

**Molecular Modeling of
the Transmembrane Domain
of Envelope Glycoproteins from Flaviviridae
Viruses**

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von
Siti Azma Jusoh

Saarbrücken
November 2010

Tag des Kolloquiums: 16. Dezember 2009

Dekan: Prof. Dr. Petra Bauer

Berichterstatter: Prof. Dr. Volkhard Helms

Prof. Dr. Richard Zimmerman

Akad. Mitarbeiter: Dr. Michael Hutter

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Saarbrücken, Dezember 2010

Acknowledgement

Foremost, I owe my deepest gratitude to my supervisor, Prof. Dr. Volkhard Helms for accepting me as his student and guiding me along my PhD journey. Thank you for patiently editing my writing and giving many useful advices. I also would like to thank his wife, Regina who helped my family and me during our initial stage in Germany. I would like to thank the group secretary, Kerstin for helping me with the paperworks.

I am thankful to the group members of Computational Biology Group for sharing memorable moments, experiences, valuable academic discussion as well as many other aspects of life especially Dr. Michael Hutter, Dr. Tihamér Geyer, Dr. Wei Gu, Mazen Ahmad, Peter Walter, Dr. Susanne Eyrisch, Mohamed Hamed Ali Fahmy, Christian Spaniol, Jennifer Metzger, Nadine Schaadt and Po-Hsien Lee.

My special appreciation goes to Barbara Hutter, Özlem Ulucan, Yvonne and Alesandro Marangon for being my closest friends.

I will not forget to thank Dr. Shirley Siu and Prof. Dr. Rainer Böckmann for teaching me MD simulations in membrane lipid bilayer. I thank Danielle, Caroline and Beate from Prof. Dr. Rainer Böckmann group and many other friends in Germany for various kinds of help.

I thank Dr. Christoph Welsch from the Johann Wolfgang Goethe University, Frankfurt for his contributions to the project described in Chapter 3. His contributions are marked specifically in the text.

I would not have been able to perform my study without financial contribution from my main sponsor, the Universiti Teknologi MARA Malaysia, and additional financial supports from the Volkswagen Foundation. I would like to acknowledge Centre for Bioinformatics of Saarland University for the computational and research supports.

Finally, I would like to thank my parents, Jusoh Ismail and Aini Mohamud for their love, care and dedication to rise me. I am grateful for my beloved husband, Mustakim Mohamed Adnan for his endless love and continuous support and also both of my sons Luqman Syahmi and Ayman Shafi for being the major drive that motivates me every single day. Thank you for my families for always being there for me.

I thank to Allah SWT for His bless and guidance through this entire journey.

Abstract

The putative transmembrane (TM) domains of the envelope glycoproteins from the family Flaviviridae consist of a highly polar segment in between two hydrophobic stretches. This type of sequence pattern does not yet exist in the database of high resolution structures of membrane proteins. Mutagenesis studies have shown that the TM domains act as membrane and signal anchors, and are responsible for heterodimerization. In hepatitis C virus (HCV), the TM domains of the envelope glycoproteins E1 and E2 were hypothesized to heterodimerize via an ion pair of Lys-Asp. Our MD simulations showed that the E1-E2 heterodimer formed by the charged residues located in the core of the lipid bilayer stabilized the helical conformation of E2. We compared the effect of other types of ion pair interactions using engineered peptides and obtained similar results. We found that an Asp amino acid had the strongest kink-inducing effect on the helix when it was located in the middle of a single-pass TM helix. The extended analyses on dengue, Japanese encephalitis, West Nile and bovine viral diarrhoea viruses again showed that their putative TM domains behave similarly. All the TM domains of the E1/prM tended to tilt and remain helical in membrane bilayer. In contrast, the TM domains of the E2/E that contain a central Asp residue were severely kinked. Altogether, these TM domains illustrated a similar structural behavior in the lipid bilayer milieu.

Kurzfassung

Die mutmaßlichen Transmembran (TM)-Domänen der Hüllglykoproteine der Familie *Flaviviridae* bestehen aus einem hochpolaren Segment zwischen zwei hydrophilen Abschnitten. Diese Sequenzmuster sind noch nicht in der Datenbank hochaufgelöster Strukturen von Membranproteinen enthalten. Gemäß Mutagenesestudien agieren die TM-Domänen als Membran- und Signalanker und sind für die Heterodimerisierung verantwortlich. Im Hepatitis C-Virus (HCV) heterodimerisieren die TM-Domänen der Hüllglykoproteine E1 und E2 möglicherweise über ein Ionenpaar zwischen Lys-Asp. Unsere MD-Simulationen zeigten, dass das E1-E2-Dimer, das durch die geladenen Residuen im Kern der Lipiddoppelschicht gebildet wird, die helikale Konformation von E2 stabilisiert. Der Effekt anderer Ionenpaarinteraktionen in künstlichen Peptiden führte zu ähnlichen Ergebnissen. Asp in der Mitte einer TM-Helix verursachte den stärksten Krümmungseffekt. Weitere Analysen mit anderen *Flaviviridae* (Dengue, Japanese encephalitis, West Nile und bovine viral diarrhoea virus) zeigten ebenfalls ein ähnliches Verhalten ihrer mutmaßlichen TM-Domänen. Alle TM-Domänen von E1/prM tendierten zur Krümmung und blieben in der Membrandoppelschicht helikal. Hingegen waren die TM-Domänen von E2/E, die ein zentrales Asp enthalten, stark gekrümmt. Insgesamt zeigten diese mutmaßlichen TM-Domänen ein ähnliches strukturelles Verhalten in der Membran.

Thesis Overview

This thesis deals with the transmembrane (TM) helix segments of the two envelope glycoproteins from Flaviviridae viruses which contain several highly polar amino acid residues located in the centre of the TM segments. Here, the structure of the thesis is outlined.

Chapter 1 provides a brief introduction of membrane proteins which covers their types, structure and functions, purposely to highlight their crucial roles in living organisms. Then, I specifically explain the mechanisms and components related to the synthesis of the helical membrane proteins. The ribosome-translocon complexes are shown to directly be involved in the biogenesis, the lateral translocation into the membrane lipids and the topological decision of the membrane proteins. Furthermore, I bring up the unresolved issues about the models of the active ribosome-translocon complexes and recent discussions regarding the controversial biological hydrophobicity scale. These fundamental issues are critical in the decision making to produce accurate prediction methods for membrane proteins. In addition, I give some information about the membrane lipids which are the residing home for the membrane proteins and their dynamic properties as well as their interaction to each other. At the end, I summarized the background of the peptides that were used in this thesis.

In Chapter 2, the technique of classical MD simulation is briefly explained including a discussion of the utilized force field. The protocols to implement a simulation of a peptide in lipid bilayer system are also described here. Some experimental data are shown side by side with respective results from other simulation studies.

Chapter 3 presents the results of our first project with the objective to study an ion pair interaction that mediates a TM helix dimer. This project was motivated by the abundant experimental data on hepatitis C virus (HCV) that suggested an important contribution of a salt-bridge for the dimerization of the putative TM helices from the E1 and E2 envelope glycoproteins. We modeled the suggested E1-E2 heterodimer with the Lys-Asp salt-bridge purposely to observe the effect on the helical structure. We explored with the MD simulation methods other possibilities that can weaken or strengthen the helix-helix interaction by performing mutations on the selected key residues as proposed by the experiments. From this, we conclude that the strong E1-E2 dimer interaction is driven by the ion pair. But the ion pair alone does not prevent the local unfolding of the helical structure. From the simulations, we clearly saw an additional hydrogen bond interaction involving another polar residue that mediates the dimer and improved the helical conformations.

Chapter 4 – The initial objective of this second project was to study three other types of salt bridges based on the same idea as the first work. The salt bridge interactions between the two helices were set up to be in the middle of the bilayer core and this time we used engineered TM sequence segments. The helical segments, named H-segment, each contained one charged residue in the centre. Therefore, the resulted interaction will solely come from the studied residues that formed the salt bridge. As expected, we showed that each type of salt bridge resulted in a stable dimer interaction throughout the 200 ns MD simulations. Also, the structural behavior of the helices that contained Lys and Asp were the same as the E1 and E2 from the HCV. Additionally, we discovered that on this time scale (more than 150 ns), several water molecules from the bulk phase were able to penetrate into the bilayer core and solvated the charged residues. Even more, they were dynamically exchanged with bulk waters. This event was illustrated by continuous replacement of water molecules in the core region of the bilayer. However, this surprising microsolvation phenomenon in hydrated hydrophobic bilayers needs to be further verified by experimental methods in order to explain in quantitative ways.

In Chapter 5, we again used the biological sequences as models to study effects of charged amino acids on the helical stability of the TM domain. Based on the simulation of helix monomers from the two previous works, we noted that the helices containing a single Asp amino acid in the centre of the transmembrane helix were locally unfolded and kinked. Therefore, one of the questions that we wanted to know was whether an Asp amino acid caused that particular effect. We used the putative TM domain of several other well known Flaviviridae viruses as models. All of them contain at least one charged residue at the middle of their TM segments. Interestingly, the simulation results showed that the Asp residue located in the centre of the transmembrane helix tend to cause disruption and kinking to the monomers. Moreover, all the E1/prM helices which do not contain an Asp remained as perfect helical structure, rather than most of the E2/E helices that were severely kinked. We also used several popular web servers which predict three-dimensional structure in order to characterize these putative TM sequences. Here we showed results from the I-Tasser prediction server that managed to provide 9 out of 10 correct 3D predicted structures similarly as been observed during the MD simulation.

Finally, we summarize the findings and propose further investigations for this study in Chapter 6.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit den Transmembran (TM)-Helix-Segmenten von zwei Glykoproteinen der Hüllen von Flavoviridae Viren, die einige hochpolare Aminosäuren enthalten, die im Zentrum des TM-Segments lokalisiert sind. Im folgenden wird der Aufbau der Arbeit erläutert.

Kapitel eins umfasst eine kurze Einführung der Membranproteine und beschreibt deren Typen, ihre Struktur und deren Funktionen insbesondere im Hinblick auf ihre wichtigen Rollen im lebenden Organismus. Im Anschluss beschreibe ich die Mechanismen und die Komponenten, die bei der Synthese der helikalen Membranproteine von Bedeutung sind. Es wird gezeigt, dass die Ribosom-Translokation-Komplexe unmittelbar an der Biogenese, der lateralen Translokation in die Membranlipide und der topologischen Ausrichtung der Membranproteine beteiligt sind. Darüber hinaus spreche ich wenig erforschte Aspekte des Modells des aktiven Ribosom-Translokation-Komplexes an sowie neuere Diskussionen über die Kontroversen bezüglich der biologischen Hydrophobizitätsskala. Diese grundlegenden Themen spielen eine entscheidende Rolle im Prozess der Entscheidungsfindung, um möglichst getreue Vorhersagemethoden für Membranproteine zu entwickeln. Zusätzlich gehe ich auf Membranlipide ein, die in Membranproteinen angesiedelt sind, sowie auf ihre dynamischen Eigenschaften und Interaktionen untereinander. Schließlich gebe ich einige Hintergrundinformationen zu Peptiden, die im Rahmen dieser Arbeit Verwendung fanden.

In Kapitel zwei wird die Technik der klassischen MD-Simulation erläutert und das verwendete Kraftfeld diskutiert. Die Protokolle zur Implementierung der Simulation eines Peptids in der Lipiddoppelschicht werden ebenfalls beschrieben. Einige experimentelle Daten werden im einzelnen dargestellt mit entsprechenden Resultaten aus anderen Simulationsstudien.

Kapitel drei präsentiert die Resultate unseres ersten Projekts, welches sich mit der Interaktion eines Ionenpaares befasst, das für die Mediation eines TM-Helix Dimers verantwortlich ist. Die Motivation für dieses Projekt beruht auf zahlreichen experimentellen Daten über das Hepatitis C Virus (HCV), die die Bedeutung einer Salzbrücke in den Vordergrund stellen, welche die Dimerisierung der TM-Helices aus den E1 und E2 Hüllglycoproteinen vermittelt. Wir modellierten dieses E1-E2 Heterodimer mit der Lys-Asp Salzbrücke in der Absicht, einen Effekt auf die helikale Struktur beobachten zu können. Wir überprüften mit Hilfe von MD-Simulationen weitere Möglichkeiten, die Helix-Helix Interaktion zu verstärken oder zu schwächen, indem ausgewählte Schlüsselresiduen, die experimentell bestimmt wurden, mutiert wurden. Aus dieser Studie schlussfolgern wir, dass die starke E1-E2 Dimer-Interaktion über das Ionenpaar vermittelt wird. Jedoch kann das Ionenpaar allein die lokale Strukturänderung der helikalen Struktur nicht verhindern. Mit Hilfe der Simulationen beobachteten wir eine zusätzliche

Wasserstoffbrücken-Interaktion, an der eine weitere polare Residue beteiligt ist, die die Bildung des Dimers vermittelt und die helikalen Konformationen stabilisiert.

Kapitel vier – Das ursprüngliche Ziel des zweiten Projekts lag in der Studie von weiteren drei Typen von Salzbrücken, die auf derselben Grundlage wie in der ersten Studie basieren. Die Salzbrückeninteraktionen zwischen zwei Helices wurden in der Mitte eine Lipid-Doppelschicht platziert. Wir verwendeten in dieser Simulation künstliche TM Sequenzsegmente. Die helikalen Segmente, genannt H-Segmente, enthalten in ihrem Zentrum jeweils eine geladene Residue. Deshalb beruht die resultierende Interaktion allein auf der untersuchten Residue, die die Salzbrücke ausbildet. Wir zeigten, dass jeder Salzbrückentyp erwartungsgemäß zu einem stabilen Dimer im Verlauf von 200ns MD-Simulationen führt. Auch gleicht das strukturelle Verhalten der Helices, die Lys und Asp enthalten, den E1 und E2 Proteinen aus HCV. Zusätzlich entdeckten wir auf dieser Zeitskala (mehr als 150ns), dass einige Wassermoleküle aus der Bulkphase in den Doppelschichtkern eindringen und die geladenen Residuen solvatisieren konnten. Überdies wurden sie dynamisch mit dem Bulkwasser ausgetauscht. Dieses Ereignis wurde begleitet von fortlaufendem Austausch von Wassermolekülen in der Kernregion der Doppelschicht. Dieses überraschende Mikrosolvatationsphänomen in hydratisierten hydrophoben Doppelschichten muss allerdings mit Hilfe weiterer experimenteller Methoden verifiziert werden um diese Beobachtung auch quantitativ zu erklären.

In Kapitel fünf verwendeten wir biologische Sequenzen als Modelle, um die Auswirkungen von geladenen Aminosäuren auf die helikale Stabilität der TM-Domäne zu untersuchen. Auf Grundlage der Simulation von Helixmonomeren aus den beiden vorangegangenen Arbeiten fanden wir heraus, dass Helices, die eine einzige Asp-Aminosäure im Zentrum der Transmembran-Helix haben, lokal denaturiert und geknickt werden. Deshalb bestand eine zu klärende Frage darin, ob ein Asparagin ebenfalls einen solchen Effekt verursachen kann. Als Modelle verwendeten wir die mutmaßliche TM-Domäne von einigen weiteren gut bekannten Flavoviridae. Alle enthalten wenigstens eine geladene Residue in der Mitte ihrer TM-Segmente. Interessanterweise zeigten die Ergebnisse der Simulation, dass die Asp-Residue im Zentrum der Transmembran-Helix dazu neigt, innerhalb der Monomere Brüche und Knicke zu verursachen. Darüber hinaus bewahren alle E1/prM Helices, die kein Asp enthielten, ihre perfekte helikale Struktur, wohingegen die meisten E2/E-Helices strukturell deutlich geknickt wurden. Wir verwendeten einige weit verbreitete Webserver zur Vorhersage von dreidimensionalen Strukturen, um diese mutmaßlichen TM-Sequenzen zu charakterisieren. Wir zeigen Ergebnisse des I-Tasser Vorhersageservers, dem es gelang, 9 von 10 korrekte 3D-Strukturen vorherzusagen, die denen aus den MD-Simulationen ähnlich sind.

Zum Schluss fassen wir die Ergebnisse in Kapitel sechs zusammen und geben einen Ausblick auf mögliche weitere Untersuchungen, die die Projekte dieser Arbeit betreffen.

Contents

Chapter 1 Introduction of the Biological Components	1
1.1 Membrane Proteins	1
1.1.1 Types of Membrane Proteins	2
1.1.2 Functions of Integral Helical Membrane Proteins	4
1.2 Biogenesis of Helical Membrane Proteins	7
1.2.1 Translocon Machinery	7
1.2.2 Constitutive Membrane Proteins	8
1.2.3 Non-constitutive membrane protein	11
1.2.4 Biological Hydrophobicity Scale	12
1.3 Membrane Lipids	14
1.4 Flaviviridae Virus family	16
1.4.1 Transmembrane Domains of the Envelope Glycoproteins	19
1.4.2 TM Domain of the E1-E2 Dimer of Hepatitis C Virus	19
1.4.3 Model of the E1-E2 Dimer Biogenesis	20
Chapter 2 Computational Methodology	22
2.1 Molecular Dynamics Simulations	22
2.2 Methods	23
2.2.1 Molecular Force Fields	25
2.2.2 Periodic Boundary Condition	27
2.2.3 Treatment of Long Range Interactions	27
2.2.4 Simulation in the Isothermal-Isobaric Ensemble	28
2.2.5 Simulation of Peptides in Lipid Bilayer Systems	30
2.2.6 Analyses of Simulations	34
2.2.7 Current State of Molecular Dynamics Simulations of Membrane Systems	36
Chapter 3 Contribution of Charged and Polar Residues for the Formation of the E1-E2 Heterodimer from Hepatitis C Virus	38
3.1 Introduction	39
3.2 Methods	41
3.2.1 Sequences	41
3.2.2 Sequence Analysis	41
3.2.3 TM Protein Prediction	42
3.2.4 Molecular Dynamics Simulations	42
3.3 Results	44
3.3.1 Sequence Analysis of the TM Domain of the E1 and E2	44
3.3.2 Identification of TM Residues by Secondary Structure Prediction Methods	44
3.3.3 MD Simulations of E1 and E2 Monomers	45
3.3.4 MD Simulations of E1-E2 Heterodimers	47
3.3.5 Mutational Analysis	49

3.3.6	Comparison of MD Structures vs the NMR structure.....	53
3.4	Discussion.....	54
3.4.1	Simulations of Heterodimer Mutants.....	57
3.4.2	GxxxG Motif.....	58
3.5	Conclusion.....	59
Chapter 4	Microsolvation of Bridging Ion Pairs in Transmembrane Helix Dimers.....	60
4.1	Introduction.....	61
4.2	Methods.....	62
4.2.1	Sequences and System Preparation.....	62
4.2.2	Peptide-Bilayer System Setup.....	63
4.2.3	Simulation Details.....	64
4.3	Results and Discussion.....	64
4.3.1	Simulation of Monomers in DMPC and DPPC Lipid Bilayers.....	65
4.3.2	Dimer Simulations.....	68
4.3.3	Water Hydration of the Membrane Core.....	71
4.3.4	Microsolvation of the Putative Transmembrane Helix Dimers from Hepatitis C Virus.....	76
4.4	Conclusion.....	79
Chapter 5	Molecular Dynamics Simulation of Putative Transmembrane Domains of Envelope Glycoproteins from Flaviviridae Viruses.....	81
5.1	Introduction.....	82
5.2	Methods.....	83
5.2.1	Sequence Analyses.....	83
5.2.2	Molecular Dynamics Simulations.....	83
5.3	Results.....	84
5.3.1	Secondary Structure Prediction.....	84
5.3.2	I-TASSER 3D Structure Prediction.....	87
5.3.3	MD Simulations of TM Helix Monomers.....	88
5.4	Discussion and Conclusion.....	90
Chapter 6	Conclusion and Outlook.....	92
References		

List of Abbreviations

DENV	dengue virus
DPPC	dipalmitoyl-phosphatidylcholine
DMPC	dimyristoyl-phosphatidylcholine
DOPC	dioleoyl-phosphatidylcholine
ER	endoplasmic reticulum
EM	electron microscopy
GPCR	G protein-coupled receptors
MD	molecular dynamics
NMR	Nuclear Magnetic Resonance
HCV	hepatitis C virus
JEV	Japanese encephalitis virus
MD	molecular dynamics
PDB	protein data bank
PME	particle mesh Ewald
POPC	palmitoyl-oleoyl-phosphatidylcholine
POPE	palmitoyl-oleoyl-phosphoethanolamine
RNA	ribonucleic acid
RMSD	root mean square deviation
SRP	signal recognition particle
TM	transmembrane
WNV	West Nile virus
YFV	yellow fever virus
Asp	Aspartic Acid
Asn	Asparagine
Cys	Cysteine
Gln	Glutamine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
Arg	Arginine
Glu	Glutamic Acid
Lys	Lysine

Chapter 1

Introduction of the Biological Components

1.1 Membrane Proteins

Membrane proteins represent up to 30% of the open-reading frames of sequenced genomes (Wallin & von Heijne, 1998) and play essential roles for biological functions such as signal transduction, solute and molecular transport across membranes, energy production, membrane and protein biogenesis, cell-cell interactions and nerve conduction. These important cellular processes make them the prime target for drug design. In fact, the membrane proteins are currently the targets for at least 50% drugs in the market (Terstappen & Reggiani, 2001)(Overington et al, 2006). Despite the growing number of structures of membrane proteins at atomic resolution, the number is still representing only less than 2% out of the total. Currently, there are only ~252 three-dimensional structures (<http://blanco.biomol.uci.edu/>) (White, 2009) of different membrane proteins determined at atomic resolution by X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy due to the difficulties with extraction and crystallization.

The main criterion contributes to the differences between membrane proteins and soluble proteins are due to their residing location. The soluble proteins reside in an aqueous phase but the membrane proteins are buried in the hydrophobic milieu of membrane lipids. Thus, the exterior surfaces of the membrane embedded proteins comprise of mostly non-polar amino acids. However, their internal packing is similar to that of soluble proteins.

1.1.1 Types of Membrane Proteins

There are two main groups of membrane proteins; the integral type and the peripheral type. The integral membrane proteins are permanently attached to the membrane lipids, where a large portion is embedded in the hydrophobic layer of the cell membrane. In contrast, the peripheral type is temporarily adhered to the biological membrane where they may interact with other proteins or directly with the membrane lipids. Examples of peripheral type of proteins are regulatory protein subunits associated with ion channels or receptor proteins. Integral membrane proteins can be again categorized into two main types; the α -helix bundles and the β -barrel types. The helix-bundle type of membrane proteins occur in most cell membranes. Indeed, ~27% of the human proteome are estimated to be α -helical transmembrane (TM) proteins (Almen et al, 2009). The β -barrel proteins where the β -strands are arranged in an anti-parallel fashion are found only in the lipid-rich cell walls of a few Gram-negative bacteria, in outer membrane of mitochondria and in chloroplasts. The bacterial porin protein family typically forms as β -barrel proteins. The porins, which contain a water-filled channel, function as a filter to transport hydrophilic molecules across the bacterial cell membrane.

The α -helix bundle membrane proteins can be further assigned to three subgroups which are a monotopic that spans half of the bilayer, a bitopic that span both of the bilayer, and a polytopic that span the entire bilayer more than once which also called multi-spanning membrane proteins (illustrated in **FIGURE 1.1**). In this thesis, the studies are focusing only on the bitopic and the polytopic types of the integral membrane proteins. Below, I will describe main characteristics of the helical membrane proteins. Several groups of membrane proteins will be briefly explained and the recent published crystal structures are shown together with their protein databank (PDB) id.

Integral Helical Membrane Proteins

The TM helices of the integral membrane proteins are composed primarily of non-polar amino acids. Typically, they consist of about 25 amino acids (1.5\AA per residue) which can comfortably span the 30\AA thick membrane lipid bilayer (Bowie, 1997)(White & Wimley, 1999). The preferred location of the amino acids in the TM segments is based on their properties. The hydrocarbon region of the membrane bilayer usually consists of non-polar amino acids such as Ala, Leu and Val. Their hydrophobic side-chains can interact well with the lipid carbon tails. On the other hand, polar and charged amino acids as for example Asp, Arg, Lys, Glu, Asn and Glu are normally present near the end of the TM segments near to the membrane interfacial region. At this location, the side chains of these amino acids can form hydrogen bonding with bulk water. Another distinct feature of the TM helix is the distribution of the Tyr, Trp and Phe residues towards the membrane interface. This clustering of aromatic amino acids of

the integral membrane proteins has been referred as ‘aromatic belt’ (Ulmschneider & Sansom, 2001)(Adamian et al, 2005).

The higher distribution of Arg and Lys in cytoplasm compared to periplasm was statistically observed in bacterial inner membrane proteins that led to the ‘positive-inside rule’ (von Heijne, 1992). The positive-inside rule is useful to predict the TM protein topology based on the amino acid sequences. When the N-terminal flanking region possesses fewer positive charges and the TM domain is relatively longer, the segment tends to form an N_{lum}/C_{cyt} orientation, in which the N-terminus is in the lumen of endoplasmic reticulum and the C-terminus is in the cytoplasm. In contrast, when the N-terminal flanking domain possesses more positive charges than the C-terminus, the TM domain adopts the opposite orientation (N_{cyt}/C_{lum}). Both orientations are possible for the bitopic TM helix and they are categorized into the type I membrane protein if contained a cleavable signal-anchor (**FIGURE 1.1**). Type II and type III contained a non-cleavable signal anchor attached to the N-terminus and located in the cytosol and lumen, respectively.

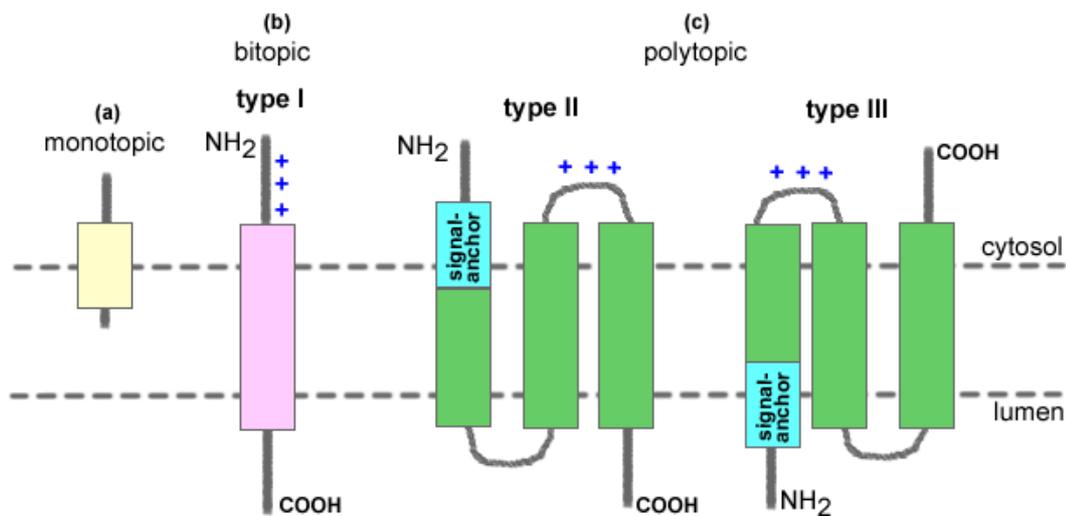


FIGURE 1.1 The positive-inside rule and topologies of membrane proteins. The dash lines indicate the membrane bilayer interface. (a), (b) and (c) are types of helical membrane proteins; monotopic, bitopic and polytopic, respectively. The positive-inside rule illustrates the bias positively charged residues in the connected loop (fewer than 60 residues) between the TM helices that located in the cytoplasm. However, it should be noted that if the length of the loop is longer, the rule is ignored. Also, type of TM helices depends on the signal anchor (cleavable or non-cleavable) and its location (cytosol or lumen).

1.1.2 Functions of Integral Helical Membrane Proteins

G-protein-coupled receptors (GPCRs)

G-protein-coupled receptors (GPCRs) are the type of integral membrane protein known as membrane receptors that play specialized roles in the communication between the cell and the extracellular signals (lights, odors and tastes) and intracellular signals (hormones and neurotransmitters) through the signal transduction process to trigger changes in the function of the cell. The family of GPCRs is the largest class of receptors in the human genome and is found only in eukaryotes. In fact, it is the largest of all protein families. GPCRs also have a great pharmacological importance since 50-60% of all approved drugs are targeting members from GPCR family (Hopkins & Groom, 2002)(Overington et al, 2006). The signature motif of GPCRs is that they comprise of 7-TM helices (**FIGURE 1.2**).

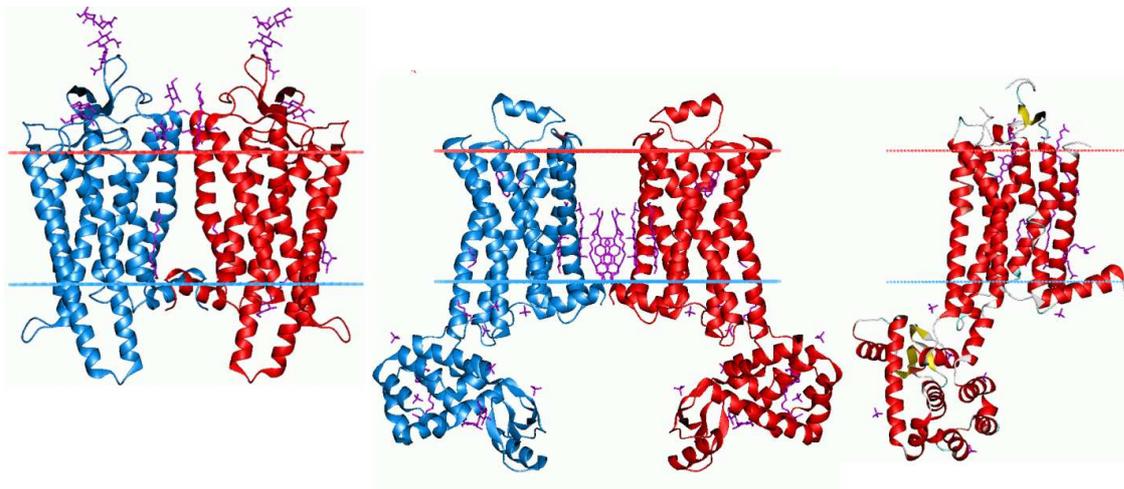


FIGURE 1.2 The GPCR family. Examples here are (a) Mammalian Rhodopsin/Opsin from *Bos taurus* (3CAP); (b) Human β_2 -adrenergic (1RH1); (c) Human adenosine A_{2A} receptor (A_{2A} -adenosine) (3EML) is the most recent solved GPCRs structure. The figures were taken from the OPM database (<http://opm.phar.umich.edu/>).

Active Transport

The transport proteins are mostly integral membrane proteins which are responsible for the movement of ions, small molecules, proteins and lipids across a biological membrane. The active transporter requires energy to function in order to transport solutes against the concentration gradients. The ATP-binding-cassette (ABC) family of membrane proteins is one example of active transporters that utilize the energy released during adenosine triphosphate (ATP) hydrolysis to translocate a wide variety of substrates across extra- and intracellular membranes such as small toxic molecules and lipids (**FIGURE**

1.3). P-glycoprotein is a member of the ABC transporter family associated with multi-drug resistance (MDR). It belongs to the human ABCB family (known as ABCB1/MDR1). P-glycoprotein functions as an efflux-pump which filters out hundreds of chemically unrelated toxins including the drugs to suppress tumor cells. This has caused problems in the treatment of cancers. The recently published crystal structure of mouse P-glycoprotein serve as the best template to study human P-glycoprotein (**FIGURE 1.3** (a)).

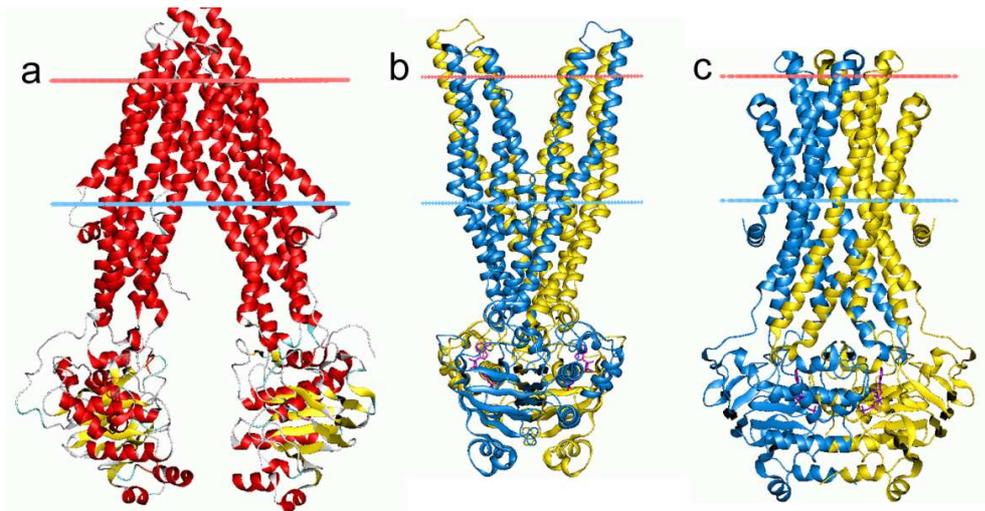


FIGURE 1.3 The ABC transporters. (a) The crystal structure of P-glycoprotein from mouse (ABCB1a) (3G5U) (Aller et al, 2009) was recently published. It currently represents the best template for human P-glycoprotein. (b) MsbA (3B60) (Ward et al, 2007) is also an MDR type of ABC transporter as well as (c) Sav1866 (2HYD) (Dawson & Locher, 2006). Both are from bacteria. The sequence identities among these three structures are only ~30% to each other. The proteins classified in the ABC-protein family are based on the sequence motif and domain organization. The figures were taken from the OPM database (<http://opm.phar.umich.edu/>).

Channels and Pores

The channel or pore type refers to the membrane proteins that have pore across the lipid bilayer that function to control the influx and outflux of the molecules or ions across the cellular membrane.

Ion channels contain a selective ion-conduction pore that can be gated by voltage or a ligand. The main function of voltage-gated ion channels in the nervous system is to process sensory signals and generating motor outputs. They have crucial roles in the initiation and propagation of nerve impulses. The first discovered structure of membrane channel was from the bacterial potassium channel, KvAP (Doyle et al, 1998). Up to now, there are several crystal structures for potassium channels for examples from eukaryote are Kv1.2 (3LUT) (see **FIGURE 1.4**), KcsA (1K4C); and bacteria: KvAP, Kirbac. The potassium channels are commonly voltage-gated and function in a tetrameric unit. Other voltage-gated

channels that already have a published high resolution structure are sodium channel (Nav), calcium channel (Cav). Ligand-gated ion channels usually form a pentameric unit.

Pore type of membrane proteins usually transport water (an example in the **FIGURE 1.4**: AQP1) and glycerol. They are known as aquaporins and glycerolporin, respectively. Similar to the ion channels, the pore type function in a passive way but highly selective based on the size of the molecules. Other than channels to regulate ions, there are also channels to regulate energy for example ATP channels and protein conducting channels (e.g. SecY/Sec61; see more in section **1.2.1**).

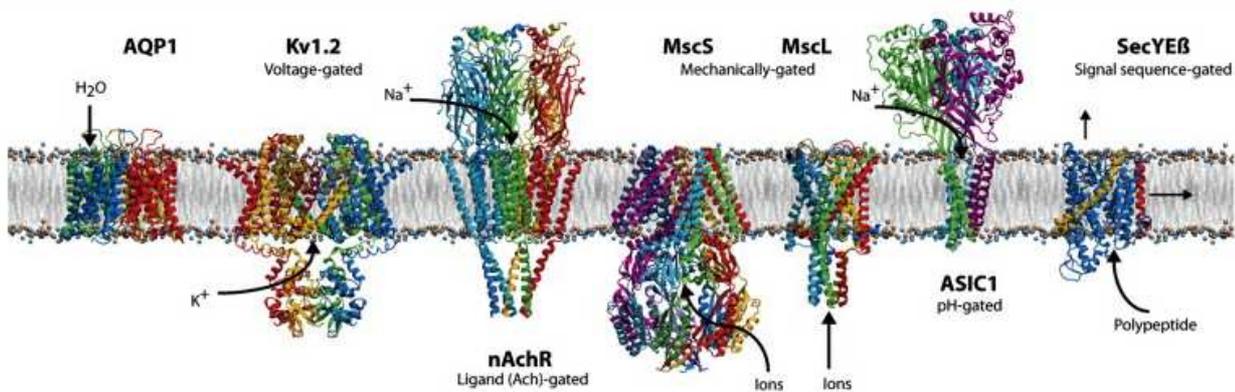


FIGURE 1.4 Structure of membrane protein channels. From left to right: The human aquaporin typically can function as a single subunit, the potassium channel is commonly consists of 4 monomers that form a functional central pore. The acetylcholine receptor (nAChR), MscS, MscL, ASIC1, SecY. The figure was adapted from (Khalili-Araghi et al, 2009); depicts the channel types that have been studied using the molecular dynamics simulation.

Other Functions Energy Generation

In plants and bacteria, membrane proteins play the main role to produce energy through the process of photosynthesis by capturing the light energy. The first high resolution structure of a membrane protein was that of the bacterial photosynthetic reaction center from purple bacteria (Deisenhofer et al, 1984). Other than that, membrane proteins also can exist as enzymes that are responsible for intramembrane proteolysis and are involved in cell immune system known as family of Membrane Associated Proteins in Eicosanoid and Glutathione metabolism (MAPEG).

1.2 Biogenesis of Helical Membrane Proteins

1.2.1 Translocon Machinery

The translocon is the core machinery involved in the translocation process of both soluble proteins and membrane proteins. However, the synthesis of membrane proteins is unique and much more complex than that of soluble proteins (Zimmermann et al, 2010)(White & von Heijne, 2008)(Rapoport, 2007)(Rapoport et al, 2004). The translocon is directly involved in recognition, orientation, lateral integration and insertion of the membrane protein. It is called Sec61 complex and SecY complex (FIGURE 1.5). These proteins are located in the eukaryotic endoplasmic reticulum (ER) and prokaryotic plasma membrane, respectively. Its main function is to discriminate between the soluble proteins that have to cross the membrane completely and membrane proteins that integrate laterally into the membrane lipids. Furthermore, the translocon complex is even directly involved in determining the membrane protein topography based on the acid amino sequence. However, most of the molecular details of this mechanism are still unclear.

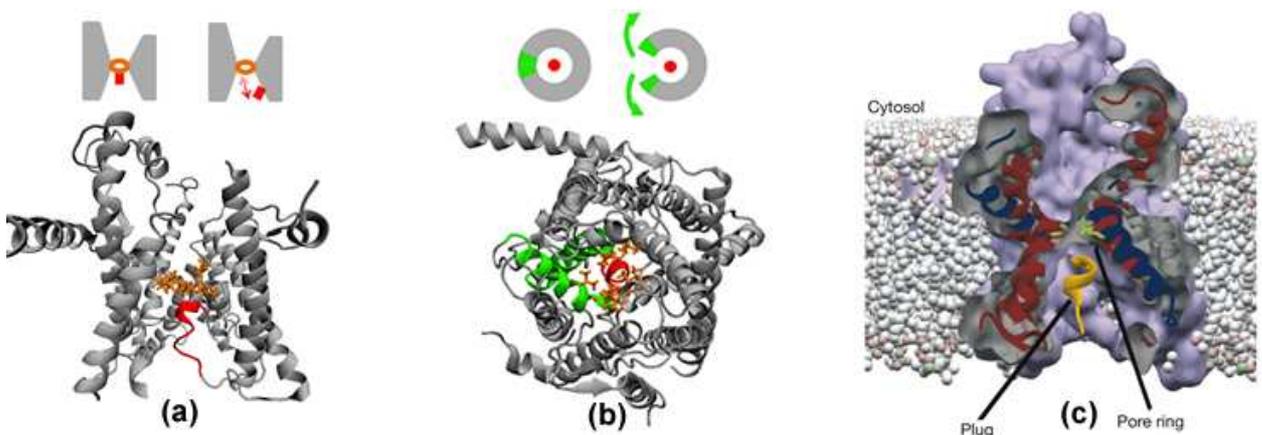


FIGURE 1.5 The SecY complex. (a) View from side, b. view from the cytosol, (adapted from (Zhang & Miller, 2010); c. Cross-sectional view of the channel from the side (Rapoport, 2007). The first crystal structure of the translocon in the closed state was solved for the archaeal SecY complex of *Methanococcus jannaschii* in the year 2004 (Berg et al, 2004). Figure (a) and (b) are illustrated with the “diagram” showing the predicted movement of the structure from the closed-state to the open-state. Note that the ‘pore ring’ is involved in filtering specific molecules to enter when the ‘plug’ is open.

Briefly, the translocon is a conserved heterotrimeric membrane protein complex consisting of α -, β -, and γ -subunit. The α - and γ -subunits are highly conserved in both the eukaryotes and the bacteria and crucial for the function of the translocon. The α -subunit mainly forms the protein translocation channel. It has an aqueous interior (Rapoport et al, 1996) (Rapoport, 2007) and is divided into two halves, TM 1-5 and TM 6-10. The loop between TM 5 and 6 at the back of the α -subunit serves as a hinge, allowing the α -subunit to open at the front, the so called "lateral gate". The 10 helices of the α -subunit form an hourglass-shaped pore that consist of cytoplasmic and external funnels, the tips of which meet about half way across the membrane. There are also six hydrophobic residues located at the tips where their side-chains orient inwardly to behave as a "pore ring". When the cytoplasmic funnel is empty, the external funnel is plugged by a short helix (**FIGURE 1.5**). Thus, not even a small molecule including a water molecule can pass through the protein-conducting channel (Saparov et al, 2007)(Gumbart & Schulten, 2006). The channel itself is a passive pore which relies on the translocon partners for example a ribosome as driving force to translocate the nascent polypeptides (Rapoport, 2007).

1.2.2 Constitutive Membrane Proteins

Like all proteins in cells, constitutive membrane proteins go through the normal translation processes in ribosome. Then depending on the signal from precursor polypeptides, the nascent peptides can get translocated by the translocon complex machinery using two ways; Co-translational and post-translational. (1) Co-translational: Briefly, this secretory pathway through the translocon complex begins when an ER signal sequence emerges from the ribosome and is recognized by the ribonucleoparticle signal-recognition particle (SRP). Then, the ribosome-nascent polypeptide-SRP complex targets to the ER membrane. The GTP hydrolysis by SRP and its receptor releases the signal sequence and subsequently transfer the complex to a protein-conducting channel of the translocon complex. The elongating polypeptide chain subsequently moves directly from the tunnel inside the ribosome into the protein-conducting channel. Furthermore, the translocon complex processes the nascent polypeptide according to the sequence segment. Soluble proteins have mostly hydrophilic segments that will cross the translocon channel. On the other hand, the membrane segments are more hydrophobic and will integrate laterally into the membrane bilayer (**FIGURE 1.6**). (2) Post-translational: The soluble proteins may also get transported through the translocon complex using this pathway. Suggested due to their weak hydrophobic signal, the precursor polypeptides escape the recognition from the SRP and complete the synthesis before the translocation (Tyedmers et al, 2000). There is no such case for the membrane proteins which use the post-translational pathway.

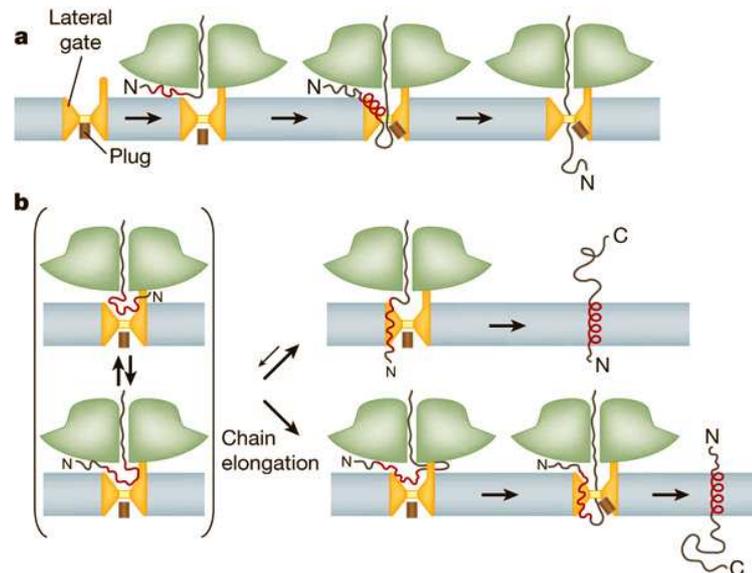


FIGURE 1.6 Biosynthesis of proteins in ribosome-translocon complexes. (a) Translocation of a soluble protein. (b) Translocation of membrane proteins. Upper panel shows a bitopic $N_{lum}-C_{cyc}$ type of membrane protein, and the lower panel shown the $N_{cyc}-C_{lum}$ type. The red line indicates the hydrophobic region of a signal sequence. (The figure taken from (Rapoport, 2007)).

The folding process of membrane proteins is still not clear. Based on the fact that most soluble proteins are able to form secondary structure in the aqueous phase prior to becoming compact, the forming of the TM helices may also occur before they integrate into the membrane lipid bilayer. Due to the energetic cost of forming the helical structure in the hydrophobic lipid environment, most likely the event takes place in the translocon environment. In fact, photo-crosslinking experiments captured a bound signal helix of about two turns near the translocon lateral exit site of translocon (Plath et al, 1998). The forming of helical structure before the integration is somehow necessary for the membrane proteins that could facilitate the translocon recognition. Thus, the translocon complex may orient the nascent helical segment according to the physico-chemical characteristics of the amino acids. For example, orient the nonpolar face of the helix to make contact with the lipid environment when they move into the membrane phase. Indeed, the orientation and position of the amino acids in the helix can serve as a coded signal whether the helix can integrate as a single helix or it needs to associate with the subsequent nascent polypeptide. Experimental studies showed that the TM helices made contact with the translocon and lipids during the synthesis (Meindl-Beinker et al, 2006). This contact will facilitate the integration and oligomerization of the membrane proteins. However, most details about the processes of the multi-spanning membrane protein oligomerization are still not clear. Most likely the tertiary structure formation does not take place in the protein conducting channel due to the size of the channel (8 Å) that seems to allow only a single helix to pass through (Berg et al, 2004)(Rapoport, 2007). In contrast, evidence from accessibility studies of fluorescence-quenching experiments showed that the aqueous pore in a functioning translocon is 40-60 Å in diameter (Hamman et al, 1997). Several low-

resolution studies of cryo-electron microscopy (cryo-EM) support this hypothesis (Hanein et al, 1996)(Beckmann et al, 1997)(Ménéret et al, 2000). A hypothetical model based on aquaporin presented a similar idea of having a central translocation pore, but formed by oligomeric translocon complexes (Pitonzo & Skach, 2006) (**FIGURE 1.7**). The central pore that contained reduced lipids could facilitate the membrane protein folding due to the increase area per lipid that correlates with the increase of water permeability (Mathai et al, 2010). The higher hydration in the lipid bilayer environment may energetically assist the process of membrane protein folding at least for the marginally hydrophobic TM helices.

Yet, there are also several studies that indicate that the translocation pore is not located at the centre of translocon complexes. For example (1) a back-to-back tetramer configuration of mammalian Sec61 complexes was obtained from cryo-EM experiments (Ménéret et al, 2005). Even though the four-subunits of the heterotetramer translocon complexes resulted to form a central pore, but the data suggest that the pore is rather depressed and could be filled with lipids (Ménéret et al, 2005). Moreover, the result showed that there was only one actively working Sec61 complex that directed the elongated precursor polypeptide. (2) In another study using cryo-EM, a ribosome-bonded SecYEG translocon complex from *Escherichia coli* was shown to form a dimer. Similarly, during the co-translocation process, only one translocon subunit laterally opened to allow the integration of the nascent polypeptide into the lipid bilayer (Mitra et al, 2005)(Driessen, 2005).

Additionally, there are numerous proteins that get associated with the translocon during the post-translocation process including translocation-associated membrane protein (TRAM) (Görlich et al, 1992), the translocation-associated protein complex (TRAP) (Wiedmann et al, 1987), the oligosaccharyl transferase (Görlich et al, 1992), the signal peptidase complex and the ER-luminal polypeptide chain binding protein (BiP) (Shibatani et al, 2005) (Dudek et al, 2005) (Dudek et al, 2009). These translocon-associated proteins are suggested to assist the lateral integration and folding of membrane proteins the endoplasmic reticulum membrane.

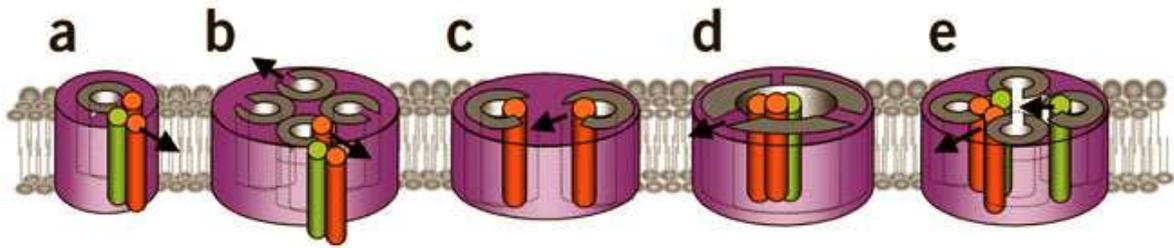


FIGURE 1.7 Models of translocon complex during the biosynthesis of membrane proteins. (a) The single and empty crystal structure of archaeal SecY complex (a homolog for Sec61 translocon complex) from *M. jannaschii* indicates that the diameter of the protein-conducting channel is about 8 Å and consists of a lateral exit site to a membrane lipid bilayer (Berg et al, 2004). The recent model of the ribosome-bonded Sec61 translocon (~11 Å) obtained from mass spectrometry also suggests the same illustration (Ménétrete et al, 2008). (b) The cryo-electron microscopy (cryo-EM) analyses showed a back-to-back tetramer configuration of Sec61 complexes (Ménétrete et al, 2005) but suggested that only one actively working Sec61 complex that directly elongates a precursor polypeptide. (c) Another cryo-EM study of SecYEG complex from *E. coli* observed a front-to-front dimer configuration of ribosome-translocon complexes. However, the opening of both SecY halves was shown not to generate one central channel. The data suggest that during the translocation of a hydrophilic segment of the nascent chain, only one of the SecY laterally opened to the bulk lipid (Mitra et al, 2005). (d) A model from fluorescence-quenching experiments showed a large central pore (40-60 Å) of functionally intact Sec61 translocon (Hamman et al, 1997). This model is supported by early low-resolution EM studies (Hanein et al, 1996)(Beckmann et al, 1997)(Ménétrete et al, 2000). (e) A hypothetical model of an oligomeric front-to-front configuration of translocon during the synthesis of multi-spanning membrane proteins that based on the topological analysis of the aquaporin 1 (AQP1). The figure was adapted from (Skach, 2009).

1.2.3 Non-constitutive membrane protein

The non-constitutive membrane proteins refer to foreign peptides as for examples antimicrobials and toxins. These peptides do not use the translocon machinery to cross the plasma membrane. There are several ways how they gain entry into the cell. For example, (1) a toxic protein, colicin forms a complex with the vitamin B₁₂ receptor, BtuB, and then recruits the OmpF porin to translocate its C-terminal toxic domain in order to kill the cell (Zakharov et al, 2006); (2) diphtheria toxins (a bacterial toxin) manage to enter and disrupt the membrane bilayer after triggering the V-ATPase proton pump protein that causes a decrease of the cell pH. The acidic environment induces the diphtheria toxin (B chain) to form a pore in the membrane that facilitates the entry of its toxic domain (A chain) into the cytoplasm (Rodnin et al), (3) melittin (bee venom), is an amphiphilic peptide that forms an α -helical structure when bound to the membrane and creates a pore in the membrane when the peptide concentration is increased (Raghuraman & Chattopadhyay, 2006). Altogether, these examples that all occur in nature infer that a single peptide alone is not independent to spontaneously get inserted into the membrane. The peptide needs to modify the normal physiological conditions of the cell (e.g. increase acidity) or/and able to manipulate the host proteins in order to gain entry into the membrane lipids.

Similarly, the spontaneous insertion of pH-low-insertion peptide (pHLIP) into the membrane is also triggered by low pH and by the protonation of Asp (there are Asp residues located in the center of the monomer and at the end of C-terminal). The pHLIP peptide showed to be soluble in aqueous phase, bind to the membrane as unstructured molecule and stably get inserted across the membrane as a α -helix monomer (Andreev et al, 2010). Hence, the two-stage model of (Popot & Engelman, 2000) is somehow still relevant to describe the entry pathway of the non-constitutive membrane proteins into the membrane lipids. We note that the case of pHLIP which contains a charged residue in the centre of the monomer and gets inserted at low pH condition is somehow similar to that of putative TM helices of envelope glycoproteins from the *Flaviviridae* virus family (Lindenbach et al, 2007). In contrast to other non-constitutive peptides, these viral envelope glycoproteins are integrated into the ER membrane through the translocon complex similarly as the constitutive membrane proteins (Cocquerel et al, 2002).

1.2.4 Biological Hydrophobicity Scale

The classical idea of Kyte and Doolittle suggests that sequence hydrophobicity is the main criterion that determines the location of TM segments (Kyte & Doolittle, 1982). Although many protein prediction methods are based on such hydrophobicity scales (von Heijne, 1992)(Rost et al, 1996)(Tusnady & Simon, 2001), it remains unclear exactly how to perfectly discriminate between TM segments and soluble segments. The present methods have an uncertain accuracy in predicting semi-hydrophobic TM segments.

The so-called biological hydrophobicity scale is based on a large-scale systematic experimental study of peptide-partitioning between the aqueous and membrane phase (**TABLE 1.1**). It describes the apparent free energy of insertion (ΔG_{app}) of each of the 20 natural amino acids when located in the centre of a TM helix (Hessa et al, 2005). This translocon-based scale has triggered a controversy because it strikingly showed that the energetic penalty for inserting charged residues (at pH=7) into the lipid bilayer when located centrally in the TM helix is not as high as expected from biophysical simulation studies (MacCallum et al, 2008)(Dorairaj & Allen, 2007)(Ulmschneider et al, 2007a). In fact, the experimental water-to-cyclohexane partitioning scale (Radzicka & Wolfenden, 1988) showed a much higher cost to bury those charged amino acids in the core region of the lipid bilayer. Interestingly, the biological hydrophobicity scale is in excellent agreement with the former Wimley-White hydrophobicity scale (water-to-POPC interface) and the Wimley-White octanol-based scale (water-to-octanol) (Wimley & White, 1996)(Wimley et al, 1996) (**TABLE 1.1**). There may be several reasons why the Wimley-White scales are correlated with the biological hydrophobicity scale: First, they used the whole residue scale and second, the water-octanol condition allowed a suitable environment for the partitioning of polar and nonpolar molecules.

TABLE 1.1 Solvation and transfer of Free Energies for acid amino (the data are in units of kcal/mol).

Residue	^a Peptide-Translocon	^b Water to POPC interface	^c Water to octanol	^d Water to cyclohexane	^e Generalized Born membrane
Asp	0.11	0.17	0.50	-1.81	0.11
Asn	2.05	0.42	0.85	6.64	3.55
Cys	-0.13	-0.24	-0.02	-1.28	0.25
Gln	2.36	0.58	0.77	5.54	3.39
Gly	0.74	0.01	1.15	-0.94	0.58
His	2.06	0.96	2.33	4.66	3.33
Ile	-0.60	-0.31	-1.12	-4.92	-0.87
Leu	-0.55	-0.56	-1.25	-4.92	-1.21
Met	-0.10	-0.23	-0.67	-2.35	0.38
Phe	-0.32	-1.13	-1.71	-2.98	-0.81
Pro	2.23	0.45	0.14	—	1.93
Ser	0.84	0.13	0.46	3.40	2.08
Thr	0.52	0.14	0.25	2.57	1.71
Trp	0.30	-1.85	-2.09	-2.33	1.12
Tyr	0.68	-0.94	-0.71	0.14	0.84
Val	-0.31	0.07	-0.46	-4.04	-0.59
Arg	2.58	0.81	1.81	14.92	11.42
Asp	3.49	1.23	3.64	8.72	23.61
Glu	2.68	2.02	3.63	6.81	27.23
Lys	2.71	0.99	2.80	5.55	20.85

^aPeptide-Translocon-based scale (Hessa et al, 2005); ^bWater-POPC interface whole residue (Wimley & White, 1996); ^cWater-octanol whole residue (Wimley et al, 1996); ^dWater-cyclohexane (Radzicka & Wolfenden, 1988); ^eGeneralized Born implicit-membrane (Ulmschneider et al, 2007a).

What could be the possible factors that reduce the cost of inserting marginally hydrophobic TM helices into the lipid bilayer? None of the current computational methods are able to clearly explain the low cost of desolvating the charged residues Arg, Lys, Asp and Glu according to the translocon-based hydrophobicity scale. The translocon-based biological hydrophobicity scale has triggered a challenge to scientists in the computational biophysical field in seeking data to explain these discrepancies. Computer simulation studies by (1) (MacCallum et al, 2008) showed that Lys, Asp and Glu adopted neutral protonation states in the core region of the lipid bilayer, but this work did not give a definitive result about the charged status of Arg; (2) (Dorairaj & Allen, 2007) simulated a long polyLeu helix with an Arg in the center and claimed that the free energy barrier for the Arg in the core bilayer might be as high as 17 kcal/mol. They suggested that Arg changes to a neutral state in the core region of membrane bilayer. This is somehow contrary to the presence of charged residues in the integral membrane proteins that is very likely required for their specific functions. For example, the S4 helix of voltage-gated ion

channels contains four Arg that are exposed to lipids and must remain charged to function (Long et al, 2005) (Hessa et al, 2005). Therefore, it is not surprising that the putative TM domain of the envelope glycoproteins from Flaviviridae virus family contain at least one charged residues located in the centre of their TM helices. In hepatitis C virus, these charged amino acids were shown experimentally to play multiple roles including ER retention, dimerization and in the virus assembly (Ciczora et al, 2007). A change from a charged to the neutral state will obviously defeat the purpose of their function.

One possible explanation for the discrepancy between the in vivo biological hydrophobicity scale and the computer simulations is the use of pure nonpolar membrane bilayers in the simulations, whereas in nature the cell membrane consists of heterogeneous components. For example it contains many different types of lipids (Johansson & Lindahl, 2009b) and integral membrane proteins (Johansson & Lindahl, 2009a). MD simulation results showed that adding more proteins in the membrane bilayer reduced the solvation energy for the charged Arg close to the experimental observation (3-5kcal/mol) (Johansson & Lindahl, 2009a). One of their findings was that the increase of the protein content in the cell membrane facilitates the TM helices to retain a certain degree of water hydration in the hydrophobic milieu. The contact between the translocon and TM helices allows the translocon to recognize the TM helices thus facilitating the integration process. Furthermore, helix-helix interaction can significantly assist the insertion of marginally hydrophobic TM helices into the lipid bilayer. The dimerization of H-segments containing polar Asn and Asp residues located in the center of the TM segment was shown to result in a low ΔG_{app} (Meindl-Beinker et al, 2006). Moreover, the putative TM segment of envelope glycoprotein E1 and E2 from HCV also has been suggested based on the experimental data to form a salt bridge before integrating to the lipid bilayer (Ciczora et al, 2007).

1.3 Membrane Lipids

Compartmentalization by biological membranes defines eukaryotic cells and organelles, prokaryotic cells and even some viruses by separating them from the aqueous solution. Thus, this principal allows each cell to perform its independent intracellular processes. By becoming the boundary of the cells, the membranes take control to regulate the transport of substances into and out of the cells. They are differentially permeable to only small substances such as water, oxygen and carbon dioxide that may diffuse through. In contrast, charged ions for example sodium and potassium, and bigger molecules require special carrier proteins or channels to actively transport them across the membranes. Big molecules such as proteins also can enter or leave cells by being incorporated into vesicles in the

processes called endocytosis and exocytosis, respectively. These responses of different mechanisms of membranes to perform specific tasks provide evidences that they are not only just passive boundaries!

According to the fluid mosaic model that was first proposed in 1972 by Singer and Nicolson (Singer & Nicolson, 1972) (Singer, 1974), biological membranes consist of heterogeneous components which are lipids, proteins and carbohydrates (**FIGURE 1.8**).

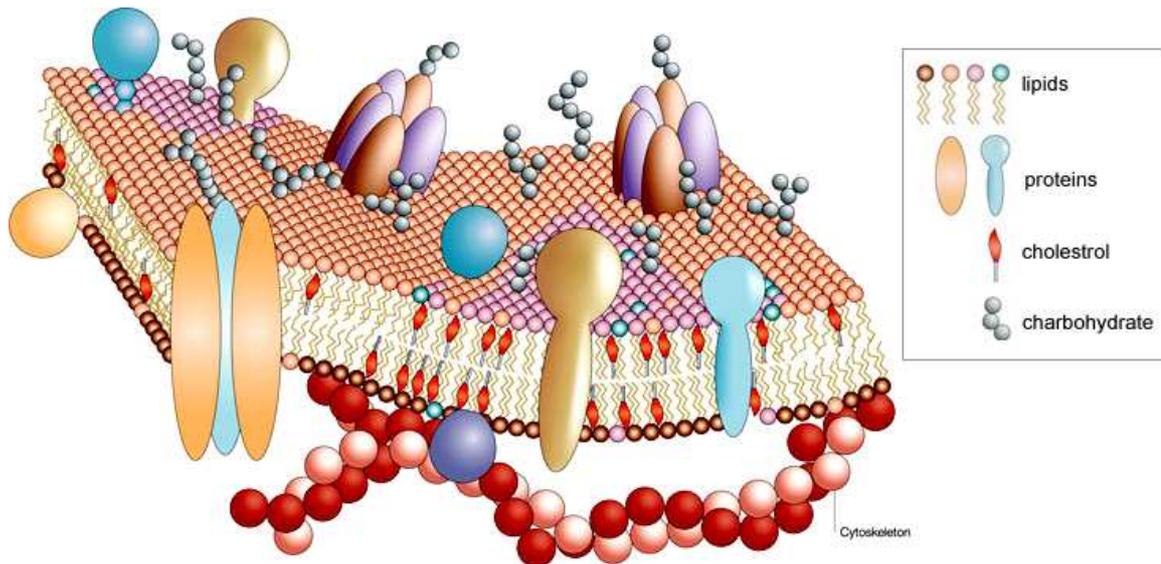


FIGURE 1.8 General model for membrane structure is based on the Singer-Nicolson ‘fluid mosaic model’ (Singer & Nicolson, 1972). The figure was adapted from (Pietzsch, 2004).

Lipid molecules are the backbone of the biological membranes. There are more than 1000 types of lipid species in cell (Sleight, 1987)(Lev, 2010). For example, erythrocytes alone contain about 100 types (Lipowsky & Sackmann, 1995) Phospholipids (glycerophospholipids) are the most abundant lipids in eukaryotic membranes. The phospholipids comprise of a glycerol backbone, which facilitates a high variability of different head groups and acyl chain combinations. The main head group classes are the phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS) and phosphatidylinositol (PI). PS and PI are the charged types of head groups. The acyl chains typically vary from 16 to 22 carbons in length and they contain 0 to 6 double bonds. In addition to the phospholipids, sphingolipids and sterols (cholesterol) are also important components in membranes. The basic unit of the membrane is a bilayer formed by phospholipids and sphingolipids. Most membranes have an asymmetric distribution of lipids in their two leaflets. The outer leaflet of the mammalian plasma membrane consists mainly of sphingolipids, PC and cholesterol, but the inner leaflet has a higher concentration of negatively charged phospholipids that can lead to a surface potential on the cytosolic site of cell membranes (Devaux & Morris, 2004). This surface potential can

affect local ion concentration as well as membrane protein conformation and function. The physico-chemical asymmetry across the bilayer leaflet is dynamically modulated according to cell specific processes. For example the increase of PS type of lipids in the outer leaflet determines that a cell is entering apoptosis (Kagan et al, 2000).

Lipids are amphiphilic molecules which consist of two distinct regions; a water soluble (hydrophilic), and a water-insoluble (hydrophobic). Due to this characteristic, they have a tendency to self-assemble in a water environment. The hydrophilic head group region tends to interact with water molecules and the hydrophobic acyl chain region prefers to pack with each other to reduce contacts with the water molecules. This results in an aggregation of a specific form of lipid assembly that is determined by the physico-chemical properties of the lipid molecules. The cone-shape lipids form inverted hexagonal structures, the cylindrical shaped ones form bilayers and the inverted cone-shaped lipids form micelles (**FIGURE 1.9**). It is important to note that these resulting shapes are caused by external condition for instance temperature, hydration or counterions (Tresset, 2009). Under normal physiological condition, hydrated phospholipids typically form bilayer structures.

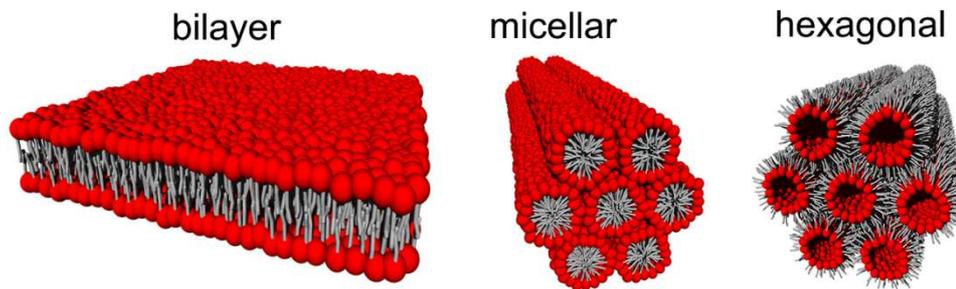


FIGURE 1.9 Phospholipid assemblies: (a) bilayers, (b) micellar hexagonal, and (c) inverted hexagonal structures. These images are adapted from (Tresset, 2009).

1.4 Flaviviridae Virus family

Flaviviridae (from the Latin word flavus which stands for, “yellow”) are named after the first human virus discovered over one century ago, the yellow fever virus (Strode, 1951). This large family of viral pathogens is responsible for causing severe diseases and mortality in humans and animals. It consists of three genera (see **FIGURE 1.10**): Flavivirus, the largest group among other Flaviviridae genres that currently has more than 70 members, classified for vector-borne disease agents such as dengue virus (DENV), japanese encephalitis (JEV), yellow fever virus (YFV) and West Nile virus (WNV);

Pestivirus is the animal pathogen (5 members), for example bovine viral diarrhea virus (BVDV); and Hepacivirus consists of the hepatitis C virus (HCV) that causes hepar-related diseases, cirrhosis and hepatocellular carcinoma in infected individuals. HCV virus includes at least 6 genotypes and numerous subtypes. GB virus A (GBV-A), GBV-B and GBV-C are shown by nucleotide and protein analyses to be most closely related to HCV than to other members of the Flaviviridae family. Therefore, they are grouped together in the Hepacivirus genus (**FIGURE 1.10**). The similar features of the viruses in Flaviviridae are the virion morphology, the genome organization and the replication strategy. Besides these similarities, however, each genus expresses distinct effects and therefore these viruses cannot be treated using the same drugs.

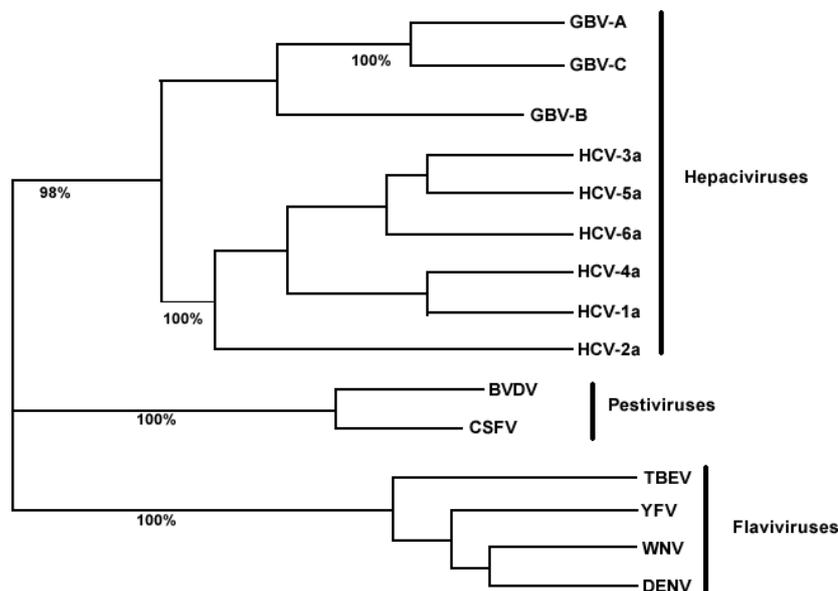


FIGURE 1.10 The Flaviviridae virus family. Shown is a phylogenetic tree based on sequence comparisons of NS3 helicase regions. Example of the flaviviruses: yellow fever virus (YFV); dengue virus (DENV), West Nile virus (WNV), and tick borne encephalitis virus (TBEV); the pestivirus genus: bovine viral diarrhea virus (BVDV) and classical swine fever (CSFV); hepacivirus genus: hepatitis C virus (HCV) including GB virus B (GBV-B); GB virus A (GBV-A) and C (GBV-C). (The figure was taken from (Thomas et al, 2005)).

At the time of writing, there are already successfully developed vaccines available against YFV, JEV and TBEV. However, there is no effective antiviral drug or even vaccines available against the DENV and HCV. The HCV antivirals are due to be released in the middle of year 2012 leucopenia (Opar, 2010). However, the new antiviral drugs, which are targeting the NS3B protein, need to be coupled with the current drugs in order to effectively function. The use of Ribavirin and pegylated interferon caused side-effects for patients such as anemia and leucopenia (Opar, 2010). Therefore, ongoing experiments need to find better solutions to treat the HCV disease.

The viruses of Flaviviridae gain entry into the cell via endocytosis after binding to the cellular receptors that are specific for viral envelope glycoproteins (Murray et al, 2008). The low pH during the endosomal pathway induces fusion between the virion envelope and the cell membrane, thus causing the nucleocapsid to uncoat and release the RNA genome. There are several experimental evidences showing that the envelope glycoproteins of these viruses are involved in the fusion stage (Lindenbach et al, 2007). Interestingly, the released virus genome is recognized by the host cell as a messenger RNA. Therefore, it uses the constitutive pathway similar to the biogenesis of membrane proteins in order to replicate (**FIGURE 1.11**).

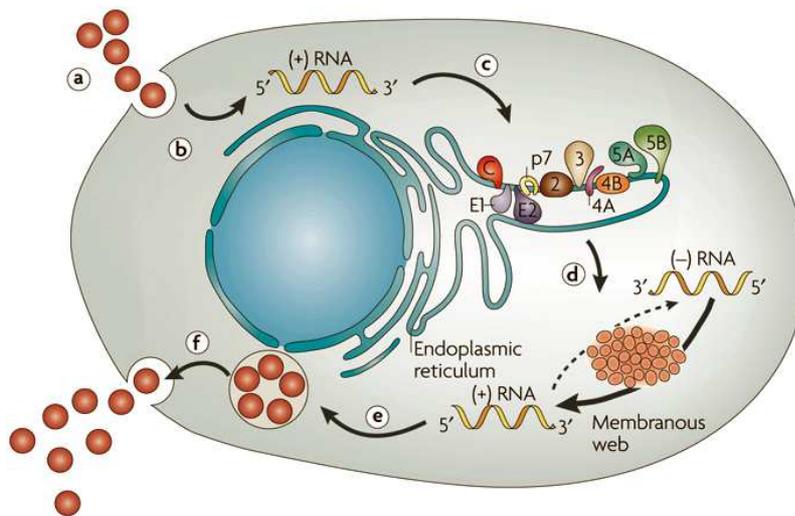


FIGURE 1.11 Processing of the hepatitis C virus. Members of Flaviviridae family likely utilize the same mechanism in order to replicate in host cells. (a) They are suggested to enter the host cellular membrane via endocytosis. (b) Then, the low pH induces fusion of the virion envelope with the cellular membrane. Followed by the uncoating of the nucleocapsid, the RNA genome is released into the cytoplasm. (c) The RNA is processed by the host translocon machinery resulting in ~3000 amino acids of polyprotein (d) RNA replication - Processing by viral and cellular enzymes releases the individual viral gene products. (e) Packaging and assembly (f) Virion maturation and final release via exocytosis. (The figure was taken from (Moradpour et al, 2007).

1.4.1 Transmembrane Domains of the Envelope Glycoproteins

The envelope glycoproteins of the family Flaviviridae have TM domains that are composed of two stretches of hydrophobic residues separated by a short segment that contains at least one fully conserved charged residue (Cocquerel et al, 2000) (**FIGURE 1.12**). The TM domains act as signal for the ER retention, are responsible for the dimerization of the envelope glycoproteins and essential for the formation of the viral envelope.

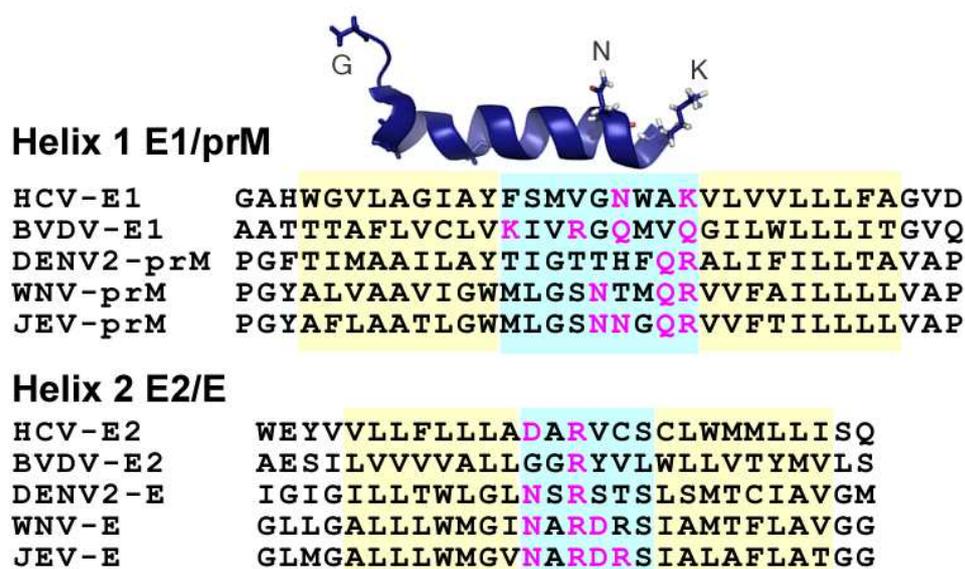


FIGURE 1.12 Sequence alignment of the putative TM domains of the envelope proteins from the Flaviviridae virus family. The domain composed of two hydrophobic segments (yellow background) separated by a short polar segment (blue background) contains highly conserved polar and charged residues (highlighted in pink). The segment of E1-HCV (1EMZ) is shown on top of the alignment. To date, this is the only three-dimensional structure available representing the envelope glycoproteins of Flaviviridae virus family. This structure of E1 (350-370) was obtained by NMR-spectroscopy constraint measured in 50% TFE (2,2,2-trifluoroethyl-1-2-d₂ alcohol).

1.4.2 TM Domain of the E1-E2 Dimer of Hepatitis C Virus

In this section, the envelope glycoproteins of HCV will be discussed in more detail due to their important role in this dissertation project. We selected this system because of the abundance of experimental data that could support our computational study of TM helix dimerization. Briefly, the HCV was discovered in the year 1989. Unfortunately, compared to the other well known members of Flaviviridae family for instant dengue and yellow fever viruses, the study of HCV lagged behind

because it is extremely difficult to propagate the HCV genome in culture. Only recently, the development of HCV pseudoparticles (HCVpp) (Bartosch et al, 2003) and the efficient amplification of cell culture system (HCVcc) (Wakita et al, 2005) have contributed to major advances in investigating the functions of HCV (Ciczora et al, 2005)(Ciczora et al, 2007).

The RNA genome of HCV contains a single long open reading frame (ORF) of approximately 9600 nucleotides encoding for a single polyprotein of about 3200 amino acids. The ORF between the 5'-noncoding region (NCR) and 3'-NCR is composed of three structural (core, E1, E2) and seven non-structural (p7, NS2-NS5B) proteins. The 5'-NCR contains an internal ribosome entry site (IRES) required for the translation of the HCV genome (Bartenschlager et al, 2004) (Penin et al, 2004). The E1 and E2 envelope glycoproteins were suggested to be responsible for the viral entry by binding to the host cell receptor (Bartenschlager et al, 2004). Their putative TM domains are an extreme example of a multifunctional membrane-spanning sequence. These domains consist of fewer than 30 amino acid residues and are composed of two hydrophobic stretches separated by a short segment containing at least one fully conserved charged amino residue (**FIGURE 1.12**). The charged residues in the centre of the TM domain have been shown to be responsible for the ER retention and the heterodimerization of the glycoproteins (Ciczora et al, 2007)(Lindenbach et al, 2007).

1.4.3 Model of the E1-E2 Dimer Biogenesis

Is the hydrophobicity signal the sole criterion used by the SRP to decide the fate of the precursor polypeptides? If yes, it must be a precise value needed to determine whether the nascent polypeptide must be co- or post-translationally processed. It is possible that marginally hydrophobic segments contain an extra signal to get selected co-translationally. Interestingly, the E1 and E2 envelope glycoproteins contain not only a signal that directs the biogenesis processes co-translationally but also a signal which determines the topology of the glycoproteins. It has been shown that hydrophobic sequences located at the C-terminal of both TM segments contains a signal sequence that is responsible for the translocation of the protein located downstream. The C-terminal half of E1 is involved in the translocation of the ectodomain of E2, and the C-terminal half of E2 is involved in the translocation and integration of p7 polyprotein (Reed and Rice, 2000)(Cocquerel et al, 2002). Due to the influence of the signal sequences, the topology of E1 and E2 envelope glycoproteins adopts that of type I TM proteins with an N-terminal ectodomains and a C-terminal hydrophobic anchor (N_{lum}/C_{cyc}) (cyt, cytosol; lum, luminal) (Cocquerel et al, 2002). Since the ectodomain of E1 and E2 are translocated into the lumen of the ER, they were suggested to adopt a hairpin structure. The experimental evidences suggest that the formation of the hairpin structure occurs before the signal sequence cleavage (Cocquerel et al, 2002).

Then, after the cleavage, the C-terminus of the TM domain is reoriented to form a single TM spanning segment (Op De Beeck et al, 2000) (Cocquerel et al, 2002) (**FIGURE 1.13**).

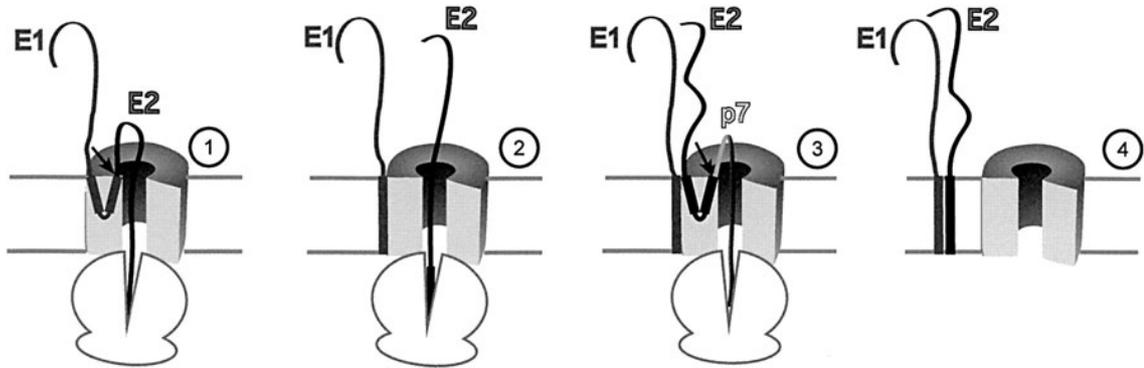


FIGURE 1.13 Model of synthesis of the E1-E2 envelope glycoproteins of HCV. (1) The N-terminus of E1 is translocated into the lumen of the ER as well as its C-terminal half which contains the signal sequence of E2 and adopts a hairpin structure. (2) After the signal sequence cleavage between E1 and E2, the C-terminal half of the TM domain of E1 is reoriented toward the cytosol resulting in a single TM spanning segment. (3) Similarly, the TM domain of E2 transiently adopts a hairpin structure to allow the translocation of p7. (4) After the signal sequence cleavage between E2 and p7, the signal sequence present in the C-terminal half of the TM domain of E2 is reoriented toward the cytosol. The TM domains of E1 and E2 form a dimer and laterally integrate into the lipid bilayer. The arrows denote the cleavage sites by the ER signal peptidase. Figure adapted from (Cocquerel et al, 2002).

Chapter 2

Computational Methodology

2.1 Molecular Dynamics Simulations

The molecular dynamics (MD) method allows atomistic descriptions of biological components evolving in time based on classical mechanics. It was first introduced by Alder and Wainwright in the year 1957 (Alder & Wainwright, 1957)(Alder & Wainwright, 1959) to study the transition of solid-fluid phases of hard spheres. Their discovery set the stage for the development of MD as a basic tool in statistical mechanics. The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon (Rahman, 1964) and for liquid water in 1974 (Stillinger & Rahman, 1974). The first protein simulation was performed in 1977 for the bovine pancreatic trypsin inhibitor (BPTI) (McCammon et al, 1977). Today, with the rapid development in the computer technology, the MD method is routinely used in a wide range of molecular research to obtain information about structural, dynamical and thermodynamic properties of complex biomolecular components as well as in chemistry and material science. Examples are the binding of ligands to soluble proteins, membrane protein aggregation in the lipid phase, interaction of protein-DNA complexes as well as the role of MD to aid in conformational sampling in experimental procedures such as X-ray crystallography and NMR structure determination.

At the present time, the MD methods are conveniently compiled as powerful software packages that can be executed on small desktop PCs up to parallel supercomputers depending on the system size and simulation length that is aim for. Popular MD software packages are CHARMM (Chemistry Harvard Macromolecular Mechanics (MacKerell et al, 1998), GROMACS (Groningen Machine for Chemical Simulations) (Hess et al, 2008)(Van Der Spoel et al, 2005), GROMOS (Groningen Molecular Simulation) (Oostenbrink et al, 2004) and NAMD (Nanoscale Molecular Dynamics) (Phillips et al, 2005). Each package usually uses its own developed force field which is a set of interaction parameters and semi-empirical rules to evaluate forces between different types of atoms in the simulation system. Perhaps the best known force fields are AMBER (Ponder & Case, 2003), CHARMM (MacKerell et al, 1998), GROMOS (Oostenbrink et al, 2004) (Berendsen et al, 1995)and OPLS (Jorgensen et al, 1996).

2.2 Methods

A rigorous MD simulation must be based on a mathematical model that correctly describes the energy of a system as a function of its structure. Ideally, structure and dynamics of molecules could be determined from electron and nuclei combinations by numerically solving the time-dependent Schrödinger equation.

$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{H}\Psi \quad (2.1)$$

In the Schrödinger equation (equation. 2.1), \hat{H} is the Hamiltonian operator, Ψ is the wave function, and \hbar is the reduced Planck constant. It describes the full dynamics of a molecular or atomic system based on the principles of quantum mechanics. Unfortunately, it is still far beyond the capacities of modern computers to apply the time-dependent Schrödinger equation to systems with more than 10 atoms. Therefore, the Born-Oppenheimer approximation is used to simplify the problem. The idea is based on the fact that the electron mass is significantly smaller than the mass of nuclei and therefore allows for their movement to be independently computed. In this scheme, the electron configuration is relaxed for every configuration of the fixed nuclei. Then, the nuclei are propagated according to a mean-field approximation. The Newton's second law of motion is used to replace the time-dependent Schrödinger equation which results in the *ab initio* MD method that is based on the movement of electron or the classical MD method that is based on the position of nuclei. The classical MD method uses parameterized analytical potentials (derived by fitting to quantum mechanical models or experiments) and allows the treatment of systems up to sizes billions of atoms (Roth et al, 2000).

Newton's equation of motion is

$$F = ma = m \frac{d^2 r}{dt^2} = -\Delta V(r_1, \dots, r_N) \quad (2.2)$$

where F_i is the force, acting on the i -th particle, m_i is the mass and a_i is the acceleration of the i -th particle. For this purpose we need to calculate the force, F_i acting on the atoms, and these are derived from the potential energy $U(r_N)$, where $r_N = (r_1, r_2, \dots, r_N)$ represents the complete set of $3N$ atomic coordinates.

An integration scheme is required to determine the time evolution of the atomic positions and velocities. Given the positions and velocities at $t = 0$ the positions and velocities at a later time t can be

obtained using a suitable algorithm. The Verlet algorithm is one of the simplest numerical methods used to integrate Newton's equations of motion. This popular integrator offers great stability as well as time-reversibility and energy conservation properties. It can be derived by writing a simple Taylor expansion

$$r(t + \Delta t) = r(t) + \Delta t \frac{d}{dt} r(t) + \frac{1}{2} \Delta t^2 \frac{d^2}{dt^2} r(t) + O(\Delta t^3) \quad (2.3)$$

where Δt is an integration time step, r is the position at a given time t , and O represents the terms of order three and higher. In the equation of motion, the first and second derivatives of position can be replaced with a velocity, v and acceleration, a respectively. By summing the respective Taylor expansion for $r(t - \Delta t)$ and truncating $O(\Delta t^3)$, the Verlet algorithm is revealed (Verlet, 1967).

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{1}{2} \frac{F}{m} \Delta t^2 \quad (2.4)$$

The verlet algorithm uses the positions and accelerations at the time t and the positions at the time $t - \Delta t$ to predict the positions at the time $t + \Delta t$, where Δt is the integration step. The velocities are obtained from the basic definition of differentiation:

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} \quad (2.5)$$

where acceleration is substituted using Newton's equation of motion to give force divided by mass. However, there is an error of the order of Δt^2 . Therefore, to obtain more accurate velocities, the Leap-Frog algorithm (Hockney & Eastwood, 1988) is used, using velocities at half time step

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + (\frac{F(t)}{m})\Delta t \quad (2.6)$$

The velocities at time t can be also computed from

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} \quad (2.7)$$

The atomic positions are then obtained from:

$$r(t + \Delta t) = r(t) + v(t + \frac{\Delta t}{2})\Delta t \quad (2.8)$$

Equivalent to the leap-frog algorithm, the velocity Verlet can yield the position, velocity and acceleration with the same step. However, the velocity Verlet is more computationally expensive because of the increased calculation required per time step.

2.2.1 Molecular Force Fields

To simulate a molecular system, a computational model must be developed to determine the energy of a system according to the current positional of all particles accounting for bonded interactions in polyatomic molecules and for interactions among the particles. The interactions are described by a force field, which constitutes a set of functions that sum up the potential energy of the system. The basic functional form of the force field consists of both intramolecular forces which describe covalently bonded interactions and intermolecular forces describing non-bonded interactions.

$$V(r_1, \dots, r_N) = E_{bonded} + E_{non-bonded} \quad (2.9)$$

The intramolecular forces are bonded interactions and consist of a bond stretching (2-body), bond angle (3-body), and dihedral angle terms (4-body) (**FIGURE 2.1**).

$$V_{bonded} = V_{bonds} + V_{angles} + V_{imp.dih} + V_{pro.dih} \quad (2.10)$$

The bonded energy is the total sum of the different interaction energies that are defined by connectivity. This stretching energy between a pair of bonded atoms is based on the Hookean spring model

$$V_{bond}(b_{ij}) = \sum 1/2 K_{ij}^b (b_{ij} - b_{ij}^0)^2 \quad (2.11)$$

where K_{ij}^b is the force constant which controls the stiffness of the bond spring, b_{ij}^0 is the equilibrium bond length, and b_{ij} is the actual bond length between atoms i and j . This equation estimates the energy associated with the vibration about the equilibrium bond length. The term describing the bending energy is represented by an angle formed by three atoms and is also based on Hooke's law.

$$V_{angle}(\theta_{ij}) = \sum 1/2 K_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2 \quad (2.12)$$

where K_{ijk}^θ is the angle-bending force constant that controls the stiffness of the angle spring, θ_{ijk}^0 is the equilibrium bond angle, and θ_{ijk} is the actual angle formed by atoms i , j and k . Thereby, this equation estimates the energy associated with the vibration about the equilibrium bond angle. The proper dihedral angle is defined as the angle formed by four atoms linearly bonded together. The torsion energy is modeled by a simple periodic function

$$V_{proper}(\varphi_{iklj}) = \sum 1/2 K_{ijkl}^\varphi (1 + \cos(n\varphi_{iklj} - \varphi^0)) \quad (2.13)$$

with K_{ijkl}^φ is the torsional barrier, φ_{iklj} represents the angle between the two planes shared by atoms i , j , k and j , k , l . In organic chemistry, the angle relates to the trans state at 180° and gauche at $\pm 60^\circ$. n

describes the periodicity of the energy function, and φ^0 is a reference torsional angle that defines the positions of the energy minima. Lastly, the improper dihedral angle depends on the position of three atoms centered around a fourth atom. It is used to enforce both planar and chiral conformations.

$$V_{improper}(\zeta_{ijkl}) = \sum 1/2 K_{ijkl}^{\zeta} (\zeta - \zeta^0)^2 \quad (2.14)$$

K_{ijkl}^{ζ} is the energy constraint, ζ is the actual angle between two planes, and ζ^0 is the reference angle.

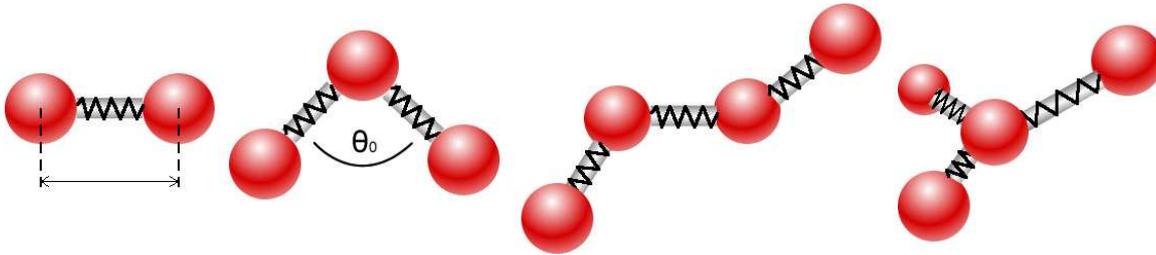


FIGURE 2.1 Schematic representation of bonded interactions. (a) Bond stretching between two bonded atoms, (b) angle bending (1,3 interaction), (c) proper dihedral and (d) improper dihedral (1,4 interaction).

The intermolecular forces are describing interactions between separate atoms. This non-bonded interaction is represented by the Lennard-Jones (LJ) and Coulomb potentials that sum all the energies of all possible interacting non-bonded atoms i and j

$$V_{non-bonded} = E_{LJ} + E_{Coulomb} \quad (2.15)$$

The Lennard-Jones potential (also known as 6-12 potential) describes the van der Waals interaction that represents a combination of repulsion (Pauli repulsion) and attraction (London dispersion) between a pair of non-bonded atoms. The repulsion occurs when distances between interacting atoms become smaller than the sum of their contact radii and causes a repulsive force proportional to $1/r^{12}$, where r is the interparticle distance. However, when the two atoms are beyond a certain distance, favorable dipole interactions are induced, which pull the atoms closer together. This attraction decays quickly with $1/r^6$. The full Lennard-Jones potential is thus written as:

$$V_{LJ} = \sum 4\epsilon_{ij} ((A_{ij}/r_{ij})^{12} - (B_{ij}/r_{ij})^6) \quad (2.16)$$

where A_{ij} is the short-range repulsive term coefficient, B_{ij} is the attractive term coefficient and r_{ij} is the actual distance between atoms i and j . The A parameter is obtained from atomic polarizability, or it can be calculated quantum mechanically. The B parameter is typically derived from crystallographic data. The second term of the non-bonded interactions is the electrostatic interaction between two charged atoms expressed by the Coulomb potential,

$$V_{Coulomb} = \sum 1/4\pi\epsilon_0\epsilon_r \cdot q_i q_j / r_{ij} \quad (2.17)$$

where q_i and q_j are the effective partial charges for atoms i and j , respectively. ϵ_0 is the permittivity of vacuum, ϵ_r is a relative dielectric constant and r_{ij} is the distance between atoms i and j (**FIGURE 2.2**). If all atoms of the system including the solvent are modeled explicitly, ϵ_r equals to 1.



FIGURE 2.2 Schematic representations of the non-bonded interactions. (a) Lennard-Jones potential and (b) Coulomb potential.

2.2.2 Periodic Boundary Condition

The limitation of the available computer speed sets the limit for the size of the system that can be simulated over useful lengths. An artificial boundary must exist to truncate the number of molecules due to the finite number of simulated atoms. However, these boundaries cause surface effects due to neglecting the interactions with particles beyond the boundaries. To reduce these effects, periodic boundary conditions are frequently used (Allen & Tildesley, 1989) in MD simulations to mimic an infinite bulk system. Consequently, molecules at one edge of the system will interact with molecules at the opposite edge.

2.2.3 Treatment of Long Range Interactions

In a large system with N atoms, there are much fewer bonded interactions to be computed in a simulation because the number of bonded interaction is linearly dependent on N . In contrast, the computation of non-bonded interactions costs much more computation time because each atom can potentially interact with all the other atoms. Therefore, the number of non-bonded interaction scales as N^2 . Due to the limited computing capacity and the finite size of the simulation cell, the interaction of the non-bonded atoms must be in practice truncated. In former times, this was generally solved by using cut-offs method. The interactions are calculated only up to a certain distance and after this cut-off distance the potential of the interaction goes to zero. The size of the simulation cell restricts the cut-off distance due to the periodic boundaries condition where the molecule would interact with two and more

copies of another molecule if the cut-off distance is too large. Thus, the cut-off distance has to be smaller than half of the shortest unit vector of the simulation cell (Allen & Tildesley, 1989). For the Lennard-Jones potential, the short-range interaction (r^{-6}) can be cut well below the symmetry distance. Unlike the Lennard-Jones interactions, the Coulomb potential of electrostatic interactions (r^{-1}) cannot be easily cut-off due to their long-range nature. Using the same cut-off method leads to serious artifacts. These effects can be reduced by shifting the force function to zero or by switching the force smoothly off at the cut-off radius. Nowadays, it is standard practice to include the long-ranged Coulombic interactions by employing a variant of the Ewald-type summation techniques that compute the energy in infinite lattices. In this work, the Particle Mesh Ewald (PME) technique (Essmann et al, 1995) was used. It is an improvement version of the original Ewald summation method which converted the sum over all pairwise interactions into real-space and Fourier space components. Because of using the fast Fourier transform (FFT) algorithm (Darden et al, 1993) (Essmann et al, 1995) it scales as $N \log N$.

2.2.4 Simulation in the Isothermal-Isobaric Ensemble

A standard MD simulation simply solving Newton's equation of motion is performed in the microcanonical ensemble where the energy and volume of the system are constant (NVE-ensemble). In such a simulation, all states have the same energy as the starting configuration. No relaxation processes to states at lower energy or transitions over energy barriers can be observed. However, most real experiments are performed at constant pressure and temperature instead at constant volume and at constant energy. In that case, the distribution of states of different energy is governed by the Boltzmann distribution. Therefore, in order to make the simulation correspond to the real experiment, one has to characterize the macroscopic equilibrium state by keeping the state parameters temperature and pressure of the modeled system to given values. In this work, pressure was always set to ambient pressure (10^5 Pa) and the temperature was set to different values between 310 and 323 K. In MD simulation, this condition is called the isothermal-isobaric ensemble or NPT-ensemble, where the number of atoms, the temperature and the pressure are kept constant. There are many methods to keep the temperature constant, one of which is the Berendsen algorithm that mimics a weak coupling to an external heat and pressure bath (Berendsen et al, 1984). In the Berendsen scheme, the temperature scaling method is essentially a direct scaling of the particle velocities but it is softened with a time constant. When coupled to a heat bath, the actual system temperature, T_0 is corrected according to

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T} \quad (2.18)$$

where t is the time, T_0 is the desired temperature and τ_T is the time constant. So the temperature deviation decays exponentially with τ . In practice, the temperature is adjusted by scaling the velocity of all particles at each time step. This results in a change of the kinetic energy, $E_{kinetic}$

$$\Delta E_{kinetic} = \sum_i^N \frac{1}{2} m (\lambda v_i^2) - \sum_i^N m v_i^2 \quad (2.19)$$

$$= (\lambda^2 - 1) \frac{1}{2} N_{df} k_b T \quad (2.20)$$

where N_{df} is the number of degrees of freedom and k_b is Boltzmann's constant. Equation 2.20 is based on the equipartition theorem; where the system temperature, T can be determined and is defined by the sum of the kinetic energy; $\Delta E_{kinetic} = \sum_i^N \frac{1}{2} m v_i^2 = \frac{1}{2} N_{df} k_b T$. The scaling of the kinetic energy corresponds to a temperature change

$$\Delta T = \frac{\Delta E_{kinetic}}{C} \quad (2.21)$$

Therefore, the scaling factor can be derived as

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t)} \right) - 1} \quad (2.22)$$

with the time step Δt and the temperature coupling time constant τ_T .

Similarly to temperature, also the pressure can be scaled. The pressure is controlled by scaling the coordinates and vectors of the simulation cell at every time step. For the isothermal-isobaric ensemble condition, the pressure must first be defined for a MD simulation

$$PV = N_f k_b T + \frac{1}{3} \sum_{i=j}^N r_{ij} F_{ij} \quad (2.23)$$

where V is the volume of the system and N_f is the number of degrees of freedom. $r_{ij} = r_i - r_j$ and F_{ij} is the force on the particle i due to particle j . Using the Berendsen pressure coupling method, the pressure, P , is scaled similarly as the temperature towards the given reference pressure P_0 .

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \quad (2.24)$$

A pressure change can be done by changing the virial through scaling of interparticle distances. The scaling matrix is given by the elements

$$\mu = 1 - \frac{\Delta t}{3\tau_p} \beta_{ij} (P_{ij}^0 - P_{ij}) \quad (2.25)$$

where τ_p is the pressure coupling time and β is the isothermal compressibility of the system. The appropriate time constant for pressure is 0.1ps or larger. A smaller time constant leads to instability of the algorithm with increased pressure and volume fluctuations.

Particle coordinates r_i are scaled as

$$\dot{r}_i = \mu r_i \quad (2.26)$$

The equations 2.25 and 2.26 are used for an anisotropic system. It can be reduced to an isotropic system if $\mu = \mu l$.

2.2.5 Simulation of Peptides in Lipid Bilayer Systems

Due to the complexity of the membrane composition, it is very difficult to experimentally characterize the membrane properties at microscopic scale. Over the years, NMR and X-ray diffraction experiments (Petrache et al, 1998)(Nagle & Tristram-Nagle, 2000) and simulations studies of membrane lipids have been done side by side. This resulted in the possibility to construct theoretical models for lipid bilayers. The current available force fields for simple lipid bilayers are able to successfully reproduce structural experimental properties such as density, heat of vaporization and chain order (Gompper & Schick, 2008). The lipid parameter set known as ‘Berger lipid’ that was constructed from a combination of GROMOS-87 bonds, angles, and dihedral (Ryckaert-Belleman dihedral for the chains), OPLS for LJ-interactions (with Berger’s adjustment for the chains), and partial charges from the work by Chiu et al has been widely used for various phospholipid simulations because it provide good results that reproduce the experimental data (Berger et al. 1997).

In order to sample the correct or realistic statistical motion of the physiological membrane system, the simulation temperature of the bilayer system needs to be set above the melting temperature to a liquid-crystalline state, L_α . Above the melting transition, lipids exist in a fluid and disordered state. The thickness of the bilayer decreases and the area per molecule increases. In contrast, below the melting temperature, a pure phospholipid bilayer is in a gel phase, which is characterized by a high chain order. For instance the chains tend to orient parallel to each other. At the physiological temperature, most abundant phospholipid species form the L_α phase. However, DPPC is one type of the lipids in biological membranes prone to exist in a gel phase at the physiological temperature. In gel phase, the highly ordered hydrocarbon chains tend to form a cone-shaped structure. The transition

temperature from L_α state to L_β of DPPC occurs at 42°C (315 K). The DPPC and DMPC are common lipid species which have been used in numerous experimental (Nagle & Tristram-Nagle, 2000) and computational (Tieleman et al, 1997) studies of lipid systems. Both of these lipids were used in the simulation works for this thesis (FIGURE 2.3).

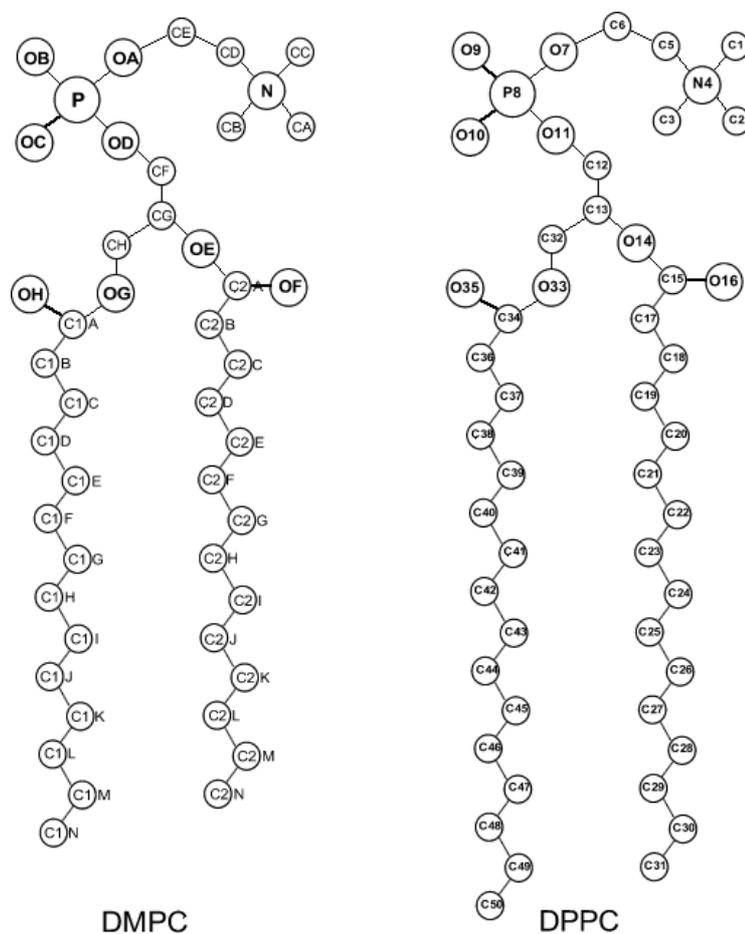


FIGURE 2.3 Common lipid types used in the MD simulations of membrane lipid bilayer systems. (Left) DMPC and (Right) DPPC lipids are shown in the united-atom representation. The CH_2 and CH_3 groups of lipid alkyl chains are modeled as united-atoms. The DMPC has slightly shorter acyl chains than DPPC but they have a different melting-transition temperature.

The isothermal-isobaric ensemble is the most appropriate for the MD simulation of a lipid bilayer that resemble the real experimental condition. The constant volume ensemble (NVE-ensemble) is not suitable for the bilayer simulations due to fluidity problem (Tieleman and Berendsen, 1996). It is also necessary to enforce the surface tension (γ) of the bilayer to a fitted value along with the normal pressure on the bilayer ($N\gamma$ PT ensemble). Otherwise, the small lipid bilayer simulations run in NPT ensemble generally do not approach the expected equilibrium state when $\gamma = 0$. In order to overcome the abnormal behavior of the particles near the boundaries, periodic boundary condition (PBC) is

implemented. For simulations in the lipid bilayer system, the usage of PBC means simulating an infinite stack of alternating layers of lipid and water.

The choice of the water model is closely related to the force field parameters used for the lipid model. In the GROMACS simulation software, the Berger lipid (Berger et al, 1997) has been implemented for MD simulations of lipid bilayers. The most recommend water model in combination with the Berger lipids is SPC (simple point charge) water model (van Buuren et al 1993). This is because the special reduced LJ-interactions between the water oxygen, O_w , and the CH_2/CH_3 groups have been optimized for the Berger lipids (Berger et al, 1997). To avoid serious unwanted artifacts caused by the PBC, a sufficient number of water molecules are necessary to simulate lipid bilayers. Generally, there should be at least about 32 waters per lipid for simulations of phospholipids with phosphocholine polar groups (Gompper & Schick, 2008). For our works that involve the study of peptides in a lipid bilayer system, the GROMOS FF53A6 force field was employed (Oostenbrink et al, 2004) in all simulations. This recently developed force field parameter has been specifically parameterized to reproduce free enthalpies of solvation in water and cyclohexane. The values correspond well to the experimental data of the 20 amino acid residues. Therefore, it has been recommended as one of the most suitable force fields to date for studies on protein folding (van Gunsteren et al, 2006). Additionally, Poger et al have shown that the use of the FF53A6 force field provides excellent agreement of lipid bilayer properties with the experimental data (Poger et al, 2010) (Poger & Mark, 2010).

Embedding of Peptides in Lipid Bilayer

There are several protocols to prepare the starting configuration for the simulation of peptides in a lipid bilayer system. The coordinates of the peptide or protein of interest can be obtained from a crystal structure, an NMR structure or an engineered model. The coordinates for the lipids can be constructed by several softwares as, for instance Visual Molecular Dynamics (VMD) (Humphrey et al, 1996). But, due to the high viscosity of fluid lipid, a quite long time is needed to equilibrate the lipids. Therefore, several methods were suggested in order to reduce the equilibration time of peptide-lipids system in MD simulations. One example is to use previously equilibrated pure lipid bilayers which can be obtained from several trusted webs.

In this thesis, we followed the protocol that was established by (Faraldo-Gómez et al, 2002) that used the equilibrated lipid bilayers as the starting configuration. The main purpose of this method is to prepare an optimized cavity that suits the peptide of interest. This can reduce the equilibration time of the system as well as providing an optimal interaction between the peptide and the surrounding lipid molecules. At the initial stage, several lipids that centrally overlap with the peptide are removed. This can be done by (1) simply putting both coordinate files together, (2) manually check with any molecular

visual graphic software and (3) manually deleting overlapping lipid molecules. The solvent-accessibility protein surfaces of the peptide was computed by the MSMS program (Sanner et al, 1996) using the probe size radius of 1.4 Å. The resulting cavity was used as an input for the MDRUN program of the modified version of GROMACS 3.1.4 package to create the optimized cavity of the peptide. Then the coordinates of the lipid bilayer are combined with the coordinate file of the peptide. This coordinate file is used as a starting structure for the production run (**FIGURE 2.4**).

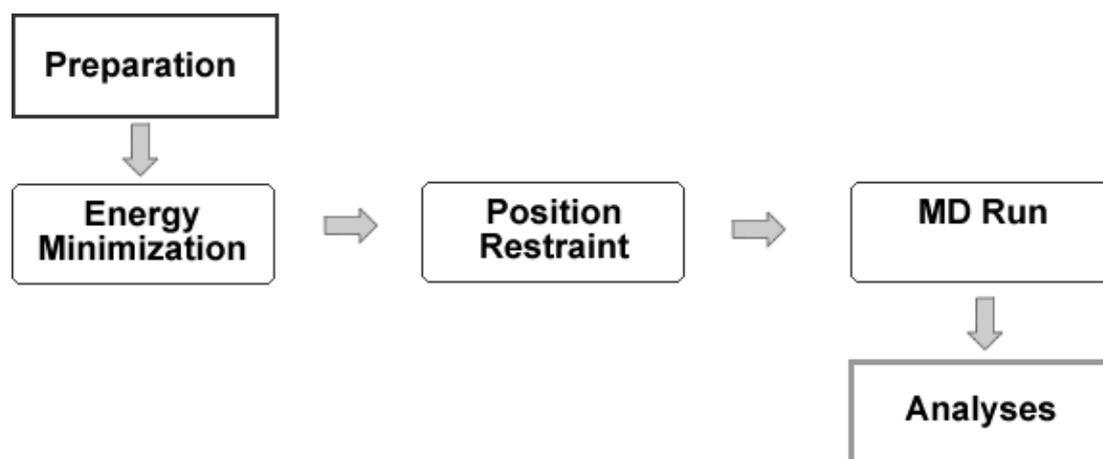


FIGURE 2.4 Schema of MD simulation process of peptide in lipid bilayer system based on GROMACS software.

Equilibration and Diffusion

Simulations of pure lipid molecules in bulk water will take approx. 10 ns to 100 ns to form a bilayer phase if starting from random solution. The equilibration of DPPC lipid bilayer in MD simulation was shown to be about 23 ns (Marrink et al, 2001). However, the equilibration of the peptide-lipid bilayer system is not easy to validate due to the absence of experimental data for the transfer free energy of lipid bilayer (Sapay & Tieleman, 2008). Although, some peptides were shown to equilibrate in nanosecond time scale, there is a high possibility that the straightforward MD is unable to cross high energetic barriers. For example, simulations of small synthetic peptides in a DOPC lipid bilayer system in 50 ns time scale were showed to be insufficient to reproduce the experimental data (Aliste & Tieleman, 2005).

The lateral diffusion coefficient of lipid bilayers in liquid phase measured from experiments is approximately $1.27\text{-}1.52 \times 10^{-7} \text{ cm}^2/\text{s}$. In simulations, the value obtained from 288 lipid molecules was $0.95 \times 10^{-7} \text{ cm}^2/\text{s}$ and $2.92 \times 10^{-7} \text{ cm}^2/\text{s}$ 72 for a lipid system (Sundararajan, 2008). MD simulations of lipid bilayers to study the diffusion of benzene in DMPC lipids were shown to provide favorable results

as that of the experimental data in about 2 ns. The rate of the diffusion for benzene molecules was found to be higher when near to the bilayer core compared to the head group region (Stouch, 1993), and can be increased in higher temperature (Bassolino-Klimas et al, 1993). However, in the case of diffusion of a peptide, the sizes of the macromolecules need to be taken into consideration.

2.2.6 Analyses of Simulations

For pure lipid bilayer systems, MD simulation methods have come to the stage where they can reproduce the experimental values. Therefore, several standard analyses are being used to compare the results from a simulation with the experimental data.

Deuterium Order Parameter

The order parameter of the lipid tails in MD simulations can be compared to values obtained from NMR experiments. From the simulations, the value can be calculated from the average fluctuation over the equilibration time based on the order parameter tensor that measures the spatial restriction of the motion of a CH vector.

$$S_{ij} = \frac{1}{2} \langle 3 \cos \theta_i \cos \theta_j - \delta_{ij} \rangle \quad (2.27)$$

where θ_i is the angle between the i th molecular axis and the bilayer normal. The brackets indicate an ensemble average. Then the deuterium order parameter S_{CD} can be calculated from

$$S_{CD} = -\frac{2}{3} S_{xx} - \frac{1}{3} S_{yy} \quad (2.28)$$

Both acyl chains of the lipid are computed separately. The experimental value of S_{CD} for DPPC lipids is 0.20 ± 0.02 (Nagle, 1993) that was computed based on the 4th to 8th CH₂ group. For comparison **FIGURE 2.5** shows the results from a recent simulations study. Towards the end of the lipid tails, the order parameter drops towards zero, demonstrating no preferential orientation.

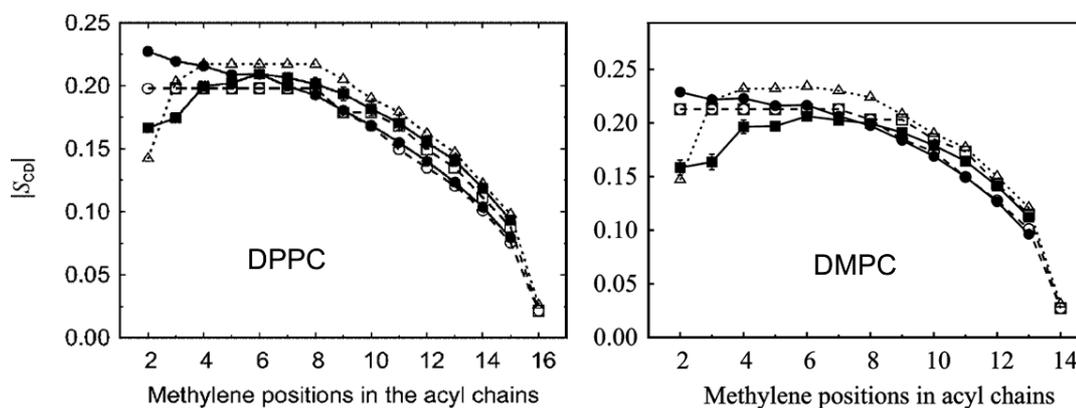


FIGURE 2.5 Deuterium order parameter of DPPC (Poger et al, 2010) and DMPC (Poger & Mark, 2010) lipids from recent simulation studies using the GROMOS FF53A6 force field.

Bilayer Thickness

The thickness of the bilayer reflects the ordering of the lipid acyl chains and thus to the state of the particular type of lipids. In an MD simulation, the bilayer thickness can be computed by averaging the headgroup-to-headgroup thickness of the bilayer at each time step based on the center of mass coordinates of each lipid head group.

Area per Lipid

The conventional approach to compute the area per lipid is to divide the total area of the simulation box by the number of lipids in one monolayer and subtract the space occupied by the solute.

TABLE 2.1 Comparing data for pure lipid bilayer properties between experiments and simulations.

Lipid bilayer type	Bilayer thickness, Temp. (nm)		Area per lipid (nm ²)	
	Experiment	Simulation	Experiment	Simulation
DMPC	^c 3.53 (30 °C) ^b 3.60 (30 °C)	^d 3.46 (30 °C)	^c 0.606 (30 °C) ^b 59.5-67.6 (30 °C)	^d 0.642 (30 °C)
DPPC	^b 3.83 (50 °C)	^a 3.60 (52 °C)	^b 0.633-0.729 (50 °C)	^a 0.600 (52 °C) ^c 0.655 (52 °C)

^a(Tieleman & Berendsen, 1996); ^b(Nagle & Tristram-Nagle, 2000); ^c(Kučerka et al, 2005); ^d(Griepner et al, 2007); ^e(Patra et al, 2004).

2.2.7 Current State of Molecular Dynamics Simulations of Membrane Systems

The X-ray crystallography technique provides the highest quality of protein structure among the available structural determination methods. But, for the integral membrane proteins the crystallization is typically carried out in detergent solutions which do not resemble their physiological environment. Thus, MD simulation of lipid bilayer system may provide a means to study the behavior of integral membrane proteins in an environment of a lipid bilayer. Moreover, the static data alone does not reveal the functional dynamics of biological processes. In contrast, the MD simulations allow monitoring the detailed motion of each molecule of a system on fast time scales which are not accessible by experiment. The MD methods have been successfully employed to study the ion selectivity permeation and the gating mechanism in potassium ion channels for examples the bacterial K channel (KcsA) (Shrivastava & Sansom 2000), the voltage-gated K channel (KvAP) (Monticelli et al, 2004), and the inward rectifier K channel (Kir) KirBac1.1 (Domene et al, 2004). Similarly, in the study of aquaglyceroporins, a standard protocol of MD simulations of POPE bilayer system managed to observe water and glycerol permeation through the human water channel aquaporin-1 (AQP1) and the homologous bacterial glycerol facilitator (GlpF), respectively (de Groot & Grubmuller, 2001). These dynamic events of ion, water and glycerol permeations through the pore domain of membrane proteins were obtained in approximately 10 ns of simulation time and were in good agreement with the experimental rate of ion permeation (Sansom et al, 2002). Commonly, MD simulation has been employed to characterize experimental structures atomistically. For the M2 helix of the nicotinic acetylcholine receptor (naAChR), MD simulation reproduced a similar structure as was illustrated by solid state NMR data of the kinked TM helix caused by the central Leu (Law et al 2000). In the field of drug engineering, the study of peptide-lipid interaction is crucial to be fully described by the all-atom MD simulation in the bilayer system (Phil and Sansom 1999). The engineered antimicrobial peptides were shown to diffuse into the lipid bilayer in approximately 30-50 ns MD simulations (Shepherd et al 2003)(Aliste & Tieleman 2005). Altogether, these evidences illustrate that the MD simulation method is a powerful method to gain knowledge about the structure and functions of membrane proteins in their natural lipid bilayer environment.

However, similarly to lab bench experiments, computational methods also have limits. The MD simulation which is based on the classical approximation cannot reproduce quantum effects as for example the formation or breaking of bonds. Moreover, the simplified partial charges that are required for the potential functions do not guarantee to reproduce the exact experimental data (Tieleman & Berendsen, 1996)(Tieleman et al, 1997). The straightforward MD simulation method is also prone to cause the protein to get stuck in local energy minima due to the high energy barriers.

Nowadays, MD methods are advancing rapidly with the continuous development of computer technologies. New dedicated hardwares and enhancement of computational algorithms have increased the simulation time up to the microsecond scale (Freddolino et al, 2008). Therefore, we can expect to reach the millisecond time scale soon. This is the time where most of the exciting biological processes occur for example the folding of soluble proteins. Computational methods promise great benefits to the pharmaceutical industry. The importance of the membrane proteins for this industry has lead to the method developments that can be optimized for membrane proteins. For example the coarse-grained MD simulations of membrane proteins in lipid bilayer was recently shown to produce reliable results and therefore this method could be one of the promising methods to study the folding membrane of protein (Sansom et al, 2008). There are also extensions for the classical MD simulation such as the replica exchange and the umbrella sampling methods that are now optimized for the membrane proteins that can improve the conformational sampling and compute the free energy, respectively (Nymeyer et al, 2005)(Chetwynd et al, 2010).

In summary, based on these increasing method developments, MD simulation may soon become a routine procedure in biology, chemistry and physic for varied purposes. Collaboration efforts between experimentalist and computational biophysicists will speed up the MD approach to reach the level of experimental accuracy.

Chapter 3

Contribution of Charged and Polar Residues for the Formation of the E1-E2 Heterodimer from Hepatitis C Virus

Published in Journal of Molecular Modeling, 16 (10): 1625-1637, 2010

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.

3.1 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 (Appel et al, 2006). 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 (Moradpour et al, 2007), particularly the HCV pseudoparticles (Bartosch et al, 2003) (HCVpp) and the first system for efficient production of infectious viral particles in cell culture (Wakita et al, 2005) (HCVcc).

HCV is the only member of the *Hepacivirus* genus which belongs to evolutionary related viruses of the *Flaviviridae* family (Francki et al, 1991)(Lindenbach et al, 2001). 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 (Matsuura & Miyamura, 1993). 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.

Binding and internalization of the HCV 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 (Lindenbach et al, 2001) and assemble as a non-covalent E1-E2 heterodimer which is retained in the endoplasmic reticulum (ER) (Dubuisson et al, 2000). These two membrane proteins are type I transmembrane (TM) proteins which are composed of a large N-terminal ectodomain towards 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 (Cocquerel et al, 2000). Interestingly, two consecutive GxxxG motifs are known within the TM of E1. The presence of the GxxxG motif in glycophorin A (GpA), a membrane protein of erythrocytes, at the helix-helix packing interface is known to be involved in the GpA homodimerization (Cuthbertson et al, 2006). 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 (Dubuisson et al, 2000)(Ciczora et al, 2007), e.g. virus entry (Ciczora et al, 2007), ER retention, as an internal signal peptide and E1-E2 heterodimerization (Cocquerel et al, 2002).

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 (Ciczora et al, 2007)(Ciczora et al, 2005). 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 (Ronecker et al, 2008). 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 (Ciczora et al, 2007).

Despite their relative abundance in the protein-coding regions of different genomes (25-30 %), 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 (White, 2004). Yet, molecular dynamics (MD) simulations of membrane proteins embedded in lipid bilayers have become quite popular and successful in the last ten years (Cuthbertson et al, 2006)(Bond & Sansom, 2003). In particular, MD simulations were applied to study the spontaneous aggregation of phospholipids around membrane proteins (Böckmann & Caffisch, 2005) or have been used to investigate the relative position of individual TM helices in lipid bilayers (Lomize et al, 2006) and their dynamic interactions with phospholipid bilayers (Matthews et al, 2006). For example, the structure of the Glycoprotein A (GpA) dimer was computationally predicted (Treutlein et al, 1992), including results from an extensive mutagenesis work (Lemmon et al, 1992) to narrow the search. The prediction was later refined, using an improved global search method (Adams et al, 1996). The subsequently determined NMR structure of the GpA dimer in micelles (MacKenzie et al, 1997) was in good agreement with the predicted structure. Furthermore, MD simulations were used to study the behavior of individual helices of bacteriorhodopsin (Woolf, 1998), the oligomerization of the helices of Vpu (Candler et al, 2005)(Fischer & Sansom, 2002), the free energy for dimerization of GpA (Hénin et al, 2005), and the protonation equilibrium of Arg residues within a TM helix (Yoo & Cui, 2008).

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.

3.2 Methods

3.2.1 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/>) (Wu et al, 2006). The E1 sequence used in this study is G³⁵⁰ AHWGVLAGIA³⁶⁰ YFSMVGWAK³⁷⁰ VLVVLLLFAG³⁸⁰ VDA. The E2 sequence is WAIKWEYVV⁷²⁰ LLFLLADAR⁷³⁰ VCSCSLWMMLL⁷⁴⁰ ISQAEA. 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 (Hessa et al, 2005). 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).

3.2.2 Sequence Analysis

This part of the project was carried out by Dr. Christoph Welsch from the Johann Wolfgang Goethe University, Frankfurt. Sequences of HCV envelope proteins were retrieved from public HCV databases, UniProtKB and euHCVdb (<http://www.euHCVdb.de>) (Combet et al, 2007). HCV genotypes have been differentiated according to a consensus proposal for a unified system of HCV genotype nomenclature (Simmonds et al, 2005). Sequence alignments were computed using CLUSTAL W (Larkin et al, 2007) and MUSCLE (Edgar, 2004), and subsequently improved by minor manual modifications using the SEAVIEW alignment editor (Galtier et al, 1996). 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.

3.2.3 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 (Rost et al, 1996), SPLIT4 (Juretic et al, 2002), HMMTOP2.0 (Tusnady & Simon, 2001), TMHMM (Krogh et al, 2001), and TMMOD (Kahsay et al, 2005). MINS2 (Park & Helms, 2008a) was used to predict the membrane insertion free energy of the TM domains of E1 and E2.

3.2.4 Molecular Dynamics Simulations

All structures used in this study were prepared as ideal α -helices. The SCWRL program (Canutescu et al, 2003) was used to position the side chain rotamers and to generate mutants. Gromacs (Hess et al, 2008) 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 3.1**) 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 (Griepner et al, 2007) 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 (Faraldo-Gómez et al, 2002). The solvent-accessible protein surfaces of the peptides were calculated by the program MSMS (Sanner et al, 1996) 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 centre 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 (Berendsen et al, 1995) 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 approximately 45 water molecules per lipid molecule, thus ensuring full hydration of the membrane (Siu et al, 2008). 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 $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) 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 (Hess et al, 2008). All monomer and dimer simulations were performed with united atom force field based on GROMOS96 (53a6) (Oostenbrink et al, 2004) for the peptides and the Berger force field (Berger et al, 1997)(Chiu et al, 1995) 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} (Berendsen et al, 1984). 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 (Hess et al, 1997) 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 grid spacing of 0.12 nm and cubic interpolation.

Analyses of the trajectories were primarily performed with tools included in the Gromacs 4.0.3 suite (Hess et al, 2008)(Berendsen et al, 1995). 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 3.1**). Helix centers of mass were computed using the coordinates of C α atoms only for the segments 5-25 (E1) and 35-55 (E2). All images in this work were prepared with the Pymol program (<http://pymol.sourceforge.net>).

TABLE 3.1 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. 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.

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

3.3 Results

3.3.1 Sequence Analysis of the TM Domain of the E1 and E2

This part of the project was carried out by Dr. Christoph Welsch from the Johann Wolfgang Goethe University, Frankfurt. 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.

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).

3.3.2 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 **FIGURE 3.1**). 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 (Park & Helms, 2008b) 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 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) when using Lys370 as center, and a border-line value of 4.3 kcal/mol for E2 (4.3 kcal/mol) when using Asp728 as center.

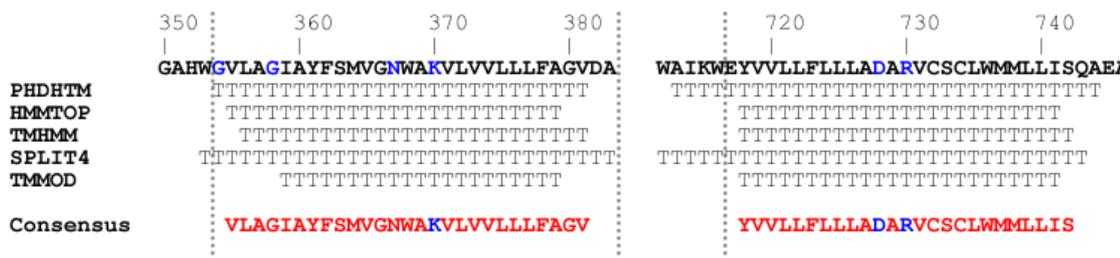


FIGURE 3.1 Results from Secondary Structure Prediction Programs. The consensus prediction is given at the bottom; the positions in the consensus sequence indicate that three or more methods gave the same results. Highlighted in blue in the consensus prediction are the charged residues Lys370, Asp728 and Arg730. The dotted lines show the segments which were used in the MD simulations.

3.3.3 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 towards the lipid bilayer interface. Only Lys370 was able to comfortably anchor to the interfacial region without affecting the helix stability (**FIGURE 3.2**). Due to its shorter side chain, Asp728, which is positioned in the centre 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 **FIGURE 3.3**) thus confirming our observations and providing further evidence that they are caused by the charged Lys370 and Asp728 residues.

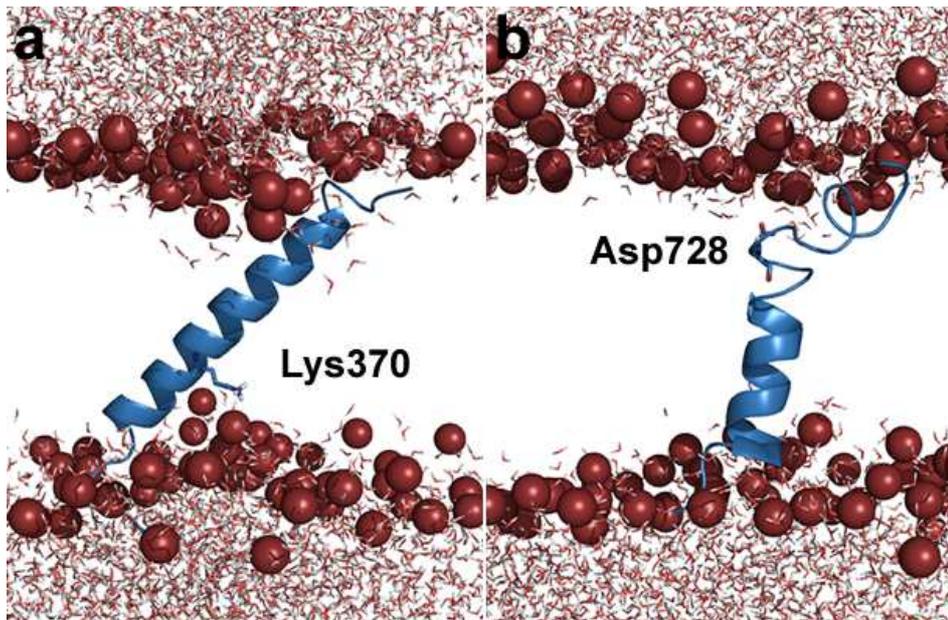


FIGURE 3.2 Final snapshots of MD simulations from the E1 and E2 monomers: (a) E1 TM segment with a charged Lys370, and (b) E2 TM segment with a charged Asp728. Lipid tails and ions are not shown for clarity. The charged Lys370 and Asp728 are shown as stick representation.

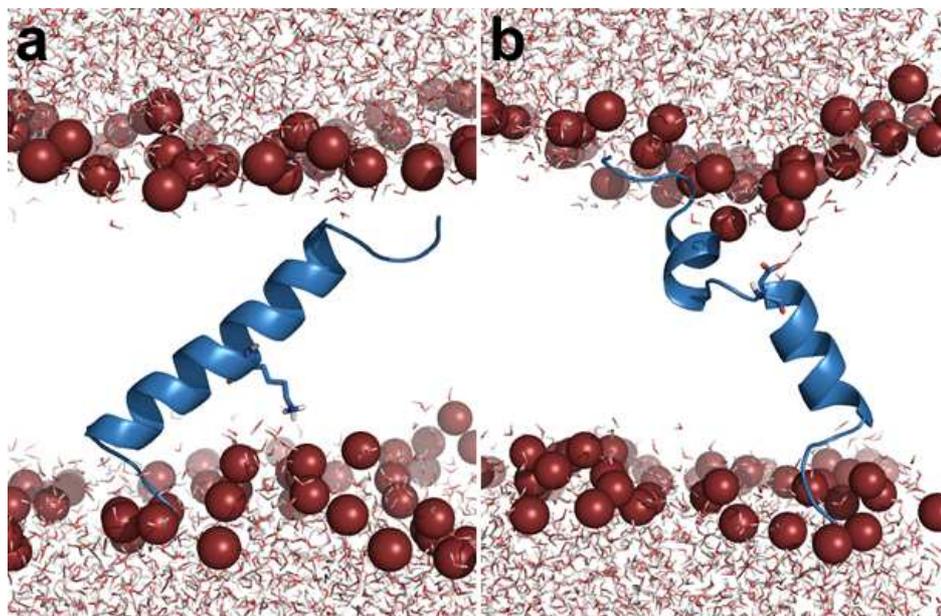


FIGURE 3.3 Final snapshot after 100 ns of MD simulations of the H-segment monomers containing a charged residue in the middle of their TM domains: (a) H-segment with a charged Lys and (b) H-segment with a charged Asp. Lipid tails and ions are not shown for clarity.

3.3.4 MD Simulations of E1-E2 Heterodimers

FIGURE 3.4 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 (G³⁵⁴VLAGIA³⁶⁰ YFSMVGWAK³⁷⁰ VLVVLLLFAG³⁸⁰VD). The E2 TM domain was observed to contain 27 residues between two polar residues at both N- and C-termini (EYVV⁷²⁰ LLFLLADAR⁷³⁰ VCSCSLWMMLL⁷⁴⁰ 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.

	360	370	380		720	730	740
E1-E2							
WILDTYPE 1	GVLGAIYFMSVGNWAKVLVLLLFAGVDA				EYVLLFLLADARVCSCSLWMMLLISQAEA		
WILDTYPE 2	*****				*****		
R730K MUTANT	*****				*****K*****		
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	GVLGAIYFMSVGNWAKVLVLLLFAGVD-				EYVLLFLLADARVCSCSLWMMLLISQ---		
TM Prediction Methods	-VLAGIAYFMSVGNWAKVLVLLLFAG---				-YVLLFLLADARVCSCSLWMMLLIS----		

FIGURE 3.4 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.

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 (**FIGURE 3.5**). **FIGURE 3.6** 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 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.

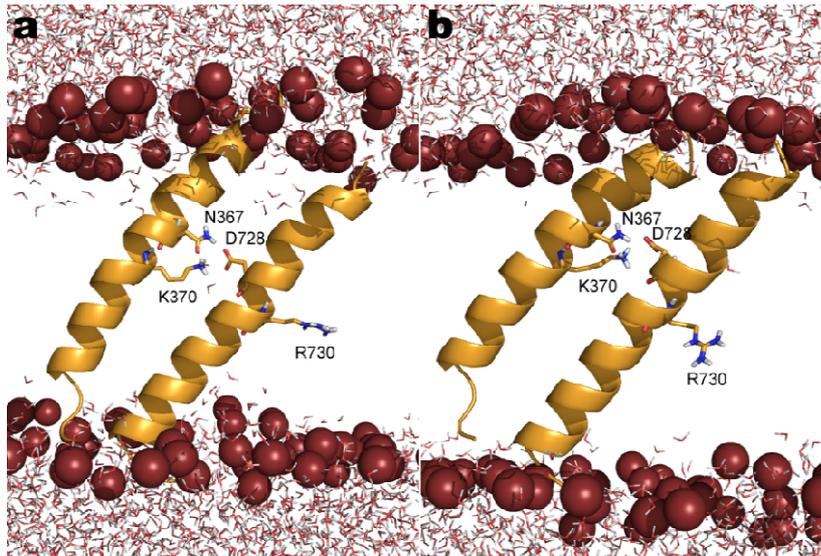


FIGURE 3.5 Final snapshots after 100 ns MD simulation of the E1-E2 heterodimers in the two wild-type simulations. The conserved residues Asn367, Lys370, Asp728 and Arg730 are highlighted as stick presentation. Lipid tails and ions are not shown for clarity.

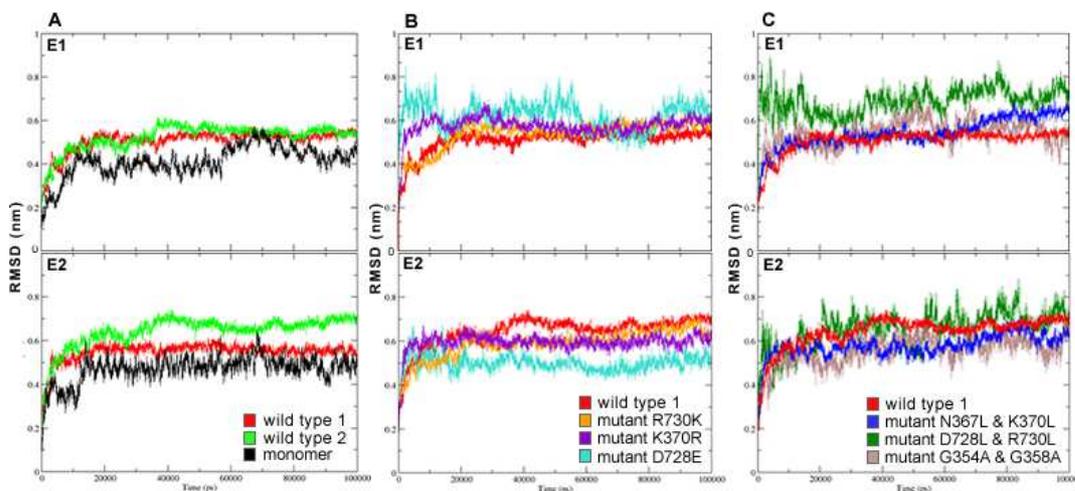


FIGURE 3.6 Root-mean-square deviations (RMSDs) of E1 and E2 TM domains of the E1-E2 wild-type and mutant heterodimers. (A) RMSDs of wild-type E1-E2 heterodimers versus the E1 and E2 simulations of isolated helices. (B) RMSDs of single mutants which contain a salt-bridge at the helix-helix packing and (C) RMSDs of double mutants. In (B) and (C), the E1-E2 wild-type 1 is shown for comparison.

The heterodimerization was clearly mediated by the salt bridge interaction of the charged Lys370 and Asp728 at the helix-helix interface. **TABLE 3.1** 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 3.2**). 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.

TABLE 3.2 Average H-bonds analyzed for the data between 80 and 100 ns of MD simulations of E1-E2 wild-types and mutants.

Residues	Wild type 1	Wild type 2	R730K	G354A & G358A	K370R	D728E	K370A
367-728	0.89 ± 0.47	1.02 ± 0.42	0.88 ± 0.41	0.80 ± 0.48	0.52 ± 0.51	0.98 ± 0.19	0.72 ± 0.47
370-728	0.02 ± 0.12	0.00 ± 0.04	0.03 ± 0.18	0.12 ± 0.34	0.78 ± 0.89	0.88 ± 0.45	0
367-370	0.90 ± 0.37	1.00 ± 0.32	0.70 ± 0.48	0.91 ± 0.56	1.42 ± 0.98	0.98 ± 0.33	0.01 ± 0.08
728-730	0	0	0	0	0	0	1.00 ± 0.56
367-730	0	0	0	0	0	0	0.49 ± 0.50
370-730	0	0	0	0	0	0	0
367,370-728,730 ^a	0.91 ± 0.48	1.02 ± 0.42	0.92 ± 0.44	0.91 ± 0.56	1.30 ± 0.94	1.86 ± 0.47	1.21 ± 0.72
TM ^b	1.85 ± 0.53	1.98 ± 0.46	1.53 ± 0.65	1.48 ± 0.72	2.30 ± 0.97	3.00 ± 0.63	3.00 ± 1.14

^a H-bonds interactions among the four residues; ^b H-bonds interactions between two helices.

3.3.5 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 (**FIGURE 3.7**) as for wild-type that are stabilized by an ion-pair interaction when started from a salt-bridged conformation. This behavior can be expected 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 3.1**). Also, the RMSD values are of similar magnitude than those of the wild-type simulations (**FIGURE 3.6**). 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.

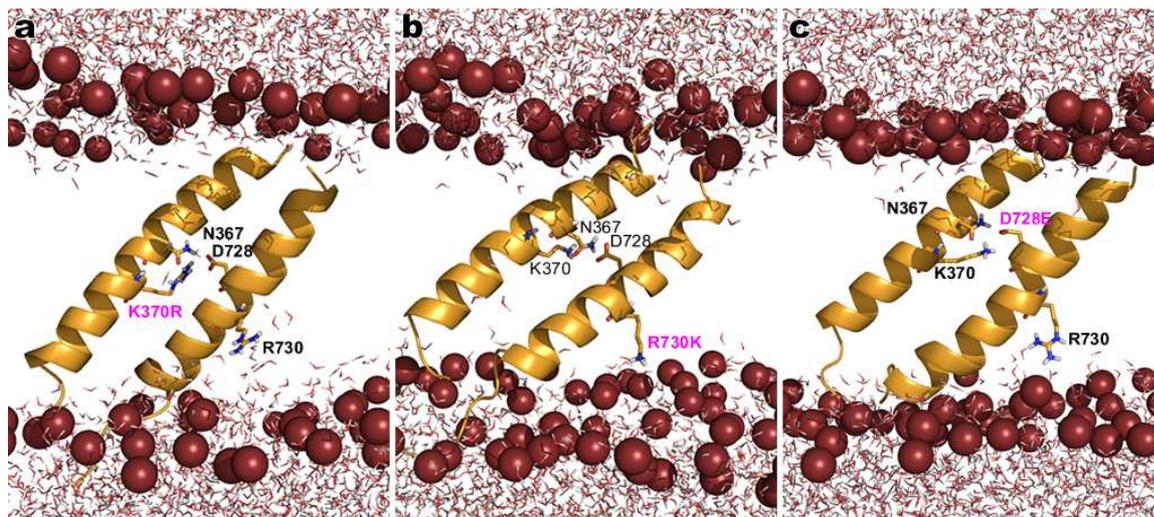


FIGURE 3.7 Final snapshots of MD simulations from single mutant dimers with a salt-bridge. (a) K370R mutant, (b) R730K mutant, and (c) D728E mutant. The conserved residues Asn367, Lys370, Asp728, Arg730 and mutated residues are shown as stick representation. Lipid tails and ions are not shown for clarity.

The MD simulations of the three double mutants (N367L & K370L, D728L & R730L, G354A & G358A) resulted in different conformations with intact TM helices (**FIGURE 3.8**). The largest structural fluctuations compared to the starting structure were observed for the D728L & R730L double mutant (**FIGURE 3.8a**). 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 α distances in **TABLE 3.3**). However, the G354A & G358A double mutant was as stable as the wild-type during the simulation (**FIGURE 3.8c**), probably due to the presence of the salt-bridge interaction.

TABLE 3.2 Structural parameters for the data between 80 and 100 ns of the MD simulations of E1-E2 wild-type heterodimer and E1-E2 mutants. Average number of H-bonds per time frame: An H-bond characterized by Donor-Hydrogen-Acceptor (D-H...A) is defined to have an 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.

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

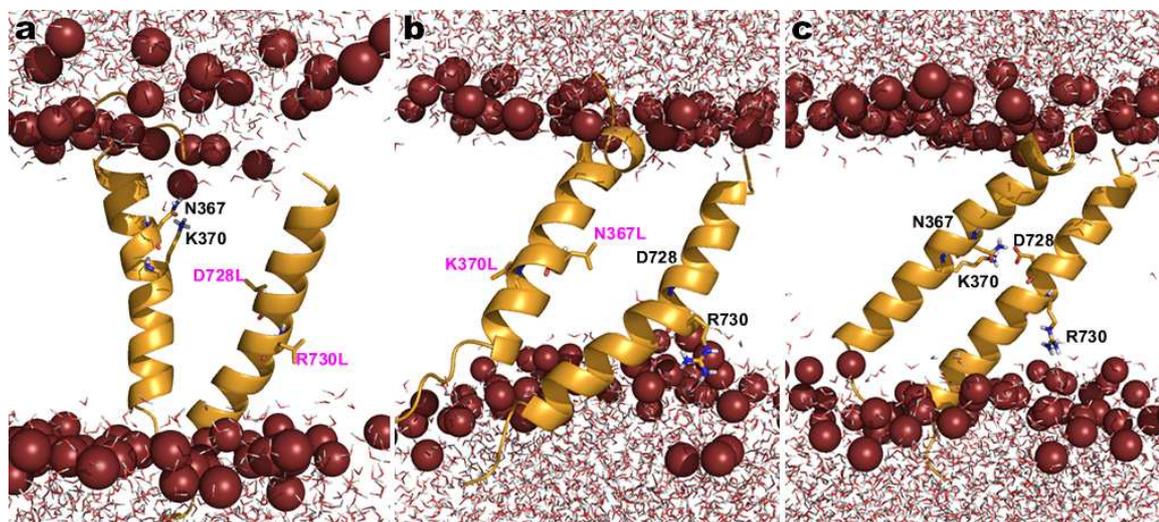


FIGURE 3.8 Final snapshots of MD simulations from the double mutant dimers. (a) D728L & R730L, (b) N367L & K370L and (c) G354A & G358A (mutated residues are not shown). The conserved residues Asn367, Lys370, Asp728, Arg730 and the replaced Leu are shown as stick representation. Lipid tails and ions are not shown for clarity.

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 3.3**). 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 3.3** and **FIGURE 3.8**). 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) (**FIGURE 3.9**). 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.

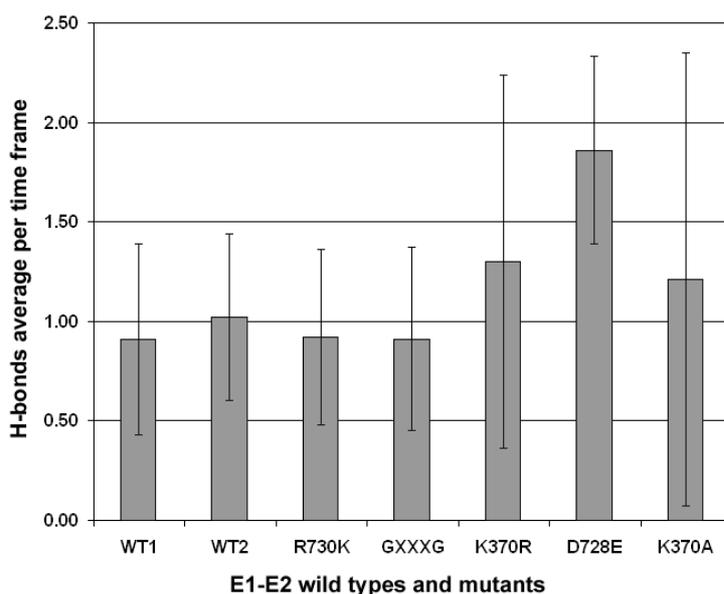


FIGURE 3.9 Inter-helical H-bond interactions for the E1-E2 wild-types and mutants.

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 (**FIGURE 3.10a**). 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) (**TABLE 3.3**) and the average number of H-bond interactions between both monomers

was similar to the other heterodimerized conformations (1.21 nm) (**FIGURE 3.9**). Here 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. This behavior caused local unfolding in the centre of the E2 of K370A mutant (**FIGURE 3.10a**). On the other hand, if Asp728 is mutated into Ala, no rotation of Arg730 in E2 was observed (**FIGURE 3.10b**).

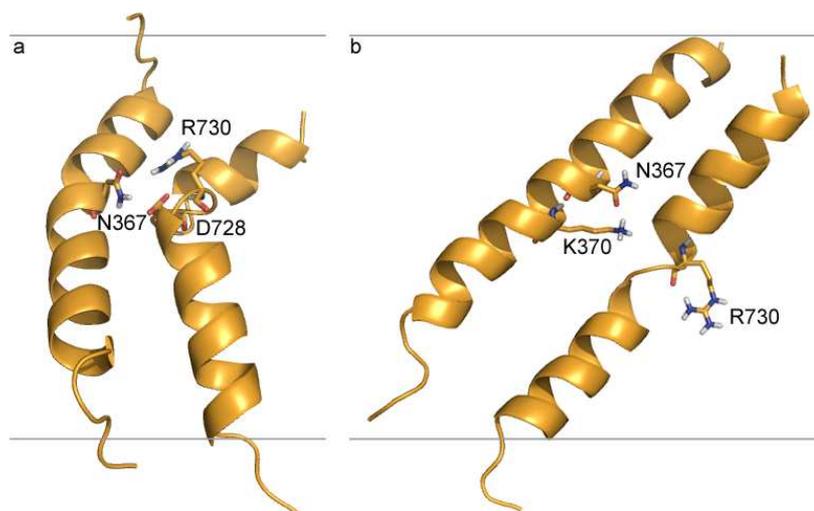


FIGURE 3.10 Final snapshot of mutants (a) K370A and (b) D728A. The conserved residues Asn367, Lys370, Asp728 and Arg730 are shown as stick representation. The mutated residues, lipid bilayer, water and ions molecules are not shown for clarity.

3.3.6 Comparison of MD Structures vs the NMR structure

As the only NMR structure available for the HCV envelope glycoproteins is a segment of E1 consisting of 21 residues (Op De Beeck et al, 2000) with the PDB-code 1EMZ.pdb, RMSD analyses were done on the same segment during the MD simulations with respect to the NMR structure (**FIGURE 3.11**). 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, whereas the RMSD of the structural ensemble derived from NOE restraints was 0.03 nm (Op De Beeck et al, 2000). Although we found a somehow larger RMSD of the 21 residue segment (Gly350 to Lys370) of 0.15 nm, these deviations are still smaller to the variation within the NMR ensemble of 24 structures (0.29 nm) (Op De Beeck et al, 2000). Thus, the E1 helical conformation derived from MD simulations is quite similar to the conformation determined by NMR in trifluoroethanol (TFE) (Op De Beeck et al, 2000). 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 (**FIGURE 3.2**), but the GxxxG motif was not located at the helix-helix interface (see discussion below).

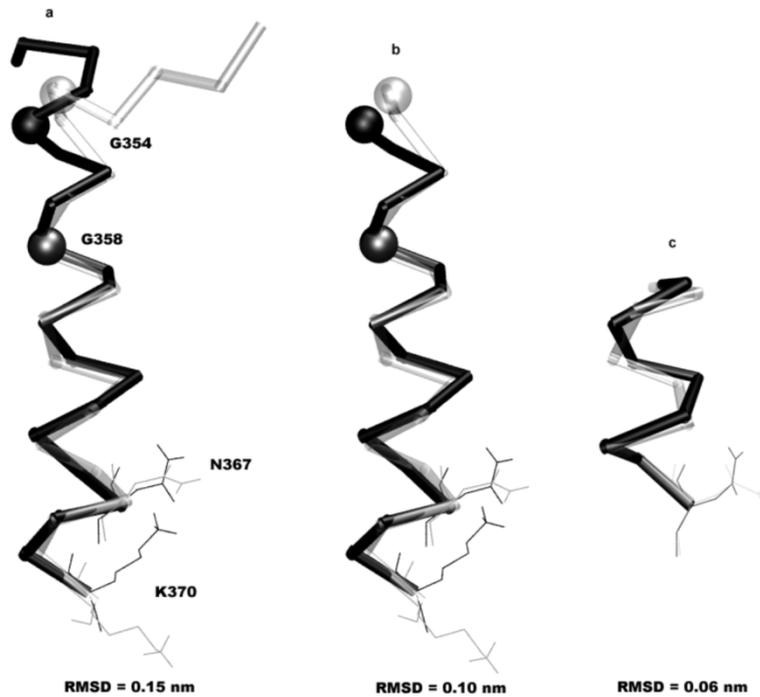


FIGURE 3.11 Superimposition of the E1 segment from the E1-E2 heterodimer wild-type to the NMR structure, 1EMZ.pdb. Coloring scheme: Black – 1EMZ.pdb; Grey – Segment of E1 from the simulation of the E1-E2 heterodimer model. RMSD values are listed below the figures; (a) Segment consists of residue G350 to K370, (b) segment from G354 to K370. Conserved residues G354, G358, N367 and K370 are highlighted as wire frame presentation, and (c) segment from I359 to N367.

3.4 Discussion

In the viral *Flaviviridae* family, at least one positively charged residue is highly conserved in both putative TM domains of the envelope glycoproteins (Cocquerel et al, 2000). 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 (Cocquerel et al, 2000)(Ciczora et al, 2007) 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 (Yoo & Cui, 2008). In spite of these considerations, an experimental study by Hessa *et al.* (Hessa et al, 2005) proved that single TM segments with a polar or charged residue in the middle of the domain were able to be inserted as membrane proteins via Sec61 translocon.

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 (Johansson & Lindahl, 2006). Moreover, Hessa *et al.* (Hessa et al, 2005) found that the biological apparent insertion free energy scale showed the highest value (3.49 kcal/mol) when Asp was placed in the middle of the TM domain of the H-segment compared to other amino acids.

Dubuisson and co-workers suggested that the E1 and E2 TM helices are inserted cooperatively into the lipid bilayer based on mutagenesis results (Ciczora et al, 2007)(Ciczora et al, 2005). 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 environment. 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 (Cocquerel et al, 2000)(Cocquerel et al, 2002), this should facilitate the entry of the E1-E2 heterodimer into the lipid bilayer environment (Cocquerel et al, 2002). 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 (Bonifacino et al, 1991). 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 towards 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 (Yoo & Cui, 2008). 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 (Freites et al, 2005).

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 (Joh et al, 2008). 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 (Meindl-Beinker et al, 2006).

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 what 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 what 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 (Im et al, 2003)(Tanizaki & Feig, 2005)(Bu & Brooks III, 2008)(Bu et al, 2007)(Lazaridis, 2003)(Mottamal et al, 2006)(Ulmschneider et al, 2007a) 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 (Nymeyer et al, 2005). However, it is not clear from experiment what is the thermodynamically most favorable state of the two isolated helices because the experiments (Ciczora et al, 2007)(Cocquerel et al, 2002) 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.

3.4.1 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. These results strongly support that the ion pair interaction between the particular charged amino acids is responsible for the inter-helical 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 (Ciczora et al, 2005). 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 (Ciczora et al, 2007)(Ciczora et al, 2005). 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 (Meindl-Beinker et al, 2006). 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% (Ciczora et al, 2007). 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 (Ciczora et al, 2007)(Ciczora et al, 2005)(Op De Beeck et al, 2000).

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 (Cocquerel et al, 2002). For soluble proteins, there exist several computational methods that can qualitatively predict the effect of protein mutations on their stability (Benedix et al, 2009)(Potapov et al, 2009). 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.

3.4.2 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 (Langosch et al, 1996)(Senes et al, 2000). Also for the E1-E2 heterodimer, mutating either Gly354 or Gly358 impaired the E1-E2 assembly (Ciczora et al, 2007). 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.

3.5 Conclusion

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 (Cocquerel et al, 2002) 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.

Chapter 4

Microsolvation of Bridging Ion Pairs in Transmembrane Helix Dimers

Submitted to BBA Biomembranes, Under revision

Charged and polar amino acids in the transmembrane domains of integral membrane proteins can be crucial for protein function and also promote helix-helix association or protein oligomerization. Yet, our current understanding is still limited on how these hydrophilic amino acids are efficiently translocated from the Sec61/SecY translocon into the cell membrane during the biogenesis of membrane proteins. In hepatitis C virus, the putative transmembrane segments of envelope glycoproteins E1 and E2 were suggested to heterodimerize via an Lys-Asp ion pair in the host endoplasmic reticulum. Therefore in this work, we carried out molecular dynamic simulations in explicit lipid bilayer and solvent environment to explore the stability of all possible bridging ion pairs using the model of H-segment helix dimers. We observed that, frequently, several water molecules penetrated from the interface into the membrane core to stabilize the charged and polar pairs. The hydration time and amount of water molecules in the membrane core depended on the position of the charged residues as well as on the type of ion pairs. Similar microsolvation events were observed in simulations of the putative E1-E2 transmembrane helix dimer of envelope glycoproteins from the hepatitis C virus. Thus this study illustrates the important contribution of water microsolvation to overcome the unfavorable energetic cost of burying charged and polar amino acids in membrane lipid bilayers.

4.1 Introduction

Helical transmembrane (TM) bundles are the predominant type of polytopic TM proteins. Their structures are assemblies of mainly hydrophobic helices. However, for functional reasons, they sometimes contain polar and charged residues even in the hydrophobic core of the membrane bilayer. Once the proteins are fully folded, these residues are shielded from the lipid environment. However, the insertion into the membrane via the Sec61/SecY translocon is an energetically challenging hurdle that these helices need to overcome. It has been suggested that cooperative insertion of multiple helices may facilitate this process (Meindl-Beinker et al, 2006). For example, TM helix dimers found in the family of *Flaviviridae* viruses (Mukhopadhyay et al, 2005) (Lindenbach et al, 2001) are stabilized by charged and polar residues in the center of the lipid bilayer (Ciczora et al, 2007). Since these are formed from only two helices, it is not possible to fully shield the charged residues from the surrounding lipid acyl chains.

The energetic cost of inserting polar and charged amino acids into lipid membrane was analyzed by several computational studies (MacCallum et al, 2008)(Ulmschneider et al, 2007b)(Yoo & Cui, 2008)(Dorairaj & Allen, 2007). However, the experimental studies by von Heijne, White and their colleagues (Hessa et al, 2005)(Hessa et al, 2009) indicated that the insertion energy for a helix monomer containing charged or polar residues is not as high as predicted from the free energy of solvation (Radzicka & Wolfenden, 1988). The efficiency to get inserted into the membrane by the translocon machinery depends strongly on the positions of the polar/charged residues with respect to the membrane and to each other and on helix-helix association (Meindl-Beinker et al, 2006). Recent experimental studies suggested that motifs from loop regions or from the nearest- neighbor TM helices can also favor the membrane insertion (Hedin et al, 2010). Also, TM helix repositioning in the membrane during the folding and oligomerization (Kauko et al, 2010) could be one of the reasons for lowering the cost of inserting the charged and polar residues. Johansson and Lindahl pointed out that high protein content in biological membranes could counterbalance the hydrophobic environment of membrane lipid bilayers (Johansson & Lindahl, 2009a). We will argue here that the remarkable efficiency of multi-spanning TM helices containing polar and charged residues to partition into the hydrophobic core of the lipid bilayer could also be explained –in part – if those residues remained partially solvated during the folding process (Krepkiy et al, 2009). The aqueous interior of the protein-conducting channel in the translocon suggests that water molecules could be co-translocated with the peptide chains. This could in fact lower the energetic cost of translocating polar amino acids during the TM protein biosynthesis.

For the envelope glycoproteins E1 and E2 of hepatitis C virus, the TM domains were suggested to heterodimerize via a salt-bridge (Cocquerel et al, 2002). Additional support for this model has been found in a recent MD simulation study that observed atomistically the contribution of polar and charged residues to the helix-helix association of the E1-E2 heterodimer (Jusoh et al, 2010). The TM segments of the E1 and E2 glycoproteins consist of two stretches of short hydrophobic residues with a short segment of highly conserved polar and charged residues in between. This pattern also occurs in the putative TM domains of other envelope glycoproteins from *Flaviviridae* viruses (Cocquerel et al, 2000). The TM segments of these viruses are believed to be not only involved in the virus entry but also responsible for the retention of the E1-E2 envelope glycoproteins in the endoplasmic reticulum membrane (Ciczora et al, 2005). These unique multifunctional roles inspired us to further investigate the roles of the polar and charged amino acids in the TM helix domains.

In this work we employed atomistic molecular dynamics simulations with explicit modeling of the lipid bilayer and water environment to explore the behavior of TM helix monomers containing a charged residue in the middle of the helix segments. Thereafter, we studied helix dimers interacting via an ion pair to observe the dynamic properties of the peptide-water-lipid bilayer system. As a model system we used the so-called H-segment that was extensively studied as a fusion TM segment by Hessa et al. (Meindl-Beinker et al, 2006)(Hessa et al, 2005)(Hessa et al, 2009)(Hessa et al, 2007). The results from the simulation of monomers demonstrated differential effects of the individual charged amino acids on the isolated TM helices in DMPC or DPPC lipid bilayers. Furthermore, we show the effect of dimerization via a salt-bridge and the position of the interacting charged residues which give rise to dynamic microsolvation events in the dehydrated membrane lipid bilayer. Similar trends for hydration were observed in the previously published simulations of the TM domain of E1-E2 envelope glycoproteins of hepatitis C virus (Jusoh et al, 2010) which were carefully re-analyzed for this work.

4.2 Methods

4.2.1 Sequences and System Preparation

This work has been inspired by the in vivo hydrophobicity scale of (Hessa et al, 2005) and (Hessa et al, 2007). Consequently, we used the same H-segment sequence as those authors comprising the 27 residues GGPG-AAAALALALXLALALAAAA-GPGG. The “X” represents the location in the TM helix monomer that was substituted by a charged residue (Arg, Lys, Glu, Asp) in this study. For the TM

helix dimers that are bridged by ion pairs, 2 sets of simulations were performed; set X14-Y14 and set X12-Y16, that differed in the location of the positively charged and negatively charged residues. The numbers indicate the positions of the charged residue in the helices.

All structures used in this study were prepared as ideal α -helices. The SCWRL program (Canutescu et al, 2003) was used to position the side-chain rotamers. For the TM helix dimers, Gromacs 4.0.3 (Hess et al, 2008) tools were used to set up paralleled H-segment dimers with the charged residues pointing to each other at the helix-helix interfaces. When constructed this way, the terminal side-chain atoms of the charged residues were separated by distances between 0.3-0.5 nm. The protonation states of the titratable side-chains were kept as found at pH 7 in aqueous solution.

4.2.2 Peptide-Bilayer System Setup

We used two different lipid bilayers as membrane environment for the simulation of the H-segment monomers. The starting geometries were constructed from a fully hydrated equilibrated lipid bilayer of 128 dimyristoyl-phosphatidylcholine (DMPC) lipids solvated with 5,673 simple point charge (SPC) water molecules (Griepner et al, 2007) and from 128 dipalmitoyl-phosphatidylcholine (DPPC) lipids solvated in 6143 SPC water molecules (Berendsen et al, 1981), respectively. A cavity of suitable size was created to accommodate one or two TM helices using the protocols of reference (Faraldo-Gómez et al, 2002). The solvent-accessible protein surfaces of the TM helices required for the cavity measurement were calculated by the program MSMS using a probe size radius of 1.4 Å (Sanner et al, 1996).

Each peptide monomer or dimer was introduced parallel to the membrane bilayer normal in the lipid membranes. In each case, 4-8 lipids were removed that severely overlapped with the peptides and the protein-lipid bilayer system was surrounded by approximately 45-50 water molecules per lipid molecule, thus ensuring full hydration of the membrane. The system was then subjected to 500 steps of energy minimization using the steepest descent algorithm in order to relax any steric conflicts generated during the setup. Na⁺ and Cl⁻ ions 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 its insertion. Subsequently, fully unrestrained production runs of at least 100 ns duration were performed for the systems.

4.2.3 Simulation Details

The DMPC and DPPC lipid bilayer interactions were described with the Berger force-field parameters (Berger et al, 1997). The TM helices were modeled with the united atom force-field GROMOS96 53a6 (Oostenbrink et al, 2005). Simulations were performed with the Gromacs 4.0.3 package (Hess et al, 2008) using 2-fs time steps. Periodic boundary conditions were used in all directions. Bonds to H atoms were constrained using the LINCS algorithms (Hess et al, 1997). For the short-range van der Waals interactions, a cutoff distance of 1.0 nm was used. The long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method with a grid spacing of 0.12 nm and cubic interpolation. The non-bonded pair list was generated every 10 steps with a cutoff of 1.0 nm. Water, lipids and peptide systems were coupled separately to temperature baths, 323 K for the DPPC and 310 K for others using the Berendsen algorithm with a time constant of $\tau_T = 0.1$ ps (Berendsen et al, 1984). The higher temperature is commonly used for DPPC simulations (Nagle & Tristram-Nagle, 2000) (Krüger & Fischer, 2008) to avoid that the lipids form a gel-like phase with increased ordering of the hydrocarbon chains. For keeping the pressure constant, semi-isotropic coupling was employed separately for the lateral and for the normal directions with Berendsen weak coupling and a $\tau_p = 1$ ps time constant. The compressibility was set to 4.5×10^{-5} bar⁻¹ (Berendsen et al, 1984).

Analyses of the trajectories were primarily performed with tools included in the Gromacs 4.0.3 suite (Hess et al, 2008). Root mean square deviations (RMSDs) analyses were based on the coordinates of all atoms of the peptides. The hydrogen bond analyses used a 0.35 nm distance cut-off between donor-acceptor atoms and required the bond angle to be between 150-180°. All protein structure images in this work were prepared with the Pymol program (<http://pymol.sourceforge.net>).

4.3 Results and Discussion

In this work, all-atom MD simulations were performed to investigate the structure and degree of internal solvation of membrane lipid bilayers containing TM helix monomers with a charged residue located in the centre of the helix and TM helices that are associated via an ion pair, respectively. As designed TM domain, we used the so-called H-segment (Hessa et al, 2005) which was prepared as an ideal helix for the starting structures of the simulations. 16 systems were simulated for at least 100 ns each.

4.3.1 Simulation of Monomers in DMPC and DPPC Lipid Bilayers

During the simulations of H-segment monomers in a DMPC lipid bilayer, all helices remained close to the conformation of an ideal α -helix (RMSD 0.06 nm), except for the H-segment containing Asp at the helix center (0.27 nm) which partially unfolded (**TABLE 4.1**). The intact H-segment monomers tilted strongly between 52.3 and 67.5 ° with respect to the membrane normal (also see **FIGURE 4.1**). The thickness of the hydrocarbon core of the bilayer membrane for the monomers in a DMPC bilayer ranged between 3.45-3.59 nm (**TABLE 4.1**). This result is in good agreement with the membrane thickness of a pure DMPC lipid bilayer (3.46 nm) obtained from MD simulation (Griepner et al, 2007) and in the experiment (Lewis & Engelman, 1983)(Kučerka et al, 2005).

TABLE 4.1 MD simulations of H-segment monomers in DMPC and DPPC lipid bilayers. All simulations were run for 100 ns of simulation time. The flanking residues (Gly-Gly-Pro-Gly) at either N or C-terminal side were not included in the analyses of RMSDs from ideal helix and the average helical angle. The membrane thickness, computed by GridMAT-MD (Allen et al, 2009), indicates the average hydrophobic thickness of the membrane, measured from the average distances between the phosphate atoms of upper and lower leaflets.

H-segment monomers	Membrane thickness (nm)	RMSDs from ideal α -helix (nm)	Average helical angle per-residue (°)	Tilting angle (°)
In DMPC				
Asp14	3.58 ± 0.40	0.27 ± 0.03	63.94 ± 16.27	Kinked
Glu14	3.45 ± 0.31	0.06 ± 0.02	98.63 ± 0.80	67.5 ± 9.3
Lys14	3.59 ± 0.32	0.06 ± 0.01	98.36 ± 0.67	59.0 ± 4.4
Arg14	3.51 ± 0.35	0.06 ± 0.02	99.33 ± 0.87	52.3 ± 12.0
In DPPC				
Asp14	3.99 ± 0.33	0.25 ± 0.01	79.63 ± 9.46	Kinked
Glu14	4.20 ± 0.44	0.05 ± 0.01	99.35 ± 0.60	47.8 ± 3.4
Lys14	4.02 ± 0.37	0.32 ± 0.02	48.76 ± 10.91	Kinked
Arg14	3.93 ± 0.42	0.06 ± 0.02	98.58 ± 0.78	53.1 ± 4.5

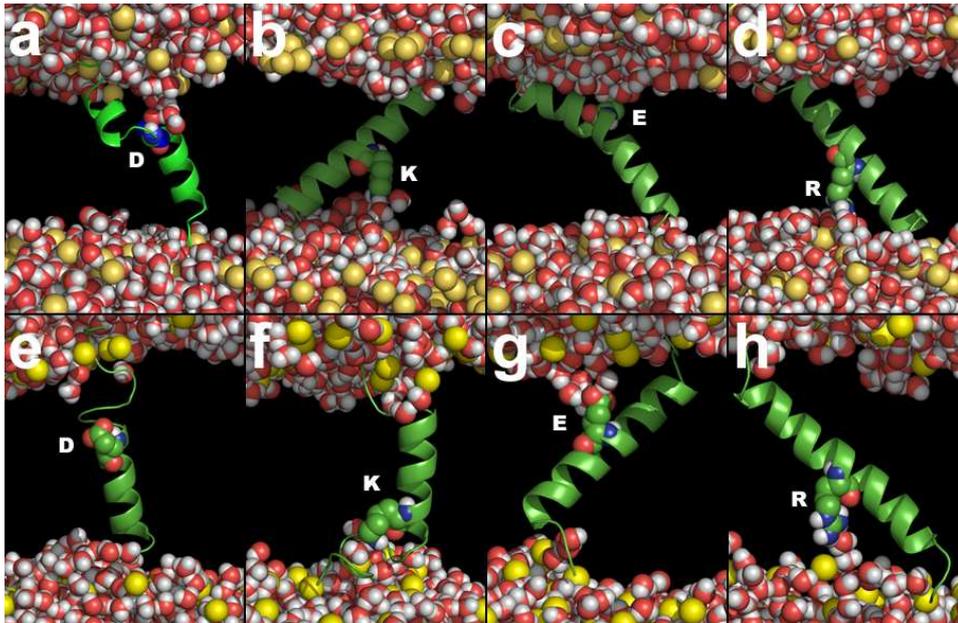


FIGURE 4.1 MD simulations of H-segment monomers containing a charged residue. (a-d) in DMPC lipids, (e-h) in DPPC lipids. The H-segment monomers are represented as helical cartoon. The charged amino acids are labeled and shown as van-der-Waals spheres. The lipid head groups are shown as yellow spheres and water molecules as red-white spheres. Lipid acyl-chains are not shown for clarity.

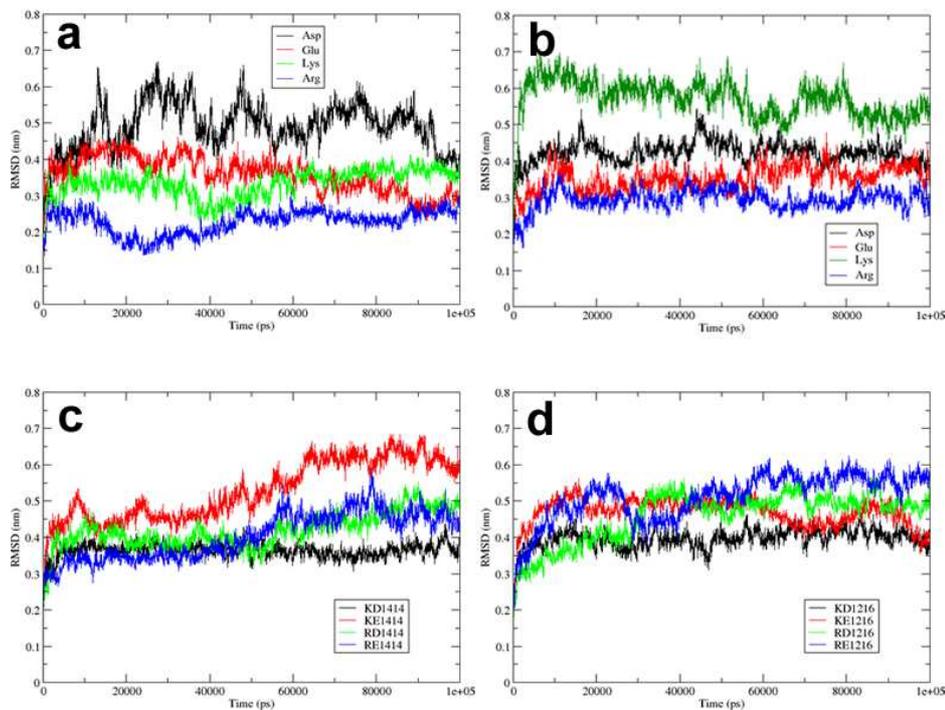


FIGURE 4.2 RMSDs of the H-segment monomers and dimers from the starting structures. (a) H-segment monomers in DMPC bilayers; (b) H-segment monomers in DPPC bilayers; (c) H-segment dimers of set X14-Y14 and (d) H-segment dimers of set X12-Y16.

In a DPPC bilayer consisting of lipid molecules with longer hydrophobic carbon-tails, one could expect that both Asp and Glu would severely bend the backbone of the helices due to their short side chains (Johansson & Lindahl, 2006). However, the only amino acid which caused the same effect as Asp was Lys. The helical conformation of the H-segment containing a charged Lys largely deviated from the starting structure (0.32 nm) and exhibited severe distortions. The other H-segment monomers containing Arg and Glu remained structurally stable (0.06 and 0.05 nm RMSD, respectively) similar to when simulated in the DMPC lipid bilayers. **FIGURE 4.2** shows the effect of each charged residue to the helix monomers during the simulation time. The average helical per-residue angle (see **TABLE 4.1**) describes the helix integrity. For an ideal α -helix, the angle should be close to 100°, but the partially unfolded helices of the H-segments with Asp in DMPC and DPPC lipids and Lys in DPPC lipids showed smaller average angles below 80°.

The unfolding of a TM helix reveals which portions of the peptide cannot be favorably accommodated in the hydrophobic lipid bilayer environment. Naturally, all charged residues in a TM helix like to interact with other polar or charged atoms. In our setup, the only chance for the centrally placed charged residues to achieve a favorable coordination of their charged side-chains is to form hydrogen bonds with water molecules or with the polar head groups of the lipids at the membrane interfacial region. For that reason, all H-segments tilted in order for their side-chains to reach the hydrophilic interface but at the same time optimized the position of the other residues according to the environment. The side chain of Asp is the shortest among the four charged residues. Even with a tilted TM helix, a centrally placed Asp side-chain cannot reach the interface region unless the ideal helix geometry is distorted causing partial unfolding. No such behavior was observed for the H-segment monomer containing a Glu amino acid which has the second shortest side-chain next to Asp. Even when simulated in the thicker DPPC lipid bilayer, the Glu amino acid positioned in the center of the helix did not affect the helical conformation along the simulation time. Surprisingly, the only amino acid causing severe distortions beside Asp in a DPPC bilayer was Lys.

The flexible side-chain of Lys is well-known to efficiently snorkel up to form hydrogen bonds with the phosphate and carbonyl groups of the phospholipids. However, in the DPPC lipid bilayer, the H-segment adopted a smaller helix tilting angle to enable the flanking anchors (Gly-Gly-Pro-Gly) on both sides to interact well with the membrane interfacial region. Therefore, the side-chain of Lys could not reach out to form hydrogen bonds with the hydrophilic region as in the DMPC bilayer. As a result this caused the H-segment to partially distort in the DPPC bilayer (**TABLE 4.1** and **FIGURE 4.1e-h**). Interestingly, we noted that the undistorted H-segments containing the Arg and Glu amino acids oriented their charged side-chains either to the N- or C-termini. In contrast, for the H-segments with Asp/Lys amino acids, the Asp and Lys side-chains oriented to the same termini in both types of lipids (**FIGURE 4.1**). Although it was noted before that the side-chain orientation of charged residues is

generally biased toward the N-terminal region (Johansson & Lindahl, 2006), the H-segments exhibited both orientations in this work.

The occurrence of polar or charged amino acids in a hydrophobic TM helix monomer destabilizes the helical structure due to the strong interaction of the polar side-chain with the hydrophilic membrane interface. Although the helix could overcome the unfolding by tilting, the helical integrity will still depend on the helix length and the sequence composition (Jaud et al, 2009). The shorter the TM helix and the more polar or charged residues exist, the less stable is the helix. In nature, the helix distortion illustrates the non-TM topology and this is supported by statistical analysis of the current high resolution structures (London & Shahidullah, 2009). No TM helix monomer or dimer containing a charged residue exists in the database so far. Interestingly, White, von Heijne and colleagues showed on the basis of *in vivo* free energy of insertion of the 20 amino acids that marginally hydrophobic TM segments can be filtered by the translocon to be integrated into the membrane lipids (Hessa et al, 2005), regardless of the post-processing TM state. Therefore, the observed unfolding behavior of the helices indicates a lower preference for the TM state (Zhao & London, 2006) which correlates to the increment of the apparent free energy of insertion. Our results from the monomer simulation are in agreement with the apparent free energy of the biological scale that assigned the highest insertion energies to H-segment monomers containing Asp (3.49 kcal/mol) followed by Lys (2.71 kcal/mol (Hessa et al, 2005). In fact, a charged Asp residue was shown by several experiments to induce partial helix unfolding when located deeper in the core of the bilayer (Caputo & London, 2004).

4.3.2 Dimer Simulations

This study was initially inspired by the putative TM helices of the envelope glycoprotein from the family of the *Flaviviridae* viruses which contain at least one positively charged residue located in between hydrophobic stretches and have been suggested to exist as monomers and/or dimer (Mukhopadhyay et al, 2005). In HCV, the putative TM helices of the E1 and E2 envelope glycoproteins were suggested to associate as a dimer via an ion pair of Lys-Asp amino acids. Here, we investigated the structural integrity of TM dimers with different types of ion pairs as well as the dynamic interaction among the components in the membrane lipid bilayer system. Two sets of H-segment dimers were simulated with four different combinations of charged residues. In the first set named X14-Y14 both charged residues were placed at the same position in the TM helices. In the second set named X12-Y16, the charged residues were located one turn apart from each other.

We measured the bilayer thickness by averaging the distances between lipid head groups in the upper and lower leaflets of the lipid membrane with the tool GridMAT-MD (Allen et al, 2009). The

DMPC bilayer thickness of the set X14-Y14 ranged between 3.62-3.73 nm and for the set X12-Y16 between 3.57-3.72 nm (TABLE 4.2). These values are larger than those observed in the monomer simulations (3.45-3.59 nm). Likely induced by the ion pair interaction between the charged residues, also the tilting angles of both partner helices decreased significantly compared to the helix monomers. The tilting angles of the helices with a positively charged amino acid were in the range 16.3 – 32.7 ° and the helices with a negatively charged amino acid tilted between 20.0 - 49.1 ° (TABLE 4.2).

TABLE 4.2 Simulation details of H-segment dimers. All simulations were run for 100 ns. X and Y denote the two helices of the TM helix dimer.

H-segment dimers	Membrane thickness (nm)	RMSDs from the ideal α -helix conformation (nm) (80-100 ns)		Tilting angle (°) (80-100 ns)	
		Helix		Helix	
		X	Y	X	Y
Set X14-Y14					
K14-D14	3.62 ± 0.54	0.09 ± 0.06	0.18 ± 0.02	28.4 ± 3.9	44.4 ± 3.9
K14-E14	3.72 ± 0.41	0.08 ± 0.01	0.23 ± 0.02	23.9 ± 3.3	49.1 ± 3.2
R14-D14	3.73 ± 0.37	0.07 ± 0.02	0.31 ± 0.008	25.9 ± 5.4	Kinked
R14-E14	3.73 ± 0.50	0.07 ± 0.02	0.30 ± 0.03	16.3 ± 5.3	Kinked
Set X12-Y16					
K12-D16	3.72 ± 0.43	0.06 ± 0.02	0.26 ± 0.02	22.0 ± 7.2	24.2 ± 4.8
K12-E16	3.57 ± 0.51	0.08 ± 0.04	1.72 ± 0.05	32.7 ± 4.1	45.6 ± 5.2
R12-D16	3.67 ± 0.41	0.10 ± 0.01	0.29 ± 0.09	29.8 ± 4.0	44.6 ± 5.5
R12-E16	3.72 ± 0.37	0.13 ± 0.02	0.18 ± 0.01	28.5 ± 4.3	20.0 ± 3.8

The RMSD analyses of both dimer helices showed a large difference between the helices containing a positively charged (Arg, Lys) and those containing a negatively charged (Glu, Asp) amino acid (TABLE 4.2). The helices with a positively charged residue stayed structurally close to an ideal helix (0.08 - 0.13 nm). On the other hand, the helices with a negatively charged Asp or Glu amino acid deviated between 0.18 to 0.31 nm from the ideal conformation. We also noted that most of the helices with the charged Asp/Glu exhibited kinking but their helix partners were stably intact as depicted in FIGURE 4.3. FIGURE 4.2 shows the RMSDs of the helix dimers from their starting structures along the simulation time compared to simulation of monomers. Each dimer from both sets deviated between 0.2-0.7 nm from its starting structure.

These findings show that the position of the charged residues in the TM helix influenced the helical conformation of the H-segment dimers. Although interacting via the same type of salt-bridge, different locations of the charged residues affected the helix-helix packing in different ways. Set X14-Y14 resulted in severe helix bending and kinking particularly for the helices which contained a negatively charged residue (FIGURE 4.3 (b), (c) and (d)). The only exception is H-segment dimer

K14-D14. However, in the set X12-Y16, the integrity of the TM helices was better maintained for the helices containing a positively charged residue, R12-D16, K12-E16 and R12-E16, compared to those in the other set X14-Y14 which contain the same salt-bridges.

In all dimer simulations of the H-segment, we observed very tight ion pair interactions of the charged residues along the simulation time. Although we found several kinked helices, they were not unfolded as in H-segment monomers. Particularly for the H-segment monomer with the negatively charged Asp residue, the ion pair interaction to the other helix partner significantly increased the TM topology state. We found a similar effect in the simulations of the E1-E2 TM dimer of HCV where the ion pair interaction increased the helix integrity and stabilization (Jusoh et al, 2010). This underlines that the presence of polar and charged residues in multi-spanning membrane proteins may in part serve to stabilize helix-helix association. Indeed, peptide dimer interactions mediated by interhelical hydrogen bonds between Asn-Asn and Asp-Asp amino acids were shown experimentally to enhance the membrane insertion efficiency (Meindl-Beinker et al, 2006).

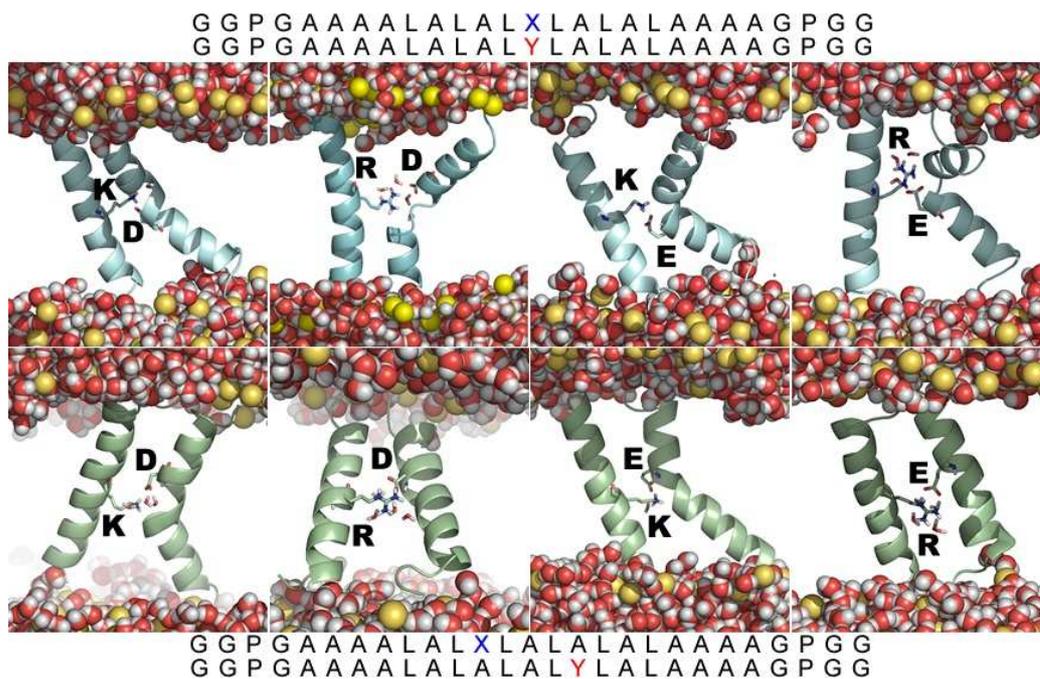


FIGURE 4.3 Dimer simulations with two different locations of the charged residue pairs, set X14-Y14 and set X12-Y16. Final snapshots after 100 ns of MD simulation of H-segment dimers with an interhelical salt-bridge interaction. The charged residues are labeled as single-letter code. Lipid phosphates are shown as yellow balls. Water molecules near the charged residues are shown in stick representation and water in the bulk phase as van der Waals spheres. Lipid acyl-chains are not shown for clarity.

4.3.3 Water Hydration of the Membrane Core

Events of water crossing and residence in pure lipid bilayers are very rare. For that reason, it is remarkable to observe water molecules which are able to reside in hydrophobic environments, particularly in the core region of the membrane lipid bilayer. Water penetration into the lipid bilayer was already reported in simulation studies (Johansson & Lindahl, 2006) and measured experimentally by solid-state NMR where waters coordinated Arg residues pointing into the lipid bilayer (Li et al, 2010). Water hydration of nonpolar cavities was also detected by NMR for the protein interleukin 1 β (Ernst et al, 1995) and by crystallography for the protein T4 lysozyme (Liu et al, 2008).

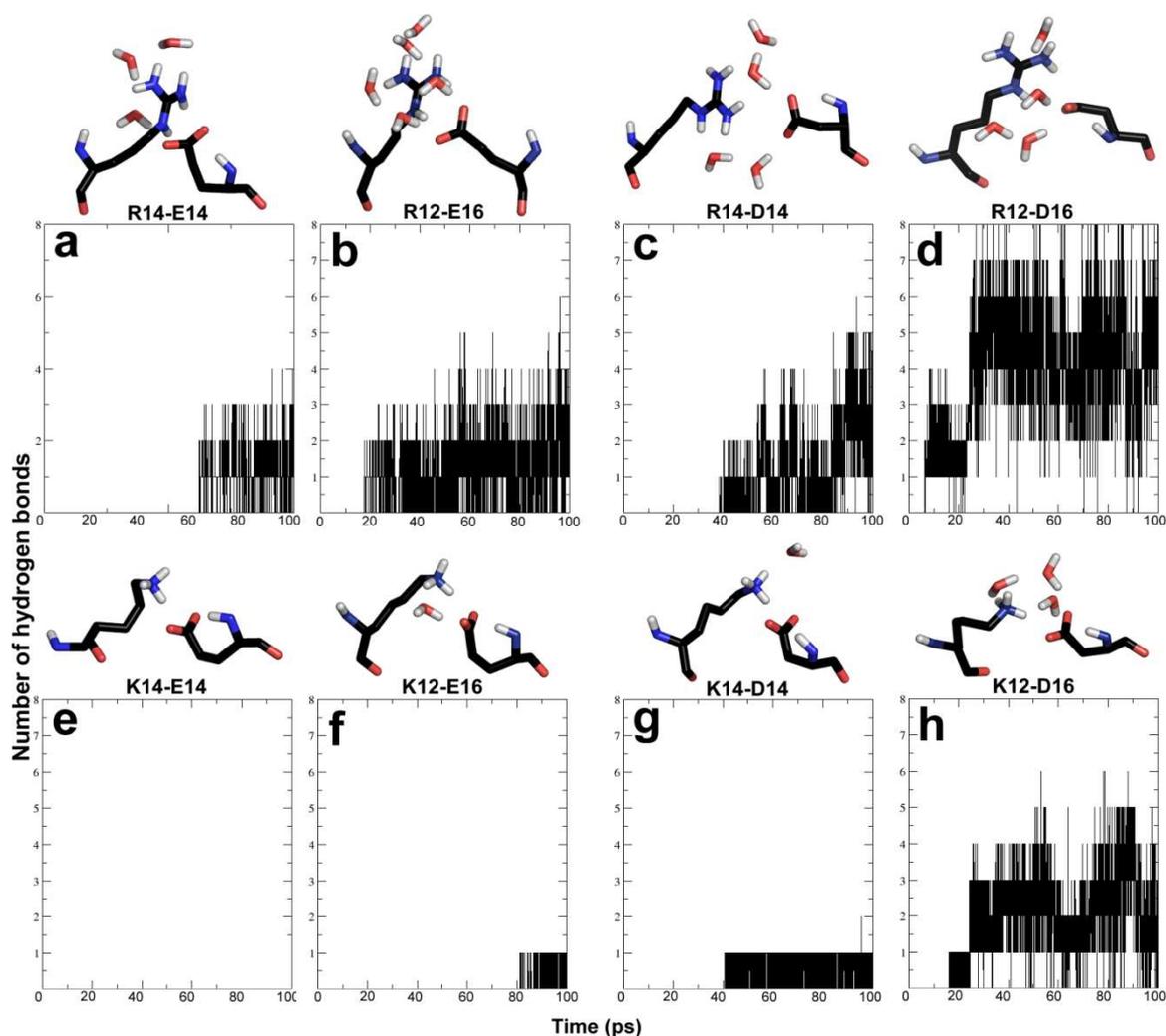


FIGURE 4.4 Hydration analyses of the charged residues in the hydrophobic core of the membrane lipid bilayer. Figures (a) to (h) are labeled according to the type of H-segment dimers. Shown are snapshots after 100 ns of MD simulation where the side-chains of the interacting charged residues are shown as stick representation together with water molecules within a distance of 0.7 nm. The graphs show the number of hydrogen bonds formed between the charged amino acids and the waters in each simulation.

Here, we observed repeatedly that the charged residues attracted several water molecules from the bulk phase to the membrane center to form hydrogen bonds (**FIGURE 4.2**). This resulted in permanent water penetration into the core of the membrane lipid bilayer. We note that, initially, the membrane core never contained any water molecules at the start of a MD simulation. The hydration level in the hydrophobic membrane core (characterized by the average number of hydrogen bonds) depended on the location of the charged residues in the membrane lipid bilayer. The deeper inside the core, the longer time water molecules needed to make contact with the charged residues. **FIGURE 4.4** clearly indicates that the TM helix dimers from the set X12-Y16 were hydrated earlier than the set X14-Y14. The fastest solvation was observed for the dimer R12-D16, where the water molecules managed to enter the hydrophobic core of the lipid bilayer within 10 ns of simulation time. The lowest hydration was found for both dimers interacting via Lys-Glu salt-bridges. Only one water molecule penetrated after 80 ns for the dimer K12-E16 whereas the dimer K14-E14 was still totally dehydrated after 100 ns of MD simulation.

The number of hydration waters also depended on the type of ion pair and the location of the charged amino acids in the TM helix (**FIGURE 4.4**). The TM helix dimers containing Lys-Asp and Lys-Glu ion pairs were not hydrated as much as dimers with Arg-Asp and Arg-Glu pairs. The average number of hydrogen bonds between the charged amino acids and the penetrating water molecules varied in each simulation. The helix dimers containing an Arg residue attracted more water molecules into the core of the membrane bilayers (3-5 water molecules), compared to the helix dimers with a Lys (1-3 water molecules). This is quite expected because Arg has more hydrogen-donor atoms in its side-chain compared to Lys (three-hydrogen donors). The helix dimers containing Arg-Asp pairs showed the same amount of water molecules in the membrane core after 100 ns of simulation, although the hydration of the dimer R14-D14 took place at a later time (~38 ns) than for the dimer R12-D16 (~6 ns).

As described, the hydrogen bonds between the charged residues and the water molecules varied due to the type of salt-bridge and the location of the charged residues. **FIGURE 4.5** summarizes the average number of hydrogen bonds observed in the membrane core in each dimer simulation. The results clearly illustrate that the average number of hydrogen bonds of the set X14-Y14 is lower than the set X12-Y16. In each case, there was at least one hydrogen bond stably connecting the interacting charged residues along the simulation time. The detailed analysis of the average number of hydrogen bonds is shown in **TABLE 4.3**. The largest number of hydrogen bonds was found for the H-segment dimer R12-D16 (~5 hydrogen bonds with waters). One reason for this could be that the charged residues are located nearer to the hydrophilic interface. Secondly, the side-chain of Arg contains the largest number of hydrogen bond donor atoms. Surprisingly, the side-chain of Asp also participated in almost the same number of hydrogen bonds to water molecules as Arg in sets K12D16 and R12D16 (**FIGURE 4.5** and **TABLE 4.3**). In the partially hydrated region of the membrane interfaces, all charged amino acids have the choice either to form hydrogen bonds with waters or with the polar

groups of lipids. Positively charged amino acids (Arg/Lys) usually act as hydrogen bond donors and negatively charged amino acids (Asp/Glu) naturally act as hydrogen bond acceptors (Johansson & Lindahl, 2006). We found that in the dehydrated region of the membrane core where the amount of water is limited, the available water molecules tend to act both as hydrogen bond acceptors and donors for the basic and acidic charged amino acids, respectively (**FIGURE 4.5**).

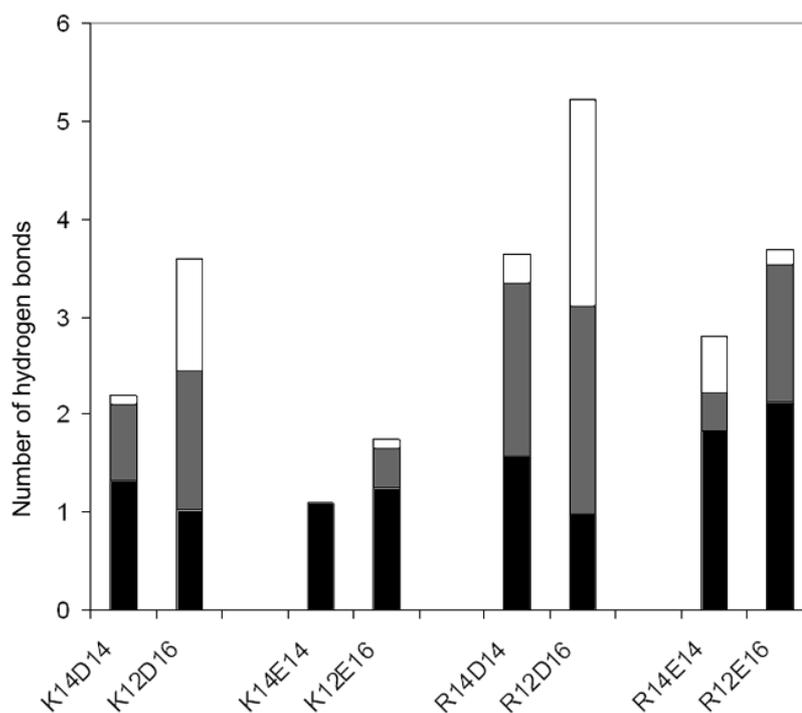


FIGURE 4.5 Average number of hydrogen bonds in each dimer simulation between the charged residues themselves and with water molecules in the core region of the membrane bilayer during 80-100 ns simulations. Black: Hydrogen bonds between the charged residues themselves; gray bars: hydrogen bonds between the positively charged residues and water molecules; white bars: hydrogen bonds between the negatively charged residues and water molecules.

TABLE 4.3 Number of hydration waters that coordinate the charged residue pairs in the center of the membrane bilayer and hydrogen bond analyses for the MD simulations of H-segment dimers. The number of hydration waters was computed in a sphere of 0.7 nm radius around the side-chains of the charged residues.

Simulations	No. of hydration waters at 100 ns	Average number of hydrogen-bonds	
		Charged residues to bulk waters (80-100ns)	Between charged-residues (0-100ns)
Lys-Asp (K14-D14)	1	0.78 ± 0.42	1.32 ± 0.52
Lys-Asp (K12-D16)	3	2.56 ± 1.32	1.03 ± 0.67
Lys-Glu (K14-E14)	0	0	1.10 ± 0.47
Lys-Glu (K12-E16)	1	0.41 ± 0.49	1.25 ± 0.68
Arg-Asp (R14-D14)	4	2.07 ± 1.28	1.57 ± 0.97
Arg-Asp (R12-D16)	4	4.25 ± 1.50	0.98 ± 0.95
Arg-Glu (R14-E14)	3	1.16 ± 0.61	1.83 ± 0.88
Arg-Glu (R12-E16)	5	1.57 ± 1.04	2.12 ± 0.77

The results from the MD simulations of helix dimers showed that once water molecules came into contact with the charged residues, several of them managed to stay throughout the simulation time. The hydration level and penetration time of water molecules differed although the dimers comprised the same type of salt-bridge. We extended the simulations of two helix dimers (R14-E14 and R12-E16) up to 200 ns to characterize the water solvation on a longer time scale. We observed that further water molecules continued to penetrate into the core of the membrane bilayer to solvate the charged residues. **FIGURE 4.6** clearly depict the increasing number of water molecules in a sphere of 0.7 nm radius around the side-chains of the charged residues along the simulation time. Again the helix dimer R12-E16 was hydrated more than the R14-E14 dimer although both comprised the same type of ion pair. Up to 8 water molecules occupied the core membrane of the R12-E16 dimer after 152 ns of simulation and their number gradually decreased to 5 water molecules on average till 200 ns (**FIGURE 4.6**). In the R14-E14 simulation the level of hydration remained around 4 water molecules till 200 ns.

Interestingly, the water molecules which were retained in the hydrophobic core of the membrane even managed to exchange with the bulk water on the 200 ns time scale. We observed several events of such dynamic water replacements in order to solvate the hydrophobic environment around the charged residues. **FIGURE 4.7** shows individual snapshots from the simulation of the R12-E16 dimer. The R12-E16 dimer was more hydrated at an earlier stage compared to the R14-E14 dimer as the charged residues are located closer to the membrane interface. In the R12-E16 simulation, the first water molecule started to enter the core region at 17 ns and jumped back into the bulk phase after 27 ns. The second water molecules went into the core at 22 ns shortly 5 ns before the first water jumped out and remained up to 103 ns. Similar water replacement events occurred throughout the simulation where new water molecules replaced the old ones. In total, 15 and 7 different water molecules were

observed that solvated the membrane core during the 200 ns simulation of the R12-E16 and R14-D14 systems, respectively.

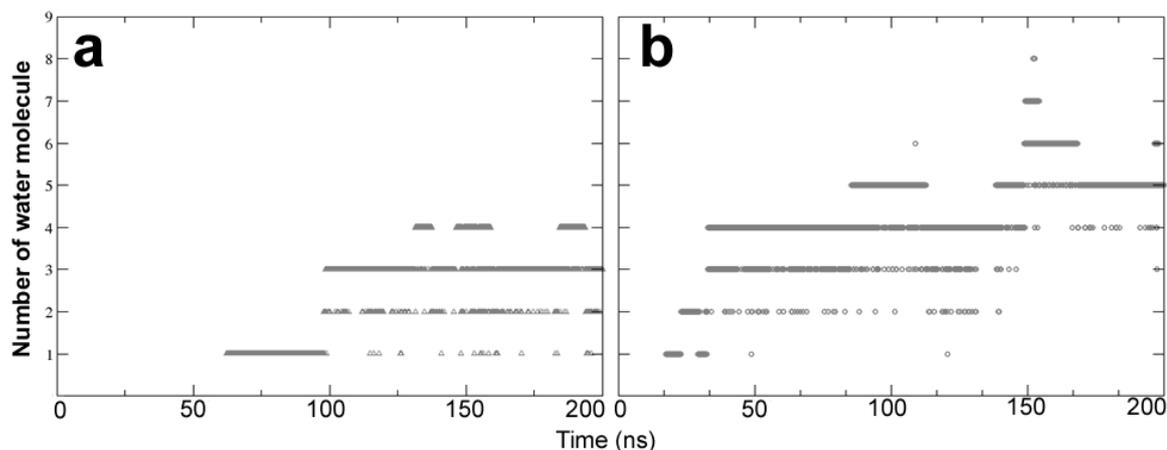


FIGURE 4.6 Hydration in the membrane core of (a) R14-E14 and (b) R12-E16. Shown is the number of water molecules per snapshot in a sphere of 0.7 nm radius around the center of mass of both charged residues in each simulation.

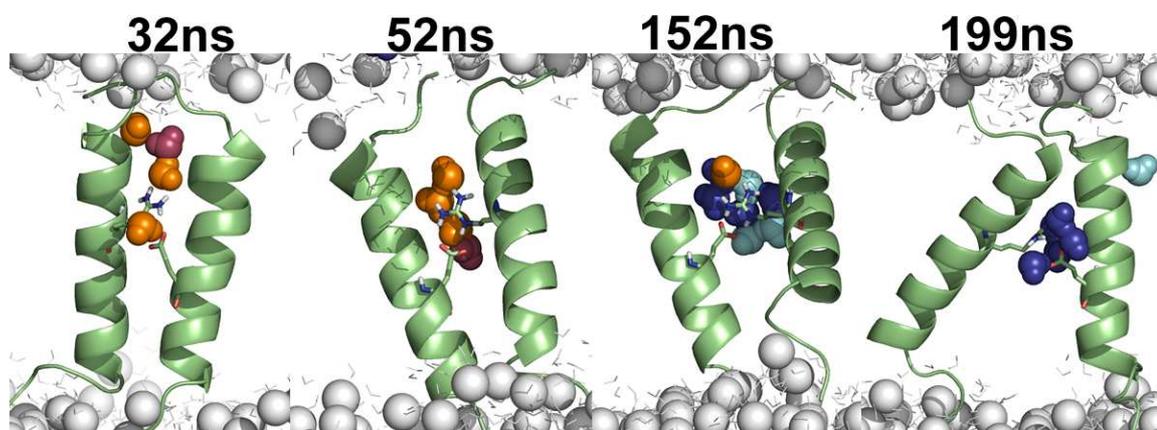


FIGURE 4.7 Microsolvation of the R12-E16 dimer in the membrane core. Snapshots are labeled according to the simulation time. In total, 15 different water molecules solvated the charged Arg and Glu ion pair during the 200 ns MD simulation. The water molecules are represented as spheres and colored based on the range of penetration time: (1) entering before 50 ns (orange); (2) entering and exiting in the interval (17 - 100ns) (maroon); (3) entering after 100 ns and exiting before 200 ns (cyan); (4) entering after 100 ns and residing in the core until 200 ns (blue). The other water molecules are shown as grey. The R12-E16 helix dimer is shown as cartoon and the side-chain of Arg and Glu are highlighted as sticks. The lipid head groups are shown as grey spheres. Lipid acyl-chains are not shown for clarity.

4.3.4 Microsolvation of the Putative Transmembrane Helix Dimers from Hepatitis C Virus

In the putative TM dimer of the E1-E2 envelope glycoproteins of hepatitis C virus, the helix dimer was suggested to interact via a salt-bridge of charged Lys-Asp amino acids. Our previous study showed that besides the salt-bridge interaction also inter-helical hydrogen bonds between the TM segments contributed to stabilizing the E1-E2 dimer (Jusoh et al, 2010). Several simulations of the TM domain of E1-E2, that were all started from the same configuration revealed three possible modes of interaction (**FIGURE 4.8**): (a) interaction via the salt bridge Lys370-Asp728 only, (b) interaction only via a hydrogen-bond between Asn367 and Asp728 and (3) the dimer interacts via both the hydrogen bond and the salt bridge where the side-chains of Asn367, Lys370 and Asp728 are oriented to the helix-helix interface in the core region of the bilayer.

The latter exhibited the most stable TM structure with a low hydration of only 2 water molecules during 100 ns of simulation (**FIGURE 4.8c**). The highest hydration was observed for the first case, where the E1-E2 dimer was interacting via hydrogen bonds formed by Asn367 and Asp728, and Lys370 faced the lipid acyl-chains. Due to the positively charged side-chain of Lys, several phosphate head groups were pulled into the core what locally distorted the lipid membrane. Among the three models, the only E1-E2 dimer model that exhibited kinking (and a locally unfolded E2 helix) was the first type. The Asn367 residue not involved in hydrogen bonds pulled bulk waters and lipid head groups into the core region and this deformed the local thickness of the membrane lipid bilayer. Therefore, it is tempting to suggest that the third model of an E1-E2 dimer interacting via both the hydrogen bond (Asn367-Asp728) and the ion pair (Lys370-Asp728) could be the one existing in nature based on the stability of the TM state and the relatively unperturbed membrane thickness. However, further experimental studies should be carried out to confirm this hypothesis.

Intensive mutagenesis works by the group of Dubuisson showed that the highly conserved Lys370 and Asp728 residues contributed to the HCV E1-E2 heterodimerization (Ciczora et al, 2005) (Ciczora et al, 2007). They noted, however, that when mutating Asp728 to Lys, the heterodimerization was still unaltered (Ciczora et al, 2005) and concluded that the ion pair is not the sole contributor to the helix-helix association. In our previous study, we also simulated a model of the putative E1-E2 dimer where Lys370 was mutated to Ala (Jusoh et al, 2010). The results from the MD simulations indicated that even in the absence of ion pair interaction, the E1-E2 dimer may still be stably associated because Asn367 formed hydrogen bonds with the side-chains of Asp728 and the oriented side-chain of Arg730. In the wild type E1-E2 dimer simulations, Asn367 contributed to the stability of the dimer as well besides Lys370 and Asp728 (Jusoh et al, 2010). **FIGURE 4.9** shows the snapshot of the putative TM domain of E1-E2 of HCV after 200 ns of simulation as well as its hydration plot during the simulation.

We observed a maximum number of four water molecules that managed to reside at one time to solvate the charged and polar residues. The hydration level is comparable to the H-segment dimers having the same type of Lys-Asp ion pair (K14-D14 and K12-D16). Eight different water molecules solvated the membrane core during the 200 ns simulation.

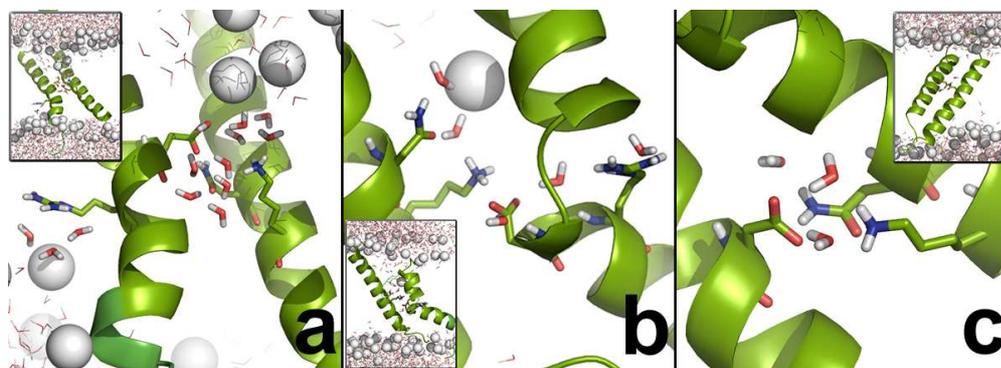


FIGURE 4.8 Models of the TM segment of the E1-E2 dimer from hepatitis C virus. The putative structures resulted from different polar/charged interactions at the helix-helix interface. a: Asn367-Asp728; b: Lys370-Asp728 and c: Asn367-Lys370-Asp728 interaction. The side-chains of Asn, Lys, Asp and Arg are shown as sticks. The E1-E2 dimers are shown as cartoon and the water molecules in a sphere of 0.7 nm radius are shown as sticks. Lipid acyl-chains are not shown for clarity. The small figures are shown to illustrate the whole system for each dimer. The snapshots shown are final conformations after 100 ns MD simulation.

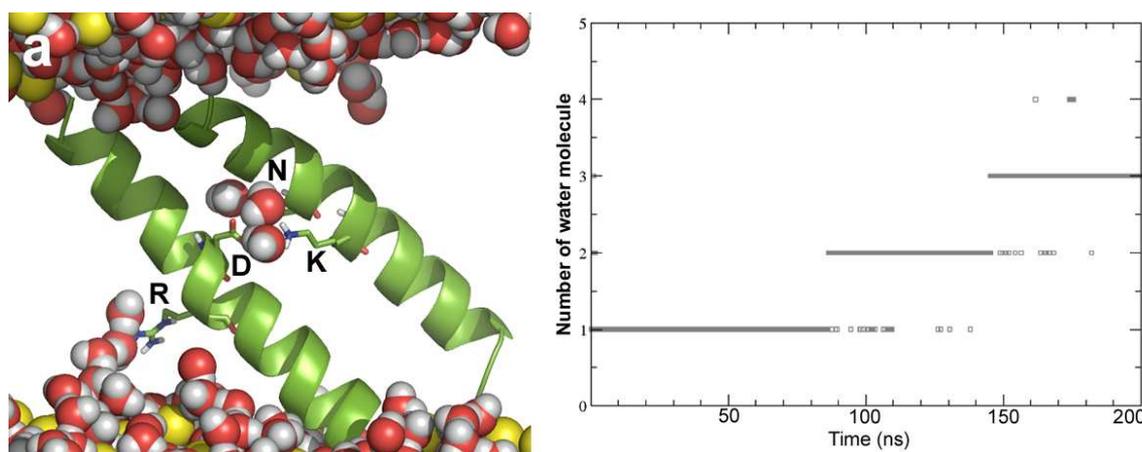


FIGURE 4.9 Microsolvation in the putative TM model of the E1-E2 dimer from hepatitis C virus. (a) Snapshot at 200 ns MD simulation of the TM domain of E1-E2 dimer; (b) The number of water molecules during the simulation in a sphere of 0.7 nm radius around the side-chains of Asn367-Lys370-Asp730.

We also analyzed the water hydration in the simulations of mutated E1-E2 dimers. As for the wild-type E1-E2 dimer, a similar increasing trend of solvation was observed for the simulation of R730K (**FIGURE 4.10a**). This is expected because the mutated residue, Arg730, is not located at the helix-helix interface. Thus the hydration is the same as in the wild type. For the doubly mutated E1-E2 dimer, where we replaced Gly354 and Gly358 by Ala, only two water molecules managed to retain at the same time (**FIGURE 4.10b**). It is interesting that although both TM dimers had the same Asn-Lys-Asp interaction at the helix-helix interface, the number of water molecules during the 100 ns time scale was different. It is possible that Gly354 and Gly358 facilitated the penetration of water molecules to solvate the highly polar residues. Therefore, a lower degree of hydration could have resulted when both glycines were mutated to Ala.

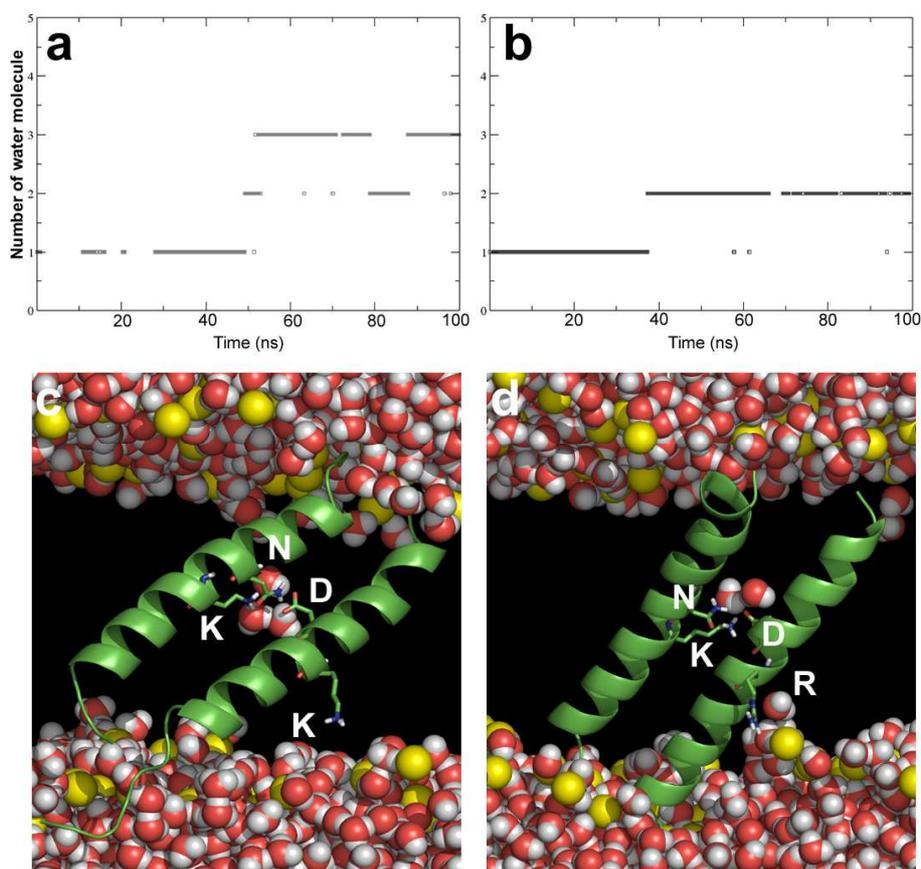


FIGURE 4.10 Microsolvation of the TM domain of the E1-E2 HCV mutated dimers. Simulation of (a) R730K dimer; (b) G354A&G358A doubly mutated dimer. The plot shows the number of water molecules interacting with the side-chains of Asn367, Lys370 and Asp728 in a sphere of 0.7 nm radius. Both mutated models contain the same Asn-Lys-Asp interaction at the helix-helix interface as the E1-E2 wild type. The snapshots in the lower panel show the final conformations after 100 ns of MD simulation.

Although water permeability across lipid membranes has been extensively studied, the mechanism still remains unclear (Mathai et al, 2010). Recent experiments showed that the penetration of water molecules correlated stronger with the area per lipid than with the chain length, saturation, or composition of the head group of the lipids (Mathai et al, 2010). In the case of an HIV1-TAT peptide, computational studies supported the model where the peptide translocates from the hydrophilic interface of the bilayer into the membrane core (Herce & Garcia, 2007). Although the peptide was highly hydrophilic and contained many charged Arg and Lys, it was able to cross the hydrophobic core of the membrane when helped by further peptides nearby.

In a pre-study, we also simulated systems with uncharged Asp and Lys in TM helix monomers and dimers. In none of these simulations, water molecules or lipid head groups penetrated into the core of the bilayer. Neither did we observe membrane deformations. However, we did not observe stable association of any helix dimer. This contradicts the experimental data about the importance of interhelical hydrogen-bonding residues for the H-segments (Meindl-Beinker et al, 2006) and of interhelical ion-pairing residues for the E1-E2 dimer of HCV (Ciczora et al, 2005). Therefore we focused on studying the charged forms of these residues here. The MD simulations of our study clearly revealed that individual water molecule from the bulk phase may enter the hydrophobic core of the membrane to coordinate polar and charged side-chains. Also, the simulation time scale of ~ 100 ns appeared long enough so that they may reversibly exchange. However we do not consider the level of hydration to be converged on this time scale. This will require substantially longer plain MD simulations or the use of simulations in the semi-grand canonical ensemble (Deng & Roux, 2008). Another possible concern is the suitability of the combination of Berger force field for the lipids and the SPC water model to study the favorability of microhydration relative to the bulk phase. Definite answers will require the availability of some experimental data possibly from solid-state NMR (Li et al, 2010).

4.4 Conclusion

The experimental free energies required for transferring charged amino acids from water to cyclohexane (Radzicka & Wolfenden, 1988) as well as the theoretical values according to the generalized Born scale (Ulmschneider et al, 2007b) are at least twice as large as the values of the biological scale of Hessa et al. (Hessa et al, 2007) and those of the Wimley-White hydrophobicity scale (Wimley & White, 1996). Recent studies demonstrated that the insertion of TM helices containing polar and charged residues into the membrane is facilitated by non-covalent interactions with motifs in loop

regions and nearest-neighbor TM helices, as well as by TM helix repositioning in the membrane during the folding and oligomerization, and by the high protein content in biological membranes.

Furthermore, this study shows based on molecular dynamics computer simulations of H-segment dimers and E1-E2 dimers from Hepatitis C virus that microsolvation of polar and charged amino acids and even ion pairs is another important factor that facilitates the oligomerization of membrane proteins and their insertion in the lipid bilayer. In the simulations, several water molecules from the bulk phase repeatedly managed to penetrate into the bilayer core where they hydrated the charged residues. These buried water molecules frequently exchanged with waters from the bulk phase on timescales of tens of nanoseconds. These observations illustrate that the very hydrophobic core of pure lipid bilayer membranes shows a significant degree of physicochemical adaptability in the presence of embedded TM helices and proteins.

Chapter 5

Molecular Dynamics Simulation of Putative Transmembrane Domains of Envelope Glycoproteins from Flaviviridae Viruses

The envelope glycoproteins of the family of Flaviviridae viruses are responsible for the initial binding of the virion to the cell membrane of the host cells before entering the host cells. During the virus biogenesis, these proteins are retained in the membrane of endoplasmic reticulum, and then they assemble with the other particles to form a mature virus. The TM domains of the envelope glycoproteins are have been shown to play multiple roles. For example, they contain a signal peptide, responsible for the endoplasmic reticulum retention and are crucial for the E1-E2 or prM-E dimerization. Unfortunately, so far no X-ray structure has been determined for the complete structure of E1/prM and E2/E envelope glycoproteins. Knowledge about these TM domains could lead to the finding of possible drug targets or vaccine candidates for the Flaviviridae viruses. Here we show that the TM segments of the E1/prM are more stable as a helix monomer compared to the TM segments of the E2/E envelope glycoproteins in lipid bilayers. Severely kinked helices were observed during the MD simulation of the TM domains of the DENV-E, WNV-E and JEV-E which are similar to the previous results for the HCV-E2 presented in Chapter 3. Comparative studies based on sequence analyses and from MD simulations show paralleled results that support the idea that the TM domains of the E1/prM and E2/E consist of a highly polar segment located in between two hydrophobic stretches.

5.1 Introduction

Dengue virus (DENV) infects more than 50 million people annually and the hepatitis C virus (HCV) is currently chronically infecting more than 170 million people worldwide (Vlachakis, 2008). They belong to the group of viruses for which till now there exists no specific antiviral therapy. These viruses are members of the family *Flaviviridae* which consists of three main genera; Flaviviruses, Hepaciviruses and Pestiviruses. The largest group is the Flavivirus that currently has more than 70 members of tick-borne or mosquitoes-borne viruses including DENV, West Nile Virus (WNV) and Japanese encephalitis virus (JEV). The hepacivirus genus consists of only one member, the HCV which is transmitted among humans by infected blood. Both Flavivirus and Hepacivirus genera are human pathogens. The Pestivirus is a genus that can only infect animals as, for example, the classical swine fever virus (CSFV) and the bovine viral diarrhea virus (BVDV). The family of *Flaviviridae* shares similarities in virion morphologies, genome organization and replication mechanism.

The envelope proteins namely E2 (in Hepacivirus and Pestiviruses) and prM and E (in Flaviviruses) are responsible for the initial binding to the host cells. The transmembrane (TM) domains that are located at the C-terminus cause the envelope glycoproteins to be retained in the membrane of the endoplasmic reticulum during the virus biogenesis. In Chapter 3, we showed that the putative TM segment of HCV-E1 was stable as a single-pass TM helix during the simulations although it contains highly polar and charged residues in its centre (Jusoh et al, 2010). In parallel, the NMR-derived structure of E1 illustrated that the polar Asn367 and the positively charged Lys370 are part of the helical region (Op De Beeck et al, 2000). In contrast, the putative TM segment of the HCV-E2 locally unfolded and kinked when it existed as a helix monomer in the membrane bilayer. The simulation study supports the experimental hypothesis that the highly conserved charged residues located in the middle of both putative TM domains of HCV-E1 and HCV-E2 are crucial for their heterodimerization (Jusoh et al, 2010)(Ciczora et al, 2007). The other members of the family *Flaviviridae* virus also show the similar TM sequence pattern. Their putative TM domains of envelope glycoproteins contain a short highly polar segment consisting of highly conserved charged and polar residues connecting two hydrophobic stretches see **FIGURE 1.12**.

Here, we extended the structural analyses of Chapter 3 to the several other members of the family of *Flaviviridae*. MD simulations were used to simulate each of the putative TM segments of E1/E2 or prM/E envelope glycoproteins as a single helix monomer. In this study, the sequences were obtained from four more viruses; DENV, JEV and WNV which represent the Flavivirus genus, and the

BVDV that represents the Pestivirus genus. The data were analyzed and compared with that of HCV of genus Hepacivirus.

5.2 Methods

5.2.1 Sequence Analyses

The sequences of the envelope glycoproteins of the Flaviviridae viruses were obtained from the UNIPROT database (<http://www.uniprot.org>). The referral id numbers from UNIPROT are described in parentheses for each of the viruses. In this study, we used sequences from HCV (P26664) and BVDV (P19711) to represent genus Hepacivirus and Pestiviruses, respectively. Sequences from three other viruses, DENV (P14337), JEV (P32886) and WNV (P06935), are representing genus Flaviviruses. Here, we were only interested in the envelope glycoprotein domains. Therefore, the corresponding sequences suggested by the UNIPROT as TM domain that are named E1/prM and E2/E proteins were further analyzed.

The sequences obtained from the UNIPROT database were initially analyzed by several secondary structure prediction servers. The TOPCONS web server (<http://topcons.cbr.su.se/>) was used to analyze the full sequences of the envelope glycoproteins. Then, putative TM segments were analyzed by a DeltaG prediction server (<http://dgpred.cbr.su.se/>) (Hessa et al, 2007) to predict their apparent free energy, ΔG_{app} . Then, the I-Tasser web server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Zhang, 2008) was used to predict the 3D structure of the putative TM segments.

5.2.2 Molecular Dynamics Simulations

An equilibrated simulation box of 128 DMPC lipids solvated in 5673 water molecules was chosen as the starting configuration for the simulations. The helix monomers were embedded into the lipid bilayer parallel to the bilayer normal by using the protocols given by (Faraldo-Gómez et al, 2002) that was described in details in the two previous chapters. Then, Na⁺ and Cl⁻ ions were added randomly to neutralize the system and provided a close-to-physiological condition of 100 mM salt. The initial helix-lipids-water system was subjected to 500 steps of energy minimization using the steepest descent

algorithm followed by 200 ps simulation with harmonic position restraints. Finally, fully unrestrained productions run were performed for a simulation length of 200 ns.

The DMPC lipids were described with the Berger force-field parameters (Berger et al, 1997). The GROMOS96 FF53A6 force field (Oostenbrink et al, 2005) was used for the peptide that was showed to produce good interaction between lipid and peptide. SPC water model was used (Berendsen et al, 1981). All simulations were carried out by using GROMACS simulation software, version 4.0.3 (Hess et al, 2008). Periodic boundary conditions were used in all directions. Electrostatic interactions were calculated explicitly at a distance smaller than 1 nm, and long range electrostatic interactions were calculated by particle-mesh Ewald summation (Darden et al, 1993). Lennard-Jones interactions cutoff was 1 nm. All bonds were constrained by using the LINCS algorithm (Hess et al, 1997) allowing for an integration time step of 2 fs. The simulation temperature was kept constant by weakly ($\tau_p = 0.1$ ps) coupling the lipids, protein, and solvent separately to a temperature bath of 310 K. Likewise, the pressure was kept constant by weakly coupling the system to a pressure bath of 1 bar.

5.3 Results

5.3.1 Secondary Structure Prediction

We used the TOPCONS web server to predict the secondary structure of the studied envelope glycoproteins. The screening of a full length combination of E1 and E2 or prM and E amino acid sequences showed that both the E1/prM and the E2/E envelope glycoproteins contain at least one TM domain (**FIGURE 5.1**). The results for the genus Flavivirus (DENV, WNV and JEV) were consistently similar to each other with two TM domains located at their C-terminal regions of the prM and E proteins. In the case of BVDV, a representative from Pestivirus, two putative TM domains were predicted for the E1 region and only one putative TM domain for the E2 region. The TOPCONS server provided a slightly different prediction for the envelope glycoprotein of HCV-E1. However, the locations of the putative TM regions of the HCV-E1 and E2 are still the same as were previously suggested (Op De Beeck et al, 2000)(Cocquerel et al, 2000). The only difference is that one more putative TM region was predicted by four predictors of TOPCONS approximately 23 residues upstream from the HCV-E1 region of interest (P26664 350-383). The HCV-E2 protein was shown to have only one putative TM domain similarly to the BVDV-E2. However, when shorter sequence segments were

given as inputs (segments with approximately 24-35 amino acid residues with charged residues located at the centre), TOPCONS predicted single-pass helices (results not shown here) for the input segments.

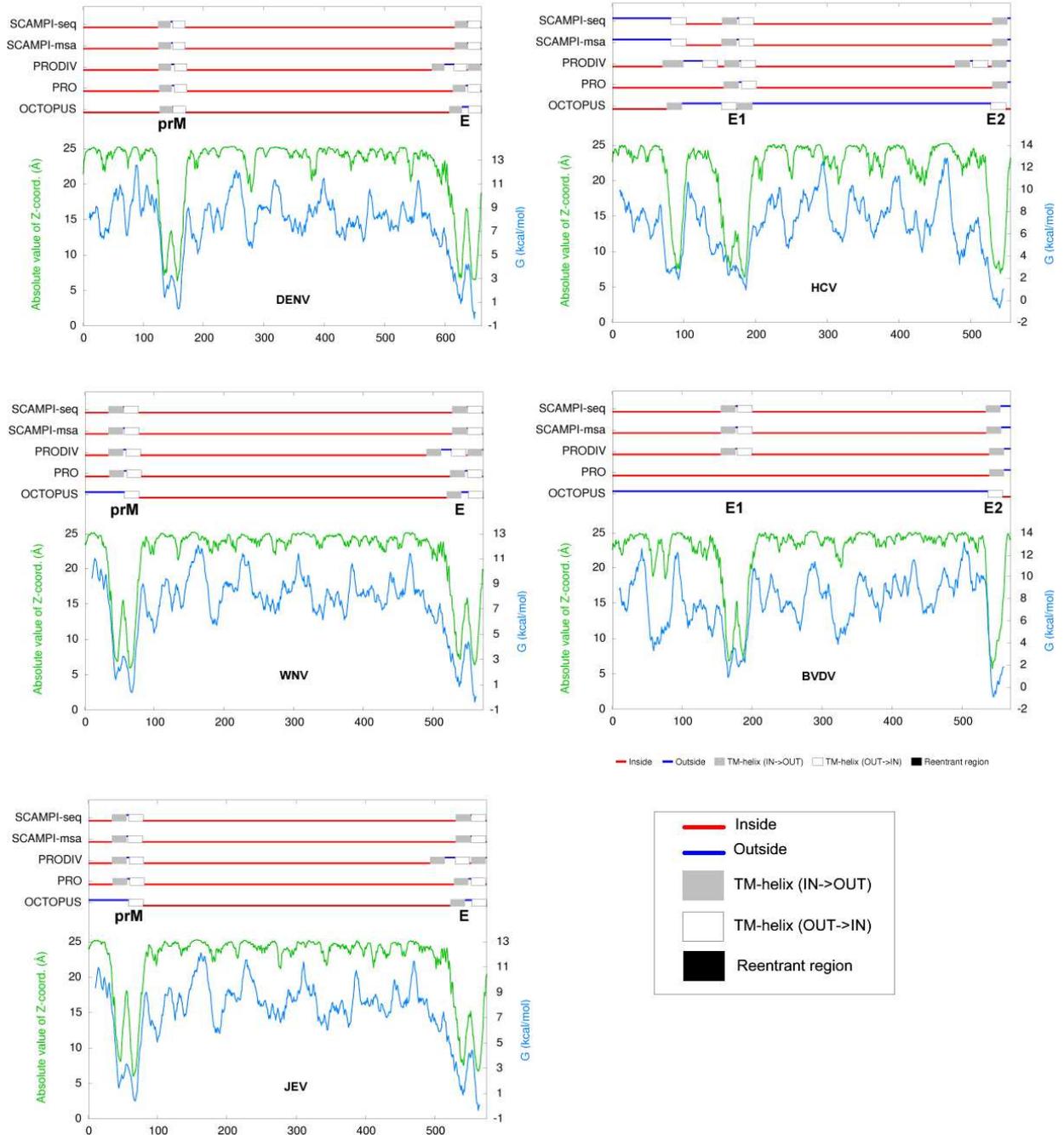


FIGURE 5.1 TOPCONS analyses for complete sequences of the E1/prM and E2/E envelope glycoproteins from Flaviviridae viruses. The left sides (upstream region) show the putative TM domains for the E1/prM and the right sides (downstream region) that of E2/E. These data were generated by the TOPCONS web server (<http://topcons.cbr.su.se>). Each result is labeled by the virus name.

For all studied segments, the ΔG_{app} values ranged from -0.60 to 2.54 kcal/mol (**TABLE 5.1**). The ΔG_{app} values for the putative TM domains obtained by the TOPCONS prediction server (**FIGURE 5.1**) are listed with the extension labels A and B in **TABLE 5.1**. The sequences with the extension label “MD” are the segments used in the MD simulations.

TABLE 5.1 DeltaG predictions of putative TM segments of the envelope glycoproteins from Flaviviridae viruses. The virus names with A, B, and C extensions indicate the suggested TM sequence segments from the TOPCONS analyses for the full sequence analyses of the envelope glycoproteins. The virus names with the extension ‘MD’ indicate the segments that were used for MD simulations. The criteria for these “MD” segments are (1) they contain a charged residue in the centre of the sequence segments, (2) their length is 23-40 amino acids, and (3) consensus results were obtained from at least three TOPCONS predictors.

Virus Genus & Type	Sequence (residue length)	ΔG (kcal/mol)
Flavivirus		
DENV-prM-A	WILRHPGFTIMAAILAYTIGT (21)	1.249
DENV-prM-B	QRALIFILLTAVAPSMTMRCI (21)	0.640
DENV-prM-MD	PGFTIMAAILAYTIGTTHFQRALIFILLTAVAP (33)	1.255
JEV-prM-A	WILRNPGYALVAAVIGWMLGS (21)	1.826
JEV-prM-B	MQRVVFAILLLLVAPAYSFNC (21)	-0.136
JEV-prM-MD	PGYAFLAATLGWMLGSNNGQVVFTILLLLVAP (33)	2.540
WNV-prM-A	WIIRNPGYAFLAATLGWMLGS (21)	2.406
WNV-prM-B	QRVVFTILLLLVAPAYSFNCL (21)	0.000
WNV-prM-MD	NPGYALVAAVIGWMLGSNTMQRVVFAILLLLVAP (34)	1.657
Pestivirus		
BVDV-E1-A	LTRIWNAATTTAFLVCLVKIV (21)	1.035
BVDV-E1-B	MVQGILWLLITGVQGHLDCK (21)	2.293
BVDV-E1-MD	RIWNAATTTAFLVCLVKIVRGQMVGILWLLITG (35)	0.917
Hepacivirus		
HCV-E1-A	LDMIAGAHWGVLAGIAYFSMV (21)	2.037
HCV-E1-B	WAKVLVLLLLFAGVDAETHVT (19)	1.939
HCV-E1-MD	GAHWGVLAGIAYFSMVGNWAKVLVLLLLFAGVDA (34)	1.687
Flavivirus		
DENV-E-A	GVSWTMKILIGVIITWIGMNS (21)	1.479
DENV-E-B	TLSVSLVVLVGVITLYLGV MV (21)	-0.403
DENV-E-MD	LGILLTWLGLNSRSTSLSMTC IAVGMVTLYL G (32)	1.275
JEV-E-A	GMSWITQGLMGALLWGMVNA (21)	1.644
JEV-E-B	RSIALAFLATGGVLVFLATNV (21)	0.051
JEV-E-MD	GALLWGMVGNARDRSIALAFLATGGVLVFLA (29)	2.053
WNV-E-A	FRSLFGGMSWITQGLLWGMGIN (26)	1.161
WNV-E-B	RSIAMTFLAVGGVLLFLSVNV (21)	-0.005
WNV-E-MD	GLLGALLWGMGINARDRSIAMTFLAVGGVLLFLSV (35)	1.707
Pestivirus		
BVDV-E2-C	ESILVVVVALLGGRYVLWLLV (21)	-0.433
BVDV-E2-MD	DYFAESILVVVVALLGGRYVLWLLVTYMLVSEKALG (37)	1.882
Hepacivirus		
HCV-E2-C	LLFLLADARVCSCLWMLLI (21)	-0.630
HCV-E2-MD	EYVVLLFLLADARVCSLWMLLIAQAEA (29)	0.070

5.3.2 I-TASSER 3D Structure Prediction

The I-TASSER prediction server predicted all the upstream TM domains (E1/prM) to exist as stable non-kinked helices (**FIGURE 5.2**). As expected, due to the absence of an ASP amino acid in the centre of the BVDV-E2 segment, no kinked behavior was observed. I-TASSER gave different results for the E2 and E types of envelope glycoproteins. The TM domains of DENV-E, WNV-E and JEV-E were predicted to centrally unfold. This behavior was probably caused by their central ASP residue. However, for the DENV-E that contain no central ASP, was also kinked as the others. Surprisingly, 4 out of 5 models of the HCV-E2 were predicted as non-kinked helices by I-TASSER. The HCV-E2 segment contains a negatively charged ASP in its centre. The structural templates that were used by the I-TASSER for modeling this target sequence were related to light-harvesting complexes or electron transport proteins (e.g. 1S5L, 1Q90, 1EHKC, 1JBO, 1DXR).

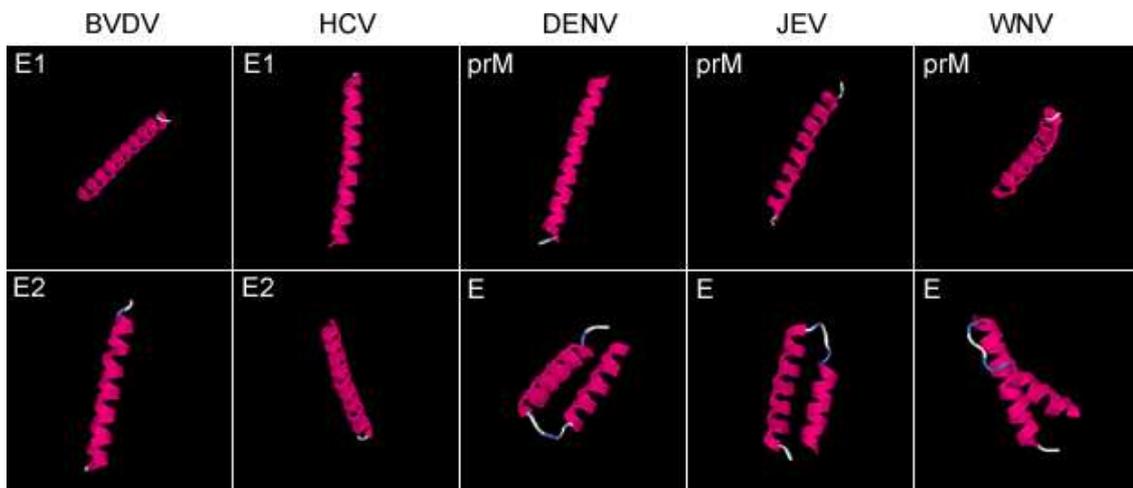


FIGURE 5.3 Representative 3D models obtained from the I-TASSER server. For the E1/prM : BVDV-E1 (2/5), HCV-E1 (1/3), DENV-prM (1/3), JEV-prM (1/3), WNV-prM (1/4) and the E2/E: BVDV-E2 (1/5), HCV-E2 (4/5), DENV-E (0/5), JEV-E (0/5) and WNV-E (0/5). The numbers in the parentheses indicate how many stable non-kinked helices were among all predicted models.

5.3.3 MD Simulations of TM Helix Monomers

We initially did several test simulations for the short hydrophobic stretches. However, they resulted in unfolded helices (results not shown here). Therefore we decided to use the longer sequence segments that contain highly polar residues located in the centre of the segment (**TABLE 5.1**, refer to those that are labeled with the extension –MD). These putative TM domains of the E1/prM were observed as stable helices during the 200 ns simulation time (**FIGURE 5.3**). Only the JEV-prM slightly kinked. In contrast, the TM domains of the E2/E severely kinked (HCV-E2, DENV-E, JEV-E and WNV-E) except the TM domain of the BVDV-E2. The RMSD analyses based on the ideal α -helix, clearly showed that the putative TM domain of E1/prM were closer to the ideal helix than the E2/E (**FIGURE 5.4**).

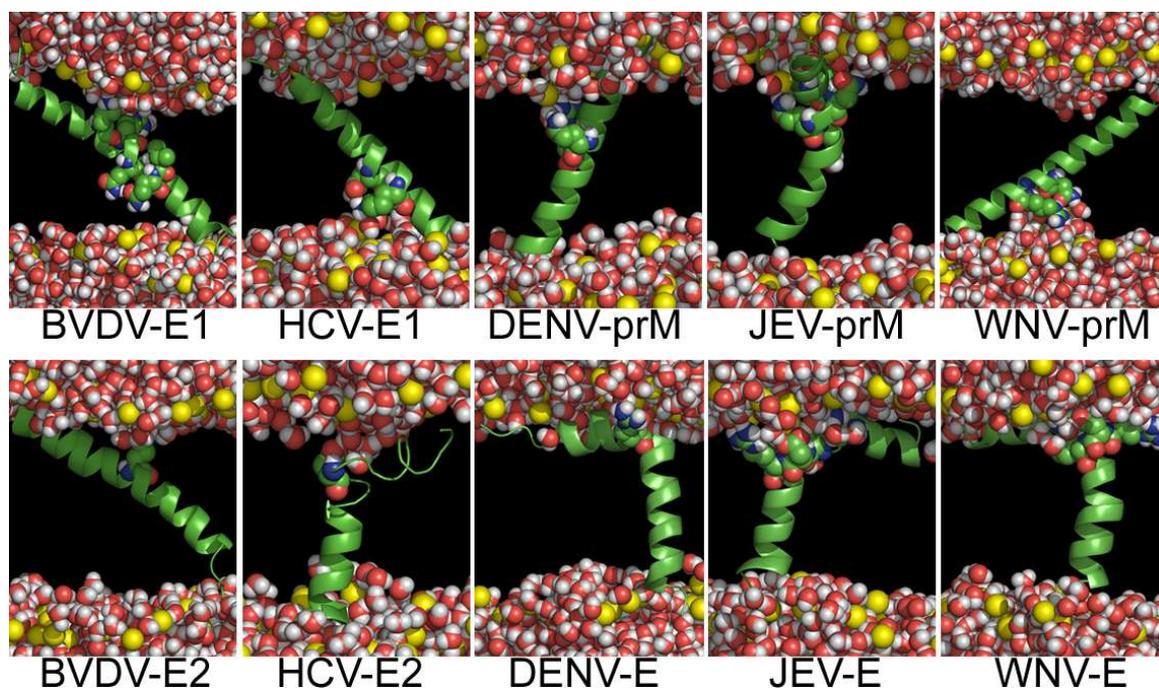


FIGURE 5.3 Final configurations after 100 ns of MD simulation of the prM/E1 and the E/E2 TM helix monomers from the Flaviviridae viruses. The TM helices are shown as helical cartoon. Bulk water, lipid head groups and the charged residues in the center of the TM helices are represented as atomic spheres. Lipid acyl chains are not shown for clarity.

The putative TM domains of E1/prM contain at least one positively charged residue. The DENV-E, WNV-E and JEV-E (Flaviviruses) contain only an Arg, HCV-E1 contains a Lys, and the BVDV-E1 contains both Lys and Arg residues (**TABLE 5.1**). During MD simulations, the side chains of their Lys and Arg residues managed to tilt and make contact with lipid head groups or bulk water.

The putative TM domains of the HCV-E2, WNV-E and JEV-E contain at least one negatively charged Asp residue. Asp which has the shortest side chain among all four charged amino acids was not efficiently anchored to the bilayer interface. Therefore this could be the possible cause for the observed kinking of the helix. We conclude that all the TM domains in this study which contain at least a single Asp in the centre of their TM segments resulted in severely kinked helices. The only exception to this 'role' is the DENV-E domain which contains no Asp. Also this peptide was observed as a kinked helix, similarly to the other members of the Flavivirus genus.

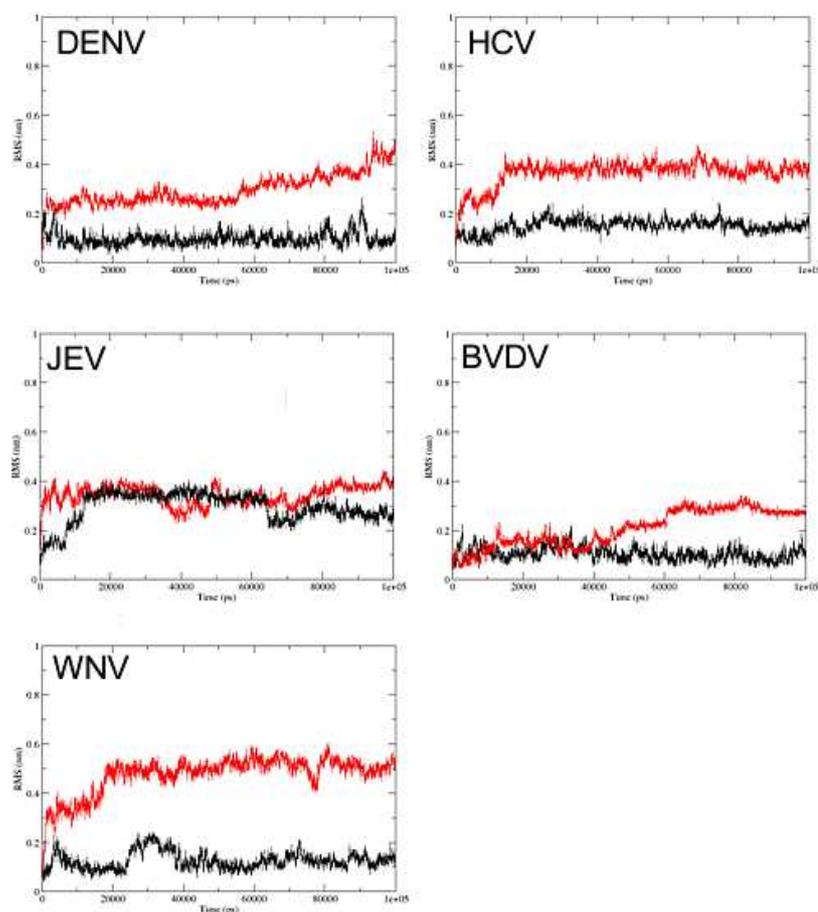


FIGURE 5.4 RMSD from the conformation of an ideal α -helix for the putative TM helices of the envelope glycoproteins from the Flaviviridae viruses. Five flanking residues each from both sides of the N and C-terminal regions were not included in the calculation. Each graph is labeled according to the virus name. Red lines refer to the TM domain of the E/prM, and the black lines refer to the TM domain of the E2/E.

The analysis of the lipid bilayer membrane thickness showed a range between 3.35 to 3.51 nm (TABLE 5.2). There is no significant difference of the membrane thickness results between the E1/prM and the E2/E

domains. The tilting angles were computed only for the unknicked TM domains. The smallest tilting angle was obtained for the TM domain of DENV-prM (**TABLE 5.2**). The other TM domains tilted by more than 42° which is obviously not in the typical range for tilting angles of helical TM proteins (White & Wimley, 1999). However, one should not forget that the ectodomain region of these envelope proteins could affect the structural behavior of the helix monomers. We speculate that, in nature, the tilting angle of these TM domains may be smaller when the large ectodomain region is present or/and due to the heterodimerization of both envelope glycoproteins.

TABLE 5.2 Structural behavior during MD simulations of putative TM helices from envelope glycoproteins of Flaviviridae viruses. The membrane thickness, computed by GridMAT-MD (Allen et al, 2009) indicates the average hydrophobic thickness of the membrane, measured from the average distances between the phosphate atoms of upper and lower leaflets.

Monomers of Flaviviruses	Membrane thickness (nm)	Tilting angle of helices (°) (80-100ns)
BVDV-E1	3.46 ± 0.42	64.6 ± 5.9
BVDV-E2	3.51 ± 0.51	78.0 ± 8.3
HCV-E1	3.42 ± 0.50	60.6 ± 4.9
HCV-E2	3.45 ± 0.54	Kinked
DENV-prM	3.50 ± 0.34	32.3 ± 5.0
DENV-E	3.45 ± 0.48	Kinked
JEV-prM	3.50 ± 0.30	42.5 ± 4.5
JEV-E	3.35 ± 0.52	Kinked
WNV-prM	3.52 ± 0.33	61.7 ± 4.0
WNV-E	3.52 ± 0.51	Kinked

5.4 Discussion and Conclusion

Sequence analyses of the envelope glycoproteins from the Flaviviridae viruses showed that these membrane proteins have TM domains that are located at the end of their C-terminal region. Full length sequence analyses by TOPCONS showed that the E1/prM envelope glycoproteins contain two putative TM domains connected by a short loop. For HCV-E1, this result is not in agreement with the experimental data that suggest that both regions exist as a single TM domain (Type I membrane protein) (Cocquerel et al, 2002)(Cocquerel et al, 2000). Moreover, the NMR-derived structure (1EMZ) of a segment from the HCV-E1 showed that the Asn367 and Lys370 residues are in the helical region (Op De Beeck et al, 2000). The structure suggests that both of these highly polar residues are interconnecting the two hydrophobic stretches and they are located in the core of the lipid bilayer. Both of

the hydrophobic segments are too short to exist as a single-pass helix in the lipid bilayer environment. In fact, we performed several test simulations for the hydrophobic stretches as helix monomers (approximately 20 amino acid residues), but they unfolded during the simulations.

The I-TASSER prediction server also predicted a model of HCV-E1 as an unkinked helix containing the polar and charged residues in the center of a segment of 26 amino acids length. I-TASSER gave similar 3D structural predictions for the other segments of E1/prM. Furthermore, the MD simulations also showed that these putative TM domains were stably retained as non-kinked helix monomers in the lipid bilayer environment. Interestingly, although these putative TM segments contain several charged amino acid residues located in the centre of the segments, their ΔG_{app} values are in the range found for other translocon integrated TM peptides and for TM segments from known 3D structures of membrane proteins (Hessa et al, 2007)(Hessa et al, 2009)(Hessa et al, 2005). This shows that the ΔG_{app} predictions also in agreement with the experimental data. However, the ΔG_{app} only illustrates the possibility of the TM domain to laterally integrate from the translocon into the membrane bilayer but not their structural conformation. Here we showed that most of the putative TM domains of the E2/E were severely kinked during the simulations. This type of kinked TM helix monomer does not exist in the current PDB database. Moreover, the unfolded and severely kinked helices indicate that they are unfavorable to remain as stable helix monomers in the lipid bilayer. We showed in Chapter 3 that the helical structure of HCV-E2 improved when it exists as a heterodimer interacting with HCV-E1 (Jusoh et al, 2010). Therefore, based on this extended study, we suggest a similar mechanism for the other members of Flaviviridae viruses. Furthermore, the E1-E2 or the prM-E envelope glycoproteins have been proposed based on experimental data to exist as a heterodimer and to be retained in the ER membrane. The HCV-E1-E2 heterodimer was shown to interact before entering the membrane bilayer environment (Cocquerel et al, 2002). Both of the TM domains of E1 and E2 were hypothesized to form a hairpin-like structure before the signal sequence is cleaved in the translocon environment. Then each of the TM domains formed a single pass type I TM helix in membrane bilayer environment (Cocquerel et al, 2002). The heterodimerization of these envelope glycoproteins was explained by the dependency of the correct folding of the E2/E in the presence of the E1/prM (Lorenz et al, 2003)(Cocquerel et al, 2002).

In conclusion, we showed the structural behavior of the putative TM helix monomers from the envelope glycoproteins of the Flaviviridae viruses in the membrane bilayer. The results presented here suggest that several TM domains of the E2/E may not exist as stable single-pass helices in nature depending on the type and total amount of the central polar and charged residues.

Chapter 6

Conclusion and Outlook

Acid amino sequences do not only play an important role for the functions of a TM domain but also determine its structural behavior. The MD simulations of the engineered monomer peptides were used as references. This study showed that a single-pass TM helix that contains at least one Asp residue located in the center severely kinked during the MD simulation. The similar strategy was applied to the TM domains of the envelope glycoproteins of the Flaviviridae viruses. These TM domains of HCV-E2, JEV-E, and WNV-E that contain an Asp amino acid were observed to be severely kinked during the simulation. However, a TM helix without a central Asp residue still has a tendency to kink if it contains several other highly polar charged residues as found for the TM domain of the DENV-E. In contrast, the TM helix monomers of HCV-E1, BVDV-E1, DENV-prM, and WNV-prM showed stable helical conformations during 200 ns of simulation time. Interestingly, the TM domain of E1/prM that is located in the upstream segments of the nascent polypeptide was shown to be more stable when expressed as a single-pass helix in bilayer compared to the TM domains of the E2/E. The unstable helical conformation of the E2/E monomers in the MD simulations could relate with the dependency of the E2/E envelope glycoprotein to the presence of the E1/prM as been suggested by the experimental data. Therefore, heterodimerization of the E1-E2 or prM-E envelope glycoproteins before entering the lipid bilayer may be the best option for the E2/E to prevent the misfolding event. Altogether, the results show that these putative TM domains of the envelope glycoproteins from Flaviviridae viruses illustrate similar structural behavior of their TM regions.

The classical MD simulation is the best unbiased method to study membrane proteins in their realistic bilayer environment. Here we clearly show that when a TM helix monomer contains charged or polar residue located in its center that helix has a tendency to tilt or severely kink depending on the type and total amounts of the highly polar residues. However, the straightforward MD simulation may not be suitable to sample alternative conformations for this type of TM helix monomer. The interaction between the charged and polar side chains with the hydrophilic atoms at the bilayer interfacial region started at the initial simulation time. This tight interaction might cause potential energy barriers to sample other type of conformations although the time of a simulation is exhaustively prolonged. Another strategy for future structural studies of this type of TM domain is to employ modern techniques to enhance conformational sampling during MD simulations as for example, the replica exchange or the umbrella sampling methods. But, we still need to keep in mind that getting more structural configurations does not mean getting correct structures. More experimental data are still needed to

verify the structural conformations for the TM domains of these envelope glycoproteins characterized in this study.

References

- Adamian L, Nanda V, DeGrado WF & Liang J (2005) Empirical lipid propensities of amino acid residues in multispan alpha helical membrane proteins. *Proteins* **59**: 496-509
- Adams PD, Engelman DM & Brünger AT (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* **26**: 257-261
- Alder BJ & Wainwright TE (1957) Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **27**: 1208-1209
- Alder BJ & Wainwright TE (1959) Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **31**: 459-466
- Aliste M & Tieleman DP (2005) Computer simulation of partitioning of ten pentapeptides Ace-WLXLL at the cyclohexane/water and phospholipid/water interfaces. *BMC Biochem.* **6**: 30
- Allen MP & Tildesley DJ (1989) Computer simulation of liquids. New York: Oxford University Press
- Allen WJ, Lemkul JA & Bevan DR (2009) GridMAT-MD: A grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.* **30**: 1952-1958
- Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL & Chang G (2009) Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. *Science* **323**: 1718-1722
- Almen M, Nordstrom K, Fredriksson R & Schioth H (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7**: 50
- Andreev OA, Karabadzhak AG, Dhammika Weerakkody, Andreev GO, Engelman DM & Reshetnyak YK (2010) pH (low) insertion peptide (pHLIP) inserts across a lipid bilayer as a helix and exits by a different path. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 4081-4086
- 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
- Bartenschlager R, Frese M & Pietschmann T (2004) Novel Insights into Hepatitis C Virus Replication and Persistence. *Adv. Virus Res.* **63**: 71-180
- 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
- Bassolino-Klimas D, Alper HE & Stouch TR (1993) Solute diffusion in lipid bilayer membranes: An atomic level study by molecular dynamics simulation. *Biochemistry* **32**: 12624-12637
- Beckmann R, Bubeck D, Grassucci R, Penczek P, Verschoor A, Blobel G & Frank J (1997) Alignment of Conduits for the Nascent Polypeptide Chain in the Ribosome-Sec61 Complex. *Science* **278**: 2123-2126
- Benedix A, Becker CM, de Groot BL, Caflisch A & Bockmann RA (2009) Predicting free energy changes using structural ensembles. *Nat. Methods* **6**: 3-4
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A & Haak JR (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**: 3684-3690

- Berendsen HJC, Postma JPM, Gunsteren WF & Hermans J (1981) Interaction models for water in relation to protein hydration. In *Intermolecular Forces* pp 331-342. The Netherlands: D. Reidel Publishing
- Berendsen HJC, van der Spoel D & van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**: 43-56
- Berg BVD, Clemons WM, Collinson I, Modis Y, Hartmann E, Harrison SC & Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* **427**: 36-44
- Berger O, Edholm O & Jähnig F (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **72**: 2002-2013
- Böckmann RA & Caflisch A (2005) Spontaneous Formation of Detergent Micelles around the Outer Membrane Protein OmpX. *Biophys. J.* **88**: 3191-3204
- Bond PJ & Sansom MS (2003) Membrane Protein Dynamics versus Environment: Simulations of OmpA in a Micelle and in a Bilayer. *J. Mol. Biol.* **329**: 1035-1053
- 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.* **10**: 2783-2793
- Bowie JU (1997) Helix packing in membrane proteins. *J. Mol. Biol.* **272**: 780-789
- Bu L, Im W & Brooks III CL (2007) Membrane Assembly of Simple Helix Homo-Oligomers Studied via Molecular Dynamics Simulations. *Biophys. J.* **92**: 854-863
- Bu L & Brooks III CL (2008) De Novo Prediction of the Structures of M. tuberculosis Membrane Proteins. *J. Am. Chem. Soc.* **130**: 5384-5385
- 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. *Biochim. Biophys. Acta, Biomembr.* **1716**: 1-10
- Canutescu AA, Shelenkov AA & Jr RLD (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12**: 2001-2014
- Caputo GA & London E (2004) Position and Ionization State of Asp in the Core of Membrane-Inserted α Helices Control Both the Equilibrium between Transmembrane and Nontransmembrane Helix Topography and Transmembrane Helix Positioning†. *Biochemistry* **43**: 8794-8806
- Chiu S, Clark M, Balaji V, Subramaniam S, Scott H & 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
- Ciczora Y, Callens N, Montpellier C, Bartosch B, Cosset F, De Beeck AO & Dubuisson J (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
- Ciczora Y, Callens N, Penin F, Pecheur E & 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. Gen. Virol.* **81**: 2372-2381

- Cocquerel L, Op de Beeck A, Lambot M, Roussel J, Delgrange D, Pillez A, Wychowski C, Penin F & Dubuisson J (2002) Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. *EMBO J.* **21**: 2893-2902
- 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
- Combet C, Garnier N, Charavay C, Grando D, Crisan D, Lopez J, Dehne-Garcia A, Geourjon C, Bettler E, Hulo C, Mercier PL, Bartenschlager R, Diepolder H, Moradpour D, Pawlotsky J, Rice CM, Trepo C, Penin F & Deleage G (2007) euHCVdb: the European hepatitis C virus database. *Nucleic Acids Res.* **35**: D363-366
- Cuthbertson JM, Bond PJ & Sansom MSP (2006) Transmembrane Helix–Helix Interactions: Comparative Simulations of the Glycophorin A Dimer†. *Biochemistry* **45**: 14298-14310
- Darden T, York D & Pedersen L (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**: 10089
- Dawson RJP & Locher KP (2006) Structure of a bacterial multidrug ABC transporter. *Nature* **443**: 180-185
- Deisenhofer J, Epp O, Miki K, Huber R & Michel H (1984) X-ray structure analysis of a membrane protein complex: Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J. Mol. Biol.* **180**: 385-398
- Deng Y & Roux B (2008) Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* **128**: 115103
- Domene C, Grottesi A & Sansom MS (2004) Filter Flexibility and Distortion in a Bacterial Inward Rectifier K^+ Channel: Simulation Studies of KirBac1.1. *Biophys. J.* **87**: 256-267
- Dorairaj S & Allen TW (2007) On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 4943-4948
- Doyle DA, Cabral JM, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT & MacKinnon R (1998) The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity. *Science* **280**: 69-77
- Driessen AJM (2005) Cell biology: Two pores better than one? *Nature* **438**: 299-300
- Dubuisson J, Duvet S, Meunier J, Op De Beeck A, Cacan R, Wychowski C & Cocquerel L (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
- Dudek J, Benedix J, Cappel S, Greiner M, Jalal C, Müller L & Zimmermann R (2009) Functions and pathologies of BiP and its interaction partners. *Cell. Mol. Life Sci.* **66**: 1556-1569
- Dudek J, Greiner M, Muller A, Hendershot LM, Kopsch K, Nastainczyk W & Zimmermann R (2005) ERj1p has a basic role in protein biogenesis at the endoplasmic reticulum. *Nat. Struct. Mol. Biol.* **12**: 1008-1014
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792-1797
- Ernst J, Clubb R, Zhou H, Gronenborn A & Clore G (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **267**: 1813-1817

- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H & Pedersen LG (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**: 8577
- Faraldo-Gómez J, Smith G & Sansom M (2002) Setting up and optimization of membrane protein simulations. *European Biophysics Journal* **31**: 217-227
- Fischer WB & Sansom MS (2002) Viral ion channels: structure and function. *Biochim. Biophys. Acta, Biomembr.* **1561**: 27-45
- 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.* 140-144
- Freddolino PL, Liu F, Gruebele M & Schulten K (2008) Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys. J.* **94**: L75-L77
- Freites JA, Tobias DJ, von Heijne G & White SH (2005) Interface connections of a transmembrane voltage sensor. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 15059-15064
- Galtier N, Gouy M & Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Bioinformatics* **12**: 543-548
- Gompper G & Schick M (2008) Soft Matter: Volume 4: Lipid Bilayers and Red Blood Cells. Wiley-VCH
- Görlich D, Prehn S, Hartmann E, Kalies K & Rapoport TA (1992) A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* **71**: 489-503
- Griepnerau B, Leis S, Schneider MF, Sikor M, Steppich D & Böckmann RA (2007) 1-Alkanols and membranes: A story of attraction. *Biochim. Biophys. Acta, Biomembr.* **1768**: 2899-2913
- de Groot BL & Grubmuller H (2001) Water Permeation Across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF. *Science* **294**: 2353-2357
- Gumbart J & Schulten K (2006) Molecular Dynamics Studies of the Archaeal Translocon. *Biophys. J.* **90**: 2356-2367
- van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholtz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA & Yu HB (2006) Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chem. Int. Ed.* **45**: 4064-4092
- Hamman BD, Chen J, Johnson EE & Johnson AE (1997) The Aqueous Pore through the Translocon Has a Diameter of 40-60 Å during Cotranslational Protein Translocation at the ER Membrane. *Cell* **89**: 535-544
- Hanein D, Matlack KES, Jungnickel B, Plath K, Kalies K, Miller KR, Rapoport TA & Akey CW (1996) Oligomeric Rings of the Sec61p Complex Induced by Ligands Required for Protein Translocation. *Cell* **87**: 721-732
- Hedin LE, Öjemalm K, Bernsel A, Hennerdal A, Illergård K, Enquist K, Kauko A, Cristobal S, von Heijne G, Lerch-Bader M, Nilsson I & Elofsson A (2010) Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context. *J. Mol. Biol.* **396**: 221-229
- von Heijne G (1992) Membrane protein structure prediction : Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**: 487-494

- Hénin J, Pohorille A & Chipot C (2005) Insights into the Recognition and Association of Transmembrane α -Helices. The Free Energy of α -Helix Dimerization in Glycophorin A. *Journal of the American Chemical Society* **127**: 8478-8484
- Herce HD & Garcia AE (2007) Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 20805-20810
- Hess B, Bekker H, Berendsen H & Fraaije J (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**: 1472, 1463
- 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
- Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH & von Heijne G (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**: 377-381
- Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH & von Heijne G (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**: 1026-1030
- Hessa T, Reithinger JH, von Heijne G & Kim H (2009) Analysis of Transmembrane Helix Integration in the Endoplasmic Reticulum in *S. cerevisiae*. *J. Mol. Biol.* **386**: 1222-1228
- Hessa T, White SH & von Heijne G (2005) Membrane Insertion of a Potassium-Channel Voltage Sensor. *Science* **307**: 1427
- Hopkins AL & Groom CR (2002) The druggable genome. *Nat. Rev. Drug Discovery* **1**: 727-730
- Humphrey W, Dalke A & Schulten K (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**: 33-38
- Im W, Feig M & Brooks III CL (2003) An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys. J.* **85**: 2900-2918
- Jaud S, Fernandez-Vidal M, Nilsson I, Meindl-Beinker NM, Hubner NC, Tobias DJ, von Heijne G & White SH (2009) Insertion of short transmembrane helices by the Sec61 translocon. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 11588-11593
- Joh NH, Min A, Faham S, Whitelegge JP, Yang D, Woods VL & Bowie JU (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**: 1266-1270
- Johansson ACV & Lindahl E (2009a) Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 15684-15689
- Johansson ACV & Lindahl E (2009b) The role of lipid composition for insertion and stabilization of amino acids in membranes. *J. Chem. Phys.* **130**: 185101-8
- Johansson AC & Lindahl E (2006) Amino-Acid Solvation Structure in Transmembrane Helices from Molecular Dynamics Simulations. *Biophys. J.* **91**: 4450-4463
- Jorgensen WL, Maxwell DS & Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**: 11225-11236

- Juretic D, Zoranic L & Zucic D (2002) Basic Charge Clusters and Predictions of Membrane Protein Topology. *J. Chem. Inf. Comput. Sci.* **42**: 620-632
- Jusoh SA, Welsch C, Siu SWI, Böckmann R & Helms V (2010) Contribution of charged and polar residues for the formation of the E1–E2 heterodimer from Hepatitis C Virus. *J. Mol. Model.* **16**: 1625-1637
- Kagan VE, Fabisiak JP, Shvedova AA, Tyurina YY, Tyurin VA, Schor NF & Kawai K (2000) Oxidative signaling pathway for externalization of plasma membrane phosphatidylserine during apoptosis. *FEBS Lett.* **477**: 1-7
- 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
- Kauko A, Hedin LE, Thebaud E, Cristobal S, Elofsson A & von Heijne G (2010) Repositioning of Transmembrane α -Helices during Membrane Protein Folding. *J. Mol. Biol.* **397**: 190-201
- Khalili-Araghi F, Gumbart J, Wen P, Sotomayor M, Tajkhorshid E & Schulten K (2009) Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* **19**: 128-137
- Krepkiy D, Mihailescu M, Freitas JA, Schow EV, Worcester DL, Gawrisch K, Tobias DJ, White SH & Swartz KJ (2009) Structure and hydration of membranes embedded with voltage-sensing domains. *Nature* **462**: 473-479
- Krogh A, Larsson B, von Heijne G & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**: 567-580
- Krüger J & Fischer WB (2008) Exploring the conformational space of Vpu from HIV-1: A versatile adaptable protein. *J. Comput. Chem.* **29**: 2416-2424
- Kučerka N, Liu Y, Chu N, Petrache HI, Tristram-Nagle S & Nagle JF (2005) Structure of Fully Hydrated Fluid Phase DMPC and DLPC Lipid Bilayers Using X-Ray Scattering from Oriented Multilamellar Arrays and from Unilamellar Vesicles. *Biophys. J.* **88**: 2626-2637
- Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105-132
- Langosch D, Brosig B, Kolmar H & Fritz H (1996) Dimerisation of the Glycophorin A Transmembrane Segment in Membranes Probed with the ToxR Transcription Activator. *J. Mol. Biol.* **263**: 525-530
- Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T & Higgins D (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948
- Lazaridis T (2003) Effective energy function for proteins in lipid membranes. *Proteins Struct. Funct. Bioinf.* **52**: 176-192
- Lemmon MA, Flanagan JM, Treutlein HR, Zhang J & Engelman DM (1992) Sequence specificity in the dimerization of transmembrane .alpha.-helices. *Biochemistry* **31**: 12719-12725
- Lev S (2010) Non-vesicular lipid transport by lipid-transfer proteins and beyond. *Nat Rev Mol Cell Biol* **11**: 739-750

- Lewis BA & Engelman DM (1983) Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *J. Mol. Biol.* **166**: 211-217
- Li S, Su Y, Luo W & Hong M (2010) Water-Protein Interactions of an Arginine-Rich Membrane Peptide in Lipid Bilayers Investigated by Solid-State Nuclear Magnetic Resonance Spectroscopy. *J. Phys. Chem. B* **114**: 4063-4069
- Lindenbach BD, Thiel H & Rice CM (2001) Flaviviridae: The viruses and their replication. In *In Fields Virology*, D.M. Knipe and P.M. Howley, editors pp 991-1042. Philadelphia: Lippincott Williams & Wilkins
- Lindenbach BD, Thiel H & Rice CM (2007) Flaviviridae: The viruses and their replication. In *Field Virology* pp 1101-1152. Philadelphia: Lippincott-Raven Publishers
- Lipowsky R & Sackmann E (1995) Structure and dynamics of membranes. Amsterdam: Elsevier
- Liu L, Quillin ML & Matthews BW (2008) Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 14406-14411
- Lomize AL, Pogozheva ID, Lomize MA & Mosberg HI (2006) Positioning of proteins in membranes: A computational approach. *Protein Sci.* **15**: 1318-1333
- London E & Shahidullah K (2009) Transmembrane vs. non-transmembrane hydrophobic helix topography in model and natural membranes. *Curr. Opin. Struct. Biol.* **19**: 464-472
- Long SB, Campbell EB & MacKinnon R (2005) Crystal Structure of a Mammalian Voltage-Dependent Shaker Family K⁺ Channel. *Science* **309**: 897-903
- Lorenz IC, Kartenbeck J, Mezzacasa A, Allison SL, Heinz FX & Helenius A (2003) Intracellular Assembly and Secretion of Recombinant Subviral Particles from Tick-Borne Encephalitis Virus. *J. Virol.* **77**: 4370-4382
- MacCallum JL, Bennett WFD & Tieleman DP (2008) Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **94**: 3393-3404
- MacKenzie KR, Prestegard JH & Engelman DM (1997) A Transmembrane Helix Dimer: Structure and Implications. *Science* **276**: 131-133
- MacKerell, Bashford D, Bellott, Dunbrack, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE et al (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J. Phys. Chem. B* **102**: 3586-3616
- Marrink SJ, Lindahl E, Edholm O & Mark AE (2001) Simulation of the Spontaneous Aggregation of Phospholipids into Bilayers. *J. Am. Chem. Soc.* **123**: 8638-8639
- Mathai JC, Tristram-Nagle S, Nagle JF & Zeidel ML (2010) Structural Determinants of Water Permeability through the Lipid Membrane. *J. Gen. Physiol.* **131**: 69-76
- Matsuura Y & Miyamura T (1993) The molecular biology of hepatitis C virus. *Semin. Virol.* **4**: 297-304
- Matthews E, Zoonens M & Engelman D (2006) Dynamic Helix Interactions in Transmembrane Signaling. *Cell* **127**: 447-450
- McCammon JA, Gelin BR & Karplus M (1977) Dynamics of folded proteins. *Nature* **267**: 585-590

- Meindl-Beinker NM, Lundin C, Nilsson I, White SH & von Heijne G (2006) Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep.* **7**: 1111-1116
- Ménéret J, Hegde RS, Aguiar M, Gygi SP, Park E, Rapoport TA & Akey CW (2008) Single Copies of Sec61 and TRAP Associate with a Nontranslating Mammalian Ribosome. *Structure* **16**: 1126-1137
- Ménéret J, Hegde RS, Heinrich SU, Chandramouli P, Ludtke SJ, Rapoport TA & Akey CW (2005) Architecture of the Ribosome-Channel Complex Derived from Native Membranes. *J. Mol. Biol.* **348**: 445-457
- Ménéret J, Neuhof A, Morgan DG, Plath K, Radermacher M, Rapoport TA & Akey CW (2000) The Structure of Ribosome-Channel Complexes Engaged in Protein Translocation. *Molecular Cell* **6**: 1219-1232
- Mitra K, Schaffitzel C, Shaikh T, Tama F, Jenni S, Brooks CL, Ban N & Frank J (2005) Structure of the E. coli protein-conducting channel bound to a translating ribosome. *Nature* **438**: 318-324
- Monticelli L, Robertson KM, MacCallum JL & Tieleman D (2004) Computer simulation of the KvAP voltage-gated potassium channel: steered molecular dynamics of the voltage sensor. *FEBS Lett.* **564**: 325-332
- Moradpour D, Penin F & Rice CM (2007) Replication of hepatitis C virus. *Nat. Rev. Microbiol.* **5**: 453-463
- Mottamal M, Zhang J & Lazaridis T (2006) Energetics of the native and non-native states of the glycoporphin transmembrane helix dimer. *Proteins Struct. Funct. Bioinf.* **62**: 996-1009
- Mukhopadhyay S, Kuhn RJ & Rossmann MG (2005) A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **3**: 13-22
- Murray CL, Jones CT & Rice CM (2008) Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis. *Nat. Rev. Microbiol.* **6**: 699-708
- Nagle J (1993) Area/lipid of bilayers from NMR. *Biophys. J.* **64**: 1476-1481
- Nagle JF & Tristram-Nagle S (2000) Structure of lipid bilayers. *Biochim. Biophys. Acta, Biomembr.* **1469**: 159-195
- 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
- Oostenbrink C, Soares TA, Vegt NFA & Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **34**: 273-284
- 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
- Op De Beeck A, Montserret R, Duvet S, Cocquerel L, Cacan R, Barberot B, Le Maire M, Penin F & Dubuisson J (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
- Opar A (2010) Excitement grows for potential revolution in hepatitis C virus treatment. *Nat. Rev. Drug Discovery* **9**: 501-503

- Overington JP, Al-Lazikani B & Hopkins AL (2006) How many drug targets are there? *Nat. Rev. Drug Discovery* **5**: 993-996
- Park Y & Helms V (2008a) Prediction of the translocon-mediated membrane insertion free energies of protein sequences. *Bioinformatics* **24**: 1271-1277
- Park Y & Helms V (2008b) MINS2: Revisiting the molecular code for transmembrane-helix recognition by the Sec61 translocon. *Bioinformatics* **24**: 1819-1820
- Patra M, Karttunen M, Hyvönen MT, Falck E & Vattulainen I (2004) Lipid Bilayers Driven to a Wrong Lane in Molecular Dynamics Simulations by Subtle Changes in Long-Range Electrostatic Interactions. *J. Phys. Chem. B* **108**: 4485-4494
- Penin F, Dubuisson J, Rey FA, Moradpour D & Pawlowsky J (2004) Structural biology of hepatitis C virus. *Hepatology* **39**: 5-19
- Petrache HI, Tristram-Nagle S & Nagle JF (1998) Fluid phase structure of EPC and DMPC bilayers. *Chem. Phys. Lipids* **95**: 83-94
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**: 1781-1802
- Pietzsch J (2004) Background: Mind the membrane. *Nature* 1-4
- Pitonzo D & Skach WR (2006) Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim. Biophys. Acta, Biomembr.* **1758**: 976-988
- Plath K, Mothes W, Wilkinson BM, Stirling CJ & Rapoport TA (1998) Signal Sequence Recognition in Posttranslational Protein Transport across the Yeast ER Membrane. *Cell* **94**: 795-807
- Poger D & Mark AE (2010) On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *J. Chem. Theory Comput.* **6**: 325-336
- Poger D, Van Gunsteren WF & Mark AE (2010) A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **31**: 1117-1125
- Ponder JW & Case DA (2003) Force Fields for Protein Simulations. In *Protein Simulations* pp 27-85. Academic Press
- Popot J & Engelman DM (2000) Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **69**: 881-922
- Potapov V, Cohen M & Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *PEDS* **22**: 553-560
- Radzicka A & Wolfenden R (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27**: 1664-1670
- Raghuraman H & Chattopadhyay A (2006) Melittin: a Membrane-active Peptide with Diverse Functions. *Biosci. Rep.* **27**: 189-223
- Rahman A (1964) Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.* **136**: A405

- Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**: 663-669
- Rapoport TA, Goder V, Heinrich SU & Matlack KE (2004) Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* **14**: 568-575
- Rapoport TA, Jungnickel B & Kutay U (1996) Protein Transport Across the Eukaryotic Endoplasmic Reticulum and Bacterial Inner Membranes. *Annu. Rev. Biochem.* **65**: 271-303
- Rodnin MV, Kyrychenko A, Kienker P, Sharma O, Posokhov YO, Collier RJ, Finkelstein A & Ladokhin AS Conformational Switching of the Diphtheria Toxin T Domain. *Journal of Molecular Biology* **In Press**
- Ronecker S, Zimmer G, Herrler G, Greiser-Wilke I & Grummer B (2008) Formation of bovine viral diarrhea virus E1-E2 heterodimers is essential for virus entry and depends on charged residues in the transmembrane domains. *J. Gen. Virol.* **89**: 2114-2121
- Rost B, Fariselli P & Casadio R (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* **5**: 1704-1718
- Roth J, Gähler F & Trebin H (2000) A molecular dynamics Run with 5180116000 particles. *Int. J. Mod. Phys. C* **11**: 317-322
- Sanner MF, Olson AJ & Spehner J (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**: 305-320
- Sansom MS, Shrivastava IH, Bright JN, Tate J, Capener CE & Biggin PC (2002) Potassium channels: structures, models, simulations. *Biochim. Biophys. Acta, Biomembr.* **1565**: 294-307
- Saparov SM, Erlandson K, Cannon K, Schaletzky J, Schulman S, Rapoport TA & Pohl P (2007) Determining the Conductance of the SecY Protein Translocation Channel for Small Molecules. *Molecular Cell* **26**: 501-509
- Sapay N & Tieleman DP (2008) Molecular Dynamics Simulation of Lipid-Protein Interactions. In *Computational Modeling of Membrane Bilayers* USA: Academic Press
- 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 beta-branched residues at neighboring positions. *J. Mol. Biol* **296**: 921-936
- Shibatani T, David LL, McCormack AL, Frueh K & Skach WR (2005) Proteomic Analysis of Mammalian Oligosaccharyltransferase Reveals Multiple Subcomplexes that Contain Sec61, TRAP, and Two Potential New Subunits. *Biochemistry* **44**: 5982-5992
- Simmonds P, Bukh J, Combet C, Deléage G, Enomoto N, Feinstone S, Halfon P, Inchauspé G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky J, Penin F, Sablon E, Shin-I T, Stuyver LJ, Thiel H, Viazov S et al (2005) Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**: 962-973
- Singer SJ (1974) The Molecular Organization of Membranes. *Annu. Rev. Biochem.* **43**: 805-833
- Singer SJ & Nicolson GL (1972) The Fluid Mosaic Model of the Structure of Cell Membranes. *Science* **175**: 720-731
- Siu SWI, Vácha R, Jungwirth P & Böckmann RA (2008) Biomolecular simulations of membranes: Physical properties from different force fields. *J. Chem. Phys.* **128**: 125103

- Skach WR (2009) Cellular mechanisms of membrane protein folding. *Nat. Struct. Mol. Biol.* **16**: 606-612
- Sleight RG (1987) Intracellular Lipid Transport in Eukaryotes. *Annu. Rev. Physiol.* **49**: 193-208
- Stillinger FH & Rahman A (1974) Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* **60**: 1545
- Stouch TR (1993) Lipid Membrane Structure and Dynamics Studied by All-Atom Molecular Dynamics Simulations of Hydrated Phospholipid Bilayers - Molecular Simulation. *Mol. Simul.* **10**: 335-362
- Strode GK (1951) Yellow Fever. New York: McGraw-Hill
- Sundararajan V (2008) Computational Modeling of Membrane Bilayers. Academic Press
- 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
- Terstappen GC & Reggiani A (2001) In silico research in drug discovery. *Trends Pharmacol. Sci.* **22**: 23-26
- Thomas HC, Lemon SM & Zuckerman AJ (2005) Viral hepatitis. Wiley-Blackwell
- Tieleman DP & Berendsen HJC (1996) Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *J. Chem. Phys.* **105**: 4871
- Tieleman DP, Marrink SJ & Berendsen HJC (1997) A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochim. Biophys. Acta, Biomembr.* **1331**: 235-270
- Tresset G (2009) The multiple faces of self-assembled lipidic systems. *PMC Biophys.* **2**: 3
- Treutlein HR, Lemmon MA, Engelman DM & Brünger AT (1992) The glycoporphin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry* **31**: 12726-12732
- Tusnady GE & Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849-850
- Tyedmers J, Lerner M, Bies C, Dudek J, Skowronek MH, Haas IG, Heim N, Nastainczyk W, Volkmer J & Zimmermann R (2000) Homologs of the yeast Sec complex subunits Sec62p and Sec63p are abundant proteins in dog pancreas microsomes. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 7214-7219
- Ulmschneider MB & Sansom MSP (2001) Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta, Biomembr.* 1-14
- Ulmschneider MB, Ulmschneider JP, Sansom MS & Di Nola A (2007a) A Generalized Born Implicit-Membrane Representation Compared to Experimental Insertion Free Energies. *Biophys. J.* **92**: 2338-2349
- Ulmschneider MB, Ulmschneider JP, Sansom MS & Nola AD (2007b) A Generalized Born Implicit-Membrane Representation Compared to Experimental Insertion Free Energies. *Biophys. J.* **92**: 2338-2349
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE & Berendsen HJC (2005) GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**: 1701-1718

- Vlachakis D (2008) The Flaviviridae Viral Family: Properties and Genome. Dimitrios Vlachakis.
- Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Krausslich H, Mizokami M, Bartenschlager R & Liang TJ (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* **11**: 791-796
- Wallin E & von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**: 1029-1038
- Ward A, Reyes CL, Yu J, Roth CB & Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 19005 -19010
- White SH (2004) The progress of membrane protein structure determination. *Protein Sci.* **13**: 1948-1949
- White SH (2009) Biophysical dissection of membrane proteins. *Nature* **459**: 344-346
- White SH & von Heijne G (2008) How Translocons Select Transmembrane Helices. *Annu. Rev. Biophys.* **37**: 23-42
- White SH & Wimley WC (1999) Membrane protein folding and stability: Physical Principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**: 319-365
- Wiedmann M, Kurzchalia TV, Hartmann E & Rapoport TA (1987) A signal sequence receptor in the endoplasmic reticulum membrane. *Nature* **328**: 830-833
- Wimley WC, Creamer TP & White SH (1996) Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host–Guest Pentapeptides. *Biochemistry* **35**: 5109-5124
- Wimley WC & White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **3**: 842-848
- Woolf TB (1998) Molecular dynamics simulations of individual alpha-helices of bacteriorhodopsin in dimyristoylphosphatidylcholine. II. Interaction energy analysis. *Biophys. J.* **74**: 115-131
- Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N & Suzek B (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* **34**: D187-D191
- Yoo J & Cui Q (2008) Does Arginine Remain Protonated in the Lipid Membrane? Insights from Microscopic pKa Calculations. *Biophys. J.* **94**: L61-L63
- Zakharov SD, Zhálnina MV, Sharma O & Cramer WA (2006) The Colicin E3 Outer Membrane Translocon: Immunity Protein Release Allows Interaction of the Cytotoxic Domain with OmpF Porin†. *Biochemistry* **45**: 10199-10207
- Zhang B & Miller TF (2010) Hydrophobically stabilized open state for the lateral gate of the Sec translocon. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 5399-5404
- Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinf.* **9**: 40
- Zhao G & London E (2006) An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. *Protein Sci.* **15**: 1987-2001
- Zimmermann R, Eyrisch S, Ahmad M & Helms V (2010) Protein translocation across the ER membrane. *Biochim. Biophys. Acta, Biomembr.* **In Press**