



**HAL**  
open science

## The disulfide bonds in glycoprotein e2 of hepatitis C virus reveal the tertiary organization of the molecule.

Thomas Krey, Jacques d'Alayer, Carlos M. Kikuti, Aure Saulnier, Laurence Damier-Piolle, Isabelle Petitpas, Daniel X. Johansson, Rajiv G. Tawar, Bruno Baron, Bruno Robert, et al.

### ► To cite this version:

Thomas Krey, Jacques d'Alayer, Carlos M. Kikuti, Aure Saulnier, Laurence Damier-Piolle, et al.. The disulfide bonds in glycoprotein e2 of hepatitis C virus reveal the tertiary organization of the molecule.. PLoS Pathogens, 2010, 6 (2), pp.e1000762. 10.1371/journal.ppat.1000762 . pasteur-00460826

**HAL Id: pasteur-00460826**

**<https://pasteur.hal.science/pasteur-00460826>**

Submitted on 2 Mar 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Disulfide Bonds in Glycoprotein E2 of Hepatitis C Virus Reveal the Tertiary Organization of the Molecule

Thomas Krey<sup>1</sup>, Jacques d'Alayer<sup>2</sup>, Carlos M. Kikuti<sup>1</sup>, Aure Saulnier<sup>3</sup>, Laurence Damier-Piolle<sup>1</sup>, Isabelle Petitpas<sup>4</sup>, Daniel X. Johansson<sup>5</sup>, Rajiv G. Tawar<sup>1</sup>, Bruno Baron<sup>6</sup>, Bruno Robert<sup>7</sup>, Patrick England<sup>6</sup>, Mats A. A. Persson<sup>5</sup>, Annette Martin<sup>3</sup>, Félix A. Rey<sup>1\*</sup>

**1** Institut Pasteur, CNRS URA3015, Unité de Virologie Structurale, Paris, France, **2** Institut Pasteur, Plate-forme d'Analyse et de Microséquençage des Protéines, Paris, France, **3** Institut Pasteur, CNRS URA3015, Unité de Génétique Moléculaire des Virus à ARN, Paris, France, **4** CNRS UPR3296, Laboratoire de Virologie Moléculaire et Structurale, Gif-sur-Yvette, France, **5** Karolinska University Hospital, Department of Medicine, Center for Molecular Medicine, Stockholm, Sweden, **6** Institut Pasteur, CNRS URA2185, Plate-forme de Biophysique des Macromolécules et de leurs Interactions, Paris, France, **7** CEA, CNRS URA2096, Institut de Biologie et de Technologie de Saclay, Gif-sur-Yvette, France

## Abstract

Hepatitis C virus (HCV), a major cause of chronic liver disease in humans, is the focus of intense research efforts worldwide. Yet structural data on the viral envelope glycoproteins E1 and E2 are scarce, in spite of their essential role in the viral life cycle. To obtain more information, we developed an efficient production system of recombinant E2 ectodomain (E2e), truncated immediately upstream its trans-membrane (TM) region, using *Drosophila melanogaster* cells. This system yields a majority of monomeric protein, which can be readily separated chromatographically from contaminating disulfide-linked aggregates. The isolated monomeric E2e reacts with a number of conformation-sensitive monoclonal antibodies, binds the soluble CD81 large external loop and efficiently inhibits infection of Huh7.5 cells by infectious HCV particles (HCVcc) in a dose-dependent manner, suggesting that it adopts a native conformation. These properties of E2e led us to experimentally determine the connectivity of its 9 disulfide bonds, which are strictly conserved across HCV genotypes. Furthermore, circular dichroism combined with infrared spectroscopy analyses revealed the secondary structure contents of E2e, indicating in particular about 28%  $\beta$ -sheet, in agreement with the consensus secondary structure predictions. The disulfide connectivity pattern, together with data on the CD81 binding site and reported E2 deletion mutants, enabled the threading of the E2e polypeptide chain onto the structural template of class II fusion proteins of related flavi- and alphaviruses. The resulting model of the tertiary organization of E2 gives key information on the antigenicity determinants of the virus, maps the receptor binding site to the interface of domains I and III, and provides insight into the nature of a putative fusogenic conformational change.

**Citation:** Krey T, d'Alayer J, Kikuti CM, Saulnier A, Damier-Piolle L, et al. (2010) The Disulfide Bonds in Glycoprotein E2 of Hepatitis C Virus Reveal the Tertiary Organization of the Molecule. *PLoS Pathog* 6(2): e1000762. doi:10.1371/journal.ppat.1000762

**Editor:** Charles M. Rice, The Rockefeller University, United States of America

**Received:** September 3, 2009; **Accepted:** January 12, 2010; **Published:** February 19, 2010

**Copyright:** © 2010 Krey et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** T.K. benefited from an Institut Pasteur "Bourse Roux" and then from an EU Marie Curie fellowship (MEIF-CT-2007-04725). C.M.K. and A.S. benefited from an ANRS post-doctoral fellowship. R.G.T. is funded by the E.U. HCVEI NTIR consortium. F.A.R. acknowledges support from ANRS, Institut Pasteur, CNRS, the European Union, TOTAL and Merck-Serono. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Thomas Krey, Carlos M. Kikuti, Laurence Damier-Piolle and Félix A. Rey are listed in a patent application in the field of "HCV-derived polypeptides" that has been filed and is pending.

\* E-mail: rey@pasteur.fr

‡ Current address: Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

## Introduction

The hepatitis C virus (HCV) is a major cause of chronic liver disease worldwide, leading to cirrhosis and hepatocellular carcinoma [1]. In spite of being the focus of intense research efforts, no vaccine is available against HCV, and current therapeutic treatments have limited efficacy and significant side effects [2]. HCV belongs to the *Flaviviridae* family of enveloped, positive-strand RNA viruses [3]. Structural studies on this virus are difficult, in part because it propagates poorly in cell culture, and particles isolated from infected patients are heterogeneous and not amenable to a detailed structural characterization. Little structural information is available on the envelope proteins, which are heavily glycosylated, display hypervariable loops, and are stabilized by numerous disulfide bridges [4]. The folding kinetics

of these proteins are slow, requiring several hours for completion of a complex process involving various ER chaperones of the infected cell [5]. These properties make their recombinant production - in a native conformation and in sufficient amounts for structural studies - a difficult endeavor. Yet structural information on the HCV envelope proteins would be extremely valuable, given that they carry the main antigenic determinants of the virus and play an essential role in cell entry by binding to specific receptors and inducing membrane fusion. HCV has indeed been shown to depend on a number of cellular molecules for entry, including CD81 [6] and the tight junction transmembrane proteins claudin 1, 6, 9 and occludin [7–9], as well as the scavenger receptor B1 (SR-B1) [10]. The LDL receptor also plays a role in HCV uptake, in line with the observation that HCV particles in infected plasma are associated with LDL species [11].

## Author Summary

Little is known about the structure of the envelope glycoproteins of the hepatitis C virus (HCV), in spite of their essential role in the viral cycle of this major human pathogen. Here, we determined the connectivity of the 9 disulfide bonds formed by the strictly conserved 18 cysteines of the ectodomain of HCV glycoprotein E2. We show that this information, together with important functional data available in the literature, impose important restrictions to the possible three-dimensional fold of the molecule. Indeed, these constraints allow the unambiguous threading of the predicted secondary structure elements along the polypeptide chain onto the template provided by the crystal structures of related flavi- and alphavirus class II fusion proteins. The resulting model of the tertiary organization of E2 shows the amino acid distribution among the characteristic class II domains, places the CD81 binding site at the interface of domains I and III, and highlights the location of a candidate fusion loop.

A direct interaction of the HCV envelope protein E2 with CD81 and SR-B1 has been demonstrated, and these interactions were shown to be necessary but not sufficient for cell entry. The mode of interaction of HCV with the claudins and occludin is not understood at present.

The HCV genome codes for a single polyprotein precursor about 3000 amino acids long, spanning the ER membrane multiple times. It contains, sequentially, the viral proteins in the order Nter-C-E1-E2-p7-NS2-NS3-NS4A/B-NS5A/B-Cter. The N-terminal 1/4<sup>th</sup> of the precursor corresponds to the structural proteins C (Core), E1 and E2 (envelope proteins 1 and 2) and p7, which functions as a proton channel. The remainder of the polyprotein contains the non-structural (NS) proteins, which have enzymatic and other activities that are necessary for virus replication. The mature viral proteins are generated by proteolytic processing of the precursor by cellular and viral proteases [3]. In particular, the envelope proteins are generated by host-cell signalases. E1 and E2 are type 1 trans-membrane (TM) proteins with a large N-terminal ectodomain and almost no cytoplasmic tail. In the best characterized HCV strain H77, E1 and E2 are 192 and 366 amino acids long and contain 6 and 11 potential N-linked glycosylation sites, respectively. Biochemical studies have shown that E1 and E2 fold as a heterodimer, which is found at the surface of viral particles and is thought to be the functional glycoprotein form [4]. There are currently 6 identified HCV genotypes further divided into several subtypes [12]. The amino acid sequence identity between envelope proteins from different genotypes is about 68% for the most distant genotypes. E2 has been shown to contain 3 hypervariable regions that can be deleted without affecting the overall fold of the protein, as assayed by binding to conformation-sensitive mAbs and CD81 [13–15].

The genomic organization of HCV is characteristic of all members of the *Flaviviridae* family [3]. In particular, the envelope proteins are present in tandem within the polyprotein precursor. This arrangement of the structural part of the genome is characteristic of viruses encoding class II fusion proteins, reviewed in [16]. These proteins have been extensively characterized, structurally and biochemically, for viruses in the flavivirus genus within the *Flaviviridae* family [17]. Class II proteins have a common tertiary structure, which has also been observed in the fusion protein of Semliki Forest virus (SFV), an alphavirus belonging to a separate family of enveloped, positive-strand RNA viruses, the

*Togaviridae* [18]. *Togaviridae* and *Flaviviridae* display the same gene order in the structural part of their genomes. There is no amino acid sequence similarity in the alpha- and flavivirus fusion proteins, however, and in spite of sharing a common fold, they are stabilized by a different pattern of disulfide bonds. Viruses within the *Flaviviridae* families have no sequence similarity across the various genera either, and the fusion proteins from each genus also appear to have their own characteristic pattern of disulfide bonds. Yet the conservation of the class II fold across viral families in the absence of sequence conservation strongly suggests that it is also conserved across the different genera within the respective families.

A further feature of class II viral fusion proteins is that they fold as a heterodimer with the upstream glycoprotein in the polyprotein precursor. This heterodimer later dissociates to drive membrane fusion upon interactions with the host cell. The first glycoprotein in the tandem thus acts as chaperone for folding the second one, which has the membrane fusion role. The chaperone function was experimentally demonstrated for the flavivirus prM [19] and the alphavirus p62 [20] glycoproteins, which precede the fusion proteins E and E1, respectively, in the precursor polyprotein. The effect on folding appears to be reciprocal, since both p62 and prM also adopt their native conformation only in presence of the respective accompanying fusion protein (unpublished observations). Importantly, heterodimerization upon folding has also been characterized for viruses belonging to other genera in the two families, and in particular for HCV [5,21].

Flavivirus E and alphavirus E1 change into a homotrimer upon interaction with lipids in the acidic environment of a target cell endosome, in a process that drives fusion of the viral and endosomal membrane and results in infection of the cell [22,23]. This process involves homodimer (E-E, flavivirus) or heterodimer (E2-E1, alphavirus) dissociation, followed by homotrimerization of E (flavivirus) or E1 (alphavirus) upon binding to lipids. The tertiary structure of class II viral fusion proteins contains predominantly  $\beta$ -sheets segregated into three distinct domains arranged linearly, resulting in a rod-like molecule. The central domain 1 (DI) is a  $\beta$ -sandwich with two long insertions in loops connecting adjacent  $\beta$ -strands. These insertions form an elongated “fusion” domain (DII), carrying the “fusion loop” in the first of the two insertions, at the distal end of the rod. The fusion loop is a segment of the polypeptide chain that inserts into the target membrane in the first step of membrane fusion. At its C-terminal end, DI is connected via a flexible linker to domain 3 (DIII), which is located at the opposite side with respect to DII, giving rise to the linear organization of the molecule. DIII plays an important role in the fusogenic conformational change, during which it relocates to the side of the molecule, resulting in the characteristic “hairpin” conformation of the protein, which drives membrane fusion. This relocation involves a considerable stretching of the segment connecting DI to DIII, the region that changes most dramatically in conformation during the fusogenic transition (reviewed in [16]).

Although there is no direct experimental evidence demonstrating the role of E2 as the HCV fusion protein, the compelling similarities to viruses with class II fusion proteins suggest that membrane fusion is at least one of its biological roles. It is worth noting, however, that while totally unrelated viruses can have structurally homologous fusion proteins (for example, rhabdoviruses, herpesviruses and baculoviruses, reviewed in [24]), related viruses can use non-homologous fusion proteins, as is the case with paramyxoviruses and rhabdoviruses, which belong to the *Mono-negavirales* order (reviewed in [25]). Yet the fact that viruses belonging to different genera in the *Flaviviridae* and *Togaviridae* families display a genomic arrangement that is the signature of

class II fusion proteins, together with the additional common features outlined above, makes it very likely that they code for envelope glycoproteins that are at least distantly related to class II proteins.

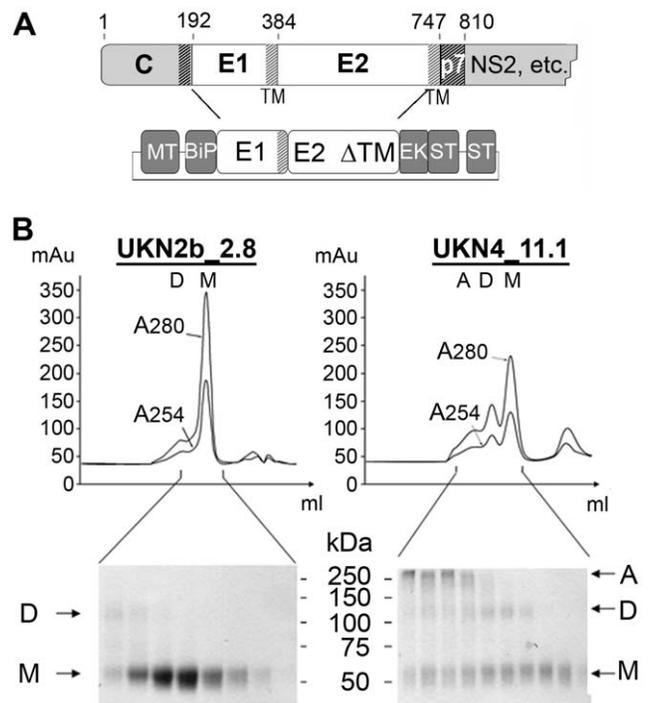
A model for E2 has actually been proposed based on the structure of the flavivirus E protein homodimer [26], although no evidence is available for homodimerization of HCV E2, which forms a heterodimer with E1 in infectious virions [4]. More importantly, this model does not take into account the location of the strictly conserved cysteine residues forming 9 disulfide bonds [27]. This model also lacks the third domain, which is important in the fusogenic transition. Moreover, it was also proposed that the membrane fusion function could be carried by HCV glycoprotein E1 (i.e., the first glycoprotein in the tandem) [28,29], in spite of the similarities with flavi- and alphaviruses discussed above, and in the absence of experimental support. Furthermore, a bioinformatics model for HCV E1 as a truncated class II protein was reported [30], postulating that E1 has the fold of DII of an alpha- or flavivirus fusion protein, but neglecting the fact that in class II proteins, DII works in conjunction with the other two domains covalently linked within the polypeptide to induce membrane fusion. The corollary is that controversial hypotheses have been reported concerning the identity of the HCV fusion protein. It is therefore important to stress that the structural studies performed over the years on viral membrane fusion proteins strongly suggest that most animal enveloped viruses encode fusion proteins belonging to one of the three currently characterized structural classes [31]. It is thus highly unlikely that HCV would have acquired a totally novel fusion machinery (for instance, one in which E1 would be the membrane fusion protein), especially when taking into account the similarities to class II proteins presented above.

In order to bring more insight into the tertiary structure of HCV E2, we report here the experimental identification of the connectivity of the 9 disulfide bonds present in the recombinant E2 ectodomain (E2e) generated by expression of the E1-E2 $\Delta$ TM portion of the HCV genome in *Drosophila* S2 cells (Fig. 1A). The absence of the transmembrane (TM) segment in E2 leads to secretion of its ectodomain after folding in the presence of E1. This approach is based on previous results leading to production of recombinant dengue virus E protein in the presence of its viral chaperone prM [32]. We tested the conformation of recombinant HCV E2e biochemically and functionally, showing that it reacts with conformation-sensitive antibodies and inhibits infection of Huh7.5 cells by infectious HCV particles (HCVcc) in a dose-dependent manner. Knowledge of the disulfide bonds, along with functional data on deletion mutants [14] and CD81 binding [26,33,34], together with secondary structure predictions, provide sufficient constraints to reconstitute the tertiary organization of the molecule. This information allowed the threading of the E2e polypeptide chain onto a class II template by matching the predicted  $\beta$ -strands. The resulting model reveals the distribution of the amino acids of HCV E2 among the different domains, maps the CD81 binding site to the DI/DIII interface, and highlights a strictly conserved segment of the polypeptide chain as a strong candidate for the HCV fusion loop.

## Results/Discussion

### Biochemical characterization of recombinant E2e

We generated stable *Drosophila* S2 cell-lines expressing the E1-E2 $\Delta$ TM segment of the precursor polyprotein (Fig. 1A) from 9 isolates spanning all 6 HCV genotypes and 4 subtypes (Table 1). In order to ensure that the recombinant E2 proteins were functional,



**Figure 1. Production and biochemical characterization of HCV E2e.** A) Schematic diagram of the HCV genome region coding for the structural proteins and the constructs used to make stable S2 cell transfectants expressing E2e. MT: inducible metallothionin promoter, BiP: *Drosophila* BiP Signal peptide, EK: enterokinase cleavage site, ST: Strep-Tag. B) Elution profile of E2e from the HCV isolates indicated from an Sdx200 size exclusion column. Bottom panels: Non-reducing SDS-PAGE analysis of the eluted fractions. Arrows indicate multimeric (A), dimeric (D) and monomeric (M) forms of the protein.  
doi:10.1371/journal.ppat.1000762.g001

we selected isolates previously tested for entry of retroviral particles pseudotyped with HCV glycoproteins (HCVpp) with the corresponding sequences [35]. Induction of expression at high cell density with CdCl<sub>2</sub> resulted in accumulation of relatively high levels of secreted E2e in the cell culture medium. We purified the protein to homogeneity from the supernatant (described in Text S1), with the yields listed in Table 1. E2e from the different isolates behaved similarly, as judged by size exclusion chromatography (SEC) followed by SDS-PAGE analysis under reducing and non-reducing conditions and Coomassie blue staining (Fig. 1). In a typical SEC profile, the majority of the protein elutes at a volume corresponding to a monomer, with additional minor peaks corresponding to disulfide linked dimers and higher multimers, which vary depending on the construct analyzed. The monomeric form was efficiently separated from the other species by pooling the corresponding fractions. Analytical ultracentrifugation and small angle X-ray scattering confirmed the monomeric state of the protein eluted in these fractions (data not shown). Once isolated, E2e from all constructs listed in Table 1 remained monomeric and showed no tendency to associate into disulfide-linked aggregates over time. The construct corresponding to the genotype 2b isolate (UKN2b\_2.8) reproducibly yielded the highest amounts of purified monomeric protein (Table 1). The construct from genotype 4 (UKN4\_11.1 isolate) yielded a significant fraction of disulfide-linked aggregates (Fig. 1), which are likely to correspond to misfolded protein. E2e from the remaining 7 constructs yielded slightly lower yields of purified, monomeric protein than did the genotype 2b construct, the lowest yields being from the genotype 6 isolate (Table 1). The SEC

**Table 1.** Production yields of E2e constructs from the 9 selected HCV isolates.

Strain/Isolate	Genotype	Genbank Accession	Average Yields *
H77	1a	GI:130461	230 µg/l
Con1	1b	GI:5420377	180 µg/l
JFH-1	2a	GI:116078059	200 µg/l
J6	2a	GI:221651	150 µg/l
UKN2b_2.8	2b	GI:58198335	300 µg/l
UKN3a_1.28	3a	GI:58198337	200 µg/l
UKN4_11.1	4	GI:58198341	190 µg/l
UKN5_14.4	5	GI:58220848	220 µg/l
UKN6_5.340	6	GI:58220846	70 µg/l

\*The yields specified correspond to the purified monomeric form of E2e. Significant differences in production yields ( $\pm 20\%$ ) were observed between independent preparations.

doi:10.1371/journal.ppat.1000762.t001

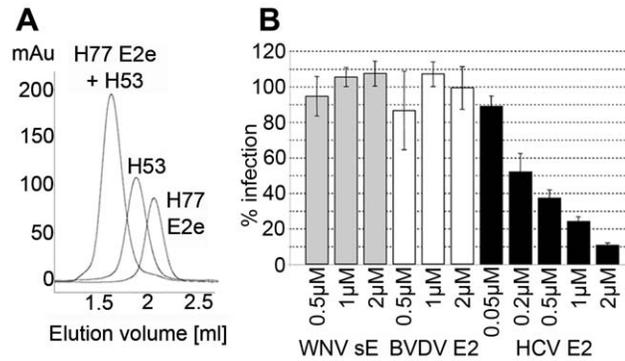
profiles from the 7 other constructs were intermediate between the two chromatograms shown in Fig. 1.

### Conformational characterization

Because of the higher production yields, we pursued most of the biochemical characterization using E2e from genotype 2b, to which we will refer to as E2e in the rest of the manuscript, except when explicitly stated. Yet because the best functionally characterized HCV strain is H77 (genotype 1a), we use the amino acid numbering corresponding to the H77 polyprotein throughout the manuscript.

Pull-down assays showed that E2e efficiently binds the CD81 large external loop (LEL), as well as conformation-sensitive mAbs CBH-4B and CBH-4D [36] (Fig. S1A). To further confirm that CD81 and the conformation-sensitive mAbs bind stoichiometrically to monomeric E2e, we used SEC to analyze the formation of various E2e/ligand complexes in defined ratios. The resulting chromatograms display a quantitative shift of the peak from monomeric protein to an E2e/mAb complex with a 2:1 stoichiometry, as expected (Figs. 2A and S1B). The SEC profile displayed in Fig. 2 shows the well-characterized conformation-sensitive mAb H53 that is specific for genotype 1a, whereas Fig. S1B shows the same analysis of E2e from the genotype 2b isolate and the human conformation-sensitive mAb CBH-4D. Similarly, SEC analysis using the Fab fragment of the corresponding mAbs under the same conditions, yielded a 1:1 E2e/Fab stoichiometry, as expected (data not shown). SEC analysis revealed that E2e also forms a stoichiometric complex with the CD81 LEL (data not shown). The monomeric fraction from all isolates listed in Table 1 yielded similar results - except perhaps for the genotype 6 isolate, which was not tested - strongly suggesting that recombinant E2e adopts a conformation closely resembling that of authentic E2 present on virions.

We further tested the ability of E2e to compete with infectious HCV particles for entry receptors, by measuring its ability to inhibit infection of Huh-7.5 cells by HCVcc (Fig. 2B). As a control, we tested in parallel the effect of the flavivirus E protein ectodomain (sE) from West Nile encephalitis virus (WNV), as well as the ectodomain of pestivirus E2 (pE2e) from the bovine viral diarrhea virus (BVDV) produced under identical conditions. In contrast to the control proteins, HCV E2e exerted a clear dose-dependent inhibition of the infection. At the lowest concentration



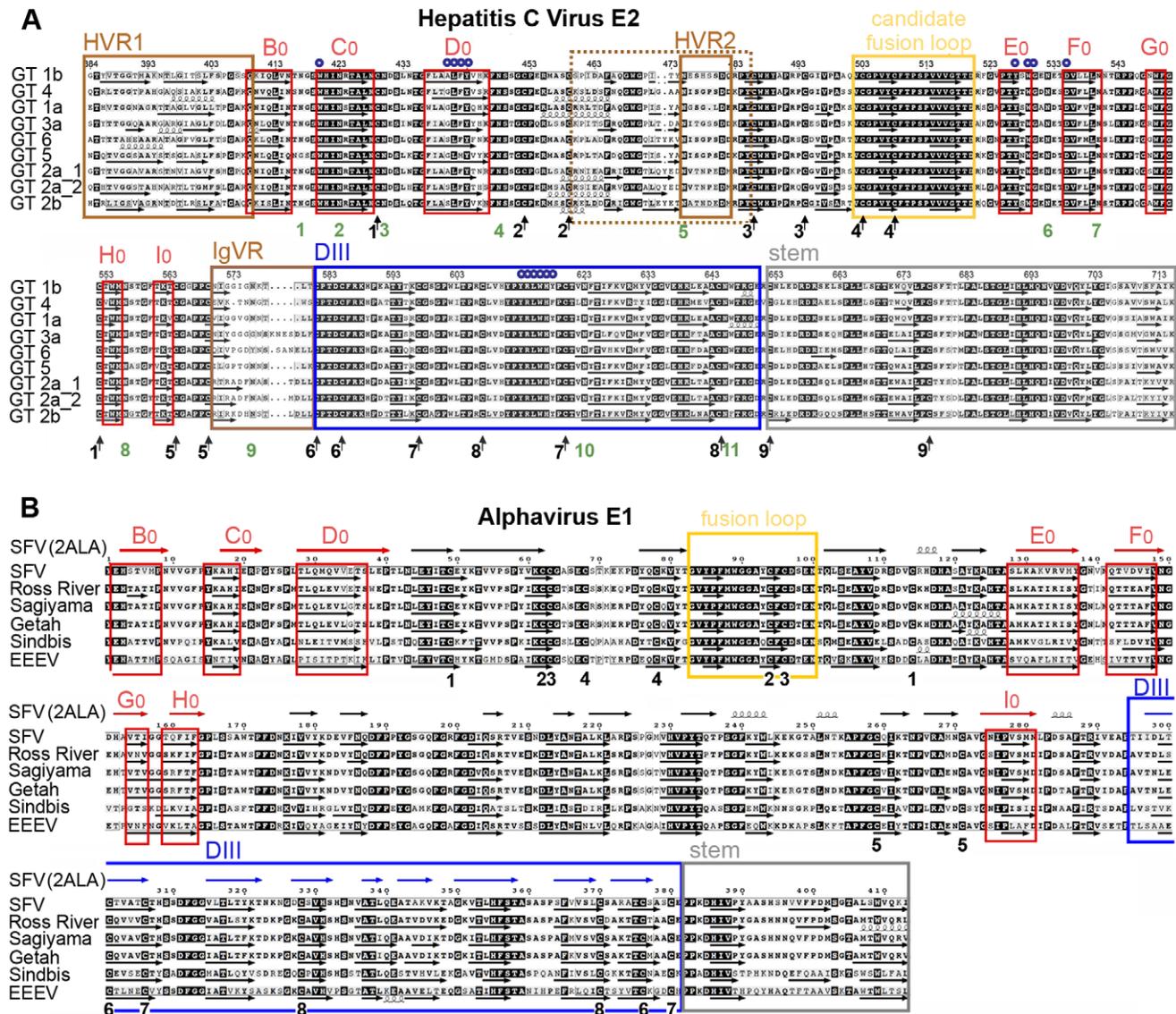
**Figure 2. Functional and conformational characterization of HCV E2e.** A) Stoichiometric complex formation between H77 E2e and mAb H53. H77-E2e, mAb H53 and a mixture of the two (molar ratio 2:1) were loaded to the column (in three different runs) (E2e~50kD, H53~150kD, complex~250kD). No peaks corresponding to either of the isolated proteins were observed in the profile of the complex, indicating a 2:1 complex stoichiometry and a high affinity of H77 E2e for mAb H53. B) Dose-dependent inhibition of infection of Huh7.5 cells by HCVcc. Huh-7.5 cells were preincubated with increasing concentrations of HCV E2e, WNV sE or BVDV E2e and subsequently infected with HCVcc in the corresponding recombinant protein concentration. The number of infected foci was determined after immunofluorescence analysis detecting intracellular HCV core antigen. The columns represent mean values of duplicates in a representative experiment; bars indicate mean deviation, 100% corresponds to the mean value of the infection in the presence of the control proteins.  
doi:10.1371/journal.ppat.1000762.g002

tested (0.05 µM), 10% inhibition was observed, which increased with protein concentration to reach 90% inhibition at 2 µM of HCV E2e. This effect is in line with the observation that E2e makes a stoichiometric complex with CD81, as described above.

### Secondary structure analyses

Computer algorithms for secondary structure prediction using amino acid alignments of E2 from all 6 HCV genotypes predict predominantly  $\beta$ -strands in E2e (Fig. 3), consistent with the fold of class II fusion proteins. We used recombinant E2e to experimentally analyze its secondary structure composition with two complementary methodologies, circular dichroism (CD), which is sensitive to the presence of  $\alpha$ -helices, and Fourier transform infrared (FTIR) spectroscopy, which readily detects  $\beta$ -sheets present in a protein. We carried out these tests in parallel with recombinant control class II envelope proteins of known structure available in the laboratory. For the CD measurements, the controls were WNV sE, ([37], PDB 2I69) and the ectodomain of glycoprotein E1 (sE1) of Chikungunya virus (CHIKV), which displays 62.5% amino acid sequence identity with the SFV E1 ectodomain, the crystal structure of which is known ([18], PDB 2Ala). Unexpectedly, the far-UV spectra of the three proteins exhibited considerable differences (Fig. 4A), the spectrum of HCV E2e being in agreement with a previous study [38]. However, deconvolution to retrieve the percentage of the various secondary-structure elements suggests similar ratios for all three proteins, indicating, in particular, only about 5%  $\alpha$ -helices in all three proteins. The strong minimum observed at 203 nm in the spectrum of HCV E2e suggests the presence of natively unfolded regions that are absent in the control proteins.

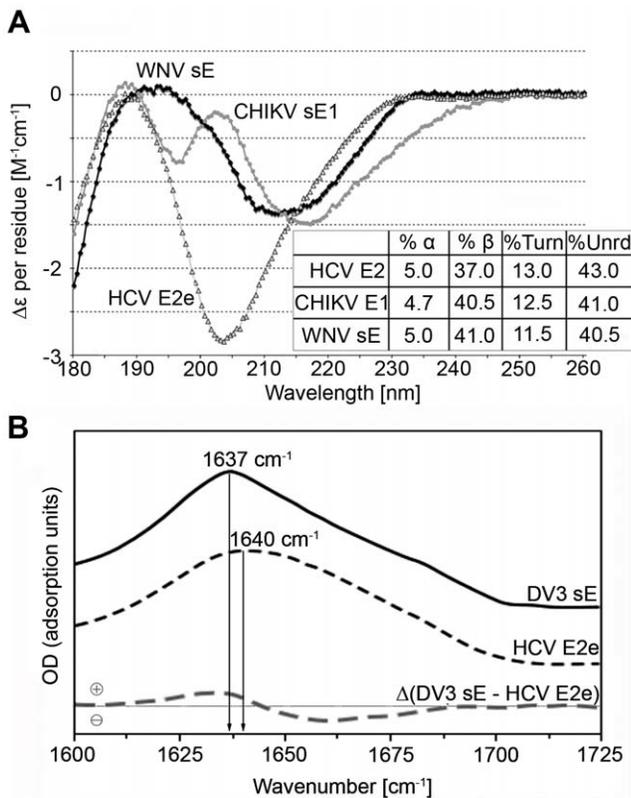
Given that circular dichroism is not the most sensitive method to determine the amount of  $\beta$ -sheet in a protein, we used FTIR spectrometry in a comparative analysis of E2e with a class II protein of known 3D structure. Because the WNV sE was not



**Figure 3. Amino acid sequence alignments and secondary structure predictions.** The secondary structure of HCV E2e and alphavirus sE1 was predicted using the program DSC [48], which was selected because the predictions matched more closely the crystallographically determined secondary structure elements of class II proteins. Arrows or spirals under each sequence indicate predicted  $\beta$ -strands or  $\alpha$ -helices, respectively. A) The main elements of the tertiary structure model of HCV E2 are framed: assigned strands in DI (red, labeled) and DIII (blue), putative fusion loop (yellow), the stem (grey) and regions that can be deleted without affecting the protein conformation (brown). The 18 cysteines, which form the 9 disulfides, are marked with arrows and numbered according to the disulfide bond (Table 2) under the sequences. N-linked glycosylation sites are numbered in green. Residues known to interact with CD81 are marked with small blue circles. The numbering corresponds to the HCV H77 polypeptide. B) Comparison with a class II fusion protein of known structure. The experimentally determined secondary structure of SFV E1 taken from the crystal structure (PDB 2ALA, [49]), shown with symbols (arrows or spirals) above the sequence alignment, was compared with secondary structure predictions for the E1 ectodomain of selected alphaviruses. Experimentally determined  $\beta$ -strands in DI, DII and DIII are colored red, black and blue, respectively. The red frames indicate the consensus predicted strands in DI, for easier comparison with panel A. Similarly the fusion loop (yellow), the region corresponding to the crystallographically identified DIII (blue), and the stem (grey) are framed. The 16 cysteines forming the 8 disulfides are numbered in black according to the disulfide bond under the sequences. The numbering starts with the first amino acid of the E1 glycoprotein. doi:10.1371/journal.ppat.1000762.g003

available at the time of the experiment, we used instead sE from dengue virus serotype 3 (DV3), for which the crystal structure is also known (PDB entry 1UZG [39]). The high-frequency region of the FTIR spectra of HCV E2e and DV3 sE is displayed in Fig. 4B. As expected, both proteins have their absorption maxima in the amide I band at  $1637$  and  $1640$   $\text{cm}^{-1}$ , respectively, close to the  $1630$   $\text{cm}^{-1}$  value typical for  $\beta$ -sheet containing polypeptides (reviewed in [40]), in agreement with the structure of the flavivirus sE and strongly indicating that HCV E2e also contains

predominantly  $\beta$ -sheets. In order to obtain a quantitative measure of the  $\beta$ -sheet content of E2e, we performed a further analysis to more precisely compare the secondary structure content of the two proteins by computing a difference spectrum after normalization to an identical area under the amide I band. The DV3/sE – HCV/E2e difference spectrum showed a positive peak at  $1630$   $\text{cm}^{-1}$ , as well as a broad negative region ranging from  $1645$  to  $1680$   $\text{cm}^{-1}$  (Fig. 4B). The value of the positive peak indicated about 14% higher  $\beta$ -sheet content for DV3 sE, which,



**Figure 4. Experimental analysis of the secondary structure of HCV E2e.** A) Far-UV CD spectra obtained with recombinant HCV E2e (empty triangles) and the controls CHIKV sE1 (grey circles) and WNV sE (black diamonds). The inset shows the estimated fraction of  $\alpha$ -helix,  $\beta$ -sheet, turns and unordered polypeptide chain for the 3 proteins. B) FTIR spectra of HCV E2e (dashed black line) and DV3 sE (solid black line) and the difference spectrum of the two after normalization (dashed grey line) in the amide I band region.

doi:10.1371/journal.ppat.1000762.g004

when using the value of 42%  $\beta$ -sheet estimated from the DV3 sE crystal structure, gives about 28%  $\beta$ -sheet for HCV E2e. The negative area of the difference FTIR spectrum indicates that the HCV E2e polypeptide displays higher relative amounts of secondary structure other than  $\beta$ -pleated sheet (random coil,  $\beta$ -turns, 3/10 helices, etc.). This difference is likely to reflect the presence of the regions that give rise to the strong minimum at 203 nm in the CD spectrum (Fig. 4A), i.e., natively unfolded segments of the polypeptide chain.

### Disulfide-bond connectivity

We determined the identity of the disulfide bridges by N-terminal sequencing together with comparative reducing/non-reducing mass spectrometry analyses of peptides obtained by trypsin digestion of E2e. For this purpose we selected E2e of three isolates, UKN2b\_2.8, H77 and JFH-1 (genotypes, 2b, 1a and 2a, respectively), which display amino acid sequences with a different pattern of predicted trypsin cleavage sites (Fig. S2). We fully deglycosylated the protein with PNGase F under denaturing conditions, then digested it with trypsin followed by separation of the resulting peptides by HPLC under reducing or non-reducing conditions. Comparison of the HPLC elution profiles enabled the identification of peaks that were affected by reduction with TCEP (asterisks in Fig. 5A). We analyzed the samples in these peaks by surface-enhanced laser desorption/ionization (SELDI) with a

time-of-flight (TOF) spectrometer (Table S1), and identified their N-terminal sequence by Edman degradation (Fig. S3). This procedure allowed the unambiguous experimental identification of 8 out of the 9 disulfide bonds present in the protein, thereby also identifying the 9<sup>th</sup> by exclusion (Table 2). This table also shows that 5 disulfides were independently identified in at least two different strains, validating the procedure. A full account of the experiments made to determine the disulfide connectivity is provided as Supplementary Information (Text S1).

### Implications for the tertiary structure of HCV E2e

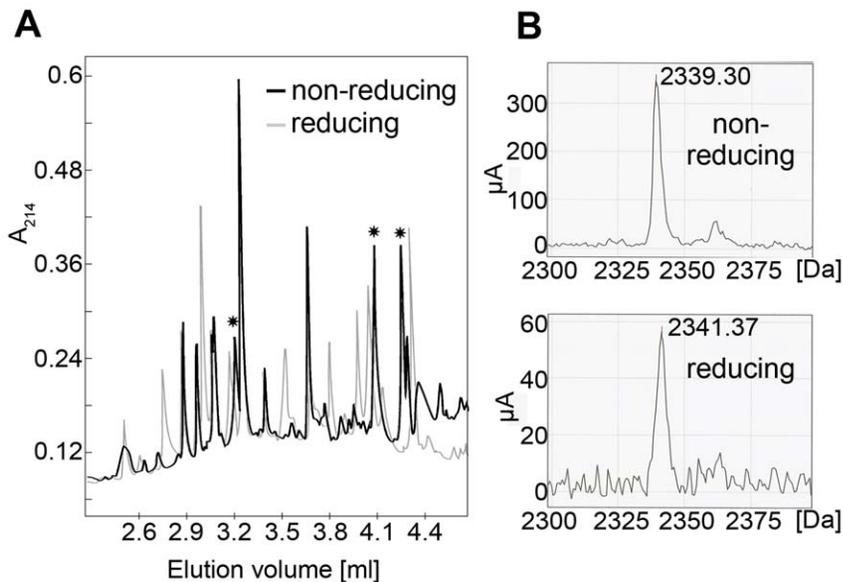
The connectivity of the disulfide bonds provides key information on distant segments of the E2 polypeptide chain that come near each other in the folded protein. This knowledge can be used in conjunction with other available data to get a better picture of the tertiary structure of the protein, namely: i) the observation that E2e is rich in  $\beta$ -sheet and that secondary structure predictions suggest regions with consensus  $\beta$ -strands along its amino acid sequence; ii) the identity of residues that are far apart in primary structure and that are known to be part of the CD81 binding site; iii) the postulate that E2 is the HCV fusion protein and therefore has a characteristic 3-domain class II fold, in agreement with the organization of its precursor polyprotein, which also implies iv) that the third domain (DIII) should be connected to DI via a linker that can extend to stabilize a post-fusion trimer. Finally, DIII should be followed by a flexible “stem” region - the presence of which has already been reported for HCV E2 [41] - connecting to the TM segment. Further information comes from the identification of “hypervariable” regions in HCV E2 that can be deleted without affecting the reactivity of the resulting deletion mutant with conformation-sensitive mAbs and with CD81 [14]. In addition, numerous reports have shown that the E2 ectodomain truncated at position 661, which is in the loop closed by disulfide 9 (Table 2), also reacts with conformation-sensitive mAbs and CD81 [38], suggesting that the downstream segment is not part of the structured ectodomain.

About one third of the E2e residues are predicted to form  $\beta$ -strands (Fig. 3), which is in overall agreement with the estimated 28%  $\beta$ -sheet content determined by FTIR spectroscopy. The pattern of predicted  $\beta$ -strands offers the possibility of threading the polypeptide chain along the template provided by the known fold of class II proteins, while simultaneously respecting all of the known constraints derived for HCV E2 by the functional studies discussed above. A useful guide for this analysis is the comparison between predicted and experimentally observed  $\beta$ -strands in the crystal structure of alpha- and flavivirus fusion proteins - for instance, in the alphavirus E1 alignment provided in Fig. 3B.

The hallmark of the tertiary structure of class II proteins is the presence of an 8-stranded ( $B_0$  through  $I_0$ ) central domain (or DI) folded as a  $\beta$ -sandwich with up-and-down topology (Fig. 6). Two insertions in this domain, in the  $D_0E_0$  and  $H_0I_0$  loops, constitute the fusion domain bearing the fusion loop in the distal part of the  $D_0E_0$  insertion. DI is followed, after strand  $I_0$ , by a flexible segment connecting to a third domain (DIII), the relocation of which is important for hairpin formation during the fusogenic conformational rearrangement of class II fusion proteins.

### Threading the HCV E2 polypeptide chain on the tertiary structure of a class II template

Functional studies have shown that deletion of the HVR1 region did not induce a loss of virus infectivity in experimentally infected chimpanzees [42], indicating that this segment cannot be part of a folded domain. We therefore began the threading process by assigning the 3 consecutive  $\beta$ -strands predicted immediately



**Figure 5. Disulfide mapping strategy.** A) Typical HPLC elution profile of a tryptic digest of E2e under non-reducing (black) and reducing (light grey) conditions, superposed to highlight the difference in mobility upon reduction. Asterisks mark peaks that disappeared upon reduction and were thus selected for further proteomic analysis. B) Mass spectrum of a sample recovered from an HPLC peak susceptible to reduction (Peak 16-3 of JFH-1 E2 identified as peptide J4 in the detailed description provided in SI). Upon reduction, a shift in molecular mass of 2Da was observed, due to addition of two hydrogen atoms upon reduction of the two cysteines. The data presented in this figure allowed the identification of disulfide 4 (Table 2). doi:10.1371/journal.ppat.1000762.g005

downstream of the HVR1 (Fig. 3, first three red boxes) to the three conserved strands in the N-terminal part of DI, i.e. B<sub>0</sub>, C<sub>0</sub> and D<sub>0</sub>. These β-strands are followed by a long intervening region that is compatible with the D<sub>0</sub>E<sub>0</sub> insertion of the class II fold. For the assignment of strands E<sub>0</sub> and F<sub>0</sub>, the available data on the residues involved in CD81 binding (small blue circles in Fig. 3) provide valuable information, since strand E<sub>0</sub> must interact with D<sub>0</sub> (see diagram in Fig. 6B). Thus, assigning E<sub>0</sub> and F<sub>0</sub> to the two consecutive strands predicted after residue 525 brings together a patch of residues that are apart in primary structure to the same face of DI, forming the site of interaction with CD81 (Fig. 6). For the assignment of the remaining β-strands, there is crucial information provided by disulfide 1. This disulfide bond connects Cys429, at the end of strand C<sub>0</sub>, with Cys552 further downstream, which therefore must be at the same end of the DI β-sandwich. This means that Cys552 must be located either at the G<sub>0</sub>H<sub>0</sub> loop,

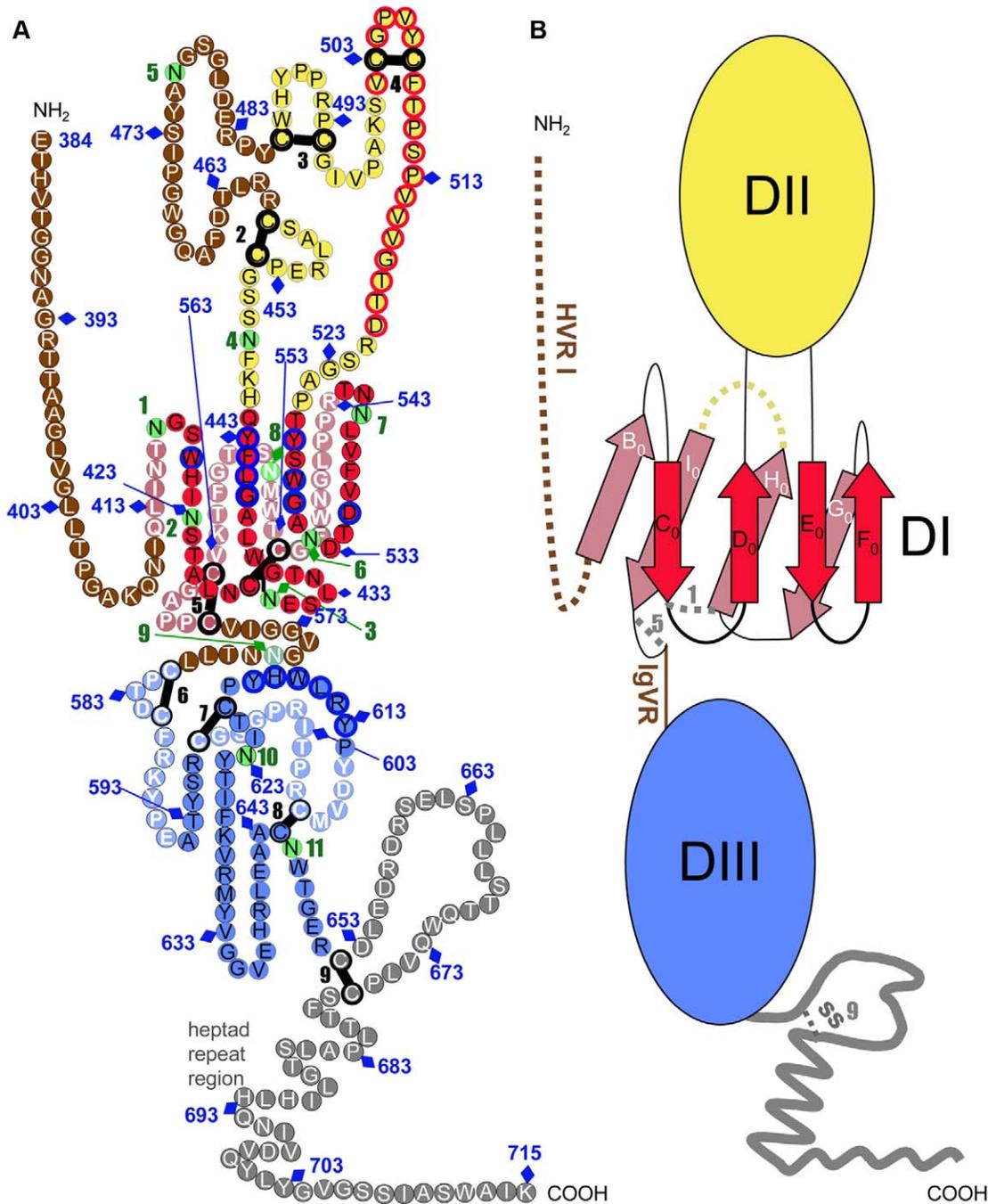
or at the end of the I<sub>0</sub> strand, if the molecule is to have a class II fold (Fig. 6B). However, after strand F<sub>0</sub>, there is a long strand predicted to span residues 549–555 (Fig. 3), which would have Cys552 in the middle. But the comparison of predicted versus experimentally determined β-strands of alphavirus E1 shows that, for several alphaviruses, the region of G<sub>0</sub> and H<sub>0</sub> is also predicted as a single long strand (Fig. 3B). Indeed, in both alphaviruses and HCV, there is a glycine residue (Gly 551 in E2e) forming a tight turn that reverses the chain orientation, going from G<sub>0</sub> into H<sub>0</sub> (some alphaviruses have two glycines at this β-turn). This shows that the prediction algorithms are not 100% reliable, suggesting that in HCV E2, Gly551 is at the G<sub>0</sub>H<sub>0</sub> turn, and that Cys552 is the first residue of strand H<sub>0</sub>. Indeed, running at the edge of the DI β-sandwich, the sequence of G<sub>0</sub> (as well as the sequence of the alphavirus B<sub>0</sub> strand, at the other end of the bottom β-sheet, Fig. 3B) appears to be less typical than the sequences of internal β-strands in a β-sheet, which are easier to predict by computer algorithms. In addition, the short connections between strands F<sub>0</sub> through H<sub>0</sub> in both alpha- and flavivirus DI are also consistent with the assignment of H<sub>0</sub> to a strand running between residues 552 and 555 in HCV E2.

Having assigned the G<sub>0</sub> and H<sub>0</sub> strands, additional considerations are necessary to assign strand I<sub>0</sub>. In alpha- and flaviviruses, I<sub>0</sub> is one of the two central β-strands of the bottom sheet of DI, and is directly followed by the linker connecting to DIII. Because it is the only strand missing to complete the 8-stranded β-sandwich, it can only make disulfide bonds to cysteines located upstream in primary sequence. In HCV, three β-strands are predicted directly downstream to the assigned H<sub>0</sub> strand: one around residue 563, one around 573, and one around 593 (Fig. 3). The strand around residue 573 is part of a segment that can be deleted without affecting protein conformation [14], indicating that it cannot be I<sub>0</sub>. The strand around 593 ends at Cys597, which forms disulfide 7 with Cys620 further downstream. Because class II proteins can have no

**Table 2. Recapitulation of the 9 disulfide bridges and the isolates in which they were identified.**

	H77	JFH-1	UKN2b_2.8
1 Cys429 - Cys552			+
2 Cys452 - Cys459	+	+	
3 Cys486 - Cys494	+		+
4 Cys503 - Cys508		+	+
5 Cys564 - Cys569	+		
6 Cys581 - Cys585	+		+
7 (Cys597 - Cys620)			
8 Cys607 - Cys644	+	+	+
9 Cys652 - Cys677	+		

doi:10.1371/journal.ppat.1000762.t002



**Figure 6. Tertiary organization of HCV E2e.** A) The linear sequence of HCV H77-E2e, numbered every 10 residue according to the polyprotein (N-terminus at position 384 and the TM region beginning at 716) is represented as a chain of beads (colored circles) labeled with the corresponding amino acid and threaded onto a class II fold as described in the text. Circles in pale and bright colors represent residues in the background and foreground of the domains, respectively labeled in white and black fonts. Disulfide bonds and glycosylation sites are indicated by thick black bars and green circles, respectively, numbered sequentially. Unstructured segments are in white font on a brown background. Residues that participate in CD81 binding are contoured in blue, and those from the putative fusion loop region in red. In DIII, a plausible arrangement of the 3 predicted  $\beta$ -strands is illustrated by placing the corresponding polypeptide segments in an antiparallel putative  $\beta$ -sheet, but unlike DI, the topological arrangement of this domain was not determined. B) Schematic diagram of the tertiary organization of HCV E2, with DI, DII and DIII in red, yellow and blue, respectively. The stem region is grey. The connectivity of the  $\beta$ -strands in DI is indicated, labeled with the standard class II nomenclature. A broken yellow line indicates the place of the second insertion in other class II proteins.  
doi:10.1371/journal.ppat.1000762.g006

interdomain disulfides - which would be incompatible with their function - this strand cannot be assigned to  $I_0$  either. Indeed, the interleaved nature of disulfide bonds 7 and 8 dictates that none of the strands predicted downstream can be in DI. The only

option compatible with a class II fold is, therefore, to assign the strand around residue 563 to  $I_0$ . This assignment implies that the long insertion in the  $H_0I_0$  loop of the alpha- and flavivirus fusion proteins is absent in HCV E2. This is in line with E2 from

HCV and pestiviruses being shorter than the alpha- and flavivirus fusion proteins by about 80–110 amino acids, i.e., roughly the length of the insertion in the H<sub>0</sub>I<sub>0</sub> loop of the latter.

The assignment of the 8 strands in HCV DI also indicates that the linker connecting DI and DIII must be between disulfides 5 and 6, encompassing the region called IgVR (“intergenotypic variable region”), which can be deleted without affecting protein conformation, at least in the prefusion form of E2. As discussed above, the segment containing disulfide 9 is likely not to be part of the structured ectodomain, further implying that DIII is comprised between disulfides 6 and 9, spanning about 70 amino acids. The presence of two long-range disulfide bonds (disulfides 7 and 8) suggests that this region is indeed structured into a separate domain. However, the secondary structure predictions point to only 3  $\beta$ -strands in this domain, indicating that the Ig-like fold of DIII in alpha- and flaviviruses may not have been maintained in HCV. Moreover, we found no obvious way to propose an Ig-like arrangement of the polypeptide chain in this domain such that it would also satisfy the constraints imposed by disulfides 7 and 8.

### Main features of the tertiary structure of HCV E2

The resulting model for the tertiary structure of E2 is presented in Fig. 6A, with the diagram of Fig. 6B highlighting, as a guide, the essential features of the resulting “class II” organization of the protein. The main features of the molecule are the following:

#### DI

This domain has an N-terminal extension in flaviviruses (which includes  $\beta$ -strand A<sub>0</sub>) with respect to alphaviruses (see review by [16]), and in HCV, the HVR1 also appears to be an N-terminal extension. DI contains disulfides 1 and 5, both at the DII distal end of the DI  $\beta$ -sandwich; i.e., at its DIII interacting end. Disulfide 5 connects two consecutive cysteines into a short loop at the end of strand I<sub>0</sub>.

The C<sub>0</sub>D<sub>0</sub>E<sub>0</sub>F<sub>0</sub>  $\beta$ -sheet (or “top” sheet) contains most of the determinants of CD81 binding (blue circles in Fig. 6A), and 5 of the 11 N-linked glycosylation sites of E2 (numbered 1, 2, 3, 6 and 7, Figs. 3 and 6A). In contrast, the B<sub>0</sub>I<sub>0</sub>H<sub>0</sub>G<sub>0</sub>  $\beta$ -sheet (or “bottom” sheet) has only site 8 (Asn 556, Fig. 6A), located in the H<sub>0</sub>I<sub>0</sub> loop, at the site of the long insertion in alpha- and flavivirus fusion proteins (yellow dotted line, Fig. 6B). The presence of an insertion in the other class II proteins suggests that there is space at this end of the barrel for a glycan chain attached to Asn556. Importantly, glycan 8 was shown to be essential for the correct folding of E2, in line with the key location in the H<sub>0</sub>I<sub>0</sub> loop in the bottom sheet. Overall, the distribution of glycans on HCV DI is compatible with the experimentally determined orientation of flavivirus E and alphavirus E1 at the virion surface, with the bottom sheet facing the viral membrane. This pattern provides additional evidence validating our assignment of the DI  $\beta$ -strands.

#### DII

This domain has two predicted glycosylation sites and three disulfide bonds (2, 3 and 4), all connecting consecutive cysteine residues very close in primary structure. In the alpha- and flavivirus counterparts, the two insertions forming DII are quite intertwined, and the second one (the H<sub>0</sub>I<sub>0</sub> insertion) acts as a scaffold supporting the D<sub>0</sub>E<sub>0</sub> insertion bearing the fusion loop at the DI-distal end. In HCV E2, the absence of the second insertion makes DII much smaller, and apparently also results in a more flexible, or disordered domain, as suggested by the absence of long-range disulfide bonds and the fact that the whole area

between disulfides 2 and 3 can be eliminated without affecting the overall conformation of the molecule.

### A putative fusion loop

The proposed tertiary organization of E2 provides a prediction for the location of the HCV E2 fusion loop, which in class II fusion proteins is a stretch of highly conserved residues within the D<sub>0</sub>E<sub>0</sub> insertion. This segment is composed mainly of non-charged residues, and is rich in glycine and non-polar amino acids. The sequence alignment highlights the region spanning residues 502–520 (red circles in Fig. 6A), which has similar characteristics and is strictly conserved within all HCV genotypes (Fig. 3). This region is thus a strong candidate for fulfilling the role of the HCV fusion loop. The smaller conserved block between residues 484–489 has been tested by site directed mutagenesis using retroviral particles pseudotyped with the HCV envelope proteins (HCVpp), suggesting that it does not play a role in membrane fusion [29]. Because in all class II fusion proteins, the fusion loop is buried at an oligomeric interface in the prefusion form, the candidate fusion loop segment is also very likely to mark a contact region with E1.

### Linker region and DIII

Our model predicts that DIII (blue), which in alpha- and flavivirus fusion proteins has an Ig-like fold, contains disulfides 6, 7 and 8, and the last two glycosylation sites, 10 and 11. It is connected to the C-terminal end of DI via the IgVR, which contains glycan 9. This linker region is such that it can be extended to allow the translocation of DIII to the side of the trimer during the fusogenic conformational change, as expected for a class II fusion protein. Yet this segment, located between disulfides 5 and 6 (highlighted in brown in Fig. 6A), can also be replaced by a GSSG linker without affecting the overall protein conformation [14], suggesting close apposition between DI and DIII at least in the prefusion form. This organization is also compatible with the observation that some of the residues important for CD81 binding map to DIII (613–618, Figs. 3 and 6), suggesting that CD81 bridges the surface of the two domains. Indeed, these two domains display an extended interaction surface in other class II fusion proteins.

### The stem region

DIII is followed in sequence by a relatively flexible but conserved region, denoted the “stem” (grey in Fig. 6), which connects to the TM segment. The stem would contain a loop that is closed by the last disulfide (number 9). A number of reports on the characterization of a protein ending at position 661 indicate that the absence of this loop does not affect the conformation of the protein and it is therefore not part of DIII. Further support for this interpretation is provided by our identification of a trypsin-resistant fragment of E2e ending at position Arg648, in between disulfides 8 and 9 (data not shown).

### Implications of the tertiary structure model of HCV E2

One important implication is that the residues interacting with CD81 are found in two domains, DI and DIII, which have to move apart during the fusogenic conformational change. This suggests that CD81 may have to dissociate away for such a conformational change to take place, or on the contrary, that its binding may help to lower the energy barrier for the conformational change to occur upon exposure to low pH in the endosomes.

An additional information from this study is the positioning of the hypervariable regions of HCV E2 in the context of the class II

fold. As suggested above, the presence of these regions is likely to be responsible for the difference in CD spectrum between HCV E2e and the other class II proteins examined (Fig. 4A), which do not contain unstructured regions. Such regions are presumably important for evading the humoral immune response of the host, given that HCV can cause chronic infection, in contrast to alpha- and flaviviruses. Another important difference to the latter are the numerous glycosylation sites in HCV E2. Our model indicates that these sites cluster in particular on the exposed face of DI. Indeed, several glycans appear to frame the CD81 binding surface, partially shielding it from recognition by circulating antibodies. These are glycans 1, 2, 6, 7, and possibly 10 and 11 (Fig. 6A). Importantly, a number of studies have pointed to a role of some of these E2 glycosylation sites modulating entry and/or CD81 binding [43,44].

Another important implication is the identification of a strong candidate region for the fusion loop. This polypeptide segment, spanning residues 502–520 of E2, has all the characteristics reported for the experimentally characterized class II fusion loops. It is strictly conserved, and is located in a region of the protein that is compatible with this function – in the D<sub>0</sub>E<sub>0</sub> insertion of DI – in spite of being part of a fusion domain (DII) that is much smaller than its alpha- and flavivirus counterpart. The fact that in the latter DII is formed by two insertions into a simple, conserved DI, suggests that they may have evolved sequentially. HCV may have thus maintained some ancient intermediate form of the class II proteins, containing only the insertion that carries the fusion loop.

The observed flexible and largely unstructured conformation of DII is likely to be due to the absence of the H<sub>0</sub>I<sub>0</sub> scaffold. The presence of the candidate HCV fusion loop relatively close to DI is another important difference. In the alpha- and flavivirus counterparts, there are about 30 intervening residues present in an extended conformation between the fusion loop and strand E<sub>0</sub>, whereas there are only a few residues in HCV. This difference is likely to be related to the intrinsic flexibility of HCV DII, since the fusion loop is unlikely to lie at the distant tip of an unstructured domain. During membrane fusion, the fusion loop could be further stabilized by interaction with the membrane proximal region of the stem, once the molecule adopts its fusogenic hairpin conformation. This organization also suggests a significantly shorter post-fusion HCV E2 trimer, compared to the other class II proteins. This would be analogous to the observed differences in the post-fusion trimers of class I viral fusion proteins, for instance from retroviruses, which are short [45], and paramyxoviruses [46] or coronaviruses [47], which display a very long hairpin conformation.

Overall, our model for the HCV E2 tertiary organization provides a structural framework to understand the antigenicity of the virion, the organization of the regions that interact with CD81, and the putative conformational changes that are likely to take place during the membrane fusion reaction to invade a target cell. In the absence of 3D structural data, our results constitute an important step to better understand the function of the HCV envelope proteins. This knowledge, in turn, can help devise possible antiviral strategies against this important pathogen. Our data also provide a handle to dissect and obtain structural data on the E2 domains separately, given that the intact ectodomain is very difficult to crystallize.

Finally, this analysis highlights the power of conducting parallel structural studies on related viruses, which provide information that can be extrapolated to other members of the respective viral families, even in the absence of sequence similarity in the corresponding proteins.

## Materials and Methods

The accompanying Supplementary Information (Text S1) describes in detail the construction of the vectors used for expression of synthetic genes coding for the E1-E2ΔTM segment of the 9 HCV isolates tested (Table 1), the protocols for production, purification and conformational characterization of recombinant HCV E2e, as well as the computer and experimental analyses used for secondary structure predictions. Finally, a detailed description of the procedures used for the experimental identification of the cysteine residues involved in 8 disulfide bonds of E2 is provided.

## Supporting Information

**Text S1** Supplementary materials and methods

Found at: doi:10.1371/journal.ppat.1000762.s001 (0.15 MB PDF)

**Table S1** Identification of disulfide bond connected HCV E2e peptides resulting from tryptic digest

Found at: doi:10.1371/journal.ppat.1000762.s002 (0.05 MB PDF)

**Figure S1** Conformational characterization of HCV E2e. A) Pull-down experiment showing that E2e specifically reacts with conformation dependent antibodies CBH-4B and CBH-4D, but not a control antibody, and binds the CD81 LEL. E2e was affinity loaded onto a Streptactin column, samples containing the respective proteins were passed through the column and the complex was eluted after washing. Elution fractions were analysed by SDS-PAGE under reducing conditions. Bands representing the E2e, CD81-LEL as well as the heavy chain (HC) and light chain (LC) of the two antibodies were observed. B) Stoichiometric complex formation between UKN2b\_2.8 E2e and mAb CBH4D. UKN2b\_2.8 E2e, mAb CBH4D and a mixture of the two (ratio 2:1) were loaded to the column (in three different runs) (E2e~50kD, H53~150kD, complex~250kD). No peaks corresponding to either of the isolated proteins were observed in the profile of the complex, indicating a 2:1 complex stoichiometry and a high affinity of UKN2b\_2.8 E2e for mAb CBH4D.

Found at: doi:10.1371/journal.ppat.1000762.s003 (1.23 MB TIF)

**Figure S2** Alignment of HCV E2 amino acid sequences from strains H77, JFH-1 and UKN2b\_2.8. Given the deamination of Asn residues by PNGase F, they are turned into Asp residues. Predicted trypsin cleavage sites (grey triangles) and N-glycosylation sites (empty diamonds) are indicated, cysteines are boxed and the respective disulfide bridges displayed (-SS-). Peptides identified after tryptic cleavage are shaded, named according to the respective isolate and numbered sequentially following the amino acid sequence of E2.

Found at: doi:10.1371/journal.ppat.1000762.s004 (2.52 MB TIF)

**Figure S3** Proteomics results for the determination of the disulfide bridges. HPLC chromatogram peaks that were selected for further proteomics analysis. Results of N-terminal sequencing and SELDI-TOF MS for the respective peaks are shown in panels A-G. Scales for the intensity (y-axis) as well as for the molecular weight (x-axis) vary considerably in the different spectra, resulting in a different appearance of the background noise. A) Peaks JFH-1 6-3 and 12-3, leading to the identification of disulfide bridges 2 (peak 6-3) and 8 (peak 12-3) in E2e of JFH-1. B) Peak JFH-1 16-3, leading to the identification of disulfide bridge 4 in E2e of JFH-1. C) Peaks UKN2b\_2.8 13-1 and 20-1, leading to the identification of disulfide bridges 2 (peak 20-1) and 8 (peak 13-1) in E2e of UKN2b\_2.8. D) Peaks UKN2b\_2.8 42-3 and 19-1, leading to the identification of disulfide bridges 1 (peak 42-3) and 6 (peak 19-1) in E2e of UKN2b\_2.8. E) Peaks H77 15-2 and 6-2, leading to the

identification of disulfide bridges 2 (peak 6-2) and 8 (peak 15-2) in E2e of H77. F) Peaks H77 26-2, which is the result of disulfide shuffling, and 32-2, leading to the identification of disulfide bridge 3 in E2e of H77. G) Peaks H77 43-2 and 33-2, leading to the identification of disulfide bridges 5 (peak 33-2) and 9 (peak 43-2) in E2e of H77.

Found at: doi:10.1371/journal.ppat.1000762.s005 (6.93 MB TIF)

## Acknowledgments

We thank J. Ball, A. Tarr, J. Dubuisson, M. Backovic, M. Kielian and F.X. Heinz for helpful discussions and critical reading of the manuscript; A.

## References

- Shepard CW, Finelli L, Alter MJ (2005) Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* 5: 558–567.
- De Francesco R, Migliaccio G (2005) Challenges and successes in developing new therapies for hepatitis C. *Nature* 436: 953–960.
- Lindenbach BD, Thiel HJ, Rice CM (2007) *Flaviviridae*: The viruses and their replication. *Fields Virology*, Fifth edition. pp 1101–1152.
- Lavie M, Goffard A, Dubuisson J (2007) Assembly of a functional HCV glycoprotein heterodimer. *Curr Issues Mol Biol* 9: 71–86.
- Merola M, Brazzoli M, Cocchiarella F, Heile JM, Helenius A, et al. (2001) Folding of hepatitis C virus E1 glycoprotein in a cell-free system. *J Virol* 75: 11205–11217.
- Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, et al. (1998) Binding of hepatitis C virus to CD81. *Science* 282: 938–941.
- Evans MJ, von Hahn T, Tschernie DM, Syder AJ, Panis M, et al. (2007) Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* 446: 801–805.
- Meertens L, Bertaux C, Cukierman L, Cormier E, Lavillette D, et al. (2008) The tight junction proteins claudin-1, -6, and -9 are entry cofactors for hepatitis C virus. *J Virol* 82: 3555–3560.
- Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, et al. (2009) Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature*.
- Scarselli E, Ansuini H, Cerino R, Roccasecca RM, Acali S, et al. (2002) The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *Embo J* 21: 5017–5025.
- von Hahn T, Rice CM (2008) Hepatitis C virus entry. *J Biol Chem* 283: 3689–3693.
- Lemon SM, Walker CM, Alter MJ, Yi MK (2007) *Hepatitis C Virus*. *Fields Virology*, Fifth edition. pp 1253–1304.
- Kato N, Ootsuyama Y, Ohkoshi S, Nakazawa T, Sekiya H, et al. (1992) Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. *Biochem Biophys Res Commun* 189: 119–127.
- McCaffrey K, Boo I, Pombourios P, Drummer HE (2007) Expression and characterization of a minimal hepatitis C virus glycoprotein E2 core domain that retains CD81 binding. *J Virol* 81: 9584–9590.
- Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, 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.
- Kielian M, Rey FA (2006) Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4: 67–76.
- Mukhopadhyay S, Kuhn RJ, Rossmann MG (2005) A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol* 3: 13–22.
- Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, et al. (2001) The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105: 137–148.
- Lorenz IC, Allison SL, Heinz FX, 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.
- Andersson H, Barth BU, Ekstrom M, Garoff H (1997) Oligomerization-dependent folding of the membrane fusion protein of Semliki Forest virus. *J Virol* 71: 9654–9663.
- Dubuisson J, Rice CM (1996) Hepatitis C virus glycoprotein folding: disulfide bond formation and association with calnexin. *J Virol* 70: 778–786.
- Kielian M (2006) Class II virus membrane fusion proteins. *Virology* 344: 38–47.
- Stiasny K, Heinz FX (2006) Flavivirus membrane fusion. *J Gen Virol* 87: 2755–2766.
- Backovic M, Jardetzky TS (2009) Class III viral membrane fusion proteins. *Curr Opin Struct Biol* 19: 189–196.
- Weissenhorn W, Hinz A, Gaudin Y (2007) Virus membrane fusion. *FEBS Lett* 581: 2150–2155.
- Yagnik AT, Lahm A, Meola A, Roccasecca RM, Ercole BB, et al. (2000) A model for the hepatitis C virus envelope glycoprotein E2. *Proteins* 40: 355–366.
- Fenouillet E, Lavillette D, Loureiro S, Krashias G, Maurin G, et al. (2008) Contribution of redox status to hepatitis C virus E2 envelope protein function and antigenicity. *J Biol Chem* 283: 26340–26348.

Tortorici for participation in initial experiments; J. Dubuisson and S. Foug for providing antibodies (H53, CBH-4D and CBH-4B); T. Wakita for the gift of pJFH1; C. Rice for the gift of Huh7.5 cells; C. Traboni for the gift of cB76.1/Huh7 cells; E. Navarro-Sanchez and K. Moncoq for preparing WNV sE, CHIKV sE1, and DV3 sE proteins, respectively.

## Author Contributions

Conceived and designed the experiments: TK JdA AS BB BR PE MAAP AM FAR. Performed the experiments: TK JdA CMK AS IP DXJ BB BR. Analyzed the data: TK JdA AS BB BR AM FAR. Contributed reagents/materials/analysis tools: TK CMK LDP RGT MAAP. Wrote the paper: TK FAR.

- Flint M, Thomas JM, Maidens CM, Shotton C, Levy S, et al. (1999) Functional analysis of cell surface-expressed hepatitis C virus E2 glycoprotein. *J Virol* 73: 6782–6790.
- Lavillette D, Pecheur EI, Donot P, Fresquet J, Molle J, et al. (2007) Characterization of fusion determinants points to the involvement of three discrete regions of both E1 and E2 glycoproteins in the membrane fusion process of hepatitis C virus. *J Virol* 81: 8752–8765.
- Garry RF, 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.
- Harrison SC (2008) Viral membrane fusion. *Nat Struct Mol Biol* 15: 690–698.
- Ivy J, Nakano E, Clements D Methods of preparing carboxy-terminally truncated recombinant flavivirus envelope glycoproteins employing drosophila melanogaster expression systems.
- Owsianka AM, Timms JM, Tarr AW, Brown RJ, Hickling TP, et al. (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.
- Drummer HE, Boo I, Maerz AL, Pombourios 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.
- Lavillette D, Tarr AW, Voisset C, Donot P, Bartosch B, et al. (2005) Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus. *Hepatology* 41: 265–274.
- Hadlock KG, Lanford RE, Perkins S, Rowe J, Yang Q, et al. (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.
- Kanai R, Kar K, Anthony K, Gould LH, Ledizet M, et al. (2006) Crystal structure of west nile virus envelope glycoprotein reveals viral surface epitopes. *J Virol* 80: 11000–11008.
- Whidby J, Mateu G, Scarborough H, Demeler B, Grakoui A, et al. (2009) Blocking hepatitis C virus infection with a recombinant form of envelope protein 2 ectodomain. *J Virol*.
- Modis Y, Ogata S, Clements D, Harrison SC (2005) Variable surface epitopes in the crystal structure of dengue virus type 3 envelope glycoprotein. *J Virol* 79: 1223–1231.
- Goormaghtigh E, Cabiaux V, Ruyschaert JM (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. I. Assignments and model compounds. *Subcell Biochem* 23: 329–362.
- Drummer HE, Pombourios 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.
- Forns X, Thimme R, Govindarajan S, Emerson SU, Purcell RH, et al. (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.
- Falkowska E, Kajumo F, Garcia E, Reinius J, Dragic T (2007) Hepatitis C virus envelope glycoprotein E2 glycans modulate entry, CD81 binding, and neutralization. *J Virol* 81: 8072–8079.
- Goffard A, Callens N, Bartosch B, Wychowski C, Cosset FL, et al. (2005) Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins. *J Virol* 79: 8400–8409.
- Skehel JJ, Wiley DC (1998) Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* 95: 871–874.
- Lamb RA, Jardetzky TS (2007) Structural basis of viral invasion: lessons from paramyxovirus F. *Curr Opin Struct Biol* 17: 427–436.
- Bartlam M, Yang H, Rao Z (2005) Structural insights into SARS coronavirus proteins. *Curr Opin Struct Biol* 15: 664–672.
- King RD, Sternberg MJ (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci* 5: 2298–2310.
- Roussel A, Lescar J, Vaney MC, Wengler G, Wengler G, et al. (2006) Structure and interactions at the viral surface of the envelope protein E1 of Semliki Forest virus. *Structure* 14: 75–86.