Structural features and interactions between transmembrane and juxtamembrane domains in membrane proteins

Surya, Wahyu

2017

Surya, W. (2017). Structural features and interactions between transmembrane and juxtamembrane domains in membrane proteins. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/69964>

<https://doi.org/10.32657/10356/69964>

NANYANG
TECHNOLOGICAL
UNIVERSITY

STRUCTURAL FEATURES AND INTERACTIONS
BETWEEN TRANSMEMBRANE AND JUXTAMEMBRANE
DOMAINS IN MEMBRANE PROTEINS

WAHYU SURYA

SCHOOL OF BIOLOGICAL SCIENCES

2017

STRUCTURAL FEATURES AND INTERACTIONS BETWEEN
TRANSMEMBRANE AND JUXTAMEMBRANE DOMAINS
IN MEMBRANE PROTEINS

WAHYU SURYA

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2017

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to my Ph.D. supervisor, A/P Dr. Jaume Torres for his expert guidance, valuable discussions, critical advice, and unrestricted support throughout my Ph.D. candidature period. I would like to acknowledge our collaborators Dr. Oscar Millet and Dr. Tammo Diercks from CIC BioGUNE, Bilbao, Spain; with whom I spent 6 weeks learning protein NMR techniques from the very beginning, and collected data for the integrin study at their facility. I would also like to thank Dr. Li Yan, a former postdoctoral fellow in our group, under whom I continued my forage into the field of protein NMR, for teaching and helping me with NMR data collection and analysis, and together co-authoring several papers and reviews on the subjects of integrins and viroporins. I also thank Dr. Antony Hardjojo and Rheza Louis for proofreading this thesis.

I am really thankful to all past and present members of Jaume's research group for the friendship, teamwork, and scientific discussions that we shared: Dr. Ardcharaporn Vararattanavech, Dr. Krupakar Parthasarathy, Dr. Gan Siok Wan, Suweeraya Limpanawat, Soon Cin Huang, Yeo Chiew Ying, Dr. Mukesh Mahajan, Chaitra Gopalakhrisnamoorthi, Monisha Thiagarajan, and Janet To. I am also thankful to the attachment students who taught me that a true sign of understanding is found in successfully teaching others: Choong Yeu Khai, Stephanie, Kathleen Koh, Yeoh Ee Ping, and Evonne Teo.

I thank all admin and research staff of SBS who have helped me in various ways (just to mention a few without undermining the rest): Dr. Ye Hong from NMR facility; Nur Eliza from Peptide Synthesis Facility; Geraldine Loke from General Store; Eugene Seetoh and Tan Kim Yin from General Office; Chong Pui Ling and Linda Chiew from Finance; and Tin Tun and Lionel Yap from Safety.

Last but not least, I would like to extend my gratitude to Prof. Lars Nordenskiöld and Prof. Daniela Rhodes for their kind moral support and encouragement during the final stages of my thesis writing, and for sharing valuable insights on writing a doctoral thesis. I would also like to thank everyone in Prof. Lars' research group, Prof. Daniela's research group, and NISB for their friendship and moral support.

TABLE OF CONTENTS

Acknowledgements	i
Table of contents	ii
List of figures	v
List of tables	vii
Abbreviations	viii
Abstract	xi
1. Introduction	1
1.1. Membrane proteins: what they are	1
1.2. Membrane proteins: what they do	2
1.3. Structural studies of membrane proteins	4
1.4. Aim of this study	9
2. Materials and Methods	10
2.1. Sample preparation	10
2.2. Secondary structure determination	13
2.3. Oligomeric state determination	14
2.4. Single channel activity measurement	15
2.5. NMR experiments and data analysis	16
3. Integrin Juxtamembrane Structure and Heterodimer Formation	20
3.1. Introduction	20
3.1.1. Structural models for integrin heterodimer: α IIB/ β 3 TM	20
3.1.2. The controversial GFFKR backbone reversal motif	23
3.1.3. The integrin α L/ β 2 TM	25
3.1.4. Aim and approach	26
3.2. Construct design and sample preparation	27
3.3. Reconstitution in bicelles and secondary structure elements	28
3.4. Solution NMR structure of integrin TM peptides	30
3.5. Reversal of the GFFKR motif backbone in α L TM domain	36
3.6. The α L/ β 2 heterodimer interface	38
3.7. Summary and future works	41
	ii

4. Coronavirus E Protein Extramembrane Structure and Homooligomer Formation	43
4.1. Introduction	43
4.1.1. The viroporins	43
4.1.2. The coronavirus E protein	44
4.1.3. Structural models for CoV E protein	45
4.1.4. The extramembrane domain of CoV E protein	49
4.1.5. Aim and approach	49
4.2. Construct optimization and sample preparation	50
4.3. Validation of the E _{TR} construct	53
4.3.1. Secondary structure of E _{TR}	53
4.3.2. Oligomeric state of E _{TR} construct	54
4.3.3. Channel activity of E _{TR} construct	56
4.4. Validation of the micellar reconstitution environment	58
4.4.1. Secondary structure of E _{TR} in detergent micelles	59
4.4.2. Oligomeric state of E _{TR} in detergent micelles	60
4.4.3. HMA binding on E _{TR} in detergent micelles	63
4.5. Solution NMR structure of E protein monomer	65
4.5.1. Overall structure	65
4.5.2. Membrane-embedded region	68
4.5.3. The structure of the extramembrane domain	70
4.6. Solution NMR model of the E protein pentamer	72
4.7. Summary	75
5. Discussions	76
5.1. Juxtamembrane structure of α L and heterodimer interface of α L/ β 2	76
5.1.1. Alternate structure for the juxtamembrane GFFKR motif of α subunit	76
5.1.2. Heterodimerization of α L/ β 2: intermediate exchange	77
5.2. Structural models for CoV E protein monomers and pentamers	78
5.2.1. The structural flexibility of the C-terminal hydrophobic region	78
5.2.2. The pentamer bundle and the ion channel activity	80
5.2.3. A pentameric model that represents the “open” form of the channel	83
5.3. Structural preference in different environments	84
5.3.1. Variations in membrane-inserted residues	84
5.3.2. Effects on secondary structure propensity	86
	iii

6. Conclusions	88
7. References	89
8. Appendix	103
9. Publications	108
10. Poster Presentations	109

LIST OF FIGURES

Figure 1-1: Folds and attachment modes of membrane proteins	2
Figure 1-2: Reconstitution environments for membrane protein	5
Figure 2-1: Schematics of protein constructs and amino acid sequences used in this study	11
Figure 3-1: Sequence alignment of the TM domain of the 18 α - and 8 β -integrin subunits	21
Figure 3-2: Proposed models of integrin α IIb/ β 3 TM heterodimer interface	22
Figure 3-3: Solution NMR models of α IIb/ β 3 TM domain	24
Figure 3-4: Expression and purification of integrin TM domain peptides	27
Figure 3-5: Secondary structure of reconstituted α L and β 2 TM peptides	29
Figure 3-6: Assigned [^1H - ^{15}N]-TROSY-HSQC of α L- and β 2-TM	30
Figure 3-7: Structural stability of α L and β 2 TM domains	33
Figure 3-8: H/D exchange behaviour of α L and β 2 TM domains in bicelles	34
Figure 3-9: PRE attenuation by 5-DSA and 16-DSA on β 2-TM	35
Figure 3-10: Solution NMR structure of integrin α L and β 2 TM domains	36
Figure 3-11: Structure of the GFFKR motif at the C-terminal juxtamembrane region of the α L TM domain	37
Figure 3-12: Integrin TM cross-titration experiment	39
Figure 3-13: Residues at the α L/ β 2 heterodimer interface	40
Figure 4-1: General features of CoV E proteins	46
Figure 4-2: Two pentameric models of SARS-CoV E protein	48
Figure 4-3: Purification of E protein	52
Figure 4-4: Secondary structure comparison between E_{TR} and E_{FL}	54
Figure 4-5: Oligomeric state of E_{TR} in SDS and PFO detergent	55

Figure 4-6: Sedimentation equilibrium analysis of E _{TR} in C14SB detergent	56
Figure 4-7: Ion channel activity measurement of E _{TR} in lipid bilayers	57
Figure 4-8: Secondary structure comparison of E _{TR} in detergent micelles	60
Figure 4-9: Gel electrophoresis of E _{TR} in detergent micelle	61
Figure 4-10: Sedimentation equilibrium analysis of E _{TR} in DPC/SDS mixed micelles	62
Figure 4-11: Identification of HMA-binding residues in micelle-reconstituted E _{TR}	64
Figure 4-12: Assigned [¹ H- ¹⁵ N]-TROSY-HSQC spectra of E protein	66
Figure 4-13: Structural model of E protein monomer in micelles	68
Figure 4-14: Determination of membrane-embedded region in E _{TR}	70
Figure 4-15: Secondary structure of the E protein extramembrane domain	72
Figure 4-16: Inter-monomeric NOEs between E _{TR} monomers in LMPG	73
Figure 4-17: Structural model of E protein pentamer	74
Figure 5-1: Pentameric model of E protein in the context of mutagenesis data	82
Figure S1: Summary of NOE connectivities	103
Figure S2: Stability of E _{TR} in SDS micelle	104
Figure S3: Screening of CoV E protein constructs	105
Figure S4: Examples of density matching experiment	107

LIST OF TABLES

Table 1-1: Detergents used in solution NMR	7
Table 3-1: Structure statistics for the selected 20 structures of α L-TM	31
Table 3-2: Structure statistics for the selected 20 structures of β 2-TM	32
Table 4-1: Structure statistics for the selected 15 structures of E _{TR} protein in DPC/SDS	67

ABBREVIATIONS

AMPS	2-acrylamido-2-methyl-1-propanesulfonate
AqpZ	Aquaporin Z
BN-PAGE	Blue native PAGE
C14SB	Myristyl (C14) sulfobetaine
CD	Circular dichroism
CoV	Coronavirus
DDM	N-dodecyl- β -D-maltopyranoside
DHPC	1,2-dihexanoyl- <i>sn</i> -glycero-3-phosphocholine
DMPC	1,2-dimyristoyl- <i>sn</i> -glycero-3-phosphocholine
DMPG	1,2-dimyristoyl- <i>sn</i> -glycero-3-phospho-(1'-rac-glycerol)
DPC	n-dodecylphosphocholine
DPhPC	1,2-diphytanoyl- <i>sn</i> -glycero-3-phosphocholine
dMTSSL	(1-acetyl-2,2,5,5-tetramethyl- Δ^3 -pyrroline-3-methyl) methanethiosulfonate
DSA	Doxyl stearic acid
DSS	Sodium 2,2-dimethyl-2-silapentane-5-sulfonate
E	Envelope
ECM	Extracellular matrix
EDC	N-ethyl-N-[3-(diethylamino)propyl] carbodiimide
EM	Electron microscopy
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
FTIR	Fourier-transform infrared
GpA	Glycophorin A
GPCR	G-protein coupled receptors

GUV	Giant unilamellar vesicle
HCV	Hepatitis C virus
HFIP	Hexafluoro-2-propanol
HMA	Hexamethylene amiloride
HPLC	High performance liquid chromatography
hRSV	Human respiratory syncytial virus
IBV	Infectious bronchitis virus
IMC	Inner membrane association clasp
IMP	Integral membrane protein
IPTG	Isopropyl- β -thio-D-glucose
LDAO	Lauryl dimethylamine oxide
LMPG	Lyso-myristoyl phosphatidylglycerol (1-myristoyl-2-hydroxy- <i>sn</i> -glycero-3-phospho-[1'- <i>rac</i> -glycerol])
MBP	Maltose binding protein
MD	Molecular dynamics
MERS	Middle East respiratory syndrome
MHV	Murine hepatitis virus
MTSSL	(1-oxyl-2,2,5,5-tetramethyl- Δ^3 -pyrroline-3-methyl) methanethiosulfonate
NMR	Nuclear magnetic resonance
OG	Octyl glucoside (n-octyl- β -D-glucopyranoside)
OMC	Outer membrane association clasp
PAGE	Polyacrylamide gel electrophoresis
PFO	Perfluorooctanoic acid
PRE	Paramagnetic relaxation enhancement
RP-HPLC	Reverse-phase HPLC

SARS	Severe acute respiratory syndrome
SDS	Sodium dodecyl sulfate
SH	Small hydrophobic
ssNMR	Solid-state NMR
SUV	Small unilamellar vesicles
TB	Terrific broth
TCA	Trichloroacetic acid
TEV	Tobacco etch virus
TFE	2,2,2-Trifluoroethanol
TGEV	Transmissible gastroenteritis virus
TM	Transmembrane
VPU	Viral protein U

ABSTRACT

Reconstitution environment plays an important role in structural determination of membrane proteins. The present study explores this theme in two systems: (1) the TM domain of integrin α L/ β 2, a single-pass integral membrane protein that heterodimerizes in α / β pairs, and (2) CoV E protein, a single-pass integral membrane protein that forms homopentameric ion channels.

Monomeric and oligomeric structural models for integrin α L/ β 2 and E protein were determined by solution NMR. The model of α L/ β 2 TM heterodimer determined herein supports the "backbone reversal" structural feature previously observed only in α IIb/ β 3 TM in bicelles, instead of an α -helix observed in the presence of micelles or organic solvent. Yet, it is possible that the two forms reflect different activation states of integrin.

Monomeric model for CoV E protein revealed conformational flexibility with functional relevance at the C-terminal extramembrane domain. The pentameric model offered possible explanations for the ion channel activity of E protein. In both systems, we found that the alternate forms observed under different reconstitution environment may reflect alternative conformation with functional relevance.

INTRODUCTION

"... the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment."

Christian B. Afinsen, Nobel Lecture on December 11, 1972

1.1 Membrane proteins: what they are

Membrane proteins occupy a niche environment in the cell: the lipid bilayer of the cellular membrane. The first cellular membrane is thought to be formed originally by the self-assembly of phospholipids in aqueous solution into bilayered vesicles, and then it is continually expanded by insertion of new phospholipid molecules (Chen and Walde, 2010). The hydrocarbon tails of phospholipids are hydrophobically excluded from the aqueous environment and face the centre of the bilayer, whereas the polar headgroup is exposed on the surface. The hydrophobic interior of the lipid bilayer causes the peptide backbone to form hydrogen bond arrays only among themselves. Thus, the protein secondary structures found inside the membrane are limited to membrane-spanning α -helical segments or β -barrel structures (Fig. 1-1).

Membrane proteins can be classified into integral and peripheral membrane proteins based on how they attach to the membrane (Fig. 1-1). Integral membrane proteins (IMPs) are permanently incorporated in the membrane and could not be solubilised without lipid, detergent, or other membrane-mimic environments. This makes membrane protein studies particularly difficult, and this will be discussed later in this chapter.

Based on the topology of their extramembrane domains, IMPs can be further classified into monotopic, bitopic, and polytopic. These models are shown in Fig. 1-1. Monotopic IMPs are anchored to one side of the membrane by a short hydrophobic domain (Fig. 1-1A). Bitopic IMPs, also known as single-pass transmembrane proteins, have one membrane-traversing α -helical domain (transmembrane domain)

and two extra-membrane domains: one on each side of the membrane (Fig. 1-1B). Polytopic IMPs have several TM domains connected by extramembrane domains, alternating between both sides of the membrane. TM domains of these proteins can be either α -helical (Fig. 1-1C) or β -strands that loop into a barrel-like form (termed β -barrel, Fig. 1-1D). In contrast, peripheral membrane proteins are soluble proteins that can attach themselves onto membranes directly via lipid or alkyl group anchors (Fig. 1-1E), or indirectly via polar interactions with the phospholipid headgroup or with an IMP. These interactions can be disrupted by extreme changes in pH or ionic strength. In this way, peripheral membrane proteins can be extracted and treated in the same way as soluble proteins. Hereafter, the term membrane protein will be used to refer only to IMPs and not peripheral membrane proteins.

Figure 1-1. Folds and attachment modes of membrane proteins. Cartoon representation of membrane proteins illustrating typical secondary structure elements found inside the membrane, and membrane-attachment modes of membrane proteins. (A) Monotopic, (B) bitopic, (C) polytopic (α -helical), (D) polytopic β -barrel, and (E) alkyl-anchored membrane protein.

1.2 Membrane proteins: what they do

Strategically positioned at the boundary between two compartments, membrane proteins act as a conduit for signal transmission and molecule transport across the

membrane, as well as a membrane anchorage site for macromolecules (Von Heijne, 2007). These simple tasks are crucial parts of a wide range of highly complex biological processes ranging from cell cycle, cell motility, neural transmission, metabolic functions, to viral infection. For this reason, membrane proteins are important drug targets. Modulation of signal transmission by G-protein coupled receptors (GPCRs), for example, is currently a major focus of drug discovery due to its involvement in the immune system, metabolic regulation, and the nervous system (Ferré et al., 2014; Nickols and Conn, 2014).

Membrane proteins involved in virus-host interaction are also attractive targets for antiviral drug discovery. In order to enter the host cell, enveloped viruses bind to specific receptors on host cell surface and subsequently trigger membrane fusion between viral envelope and host cell membrane. Development of HIV entry inhibitors, for example, is aimed at blocking the interaction between HIV gp120 and its host receptors CXCR4 and CCR5, or preventing the membrane fusion process mediated by HIV gp41 (De Clercq, 2002; Root et al., 2001).

One successful example of modulating membrane protein function to prevent viral infection can be found in the inhibition of the M2 proton channel function by adamantane compounds (e.g. amantadine and rimantadine) (Wang et al., 1993). The entry of influenza A into host cells depends on the acidification of the viral interior during its transport in the acidic endosomes (Zhirnov, 1990, 1992). By preventing M2-mediated acidification of the virion, adamantane drugs prevent infection by influenza A virus (Pinto and Lamb, 2007). Examination of the M2 proton channel structure revealed that amantadine binds to and obstructs the channel formed by a tetrameric bundle, thus preventing passage of protons (Cady et al., 2010; Stouffer et al., 2008). Furthermore, analysis on influenza A strains resistant to adamantane revealed some mutations at the transmembrane (TM) domain of M2, consistent with the structural model (Balannik et al., 2009). This finding, among many others, leads to the realization that rational drug design and discovery would benefit from the availability of high-resolution structure of membrane proteins and subsequent structure-activity relationship studies.

1.3 Structural studies of membrane proteins

Determining the high-resolution structure of membrane proteins is evidently very challenging, as less than 3% of deposited PDB structures are of membrane proteins (<http://www.rcsb.org>) despite comprising approximately 30% of the human proteome (Boyd et al., 1998; Krogh et al., 2001; Wallin and Von Heijne, 1998). Since the publication of the first high-resolution structure of *Rhodospseudomonas viridis* photosynthetic reaction centre in 1985 (Deisenhofer et al., 1985), only about 635 unique membrane protein structures have been determined; ~20% of which were solved by NMR and the rest were by X-ray crystallography (<http://blanco.biomol.uci.edu/mpstruc/>, <http://www.drorlist.com/nmr/MPNMR.html>).

The main challenge in structural determination of membrane proteins ultimately lies on their successful extraction and reconstitution into a suitable environment. This environment must, in effect, mimic the properties of the plasma membrane, within which the hydrophobic parts of a membrane protein are embedded. Following this line of thought, the lipid bilayer should be the most ideal reconstitution environment for a membrane protein. However, this system severely limits the range of applicable techniques and also the questions that can be answered.

Membrane proteins can be reconstituted into small unilamellar vesicles (SUV) made of phospholipids, which provide a native-like membrane environment (Fig. 1-2A). SUVs are excellent systems for functional studies of ion channels (e.g. (Heginbotham et al., 1998; Pinto et al., 1992)), but they are less amenable for more detailed structural studies. The high volume-to-surface ratio of SUV means that the protein concentration is low and is thus unsuitable for a low sensitivity technique such as NMR (Henry and Sykes, 1994; Warschawski et al., 2011). More importantly, the large protein-vesicle complex tumbles very slowly in solution, leading to a significant line broadening (Watts and Spooner, 1991).

Membrane protein in stacked lipid bilayers (Fig. 1-1B) can be studied by using solid-state NMR, but the technique remains highly challenging due to its low resolution and sensitivity. These are particularly important in the case of membrane proteins, as their ^{13}C spectra often gets very crowded and unresolvable due to the high abundance of hydrophobic amino acids and highly similar helical structures throughout the protein (McDermott, 2009). To-date, only around 40 unique membrane protein structures

have been studied by using solid-state NMR (<http://drorlist.com/nmr/SPNMR.html>), but the field is still growing.

Figure 1-2. Reconstitution environments for membrane protein. (A) Typical 100-nm diameter small unilamellar lipid vesicle, with the lipid bilayer in expanded view; (B) stacked lipid bilayer; (C) isotropic detergent micelles; (D) high q-ratio (lipid-to-detergent ratio) bicelles, magnetically aligned (the direction of the magnetic field (B_0) is indicated); (E) isotropic, low q-ratio bicelles. (F) Example of lipids and detergents commonly used to prepare bicelles.

The lipid cubic phase, a much less hydrated version of the lipid bilayer, is often used as crystallization medium for membrane proteins (Rummel et al., 1998). It has been used with great success to crystallize GPCRs, among other proteins (reviewed in (Caffrey et al., 2012)); and to date, more than 80 unique proteins have been crystallized in this system (<http://cherezov.usc.edu/structures.htm>). However, this method is technically challenging and problematic, mainly due to the high viscosity of the lipid cubic phase. In addition, inherent to crystallization of any protein, the bottleneck lies in getting a protein crystal to form, which depends heavily on screening a large number of conditions (Moraes et al., 2014).

There are other ways to provide a hydrophobic environment without having to deal with lipids, e.g. organic solvents and detergents. Organic solvents, such as trifluoroethanol (TFE) and hexafluoroisopropanol (HFIP), can completely solubilise membrane proteins in the absence of lipid (Damberg et al., 2001; Henry and Sykes, 1994). However, in these environments, the characteristic demarcation between aqueous and hydrophobic environment found in lipid bilayer is missing. In addition, tertiary and quaternary interactions cannot be preserved in organic solvents. Consequently organic solvent is not an ideal system for the hydrophobic membrane proteins, and especially for those that have several interacting TM domains (Warschawski et al., 2011).

Detergent micelles are more native-like than organic solvents as they provide an amphipathic environment for membrane protein incorporation (Fig. 1-2C) (Linke, 2009). Unfortunately, the bound detergent adds 20-40 kDa to the molecular weight of the complex, thus imposing further size limitation on membrane proteins tenable to NMR measurement. Recent methodological developments such as TROSY-based sequences coupled with protein deuteration and high-field NMR spectrometers (Pervushin et al., 1997) as well as various isotope labelling schemes (Tugarinov et al., 2006) have extended the protein size limitation to beyond 100 kDa. These advances have enabled routine solution NMR studies of membrane protein in detergent micelles (Table 1-1).

Detergent (abbrev.)	Protein	Reference
Sodium dodecylsulfate (SDS)	Human chemokine receptor CCR5 TM4 Human islet amyloid polypeptide Human Na,K-ATPase regulatory protein FXYD2b	(Miyamoto and Togiya, 2011) (Nanga et al., 2011) (Gong et al., 2015)
N-dodecylphosphocholine (DPC)	Severe acute respiratory syndrome coronavirus E protein TM domain Human respiratory syncytial virus small hydrophobic protein SH Human erythropoietin receptor EpoR TM domain	(Parthasarathy et al., 2012) (Gan et al., 2012) (Li et al., 2014a)
1,2-dihexanoyl- <i>sn</i> -glycero-3-phosphocholine (DHPC)	Influenza A M2 <i>Klebsiella pneumoniae</i> outer membrane protein OmpA <i>Natromonas pharaonis</i> sensory rhodopsin pSRII	(Schnell and Chou, 2008) (Renault et al., 2009) (Gautier et al., 2010)
Lauryldimethylamine-N-oxide (LDAO)	Human VDAC1 (voltage-dependent anion-selective channel) Human neuronal acetylcholine receptor (nAChR) $\alpha 4\beta 2$ TM domain	(Bayrhuber et al., 2008) (Bondarenko et al., 2012)
Lyso-myristoylphosphatidylglycerol (LMPG)	Human KCNE1 (potassium channel regulator) <i>Escherichia coli</i> histidine kinase receptors TM domain	(Kang et al., 2008) (Maslennikov et al., 2010)

Table 1-1. Detergents used in solution NMR. Some of the most commonly used detergents in solution NMR of membrane proteins. Examples of protein structures solved in each detergents are listed. List obtained and adapted from <http://www.drorlist.com/nmr/MPNNMR.html>.

Further improvement to the micellar system comes from mixing lipids into detergent micelles result in a hybrid system called bicelles (Fig. 1-2D & E). Bicelles are typically composed of ‘long-chain’ lipids, e.g. DMPC (see abbreviations section); and ‘short-chain’, detergent-like lipids, e.g. DHPC; or detergents such as CHAPSO (Fig. 1-2E) (Sanders II and Prestegard, 1990; Sanders II and Schwonek, 1992). The lipid and detergent components segregate such that the lipids form a planar disc and the detergents form hemi-micelles covering the sides (Glover et al., 2001; Sanders II and Schwonek, 1992; Vold and Prosser, 1996). This segregation confers the planarity of lipid bilayers while still maintaining a relatively small size (Vold et al., 1997).

The amount of lipid in bicelles is usually quantified by the lipid-to-detergent ratio, termed the *q ratio*. Large bicelles with $q > 2.5$ spontaneously align to a magnetic field (Fig. 1-2D), making it very useful for solid-state NMR, where bilayer alignment yields improvements in spectral resolution (De Angelis et al., 2004). Smaller bicelles with $0.1 < q < 1.0$ are relatively fast-tumbling and isotropic in solution (Vold et al., 1997). These small, isotropic bicelles have been used in solution NMR studies, such as for integrin $\alpha\text{IIb}/\beta 3$ TMD (Lau et al., 2009), receptor tyrosine kinase ErbB4 TM domain (Bocharov et al., 2012), and bacterial mercury transporter MerF (Tian et al., 2014).

As mentioned, the inclusion of lipids in bicelles gives it the advantage of having the planarity of lipid bilayers, compared to detergent micelles. Due to the small radius of detergent micelles, they have much higher surface curvature than lipid bilayers, and this curvature can affect the structure of membrane-bound protein segments (Chou et al., 2002; Marcotte and Auger, 2005). For example, by using dipolar coupling measurement, DHPC micelle-bound peptide fragment of HIV gp41 was demonstrated to experience micelle-induced curvature, whereas in bicelles, the peptide was only slightly curved away from the surface (Chou et al., 2002). A smaller effect was observed in the solution NMR solution of beta-barrel protein OmpX in DMPC/DMPG nanodiscs: the β -strands extend on average two residues longer than in DPC and DHPC micelles, but the rest of the structure was largely similar (Hagn et al., 2013).

The presence of lipids in bicelles is also thought to improve the stability of membrane proteins. Membrane proteins transferred from lipid bilayers into detergent micelles have been found to sometimes become unstable, especially those forming complexes

(Bowie, 2001; Seddon et al., 2004). Upon delipidation, the protein complex may fall apart and cease to function, as was observed in cytochrome b6f (Breyton et al., 1997) and diacylglycerol kinase (Zhou et al., 2001b). In contrast, CHIF (FXVD4, Na,K-ATPase regulatory protein) was found to be structurally similar in micelle (solution NMR) and in lipid bilayer (ssNMR) (Franzin et al., 2007).

These examples mean that eventually the choice of membrane mimic environment needs to be evaluated in a case-by-case basis, keeping in mind the functional relevance of the structure obtained.

1.4 Aim of this study

As illustrated by examples in the previous section, the choice of membrane mimic environment may or may not influence the structure of a membrane protein. Some structural features may have a strict requirement for a certain, specific environment, which is possibly related to the function of that particular structural element.

In this study I will examine this issue by using two different single-pass membrane proteins. The first is human integrin α L/ β 2 TM domain: a pair of heterodimeric transmembrane domains that mediates signal transmission across the membrane via its association and dissociation. The second is a coronavirus envelope (E) protein: a homopentameric membrane protein that conducts ions through a pore formed by a pentameric bundle of α -helices.

The structures of the juxtamembrane region of integrin, and the extramembrane region of E proteins are subjects of controversy, which I will elaborate in detail in chapters 3 and 4, respectively. The main issue leading to the discrepancies seems to be the reconstitution environment used for structure determination. The present study attempts to clarify the problem by using solution NMR to solve the structure of the forementioned two membrane proteins.

MATERIALS AND METHODS

2.1 Sample preparation

Integrin TM domain constructs. The DNA sequence encoding human α L integrin TM domain residues 1057-1103 was cloned into a pET-28b expression vector (α L-TM, Fig. 2-1A). Human β 2 integrin TM domain constructs were screened at the NTU-SBS Protein Production Platform: six constructs were prepared from a combination of 3 starting N-terminal residues (675, 679, and 683), and 2 ending C-terminal residues (718 and 721). These constructs were cloned into pNIC28-Bsa4 expression vector. The best expressing construct consisting of residues 683-718 were used in the present work (β 2-TM, Fig. 2-1B). Constructs carrying G685F and G689F point mutations (in β 2-TM) were separately prepared by site-directed mutagenesis using the relevant pair of primers.

SARS-CoV E constructs. The DNA sequence encoding the full-length SARS-CoV E protein (E_{FL}) was cloned into pTBMale expression vector. The resulting construct carries an N-terminal 6-His tag and a MBP fusion tag cleavable by tobacco etch virus (TEV) protease (Fig. 2-1). The DNA sequence of SARS-CoV E residues 8-65 (E_{TR}) was cloned into pNIC28-Bsa4 expression vector (E_{TR} , Fig. 2-1).

In both E_{FL} and E_{TR} , all three native cysteines (C40, C43, and C44) were mutated into alanines. In addition, two E_{FL} mutants previously described by Cohen *et al.* (Cohen *et al.*, 2011) were prepared by site-directed mutagenesis: P54A (E_{P54A}) and V56A-Y57A-V58A-Y59A (E_{4ALA}). For Paramagnetic Relaxation Enhancement (PRE) experiments, a S60C point mutation was introduced into the E_{TR} construct by site-directed mutagenesis using appropriate sets of primers ($E_{TR-S60C}$).

All constructs carry an N-terminal 6-His tag, followed by a TEV protease cleavage site to facilitate removal of the tag.

Figure 2-1. Schematics of protein constructs and amino acid sequences used in this study. (A) Human integrin αL-TM, (B) β2-TM (C) SARS-CoV E full-length, and (D) truncated. White bars represent parts that are eventually removed by TEV protease digestion, whereas yellow bars represent parts that are used in experiments. Red letters indicate mutations introduced into the protein .

Protein expression. Plasmids prepared by the Protein Production Platform were transformed into *E. coli* strain BL21(DE3) T1R Rosetta cells. Other plasmids prepared in-house were transformed into *E. coli* strain BL21(DE3) codon-plus cells. For expression of non-labelled proteins, the *E. coli* culture was grown in terrific broth (TB) media at 37°C until an OD₆₀₀ of 2. Protein expression was induced overnight at 18°C by addition of 0.5 mM IPTG (for E proteins), or at 25°C with 1 mM IPTG (for integrin TMs). The cells were harvested by centrifugation at 7500 × g and stored at -80 °C until used.

Stable isotope labelling. Expression of stable-isotope labelled protein was achieved in M9 minimal media supplemented with an appropriate combination of [¹⁵N] NH₄Cl, [¹³C] or [¹³CD] glucose, and D₂O according to the desired labelling scheme. The culture was initially grown in LB media until an OD₆₀₀ of 0.7 was reached. The cells were collected by centrifugation, and resuspended in 1/4 volume of M9 minimal media to achieve a high density culture (Sivashanmugam et al., 2009; Studier, 2005). The cultures were further grown for 1 hour before protein expression was induced overnight at 25°C with 1 mM IPTG for integrin TMDs, or 6 hours at 18°C with 0.5 mM IPTG for E protein constructs. The cells were harvested by centrifugation at 7500 × g and stored at -80 °C until used.

Protein purification. Frozen cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 300 mM NaCl, 5 mM imidazole, 2 mM β-mercaptoethanol and 10% glycerol) containing 1 mM PMSF and 1.5% Triton X-100. Complete lysis was achieved by sonication and microfluidization. The crude cell lysate was clarified by centrifugation at 40,000 × g and applied onto pre-equilibrated Ni-NTA resin (Bio-Rad Profinity IMAC Ni²⁺-charged). The resin was washed with 20 mM Tris pH 8.0, 300 mM NaCl, 25 mM imidazole, 2 mM β-mercaptoethanol and 10% glycerol.

The integrin TMs were cleaved on-column by TEV protease after the resin was further equilibrated with 20 mM Tris pH 8.0, 300 mM NaCl, 25 mM imidazole, 2 mM β-mercaptoethanol, 10% glycerol and 1.8 mM N-dodecyl-β-D-maltopyranoside (DDM). The cleaved peptides were collected, TCA-precipitated, lyophilized, and extracted by using methanol. The identity and purity of the peptides were confirmed by SDS-PAGE and MALDI-TOF MS. When higher purity is desired, the methanol extract was further purified by using reversed-phase HPLC. Further purification was

achieved by applying isopropanol-acetonitrile (4:1 (v/v) with 0.1% TFA) linear gradient on a Phenomenex Jupiter C18 semipreparative column (250 × 10 mm, 300 Å pore size, 5 µm particle size). The identity and purity of peptide fractions were confirmed by SDS-PAGE and MALDI-TOF MS. The peptide fractions were lyophilized in the presence of 1 mM HCl and stored at -20°C.

The SARS-CoV E proteins were eluted in 20 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol, 10% glycerol and 5 mM myristyl sulfobetaine (C14SB). The E_{TR} was directly TCA-precipitated and lyophilized, while the E_{FL} was cleaved off-column by TEV protease prior to precipitation and lyophilization. Further purification was achieved by using reversed-phase HPLC on a Phenomenex Jupiter C4 semipreparative column (250 × 10 mm, 300 Å pore size, 5 µm particle size). The lyophilized peptides were dissolved with 1% TFA in acetonitrile and separated under isopropanol-acetonitrile linear gradient (4/1 (v/v) with 0.1% TFA). The identity and purity of peptide fractions were confirmed by SDS-PAGE and MALDI-TOF MS. The peptide fractions were lyophilized in the presence of 1 mM HCl and stored at -20°C until used.

2.2 Secondary structure determination

Fourier-Transform Infrared spectroscopy. Sample preparation, data collection and H/D exchange were performed essentially as described (Torres et al., 2006) on a Nicolet Nexus 470 FT-IR spectrometer (Madison, USA) equipped with an MCT/A detector cooled in liquid N₂. The peptides were incorporated into multilamellar vesicles (MLV) by mixing dimyristoyl phosphocholine (DMPC, Avanti Polar Lipids) and lyophilized peptide in hexafluoroisopropanol (HFIP) at 50:1 molar ratio. The mixture was vacuum-dried overnight and resuspended in water by repeated freeze-thawing and vortexing cycle. The MLV mixture was subsequently deposited onto a 25-reflection germanium plate, which was mounted onto a vertical ATR accessory (Graseby Specac). Typically 200-replicate spectra were collected after 1 hour purging with dry, CO₂-free air. For H/D exchange experiment, the purge line was saturated with D₂O vapour. Fourier self-deconvolution was performed using the following parameters: FWHH: 20 cm⁻¹ and narrowing factor, k = 1.5) (Kauppinen et al., 1981).

Circular Dichroism. CD data was acquired on a Chirascan CD Spectrometer (Applied Photophysics) using a 0.2 mm quartz cuvette (Hellma). The spectra were acquired from 180 to 260 nm with 1 nm spectral bandwidth and 3 replicates per spectra. Data points with excessive absorbance values were excluded from analysis. After baseline subtraction, the data were analyzed in Dichroweb (Whitmore and Wallace, 2004) by using CDSSTR algorithm (Compton and Johnson, 1986) and SMP180 reference set (Abdul-Gader et al., 2011).

The E_{TR} peptide samples were dissolved at 1 mg/ml in 20 mM sodium phosphate pH 5.5, 50 mM NaCl, and 100 mM DPC, with or without 50 mM SDS. The integrin α L-TM and β 2-TM peptides were reconstituted at 1 mg/ml in 50 mM sodium phosphate pH 6.5, 3% w/v DHPC-DMPC bicelles with q ratio 0.3 (refer to NMR sample preparation for details).

2.3 Oligomeric state determination

Gel electrophoresis. Standard SDS-PAGE was performed in 13.5% Tris-glycine gel with TGS running buffer and stained with Coomassie Blue G-250. SDS-NuPAGE was performed in 4-12% NuPAGE® Bis-Tris gel (Invitrogen) with NuPAGE® MES SDS running buffer and stained with SimplyBlue™ SafeStain (Invitrogen) according to the manufacturer's protocol.

To perform electrophoresis in the presence of perfluorooctanoic acid (PFO) detergent, we modified Invitrogen's SDS-NuPAGE protocol by replacing SDS with PFO. Lyophilized peptides were dissolved in sample buffer containing 4% PFO and heated at 65°C for 5 min prior to loading. The gel was run at 80 V for 2-3 hours with MES running buffer containing 0.5% PFO.

Blue-native PAGE (BNPAGE) was performed as described previously (Gan et al., 2011). Lyophilized peptide was solubilized at 0.1 mM in a sample buffer containing 25 mM SDS and either 25, 50, or 100 mM dodecyl phosphocholine (DPC). *E. coli* aquaporin Z (AqpZ, 24 kDa) in 20 mM SDS (heated at 65°C for 10 min) was included as an additional molecular weight marker.

Sedimentation equilibrium. In a sedimentation equilibrium experiment, the sample is spun in an analytical ultracentrifuge at a speed where sample diffusion can still take place, until equilibrium point between diffusion and sedimentation has been reached. The resulting radial distribution profile was then fitted to different oligomer self-equilibrium models and evaluated mainly based on reduced chi-square of the fit and the fitting residuals.

The sedimentation equilibrium experiments were performed using a Beckman XL-I analytical ultracentrifuge at 20°C and the sample absorbance was monitored at 280 nm.

Lyophilized E_{TR} peptides were dissolved at OD₂₈₀ of 0.3, 0.5, and 0.8 (12 mm pathlength) in 20 mM sodium phosphate pH 5.5, 50 mM NaCl, and (1) 5 mM C14SB, or (2) 100 mM DPC and 25 mM SDS (4:1 DPC-to-SDS molar ratio). To match the density of the DPC/SDS mixture, 65% D₂O was included (see Appendix for more details).

The samples were centrifuged in six-channel charcoal-filled Epon centerpieces with quartz windows. The radial distribution profile of the absorbance was acquired after equilibrium has been reached, as tested by using HeteroAnalysis software. The data were processed and fitted to several monomer-N-mer models by using non-linear least-squares fitting method in SEDFIT and SEDPHAT software (Schuck, 2003). The model with the lowest global-reduced chi-square value and with randomly distributed fitting residuals was taken as the best model.

2.4 Single channel activity measurement

Ion channel activity of E_{TR} was measured by using Nanion Port-a-Patch®. Briefly, giant unilamellar vesicles (GUV) of 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) containing 10% cholesterol was prepared in 1 M sorbitol by using Nanion Vesicle Prep Pro® as per the manufacturer's protocol. Into a 100 µl GUV solution, 0.5 µl of E_{TR} peptide solution in ethanol (0.1 mg/ml) was added and incubated for 1 hour at room temperature. The E_{TR}-containing GUVs were subsequently deposited onto 6-8 MΩ NPC©-1 chips (Nanion). The conductance was measured under symmetrical buffer 10 mM HEPES pH 4, 500 mM NaCl.

2.5. NMR experiments and data analysis

NMR sample preparation. Integrin TMs were incorporated into bicelles by mixing lyophilized peptides and DMPC in methanol, followed by removing the methanol under N₂ stream and further drying under vacuum overnight. The dried peptide-lipid was solubilised in 50 mM potassium phosphate pH 6.5 containing DHPC. The mixture was vortexed and sonicated until it was visibly clear. Typically, the DHPC-to-DMPC ratio (q ratio) used was 0.3, and the total DHPC-DMPC content was 3% (w/v). At 0.6 mM of peptide, this corresponds to approximately 1:100 peptide-to-lipid molar ratio.

Methanol-dried E protein was reconstituted by vortex and sonication in 20 mM sodium phosphate pH 5.5 and 50 mM NaCl containing (1) 50 mM SDS or (2) 200 mM DPC, 50 mM SDS, or (3) 200 mM LMPG. For determination of membrane-embedded region, reconstituted ¹⁵N-labeled E_{TR} was added to methanol-dried 5-doxyl stearic acid (5-DSA) or 16-DSA. For cysteine-based paramagnetic relaxation enhancement (PRE) experiments, reconstituted ¹⁵N-labeled E_{TR}-S60C was treated with 0.8 mM DTT and aliquoted for labeling with (1-oxyl-2,2,5,5-tetramethyl-Δ³-pyrroline-3-methyl) methanethiosulfonate (MTSSL) (Toronto Research Chemicals Inc.) and its diamagnetic analogue: (1-acetyl-2,2,5,5-tetramethyl-Δ³-pyrroline-3-methyl) methanethiosulfonate (dMTSSL, Toronto Research Chemicals Inc.). Briefly, the sample was mixed with 10-fold molar excess of the reagent, vortexed for 30 minutes, and incubated overnight at room temperature. The excess reagents were removed in 4 rounds of washing by centrifugal filtration.

Partial alignment of the protein/micelle and protein/bicelle complexes relative to magnetic field was obtained by using stretched polyacrylamide hydrogels (Tycko et al., 2000; Ulmer et al., 2003). Briefly, the peptide-bicelle solution was co-polymerized with 4.2% or 4.6% (w/v) solution of acrylamide/bis-acrylamide (49:1) and 6% (molar ratio) 2-acrylamido-2-methyl-1-propanesulfonate (AMPS). The peptide-micelle complex was soaked into dried, pre-equilibrated 7% polyacrylamide gel prepared with similar acrylamide ratios as mentioned above. The hydrated gels were then radially compressed into 4.2 mm tube.

NMR experiments. NMR data were collected at 303K in Bruker Avance-III 800 (in collaboration with Dr. Ocsar Millet and Tammo Diercks from CIC BioGUNE, Spain)

and Avance-II 700 (in NTU-SBS) NMR spectrometer with cryogenic probes. Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as the internal reference for ^1H nuclei. The chemical shifts of ^{13}C and ^{15}N nuclei were calculated from the ^1H chemical shifts. The NMR data were processed using TopSpin 3.1 (www.bruker-biospin.com) and analyzed using CARA (www.nmr.ch).

To identify membrane-embedded residues, the samples were lyophilized overnight and reconstituted in 99% D_2O . Immediately after reconstitution, a 2D [^1H - ^{15}N]-TROSY-HSQC was collected. For HNOE measurements, spectra measuring ^1H - ^{15}N heteronuclear steady-state NOE were acquired with and without a 3s period of proton saturation.

Integrin TMs. Sequence-specific assignment of backbone $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}^{\alpha}$ of $\alpha\text{L-TM}$ was achieved by using 2D [^1H - ^{15}N]-TROSY-HSQC, 3D TROSY-optimized HNCO, HNCA, HN(CO)CA, HN(CA)CB, and HN(COCA)CB experiments on a fully-deuterated $^{15}\text{N}/^{13}\text{C}$ -labeled $\alpha\text{L-TM}$. Side-chain resonances were assigned using 3D ^{15}N -resolved NOESY-HSQC (120 ms mixing time), (H)CCH-TOCSY and ^{13}C -resolved NOESY-HSQC (120 ms mixing time) on non-deuterated and partially-deuterated $^{15}\text{N}/^{13}\text{C}$ -labeled $\alpha\text{L-TM}$ sample in bicelles.

Sequence-specific assignment of backbone $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}^{\alpha}$ of $\beta 2\text{-TM}$ was achieved by using 2D [^1H - ^{15}N]-TROSY-HSQC, 3D TROSY-optimized HNCO, HNCA and HN(CO)CA experiments on a fully-deuterated $^{15}\text{N}/^{13}\text{C}$ -labeled $\beta 2\text{-TM}$ sample in bicelles. Side-chain resonances were assigned using 3D ^{15}N -resolved NOESY-HSQC (120 ms mixing time), (H)CCH-TOCSY and ^{13}C -resolved NOESY-HSQC (120 ms mixing time) on non-deuterated and partially-deuterated $^{15}\text{N}/^{13}\text{C}$ -labeled $\beta 2\text{-TM}$ sample in bicelles.

Integrin α/β cross-titration experiments were performed by adding increasing amounts of lyophilized, non-labelled $\beta 2\text{-TM}$ to $50\text{ }\mu\text{M}$ ^{15}N -labeled $\alpha\text{L-TM}$ in bicelles, and *vice versa*, from 1:0.5 to 1:2 labelled/non-labelled molar ratio. At each ratio, a [^1H - ^{15}N]-TROSY-HSQC was collected and examined for chemical shift perturbations.

SARS-CoV E proteins. Sequence-specific assignment of backbone $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}^{\alpha}$ of E_{TR} was achieved by using 2D [^1H - ^{15}N]-TROSY-HSQC, 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, and HNCACB experiments on a $^{15}\text{N}/^{13}\text{C}$ -labeled

E_{TR} protein in micelles. Side-chain resonances were assigned using 3D ¹⁵N-resolved NOESY-HSQC (80, 100 and 150 ms mixing time), (H)CCH-TOCSY and ¹³C-resolved NOESY-HSQC (120 ms mixing time). Membrane-embedded region was determined by recording ¹⁵N-HSQC spectra in the presence and absence of 1 mM 5-doxyl stearic acid (5-DSA) or 16-DSA.

Long-distance restraints were obtained from the measured PRE effect using the previously described procedures (Battiste and Wagner, 2000; Liang et al., 2006; Van Horn et al., 2010). The PRE effect was measured using ¹⁵N-HSQC spectra of E_{TR}-S60C mutant before spin labelling and after MTSSL and dMTSSL labelling. The intensities of cross-peaks in the MTSSL (I_p) and dMTSSL (I_d) were calculated in CARA program. The correlation time was set to 10 ns. The ratios of intensities (I_p/I_d) were normalized against a set of 8 highest I_p/I_d ratios, which were assumed to belong to peaks unaffected by PRE. For peaks with the ratios below 0.15, no lower distance restraints were used, whereas upper restraints were set to 15 Å. For peaks with the ratios above 0.9, only the upper restraints of 25 Å were utilized. For peaks with the ratios between 0.15 and 0.9, the upper and lower distance restraints were generated using ± 3 Å margins.

Inter-monomeric restraints were obtained from ¹⁵N-resolved NOESY-HSQC (250 ms mixing time) of 0.6 mM asymmetrically deuterated E_{TR} samples reconstituted in 200 mM LMPG. Briefly, a 1:1 mixture of ¹⁵N/D- and ¹³C/H-labelled E_{TR} (ND/C) results in the observation of strictly inter-monomeric NOEs due to ¹H and ¹⁵N being present in separate peptide chains. Identical spectra were collected from 0.6 mM ¹⁵N/D-labelled E_{TR} sample alone to control for any intra-monomeric NOEs observed due to incomplete deuteration.

NMR data analysis and structure calculation. Sequential backbone and sidechain assignments were performed in CARA (<http://cara.nmr.ch/>, (Keller, 2004)). Backbone dihedral angle restraints (ϕ and ψ) were derived from ¹³C', ¹³C ^{α} , ¹³C ^{β} , ¹H ^{α} and ¹H ^{β} chemical shift values (where available) using TALOS (Cornilescu et al., 1999). The short-range and medium range NOE connectivities were used to establish the sequence-specific ¹H NMR assignment and to identify elements of the regular secondary structure. Hydrogen bonds were derived from the H/D exchange experiment and NOE connectivity. Structure calculations were performed using

CYANA 3.0 (Guntert et al., 1997; Herrmann et al., 2002) and visualized using MOLMOL (Koradi et al., 1996) and PyMOL (Delano Scientific). CNS 1.3 (Brunger, 2007; Brunger et al., 1998) was used to refine the monomeric structure using standard simulated annealing protocol.

The oligomeric models were built by docking the monomeric structure using HADDOCK software (De Vries et al., 2007; Dominguez et al., 2003). Inter-monomeric restraints were used as both ambiguous and unambiguous distance restraints for the docking. A standard docking procedure was employed: the first stage consists of rigid-body energy minimization generating 1000 structures, and the second stage consists of semi-flexible simulated annealing of the top 200 structures from stage 1. The final 200 structures were clustered by RMSD and ranked by internal HADDOCK scoring scheme (consists of weighted sum of various energy terms).

INTEGRIN JUXTAMEMBRANE STRUCTURE AND HETERODIMER FORMATION

3.1 Introduction

In this chapter I will present a structural study on integrin α_L/β_2 TM domain heterodimer. Primarily, this study aims at determining its juxtamembrane structure, and in addition whether a polar interaction assists its heterodimerization.

The integrin protein family are type I bitopic TM receptors which heterodimerize into pairs of α - and β -subunits. In humans, there are 18 α - and 8 β -integrin subunits that assort into 24 α/β heterodimer pairs (Hemler, 1999).

As bitopic membrane proteins, integrins possess a single α -helical TM domain that connects a large extracellular matrix (ECM) binding extracellular domain and a short C-terminal cytoplasmic tail (Carman and Springer, 2003). This allows integrin to function as a bidirectional signal transducer: classical outside-in signalling as a result of ECM binding, and inside-out signalling triggered by separation of the cytoplasmic tails, the TM domain, and the entire α/β heterodimer, leading to modulation of its ECM-binding affinity (Hantgan et al., 1999; Kim et al., 2003; Travis et al., 2003).

Destabilization of the α/β heterodimer by deletion or mutations was predicted and has been found to activate integrin (Lu and Springer, 1997; Lu et al., 2001; O'Toole et al., 1994). Conversely, heterodimer stabilization, e.g. by introducing disulfide bonds, inactivates integrin. By using this strategy, the α/β interfacial residues have been elucidated through a disulfide scanning experiment on the α_{IIb}/β_3 TM domain (Luo et al., 2004).

3.1.1 Structural models for integrin heterodimer: α_{IIb}/β_3 TM

Two conserved motifs important for integrin α/β heterodimerization can be found in their TM domain: the small-xxx-small amino acid motif, and the GFFKR motif (Fig. 3-1). The small-xxx-small amino acid motif (GxxxG-like motif) was first described in the dimerization of glycophorin A TM domain, and has since been found as

dimerization motif in TM domains of other proteins (Arkin and Brunger, 1998; Langosch et al., 1996; Lemmon et al., 1994; MacKenzie et al., 1997; Russ and Engelman, 2000; Senes et al., 2000). In integrins, this motif is found in both α and β TM domains, and is thought to mediate their heterodimerization. The GFFKR motif, on the other hand, is found only in α TM domain. Mutations and deletions in this highly conserved motif were found to activate integrin α IIB/ β 3 via the separation of their TM domains (Kim et al., 2003; Luo et al., 2004; O'Toole et al., 1994). These motifs were later found to be structurally important in α/β TM heterodimerization (Schneider and Engelman, 2004).

α 1	LAIQISKDGLPG---RVPLWVILL SAFAG LLLLMLLILALWKI GFFKR PLK-----KK
α 2	PLMIMKPDEK----AEVPTGVIIG SIIAG LLLLLALVAILWKL GFFKR KYEKMTKNPD
α 3	FSVDIDSELVEELPAEIELWLVLVAV GAG LLLLGLIILLWKC GFFKR TRY-YQIMPK
α 4	AHVLLEGLHHQRPKRYFTIVIISS SLLLGL IVLLISYVMWK GFFKR QYK-SILQEE
α 5	VATAVQWTKAEGSY-GVPLWIIIL AILFGL LLLLGLLIYILYKL GFFKR SL-----PY
α 6	VRVTVPFSKTVAQYSGVPWWIIL AILAG ILMLALLVFILWKC GFFKR SRYS-DDSVPR
α 7	IPVMVYLDPMVAEGVPWWVILL AVLAG LLVLALLVLLLWKM GFFKR AKHPEATVPQ
α 8	IKTSVIWATPNVSF-SIPLWVIL AILLGL LVLAITLALWKC GFFD RAR-----PP
α 9	VTVVFEALHNLEPRGYVVGWIIAI ISLLVG ILIFLLLAVLLWKM GFFR RRYK-EIIEAE
α 10	WSESLLEVQTRP-ILISLWILIG SVLGG LLLLALLVFCLWKL GFFA HKKI-----PE
α 11	IVFEISKQED----WQVPIWIIIV STLGG LLLLALLVLALWKL GFFR SARR--RREPG
α V	VTTNVTWGIQPAPM-PVPVWVIL AVLAG LLLLAVLVFVMYRM GFFKR VR-----PP
α D	MRAQMEMVLE-EDEVYNAIPIIM GSSVG ALLLLALITATLYKL GFFKR HYK-EMLEDK
α E	TKITVVFLKDEK---YHSLPIIIK SGVGG LLVLIVILVILFKC GFFKR KYQQNLNLESI
α M	VRSQTETKVE-PFEVNPPLPLIV GSSVG LLLLALITAALYKL GFFKR QYK-DMMSE-
α X	MRAQTTTVLE-KYKVHNPTPLIV GSSIG LLLLALITAVLYKV GFFKR QYK-EMMEEA
α IIB	VWTQLLRALKEER---AIPWVVLV GVLGG LLLLTILVLAMWKV GFFKR NR-----PP
α L	AQVVMKVDVVYEK---QMLYLYVL SGIGG LLLLLLIFIVLYKV GFFKR NLKEKMEAGR
β 2	YVDESRECVAGPNIAAIV GTVAG IVLIGILLVVIWKALIHLSDLREYRRF
β 3	VVEE-PECPKGPDIILVLL SVMGA ILLIGLAALLIWKLLITIHDRKEFAKF
β 1	VV-ENPECPTGPDIIPIV AGV VAGIVLIGLALLIWKLLMIHDRREFAKF
β 4	LVHKKKDCPPG-SFWWLIPLLLLLLPLALLLLLCWKYCACCKACLALLPC
β 5	VLRE-PEGNTPNAMTILL AVVGS ILLVGLALLAIWKLLVTIHDRREFAKF
β 6	SINE-KDCPKPPNIPMIM LGVS LAILLIGVVLLCIWKLLVSFHDRKEVAKF
β 7	VRPQE---KGADHTQAIVL GCVGG IVAVGLGLVLAYRLSVEIYDRREYSRF
β 8	YVDQTSECFSSPSYLRIFFIIFIVTFLIGLLKVLIIRQVILQWNSNLIKSS

Figure 3-1. Sequence alignment of the TM domain of the 18 α - and 8 β -integrin subunits. The GxxxG-like motif and GFFKR motif are indicated in bold. Grey box mark the approximate TM domain. Illustration adapted with modifications (Lau et al., 2009).

The first structural models for α/β TM interaction in its resting-state as well as in its activated-state were initially obtained from computational (Gottschalk et al., 2002)

and electron microscopy (EM) (Hantgan et al., 1999) studies. The two models are herein illustrated in Fig. 3-2A. In the first model (model I in Fig. 3-2A), α/β integrin chains interact through their GxxxG-like motifs. These motifs face each other and form the heterodimer interface (Fig. 3-2A, green and red spheres).

In the second model (model II in Fig. 3-2A), only the α -chain GxxxG-like motif participates in the heterodimer interface. The GxxxG-like motif on β -integrin side is not participating because its TM is rotated approximately 90° relative to its orientation in the first model (Adair and Yeager, 2002; Gottschalk et al., 2002). The second model has currently been accepted as the model for resting-state $\alpha\text{IIb}/\beta 3$. This is supported by data from disulfide scanning of the TM domain in full-length integrin (Fig 3-2B) (Luo et al., 2004; Zhu et al., 2009), mutations on the TM domain that resulted in integrin activation (Partridge et al., 2005), and solution NMR structure of the $\alpha\text{IIb}/\beta 3$ TM domain in bicelles (Lau et al., 2009).

Figure 3-2. Proposed models of integrin $\alpha\text{IIb}/\beta 3$ TM heterodimer interface. (A) The two models initially proposed for the $\alpha\text{IIb}/\beta 3$ heterodimer. Green and red spheres indicate the position of the residues in the GxxxG-like motif. (B) Relative orientation of the two TMs in the currently accepted model, based on a cysteine scanning experiment (Luo et al., 2004). Solid and dotted lines indicate constitutive and inducible disulfide formation, respectively. The first residue in each GxxxG-like motif is indicated by the arrowheads. Illustrations modified from (Luo et al., 2004) and (Lin et al., 2006).

The solution NMR structure of the α IIb/ β 3 TM domain (Fig. 3-3) revealed the existence of two interaction elements in the α/β TM heterodimer interaction (Lau et al., 2008; Lau et al., 2009; Zhu et al., 2009). The first, starting from the N-terminus, is a GxxxG-like interaction, centred on G972xxxG976 of α IIb and G708 of β 3. This forms a so-called outer-membrane association clasp (OMC because it is located at the extracellular half of the membrane. Afterwards, there is a short stretch of separation between the TMs, and further to the C-terminus the two TMs rejoin to form the second interaction element: the inner-membrane association clasp (IMC). In this clasp, F992-F993 of α IIb packs against the β 3 subunit, followed by a salt bridge between R995 of α IIb and D723 of β 3. This packing is made possible by the turning back of the peptide backbone (backbone reversal) and the reinsertion of the GFFKR motif of α IIb (G991-R995) into the membrane (Lau et al., 2009).

3.1.2 The controversial GFFKR backbone reversal motif

The conservation of the GFFKR motif among integrin α chains (Fig. 3-1) suggests that the backbone reversal, if truly present, is a conserved structural motif. However, at the beginning of this study, this structure had only been observed for α IIb in bicelles (Lau et al., 2008; Lau et al., 2009), whereas in organic solvent, the GFFKR motif formed an extended α -helix instead (Yang et al., 2009). This discrepancy could simply be due to the reconstitution environment used, which may lead to artefacts. Each reconstitution environment may be biased towards one functional form or the other, and thus the two forms of the GFFKR motif may reflect the different activation states of integrin: the backbone reversal in the resting state (α/β heterodimer) and the extended α -helix in the activated state (dissociated).

This possibility was considered because following their activation, integrins are thought to form homomultimer clusters (Gottschalk et al., 2002; Li et al., 2001). Different membrane mimics have been demonstrated to modulate the homomeric versus heteromeric balance between α and β -subunits (Kim et al., 2004; Luo et al., 2005; Wang et al., 2011). Specifically, a heteromeric interaction is favoured in the presence of lipid, whereas homodimers are more dominant in micelles and in organic solvents (Suk et al., 2012). As both the extended α -helix form and the homomultimeric state, i.e. the post-activation form of integrin α IIb, were observed in

organic solvents, it is possible that the activated (dissociated) form of integrin α adopts the extended α -helix structure at its juxtamembrane region.

Figure 3-3. Solution NMR models of α IIb/ β 3 TM domain. Cartoon representation of α IIb/ β 3 TM domain (PDB ID 2K9J) (Lau et al., 2009). Grey box indicate the approximate extent of the plasma membrane. Residues forming the OMC are shown in spheres, with the GxxxG-like motif highlighted in red. Residues forming the IMC are shown as sticks, with the salt bridge highlighted in blue.

In addition, the extended α -helical form could be relevant in integrin α binding to its partners. Calcium-and-integrin-binding protein 1 (CIB1) has been suggested to regulate the activation of α IIb/ β 3 by binding to α IIb in competition with talin, an

activator of integrin (Yuan et al., 2006). The C-terminal tail of α IIb (L983-Q1008) bound to CIB1 was found to adopt an α -helical form (Huang and Vogel, 2012).

Finally, the backbone reversal has also not been observed in the absence of the TM domain. The GFFKR motif in α IIb, α M, α L, α X or α 4, has been shown to form an α -helix instead (Bhunia et al., 2009; Chua et al., 2013; Chua et al., 2011; Chua et al., 2012; Vinogradova et al., 2000; Vinogradova et al., 2002). This may partially be explained by the absence of the preceding TM domain in the construct used, or it might also suggest some sort of structural flexibility between the backbone reversal form and the extended α -helical form.

3.1.3 The integrin α L/ β 2 TM

Since the GFFKR motif is conserved among integrin α subunits, so its structure is likely to be conserved as well. Thus, by characterizing the structure of other α/β integrin pairs, we can find support as to whether the GFFKR backbone reversal truly exists, and whether it is structurally conserved. In this study, we chose to work with the integrin α L/ β 2 pair, also known as LFA-1 (lymphocyte function-associated antigen), which plays an important role in leukocyte adhesion and migration during extravasation (Kinashi, 2007). Chemokine-based signals induce a cascade leading to an inside-out signaling which increases α L/ β 2 affinity towards its ligand ICAM-1 (intracellular adhesion molecule) present on vascular endothelium, and thus maintains lymphocyte adhesion (Butcher et al., 1999; Katagiri et al., 2003). At the lymphocyte trailing edge, α L/ β 2 eventually becomes deactivated and recycled through an endosomal transport mechanism to the surface of the leading edge, where it will become activated again (Caswell and Norman, 2006). This cycle of integrin-mediated adhesion-deadhesion helps the lymphocyte to migrate towards the chemokine-based signal.

The GFFKR motif appears to play an important role in the leukocyte adhesion and migration, as its deletion decreases recycled α L/ β 2 level (Lu and Springer, 1997) and resulted in a constitutive α L/ β 2-mediated cell adhesion in a transgenic mouse model (Semmrich et al., 2005). The heterodimer model based on solution NMR structure of α IIb/ β 3 TM domain is consistent with these findings, as the GFFKR backbone reversal motif forms the IMC that stabilizes the heterodimer. However, whether this structural motif is also present in α L/ β 2 is still unknown.

Residues at the interacting face of α L/ β 2 are thought to be homologous to that determined for α IIb/ β 3. In addition, we have proposed that a polar interaction exists between S1071 of α L and T686 of β 2 (Chng and Tan, 2011; Vararattanavech et al., 2009; Vararattanavech et al., 2010). These two residues are located at the GxxxG-like motif of the two TM domains, posing it as an additional part of the OMC interaction elements. In the low dielectric environment of the membrane, polar interactions are much stronger (Curran and Engelman, 2003; Zhou et al., 2001a). Therefore this polar interaction may further facilitate the heterodimer re-formation and α L/ β 2 deactivation. This is consistent with the role of integrin α L/ β 2 in lymphocytes, where adhesion must be reversible for migration to occur.

3.1.4 Aim and approach

The rest of this chapter describes our efforts and results in solving the structure of the α L/ β 2 heterodimer by using solution NMR in bicelles. Bicelles were chosen as they more closely mimic a membrane-like environment, and to promote heterodimer formation. Specifically, we would like to find out whether the α L/ β 2 GFFKR motif also undergoes backbone reversal, and whether a polar interaction at the GxxxG-like motif of α L/ β 2 exists. To this end, we have successfully expressed and purified the α L-TM and β 2-TM domains. Later, we have used solution NMR techniques to determine their structure in phospholipid bicelles. Putative α L/ β 2 interfacial residues have also been determined by titration experiments between one of the TM (labelled) with its unlabelled partner.

3.2 Construct design and sample preparation

The integrin α L-TM and β 2-TM constructs were designed to cover the entire TM domain, based on the previous α IIb/ β 3 study (Lau et al., 2009), with additional several residues as the termini residues are usually rather flexible (construct sequence in Fig 2-1). The TM peptides were produced through heterologous expression in *E. coli* as 6-His tagged peptides. The peptides were captured from the crude lysate by Ni^{2+} -affinity resin, followed by on-column enzymatic digestion using TEV protease (recognition site is placed between the 6-His tag and the TM peptide) to separate the TM peptides from the 6-His tag. The released TM peptides have slight yet discernible MW decrease in SDS-PAGE upon the removal of the 6-His tag (Fig. 3-4A, TEV).

Figure 3-4. Expression and purification of integrin TM domain peptides. Integrin TM peptides were heterologously expressed in *E. coli* cells and purified by Ni-NTA affinity purification. Affinity tags (6-His) were removed by on-column TEV-protease digestion, and the peptides were TCA-precipitated, methanol-extracted, and purified by RP-HPLC. (A) SDS-PAGE of the Ni-NTA-eluted (elution) and the TEV protease-cleaved peptides (TEV). (B) SDS-PAGE of the methanol-extracted peptides (meOH) and the RP-HPLC-purified peptides (HPLC) from fractions indicated by arrows in (C) HPLC chromatogram. (D) MALDI TOF/MS spectra of the purified integrin TM peptides.

Following TCA-precipitation of the entire mixture, pure TM peptides can often be extracted from the precipitate by using methanol (Fig. 3-4B, meOH). If the methanol extracts were not sufficiently pure, they were further purified by RP-HPLC (Fig. 3-

4B, HPLC and Fig. 3-4C). Finally, the identity of the peptides was confirmed once again by MALDI/TOF-MS based on their molecular weight (Fig. 3-4D).

Unlabelled peptide samples were produced for CD and FTIR measurements. Samples for solution NMR experiments consists of ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labeled peptides with varying deuteration levels (50% and 99%) which were expressed in M9 media and purified in the same way as the unlabelled peptides.

3.3 Reconstitution in bicelles and secondary structure elements

A properly reconstituted single-pass TM domain would form a membrane-embedded α -helix, instead of random coils due to the precipitation of hydrophobic peptides. To determine whether αL -TM and β2 -TM peptides can be reconstituted successfully into lipid bilayers and into bicelles, we examined their secondary structures by using CD and FTIR spectroscopy, and confirmed whether the peptides are inserted into the membrane by H/D exchange experiments.

Firstly, we collected FTIR spectra of αL -TM and β2 -TM peptides reconstituted in lipid bilayers (Fig. 3-5A, blue). The amide I bands of both peptides are centred at 1655 cm^{-1} , indicating that they are mainly α -helical. Further examination by Fourier-self deconvolution of the amide I band shows that $> 90\%$ of both peptides are α -helical (Fig. 3-5A, red).

Next, H/D exchange experiments were performed on lipid-reconstituted TM peptides (Fig. 3-5B). Normalized amide II band area ratio before and after infusion with D_2O vapour shows that about 70% (~ 35 residues) of αL -TM and $\sim 62\%$ (~ 28 residues) of β2 -TM were exchange-protected. This protection from H/D exchange indicates that the TM peptides were inserted into the lipid bilayer.

Finally, we collected CD spectra of αL -TM and β2 -TM peptides reconstituted in DMPC/DHPC bicelles (Fig 3-5C). Fitting of the CD spectra resulted in estimated 60% helical content in both peptides (~ 30 residues for αL -TM and ~ 23 residues for β2 -TM). The difference in estimated helical content in CD and FTIR is rather large. One reason might be the less-than-optimal CD spectra fitting due to parts of the spectra ($< 205\text{ nm}$) being unusable from excessive absorbance (mainly contributed by

the bicelle). Another possible reason might be the inherently different water content of the two samples: a bicelle solution (CD) would have much higher water content than a stacked lipid bilayer (FTIR). This would mean that the α -helical content is more abundant in the less hydrated environment. Whether or not this is translated into an α -helical form for the GFFKR motif cannot be ascertained from the FTIR data.

As membrane insertion cannot be examined by using CD, the question of whether bicelle-reconstituted TM peptides are inserted into the bicelle will be revisited when the NMR data are presented and discussed. Nevertheless, both the CD and FTIR data satisfactorily show that α L-TM and β 2-TM peptides can be successfully reconstituted into lipid bilayers and bicelles.

Figure 3-5. Secondary structure of reconstituted α L and β 2 TM peptides. Integrin peptides were reconstituted in DMPC lipid bilayer at 1:50 peptide-to-lipid ratio for secondary structure determination by FTIR, or in DMPC/DHPC bicelles ($q = 0.3$) for the same by CD. (A) FTIR amide I band of the integrin peptides centred at 1655 cm^{-1} (blue). The corresponding Fourier-self deconvoluted spectra show the proportion of each major peaks that form the amide I band (red). (B) FTIR amide I and amide II bands of the integrin peptides during H/D exchange experiment: blue with residual H_2O , and green after D_2O vapour exposure. (C) CD spectra of the integrin peptides showing two minima at 208 and 222 nm, typical characteristics of an α -helix.

3.4 Solution NMR structure of integrin TM peptides

Solution NMR data of the integrin TM peptides were obtained using bicelle-reconstituted peptide samples. We initially confirmed that $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra of ^{15}N -labeled peptides showed good peak dispersion (Fig. 3-6). Afterwards, we collected a series of 3D-spectra to assign backbone and sidechain resonances, and to obtain NOE distance restraints for structure calculation.

Figure 3-6. Assigned $[^1\text{H}-^{15}\text{N}]$ -TROSY-HSQC of $\alpha\text{L-TM}$ and $\beta 2\text{-TM}$. $[^1\text{H}-^{15}\text{N}]$ -TROSY-HSQC spectra of (A) $\alpha\text{L-TM}$ and (B) $\beta 2\text{-TM}$ reconstituted in DMPC/DHPC ($q=0.3$) bicelles, in 50 mM K-PO_4 pH 6.5. Cross-peaks are labelled by one-letter residue type and number.

NOE distance restraints, used to calculate the structures of $\alpha\text{L-TM}$ and $\beta 2\text{-TM}$, were obtained from ^{15}N -NOESY-HSQC and ^{13}C -NOESY-HSQC spectra, respectively. A total of 20 structures have been calculated for $\alpha\text{L-TM}$ and $\beta 2\text{-TM}$, and their structure statistics are summarized in Table 3-1 and 3-2, respectively. The structure of integrin $\alpha\text{L-TM}$ has been deposited on the Protein Data Bank (PDB) with ID 2m3e, and the assigned chemical shifts have been deposited on the Biological Magnetic Resonance Bank (BMRB) with ID 18958.

Table 3-1. Structure statistics for the selected 20 structures of α L-TM

NMR restraints	
Total unambiguous distance restraints	600
Intra residual	253
Sequential ($ i - j = 1$)	155
Short-range ($ i - j \leq 1$)	408
Medium ($2 \leq i - j \leq 4$)	168
Long range ($ i - j \geq 5$)	24
Dihedral angle restraints	46
RDC restraints	47
Hydrogen bond restraints ^a	20
RMSD from the average atomic coordinates (residues 1065-1093, Å) ^b	
Backbone atoms	0.23 ± 0.05
All heavy atoms	0.65 ± 0.07
Ramachandran analysis (%)	
Residues in most favored regions	75.7
Residues in additional allowed regions	24.1
Residues in generously allowed regions	0.2
Residues in disallowed regions	0.0

^aBackbone hydrogen bonds of α -helix are applied for regions confirmed to be α -helix according to local NOE pattern and H^N-H₂O chemical exchange experiment.

^bStatistics are calculated and averaged over an ensemble of the 20 structures with lowest target function from CYANA.

Table 3-2. Structure statistics for the selected 20 structures of β 2-TM

NMR restraints	
Total unambiguous distance restraints	421
Intra residual	151
Sequential ($ i - j = 1$)	158
Short-range ($ i - j \leq 1$)	309
Medium ($2 \leq i - j \leq 4$)	110
Long range ($ i - j \geq 5$)	2
Dihedral angle restraints	64
Hydrogen bond restraints ^a	28
RDC restraints	44
RMSD from the experimental residual dipolar couplings (Hz)	
$^1D_{NH}$	0.44 ± 0.03
RMSD from the average atomic coordinates (residues 683-704, Å) ^b	
Backbone atoms	0.28 ± 0.11
All heavy atoms	0.55 ± 0.13
Ramachandran analysis (%)	
Residues in most favored regions	91.8
Residues in additional allowed regions	8.2
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0

^aBackbone hydrogen bonds of α -helix are applied for regions confirmed to be α -helix according to local NOE pattern and H^N -H₂O chemical exchange experiment.

^bStatistics are calculated and averaged over an ensemble of the 15 structures with lowest target function from CYANA.

Before discussing the calculated structure of α L-TM and β 2-TM in bicelles, we first examined the structural stability and membrane insertion of these TM domains in bicelles.

Firstly, steady-state [^1H , ^{15}N] heteronuclear NOE measurements on α L-TM and β 2-TM (Fig. 3-6A & B) indicate that the majority of the TM peptides form well-folded and stable structure (high I/I₀ ratio), and only towards the termini they become quite flexible. Flexibility at the N- and C-terminus are common and to be expected, thus the construct design that was well beyond the TM domain ensured that the region of interest adopts a stable structure.

Figure 3-7. Structural stability of α L and β 2 TM domains. Steady-state [^1H , ^{15}N] heteronuclear NOE data of (A) α L-TM and (B) β 2-TM reconstituted in DMPC/DHPC bicelles indicating parts of the structure that are more stable (I/I₀ ~ 1) or more flexible (low I/I₀). Corresponding amino acid sequence of each TM peptide is shown for clarity. N-terminal residues indicated in grey in α L-TM (SNA) and β 2-TM (SM) are what remained of the 6-His tag after TEV cleavage.

Next, we revisited the question of whether α L-TM and β 2-TM are inserted into bicelles. Specifically, we determined the membrane-embedded residues in the TM peptides by H/D exchange experiments. Upon exchanging H₂O to D₂O, [¹H-¹⁵N] HSQC spectra show that in both TMs there is a stretch of ~20 exchange-protected residues (Fig. 3-7A) forming the central part of the TM: in α L-TM residues V1069-V1089 and in β 2-TM residues V687-L704 (Fig. 3-7B). Notably, F1092 of α L-TM is also protected from H/D exchange. The significance of this will be discussed in the next section.

Figure 3-8. H/D exchange behaviour of α L and β 2 TM domains in bicelles. Membrane-embedded regions of α L- and β 2-TM domains were determined by H/D exchange experiments. Bicelle-reconstituted α L- and β 2-TM samples were lyophilized and resuspended in D₂O. (A) [¹H-¹⁵N] HSQC spectra of bicelle-reconstituted α L- and β 2-TM domains in D₂O. Visible cross-peaks (labelled) are protected from H/D exchange due to being embedded in the bicelles. (B) These embedded residues are coloured red in the amino acid sequence, and where these residues form a stretch of ~20 amino acids they are designated as the TM region.

The TM region determined herein for β 2-TM is rather short, so we examined this further by using paramagnetic reagents 5-DSA and 16-DSA (Fig 3-7C). Addition of 5-DSA caused intensity reduction in residues 675-685 and 702-703, marking the residues located near or at the membrane surface. The addition of 16-DSA caused a more pronounced intensity reduction in the stretch of residues 683-700, confirming

that these residues are well inside the membrane. This data suggests that the membrane-embedded region of β 2-TM may actually be close to 25 residues long. This is in close agreement with the ~28 H/D exchange-protected residues obtained for β 2-TM in lipid bilayers. Further, in comparison with the 5-DSA and 16-DSA PRE experiments, the number of membrane-embedded residues may be underestimated in H/D exchange experiments using bicelles, and overestimated in H/D exchange experiments using lipid bilayers.

Figure 3-9. PRE attenuation by 5-DSA and 16-DSA on β 2-TM. PRE effect was quantified by ratio of peak intensity upon addition of 5-DSA and 16-DSA to bicelle-reconstituted β 2-TM. The corresponding amino acid sequence of β 2-TM is shown above with H/D exchange-protected residues indicated in red for comparison.

Having examined the flexibility of the structure and the extent of membrane insertion, we now turn to the calculated structure of integrin TM peptides. In agreement with CD and FTIR data, the solution NMR structures of the two peptides show a single α -helix: α L from L1065-V1089 (Fig. 3-10A) and β 2 from I679-D709 (Fig. 3-10B). The length of the α -helices in these models is in close agreement to the helical content estimated by CD spectroscopy with bicelles, but considerably shorter than the one estimated by FTIR spectroscopy with lipid bilayers. This difference is either due to the inaccuracy of the method, or the reconstitution media used (bicelles vs lipid bilayers). In lipid bilayers, the α -helix could be more extended and tilted than in bicelles, possibly due to the lower hydration inherent in stacked lipid bilayers used for FTIR measurement. Different level of hydration issue also account for the lower

number of H/D exchange-protected residues observed in [^1H - ^{15}N] HSQC (Fig 3-8) than that determined by FTIR in lipid bilayers.

Figure 3-10. Solution NMR structure of integrin αL and $\beta 2$ TM domains. Cartoon representation showing ensemble of 20 structures calculated by simulated annealing, employing NOE restraints obtained for (A) αL -TM and (B) $\beta 2$ -TM in DMPC-DHPC bicelles. Side chains are shown as blue lines. Residues at the boundary of the α -helix are indicated.

3.5 Reversal of the GFFKR motif backbone in αL TM domain

Apart from the α -helical region, the rest of the peptide is featureless, except at the few residues after V1089 in αL -TM, i.e. at the GFFKR motif. This motif adopts a structural feature similar to that previously found in αIIb TM domain (Lau et al., 2009), which we could observe through NOE connectivities between L1086, V1089,

F1091, and F1092 (Fig. 3-11A). These connectivities indicate a close hydrophobic packing between the two Phe residues against the α -helix, which is shown in the model (Fig. 3-11B). In addition, F1083 also packs against F1092 and L1086, and together forms a hydrophobic surface that can interact with β 2-TM. This packing is made possible by the peptide backbone reversal with respect to its preceding TM domain. This brings F1091 and F1092 back into the membrane and into close proximity to the TM domain. This backbone reversal is also supported by the observed protection of F1091 backbone from H/D exchange, indicating its insertion into the membrane.

We note that the structure of the GFFKR motif of α L-TM is quite stable, as indicated by the HNOE measurement presented earlier (Fig. 3-6). Thus, in a bicelle this motif is probably not prone to spontaneous structural conversion into, for example, an α -helical extension of the preceding TM domain as observed for α L-TM in organic solvent (Yang et al., 2009). However, it remains to be shown if the GFFKR motif can interconvert between the backbone reversal and the α -helical form under different conditions, e.g. in the presence of α L interacting partners that bind to its C-terminal tail.

Therefore, the backbone reversal at the GFFKR motif previously observed in α Ib-TM (Lau et al., 2009) is now also observed in α L-TM, i.e. when both α -TM peptides were reconstituted in bicelles. Superimposition of α Ib and α L models shows that the structure of GFFKR motif in α Ib and α L is very similar (Fig. 3-11C), and supports that the GFFKR motif is a structurally conserved motif in α -integrins.

Figure 3-11. Structure of the GFFKR motif at the C-terminal juxtamembrane region of the α L TM domain in bicelle. (A) Four selected strips from ^{15}N -NOESY-HSQC and ^{13}C -NOESY-HSQC showing NOE connectivities between side-chains of Leu-1086, Val-1089, Phe-1091, and Phe-1092. (B) Cartoon representation of the C-terminal juxtamembrane region of α L-TM. Side-chains for which NOE contacts have been observed are shown in sticks representation. (C) Superposition between the juxtamembrane region of α L-TM (red) and α IIb-TM (blue, PDB: 2K1A) showing structural similarity between the two models.

3.6 The α L/ β 2 heterodimer interface

To determine whether a polar interaction exists at the OMC of α L/ β 2 heterodimer, we titrated unlabelled peptide to its ^{15}N -labelled partner. Changes in the $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC spectra indicate which residues are involved in the α L/ β 2 heterodimer interface.

Titration of unlabelled β 2-TM into labelled α L-TM revealed the interfacial residues in α L side: L1067, Y1068, S1071, G1072, G1074, G1075, Y1087 and F1091 (Fig. 3-12A). When β 2-TM mutated at the GxxxG-like motif G685F and G689F were used, identical results were obtained (not shown), indicating that the small residues (“G”) in the GxxxG-like motif of β 2 do not participate in forming the heterodimer interface. This was further confirmed by the titration of unlabelled α L-TM into labelled β 2-TM:

perturbations were observed in A680, V683, G684, and side chain of W701 (Fig. 3-12B), and not in the Gly residues forming the GxxxG-like motif (G685 and G689).

Figure 3-12. Integrin TM cross-titration experiment. [^1H - ^{15}N]-HSQC showing residues affected by addition of lyophilized, unlabelled TM peptide into its ^{15}N -labelled partner reconstituted in DMPC/DHPC ($q = 0.3$) bicelles: (A) unlabelled $\beta\text{2-TM}$ added to ^{15}N -labelled $\alpha\text{L-TM}$ (red: before addition, blue: after addition at 2:1 molar ratio) and (B) unlabelled $\alpha\text{L-TM}$ added to ^{15}N -labelled $\beta\text{2-TM}$ (blue: before addition, red: after addition at 2:1 molar ratio).

The residues involved in the $\alpha\text{L}/\beta\text{2}$ interaction determined from cross-titration data were mapped onto the solution NMR model of $\alpha\text{L-TM}$ and $\beta\text{2-TM}$ (Fig. 3-13). In the N-terminal end of the TM domain (the OMC interface) αL interacts with β2 through its GxxxG-like motif, i.e. S1071/G1075 (Fig. 3-13A). However, G685/G689 which forms this motif in $\beta\text{2-TM}$ is oriented away from the interface (Fig. 3-13B), as previously observed in the $\alpha\text{IIb}/\beta\text{3}$ model (Lau et al., 2009).

In this part of the TM domain, i.e. the OMC region, we proposed previously the presence of a polar interaction between S1071 (αL) and T686 (β2) (Vararattanavech et al., 2010). Indeed, S1071 is affected by the heterodimerization, but a similar effect was not observed in T686. Because only the backbone amide was observed in [^1H - ^{15}N]-HSQC, it is possible that the changes occur only in the sidechains and not in the backbone. Nevertheless, these data lend support to the interaction between S1071 and T686, although still insufficient to decide whether it is a polar interaction.

Figure 3-13. Residues at the α L/ β 2 heterodimer interface. Top view and front view of the α L/ β 2 heterodimer interface residues (sticks representation) in (A) α L and (B) β 2 as determined by cross-titration of unlabelled peptide to its unlabelled partner (Fig. 3-10). The interfacial residues are located in one side of the TM. Residues indicated in red are the small residues ("G") of the GxxxG-like motif. The "G" residues of α L (S1071/G1075) participates in the interface, whereas those of β 2 (G685/G689) are not.

In the C-terminal end of the TM domain, i.e. the IMC region, F1091 and Y1087 (α L) interact with W701 (β 2). All of the three aromatic amino acids possibly stack together due to pi interactions, but this is yet to be shown. Nevertheless, residue F1091 is part of the GFFKR motif, so this observation adds structural explanation as to why deletion of this motif disrupts α L/ β 2 heterodimerization, manifesting as decreased recycled α L/ β 2 level (Lu and Springer, 1997) and constitutive α L/ β 2-mediated cell adhesion in a transgenic mouse model (Semmrigh et al., 2005).

We could not observe any perturbations for R1094 and D709, which are homologous to the residues forming the α Ib/ β 3 salt bridge R995-D723. On the α L side, this is due to R1094 ^1H - ^{15}N cross-peak being located at a particularly crowded spectral region, thus making the observation of any perturbations difficult. However, N1095 was affected (Fig 3-12A), suggesting that there could be an interaction with β 2-TM nearby this residue.

On the β 2 side, there are no perturbations detected at or around residue D709. This could be due to the β 2 construct being too short to allow such interaction to occur, or because no salt bridge is formed between the α L/ β 2 TM domain. The first possibility was tested by using a longer β 2-TM construct spanning residue 675-713 (3 residues longer), but titration experiment with unlabelled α L-TM gave a same result (not shown). The second possibility is at odds with the perturbations observed at residue N1095 which suggests that there is some interaction in this region. Thus, it is likely that the perturbations occur only in the sidechain of D709 and not the backbone. This can potentially be observed by monitoring the cross-titration experiment in [^1H - ^{13}C]-HSQC spectra, as long as the spectra do not get too crowded.

3.7 Summary and future works

In summary, herein I have presented a solution NMR model for α L/ β 2 TM heterodimer in bicelle which is very similar to the previously published α Ib/ β 3 TM solution NMR model in bicelle (Lau et al., 2009). These data suggest that both the general α / β heterodimer interface and the GFFKR backbone reversal motif at the juxtamembrane region of α -integrins are features that may be commonly found in all α / β integrin pairs. The latter has so far been observed only when α / β integrin TM domain are reconstituted in bicelles, and not in micelles or organic solvent.

However, there is still not enough evidence to support or reject that a polar interaction exists at the α L/ β 2 heterodimer interface, as we have previously proposed (Vararattanavech et al., 2010). Further details on the heterodimer interface, e.g. in the form of intermonomeric NOE contacts, are needed. Polar interaction can then be inferred from the proximity between the moieties involved. Inter-monomeric NOEs can be obtained, for example, by using the selective methyl labelling method

(selectively protonating methyl groups in otherwise fully-deuterated peptides), which has been used to determine the heterodimer structure of α IIb/ β 3 in bicelles (Lau et al., 2009). Alternatively, inter-monomeric NOEs can be obtained by asymmetrically deuterating the integrin TM peptides, i.e. ^{13}C -labelling in one peptide and $^{15}\text{N}/\text{D}$ -labelling in the other peptide. Only inter-monomeric NOEs will be observed due to physical separation between $^1\text{H}^{\text{N}}$ and ^1H in different peptide chains. This method has been implemented to obtain inter-monomeric NOEs between SARS-CoV E monomers, as described in the next chapter.

CORONAVIRUS E PROTEIN EXTRAMEMBRANE STRUCTURE AND HOMOOLOGOMER FORMATION

4.1 Introduction

In this chapter, I will present a structural study on the coronavirus (CoV) envelope (E) proteins, the second membrane protein analysed in this thesis. CoV E proteins belong to a class of protein known as viroporins, whose structure, like the integrin, is also dependent on its reconstitution environment.

4.1.1 The viroporins

Viroporins are a class of small membrane proteins of viral origin that alters membrane permeability to ions or small molecules (Carrasco, 1995). This is achieved by the oligomerization of viroporin monomers, which are usually single-pass transmembrane proteins, into channels or pores. The most well-known and extensively studied viroporin is the influenza A M2 tetrameric proton channel (Pinto et al., 1992). Its structure has been solved in various environments, such as by solution NMR with DHPC detergent micelles (Schnell and Chou, 2008), solid-state NMR with oriented lipid bilayers (Nishimura et al., 2002; Sharma et al., 2010; Stouffer et al., 2008), as well as X-ray crystallography, in OG detergent (Hong and DeGrado, 2008). In these environments, M2 protein shows different TM domain orientations and lengths, as well as inhibitor interaction sites, all of which are important in explaining the mechanistic details of M2 proton transport (Cross et al., 2011; Sharma et al., 2010).

Only a handful other viroporins have detailed structural information, such as hepatitis C virus p7 protein in DPC micelles (Ouyang et al., 2013), HIV-1 viral protein U (VPU) in DHPC micelles and oriented lipid bilayers (Park et al., 2003), and human respiratory syncytial virus (hRSV) small hydrophobic protein (SH) in DPC micelles and phospholipid bicelles (Gan et al., 2012; Li et al., 2014c). Some of these viroporins have been studied in more than one environments, allowing comparison between structures obtained in different membrane mimic systems. For example, HIV-1 VPU TM structures determined by dipolar wave measurement in detergent micelles and in

lipid bilayers are identical, whereas the extramembrane structure of SH protein is different in micelles versus in bicelles (Li et al., 2014c). Another viroporin whose structure is dependent on reconstitution environment, which will be discussed this chapter, is the coronavirus envelope (E) protein.

4.1.2 The coronavirus E protein

Coronaviruses (CoV) are lipid-enveloped viruses that can be classified into four groups designated as α , β , γ , and δ -coronaviruses (Enjuanes et al., 2000). Members of the second group (β -coronaviruses) include the CoVs responsible for the severe acute respiratory syndrome (SARS) and Middle-East respiratory syndrome (MERS), which produced a near pandemic in 2003 and 2013, respectively. Other frequently studied CoVs that infect animals include the porcine transmissible gastroenteritis virus (TGEV) from α -coronaviruses, murine hepatitis virus (MHV) from β -coronaviruses, and avian infectious bronchitis virus (IBV) from γ -coronaviruses.

CoV E proteins are short (76-109 amino acids), integral membrane proteins that are just barely present in the CoV virion, as compared to more abundant proteins such as the spike (S) and membrane (M) protein (Corse and Machamer, 2000; Godet et al., 1992; Liao et al., 2006; Vennema et al., 1996; Yu et al., 1994). Instead, E protein can be found more abundantly in internal membranes of infected cells, especially the ER-Golgi intermediate compartment (ERGIC), where virions assemble and bud into the lumen (Lopez et al., 2008; Nieto-Torres et al., 2011). The E protein has been thought to drive this virion budding process, as co-expression of E and M protein (and not M protein alone) has been found to produce viral-like particles (VLP) in the absence of other viral proteins (Baudoux et al., 1998; Bos et al., 1996; Corse and Machamer, 2000; Vennema et al., 1996).

Deletion of E protein generally attenuates the virus. TGEV and MERS-CoV without E protein (ΔE) were shown to accumulate immature virions in the secretory vesicles and no infectious particles could be recovered (Almazán et al., 2013; Ortego et al., 2002; Ortego et al., 2007). MHV ΔE grows with lower titer and produces smaller plaque (Kuo and Masters, 2003; Kuo et al., 2007). Similarly, SARS-CoV ΔE also grows with lower titer and produces higher proportion of aberrantly-shaped virions (DeDiego et al., 2007). Both SARS-CoV ΔE and MERS-CoV ΔE have been found to

be highly attenuated *in vivo* and potentially can be made into vaccines (Almazán et al., 2013; DeDiego et al., 2008; Lamirande et al., 2008).

The attenuation of various CoV ΔE indicates that E protein has an important role in pathogenesis but exactly how is not yet known. An important clue to this lies on the ion channel activity of the E protein, as SARS-CoV carrying E protein with mutations that abolish its ion channel activity has been shown to also be attenuated *in vivo* (Nieto-Torres et al., 2014). These ion-channel inactivating mutations are located in TM-domain residues N15 and V25, specifically N15A and V25F (Torres et al., 2007; Verdiá-Báguena et al., 2012). A homologous mutation in IBV E, i.e. T16A, decreases secretory pathway disruption, which is another hallmark of CoV infection (Ruch and Machamer, 2012). This mutation also inactivates the ion channel activity of IBV E (manuscript in preparation).

E protein appears to suppress apoptosis and upregulate inflammatory responses, thus resulting in the cellular and tissue pathology observed in CoV infection (DeDiego et al., 2014; DeDiego et al., 2011). This also seems to be related to the ion channel activity of E protein, as SARS-CoV E ion channel activity has been demonstrated to activate the inflammasome and in turn increases the level of inflammatory cytokines IL-1 β , TNF, and IL-6 (Nieto-Torres et al., 2015; Nieto-Torres et al., 2014).

In order to better understand the function of the CoV E protein, structural information is indispensable. Just as the structure of influenza A M2 protein helps us to understand the mechanism of its proton channel activity inhibition by adamantane drugs, the structure of E protein will help us to understand its ion channel activity, and lay grounds for rational drug design. We will now see how far studies on the E protein structure determination have progressed.

4.1.3 Structural models for CoV E protein

CoV E proteins are quite diverse in their amino acid sequence (Fig. 4-1A for sequence alignment). Among the few conserved residues, there is a homologous series of polar residue at the TM region (Fig. 4-1, red arrow), which in SARS-CoV E and IBV E is important for their ion channel activity, and a completely conserved proline residue at the C-terminal region (Fig. 4-1A, green arrow).

Figure 4-1. General features of CoV E proteins. (A) Representative CoV E protein sequences from α , β , and γ -CoV aligned using Clustal Omega software. Consensus symbols "*", ":", and "." indicate fully conserved, high similarity, and low similarity residues, respectively. The position of the polar residue in the TM and the conserved proline residue are indicated by the red and green arrowheads, respectively. (B) Schematics of structural features shared by CoV E proteins: a short N-terminus followed by TM domain, juxtamembrane cysteines (Cys), a hydrophobic region with β -turn- β motif, and a variable-length C-terminal tail. (C) Hydrophobicity-based TM domain probability (TMHMM) is shown as vertical bars above each residue. Horizontal bars below the protein sequence indicate the consensus secondary structure prediction colour-coded to secondary structure elements. Dotted vertical lines indicate the position of the polar residue in the TM and the conserved proline residue.

Despite their low sequence similarity, CoV E proteins share a general architecture: a short N-terminal extramembrane segment, followed by a TM domain, a juxtamembrane cluster of palmitoylated cysteines, a ~25 residues hydrophobic region, and finally a C-terminal tail of variable length between CoVs (Fig. 4-1B).

Consensus secondary prediction (https://npsa-prabi.ibcp.fr/NPSA/npsa_seccons.html) using MLRC (Guermeur et al., 1999), DLC (King and Sternberg, 1996), and PHD (Rost and Sander, 1993) algorithms shows several secondary structure elements that are also shared among CoV E proteins: a helical (TM) domain, followed by a β -turn- β motif centred on a completely conserved proline residue at the hydrophobic region (except in α -CoVs such as TGEV E), and a C-terminal random coil (Fig. 4-1B). The hydrophobicity of the hydrophobic region varies from one CoV to the other: in IBV E, for example, the region is hydrophobic enough to be predicted as a second TM (TMHMM, <http://www.cbs.dtu.dk/services/TMHMM/>).

One way to determine the actual number of TM domain in CoV E proteins is by examining its topology in cells. However, this may lead to artefacts when using transfected cells or affinity tags that may alter the observed topology. A recent study using untagged SARS-CoV E in infected cells has determined that it has a cytoplasmic C-terminus and a luminal N-terminus topology, indicating the presence of a single TM domain (Nieto-Torres et al., 2011). A similar topology has also been observed in untagged MHV E (Raamsman et al., 2000) and IBV E (Corse and Machamer, 2000; Ruch and Machamer, 2012). This topology means that there are an odd number of TM domains, i.e. one, as there are not enough hydrophobic residues to form three TM domains. This also implies that the C-terminal hydrophobic region cannot be a TM domain, even in the case of IBV E. Thus, only the N-terminal hydrophobic region constitutes the TM domain.

The TM domain of E proteins oligomerizes into a pentameric bundle with a central channel that allows ion to pass through, albeit with poor selectivity (Verdiá-Báguena et al., 2012; Wilson et al., 2006). The pentameric arrangement is one of several homo-oligomer models proposed from MD simulation of symmetric α -helices based on evolutionary conservation data of CoV E proteins (Torres et al., 2005). Other possible models described therein include dimers and trimers, but these cannot explain the ion channel activity of E protein. This leaves us with the pentamer models, of which, two

possible TM orientations have been proposed and designated as pentamer A and pentamer B (Fig. 4-2).

Figure 4-2. Two pentameric models of SARS-CoV E protein. Pentameric models (A and B) for E protein TM domain obtained by MD simulation (Torres et al., 2005). The two key residues for ion channel activity, N15 and V25, are shown in the pentamer bundle as spheres. Their orientation in the TM is also shown in slices through the pentamer bundle (boxes).

The two models are related to each other by a $\sim 50^\circ$ TM rotation, as illustrated in Fig. 4-2 using the positions of N15 and V25. Mutations at these two residues abolish E protein channel activity (Parthasarathy et al., 2008). In both models, N15 side chain is oriented towards the channel lumen, although in pentamer B it is slightly away from the lumen, as compared to in pentamer A in which N15 directly faces the lumen. V25 in pentamer A also faces the lumen, whereas in pentamer B V25 is at the TM-TM interface. The luminal orientation of N15 and V25 in pentamer A can thus directly explain why N15A and V25F mutations lead to channel inactivation. As for pentamer B, the effect is probably a result of overall pentamer destabilization due to changes in interhelical contact. This comparison has led us to consider pentamer A as a possible model for the closed state, whereas pentamer B would be the open state.

Experimental data have not been able to satisfactorily determine which of the two models are closer to the native conformation. Site-specific FTIR measurements using $^{13}\text{C}=^{18}\text{O}$ label on a lysine-terminated synthetic TM domain of E protein produced a hybrid of model A and model B, with residues N-terminus to V25 consistent with

model B, and residues C-terminus to V25 consistent with model A (Torres et al., 2006). When the TM peptide was not lysine-terminated, residues 21-24 were oriented according to model A (Parthasarathy et al., 2008), and thus model A was taken as the correct model of E protein pentamer. A later model was built from the solution NMR data of synthetic E protein residues 8-38, incorporating 2 inter-monomeric NOEs and used model A as a template (Pervushin et al., 2009). All these models, however, still lack the extramembrane domain which may influence the TM orientation.

4.1.4 The extramembrane domain of CoV E protein

As mentioned before, the C-terminal extramembrane domain of E proteins was predicted to consist of β -turn- β motif and a random coil (Fig. 4-1). The β -turn- β motif is present in all CoV E protein sequences (except in α -CoV), and at its centre there is a completely conserved proline residue. This motif has been shown to be both sufficient and necessary to direct the localization of a membrane protein to Golgi membranes and thus seems to be important for the sequestration of CoV E proteins to the ERGIC membranes (Cohen et al., 2011).

FTIR measurement on synthetic peptide fragments of different parts of the SARS-CoV E C-terminal extramembrane domain showed that these fragments fold according to their predicted secondary structure (Surya et al., 2013). Specifically, a synthetic peptide containing residues 46-60 of SARS-CoV E is β -structured whereas that containing residues 59-76 forms a random coil. However, this is in contrast with the finding that expressed full-length SARS-CoV E forms mostly α -helix, as determined by CD in SDS and in DPC micelles, and by FTIR in lipid bilayers (Parthasarathy et al., 2012). These observations suggest that in the context of full-length E protein, individual structural elements may fold differently due to the presence of the other parts of the protein. We might also hazard a guess that the extramembrane domain influences the orientation of the TM domain. It is therefore desirable to solve the structure of the E protein with its extramembrane domain present.

4.1.5 Aim and approach

In the rest of this chapter, I will describe our efforts in solving the structure of CoV E proteins. Specifically, we want to know whether a β -turn- β motif really exists in the

C-terminal extramembrane domain, and whether the TM orient according to pentamer A or pentamer B model.

Unfortunately, the full-length SARS-CoV E protein has a very poor solubility in detergents and tends to aggregate. We approached this issue from two different directions: to study homologous E proteins from other CoVs (which share similar secondary structure prediction, Fig. 4-1), or to mutate/truncate the SARS-CoV E protein to improve its solubility. A mutagenesis approach proved to be quite successful in solving of the structure of another viroporin, HCV p7, where non-conserved residues were mutated to improve NMR spectral quality (Ouyang et al., 2013). The present study uses a combination of mutating cysteine residues, which do not contribute to oligomerization (Parthasarathy et al., 2012), and truncating the SARS-CoV E construct, which will be presented in chapter 4.2. The question of whether the truncation affects the structure of E protein is explored in chapter 4.3.

The structure of E protein was then solved by solution NMR in detergent micelles (Li et al., 2014b). Although bicelles are considered as the more ideal system, the spectral quality of E protein in bicelles is very poor, therefore detergent micelles were used. The structure of E protein monomer was finally determined in DPC/SDS mixed micelles, which gives a balance between good spectral quality due to SDS and the similarity of DPC to lipid bilayer. The structure of E protein pentamer was determined in LMPG micelles, which is more lipid-like but gave only a moderate spectral quality. The suitability of these systems is examined and discussed in chapter 4.4. Finally, chapter 4.5 and 4.6 presents the monomeric and pentameric structure of E_{TR}, respectively.

4.2 Construct optimization and sample preparation

Initially, we attempted to express the full-length SARS-CoV E protein (E_{FL}) fused to a maltose binding protein (MBP) expression tag (MBP-E, construct sequence in Fig. 2-1). MBP binds to amylose and thus allows for affinity purification. In addition, the construct was also 6-His tagged at the N terminus to enable purification by using Ni²⁺-affinity chromatography. This proved to be a wise choice as the affinity of MBP-

E to amylose resin turned out to be very weak (data not shown), possibly due to the interference of detergent.

The 42-kDa MBP tag was removed by utilizing TEV protease cleavage on the linker peptide between the MBP and the E protein (Fig. 4-3A), and the E protein was separated from the mixture by RP-HPLC (Fig. 4-3B). Unfortunately, the uncleaved protein (MBP-E) co-eluted with the cleaved E protein (Fig. 4-3B & C), and the uncleaved MBP-E could not be completely removed due to the incomplete TEV protease cleavage, despite optimization attempts.

The co-elution of MBP-E with E protein probably occurred due to the interaction between the uncleaved portion of E protein in MBP-E with the cleaved E protein. This notion is supported by the ladder-like pattern observed above the MBP-E protein bands in SDS-PAGE of HPLC fractions (Fig. 4-3B, white box). This interaction might be the same one driving pentamerization of E protein through its TM domain, or it might be a non-specific disulfide formation by the juxtamembrane cysteines, as has been suggested previously (Parthasarathy et al., 2012).

Therefore, to avoid this issue, we need to either weaken the oligomerization tendency of the E protein, or prevent the disulfide formation. Indeed, when all three cysteines were mutated into alanines (C40A/C43A/C44A), we were able to successfully purify the full-length E protein (Fig. 4-3D).

Although these cysteines have been shown to be palmitoylated in MHV E (Lopez et al., 2008), IBV E (Corse and Machamer, 2002) and SARS-CoV E (Liao et al., 2006), they seem to be dispensable at least for the case of MHV E (Lopez et al., 2008). These cysteine residues also do not seem to contribute to the oligomerization of TGEV E (Godet et al., 1992), MHV E (Lopez et al., 2008), SARS-CoV E, and IBV E (Parthasarathy et al., 2012). Therefore, we can reasonably expect that the structure of a cysteine-less E protein would still be relevant.

After mutating the cysteines to alanines, however, the full-length E protein is still very poorly soluble, and has low expression yield. The amount of unlabelled full-length E protein was only sufficient for FTIR measurements. ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labelled full-length E samples were also produced to obtain an ^{15}N -HSQC and a 3D-HNCA spectrum in SDS micelles.

Figure 4-3. Purification of E protein. (A) SDS-PAGE showing the removal of the MBP tag from MBP-E by TEV protease digestion. The digestion progress is shown at 0 hr and 18 hrs after adding TEV protease to MBP-E protein. (B) RP-HPLC chromatogram of E protein purification after TEV cleavage, with five collected fractions (numbered). (C) SDS-PAGE of the RP-HPLC fractions showing the co-elution of the uncleaved MBP-E and the cleaved E protein in fractions 2-5. White box highlights the ladder-like pattern that suggests interaction between the cleaved and uncleaved E protein. (D) SDS-PAGE showing RP-HPLC fraction of E protein with C40A/C43A/C44A mutations, which improved the purification result.

Since the MBP-E fusion protein strategy was successful, albeit with low yield, we tried the same strategy to express and purify E protein from other CoVs, i.e. MHV E, TGEV E, IBV E, and MERS-CoV E. Despite having very low sequence similarity (~20%), all CoV E proteins are predicted to share similar secondary structure (Fig. 4-

1. Among these, only the MHV E and MERS-CoV E could be expressed and purified, but their yield was even lower than for the SARS-CoV E.

To improve the yield and solubility of E protein samples, we screened the expression of truncated MHV E, TGEV E, IBV E, and SARS-CoV E protein constructs at the NTU/SBS Protein Production Platform. Screening results showed two constructs that were expressed with sufficient yield (Fig. S1, appendix): a truncated SARS-CoV E construct (8-65, denoted E_{TR}) and a full-length IBV E construct. These two constructs could be expressed and purified with sufficient yield to prepare solution NMR samples. The present study will describe only the E_{TR} construct, whereas the IBV E construct has been relegated to future studies.

4.3 Validation of the E_{TR} construct

The use of truncated constructs for structural studies raises several questions. For example, does the truncation cause significant structural changes? In that case, is the structure obtained still relevant? These questions were addressed through comparison of secondary structures, examination of the oligomeric state, and channel conductivity measurements on the E_{TR} construct.

4.3.1 Secondary structure of E_{TR}

Firstly, we compared the secondary structure of SARS-CoV E_{TR} and E_{FL} when reconstituted in lipid bilayers. E_{FL} was previously shown to form mainly α -helix after reconstitution into lipid bilayers or detergent micelles (Parthasarathy et al., 2012). Herein, we have reproduced these results using expressed E_{FL}. The amide I bands in the FTIR spectra of E_{TR} and E_{FL} are almost superimposable (Fig. 4-4A, solid lines), indicating that the secondary structure composition of the two proteins are identical. Their amide I bands are centred at 1655 cm⁻¹, a characteristic of α -helical conformation. Fourier self-deconvolution of the amide I band (Fig. 4-4A, dotted lines) indicate that the two proteins are indeed mostly α -helical, and the rest of their secondary structure composition are identical.

When reconstituted in DPC micelles, E_{TR} is also mainly α -helical, as indicated by two minima at 208 and 222 nm in its CD spectra (Fig. 4-4B). The 190 nm maxima could

not be observed due to excessive absorbance. Thus, both FTIR and CD spectra indicate that the E_{TR} construct is mainly α -helical and globally it has a similar secondary structure composition as the full-length protein.

Figure 4-4. Secondary structure comparison between E_{TR} and E_{FL} . (A) Amide I band of FTIR spectra of E_{FL} (blue) and E_{TR} (red) reconstituted in DMPC lipid bilayers at 1:50 peptide-to-lipid molar ratio (solid lines). Dotted lines show the Fourier-self deconvolution of the amide I band to estimate the proportion of secondary structure-specific sub-bands that constitute the band. (B) CD spectra of E_{TR} reconstituted in 100 mM DPC, 20 mM Na-PO₄, 50 mM NaCl. Spectral region below 195 nm is excluded due to excessive absorbance.

4.3.2 Oligomeric state of E_{TR} construct

Next, we queried the oligomeric state of the E_{TR} construct. E_{FL} , as well as the TM domain of E protein (E_{TM}), has repeatedly been shown to form pentamers (Parthasarathy et al., 2012; Verdiá-Báguena et al., 2012). Because the TM domain alone is sufficient to drive pentamerization, it is highly likely that E_{TR} also forms pentamers. However, as we have speculated that the extramembrane domain may influence the structure of the rest of the protein, it becomes necessary to confirm the oligomeric state of E_{TR} .

Firstly, we examined the oligomeric state of E_{TR} by using gel electrophoresis. In the presence of SDS, E_{TR} migrates as a single band at the expected molecular weight of a monomer (9 kDa, Fig. 4-5A), whereas in the presence of perfluoro-octanoic acid (PFO), a relatively milder detergent that supports oligomerization of transmembrane domains, E_{TR} shows a single band at a molecular weight consistent with a pentamer

(45 kDa, Fig. 4-5B). These are consistent with previous observations that E_{FL} is monomeric in SDS, and pentameric in PFO detergent (Parthasarathy et al., 2012).

Figure 4-5. Oligomeric state of E_{TR} in SDS and PFO detergent. (A) SDS-NuPAGE and (B) PFO-NuPAGE of E_{TR} protein. PFO-NuPAGE was performed essentially the same way as SDS-NuPAGE (Invitrogen) by substituting SDS with PFO detergent in the sample buffer and running buffer. Number of asterisks (*) for monomer indicates the estimated oligomeric state of the E_{TR} bands, based on comparison with molecular weight ladder.

Next, we examined the oligomeric state of E_{TR} by using sedimentation equilibrium analytical ultracentrifugation. When reconstituted in C14SB micelles, E_{TR} forms a radial distribution profile that fits best to a monomer-pentamer model (Fig. 4-6A) with log Ka = 16. This is slightly weaker than the value obtained for E_{FL} (log Ka = 17.2) (Parthasarathy et al., 2012) and E_{TM} (log Ka = 17.1) (Parthasarathy et al., 2008). This suggests that the missing residues (due to truncation) have negligible participation in the pentamerization of E protein. Nevertheless, the comparison of global-reduced chi-square values from other oligomeric models (Fig. 4-6B) shows that the pentameric model gives the best fit.

Thus, both gel electrophoresis and sedimentation equilibrium data indicate that the E_{TR} construct oligomerizes into pentamers in mild detergents such as C14SB and PFO.

4.3.3 Channel activity of E_{TR} construct

E protein allows passage of ion through the channel formed by its pentameric bundle. The channel activity of synthetic E_{TM} and expressed E_{FL} has previously been demonstrated (Parthasarathy et al., 2012; Verdiá-Báguena et al., 2012; Wilson et al., 2004) and this ion channel activity can be inhibited by HMA (Wilson et al., 2006). Because the channel pore is formed by the TM domain, and because we have seen in the previous section that E_{TR} is also pentameric, we can reasonably expect that E_{TR} also conducts ion. To confirm this, we have measured the conductance of E_{TR} in lipid bilayers.

Figure 4-6. Sedimentation equilibrium analysis of E_{TR} in C14SB detergent. The oligomeric state of E_{TR} in 5 mM C14SB was examined by analytical ultracentrifugation. (A) Sedimentation equilibrium data of E_{TR} at three different speeds (circles, colour-coded to the centrifugation speed) fitted to a monomer-pentamer self-equilibrium model (solid lines). The fitting residuals are plotted in smaller panels underneath each fit. (B) The global reduced chi-square values from fitting to different monomer-N-mer models, indicating the fit quality of each model.

Figure 4-7. Ion channel activity measurement of E_{TR} in lipid bilayers. (A) Selected current recordings of E_{TR} in DPhPC bilayer with symmetrical buffer (0.5 M NaCl, 10 mM HEPES pH 4) from -60 mV to 60 mV holding potential. (B) The I/V plot of the average current measured at each holding potential. The channel conductance is calculated from the slope of the linear regression of the I/V plot. (C) Current recording at 60 mV showing the effect of addition of 10 μ M HMA at the indicated time point, after which the current measured across the membrane drops to zero.

We reconstituted E_{TR} into DPhPC lipid bilayers and measured its conductance in 0.5 M NaCl, pH 4.0 symmetrical condition (Fig. 4-7A). The conductance was calculated

from the slope of the I/V plot to be 390 ± 20 pS (Fig. 4-7B). This is significantly higher than previously reported for synthetic E_{FL} (190 ± 60 pS) and E_{TM} (180 ± 120 pS), possibly due to different pH. The ion channel activity was completely inhibited when 10 μ M HMA was added (Fig. 4-7C). These behaviours are identical to that of E_{FL} and E_{TM} , and thus confirm that E_{TR} can form a functional ion channel when reconstituted into lipid bilayers. This demonstrates the relevance of the E_{TR} construct from a functional point of view.

In summary, we have confirmed that the truncations in E_{TR} do not significantly affect its secondary structure, oligomeric state, and ion channel activity. Thus, we concluded that the E_{TR} construct is a suitable construct to study the structure of CoV E protein.

4.4 Validation of the micellar reconstitution environment

Having confirmed that E_{TR} is a suitable construct, we collected [^1H - ^{15}N]-HSQC spectra of E_{TR} in various detergents and bicelles to find a suitable reconstitution environment for solution NMR experiments. Among various detergents tested, including bicelles, only SDS resulted in excellent, workable spectral quality, in terms of both peak number and dispersion. LMPG, a lipid-like detergent, give a moderate spectral quality, but still insufficient for backbone and sidechain assignment.

DPC is very often used for solution NMR study of membrane proteins and is considered a good membrane-mimic environment, especially when compared to SDS. Although DPC alone did not produce a good spectrum, we found that DPC could be added into SDS up to a 4:1 DPC-to-SDS molar ratio without resulting in significant spectral deterioration.

We then examined the suitability of the micelle system to obtain the E_{TR} structure. In this section we will discuss the influence of DPC/SDS and LMPG to the secondary structure and the oligomeric state of E_{TR} . In addition, as an alternative to channel activity, which cannot be tested in detergent micelles, we determined the binding of the E_{TR} ion channel inhibitor HMA.

4.4.1 Secondary structure of E_{TR} in detergent micelles

Firstly, we collected CD spectra of E_{TR} in SDS, DPC/SDS (2:1 molar ratio) micelles, and DMPC/DHPC bicelles. All three spectra are almost identical and superimposable to the CD spectra of E_{TR} in DPC collected previously, with minima at 209 and 222 nm characteristic of an α -helix (Fig. 4-8A). This indicates that the secondary structure composition of E_{TR} do not vary significantly when it is reconstituted in bicelles or in detergent micelles.

Next, after collecting 3D-HNCA spectra of E_{TR} in SDS, DPC/SDS (4:1 molar ratio), and LMPG, as well as E_{FL} in SDS, we compared their secondary $^{13}\text{C}\alpha$ chemical shifts, i.e. the difference from tabulated $^{13}\text{C}\alpha$ chemical shifts for random coil, which are highly correlated with secondary structure (Spera and Bax, 1991; Xu and Case, 2001). Specifically, α -helical structures would show higher $^{13}\text{C}\alpha$ chemical shifts than random coil, whereas β -strands would show lower.

The $^{13}\text{C}\alpha$ values of the E protein constructs in these micelles are generally higher than that of a random coil, indicating an overall α -helical structure (Fig. 4-8B). In particular, residues with positive secondary $^{13}\text{C}\alpha$ values are clustered in two regions, i.e. at residues 15-35 and 55-63, indicating the presence of two separate α -helices connected by a flexible linker. The first helix is likely to be the TM domain, considering its position which coincides with TMHMM (Krogh et al., 2001) prediction (coloured bars, Fig. 4-8B), and its length (~20 residues) which is suitable for a membrane-spanning helix. On the other hand, the second helix is too short to be a second TM domain, and its presence is quite unexpected, considering the consensus secondary structure prediction. We note that residues 66-76 of E_{FL}, which are not present in E_{TR}, show minimal $^{13}\text{C}\alpha$ deviation from a random coil consistent with the secondary structure prediction of that segment and the FTIR data of a synthetic peptide corresponding to residues 59-76 (Surya et al., 2013).

The $^{13}\text{C}\alpha$ shifts of E_{TR} reconstituted in SDS are almost identical to their counterparts in E_{FL} (Fig. 4-8B) indicating that in this detergent, E_{TR} adopts the same conformation as the full-length E protein. The $^{13}\text{C}\alpha$ values of E_{TR} in DPC/SDS and LMPG are also identical to E_{TR} in SDS (Fig. 4-8B), especially in residues 17-35 i.e. the first helix, indicating that the backbone structure of E_{TR} do not vary significantly in these three micellar systems.

Some minor differences still can be observed between the three detergent micelles. For example, there is a slight increase in helicity (higher $^{13}\text{C}\alpha$ values) at residues 8-16 of E_{TR} reconstituted in DPC/SDS and LMPG, suggesting that the first helix may extend longer towards the N-terminus in these systems. The helicity also increased significantly in residues 52-57 of E_{TR} reconstituted in LMPG, indicating that the second helix is more extended in LMPG.

Taken together, the comparison of CD spectra and $^{13}\text{C}\alpha$ values suggest that the backbone structure of E_{TR} in DPC/SDS and LMPG do not vary significantly with E_{FL} under the conditions tested herein. Therefore, the structure of E_{TR} obtained in DPC/SDS and LMPG is likely to be representative of the structure of the full-length E protein.

Figure 4-8. Secondary structure comparison of E_{TR} in detergent micelles. (A) Comparison between CD spectra of E_{TR} reconstituted in various detergent micelles (100 mM DPC, 25 mM SDS, 100 mM DPC/25 mM SDS) and in DMPC/DHPC bicelles ($q = 0.3$). The spectra were truncated to 195 nm due to excessive absorbance; in bicelles, region with excessive absorbance extends up to 205 nm. (B) Comparison between secondary $^{13}\text{C}\alpha$ values (the deviation of $^{13}\text{C}\alpha$ chemical shifts from tabulated random coil values) of E protein constructs reconstituted in various detergent micelles (50 mM SDS, 200 mM DPC/50 mM SDS, and 200 mM LMPG). The consensus secondary structure prediction of SARS-CoV E (coloured bars) is reproduced from Fig. 4.1: red indicate β -strand, yellow indicate random coil, and blue indicate α -helix.

4.4.2 Oligomeric state of E_{TR} in detergent micelles

Next, we examined the oligomeric state of E_{TR} by gel electrophoresis and sedimentation equilibrium experiments. When reconstituted in DPC/SDS mixtures

with different molar ratios, E_{TR} migrates as several bands: at 1:1 and 2:1 DPC/SDS molar ratios, dimer and trimer bands were observed, whereas at 4:1 molar ratio tetramer and pentamers could be observed (Fig. 4-9A). Thus, the addition of DPC to SDS results in E_{TR} oligomerization, and with increasing DPC-to-SDS ratio the size of the oligomer increases. However, at any DPC-to-SDS ratio more than one oligomeric state could be observed. When reconstituted in LMPG, E_{TR} mostly form pentamers which break down into smaller oligomers when the LMPG-to-peptide molar ratio is increased (Fig.4-9B).

Figure 4-9. Gel electrophoresis of E_{TR} in detergent micelle. (A) BN-PAGE of E_{TR} in DPC/SDS mixture at different DPC-to-SDS molar ratios as indicated. (B) BN-PAGE of E_{TR} in LMPG at different peptide-to-LMPG molar ratios as indicated. Asterisks indicate the positions of individual bands. Molecular weight ladder do not necessarily represent the actual molecular weight of the protein due to anomalous migration effect from detergents bound to proteins, as shown by AqpZ protein in SDS micelles which show monomer (~24 kDa) and tetramer bands (~95 kDa).

Figure 4-10. Sedimentation equilibrium analysis of E_{TR} in DPC/SDS mixed micelles. (A & B) Global reduced chi-square values from fitting different oligomeric models (X-axis indicates the number of subunits) to the sedimentation equilibrium data of E_{TR} in DPC/SDS micelles (4:1 and 2:1 molar ratios, 50 mM SDS constant). (C) The sedimentation equilibrium profile of E_{TR} in DPC/SDS micelles (4:1 molar ratio, 50 mM SDS constant) at different speeds (colour-coded circles) fitted to monomer-pentamer model (black lines). The fitting residuals are shown in smaller panels underneath each fit.

Consistent with the electrophoresis data, sedimentation equilibrium data of E_{TR} in 2:1 and 4:1 DPC/SDS molar ratios could not be reliably fitted to a single oligomeric

model. Comparison of global reduced chi-square values from different oligomeric models shows that all models larger than a tetramer fits the data almost equally well (Fig. 4-10A & 4-10B). Fitting to a monomer-pentamer model, for example, (Fig. 4-10C) is qualitatively good as judged from the fitting residuals. However, the same can also be said for the monomer-tetramer and monomer-hexamer models (not shown) and thus cannot be excluded. Therefore, it is likely that E_{TR} forms a mixture of several oligomeric states in DPC/SDS.

Together with the electrophoresis data, these indicate that when reconstituted in DPC/SDS, E_{TR} pentamers become destabilized and other, possibly non-specific, oligomeric forms can be observed. Unfortunately the oligomeric state of E_{TR} in LMPG cannot be studied by sedimentation equilibrium due to technical reasons, which are explained in Appendix 8.2.

4.4.3 HMA binding on E_{TR} in detergent micelles

In section 4.3.3, we have confirmed the pentamerization of E_{TR} in lipid bilayer by ion channel activity measurement. This cannot be done in micelles, which does not form partitions, so we used an indirect means by measuring the binding of the ion channel inhibitor HMA to E_{TR}. We expect that HMA will not bind when E_{TR} does not form the ion-conducting pentameric channel, e.g. when E_{TR} is reconstituted in SDS and is monomeric.

To determine whether HMA binds to E_{TR}, we titrated HMA to E_{TR} reconstituted in SDS, DPC/SDS, and LMPG, and measured the chemical shift perturbation (CSP) values from [¹H-¹⁵N]-HSQC spectra (Fig. 4-11). For E_{TR} reconstituted in SDS, minimal to no perturbation were observed, with only 0.006 ppm average CSP value even at 10:1 HMA/E_{TR} molar ratio, i.e. 1 mM HMA (Fig. 4-11, blue), suggesting no significant interaction between HMA and E_{TR}. On the other hand, in DPC/SDS the average CSP value increases to 0.013 ppm at 2:1 HMA/E_{TR} molar ratio, i.e. 0.2 mM HMA (Fig. 4-11, red). The most affected residues are clustered at the ends of the TM domain, strongly suggesting HMA binding to E_{TR}. Specifically, the affected residues at the N-terminal end of the TM domain are E8, G10, T11, V14, N15, and S16, and at the C-terminal end is L37. There are also some affected residues at the C-terminus, but these are possibly an artefact due to the inherent flexibility of the terminal residues.

For E_{TR} reconstituted in LMPG, the average CSP value increases further to 0.018 ppm (Fig. 4-11, green). Strongly affected residues are clustered near the ends of the TM domain: T9, L12, I13, and A36. Interestingly, V47 at the extramembrane domain is also affected, despite being further away from the TM domain. This indicates either the presence of an external HMA binding site, i.e. not in the TM domain, or a conformational change that affects the extramembrane domain. This also suggests that the structure of E_{TR} pentamer in LMPG is slightly different than in DPC/SDS, at least at the extramembrane domain.

Figure 4-11. Identification of HMA-binding residues in micelle-reconstituted E_{TR} . HMA-binding residues were identified from significant chemical shift perturbation values upon addition of HMA to 0.1 mM E_{TR} in 200 mM SDS (1 mM HMA, blue), 200 mM DPC/50 mM SDS (0.2 mM HMA, red), and 200 mM LMPG (0.2 mM HMA, green). Grey box indicate the approximate location of the TM domain.

Although details differ slightly, these titration data are consistent with previous observation in E_{TM} reconstituted in DPC micelle where N15, S16, and R38 were identified as HMA-binding residues (Pervushin et al., 2009). Thus, there are at least

two HMA binding pockets: at the N-terminal end of the TM around N15, and at the C-terminal end of the TM around residues 36-38. Differences in affected residues are probably due to different construct length and reconstitution environment.

Combined observation thus suggests that the structure and oligomeric form of E_{TR} in DPC/SDS and LMPG are not significantly affected by the detergent micelles. However, in addition to the pentameric form, other minor oligomeric forms potentially exist as well. Thus, the E_{TR} structure obtained in these detergents should still be a valid model for E protein structure. Due to the closer similarity between LMPG and phospholipids (as compared to DPC/SDS), it seems reasonable to treat the structure obtained in DPC/SDS as the monomeric structure. The pentameric model will be built from the structure obtained in LMPG.

4.5 Solution NMR structure of E protein monomer

A series of 2D and 3D spectra of E_{TR} reconstituted in DPC/SDS and in LMPG micelles were collected for backbone and sidechain resonance assignments (the [¹H-¹⁵N]-TROSY-HSQC spectra are shown in Fig. 4-12). NOE distance restraints were obtained from ¹⁵N-NOESY-HSQC and ¹³C-NOESY-HSQC spectra. A total of 15 structures have been calculated and the structure statistics are summarized in Table 4-1.

4.5.1 Overall structure

E_{TR} reconstituted in DPC/SDS forms an L-shaped structure consisting of two α -helices connected by a random coil linker (Fig. 4-13A). The first helix spans residue 13-42, with a slight bend around residues 26-30. This bend is mediated by a tentative hydrogen bond between T30 H γ and F26 carbonyl oxygen, which are only 2.0 Å apart. The second helix starts at the conserved P54 and extends to the C-terminus.

The structure of E_{TR} reconstituted in LMPG (Fig. 4-13B) is partly similar to that in DPC/SDS. Superimposition of the structures obtained in these two systems are shown in Fig. 4-13C, and a schematics comparing the secondary structures is shown in Fig. 4-13D. In the two detergents, E_{TR} form an L-shaped structure with α -helix at both N- and C-terminus. In LMPG, the first helix is slightly shorter and ended at L37, whereas

Table 4-1. Structure statistics for the selected 15 structures of E_{TR} protein in DPC/SDS

NMR restraints	
Total unambiguous distance restraints	2258
Intra residual	1334
Sequential ($ i - j = 1$)	230
Short-range ($ i - j \leq 1$)	1564
Medium ($2 \leq i - j \leq 4$)	321
Long range ($ i - j \geq 5$)	11
Dihedral angle restraints	85
Hydrogen bond restraints ^a	20
RDC restraints	88
PRE restraints	
RMSD from the experimental residual dipolar couplings (Hz)	
¹ D _{NH}	0.44±0.03
RMSD from the average atomic coordinates (residues 12-63, Å) ^b	
Backbone atoms	0.27 ± 0.11
All heavy atoms	0.70 ± 0.13
Ramachandran analysis (%)	
Residues in most favored regions	87.5
Residues in additional allowed regions	12.5
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0

^aBackbone hydrogen bonds of α -helix are applied for regions confirmed to be α -helix according to local NOE pattern and H^N-H₂O chemical exchange experiment.

^bStatistics are calculated and averaged over an ensemble of the 15 structures with lowest target function from CYANA.

that the helix extends outside the micelle. Nevertheless, this indicates that the N-terminal helix of E_{TR} is embedded in micelles, and therefore it is the TM domain of E protein. This also indicates that there is only one TM domain in E protein.

These findings were further confirmed through PRE measurement with 5-DSA and 16-DSA: fatty acids with paramagnetic labels located near the surface of the membrane and at the centre/interior of the membrane, respectively. Unfortunately, PRE measurement for E_{TR} in LMPG micelles did not give meaningful result, and only data from E_{TR} in DPC/SDS are available.

Upon addition of 5-DSA into DPC/SDS-reconstituted E_{TR}, residues 11-20 and 40-55 show reduction in peak intensity (Fig. 4-14C). This indicates that the TM domain starts at least at F20 and ends at A40, slightly longer than the H/D exchange data. Slight periodicity in residues 15-20 and 40-55 suggests that these residues might be located at or close to the surface of the micelle, i.e. these are the juxtamembrane residues.

Addition of 16-DSA causes intensity reduction in residues 19-40 (Fig. 4-14C), indicating that these residues are fully embedded inside the micelle, in agreement with the H/D exchange and 5-DSA titration data. The rest of the C-terminal tail shows periodic intensity reduction, further suggesting that the C-terminal tail is close to the micelle surface. This periodicity also suggests the orientation of the C-terminal helix relative to the micelle surface. Residues V52, T55, Y59, and K63, which have the lowest intensity, possibly are oriented towards the micelle.

Thus, the H/D exchange and PRE data indicate that the N-terminal helix of E_{TR} is the TM domain of E protein, and that the C-terminal helix is located at the micelle surface.

Figure 4-14. Determination of membrane-embedded region in E_{TR} . [^1H - ^{15}N]-HSQC spectra of E_{TR} reconstituted in (A) DPC/SDS and (B) LMPG in the presence of 99% D_2O . Only cross-peaks of residues protected from H/D exchange appears in the spectra. (C) Relative cross-peaks intensity of E_{TR} reconstituted in DPC/SDS after addition of 1 mM 5-DSA (black) or 16-DSA (red). Residues with lower relative intensity indicate their close proximity to the paramagnetic groups: 5-DSA is at the micelle surface, whereas 16-DSA is at the micelle interior.

4.5.3 The structure of the extramembrane domain

Consensus secondary structure prediction suggests that the presence of a β -turn- β motif centred at the conserved P54 residue. This is supported by the structure of a synthetic peptide of SARS-CoV E residues 46-60, which is completely β -strand in lipid bilayers. However, FTIR amide I band analysis of both E_{TR} and E_{FL} (Fig. 4-4) indicate a much lower proportion of β -structure than what was expected (~20% by

prediction). The solution NMR structure of E_{TR} in DPC/SDS and LMPG (Fig. 4-14) also do not show the presence of any β -strands. Instead of a β -turn- β , the residues form a random coil prior to P54 and an α -helix after P54.

This discrepancy may be explained by proposing that only a minor population of E protein adopts the β -turn- β conformation. To test this, we expressed and purified E_{FL} constructs with mutations in the predicted β -turn- β motif: the P54A mutation (in E_{P54A} construct) and V56A/Y57A/V58A/Y59A mutations (“4ALA” mutations, in E_{4ALA} construct). Both mutations have been shown to abolish the ability of SARS-CoV E C-terminal tail to direct the Golgi localization of a membrane protein through removal of the conserved proline or disruption of the β -turn- β motif (Cohen et al., 2011). Whereas P54A mutation does not change the secondary structure prediction of E protein, the 4ALA mutations completely change the prediction of this region from β -strand to α -helix (Fig. 4-15A).

Consistently, FTIR spectra collected in lipid bilayers show that the amide I band of E_{P54A} is identical to that of E_{FL}, but the amide I band of E_{4ALA} show reductions at 1635 cm⁻¹ and 1690 cm⁻¹ (Fig. 4-15A). These side-bands are indicative of β -strands (Tamm and Tatulian, 1997), suggesting that there is indeed a proportion of β -strands in the E protein population, albeit very low, which further diminish in E_{4ALA} when some of the bulky residues in the β -turn- β motif is mutated.

Based on these observations, it is likely that the residues around P54 undergo conformational exchange between β -turn- β and α -helix, with the α -helix form being the more populated conformation (Fig. 4-15B). The β -turn- β motif is present only in a minor population of E protein, thus explaining the lack of β -strand contribution to the amide I band of E protein in lipid bilayers. In both DPC/SDS and LMPG micelles, the proportion of β -turn- β motif must also be very low, as only the α -helix form is observed in the solution NMR structure of E_{TR} in these micelles.

Figure 4-15. Secondary structure of the E protein extramembrane domain. (A) Amide I FTIR spectra of E_{FL}, E_{P54A}, and E_{4ALA} reconstituted in DMPC lipid bilayers. The arrows indicate regions that decreased in intensity upon mutating some residues adjacent to P54 into alanines (E_{4ALA}). (B) Structural models illustrating the conformational exchange proposed to take place in the residues surrounding P54. The α -helical form is the more populated conformation, which was observed in the solution NMR structure of E_{TR} in DPC/SDS and in LMPG, as compared to the β -turn- β motif.

4.6 Solution NMR model of the E protein pentamer

The pentameric model of E protein presented herein was derived from the solution structure of E_{TR} in LMPG. Inter-monomeric distance restraints were obtained from comparison between ¹⁵N-NOESY-HSQC spectra of ¹⁵N/D-labelled E_{TR} (ND) and of a mixture of ¹⁵N/D- and ¹³C/H-labelled E_{TR} (ND/C). ND/C spectra theoretically should show only inter-monomeric NOEs due to ¹H and ¹⁵N being present in physically separate monomers, i.e. ¹H in the ¹⁵N/D-labelled sample is replaced by D. However,

intra-monomeric NOEs can also show up in the ND/C spectra due to incomplete deuteration of the ND sample, thus the ND spectra is collected as a control.

Figure 4-16. Inter-monomeric NOEs between E_{TR} monomers in LMPG. List of inter-monomeric NOE contacts found by comparing ^{15}N -NOESY-HSQC spectra (250 ms mixing time) of $^{15}\text{N}/\text{D}$ -labelled E_{TR} (ND) and of a mixture of $^{15}\text{N}/\text{D}$ - and $^{13}\text{C}/\text{H}$ -labelled E_{TR} (ND/C) reconstituted in 200 mM LMPG. Selected strips show the first two NOE contacts whereas the rest are shown in Fig. S4. Peaks that appear in ND/C strip but not in ND strip are inter-monomeric NOEs, whereas those showing in both ND/C and ND strips are intra-monomeric NOEs. A strip from $^{15}\text{N}/^{13}\text{C}$ -labelled E_{TR} sample (NC) is also shown for reference.

From these spectra, 10 inter-monomeric NOEs were obtained; 8 are located at the TM domain and 2 at the C-terminal extramembrane domain (Fig. 4-16 and Fig. S5). These were used as distance restraints for docking five E_{TR} subunits into a pentamer by using HADDOCK software (De Vries et al., 2007; Dominguez et al., 2003). Following standard docking procedure, 200 final structures were generated and clustered with a $7.5\text{\AA}$ RMSD cut-off. The highest-ranked cluster was chosen as the pentameric model of E protein (Fig. 4-17).

Figure 4-17. Structural model of E protein pentamer. (A) Top and (B) side view of E protein pentamer obtained from docking five monomeric models of E_{Tr} in LMPG into a pentamer. Docking was performed by using HADDOCK software (De Vries et al., 2007; Dominguez et al., 2003), incorporating 10 inter-monomeric NOE restraints: 8 in the TM domain and 2 in the extramembrane domain. The orientation of residues N15 and V25 are shown in slices through the pentamer (inset); the residues themselves are in stick representation.

In this model, the E_{TR} subunits are arranged into a left-handed pentameric bundle with each C-terminal extramembrane tail extending towards the neighbouring subunit. The orientation of the TM bundle is roughly consistent with the pentamer B model, as shown by the bundle cross-sections at residues N15 and V25 (Fig. 4-17C). The C-terminal tail seems to lie along the surface of the detergent micelle, following its curvature. In a lipid bilayer, which is more planar, the C-terminal tail would probably also lie along the surface, resulting in an almost perpendicular orientation relative to the TM bundle.

4.7 Summary

In this chapter I have presented the monomeric and pentameric models of SARS-CoV E protein derived from the solution NMR structure of E_{TR} in DPC/SDS and in LMPG micelles, respectively. The validity of these models, which were built using a truncated protein in detergent micelles, was examined from a structural point of view by comparison of secondary structure and oligomeric form, and from a functional point of view by ion channel activity measurement and HMA (ion channel inhibitor) binding site determination.

Specifically, in these models, the E protein structure consists of a N-terminal helical TM domain, and a C-terminal extramembrane helix which is probably peripheral to the micelle surface. The structure of the C-terminus around the conserved P54 residue is a random coil and an α -helix, with a minor population forming a β -turn- β .

The pentameric model of E protein consists of a left-handed bundle consistent with the previously proposed “pentamer B” model derived from simulation of rigid helices based on sequence conservation data of E proteins.

DISCUSSIONS

5.1 Juxtamembrane structure of α L and heterodimer interface of α L/ β 2

In chapter 3, I have presented a structural model for the integrin α L and β 2 TM domains and their heterodimer interface. The model presented herein is very similar to that of α IIb/ β 3 TM domain heterodimer in bicelles published previously (Lau et al., 2009), in terms of the structure of the juxtamembrane GFFKR motif, as well as the TM-TM interface of the heterodimer.

5.1.1 Alternate structure for the juxtamembrane GFFKR motif of α subunit

The GFFKR motif is highly conserved in integrin α TM domain, especially the GFF part which is completely conserved among all α subunits. Deletions and mutations of this motif in α IIb/ β 3 prevents its heterodimerization, and results in constitutive activation (Luo et al., 2004; O'Toole et al., 1994).

In α L, this motif undergoes backbone reversal (reverse turn) in bicelles, identical to that observed in α IIb. This is a very unusual conformation and invited a lot of doubt when it was first observed in α IIb TM monomers reconstituted in bicelles (Lau et al., 2008), and later in α IIb/ β 3 heterodimers in bicelles (Lau et al., 2009). Rosetta docking with distance restraints obtained from disulfide scanning experiment also found the reverse turn in the lowest energy structure (Zhu et al., 2009).

In the absence of lipids, the GFFKR motif of α 1/ β 1 in LDAO (lauryl dimethylamine oxide) micelles was found to adopt a helical conformation (Lai et al., 2013). Similarly, α IIb/ β 3 in 50% ACN also had a helical GFFKR motif (Yang et al., 2009), although CS-Rosetta docking implementing backbone restraints obtained therein produced a structure with reverse turn as the lowest energy conformer, with a minor population having the helical GFFKR motif. This suggests that the two forms, i.e. the helical form and the reverse turn, may be in a conformational equilibrium.

It is possible that the helical form and the reverse turn correspond to the different activation state of integrin, although which conformation corresponds to which state is

not immediately clear. One argument proposed that the helical form represents the heterodimer/resting states as it seems to have better accessibility to integrin α binding partners (Yang et al., 2009). This seems plausible because the reverse turn could already be observed in α IIb (Lau et al., 2008) and α L (present study) monomers, in the absence of their β partners, which indicates that in the dissociated state the GFFKR motif adopts the reverse turn conformation.

On the other hand, the reverse turn has so far been observed only in bicelles, where the α/β heterodimer population is the highest (Suk et al., 2012). In detergents and organic solvents, the heterodimer population is relatively low compared to population of monomers and homomultimers. Thus from this point of view, the helical form observed in detergents and organic solvents may correspond to the dissociated/active state, whereas the reverse turn corresponds to the heterodimer/resting state.

Nevertheless, this controversy does not seem resolvable with the current state of knowledge. We note that in both α IIb/ β 3 (Lau et al., 2009) and α L/ β 2 (present study) the cytoplasmic tail has been truncated. Whether or not this affects the structure of the juxtamembrane region is unknown, and may be worth investigating.

5.1.2 Heterodimerization of α L/ β 2: intermediate exchange

Cross-titration experiments revealed a series of residues whose backbone resonances were affected by heterodimerization. The position of these residues in the TM indicate an α L/ β 2 TM orientation in which the small residues of the GxxxG-like motif in α L TM domain faces the β 2 TM domain, but not vice versa. Among these interfacial residues, there is the S1071 of α L, which we proposed to form polar interaction with T686 of β 2 (Vararattanavech et al., 2010), but this information is still insufficient to determine whether the two residues form a polar interaction. More information on the heterodimer interface is necessary, e.g. by monitoring the cross-titration experiment with [^1H - ^{13}C] HSQC and secondary ^{13}C chemical shift, and in the form of inter-monomeric NOE distance restraints.

The α L/ β 2 interfacial residues were identified by the attenuation of their cross-peaks in the [^1H - ^{15}N] HSQC spectra, which indicates an intermediate exchange rate of heterodimerization. Intermediate exchange was also observed for α 1/ β 1 TM heterodimer in LDAO micelles (Lai et al., 2013), but α IIb/ β 3 exhibited slow exchange

rate in DHPC/POPC bicelles, as a new set of resonances appeared during cross-titration experiment (Lau et al., 2009). The exchange rate shifted from slow to intermediate for bicelles in which less $\alpha\text{IIb}/\beta 3$ heterodimer population was observed, e.g. in CHAPS/DMPG bicelles where severe peak attenuation was observed (Suk et al., 2012).

Unfortunately the relative $\alpha\text{IIb}/\beta 3$ heterodimer proportion in DHPC/DMPC bicelles, used in the present study of $\alpha\text{L}/\beta 2$, was not determined in that study. Regardless, it is likely that $\alpha\text{L}/\beta 2$ has a different bicelle mixture preference than $\alpha\text{IIb}/\beta 3$, considering that the two integrins have different tissue distribution (the former in leukocytes, the latter in platelets), and thus possibly different native lipid environment as well. Yet, it seems plausible that the intermediate exchange rate observed for $\alpha\text{L}/\beta 2$ is due to the reconstitution environment being suboptimal for heterodimerization. A different bicelle composition, which needs to be screened, may shift the exchange rate from intermediate to slow. Slow exchange rate has been linked to higher heterodimer proportion (Suk et al., 2012), and this would certainly be useful for obtaining inter-monomeric NOE restraints to build a more detailed model of the $\alpha\text{L}/\beta 2$ heterodimer.

5.2 Structural models for CoV E protein monomers and pentamers

In chapter 4, I have presented the structural models for SARS-CoV E protein monomers and pentamers. These were based on solution NMR data of E_{TR} in DPC/SDS and in LMPG micelles, respectively. Briefly, E_{TR} monomers consist of an L-shaped structure consisting of an N-terminal TM helix and a short C-terminal helix, connected by a flexible linker which in LMPG forms yet another short helix. In the pentamer, the E_{TR} monomers are arranged into a left-handed bundle formed mainly by its TM domain, with its C-terminal tail wrapping around the bundle.

5.2.1 The structural flexibility of the C-terminal hydrophobic region

The C-terminal hydrophobic region of CoV E proteins has a propensity to form β -hairpin centred on its fully conserved proline residue, as suggested by the consensus secondary structure prediction (Fig. 4-1). Herein we observed that this region forms different structures when studied in isolation or together with the rest of the protein. A synthetic peptide corresponding to residues 46-60 forms 100% β -strand in the

presence of lipid bilayers (Surya et al., 2013). However, in the context of E_{TR} structure determined in the present study, this region consists of a coil and a helix. The similarity between the FTIR spectra of E_{TR} and E_{FL} in lipid bilayers, and the low contribution of β -strands in their amide I bands indicate that the helical form is not due to the micellar environment used for solution NMR study of E_{TR}, or due to the truncation of residues 66-76 from E protein. The presence of the rest of the E protein seems to influence the folding of the C-terminal hydrophobic region, suggesting proximity and contacts between this region and the rest of the protein.

One possibility is that the orientation of the C-terminal hydrophobic region becomes restricted when it is connected to the TM domain, and thus affects its structure. For example, the juxtamembrane domain of rat epidermal growth factor receptor Neu shows different dynamics and membrane interaction profile when studied in isolation or connected to the TM domain (Matsushita et al., 2013).

If the Val and Tyr residues in the C-terminal hydrophobic region are mutated into Ala, the secondary structure prediction of this region changes to α -helix (Cohen et al., 2011). E protein with these mutations shows a noticeable reduction in β -strands contribution to FTIR amide I bands. This indicates that a small population of E protein may indeed have a β -hairpin in its C-terminal hydrophobic region. Thus the two forms (the α -helix and β -hairpin) may coexist in equilibrium, with the α -helix being the lower energy conformation.

Interestingly, the C-terminal hydrophobic region of E protein has been demonstrated to direct Golgi accumulation of a membrane protein, specifically VSV-G fused to C-terminal tail of SARS-CoV E (Cohen et al., 2011). When the abovementioned mutations were introduced, the Golgi localization is no longer observed and the protein is directed to the plasma membrane. A similar effect can also be observed when P54A mutation was introduced, even though this mutation does not affect the consensus secondary structure prediction. The amide I bands of E_{FL} with P54A mutation is also identical to the wild-type, suggesting that the proportion of E protein with β -hairpin is unaffected. Therefore, the β -hairpin and the conserved P54 residue are both necessary for the Golgi-targeting motif.

The Golgi-targeting motif is important because E protein is known to accumulate in the ERGIC of infected cells (Nieto-Torres et al., 2011), where CoV virions bud into

the lumen (Narayanan et al., 2000). In cells infected with CoVs lacking E protein, immature virions can be observed accumulating along the secretory pathway and not being released (Almazán et al., 2013; Ortego et al., 2007; Ruch and Machamer, 2011).

Co-transfection of E protein with M protein, the most abundant protein in the CoV virion, is sufficient to drive the formation of viral-like particles (Baudoux et al., 1998; Bos et al., 1996; Corse and Machamer, 2000; Vennema et al., 1996). These two proteins have been found to interact through the C-terminal extramembrane domain of E protein (Baudoux et al., 1998; Corse and Machamer, 2002; Lim and Liu, 2001; Lim et al., 2001). These observations point towards the role of E protein, specifically its C-terminal hydrophobic region, in the budding and release of CoV virions. The structural flexibility of the hydrophobic region suggests some sort of catalytic process, possibly involving membrane rearrangement, e.g. membrane fission during the virion budding process. This could be mediated by a conformational exchange from β -hairpin to α -helix, or by the two C-terminal helices forming an α -hairpin. A similar role has been proposed for the N-terminal hydrophobic region, which in very high peptide-to-lipid ratio seems to adopt an α -hairpin structure (Arbely et al., 2004; Khattari et al., 2006).

E protein from several CoVs are functionally interchangeable despite their low sequence identity (Kuo et al., 2007). It is likely that all CoV E proteins share the same structural features, although this is yet to be shown beyond their secondary structure prediction. It will be very interesting to compare the structure of E proteins from different CoVs. Study on the full-length IBV E protein construct is underway and is expected to add further to our understanding on the structure of E protein and its function.

5.2.2 The pentamer bundle and the ion channel activity

The TM domain of E proteins also plays an important role in CoV pathogenesis: mice infected with SARS-CoV carrying defective E protein, i.e. ion channel inactive, recovered and survived the infection (Nieto-Torres et al., 2014). Mice infected with wild-type SARS-CoV typically show signs of lung edema and elevated inflammatory cytokines, all of which are absent during infection with SARS-CoV with ion-channel-defective E protein.

E proteins conduct ions through a pore formed by the pentameric arrangement of its TM domain (Verdiá-Báguena et al., 2012). Its ion channel activity can be inhibited by HMA (Wilson et al., 2006), and herein we have identified two potential HMA binding sites located at the ends of the E protein pentamer bundle, i.e. N15 and L37 at the N- and C-terminal end of the TM, respectively.

Residue N15 has been demonstrated to be important, as N15A mutation completely abolish ion channel activity in SARS-CoV E (Parthasarathy et al., 2008). When this mutation was introduced into the virus, as mentioned earlier, the virus becomes attenuated – but in some cases acquired a N15D compensatory mutation which restored both ion channel activity and virulence (Nieto-Torres et al., 2014). The N-to-D mutation preserves the polar nature of this residue, which is homologously polar in other CoV E proteins (Fig. 4-1), further confirming the importance of the polar residue at this position in E protein TM domain.

V25F is another mutation that abolishes channel activity of SARS-CoV E (Parthasarathy et al., 2008). When this mutation was introduced into the virus, it was relatively easier for the virus to evolve compensatory mutations and regained virulence (Nieto-Torres et al., 2014). One of these mutations, V25L, reverted the V25F mutation to a similar residue type. The rest, however, are scattered throughout the TM domain: L19A, F26C, L27S, T30I, and L37R.

The model for E protein pentamer presented in the current study can offer some explanations on how V25F mutation inactivates the ion channel and how the compensatory mutations can restore channel activity (Fig. 5-1). In this model, residue V25 is oriented towards the TM-TM interface, thus V25F mutation can be expected to directly affect the intermonomeric contacts and the pentamer stability. In this way V25F can abolish channel activity, although the exact mechanism, e.g. by destabilizing the pentamer or instead by compacting the pentamer in a way that causes channel blockage, is not immediately obvious.

Figure 5-1. Pentameric model of E protein in the context of mutagenesis data. Cartoon representation shows the TM domain of E protein according to the model for E protein pentamer built in this study. Yellow-coloured residues show the position of compensatory mutations which recovered ion channel activity of E protein and virulence of SARS-CoV in mice model (Nieto-Torres et al., 2014)

Residues F26, L27, and T30 are clustered close to V25 (Fig. 5-1). Thus, mutation on one of these residues possibly restores the pentamer stability through a direct interaction with V25F. Residue L37, which faces the TM-TM interface, could also be expected to affect the pentamer stability and thus indirectly compensates for the destabilization by V25F mutation. The luminal L19 is less straightforward to explain, but it possibly affects the shape or the size of the channel. Thus the pentameric model of E protein obtained herein is in agreement with the mutagenesis data, and can explain the ion channel activity of E protein.

5.2.3 A pentameric model that represents the “open” form of the channel

The pentameric model presented herein is in agreement with the “pentamer B” model, i.e. one of the two pentameric models for E protein which differs by their relative TM orientation (pentamer A and B, Fig. 4-2) built using evolutionary conservation data (Torres et al., 2005). However, studies using synthetic TM peptides (E_{TM}) converged on pentamer A as the correct model (Parthasarathy et al., 2008; Torres et al., 2006). Several reasons may account for this discrepancy, from the construct length difference to the membrane mimic environment used.

The present study uses a truncated E protein from residues 8-65 (E_{TR}), which includes some part of the extramembrane domain. The TM domain alone is sufficient for pentamerization and ion channel activity (Verdiá-Báguena et al., 2012), and the monomer-pentamer equilibrium constant of E_{TM} ($\log K_a \sim 17$) does not differ from the full-length protein (E_{FL}) (Parthasarathy et al., 2008; Parthasarathy et al., 2012). These suggest that the conformation of the pentamer is not determined by the extramembrane domain alone.

However, two inter-monomeric NOE contacts at the extramembrane domain of E_{TR} were observed, and had been used as distance restraints to build the pentamer model in the present study. As the orientation of the TM domain in the E_{TR} model is different from that in the E_{TM} , this suggests that the extramembrane domain might actually influence the conformation of the pentamer. By logical extension, though, the residues truncated in E_{TR} (66-76) could also influence the TM orientation if these residues also form inter-monomeric contacts. In that case, the TM orientations of E_{TR} , E_{FL} , and E_{TM} could be different from each other.

Two observations suggest that the TM orientation of E_{FL} and E_{TM} are similar, but different from E_{TR} . Firstly, the monomer-pentamer equilibrium constant fitted for E_{TR} ($\log K_a \sim 16$) is lower than that of E_{FL} and E_{TM} . Secondly, the conductivity of E_{TR} protein in lipid bilayer was significantly higher than E_{FL} and E_{TM} . These also suggest that the E_{TR} model (pentamer B) represents an “open” form of the channel, whereas E_{TM} model (pentamer A) is closer to the “closed” form.

Whether these speculations are correct is yet to be determined, e.g. from the structure of the full-length E protein. The full-length SARS-CoV E protein is poorly soluble, but it might be amenable by solid-state NMR, which is suitable for insoluble proteins. In addition, work on determining the structure of the full-length IBV E protein is currently underway.

5.3 Structural preference in different environments

Whether the structure of a membrane protein will be the same or different when reconstituted in different membrane mimic systems cannot always be known *a priori*. In the current study, two systems have been utilized: detergent micelles and phospholipid bicelles. The structural features of integrin α/β are different in these two systems, i.e. the GFFKR motif forms helical structure in micelles and backbone reversal in bicelles. On the other hand, the structure of E protein in micelles and in bicelles is fairly similar, at least in terms of secondary structures and backbone conformation. There are various biophysical methods that can be used to compare the structures in different environment, as exemplified in current study. To some extent, these methods can reveal subtle differences between structures found in different reconstitution systems. Two of these, i.e. membrane insertion and secondary structure, will be discussed in this section.

5.3.1 Variations in membrane-inserted residues

When studying membrane proteins, we are interested in determining which residues are embedded inside the membrane. Various prediction tools, e.g. TMHMM (Krogh et al., 2001), exist for this purpose and generally give excellent prediction results. In some cases, prediction results might not always agree with *in vivo* observations.

In the case of E protein, for instance, two TM domains have been predicted for IBV E. This was based on the hydrophobicity of the C-terminal hydrophobic domain of E protein, which is very high in IBV E and relatively low in SARS-CoV E, MERS-CoV E, MHV E, and TGEV E (Fig. 4-1). However, studies using untagged IBV E protein found that the N- and C-terminus are not in the same topological domain, indicating that it has only one TM domain (Corse and Machamer, 2000; Ruch and Machamer, 2012).

In vivo observations using different methods and in different reconstitution environments may also report different numbers of membrane-inserted residues. For example, in current study, H/D exchange in lipid bilayers measured by comparison of FTIR amide II bands always gave the highest number membrane-inserted residues, followed by 5-DSA/16-DSA PRE experiment in bicelles/micelles, and finally H/D exchange monitored by [^1H - ^{15}N]-HSQC. Several reasons may account for this particular ordering.

Firstly, stacked lipid bilayers used in FTIR is relatively dried, albeit with residual interstitial water, whereas bicelles and micelles are fully hydrated. Thus the lower hydration in stacked lipid bilayers may overestimate the number of membrane-embedded residues in an H/D exchange experiment. Or conversely, the higher hydration in micelles/bicelles could underestimate the number of membrane-embedded residues (reviewed in (Cross et al., 2013) and (Zhou and Cross, 2013)). This effect may be even more pronounced in micelles, which have higher curvature than bicelles and lipid bilayers.

Excessive hydration in membrane mimic environments has been observed in other membrane protein systems as well. For example, H/D exchange rate in the influenza A M2 TM domain reconstituted in Triton X-100 detergent micelles was found to be substantially faster than in DMPC liposomes (Hansen et al., 2002). Similarly, H/D exchange could be observed in the TM residues of the histidine kinase receptor KdpD reconstituted in LMPG micelles (Maslennikov et al., 2010).

Secondly, H/D exchange experiments are usually time-sensitive due to the diffusion of water into the hydrophobic interior. The timescale involved can be quite different depending on how D_2O is introduced into the sample. In FTIR, the timescale can be quite long because it uses N_2 gas stream saturated with D_2O vapour. On the other

hand, micelles/bicelles are fully rehydrated in liquid D₂O, making it easier for some membrane-inserted residues to undergo exchange, especially those near the surface of the membrane. Indeed, H/D exchange half-life of juxtamembrane residues generally decrease exponentially as they get closer to the surface (Demmers et al., 2000). The difference in timescale could be reconciled by using time-course experiments coupled with exchange quenching (Czerski et al., 2000; Dempsey, 2001).

Thus H/D exchange experiments may not be the best choice when hydration and time are of concern. Other alternative methods such as lipid-soluble 5-DSA/16-DSA, or water-soluble Mn²⁺EDDA²⁻ and Gd³⁺ paramagnetic reagents are not hydration-dependent or time-dependent, and depend only on the thickness of the reconstitution system. The thickness of DMPC/DHPC bicelles, for example, is 60.8Å (Andersson and Mäler, 2006), which is quite similar to fully hydrated DMPC bilayers at 62.6Å (Kučerka et al., 2005). Thus determination of membrane-embedded residues in bicelles can be quite similar to that in lipid bilayers when using paramagnetic reagents.

5.3.2 Effects on secondary structure propensity

Secondary structure is another information that can be readily predicted from the amino acid sequence to a great accuracy. However, different secondary structures may form due to influence of reconstitution environment, or tertiary contacts from other parts of the protein.

In the case of integrin α , for example, its GFFKR juxtamembrane motif has been observed to form backbone reversal which rests upon the preceding TM domain. In the absence of the preceding TM domain, the backbone reversal can hardly be expected. In addition, the backbone reversal has been observed only in bicelles, whereas in micelles and organic solvent it has been observed to form α -helix.

In the case of E protein, a stretch of hydrophobic residues in the C-terminal has been predicted to form a β -hairpin motif. A synthetic peptide corresponding to this region has been observed to form completely β -strand. But in the context of a longer protein construct, this region folds into an α -helix.

In both cases, we have discussed and concluded that the two forms observed could possibly co-exist in a conformational equilibrium, and these may reflect different

states in the function of the protein. Indeed, proteins often do not form a single, rigid structure *in vivo*, instead the true conformation dynamically change depending on their functional state and immediate environment. Different reconstitution environments may be biased towards one conformation or the other, and a task is left to the researchers to determine and validate whether the structure they have obtained is functionally relevant.

CONCLUSIONS

The current study explores the theme of reconstitution environment and its effects on the structure of membrane proteins. To this end, two systems have been chosen: (1) the TM domain of integrin α L/ β 2, a single-pass integral membrane protein that heterodimerize in α / β pairs, and (2) CoV E proteins, a single-pass integral membrane protein that form homopentameric ion channels.

Firstly, a solution NMR model for α L/ β 2 TM heterodimer in bicelles determined herein is very similar to a previously published solution NMR model for α IIB/ β 3 TM in bicelle (Lau et al., 2009). This similarity suggests that both the general α / β heterodimer interface and the GFFKR backbone reversal motif at the juxtamembrane of α -integrins are features that may be commonly found in all α / β integrin pairs. The discrepancy between the backbone reversal conformation, which has so far been observed only in bicelles, and the α -helical form observed in micelles or organic solvent, possibly corresponds to two different activation states of integrin.

The α L/ β 2 heterodimer model presented herein, however, does not have enough details to determine whether a polar interaction exists at the α L/ β 2 heterodimer interface, as we have previously proposed. A more detailed model of the heterodimer interface built using intermonomeric NOE restraints, is still needed.

For the second system, the solution NMR models for monomeric and pentameric forms of SARS-CoV E protein has been presented. In these models, the E protein structure consists of a N-terminal helical TM domain, and a membrane-peripheral C-terminal extramembrane helix. The C-terminal hydrophobic domain was found to form an α -helix, with a minor population forming a β -strand.

In the pentameric model, E protein TM domains form a left-handed bundle consistent with previously published “pentamer B” model built computationally using evolutionary conservation data of E proteins. The orientation of the TM domain is in agreement with mutagenesis data of SARS-CoV in mouse model, but the extramembrane domain in this model still needs some refinement.

REFERENCES

- Abdul-Gader, A., Miles, A.J., Wallace, B.A., (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics* **27**, 1630-1636.
- Adair, B.D., Yeager, M., (2002) Three-dimensional model of the human platelet integrin $\alpha\text{IIb}\beta 3$ based on electron cryomicroscopy and x-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14059-14064.
- Almazán, F., Dediego, M.L., Sola, I., Zuñiga, S., Nieto-Torres, J.L., Marquez-Jurado, S., Andrés, G., Enjuanes, L., (2013) Engineering a replication-competent, propagation-defective middle east respiratory syndrome coronavirus as a vaccine candidate. *mBio* **4**.
- Andersson, A., Måler, L., (2006) Size and Shape of Fast-tumbling Bicelles as Determined by Translational Diffusion. *Langmuir* **22**, 2447-2449.
- Arbely, E., Khattari, Z., Brotons, G., Akkawi, M., Salditt, T., Arkin, I.T., (2004) A Highly Unusual Palindromic Transmembrane Helical Hairpin Formed by SARS Coronavirus E Protein. *Journal of molecular biology* **341**, 769-779.
- Arkin, I.T., Brunger, A.T., (1998) Statistical analysis of predicted transmembrane α -helices. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1429**, 113-128.
- Balannik, V., Wang, J., Ohigashi, Y., Jing, X., Magavern, E., Lamb, R.A., DeGrado, W.F., Pinto, L.H., (2009) Design and Pharmacological Characterization of Inhibitors of Amantadine-Resistant Mutants of the M2 Ion Channel of Influenza A Virus. *Biochemistry* **48**, 11872-11882.
- Battiste, J.L., Wagner, G., (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365.
- Baudoux, P., Carrat, C., Besnardeau, L., Charley, B., Laude, H., (1998) Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes. *Journal of Virology* **72**, 8636-8643.
- Bayrhuber, M., Meins, T., Habeck, M., Becker, S., Giller, K., Villinger, S., Vonnrhein, C., Griesinger, C., Zweckstetter, M., Zeth, K., (2008) Structure of the human voltage-dependent anion channel. *Proc Natl Acad Sci U S A* **105**, 15370-15375.
- Bhunia, A., Tang, X.Y., Mohanram, H., Tan, S.M., Bhattacharjya, S., (2009) NMR solution conformations and interactions of integrin $\alpha\text{L}\beta 2$ cytoplasmic tails. *Journal of Biological Chemistry* **284**, 3873-3884.
- Bocharov, E.V., Mineev, K.S., Goncharuk, M.V., Arseniev, A.S., (2012) Structural and thermodynamic insight into the process of "weak" dimerization of the ErbB4 transmembrane domain by solution NMR. *Biochimica et biophysica acta* **1818**, 2158-2170.

- Bondarenko, V., Mowrey, D., Tillman, T., Cui, T., Liu, L.T., Xu, Y., Tang, P., (2012) NMR structures of the transmembrane domains of the $\alpha 4\beta 2$ nAChR. *Biochimica et biophysica acta* **1818**, 1261-1268.
- Bos, E.C.W., Luytjes, W., Van Der Meulen, H., Koerten, H.K., Spaan, W.J.M., (1996) The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. *Virology* **218**, 52-60.
- Bowie, J.U., (2001) Stabilizing membrane proteins. *Current Opinion in Structural Biology* **11**, 397-402.
- Boyd, D., Schierle, C., Beckwith, J., (1998) How many membrane proteins are there? *Protein Science* **7**, 201-205.
- Breyton, C., Tribet, C., Olive, J., Dubacq, J.P., Popott, J.L., (1997) Dimer to monomer conversion of the cytochrome b6f complex: Causes and consequences. *Journal of Biological Chemistry* **272**, 21892-21900.
- Brunger, A.T., (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* **2**, 2728-2733.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905-921.
- Butcher, E.C., Williams, M., Youngman, K., Rott, L., Briskin, M., (1999) *Lymphocyte trafficking and regional immunity*, Advances in Immunology, pp. 209-253.
- Cady, S.D., Schmidt-Rohr, K., Wang, J., Soto, C.S., DeGrado, W.F., Hong, M., (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463**, 689-692.
- Caffrey, M., Li, D., Dukkipati, A., (2012) Membrane Protein Structure Determination Using Crystallography and Lipidic Mesophases - Recent Advances and Successes. *Biochemistry* **51**, 6266-6288.
- Carman, C.V., Springer, T.A., (2003) Integrin avidity regulation: are changes in affinity and conformation underemphasized? *Current Opinion in Cell Biology* **15**, 547-556.
- Carrasco, L., (1995) *Modification of Membrane Permeability by Animal Viruses*. In: Karl Maramorosch, F.A.M., Aaron, J.S. (Eds.), *Advances in Virus Research*. Academic Press, pp. 61-112.
- Caswell, P.T., Norman, J.C., (2006) Integrin Trafficking and the Control of Cell Migration. *Traffic* **7**, 14-21.
- Chen, I.A., Walde, P., (2010) From Self-Assembled Vesicles to Protocells. *Cold Spring Harbor Perspectives in Biology* **2**.
- Chng, C.P., Tan, S.M., (2011) Leukocyte integrin $\alpha L\beta 2$ transmembrane association dynamics revealed by coarse-grained molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **79**, 2203-2213.
- Chou, J.J., Kaufman, J.D., Stahl, S.J., Wingfield, P.T., Bax, A., (2002) Micelle-Induced Curvature in a Water-Insoluble HIV-1 Env Peptide Revealed by NMR

Dipolar Coupling Measurement in Stretched Polyacrylamide Gel. *Journal of the American Chemical Society* **124**, 2450-2451.

Chua, G.L., Patra, A.T., Tan, S.M., Bhattacharjya, S., (2013) NMR Structure of Integrin alpha4 Cytosolic Tail and Its Interactions with Paxillin. *PLoS One* **8**, e55184.

Chua, G.L., Tang, X.Y., Amalraj, M., Tan, S.M., Bhattacharjya, S., (2011) Structures and interaction analyses of integrin alphaMbeta2 cytoplasmic tails. *Journal of Biological Chemistry* **286**, 43842-43854.

Chua, G.L., Tang, X.Y., Patra, A.T., Tan, S.M., Bhattacharjya, S., (2012) Structure and binding interface of the cytosolic tails of alphaXbeta2 integrin. *PLoS One* **7**, e41924.

Cohen, J.R., Lin, L.D., Machamer, C.E., (2011) Identification of a Golgi Complex-Targeting Signal in the Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Envelope Protein. *J Virol* **85**, 5794-5803.

Compton, L.A., Johnson, W.C., (1986) Analysis of protein circular-dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical biochemistry* **155**, 155-167.

Cornilescu, G., Delaglio, F., Bax, A., (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of biomolecular NMR* **13**, 289-302.

Corse, E., Machamer, C.E., (2000) Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *Journal of Virology* **74**, 4319-4326.

Corse, E., Machamer, C.E., (2002) The cytoplasmic tail of infectious bronchitis virus E protein directs golgi targeting. *Journal of Virology* **76**, 1273-1284.

Cross, T.A., Murray, D.T., Watts, A., (2013) Helical Membrane Protein Conformations and their Environment. *European biophysics journal : EBJ* **42**, 10.1007/s00249-00013-00925-x.

Cross, T.A., Sharma, M., Yi, M., Zhou, H.-X., (2011) Influence of Solubilizing Environments on Membrane Protein Structures. *Trends in biochemical sciences* **36**, 117-125.

Curran, A.R., Engelman, D.M., (2003) Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Current Opinion in Structural Biology* **13**, 412-417.

Czerski, L., Vinogradova, O., Sanders, C.R., (2000) NMR-Based Amide Hydrogen–Deuterium Exchange Measurements for Complex Membrane Proteins: Development and Critical Evaluation. *Journal of Magnetic Resonance* **142**, 111-119.

Damberg, P., Jarvet, J., Gräslund, A., (2001) [13] - *Micellar Systems as Solvents in Peptide and Protein Structure Determination*. In: Thomas L. James, V.D., Uli, S. (Eds.), *Methods in Enzymology*. Academic Press, pp. 271-285.

De Angelis, A.A., Nevzorov, A.A., Park, S.H., Howell, S.C., Mrse, A.A., Opella, S.J., (2004) High-resolution NMR spectroscopy of membrane proteins in aligned bicelles. *Journal of the American Chemical Society* **126**, 15340-15341.

De Clercq, E., (2002) Strategies in the design of antiviral drugs. *Nat Rev Drug Discov* **1**, 13-25.

- De Vries, S.J., van Dijk, A.D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., Bonvin, A.M., (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: structure, function, and bioinformatics* **69**, 726-733.
- DeDiego, M.L., Pewe, L., Alvarez, E., Rejas, M.T., Perlman, S., Enjuanes, L., (2008) Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE-2 transgenic mice. *Virology* **376**, 379-389.
- DeDiego, M.L., Nieto-Torres, J.L., Regla-Nava, J.A., Jimenez-Guardeño, J.M., Fernandez-Delgado, R., Fett, C., Castaño-Rodriguez, C., Perlman, S., Enjuanes, L., (2014) Inhibition of NF- κ B-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. *Journal of Virology* **88**, 913-924.
- DeDiego, M.L., Álvarez, E., Almazán, F., Rejas, M.T., Lamirande, E., Roberts, A., Shieh, W.J., Zaki, S.R., Subbarao, K., Enjuanes, L., (2007) A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *Journal of Virology* **81**, 1701-1713.
- DeDiego, M.L., Nieto-Torres, J.L., Jiménez-Guardeño, J.M., Regla-Nava, J.A., Álvarez, E., Oliveros, J.C., Zhao, J., Fett, C., Perlman, S., Enjuanes, L., (2011) Severe acute respiratory syndrome coronavirus envelope protein regulates cell stress response and apoptosis. *PLoS Pathogens* **7**.
- Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H., (1985) Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3Å resolution. *Nature* **318**, 618-624.
- Demmers, J.A.A., Haverkamp, J., Heck, A.J.R., Koeppe, R.E., Killian, J.A., (2000) Electrospray ionization mass spectrometry as a tool to analyze hydrogen/deuterium exchange kinetics of transmembrane peptides in lipid bilayers. *Proceedings of the National Academy of Sciences* **97**, 3189-3194.
- Dempsey, C.E., (2001) Hydrogen exchange in peptides and proteins using NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39**, 135-170.
- Dominguez, C., Boelens, R., Bonvin, A.M., (2003) HADDOCK: A protein–protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* **125**, 1731-1737.
- Durchschlag, H., (1986) *Specific Volumes of Biological Macromolecules and Some Other Molecules of Biological Interest*. In: Hinz, H.-J. (Ed.), *Thermodynamic Data for Biochemistry and Biotechnology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 45-128.
- Enjuanes, L., Brian, D., Cavanagh, D., Holmes, K., Lai, M.M.C., Laude, H., Masters, P., Rottier, P., Siddell, S.G., Spaan, W.J.M., Taguchi, F., Talbot, P., (2000) *Coronaviridae*. In: van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carsten, E.B., Estes, M.K., Lemon, S.M., McGeoch, D.J., Maniloff, J., Mayo, M.A., Pringle, C.R., Wickner, R.B. (Eds.), *Virus taxonomy. Classification and nomenclature of viruses*. Academic Press, San Diego, pp. 835-849.
- Ferré, S., Casadó, V., Devi, L.A., Filizola, M., Jockers, R., Lohse, M.J., Milligan, G., Pin, J.P., Guitart, X., (2014) G protein-coupled receptor oligomerization revisited: Functional and pharmacological perspectives. *Pharmacological Reviews* **66**, 413-434.

- Fleming, K.G., (2005) *Analysis of Membrane Proteins Using Analytical Ultracentrifugation*, Analytical Ultracentrifugation: Techniques and Methods, pp. 432.
- Franzin, C.M., Teriete, P., Marassi, F.M., (2007) Structural similarity of a membrane protein in micelles and membranes. *Journal of the American Chemical Society* **129**, 8078-8079.
- Gan, S.W., Vararattanavech, A., Nordin, N., Eshaghi, S., Torres, J., (2011) A cost-effective method for simultaneous homo-oligomeric size determination and monodispersity conditions for membrane proteins. *Analytical biochemistry* **416**, 100-106.
- Gan, S.W., Tan, E., Lin, X., Yu, D., Wang, J., Tan, G.M.Y., Vararattanavech, A., Yeo, C.Y., Soon, C.H., Soong, T.W., Pervushin, K., Torres, J., (2012) The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels. *Journal of Biological Chemistry* **287**, 24671-24689.
- Gautier, A., Mott, H.R., Bostock, M.J., Kirkpatrick, J.P., Nietlispach, D., (2010) Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nature structural & molecular biology* **17**, 768-774.
- Glover, K.J., Whiles, J.A., Wu, G., Yu, N.J., Deems, R., Struppe, J.O., Stark, R.E., Komives, E.A., Vold, R.R., (2001) Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophysical Journal* **81**, 2163-2171.
- Godet, M., L'Haridon, R., Vautherot, J.-F., Laude, H., (1992) TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology* **188**, 666-675.
- Gong, X.M., Ding, Y., Yu, J., Yao, Y., Marassi, F.M., (2015) Structure of the Na,K-ATPase regulatory protein FXVD2b in micelles: implications for membrane-water interfacial arginines. *Biochimica et biophysica acta* **1848**, 299-306.
- Gottschalk, K.E., Adams, P.D., Brunger, A.T., Kessler, H., (2002) Transmembrane signal transduction of the α IIb β 3 integrin. *Protein Science* **11**, 1800-1812.
- Guermeur, Y., Geourjon, C., Gallinari, P., Delvigne, G., (1999) Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* **15**, 413-421.
- Guntert, P., Mumenthaler, C., Wuthrich, K., (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of molecular biology* **273**, 283-298.
- Hagn, F., Etzkorn, M., Raschle, T., Wagner, G., (2013) Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal of the American Chemical Society* **135**, 1919-1925.
- Hansen, R.K., Broadhurst, R.W., Skelton, P.C., Arkin, I.T., (2002) Hydrogen/deuterium exchange of hydrophobic peptides in model membranes by electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **13**, 1376-1387.

Hantgan, R.R., Paumi, C., Rocco, M., Weisel, J.W., (1999) Effects of ligand-mimetic peptides Arg-Gly-Asp-X (X = Phe, Trp, Ser) on α IIb β 3 integrin conformation and oligomerization. *Biochemistry* **38**, 14461-14474.

Heginbotham, L., Kolmakova-Partensky, L., Miller, C., (1998) Functional Reconstitution of a Prokaryotic K(+) Channel. *The Journal of General Physiology* **111**, 741-749.

Hemler, M., (1999) *Extracellular Matrix, Anchor, and Adhesion Proteins*. Oxford University Press, Oxford, Great Britain.

Henry, G.D., Sykes, B.D., (1994) [18] *Methods to study membrane protein structure in solution*. In: Thomas L. James, N.J.O. (Ed.), *Methods in Enzymology*. Academic Press, pp. 515-535.

Herrmann, T., Guntert, P., Wuthrich, K., (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of molecular biology* **319**, 209-227.

Hong, M., DeGrado, W.F., (2008) Structural basis for proton conduction and inhibition by the influenza M2 protein. *Nature* **451**, 596-599.

Huang, H., Vogel, H.J., (2012) Structural Basis for the Activation of Platelet Integrin α IIb β 3 by Calcium- and Integrin-Binding Protein 1. *Journal of the American Chemical Society* **134**, 3864-3872.

Kang, C., Tian, C., Sonnichsen, F.D., Smith, J.A., Meiler, J., George, A.L., Jr., Vanoye, C.G., Kim, H.J., Sanders, C.R., (2008) Structure of KCNE1 and implications for how it modulates the KCNQ1 potassium channel. *Biochemistry* **47**, 7999-8006.

Katagiri, K., Maeda, A., Shimonaka, M., Kinashi, T., (2003) RAPL, a Rap1-binding molecule that mediates Rap1-induced adhesion through spatial regulation of LFA-1. *Nature Immunology* **4**, 741-748.

Kauppinen, J.K., Moffatt, D.J., Cameron, D.G., Mantsch, H.H., (1981) NOISE IN FOURIER SELF-DECONVOLUTION. *Applied Optics* **20**, 1866-1879.

Keller, R., (2004) The computer aided resonance assignment tutorial. *Cantina, Goldau* **15**.

Khattari, Z., Brotons, G., Akkawi, M., Arbely, E., Arkin, I.T., Salditt, T., (2006) SARS Coronavirus E Protein in Phospholipid Bilayers: An X-Ray Study. *Biophysical Journal* **90**, 2038-2050.

Kim, M., Carman, C.V., Springer, T.A., (2003) Bidirectional transmembrane signaling by cytoplasmic domain separation in integrins. *Science* **301**, 1720-1725.

Kim, M., Carman, C.V., Yang, W., Salas, A., Springer, T.A., (2004) The primacy of affinity over clustering in regulation of adhesiveness of the integrin α (L) β 2. *The Journal of cell biology* **167**, 1241-1253.

Kinashi, T., (2007) Integrin regulation of lymphocyte trafficking: lessons from structural and signaling studies. *Adv Immunol* **93**, 185-227.

King, R.D., Sternberg, M.J.E., (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Science* **5**, 2298-2310.

- Koradi, R., Billeter, M., Wuthrich, K., (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-55, 29-32.
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹. *Journal of molecular biology* **305**, 567-580.
- Kučerka, N., Liu, Y., Chu, N., Petrache, H.I., Tristram-Nagle, S., Nagle, J.F., (2005) Structure of Fully Hydrated Fluid Phase DMPC and DLPC Lipid Bilayers Using X-Ray Scattering from Oriented Multilamellar Arrays and from Unilamellar Vesicles. *Biophysical Journal* **88**, 2626-2637.
- Kuo, L., Masters, P.S., (2003) The small envelope protein E is not essential for murine coronavirus replication. *Journal of Virology* **77**, 4597-4608.
- Kuo, L., Hurst, K.R., Masters, P.S., (2007) Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function. *Journal of Virology* **81**, 2249-2262.
- Lai, C., Liu, X., Tian, C., Wu, F., (2013) Integrin $\alpha 1$ Has a Long Helix, Extending from the Transmembrane Region to the Cytoplasmic Tail in Detergent Micelles. *PLoS One* **8**, e62954.
- Lamirande, E.W., DeDiego, M.L., Roberts, A., Jackson, J.P., Alvarez, E., Sheahan, T., Shieh, W.J., Zaki, S.R., Baric, R., Enjuanes, L., Subbarao, K., (2008) A live attenuated severe acute respiratory syndrome coronavirus is immunogenic and efficacious in Golden Syrian hamsters. *Journal of Virology* **82**, 7721-7724.
- Langosch, D., Brosig, B., Kolmar, H., Fritz, H.-J., (1996) Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *Journal of molecular biology* **263**, 525-530.
- Lau, T.L., Dua, V., Ulmer, T.S., (2008) Structure of the integrin αIIb transmembrane segment. *Journal of Biological Chemistry* **283**, 16162-16168.
- Lau, T.L., Kim, C., Ginsberg, M.H., Ulmer, T.S., (2009) The structure of the integrin $\alpha \text{IIb} \beta 3$ transmembrane complex explains integrin transmembrane signalling. *The EMBO journal* **28**, 1351-1361.
- le Maire, M., Champeil, P., Møller, J.V., (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1508**, 86-111.
- Lemmon, M.A., Treutlein, H.R., Adams, P.D., Brünger, A.T., Engelman, D.M., (1994) A dimerization motif for transmembrane α -helices. *Nature Structural Biology* **1**, 157-163.
- Li, Q., Wong, Y.L., Huang, Q., Kang, C., (2014a) Structural insight into the transmembrane domain and the juxtamembrane region of the erythropoietin receptor in micelles. *Biophys J* **107**, 2325-2336.
- Li, R., Babu, C.R., Lear, J.D., Wand, A.J., Bennett, J.S., DeGrado, W.F., (2001) Oligomerization of the integrin $\alpha \text{IIb} \beta 3$: roles of the transmembrane and cytoplasmic domains. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12462-12467.

- Li, Y., Surya, W., Claudine, S., Torres, J., (2014b) Structure of a Conserved Golgi Complex-targeting Signal in Coronavirus Envelope Proteins. *Journal of Biological Chemistry* **289**, 12535-12549.
- Li, Y., To, J., Verdià-Baguena, C., Dossena, S., Surya, W., Huang, M., Paulmichl, M., Liu, D.X., Aguilella, V.M., Torres, J., (2014c) Inhibition of the human respiratory syncytial virus small hydrophobic protein and structural variations in a bicelle environment. *Journal of Virology* **88**, 11899-11914.
- Liang, B., Bushweller, J.H., Tamm, L.K., (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* **128**, 4389-4397.
- Liao, Y., Yuan, Q., Torres, J., Tam, J.P., Liu, D.X., (2006) Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology* **349**, 264-275.
- Lim, K.P., Liu, D.X., (2001) The missing link in coronavirus assembly. Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins. *Journal of Biological Chemistry* **276**, 17515-17523.
- Lim, K.P., Xu, H.Y., Liu, D.X., (2001) *Physical interaction between the membrane (M) and envelope (E) proteins of the coronavirus avian infectious bronchitis virus (IBV)*, Advances in Experimental Medicine and Biology, pp. 595-602.
- Lin, X., Tan, S.M., Law, S.K.A., Torres, J., (2006) Unambiguous prediction of human integrin transmembrane heterodimer interactions using only homologous sequences. *Proteins-Structure Function and Bioinformatics* **65**, 274-279.
- Linke, D., (2009) *Chapter 34 Detergents: An Overview*. In: Richard, R.B., Murray, P.D. (Eds.), *Methods in Enzymology*. Academic Press, pp. 603-617.
- Lopez, L.A., Riffle, A.J., Pike, S.L., Gardner, D., Hogue, B.G., (2008) Importance of conserved cysteine residues in the coronavirus envelope protein. *Journal of Virology* **82**, 3000-3010.
- Lu, C.F., Springer, T.A., (1997) The alpha subunit cytoplasmic domain regulates the assembly and adhesiveness of integrin lymphocyte function-associated antigen-1. *Journal of Immunology* **159**, 268-278.
- Lu, C.F., Takagi, J., Springer, T.A., (2001) Association of the membrane proximal regions of the α and β subunit cytoplasmic domains constrains an integrin in the inactive state. *Journal of Biological Chemistry* **276**, 14642-14648.
- Luo, B.H., Springer, T.A., Takagi, J., (2004) A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biology* **2**, 776-786.
- Luo, B.H., Carman, C.V., Takagi, J., Springer, T.A., (2005) Disrupting integrin transmembrane domain heterodimerization increases ligand binding affinity, not valency or clustering. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3679-3684.
- MacKenzie, K.R., Prestegard, J.H., Engelman, D.M., (1997) A transmembrane helix dimer: Structure and implications. *Science* **276**, 131-133.

- Marcotte, I., Auger, M., (2005) Bicelles as model membranes for solid- and solution-state NMR studies of membrane peptides and proteins. *Concepts in Magnetic Resonance Part A* **24A**, 17-37.
- Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mors, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y.H., Choe, S., (2010) Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. *Proc Natl Acad Sci U S A* **107**, 10902-10907.
- Matsushita, C., Tamagaki, H., Miyazawa, Y., Aimoto, S., Smith, S.O., Sato, T., (2013) Transmembrane helix orientation influences membrane binding of the intracellular juxtamembrane domain in Neu receptor peptides. *Proceedings of the National Academy of Sciences* **110**, 1646-1651.
- Mayer, G., Ludwig, B., Müller, H.-W., Van den Broek, J., Friesen, R., Schubert, D., (1999) *Studying membrane proteins in detergent solution by analytical ultracentrifugation: different methods for density matching*, Analytical Ultracentrifugation V. Springer, pp. 176-181.
- McDermott, A., (2009) Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR. *Annual Review of Biophysics* **38**, 385-403.
- Miyamoto, K., Togiya, K., (2011) Solution structure of LC4 transmembrane segment of CCR5. *PLoS One* **6**, e20452.
- Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., Stewart, P.D.S., (2014) Membrane protein structure determination — The next generation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838**, 78-87.
- Nanga, R.P., Brender, J.R., Vivekanandan, S., Ramamoorthy, A., (2011) Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica et biophysica acta* **1808**, 2337-2342.
- Narayanan, K., Maeda, A., Maeda, J., Makino, S., (2000) Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *Journal of virology* **74**, 8127-8134.
- Nickols, H.H., Conn, P.J., (2014) Development of allosteric modulators of GPCRs for treatment of CNS disorders. *Neurobiology of Disease* **61**, 55-71.
- Nieto-Torres, J.L., DeDiego, M.L., Álvarez, E., Jiménez-Guardaño, J.M., Regla-Nava, J.A., Llorente, M., Kremer, L., Shuo, S., Enjuanes, L., (2011) Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology* **415**, 69-82.
- Nieto-Torres, J.L., Verdiá-Báguena, C., Jimenez-Guardaño, J.M., Regla-Nava, J.A., Castaño-Rodriguez, C., Fernandez-Delgado, R., Torres, J., Aguilera, V.M., Enjuanes, L., (2015) Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology* **485**, 330-339.
- Nieto-Torres, J.L., DeDiego, M.L., Verdiá-Báguena, C., Jimenez-Guardaño, J.M., Regla-Nava, J.A., Fernandez-Delgado, R., Castaño-Rodriguez, C., Alcaraz, A., Torres, J., Aguilera, V.M., Enjuanes, L., (2014) Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity Promotes Virus Fitness and Pathogenesis. *PLoS Pathogens* **10**.

- Nishimura, K., Kim, S., Zhang, L., Cross, T.A., (2002) The closed state of a H⁺ channel helical bundle combining precise orientational and distance restraints from solid state NMR. *Biochemistry* **41**, 13170-13177.
- O'Toole, T.E., Katagiri, Y., Faull, R.J., Peter, K., Tamura, R., Quaranta, V., Loftus, J.C., Shattil, S.J., Ginsberg, M.H., (1994) Integrin cytoplasmic domains mediate inside-out signal transduction. *The Journal of cell biology* **124**, 1047-1059.
- Ortego, J., Escors, D., Laude, H., Enjuanes, L., (2002) Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. *Journal of Virology* **76**, 11518-11529.
- Ortego, J., Ceriani, J.E., Patiño, C., Plana, J., Enjuanes, L., (2007) Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology* **368**, 296-308.
- Ouyang, B., Xie, S., Berardi, M.J., Zhao, X., Dev, J., Yu, W., Sun, B., Chou, J.J., (2013) Unusual architecture of the p7 channel from hepatitis C virus. *Nature* **498**, 521-525.
- Park, S.H., Mrse, A.A., Nevzorov, A.A., Mesleh, M.F., Oblatt-Montal, M., Montal, M., Opella, S.J., (2003) Three-dimensional structure of the channel-forming transmembrane domain of virus protein "u" (Vpu) from HIV-1. *J Mol Biol* **333**, 409-424.
- Parthasarathy, K., Lin, X., Tan, S.M., Law, S.K., Torres, J., (2008) Transmembrane helices that form two opposite homodimeric interactions: an asparagine scan study of alphaM and beta2 integrins. *Protein Science* **17**, 930-938.
- Parthasarathy, K., Lu, H., Surya, W., Vararattanavech, A., Pervushin, K., Torres, J., (2012) Expression and purification of coronavirus envelope proteins using a modified β -barrel construct. *Protein Expression and Purification* **85**, 133-141.
- Partridge, A.W., Liu, S., Kim, S., Bowie, J.U., Ginsberg, M.H., (2005) Transmembrane domain helix packing stabilizes integrin α IIb β 3 in the low affinity state. *Journal of Biological Chemistry* **280**, 7294-7300.
- Pervushin, K., Riek, R., Wider, G., Wüthrich, K., (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94**, 12366-12371.
- Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F.L., Yu, D., Vararattanavech, A., Tuck, W.S., Ding, X.L., Torres, J., (2009) Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathogens* **5**.
- Pinto, L.H., Lamb, R.A., (2007) Controlling influenza virus replication by inhibiting its proton channel. *Molecular BioSystems* **3**, 18-23.
- Pinto, L.H., Holsinger, L.J., Lamb, R.A., (1992) Influenza virus M2 protein has ion channel activity. *Cell* **69**, 517-528.
- Raamsman, M.J.B., Krijnse Locker, J., De Hooge, A., De Vries, A.A.F., Griffiths, G., Vennema, H., Rottier, P.J.M., (2000) Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. *Journal of Virology* **74**, 2333-2342.
- Renault, M., Saurel, O., Czaplicki, J., Demange, P., Gervais, V., Lohr, F., Reat, V., Piotto, M., Milon, A., (2009) Solution state NMR structure and dynamics of

- KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications. *J Mol Biol* **385**, 117-130.
- Reynolds, J.A., Tanford, C., (1976) Determination of molecular weight of the protein moiety in protein-detergent complexes without direct knowledge of detergent binding. *Proceedings of the National Academy of Sciences* **73**, 4467-4470.
- Root, M.J., Kay, M.S., Kim, P.S., (2001) Protein Design of an HIV-1 Entry Inhibitor. *Science* **291**, 884-888.
- Rost, B., Sander, C., (1993) Prediction of Protein Secondary Structure at Better than 70% Accuracy. *Journal of molecular biology* **232**, 584-599.
- Ruch, T.R., Machamer, C.E., (2011) The hydrophobic domain of infectious bronchitis virus E protein alters the host secretory pathway and is important for release of infectious virus. *Journal of virology* **85**, 675-685.
- Ruch, T.R., Machamer, C.E., (2012) A single polar residue and distinct membrane topologies impact the function of the infectious bronchitis coronavirus E protein. *PLoS Pathogens* **8**.
- Rummel, G., Hardmeyer, A., Widmer, C., Chiu, M.L., Nollert, P., Locher, K.P., Pedruzzi, I.I., Landau, E.M., Rosenbusch, J.P., (1998) Lipidic Cubic Phases: New Matrices for the Three-Dimensional Crystallization of Membrane Proteins. *J Struct Biol* **121**, 82-91.
- Russ, W.P., Engelman, D.M., (2000) The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296**, 911-919.
- Sanders II, C.R., Prestegard, J.H., (1990) Magnetically orientable phospholipid bilayers containing small amounts of a bile salt analogue, CHAPSO. *Biophysical Journal* **58**, 447-460.
- Sanders II, C.R., Schwonek, J.P., (1992) Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid- state NMR. *Biochemistry* **31**, 8898-8905.
- Schneider, D., Engelman, D.M., (2004) Involvement of Transmembrane Domain Interactions in Signal Transduction by α/β Integrins. *Journal of Biological Chemistry* **279**, 9840-9846.
- Schnell, J.R., Chou, J.J., (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451**, 591-595.
- Schuck, P., (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Analytical biochemistry* **320**, 104-124.
- Seddon, A.M., Curnow, P., Booth, P.J., (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et biophysica acta* **1666**, 105-117.
- Semmrigh, M., Smith, A., Feterowski, C., Beer, S., Engelhardt, B., Busch, D.H., Bartsch, B., Laschinger, M., Hogg, N., Pfeffer, K., Holzmann, B., (2005) Importance of integrin LFA-1 deactivation for the generation of immune responses. *J. Exp. Med.* **201**, 1987-1998.
- Senes, A., Gerstein, M., Engelman, D.M., (2000) Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and association with β -branched residues at neighboring positions. *Journal of molecular biology* **296**, 921-936.

- Sharma, M., Yi, M., Dong, H., Qin, H., Peterson, E., Busath, D.D., Zhou, H.X., Cross, T.A., (2010) Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. *Science* **330**, 509-512.
- Sivashanmugam, A., Murray, V., Cui, C.X., Zhang, Y.H., Wang, J.J., Li, Q.Q., (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* **18**, 936-948.
- Spera, S., Bax, A., (1991) Empirical correlation between protein backbone conformation and C.alpha. and C.beta. ¹³C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490-5492.
- Stouffer, A.L., Acharya, R., Salom, D., Levine, A.S., Di Costanzo, L., Soto, C.S., Tereshko, V., Nanda, V., Stayrook, S., DeGrado, W.F., (2008) Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* **451**, 596-599.
- Studier, F.W., (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207-234.
- Suk, J.-E., Situ, A.J., Ulmer, T.S., (2012) Construction of Covalent Membrane Protein Complexes and High-Throughput Selection of Membrane Mimics. *Journal of the American Chemical Society* **134**, 9030-9033.
- Surya, W., Samsó, M., Torres, J., (2013) *Structural and Functional Aspects of Viroporins in Human Respiratory Viruses: Respiratory Syncytial Virus and Coronaviruses*.
- Tamm, L.K., Tatulian, S.A., (1997) Infrared spectroscopy of proteins and peptides in lipid bilayers. *Quarterly reviews of biophysics* **30**, 365-429.
- Tian, Y., Lu, G.J., Marassi, F.M., Opella, S.J., (2014) Structure of the membrane protein MerF, a bacterial mercury transporter, improved by the inclusion of chemical shift anisotropy constraints. *J Biomol NMR* **60**, 67-71.
- Torres, J., Wang, J., Parthasarathy, K., Liu, D.X., (2005) The transmembrane oligomers of coronavirus protein E. *Biophysical Journal* **88**, 1283-1290.
- Torres, J., Parthasarathy, K., Lin, X., Saravanan, R., Kukol, A., Ding, X.L., (2006) Model of a putative pore: The pentameric α -helical bundle of SARS coronavirus E protein in lipid bilayers. *Biophysical Journal* **91**, 938-947.
- Torres, J., Maheswari, U., Parthasarathy, K., Ng, L., Ding, X.L., Gong, X., (2007) Conductance and amantadine binding of a pore formed by a lysine-flanked transmembrane domain of SARS coronavirus envelope protein. *Protein Science* **16**, 2065-2071.
- Travis, M.A., Humphries, J.D., Humphries, M.J., (2003) An unraveling tale of how integrins are activated from within. *Trends Pharmacol Sci* **24**, 192-197.
- Tugarinov, V., Kanelis, V., Kay, L.E., (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protocols* **1**, 749-754.
- Tycko, R., Blanco, F.J., Ishii, Y., (2000) Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole–Dipole Couplings in High-Resolution Biomolecular NMR. *Journal of the American Chemical Society* **122**, 9340-9341.

- Ulmer, T.S., Ramirez, B.E., Delaglio, F., Bax, A., (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *Journal of the American Chemical Society* **125**, 9179-9191.
- Van Horn, W.D., Beel, A.J., Kang, C., Sanders, C.R., (2010) The impact of window functions on NMR-based paramagnetic relaxation enhancement measurements in membrane proteins. *Biochimica et biophysica acta* **1798**, 140-149.
- Vararattanavech, A., Lin, X., Torres, J., Tan, S.M., (2009) Disruption of the integrin alphaLbeta2 transmembrane domain interface by beta2 Thr-686 mutation activates alphaLbeta2 and promotes micro-clustering of the alphaL subunits. *Journal of Biological Chemistry* **284**, 3239-3249.
- Vararattanavech, A., Chng, C.P., Parthasarathy, K., Tang, X.Y., Torres, J., Tan, S.M., (2010) A transmembrane polar interaction is involved in the functional regulation of integrin $\alpha\text{L}\beta\text{2}$. *Journal of molecular biology* **398**, 569-583.
- Vennema, H., Godeke, G.J., Rossen, J.W.A., Voorhout, W.F., Horzinek, M.C., Opstelten, D.J.E., Rottier, P.J.M., (1996) Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *The EMBO journal* **15**, 2020-2028.
- Verdiá-Báguena, C., Nieto-Torres, J.L., Alcaraz, A., DeDiego, M.L., Torres, J., Aguilella, V.M., Enjuanes, L., (2012) Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. *Virology* **432**, 485-494.
- Vinogradova, O., Haas, T., Plow, E.F., Qin, J., (2000) A structural basis for integrin activation by the cytoplasmic tail of the alpha(IIb)-subunit. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 1450-1455.
- Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T.A., Plow, E.F., Qin, J., (2002) A structural mechanism of integrin alpha(IIb)beta(3) "inside-out" activation as regulated by its cytoplasmic face. *Cell* **110**, 587-597.
- Vold, R.R., Prosser, R.S., (1996) Magnetically Oriented Phospholipid Bilayered Micelles for Structural Studies of Polypeptides. Does the Ideal Bicelle Exist? *Journal of Magnetic Resonance, Series B* **113**, 267-271.
- Vold, R.R., Prosser, R.S., Deese, A.J., (1997) Isotropic solutions of phospholipid bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides. *Journal of biomolecular NMR* **9**, 329-335.
- Von Heijne, G., (2007) The membrane protein universe: What's out there and why bother? *Journal of Internal Medicine* **261**, 543-557.
- Wallin, E., Von Heijne, G., (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **7**, 1029-1038.
- Wang, C., Takeuchi, K., Pinto, L.H., Lamb, R.A., (1993) Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. *Journal of Virology* **67**, 5585-5594.
- Wang, W., Zhu, J.Q., Springer, T.A., Luo, B.H., (2011) Tests of Integrin Transmembrane Domain Homo-oligomerization during Integrin Ligand Binding and Signaling. *Journal of Biological Chemistry* **286**, 1860-1867.

- Warschawski, D.E., Arnold, A.A., Beaugrand, M., Gravel, A., Chartrand, É., Marcotte, I., (2011) Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 1957-1974.
- Watts, A., Spooner, P.J.R., (1991) Phospholipid phase transitions as revealed by NMR. *Chemistry and Physics of Lipids* **57**, 195-211.
- Whitmore, L., Wallace, B.A., (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668-W673.
- Wilson, L., Gage, P., Ewart, G., (2006) Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. *Virology* **353**, 294-306.
- Wilson, L., McKinlay, C., Gage, P., Ewart, G., (2004) SARS coronavirus E protein forms cation-selective ion channels. *Virology* **330**, 322-331.
- Xu, X.-P., Case, D., (2001) Automated prediction of ^{15}N , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and $^{13}\text{C}'$ chemical shifts in proteins using a density functional database. *J. Biomol. NMR* **21**, 321-333.
- Yang, J., Ma, Y.-Q., Page, R.C., Misra, S., Plow, E.F., Qin, J., (2009) Structure of an integrin $\alpha\text{IIb}\beta 3$ transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 17729-17734.
- Yu, X., Bi, W., Weiss, S.R., Leibowitz, J.L., (1994) Mouse Hepatitis Virus Gene 5b Protein Is a New Virion Envelope Protein. *Virology* **202**, 1018-1023.
- Yuan, W., Leisner, T.M., McFadden, A.W., Wang, Z., Larson, M.K., Clark, S., Boudignon-Proudhon, C., Lam, S.C., Parise, L.V., (2006) CIB1 is an endogenous inhibitor of agonist-induced integrin $\alpha\text{IIb}\beta 3$ activation. *J Cell Biol* **172**, 169-175.
- Zhirnov, O.P., (1990) Solubilization of matrix protein M1/M from virions occurs at different pH for orthomyxo- and paramyxoviruses. *Virology* **176**, 274-279.
- Zhirnov, O.P., (1992) Isolation of matrix protein M1 from influenza viruses by acid-dependent extraction with nonionic detergent. *Virology* **186**, 324-330.
- Zhou, F.X., Merianos, H.J., Brunger, A.T., Engelman, D.M., (2001a) Polar residues drive association of polyleucine transmembrane helices. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2250-2255.
- Zhou, H.-X., Cross, T.A., (2013) Influences of Membrane Mimetic Environments on Membrane Protein Structures. *Annual Review of Biophysics* **42**, 361-392.
- Zhou, Y., Lau, F.W., Nauli, S., Yang, D., Bowie, J.U., (2001b) Inactivation mechanism of the membrane protein diacylglycerol kinase in detergent solution. *Protein Science* **10**, 378-383.
- Zhu, J., Luo, B.H., Barth, P., Schonbrun, J., Baker, D., Springer, T.A., (2009) The structure of a receptor with two associating transmembrane domains on the cell surface: integrin $\alpha\text{IIb}\beta 3$. *Molecular cell* **34**, 234-249.

APPENDIX

Figure S1. Summary of NOE connectivities. Sequential and medium-range NOE connectivities in (A) α L-TM in bicelles, (B) β 2-TM in bicelles, and (C) E_{TR}. NOE connectivities are displayed as lines below the respective residues.

Figure S2. Stability of E_{TR} in SDS micelle. The stability of E_{TR} backbone structure when reconstituted in 50 mM SDS was examined by HNOE experiment.

Figure S3. Screening of CoV E protein constructs. We screened SARS-CoV E, IBV E, TGEV E, and MHV E protein constructs with various truncations at NTU-SBS Protein Production Platform. Coloured bars indicate sequence coverage of each construct, and colour-coded for qualitative expression yield: red indicate no expression, orange and yellow indicate low expression, and green indicate good expression yield.

Density matching in sedimentation equilibrium analytical ultracentrifugation

Sedimentation equilibrium analysis of membrane proteins is complicated by the presence of detergent which typically has a different density than the membrane protein being studied. Density of proteins, or more often expressed in inverse density or partial specific volume, is on average $0.735 \text{ cm}^3/\text{g}$ (Durchschlag, 1986), whereas detergents are generally less dense, for example DDM is $0.826 \text{ cm}^3/\text{g}$ and C8E5 is $0.993 \text{ cm}^3/\text{g}$ (Fleming, 2005). Thus bound detergent alters the density of the protein-detergent complex. The density of the protein-detergent-complex can be calculated as a linear combination of the density of protein and of detergent, with constants that depend on the molar ratio of detergent bound to the protein. The amount of bound detergent needs to be determined empirically, for example using radioactively labelled detergent probes (le Maire et al., 2000).

Alternatively, the contribution of detergent can be made negligible by matching the density of the buffer to that of the detergent (“density matching”) (Reynolds and Tanford, 1976). This is usually done by adding a small amount of D_2O into the buffer. Other density-altering substances can also be used, e.g. glycerol or sucrose (Mayer et al., 1999), but these may affect the protein and detergent hydration (Fleming, 2005). However, only a few detergents are amenable to density matching by D_2O , i.e. those with densities between 100% water ($1.0 \text{ cm}^3/\text{g}$) and 100% D_2O ($0.9 \text{ cm}^3/\text{g}$). These include C8E5 (0% D_2O), C14SB (~15% D_2O), and DPC (~55% D_2O) (Fleming, 2005). Consequently, detergents with higher density than D_2O cannot be density-matched, e.g. DDM ($0.826 \text{ cm}^3/\text{g}$) and SDS ($0.870 \text{ cm}^3/\text{g}$). The density of LMPG was calculated by using Durchschlag method (Durchschlag, 1986) to be $0.840 \text{ cm}^3/\text{g}$ (Rieth M. D. and K. J. Glover, personal communication), and thus also cannot be density-matched using D_2O .

Under intense centrifugal field, detergent micelles sediment in buffers with lower density, and float in buffers with higher density. Density matching is done by centrifuging micelles in buffer containing different D_2O concentrations. The neutral density, i.e. where micelles neither sediment nor float, can be found by interpolation of the micelle distribution slopes. Examples of density matching for detergents used in this study are shown for DPC/SDS (Fig. S3A) and C14SB (Fig. S3B).

Figure S4. Examples of density matching experiment. Sedimentation equilibrium analytical ultracentrifugation for detergent micelles were monitored using interference optics, as detergents do not absorb in the UV/Vis range. Interference measurement results in fringe shift units, which is directly proportional to solute concentration. Sedimentation equilibrium data are shown for (A) DPC/SDS at 4:1 molar ratio and (B) C14SB in three D₂O concentrations (left to right). In the three chosen D₂O concentrations, the detergent micelles can be observed to either sediment, almost at neutral buoyancy, or float. The slope was then plotted against the D₂O concentration (rightmost panel), fitted linearly, and the X-axis intercept is interpreted as the density-matched D₂O concentration, i.e. ~66% for DPC/SDS, and 30.3% for C14SB.

PUBLICATIONS

Surya W, Li Y, Torres J (2017) Pentameric Model of SARS-CoV E Protein. Manuscript in preparation.

Li Y, Surya W, Claudine S, Torres J (2014) Structure of a Conserved Golgi Complex-targeting Signal in Coronavirus Envelope Proteins. *Journal of Biological Chemistry* 289: 12535-12549.

Surya W, Samso' M, Torres J (2013) Structural and Functional Aspects of Viroporins in Human Respiratory Viruses: Respiratory Syncytial Virus and Coronaviruses.

Surya W, Li Y, Millet O, Diercks T, Torres J (2013) Transmembrane and Juxtamembrane Structure of α L Integrin in Bicelles. *PLoS ONE* 8: e74281.

Parthasarathy K, Lu H, Surya W, Vararattanavech A, Pervushin K, et al. (2012) Expression and purification of coronavirus envelope proteins using a modified β -barrel construct. *Protein Expression and Purification* 85: 133-141.

POSTER PRESENTATIONS

Surya W, Li Y, Claudine, S, Torres J. Structure of a Conserved Golgi Complex-targeting Signal in Coronavirus Envelope Proteins. Poster session presented at the 4th Asia Pacific Protein Association (APPA) Conference; 2014 May 17-20; Jeju, South Korea.