LETTER

doi:10.1038/nature13117

Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2

Abdul Ghafoor Khan[1], Jillian Whidby[1], Matthew T. Miller[1], Hannah Scarborough[2], Alexandra V. Zatorski[1], Alicja Cygan[1], Aryn A. Price[2], Samantha A. Yost[1], Caitlin D. Bohannon[2], Joshy Jacob[2], Arash Grakoui[2,3] & Joseph Marcotrigiano[1]

Hepatitis C virus (HCV) is a significant public health concern with approximately 160 million people infected worldwide[1]. HCV infection often results in chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. No vaccine is available and current therapies are effective against some, but not all, genotypes. HCV is an enveloped virus with two surface glycoproteins (E1 and E2). E2 binds to the host cell through interactions with scavenger receptor class B type I (SR-BI) and CD81, and serves as a target for neutralizing antibodies[2–4]. Little is known about the molecular mechanism that mediates cell entry and membrane fusion, although E2 is predicted to be a class II viral fusion protein. Here we describe the structure of the E2 core domain in complex with an antigen-binding fragment (Fab) at 2.4 Å resolution. The E2 core has a compact, globular domain structure, consisting mostly of β-strands and random coil with two small α-helices. The strands are arranged in two, perpendicular sheets (A and B), which are held together by an extensive hydrophobic core and disulphide bonds. Sheet A has an IgG-like fold that is commonly found in viral and cellular proteins, whereas sheet B represents a novel fold. Solution-based studies demonstrate that the full-length E2 ectodomain has a similar globular architecture and does not undergo significant conformational or oligomeric rearrangements on exposure to low pH. Thus, the IgG-like fold is the only feature that E2 shares with class II membrane fusion proteins. These results provide unprecedented insights into HCV entry and will assist in developing an HCV vaccine and new inhibitors.

HCV envelope glycoprotein 2 (E2) is a type I transmembrane protein with an amino-terminal ectodomain connected to a carboxy-terminal transmembrane helix through an amphipathic, α-helical stem (Fig. 1a)[5,6]. E2 is highly modified post-translationally with 9–11 *N*-linked glycosylation sites and 18 cysteine residues that are conserved across all genotypes. For ease of comparison with other genotypes, we refer to the cysteines and *N*-linked glycosylation sites as C1 to C18 and N1 to N11, respectively, with residue numbers from the J6 genotype (2a) given in parentheses. Full-length, E2 ectodomain (eE2) (384–656) was produced in *N*-acetylglucosaminyltransferase I-negative (GnTI−) HEK293T cells by a lentiviral expression system and grown in an adherent cell bioreactor. The resulting eE2 protein is monomeric as determined by non-reducing SDS–polyacrylamide gel electrophoresis (PAGE) and size-exclusion chromatography (Extended Data Fig. 1).

Solution-based studies using limited proteolysis and hydrogen deuterium exchange demonstrated that approximately 80 amino acids on the N terminus (384–463) from hypervariable region (HVR) 1 through to HVR2 are exposed and flexible. This region includes conserved sequences implicated in binding to the cellular receptors (SR-BI and CD81) as well as several epitopes for neutralizing antibodies (Fig. 1 and Extended Data Figs 2 and 3)[7–11]. Various N-terminal deletions were produced to minimize regions of disorder while preserving an even number of cysteines, potentially allowing for intramolecular disulphide-bond formation. All constructs were screened for aggregation by non-reducing SDS–PAGE and size-exclusion chromatography. E2 core (456–656) is soluble, monomeric and maintains similar secondary structure content when compared with eE2 as determined by reactivity towards HCV-infected patient sera (Extended Data Fig. 4a, b) and circular dichroism (data not shown). However, in contrast to eE2, CD81 binding affinity and the efficiency of inhibition of HCV cell culture (HCVcc) entry was diminished for the E2 core (Extended Data Fig. 4c–e). This indicates that the N terminus of eE2 is critical for CD81 interaction and probably undergoes a transition from disorder to order on binding. Alternatively, the N-terminal region may also be ordered through interactions with other factors, for example, E1, apolipoproteins, lipids, cellular receptors, or antibodies.

Monoclonal antibodies were generated against recombinant eE2 and crystals of deglycosylated E2 core were produced in complex with a Fab (2A12) and diffracted to 2.4 Å resolution (Fig. 1 and Extended Data Table 1). The complex structure was determined by molecular replacement using a Fab structure followed by iterative rounds of model building and refinement. The E2 core domain has a globular fold, consisting of mostly β-strands and random coil with two short α-helices, which is consistent with previous spectroscopic studies of eE2 (refs 12, 13). The protein contains two, four-stranded antiparallel β-sheets (termed sheets A and B), the planes of which are approximately perpendicular to each other. The four strands of the N-terminal β-sheet (sheet A) are stabilized by two disulphide bonds, between strands 1 and 3 (C7 (510) and C8 (554)) and the N-terminal loop with strand 4 (C5 (496) and C9 (566)). The loop between strands 2 and 3 contains sequences implicated in CD81 binding and is flexible, similar to the N-terminal CD81 binding sites, which were deleted[14,15]. After strand 4, the polypeptide continues into a long, disordered loop before forming the first short helix (H1) followed by the second β-sheet (sheet B). A second short α-helix (H2) is located between strands 6 and 7. A disulphide bond (C14 (611) and C16 (648)) between strand 6 and the C-terminal strand 8 further stabilizes the fold. The C-terminal strands (7 and 8) are the longest within the protein with approximately nine amino acids each and encompass the 2A12-binding site. 2A12 does not neutralize HCV infection, indicating that the epitope is either buried within the particle or incapable of preventing entry (Extended Data Fig. 4f). The two β-sheets are held together by (1) two disulphide bonds, connecting the loops before strand 1 and after H2 (C4 (488) with C15 (624)) as well as the loops after strand 4 and before H1 (C10 (571) and C13 (601)), and (2) an extensive hydrophobic core consisting of numerous aromatic residues (Extended Data Fig. 5).

HCV belongs to the genus *Hepacivirus* of the Flaviviridae family. Other members of the family include the flavivirus and pestivirus genera, which consist of arthropod-borne viruses and important livestock pathogens, respectively[16]. The flavivirus envelope glycoprotein (E) is a class II fusion protein and HCV E2 was expected to have a similar fold[13,17,18]. All class II fusion proteins have a common elongated structure, consisting of predominantly β-sheets, and exist as homo- or

[1]Center for Advanced Biotechnology and Medicine, Department of Chemistry and Chemical Biology, Rutgers University, 679 Hoes Lane West, Piscataway, New Jersey 08854, USA. [2]Division of Microbiology and Immunology, Emory Vaccine Center, Emory University School of Medicine, 100 Woodruff Circle, Atlanta, Georgia 30322, USA. [3]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, 100 Woodruff Circle, Atlanta, Georgia 30322, USA.

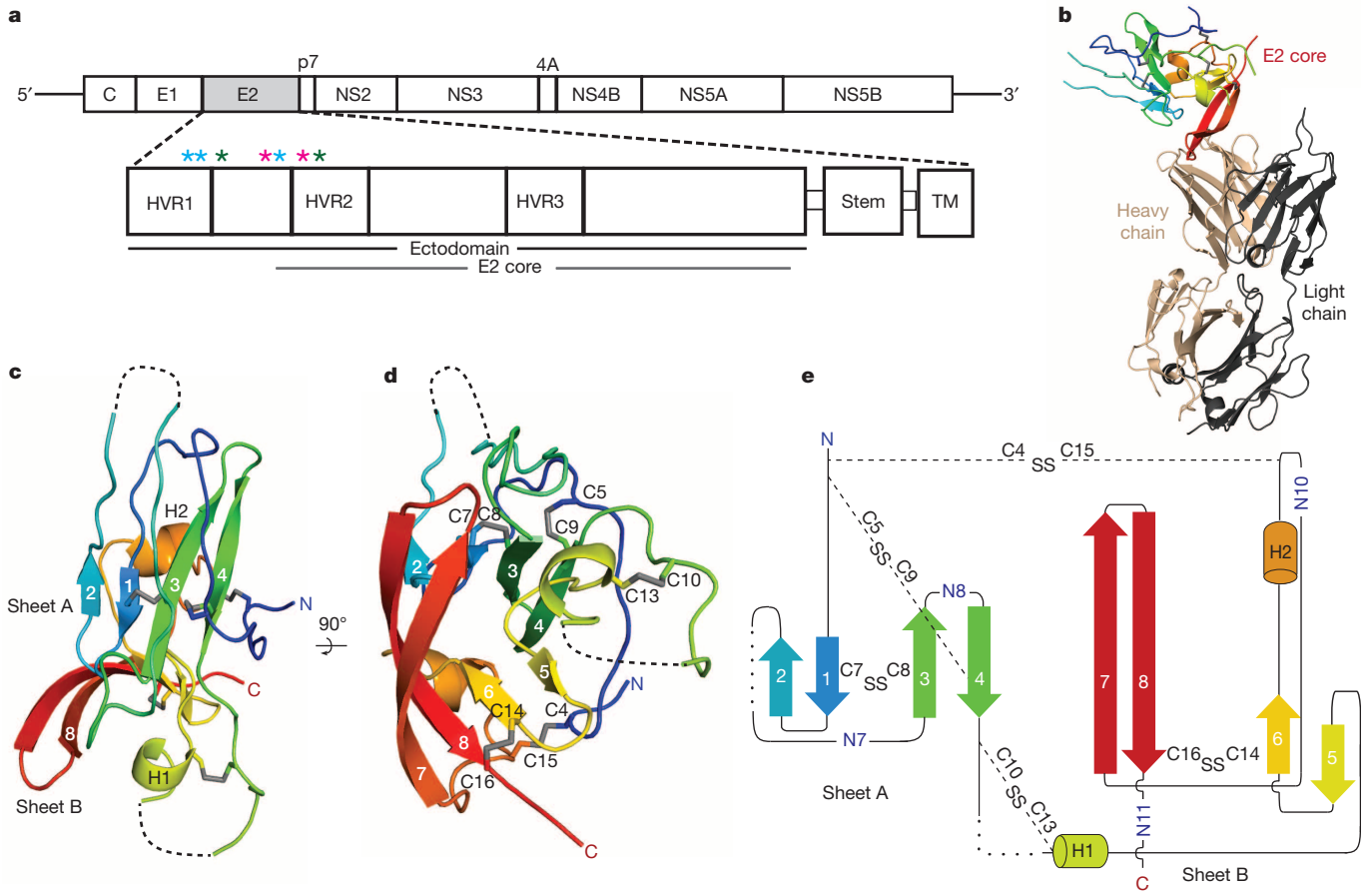


Figure 1 | Overview of HCV E2. **a**, Schematic representation of the HCV genome and E2 domain organization. Full-length eE2 and the crystallization construct are indicated by the black and grey bars, respectively. C, capsid protein; non-structural protein (NS) 2-5B. Asterisks indicate the location of trypsin (blue), chymotrypsin (green) and GluC (magenta) cleavage sites. **b–d**, Ribbon diagram of the E2 core domain bound to Fab 2A12 (**b**) and alone (**c** and **d**). The view in **d** is a 90° rotation about a horizontal axis from **c**. The E2 polypeptide chain is coloured from the N terminus (blue) to C terminus (red). **e**, Topology diagram of E2 core domain, detailing secondary structure elements, disulphide bonds (dashed lines labelled with SS), *N*-linked glycosylation sites and regions of disordered polypeptide (dotted lines).

heterodimers with the membrane-fusion, hydrophobic peptide buried at the dimer interface at neutral pH. On receptor binding and/or exposure to low pH, these proteins undergo self-rearrangement into stable trimers, exposing the fusion peptide and resulting in viral and host membrane fusion. Despite containing a similar extended organization, the recent structure of the pestivirus bovine viral diarrhoea virus (BVDV) E2 glycoprotein does not represent a typical class II fusion protein fold and lacks an apparent fusion peptide, indicating that it is unlikely to be a class II fusion protein[19,20].

Similar to the flavivirus and pestivirus glycoproteins, the HCV E2 core secondary structure consists of predominantly β-sheets and random coil. However, E2 core is a monomer with a compact globular shape, in contrast to the extended structures reported in other viruses. Solution-based small-angle X-ray scattering (SAXS) was used to correlate the crystallographic core domain structure with fully glycosylated eE2 and various fragments. The *ab initio* SAXS envelopes of E2 core and eE2 are similar, with approximately the same radius of gyration (R_g) (Fig. 2a, b and Extended Data Table 2). Glycosylation, which is missing in the E2 core crystal structure, represents roughly one-third of the mass and accounts for the unmodelled areas of the envelopes. Notably, neither the R_g nor the elution profiles on size-exclusion chromatography for fully glycosylated eE2 and E2 core changed significantly at pH 5.0 (Extended Data Fig. 6a, b). These results indicate that unlike class II membrane fusion proteins, E2 does not undergo significant structural rearrangements on exposure to low pH.

SAXS was used to investigate the CD81 binding region on the E2 ectodomain. To simplify data interpretation, eE2(ΔHVR1) was used, as HCV lacking HVR1 remains infectious[21]. The binding site of CD81 was identified by superimposing the SAXS envelopes of eE2(ΔHVR1) alone and in complex with CD81-LEL (Fig. 2c–e). Although CD81-LEL is a dimer in solution (Extended Data Fig. 6c), the extra density in the SAXS envelope is more consistent with monomeric binding; however, a dimer cannot be ruled out.

HCV E2 is modified by *N*-linked glycosylation, which is necessary for proper folding and immune invasion. E2 from the J6 genotype has 11 glycosylation sites. Four of the glycosylation sites are in the flexible N-terminal region, which were deleted, and seven are in the core domain (N5–N11). The location of N7, N8, N10 and N11 are modelled in the final E2 core structure. All of these glycans are present in loop areas, indicating that these sites are solvent exposed and flexible. Mutagenesis studies in HCVcc have shown that N6, N8 and N10 are integral for virus infectivity. Removal of the N6 site results in improved CD81 binding, whereas N8 and N10 mutations destabilize the protein and cause defective particle production[22]. Both sheets have one critical glycosylation site: N8 in sheet A and N10 in sheet B. All four of the observed glycosylation sites are on the periphery of the core and are located on a highly basic surface (Fig. 3). The opposite surface is predominantly hydrophobic and highly conserved when compared to the basic surface. Furthermore, the epitope for antibodies (AR1, AR3A and AR3C) that inhibit E1E2 binding to CD81 is located at the interface of

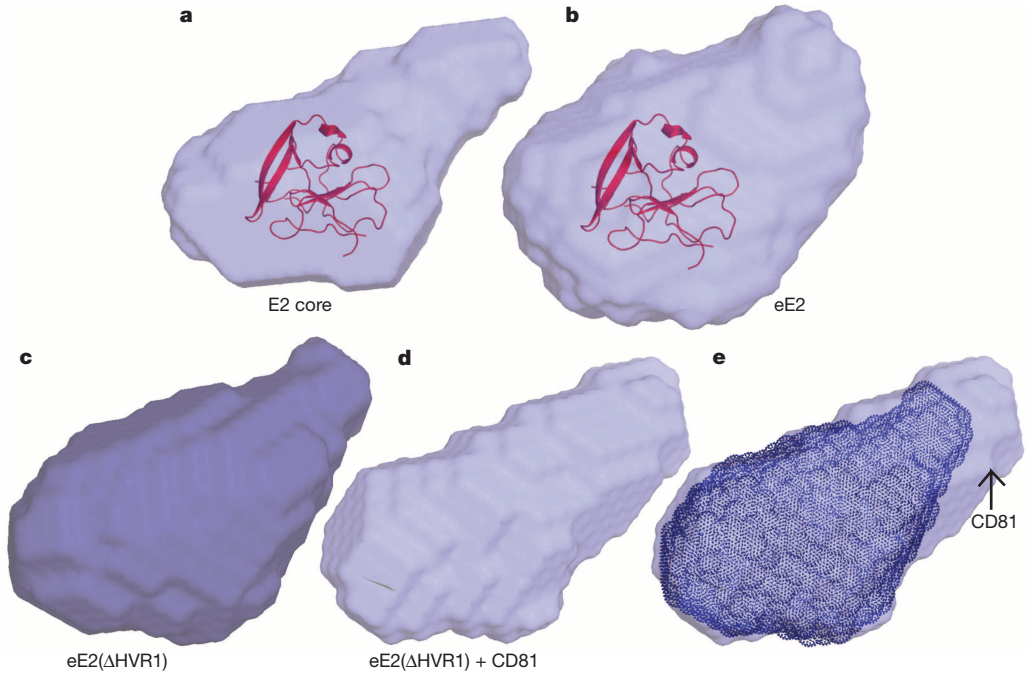


Figure 2 | *Ab initio* SAXS envelopes of E2 core, eE2(ΔHVR1) and eE2. **a**–**d**, SAXS envelopes of glycosylated E2 core (**a**), eE2 (**b**), eE2(ΔHVR1) (**c**) and eE2(ΔHVR1) in complex with CD81-LEL (**d**). The E2 core domain structure has been fitted into **a** and **b**. **e**, Superposition of the SAXS envelopes of eE2(ΔHVR1) alone (**c**) and in complex with CD81-LEL (**d**), highlighting the approximate position of CD81-LEL.

the hydrophobic and basic surfaces, including the N7 glycosylation site (Extended Data Fig. 7a). Interestingly, N7 is only 7 residues away from N6, which has a critical effect on CD81 binding. Epitopes for antibodies (that is, AR5) that block E1E2 heterodimerization are also found on the hydrophobic surface, making it highly plausible that this surface is interacting with E1 in the context of the viral particle[4].

The precise roles played by E1 and E2 in membrane fusion are not fully understood. It has been predicted that amino acids 262–290 in E1

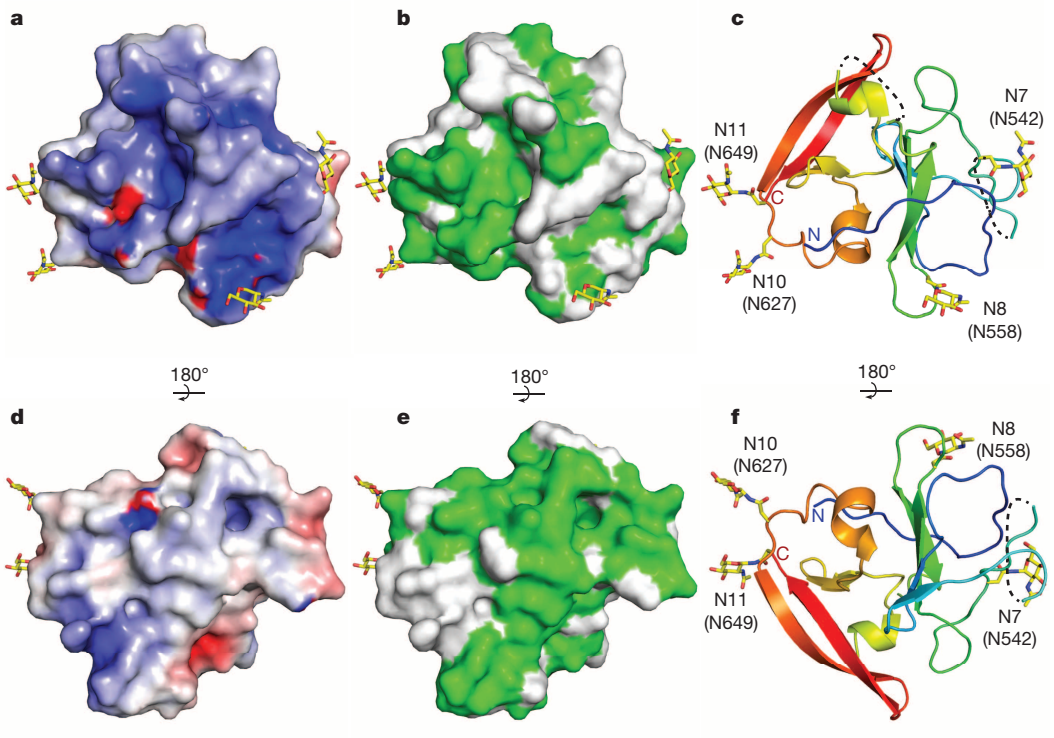


Figure 3 | Surface features of E2. The surface of the E2 core domain coloured for electrostatic potential (**a** and **d**)—blue (basic), white (neutral), red (acidic) at $\pm 5\,kT\,e^{-1}$—and sequence identity (**b** and **e**) (green) from the alignment in Extended Data Fig. 2. **c**, **f**, Ribbon diagrams highlighting the location of the *N*-linked glycosylation sites. The orientations of **d**–**f** as well as **a**–**c** are identical. The orientation in **d**–**f** is rotated 180° about a horizontal axis from the view in **a**–**c**.

as well as 416–430, 504–522 and 604–624 in E2 are important for fusion[13,23,24]. In the structure, the potential fusion regions in E2 (504–522 and 604–624) are located in secondary structure elements within the hydrophobic core and therefore unlikely to serve as the fusion peptide. Furthermore, size-exclusion chromatography and SAXS analyses at low pH indicate that E2 does not undergo oligomeric or structural rearrangement. Thus, it seems unlikely that E2 has a direct role in membrane fusion. However, it is possible that E1 alone or the E1E2 heterodimer has a major role in the fusion process.

Structural comparison of the HCV E2 core domain with all known folds in the Protein Data Bank using the Dali server[25] identified proteins with IgG-like folds similar to the N-terminal sheet A, none of which is a class II fusion protein, although IgG-like folds are common in these proteins. The server failed to identify any statistically significant structures to sheet B, suggesting a novel fold. During the review process of this manuscript, a structure for HCV E2 from genotype 1a was published[26]. The core domain of both structures is highly similar with a root mean squared deviation of 0.8 Å for similar carbon α atoms. Our biochemical and structural data provide valuable information towards defining the role of E2 and establish a foundation for further studies in understanding HCV entry and infection.

METHODS SUMMARY

Recombinant, full-length E2 ectodomain (384–656, genotype J6), eE2(ΔHVR1) (413–656) and E2 core (456–656) were produced in HEK293T GnTI− cells[27] by a lentiviral expression system with a C-terminal, protein-A tag. Stable cell lines were grown in an adherent cell bioreactor. Collected supernatants were purified over IgG sepharose and the E2 proteins were eluted by PreScission Protease digestion. The eluted protein was deglycosylated with endoglycosidase H (EndoH) and purified to homogeneity by heparin affinity and size-exclusion chromatography. Monoclonal antibody 2A12 secreted by hybridoma cells maintained in suspension culture was purified over protein G resin. Fab was generated using papain digestion and subtractive protein A chromatography. E2 and 2A12 Fab complex was purified by size-exclusion chromatography and crystals were obtained in 18% (w/v) PEG 3350, 0.5 M $MgCl_2$, 0.1 M HEPES (pH 7.5), 15% dioxane and 2% formamide at 20 °C by the hanging-drop vapour diffusion method. The crystals belong to space group $P2_12_12$ with cell parameters $a = 85.96$ Å, $b = 194.57$ Å and $c = 37.92$ Å. The structure was determined by molecular replacement to 2.4 Å resolution using a mouse Fab (Protein Data Bank accession 2GSI). The final model has an R_{work} and R_{free} of 0.217 and 0.269, respectively. SAXS measurements were performed on eE2, E2 core, eE2(ΔHVR1) and eE2(ΔHVR1)/CD81-LEL. Data were analysed using BioXTAS RAW[28] and applications from the ATSAS[29] program suite.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 November 2013; accepted 31 January 2014.
Published online 19 February 2014.

1. Lavanchy, D. Evolving epidemiology of hepatitis C virus. *Clin. Microbiol. Infect.* **17,** 107–115 (2011).
2. Pileri, P. *et al.* Binding of hepatitis C virus to CD81. *Science* **282,** 938–941 (1998).
3. Scarselli, E. *et al.* The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* **21,** 5017–5025 (2002).
4. Sautto, G., Tarr, A. W., Mancini, N. & Clementi, M. Structural and antigenic definition of hepatitis C virus E2 glycoprotein epitopes targeted by monoclonal antibodies. *Clin. Dev. Immunol.* **2013,** 450963 (2013).
5. Michalak, J. P. *et al.* Characterization of truncated forms of hepatitis C virus glycoproteins. *J. Gen. Virol.* **78,** 2299–2306 (1997).
6. Drummer, H. E. & Poumbourios, P. 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 (2004).
7. Wahid, A. & Dubuisson, J. Virus-neutralizing antibodies to hepatitis C virus. *J. Viral Hepat.* **20,** 369–376 (2013).
8. Keck, Z. Y. *et al.* Human monoclonal antibodies to a novel cluster of conformational epitopes on HCV E2 with resistance to neutralization escape in a genotype 2a isolate. *PLoS Pathog.* **8,** e1002653 (2012).
9. Kong, L. *et al.* Structure of hepatitis C virus envelope glycoprotein E2 antigenic site 412 to 423 in complex with antibody AP33. *J. Virol.* **86,** 13085–13088 (2012).
10. Kong, L. *et al.* Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proc. Natl Acad. Sci. USA* **109,** 9499–9504 (2012).
11. Deng, L. *et al.* Structural evidence for a bifurcated mode of action in the antibody-mediated neutralization of hepatitis C virus. *Proc. Natl Acad. Sci. USA* **110,** 7418–7422 (2013).
12. Whidby, J. *et al.* Blocking hepatitis C virus infection with recombinant form of envelope protein 2 ectodomain. *J. Virol.* **83,** 11078–11089 (2009).
13. Krey, T. *et al.* The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLoS Pathog.* **6,** e1000762 (2010).
14. Keck, Z. Y. *et al.* Analysis of a highly flexible conformational immunogenic domain A in hepatitis C virus E2. *J. Virol.* **79,** 13199–13208 (2005).
15. Rothwangl, K. B., Manicassamy, B., Uprichard, S. L. & Rong, L. Dissecting the role of putative CD81 binding regions of E2 in mediating HCV entry: putative CD81 binding region 1 is not involved in CD81 binding. *Virol. J.* **5,** 46 (2008).
16. Lindenbach, B. D., Thiel, H.-J. & Rice, C. M. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 1101–1152 (Lippincott Williams & Wilkins, 2007).
17. White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* **43,** 189–219 (2008).
18. Vaney, M. C. & Rey, F. A. Class II enveloped viruses. *Cell. Microbiol.* **13,** 1451–1459 (2011).
19. El Omari, K., Iourin, O., Harlos, K., Grimes, J. M. & Stuart, D. I. Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. *Cell Rep.* **3,** 30–35 (2013).
20. Li, Y., Wang, J., Kanai, R. & Modis, Y. Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. *Proc. Natl Acad. Sci. USA* **110,** 6805–6810 (2013).
21. Forns, X. *et al.* 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. USA* **97,** 13318–13323 (2000).
22. Helle, F. *et al.* Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. *J. Virol.* **84,** 11905–11915 (2010).
23. Lavillette, D. *et al.* 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 (2007).
24. Li, H. F., Huang, C. H., Ai, L. S., Chuang, C. K. & Chen, S. S. Mutagenesis of the fusion peptide-like domain of hepatitis C virus E1 glycoprotein: involvement in cell fusion and virus entry. *J. Biomed. Sci.* **16,** 89 (2009).
25. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).
26. Kong, L. *et al.* Hepatitis C virus E2 envelope glycoprotein core structure. *Science* **342,** 1090–1094 (2013).
27. Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl Acad. Sci. USA* **99,** 13419–13424 (2002).
28. Nielsen, S. S., Moller, M. & Gillilan, R. E. High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell. *J. Appl. Crystallogr.* **45,** 213–223 (2012).
29. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).

Acknowledgements We acknowledge access to the X25 beamline at NSLS and thank the NSLS staff. NSLS is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. We thank J. Tainer and J. Perry for their support and access to SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory. We thank E. Arnold, J. Bonanno, S. K. Burley, J. Chiu, D. Comoletti, E. Elrod, F. Jiang, S. Khare, P. Lobel, A. Shatkin, A. Stock, J. Shires and A. Thanou for providing helpful comments and assistance. Special thanks to C. M. Rice for providing J6 HCV clone and support. This work was supported by a Yerkes Research Center Base Grant RR-00165 (A.G.) and NIH grants P50 GM103368 (J.M.), R01 AI080659 (J.M.) AI070101 (A.G.) and DK083356 (A.G.). A.G.K. was supported by a grant from the New Jersey Commission on Cancer Research (DFHS13CRP001).

Author Contributions The project was initiated, designed and supervised by A.G. and J.M. A.G.K., J.W., A.V.Z., A.C. and S.A.Y. designed protein constructs and established purification protocols. A.G.K. prepared all protein crystals. The mouse monoclonal antibody was produced by H.S. and A.G., and was sequenced by C.D.B. and J.J. A.A.P. and A.G. performed the virus neutralization and patient sera ELISA. J.M., M.T.M. and A.G.K. collected, processed and analysed the X-ray crystallographic and SAXS data. A.G.K., J.W. and J.M. wrote the paper with contributions from all authors.

Author Information The coordinates and structure factors have been deposited to the Protein Data Bank under accession code 4NX3. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. (jmarco@cabm.rutgers.edu) or A.G. (arash.grakoui@emory.edu).

METHODS

J6 E2 expression. eE2, eE2(ΔHVR1) and E2 core domain encompasses residues 384–656, 413–656 and 456–656 from the HCV J6 genome, respectively. Owing to incomplete deglycosylation at N7 (542) with EndoH, the crystallization construct contained an asparagine to glutamine mutation at this position. The expression constructs consisted of a CMV promoter, a prolactin signal sequence, E2 fragment, PreScission Protease cleavage site and a C-terminal protein-A (ProtA) tag. The entire prolactin-E2-ProtA sequence was PCR amplified and cloned into the pJG lentiviral vector (J. Shires).

Wild-type and GnTI− HEK293T cells[27] (provided by D. Comoletti) were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) at 37 °C with 5% CO_2. One day before the planned transfection, a single T-225 monolayer flask was seeded with 6.0×10^6 HEK293T cells. 90 μg pJG-E2, 60 μg psPAX2 (HIV Gag-Pol packaging vector), 30 μg pMD2.G (VSV glycoprotein vector) and 450 μl of 2 M $CaCl_2$ were mixed and brought to a final volume of 4.5 ml with ddH_20. 4.5 ml of 2× HEPES buffered saline was added at room temperature. After a 2-min incubation, the mixture was added directly to HEK293T cells. After 6–8 h, the media was replaced with DMEM with 10% FBS and 1% antibiotic/antimycotic (A/A) media and incubated for another 70 h.

Two days after transfection, 10,000 GnTI− HEK293T cells were seeded into a single well of a 96-well plate. The supernatant from the transfection, containing the recombinant lentiviruses, was collected and centrifuged for 30 min at 4,000*g* at 4 °C to pellet large cellular debris. Clarified supernatant was transferred to a Beckman Ultracentrifuge tube and virus was pelleted for 1.5 h at 25,000 r.p.m. (80,000*g*) at 4 °C in an SW28 rotor. Supernatant was discarded and the pellet re-suspended in 120 μl of DMEM containing 20% FBS, 1% A/A, and 8 μg ml^{-1} of polybrene. 60 μl of virus suspension was added to the prepared GnTI− HEK293T cells and incubated overnight. Infected cells were expanded and ultimately seeded into an adherent cell bioreactor (Cesco Bioengineering) for long-term growth and protein production.

eE2, eE2(ΔHVR1) and E2 core purification. Cell supernatant containing E2-ProtA was centrifuged for 10 min at 7,000*g*, filtered through a 0.22-μm membrane, and loaded onto an IgG FF column (GE Healthcare). The column was extensively washed with 20 mM sodium phosphate pH 7.0 and then equilibrated with 20 mM HEPES pH 7.5, 250 mM NaCl and 5% glycerol. PreScission Protease was added into the column at approximately 400 μg l^{-1} of supernatant and incubated overnight at 4 °C. For deglycosylation, the pH of the protein solution was adjusted using 1 M sodium citrate pH 5.5 to a final concentration of 100 mM. EndoH was added at a ratio of 1 mg per 2 mg of E2 and the reaction was incubated at room temperature for 3–4 h. The deglycosylated proteins were desalted into 20 mM HEPES pH 7.5, 50 mM NaCl, and 5% glycerol and purified by heparin affinity followed by size-exclusion chromatography over a Superdex200 column. Final yields for all E2 proteins averaged 30 mg l^{-1} of supernatant.

Crystallization. A 1.1:1 molar ratio of E2 core to Fab was incubated for 1–2 h at 4 °C. The complex was purified over a Superdex200 column equilibrated with 20 mM HEPES pH 7.5 and 100 mM NaCl. The complex was concentrated to 5–7 mg ml^{-1} and crystals were grown by the hanging-drop vapour diffusion method. Briefly, 2.5 μl of complex was mixed with an equal volume of reservoir solution, comprising of 22% (w/v) PEG 3350, 0.5 M $MgCl_2$, 0.1 M HEPES pH 7.5, and 15% (v/v) dioxane. Initially, clusters of plate-like crystals grew in 3–4 days. Single, plate-like crystals were obtained via microseeding using a similar reservoir solution supplemented with 2% (v/v) formamide. Crystals were cryoprotected using reservoir solution with 24% (v/v) ethylene glycol and flash cooled in liquid nitrogen. Data were collected at a wavelength of 0.979 Å using beamline X25 of the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory.

Structure determination and refinement. The crystals belong to space group $P2_12_12$ with cell parameters $a = 85.96$ Å, $b = 194.57$ Å, $c = 37.92$ Å. Phases were determined by the molecular replacement method using PHENIX[30] and the coordinates from chains A and B from PDB entry 2GSI. Unambiguous placement of the Fab heavy and light chains provided the necessary phases to extend the map to cover E2 core domain using iterative rounds of model building and density modification by COOT[31], PHENIX, REFMAC[32] and PARROT[33]. The final model was built to a resolution of 2.4 Å, comprising residues 492–522, 538–571 and 596–649 of E2 from the J6 genome, 1–217 of 2A12 light chain, and 1–133 and 136–218 of 2A12 heavy chain with two *N*-linked, *N*-acetylglucosamine, six molecules of formamide and 141 solvents molecules. The model coordinates were refined to R_{work} 0.217 and R_{free} 0.269. Model validation demonstrated 95.0% of the residues located in the most favourable region of the Ramachandran plot with 4.8% in the generously allowed regions[30]. Statistics of the data processing and structure refinement are summarized in Extended Data Table 1.

Small angle X-ray scattering (SAXS). Glycosylated E2 proteins were purified over IgG and anion exchange columns. The proteins were equilibrated with either pH 7.5 buffer (50 mM HEPES pH 7.5, 250 mM NaCl and 1% glycerol) or pH 5.0 buffer (50 mM sodium citrate pH 5.0, 250 mM NaCl and 1% glycerol) by Superdex200 gel filtration column. Glycosylated eE2(ΔHVR1) alone or complex with CD81-LEL (1:2 molar ratios) was purified using pH 7.5 buffer by gel filtration chromatography. Three concentrations of each protein were prepared along with their respective buffers as background control. SAXS data was collected on the SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory. Sample analysis and processing was performed using BioSAXS RAW[28], ATSAS[29], and GNOM[34]. The *ab initio* models were calculated using the application DAMMIF[35]. Consensus models and the normalized spatial discrepancy (NSD) values were calculated by averaging 10 *ab initio* models using the application DAMAVER[36]. X-ray structures of CD81 (PDB ID 1G8Q), HCV E2 core and *ab initio* models were aligned using the application SUPCOMB[37].

Hydrogen deuterium exchange. HD exchange experiments were conducted as described previously[38]. Briefly, 5 μl of deglycosylated eE2 (1.5 mg ml^{-1}), in 200 mM NaCl, 20 mM HEPES pH 7.5, was incubated with 15 μl of the same buffer made with 99.96% ^2H_2O (Cambridge Isotope Laboratories) for 10, 100, or 1,000 s and quenched in 30 μl of 2 M urea, 0.8% formic acid and 50 mM tris(2-carboxyethyl) phosphine (TCEP). The reaction mixture was immediately frozen on dry ice until injection. For the zero time point experiment, the protein was incubated in the buffer made with ^1H_2O and then quenched and frozen. To correct background exchange, a completely deuterated sample was produced by incubating the protein with 100 mM TCEP in 99.96% ^2H_2O overnight before being quenched and frozen. Dionex RSLC with a C18 column (2.1 × 50 mm, 3 μm, Q-C18, 150A, CMP Scientific) and LTQ Velos Orbitrap pro were used for LC–MS analysis. The mass was measured using Orbitrap with resolution of 60,000 and mass range from 300 to 2,000 *m*/*Z*. The LC–MS data were analysed using HDexaminer 1.2.0 (Serra Analytics) with manual checking of each peptide afterwards.

Limited proteolysis. 8–10 μg of deglycosylated eE2 protein was mixed with trypsin, chymotrypsin or GluC at 1:120 (w/w) ratio (endopeptidase:E2) and incubated at room temperature. Samples were taken at noted time points and analysed by reducing SDS–PAGE, mass spectrometry and N-terminal sequencing.

Production of monoclonal antibody 2A12. Six-to-eight-week-old female BALB/c mice were immunized intraperitoneally with 50 μg eE2 in either complete Freund's adjuvant (first immunization only), or incomplete Freund's adjuvant bi-weekly for 8 weeks. A final immunization with 50 μg of eE2 was given intravenously 4 days before collection of splenocytes. Hybridomas were generated using a cloned HAT-sensitive mouse myeloma cell line as a fusion partner. Proliferating hybridomas were screened for their ability to bind eE2 via ELISA, at which point 2A12 was positively identified. Monoclonal antibodies were generated in the laboratory of A. Grakoui (IACUC protocol number YER-2002369-070816GN).

Generation and purification of 2A12 Fab. Hybridoma cells were expanded to a final volume of 2 l in spinner flasks at 100 r.p.m. using Iscove's Modified Dulbecco's Medium, 10% ultra-low IgG FBS, 1% A/A, and 10 mM HEPES (Life Technologies). Cells were collected at $2–3 \times 10^6$ cells per ml, centrifuged for 10 min at 7,000*g*, filtered through a 0.22-μm membrane, and loaded onto a Protein G column (GE Healthcare Life Sciences). After completion, the column was washed with 20 mM sodium phosphate (pH 7.0) followed by phosphate buffered saline (PBS). The antibody was eluted with 0.05% TFA in 2 ml fractions into tubes containing 100 μl of 1 M Tris pH 7.5 for immediate pH neutralization. The eluted antibody was dialysed into 20 mM sodium phosphate pH 7.0 and 10 mM EDTA. Insoluble papain was added to the antibody at 0.15 mg per 1 mg of antibody. Freshly prepared L-cysteine was added to the reaction to a final concentration of 20 mM and mixed at 37 °C for 2 h. The papain was removed by centrifugation at 3,500*g* for 2 min and filtration through a 0.22-μm membrane. Fab was purified by subtractive chromatography over Protein A FF column and desalted into 20 mM Tris pH 8.0.

Sequencing Ig H and L chain gene segments of 2A12 antibody. Total RNA isolated from 2A12 hybridoma cells was reverse transcribed into cDNA using random hexamers. Expressed heavy (H) and light (L) chains were amplified using standard primers that are complimentary to all murine H and L chain gene segments[39]. The PCR products were sequenced either directly or following cloning into pCR 2.1-TOPO vector (Life Technologies).

CD81 purification and binding assays. Human CD81-LEL (residues 112–202) was produced as a fusion with C-terminal ProtA tag in HEK293T cells using the same lentiviral expression system described for eE2. Cell culture supernatants were loaded onto an IgG FF column, washed with 20 mM sodium phosphate pH 7.0, eluted with 100 mM sodium citrate pH 3.0 containing 20 mM KCl and immediately neutralized with 1 M Tris pH 9.0. The ProtA tag was cleaved by PreScission Protease in a ratio of 1:50 (w/w) followed by overnight dialysis in 20 mM HEPES pH 7.5, 250 mM NaCl, and 5% glycerol. High-purity CD81 protein was obtained by anion exchange and size-exclusion chromatography.

For binding studies, a 96-well plate (Nalgene Nunc, Thermo Fisher Scientific) was coated with 50 μg of CD81-LEL overnight at 4 °C. All experiments were

duplicated against BSA as a negative control. Plates were washed three times with PBS containing 0.05% Tween 20 (PBS-T) and blocked with 3% (w/v) BSA in PBS-T for 1 h at room temperature. 50 µl of eE2 or E2 core at different concentrations was added to appropriate wells and incubated overnight at 4 °C. On day 3, the wells were washed three times with PBS-T and incubated with monoclonal antibody 2A12 cell supernatant for 1 h at room temperature. Plates were washed three times with PBS-T and incubated with anti-mouse-HRP conjugated antibody for 1 h at room temperature. Finally, the plate was washed five times with PBS-T. 50 µl of TMB substrate (ThermoFisher Scientific) was added to each well and incubated for 5 min, followed by the addition of 50 µl of 2 M sulphuric acid to stop the reaction. Absorbance readings were acquired at 450 nm using Softmax Pro software on a Spectra Max 250 (Molecular Devices).

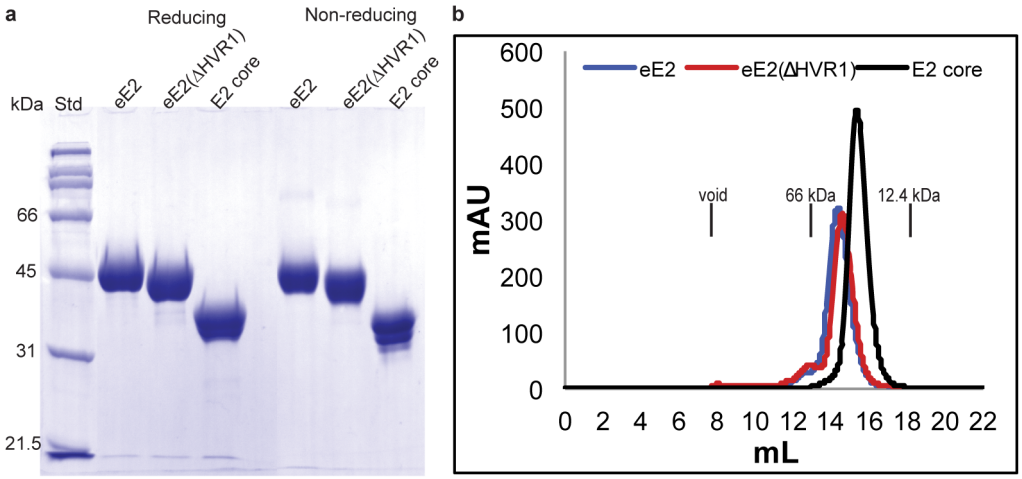
Neutralization assay. Huh-7.5 cells were maintained in DMEM containing 10% FBS (Hyclone) and 100 µg ml^{-1} of penicillin and streptomycin (Cellgro) at 37 °C in 5% CO_2. Naive Huh-7.5 cells were seeded at 6,000 cells per well in a 96-well plate. The following day, 100 µl of 2C1, 2A12, or H113 serially diluted in complete media were added per well at various concentrations. In parallel, 100 µl of eE2, E2 core, gp140, or CD81-LEL serially diluted in complete DMEM were added at varying concentrations beginning at 100 µg ml^{-1}. Cells were then infected with 100 µl of genotype 2a virus Cp7 encoding the *Renilla* luciferase gene[40]. 72 h after infection, relative light units were measured on a Clarity 4.0 luminometer (Biotek) using the *Renilla* Luciferase Assay System (Promega).

Assessment of cellular cytotoxicity. Huh-7.5 cells were incubated with varying concentrations of protein as described above, beginning at 100 µg ml^{-1}. After 72 h, cells were washed once with PBS, treated with trypsin, and collected in 100 µl PBS. Cells were stained with 7-AAD according to the manufacturer's instructions (BD Biosciences) and analysed using a BD LSR II and FlowJo software (Tree Star).

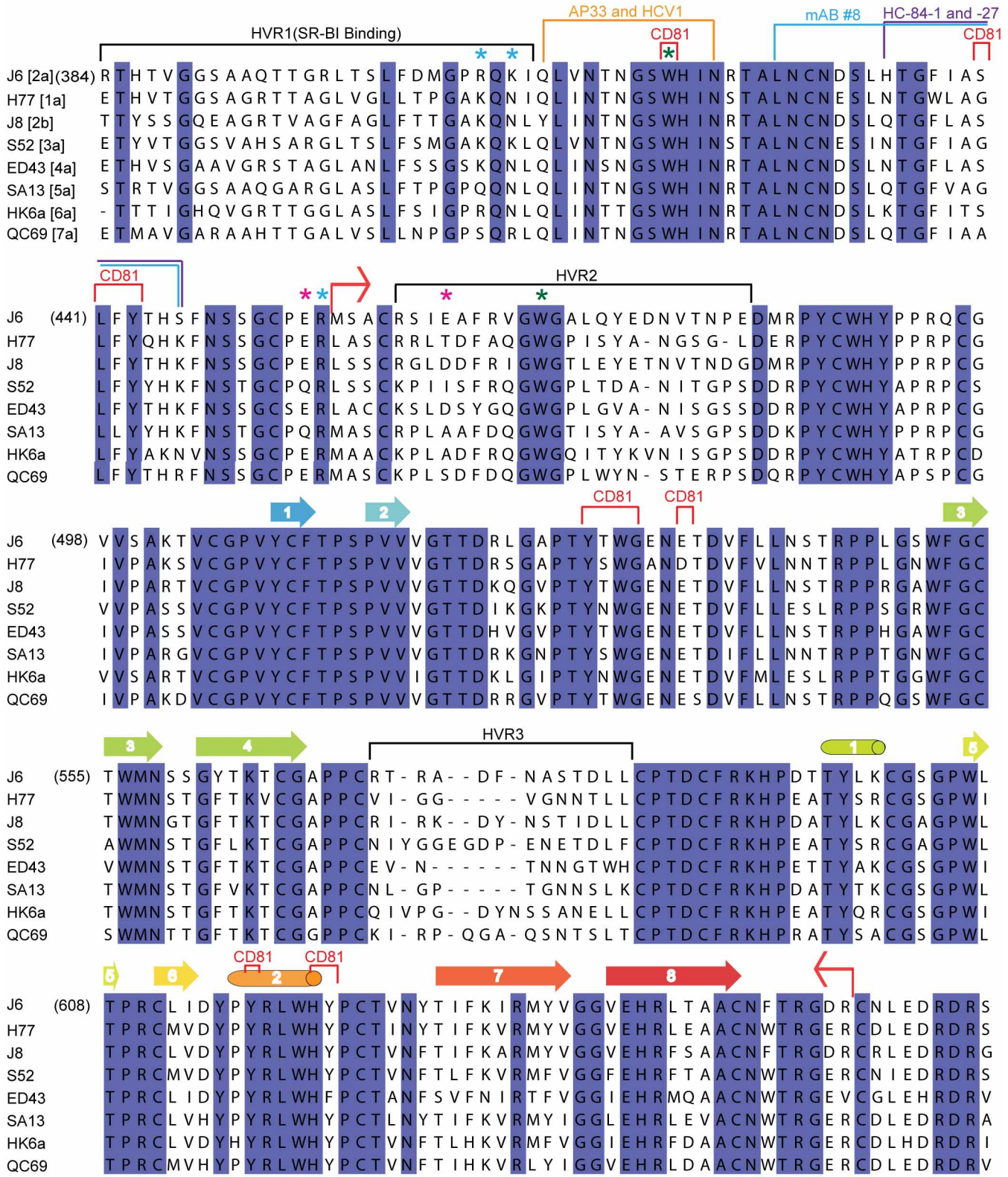
Human plasma ELISA. 96-well enzyme immunoassay plates (ThermoFisher Scientific) were coated overnight at 4 °C with 50 µl of a 1 µg ml^{-1} solution of eE2 or E2 core diluted in 0.1 M Na_2CO_3. Plates were washed twice with PBS-T and then blocked for 1 h at 37 °C in PBS-T containing 10% fetal calf serum (HyClone). Blood samples were collected in heparin tubes (Becton Dickinson) and plasma was isolated and frozen at −80 °C. Plasma was serially diluted in a binding solution composed of 0.1% (v/v) normal goat serum in PBS-T (Jackson ImmunoResearch Laboratories). 100 µl of sample were added per well and incubated at room temperature for 90 min. After eight washes with PBS, 100 µl of mouse anti-human IgG biotin antibody (Mabtech) diluted 1:20,000 in binding solution were added per well and incubated 1 h at room temperature. Following five additional washes with PBS, 100 µl streptavidin-horseradish peroxidase (HRP) conjugate was added to each well at a 1:2,000 dilution in binding buffer and incubated for 45 min at room temperature (Mabtech). Absorbance was measured and analysed using a VersaMax microplate reader and SoftMax Pro software (Molecular Devices) following five washes and the addition of tetramethylbenzidine substrate solution (Ebioscience). Human sera were isolated from whole-blood samples, and informed consent was obtained for all subjects (IRB no. 1358-2004, Emory University School of Medicine, principal investigator A. Grakoui).

Alignment. Secondary structures were assigned using the program DSSP[41]. Sequences were obtained from the National Center for Biotechnology Information (NCBI) using the following accession numbers: J6 ADV40003.1, H77 ACA53555.1, J8 P26661.3, S52 AEB71616.2, ED43 AEB71617.2, SA13 AEB71618.2, HK6a AEB71619.2, QC69 ACM69041.1. The E2 sequences were aligned with multiple alignment using fast Fourier transform (MAFFT)[42] and edited for figure generation using JalView version 2 (ref. 43).

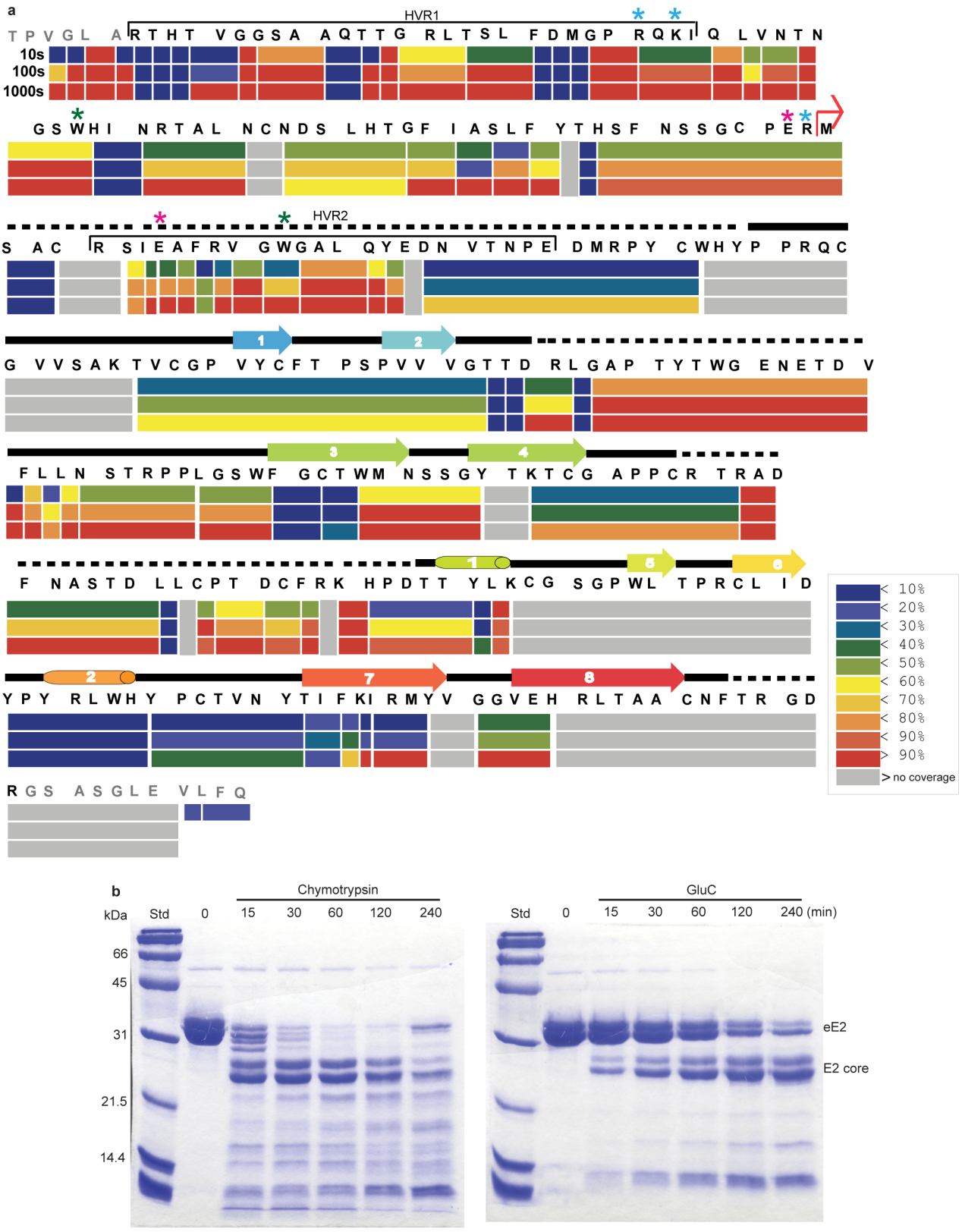
30. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
32. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
33. Zhang, K. Y., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods Enzymol.* **277,** 53–64 (1997).
34. Semenyuk, A. V. & Svergun, D. I. GNOM - a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24,** 537–540 (1991).
35. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 (1999).
36. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36,** 860–864 (2003).
37. Kozin, M. B. & Svergun, D. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34,** 33–41 (2001).
38. Sharma, S. *et al.* Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry. *Proteins* **76,** 882–894 (2009).
39. Tiller, T., Busse, C. E. & Wardemann, H. Cloning and expression of murine Ig genes from single B cells. *J. Immunol. Methods* **350,** 183–193 (2009).
40. Mateu, G., Donis, R. O., Wakita, T., Bukh, J. & Grakoui, A. Intragenotypic JFH1 based recombinant hepatitis C virus produces high levels of infectious particles but causes increased cell death. *Virology* **376,** 397–407 (2008).
41. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).
42. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
43. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).



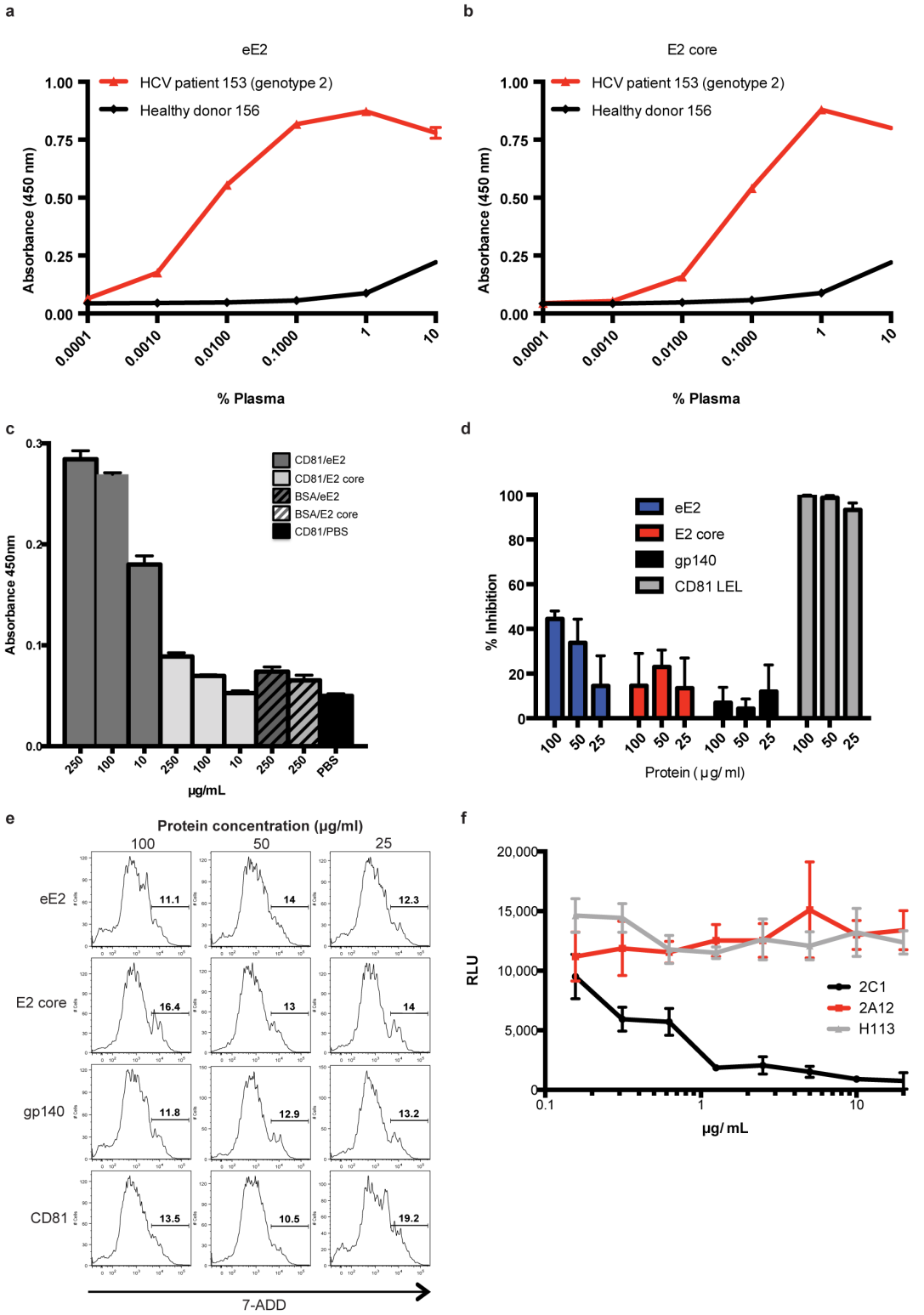
Extended Data Figure 1 | eE2, eE2(ΔHVR1) and E2 core are highly soluble and monomeric in solution. **a**, A comparison of proteins under reducing and non-reducing conditions is shown by a 10% SDS–PAGE gel with protein standards (Std). **b**, Size-exclusion chromatography of eE2, eE2(ΔHVR1) and E2 core proteins on a Superdex200 gel filtration column. The elution positions of the void volume (>200 kDa), albumin (66 kDa) and cytochrome C (12.4 kDa) are indicated. Molecular masses of eE2, eE2(ΔHVR1) and E2 core are ~46 kDa, ~42 kDa and ~32 kDa, respectively.



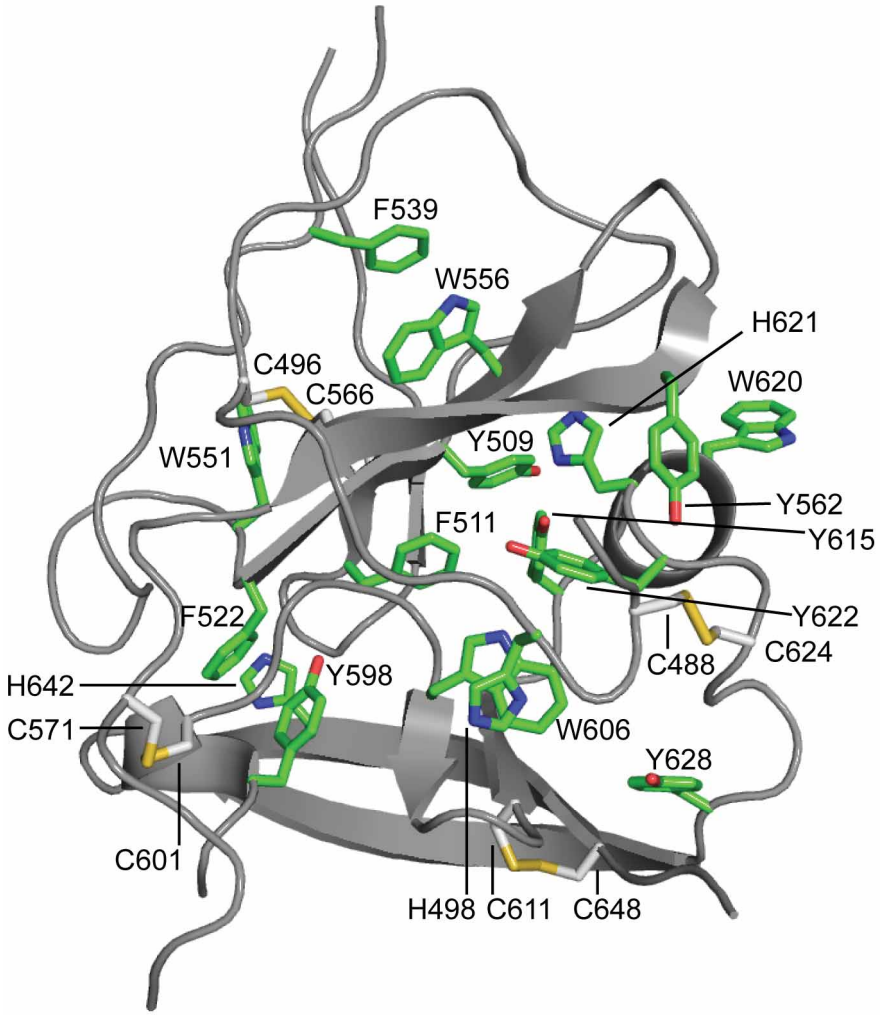
Extended Data Figure 2 | eE2 sequence alignment. Red bent arrows indicate the N- and C-terminal boundaries of the E2 crystallization construct. Cylinders and arrows represent α-helices and β-strands, respectively, and are coloured according to cartoon representation in Fig. 1. CD81 binding regions are bracketed in red; hypervariable regions are bracketed in black. SR-BI binds to HVR1. The asterisks indicate the location of trypsin (blue), chymotrypsin (green) and GluC (magenta) cleavage sites. The binding sites of neutralizing antibodies for which structural information is available are coloured orange for HCV1 and AP33, blue for mAb 8, and purple for HC34-1 and HC34-17.



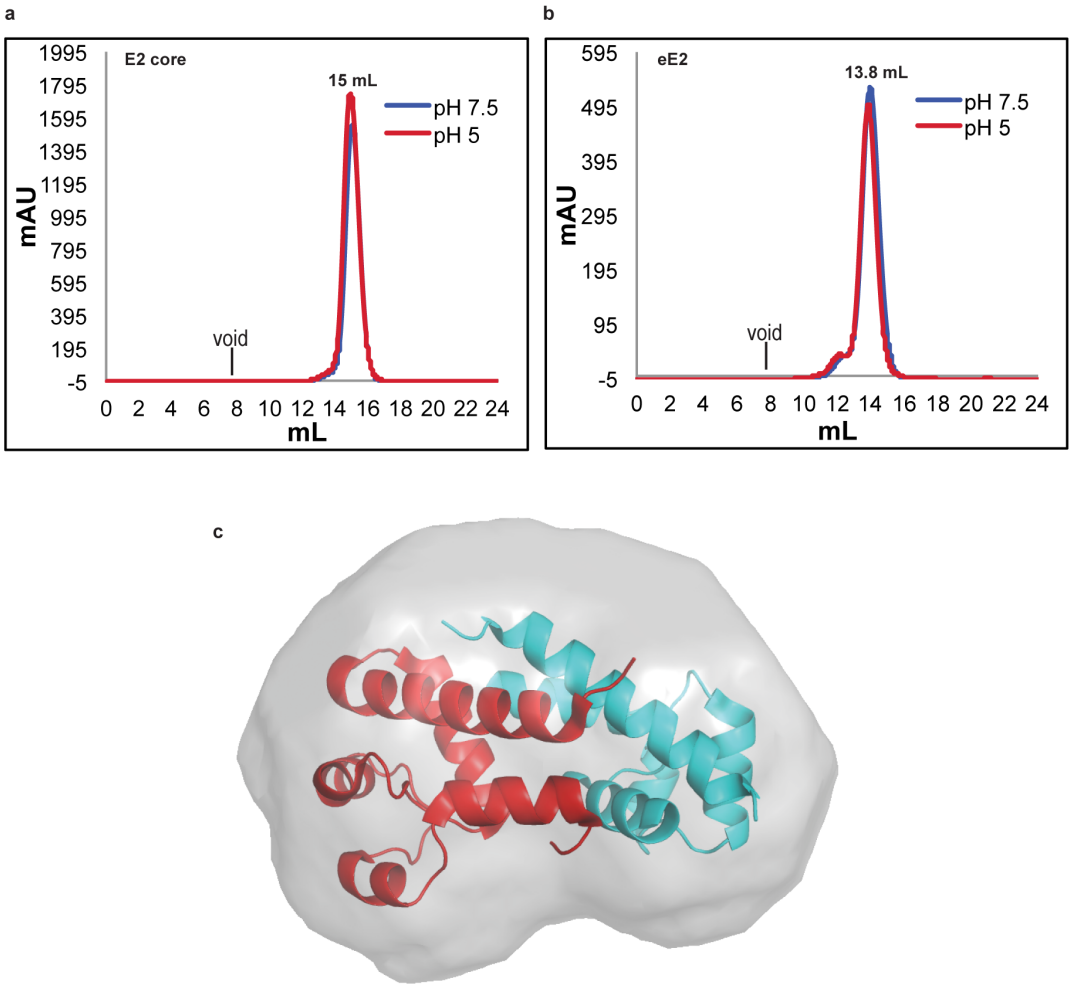
Extended Data Figure 3 | Hydrogen deuterium exchange and limited proteolysis of eE2. a, The percentage hydrogen deuterium exchange shown at 10, 100 and 1,000 s time points. The secondary structure of E2 core is placed above to emphasize flexible regions. A red arrow indicates the E2 core N terminus. Extra residues (grey) on N and C terminus come from the vector. Potential cleavage sites for trypsin (blue), chymotrypsin (green) and GluC (magenta) are indicated by asterisks. The colour pattern indicates the percentage of exchange. Grey areas are the regions of no coverage. **b**, Digestion of deglycosylated eE2 with chymotrypsin (left) and GluC (right) reveals a shift from the ~35-kDa untreated protein (0 min) to ~25 kDa after digestion. Samples were taken at the indicated time points and analysed by reducing 12% SDS–PAGE gel. Molecular mass protein standards (Std) are indicated. The bands were analysed by N-terminal sequencing and mass spectrometry.



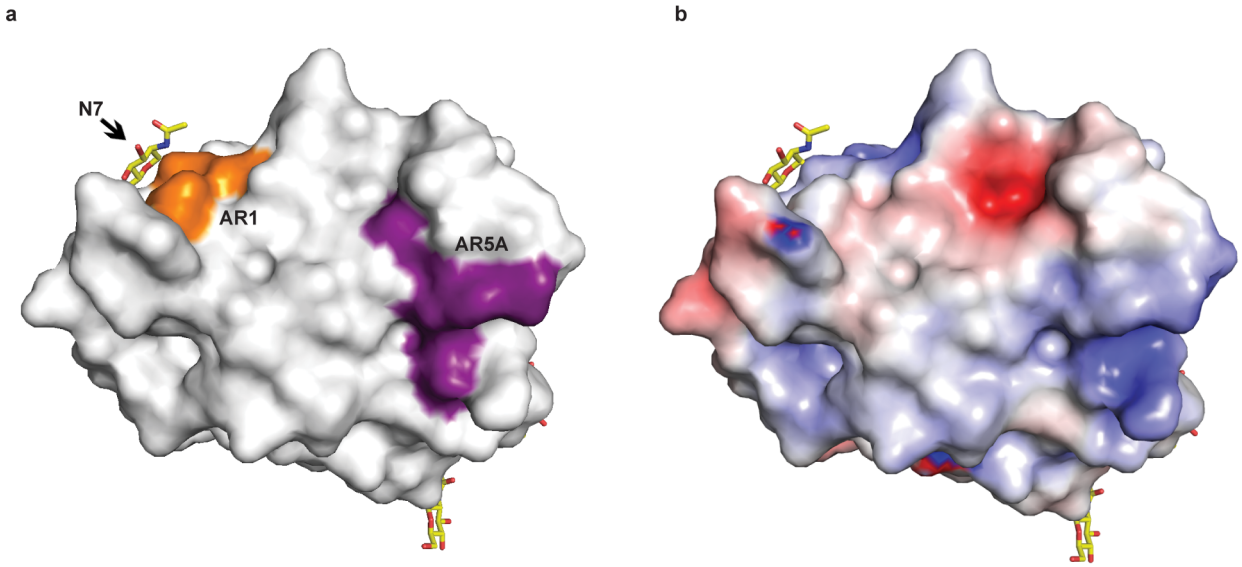
Extended Data Figure 4 | Functional analyses of eE2 and E2 core. **a**, Antibodies from patient sera infected with HCV genotype 2 show a concentration-dependent binding to eE2 (red) whereas healthy donor sera exhibit only background binding (black). **b**, Similar binding is observed for E2 core. The measurements were done in triplicate with the error bars representing the standard error of the mean (s.e.m.). **c**, E2 core (light grey) shows reduced binding to CD81 when compared to eE2 (dark grey) by an ELISA. Bars with stripes indicate E2 binding to a negative control, BSA. The solid black bar indicates CD81 binding to PBS, used to verify the absence of background. The measurements were done in triplicate with the error bars representing the s.e.m. **d**, eE2 (blue) and CD81-LEL (positive control, grey) inhibit the infection. E2 core (red) shows reduced inhibition. HIV gp140 (black) expressed in the same system was used as a negative control. The measurements were done in triplicate with the error bars representing the s.e.m. **e**, To rule out the possibility of toxic effects from the recombinant proteins, the cell viability was measured as described in Methods, using similar protein concentrations as in **d**. **f**, In an ELISA, 2A12 (red), and an irrelevant antibody, H113 (grey), fail to neutralize HCVcc infection. 2C1 (positive control, black), a mouse monoclonal antibody that binds to the disordered N-terminal region of eE2, blocks infection. The measurements were done in triplicate with the error bars representing the s.e.m.



Extended Data Figure 5 | E2 core contains an extensive hydrophobic core. Sheets A and B are held together by an extensive hydrophobic core composed of mostly aromatic amino acids (green) and five disulphide bonds (yellow).



Extended Data Figure 6 | eE2 and E2 core do not undergo oligomeric changes at low pH. **a**, **b**, An overlay of E2 core (**a**) and eE2 (**b**) elution profiles from Superdex200 gel filtration at pH 7.5 (blue) and pH 5.0 (red). The expected void volume and observed elution positions of individual proteins are indicated. **c**, The SAXS envelope of CD81-LEL fit with a dimer crystal structure (PDB 1G8Q). The individual proteins of the CD81-LEL dimer are coloured red and blue.



Extended Data Figure 7 | Epitope mapping of conformational antibodies on E2 core surface. **a**, Surface epitopes of AR1 (orange) are shown. AR1 blocks the E1E2 heterodimer binding to CD81. AR5A (purple) inhibits E1E2 heterodimerization and is mapped on a well-conserved hydrophobic surface of the core. **b**, Surface of E2 core coloured by electrostatic potential. The view in **a** and **b** is identical.

Extended Data Table 1 | Summary of the X-ray crystallographic analyses

Data collection	
Wavelength	0.979 Å
Space group	$P2_12_12$
Cell dimensions	
a, *b*, *c* (Å)	85.96, 194.57, 37.92
α, β, γ (°)	90, 90, 90
Resolution (Å)	24.69-2.40 (2.49-2.40)*
R_{merge}	0.118 (1.042)
R_{pim}	0.054 (0.490)
I/σ	9.6 (1.9)
Completeness (%)	100 (100)
Redundancy	5.7 (5.3)
Refinement	
Resolution (Å)	24.69-2.40
No. unique reflections	24349 (2517)
R_{work}/R_{free}	0.217/0.269
No. atoms	
Protein/glycans	4162
Solvent	137
Ions	18
B-factors (Å^2)	
Protein/glycan	54
Solvent	45
Ions	47
R.M.S deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.14
Ramachandran favored (%)	95.0
Ramachandran outliers (%)	0.2

* Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | Summary of SAXS analyses

Protein	R_g (Å)	D_{max} (Å)	NSD
E2 core	27.8±0.03	95	0.705±0.026
eE2(ΔHVR1)	28.9±0.05	101	0.696±0.075
eE2	28.2±0.02	84	0.702±0.054
eE2(ΔHVR1) + CD81	36.8±0.22	127	0.714±0.026
CD81	18.8±0.04	64	0.554±0.007
E2 core pH 5.0	27.8±0.07	95	-
eE2 pH 5.0	27.8±0.07	95	-