Published in final edited form as:

J Biomol Struct Dyn. 2015 ; 33(8): 1682–1694. doi:10.1080/07391102.2014.967300.

Structure–function analysis of hepatitis C virus envelope glycoproteins E1 and E2

Aparajita Nayak^{a,1}, Nagarajan Pattabiraman^{b,1}, Numrah Fadra^a, Radoslav Goldman^c, Sergei L. Kosakovsky Pond^d, and Raja Mazumder^{a,e,*}

^aDepartment of Biochemistry and Molecular Medicine, George Washington University, Washington, DC 20037, USA

^bMolbox LLC, 8115 Fenton Street, Suite 304, Silver Spring, MD 20910, USA

^cDepartment of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 20007, USA

^dDepartment of Medicine, University of California at San Diego, La Jolla, CA 92093, USA

^eMcCormick Genomic and Proteomic Center, George Washington University, Washington, DC 20037, USA

Abstract

Hepatitis C virus (HCV) is the leading cause of chronic liver disease in humans. The envelope proteins of HCV are potential candidates for vaccine development. The absence of three-dimensional (3D) structures for the functional domain of HCV envelope proteins [E1.E2] monomer complex has hindered overall understanding of the virus infection, and also structure-based drug design initiatives. In this study, we report a 3D model containing both E1 and E2 proteins of HCV using the recently published structure of the core domain of HCV E2 and the functional part of E1, and investigate immunogenic implications of the model. HCV [E1.E2] molecule is modeled by using aa205–319 of E1 to aa421–716 of E2. Published experimental data were used to further refine the [E1.E2] model. Based on the model, we predict 77 exposed residues and several antigenic sites within the [E1.E2] that could serve as vaccine epitopes. This study identifies eight peptides which have antigenic propensity and have two or more sequentially exposed amino acids and 12 singular sites are under negative selection pressure that can serve as vaccine or therapeutic targets. Our special interest is ₂₈₅FLVGQLFTFSPRRHW₂₉₉ which has five negatively selected sites (L286, V287, G288, T292, and G303) with three of them sequential and four amino acids exposed (F285, L286, T292, and R296). This peptide in the E1 protein maps to dengue envelope vaccine target identified previously by our group. Our model provides for the first time an overall view of both the HCV envelope proteins thereby allowing researchers explore structure-based drug design approaches.

*Corresponding author. mazumder@gwu.edu.

¹Equal contributors. Should be considered as first authors.

Supplemental data

Supplemental data for this article can be accessed here <http://dx.doi.10.1080/07391102.2014.967300>.

Keywords

HCV; envelope proteins; glycoproteins; structural analysis; 3D model

Introduction

Hepacivirus is an enveloped positive-sense single stranded RNA virus (+ssRNA virus) that belongs to the flaviviridae family. Hepatitis C virus (HCV), the type species of the genus *Hepacivirus*, is a highly pathogenic virus causing liver cirrhosis and hepatocellular carcinoma (Ashfaq, Javed, Rehman, Nawaz, & Riazuddin, 2011; Qureshi, 2007). The HCV genome polyprotein is cleaved into 11 viral proteins, out of which five are structural (core protein p21, core protein p19, envelope glycoprotein E1, envelope glycoprotein E2 and core protein p7) and six are non-structural (protease NS2–3, serine protease NS3, non-structural protein 4A, non-structural protein 4B, non-structural protein 5A, and RNA-directed RNA polymerase) (Lidenbach, Thiel, & Rice, 2007; Krekulova, Rehak, & Riley, 2006). The envelope glycoprotein (E protein in *flavivirus* and E1 and E2 protein in *Hepacivirus*) is the principal protein of the virion in all of flaviviridae family, which mediates the virus attachment and membrane fusion, thereby playing an important role during infection of host cells, and is considered to be a potential vaccine and therapeutic target in the flaviviridae family (Geiss, Stahla, Hannah, Gari, & Keenan, 2009; Helle & Dubuisson, 2008; Houghton & Abrignani, 2005; Kachko et al., 2011; Law et al., 2013; Stamataki, Coates, Abrignani, Houghton, & McKeating, 2011; Voisset & Dubuisson, 2004). Several experimental studies suggest that E1 and E2 are transmembrane glycoproteins bearing an N-terminal ectodomain and a C-terminal transmembrane domain. They form a multi-subunit complex at the surface of the virus particle which facilitates entry of the virus particle into the host cell, a key step involved in HCV infection (Cocquerel, Voisset, & Dubuisson, 2006; Op De Beeck, Cocquerel, & Dubuisson, 2001).

The envelope protein is the major component of the external surface of the virion and represents the dominant virus antigen, evoking protective immune responses. HCV has two envelope glycoproteins namely E1 and E2 (unlike dengue virus and other *Flaviviruses* which have only one). The E1 glycoprotein ranges from position 192 in the polyprotein to 383 (192 amino acid long), while E2 glycoprotein is from 384 to 746 (363 amino acid long) [amino acid numbers are based on UniProtKB entry POLG_HCV1 (UniProt Consortium, 2012)]. In *Flaviviruses*, the newly synthesized prM and E proteins associate to form non-covalent heterodimers (Wengler, 1989). Subsequently prM-E are dissociated due to the cleavage of prM by the golgi-resident protease, furin (Stadler, Allison, Schlich, & Heinz, 1997) which leads to the formation of E protein homodimer (Stiasny, Allison, Marchler-Bauer, Kunz, & Heinz, 1996). While it is well established that the E1 and E2 proteins are one of the major proteins responsible for the pathogenicity and immunogenic properties of HCV with the E2 core domain having some similarities with *flavivirus* envelope protein, such as the core secondary structure consisting of predominantly beta-sheets and random coil (Khan et al., 2014; Kong et al., 2013; Yu et al., 2007); identification of exact residues/regions that are critical for virus survival and propagation is still being actively researched (Lavie et al., 2014; Sabahi, Uprichard, Wimley, Dash, & Garry, 2014). Results from our study is expected

to assist in the determination of the E1 and E2 complex structure using similar approaches used to solve the E1 and E2 complex of *Alphaviruses* (Li, Jose, Xiang, Kuhn, & Rossmann, 2010; Voss et al., 2010).

Although the recent X-ray structure of core E2 shows a unique compact globular structure for core E2, the precise roles played by E1 and E2 in membrane fusion are still not fully understood (Khan et al., 2014). Several recent articles discuss the unexpected structural features of the recently solved E2 protein, but also remind us that the protein expressed had to be modified for crystallization (Sabahi et al., 2014; Vieyres, Dubuisson, & Pietschmann, 2014) and hence, alternate conformations may exist. Size-exclusion chromatography and SAXS analyses suggest that it is unlikely that E2 has a direct role in membrane fusion, although it is possible that the E1 and E2 monomer complex (represented hereon as [E1.E2]) could play a major role in the fusion process (Khan et al., 2014) or E1 is the actual fusion protein (Vieyres et al., 2014). The formation of a tetramer of E1 and E2 (or the dimer of [E1.E2]) which is morphologically and possibly functionally similar to homodimer structure of *flavivirus* E protein has been suggested based on cryo-electron microscopy (Yu et al., 2007).

So far, there has been no structural bioinformatics analysis of the entire envelope protein (E1 plus E2) and hence, represents a gap in our understanding of the molecule. Recent work by Rychlowska et al. (2011) and others (Callens et al., 2005; Dubuisson & Rice, 1996; Drummer, Boo, & Pountourios, 2007; Flint, Maidens. et al., 1999; Flint, Thomas, et al., 1999; Fournillier et al., 2001; Garry & Dash, 2003; Grove et al., 2008; Helle et al., 2007; Jacob, Keck, Olson, Fong, & Tomer, 2008; Lavie, Goffard, & Dubuisson, 2006; Lavillette et al., 2007; Maurin et al., 2011; Yagnik et al., 2000; Yi, Nakamoto, Kaneko, Yamashita, & Murakami, 1997) has helped in identifying some of the functionally important residues, and therefore provides ample information for a comprehensive structural bioinformatics analysis that can be built upon the existing knowledge. It has been described that the HCV glycan shield reduces the immunogenicity of the envelope proteins, and masks conserved neutralizing epitopes at their surface (Helle, Duverlie, & Dubuisson, 2011). Extensive research in the development of a protective vaccine against HCV suggests that the vaccine needs to stimulate production of antibodies that exhibit antiviral activity and potent HCV specific T cell responses (Fournillier et al., 2001).

The guiding principles and motivation to generate a functional monomer model of [E1.E2] are as follows. Recently, a three-dimensional (3D) structure of HCV envelope glycoprotein E2 core bound to an antibody using single-crystal X-ray diffraction study has been reported (Khan et al., 2014; Kong et al., 2013). 3D models were reported for HCV E2 domain using homology with dengue envelop protein (Yagnik et al., 2000) and *de novo* modeling (Afzal, Idrees, & Hussain, 2014). Majority of flaviviridae envelope proteins belong to class II fusion protein, and fold into similar shapes containing three domains (Figure 1(A)); in addition, on comparing the 3D folding of the envelope proteins of dengue, (Modis, Ogata, Clements, & Harrison, 2003; Nayak et al., 2009) West Nile virus (Kanai et al., 2006), tick-borne encephalitis virus (Rey, Heinz, Mandl, Kunz, & Harrison, 1995), E1 of Semliki Forest virus (Roussel, Lescar, Vaney, Wengler, & Rey, 2006), glycoprotein of Chikungunya virus (Voss et al., 2010), and envelope protein E1 of Rubella Virus (DuBois et al., 2013), the folding of N-

terminal residues (around 120 amino acids; blue colored ribbon in Figure 1(B)) for all of them is such that the N-terminal interacts with domain I, continues to span across domain II and the amino acids closer to C-terminal interacts with domain III. The differences between the above envelope proteins are in the position and orientation of domain II containing the membrane fusing sequences with respect to the other two domains (I and III). Thus, it is reasonable to assume that the 120 amino acids at the N-terminal of E1 could be either a part of a contiguous polypeptide of the envelope protein as observed in dengue or could possibly come from two different regions of a polypeptide to form a functional envelope protein (as in HCV). In addition, the amino acids between these two functional segments of E1 and E2 could be anchored in the membrane, and cleaved to form a functional envelope protein as in case of HCV envelope protein.

In this study, we focus on building a functional knowledge guided 3D model of the HCV envelop [E1.E2] using the majority of the ectodomain sequences of E1 and E2 by using the published crystal structure of E2 core (PDB ID: 4MWF) and dengue virus serotype-2 envelope glycoprotein (PDB ID: 1OAN) as template for E1 and a part of E2. Based on information available from the model, we also predict the sequential epitopes and exposed/buried residues to identify antigenic regions of the envelope proteins which can serve as priority experimental targets for the development of vaccines and therapeutics against HCV. This study identifies potential targets in E1 and E2 proteins which could aid in the development of a prophylactic HCV vaccine by using a combination of bioinformatics methods such as modeling, selection pressure studies, and immunoinformatics methods developed earlier and applied in the discovery and validation of such targets in other viruses (Mazumder et al., 2007; Sagripanti, Mazumder, & Wu, 2011).

Materials and methods

Amino acid residue numbers mentioned in the manuscript are based on UniProtKB accession: P26664 (POLG_HCV1) protein sequence unless otherwise noted.

Model building and evaluation

For the modeling of [E1.E2] monomer the sequences of the HCV genotype 1 E1 and E2 proteins were selected. The modeling of [E1.E2] monomer was carried out in three steps. 1) In the reported crystal structure of HCV E2 core, 41 residue loop from 452 to 492, 14 residue loop from 574 to 577, and 11 residues loop from 586 to 596 are missing. Missing loops were built using the loop-building module in MODELLER (Fiser & Sali, 2003). The final model was energy minimized using GROMOS (Christen et al., 2005) as implemented in DeepView (Schwede, Kopp, Guex, & Peitsch, 2003) using constraint on c-alpha atoms. 2) Using two alignments between E1 of HCV1 (residues 205–319) and the corresponding crystal structure of dengue (residues 1–118 in 1OAN) as template and between E2 of HCV1 (residues 646–716) and the corresponding crystal structure of dengue (residues 324–394 in 1OAN). The homology models of these two fragments were generated by MODELLER, and energy minimized using constraint on c-alpha atoms. 3) The model for the functional [E1.E2] monomer was generated from the E2 and the above two fragments' models were assembled guided by the dengue envelop protein crystal structure 1OAN. During this part of

the modeling, all the sugars found in the E2 core crystal structure were included. The assembled [E1.E2] monomer model was energy minimized without any constraints to generate the final [E1.E2] monomer model. The final energy minimized models were evaluated for stereo-chemical qualities with PROCHECK (Laskowski, Moss, & Thornton, 1993) and with QMEAN (Benkert, Kunzli, & Schwede, 2009). The final minimized model of the [E1.E2] monomer was visualized, compared, and annotated using UCSF CHIMERA (Pettersen et al., 2004).

Prediction of exposed, buried residues, and antigenic epitopes

Surface for each residue was computed from the exposed and occluded (buried) surface areas of each residue in the model using the program OS (Pattabiraman, Ward, & Fleming, 1995). The percentages of exposed and buried surface areas for each residue were calculated from the ratios of exposed and occluded surface areas over the total surface area, respectively, and multiplied by 100. For each residue, if the percentage of the exposed surface ratio is greater than the buried surface ratio then the residue is labeled as exposed, and vice versa for a buried residue. The antigenic determinants were predicted using predicted antigenic peptide server (<http://imed.med.ucm.es/Tools/antigenic.html>) based on the algorithm by Kolasker and Tongaonkar (Kolasker & Tongaonkar, 1990), and the NetCTLpan server (Stranzl, Larsen, Lundegaard, & Nielsen, 2010) was used to perform the CTL epitope predictions. NetCTLpan integrates prediction of peptide MHC class I binding; proteasomal C terminal cleavage and TAP transport efficiency. A peptide length of 9 mer was chosen because most HLA molecules have a strong preference for binding 9mer peptides. The HLA supertype representative with 12 different allele types was chosen, because the class I MHC molecules display high polymorphism. A percent rank of .8 was set for prediction threshold. These epitopes were then mapped onto the functionally important amino acids including those residues which were predicted to be exposed on the surface of both envelope proteins.

Selection pressure studies

We created a curated set of 2565 sequences by performing a TBLASTN (NCBI_Resource_Coordinators, 2014) search using the protein sequences of the six genotypes of different isolates of E1 and E2 that were retrieved from UniProtKB (UniProt_Consortium, 2012). The TBLASTN search was then used as a guide to retrieve the nucleotide sequences from the BLASTN results. Sequences were aligned using a codon-aware algorithm, (Kumar, Nei, Dudley, & Tamura, 2008) and analyzed for evidence of natural selection using the fixed effects likelihood (FEL) (Kosakovsky Pond & Frost, 2005) and the mixed effects model of evolution (MEME) (Murrell et al., 2012) phylogenetic methods. FEL was used to identify individual sites in the multiple sequence alignment where the mean (over the phylogenetic tree) ratio of non-synonymous (dN) and synonymous (dS) substitution rates (dN/dS) was significantly ($p < 10^{-5}$, likelihood ratio test) <one. $dN/dS < 1$ implies that purifying selection is removing non-synonymous mutations faster than expected under neutrality [$dN/dS = 1$, e.g. see (Delpont, Scheffler, & Seoighe, 2009)]. In a large tree, bursts of episodic positive selection, which affect only a small subset of branches, can be masked by long periods of purifying selection (Murrell et al., 2012). To guard against this possibility, we further screened the alignments of evidence of episodic diversifying selection

using MEME. A site would be classified as negatively selected if two conditions were met: (i) FEL reported $dN/dS < 1$ ($p < 10^{-5}$) and (ii) MEME did not find evidence of episodic diversifying selection ($p > .05$, likelihood ratio test).

Results and discussion

Model building of HCV [E1.E2] monomer

Based on UniProtKB annotation and our analysis, the envelope protein residues 359–379 are helical transmembrane region and buried into the membrane; and the envelope protein is cleaved between 383 and 384 by a cellular signal peptidase into two polypeptides, E1 and E2. During the formation of E1 and E2 functional complex, the above helix and its neighboring residues should not interfere with the folding and the formation of E1 and E2 complex. Residues 384–413 of E2 are the HVR1. In HCV infected chimpanzees, the deletion of HVR1 of E2 did not affect viral infectivity (Forns et al., 2000) and this region is not part of the folded domain (Kong et al., 2013; Krey et al., 2010). Based on our previously published methods in structure-guided sequence alignments (Marchler-Bauer et al., 2003; Mazumder et al., 2007; Mazumder & Vasudevan, 2008; Mazumder, Vasudevan, & Nikolskaya, 2008), we built the 3D homology model by manually curating the alignment in Cn3D using the conserved motif strategy (Chakrabarti & Sowdhamini, 2004) not only in the primary sequences, but also in the secondary structure as well as in the 3D folds between envelope proteins. The editing of the alignment was guided by experimentally validated functional information on specific amino acid sites (Supplementary Table S1). Residues 205–319 of E1 and 421–716 of E2 proteins were aligned with the E protein of dengue with the residues between 1–118 and 324–394 (using residue numbers from PDB structure: 1OAN), respectively. During the alignment, the sequences of beta-strand structures in the crystal structure of E protein of dengue were not broken. The final alignment generated based on this method is shown in Figure 2. In panel A of Figure 2, shows the E1 regions marked as fusion peptide and vaccine target (Sagripanti et al., 2011) which aligns with dengue E protein with comparisons in a recent review that argues that E1 might actually be the fusion protein (Vieyres et al., 2014). Initial prediction of the secondary structures of sequences for E1 (aa205–319) and for E2 (aa421–716) glycoproteins using the GOR4 program (Gibrat, Garnier, & Robson, 1987) reveals that around 36% of the total residues are extended/beta-strand and around 4% are alpha-helices, which is similar to the secondary structures of the corresponding sequence of *Flavivirus* E proteins and also matches the solved HCV E2 protein structure which consists mostly of beta-strands and random coil with two small alpha-helices (Khan et al., 2014). Finally, the solved structure of E2 core domain (Kong et al., 2013) was used to generate the final [E1.E2] model.

Figure 3 shows a ribbon representation of the generated model for the functional domain of the envelope protein [E1.E2] monomer. The magenta ribbon represents a part of the HCV1 E1 (amino acid residues from 205–319), the cyan, blue and the green-colored ribbons represent a part of the HCV1 E2 (amino acid residues 421–716). In this figure, the C-terminal break at A319 for E1 and the N-terminal break at H421 for E2 are shown by arrows and labeled. The X-ray crystal structure of the E2 core is represented by cyan ribbons. Segments modeled are shown by magenta, blue, and green-colored ribbons in Figure 3. In

our model of [E1.E2] monomer, about 140 residues are extended/beta-strand (35%) and 24 residues are helical (~6%). Although the number for extended/beta-strand is slightly higher than what is predicted by GOR4 program using sequence of E1 and E2, it overall agrees with the results of the solved structure. In order to evaluate the quality of models, we used two programs, PROCHECK (Laskowski, MacArthur, Moss, & Thornton, 1993) and QMEAN (Benkert et al., 2009). Although, evaluating quality of models or even solving 3D structures using a variety of software has become standard practice, the utility of these evaluations are often contextual (Bhattacharya, Tejero, & Montelione, 2007; Brown & Ramaswamy, 2007). PROCHECK provides the comparison of geometrical/stereochemical parameters of the model to that of the analysis of high-resolution 118 structures of at least 2.0 Å. From the structural analysis of our [E1.E2] monomer model by PROCHECK, we found that 93.8% residues of the model are in the most favored regions and additionally allowed regions, ~4.7% in generously allowed regions, and ~1.5% in the disallowed regions in the Ramachandran plot. Furthermore, the overall average score for our model by PROCHECK is closer to -.386 (average *G*-factor for dihedral angles is -.702 and that of main-chain covalent forces is -.14) with the overall average is -.421 which is above -.5. In contrast to PROCHECK, QMEAN provides the c β interaction energy, all-atom pairwise, solvation, and torsion angle energies as well as the secondary structure and solvent accessibility agreement. For our model, lower z-score was reported for the c β interaction energy (-2.83), solvent accessible agreement (-3.33), and torsion angle energy (-3.38), and the other parameters are lower than -5.0. The overall z-score for the model is -4.97 and the QMEAN score is .344. As more and more data available on the E1 and E2 protein, the quality of the model can be improved. Thus, the model for the functional [E1.E2] monomer reported by this study can be used as a working model with quality sufficient for the prediction of the sequential antigenic epitopes.

Structural features of HCV [E1.E2] monomer model

Several studies show that the E1 and E2 work together as a single molecule in order to mediate HCV entry into host cells (Goffard et al., 2005; Lavie et al., 2006). In this study, we have taken the first attempt to build a functional model of [E1.E2] monomer (Figure 3). The amino acids from 320–425, that contains the cleavage site (aa383–384) and the transmembrane helix (aa359–379) in E1, were not included in our model and were shown by a break between amino acids 319 and 421 in Figure 3. The location of the break was chosen such that the amino acids (aa320–420) not included in our model can fold into a separate domain without interfering with the other domains and also able to insert the transmembrane helix into the membrane. As observed in case of dengue E protein, the [E1.E2] monomer folds into three distinct domains. Maurin et al. (2011) showed that the motif (aa219–221) of E1 is essential for [E1.E2] formation (Maurin et al., 2011). In our model, the A219-L220-L221 region of E1 forms a stable antiparallel beta-strand and A219 is exposed as well as L220 and L221 are partially exposed. This hydrophobic patch could be important to form a [E1.E2].[E1.E2] dimer. The 41 residue loop in E2 (from 452 to 492) that was not observed in the crystal structure is required to guide the folding of E1 as seen in Figure 3. The C-terminal amino acids from 646 to 716 that were not included in the x-ray structure of the core E2 interacts with the N-terminal of E1.

Due to the disulfide bond between Cys452 and Cys620 in E2, the solved X-ray structure forms a compact globular structure which allows E1 to fold around that core to form a class II like fusion protein structure (Wahid et al., 2013). It is interesting to note that in our model the disulfide bonds (Cys207 to Cys226 and Cys272 to Cys306) in E1 are predicted to be important for folding of E1, and forming a functional [E1.E2] monomer were also reported to be important virus assembly and stability of the virion through mutation analysis (Wahid et al., 2013).

N-linked glycosylation is another important protein modification of viral proteins (Hart, 1992). N-linked glycans play a major role in protein folding, stability, entry functions, ER retention function, and cell receptor fusion (Hebert, Zhang, Chen, Foellmer, & Helenius, 1997). All dengue viral species contain two N-linked glycosylation sites at positions 67 and 153 (347 and 433, respectively, according to POLG_DEN26 numbering); both glycosylation sites are functionally and structurally useful (Rey, 2003). But other species of *Flavivirus* contain only one glycosylation site at position 153. The presence of the N-linked glycosylation site at position 67 is predicted to be responsible for higher immunogenic reactions in humans (Barker, Mazumder, Vasudevan, Sagripanti, & Wu, 2009). HCV [E1.E2] contains several N-linked glycosylation sites (NGS). The NGS such as N209 and N234 in domain I are important for induction of humoral and cellular immune responses (Fournillier et al., 2001). Site N556 is predicted to be needed for proper folding in the ER, secretory pathway, and for the formation of its antigenic structure (Lavie et al., 2006). In the sequence region that is used in the model there are four such sites (out of five) for E1 (205, 209, 234, and 305) and nine (out of 11) for E2 (430, 448, 476, 532, 540, 556, 576, 623, and 645) (Helle et al., 2010; Jacob, Perdivara, Przybylski, & Tomer, 2008; Khan et al., 2014). The increase in the number of N-linked glycosylation in HCV envelope proteins (E1: 3 N-linked glycosylation sites and E2: 11 N-linked glycosylation sites) compared with the envelop protein of dengue (2 N-linked glycosylation sites) may be due to the exposure of E1 and E2 polypeptide before they form a compact [E1.E2] structure. N205, N234, and N305 for E1 are partially exposed and for E2, N430, N532, N577, N623, and N695 are exposed. From our model, it is clear that N430, N556, and N623 are critical for the folding of E1 and interact with E2. The sugar attached to these sites will not only anchor E2 into the membrane, but also guide the E1 to fold over E2 and expose the peptide bond between 383 and 384 to be cleaved. Experimentally, it has been shown that the residues from aa476 to 494 of E2 including the “WHY” motif (aa487–489) form a loop region and are important for association of E1 and E2 (Yi et al., 1997). Within this region, the glycosylation site at N476 and the HVR-2 (from aa474–482) of E2 are present which participate in the association of E1 and E2. An overview of additional well known functional sites mapped to the model provides a glimpse of how the model can be used to better understand the envelope proteins and their interactions. We believe from the number of favorable residue–residue interactions including backbone–backbone hydrogen between E1 and E2 in the [E1.E2] monomer, the folding of E1 polypeptide is guided by the folded E2 polypeptide. Table 1 lists the residues involved at the interface of E1 and E2. The interactions between these residues at the interface form a stable [E1.E2] monomer.

Prediction of exposed residues

To identify exposed residues, occluded (buried) molecular surface calculations were performed for the [E1.E2] monomer. A total of 77 amino acid residues (26 from E1 and 51 from E2) for the [E1.E2] monomer which have exposed surface area ratio more than 50% were identified. Three or more residues in our model with exposed surface ratio more than 50% are listed as bold underline in Table 2. In E1, we have identified three regions containing four or more contiguous amino acid sequences with greater than 50% exposed surface ratio in the monomer. They are (1) 238-CWV-240, (2) 274-ALY-VGD-279, and (3) 299-WTTQDC-304. In E2, we have identified six regions containing three or more contiguous amino acid sequences with greater than 50% exposed surface ration. They are (1) 442-FYQ-444, (2) 531-AND-533, (3) 587-RKY-589, (4) 625-TIFK-628, (5) 681-TLPA-684, and (6) 706-GSS-708. In our model of [E1.E2] monomer, the residues from 296–312 map to the fusion loop of dengue. In dengue and other *Flaviviruses*, it has been shown that this highly conserved fusion loop region in domain II is critical for infectivity, and broadly neutralizing human monoclonal antibodies target this region (Costin et al., 2013). In our model eight residues in this region are exposed with four of them being sequential (Table 2).

Prediction of the sequential antigenic epitopes

Several researchers have suggested that HCV envelope glycoprotein is potential target for vaccine development (Beyene, Basu, Meyer, & Ray, 2002; Law et al., 2013). Most of the prediction of epitopes are based on E2 sequence (Ikram, Anjum, & Tahir, 2014) not for the HCV [E1.E2] complex. Our result shows 18 sequential epitopes present in the sequence of [E1.E2] that is part of our model (Table 3). There are fifteen out of eighteen sites which have at least one amino acid that is exposed and ten sites have two or more amino acids that are exposed. In HCV the region around 299 has been implicated as essential for human neutralizing monoclonal antibody action (Lavillette et al., 2007; Meunier et al., 2008). Recently, Kachko et al. identified regions aa264–318 of E1 and aa448–483 and aa496–515 of E2 to be important for neutralization (Kachko et al., 2011). Several amino acids in these regions are exposed in our model with majority of them in the aa264–318 region (Table 2). All of the epitopes shown in Table 3 are also experimentally validated and are present in immune epitope database (IEDB) either as the full peptide or as a sub-sequence (Salimi, Fleri, Peters, & Sette, 2012). An evaluation of the peptides available from IEDB reveals that almost the entire E1 and E2 sequence is represented as epitopes in the database, thus making it difficult for an experimentalist to choose which ones to concentrate on for further studies. What we provide in this study is a way to prioritize targets for further experimentation based on location on the 3D model, available functional information, sequence conservation, and negatively selected sites. This methodology is similar to our diagnostics and vaccine target discovery pipeline that we used for dengue E protein (Mazumder et al., 2007; Sagripanti et al., 2011) where we identified specific targets. Integration of results from all the analysis performed in this study, we identified 20 singular sites (L275, Y276, F285, L286, Q302, G303, I308, Y309, D533, S557, T558, P582, T583, R587, F627, E661, L662, P683, S707, and S708) which have antigenic propensity, have two or more sequential amino acids exposed, and are highly conserved within all prototypes of HCV.

Selection pressure studies

In E1, 48 of the total 192 residues were identified to be under negative selection pressure and 46 residues were positively selected. For E2, we found 90 negatively selected sites and 45 positively selected sites. So, for E1, ~25% of the protein is under negative or purifying selection and 10% of the protein is under positive or pervasive selection. For E2, 25% of the protein is negatively selected and 12.3% of the protein is positively selected (Table 4). These results show that E1 and E2 both have undergone reasonable amount of negative selection which could be exploited for development of vaccines using these negatively selected sites as potent drug targets. Taken together the 48 and 90 negatively selected sites in E1 and E2, respectively, could be of special interest as potential candidate sites for developing vaccines against HCV (Figure 4 and Supplementary Table 1). Of the 20 singular sites mentioned in the previous section, 12 sites are under negative selection pressure (Y276, G303, D533, R587, F627, P683, S707, and S708), and therefore they and their corresponding eight peptides (²⁶⁸SATLCSALY²⁷⁶, ²⁸⁵FLVGQLFTFSPRRHW²⁹⁹, ³⁰²QGCNCSIYPGH³¹², ⁵²¹RSGAPTYSWGENDTDV⁵³⁸, ⁵⁷⁷NTLHCPTDCFR⁵⁸⁷, ⁵⁹⁴YSRCGSGPWITPRLVDYPYRLWHYPCTINYTIFKIRMY⁶³², ⁶⁸²LPALSTGLIHL⁶⁹², ⁶⁹⁵NIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARV⁷³¹) can be considered as vaccine targets. Of special interest is ²⁸⁵FLVGQLFTFSPRRHW²⁹⁹ which maps to the region of dengue protein identified as vaccine target by our group (Mazumder et al., 2007; Sagripanti et al., 2011) (Figure 2). This peptide has five negatively selected sites (L286, V287, G288, T292, and G303) with three of them sequential and 4 amino acids exposed (F285, L286, T292, R296). Recent work has identified sites 502 to 520 playing a key role in cell entry by influencing the association of the viral particle with co-receptors and neutralizing antibodies (Lavie et al., 2014). It is interesting to note that several of the sites in this region are identified in our study as negatively selected (Figure 4).

Conclusion

In this paper, we describe a functional knowledge guided monomer model for the majority of the ectodomain sequences of E1 and E2 by using the crystal structure of the core domain of E2 and one of the class II fusion proteins, dengue virus serotype-2 envelope glycoprotein (1OAN) as template. The [E1.E2] model provide for the first time predictions of priority targets that can be viewed on the structure. Due to the compact core domain of E2 and based on our analysis, we predict that the N-terminal 115 amino acid residues of E1 forms a part of domains I and II. Also, the folding of part of E1 protein is guided by the folding of E2. In HCV, an additional 112 amino acids exist between the functional part of E1 and E2 that are involved in virus infection and a part of them forms a transmembrane domain and is inserted into a membrane. Due to this, it might be difficult to obtain a single crystal of the whole length E1 and E2. It might be possible to verify our homology model by including window of 10–15 amino acids from 319 in E1 for the expression and crystallization following the procedure for dengue envelope protein. Another method to verify this model will be to synthesize overlapping peptides within 115 amino acids in the N-terminal part of E1 and compete against the binding site of the corresponding peptides in E1–E2. As more and more biochemical and biological information are made available, this model will be further refined.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. CR Vinayaka for his comments and Dr. MR Pillai for providing support to AN. We would also like to thank Dr B Korba and Dr M Major for initial discussions on HCV and envelope protein characteristics.

Funding

This work was supported in part by NIH grants RO1 CA135069, U01 CA168926, DA034978, GM093939, the BIT Core of the University of California San Diego Center for AIDS Research (P30 AI036214), and funds from GWU.

References

- Afzal S, Idrees M, Hussain M. De Novo modeling of envelope 2 protein of HCV isolated from Pakistani patient and epitopes prediction for vaccine development. *Journal of Translational Medicine*. 2014; 12:1–9. [PubMed: 24387290]
- Ashfaq UA, Javed T, Rehman S, Nawaz Z, Riazuddin S. An overview of HCV molecular biology, replication and immune responses. *Virology Journal*. 2011; 8:1–10. [PubMed: 21205327]
- Barker WC, Mazumder R, Vasudevan S, Sagripanti JL, Wu CH. Sequence signatures in envelope protein may determine whether flaviviruses produce hemorrhagic or encephalitic syndromes. *Virus Genes*. 2009; 39:1–9. [PubMed: 19283462]
- Benkert P, Kunzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Research*. 2009; 37:W510–W514. [PubMed: 19429685]
- Beyene A, Basu A, Meyer K, Ray R. Hepatitis C virus envelope glycoproteins and potential for vaccine development. *Vox Sanguinis*. 2002; 83(Suppl 1):27–32. [PubMed: 12617098]
- Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins*. 2007; 66:778–795. [PubMed: 17186527]
- Brown EN, Ramaswamy S. Quality of protein crystal structures. *Acta Crystallographica Section D: Biological Crystallography*. 2007; 63:941–950. [PubMed: 17704562]
- Callens N, Ciczora Y, Bartosch B, Vu-Dac N, Cosset FL, Pawlotsky JM, Dubuisson J. Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein E2 contribute to virus entry. *Journal of Virology*. 2005; 79:15331–15341. [PubMed: 16306604]
- Chakrabarti S, Sowdhamini R. Regions of minimal structural variation among members of protein domain superfamilies: Application to remote homology detection and modelling using distant relationships. *FEBS Letters*. 2004; 569:31–36. [PubMed: 15225604]
- Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, van Gunsteren WF. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*. 2005; 26:1719–1751. [PubMed: 16211540]
- Cocquerel L, Voisset C, Dubuisson J. Hepatitis C virus entry: Potential receptors and their biological functions. *Journal of General Virology*. 2006; 87:1075–1084. [PubMed: 16603507]
- Costin JM, Zaitseva E, Kahle KM, Nicholson CO, Rowe DK, Graham AS, Isern S. Mechanistic study of broadly neutralizing human monoclonal antibodies against dengue virus that target the fusion loop. *Journal of Virology*. 2013; 87:52–66. [PubMed: 23077306]
- Delpont W, Scheffler K, Seoighe C. Models of coding sequence evolution. *Brief Bioinform*. 2009; 10:97–109. [PubMed: 18971241]
- Drummer HE, Boo I, Pountourios P. Mutagenesis of a conserved fusion peptide-like motif and membrane-proximal heptad-repeat region of hepatitis C virus glycoprotein E1. *Journal of General Virology*. 2007; 88:1144–1148. [PubMed: 17374757]
- DuBois RM, Vaney MC, Tortorici MA, Kurdi RA, Barba-Spaeth G, Krey T, Rey FA. Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature*. 2013; 493:552–556. [PubMed: 23292515]

- Dubuisson J, Rice CM. Hepatitis C virus glycoprotein folding: Disulfide bond formation and association with calnexin. *Journal of Virology*. 1996; 70:778–786. [PubMed: 8551615]
- Fiser A, Sali A. Modeller: Generation and refinement of homology-based protein structure models. *Methods in Enzymology*. 2003; 374:461–491. [PubMed: 14696385]
- Flint M, Maidens C, Loomis-Price LD, Shotton C, Dubuisson J, Monk P, McKeating JA. Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. *Journal of Virology*. 1999; 73:6235–6244. [PubMed: 10400713]
- Flint M, Thomas JM, Maidens CM, Shotton C, Levy S, Barclay WS, McKeating JA. Functional analysis of cell surface-expressed hepatitis C virus E2 glycoprotein. *Journal of Virology*. 1999; 73:6782–6790. [PubMed: 10400776]
- Forns X, Thimme R, Govindarajan S, Emerson SU, Purcell RH, Chisari FV, Bukh J. Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees. *Proceedings of the National Academy of Sciences*. 2000; 97:13318–13323.
- Fournillier A, Wychowski C, Boucreux D, Baumert TF, Meunier JC, Jacobs D, Inchauspe G. Induction of hepatitis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites. *Journal of Virology*. 2001; 75:12088–12097. [PubMed: 11711599]
- Garry RF, Dash S. Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins. *Virology*. 2003; 307:255–265. [PubMed: 12667795]
- Geiss BJ, Stahla H, Hannah AM, Gari AM, Keenan SM. Focus on flaviviruses: Current and future drug targets. *Future Medicinal Chemistry*. 2009; 1:327–344. [PubMed: 20165556]
- Gibrat JF, Garnier J, Robson B. Further developments of protein secondary structure prediction using information theory new parameters and consideration of residue pairs. *Journal of Molecular Biology*. 1987; 198:425–443. [PubMed: 3430614]
- Goffard A, Callens N, Bartosch B, Wychowski C, Cosset FL, Montpellier C, Dubuisson J. Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins. *Journal of Virology*. 2005; 79:8400–8409. [PubMed: 15956584]
- Grove J, Nielsen S, Zhong J, Bassendine MF, Drummer HE, Balfe P, McKeating JA. Identification of a residue in hepatitis C virus E2 glycoprotein that determines scavenger receptor BI and CD81 receptor dependency and sensitivity to neutralizing antibodies. *Journal of Virology*. 2008; 82:12020–12029. [PubMed: 18829747]
- Hart GW. Glycosylation. *Current Opinion in Cell Biology*. 1992; 4:1017–1023. [PubMed: 1485955]
- Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A. The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin. *The Journal of Cell Biology*. 1997; 139:613–623. [PubMed: 9348279]
- Helle F, Dubuisson J. Hepatitis C virus entry into host cells. *Cellular and Molecular Life Sciences*. 2008; 65:100–112. [PubMed: 17914604]
- Helle F, Duverlie G, Dubuisson J. The hepatitis C virus glycan shield and evasion of the humoral immune response. *Viruses*. 2011; 3:1909–1932. [PubMed: 22069522]
- Helle F, Goffard A, Morel V, Duverlie G, McKeating J, Keck ZY, Voisset C. The neutralizing activity of anti-hepatitis C virus antibodies is modulated by specific glycans on the E2 envelope protein. *Journal of Virology*. 2007; 81:8101–8111. [PubMed: 17522218]
- Helle F, Vieyres G, Elmkrief L, Popescu CI, Wychowski C, Descamps V, Dubuisson J. Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. *Journal of Virology*. 2010; 84:11905–11915. [PubMed: 20844034]
- Houghton M, Abrignani S. Prospects for a vaccine against the hepatitis C virus. *Nature*. 2005; 436:961–966. [PubMed: 16107836]
- Iacob RE, Keck Z, Olson O, Fong SK, Tomer KB. Structural elucidation of critical residues involved in binding of human monoclonal antibodies to hepatitis C virus E2 envelope glycoprotein. *Biochimica et Biophysica Acta (BBA)*. 2008; 1784:530–542. [PubMed: 18230369]
- Iacob RE, Perdivara I, Przybylski M, Tomer KB. Mass spectrometric characterization of glycosylation of hepatitis C virus E2 envelope glycoprotein reveals extended microheterogeneity of N-glycans. *Journal of the American Society for Mass Spectrometry*. 2008; 19:428–444. [PubMed: 18187336]

- Ikram A, Anjum S, Tahir M. *In silico* identification and conservation analysis of B-cell and T-cell epitopes of hepatitis C virus 3a genotype enveloped glycoprotein 2 from Pakistan: A step towards heterologous vaccine design. *Hepatitis Monthly*. 2014; 14:e9832. [PubMed: 24976845]
- Kachko A, Kochneva G, Sivolobova G, Grazhdantseva A, Lupan T, Zubkova I, Major ME. New neutralizing antibody epitopes in hepatitis C virus envelope glycoproteins are revealed by dissecting peptide recognition profiles. *Vaccine*. 2011; 30:69–77. [PubMed: 22041300]
- Kanai R, Kar K, Anthony K, Gould LH, Ledizet M, Fikrig E, Modis Y. Crystal structure of west nile virus envelope glycoprotein reveals viral surface epitopes. *Journal of Virology*. 2006; 80:11000–11008. [PubMed: 16943291]
- Khan AG, Whidby J, Miller MT, Scarborough H, Zatorski AV, Cygan A, Marcotrigiano J. Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature*. 2014; 509:381–384. [PubMed: 24553139]
- Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*. 1990; 276:172–174. [PubMed: 1702393]
- Kong L, Giang E, Nieuwma T, Kadam RU, Cogburn KE, Hua Y, Law M. Hepatitis C virus E2 envelope glycoprotein core structure. *Science*. 2013; 342:1090–1094. [PubMed: 24288331]
- Kosakovsky P, Frost SD. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*. 2005; 22:1208–1222. [PubMed: 15703242]
- Krekulova L, Rehak V, Riley LW. Structure and functions of hepatitis C virus proteins: 15 years after. *Folia Microbiologica*. 2006; 51:665–680. [PubMed: 17455808]
- Krey T, d'Alayer J, Kikuti CM, Saulnier A, Damier-Piolle L, Petitpas I, Rey FA. The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLoS Pathogens*. 2010; 6:e1000762. [PubMed: 20174556]
- Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*. 2008; 9:299–306. [PubMed: 18417537]
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereo-chemical quality of protein structures. *Journal of Applied Crystallography*. 1993; 26:283–291.
- Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *Journal of Molecular Biology*. 1993; 231:1049–1067. [PubMed: 8515464]
- Lavie M, Goffard A, Dubuisson J. HCV glycoproteins: Assembly of a functional E1–E2 heterodimer. In: Tan, SL., editor. *Hepatitis C viruses: Genomes and molecular biology*. Norfolk: Horizon Bioscience; 2006. p. 121–150.
- Lavie M, Sarrazin S, Montserret R, Descamps V, Baumert TF, Duverlie G, Dubuisson J. Identification of conserved residues in hepatitis C virus envelope glycoprotein E2 that modulate virus dependence on CD81 and SRB1 entry factors. *Journal of Virology*. 2014; 88:10584–10597. [PubMed: 24990994]
- Lavillette D, Pecher EI, Donot P, Fresquet J, Molle J, Corbau R, Cosset FL. 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. *Journal of Virology*. 2007; 81:8752–8765. [PubMed: 17537855]
- Law JL, Chen C, Wong J, Hockman D, Santer DM, Frey SE, Houghton M. 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*. 2013; 8:e59776. [PubMed: 23527266]
- Li L, Jose J, Xiang Y, Kuhn RJ, Rossmann MG. Structural changes of envelope proteins during alphavirus fusion. *Nature*. 2010; 468:705–708. [PubMed: 21124457]
- Lidenbach, BD., Thiel, HJ., Rice, CM. *Flaviviridae: The viruses and their replication*. In: DM, K., PM, H., editors. *Fields virology*. 5. Vol. 1. Philadelphia, PA: Lippincott-Raven; 2007. p. 1101–1152.
- Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Bryant SH. CDD: A curated Entrez database of conserved domain alignments. *Nucleic Acids Research*. 2003; 31:383–387. [PubMed: 12520028]

- Maurin G, Fresquet J, Granio O, Wychowski C, Cosset FL, Lavillette D. Identification of interactions in the E1E2 heterodimer of hepatitis C virus important for cell entry. *Journal of Biological Chemistry*. 2011; 286:23865–23876. [PubMed: 21555519]
- Mazumder R, Hu ZZ, Vinayaka CR, Sagripanti JL, Frost SD, Kosakovsky Pond SL, Wu CH. Computational analysis and identification of amino acid sites in dengue E proteins relevant to development of diagnostics and vaccines. *Virus Genes*. 2007; 35:175–186. [PubMed: 17508277]
- Mazumder R, Vasudevan S. Structure-guided comparative analysis of proteins: Principles, tools, and applications for predicting function. *PLoS Computational Biology*. 2008; 4:e1000151. [PubMed: 18818720]
- Mazumder R, Vasudevan S, Nikolskaya AN. Protein functional annotation by homology. *Methods in Molecular Biology*. 2008; 484:465–490. [PubMed: 18592196]
- Meunier JC, Russell RS, Goossens V, Priem S, Walter H, Depla E, Purcell RH. Isolation and characterization of broadly neutralizing human monoclonal antibodies to the E1 glycoprotein of hepatitis C virus. *Journal of Virology*. 2008; 82:966–973. [PubMed: 17977972]
- Modis Y, Ogata S, Clements D, Harrison SC. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proceedings of the National Academy of Sciences*. 2003; 100:6986–6991.
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*. 2012; 8:e1002764. [PubMed: 22807683]
- Nayak V, Dessau M, Kucera K, Anthony K, Ledizet M, Modis Y. Crystal structure of dengue virus type 1 envelope protein in the postfusion conformation and its implications for membrane fusion. *Journal of Virology*. 2009; 83:4338–4344. [PubMed: 19244332]
- NCBI_Resource_Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Research*. 2014; 42:D7–17. [PubMed: 24259429]
- Op De Beeck A, Cocquerel L, Dubuisson J. Biogenesis of hepatitis C virus envelope glycoproteins. *Journal of General Virology*. 2001; 82:2589–2595. [PubMed: 11602769]
- Pattabiraman N, Ward KB, Fleming PJ. Occluded molecular surface: Analysis of protein packing. *Journal of Molecular Recognition*. 1995; 8:334–344. [PubMed: 9052974]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera? A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 2004; 25:1605–1612. [PubMed: 15264254]
- Qureshi SA. Hepatitis C virus–biology, host evasion strategies, and promising new therapies on the horizon. *Medicinal Research Reviews*. 2007; 27:353–373. [PubMed: 16958135]
- Rey FA. Dengue virus envelope glycoprotein structure: New insight into its interactions during viral entry. *Proceedings of the National Academy of Sciences*. 2003; 100:6899–6901.
- Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature*. 1995; 375:291–298. [PubMed: 7753193]
- Roussel A, Lescar J, Vaney MC, Wengler G, Rey FA. Structure and interactions at the viral surface of the envelope protein E1 of Semliki Forest virus. *Structure*. 2006; 14:75–86. [PubMed: 16407067]
- Rychlowska M, Owsianka AM, Fong SK, Dubuisson J, Bienkowska-Szewczyk K, Patel AH. Comprehensive linker-scanning mutagenesis of the hepatitis C virus E1 and E2 envelope glycoproteins reveals new structure-function relationships. *Journal of General Virology*. 2011; 92:2249–2261. [PubMed: 21697343]
- Sabahi A, Uprichard SL, Wimley WC, Dash S, Garry RF. Unexpected structural features of the hepatitis C virus envelope protein 2 ectodomain. *Journal of Virology*. 2014; 88:10280–10288. [PubMed: 24991010]
- Sagripanti J-L, Mazumder R, Wu CH. Amino acid sites in Flavivirus E proteins useful for development of diagnostics and vaccines. USPTO. USA. 2011 US7943148B1.
- Salimi N, Fleri W, Peters B, Sette A. The immune epitope database: A historical retrospective of the first decade. *Immunology*. 2012; 137:117–123. [PubMed: 22681406]
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*. 2003; 31:3381–3385. [PubMed: 12824332]
- Stadler K, Allison SL, Schalich J, Heinz FX. Proteolytic activation of tick-borne encephalitis virus by furin. *Journal of Virology*. 1997; 71:8475–8481. [PubMed: 9343204]

- Stamatakis Z, Coates S, Abrignani S, Houghton M, McKeating JA. Immunization of human volunteers with hepatitis C virus envelope glycoproteins elicits antibodies that cross-neutralize heterologous virus strains. *Journal of Infectious Diseases*. 2011; 204:811–813. [PubMed: 21788452]
- Stiasny K, Allison SL, Marchler-Bauer A, Kunz C, Heinz FX. Structural requirements for low-pH-induced rearrangements in the envelope glycoprotein of tick-borne encephalitis virus. *Journal of Virology*. 1996; 70:8142–8147. [PubMed: 8892942]
- Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: Pan-specific MHC class I pathway epitope predictions. *Immunogenetics*. 2010; 62:357–368. [PubMed: 20379710]
- UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*. 2012; 40:D71–75. [PubMed: 22102590]
- Vieyres G, Dubuisson J, Pietschmann T. Incorporation of hepatitis C virus E1 and E2 glycoproteins: The keystones on a peculiar virion. *Viruses*. 2014; 6:1149–1187. [PubMed: 24618856]
- Voisset C, Dubuisson J. Functional hepatitis C virus envelope glycoproteins. *Biology of the Cell*. 2004; 96:413–420. [PubMed: 15325070]
- Voss JE, Vaney MC, Duquerroy S, Vonnrhein C, Girard-Blanc C, Crublet E, Rey FA. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature*. 2010; 468:709–712. [PubMed: 21124458]
- Wahid A, Helle F, Descamps V, Duverlie G, Penin F, Dubuisson J. Disulfide bonds in hepatitis C virus glycoprotein E1 control the assembly and entry functions of E2 glycoprotein. *Journal of Virology*. 2013; 87:1605–1617. [PubMed: 23175356]
- Wengler G. Cell-associated west nile flavivirus is covered with E+pre-M protein heterodimers which are destroyed and reorganized by proteolytic cleavage during virus release. *Journal of Virology*. 1989; 63:2521–2526. [PubMed: 2724410]
- Yagnik AT, Lahm A, Meola A, Roccasecca RM, Ercole BB, Nicosia A, Tramontano A. A model for the hepatitis C virus envelope glycoprotein E2. *Proteins: Structure, Function, and Genetics*. 2000; 40:355–366.
- Yi M, Nakamoto Y, Kaneko S, Yamashita T, Murakami S. Delineation of regions important for heteromeric association of hepatitis C virus E1 and E2. *Virology*. 1997; 231:119–129. [PubMed: 9143310]
- Yu X, Qiao M, Atanasov I, Hu Z, Kato T, Liang TJ, Zhou ZH. Cryo-electron microscopy and three-dimensional reconstructions of hepatitis C virus particles. *Virology*. 2007; 367:126–134. [PubMed: 17618667]

Figure 1. Ribbon representation of dengue virus envelope protein (PDB id: 1OAN) (A) showing three structural domains (I, II, and III) identified in the crystal structure. (B) The dark blue-colored ribbon is 120 amino acids at the N-terminal (see text).

(A)

1OAN	1	MRCIGISNRDFVEGVSGGSWVDIVLEHGSVCVTTMAK-----NKPTLDFEL	45
POLG_HCV1	205	NDC---PNSSIVYEAA-----DAILHTPGCVPCVREGNASRCWVAMTPTV	246
		C N V D *L CV * * *	
fusion peptide/loop (HCV)			
1OAN	46	IKTEAKQPATLRKYCI EAKLTNTT TESRCPTQGEPTLN EEQ DKRFVCKHS	95
POLG_HCV1	247	ATRDGKLPATQLRRH ITDLLVGSATL--- CSALYVGDLCGSVFLVGQLFTF	293
		K PAT : I * T C L	
Vaccine target (Dengue)			
1OAN	96	MVDRGW---GNGCGLFGKGGIVTCAM	118
POLG_HCV1	294	SPRRHWTTQGCNC SIYPGH ITGHRMA	319
		R W G C *#	
(B)			
1OAN	324	VQYEGDGSPCKIPFEIMDL---EKRHVLGRLITVNPIVTEKDSPVNIEA	369
POLG_HCV1	646	WTRGE---RCDLED RDRSELSPL LLTTTQWQVLPCSF TTLP ALSTGLIH	691
		C * * T *	
1OAN	370	EPPFGDSYIIIGVEPGQLKLNWFKK	394
POLG_HCV1	692	LHQNIVDVQYLYGVGSSIASWAIKW	716
		* * K	

Figure 2. Function and structure-guided manually curated pairwise alignment of template dengue virus envelope protein structure (PDB id: 1OAN) and [E1.E2] from POLG_HCV1 for two fragments (A) between amino acid residues 1–118 (dengue) and 205–319 (HCV1) and (B) between 324–394 (dengue) and 646–716(HCV1) used in the homology modeling.

Figure 3. Model of HCV envelope proteins [E1.E2]. Three domains of the [E1.E2] monomer in ribbon representation. Arrows show the polypeptide breaks for E1 and E2.

Numbers indicate amino acid position/position in polyprotein

Experimentally validated sites are in red. Negatively selected sites in purple and bold.

Sites that are in italics are predicted/known N-glycosylation sites that are negatively selected

Sites conserved in all 6 genotypes are underlined

N- and O-glycosylation sites in E2 that are critical for HCVpp entry are in italics and bold [pmid: 17507469]

Figure 4.

N- and O- glycosylation sites in [E1.E2] including all negatively selected sites.

Table 1

Residues at the interface of E1 and E2 in the [E1.E2] monomer model.

From E1 (54 residues)	From E2 (61 residues)
N205, D206, C207, P208, N209, S211, V212, T223, P224, G225, C226, P228, V230, E232, G233, N234, R237, A241, P244, T245, V246, T248, R249, D250, G251, K252, L253, P254, A255, T256, Q257, L258, R259, R260, H261, I262, D263, L264, L265, V266, G267, S268, C281, V287, G288, Q289, F293, P295, I313, T314, G315, H316, R317, M318	N448, S449, S450, G451, C452, P453, D464, I472, S473, Y474, N476, G477, E482, W487, H488, Y489, P490, R492, P493, C494, I496, V497, N540, N541, T542, R543, P544, P545, W549, F550, F560, T561, K562, V563, K588, A592, Y594, S595, R596, C597, G598, P619, T621, V633, G634, D656, R657, V674, L675, P676, C677, S678, F679, T680, L682, L685, S686, T687, L702, Y703, V705

Table 2

Amino acid residues with greater than 50% exposed surface ratio* for the monomer model of [E1.E2]. Residues from E1 are colored red. Three or more consecutive residues exposed are bolded and underlined.

Amino acid symbol and number based on UniProtKB ID POLG_HCV1			
A216	T300	Y485	R630
D218	T301	R521	R639
A219	Q302	A524	W646
A235	G303	P525	L662
C238	S307	Y527	S663
W239	Y309	E531	L665
V240	H421	N532	T681
V246	I422	D533	L682
T270	N430	N577	P683
A274	E431	L579	A684
L275	L433	R587	S686
Y276	F442	K588	Q694
Y277	Y443	Y589	N695
G278	H444	E591	V697
D279	K446	P612	G706
V284	L456	N623	S707
F285	C459	T625	S708
L286	Q467	I626	A710
T292	Y474	F627	I714
R296	P480	K628	W716

* Based on a method reported by Pattabiraman et al. (55).

Table 3

The predicted sequential epitopes in [E1.E2] monomer model. Method integrates prediction of peptide MHC class I binding, proteasomal C-terminal cleavage, and TAP transport efficiency.

Peptides with antigenic propensity	
From E1	
208–231	PNSSIVYE <u>A</u> <u>A</u> <u>D</u> AILHTPGCVPCVR ^{a,b}
243–253	TPT <u>V</u> ATRDGKL
256–265	TQLRRHIDLL
268–276	SATLCS <u>A</u> <u>L</u> <u>Y</u> ^b
285–299	FLVGQLF <u>T</u> FSP <u>R</u> RRHW ^b
302–312	<u>Q</u> GCNC <u>S</u> <u>I</u> YPGH ^b
316–324	HRMAWDMMM
From E2	
434–447	NTGWLAGL <u>F</u> <u>Y</u> <u>Q</u> <u>H</u> <u>K</u> F ^b
454–462	ERLAS <u>C</u> <u>R</u> <u>R</u> <u>L</u> ^b
482–517	QRP <u>Y</u> CWHYPPKPCGIVPAKSVCGPVYCFTPSPVVVG
521–538	<u>R</u> SGAP <u>T</u> <u>Y</u> SWGANDTDVfv ^b
546–562	<u>L</u> GNWFGCTWMNSTGFTK
577–587	<u>N</u> TLHCPTDC <u>F</u> <u>R</u> ^b
594–632	YSRCGSGPWITPRCLVDY <u>P</u> YRLWHYPCTI <u>N</u> <u>Y</u> <u>T</u> <u>I</u> <u>F</u> <u>K</u> IRMY ^b
654–662	LED ^R DRSEL
666–674	LLTTTQWQV
682–692	<u>L</u> <u>P</u> <u>A</u> LSTGLIHL ^b
695–731	<u>N</u> IVDVQYLYGV <u>G</u> <u>S</u> <u>S</u> IASWAIK <u>W</u> EYVLLFLLADARV ^b

^a Amino acid in bold underline has greater than 50% of its exposed surface area ratio with respect to the total surface of the amino acid.

^b Peptides which have two or more exposed amino acids in the [E1.E2] model.

Table 4

Positively and negatively selected sites in E1 and E2.

Protein ^a	Sites ^b (amino acids and nucleotides)	Total ^c	Positive selection ^d	Negative selection ^e
E1	Amino acids	192	46	48
	Nucleotides	576	138	144
E2	Amino acids	363	45	90
	Nucleotides	1089	135	270

^aProtein name.^bAmino acids and nucleotide sites in E1 and E2.^cTotal number of positively or negatively selected amino acids and nucleotides in E1 and E2.^dNumber of positively selected sites (amino acids and nucleotides) in E1 and E2.^eNumber of negatively selected sites (amino acids and nucleotides) in E1 and E2.