1987). Other fusion proteins such as those of alphaviruses and flaviviruses are synthesized together with a "chaperone protein" which

assists in folding the fusion protein into the metastable form (Andersson *et al.*, 1997; Lorenz *et al.*, 2002).

The pre-fusion form exists at a local energy minimum and is thus metastable. The energy barrier around the local minimum prevents the pre-fusion form from spontaneous refolding to a more stable post-fusion form and thus releasing the stored energy. The transition between the pre-fusion and post-fusion forms requires a specific trigger that lowers this energy barrier. Despite the common mechanism to merge lipid bilayers used by all viral fusion proteins characterized so far, the specific trigger depends on the individual virus. In general, the fusion mechanisms are classified into pH independent or pH dependent. The triggers described to date include low pH, receptor binding, or a combination of both. Thus, the ligand binding - whether it is a receptor or it is a proton - is a crucial factor for initiating membrane fusion.

Low pH is the main trigger for a number of viruses (e.g. orthomyxo-, alpha-, flavi, bunya-, arenaviruses). In this case, the virus is internalized by one of the endocytic pathways, for example, clathrin-mediated endocytosis, which transports the virus together with their receptors into early and late endosomes. After being endocytosed, the viruses fuse with early or late endosomes depending on the pH optimal for conformational changes of the fusion protein.

Other viruses (e.g. retroviruses or paramyxoviruses) initiate the fusion at neutral pH directly at the cell surface after the fusion protein interactions with the host cell receptor. In this case, the binding of the virus to its receptor activates the fusogenic potential of the virus envelope proteins by inducing certain conformational changes. In the case of HIV-1 membrane fusion is triggered by receptor plus co-receptor binding. HIV envelope proteins gp120 and gp41 are synthesized as a single 160 kD protein precursor which is posttranslationally cleaved into a surface subunit (gp120) mediating receptor binding and a trans-membrane subunit (gp41) mediating membrane fusion. Both proteins remain associated until gp120 binds to CD4 receptor on CD4+ T lymphocytes. This causes structural rearrangements in gp120 and its further interaction with cellular co-receptors, which leads to gp120 dissociation from gp41. The dissociation of gp120 is accompanied by a conformational change in gp41, which exposes the fusion peptide and leads to membrane fusion (reviewed in (Doms & Moore, 2000)). Paramyxo- and herpesviruses, similarly, require receptor binding to a separate attachment/receptor protein, which indirectly activates the fusion protein (reviewed in (Chang & Dutch, 2012; Connolly *et al.*, 2011)).

Some viruses use an intermediate mechanism between the two described above. For example, avian retroviruses employ both receptor binding and low pH to induce the fusion (Mothes *et al.*, 2000). Receptor binding induces conformational changes within the fusion protein, thus allowing it to become sensitive to the acidic pH of endocytic compartments.

For many fusion proteins, the interactions with receptor or with protons trigger a conformational change that may involve changes of the oligomeric state (Allison *et al.*, 1995; Baquero *et al.*, 2013; Stiasny *et al.*, 2002). These rearrangements lead first to the exposure of the fusion peptide and its projection toward the target membrane, and subsequently the folding back of the C-terminal region of the protein onto a trimeric N-terminal region resulting in a final post-fusion conformation with both fusion peptide and transmembrane domain anchored in the merged viral and cellular membranes.

Structures of viral fusion proteins

A detailed characterization of the 3D structures of a number of viral fusion proteins has contributed substantially to our understanding of the molecular mechanisms of specific fusion reactions. Based on their structural and functional characteristics, viral fusion proteins have been classified into three classes (class I, class II and class III) (reviewed in (Backovic & Jardetzky, 2011; Harrison, 2008; Kielian & Rey, 2006; Plemper, 2011; White *et al.*, 2008)). Although the sequences of different viral fusion proteins vary considerably even within the same class, each class shares certain structural features and undergoes similar rearrangements during the fusion of the virus envelope and host cell membrane. Examples of the atomic structures of class I, II and III fusion proteins in their pre-fusion and post-fusion forms are provided in **Figure 2**.

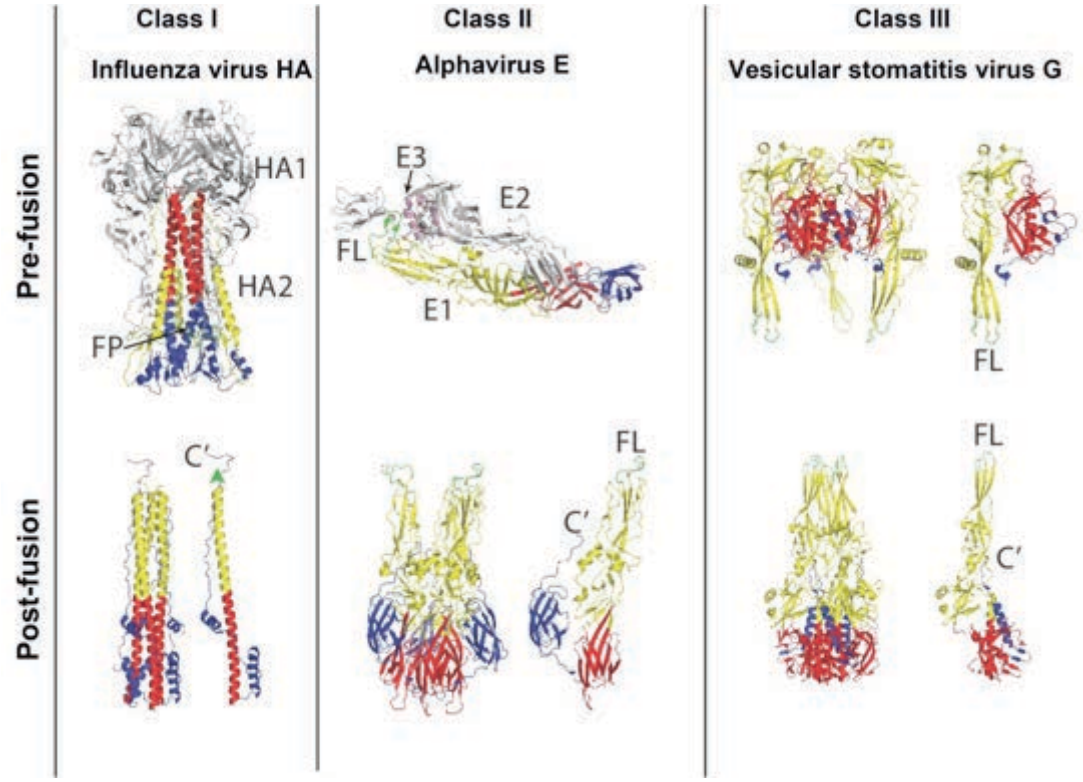


Figure 2. Examples of the atomic structures of class I, II and III fusion proteins in their pre-fusion and post-fusion forms. The domains bearing the fusion peptide/loop (itself colored in green) are colored in yellow. The domains forming the core of the post-fusion trimers are shown in red and blue. The viral membranes would be located at the bottom of the proteins in their pre-fusion forms. For the post-fusion forms both the trimers and protomers composing the trimers are depicted. The merged viral and cellular membranes would be situated at the top of the proteins in their post-fusion forms. Left panel: the influenza hemagglutinin (HA) in its pre-fusion (PDB code 1RUZ) and post-fusion (PDB code 1QU1) forms. HA1 and HA2 are indicated in grey and in colors, respectively. Only HA2 is shown in the post-fusion form. The fusion peptide is not present in the structure, thus, it is specified by a green triangle. Middle panel: the alphavirus fusion protein E1 in its pre-fusion and post-fusion forms. The pre-fusion structure depicts a mature E1-E2-E3 glycoprotein complex of the Chikungunya virus with E1 in colors, E2 in grey, and E3 in pink (PDB code 1RUZ). The post-fusion structure shows the Semliki Forest virus fusion glycoprotein E1 (PDB code 1RER). Right panel: The rhabdovirus vesicular stomatitis G in its pre-fusion (PDB code 2J6J) and post-fusion (PDB code 2CMZ) forms. Both a trimer and a protomer of the pre-fusion form are displayed. Adapted from (Kielian, 2014).

Class I viral fusion proteins are trimers in their pre-fusion and post-fusion forms. They are synthesized as a precursor, which oligomerizes and is then cleaved into two subunits that often remain linked by a disulphide bond. This cleavage is in general mediated by a cellular protease (like furin) in the host secretory pathway (reviewed in (Klenk & Garten, 1994), but it may also occur during entry into a target cells, via endosomal proteases such as chathepsins (like in Ebola virus). Cleavage of the precursor occurs right upstream of the fusion peptide, leaving the fusion protein in a metastable pre-fusion form. The pre-fusion structures of the majority of class I fusion proteins reveal a trimer with a large globular head region and a long α-helical coiled-coil stalk region. External triggers, such as a receptor binding or environmental acidification in the endosomal compartment, induce a conformational change

that results in exposure of a previously hidden fusion peptide (which is in general buried at the trimer interface) and its insertion into the target membrane.

A characteristic features of all class I fusion proteins is a central, parallel coiled coil, identifiable in the sequence by a heptad repeat (HR) motif located downstream of the fusion peptide (in many cases, there is also a second HR motif adjacent to the viral transmembrane region (Chambers *et al.*, 1990)). HRs consists of a pattern of seven amino acids (*abcdefg*) that are repeated in sequence, in which amino acid residues at positions *a* and *d* are typically hydrophobic. This feature of HRs favors a tendency to form α-helical coiled coils (Bruccoleri *et al.*, 1986). The HR segments found in class I fusion proteins play a very important role in the membrane fusion reaction (**Figure 3**). Upon triggering, the N-terminal HR segment first extends towards the target membrane, resulting in the insertion of the fusion peptide into the membrane and formation of a central N-helical coiled-coil (pre-hairpin intermediate). This structural reorganization is followed by the inversion of the C-terminal HR helix (which sometimes can be just an extended segment, like in influenza HA) that packs into the grooves of the central N-terminal coiled-coil and forms a very stable six-helix bundle (6HB) (Lamb & Jardetzky, 2007).

In this conformation, the fusion peptide and the transmembrane domain are brought into the same end of the trimer, which leads to destabilization and merging of viral and cellular membranes. The central coiled coil is a characteristic feature of the post-fusion conformation of all class I fusion proteins. Its length as well as the region where the polypeptide “turns” to make the hairpin varies significantly among class I fusion proteins. The majority of these proteins (except those found in lentiviruses) contain a conserved stutter, which breaks the HR pattern of the coiled-coil in the form of four-residue insertion (“*defg*” introduced between two “*abcdefg*” repeats). Fusion proteins containing coiled-coils can be aligned based on a stutter position allowing a comparison of the relative positions of the membrane interacting elements (i.e. fusion loop and transmembrane domain) with respect to the conserved core of the molecule (Igonet *et al.*, 2011).

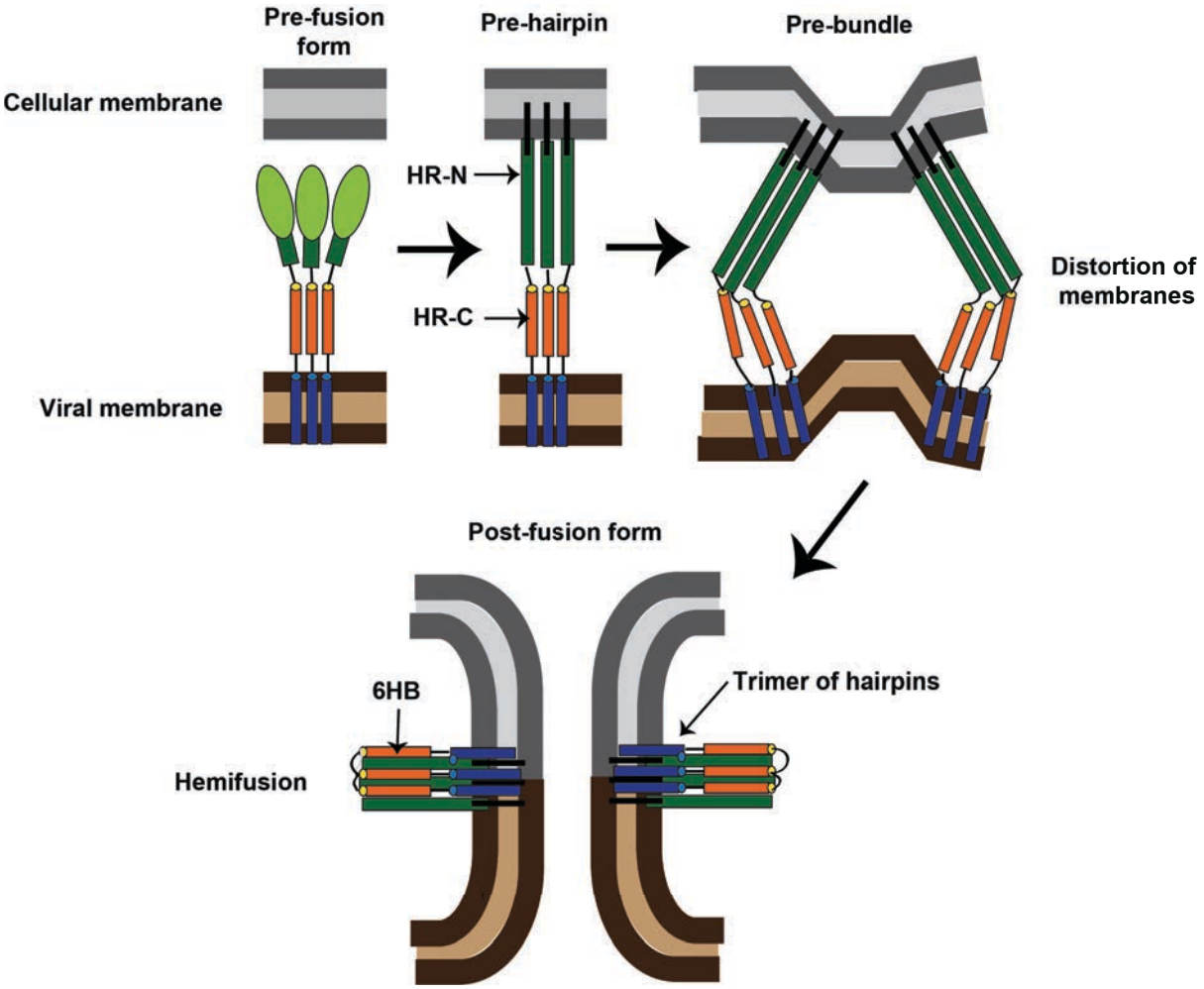


Figure 3. A schematic model for class I fusion protein induced membrane fusion. Class I fusion proteins contain two hydrophobic domains: fusion peptide (colored in black) and transmembrane domain (colored in blue) adjacent to N-terminal and C-terminal HRs (HR-N and HR-C), respectively. In pre-fusion trimer of class I proteins, the fusion peptide is solvent inaccessible. After binding to a receptor or exposure to the low pH, the protein undergoes a conformational change: N-terminal HR (HR-N) (colored in green) extends towards the target membrane resulting in the insertion of the hydrophobic fusion peptide (colored in black) into the host cell membrane. This results in formation of an extended pre-hairpin intermediate containing a central N-helical coiled-coil. Several fusion proteins may mediate fusion process. Subsequently, structural rearrangement of the fusion protein begins during which HR-N and HR-C are gradually brought together and resulting in a bending of the host and viral cell membranes. Several fusion proteins may mediate this process. The prehairpin intermediate resolves to the post-fusion trimer in which HR-N and HR-C are packed in an antiparallel manner forming a six-helix bundle (6HB). Such a conformation brings the fusion peptide into close proximity to the transmembrane domain (hairpin formation), which leads to a close apposition of the cellular and viral membranes and subsequently the actual membrane-fusion.

Comparison of the three-dimensional structures of the central cores of different class I fusion proteins shows that there are two subtypes of these proteins, and it is not clear whether the two have converged from different origins into a central coiled-coil pattern, or whether they diverged from a common ancestor (Igonet *et al.*, 2011). One subtype includes the fusion proteins from paramyxoviruses and coronaviruses which may have evolved from a common ancestor, but not enough structural information is available to confirm this hypothesis, since

the 3D structure of the coronavirus intact spike protein ectodomain in its pre-fusion form is not known. The fusion proteins of retro-, lenti-, filo-, orthomyxo- and arenaviruses appear to have diverged from a common ancestor, and belong to the second subtype of class I fusion proteins, with a short "turn" region of the hairpin.

All class II viral fusion proteins described to date belong to viruses within the families *Flaviviridae*, *Togaviridae* and *Bunyaviridae* (reviewed in (Kielian, 2006; Kielian & Rey, 2006; Modis, 2014). The class II fusion proteins are co-translated with the second N-terminally-located envelope protein (p62 for alphaviruses, prM for flaviviruses and Gn for bunyaviruses), which functions as a chaperone in folding and transport of the fusion protein. Class II fusion proteins form homo- or heterodimers lying parallel to the viral membrane in their pre-fusion conformation, but after exposure to acidic pH in the endosomes they rearrange to form the post-fusion homotrimers.

In contrast to class I fusion proteins in which the main secondary structure element is an α-helix, class II fusion proteins are composed essentially of β-sheets and have a three-domain architecture in both their pre-fusion and post-fusion conformations. The N-terminal domain I is a β-barrel with up-and-down topology. An elongated domain II, which is formed by two segments extending from domain I, consists mostly of the β-strands and includes a fusion loop at its tip, which is buried at the dimer interface. Domain III, positioned at the opposite end of domain I, has an immunoglobulin-superfamily fold. During membrane fusion the overall fold of the domains remain the same but their relative orientation to one another changes. The most significant rearrangement is the foldback of domain III towards the fusion loop positioning the transmembrane domain and the fusion loop at the same end of the molecule.

All viruses from the genus *Flavivirus* within the family *Flaviviridae* carry a class II fusion protein. Based on a putative common signature and the similar genome organization amongst members of this family it was therefore predicted that viruses belonging to other genera within this family (i.e. hepaciviruses and pestiviruses) also encode class II fusion proteins. However, the recently determined structures of hepatitis C E2 and bovine viral diarrhea virus E2 revealed an unexpected fold of these two proteins unrelated to class II viral fusion proteins (El Omari *et al.*, 2013; Khan *et al.*, 2014; Kong *et al.*, 2013; Li *et al.*, 2013). This discovery suggests that E1 is likely to be the fusion effector protein in these viruses, which apparently display a fusion protein belonging to a separate structural class, which has not been fully characterized. Indeed, biochemical data and primary sequence analysis indicates that E1 lacks features characteristic of any the three currently described viral fusion

protein classes. These findings suggest that viruses within the same family have evolved to acquire envelope proteins from different sources.

Class III viral fusion proteins are five-domain molecules composed of both secondary structure elements: α-helices and β-strands (reviewed in (Backovic & Jardetzky, 2011)). The known class III fusion proteins (rhabdovirus G protein (Roche *et al.*, 2006), herpesvirus gB (Backovic *et al.*, 2009; Heldwein *et al.*, 2006), and baculovirus gp64 (Kadlec *et al.*, 2008)) belong to different viral families. Although the three-dimensional organization of these proteins differs significantly from those of class I and class II, some common features with the other classes can be identified. For example, in their post-fusion form they display a central trimeric α-helical coiled-coil similar to those found in class I fusion proteins and which is downstream of a "fusion domain". which is a long β-strand-rich domain reminiscent of the domain II of class II proteins, with two internal fusion loops. A special feature of the best characterized class III fusion protein, VSV G, is that the conformational change induced by low pH is reversible if a target membrane is not available to induce the membrane fusion reaction (Baquero *et al.*, 2013; Gaudin *et al.*, 1991).

Comparison of the viral fusion proteins of different classes is provided in **Table 1**.

Table 1. Comparison of viral fusion proteins of different classes.

Property	**Class I**	**Class II**	**Class III**
Major secondary structure	α-helix	β-sheet	α-helix and β-sheet
Oligomeric state of pre-fusion form	Trimer	Homodimer or heterodimer	Trimer
Requires proteolytic processing for activation	Yes (of fusion protein)	Yes (of companion/chaperone protein) ?	No
Fusion trigger	Low pH, receptor(s) binding, receptor binding plus low pH	Low pH	Low pH or receptor binding
Location of the fusion peptide	Peptide at N-terminus, buried in the trimer interface	Internal loop buried at the dimer interface	Internal loop buried at the trimer interface (except VSV G)
Post-fusion structure	Trimer of hairpins with a central with a central α-helical coiled-coil (six helix bundle)	Trimer of hairpins composed of mainly β-sheets	Trimer of hairpins with a central α-helical coiled-coil and domains composed of β-

			sheets
Type of transmembrane protein	Type I	Type I	Type I
Examples	Influenza HA, respiratory syncytial virus F, HIV gp41	Tick-borne encephalitis virus E, Semliki Forest virus E1, Rift Valley fever virus Gc	Vesicular stomatitis virus G, Herpes simplex virus gB, baculovirus gp64

The role of viral envelope proteins in immune evasion

In addition to their functional role in mediating cell attachment and membrane fusion, viral envelope proteins play an important role in evasion of the host's immune system. Being exposed on the surface of the virion, the envelope proteins are the primary targets of the humoral immune response. Antibodies binding to the epitopes within the domains essential for the viral glycoprotein function (i.e. receptor recognition or induction of the membrane fusion) inhibit viral entry. The presence of such neutralizing antibodies is crucial for viral clearance and a robust immunity against viral infections.

The function of the majority of existing preventive and therapeutic vaccines against diseases caused by enveloped viruses is based on induction of neutralizing antibodies directed against the viral glycoproteins. Many viruses, however, including major human pathogens such as HIV and hepatitis C, have evolved to use elaborate strategies to escape and/or suppress the adaptive and innate immune systems of their hosts. Epitope masking, high mutation rate resulting in altering the antigens within envelope proteins, and cell-to-cell transmission are just a few of many viral envelope protein-associated mechanisms employed by viruses to evade host defenses. As a result, understanding the structure and behavior of viral envelope proteins is crucial to the development of more efficient methods of combatting viral infections.

Chapter I

Hepaciviruses and their entry to target cells

Introduction

Hepatitis C Virus

In the 1970s the increasing concerns of physicians and epidemiologists about the existence of a chronic non-A, non-B hepatitis (Choo *et al.*) transmitted via blood prompted intensive research efforts to identify the causative agent of this new type of hepatitis (Alter *et al.*, 1975a; Alter *et al.*, 1975b; Feinstone *et al.*, 1975; Prince *et al.*, 1974). In 1989 medical researchers identified this pathogen as a new virus related to flaviviruses and pestiviruses, which was subsequently named hepatitis C virus (HCV) (Choo *et al.*, 1989). HCV was classified as the only member of the new genus *Hepacivirus* within the *Flaviviridae* family.

The development of effective diagnostic tests for HCV revealed that the virus is widespread globally. According to the estimations of the World Health Organization ~185 million people are chronically infected with HCV worldwide, and more than 35000 people die from liver diseases caused by the virus every year (Mohd Hanafiah *et al.*, 2013; Thomas, 2013).

HCV is primarily transmitted via blood whereas sexual and vertical transmissions are infrequent (Alter, 1997). Although the introduction for blood screening for HCV in 1990 reduced the number of new infections, 3–4 million people are still infected with HCV every year. The virus spreads mainly due to poor accessibility or quality of diagnostics in certain regions, and to intravenous drug injections in developed countries (Drucker *et al.*, 2001).

HCV pathogenesis

The first six months following an initial HCV infection is referred to as the acute phase, which is asymptomatic in 80% of infected individuals. In the other cases, the acute phase of the disease is associated with jaundice and symptoms similar to those of other mild

infections, such as fever, fatigue, decreased appetite and nausea. Acute HCV infection may resolve itself spontaneously in ~25% of the cases (**Figure 4**). Unfortunately, it progresses to chronic infection in most patients (Maasoumy & Wedemeyer, 2012). Viral clearance is more efficient in people who develop symptoms during the acute phase, whereas persistent HCV infection usually establishes in the case of an asymptomatic acute phase (Gerlach *et al.*, 2003).

HCV infection is marked by an increased concentration of alanine aminotransferase (ALT) up to ten times the normal limit, however, very rarely results in fulminant hepatic failure. In those rare cases, fulminant hepatic failure is likely to be caused by highly virulent strains (Farci *et al.*, 1996a). The only HCV strain (JFH1) capable of replicating in cell cultures was isolated from a patient with such a fulminant hepatitis (Wakita *et al.*, 2005). Production of antibodies (seroconversion) against HCV is usually delayed for up to 3 months. In contrast, viral RNA is detected within 1-2 weeks of HCV exposure, therefore, PCR techniques for detection of viral RNA are usually utilized for diagnosis of the infection.

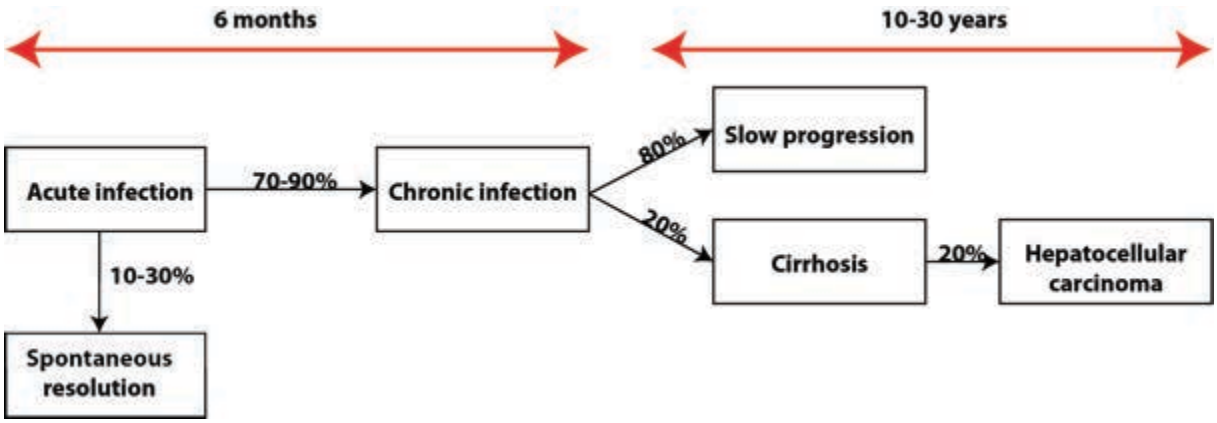


Figure 4. Flow chart of the clinical course of HCV infection.

Chronic HCV infection may remain symptomless for decades and slowly progresses towards chronic liver disease. The outcome of chronic HCV infection depends on a number of determinants such as viral genotype, the person's age, race, gender, and lifestyle. Chronically infected individuals have a high risk of developing liver cirrhosis and hepatocellular carcinoma with a mortality rate of 1-5% without treatment. HCV is estimated to be the underlying cause of liver cancer in 25% of liver cancer patients and the most common cause of liver transplantation in the world (Brown, 2005; Muhlberger *et al.*, 2009; Verna & Brown, 2006). However, the reinfection of the liver graft usually results in damage and loss of the

new liver and the need for re-transplantation. (Hsu *et al.*, 2013). The level of liver enzymes ALT and viremia do not necessarily correlate with the degree of liver injury. During the chronic stage of the disease a lower but relatively stable viral load of ~10^{12} virions per patient per day is commonly detected (Neumann *et al.*, 1998).

Animal models for HCV

HCV has a very narrow host tropism with chimpanzees being the only primates beside humans susceptible to HCV infection. As a result, the chimpanzee has been the primary animal model used to study HCV for more than a decade (Bukh, 2004). Studies in chimpanzees have provided valuable insights about many important aspects of HCV infection such as the course of infection and immune responses to the virus. Nevertheless, differences between HCV infection in chimpanzees and humans exists: exposure of chimpanzees to HCV less frequently results in chronic infection, and in case of chronic infection animals do not normally develop a significant liver disease. Due to ethical issues, the use of these animals is forbidden now, encouraging HCV researchers to search for other animal models.

The tree shrew (*Tupaia belangeri*), a small non-primate mammal, has also been discovered to be susceptible to HCV infection. HCV causes mild hepatitis in these animals with relatively low viral loads and the infection is self-limited in the majority of cases (Xie *et al.*, 1998). Three years after infection, however, liver stearosis, fibrosis and cirrhosis have been observed in some animals (Amako *et al.*, 2010). Though these results are promising, the use of the tree shrews as an HCV model animal is limited by their genetical heterogeneity as outbred species.

A chimeric uPA/SCID (urokinase plasminogen activator/ severe combined immunodeficiency) mouse model harboring human hepatocytes is the most frequently used small animal model for HCV infection (Meuleman & Leroux-Roels, 2008) and has been particularly valuable for testing antiviral compounds. However, the main limitations of this model are that uPA/SCI mice are difficult to generate and have a high mortality rate. In addition, using human hepatocytes within a heterologous organism requires that the mice are immunodeficient in order to prevent rejection of a graft, which does not allow to study the role of the immune system in the pathogenicity of HCV.

During recent years new small animal models overcoming these shortcomings have been developed. Genetically humanized mice expressing human receptors supporting viral

entry were shown to produce new infectious particles allowing investigation of the entire replication cycle of HCV (Dorner *et al.*, 2013). Another model using humanized mice (HIS/HuHEP) possessing a human immune system (HIS) and human hepatocytes (HuHEP), represents a valuable tool to study human immune responses to HCV and to evaluate vaccine efficacy (Yusa *et al.*, 2011). Despite the growing availability of different mouse models no universal small animal model addressing all aspects of HCV infection exists. Therefore, the chimpanzee represents the most relevant model for the study of HCV infection in humans, however, they use in biomedical research is banned.

GBV-B as a surrogate model for HCV

In 1966, in the course of searching a nonhuman primate susceptible for the agents causing human hepatitis, the serum of a surgeon diagnosed with acute hepatitis was used to inoculate tamarins, a small New World monkey (*Saguinus labiatus*). These monkeys developed acute hepatitis, and the serum from the inoculated animals was infectious when passed into a new group of tamarins (Deinhardt *et al.*, 1967). This agent causing hepatitis in tamarins was named GB agent based on the initials of the surgeon, George Barker, from whom the serum was obtained.

Already in the 1970s it was revealed that GB agent is a virus, and might be related to HCV (Almeida *et al.*, 1976; Deinhardt *et al.*, 1975; Parks *et al.*, 1969). However, GB agent was characterized only in 1995 when two flavivirus-like genomes (GB-A and GB-B) were identified in the serum of tamarins inoculated with infectious serum from the 11th passage of the GB agent (Simons *et al.*, 1995). The genomes of these viruses were cloned, and the viruses were named GB virus A (GBV-A), and GB virus B (GBV-B) respectively. Only the GBV-B genome was isolated from the liver of the animals inoculated with the GB agent, thus, it was concluded that GBV-B is the hepatotropic virus causing hepatitis in tamarins.

The amino acid sequence analysis of GBV-B revealed that it is phylogenetically closely related to HCV (28% amino sequence identity across the whole open reading frames) (Muerhoff *et al.*, 1995). It has been concluded that GBV-B was originally a New World primate virus and that the human serum had been inoculated into already infected tamarins. The fact that chimpanzees are not susceptible to GBV-B also argues against its human origin (Bukh *et al.*, 2001; Tabor *et al.*, 1980). However, GBV-B has never been directly isolated from tamarins in the wild, probably due to rapidly resolving infection and the difficulty of

studying colonies of the animals in nature. Subsequent studies showed that GBV-B is able to replicate in some other New World monkeys such as marmosets (*Callitrichidae* family) (Lanford *et al.*, 2003) and owl monkeys (*Cebidae* family) (Bukh *et al.*, 2001), albeit less efficiently than in tamarins. The peak viral titers reach 10^5-10^8 genome equivalents/ml of serum in tamarins but are 10^2-10^3ge/ml lower in marmosets and owl monkeys (Bright *et al.*, 2004; Bukh *et al.*, 2001; Lanford *et al.*, 2003).

The natural course of the infection of GBV-B in susceptible New World monkeys is substantially different from that of HCV in humans. While HCV infection usually leads to chronic hepatitis, GBV-B infection causes an acute hepatitis, which resolves spontaneously within 1-3 months. Occasionally, GBV-B also induces prolonged viremia (more than 6 months), and a disease resembling chronic HCV infection (Jacob *et al.*, 2004; Weatherford *et al.*, 2009). Moreover, in some studies the virus was still detected in the serum of tamarins or marmosets infected by intrahepatic injection of GBV-B RNA transcripts for more than 1-2 years (Iwasaki *et al.*, 2011; Martin *et al.*, 2003; Nam *et al.*, 2004; Takikawa *et al.*, 2006).

The discovery of GBV-B, the closest relative to HCV at the time, raised great interest among researchers because it was believed that studies of GBV-B might help to complement certain less understood/difficult to study aspects of HCV. The fact that GBV-B is closely phylogenetically related to HCV, and is able to occasionally cause chronic hepatitis in non-human primates have led to suggestions that GBV-B infection in small New World monkeys can serve as a surrogate model to study HCV pathogenesis and host-pathogen interactions. This has greatly fostered the molecular studies of GBV-B and characterization of the virus. Since the majority of studies on GBV-B are based on prior knowledge obtained for HCV, data on both virus systems will be discussed in parallel throughout the next sections.

Genome organization of *Hepaciviruses*

Both HCV and GBV-B have been classified in the genus *Hepacivirus* in the *Flaviviridae* family. Besides hepaciviruses, the *Flaviviridae* family includes three other genera: *Flavivirus* (e.g. Dengue virus, Yellow fever virus, West Nile virus, tick-borne encephalitis virus), *Pestivirus* (e.g. Bovine viral diarrhea virus), and *Pegivirus* (e.g. GBV-A). Recently, the *Hepacivirus* genus was expanded to include non-primate hepaciviruses discovered in dogs, horses and rodents (Burbelo *et al.*, 2012; Drexler *et al.*, 2013; Kapoor *et al.*, 2011). Though all members of *Flaviviridae* family share some common features,

particularly in their genome organization and replication strategy, they differ considerably in their virulence and tropism.

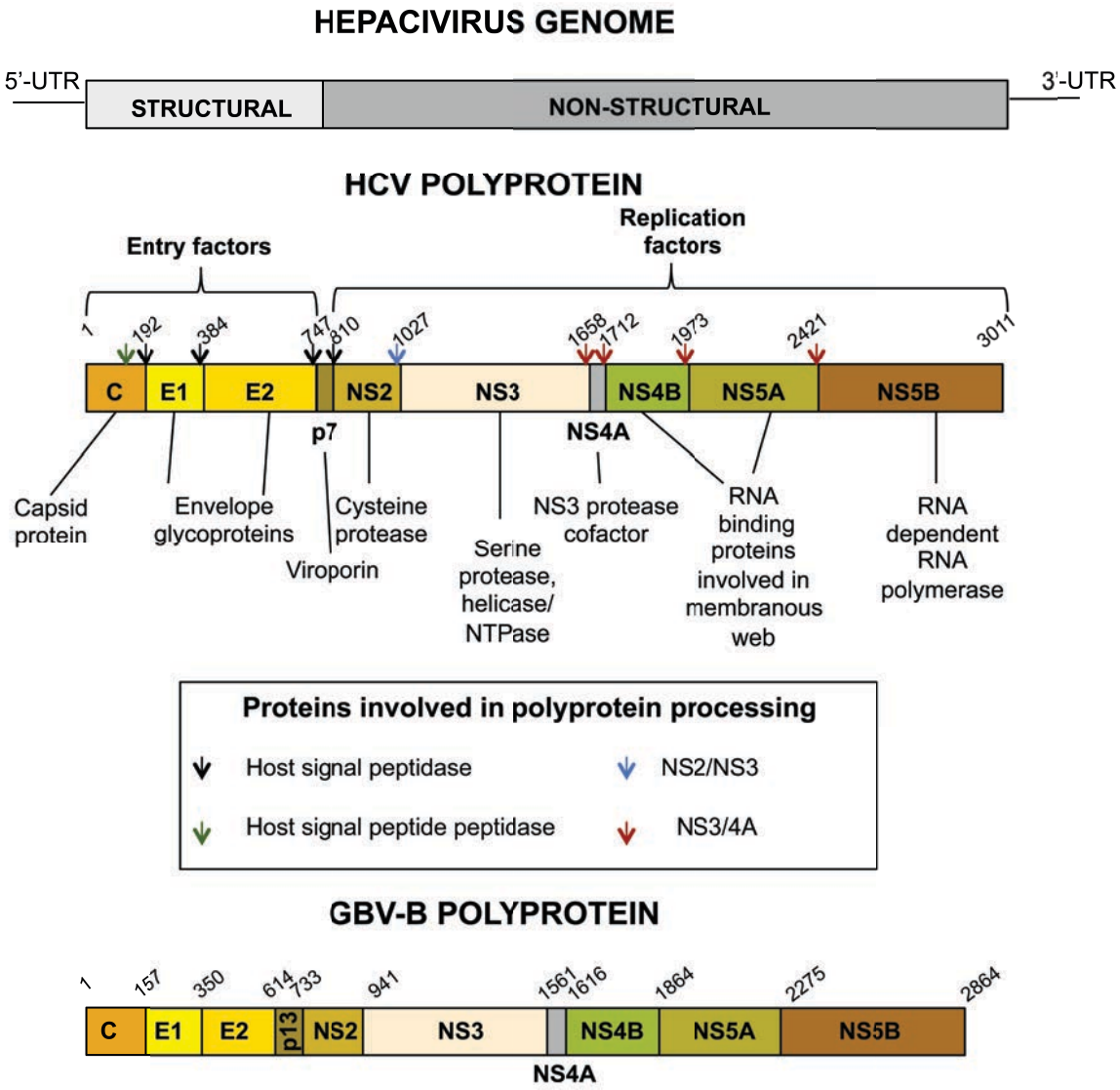


Figure 5. Schematic illustration of hepacivirus genome (A) and HCV and GBV-B polyproteins (B). Cleavage of the polyprotein by host and viral proteases are indicated by arrow heads (black-host signal peptidases, red-NS2/3 protease and blue-NS3/4A protease generates the mature viral proteins: the structural proteins core (C), E1 and E2, and the non-structural proteins p7/p13, NS2-5B. Cleavage of the polyprotein by host and viral proteases and protein functions are indicated only for HCV, however, they also apply to GBV-B. C protein requires an additional cleavage by host signal peptide peptidase to release a mature C protein (green arrow).

Like other members of the family *Flaviviridae*, HCV and GBV-B have an uncapped single stranded positive-sense RNA genome containing a single open reading frame (ORF) that encodes a polyprotein of ~3000 amino acid residues (**Figure 5**). This polyprotein is cleaved co- and post-translationally into structural and non-structural proteins by host and viral proteases (Muerhoff *et al.*, 1995; Simons *et al.*, 1995). The ORFs of both HCV and

GBV-B are flanked by 5' and 3' untranslated regions (UTRs), which form highly structured RNA elements crucial for protein translation and viral replication (Iizuka *et al.*, 1994; Wang *et al.*, 1993). Despite relatively low amino acid identity between the HCV and GBV-B proteins (~28%), the main functions of the corresponding proteins are conserved.

Structural proteins

The structural proteins (i.e. the core (C) protein) and two envelope glycoproteins (E1 and E2) are encoded in the N-terminal moiety of the ORF. As in other members of the *Flaviviridae* family, the envelope glycoproteins are found in tandem within the polyprotein. The structural part of the polyprotein is processed by host signal peptidases liberating the individual proteins, which are then integrated into a virion. The core (C) protein is a RNA-binding protein that forms the nucleocapsid of the virion. E1 and E2 proteins are essential components of the virion envelope and are crucial for viral entry. Since the envelope proteins of HCV and GBV-B are of particular interest in this thesis due to their role in virus entry, they will be described later in a separate section.

In between the structural and non-structural proteins, HCV and GBV-B encodes a small hydrophobic protein p7 and p13, respectively, which has been assigned neither to structural nor to non-structural proteins. p7, encoded by HCV, is a small hydrophobic protein belonging to the viroporin family. After oligomerization p7 forms hexameric or heptameric hydrophilic channels in host cell membranes that modify membrane permeability to ions (Clarke *et al.*, 2006; Griffin *et al.*, 2003; Luik *et al.*, 2009; Montserret *et al.*, 2010). GBV-B also contains a p7 homologue, a small protein p13, which is believed to have a similar function to HCV p7 (Ghibaudo *et al.*, 2004). p13 is processed into two components (p6 and p7) (Takikawa *et al.*, 2006). Interestingly, GBV-B lacking p6 is fully infectious *in vivo.* This finding led to attempts to generate a GBV-B/HCV chimera that contains HCV p7 instead of GBV-B p13. GBV-B/HCV chimeras containing HCV p7 instead of the full sequence of GBV-B p13 or N-terminal p6 portion were shown to be infectious in marmosets; although they replicated to low levels (Griffin, 2008).

Non-structural proteins

The majority of the non-structural proteins of HCV and GBV-B are primarily involved in viral genome amplification by participating in the viral replication complex. NS2 is a cysteine protease that catalyzes the cleavage between NS2 and NS3 (Grakoui *et al.*, 1993; Hijikata *et al.*, 1993a; Kolykhalov *et al.*, 1996; Lohmann *et al.*, 1999). NS3 is a multifunctional protein encoding chymotrypsin-like serine protease in the N-terminal third of the protein and nucleoside triphosphotase (NTPase)/RNA helicase in the remaining two thirds (Kim *et al.*, 1995; Suzich *et al.*, 1993) (Bartenschlager *et al.*, 1993; Hahm *et al.*, 1995; Tai *et al.*, 1996). NS4A serves as a cofactor for the protease activity of NS3 (Satoh *et al.*, 1995). NS3 and NS4A form a non-covalent complex that mediates the processing of the HCV polyprotein to individual mature proteins downstream of NS3 (Miller & Purcell, 1990). NS4B is an integral membrane protein that induces formation of membranous vesicles or invaginations (so called membranous web), where the viral genome replication occurs (Egger *et al.*, 2002; Gosert *et al.*, 2003). NS4B is considered to play a key role in the assembly of other non-structural proteins into replication complexes (Blight, 2011; Butkiewicz *et al.*, 2000). NS5A is a RNA-binding phosphoprotein involved in HCV genome replication and viral particle assembly (Egger *et al.*, 2002; Lohmann *et al.*, 1999). It is associated with the RNA-dependent RNA polymerase NS5B, which is the key enzyme of the HCV replication complex (Behrens *et al.*, 1996; Lohmann *et al.*, 1997).

HCV genotypes

HCV isolated from different patients shows a remarkable genetic diversity, which led to the assignment of different genotypes and subtypes. According to the most recent classification, HCV is currently grouped into 7 main genotypes (1 to 7) and 67 different subtypes (designated a, b, c, etc.) (Smith *et al.*, 2014). Different HCV genotypes are associated with distinct geographical areas, antigenic properties, disease progression, and response to therapy. The different genotypes of HCV can vary 30-35% in their nucleotide sequence (20-25% among the subtypes). The 5'UTR and 3'UTR are the most conserved regions among different genotypes and are used in HCV detection by PCR assays. The viral glycoproteins E1 and E2 show the highest degree of sequence variation. Genotypes 1-3 have the widest geographical distribution with HCV subtypes 1a and 1b being the most common.

The most likely driving force for the origin of different HCV genotypes was the selective pressure of the host immune system (Pang *et al.*, 2009) and/or geographical isolation (Grenfell *et al.*, 2004). Moreover, in infected individuals HCV exists as a mixture of related but genetically distinct variants called quasispecies (Law *et al.*, 2008). Quasisipecies are continuously generated during HCV replication largely as a result of the errors made by viral RNA-dependent RNA polymerase, which lacks proof-reading activity (Bukh *et al.*, 1995).

Current treatment options

The treatment of chronic HCV infection has fundamentally changed during the last few years. Until very recently the traditional therapy for chronic HCV infection was a combination of pegylated interferon-α (PEG-IFN) and the broad-spectrum antiviral agent ribavirin (RBV) (Glue *et al.*, 2000; Manns *et al.*, 2001; McHutchison & Fried, 2003). This therapy, however, was not only poorly tolerated (Russo & Fried, 2003) and therefore applied to only a small percentage of the patients, but also gave a sustained virological response in only ~50% of the treated individuals depending on the HCV genotype (Fried *et al.*, 2002; Hadziyannis *et al.*, 2004; Manns *et al.*, 2001).

Approval of a number of direct-acting antivirals (DAAs) during the last few years has tremendously advanced HCV therapy. NS3/4A serine protease inhibitors biceprevir (Chang *et al.*, 2012) and telaprevir (Forestier & Zeuzem, 2012)) were the first two commercialized DAAs. They have significantly improved the patient cure rates when used in combination with PEG-IFN and RBV (McHutchison *et al.*, 2009). However, the regimens containing biceprevir and telaprevir are efficient only in the treatment of HCV genotype 1 infections. Moreover, they are associated with significant toxicity and a burdensome dosing schedule. Another serious issue is the emergence of drug-resistance viral variants (Aghemo *et al.*, 2013; Ferenci & Reddy, 2011; Wyles, 2012; Zhu & Chen, 2013).

In 2014 this list was supplemented with two new NS3 protease inhibitors (simeprevir and faldaprevir), a non-nucleoside polymerase inhibitor (sofosbuvir) and an NS5a replication complex inhibitor daclatasvir. In the majority of cases, the treatment with the DAAs still includes RBV, however, in most cases it allows the exclusion of PEG-IFN (Muir, 2014). Treatment with a combination of drugs from different classes has shown very promising results. For example, 12-week triple therapy including the nucleotide polymerase inhibitor sofosbuvir, a NS5A inhibitor such as daclatasvir or ledipasvir, and RBV resulted in sustained

virological response rates close to 100% regardless of HCV genotype (Sulkowski *et al.*, 2014) (Afdhal *et al.*, 2014). The development of treatments eliminating RBV is under way with a combination tablet of ledipasvir/sofosbuvir for the treatment of genotype 1 HCV infection being submitted for FDA approval at the beginning of 2014.

The development of new compounds to fight chronic HCV infection is a very rapidly evolving field, with a number of novel DAAs already in late-stage clinical trials. This is expected to result in well-tolerated all-oral regimens in the near future (De Clercq, 2014). However, the selection of resistance-associated variants is a possible threat to the success of these new therapies and has to be carefully evaluated (Poveda *et al.*, 2014).

Although tremendous progress has been made in the treatment of chronic HCV infection since its discovery, the development of efficient prophylactic or/and therapeutic vaccines advanced much slower. Currently, there is no vaccine against HCV, although the need for a safe and efficient prophylactic and/or therapeutic vaccine is evident. The treatment of the infection is not only associated with side effects making it unsuitable for many patients, but is also very expensive (a course of treatment with sofosbuvir can cost from $84,000 to $168,000). Moreover, most people are not aware that they are infected until the late stages of the disease when they experience noticeable complications. To control HCV epidemics globally, a preventive vaccine would be the most efficient means. Development of an effective vaccine requires a thorough understanding of the immune responses against a pathogen. Vaccination against a number of viral infections is based on induction of neutralizing antibodies. However, the extreme genetic diversity of HCV greatly impedes development of a vaccine, which in order to be successful must be effective against different isolates. Therefore, characterization of the mode of HCV neutralization by antibodies and identification of conserved B cell epitopes has been an important HCV research area.

Experimental systems for studying HCV and GBV-B entry *in vitro*

Since the discovery of HCV, the studies of the virus cycle and host-pathogen interactions were hampered by difficulties in propagating the virus *in vitro*. The attempts to grow HCV in primary hepatocytes or established cell lines resulted only in low-level replication, which was not sufficient for investigation of the full viral cycle (Carloni *et al.*, 1993; Fournier *et al.*, 1998; Iacovacci *et al.*, 1993). The A first infectious cDNA clone of HCV was generated in 1997, but it replicated only in chimpanzees (Yanagi *et al.*, 1997).

The first breakthrough in the development of a HCV permissive cell culture system was the establishment of the subgemomic replicon system (Lohmann *et al.*, 1999). The subgenomic replicons could self-replicate in human hepatoma cell cultures and were extensively used to characterize HCV replication. However, the subgenomic replicons lacked the structural protein genes, and thus were not able to produce infectious viral particles. Therefore, other cell culture models were required to obtain better understanding of HCV entry and viral particle assembly.

Studies of HCV entry were greatly facilitated first by the development of retroviral HCV pseudoparticles (HCVpp) and, a few years later, infectious cell-cultured virus (HCVcc), the two most widely used tools to investigate various aspects of the HCV entry pathway. HCVpp are replication-deficient retroviruses carrying non-modified HCV glycoproteins (Bartosch *et al.*, 2003b; Drummer *et al.*, 2003; Hsu *et al.*, 2003). They are produced in human embryonic kidney (HEK) 293T cells by co-transfection of vectors encoding the gag-pol proteins of human immunodeficiency virus or murine leukemia virus, the HCV glycoproteins, and a packaging-competent retroviral genome harboring a reporter gene such as luciferase. The production of HCVpp is relatively easy and it is possible to incorporate patient-derived glycoproteins, facilitating the analysis of cross-neutralizing antibodies (Bartosch *et al.*, 2003b; Tarr *et al.*, 2007b). Moreover, the use of HCVpp allows the study of HCV entry into cells that are not able to support HCV replication (Evans *et al.*, 2007; Ploss *et al.*, 2009). HCVpp have been used for the identification of a number of cell receptors and attachment factors used by the virus. However, this system fails to mimic the association of infectious virus particles with lipoproteins, and the retroviral particles assemble in post-Golgi compartments or/and at the plasma membrane, which causes changes to the glycosylation pattern and oligomerization of the HCV glycoproteins compared with infectious virus particles. These features may have an impact on studies concentrating on antibody neutralization and the role of HCV receptors involved in lipid metabolism.

An important milestone in HCV research was the development of HCVcc, representing the first tissue culture system that allowed the completion of the viral cycle (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005; Zhong *et al.*, 2005). HCVcc are based on an infectious clone derived from the Japanese Fulminant Hepatitis-1 (JFH-1) strain of genotype 2a. This system allows studies of all steps of the HCV cycle, from viral entry to virion assembly and release. HCVcc have been shown to be infectious in cell cultures (mostly human hepatoma Huh-7 cells) and *in vivo* (in chimpanzees and transgenic mice with human liver xenografts) (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005). In recent years, further

improvements of the HCVcc system have been made. First, chimeric genomes harboring the structural genes of other HCV genotypes were constructed and were shown to produce infectious virions (Gottwein *et al.*, 2009; Jensen *et al.*, 2008; Pietschmann *et al.*, 2006; Scheel *et al.*, 2008; Yi *et al.*, 2007). Second, the incorporation of reporter genes resulted in an easier and more quantitative way to quantify the number of infectious particles (Gottwein *et al.*, 2011; Gottwein *et al.*, 2007; Koutsoudakis *et al.*, 2006; Lindenbach *et al.*, 2005). Hepatocytes in the liver are polarized, i.e. their basal surface faces the circulation while adjacent hepatocytes and hepatocytes face bile canaliculi form the basal surface (Decaens *et al.*, 2008). Since HCV engages receptors at the basal surface of hepatocytes, polarization limits HCV entry probably by restricting mobility of CD81 receptor (Harris *et al.*, 2013; Mee *et al.*, 2009). One of the drawbacks of HCVcc is that human hepatoma cell lines like Huh-7 have no or poor polarization, though some advances were made to overcome this problem (Decaens *et al.*, 2008; Kambara *et al.*, 2012; Narbus *et al.*, 2011). Another issue is that HCVcc particles differ in lipoprotein composition from the viral particles found in the serum of HCV infected patients (Icard *et al.*, 2009; Nielsen *et al.*, 2006).

Another important tool for studying HCV entry, and which has been extensively used for describing HCV interaction with receptors and neutralization by antibodies, consists in the use of different versions of the soluble ectodomain of the glycoprotein E2 expressed in mammalian or insect cells (Michalak *et al.*, 1997; Pileri *et al.*, 1998; Scarselli *et al.*, 2002; Spaete *et al.*, 1992) (Krey *et al.*, 2010).

Since each of the available systems for studying HCV entry have different shortcomings, only the use of a combination of approaches has allowed researchers to uncover important components of the virus entry process. Nevertheless, the development of new host/virus systems or improvement of the existing ones remains essential for further investigation of HCV entry.

The first infectious clone of GBV-B was reported in 1999, and corresponded to the full-length GBV-B genome (Bukh *et al.*, 1999). The intrahepatic inoculation of tamarins with RNA transcripts of this clone resulted in high viral titers of infectious virus in the serum of the animals. GBV-B has also been reported to infect and replicate in primary cultures of hepatocytes from tamarins or from marmosets (Beames *et al.*, 2000; Bright *et al.*, 2004). However, the investigation of different aspects of the GBV-B life cycle in primary cultures is difficult due to variability between preparations, and requirements for special conditions and handling.

Infectious cycle of HCV and GBV-B

The infectious cycle of HCV and GBV-B includes a number of steps, with the major events being: 1) cell attachment and entry into the host cell, 2) translation of viral proteins and replication of viral RNA, and 3) assembly of new viral particles and their release from the host cell (**Figure 6**).

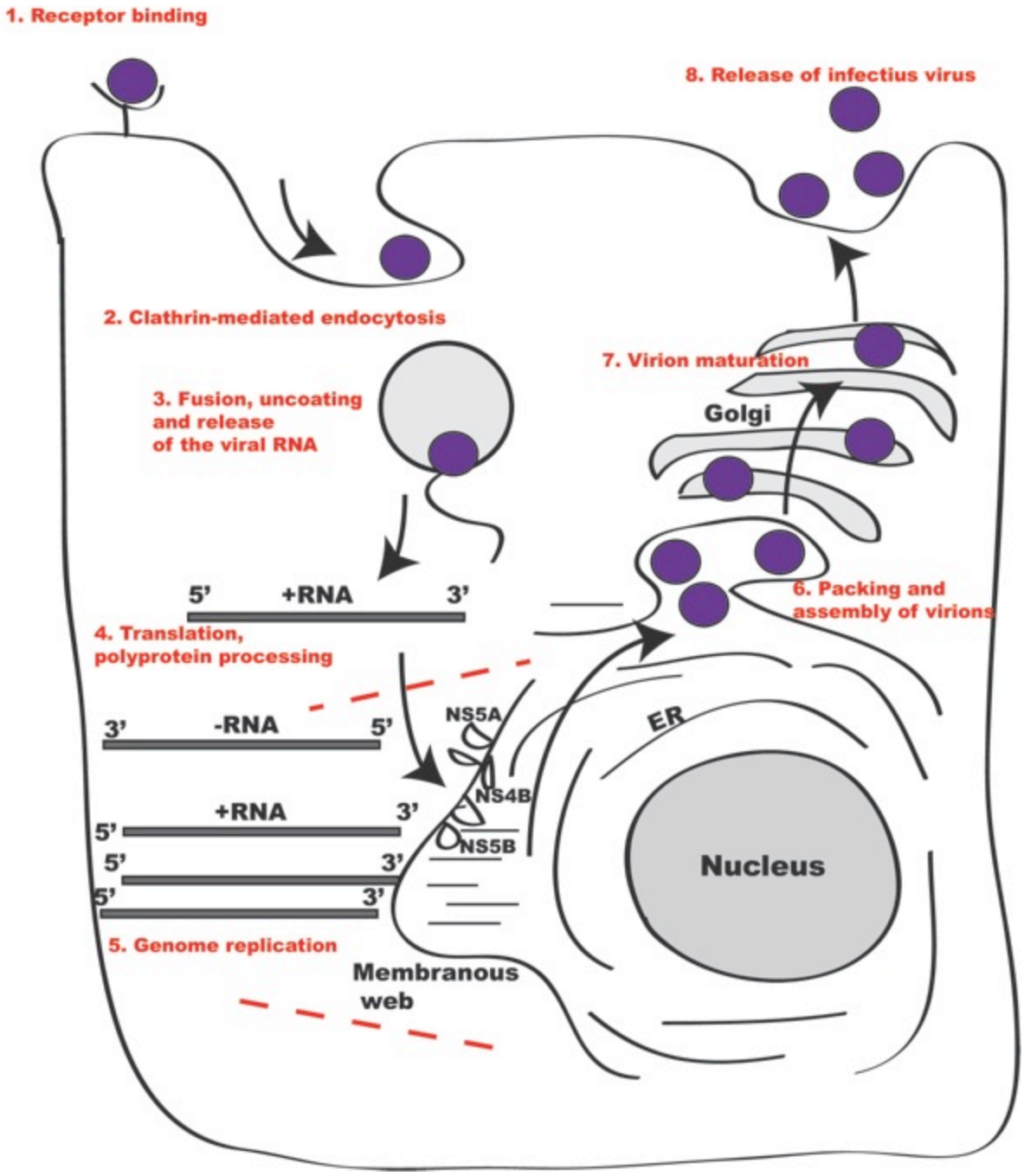


Figure 6. HCV replication cycle (simplified representation). The cycle starts by virus entry into the cell, which involves the interaction with a number of receptors and occurs by clathrin-mediated endocytosis. In early endosomes, low pH induces the fusion of viral and endosomal membranes to release viral genome. This step is mediated by the viral glycoproteins. +RNA genome is subsequently translated into a single precursor

polyprotein, which is processed into the individual mature proteins. Replication of the genome occurs in a membranous web. Assembly of progeny virions takes place at the ER membrane. Subsequently, maturation of the viral particles occurs when they travel through the secretory pathway before being released from the cell.

HCV Entry

HCV cell entry is a multi-step process that can be subdivided into three main events: 1) virus attachment to the cell, 2) clathrin-mediated endocytosis of the viral particle, and 3) fusion of viral and cellular membranes. The key players in HCV cell entry are the envelope glycoproteins E1 and E2. Viruses that use clathrin-mediated endocytosis are typically internalized after interaction with cellular receptors. One of the hallmarks of HCV internalization is that it requires a large number of viral and cellular factors to initiate endocytosis.

Due to their importance in host-cell interaction, viral factors such as the envelope glycoproteins or the composition of the HCV particle will be discussed first followed by an introduction of the cellular factors that are engaged in HCV entry.

HCV particle

The composition of the HCV particle plays a major role in virus-host interactions and is a key element of the sophisticated entry mechanism of this virus into the target cell. The HCV particle contains both viral and host-derived components, which results in a complex composition and morphology of the particle.

A unique feature of HCV is that the virion circulates in the bloodstream of the infected individuals in complex with host lipoproteins. Patients' sera contain distinct viral populations. A very low-density fraction is associated with very-low density lipoproteins (VLDL) and low density-lipoproteins (LDL), which results in a low buoyant density (<1.10 g/mL) of the particles, which is atypical for an enveloped RNA virus. An intermediate fraction (1.10-1.21 g/ml) harbors high-density lipoproteins (HDL). Naked capsids and virions associated with antibodies constitutes the densest viral population (>1.21 g/ml) (Hijikata *et al.*, 1993b; Kanto *et al.*, 1995; Merz *et al.*, 2011; Prince *et al.*, 1996; Thomssen *et al.*, 1992; Thomssen *et al.*, 1993).

Interestingly, the infectivity of HCV lipoviral particles is inversely correlated with their density, with the light-density fractions exhibiting a higher specific infectivity *in vivo*

and in cell culture than the high-density fractions (Andre *et al.*, 2002). A number of lipoproteins such as apolipoprotein A-I (apoA-I), apoE, apoB and apoC-I has been identified to be incorporated into HCV lipoviral particles (Andre *et al.*, 2002; Chang *et al.*, 2007; Diaz *et al.*, 2006; Meunier *et al.*, 2008a; Thomssen *et al.*, 1992). However, the lipoprotein composition differs depending on the host cell and the system in which the particles were produced (Bartenschlager *et al.*, 2011; Lindenbach *et al.*, 2006; Podevin *et al.*, 2010).

Lipoproteins were proposed to have several roles in HCV entry and contribute to the complexity of this process. Firstly, they mask envelope proteins from the host's immune system (Bankwitz *et al.*, 2010; Dreux *et al.*, 2006; Grove *et al.*, 2008; Prentoe *et al.*, 2011). Secondly, lipoproteins act as host-derived ligands on the surface of the particles to facilitate interactions with the target cells (Mazumdar *et al.*, 2011). Finally, they were also reported to bind to cellular receptors/attachment factors (e.g., glycosaminoglycans, LDL receptor and SR-BI) that will be described later in more detail.

Due to difficulties producing and purifying the viral particles, the architecture of HCV particles still remains elusive. Recent cryo-EM and cryo-ET studies indicate that HCV particles are spherical, heterogeneous in size (40-100 nm in diameter), and contain spike-like projections. It seems that apoE is better exposed on the particle surface than E2 (Catanese *et al.*, 2013). The arrangement of viral glycoproteins on HCV particles is not defined yet, however, it seems to be different from the well ordered and symmetrical glycoprotein shell typically observed for viruses belonging to the genus Flavivirus within the family *Flaviviridae*.

HCV envelope glycoproteins E1 and E2

E1 and E2 are type I trans-membrane proteins composed of an N-terminal ectodomain of ~160 and 334 amino acids respectively, and a short C-terminal membrane spanning domain of ~30 amino acids. It is believed that E1 and E2 function as non-covalent heterodimers (Deleersnyder *et al.*, 1997; Dubuisson *et al.*, 1994; Op De Beeck *et al.*, 2004). However, more recent studies reported that virion-associated E1 and E2 form high molecular weight disulfide-linked complexes, whereas the intracellular E1 and E2 exist as non-covalent heterodimers (Vieyres *et al.*, 2010). The two glycoproteins heterodimerize in the ER even if they are expressed in *trans* (Ralston et al., 1993). Separately, E1 and E2 are not sufficient for HCVpp infectivity (Bartosch *et al.*, 2003b).

The E1 and E2 ectodomains are heavily glycosylated and possess 4-5 and 11 N-linked glycans, respectively (Goffard *et al.*, 2005). 4 glycosylation sites of E1 and 9 in E2 are highly conserved, indicating that they are required in the virus life cycle (Helle *et al.*, 2007). The glycans have been reported to play a role in proper folding of E1 and E2, and HCV entry (Falkowska *et al.*, 2007; Goffard *et al.*, 2005; Helle *et al.*, 2007; Helle *et al.*, 2010).

The transmembrane domains of E1 and E2 not only function as the membrane anchors but are also important for ER retention (Cocquerel *et al.*, 1999; Cocquerel *et al.*, 1998) and heterodimerization of E1 and E2 (Ciczora *et al.*, 2007; Cocquerel *et al.*, 2002; Op De Beeck *et al.*, 2000).

The E2 glycoprotein has gained a lot of attention in HCV research for several reasons: 1) It directly interacts with a number of host cell membrane proteins, leading to virus entry, 2) Most HCV neutralizing antibodies are directed against E2, and 3) A recombinant soluble E2 lacking the trans-membrane domain adopts a native fold and inhibits HCVcc infection.

The E2 glycoprotein is composed of a receptor binding domain (aa residues 384–661) connected to the transmembrane domain via a stem region (aa residues 716–746). The stem region contains a heptad repeat (residues 675–699) segment that has been shown to be important for E1E2 heterodimerization and is supposed to be involved in the reorganization of the E1E2 complex during the fusion process (Albecka *et al.*, 2011; Drummer & Poumbourios, 2004; Perez-Berna *et al.*, 2006).

HCV E2 exhibits high genetic variability. The most distant HCV genotypes can differ in their E2 amino acid sequence by up to 20%. The ectodomain of E2 has been reported to contain four hypervariable regions (HVR1-3) and the inter-genotype variable region (igVR) (Hijikata *et al.*, 1991; Troesch *et al.*, 2006; Weiner *et al.*, 1991). HVR1 is a 27 amino acid region (aa residues 384–410 of the viral polyprotein) at the N-terminus of E2 that has been demonstrated to play an important role in virus entry, antibody neutralization, and disease outcome (Bartosch *et al.*, 2003c; Farci *et al.*, 2000; von Hahn *et al.*, 2007). The other two HVRs, HVR2 and HVR3, span aa residues 474–482 and 431–466, respectively. HVR3 has been reported to overlap with the antigenic region (epitope II) targeted by some broadly neutralizing antibodies (Lapierre *et al.*, 2011). The fourth hypervariable region, the igVR, spans aa residues 570–580) (McCaffrey *et al.*, 2007). HVR2 and igVR have been suggested to be important for E1-E2 heterodimerization and infectivity (McCaffrey *et al.*, 2011). Three regions, HVR1, HVR2 and igVR, can be deleted in a recombinant soluble E2 without affecting the native overall fold of the glycoprotein core (McCaffrey *et al.*, 2007), deletion of

all three regions result in the loss of infectivity of the HCVcc, HCVpp or *in vivo* produced virus (Bankwitz *et al.*, 2010; Forns *et al.*, 2000; McCaffrey *et al.*, 2011).

Recently, the structure of the core ectodomain of HCV E2 was independently reported by two different laboratories (Khan *et al.*, 2014; Kong *et al.*, 2013). Kong and colleagues crystallized E2 spanning aa residues 412-645 (according to the numbering based on strain H77 (genotype 1a) polyprotein) in complex with a Fab derived from the human broadly neutralizing antibody AR3C (PDB ID 4MWF) (Kong *et al.*, 2013). To facilitate the crystallization, the protein was truncated at the N and C termini (resulting in removal of HVR1, stem region, and transmembrane domain) and had the HVR2 region substituted by a short linker. This modified version of E2 was designated HCV E2 core (E2c).

Khan *et al*. determined the crystal structure of a slightly different HCV E2c in complex with a non-neutralizing murine Fab 2A12 (PDB ID 4NX3) (Khan *et al.*, 2014). E2c crystallized by Khan *et al.* is shorter at N-terminus and encompasses aa residues 456-656 (J6 strain, genotype 2a). Its architecture is identical to the crystal structure reported by Kong and co-workers with a root mean square deviation (RMSD) of 0.8 Å between the corresponding α carbon atoms of the two structures.

Overall, the E2c structure shows of a central Ig-fold β-sandwich domain (residues 492 to 566), which is covered by a layer of loops, short helices, and β-sheets on the front and the back faces (**Figure 7**). The E2c β-sandwich is composed of 6 strands, 4 strands forming an inner sheet and two strands forming a solvent exposed outer sheet. E2c contains many regions which are disordered or do not possess regular secondary structure. E2c crystallized by Kong *et al*. lacks high resolution data for the region spanning aa 586-596 and the regions spanning aa 456-491 (HVR2 and HVR3), 523-538 and 572-595 (igVR) are absent in the E2 core structure determined by Khan *et al*.

To obtain insight into the full-length E2 structure, negative stain electron microscopy (EM) of the complete E2 ectodomain bound to a Fab AR2A was performed (Kong *et al.*, 2013). The EM reconstruction indicated that the full-length E2 ectodomain has a globular compact shape. Similarly, Khan and colleagues performed solution based small-angle X-ray scattering (SAXS) to correlate the dimension of the crystallized E2c with the dimensions of glycosylated full-length E2 ectodomain. The similar gyration radius of the *ab initio* SAXS envelopes of both proteins also suggests that the full-length ectodomain possess a globular fold (Khan *et al.*, 2014).

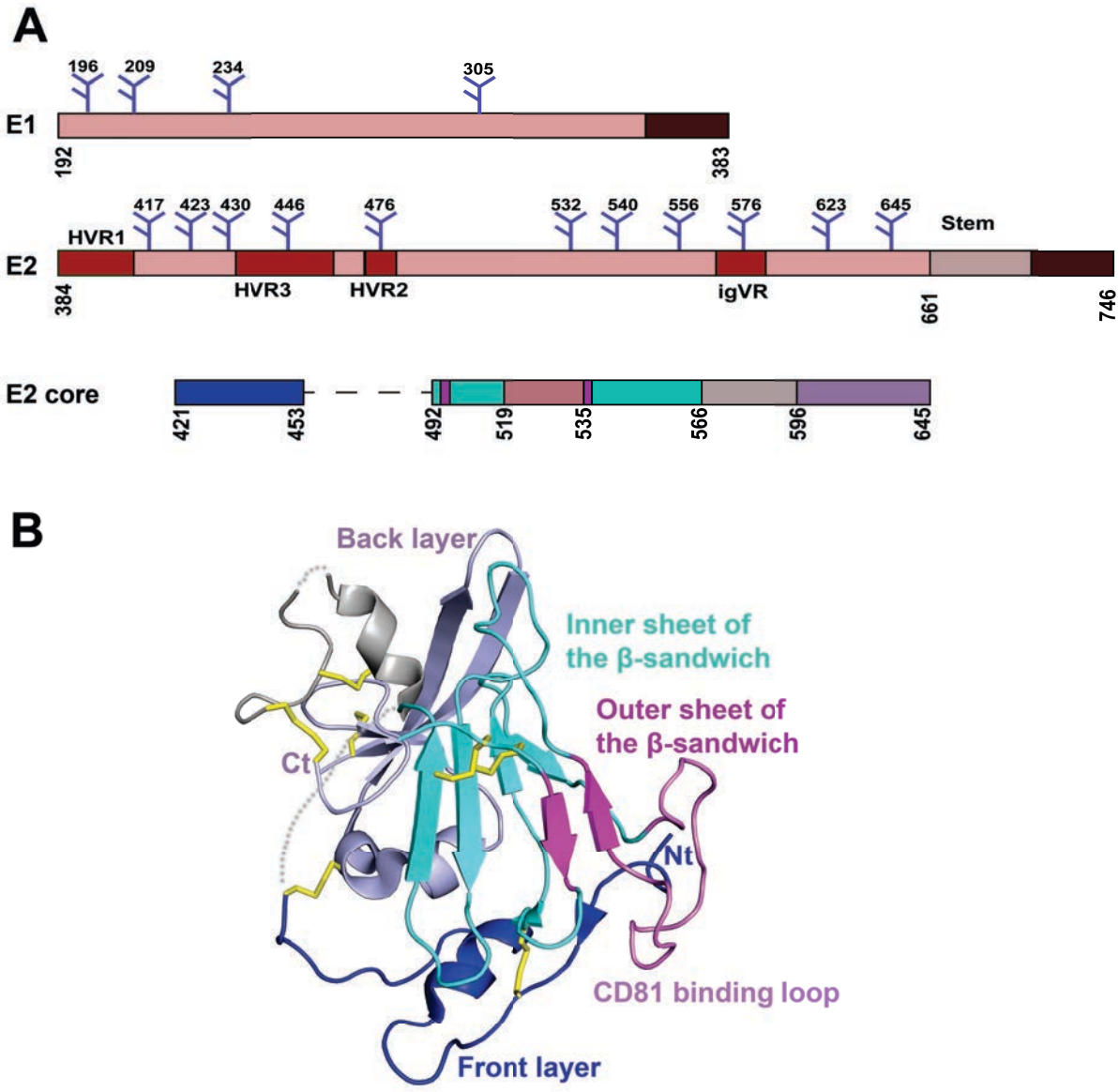


Figure 7. Structure of HCV E2 core fragment. (A) Schematic representation of HCV E1 and E2 glycoproteins and the E2 core domain crystallized by Kong *et al*. The conserved glycosylation sites are depicted as branched tree. The transmembrane domains are shown in dark red. The hypervariable regions within E2 are depicted in red. Schematic representation of the E2 core is colored as in B. (B) Crystal structure of the E2 core domain displayed as cartoon and colored by main structural components: the inner sheet of the Ig β-sandwich (cyan), the outer sheet of the Ig β-sandwich (magenta), the CD81 receptor binding loop (519-535) (purple), the front layer (blue) and the back layer (violet). Disulphides are displayed as yellow sticks. Disordered regions are shown as dotted lines.

HCV E2 does not show similarity to an extended, multi-domain class II fusion protein fold found in flavivirus envelope fusion proteins despite being closely related. Moreover, in contrast to flavivirus glycoproteins, no structural or oligomeric rearrangements were observed in HCV E2 by solution based small-angle X-ray scattering (SAXS) analysis on protein exposure to pH5 (Khan *et al.*, 2014). Interestingly, the pestivirus bovine viral diarrhea virus 1 (BVDV-1) E2 structure shows divergent organization from both HCV E2 and flavivirus E

glycoprotein structures (El Omari *et al.*, 2013; Li *et al.*, 2013) (**Figure 8**). Pestivirus E2 is composed of four β-sandwich domains A to D, arranged linearly from the N to the C terminus and likewise does not have the characteristic class II fusion protein fold. Both HCV and BVDV-1 E2 lack a lipophilic fusion peptide and most likely function as cell-attachment and receptor binding proteins and do not play a role in fusion of the cellular and viral membranes. It is possible that hepaciviruses and pestiviruses evolved to use different mechanisms for membrane fusion than the flaviviruses, which are prototypic representatives of the *Flaviviridae* family.

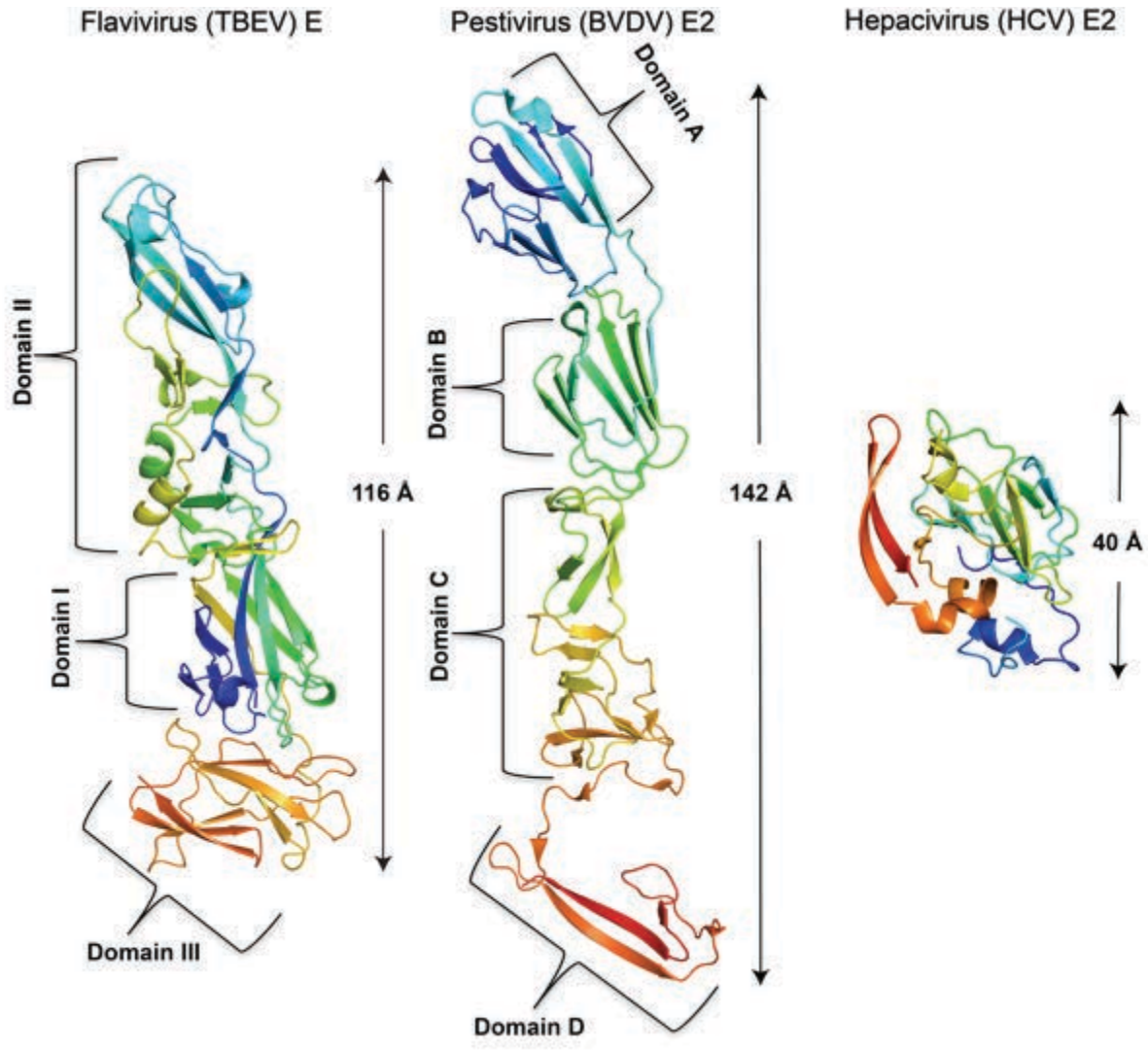


Figure 8. Comparison of the crystal structures of the major envelope glycoproteins in the *Flaviviridae* family: tick-born encephalitis virus (TBEV) E protein for the flaviviruses (PDB ID 1SVB) (Rey *et al.*, 1995), BVDV E2 protein for the pestiviruses (PDB ID 2YQ2 (El Omari *et al.*, 2013) and HCV E2 protein for the hepaciviruses (PDB ID 4MWF (Kong *et al.*, 2013)). The structures are depicted as cartoon and are ramp-colored from N-terminus (blue) to C-terminus (red) through green. TBEV E2 is a prototype class II fusion protein, folded into three domains, domain II harboring the fusion peptide. BVDV E2 is folded into 4 domains, arranged linearly from the N- to the C-terminus. HCV E2 adopts a compact and globular conformation.

Compared with E2, little is known about the smaller HCV glycoprotein E1 (acids aa 192-383 of the strain H77 polyprotein). For a long time it was considered to play a chaperone-like role in the folding of E2 (Michalak *et al.*, 1997). However, since the recently reported structure of HCV E2 argues against E2 being a fusion protein, HCV E1 became a promising candidate for carrying out the membrane fusion activity. E1 has been studied in the context of E1-E2 heterodimer as the folding of E1 in the absence of E2 was reported to be difficult to achieve (Botti *et al.*, 2011; Merola *et al.*, 2001; Michalak *et al.*, 1997). Though E1 is much shorter than prototypic class II fusion proteins, it has been suggested previously to be a truncated class II fusion protein and to contain a putative fusion peptide (Drummer *et al.*, 2007; Flint *et al.*, 1999b; Garry & Dash, 2003). However, recently reported crystal structure of the N-terminal domain (residues 1–79) of the HCV E1 ectodomain revealed a covalently linked homodimer which does not resemble the expected truncated class II fusion protein fold (El Omari *et al.*, 2014).

Interestingly, one study indicated that secretion of recombinant E1 protein requires removal of the C-terminal hydrophobic domain after aa 340 as well as an internal hydrophobic region spanning aa 263-289 (Matsuura *et al.*, 1994). This internal hydrophobic sequence has been suggested to have similarities with the flavivirus and paramyxovirus fusion peptides. It is known that fusion peptides may anchor the protein to cellular membranes and prevent protein secretion (Paterson & Lamb, 1987). Moreover, peptides comprising aa 267-284 and aa 274-291 were shown to interfere with HCVcc infectivity (Cheng *et al.*, 2008) Nevertheless, deciphering the fusion mechanism used by HCV requires more structural data on the smaller glycoprotein E1.

Very little is known about the corresponding GBV-B glycoproteins. GBV-B E2 is smaller than HCV E2 (264 and 363 aa residues, respectively). It is also less glycosylated containing 6 putative N-linked glycosylation sites versus 11 found in HCV E2 (Ghibaudo *et al.*, 2004). GBV-B E1 contains 192 aa residues (193 aa residues in HCV E1) and harbors three putative N-linked glycosylation sites. Although GBV-B and HCV glycoproteins share a similar hydropathy profiles he amino acid sequence identity between them is rather low (~28% between GBV-B and HCV E1 and ~15% between GBV-B and HCV E2). GBV-B glycoproteins are believed to mediate viral entry to target cells, however, no experimental results have been reported to support this hypothesis.

HCV attachment

In order to interact with specific cellular receptors, HCV needs to be brought into close proximity with the target cell surface. The initial attachment of HCV to cells is mediated by low affinity interactions with low-density-lipoprotein receptor (LDL-R), cell surface glycosaminoglycans (GAGs) and C-type lectins such as liver or dendritic cell specific intracellular adhesion molecule 3-grabbing nonintegrins (L/DC-SIGN).

GAGs are long negatively charged polysaccharides found at the surface of cells. It has been demonstrated that HCVpp, HCVcc, and recombinant E1 and E2 can bind GAGs such as highly sulfated heparansulfate proteoglycans (Barth *et al.*, 2003; Barth *et al.*, 2006; Jiang *et al.*, 2012; Kobayashi *et al.*, 2012; Koutsoudakis *et al.*, 2006; Morikawa *et al.*, 2007). The interaction between HCV and GAGs is likely mediated by the positively charged N-terminal portion of E2 (Penin *et al.*, 2001). Moreover, there is some evidence that apoE can contribute to HCV attachment via GAGs (Jiang *et al.*, 2012). LDL-R was initially proposed to works as an attachment factor for the lipoproteins associated with the HCV particle (Agnello *et al.*, 1999). LDL-R is able to recognize apoB and apoE on the HCV lipoviral particles (Huang *et al.*, 2007). Since E1 was shown to interact both with apoE and apoB, the association of these lipoproteins with E1 may facilitate virus entry through LDL-R (Mazumdar *et al.*, 2011). However, another study suggests that LDL-R is involved in non-productive HCV entry, which does not lead to viral infection, and rather plays a role in HCV replication. The observation that soluble LDL-R blocks HCVcc infectivity might be related to the fact that it affects the interaction of ApoE present in HCV lipoviroparticles with cell-surface heparan sulfate proteoglycans (Albecka *et al.*, 2012).

L-SIGN and DC-SIGN recognize mannose-type oligosaccharides present on glycoproteins. Since L-SIGN is found on endothelial cells and on liver sinusoid cells, and DC-SIGN is expressed in dendritic cells, they are likely to serve as tissue-specific capture receptors (Cormier *et al.*, 2004a; Gardner *et al.*, 2003; Lozach *et al.*, 2003).

Essential cellular factors for HCV entry

The first and the best-characterized cellular entry factor for HCV entry is the human tetraspanin **CD81** (CD = cluster of differentiation). CD81 is expressed on most human cells and is involved in a number of cellular processes (Jones *et al.*, 2011; Rocha-Perugini *et al.*,

2008; van Spriel, 2011). CD81 is a type III membrane protein (defined by multiple transmembrane domains) belonging to the tetraspanin family. It possesses short intracellular N and C termini, two extracellular loops (a long and a short one), and four transmembrane domains. The long extracellular loop (McLellan *et al.*) interacts with HCV E2 and is critical for virus entry (Pileri *et al.*, 1998). The requirement of CD81 for HCV entry was confirmed by a number of studies using different techniques such as anti-CD81 antibodies and inhibition of infection by soluble recombinant CD81-LEL (Bartosch *et al.*, 2003c; Heo *et al.*, 2006; McKeating *et al.*, 2004; Meuleman *et al.*, 2008; Molina *et al.*, 2008; Zhang *et al.*, 2004). CD81 is also important in defining the tropism of HCV for human cells; however, recent studies indicate that it is not the sole factor participating in species restriction (Bartosch *et al.*, 2003c; Cormier *et al.*, 2004b; Masciopinto *et al.*, 2002; Meola *et al.*, 2000; Rocha-Perugini *et al.*, 2009). CD81 is not thought to be involved in initial HCV virion binding to target cells and most likely acts in post-attachment steps of viral entry (Cormier *et al.*, 2004b).

Disrupting the interaction of CD81 and E2 prevents HCV entry. Most of the neutralizing antibodies block HCV infection by preventing E2 binding to this receptor. Therefore, the regions of E2 responsible for the interaction between the two proteins have been extensively studied, revealing several discontinuous regions involved in CD81 binding (Drummer *et al.*, 2006; Owsianka *et al.*, 2006). The E2 binding site within CD81-LEL is conformational and amino acid F186 has been identified to be particularly critical for binding E2 (Drummer *et al.*, 2002; Higginbottom *et al.*, 2000).

HCV particles have been shown to tolerate acidic environments. Interestingly, treatment of HCV with soluble CD81 can induce HCV fusion with the cell plasma membrane at low pH (Sharma *et al.*, 2011). Therefore, it is likely that the interaction of CD81 with HCV glycoproteins induces conformational changes in the glycoproteins that render them capable of low-pH dependent fusion.

Kong and co-workers provided the first insight into CD81 binding to E2 by negative stain electron microscopy (EM) reconstruction of a ternary complex between E2 ectodomain, CD81 LEL and a Fab AR2A. The EM density for CD81 corresponded to the CD81 dimer interacting with the E2c front layer in which the residues previously shown to be critical for CD81 binding are located (Kong *et al.*, 2012b).

SR-BI (Scavenger receptor BI) is another indispensable cellular factor for HCV entry. This protein is expressed on hepatocytes where it functions as a receptor for different classes of lipoproteins (Krieger, 2001). SR-BI interacts with lipoproteins displayed on the HCV particles (Catanese *et al.*, 2007; Dao Thi *et al.*, 2012; Maillard *et al.*, 2006), but also

participates in post-binding events. It has been demonstrated to bind to the HVR1 region in E2 (Scarselli *et al.*, 2002), which presumably leads to exposure of the CD81 binding site and subsequent interaction between CD81 and E2 (Bankwitz *et al.*, 2010). Also, SR-BI is believed to facilitate HCV internalization through its lipid transfer activity (Dreux *et al.*, 2006).

The tight junction proteins claudin-1 (**CLDN1**) and occludin (**OCLN**), involved in the formation of tight junctions, are two additional HCV entry factors (Evans *et al.*, 2007). They form a barrier regulating permeability of endothelial and epithelial cells and also are important in maintaining cell polarity.

CLDN1 is present not only in tight junctions but also in lower amounts at the basolateral surface of hepatocytes. CLDN1 does not interact directly with the HCV glycoproteins, but contributes to HCV entry via association with CD81 (Harris *et al.*, 2010). OCLN has been also implicated in HCV entry, since silencing its gene in permissive cells rendered them resistant to both HCVpp and HCVcc infection (Ploss *et al.*, 2009). CLDN1 and OCLN have been shown to contribute to the tropism of HCV to human cells together with CD81 and SR-BI (Ploss *et al.*, 2009) (Dorner *et al.*, 2011).

The observation that HCV patients tend to accumulate iron in the liver led to identification of the iron uptake receptor, transferrin receptor (**TfR1**), as one more essential cellular factor for HCV entry (Martin & Uprichard, 2013). A TfR1 knock-down or its inhibition by antibodies, prevents HCV infection and kinetic studies suggested that TfR1 acts downstream of the interaction with CD81, but the exact molecular mechanism of how TfR1 is involved in virus entry remains to be determined.

Niemann-pick C1-like 1 (**NPC1L1**) protein was also included in the list of indispensable host factors for HCV entry. In contrast to other entry factors that are located at the basolateral surface of hepatocytes or in tight junctions, NPC1L1 is found on the apical side where its role is the reabsorption of unesterified cholesterol in bile. NPC1L1 associated with cholesterol is endocytosed (Jia *et al.*, 2011). Silencing NPC1L1 or using specific inhibitor ezetimibe or antibodies directed against NPC1L1 impairs HCVcc entry in cell culture (Sainz *et al.*, 2012). The precise role of NPC1L1 in HCV entry is unknown but it might be related to cholesterol uptake.

In addition, efficient HCV entry depends on host cell kinases that are involved in specific signal transduction pathways. For example, protein kinase A (PKA) and two receptor tyrosine kinases - epidermal growth factor receptor (EGFR) and ephrin type A receptor 2 (EPHA2) – signaling is important for the interaction between CD81 and CLDN1 (Farquhar *et al.*, 2008; Lupberger *et al.*, 2011). (Simmonds *et al.*, 1993).

No experimental evidence has been reported to illustrate, which cellular factors are required for GBV-B entry.

An integrated overview of HCV entry

Although the exact role of each HCV entry factor has not been deciphered yet, current knowledge allows suggesting a potential HCV entry model. Circulating viral particles are guided to permissive cells by their interaction with L-SIGN and DC-SIGN that are believed to function as “capture receptors”. Then GAGs and LDL-R facilitate initial attachment of HCV particles to target cells, likely occurring via binding of the lipoprotein components of HCV virions to LDL-R and GAGs and/or direct interaction with the envelope glycoproteins. After the initial attachment to hepatocytes, HCV particles interact with SR-BI and CD81. It has been proposed that the interaction with SR-BI occurs upstream of the binding to CD81. The HCV-receptor complex may then laterally migrate to the tight junctions or trigger cytoskeletal rearrangements, which results in bringing the virus to sites containing CLDN-1 and OCLN. There is then formation of an E2-CD81-CLDN-1 complex, which facilitates internalization of the virus particle via clathrin-mediated endocytosis. OCLN is also internalized via clathrin-mediated endocytosis though its role in the entry process is not clear yet. The virus particle is transported to the endosomes, where pH-dependent fusion of the viral envelope with endosomal plasma membranes is believed to take place. The fusion mechanism of HCV remains a grey area. Nevertheless, the HCV entry process is definitely more complex than described. For example, the exact functions of recently identified entry factors TfR1, NPC1L1, EGFR and EPHA2 still need to be elucidated in order to complete a scheme of HCV entry.

HCV replication and assembly

Once the viral particle has been internalized into endosomes, fusion of viral and cellular membranes takes place, triggered by environmental acidic pH in the endosomes as described for other viruses (Meertens *et al.*, 2006) (Lavillette *et al.*, 2006). As a result, the capsid is released into the cytoplasm and dissociates from the viral genome. The viral genome directly serves as a messenger RNA (mRNA) for translation of the viral polyprotein. An

internal ribosome entry site (IRES) element in the 5′-UTR of the viral RNA initiates the translation by binding to ribosomes. The polyprotein, which is inserted into the ER membrane, is thereafter processed by host and viral proteases generating the viral proteins. NS3, 4A, 4B, 5A, and 5B as well as cis-acting elements in the viral RNA are involved in the viral genome replication. The replication sites are located in a vesicular membrane structure designated as the membranous web (reviewed in (Niepmann, 2013)).

Viral RNA synthesis proceeds via a negative-strand RNA intermediate, which serves as a template for production of the positive-sense, single-stranded RNA genome, which is encapsidated by the core (C) protein into nucleocapsids. The exact mechanism for HCV particle assembly is not completely clear. HCV core protein accumulates around lipid droplets (LDs), where nucleocapsid formation and packaging of the viral genome are thought to occur. It has been assumed that the nucleocapsids acquire their lipid envelopes containing the glycoproteins E1 and E2 by budding into the ER lumen. The release of the viral particles from the cell is believed to occur through the cellular secretory pathway (reviewed in (Lindenbach, 2013; Lohmann, 2013; Niepmann, 2013)).

Role of neutralizing antibodies in HCV infection

During the last decade evidence has accumulated that supports the role of the humoral immune response in controlling HCV infection. Initially it was believed that only cell mediated immune response was important as indicated by a number of studies demonstrating correlation between viral clearance and strong cell mediated immune response. The importance of the other part of the adaptive immune system in the control of HCV infection, the humoral immune response, has been more difficult to study due to 1) the absence of a convenient cell culture system that allows evaluation of the neutralizing activity of anti-HCV antibodies together with their autologous virus isolate and 2) the heterogeneity of patient cohorts.

Antibodies against both structural and non-structural proteins are elicited in HCV infected patients (Sillanpaa *et al.*, 2009). Typically, HCV RNA can be detected within 1–3 weeks post-infection, but antibody responses are delayed for 7-10 weeks post-infection in both self-resolving and chronically evolving HCV (Bowen & Walker, 2005). In general, antibodies against HCV glycoproteins appear later and at lower titers than those to the non-structural proteins. The majority of raised antibodies are restricted to the IgG1 isotype,

suggesting that the immunoglobulin class-switching process might also be disturbed (Chen *et al.*, 1999b; Netski *et al.*, 2005).

Evidence that HCV specific antibodies can contribute to control HCV *in vivo* is derived from studies carried out in chimpanzees and mice. One of the earliest studies showed that HCV pre-treated with serum from a chronically infected individual was not infectious when inoculated into chimpanzees, suggesting the presence of neutralizing antibodies in the serum (Farci *et al.*, 1994). In another study, chimpanzees immunized with recombinant HCV glycoproteins E1 and E2 were protected against experimental challenge with a homologous HCV strain, which correlated with the presence of antibodies against HCV envelope glycoproteins (Choo *et al.*, 1994). In addition, the chimpanzee with the highest-level of E2-specific antibody response after vaccination was also protected against heterologous virus challenge (Youn *et al.*, 2005). One more study revealed that the monoclonal antibody HCV1, directed against E2 amino acids 412-423, was able to completely prevent HCV infection when administered to a naïve chimpanzee (Morin *et al.*, 2012).

The studies in mice also showed that passive administration of human anti-HCV polyclonal antibodies or neutralizing anti-E2 mAbs can protect human liver-chimeric mice against heterologous virus challenge (Law *et al.*, 2008; Meuleman *et al.*, 2012). Immunization of immunocompetent humanized mice with a vaccinia virus vector expressing structural HCV proteins elicited humoral response protecting some animals against heterologous challenge (Dorner & Ploss, 2011).

One of the first reports revealing a protective effect of anti-HCV antibodies in humans is a retrospective cohort study of patients who underwent liver transplantation. Patients receiving polyclonal immunoglobulins against hepatitis B virus surface antigen (HBIGs) had a lower incidence of HCV infection after transplantation in comparison with those not receiving HBIGs. HBIGs used for the treatment were shown to contain anti-HCV antibodies implying a passive transfer of anti-HCV antibodies to the patients receiving HBIGs (Feray *et al.*, 1998). HCV infection in hypogammaglobulinaemic patients is associated with rapid disease progression (Bjoro *et al.*, 1994). In addition, rituximab-combination chemotherapy results in exacerbation of HCV infection due to depletion of B cells (Ennishi *et al.*, 2008). One of the rare cases when HCV infection was cleared after >65 weeks without therapy was associated with antibodies neutralizing the infecting HCV genotype and the absence of viral escape mutants (Raghuraman *et al.*, 2012).

The quality and magnitude of the antibody response in the early stages of an HCV infection has been shown to predetermine the course of the infection. Several studies showed

that individuals who rapidly develop a broad neutralizing antibody (nAb) response after HCV exposure are more likely to resolve the infection (Dowd *et al.*, 2009; Logvinoff *et al.*, 2004; Osburn *et al.*, 2014; Pestka *et al.*, 2007). Pestka and colleagues studied a cohort of individuals infected with a single-source HCV genotype 1b inoculum and discovered that the induction of high titers of cross-reactive nAbs during the acute phase of infection correlated with viral clearance (Pestka *et al.*, 2007). In contrast, a weak antibody response, or none at all, was detected during the acute phase in patients who later developed chronic disease. In chronically infected patients the antibody response increased over time and high antibody titers could be detected after 10 or more years post-infection. A recent study of a cohort of HCV infected patients showed a similar pattern, i.e. a broad nAb response in an early stage of infection was associated with spontaneous recovery (Osburn *et al.*, 2014). Moreover, this study also revealed that the breadth of nAb response during acute infection is not related to the infection genotype. Thus, the early appearance of broad nAb response is contributing to the efficient control of HCV infection.

In conclusion, in persistently infected patients antibody responses are delayed and are less broad. While the nAb response diminishes after the clearance of the virus and is undetectable in the majority of patients after 10-17 years, nAb responses are gradually broadened in patients with persistent infection. Thus, high titers of cross-reactive neutralizing antibodies can be detected in the majority of individuals with a chronic HCV infection, but they are not able to clear the virus.

Beside their direct neutralizing activity, a role of Fc-effector function of anti-HCV antibodies in controlling HCV infection is emerging. Anti-HCV antibodies have been reported to mediate antibody dependent cytotoxicity, complement-dependent cytotoxicity, and antibody-dependent phagocytosis by cells carrying Fcγ receptors (Meyer *et al.*, 2002; Nattermann *et al.*, 2005).

Antibodies targeting HCV glycoproteins

The most common way in which nAbs prevent infection is by binding directly to virus particles and by interfering with viral entry or post entry processes such as viral uncoating. NAbs usually block virus entry due to direct competition for binding residues involved in receptor interactions or steric hindrance and shielding of receptor binding sites (Corti & Lanzavecchia, 2013). The nAbs elicited during HCV infection primarily target linear or

conformational epitopes within the envelope glycoproteins E1 and E2 and block interaction with the cellular receptors SR-BI and CD81 (**Figure 9**).

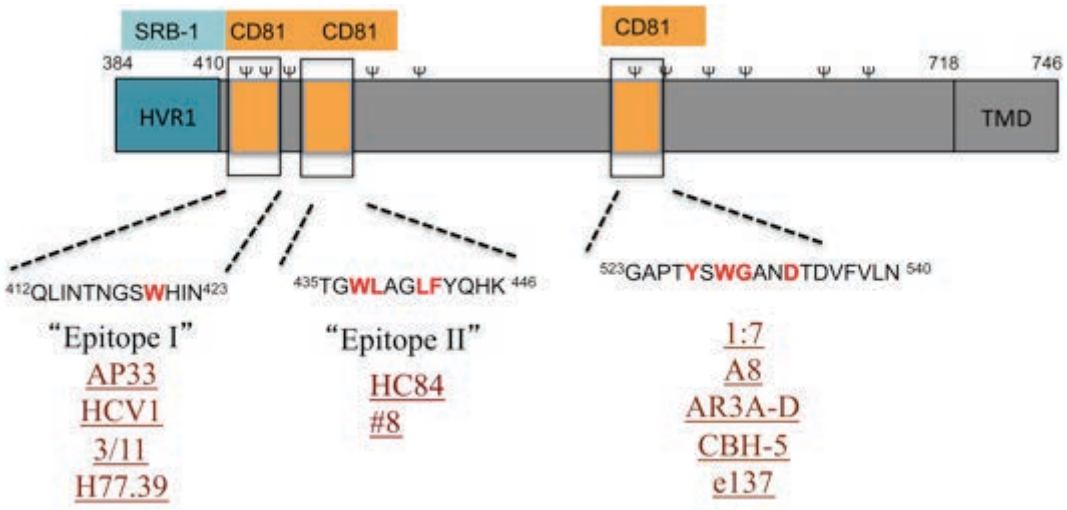


Figure 9. Main regions containing neutralizing epitopes within HCV E2 glycoprotein. Hypervariable region 1 (HVR1) is a prime target for neutralizing antibodies, although antibodies targeting it tend to be genotype and isolate specific. Antibodies that show broadly neutralizing activity are directed against three conserved discontinuous regions (orange) within E2: 1) epitope I (aa 412-423), 2) epitope II (aa 435-446) and 3) region comprising aa 523-540. These three regions overlap with a composite CD81 binding site, thus, antibodies targeting them inhibit E2 interaction with CD81. Several neutralizing antibodies targeting each of the three regions are listed below.

Neutralizing antibodies against HVR1

HVR1, spanning the first 27 amino acids at the N-terminus of E2 (aa384-410), is an immunodominant region and a major target for nAbs. HVR1 plays an important role in SR-BI binding (see section Essential cellular factors for HCV entry). Anti-HVR antibodies have been observed *in vivo* and were shown to neutralize HCV infection. But due to high genetic variability of this region, these nAbs exhibit isolate-specific neutralization (Kato *et al.*, 1994; Kato *et al.*, 1993; Taniguchi *et al.*, 1993; Weiner *et al.*, 1992; Zucchelli *et al.*, 2001). The protective potential of anti-HVR1 antibodies was demonstrated in chimpanzees. An antibody raised against a peptide located in HVR1 protected chimpanzees against the challenge with HCV carrying the autologous HVR1 sequence (Farci *et al.*, 1996b).

Anti-HVR1 antibodies are either neutralizing or non-neutralizing depending on the location of their epitope within HVR1. Two linear antigenic regions have been described within the HVR1, one located at the N-terminus and the other one located rather at the C-terminus, but only the one located at the C-teminus contains neutralizing epitopes. For example, the epitope of mAb 9/27 has been mapped to aa 396-407 within HVR1 and this

potent nAb blocks SR-BI binding to E2 and neutralizes infectivity of genotype 1a HCVpp. In contrast, rat mAbs 6/16, 7/59, 6/82, which target aa 384-395, and are non-neutralizing (Bartosch *et al.*, 2003a; Hsu *et al.*, 2003).

Some studies have claimed that resolved HCV infections correlate with stable HVR1 sequences and early induction of anti-HVR1 antibodies, while considerable sequence changes within HVR1 are detected in the case of chronic infections (Farci *et al.*, 2000; Ray *et al.*, 1999; Zibert *et al.*, 1997). However, Liu and colleagues contradict this hypothesis by showing that a spontaneous clearance of HCV can be associated with a rapid evolution of the HVR1 sequence, but which is driven by a strong and early nAb response during the acute phase of infection. On the other hand, the changes within the HVR1 sequence were more convergent and occurred at a slower rate due to the later and weaker nAb response in chronically evolving infection (Liu *et al.*, 2012). Therefore, it is likely that a strong nAb response directed against other, more conserved, epitopes within E2 during an early stage of infection may have more impact on the outcome of the infection.

Neutralizing antibodies interfering with CD81 binding to E2

The majority of broadly nAbs (bnAbs) are directed against both linear and conformational epitopes overlapping with the CD81 binding site within the envelope glycoprotein E2 and block interaction between CD81 and E2. Specifically, their epitopes include at least one of the highly conserved residues critical for CD81 binding (i.e. W420, Y527, W529, G530 and D535) as well as the amino acid segment 436-GWLAGLFY-443 (Drummer *et al.*, 2006; Owsianka *et al.*, 2006).

Two linear epitopes within E2, designated epitope I and II, have been described to elicit bnAbs. Epitope I is located immediately downstream of HVR1 (aa 412-423) and is recognized by a mouse monoclonal antibody (mAb) AP33 exhibiting broad neutralizing activities against all major genotypes and isolates (Owsianka *et al.*, 2005). Some other nAbs (bnAbs) broadly targeting this region include the rat mAb 3/11 (Flint *et al.*, 1999a; Tarr *et al.*, 2006), mAb HCV1 and 95-2 isolated from a transgenic mice containing human antibody genes (Broering *et al.*, 2009) as well as a subset of human mAbs designated HC33-related mAbs (Keck *et al.*, 2013). The bnAb HCV1 has been shown to prevent and treat HCV infection in chimpanzees (Morin *et al.*, 2012). For all these bnAbs tryptophan residue 420, which is conserved among all HCV genotypes and is required for interaction with CD81

(Owsianka *et al.*, 2006), has been reported to be an essential contact residue. However, human nAbs targeting this epitope are very rare. Different groups have reported that they could identify such mAbs in only 2-3% (Tarr *et al.*, 2007a; Tarr *et al.*, 2012) or 15% (Keck *et al.*, 2013) of sera samples from infected individuals. The observed variation probably is related to differences in the studied populations of the infected individuals and their stages of HCV infection. Keck and colleagues have suggested that the observed low frequency of mAb responses to aa 412 to 423 is due to shielding this site by N-glycans and /or diversion of the immune response from this site to more immunodominant regions (Keck *et al.*, 2013).

Another conserved epitope targeted by broadly neutralizing antibodies encompasses amino acids 434-446 of E2, and is referred to as epitope II. The sequence 436-GWLAGLFY-443 within epitope II has been reported to contribute to CD81 binding (Drummer *et al.*, 2006) suggesting a direct overlap with the CD81 binding site. This region contains both variable and conserved residues. As a result, both neutralizing and non-neutralizing antibodies recognizing this antigenic site have been identified (Duan *et al.*, 2012; Hsu *et al.*, 2003; Keck *et al.*, 2012; Tarr *et al.*, 2012) (Zhang *et al.*, 2007; Zhang *et al.*, 2009). Initial studies suggested that non-neutralizing antibodies targeting this region interfere with neutralization mediated by bnAbs directed to epitope I (Zhang *et al.*, 2007; Zhang *et al.*, 2009). Subsequent studies, however, were unable to confirm this relationship between antibodies binding to the two epitopes, and revealed mainly additive neutralization when both sets of antibodies were combined (Keck *et al.*, 2013; Tarr *et al.*, 2012).

nAbs directed against epitope II were reported by a number of different groups (Duan *et al.*, 2012; Hsu *et al.*, 2003; Keck *et al.*, 2012; Tarr *et al.*, 2012). A panel of epitope II specific human mAbs HC84-1–HC84-27 isolated by Keck *et al.* has been demonstrated to not only possess broad neutralizing activity but also to be resistant to neutralization escape in a cell culture system. Though the epitope of HC84-related mAbs is predominately linear, alanine scanning mutagenesis experiments revealed that some of them also bind to residues in a second E2 segment spanning aa611-613.

The core-binding region of the majority of human bnAbs is the E2 region encompassing aa 523-540 (reviewed in Ball et al., 2014, Edwards et al., 2012. This region is targeted by many well-characterized human bnAbs, e.g., A8 (Johansson *et al.*, 2007), e137 (Perotti *et al.*, 2008), CBH-7 (Hadlock *et al.*, 2000), HC11, HC-1 (Keck *et al.*, 2008b), AR3A (Law *et al.*, 2008). Importantly, human bnAbs targeting this E2 segment are exclusively conformation-sensitive and for some of them, such as the human mAbs AR3C and e137, or an alpaca nanobody D03 (Tarr *et al.*, 2013), alanine scanning mutagenesis suggested also some

contact residues located within epitope I and/or epitope II. The conserved amino acid residues Y527, W529, G530 and D535 within this region are critical for CD81 interaction with E2 (Owsianka *et al.*, 2006). The bnAbs directed against aa 523-540 have been described to bind to at least two out the four residues G523, W529, G530 and D535 and, as a result, to inhibit the interaction with CD81. Human bnAbs to overlapping epitopes located within this site have been isolated from a number of different patients, indicating that this region is immunogenic, although they typically appear later during the natural course of HCV infection. However, murine antibodies binding to aa 523-540 are rare and they also generally exhibit conformation dependent binding to E2 (for example, mAb 9/75 (Hsu *et al.*, 2003), H35 and H48 (Cocquerel *et al.*, 1998; Owsianka *et al.*, 2006)). To date, only a few murine nAbs, recognizing a linear epitope within aa 523 -540 have been reported, and include mAb 1H8 (aa 524-529) (Zhao *et al.*, 2014) and mAb 2/64a (Hsu *et al.*, 2003).

Other neutralizing epitopes within E2

Although the majority of nAbs target epitopes within the CD81 binding site or HVR1 region, some nAbs recognize epitopes outside those regions. AR4A and AR5A are two nAbs that specifically recognize the E1E2 heterodimer (Giang *et al.*, 2012). E2 residues D698 and R639 are essential for binding of those nAbs, indicating that this region outside the CD81 binding site is able to elicit a nAb response. However, the exact neutralization mechanism used by these nAbs remains to be determined.

Kachko and colleagues also found, after vaccination of mice with recombinant E1E2, that aa 448-483 and aa 496-515 of E2 contain neutralizing epitopes (Kachko *et al.*, 2011). While the region encompassing aa 448-483 is poorly conserved among different HCV isolates, the segment spanning aa 496-515 is highly conserved. Nevertheless, in a natural infection, antibodies against this region are rarely observed, suggesting that it is less immunogenic when displayed on native virions.

Non-neutralizing anti-HCV E2 antibodies

As mentioned previously, HCV infection also induces antibodies that do not interfere with the viral cycle. Non-neutralizing antibodies (non-nAbs) recognizing linear epitopes

within the N-terminal region of HVR1 and other regions within E2 have been described (Clayton *et al.*, 2002; Flint *et al.*, 1999a; Hsu *et al.*, 2003). In addition, six human non-nAbs designated CBH-4D, -4B, -4G, -20, -21 and -22, recognizing a specific region within E2 known as "antigenic domain A", were shown to be conformation-sensitive (Hadlock *et al.*, 2000; Keck *et al.*, 2005; Keck *et al.*, 2004). Those non-nAbs do not cross-compete with nAbs directed against the CD81 binding site, implying that their epitopes are located in a spatially different surface of the E2 glycoprotein. Antigenic domain A is more sensitive to low-pH treatment than the regions involved in CD81 binding, as indicated by significantly decreased binding of non-nAbs to low-pH-treated HCVpp. It was suggested that low pH induces conformational change within antigenic domain A, which may be related to E1E2 disassociation (Keck *et al.*, 2005).

Anti-HCV E1 antibodies

The envelope glycoprotein E1 can also be targeted by antibodies, although these antibodies are infrequent in patient sera (Leroux-Roels *et al.*, 1996; Penin *et al.*, 2001). The low immunogenicity of E1 may be due to the shielding of immunogenic regions of E1 by the E2 glycoprotein or glycan moieties, or due to immunodominance of E2 (Fournillier *et al.*, 2001; Garrone *et al.*, 2011). In addition, the possibility that E1 antibodies remain undetected because of technical issues cannot be ruled out, particularly in view of the fact that E1 misfolds in the absence of E2 (Michalak *et al.*, 1997).

Despite difficulties triggering the anti-E1 antibody response, several neutralizing anti-E1 antibodies have been described. MAb H-111 targeting aa 192-211 can neutralize HCV genotypes 1a, 1b, 2b and 3a, and reduce virus entry into Molt-4 cells. The human mAbs IGH505 and IGH526, which recognize a linear epitope spanning E1 aa 313-327, have been shown to possess broadly neutralizing activity (Meunier *et al.*, 2008b).

An E1 protein vaccine has been explored in several trials and was shown to induce anti-E1 antibody responses (Garrone *et al.*, 2011; Leroux-Roels *et al.*, 2004; Nevens *et al.*, 2003; Verstrepen *et al.*, 2011). Verstrepen *et al.* demonstrated that vaccination with a HCV E1 subunit vaccine elicits nAbs in chimpanzees and protects the vaccinated animals against experimental infection (Verstrepen *et al.*, 2011). In another study, vaccination of chronically infected patients with HCV E1 did not have any effect on viral RNA levels, but did result in improved liver fibrosis in some patients (Nevens *et al.*, 2003). Garrone et al. studied a prime-

boost vaccination strategy in chimpanzees using VLPs pseudotyped with E1 and E1E2. These VLPs were generated by coexpressing the Gag protein of Moloney murine leukemia virus (MLV) with the E1 and E2 envelope proteins of HCV in mammalian cells. These constructs lacked the MLV enzymes (protease, reverse transcriptase,integrase) to make them safer, as with other vaccine platforms. Intriguingly, induced anti-E1 antibodies were non-neutralizing unless the animals were boosted with the VLPs pseudotyped with E1E2 (Garrone *et al.*, 2011).

The neutralization mechanism of anti-E1 antibodies is still unknown. Since the E2 structure revealed that it is unlikely to carry membrane fusion activity (see above), E1 may represent the HCV fusion protein. It is possible that the currently identified anti-E1 nAbs may possibly inhibit the fusion step.

HCV immune escape strategies

HCV is able to persist in an infected patient despite the presence of nAbs, implying that the virus evades the host immune response. Several immune evasion strategies to escape from the nAb response have been described for HCV.

Mutational escape

The primary reason for the virus' persistence in chronically infected persons despite the presence of nAbs is the rapid evolution of neutralization resistant viral variants via mutations in neutralizing epitopes. The error-prone genome amplification strategy leads to the generation of a diverse but related population of viral variants, referred to as quasispecies. It has been estimated that more than 10^{12} quasispecies can be generated in an infected individual each day. Because the HCV RNA-dependent RNA polymerase lacks proof-reading activity, the error rate is about $\sim 10^{-4}$ base substitutions per site per virus generation (Cuevas *et al.*, 2009), meaning that on average, each replicate genome will have 1 mutation.

Amino acid substitutions occur at the highest rate in the envelope glycoproteins E1 and E2, notably the main targets of nAbs. As a result, viral variants capable of escaping immune surveillance appear in the virus population. Since such viral mutants have a selective advantage over neutralization sensitive viral variants, they quickly become the dominant circulating species. Not surprisingly, the nAb response lags behind due to a continual

evolution of the envelope glycoprotein sequence. For instance, the serum samples from infected patients were shown to be inefficient for neutralizing the coexisting virus but were able to neutralize the viral variants from earlier time points (Dowd *et al.*, 2009; von Hahn *et al.*, 2007).

Positively selected mutations mostly concentrate within or close to the binding sites of cellular receptors and nAb epitopes (Brown *et al.*, 2005; Brown *et al.*, 2007). As described earlier, HVR1 is the most variable and immunogenic region of the E2 glycoprotein. Immune pressure has been suggested to be a driving force of mutation of HVR1 and one of the main contributors to the establishment of chronic infections. Despite a high rate of amino acid substitutions in HVR1, it retains the overall basic charge, which may be crucial for HVR1 interaction with negatively charged entry factors on the cell surface (Penin *et al.*, 2001). Mutations that result in an increase of the basic charge within HVR1 are associated with improved viral infectivity (Callens *et al.*, 2005).

The discontinuous CD81 binding region contains a number of universally conserved residues among different HCV isolates. As they are crucial for virus entry, viral variants harboring substitutions of those conserved residues are non-viable. As a consequence, HCV evolved to use additional evasion strategies in order to prevent neutralization by nAbs targeting CD81 binding sites.

Shielding

The CD81 binding site is the most conserved part of the E2 glycoprotein, thus, HCV has developed various mechanisms for shielding this site from the nAb response. One of the mechanisms employed by HCV is shielding of the conserved CD81 binding region by the HVR1, which protects this site from recognition by nAbs.

HVR1 has been shown to be non-essential for HCV infectivity, even though the removal of this region has a negative effect on viral replication (Forns *et al.*, 2000). Several studies imply that HVR-1 masks the CD81 binding region within E2, because viruses lacking the HVR1 are more sensitive to neutralization by nAbs targeting the CD81 binding site (Bankwitz *et al.*, 2010; Prentoe *et al.*, 2011; Zucchelli *et al.*, 2001).

A second mechanism to mask the CD81 binding site is via the so-called “glycan shield” provided by the extensive glycosylation. HCV glycoproteins E1 and E2 carry 11 and 4-5 (depending on the strain) glycosylation sites, respectively. Nine out of eleven N-linked

glycosylation sites within E2 are conserved across HCV genotypes. The long and flexible carbohydrate chains have been implicated to interfere with the access of nAbs to the important neutralizing epitopes (Falkowska *et al.*, 2007; Helle *et al.*, 2007; Helle *et al.*, 2010). Site-directed mutagenesis revealed that ablation of any of the five N-linked glycans on E2 (E2N1, E2N2, E2N4, E2N6 and E2N11; positions 417, 423, 448, 532 and 645) resulted in an increased sensitivity of HCVcc to neutralization by nAbs recognizing the CD81 binding site. Specifically, the glycans E2N1, E2N2, E2N4 and E2N6 were suggested to shield the CD81 binding region. E2N7 (position 540) is also considered to contribute to the masking of the CD81 binding site, although it is absent in genotypes 3 and 6. In addition, changes in N-linked glycosylation can also lead to escape from nAbs. Asparagine 417 within epitope I has been described to be glycosylated in the context of virions. Also, a virus bearing a substitution of asparagine 417 to serine or threonine was shown to have the glycan attachment site shifted from N417 to N415, giving rise to a virus that is resistant to a number of broadly neutralizing nAbs (Pantua *et al.*, 2013).

Lipoproteins associated with HCV particles are also considered to protective HCV particles from the nAb response. HCV circulates in patient sera as lipoviralparticles (LVPs). The low density LVPs, which are associated with LDL or VLDL such as apoB and apoE are the most infectious (Andre *et al.*, 2002). In contrast, the high density LVPs are found to be associated with Igs indicated that they are recognized easier by antibodies (Hijikata *et al.*, 1993b). The density of viral particles has been demonstrated to correlate with sensitivity to nAbs, suggesting that lipoproteins modulate the accessibility of neutralizing epitopes (Tao *et al.*, 2009) (Grove *et al.*, 2008). HCVcc lipoprotein content can be reduced by certain mutations in E2, resulting in viral particles more sensitive to neutralization by anti-E2 nAbs. In addition, the increased sensitivity to neutralization of the viral particles produced in serum-free medium might also be a consequence of an altered lipoprotein composition such as reduced level of apoE associated with the virions (Akazawa *et al.*, 2011).

The high density lipoproteins (HDL) of human serum alter the antibody-mediated neutralization of HCV by enhancing internalization of the virus via HVR1 and SR-BI-dependent mechanism. Inhibiting the lipid transfer function of SR-BI significantly reduces the nAb titers required for complete inhibition of viral infectivity. Due to acceleration of viral entry, nAbs have a reduced time window to interact with the viral particles (Dreux *et al.*, 2006). Interestingly, the HDL related enhancement of viral entry affects in particular nAb directed against the CD81 binding region, suggesting that such nAbs are probably the most efficient at the early stage virus/cell interactions.

Cell-to-cell transmission

HCV is also able to spread by direct transmission between contacting cells in a tissue. This transmission mode exploits cell-cell contacts and has been observed for a number of enveloped viruses, including HIV, human T-cell lymphotropic virus (HTLV), measles virus and herpes viruses. Cell-to-cell transmission not only results in a rapid spread of the infection but also protects virus from the exposition to circulating nAbs as viral particles are released only to a confined space between the cells where antibodies are less likely to access (reviewed in (Mothes *et al.*, 2010)). HCV has been shown to use this infection route by its ability to spread from cell to cell in the presence of polyclonal and monoclonal nAbs targeting HCV envelope glycoproteins in the surrounding milieu (Brimacombe *et al.*, 2011; Timpe *et al.*, 2008; Witteveldt *et al.*, 2009).

The exact mechanism of HCV cell-to-cell transmission is unknown and is likely to differ in certain aspects from the entry mechanism of extracellular virus. However, so far it has been demonstrated that both entry mechanisms share the cellular entry factors CD81, SR-BI, claudin-1, occludin, NPC1L1, EGFR, and EphA2 (Barretto *et al.*, 2014; Brimacombe *et al.*, 2011; Lupberger *et al.*, 2011). SR-BI seems to play the most important role in this transmission mode since HCV entry inhibitors and mAbs targeting SRB-I have been shown to have the largest effect on blocking cell-to-cell transmission not only *in vitro* but also *in vivo* (Brimacombe *et al.*, 2011; Meuleman *et al.*, 2012).

In contrast, the role of CD81 is still controversial, since some studies revealed that at least a fraction of cell-to-cell transmission of HCV could occur in a CD81-independent manner (Jones *et al.*, 2010; Witteveldt *et al.*, 2009). The involvement of CD81 in cell-to-cell transmission is supported by the fact that inhibiting the interaction between CD81 and E2 also restrains cell-to-cell spread of HCV. For example, an anti-CD81 mAb blocks this cell-to-cell transmission in a dose-dependent manner (Fofana *et al.*, 2013). Furthermore, an anti-E2 camelid nanobody interfering with E2 binding to CD81 has also been shown to block cell-to-cell-transmission of HCV (Tarr *et al.*, 2013).

In chronic infections, cell-to-cell transmission may prove advantageous for the virus compared to the cell-free virus, given the fact that it is not accessible for neutralization by the nAbs present in the majority of chronically infected patients. Direct HCV cell-to-cell spread may pose limitations on antibody-based therapeutics. Therefore, development of nAbs capable to neutralize both cell-free virus and cell-to-cell transmission would be desirable. Such antibody-based therapeutics would be especially relevant in prevention of re-infection of

the liver allograph in chronic HCV patients.

Interfering non-neutralizing antibodies and other immune evasion mechanisms

Both nAbs and non-nAbs are detected in the serum of HCV infected patients. As described earlier, non-nAbs targeting epitope II were suggested to disrupt the neutralizing activity of nAbs directed against epitope I (Duan *et al.*, 2012; Zhang *et al.*, 2007; Zhang *et al.*, 2009). However, this area still needs further investigation as different studies have revealed contradictory results (Keck *et al.*, 2013; Tarr *et al.*, 2012).

In addition, HCV was proposed to infect B lymphocytes and induce hypermutations of heavy-chain immunoglobulins, which consequently can lower the affinity of nAbs and negatively affect their neutralizing activity (Machida *et al.*, 2008). Also, exosome-mediated transmission of HCV has recently been identified as another immune evasion strategy. Exosomes derived from HCV infected cells were demonstrated to contain viral particles and able to transmit infection in the presence of nAbs.

3D structures of the main neutralizing epitopes

A 3D structure of conserved epitopes targeted by nAbs can provide important information about the neutralization mechanisms employed by specific nAbs and assist in the development of efficient vaccines or other therapeutics. However, until very recently, a 3D model of HCV E2 was not available, making it difficult to understand the molecular basis of HCV neutralization by nAbs.

Antibodies recognizing linear epitopes, and in exceptional cases also conformational epitopes, often bind to synthetic peptides corresponding to the epitope sequence. Therefore, structural information about a linear antibody epitope can be obtained by co-crystallizing an antibody fragment in complex with a synthetic epitope peptide. In most cases, antibody fragments (Fab or scFv) are used, since they are more amenable to crystallization than full IgG molecules. This strategy can help to overcome difficulties in structure determination for an entire antigen-antibody complex, as is often the case for viral glycoproteins. Antibody fragment/peptide co-crystallization was employed to gain insight into the atomic structure of the epitopes for a number of proteins from different pathogens and in the case of HCV for the

structural characterization of epitopes I and II.

Crystal structure of epitope I

An epitope I peptide was co-crystallized with Fab fragments derived from the antibodies AP33, HCV1, 3/11, hu5B3.v3 and humanized AP33 (MRCT10.v362) (Kong *et al.*, 2012a; Pantua *et al.*, 2013; Potter *et al.*, 2012) (Meola A, 2014). The first reported crystal structure of epitope I was determined by co-crystallizing nAb HCV1 in complex with a peptide corresponding to aa 412-423 of E2 (Kong *et al.*, 2012b). The crystal structure revealed that the peptide adopts an extended β-hairpin conformation with a hydrophilic face exposed to solvent and a hydrophobic face interacting with the Fab. L413, N415, G418 and W420 were identified as the main contact residues of the peptide to the Fab and all of them were shown to be crucial for HCV1 mAb binding to E2. In addition, they are highly conserved among different HCV isolates. Normally it would be expected that the hydrophobic face of epitope I should be buried in the interface with the protein in context of the full-length folded E2. However, in that case it would be not accessible for the binding of the antibody due to steric clashes. The hydrophilic face of the β-hairpin harbors two N-linked glycans at N417 and N423 and, thus, should also be exposed. Therefore, epitope I was suggested to extend away from the folded core of the protein resulting in both sides of the β-hairpin being solvent exposed.

The Epitope I peptide was also crystallized with a Fab fragment derived from the mouse bnAb AP33 (Kong *et al.*, 2012a). The peptide co-crystallized with AP33 Fab adopted a very similar conformation to the one observed in complex with HCV1 Fab, although the CDRs of the antibodies are remarkably different (**Figure 10**). Moreover, the same residues of epitope I, i.e. L413, N415, G418 and W420, are buried at the interface with both antibodies and are crucial for antibody binding. The AP33 epitope has been suggested to be partially conformational based on the fact that the antibody binds weaker to denatured E2 than to intact E2 (Tarr *et al.*, 2006). Moreover, it was shown to cross-compete with conformation-sensitive nAbs HC-1, HC-11 and CBH-5, which interact with most of the critical binding residues within aa 523-535, indicating some degree of sterical hindrance between AP33 and those nAbs.

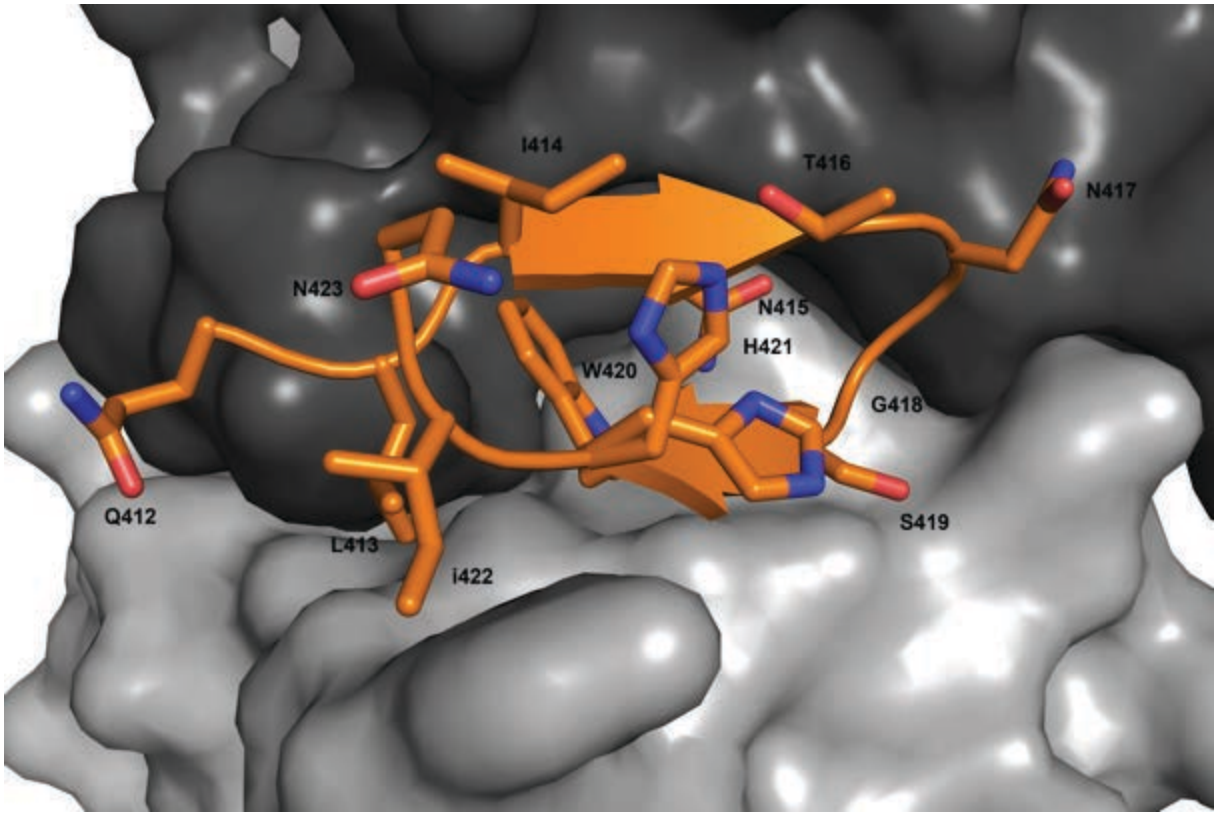


Figure 10. Epitope I peptide structure in complex with the Fab derived from the neutralizing antibody AP33. (PDB ID 4G6A). View on the paratope of Fab AP33 in complex with the epitope I peptide. The peptide is shown as cartoon with side chains as sticks and colored by atom type (orange, blue and red for carbon, nitrogen and oxygen, respectively. The molecular surface of the light chain and heavy chain are colored in light grey and dark grey, respectively.

Pantua et al. also reported crystal structures of an epitope I peptide in complex with Fab fragments derived from humanized AP33 antibody called MRCT10.v362 and another humanized mouse antibody hu5B3.v3, which superposed well with a β-hairpin conformation observed in complexes with HCV1 and AP33 Fabs (Pantua *et al.*, 2013).

Neither of the two currently reported crystal structures of the E2 core molecule encompasses epitope I, although in one case, this segment was present in the construct used for crystallization, suggesting that this region is highly flexible and mobile. Recently, the crystal structure of an epitope I peptide in complex with a Fab fragment derived from the rat nAb 3/11 mAb has been determined. In contrast to all the other reported structures, the epitope I peptide adopts a completely different extended conformation in this complex (Meola A, 2014). This is a strong indication that epitope I may exhibit intrinsic flexibility at the surface of infectious HCV particles.

Epitope II

Two groups determined the 3D structure of epitope II independently by co-crystallizing the corresponding peptide (E2 aa 434-446) in complex with nAbs (Deng *et al.*, 2013; Krey *et al.*, 2013) (Deng *et al.*, 2014). In the complexes with the Fabs from human nAbs HC84-1 and HC84-27, the peptide adopted a short α-helical turn encompassing aa 437-442 with aa 443-446 present in an extended conformation (Krey *et al.*, 2013) (**Figure 11**). The crystal structures revealed that L441, F442, Y443, and K446 form extensive interactions with the nAbs, which is in agreement with alanine scanning mutagenesis results. In addition, NAb HC84-27 was previously described to utilize W616 as an additional contact residue. Interestingly, some extra electron density in the HC84-27/epitope II peptide complex was observed that could not be attributed to epitope II, but was proposed to provide a glance on the 3D arrangement of these two parts of the epitope.

The aromatic side chains of F442 and Y443, as well as the aliphatic side chains of L438, A439 and L441, are all on one side of the epitope II α-helix. The residues L441, F442 and Y443 have also been previously described to be crucial for CD81 binding (Drummer *et al.*, 2006). The structures of the epitope II peptide show that these three residues form a hydrophobic protrusion, which might interact with a hydrophobic surface patch of CD81. In addition, F442 and Y443 are in close proximity with glycan N4 (aa residue N448), which is in line with this glycan shielding the CD81 binding site. L441 and Y443 are highly conserved among epitope II sequences from different HCV isolates, while F442 is conserved in only ~60% of the sequences. Notably, in the other 40% of sequences, F442 is substituted by other large hydrophobic residues, which most likely retain surface complementarity between the hydrophobic protrusion of epitope II and the CD81 binding patch. Involvement of the critical CD81 binding residues L441, Y442 and F443 in the interaction with the nAbs HC84-1 and HC84-27 may explain, why no neutralization escape has been observed for those nAbs.

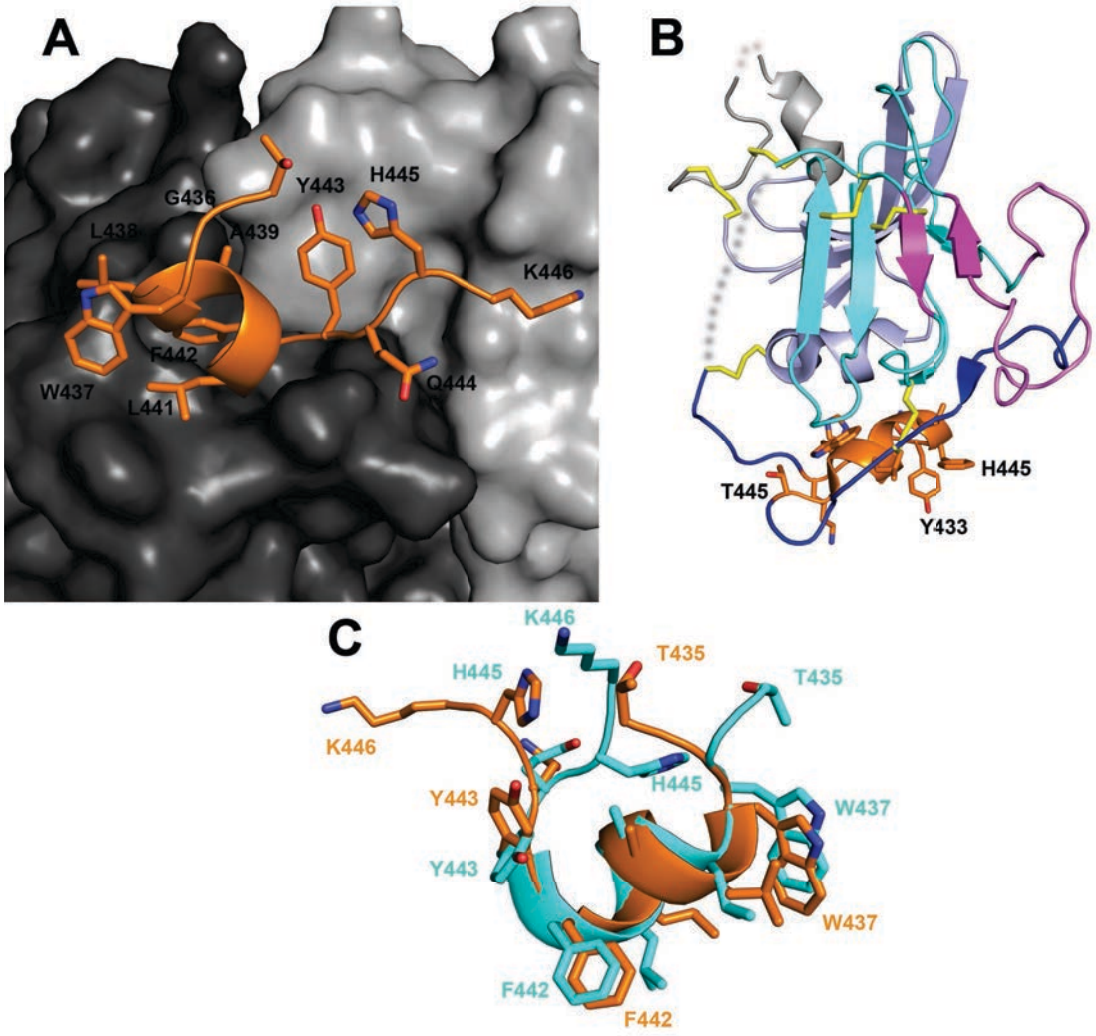


Figure 11. Epitope II peptide structure in complex with the HC84-1 Fab (PDB ID 4JZN). (A) View on the paratope of the epitope II / Fab HC84-1 complex. The peptide is shown as cartoon with side chains as sticks and colored by atom type (orange, blue and red for carbon, nitrogen and oxygen, respectively). The molecular surface of the light and heavy chains are colored in light and dark grey, respectively). **(B)** The conformation of epitope II adopted in crystal structure of E2 core. The epitope II is shown in orange, with the rest of E2 core colored as described in **Figure 7**. Disulphides are displayed as yellow sticks. (C) Superimposition of epitope II from HC84-1/epitope II and E2 core crystal structures colored in orange and cyan, respectively.

In parallel, Deng et al. obtained the crystal structure of the epitope II peptide (aa 430-446) in complex with Fab #8, derived from the murine genotype 1a specific nAb #8 (Deng *et al.*, 2013). The crystal structure revealed a similar conformation to the one observed in complex with nAbs HC84-1 and HC84-27, with the N-terminal part of the peptide in an extended conformation and the C-terminal part folding into a 1.5-turn α-helix. The binding of nAb #8 to the epitope II peptide is mostly dependent on hydrophobic residues (W437 and L438), which are different to the ones involved in HC84 interaction, providing a possible explanation for the restriction of the neutralization activity of nAb #8 to genotype 1a.

Recently, a crystal structure of epitope II in complex with a Fab derived from non-nAb #12, which is nearly identical to the Fab #8/epitope II structure, was also reported (Deng *et al.*, 2014). Both mAb #8 and mAb #12 use the residues W437 and L438 for the interaction with epitope II peptide. The main difference between the two structures is that the N-terminal segment of epitope II (aa 434–436) adopts different conformations depending on the antibody, turning at a flexible G436 residue that acts as hinge.

The epitope II peptide in complex with the HC84 antibodies adopts a similar conformation as in the crystal structure of the E2 core (Kong *et al.*, 2013), where epitope II also forms an α-helix at aa 437-442 designated α1. As depicted in **Figure 11 (C),** the helical parts of epitope II from both structures are superimposable. α1 helix is a part of the E2 core front layer and has been suggested to be a portion of CD81 binding site together. In the E2 core structure, the conserved residues L441, F442 and Y443 previously identified to be crucial for E2-CD81 interaction form a solvent exposed hydrophobic cluster, which strongly suggests that this structural feature of the E2 surface participates in interaction with CD81 (Drummer *et al.*, 2006). As these residues are solvent exposed they are also accessible for binding of HC84 antibodies.

The helical parts of the epitope II structures obtained in complex with the Fab fragments derived from mAbs #8 and #12 also superimpose well with α1 helix. As mentioned above, the crystal structures of the Fabs derived from nAb #8 and non-nAb #12 in complex with epitope II peptide reveal that the residues W437 and L438 are crucial for the interaction between those mAbs and epitope II peptide. This implies that W437 and L438 residues have to be accessible on E2 surface in order for nAb #8 non-nAb #12 to bind to the glycoprotein. This is in disagreement with these residues being on the side of α1 helix pointing towards the E2 hydrophobic core. As a result, those mAbs would have severe steric hindrance with the rest of the protein for binding to α1 helix in this orientation.

Since α1 helix packs relatively loosely against the major part of E2, it is possible that this region within the protein may undergo local conformational changes. Deng *et al*. suggest that epitope II can exist in an open and closed state in the context of the E2 glycoprotein, inducing antibodies against both conformations. MAbs #8 and #12 binds to epitope II in an open conformation in which the residues W437 and L438 are solvent exposed and, thus, available for binding of those mAbs. In contrast, mAbs AR3C and HC84 recognize epitope II in its closed conformation, which is likely to be closer to the one interacting with CD81.

E2 region spanning aa 523-540

A number of amino acid residues critical for CD81 binding (Y527, W529, G530 and D535) are located within the E2 region comprising aa 523-540 (Owsianka *et al.*, 2006). Although this region is targeted by the majority of human bnAbs, the strategy of co-crystallization epitope peptide with any of the known bnAbs was not applied due to the conformation-sensitive nature of those nAbs. Therefore, structural information on this important antigenic region was not available until the recently reported E2 core crystal structure in complex with neutralizing AR3C.

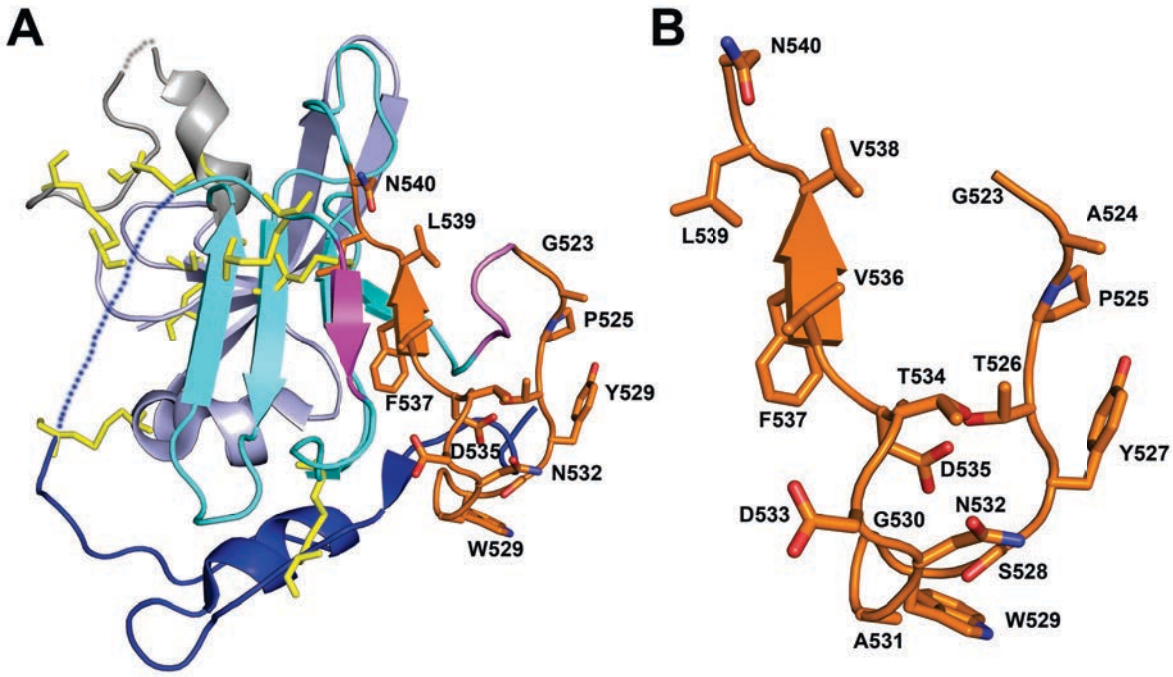


Figure 12. Structure of E2 antigenic region spanning aa 523-540. (A) The conformation of aa 523-540 adopted in crystal structure of E2 core. The region comprising aa 523-540 is shown as cartoon with side chains as sticks and colored by atom type (orange, blue and red for carbon, nitrogen and oxygen, respectively). The rest of E2 core structure is colored as described in Figure 7. Disulphides are displayed as sticks and colored in yellow. (B) A closer view of the region comprising aa 523-540 in E2 core crystal structure.

The crystal structure of the E2 core reveals that aa 523-535 form a big, bi-lobed surface loop further extending as a short β-strand (aa 536-538) (**Figure 12**). This β-strand is part of an Ig-like domain forming the inner core of the protein. The last residues of this region (aa 539-540) adopt an extended conformation. Y527, G530 and D535, which have been implicated in CD81 binding, are solvent exposed and, as a result, would be available for the interaction with the receptor.

AR3C Fab binds to a large part of the front layer of the E2 (aa 421-446) (**Figure 13**). In addition, AR3C Fab interacts with residues S528, W529 and A531 residing within the CD81 binding loop as well as a few other residues (P515 and Y613). Unexpectedly, the majority of the residues previously shown to be essential for nAb AR3C binding by alanine scanning mutagenesis of E2 (S424, P525, G530 and D535) do not participate in direct interaction between the antibody and the glycoprotein. However, they might be important for the correct fold of the E2 regions compromising the AR3C epitope.

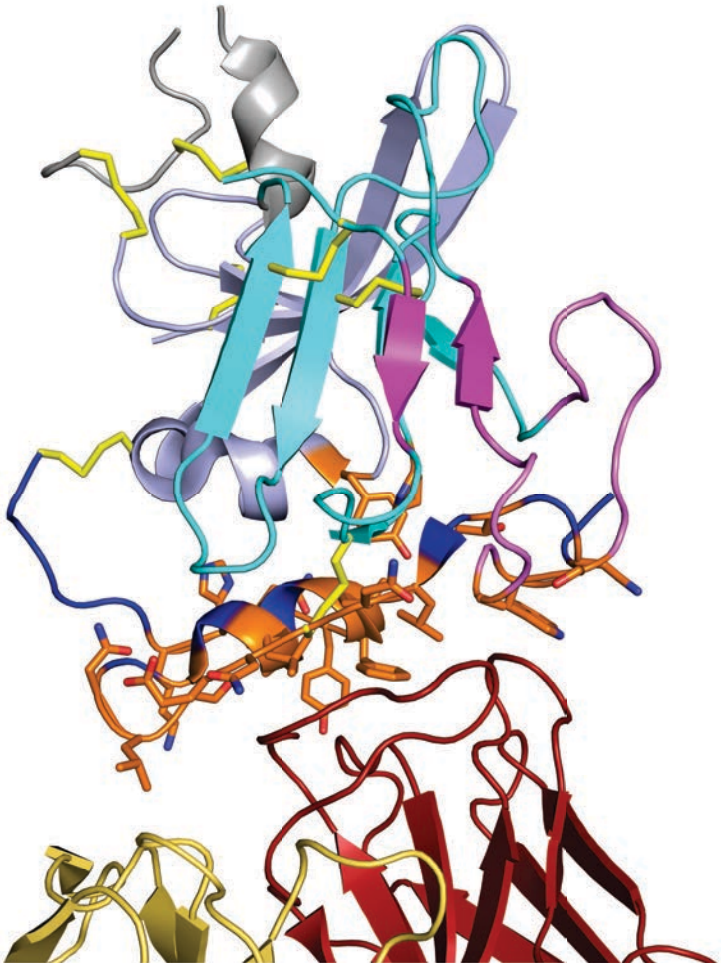


Figure 13. HCV E2 core interaction with Fab AR3C. Both AR3C Fab and E2 core are displayed as cartoon. E2 core is colored by structural components (see Figure 7) and the Fab heavy and light chains are colored in dark red and yellow, respectively. The AR3C epitope on the E2 core protein is shown as cartoon with side chains as sticks and colored by atom type (orange, blue and red for carbon, nitrogen and oxygen, respectively).

The heavy chain of AR3C dominates binding to E2 and accounts for ~86% of the buried surface area in the E2-AR3C interface. Importantly, the buried surface area of the AR3C epitope encompasses almost exclusively residues that are 80-100% conserved in all HCV genotypes and also covers a number of the residues (G536, L438, L441, F442 and

W529) critical for CD81 binding as established by site-directed mutagenesis. Thus, the broad neutralization of mAb AR3C can be explained by an overlapping epitope that directly competes for contact residues with CD81 in epitopes I and II, and sterically blocks access to the CD81 binding loop.

Therapeutic potential of neutralizing antibodies

Despite the recent advances in HCV treatment, vaccination would still represent the best way to reduce the global burden of HCV. However, to date no licensed vaccine exists against this virus. Most of the successful vaccines are based primarily on the induction of potent nAb responses (Lambert *et al.*, 2005) and to date a number of studies have demonstrated that nAbs are capable of controlling HCV infection (as described above). This is encouraging for the potential development of at least a partially anti-HCV vaccine ideally capable of inducing long-term B-cell and T-cell memory responses. Unfortunately, a number of difficulties, such as the emergence of neutralization escape variants, HCV cell-to-cell transmission, and the ability of HCV to re-infect previously exposed persons, makes the development of an efficient HCV vaccine a major challenge.

The traditional approach of using inactivated viruses for vaccination against HCV has received little attention due to the difficulties in producing large quantities of infectious HCV particles in cell culture. Akazawa *et al.*, who used the inactivated genotype 2 (J6/JFH-1) HCVcc to immunize mice, demonstrated the feasibility of an inactivated whole virus vaccine. Induced NAbs were able to neutralize genotypes 1a, 1b and 2a *in vitro* and could protect human liver chimeric uPA-SCID mice from experimental challenge with the lowest virus dose. Interestingly, immunization of mice with inactivated HCVcc resulted in more efficiently neutralizing serum than vaccination with recombinant E1 and E2, implying that HCV particles may be more immunogenic than just the individual recombinant envelope proteins. In addition, another study demonstrated that VLPs pseudotyped with HCV envelope glycoproteins induced high titer of bnAbs in mice and macaques (Garrone *et al.*, 2011).

The only anti-HCV vaccine tested in humans was developed by Chiron Corporation (now Novartis, Basel, Switzerland). It used as immunogen a heterodimer of recombinant E1 and E2 produced in mammalian cells. In initial trials this vaccine induced an anti-E2 antibody response in chimpanzees, which was protective against challenge with homologous virus (Choo *et al.*, 1994). However, the challenge with a heterologous virus strain protected only

one chimpanzee of nine, indicating that in most vaccinated animals the induced immune response was not sterilizing (Houghton & Abrignani, 2005). A recombinant vaccine containing genotype 1a glycoproteins E1/E2 combined with an oil–water adjuvant has also been assessed in human volunteers. The initial results revealed that approximately half of the vaccinated people had antibodies against HVR1 and some also possessed antibodies against epitope I, and epitope II as well as the E1 region 313-327.

Moreover, vaccinees did not experience any significant adverse events (Frey *et al.*, 2010; Ray *et al.*, 2010). In a follow-up study, the serum of one out of sixteen persons vaccinated with a single HCV strain of genotype 1a possessed cross-neutralizing activity of all seven HCV genotypes (Law *et al.*, 2013). These results prove that bnAbs can be induced by recombinant E1/E2 vaccines at least in some individuals. However, a number of difficulties associated with the development of an efficient vaccine, such as relatively low titers of bnAbs and average levels of neutralization, need to be addressed. One problem is that in the case of immunization with unmodified E2 or E1E2, HVR1 is immunodominant, implying that engineered immunogens lacking HVR1 may be needed. Finally, due to the multiple escape mechanisms of HCV from the immune response, targeting a single viral component is not likely to be sufficient. As a result, vaccine development efforts should focus on designing immunogens that are capable of inducing nAbs targeting several epitopes on the envelope proteins. In addition, it would be interesting to better analyze the nature and specificity of nAbs during the acute phase of self-resolving HCV infection, since those antibodies are likely to have the highest protective potency. One study has revealed the higher prevalence of a mAbs recognizing discontinuous epitopes (E1 (aa 297-306)-E2 (aa 480-494)-E2 (aa613-621)) in patients who spontaneously cleared infection (Ndongo *et al.*, 2010).

Hepatocellular carcinoma and liver cirrhosis are the common complications of chronic HCV infection and require liver transplantation as the ultimate treatment. Unfortunately, reinfection of the liver allograft is almost inevitable, which commonly results in rapid progression to cirrhosis, allograft failure and death (Crespo *et al.*, 2012; Gallegos-Orozco *et al.*, 2009). Serum HCV RNA levels decrease after removal of the infected liver, and the circulating virus infects the transplanted organ within a few days (Powers *et al.*, 2006). No efficient and well-tolerated prophylaxis for HCV infection of the graft exists. Therefore, antibody-based therapy before and immediately after transplantation would be greatly beneficial. When applied before transplantation it would reduce the circulating virus, which would lessen the possibility of re-infection of the transplanted donor liver. Administration of

the anti-HCV antibodies after liver transplantation would aim to completely eradicate HCV before the re-establishment of the chronic disease within the allograft.

MBL-HCV1 (Medarex, Inc., a subsidiary of Bristol-Myers Squibb) is a humanized version of mAb HCV1 recognizing epitope I (aa 412-423). Previously, MBL-HCV1 had been shown to prevent HCV infection and to suppress the viral load in chronically infected chimpanzees (Morin *et al.*, 2012). The original murine HCV1 has been crystallized with its epitope peptide (see above) (Kong *et al.*, 2012b). It is the only mAb so far evaluated in a phase II clinical trial to demonstrate an ability to prevent HCV re-infection of an allograft after liver transplantation. This mAb significantly reduced viral load in all 6 treated patients for a period ranging from 1 to 4 weeks, and delayed time to HCV rebound compared with the group receiving a placebo (Chung *et al.*, 2013).

The rebound of HCV infection correlated with the appearance of resistant viral variants indicating that monotherapy is not sufficient to completely clear the virus. MBL-HCV1 resistant variants contained substitution at position 415 (N415D, N415K, N415S) or 417 (N417S) of the epitope I (Babcock *et al.*, 2014). The available crystal structures of epitope I with HCV1 and AP33 Fabs indicate that N415 stabilizes the β-hairpin structure by forming a hydrogen bond with G418 (Kong *et al.*, 2012a; Potter *et al.*, 2012). In addition, Pantua and colleagues observed that N417S substitution results in the shift of glycosylation site N-X-S from N417 to N415, which in turn blocks the neutralization by anti-epitope I nAbs (Pantua *et al.*, 2013). Thus, mutation at positions N415 or N417 are likely responsible for the observed viral escape from mAb MBL-HCV1 by disrupting the conformation of the β-hairpin. Interestingly, both mutations were never detected in the same virus (Babcock *et al.*, 2014). The mutations at N415 and N417 have been also observed to interfere with neutralization activity of other mAbs targeting epitope I (Gal-Tanamy *et al.*, 2008; Keck *et al.*, 2014). Moreover, the glycosylation site shift has been also observed to occur spontaneously when passaging HCVcc in the absence of selection by nAbs targeting this region (Keck *et al.*, 2011; Keck *et al.*, 2012). The combination of MBL-HCV1 with HCV NS3 protease inhibitor telaprevir was demonstrated to suppress the appearance of resistant viral variants to both agents *in vitro*. Such a therapeutic regime, including a second direct-acting antiviral, can potentially be applied in HCV patients undergoing liver transplantation and will possibly allow to reduce the required antibody dose administered to patients, which in the presented study was relatively high (50 mg/kg).

Immune responses to GBV-B virus

Studies of immune responses against GBV-B infection in tamarins and marmosets have revealed that animals with resolved infection usually are not completely protected against re-infection (Beames *et al.*, 2000; Bright *et al.*, 2004; Bukh *et al.*, 2008). Tamarins re-challenged with GBV-B resolved the infection earlier and had lower peak viral titers compared with the animals with the primary infection. Sterilizing immunity was observed only in one tamarin. In addition, re-challenge did not progress to hepatitis in most animals (Bukh *et al.*, 2008). Similar results were observed in marmosets. Viral clearance in these animals correlated with the appearance of virus-specific T-cell responses mostly directed against NS3 and NS4A (Woollard *et al.*, 2008). nAbs were not detected in marmosets neither at the time of viral clearance nor a few weeks after clearance (although the possibility that it is related with the lack of a good detection system exists). In this study, marmosets had a protective immunity against re-infection with GBV-B when they were re-challenged several months after resolving the primary infection, which was associated with the increased T-cell responses (Woollard *et al.*, 2008). These studies might lead to the development of a small-animal model for HCV, although more work is required to evaluate the relevance for testing vaccine strategies against HCV.

I. Structural Characterization of the GB virus B Envelope Glycoprotein E2

Background

My thesis project was initially focused on structural studies of the envelope glycoprotein E2 of GB virus B (GBV-B), which infects tamarins and is a close relative of the hepatitis C virus (HCV). For many years the crystallization of HCV glycoproteins, which mainly focused on E2, has been a serious challenge. The crystal structure of a core fragment of HCV E2 was determined just recently (Khan *et al.*, 2014; Kong *et al.*, 2013), showing that is has a fold that does not resemble a viral fusion protein. In particular, it does not display a "class II fusion" fold, as predicted by comparison to other members of the *Flaviviridae* family, in the flavivirus genus.

Because the envelope proteins of GBV-B are less extensively glycosylated than those of HCV, we expected that it would provide an opportunity to carry out structural studies that would be more straightforward than the study of the HCV glycoproteins. In addition, a comparative analysis would provide insight to better understand their common function. The crystallization of GBV-B E2 turned out to be as – or perhaps more – challenging than the studies of HCV E2. Although diffracting crystals were not obtained and we were therefore not able to determine the 3D structure, we were able to show that the recombinant soluble fragment of GBV-B E2 can inhibit infection by GBV-B. Therefore, I will present first the data showing the inhibition of infection by the soluble GBV-B E2 fragment and then, for interested readers, I will describe in detail the efforts to crystallize the GBV-B E2 ectodomain.

Objectives

The objective of this study was to determine the structure of the GBV-B envelope glycoprotein E2 by X-ray crystallography. The structural characterization of GBV-B E2 was expected to help with the identification of common elements between GBV-B and HCV that could provide important clues in terms of both evolution and of function.

Results

Infection inhibition by the recombinant GBV-B E2e

A number of different constructs of the recombinant GBVB E2 ectodomain (E2e) were produced during the study. However, all of the mAbs we obtained from the immunization of mice with the recombinant GBV-B E2e appeared to recognize linear epitopes, as they reacted with E2 in Western blot, thereby making it difficult to verify the correct conformation of our recombinant GBV-B E2e constructs. Such verification can be done by measuring the recombinant GBV-B E2e ability to inhibit infection of primary hepatocytes from marmosets, which likely occurs by competition with infectious GB virus B for entry receptors. We therefore initiated a collaboration with the laboratory of Annette Martin at Institut Pasteur, who studies GBV-B entry and replication. We tried to inhibit infection with GBV-B $E2_{350\text{-}581}$, which represents the full-length ectodomain of GBV-B E2, and a shorter variant of GBV-B ectodomain, $E2_{350\text{-}540}$ **(Figure 14).**

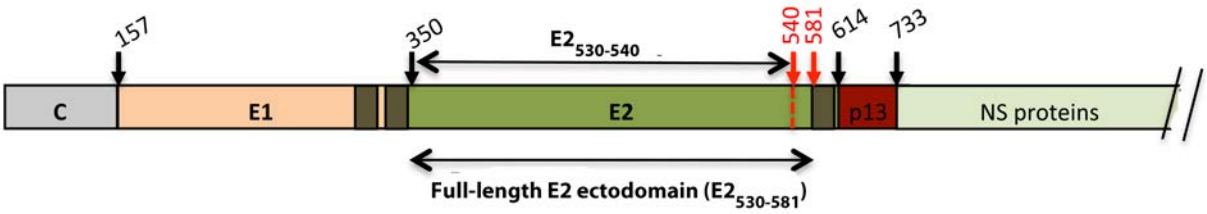


Figure 14. The diagram showing the organization of GBV-B polyprotein indicating GBV-B E2 full-length ectodomain ($E2_{530\text{-}581}$) and the shorter variant of the E2 ectodomain ($E2_{530\text{-}540}$) used in the experiments to inhibit GBV-B infection of primary hepatocytes from marmosets. The transmembrane domains are depicted as grey boxes. The expression system of those constructs is described in detail in the following section.

As a control, we tested in parallel the effect of the ectodomain of glycoprotein E2 from the bovine viral diarrhea virus (BVDV), which belongs to the *Flavivirus* genus of the *Flaviviridae*, produced under identical conditions. BVDV E2 has no detectable sequence similarity with GBVB E2. In contrast to the control protein, GBV-B $E2_{350\text{-}581}$ exerted a clear inhibition of the infection **(Figure 15**). However, the shorter variant of GBV-B E2e ($E2_{350\text{-}540}$) was unable to block the infection.

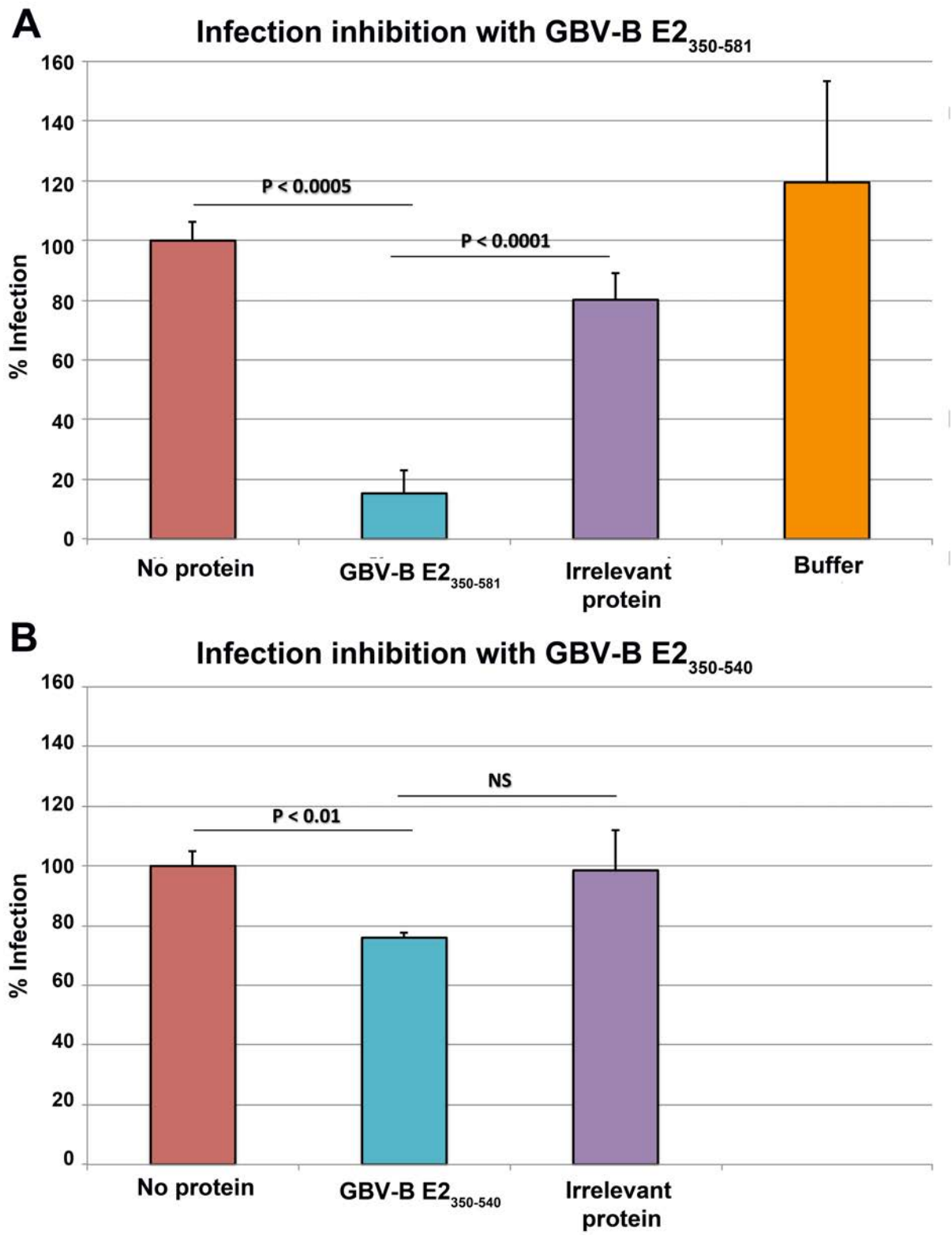


Figure 15. Inhibition of infection of primary hepatocytes from marmosets with GBV-B $E2_{350-581}$. (A) and GBV-B $E2_{350-540}$ (B). Controls: 1) no protein added 2) BVDV E2 (irrelevant protein) added 3) only the buffer of the proteins added. The experiment was performed by Caroline Marnata, a former PhD student in Annette Martin's laboratory.

These results indicated that at least the recombinant full-length GBV-B E2e adopts a conformation that is similar to the one present in virions. Also, it suggested that a receptor-binding site, or segments that contribute to this site, might reside within the last 41 GBV-B E2e amino acid residues.

Production of a soluble GBV-B E2 ectodomain

In order to perform crystallization trials, a large quantity of soluble material (in this case, the GBV-2 ectodomain) is required. Therefore, we first undertook to develop an expression system for the soluble ectodomain of GBV-B E2 (GBV-B E2e) based on the Drosophila Expression System, which can produce sufficient amounts of the protein secreted into the supernatant from the corresponding stable *Drosophila* S2 cell line. The expression construct for the soluble ectodomain of GBV-B E2 contained the E1-E2ΔTM portion of the genome, codon-optimized (for *Drosophila melanogaster*) in a synthetic DNA. The absence of the transmembrane (TM) segment in E2 allowed secretion of its ectodomain after folding in the presence of E1, which was proposed to work as a chaperone. The expression vector also contained a double Strep-Tag allowing a simple two-step purification strategy, downstream from an enterokinase cleavage site, which enabled the proteolytic cleavage of the tag for crystallization purposes. Initially GBV-B E2e was truncated at the amino acid residue K581 (GBV-B polyprotein numbering) (UniProtKB accession number NP_056931) eliminating the predicted TM domain (residues 582-605) and the C-terminal cytoplasmic tail domain present in the full-length protein. Later, however, another construct shorter by 17 amino acid residues upstream from the predicted TM domain of E2 (i.e. truncated at Q564) was chosen for crystallization because it was expressing better (**Figure 2**). Further in the work this construct is referred as GBV-B E2e.

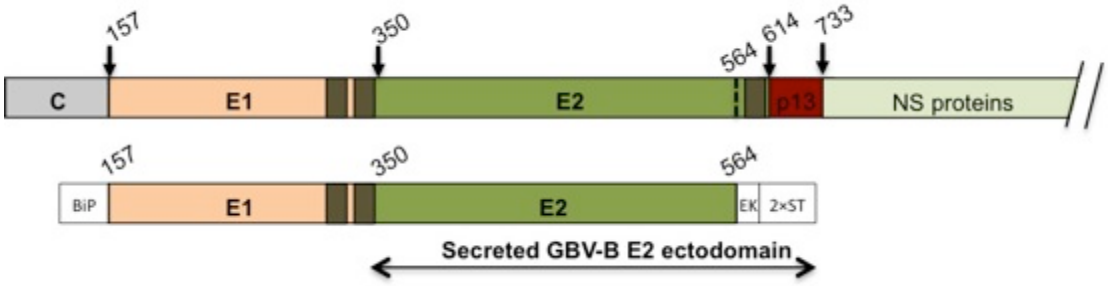
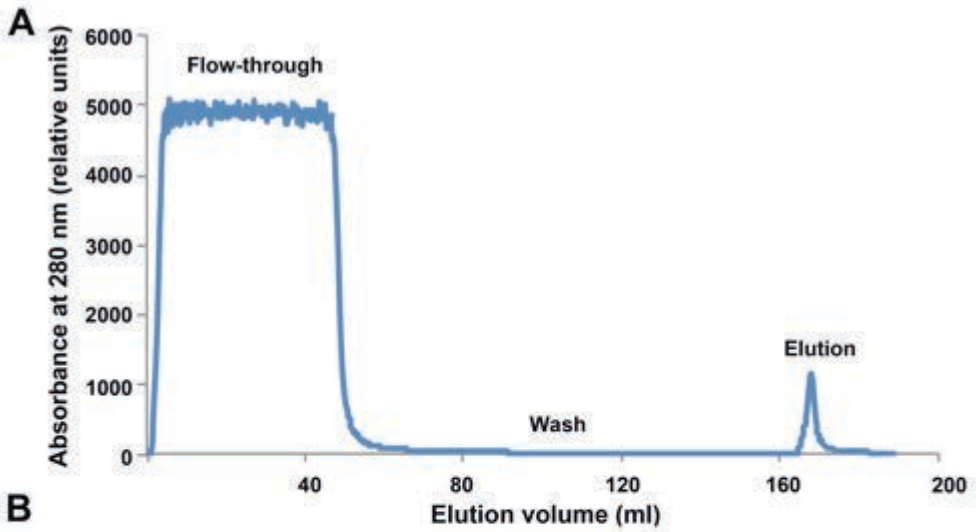
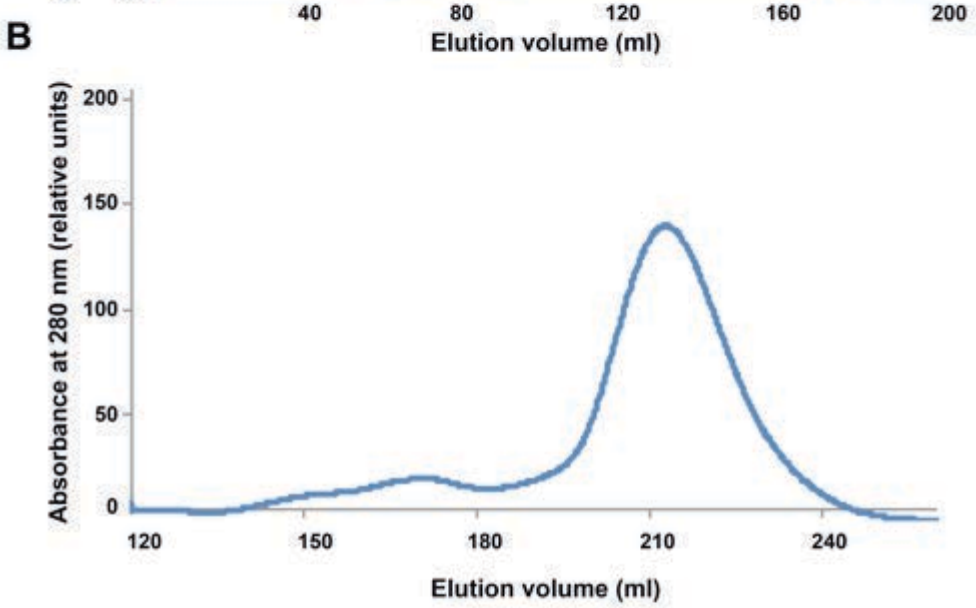


Figure 16. GBV-B E2 ectodomain (E2e) construct. The top diagram shows the organization of GBV-B polyprotein. The transmembrane domains are depicted as grey boxes. The bottom diagram shows the construct inserted in the plasmid for production in *Drosophila* S2 cells with BiP signal sequence in frame and an

enterokinase cleavage site (EK) followed by a double Strep-tag (2×ST). The numbers corresponds to the polyprotein numbering, position 350 being E2 residue 1. This construct was designed by analogy to constructs used in our laboratory to produce the E protein from the dengue virus, which has first the chaperone glycoprotein prM followed by the fusion protein E. This was done before it was known that hepaciviruses do not have class II fusion proteins.

This system normally yielded about 400 µg of purified, monomeric GBV-B E2e per litre of cell culture. Representative protein elution profiles from Streptactin affinity purification and size exclusion chromatography (SEC) columns are displayed in **Figure 17**. Although we carried out extensive crystallization trials with this protein, we were not successful in obtaining crystals suitable for structural studies.


A
Absorbance at 280 nm (relative units)
6000
5000
4000
3000
2000
1000
0
Flow-through
Wash
Elution
40
80
120
160
200
Elution volume (ml)


B
Absorbance at 280 nm (relative units)
200
150
100
50
0
120
150
180
210
240
Elution volume (ml)

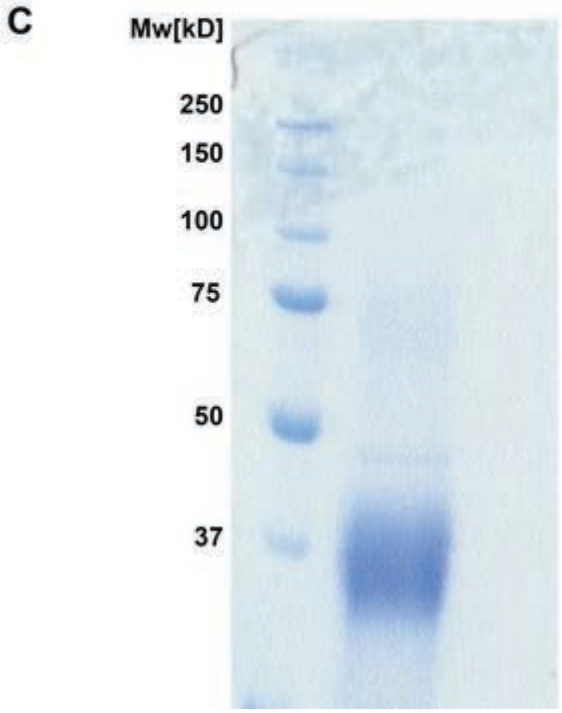

C
Mw[kD]
250
150
100
75
50
37

Figure 17. Purification of GBV-B E2e. (A and B) GBV-B E2e is purified by affinity chromatography using a Streptactin Superflow column and further separated by size exclusion chromatography (SEC) using HiLoad 26/60 Superdex 200 column (GE Healthcare Life Sciences). Protein elution was monitored by absorbance at 280 nm (blue curves). (C) The purity of the pooled and concentrated fractions of SEC was determined by SDS-PAGE under non-reducing conditions followed by Coomassie staining.

Crystallization trials of GBV-B E2e with antibody fragments

Multi-domain, heavily glycosylated viral envelope proteins are often difficult to crystallize, and different strategies can be applied to achieve crystallization in these cases. One of these strategies is the use of crystallization "chaperones", which are ligands that promote the formation of a crystal lattice when assayed in complex with the protein being studied. We decided to use this strategy and so, to obtain specific ligands for GBV-B E2e to perform co-crystallization experiments.

For this purpose, we immunized mice with the recombinant ectodomain and obtained a number of monoclonal antibodies (mAbs) specifically recognizing the glycoprotein. The main characteristics of these mAbs are summarized in **Table 2**. I determined the antibody sequence from the hybridoma cDNA and cloned the variable domains into a plasmid for production of recombinant antibody fragments (Fabs and single chain variable fragments (scFvs)) in *Drosophila* S2 cells. This enabled me to produce and purify large quantities of these ligands. Competition analysis of those antibody fragments by SEC revealed that three (E19.4, F7.7 and F16.1) out of five mAbs cross-compete for binding to the same antigenic region on GBV-B E2 (**Figure 18**). The mAb F24.3 had a lower affinity for GBV-B E2e, rendering it more difficult to determine cross-competition patterns with other antibodies by SEC. Two of the mAbs (C23.21 and D18.6) were found to bind the affinity purification tag (double *Strep*-tag® II) (http://www.iba-lifesciences.com/strep-tag.html), which had not been proteolytically removed from the protein used for immunizations. A stable monoclonal HEK293T cell line expressing the Fab derived from C23.21 mAb was established, enabling the production of this ligand, and which is used for very specific recognition of the Strep-Tag on many different proteins.

Table 2. Characteristics of the monoclonal antibodies (mAbs) against GBV-B E2e. In total 8 mAbs were obtained after the immunization of mice. Two of the mAbs recognized the double Strep-tag fused to GBV-B E2e. The other two were of an isotype IgM and although the Fab derived from the G8.3 mAb sequence was produced, binding of this Fab to GBV-B E2e was detected neither by SEC nor by ELISA (the antibody production platform of the Institut Pasteur). The remaining four mAbs specifically recognized E2e. The

corresponding Fabs and scFvs derived from the sequences of those mAbs were produced for co-crystallization trials. However, three of them were cross-competing indicating that they bind to the same antigenic region of E2e. The affinity of the mAbs was measured by ELISA by Farida Nato (antibody production platform of Institut Pasteur).

Antibody	**Isotype**	**Affinity**	**Antigenic region of E2e**	**Ligands produced**
C23.21	IgG1, κ	1.4×10^{-10}	Strep-tag	Fab
D18.6	IgG1, κ	1.6×10^{-8}	Strep-tag	-
E19.4	IgG1, κ	1.2×10^{-8}	A	Fab and scFv
F7.7	IgG1, κ	2.3×10^{-8}	A	Fab and scFv
F16.1	IgG1, κ	1.0×10^{-8}	A	Fab and scFv
F24.3	IgG1, κ	1.03×10^{-7}	B	Fab and scFv
G8.3	IgM, κ	8.2×10^{-9}	?	-
A18.6	IgM, κ	2.48×10^{-8}	?	-

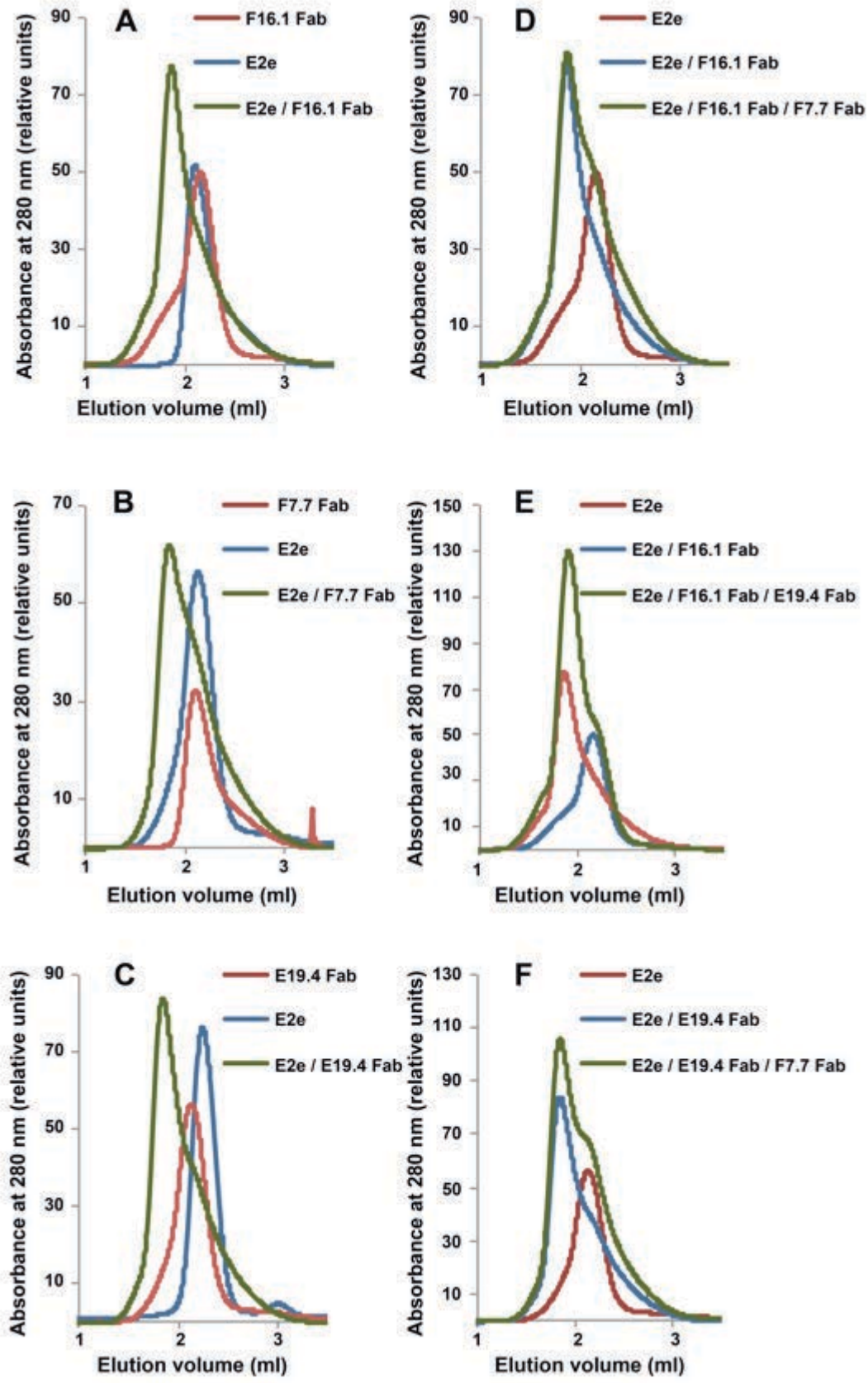


Figure 18. Complex formations between GBV-B E2e and Fab fragments derived from different mAbs. (A, B and C) E2e, the respective Fab fragment, and a mixture of the two pre-incubated overnight at 4°C (molar ratio 1:1) were loaded to a Sdx200 size exclusion column (in three different runs) (E2e ~ 40kD, Fab ~ 50kD, binary

complex ~ 90kD). Appearance of a peak at higher molecular weight indicated binary complex formation. (D, E and F). Cross-competition analysis of Fabs derived from conformation-sensitive mAbs. No ternary complex (140kD) was observed when two Fabs were pre-incubated together with E2e.

Notably, the three cross-competing mAbs shared an unusual, conserved cysteine residue in the complementarity-determining region of the heavy chain (CDR H2). In order to analyze the impact of the free cysteine, I crystallized one F16.1 Fab alone and refined this structure to ~2Å. The molecular surface representation of the Fab fragment revealed that the cysteine residue is entirely buried and most likely not involved in interaction with GBV-B E2 (**Figure 19**). However, we observed the appearance of covalent dimers of the antibody fragments over time upon storage at 4°C, due to reactivity of the free cysteine. The observed reactivity of this buried cysteine residue was likely due to the flexibility of CDR H2. Because this dimerization makes them unsuitable for co-crystallization trials, I mutated the free cysteine to serine. Control experiments confirmed that this mutation did not affect binding of the antibody fragments to GBV-B E2.

The co-crystallization trials of GBV-B E2e were performed with each of the three antibodies binding to the antigenic region A (F16.1, F7.7 and E19.4) both as Fabs and scFvs. I also tried to crystallize GBV-B E2e with the antibody fragments derived from mAb F24.3 binding to the antigenic region B alone or in combination with the antibody fragments targeting the antigenic region A because a single antibody fragment bound to the protein may not be enough to obtain diffraction quality crystals. Nevertheless, none of the co-crystallization trials of GBV-E2e in complex with antibody fragments yielded crystals. In this context, we performed a phage display to screen a synthetic scFv library based on human sequences (in collaboration with Jean-Luc Jestin in the laboratory) leading to the identification of three additional human scFvs (C11, C6 and C7) that interact with GBV-B E2. The selected scFvs were also expressed in *Drosophila melanogaster* S2 cells and purified for co-crystallization experiments with GBV-B E2. Unfortunately, the affinity of these scFvs was not high enough to detect their binding to GBV-B E2 by analytical SEC. In spite of extensive crystallization trials of GBV-B E2e in binary or ternary complexes with all of the above mentioned antibody fragments (the scFvs C11, C6 and C7 as well as Fabs and scFvs binding to the antigenic region A and B) in different combinations, we did not obtain diffraction quality crystals.

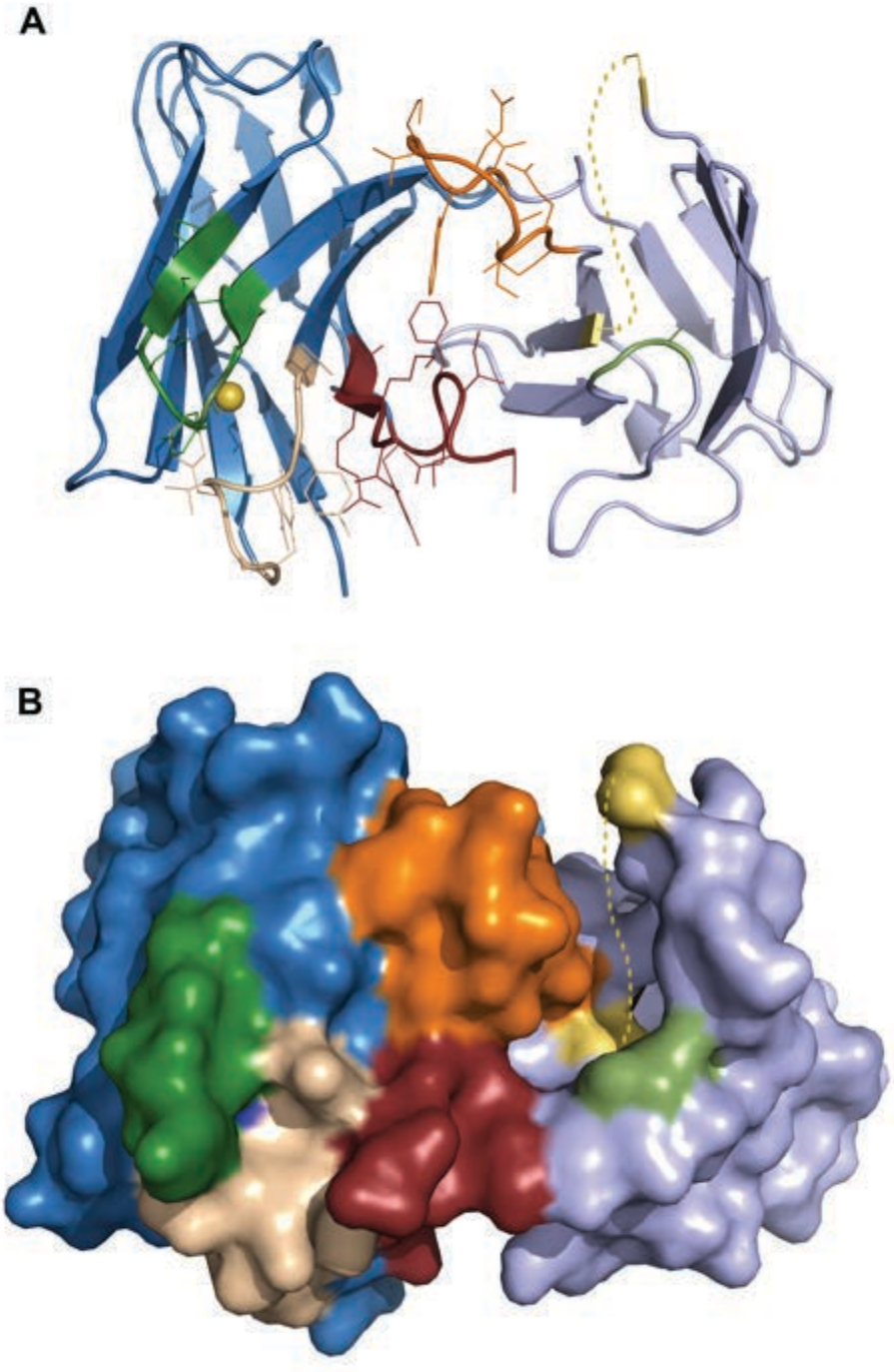


Figure 19. F16.1 Fab crystal structure. The crystal structure of Fab fragment F16.1 was refined to ~2Å. (A) A cartoon representation, and (B) the molecular surface of the variable region, both viewed from the top. The light and heavy chain are displayed in light and dark blue respectively. The CDR loops are colored in dark yellow (CDR-H1), dark green (CDR-H2), dark red (CDR-H3), light yellow (CDR-L1), light green (CDR-L2), and orange (CDR-L3). The CDR-L2 region showed a high degree of disorder and could therefore not be modeled. It is thus represented by a yellow dashed line. The side chains of the residues within the CDR loops are shown as lines, and the sulfur atom of the free cysteine residue in CDR-H2 is shown as sphere (left). The molecular surface representation indicates that this cysteine residue is entirely buried in a crystal structure.

Limited proteolysis of GBV-B E2e

In parallel, I carried out limited proteolysis experiments on GBV-B E2 in order to identify possible protease-resistant fragments that could be more amenable to crystallization than intact E2e. GBV-B E2e was subjected to limited proteolysis using trypsin and two resistant fragments were observed by SDS-PAGE (**Figure 20**).

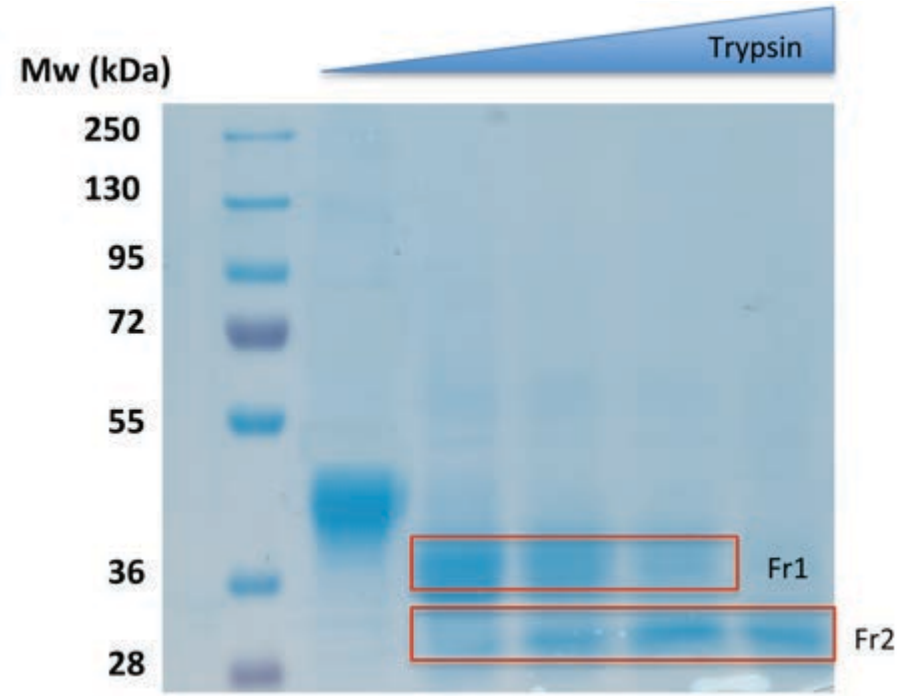


Figure 20. Limited proteolysis of GBV-B E2e with trypsin. The bar above indicates the increasing amounts of trypsin used in proteolysis of GBV-B E2e. Two degradation resistant fragments were observed by SDS-PAGE and were further characterized by N-terminal sequencing and SELDI-TOF (surface-enhanced laser desorption ionization time-of-flight) mass spectrometry. SDS-PAGE was performed under non-reducing conditions followed by Coomassie staining.

N-terminal sequencing analysis of the trypsin treated protein samples showed that both fragments retained the N-terminus of the intact protein, indicating that the protease had removed the C-terminal end of GBV-B E2.

I also performed limited proteolysis with trypsin for GBV-B E2 in a complex with the recombinant antibody fragments (F7.7 Fab and F24.3 scFv). Our assumption was that these ligands might protect the protease sensitive parts of GVB-B E2, which are cleaved off when the protein alone is subjected to limited proteolysis. However, we observed the same digestion pattern resulting in identical fragments as with the GBV-B E2 alone (**Figure 21**).

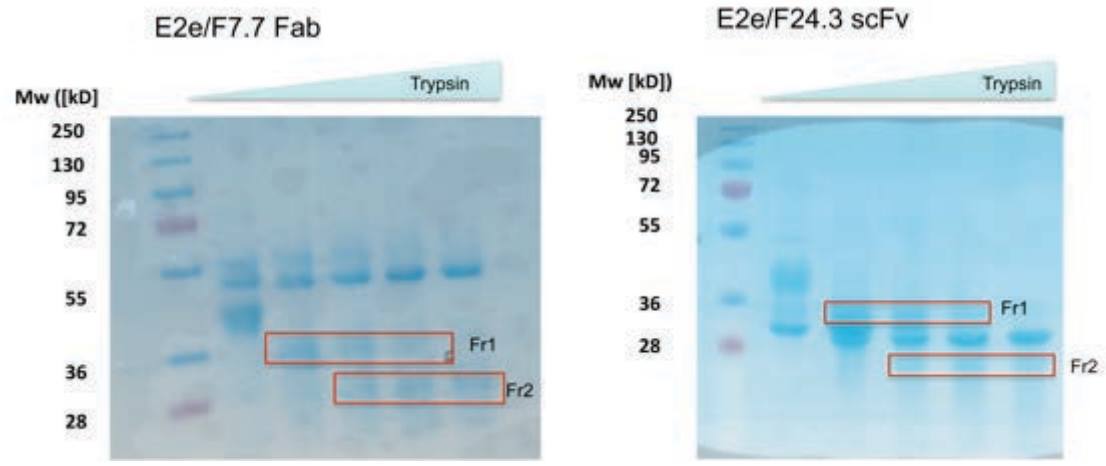


Figure 21. Limited proteolysis of E2e/F7.7 Fab (left panel) and E2e/F24.3 scFv complexes with trypsin. E2e complexes with F7.7 Fab or F24.3 scFv were subjected to limited proteolysis using trypsin. The bars above indicate the increasing amount of trypsin used. Two trypsin resistant fragments (framed by red boxes) were observed by SDS-PAGE and were further characterized by N-terminal sequencing, and SELDI-TOF mass spectrometry revealing that the fragments are identical to the ones obtained by the proteolysis of E2e alone. SDS-PAGE was performed under non-reducing conditions followed by Coomassie staining.

Based on proteomic analysis of these fragments and secondary structure predictions, we decided to clone and produce three different C-terminally truncated E2 species ($E2_{350\text{-}482}$, $E2_{350\text{-}540}$ and $E2_{350\text{-}513}$) **(Figure 22**).

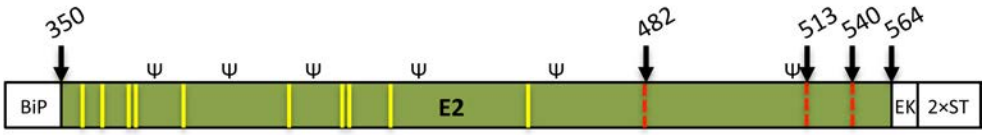


Figure 22. GBV-B E2e constructs truncated at the C-terminus. Red dashed lines indicate the borders of GBV-B E2e C-terminally truncated constructs produced based on proteomic analysis of trypsin resistant fragments and secondary structure predictions. Glycosylation sites and cysteine residues are labeled with Ψ and yellow lines, respectively. The numbering corresponds to the precursor polyprotein.

The question arose if the recombinant antibody fragments would still bind to the new variants of GBV-B E2 I was generating. To test this question, interaction between trypsin treated GBV-B E2e and the recombinant Fab F7.7 was evaluated by SEC (**Figure 23**). Several peak fractions were analyzed by N-terminal sequencing, revealing that the N-terminus of GBV-B E2 co-eluted together with the N-termini of the light and the heavy chain of the Fab. These results indicated that the epitope of the Fab is located within the trypsin resistant fragments of GBV-B E2e.

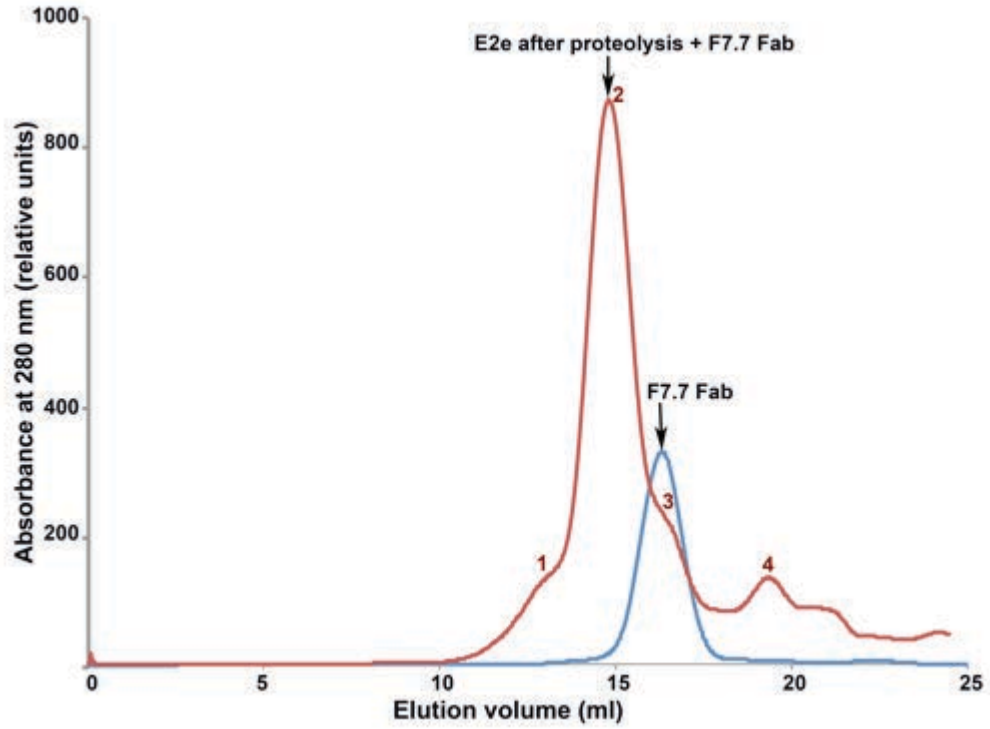


Figure 23. Interaction between trypsin treated GBV-B E2e and the recombinant Fab F7.7. F7.7 Fab and a preincubated mixture of the trypsin treated E2e and F7.7 Fab (molar ratio 1:1) were loaded to a Sdx200 size exclusion column (in two different runs). The fractions 1, 2, 3 and 4 were subjected to N-terminal sequencing. The fractions 1, 2, and 3 contained the N-terminus of GBV-B E2e, as well as the N-termini of the light and heavy chains of the Fab, indicating that the trypsin treated E2e and F7.7 Fab co-eluted together. The analysis of fraction 4 revealed the N-terminus of the C-terminal fragment.

All of the new C-terminally truncated variants of E2e were monomeric in SEC. The shortest construct GBV-B $E2_{350\text{-}482}$ (called E2ΔC) was chosen as a new target for crystallization given that it contained all cysteine residues, and therefore likely all disulfide bridges, and 5 out of 6 N-linked glycosylation sites present in GBV-B E2e (**Figure 22**). Moreover, it also retained the epitopes for all four mAbs described above. However, the initial co-crystallization trials of GBV-B E2ΔC with the antibody fragments described previously did not yield crystals.

Crystallization of deglycosylated GBV-B E2ΔC

Although GBV-B E2 is less extensively glycosylated than its HCV counterpart (6 predicted N-linked glycans in GBV-B E2 instead of 11 in HCV E2), the glycans present may hinder crystallization of the protein given the high degree of flexibility of the sugar chains. One way of avoiding this problem is enzymatic deglycosylation of the protein after

purification from the supernatant. This is preferable to mutating the N-glycosylation sites since glycosylation may be important during folding in the endoplasmic reticulum. In addition, the Strep tag, which is fused at the C-terminal end of the recombinant glycoprotein as well as to the Fab C-terminal end, may also interfere with crystallization. Therefore, I developed a protocol for deglycosylation and the enzymatic removal of the affinity tag from GBV-B E2ΔC and the Fabs. The double Strep affinity tag was removed from E2ΔC and the Fab by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). The amount of enterokinase required to achieve complete removal of the tag was optimized for each protein in a small-scale reaction (**Figure 24**). The reaction was then scaled up in a linear manner and the protein without the Strep-tag was purified from the reaction mixture.

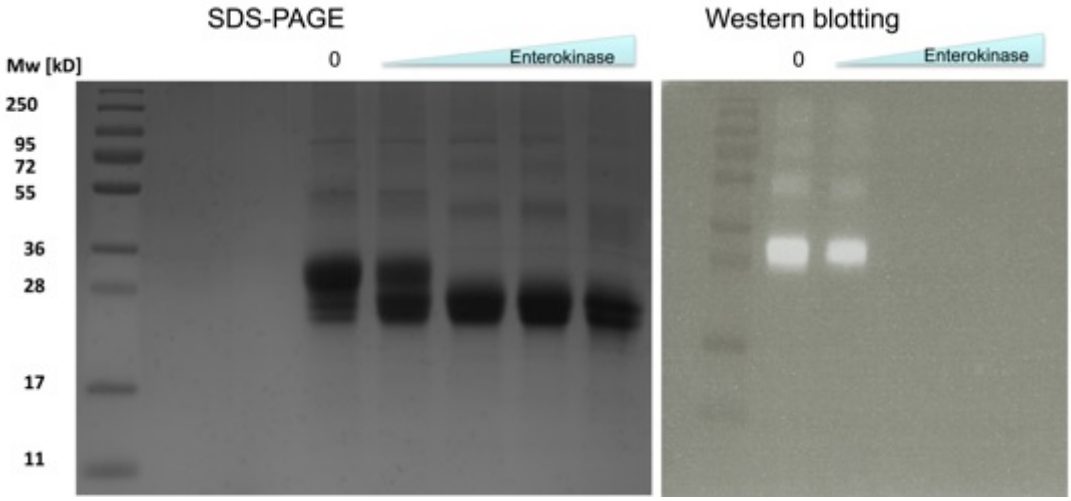


Figure 24. Enzymatic removal of the Strep affinity tag from E2ΔC. E2ΔC was incubated with increasing amounts of enterokinase as indicated by the bar above. E2ΔC mobility shift corresponding to the protein without the tag was observed on the SDS-PAGE gel (left panel). A Western blotting of the same samples was performed using an anti-Strep antibody (right panel). The ratio of protein: enterokinase for which no signal was detected in Western blotting was chosen for linear upscaling of the reaction. A control sample (without enterokinase treatment) is labeled by (0).

The deglycosylation of E2ΔC after the removal of the affinity tag was attempted using deglycosylases PNGase F and EndoH. The extent of deglycosylation of E2ΔC by different endodeglycosidases was assessed by mobility shift of the deglycosylated protein versus the intact glycoprotein on SDS-PAGE gels. A clear mobility shift of E2ΔC deglycosylated by PNGase F (**Figure 25**) was observed on SDS-PAGE gel. Deglycosylation of E2ΔC by EndoH did not have any evident effect on protein mobility on the SDS-PAGE gel indicating that this endodeglycosidase most likely is not able to remove the sugars from the protein. However,

the protein deglycosylated with PNGase aggregated as indicated by SEC. Therefore, I chose an alternative approach of deglycosylating E2ΔC in a complex with F16.1 Fab which helped to prevent aggregation. The SEC profile of the deglycosylated E2ΔC/F16.1 Fab is depicted in **Figure 26.**

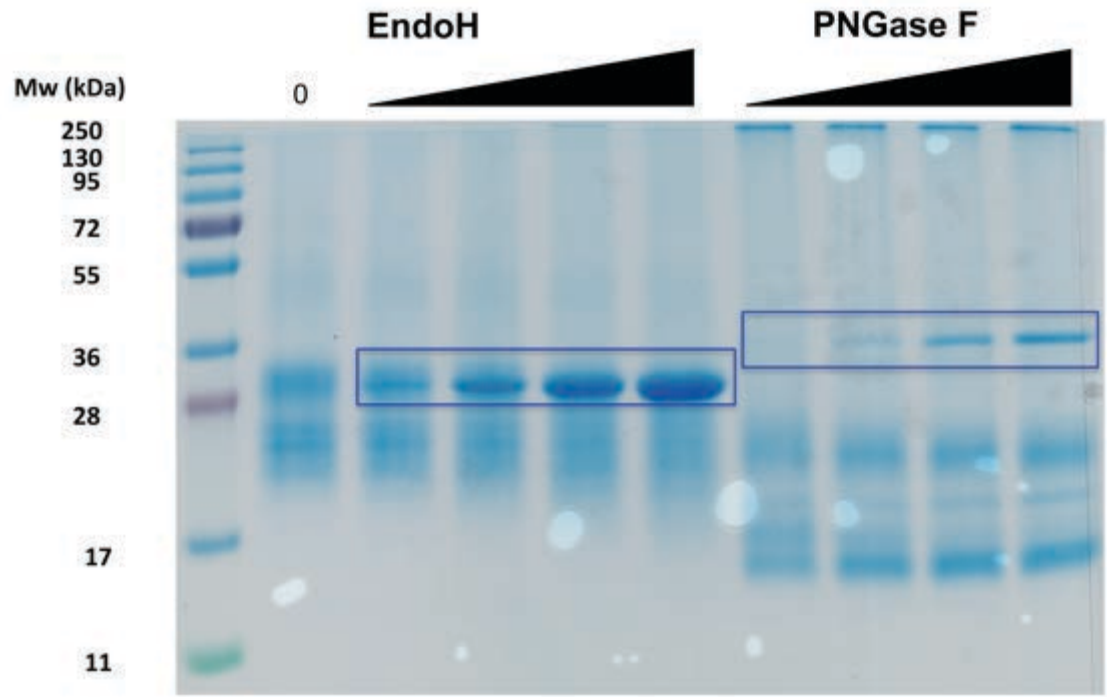


Figure 25. Enzymatic deglycosylation of E2ΔC with EndoH and PNGase F. E2ΔC was incubated with PNGase F over night at 30°C using different E2ΔC:endodeglycosidase ratios. The aliquots were removed and analyzed on SDS-PAGE gel under non-reducing conditions for the extent of deglycosylation. A clear mobility shift of E2ΔC deglycosylated by PNGase F versus non-deglycosylated E2ΔC (labeled (0)) was observed. Deglycosylation of E2ΔC by EndoH did not have any evident effect on protein mobility on the SDS-PAGE gel. Lane 1: protein ladder, lane 2: non-deglycosylated E2ΔC. EndoH and PNGase appear on the gel as 29 kD and 36 kD bands respectively (framed in blue).

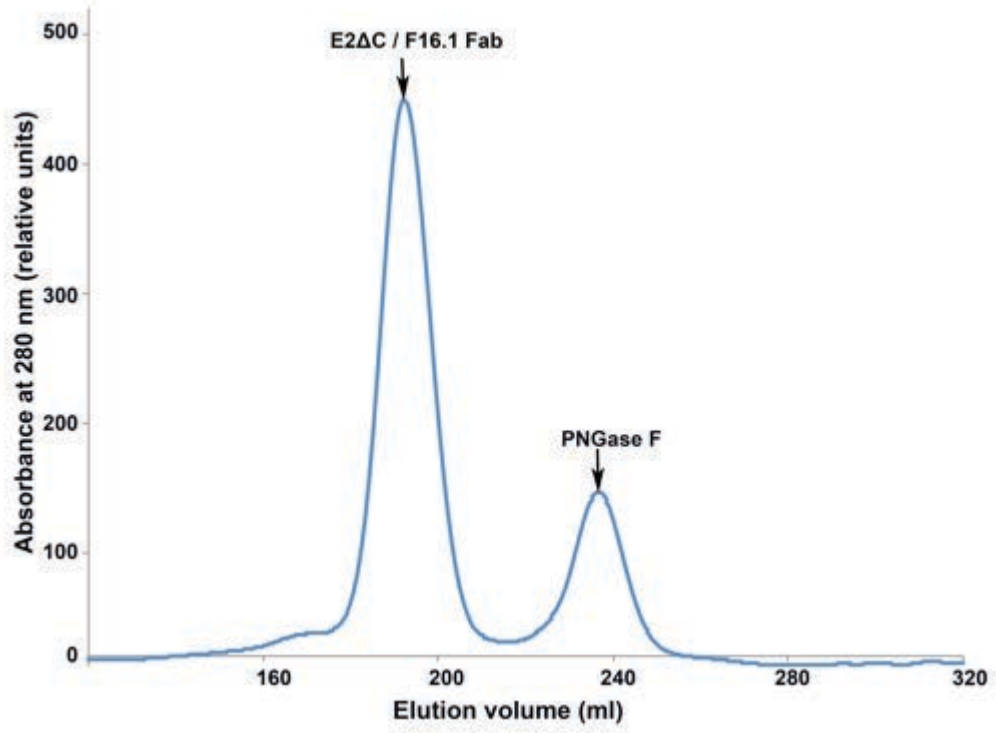


Figure 26. SEC profile of the deglycosylated E2ΔC/F16.1 Fab. E2ΔC after the removal of the Strep-tag was preincubated with F16.1 Fab over night. The preformed complex was subjected to deglycosylation by PNGase F at 30°C over night. After deglycosylation with PNGase F, E2ΔC/F16.1 Fab was separated from the endodeglycosidase by SEC. The first peak in the chromatogram corresponds to the deglycosylated E2ΔC, while the second peak corresponds to PNGase F.

The crystallization trials of the deglycosylated GBV-B E2ΔC in complex with F16.1 Fab as well as in a binary complex with F7.7 and F24.3 Fabs yielded crystals under several conditions (**Figure 27**), but the diffraction analysis and the determination of the crystal structure revealed that they contained only the Fab fragment. These crystals grew, however, at pH 4.6, which potentially leads to the dissociation of the Fab/glycoprotein complex. Therefore, I performed an interaction analysis of the complex at acidic and neutral pH, which revealed that the complex indeed dissociates at pH 4.6. This suggested that only the crystals growing at neutral or alkaline pH possibly contain glycoprotein in complex with a Fab molecule. Next, we decided to use the Fab crystals to microseed into a solution containing the purified complex to see if they would promote crystallization of the complex under conditions at neutral pH. I obtained multiple hits but most of them grew at acidic pH, meaning that the crystals were again crystals of the Fab alone. Nevertheless, crystals were also obtained in some conditions at neutral pH. These crystals were further optimized and subjected to X-ray diffraction analysis, which again revealed that they also contained only the Fab.

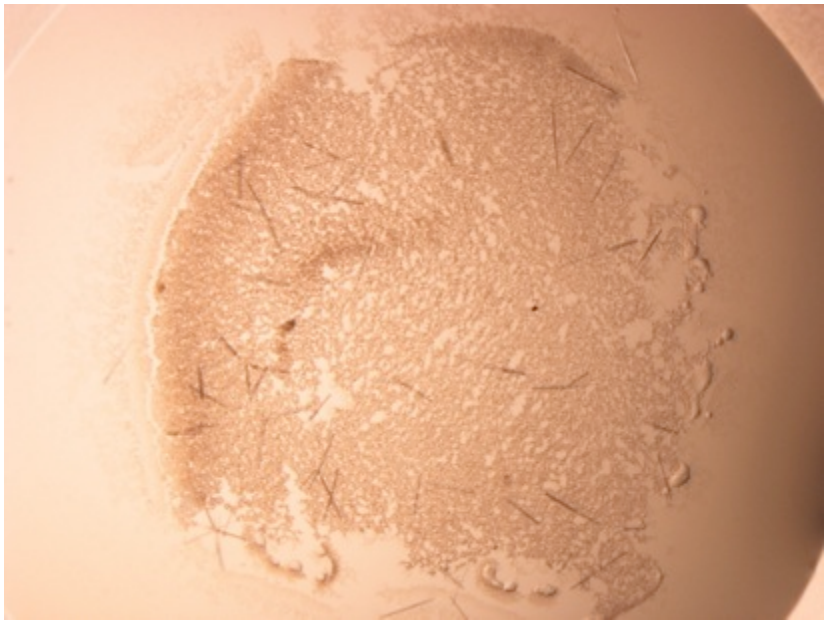


Figure 27. Photographs of the crystals obtained in crystallization trials of GBV-B E2ΔC complex with F16.1 Fab. These needle-like crystals grew in a condition containing sodium acetate pH4.6 as a buffer. The crystals diffracted to ~2Å resolution, however, they contained only the Fab molecules.

Crystallization trials of the C-terminal domain of GBV-B E2e

In view of the difficulty obtaining crystals of GBV-B E2ΔC, crystallization of a deletion mutant consisting only of the C-terminal residues 483-564 was attempted. We assumed that it might correspond to domain III as in other class II fusion proteins. Domain III of flaviviruses and alphaviruses has an Ig-like fold. It can be produced in *E.coli* as inclusion bodies and refolded. Moreover, the C-terminal fragment of GBV-B E2e (called $E2_{483\text{-}564}$) did not contain any disulphide bonds making *E.coli* a preferable expression system. I therefore expressed $E2_{483\text{-}564}$ in *E.coli* as inclusion bodies and tried to refold it from the solubilized inclusion bodies without success. Therefore, I cloned and expressed the protein in *Drosophila* S2 cells. Although I obtained many aggregates in SEC (**Figure 28A**), the elution profile showed a fraction corresponding to monomeric $E2_{483\text{-}564}$ and which was hopefully correctly folded. The fractions from SEC corresponding to the monomeric $E2_{483\text{-}564}$ were pooled and concentrated. The concentrated protein was again loaded to a SEC column to verify if it stayed monomeric after concentration. The protein eluted as a single symmetric peak containing the monomeric $E2_{483\text{-}564}$ as judged by comparison of its elution volume with those of known protein standards (**Figure 28B**).

Unfortunately, the crystallization trials of $E2_{483\text{-}564}$ did not yield crystals suitable for X-ray diffraction experiments.

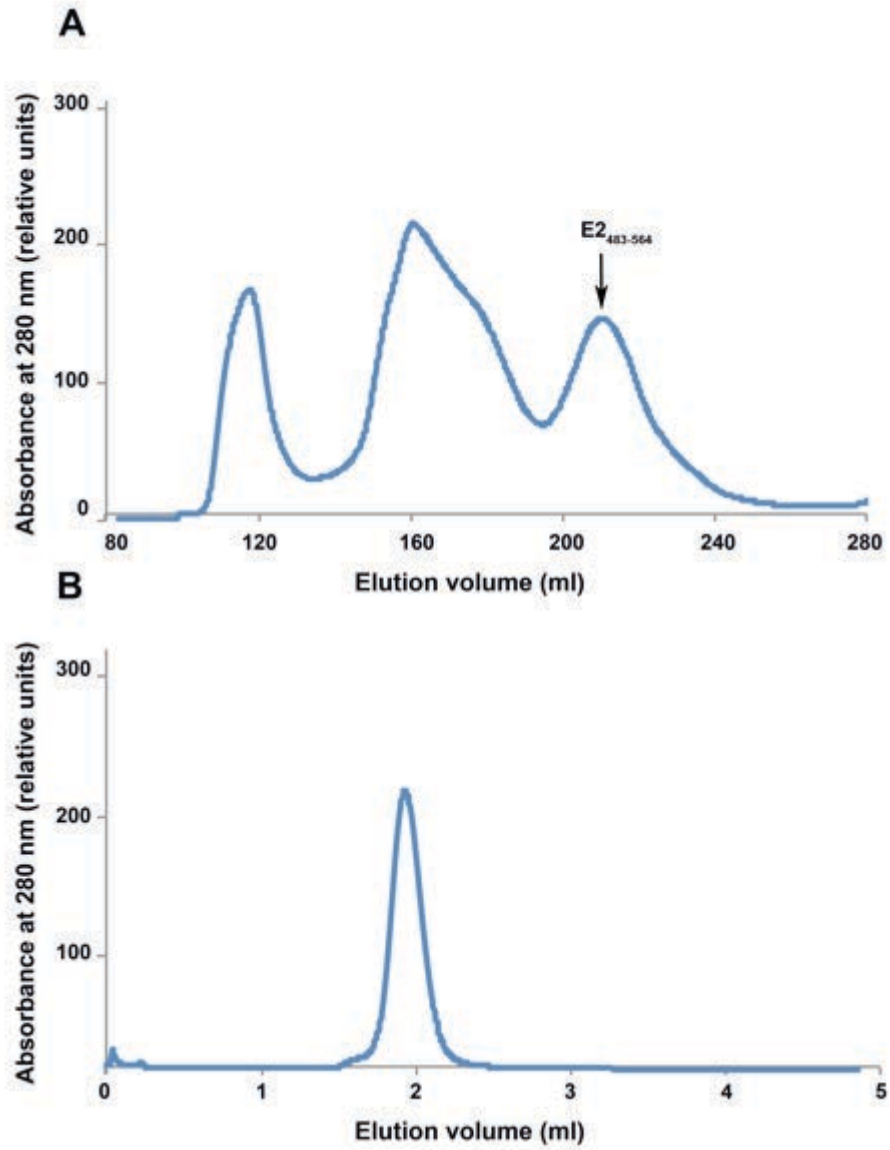


Figure 28. SEC profiles of the C-terminal fragment of GBV-B $E2_{483-564}$. (A) After the Streptactin affinity purification step the protein was further separated by SEC using a HiLoad 26/60 Superdex 75 column (GE Healthcare Life Sciences). A large proportion of the protein eluted as large molecular weight species, however, the peak corresponding to the monomeric $E2_{483-564}$ was also obtained. (B) The pooled and concentrated fractions corresponding to the monomeric $E2_{483-564}$ from (A) were analyzed by SEC using a Superdex 5/150 column. The protein eluted as a single monomeric peak as judged by comparison of its elution volume with those of known protein standards.

Production of GBV-B E2 domains based on the crystal structure of BVDV E2

At the beginning of 2013, the structure of the glycoprotein E2 of BVDV was reported by two different laboratories (PDB accession code 2YQ2) {El Omari, 2013 #131; Li, 2013 #130}. The reported structure revealed that pestivirus E2 does not have the characteristic class

II fusion protein fold like flaviviruses and alphaviruses {Kielian, 2006 #218}, but is composed of four β-sandwich domains A to D, arranged linearly from the N to the C terminal end. It has been suggested that pestivirus E2 may be the receptor-binding protein and is not involved in direct fusion of viral and cellular membranes. The structure of HCV E2 core was still unknown at the time. We hypothesized that a structural homology between the glycoprotein counterparts of hepaciviruses and pestiviruses is conceivable since viruses in the pestivirus and hepacivirus genera appear closer to each other than to the flavivirus genus.

To test this hypothesis, we tried to identify the corresponding domains in GBV-B E2 based on cysteine positions and previously obtained proteolysis data. I made three new constructs of GBV-B E2 for production of the putative domains A ($E2_{350-416}$), B ($E2_{417-523}$) and AB ($E2_{350-523}$) in *Drosophila melanogaster* S2 cells. The stable cell lines for these constructs were established and the proteins were expressed and purified. The yield of the putative domain A was ~100 µg/L and the SEC profile indicated that the protein is most likely a monomer, although further characterization of its oligomerization state has not been performed. A putative domain B was expressed in high amounts (~20 mg/L), but SEC analysis revealed that the protein was aggregated. The construct for the putative AB domain protein was produced at levels similar to GBV-B $E2_{350-482}$ (~300 µg/L) and ran as a monomer in SEC. The fact that the putative domain B was shown to be aggregated in SEC, and thus probably not folded correctly, indicated that either we did not succeed in identifying the correct borders for domains A and B, or that the pestivirus model does not apply to hepaciviruses (which was recently confirmed by HCV E2 core crystal structure) (Khan *et al.*, 2014; Kong *et al.*, 2013). Based on BVDV E2 and HCV E2 crystal structures, GBV-B E2 is likely to display receptor binding function but not membrane fusion. It is possible that it has a completely unrelated fold to BVDV as well, as the hepacivirus E2 as viral attachment glycoproteins would be much less evolutionary conserved than the fusion glycoproteins.

Discussion

Hepacivirus fusion machinery is still unknown

The *Flaviviridae* family consists of four genera: *Flavivirus*, *Hepacivirus*, *Pestivirus* and *Pegivirus*. Until 2013, the structural information about envelope glycoproteins was available only for flaviviruses, which encode class II fusion proteins catalyzing viral and cellular membrane fusion (reviewed in (Kielian, 2006). The structures of envelope glycoproteins of other members of the *Flaviviridae* family (hepaciviruses, pestiviruses and pegiviruses) remained unknown until recently despite intensive research efforts worldwide. Most of attempts were directed towards crystallization of the E2 glycoprotein of the major human pathogen HCV.

Based on a number of key characteristics such as genome organization between flaviviruses and hepaciviruses, HCV E2 was postulated to have a class II fusion protein architecture (Garry & Dash, 2003; Krey *et al.*, 2010). Nevertheless, the smaller size of E2 glycoproteins of hepaciviruses suggested that they would not represent typical class II fusion proteins.

At the beginning of my thesis (September, 2010), GBV-B was the only virus other than HCV classified as a hepacivirus and, thus, the closest relative for HCV. Therefore, we chose to pursue the structural characterization of GBV-B E2 hoping that the structure of GBV-B E2 would provide important insights into the structure and function of hepacivirus glycoproteins. GBV-B E2, being less glycosylated and smaller than its HCV counterpart, was expected to be more amenable to crystallization than HCV E2. However, all attempts to crystallize the soluble ectodomain of GBV-B E2 (GBV-B E2e) remained unsuccessful despite the fact that I applied different strategies (such as co-crystallization with antibody fragments, deglycosylation and crystallization of the protein fragments obtained by limited proteolysis).

Our main finding during this project (a result of our collaboration with Annette Martin at the Institut Pasteur) was that GBV-B $E2_{350\text{-}581}$, which represents the full-length recombinant ectodomain of GBV-B E2, is able to inhibit infection of primary marmoset hepatocytes. This finding not only indicates that the recombinant full-length GBV-B E2e adopts a conformation that is similar to the one present on virions, but also implies that GBV-B E2 is responsible for binding to a receptor on the surface of host cells - as are E2 proteins of other members of hepaci- and pestiviruses.

The shorter variant of GBV-B E2e ($E2_{350-540}$) has also been tested for its ability to compete with infectious GBV-B for entry receptors, but inhibition was not observed. These results suggest that the C-terminal residues 541-581 of GBV-B E2 are involved in receptor binding or are required for the protein to adopt an active conformation. However, the inhibition assays with $E2_{350-540}$ and $E2_{350-581}$ were done as separate experiments at different times, and it can not be ruled out that the observed differences in inhibition are due to variations in experimental conditions. Testing all different C-terminal deletion mutants in the same experiment would allow further verification of the observed results. The receptor/s involved in GBV-B entry are not yet identified, and the recombinant GBV-B $E2_{350-581}$ could potentially be used for identification of a cellular receptor.

The main problem encountered while working on this project was the low crystallizability of GBV-B E2e, either by itself or in complex with ligands. Of note, obtaining the crystal structure of the HCV E2 core required designing and expressing 41 different soluble E2 constructs in order to identify 7 constructs that were further screened with a number of E2-specific Fabs in crystallization trials (Kong *et al.*, 2013). Although it is not known yet if HCV and GBV-B E2 display similar folds, HCV E2 crystallization illustrates the degree of difficulty for crystallizing hepacivirus glycoproteins, which seems to also apply to GBV-B. Another major obstacle was the low expression level of the protein (0.3-0.5 mg/l) which was a limiting factor in performing crystallization trials. Eventually, no useful crystallization chaperons were identified given that: 1) three out of four mAbs cross-competed for binding to E2 and 2) the complex of GBV-B E2e with these antibody fragments appeared to be unstable at acidic pH, which limited the number of screened crystallization conditions.

Several strategies could still be explored in order to obtain GBV-B E2e crystals. It would be beneficial to obtain more ligands for GBV-B E2 that might help in crystallization of this protein. Recently, as a result of the collaboration with Annette Martin (Pasteur Institute, Paris) and Nicola J. Rose (National Institute for Biological Standards and Control, Hertfordshire, UK), B-lymphocytes from a marmoset infected with GBV-B were obtained. Antibody genes could be isolated from these B-lymphocytes and used to create an antibody fragment library by combinatorial phage display technology. Screening of the library would allow identification of new GBV-B E2 specific ligands.

Another feature that might be responsible for hindering crystallization of GBV-B E2e is its glycosylation. The glycoprotein was partially deglycosylated with PNGase F but its solubility was reduced, becoming prone to aggregation, which made it unsuitable for crystallization trials. After deglycosylation of E2e, I added the Fab fragment in order to form

a complex of the deglycosylated E2e, which is more soluble than the deglycosylated E2 alone, although this did not result in crystals. An alternative strategy to protein deglycosylation is the targeted knockout of N-linked glycosylation sites by mutagenesis of the asparagine residue to aspartic acid. This strategy would allow us to determine whether any of the glycans are not required for folding and secretion of the glycoprotein. Such mutant/mutants would constitute promising new target/s for crystallization. Alternatively, a combination of peptide-*N*-glycosidases EndoH and EndoD could be tried as an alternative to PNGase. Since these enzymes cleave between the first and second N-acetylglucosamine moiety attached to the asparagine, the deglycosylated protein may be more soluble. Another option would be to express the protein in the presence of the N-linked glycosylation inhibitor tunicamycin, which is often used to reduce glycosylation of the recombinant proteins used for crystallization trials.

The atomic structures of the heterodimeric Chikungunya virus envelope glycoproteins (Voss *et al.*, 2010) were determined by using a construct that contained covalently linked ectodomains of p62 and E1. This strategy could be potentially exploited for GBV-B E1 and E2 glycoproteins which would require replacing the transmembrane region of E1 by a flexible $(GGGGS)_4$ linker that may allow the secretion of a covalently linked soluble E1-E2 heterodimer.

Moreover, other biophysical methods such as nuclear magnetic resonance (NMR) might be used to provide additional structural information. Francois Bontems in our laboratory has developed a protocol for efficient labeling of recombinant proteins in S2 cells with stable isotopes (^{15}N and ^{13}C) for NMR (Meola *et al.*, 2014).

In 2013, the atomic structures of BVDV E2 ectodomain and a core fragment of HCV E2 were reported (El Omari *et al.*, 2013; Khan *et al.*, 2014; Kong *et al.*, 2013; Li *et al.*, 2013). Unexpectedly, both proteins were found to have novel folds unrelated to the class II fusion protein fold. HCV and BVDV E2 structures revealed that these proteins are not likely to function as fusogens as they lack structural features of fusion proteins, in particular a distinguishable fusion peptide or fusion loop. It remains an open question if the GBV-B E2 fold resembles the folds of HCV or BVDV. The trials to express putative GBV-B E2e domains based on BVDV E2e structure were unsuccessful, implying that either GBV-B E2 might have yet a different fold, or we did not identify the correct borders of the domains. Alternatively, GBV-B E2 might have a fold similar to the globular fold of HCV E2, which contains an Ig-like β-sandwich. It is possible, however, that GBV-B E2 harbors a fold unrelated to the ones observed in HCV and BVDV. HCV and BVDV E2 have been shown to

play a role in virus entry as receptor-binding proteins (interacting with CD81 and CD46, respectively) (Cormier *et al.*, 2004b; Maurer *et al.*, 2004; Pileri *et al.*, 1998). Usually receptor-binding viral envelope proteins are to a less degree evolutionary conserved than fusion proteins, so these proteins could evolve to have different folds even in phylogenetically related viruses. Elucidation of HCV and BVDV E2 protein architecture also raised the hypothesis that E1 glycoproteins of hepaciviruses and pestiviruses are likely to be the fusion protein in these two viral genera. A putative fusion motif has been identified in both of these proteins (El Omari *et al.*, 2013; Li *et al.*, 2013; (Drummer *et al.*, 2007). However, it is unlikely that E1 has a class I, II or III fusion protein fold for several reasons: it is much smaller (~180 aa) than all fusion proteins of known structure and is not likely to span the distance between viral and cellular membranes with any of the above folds; and it is also not likely to adopt a class I fusion protein fold because it lacks important features typical for these proteins such as proteolytic activation and heptad repeat motif involved in forming a helical bundle. These observations have been supported by recently reported crystal structure of the N-terminal domain of HCV E1 revealing that its fold is not related to any class I, II or III fusion proteins (El Omari *et al.*, 2014). Nevertheless, it is possible that in spite of the different folds of E2 proteins observed in pestiviruses and hepaciviruses, the E1 protein still has a conserved architecture in both genera. In light of the current information about hepacivirus and pestivirus glycoproteins, it seems that viruses belonging to these genera have evolved to use a novel fusion machinery which still remains to be deciphered. To achieve this, further structural studies of hepacivirus and pestivirus glycoproteins, in particular of E1E2 complex, is needed.

Materials and methods

Constructs used or generated in the study

All GBV-B E2 glycoprotein constructs were inserted into the pT350 vector (described in Supplementary materials and methods) and expressed in *Drosophila melanogaster* S2 cells. GBV-B $E2_{350-564}$ (also encoding E1) and GBV-B $E2_{350-581}$ were generated in the lab by Christine Girard-Blanc and Thomas Krey before the PhD project was started by using a synthetic gene of GBV-B E2 ectodomain. The constructs of the proteolytic fragments of GBV-B E2 (GBV-B E1-$E2_{350-482}$ called E2ΔC), E1-$E2_{350-540,}$ and E1-$E2_{350-513,}$ as well as putative domains A ($E2_{350-416}$), B ($E2_{417-523}$) and AB ($E2_{350-523}$) were produced by deletion mutagenesis.

Deletion mutagenesis was performed using inverse PCR with two inverted tail-to-tail primers to amplify the entire plasmid sequence excluding the region to be deleted (Imai *et al.*, 1991). PCR was performed using Phusion High-Fidelity DNA polymerase (Finnzymes) (see Supplementary methods for PCR conditions). The PCR product was then incubated with 30 U DpnI (NEB) for 1-2 h at 37°C to digest the parental DNA and purified using Macherey Gel and PCR Clean-up Kit (Macherey-Nagel). Purified DNA was eluted in 20 µl water and the 5'-OH group of DNA was phosphorylated using T4 Polynucleotide Kinase (Thermo Fisher Scientific) according to the manufacturer's protocol. The purified DNA was self-ligated and transformed into competent Top 10 *E. coli* cells. The primers used to generate the constructs are provided in **Table 3** at the end of this chapter**.**

Generation and production of recombinant Fab and scFv molecules

Antibody fragments (scFv and Fab) are routinely used in co-crystallization studies with peptides and proteins. They contain the entire antigen binding region, and thus the specificity of the parent mAb. Fab molecules are composed of a light chain (LC) and approximately half of the Ig heavy chain (HC), termed the 'Fd fragment'. Both LC and Fd fragment contain a variable domain and a constant domain, which is called L_C and C_{H1} respectively. scFv molecules are composed of the variable regions of LC and HC (V_L and V_C) (**Figure 29**).

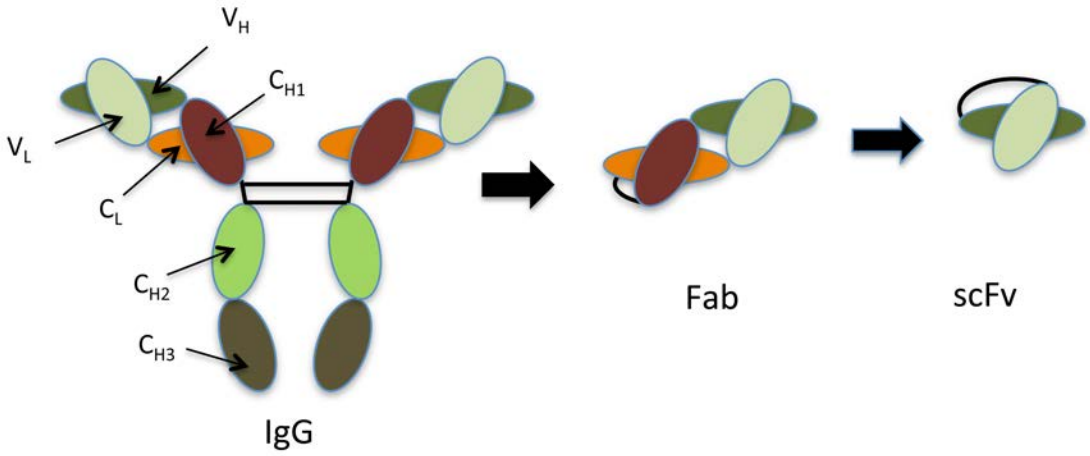


Figure 29. Representation of IgG, Fab and scFv molecules. Monoclonal antibodies can be used in form of a Fab (fragment antigen binding) or an scFv (single chain variable fragment) for co-crystallization.

Anti-GBV-B E2 mAbs were obtained by immunizing mice with the recombinant GBV-B $E2_{350-564}$. The isotype of MAbs was detected using a Pierce Rapid Elisa Mouse mAb Isotyping Kit (Thermo Fisher Scientific). Total RNA from mouse hybridoma cells was isolated using a RNeasy Mini kit (Qiagen) and the first-strand cDNA synthesis was carried out using SuperScript III Reverse Transcriptase (Invitrogen) according to the protocols provided by the manufacturers. The sequences encoding the Fd (V_H-C_{H1}) and LC (V_L-C_L) were amplified by PCR using Phusion polymerase (Finnzymes) and one of the pairs of the primers (**Table 3**) annealing to the 5′ and 3′ ends of the Fd and LC segments of mouse Igs.

The functional sequences of the LC genes were not obtained with these primers for mAbs F24.3 and F7.7. Instead of the functional sequences of the LC an aberrant kappa LC transcript coexisting in the hybridoma was obtained. The aberrant LC transcript has a premature termination codon at position 105 (according to the Kabat numbering system) and it is not translated into fully functional LC. Often hybridomas express the aberrant LC transcript to a greater extent than the functional LC gene which makes identifying the functional LC transcript difficult.

Based on two reports in the literature (Juste *et al.*, 2006; Yuan *et al.*, 2004), a specific strategy to obtain the functional LC sequences was created and applied. First, the N-terminal sequencing for F24.3 and F7.7 LCs was carried out in order to design the gene specific primers. Then the PCR using cDNA isolated from the hybridomas was performed with a molar excess of a primer complementary to the CDR3 of the aberrant LC sequence in addition to the primers designed to amplify the functional LC. The obtained PCR products were digested with *Bci*VI (NEB) restriction endonuclease. The *Bci*VI restriction site is present in

the aberrant kappa LC transcript, however, it is absent in the majority of known LC sequences.

The PCR products of ~750 bp (corresponding to the size of LC and Fd sequences) were cloned in the pCR-Blunt II-Topo vector (Zero Blunt PCR Topo cloning, Invitrogen) and sequenced. Once the whole coding sequences for Fd and LC were known, new gene-specific primers were designed.

The pMT-Fab-Strep vector contains two multiple cloning sites allowing co-expression of two target genes (Backovic *et al*., 2010). The pMT-scFv-Strep vector also contains two multiple cloning sites separated by a linker sequence which allows insertion of V_H and V_L genes in a way that they are joined by a linker sequence (Gilmartin et al., 2011). The genes coding for V_H and V_L or Fd and LC sequences were amplified by PCR. The gene encoding LC was inserted into pMT-Fab-Strep between the BglII and XbaI restriction sites, while the Fd gene was inserted into the vector between the AvrII and NheI restriction sites. The amplified V_L gene was cloned into the pMT-scFv-Strep vector using NheI and NotI restriction sites while the V_H gene was inserted between the NcoI and KpnI restriction sites.

Transfection of *Drosophila* S2 cells, generation of the stable cell lines, and production and purification of the recombinant Fab and scFv molecules were carried out as described in Supplementary materials and methods.

Site directed-mutagenesis of the free cysteine residue to serine residue in GBV-B E2 antibody fragments F16.1, F7.7 and E19.4 was performed using QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies) according to a protocol provided by the manufacturer.

Deglycosylation of GBV-B E2ΔC

Production of endodeglycosidases PNGase, EndoH and EndoD is described in Supplementary materials and methods. GBV-B E2ΔC was deglycosylated with PNGase at 30 ^{o}C for 16 h at PNGaseF:glycoprotein ratio on a weight basis 1:3. Deglycosylation was carried out at the final concentration of GBV-B E2 of 0.167 mg/ml. The deglycosylated protein had a relatively low solubility. Therefore, the deglycosylation was performed at low concentration of the protein in order to avoid the aggregation of the deglycosylated protein. To improve the solubility of the deglycosylated protein, anti-GBV-B E2 Fab was added to the reaction mixture after 16 h and the proteins incubated at room temperature for an additional 2 h. Deglycosylated GBV-B E2/Fab complex was separated from PNGase F by SEC on Superdex 200 26/60 column (GE Healthcare). The purified complex was concentrated and used for

further experiments (removal of the Strep-tag by enterokinase cleavage and crystallization trials).

Removal of the double Strep affinity tag

A C-terminal double Strep tag preceded by an enterokinase recognition site was removed from the recombinant antibody fragments (Fabs and scFvs) or GBV-B E2ΔC/Fab complexes by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). The detailed protocol is provided in Supplementary Materials and Methods.

SEC analysis of glycoprotein complexes with the antibody fragments

20-30 µg of the soluble glycoprotein and an equimolar amount of the antibody fragment (Fab or scFv) were incubated as isolated proteins as well as in complex for 16 h at 4°C followed by analysis on Superdex 200 5/150 column (GE Healthcare, Uppsala, Sweden).

In order to analyze cross-competition between two different mAbs, the above protocol was applied with the only difference being that the glycoprotein was incubated together with two Fabs and analyzed by SEC for a ternary complex formation.

Limited proteolysis

Purified GBV-B E2e protein alone or in complex with antibody fragments was incubated with increasing concentrations of TPCK (L-1-tosylamide-2-phenylmethyl chloromethyl ketone)-treated trypsin (Sigma) for 2h at 37 ^{o}C. Digestion was stopped by adding PMSF at a final concentration of 0.1 mM. The cleaved protein was subsequently analyzed by SDS-PAGE stained with Coomassie Blue. In addition, target protein bands were analyzed by N-terminal sequencing (a facility at Institut Pasteur). SELDI-TOF analysis of the digested protein was also performed.

Direct Enzyme-linked Immunosorbent Assay (ELISA)

The binding of the anti-GBV-B E2 Fabs to GBV-B E2 proteolytic fragments was assessed by ELISA. Nunc Maxisorp 96 well plate (Immunosorp, Nunc, Denmark) was coated

with 50 µl/well of 1 µg/ml glycoprotein in PBS over night at 4^{o}C. Next day, 50 µl/well of monoclonal antibodies serially diluted in 0.5% bovine gelatin+0.1% Tween in PBS were added, and the plates were incubated at room temperature for 2 h. The dilutions of the antibodies used in the assay were 50.00, 25.00, 12.50, 6.25, 3.13, 1.56 and 0.78 µg/ml. Non-relevant monoclonal antibody (negative control) was also included on each plate. After washing the plates 5× with PBS-Tween (0.05%), 50 µl of horseradish peroxidase conjugated anti-mouse total IgG (0.4 mg/ml from Thermo Fisher Scientific) diluted 1/10000 in 0.5% bovine gelatin+0.1% Tween in PBS was added to each well. Then the plates were incubated for 1 h at 37^{o}C. Subsequently, the plates were washed as described above. The bound antibodies were detected by adding 100 µl/well of 3,3',5,5'-Tetramethylbenzidine (TMB) Microwell Peroxidase Substrate System (KPL, Gaithersburg, MD) solution prepared according to the instructions included with the product. The reaction was stopped after 4 min with 100 µl 2M H_2SO_4. The absorbance at 450 nm was measured by VICTOR 1420 Multilabel Counter (Perkin Elmer,Waltham, MA).

Detection of free cysteine residues in the antibody fragments

Free cysteine residues in anti-E2 antibody fragments were detected using the DTNB-thiols assay. The detailed protocol is provided in Supplementary materials and methods.

Crystallization

Crystallization screening, crystal optimization and crystal cryo-protection techniques are in detailed described in Supplementary materials and methods

Table 3. Selected protein constructs and PCR primers.

Protein	Construct	Forward primer (5'to 3'direction)	Reverse primer (5'to 3'direction)	Construct generation
GBV-B $E2_{350-482}$ (E2ΔC)	pT350/ $E2_{350-482}$(E2ΔC)	TTCGAAGACGATGACGATAAGGCCGGTTG	TTTGCTGCCGGGGTACTGCAGGATG	Deletion using pT350/E1-$E2_{350-564}$
GBV-B $E2_{350-540}$	pT350/$E2_{350-540}$	TTCGAAGACGATGACGATAAGGCCGGTTG	GGAATTCACGGGGTCGTAGAAGTACACG	Deletion using pT350/E1-$E2_{350-564}$
GBV-B $E2_{350-513}$	pT350/$E2_{350-513}$	TTCGAAGACGATGACGATAAGGCCGGTTG	CTGGGGCACCTGCAGCCAG	Deletion using pT350/E1-$E2_{350-564}$
GBV-B $E2_{350-416}$ (domain A)	pT350/$E2_{350-416}$	TTCGAAGACGATGACGATAAGGCCGGTTG	CCTAGCCCACCTCTCCGGTGGTAGGAGAG	Deletion using pT350/ $E2_{350-581}$
GBV-B $E2_{417-523}$ (domain B)	pT350/$E2_{417-523}$	ATGGTTAAATTCAAAAATAACACATGGGG	CCCGAGCGAGAGGCCAACAAAG	Deletion using pT350/ $E2_{350-523}$
GBV-B $E2_{350-523}$ (domain AB)	pT350/$E2_{350-523}$	TTCGAAGACGATGACGATAAGGCCGGTTG	CCTAGCCCACCTCTCCGGTGGTAGGAGAG	Deletion using pT350/ $E2_{350-581}$
$E2_{483-564}$	pT350/$E2_{483-564}$	ATGGTTAAATTCAAAAATAACACATGGGG	CCCGAGCGAGAGGCCAACAAAG	Deletion using pT350/E1-

	F16.1_HC			cloning
E19.4_LC	pCR™-Blunt/ E19.4_LC	CCAGTTCCGAGCTCGTGATGACA CAGTCTCCA	GCGCCGTCTAGAATTAACACTCATTCCTGTTG AA	Blunt-cloning
E19.4_HC	pCR™-Blunt/ E19.4_HC	GAGGTCCAGCTCGAGCAGTCTGGACC	AGGCTTACTAGTACAATCCCTGGGCACAAT	Blunt-cloning
F7.7_LC	pCR™-Blunt/ F7.7_LC	AAAAAGATCTGACATCGTGCTGACACAGT CTCCA	GCGCCGTCTAGAATTAACACTCATTCCTGTTG AA	Blunt-cloning
F7.7_HC	pCR™-Blunt/ F7.7_HC	GAGGTCCAGCTCGAGCAGTCTGGACC	AGGCTTACTAGTACAATCCCTGGGCACAAT	Blunt-cloning
F24.3_LC	pCR™-Blunt/ F24.3_LC	AAAAAAAGATCTGATATCCAGATGACACA GACTACWTCCTCC	GCGCCGTCTAGAATTAACACTCATTCCTGTTG AA	Blunt-cloning
F24.3_HC	pCR™-Blunt/ F24.3_HC	GAGGTAAAGCTCGAGGAGTCTGGAGG	AGGCTTACTAGTACAATCCCTGGGCACAAT	Blunt-cloning
F16.1Fab	pMT/F16.1Fab_LC	AAAAAGATCTGAGCTCGTGATGACACAG	AAAATCTAGATTAACACTCATTCCTGTTGAAG C	Restriction cloning (BglII-XbaI)
	pMT/F16.1Fab_LC _HC		TTTTTTGCTAGCACCACAATCCCTGGGCACAA TTTTCTTGTCCACCTTGGT	Restriction cloning (AvrII-NheI)
E19.4Fab	pMT/E19.4Fab_LC	AAAAAGATCTGAGCTCGTGATGACACAG	AAAATCTAGATTAACACTCATTCCTGTTGAAG C	Restriction cloning (BglII-XbaI)
	pMT/E19.4Fab_LC _HC	AAAACCTAGGGGAGGTTCAGCTCGAGCAG	TTTTTTGCTAGCACCACAATCCCTGGGCACAA TTTTCTTGTCCACCTTGGT	Restriction cloning (AvrII-NheI)
F7.7Fab	pMT/F7.7Fab_LC	AAAAAGATCTGACATCGTGCTGACACAGT CTCCA	AAAATCTAGATTAACACTCATTCCTGTTGAAG C	Restriction cloning (BglII-XbaI)
	pMT/F7.7Fab_LC_ HC		TTTTTTGCTAGCACCACAATCCCTGGGCACAA TTTTCTTGTCCACCTTGGT	Restriction cloning (AvrII-NheI)
F24.3Fab	pMT/F24.3Fab_LC	AAAAAAAGATCTGATATCCAGATGACACA GACTACWTCCTCC	AAAATCTAGATTAACACTCATTCCTGTTGAAG C	Restriction cloning (BglII-XbaI)
	pMT/F24.3Fab_LC	AAAACCTAGGGGAGGTAAAGCTCGAGGAG	TTTTTTGCTAGCACCACAATCCCTGGGCACAA	Restriction

	_HC		TTTTCTTGTCCACCTTGGT	cloning (AvrII-NheI)
F16.1scFv	pMT/F16scFv_LC	AAAAAAGCTAGCGAGCTCGTGATGACACA GTCTCCAG	TTTTTTGCGGCCGCTTTGATTTCCAGCTTGGT GCCTC	Restriction cloning (NheI-NotI)
	pMT/F16scFv_LC_ HC	AAAAAACCATGGGCGAAGTTCAGCTCGAG CAGTCTGG	TTTTTTGGTACCTGAGGAGACTGTGAGAGTGG TGC	Restriction cloning (NcoI-KpnI)
E19.4scFv	pMT/E19.4scFv_L C	AAAAAAGCTAGCGAGCTCGTGATGACACA GTCTCCAG	TTTTTTGCGGCCGCTTTGATTTCCAGCTTGGT GCCTC	Restriction cloning (NheI-NotI)
	pMT/E19.4scFv_L C_HC	AAAATCATGAGCGAGGTTCAGCTCGAGCA GTCTG	TTTTTTGGTACCTGAGGAGACTGTGAGAGTGG TGC	Restriction cloning (BspHI-KpnI)
F7.7scFv_LC	pMT/F7.7scFv_LC	AAAAAAGCTAGCGACATCGTGCTGACACA GTCTCCAC	TTTTTTGCGGCCGCTTTGCTTTCCAGCTTGGT GCCTC	Restriction cloning (NheI-NotI)
	pMT/F7.7scFv_LC _HC	AAAATCATGAGCGAGGTCCAGCTCGAGCA GTCT	TTTTTTGGTACCTGAGGAGACTGTGAGAGTGG TGC	Restriction cloning (BspHI-KpnI)
F24.3scFv	pMT/F24.3scFv_L C	AAAAAAGCTAGCGATATCCAGATGACACA GACTACTTCCTCC	TTTTTTGCGGCCGCTTTGATTTCCAGCTTGGT GCCTC	Restriction cloning (NheI-NotI)
	pMT/F24.3scFv_L C_HC	AAAAAACCATGGGCGAGGTAAAGCTCGAG GAGTCTGGAG	TTTTTTGGTACCTGAGGAGACTGTGAGAGTGG TGC	Restriction cloning (NcoI-KpnI)
F7.7Fab C_to_S_mut		GAGTGGATTGGATATATTAAT**TCT**CACAG TGTTACTTCAACCTAC	GTAGGTTGAAGTAACACTGTG**AGA**ATTAATAT ATCCAATCCACTC	Site-directed mutagenesis

Enzyme restriction sites and codons encoding mutated amino acids are color-coded: BglII site – yellow, XbaI site – bright green, AvrII site – magenta, NheI site – light blue, KpnI site – grey, BspHI site – red, NcoI codon – green. A codon encoding mutated cysteine to serine is underlined. Light chain (LC) and heavy chain (HC) sequences cloned to pCR™-Blunt were amplified using cDNA isolated from hybridomas expressing the corresponding monoclonal antibodies.

II. The Structure Of the Hepatitis C Virus Envelope Glycoprotein E2 Antigenic Site 529-540 in Complex With Antibody DAO5

Background

In the Introduction about HCV, I have described a number of crystal structures of HCV anti-E2 antibodies in complex with its peptide epitopes. This led to structural characterizations of two important epitopes (epitope I and epitope II) targeted by neutralizing antibodies (Deng *et al.*, 2014; Deng *et al.*, 2013; Kong *et al.*, 2012a; Krey *et al.*, 2013; Pantua *et al.*, 2013; Potter *et al.*, 2012). The third region (aa 523-540 following HCV polyprotein numbering) is recognized almost exclusively by conformation sensitive antibodies that do not bind peptides. Four residues within this region (G523, W529, G530 and D535) are recognized by a number of conformation-dependent human monoclonal antibodies - as determined by alanine-scanning mutagenesis - and are involved in CD81 binding (Owsianka *et al.*, 2006). At the time of this project, structural characterization of this third region had not been reported. Together with our collaborators from MRC-University of Glasgow Centre for Virus Research (Arvind Patel and Ania Owsianka) we obtained a murine antibody DAO5 recognizing a linear epitope within aa 523-540 of HCV E2. To gain insight into this important, uncharacterized region of E2, we applied a strategy of co-crystallization of DAO5 antibody in complex with its peptide epitope.

Objective

The objective of this project was to determine the crystal structures of antibody fragments (scFv and Fab) from the murine antibody DAO5 in complex with its peptide epitope corresponding to HCV E2 aa 529-540 derived from two different HCV genotypes. In addition, we aimed to prove that this antibody recognizes the glycoprotein E2 in its native

conformation as present in the virion. The study was intended to provide a first glimpse into this important region of E2 for virus neutralization and vaccine design efforts.

Results

Expression and purification of DAO5 scFv and DAO5 Fab

Monoclonal antibody (mAb) DAO5 was generated and its epitope characterized by our collaborators in Glasgow (Arvind Patel and Ania Owsianka). In order to raise broadly neutralizing antibodies against HCV E2, mice were immunized and boosted with a series of recombinant soluble HCV E2 derived from different genotypes produced in our laboratory. The epitope mapping studies revealed that DAO5 mAb binds specifically to E2 residues 533-ETDVMLLN-540 with residue D535 being an essential contact residue for DAO5 mAb. The main features of DAO5 mAb as a result of the characterization by Arvind Patel and Ania Owsianka are summarized in **Table 1**.

Table 4. Summary of DAO5 mAb characterization.

Feature	Result	Method
Epitope	533-ETDVMLLN-540	Phage display+ELISA
Essential residues for binding	D535	Alanine scanning mutagenesis of HCVcc
Binding to the genotypes	2a, 2b, 3a, 4	ELISA
Competition analysis	Does not cross-compete with other mAbs binding to D535	ELISA
Neutralization	no	HCVpp, HCVcc
Binding to viral particles	Initial results indicate that it does bind but it needs to be confirmed	Immunopreciptation with HCVcc and HCVpp
Conformational mAb	no	Western blot, phage display of peptides

For crystallization studies with the peptides corresponding to DAO5 mAb epitope, we produced the recombinant DAO5 scFv and DAO5 Fab in *Drosophila melanogaster* S2 cells. Representative protein elution profiles from SEC are shown in **Figure 30**. DAO5 Fab eluted as a single peak corresponding to a monomeric Fab. Although the majority of the DAO5 scFv eluted from a SEC column at a volume (225-260 ml) corresponding to monomer, some extra peaks corresponding to dimeric and higher oligomeric scFv could be observed. In order to get homogenous monomeric DAO5 scFv for crystallization trials, the fractions from the peak corresponding to the monomeric scFv were pooled, concentrated and subjected to the second

SEC. As judged from the SEC profile, the second SEC resulted in a clear separation of monomeric scFv from other oligomeric species. The final yield of the DAO5 Fab and DAO5 scFv was ~21 and ~8.5 mg per liter of culture supernatant.

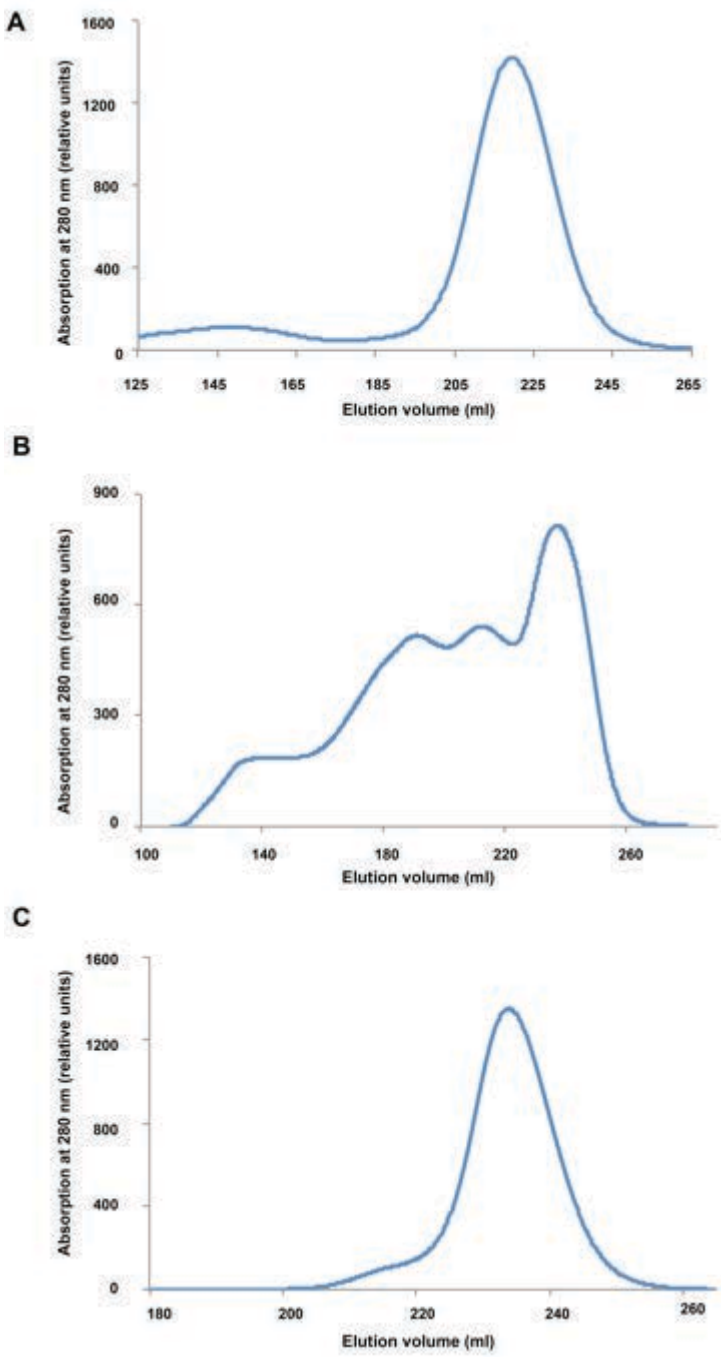


Figure 30. Size-exclusion chromatography (SEC) of DAO5 Fab (A) and DAO5 scFv (B and C). Separation by SEC was performed using HiLoad 26/60 Superdex 200 column (GE Healthcare Life Sciences) at a flow speed of 2 ml/min in 10 mM Tris-HCl pH 8.0, 150 mM NaCl. Protein elution was monitored by absorbance at 280 nm (blue curves). Chromatogram A reveals a single major peak for DAO5 Fab corresponding to a monomeric Fab (~50 kD). For DAO5 scFv, the first separation by SEC (B) resulted in a major peak corresponding to a monomeric scFv (~26 kD) and some additional overlapping peaks corresponding to dimeric and higher oligomeric scFvs. Fractions corresponding to the monomeric scFv were pooled, concentrated and then analyzed again by SEC. DAO5 scFv eluted as a single major peak (C) corresponding to a monomeric scFv, which

indicated a successful separation of monomeric scFv from higher oligomeric state species, by removing the fractions to the left (smaller volumes) of the peak.

Crystallization and structure determination of DAO5 scFv and DAO5 Fab in complex with the epitope HCV E2$^{529-540}$ peptides

Two peptides spanning the residues 529-540 from HCV genotype 1b isolate J4 and 2a isolate JFH-1 (referred to as peptide_J4 and peptide_JFH, respectively) were selected for crystallization with DAO5 antibody fragments. The sequences of the peptides are provided in **Table 5**.

Table 5. The sequences of the peptides selected for crystallization studies with DAO5 antibody fragments. The peptides span the residues 529-540 from HCV genotype 1b isolate J4 (peptide_JFH) and 2a genotype isolate JFH-1 (peptide_JFH) and differ by one amino acid at the position 537 (colored in red).

Peptide	Sequence
Peptide_JFH	WGENETDV**F**LLN
Peptide_J4	WGENETDV**M**LLN

The HCV E2 residues 529-540 are highly conserved among different HCV genotypes (**Table 6**). Therefore, the selected peptides differ only by one amino acid at the position 537 (M in the peptide_J4 and F in the peptide_JFH-1). The rationale to use these two peptides was that they represent the DAO5 epitope from the two most phylogenetically distant HCV genotypes.

For crystallization trials with DAO5 antibody fragments, we chose peptides longer by 4 residues at the N-terminal end (i.e. the peptides spanning aa 529-540) than the central DAO5 mAb epitope (HCV E2 residues 533-540) because the residues W529 and G530 belong to the amino acids that are conserved across all genotypes and are critical for CD81 binding. We were hoping that we could probably see the electron density for those residues in the crystal structures, which would give more information about this important HCV E2 region.

Table 6. The sequences of HCV E2$^{529-540}$ from six major HCV genotypes. The amino acid residues that have been shown to be critical for CD81 binding are displayed in bold. Gt: genotype. JFH 2a: GenBank accession number AB047645. H77 1a: GenBank accession number NC_004102. Gt 1b (J4): GenBank accession number AF054259. Gt 3a: GenBank accession number D28917. Gt 4a: GenBank accession number DQ41878.Gt 5a: GenBank accession number Y13184. Gt 6: GenBank accession number AY859526. The residues critical for CD81 binding (Owsianka *et al.*, 2006)are colored by red. The residues that are variable among six major HCV genotypes are colored in grey.

HCV genotype	Sequence of HCV E2$^{529-540}$
JFH 2a	**WG**ENET**D**VFLLN
H77 1a	**WG**ANDT**D**VFVLN
Gt 1b (J4)	**WG**ENET**D**VMLLN

Gt 3a	**WG**ANKT**D**VFLLE
Gt 4a	**WG**ENES**D**VFLLN
Gt 5a	**WG**SNET**D**ILLLN
Gt6	**WG**ENET**D**VFMLE

Crystallization of DAO5 scFv in complex with the epitope HCV E2$^{529-540}$ peptides

First, we tried to co-crystallize DAO5 scFv with the peptide_J4 and peptide_JFH. However, no crystals of DAO5 scFv in complex with the peptides were obtained. Therefore, we tried to induce the crystallization of DAO5 scFv-peptide complexes by microseed matrix screening using unliganded DAO5 scFv crystals as a source of the microseed stock. Unfortunately, this microseeding technique did not promote the crystallization of DAO5 scFv-peptide complexes.

An alternative strategy to obtain crystals of protein–ligand complexes is soaking ligands into protein crystals. DAO5 scFv was previously crystallized (space group P41 21 2) and its 3D structure at 2.05 Å determined by Thomas Krey (unpublished data). These crystals allowed soaking because the analysis of crystal lattice contacts indicated that the antigen-binding site is accessible and might be able to accommodate the peptide. Moreover, the crystals had a high solvent content, which is favorable for peptide diffusion. To obtain a number of diffraction quality crystals for soaking experiments, the crystallization conditions were further refined in 24 well plates. DAO5 scFv crystals were grown by hanging-drop vapor diffusion method in drops containing 1 µl DAO5 scFv (9 mg/ml) and 1 µl of reservoir solution composed of 100 mM MES pH 6.5, 26-29% PEG 400 and 20-150 mM sodium acetate. Long rod-shaped crystals formed in all conditions during three days (**Figure 31**). DAO5 scFv crystals were transferred to the mother liquor supplemented with 0.1 mM peptide for 5 h (with peptide_J4) or overnight (with peptide_JFH). DAO5 scFv crystals were stable in the soaking solutions containing the peptides and no evident changes in DAO5 scFv crystal morphology were observed (determined by visual inspection) even after overnight soaking. Since the cryoprotectant was present in a soaking solution (26-29% PEG 400), soaked DAO5 scFv were directly flash-frozen in liquid nitrogen. Diffraction data sets at 2.0 and 1.9 Å were collected for DAO5 scFv crystals soaked in peptide_J4 and peptide_JFH solutions, respectively. (**Figure 32 and Figure 33**).

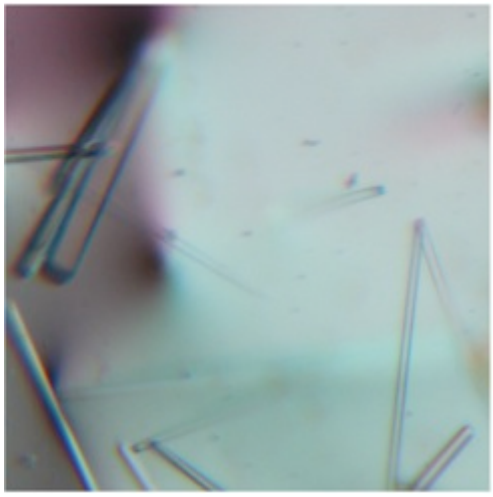


Figure 31. DAO5 scFv crystals used in soaking with the peptides.

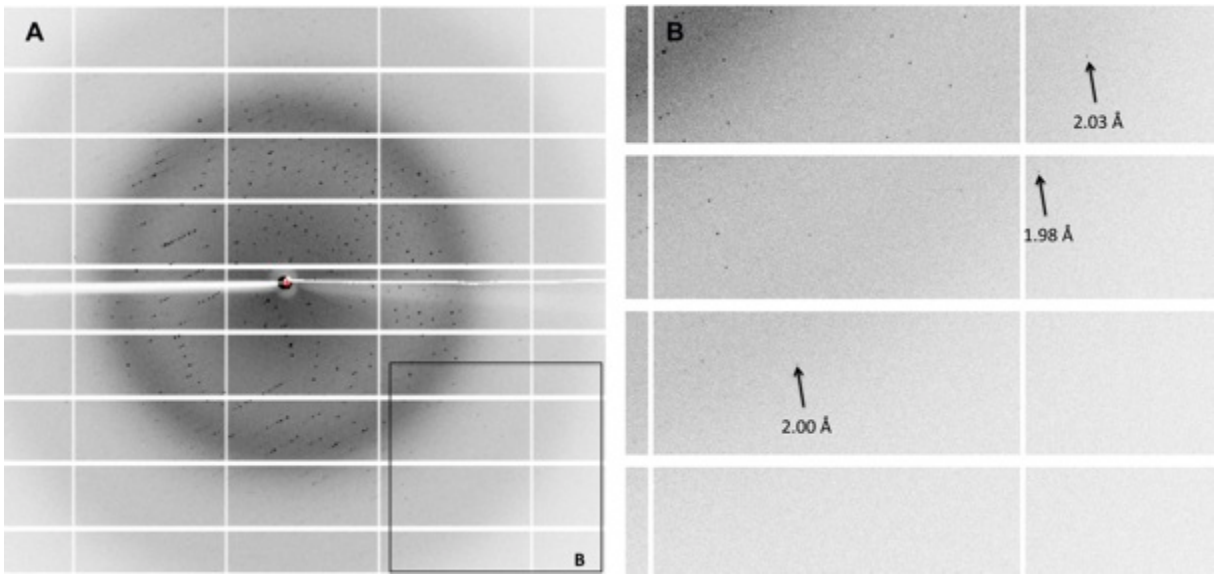


Figure 32. (A) X-ray diffraction image of the DAO5 scFv crystal soaked in peptide_J4 solution. (B) High resolution reflections of the same diffraction image and corresponding resolution.

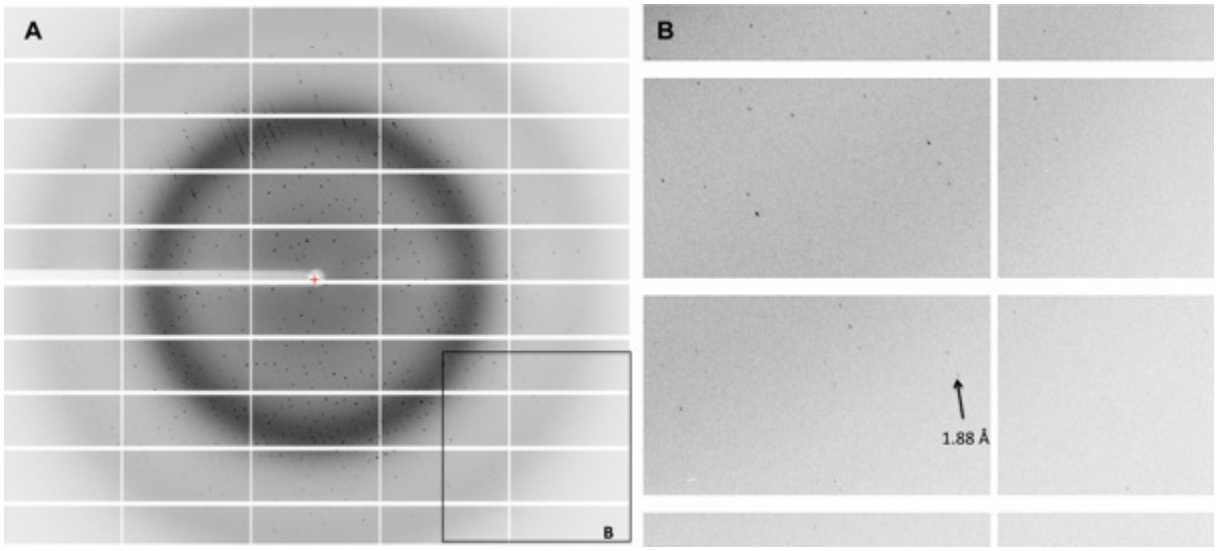


Figure 33. (A) X-ray diffraction image of the DAO5 scFv crystal soaked in peptide_JFH solution. (B) High resolution reflections of the same diffraction image and corresponding resolution.

The structures of both complexes were determined by molecular replacement method as described in detail in the section of Materials and Methods. Each crystallographic asymmetric unit contained two DAO5 scFv-peptide complexes. The final electron density map of the peptides allowed to build an atomic model of the peptide residues 530-540 and 532-540 for the first and the second copy of the peptide in the asymmetric unit, respectively (**Figure 34).** The overview of the structures of the complexes is shown in **Figure 35**. Crystallographic data collection and refinement statistics are summarized in **Table 7**.

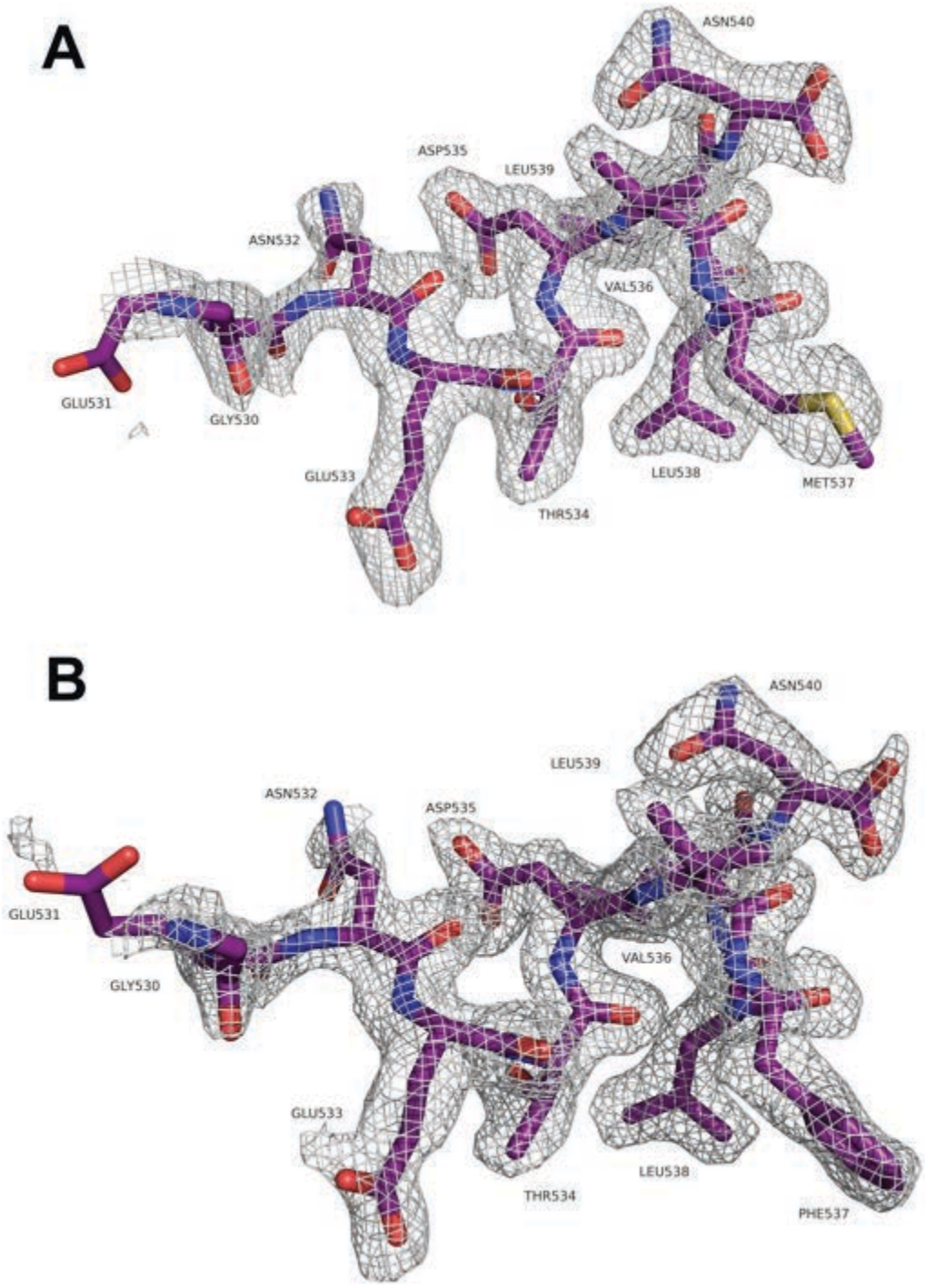


Figure 34. Composite omit maps around the peptide_J4 (A) and the peptide_JFH (B) calculated after refinement of the crystal structures of DAO5 scFv in complex with the peptide J4, and the peptide JFH, respectively. The omit maps are shown as grey mesh contoured at 1 σ level. The density of the central α-helix is well defined for both peptides and allowed unambiguous placement of the peptide.

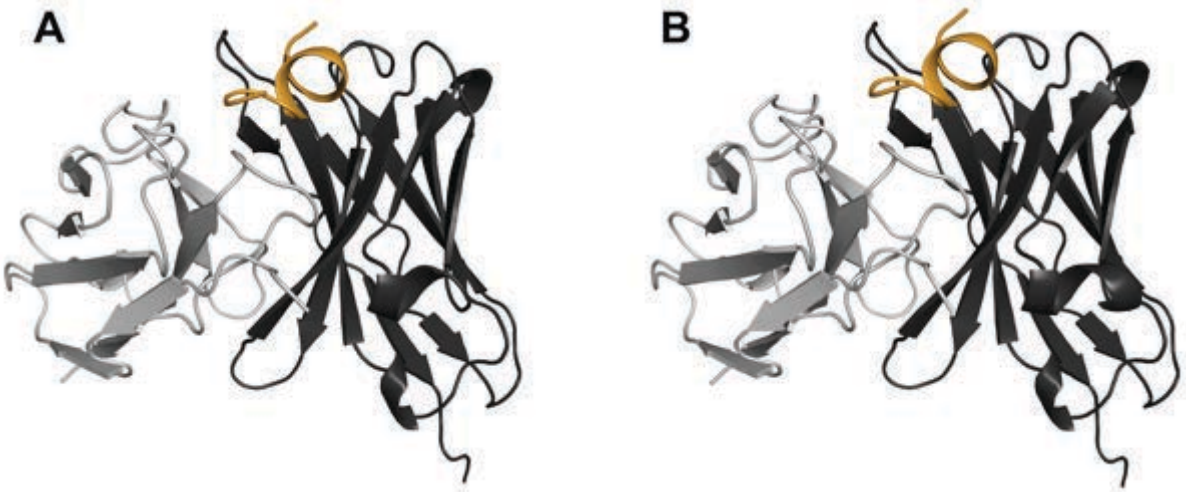


Figure 35. Crystal structures of DAO5 scFv in complex with peptide_J4 (A) and peptide_JFH (B). The crystal structures of the DAO5 scFv in complex with peptide_J4 and peptide _JFH were determined and refined to 2.0 and 1.7 Å resolution, respectively. The crystal structures are shown as cartoon. The light chain and heavy chain are colored in light grey and dark grey, respectively. The peptide is colored in yellow.

Crystallization of DAO5 Fab in complex with the epitope HCV E2$^{529\text{-}540}$ peptides

It has been shown that the structure of a protein in complex with a ligand may differ depending on whether it was obtained by soaking or co-crystallization experiments (Hiller *et al.*, 2006; Zhu *et al.*, 1999) as the binding of a ligand to a protein in solution could be different. Moreover, regardless of how the crystals of the complex were obtained, crystal-packing contacts could influence the conformation of the ligand.

In order to exclude a putative influence of the crystalline environment on the peptide conformation, I also performed co-crystallization trials for the DAO5 Fab in complex with the J4 peptide. To form the DAO5 Fab/peptide complex, 6-fold molar excess of the peptide dissolved in 20 mM Tris pH 9 was added to the protein solution (18 mg/ml of Fab+3 mg/ml peptide). The complex was incubated overnight at 4°C. Crystal screening was performed using commercially available crystallization screening kits as described in Section 2. A cluster of needle-like crystals was obtained in condition 10% PEG8k 200 mM ZnAc, 100 mM pH 6.5.

Since the quality of the crystals was not suitable for X-ray diffraction measurement, the crystal growth condition was further optimized in 24-well plates by hanging-drop vapor diffusion method by fine-tuning pH, salt and PEG concentration as well as protein

concentration. In addition, streak-seeding experiments were performed. Rod shaped crystals were obtained after the streak-seeding above reservoir solutions composed of 8% PEG8k, 350 mM ZnAc, 100 mM pH 6.5 as well as 8% PEG8k, 350 mM ZnAc, 100 mM Tris pH 8. The crystals from these two conditions were cryoprotected in 20% glycerol and subjected to X-ray diffraction experiments on a home source, a Rigaku MicroMax-O7 copper anode generator with a Mar345 image plate detector (a Platform for crystallography at the Pasteur Institute). However, all tested crystals diffracted just to ~8 Å, which was not sufficient for the structure determination.

To promote crystal growth, microseed matrix-screening was performed with the seed stock prepared from the crystals obtained in condition 11% PEG 8k, 350 mM ZnAc, and 100 mM pH6.5. The crystals grew in ~100 different conditions (mostly PEG or salt based). The conditions that yielded the morphologically best crystals were selected for refinement in 24 well plates. The diffraction-quality crystals obtained in conditions containing 10% PEG 20K, 100 mM $MgCl_2$, Tris pH 8.5; 600 mM $(NH_4)_2SO_4$, 100 mM MES pH 6 ; 600 mM $LiSO_4$, 100 mM Tris pH 8.5, 50 mM $NiCl_2$ were cryoprotected using reservoir solutions supplemented with 25% glycerol and tested for X-ray diffraction at the synchrotron Soleil beamline Proxima I and the Swiss Light Source beamline PXI. The crystals grown in condition 600 mM $LiSO_4$, 100 mM Tris pH 8.5, 50 mM $NiCl_2$ diffracted X-rays to ~3.7-4 Å, but the diffraction images were difficult to index due to multiple crystal lattices.

All co-crystallization experiments of DAO5 Fab with peptide J4 defined above were performed with DAO5 Fab containing a C-terminal double Strep tag. The affinity tags sometimes interfere with crystal lattice formation. Therefore, the C-terminal double Strep tag, preceded by an enterokinase recognition site, was removed from DAO5 Fab by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). The amount of enterokinase required to achieve complete removal of the tag was optimized in a small-scale reaction (Section 2). The reaction was scaled up in a linear manner and the DAO5 Fab without the Strep-tag was purified from the reaction mixture (Section 2). DAO5 Fab-peptide J4 complex was formed and the crystallization screening performed as described previously. Rod-shaped diffraction quality crystals appeared after 5 days in a drop containing 0.2 µl of the complex (10 mg/ml of Fab+1.72 mg/ml peptide in 10 mM TRIS pH 8.0, 150 mM NaCl) mixed with an equal amount of reservoir solution containing 20% PEG3350 and 200 mM sodium thiocyanate (**Figure 36)** The crystals were flash-frozen in liquid nitrogen after transferring them to a cryo-protective solution containing the mother liquor and 20% (v/v)

glycerol. A diffraction data set at 1.5 Å was collected from a single crystal on beamline Proxima I at the synchrotron Soleil (**Figure 37).**

Figure 36. DAO5 Fab/peptide J4 crystal obtained by co-crystallizing the peptide J4 with the DAO5 Fab from which the Strep-tag was removed by specific proteolytic cleavage with enterokinase. The crystal was used to collect an X-ray diffraction data set at 1.5 Å on a beamline Proxima I at the synchrotron Soleil.

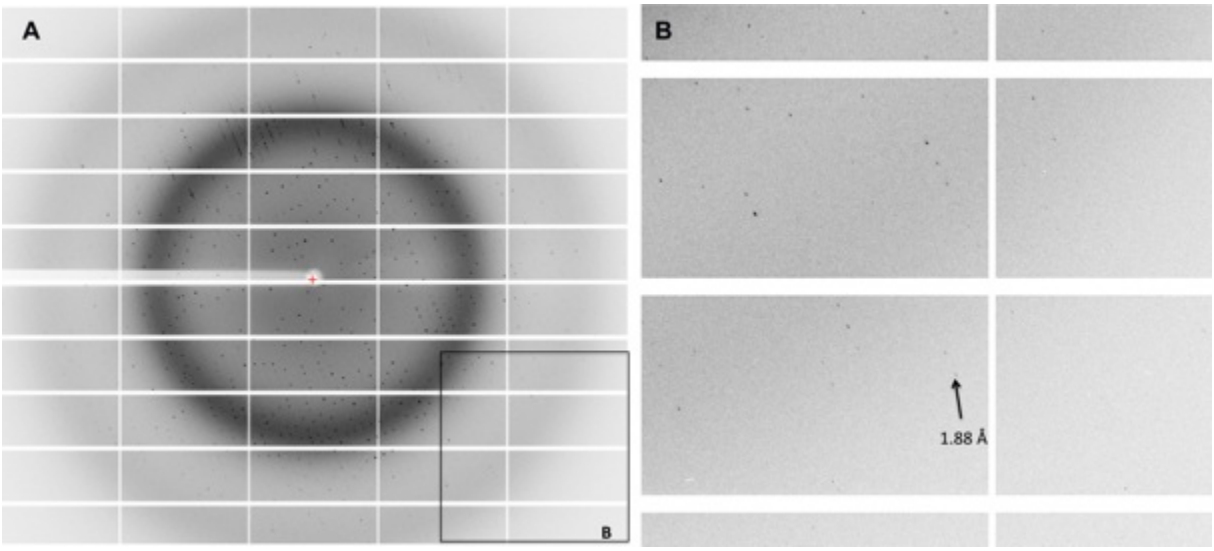


Figure 37. (A) X-ray diffraction image of the DAO5 Fab crystal co-crystallized with the peptide_J4. (B) High resolution reflections of the same diffraction image and corresponding resolution.

The DAO5 Fab-peptide J4 crystal belonged to space group P1 21 1 with unit cell parameters a=48.678, b=80.717, c=54.565, α=90.00, β=95.64, γ=90.00 and contained one complex per asymmetric unit. The Matthews' coefficient (V_m) for the DAO5 Fab-peptide J4 crystals was estimated to be 2.18 $Å^3$/Da (corresponding to a solvent content of 43.51%). Usually more tightly packed crystals (containing lower solvent content) tend to diffract X-rays to higher resolution (Kantardjieff & Rupp, 2003). This observation held true for DAO5

Fab-peptide and DAO5 scFv-peptide crystals (solvent content 43.51% and 63% accordingly) with DAO5 Fab-peptide-J4 crystal diffracting X-rays better. The data was integrated and scaled as described in Materials and Methods.

The resolution limit of 1.7 Å was chosen for the refinement of DAO5 Fab-peptide_J4 model. Though well-defined positive electron density for a peptide was already visible in proximity of the CDRs after molecular replacement, the atomic model of the peptide was built after the coordinates of the Fab were fully refined. The final electron density allowed unambiguous fitting of the J4 peptide residues 532-540. The overview of the structure of the complex is shown in **Figure 38.** Crystallographic data collection and refinement statistics are summarized in **Table 7**.

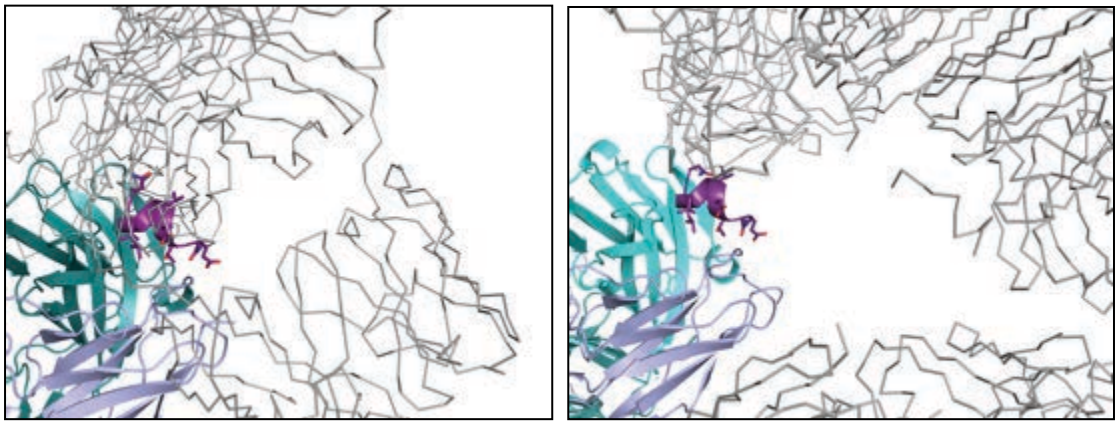
Figure 38. Crystal structure of DAO5 Fab in complex with peptide_J4. The crystal structure of the DAO5 Fab in complex with peptide_J4 was determined and refined 1.5 Å resolution and is shown as cartoon. The light and heavy chains are colored in light and dark grey, respectively. The peptide is colored in yellow.

Table 7. Data collection and refinement statistics for the DAO5 scFv and DAO5 Fab crystals in complex with the indicated peptides. Values in parentheses represent those in the highest resolution bin.

	DAO5 scFv-peptide J4	DAO5 scFv-peptide JFH	DAO5 Fab-peptide J4
Data collection			
Space group	P41 21 2	P41 21 2	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	155.94 155.94 61.51	155.69 155.69 61.65	48.68 80.71 54.57
a, b, g (°)	90.00 90.00 90.00	90.00 90.00 90.00	90.00, 95.64 90.00
Resolution (Å)	49.31-2.00 (2.11-2.00)	49.23-1.90 (2.00-1.90)	48.44-1.70 (1.79-1.70)
Solvent content (%) (molecules per asymmetric unit)	63 (2)	63 (2)	43.51 (1)
R_{merge}	0.107 (0.516)	0.137 (0.614)	0.04 (0.228)
Total number of observations	298752 (18087)	261291 (37230)	255101 (36579)
Total number of unique reflections	50405 (6158)	58977 (8629)	44810 (6417)
I / sI	10.3 (1.9)	9.5 (3.2)	27.6 (6.8)
Completeness (%)	97.6 (83.5)	98.5 (99.8)	97.1 (95.6)
Redundancy	5.9 (2.9)	4.4 (4.3)	5.7 (5.7)
Refinement			
Resolution (Å)	20.16-2.00: (2.05-2.00)	38.00-1.90 (1.95-1.90)	48.44-1.70 (1.74-1.70)
No. reflections	47116	58950	44787
R_{work} / R_{free}	0.1979/0.2176	0.1875/0.2039	0.1774/0.2032
No. atoms			
Protein	3727	3728	3361
Water	269	326	329
B-factors			
Wilson B-factor (Å^2)	28.69	19.77	19.53
Average B-factor (Å^2)	28.67	20.84	20.02
R.m.s. deviations			
Bond lengths (Å)	0.010	0.010	0.010
Bond angles (°)	1.08	1.06	1.12
Ramachandran statistics[#]			
Favored (%)	96.62	97.03	97.9
Number of outliers	0	0	1

[#] Ramachandran statistic according to Molprobity server.

The crystal packing interfaces in proximity of the peptide_J4 differ considerably in the DAO5 Fab-peptide J4 and DAO5 scFv-peptide JFH crystals. DAO5 Fab-peptide J4 crystal has denser packing in comparison with the DAO5 scFv-peptide crystals. As a result, the peptide J4 environment in the DAO5 Fab-peptide crystal is crowded by symmetry-mates with some of them packed close to the antigen-binding site (**Figure 39**).



Crystal packing of DAO5 Fab/peptide J4
Space group: P 1 2_1 1

Crystal packing of DAO5 scFv/peptide J4
Space group: P 4_1 2_1 2

Figure 39. Crystal packing of interfaces for DAO5 Fab and DAO5 scFv in complex with the peptide J4. One molecule of the complex is shown as cartoon (with side chains displayed for the peptide J4). The light chain and the heavy chain are colored in light blue and cyan, respectively, and the peptide is colored purple. Symmetry mates are shown in light grey as Cα traces. The peptide J4 has completely different packing interfaces in DAO5 scFv/peptide J4 and DAO5 Fab/peptide J4 crystals but adopts the same conformation, indicating that the crystal packing does not have an effect on the peptide conformation.

Nevertheless, the peptide J4 conformation is very similar in both crystal forms as indicated by the root mean square deviation (rmsd) values calculated between each pair of aligned residues 532-540 (**Figure 40**). The main difference observed between the two peptide structures is the different side chain conformations of residues M537 and N540. It is widely accepted that if the same ligand conformation is found in multiple protein-ligand crystal forms, the conformation is not likely to be induced by the crystalline environment and thus represents a biologically relevant interaction. In conclusion, analysis of the crystal packing interfaces revealed unrelated interfaces for Fab and scFv complexes respectively, indicating that the J4 peptide reflects the conformation that is recognized by the functional monoclonal antibody in the context of the native glycoprotein.

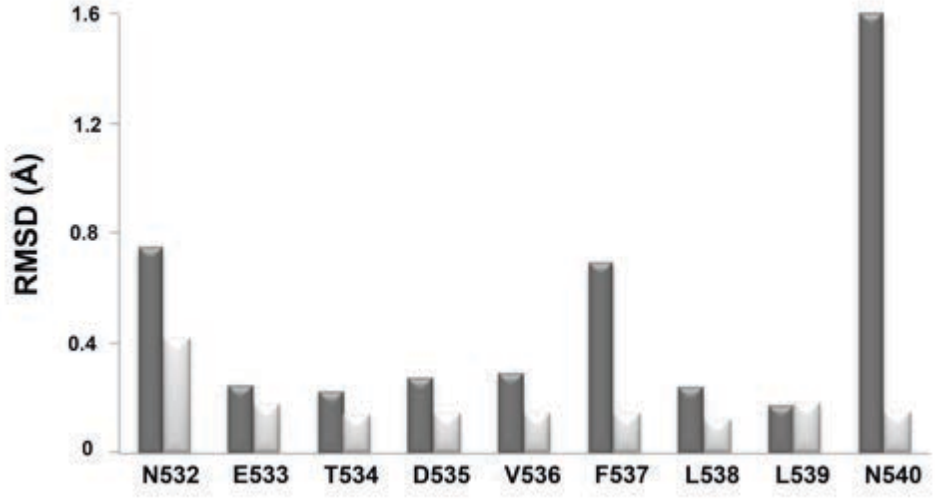


Figure 40. Root mean square deviation (rmsd) calculated after superposition of peptide J4 in complex with scFv and Fab, respectively using Chimera including all atoms (dark grey) or main chain atoms only (light grey) in the calculation and represented per residue. The peptides were superposed using Superpose 1.0.

Comparison of the epitope HCV E2$^{529\text{-}540}$ peptide structures

The crystallization of DAO5 scFv-peptide and DAO5 Fab peptide complexes yielded in total five atomic models of the peptides (two copies of each scFv complex per asymmetric unit and one copy of the DAO5 Fab in complex with the J4 peptide) that are listed in **Table 8**.

Table 8. Summary of the obtained crystal structures of the epitope HCV E2$^{529\text{-}540}$ peptides.

Number	Peptide	Crystal	Modeled residues	Average B factor for the peptide residues 532-540
1	Peptide J4	DAO5 scFv-peptide_J4 (asymmetric unit 1)	530-GENETDVMLLN-540	32.66
2	Peptide J4	DAO5 scFv-peptide_J4 (asymmetric unit 2)	532-NETDVMLLN-540	36.53
3	Peptide JFH	DAO5 scFv-peptide_JFH (asymmetric unit 1)	530-GENETDVFLLN-540	23.38
4	Peptide JFH	DAO5 scFv-peptide_JFH (asymmetric unit 2)	532-NETDVMLLN-540	26.66
5	Peptide J4	DAO5 Fab-peptide_J4	532-NETDVMLLN-540	25.46

Since the peptides differed in the number of modeled residues at the N-terminal end, only residues that were present in all five peptide structures (aa 532-540) could be used for comparison. Superposition of the segment 532-540 from all five available crystal structures of

the peptide epitope (two copies of each scFv complex per asymmetric unit and one copy of the DAO5 Fab in complex with the J4 peptide) revealed an rmsd of 0.114 Å calculated over the backbone atoms of the peptide, which confirms identical peptide conformations (**Figure 41**).

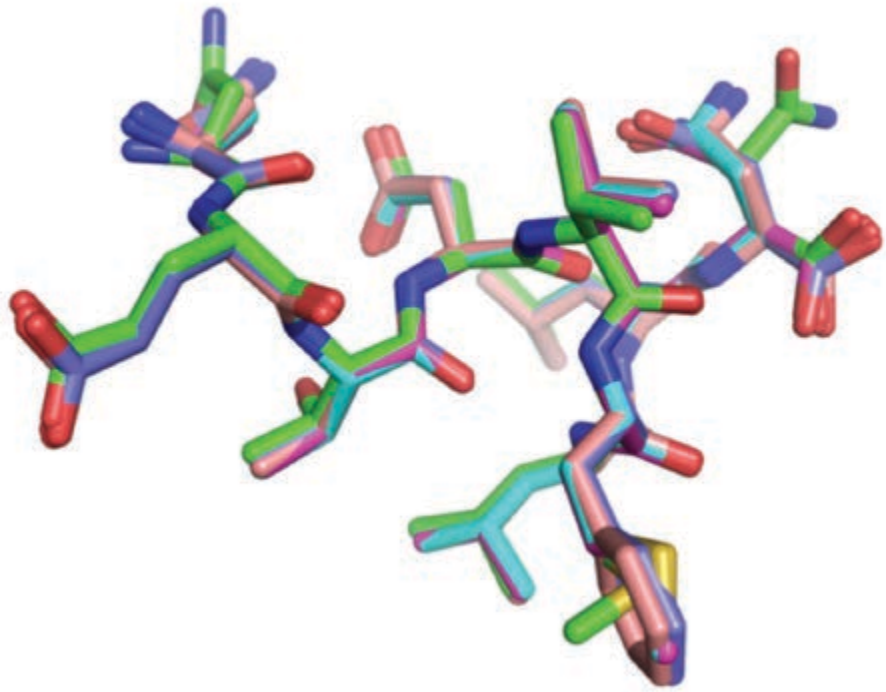


Figure 41. Superposition of five atomic models of the peptides (two copies of each scFv complex per asymmetric unit and one copy of the DAO5 Fab in complex with the J4 peptide) using Superpose 1.0.

The temperature factor (B-factor) analysis of the segment 532-540 of all five independent peptide structures indicates that peptide J4 (Nr.5 in **Table 8**) of DAO5 Fab-peptide crystal has the lowest average B-factor value. Also, it showed that the longer peptides (Nr.1 and Nr.3) encompassing aa 530-540 in the first asymmetric unit of DAO5 scFv-peptide crystals have the lower average B-factor values in comparison with peptides Nr.2 and Nr.4 in the second asymmetric unit, implying that they are more ordered.

Peptide conformation and its binding to DAO5 scFv

We did not observe electron density for W529 in any of the crystal structures, which is consistent with this residue not being part of the DAO5 mAb epitope, as shown by alanine

scanning mutagenesis of E2. The electron density for the residues N532-N540 was clearly defined in all five complexes. The residues 530-531 are visible only in the electron density map of the J4 peptide or the JFH-1 peptide complex with DAO5 scFv. Moreover, the electron density for G530 and E531 in both peptides is present only in the same more ordered counterpart of the scFv-peptide complex in the asymmetric unit. All together, it suggests that the N-terminal residues G530 and E531 are highly flexible and most likely stabilized by crystal packing of scFv molecules. Thus we conclude that the interactions they form with DAO5 scFv in the crystal structures are nonspecific and do not participate in the natural interactions of DAO5 mAb with its epitope. Therefore, they will not be discussed further in this study. The peptide forms one α-helical turn comprising the residues 535-539 (535-DVM/FLL-539) which is stabilized by canonical intramolecular hydrogen bonds of an α-helix (**Table S2**). At the N-terminus, the residues 533-534 make a hydrogen bonded turn, and the chain continues as an extended loop (residues G530-N532).

The J4 and JFH-1 peptide interaction with the paratope buries an area of 734 $Å^2$ and 723 $Å^2$ of the peptide and 645 $Å^2$ and 656.2 $Å^2$ of the antibody, respectively. For both peptides, the total buried surface area amounts to 1379 $Å^2$ in the interface, which is very close to the interface (1680±260 $Å^2$) described for other antibody-antigen complexes (Lo Conte *et al.*, 1999). The shape complementarity index is 0.810 and 0.771, respectively, for the J4 and JFH-1 peptides, similar to indices reported for other antibody-peptide complexes (Lawrence & Colman, 1993). In both structures, the peptides bind to a cleft between the variable regions of the heavy chain (HC) and the light chain (LC) and interact with all six CDR loops of the antibody.

As illustrated by the analysis of surface hydrophobicity of epitope and paratope, the α-helical turn of the peptide establishes mainly hydrophobic interactions with the paratope (**Figure 42**). The hydrophobic residues M/F537, L538 and L539 within the short α-helical turn insert deeply into a hydrophobic groove formed by aromatic and aliphatic residues within the complementarity determining region 3 of the heavy chain (CDR-H3) (F99, Y103, P104 and Y105), CDR-H2 (A59) and all CDR loops of the LC (Y32, Y50, W94 and L96) (**Table S1**). Calculation of the solvent-accessible surface area reveals that more than 94% of the solvent-accessible surface area of the residues 537M/F-539L is buried in this hydrophobic groove (Figure 14). Moreover, the side chain of K52 within the CDR-H2 loop forms an extensive hydrogen-bonding network with the main chain carbonyl groups of residues V536, M/F537 and N540, which further stabilizes the interaction of the α-helical turn with the heavy chain.

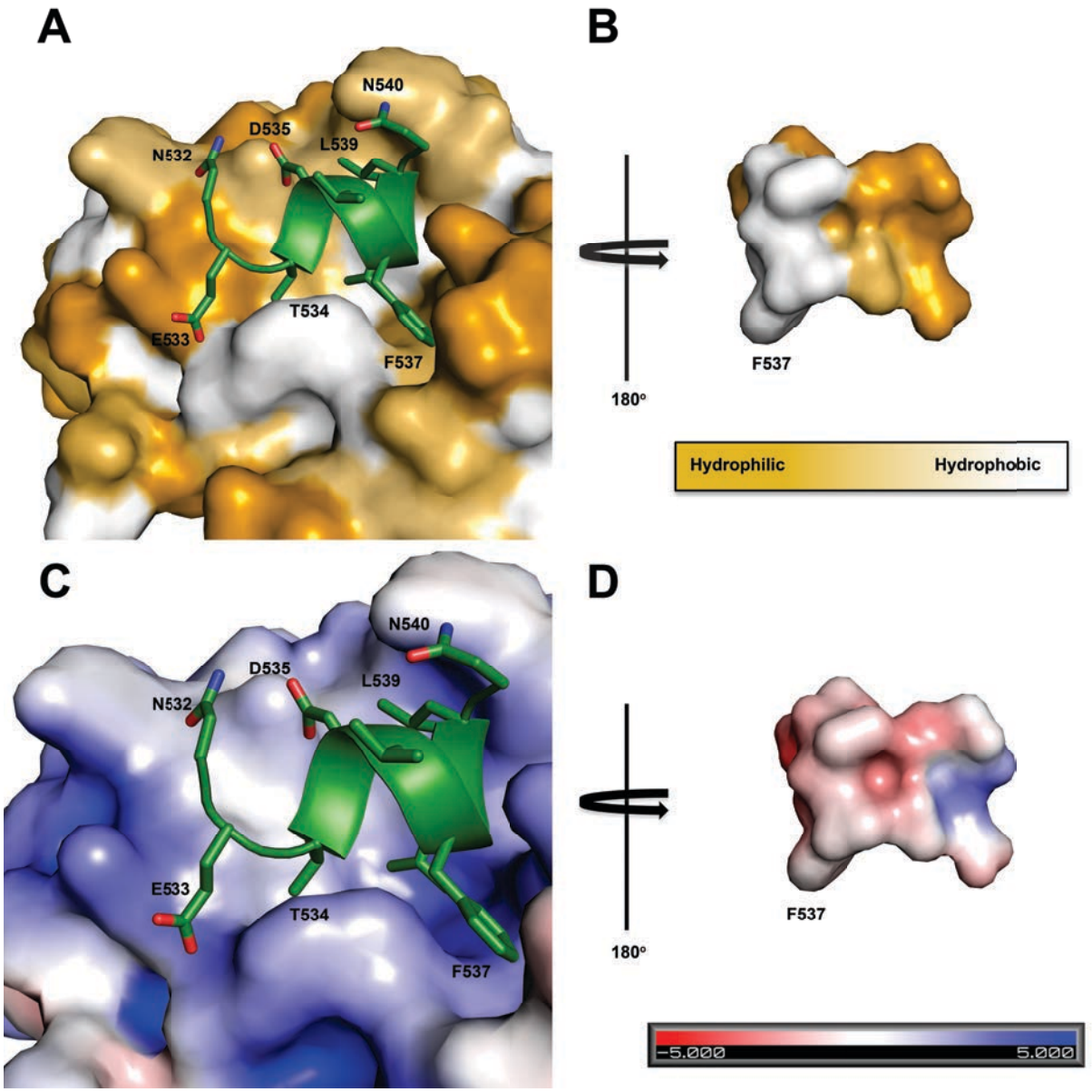


Figure 42. Interaction between DAO5 scFv and the epitope peptide_JFH. DAO5 scFv (A and B) is shown as molecular surface. The peptide (A and B) is shown as a cartoon with side chains displayed as sticks and colored by atom-type (green, red and blue for carbon, oxygen and nitrogen, respectively). (A) The molecular surface of DAO5 scFv and the epitope peptide are colored according to a normalized hydrophobicity scale from white (hydrophobic) to bright yellow (hydrophilic). (B) Electrostatic potential [−5 kT/e (red) to 5 kT/e (blue)] across the molecular surface of DAO5 scFv and the peptide epitope calculated using the adaptive Poisson-Boltzmann solver. The molecular surface of the peptide in A and B is shown looking from the paratope.

In the case of JFH-1 peptide, in which M537 is replaced by F537, the side chain of K52 is also involved in a cation-π interaction with F537. Cation–π interaction is a predominantly electrostatic interaction between the face of an electron-rich π –system (e.g. aromatic amino acid side chain) and an adjacent cation (e.g. protonated arginine or lysine side chain), which has been discovered to play an important stabilizing role in protein-protein

interaction (Gallivan & Dougherty, 1999). In addition, F537 is engaged in aromatic stacking interaction with W94 within CDR-H3. F537 inserts in the same hydrophobic pocket as M537 with 90.2% of solvent-accessible surface area buried (**Figure 43**). M537 has a different side chain conformation in the complex with DAO5 scFv and DAO5 Fab, suggesting a higher structural flexibility within this pocket than the bulkier F537. This flexibility is further supported by the higher B-factor for the M537 side chain compared to other residues within the α-helical turn. Also, it is known that methionine residues have very often several conformations, even when they are buried in the hydrophobic core of a globular protein, whereas phenylalanines do not have such features.

Heavy chain and light chain contribute differently to the binding of the peptide segment V536-N540 into the hydrophobic pocket, burying ~351 $Å^2$ and ~100 $Å^2$ solvent-accessible surfaces of these residues, respectively. This confirms that the heavy chain dominates the interaction between the α-helical turn and the hydrophobic antigen-binding groove.

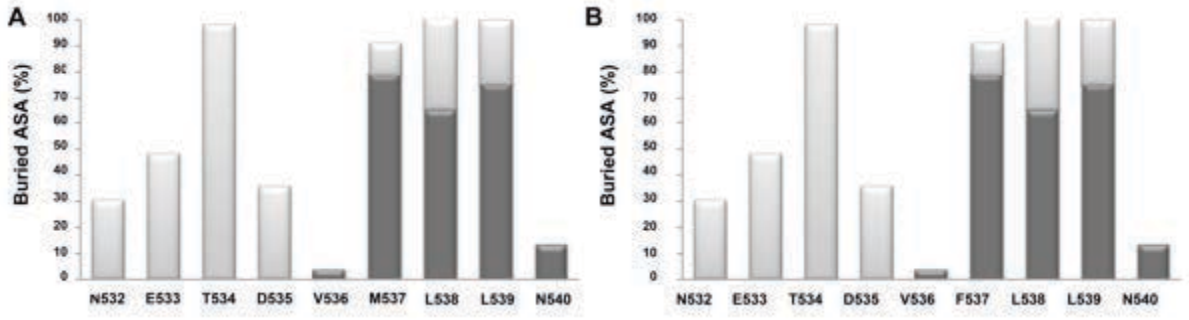


Figure 43. Percentages of accessible surface area (ASA) of peptide J4 (A) and peptide JFH (B) buried in the complex with DAO5 scFv, calculated using PISA and represented per residue as stacked columns for heavy (dark grey) and light (light grey) chains of DAO5 scFv.

The N-terminal peptide residues N532-D535 make contacts exclusively with the light chain. The contact surface here is more planar, which results in a smaller buried surface area on the peptide (~200 $Å^2$). Both side chain carboxyls of D535 form hydrogen bonds with the hydroxyl group of Y52, consistent with the results of E2 alanine-scanning mutagenesis identifying D535 as an essential contact residue for binding of the DAO5 mAb. Since the hydrophobic V536 side chain is completely exposed and not involved in any antibody interactions, D535 is likely required to stabilize the interaction of the solvent-exposed part of the α-helical turn with the paratope. Moreover, its position in the junction between N-terminal

and C-terminal portions of the peptide suggests that the interaction of D535 with Y52 anchors the epitope in the binding position.

The other residues in the N-terminal stretch N532-D535 also form hydrogen bonds to the light chain. E533 establishes a salt bridge with K93. In the J4 peptide complex with DAO5 scFv, it also forms two main-side chain hydrogen bonds with W94. In the J4 peptide co-crystal structure, an extra hydrogen bond between N532 and N92 is observed. T534 is the only residue within N532-D535 peptide portion fixed in a deeper pocket of the antigen-binding site. Within this pocket, the side chain of T534 is stabilized by a hydrogen bond with the carbonyl oxygen of T91. Upon antibody binding, 99% of its solvent-accessible surface area is buried in the interaction interface with the light chain (**Figure 43**). T534 substitution to S534 most likely does not disrupt this hydrogen-bonding pattern: E2 of HCV genotype 4 possesses this substitution and is still recognized by mAb DAO5.

The B-factor analysis indicates a stable and strong interaction with the paratope (**Figure 43**). In contrast, the residues at the termini of the peptide (N532, E533 and N540) have higher B-factors compared to all other peptide residues. The root mean square deviation (rmsd) between the two peptides is also higher at the N- and C-termini (**Figure 45**). At least for residues N532 and N540, an increased flexibility can be expected since they are normally glycosylated in native E2. The modeling of the hypothetical glycan chains attached to N532 and N540 reveals that the helical peptide conformation is compatible with the N-linked glycosylation of those residues in the native glycoprotein (**Figure 46**).

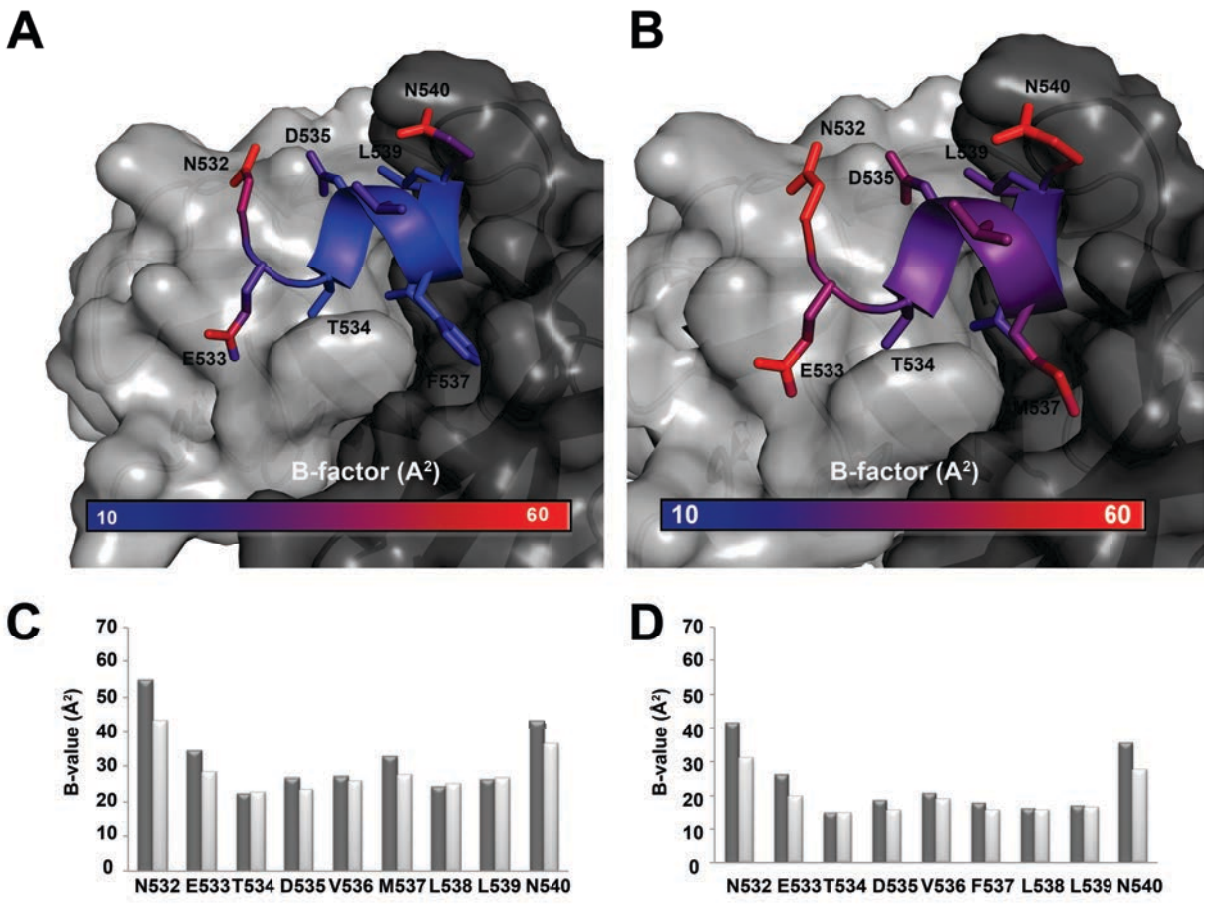


Figure 44. Peptide J4 and peptide JFH colored by temperature factor (B-factor) per atom according to a scale from blue (low B-factor) to red (high B-factor). The peptide (A and B) is shown as cartoon with side chains as sticks. DAO5 scFv is shown as molecular surface with the light chain and heavy chain colored in light grey and dark grey, respectively. Average temperature factors of peptide J4 (C) and peptide JFH (D) in complex with DAO5 scFv (light grey) calculated per residue including all atoms (dark grey) or main chain atoms only (light grey).

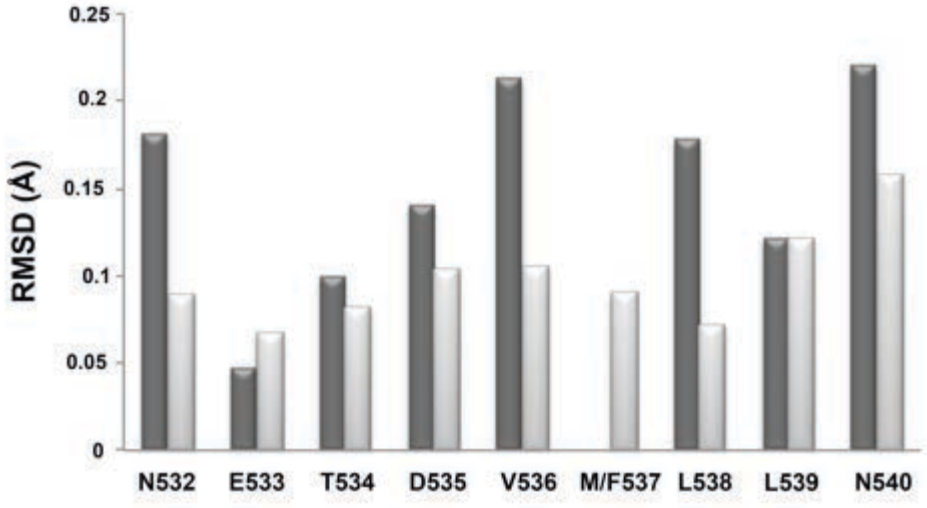


Figure 45. Root mean square deviation (rmsd) upon superposition of the peptide J4 and peptide JFH in complex with DAO5 scFv calculated using Chimera including all atoms (dark grey) or main chain atoms only (light grey) in the calculation and represented per residue. The peptides were superposed using Superpose 1.0.

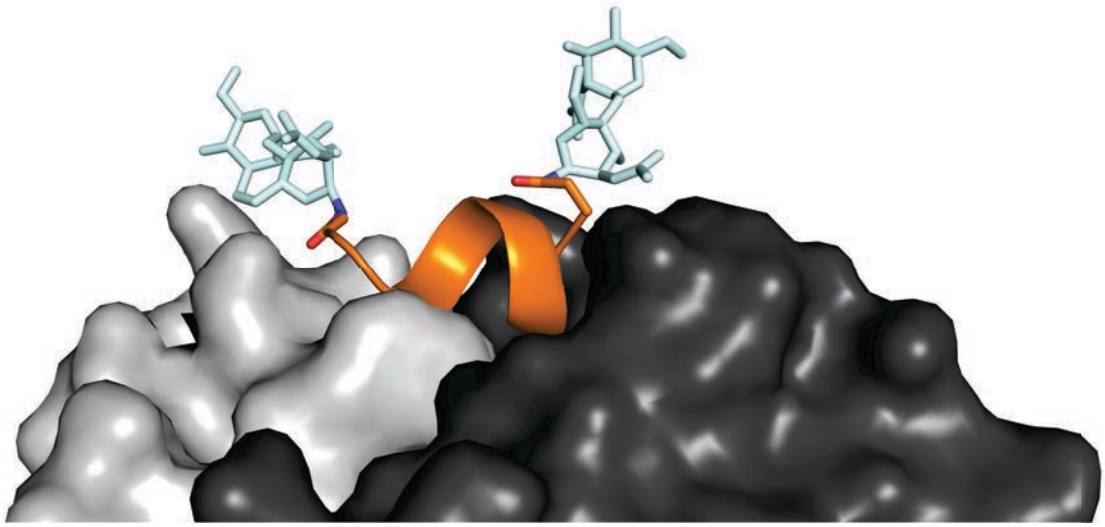


Figure 46. Compatibility of the peptide_J4 conformation in complex with DAO5 scFv with N-linked glycosylation. Hypothetical glycan chains containing two N-acetylglucosamin moieties (light blue) are modeled to be attached to ND2 atoms of N532 and N540. The peptide is displayed as cartoon and colored in orange. The side chains of N532 and N540 are displayed as sticks and ND2 atoms, to which the glycans are linked, and OD1 atoms are colored in blue and red, respectively.

The stark difference between conformations of aa 532-540 in the E2 core fragment and DAO5/peptide crystal structures

The conformation of residues 532-540 in the JFH-1 peptide structure was compared to the one of the corresponding peptide in E2 crystal structure in complex with the Fab derived

from neutralizing antibody AR3C (PDB 4MWF) (Kong *et al.*, 2013). AR3C Fab also requires D535 for binding to E2 (Law *et al.*, 2008). In the E2 structure published by Kong and colleagues, amino acid residues 532-540 adopts a completely different extended conformation (with aa 536-538 forming a β-strand) than the equivalent peptide in complex with DAO5 scFv (with aa 535-539 forming an α-helix) (**Figure 47**). The extended conformation of the peptide in the context of the E2 core fragment spans 20.6 Å while the helical conformation observed in the epitope peptide in complex with DAO5 scFv spans only 11.6 Å, respectively. T534, F537 and L539 that inserts deeply in the antigen-binding groove of DAO5 are completely buried in E2 core meaning that these residues would not be accessible for DAO5 binding. Thus, a change of conformation of this E2 core region would be required in order for DAO5 to bind to the protein. Intriguingly, a similar situation has been observed for the epitope II as the binding of mAbs #8 and #12 to E2 core is also impossible unless there is a change of conformation of the epitope from the closed to the open conformation (see Introduction) (Deng *et al.*, 2014).

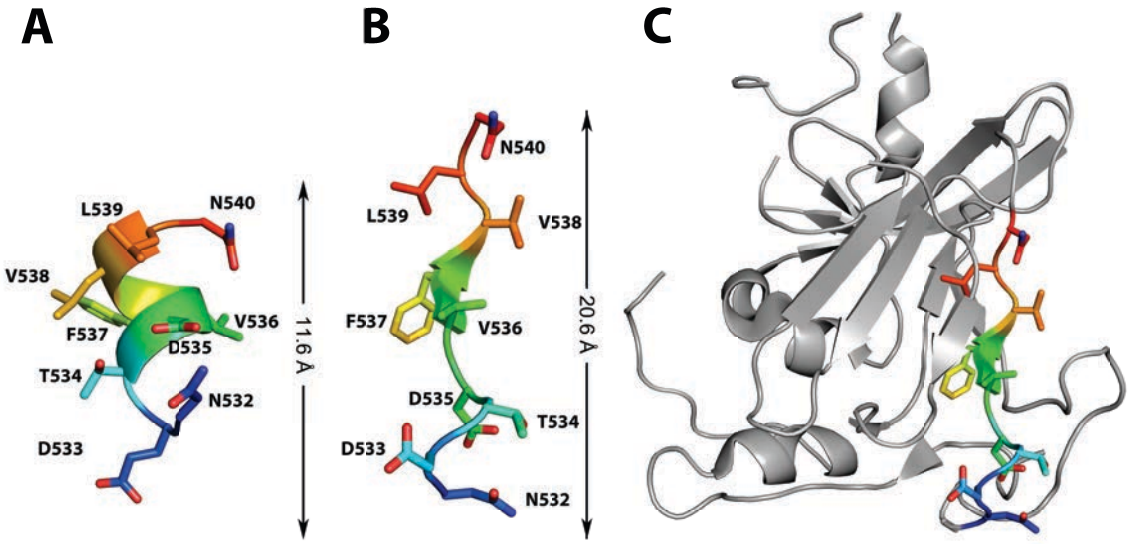


Figure 47. The conformation of residues 532-540 in the JFH-1 peptide structure and the corresponding peptide in the context of an E2 core fragment (PDB 4MWF). (A and C) Epitope aa 532-540 conformation in E2 core fragment. Note that F537 and L539 residues are buried in E2 crystal structure indicating that it must become exposed in E2 in order for DAO5 mAb to bind. (B) Epitope aa 532-540 (peptide_JFH) conformation observed in DAO5 scFv/peptide_JFH structure. All side chain residues of the epitope aa 530-540 are displayed as sticks and colored by atom-type (orange and red for oxygen and nitrogen, respectively, carbon atoms are ramp-colored from N-terminus (blue) to C-terminus (red) through green. E2 core structure (except aa 530-540) is depicted as cartoon and colored in grey.

Two populations of sE2ΔHVR1 can be distinguished by DAO5 and e137 antibody fragments

Due to observed discrepancy of the conformation of the epitope spanning aa 532-540 in E2 core/AR3C Fab and DAO5 scFv/peptide_JFH complexes, I decided to verify if DAO5 scFv is able to bind to the E2 core. For these experiments, I selected the recombinant soluble E2ΔHVR1 (sE2ΔHVR1) expressed in S2 cells. The recombinant sE2ΔHVR1 was chosen because it lacks HVR1 region but comprises the entire E2 core and, thus, is similar to the construct crystallized by Kong *et al*. The protein elutes as a single monomeric peak in SEC (**Figure 48**). The correct folding of sE2ΔHVR1 was verified by binding of the Fabs derived from conformation-sensitive mAbs (described later). Also, it has been previously described to induce broadly neutralizing antibodies after immunization, which is a strong indication that the protein is correctly folded (Tarr *et al.*, 2013).

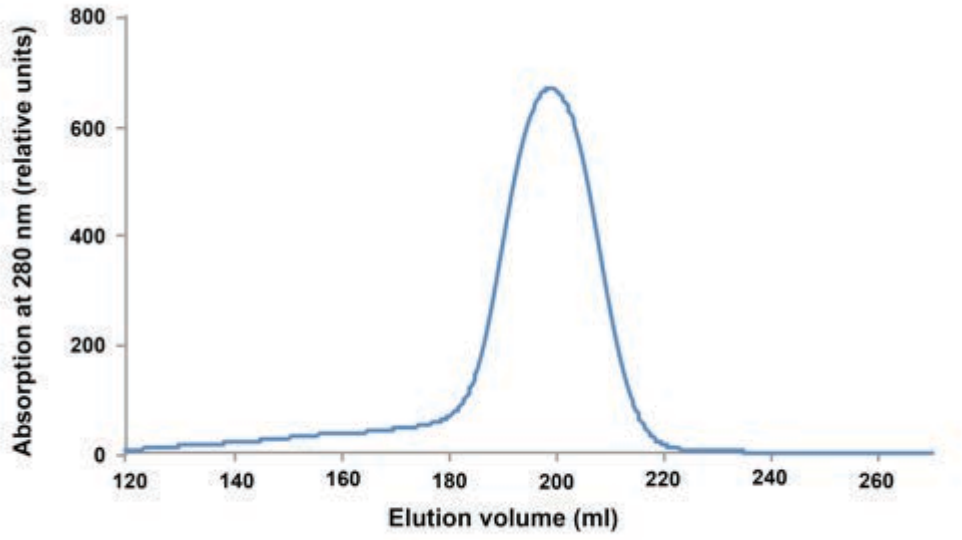


Figure 48. Size-exclusion chromatography (SEC) of sE2ΔHVR1. Separation by SEC was performed using HiLoad 26/60 Superdex 200 column (GE Healthcare Life Sciences) at a flow speed of 1 ml/min in 10 mM Tris-HCl pH 8.0, 150 mM NaCl. Protein elution was monitored by absorbance at 280 nm (blue curves). The chromatogram reveals a single major peak for sE2ΔHVR1, corresponding to a monomeric protein (~39 kD not counting the sugar chains).

Binding of DAO5 scFv to sE2ΔHVR1 was tested in a pull-down assay. sE2ΔHVR1 was bound to a Streptactin mini column and then an equimolar amount of DAO5 scFv lacking the Strep-tag was added. The eluted fraction was analyzed by SDS-PAGE under non-reducing conditions (**Figure 49**). Two bands, one representing sE2ΔHVR1 and another one DAO5 scFv were observed. However, the DAO5 scFv band was much weaker than sE2ΔHVR1

indicating that only a small fraction of added DAO5 scFv was bound to sE2ΔHVR1. Thus, the interaction between sE2ΔHVR1and DAO5 scFv in pull-down assay was not stoichiometric suggesting that a fraction of sE2ΔHVR1 might adopt the conformation not compatible with binding of DAO5 scFv.

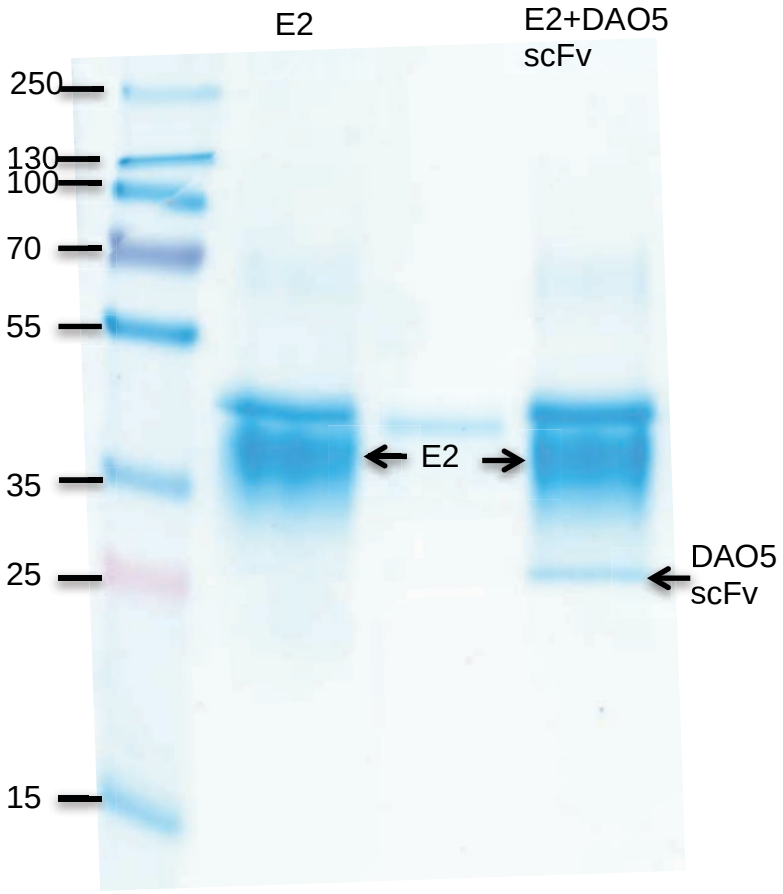


Figure 49. Pull-down experiment showing that sE2ΔHVR1 specifically reacts with DAO5 scFv. sE2ΔHVR1 was affinity loaded onto a Streptactin column, DAO5 scFv lacking the tag was passed through the column and the proteins were eluted from the column after washing. Elution fraction was analyzed by SDS-PAGE under non-reducing conditions (lane 4) and stained by Coomassie Blue. The bands corresponding to both the sE2ΔHVR1 and DAO5 scFv were observed in the elution fraction indicating DAO5 scFv binding to sE2ΔHVR1. Lane 1: Page Ruler Prestained Marker. Lane 2: sE2ΔHVR1. Lane 3: empty, lane 4: elution fraction from the column.

These results together with the observed discrepancy of the conformation of the epitope spanning aa 532-540 in E2 core/AR3C Fab and DAO5 scFv/peptide_JFH complexes provided some implications that this epitope can exist in different conformations in the E2 glycoprotein. To further explore this interpretation, I chose to repeat the above-described experiment including a Fab fragment which potentially recognizes a different conformation of aa 532-540 than DAO5 scFv.

As we did not have AR3C Fab, we selected a neutralizing conformation-sensitive Fab

e137 (Perotti *et al.*, 2008) as a substitute. Alanine scanning mutagenesis of E2 has previously indicated that both Fabs use similar residues including D535 for binding to E2 (Law *et al.*, 2008) and as a result should recognize closely related epitopes. Since both DAO5 scFv and e137 Fab use D535 as an essential binding residue, they should cross-compete with each other unless they recognize distinct conformation of aa 532-540.

Intriguingly, simultaneous binding of both DAO5 scFv and e137 Fab to sE2ΔHVR1 could be detected in a pull-down assay on the Streptactin affinity column (**Figure 50A).** In this assay sE2ΔHVR1 was affinity bound to the column. DAO5 scFv and e137 Fab (both lacking the Strep-tags) were loaded on the column in separate steps (DAO5 scFv in the first step and e137 in the second step including a washing step in between). The molar excess of both DAO5 scFv and e137 Fab were used to make sure that they saturate all available binding sites in sE2ΔHVR1. Surprisingly, the eluted fractions contained all three proteins, meaning that both DAO5 scFv and e137 were able to bind to sE2ΔHVR1 despite the fact that they both should cross-compete for binding to D535. Thus, DAO5 scFv was not able to recognize all sE2ΔHVR1 molecules when it was passed through the column, indicating that its epitope was occluded or present in a different conformation in some sE2ΔHVR1 molecules at a given moment. However, the binding of Fab e137 demonstrated that the population of molecules that was not bound by scFv DAO5 was available for binding by Fab e137, suggesting that DAO5 and e137 mAbs should recognize distinct conformations of the same epitope in sE2ΔHVR1.

I also repeated the pull-down experiment by replacing the order in which scFv DAO5 and Fab e137 were loaded on the column, i.e. loading Fab e137 in the first step followed by scFv DAO5. Again, the eluted fraction contained all three proteins, indicating that Fab e137 also binds only to the fraction of the molecules that presents its epitope in a certain conformation at a given moment.

This observation might be related with the conformational flexibility of the region spanning aa 532-540: at a given time this epitope is present in a few or more distinct conformations, thus, there is a fraction of sE2ΔHVR1 molecules in the conformation recognized by DAO5 mAb and the fraction of the molecules in the conformation recognized by mAb e137.

In order to confirm our hypothesis that the epitope spanning aa 532-540 is flexible, we also performed a similar experiment to the above-described pull-down using surface plasmon resonance (SPR) (**Figure 50B**). sE2ΔHVR1 was immobilized on the chip via the Strep-tag and then Fab DAO5, e137 or a control Fab (after removing the Strep-tag) was injected over

the glycoprotein followed by the second Fab. In all cases, the second Fab was injected when a saturation of all accessible binding sites on sE2ΔHVR1 was reached with the first injected Fab. Similar to the pull-down assay, binding of both Fabs DAO5 and e137 was observed. Of note, the binding of both of these Fabs together corresponded to ~ 400 response units while ~800 response units of sE2ΔHVR1 were immobilized non-covalently via the Strep-tag. Therefore, a much higher response would be expected if 100% of immobilized sE2ΔHVR1 were recognized when both Fabs are injected. Non-covalent immobilization via the Strep-tag should neither affect the conformation of the glycoprotein nor sterically interfere with the accessibility of the Fabs to the epitope spanning aa 532-540. Therefore, the lower than expected response might indicate that due to flexibility of the epitope, a fraction of sE2ΔHVR1 molecules had the conformation of the epitope recognized neither by Fab DAO5 nor by Fab e137.

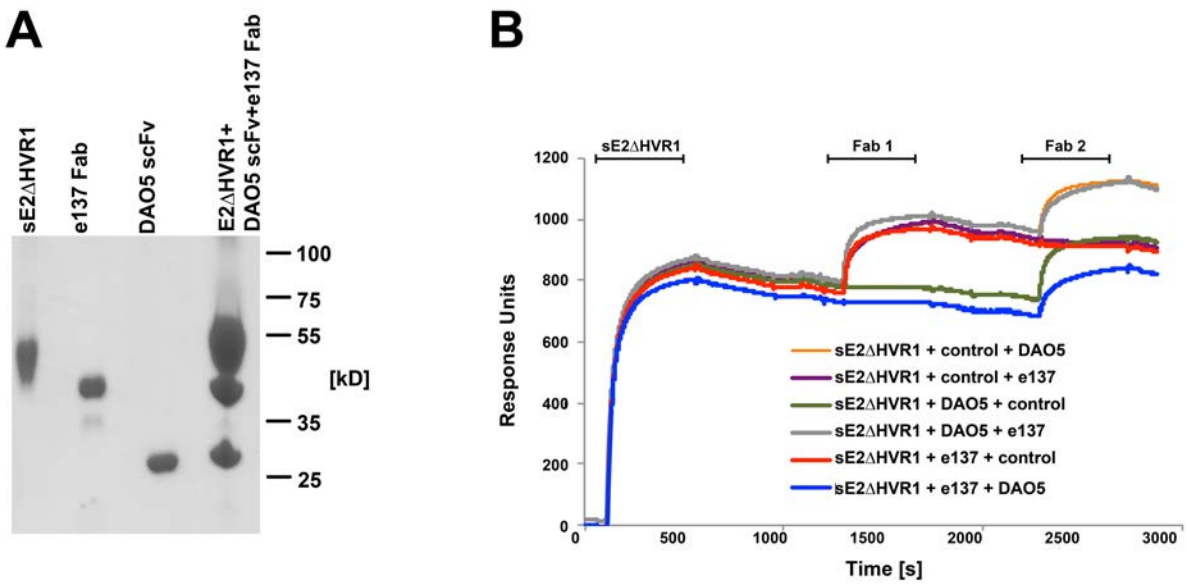


Figure 50. Two populations of sE2ΔHVR1 can be distinguished by Fabs DAO5 and e137. (A) A pull-down experiment similar to Figure 5A, immobilizing sE2ΔHVR1 on a Streptactin column followed by a molar excess of DAO5 scFv and subsequently by e137 Fab. Elution of the complex revealed binding of both antibody fragments to immobilized sE2ΔHVR1. Given that no ternary complex containing sE2ΔHVR1/DAO5/e137 was observed (see Figure 5C), this suggests the presence of two populations of E2 molecules - one recognized by DAO5 and the second one recognized by e137. (B) Real-time SPR analysis of Fab binding to immobilized sE2ΔHVR1 recording the binding response in resonance units (RU) as a function of time. Fabs lacking the Strep-tag (DAO5, e137 or a control Fab) were injected over HCV sE2ΔHVR1 immobilized using an anti-Strep-tag antibody at a flow rate of 5 µL/ml. After saturation, a second Fab lacking the Strep-tag (DAO5, e137 or a control Fab) was injected. The association/dissociation time-course profiles support the presence of two distinct populations recognized by either DAO5 or e137.

DAO5 antibody fragments interact with their cognate antigen in a temperature dependent manner

I also further analyzed the complex formation between DAO5 Fab and sE2ΔHVR1 by SEC. DAO5 Fab and sE2ΔHVR1were mixed at equimolar ratio and incubated for 18-24 h at 4°C followed by analysis on a Superdex 200 10/300 column (column volume 24 ml, GE Healthcare). The complex of sE2ΔHVR1/DAO5 Fab eluted considerably earlier than the respective individual proteins (**Figure 51A).** However, a significant amount of sE2ΔHVR1 and DAO5 Fab did not form a complex and eluted at volumes corresponding to the molecular mass of the individual proteins, which was in agreement with the results obtained in the pull-down assay with DAO5 scFv. To rule out the possibility that sE2ΔHVR1 and DAO5 Fab were not mixed at exactly equimolar ratio due to pipetting errors, the experiment was repeated a second time using exactly the same amounts of the proteins. The complex formation this time was carried out at room temperature over night. Interestingly, SEC analysis revealed that there was less free sE2ΔHVR1 and DAO5 Fab in the mixture after the complex formation at room temperature (**Figure 51B**).

Detection of free sE2ΔHVR1 and DAO5 Fab beside the complex in SEC could be related with relatively low affinity between the two proteins. Complex stability is known to depend on the affinity between two proteins. For lower affinity interactions, a complex can dissociate on the column since it is diluted during SEC. In such a case, a complex reaches equilibrium with individual proteins on the column. If the eluted complex is concentrated and analyzed by SEC the second time, the equilibrium is established again, i.e. the fraction of the complex dissociates to individual proteins. To assess the stability of sE2ΔHVR1/DAO5 Fab complex, I pooled the fractions containing the complex, concentrated it, and loaded it on the column. Only a negligible amount of free sE2ΔHVR1 and DAO5 Fab were eluted, indicating that the complex is stable once it is formed (**Figure 51C and D**).

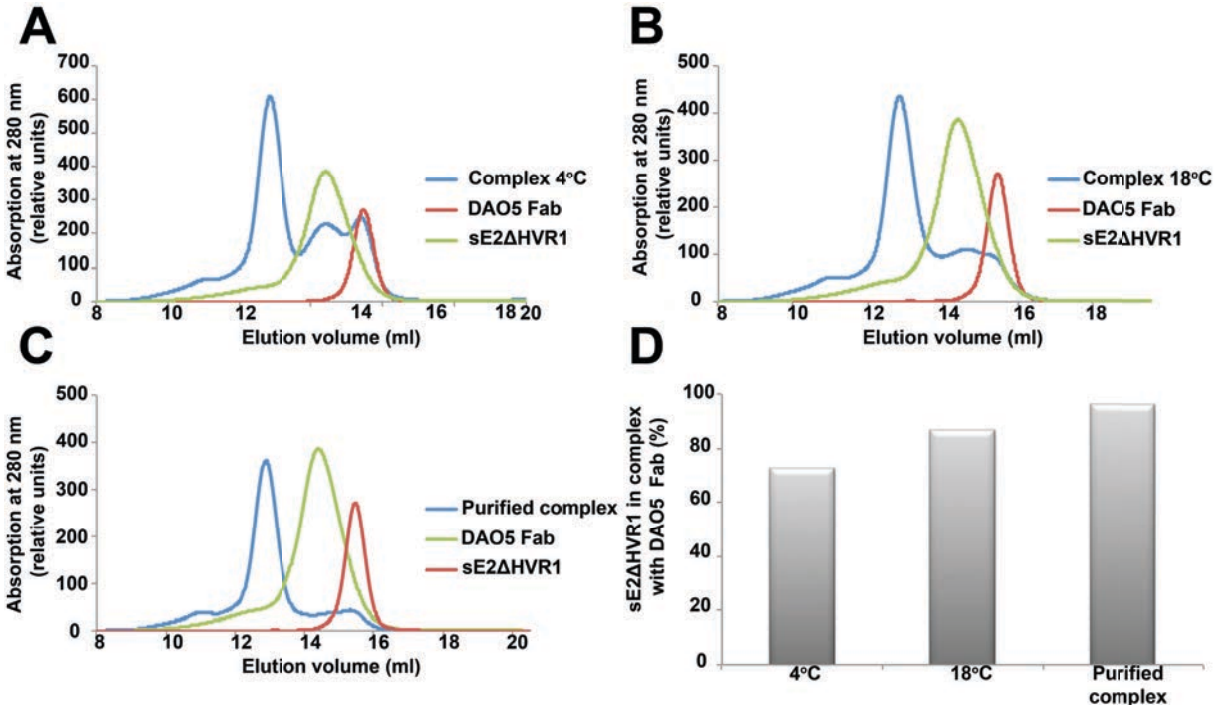


Figure 51. Stoichiometric complex formation between HCV sE2ΔHVR1 and DAO5 Fab. sE2ΔHVR1, DAO5 Fab and a mixture of the two (molar ratio 1:1) were loaded to the column (in three different runs) (E2 ΔHVR1:39 kD, DAO5 Fab:51 kD, complex: 90 kD). The complex was formed by incubating mixed proteins at 4°C over night (A) or room temperature over night (B). Peak fractions of the eluted complex from (B) were concentrated and loaded on the column again (C). The peak fraction of sE2ΔHVR1/DAO5 Fab was injected on SEC column. No significant peaks corresponding to either of the isolated proteins were observed in the profile of the purified complex, indicating that the complex of sE2ΔHVR1/DAO5 Fab is stable once it is formed. (D) The fraction of E2 ΔHVR1 in complex with the respective Fab was calculated by integrating areas under the curves in A, B and C using UNICORN control software (GE Healthcare).

The initial analysis of complex formation by SEC already revealed that the incubation temperature might have an effect on the DAO5 Fab binding to sE2ΔHVR1. Therefore, I performed a more detailed investigation of the temperature effect on the interaction. The formation of sE2ΔHVR1/DAO5Fab complex was carried out at four different temperatures (4°C, 18°C, 30°C and 37°C) overnight and subsequently analyzed by SEC on a Superdex 200 10/30 column (GE Healthcare). Clear temperature-dependence of the sE2ΔHVR1/DAO5 Fab complex formation was observed (**Figure 51A**). The calculation of the fraction of sE2ΔHVR1 in complex with DAO5 was calculated based on the area under the curves (**Figure 51B)**). Approximately 90% of sE2ΔHVR1were in complex with DAO5 Fab if the proteins were pre-incubated at 30°C or 37°C, while at 4°C the fraction of the sE2ΔHVR1in complex was less than 60%.

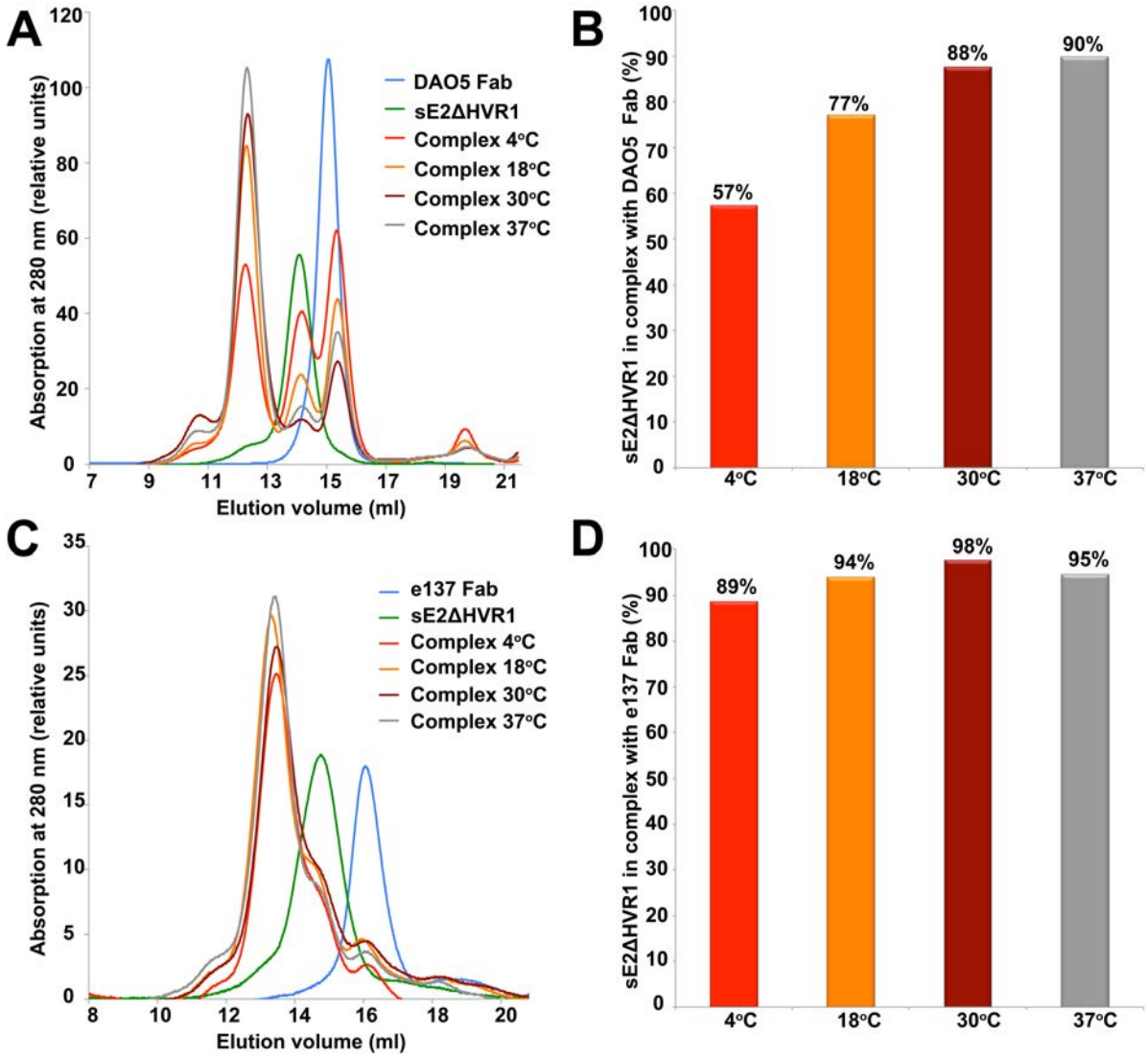


Figure 52. Complex formation between HCV E2 ΔHVR1 and Fab fragments derived from DAO5 and e137 mAbs, respectively, at different temperatures. (A) The mixture of E2 ΔHVR1 and the respective Fab fragment (molar ratio 1:1) was incubating at 4°C, 18 °C, 30°C and 37°C, respectively, overnight. E2 ΔHVR1, the Fab and a mixture of the two were loaded to the column (in different runs) (sE2ΔHVR1: 39 kD, Fab: 51 kD, complex: 90 kD). (B and D) The fraction of E2 ΔHVR1 in complex with the respective Fab was calculated based on the area under the curve using UNICORN control software (GE Healthcare). The columns are colored according to the curves in (A and C): E2/ Fab complex formed at 4°C (red), E2/DAO5 Fab complex formed at 18°C (orange), E2/DAO5 Fab complex formed at 30°C (dark red), E2/DAO5 Fab complex formed at 37°C (grey).

I also used a conformation-sensitive e137 Fab as a control to check if the temperature dependence observed for DAO5 Fab binding to sE2ΔHVR1 is not due to partial denaturation of the glycoprotein. In addition to residues within epitope I and epitope II, e137 mAb interacts with residues within the aa 523-540 segment. D535 has been also characterized as an essential binding residue for this mAb. As DAO5 Fab recognizes a linear epitope, the observed temperature dependence could be due to the fact that DAO5 epitope gets more exposed as the sE2ΔHVR1 becomes partially denatured during the overnight incubation at

higher temperatures. The complexes of sE2ΔHVR1 with e137 Fab were formed and analyzed by SEC in the same way as described earlier for sE2ΔHVR1 /DAO5 Fab complexes. No temperature-dependent binding was detected for e137 indicating that the overall conformation of the sE2ΔHVR1 is not affected by incubation of the protein at higher temperatures overnight (Figure 22).

In conclusion, the flexibility of the epitope is also supported by the observed temperature-dependent binding of DAO5 mAb. Since at elevated temperatures the mobility of the flexible epitope is likely to increase, DAO5 mAb is able to bind to all sE2ΔHVR1 molecules given enough time.

DAO5 binding to sE2ΔHVR1 might happen due to molecular breathing of HCV E2 protein

Based on the results of the complex formation observed in the SEC and pull-down assay, we developed a hypothesis that the epitope of DAO5 mAb might become exposed due to so-called molecular breathing of E2 protein.

Sabo and colleagues have demonstrated that the neutralization potency of some anti-HCV E2 antibodies depends on temperature and pre-incubation time (Sabo *et al.*, 2012). Some mAbs were not able to neutralize HCV when they were pre-incubated with the virus at 4^{o}C for 1 hour, but were neutralizing if this step was performed at 37 or 40°C. In addition, the neutralizing activity of these mAbs was improved if the pre-binding step was extended to 8h. The difference between those mAbs in neutralizing activity was not associated with the change of binding kinetics or antibody-virus aggregation at different temperatures. Furthermore, immune sera from acute or chronically infected patients also displayed enhanced neutralization when pre-incubated at elevated temperatures and longer periods. Immunoprecipitation studies with those mAbs revealed that greater amounts of HCV were immunoprecipitated when pre-incubation was carried out 37 or 40^{o}C in comparison with 4^{o}C. These results suggested that temperature is likely to affect epitope exposure of the studied mAbs.

Moreover, prior studies revealed that the virions of some flaviviruses such as West Nile virus and Dengue virus display a dynamic motion in their envelope proteins, which is referred to as virus "breathing" (Dowd *et al.*, 2011; Lok *et al.*, 2008). Virus "breathing" was discovered based on antibody binding studies and shown to significantly modulate epitope accessibility at the surface of West Nile and Dengue virus particles, allowing antibodies to

bind otherwise obscured epitopes. Increased temperatures can promote dynamic motion, which is illustrated by the fact that panels of mono- and polyclonal antibodies specific against these viruses have been shown to neutralize in a temperature dependent manner.

In addition, prolonged interaction times between antibody and virus also enhanced neutralization activity, as it is likely to result in engagement of less accessible epitopes through virus "breathing". It has been shown that neutralization of some flaviviruses requires antibody binding to exceed a threshold of bound antibody molecules per virion, likely related to the number of symmetry related epitopes that need to be occupied for neutralization (Diamond *et al.*, 2008; Pierson *et al.*, 2007). In agreement with this observation, longer time intervals of interaction dramatically improved the neutralization potency of the mAbs, which were defined as weakly or non-neutralizing by standard neutralization assays.

The dynamic motion of the flavivirus virions is supported by high resolution structures available for the envelope glycoprotein E in conjunction with electron microscopy maps of virions at high resolution. For example, neutralizing mAb 1A1D-2 - which binds to Dengue virions only at 37°C, but not at 4°C - binds to a β-strand in domain III of E protein, which is partially occluded in the context of the virion, because E proteins cluster tightly around the 3-fold symmetry axes on the mature virion (Lok *et al.*, 2008). Cryo-electron microscopy revealed that Fab 1A1D-2 binds to two out of the three E proteins arranged around the 3-fold symmetry axes, thereby trapping E in the conformation differing from the one present on the mature virion. This implies that an elevated temperature increases the mobility of the E proteins at the vertices of the virion, exposing otherwise hidden epitopes and providing a possibility for the antibody to capture it in this conformation. Once the antibody is bound to its epitope it is likely to stabilize that particular conformation by restricting the "breathing" motions because of the steric hindrance with the adjacent E proteins. The role of virus "breathing" *in vivo* on antibody neutralization has not been investigated yet, thus it is not clear if an elevated body temperature or the virus entry rate into target cells *in vivo* might affect the efficiency of antibody neutralization.

HCV is closely related to flaviviruses, thus a phenomenon of virion "breathing" might also be characteristic to this virus. Unfortunately, the organization of HCV envelope proteins on the virion surface is unknown making it difficult to assess if temperature can facilitate dynamic motions of the virion and, as a result, alter epitope accessibility. However, the study of Sabo *et al.* strongly supports the possibility that HCV virions may also undergo "breathing" motions.

Verification of sE2ΔHVR1 folding

Since DAO5 recognizes a linear epitope within HCV E2, it is essential to prove that recombinant sE2ΔHVR1 is correctly folded and, thus, represents the conformation of E2 present on HCV virions. Usually the binding of conformation-sensitive antibodies to a recombinant protein is used to confirm that the recombinant protein adopts a relevant conformation. Therefore, I tested the complex formation between sE2ΔHVR1 and a number of recombinant Fabs derived from neutralizing antibodies, including conformation-sensitive ones, by SEC analysis (**Figure 53**). Fabs e137 (Perotti *et al.*, 2008), A8 (Johansson *et al.*, 2007), CBH-7 (Hadlock *et al.*, 2000), CBH-23 (Hadlock *et al.*, 2000), HC-1 (Broering *et al.*, 2009) and HC84-1 (Keck *et al.*, 2012) recognize at least a few residues within aa 523-540 of HCV E2, while mAbs AP33 (Tarr *et al.*, 2006) and HC84-1 (Keck *et al.*, 2012) recognize epitope I and epitope II, respectively. The SEC analysis demonstrated that all Fabs derived from conformation-sensitive mAbs efficiently bind to the recombinant sE2ΔHVR1.

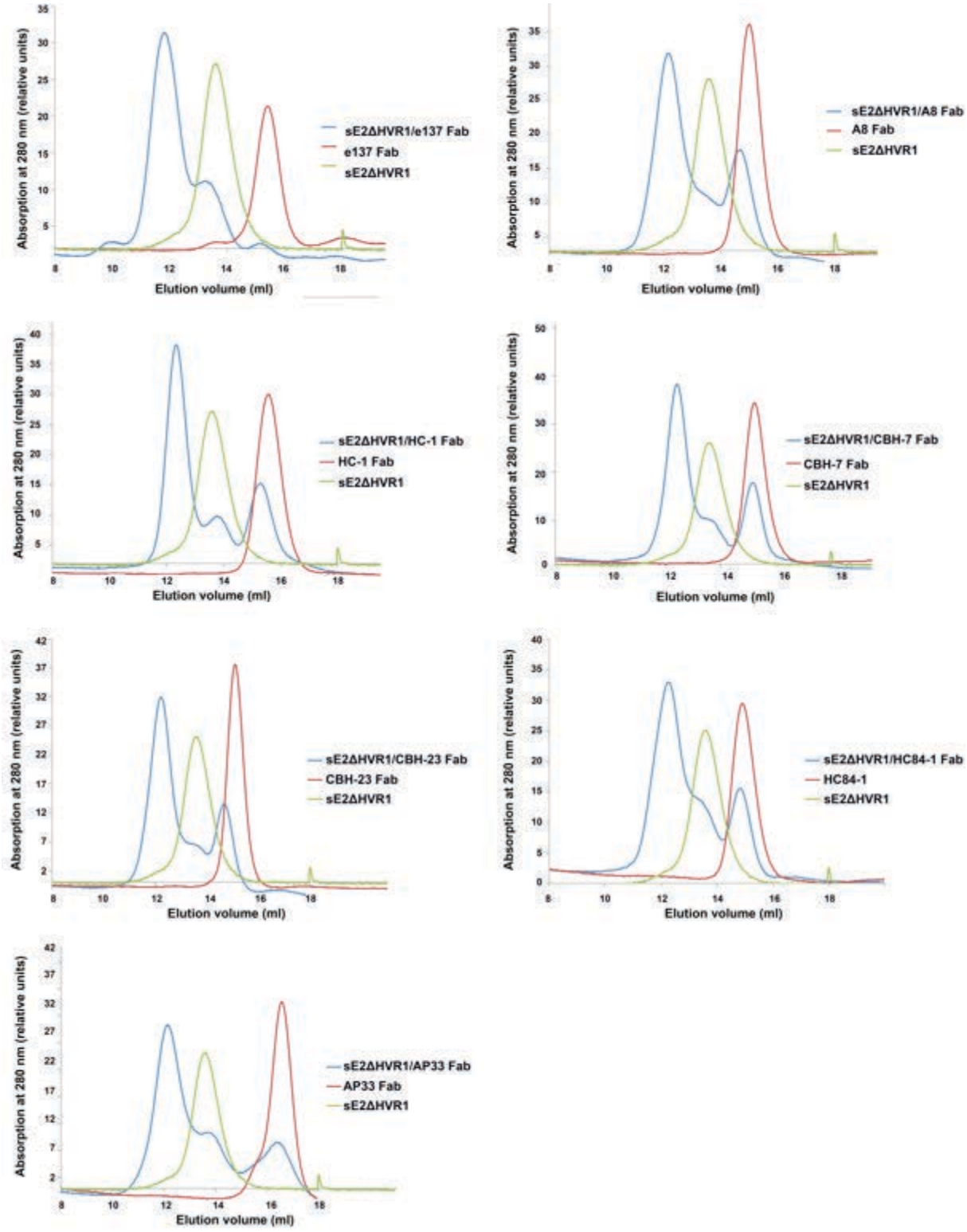


Figure 53. Stoichiometric complex formation between HCV sE2ΔHVR1 and the Fabs derived from conformation-sensitive mAbs: (A) e137 (B) A8 (C) CBH-7 (D) CBH-23 (E) HC-1 (F) AP33 (G) HC84-1. E2 ΔHVR1, one of the Fabs and a mixture of the two (molar ratio 1:1) were loaded to the column (in three different runs) (sE2ΔHVR1: 39 kD (green curve), Fab: 51 kD (red curve), complex: 90 kD (blue curve). The complexes were formed by incubating mixed proteins at room temperature overnight.

Structurally intact conformation of sE2ΔHVR1 was also evaluated in a pull-down

assay using a Fab derived from non-neutralizing conformation-dependent mAb CBH-4D. mAb CBH-4D is binding to a different conformational antigenic domain on E2 (Hadlock *et al.*, 2000; Keck *et al.*, 2004) than DAO5 mAb and is not expected to cross-compete for binding to sE2ΔHVR1. Cross-competition between those two mAbs was tested in a pull down-assay using DAO5 scFv and CBH-4D Fab (**Figure 54**). The purified complex of sE2ΔHVR1 (containing the Strep-tag) and DAO5 scFv (without the Strep-tag) was affinity bound on Streptactin column and then CBH-4D Fab (without the Strep-tag) passed through the column. The eluted fraction contained all three proteins (sE2ΔHVR1, DAO5 scFv and CBH-4D Fab) indicating that DAO5 induces just a local conformational change in sE2ΔHVR1 and do not interfere with the binding of CBH-4D Fab (**Figure 54**).

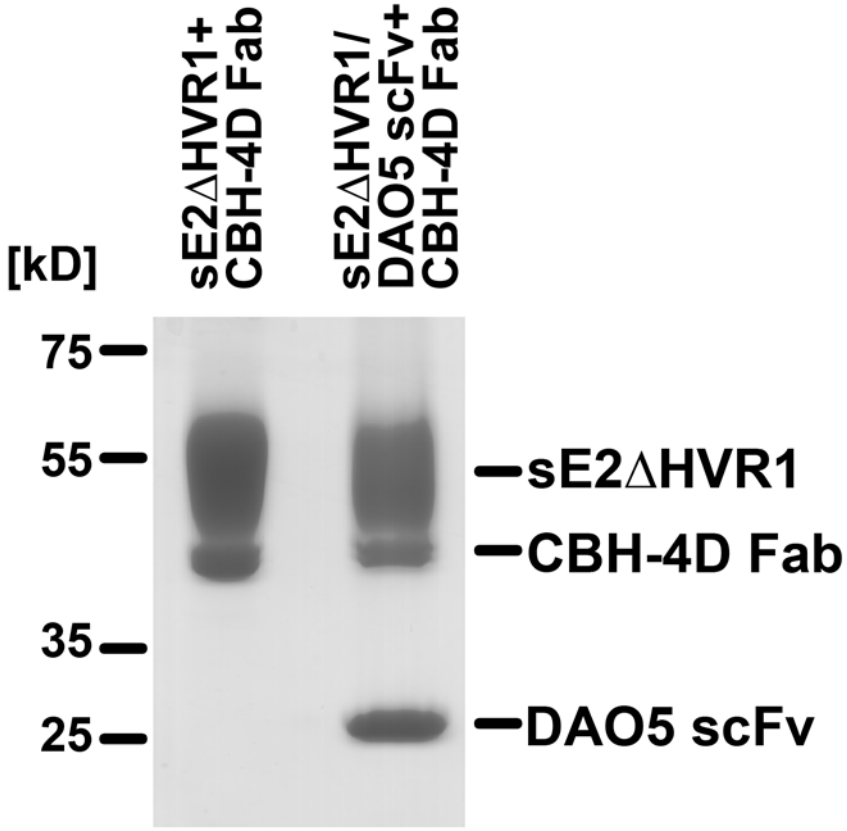


Figure 54. Pull-down experiment showing that sE2ΔHVR1 specifically reacts with DAO5 scFv and CBH-4D Fab. sE2ΔHVR1 or sE2ΔHVR1/DAO5 scFv complex were affinity loaded onto two separate Streptactin columns followed by DAO5 scFv lacking the Strep-tag and the proteins were eluted from both columns. Elution fractions were analyzed by SDS-PAGE under non-reducing conditions and stained by Coomassie Blue. Bands corresponding to all three proteins (sE2ΔHVR1, CBH4D and DAO5 scFv) were observed in the elution fraction from the column onto which sE2ΔHVR1/DAO5 scFv complex was loaded indicating that both DAO5 scFv and CBH4D Fab bind to sE2ΔHVR1.

DAO5 mAb cross-competition with other anti-HCV mAbs

As the crystal structures of DAO5 mAb peptide epitope revealed, the antibody interacts with D535. Though this residue is not completely buried in the interface with the DAO5 antibody fragments, it is not likely to be accessible for binding to another mAb that uses D535 as contact residue while it is interacting with the DAO5 Fab. To confirm that DAO5 mAb indeed cross-compete with other mAbs (e137, A8, CBH-7, CBH-23 and HC-1), which have been described to require D535 for binding to E2, I performed cross-competition analysis using SEC. Since the formation of sE2ΔHVR1/DAO5 Fab was almost equally efficient at 30°C and 37°C, I incubated sE2ΔHVR1 plus DAO5 Fab at 30°C overnight to obtain sE2ΔHVR1/DAO5 Fab complex. The excess of unbound DAO5 Fab was removed by SEC. The purified sE2ΔHVR1/DAO5 Fab complex was used to set up the complexes with Fabs e137, A8, CBH-7, CBH-23 and HC-1. In addition, I also set up complexes with AP33 and HC84-1 Fabs that bind to epitope I and epitope II, respectively, and cross-compete with CD81.

The complexes were analyzed by SEC after overnight incubation of the mixed proteins at room temperature (**Figure 55**). For the complexes between sE2ΔHVR1/DAO5 Fab complex e137, A8, CBH-7, CBH-23, HC-1 and HC84-1 two peaks, one corresponding to the molecular weight of sE2ΔHVR1/Fab complex and the second corresponding to the molecular weight of a Fab molecule, were detected in SEC. The ternary complex was observed only with Fab AP33, indicating that Fab DAO5 does not sterically block the binding of Fab AP33 to epitope I. However, DAO5 Fab interfered with binding of the epitope II Fab HC84-1.

The described competition analysis by SEC showed that once a stable sE2ΔHVR1/DAO5 Fab complex is formed, other Fabs (including Fab e137) binding to the same epitope in sE2ΔHVR1 are excluded from binding to the glycoprotein. This means that Fab DAO5 locks the epitope in a certain conformation that makes its replacement by other Fabs difficult.

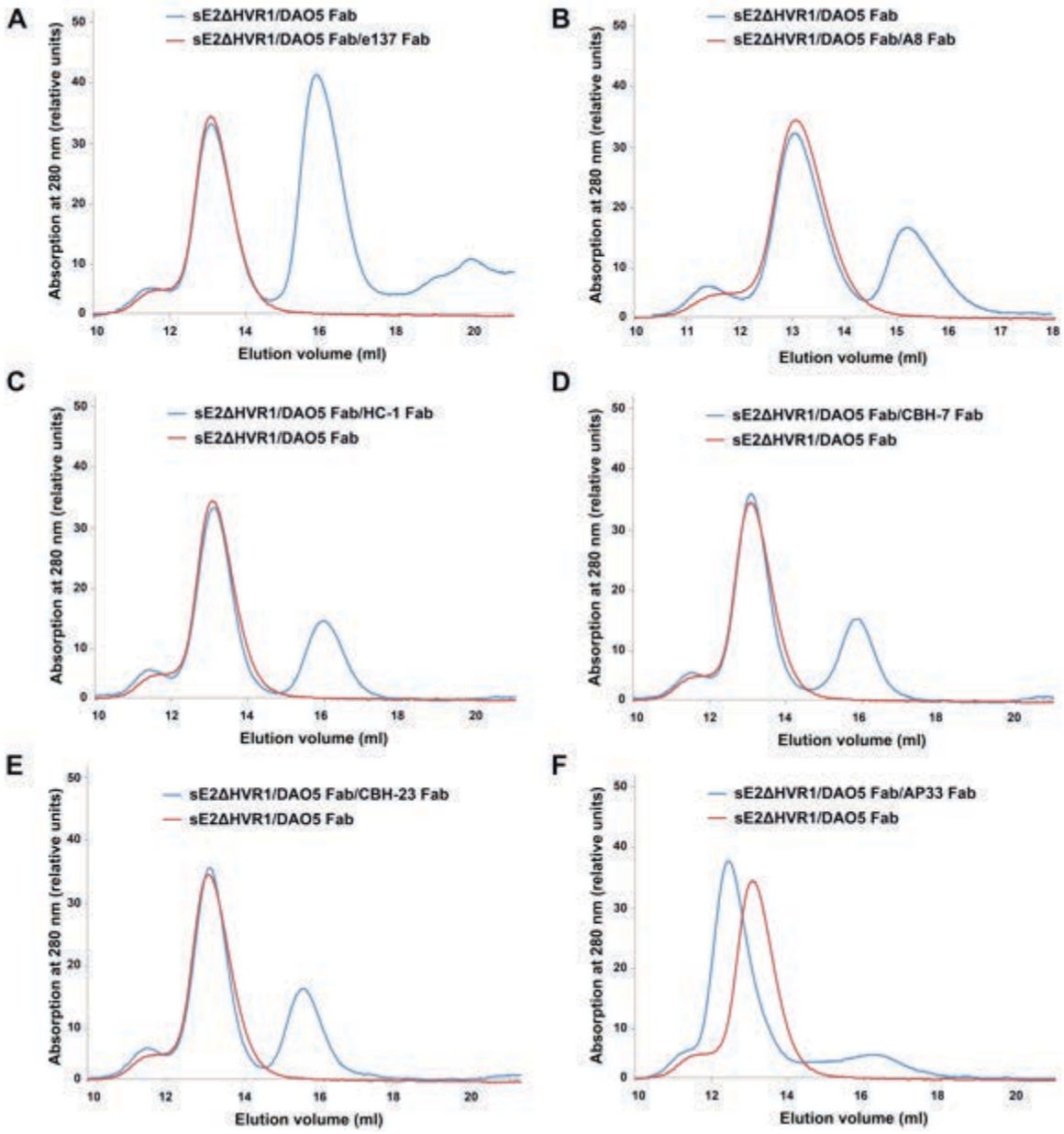


Figure 55. Cross-competition analysis of DAO5 Fab with the Fabs derived from conformation-sensitive mAbs: (A) e137 (B) HC-1 (C) CBH-23 (D) CBH-7 (E) A8 (F) AP33. Purified complex of E2 ΔHVR1/DAO5 Fab and a mixture of the latter with one of the above mentioned Fabs (molar ratio 1:1) were loaded to the column (in two different runs) (E2 ΔHVR1/DAO5 Fab:90 kD, Fab:51 kD, E2 ΔHVR1/DAO5 Fab/AP33 Fab 141 kD). The complexes were formed by incubating mixed proteins at room temperature overnight.

DAO5 mAb cross-competition with the receptor CD81

Our partners in Glasgow have characterized DAO5 mAb as non-neutralizing antibody using HCVcc and HCVpp. Inability of DAO5 mAb to block the infection suggests that it does not prevent E2 interaction with the cellular receptor CD81. However, it is surprising given the fact that D535 is required for binding of both CD81 and DAO5 mAb. I further verified if DAO5 Fab cross-competes with CD81 by using SEC. The complexes between the purified

sE2ΔHVR1/DAO5 Fab complex or only sE2ΔHVR1 and the recombinant large extracellular loop (McLellan *et al.*) of CD81 was formed overnight and analyzed by SEC on a Superdex 200 10/300 column the following day. Since the molecular weight of CD81 LEL is only ~10 kD, the shift of sE2ΔHVR1/DAO5 Fab complex would not be visible in SEC even if CD81 binds. Therefore, the peak fractions of the complexes were concentrated and analyzed by SDS-PAGE. No binding of CD81 LEL to sE2ΔHVR1/DAO5 Fab complex was detected as indicated by SDS-PAGE of eluted fraction of sE2ΔHVR1/DAO5 Fab complex plus CD81 LEL (**Figure 56).** In contrast, CD81 LEL was binding to sE2ΔHVR1 alone in the same assay.

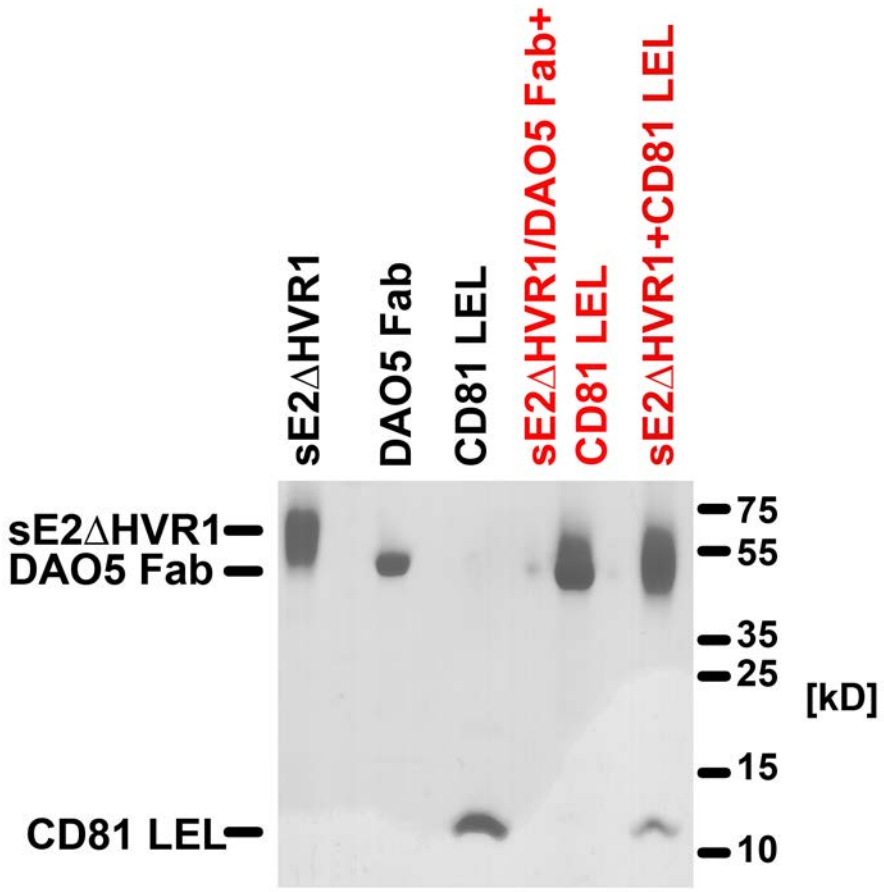


Figure 56. DAO5 Fab cross-competition with CD81 LEL for binding to sE2ΔHVR1. sE2ΔHVR1 and sE2ΔHVR1/DAO5 Fab complex was pre-incubated with CD81 LEL overnight at room temperature. The complexes were separated by size-exclusion chromatography and the peak fractions were concentrated and analyzed by SDS-PAGE under non-reducing conditions followed by Coomassie staining. sE2ΔHVR1/DAO5 Fab does not bind the CD81 LEL, suggesting a direct interference between DAO5 and CD81 binding in spite of the lack of neutralizing activity observed for mAb DAO5.

These results are in agreement with D535 residue being an essential binding residue for both DAO5 mAb and CD81 LEL. Thus, DAO5 mAb is interfering with CD81 binding though it is not able to neutralize viral particles. Notably, our collaborators at Glasgow

performed the neutralization experiments by incubating DAO5 mAb with HCVpp and HCVcc at 37^{o}C for 1h. This incubation time could be too short to achieve neutralization given the fact that the temperature dependent binding between DAO5 antibody fragments and sE2ΔHVR1 was observed after overnight pre-incubation of the proteins. Unfortunately, the investigation of neutralization using longer incubation times of DAO5 mAb with HCVpp and HCVcc at 37^{o}C was not possible due to instability of the viral particles.

All the above-described results together suggest that HCV E2 region aa 530-540, which is crucial for the glycoprotein-CD81 interaction and, thus, infection of cells, is very flexible. The possible benefits of this flexibility for virus infection will be discussed in the next section.

Supplementary data

Table S1. Fab – peptide interactions.

	Fab residue	Peptide residue	Dist. [Å]
HEAVY CHAIN			
DAO5 Fab/peptide_J4			
Hydrogen bonds			
main chain - side chain			
	Lys 52 NZ	Val 536 O	2.94
	Lys 52 NZ	Met 537 O	3.25
	Lys 52 NZ	Asn 540 O	2.87
Hydrophobic Interactions			< 5
	Ala 59	Met 537	
	Phe 99	Leu 538	
	Phe 99	Leu 539	
	Tyr 103	Leu 539	
	Pro 104	Leu 539	
	Tyr 105	Leu 539	
DAO5 scFv/peptide_J4			
Hydrogen bonds			
main chain - side chain			
	Lys 52 NZ	Val 536 O	2.88
	Lys 52 NZ	Met 537 O	3.29
	Lys 52 NZ	Asn 540 O	2.81
Hydrophobic Interactions			
	Ala 59	Met 537	
	Phe 99	Leu 538	
	Phe 99	Leu 539	
	Tyr 103	Leu 539	
	Pro104	Leu 539	
	Tyr 105	Leu 539	
DAO5 scFv/peptide_JFH			
Hydrogen bonds			
main chain - side chain			< 5
	Lys 52 NZ	Val 536 O	3.05
	Lys 52 NZ	Phe 537 O	2.87
	Lys 52 NZ	Asn 540 O	2.74
Hydrophobic Interactions			
	Ala 59	Phe 537	
	Phe 99	Leu 538	
	Tyr 103	Leu 539	
	Pro 104	Leu 539	
	Tyr 105	Leu 539	
	Phe 99	Leu 539	
Aromatic-aromatic interactions			4.5-7
	Trp 94	Phe 537	5.42
			< 6
Cation-Pi interactions			
	Lys 52	Phe 537	5.68

Table S1. Fab – peptide interactions continued.

LIGHT CHAIN

	Fab residue	Peptide residue	Dist. [Å]
DAO5 Fab – peptide 1			
Hydrogen bonds			
main chain - side chain			
	Thr 91 O	Thr 534 OG1	2.63
	Trp 94 N	Glu 533 OE2	2.89
side chain - side chain			
	Tyr 32 OH	Asp 535 OD1	2.71
	Tyr 32 OH	Asp 535 OD2	3.28
	Lys 93 NZ	Glu 533 OE1	3.5
Hydrophobic Interactions			< 5
	Trp 94	Met 537	
	Trp 94	Leu 538	
	Leu 96	Leu 538	
	Tyr 32	Leu 539	
	Tyr 50	Leu 539	
Ionic interactions			< 5
	Lys 93	Glu 533	
DAO5 scFv /peptide_J4			
Hydrogen bonds			
main chain - side chain			
	Asn 92 OD1	Glu 531 O	3.05
	Lys 93 NZ	Gly 530 O	3.32
	Trp 94 N	Glu 533 OE2	2.84
	Trp 94 NE1	Glu 533 O	3.31
	Thr 91 O	Thr 534 OG1	2.64
	Asn 92 ND2	Glu 531 O	3.48
side chain - side chain			
	Tyr 30 OH	Glu 531 OE1	2.89
	Tyr 32 OH	Asn 532 OD1	3.46
	Tyr 32 OH	Asp 535 OD1	2.82
	Tyr 32 OH	Asp 535 OD2	3.05
	Asn 92 OD1	Asn 532 OD1	3.48
Hydrophobic Interactions			< 5
	Tyr 32	Leu 539	
	Tyr 50	Leu 539	
	Trp 94	Phe 537	
	Trp 94	Leu 538	
	Leu 96	Leu 538	
Ionic Interactions			< 6
	Lys 93	Glu 531	
	Lys 93	Glu 533	
DAO5 scFv – peptide_JFH			
Hydrogen bonds			

main chain - side chain			
	Asn 92 OD1	Glu 531 O	3.46
	Asn 92 ND2	Glu 531 O	2.94
	Lys 93 NZ	Gly 530 O	3.13
	Trp 94 N	Glu 531 OE2	2.86
	Trp 94 NE1	Glu 531 O	3.35
	Thr 91 O	Thr 534 OG1	2.65
side chain - side chain			
	Tyr 32 OH	Asn 532 OD1	3.36
	Tyr 32 OH	Asn 532 ND2	3.26
	Tyr 32 OH	Glu 531OE1	2.99
	Tyr 32 OH	Asp 535 OD1	2.72
	Tyr 32 OH	Asp 535 OD2	3.03
Hydrophobic Interactions			< 5
	Tyr 32	Leu 539	
	Tyr 50	Leu 539	
	Trp 94	Phe 537	
	Trp 94	Leu 538	
	Leu 96	Leu 538	
Ionic Interactions			< 6
	Lys 93	Glu 533	
	Lys 93	Glu 531	

Table S2. Intrapeptide interactions.

	Contact residue 1	Contact residue 2	Dist. [Å]
Peptide_J4 (complexed with DAO5 Fab)			
Hydrogen bonds			
main chain - main chain			
	Thr 534 N	Asn 532 O	3.12
	Asp 535 N	Asn 532 O	3.27
	Val 536 N	Thr 534 O	3.39
	Met 537 N	Thr 534 O	3.30
	Met 537 N	Asp 535 O	3.31
	Met 538 N	Thr 534 O	2.85
	Leu 538 N	Asp535 O	3.41
	Leu 539 N	Asp 535 O	2.78
	Asn 540 N	Asp 535 O	3.14
Hydrophobic Interactions			< 5
	Met 537	Leu 538	
	Leu 538	Leu 539	
Peptide_J4 (complexed with DAO5 scFv)			
Hydrogen bonds			
main chain - main chain			
	Thr 534 N	Asn 532 O	3.26
	Asp 535 N	Asn 532 O	3.40
	Val 536 N	Thr 534 O	3.31
	Met 537 N	Thr 534 O	3.49
	Met 537 N	Asp 535 O	3.48
	Leu 538 N	Thr 534 O	2.88
	Leu 539 N	Asp 535 O	2.84
	Asn 540 N	Asp 535 O	3.17
main chain - side chain			
	Asn 540 OD1	Asp 535 O 1	3.11
	Asn 540 OD1	Asp 535 O 2	3.11
side chain – side chain			
	Asn 532 ND2	Asp 535 OD2 1	3.35
Hydrophobic Interactions			< 5
	Met 537	Leu 538	
	Leu 538	Leu 539	
Peptide_JFH (complexed with DAO5 scFv)			
Hydrogen bonds			
main chain - main chain			
	Thr 534 N	Asn 532 O	3.25
	Asp 535 N	Asn 532 O	3.36
	Val 536 N	Thr 534 O	3.38
	Met 537 N	Thr 534 O	3.29
	Met 537 N	Asp 535 O	3.48
	Leu 538 N	Thr 534 O	2.96
	Leu 539 N	Asp 535 O	2.85
	Asn 540 N	Asp 535 O	3.28
main chain - side chain			
	Asn 540 OD1	Asp 535 O	3.36

Hydrophobic Interactions			< 5
	Phe 537	Leu 538	
	Leu 538	Leu 539	

Discussion

Intrinsic structural flexibility of the main antigenic region in HCV E2 glycoprotein

HCV E2 interaction with the host entry factor CD81 is an essential step in the HCV entry process. The majority of HCV neutralizing antibodies interfere with CD81-mediated virus entry, suggesting that a conserved CD81 binding site could represent a potential candidate for immunogen design. The CD81 binding site on the E2 surface is composed of discontinuous segments and the epitopes of most broadly neutralizing antibodies overlap with these segments or reside in close vicinity. First structural insights into those antigenic regions came from crystal structures of E2-specific antibodies in complex with synthetic peptides mimicking their respective epitopes (Kong *et al.*, 2012a; Kong *et al.*, 2012b) (Deng *et al.*, 2014; Deng *et al.*, 2013; Krey *et al.*, 2013; Pantua *et al.*, 2013; Potter *et al.*, 2012).

However, the most comprehensive analysis of structural basis of HCV neutralization was gained from a recent study of the structure of a core fragment of E2 glycoprotein in complex with a Fab fragment of the neutralizing antibody AR3C (Kong *et al.*, 2013). The structure of this complex has significantly contributed to our understanding of antibody-mediated neutralization of HCV by providing evidence that AR3C Fab binds to the same E2 surface interface as CD81. The crystal structure also revealed that more than half of the E2 core residues, especially in the solvent exposed areas, are disordered or present in loops suggesting the overall high degree of flexibility of the glycoprotein. Moreover, it implies that some regions of E2 involved in CD81 binding might depend on association with AR3C Fab for their stabilization. The segments comprising CD81 binding site as well as the major antigenic regions targeted by neutralizing antibodies reside in the flexible area of the E2 protein.

In the presented thesis I describe the crystal structure of a peptide encompassing aa 529-540 in complex with the non-neutralizing mAb DAO5. In our structure the epitope peptide forms an α-helix at the C-terminal end of the peptide (D535-L539) while the N-terminal region (N532-T534) has an extended conformation. The observed conformation of the peptide epitope was unexpected given that this region adopts a ~20Å long, extended conformation encompassing a short β-strand (aa 536-538) in complex with neutralizing antibody ARC3. The epitope peptide in complex with DAO5 Fab and DAO5 scFv crystallized in different space groups but adopted an identical conformation as indicated by RMSD values

between two structures suggesting that the observed conformation is not due to crystal packing. Moreover, DAO5 scFv was crystallized with the peptide epitope from two different HCV genotypes (genotype 2a (strain JFH-1) and genotype 1b (strain J4)), which also revealed almost identical conformations.

The structure of the peptide bound to DAO5 mAb fragments suggests that residues F/M537 and L539 would be accessible on the E2 surface as these residues insert into the hydrophobic groove formed by the CDRs of DAO5 mAb. However, the crystal structure of the E2 core fragment reveals that these residues point towards the hydrophobic core of E2. The stark contrast between the α-helical conformation of the peptide epitope in complex with DAO5 and the extended β-strand conformation observed in the structure of the E2 core fragment indicate the dramatic structural flexibility of this crucial CD81 binding site.

By SPR and pull-down experiments we demonstrated that a given solution of recombinant sE2ΔHVR1 can bind both DAO5 Fab and a Fab fragment derived from the human neutralizing conformation-sensitive mAb e137 simultaneously. e137 mAb is a broadly neutralizing conformation-sensitive human mAb requiring D535 for binding to E2. Since e137 and AR3C mAbs use a number of the same contact residues within E2 as previously indicated by alanine scanning mutagenesis, we believe that both of them bind to E2 when aa 536-538 adopts the β-strand conformation. In contrast, DAO5 mAb selectively binds to sE2ΔHVR1 when its epitope within the protein adopts the α-helical conformation.

Our data demonstrating temperature-dependent binding of DAO5 antibody fragments to the recombinant sE2ΔHVR1 further supports the structural flexibility of this epitope. Certainly at higher temperatures the motion, i.e. flexibility, of the epitope increases resulting in a higher number of events when the epitope adopts the α-helical conformation recognized by DAO5 mAb. The fact that we did not observe temperature-dependent binding of e137 Fab suggests that this region in the sE2ΔHVR1 fluctuates mostly around the β-strand conformation while the energetic barrier to reach the α-helical conformation is higher and requires additional activation energy. We also demonstrated that the binding of the antibody locks the epitope in that particular conformation and prevents the binding of the second antibody recognizing a different conformation of the epitope. None of the tested human conformation-sensitive mAbs using D535 as an essential contact residue was able to bind to the purified sE2ΔHVR1/DAO5 Fab complex. Thus, the two observed conformations of the epitope are likely to be in dynamic equilibrium, which can be shifted in either direction by antibody binding. The energetic barrier for the epitope to adopt the α-helical conformation

should be higher, which is in agreement with the fact that the equilibrium towards this conformation can be shifted at physiological temperature in the presence of DAO5 mAb.

The β-strand conformation is also likely to be a predominant conformation because it is stabilized by the overall E2 core fold. The crystal structure of E2 core fragment shows that the short β-strand spanning aa 536-538 is part of the central β-sandwich, which resembles an IgC2-like domain and consists of four strands forming an inner sheet and two strands forming an outer sheet. The outer sheet is composed of two anti-parallel strands comprising aa 536-538 and aa 496-498, respectively. The interaction between these two strands includes hydrophobic contacts between I496 and V536, I496 and V538, V497 and F537, and P498 and V536 as well as the main chain-main chain hydrogen bond between V497 and F537. Most likely, the conformation of the β-strand spanning aa 496-498 would not be stabilized and retained if aa 535-539 adopted the α-helical conformation. Alternatively, it would require a change in the interactions between these two stretches of amino acids. Since the β-strands composing the outer sheet are short and do not form a number of stabilizing hydrogen bonds characteristic of longer β-sheets, the interactions between the strands are not strong enough to fix them in this conformation. The binding of AR3C Fab seems to have a stabilizing effect on this region: the Fab captures it in a certain (presumably more-stable) conformation and prevents it from structural fluctuation between different conformations. In conclusion, our results imply that the outer sheet of the IgC2-like domain composing the E2 core is displays a higher degree of flexibility. This is also supported by the fact that this region is disordered in E2 core crystal structure determined by Khan and colleagues (Khan *et al.*, 2014).

One possible explanation for the presence of two different conformations within the recombinant sE2ΔHVR1 is that DAO5 mAb binds to partially denatured or misfolded protein. It is known that HCV glycoproteins contain a number of disulphide bridges, thus, the folding of the proteins is difficult and often results in the aggregation of a fraction of the protein when expressed *in vitro* (reviewed in (Op De Beeck *et al.*, 2001)). Though sE2ΔHVR1 has been demonstrated to be recognized by human conformation-sensitive antibodies and induce neutralizing antibodies after immunization (Tarr *et al.*, 2013), we cannot completely rule out the possibility that a small fraction of the immunogen used to raise DAO5 mAb was misfolded or denatured. Giving the fact that DAO5 mAb recognizes a linear epitope, verification of whether or not it binds to the natively folded glycoprotein is essential. Neutralization of infection is the most straightforward way to demonstrate the binding to the envelope glycoprotein in its native conformation, as it is present on virions. However, DAO5 mAb is non-neutralizing, eliminating neutralization assay as a method to prove that DAO5

mAb recognizes the natively folded protein at the surface of virus particles. Nevertheless, we showed that the antigen recognized by DAO5 mAb is likely to be natively folded glycoprotein. First, we measured binding of the conformation-sensitive, non-neutralizing human antibody CBH-4D to sE2ΔHVR1/DAO5 scFv complex by pull-down assay. The obtained results indicated that sE2ΔHVR1 in complex with DAO5 scFv bound CBH-4D Fab as efficiently as the glycoprotein alone, which is a strong indication that sE2ΔHVR1 recognized by DAO5 scFv is not fully denatured. In addition, these results imply that DAO5 mAb binding does not affect the overall fold of the glycoprotein, which could be a possibility given the fact that aa 536-538 might not be the part of an IgC2-like domain anymore in the α-helical conformation.

Our collaborators in Glasgow also demonstrated that HCVcc particles could be immunoprecipitated with DAO5 mAb and were able to detect E2 in immunoprecipitated material. However, DAO5 may bind to immature forms of E2 that are secreted or found in the exosomes instead of the E2 displayed on viral particles. To confirm if the immunoprecipitated material actually contains viral particles our collaborators in Glasgow are planning to perform additional experiments: 1) test the infectivity of immunoprecipitated material, 2) try to quantitate viral RNA in immunoprecipitated material, 3) try to detect the core protein in immunoprecipitated material using anti-core antibodies.

It is possible but still unproven that the flexibility of the epitope observed in recombinant sE2ΔHVR1 also applies for E2 displayed at the surface of viral particles. Since DAO5 mAb is non-neutralizing, we still lack evidence that E2 displayed on the virions can adopt the α-helical conformation. Little is known about the virion structure of HCV and how the glycoprotein complexes are arranged in the viral envelope. It is possible that the presence of E1 and/or packing of the glycoprotein complexes on the virion surface might have a stabilizing effect and reduce the flexibility of the epitope. Therefore, there are two possibilities: 1) DAO5 mAb does not bind to the viral particles at all because the tertiary/quaternary structural constraints limit the flexibility of the epitope on viral particles and it does not adopt the α-helical conformation, 2) E2 displayed on the virions can adopt the α-helical conformation, however, due to the high energetic barrier to reach this conformation these events are relatively rare.

We could demonstrate that DAO5 mAb cross-competes with CD81 LEL for binding to sE2ΔHVR1 if all available sE2ΔHVR1 molecules are in complex with DAO5 Fab. This finding is contradictory to the fact that DAO5 is non-neutralizing, since inhibiting CD81 interaction with E2 normally prevents infection. There are several possible explanations why

DAO5 mAb is not able to neutralize virus infectivity. If the epitope does not adopt the α-helical conformation at the surface of viral particles, this explains why DAO5 does not neutralize HCVcc or HCVpp. Another possible explanation is that DAO5 mAb binds only to some glycoprotein complexes displayed on the virions (because this event is rare) while the majority of the glycoproteins are not bound by the antibody and, thus, can interact with CD81 and subsequently initiate virus entry. It has been shown that for some viruses neutralization occurs only when virions are bound by a number of antibodies that exceeds a required threshold. For example for flaviviruses ~30 antibody molecules per virion are required for efficient neutralization (Pierson *et al.,* 2007). Stoichiometric requirements for neutralization of HCV particles are unknown. Inhibition of infection, however, might require covering most of the available CD81 binding sites on the virion. Since neutralization potency of the antibody depends on a number of factors including the number of accessible epitopes on the surface of the virion, it is likely that DAO5 mAb is not neutralizing because it does not reach the required stoichiometry for neutralization.

Our preliminary results showed that HCVcc could be immunoprecipitated with DAO5 mAb, which would indicate that the antibody binds to the virions. To prove the binding of the mAb to the virions, we are planning to perform additional experiments. In addition to the identification of the origin of E2 in immunoprecipitated material as described above, we will also verify DAO5 mAb binding to the purified HCVpp particles by SPR. Furthermore, it might be possible to address this question by doing an additive neutralization by combining e137 and DAO5 mAbs. If we reach a higher level of neutralization by combining both mAbs, it would be a strong indication that DAO5 mAb contributes to the virus neutralization. Alone, however, it is not able to reach the stoichiometric requirements for neutralization. Also, we would like to verify if the observed flexibility of the recombinant glycoprotein ectodomain is related with the protein expression system or the chosen construct. Therefore, we are going to produce sE2ΔHVR1, the full-length sE2 and E2 core fragment described by Kong *et al*. in HEK293F cells and test if the epitope flexibility is similar for the produced proteins.

Intriguingly, the flexibility of epitope I and epitope II, two other regions involved in CD81 binding, has been recently demonstrated. Epitope I (aa 412-423) has been shown to adopt at least two different conformations. A recent crystal structure of the epitope I (aa 412-423) peptide in complex with the neutralizing Fab 3/11 revealed an extended conformation of the peptide (Meola A, 2014) that it is in stark contrast to the previously reported β-hairpin obtained for the same peptide in complex with mAbs AP33, HCV1 and Hu5B3.v3 (Kong *et al.*, 2012b; Pantua *et al.*, 2013; Potter *et al.*, 2012). Although available crystal structures of the

epitope II (aa 427-446) so far suggest that this epitope retains its overall conformation (an N-terminal loop and a C-terminal 1.5-turn α-helix), the spatial arrangement of these components is different in complexes with distinct mAbs (Deng *et al.*, 2014; Deng *et al.*, 2013) (Krey *et al.*, 2013). Therefore, the epitope seems to undergo a transition process between open and closed conformations in the context of the E2 core molecule (Kong *et al.*, 2013) Deng, 2014 #665). Deng *et al*., claim that epitope II can induce both neutralizing and non-neutralizing epitopes depending on which conformation of E2 on the virions is presented to the host immune system (Deng *et al.*, 2014). Our results further support the prevailing hypothesis that the CD81 binding region within E2 is highly flexible. Although more evidence is still needed, it is very likely that this flexibility is characteristic not only for *in vitro* produced E2 but also for the E2 on the virion surface.

Whether the flexibility of CD81 binding region plays a substantial role in the natural course of HCV infection remains to be determined. Further, I provide the possible interpretations how this phenomenon might impact the course of HCV infection.

It is tempting to speculate that the flexibility of CD81 binding region may have significant outcomes in HCV infection and disease progression since the virus may use it as a strategy to escape from the host immune system. Another major human pathogen, human immunodeficiency virus (HIV), has also been reported to have a flexible receptor-binding site within envelope glycoprotein gp120 (Myszka *et al.*, 2000). The unoccupied CD4 receptor-binding site has been shown to be disordered or exist in substantially different conformation in comparison with this site bound by the receptor. This is in agreement with the finding that the binding of CD4 receptor site-specific antibodies is accompanied by large negative changes in entropy, which is indicative of protein folding (Kwong *et al.*, 2002). This structural flexibility has been suggested to be used by viruses as an additional immune evasion strategy, helping viruses to avoid antibody-mediated neutralization. It appears likely that the CD81 receptor-binding site within HCV E2 preexists in several conformations in the state unbound by CD81. As a result, the conformational fluctuations may create structural heterogeneity within the receptor-binding region despite its high degree of sequence conservation, which in turn may reduce immunogenicity of this important antigenic site and impede the efficiency with which neutralizing antibodies are elicited. Therefore, the structural transitions of the CD81 binding site within E2 may be beneficial to the virus allowing it to simultaneously retain its receptor-binding function and to evade humoral immune response.

In addition, the structural flexibility of the CD81 binding site may lead to induction of substantial amounts of non-neutralizing antibodies against it. It has already been demonstrated

that epitope II elicits both neutralizing and non-neutralizing antibodies (Duan *et al.*, 2012). Non-neutralizing antibodies have been reported to have detrimental effects for the outcome of a number of viral infections as they can interfere with the binding of neutralizing antibodies to conserved epitopes due to steric hindrance or contribute to infection through interaction with Fc receptors or complement receptors (Takada & Kawaoka, 2003). These host immune system escape mechanisms facilitated by non-neutralizing antibodies may also be valid in the case of HCV infection. First, non-neutralizing antibodies against the CD81 binding loop may prevent the binding of neutralizing antibodies to conserved epitopes due to steric hindrance. It has been described that non-neutralizing antibodies against epitope II interfere with neutralization activity of antibodies specific to a proximal epitope I (Duan *et al.*, 2012; Zhang *et al.*, 2007; Zhang *et al.*, 2009). Our cross-competition experiments with recombinant sE2ΔHVR1 revealed that DAO5 mAb completely abrogates the binding of the neutralizing conformation-sensitive antibodies harboring essential contact residues within the antigenic region aa 523-540. If non-neutralizing antibodies such as DAO5 mAb are elicited during natural HCV infection, they could cross-compete with the neutralizing antibodies and reduce their neutralization potency.

In conclusion, our results indicate that the previously reported structural flexibility epitope I and epitope II composing CD81-binding site further extends to the antigenic site spanning aa 529-540 at least in the soluble glycoprotein. As these three discontinuous segments comprise the CD81 binding site on the surface of E2 glycoprotein, our study supports the hypothesis that the entire CD81 binding site is highly flexible. We have demonstrated that the antigenic site spanning aa 529-540 undergoes structural fluctuations in the recombinantly produced sE2. Further examination of DAO5 mAb binding to HCV particles should provide us with better understanding if the flexibility of this region is also characteristic for E2 displayed on the virions. The intrinsic structural flexibility of epitope I and epitope II has already been proven to occur at the surface of infectious particles. Therefore, we believe that this feature also applies to the third segment of E2 (aa 529-540) involved in CD81 binding. Notably, the two available E2 core structures (PDB IDs 4MWF and 4NX3) differ in their disulphide bridge pattern and share only three disulfide bonds (C494-C564, C508-C552 and C607-C644). Castelli and colleagues has recently assessed the impact of each cysteine residue in E2 expressed in near-native conditions on the binding of conformation sensitive mAbs. This study revealed the presence of alternative cysteine disulfide pairs than the ones observed in E2 core crystal structures obtained by Kong *et al*., and Khan *et al* (Castelli *et al.*, 2014). These observations suggest that E2 glycoprotein is

extremely flexible and the two crystallized E2 core domains may be just a snapshot of a few E2 forms existing at the surface of viral particles.

Normally receptor-binding sites are the most susceptible parts of viral glycoproteins for antibody-mediated neutralization. The conformational flexibility of the CD81 binding site might translate to inefficiency in generating neutralizing antibodies against this site as well as the reduced neutralization potency of these antibodies. This is in agreement with the fact that broadly neutralizing antibodies are usually not detected at the early stages of HCV infection (Logvinoff *et al.*, 2004). A better understanding of the properties of HCV E2 that contribute to neutralization escape of the virus are crucial for development of vaccines or other therapeutics targeting the receptor-binding. Our results imply that the CD81 binding site might be not an ideal candidate for creating a vaccine due to the high degree of flexibility. Presumably, conformational stabilization of the CD81 binding site on E2 may be beneficial in the development an efficient vaccine.

Materials and Methods

Constructs used or generated in the study

The HCV E2ΔHVR1 expression construct (pT350/ E2ΔHVR1) was previously described, and the produced protein was extensively characterized (Tarr *et al.*, 2013).

Anti-HCV E2 mAb DAO5 was obtained by immunizing mice with HCV E2 recombinant proteins from different genotypes. The sequences of mAb were determined by Arvind Patel and Ania Owsianka (Glasgow, UK). The construct expressing a recombinant DAO5 scFv was generated in the lab. The construct expressing recombinant DAO5 Fab was generated as described in a following chapter.

The sequences of anti-HCV E2 mAb were previously published. Anti-HCV CBH-4D (Hadlock *et al.*, 2000), e137 (Perotti *et al.*, 2008), A8 (Allander *et al.*, 2000; Johansson *et al.*, 2007), HC84-1 (Keck *et al.*, 2012; Krey *et al.*, 2013), CBH-23 (Keck *et al.*, 2012), CBH-7 (Keck *et al.*, 2004), HC-1 (Keck *et al.*, 2008a; Wang *et al.*, 2011) and AP33 (Potter *et al.*, 2012; Tarr *et al.*, 2006) Fabs were previously produced recombinantly or by papain digest.

Generating DAO5 Fab construct

DAO5 V_L and V_H were previously cloned into the pMT-scFv-Strep vector in the lab. In order to generate the sequences of the LC (V_L-C_L) and the Fd fragment (V_H-C_{H1}) which could be cloned into the pMT-Fab-Strep vector, the V_L and V_H of DAO5 were fused, respectively, to the L_C and the C_{H1}, of previously characterized F16 Fab. The fusion was carried out by PCR-driven overlap extension (Heckman & Pease, 2007), which is illustrated in **Figure 57**. The resulted PCR products were cloned into the Fab expression vector for production of recombinant DAO5 Fab in *Drosophila melanogaster* S2 cells. The sequences of the primers are:

Fwd_DAO5_HC_fusion	GTGATCGTGAGTAGTGCCAAAACGACACCCCCATCTG
Rev_DAO5_HC_fusion	ACTACTCACGATCACGGATGTTCC
Fwd_DAO5_LC_fusion	GAAGCTGGAGCTGAAGCGGGCTGATGCTGCACCAACTGTATCC
Rev_DAO5_LC_fusion	CTTCAGCTCCAGCTTCGTTCC

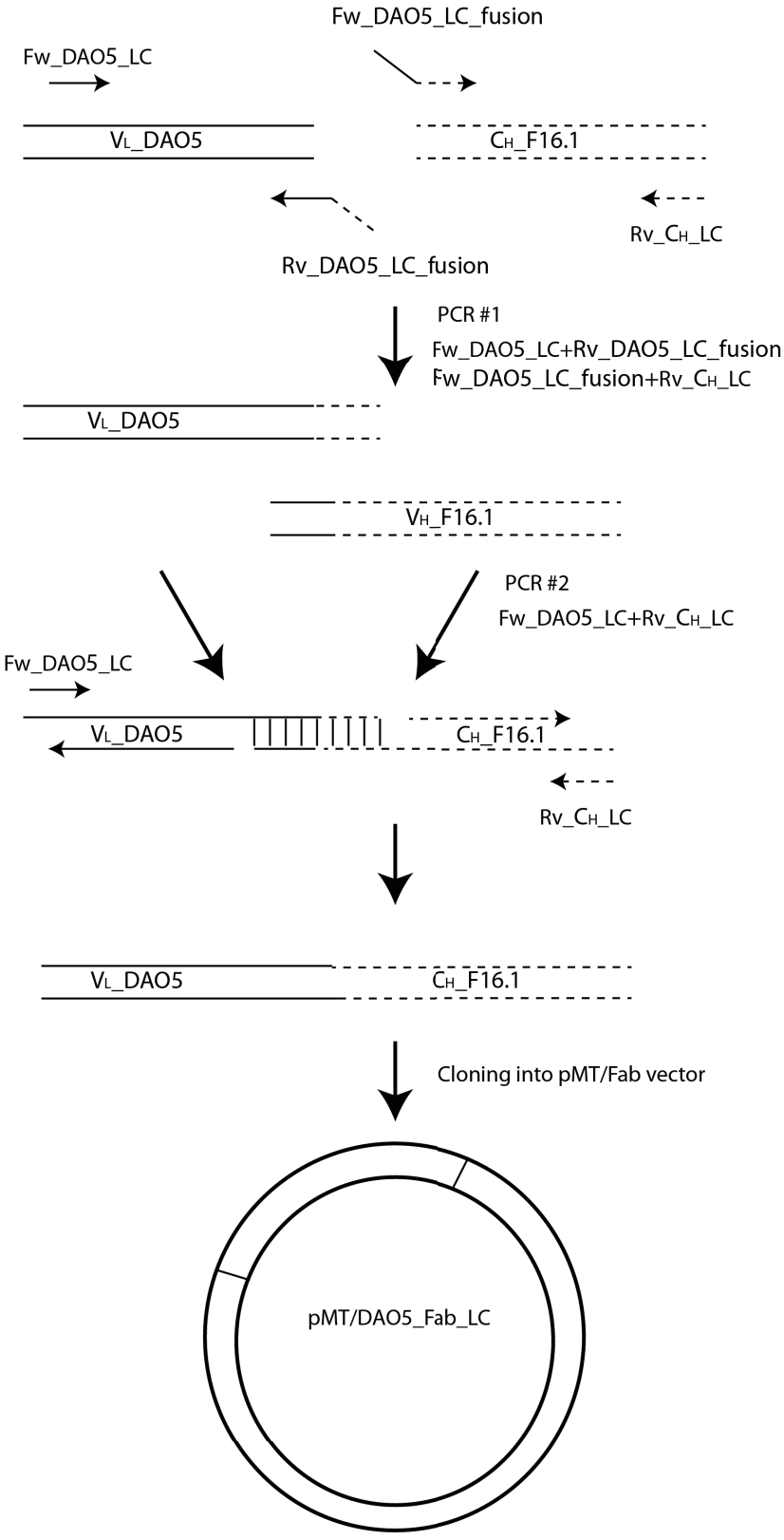


Figure 57. Generating DAO5 Fab construct. Chimeric gene product between DAO5 F16 Fab V_L and F16 Fab C_L was generated by two PCRs. Primers Fw_DAO5_LC_fusion and Rv_DAO5_LC_fusion contained overlapping sequences to DAO5 V_L and F16.1 Fab C_L, respectively (solid line and dashed line). In the first PCR two gene products were generated containing the junction segments between V_L and C_L. The second PCR generated the hybrid product DAO5 V_L- F16 Fab C_H which was inserted into a pMT-Fab-Strep vector by restriction cloning. Similarly, DAO5 V_H- F16 Fab C_{H1} gene was generated and inserted into the same vector (not shown).

CD81 LEL was produced based on the previously described protocol (Kitadokoro *et al.*, 2001). CD81-LEL with a hexahistidine tag (6His) was expressed in *E. coli* as a fusion protein with the IgG binding domain of the *Staphylococcus aureus* protein A (ProtA). The hinge region between CD81-LEL-ProtA contained a thrombin cleavage site.

6 l of TB medium containing 50 pg/ml kanamycin were inoculated 30/1000 with an overnight culture of *E. coli* Rosetta-gami (DE3) strain transformed with CD81-LEL-ProtA in pET28. After growing the cells at 37 ^{o}C to an OD550 of 0.78 the culture was cooled to room temperature. Expression was induced with 0.5 IPTG and the culture was incubated for 20 hours at 28°C. The cells were harvested by centrifugation at 3000 g for 12 min and EndoH was purified from the periplasm of the bacteria. The pelleted cells were resuspended in 2.5% of the original culture volume in ice cold 20% sucrose, 0.1 M Tris pH 8 and 1 mM EDTA and pelleted again by centrifugation at 12000 g for 15 min. Then the cells were resuspended in equivalent volume of ice-cold water and incubated for 30 min on ice. After pelleting the cells at 70000g for 25 min, the supernatant was collected and filtered through 0.22 μM cut-off membrane. The supernatant was loaded onto IgG Sepharose FF (GE) column (~7ml resin) equilibrated with 50 mM TRIS pH 7.6, 150 mM NaCl, 0.05% Tween 20. After washing the column with the same buffer, the protein was eluted with 500 mM Na-Acetate pH 3.4. Elution fractions (2.5ml) contained 1.5ml of TRIS pH 9.0 to neutralize pH. The elution fractions were pooled and concentrated to ~1 ml using Vivaspin centrifugal concentrator (Sartorius). 20ml of 200 mM TRIS pH 8.5 were added and the protein was re-concentrated to ~1 ml. CD81-LEL-ProtA was cleaved with thrombin using Thrombin Cleavage Capture kit (EMD Millipore, Darmstadt, Germany) according to the manufacture's instructions. 1 μl of thrombin per 1 mg of CD81-LEL-ProtA was used. CD81-LEL was further purified using Ni^{2+}ion affinity chromatography followed by SEC on Superdex 75 16/60 column (GE Healthcare).

Removal of the double Strep affinity tag from the recombinant antibody fragments

A C-terminal double Strep tag preceded by an enterokinase recognition site was removed from the recombinant antibody fragments and by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). The detailed protocol is provided in Supplementary Materials and Methods.

Peptides and complex formation

Synthetic peptides comprising HCV E2 residues 529-540 of the J4 strain (GENETDVMLLN) and JFH strain (GENETDVMLLN) were ordered from GenScript and dissolved in 20 mM Tris pH 9 at 10 mg/ml. The DAO5 Fab/peptide complex was formed by mixing protein with the peptide at 1:6 molar ratio (10 mg/ml of Fab+1.72 mg/ml peptide). The complex formation was carried out overnight at 4°C.

Crystallization and crystal soaking in peptide solution

Crystallization conditions for the DAO5 Fab in complex with the peptide_J4 were screened by a Mosquito robot (TTP LabTech Ltd, Royston, UK) at 293 K using the sitting-drop vapor-diffusion method. Rod-shaped diffraction quality crystals appeared after 5 days in a drop containing 0.2 µl of the complex (10 mg/ml of Fab+1.72 mg/ml peptide in 10 mM TRIS pH 8.0, 150 mM NaCl) mixed with an equal amount of reservoir solution containing 20% PEG3350 and 200 mM sodium thiocyanate. The crystals were flash-frozen in liquid nitrogen after transferring crystals to a cryo-protective solution containing the mother liquor and 20% (v/v) glycerol.

Crystals of DAO5 scFv in complex with the peptide_J4 and peptide_JFH were obtained by soaking experiments of unliganded DAO5 scFv crystals obtained by hanging-drop vapor diffusion method in drops containing 1 µl DAO5 scFv (9 mg/ml in in 10 mM TRIS pH 8.0, 150 mM NaCl) and 1 µl of reservoir solution (29% PEG 400, 200 mM MES pH 6.5 and 150 mM sodium acetate). Unliganded DAO5 scFv crystals were transferred to the mother liquor supplemented with 0.1 mM peptide for 5 h (with peptide_J4) or overnight (with peptide_JFH) and were directly flash-frozen in liquid nitrogen.

Structure determination of DAO5 antibody fragments in complex with the peptides

Data were collected at the Synchrotron Soleil beamline Proxima 1, the Swiss Light source beamline PX I at 100 K using a Pilatus detector (Dectris, Baden, *Switzerland)*. The programs iMosflm (Leslie, 2006; Powell, 1999) or xdsme (https://code.google.com/p/xdsme/) were used to determine the optimum orientation of the crystal for the complete data set collection. X-ray diffraction data were processed using XDS (Kabsch, 2010). Scaling and

reduction of data were performed using Pointless (Evans, 2006) and programs from CCP4 suite (CCP4, 1994).

The structure of DAO5 scFv was previously determined by molecular replacement with the program Phaser (McCoy *et al.*, 2007) by Thomas Krey (space group $P4_12_12$.). A search model used for molecular replacement was assembled from the V_L region of a Fab with PDB accession code 1FH5 and the V_H region of a Fab with PDB accession code 3RHV.

In the case of the data set for DAO5 scFv-peptide-JFH crystal, the $P4_122$ space group was suggested by POINTLESS, but the actual space group of the DAO5 scFv crystals used in soaking experiments was previously determined to be $P4_12_12$. However, detection of systematic absences is sometimes is unreliable because they may be missing from the data set if they lie in the blind region (i.e. along spindle rotation axis in the data collection). This situation is more common for long needle or rod-shaped crystals; thus, it was likely to occur during the data collection on the DAO5 scFv-peptide-JFH crystal. Therefore, the list of reflections was analyzed to check if there are any systemic absences along the k reciprocal axis. The systematic absence of the spots with Miller indices 0k0 where k=2n+1 was observed suggesting a two-fold axis along the b-axis in real space. The output from SCALA revealed unusually high R_{merge} values for the data set of DAO5 scFv-peptide_JFH crystal when all 800 frames were used. Therefore, the R_{merge} was analyzed versus the batch number to determine if the data quality had declined during the collection time. Indeed, an increase in R_{merge} was detected for the second half of the batches, likely as a result of radiation damage. Since the crystal belonged to a high symmetry space group ($P4_12_12$), the batches with high R_{merge} could be excluded from the data set without affecting its completeness. The data processing, scaling and reduction was performed on the first 300 out of 800 frames.

Since the structure of unliganded DAO5 scFv was already available, the phases for the structures of DAO5 scFv in complex with the peptides were calculated by the molecular replacement method with the program Phaser (McCoy *et al.*, 2007) using the previously determined coordinates from the native crystal structure of DAO5 scFv as a search model. Each crystallographic asymmetric unit contained two DAO5 scFv/peptide complexes.

Fab fragments have a flexible linker region (elbow angle) between the variable domains (V_L and V_H) and the constant domains (C_{H1} and C_L), which requires a Fab search model with an approximately correct elbow angle for molecular replacement. Alternatively, the individual CH_1/C_L and V_H/V_L domains of Fab (excluding the residues composing the elbow angle) can be used as search models. The molecular replacement solution for the DAO5 Fab was obtained by using two ensembles as a search model: the V_LV_H region of

DAO5 scFv, and the C_HC_L region from anti-GBV-B E2 F16.1 Fab (unpublished). The residues composing the elbow angle were excluded from the search model. Each crystallographic asymmetric unit contained one DAO5 scFv/peptide_JFH complex. The resolution limit of 1.7 Å was chosen for the refinement of DAO5 Fab-peptide_J4 model.

Refinement for all three crystal structures was implemented using AutoBuster (Bricogne G, 2010). Because DAO5 scFv/peptide_J4 and DAO5 scFv/peptide_JFH complexes contained two molecules per asymmetric unit, non-crystallographic symmetry (NCS) restraints were applied throughout model refinement. The same *R-free* test set which was used in the refinement of previously determined DAO5 scFv crystal structure had to be maintained throughout the refinement of the models of the DAO5 scFv-peptide complexes. The datasets of the DAO5 scFv-peptide complexes were of higher resolution (1.9 and 2.00 Å in compared to 2.05 Å for native DAO5 scFv crystals). Therefore, the R-free test set of unliganded DAO5 scFv contained only the reflections going to 2.05 Å. The existing R-free test set was extended to the entire resolution range of DAO5 scFv-peptide complexes by adding reflections from the highest resolution bins of those datasets.

Manual model building was carried out using Coot (Emsley *et al.*, 2010) Although after molecular replacement, electron density maps of all three complexes revealed unambiguous density for the peptide in the antigen-binding site, the peptides were omitted in the initial refinement and model building cycles and were built into the electron density in the antigen binding site after the refinement and building of the scFv/Fab structures was completed. Water molecules were added by using AutoBuster and verified manually. Details of the statistics of data collection and refinement are presented in the chapter Results.

Crystal structure analysis

Multiple structural superposition of the peptides from three crystal structures and calculations of root mean square deviation (RMSD) over all atoms, CA atoms, and backbone atoms (N, CA, C, O) were performed using SuperPose 1 (Maiti *et al.*, 2004). To compare the crystal structures of two peptides (peptide_J4 in Fab and scFv structures or peptide_J4 and peptide_JFH in scFv structures), the corresponding peptides were superposed using USSF Chimera (Pettersen *et al.*, 2004) and RMSD were calculated between every residue in two peptides including all atoms or only backbone atoms (N, CA, C, O). Surface complementarity coefficients were calculated using SC from the CCP4 suite. Electrostatic potentials were calculated using the adaptive Poisson Boltzmann solver (Baker *et al.*, 2001). The protein

interaction calculator (Molina *et al.*) (Tina et al., 2007) was used to identify the interactions between the peptides and scFv/Fab. The surface area buried upon complex formation for the interfaces and for the individual residues within the peptide was estimated using the PISA server (Krissinel, 2007). Figures of the crystal structures were prepared in the PyMOL Molecular Graphics System (www.pymol.org). Composite omit maps were generated using CNSsolve (Brunger *et al.*, 1998).

SEC analysis of sE2ΔHVR1 complexes with the antibody fragments

20-30 µg of sE2ΔHVR1 and an equimolar amount of the antibody fragment (Fab or scFv) were incubated as isolated proteins as well as in complex for 16 h at certain temperature (4°C, 18°C, 30°C or 37°C) followed by analysis on a Superdex 200 10/300 column (GE Healthcare, Uppsala, Sweden).

sE2ΔHVR1/DAO5 Fab complex purified by SEC was used. The second Fab was added to sE2ΔHVR1/DAO5 Fab complex and the proteins were incubated for 16 h at 4°C before analyzing them on a Superdex 200 10/300 column (GE Healthcare).

Cross-competition analysis of DAO5 Fab and CD81 LEL by SEC

100 µg of E2ΔHVR1 and pre-formed sE2ΔHVR1/DAO5 Fab complex was incubated with a molar excess of CD81 LEL over night at room temperature. The next day the complexes as well as individual proteins (E2ΔHVR1 and CD81 LEL) were loaded onto a Superdex 200 10/300 column (GE Healthcare, Uppsala, Sweden) and the peak fractions were collected. The peak fractions were concentrated and analyzed by SDS-PAGE under non-reducing conditions. The gel was stained with Coomassie Blue.

Pull-down experiments

Pull-down experiments were performed to analyze the conformation and cross-competition profile of sE2ΔHVR1/DAO5 complex. The pull-down assays were performed on a Streptactin Superflow mini column (0.2ml column bed volume) using the Strep-tagged E2ΔHVR1 and the Fabs without the Strep-tag.

To analyze the cross-competition between mAb CBH-4D and DAO5, 400 µg of pre-

formed E2ΔHVR1/DAO5 scFv complex was affinity bound to a Streptactin Superflow mini column and washed with 10 column volumes of washing buffer. Subsequently, an excess of conformation dependent Fab CBH-4D Fab was loaded onto the column followed by a washing step. The complex was eluted in 4.5 column volumes of elution buffer.

To analyze the cross-competition between mAb e137 and DAO5, a molar excess of e137 Fab was passed through the column with affinity bound E2ΔHVR1 followed by DAO5 scFv. The column was washed with 10 column volumes after addition of each protein. The complexes were eluted as described above. The experiment was also performed by first loading a molar excess of DAO5 scFv followed by e137 Fab. 15 µl of elution fractions were analyzed by SDS-PAGE followed by Coomassie Blue staining.

Surface plasmon resonance (SPR) experiments

SPR experiments were performed on a Biacore 2000 system, equilibrated at 18°C in 10 mM Tris pH 8.0, 500 mM NaCl buffer using a CM5 sensor chip with a density of around 17000 response units (RU) of covalently immobilized anti-Strep antibody. E2ΔHVR1 used in the experiments contained the Strep-tag allowing its capture on the chip via the Strep-tag.

The anti-Strep antibody was immobilized via primary amines using the following protocol:

1) The sensor chip surface was activated with a freshly prepared 1:1 mixture of 0.4 M EDC (N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride) (GE Healthcare) and 0.1 M NHS (N-hydroxysuccinimide) (GE Healthcare) in water;

2) Monoclonal anti-Strep antibody diluted in PBS to 37 µg/ml in 10 mM sodium acetate, pH 5.5 (Biacore, Uppsala, Sweden) was injected into the experimental flow cells of the sensor at a flow rate 5 µl/min for 20 min in order to saturate the surface.

3) The sensor chip surface was deactivated with 1 M 2-aminoethanol pH 8.5 (GE Healthcare) by injecting a reagent at flow rate 5 µl/min for 12 min)

Diluted E2ΔHVR1 were injected over the sensor chip (3 min, 5 µl/min). After 5 min of dissociation time, DAO5 Fab and e137 Fab were sequentially injected over pT424 (2 min each, 5 µl/min). The injections of the Fabs were also carried out in reversed order (e137 Fab

followed by DAO5 Fab). The second Fab was injected after the first Fab reached the saturation of the surface with non-covalently immobilized pT424.

E2ΔHVR1 / DAO5 Fab and E2ΔHVR1 / e137 Fab complexes pre-formed by incubating the proteins overnight at 30 ^{o}C were injected over the sensor chip (4 min, 5 µl/min). After 5 min of dissociation time, DAO5 Fab or e137 Fab were injected over E2ΔHVR1 / DAO5 Fab and E2ΔHVR1 / e137 Fab complexes, respectively (6 min, 5 µl/min).

All Fabs used in the experiment did not contain the Strep-tag in order to measure a specific binding of the Fab to E2ΔHVR1 and E2ΔHVR1 / DAO5 Fab and E2ΔHVR1 / e137 complexes immobilized non-covalently on the chip via anti-Strep antibody. The second Fab was injected after the first Fab reached a saturation state. After injections of the Fabs, the sensor chip surface was regenerated using glycine-HCl pH 2.0 (2 min, 5 µl/min) and 0.1% SDS (1min, 5 µl/min). All injections were carried out in triplicate. The data were processed using Scrubber software (BioNavis, Ylöjärvi, Finland) and double referenced by subtraction of the blank surface and buffer-only injection before local fitting of the data.

Chapter II.

Baculovirus Envelope Protein F

Introduction

Baculoviridae family

Baculoviruses have been known for many centuries since the first written accounts describing the infection of Chinese silkworms. These large rod-shaped (30–60 × 250–300 nm) viruses belong to the virus family *Baculoviridae* and are pathogenic almost exclusively to insects of the order *Lepidoptera*, *Hymenoptera* and *Diptera*. The genomes of baculoviruses consist of a large (80-180 kbp) circular double-stranded DNA genome containing approximately 120 to 160 open reading frames. Classically, baculoviruses were divided into two genera, nuclear polyhedrosis viruses (NPVs) and granulosis viruses, based on virion morphology. The current classification, however, is based on phylogeny of baculoviruses and classifies its members into four genera: *Alphabaculovirus* (lepidopteran-specific NPVs), *Betabaculovirus* (lepidopteran-specific granulosis viruses), *Gamabaculovirus* (hymenopteran-specific NPV) and *Deltabaculovirus* (dipteran-specific NPV) (Jehle *et al.*, 2006) (**Figure 58**). The *Alphabaculovirus* genus has been further subdivided into two groups (group I and group II), which correlates with the usage of two different membrane fusion proteins, gp64 and F protein, respectively. Protein "F" stands for "fusion" protein, and has amino acid sequence features characteristic of the paramyxovirus fusion F. The structural characterization of the baculovirus F protein is the subject of the second part of my PhD thesis.

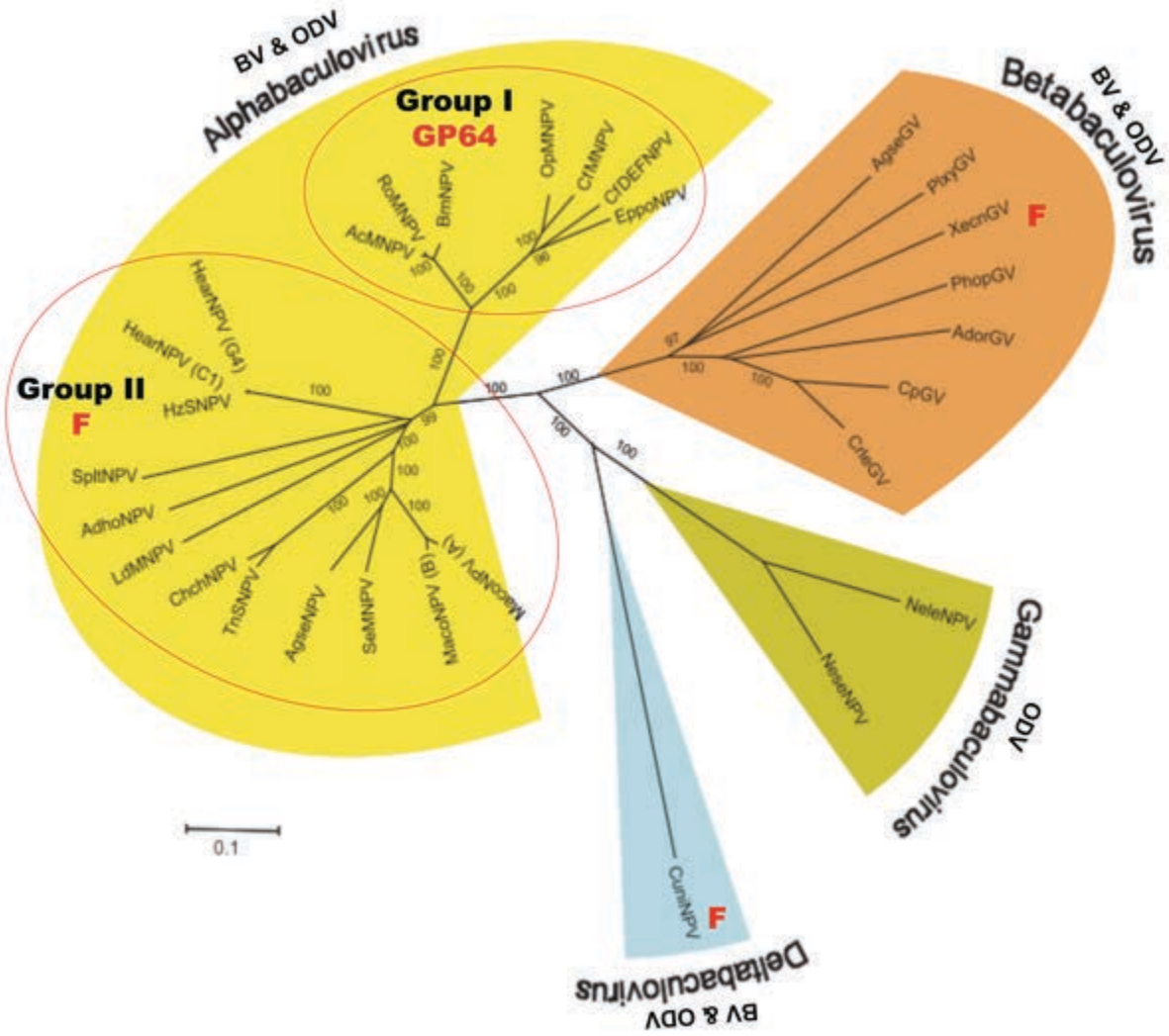


Figure 58. The baculovirus classification. The baculovirus tree consists of 4 major clades based on alignment of amino acid positions of sequenced baculoviruses. The best-studied genus is *Alphabaculovirus*, which includes all the lepidopteran-specific nucleopolyhedrovirus. Alphabaculoviruses can be subdivided into two groups, group I and group II, based on the membrane fusion protein, which in budded virions (BV) of Group I is GP64 and in Group II F, which is also version of F. Gammabaculoviruses exist only as occlusion derived viruses (ODV). Adapted from (Jehle *et al.*, 2006).

Baculoviruses have a number of useful applications. First of all, since they are natural enemies of many insects, they can be applied as bioinsecticides. They are also used as efficient eukaryotic expression vectors for production of recombinant proteins and vaccines. Recently, certain baculoviruses have been demonstrated to represent potential gene therapy vectors due to their capacity to transduce mammalian cells (Airenne *et al.*, 2013).

Replication cycle

During their lifecycle the baculoviruses commonly can be found in two distinct phenotypes: occlusion-derived virus (ODV), and budded virus (BV). ODVs are immobilized in a protein matrix (polyhedrin or granulin) and are responsible for the infection of the midgut epithelial cells (Coulibaly *et al.*, 2009) (**Figure 59**). NPVs form large (1–15 μm) polyhedral inclusion bodies, and based on the number of the nucleocapsids in the inclusion body, are subdivided to single (S) NPVs or multiple (M) NPVs. Inclusion bodies of granuloviruses are small and contain only a single virion. The virus released from the midgut epithelial cells is BV, which is highly infectious and is responsible for systemic infection of an insect.

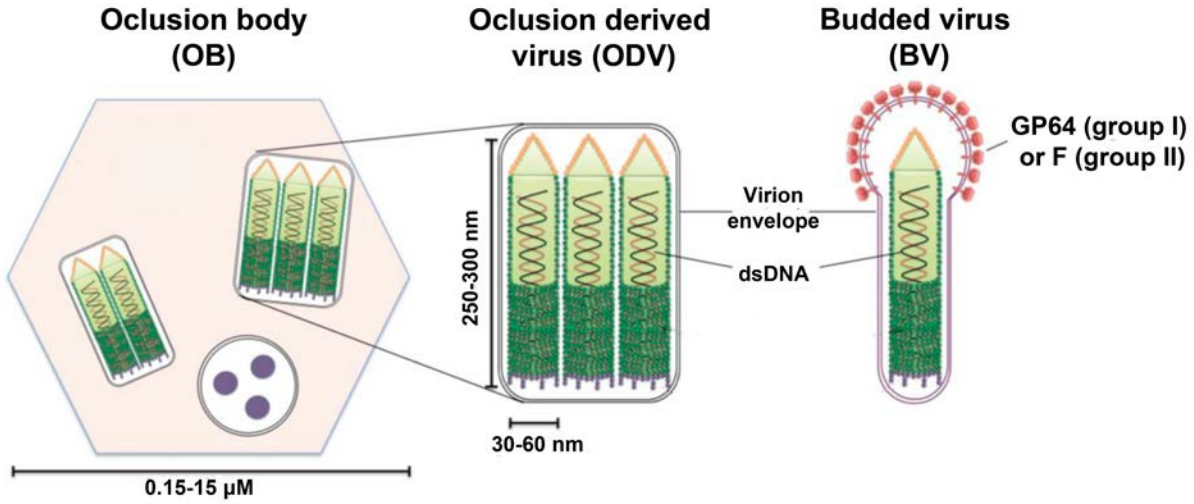


Figure 59. Structure of baculovirus occlusion bodies (OB) of nucleopolyhedroviruses (NPVs), occlusion-derived virion (ODV) and budded virion (BV) and form occlusion bodies (OBs). ODVs are responsible for the infection of the midgut epithelial cells. BV envelope contains the fusion protein GP64 (NPV group I) or F protein (NPV group II). BVs are responsible for systemic infection of an insect. Adapted from (Au *et al.*, 2013).

Since insect populations are seasonal, baculoviruses evolved a strategy to survive in the environment until their hosts reappear by immobilizing the virus in a protein matrix (polyhedrin or granulin) of the occlusion bodies. These occlusion bodies are crystalline, and the crystals diffract to high resolution resulting in determination of the structure of baculovirus polyhedra determined by X-ray crystallography (Coulibaly *et al.*, 2009). A baculoviral infection starts with the ingestion of occlusion bodies by a larva. In the midgut of the insect, occlusion bodies dissolve due to the alkaline pH liberating ODVs that infect the midgut columnar epithelial cells. When nucleopcapsids reach the nucleus of the cells, the DNA is released and expression and replication of the viral genome occurs.

The viral envelope proteins (gp64 or F) are transported to the plasma membrane of the infected cell (**Figure 60**). The nucleocapsids assemble in the nucleus and migrate towards the plasma membrane, where they bud and acquire the viral glycoproteins. BVs propagate the infection throughout the insect, entering cells via clathrin-mediated endocytosis (Long *et al.*, 2006a). They fuse with the endosomal membranes in an acidic pH-induced membrane fusion reaction, as with many other viruses. The released nucleocapsids are transported to the nucleus, where transcription and replication is initiated. In the late stages of infection, the occlusion bodies are formed in the nuclei of infected cells and are released from the cells after the death and disintegration of the host (Rohrmann, 2013).

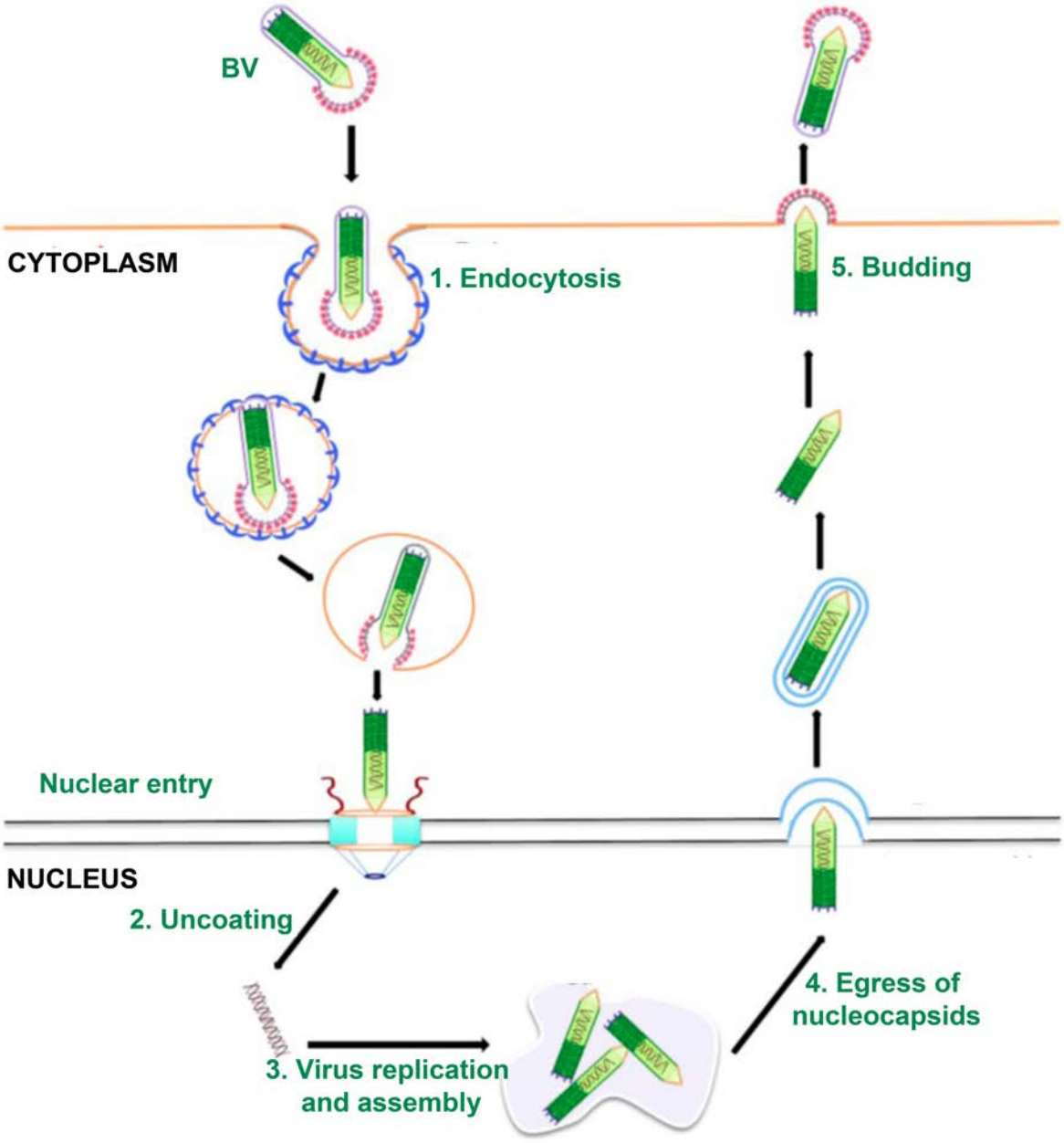


Figure 60. Budded virus infection cycle. See text for detailed description. Adapted from (Au *et al.*, 2013).

Baculovirus envelope fusion proteins

The entry of baculovirus BVs into the host cells is mediated by a specific envelope glycoprotein, either gp64 or F (Blissard & Rohrmann, 1989; WF *et al.*, 2000; Whitford *et al.*, 1989). Both proteins have similar biological functions, and have been demonstrated to be involved in receptor binding, low-pH dependent fusion of virus and cellular membranes, and efficient budding (Blissard & Wenz, 1992; Hefferon *et al.*, 1999; Long *et al.*, 2006b; Oomens & Blissard, 1999; WF *et al.*, 2000). However, the molecular basis of action of these proteins is significantly different.

The three-dimensional structure of gp64 has been determined by X-ray crystallography, revealing that it belongs to the structural class III fusion proteins, which do not require proteolytic cleavage in order to be activated (Kadlec *et al.*, 2008). The F protein has been suggested to perform an analogous function to gp64 based on the fact that infectivity *gp64*-null *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) can be restored by the introducing F (Lung *et al.*, 2002).

F proteins are not only more widespread within *Baculoviridae* family but also more diverse (with amino acid identity 20-40% and >74% for F and gp64, respectively). Phylogenetic analyses imply that the F protein is likely to be an ancestral fusion protein of baculoviruses while gp64 has been incorporated into the baculovirus genome relatively recently (Jiang *et al.*, 2009; Pearson & Rohrmann, 2002). Group I alphabaculoviruses still encode a non-fusogenic F protein homologue (F-like protein) which most likely lost its fusogenic function due to the acquisition of gp64 during evolution (Lung *et al.*, 2002; Wang *et al.*, 2008).

Characteristics of baculovirus F protein

One of the most studied F proteins comes from the group I alphabaculovirus *Spodoptera exigua* multicapsid nucleopolyhedrovirus (SeMNPV). SeMNPV is very species-specific and infects only the larvae of the beet armyworm (*Spodoptera exigua*). The beet armyworm is a pest for many cultivated crops including vegetables, flowers, and cotton, causing significant economic losses. SeMNPV is licensed as a bioinsecticide and is quite effective when applied in greenhouses, where it is protected from ultraviolet light from the sun (Bianchi *et al.*, 2002). The complete SeMNPV genome was sequenced in 1999, revealing

that it lacked a homologue of *gp64*. The ORF8 of SeMNPV encoded another envelope protein, which was shown to be sufficient to mediate membrane fusion in syncytium formation assays at low pH (WF *et al.*, 2000). This functional homologue of gp64 was called fusion (F) protein.

Biochemical characterization of SeMNPV F protein has revealed that this protein contains the features characteristic to class I fusion proteins. SeMNPV F protein is translated as ~76kD precursor called F_0. As for other class I fusion viral proteins, the precursor is posttranslationally cleaved. In the case of SeMNPV F, this cleavage is mediated by subtilisin-like endoprotease furin. The mutation of the furin cleavage site RSKR (amino acid residues 145-149) results in a loss of the fusogenic activity, indicating that the cleavage is essential to generate a fusion-competent protein (Westenberg *et al.*, 2002).

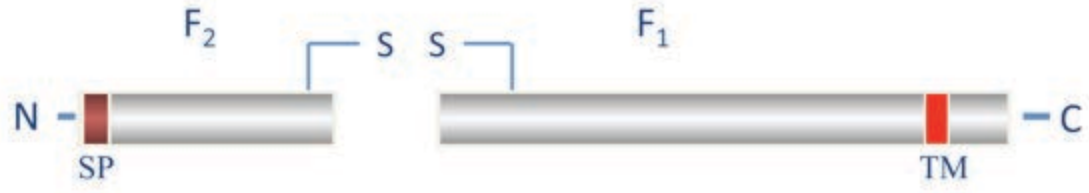


Figure 61. SeMNPV F protein after furin cleavage. Furin cleavage generates two disulphide-linked subunits F_2 and F_1. The larger F_1 subunit is membrane anchored by its transmembrane domain. TM: transmembrane domain, SP: signal peptide.

The furin cleavage yields two disulphide-linked subunits: a small N-terminal subunit F_2 (theoretical molecular weight ~15kD), and a bigger membrane-anchored C-terminal subunit F_1 (theoretical molecular weight ~59kD) (**Figure 61**). F_2 subunit contains only one cysteine residue C94, which must pair with one of the ten cysteine residues in the F2 subunit located upstream from the predicted transmembrane domain. Thus, the ectodomain of the F2 subunit contains a single cysteine residue.

The class I fusion proteins form homotrimers at the surface of a viral particle. The exact oligomeric state of SeMNPV F is unknown but it is predicted to be trimeric based on the fact that the F protein of closely related group II alphabaculovirus, *Helicoverpa armigera* multicapsid nucleopolyhedrovirus (HearMNPV), assembles as trimers on the virus envelope (Long *et al.*, 2006b). Some viral fusion proteins have been shown to form higher-order oligomers via covalent disulphide links. Whether or not the free cysteine residue in the F2 subunit is involved in oligomerization of SeMNPV F by making interprotein disulphide linkages remains to be determined.

The furin cleavage occurs upstream of the hydrophobic sequence located at the N-terminus of the membrane anchored F1 subunit. The stretch of the first 18 amino acid residues at the N-terminus of F1 subunit (150-GLFNFMGHVDKYLFGIMDS-168) has been suggested to represent a putative fusion peptide because it contains features characteristic of previously described viral fusion peptides (White, 1992): 1) it is hydrophobic, 2) it can form an amphipathic helix with conserved glycines at one side, and 3) it shows a high degree of conservation among baculovirus F proteins. However, some differences with vertebrate viral fusion proteins can be identified such as the absence of alanine residues and a higher number of polar residues. The deletion of amino acid residues 151-170 in F protein resulted in the loss of virus infectivity even though the protein was incorporated into viral particles and was cleaved by furin, supporting the role of this amino acid stretch as a fusion peptide (Westenberg *et al.*, 2004).

The structural similarity between baculovirus F and paramyxovirus F proteins

The SeMNPV F protein has been predicted to share structural features with the paramyxovirus F protein. In contrast to baculoviruses, which are DNA viruses, paramyxoviruses are single-stranded, negative-sense RNA viruses. The envelope glycoprotein F is responsible for virus fusion with the cellular membrane. Paramyxovirus F proteins are also synthesized as F_0 precursors and cleaved in the trans-Golgi complex by furin into two subunits, F_1 and F_2, with the fusion peptide located at the newly generated N-terminus of the F_1 subunit.

The important difference between the entry of baculoviruses and paramyxoviruses is that the latter fuse at the cell surface in a pH-independent manner, while baculoviruses mediate fusion in the endosomes in an acidic-pH environment. Moreover, paramyxoviruses uses a separate attachment protein for binding to the cell surface, in contrast to baculoviruses, which have both membrane fusion and receptor binding activities in the same envelope glycoprotein F. The interaction of the paramyxovirus attachment protein with the cellular receptor triggers conformational changes in the prefusion F trimers, which eventually lead to the fusion of viral and cellular membranes (reviewed in (Bossart *et al.*, 2013; Jardetzky & Lamb, 2014; Plattet & Plemper, 2013).

Of all paramyxoviruses, the F protein of human respiratory syncytial virus (RSV) was identified as having the highest amino acid identity (12%) to SeMNPV F. Despite low amino acid sequence identity, these two proteins display conserved positioning of the furin cleavage sites, the secondary structure elements, and the transmembrane domains which is an indication of homology between their corresponding genes.

The crystal structures of F in pre-fusion and/or post-fusion from has been determined for a number of paramyxoviruses: human parainfluenza virus 3 (hPIV3), human parainflueza virus 5 (hPIV5), New Castle disease virus (NDV) fusion protein The crystal structures of RSV F in its post-fusion and pre-fusion forms are available (Swanson *et al.*, 2011) (McLellan *et al.*, 2013) revealing a class I fusion protein fold observed in previously solved F structures from other paramyxoviruses (human parainfluenza virus 3 (hPIV3), parainflueza virus 5 (PIV5), Newcastle disease virus (NDV), Simian virus 5 (SV5) and metapneumovirus (MV)) ((Baker *et al.*, 1999; Chen *et al.*, 2001; Swanson *et al.*, 2010; Welch *et al.*, 2012; Wen *et al.*, 2012; Yin *et al.*, 2005). The RSV F post-fusion trimer (**Figure 62**) has an elongated shape with a globular head domain on the top and long stalk domain constituted by intertwined α-helices that form a stable 6HB at the membrane proximal end of the molecule. The helices of the 6HB are composed by two HRs: HRA adjacent to the fusion peptide, and HRB located upstream from the transmembrane region. HRAs of three protomers form a central core of a triple-stranded coiled-coil creating three grooves into which the C-terminal HRBs pack antiparallel to the central core. In contrast, the most prototypical class I fusion protein, influenza HA, forms just a small 6HB at the membrane distal end of the protein as it contains just an extended segment the C-terminal HR helix that packs into the grooves of the central N-terminal coiled-coil. Each protomer of RSV F in its post-fusion form is composed of three domains, termed DI, DII and DIII. The globular head domain is mostly composed of D1 and DII. At the base of the head domain is DIII, which carries a long HRA helix that extends down and intertwines with HRA helices of the other two protomers to form a central coiled-coil of the 6HB. HRB helices extending from DII interact with the central coiled-coil to form the outer helices of a 6HB. The fusion peptide is located at the N-terminus of HRA while the transmembrane region (not present in the crystal structure) is positioned at the C-terminus of HRB. These two elements are located at the bottom of the stalk and are inserted into the cellular membrane in the full-length F.

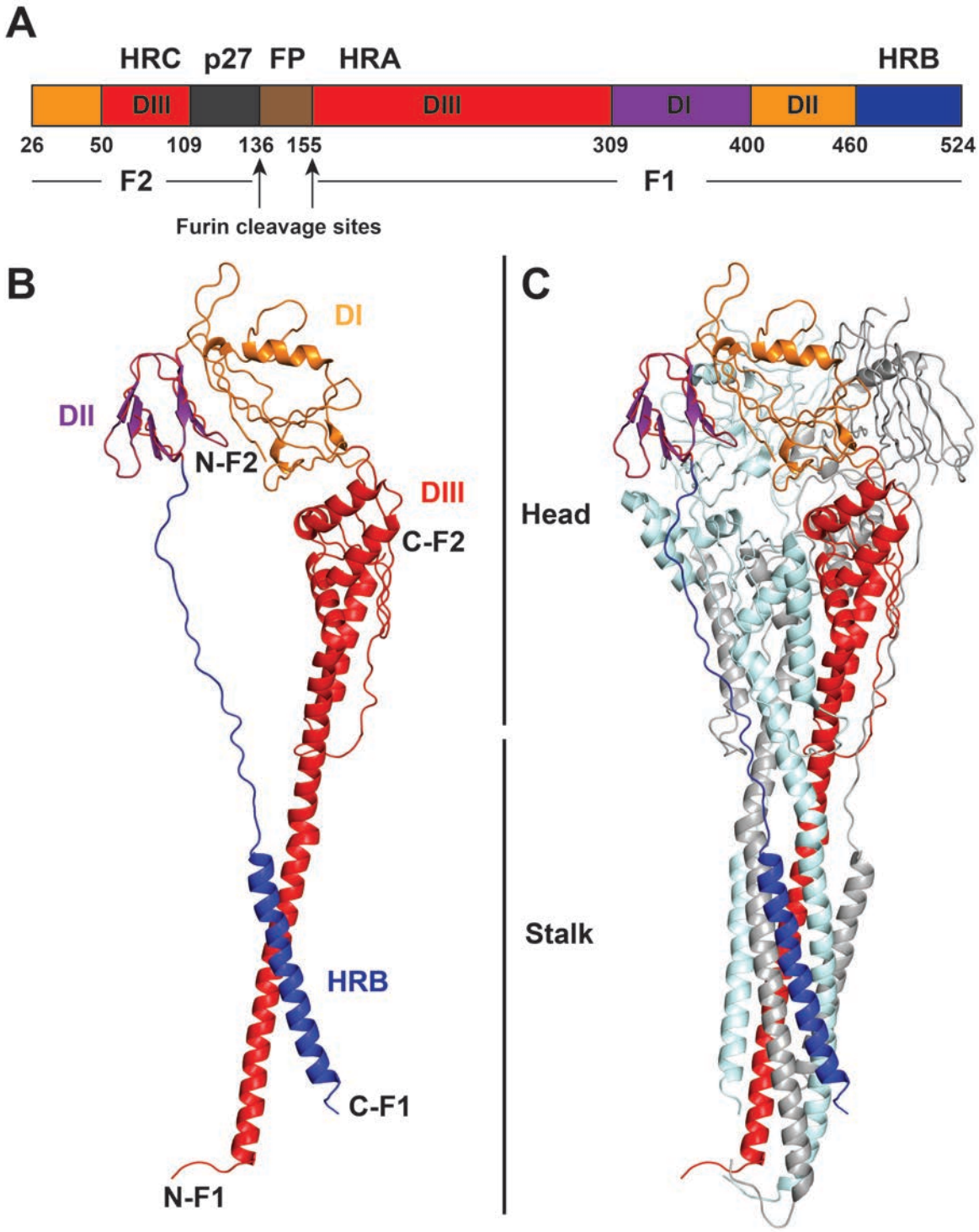


Figure 62. RSV F ectodomain structure in its post-fusion form. (A) Linear diagram of the RSV F ectodomain. The furin cleavage sites are indicated by black arrows. FP: fusion peptide; HRA, -B and -C: heptad repeats; DI-DIII domains I-III, p27: peptide removed by furin cleavage. (B) Cartoon representation of one protomer of the post-fusion trimer colored by domains as in A. (C) Cartoon representation of the post-fusion trimer with one protomer colored by domains as in A and the remaining two colored in grey and light blue.

The corresponding putative HRs (HR1 and HR2) are also found in SeMNPV F protein. They are predicted to encompass amino acid residues 174-202 and 521-549

respectively. The three-dimensional structure of SeMNPV F is predicted to be similar to that of RSV F (Misseri *et al.*, 2003).

Receptor-binding function of baculovirus F

The insect cellular receptor used by baculoviruses has not yet been identified. Some studies indicate that gp64 and F proteins of baculoviruses interact with distinct insect cell receptors (Hefferon *et al.*, 1999; Westenberg *et al.*, 2007; Wickham *et al.*, 1992). The prototypic member of group II alphabaculoviruses *Autographa californica* (AcMNPV) containing gp64 is able to transduce a number of mammalian cell types through the endocytosis pathway, indicating that that the cells carry the receptor recognized by gp64 at their surface. gp64, however, is also able to enter mammalian cells via direct fusion with the plasma membrane under low pH (Dong *et al.*, 2010). gp64-null Autographa californica (AcMNPV) pseudotyped with baculovirus F is unable to enter mammalian cells suggesting that the F protein receptor is not present on mammalian cells.

F-null group II baculovirus pseudotyped with gp64 and non-fusogenic F protein (mimicking the composition of the envelope glycoproteins in group I alphabaculoviruses), however, has been demonstrated to successfully enter insect cells by endocytosis. In this study, non-fusogenic F protein was the primary contributor to the binding to the cells while gp64 was only a minor contributor to the binding process. These results imply that F-like protein may be a major receptor binding protein for entry to insect cells (Wang *et al.*, 2014).

Relationship between insect retroviruses and baculoviruses

Retroelements with long-terminal repeats (LTRs) are found in the majority of eukaryotic genomes. The genomes of all cells contain a number of transposable elements integrated into their genomes. Exogenous retroviruses, endogenous retroviruses (ERVs), and LTR-retrotransposons are one of the sources of retroelements. These transposable elements have been described as being able to cross species barriers by horizontal transfer (Jordan *et al.*, 1999). The insect retroelements encoding an envelope (*env*) gene have been classified into the *Errantivirus* genus of the *Metaviridae* family and their *env* gene has been shown to share common ancestry with the gene encoding baculovirus F (Malik *et al.*, 2000). The sequence

similarity is the highest in the region that includes the furin cleavage signal and a predicted fusion peptide.

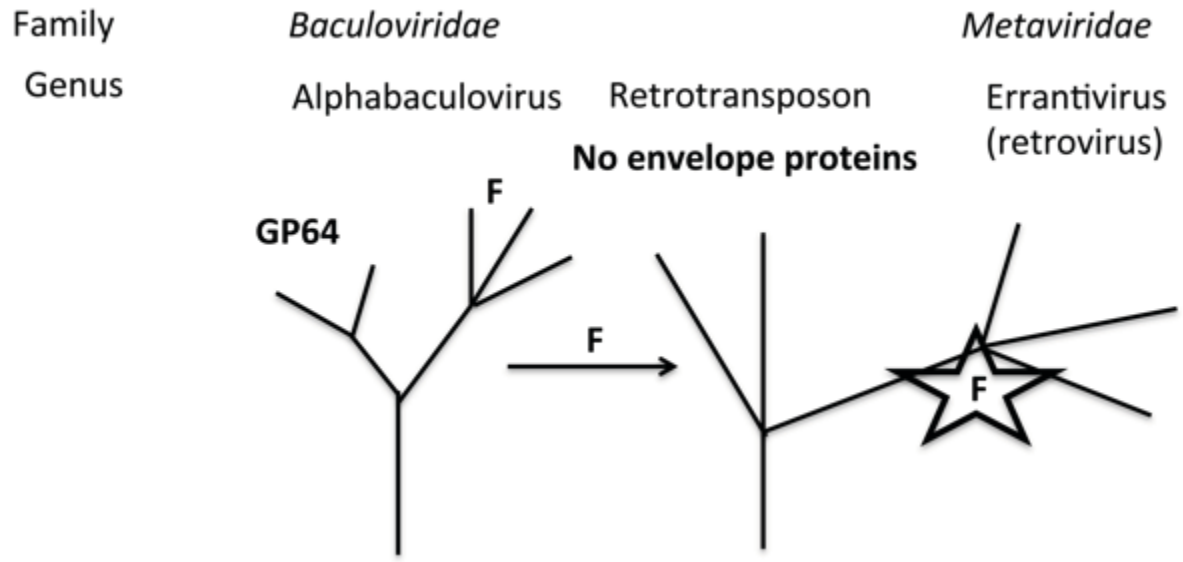


Figure 63. Baculovirus F protein orthologs are found in genomes of some insect retroviruses (errantiviruses). Evolutionary, they most likely originated from a retrotrasposons that incorporated F protein from baculovirus.

The baculovirus origin of *env* genes found in insect errantiviruses has been proposed based on the fact that both errantiviruses and baculoviruses have the same hosts and LTR-retrotransposons have been previously found incorporated into baculovirus genomes (Friesen & Nissen, 1990) (**Figure 63**). As a result, *env* genes could have been acquired via intragenomic recombination events that occurred after integration of LTR-retrotransposons into the genome of baculoviruses (Pearson and Rohrmann, 2002)(Malik *et al.*, 2000). *Drosophila melanogaster* genome encodes a number of errantiviruses with *gypsy* being the most studied retrovirus-like element in this organism. In contrast to retroviruses, errantiviruses are considered to be non-infectious. However, *gypsy* Env has been demonstrated to localize to the cell membrane of insect cells and to posses fusogenic properties (Song *et al.*, 1997) (Kim *et al.*, 1994) (Misseri *et al.*, 2004).

Cellular orthologs of baculovirus F protein

In addition to insect retroviruses, the F protein gene was identified in the genomes of four dipteran species: *Anopheles gambiae*, and the fruit flies *Drosophila melanogaster*, *Drosophila yakuba*, and *Drosophila pseudoobscura* (Lung & Blissard, 2005). Phylogenetic

studies have suggested that these F-like genes (named *iris*) were incorporated into the *Drosophila* genome 25 million years ago from endogenous retroviruses. However, the Iris found in *Drosophila* species *(D.melanogaster, D. yakuba and D. pseudoobscura)* lack the structural elements of a fusion protein such as a predicted furin cleavage site, a predicted fusion peptide, and a coiled-coil domain. Moreover, a study that analyzed a possible membrane fusion activity of *D.melanogaster* Iris did not detect this protein as capable to mediate membrane fusion and showed that it localized to mitochondria and not to the cellular membrane as baculovirus F or gypsy Env (Lung & Blissard, 2005). The *Anopheles gambiae* F protein has a potential furin cleavage site but the cleavage at this site has not yet been demonstrated.

The ectodomains of lepidopteran baculovirus F proteins contain 10 conserved cysteine residues, but only 6 of them (C5-C10) are found in Drosophila Iris and the F protein from mosquito baculovirus CuniNPV. In *Anopheles gambiae* mosquitoes, C9 is also missing. The spacing between those cysteine residues in lepidopteran baculovirus F proteins and insect cellular F protein is conserved. In addition, 6 highly conserved non-cysteine residues were identified between the cellular F and lepidopteran baculovirus F. Those conserved residues are located in the central and C-terminal portion of F and reside within the borders of so-called domain of unknown function (DUF3609) (conserved domain accession: pfam12259). This domain of ~360 amino acids in length has been previously recognized in eukaryotes and in viruses.

Although it has been proposed that *iris* was incorporated into the *Drosophila* genome from endogenous viruses, the opposite hypothesis should also be considered. It could be that the *f* gene in baculoviruses was acquired directly or indirectly from their insect hosts and evolved to acquire fusion activity. Although baculoviruses infecting the *Drosophila* and *Anopheles* species are not known, baculoviruses infecting other dipterans have been reported (Becnel *et al.*, 2001; Federici, 1980). It could be that ancestral baculoviruses existed only as occlusion derived virions and were able to replicate only in the epithelial cells of the insect midgut (which is still the case for sawfly baculoviruses). By acquiring a cellular *f* gene they were able to infect the insect hemocel, which resulted in the evolution of a BV phenotype.

The indirect route of acquiring the *f* gene in baculoviruses might involve insect endogenous viruses. As mentioned earlier, insect endogenous viruses encode *f-like* genes that could potentially be of cellular origin. Combined with the fact that they can insert into the baculovirus genome, insect endogenous viruses may represent a source for the indirect transfer of the *f-gene* (Fraser *et al.*, 1985; Malik *et al.*, 2000; Miller & Miller, 1982).

Regardless, *f* gene homologues are very likely to be present in many insects including not only the four above-mentioned dipteran insects but also lepidopteran insects. Given the fact that the similarity of the F proteins of baculoviruses is rather low (less that 20% amino acid identity in some cases) it is possible that cellular F proteins are also divergent and still remain to be identified in insect genomes.

Structural characterization of a baculovirus fusion protein ectodomain

Background

As mentioned in the introduction, group II alphabaculoviruses use protein F for fusion, whereas group I use GP64. Amino acid sequence analyses indicate that the F protein of baculoviruses displays a class I viral fusion protein fold and that it is related to the paramyxovirus fusion protein F. Therefore, it is interesting to understand the organization of F protein counterparts from DNA viruses such as the baculoviruses, in order to provide insight into their evolution.. Paramyxovirus F protein appears related to the spike protein of the coronaviruses with a large intervening domain between N-terminal and C-terminal HRs. These proteins are thus more distant from, and perhaps not true structural homologs to, other class I fusion proteins characterized to date such as those from retro-, filo-, arena- and influenza viruses. Importantly, there are no DNA viruses known to encode a class I fusion protein, and a crystal structure of SeMNPV F would provide important insight into evolutionary aspects relating class I viral fusion proteins from RNA and DNA viruses.

Objectives

The aim of our research was to obtain structural insight into the fusion process of the group II alphabaculovirus by determining the structure of the soluble F ectodomain of SeMNPV by X-ray crystallography. This project was initiated as a collaborative research project within the 7th framework program-Marie Curie Initial Training (Vecerek *et al.*) network Virus Entry between our laboratory and our partners from Utrecht University, Netherlands.

Results

Production and purification of SeFe

Together with our partners from Utrecht University, we established an efficient expression system for to obtain the soluble ectodomain of the SeMNPV F protein, called SeFe in *Drosophila S2* cells. The construct containing the SeFe encompasses aa residues A18-T553 (aa 1-17 comprise a signal peptide) of the full-length SeMNPV F protein (UniProtKB accession number Q9J8C6). It lacks the predicted transmembrane (TM) anchor domain (residues 580-602) and the C-terminal cytoplasmic tail (CT) domain (residues 603-665) present in the full-length protein (**Figure 64**) in order to allow secretion from cells. Our first approach was to crystallize SeFe in its pre-fusion form. To stabilize the pre-fusion form, the furin cleavage site was mutated (RRSKR to SGSKK) to prevent proteolytic processing into F1 and F2 subunits by furin in the transfected S2 cells.

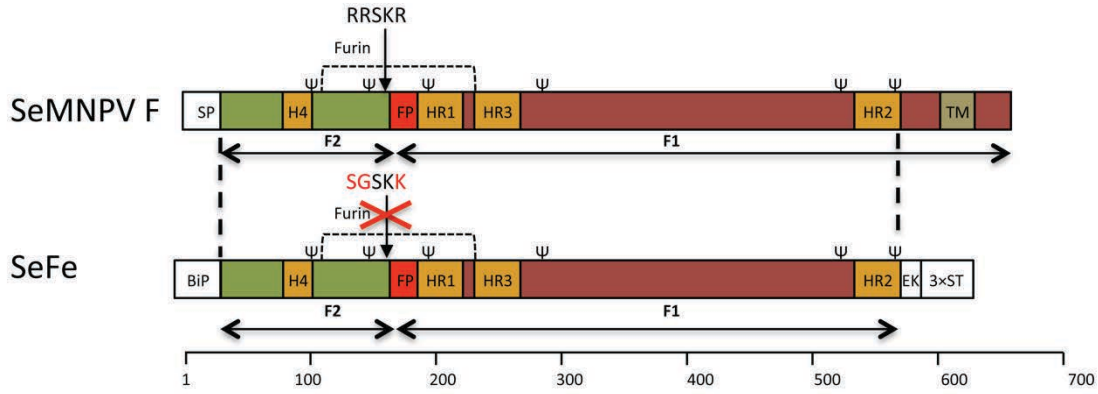


Figure 64. Primary structure of *Se*MNPV F and the expression system of its ectodomain (SeFe) in *Drosophila melanogaster* S2 cells. The domains indicated are F2 and F1, the fusion peptide (FP), the signal peptide (SP), three heptad repeat regions (HR1, HR2, HR3 and H4), the transmembrane domain (TM), enterokinase cleavage site (EK), triple strep-tag (3×ST) and *Drosophila* secretion signal (BiP). Predicted *N*-glycosylation sites are marked by (Ψ), furin cleavage site in SEMNPV F (RRSKR) and the mutated furin cleavage site in SeFe (SGSKK) are indicated by black arrows. The disulfide bridge connecting domains F2 and F1 is shown as a thin dashed line. The thick dashed line indicates the borders of the F ectodomain cloned into the expression vector. The scale below corresponds to amino acid numbering of SeMNPV F (UniProtKB accession number Q9J8C6).

The large-scale expression and purification of SeFe was performed using standard procedures (see section 2). Isolation of pure SeFe employed a combination of Streptactin affinity column (**Figure 65A**) and size exclusion chromatography using a HiLoad 26/60 Superdex 200 gel filtration column (GE Healthcare) (**Figure 65B**). The protein eluted from

the gel filtration column as a single peak corresponding to monomeric SeFe (**Figure 65B**). The final yield of SeFe was ~10 mg per litre of culture supernatant.

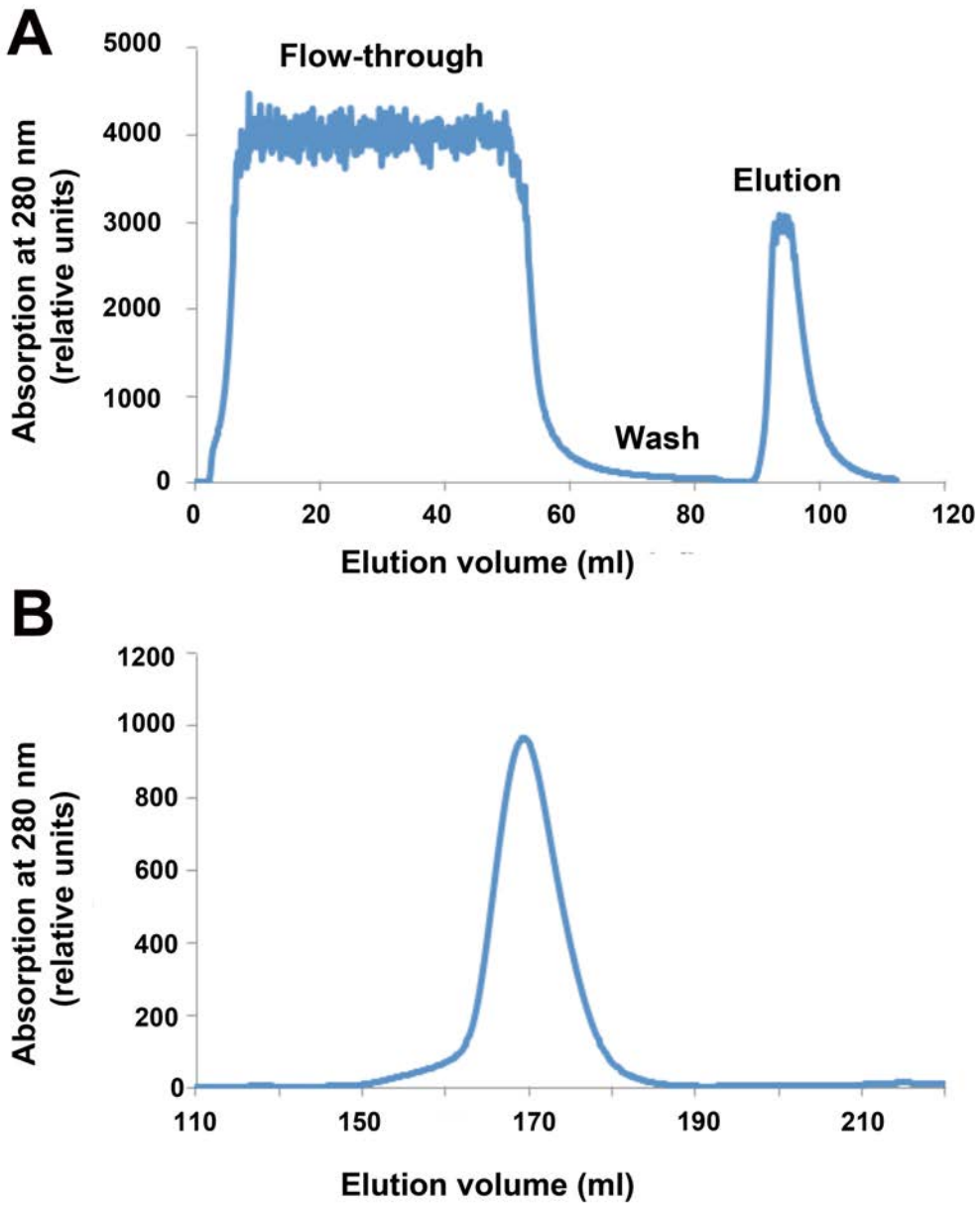


Figure 65. Purification of SeFe. The concentrated supernatant was loaded on 8 ml Streptactin column (A). After washing step, the Strep-tagged protein was eluted with 2.5 mM desthiobiotin. The fractions of the eluent were pooled and subjected to size exclusion chromatography (SEC) (B). Separation by SEC was performed using HiLoad Superdex 200 26/60 column (GE Healthcare) at a flow speed of 2 ml/min in 5 mM Hepes pH 7.0 150 mM NaCl and 0.1 mM EDTA. Protein elution was monitored by absorbance at 280nm (blue curves). Chromatogram B reveals a single major peak for SeFe corresponding to the monomeric protein.

Initially Scott Jeffers in our laboratory crystallized monomeric SeFe at neutral pH and obtained crystals (most likely corresponding to the pre-fusion form of SeFe) but the crystals diffracted only to ~8 Å, which was not sufficient to determine the crystal structure. The

crystallized protein was monomeric as judged by size exclusion chromatography, suggesting that it may represent a pre-fusion conformation. Crystallizing viral fusion proteins in their pre-fusion form is usually challenging because they are metastable. Based on the fact that the paramyxovirus F protein is a stable homotrimer in its post-fusion conformation (Swanson *et al.*, 2011), we decided to attempt crystallization of the SeFe in its post-fusion form as an alternative strategy.

Trypsin proteolysis of SeFe yields a trypsin-resistant fragment

The main problem in crystallizing the post-fusion conformation of viral fusion proteins is that the fusion peptide is exposed, usually resulting in aggregation of the protein. Our collaborators from Utrecht University established a protocol to obtain a trypsin resistant fragment of SeFe. Furthermore, they demonstrated that lowering the pH from pH7 to pH5 resulted in an altered oligomeric state of the trypsin resistant fragment, suggesting the formation of a post-fusion trimer. N-terminal sequencing results, together with SDS-PAGE and Western blotting analysis of the trypsin-resistant SeFe trimer, revealed that:

1. Trypsin cleavage occurs in the region between the predicted HR1 and HR3 regions at residues R210 and R212 (210-RéMRéD-213) (based on N-terminal sequencing).
2. Trypsin also cleaves at the mutated furin cleavage site at residues K147 or K148 (SGSKéKé) (based on SDS-PAGE).
3. The Strep-tag can no longer be detected by anti-strep antibodies in the Western blot, thus it is completely removed by trypsin cleavage.
4. The trypsin cleavage at the C-terminal end of the F1 subunit most likely occurs at one of the trypsin target sites located at either end of the HR2 domain.

Although trypsin cleavage results in the removal of the internal protein part composed of the fusion peptide and the HR1 domain, F1 and F2 still remain linked by the disulphide bridge (**Figure 66**).

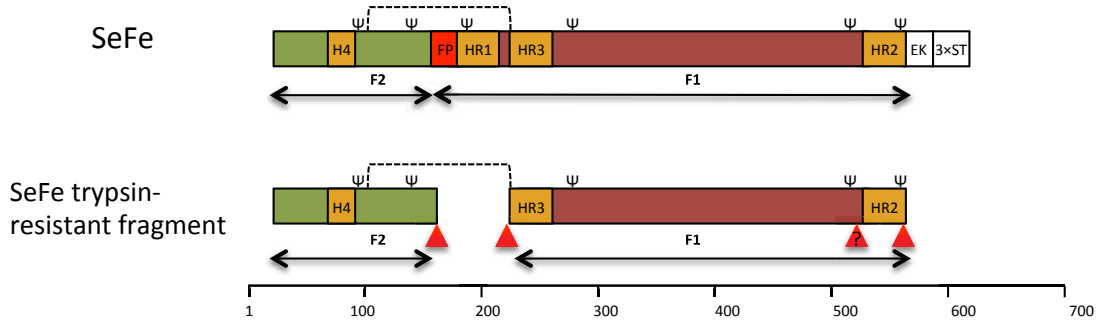


Figure 66. Schematic representation of SeFe and its trypsin-resistant fragment. The trypsin cleavage sites are indicated with red arrowheads. The domains indicated are F2 and F1, the fusion peptide (FP), three heptad repeat regions (HR1, HR2, HR3 and H4), enterokinase cleavage site (EK), and triple strep-tag (3×ST). Predicted *N*-glycosylation sites are marked by (Ψ). The disulfide bridge connecting domains F2 and F1 is shown as a thin dashed line.

I undertook to reproduce the trypsin cleavage of SeFe in our laboratory. The limited trypsin proteolysis with increasing concentrations of trypsin led to the formation of a predominant product, which was observed as a single band of ~50 kD in SDS-PAGE under non-reducing conditions (**Figure 67**).

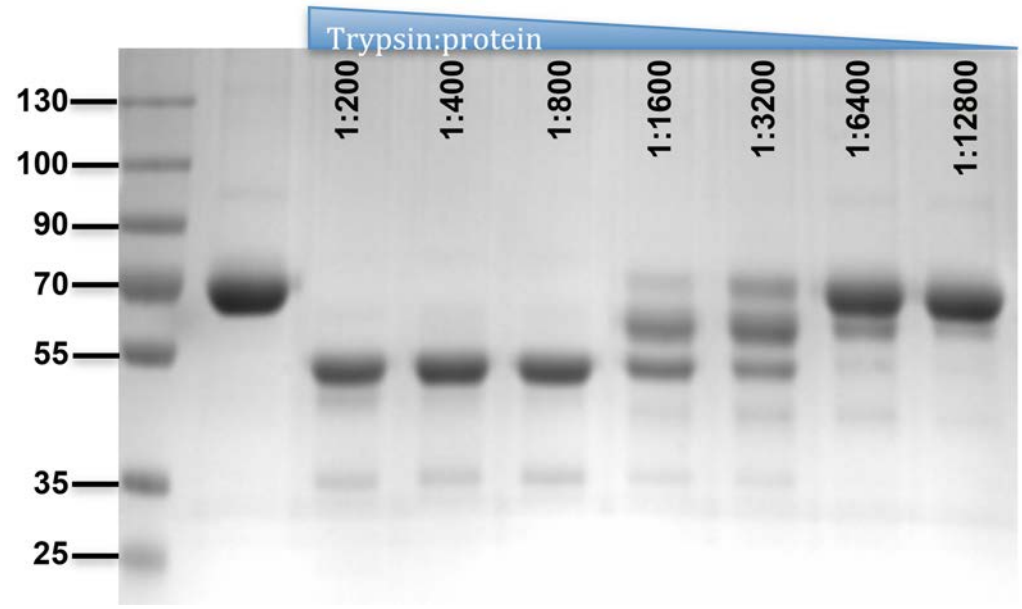


Figure 67. Limited proteolysis of SeFe by trypsin. SeFe was mixed with trypsin at different ratios, incubated at 22°C for 30 min and analyzed by SDS-PAGE under non-reducing conditions. The gel was stained with Coomassie blue. A predominant product of ~50 kD was observed by SDS-PAGE when using trypsin:SeFe ratio from 1:800 to 1:200. Lane 1: protein ladder.

On the intact virions, furin-cleaved baculovirus F proteins mediate membrane fusion upon exposure to acidic pH encountered during endocytosis(WF *et al.*, 2000). Therefore, I investigated the effect of acid treatment of the SeFe and the SeFe trypsin-resistant fragment.

Both proteins were exposed to pH5.5 or pH7 and subsequently analyzed by SEC and multi-angle laser light scattering (MALLS). The SeFe trypsin resistant fragment at pH5.5 eluted earlier from the SEC column than the SeFe trypsin resistant fragment at pH7, suggesting the change in its oligomerization at acidic pH (**Figure 68**).

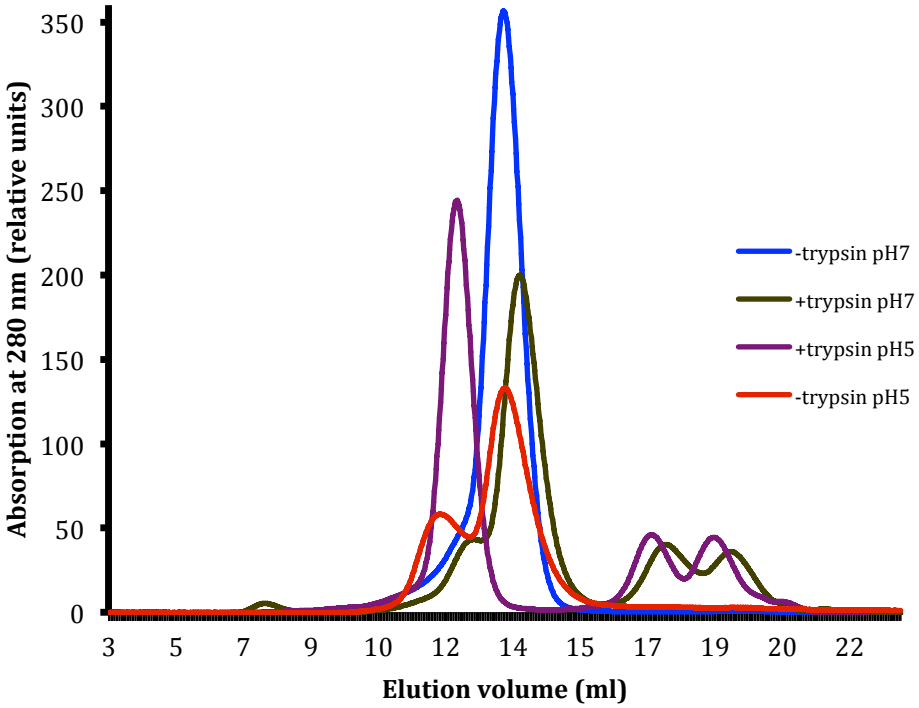


Figure 68. Effect of acid treatment on SeFe and SeFe trypsin resistant fragment. Non-digested or trypsin digested SeFe was exposed to neutral (pH7) or acidic (pH5.5) pH and analyzed by SEC using Superdex 200 10/300 column (GE Healthcare). SeFe trypsin resistant fragment at pH5.5 eluted earlier from the SEC column than the SeFe trypsin resistant fragment at pH7, suggesting that there was a change in oligomerization at acidic pH.

The analysis of the oligomeric state of the SeFe and SeFe trypsin resistant fragment by MALLS confirmed that the initially crystallized protein (SeFe, pH7) was a monomer while the trypsin resistant fragment forms a trimer at acidic pH (**Table 9**). The acid-induced trimerization of the trypsin resistant SeFe fragment was irreversible since the titration of pH back to neutral prior to MALLS analysis did not lead to trimer dissociation.

Table 9. MALLS analysis of the oligomeric state of SeFe and SeFe trypsin resistant fragment at neutral (pH7) or acidic (pH5.5) pH.

	Molecular weight [kD]	**Mass fraction (%)**	**Oligomeric state**
- Trypsin pH7	95.4 (±5.5%)	100	Monomer
- Trypsin pH5	237.4 (±0.1%) and 91.3 (±0.2%)	36.4 and 63.6	Trimer and monomer
+ Trypsin pH7	77.5(±0.2%)	100	Monomer
+ Trypsin pH5	188.0 (±1.9%)	100	Trimer

We decided to use this protease resistant trimer for crystallization as it most likely represents the stable post-fusion conformation. I made a large-scale preparation of the trimeric trypsin-resistant fragment of SeFe (called SeFet) and purified it by SEC on a Superdex 200 26/60 column, from which the protein eluted in a single symmetric peak corresponding to a SeFet trimer, as expected **(Figure 69).**

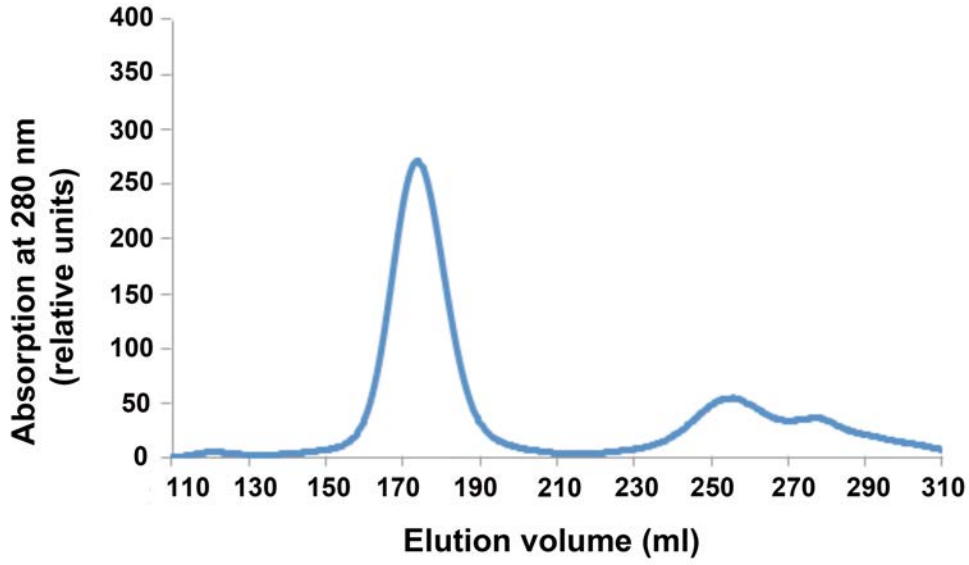


Figure 69. Large scale SeFet formation. The trypsin proteolysis reaction of SeFe and acid induced trimerization were upscaled in a linear manner. SeFet was purified from the reaction mixture by SEC on Superdex 200 26/60 column (GE Healthcare). SeFet eluted from SEC column as a single symmetric peak corresponding to the trimeric SeFe trypsin resistant fragment.

Crystallization of SeFet

The crystallization screening for SeFet was performed as described in Materials and Methods at protein concentration 6 and 10 mg/ml. The protein crystallized in a number of

conditions that were further optimized in 24-well plates. The crystals that were subjected to X-ray diffraction analysis are listed in **Table 10**. Most of the crystals diffracted to 10-12 Å or lower resolution, with the exception of one crystal form that diffracted to 6.5 Å, which was sufficient for determination of the space group ($P2_12_12_1$) and unit cell parameters (a=109.68, b=346.24, c=111.34). In addition, we observed a peak in self-rotation function at 120 degrees indicating that the crystallized molecule is a trimer. According to the Matthew's coefficient, the unit cell probably accommodated three trimers per asymmetric unit. These crystals of SeFet formed after three weeks in 2µl hanging drops by vapor diffusion against a reservoir solution containing 18% PEG 6000 and 100 mM Tris pH8 (1:1 protein-to-reservoir-solution ratio) (**Figure 70**) and reproducibly diffracted to 6.5-7 Å resolution.

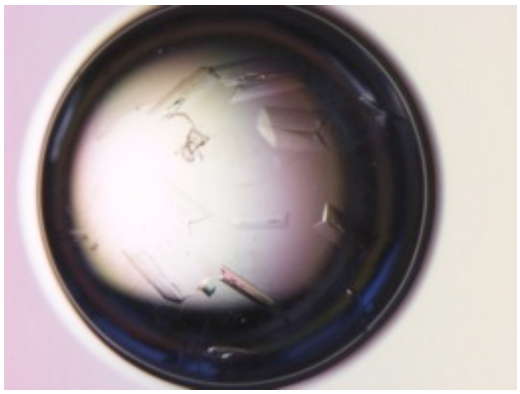


Figure 70. Best diffracting SeFet crystals. These plate-shaped crystals of SeFet were formed after three weeks in 2µl hanging drops by vapor diffusion against reservoir solution containing 18% PEG 6000 and 100 mM Tris pH8 (1:1 protein-to-reservoir-solution ratio) and reproducibly diffracted to 6.5-7 Å resolution.

In order to obtain better diffracting crystals, I performed a number of optimization experiments (microseeding and streak seeding, crystallization at different protein concentrations, testing crystallization additives, crystallization at 4˚C, crystallization of the protein after an extra purification step by ion-exchange chromatography and vapor diffusion dehydration). By vapor diffusion dehydration of the crystals for 2 days above the reservoir solution containing 23% PEG 6000, 100 mM Tris pH8, I was able to improve resolution and collect a complete dataset to ~5.5 Å.

Table 10. The crystals of SeFet subjected to X-ray diffraction experiments. The ability of the crystals to diffract X-rays were tested at Proxima I line, Synchrotron Soleil, France or ID 23 line European Synchrotron Radiation Facility (ESRF), France.

Crystallization condition	**Concentration (mg/ml)**	**Resolution (Å)**
18% PEG 6000, 100 mM Tris pH 8	6	6.5
16% PEG 4000, 100 mM Tris pH 7.5, 15% 2-propanol	6	12
18% PEG 6000, 100 mM Tris pH 8	6	8
14% PEG 8000, 100 mM Tris pH 7.5	6	7
23% PEG 6000, 100 mM Tris pH 7.5	6	9
12% PEG 6000, 10 mM NaAc, 40% ethanol	6	13
20% PEG 6000, 100 mM imidazole pH 8	6	10
26.5 % PEG 8000, 100 mM Hepes pH 7.5, 10% 2-propanol, 100 mM NaAc	6	20
20.5% PEG 10000, 100 mM Tris pH7.5	6	10
19% PEG 4000, 100 mM NaAc, 100 mM Hepes pH7.5	6	20
15% PEG 8000, 40 mM potassium phosphate monobasic	6	>30
16% PEG 6000, 100 mM NaCl, 100 mM Tris pH 7.5	6	25
19% PEG 6000, 200 mM NaCl, 100 mM Tris pH 8	6	9
19% PEG 6000, 100 mM NaCl, 100 mM Tris pH 8	6	10
16% PEG 8000, 170 mM NaAc, 90 mM sodium cacodylate pH 6.5	6	15
16% PEG 8000, 170 mM NaAc	6	9
14% PEG 8000, 100 mM Tris pH 7.5	6	
17.4% PEG 3350, 200 mM NaCl, 200 mM $MgCl_2$	6	>30

Crystallization of the deglycosylated SeFet

SeFe contains six predicted N-linked glycosylation sites (N86, N132, N179, N270, N508 and N551) (**Figure 71**). One of the advantages of the expression of the recombinant proteins for crystallization in S2 cells is that N-linked glycosylation in *Drosophila* is less complex and more homogenous than in mammalian cells. Proteins expressed in S2 cells have high-mannose N-linked glycosylation and are not sialylated. In general, a deglycosylated protein might form a more rigid protein lattice than a protein containing all sugar chains and, thus, yield better diffracting crystals. Although one or two sugars are removed in trypsin resistant fragments because they are present in the parts of the protein cleaved off by trypsin, the rest of the sugars might still hinder the formation of a rigid protein lattice. As a result, I attempted to deglycosylate SeFet with different endoglysosidases for crystallization trials.

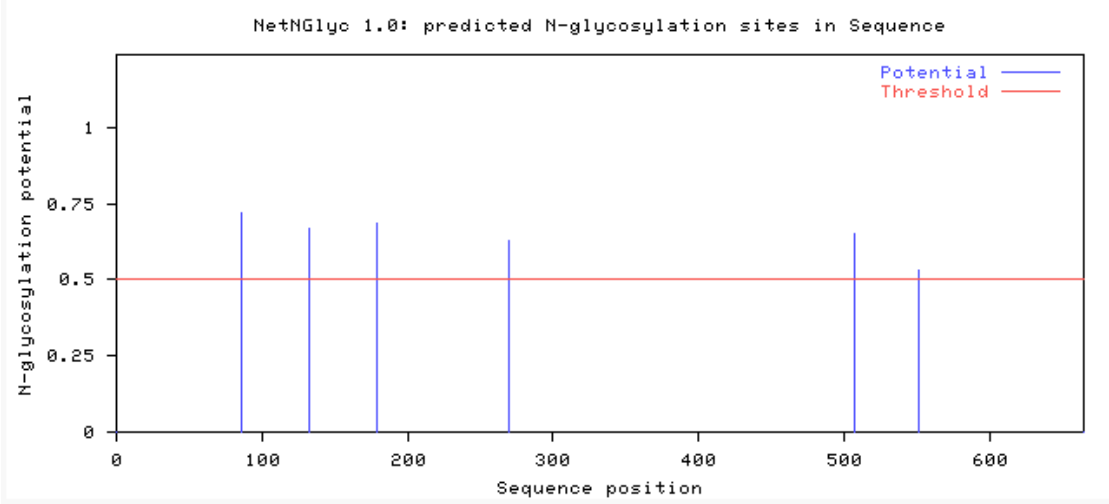


Figure 71. Putative glycosylation sites in SeMNP F protein predicted by NetN Glyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc). The graph illustrates predicted N-glycosylation sites across the protein chain. Positions with potential (vertical lines) crossing the threshold (horizontal lines at 0.5) are predicted glycosylated.

The deglycosylases PNGase F, EndoH and EndoD were produced using periplasmic expression in *E.coli* and purified as described in Materials and Methods. The extent of deglycosylation of SeFet by different endodeglycosidases was assessed by mobility shift of the deglycosylated protein versus the intact glycoprotein on SDS-PAGE gels. A clear mobility shift of SeFet deglycosylated by PNGase F (**Figure 72**) as well as EndoD/EndoH (not shown) was observed. Deglycosylation of SeFet by EndoH alone (not shown) did not have any evident effect on protein mobility on the SDS-PAGE gel, indicating that this endodeglycosidase most likely is not able to remove the sugars from the protein.

Subsequently, I performed a large-scale preparation of SeFet deglycosylated by PNGase F. After deglycosylation, SeFet was separated from PNGase F by SEC (**Figure 73A).** In addition, a large-scale preparation of SeFet deglycosylated by EndoD/EndoH was carried out. EndoD/EndoH were removed from the reaction mixture by Ni-affinity chromatography and subsequent SEC (**Figure 73B**). The crystallization screening of SeFet deglycosylated with PNGase F and EndoD/EndoH was performed as described in Materials and Methods at protein concentrations of 4 and 6 mg/ml, respectively. Since the endodeglycosidase EndoD was obtained later during the project, the crystallization of SeFet deglycosylated with EndoD/EndoH will be described later.

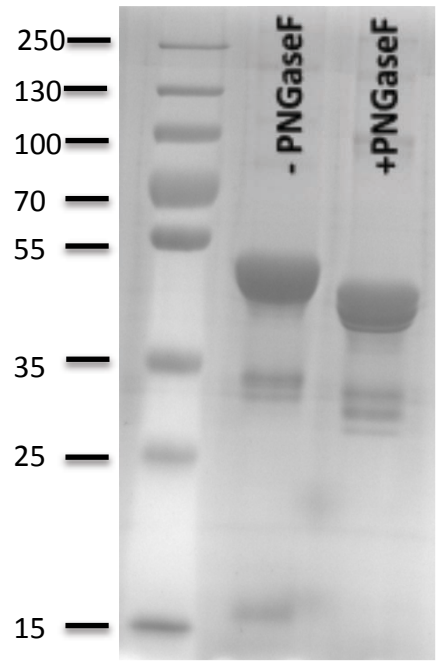


Figure 72. Enzymatic deglycosylation of SeFet with PNGase F. SeFet was incubated with PNGase F overnight at 37°C using the ratio 1:16 of endoglycosidase:SeFet on a weight basis. A clear mobility shift of SeFet deglycosylated by PNGase F versus non-deglycosylated SeFet was observed. Lane 1: protein ladder, lane 2: non-deglycosylated SeFet, lane 3: SeFet deglycosylated with PNGase F.

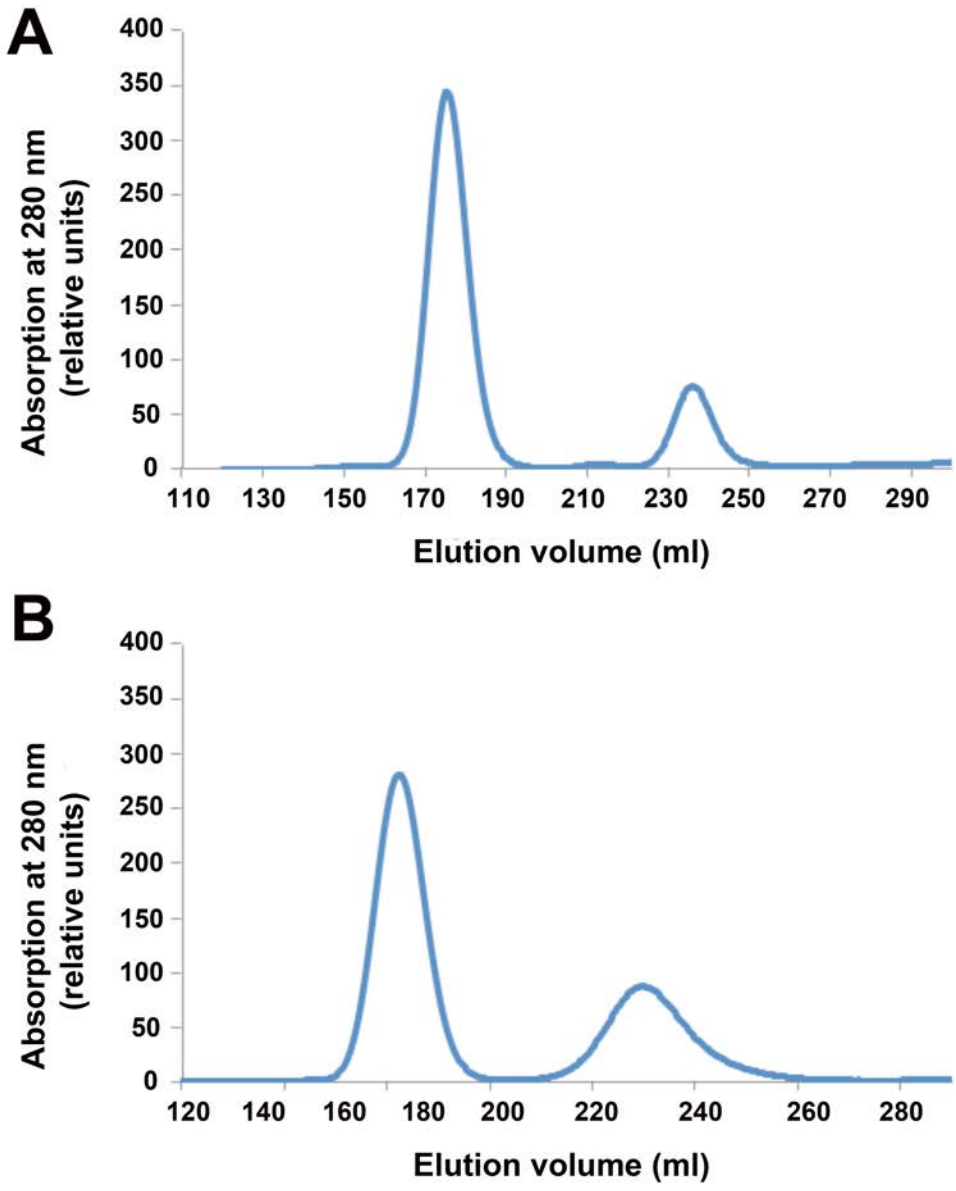


Figure 73. Purification of deglycosylated SeFet. After deglycosylation with PNGase F, SeFet was separated from the endodeglycosidase by SEC (A). The first peak in the chromatogram corresponds to the deglycosylated SeFet, while the second peak corresponds to PNGase F. After deglycosylation with EndoD/EndoH, SeFet was separated from endodeglycosidases by Ni-affinity chromatography and subsequent SEC (B). The first peak in the chromatogram corresponds to the deglycosylated SeFet, while the second peak corresponds to the fraction of EndoH that was not fully removed by Ni-affinity chromatography.

SeFet deglycosylated with PNGase F crystallized under fewer conditions, which and were different from those of the crystals of non-deglycosylated protein. The crystallization conditions were further optimized in 24-well plates and the best crystals subjected to X-ray diffraction. However, none of the crystals of deglycosylated protein diffracted to higher than 8Å resolution. In addition, the deglycosylated protein did not crystallize any more in the condition in which the best diffracting crystals of the sugar-containing SeFet were obtained

(16% PEG 6000, 100 mM Tris pH8). Therefore, I tried to induce the crystal growth of deglycosylated protein under these conditions by seeding with the crystals of the glycosylated protein as a seed source. Although seeding helped to obtain the crystals of deglycosylated SeFet under the same conditions of the not-deglycosylated protein, the resulting crystals again only diffracted to about 8Å resolution (**Table 11**).

Table 11. Crystals of the deglycosylated SeFet subjected to X-ray diffraction at Proxima I line, Synchrotron Soleil, France or PXI line, Swiss Light Source (SLS), Switzerland.

Crystallization condition	Concentration (mg/ml)	Resolution
21 % PEG 3350	4	30
19 % PEG 3350, 200 mM ammonium formate	3	3.8
19 % PEG 3350, 200 mM sodium formate	3	5
18 % PEG 3350, 200 mM ammonium chloride	2	2.7
18 % PEG 6000, 100 mM Tris pH8	4	8
20 % PEG 2000 mono-methyl polyethylene glycol	4	10
22 % PEG 3350, 200 mM sodium nitrate	4	10
19 % PEG 3350, 200 mM ammonium fluoride	4	6
9% PEG 4000, 100 mM imidazole pH 8 and 30% MPD	4	10

To obtain different crystals of deglycosylated SeFet, I used microseed matrix screening to sample many more crystallization conditions. This is a seeding technique where crystals grown in one condition are ground to use as seeds, and are then distributed by a robotized procedure into hundreds of different crystallization conditions. The crystals of deglycosylated SeFet grown in 9% PEG 4000, 100 mM imidazole pH 8 and 30% 2-methyl-2,4-pentanediol (MPD) were used as a seed stock for microseed matrix screening. This technique helped to obtain crystals in different crystallization conditions that were further optimized in 24-well plates and subjected to X-ray diffraction analysis.

For one crystal form (**Figure 74**) (which grew in 18% PEG 3350, 200 mM ammonium chloride) we were able to collect a complete native data set at 2.7 Å (space group P1) (). The unit cell parameters (a=73.94, b=74.86, c=78.11, α=94.5, β=114.3, γ=114.2) could accommodate one trimer per asymmetric unit. The Matthews' coefficient (V_m) was estimated to be 2.04 $Å^3$/Da (corresponding to a solvent content of 39.69%). The self-rotation function (**Figure 76**) clearly showed the existence of a 3-fold non-crystallographic symmetry (NCS) axis. This crystal form was reproducible and grew in ~48 h.

The quality and morphology of the crystals were very tightly dependent on the seed stock dilution, protein concentration, and PEG 3350 concentration used. Native crystals reproducibly diffracted to ~3 Å. All tested cryoprotectants (PEG 400, glycerol, MPD and ethylene glycol) were suitable for cryopreservation of the crystals. Screening for

crystallization additives was performed, expecting to further improve the diffraction of the crystals but the identified additives did not enhance the diffraction of the crystals. It was possible to grow crystals using pH gradient from 6-8. However, pH did not have significant influence on crystal morphology or diffraction. Crystals growing at 4°C also did not help to improve the diffraction either.

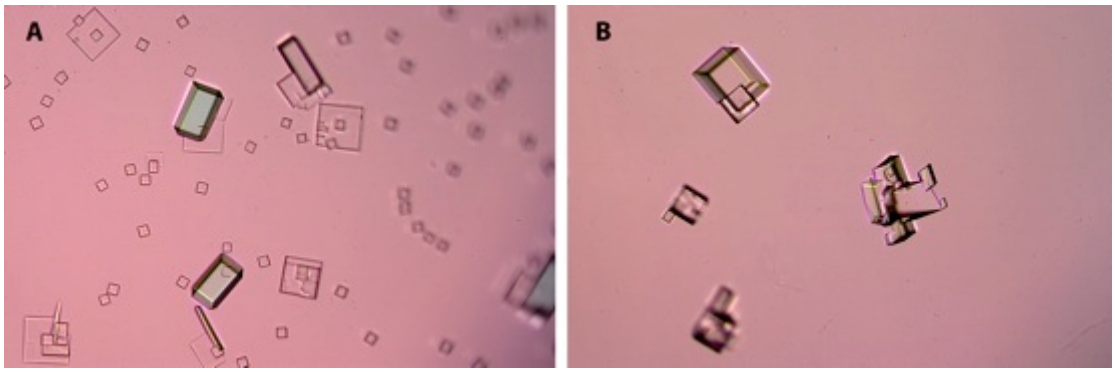


Figure 74. Photographs of typical SeFet crystals obtained after seeding. (A and B). The crystals were grown at 2 mg/ml SeFet in 2.5 µl hanging drops (1:1:0.5 protein:reservoir solution:seed stock ratio) by vapor diffusion against reservoir solution containing 16-18% PEG 3350 and 200 mM ammonium chloride and diffracted to ~3 Å resolution. The crystallization drops usually used to contain different quality crystals. The crystals had a tendency to grow as clusters of multiple crystals (B).

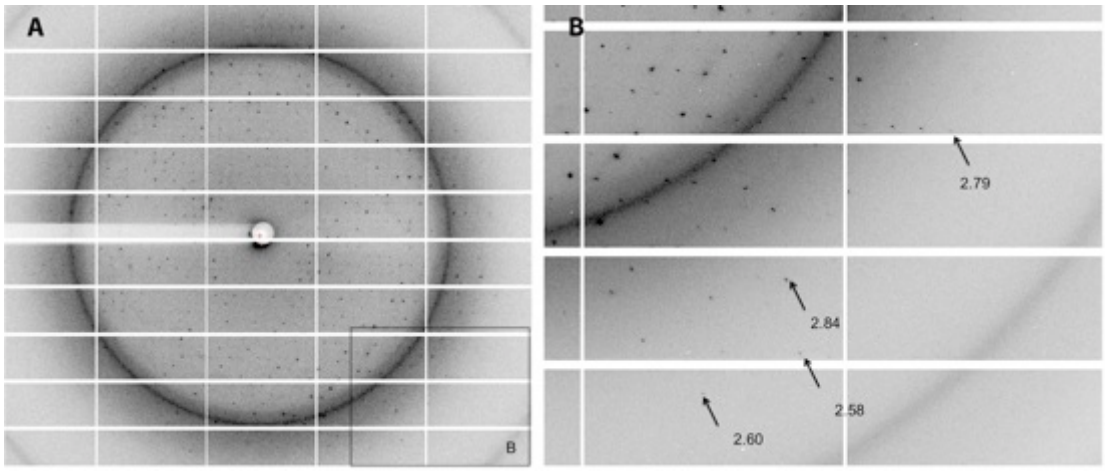


Figure 75. An X-ray diffraction pattern from the best native deglycosylated SeFet crystal on a Pilatus detector (Dectris, Baden, Switzerland). (A) High resolution reflections of the same diffraction image and corresponding resolution (B). The data were collected on the PXI beam line at the SLS in Switzerland.

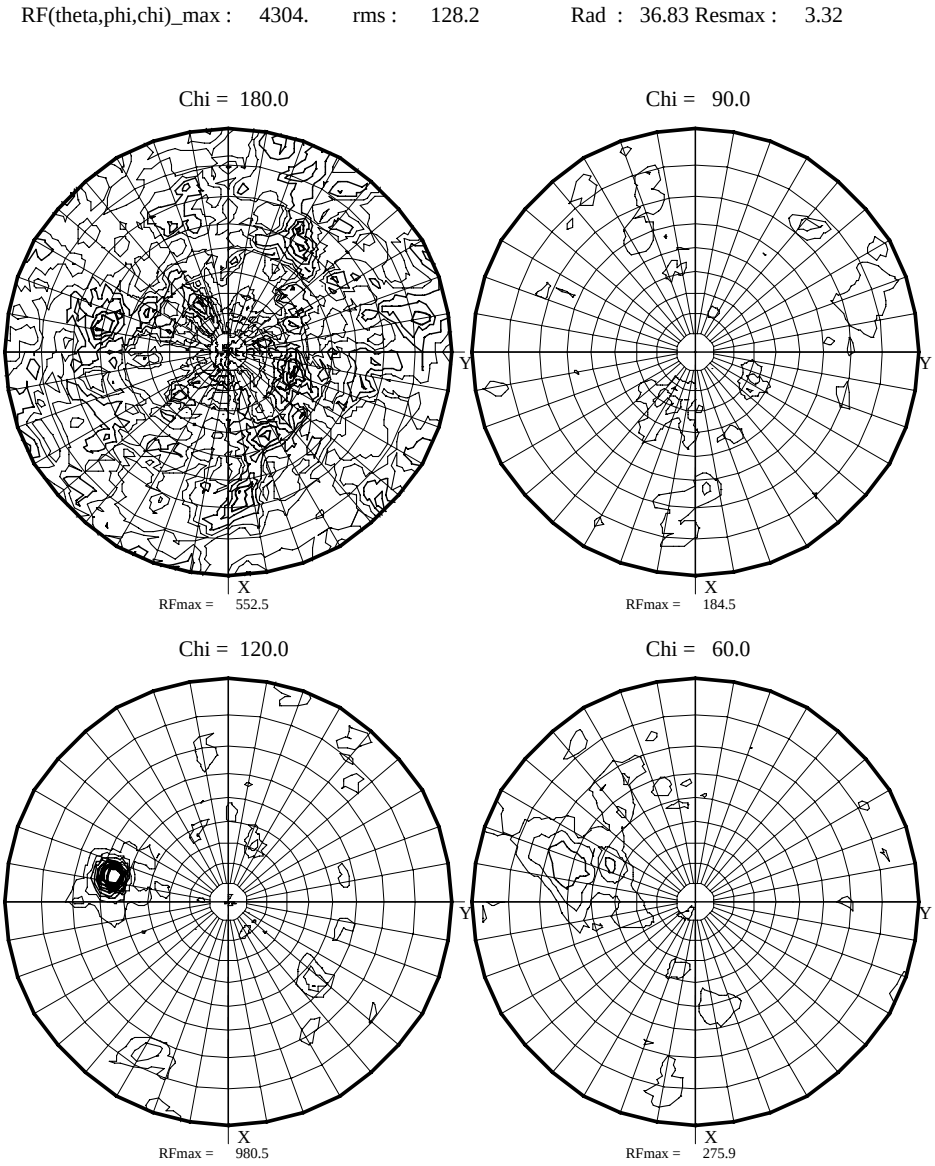


Figure 76. Self-rotation function of SeFet P1 crystal form as determined from the program MOLREP (Vagin & Teplyakov, 2010). The peak on the sections (chi = 120) indicates the presence of non-crystallographic 3-fold axis.

Phasing of SeFet crystals

In order to obtain phase information, we resorted to experimental phasing given that the paramyxovirus F model was not sufficient to obtain accurate phases by molecular replacement. I searched for heavy atom derivatives of the crystals by soaking them in various heavy atom solutions. Derivatization depends on parameters such as the exposure of functional groups, local chemical environment, and ionization state. SeFe contains an odd number of cysteines so the first choice was soaking the crystals with different mercury compounds that are known to specifically react with free cysteine thiols. In addition, I tried to

co-crystallize SeFet in the presence of the same mercury compounds. Unfortunately, neither of the two approaches proved to be successful. I tested the availability of free cysteine by a colorimetric assay using 5,5′-Dithiobis(2-nitrobenzoic acid) (DTNB) reagent (described in section 2). A lower signal than expected for three free thiols indicated that the cysteines in the SeFet are difficult to access, which could explain the failure to derivatize SeFet crystals with mercury compounds. In parallel, I evaluated a number of other heavy atom compounds, using at least two different concentrations for screening, and also testing different soak times. The heavy atom compounds tried in the derivatization of SeFet crystals are listed in **Table 12**.

Table 12. The heavy atom compounds tried in the derivatization of SeFet crystals.

N^o	**Heavy atom compound**	**Anomalous signal**
1.	Thimerosal	No
2.	Ethyl Mercuric Phosphate	No
3.	Methylmercury (II) chloride	No
4.	Mercury (II) chloride	No
5.	Mercury (II) cyanide	No
6.	Mercury (II) acetate	No
7.	Potassium tetracyanoplatinate (II) hydrate	No
8.	Potassium tetrachloroplatinate (II)	Yes
9.	Dipotassium hexachlororhenate	maybe weak
10.	Lead(II) acetate trihydrate10	No
11.	Sodium tetrachloraurate (III) dihydrate	Yes
12.	Gold (I) potassium cyanide	No
13.	Potassium tetrachloraurate (III) hydrate	No
14.	Neodymium trichloride, hexahydrate	no
15.	Sammarium (III) acetate	maybe weak
16.	Hexatantalum tetradecabromide	yes

Initially, we succeeded in obtaining heavy atom derivatives with potassium tetrachloroplatinate (II) ($K_2PtCl_{4)}$ and sodium tetrachloraurate (III) ($NaAuCl_4$), but in both cases the derivatives diffracted only to low resolution (~5-6 Å) and the anomalous signal was not strong enough to obtain an initial set of phases. In addition, the derivatized crystals were highly non-isomorphous with the native crystals, excluding the multiple isomorphous replacement method as a possible phasing strategy. Furthermore, the fast decay and low symmetry made it difficult to determine accurate phases experimentally. The data sets were often not complete due to radiation damage, as data collection required a long period of time to obtain enough redundancy in this low symmetry P1 space group (at least 360° oscillation in inversed-beam collection mode). As a result, the low multiplicity of the measured intensities (Friedel pairs are measured only once if collected 360°) resulted in poor measurement precision. We therefore initiated a collaboration with Pierre Legrand from the synchrotron Soleil who is very experienced in experimental phasing.

First, we optimized the data collection strategy by attenuating the beam to reduce radiation damage, which allowed us to use the Multiple Anomalous Diffraction (MAD) method by collecting data in small wedges in the inverse beam mode while alternating between several wavelengths. We also collected a single-wavelength anomalous diffraction (Drexler *et al.*) (Drexler *et al.*) data set at 3.1 Å at the sulphur edge on native crystals, which allowed us to identify the locations of the disulfide bonds. But even after optimization of the data collection strategy, the anomalous signal of the crystals derivatized with K_2PtCl_4 or $NaAuCl_4$ were not strong enough to obtain an initial set of phases good enough to determine the structure.

Next we tried using the hexatantalum tetradecabromide ($Ta_6Br_{12}^{+}$) cluster compound, which is known to be a powerful derivatization reagent for phasing crystals at low resolution. The soaking of SeFet crystals in 1mM $Ta_6Br_{12}^{+}$ over night led to the incorporation of the compound into the crystals, which was apparent due to the change of the crystal color. We applied the double-inflection MAD collection strategy using one crystal derivatized with $Ta_6Br_{12}^{+}$ and detected significant anomalous differences up to ~6 Å. This data set was used to locate the heavy atom sites in the unit cell of the crystal, and to calculate an initial set of phases, which resulted in an interpretable electron-density map with clearly identifiable helices resembling six-helix bundle characteristic for class I fusion proteins (**Figure 77**).

We then applied density modification techniques to further improve the map and started manual-building of the initial model as polyalanine chains within this map, along with iterative cycles of electron density modification (solvent flattening, non-crystallographic symmetry (NCS) averaging). The initial maps indicated that the SeFet molecule is a very flexible trimer, with the angles between three protomers varying along the trimer axis. As a result, NCS averaging of the SeFet crystals required defining several masks and NCS operators for different parts of the molecule. In addition, the low solvent content of the crystals was not beneficial for solvent-flattening density modification, which is the more powerful the larger the solvent content. The phases were transferred to the native SeFe crystal (P1) by molecular replacement, which resulted in an electron density map at ~3 Å resolution and the model further improved. However, the quality of the map was still not good enough for building a full model of SeFet.

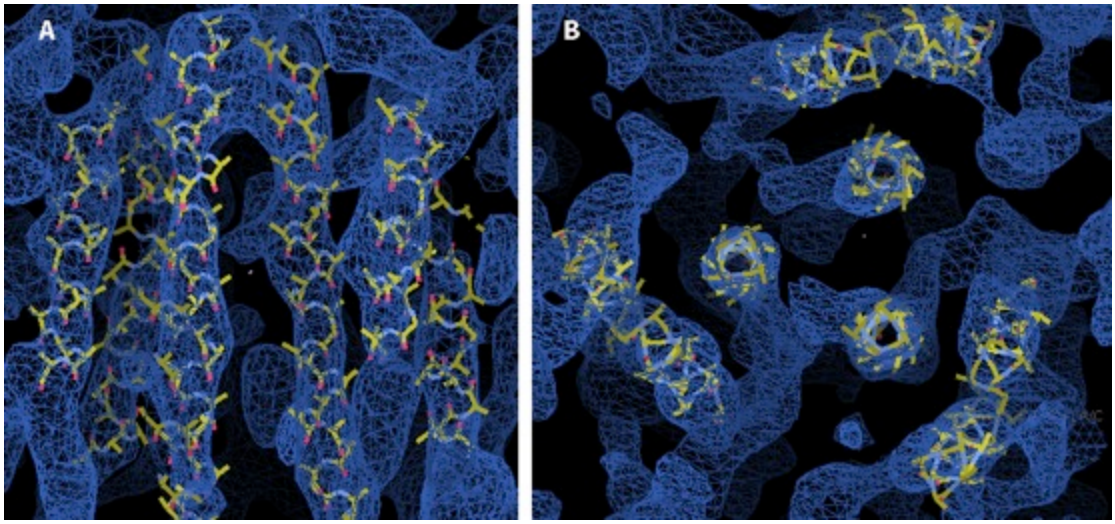


Figure 77. The electron density map of SeFet with the first helices built in alanine residues. The side view of SeFet helices (A). The view along the trimer axis (B).

The main breakthrough in the experimental phasing of the SeFet crystals was obtaining a higher symmetry crystal form (space group $P4_32_12$) with cell parameters (a=80.3, b=80.3, c=478.7, α=90, β=90, γ=90) that could accommodate one trimer of SeFe per asymmetric unit. This crystal form grew under identical conditions as the P1 crystal form. However, instead of the seed stock solution containing the seeds, the same solution (9% PEG 4000, 100 mM imidazole pH 8 and 30% MPD) without the seeds was used to set up the crystallization drops. Those crystallization drops were intended to be used for streak seeding, but self-nucleated crystals appeared after ~12 h (**Figure 78**). Though these crystals were difficult to reproduce, and approximately only one out of 20 crystals diffracted, we managed to collect a complete native data set at 2.9 Å as well as double-inflection MAD data on a derivative with $Ta_6Br_{12}^{+}$ that diffracted to ~3.6 Å.

We were able to resolve the individual tantalum atoms in the cluster (using double-inflection MAD data for a crystal form $P4_32_12$), which resulted in an improved anomalous signal in comparison with the anomalous signal of the cluster as a super-atom. We also applied multi-crystal averaging to further improve the experimental electron density map. The multi-crystal averaging was complicated due to difficulties finding the right symmetry operators between SeFet molecules in different crystal forms. Nevertheless, performing a few multi-crystal averaging cycles resulted in a better-quality map. The experimental electron density map calculated from this MAD dataset was further used for model building. The phases were gradually extended to 2.7 Å (the best native data set in P1 crystal form).

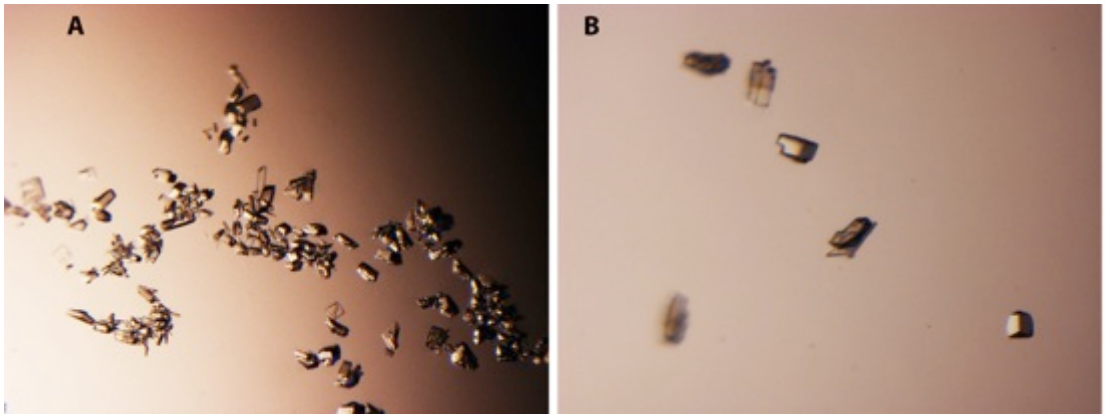


Figure 78. Deglycosylated SeFet crystals of $P4_32_12$ space group (A and B). Self-nucleated crystals appeared after ~12 h in 2.5 µl hanging drops (1:1:0.5 protein:reservoir solution:seed stock solution ratio) by vapor diffusion against reservoir solution containing 16-18% PEG 3350 and 200 mM ammonium chloride. Instead of seed stock solution containing the seeds the same solution (9% PEG 4000, 100 mM imidazole pH 8 and 30% MPD) without the seeds was used to set up the crystallization drops.

I also introduced an anomalous scatterer into the SeFet cell culture (i.e. I incorporated selenomethionine (SeMet) replacing methionine within the native protein). The incorporation of SeMet into proteins expressed in insect cells is difficult in comparison to proteins produced in *E. coli*, where the incorporation of SeMet is usually 100 %. Moreover, the yields of SeMet labeled protein are usually lower than those of a native protein due to SeMet toxicity to the cells. Although the expression levels of SeMet-labeled SeFet dropped significantly in comparison with the native protein (0.150 mg of labeled SeFet from 1 L of cell culture supernatant), it was enough to obtain crystals. One of these crystals diffracted to 3.2 Å and was used to collect a SAD data set. SeMet incorporated in SeFe gave a good anomalous signal with usable SAD phases to about 4 Å. In total, approximately 30 SeMet sites were identified (36 methionines in the SeFet trimer) meaning that SeMet incorporation into SeFe was more than 80%. This data set was used to identify the position of the methionines in the protein, which together with previously obtained information about the position of the disulphide bonds helped in assigning the correct amino acid sequence during model building (**Figure 79**). Moreover, the SeMet SAD data provided us with another set of good experimental phases.

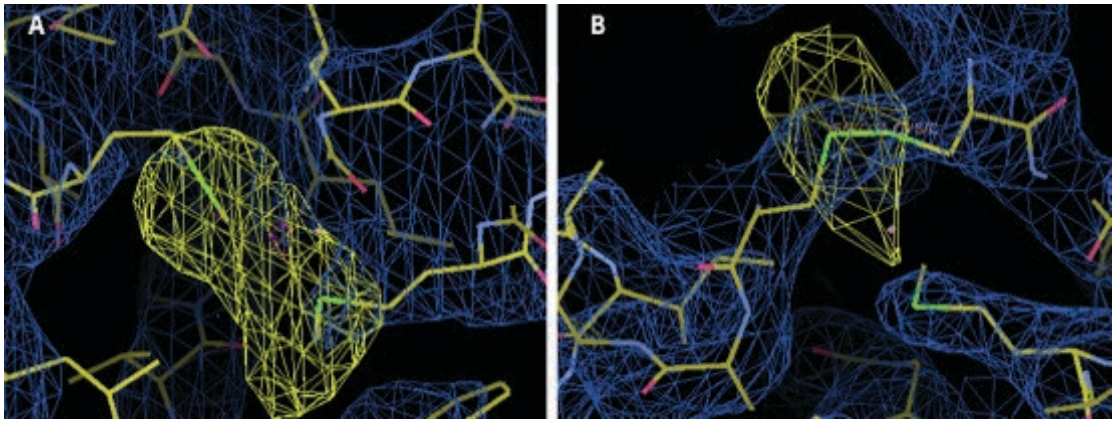


Figure 79. Electron density for the anomalous scatterers (selenium and sulphur) calculated using AnoDe (Thorn & Sheldrick, 2011). (A) Electron density maps of SeFet showing the density (yellow) for selenium within selenomethionine residues (SeMet-SAD data set for P1 crystal form) and (B) for sulphur within the disulphide bond between two cysteine residues (S-SAD data set for P1 crystal form).

In addition, I obtained crystals of SeFet deglycosylated with EndoH/EndoD. One crystal form, which grew in 14% PEG 4000, 100 mM Tris pH8.5, 200 mM lithium sulphate diffracted X-ray to 3.4 Å resolution. The crystals belonged to P321 space group with cell parameters (a=66.92, b=66.92, c=180.79, α=90, β=90, γ=120). I also tried to prepare heavy atom derivatives of this crystal form, but soaking the crystals in heavy atom solutions resulted in loss of diffraction.

The model building and refinement of SeFet was performed alternating between the electron density maps in different crystal forms. During refinement, NCS restraints and TLS groups were applied. The SeFet model was built and fully refined in the P1 crystal form (using the best native data set of 2.7 Å) and the $P4_32_12$ crystal form (using the best native data set of 2.9 Å). Data collection and refinement statistics are summarized in **Table 13**.

Table 13. Data collection and refinement statistics for the native SeFe crystals. Values in parentheses represent those in the highest resolution bin.

	SeFe (native)	SeFe (native)
Data collection		
Space group	P1	$P4_32_12$
Cell dimensions		
a, *b*, *c* (Å)	73.670, 75.080, 78.420	80.230, 80.230, 477.860
a, b, c (°)	94.06, 114.52, 114.72	90.00, 90.00, 90.00
Resolution (Å)	41.38-2.70 (2.85-2.70)	48.78-2.9 (3.005-2.901)
Solvent content (%) (molecules per asymmetric unit)	39.69 (1)	42.41 (1)
R_{merge}	0.071 (0.227)	0.15 (0.222)
Total number of observations	89018 (5710)	312580 (48778)
Total number of unique reflections	32435 (3748)	36060 (3481)
I / sI	9.5 (1.3)	9.91 (1.07)
Completeness (%)	89.3 (72.4)	99.87 (99.40)
Redundancy	2.7 (1.5)	4.2 (2.5)
Refinement		
Resolution (Å)	40.85-2.70 (2.79-2.70)	48.78-2.90 (2.98-2.90)
No. reflections	47116	36059
R_{work} / R_{free}	0.2295/0.2538	0.2144/ 0.2456
No. atoms		
Protein	9358	9442
Water	15	23
B-factors		
Wilson B-factor (Å^2)	77.18	106.48
Average B-factor (Å^2)	89.53	103.48
R.m.s. deviations		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	0.99	1.08
Ramachandran statistics[#]		
Favored (%)	94.46	94.81
Number of outliers	2	2

[#] Ramachandran statistic according to Molprobity server.

Key features of SeFet crystal structure

The model of SeFet at 2.9 Å resolution built in the tetragonal crystal form is shown in **Figure 80**. The SeFet crystal structure is consistent with the classification of this protein as a class I fusion glycoprotein as predicted from the primary sequence. The structure of SeFet possesses structural features characteristic to other class I fusion proteins in their post-fusion

form. It is composed of three tightly intertwined protomers that form a “globular head domain”, which is involved in receptor binding in many class I fusion proteins. Below the head, the trimer forms a “stalk region”, which is known to assemble into a 6HB when the class I fusion proteins adopt the post-fusion structure during membrane fusion. The protomer of SeFet is wedge-shaped and consists of a globular, predominantly β-sheet-containing head domain, and a stalk region formed mostly by α-helices (**Figure 80**). ~20% of the surface area of each protomer is buried in the trimer interface, resulting in a total buried surface area of 14370 $Å^2$. A number of inter-chain salt bridges stabilize the head region of the trimer, in addition to inter-chain hydrogen bonds and hydrophobic interactions. In contrast, the stalk region is mainly stabilized by inter-chain hydrogen bonds and hydrophobic interactions. Each protomer contains five disulphide bridges and one free cysteine residue (C229). A disulphide bond between residues C94 and C218 keeps the SeFe subunits F1 and F2 connected. Topology diagram of SeFet protomer in the post-fusion form is shown in **Figure 81**.

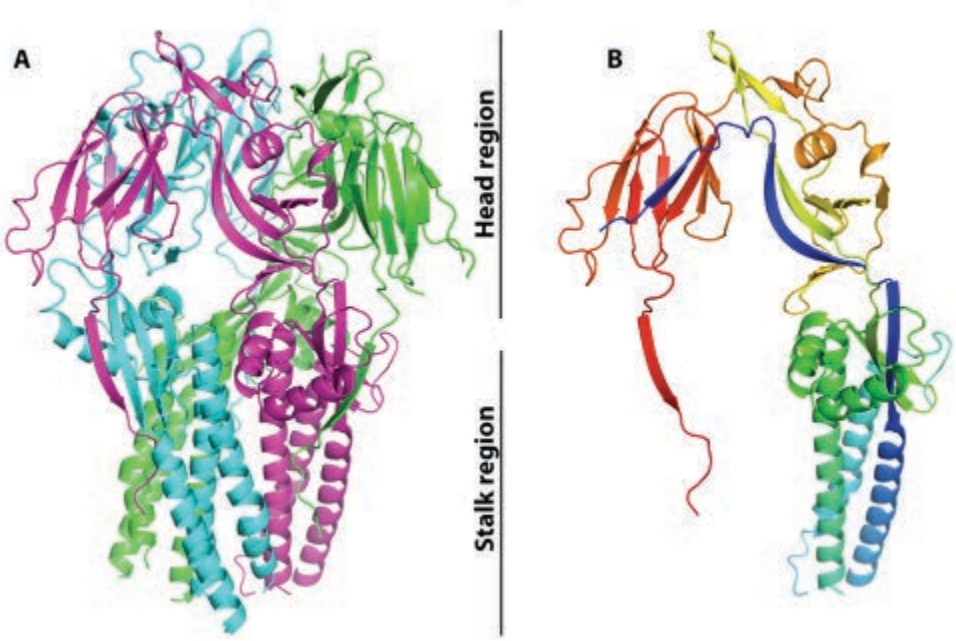


Figure 80. Crystal structure of SeFet. (A) Cartoon representation of the crystal structure of SeFet with each subunit in different colors. (B) A single SeFe protomer ramp-colored from blue (N terminus) to red (C terminus), through cyan, green, yellow and orange. The head and stalk regions are indicated in the middle.

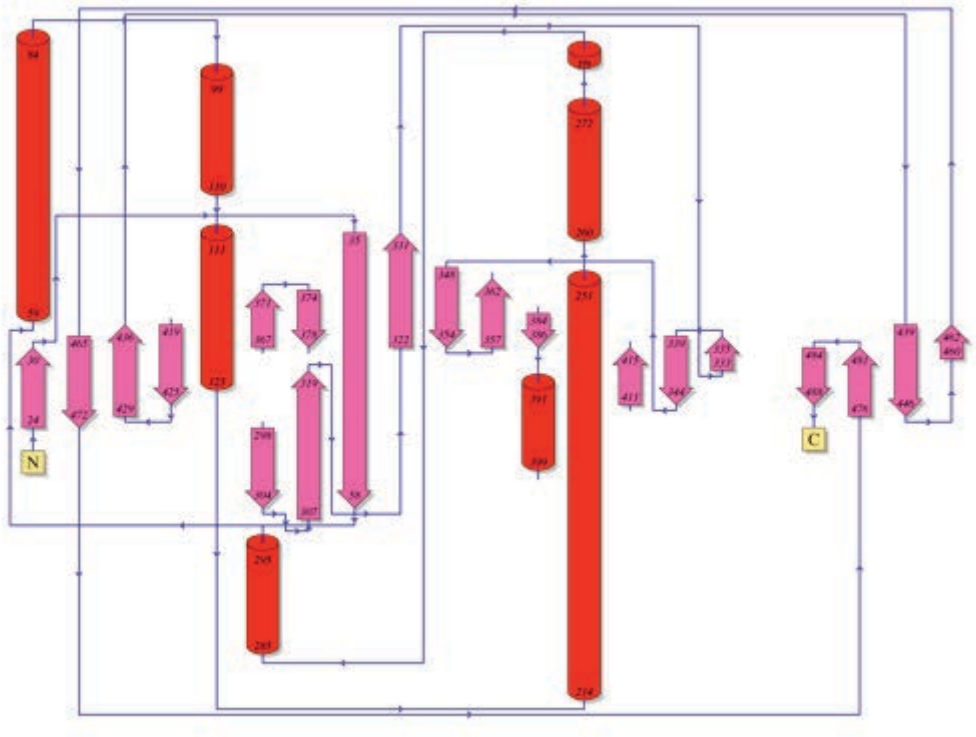


Figure 81. Topology diagram of SeFet protomer in the post-fusion form. α helices are shown as red cylinders, and β strands are shown as pink arrows. Amino acid numbers at the boundaries of each secondary structure element correspond to the full length SeMNPV F numbering (including secretion signal). The topology diagram was generated using the program PDBsum (Laskowski, 2009) for chain C of the SeFet crystal structure.

Comparison of SeFet and RSV F

I compared the SeFet structure with other protein structures available in the PDB by using a DALI server (Holm & Rosenstrom, 2010). As expected, the DALI server assigned the highest Z-scores to the fusion proteins from paramyxoviruses with a Z-score of 11.7 for RSV F protein in its post-fusion conformation (**Table 14**). A Z-score above 2 indicates that structures have significant similarities, and have similar folds. The superposition of SeFet and RSV F indeed reveals that the proteins are very similar overall in domain organization (**Figure 82**).

Table 14. Z-scores assigned by DALI server using SeFet as a query protein structure.

DALI Z-score	Protein	Reference
11.7	Respiratory syncytial virus F protein	(Swanson *et al.*, 2011)
11.1	Parainfluenza virus 3 F protein	(Yin *et al.*, 2005)

11.1	Newcastle disease virus F protein	(Swanson *et al.*, 2010)
10.5	Parainfluenza virus 5 F protein	(Yin *et al.*, 2006)
9.0	Human metapneumovirus F protein	(Wen *et al.*, 2012)

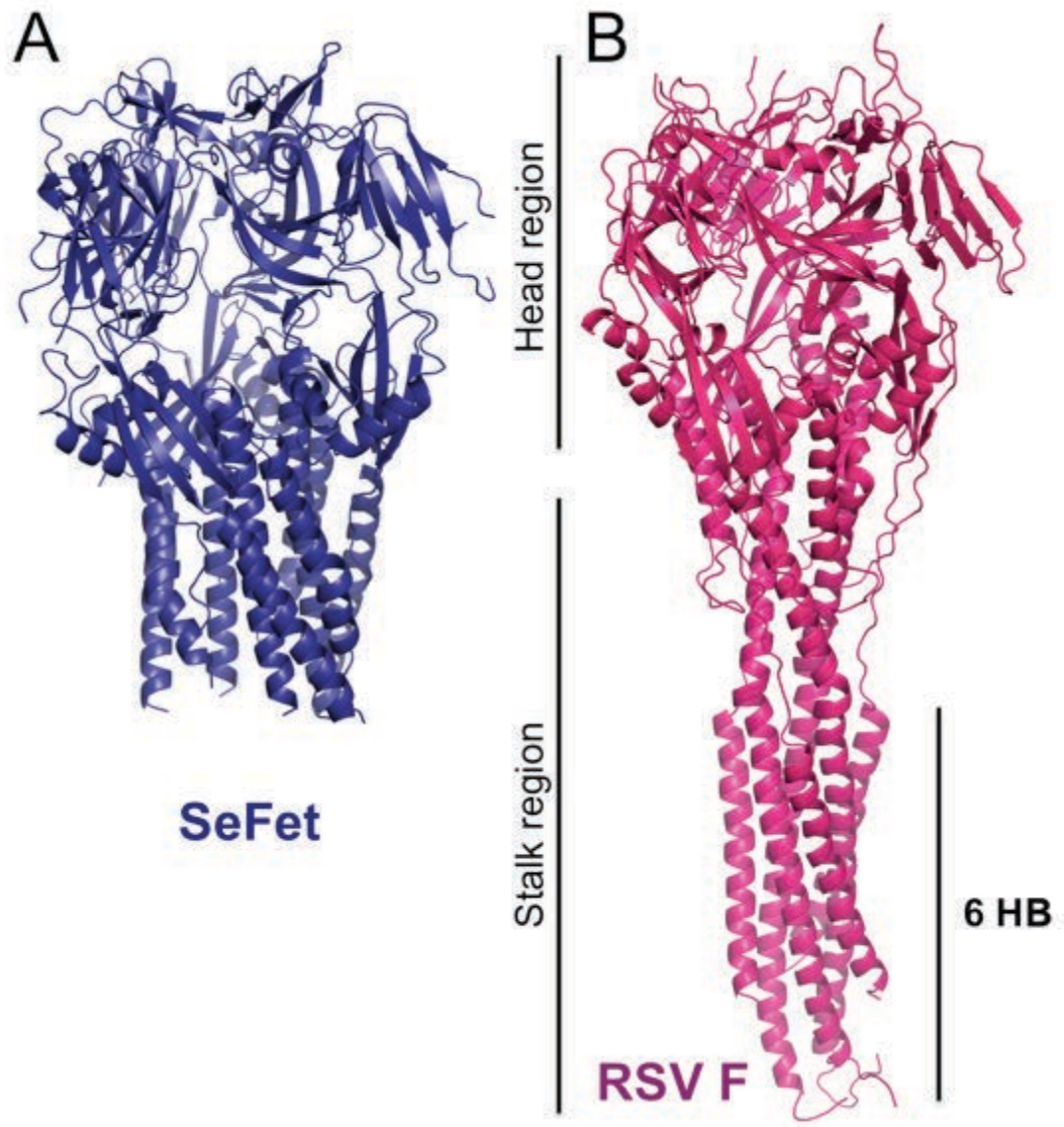


Figure 82. Comparison of the RSV F and SeFet structures. (A) Structure of SeFet (B) Structure of the RSV F post-fusion trimer (PDB ID code 3RKI). Note the absence of the N-terminal region 6HB in SeFet structure which leads to a shortening of the stalk region. The SeFet structure displays close overall correspondence with the RSV F structure in the head and the first half of the stalk.

Comparison of the superposed post-fusion trimer of RSV F and SeFet side by side (**Figure 82**) shows that the major difference between the two structures is located at the base of the stalk region. The RSV F structure extends further to reveal an intact 6HB formed by the HRA and HRB regions, while the structure of SeFet lacks this domain due to trypsin cleavage

before low pH induced trimerization (**Figure 83**). One of the missing regions in the SeFet structure, from amino acid 137 to 213, includes the fusion peptide and the HR1. Also missing is the region from amino acid 509 to 553, which includes the HR2 domain. As HR1 and HR2 correspond to the N-terminal part of HRA and HRB, respectively, in RSV F, this explains why the 6HB is not present in the crystal structures of trypsin treated low pH induced SeFet. In addition, no electron density was observed for residues 86 to 93, connecting two α-helices in the F2 subdomain. Those amino acid residues are likely to form a flexible loop that is disordered in the crystal structure. **Figure 83** shows missing regions in SeFet inserted as dashed lines. The HRA helix in RSV F extends further down the stalk by about 12 additional helical turns (45 residues), which corresponds approximately to the number of residues (42) between the fusion peptide and the N-terminal end of HR3 in SeFet. Thus, the stalk region in SeFet should be extended to approximately the same extent as in RSV F in a non-proteolysed SeFet structure.

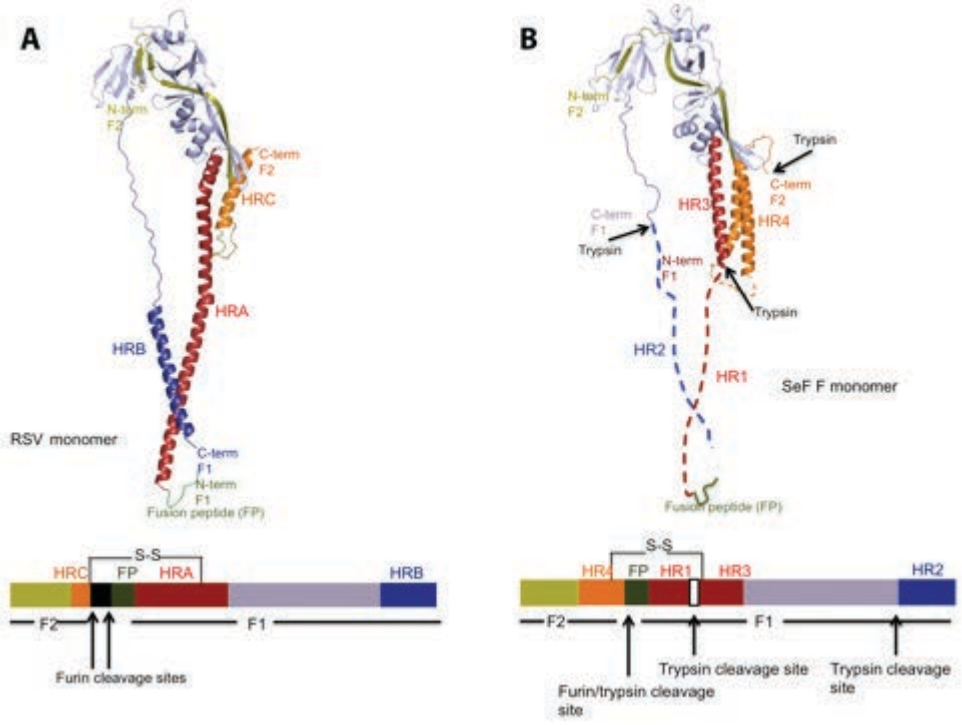


Figure 83. Comparison of SeFet and RSV F monomer structures colored by key location within the primary sequences. The domains indicated are F2 and F1, the fusion peptide (FP), and the heptad repeat regions (HR1, HR2, HR3 and H4 in SeFe; and HRA, HRB and HRC in RSV F, respectively). Furin and trypsin cleavage sites are marked by black arrows. The missing regions in SeFet are drawn as dashed lines.

Another obvious difference between the superposed models of SeFet and RSV F is present at the beginning of the stalk region. SeFe has a larger F2 domain, which results in an

additional α-helix at the C-terminus of the F2 domain which, in turn, packs against HR3 and HR4 helices (**Figure 84**). Therefore, the upper part of the stalk region in SeFet is composed of 9 helices in total and is broader than the corresponding region of the RSV F trimer, which consists only of 6 helices (HRA and HRC from each monomer).

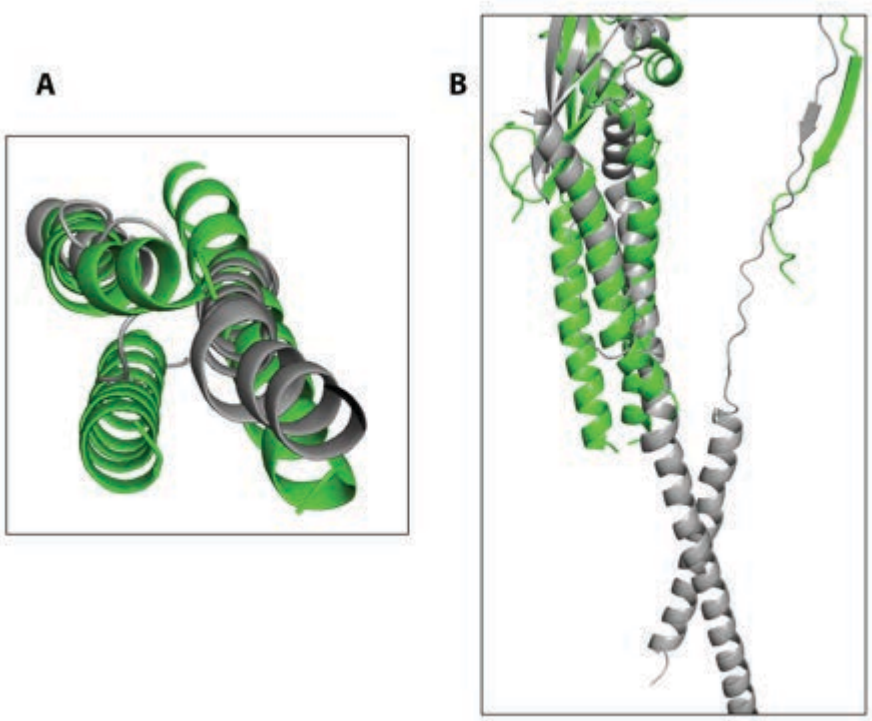


Figure 84. SeFe and RSV F comparison. (A and B) Cartoon representation of the superposed SeFe (green) and RSV F (grey) monomers showing that SeFe contains an additional α-helix at the C-terminus of the F2 domain, which packs against the HR3 and HR4 helices. In the F2 domain of RSV F, the corresponding helix is missing. (A) Superposed SeFe and RSV F viewed along the three-fold axis from the bottom of the stalk, clipped by planes that cut at the N- and C-terminal ends of the HRA and HRC helices of SeFe. (B) Superposed SeFe and RSV F viewed perpendicular to the three-fold axis.

The structures of the majority of the post-fusion viral fusogenic proteins containing trimeric coiled coils have a characteristic feature called an x-layer type stutter. The stutter is a perturbation of the heptad repeat pattern of the coiled coil due to the insertion of a four-residue motif “defg” between two “abcdefg” repeats (Lupas *et al.*, 1995). The stutter results in an opening of the coiled coil and has been shown to be a useful reference for superposing the central coiled coils of the fusion proteins from different viral families (Igonet *et al.*, 2011). Such superposition allows the comparison of the relative positions of the membrane-interacting elements (i.e. fusion loop and TM region).

The stutter position is also conserved in the *Paramyxoviridae* family with the residues 199-IDKQ-202 corresponding to the stutter in RSV F protein. The alignment of SeFe and RSV F sequences reveals that the RSV F stutter aligns with the residues 214-ADKQ-217 in

SeFe (**Figure 85**). All four residues 214-ADKQ-217 are present in SeFet structure only in chain B because of the trypsin cleavage upstream (210-RMRDADKQ-217). Nevertheless, the residues 214-ADKQ-217 superpose with the residues 199-IDKQ-202 in RSV F (**Figure 86**). Thus, the residues 214-ADKQ-217 in SeFe correspond to the stutter in the post-fusion form of the baculovirus F protein.

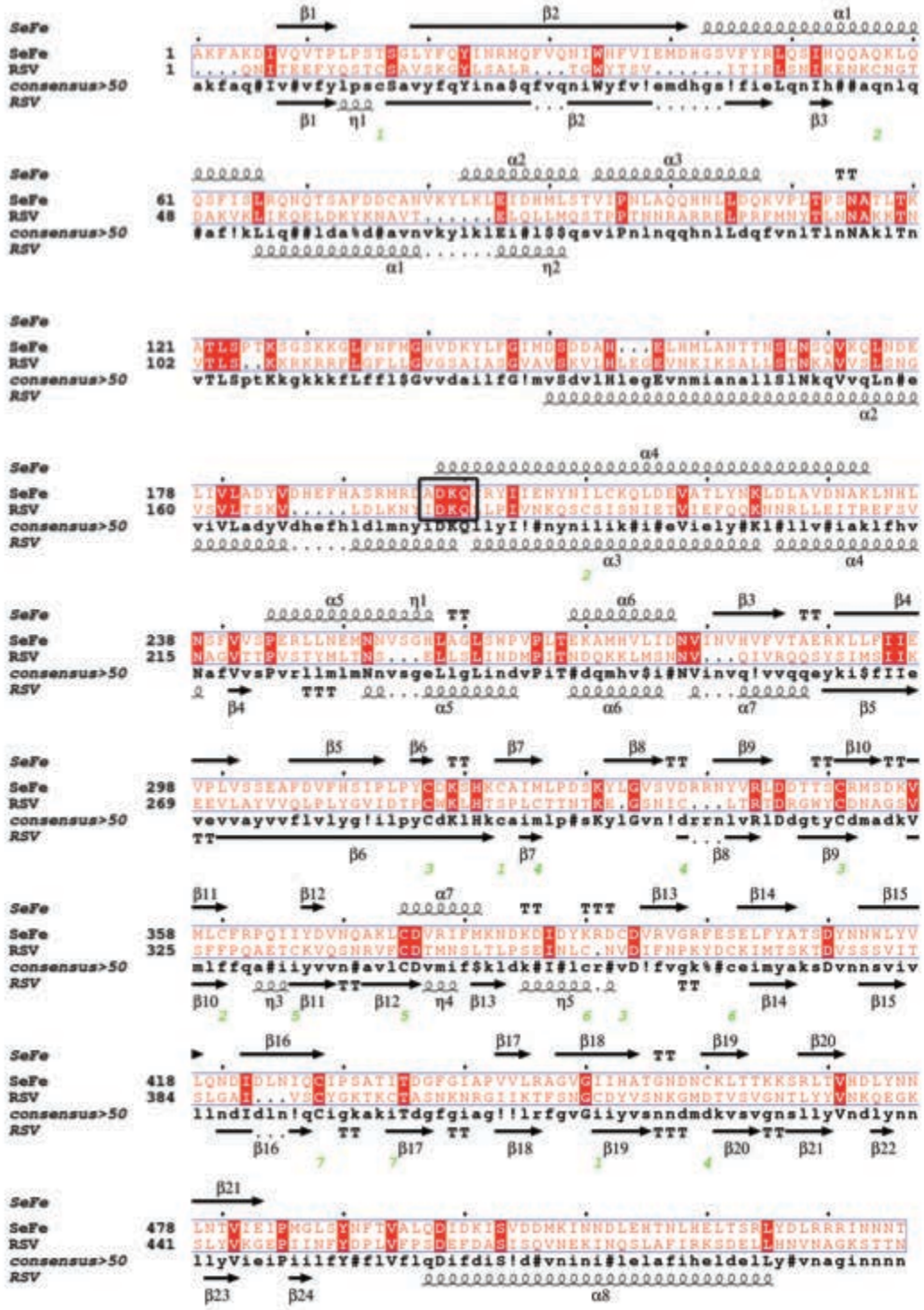


Figure 85. Sequence alignment of the F ectodomains of SeMNPV and RSV. The sequences were aligned using MultAlin (Corpet, 1988); http://multalin.toulouse.inra.fr/) and formatted with ESPript 3 (Gouet *et al.*, 2003); http://espript.ibcp.fr/). Numbering starts at the first residue of the ectodomains after the signal peptide removal. The conserved residues are highlighted in red. The conserved position of the x layer type stutter is framed by a black box. Secondary structure features of SeFe are indicated above the sequence, and secondary structure features of RSV F are indicated below the sequence. Green numbers designate residues that form disulfide bonds with the same number for each partner in a disulfide-linked pair.

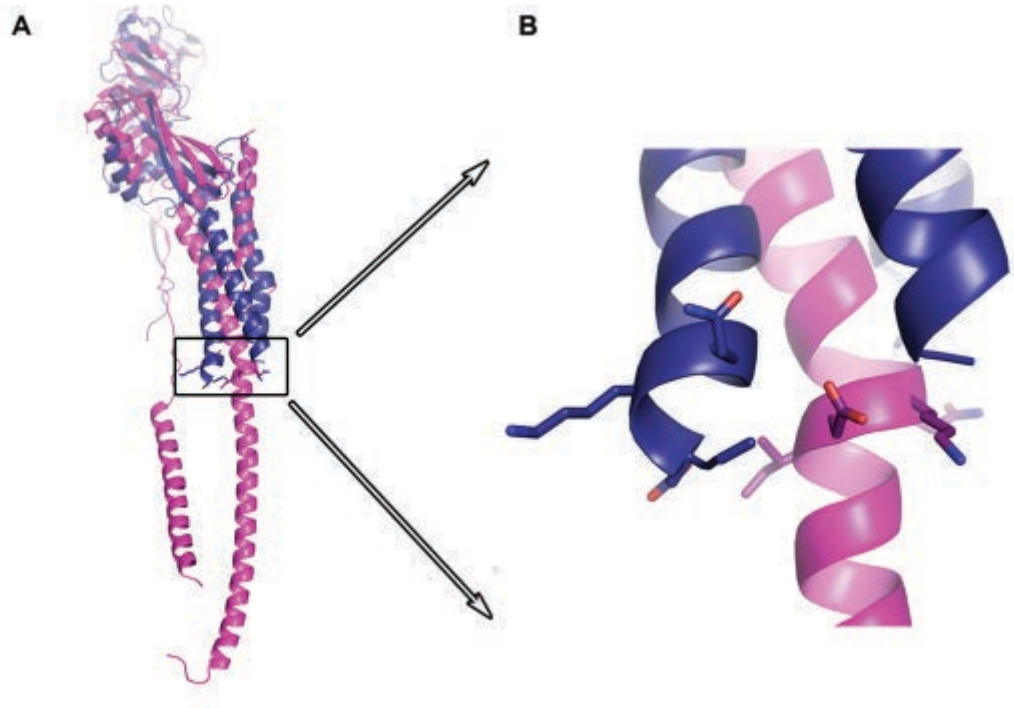


Figure 86. Stutter conservation in RSV F (magenta) and SeMNPV F (blue). (A) The position of the stutter in the superposed RSV F and SeFe models is framed by the black box. (B) A closer view of the stutter in the superposed models. The side chains of the four residues composing the x-layer stutter are displayed as sticks.

The surface of both molecules lack extensive positively or negatively charged areas (**Figure 87**). The most obvious difference between the electrostatic profiles of SeFet and RSV F is that the top of the head domain of SeFet seems to be more negatively charged. The hydrophobicity of the surface of SeFet and RSV F prevents the identification of any distinctive features between the two molecules (**Figure 88**).

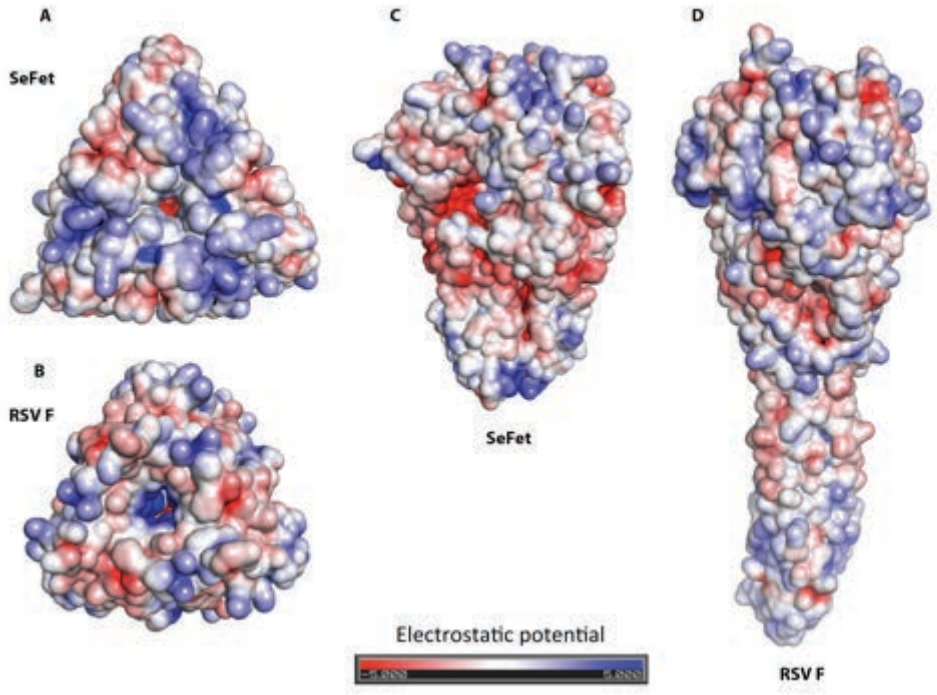


Figure 87. Electrostatic properties of the surfaces of SeFet (A and C) and RSV F (B and D). The proteins are shown as solvent-accessible surfaces colored by electrostatic potential calculated using the adaptive Poisson-Boltzmann solver and contoured at ± 5 kT/e (red, acidic/negative; blue, basic/positive). (A and B) Surface representation viewed along the three-fold axis from the top of the head. (C and D) Surface representation viewed perpendicular to the three-fold axis.

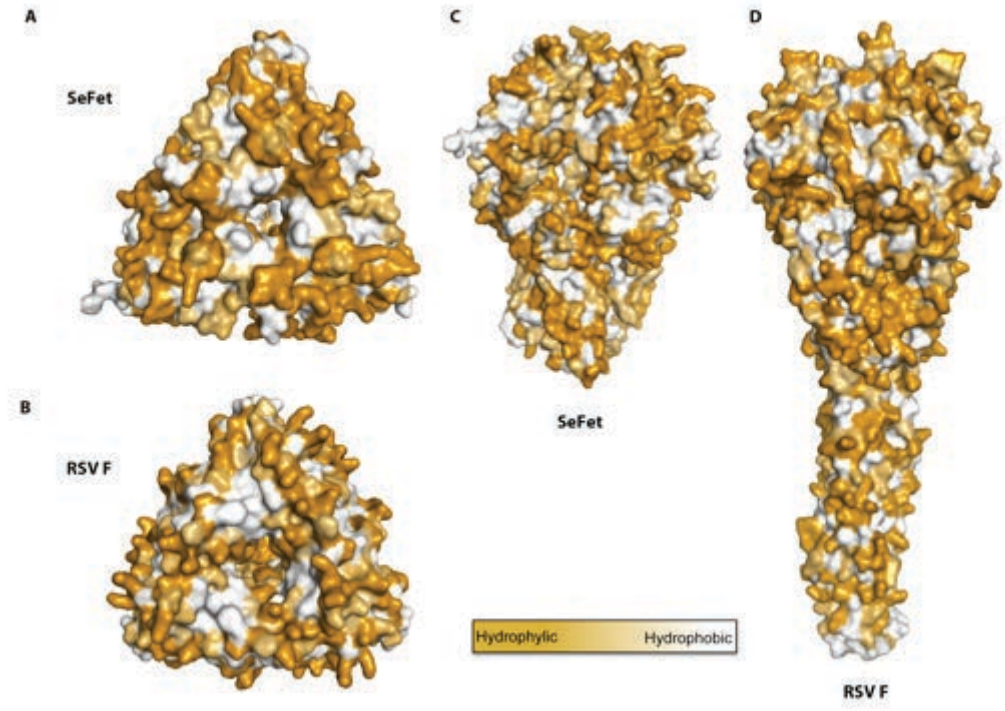


Figure 88. Hydrophobicity of the SeFet (A and C) and RSV F (B and D) surfaces. The proteins are shown as solvent-accessible surfaces colored according to a normalized hydrophobicity scale from white (hydrophobic) to bright yellow (hydrophilic). (A and B) Surface representations viewed along the three-fold axis from the top of the head. (C and D) Surface representations viewed perpendicular to the three-fold axis.

Conserved residues in baculovirus F

It has been proposed that baculovirus F protein also acts as a receptor binding protein (Westenberg *et al.*, 2007). Thus, it is possible that the F protein surface-exposed residues involved in receptor binding are evolutionarily conserved among different baculovirus F proteins. Although it is not known if the post-fusion form is still able to interact with a receptor, the residues interacting with the receptor in the post-fusion form of F protein are likely to stay exposed after the rearrangement of the protein in the post-fusion form.

In order, to identify residues that potentially could be involved in the receptor binding, I analyzed which surface exposed residues of the F protein are conserved among the 39 different baculovirus F proteins. The conserved residues were mapped using CONSURF (Glaser *et al.*, 2003). The SeFet surface colored by residue conservation is shown in **Figure 89**. The conserved residues are scattered around the whole protein. As a result, there is no easily identifiable highly conserved site. The residues around the axial canal of SeFet form the most distinct highly conserved patch. In addition, there are some conserved surface exposed residues on the side of the SeFet head region as well as the stalk region. The analysis yielded several clearly conserved residue patches on the SeFet surface that could potentially act in receptor recognition. One pretty strong cluster is present at the top of the trimer axis (**Figure 89F**). In order to further investigate the role of the conserved residues exposed on the surface of SeFet, mutagenesis of those residues combined with cell-binding/infectivity assays should be performed, but is may be wise to await for the structure of the pre-fusion form in order to carry out such an experiment.

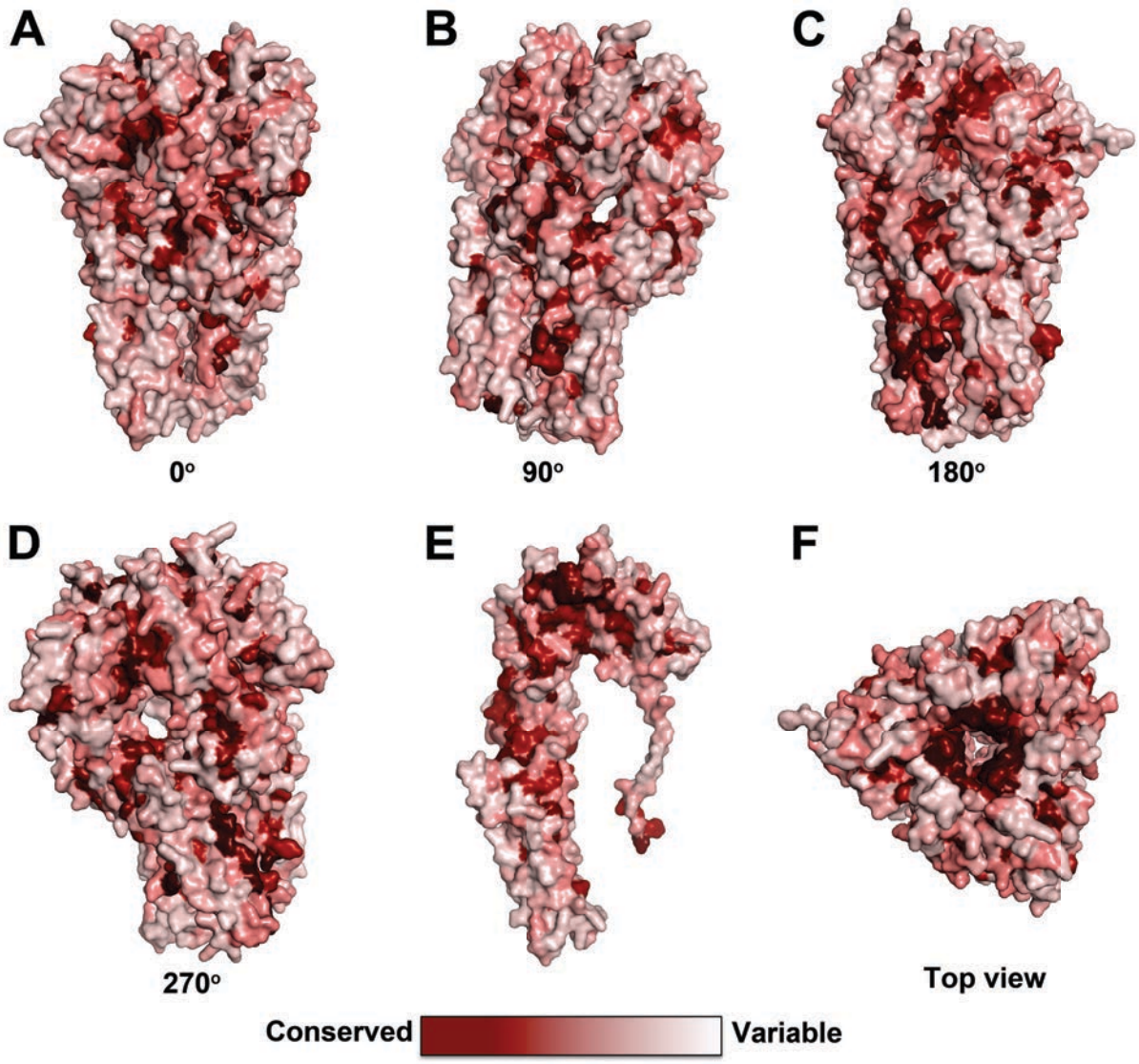


Figure 89. Molecular surface of SeFe colored by sequence conservation among SeFe and 39 F proteins from different baculoviruses analyzed by the program ConSurf (Glaser *et al.*, 2003). The colors vary from dark red for highly conserved residues to white for residues with little conservation. (A, B, C and D) Surface representation of SeFet viewed perpendicular to the three-fold axis at angles 0, 90, 180 and 270°; (E) surface representation viewed along the three-fold axis from the top of the head; (F) surface representation of a monomer viewed perpendicular to the three-fold axis.

Cellular F homologues

The PSI-BLAST search for remote homology using the SeFe sequence (including only the residues present in the 3D model of SeFet) yielded proteins containing an approximately 360 amino acid long domain of unknown function (DUF3609). DUF 3609 has been previously found in the F proteins of baculoviruses as well as in the coding sequences of *env* genes of endogenous insect retroviruses (Rohrmann & Karplus, 2001). In addition, it was also

identified in eukaryotic genomes of *Drosophila* and mosquito *Anopheles* genomes encoding cellular proteins (Malik & Henikoff, 2005). However, the PSI-BLAST search also yielded the protein sequences from other eukaryotes that have not yet been proposed to display homology with baculovirus F proteins (**Figure 91**). The highest degree of conservation is present within the sequences corresponding to the DUF3609 domain (**Figure 90**). All those sequences were found in insect genomes with the exception of *Branchiostoma floridae* (Florida lancelet) and *Lottia gigantea* (Giant owl limpet). As a result, I looked for conserved surface exposed residues between the SeFe and cellular F-homologues within DUF3609 domain (**Figure 90**). The analysis reveals that a number of highly conserved residues are exposed on the protein surface forming a few distinct patches. The functional and/or structural importance of these evolutionary conserved residues in DUF3609 domain still remains to be elucidated.

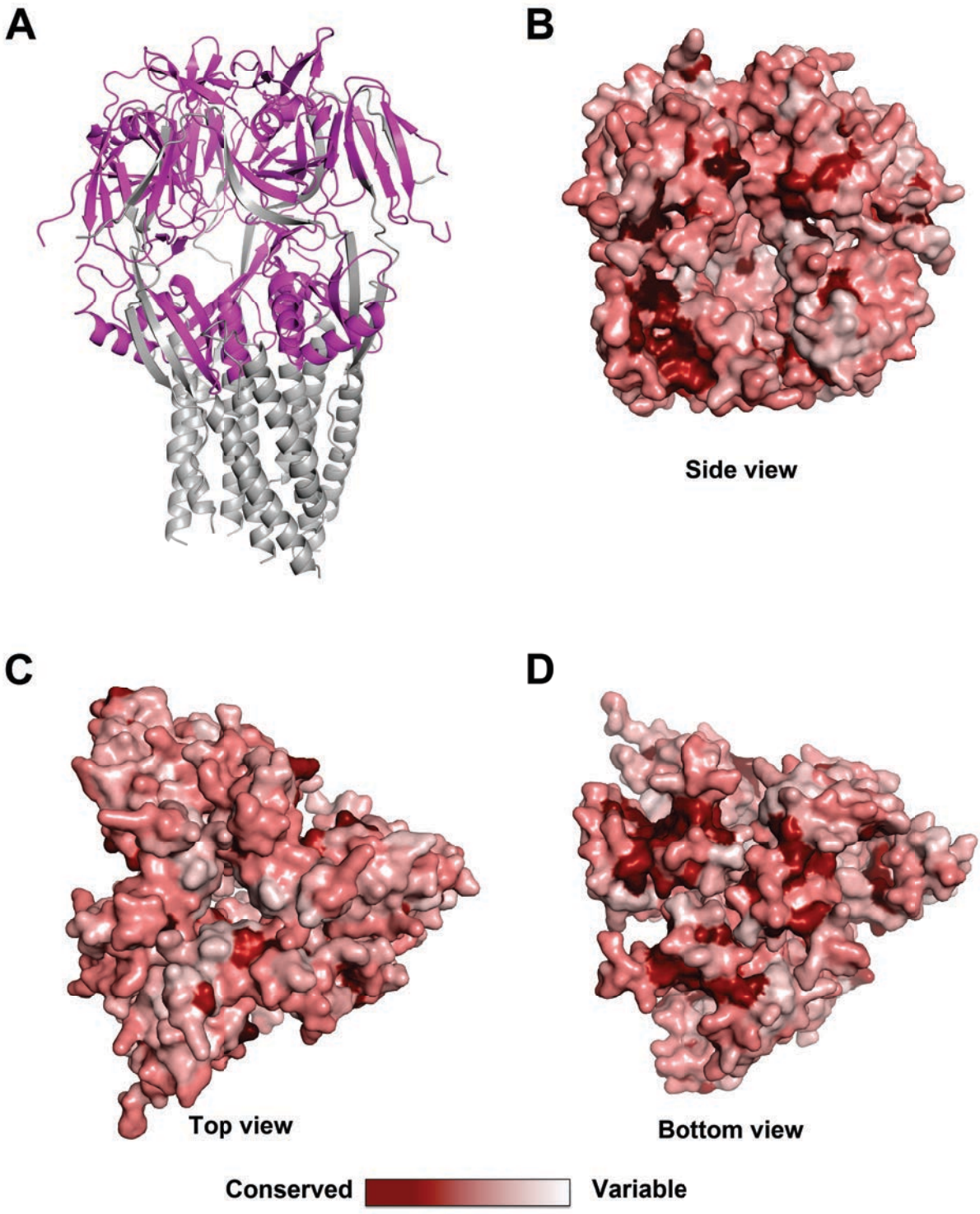
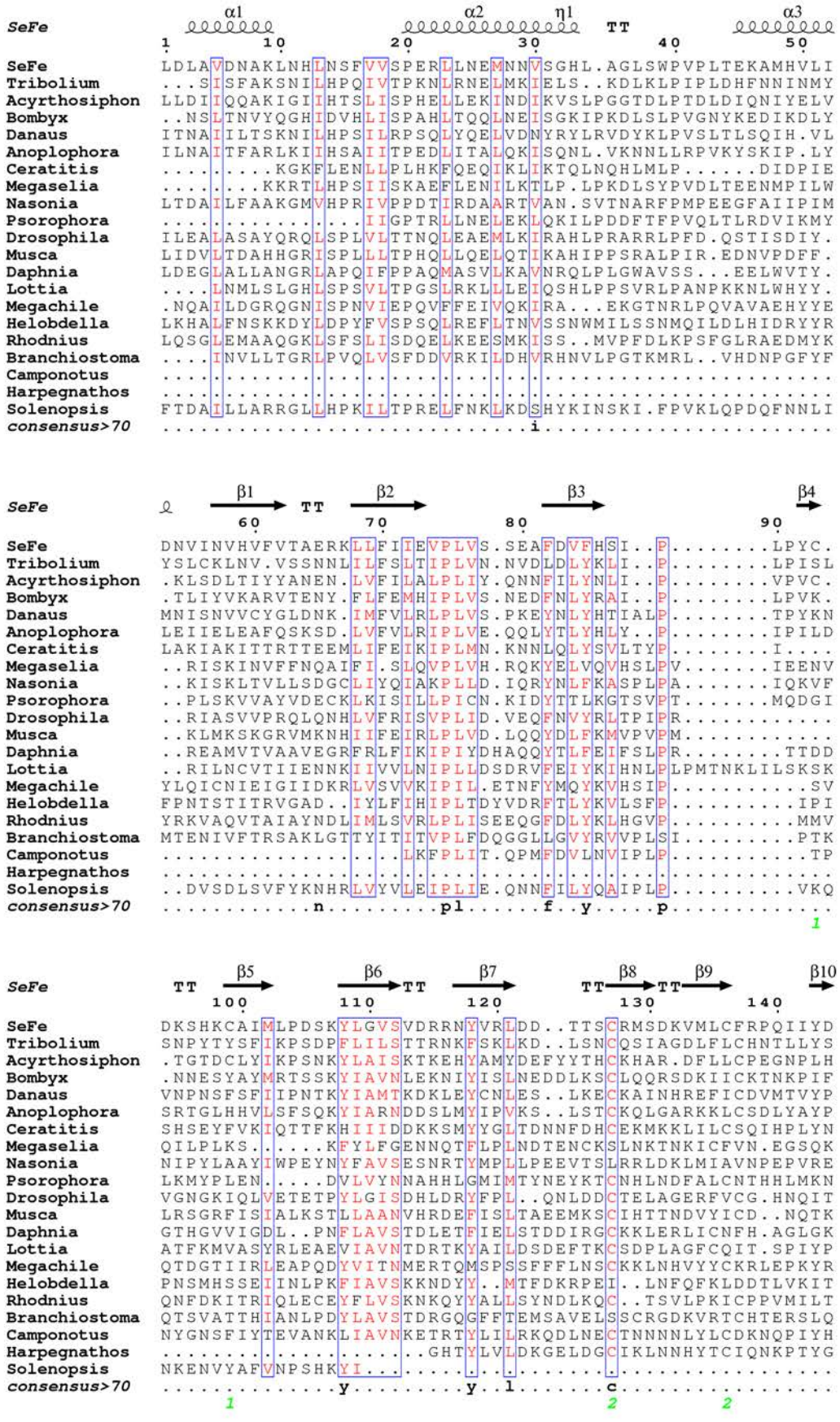


Figure 90. DUF3609 domain. (A) Cartoon representation of SeFet with the sequence of DUF3609 domain colored in magenta. (B, C and D) Molecular surface of DUF3609 domain colored by sequence conservation among SeFe and 20 cellular F homologues analyzed by the program ConSurf (Glaser *et al.*, 2003). The colors vary from dark red for highly conserved residues to white for residues with little conservation.


SeFe
α1 α2 η1 TT α3
1 10 20 30 40 50
SeFe LDLAVDNAKLNHLNSFVVSPERLLNEMNNVSGHL.AGLSWPVPLTEKAMHVLI
Tribolium ...SISFAKSNILHPQIVTPKNLRNELMKIELS..KDLKLPIPLDHFNNINMY
Acyrthosiphon LLDIIQQAKIGIIHTSLISPHELLEKINDIKVSLPGGTDLPTDLDIQNIYELV
Bombyx ..NSLTNVYQGHIDVHLISPAHLTQQLNEISGKIPKDLSLPVGNYKEDIKDLY
Danaus ITNAIILTSKNILHPSILRPSQLYQELVDNYRYLRVDYKLPVSLTLSQIH.VL
Anoplophora ILNAITFARLKIIHSAIITPEDLITALQKISQNL.VKNNLLRPVKYSKIP.LY
Ceratitis .....KGKFLENLLPLHKFQEQIKLIKTQLNQHLMLP.....DIDPIE
Megaselia .....KKRTLHPSIISKAEFLENILKTLP.LPKDLSYPVDLTEENMPILW
Nasonia LTDAILFAAKGMVHPRIVPPDTIRDAARTVAN.SVTNARFPMPEEGFAIIPIM
Psorophora .....IIGPTRLLNELEKLQKILPDDFTFPVQLTLRDVIKMY
Drosophila ILEALASAYQRQLSPLVLTTNQLEAEMLKIRAHLPRARRLPFD.QSTISDIY.
Musca LIDVLTDAHHGRISPLLLTPHQLLQELQTIKAHIPPSRALPIR.EDNVPDFF.
Daphnia LDEGLALLANGRLAPQIFPPAQMASVLKAVNRQLPLGWAVSS...EELWVTY.
Lottia ....LNMLSLGHLSPSVLTPGSLRKLLLEIQSHLPPSVRLPANPKKNLWHYY.
Megachile .NQAILDGRQGNISPNVIEPQVFFEIVQKIRA...EKGTNRLPQVAVAEHYYE
Helobdella LKHALFNSKKDYLDPYFVSPSQLREFLTNVSSNWMILSSNMQILDLHIDRYYR
Rhodnius LQSGLEMAAQGKLSFSLISDQELKEESMKISS..MVPFDLKPSFGLRAEDMYK
Branchiostoma ....INVLLTGRLPVQLVSFDDVRKILDHVRHNVLPGTKMRL..VHDNPGFYF
Camponotus .....
Harpegnathos .....
Solenopsis FTDAILLARRGLLHPKILTPRELFNKLKDSHYKINSKI.FPVKLQPDQFNNLI
consensus>70 .....i.....
SeFe
β1 TT β2 β3 β4
60 70 80 90
SeFe DNVINVHVFVTAERKLLFIIEVPLVS.SEAFDVFHSI..P.....LPYC.
Tribolium YSLCKLNV.VSSNNLILFSLTIPLVN.NVDLDLYKLI..P.....LPISL
Acyrthosiphon .KLSDLTIYYANEN.LVFILALPLIY.QNNFILYNLI..P.....VPVC.
Bombyx .TLIYVKARVTENY.FLFEMHIPLVS.NEDFNLYRAI..P.....LPVK.
Danaus MNISNVVCYGLDNK.IMFVLRLPLVS.PKEYNLYHTIALP.....TPYKN
Anoplophora LEIIELEAFQSKSD.LVFVLRIPLVE.QQLYTLYHLY..P.....IPILD
Ceratitis LAKIAKITTRTTEEMLIFEIKIPLMN.KNNLQLYSVLTYP.....I....
Megaselia ..RISKINVFFNQAIFI.SLQVPLVH.RQKYELVQVHSLPV.....IEENV
Nasonia ..KISKLTVLLSDGCLIYQIAKPLLD.IQRYNLFKASPLPA.....IQKVF
Psorophora ..PLSKVVAYVDECKLKISILLPICN.KIDYTTLKGTSVPT.....MQDGI
Drosophila ..RIASVVPRQLQNHLVFRISVPLID.VEQFNVYRLTPIPR.....
Musca ..KLMKSKGRVMKNHIIFEIRLPLVD.LQQYDLFKMVPVPM.....
Daphnia ..REAMVTVAAVEGRFRLFIKIPIYDHAQQYTLFEIFSLPR.....TTDD
Lottia ..RILNCVTIIENNKIIVVLNIPLLDSDRVFEIYKIHNLPLPMTNKLILSKSK
Megachile YLQICNIEIGIIDKRLVSVVKIPIL.ETNFYMQYKVHSIP.....SV
Helobdella FPNTSTITRVGAD..IYLFIHIPLTDYVDRFTLYKVLSFP.....IPI
Rhodnius YRKVAQVTAIAYNDLIMLSVRLPLISEEQGFDLYKLHGVP.....MMV
Branchiostoma MTENIVFTRSAKLGTTYITITVPLFDQGGLLGVYRVVPLSI.....PTK
Camponotus .....LKFPLIT.QPMFDVLNVIPLP.....TP
Harpegnathos .....
Solenopsis ..DVSDLSVFYKNHRLVYVLEIPLIE.QNNFILYQAIPLP.....VKQ
consensus>70 .....n.....pl.....f..y.....p.....
1
SeFe
TT β5 β6 TT β7 TT β8 TT β9 β10
100 110 120 130 140
SeFe DKSHKCAIMLPDSKYLGVSVDRRNYVRLDD..TTSCRMSDKVMLCFRPQIIYD
Tribolium SNPYTYSFIKPSDPFLILSTTRNKFSKLKD..LSNCQSIAGDLFLCHNTLLYS
Acyrthosiphon .TGTDCLYIKPSNKYLAISKTKEHYAMYDEFYYTHCKHAR.DFLLCPEGNPLH
Bombyx .NNESYAYMRTSSKYIAVNLEKNIYISLNEDDLKSCLQQRSDKIICKTNKPIF
Danaus VNPNSFSFIIPNTKYIAMTKDKLEYCNLES..LKECKAINHREFICDVMTVYP
Anoplophora SRTGLHHVLSFSQKYIARNDDSLMYIPVKS..LSTCKQLGARKKLCSDLYAYP
Ceratitis SHSEYFVKIQTTFKHIIIDDKKSMYYGLTDNNFDHCEKMKKLILCSQIHPLYN
Megaselia QILPLKS.....KFYLFGENNQTFLPLNDTENCKSLNKTNKICFVN.EGSQK
Nasonia NIPYLAAYIWPEYNYFAVSESNRTYMPLLPEEVTSLRRLDKLMIAVNPEPVRE
Psorophora LKMYPLEN.....DVLVYNNAHHLGMIMTYNEYKTCNHLNDFALCNTHHLMKN
Drosophila VGNGKIQLVETETPYLGISDHLDRYFPL..QNLDDCTELAGERFVCG.HNQIT
Musca LRSGRFISIALKSTLLAANVHRDEFISLTAEEMKSCIHTTNDVYICD..NQTK
Daphnia GTHGVVIGDL..PNFLAVSTDLETFIELSTDDIRGCKKLERLICNFH.AGLGK
Lottia ATFKMVASYRLEAEVIAVNTDRTKYAILDSDEFTKCSDPLAGFCQIT.SPIYP
Megachile QTDGTIIRLEAPQDYVITNMERTQMSPSSFFFLNSCKKLNHVYYCKRLEPKYR
Helobdella PNSMHSSEIINLPKFIAVSKKNDYY..MTFDKRPEI..LNFQFKLDDTLVKIT
Rhodnius QNFDKITRIQLECEYFLVSKNKQYYALLSYNDLKQC..TSVLPKICPPVMILT
Branchiostoma QTSVATTHIANLPDYLAVSTDRGQGFFTEMSAVELSSCRGDKVRTCHTERSLQ
Camponotus NYGNSFIYTEVANKLIAVNKETRTYLILRKQDLNECTNNNNLYLCDKNQPIYH
Harpegnathos .....GHTYLVLDKGELDGCIKLNNHYTCIQNKPTYG
Solenopsis NKENVYAFVNPSHKYI.....
consensus>70 .....y.....y..l.....c.....
1 2 2

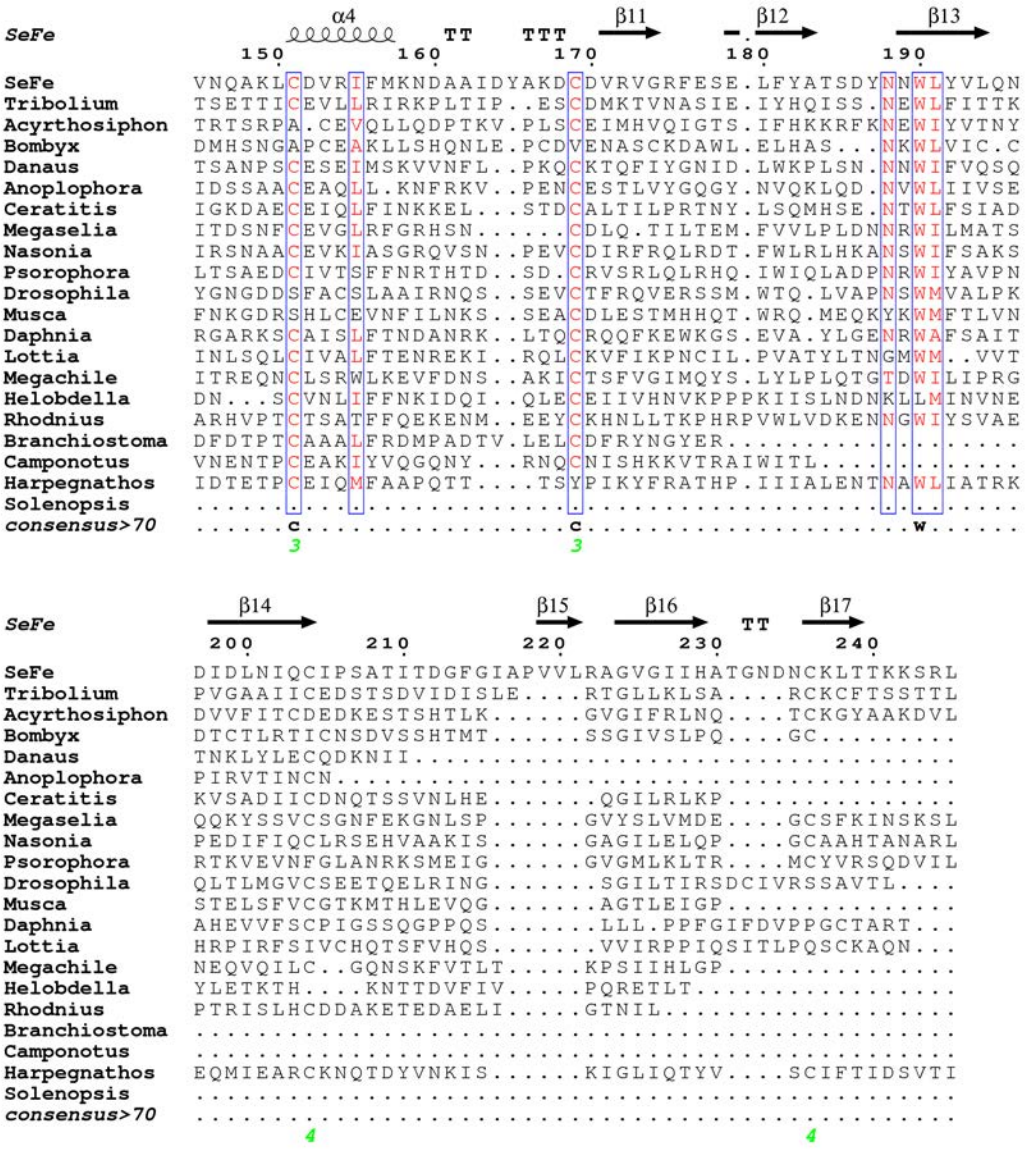


Figure 91. Multiple sequence alignment of SeFe and cellular F homologues from different species. The aligned sequences: *Megachile rotundata* (alfalfa leaf cutting bee) UniRef90_UPI000258F7DF, *Helobdella robusta* (Californian leech) UniRef90_T1FA15, *Acyrthosiphon pisum* (Pea aphid) UniRef90_J9LX66_, *Megaselia scalaris* (Humpbacked fly) UniRef90_T1GHP5_, *Psorophora albipes* (Psorophora mosquitoe) UniRef90_T1DI06, Nasonia vitripennis (Parasitic wasp) UniRef90_K7JGM3, *Branchiostoma floridae* (Florida lancelet) UniRef90_C3ZJR6, *Drosophila lutescens* (Fruit fly) UniRef90_Q30CL5, *Tribolium castaneum* (Red flour beetle) UniRef90_D7GY13, *Harpegnathos saltator* (Jerdon's jumping ant) UniRef90_E2BQZ1, *Ceratitis capitata* (Mediterranean fruit fly) UniRef90_W8AJR2, *Daphnia pulex* (Water flea) UniRef90_E9HB17, *Bombyx mori* (Silk moth) UniRef90_H9JCV6, *Lottia gigantea* (Giant owl limpet) UniRef90_V4AHE8, *Danaus plexippus* (Monarch butterfly) UniRef90_G6CIW0, *Camponotus floridanus* (Florida carpenter ant) UniRef90_E2A5H8, *Anoplophora glabripennis* (Asian longhorn beetle) UniRef90_V5I841, *Musca domestica* (House fly) UniRef90_T1PDA3, *Rhodnius prolixus* (Triatomid bug) UniRef90_T1HQD7, *Solenopsis invicta* (Red imported fire ant) UniRef90_E9J0H3. Green numbers designate residues that form disulfide bonds, with the same number for each partner in a disulfide-linked pair.

Structural studies of SeF full-length ectodomain

In order to obtain structural information about 6HB which is lacking in SeFet crystal structure, I attempted to obtain and crystallize the post-fusion trimer of the full-length SeF ectodomain. Previous observations have shown that low-pH triggered trimerization of the SeFe containing a wild type fusion peptide results in exposure of the hydrophobic fusion peptide, which leads to protein aggregation. Trypsin treatment of SeFe harbouring a mutated furin cleavage site yielded a proteolytic fragment of the protein lacking the fusion peptide, allowing successful acid-induced trimerization. It has been described in the literature that the post-fusion forms of viral fusion protein ectodomains can be obtained when the hydrophobic fusion peptide is deleted or its hydrophobic nature altered. Therefore, in order to achieve the trimerization of the full-length SeFe, I produced three new constructs of SeFe (all harboring a wild type furin cleavage site): 1) SeFeΔ1 containing the deletion of the fusion peptide residues 150-GLFNFMGHV-158 2) SeFeΔ2 containing the deletion of the fusion peptide residues 150-GLFNFMGHVDKYLF-163 and 3) SeFe-mut containing the hydrophobic residues within the fusion peptide replaced by the hydrophilic ones (150-GLFNFMGHV-158è150-GQTNSHGHN-158). In SeFeΔ1 and SeFeΔ2 constructs a short GGS linker was introduced instead of the deleted fusion peptide sequences. All proteins were expressed in *Drosophila* S2 cells yielding 6-10 mg/L of supernatant. SEC purification of SeFeΔ1 and SeFeΔ2 revealed that these proteins are monomeric at pH 8. In contrast, SeFe-mut formed both monomers and trimers at pH8. The SeFe-mut trimer was stable at pH 8 and presumably represented a pre-fusion trimer. This trimer was subjected to crystallization trials and crystals were obtained, albeit diffracting only to ~10 Å. The formation of post-fusion trimers was attempted with monomeric forms of SeFeΔ1, SeFeΔ2 and SeFe-mut by incubating them in 30 mM NaAc pH 5. Incubation of SeFeΔ2 at room temperature for two hours in 30 mM NaAc was sufficient to achieve a complete trimerization, while SeFeΔ1 and SeFe-mut required overnight incubation at room temperature for the formation of a post-fusion trimer.

Intensive crystallization trials with these proteins were performed, which resulted in obtaining crystals under different crystallization conditions. Unfortunately, all tested crystals diffracted to low resolution. The best diffraction (6.5 Å) was obtained for the SeFeΔ1 trimer after removal of the Strep-tag and deglycosylation of the protein with EndoD (space group I222; cell unit dimensions: a=111.13 b=117.09 c=348.37 $\alpha = \beta = \gamma = 90.000$ deg.). The packing of this new crystal form provides an extra space for accommodating the extension of the trimer, which is lacking in the trypsin truncated crystal form. Moreover, the analysis of the

crystal lattice shows that the crystal packing would not be stable without the extra chains joining the consecutive layers in the direction of the long axis (c=348.0) (**Figure 92**). At the moment I am trying to further improve the resolution of this crystal form, which would allow the extension of the current model of SeFet.

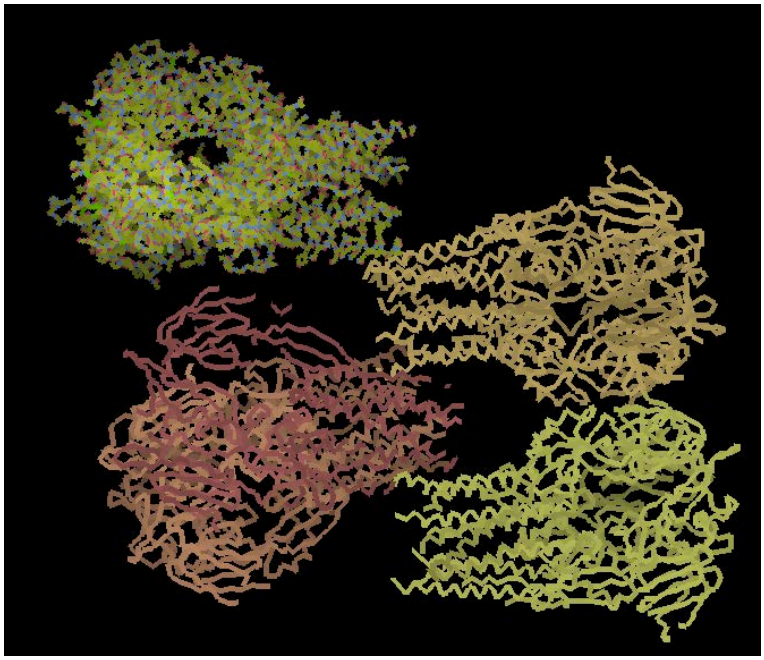


Figure 92. Crystal packing of SeFeΔ1 timer. The model of SeFet (truncated post-fusion trimer) was used for molecular replacement. Crystal packing analysis reveals an extra space for extension of the molecule.

To further delineate the structure of the full-length SeFet post-fusion form, we performed electron microscopy (EM) studies on the SeFe-mut post-fusion trimer. EM was performed by Xiaokang Zhang, a post-doc in our laboratory (**Figure 93**). Docking SeFet crystal structure into the 30-50 Å resolution reconstruction map revealed an extra extensive density corresponding to the 6HB which is absent in the crystal structure (**Figure 94**).

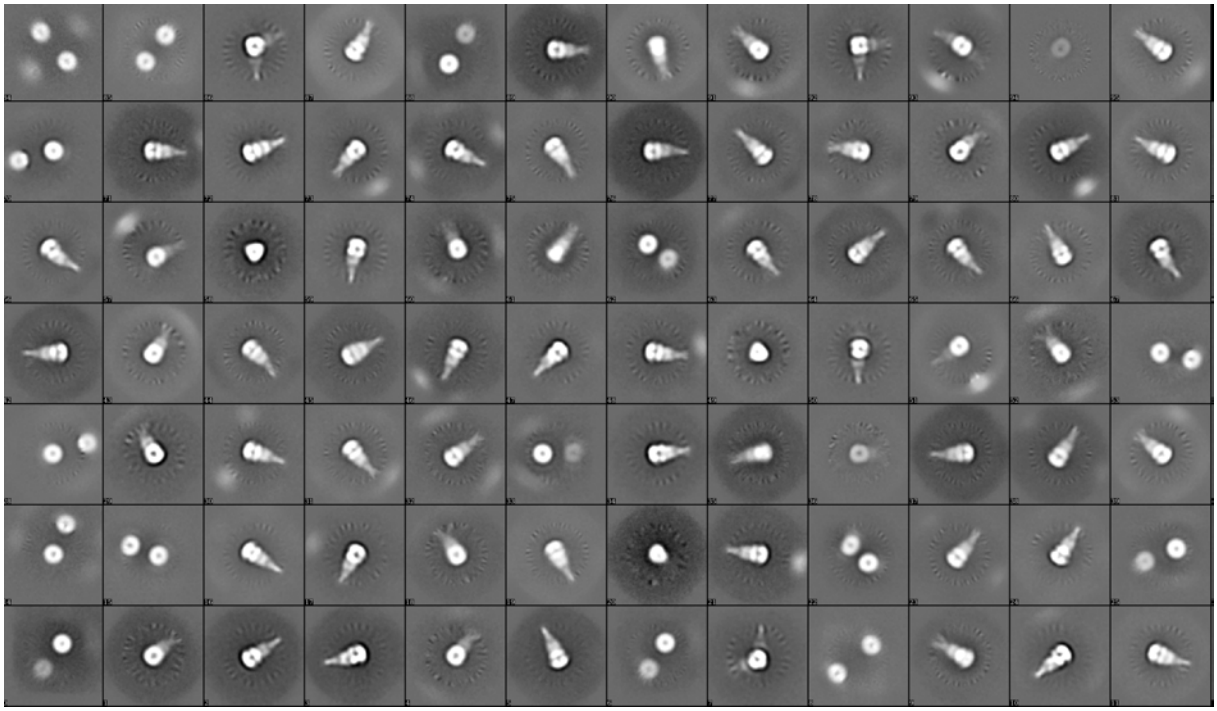


Figure 93. Reference-free 2D class averages of negatively stained SeFe-mut post-fusion trimer.

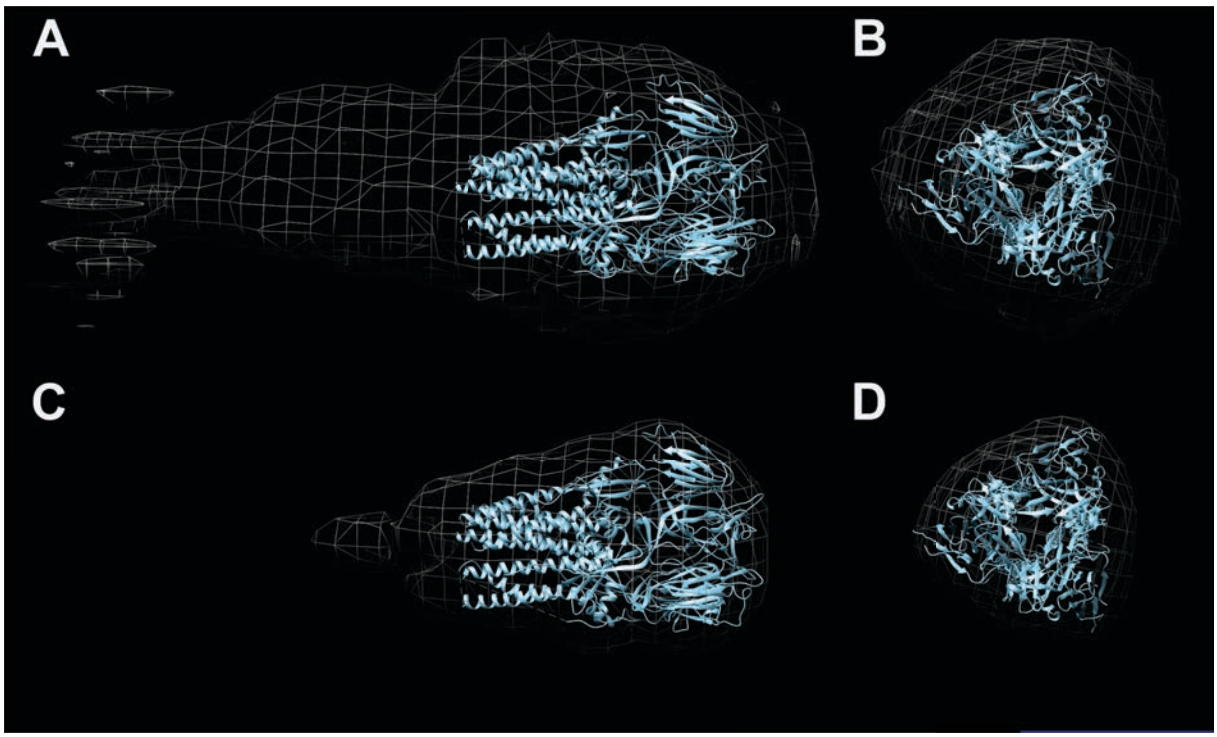


Figure 94. EM reconstruction of SeFe-mut post-fusion trimer. The crystal structure of SeFet (displayed as cartoon) is fitted within the electron density. (A and B) Side view and top view, respectively, at contour level of the density 0.651. (C and D) Side view and top view, respectively, at contour level of the density 0.162.

Supplementary data

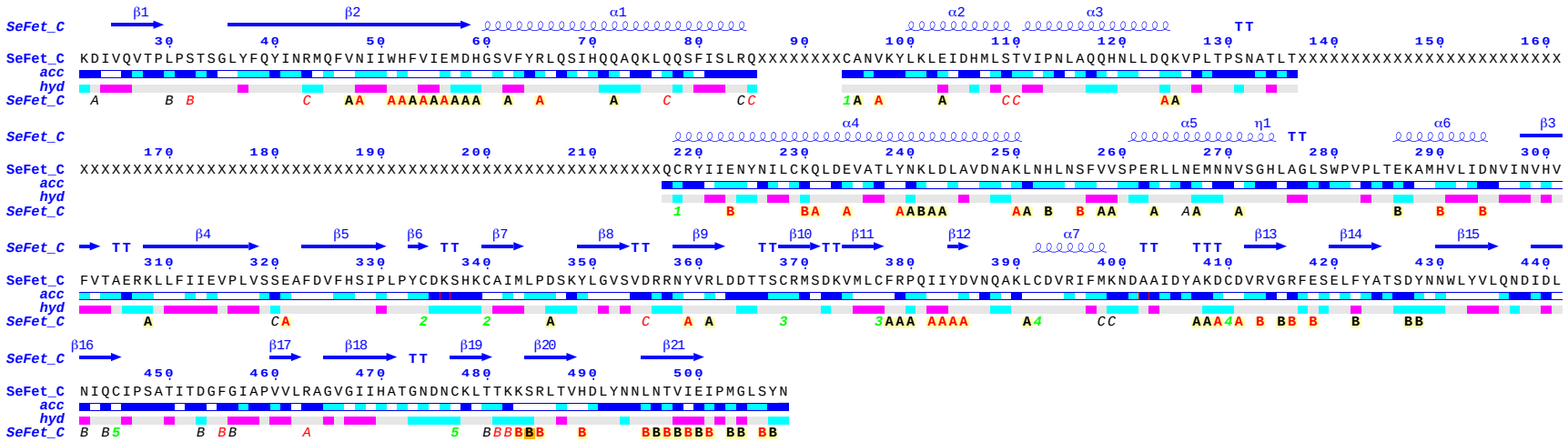


Figure 95. The amino acid sequence of the SeFe model adorned with secondary structure elements by ENDscript 2 (helices with squiggles, β-strands with arrows and turns with TT letters)(Gouet *et al.*, 2003). Green numbers designate residues that form disulfide bonds, with the same number for each partner in a disulfide-linked pair. Solvent accessibility is rendered by a first bar below the sequence (blue is accessible, cyan is intermediate, white is buried) and hydropathy by a second bar below (pink is hydrophobic, white is neutral, cyan is hydrophilic).

Discussion

Evolutionary links revealed by baculovirus F protein structure

Together with our collaborators from Utrecht University we characterized the trypsin-truncated SeMNPV F ectodomain (SeFe) and showed that low pH-induces conformational changes within the protein leading to formation of a stable post-fusion trimer. After intensive crystallization attempts, diffracting crystals of the post-fusion trimer of SeFe trypsin fragment (SeFet) were obtained. The structure determination of SeFet confirmed the previous predictions that baculovirus F protein exhibits a class I fusion protein fold.

The recombinant SeFe contained a mutated furin cleavage site in order to prevent furin cleavage and, as a result, stabilize the protein in its pre-fusion form. Based on the knowledge that furin cleavage occurs in the trans-Golgi network, the mutation was not expected to impede protein folding. In addition, available pre-fusion structures of other class I fusion proteins (influenza HA and paramyxovirus F) in their cleaved and uncleaved forms, are largely superimposable (Chen *et al.*, 1999a; Welch *et al.*, 2012). Proteolytic activation is required for the majority of class I fusion proteins to induce membrane fusion, i.e. for conformational changes to reach the post-fusion state. When altering the furin cleavage site, two arginine residues were introduced within this site in order to allow proteolytic activation of the protein by trypsin cleavage. We have previously observed that the wild type SeFe cleaved by furin aggregates if incubated at low pH, which could be explained by exposure of the hydrophobic fusion peptides that cluster together in the absence of cellular membrane. As illustrated by examples of other class I fusion proteins, obtaining a soluble form of post-fusion trimer usually requires removal of the fusion peptide (Ruigrok *et al.*, 1988; Swanson *et al.*, 2011).

Trypsin cleavage of the mutated SeFe yielded a stable proteolytic fragment of the protein lacking the fusion peptide, which was able to form a trimer when exposed to low pH. This trimer was presumably corresponding to the post-fusion form based on the observed low pH dependent fusion of the baculovirus F protein (Westenberg *et al.*, 2002). The conformational changes induced in the SeFe trypsin fragment by low pH were irreversible, which is consistent with the transition of a metastable pre-fusion form to a more stable post-fusion conformation, which is a general feature of the

majority of viral fusion proteins. SeFe with knocked-down furin cleavage site expressed in S2 cells was secreted as a monomer, which is in contrast to the observed trimeric state of the full-length protein (Long *et al.*, 2006b). The monomeric state of SeFe in solution could indicate the importance of the transmembrane (Holland *et al.*) domain in protein trimerization. Both class I and class III fusion proteins lacking TM domain have been reported to trimerize inefficiently (Albertini *et al.*, 2012; Bullough *et al.*, 1994; Singh *et al.*, 1990; Smith *et al.*, 2013; Wurdinger *et al.*, 2005). TM-TM interactions have been suggested to be important for the stability of pre-fusion forms of paramyxovirus F proteins (Smith *et al.*, 2013). The crystal structures of pre-fusion forms of parainfluenza virus 5 (PIV 5) F protein and RSV F protein trimers were obtained only when the TM domain was replaced with a trimerization domain (GCN4t or T4 fibritin, respectively) (McLellan *et al.*, 2013; Yin *et al.*, 2006). Trimeric interfaces of paramyxovirus F and influenza HA proteins are significantly different in their pre-and post-fusion forms (Bullough *et al.*, 1994; Wilson *et al.*, 1981; Yin *et al.*, 2005; Yin *et al.*, 2006) revealing weaker interprotomer interactions in the pre-fusion forms. Post-fusion forms of these proteins are stabilized via hydrophobic interactions of the HR regions forming the 6-helix bundle. The trypsin cleaved SeFe formed a stable post-fusion trimer after acid-induced conformational changes, which also promoted transition from a monomeric to a trimeric state.

The crystal structure revealed that SeFet lacks HR1 and HR2 regions and, as a result, does not form a 6-helix bundle, which is believed to contribute the most to the stability of the post-fusion state of class I fusion proteins. Instead, the assembly and stabilization of SeFet seems to depend mainly on a trimeric coiled coil formed by the HR3 segments. We complemented our crystal structure with a negative stain EM reconstruction of the post-fusion form of the full-length SeFe ectodomain. The visualization of the molecule in the EM map revealed that the molecule has a shape and size characteristic of the post-fusion trimer of the paramyxovirus F protein and allowed us to confirm the presence of the six-helix bundle at the trimer axis. The trypsin-truncated SeFet crystal structure was docked into the EM map, clearly revealing extra density corresponding to the 6-helix bundle, which is absent in the crystal structure. Taken together, the EM density map and the X-ray crystallography data provide evidence that baculovirus F protein is a typical class I fusion protein.

SeFe requires low pH in order to undergo conformational changes from the pre-fusion to the post-fusion form. We did not observe formation of post-fusion

trimeric forms at neutral pH for either trypsin-cleaved or uncleaved ectodomain. This is in contrast to F proteins from paramyxoviruses, which are able to catalyze the membrane fusion reaction at neutral pH. Thus, despite structural similarities between baculovirus F and paramyxovirus F proteins, conformational changes within these proteins leading to the fusion reaction are induced by different triggers. Conformational changes in paramyxovirus F proteins from PIV3 and PIV5 are triggered by the HN attachment protein and in others like RSV and SV5 F proteins upon receptor binding (reviewed in (Bossart *et al.*, 2013)). The secreted hPIV3 F ectodomain with a knocked-down furin cleavage site has been shown to adopt primarily the post-fusion conformation instead of the expected pre-fusion form. This indicates that the energetic barrier for transition of paramyxovirus F ectodomain from pre-fusion to post-fusion form is rather low, once the stabilizing attachment protein is removed.

Remarkably, SeFe contains 20 histidine residues compared to only 3 found in the RSV F ectodomain. The protonation of histidine residues has been reported to trigger conformational changes of pH-dependent fusion proteins leading to membrane fusion (Kampmann *et al.*, 2006). In the post-fusion form of viral fusion proteins, a number of conserved histidine residues often tend to form salt bridges with negatively charged residues that are supposed to have a stabilizing effect. Low numbers of histidine residues in RSV F ectodomain are in line with its low-pH independent fusion mechanism. Surprisingly, despite a high number of histidine residues, SeFet crystal structure indicates that none of these residues participate in salt bridges, suggesting that they have only a more general effect of increasing surface charge.

The SeFet crystal structure represents the first crystal structure of a class I fusion protein from a DNA virus. Moreover, it also confirmed the predicted similarities between baculovirus F and paramyxovirus F proteins, with the post-fusion form of RSV F protein (PDB accession code: 3RKI) having the highest DALI score. The conservation of the structural features between baculovirus F and paramyxovirus F proteins suggests that both proteins have a common ancestor, which has been already predicted from the sequence similarity (Misseri *et al.*, 2003). Remarkably, baculoviruses and paramyxoviruses are distant viral families. First of all they belong to different groups of viruses: group I (double-stranded DNA viruses) and group V (negative-sense single stranded RNA viruses), respectively (Baltimore, 1971). Second, they have different hosts: baculoviruses infect invertebrates, primarily

insects, while paramyxoviruses infect vertebrates and have a very broad host range. Finally, these viruses use different entry mechanisms to their host cells: baculovirus F proteins function both as receptor binding and fusion proteins and induce fusion in the endosomes upon acidification while most of paramyxoviruses have separate receptor and fusion proteins and fuses directly with the cellular membranes at neutral pH. VSV G and herpes simplex virus glycoprotein B is another example of the homologous fusion protein in very different viruses, one RNA and the other one DNA (Heldwein *et al.*, 2006; Roche *et al.*, 2006). Moreover, they do not infect the same hosts either. Therefore, the route by which baculoviruses and paramyxoviruses acquired the same common ancestral fusion protein is uncertain. The origin of RNA and DNA viruses is still debatable (Forterre, 2006; Holmes, 2003), therefore, it is not clear if paramyxoviruses and baculoviruses acquired F proteins independently during evolution (probably from a cellular F ortholog) or if one of the two viral families was the source of the F protein.

Intriguingly, the DUF3609 domain found in baculovirus proteins has been also discovered in some cellular proteins and retroviral elements suggesting that that they evolved from the same ancestral protein, which could be either viral or eukaryotic origin (Lung & Blissard, 2005). The SeFet crystal structure revealed that the DUF3609 domain resides within the globular head domain of SeFet. Therefore, it is unlikely that the DUF3609 domain is involved in membrane fusion. Presumably, it plays a role in cellular receptor recognition. Analysis of the conserved solvent-exposed residues in the DUF3609 domain revealed a few conserved patches on the surface of the domain that could potentially represent a receptor binding site and/or some other unidentified functional site. The fusion activity of currently identified F cellular proteins from *Drosophila melanogaster*, *Drosophila yakuba*, *Drosophila pseudoobscura* and *Anopheles gambiae* have not yet been demonstrated. Although it is possible that the fusion activity was not observed due to assay conditions and these proteins do contain other domains than DUF3609 mediating fusion, it is also likely that cellular F proteins have completely different function.

The prevalence of *f-like* genes in the genomes of many insects suggests that baculovirus *f* gene or at least a DUF3609 domain could be acquired directly or indirectly from an insect host. The indirect route might include insect endogenous retroviruses encoding f-like genes as they have been reported to be able to insert into the baculovirus genome (Malik *et al.*, 2000; Rohrmann & Karplus, 2001). However,

the direction of horizontal gene transfer is ambiguous given the fact that errantiviruses have been proposed to contain the *env* gene that they acquired from baculoviruses (Malik *et al.*, 2000).

We further extended the search of F-like proteins in different organisms using the SeFe sequence. A PSI-BLAST search yielded previously unidentified F-like protein sequences in the genomes of a number of eukaryotes, mainly insects (phylum *Arthropoda*). Intriguingly, putative homologous proteins to baculovirus F were found in *Branchiostoma floridae* (Florida lancelet) and *Lottia gigantea* (Giant owl limpet) belonging to phyla *Chordata* and *Mollusca*, respectively. *Branchiostoma floridae* represents one of the closest living invertebrate relatives of vertebrates (Delsuc *et al.*, 2006). Although lancelets split from vertebrates more than 520 million years ago, some studies revealed that vertebrates have employed the genes found in lancelets and changed their regulation or even function (Holland *et al.*, 2008). If vertebrates also contain yet unidentified *f* genes, the *f* gene found in *Branchiostoma floridae* may represent a common ancestor from which they evolved. Our results indicate that *f-like* genes might be even more widespread than previously thought. It is important to recognize, however, that more detailed analysis is required to distinguish if these genes are not part of a previously integrated retrovirus genome. Determination of the structures of F-related cellular proteins would be helpful to elucidate the evolutionary link between viral and cellular F proteins. In addition, it might also help to understand the function of cellular F proteins.

Baculovirus F protein not only functions as a fusion protein but also interacts with unknown cellular receptors at the surface of host cells. In order, to identify residues that potentially could be involved in the receptor binding, we attempted to identify surface exposed residues conserved among the 39 different baculovirus F proteins. The analysis revealed several well-defined conserved patches on the surface of the SeFet post-fusion trimer. Nevertheless, since the F protein should interact with the cellular receptor before the fusion, the structure of pre-fusion form of F would be more appropriate to identify the conserved surface residues potentially involved in receptor recognition.

Though our EM data of the full-length SeFe ectodomain post-fusion trimer revealed the presence of a 6-helix bundle in this molecule, the crystal structure of the full-length molecule would provide more insights into the fusion mechanism of baculovirus F protein. We have produced several constructs of SeFe allowing us to

successfully form the full-length post-fusion trimer. The crystallization of these SeFet variants is underway. Though crystals of the full-length SeFet in its post-fusion form have already been obtained, they diffracted to low resolution (~7 Å). In addition, we observed that the full-length SeFe containing the fusion peptide in which the hydrophobic residues were mutated to hydrophilic ones yields a stable trimer at neutral pH in addition to a monomer. This trimer presumably corresponds to the pre-fusion trimer. Initial crystallization trials of the putative pre-fusion trimer have been performed and the crystals obtained, however, the crystals diffracted only to ~10 Å.

In summary, we have developed the protocols to form a stable trimer of post-fusion baculovirus F ectodomain and determined the crystal structure of trypsin truncated SeFet. The formation of a trimer is induced by low pH treatment, which induces irreversible conformational changes within the protein and the transition from monomeric pre-fusion state to a stable trimeric post-fusion state. In order to prevent the pH triggered aggregation of the protein, the fusion peptide must be removed or hydrophobic residues within the fusion peptide must be replaced by the hydrophilic ones. The crystal structure of trypsin-truncated SeFet confirmed the evolutionary ties between baculovirus and paramyxovirus F proteins. Our results also suggest that solving the structures of F homologues found in eukaryotes is important for a better understanding of the evolutionary link between viral and cellular F proteins.

Materials and Methods

Constructs used or generated in the study

The gene encoding a soluble ectodomain of baculovirus F protein encoding amino acid residues 18-553 of the full-length SeMNPV F protein (UniProtKB accession number Q9J8C6) was inserted into pT350 (described in Supplementary materials and methods) by Qiushi Wang (Utrecht University, the Netherlands). The residues 1-17 were excluded from the construct because they comprise the signal peptide. This construct exceptionally contained a triple Strep-tag instead of the double Strep-tag. The stable cell line was generated by a previous lab member Scott Jeffers.

SeFeΔ1 containing the deletion of the fusion peptide residues 150-GLFNFMGHV-158 2) SeFeΔ2 containing the deletion of the fusion peptide residues 150-GLFNFMGHVDKYLF-163 and 3) SeFe-mut containing the hydrophobic residues within the fusion peptide replaced by the hydrophilic ones (150-GLFNFMGHV-158è150-GQTNSHGHN-158) were generated site-directed ligase independent mutagenesis (SLIM)(Chiu *et al.*, 2008). In SeFeΔ1 and SeFeΔ2 constructs a short GGS linker was introduced at the same time instead of the deleted fusion peptide sequences.

Limited proteolysis

Purified SeFe protein in HNE buffer (0.15 NaCl, 0.1 M EDTA, 5 mM Hepes pH7) at 1 mg/ml was incubated with the indicated concentrations of TPCK (L-1-tosylamide-2-phenylmethyl chloromethyl ketone)-treated trypsin (Sigma) at a trypsin:SeFe ratio of 1:800 (w/w) for 30 min at 23 ^{o}C Digestion was terminated by addition of PMSF to the final concentration of 0.2 mM (Sigma). The cleaved protein was subsequently analyzed by SDS-PAGE stained with Coomassie Blue. Target protein bands were analyzed by N-terminal sequencing.

Acid-treatment of SeFe samples

Different variants of purified SeF ectodomain protein as well as the (SeFe, SeFe trypsin truncated fragment, SeFe-mut, SeFeΔdel1 and SeFeΔdel2) were exposed to low pH by adding NaAc pH5 to a final concentration of 30 mM and incubated over night at +4 ^{o}C.

Multiangle light-scattering analysis

Multi-Angle Light Scattering (MALLS) is an analytical technique for determining absolute molar masses and the average size of particles in solution of all types of macromolecules including proteins by calculating the amount of scattered light at different angles. The sensitivity of the light scattering detector becomes increases with molar mass of the analyte. Thus it is an excellent tool for detecting oligomeric state of the proteins. The higher the aggregation number, the more sensitive the detector becomes. MALLS detector is often coupled downstream to a chromatographic system allowing using SEC, which together provides means for measuring the molar mass, size, and distribution (http://www.wyatt.eu/index.php?id=multi-angle-light-scattering).

The absolute molecular masses of different SeFe samples were determined by gel filtration combined with detection using MALLS and refractrometry (Wyatt, 1998). Purified protein (90 µg) was loaded onto a Superdex 200 10/300 column (GE Healthcare) connected to an MALS instrument and an interferometric refractometer (DAWN HELEOS II, Wyatt Technologies, Santa Barbara, CA). The column was equilibrated either at pH5 (10 mM NaAc pH5, 150 mM NaCl, 0.1 mM EDTA) or pH7 (5 mM HEPES pH7, 150 mM NaCl, 0.1 mM EDTA), respectively. The absolute molecular masses were calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA).

Deglycosylation of SeFet

Production of endodeglycosidases PNGase, EndoH and EndoD is described in Supplementary materials and methods. For crystallization trials, SeFet was deglycosylated with PNGase F using the ratio 1:16 of endoglycosidase:SeFet on a

weight basis. For deglycosylation of SeFet with EndoD-EndoH, the ratio 1:4:28 (SeFet:EndoD:EndoH) on a weight basis was used. Deglycosylation reactions were carried out at 37^{o}C. Deglycosylated SeFet was separated from PNGase by SEC on a Superdex 200 column (GE Healthcare). EndoD and EndoH were removed from the reaction mixture by Ni^{2+}ion affinity chromatography.

Removal of the Strep affinity tag

A C-terminal triple Strep tag preceded by an enterokinase recognition site was removed from the monomeric SeFe-mut, SeFeΔdel1 and SeFeΔdel2 prior trimerization at low pH by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). The detailed protocol is provided in Supplementary Materials and Methods.

Evaluation of accessibility of free cysteine residues in SeFet

The accessibility of the cysteine residues in SeFet was evaluated by DTNB-thiols assay. The detailed protocol is provided in Supplementary materials and methods.

Crystallization

Crystallization screening, crystal optimization and crystal cryo-protection techniques are in detailed described in Supplementary materials and methods. The best-diffracting crystals of SeFet deglycosylated with PNGase (P1 crystal form) were grown at 2 mg/ml SeFet in 2.5 µl hanging drops (1:1:0.5 protein:reservoir solution:seed stock volume ratio) by vapor diffusion against a reservoir solution containing 16-18% PEG 3350 and 200 mM ammonium chloride. The crystals of deglycosylated SeFet grown in 9% PEG 4000, 100 mM imidazole pH 8 and 30% 2-methyl-2,4-pentanediol (MPD) were used as a seed stock for microseeds. The $P4_32_12$ crystal form grew under the same conditions when instead of the seed stock just a solution of 9% PEG 4000, 100 mM imidazole pH 8 and 30% 2-methyl-2,4-pentanediol (MPD) was used.

Dehydration of the crystals

The dehydration experiments were performed on SeFet crystals to improve the diffraction resolution of those crystals. Dehydration removes excess solvent from the crystals which may result in tighter packing of protein molecules and a subsequent increase of the X-ray diffraction of the crystals. Two techniques were applied for the dehydration of SeFet crystals: 1) the cover slip with a hanging drop containing the crystals was transferred over a reservoir solution with a higher percentage of precipitant and allowed to equilibrate for 3 days, and 2) the crystals were directly transferred into a dehydrating solution consisting of the mother liquor with a higher percentage of precipitant, and dehydrated over a reservoir solution containing the same dehydrating solution for 3 days.

Soaking crystals in heavy atom solutions

In order to obtain the heavy atom derivative of SeFet crystals, the crystals were soaked in a number of different heavy atom compounds listed in Table 3 in the Results chapter. All compounds except hexatantalum tetradecabromide ($Ta_6Br_{12}^{2+}$) were available from Heavy atom screens (Hampton Research). $Ta_6Br_{12}^{2+}$was kindly provided by Gérard Bricogne (Global Phasing Limited, Cambridge, UK). $Ta_6Br_{12}^{2+}$ has been reported in the literature to be a powerful derivatization agent.

For soaking experiments, 100 mM stock solutions of different heavy atom compounds were prepared in water. The heavy atom soak solution for soaking SeFet crystals was composed of mother liquor containing 20 % (v/v) glycerol and a heavy compound at a specific concentration. Various concentrations of the heavy atom compound (0.5-20 mM) as well as various soaking times (from 10 min to 18 h) were tried. 2 µl of the heavy atom soak solution was pipetted onto a new siliconized cover slip and the crystals were transferred from the mother liquor to the drops of the heavy atom soak solution. The cover slip with a drop was placed above a well containing mother liquor supplemented with 20 % (v/v) glycerol.

SeFet crystals derivatized with $Ta_6Br_{12}^{+}$ were obtained by soaking crystals in a soak solution containing 1 mM $Ta_6Br_{12}^{+}$ for 18 h. The crystals were back-soaked by transferring the crystals into mother liquor containing 20% glycerol (v/v) and flash-frozen in liquid nitrogen.

Production of selenomethionine substituted SeFe

A selenomethionine (SeMet) substituted SeFe was produced in ESF-921 serum-free medium methionine-free medium (Expression Systems) supplemented with L-SeMet. Initially, 3 l of the S2-SeFe-expressing cells were grown in Insect Express medium until it reached the density of ~20×10^6/ml. The cells were collected by centrifugation for 5 min at 200 g and resuspended in 1.5 L of ESF-921 serum-free methionine-free medium supplemented with 0.8 g/L L-cysteine. After 4 h of starving, 300 mg/L L-SeMet (Acros Organics, Geel, Belgium) and 4 μM $CdCl_2$ were added. An extra 300 mg/L L-SeMet were added at day 3. The supernatant was harvested 5 days after induction. The substituted protein was purified in the same way as the native protein but the yields were 10-fold lower.

Structure determination of SeFet

Native as well as MAD (Multi-wavelength anomalous dispersion) and SAD (Single-wavelength anomalous diffraction) data sets were collected at the Synchrotron Soleil beamline Proxima 1, the Swiss Light source beamline PX I or European Synchrotron Radiation Facility (ESRF) beamlines ID23-1, ID14-4 and ID23-2 at 100 K. Typically, diffraction data were collected using Pilatus (Dectris, Baden, *Switzerland*) or charge-coupled device (CCD)-based detectors. Programs iMosflm (Leslie, 2006; Powell, 1999) or xdsme were used to determine the optimum orientation of the crystal for the complete data set collection. The best native data sets of X-ray diffraction data for $P4_32_12$ and P1 crystal forms (2.9 Å and 2.7 Å resolution, respectively) were processed using XDS (Kabsch, 2010).

The initial experimental phases were obtained by the MAD method using the $Ta_6Br_{12}^{2+}$ derivative of P1 crystal forms. Data for this heavy-atom derivative were collected on a single crystal using an inverse beam data collection strategy with wedges of 10 degrees while cycling through three wavelengths (peak, remote inflection). This strategy was used in order to optimize the measurement of Friedel pairs. In addition, due to crystal sensitivity to radiation damage, the collection of MAD data sets was performed with reduced beam intensity. The data sets obtained at each wavelength were processed using XDS and scaled together using XSCALE (Kabsch, 2010). SHELX C/D (Schneider & Sheldrick, 2002) within SHARP

(Bricogne *et al.*, 2003) was used to locate the heavy atom sites (Schneider & Sheldrick, 2002). Experimental phasing was performed with SHARP followed by solvent flattening, which provided initial experimental phases. The obtained map was used to build an initial model in polyalanines. NCS averaging was applied to further improve the map. Better phases were obtained with a MAD experiment using the $Ta_6Br_{12}^{2+}$ derivative of the $P4_32_12$ crystal form. This data set was collected using an inverse beam data collection strategy with wedges of 10 degrees while cycling through the two wavelengths of falling and mounting inflection (1.255070 (9879 eV), 1.254140 (9886eV). In the case of the $P4_32_12$ crystal form $Ta_6Br_{12}^{2+}$ derivative, individual Ta atoms were identified using SHARP.

The phases in $P4_32_12$ crystal form were combined with the phases of the initial model built in P1 crystal form. After applying density modification and multi-crystal averaging techniques, better quality maps were obtained for both crystal forms allowing the complete building of the model. NCS averaging and multi-crystal averaging was performed using the DM (CCP4 suite), and phenix.multi_crystal_average (Terwilliger, 2002). The building of the model was performed manually in Coot (Emsley *et al.*, 2010) alternating between the electron density maps in different crystal forms. Manual building was supported by anomalous data for a SeMet derivative from the $P4_32_12$ crystal form collected at the peak wavelength as well as the highly redundant Sulfur-SAD data set at a wavelength of 1.7995 Å from the native crystal in P1 form. The density for the anomalous scatterers (selenium and sulphur) was calculated using AnoDe (Thorn & Sheldrick, 2011). Refinement was performed using AutoBuster (Bricogne G, 2010) against both the P1 and the P4 data sets imposing NCS restraints and TLS groups. Refinement was monitored following the R_{free} R_{work} and R_{free} values. R_{free} values were calculated for a random subset (5 %) of reflections omitted from refinement. Water molecules were added manually using Coot. Throughout the refinement, a structure-validation web service MolProbity (Davis *et al.*, 2004) was used to monitor all-atom contact analysis as well as Ramachandran and rotamer distributions.

Crystal structure analysis

Electrostatic potentials were calculated using the adaptive Poisson Boltzmann solver (Baker *et al.*, 2001). The protein interaction calculator (Molina *et al.*) (Tina et

al., 2007) was used to identify the interactions between the protomers of SeFet. The surface area buried upon trimer formation was estimated using the PISA server (Krissinel, 2007). The topology diagram of SeFet protomer was generated using program PDBsum (Laskowski, 2009). SeFet structure comparison with other protein structures available in the PDB was carried out using DALI server (Holm & Rosenstrom, 2010). Sequence alignments were performed using MultAlin (Corpet, 1988); http://multalin.toulouse.inra.fr/) and formatted with ESPript 3 (Gouet *et al.*, 2003); http://espript.ibcp.fr/). Sequence conservation among SeFe and F proteins from different baculoviruses or different putative cellular F orthologs was analyzed by the program ConSurf (Glaser *et al.*, 2003). Figures of the crystal structures were prepared in the PyMOL Molecular Graphics System (www.pymol.org).

Final Discussion

My thesis research was initially focused on the structural analysis of the GB virus B envelope protein E2. Despite extensive efforts described in this thesis I was unable to obtain diffraction-quality crystals of this protein. During the course of this project I did, however, learn a number of important techniques such as construct design and optimization, protein expression and purification using several different techniques, protein deglycosylation, limited proteolysis, protein crystallization as well as generation and expression of ligands in form of antibody fragments.

As a result, the experience obtained during this initial project was extremely useful and subsequently applied to the following two projects I pursued. The first of these projects focused on the characterization of an important neutralizing epitope within HCV E2 and the second project focused on the structural analysis of the baculovirus F fusion protein.

In the course of the first project, I determined three structures of antibody fragments in complex with a synthetic peptide mimicking the HCV E2 epitope spanning aa 529-540 by molecular replacement. This experience allowed me to improve my skills in this phasing technique that is frequently used in macromolecular crystals.

In the second project, I determined the structure of a trypsin truncated fragment of the baculovirus F protein using experimental phasing techniques such as Multi-wavelength Anomalous Dispersion (Albertini *et al.*) and Single wavelength anomalous dispersion (Drexler *et al.*). As described in this thesis, the phasing of baculovirus F protein crystals was a difficult case that gave me the opportunity to learn experimental phasing of challenging crystallization targets. Already the challenge to obtain diffraction quality crystals of the baculovirus F protein fragment allowed me to practice different crystal optimization techniques. Therefore, these projects contributed greatly to my profound training in structural biology approaches.

Neither GBV-B nor baculovirus are viruses that infect humans, but the structural characterization of the envelope glycoproteins of these viruses can provide important insights, including both evolutionary and functional aspects, to understand the mechanism-of-action of their counterparts in important human pathogens. For example, the same fusion protein fold can be found in different viral families (e.g.

class II fusion proteins in *Flaviviridae*, *Bunyaviridae* and *Togaviridae*). However, as the crystal structures of HCV and BVDV E2 glycoproteins reveal, in some cases structurally divergent viral glycoproteins might exist within the same virus family.

The evolutionary origin of fusion proteins is still not clear and debatable. At the moment there is not a single general principle explaining the origin of these proteins. In some viruses the similar overall fold of the fusion protein seems to be the result of divergent evolution (i.e., evolution from a common ancestor), while in other viruses the similar overall fold likely results from convergent evolution (i.e., a similar overall fold implying a specific fusion mechanism evolved in parallel from distinct ancestor proteins).

For example, class II fusion proteins from the *Flaviviridae*, *Bunyaviridae* and *Togaviridae* families appear to have evolved from a common ancestor. The amino acid sequences of these proteins do not display any sequence similarity, but the proteins adopt the same overall fold. Structural relationships between class I fusion proteins from paramyxoviruses and baculoviruses described in this thesis offer a second example of fusion proteins with a common ancestor, although their structural homology was predicted from sequence analysis. VSV G and herpes simplex virus glycoprotein B belonging to class III fusion proteins also represent homologous fusion proteins and may have a common ancestor.

The role of divergent evolution in the origin of fusion proteins can be deduced from the fact that structures of different class I fusion proteins can have distinct characteristics despite the fact that they use a common fusion mechanism. Post-fusion forms of class I fusion proteins from retro-, filo-, arena- and influenza viruses do not harbor a globular head domain found in paramyxoviruses or coronaviruses although all of these proteins form a trimeric coiled coil, indicating that they are not true structural homologs. Therefore, they most likely do not have a common ancestry and evolved independently to mediate membrane fusion. Interestingly, class I viral fusion proteins share an α-helical coiled coil architecture and a similar fusion mechanism driven by the zippering of the coiled coils with the SNARE family of intracellular vesicle fusion proteins (Donald *et al.*, 2011; Skehel & Wiley, 1998). This is also considered to be an example of convergent evolution.

The fact that homologous viral fusion proteins are found in evolutionary distant viral families (e.g., F proteins of paramyxoviruses and baculoviruses, which

are RNA and DNA viruses, respectively) raises the question how viral fusion proteins were distributed across different viral families. There are two prevailing hypothesis explaining this: 1) the origin of the fusion proteins is viral and different viruses acquired their fusion proteins by horizontal virus-to-virus transfer, and 2) viruses acquired their fusion machinery from their hosts (host-to-virus horizontal transfer).

While the true origin of viral fusion proteins remains to be elucidated, an extensive horizontal gene transfer between different viruses as well as between viruses and cells becomes more and more evident. In some cases, direct transfer of cellular fusion protein genes between the viral genome and the germline is very likely. The best example for this transfer are the syncytins, a protein family catalyzing syncytial cell layer formation during placental development that is derived from envelope proteins of endogenous retroviruses (reviewed in (Mi *et al.*, 2000)). Syncytin genes from different endogenous viruses were captured and domesticated by various mammalian species independently on different occasions during evolution. Another example that is less obvious is the *C. elegans* cell-cell fusion protein EFF-1 (Perez-Vargas *et al.*, 2014), which was recently reported to adopt a class II fusion protein fold and was possibly acquired from endogenous retroviruses carrying a bunyavirus-like glycoprotein (Frame *et al.*, 2001; Malik *et al.*, 2000) - a family that also carries class II fusion proteins (Dessau & Modis, 2013). On the other hand, in some cases virus-to-virus horizontal transfer of a common ancestral membrane fusion protein seems to be more likely (e.g., baculoviruses and paramyxovirus F proteins)

Given the complicated relationship between viruses and their hosts it is difficult to elucidate the origin of specific fusion proteins. Which of these mechanisms apply in the case of baculovirus F protein remains to be answered. However, the benefit of structural studies on viral glycoproteins to understand these evolutionary relationships cannot be debated anymore.

Supplementary Materials and Methods

Cell lines

The Drosophila Schneider 2 (S2) cell line was purchased from Invitrogen. This cell line was originally derived from a primary culture of late stage Drosophila melanogaster S2 embryos. Drosophila Schneider 2 (S2) cells were grown at 28°C in a normal atmosphere in a Schneider's Drosophila medium (Invitrogen, Carlsbad, USA) supplemented with 10% fetal bovine serum (FBS), 50 U/ml penicillin, and 50 µg/ml streptomycin. The cells were maintained applying standard protocols provided by Invitrogen. A remarkable feature of S2 cells is that they are able to integrate a high copy number of plasmids into their genome which renders them suitable for high-level expression of recombinant proteins. Stable cell lines expressing the recombinant proteins were generated by co-transfection of the expression and selection plasmids as described later. Stable cell lines were adapted to, and cultured in, serum-free Insect Xpress media (Lonza, Basel, Switzerland) which was also used for protein production. S2 cells grow as a loose, semi-adherent monolayer in tissue culture flasks and are, thus, easily adapted to grow in suspension. For protein production, stable, suspension-adapted cells were propagated in spinner flasks of different sizes (from 1 to 4 l) under continuous agitation at 90 RPM.

Expression vectors

For expression the recombinant proteins in Drosophila cells, a modified version of the pMT/BiP/V5-His vector (Invitrogen) plasmid (Invitrogen) designated pT350 was used. pT350 contains a double Strep-tag (IBA, http://www.iba-go.com/) with a linker region (GlyGlySer)4 in between for efficient purification of recombinant protein. An enterokinase cleavage site (DDDDK↑X) is introduced upstream of the Strep-tag to allow specific removal of the tag. The vector contains the inducible metallothionein promoter which is induced upon addition of divalent cations (Cu^{2+} or Cd^{2+}). A gene of recombinant protein is cloned in frame with *Drosophila* BiP signal sequence at the N-terminus which serves for translocation of recombinant protein into the endoplasmic reticulum lumen and directs the protein through the secretory pathway of S2 cells into the culture medium.

The modified versions of the pT350 vector were used for expression of recombinant Fab and scFv molecules as previously described (Backovic *et al.*, 2010; Gilmartin *et al.*, 2012). The vectors are designated pMT-Fab-Strep and pMT-scFv-Strep, respectively.

Polymerase chain reaction

Polymerase chain reaction (PCR) for amplification of a specific insert for restriction cloning or blunt-cloning was performed using Phusion High-Fidelity DNA polymerase (Finnzymes) as recommended by the producer. Typically the PCR reaction contained: 0.5 µM of each primer (MWG Operon), 50 ng template plasmid DNA, 200 µM dNTPs, 1 U Phusion High-Fidelity DNA polymerase (Finnzymes), 3% DMSO, and 5×Phusion HF Buffer in a total volume of 50 µl. PCR was performed under the following conditions:

Step	**Temperature, ^{o}C**	**Time**	**Number of cycles**
Initial denaturation	98	3 min	1
Denaturation Primer annealing Extension	98 X[#] 72	10 s 40 s 30s/kb	40
Final extension	72	10 min	1
Cooling	4	∞	1

[#]Primer annealing temperature was calculated according to the nearest-neighbour method (Breslauer *et al.*, 1986)

DNA restriction digest, ligation and transformation

Purified PCR products and appropriate vectors were digested for 1–2 hours with the corresponding restriction enzymes (New England Biolabs, Ipswich, MA) according to the manufacturer's protocols. The digested DNA was verified by agarose gel electrophoresis and purified using Macherey Gel and a PCR Clean-up Kit (Macherey-Nagel). Linearized vectors were de-phosphorylated by Antarctic Phosphatase (NEB) applying a protocol provided by the manufacturer. The DNA vector and insert were ligated (3:1 molar ratio of insert DNA termini to vector DNA) for 1 h at 16–25 °C using T4 DNA ligase (New England Biolabs). 2-5 µl of the ligation mixture was used to transform 50 µl of TOP10 DH5α chemically competent cells (Invitrogen) using standard procedures. 50-300 µl of cell suspension was spread

on LB -agar plates containing an appropriate antibiotic and incubated for 16–20 hours at 37 °C. Typically one colony of the transformed DH5α cells was inoculated into 4 ml of LB supplemented with antibiotic and grown with agitation for 16-18 h at 37 °C. The plasmid DNA was purified using a NucleoSpin Miniprep Kit (Macherey-Nagel) as described in the kit manual. To verify the presence of an insert the purified plasmid DNA was digested with appropriate restriction enzymes and analyzed by agarose gel electrophoresis. The plasmid DNA containing an insert of the expected size was sequenced for final validation (GATC Biotech AG, Constance, Germany or Sequencing Facility of Institut Cochin, Paris, France).

Production of recombinant proteins in *Drosophila S2* cells

For the large scale production of the recombinant proteins in *Drosophila S2* cells, the corresponding stable cell line was expanded at 28°C to a final volume of 1-4 l in spinner flasks depending on recombinant protein expression level. The cells were induced with 4 μM $CdCl_2$ when they reached the density $>8\times10^6$ cells/ml. After 6-10 days the cells were pelleted by centrifugation for 15 min at 15 000 × g and the supernatant concentrated using a Vivaflow 200 flip filtration concentration system with 10 kD cutoff membrane (Sartorius Stedim Biotech GmbH, Goettingen, Germany) to ~50 ml. The pH of the concentrated supernatant was adjusted by adding Tris pH 8 to a final concentration of 100 mM. Avidin was added to 15 μg/ml to prevent binding of the biotin present in the medium.

The supernatant was cleared by centrifugation at 20 000 × g for 30 min and filtered with 0.2 mm cutoff membrane and loaded onto a Streptactin Superflow column (IBA, Gottingen, Germany). The column was washed with 5 column volumes (CV) of 0.1 M Tris pH 8, 0.15 M NaCl, 1 mM EDTA, and a recombinant protein was eluted with 2 CV of the elution buffer containing 2.5 mM desthiobiotin, 0.1 M Tris pH 8, 0.15 M NaCl, 1 mM EDTA. The eluate was further purified by size exclusion chromatography using Superdex 200 or Superdex 75 column (GE Healthcare, Uppsala, Sweden). Typically, SEC was performed at 15 in 10 mM Tris pH 8 and 150 mM NaCl. In the case of SeFe, a monomer was purified by SEC in 5 mM Hepes, 150 mM NaCl, and 0.1 mM EDTA (HNE) buffer. Protein fractions were analyzed by SDS-PAGE. Based on the SDS-PAGE analysis fractions containing >95 % pure protein were pooled and concentrated to approximately 20 mg/ml using a Vivaspin

centrifugal concentrator (Sartorius Stedim Biotech GmbH, Goettingen, Germany).

Protein concentrations were determined from the absorbance at 280 nm on a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) using calculated molar extinction coefficients from a program ProtParam (Gasteiger et al., 2005). Purified recombinant proteins were immediately used for subsequent experiments or stored at +4 °C.

Production of endodeglycosidases and deglycosylation

For expression of PNGase, 3 l of LB medium containing 50 pg/ml ampicillin was inoculated 1/100 with an overnight culture of *E. co*li BL21 (DE3) strain transformed with pBlueScript (Agilent Technologies)-EndoH. After growing the cells at 37 ^{o}C to an OD550 of 0.5-0.6 the culture was cooled to room temperature. Expression was induced with 0.5 mM isopropyl- I-thio-P-D-galactoside (IPTG) and the culture was incubated overnight at 30°C.

The cells were harvested by centrifugation at 6000 g for 15 min and EndoH was purified from the periplasm of the bacteria. The pelleted cells were resuspended in 5% of the original culture volume in ice cold 20% sucrose, 0.1 M Tris pH 8 and 1 mM EDTA and pelleted again by centrifugation at 12000g 20 min. Then the cells were resuspended in equivalent volumes of ice-cold water+Protease Cocktail (Sigma) and incubated for 10 min on ice. After pelleting the cells at 16000g for 1h the supernatant was collected, adjusted to 50mM TRIS pH 8.0, 1mM $MgCl_{2,}$ and filter through a 0.22 μM cut-off membrane. The supernatant was loaded onto 5 ml of pre-packed HP Nickel-sepharose (GE Healthcare) at 3.0 ml/min. After washing with 40mM imidazole, 50mM Tris 8.0, and 500mM NaCl, the protein was eluted with a two-step gradient with 75 mM imidazole, 50mM Tris 8.0, 500mM NaCl and 300 mM imidazole, 50mM Tris 8.0, 500mM NaCl. The elution fractions containing EndoH were pooled and further purified by SEC on a Superdex 200 26/60 column (GE Healthcare).

Essentially the same procedure was followed for the expression of EndoH except that IPTG induction was performed at 20 hours at 20°C.

For expression of EndoD, 2 l of LB medium containing 50 pg/ml ampicillin were inoculated 1/30 with an overnight culture of *E. co*li BL21 (DE3) strain

transformed with pET28a (Novagen)-EndoD. After growing the cells at 37 °C to an OD550 of 1.0 the culture was cooled to room temperature. Expression was induced with 0.5 mM isopropyl- I-thio-P-D-galactoside (IPTG) and the culture was incubated for 20 h at 20 °C. The cells were harvested by centrifugation at 4000 g for 15 min and EndoD was purified from the cytoplasm of the bacteria. The pelleted cells were resuspended in 25ml 40 mM imidazole, 50 mM Tris 8.0, 500 mM NaCl buffer and homogenized by two passages through an Emulsiflex-C5 homogenizer (Avestin). The soluble fraction was separated by centrifugation at 35000 g for 30 min and filtered through 0.22 μM cut-off membrane. EndoD was purified from the supernatant applying the same protocol as for PNGase and EndoH.

Typically, deglycosylation of the native glycoproteins was carried out overnight (approximately 16 hours) at 37°C in 50 mM sodium citrate pH5.5 (Endo H and EndoD) or 50 mM Tris pH8. Initially, the amount of each endodeglycosidase used for deglycosylation of the target proteins was optimized in small-scale reactions at 37 ^{o}C. The glycoprotein:endodeglycosylase ratio at which maximal deglycosylation was achieved as based on SDS-PAGE analysis was chosen and the reaction was upscaled in a linear manner.

Removal of the Strep affinity tag

A C-terminal Strep tag preceded by an enterokinase recognition site was removed from the recombinant proteins by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). To cleave the tag from the recombinant proteins, typically 1 unit of EKMax enterokinase was added to a protein solution at 0.66 mg/ml in 50 mM Tris pH 8, 50 mM NaCl, 1 mM $CaCl_2$ and the reaction mixture was incubated at 37°C. After 16 h 0.1 mM phenylmethylsulphonyl fluoride (PMSF) was added in order to inactivate the protease. To remove any residual molecules harboring Strep tag, the reaction mixture was loaded on to a Streptactin column and the flow-through containing the protein molecules without the double Strep tag collected. Subsequently, the collected flow-through was loaded onto the Superdex 200 column (GE Healthcare, Uppsala, Sweden). The fractions containing purified protein without the Strep-tag were pooled and the protein was concentrated to approximately 25 mg/ml in 10 mM Tris pH 8 and 150 mM NaCl buffer.

SDS-PAGE and Western blot analysis

For electrophoresis under reducing and non-reducing conditions, samples in Laemmli sample buffer (125 mM Tris-HCl, 2% sodium dodecyl sulfate (SDS), 10% glycerol and 0.001% bromophenol blue, pH 6.8) with or without 100 mM DTT were denatured for 5 min at 98°C. Protein samples were subjected to SDS-PAGE electrophoresis using standard methods (Sambrook and Russell, 2000) and stained with Coomassie Blue (Bio-Rad) or subsequently transferred to PVDF membranes (Bio-Rad) by wet electrophoresis transfer (Bio-Rad).

The membrane was blocked for 1h at room temperature in PBS-T (PBS with the addition of 0.1% Tween 20) containing 5% dry milk. For detection of strep-tagged proteins, the membrane was first incubated for 1 h at room temperature with mouse-anti-Strep IgG (*Strep*MAB-Classic**,** IBA) for 1h at a dilution of 1:3000 in blocking buffer. After washing the membrane 3 times with PBS-T for 5 min, rabbit anti-mouse IgG conjugated with horseradish peroxidase was added at a dilution of 1:10000 in blocking buffer and the membrane was incubated for 1h. After washing the membrane three times for 15 min each time in PBS-T and once for 15 min in PBS, proteins were detected using the ECL kit (GE Healthcare).

DTNB-Thiols Assay

Free cysteine residues in proteins were detected using the DTNB-thiols assay which measures sulfhydryl groups with 5-5dithiobis[2nitrobenzoic acid] (DTNB) reagent forming a measurable yellow-colored product when it reacts with sulfhydryls. 750 µM DTNB in 100 mM Tris pH8.0 was used as a working solution. 5 µl of L-cysteine standards (500 µM, 750 µM, 1 mM 1.25 mM, 1.5 mM, 1.75 mM, 2 mM, 2.5 mM, 3 mM and 3.5 mM) were mixed with 160 µl working solution and 35 µl of buffer (10 mM Tris, 150 mM NaCl) in microtiter wells and incubated at room temperature for 10 min in order to obtain a standard curve. Similarly, 40 µl of diluted protein sample was mixed with 160 µl working solution and incubated at room temperature for 10 min. After 10 min absorbance of the solutions was measured at 412 nm. Absorbance values of L-cysteine standards were plotted versus concentration to generate a standard curve. The concentration of the sulfhydryl groups in protein samples was estimated from the standard curve.

Screening for crystallization conditions

To assess the appropriate protein concentration for crystallization screens, a pre-crystallization experiment using PCT Pre-Crystallization Test kit (Hampton Research) was performed for each new protein sample. Initial crystallization trials were carried out in 96-well crystallization plates at 293 K using the sitting drop-vapor diffusion method. Liquid nanoliter scale pipetting was performed using a Mosquito robot (TTP LabTech Ltd, Royston, UK). Sitting drops contained 200 nl protein solution and 200 nl reservoir solution. To identify initial crystallization conditions screening initially was performed using the commercially available crystallization screening kits listed below:

- Crystal screen 1 and 2 (Hampton Research, HR2-110 and HR2-112)
- Wizard I and II (Jena Biosciences, EBS-WIZF)
- JBS Screen 1-8 bulk (Jena Biosciences, CS-101L-108L)
- Structure Screen 1 and 2 (Molecular Dimensions, MD1-01 and MD1-02)
- PEG/Ion Screen (Hampton Research, HR2-126)
- Crystal screen CRYO (Hampton Research, HR2-122)

In total, 576 different reservoir solutions were tested during initial screening. Crystallization plates were placed and sitting drops were regularly imaged using a Rock Imager 1000 (Formulatrix, MA, USA). Crystallization conditions for SeFe as well as DAO5 Fab and DAO5 scFv in complex with the peptides are provided in the Results chapter.

Optimization of crystallization conditions

Multiple, small or poorly X-ray diffracting crystals obtained in 96 well plates during initial screening were further optimized using the hanging drop method in 24-well VDX plates (Hampton Research). Typically, drops of 2 µl (containing 1 µl of protein solution and 1 µl of reservoir solution) were placed on siliconized cover slides (Hampton Research) over 0.75 ml well solution. Optimization was performed around

the conditions that yielded crystals during initial screening. Optimization strategies included: 1) varying precipitant and/or other compounds concentration, 2) varying the pH, 3) screening with different protein concentrations, 4) performing crystallization at different temperatures, 5) using different compounds from Hampton Additive Screens 1–3 (Hampton Research) as additives to original crystallization condition (in 96 well plates), and 6) using seeding. Different methods of seeding are described in the following section.

Seeding techniques

It is believed that the optimal conditions for crystal nucleation can be different from the ones optimal for crystal growth (Kam *et al.*, 1978). Introduction of the seeds into crystallization drops provides a preformed crystal surface which may further promote the ordered assembly of molecules at a lower degree of supersaturation than needed for nucleation (Bergfors, 2003). Seeding approach has been demonstrated to be efficient in crystal optimization or promoting crystal growth (Korkhin *et al.*, 1996) (Stura, 1991).

In order to apply seeding techniques, seed stocks were prepared using the crystals from 4-6 crystallization drops. The crystals were transferred to a tube containing 50-100 µl of reservoir solution. Several 425-600 µm glass beads (Sigma) were added and the crystals were crushed by stirring for 5-10 min using a laboratory vortex. The seed-stocks were stored at +4°C. Three seeding techniques (direct seeding, streak seeding, and microseed matrix seeding) were applied.

1) In the case of direct seeding, serial 10-fold dilutions were prepared from the seed stock. The crystallization drops were set up by mixing 1 µl of reservoir solution, 0.5 µl of diluted seeds, and 1 µl of protein solution. The seed stock dilution yielding the best quality crystals was chosen for growing crystals for soaking or/and X-ray diffraction analysis.
2) Streak seeding was performed using a cat whisker on the drops pre-equilibrated for at least 8 h.
3) Microseed Matrix Screening (MMS) is a seeding technique where the seed of the crystals grown in one condition are seeded into new

conditions as part of the screening procedure (D'Arcy *et al.*, 2007). This technique has been successfully applied to promote crystallization of different macromolecules (Ireton & Stoddard, 2004) as well as antigen-antibody complexes (Obmolova *et al.*, 2010) . MMS can be further subdivided into self-seeding MMS (crystals of the same protein are used as seed-stock) and cross-seeding MMS (crystals of the related protein -- a mutant or a homologous protein -- are used as seed stock) (Walter *et al.*, 2008). Typically, MMS was performed by adding seeds during screening with the commercial screening kits described above. The drops consisted of 200 nl protein solution, 200 nl reservoir solution, and 100 nl of solution containing seeds. The specific application of seeding techniques is described in a chapter Results.

Cryo-protection and freezing of crystals

Cryo-protection of protein crystals and cooling during data collection is necessary to reduce radiation damage when crystals are irradiated with high intensity X-ray sources (reviewed in (Garman & Owen, 2006)). Good cryoprotectants prevent the formation of crystalline ice in the cooled crystals and results in vitrification of the water molecules. The cryo-protection method was optimized depending on the crystal growth conditions. 20% glycerol (v/v) in mother liquor was used for cryoprotection of a majority of the crystals. Crystals growing in conditions containing 20-30% PEG400, MPD or glycerol were frozen using a mother liquor as a cryo-protectant solution. Cryo-cooling procedures typically included three steps: 1) crystals were transferred from the crystallization drops to the drops containing cryo-protectant solution using a nylon cryoloop, 2) the crystals were kept in the cryoprotectant solution for several seconds, and 3) the crystals were fished using a cryoloop and flash frozen by plunging into liquid nitrogen (-196 °C).

References

Afdhal, N., Reddy, K. R., Nelson, D. R., Lawitz, E., Gordon, S. C., Schiff, E., Nahass, R., Ghalib, R., Gitlin, N., Herring, R., Lalezari, J., Younes, Z. H., Pockros, P. J., Di Bisceglie, A. M., Arora, S., Subramanian, G. M., Zhu, Y., Dvory-Sobol, H., Yang, J. C., Pang, P. S., Symonds, W. T., McHutchison, J. G., Muir, A. J., Sulkowski, M. & Kwo, P. (2014). Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection. *The New England journal of medicine* **370**, 1483-1493.

Aghemo, A., Degasperi, E. & Colombo, M. (2013). Directly acting antivirals for the treatment of chronic hepatitis C: unresolved topics from registration trials. *Dig Liver Dis* **45**, 1-7.

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. *Proc Natl Acad Sci U S A* **96**, 12766-12771.

Airenne, K. J., Hu, Y. C., Kost, T. A., Smith, R. H., Kotin, R. M., Ono, C., Matsuura, Y., Wang, S. & Yla-Herttuala, S. (2013). Baculovirus: an insect-derived vector for diverse gene transfer applications. *Mol Ther* **21**, 739-749.

Akazawa, D., Morikawa, K., Omi, N., Takahashi, H., Nakamura, N., Mochizuki, H., Date, T., Ishii, K., Suzuki, T. & Wakita, T. (2011). Production and characterization of HCV particles from serum-free culture. *Vaccine* **29**, 4821-4828.

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

Albecka, A., Belouzard, S., Op de Beeck, A., Descamps, V., Goueslain, L., Bertrand-Michel, J., Terce, F., Duverlie, G., Rouille, Y. & Dubuisson, J. (2012). Role of low-density lipoprotein receptor in the hepatitis C virus life cycle. *Hepatology* **55**, 998-1007.

Albertini, A. A., Merigoux, C., Libersou, S., Madiona, K., Bressanelli, S., Roche, S., Lepault, J., Melki, R., Vachette, P. & Gaudin, Y. (2012). Characterization of monomeric intermediates during VSV glycoprotein structural transition. *PLoS Pathog* **8**, e1002556.

Allander, T., Drakenberg, K., Beyene, 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 inhibit its interaction with CD81. *J Gen Virol* **81**, 2451-2459.

Allison, S. L., Schalich, J., Stiasny, K., Mandl, C. W., Kunz, C. & Heinz, F. X. (1995). Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J Virol* **69**, 695-700.

Almeida, J. D., Deinhardt, F., Holmes, A. W., Peterson, D. A., Wolfe, L. & Zuckerman, A. J. (1976). Morphology of the GB hepatitis agent. *Nature* **261**, 608-609.

Alter, H. J., Holland, P. V., Morrow, A. G., Purcell, R. H., Feinstone, S. M. & Moritsugu, Y. (1975a). Clinical and serological analysis of transfusion-associated hepatitis. *Lancet* **2**, 838-841.

Alter, H. J., Holland, P. V. & Purcell, R. H. (1975b). The emerging pattern of post-transfusion hepatitis. *Am J Med Sci* **270**, 329-334.

Alter, M. J. (1997). The epidemiology of acute and chronic hepatitis C. *Clin Liver Dis* **1**, 559-568, vi-vii.

Amako, Y., Tsukiyama-Kohara, K., Katsume, A., Hirata, Y., Sekiguchi, S., Tobita, Y., Hayashi, Y., Hishima, T., Funata, N., Yonekawa, H. & Kohara, M. (2010). Pathogenesis of hepatitis C virus infection in Tupaia belangeri. *J Virol* **84**, 303-311.

Anderson, R., King, A. D. & Innis, B. L. (1992). Correlation of E protein binding with cell susceptibility to dengue 4 virus infection. *J Gen Virol* **73 (Pt 8)**, 2155-2159.

Andersson, H., Barth, B. U., Ekstrom, M. & Garoff, H. (1997). Oligomerization-dependent folding of the membrane fusion protein of Semliki Forest virus. *J Virol* **71**, 9654-9663.

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

Au, S., Wu, W. & Pante, N. (2013). Baculovirus nuclear import: open, nuclear pore complex (NPC) sesame. *Viruses* **5**, 1885-1900.

Babcock, G. J., Iyer, S., Smith, H. L., Wang, Y., Rowley, K., Ambrosino, D. M., Zamore, P. D., Pierce, B. G., Molrine, D. C. & Weng, Z. (2014). High-throughput sequencing analysis of post-liver transplantation HCV e2 glycoprotein evolution in the presence and absence of neutralizing monoclonal antibody. *PLoS One* **9**, e100325.

Backovic, M. & Jardetzky, T. S. (2011). Class III viral membrane fusion proteins. *Adv Exp Med Biol* **714**, 91-101.

Backovic, M., Johansson, D. X., Klupp, B. G., Mettenleiter, T. C., Persson, M. A. & Rey, F. A. (2010). Efficient method for production of high yields of Fab fragments in Drosophila S2 cells. *Protein Eng Des Sel* **23**, 169-174.

Backovic, M., Longnecker, R. & Jardetzky, T. S. (2009). Structure of a trimeric variant of the Epstein-Barr virus glycoprotein B. *Proc Natl Acad Sci U S A* **106**, 2880-2885.

Baker, K. A., Dutch, R. E., Lamb, R. A. & Jardetzky, T. S. (1999). Structural basis for paramyxovirus-mediated membrane fusion. *Mol Cell* **3**, 309-319.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol Rev* **35**, 235-241.

Bankwitz, D., Steinmann, E., Bitzegeio, J., Ciesek, S., Friesland, M., Herrmann, E., Zeisel, M. B., Baumert, T. F., Keck, Z. Y., Foung, S. K., Pecheur, E. I. & Pietschmann, T. (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.

Baquero, E., Albertini, A. A., Vachette, P., Lepault, J., Bressanelli, S. & Gaudin, Y. (2013). Intermediate conformations during viral fusion glycoprotein structural transition. *Curr Opin Virol* **3**, 143-150.

Barretto, N., Sainz, B., Jr., Hussain, S. & Uprichard, S. L. (2014). Determining the involvement and therapeutic implications of host cellular factors in hepatitis C virus cell-to-cell spread. *J Virol* **88**, 5050-5061.

Bartenschlager, R., Ahlborn-Laake, L., Mous, J. & Jacobsen, H. (1993). Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J Virol* **67**, 3835-3844.

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

Barth, H., Schafer, C., Adah, M. I., Zhang, F., Linhardt, R. J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T. H., Depla, E., Von Weizsacker, F., Blum, H. E. & Baumert, T. F. (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. & Baumert, T. F. (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. *Proc Natl Acad Sci U S A* **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., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A. & Cosset, F. L. (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.

Beames, B., Chavez, D., Guerra, B., Notvall, L., Brasky, K. M. & Lanford, R. E. (2000). Development of a primary tamarin hepatocyte culture system for GB virus-B: a surrogate model for hepatitis C virus. *J Virol* **74**, 11764-11772.

Becnel, J., White, S., Moser, B., Fukuda, T., Rotstein, M., Undeen, A. & Cockburn, A. (2001). Epizootiology and transmission of a newly discovered baculovirus from the mosquitoes Culex nigripalpus and C. quinquefasciatus. *J Gen Virol* **82**, 275-282.

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

Bergfors, T. (2003). Seeds to crystals. *J Struct Biol* **142**, 66-76.

Bianchi, F. J. J. A., Vlak, J. M. & van der Werf, W. (2002). Evaluation of the control of beet armyworm, Spodoptera exigua, with baculoviruses in greenhouses using a process-based simulation model. *Biol Control* **24**, 277-284.

Bjoro, K., Froland, S. S., Yun, Z., Samdal, H. H. & Haaland, T. (1994). Hepatitis C infection in patients with primary hypogammaglobulinemia after treatment

with contaminated immune globulin. *The New England journal of medicine* **331**, 1607-1611.

Blight, K. J. (2011). Charged residues in hepatitis C virus NS4B are critical for multiple NS4B functions in RNA replication. *J Virol* **85**, 8158-8171.

Blissard, G. W. & Rohrmann, G. F. (1989). Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *Virology* **170**, 537-555.

Blissard, G. W. & Wenz, J. R. (1992). Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *J Virol* **66**, 6829-6835.

Bossart, K. N., Fusco, D. L. & Broder, C. C. (2013). Paramyxovirus entry. *Adv Exp Med Biol* **790**, 95-127.

Botti, V., Bianchi, A., Foung, S. K. & Merola, M. (2011). The hepatitis C virus E1 glycoprotein undergoes productive folding but accelerated degradation when expressed as an individual subunit in CHO cells. *PLoS One* **6**, e23838.

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

Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A* **83**, 3746-3750.

Bricogne G, B. E., Brandl M, Flensburg C, Keller P, et al. (2010). BUSTER version 2.9. *Cambridge (United Kingdom): Global Phasing Ltd*.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* **59**, 2023-2030.

Bright, H., Carroll, A. R., Watts, P. A. & Fenton, R. J. (2004). Development of a GB virus B marmoset model and its validation with a novel series of hepatitis C virus NS3 protease inhibitors. *J Virol* **78**, 2062-2071.

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., Tellinghuisen, T. L., Wong-Staal, F., Balfe, P. & McKeating, J. A. (2011). Neutralizing antibody-resistant hepatitis C virus cell-to-cell transmission. *J Virol* **85**, 596-605.

Broering, T. J., Garrity, K. A., Boatright, N. K., Sloan, S. E., Sandor, F., Thomas, W. D., Jr., Szabo, G., Finberg, R. W., Ambrosino, D. M. & Babcock, G. J. (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., 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., Tarr, A. W., McClure, C. P., Juttla, V. S., Tagiuri, 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.

Brown, R. S. (2005). Hepatitis C and liver transplantation. *Nature* **436**, 973-978.

Bruccoleri, R. E., Novotny, J., Keck, P. & Cohen, C. (1986). Two-stranded alpha-helical coiled-coils of fibrous proteins: theoretical analysis of supercoil formation. *Biophys J* **49**, 79-81.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921.

Bukh, J. (2004). A critical role for the chimpanzee model in the study of hepatitis C. *Hepatology* **39**, 1469-1475.

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

Bukh, J., Apgar, C. L. & Yanagi, M. (1999). Toward a surrogate model for hepatitis C virus: An infectious molecular clone of the GB virus-B hepatitis agent. *Virology* **262**, 470-478.

Bukh, J., Engle, R. E., Govindarajan, S. & Purcell, R. H. (2008). Immunity against the GBV-B hepatitis virus in tamarins can prevent productive infection following rechallenge and is long-lived. *J Med Virol* **80**, 87-94.

Bukh, J., Miller, R. H. & Purcell, R. H. (1995). Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Semin Liver Dis* **15**, 41-63.

Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37-43.

Burbelo, P. D., Dubovi, E. J., Simmonds, P., Medina, J. L., Henriquez, J. A., Mishra, N., Wagner, J., Tokarz, R., Cullen, J. M., Iadarola, M. J., Rice, C. M., Lipkin, W. I. & Kapoor, A. (2012). Serology-enabled discovery of genetically diverse hepaciviruses in a new host. *J Virol* **86**, 6171-6178.

Butkiewicz, N., Yao, N., Zhong, W., Wright-Minogue, J., Ingravallo, P., Zhang, R., Durkin, J., Standring, D. N., Baroudy, B. M., Sangar, D. V., Lemon, S. M., Lau, J. Y. & Hong, Z. (2000). Virus-specific cofactor requirement and chimeric hepatitis C virus/GB virus B nonstructural protein 3. *J Virol* **74**, 4291-4301.

Callens, N., Ciczora, Y., Bartosch, B., Vu-Dac, N., Cosset, F. L., Pawlotsky, J. M., Penin, F. & Dubuisson, J. (2005). Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein e2 contribute to virus entry. *J Virol* **79**, 15331-15341.

Carloni, G., Iacovacci, S., Sargiacomo, M., Ravagnan, G., Ponzetto, A., Peschle, C. & Battaglia, M. (1993). Susceptibility of human liver cell cultures to hepatitis C virus infection. *Arch Virol Suppl* **8**, 31-39.

Castelli, M., Clementi, N., Sautto, G. A., Pfaff, J., Kahle, K. M., Barnes, T., Doranz, B. J., Dal Peraro, M., Clementi, M., Burioni, R. & Mancini, N. (2014). HCV E2 core structures and mAbs: something is still missing. *Drug Discov Today*.

Catanese, M. T., Graziani, R., von Hahn, T., Moreau, M., Huby, T., Paonessa, G., Santini, C., Luzzago, A., Rice, C. M., Cortese, R., Vitelli, A. & Nicosia, A. (2007). High-avidity monoclonal antibodies against the human scavenger class B type I receptor efficiently block hepatitis C virus infection in the presence of high-density lipoprotein. *J Virol* **81**, 8063-8071.

Catanese, M. T., Uryu, K., Kopp, M., Edwards, T. J., Andrus, L., Rice, W. J., Silvestry, M., Kuhn, R. J. & Rice, C. M. (2013). Ultrastructural analysis of hepatitis C virus particles. *Proc Natl Acad Sci U S A* **110**, 9505-9510.

Chambers, P., Pringle, C. R. & Easton, A. J. (1990). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J Gen Virol* **71 (Pt 12)**, 3075-3080.

Chang, A. & Dutch, R. E. (2012). Paramyxovirus fusion and entry: multiple paths to a common end. *Viruses* **4**, 613-636.

Chang, K. S., Jiang, J., Cai, Z. & Luo, G. (2007). Human apolipoprotein e is required for infectivity and production of hepatitis C virus in cell culture. *J Virol* **81**, 13783-13793.

Chang, M. H., Gordon, L. A. & Fung, H. B. (2012). Boceprevir: a protease inhibitor for the treatment of hepatitis C. *Clin Ther* **34**, 2021-2038.

Chen, J., Skehel, J. J. & Wiley, D. C. (1999a). N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc Natl Acad Sci U S A* **96**, 8967-8972.

Chen, L., Gorman, J. J., McKimm-Breschkin, J., Lawrence, L. J., Tulloch, P. A., Smith, B. J., Colman, P. M. & Lawrence, M. C. (2001). The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion. *Structure* **9**, 255-266.

Chen, M., Sallberg, M., Sonnerborg, A., Weiland, O., Mattsson, L., Jin, L., Birkett, A., Peterson, D. & Milich, D. R. (1999b). Limited humoral immunity in hepatitis C virus infection. *Gastroenterology* **116**, 135-143.

Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J. & Marks, R. M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* **3**, 866-871.

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

Chernomordik, L. V. & Kozlov, M. M. (2003). Protein-lipid interplay in fusion and fission of biological membranes. *Annu Rev Biochem* **72**, 175-207.

Chernomordik, L. V. & Kozlov, M. M. (2008). Mechanics of membrane fusion. *Nat Struct Mol Biol* **15**, 675-683.

Chiu, J., Tillett, D., Dawes, I. W. & March, P. E. (2008). Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *J Microbiol Methods* **73**, 195-198.

Choo, Q. L., Kuo, G., Ralston, R., Weiner, A., Chien, D., Van Nest, G., Han, J., Berger, K., Thudium, K., Kuo, C. & et al. (1994). Vaccination of chimpanzees against infection by the hepatitis C virus. *Proc Natl Acad Sci U S A* **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, R. T., Gordon, F. D., Curry, M. P., Schiano, T. D., Emre, S., Corey, K., Markmann, J. F., Hertl, M., Pomposelli, J. J., Pomfret, E. A., Florman, S., Schilsky, M., Broering, T. J., Finberg, R. W., Szabo, G., Zamore, P. D., Khettry, U., Babcock, G. J., Ambrosino, D. M., Leav, B., Leney, M., Smith, H. L. & Molrine, D. C. (2013). Human monoclonal antibody MBL-HCV1 delays HCV viral rebound following liver transplantation: a randomized controlled study. *Am J Transplant* **13**, 1047-1054.

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

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

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

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

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

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

Connolly, S. A., Jackson, J. O., Jardetzky, T. S. & Longnecker, R. (2011). Fusing structure and function: a structural view of the herpesvirus entry machinery. *Nat Rev Microbiol* **9**, 369-381.

Cormier, E. G., Durso, R. J., Tsamis, F., Boussemart, L., Manix, C., Olson, W. C., Gardner, J. P. & Dragic, T. (2004a). L-SIGN (CD209L) and DC-SIGN (CD209) mediate transinfection of liver cells by hepatitis C virus. *Proc Natl Acad Sci U S A* **101**, 14067-14072.

Cormier, E. G., Tsamis, F., Kajumo, F., Durso, R. J., Gardner, J. P. & Dragic, T. (2004b). CD81 is an entry coreceptor for hepatitis C virus. *Proc Natl Acad Sci U S A* **101**, 7270-7274.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-10890.

Corti, D. & Lanzavecchia, A. (2013). Broadly neutralizing antiviral antibodies. *Annu Rev Immunol* **31**, 705-742.

Coulibaly, F., Chiu, E., Gutmann, S., Rajendran, C., Haebel, P. W., Ikeda, K., Mori, H., Ward, V. K., Schulze-Briese, C. & Metcalf, P. (2009). The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. *Proc Natl Acad Sci U S A* **106**, 22205-22210.

Crespo, G., Marino, Z., Navasa, M. & Forns, X. (2012). Viral hepatitis in liver transplantation. *Gastroenterology* **142**, 1373-1383 e1371.

Cuevas, J. M., Gonzalez-Candelas, F., Moya, A. & Sanjuan, R. (2009). Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. *J Virol* **83**, 5760-5764.

D'Arcy, A., Villard, F. & Marsh, M. (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr* **63**, 550-554.

Danieli, T., Pelletier, S. L., Henis, Y. I. & White, J. M. (1996). Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. *J Cell Biol* **133**, 559-569.

Dao Thi, V. L., Granier, C., Zeisel, M. B., Guerin, M., Mancip, J., Granio, O., Penin, F., Lavillette, D., Bartenschlager, R., Baumert, T. F., Cosset, F. L. & Dreux, M. (2012). Characterization of hepatitis C virus particle subpopulations reveals multiple usage of the scavenger receptor BI for entry steps. *J Biol Chem* **287**, 31242-31257.

Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615-619.

De Clercq, E. (2014). Current race in the development of DAAs (direct-acting antivirals) against HCV. *Biochem Pharmacol* **89**, 441-452.

Decaens, C., Durand, M., Grosse, B. & Cassio, D. (2008). Which in vitro models could be best used to study hepatocyte polarity? *Biol Cell* **100**, 387-398.

Deinhardt, F., Holmes, A. W., Capps, R. B. & Popper, H. (1967). Studies on the transmission of human viral hepatitis to marmoset monkeys. I. Transmission of disease, serial passages, and description of liver lesions. *J Exp Med* **125**, 673-688.

Deinhardt, F., Peterson, D., Cross, G., Wolfe, L. & Holmes, A. W. (1975). Hepatitis in marmosets. *The American journal of the medical sciences* **270**, 73-80.

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.

Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965-968.

Deng, L., Ma, L., Virata-Theimer, M. L., Zhong, L., Yan, H., Zhao, Z., Struble, E., Feinstone, S., Alter, H. & Zhang, P. (2014). Discrete conformations of epitope II on the hepatitis C virus E2 protein for antibody-mediated neutralization and nonneutralization. *Proc Natl Acad Sci U S A*.

Deng, L., Zhong, L., Struble, E., Duan, H., Ma, L., Harman, C., Yan, H., Virata-Theimer, M. L., Zhao, Z., Feinstone, S., Alter, H. & Zhang, P. (2013). Structural evidence for a bifurcated mode of action in the antibody-mediated neutralization of hepatitis C virus. *Proc Natl Acad Sci U S A* **110**, 7418-7422.

Dessau, M. & Modis, Y. (2013). Crystal structure of glycoprotein C from Rift Valley fever virus. *Proc Natl Acad Sci U S A* **110**, 1696-1701.

Diamond, M. S., Pierson, T. C. & Fremont, D. H. (2008). The structural immunology of antibody protection against West Nile virus. *Immunol Rev* **225**, 212-225.

Diaz, O., Delers, F., Maynard, M., Demignot, S., Zoulim, F., Chambaz, J., Trepo, C., Lotteau, V. & Andre, P. (2006). Preferential association of Hepatitis C virus with apolipoprotein B48-containing lipoproteins. *J Gen Virol* **87**, 2983-2991.

Doms, R. W. & Moore, J. P. (2000). HIV-1 membrane fusion: targets of opportunity. *J Cell Biol* **151**, F9-14.

Donald, J. E., Zhang, Y., Fiorin, G., Carnevale, V., Slochower, D. R., Gai, F., Klein, M. L. & DeGrado, W. F. (2011). Transmembrane orientation and

possible role of the fusogenic peptide from parainfluenza virus 5 (PIV5) in promoting fusion. *Proc Natl Acad Sci U S A* **108**, 3958-3963.

Dong, S., Wang, M., Qiu, Z., Deng, F., Vlak, J. M., Hu, Z. & Wang, H. (2010). Autographa californica multicapsid nucleopolyhedrovirus efficiently infects Sf9 cells and transduces mammalian cells via direct fusion with the plasma membrane at low pH. *J Virol* **84**, 5351-5359.

Dorner, M., Horwitz, J. A., Donovan, B. M., Labitt, R. N., Budell, W. C., Friling, T., Vogt, A., Catanese, M. T., Satoh, T., Kawai, T., Akira, S., Law, M., Rice, C. M. & Ploss, A. (2013). Completion of the entire hepatitis C virus life cycle in genetically humanized mice. *Nature* **501**, 237-241.

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

Dorner, M. & Ploss, A. (2011). Deconstructing hepatitis C virus infection in humanized mice. *Ann N Y Acad Sci* **1245**, 59-62.

Dowd, K. A., Jost, C. A., Durbin, A. P., Whitehead, S. S. & Pierson, T. C. (2011). A dynamic landscape for antibody binding modulates antibody-mediated neutralization of West Nile virus. *PLoS Pathog* **7**, e1002111.

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., Pietschmann, T., Granier, C., Voisset, C., Ricard-Blum, S., Mangeot, P. E., Keck, Z., Foung, S., Vu-Dac, N., Dubuisson, J., Bartenschlager, R., Lavillette, D. & Cosset, F. L. (2006). High density lipoprotein inhibits hepatitis C virus-neutralizing antibodies by stimulating cell entry via activation of the scavenger receptor BI. *J Biol Chem* **281**, 18285-18295.

Drexler, J. F., Corman, V. M., Muller, M. A., Lukashev, A. N., Gmyl, A., Coutard, B., Adam, A., Ritz, D., Leijten, L. M., van Riel, D., Kallies, R., Klose, S. M., Gloza-Rausch, F., Binger, T., Annan, A., Adu-Sarkodie, Y., Oppong, S., Bourgarel, M., Rupp, D., Hoffmann, B., Schlegel, M., Kummerer, B. M., Kruger, D. H., Schmidt-Chanasit, J., Setien, A. A., Cottontail, V. M., Hemachudha, T., Wacharapluesadee, S., Osterrieder, K., Bartenschlager, R., Matthee, S., Beer, M., Kuiken, T., Reusken, C., Leroy, E. M., Ulrich, R. G. & Drosten, C. (2013). Evidence for novel hepaciviruses in rodents. *PLoS Pathog* **9**, e1003438.

Drucker, E., Alcabes, P. G. & Marx, P. A. (2001). The injection century: massive unsterile injections and the emergence of human pathogens. *Lancet* **358**, 1989-1992.

Drummer, H. E., Boo, I., Maerz, A. L. & Poumbourios, P. (2006). A conserved Gly436-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., Boo, I. & Poumbourios, P. (2007). Mutagenesis of a conserved fusion peptide-like motif and membrane-proximal heptad-repeat region of hepatitis C virus glycoprotein E1. *J Gen Virol* **88**, 1144-1148.

Drummer, H. E., Maerz, A. & Poumbourios, P. (2003). Cell surface expression of functional hepatitis C virus E1 and E2 glycoproteins. *FEBS Lett* **546**, 385-390.

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.

Drummer, H. E., Wilson, K. A. & Poumbourios, P. (2002). Identification of the hepatitis C virus E2 glycoprotein binding site on the large extracellular loop of CD81. *J Virol* **76**, 11143-11147.

Duan, H., Kachko, A., Zhong, L., Struble, E., Pandey, S., Yan, H., Harman, C., Virata-Theimer, M. L., Deng, L., Zhao, Z., Major, M., Feinstone, S. & Zhang, P. (2012). Amino acid residue-specific neutralization and nonneutralization of hepatitis C virus by monoclonal antibodies to the E2 protein. *J Virol* **86**, 12686-12694.

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

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.

El Omari, K., Iourin, O., Harlos, K., Grimes, J. M. & Stuart, D. I. (2013). Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. *Cell Rep* **3**, 30-35.

El Omari, K., Iourin, O., Kadlec, J., Sutton, G., Harlos, K., Grimes, J. M. & Stuart, D. I. (2014). Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. *Nat Commun* **5**, 4874.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.

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, G. & Wilson, K. S. (1999). A MAD experiment performed at the white line of the iridium LIII absorption edge in lysozyme. *Acta Crystallogr D Biol Crystallogr* **55**, 67-76.

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

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.

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., Shimoda, A., Govindarajan, S., Cheung, L. C., Melpolder, J. C., Sacher, R. A., Shih, J. W. & Purcell, R. H. (1996a). Hepatitis C virus-associated fulminant hepatic failure. *The New England journal of medicine* **335**, 631-634.

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. *Proc Natl Acad Sci U S A* **91**, 7792-7796.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J. C., Strazzera, A., Chien, D. Y., Munoz, S. J., Balestrieri, A., Purcell, R. H. & Alter, H. J. (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., Shapiro, M., Alter, H. J. & Purcell, R. H. (1996b). Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc Natl Acad Sci U S A* **93**, 15394-15399.

Farquhar, M. J., Harris, H. J., Diskar, M., Jones, S., Mee, C. J., Nielsen, S. U., Brimacombe, C. L., Molina, S., Toms, G. L., Maurel, P., Howl, J., Herberg, F. W., van Ijzendoorn, S. C., Balfe, P. & McKeating, J. A. (2008). Protein kinase A-dependent step(s) in hepatitis C virus entry and infectivity. *J Virol* **82**, 8797-8811.

Federici, B. A. (1980). Mosquito baculovirus: sequence of morphogenesis and ultrastructure of the virion. *Virology* **100**, 1-9.

Feinstone, S. M., Kapikian, A. Z., Purcell, R. H., Alter, H. J. & Holland, P. V. (1975). Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N Engl J Med* **292**, 767-770.

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.

Ferenci, P. & Reddy, K. R. (2011). Impact of HCV protease-inhibitor-based triple therapy for chronic HCV genotype 1 infection. *Antivir Ther* **16**, 1187-1201.

Flint, M., Maidens, C., Loomis-Price, L. D., Shotton, C., Dubuisson, J., Monk, P., Higginbottom, A., Levy, S. & McKeating, J. A. (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., Xiao, F., Thumann, C., Turek, M., Zona, L., Tawar, R. G., Grunert, F., Thompson, J., Zeisel, M. B. & Baumert, T. F. (2013). A novel monoclonal anti-CD81 antibody produced by genetic immunization efficiently inhibits Hepatitis C virus cell-cell transmission. *PLoS One* **8**, e64221.

Forestier, N. & Zeuzem, S. (2012). Telaprevir for the treatment of hepatitis C. *Expert Opin Pharmacother* **13**, 593-606.

Forns, X., Thimme, R., Govindarajan, S., Emerson, S. U., Purcell, R. H., Chisari, F. V. & Bukh, J. (2000). Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees. *Proc Natl Acad Sci U S A* **97**, 13318-13323.

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res* **117**, 5-16.

Fournier, C., Sureau, C., Coste, J., Ducos, J., Pageaux, G., Larrey, D., Domergue, J. & Maurel, P. (1998). In vitro infection of adult normal human

hepatocytes in primary culture by hepatitis C virus. *J Gen Virol* **79 (Pt 10)**, 2367-2374.

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.

Frame, I. G., Cutfield, J. F. & Poulter, R. T. (2001). New BEL-like LTR-retrotransposons in Fugu rubripes, Caenorhabditis elegans, and Drosophila melanogaster. *Gene* **263**, 219-230.

Fraser, M. J., Brusca, J. S., Smith, G. E. & Summers, M. D. (1985). Transposon-mediated mutagenesis of a baculovirus. *Virology* **145**, 356-361.

Frey, S. E., Houghton, M., Coates, S., Abrignani, S., Chien, D., Rosa, D., Pileri, P., Ray, R., Di Bisceglie, A. M., Rinella, P., Hill, H., Wolff, M. C., Schultze, V., Han, J. H., Scharschmidt, B. & Belshe, R. B. (2010). Safety and immunogenicity of HCV E1E2 vaccine adjuvanted with MF59 administered to healthy adults. *Vaccine* **28**, 6367-6373.

Fried, M. W., Shiffman, M. L., Reddy, K. R., Smith, C., Marinos, G., Goncales, F. L., Jr., Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J. & Yu, J. (2002). Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *The New England journal of medicine* **347**, 975-982.

Friesen, P. D. & Nissen, M. S. (1990). Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. *Mol Cell Biol* **10**, 3067-3077.

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

Gallegos-Orozco, J. F., Yosephy, A., Noble, B., Aqel, B. A., Byrne, T. J., Carey, E. J., Douglas, D. D., Mulligan, D., Moss, A., de Petris, G., Williams, J. W., Rakela, J. & Vargas, H. E. (2009). Natural history of post-liver transplantation hepatitis C: A review of factors that may influence its course. *Liver Transpl* **15**, 1872-1881.

Gallivan, J. P. & Dougherty, D. A. (1999). Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* **96**, 9459-9464.

Gardner, J. P., Durso, R. J., Arrigale, R. R., Donovan, G. P., Maddon, P. J., Dragic, T. & Olson, W. C. (2003). L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. *Proc Natl Acad Sci U S A* **100**, 4498-4503.

Garman, E. F. & Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **62**, 32-47.

Garrone, P., Fluckiger, A. C., Mangeot, P. E., Gauthier, E., Dupeyrot-Lacas, P., Mancip, J., Cangialosi, A., Du Chene, I., LeGrand, R., Mangeot, I., Lavillette, D., Bellier, B., Cosset, F. L., Tangy, F., Klatzmann, D. & Dalba, C. (2011). A prime-boost strategy using virus-like particles pseudotyped for HCV proteins triggers broadly neutralizing antibodies in macaques. *Sci Transl Med* **3**, 94ra71.

Garry, R. F. & Dash, S. (2003). Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins. *Virology* **307**, 255-265.

Gaudin, Y., Tuffereau, C., Segretain, D., Knossow, M. & Flamand, A. (1991). Reversible conformational changes and fusion activity of rabies virus glycoprotein. *J Virol* **65**, 4853-4859.

Gerlach, J. T., Diepolder, H. M., Zachoval, R., Gruener, N. H., Jung, M. C., Ulsenheimer, A., Schraut, W. W., Schirren, C. A., Waechtler, M., Backmund, M. & Pape, G. R. (2003). Acute hepatitis C: high rate of both spontaneous and treatment-induced viral clearance. *Gastroenterology* **125**, 80-88.

Ghibaudo, D., Cohen, L., Penin, F. & Martin, A. (2004). Characterization of GB virus B polyprotein processing reveals the existence of a novel 13-kDa protein with partial homology to hepatitis C virus p7 protein. *J Biol Chem* **279**, 24965-24975.

Giang, E., Dorner, M., Prentoe, J. C., Dreux, M., Evans, M. J., Bukh, J., Rice, C. M., Ploss, A., Burton, D. R. & Law, M. (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., 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.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163-164.

Glue, P., Rouzier-Panis, R., Raffanel, C., Sabo, R., Gupta, S. K., Salfi, M., Jacobs, S. & Clement, R. P. (2000). A dose-ranging study of pegylated interferon alfa-2b and ribavirin in chronic hepatitis C. The Hepatitis C Intervention Therapy Group. *Hepatology* **32**, 647-653.

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.

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

Gottwein, J. M., Jensen, T. B., Mathiesen, C. K., Meuleman, P., Serre, S. B., Lademann, J. B., Ghanem, L., Scheel, T. K., Leroux-Roels, G. & Bukh, J. (2011). Development and application of hepatitis C reporter viruses with genotype 1 to 7 core-nonstructural protein 2 (NS2) expressing fluorescent proteins or luciferase in modified JFH1 NS5A. *J Virol* **85**, 8913-8928.

Gottwein, J. M., Scheel, T. K., Hoegh, A. M., Lademann, J. B., Eugen-Olsen, J., Lisby, G. & Bukh, J. (2007). Robust hepatitis C genotype 3a cell culture releasing adapted intergenotypic 3a/2a (S52/JFH1) viruses. *Gastroenterology* **133**, 1614-1626.

Gottwein, J. M., Scheel, T. K., 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.

Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, 3320-3323.

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

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A. & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327-332.

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. & Marsh, M. (2011). The cell biology of receptor-mediated virus entry. *J Cell Biol* **195**, 1071-1082.

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.

Hadlock, K. G., Lanford, R. E., Perkins, S., Rowe, J., Yang, Q., Levy, S., Pileri, P., Abrignani, S. & Foung, S. K. (2000). Human monoclonal antibodies that inhibit binding of hepatitis C virus E2 protein to CD81 and recognize conserved conformational epitopes. *J Virol* **74**, 10407-10416.

Hadziyannis, S. J., Sette, H., Jr., Morgan, T. R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H., Jr., Bernstein, D., Rizzetto, M., Zeuzem, S., Pockros, P. J., Lin, A. & Ackrill, A. M. (2004). Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* **140**, 346-355.

Hahm, B., Han, D. S., Back, S. H., Song, O. K., Cho, M. J., Kim, C. J., Shimotohno, K. & Jang, S. K. (1995). NS3-4A of hepatitis C virus is a chymotrypsin-like protease. *J Virol* **69**, 2534-2539.

Harris, H. J., Clerte, C., Farquhar, M. J., Goodall, M., Hu, K., Rassam, P., Dosset, P., Wilson, G. K., Balfe, P., Ijzendoorn, S. C., Milhiet, P. E. & McKeating, J. A. (2013). Hepatoma polarization limits CD81 and hepatitis C virus dynamics. *Cell Microbiol* **15**, 430-445.

Harris, H. J., Davis, C., Mullins, J. G., Hu, K., Goodall, M., Farquhar, M. J., Mee, C. J., McCaffrey, K., Young, S., Drummer, H., Balfe, P. & McKeating, J. A. (2010). Claudin association with CD81 defines hepatitis C virus entry. *J Biol Chem* **285**, 21092-21102.

Harrison, S. C. (2008). Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698.

Heckman, K. L. & Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**, 924-932.

Hefferon, K. L., Oomens, A. G., Monsma, S. A., Finnerty, C. M. & Blissard, G. W. (1999). Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. *Virology* **258**, 455-468.

Heldwein, E. E., Lou, H., Bender, F. C., Cohen, G. H., Eisenberg, R. J. & Harrison, S. C. (2006). Crystal structure of glycoprotein B from herpes simplex virus 1. *Science* **313**, 217-220.

Helle, F., Goffard, A., Morel, V., Duverlie, G., McKeating, J., Keck, Z. Y., Foung, S., Penin, F., Dubuisson, J. & Voisset, C. (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. & Dubuisson, J. (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., Lee, S. M., Bartosch, B., Cosset, F. L. & Kang, C. Y. (2006). Hepatitis C virus E2 links soluble human CD81 and SR-B1 protein. *Virus Res* **121**, 58-64.

Higginbottom, A., Quinn, E. R., Kuo, C. C., Flint, M., Wilson, L. H., Bianchi, E., Nicosia, A., Monk, P. N., McKeating, J. A. & Levy, S. (2000). Identification of amino acid residues in CD81 critical for interaction with hepatitis C virus envelope glycoprotein E2. *J Virol* **74**, 3642-3649.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S. & Shimotohno, K. (1991). Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochem Biophys Res Commun* **175**, 220-228.

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

Hijikata, M., Shimizu, Y. K., Kato, H., Iwamoto, A., Shih, J. W., Alter, H. J., Purcell, R. H. & Yoshikura, H. (1993b). Equilibrium centrifugation studies of hepatitis C virus: evidence for circulating immune complexes. *J Virol* **67**, 1953-1958.

Hiller, N., Fritz-Wolf, K., Deponte, M., Wende, W., Zimmermann, H. & Becker, K. (2006). Plasmodium falciparum glutathione S-transferase--structural and mechanistic studies on ligand binding and enzyme inhibition. *Protein Sci* **15**, 281-289.

Holland, L. Z., Albalat, R., Azumi, K., Benito-Gutierrez, E., Blow, M. J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L. J., Ferrier, D. E., Garcia-Fernandez, J., Gibson-Brown, J. J., Gissi, C., Godzik, A., Hallbook, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G. W., McHardy, A. C., Meulemans, D., Nonaka, M., Olinski, R. P., Pancer, Z., Pennacchio, L. A., Pestarino, M., Rast, J. P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J. K., Zhang, Q., Zmasek, C. M., de Jong, P. J., Osoegawa, K., Putnam, N. H., Rokhsar, D. S., Satoh, N. & Holland, P. W. (2008). The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* **18**, 1100-1111.

Holm, L. & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549.

Holmes, E. C. (2003). Molecular clocks and the puzzle of RNA virus origins. *J Virol* **77**, 3893-3897.

Houghton, M. & Abrignani, S. (2005). Prospects for a vaccine against the hepatitis C virus. *Nature* **436**, 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. *Proc Natl Acad Sci U S A* **100**, 7271-7276.

Hsu, S. H., Yeh, M. L. & Wang, S. N. (2013). New insights in recurrent HCV infection after liver transplantation. *Clin Dev Immunol* **2013**, 890517.

Huang, H., Sun, F., Owen, D. M., Li, W., Chen, Y., Gale, M., Jr. & Ye, J. (2007). Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins. *Proc Natl Acad Sci U S A* **104**, 5848-5853.

Iacovacci, S., Sargiacomo, M., Parolini, I., Ponzetto, A., Peschle, C. & Carloni, G. (1993). Replication and multiplication of hepatitis C virus genome in human foetal liver cells. *Res Virol* **144**, 275-279.

Icard, V., Diaz, O., Scholtes, C., Perrin-Cocon, L., Ramiere, C., Bartenschlager, R., Penin, F., Lotteau, V. & Andre, P. (2009). Secretion of hepatitis C virus envelope glycoproteins depends on assembly of apolipoprotein B positive lipoproteins. *PLoS One* **4**, e4233.

Igonet, S., Vaney, M. C., Vonrhein, C., Bricogne, G., Stura, E. A., Hengartner, H., Eschli, B. & Rey, F. A. (2011). X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proc Natl Acad Sci U S A* **108**, 19967-19972.

Iizuka, M., Chiba, M., Masamune, O., Kaga, E., Nakagomi, T. & Nakagomi, O. (1994). A highly conserved genomic RNA constellation of Japanese isolates of human rotaviruses carrying G serotype 3 and P serotype 9. *Res Virol* **145**, 21-24.

Imai, Y., Matsushima, Y., Sugimura, T. & Terada, M. (1991). A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res* **19**, 2785.

Ireton, G. C. & Stoddard, B. L. (2004). Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr* **60**, 601-605.

Iwasaki, Y., Mori, K., Ishii, K., Maki, N., Iijima, S., Yoshida, T., Okabayashi, S., Katakai, Y., Lee, Y. J., Saito, A., Fukai, H., Kimura, N., Ageyama, N., Yoshizaki, S., Suzuki, T., Yasutomi, Y., Miyamura, T., Kannagi, M. & Akari, H. (2011). Long-Term Persistent GBV-B Infection and Development of a Chronic and Progressive Hepatitis C-Like Disease in Marmosets. *Front Microbiol* **2**, 240.

Jacob, J. R., Lin, K. C., Tennant, B. C. & Mansfield, K. G. (2004). GB virus B infection of the common marmoset (Callithrix jacchus) and associated liver pathology. *J Gen Virol* **85**, 2525-2533.

Jardetzky, T. S. & Lamb, R. A. (2014). Activation of paramyxovirus membrane fusion and virus entry. *Curr Opin Virol* **5C**, 24-33.

Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257-1266.

Jensen, T. B., Gottwein, J. M., Scheel, T. K., Hoegh, A. M., Eugen-Olsen, J. & Bukh, J. (2008). Highly efficient JFH1-based cell-culture system for hepatitis C virus genotype 5a: failure of homologous neutralizing-antibody treatment to control infection. *J Infect Dis* **198**, 1756-1765.

Jia, L., Betters, J. L. & Yu, L. (2011). Niemann-pick C1-like 1 (NPC1L1) protein in intestinal and hepatic cholesterol transport. *Annu Rev Physiol* **73**, 239-259.

Jiang, J., Cun, W., Wu, X., Shi, Q., Tang, H. & Luo, G. (2012). Hepatitis C virus attachment mediated by apolipoprotein E binding to cell surface heparan sulfate. *J Virol* **86**, 7256-7267.

Jiang, Y., Deng, F., Rayner, S., Wang, H. & Hu, Z. (2009). Evidence of a major role of GP64 in group I alphabaculovirus evolution. *Virus Res* **142**, 85-91.

Johansson, D. X., Voisset, C., Tarr, A. W., Aung, M., Ball, J. K., Dubuisson, J. & Persson, M. A. (2007). Human combinatorial libraries yield rare antibodies that broadly neutralize hepatitis C virus. *Proc Natl Acad Sci U S A* **104**, 16269-16274.

Jolly, C. L. & Sattentau, Q. J. (2013). Attachment factors. *Adv Exp Med Biol* **790**, 1-23.

Jones, C. T., Catanese, M. T., Law, L. M., Khetani, S. R., Syder, A. J., Ploss, A., Oh, T. S., Schoggins, J. W., MacDonald, M. R., Bhatia, S. N. & Rice, C. M. (2010). Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system. *Nat Biotechnol* **28**, 167-171.

Jones, E. L., Demaria, M. C. & Wright, M. D. (2011). Tetraspanins in cellular immunity. *Biochem Soc Trans* **39**, 506-511.

Jordan, I. K., Matyunina, L. V. & McDonald, J. F. (1999). Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc Natl Acad Sci U S A* **96**, 12621-12625.

Juste, M., Muzard, J. & Billiald, P. (2006). Cloning of the antibody kappa light chain V-gene from murine hybridomas by bypassing the aberrant MOPC21-derived transcript. *Anal Biochem* **349**, 159-161.

Kabsch, W. (2010). Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132.

Kachko, A., Kochneva, G., Sivolobova, G., Grazhdantseva, A., Lupan, T., Zubkova, I., Wells, F., Merchlinsky, M., Williams, O., Watanabe, H., Ivanova, A., Shvalov, A., Loktev, V., Netesov, S. & Major, M. E. (2011). New neutralizing antibody epitopes in hepatitis C virus envelope glycoproteins are revealed by dissecting peptide recognition profiles. *Vaccine* **30**, 69-77.

Kadlec, J., Loureiro, S., Abrescia, N. G., Stuart, D. I. & Jones, I. M. (2008). The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. *Nat Struct Mol Biol* **15**, 1024-1030.

Kam, Z., Shore, H. B. & Feher, G. (1978). On the crystallization of proteins. *J Mol Biol* **123**, 539-555.

Kambara, H., Fukuhara, T., Shiokawa, M., Ono, C., Ohara, Y., Kamitani, W. & Matsuura, Y. (2012). Establishment of a novel permissive cell line for the propagation of hepatitis C virus by expression of microRNA miR122. *J Virol* **86**, 1382-1393.

Kampmann, T., Mueller, D. S., Mark, A. E., Young, P. R. & Kobe, B. (2006). The Role of histidine residues in low-pH-mediated viral membrane fusion. *Structure* **14**, 1481-1487.

Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-1871.

Kanto, T., Hayashi, N., Takehara, T., Hagiwara, H., Mita, E., Naito, M., Kasahara, A., Fusamoto, H. & Kamada, T. (1995). Density analysis of

hepatitis C virus particle population in the circulation of infected hosts: implications for virus neutralization or persistence. *J Hepatol* **22**, 440-448.

Kapoor, A., Simmonds, P., Gerold, G., Qaisar, N., Jain, K., Henriquez, J. A., Firth, C., Hirschberg, D. L., Rice, C. M., Shields, S. & Lipkin, W. I. (2011). Characterization of a canine homolog of hepatitis C virus. *Proc Natl Acad Sci U S A* **108**, 11608-11613.

Kato, 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.

Keck, Z., Wang, W., Wang, Y., Lau, P., Carlsen, T. H., Prentoe, J., Xia, J., Patel, A. H., Bukh, J. & Foung, S. K. (2013). Cooperativity in virus neutralization by human monoclonal antibodies to two adjacent regions located at the amino terminus of hepatitis C virus E2 glycoprotein. *J Virol* **87**, 37-51.

Keck, Z. Y., Angus, A. G., Wang, W., Lau, P., Wang, Y., Gatherer, D., Patel, A. H. & Foung, S. K. (2014). Non-random Escape Pathways from a Broadly Neutralizing Human Monoclonal Antibody Map to a Highly Conserved Region on the Hepatitis C Virus E2 Glycoprotein Encompassing Amino Acids 412-423. *PLoS Pathog* **10**, e1004297.

Keck, Z. Y., Li, T. K., Xia, J., Bartosch, B., Cosset, F. L., Dubuisson, J. & Foung, S. K. (2005). Analysis of a highly flexible conformational immunogenic domain a in hepatitis C virus E2. *J Virol* **79**, 13199-13208.

Keck, Z. Y., Li, T. K., Xia, J., Gal-Tanamy, M., Olson, O., Li, S. H., Patel, A. H., Ball, J. K., Lemon, S. M. & Foung, S. K. (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. (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. (2004). Hepatitis C virus E2 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. (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., Xia, J., Wang, Y., Wang, W., Krey, T., Prentoe, J., Carlsen, T., Li, A. Y., Patel, A. H., Lemon, S. M., Bukh, J., Rey, F. A. & Foung, S. K. (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.

Khan, A. G., Whidby, J., Miller, M. T., Scarborough, H., Zatorski, A. V., Cygan, A., Price, A. A., Yost, S. A., Bohannon, C. D., Jacob, J., Grakoui, A. & Marcotrigiano, J. (2014). Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature*.

Kielian, M. (2006). Class II virus membrane fusion proteins. *Virology* **344**, 38-47.
Kielian, M. (2014). Mechanisms of Virus Membrane Fusion Proteins. *Annual Review of Virology* **1**, 171-189.
Kielian, M., Chanel-Vos, C. & Liao, M. (2010). Alphavirus Entry and Membrane Fusion. *Viruses* **2**, 796-825.
Kielian, M. & Rey, F. A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* **4**, 67-76.
Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Purd'homme, N. & Bucheton, A. (1994). Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **91**, 1285-1289.
Kim, D. W., Gwack, Y., Han, J. H. & Choe, J. (1995). C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem Biophys Res Commun* **215**, 160-166.
Kitadokoro, K., Bordo, D., Galli, G., Petracca, R., Falugi, F., Abrignani, S., Grandi, G. & Bolognesi, M. (2001). CD81 extracellular domain 3D structure: insight into the tetraspanin superfamily structural motifs. *Embo J* **20**, 12-18.
Klenk, H. D. & Garten, W. (1994). Host cell proteases controlling virus pathogenicity. *Trends Microbiol* **2**, 39-43.
Kobayashi, F., Yamada, S., Taguwa, S., Kataoka, C., Naito, S., Hama, Y., Tani, H., Matsuura, Y. & Sugahara, K. (2012). Specific interaction of the envelope glycoproteins E1 and E2 with liver heparan sulfate involved in the tissue tropismatic infection by hepatitis C virus. *Glycoconj J* **29**, 211-220.
Kolykhalov, A. A., Feinstone, S. M. & Rice, C. M. (1996). Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol* **70**, 3363-3371.
Kong, L., Giang, E., Nieusma, T., Kadam, R. U., Cogburn, K. E., Hua, Y., Dai, X., Stanfield, R. L., Burton, D. R., Ward, A. B., Wilson, I. A. & Law, M. (2013). Hepatitis C virus E2 envelope glycoprotein core structure. *Science* **342**, 1090-1094.
Kong, L., Giang, E., Nieusma, T., Robbins, J. B., Deller, M. C., Stanfield, R. L., Wilson, I. A. & Law, M. (2012a). Structure of hepatitis C virus envelope glycoprotein E2 antigenic site 412 to 423 in complex with antibody AP33. *J Virol* **86**, 13085-13088.
Kong, L., Giang, E., Robbins, J. B., Stanfield, R. L., Burton, D. R., Wilson, I. A. & Law, M. (2012b). Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proc Natl Acad Sci U S A* **109**, 9499-9504.
Korkhin, Y., Frolow, F., Bogin, O., Peretz, M., Kalb, A. J. & Burstein, Y. (1996). Crystalline alcohol dehydrogenases from the mesophilic bacterium Clostridium beijerinckii and the thermophilic bacterium Thermoanaerobium brockii: preparation, characterization and molecular symmetry. *Acta Crystallogr D Biol Crystallogr* **52**, 882-886.
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.
Krey, T., d'Alayer, J., Kikuti, C. M., Saulnier, A., Damier-Piolle, L., Petitpas, I., Johansson, D. X., Tawar, R. G., Baron, B., Robert, B., England, P., Persson, M. A., Martin, A. & Rey, F. A. (2010). The disulfide bonds in

glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLoS Pathog* **6**, e1000762.

Krey, T., Meola, A., Keck, Z. Y., Damier-Piolle, L., Foung, S. K. & Rey, F. A. (2013). Structural basis of HCV neutralization by human monoclonal antibodies resistant to viral neutralization escape. *PLoS Pathog* **9**, e1003364.

Krieger, M. (2001). Scavenger receptor class B type I is a multiligand HDL receptor that influences diverse physiologic systems. *J Clin Invest* **108**, 793-797.

Krissinel, E. (2007). On the relationship between sequence and structure similarities in proteomics. *Bioinformatics* **23**, 717-723.

Kwong, P. D., Doyle, M. L., Casper, D. J., Cicala, C., Leavitt, S. A., Majeed, S., Steenbeke, T. D., Venturi, M., Chaiken, I., Fung, M., Katinger, H., Parren, P. W., Robinson, J., Van Ryk, D., Wang, L., Burton, D. R., Freire, E., Wyatt, R., Sodroski, J., Hendrickson, W. A. & Arthos, J. (2002). HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678-682.

Lamb, R. A. & Jardetzky, T. S. (2007). Structural basis of viral invasion: lessons from paramyxovirus F. *Curr Opin Struct Biol* **17**, 427-436.

Lambert, P. H., Liu, M. & Siegrist, C. A. (2005). Can successful vaccines teach us how to induce efficient protective immune responses? *Nat Med* **11**, S54-62.

Lanford, R. E., Chavez, D., Notvall, L. & Brasky, K. M. (2003). Comparison of tamarins and marmosets as hosts for GBV-B infections and the effect of immunosuppression on duration of viremia. *Virology* **311**, 72-80.

Lapierre, P., Troesch, M., Alvarez, F. & Soudeyns, H. (2011). Structural basis for broad neutralization of hepatitis C virus quasispecies. *PLoS One* **6**, e26981.

Laskowski, R. A. (2009). PDBsum new things. *Nucleic Acids Res* **37**, D355-359.

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.

Law, J. L., Chen, C., Wong, J., Hockman, D., Santer, D. M., Frey, S. E., Belshe, R. B., Wakita, T., Bukh, J., Jones, C. T., Rice, C. M., Abrignani, S., Tyrrell, D. L. & Houghton, M. (2013). A hepatitis C virus (HCV) vaccine comprising envelope glycoproteins gpE1/gpE2 derived from a single isolate elicits broad cross-genotype neutralizing antibodies in humans. *PLoS One* **8**, e59776.

Law, M., Maruyama, T., Lewis, J., Giang, E., Tarr, A. W., Stamataki, Z., Gastaminza, P., Chisari, F. V., Jones, I. M., Fox, R. I., Ball, J. K., McKeating, J. A., Kneteman, N. M. & Burton, D. R. (2008). Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. *Nat Med* **14**, 25-27.

Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946-950.

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., Philippe, 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.

Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57.

Li, Y., Wang, J., Kanai, R. & Modis, Y. (2013). Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. *Proc Natl Acad Sci U S A* **110**, 6805-6810.

Lindenbach, B. D. (2013). Virion assembly and release. *Curr Top Microbiol Immunol* **369**, 199-218.

Lindenbach, B. D., Evans, M. J., Syder, A. J., Wolk, B., Tellinghuisen, T. L., Liu, C. C., Maruyama, T., Hynes, R. O., Burton, D. R., McKeating, J. A. & Rice, C. M. (2005). Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623-626.

Lindenbach, B. D., Meuleman, P., Ploss, A., Vanwolleghem, T., Syder, A. J., McKeating, J. A., Lanford, R. E., Feinstone, S. M., Major, M. E., Leroux-Roels, G. & Rice, C. M. (2006). Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro. *Proc Natl Acad Sci U S A* **103**, 3805-3809.

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.

Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-2198.

Logvinoff, C., Major, M. E., Oldach, D., Heyward, S., Talal, A., Balfe, P., Feinstone, S. M., Alter, H., Rice, C. M. & McKeating, J. A. (2004). Neutralizing antibody response during acute and chronic hepatitis C virus infection. *Proc Natl Acad Sci U S A* **101**, 10149-10154.

Lohmann, V. (2013). Hepatitis C virus RNA replication. *Curr Top Microbiol Immunol* **369**, 167-198.

Lohmann, V., Korner, F., Herian, U. & Bartenschlager, R. (1997). Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J Virol* **71**, 8416-8428.

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.

Lok, S. M., Kostyuchenko, V., Nybakken, G. E., Holdaway, H. A., Battisti, A. J., Sukupolvi-Petty, S., Sedlak, D., Fremont, D. H., Chipman, P. R., Roehrig, J. T., Diamond, M. S., Kuhn, R. J. & Rossmann, M. G. (2008). Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. *Nat Struct Mol Biol* **15**, 312-317.

Long, G., Pan, X., Kormelink, R. & Vlak, J. M. (2006a). Functional entry of baculovirus into insect and mammalian cells is dependent on clathrin-mediated endocytosis. *J Virol* **80**, 8830-8833.

Long, G., Westenberg, M., Wang, H., Vlak, J. M. & Hu, Z. (2006b). Function, oligomerization and N-linked glycosylation of the Helicoverpa armigera single nucleopolyhedrovirus envelope fusion protein. *J Gen Virol* **87**, 839-846.

Lorenz, I. C., Allison, S. L., Heinz, F. X. & Helenius, A. (2002). Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. *J Virol* **76**, 5480-5491.

Lozach, P. Y., Lortat-Jacob, H., de Lacroix de Lavalette, A., Staropoli, I., Foung, S., Amara, A., Houles, C., Fieschi, F., Schwartz, O., Virelizier, J. L., Arenzana-Seisdedos, F. & Altmeyer, R. (2003). DC-SIGN and L-SIGN are

high affinity binding receptors for hepatitis C virus glycoprotein E2. *J Biol Chem* **278**, 20358-20366.

Luik, P., Chew, C., Aittoniemi, J., Chang, J., Wentworth, P., Jr., Dwek, R. A., Biggin, P. C., Venien-Bryan, C. & Zitzmann, N. (2009). The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. *Proc Natl Acad Sci U S A* **106**, 12712-12716.

Lung, O. & Blissard, G. W. (2005). A cellular Drosophila melanogaster protein with similarity to baculovirus F envelope fusion proteins. *J Virol* **79**, 7979-7989.

Lung, O., Westenberg, M., Vlak, J. M., Zuidema, D. & Blissard, G. W. (2002). Pseudotyping Autographa californica multicapsid nucleopolyhedrovirus (AcMNPV): F proteins from group II NPVs are functionally analogous to AcMNPV GP64. *J Virol* **76**, 5729-5736.

Lupas, A., Muller, S., Goldie, K., Engel, A. M., Engel, A. & Baumeister, W. (1995). Model structure of the Omp alpha rod, a parallel four-stranded coiled coil from the hyperthermophilic eubacterium Thermotoga maritima. *J Mol Biol* **248**, 180-189.

Lupberger, J., Zeisel, M. B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C. J., Turek, M., Gorke, S., Royer, C., Fischer, B., Zahid, M. N., Lavillette, D., Fresquet, J., Cosset, F. L., Rothenberg, S. M., Pietschmann, T., Patel, A. H., Pessaux, P., Doffoel, M., Raffelsberger, W., Poch, O., McKeating, J. A., Brino, L. & Baumert, T. F. (2011). EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat Med* **17**, 589-595.

Maasoumy, B. & Wedemeyer, H. (2012). Natural history of acute and chronic hepatitis C. *Best Pract Res Clin Gastroenterol* **26**, 401-412.

Machida, K., Kondo, Y., Huang, J. Y., Chen, Y. C., Cheng, K. T., Keck, Z., Foung, S., Dubuisson, J., Sung, V. M. & Lai, M. M. (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.

Maiti, R., Van Domselaar, G. H., Zhang, H. & Wishart, D. S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* **32**, W590-594.

Malik, H. S. & Henikoff, S. (2005). Positive selection of Iris, a retroviral envelope-derived host gene in Drosophila melanogaster. *PLoS Genet* **1**, e44.

Malik, H. S., Henikoff, S. & Eickbush, T. H. (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* **10**, 1307-1318.

Manns, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M. & Albrecht, J. K. (2001). Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* **358**, 958-965.

Martin, A., Bodola, F., Sangar, D. V., Goettge, K., Popov, V., Rijnbrand, R., Lanford, R. E. & Lemon, S. M. (2003). Chronic hepatitis associated with GB

virus B persistence in a tamarin after intrahepatic inoculation of synthetic viral RNA. *Proc Natl Acad Sci U S A* **100**, 9962-9967.

Martin, D. N. & Uprichard, S. L. (2013). Identification of transferrin receptor 1 as a hepatitis C virus entry factor. *Proc Natl Acad Sci U S A* **110**, 10777-10782.

Masciopinto, F., Freer, G., Burgio, V. L., Levy, S., Galli-Stampino, L., Bendinelli, M., Houghton, M., Abrignani, S. & Uematsu, Y. (2002). Expression of human CD81 in transgenic mice does not confer susceptibility to hepatitis C virus infection. *Virology* **304**, 187-196.

Matsuura, Y., Suzuki, T., Suzuki, R., Sato, M., Aizaki, H., Saito, I. & Miyamura, T. (1994). Processing of E1 and E2 glycoproteins of hepatitis C virus expressed in mammalian and insect cells. *Virology* **205**, 141-150.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* **33**, 491-497.

Maurer, K., Krey, T., Moennig, V., Thiel, H. J. & Rumenapf, T. (2004). CD46 is a cellular receptor for bovine viral diarrhea virus. *J Virol* **78**, 1792-1799.

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.

McCaffrey, K., Boo, I., Poumbourios, P. & Drummer, H. E. (2007). Expression and characterization of a minimal hepatitis C virus glycoprotein E2 core domain that retains CD81 binding. *J Virol* **81**, 9584-9590.

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.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

McCune, J. M., Rabin, L. B., Feinberg, M. B., Lieberman, M., Kosek, J. C., Reyes, G. R. & Weissman, I. L. (1988). Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* **53**, 55-67.

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. *The New England journal of medicine* **360**, 1827-1838.

McHutchison, J. G. & Fried, M. W. (2003). Current therapy for hepatitis C: pegylated interferon and ribavirin. *Clin Liver Dis* **7**, 149-161.

McKeating, J. A., Zhang, L. Q., Logvinoff, C., Flint, M., Zhang, J., Yu, J., Butera, D., Ho, D. D., Dustin, L. B., Rice, C. M. & Balfe, P. (2004). Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. *J Virol* **78**, 8496-8505.

McLellan, J. S., Chen, M., Leung, S., Graepel, K. W., Du, X., Yang, Y., Zhou, T., Baxa, U., Yasuda, E., Beaumont, T., Kumar, A., Modjarrad, K., Zheng, Z., Zhao, M., Xia, N., Kwong, P. D. & Graham, B. S. (2013). Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113-1117.

Mee, C. J., Harris, H. J., Farquhar, M. J., Wilson, G., Reynolds, G., Davis, C., van, I. S. C., Balfe, P. & McKeating, J. A. (2009). Polarization restricts hepatitis C virus entry into HepG2 hepatoma cells. *J Virol* **83**, 6211-6221.

Meertens, L., Bertaux, C. & Dragic, T. (2006). Hepatitis C virus entry requires a critical postinternalization step and delivery to early endosomes via clathrin-coated vesicles. *J Virol* **80**, 11571-11578.

Meola, A., Deville, C., Jeffers, S. A., Guardado-Calvo, P., Vasiliauskaite, I., Sizun, C., Girard-Blanc, C., Malosse, C., van Heijenoort, C., Chamot-Rooke, J., Krey, T., Guittet, E., Petres, S., Rey, F. A. & Bontems, F. (2014). Robust and low cost uniform N-labeling of proteins expressed in Drosophila S2 cells and Spodoptera frugiperda Sf9 cells for NMR applications. *J Struct Biol*.

Meola, A., Sbardellati, A., Bruni Ercole, B., Cerretani, M., Pezzanera, M., Ceccacci, A., Vitelli, A., Levy, S., Nicosia, A., Traboni, C., McKeating, J. & Scarselli, E. (2000). Binding of hepatitis C virus E2 glycoprotein to CD81 does not correlate with species permissiveness to infection. *J Virol* **74**, 5933-5938.

Meola A, T. A. W., England P, Meredith L W, McClure C P, Foung S K H, McKeating J A, Ball, J K, Rey F A, Krey T (2014). Structural Flexibility of a Highly Conserved Neutralizing Epitope in Hepatitis C Virus Glycoprotein E2 In *21st International Symposium on Hepatitis C and Related Viruses*

Mercer, J., Schelhaas, M. & Helenius, A. (2010). Virus entry by endocytosis. *Annu Rev Biochem* **79**, 803-833.

Merola, M., Brazzoli, M., Cocchiarella, F., Heile, J. M., Helenius, A., Weiner, A. J., Houghton, M. & Abrignani, S. (2001). Folding of hepatitis C virus E1 glycoprotein in a cell-free system. *J Virol* **75**, 11205-11217.

Merz, A., Long, G., Hiet, M. S., Brugger, B., Chlanda, P., Andre, P., Wieland, F., Krijnse-Locker, J. & Bartenschlager, R. (2011). Biochemical and morphological properties of hepatitis C virus particles and determination of their lipidome. *J Biol Chem* **286**, 3018-3032.

Meuleman, P., Catanese, M. T., Verhoye, L., Desombere, I., Farhoudi, A., Jones, C. T., Sheahan, T., Grzyb, K., Cortese, R., Rice, C. M., Leroux-Roels, G. & Nicosia, A. (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.

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

Meunier, J. C., Russell, R. S., Engle, R. E., Faulk, K. N., Purcell, R. H. & Emerson, S. U. (2008a). Apolipoprotein c1 association with hepatitis C virus. *J Virol* **82**, 9647-9656.

Meunier, J. C., Russell, R. S., Goossens, V., Priem, S., Walter, H., Depla, E., Union, A., Faulk, K. N., Bukh, J., Emerson, S. U. & Purcell, R. H. (2008b). Isolation and characterization of broadly neutralizing human monoclonal antibodies to the e1 glycoprotein of hepatitis C virus. *J Virol* **82**, 966-973.

Meyer, K., Basu, A., Przysiecki, C. T., Lagging, L. M., Di Bisceglie, A. M., Conley, A. J. & Ray, R. (2002). Complement-mediated enhancement of antibody function for neutralization of pseudotype virus containing hepatitis C virus E2 chimeric glycoprotein. *J Virol* **76**, 2150-2158.

Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X. Y., Edouard, P., Howes, S., Keith, J. C., Jr. & McCoy, J. M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785-789.

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 (Pt 9)**, 2299-2306.

Miller, D. W. & Miller, L. K. (1982). A virus mutant with an insertion of a copia-like transposable element. *Nature* **299**, 562-564.

Miller, R. H. & Purcell, R. H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc Natl Acad Sci U S A* **87**, 2057-2061.

Misseri, Y., Cerutti, M., Devauchelle, G., Bucheton, A. & Terzian, C. (2004). Analysis of the Drosophila gypsy endogenous retrovirus envelope glycoprotein. *J Gen Virol* **85**, 3325-3331.

Misseri, Y., Labesse, G., Bucheton, A. & Terzian, C. (2003). Comparative sequence analysis and predictions for the envelope glycoproteins of insect endogenous retroviruses. *Trends Microbiol* **11**, 253-256.

Modis, Y. (2014). Relating structure to evolution in class II viral membrane fusion proteins. *Curr Opin Virol* **5**, 34-41.

Mohd Hanafiah, K., Groeger, J., Flaxman, A. D. & Wiersma, S. T. (2013). Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* **57**, 1333-1342.

Molina, S., Castet, V., Pichard-Garcia, L., Wychowski, C., Meurs, E., Pascussi, J. M., Sureau, C., Fabre, J. M., Sacunha, A., Larrey, D., Dubuisson, J., Coste, J., McKeating, J., Maurel, P. & Fournier-Wirth, C. (2008). Serum-derived hepatitis C virus infection of primary human hepatocytes is tetraspanin CD81 dependent. *J Virol* **82**, 569-574.

Montserret, R., Saint, N., Vanbelle, C., Salvay, A. G., Simorre, J. P., Ebel, C., Sapay, N., Renisio, J. G., Bockmann, A., Steinmann, E., Pietschmann, T., Dubuisson, J., Chipot, C. & Penin, F. (2010). NMR structure and ion channel activity of the p7 protein from hepatitis C virus. *J Biol Chem* **285**, 31446-31461.

Morikawa, K., Zhao, Z., Date, T., Miyamoto, M., Murayama, A., Akazawa, D., Tanabe, J., Sone, S. & Wakita, T. (2007). The roles of CD81 and glycosaminoglycans in the adsorption and uptake of infectious HCV particles. *J Med Virol* **79**, 714-723.

Morin, T. J., Broering, T. J., Leav, B. A., Blair, B. M., Rowley, K. J., Boucher, E. N., Wang, Y., Cheslock, P. S., Knauber, M., Olsen, D. B., Ludmerer, S. W., Szabo, G., Finberg, R. W., Purcell, R. H., Lanford, R. E., Ambrosino, D. M., Molrine, D. C. & Babcock, G. J. (2012). Human monoclonal antibody HCV1 effectively prevents and treats HCV infection in chimpanzees. *PLoS Pathog* **8**, e1002895.

Mothes, W., Boerger, A. L., Narayan, S., Cunningham, J. M. & Young, J. A. (2000). Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein. *Cell* **103**, 679-689.

Mothes, W., Sherer, N. M., Jin, J. & Zhong, P. (2010). Virus cell-to-cell transmission. *J Virol* **84**, 8360-8368.

Muerhoff, A. S., Leary, T. P., Simons, J. N., Pilot-Matias, T. J., Dawson, G. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Desai, S. M. &

Mushahwar, I. K. (1995). Genomic organization of GB viruses A and B: two new members of the Flaviviridae associated with GB agent hepatitis. *J Virol* **69**, 5621-5630.

Muhlberger, N., Schwarzer, R., Lettmeier, B., Sroczynski, G., Zeuzem, S. & Siebert, U. (2009). HCV-related burden of disease in Europe: a systematic assessment of incidence, prevalence, morbidity, and mortality. *BMC Public Health* **9**, 34.

Muir, A. J. (2014). The rapid evolution of treatment strategies for hepatitis C. *Am J Gastroenterol* **109**, 628-635; quiz 636.

Myszka, D. G., Sweet, R. W., Hensley, P., Brigham-Burke, M., Kwong, P. D., Hendrickson, W. A., Wyatt, R., Sodroski, J. & Doyle, M. L. (2000). Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci U S A* **97**, 9026-9031.

Nam, J. H., Faulk, K., Engle, R. E., Govindarajan, S., St Claire, M. & Bukh, J. (2004). In vivo analysis of the 3' untranslated region of GB virus B after in vitro mutagenesis of an infectious cDNA clone: persistent infection in a transfected tamarin. *J Virol* **78**, 9389-9399.

Narbus, C. M., Israelow, B., Sourisseau, M., Michta, M. L., Hopcraft, S. E., Zeiner, G. M. & Evans, M. J. (2011). HepG2 cells expressing microRNA miR-122 support the entire hepatitis C virus life cycle. *J Virol* **85**, 12087-12092.

Nattermann, J., Schneiders, A. M., Leifeld, L., Langhans, B., Schulz, M., Inchauspe, G., Matz, B., Brackmann, H. H., Houghton, M., Sauerbruch, T. & Spengler, U. (2005). Serum antibodies against the hepatitis C virus E2 protein mediate antibody-dependent cellular cytotoxicity (ADCC). *J Hepatol* **42**, 499-504.

Ndongo, N., Berthillon, P., Pradat, P., Vieux, C., Bordes, I., Berby, F., Maynard, M., Zoulim, F., Trepo, C. & Petit, M. A. (2010). Association of anti-E1E2 antibodies with spontaneous recovery or sustained viral response to therapy in patients infected with hepatitis C virus. *Hepatology* **52**, 1531-1542.

Netski, D. M., Mosbruger, T., Depla, E., Maertens, G., Ray, S. C., Hamilton, R. G., Roundtree, S., Thomas, D. L., McKeating, J. & Cox, A. (2005). Humoral immune response in acute hepatitis C virus infection. *Clin Infect Dis* **41**, 667-675.

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-alpha 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., Depla, E., Dincq, S., Vander Stichele, C., Maertens, G. & Hulstaert, F. (2003). A pilot study of therapeutic vaccination with envelope protein E1 in 35 patients with chronic hepatitis C. *Hepatology* **38**, 1289-1296.

Nielsen, S. U., Bassendine, M. F., Burt, A. D., Martin, C., Pumeechockchai, W. & Toms, G. L. (2006). Association between hepatitis C virus and very-low-density lipoprotein (VLDL)/LDL analyzed in iodixanol density gradients. *J Virol* **80**, 2418-2428.

Niepmann, M. (2013). Hepatitis C virus RNA translation. *Curr Top Microbiol Immunol* **369**, 143-166.

Obmolova, G., Malia, T. J., Teplyakov, A., Sweet, R. & Gilliland, G. L. (2010). Promoting crystallization of antibody-antigen complexes via microseed matrix screening. *Acta Crystallogr D Biol Crystallogr* **66**, 927-933.

Oomens, A. G. & Blissard, G. W. (1999). Requirement for GP64 to drive efficient budding of Autographa californica multicapsid nucleopolyhedrovirus. *Virology* **254**, 297-314.

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., Montserret, R., Duvet, S., Cocquerel, L., Cacan, R., Barberot, B., Le Maire, M., Penin, F. & Dubuisson, J. (2000). The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J Biol Chem* **275**, 31428-31437.

Op De Beeck, A., Voisset, C., Bartosch, B., Ciczora, Y., Cocquerel, L., Keck, Z., Foung, S., Cosset, F. L. & Dubuisson, J. (2004). Characterization of functional hepatitis C virus envelope glycoproteins. *J Virol* **78**, 2994-3002.

Osburn, W. O., Snider, A. E., Wells, B. L., Latanich, R., Bailey, J. R., Thomas, D. L., Cox, A. L. & Ray, S. C. (2014). Clearance of hepatitis C infection is associated with the early appearance of broad neutralizing antibody responses. *Hepatology*.

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. M., Timms, J. M., Tarr, A. W., Brown, R. J., Hickling, T. P., Szwejk, A., Bienkowska-Szewczyk, K., Thomson, B. J., Patel, A. H. & Ball, J. K. (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.

Pang, P. S., Planet, P. J. & Glenn, J. S. (2009). The evolution of the major hepatitis C genotypes correlates with clinical response to interferon therapy. *PLoS One* **4**, e6579.

Pantua, H., Diao, J., Ultsch, M., Hazen, M., Mathieu, M., McCutcheon, K., Takeda, K., Date, S., Cheung, T. K., Phung, Q., Hass, P., Arnott, D., Hongo, J. A., Matthews, D. J., Brown, A., Patel, A. H., Kelley, R. F., Eigenbrot, C. & Kapadia, S. B. (2013). Glycan shifting on hepatitis C virus (HCV) E2 glycoprotein is a mechanism for escape from broadly neutralizing antibodies. *J Mol Biol* **425**, 1899-1914.

Parks, W. P., Melnick, J. L., Voss, W. R., Singer, D. B., Rosenberg, H. S., Alcott, J. & Casazza, A. M. (1969). Characterization of marmoset hepatitis virus. *J Infect Dis* **120**, 548-559.

Paterson, R. G. & Lamb, R. A. (1987). Ability of the hydrophobic fusion-related external domain of a paramyxovirus F protein to act as a membrane anchor. *Cell* **48**, 441-452.

Pearson, M. N. & Rohrmann, G. F. (2002). Transfer, incorporation, and substitution of envelope fusion proteins among members of the Baculoviridae, Orthomyxoviridae, and Metaviridae (insect retrovirus) families. *J Virol* **76**, 5301-5304.

Penin, F., Combet, C., Germanidis, G., Frainais, P. O., Deleage, G. & Pawlotsky, J. M. (2001). Conservation of the conformation and positive charges of

hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. *J Virol* **75**, 5703-5710.

Perez-Berna, A. J., Moreno, M. R., Guillen, J., Bernabeu, A. & Villalain, J. (2006). The membrane-active regions of the hepatitis C virus E1 and E2 envelope glycoproteins. *Biochemistry* **45**, 3755-3768.

Perez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B. & Rey, F. A. (2014). Structural basis of eukaryotic cell-cell fusion. *Cell* **157**, 407-419.

Perotti, M., Mancini, N., Diotti, R. A., Tarr, A. W., Ball, J. K., Owsianka, A., Adair, R., Patel, A. H., Clementi, M. & Burioni, R. (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., Viazov, S., Rispeter, K., Blum, H. E., Roggendorf, M. & Baumert, T. F. (2007). Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. *Proc Natl Acad Sci U S A* **104**, 6025-6030.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.

Pierson, T. C., Xu, Q., Nelson, S., Oliphant, T., Nybakken, G. E., Fremont, D. H. & Diamond, M. S. (2007). The stoichiometry of antibody-mediated neutralization and enhancement of West Nile virus infection. *Cell Host Microbe* **1**, 135-145.

Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M., Cosset, F. L. & Bartenschlager, R. (2006). Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc Natl Acad Sci U S A* **103**, 7408-7413.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G. & Abrignani, S. (1998). Binding of hepatitis C virus to CD81. *Science* **282**, 938-941.

Plattet, P. & Plemper, R. K. (2013). Envelope protein dynamics in paramyxovirus entry. *MBio* **4**.

Plemper, R. K. (2011). Cell entry of enveloped viruses. *Curr Opin Virol* **1**, 92-100.

Ploss, A., Evans, M. J., Gaysinskaya, V. A., Panis, M., You, H., de Jong, Y. P. & Rice, C. M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457**, 882-886.

Podevin, P., Carpentier, A., Pene, V., Aoudjehane, L., Carriere, M., Zaidi, S., Hernandez, C., Calle, V., Meritet, J. F., Scatton, O., Dreux, M., Cosset, F. L., Wakita, T., Bartenschlager, R., Demignot, S., Conti, F., Rosenberg, A. R. & Calmus, Y. (2010). Production of infectious hepatitis C virus in primary cultures of human adult hepatocytes. *Gastroenterology* **139**, 1355-1364.

Potter, J. A., Owsianka, A. M., Jeffery, N., Matthews, D. J., Keck, Z. Y., Lau, P., Foung, S. K., Taylor, G. L. & Patel, A. H. (2012). Toward a hepatitis C virus vaccine: the structural basis of hepatitis C virus neutralization by AP33, a broadly neutralizing antibody. *J Virol* **86**, 12923-12932.

Poveda, E., Wyles, D. L., Mena, A., Pedreira, J. D., Castro-Iglesias, A. & Cachay, E. (2014). Update on hepatitis C virus resistance to direct-acting antiviral agents. *Antiviral Res* **108C**, 181-191.

Powell, H. R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* **55**, 1690-1695.

Powers, K. A., Ribeiro, R. M., Patel, K., Pianko, S., Nyberg, L., Pockros, P., Conrad, A. J., McHutchison, J. & Perelson, A. S. (2006). Kinetics of hepatitis C virus reinfection after liver transplantation. *Liver Transpl* **12**, 207-216.

Prentoe, J., Jensen, T. B., Meuleman, P., Serre, S. B., Scheel, T. K., 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.

Prince, A. M., Brotman, B., Grady, G. F., Kuhns, W. J., Hazzi, C., Levine, R. W. & Millian, S. J. (1974). Long-incubation post-transfusion hepatitis without serological evidence of exposure to hepatitis-B virus. *Lancet* **2**, 241-246.

Prince, A. M., Huima-Byron, T., Parker, T. S. & Levine, D. M. (1996). Visualization of hepatitis C virions and putative defective interfering particles isolated from low-density lipoproteins. *J Viral Hepat* **3**, 11-17.

Raghuraman, S., Park, H., Osburn, W. O., Winkelstein, E., Edlin, B. R. & Rehermann, B. (2012). Spontaneous clearance of chronic hepatitis C virus infection is associated with appearance of neutralizing antibodies and reversal of T-cell exhaustion. *J Infect Dis* **205**, 763-771.

Ralston, R., Thudium, K., Berger, K., Kuo, C., Gervase, B., Hall, J., Selby, M., Kuo, G., Houghton, M. & Choo, Q. L. (1993). Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses. *J Virol* **67**, 6753-6761.

Ray, R., Meyer, K., Banerjee, A., Basu, A., Coates, S., Abrignani, S., Houghton, M., Frey, S. E. & Belshe, R. B. (2010). Characterization of antibodies induced by vaccination with hepatitis C virus envelope glycoproteins. *J Infect Dis* **202**, 862-866.

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.

Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. *Nature* **375**, 291-298.

Rocha-Perugini, V., Lavie, M., Delgrange, D., Canton, J., Pillez, A., Potel, J., Lecoeur, C., Rubinstein, E., Dubuisson, J., Wychowski, C. & Cocquerel, L. (2009). The association of CD81 with tetraspanin-enriched microdomains is not essential for Hepatitis C virus entry. *BMC Microbiol* **9**, 111.

Rocha-Perugini, V., Montpellier, C., Delgrange, D., Wychowski, C., Helle, F., Pillez, A., Drobecq, H., Le Naour, F., Charrin, S., Levy, S., Rubinstein, E., Dubuisson, J. & Cocquerel, L. (2008). The CD81 partner EWI-2wint inhibits hepatitis C virus entry. *PLoS One* **3**, e1866.

Roche, S., Bressanelli, S., Rey, F. A. & Gaudin, Y. (2006). Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* **313**, 187-191.

Roche, S. & Gaudin, Y. (2002). Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers. *Virology* **297**, 128-135.

Rohrmann, G. F. (2013). In *Baculovirus Molecular Biology: Third Edition*. Bethesda (MD).

Rohrmann, G. F. & Karplus, P. A. (2001). Relatedness of baculovirus and gypsy retrotransposon envelope proteins. *BMC Evol Biol* **1**, 1.

Ruigrok, R. W., Aitken, A., Calder, L. J., Martin, S. R., Skehel, J. J., Wharton, S. A., Weis, W. & Wiley, D. C. (1988). Studies on the structure of the influenza virus haemagglutinin at the pH of membrane fusion. *J Gen Virol* **69 (Pt 11)**, 2785-2795.

Russo, M. W. & Fried, M. W. (2003). Side effects of therapy for chronic hepatitis C. *Gastroenterology* **124**, 1711-1719.

Sabo, M. C., Luca, V. C., Ray, S. C., Bukh, J., Fremont, D. H. & Diamond, M. S. (2012). Hepatitis C virus epitope exposure and neutralization by antibodies is affected by time and temperature. *Virology* **422**, 174-184.

Sainz, B., Jr., Barretto, N., Martin, D. N., Hiraga, N., Imamura, M., Hussain, S., Marsh, K. A., Yu, X., Chayama, K., Alrefai, W. A. & Uprichard, S. L. (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.

Satoh, S., Tanji, Y., Hijikata, M., Kimura, K. & Shimotohno, K. (1995). The N-terminal region of hepatitis C virus nonstructural protein 3 (NS3) is essential for stable complex formation with NS4A. *J Virol* **69**, 4255-4260.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. & Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *Embo J* **21**, 5017-5025.

Scheel, T. K., Gottwein, J. M., Jensen, T. B., Prentoe, J. C., Hoegh, A. M., Alter, H. J., Eugen-Olsen, J. & Bukh, J. (2008). Development of JFH1-based cell culture systems for hepatitis C virus genotype 4a and evidence for cross-genotype neutralization. *Proc Natl Acad Sci U S A* **105**, 997-1002.

Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779.

Schneider-Schaulies, J. (2000). Cellular receptors for viruses: links to tropism and pathogenesis. *J Gen Virol* **81**, 1413-1429.

Sharma, N. R., Mateu, G., Dreux, M., Grakoui, A., Cosset, F. L. & Melikyan, G. B. (2011). Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion. *J Biol Chem* **286**, 30361-30376.

Sillanpaa, M., Melen, K., Porkka, P., Fagerlund, R., Nevalainen, K., Lappalainen, M. & Julkunen, I. (2009). Hepatitis C virus core, NS3, NS4B and NS5A are the major immunogenic proteins in humoral immunity in chronic HCV infection. *Virol J* **6**, 84.

Simmonds, P., Holmes, E. C., Cha, T. A., Chan, S. W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. & Urdea, M. S. (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol* **74 (Pt 11)**, 2391-2399.

Simons, J. N., Pilot-Matias, T. J., Leary, T. P., Dawson, G. J., Desai, S. M., Schlauder, G. G., Muerhoff, A. S., Erker, J. C., Buijk, S. L., Chalmers, M. L. & et al. (1995). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci U S A* **92**, 3401-3405.

Singh, I., Doms, R. W., Wagner, K. R. & Helenius, A. (1990). Intracellular transport of soluble and membrane-bound glycoproteins: folding, assembly and secretion of anchor-free influenza hemagglutinin. *Embo J* **9**, 631-639.

Skehel, J. J. & Wiley, D. C. (1998). Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* **95**, 871-874.

Skehel, J. J. & Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* **69**, 531-569.

Smith, A. E. & Helenius, A. (2004). How viruses enter animal cells. *Science* **304**, 237-242.

Smith, D. B., Bukh, J., Kuiken, C., Muerhoff, A. S., Rice, C. M., Stapleton, J. T. & Simmonds, P. (2014). Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology* **59**, 318-327.

Smith, E. C., Smith, S. E., Carter, J. R., Webb, S. R., Gibson, K. M., Hellman, L. M., Fried, M. G. & Dutch, R. E. (2013). Trimeric transmembrane domain interactions in paramyxovirus fusion proteins: roles in protein folding, stability, and function. *J Biol Chem* **288**, 35726-35735.

Song, S. U., Kurkulos, M., Boeke, J. D. & Corces, V. G. (1997). Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of Drosophila. *Development* **124**, 2789-2798.

Spaete, R. R., Alexander, D., Rugroden, M. E., Choo, Q. L., Berger, K., Crawford, K., Kuo, C., Leng, S., Lee, C., Ralston, R. & et al. (1992). Characterization of the hepatitis C virus E2/NS1 gene product expressed in mammalian cells. *Virology* **188**, 819-830.

Stiasny, K., Allison, S. L., Schalich, J. & Heinz, F. X. (2002). Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *J Virol* **76**, 3784-3790.

Stura, E. A. W., Ian A. (1991). Applications of the streak seeding technique in protein crystallization. *Journal of Crystal Growth* **110**, 270-282.

Sulkowski, M. S., Gardiner, D. F., Rodriguez-Torres, M., Reddy, K. R., Hassanein, T., Jacobson, I., Lawitz, E., Lok, A. S., Hinestrosa, F., Thuluvath, P. J., Schwartz, H., Nelson, D. R., Everson, G. T., Eley, T., Wind-Rotolo, M., Huang, S. P., Gao, M., Hernandez, D., McPhee, F., Sherman, D., Hindes, R., Symonds, W., Pasquinelli, C. & Grasela, D. M. (2014). Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. *The New England journal of medicine* **370**, 211-221.

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.

Swanson, K., Wen, X., Leser, G. P., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. (2010). Structure of the Newcastle disease virus F protein in the post-fusion conformation. *Virology* **402**, 372-379.

Swanson, K. A., Settembre, E. C., Shaw, C. A., Dey, A. K., Rappuoli, R., Mandl, C. W., Dormitzer, P. R. & Carfi, A. (2011). Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers. *Proc Natl Acad Sci U S A* **108**, 9619-9624.

Tabor, E., Peterson, D. A., April, M., Seeff, L. B. & Gerety, R. J. (1980). Transmission of human non-A, non-B hepatitis to chimpanzees following failure to transmit GB agent hepatitis. *J Med Virol* **5**, 103-108.

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.

Takada, A. & Kawaoka, Y. (2003). Antibody-dependent enhancement of viral infection: molecular mechanisms and in vivo implications. *Rev Med Virol* **13**, 387-398.

Takikawa, S., Engle, R. E., Emerson, S. U., Purcell, R. H., St Claire, M. & Bukh, J. (2006). Functional analyses of GB virus B p13 protein: development of a recombinant GB virus B hepatitis virus with a p7 protein. *Proc Natl Acad Sci U S A* **103**, 3345-3350.

Taniguchi, S., Okamoto, H., Sakamoto, M., Kojima, M., Tsuda, F., Tanaka, T., Munekata, E., Muchmore, E. E., Peterson, D. A. & Mishiro, S. (1993). A structurally flexible and antigenically variable N-terminal domain of the hepatitis C virus E2/NS1 protein: implication for an escape from antibody. *Virology* **195**, 297-301.

Tao, W., Xu, C., Ding, Q., Li, R., Xiang, Y., Chung, J. & Zhong, J. (2009). A single point mutation in E2 enhances hepatitis C virus infectivity and alters lipoprotein association of viral particles. *Virology* **395**, 67-76.

Tarr, A. W., Lafaye, P., Meredith, L., Damier-Piolle, L., Urbanowicz, R. A., Meola, A., Jestin, J. L., Brown, R. J., McKeating, J. A., Rey, F. A., Ball, J. K. & Krey, T. (2013). An alpaca nanobody inhibits hepatitis C virus entry and cell-to-cell transmission. *Hepatology* **58**, 932-939.

Tarr, A. W., Owsianka, A. M., Jayaraj, D., Brown, R. J., Hickling, T. P., Irving, W. L., Patel, A. H. & Ball, J. K. (2007a). 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., Szwejk, A., Ball, J. K. & Patel, A. H. (2007b). Cloning, expression, and functional analysis of patient-derived hepatitis C virus glycoproteins. *Methods Mol Biol* **379**, 177-197.

Tarr, A. W., Owsianka, A. M., Timms, J. M., McClure, C. P., Brown, R. J., Hickling, T. P., Pietschmann, T., Bartenschlager, R., Patel, A. H. & Ball, J. K. (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., Jayaraj, D., Brown, R. J., McKeating, J. A., Irving, W. L. & Ball, J. K. (2012). 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.

Terwilliger, T. C. (2002). Rapid automatic NCS identification using heavy-atom substructures. *Acta Crystallogr D Biol Crystallogr* **58**, 2213-2215.

Thomas, D. L. (2013). Global control of hepatitis C: where challenge meets opportunity. *Nat Med* **19**, 850-858.

Thomssen, R., Bonk, S., Propfe, C., Heermann, K. H., Kochel, H. G. & Uy, A. (1992). Association of hepatitis C virus in human sera with beta-lipoprotein. *Med Microbiol Immunol* **181**, 293-300.

Thomssen, R., Bonk, S. & Thiele, A. (1993). Density heterogeneities of hepatitis C virus in human sera due to the binding of beta-lipoproteins and immunoglobulins. *Med Microbiol Immunol* **182**, 329-334.

Thorn, A. & Sheldrick, G. M. (2011). ANODE: anomalous and heavy-atom density calculation. *J Appl Crystallogr* **44**, 1285-1287.

Timpe, J. M., Stamataki, Z., Jennings, A., Hu, K., Farquhar, M. J., Harris, H. J., Schwarz, A., Desombere, I., Roels, G. L., Balfe, P. & McKeating, J. A. (2008). Hepatitis C virus cell-cell transmission in hepatoma cells in the presence of neutralizing antibodies. *Hepatology* **47**, 17-24.

Troesch, M., Meunier, I., Lapierre, P., Lapointe, N., Alvarez, F., Boucher, M. & Soudeyns, H. (2006). Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein. *Virology* **352**, 357-367.

Vagin, A. & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22-25.

van Spriel, A. B. (2011). Tetraspanins in the humoral immune response. *Biochem Soc Trans* **39**, 512-517.

Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R. & Blasi, U. (2008). The C-terminal domain of Escherichia coli Hfq is required for regulation. *Nucleic Acids Res* **36**, 133-143.

Verna, E. C. & Brown, R. S., Jr. (2006). Hepatitis C virus and liver transplantation. *Clin Liver Dis* **10**, 919-940.

Verstrepen, B. E., Depla, E., Rollier, C. S., Mares, G., Drexhage, J. A., Priem, S., Verschoor, E. J., Koopman, G., Granier, C., Dreux, M., Cosset, F. L., Maertens, G. & Heeney, J. L. (2011). Clearance of genotype 1b hepatitis C virus in chimpanzees in the presence of vaccine-induced E1-neutralizing antibodies. *J Infect Dis* **204**, 837-844.

Vieyres, G., Thomas, X., Descamps, V., Duverlie, G., Patel, A. H. & Dubuisson, J. (2010). Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. *J Virol* **84**, 10159-10168.

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.

Voss, J. E., Vaney, M. C., Duquerroy, S., Vonrhein, C., Girard-Blanc, C., Crublet, E., Thompson, A., Bricogne, G. & Rey, F. A. (2010). Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **468**, 709-712.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H. G., Mizokami, M., Bartenschlager, R. & Liang, T. J. (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**, 791-796.

Walter, T. S., Mancini, E. J., Kadlec, J., Graham, S. C., Assenberg, R., Ren, J., Sainsbury, S., Owens, R. J., Stuart, D. I., Grimes, J. M. & Harlos, K. (2008). Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 14-18.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol* **115**, 90-112.

Wang, C., Sarnow, P. & Siddiqui, A. (1993). Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J Virol* **67**, 3338-3344.

Wang, M., Tan, Y., Yin, F., Deng, F., Vlak, J. M., Hu, Z. & Wang, H. (2008). The F-like protein Ac23 enhances the infectivity of the budded virus of gp64-null Autographa californica multinucleocapsid nucleopolyhedrovirus pseudotyped with baculovirus envelope fusion protein F. *J Virol* **82**, 9800-9804.

Wang, M., Wang, J., Yin, F., Tan, Y., Deng, F., Chen, X., Jehle, J. A., Vlak, J. M., Hu, Z. & Wang, H. (2014). Unraveling the entry mechanism of baculoviruses and its evolutionary implications. *J Virol* **88**, 2301-2311.

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.

Weatherford, T., Chavez, D., Brasky, K. M. & Lanford, R. E. (2009). The marmoset model of GB virus B infections: adaptation to host phenotypic variation. *J Virol* **83**, 5806-5814.

Weiner, A. J., Brauer, M. J., Rosenblatt, J., Richman, K. H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q. L., Houghton, M. & et al. (1991). Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. *Virology* **180**, 842-848.

Weiner, A. J., Geysen, H. M., Christopherson, C., Hall, J. E., Mason, T. J., Saracco, G., Bonino, F., Crawford, K., Marion, C. D., Crawford, K. A. & et al. (1992). Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc Natl Acad Sci U S A* **89**, 3468-3472.

Welch, B. D., Liu, Y., Kors, C. A., Leser, G. P., Jardetzky, T. S. & Lamb, R. A. (2012). Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein. *Proc Natl Acad Sci U S A* **109**, 16672-16677.

Wen, X., Krause, J. C., Leser, G. P., Cox, R. G., Lamb, R. A., Williams, J. V., Crowe, J. E., Jr. & Jardetzky, T. S. (2012). Structure of the human metapneumovirus fusion protein with neutralizing antibody identifies a pneumovirus antigenic site. *Nat Struct Mol Biol* **19**, 461-463.

Westenberg, M., Uijtdewilligen, P. & Vlak, J. M. (2007). Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. *J Gen Virol* **88**, 3302-3306.

Westenberg, M., Veenman, F., Roode, E. C., Goldbach, R. W., Vlak, J. M. & Zuidema, D. (2004). Functional analysis of the putative fusion domain of the baculovirus envelope fusion protein F. *J Virol* **78**, 6946-6954.

Westenberg, M., Wang, H., WF, I. J., Goldbach, R. W., Vlak, J. M. & Zuidema, D. (2002). Furin is involved in baculovirus envelope fusion protein activation. *J Virol* **76**, 178-184.

WF, I. J., Westenberg, M., Goldbach, R. W., Blissard, G. W., Vlak, J. M. & Zuidema, D. (2000). A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. *Virology* **275**, 30-41.

White, J. M. (1992). Membrane fusion. *Science* **258**, 917-924.

White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* **43**, 189-219.

Whitford, M., Stewart, S., Kuzio, J. & Faulkner, P. (1989). Identification and sequence analysis of a gene encoding gp67, an abundant envelope glycoprotein of the baculovirus Autographa californica nuclear polyhedrosis virus. *J Virol* **63**, 1393-1399.

Wickham, T. J., Shuler, M. L., Hammer, D. A., Granados, R. R. & Wood, H. A. (1992). Equilibrium and kinetic analysis of Autographa californica nuclear

polyhedrosis virus attachment to different insect cell lines. *J Gen Virol* **73 (Pt 12)**, 3185-3194.

Wiley, D. C. & Skehel, J. J. (1987). The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem* **56**, 365-394.

Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature* **289**, 366-373.

Witteveldt, J., Evans, M. J., Bitzegeio, J., Koutsoudakis, G., Owsianka, A. M., Angus, A. G., Keck, Z. Y., Foung, S. K., Pietschmann, T., Rice, C. M. & Patel, A. H. (2009). CD81 is dispensable for hepatitis C virus cell-to-cell transmission in hepatoma cells. *J Gen Virol* **90**, 48-58.

Woollard, D. J., Haqshenas, G., Dong, X., Pratt, B. F., Kent, S. J. & Gowans, E. J. (2008). Virus-specific T-cell immunity correlates with control of GB virus B infection in marmosets. *J Virol* **82**, 3054-3060.

Wurdinger, T., Verheije, M. H., Broen, K., Bosch, B. J., Haijema, B. J., de Haan, C. A., van Beusechem, V. W., Gerritsen, W. R. & Rottier, P. J. (2005). Soluble receptor-mediated targeting of mouse hepatitis coronavirus to the human epidermal growth factor receptor. *J Virol* **79**, 15314-15322.

Wyatt, P. J. (1998). Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation. *J Colloid Interface Sci* **197**, 9-20.

Wyles, D. L. (2012). Beyond telaprevir and boceprevir: resistance and new agents for hepatitis C virus infection. *Top Antivir Med* **20**, 139-145.

Xie, Z. C., Riezu-Boj, J. I., Lasarte, J. J., Guillen, J., Su, J. H., Civeira, M. P. & Prieto, J. (1998). Transmission of hepatitis C virus infection to tree shrews. *Virology* **244**, 513-520.

Yanagi, M., Purcell, R. H., Emerson, S. U. & Bukh, J. (1997). Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proc Natl Acad Sci U S A* **94**, 8738-8743.

Yi, M., Ma, Y., Yates, J. & Lemon, S. M. (2007). Compensatory mutations in E1, p7, NS2, and NS3 enhance yields of cell culture-infectious intergenotypic chimeric hepatitis C virus. *J Virol* **81**, 629-638.

Yin, H. S., Paterson, R. G., Wen, X., Lamb, R. A. & Jardetzky, T. S. (2005). Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc Natl Acad Sci U S A* **102**, 9288-9293.

Yin, H. S., Wen, X., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. (2006). Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature* **439**, 38-44.

Youn, J. W., Park, S. H., Lavillette, D., Cosset, F. L., Yang, S. H., Lee, C. G., Jin, H. T., Kim, C. M., Shata, M. T., Lee, D. H., Pfahler, W., Prince, A. M. & Sung, Y. C. (2005). Sustained E2 antibody response correlates with reduced peak viremia after hepatitis C virus infection in the chimpanzee. *Hepatology* **42**, 1429-1436.

Yuan, X., Gubbins, M. J. & Berry, J. D. (2004). A simple and rapid protocol for the sequence determination of functional kappa light chain cDNAs from aberrant-chain-positive murine hybridomas. *J Immunol Methods* **294**, 199-207.

Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., Miranda, E., Ordonez, A., Hannan, N. R., Rouhani, F. J., Darche, S., Alexander, G., Marciniak, S. J., Fusaki, N., Hasegawa, M., Holmes, M. C., Di Santo, J. P., Lomas, D. A., Bradley, A. & Vallier, L. (2011). Targeted

gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391-394.

Zhang, J., Randall, G., Higginbottom, A., Monk, P., Rice, C. M. & McKeating, J. A. (2004). CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J Virol* **78**, 1448-1455.

Zhang, P., Wu, C. G., Mihalik, K., Virata-Theimer, M. L., Yu, M. Y., Alter, H. J. & Feinstone, S. M. (2007). Hepatitis C virus epitope-specific neutralizing antibodies in Igs prepared from human plasma. *Proc Natl Acad Sci U S A* **104**, 8449-8454.

Zhang, P., Zhong, L., Struble, E. B., Watanabe, H., Kachko, A., Mihalik, K., Virata-Theimer, M. L., Alter, H. J., Feinstone, S. & Major, M. (2009). Depletion of interfering antibodies in chronic hepatitis C patients and vaccinated chimpanzees reveals broad cross-genotype neutralizing activity. *Proc Natl Acad Sci U S A* **106**, 7537-7541.

Zhao, Z., Zhong, L., Elrod, E., Struble, E., Ma, L., Yan, H., Harman, C., Deng, L., Virata-Theimer, M. L., Liu, P., Alter, H., Grakoui, A. & Zhang, P. (2014). A neutralization epitope in the hepatitis C virus E2 glycoprotein interacts with host entry factor CD81. *PLoS One* **9**, e84346.

Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D. R., Wieland, S. F., Uprichard, S. L., Wakita, T. & Chisari, F. V. (2005). Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* **102**, 9294-9299.

Zhu, D. W., Campbell, R., Labrie, F. & Lin, S. X. (1999). Crystallization and preliminary crystal structure of the complex of 17beta-hydroxysteroid dehydrogenase with a dual-site inhibitor. *J Steroid Biochem Mol Biol* **70**, 229-235.

Zhu, Y. & Chen, S. (2013). Antiviral treatment of hepatitis C virus infection and factors affecting efficacy. *World J Gastroenterol* **19**, 8963-8973.

Zibert, A., Meisel, H., Kraas, W., Schulz, A., Jung, G. & Roggendorf, M. (1997). Early antibody response against hypervariable region 1 is associated with acute self-limiting infections of hepatitis C virus. *Hepatology* **25**, 1245-1249.

Zucchelli, S., Roccasecca, R., Meola, A., Ercole, B. B., Tafi, R., Dubuisson, J., Galfre, G., Cortese, R. & Nicosia, A. (2001). Mimotopes of the hepatitis C virus hypervariable region 1, but not the natural sequences, induce cross-reactive antibody response by genetic immunization. *Hepatology* **33**, 692-703.

Appendix

Crystallography techniques and terms used in the thesis

Vapor diffusion crystallization

Vapor diffusion is the most frequently used method of protein crystallization. In this technique, a small droplet containing purified protein is mixed with a similar volume of crystallization solution typically containing a buffer, a precipitant and salt. The resulting droplet is allowed to equilibrate in a reservoir containing the same crystallization solution as added to the protein droplet.

Initially, the droplet of protein solution contains a lower concentration of precipitant and/or salt than the reservoir solution, therefore, water evaporates from the drop into the reservoir. This results in a gradual increase of both protein and precipitant in the droplet which may lead to crystal growth.

The two most common formats of vapor diffusion are hanging-drop and sitting-drop. In the case of the hanging-drop format a droplet of protein solution is placed on a siliconized cover slip, which is then inverted and sealed over a reservoir containing the crystallization solution. Sitting-drop crystallization set up involves placing a droplet on a small pedestal surrounded by the crystallization solution and then the chamber is sealed. Crystallization robots used for setting up crystallization droplets are very useful for screening a large number of crystallization conditions. The main advantage of these robots is that they can handle very small volumes of protein solution, and the reproducibility of the results.

Non-Crystallographic Symmetry

Non-crystallographic symmetry (NCS) exists when more than one copy of a molecule is present in the asymmetric unit of the crystal (for example, more than one

protomer, complex, etc). These objects usually adopt the same folds which can be exploited in density modification procedures and structure refinement.

The Matthews Coefficient

The Matthews Coefficient V_m is the ratio between the volume of the asymmetric crystallographic unit and the molecular weight of the molecules in the unit cell:

$$V_m = V(\text{unit cell})/MW \times Z \times X$$

Z is the number of asymmetric units in the unit cell, X is the number of molecules in the asymmetric unit, and MW is the molecular weight of a macromolecule (Matthews, 1968). It is measured in $Å^3Da^{-1}$ and usually values 1.5-6 $Å^3Da^{-1}$.V_m can be easily converted to the solvent volume of the crystal by the simple relationship:

$$x(p) = 1.66 \times \upsilon / V_m$$

x(p) is the fraction of the asymmetric volume occupied by a protein molecule and υ is the partial specific volume of the protein which is 0.74 cm^3/g for proteins. As a result, the fraction of solvent can be calculated as:

$$x(s) = 1 - x(p)$$

Proteins having a lower solvent content tend to diffract better.

Self-rotation function

The self-rotation function is a means of detecting if the molecules in the asymmetric unit are related by rotational symmetry axis. A map in stereographic projection of self-rotation function may provide information about the oligomeric

state of the crystallized protein as proper rotational NCS results in peaks in the $\kappa=360/n$ degree section. However, the peaks arising from NCS are sometimes difficult to distinguish from the peaks resulting from crystallographic symmetry. The calculation of self-rotation function requires unit cell parameters, space group, and observed structure factors.

Solving the phase problem

Single-wavelength and multi-wavelength anomalous diffraction

Single-wavelength and multi-wavelength anomalous diffraction/dispersion (SAD/MAD) are techniques for phase angle determination in X-ray crystallography. Both SAD and MAD exploit anomalous scattering of certain atoms near their X-ray absorption edges. If a wavelength near the element absorption edge is used, the recorded intensities of symmetry related reflections hkl and -h-k-l (called Friedel's pairs) are no longer equal.

Anomalous scattering is stronger for the heavier elements than for the light elements in the periodic system. Thus, the anomalous scattering of typical protein atoms such as nitrogen, carbon and oxygen do not contribute to anomalous scattering at the wavelengths used in X-ray crystallography. In order to observe anomalous diffraction, heavy atoms are introduced into protein crystals. When X-ray wavelength close to the element absorption edge is used, the obtained difference in intensity between Friedel's pairs can be exploited for determination of initial phases for a structure.

MAD requires chosing the wavelenghts for data collection, which optimize the difference of intensities between Friedel's pairs (anomalous differences) and the difference of the same intensity at different X-ray energies (dispersive differences). Typically MAD data are collected for at least two wavelengths (**Figure 96**): the absorption (f'') peak ((λ_1), and at the point of inflection on the absorption curve (λ_2) where the dispersive term f' (a derivative of the f'' curve) has its minimum. If the third wavelength is used, data are usually collected at a remote wavelength (λ_3) to maximize the dispersive difference to λ_2. Since the absorption of the heavy atom is affected by its environment within the protein, the absorption curve should be

recorded by performing a fluorescence scan on the crystal at a synchrotron. The absorption curve for selenomethionine recorded on SeFet crystal with incorporated SeMet is showed in Figure X.

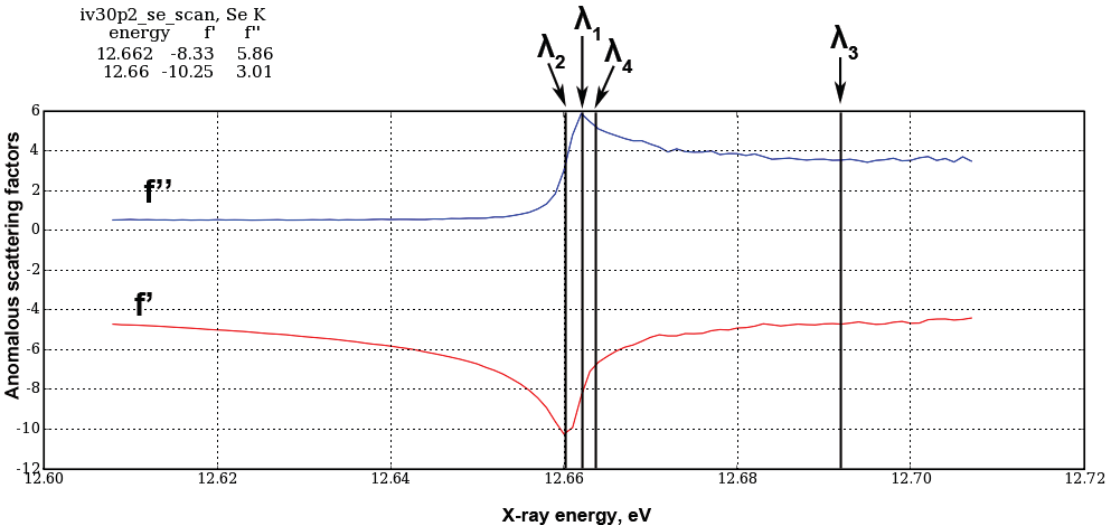


Figure 96. Fluorescense scan of SeFet crystal derivativized with SeMet in order to determine experimental values for f' and f'' as a function of X-ray energy. X-ray energy in keV = 12.398/λ in Å. A MAD data set is typically collected at three wavelengths: the absorption (f'') peak ((λ_1), the point of inflection on the absorption curve, and at a remote wavelength (λ_3). A double inflection MAD data set is collected at the point of rising inflection (λ_2) and at the point of falling inflection (λ_4).

Anomalous intensity differences are usually small (just a few percentage points), therefore, high multiplicity of the measurements and data completeness are important in order to increase the statistical significance. In processing MAD data, the first step is locating the heavy atoms in the unit cell (the heavy atom substructure).

SAD data, in contrast to MAD data, are collected on a single wavelength (at the absorption (*f''*) peak (λ_1)). Thus, SAD provides only measurements for anomalous, or Friedel pair, differences. SAD requires resolving two ambiguities: phase angle ambiguity and substructure-handedness. The ambiguity of the phase remains in SAD due to two maxima in each probability distribution, however, it can be successfully broken by density-modification procedures. One advantage of this technique is that the crystal spends less time in the beam while collecting data, which reduces potential radiation damage of the crystal.

Moreover, SAD can also use the anomalous scattering coming from intrinsic scatterers present in the protein such as sulfur atoms in cysteines and methionines. This technique is referred as **sulfur-SAD (S-SAD)**. This requires very high precision

in the measurement of reflections since the difference in intensities between the Friedel's pairs resulting from the anomalous scattering of sulfur is very low.

Molecular replacement

Molecular replacement is a technique used to solve phase problem in X-ray crystallography and can be applied when a search model with sufficient structural similarity to the crystallized molecule is available. During molecular replacement the proper orientation and location of the search model in the target unit cell is determined. RMSD calculated for the main chain atoms between a search model and a target molecule should not exceed 1.5-2 Å.

Inverse-beam collection mode

Inverse-beam collection mode is often used to collect anomalous diffraction data. Typically a data set is divided into small wedges. A thin wedge of data (5-20 degrees) is collected at phi and phi+180 degrees and then the crystal is rotated to record the equivalent wedge away from the current wedge (**Table 1**). If MAD is used it is combined with cycling between different wavelengths. The advantage of this collection mode is that the reflections hkl and -h-k-l are recorded close in time which results in more precise measurement of the intensity difference between Friedel's pairs.

Table 15. Inverse-beam collection mode example.

Start	Wedge	Energy
Phi=0 ^{o}	0-10^{o}	Peak
Phi=180 ^{o}	180-190 ^{o}	Peak
Phi=0 ^{o}	0-10^{o}	Inflection
Phi=180 ^{o}	180-190 ^{o}	Inflection
Phi=10 ^{o}	10-20^{o}	Peak
Phi=190 ^{o}	190-200 ^{o}	Peak
Phi=10 ^{o}	0-10^{o}	Inflection
Phi=190 ^{o}	190-200 ^{o}	Inflection
	Etc.	

Double inflection MAD

Double inflection MAD is a modified version of the traditional MAD data set collection mode. In this case, the MAD data set is collected on two inflection points: rising infection (λ_2) and falling inflection (λ_4). Double inflection MAD has been described to have certain advantages over traditional MAD data set collection mode and may result in better experimental maps (Evans & Wilson, 1999).

Density modification techniques

Density modification is a set of techniques for improving the quality of electron density maps. These techniques usually exploit prior knowledge of the "correct electron density map". Density modification techniques aim to improve phase estimates at moderate resolution and can be very useful with experimental (MAD, SAD, MIR etc) phases. In addition, it could also help to reduce phase bias of molecular replacement solutions.

Solvent-flattening density modification

Solvent flattening is based on the fact that the electron density is rather flat in the solvent region (due to thermal motion and disorder of solvent molecules) and contains little variation. Therefore, the electron density within this region can be set to a constant value (typical value is $0.33e^-/Å^{3)}$. The method requires defining the protein–solvent boundary, which was developed by Wang (Wang, 1985).

NCS map averaging

NCS map averaging is a density modification technique used to improve the initial electron density maps obtained by experimental phasing when an electron density is averaged across NCS related objects. NCS averaging requires defining NCS, i.e. finding NCS operators and defining envelopes.

Multi-crystal averaging

Multi-crystal (or cross-crystal) averaging is a density modification technique when averaging of electron density from the same molecule is performed between different crystal forms or non-isomorphous crystals.

TLS (Translation/Libration/Screw) refinement

In refinement of macromolecular structures it is important to retain a realistic data to parameter ratio. Therefore, the choice of atomic displacement parameters (the temperature factors or B-factors), which describethe vibration of an atom, should be carefully considered. Typically isotropic temperature factors are used to describe the motion of atoms in crystals that diffract to lower than 2 Å resolution assuming that atoms vibrate the same in all directions, which results in only one parameter defining the temperature factor.

Nevertheless, atoms in crystals diffracting to lower resolution also move anisotropically, i.e. not equally to all directions. Moving to anisotropic temperature factor for individual atoms, however, results in a six-fold increase of the number of parameters. TLS refinement is based on defining an entire molecule or a domain as a TLS group. This provides a good approximation of anisotropy but describes anisotropic motion with much fewer parameters (20 for an entire group of atoms versus six for an individual atom). TLS refinement has been described to improve R and R free values by several percentage points.

Table of Figures

Table of Tables

Abstract

Viral glycoproteins are responsible for the two major steps in entry into host cells by enveloped viruses: 1) attachment to cellular receptor/s and 2) fusion of the viral and cellular membranes.

My thesis concentrated first on the structural analysis of the major envelope glycoprotein E2 of two hepaciviruses: GB virus B (GBV-B) and hepatitis C virus (HCV). Crystallization of the GBV-B E2 ectodomain remained unsuccessful, but the characterization of truncated versions of E2 suggested an important role of its C-terminal moiety in receptor binding. In parallel, I co-crystallized a synthetic peptide mimicking HCV E2 with an antibody fragment directed against the major receptor-binding loop of E2 that is targeted by broadly neutralizing antibodies. The structure unexpectedly revealed an α-helical peptide conformation, which is in stark contrast to the extended conformation of this region observed in the structure of an E2 core fragment. Together with further biochemical evidence this suggests an unanticipated structural flexibility within this region in the context of the soluble E2 ectodomain.

Secondly, I focused on the structural analysis of the baculovirus glycoprotein F. I determined the crystal structure of the post-fusion trimer of a trypsin-truncated F fragment. This structure confirmed previous predictions that baculovirus F protein adopts a class I fusion protein fold and is homologous to the paramyxovirus F protein. Baculovirus F is therefore the first class I fusion protein encoded by a DNA virus. My results support the hypothesis that F proteins may have a common ancestor and imply interesting evolutionary links between DNA and RNA viruses and their hosts.

Résumé

Les glycoprotéines virales sont impliquées dans les deux principales étapes permettant aux virus enveloppés de pénétrer leurs cellules hôtes : l'attachement des virus aux récepteurs cellulaires et la fusion des membranes virale et cellulaire.

Lors de ma thèse, je me suis tout d'abord attachée à l'étude structurale de la principale glycoprotéine, E2, de deux hépacivirus : la forme B du virus GB (GBV-B) et le virus de l'hépatite C (HCV). Mes tentatives de cristallisation de l'ectodomaine de la protéine E2 du GBV-B sont restées vaines, mais l'analyse des propriétés de fragments de la protéine a suggéré un rôle de son extrémité C-terminale dans la liaison à son récepteur. En parallèle, j'ai co-cristallisé un fragment synthétique de la protéine E2 du HCV avec un fragment d'anticorps dirigé contre la principale boucle de liaison à son récepteur, cible de nombreux anticorps neutralisants. De façon surprenante le peptide forme une hélice α, en nette contradiction avec la conformation étendue qu'il adopte dans un fragment du cœur de E2. Associé à des données biochimiques, cela suggère une flexibilité inattendue de cette région dans le contexte de l'ectodomaine d'E2.

Dans un second temps, je me suis intéressée à la glycoprotéine F des baculovirus. J'ai résolu la structure du trimère d'un fragment tryptique de F dans sa conformation post-fusion. Cette structure a validé une prédiction selon laquelle la protéine F était une protéine de fusion de classe I homologue a celle des paramyxovirus. La protéine F des baculovirus est donc le premier exemple d'une protéine de fusion de classe I encodée par un virus à ADN. Mes résultats confortent donc l'hypothèse que toutes les protéines F ont un ancêtre commun et suggèrent un lien évolutif intéressant entre les virus à ADN, à ARN et leurs hôtes.