

Contribution of charged and polar residues for the formation of the E1–E2 heterodimer from Hepatitis C Virus

Siti Azma Jusoh · Christoph Welsch · Shirley W. I. Siu · Rainer A. Böckmann · Volkhard Helms

Received: 24 November 2009 / Accepted: 25 January 2010 / Published online: 2 March 2010
© Springer-Verlag 2010

Abstract The transmembrane domains of the envelope glycoprotein E1 and E2 have crucial multifunctional roles in the biogenesis of hepatitis C virus. We have performed molecular dynamics simulations to investigate a structural model of the transmembrane segments of the E1–E2 heterodimer. The simulations support the key role of the Lys370–Asp728 ion pair for mediating the E1–E2 heterodimerization. In comparison to these two residues, the simulation results also reveal the differential effect of the conserved Arg730 residue that has been observed in experimental studies. Furthermore, we discovered the formation of *inter*-helical hydrogen bonds *via* Asn367 that stabilize dimer formation. Simulations of single and double mutants further demonstrate the importance of the ion-pair

and polar interactions between the interacting helix monomers. The conformation of the E1 fragment in the simulation of the E1–E2 heterodimer is in close agreement with an NMR structure of the E1 transmembrane segment. The proposed model of the E1–E2 heterodimer supports the postulated cooperative insertion of both helices by the translocon complex into the bilayer.

Keywords Charged amino acids · Hepatitis E1-E2 envelope glycoproteins · Molecular dynamics simulation · Salt-bridges · Transmembrane helix

Introduction

Hepatitis C virus (HCV) is estimated to have infected at least 170 million people worldwide and is a major cause of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma [1]. Until recently, experimental studies on HCV were limited due to lack of efficient cell culture systems for the virus amplification. However, this situation has changed with the development of novel *in vitro* systems [2], particularly the HCV pseudoparticles [3] (HCVpp) and the first system for efficient production of infectious viral particles in cell culture [4] (HCVcc).

HCV is the only member of the *Hepacivirus* genus which belongs to evolutionary related viruses of the *Flaviviridae* family [5, 6]. The virus genome contains a long open reading frame of more than 9600 nucleotides that is translated into a single polyprotein of approximately 3000 amino acids length [7]. The open reading frame between the 5'-non coding region (NCR) and 3'-NCR is composed of the structural core protein and the two envelope glycoproteins E1 and E2, the p7 ion channel and at least six non-structural proteins.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0672-1) contains supplementary material, which is available to authorized users.

S. A. Jusoh · V. Helms (✉)
Center for Bioinformatics, Saarland University,
66041 Saarbruecken, Germany
e-mail: volkhard.helms@bioinformatik.uni-saarland.de

S. A. Jusoh
Faculty of Pharmacy, Universiti Teknologi MARA,
40450 Shah Alam, Malaysia

C. Welsch
Department of Internal Medicine I,
Johann Wolfgang Goethe-University Hospital,
60590 Frankfurt am Main, Germany

S. W. I. Siu · R. A. Böckmann
Computational Biology, Universität Erlangen-Nürnberg,
BTE Gebäude,
Erwin-Rommel-Str.3,
91058 Erlangen, Germany

Binding and internalization of the hepatitis C virus are essential steps in the viral replication cycle mediated by the envelope glycoproteins E1 and E2. The E1 and E2 proteins are released by host signal peptidase cleavages [6] and assemble as a non-covalent E1–E2 heterodimer which is retained in the endoplasmic reticulum (ER) [8]. These two membrane proteins are type I transmembrane (TM) proteins which are composed of a large N-terminal ectodomain toward the ER lumen and a C-terminal hydrophobic anchor. The membrane-spanning segments for both E1 and E2 are located at the C-termini and predicted to be less than 30 amino acids long with two stretches of hydrophobic residues separated by a short polar segment with at least one highly conserved charged residue [9]. Interestingly, two consecutive GxxxG motifs are known within the TM of E1. The presence of the GxxxG motif in glycoporphin A (GpA), a membrane protein of erythrocytes, at the helix-helix packing interface is known to be involved in the GpA homodimerization [10]. Experimental studies demonstrated that the TM domains of E1 and E2 are not just membrane anchors, but play important multifunctional roles during the biogenesis of HCV [8, 11], *e.g.*, virus entry [11], ER retention, as an internal signal peptide and E1–E2 heterodimerization [12].

In particular, Gly354, Gly358, and the conserved charged residues in the TM region Lys370, Asp728, and Arg730 were shown to be involved in E1–E2 heterodimerization. Different experiments, mutagenesis studies of alanine scanning insertion, site-directed mutagenesis and tryptophan replacement suggest a salt-bridge interaction between Lys370 and Asp728 at the helix-helix dimer interface, which strongly contributes to the E1–E2 heterodimerization [11, 13]. The charged residues in the TM domain of E1 and E2 glycoproteins of bovine viral diarrhea virus (BVDV) were also claimed to be responsible for the heterodimerization [14]. So far, this hypothesis has not yet been confirmed by structure determination methods. In contrast to these residues, Arg730 was shown to play a minor role for the assembly of the E1–E2 envelope glycoprotein [11].

Despite their relative abundance in the protein-coding regions of different genomes (20–25%), only a few high-resolution structures of membrane proteins could be determined so far due to the difficulty of membrane protein crystallization in the lipid bilayer environment [15]. Yet, molecular dynamics (MD) simulations of membrane proteins embedded in lipid bilayers have become quite popular and successful in the last ten years [10, 16]. In particular, MD simulations were applied to study the spontaneous aggregation of phospholipids around membrane proteins [17] or have been used to investigate the relative position of individual transmembrane helices in lipid bilayers [18] and their dynamic interactions with phospholipid bilayers [19].

For example, the structure of the Glycoprotein A (GpA) dimer was computationally predicted [20], including results from an extensive mutagenesis work [21] to narrow the search. The prediction was later refined, using an improved global search method [22]. The subsequently determined NMR structure of the GpA dimer in micelles [23] was in good agreement with the predicted structure. Furthermore, MD simulations were used to study the behavior of individual helices of bacteriorhodopsin [24], the oligomerization of the helices of Vpu [25, 26], the free energy for dimerization of GpA [27], and the protonation equilibrium of Arg residues within a TM helix [28].

The principal aim of the present study was to identify critical regions and crucial residues within HCV envelope proteins for the formation of the E1–E2 heterodimer. Thus, we performed atomistic MD simulations for the putative TM domain of the E1–E2 heterodimer from HCV. Our results provide, for the first time, an atomic structural and dynamic model for the TM domain of the E1–E2 heterodimer. The simulations reveal the importance of the ion-pair interaction and of additional *inter*-helical hydrogen bonds in the middle of the helix interfacial region for the structural integrity of the heterodimer. Furthermore, we confirmed the locations of the conserved residues which are in good agreement with the experimental studies.

Methods

Sequences

The protein sequences used for MD simulations of E1 and E2 from the hepatitis C virus genome polyprotein were obtained from the UniProtKB/Swiss-Prot database (<http://au.expasy.org/uniprot/>) [29]. The E1 sequence used in this study is G³⁵⁰AHWGVLGIA³⁶⁰YFSMVGWAK³⁷⁰VLVLLLLFAG³⁸⁰VDA. The E2 sequence is WAIKWEYVV⁷²⁰LLFLLLDAR⁷³⁰VCSCLWMLL⁷⁴⁰ISQAEA. Both sequences are from HCV genotype 1a.

We also used test segments, named H-segments, which were used to study apparent membrane-transfer free energies of each of the 20 naturally occurring amino acids [30]. The H-segments were prepared as ideal α -helices, which contained a charged amino acid in the middle of their TM helix. MD simulations of the H-segments were compared to the results of E1 and E2 monomer simulations (see below).

Sequence analysis

Sequences of HCV envelope proteins were retrieved from public HCV databases, UniProtKB and euHCVdb (<http://www.euHCVdb.de>) [31]. HCV genotypes have been differentiated according to a consensus proposal for a

unified system of HCV genotype nomenclature [32]. Sequence alignments were computed using CLUSTAL W [33] and MUSCLE [34], and subsequently improved by minor manual modifications using the SEAVIEW alignment editor [35]. A comprehensive sequence analysis was performed in 604 HCV E1 sequences (HCV genotype 1: 476, other genotypes: 128) and in 569 HCV E2 sequences (HCV genotype 1: 444; other genotypes: 125). We deduced amino acid polymorphisms in the E1 and E2 TM domains including all sites associated with E1–E2 heterodimerization investigated in this study.

TM protein prediction

Five prediction methods for helical membrane proteins were employed to determine the start and end points of the E1 and E2 TM regions: PHDHTM [36], SPLIT4 [37], HMMTOP2.0 [38], TMHMM [39], and TMMOD [40]. MINS2 [41] was used to predict the membrane insertion free energy of the TM domains of E1 and E2.

MD simulations

All structures used in this study were prepared as ideal α -helices. The SCWRL program [42] was used to position the side chain rotamers and to generate mutants. Gromacs [43] tools were used to set up paralleled dimers aligned along the membrane normal with a salt bridge interaction at their helix-helix interfaces. In this conformation, the side chains of the charged residues were within 5 to 6 Å distance to each other (see Table 1) and Asn367 forms an *inter*-helical hydrogen bond. MD simulations of the E1–E2 heterodimer were done twice and each simulation was assigned different starting velocities.

A snapshot of a fully hydrated equilibrated lipid bilayer containing 128 DMPC lipids [44] solvated with 5,673 simple point charge (SPC) water molecules was used as a starting point for all MD simulations. A cavity within the bilayer was created using the protocols of reference [45]. The solvent-accessible protein surfaces of the peptides were calculated by the program MSMS [46] using a probe size radius of 1.4 Å. The solvent-accessible surfaces of the peptides were used as templates for estimating the volume of the necessary cavity. In each case, 4–6 lipids in the center of the projected hole were removed to avoid overlaps of lipids with the protein. 200 ps of simulation with a modified version of the Gromacs version 3.1.4 [47] were performed to create the protein cavity in the DMPC lipid bilayer. Each peptide sequence was embedded into the DMPC bilayer using a cavity of suitable size. The mixed protein-lipid bilayer system was surrounded by approx. 45 water molecules per lipid molecule, thus ensuring full hydration of the membrane [48]. The protein/lipid/water

system was then subjected to 500 steps of energy minimization using the steepest descent algorithm. Ions (Na^+ and Cl^-) were added to neutralize the system and to achieve close-to-physiological conditions at ~ 0.1 M NaCl. This was followed by a 200 ps MD run with harmonic position restraints (force constant 1000 kJ mol^{-2}) applied to all heavy atoms of the protein. This procedure allowed the lipids and the water molecules to relax around the protein after insertion of the protein. Subsequently, fully unrestrained production runs of 100 ns duration were performed for the protein/lipid systems.

All simulations were performed using the Gromacs 4.0.3 package [43]. All monomer and dimer simulations were performed with united atom force field based on GRO-MOS96 (53a6) [49] for the peptides and the Berger force field [50, 51] for the phospholipids. Periodic boundary conditions were used in all directions. The system was coupled to a temperature bath at 310 K separately for the protein, the lipids, and the water/ions with a coupling constant of 0.1 ps^{-1} [52]. For the pressure, semi-isotropic coupling was employed separately for the lateral and for the normal directions with a coupling time $\tau_p = 1 \text{ ps}$. The compressibility was set to $4.5 \times 10^{-5} \text{ bar}^{-1}$. Covalent bonds to H-atoms were constrained using the LINCS algorithm [53] and an integration step size of 2 fs was used. The non-bonded pair list was generated every 10 steps with a cutoff of 1.0 nm. For short range van der Waals interactions, a cutoff distance of 1.0 nm was used. The long-range electrostatics interactions were treated using the Particle-Mesh Ewald method with a real space cutoff of 1.0 nm and a grid spacing of 0.12 nm with cubic interpolation.

Analyses of the trajectories were primarily performed with tools included in the Gromacs 4.0.3 suite [43, 47]. Root mean square deviation (RMSD) analyses were based on atoms of the protein backbone. Salt bridge contacts were defined by monitoring the average distance between the side chains (see Table 1). Helix centers of mass were computed using the coordinates of $\text{C}\alpha$ atoms only for the segments 5–25 (E1) and 35–55 (E2). All images in this work were prepared with the Pymol program (pymol.sourceforge.net).

Results

Sequence analysis of the TM domains of E1 and E2

The conserved residues Gly354, Gly358, Lys370, Asp728, and Arg730 were predicted to be located in the TM region of E1 and E2 in HCV genotype 1a from the UNIPROT database. A comprehensive sequence analysis was carried out to investigate the natural polymorphisms occurring at these particular amino acid sites.

Table 1 Salt-bridges between E1–E2 wild-types and mutants. Given are average values for the data between 80 and 100 ns of the MD simulations

Wild-types & mutants of E1–E2 heterodimers	Interacting residues	Interacting atoms	Average salt bridge distances (nm)
Wild type 1	Lys370...Asp728	NZ:HZ...OD:CG	0.31±0.03
Wild type 2	Lys370...Asp728	NZ:HZ...OD:CG	0.32±0.02
Mutant R730K	Lys370...Asp728	NZ:HZ...OD:CG	0.30±0.03
Mutant G354A & G358A	Lys370...Asp728	NZ:HZ...OD:CG	0.30±0.03
Mutant K370R	Arg370...Asp728	NH:HH...OD:CG	0.34±0.05
Mutant D728E	Lys370...Glu728	NZ:HZ...OE:CD	0.27±0.01

• A salt-bridge distance is calculated by averaging the distances between the hydrogen and oxygen atoms from an amine/carboxyl group of E1 and the carboxyl group of E2

• Only mutants with a predicted salt-bridge at the helix-helix interface are calculated

We found the Lys370 in E1 being only once replaced by Arg in HCV genotype 1. All other genotypes investigated showed no polymorphism at this site in E1. The residues Asp728 and Arg730 are highly conserved in HCV E2 genotype 1. We found a non-conservative polymorphism only at position 728. The polar residue Asp was replaced by the aromatic and non-polar residue Tyr. Again no polymorphism at 728 or 730 was found in genotypes 2, 3 or 5, whereas Gly728 and Lys730 were found in genotype 4 once respectively, and Val728 and His730 in genotype 6 once respectively. Overall, polymorphisms at Lys370, Asp728 and Arg730 have been observed only exceptionally.

Two consecutive GxxxG motifs are present in the TM segment of E1. Gly350 and Gly354 were found to be highly conserved in all genotypes investigated. The second motif showed the conservative polymorphism Gly358Ala in genotypes 1, 5 and 6. Gly358 was conserved in genotypes 2, 3, and 4. Genotype 1 showed an Ala twice at 358. Only Ala358 was found in genotype 5. Genotype 6 showed Ala358 in the majority of sequences investigated (30 over 43).

Identification of TM residues by secondary structure prediction methods

We used five different methods for secondary structure prediction of the TM domains of E1 and E2 (see Fig. S1). This gave predicted TM helices of 21 to 31 amino acids length for the TM domain of E1. The consensus segment predicted by at least three out of five methods ranges from Val355 to Ala379. For the TM domain of the E2 glycoprotein, the consensus segment assigned by at least three methods ranges from Tyr718 to Ser742. Interestingly, all methods placed the conserved charged residues Lys370, Asp728, and Arg730 in the middle part of the TM domains.

The MINS2 [54] method that is based on amino acid frequencies and calibrated against the dataset of Hessa et al. (2007) was applied to compute membrane insertion free

energies of TM segments. Compared to the threshold of 3.5 kcal mol⁻¹ for observed TM helices in known structures of helical membrane proteins, MINS2 gives a favorable insertion free energy for the isolated TM segments of E1 (1.8 kcal mol⁻¹) when using Lys370 as center, and a borderline value of 4.3 kcal mol⁻¹ for E2 (4.3 kcal mol⁻¹) when using Asp728 as center.

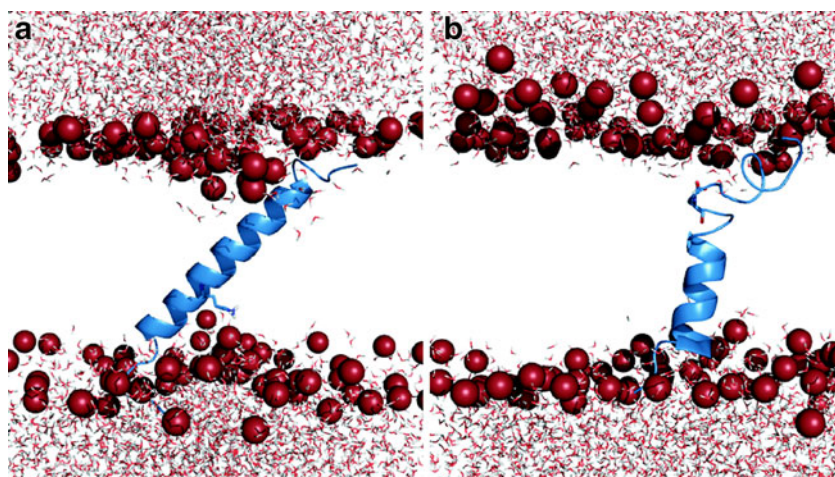
MD simulations of E1 and E2 monomers

MD simulations were carried out to investigate the behavior of the monomeric TM segments containing a charged residue in the middle of the helices. We observed that during the 100 ns MD simulations the charged residues Lys370 from E1 and Asp728 from E2 were attracted toward the lipid bilayer interface. Only Lys370 was able to comfortably anchor to the interfacial region without affecting the helix stability (Fig. 1). Due to its shorter side chain, Asp728, which is positioned in the center of the TM domain of E2, was not able to anchor to the interface region. Moreover, its strained conformation led to disruption of the α -helical conformation of the N-terminal half of the E2 monomer. In contrast the E1 helix segment was stable along the simulation time. Analogous simulations of H-segment monomers containing Lys and Asp amino acids in the middle of the TM segments, respectively, gave similar results (see Fig. S2) thus confirming our observations and providing further evidence that they are caused by the charged Lys370 and Asp728 residues.

MD simulations of E1–E2 heterodimers

Fig. 2 shows the consensus TM assignment based on the analysis of all MD simulations. Both the MD simulations of the individual helices and of the heterodimers indicate that the TM domain of E1 consists of 29 residues ranging from Gly354 to Gly380 (G³⁵⁴VLGIA³⁶⁰YFSMVGWAK³⁷⁰VLVLLLLFAG³⁸⁰VD). The E2 TM domain was observed

Fig. 1 Final snapshots of MD simulations from the E1 and E2 monomers (a) E1 TM segment with charged Lys370 and (b) E2 TM segment with charged Asp728. Lipid tails and ions are not shown for clarity. The charged Lys370 and Asp728 are shown as stick representation



to contain 27 residues between two polar residues at both N- and C-termini (EYVV⁷²⁰ LFLLLADAR⁷³⁰ VCSCLLWMLL⁷⁴⁰ ISQ). These are Glu717 and Tyr718 at the N-terminus and Ser742 and Gln743 within the C-terminal region. The consensus from the secondary structure prediction methods agrees closely with the consensus of the MD simulations sequences (Fig. 2).

To investigate the dynamics of the TM domain of the modeled E1–E2 dimer structure (see methods section), we performed two MD simulations of the E1–E2 wild type heterodimer with different starting velocities. These were named WT1 and WT2 in the subsequent tables and figures. Both simulations resulted in similar stable final E1–E2 conformations (Fig. S4). Fig. S3 shows root-mean-square deviations (RMSDs) of each monomer in the simulations of the E1–E2 heterodimer with respect to the perpendicular starting conformation. The RMSD values of the entire structures stabilize after around 20 ns between 0.5 and 0.7 nm which is mainly due to a tilting motion of one peptide with respect to its initial perpendicular orientation in order to find an optimal position in the membrane environment. The tilting motion observed matches with the fact that secondary structure prediction assigned TM segments of 25 to 30 residue length (see above). RMSD analyses also indicate that both simulations of the E1–E2 heterodimer showed smaller fluctuations than the simulations of the E1 and E2 monomers.

The heterodimerization was clearly mediated by the salt bridge interaction of the charged Lys370 and Asp728 at the helix-helix interface. Table 1 shows average distances between the functional groups (atoms Lys370-NZ:HZ and Asp728-CG:OD) to measure the stability of the Lys-Asp ion-pair. The distance was found to be stable at 0.30–0.32 nm in the WT1 and WT2 simulations. Due to the helical periodicity, Arg730, being two positions away from the central Asp728, pointed into the opposite position and faced the hydrophobic lipid bilayer to anchor to the lipids polar interface.

Apart from formation of the central ion pair, we also observed formation of additional *inter*-helical H-bonds (see Table S1). This appears to be a novel finding related to the formation of the E1–E2 dimer. For the wild-type, about 1±0.4 H-bonds are formed between Asn367 and Asp728.

Mutational analysis

MD simulations of E1–E2 single and double mutants were carried out to analyze the naturally occurring polymorphisms and to confirm the contributions of the conserved amino acids of the E1 and E2 TM segments. All three single mutants with a salt-bridge (R730K, K370R, and D728E) were set-up independently as for wild-type and maintained stable heterodimers during the simulations (Fig. 3) as for wild-type that are stabilized by an ion-pair interaction when started from a salt-bridged conformation. This behavior can be expected

Fig. 2 TM residues of E1 and E2 resulting from 100 ns of MD simulations are compared to the results of secondary structure prediction methods. The consensus prediction resulting from the MD simulations and secondary structure methods are given at the bottom

	360	370	380	720	730	740
E1-E2	GVL AGIAYFSMVG N AKVLV V LLLFAG V DA			EYV LLFLLADAR V CS L IN M LLIS Q AE A		
WILDTYPE 1	*****			*****		
WILDTYPE 2	*****			*****		
R730K MUTANT	*****			***** R *****		
D728E MUTANT	*****			***** E *****		
K370R MUTANT	***** R *****			*****		
K370A MUTANT	***** A *****			*****		
N367L & K370L MUTANT	***** L * L *****			*****		
D728L & R730L MUTANT	*****			***** L * L *****		
G354A & G358A MUTANT	A *** A *****			*****		
Consensus						
MD Simulations	GVL A GIAYFSMVG N AKVLV V LLLFAG V D-			EYV L LLFLLADAR V CS L IN M LLIS Q ---		
TM Prediction Methods	-V L AGIAYFSMVG N AKVLV V LLLFAG ---			-YV V LLFLLADAR V CS L IN M LLIS ---		

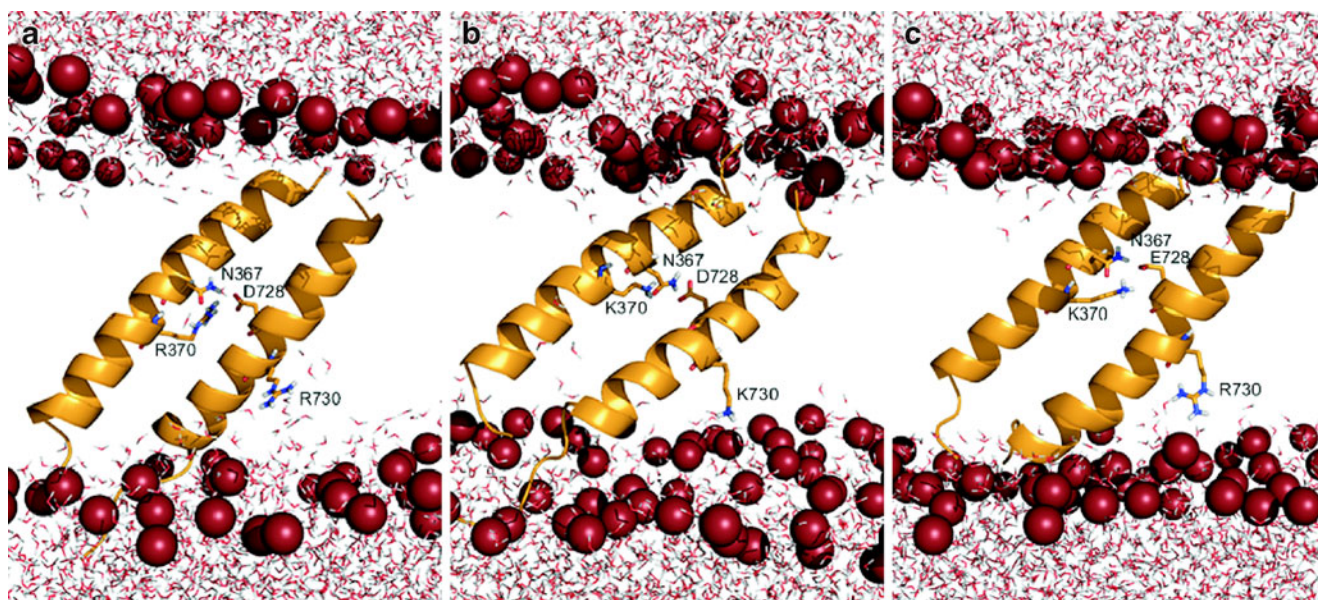


Fig. 3 Final snapshots of MD simulations from single mutant dimers with a salt-bridge. **(a)** K370R mutant, **(b)** R730K mutant and **(c)** D728E mutant. The conserved residues Asn367, Lys370, Asp728,

Arg730 and mutated residues are shown as stick representation. Lipid tails and ions are not shown for clarity

due to the conservative nature of the mutation. The charged residues of all single mutants with an ion-pair interaction were in close atomic contact as for the wild-type (0.27–0.34 nm distance) (Table 1). Also, the RMSD values are of similar magnitude than those of the wild-type simulations (Fig. S3). Interestingly, despite having a longer side chain than Asp, the replaced Glu residue of the D728E mutant showed the shortest average distance (0.27 nm). The longer side chain of Glu apparently allows for an optimal contact with the Lys730 side chain. On the other hand, the K370R mutant had the largest average salt-bridge distances (0.34 nm), which may be caused by the long and bulky side chain of the mutated Arg.

The MD simulations of the three double mutants (N367L & K370L, D728L & R730L, G354A & G358A) resulted in different conformations with intact TM helices (Fig. 3 and Fig. 4). The largest structural fluctuations compared to the starting structure were observed for the D728L & R730L mutant (Fig. S3c). Mutating the conserved residues N367 and K370 in E1, and D728 and R730 in E2 led to a partial separation of the two helices, see $C\alpha$ distances in Table 2 and Fig. 4. However, the double mutant G354A & G358A was as stable as the wild-type during the simulation (Fig. 4c), probably due to the presence of the salt-bridge interaction.

The tight heterodimerization of the ion-pair stabilized helix dimers is also reflected by a close distance between the centers of mass of the two helices (Table 2). All single mutants with a salt-bridge (K370R, R730K and D728E), and the G354A & G358A double mutant showed close distances (1.08–1.23 nm) as the two wild type simulations (1.08–1.10 nm).

On the other hand, the double mutants (N367L & K370L and D728L & R730L) showed much larger separations (1.31–1.76 nm) reflecting the absence of an ion pair interaction or of other stabilizing *inter*-helical interactions (Table 2 and Fig. 4). Most simulations showed tilting angles of the two helices around 40° to 60°. The only exception is the E1 monomer in the double mutant D728L & R730L that is almost straight (13°) as this mutant dissociated.

As discussed before, additional *inter*-helical H-bonding was observed to stabilize the helix dimer and prevent helix kinking or partial unfolding. For the mutants with a salt bridge (R730K and G354A & G358A) a similar average number of *inter*-helical H-bonds was found as for wild-type (0.91–1.02) (Fig. 5). The highest average number of *inter*-helical H-bonds resulted from the D728E mutant (1.86) followed by the K370R mutant (1.30) indicating more favorable contacts.

To clarify the function of the salt bridge interaction at the helix-helix interface, we mutated Lys370 to Ala which removes the ability to form an ion-pair between the helix monomers. Interestingly, even in the absence of an ion-pair interaction, the K370A mutant was heterodimerized during the simulation. The distance between the helix monomers is the closest one found (0.97 nm) and the average number of H-bond interactions between both monomers was similar to the other heterodimerized conformations (1.21 nm) (Table S1 and Fig. 5). We found that Asp728 made very stable contacts with Arg730 so that Arg730 turned around and was now located at the helix-helix interface. Arg730 then formed an H-bond with Asn367 with 72% occupancy.

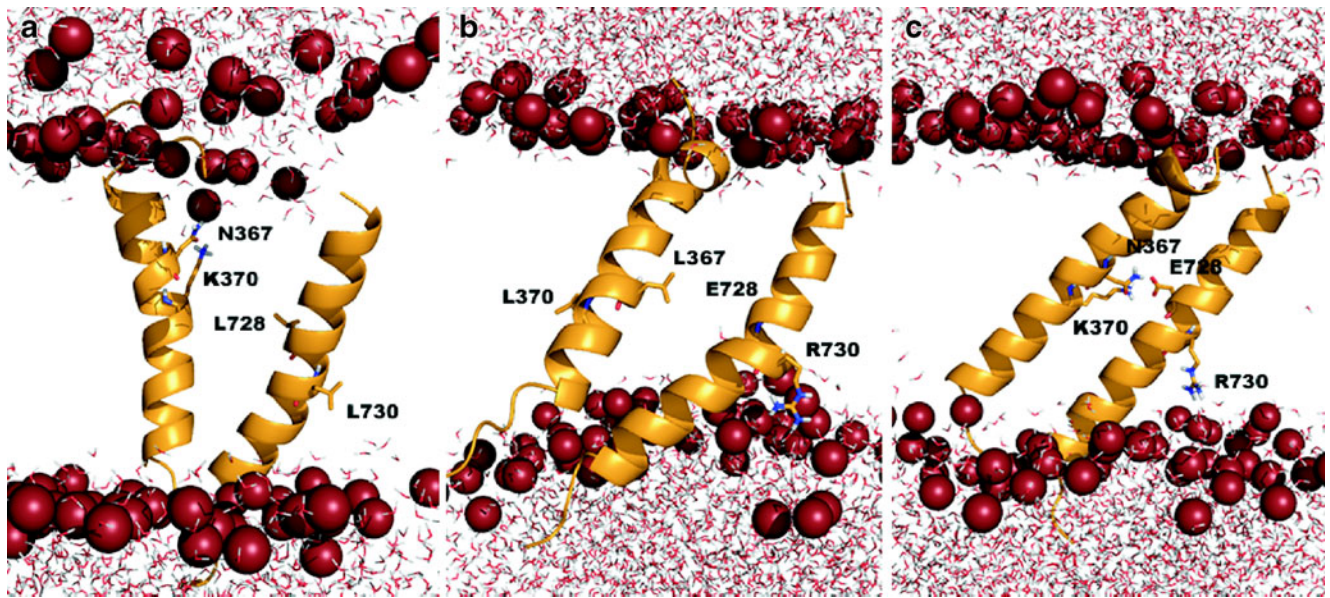


Fig. 4 Final snapshots of MD simulations from the double mutant dimers. (a) D728L & R730L, (b) N367L & K370L and (c) G354A & G358A. The conserved residues Asn367, Lys370, Asp728, Arg730

and the replaced Leu are shown as stick representation. Lipid tails and ions are not shown for clarity

This behavior caused local unfolding in the center of the E2 of K370A mutant (Fig. S4). On the other hand, if Asp728 is mutated into Ala, no rotation of Arg730 in E2 was observed (Fig. S5).

Comparison of MD structures versus the NMR structure

As the only NMR structure available for the HCV envelope glycoproteins is a segment of E1 consisting of 21 residues

[55] with the PDB-code 1EMZ.pdb, RMSD analyses were done on the same segment during the MD simulations with respect to the NMR structure (Fig. 6). We compared the RMSD of the backbone atoms of the E1 TM segment (Gly350–Lys370). The central part (354–370) formed a well defined α -helix in the simulation. The average conformation from residues 359–367 in the simulation of the E1–E2 dimer has an RMSD of 0.06 nm compared to the NMR structure (Fig. 6), whereas the RMSD of the

Table 2 Structural parameters for the data between 80 and 100 ns of the MD simulations of E1–E2 wild-type heterodimer and E1–E2 mutants

Wild-type E1–E2	CA distances between E1 and E2 helices (nm)	TM tilt angles E1 (°)	TM tilt angles E2 (°)
Wild-type 1 (K...D)	1.10±0.03	41.3±5.1	40.8±4.7
Wild-type 2 (K...D)	1.08±0.02	47.3±4.0	50.2±4.1
Single mutants with a salt-bridge			
R730K (K...D)	1.08±0.03	50.0±3.7	67.2±1.4
K370R (R...D)	1.17±0.03	59.5±2.2	42.8±4.1
D728E (K...E)	1.23±0.03	41.0±3.1	41.9±1.7
Single mutants without salt-bridge			
K370A	0.97±0.02	60.0±3.0	Kinked
D728A	1.07±0.03	40.4±4.1	Kinked
Double mutants			
D728L & R730L	1.76±0.04	13.3±4.2	43.3±4.1
N367L & K370L	1.31±0.05	47.7±3.9	52.0±5.2
G354A & G358A (K...D)	1.11±0.04	45.7±4.0	45.1±3.7

• Average number of H-bonds per time frame. A H-bond (D–H...A) is defined to have a H...A distance less than 3.5 Å and a D–H...A angle greater than 120°

• Interacting residues at the helix-helix packing are shown for each dimer

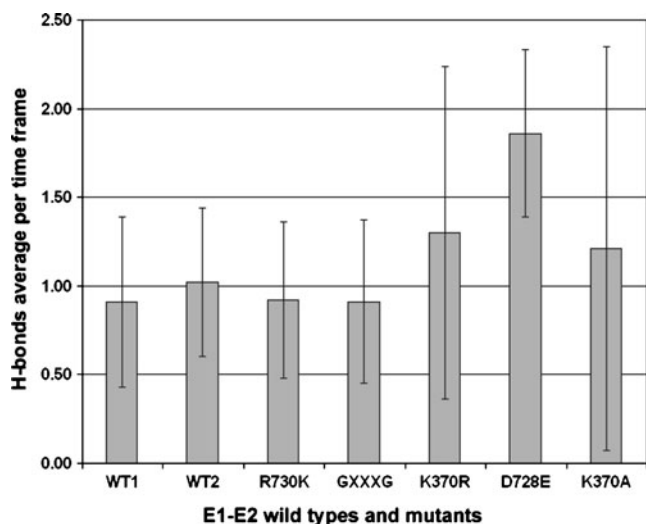
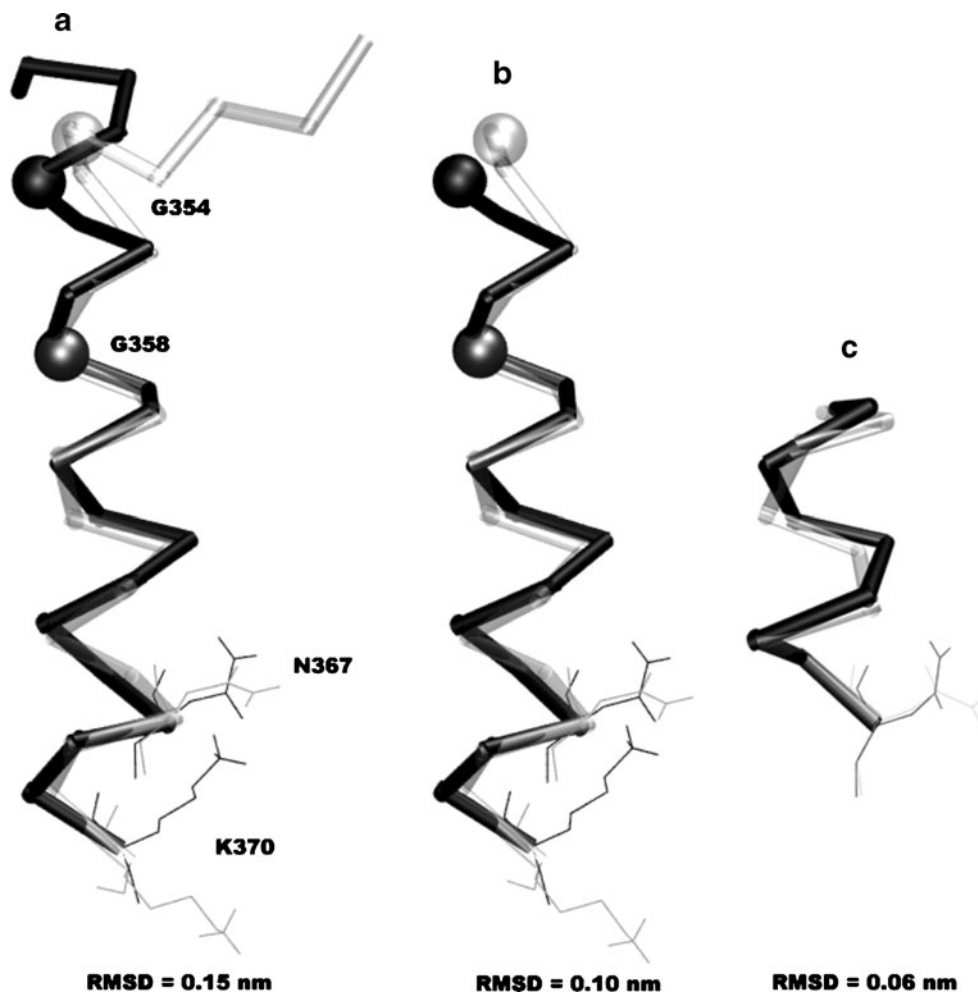


Fig. 5 Inter-helical H-bond interactions for the E1–E2 wild-types and mutants

structural ensemble derived from NOE restraints was 0.03 nm [55]. Although we found a somehow larger RMSD of the 21 residue segment (Gly350 to Lys370) of 0.15 nm,

Fig. 6 Superimposition of the E1 segment from the E1–E2 heterodimer wild-type to the NMR structure, 1EMZ.pdb. Coloring scheme: Black-1EMZ.pdb; Grey-Segment of E1 from the simulation of the E1–E2 heterodimer model. RMSD values are listed below figures. (a) Segment consists of residues G350 to K370. (b) Segment from G354 to K370. Conserved residues G354, G358, N367 and K370 are highlighted as wire frame presentation. (c) Segment from Ile 359 to Asn 367



these deviations are still smaller than the variation within the NMR ensemble of 24 structures (0.29 nm) [55]. Thus, the E1 helical conformation derived from MD simulations is quite similar to the conformation determined by NMR in trifluoroethanol (TFE) [55]. As expected, the largest RMS fluctuations were observed for the residues at the helix ends near the membrane bilayer interface. Although previous NMR studies did not show the segment Gly354 to Gly358 to be α -helical, we found that this segment is in stable α -helical conformations in the MD simulations on the investigated timescale.

The segment between Gly354 and Gly358 was observed to be in the TM region during the MD simulations (Fig. 2), but the GxxxG motif was not located at the helix-helix interface (see discussion below).

Discussion

In the viral *Flaviviridae* family [9], at least one positively charged residue is highly conserved in both putative TM domains of the envelope glycoproteins. Polymorphism

analysis of the conserved residues G354, G558, Lys370, Asp728, and Arg730 in all HCV genotypes indicates that mutations rarely occur at these particular sites. The data analyzed in this study confirms previous findings [9, 11] that these conserved residues are crucial for the viral specific functions of the E1 and E2 envelope proteins.

Having a charged residue in the middle of a TM domain would be energetically unfavorable for an isolated α -helix. Instead, these residues would probably appear neutral by shifting their pKa values at an energetic expense or they try to position their charged side-chain into the polar head-group region [28]. In spite of these considerations, an experimental study by Hessa et al. [30] proved that single transmembrane segments with a polar or charged residue in the middle of the domain were able to be inserted as membrane proteins *via* Sec61 translocon.

Simulation of monomers

The MD simulations of helix monomers revealed that the charged Lys370 and Asp728 had different effects on the TM segments of E1 and E2 monomers, respectively, if they were placed as isolated helices in a membrane lipid bilayer. The TM segment of the E1 helix was stable during the simulation, whereas the N-terminal half of the TM segment of E2 was disrupted, possibly due to the shorter side chain of Asp730. Subsequent MD simulations of H-segments containing a charged Lys or Asp showed a similar behavior. Asp residues were previously shown to induce stronger distortions in α -helices compared to basic residues [56]. Moreover, Hessa et al. [30] found that the biological apparent insertion free energy scale showed the highest value (3.49 kcal mol⁻¹) when Asp was placed in the middle of the TM domain of the H-segment compared to other amino acids.

Simulation of E1–E2 wild type

Dubuisson and co-workers suggested that the E1 and E2 TM helices are inserted cooperatively into the lipid bilayer based on mutagenesis results [9, 11, 13]. Here, we put this hypothesis on stable energetic and structural grounds based on extensive MD simulations of wild-type and mutant heterodimers. Indeed, favorable salt-bridge and H-bonding interactions between the TM segments of E1 and E2 contribute to stabilization of the dimer conformation in lipid bilayers. As mentioned above, the E2 monomer containing the charged Asp728 unfolded partially during MD simulations. However, when simulated as part of the E1–E2 heterodimer, the E2 maintained its stable α -helical structure. This is a strong indication that the dimer conformation of the E1 and E2 envelope glycoproteins is a favorable arrangement even in a hydrophobic environ-

ment. If the ion-pair of Lys370 and Asp728 at the helix-helix interface is already established in the translocon or near to its exit, as suggested before [12], this should facilitate the entry of the E1–E2 heterodimer into the lipid bilayer environment [12]. Moreover, the stability of this ion-pair interaction may serve as a kinetic barrier against the E1–E2 heterodimer dissociation. This role is in agreement with the suggested function of one or more hydrophilic residues which were observed in other TM domains to be responsible for the ER retention [57]. Since Lys370 and Asp728 were located at the helix-helix interfacial region in our model, Arg730 was positioned oppositely where it faced the lipid tails. To optimize its position, the positively charged side chain of Arg730 oriented its guanidinium group toward the polar region of the lipid bilayer. Molecular simulations previously showed that Arg adjust energetically in the membrane environment and its long side chain is likely to remain positively charged in lipid bilayers [28]. Also, the charged Arg residues in the voltage sensor domain of potassium channels behaved such that the Arg residues were stabilized by the polar head groups of lipids and water molecules [58].

The atomistic observation from the MD trajectories also reveals a so far unreported *inter*-helical H-bond contributed by Asn367 which also contributes to stabilize the structure of the E1–E2 heterodimer. *Inter*-helical H-bonds are known to be of particular importance for the formation of secondary or tertiary structure in the hydrophobic membrane center with low dielectric environment [59]. A recent report from von Heijne and co-workers also demonstrated that engineered TM domains with *inter*-helical interactions mediated by polar residues are more efficiently inserted into the lipid bilayer [60].

We now discuss the relevance of the sampled dimer conformations. At the start of each simulation, the two helices were arranged parallel to the membrane normal with the ion-pairing residues facing each other. The simulations then showed that these initial orientations are stable on the time scale of the simulations which supports the experimental finding that the E1–E2 helices are inserted by the translocon with the ion-pair already formed. A situation of an E1–E2-dimer with one or both of the helices turned by 90 or 180 degrees, for example, likely never occurs in nature. However, as we clearly did not sample the range of possible orientations, we cannot address whether the generated models correspond to the thermodynamically most favorable orientation of the isolated E1–E2 helices. That would require sampling a large range of orientations over long simulation times which is currently infeasible by plain MD simulations in explicit bilayers. The simulated 100 ns time scale is clearly not sufficient for entire helices to turn around their axis in a lipid bilayer. As an alternative, using an implicit-solvent representation of the membrane

[61–67] would allow for a more complete sampling and for faster orientational relaxation. Besides, replica-exchange simulations allow to speed up the penetration into membranes and re-orientations [68]. However, it is not clear from experiment which is the thermodynamically most favorable state of the two isolated helices because the experiments [9, 11, 12] were always performed on the full E1–E2 proteins with the external domains present. So it is in fact possible that dimerization is only stable with the external domains present.

Simulations of heterodimer mutants

To further confirm the location of the important residues of the TM domains of E1 and E2, we carried out simulations of several conservative mutants. The single mutants R730K, K370R and D728E, which contain a salt bridge between the helices, were all heterodimerized during the simulations with stable α -helical conformations (Fig. 3). These results strongly support that the ion pair interaction between the particular charged amino acids is responsible for the *inter*-helix interaction. The distances between the charged side chains of these single mutants are similar to the E1–E2 wild-type (0.27–0.34 nm). These results are in perfect agreement with experimental findings which reported that R730K and D728E mutants form heterodimers similarly to the E1–E2 wild-type [13]. Our analysis of natural polymorphisms indicates that the R730K mutant occurred once in genotype 4 of HCV which is very rare. On the other hand, the R730K mutants only resulted in a slightly reduced incorporation and infectivity of E1–E2 proteins into HCVpp compared to the E1–E2 wild-types. Here, Lys led to a similar dimer conformation than with Arg730 since both are positively charged amino acids. We note, however, that in this structural model, with a salt-bridge stabilized heterodimer, Arg730 is not located at the helix-helix interface and its mutation should not affect dimerization. In contrast, the infectivity of the D728E mutant was strongly reduced, however without affecting the formation of heterodimers [11, 13]. This indicates that even conservative mutations that can be expected to maintain the salt-bridge interaction may lead to different biological function such as viral entry. One may therefore speculate that placing the longer side chain of Glu between the two helices may affect the helical packing although this is not apparent in the simulations.

In a second set of mutant simulations, we mutated Lys370 to Ala to investigate the effect of removing the salt-bridge on the E1–E2 heterodimerization. Interestingly, the K370A mutant still managed to remain heterodimerized during the simulation. Arg730 turned around to interact with Asn367 so that the average number of H-bonds between the E1–E2 helices increased compared to the

wild-type. *Inter*-helix H-bonding of polar amino acids was recently studied experimentally by systematically constructing H-segment dimers [60]. This work concluded that polar *inter*-helix interactions increase the translocon insertion efficiency of both helices. However, this rotation of Arg730 caused severe rearrangements of the backbone conformation in the central part of the E2 TM helix. In the experimental setting, mutation of Lys370 led to reduced heterodimerization to about 50% [11]. On the other hand, mutation of Asp728 severely reduced the E1–E2 heterodimer biogenesis to about 10 to 20% when replaced with hydrophobic amino acids such as Leu, Ala or Trp [11, 13, 55].

In a third, final set of double mutants, the central residues at 367 and 370 or 728 and 730 were replaced by leucine residues. Both double mutants resulted in significantly enlarged distances between the TM helix monomers compared to the wild-type and to the mutants containing a salt-bridge. The E2 double mutant D728L/R730L residue gave a larger average distance (1.76 Å) between the helix dimer than the E1 double mutant N367L/K370L (1.31 Å). Interestingly, these results are again in line with the experimental study, which reported a differential effect of both double mutations [12]. For soluble proteins, there exist several computational methods that can qualitatively predict the effect of protein mutations on their stability [69, 70]. It is certainly feasible to transfer these methods to the area of TM proteins. Up to now, however there is a lack of quantitative experimental data on the thermodynamics stability of TM helix bundles and respective mutants against which such computational methods can be calibrated.

GxxxG motif

For the GpA homodimer, the GxxxG motif at the helix dimer interface has been shown to play an important role for the homodimerization [71, 72]. Also for the E1–E2 heterodimer, mutating either Gly354 or Gly358 impaired the E1–E2 assembly [11]. In the structural model of the E1–E2 heterodimer developed in this study, however, the Gly350, Gly354, and Gly358 residues are not located at the helix interfacial region. Therefore, we did not observe any possible interaction between the GxxxG motif of E1 and the residues from the TM domain of E2. However, this does not exclude the probability of GxxxG segments to heterodimerize at the ectodomain region of the E2 glycoprotein. The E1 helix conformation agrees nicely with an experimental structure of E1 solvated in TFE. Whereas the NMR analysis revealed an unwinding of the N-terminal end of the E1 helix between Gly354 and Gly358, this region stayed intact in an α -helical conformation during the heterodimer simulations.

Conclusions

This study puts the assignment of the TM domains of E1 and E2 on a firm basis. The structural model explains the roles of the highly conserved positively and negatively charged residues in the family of *Flaviviridae* glycoproteins. The stability of the ion pair supports the hypothesis [12] that membrane insertion at the translocon complex occurs cooperatively for the E1 and E2 helices. Otherwise, having unpaired charged residues in the middle of a membrane bilayer would be thermodynamically unfavorable. The emerging structural model of the helix dimer shows the importance of the Lys370–Asp728 ion pair at the center of the lipid bilayer for the formation of the E1–E2 heterodimer.

Acknowledgments We are grateful for the financial support of the Universiti Teknologi MARA, Malaysia and by grant I 80469 by the Volkswagen foundation and for the research facilities from The Center for Bioinformatics Saarland University. We also thank the Deutsche Forschungsgemeinschaft Graduate School 1276/1 “Structure Formation and Transport in Complex Systems” for collaboration.

References

- Appel N, Schaller T, Penin F, Bartenschlager R (2006) From structure to function: New insights into hepatitis C virus RNA replication. *J Biol Chem* 281:9833–9836. doi:10.1074/jbc.R500026200
- Moradpour D, Penin F, Rice CM (2007) Replication of hepatitis C virus. *Nat Rev Microbiol* 5:453–463. doi:10.1038/nrmicro1645
- Bartosch B, Dubuisson J, Cosset F (2003) Infectious hepatitis C virus pseudo-particles containing functional E1–E2 envelope protein complexes. *J Exp Med* 197:633–642. doi:10.1084/jem.20021756
- Wakita T et al. (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* 11:905–905. doi:10.1038/nm1268
- Francki R, Fauquet C, Knudson D, Brown F (1991) Classification and nomenclature of viruses. Fifth report of the international Committee on taxonomy of viruses. *Arch Virol Suppl* 2:140–144
- Lindenbach BD, Thiel HJ, Rice CM (2001) *Flaviviridae: the viruses and their replication*. In: Knipe DM, Howley PM (eds) *Fields virology*, 5th edn. Lippincott Williams & Wilkins, Philadelphia, pp 991–1041
- Matsuura Y, Miyamura T (1993) The molecular biology of hepatitis C virus. *Semin Virol* 4:297–304. doi:10.1006/smv.1993.1026
- Dubuisson J et al. (2000) Glycosylation of the hepatitis C virus envelope protein E1 is dependent on the presence of a downstream sequence on the viral polyprotein. *J Biol Chem* 275:30605–30609. doi:10.1074/jbc.M004326200
- Cocquerel L, Wychowski C, Minner F, Penin F, Dubuisson J (2000) Charged residues in the transmembrane domains of hepatitis C virus glycoproteins play a major role in the processing, subcellular localization, and assembly of these envelope proteins. *J Virol* 74:3623–3633
- Cuthbertson JM, Bond PJ, Sansom MSP (2006) Transmembrane helix-helix interactions: comparative simulations of the glycoporphin A dimer. *Biochemistry* 45:14298–14310. doi:10.1021/bi0610911
- Ciczora Y, Callens N, Penin F, Pecheur EI, Dubuisson J (2007) Transmembrane domains of hepatitis C virus envelope glycoproteins: Residues involved in E1E2 heterodimerization and involvement of these domains in virus entry. *J Virol* 81:2372–2381. doi:10.1128/JVI.02198-06
- Cocquerel L et al. (2002) Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. *EMBO J* 21:2893–2902. doi:10.1093/emboj/cdf295
- Ciczora Y et al. (2005) Contribution of the charged residues of hepatitis C virus glycoprotein E2 transmembrane domain to the functions of the E1E2 heterodimer. *J Gen Virol* 86:2793–2798. doi:10.1099/vir.0.81140-0
- Ronecker S, Zimmer G, Herrler G, Greiser-Wilke I, Grummer B (2008) Formation of bovine viral diarrhoea virus E1–E2 heterodimers is essential for virus entry and depends on charged residues in the transmembrane domains. *J Gen Virol* 89:2114–2121. doi:10.1099/vir.0.2008/001792-0
- White SH (2004) The progress of membrane protein structure determination. *Protein Sci* 13:1948. doi:10.1110/ps.04712004
- Bond PJ, Sansom MSP (2003) Membrane protein dynamics versus environment: simulations of OmpA in a micelle and in a bilayer. *J Mol Biol* 329:1035–1053. doi:10.1016/S0022-2836(03)00408-X
- Böckmann RA, Caffisch A (2005) Spontaneous formation of detergent micelles around the outer membrane protein OmpX. *Biophys J* 88:3191–3204. doi:10.1529/biophysj.105.060426
- Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI (2006) Positioning of proteins in membranes: a computational approach. *Protein Sci* 15:1318–1333. doi:10.1110/ps.062126106
- Matthews EE, Zoonens M, Engelman DM (2006) Dynamic helix interactions in transmembrane signaling. *Cell* 127:447–450. doi:10.1016/j.cell.2006.10.016
- Treutlein HR, Lemmon MA, Engelman DM, Brunger A (1992) The glycoporphin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry* 31:12726–12732. doi:10.1021/bi00166a002
- Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM (1992) Sequence specificity in the dimerization of transmembrane α -helices. *Biochemistry* 31:12719–12725. doi:10.1021/bi00166a002
- Adams PD, Engelman DM, Brünger AT (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycoporphin A obtained through global searching. *Proteins* 26:257–261. doi:10.1002/(SICI)1097-0134(199611)26:3<257::AID-PROT2>3.0.CO;2-B
- MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: structure and implications. *Science* 276:131–133. doi:10.1126/science.276.5309.131
- Woolf TB (1998) Molecular dynamics simulations of individual α -helices of bacteriorhodopsin in dimyristoylphosphatidylcholine. II. Interaction energy analysis. *Biophys J* 74:115–131. doi:10.1016/S0959-440X(99)80015-3
- Candler A, Featherstone M, Ali R, Maloney L, Watts A, Fischer WB (2005) Computational analysis of mutations in the transmembrane region of Vpu from HIV-1. *BBA-Biomembranes* 1716:1–10. doi:10.1016/j.bbamem.2005.07.012
- Fischer WB, Sansom MSP (2002) Viral ion channels: structure and function. *BBA-Biomembranes* 1561:27–45. doi:10.1016/S0304-4157(01)00009-0
- Hénin J, Chipot C, Pohorille A (2005) Insights into the recognition and association of transmembrane α -helices. The free energy of α -helix dimerization in glycoporphin A. *J Am Chem Soc* 127:8478–8484. doi:10.1021/ja050581y
- Yoo J, Cui Q (2008) Does Arginine remain protonated in the lipid membrane? Insights from microscopic pKa calculations. *Biophys J* 94:61–63. doi:10.1529/biophysj.107.122945
- Wu CH et al. (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* 34:187–191. doi:10.1093/nar/gkj161

30. Hessa T et al. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381. doi:10.1038/nature03216
31. Combet C et al. (2007) euHCVdb: the European hepatitis C virus database. *Nucleic Acids Res* 35:D363. doi:10.1093/nar/gk1970
32. Simmonds P et al. (2005) Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 42:962–973. doi:10.1002/hep.20819
33. Larkin MA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947. doi:10.1093/bioinformatics/btm404
34. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. doi:10.1093/nar/gkh340
35. Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Bioinformatics* 12:543–548. doi:10.1093/bioinformatics/12.6.543
36. Rost B, Fariselli P, Casadio R (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci* 5:1704–1718. doi:10.1002/pro.5560050824
37. Juretic D, Zoranic L, Zucic D (2002) Basic charge clusters and predictions of membrane protein topology. *J Chem Inf Comp Sci* 42:620–632. doi:10.1021/ci10263s
38. Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849–850. doi:10.1093/bioinformatics/17.9.849
39. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. doi:10.1006/jmbi.2000.4315
40. Kahsay RY, Gao G, Liao L (2005) An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. *Bioinformatics* 21:1853–1858. doi:10.1093/bioinformatics/bti303
41. Park Y, Helms V (2008) Prediction of the translocon-mediated membrane insertion free energies of protein sequences. *Bioinformatics* 24:1271–1277. doi:10.1093/bioinformatics/btn114
42. Canutescu AA, Shelenkov AA, Dunbrack RL (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* 12:2001–2014. doi:10.1110/ps.03154503
43. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447. doi:10.1021/ct700301q
44. Griepernau B, Leis S, Schneider MF, Sikor M, Steppich D, Bockmann RA (2007) 1-Alkanols and membranes: a story of attraction. *Biochim Biophys Acta* 1768:2899–2913. doi:10.1016/j.bbame.2007.08.002
45. Faraldo-Gómez JD, Smith GR, Sansom MS (2002) Setting up and optimization of membrane protein simulations. *Eur Biophys J* 31:217–227. doi:10.1007/s00249-002-0207-5
46. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient. *Biopolymers* 38:305–320. doi:10.1002/(SICI)1097-0282(199603)38:3<305::AID-BIP4>3.0.CO;2-Y
47. Berendsen HJC, van der Spoel D, van Drunen R (1995) Gromacs—a message-passing parallel molecular-dynamics implementation. *Comput Phys Commun* 91:43–56. doi:10.1016/0010-4655(95)00042-E
48. Siu SWI, Vácha R, Jungwirth P, Böckmann RA (2008) Biomolecular simulations of membranes: physical properties from different force fields. *J Chem Phys* 128:125103. doi:10.1063/1.2897760
49. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676. doi:10.1002/jcc.20090
50. Berger O, Edholm O, Jahnig F (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* 72:2002–2013. doi:10.1016/S0006-3495(97)78845-3
51. Chiu SW, Clark M, Balaji V, Subramaniam S, Scott HL, Jakobsson E (1995) Incorporation of surface tension into molecular dynamics simulation of an interface: a fluid phase lipid bilayer membrane. *Biophys J* 69:1230–1245. doi:10.1016/S0006-3495(95)80005-6
52. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684. doi:10.1063/1.448118
53. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
54. Park Y, Helms V (2008) MINS2: revisiting the molecular code for transmembrane-helix recognition by the Sec61 translocon. *Bioinformatics* 24:1819–1820. doi:10.1093/bioinformatics/btn255
55. De Beek AO et al. (2000) The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J Biol Chem* 275:31428–31437. doi:10.1074/jbc.M003003200
56. Johansson ACV, Lindahl E (2006) Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. *Biophys J* 91:4450–4463. doi:10.1529/biophysj.106.092767
57. Bonifacino JS, Cosson P, Shah N, Klausner RD (1991) Role of potentially charged transmembrane residues in targeting proteins for retention and degradation within the endoplasmic reticulum. *EMBO J* 1991:2783–2793
58. Freitas JA, Tobias DJ, von Heijne G, White SH (2005) Interface connections of a transmembrane voltage sensor. *Proc Natl Acad Sci USA* 102:15059–15064. doi:10.1073/pnas.0507618102
59. Joh NHJ et al. (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* 453:1266–1270. doi:10.1038/nature06977
60. Meindl-Beinker NM, Lundin C, Nilsson IM, White SH, von Heijne G (2006) Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep* 7:1111. doi:10.1038/sj.embor.7400818
61. Im W, Feig M, Brooks CL (2003) An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys J* 85:2900–2918. doi:10.1016/S0006-3495(03)74712-2
62. Tanizaki S, Feig M (2005) A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J Chem Phys* 122:124706. doi:10.1063/1.1865992
63. Bu L, Brooks RCL (2008) De novo prediction of the structures of M. tuberculosis membrane proteins. *J Am Chem Soc* 130:5384–5385. doi:10.1021/ja710213p
64. Bu L, Im W, Brooks C (2007) Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations. *Biophys J* 92:854–863. doi:10.1529/biophysj.106.095216
65. Lazaridis T (2003) Effective energy function for proteins in lipid membranes. *Proteins* 52:176–192. doi:10.1002/prot.10410
66. Mottamal M, Zhang J, Lazaridis T (2006) Energetics of the native and non-native states of the glycoporphin transmembrane helix dimer. *Proteins* 62. doi:10.1002/prot.20844
67. Ulmschneider M, Ulmschneider J, Sansom M, Di Nola A (2007) A generalized born implicit-membrane representation compared to experimental insertion free energies. *Biophys J* 92:2338–2349. doi:10.1529/biophysj.106.081810
68. Nymeyer H, Woolf TB, Garcia AE (2005) Folding is not required for bilayer insertion: replica exchange simulations of an α -helical

- peptide with an explicit lipid bilayer. *Proteins* 59:783–790. doi:[10.1002/prot.20460](https://doi.org/10.1002/prot.20460)
69. Benedix A, Becker CM, de Groot BL, Caflich A, Bockmann RA (2009) Predicting free energy changes using structural ensembles. *Nat Meth* 6:3–4. doi:[10.1038/nmeth0109-3](https://doi.org/10.1038/nmeth0109-3)
70. Potapov V, Cohen M, Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel* 22:553–560. doi:[10.1093/protein/gzp030](https://doi.org/10.1093/protein/gzp030)
71. Langosch D, Brosig B, Kolmar H, Fritz HJ (1996) Dimerisation of the glycoporphin A transmembrane segment in membranes probed with the ToxR transcription activator. *J Mol Biol* 263:525–530. doi:[10.1006/jmbi.1996.0595](https://doi.org/10.1006/jmbi.1996.0595)
72. Senes A, Gerstein M, Engelman DM (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with β -branched residues at neighboring positions. *J Mol Biol* 296:921–936. doi:[10.1006/jmbi.1999.3488](https://doi.org/10.1006/jmbi.1999.3488)