A Model for the Hepatitis C Virus Envelope Glycoprotein E2

Asutosh T. Yagnik, Armin Lahm, Annalisa Meola, Rosa Maria Roccasecca, Bruno B. Ercole, Alfredo Nicosia, and Anna Tramontano*

Istituto di Ricerche di Biologia Molecolare P. Angeletti, Pomezia (Rome), Italy

ABSTRACT **Several experimental studies on hepatitis C virus (HCV) have suggested the envelope glycoprotein E2 as a key antigen for an effective vaccine against the virus. Knowledge of its structure, therefore, would present a significant step forward in the fight against this disease. This paper reports the application of fold recognition methods in order to produce a model of the HCV E2 protein. Such investigation highlighted the envelope protein E of Tick Borne Encephalitis virus as a possible template for building a model of HCV E2. Mapping of experimental data onto the model allowed the prediction of a composite interaction site between E2 and its proposed cellular receptor CD81, as well as a heparin binding domain. In addition, experimental evidence is provided to show that CD81 recognition by E2 is isolate or strain specific and possibly mediated by the second hypervariable region (HVR2) of E2. Finally, the studies have also allowed a rough model for the quaternary structure of the envelope glycoproteins E1 and E2 complex to be proposed. Proteins 2000;40:355–366.** © 2000 Wiley-Liss, Inc.

Key words: hepatitis C virus; protein structure prediction; fold recognition; E1/E2 association; CD81 binding; heparin binding

INTRODUCTION

Hepatitis C virus (HCV)[1] is the major aetiological agent of both community and post-transfusionaly acquired non-A, non-B viral hepatitis.[1–3] Approximately 90% of patients develop chronic hepatitis,[3,4] of which 20–30% progress onto liver cirrhosis.[3,5] All cases of infection carry an increased risk of hepatocellular carcinoma which may be further exacerbated by co-infection with hepatitis B and/or high alcohol consumption.[6] Presently the only available therapies are interferon-α (IFN) on its own[7,8] or in combination with ribavirin.[9] Such treatments are expensive, show low-response rates, and carry the risk of significant side effects. Consequently, the development of a vaccine against hepatitis C remains a high priority goal.

The HCV genome is composed of a single-stranded positive sense RNA of approximately 9.5 kB, encoding for a single polyprotein of between 3010 and 3033 amino acids.[10,11] Owing to the similarity with the Flaviviruses and Pestiviruses,[12,13] in terms of genomic organization, HCV has been classified as a separate genus in the *Flaviviridae* family.[14] A combination of host and viral peptidases are involved in polyprotein processing to give at least nine different proteins.[15–19] The predicted structural components of the virus comprise the core (C) (~21 kDa) and two heavily N-glycosylated envelope glycoproteins,[20] E1 (~31 kDa) and E2 (~70 kDa),[21] starting at polyprotein positions 191 and 384, respectively.[15] Both are believed to be type 1 transmembrane proteins, with N-terminal ectodomains and C-terminal hydrophobic anchors,[22] and together are expected to form the viral envelope. Several studies have shown E1 and E2 form complexes, although the exact nature of this interaction is still poorly understood.[23–25]

HCV envelope glycoproteins are deemed important since chimpanzees immunized with purified recombinant E1/E2 heterodimeric proteins have been shown to educe limited protection against challenge with homologous virus.[26] It has also been reported that the presence of neutralizing antibody against HCV E2 correlates with protection from HCV infection,[27] suggesting that it is a fundamental candidate antigen for a vaccine against hepatitis C virus.

The N-terminal 27 residues of E2 (aa 384–410) show a very high degree of variation, both within isolates and genotypes, and this portion of the sequence has been termed hypervariable region 1 (HVR1).[13,28] Several experimental data exist implicating HVR1, at least in part, for escape of E2 from the host immune system by way of creating mutants,[7,29,30] and efforts have been made to generate cross-reactive mimotope libraries of HVR1 for use in vaccine design.[31]

Despite the recent work of Lohmann et al.,[32] who succeeded in designing a self-replicating viral RNA, the bottleneck in viral infection and inhibition studies is the lack of an efficient cell culture infection system. Therefore, knowledge of the three-dimensional structure of HCV E2 would be of great value in the quest for a vaccine, in explaining existing data and in designing novel experiments. However, this is quite a challenging problem, both from an experimental and theoretical point of view. For example, current understanding of HCV envelope proteins is based on mammalian cell culture transient expression assays with viral and non-viral vectors.[33] These systems produce very low levels of heterogeneous protein due to glycosylation and aggregation, and it is difficult to distinguish between molecules that undergo productive and non-productive folding.[34,35] Moreover, traditional techniques such as secondary structure prediction, fold recognition and homology modeling, albeit often successful in cases where sufficient information about the protein fam-

*Correspondence to: Anna Tramontano, IRBM P. Angeletti, Via Pontina Km 30.600, 00040 Pomezia (Roma), Italy. E-mail: Tramontano@irbm.it

Received 14 December 1999; Accepted 23 March 2000

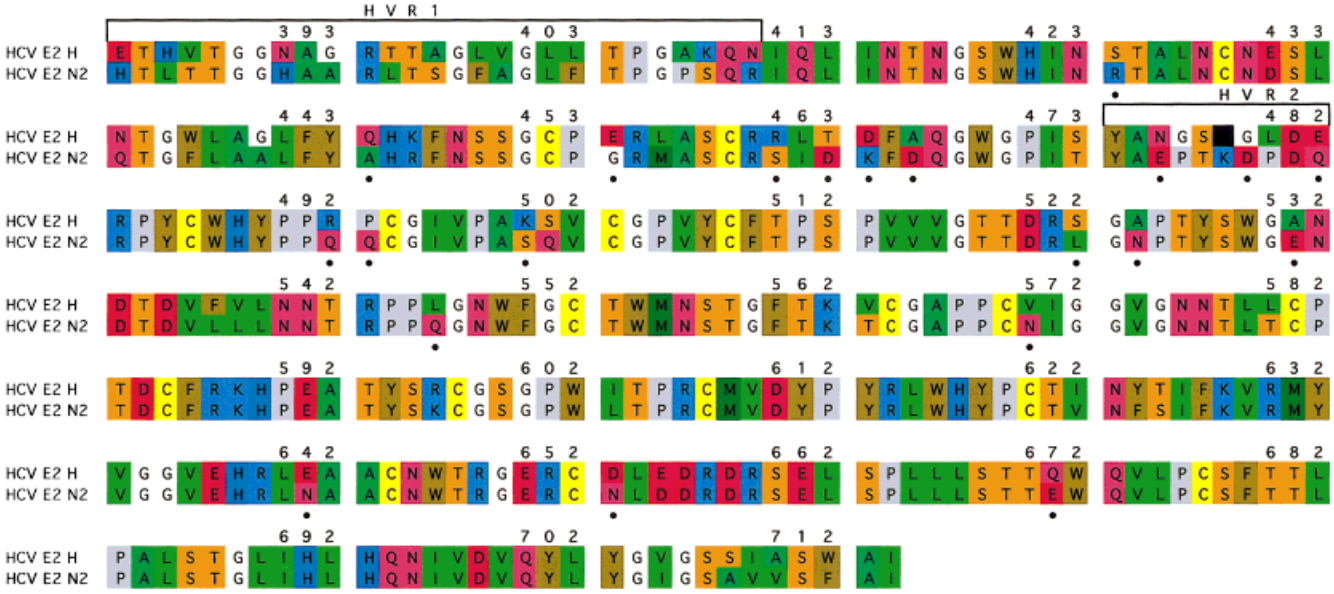

Fig. 1. A sequence alignment between the HCV strain H E2 sequence used to create the model (GenBank Accession Number AF011751) and the N2 strain E2 sequence (GenBank Accession Number D13406). The hypervariable regions (HVR1 and HVR2) are indicated above the alignment and positions marked below with a dot represent those amino acids found to be significantly different between the two sequences.

ily is known,[36] are hampered by the high degree of similarity between different E2 sequences and no detectable sequence identity with any protein of known structure. Recently, however, some additional information became available when three new GB viruses were added to the list of *Flaviviridae*.[37] GB viruses A (GBV-A) and B (GBV-B) were isolated from tamarins[38] and GBV-C from a human specimen.[39] A genomic sequence analysis has shown GBV-A and GBV-C to be closely related, with a more distant relationship to HCV.[40] GBV-B is instead the sole member of a separate subgroup that bears the same resemblance both to GBV-A/C and to HCV.

A significant amount of experimental data on the HCV E2 protein have recently accumulated and it was felt that these, together with the progress in fold recognition techniques, could be exploited in order to build a set of reasonable models for HCV E2. Since the E2 sequences of the related GB viruses can be expected to be functionally equivalent, and therefore to preserve some structural similarity, a range of various fold recognition methods was applied to each of them, as well as the HCV E2 sequence. This highlighted common folds found by the different fold recognition techniques on each of the diverse sequences. These results were used to initially build two alternative models of HCV E2, from which the one based on the envelope protein E of Tick Borne Encephalitis (TBEV) was selected for its ability to explain existing experimental data.

It has been reported that HCV E2 binds human CD81,[41] a tetraspanin widely expressed on most tissues, that E1 and E2 form a complex of as yet unknown stoichiometry,[23–25] and that HCV might contain a heparin binding domain.[42,43] Mapping and interpreting the relevant experimental data in the three-dimensional context of the model allowed several new conclusions to be drawn: a composite CD81 binding site could be proposed and the quaternary structure of the E1/E2 complex was predicted, based on the dimeric nature of the TBEV template. Furthermore, it was shown that E2 specifically binds heparin and a potential binding site for this interaction is proposed in the three-dimensional model.

To complement existing data, binding of E2 from two different HCV isolates to CD81 was measured and the results correlated with sequence variations, some of which overlapped with the hypervariable region 2 (HVR2) of E2. This latter result is particularly intriguing since it implies that strain or even isolate specific characteristics could play a role in both HCV infectivity and tropism, features of central importance for the development of HCV vaccines.

MATERIALS AND METHODS

Production and Normalization of E2 Proteins

H and N2 strain E2 proteins (see Fig. 1 for aa sequences) were constructed by PCR as truncation forms at aa 683

Fig. 2. The sequence alignment between HCV E2 and 1SVB used to construct the model, showing the regions of secondary structure (SS) as predicted by PHD for HCV E2 and calculated by DSSP for 1SVB.[89] Sequence numbering in HCV E2 is according to the polyprotein position. SS types H/h (beige color) denote an alpha helix and E/e (violet color) identify a beta strand, with the lowercase letters indicating a PHD prediction reliability less than 5. The domain boundaries in 1SVB are clearly marked with Roman numerals and arrows, as is the cd loop (aa 98–113).[65]

Fig. 3. Schematic ribbon representation of the HCV E2 model based on 1SVB (TBEV E protein) indicating regions involved in CD81 binding (orange and magenta), the HVR1 (green), and aa 612–620 (red). Within these regions, spheres highlight the locations of the predicted site of E1/E2 interaction (cyan), the HVR2 region (blue), and the site of interaction with mAb 6/41a (orange) and mAb 6/53 (magenta).

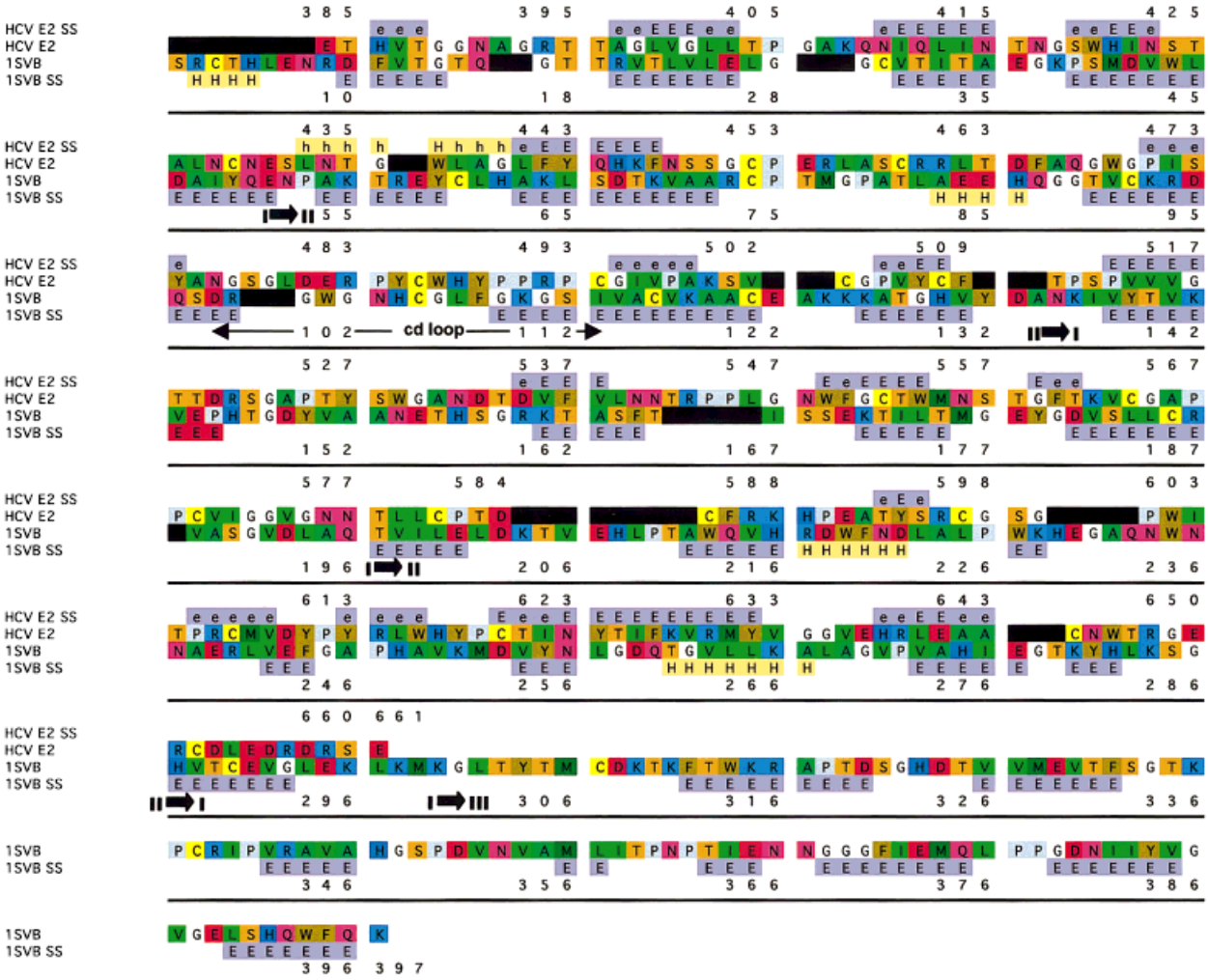

Figure 2.

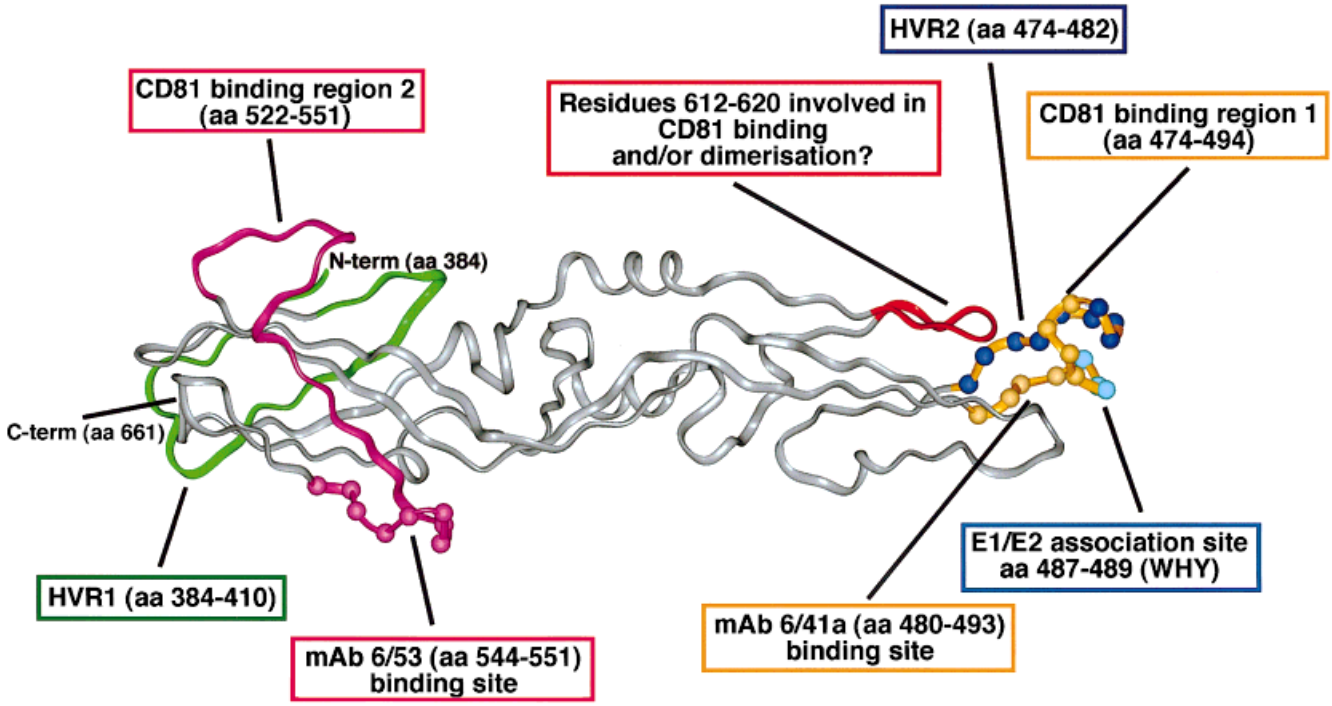

Figure 3.

and 684, respectively, with an additional tag of six histidine residues at the C-terminus. E2 H corresponds to the genotype 1a isolate H77[44] (GenBank Account Number AF011753), and E2 N2 to the genotype 1b isolate N2[45] (GenBank Account Number D13406); 293 cells (Human Embrionic Kidney, ATCC) were transfected with the E2 plasmids and the protein was harvested after 48 hr as crude cell extract in lysis buffer (1% tritonX100/20 mM Tris HCl, pH 7.5/150 mM NaCl/1 mM EDTA supplemented by 1 tablet/50 ml of protease inhibitor cocktail tablets; Boehringer Mannheim, 1-697-498).

The H strain of E2, truncated at aa 661, was employed for the heparin binding studies. E2 proteins used in heparin binding assays were produced as crude cell extracts similar to those used in CD81 binding studies, but replacing in the lysis buffer 20 mM Tris HCl pH 7.5 with 50 mM sodium phosphate buffer pH 7.5.

The relative amount of E2 present in different preparations was determined as follows. ELISA plates were coated with GNA (Lectin from *Galanthus Nivalis*, Sigma L 8275) diluted to 1 μg/well in PBS. After an o/n incubation at 4°C, plates were washed in 0.05% Tween 20/PBS and unspecific binding sites were saturated with BSA buffer (2.5% BSA/ 0.05% Tween 20/0.05% NaN_3 in PBS). Serial dilutions of E2 cell extracts were added to the plates in the final amount of 100 μl/well in BSA buffer and incubated for 2 hr at room temperature (RT). After washing with 0.5% Tween 20/PBS, 100 μl/well of anti-his tag mouse mAb (QIAGEN 34570) diluted 1/400 in BSA buffer were added and incubated for further 2 hr RT. Plates were washed and incubated for 1 hr RT with 100 μl/well of alkaline phosphatase conjugated secondary antibody (goat anti-mouse IgG Sigma A7434) diluted 1/2,000 in BSA buffer. After washing, alkaline phosphatase was revealed by incubation at 37°C with a 1 mg/ml solution of *p*-nitrophenyl phosphate in ELISA substrate buffer (10% diethanolamine buffer, 0.5 mM $MgCl_2$, pH 9.8). Results were expressed as the difference between OD_{405nm} and OD_{620nm} by an automated ELISA reader (Labsystems Multiskan Bichromatic, Helsinki, Finland).

CD81 Binding and mAb Competition Studies

ELISA plates (Nunc maxisorp, Roskiilde, Denmark) were coated with 1 μg/well of human CD81 diluted in PBS (2 mM Na_2HPO_4 H_2O/16 mM Na_2HPO_4 $2H_2O$ /150 mM NaCl). The second extracellular loop of human CD81 (residues 114–200) was expressed as a C-terminal fusion to Glutathione S-Transferase (GST) and was purified on glutathione sepharose matrix (Amersham Pharmacia Biotech AB, 17-0756-01). GST was employed as a control, using 3 μg/well.

Following o/n incubation at 4°C, plates were washed with 0.05% Tween 20/PBS and non-specific binding sites were blocked with 300 μl/ well of milk buffer (5% dry non-fat milk/ 0.05% Tween 20/ 0.05% NaN_3 in PBS) for 1 hr at 37°C. E2 proteins were diluted in milk buffer supplemented by 50 μg/ml of GST, preincubated 1 hr RT, and added in a final volume of 250 μl/well to CD81 plates for an o/n incubation at 4°C. After extensive washing with 0.05% Tween 20/PBS, 100 μl/well of anti-his tag mouse mAb (QIAGEN 34670) diluted 1/400 in 2.5% BSA/PBS was added and incubated for 3 hr at 4°C. The plates were washed, 100 μl/well of alkaline phosphatase conjugated secondary antibody (goat anti-mouse IgG Sigma A7434) diluted 1/2,000 in milk buffer added, and then incubated for 3 hr at 4°C. The plates were finally developed as described above.

For the competition experiments, mAbs were added in the indicated amount to the E2 proteins preincubation mix. mAb 166.F3 was generated in mice upon immunization with protein from the N2 viral isolate.

Heparin Binding Studies

Fifty microliters of E2 cell extracts were incubated o/n at 4°C with different amounts of either Heparin Sepharose CL6B (Pharmacia) or S-Sepharose (Pharmacia) preequilibrated in the lysis buffer. After centrifugation at 14 krpm for 15 min, flow-through was collected and the resin washed three times with 1 ml of lysis buffer. Heparin bound proteins were eluted by addition of SDS-polyacrylamide gel sample buffer, with and without 30 mM DTT and 2% β-mercaptoethanol.

For Western blot analysis, samples were separated on SDS-polyacrylamide gel (4–15% gradient precasted gel, Biorad) and transferred to nitro-cellulose membrane. After blocking with 5% dry non-fat milk/0.25% Tritonx100/ 0.02% Tween 20 in TBS, the membrane was incubated for 1 hr RT with anti-his tag mouse mAb diluted 1/250 in 2% BSA/0.25% Triton ×100/0.02% Tween 20 in 1× TBS. Following washing, bound antibody was detected with an alkaline phosphatase goat anti-mouse IgG (Sigma A7434) diluted 1/2,000 and developed with chromogenic alkaline phosphatase substrates NBT and BCIP (Sigma).

Sequence Alignment and Secondary Structure Prediction

A set of 50 HCV polyprotein sequences were collected using a BLAST46 search in the non-redundant NCBI database.[47] The E2 ectodomain regions (aa 384–714) were extracted from these and a subset created with maximum 70% pairwise sequence identity, following alignment using CLUSTALW (v. 1.7).[48] This alignment was used as input to the secondary structure prediction program, PHD.[49–52] Secondary structure predictions were similarly derived for the GBV-A, GBV-B, and GBV-C sequences. For GBV-A and GBV-C, 70% and 90% redundant multiple sequence alignments were used respectively, whereas the sole known GBV-B sequence was used.

Fold Recognition and Cross Comparison

The sequences of HCV E2 strain H (GenBank Accession Number AF011751) (see Fig. 1 for aa sequence), GBV-A (GenBank Accession Number AF023425), GBV-B (GenBank Accession Number U22304), and GBV-C (GenBank Accession Number AB003288) were chosen for fold recognition studies, using TOPITS,[53] THREADER2 (v. 2.1),[54,55] and both ProFIT,[56,57] and ProSup[58] from within ProCyon (v. 2.0).[59] THREADER2 outputs were analyzed by using

TABLE I. Relative Positions of mAb Epitopes,[34,76] Hypervariable Regions, the Putative E1/E2 Association Site, and Potential N-Glycosylation Sites

Feature	Residue range	Position in model
mAb 6/82a	384–391	Very exposed
mAb 7/16b	436–447	Partially exposed
mAb 166.F3	459–491	Very exposed
mAb 6/1a	464–471	Very exposed
mAb 6/41a	480–493	Very exposed
mAbs Cet-1 to -6	528–546	Very exposed
mAb 6/53	544–551	Very exposed
HVR1	384–410	Very exposed
HVR2	474–482	Very exposed
E1/E2 association site	487–489	Very exposed
N-X-S glycosylation site	417–419	Very exposed
N-X-S glycosylation site	430–432	Very exposed
N-X-S glycosylation site	448–450	Very exposed
N-X-S glycosylation site	476–478	Very exposed
N-X-T glycosylation site	423–425	Buried
N-X-T glycosylation site	532–534	Exposed
N-X-T glycosylation site	540–542	Buried
N-X-T glycosylation site	556–558	Exposed
N-X-T glycosylation site	576–578	Very exposed
N-X-T glycosylation site	623–625	Very exposed
N-X-T glycosylation site	645–647	Mostly buried

the graphical interface TAN.[60] Shuffled randomization tests were performed using the top scoring 50 results from THREADER2.

For each fold recognition method the results were classified in terms of the top scoring 30 hits (only top 7 in shuffled THREADER2; in ProCyon high scoring small structures were discarded). Hierarchical classification was performed based on pair-wise potential energy scores. For each hit to a particular PDB[61] entry, a manual assignment of CATH[62] (domain list caths_list.oct19), and SCOP[63] (release 1.37 pdb100d) classification names and numbers was made. The frequency of occurrence of each SCOP classification type in GBV-A and HCV was analyzed (using the Class, Fold, and Superfamily identification numbers). The shuffled THREADER2 alignments were classified into regions which matched or mismatched the predicted PHD output, allowing specific SCOP classes to be selected as potential candidates for model building.

The sequence of the envelope glycoprotein E from TBEV[64,65] was studied using identical techniques of fold recognition and secondary structure prediction as for the HCV and GB viruses.

Model Building and Analysis

Model generation initiated from a shuffled THREADER2 alignment between the target HCV E2 strain H sequence and that of the template structure, and which had subsequently been optimized for secondary structure overlap. Substitution of amino acid residues and modeling of insertions or deletions in the target structure were performed using the InsightII software.[66] The model was regularized using 100 steps of the Steepest Descents algorithm with the CVFF forcefield.[67]

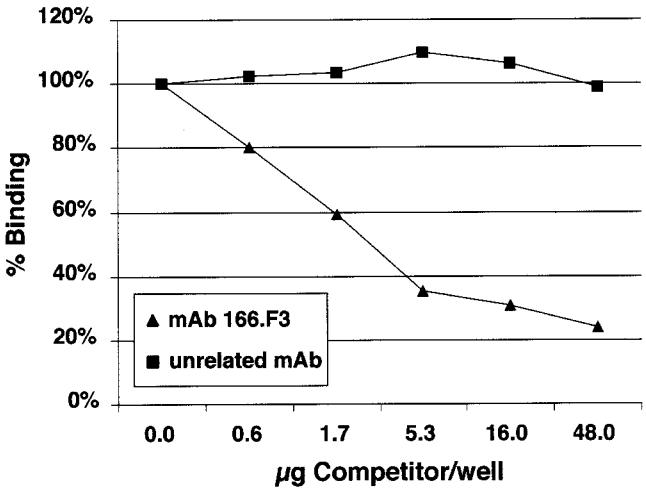

Fig. 4. Inhibition of the binding of N2 strain E2 protein to human CD81 with anti-E2 mAb 166.F3 and, as control, with an unrelated monoclonal antibody. Results are expressed as percentage of inhibition.

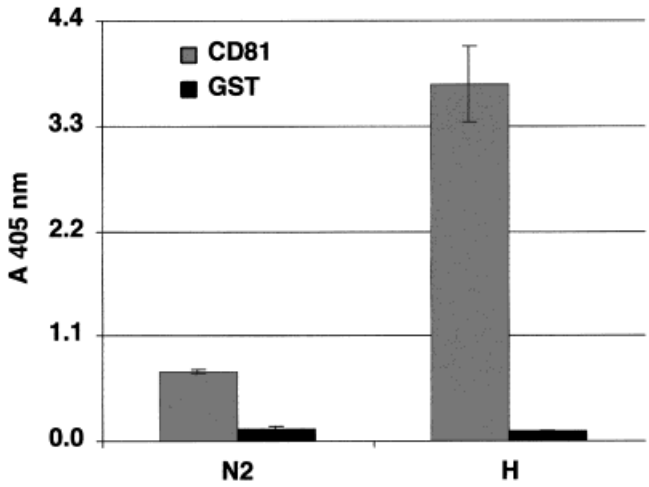

Fig. 5. Binding of N2 and H strain E2 proteins to human CD81 (hCD81) and GST control. The relative amounts of E2 have been normalized as described in Materials and Methods. Average values from two independent experiments are reported, with corresponding standard deviations.

The PepPlot function within GCG (v. 10.0)[68] was used to create a Kyte and Doolittle hydropathy plot[69] of the HCV E2 sequence, using a nine residue window. Relative residue accessibility was calculated for the HCV E2 monomer and dimer models using NACCESS.[70,71] The single residue values thus obtained were averaged over a nine residue window and plotted together with the results from the hydropathy calculation.

Surface potential calculations were performed from within the MOLMOL software (v. 2.6),[72] using a finite difference[73] based algorithm by Honig et al.[74] to solve the Poisson-Boltzmann equation. The computations, employing heavy atoms only, were performed using a "simple charge" representation of atomic charges and default software settings for the remaining variables. Graphic

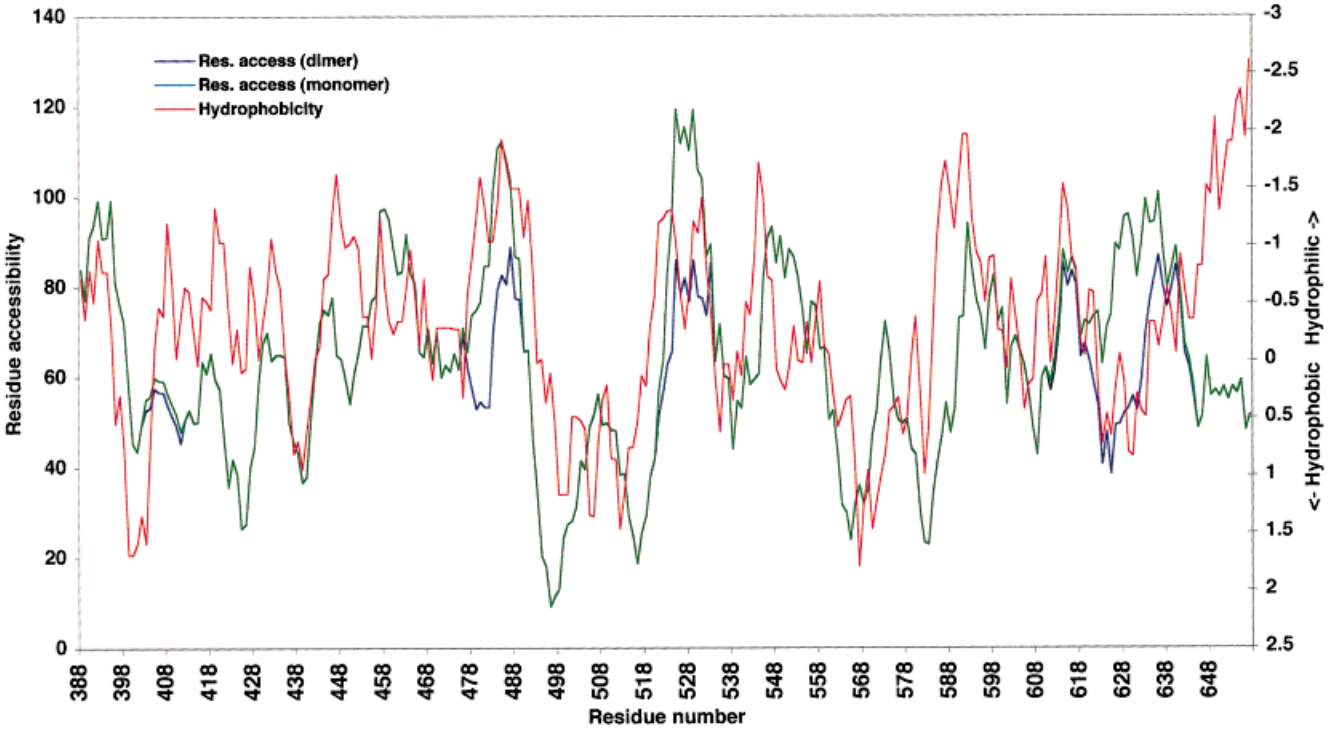

Fig. 6. Graph plotting hydropathy vs. residue accessibility for the monomeric and dimeric models of HCV E2. Hydrophobicity increases on descending the Y-axis, whereas residue accessibility decreases.

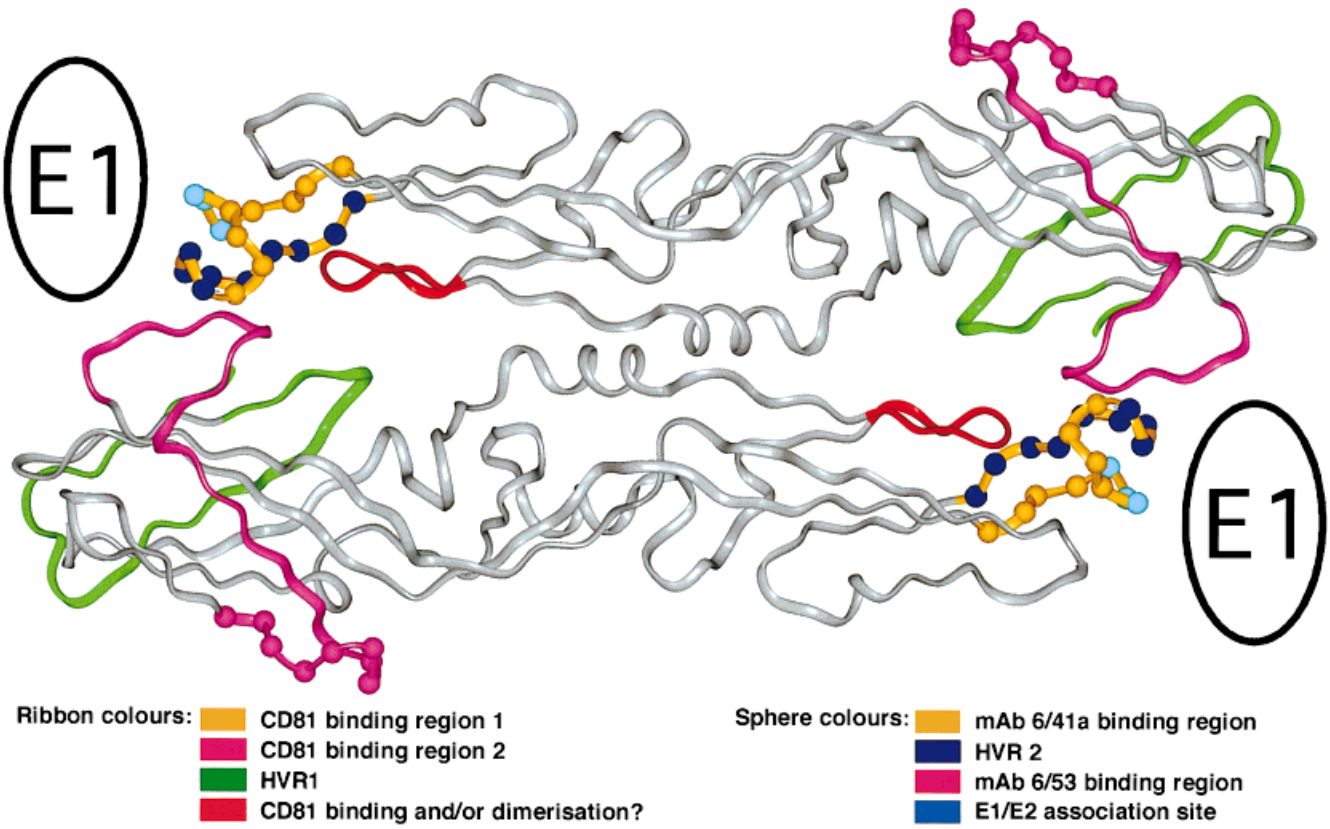

Fig. 7. A unified model for the HCV envelope glycoproteins E1 and E2. The model depicts E2 as a head-to-tail homodimer, possibly covalently linked, and forming a heterodimeric association with E1.

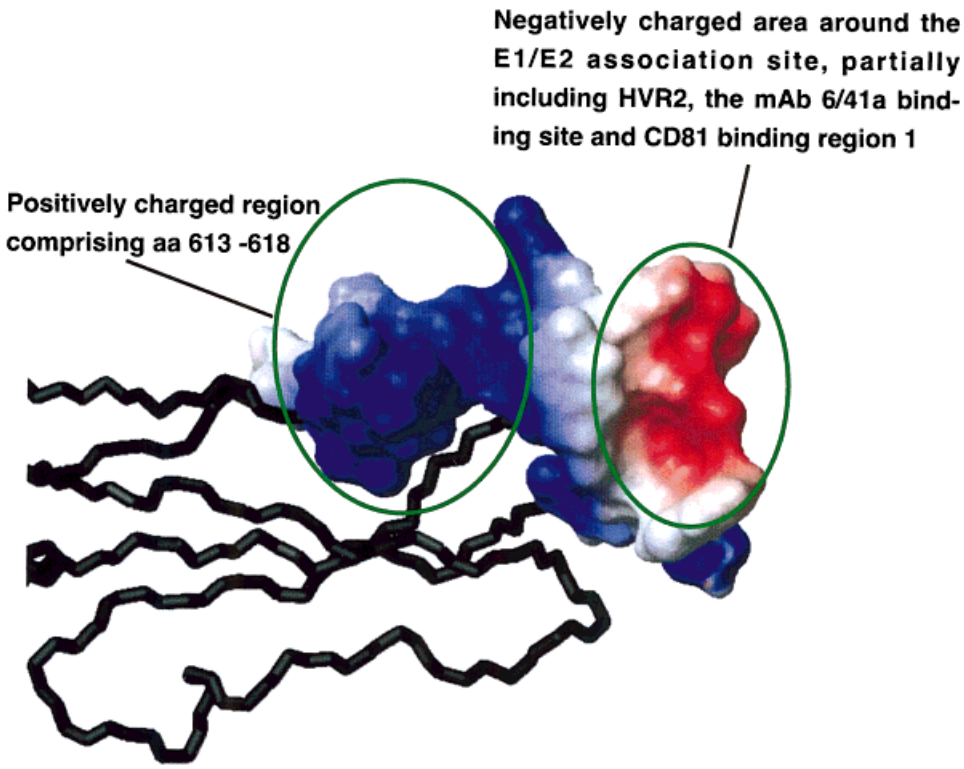

Fig. 8. A surface potential calculation of the area surrounding the E1/E2 association site in E2 showing it is negatively charged. The adjacent area comprising residues 613–618 forms part of the positively charged region spanning aa 612–620 and may be involved in CD81 binding and/or dimerization. Figure prepared using the program MOLMOL.[72]

visualization of the results was possible in MOLMOL by performing a molecular surface calculation using heavy atoms only, followed by coloring according to the surface potential map computed previously.

RESULTS AND DISCUSSION

Structure Prediction

Secondary structure prediction of the ectodomain of HCV E2, using only HCV sequences, suggested an overall low secondary structure content (~37%) of predominantly β-strands. Similar results were obtained for the GBV-A and GBV-C E2 sequences, with the exception of a few additional weakly predicted α-helices. Owing to the low pairwise sequence identity to HCV E2 (<20%), attempts to align these sequences using sequence information and/or through their predicted secondary structure were unsuccessful and gave ambiguous results. Predictions for the single GBV-B E2 sequence again indicated a predominantly β-type secondary structure, although with much lower reliability scores. As for GBV-A and GBV-C, no reliable sequence alignment could be established between GBV-B E2 and HCV E2, also hampered by the fact that the latter sequence is approximately 70 residues longer than the corresponding GBV-B E2 ectodomain. For these reasons, the GBV-B sequence was not considered any further. The conclusion from the secondary structure predictions was therefore a qualitative consensus indicating a relatively low secondary structure content of mainly β-type for HCV E2.

Fold recognition techniques were then applied to search for potential target folds which could represent a starting point for model building. The use of diverse software in this case was intended primarily to discover if any particular fold was common to more than one technique, thus making it a more likely candidate, as well as to somewhat counter the occurrence of false positives and possible "software bias" towards particular fold families.

The TOPITS results for HCV, GBV-A, and GBV-C (data not shown) showed many β-folds as expected, due to the dependence of the algorithm on the predicted secondary structure. For these sequences TOPITS also identified the envelope glycoprotein E from TBEV (PDB code 1SVB)[64,65] which was particularly interesting because of its functional similarity to HCV E2. The hypothesis that E2 adopts a β-fold was then independently confirmed by analyzing the THREADER2 and ProCyon results for the HCV, GBV-A, and GBV-B sequences. In the case of THREADER2, although acid proteases scored highly for all target sequences, chymosin from bovine (PDB code 1CMS)[75] was found for both HCV and GBV-C. A control experiment was also performed using the TBEV E sequence as a target for the THREADER2 and ProCyon software to verify whether the programs would select similar folds for TBEV as those found for HCV and GBV-A. This was indeed the case, indicating that these proteins are similar from the point of view of the parameters in these fold recognition methods.

A thorough comparison of all the results obtained by the different methods on the HCV, GBV-A, GBV-C, and TBEV E sequences suggested two things: first, the PDB entry 1CMS was a candidate structure for model building since it was found as a high scoring hit with all sequences except

GBV-A, and the alignment with HCV E2 was reasonable in terms of residue identity, number of gaps, and secondary structure agreement (based on number of matching and mismatching regions); and second, since the TBEV E sequence had found folds in common with HCV, GBV-A, and GBV-C, it too could be considered a likely possibility for model building of HCV E2. Biologically, the functionally closely related TBEV envelope glycoprotein E would be expected to retain a distant structural similarity to HCV E2, thus reinforcing its candidature for model building.

The HCV E2 Model

In the final analysis, a protein model may be assessed by three means. The first includes checking the structure to ascertain whether simple physicochemical rules are obeyed (e.g., hydrophobic residues buried in the core and hydrophilic surface exposed) and employing software to perform stereochemical checks. However, this gives a broad overview of the model and only highlights major potential problems, especially in cases of models derived from fold recognition results. A second and more informative technique is to map all known experimental information onto the structure and determine the extent of agreement between the two. For example, epitopes and N/O-glycosylation regions are expected to be exposed and cysteine residues known to form disulphide bridges should be within bonding distance. The final method is to use the model to predict and then test protein behavior under specific experimental conditions.

Sequence alignments from which to build HCV E2 models based on 1CMS and 1SVB were obtained from shuffled THREADER2 runs, with subsequent manual optimization for good overlap between the predicted and observed secondary structure elements. However, in the case of 1CMS, this placed monoclonal antibody (mAb) binding regions (aa 436–447,[34] aa 528–546[76]) and the second hypervariable region (HVR2, aa 474–482) within the inaccessible core of the protein. Therefore, although this model seemed reasonable from a protein architecture viewpoint, the conclusion was that it most likely did not reflect the general fold of the protein.

In the case of the second candidate template, 1SVB, the sequence alignment (Fig. 2) and resulting model (Fig. 3) attempted to retain as much as possible of the original central β-barrel (domain I), in accordance with the dominant type of secondary structure predicted. In doing so, four loop insertions of varying lengths had to be accomodated in the original β-barrel. The dimerization domain II of 1SVB required several deletions in the HCV E2 model and one insertion, which was acceptable since this domain contains several unusually long loop regions.

The current model of E2 (strain H) terminates at aa 661 (1SVB aa 297), prior to the start of 1SVB domain III (Figs. 2, 3). In fact, the 661-truncated E2 protein is sufficient to bind CD81,[34,41] to be exported [77,78] and to heterodimerize with E1.[25] Therefore, it is reasonable to assume that aa 384–661 represent the structural core of a functional E2 protein and the model reflects this thinking. The equivalent portion of the C-terminal IG-like domain III of 1SVB is missing in HCV E2 and residues 662 to 714 should serve as a membrane-anchor, analogous to the TBEV E residues 396–447.[79]

Mapping known experimental observations onto the model showed it was generally consistent with such data (Table I): mAb binding sites were mostly on the surface of the protein, apart from a small stretch of residues (aa 436–440) at the edge of one mAb recognition site, as were most of the N-X-S/T potential glycosylation sites. Further analysis of the model allowed some conclusions to be drawn about the biological properties of E2, in particular identification of the CD81 and heparin binding sites as well as a scheme for the formation of an E1/E2 complex.

CD81 Binding Region

It is known that binding of mAb 6/41a (epitope: aa 480–493) or mAb 6/53 (epitope: aa 544–551) (Table I) to E2 inhibits CD81 binding, while that of mAb 7/16b (epitope: aa 436–447) or mAb 6/1a (epitope: aa 464–471) (Table I) does not.[34] Similar experiments using mAb 166.F3, whose epitope maps in the region aa 459–491 and which specifically recognizes the N2 strain of E2, showed it competes for binding with CD81 (Fig. 4). This prompted further tests to investigate whether different strains of HCV E2 bound CD81 with varying affinities. The results demonstrated that this was indeed the case, since E2 strain N2 showed greatly reduced binding efficiency compared to strain H (Fig. 5). Therefore, the CD81 binding site on E2 seems, at least in part, to be characterized by sequence differences between the N2 and H strain E2 proteins. Excluding the HVR1 region from the analysis, comparison of the two sequences (Fig. 1) highlighted aa positions 424, 444, 454, 461, 463, 464, 466, 476, 479, 482, 492, 493, 500, 522, 524, 531, 546, 570, 641, 653, 655, and 671 as being significantly different.

In conjuction with the model, the combined data thus implied a composite CD81 binding site including residues 474–494 (around the HVR2 region) (binding region 1), and 522–551 (binding region 2) (Fig. 3). Both of these regions are exposed and, importantly, come close to each other in a dimeric E2 model (see below).

Closer analysis of the predicted CD81 binding areas showed both contained a high number of charged polar residues. Examination of a CD81 multiple sequence alignment (data not shown) highlighted the second extracellular loop of CD81 (aa 114–200) as similarly having a high concentration of charged or polar residues. Differences between the CD81 sequences from human and green monkey, the latter of which does not bind HCV E2, also map in this region (aa positions 154, 163, 186, 188, and 197),[41] as does the single difference between CD81 from human and chimpanzee (aa 197), the only known species liable to HCV infection. All these differences are therefore located in the region of CD81, which has been experimentally demonstrated to be sufficient for E2 binding.[41]

E2 Homodimerization and the E1/E2 Association Site

The model of HCV E2 predicts probable disulphide bridges between C429–C644, C452–C486, C552–C564, and C597–C607, as well as one between either C508–C581 or C581–C585. The majority of these are in the region corresponding to the dimerization domain II in 1SVB, presumably instilling stability in this part of the elongated structure. The complete E2 ectodomain sequence (aa 384–714) contains 18 cysteines, allowing nine potential cys-cys bridges. However, the aa 661 truncated sequence contains an odd number of 17 cysteine residues, thus inferring eight or less intramolecular, and possibly one or more intermolecular, disulphide bridges. Further evidence that HCV E2 might not be monomeric was obtained by comparing hydropathy and residue accessibility along the chain. The corresponding plot showed relatively good agreement between the two graphs (Fig. 6) except for the region between aa 613–635, which contained several highly accessible hydrophobic residues, and the region aa 638–661, which showed the opposite situation.

Extrapolating knowledge from the TBEV structure onto HCV led to the idea that E2 may similarly exist as a head-to-tail homodimer. A disulphide linked covalent homodimeric structure could explain the odd number of cysteine residues in the aa 661 truncated form of E2. The most likely candidate for such covalent interaction was C620 from one monomer, although it was unclear which other residue may be involved from the other monomer since all cysteines were too distant in the approximate model.

An HCV E2 homodimer (Fig. 7) was created using the 1SVB dimer as a reference. The corresponding hydropathy vs. residue accessibility graph (Fig. 6) showed generally better agreement (especially in the region aa 613–635), thus supporting the dimer hypothesis. The analysis also suggested care be taken in the interpretation of any data within the region aa 638–661, where the overlap between the two graphs values did not improve. Whether this reflects an incorrectly predicted local conformation, an erroneous local sequence alignment, or other factors was not clear.

Of particular note is the alignment between aa 452–495 in HCV E2 and 1SVB aa 74–114, where one of the original 1SVB disulphide bridges (C74–C105) was retained (Fig. 2). This region contains the "cd" loop in 1SVB (aa 98–113),[65] which is almost completely conserved in Flaviviruses and has been proposed as important for fusogenic activity.[80,81] In the corresponding region of the HCV E2 model (aa 476–494) one finds a hypervariable region (HVR2) followed by a highly conserved stretch of residues (see Figs. 1, 2). Although very little is known about the HCV cell fusion process, the alignment between 1SVB and HCV in this region would suggest residues 476–494 in E2 may play a role in viral fusion or a related process. Since the cd loop in 1SVB also forms hydrophobic contacts with domains I and III of the other subunit,[65] extending the analogy between HCV and 1SVB could therefore mean that E1 occupies a position equivalent to domain III of the TBEV E protein. Strong support for this hypothesis comes from the observation that the E2 segment aa 476-494 includes the "WHY" motif (aa 487-489) (Figures 3 & 7), which has been shown experimentally to be important for heteromeric association between E1 and E2.[82] A surface potential calculation of the area, partially including the neighboring HVR2 region, showed it contained a localized negative charge (Fig. 8) and suggested it may interact with a positively charged region in the E1 protein.

Further examination of the HCV E2 sequence and model showed the positively charged region spanning aa 612–620 (PYRLWHYPC) (Fig. 8) was highly similar to aa 484–494 (PYCWHYPPRPC) within CD81 binding region 1 and the region possibly important for viral fusion (aa 476–494). These two regions are also physically very close in the model (Figs. 3, 7), leading to the hypothesis that aa 613–618 may also be involved in CD81 binding and/or in E1/E2 or E2/E2 dimerization. Although it cannot be excluded that this may be due to partial sequence duplication within HCV E2, a dotplot analysis did not suggest this was the case.

The Heparin Binding Domain

It has been reported that binding of Dengue Flavivirus to its cellular receptor can be blocked by heparin and suramin.[42] Since suramin also blocks HCV binding to hepatoma cells in vitro,[43] in a mode distinct from the non-specific action of any polyanionic compound, the existence of a heparin binding motif in the HCV envelope had been proposed.[42,43] Therefore the ability of E2 protein from crude cell extracts to bind heparin cross-linked to a sepharose matrix was tested. The results showed that E2 bound heparin but, under the same conditions, did not bind the control polyanionic S-Sepharose (Fig. 9), indicating the presence of a specific binding motif within the E2 structure. It was noted that the E2 protein was not completely depleted by the heparin column (Fig. 9) but also detected in the flow-through. However, the matrix bound proteins did not differ from the unbound flow-through fraction in terms of disulphide dependent aggregates. Essentially the same ratio between bound and unbound protein was present also under non-reducing conditions (data not shown). Although the presence of disulphide independent misfolded aggregates and/or glycosylation forms which are not able to bind heparin cannot be ruled out, limits of the assay design are the most likely explanation for the unbound fraction.

An analysis of heparin binding domains within SCOP[63] suggested they are often located in regions rich in positively charged residues, with cys-cys disulphide linkages and possibly containing one or two small β-strands and a small α-helix. In the current model such a motif exists in the region aa 559–614, suggesting it as a possible region for the heparin binding site.

A Unified Model for the HCV Envelope Glycoproteins

Bearing in mind all the available experimental data, an overall picture of E2 and its possible interaction with E1

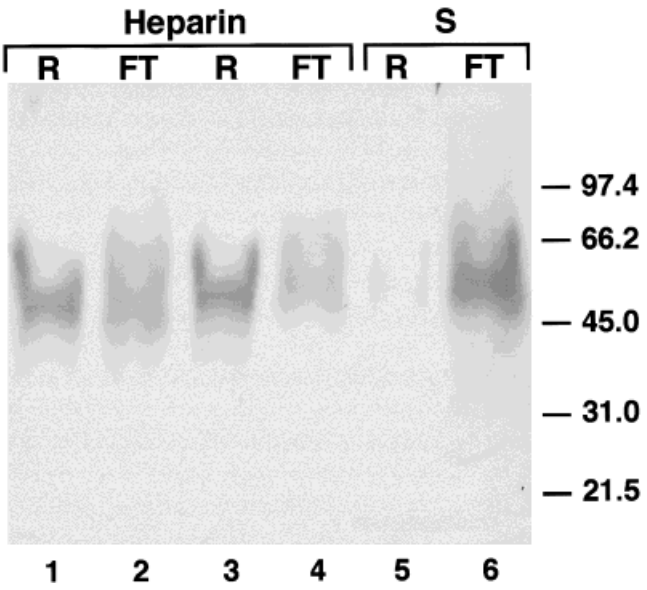

Fig. 9. Heparin binding of H strain E2, with molecular mass of proteins (kDa) indicated on the right. Under reducing conditions lysates of E2 expressing cells were incubated with increasing amount of Heparin sepharose: lane 1– 50 μl of 50% slurry gel resin (R) and lane 2 respective flow-through (FT); lane 3–150 μl of 50% slurry gel and lane 4 respective flow-through. With S-sepharose control matrix: lane 5–150 μL of 50% slurry gel and lane 6 respective flow-through.

began to emerge. The head-to-tail E2 homodimer, perhaps covalently linked, would explain the presence of two spatially distinct CD81 binding regions (Fig. 3). The E1 protein would then be located at the "end" of each E2 dimerization domain, thus forming a homodimeric pair of heterodimers (Fig. 7). This set-up places part of E1 close to the CD81 binding regions on E2 and may explain reports that it too is involved in CD81 binding.[83]

Since the model of HCV E2 envisages a homodimer based on the TBEV E structure, it would be reasonable to assume that its physical relationship to the viral membrane is also similar. On initial examination this would place HVR1 and part of HVR2 on the lower face of the protein, facing the membrane. However, several pieces of evidence suggest the model is still reasonable. A number of reports have shown the elongated, flat TBEV structure rearranges to form trimeric spikes at low pH, thus presumably exposing part of the underside of domain II and the fusion peptide loop (aa 98–113).[84–86] Under such circumstances, in HCV E2 the proposed CD81 binding region 1 and HVR2 could become more exposed, as expected. An associated question concerns the extent to which the relative orientation between domains I and II is conserved between TBEV E and HCV E2. However, domain modeling is still beyond the limits of present computational techniques, especially in such cases as described here where essentially no sequence homology is present. Nonetheless, even if small rearrangements of the relative domain positioning occurred, the features of the proposed dimeric E2 model should be conserved. It is therefore also conceivable that HVR1, one of the principal neutralization determinants of HCV, might become more accessible to antibody recognition, thus explaining its strong immunogenic character.

Extrapolating from knowledge of the closely related genomic organization of GB viruses, it would be reasonable to expect that the gross features of the model presented here for HCV should also apply for the E2 proteins of these viruses. An exception seems to be GBV-B, which, despite its overall polyprotein sequence similarity with HCV, shows a major difference in that the ectodomain of its E2 protein equivalent, is about 70 aa shorter than that of GBV-A/C and HCV. The resulting low sequence homology with HCV E2 and complete lack of experimental data for GBV-B E2 does not allow this overall genomic similarity with HCV to be further exploited.

CONCLUSIONS

Similarities in genomic organization between the Flaviviral, Pestiviral, and HCV sequences have been known since HCV was discovered. In some cases these directly translate into functional and, extrapolating from conserved sequence motifs, even into structural similarities. Examples are the NS3-protease important for viral polyprotein processing[87] and the NS5B RNA-dependent RNA polymerase necessary for viral replication.[88] The extent to which functional and, in particular, structural equivalences can be inferred from an overall related genomic organization is however not clear, especially at undetectable levels of amino acid sequence similarity. Besides providing a platform for designing experiments, the present model of the HCV E2 ectodomain precisely addresses this aspect of viral evolution. Confirmed or rejected at a later stage, the model and its implied resemblance to TBEV E will allow steps forward in answering these questions.

Site-directed mutagenesis experiments can be used to test both the proposed location and strain specificity of CD81 binding. In addition, covalent E2/E2 homodimerization could be tested by mutating C620 to alanine as could the impact of aa 613–618 (YRLWHY) on homo/- or heterodimerization. Presently, these experiments are hampered by the lack of efficient techniques to measure E1/E2 and E2/E2 dimerization.

Although many aspects regarding the function and properties of HCV E2 still remain unresolved, the model has nonetheless succeeded in explaining a great amount of experimental data and, most importantly, suggested ways forward for experimental investigations. Last but not least, these in turn may furnish essential information towards on-going vaccine efforts.

ACKNOWLEDGMENTS

This work was supported by an EU TMR fellowship to Asutosh T. Yagnik. The authors would like to thank Giovanni Migliaccio for his many useful discussions on the subject and Shoshana Levy for providing the GST-CD81 expression plasmid.

REFERENCES

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 1989;244:359–362.

2. Alter HJ et al. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. N Engl J Med 1989;321:1494–1500.
3. Alter MJ et al. The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. N Engl J Med 1992;327:1899–1905.
4. Dittmann S, Roggendorf M, Durkop J, Wiese M, Lorbeer B, Deinhardt F. Long-term persistence of hepatitis C virus antibodies in a single source outbreak. J Hepatol 1991;13:323–327.
5. Gish RG, Lau JYN. Hepatitis C virus: eight years old. Viral Hep Rev 1997;13:17–37.
6. Di Bisceglie AM. Hepatitis C and hepatocellular carcinoma. Semin Liv Dis 1995;15:64–69.
7. Major ME, Feinstone SM. The molecular virology of hepatitis C. Hepatology 1997;25:1527–1538.
8. Davis GL. Interferon treatment of chronic hepatitis C. Am J Med 1994;96:41S–46S.
9. Reichard O, Schvarcz R, Weiland O. Therapy of hepatitis C: alpha interferon and ribavirin. Hepatology 1997;26:108S–111S.
10. Houghton M, Weiner A, Han J, Kuo G, Choo QL. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology 1991;14:381–388.
11. Takamizawa A et al. Structure and organization of the hepatitis C virus genome isolated from human carriers. J Virol 1991;65:1105–1113.
12. Miller RH, Purcell RH. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc Natl Acad Sci USA 1990;87:2057–2061.
13. Weiner AJ et al. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 1991;180:842–848.
14. Francki RIB, Fauquet CM, Knudson DL, Brown F. Classification and nomenclature of viruses: fifth report of the international committee on taxonomy of viruses. Arch Virol 1991;2:223.
15. Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc Natl Acad Sci USA 1991;88:5547–5551.
16. Hijikata M et al.. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J Virol 1993;67:4665–4675.
17. Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J Virol 1993;67:2832–2843.
18. Tomei L, Failla C, Santolini E, De Francesco R, La Monica N. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. J Virol 1993;67:4017–4026.
19. Lin C, Lindenbach BD, Pragai BM, McCourt DW, Rice CM. Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. J Virol 1994;68:5063–5073.
20. Miyamura T, Matsuura Y. Structural proteins of hepatitis C virus. Trends Microbiol 1993;1:229–231.
21. Lagging LM, Meyer K, Owens RJ, Ray R. Functional role of hepatitis C virus chimeric glycoproteins in the infectivity of pseudotyped virus. J Virol 1998;72:3539–3546.
22. Rice CM. *Flaviviridae*: the viruses and their replication. In: Fields BN et al., editors. Fields virology, 3rd edition. Philadelphia: Raven; 1996. p 931–960.
23. Grakoui A, Wychowski C, Lin C, Feinstone SM, Rice CM. Expression and identification of hepatitis C virus polyprotein cleavage products. J Virol 1993;67:1385–1395.
24. Ralston R et al. Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses. J Virol 1993;67:6753–6761.
25. Patel J, Patel AH, McLauchlan J. Covalent interactions are not required to permit or stabilize the non-covalent association of hepatitis C virus glycoproteins E1 and E2. J Gen Virol 1999;80:1681–1690.
26. Choo QL et al. Vaccination of chimpanzees against infection by the hepatitis C virus. Proc Natl Acad Sci USA 1994;91:1294–1298.
27. Rosa D et al. A quantitative test to estimate neutralizing antibodies to the hepatitis C virus: cytofluorimetric assessment of envelope glycoprotein 2 binding to target cells. Proc Natl Acad Sci USA 1996;93:1759–1763.
28. Mizushima H, Hijikata M, Asabe S, Hirota M, Kimura K, Shimotohno K. Two hepatitis C virus glycoprotein E2 products with different C termini. J Virol 1994;68:6215–6222.
29. Taniguchi S et al. A structurally flexible and antigenically variable N-terminal domain of the hepatitis C virus E2/NS1 protein: implication for an escape from antibody. Virology 1993;195:297–301.
30. Shimizu YK, Hijikata M, Iwamoto A, Alter HJ, Purcell RH, Yoshikura H. Neutralizing antibodies against hepatitis C virus and the emergence of neutralization escape mutant viruses. J Virol 1994;68:1494–1500.
31. Puntoriero G et al. Towards a solution for hepatitis C virus hypervariability: mimotopes of the hypervariable region 1 can induce antibodies cross-reacting with a large number of viral variants. EMBO J 1998;17:3521–3533.
32. Lohmann V, Korner F, Koch J, Herian U, Theilmann L, Bartenschlager R. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. Science 1999;285:110–113.
33. Dubuisson J. Folding, assembly and subcellular localization of hepatitis C virus glycoproteins. Curr Top Microbiol Immunol 2000;242:135–148.
34. Flint M et al. Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. J Virol 1999;73:6235–6244.
35. Flint M et al. Functional characterization of intracellular and secreted forms of a truncated hepatitis C virus E2 glycoprotein. J Virol 2000;74:702–709.
36. Sternberg MJ, Bates PA, Kelley LA, MacCallum RM. Progress in protein structure prediction: assessment of CASP3. Curr Opin Struct Biol 1999;9:368–373.
37. Zuckerman AJ. The new GB hepatitis viruses. Lancet 1995;345:1453–1454.
38. Simons JN et al. Identification of two flavivirus-like genomes in the GB hepatitis agent. Proc Natl Acad Sci USA 1995;92:3401–3405.
39. Simons JN et al. Isolation of novel virus-like sequences associated with human hepatitis. Nat Med 1995;1:564–569.
40. Zuckerman AJ. Alphabet of hepatitis viruses. Lancet 1996;347:558–559.
41. Pileri P et al. Binding of hepatitis C virus to CD81. Science 1998;282:938–941.
42. Chen Y et al. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulphate. Nat Med 1997;3:866–871.
43. Garson JA, Lubach D, Passas J, Whitby K, Grant PR. Suramin blocks hepatitis C binding to human hepatoma cells *in vitro*. J Med Virol 1999;57:238–242.
44. Kolykhalov AA, Agapov EV, Blight KJ, Mihalik K, Feinstone SM, Rice CM. Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. Science 1997;277:570–574.
45. Esumi M et al. Immunoreactive core peptides of hepatitis C virus produced in *Escherichia coli* and *in vitro* DNA amplification-restricted transcription-translation system. J Virol Methods 1996;59:91–98.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
47. Benson DA et al. GenBank. Nucleic Acid Res 1999;27:12–17.
48. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acid Res 1994;22:4673–4680.
49. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993;232:584–599.
50. Rost B, Sander C. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. Proc Natl Acad Sci USA 1993;90:7558–7562.
51. Rost B, Sander C, Schneider R. PHD: an automatic mail server for protein secondary structure prediction. Comput Appl Biosci 1994;10:53–60.
52. Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 1994;19:55–72.
53. Rost B. TOPITS: threading one-dimensional predictions into three-dimensional structures. In: Rawlings C, Clark D, Altman R, Hunter L, Lengauer T, Wodak S, editors. Third international

conference on intelligent systems for molecular biology. Cambridge, UK: AAAI Press; 1995. p 314–321.
54. Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. Nature 1992;358:86–89.
55. THREADER2, Version 2.1. Jones DT; 1998. Avaliable from Department of Biological Sciences, University of Warwick, Coventry, UK. http://globin.bio.warwick.ac.uk/~jones
56. Flöckner H, Braxenthaler M, Lackner P, Jaritz M, Ortner M, Sippl MJ. Progress in fold recognition. Proteins 1995;23:376–386.
57. Sippl MJ, Flöckner H. Threading thrills and threats. Structure 1996;4:15–19.
58. Feng ZK, Sippl MJ. Optimum superimposition of protein structures: ambiguities and implications. Fold Des 1996;1:123–132.
59. ProCyon, Version 2.0. Salzburg, Austria: ProCeryon Biosciences GmbH; 1998. The ProCyon software previously from MJ Sippl at Center of Applied Molecular Engineering, Salzburg, Austria, is now known as ProCeryon. http://www.proceryon.com
60. Threading ANalyst (TAN): a program to evaluate threading results. Miller RT, Thornton JM; 1995. Available from within THREADER2 (see reference 55) or contact RT Miller (rmiller@sanbi.ac.za).
61. Bernstein FC et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112: 535–542.
62. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH: a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.
63. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
64. Heinz FX et al. The flavivirus envelope protein E: isolation of a soluble form from tick-borne encephalitis virus and its crystallization. J Virol 1991;65:5579–5583.
65. 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.
66. InsightII, Version 98.0. San Diego: Molecular Simulations, Inc.; 1998.
67. Consistent Valence Force-Field (CVFF). Molecular Simulations, Inc. Available as part of the Discover program from within the InsightII software suite (see reference 66).
68. Wisconsin Package, Version 10.0. Wisconsin, USA: Genetics Computer Group (GCG); 1999.
69. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
70. Hubbard SJ, Campbell SF, Thornton JM. Molecular recognition: conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. J Mol Biol 1991;220:507–530.
71. NACCESS, Version 2.1.1. Hubbard SJ, Thornton JM; 1993. Now available from Inpharmatica, Ltd., London, UK.
72. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55, 29–32.
73. Warwicker J, Watson HC. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. J Mol Biol 1982;157: 671–679.
74. Honig B, Nicholls A. A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comp Chem 1991;12:435–445.
75. Gilliland GL, Winborne EL, Nachman J, Wlodawer A. The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. Proteins 1990;8:82–101.
76. Lee JW, Kim K, Jung SH, Lee KJ, Choi EC, Sung YC, Kang CY. Identification of a domain containing B-cell epitopes in hepatitis C virus E2 glycoprotein by using mouse monoclonal antibodies. J Virol 1999;73:11–18.
77. Spaete RR, Alexander D, Rugroden ME, Choo QL, Berger K, Crawford K, Kuo C, Leng S, Lee C, Ralston R, Thudium K, Tung JW, Kuo G, Houghton M. Characterization of the hepatitis C virus E2/NS1 gene product expressed in mammalian cells. Virology 1992;188:819–830.
78. Selby MJ, Glazer E, Masiarz F, Houghton M. Complex processing and protein:protein interactions in the E2:NS2 region of HCV. Virology 1994;204:114–122.
79. Allison SL, Stiasny K, Stadler K, Mandl CW, Heinz FX. Mapping of functional elements in the stem-anchor region of tick-borne encephalitis virus envelope protein E. J Virol 1999;73:5605–5612.
80. Roehrig JT, Hunt AR, Johnson AJ, Hawkes RA. Synthetic peptides derived from the deduced amino acid sequence of the E-glycoprotein of Murray Valley encephalitis virus elicit antiviral antibody. Virology 1989;171:49–60.
81. Volkova TD, Vorovitch MF, Ivanov VT, Timofeev AV, Volpina OM. A monoclonal antibody that recognizes the predicted tick-borne encephalitis virus E protein fusion sequence blocks fusion. Arch Virol 1999;144:1035–1039.
82. 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.
83. Flint M et al. Functional analysis of cell surface-expressed hepatitis C virus E2 glycoprotein. J Virol 1999;73:6782–6790.
84. Allison SL, Schalich J, Stiasny K, Mandl CW, Kunz C, Heinz FX. Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. J Virol 1995;69:695–700.
85. Helenius A. Alphavirus and flavivirus glycoproteins: structures and functions. Cell 1995;81:651–653.
86. Stuart D, Gouet P. Viral envelope glycoproteins swing into action. Structure 1995;3:645–648.
87. Ryan MD, Monaghan S, Flint M. Virus-encoded proteinases of the Flaviviridae. J Gen Virol 1998;79:947–959.
88. Bressanelli S et al. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc Natl Acad Sci USA 1999;96: 13034–13039.
89. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.