

Structural characterisation of human tetraspanins and their interaction with cholesterol and gangliosides

Luke Matthew Broadbent
Doctor of Philosophy

Aston University
May 2022

©Luke Matthew Broadbent, 2022

Luke Matthew Broadbent asserts their moral right to be identified as the author of this thesis

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright belongs to its author and that no quotation from the thesis and no information derived from it may be published without permission or acknowledgement

Aston University

Structural characterisation of human tetraspanins and their interaction with cholesterol and gangliosides

Luke Matthew Broadbent

Doctor of Philosophy

2022

Thesis Summary

Tetraspanins are a family of membrane proteins that play a role in various functions, such as cell migration, signal transduction and intracellular trafficking. They are organisers within the membrane, forming tetraspanin-enriched microdomains (TEMs) comprised of tetraspanins, partner proteins, cholesterol and gangliosides.

Knowledge of the precise interactions between these molecules is limited. Understanding more about these interactions would add to the body of knowledge about how TEMs are formed. With this knowledge these interactions could be disrupted to prevent negative biological events associated with TEMs, such as microbial infections. To study these interactions CD81, which interacts with hepatitis C virus, was used as a model tetraspanin.

The purification of CD81 expressed in *Pichia pastoris* in styrene-maleic acid lipid particles (SMALPs) was optimised by including imidazole in the binding buffer and increasing the concentration of imidazole in the purification wash steps. CD81 mutants were created to study cholesterol-dependent conformational change using electron paramagnetic resonance. Protein-ligand docking was used to investigate the interaction between the two tetraspanins, CD81 and CD82, and the sugar residues in gangliosides. Asp122 in CD81 and Asp37 and Ser135 in CD82 were identified as interaction sites.

The basic structure of tetraspanins appears to be ubiquitous but a variable region in the large extracellular loop (LEL) requires further research because this is a crucial area for tetraspanin interactions. LELs of some human tetraspanins were modelled and their disulfide bond arrangement was analysed, finding a different arrangement in the subset of tetraspanins referred to as TspanC6-CxCs.

Sequence and structural alignment of human tetraspanins highlighted conserved residues within different structural regions that were used as anchor residues to build a universal amino acid numbering system.

Keywords: Tetraspanin, tetraspanin-enriched microdomains, cholesterol, ganglioside, multiple sequence alignment, protein-ligand docking, protein purification

This thesis is dedicated to my Nan

Acknowledgements

I would like to thank Professor Roslyn Bill for being crazy enough to give me the opportunity to do this PhD. I would also like to thank her and my associate supervisor, Dr Alice Rothnie, for their continued support and encouragement throughout my PhD. Thanks to Dr John Simms, who essentially became a third supervisor during the pandemic, for offering advice on all computational work. I knew next to nothing about it two years ago so without his help life would've been a lot harder.

I'd like to thank Dr Hoor Ayub and Dr Michelle Clare for their guidance at the start of my PhD. I'm grateful to Hoor for showing me everything to do with growing, expressing and purifying CD81. Thanks to Dr Philip Kitchen who has always been willing to discuss my project and offer help.

I'd like to thank everyone who has worked in MB347 during my PhD, particularly Thanos, Ahmed, Ravneet, Ben and Lucas. I'd like to thank everyone who has been a member of the Aston Centre for Membrane Proteins and Lipids Research (AMPL), particularly Aneel, Olivia, Cristina, Alexis and Peer. Anyone named in this paragraph has made the experience better because of the great conversations we have had and continue to have.

I'd like to thank Jiteen Ahmed, and the rest of the technicians, who were always available to help with anything that was causing a problem in the lab. A big thanks to all the cleaners who tirelessly walked round the building cleaning doors and door handles so that it would be safe enough for us to work when we were allowed back in after lockdown.

Finally, thanks to all my family and friends who have had to put up with a lot longer than everyone mentioned above.

Abbreviations

ADAM10 – A disintegrin and metalloproteinase domain-containing protein 10

ANOVA – Analysis of variance

AOX – Alcohol oxidase

APS – Ammonium persulfate

AQP1 – Aquaporin 1

β -OG – Octyl β -D-glucopyranoside

BCR – B cell receptor

BLAST – Basic local alignment search tool

BLOSUM – BLOcks SUBstitution Matrix

BMGY – Buffered complex glycerol medium

BMMY – Buffered complex methanol medium

CD81 – Cluster of differentiation 81

CHO – Chinese hamster ovary cells

CHS – Cholesteryl hemisuccinate

CLDN – Claudin

CMC – Critical micelle concentration

CRAC – Cholesterol Recognition/interaction Amino acid Consensus sequence

Cryo-EM – Cryogenic electron microscopy

cwEPR – Continuous-wave electron paramagnetic resonance

DDM – *n*-dodecyl- β -D-maltopyranoside

DEER – Double electron-electron resonance

DM – Decyl- β -D-maltopyranoside

DMEM – Dulbecco's modified eagle medium

E. coli – *Escherichia coli*

ECL – Enhanced chemiluminescence

EDTA – Ethylenediaminetetraacetic acid

EGFR – Epidermal growth factor receptor

ELISA – Enzyme-linked immunosorbent assay

EPR – Electron paramagnetic resonance

FBS – Fetal bovine serum

GD1a – Disialogangliode-GD1a

GD1b – Disialoganglioside-GD1b

GD3 – Ganglioside GD3

GFP – Green fluorescence protein

GIPC – Glycosylionositolphosphoceramides

GM1 – Monosialotetrahexosylganglioside

GM2 – Ganglioside GM2

GM3 – Monosialodihexosylganglioside

GPCR – G protein-coupled receptor

GT1b – Trisialoganglioside-GT1b

HCV – Hepatitis C virus

HCVpp – Hepatitis C virus pseudoparticles

HEK293 – Human embryonic kidney 293 cells

HEPES – 2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid

HIV – Human immunodeficiency virus

hMOR – Human mu opioid receptor

HPV – Human papillomavirus

HRP – Horseradish peroxidase

IgG – Immunoglobulin G

LacCer – Lactosylceramide

LB – Luria-Bertani

LDL-R – Low-density lipoprotein receptor

LEL – Large extracellular loop

M β CD – Methyl- β -cyclodextrin

MSA – Multiple sequence alignment

MTSSL – Methanethiosulfanate spin label

Ni-NTA – Nitriloacetic acid

NMR – Nuclear magnetic resonance

OCLN – Occludin

P. pastoris – *Pichia pastoris*

p-null – Palmitoylation-null

PAGE – Polyacrylamide gel electrophoresis

PBS – Phosphate buffered saline

PBS-T – Phosphate buffered saline Tween-20

PCR – Polymerase chain reaction

PDB – Protein data bank

PEG – Polyethylene glycol

PEI – Polyethylenimine

PELDOR – Pulsed electron-electron double resonance

PKC- α – Protein kinase C alpha

PMEL – Premelanosome protein

POPC – 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

PRPH2 – Peripherin-2

PSA – Pairwise sequence alignment

PVDF – Polyvinylidene fluoride

QMEAN – Qualitative model energy analysis

RIPA – Radioimmunoprecipitation assay buffer

RMSD – Root-mean-square deviation

ROM1 – Rod outer segment membrane protein 1

S. cerevisiae – *Saccharomyces cerevisiae*

SDS – Sodium dodecyl sulfate

SDSL – Site-directed spin labelling

sE2 – Soluble E2

SEL – Small extracellular loop

SMA – Styrene-maleic acid

SMALP – Styrene-maleic acid lipid particles

SMA_{nh} – Styrene-maleic anhydride

smFRET – Single molecule fluorescence resonance energy transfer

SMI – Styrene-maleimide

SR-B1 – Scavenger receptor class B type 1

TEM – Tetraspanin-enriched microdomain

Tet8 – Tetraspanin-8

TLC – Thin layer chromatography

TM – Transmembrane domain

Tspan – Tetraspanin

UP1a – Uroplakin-1a

YPD – Yeast peptone dextrose

YNB – Yeast nitrogen base

Units

Å	Angstrom
°C	Celsius
g	Gram
g	Gravitational force
h	Hour
kb	Kilobases
kcal/mol	Kilocalorie per mole
kDa	Kilo Dalton
µg	Microgram
µl	Microlitre
mg	Milligram
ml	Millilitre
mM	Millimole
M	Mole
ng	Nanogram
nm	Nanometer
psi	Pounds per square inch
rpm	Revolutions per minute
V	Volts

Table 1. Standard amino acid abbreviations and their structures. List of standard amino acid abbreviations used interchangeably throughout this thesis (structures from BioRender.com)

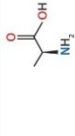
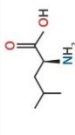
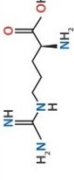
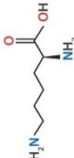
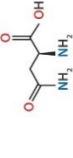
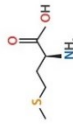
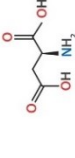
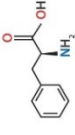
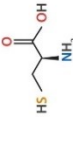
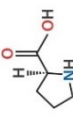
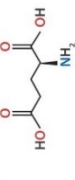
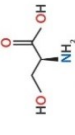
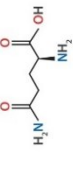
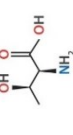
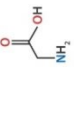
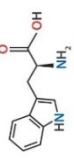
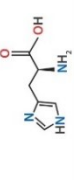
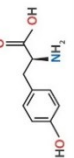
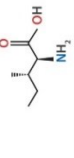
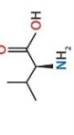
Amino acid	3-letter	1-letter	Structure	Amino acid	3-letter	1-letter	Structure
Alanine	Ala	A		Leucine	Leu	L	
Arginine	Arg	R		Lysine	Lys	K	
Asparagine	Asn	N		Methionine	Met	M	
Aspartic acid	Asp	D		Phenylalanine	Phe	F	
Cysteine	Cys	C		Proline	Pro	P	
Glutamic acid	Glu	E		Serine	Ser	S	
Glutamine	Gln	Q		Threonine	Thr	T	
Glycine	Gly	G		Tryptophan	Trp	W	
Histidine	His	H		Tyrosine	Tyr	Y	
Isoleucine	Ile	I		Valine	Val	V	

Table of Contents

Abbreviations	5
List of Figures	16
List of Tables	25
Chapter 1 – Introduction.....	26
1.1 The plasma membrane.....	26
1.1.1 Phospholipids.....	26
1.1.2 Glycolipids.....	27
1.1.3 Cholesterol.....	29
1.2 Membrane proteins.....	31
1.3 Tetraspanins.....	33
1.3.1 Multiple sequence alignment of tetraspanin amino acid sequences	35
1.3.2 Generic amino acid numbering for protein families.....	36
1.4 The tetraspanin, CD81.....	38
1.5 Structures of tetraspanins.....	39
1.5.1 Electron paramagnetic resonance to probe protein conformational change.....	41
1.5.2 Protein structure prediction	43
1.6 Molecular Function of Tetraspanins.....	44
1.6.1 The role of tetraspanins in trafficking.....	44
1.6.2. Tetraspanins interaction with integrins	46
1.6.3 The role of CD81 and CD9 in cell-cell fusion	47
1.7 Tetraspanin-enriched microdomains.....	47
1.7.1 Tetraspanin interactions with gangliosides	48
1.7.2 Tetraspanin interactions with cholesterol.....	50
1.7.3 Protein-ligand docking.....	53
1.8 Expression of membrane proteins	54
1.8.1 Expression of recombinant tetraspanins in <i>Pichia pastoris</i>	54
1.8.2 Expression of recombinant tetraspanins in mammalian cells.....	56
1.9 Solubilisation of tetraspanins	57

1.9.1 Solubilisation of tetraspanins with detergents.....	57
1.9.2 Solubilisation of tetraspanins with styrene-maleic acid	60
1.10 Aims and objectives.....	63
Chapter 2 – Materials and Methods	64
2.1 Materials.....	64
2.2 <i>Pichia pastoris</i> growth.....	65
2.2.1 Stock solutions and growth media	65
2.2.3 Expression of recombinant CD81-p-null	67
2.3 <i>Pichia pastoris</i> membrane solubilisation	67
2.3.1 <i>P. pastoris</i> membrane preparation	67
2.4 CD81 Purification.....	69
2.4.1 SMALP-CD81 Purification	69
2.4.2 DDM-CD81 Purification	70
2.5 SDS-PAGE Gel Electrophoresis	71
2.6 Western Blotting	72
2.7 Site-Directed Mutagenesis.....	74
2.8 Transformations.....	76
2.8.1 <i>E. coli</i> Transformation	76
2.8.2 Yeast Transformation	77
2.9 CD81 expression in HEK293T cells.....	78
2.9.1 HEK293T cell growth	78
2.9.2 PEI transfection of CD81 in HEK293 cells	78
2.9.3 HEK293 cell lysis	79
2.10. Tetraspanin sequence and structural alignment.....	80
2.10.1 Alignment of human tetraspanin TMs	80
2.10.2 Alignment of human tetraspanins LEL regions	81
2.10.3 Alignment of 3D structures of tetraspanins	82
2.11 Protein structure prediction of human tetraspanin LELs	83
2.12 Disulfide bond prediction in human tetraspanin LELs.....	84

2.13. Homology modelling of CD82	85
2.14. Docking ganglioside sugars to tetraspanins.....	86
2.14.1 Prediction of tetraspanin binding pockets	86
2.14.2 Docking of ganglioside sugars with Galaxy7TM	87
Chapter 3 – Establishing a universal numbering system for human tetraspanins	89
3.1 Sequence and structural alignment of the TM1 domains of human tetraspanin	89
3.2 Sequence and structural alignment of the TM2 domains of human tetraspanins.....	93
3.3 Sequence and structural alignment of the TM3 domains of human tetraspanins.....	96
3.4 Sequence and structural alignments of the TM4 domains of human tetraspanins....	100
3.5 Anchor residue in the 4 transmembrane regions of tetraspanins	104
3.6 Sequence and structural alignments of the α helix of the large extracellular loop of human tetraspanins	105
3.7 Sequence and structural alignments of the β helix of the large extracellular loop of human tetraspanins	106
3.8 Sequence and structural alignments of the ϵ helix of the large extracellular loop of human tetraspanins	108
3.9 A universal tetraspanin numbering system	111
3.10 Conclusion.....	112
Chapter 4 – Characterisation of the loops in the LELs of tetraspanins	113
4.1 Sequence Alignment of the loops in the LELs of tetraspanins.....	113
4.2 Cysteine arrangement in the LEL of human tetraspanins.....	115
4.3 Disulfide bond formation in the LEL of CD81	116
4.4 Sequence conservation in the loops of the LEL of TspanC4s	119
4.5 Structural characterisation and sequence conservation of the LEL in TspanC6-CCs120	
4.6 Structural characterisation and sequence conservation of the LEL in TspanC6-CxCs	122
4.7 Structural characterisation and sequence conservation of the LEL in TspanC6-CxxxCs	126
4.7 Structural characterisation and sequence conservation of the LEL in TspanC6 β s ...	129
4.7 Structural characterisation and sequence conservation of the LEL in TspanC8s	132

4.8 Conservation in the SEL of tetraspanins by subfamily	135
4.8 Conclusion.....	137
Chapter 5 – Expression and purification of the tetraspanin, CD81.....	139
5.1 Solubilisation Efficiency of Autoclaved SMA2000	139
5.2 Optimisation of the Purification of CD81-p-null Solubilised with SMA2000.....	141
5.3 CD81-p-null Mutations for PELDOR Experiments.....	148
5.4 Purification of CD81-p-null-C104A-D128C Double Mutant.....	156
5.5 Small Scale Size Exclusion Chromatography of CD81-C104A-D128C	162
5.6 Conclusion.....	166
Chapter 6 – Interaction between ganglioside sugars and tetraspanins.....	167
6.1 Docking ganglioside sugars to closed CD81	167
6.2 Mutating Asp122 in CD81 in <i>P. pastoris</i>	171
6.3 Expression of CD81 in HEK293T cells.....	175
6.4 Docking ganglioside sugars to closed CD82.....	177
6.5 Conclusion.....	184
Chapter 7 – Discussion	185
7.1 Universal tetraspanin numbering system	185
7.2 Greater sequence conservation in loops of the LEL is apparent when categorised into subfamilies based on cysteine arrangement.....	189
7.3 Disulfide bond pattern differs by tetraspanin subfamily based on cysteine arrangement in the LEL	190
7.4 Solubilisation of CD81 with autoclaved SMA2000 is comparable to solubilisation with refluxed SMA2000.....	193
7.5 Purification of CD81-SMALPs with Ni-NTA resin can be improved by increasing imidazole in the wash steps and adding imidazole to the binding buffer	195
7.6 Purification of CD81-C104A-D128C improves when solubilised with DDM instead of SMA2000	196
7.7 Future work on probing the interaction of cholesterol with tetraspanins and its significance for conformational change.....	197
7.8 Asp122, in CD81, and Asp37 & Ser135, in CD82, bind ganglioside sugar residues.	199

7.9 Overall Conclusion	202
Chapter 8 – References	204

List of Figures

Chapter 1

Figure 1.1	The structure and arrangement of phospholipids in the plasma membrane	27
Figure 1.2	The chemical structure of the ganglioside GM3	28
Figure 1.3	The chemical structure of cholesterol	30
Figure 1.4	Illustration of how membrane proteins interact with the plasma membrane	31
Figure 1.5	Illustration of the main categories of membrane protein	32
Figure 1.6	Basic structure of a tetraspanin	33
Figure 1.7	Formation of disulfide bonds in proteins	34
Figure 1.8	BLOSUM62 score matrix	36
Figure 1.9	Generic numbering of amino acids in GPCRs	37
Figure 1.10	Hepatitis C viral entry showing the importance of CD81	38
Figure 1.11	Solved full length tetraspanin structures	39
Figure 1.12	Early full length tetraspanin structures	40
Figure 1.13	Attachment of MTSSL to a cysteine residue	42
Figure 1.14	The interaction of CD81 with CD19 in resting and activated B cells	45
Figure 1.15	The interaction of CD151 with integrins	46
Figure 1.16	Illustration of a tetraspanin-enriched microdomain	48
Figure 1.17	The chemical structure of the ganglioside GM2	49
Figure 1.18	The interaction between cholesterol and CD81 in the CD81 cavity	51
Figure 1.19	Glucose molecule docked to closed CD81 in Galaxy7TM	54
Figure 1.20	Illustration of the process of growing <i>P. pastoris</i> cells for membrane protein overexpression	55
Figure 1.21	Formation of detergent micelles in solution	58

Figure 1.22	Illustration of how high concentration of detergent is required to solubilise membrane proteins	59
Figure 1.23	The chemical structure of n-dodecyl- β -D-maltopyranoside	60
Figure 1.24	Hydrolysis of styrene-maleic anhydride to styrene-maleic acid	61
Figure 1.25	Solubilisation of membrane proteins using SMA	62
Chapter 2		
Figure 2.1	The pPICZB plasmid map	64
Figure 2.2	The pEF6.A plasmid map	65
Figure 2.3	The CD81-p-null amino acid sequence	74
Figure 2.4	The input page on Robetta with TrRosetta	83
Figure 2.5	The input page for SSBondPre	84
Figure 2.6	The SSBondPre prediction of disulfide bonds in ROM1	85
Figure 2.7	The input page for DeepSite	86
Figure 2.8	The DeepSite results page	87
Figure 2.9	The input page for Galaxy7TM	88
Chapter 3		
Figure 3.1	Structure of CD81 with regions labelled	89
Figure 3.2	Determination of the TM1 region of human tetraspanins	90
Figure 3.3	Sequence alignment and conservation of the TM1 domain in human tetraspanins	91
Figure 3.4	BLOSUM62 scores for each amino acid position in TM1 of human tetraspanins	92
Figure 3.5	Sequence and structural alignment of the TM1 domain of human tetraspanins	93
Figure 3.6	Determination of the TM2 region of human tetraspanins	94
Figure 3.7	Sequence alignment and conservation of the TM2 domain in	

	human tetraspanins	95
Figure 3.8	Sequence and structural alignment of the TM2 domain of human tetraspanins	96
Figure 3.9	Determination of the TM3 region of human tetraspanins	97
Figure 3.10	Sequence alignment and conservation of the TM3 domain in human tetraspanins	98
Figure 3.11	BLOSUM62 scores for each amino acid position in TM3 of human tetraspanins	99
Figure 3.12	Sequence and structural alignment of the TM3 domain of human tetraspanins	100
Figure 3.13	Determination of the TM4 region of human tetraspanins	101
Figure 3.14	Sequence alignment and conservation of the TM4 domain in human tetraspanins	102
Figure 3.15	Sequence alignment and residues with >70% conservation of the TM4 domain in human tetraspanins	103
Figure 3.16	Sequence and structural alignment of the TM4 domain of human tetraspanins	104
Figure 3.17	Ball and stick representation of CD81 showing the position of the anchor residues in the TMs of human tetraspanins	105
Figure 3.18	Sequence and structural alignment of the α helix in the large extracellular loop of human tetraspanins	106
Figure 3.19	Sequence and structural alignment of the β helix in the large extracellular loop of human tetraspanins	107
Figure 3.20	Sequence alignment and conservation of the ϵ helix in the large extracellular loop of human tetraspanins	109
Figure 3.21	BLOSUM62 scores for each amino acid position in the ϵ helix	

	in the large extracellular loop of human tetraspanins	110
Figure 3.22	Sequence and structural alignment of the ϵ helix in the large extracellular loop of human tetraspanins	111
Figure 3.23	A universal tetraspanin residue numbering system	112
Chapter 4		
Figure 4.1	Location of the loop in the LEL of closed CD81	113
Figure 4.2	Sequence alignment and conservation of the loops between the β helix and the ϵ helix in the LELs of human tetraspanins	114
Figure 4.3	The arrangement of cysteines in the LELs of human tetraspanins	116
Figure 4.4	Disulfide bond pattern in the loop of the LEL in the closed structure of CD81	117
Figure 4.5	Disulfide bond pattern in the loop of the LEL in the open structure of CD81	118
Figure 4.6	Disulfide bond pattern in the loop of the LEL in the closed structure of CD81 and sequence alignment of the loops of TspanC4s	119
Figure 4.7	Modelled structure of the LEL of CD151 with secondary structure prediction	120
Figure 4.8	Disulfide bond pattern in the loop of the LEL in the modelled structure of CD151 and sequence alignment of the loops of TspanC6-CCs	122
Figure 4.9	Modelled structure of the LEL of CD37 with secondary structure prediction	123
Figure 4.10	Disulfide bond pattern in the loop of the LEL in the modelled structure of CD37 and sequence alignment of the loops of TspanC6-CxCs	124

Figure 4.11	Disulfide bond pattern in the loop of the LEL in the modelled structure of CD82, Tspan19 and Tspan8	125
Figure 4.12	Modelled structure of the LEL of Tspan13 with secondary structure prediction	127
Figure 4.13	Disulfide bond pattern in the loop of the LEL in the modelled structure of Tspan13 and sequence alignment of the loops of TspanC6-CxxxCs	128
Figure 4.14	Modelled structure of the LEL of ROM1 with secondary structure prediction	129
Figure 4.15	Disulfide bond pattern in the loop of the LEL in the modelled structure of ROM1 and sequence alignment of the loops of TspanC6βs	131
Figure 4.16	Modelled structure of the LEL of Tspan15 with secondary structure prediction	132
Figure 4.17	Comparison of the modelled Tspan15 LEL and the solved structure of Tspan15	133
Figure 4.18	Disulfide bond pattern in the loop of the LEL in the modelled structure of Tspan15 and sequence alignment of the loops of TspanC8s	134
Figure 4.19	Sequence alignment and conservation of the SELs of human tetraspanins	135
Figure 4.20	Sequence alignment and conservation of the SELs by subfamily of human tetraspanins	136
Chapter 5		
Figure 5.1	Comparison of the solubilisation efficiency of autoclaved SMA2000 compared with refluxed SMA2000	140

Figure 5.2	Comparison of SMA2000 stuck to the bottom of the bottle after first the autoclave cycle	141
Figure 5.3	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in <i>P. pastoris</i>	142
Figure 5.4	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in <i>P. pastoris</i> with the volume of the wash steps doubled	143
Figure 5.5	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in <i>P. pastoris</i> with the concentration of imidazole in the wash steps increased	144
Figure 5.6	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in <i>P. pastoris</i> with 20 mM imidazole in the binding buffer and the volume of the wash steps doubled	145
Figure 5.7	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in <i>P. pastoris</i> with 20 mM imidazole in the binding buffer and the increased imidazole in the wash steps	146
Figure 5.8	Western blot analysis of the optimised CD81-p-null purification	147
Figure 5.9	SDS-PAGE analysis of a small scale SMA2000 purification of CD81 using the optimised method with increased elution steps	148
Figure 5.10	Agarose gel of PCR product after site-directed mutagenesis of CD81-p-null to create single cysteine mutants	150
Figure 5.11	Single cysteine mutant sequences compared to human CD81	151
Figure 5.12	Low salt LB agar plates with transformed cells after site-directed mutagenesis	152
Figure 5.13	Double cysteine mutant sequences compared to human CD81	153

Figure 5.14	Western blot analysis of CD81-p-null-C97A and CD81-p-null-C97A-D128C small scale expression	154
Figure 5.15	Western blot analysis of CD81-p-null-C104A and CD81-p-null-C104A-D128C small scale expression	155
Figure 5.16	Western blot analysis of CD81-p-null-C97A-D128C and CD81-p-null-C104A-D128C medium scale expression	156
Figure 5.17	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i>	157
Figure 5.18	SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i> solubilised at 120 mg/ml membrane concentration	158
Figure 5.19	SDS-PAGE analysis of a small scale SMA2000 purification with Qiagen Ni-NTA resin of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i>	159
Figure 5.20	SDS-PAGE analysis of a small scale SMA2000 purification with Generon Ni-NTA resin of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i>	160
Figure 5.21	SDS-PAGE analysis of a small scale DDM purification of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i>	161
Figure 5.22	SDS-PAGE analysis of a small scale DDM purification with double the volume of solubilisation buffer of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i>	162
Figure 5.23	Size exclusion chromatogram of CD81-p-null-C104A-D128C concentrated to 500 μ l	163

Figure 5.24	SDS-PAGE analysis of a small scale size exclusion chromatography of CD81-p-null-C104A-D128C expressed in <i>P. pastoris</i>	164
Figure 5.25	Size exclusion chromatogram of CD81-p-null-C104A-D128C concentrated to 200 μ l	165
Figure 5.26	SDS-PAGE analysis of a small scale size exclusion chromatography of CD81-p-null-C104A-D128C concentrated to 200 μ l expressed in <i>P. pastoris</i>	165
Chapter 6		
Figure 6.1	Binding pocket of closed CD81	163
Figure 6.2	Interaction of ganglioside sugars with closed CD81	169
Figure 6.3	Overlaying β -D-galactose and lactose with closed CD81	170
Figure 6.4	Agarose gels of the PCR product after site-directed mutagenesis of CD81-p-null to mutate Asp122 to Ala	172
Figure 6.5	Asp122Ala mutant sequence compared to wild-type human CD81	173
Figure 6.6	Transformed CD81-p-null-D122A mutants on YPD agar plates with varying concentrations of zeocin	174
Figure 6.7	Expression of CD81-p-null-D122A mutants	175
Figure 6.8	Expression of wild-type CD81 in HEK293T cells	177
Figure 6.9	Homology model of closed CD82 in the membrane	178
Figure 6.10	Ramachandran plot and QMEAN Z-scores of CD82 homology model	179
Figure 6.11	QMEANBrane analysis of the CD82 homology model	180
Figure 6.12	Binding pocket of CD82	181
Figure 6.13	Interaction of ganglioside sugars with closed CD82	182
Figure 6.14	Overlaying β -D-galactose and lactose interacting with closed CD82	183

Chapter 7

Figure 7.1	Position of anchor residues in the TM and LEL regions of CD81	186
Figure 7.2	The proposed new naming system of the structures in the loop region in the LEL of human tetraspanins	192

List of Tables

Table 1	Standard amino acid abbreviation and their structures	10
Chapter 1		
Table 1.1	Human tetraspanins and their interactions	52
Chapter 2		
Table 2.1	SDS-PAGE 12% separating gel recipe	71
Table 2.2	SDS-PAGE 4% stacking gel recipe	72
Table 2.3	Primary and secondary antibodies for Western blots	73
Table 2.4	Primers used in mutations of CD81-p-null	75
Table 2.5	Recipe for the PCR reaction for site directed mutagenesis	75
Table 2.6	PCR reaction temperature and timings for site-directed mutagenesis	76
Table 2.7	Uniprot codes for human tetraspanins	80
Table 2.8	Amino acid groups in the Sequence Manipulation Suite: Color Align Conservation	81
Chapter 6		
Table 6.1	Docking energy of galactose and lactose with closed CD81	171

Chapter 1 – Introduction

This thesis is an investigation into the structural characteristics of human tetraspanins and their interactions with cholesterol and gangliosides in the plasma membrane. Modelled structures and sequence conservation were investigated to look at the structural characteristic of human tetraspanin large extracellular loops (LELs) and create a universal residue numbering system. Mutants of CD81 were created to investigate its conformational change using electron paramagnetic resonance (EPR) and protein-ligand docking was used to find residues in CD81 and CD82 that interact with the sugar molecules that make up part of a ganglioside.

1.1 The plasma membrane

The plasma membrane is a bilayer of lipids of various classes, although approximately 40% of the bilayer is made up of phospholipid in terms of mass forming a phospholipid bilayer that is approximately 5-10 nm in depth (van Meer & de Kroon, 2011). The primary function of the membrane is to separate the intracellular and extracellular regions of a cell, providing protection to the cell from external forces and playing a role in allowing specific molecules to enter and exit the cell (Chernomordik et al., 1995). The three main classes of lipids in the membrane are phospholipids, glycolipids and cholesterol (Andersen & Koeppe, 2007).

1.1.1 Phospholipids

Phospholipids comprise a negatively charged phosphate-linked head group attached to glycerol, forming the hydrophilic part of the lipid, with a hydrophobic tail consisting of two fatty acid acyl chains (Figure 1.1A). Phospholipids spontaneously arrange themselves into a bilayer in water because the hydrophobic tails bury themselves to avoid contact with water, while the hydrophilic head interacts with the water via hydrogen bonds. The hydrophobic tails point inwards into the membrane facing each other while the hydrophilic head group points outwards towards either the intracellular region or extracellular region creating a hydrophobic region in the centre of the membrane (Seelig & Seeling, 1974) (Figure 1.1B).

The acyl chains in the tail can exist in either a saturated or unsaturated state. Unsaturated chains have one or more double bonds producing a kink in the chain which increases the flexibility and fluidity of the acyl chains (de Kroon, 2007). The level of unsaturation in the acyl chains, along with the length of the acyl chains and the different head groups attached to the phosphate region contribute the high diversity in phospholipids (Matsumoto et al., 2006).

Ever since the early 1970s the membrane has been seen in terms of a fluid mosaic model (Singer & Nicholson, 1972). The fluid mosaic model is so called because from above the membrane resembles a mosaic because it consists of phospholipids, glycolipids, proteins, and

cholesterol. The membrane is also considered to be a fluid region, although its fluidity is dependent on numerous factors, such as how much cholesterol is in the membrane, the presence of lipoprotein complexes or its interaction with either the intracellularly located cytoskeleton or the extracellular matrix (Andersen & Koeppe, 2007; Engelman, 2005; Vereb et al., 2003).

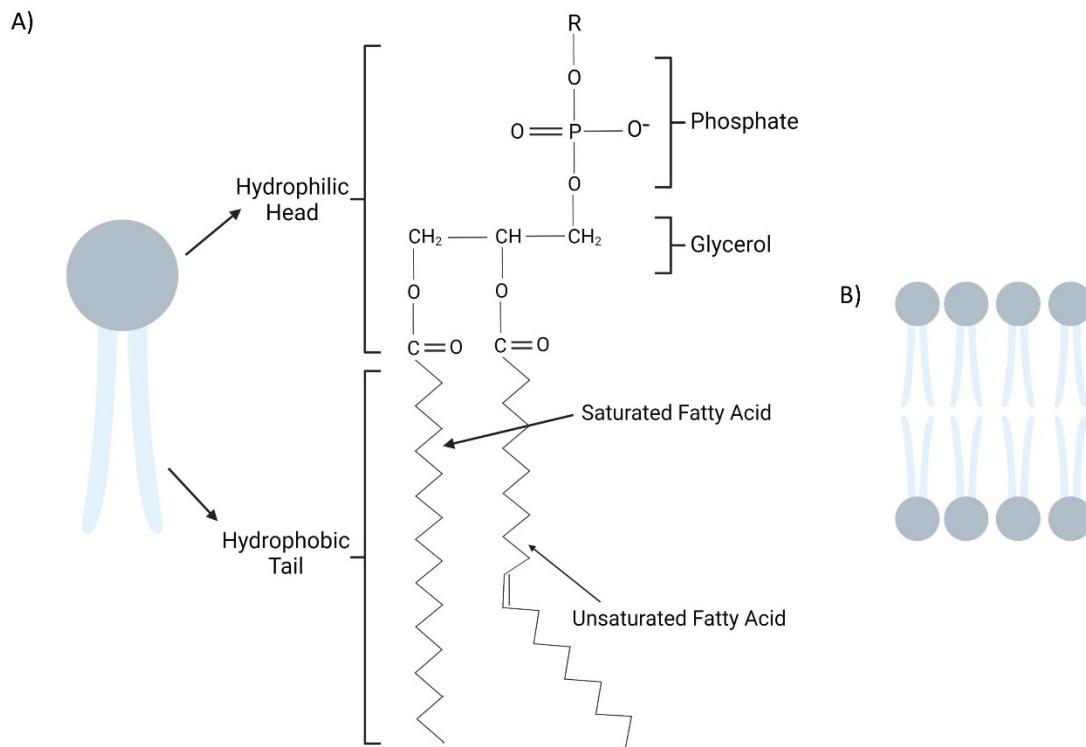


Figure 1.1. The structure and arrangement of phospholipids in the plasma membrane. (A) The chemical structure of a phospholipid showing the hydrophilic head consisting of a negatively charged phosphate head and glycerol. It has a hydrophobic tail consisting of two fatty acid acyl chains. One is saturated and the other is unsaturated due to the presence of a double bond in the chain which causes a kink in the chain, providing the phospholipid with greater flexibility and fluidity. (B) The arrangement of phospholipids in the plasma membrane where they have their hydrophilic head pointing out and their fatty acid acyl chain point in, forming a phospholipid bilayer (Created with BioRender.com).

1.1.2 Glycolipids

Glycolipids are another lipid found in large numbers in the membrane, consisting of at least one saccharide residue which is linked via a glycosidic bond to a lipid moiety (Malhotra, 2012). The lipid moiety is located in the plasma membrane, while the sugars residues that form the head of the glycolipid are situated near the membrane surface (Morrow et al., 1995). It is the sugar residues in the head which can, depending on how many there are in a particular species of glycolipid, extend prominently out the membrane which allow glycolipids to play a crucial role in cell-cell interactions. Some cell-cell interactions are as a result of carbohydrate-

carbohydrate interactions between glycolipids from different cells (Varki et al., 1993). The migration of tumour cells can be aided by carbohydrate-carbohydrate interactions. Specifically, the interaction between the ganglioside GM3 (Figure 1.2) and LacCer results in cell adhesion between melanoma cells and endothelial cells (Malhotra, 2012). Adhesion and tumour cells migration occurs when the GM3-LacCer interaction takes place and activates endothelial cells leading to the expression selectin receptors. Another example of a carbohydrate-carbohydrate interaction is the interaction between the lipid head groups of galactosyl ceramide and cerebroside sulfate in the myelin sheath, resulting in an improvement in the stability of the myelin sheath (Schnaar et al., 2004). These carbohydrate-carbohydrate interactions are quick and weak in comparison to protein-protein interactions (Malhotra, 2012).

One subset of glycolipids are the glycosphingolipid family with a smaller subset of that family called gangliosides. Gangliosides are comprised of a ceramide lipid tail, which is made up of sphingosine and fatty acid, and a glycan head containing various numbers of sugars and at least one sialic acid (Schnaar et al., 2014) (Figure 1.2).

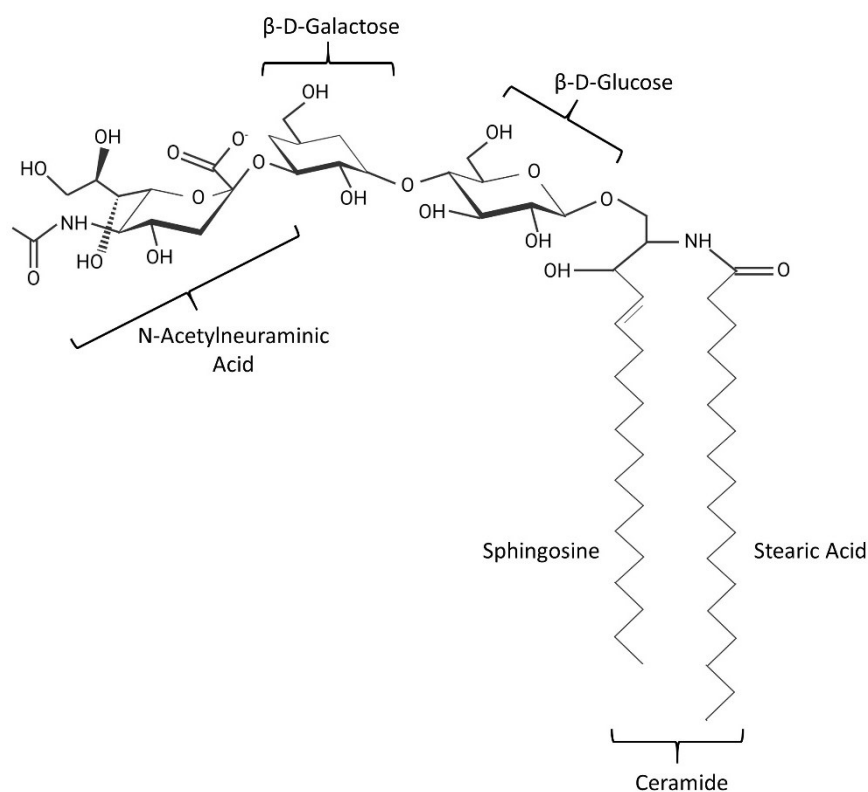


Figure 1.2. The chemical structure of the ganglioside GM3. The structure of GM3, a ganglioside common to regions outside the brain. It has a ceramide tail that is situated in the membrane consisting of a stearic acid and a shorter sphingosine. The glycan head is attached to the ceramide by a glycosidic bond. The glycan head contains sugar residues. In GM3, it comprises β -D-glucose residue and a β -D-galactose residue which is attached to sialic acid, providing it with a negative charge. Different gangliosides have different numbers of sugar and sialic acids residues. The head extends out of the membrane into the extracellular region (Created with BioRender.com).

The different numbers of sugars and their arrangements in the glycan head, as well as differing lengths and saturation of sphingosine and fatty acid in the ceramide tail mean that there are >200 forms of ganglioside (Yu et al., 2011). This high diversity of ganglioside structures leads to a different ganglioside composition in different tissues, even though most gangliosides are found in the brain. There are 10-30 fold more gangliosides found in the brain than other tissue (Sipione et al., 2020). Of the gangliosides found in the brain 90% are either GM1, GD1a, GD1b or GT1b (Tettamanti et al., 1973), while GM3 is the dominant ganglioside found in other regions of the body (Prokazova et al., 2009). Gangliosides with saturated tails are more common in places like the liver and kidney but they are less common in the brain (Hama, 2010).

Gangliosides are only found in the outer leaflet of the lipid bilayer and the ceramide tail is situated in the membrane with the glycan head poking out the membrane into the extracellular region allowing for *cis* and *trans* interactions, most commonly via H-bonds between the sugars in the glycan head and the interacting molecule (Imberty & Varrot, 2008). Gangliosides can interact with membrane proteins and receptors from other cells via glycan binding sites (Lopez & Schnaar, 2009). Foreign bodies, such as microbes and toxins, are also capable of interacting with gangliosides which can lead to infection (Merritt et al. 1998). For instance, the malaria parasite, *Plasmodium falciparum*, interacts with sialic acid and/or gangliosides (Cuttillo et al. 2020) and the cholera toxin subunit B, and the similar *E. coli* heat-labile toxin subunit B, interact with GM1 (Spangler, 1992).

1.1.3 Cholesterol

Cholesterol is an asymmetric biomolecule consisting of four hydrocarbon rings, an eight carbon isooctyl chain for a tail and a hydroxyl group at its head (Figure 1.3). This hydroxyl group can form hydrogen bonds with polar amino acids. The arrangement of the four hydrocarbon rings forms two distinct faces. There is a “flat” α face and a “bumpy” β face with two methyl groups. The β face interacts with Leu, Ile and Val via van der Waals forces, while amino acids with aromatic rings can interact with either face of cholesterol (Fantini & Barrentes, 2012).

Cholesterol interacts with membrane proteins at sites that have specific amino acid sequences. The Cholesterol Recognition/interactions Amino acid Consensus sequence (CRAC) is one such sequence. The CRAC domain consists of (L/V)-X1-5-(Y)-X1-5-(K/R) (Li & Papadopoulos, 1998). The Leu/Val at the beginning of the CRAC domain interacts with the isooctyl tail of cholesterol, while the Tyr in the middle interacts with the hydrocarbon rings and the Lys/Arg at the end interacts with the hydroxyl group. The “X1-5” between the defined amino acids represents a consecutive sequence of any 1-5 amino acids. Another cholesterol binding

domain is the CARC domain, named because the sequence of amino acids is the CRAC domain reversed with the addition of Phe in the middle. The CARC domain consists of (K/R)-X1-5-(Y/F)-X1-5-(L/V) (Baier et al., 2011).

Cholesterol often associates with glycosphingolipids, such as gangliosides, in the exoplasmic leaflet of the membrane in areas known as lipid rafts (Simons & Toomre, 2000). Glycolipids appear to prefer such regions because as much as 70% of the total amount of glycolipids in any one cell will be situated in one of these cholesterol-rich rafts (Malhotra, 2012). Glycosphingolipids interact with each other in the membrane via their carbohydrate heads leaving cholesterol to interact with, and fill in the gaps of, the hydrophobic lipid tails of the glycosphingolipids (Brown & London, 2000). These raft regions contain few phospholipids because the cholesterol-rich regions are more rigid. The cholesterol and glycosphingolipid interaction leads to partitioning of the membrane into more fluid regions that are abundant in phospholipids and other, more rigid, regions abundant in cholesterol and glycosphingolipids (Shaul & Anderson, 1998).

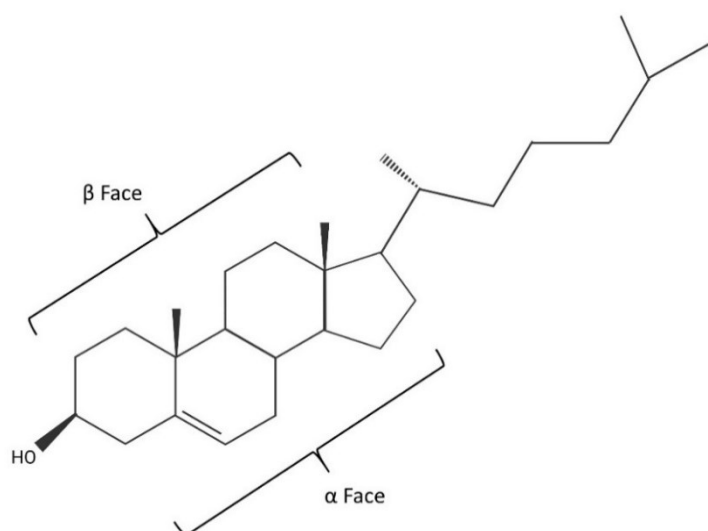


Figure 1.3. The chemical structure of cholesterol. Illustration of the basic structure of cholesterol with a hydroxyl group (-OH) at its head. The core of the molecule is made up of four hydrocarbon rings, forming two faces. There is an α face that is “flat” and a β face which is “bumpy” due to two methyl groups. Attached to the fourth hydrocarbon ring is an isooctyl chain for a tail (Created with BioRender.com).

The headgroup of gangliosides, such as GM1, adopt different conformations at the membrane surface to aid interactions with cholesterol (Lingwood, 2011). The presence of cholesterol can also help with the clustering of gangliosides. For instance, GM1 clustering is increased in GM1/sphingomyelin/cholesterol membranes when compared to a GM1/POPC membrane (Mori et al., 2012). The hydroxyl group of cholesterol forms a hydrogen bond with the

glycosidic bond that links the ceramide to sugars in the head group of GM1 causing the head of GM1 to tilt towards the membrane (Fantini et al., 2013).

1.2 Membrane proteins

Membrane proteins are ubiquitous in membranes and their roles in activities such as cell proliferation, cell signalling and the transport of ions and other molecules mean that membrane proteins are crucial for the survival of cells. They can also be the source of harm to the host when damaging mutations occur, as well as pathogens using them as an entry point into host cells.

Approximately 30% of all proteins encoded in the human genome are membrane proteins (Wallin & von Heijne, 1998), yet they are the target for approximately 60% of drugs highlighting their importance and the need for further study of them (Overington et al., 2006). In terms of the overall molecular mass of the membrane, membrane proteins and lipids each account for approximately 50% of the membrane (Frick & Schmidt, 2019).

Membrane proteins can either be integral, peripheral, or lipid-anchored in their interaction with the plasma membrane (Figure 1.4). Integral membrane proteins are fully embedded in the membrane, interacting with the hydrophobic core via bonds to the acyl chains of phospholipids and they usually traverse the whole membrane (Frick & Schmidt, 2019).

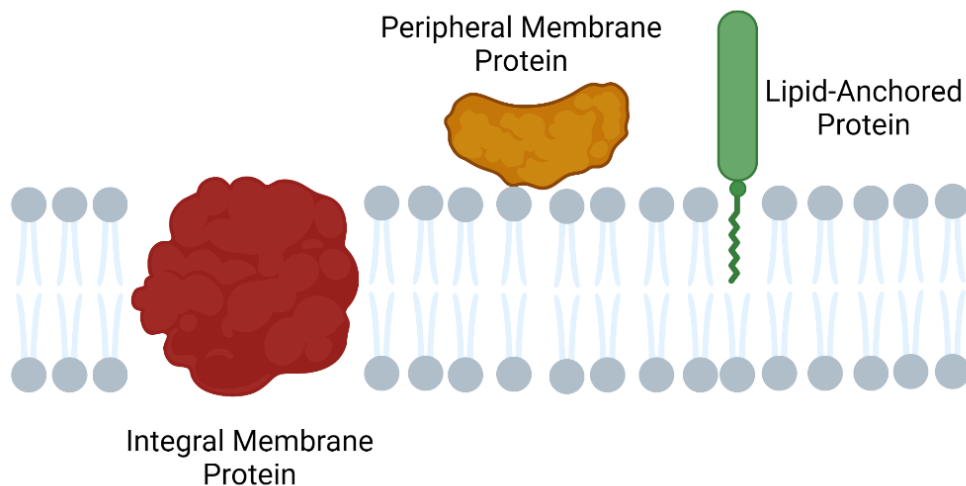


Figure 1.4. Illustration of how membrane proteins interact with the plasma membrane. Integral membrane proteins (red) are embedded in the membrane, interacting with the fatty acid acyl chains of the phospholipids. Peripheral membrane proteins (orange) do not embed in the membrane. Instead, they interact with phospholipid head groups on the surface of the membrane. Lipid-anchored membrane proteins (green) interact with the membrane via a lipid bound to the protein which embeds itself into the membrane (Created with BioRender.com).

Peripheral membrane proteins are not embedded in the membrane. They directly interact with phospholipid head groups via electrostatic and other non-covalent interactions. Unlike integral membrane proteins which cannot be easily displaced from the membrane, peripheral membrane proteins can be displaced from the membrane due to their susceptibility to changes in the environment, such as pH (Pieper et al., 2009). An example of a peripheral membrane protein is dihydroorotate dehydrogenase, an enzyme crucial for DNA and RNA construction because it is part of the biosynthesis process of pyrimidine (Khutornenko et al., 2010). Lipid-anchored membrane proteins have no direct interaction with the membrane, so they are not embedded in the membrane either. They interact with the membrane via a lipid that is covalently bound to the protein (Hentschel et al., 2016). One such family of lipid-anchored membrane proteins are alkaline phosphatases which are responsible for regulating nucleotides by hydrolysing them to nucleosides (Montenegro et al., 2014).

Integral membrane proteins fall into the four main categories, transporters and channels, receptors, enzymes and anchors (Figure 1.5). Transporters and channels are pore-forming membrane proteins that span the membrane. The ability to form a pore which may open, and close, allows for the transport of small molecules across the membrane, such as ions, amino acids and water (Drew & Boudker, 2016).

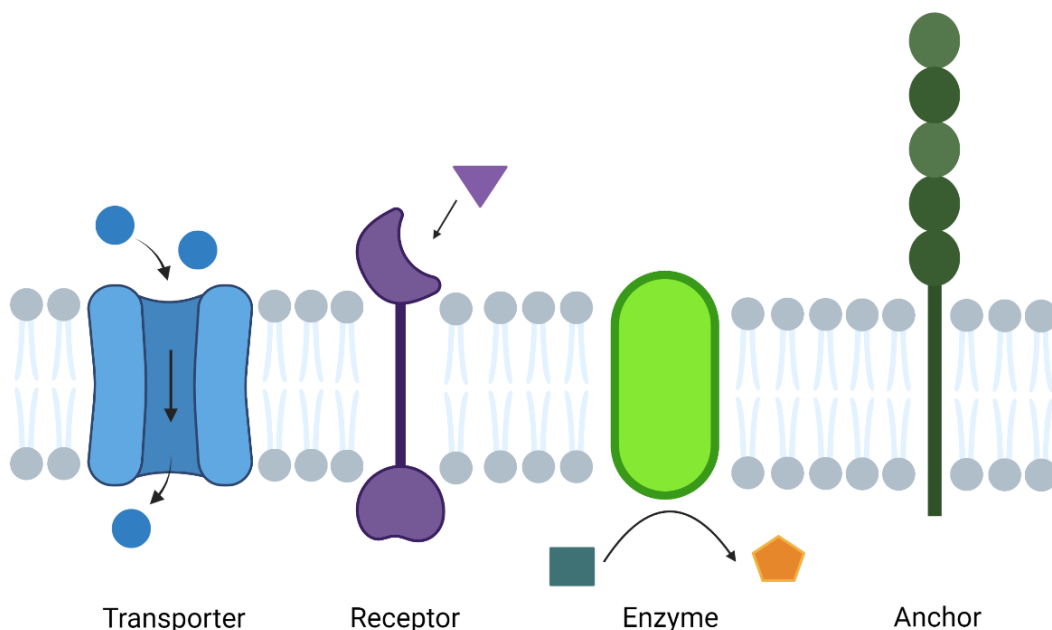


Figure 1.5. Illustration of the main categories of membrane protein. Transporters and channels (blue) are pore-forming membrane proteins capable of transporting small across the membrane via the pore. Receptors (purple) bind a ligand in a ligand binding site in their extracellular domains which leads to the start of intracellular processes. Enzymes (light green) catalyse the transformation of a molecule from one form to another. Anchors (dark green) provide stability by interacting with other proteins from other cells or by interacting with intracellular molecules, such as actin to provide cytoskeletal stability (Created with BioRender.com).

Receptors, such as G protein-coupled receptors (GPCRs), have a ligand binding site in their extracellular domain which binds a ligand which results in the activation of an intracellular process (Rosenbaum et al., 2009). Enzymes in the membrane catalyse the chemical reaction of a molecule from one form to another, like receptor tyrosine kinases (Lemmon & Schlessinger, 2010). Anchors act by ensuring that cells stay in a particular place by interacting with other membrane proteins from other cells, such as in the case of occludin (OCLN) which is in the tight junction binding to OCLN from adjacent cells (Cummins, 2012).

1.3 Tetraspanins

Tetraspanins are membrane proteins whose distinguishing feature is the presence of four transmembrane (TM) regions and a Cys-Cys-Gly motif in the LEL. There are 33 known tetraspanins in humans that are relatively small proteins ranging in size with a molecular mass of 20-30 kDa (Hemler, 2008). The 33 tetraspanins are expressed throughout the human, playing a role in a variety of biological phenomena, such as cell migration, signal transduction and intracellular trafficking (Florin & Lang, 2018). Tetraspanins are also used by pathogens to gain entry to human cells, such as hepatitis C binding to CD81 to infect hepatocytes (Fénéant et al., 2014).

Tetraspanins, except for Tspan10, have short intracellular N-terminal tails followed by the first transmembrane domain (TM1) which is connected to the second transmembrane domain (TM2) by a small extracellular loop (SEL). A small intracellular loop (SIL) joins TM2 to the third transmembrane domain (TM3) which connects to the fourth transmembrane domain (TM4) by the large extracellular loop (LEL) (Figure 1.6).

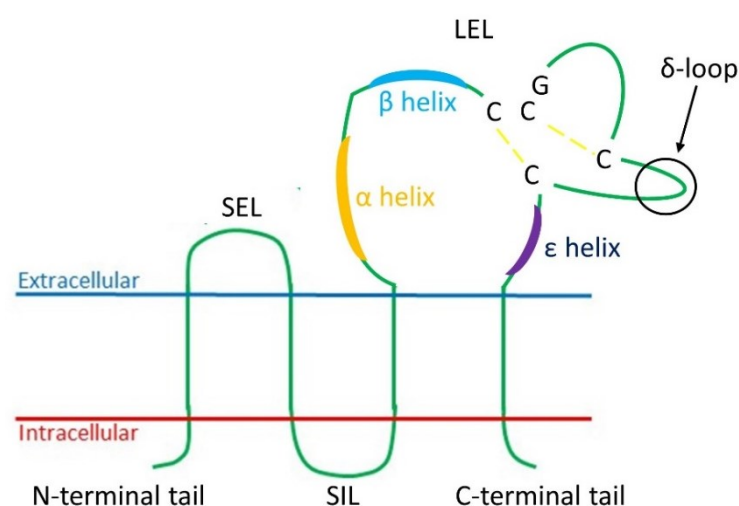


Figure 1.6. Basic structure of a tetraspanin. Tetraspanins have four transmembrane domains linked by a SEL and LEL in the extracellular region and a SIL in the intracellular region. The LEL contains the characteristic CCG-motif whose Cys residues form disulfide bonds with other Cys residues in the LEL. The LEL contains the α helix, the β helix, the ϵ helix and the δ -loop between the β helix and the ϵ helix.

All human tetraspanins have α , β and ϵ helices in the LEL with differing structures between the β and ϵ helices. TM4 is followed, in most cases, by a short intracellular C-terminal tail (Hemler, 2008). The defining feature of a tetraspanin is the conserved Cys-Cys-Gly motif (CCG-Motif) situated in the LEL which forms disulfide bonds with other Cys residues in the LEL to stabilise the structure of the LEL (Kitadokoro et al., 2001).

Disulfide bonds are bonds between two Cys residues because Cys residues have a sulfhydryl group which can form a S-S bond when oxidised that provide stability and structural integrity in proteins (Weidemann et al., 2020) (Figure 1.7). Disulfide bonds can be formed intramolecularly, such as in the case of the bonds formed in tetraspanins LELs, or they can be formed intermolecularly between proteins, such as the disulfide bonds formed between the tetraspanins, ROM1 and PRPH2 (Goldberg & Molday, 1996).

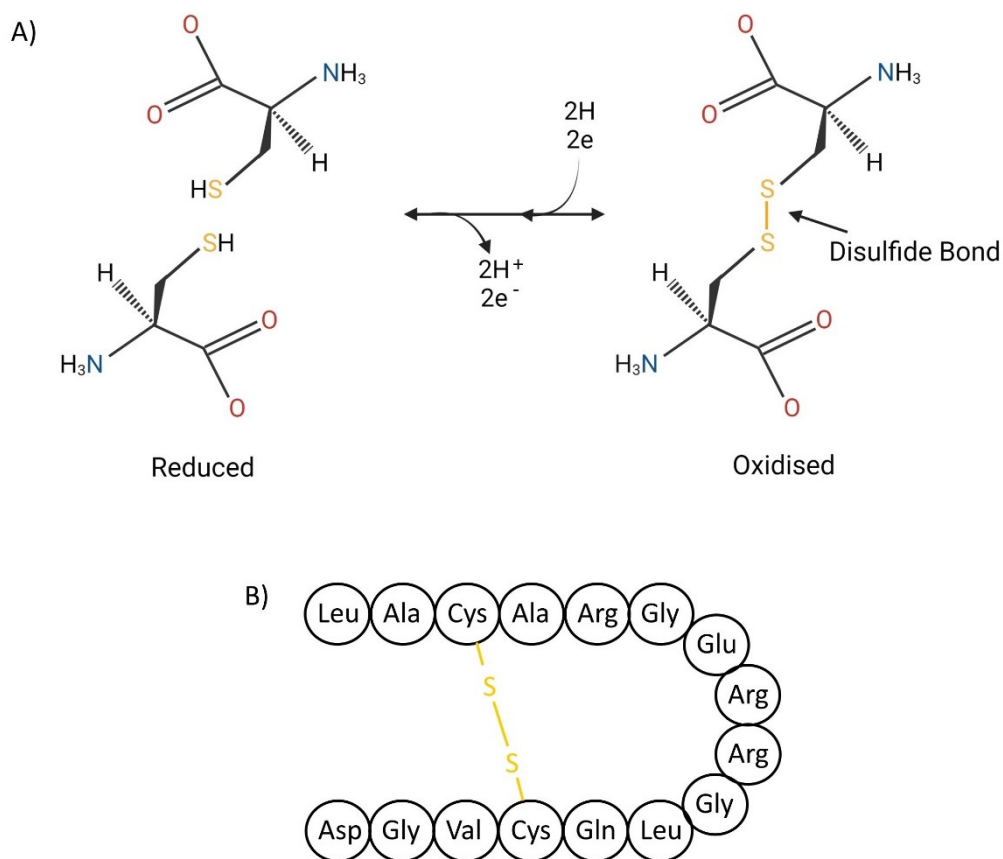


Figure 1.7. Formation of disulfide bonds in proteins. (A) The chemical reaction of two reduced cysteine residues undergoing oxidation to form a disulfide bond between the two cysteine residues (B) Example of a disulfide bond between two Cys residues in a peptide chain (Create with BioRender.com).

Disulfide bonds can be cleaved, changing the conformation of the protein which influences binding to the protein and its function (Weidemann et al., 2020). One such instance of this occurring is in the blood protein plasminogen which, with the aid of urokinase, is converted to

plasmin which contains a Cys512-Cys536 disulfide bond. The bond is cleaved, changing the conformation of the protein, and effectively opening up the peptide backbone to expose the Lys486, Arg474 and Arg530 residues for proteolysis (Butera et al., 2014). Other post-translational modifications that can be applied to tetraspanins are ubiquitination in the intracellular region which has been linked to Tspan6 downregulation (Wang et al., 2012), N-linked glycosylation (Marjon et al., 2016) and palmitoylation (Charrin et al., 2002).

1.3.1 Multiple sequence alignment of tetraspanin amino acid sequences

Multiple sequence alignments are a bioinformatics technique done to compare the amino acid, DNA or RNA sequences of specific proteins or genomes to look at molecular phylogeny or, in the case of amino acid alignment, it can be used to highlight conserved regions that may be biologically significant (Chowdhury & Garai, 2017). Amino acid sequence alignments are more difficult than DNA sequence alignments because of the increased complexity and possible alignments that arise from there being 20 amino acids compared to four nucleotides (Daugelaite et al., 2013).

The two types of alignment available are pairwise sequence alignment (PSA) and multiple sequence alignment (MSA). The difference between the two is that PSA only aligns two sequences, whereas MSA aligns more than two sequences. MSA is ideal for looking at protein families, such as the tetraspanin family, because there are 33 known human tetraspanins and MSAs highlight conserved regions which may have some structural and/or functional significance (Do & Katoh, 2008).

One method employed by MSA programs is progressive alignment where the best pairwise alignment of the submitted sequences is found and sequences are then added to the alignment in a progressive manner based on the initial pairwise alignment and subsequent alignments (Chowdhury & Garai, 2017). A consensus sequence is generated as the progressive alignment is built which aids the addition of the next sequence to the alignment. Accuracy in the early stages of the progressive alignment is crucial since errors at the start of the alignment will be exacerbated later when sequences are incorrectly aligned due to errors in the consensus alignment (Daugelaite et al., 2013).

Aligned sequences are subject to a scoring method, such as BLOcked SUBstitution Matrix (BLOSUM) (Henikoff & Henikoff, 1992) which is a score given to a certain position within the sequence based on its similarity throughout that position. A commonly used BLOSUM scoring matrix is BLOSUM62 which is used to score sequences that have a sequence similarity <62% (Figure 1.8). Each amino acid substitution is given a score based on whether the likelihood of that substitution occurring more likely than by chance.

A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	0	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-2	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Figure 1.8. BLOSUM62 score matrix. Amino acid substitution scores in the BLOSUM62 score matrix ranging from a score of 11 for Trp-Trp (W-W) to a score of -4 for several hydrophobic-hydrophilic and small-large changes. Positive scores are coloured in yellow.

Amino acids with similar structures and properties, such as Val to Ile, are given a positive number, while amino acids that do not have similar structure and properties, such as Asp to Arg, are given a negative number. Larger positive scores are given to amino acids that are less commonly found in proteins. For instance, amino acids like Ala, Leu and Gly are common, whereas the likes of Trp, Cys and Met are not as common (Pearson, 2013). To highlight this Ala-Ala in the BLOSUM62 score matrix is given a score of 4, whereas Trp-Trp is given a score of 11 (Figure 1.8).

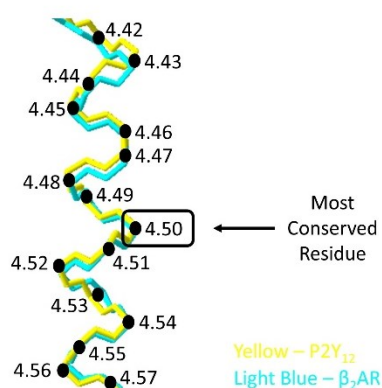
Highlighting conserved regions of tetraspanins can guide further research as conserved regions often are significant for function. Highlighting the conserved regions of tetraspanins can also lead to the production of a generic amino acid numbering system, much like the systems currently in use for G protein-coupled receptors (GPCRs).

1.3.2 Generic amino acid numbering for protein families

G protein-coupled receptors (GPCRs) are a large family of membrane proteins with seven TM domains involved in a range of interactions and processes (Basith et al., 2018). Having been widely studied a generic residue numbering system was created for them giving each position

in the TM domains a generic number that will be the same in each GPCR allowing for easier comparisons between the residues in different GPCRs.

The most used numbering system is the Ballesteros-Weinstein system (Ballesteros & Weinstein, 1995). Ballesteros-Weinstein numbers are based around a conserved residue and each position in the TMs is given a number composed of two numbers separated by a dot. The most conserved residue in a TM is found and if that residue were in TM4 then the most conserved residue would be given the number, 4.50. The first number, before the dot, indicates the TM number in which that residue is located and the second number, after the dot, indicates where the residue is in relation to the most conserved residue. The most conserved residue in the Ballesteros-Weinstein numbering system is assigned the number '50', therefore, 1.50 refers to the most conserved residue in TM1. A residue located four residues before the most conserved residue would be assigned the number 4.46, while a residue located four residues after the most conserved residue would be assigned the number 4.54 (Figure 1.9).



GPCR	4.42	4.43	4.44	4.45	4.46	4.47	4.48	4.49	4.50	4.51	4.52	4.53	4.54	4.55	4.56	4.57
P2Y ₁₂	<u>A141</u>	K142	I143	L144	S145	V146	V147	I148	<u>W149</u>	A150	F151	M152	F153	<u>L154</u>	L155	<u>S156</u>
β ₂ AR	<u>A150</u>	R151	V152	I153	I154	L155	M156	V157	<u>W158</u>	I159	V160	S161	G162	<u>L163</u>	T164	<u>S165</u>

Figure 1.9. Generic numbering of amino acids in GPCRs. Using the Ballesteros-Weinstein numbering system amino acids position in TM4 of P2Y12 and β2AR can be accurately described in relation to the most conserved residue. The most conserved residue is assigned the number 4.50 and the amino acids either side of it numbered according to their position in relation to 4.50. Amino acids before 4.50 are given a number that decreases, while amino acids after 4.50 are given a number that increases. Conserved residues between the two proteins are displayed (bold and underlined), demonstrating that despite little sequence conservation, their structure is well conserved allowing for accurate residue numbering based on a combination of sequence and structure.

Other numbering systems for GPCRs have been devised which use different numbers and methods to highlight the central residue which acts as the reference point. One such system uses '16' rather than '50' as the indicator of the most conserved residue and it does not include a dot between the TM number and the residue position so that it is easier for computers to

read (Oliveira et al., 1993). Two other systems do not determine the reference residue by conservation. Instead, the reference residue is found based on depth in the membrane, so the reference point will be the same distance inside the membrane (Baldwin et al., 1997; Schwartz et al., 1994). Both methods do not use an Arabic numeral for the TM number because they use Roman numerals from I-VII. The Oliveira, Baldwin and Schwartz numbering systems have been used less over time because as more GPCR structures were solved it became clear that GPCR TMs were different lengths and their inclinations meant that their angles in the membrane were different (Isberg et al., 2015).

In this thesis MSA techniques will be utilised to devise an amino acid numbering system for human tetraspanins that will allow tetraspanin researchers to determine the position of a given amino acid efficiently.

1.4 The tetraspanin, CD81

A lot of what is known about CD81 has come through research into its role in hepatitis C infection (HCV). HCV, like many other pathogens, uses tetraspanin-enriched microdomains to bind to and enter the cell (Florin & Lang, 2018). HCV is a positive-sense single stranded RNA virus belonging to the flavivirus genus which is the cause of approximately 58 million HCV infections each year (Yazdani et al., 2022).

Ten viral proteins are coded for in the HCV genome which includes E1 and E2, two glycoproteins expressed on the viral envelope which interact with host cells playing a crucial role in viral attachment and entry (Timpe et al., 2007). HCV E2 can bind to CD81 which results in the relocation of the CD81-HCV to the tight junction where CD81 binds to the tight junction protein claudin 1 (CLDN1) (Brazzoli et al., 2008). Both CD81 and CLDN1, along with scavenger receptor class B type 1 (SR-B1) and occludin (OCLN), another tight junction protein, are essential for HCV entry (Bartosch et al., 2003; Timpe et al., 2007; Benedicto et al., 2009) (Figure 1.10).

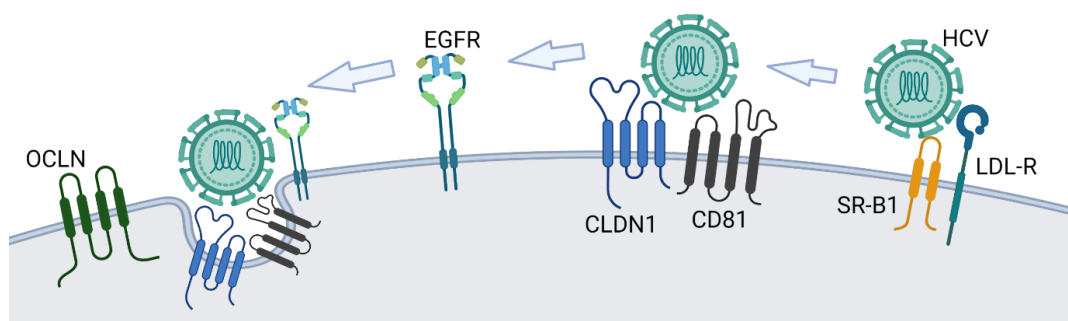


Figure 1.10. Hepatitis C viral entry showing the importance of CD81. HCV glycoproteins E1 and E2 bind to SR-B1 and LDL-R before interacting with both CD81 and CLDN1. EGFR is activated and the HCV virus is moved along with CD81 and CLDN1 to the tight junction where OCLN is situated (Created with BioRender.com).

1.5 Structures of tetraspanins

To date, four full length structures of tetraspanins have been solved. The first high resolution structure solved was CD81, also known as target of the antiproliferative antibody 1 (TAPA1) and Tspan28, at a 2.96 Å resolution by X-ray crystallography (Figure 1.11A). This solved structure is what has become known as the ‘closed’ conformation (Zimmerman et al., 2016). Somewhat unexpectedly, the TM region of CD81 in the closed conformation is in a cone-like structure with TM1-TM2 and TM3-TM4 situated close to each other. The cone-like structure results in a large hydrophobic cavity that was filled in the solved structure by a cholesterol molecule which was bound to CD81. The LEL bends at the intersection of where the LEL meets TM3 and TM4 over the cavity, in a similar manner to a hinged lid.

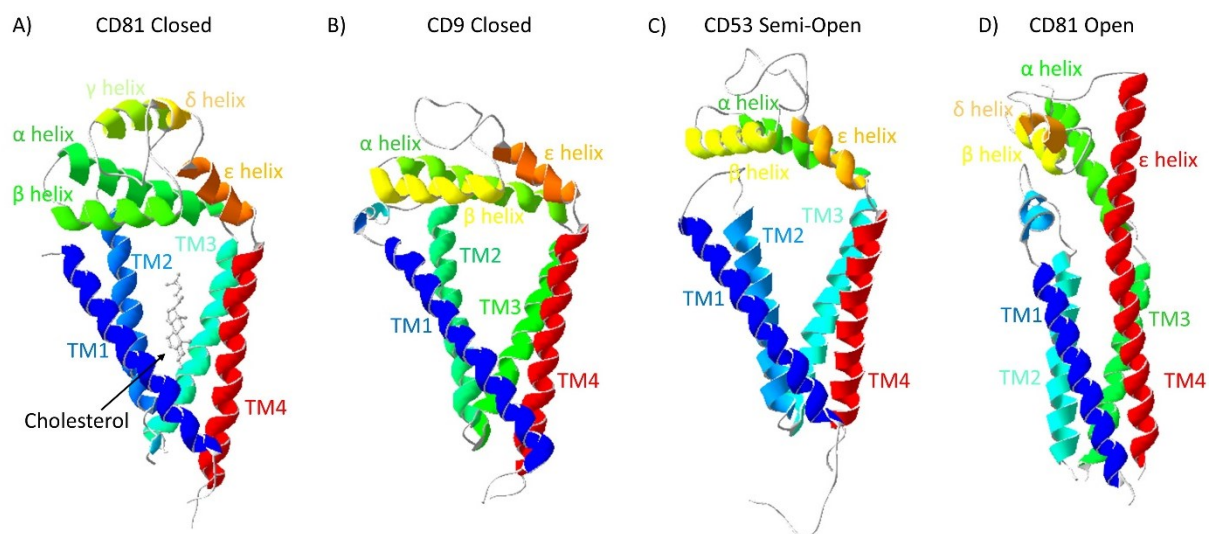


Figure 1.11. Solved full length tetraspanin structures. High resolution solved structures of full length tetraspanins. (A) CD81 closed structure (PDB: 5tcx) with a cholesterol molecule in the central cavity obtained by X-ray crystallography at a resolution of 2.96 Å (Zimmerman et al., 2016). (B) CD9 closed structure (PDB: 6k4j) obtained by X-ray crystallography at a resolution of 2.70 Å (Umeda et al., 2020) (C) CD53 semi-open structure (PDB: 6wvg) obtained by X-ray crystallography at a resolution of 2.90 Å (Yang et al., 2020) (D) CD81 open structure (PDB: 7jic) obtained by Cryo-EM at a resolution of 3.80 Å (Susa et al., 2021).

The Zimmerman et al. (2016) closed CD81 structure was surprising because a previously modelled structure of CD81 indicated that the TMs would be more closely packed together rather than forming a large cavity (Seigneuret, 2006) (Figure 1.12A). Furthermore, the LEL does not fold over the TM domain. Instead, the α and ϵ helices extend upwards away from the plasma membrane. The Seigneuret model (2006) was similar to a low resolution (6 Å) cryogenic electron microscopy (cryo-EM) structure of a mouse uroplakin (Min et al., 2006) (Figure 1.12B), which was the first experimentally solved full length tetraspanin structure.

More recently, two more tetraspanins have had their structures solved. Both CD9, also known as Tspan29, and CD53, also known as Tspan25, have a cone-like structure in the TM region like the closed CD81 structure (Figure 1.11B). The CD9 closed structure was obtained by X-ray crystallography at a resolution of 2.70 Å (Umeda et al., 2020) and its basic structure is almost identical to the closed CD81 structure. The LEL bends over the TM region, parallel with the membrane. The most notable difference is the lack of helical structures in the LEL between the β and ϵ helices that are present in the closed CD81 structure.

The structure elucidated for CD53 at a resolution of 2.90 Å by X-ray crystallography, however, adopts a slightly different conformation to that seen in the CD81 and CD9 structure (Figure 1.11C). It has the same cone-like structure in the TM region but the LEL does not bend parallel to the membrane (Yang et al., 2020). It bends closer to a 45° angle, approximately halfway between the position the LEL adopts in the CD81 and CD9 structure and the Seigneuret model. The conformation CD53 was captured in has been referred to as semi-open.

The fourth full length human tetraspanin structure is another of CD81 at a resolution of 3.80 Å obtained using cryo-EM (Susa et al., 2021) (Figure 1.11D). In this one, however, there is no large cavity in the TM region because the TMs are packed closer together and the LEL extends straight up away from the membrane exactly like the Seigneuret model (Figure 1.12A). This new CD81 conformation is known as the open conformation.

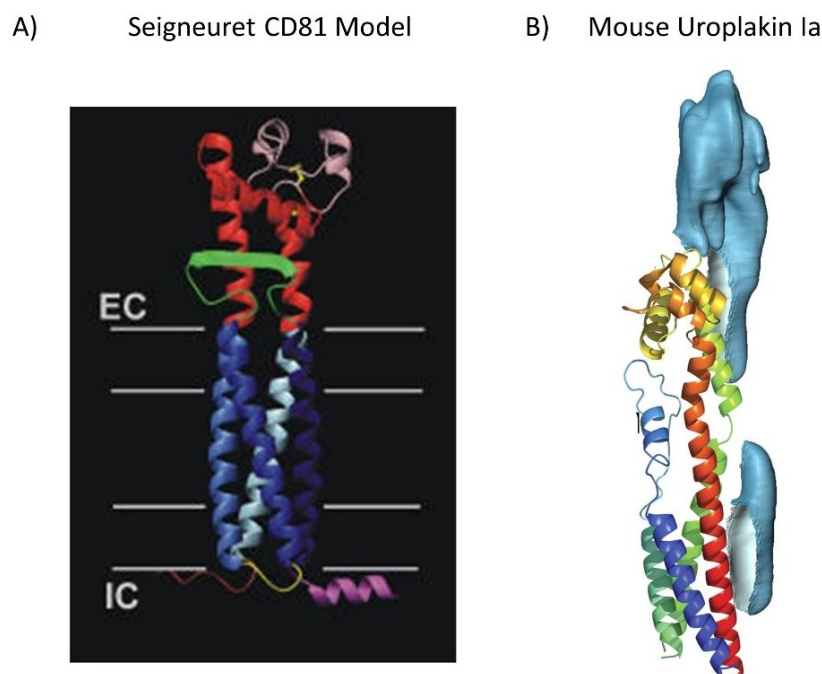


Figure 1.12. Early full length tetraspanin structures. (A) Full length modelled structure of CD81 (Seigneuret, 2006). (B) Full length structure of mouse uroplakin Ia obtained by cryo-EM at a resolution of 6 Å (Min et al., 2006).

Three of the full-length tetraspanin structures have this cone-like structure in the closed conformation (Figures 1.11A, 1.11B & 1.11C), although some doubts surrounding this structure exist when the lateral pressure of the membrane bilayer is taken into consideration (Cantor, 1999). While the net lateral pressure within the bilayer is equal to zero there are localised regions within the bilayer that are subject to lateral pressure as high as hundreds of atmospheres (Gullingsrud & Schulten, 2004). It is these localised differences in lateral pressure that can have an impact on protein conformation (Cantor, 1999), which is a phenomenon found in mechanosensitive channels, such as MscL (Perozo et al., 2002). Lateral pressure within the bilayer can also be responsible for spontaneous curvature found in the membrane (Israelachvili et al., 1977). This is of particular importance for tetraspanins because of their propensity to found in curved regions of the membrane and/or vesicles (Umeda et al., 2020). The mechanics of this cone-like structure and its transition to the open conformation are unexplored in the laboratory. To date, all work on the transition has been done computationally. One method to investigate this conformational change is with electron paramagnetic resonance (EPR).

1.5.1 Electron paramagnetic resonance to probe protein conformational change

Until recently, much of structural biology relied on X-ray crystallography to provide structural information on proteins but the major drawback of this method is the lack of dynamic information obtained. One method that can be used to look at protein dynamics and their change in conformation is EPR (Columbus & Hubbell, 2002).

EPR uses spin labels to measure the intramolecular or intermolecular distances between two spin labels. One way in which these measurements can be done is using continuous wave EPR (cwEPR) which is performed using constant microwave radiation at a specifically defined frequency (Glaenger et al., 2018). Another commonly used method is pulsed electron-electron double resonance (PELDOR), also known as double electron-electron resonance (DEER), where a pulsed, rather than a continuous, wave is applied to measure the distance between two spin labels in the same molecule, or in different molecules (Claxton et al., 2015). PELDOR can be used to measure distances between 15-80 Å, although most proteins are more amenable to a lower high of 60 Å, while cwEPR is ideal for distances <20 Å (Hubbell et al., 2000).

Site-directed spin labelling (SDSL) is used to introduce a spin label that contains a small radical at a specific site in a protein (Altenbach et al., 1990). One commonly used spin label is the methanethiosulfanate spin label (MTSSL) which forms a disulfide bond with cysteine (Sahu & Lorigan, 2018) (Figure 1.13). Cysteine residues native to a protein can be used to

form a disulfide bond with the spin label or cysteines can be mutated into a protein in a specific location that may be better in terms of accessibility or to stay within the distance range that is detectable (Glazner et al., 2018).

EPR is not alone as a technique used to research protein conformational change with nuclear magnetic resonance (NMR) and single-molecule fluorescence resonance energy transfer (smFRET) frequently used. EPR, however, is more sensitive than NMR by 3 orders of magnitude (Sahu & Lorigan, 2018). The advantage of EPR over smFRET is that the spin labels used for EPR are smaller than the large labels used for smFRET and the same spin label can be used for both sites that need labelling when doing EPR, whereas smFRET requires different labels for each site (Glazner et al., 2018).

Many proteins are not static molecules that stay in one conformational state; therefore, it is important to use methods that can probe the dynamics that underpin conformational change. It has been proposed that CD81 undergoes significant conformational change from an open to a closed state, and vice versa, depending on its interaction with cholesterol (Zimmerman et al., 2016). EPR can be used to look at this mechanism experimentally to confirm what has been proposed based on molecular modelling.

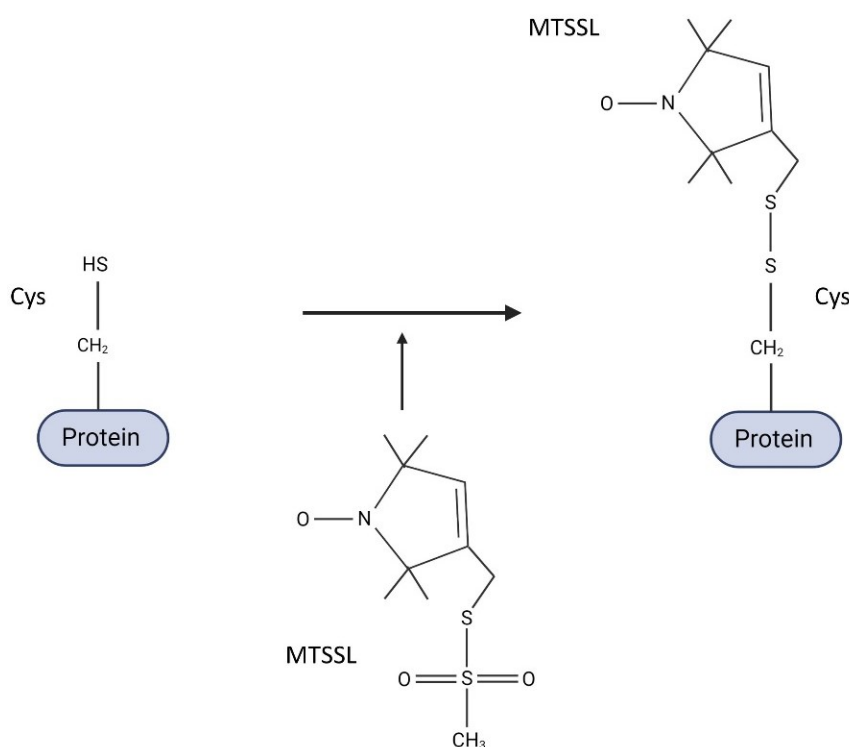


Figure 1.13. Attachment of MTSSL to a cysteine residue. A spin label, MTSSL, is attached to a free cysteine residue by disulfide bond to be able to perform EPR experiments (Created with BioRender.com).

There are unanswered questions with regards to the mechanism that dictates how CD81, and other tetraspanins, may change between the open and closed conformation but questions remain about tetraspanins structures in general. Only three tetraspanins have had their full-length structures solved, which means that there are 30 more human tetraspanins that have not had their structures solved. This is something that needs addressing so that the picture can become clearer with regards to common and unique structural elements in tetraspanins. It is possible to use the computational method of protein structure prediction to try and advance our knowledge in this field.

1.5.2 Protein structure prediction

Membrane proteins, such as tetraspanins, are important due to their ubiquity and their propensity to be targeted by drug. To be able to understand the interaction between drugs and membrane proteins and/or design new or better drugs a structure of the protein is useful. Despite their importance only approximately 3% of crystal structures in the Protein Data Bank (PDB) are membrane proteins (Xia et al., 2018). Newer methods of obtaining structures, such as Cryo-EM have opened the door for more membrane protein structures but it is an aspect of structural biology that needs addressing.

One way to overcome the problem is to use structure prediction programs to produce structures. Structure prediction methods broadly fall into one of two categories. Either a template-based method or template-free method is employed. In both cases an amino acid sequence can be submitted but in the case of template-based structure prediction similar sequences are searched for that have had their structures solved (Deng et al., 2018). This method forms the basis of homology modelling and is built upon the idea that there are only a limited number of structural folds in proteins and most of them will have already been documented in solved structures (Zhang & Skolnick, 2005).

SWISS-MODEL (Guex & Peitsch, 1997) is a popular homology modelling program that uses a template-based approach. It works by taking the input sequence and using BLAST (Altschul et al., 1997) to search for potential templates based on sequence similarity, followed using HHblits (Remmert et al., 2011) to align the input sequence of the sequence of the template. The homology model is constructed based on the constraints of the atomic co-ordinates from the alignment with any gaps in the alignment being filled in by taking suitable fragments of protein structure from the fragment library (Deng et al, 2018). Another program that has the option of using template-based structure design is transform-restrained Rosetta (trRosetta) but trRosetta can also use a template-free method.

Template-free methods, such as trRosetta (Yang et al., 2020; Du et al., 2021), often still use some form of template, making use of small structural regions in solved structures. Unlike the

template-based method, however, there is no global template used that covers the whole sequence of the protein (Deng et al., 2018). The input sequence is used to create an MSA consisting of related proteins or proteins that contain specific regions that align well with the input sequence. The MSA is then used to compile information of secondary structure, torsion angles and potential interactions between different residues in the protein (Kuhlman & Bradley, 2019).

1.6 Molecular Function of Tetraspanins

1.6.1 The role of tetraspanins in trafficking

The molecular function of tetraspanins is varied. One function that they are involved in is trafficking, and in some cases, it appears to be the main function of some tetraspanins. One such tetraspanin is CD63 which possesses a tyrosine-based internalisation motif which aids the trafficking of partner proteins (Charrin et al., 2014). Two proteins that interact with CD63 and are trafficked are synaptotagmin, which is involved in lysosome exocytosis and membrane repair, and PMEL17, which is involved in melanogenesis (Flannery et al., 2010; van Niel et al., 2011).

The tetraspanins that have eight Cys residues in their LEL (Tspan5, Tspan10, Tspan14, Tspan15, Tspan17 and Tspan33) belong to the TspanC8 family and they are known for interacting with, and trafficking ADAM10, a metalloproteinase (Prox et al., 2012). The interaction facilitates the trafficking of ADAM10 from the endoplasmic reticulum to either late endosomes or the membrane.

CD81 is also involved in B cell receptor (BCR) signalling, forming part of the B cell co-receptor complex with CD19 and CD21 (Carter & Barrington, 2004). CD19 is a transmembrane protein consisting of just 1 TM and its interaction with CD81 was first seen using co-immunoprecipitation (Bradbury et al., 1992). CD19 is crucial for both BCR dependent and independent signalling due to its ability to lower the threshold for these signalling events (Susa et al., 2020). When CD19 is stimulated it leads to a lower threshold by at least two orders of magnitude for dependent or independent B cell activation (Carter & Fearon, 1992).

The importance of CD81 for correct CD19 function has been demonstrated in CD81-null mice because these mice have a lower level of CD19 surface expression despite normal B cell development (Susa et al., 2020; Miyazaki et al., 1997). CD19 interaction is CD81 specific because the knockout of other tetraspanins has no effect on normal CD19 expression and expressing CD81 in CD81-deficient mice leads to CD19 expression level returning to normal but this is not seen when CD9 is expressed instead of CD81 (Levy, 2014).

The interaction between CD81 and CD19 is believed to be a dynamic interaction whereby the presence of the interaction is required for B cell activation but once the B cells are activated CD19 is no longer interacting with CD81 (Susa et al., 2020). This dynamic interaction can be exemplified by the role CD81 plays in the diffusion of CD19 (Figure 1.14). CD81 acts to prevent CD19 movement in the membrane, almost tethering CD19 to a specific location in the membrane (Cherukri et al., 2004). In this state the B cells are inactivated. B cells lacking CD81, however, show an increase in the speed at which CD9 moves within the membrane (Mattila et al., 2013).

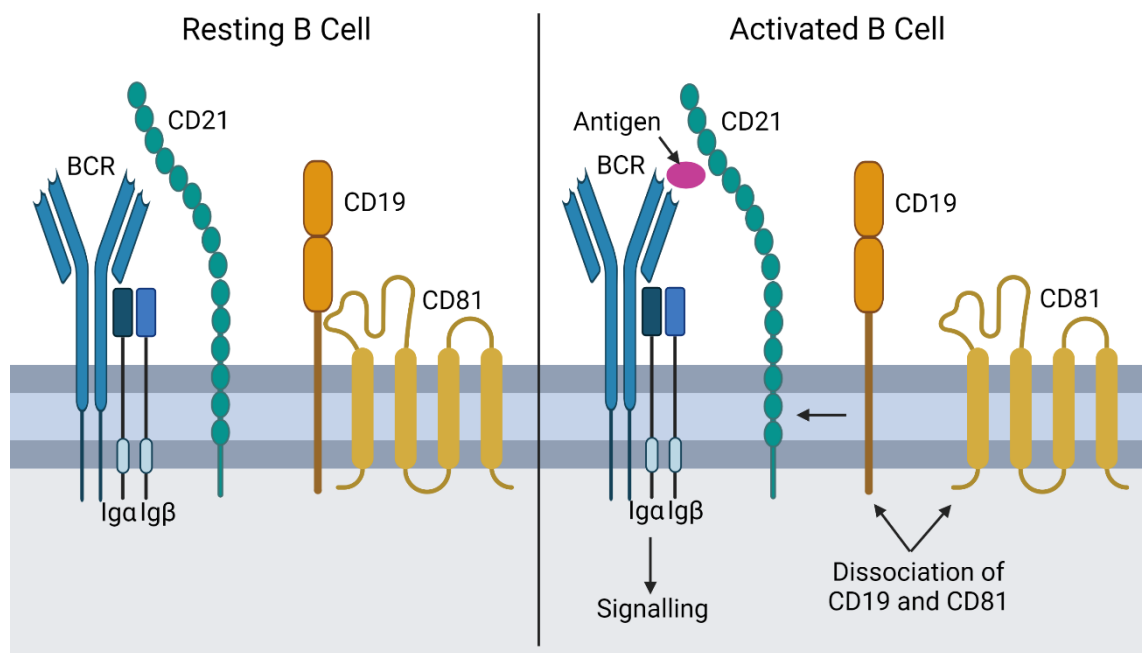


Figure 1.14. The interaction of CD81 with CD19 in resting and activated B cells. Proposed model of CD81 and CD19 interaction in B cells where CD19 diffusion in the membrane is controlled by its interaction with CD81 (Susa et al., 2020). In resting B cells there is a direct interaction between CD81 and CD19 but in an activated B cell, CD19 dissociates from CD81 and associates with the B cell receptor complex (Created with BioRender.com).

While the hypothesis put forward by Susa et al. (2020) proposes that activation of B cells is dependent on dissociation of CD19 from CD81 it is also possible that downstream signalling plays a role in the CD81-CD19 and CD19-BCR interactions (Figure 1.14). In this case the dissociation of CD81 and CD19 is because of CD81 relocating to the immune synapse of activated B cells (Mittelbrunn et al., 2002). It has also been shown that integrins can be found at the immune synapse of activated B cells interacting with CD81 (Levy et al., 1998).

1.6.2. Tetraspanins interaction with integrins

Tetraspanins have also been found to interact with integrins. CD151, for instance, interacts with the laminin-binding integrins $\alpha 3\beta 1$, $\alpha 6\beta 1$ and $\alpha 6\beta 4$ which leads to the activation of downstream signalling molecules, such as Akt, BCAR1 and Rho family GTPases (Charrin et al., 2014). In these instances, CD151 is not directly involved in strengthening adhesion but it can play a direct role (Figure 1.15).

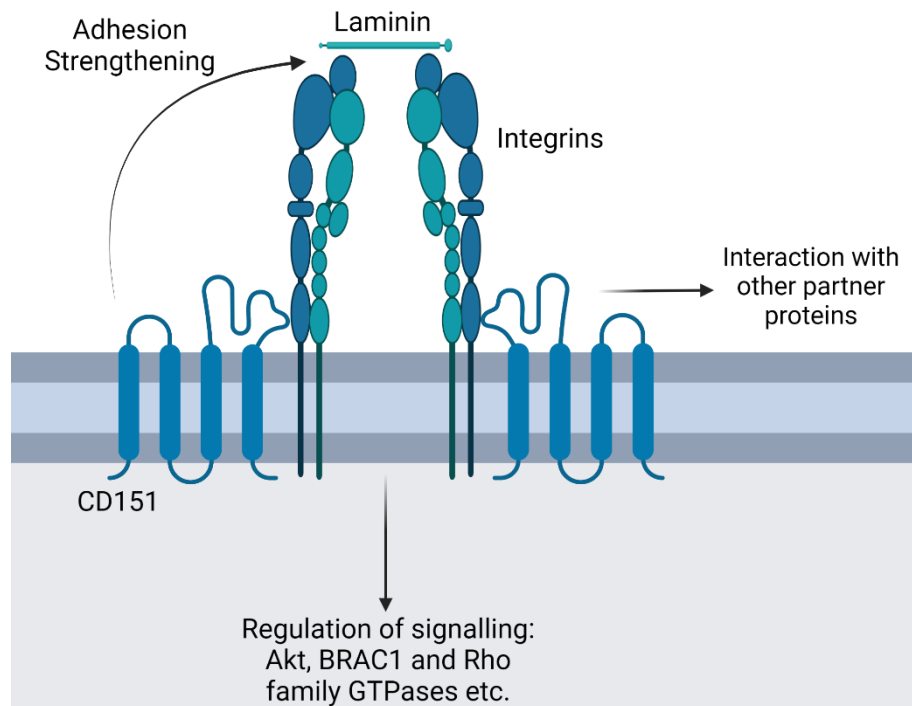


Figure 1.15. The interaction of CD151 with integrins. CD151 is directly involved in adhesion strengthening via its interaction with integrins, as well as this interaction regulating the signalling of downstream molecules, such as Akt, BRAC1 and Rho family GTPases.

CD151 can directly strengthen adhesion by interacting with integrins and clustering receptors together within a specific region of the membrane (Puklin-Faucher & Sheetz, 2009). One example of this can be found when CD151 interacts with the platelet fibrinogen receptor and $\alpha \text{IIb}\beta 3$ (Lau et al., 2004). Another example involving two other tetraspanins, CD81 and CD37, controls the adhesion of lymphoid B cells to integrin $\alpha 4\beta 1$ ligands (Charrin et al., 2014). The importance of this interaction has been shown because in CD37-null B cells, integrin $\alpha 4\beta 1$ activity is disrupted due to a difference in where integrin $\alpha 4\beta 1$ is localised within the membrane (van Spriel et al., 2012).

1.6.3 The role of CD81 and CD9 in cell-cell fusion

CD81 and CD9 share a sequence similarity of 45% and have common partner proteins, such as EWI-F/CD9P-1 and EWI-2 (Charrin et al., 2003). These similarities mean that they are involved in cell-fusion. Their interaction, crucial for the interaction between a sperm and an egg, is a well understood molecular event. Eggs that are CD9 deficient are unable to fuse to a sperm (Miyado et al., 2000) which can partly be explained by the change of shape of egg microvilli. In CD9-deficient eggs the microvilli have a different length, thickness, and density. The thickness of the microvilli, for instance, are twice the radius of that seen in normal eggs which is problematic because smaller microvilli are better for fusion (Runge et al., 2007).

The CD81-CD9 interaction also acts as a negative regulator of muscle cell fusion and in the repair of muscle cells (Charrin et al., 2013). The repair of muscle is, to some degree, dependent on cells, such as myofibres to fuse but this is diminished, leading to abnormal muscle regeneration when CD81 and CD9 are absent. The same occurs when the CD81 and CD9 partner protein, CD9P-1, is silenced (Charrin et al., 2013), highlighting the importance of tetraspanins as organisers within the membrane that ensure that partner proteins are correctly localised.

1.7 Tetraspanin-enriched microdomains

Tetraspanin-enriched microdomains (TEMs) are regions within the plasma membrane that are abundant in tetraspanins, cholesterol, gangliosides, and partner proteins like epidermal growth factor receptor (EGFR) and integrins (Zuidscherwoude et al., 2015) (Figure 1.16). Due to the number of biomolecules that conglomerate together in specific regions TEMs are crucial for membrane organisation, with strong interactions formed between tetraspanins and non-tetraspanin proteins (Florin & Lang, 2018). Weaker interactions exist between tetraspanins and other tetraspanins via the δ -loop in the LEL with homo-tetraspanin interactions preferred to hetero-tetraspanin interactions (Kovalenko et al., 2004). Regions of the membrane high in cholesterol and gangliosides are 10-200 nm in size and more rigid than other areas which will have more phospholipids (Pike, 2009; Shaul & Anderson, 1998).

The regions are stabilised by palmitoylated regions of tetraspanins (Charrin et al., 2002). As membrane organisers, palmitoylation is a post-translational modification of proteins which involves the covalent attachments of palmitic acid to a Cys residue, aiding the relocation of the palmitoylated protein to regions of the membrane that are cholesterol rich (Linder & Deschens, 2003). Palmitoylation is important for maintaining the integrity of TEMs. Non-palmitoylated CD9 does not interact with CD53 or CD81 (Charrin et al., 2002), while non-palmitoylated CD151 weakens TEMs (Berditchevski et al., 2002). With each additional Cys

residue linked to palmitoylation in CD81 mutated, its interaction with CD151 becomes weaker (Zhu et al., 2012).

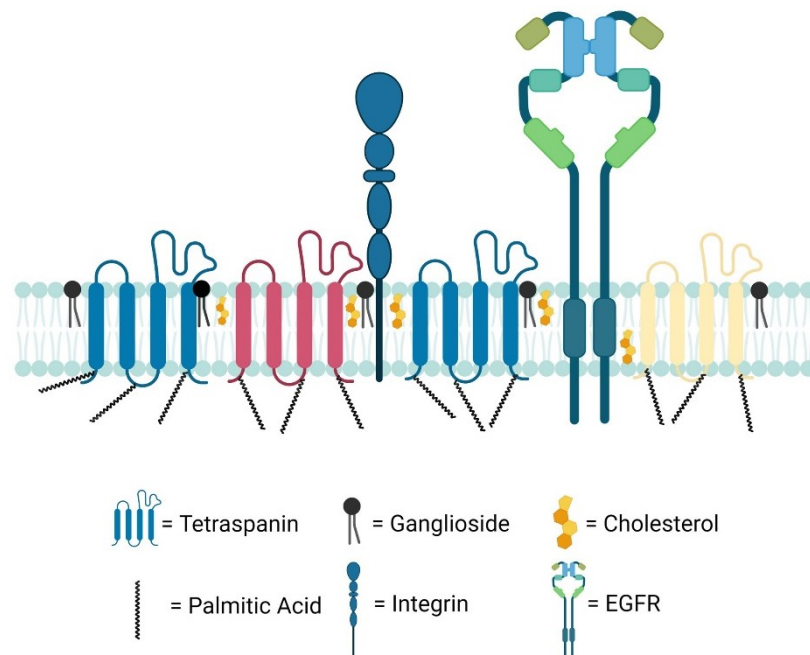


Figure 1.16. Illustration of a tetraspanin-enriched microdomain. A basic diagram of a TEM showing the accumulation of tetraspanins, along with partner proteins, such as integrins and EGFR, in a specific region of the plasma membrane. Cholesterol, which is mostly found in the outer leaflet of the plasma membrane, and gangliosides, which are only found in the outer leaflet of the plasma membrane, and palmitic acid attached to tetraspanins help these proteins to interact and stabilise TEMs (Created with BioRender.com).

1.7.1 Tetraspanin interactions with gangliosides

Much of our understanding of tetraspanin-ganglioside interactions comes from the study of the tetraspanin CD82. For instance, the ganglioside GM2 interacts with CD82 and while no direct interaction has been established between GM3 and CD82 it is known that GM2 (Figure 1.17), GM3 and CD82 form a complex (Todeschini et al., 2007; Todeschini et al., 2008).

Depletion of GM3 disrupts the CD82-EGFR-Caveolin-1-PKC- α complex and the CD82-EGFR-Caveolin-1 (Wang et al., 2007). EGFR interacts with a series of gangliosides with a preference for interacting with GM3 (Mijlan et al., 2002) which may be the reason why these complexes are disturbed given that no direct interaction between GM3 and CD82 has been demonstrated. While CD82 does not directly interact with GM3, an interaction between GM3 and the two tetraspanins, CD9 and CD81 has been shown (Kawakami et al., 2002; Toledo et al., 2004).

TEMs are negatively affected when gangliosides are depleted with a ~25% decrease in the association of CD82 with CD9 and a ~20% decrease between CD82 and $\alpha 3$ integrin subunit

in HB2/CD82 cells (Odintsova et al., 2006). The biggest impact of the depletion of gangliosides is seen on the interaction between CD82 and CD151 which results in a four-fold reduction in their interaction. There was, however, no impact on the interaction between CD9 and CD151 under the same conditions (Odintsova et al., 2006).

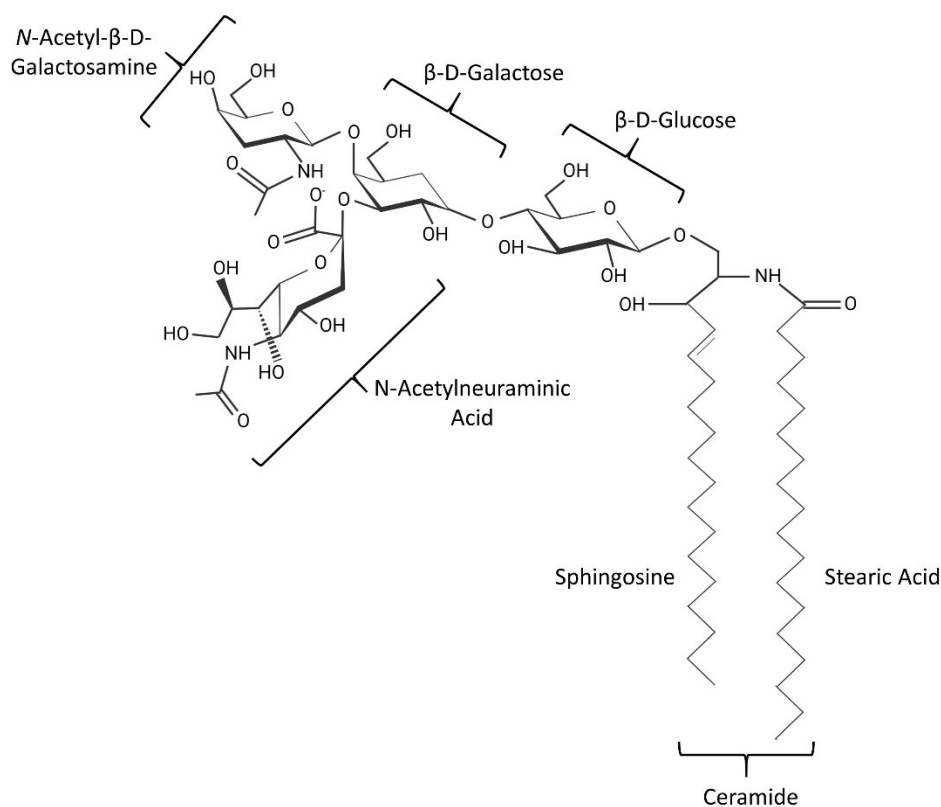


Figure 1.17. The chemical structure of the ganglioside GM2. The structure of GM2, a ganglioside common to regions outside the brain. It has a ceramide tail that is situated in the membrane consisting of a stearic acid and a shorter sphingosine. The glycan head is attached to the ceramide by a glycosidic bond. The glycan head contains sugar residues. In GM2, it comprises a β -D-glucose residue, a β -D-galactose residue attached to sialic acid, and N-acetyl- β -D-galactosamine (Created with BioRender.com).

With CD82, and to a lesser extent CD9, providing most of what is known about tetraspanin-ganglioside interaction it would be beneficial to further probe the interaction of gangliosides with CD81. The only experimental data on ganglioside-CD81 interactions is that CD81 interacts with GM3 (Toledo et al., 2004). Modelling of open CD81 indicates that it preferentially interacts with gangliosides over phosphatidylcholine species (Schmidt et al., 2016). It suggests that gangliosides mostly interact with the SEL, TM1 and TM2, while a small but not insignificant interaction occurs between the δ -loop in the LEL which may help to stabilise that region of the protein.

Plants lack gangliosides however, they do possess glycosylinositolphosphoceramides (GIPCs) which share some similarities with gangliosides, such as having a ceramide backbone, sugar residues in the extracellular region and it is negatively charged (Gronnier et al., 2016). In *Arabidopsis* rosette leaves the tetraspanin Tet8 plays a crucial role in GIPC interaction and, consequently, extracellular vesicle formation. In *Arabidopsis* rosette leaves with Tet8 knocked out there was a ~40% decrease in the formation of extracellular vesicles and, while the level of all other lipids present in the leaf were unchanged or increased in the knockout plant there was a ~22% reduction in the level of GIPCs (Liu et al., 2020). Tetraspanins are present in extracellular vesicles and gangliosides play a crucial role in membrane curvature (Bari et al., 2011).

Current knowledge of tetraspanin and ganglioside interactions is limited. It is known that the two molecules are present in tetraspanin-enriched microdomains and that some tetraspanins interact directly with some gangliosides (Odintsova et al., 2006). Despite this, little is known about the precise nature of these interactions with regards to which amino acids in tetraspanins are responsible for interacting with gangliosides. Disrupting the tetraspanin-ganglioside interface by removing gangliosides from the membrane has a negative effect on tetraspanin proteins and their ability to form complexes with partner proteins (Odintsova et al., 2006; Wang et al., 2007). Discovering the mechanism of interaction between tetraspanins and gangliosides could pave the way to therapeutics that target this interaction to treat infectious diseases and some cancers which utilise TEMs and is one focus of this thesis.

1.7.2 Tetraspanin interactions with cholesterol

It is known that CD9, CD81 and CD82 have all been found to bind cholesterol (Charrin et al., 2003; Silvie et al., 2006). The importance of tetraspanin-cholesterol interactions is further exemplified by the fact that depletion of cholesterol from the membrane using methyl- β -cyclodextrin or sequestration using saponin disrupts the interaction between tetraspanins (Charrin et al., 2003). These findings are not surprising because of the propensity for cholesterol to associate in TEMs alongside tetraspanins. Given that TEMs are occupied by tetraspanins and cholesterol it is likely that most tetraspanins have some sort of interactions with cholesterol. Four CRAC domains have been discovered in the tetraspanin CD82 and it appears to be a common feature in tetraspanins. Out of 33 known human tetraspanins, 29 of them have at least one cholesterol binding domain (Table 1.1) (Huang et al., 2020).

It had been predicted that CD81 can change conformation from an open to closed state and vice versa. Molecular modelling suggests that the cholesterol molecule that was bound inside the large cavity is responsible for the conformational change (Zimmerman et al., 2016). It has been proposed that when cholesterol is present in the cavity, CD81 is found in the closed

conformation and when cholesterol is removed it changes to the open conformation (Zimmerman et al., 2016; Palor et al., 2019). Asn18 in TM1 and Glu219 have been identified as the binding site for cholesterol (Zimmerman et al., 2016) (Figure 1.18).

Homology modelling and molecular dynamics experiments on another tetraspanin, CD151, show a similar mechanism of action. Cholesterol forms hydrogen bonds with Tyr23 in TM1 of CD151 and the backbone of Val233 in TM4 (Purushothaman & Thiruvencatam, 2019). Like modelling on CD81, CD151 adopts the closed conformation when cholesterol is docked in the cavity but then changes to the open conformation when cholesterol is removed.

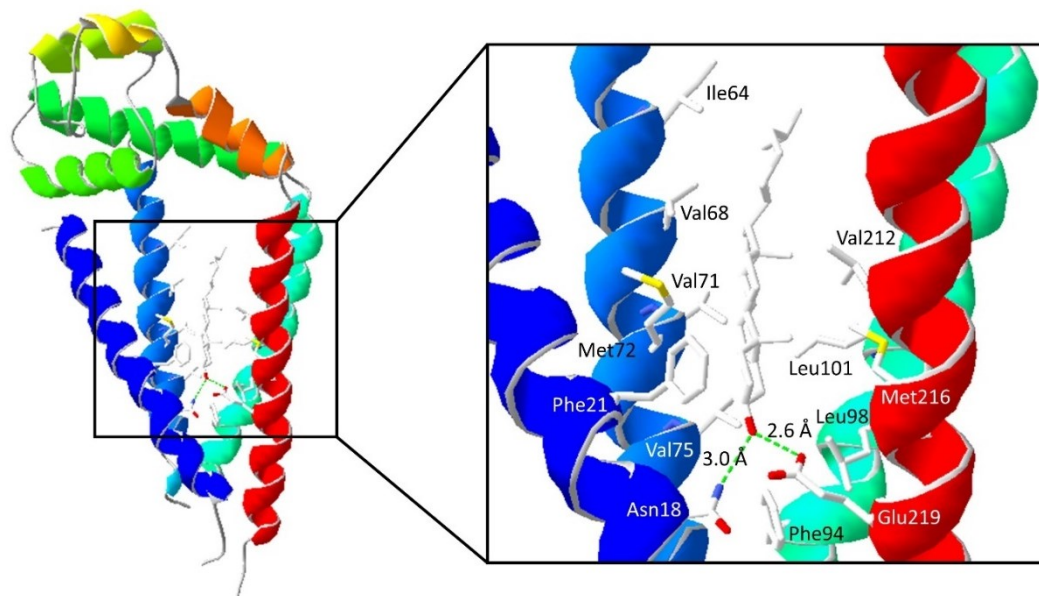


Figure 1.18. The interaction between cholesterol and CD81 in the CD81 cavity. The solved CD81 structure (Zimmerman et al., 2016) had a cholesterol molecule bound in the central cavity. All amino acids within 4 Å of the cholesterol molecule are highlighted with attention paid to Asn18 and Glu219 which form a hydrogen bond with the hydroxyl group of cholesterol. Illustration made with Swiss-Pdbviewer (Guex & Peitsch, 1997) based on a figure from Zimmerman et al. (2016).

TEMs are exploited as entry points into cells by pathogens, such as hepatitis C virus (HCV), human immunodeficiency virus (HIV) and human papillomavirus (HPV) (Florin & Lang, 2018). The role of cholesterol in HCV infection is clear. Depleting cholesterol from the plasma membrane using methyl- β -cyclodextrin has an inhibitory effect on HCV infection (Kapadia et al., 2007). Cells depleted of cholesterol cannot form TEMs which are central to how HCV enters host cells (Felmlee et al., 2013).

While our knowledge about the constituents of TEMs is advanced there is still a lack of detail about the specific interactions between molecules such as tetraspanins and gangliosides. Even the information that is known is only known about a select few tetraspanins, such as CD81, CD9, CD82 and CD151 (Table 1.1). There are still large gaps about how many of the other tetraspanins interact, if at all, with cholesterol and gangliosides.

Deciphering more about how TEMs are formed, particularly with regards to how tetraspanins, cholesterol and gangliosides interact with each other may open the door to research that disrupts these microdomains to help alleviate HCV infection and is examined in this thesis. Determining how some of them interact with these molecules could pave the way for research into similar interactions into lesser researched tetraspanins. Looking at interactions between a protein and a lipid can be performed *in silico* by making use of protein docking programs.

Table 1.1. Human tetraspanins and their interactions. A list of all known human tetraspanins and whether they are a known interactor of cholesterol, ganglioside or a microbe that causes an infection. The “Cholesterol binding domain” column refers to how many CRAC/CARC/CRAC-like or CARC-like domains are in each tetraspanin. “Yes” means that there is a known interaction, whereas “-” means that there no known interaction.

Tspan	Known Cholesterol Interaction	Cholesterol Binding Domains	Ganglioside Interaction	Microbial Infection
CD81	Yes	1	Yes	Yes
CD9	Yes	1	Yes	Yes
CD53	-	3	-	-
CD82	Yes	4	Yes	Yes
CD151	Yes	2	Yes	Yes
CD63	-	1	-	Yes
CD37	-	2	-	Yes
Tspan1	-	2	-	-
Tspan2	-	3	-	-
Tspan3	-	3	-	-
Tspan4	Yes	3	-	-
Tspan5	-	3	-	-
Tspan6	-	4	-	-
Tspan7	Yes	2	-	Yes
Tspan8	-	3	-	Yes
Tspan9	-	3	-	Yes
Tspan10	-	0	-	-
Tspan11	-	2	-	-
Tspan12	-	1	-	-
Tspan13	-	1	-	-
Tspan14	-	3	-	Yes
Tspan15	-	3	-	-
Tspan16	-	1	-	-
Tspan17	-	3	-	-
Tspan18	-	3	-	-
Tspan19	-	3	-	-
Tspan31	-	0	-	-
Tspan32	-	2	-	-
Tspan33	-	2	-	-
UP1a	-	1	-	Yes
UP1b	-	0	-	-
ROM1	-	0	-	-
PRPH2	-	4	-	-

1.7.3 Protein-ligand docking

Molecular docking is a technique used in computational biology to investigate the interaction between two molecules, such as protein-protein interaction or protein-ligand interactions (Pietro-Martinez et al., 2018). Docking is based upon the molecular mechanics that take into consideration aspects of molecular interactions, such as torsion angles and charges (Lopes et al., 2015).

At one time it was accepted that a ligand fits a protein perfectly and vice versa with no conformational change required by either molecule and it is on this basis that rigid docking programs operate. The premise of rigid docking is that the two molecules fit together like a “lock-and-key”, therefore both the protein and the ligand are rigid, unable to change conformation. The affinity with which they bind is seen as a measure of how well they fit together (Mezei, 2003).

Small differences, however, between the structures of proteins that are ligand free and those that are interacting with a docked ligand indicate that neither the protein, nor the ligand, can be considered rigid (Totrov & Abagyan, 1994). Rather than using rigid docking methods it would be better to use “implied-fit” methods where the protein and the ligand are both considered to be flexible molecules (Hammes, 2002). Many proteins undergo large conformational changes, like the proposed conformational change between open and closed in CD81 (Zimmerman et al., 2016). Another aspect to consider is that amino acid side chains within a protein are often flexible, altering the interaction site for the ligand to dock (Pagadala et al., 2017).

A program that uses this flexible docking method is Galaxy7TM (Lee & Seok, 2016) which is predominantly used for GPCR-ligand docking but non-GPCR proteins can also be submitted. In Galaxy7TM the receptor, both backbone and sidechains, and the ligand are both treated as fully flexible molecules (Figure 1.19).

Full flexibility helps to create 30 structures based on the receptor that is submitted to the server. The 30 structures are made after the initial structure is perturbed 200 times. The ligand can then be docked to the range of structures that have been created using GalaxyDock (Shin et al., 2013). To dock the ligand, it is possible to submit residues thought to interact with it. In this instance any residue $<4 \text{ \AA}$ away from the centre of the residues thought to interact is considered as a potential binding site. Refinement of 120 receptor-ligand complexes is completed by GalaxyRefine (Heo et al., 2013). Sidechains are repacked, as well as the complete receptor structure and the ligand undergoing relaxation to allow for full flexibility of both molecules (Lee & Seok, 2016).



Figure 1.19. Glucose molecule docked to closed CD81 in Galaxy7TM. An example of the output from a run Galaxy7TM (Lee & Seok, 2016) run showing the molecule (glucose) docked to the receptor (CD81). Interacting amino acids are shown. Overall structure flexibility is shown with the original structure uploaded (grey) and the overlaid refined structure (green).

1.8 Expression of membrane proteins

Studying protein-lipid interactions is not confined to computational experiments. Membrane proteins, such as tetraspanins, can be overexpressed in different expression systems, followed by solubilisation and purification. After which proteins and lipids can be bound together, although some expression and solubilisation methods are more amenable to studying native protein-lipid interactions.

1.8.1 Expression of recombinant tetraspanins in *Pichia pastoris*

To study the biochemical and biophysical aspects of a membrane protein it is advantageous to overexpress the membrane protein using an expression system. The protein can then be purified and analysed. One such expression system is the methylotrophic yeast, *Pichia pastoris* (Figure 1.20), which has previously been used to prepare purified samples of CD81 (Ayub et al., 2020). *P. pastoris* was also used as the expression system when the CD53 structure was solved, demonstrating that the structural integrity of tetraspanins can be maintained in this expression system (Yang et al., 2020).

Using other common expression systems, such as the bacteria *Escherichia coli*, can sometimes be problematic because some membrane proteins can be toxic to the cells

(Gubellini et al., 2011). Toxicity of tetraspanins when attempting overexpression in *E. coli* has been encountered. Expressing shortened versions of the tetraspanin CD151, also known as Tspan24, that contain LEL-TM4 and TM3-LEL-TM4 is possible but expression of full-length CD151 proved to be toxic to *E. coli* cells (Purushothaman & Thiruvengatam, 2019).

P. pastoris, therefore, can be an attractive alternative to *E. coli* when expressing tetraspanins. CD81 has been expressed in using a *P. pastoris* expression vector, pPICZB, which incorporates a 6x His tag at the 3' end (Jamshad et al., 2008). To prevent the aggregation of CD81 and its interaction other proteins six intracellular Cys residues have been mutated to Ala to stop palmitoylation. Making these alterations to the protein has no detrimental effects on the stability and function (Berditchevski & Odintsova, 1999).

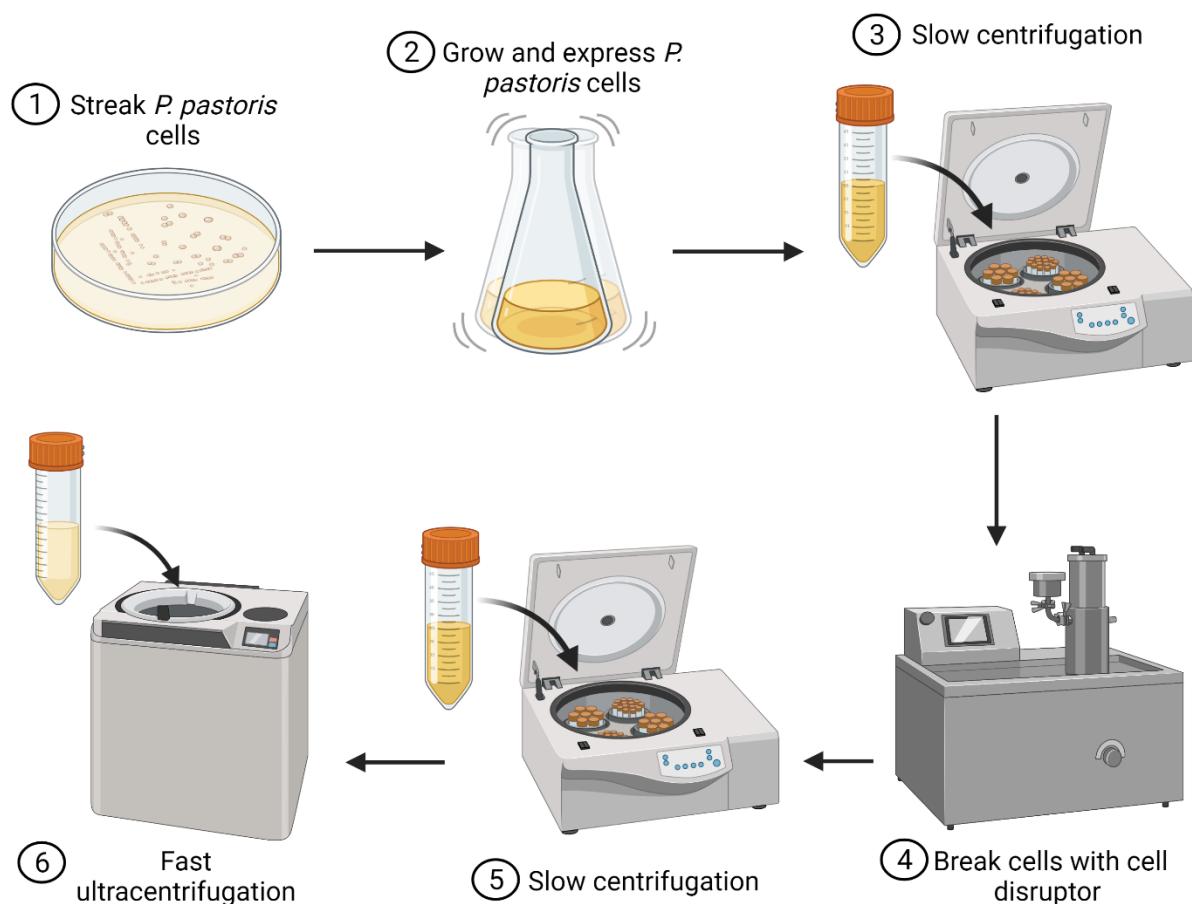


Figure 1.20. Illustration of the process of growing *P. pastoris* cells for membrane protein overexpression. (1) *P. pastoris* cells that have been transformed with a plasmid containing a gene for the protein of interest are streaked on an agar plate. (2) A colony is picked from the plate and grown in liquid media. (3) After cell growth the cells are harvested using a slow speed centrifugation ($\sim 4,000 \times g$). (4) Harvested cells are resuspended in a suitable buffer and broken using high pressure. (5) The broken cells are centrifuged at slow speed ($\sim 4,000 \times g$) to remove debris and unbroken cells. (6) The supernatant from step 5 is centrifuged at high speed ($\sim 100,000 \times g$) to obtain a pellet which contains the membrane that is then resuspended in a suitable buffer (Created with BioRender.com).

The benefits of using the *P. pastoris* expression include the fact that it can perform post-translational modifications, such as phosphorylation, glycosylation and disulfide bond formation (Daly & Hearn, 2005). Disulfide bond formation is particularly useful when expressing tetraspanins because of the disulfide bonds formed between Cys residues in the LEL (Kitakodoro et al., 2001).

Another benefit of *P. pastoris* is that it is a relatively cheap expression system and it can be grown in media which uses methanol as a carbon source. Using methanol as a carbon source is beneficial because the risk of contamination is lowered because contaminants that are methanol intolerant are not able to grow (Routledge et al., 2016). The *P. pastoris* cells possess two alcohol oxidase genes, *AOX1* and *AOX2*, which are dependent on methanol. The inclusion of methanol in the induction media induces *AOX1*, while other carbon sources, such as glycerol used during the initial growth stages represses *AOX1* (Routledge et al., 2016).

While there are many benefits of using *P. pastoris* to overexpress CD81 there may also be some limitations. For instance, CD81 is known to interact with cholesterol (Charrin et al., 2003; Zimmerman et al., 2016) but *P. pastoris* cells do not contain cholesterol (Grillitsch et al., 2014). Instead of cholesterol, *P. pastoris* cells synthesise ergosterol. The amount of ergosterol in the *P. pastoris* membrane is low in comparison to the other commonly used yeast expression system, *Saccharomyces cerevisiae*. *S. cerevisiae* membranes have ~400 µg of ergosterol per mg of protein in the membrane which is eight times as much as the ~50 µg/mg of protein which is commonly found in *P. pastoris* membranes (Zinser et al., 1993; Grillitsch et al., 2014). It is possible to utilise a strain of *P. pastoris* that has been engineered to synthesise cholesterol (Hirz et al., 2013). Another expression system that can be used if the presence of cholesterol is important is mammalian cell culture.

1.8.2 Expression of recombinant tetraspanins in mammalian cells

Mammalian cell expression for membrane proteins offers many advantages. The biggest advantage of expressing a mammalian protein of interest in mammalian cells is that the protein is expressed in an environment that is suited to its expression. Gene synthesis, processing and post-translational modifications that can be tricky or impossible for other expression systems can be done with ease in mammalian cells (Verma et al., 1998). Despite these obvious advantages using mammalian cells does not come without its problems. It is an expensive expression system, especially when compared to a cheap system like yeast. It requires more hands-on work by the researcher, there is a risk of viral contamination and scaling up to produce high quantities of protein can be difficult. Part of the issue with scalability is that many cells used, such as human embryonic kidney cells (HEK293) exist as a 2D

monolayer which means that high density growth seen in bacteria and yeast is not easy (Forstner et al., 2007).

Commonly used cells are HEK293T cells and Chinese hamster ovary cells (CHO) which can be transfected using either polyethylenimine (PEI) or calcium phosphate (Khan, 2013). HEK293 cells are particularly amenable to the use of PEI with expression of green fluorescent protein (GFP) reaching 50-80% when transfected using this method (Huh et al., 2007). Transfections can either be stable or transient with the difference between the two being that in stable transfections the foreign DNA introduced to the host cells is integrated into the host cell genome allowing for long term expression of the protein of interest, whereas there is no integration of the DNA in transient transfections (Fus-Kujawa et al., 2021).

CD81 has previously been shown to express endogenously in HEK293 cells (Cevik et al., 2012), so these cells fulfil the requirements to be able to produce CD81 efficiently and correctly. When looking at the interaction between tetraspanins and gangliosides HEK293 offer a promising expression system because they express many of the main gangliosides. GM3 and GM2 are known interactors with tetraspanins and both of these are expressed in HEK293 cells, as well as GD3 and GT1b but not GM1 (Cho et al., 2010).

1.9 Solubilisation of tetraspanins

After expressing membrane proteins, such as tetraspanins, in a suitable expression system, they can be extracted from the membrane for further study. The protein can either be extracted using a detergent which removes almost all native lipids from around the protein or the newer technology, styrene-maleic acid (SMA) copolymers, which extract the protein with its native lipids.

1.9.1 Solubilisation of tetraspanins with detergents

CD81 has previously been solubilised with detergents and the resulting purified protein has been used downstream for techniques such as, analytical ultracentrifugation, circular dichroism spectroscopy, ELISA and dynamic light scattering (Jamshad et al., 2008; Ayub et al., 2020). The three tetraspanins that have had their full-length structures solved were all solubilised in a detergent (Zimmerman et al., 2016; Umeda et al., 2020; Yang et al., 2020) further demonstrating that tetraspanins are amenable to detergent solubilisation.

After expressing a tetraspanin in a suitable expression system it needs to be solubilised so that it can be purified for further downstream analysis. Membrane proteins are routinely solubilised using amphipathic detergents that mimic the plasma membrane. Detergent monomers are amphipathic due to their hydrophilic head group and their hydrophobic alkyl tail (Prive, 2007). These properties are like those found in lipids in the lipid bilayer, therefore

detergents are extensively used to extract membrane proteins in an environment like that which is found in the membrane.

In solution detergent monomers will come together to form a detergent micelle which has a hydrophobic core and a hydrophilic exterior because the tails of the detergent monomers point inwards towards each other (Oliver et al., 2014). At low detergent concentrations detergent monomers will form a monolayer at the air-water interface until the critical micelle concentration (CMC) is reached by increasing the concentration of detergent (Moraes et al., 2017). The CMC is the concentration at which the detergents will form micelles. Once the concentration reaches the CMC the detergent monomers forming the monolayer will partition into the water and eventually form micelles to protect their hydrophobic tails from the water (Figure 1.21).

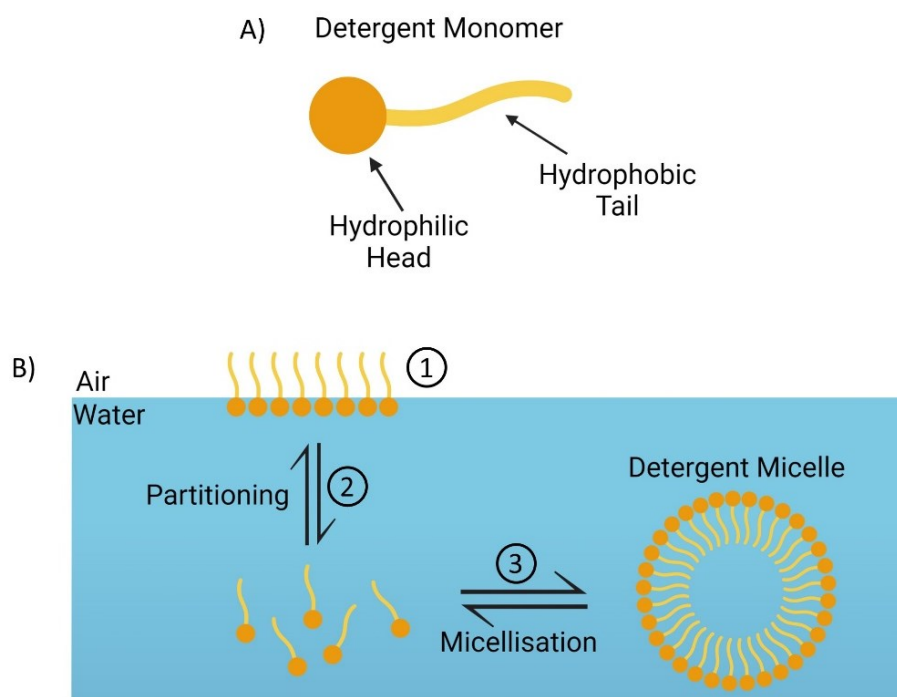


Figure 1.21. Formation of detergent micelles in solution. (A) Basic diagram of a detergent monomer showing the hydrophilic head and hydrophobic tail. (B) (1) At low CMC the detergent monomers form a monolayer at the air-water interface with their hydrophilic heads entering the water while their hydrophobic tails remain outside the water. (2) Increasing the concentration of detergent leads to partitioning. (3) Detergent monomers that are in solution at a concentration above their CMC will form micelles with a hydrophilic outside and hydrophobic inside to protect their hydrophobic tails from the surrounding water environment (Created with BioRender.com).

Low concentrations of detergent will not disrupt the membrane because the CMC is not reached, and micelles cannot be formed. Detergent monomers, however, will intermittently bind to the membrane. Increasing the concentration of detergent will disrupt the membrane and allow for the encapsulation of membrane proteins in detergent micelles (Kalipatnapu & Chattopadhyay, 2005) (Figure 1.22).

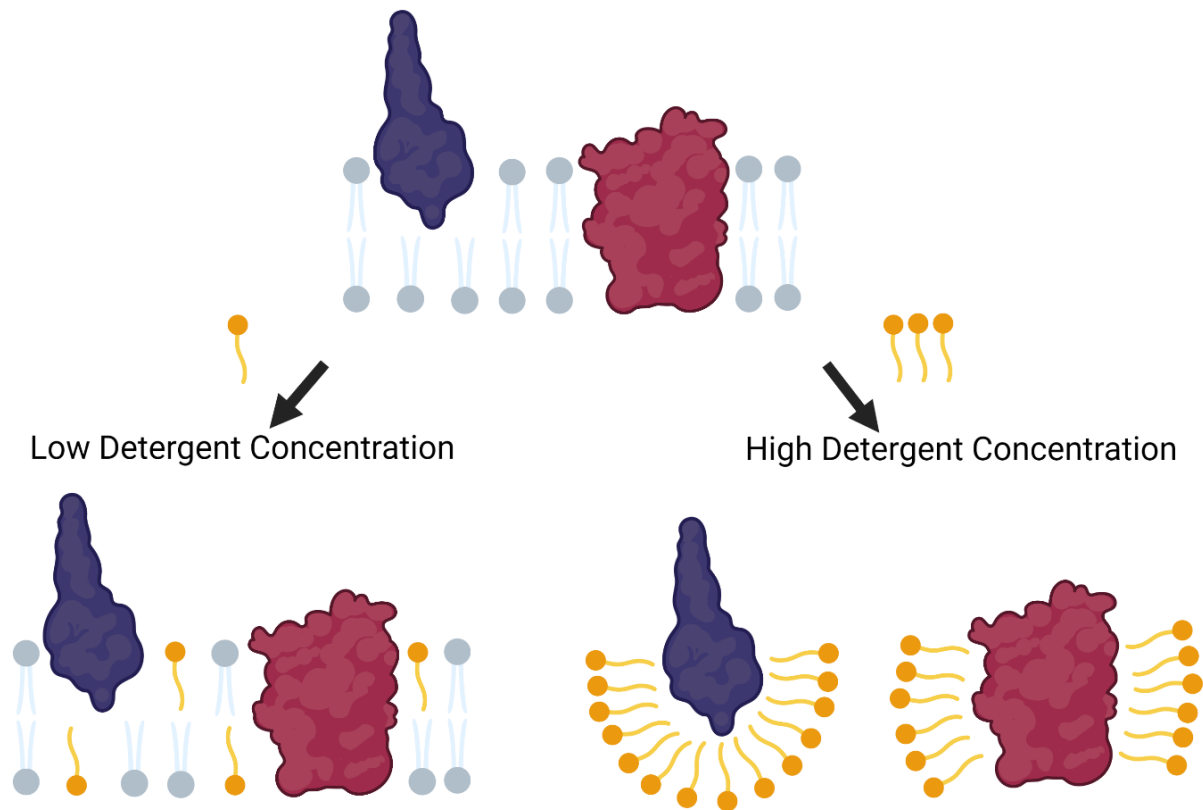


Figure 1.22. Illustration of how high concentration of detergent is required to solubilise membrane proteins. At low concentrations of detergent, below the CMC, monomers will be able to bind to and insert into the membrane, but it will be unable to solubilise membrane proteins. Increasing the concentration of detergent to a higher concentration, above the CMC, means that membrane proteins can be solubilised (Created with BioRender.com).

Commonly used detergents are non-ionic 'mild' detergents, such as *n*-dodecyl- β -D-maltopyranoside (DDM), *n*-decyl- β -D-maltopyranoside (DM) and *n*-octyl- β -D-glucopyranoside (β -OG). These are mild in nature and are less likely to disrupt any protein-protein interactions, unlike harsher detergents. They only disrupt interactions between lipids and between proteins and lipids (Moraes et al., 2014). The chemical structures of detergents are a determining factor for its CMC. The hydrophilic head group of detergent monomers interacts with the protein, while the CMC is determined by the length of the alkyl chain. Detergents with longer alkyl chains like DDM, which has a 12-carbon chain, have a much lower CMC (0.0087%) (Figure

1.23) when compared to DM and β -OG which have 10 carbons and 8 carbons in their respective chains and higher CMCs (0.087% and 0.53% respectively) (Prive, 2007).

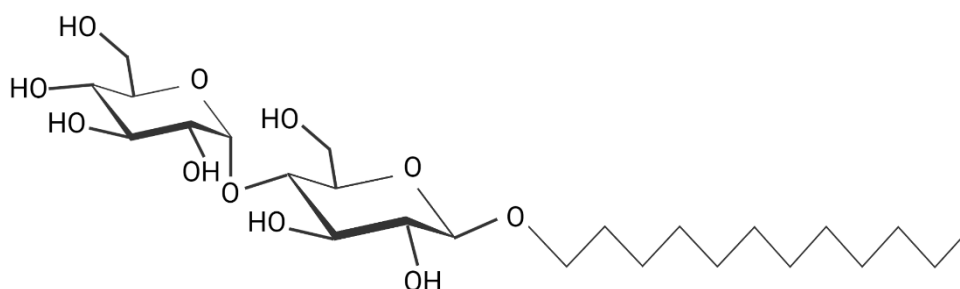


Figure 1.23. The chemical structure of *n*-dodecyl- β -D-maltopyranoside. Structure of DDM showing the 12-carbon hydrophobic chain as the tail and a hydrophilic maltose head consisting of two glucose molecules (Created with BioRender.com).

While the amphipathic nature of detergents provides a membrane-like environment for solubilised proteins it does not provide a native environment. Some membrane proteins are reliant on the native lipids that surround and interact with them for their stability and function (Seddon et al., 2004). For example, the maltose ABC transporter MalFGK2 which is kept in its preferred low energy conformation state by phospholipids with long acyl chains, whereas phospholipids with short acyl chains lead to the protein becoming unstable (Bao et al., 2013). Therefore, removing a membrane protein from its native lipid environment can have a detrimental impact on stability and function. It is possible, however, to extract the membrane protein in its native lipid environment using newly developed styrene-maleic acid (SMA) copolymers.

1.9.2 Solubilisation of tetraspanins with styrene-maleic acid

Using SMA copolymers means that the membrane protein can be solubilised with its surrounding lipids intact, forming SMA lipid particles (SMALPs) nanodiscs (Overduin & Esmaili, 2019). SMALP nanodiscs are formed because of the styrene ring, part of the SMA copolymer, inserting into the membrane between the acyl chains of phospholipids (Orekhov et al., 2019).

Commonly used SMA copolymers have a styrene-to-maleic acid ratio of 2:1 or 3:1 which means that there are two or three styrenes for every maleic acid (Morrison et al., 2016). SMA2000 is a 2:1 copolymer extensively used which is approximately 7.5 kDa in size on average and has been used previously to efficiently solubilise CD81 expressed in *P. pastoris* (Ayub et al., 2020).

SMA copolymers are comprised of a series of styrene and maleic acid residues in anhydride form. Styrene-maleic anhydride (SMA_{nh}) is hydrophobic, so it is unable to insert itself into the plasma membrane and solubilise lipids and membrane proteins (Lee et al., 2016). To be able to solubilise the plasma membrane it is vital that SMA_{nh} is hydrolysed to form SMA which is amphiphilic and soluble (Figure 1.24). The most common method used to hydrolyse SMA_{nh} is reflux, but it is also possible to autoclave on a conventional 121°C liquid cycle (Kopf et al., 2019).

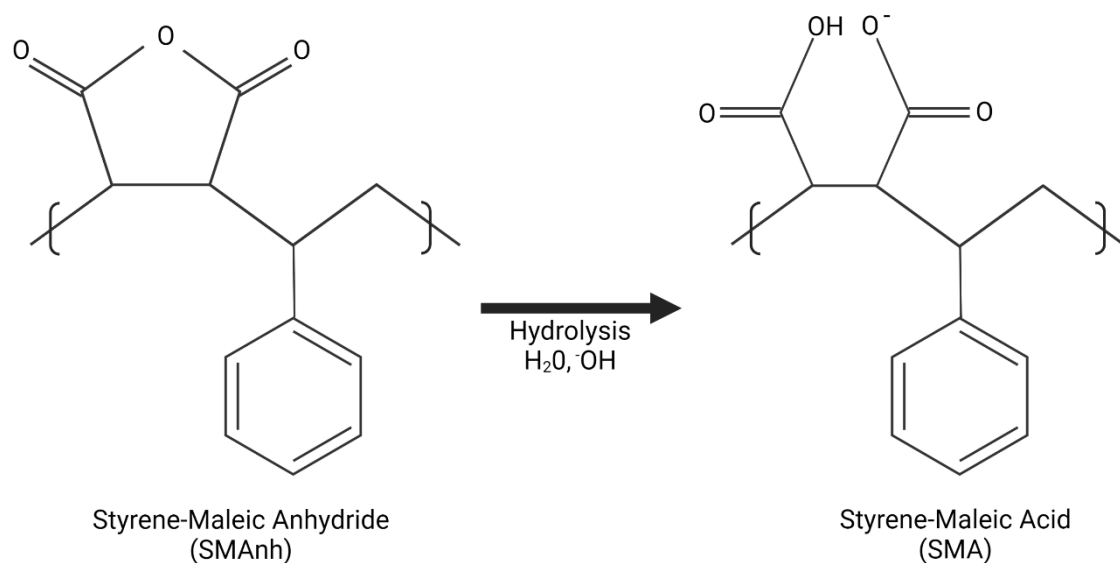


Figure 1.24. Hydrolysis of styrene-maleic anhydride to styrene-maleic acid. The hydrophobic styrene-maleic anhydride is hydrolysed by refluxing or autoclaving to the amphiphilic and soluble styrene-maleic acid which can solubilise membranes (Created with BioRender.com).

Working with SMA comes with benefits and limitations. Aside the obvious benefit of being able to solubilise membrane proteins in their native lipid environment (Figure 1.25), another benefit is that when solubilising SMA is indiscriminate in what it solubilises with no preference for any phospholipid (Overduin & Esmaili, 2019). As a result, subsequent studies can be carried out using techniques like thin layer chromatography (TLC) and/or mass spectrometry to analyse which lipids have been solubilised with the membrane protein. Another benefit of working with SMA is that the solubilised membrane protein will remain stable at room temperature for days, unlike detergent solubilised membrane proteins which usually need to be kept at 4°C (Gulamhussein et al., 2019).

Despite these positives, SMA can come with some problems. For instance, it is susceptible to divalent cations, although styrene-maleimide (SMI), which substitutes the maleic acid found

in SMA for maleimide, and disobutylene-maleic acid, which substitutes styrene in SMA for disobutylene, can be used to reduce cation sensitivity (Hall et al., 2018; Oluwole et al., 2017). Even though some nanodiscs formed can be as large as 30 nm in diameter, the average size is approximately 10 nm which may be problematic when working with large proteins (Craig et al., 2016).

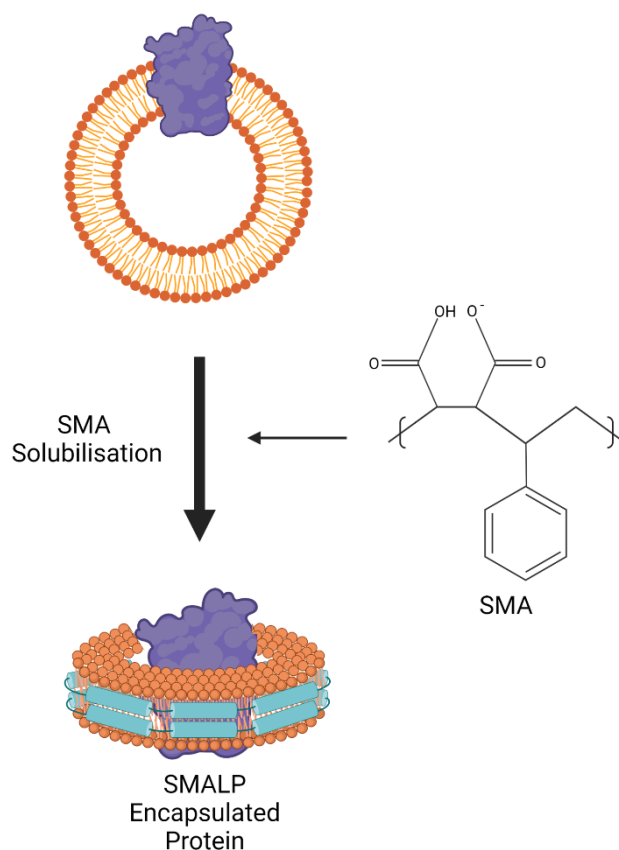


Figure 1.25. Solubilisation of membrane proteins using SMA. Overexpressed membrane proteins are solubilised using SMA co-polymers at an approximate concentration of 2.5% to extract the protein from the membrane still surrounded by its native lipids (Created with BioRender.com).

Using SMA has provided new insights into certain membrane proteins. There have been developments in recent years with regards to the use of Cryo-EM of membrane proteins encapsulated in SMALPs to determine their structure. For instance, this technique was used to determine the structure of the *Flavobacterium johnsoniae* protein, Alternative Complex III, at a resolution of 3.4Å with some annular lipids detected (Sun et al., 2018). ArcB has also had its structure solved after being solubilised with SMA at a resolution of 3.2Å (Qiu et al., 2018). Its structure showed a central cavity that was occupied by 24 lipids. Both examples demonstrate the advantages of solubilising membrane proteins in SMALPs and retaining some of the lipids that surround and interact with the proteins in biological membranes.

Solubilising membrane proteins in SMALPs can also have a positive effect on the function of the protein. Solubilising with detergents can have a detrimental effect on, such as on the rhomboid protease GlpG but when it is solubilised using a polymer its functional activity is closer to the activity seen when it is in the membrane (Barniol-Xicotá & Verhelst, 2018).

The protein studied in this thesis is CD81, which is approximately 25 kDa, therefore 10 nm diameter nanodiscs have proven suitable for solubilisation (Ayub et al., 2020). CD81 has previously only been solubilised using SMA2000 that has been prepared by refluxing, therefore part of this thesis will look at the solubilisation efficiency of SMA2000 prepared using the autoclave method.

1.10 Aims and objectives

The objectives of this thesis were:

- To create a universal residue numbering system for human tetraspanins.
- To characterise the structures of the LELs of tetraspanin subfamilies based on their Cys patterns in the LEL.
- To optimise the purification of CD81 expressed in *P. pastoris* using SMA2000 prepared by autoclaving.
- To purify CD81 mutants that can be used for PELDOR experiments to investigate the conformational change in CD81.
- To investigate the interaction between tetraspanins and gangliosides using Galaxy7TM.

Chapter 2 – Materials and Methods

2.1 Materials

This project used *Pichia pastoris* X33 cells that have been transformed with human CD81-p-null in the pPICZB expression plasmid containing a 6x His tag on the C-terminus (Figure 2.1) (Jamshad et al., 2008). Cells were stored at -80°C.

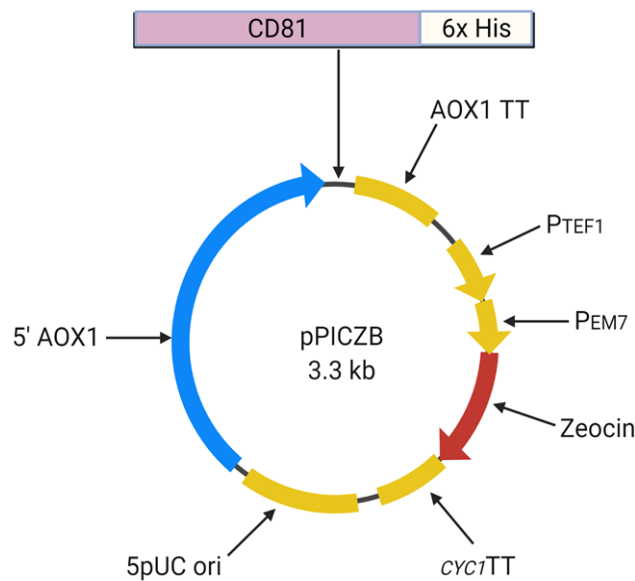


Figure 2.1. The pPICZB plasmid map. pPICZB is a vector 3328 bp in length containing the CD81-p-null gene. CD81-p-null has had all palmitoylation sites mutated out and it was cloned into the vector with *EcoRI* and *XhoI*. The plasmid possesses a zeocin resistance gene for growth on zeocin resistance agar plates and the *AOX1* promoter to induce expression in media containing methanol. At the C-terminal of CD81-p-null is a 6x Histidine tag to aid purification with Ni-NTA resins and expression analysis when doing Western blots.

SMA 2000 copolymer at a ratio of 2:1 was purchased from Cray Valley for SMA solubilisations, while n-dodecyl-B-D-maltoside (ThermoFisher) was purchased as pure powder and cholesteryl hemisuccinate (Sigma-Aldrich) were used for detergent solubilisation.

One Shot™ TOP10 Chemically Competent *E. coli* cells were purchased from ThermoFisher Scientific and stored at -80°C.

Zeocin was purchased in 1.25 ml tubes at a concentration of 100 mg/ml from ThermoFisher Scientific and stored at -20°C.

HEK293T cells were obtained from ATCC (CRL-3216™) and stored in liquid nitrogen until their use.

This project used a pEF6.A plasmid that contains wild-type CD81 (Figure 2.2) so that HEK293T cells can be transfected with it to produce exogenous CD81. The plasmid was a gift from Dr Mike Tomlinson at the University of Birmingham.

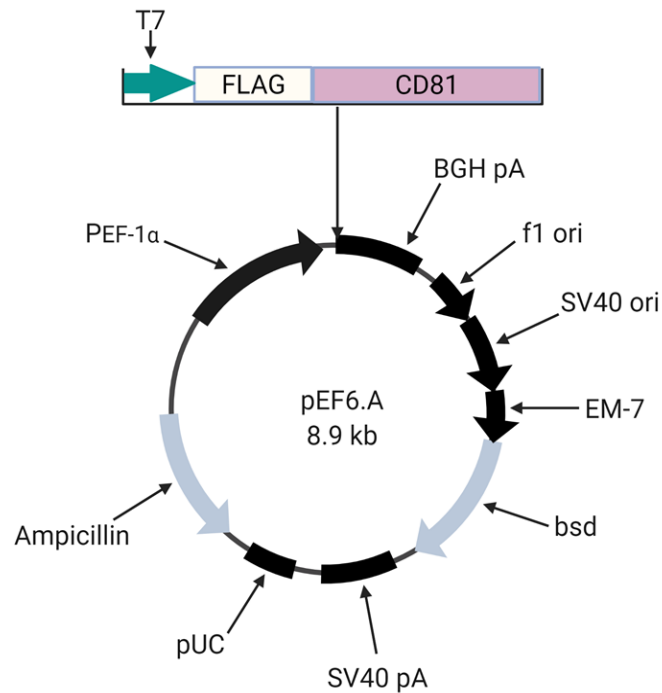


Figure 2.2. The pEF6.A plasmid map. pEF6.A is a vector 8980 bp in length containing the wild-type CD1 gene. The plasmid contain a blasticidin and an ampicillin resistance gene. At the N-terminal of CD81-p-null is a FLAG-tag to aid purification and expression analysis when doing Western blots.

2.2 Pichia pastoris growth

2.2.1 Stock solutions and growth media

Stock Solutions

10x Yeast Nitrogen Base (YNB) with Ammonium Sulphate and without Amino Acids

134 g YNB with ammonium sulphate and without amino acids was dissolved in 1 L of water. To fully dissolve the solution was heated at approximately 70°C for 20-30 minutes. It was filter sterilised using a 0.22 µm filter and stored at 4°C.

0.02% Biotin

20 mg of biotin was dissolved in 100 ml of water and filter sterilised using a 0.22 µm filter and stored at 4°C.

20% Dextrose

200 g of D-glucose was dissolved in 1 L of water and autoclaved on a liquid cycle at 121°C for 20 minutes and then stored at 4°C.

5% Methanol

5 ml of 100% methanol was mixed with 95 ml of water and filter sterilised using a 0.22 µm filter and stored at 4°C.

10% Glycerol

100 ml of glycerol was mixed with 900 ml of water and autoclaved on a liquid cycle at 121°C for 20 minutes and then stored at room temperature.

1 M Potassium Phosphate Buffer, pH 6.0

132 ml of 1 M K_2HPO_4 was mixed with 868 ml of 1 M KH_2PO_4 and the pH was adjusted to 6.0 ± 0.1 if required using phosphoric acid or KOH. It was autoclaved on a liquid cycle at 121°C for 20 minutes and stored at room temperature.

Growth Media

Yeast Extract Peptone Dextrose Medium (YPD)

1% yeast extract, 2% peptone, 2% dextrose, 2% agar, 100 mg/ml zeocin

In a 1 L solution 10 g yeast of yeast extract and 20 g peptone were dissolved in 900 ml of water. When making plated 20 g of agar is also added. It was autoclaved on a liquid cycle at 121°C for 20 minutes. The solution was allowed to cool to ≈60°C and then 100 ml of 20% dextrose was added. 100 mg/ml of zeocin was added as appropriate for a final concentration of 100 µg/ml.

Buffered Glycerol-complex Medium (BMGY) and Buffered Methanol-complex Medium (BMMY)

1% yeast extract, 2% peptone, 100 mM phosphate buffer, pH 6.0, 1.34% YNB, $4 \times 10^{-5}\%$ biotin, 1% glycerol or 0.5% methanol.

In a 1 L solution 10 g of yeast extract and 20 g of peptone were dissolved in 700 ml of water. This was autoclaved on a liquid cycle at 121°C for 20 minutes and allowed to cool to room temperature. The solution was made up to 1 L by adding 100 ml of 1 M potassium phosphate buffer (pH 6.0), 100 ml 10X YNB, 2 ml of 0.02% biotin and 100 ml of 10% glycerol (for BMGY) or 100 ml 5% methanol (for BMMY). The solution was stored at 4°C for up to two months.

2.2.3 Expression of recombinant CD81-p-null

Large scale expression of recombinant CD81-p-null was done in large baffled shake flasks. One *P. pastoris* colony was added to 50 ml of BMGY containing 100 µg/ml of zeocin in a 250 ml baffled flask and incubated at 30°C and 220 rpm overnight. The next day 5 ml of this seed culture was added to 200 ml of BMGY in a 1 L baffled flask at 30°C and 220 rpm for 24 hours. Cells in 200 ml BMGY were grown to $OD_{600} \approx 8-10$ and pelleted by centrifugation (4000 rpm for 10 minutes). The pellet was resuspended in 500 ml of BMMY to induce recombinant CD81-p-null expression in a 2 L baffled flask at 30°C and 220 rpm for 22-24 hours. After 22-24 hours cells were supplemented with 5 ml 100% methanol (1% v/v) and grown for a further 22-24 hours. After which cells were harvested by centrifugation (4000 rpm for 20 minutes at 4°C) and stored at -80°C.

2.3 *Pichia pastoris* membrane solubilisation

2.3.1 *P. pastoris* membrane preparation

Buffers

Breaking Buffer

5% glycerol, 2mM EDTA, 100 mM sodium chloride, 50 mM sodium phosphate monobasic (NaH_2PO_4), 50 mM sodium phosphate dibasic (Na_2HPO_4), pH 7.4, 50 mM sodium phosphate dibasic (Na_2HPO_4), pH 7.4

In a 1 L solution 50 ml of 5 % glycerol, 4 ml of 0.5 M EDTA, 5.84 g NaCl, 11.3 ml of 1 M NaH_2PO_4 and 38.7 ml of 1 M Na_2HPO_4 were added to 800 ml of water and the pH was adjusted to 7.4 and made up to 1 L. The solution was autoclaved on a liquid cycle at 121°C for 20 minutes and then stored at 4°C.

Resuspension Buffer (Buffer A)

20mM HEPES, 50mM sodium chloride, 10% glycerol, pH 7.0

In a 1 L solution 4.76 g of HEPES, 2.92 g of NaCl and 100 ml of glycerol were added to 800 ml water and the pH was adjusted to 7.0 and made up to 1 L. The solution was autoclaved on a liquid cycle at 121°C for 20 minutes and then stored at 4 °C.

P. Pastoris Cell Lysis

P. pastoris cells frozen at -80°C were lysed using the Emulsiflex-C3 cell disruptor (Avestin). First the cell pellet was mixed with ice-cold breaking buffer at a ratio of 1:3. For example if there the cell pellet weighed 50 g then 150 ml of breaking buffer would be used. One protease inhibitor cocktail tablet was dissolved for each 50 ml of breaking buffer, so in this scenario 3 tablets would be dissolved. The cells were run through the cell lyser approximately 5 times at 20,000 to 25,000 psi. Unbroken cells and debris were removed by centrifugation (10,000 x g for 10 minutes at 4°C) and then the supernatant was collected and ultracentrifuged (100,000 x g for 45 minutes at 4°C). The pellet was resuspended in Buffer A to give a wet pellet concentration of 160 mg/ml and homogeniser using a glass homogeniser. The membrane preparation was stored at -80°C for future use.

Solubilisation Buffer

20 mM HEPES, 200 mM sodium chloride, 10% glycerol, pH 8.0

In a 1 L solution 4.76 g of HEPES, 11.68 g of sodium chloride and 100 ml of glycerol were added to 700 ml of water and the pH was adjusted to 8.0 with sodium hydroxide before making the solution up to 1 L.

SMA Solubilisation

SMA^{nh} co-polymer was refluxed or autoclaved to hydrolyse the co-polymer as previously described (Lee et al., 2016; Rothnie, 2016; Kopf et al., 2019). When autoclaving, the SMA was autoclaved for two and three cycles so that the solubilisation efficiency of each could be tested. Yeast membranes were solubilised with SMA 2000. *P. pastoris* membranes expressing CD1-p-null at a wet pellet of 160 mg/ml were diluted four three fold with solubilisation buffer and incubated with 2.5% of the SMA polymer for 1 hour at room temperature on a roller. Solubilised material was ultracentrifuged by ultracentrifugation (100,000 × g for 20 minutes at 4 °C) to give a supernatant containing soluble CD81-p-null.

Detergent Solubilisation

P. pastoris membranes expressing CD81-p-null were solubilised with DDM (n-dodecyl-β-D-maltoside) in solubilisation buffer supplemented with 1% DDM and 0.1% cholesteryl hemisuccinate. Solubilisation was performed for 1 hour at 4°C using a magnetic stirrer. Solubilised material was ultracentrifuged by ultracentrifugation (100,000 × g for 1 hour at 4 °C) to give a supernatant containing soluble CD81-p-null.

2.4 CD81 Purification

2.4.1 SMALP-CD81 Purification

Ni-NTA resin and buffers

The optimised protocol for small scale purification used 3 ml of membrane preparation at a concentration of 60 mg/ml. 500 µl of Ni-NTA resin was used for the small scale purification.

Wash 1 buffer– 10 ml of solubilisation buffer supplemented with 30 mM imidazole

Wash 2 buffer – 5 ml of solubilisation buffer supplemented with 60 mM imidazole

Elution buffer – 1 ml of solubilisation buffer supplemented with 300 mM imidazole

Purification Method

Solubilised membranes were incubated with Ni-NTA resin that had been washed twice with water and once with the solubilisation buffer supplemented with 20 mM imidazole at 4°C overnight on a rocker. The next day the solution was poured through a chromatography column that had been washed with sodium hydroxide, then water and finally the solubilisation

buffer. The flow through was collected. The resin was washed with 20x the resin volume of wash 1 buffer, followed by a second wash of 10x the resin volume of wash 2 buffer. Finally, CD81-p-null was eluted off the column with 5x the resin volume with the elution buffer. Elution fractions were collected in 200 µl volumes.

2.4.2 DDM-CD81 Purification

Ni-NTA Resin and Buffers

Small scale purification used 3 ml of membrane at a concentration of 180 mg/ml mixed with 24 ml of solubilisation buffer. 500 µl of Ni-NTA resin was used for the small scale purification.

Wash 1 buffer – 5 ml of solubilisation buffer supplemented with 30 mM imidazole, 0.1% DDM and 0.01% CHS

Wash 2 buffer – 5 ml of solubilisation buffer supplemented with 40 mM imidazole, 0.1% DDM and 0.01% CHS

Elution buffer – 2.5 ml of Solubilisation buffer supplemented with 300 mM imidazole, 0.1% DDM and 0.01% CHS

Purification method

Solubilised membranes were incubated with Ni-NTA resin that had twice been washed with water and once with the detergent solubilisation buffer at 4°C for two hours while gently stirring. The solution was poured through a chromatography column which had been washed with sodium hydroxide, water and the detergent solubilisation buffer. The flow through was collected. The resin was washed 10x resin volume with wash 1 buffer, followed by a second wash of 10x resin volume with wash 2 buffer. Finally, the CD81 was eluted off the column with 5x resin volume of elution buffer. The elutions were collected in 500 µl volumes.

2.5 SDS-PAGE Gel Electrophoresis

Buffers

4x Laemmli sample buffer

4x Laemmli buffer was prepared using 2.4 ml 1 M tris pH 6.8, 0.8 ml 10% sodium dodecyl sulphate (SDS) stock, 4 ml glycerol, 0.01% bromophenol blue, 1ml β -mercaptoethanol and 2.8 ml water. The solution was thoroughly mixed and stored at room temperature.

Running Buffer

100 ml of 10x Tris/Glycine/sodium dodecyl sulfate (SDS) buffer (National Diagnostics) containing 0.25 M tris, 1.92 M glycine and 1% SDS was diluted in 900 ml water.

SDS-PAGE Method

12% separating gels were prepared with reference to the recipe in (Table 2.1). APS and TEMED added last due to their quick polymerisation properties. The gel solution was poured and topped with isopropanol to ensure that the gel was level. After the separating gel (Table 2.2) had set the isopropanol was removed with filter paper to ensure all isopropanol was removed and the 4% stacking gel was poured on top of the separating gel and a 10 well comb was inserted into the solution and removed once set.

Table 2.1. SDS-PAGE 12% separating gel recipe.

Ingredient	Volume (to make 4 gels)
ddH ₂ O	7.2 ml
30% Polyacrylamide	9 ml
1.5 M Tris-HCl, pH 8.8	6 ml
10% SDS	240 μ l
Ammonium persulphate (APS)	80 μ l
Tetramethyl ethylenediamine (TEMED)	18 μ l

A 30 μ l sample was mixed with 10 μ l of 4x Laemmli sample buffer and heated at 95°C for 10 minutes. 25 μ l of sample was loaded into the gel alongside 3 μ l of protein ladder (Fisher). 1 L of Running buffer was poured into the gel tank and the gel was run at 160 volts for 1 hour or

until the dye front had reached the bottom of the gel. SDS-PAGE gels were then stained by incubated them with InstantBlue Coomassie protein stain (Expedeon) at room temperature and with gentle rocking.

Table 2.2. SDS-PAGE 4% stacking gel recipe.

Ingredient	Volume (to make 4 gels)
ddH ₂ O	6.2 ml
30% Polyacrylamide	1.4 ml
0.5 M Tris-HCl, pH 6.8	2.6 ml
10% SDS	50 µl
Ammonium persulphate (APS)	40 µl
Tetramethyl ethylenediamine (TEMED)	10 µl

2.6 Western Blotting

Buffers

Western Transfer Buffer

100 ml of 10x Tris/Glycine buffer (National Diagnostics) containing 0.25 M tris and 1.92 glycine was diluted with 700 ml of water and 200 ml of absolute methanol.

Phosphate Buffered Saline (PBS)

10 PBS tablets were dissolved in 1 L of water.

Blocking Buffer

20% blocking buffer was prepared by adding 2 g of Blotting-Grade Block Non-Fat Dry Milk (Bio-Rad) in 10 ml of PBS.

PBS-T Wash Buffer

1 L of PBS was mixed with 2 ml of tween-20 (0.2%).

Antibodies

The following antibodies were used in this thesis:

Table 2.3. Primary and secondary antibodies for Western blots. Primary antibodies were used against the 6xHis tag at the C-terminal of CD81-p-null, the LEL epitope in CD81 and the FLAG-tag at the N-terminal of wild-type CD81. The secondary antibodies bound to the primary antibodies to allow chemiluminescence.

Primary antibody	Dilution
6xHis monoclonal antibody (albumin free)	1:5,000
Anti-CD81 2s131	1:100
Rabbit anti-FLAG antibody (Sigma-Aldrich)	1:5,000
Secondary antibody	Dilution
Anti-mouse HRP-conjugated IgG (Cell Signalling Technology)	1:2,000
Anti-rabbit HRP-conjugated antibody (Cell Signalling Technology)	10 μ l

Western Blotting Method

Immediately after SDS-PAGE the gel was taken and the proteins were transferred from the gel to a 0.2 μ m nitrocellulose membrane (ThermoScientific). The membrane along with four fibre pads and four filter papers (Whatman 3mm chromatography paper) were soaked for approximately 10 minutes in Western transfer buffer.

The transfer was prepared in the following order:

- 1 fibre pad
- 2 filter papers
- SDS-PAGE gel
- 0.2 μ m nitrocellulose membrane
- 2 filter papers
- 1 fibre pad
-

This preparation was loaded into the gel tank alongside an ice block and 1 L of Western transfer buffer was poured into the tank. The transfer was run at 90 volts for 75 minutes.

After the transfer the nitrocellulose membrane was removed and placed in a 50 ml Falcon tube with 10 ml of blocking buffer on a rocker for 1 hour at room temperature. The membrane was then removed from the tube and washed 3x for 15 minutes each time with PBS-T buffer. The membrane was then incubated in 10 ml of PBS-T with the appropriate concentration of primary antibody overnight at 4°C on a rocker. The following day the membrane was washed 3x for 15 minutes each time with PBS-T buffer. The membrane was then incubated in 10 ml PBS-T with the appropriate concentration of secondary antibody for 1 hour at room temperature. After that the membrane was washed 5x for 5 minutes each with PBS-T buffer. 4 ml of ECL Western Blotting Substrate (Pierce) was poured onto the membrane at the membrane was visualised using a G:Box Chemi XRQ (Syngene).

2.7 Site-Directed Mutagenesis

Plasmid

pPICZB plasmid containing human CD81-p-null (Figure 2.3) was used as a template for mutagenesis (Fig. 2.1) (Jamshad et al., 2008).

CD81	<u>MGVEGCTKCIKYLLEFVFNFEVFWLAGGVILGVALWLRHDPQTTNLLYLELG</u>	50
CD81-p-null	<u>MSVEGATKAIKYLLEFVFNFEVFWLAGGVILGVALWLRHDPQTTNLLYLELG</u>	50
CD81	<u>DKPAPNTFYVGIYILIAVGAVMMFVGFLGAYGAIQESQALLGTTFFTCCLVI</u>	100
CD81-p-null	<u>DKPAPNTFYVGIYILIAVGAVMMFVGFLGAYGAIQESQALLGTTFFTCCLVI</u>	100
CD81	<u>LFACEVAAGIWGFVNKDQIAKDVKQFYDQALQQAVVDDDDANNAKAVVKTFL</u>	150
CD81-p-null	<u>LFACEVAAGIWGFVNKDQIAKDVKQFYDQALQQAVVDDDDANNAKAVVKTFL</u>	150
CD81	<u>HETLDCCGSSTLTALTTSVLKNNLCPSPGSNII SNLFKEDCHQKIDDLFSG</u>	200
CD81-p-null	<u>HETLDCCGSSTLTALTTSVLKNNLCPSPGSNII SNLFKEDCHQKIDDLFSG</u>	200
CD81	<u>KLYLIGIAAIVVAVIMIFEMILSMVLCCGIRNSSVY-----</u>	236
CD81-p-null	<u>KLYLIGIAAIVVAVIMIFEMILSMVLAAGIRNSSVYHHHHHHH</u>	242

Figure 2.3. The CD81-p-null amino acid sequence. Amino acid sequence of CD81-p-null (Jamshad et al., 2008) with the wild-type CD81 sequence. Mutations are highlighted in the red boxes. Six Cys residues were mutated to Ala to make the protein palmitoylation deficient. A further Gly to Ser mutation was done in the second position. The 6x His tag can be seen at the end of the CD81-p-null sequence in the blue box.

Primers

The following primers were ordered from Eurofins Genomics for site-directed mutagenesis:

Table 2.4. Primers used in mutations of CD81-p-null. Forward and reverse primers for C97A, C104A, D122A and D128C mutations. C97A, C104A and D128C mutations were used in PELDOR work and D122A was used for CD81-ganglioside interactions.

Mutation	Primer	Sequence
C97A	Forward	5'GGGGACGTTCTTCACCGCGCTGGTCATCCTGTTT 3'
	Reverse	5' CAAACAGGATGACCAGCGCGGTGAAGAACGTCCCC 3'
C104A	Forward	5' CTGGTCATCCTGTTTGCAGCGGAGGTGGCAGCAGGTATC 3'
	Reverse	5' GATACCTGCTGCCACCTCCGCTGCAAACAGGATGACCAG 3'
D122A	Forward	5' CAGATCGCCAAGGCTGTGAAGCAGTTC 3'
	Reverse	5' GAACTGCTTCACAGCCTTGCGATCTG 3'
D128C	Forward	5' GTGAAGCAGTTCTATTGCCAGGCCCTACAGC 3'
	Reverse	5' GCTGTAGGGCCTGGCAATAGAACTGCTTCAC 3'

The AOX1 Forward Primer (5' GACTGGTTCCAATTGACAAGC 3') was used for sequencing.

PCR Reaction

The following PCR mix was created:

Table 2.5. Recipe for the PCR reaction for site directed mutagenesis.

Ingredients	Volume
PfuUltra II Hotstart 2x Master Mix (Agilent)	25 μ l
Forward primer (10 μ M)	1 μ l
Reverse primer (10 μ M)	1 μ l
pPICZB CD81-p-null plasmid (100 ng/ μ l)	1 μ l
ddH ₂ O	22 μ l

The following PCR reaction conditions were set up:

Table 2.6. PCR reaction temperature and timings for site-directed mutagenesis.

Step	Temperature	Time
1	95°C	2 minutes
2	95°C	20 seconds
3	65°C	20 seconds
4	72°C	60 seconds
5	Repeat 2 nd , 3 rd and 4 th step 18 times	
6	72°C	3 minutes
7	4°C	Until collection

Agarose Gel electrophoresis

A 1% agarose gel was made by mixing 1 g of agarose with 50x TAE (ThermoFisher) diluted into 100 ml of water. The mixture was boiled in a microwave until the agarose had fully dissolved. The solution was cooled to 50°C and 3 µl of 20,000x RedSafe™ (Intron Biotechnology) was added. The gel was poured and an eight well comb was inserted and removed when set. 1x TAE was poured into the gel tank until the gel was completely covered and 6 µl of sample (5 µl of DNA and 1 µl of 6x loading dye (New England Biolabs) was loaded into each well alongside the DNA ladder. The gel was run at 80 volts for 1 hour. The gel was removed from the tank and visualised on a G:Box Chemi XRQ (SynGene).

2.8 Transformations

2.8.1 *E. coli* Transformation

Low Salt LB Media

In a 1 L solution 10 g tryptone, 5 g sodium chloride and 5 g yeast extract (15 g agar if making plates) were dissolved into 950 ml of water and the pH was adjusted to 7.5 with sodium hydroxide before making the volume up to 1 L. The solution was autoclaved on a liquid cycle at 121°C for 20 minutes and then stored at room temperature. Zeocin at a final concentration of 100 µg.ml was added as appropriate.

Transformation

2 µl of mutated DNA (50 ng/µl) was mixed with 200 µl of TOP10 competent cells and incubated on ice for 30 minutes. The mixture was then transferred to a 42°C heat block for heat shock of bacterial cells for 45 seconds and then placed immediately back on ice for 10 minutes. 1 ml of low salt LB media was added the heat shocked bacteria and incubated at 37°C and 200 rpm for 1 hour. After which the 200 µl cells were plated a low salt LB agar plate and left to grow for 16 hours at 37°C.

Miniprep

The next day a colony was picked and grown in 5 ml of low salt LB with the appropriate concentration of zeocin at 37°C for 16 hours at 200 rpm. After growth the cells were pelleted and the supernatant discarded. A miniprep was then performed according to the protocol in the GeneJET Plasmid Miniprep Kit (Thermo Scientific).

Sequencing

15 µl of miniprepmed plasmid (50 ng/µl) was mixed with 2 µl of *AOX1* forward primer and sent to Eurofins Genomics for sequencing. Results were translated using ExPasy's Translate tool and the sequence was aligned with wild type human CD81 using Clustal Omega.

2.8.2 Yeast Transformation

For yeast transformations, 2 µg of mutated DNA was added to 25 µl of 2 mg/ml salmon sperm DNA, 36 µl 1 M lithium chloride, 200 µl of competent *P. pastoris* X33 cells and 240 µl 50% PEG solution. Tubes were vortexed to ensure they were mixed and incubated at 30°C for 30 min. Cells were then heat shocked at 42°C for 20 minutes. Cells were centrifuged at 6,000 rpm and resuspended in 1 ml YPD and allowed to recover at 30°C for 2 hours before plating 100 µl onto YPD agar plates and left to grow at 30°C for 3-4 days.

2.9 CD81 expression in HEK293T cells

2.9.1 HEK293T cell growth

Cell revival

HEK293T cells in a cryovial that were stored in liquid nitrogen were revived by warming the cells rapidly by holding it in the hands. The defrosted cells in the vial were slowly added to a T75 flask containing 5 ml of media. The flask containing the cells were incubated at 37°C in 5% CO₂.

Passaging cells

Dulbecco's Modified Eagle's Medium (DMEM) with Fetal Bovine Serum (FBS) added for a final concentration of 10% FBS was pre-warmed, along with autoclaved PBS and trypsin in a 37°C incubator for approximately 30 minutes before use. Media was removed from the flask and the cells were washed with 10 ml of PBS to ensure the removal of all old media. 2 ml of trypsin was added to the flask and incubated at 37°C for 5 minutes. After 5 minutes 10 ml of DMEM with FBS media was added to stop trypsinisation. A new flask was prepared by adding fresh DMEM with FBS media into the flask and trypsinised cells to a concentration of 1:5 (for 2 days before next passage) or 1:10 (for 3 days before next passage) and grown until cells reached a confluency of 80%. The new flask with cells was placed in the incubator at 37°C in 5% CO₂.

Counting cells

HEK293 cells were counted by taking 90 µl of cells in media and mixing them with 10 µl of trypan blue in an Eppendorf tube. 10 µl of the cells-trypan mixture were pipetted onto a hemocytometer and cells were counted under a microscope.

2.9.2 PEI transfection of CD81 in HEK293 cells

Preparation of PEI solution

A stock polyethylenimine (PEI) solution at a concentration of 1 mg/ml was made by taking approximately 30-40 mg of PEI weighed on a fine balance. The PEI was transferred to a 50

ml falcon tube and distilled water was added to a few ml below the required amount for a 1mg/ml solution. The tube was left overnight on a roller to ensure it is thoroughly mixed. The next day the solution was adjusted to pH 7.5 and filter sterilised under a tissue culture hood with a 0.22 µm filter. The stock PEI solution was stored at 4°C.

PEI transfection

One day before transfection 500,000 and 800,000 HEK293 cells in 2 ml of DMEM were plated in a 6-well plate and placed in the incubator at 37°C in 5% CO₂. The following day 2 µg of pEF6.2-wtCD81 DNA was added to 100 µl of serum free media in a 1.5 ml Eppendorf tube. DNA was mixed with the media by gently flicking the tube. 12 µl of the stock PEI solution was added to the serum free media and DNA mixture and mixed by vortexing in pulses. So, 15 one second pulses on the vortex to mix the solution. This solution was incubated at room temperature for 10 minutes. 0.6 ml of full media was added to the solution and mix. Media in the 6-well plate after one day of growth was removed and the complete transfection mix was added dropwise to the cells. Cells were placed in the incubator at 37°C in 5% CO₂ for 2 hours. After 2 hours the 2 ml of full media was supplemented, and cells were placed in the incubator at 37°C in 5% CO₂ for 24 hours.

2.9.3 HEK293 cell lysis

Preparation of RIPA buffer

150 mM sodium chloride, 50 mM Tris-HCL, pH 8.0, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS

100 ml of radioimmunoprecipitation assay (RIPA) buffer was prepared by mixing 3 ml of 5 M sodium chloride, 5 ml of 1 M Tris-HCl, 1 ml of Triton X-100, 5 ml of 10% sodium deoxycholate and 1ml of 10% SDS. The solution was made up to 100 ml with ddH₂O and then stored at 4°C.

Cell lysis

RIPA buffer containing a protease inhibitor cocktail (now referred to as the lysis buffer) was placed on ice. In the meantime, media was removed from cells and they were washed twice with ice-cold 1 ml of PBS. 250 µl of lysis was added to each well and pipetted numerous times onto the cells with the plate tilted on an angle to ensure that all cells are dislodged. The 250

µl lysis buffer with cells was transferred to Eppendorf tubes and incubated at 4°C on a spinning wheel for 45 minutes. After 45 minutes it was centrifuged for 10 minutes at 14,000 rpm at 4°C. The supernatant was collected and mixed with LSB for Western blot analysis.

2.10. Tetraspanin sequence and structural alignment

Tetraspanin sequences

The following tetraspanin sequences used in this thesis were taken from Uniprot:

Table 2.7. Uniprot codes for human tetraspanins.

Uniprot Code	Tetraspanin	Uniprot Code	Tetraspanin
P60033	CD81	A1L157	Tspan11
P21926	CD9	O95859	Tspan12
P19397	CD53	O95857	Tspan13
P27701	CD82	Q8NG11	Tspan14
P48509	CD151	O95858	Tspan15
P08962	CD63	Q9UKR8	Tspan16
P11049	CD37	Q96FV3	Tspan17
O60635	Tspan1	Q96SJ8	Tspan18
O60636	Tspan2	P0C672	Tspan19
O60637	Tspan3	Q12999	Tspan31
O14817	Tspan4	Q96QS1	Tspan32
P62079	Tspan5	Q86UF1	Tspan33
O43657	Tspan6	O00322	UP1a
P41732	Tspan7	O75841	UP1b
P19075	Tspan8	Q03395	ROM1
O75954	Tspan9	P23942	PRPH2
Q9H1Z9	Tspan10		

2.10.1 Alignment of human tetraspanin TMs

Sequence alignment of human tetraspanin TMs

The sequences from Table 2.7 were aligned with Clustal Omega using the default settings (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers et al., 2011). TM regions were located in

the MSA using a combination of the three known full-length structures of CD81 (PDB: 5tcx), CD9 (PDB: 6k4j) and CD53 (PDB: 6wvg). The transmembrane helices prediction server TMHMM 2.0 was also used (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) (Moller et al., 2001; Krogh et al., 2001). Any gaps in the sequence alignment in the TM regions were manually removed.

Analysis of TM sequence alignments

Aligned sequence were analysed in JalView (Waterhouse et al., 2009) with particular attention paid to the “Quality” score which is analogous to the BLOSUM62 score (Henikoff & Henikoff, 1992). These scores were taken and displayed in this thesis as BLOSUM62 scores to help determine conserved residues. The Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) was used to analyse and visualise the conservation in TM domains. The sequence conservation threshold was set at >70% or >80%. In some instances, amino acids were grouped together based on chemical and structural similarity.

The amino acid groups in the Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) were as follows:

Table 2.8. Amino acid groups in the Sequence Manipulation Suite: Color Align Conservation.

Group	Amino Acids
1	Gly, Ala, Val, Leu, Ile
2	Phe, Tyr, Trp
3	Cys, Met
4	Lys, Arg, His
5	Asp, Glu, Asn, Gln
6	Ser, Thr
7	Pro

If the conservation threshold was set at >80% conservation, for example, then positions that had one amino acid with >80% were coloured with a black background, while positions that consisted of >80% of amino acids belonging to one of the groups (Table 2.8) were coloured with a grey background.

2.10.2 Alignment of human tetraspanins LEL regions

The same human tetraspanin amino acid sequence from Uniprot were used, as above (Table 2.7), and the different region of the LEL, the α helix, the β helix, the variable loop and the ϵ

helix were located using the known structures of CD81 (PDB: 5tcx), CD9 (PDB 6k4j) and CD53 (PDB: 6wvg). For all other tetraspanins with no structural information available the secondary structures were predicted using the prediction server PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>) (McGuffin et al., 2000; Buchan & Jones, 2019). The amino acid sequences of each human tetraspanin was put into the PSIPRED 4.0 web server and predicted helical structures in the LEL were noted so that each specific structure could be isolated and analysed for sequence conservation.

In some instances, significant places in the sequences were used as a cut-off point for different structural regions. For instance, the end of the β helix was determined, in all human tetraspanins, to the first Cys residue in the CCG-motif at the start of the loop in the LEL region. The start of the ϵ helix was determined, in all human tetraspanins, to be the final Cys residue in the loop of the LEL region. These specific points in the sequences make good places to start or end a specific structural region because they are ubiquitous across all human tetraspanins.

Each helical structure in the LEL was aligned and analysed in JalView (Waterhouse et al., 2012) with the “Quality” highlighted as it is analogous to the BLOSUM62 score (Henikoff & Henikoff, 1992). The Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) was used to analyse and visualise the conservation in the helical LEL regions. The sequence conservation threshold was set at >80%.

2.10.3 Alignment of 3D structures of tetraspanins

To analyse the structural similarity of the TM domains and helical structure in the LEL the three full-length structures of CD81 (PDB:5tcx), CD9 (PDB: 6k4j) and CD53 (PDB: 6wvg) were used. In Swiss-pdbviewer (Guex & Peitsch, 1997) each specific region was isolated and locally superimposed on top of each other. For example, CD81 was opened in Swiss-pdbviewer and TM1 was isolated by removing all amino acids in the structure except for the amino acids in TM1. To do this all, amino acids that were to be removed were highlighted and then removed by selecting “Build” and then “Remove selected residues...”. The same was done with CD9 and CD53. These regions were then superimposed on top of each using the “Magic Fit” function in Swiss-pdbviewer. The RMSD between was calculated using the “Calculate RMSD...” function. The RMSD was calculated between CD81-CD9, CD81-CD53 and CD9-CD53 and then the average of these three was calculated.

2.11 Protein structure prediction of human tetraspanin LELs

Amino acids sequences of human tetraspanin LELs from at least one of each tetraspanin subfamily were entered in the submission page on Robetta (<https://robetta.bakerlab.org/submit.php>). The LEL sequences were taken from the end of TM3 to the start of TM4. The TrRosetta (TR) program was selected from the modelling options underneath the box where the sequence is entered (Figure 2.4).

Robetta Project ▾ Structure Prediction ▾ ☰

Submit a job for structure prediction
Please do not submit jobs under different user accounts. Such jobs will be removed.

Required

Target Name ⓘ ROM1_LEL

Protein sequence ⓘ
PGSLDEALEEGLVTALAHYKDTEVPGHCQAKRLVDELQLRYHCCGRHGKDFWGV
QWVSSRYLDPGDRDVADRIQSNVEGLYLTGVPFSCCNPHSPRPCLQNRLLSDSYAHPLFD
PRQPNQNLWAOGCHEVLEHLQD

or upload FASTA Choose File No file chosen

Optional

RoseTTAFold ⓘ TR ⓘ CM ⓘ AB ⓘ Predict domains ⓘ

Upload MSA ⓘ Choose File No file chosen

Submit 3 + 2 = Keep private ⓘ

Figure 2.4. The input page on Robetta with TrRosetta. The LEL amino acid sequence for ROM1 entered in the input page on Robetta. “TR” is selected to use the TrRosetta prediction server (Yang et al., 2020, Du et al., 2021). At the time of entering the sequence TrRosetta was the most accurate prediction server on Robetta. It is now RoseTTAFold.

2.12 Disulfide bond prediction in human tetraspanin LELs

To analyse the LEL structures created with TrRosetta the structures were uploaded to the SSBondPre prediction server (Gao et al., 2020) to predict disulfide bonds in the LEL. The PDB file for the predicted structure was uploaded and submitted to the server (Figure 2.5).

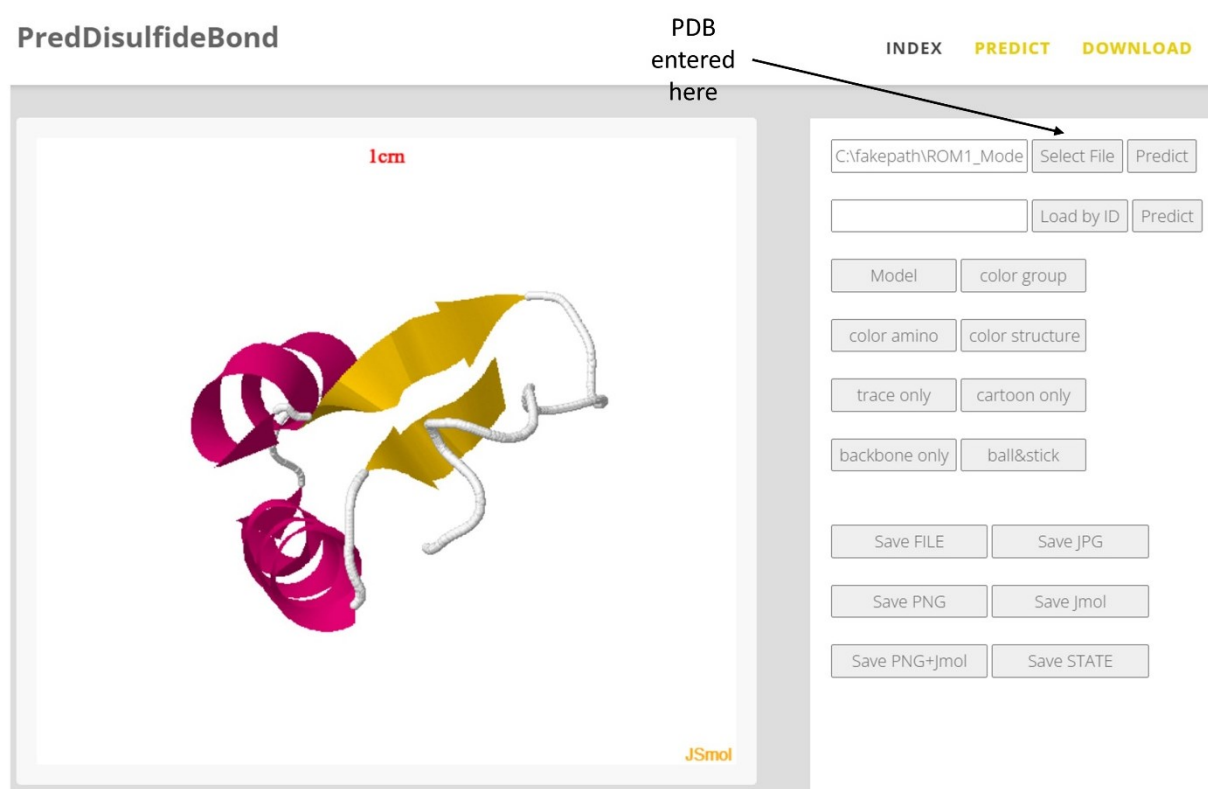


Figure 2.5. The input page for SSBondPre. The ROM1 predicted LEL structure was uploaded to the SSBondPre server (Gao et al., 2020) using the “Select File” option. After uploading the “Predict” button to the right of “Select File” was pressed.

The results of the SSBondPre were analysed with attention paid to disulfide bonds with an energy score of 0 (Figure 2.6). SSBondPre is prediction primarily used to look for residues that can be mutated to Cys to form new disulfide bonds, therefore these mutations will cause a change in energy. Looking at disulfide bonds between native Cys residues means that there will be no change in energy, therefore predicted disulfide bonds in the LEL structures were located by selected bonds with an energy score of 0.

	A	B	C	D	E	F
1		key	probability	entropy	energy	
2	0	CYSA43-VALA131	0.995	-57.9369	-1.5518	
3	1	CYSA43-CYSA128	0.994	-57.5043	0	
4	2	SERA96-ASNA122	0.994	-42.7317	-2.0615	
5	3	LYSA20-TRPA57	0.991	-47.1318	-0.7085	
6	4	ASNA103-ALAA110	0.991	-26.3674	-1.7261	
7	5	TYRA19-VALA34	0.99	-35.8721	0.8858	
8	6	CYSA92-CYSA100	0.989	-28.0327	0	
9	7	LEUA12-ARGA46	0.988	-46.0772	-0.5761	
10	8	HISA18-ALAA30	0.987	-33.0893	-0.5712	
11	9	TRPA52-VALA87	0.981	-46.4387	-1.3926	
12	10	CYSA100-LEUA123	0.98	-41.2027	-1.1066	
13	11	ALAA15-VALA34	0.979	-38.8201	0.7549	
14	12	CYSA44-CYSA91	0.978	-50.1152	0	
15	13	LEUA81-PHEA89	0.974	-28.0327	-0.6388	
16	14	ASPA85-SERA106	0.973	-40.0682	-0.4272	
17	15	ASNA103-PHEA114	0.972	-32.0042	-0.9227	
18	16	TYRA49-GLNA102	0.97	-51.6135	-1.4197	
19	17	VALA24-CYSA28	0.943	-19.3885	-0.6973	
20	18	HISA18-HISA27	0.938	-29.5016	-0.5052	
21	19	LEUA8-LEUA133	0.925	-62.3139	-0.1114	

Figure 2.6. The SSBondPre prediction of disulfide bonds in ROM1. The Excel file of predicted disulfide bonds by the SSBondPre prediction server (Gao et al., 2020). Energy scores of 0 were highlighted as potentially native disulfide bonds.

2.13. Homology modelling of CD82

The amino acid sequence of CD82 (Uniprot: P27701) was uploaded to SWISS-MODEL (Waterhouse et al., 2018) using CD9 (PDB: 6k4j) as a template. The structure assessment feature on SWISS-Model was used to analyse the structure that was created. Attention was paid to the Ramachandran plot (Ramachandran et al., 1963) but more attention was to the QMEANBrane analysis (Studer et al., 2014). Due to the difficulty of modelling large unstructured loops like the one seen in CD82 the QMEANBrane score provided a score for the local quality of the structure produced. QMEANBrane gives a score from 0-1 with scores lower than 0.50 considered to be areas of lower quality.

2.14. Docking ganglioside sugars to tetraspanins

2.14.1 Prediction of tetraspanin binding pockets

CD81 (PDB: 5tcx) and CD82 (modelled for this thesis) were uploaded to DeepSite (<https://playmolecule.com/deepsite/>) (Jimenez et al., 2017), a server which utilises neural networks to find binding pockets in a protein (Figure 2.7).

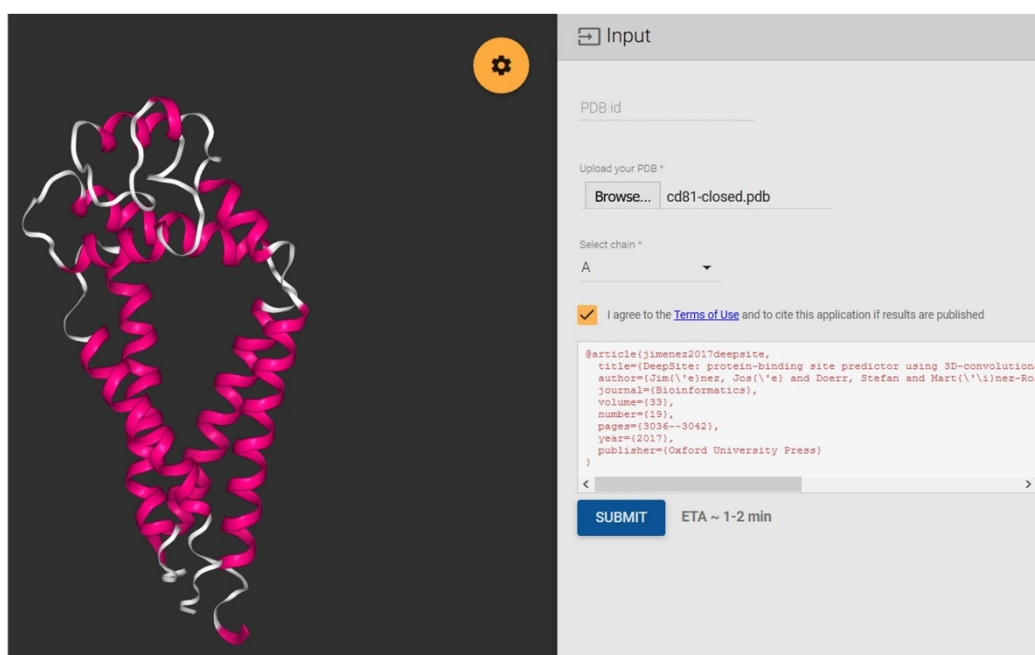


Figure 2.7. The input page for DeepSite. The closed CD81 structure (PDB: 5tcx) was uploaded to the DeepSite server (Jimenez et al., 2017) and submitted.

Potential binding sites were analysed by looking for binding pockets that had the highest score but were located at, or around, the membrane-water interface because this is where the ganglioside sugars will be situated when they are interacting with a protein (Figure 2.8).

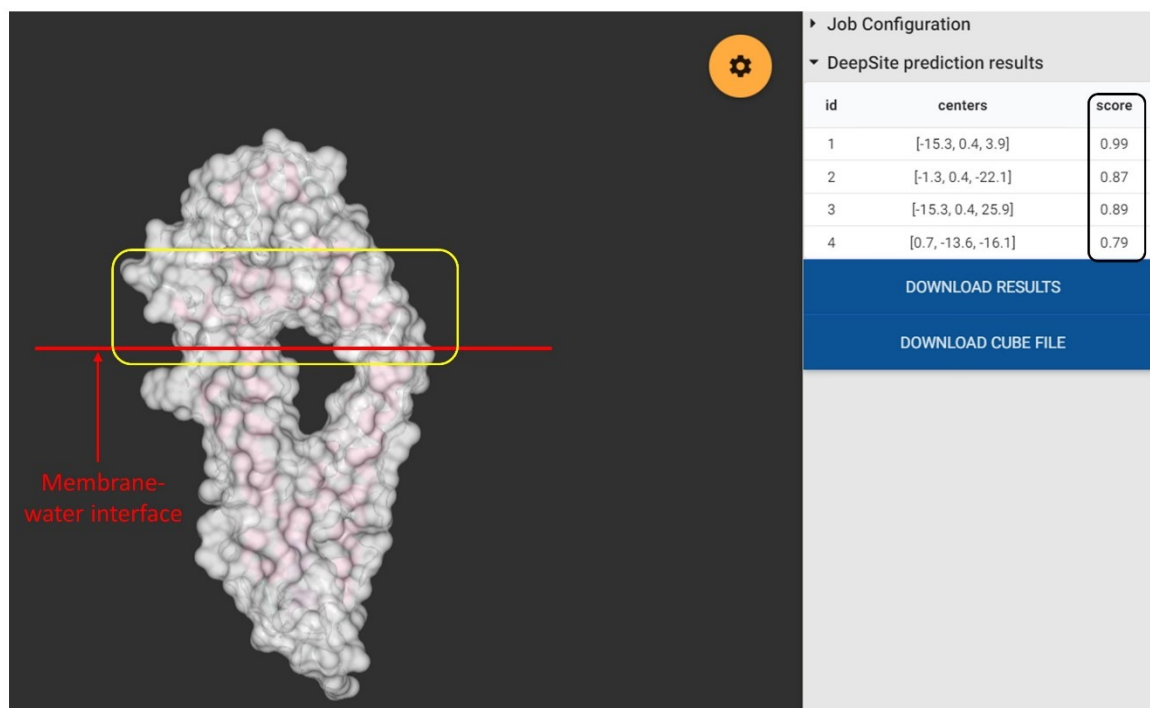


Figure 2.8. The DeepSite results page. DeepSite (Jimenez et al., 2017) results showing four potential binding pockets, their coordinates, and their scores. Since ganglioside sugars are situated around the membrane water interface (red line), binding pockets around this area (yellow box) were investigated further.

Clicking on the binding pocket under “DeepSite prediction results” zoomed in on the binding pocket on the ribbon structure. After finding suitable binding pocket at the membrane-water interface the amino acids within in the binding pocket were noted.

2.14.2 Docking of ganglioside sugars with Galaxy7TM

The Ganglioside sugars (glucose, galactose and disaccharide or glucose and galactose) were drawn using Corina (https://www.mn-am.com/online_demos/corina_demo_interactive) so that each one could be docked with CD81 (PDB: 5tcx) and CD82 (modelled in this thesis) separately. The amino acids found in the binding pocket from DeepSite (Jimenez et al., 2017) analysis were entered into the Galaxy7TM server (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=7TM>) (Seok & Lee, 2016) to dock the sugars in the binding pocket (Figure 2.9).

User Information	
Job name	<input type="text" value="CD81 Closed Galactose"/>
E-mail address (Optional)	<input type="text" value="093162849@aston.ac.uk"/>
Input GPCR and ligand structures	
PDB File (≤1000 AA)	<input type="button" value="Browse..."/> <input type="text" value="cd81-closed.pdb"/> Protein Structure File (allowed file extensions: pdb, txt)
Sequence File (Optional) (≤1000 AA)	<input type="button" value="Browse..."/> No file selected. Protein Sequence File (allowed file extensions: fa, fasta, seq)
Ligand File (≤150 atoms)	<input type="button" value="Browse..."/> <input type="text" value="B-D-Galactose.pdb"/> Ligand Structure File (allowed file extensions: mol2, pdb, xyz) <i>Note: Ligand structure with stereochemically wrong topology might results in inaccurate docking. (e.g. 2D-projected structure of non-planar ligands)</i>
Binding pocket residues (Optional)	
Binding pocket residues (≤10 res)	<input type="text" value="108,111,119,122,123"/> Residue numbers should follow the input PDB file residue numbering. Up to 10 residue numbers can be submitted in integers separated by commas (example: 51,64,78).
Refinement option	
Energy function	<input checked="" type="radio"/> GPCR (default) <input type="radio"/> Soluble protein (binding pocket residue numbers must be assigned above)
Submit	
<input type="button" value="submit"/> <input type="button" value="reset"/>	

Figure 2.9. The input page for Galaxy7TM. CD81 was uploaded to the Galaxy7TM web server (red box), as well as one of the ganglioside sugars as the ligand file (green box). The residues situated in the binding pocket that was found using DeepSite were entered in the “Binding pocket residues” section (light blue box).

Results from the docking were analysed by looking for docked sugars that align with each. For instance, the ideal situation would be to have a glucose, galactose and lactose molecule that have been docked in the place on the protein and overlay perfectly. To analyse this PDB files from the Galaxy7TM docking results were analysed in Swiss-pdbviewer (Guex & Peitsch, 1997). Sugars were overlaid to find sugars that were docked in a similar position. Using the LIGPLOT (Wallace et al., 1996) output on Galaxy7TM as a guide 2D images of sugar-protein interactions were drawn showing hydrogen bonds and hydrophobic interactions.

Amino acids involved in hydrogen bonds with the sugar residues were mutated to Ala by using the “Mutate” function in Swiss-pdbviewer (Guex & Peitsch, 1997). The mutated protein was uploaded to Galaxy7TM again, as well the ganglioside sugar residues. The docking energies of all 10 docked results were noted and the average docking energy was calculated.

Chapter 3 – Establishing a universal numbering system for human tetraspanins

Despite having three solved full length tetraspanins structure (CD81, CD9 and CD53) there is still much to be discovered with regards to the structure and function of tetraspanins. In order to shed light on some gaps in our knowledge the sequences and structures were aligned in combination to add greater clarity. In doing so, it has prompted the introduction of a universal numbering system for human tetraspanins to enable researchers to quickly and accurately describe the location of an amino acid. Introducing a universal numbering system to the field of tetraspanin research requires ‘anchor’ residues in each structured region (Figure 3.1) of tetraspanins to build the system around.

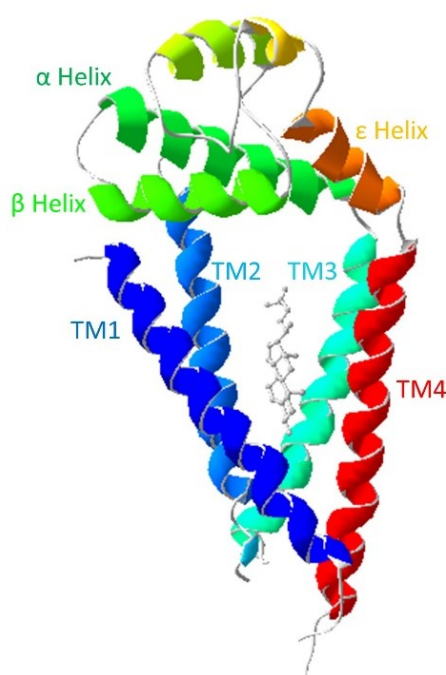


Figure 3.1. Structure of CD81 with regions labelled. The structure of CD81 (PDB: 5tcx) with the regions of the protein labelled. TM1 = Blue, TM2 = Light Blue, TM3 = Cyan, TM4 = Red, α Helix = Green, β Helix = Light Green, ϵ Helix = Orange.

3.1 Sequence and structural alignment of the TM1 domains of human tetraspanin

To find anchor residues for a universal numbering system the first step was to highlight the TM regions in human tetraspanins. Protein sequences of all known human tetraspanins were obtained from Uniprot and they were aligned using Clustal Omega (Sievers et al., 2011). The region of the alignment likely to include TM1 was located at the beginning of the sequences because it is known that all but a few human tetraspanins have small N-terminal tails (Figure 3.2A). Using the web server TMHMM (Moller et al., 2001) a prediction of amino acids that form TM1 in CD81 was produced (Figure 3.2B).

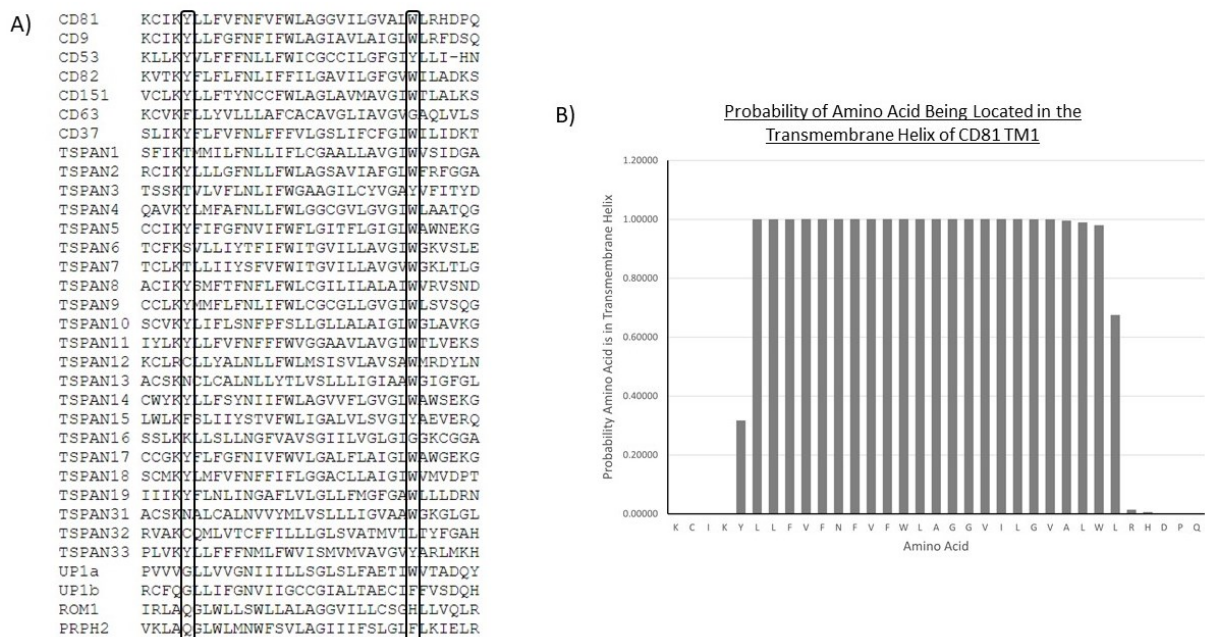


Figure 3.2. Determination of the TM1 region of human tetraspanins. (A) Sequence alignment of the regions containing TM1 of all known human tetraspanins aligned using Clustal Omega (Sievers et al., 2011) with residues that form part of an “aromatic belt” in the interfacial region highlighted in the black boxes. (B) Prediction of the probability that an amino acid in CD81 is likely to form part of the TM1 transmembrane helix using TMHMM (Moller et al., 2001).

The prediction indicates 22 amino acids with a high probability of forming part of TM1 starting at Leu-Leu-Phe and ending at Ala-Leu-Trp (Figure 3.2B). It was previously reported when the CD81 structure was elucidated and when subsequent tetraspanin structures have been solved that the Trp residue at position 23 is not part of the TM region (Zimmermann et al., 2016; Susa et al., 2021). The Trp residue is highly conserved with 70% sequence conservation (Figure 3.2A, 2nd black box) and as an aromatic residue Trp residues are often found in the interfacial region forming an “aromatic belt” (Braun & von Heijne, 1999). Combining the likelihood that the Trp residue is situated at the membrane-water interface and that previous literature has not included Trp in the TM1 it was omitted from further TM1 alignment analysis.

Aligning the TM1 regions of the known 33 human tetraspanins shows a high degree of conservation (Figure 3.3). Of the 21 amino acid positions that comprise TM1, 8 of them have >50% sequence conservation for one amino acid. Leu in the 2nd position (76% sequence conservation), Asn in the 6th position (82%), Phe in the 9th position (70%), Trp in the 10th position (55%), Leu in the 11th position (58%), Gly in the 13th position (82%), Leu in the 17th position (58%) and Gly in the 20th position (73%) all have >50% sequence conservation.

CD81	LLFVFNFWLGGVILGVAL
CD9	LLFGNFI FWLAGI AVLAI GL
CD53	VLFFNLLFWICGCCILCFGI
CD82	FLFLENLIFFILGAVILGFGV
CD151	LLFTYNCCFWIAGLAVMAVGI
CD63	LLYVLLLAFCACAVGLI AVGV
CD37	FLFVFNLFFFVLGSLIFCFGI
TSPAN1	MMILENLLIFL CGAALLAVGI
TSPAN2	LLLGENLLFWL AGSAVIAFGL
TSPAN3	VLVFNLI FWGAAGILCYVGA
TSPAN4	LMFAENLLFWLGGCGVLGVGI
TSPAN5	FIFGENVIFWFLGITFLCIGL
TSPAN6	VLLIYTFIFWITGVILLAVGI
TSPAN7	LLIIYSFVFWITGVILLAVGV
TSPAN8	SMFTENLFWLCCGILILALAI
TSPAN9	MMFLENLIFWLCCGGLLGVGI
TSPAN10	LIFLSNFFSLLGLLALAI GL
TSPAN11	LLFVFNFFWVGGAAVIAVGI
TSPAN12	LLYALNLLFWIMSI SVLAVSA
TSPAN13	CLCAINLLYTLVSLILICIAA
TSPAN14	LLFSYNI FWLAGVFLCVGL
TSPAN15	SLIIYSTVFWLIGALVLSVGI
TSPAN16	LLSLNFGFVAVSGIIVGLGI
TSPAN17	FLFGENIVFWVLGALFLAIGL
TSPAN18	LMFVNEFFIFLGGACLLAIGI
TSPAN19	FLNLTNGAFLVLGLLEFMCFGA
TSPAN31	ALCAINVVYMLVSLILIGVAA
TSPAN32	QMLVTCFFILLLGLSVATMVT
TSPAN33	LLFFFNMLFWVISMVAVGV
UP1a	LLVVGNI ILLSGLSLAETI
UP1b	LLIFGNVIIGCCGIALTAECI
ROM1	GLWLLSWLLADAGGVILLCSG
PRPH2	GLWLMNWFVLAGIIFSLGL

Figure 3.3. Sequence alignment and conservation of the TM1 domain in human tetraspanins. The TM1 domains of all known human tetraspanins were determined using TMHMM (Moller et al., 2001) and were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

Two of these positions, Asn in the 6th position and Gly in the 13th position, have the highest sequence conservation at 82% (Figure 3.3). These two residues form part of the highly conserved NGG7 motif found in TM1 of human tetraspanins (Kovalenko et al., 2005). In trying to find an ‘anchor’ residue in the TM1 region on which to base a proposed universal numbering system, further analysis of the sequence alignment was needed. BLOSUM62 scores were taken from JalView (Waterhouse et al., 2009) to look for conserved positions. BLOSUM62 long-odds score represents the biological probability of a substitution relative to the chance probability of the substitution. The substitution of chemically similar amino acids receives a positive score, and the substitution of chemically dissimilar amino acids receives a negative one (Henikoff & Henifoff, 1992).

The BLOSUM62 scores of each position in TM1 shows 9 positions with a BLOSUM62 score of >150 (Figure 3.4). The three most significant of these positions are the 2nd position, the 6th position and the 13th position. The 2nd position which is 76% Leu possesses the highest BLOSUM62 score of 204.36. Of the other two highlighted positions it is the 13th position which is 82% Gly which has a higher BLOSUM62 score of 192.19 compared to a score of 160.22 for the 6th position with is 82% Asn (Figure 3.4). The higher BLOSUM62 score means that the Gly at the 13th position is more often found by chance at this position than the Asn at the 6th position.

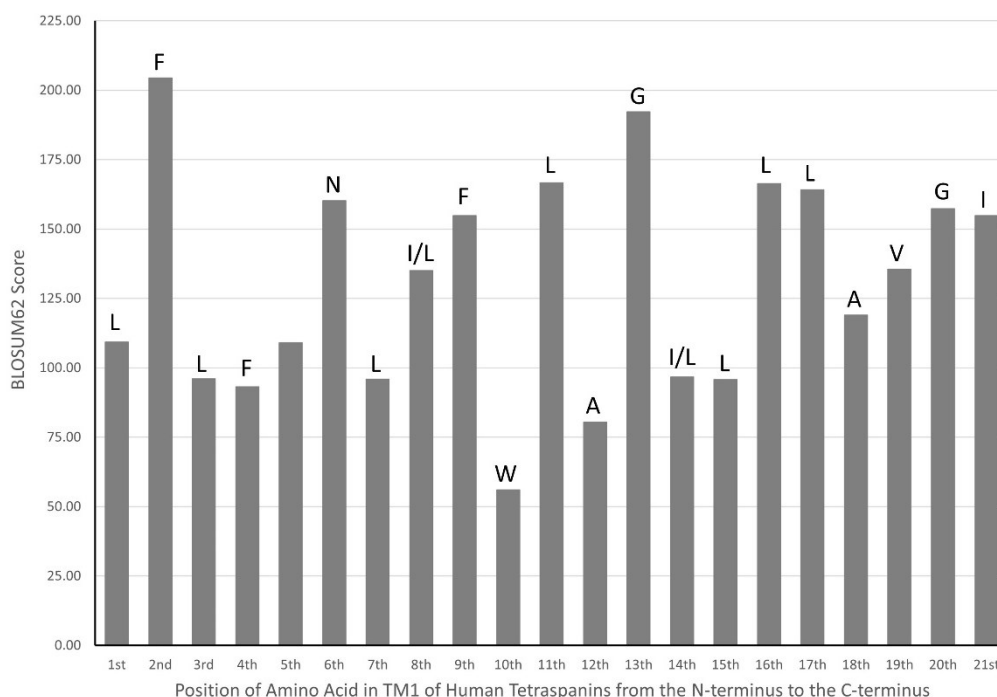


Figure 3.4. BLOSUM62 scores for each amino acid position in TM1 of human tetraspanins. The aligned TM1 domains were analysed in JalView (Waterhouse et al., 2009) to obtain BLOSUM62 scores. BLOSUM62 scores were plotted against each position located in TM1 with the residues located at the N-terminal end of TM1 and moving right leads to residues located at the C-terminal end of TM1. The most conserved residue in each position is written above the corresponding bar on the graph.

Analysis of the primary amino acid sequences provides a limited view of the tetraspanins family. The primary sequence was therefore examined in conjunction with the known full-length structures of human tetraspanins. To date, three human tetraspanins have had their full-length structures solved (CD81, CD9 ad CD53). The isolated TM1 regions of these three structures were locally aligned in DeepView and the RMSDs of each pairing were determined. The average of the three pairings was calculated giving a RMSD of 0.56 Å. This low RMSD indicates that the TM1 regions are structurally similar (Figure 3.5B).

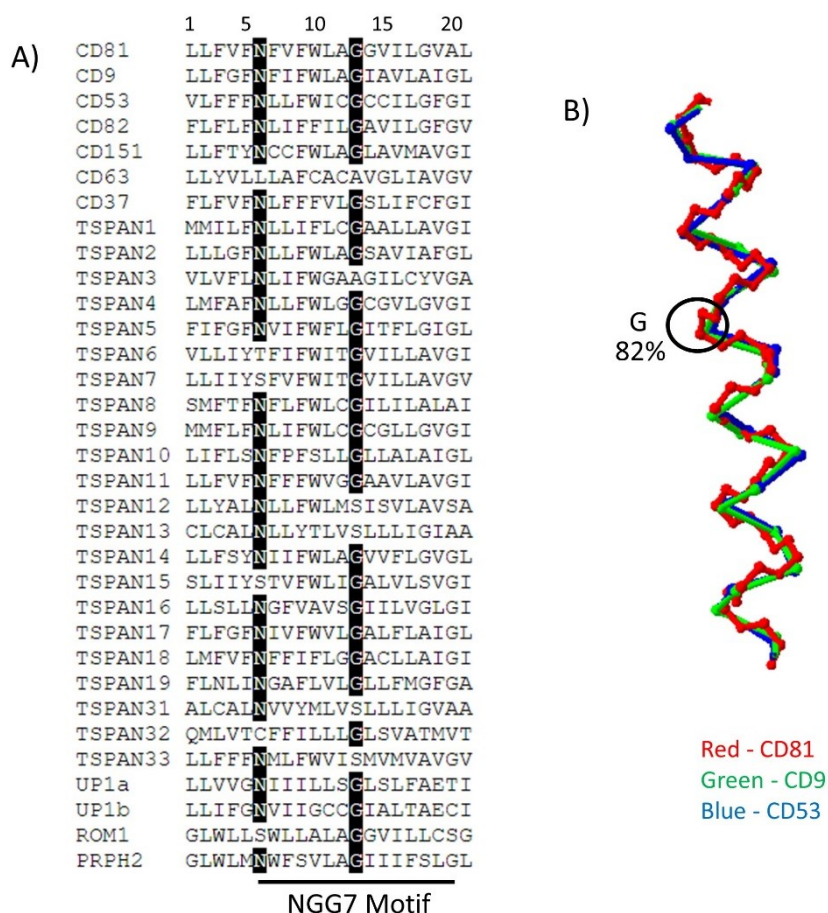


Figure 3.5. Sequence and structural alignment of the TM1 domain of human tetraspanins. (A) The sequence alignment of TM1 in which residues with >80% sequence conservation for one amino acid are highlighted in black. This shows two conserved residues; Asn in the 6th position (82%) and Gly in the 13th position (82%). The NGG7 motif that the Asn and Gly residue are constituents of is labelled underneath the sequence alignment. (B) The 3D structures of the TM1 domains of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 0.56 Å between the three structures was calculated. As the anchor residue, Gly (82% conserved) in the 13th position is highlighted in the black circle.

Given the Gly residue in the 13th position exhibits a high level of conservation and it occupies a similar position in the known 3D structures (Figure 3.5B, black circle) it was selected as the anchor residue in TM1 for the universal numbering system.

3.2 Sequence and structural alignment of the TM2 domains of human tetraspanins

The amino acids located in TM2 of CD81 were predicted using TMHMM (Moller et al., 2001) with 16 residues showing high probability of forming part of TM2 starting from Ile-Tyr-Ile to Val-Gly-Phe (Figure 3.6B). When looking at >80% conservation of the hydrophobic residues (Gly, Ala, Val, Leu and Ile) in the TM2 region the predicted Ile and Tyr residues at the start of the TM are not well conserved (Figure 3.6A). There is a >80% conservation of hydrophobic residues in the 13 positions highlighted between the two black boxes in Figure 3.6A. The Ala

and Ile in CD81 closest to the second black box have low probabilities in the TMHMM prediction (Figure 3.6B). They are, however, included in the TM2 region because they are flanked on their C-terminal side by a position which is occupied in 63% of cases by the positively charged Lys or Arg which can be found “snorkelling” at the membrane-water interface (Kim et al., 2011).

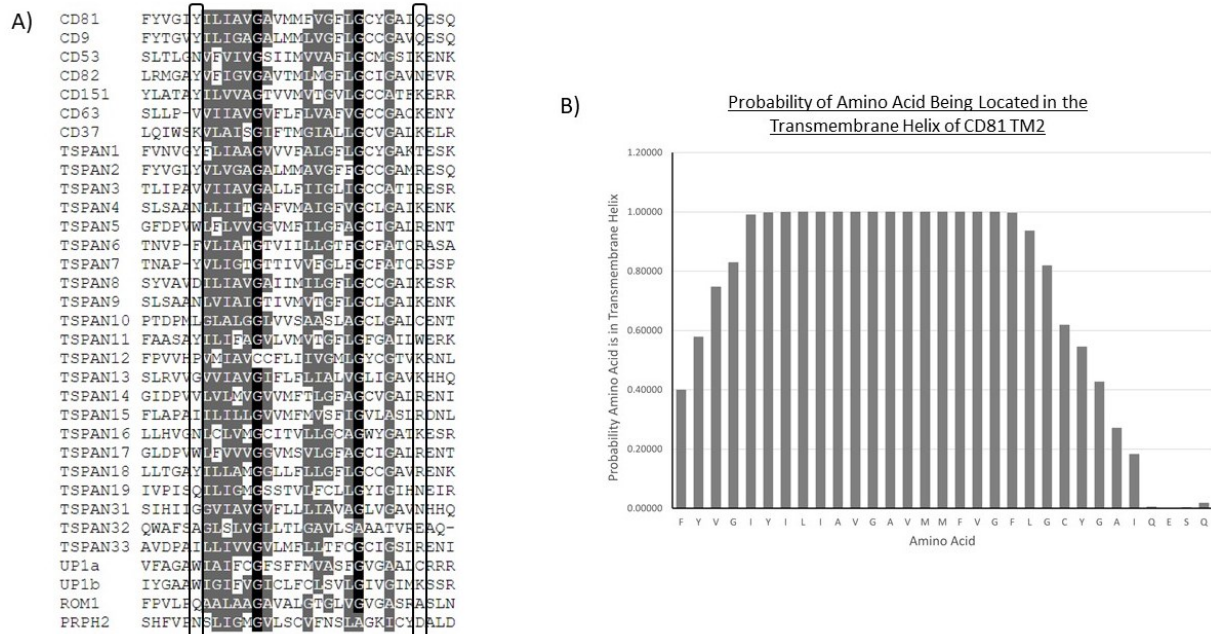


Figure 3.6. Determination of the TM2 region of human tetraspanins. (A) Sequence alignment of the regions containing TM2 of all known human tetraspanins aligned using Clustal Omega (Sievers et al., 2011) with potential “snorkelling” residues at the membrane-water interface and residues that form part of an “aromatic belt” in the interfacial region highlighted in the black boxes. Hydrophobic residues with >80% conservation are highlighted in black for a single amino acid at that position and in grey if the residue is one of Gly, Ala, Val, Leu or Ile. (B) Prediction of the probability that an amino acid in CD81 is likely to form part of the TM2 transmembrane helix using TMHMM (Moller et al., 2001).

Like TM1, TM2 has a high degree of conservation with 9 of the 21 positions having >50% sequence conservation of a specific amino acid (Figure 3.7). Leu in the 2nd position (55% sequence conservation), Ile in the 3rd position (64%), Gly in the 6th position (97%), Gly in the 13th position (61%), Phe in the 14th position (52%), Gly in the 16th position (94%), Cys in the 17th position (64%), Gly in the 19th position (70%) and Ala in the 20th position (58%) all have >50% sequence conservation.

CD81	ILIAVGVMMFVGFVGGYGA
CD9	ILIGAGALMMLVGFVGGGAV
CD53	VIVVGSIIIMVVAFLGCMGSI
CD82	VIVGVAVTMMVGFVGGYGA
CD151	ILVVAGTVVMVTGVLGCCATF
CD63	VIIAVGVFLFLVAVFGCCGAC
CD37	VLAISGIFTMGIALLGCVGAL
TSPAN1	ELIAAGVVVFALGFVGGYGA
TSPAN2	VLVAGALMMAVGVVGGGAM
TSPAN3	VIIAVGALLFTIIGLVGGATI
TSPAN4	LLIITGAFVMAIGFVGGYGA
TSPAN5	ELVVVGGVVFILGFVGGYGA
TSPAN6	VLIATGTVIILVGGYGA
TSPAN7	VLIATGTVIILVGGYGA
TSPAN8	ILIAVGAIIIMVVAFLGCMGSI
TSPAN9	LVIAIGTIVMVTGVLGCCGAI
TSPAN10	GLALGGLVVSAAVSLAGCLGAL
TSPAN11	ILIFAGVLMVVTGVLGFGAIL
TSPAN12	VMIAVCCFLIIVGMVGGYGA
TSPAN13	VVIAVGIIFLFLIALVGLIGAV
TSPAN14	LVLVGGVVFVFLGFVGGYGA
TSPAN15	ILILLGVVFMVSVFVGLASL
TSPAN16	LCVVMGCITVLLGCVGGYGA
TSPAN17	LEVVVGGVMSVLFVGGYGA
TSPAN18	ILLAMGGLLFLVGGYGA
TSPAN19	ILIGMGSSTVLFCLLGYIGH
TSPAN31	GVIAGVFLLLIIVAGLVGAV
TSPAN32	GLVGLLTLGAVLSAAATVR
TSPAN33	LLIVVGVLMFLLTFVGGYGA
UP1a	IAIFCGFSFFMVASFVGAAL
UP1b	IGIFVVICLFLSVLVGGYGA
ROM1	AALAAGAVALGTGLVGGYGA
PRPH2	SLVGMVLSVFNLSVGGYGA

Figure 3.7. Sequence alignment and conservation of the TM2 domain in human tetraspanins. The TM2 domains of all known human tetraspanins were determined using TMHMM (Moller et al., 2001) and were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

The 6th residue in TM2, located near the extracellular region, has the highest conservation with 97% (Figure 3.8A). All 33 human tetraspanins have a Gly residue at the 6th position except for Tspan12 which has a Cys residue. The Gly at the 6th position forms the start of the GGA7 motif found in TM2 of human tetraspanins with the less well conserved Gly residue at the 13th position and the Ala at the 20th position (Kovalenko et al., 2005). The superimposed structures of the TM2 region of CD81, CD9 and CD53 have an average RMSD of 0.93 Å (Figure 3.8B).

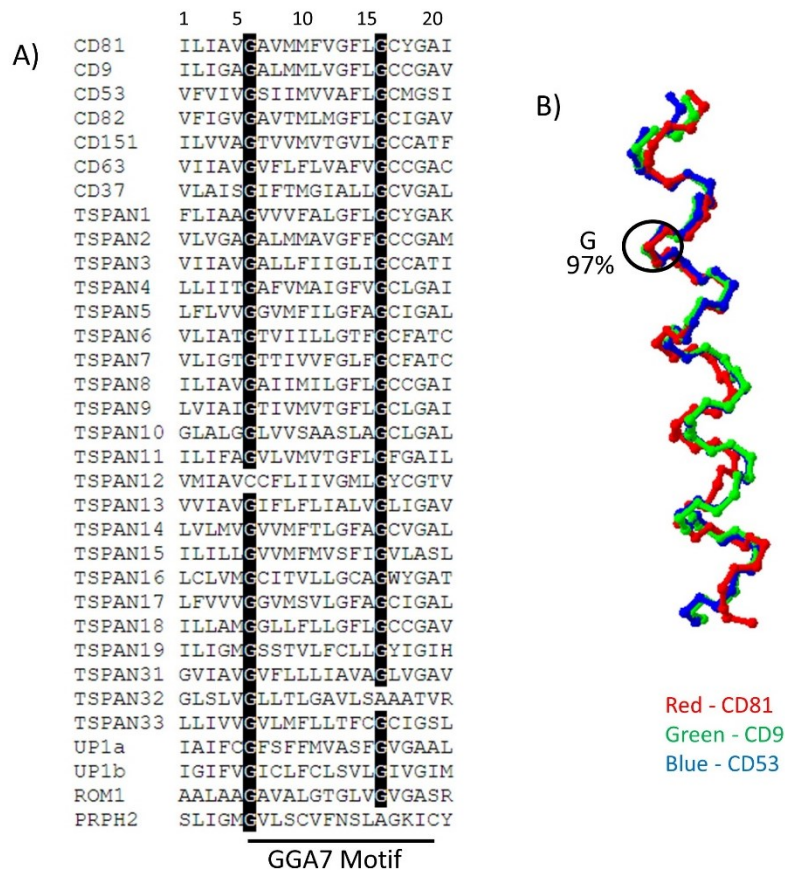


Figure 3.8. Sequence and structural alignment of the TM2 domain of human tetraspanins. (A) The sequence alignment of TM2 showing only residues with >80% sequence conservation for one amino acid highlighted in black. This shows the two conserved Gly residues in the 6th position (97%) and the 16th position (94%). The GGA7 motif of which the 97% conserved Gly in the 6th position is a constituent is labelled underneath the sequence alignment. (B) The 3D structures of the TM2 domains of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 0.93 Å between the three structures was calculated. The proposed anchor residue, Gly (97% conserved) in the 6th position is highlighted with a black circle.

Due to the high level of conservation seen at the 6th residue of TM2 and the similar position that the three residues at this position occupy in the known 3D structures (Figure 3.8B, black circle) it was selected as the anchor residue in TM2 for the universal numbering system.

3.3 Sequence and structural alignment of the TM3 domains of human tetraspanins

The human tetraspanin TM3 region is connected to TM2 by a short chain of approximately five amino acids that forms the small intracellular loop. The TM3 was defined using tetraspanin sequence alignments. Residues showing >80% sequence conservation for hydrophobic residues (Gly, Ala, Val, Leu and Ile) usually found in TM domains are highlighted (Fig 3.9A). TM3 has a hydrophobic region starting immediately after the Cys residue in CD81 with Leu-Leu-Gly but the TMHMM (Moller et al., 2001) analysis only predicts with relatively high confidence that the second Leu residue is the start of the TM helix (Figure 3.9B).

According to the TMHMM analysis, TM3 then extends to Ile-Trp-Gly. The possibility of TM3 going further is illustrated by the Gly-Phe-Val region of CD81 (Figure 3.9A), however the Val residue only has a 0.20910 probability of being in the TM helix (Figure 3.9B). Given that Phe, like Trp, is an aromatic residue that can exist at the membrane-water interface, Ile-Trp-Gly was defined as the end of the TM3 helix.

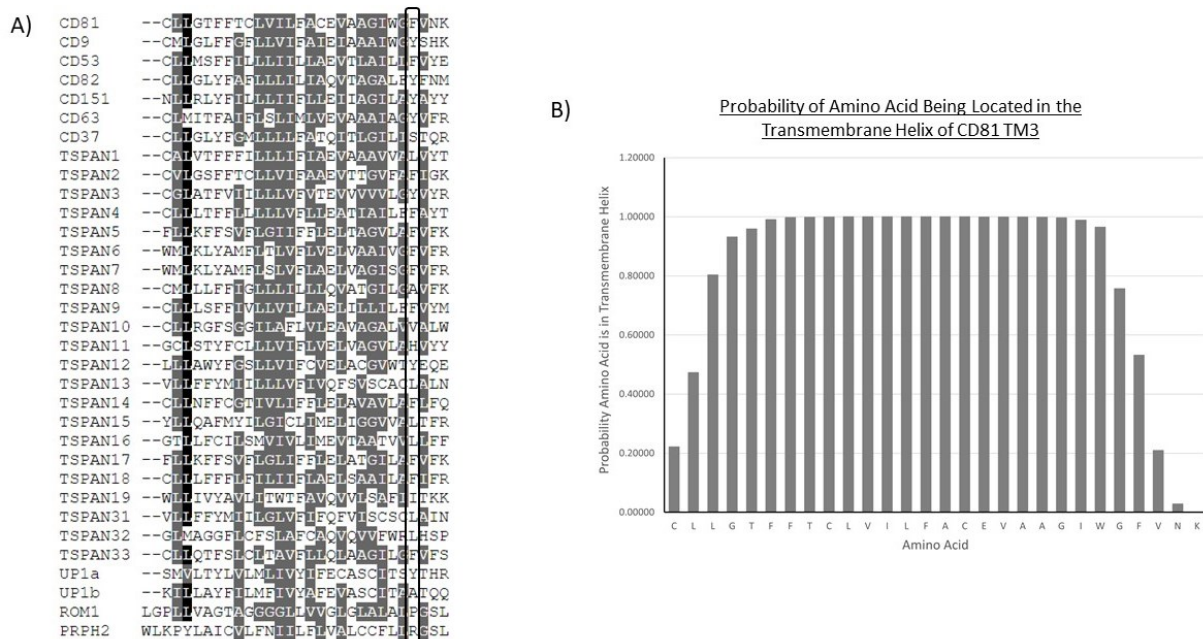


Figure 3.9. Determination of the TM3 region of human tetraspanins. (A) Sequence alignment of the regions containing TM3 of all known human tetraspanins aligned using Clustal Omega (Sievers et al., 2011) with residues that form part of an “aromatic belt” in the interfacial region highlighted in the black box. Residues with >80% conservation are highlighted in black if there is a single amino acid at that position and in grey if the hydrophobic residue is Gly, Ala, Val, Leu or Ile. (B) Prediction of the probability that an amino acid in CD81 is likely to form part of the TM3 transmembrane helix using TMHMM (Moller et al., 2001).

TM3 exhibits slightly lower levels of conservation compared to TM1 and TM2 (Figure 3.10). There are 7 of its 23 positions with >50% sequence conservation for one amino acid. Leu in the 1st position (64% sequence conservation), Leu in the 2nd position (88%), Phe in the 5th position (52%), Leu in the 9th position (70%), Leu in the 10th position (52%), Phe in the 13th position (64%) and Glu in the 16th position (70%).

```

CD81      LLGTFFTCILVILFACEVVAAGIWG
CD9       MGLFFGFLLVIFALETAAAIWG
CD53      LLMSEFILLIILLAEVTLAILL
CD82      LLGLYFAELLLILIAQVTAGALF
CD151     LLRLYFILLIIFLLEIAGILA
CD63      LMITFAIFLSLIMLVEVAAAIAAG
CD37      LLGLYFGMLLLLFATQITLGIIL
TSPAN1   ALVTFFFILLIFIAEVAADVVA
TSPAN2   VLGSEFTCLLVIFAAEVTTGVFA
TSPAN3   GLATFVILLLVFVTEVVVVVLG
TSPAN4   LLLTFELLLLVFLEATIAILF
TSPAN5   LLKFFSVELGIIFLEITAGVLA
TSPAN6   MLKLYAMFLLVFLVELVAIVG
TSPAN7   MLKLYAMFSLVFLAETVAGISG
TSPAN8   MLLFFFIGLLILLQVATGILG
TSPAN9   LLLSEFIVLLVILLAEIILLILF
TSPAN10  LLRGFSGGILAFVLVLEAVAGALV
TSPAN11  CLSTYFCILLVIFLVELVAGVLA
TSPAN12  LLAWYFGSLLVIFCVELACGVWT
TSPAN13  LLFFYMIILLLVFIVCFVSCAC
TSPAN14  LLNFFCGTIVLIFFLVAVVLA
TSPAN15  LLQAEMYILGICLIMEIIGGVVA
TSPAN16  TLLFCILSMVIVLIMEVTAATVV
TSPAN17  LLKFFSVFLGLIFFLEIATGILA
TSPAN18  LLLFFFLIILIFLAETSAAILA
TSPAN19  LLIVYAVLITWTFVAVQVLSAFI
TSPAN31  LLFFYMIILGLVFIQFVISCSC
TSPAN32  LMAGGFLLCFSLAFCAQVQVFWR
TSPAN33  LLQTFSLCLTAVFLLQLAAGILG
UP1a     MVLTYLVLMILIVYIFECASCITS
UP1b     ILLAYFILMFIVYAFEVASCITA
ROM1     LLVAGTAGGGLLLVVGLGLALAL
PRPH2    PYLAICVLENIILFLVALCCFLL

```

Figure 3.10. Sequence alignment and conservation of the TM3 domain in human tetraspanins. The TM3 domains of all known human tetraspanins were determined using TMHMM (Moller et al., 2001) and were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

The standout residue in TM3 is the 88% conserved Leu residue in the 2nd position. When deciding on potential anchor residues it is better to look for residues located nearer to the centre of the TM region. Leu in the 9th position and Glu in the 16th position are the next highest with 70% conservation and both are located nearer the centre of the TM region. To distinguish between the two and find an anchor residue, one method that can be used is to look at the sequence conservation when grouping Glu and Gln together due to their structural similarity (despite Glu being negatively charged in a medium to high pH environment). The similarity between the two residues is exemplified by the positive score of 2 given to this substitution in the BLOSUM62 score matrix (Henikoff & Henikoff, 1992). Another method that can be used to help determine an anchor residue is looking at the BLOSUM62 scores.

The highest BLOSUM62 score in TM3, unsurprisingly, is in the 2nd position with a score of 203.26. There next three highest scoring positions are the 9th, 12th and 16th positions (Figure

3.11). Of these three the 9th position has the highest BLOSUM62 with 179.39, while the other candidate anchor residue, the 16th position, has the lowest score of the three with 172.48.

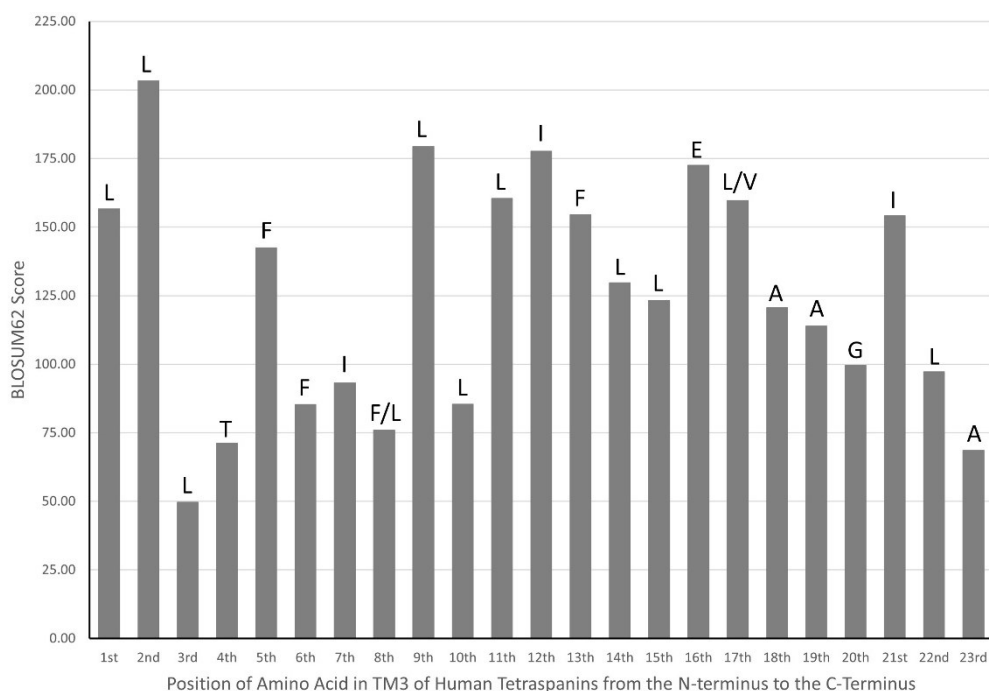


Figure 3.11. BLOSUM62 scores for each amino acid position in TM3 of human tetraspanins. The aligned TM3 domains were analysed in JalView (Waterhouse et al., 2009) to obtain BLOSUM62 scores. BLOSUM62 scores were plotted against each position located in TM3 with the residues located at the N-terminal end of TM1 and moving right to residues located at the C-terminal end of TM3. The most conserved residue in each position is written above the corresponding bar on the graph.

Analysing the sequence alignment by setting it so that all amino acids are treated separately except for Glu and Gln which are grouped together shows 94% conservation of Glu/Gln at the 16th position (Figure 3.12A). The Leu residue in the 9th position can be grouped with Ile because they are similar and, like Glu and Gln, have a positive score of 2 in the BLOSUM62 score matrix. Leu/Ile have 82% conservation in the 9th position.

Due to the high conservation of Glu/Gln of 94% when grouped together and the recognisable feature of a charged residue in the transmembrane region, the 16th position makes a good candidate for an anchor residue. The locally aligned TM3 structures of CD81, CD9 and CD53 have an average RMSD of 0.77 Å and the Glu residue found at this position in all three proteins is in a structurally similar place (Figure 3.12B, black circle).

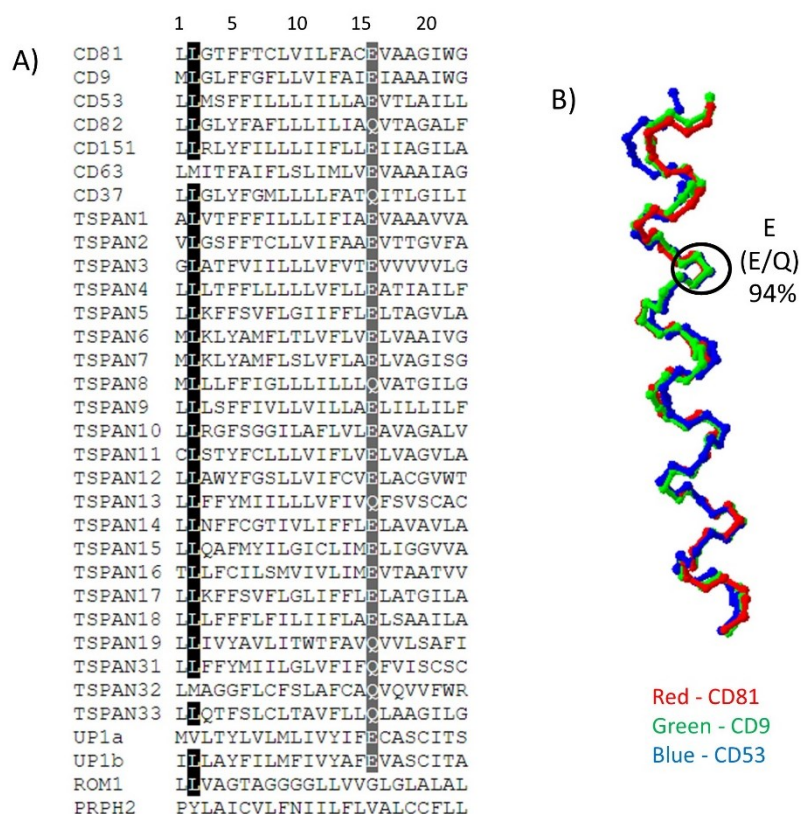


Figure 3.12. Sequence and structural alignment of the TM3 domain of human tetraspanins. (A) The sequence alignment of TM3 showing residues with >80% sequence conservation for one amino acid highlighted in black at Leu in the 2nd position (88% conservation). Residues with >80% sequence conservation of Glu/Gln are highlighted in grey at Glu/Gln in the 16th position (94% conservation). (B) The 3D structures of the TM3 domains of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 0.77 Å between the three structures was calculated. As the anchor residue, Glu/Gln (94% conserved) in the 16th position is highlighted in the black circle.

3.4 Sequence and structural alignments of the TM4 domains of human tetraspanins

The human tetraspanin TM4 region immediately follow the ϵ helix in the LEL. To locate the TM4 region the sequences were aligned and positions with >80% sequence conservation for hydrophobic residues (Gly, Ala, Val, Leu and Ile) usually found in TM regions were highlighted (Figure 3.13A). Sequence conservation of hydrophobic residues in TM4 is not as high in the other three TM regions with just eight positions within the region that contains TM4 showing >80% conservation. This region begins at Ile-Gly-Ile in CD81 and these three positions are well conserved for hydrophobicity. TMHMM (Moller et al., 2001) analysis indicates that these three residues are situated in the TM domain but the preceding Leu residue in CD81 has a >0.6 probability of being in the TM region (Figure 3.13B). When also considering the CD81 structure (PDB: 5tcx) the helical structure is likely to begin at either the preceding Tyr residue or the Leu before the Tyr. These residues may form part of the same helical structure, but they may not be in the TM domain, however, the Leu-Tyr-Leu of CD81 was taken to be the start of

TM4. The end of the TM4 domain extends to the Met-Val-Leu region in CD81 when analysing the sequence with TMHMM. All three of these residues have a probability >0.7. It is possible that one or both Cys residues that follow the Met-Val-Leu region are in TM because some of the other tetraspanins will have TM4 regions containing one or two Cys residues at its C-terminal end, such as CD151, CD63 and Tspan11 (Figure 3.13A). Nevertheless, the Met-Val-Leu region was taken to be the end of TM4.

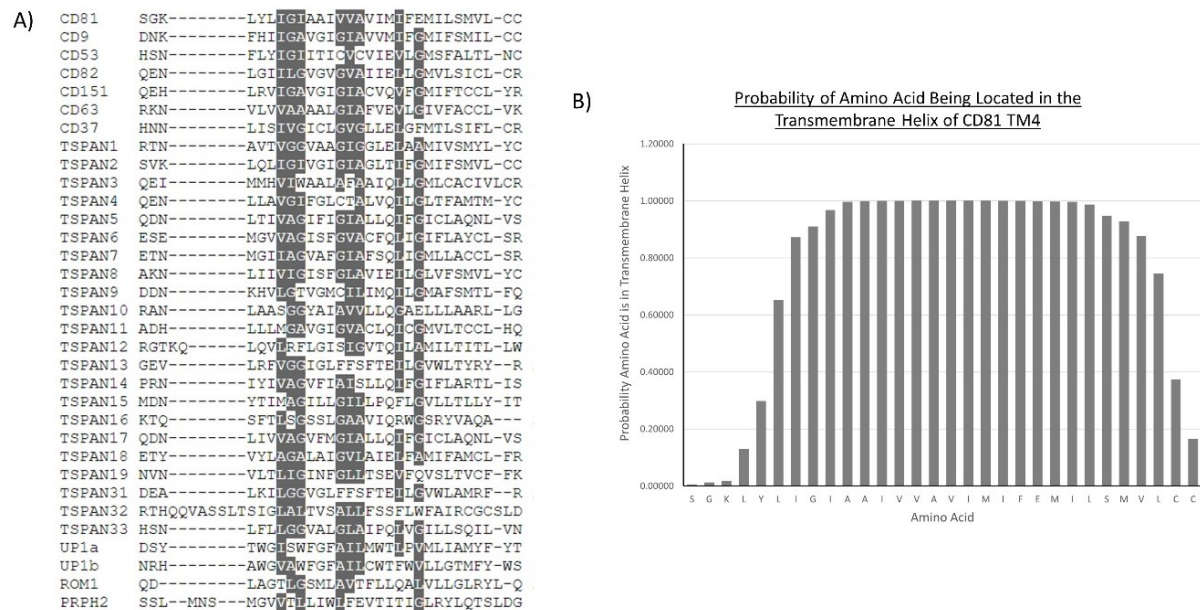


Figure 3.13. Determination of the TM4 region of human tetraspanins. (A) Sequence alignment of the regions containing TM4 of all known human tetraspanins aligned using Clustal Omega (Sievers et al., 2011). Residues with >80% conservation are highlighted in black if there is a single amino acid at that position and in grey if the hydrophobic residue is Gly, Ala, Val, Leu or Ile. (B) Prediction of the probability that an amino acid in CD81 is likely to form part of the TM4 transmembrane helix using TMHMM (Moller et al., 2001).

TM4 in human tetraspanins is the most conserved of all four TMs with just 4 out of the 23 positions having >50% sequence conservation (Figure 3.14), Gly in the 6th position (52% sequence conservation), Gly in the 10th position (55%), Gly in the 18th position (67%) and Leu in the 21st position (52%).

Analysis of the sequence alignments was performed by looking for >80% sequence conservation with similar amino acids grouped together using the default setting on the Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000). Hydrophobic residues (Gly, Ala, Val, Leu and Ile) are one group, aromatic residues (Phe, Tyr and Trp) are a second group, sulfur containing residues (Cys and Met) are a third group, positively charged residues (Lys, Arg and His) are a fourth group, negatively charged and structurally similar

residues (Asp, Glu, Asn, Gln) are a fifth group, polar, uncharged residues (Ser and Thr) are a sixth group and Pro is on its own.

```

CD81  LYLIGIAAIVVAVIMIFEMILSMVL
CD9   FHIIGAVGIGIAVVMIFGMIFSMIL
CD53  FLYIGIITICVCVIEVLGMSFALTL
CD82  LGIILGVGVGVAIIIEELGMVLSICL
CD151 LRVIGAVGIGIACVQVFGMIFTCCCL
CD63  VLVVAAAALGIAFVEVLGIVFACCL
CD37  LISIVGICLGVGLLELGFMTLSIFL
TSPAN1 AVTVGGVAAGIGGLELAAIVSMYL
TSPAN2 LQLIGIVGIGIAGLTIFGMIFSMVL
TSPAN3 MMHVIAAALAFAAIQLLGMLCACIV
TSPAN4 LLAVGIFGLCTALVQIILGLTFAMTM
TSPAN5 LTIIVAGIFIGIALLQIFGICLAQNL
TSPAN6 MGVVAGISFGVACFQLIGIFLAYCL
TSPAN7 MGIILAGVAFGIAFSQLIGMLLACCL
TSPAN8 LIIIVIGISFGLAVIEILGLVFSMVL
TSPAN9 KHVLGTVGMCIIIMQIILGMAFSMTL
TSPAN10 LAASGGYATAVVLLQGAELLLAARL
TSPAN11 LLLMGAVGIGVACLQICGMVLTCCCL
TSPAN12 LQVIRFLGISIGVTQIILAMILTITL
TSPAN13 LRFVGGIGLFFSFTEILGVWLTYRY
TSPAN14 IYIVAGVFIATSLIQIFGIFLARTL
TSPAN15 YTIMAGILLGILLPQFLGVLLTLLY
TSPAN16 SFTLISGSSSLGAAVIQRWGSRYVAQA
TSPAN17 LIVVAGVFMGIALLOIFGICLAQNL
TSPAN18 VYLAGALAIQVLAIELEFAMIFAMCL
TSPAN19 VLTLLIGINFGLLTSEVQVSLTVCF
TSPAN31 LKILGGVGLFFSFTEILGVWLAMRF
TSPAN32 SIGLALTVSALLFSSFLWF AIRCGC
TSPAN33 LFLLGVALGLAIPQLVGIILLSQIL
UP1a  TWGHSWFGFALLMWTLPVMLIAMYF
UP1b  AWGVAVFGFALLCWTFFVLLGTMFY
ROM1  LAGTLGSMLAVTFLLQALVLLGLRY
PRPH2  MGVVTLIWLFEVTITIGLRYLQTS

```

Figure 3.14. Sequence alignment and conservation of the TM4 domain in human tetraspanins. The TM4 domains of all known human tetraspanins were determined using TMHMM (Moller et al., 2001) were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved for a single amino acids (highlighted in black) or similar amino acids (highlighted in grey). Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

Even doing this does not provide any obvious candidates for an anchor residue in TM4. There are eight positions that have >80% sequence conservation when amino acids such as Gly, Ala, Val, Leu and Ile are grouped but to reach that threshold at least three of those amino acids are needed in each of the eight positions (Figure 3.14). This does not lead to an obvious anchor residue.

Lowering the threshold required for sequence conservation by 10% to >70% results in three more positions (Figure 3.15), as well as the eight seen previously when the threshold was >80%. Two of them in the 3rd and 7th position consist largely of the hydrophobic residues found in the large numbers in the eight positions previously described. The third new position

highlighted is the 15th position which is 75% Gln/Glu (Figure 3.15, black box). The Gln residue is 45% conserved and Glu is 30% conserved.

```

CD81  LYLIGIAAIVVAVIMIFEMILSMVL
CD9   FHIIGAVGIGIAVVMIFGMIFSMIL
CD53  FLYIGIITICVCVIEVLGMSFALTL
CD82  LGIILGVGVGVATIEELGMVLSICL
CD151 LRVIGAVGIGIACVQVFGMIFTCCL
CD63  VLVVAAAALGIAFVEVLGIVFACCL
CD37  LISIVGICLVGLIEELGFM TLSIFL
TSPAN1 AVTVGGVAAGIGGIEEAAMIVSMYL
TSPAN2 LQLIGIVGIGIAGITIFGMIFSMVL
TSPAN3 MMHVITWAALAFATQLLGM LCACIV
TSPAN4 LLAVGIFGLCTALVQILGLTFAMTM
TSPAN5 LTIVAGIFIGIALIQIFGICLAQNL
TSPAN6 MGVVAGISFGVACHQLIGIFLAYCL
TSPAN7 MGI IAGVAFGIAFSQLIGM LLACCL
TSPAN8 LIIVIGISFGLAVIEILGLVFSMVL
TSPAN9 KHVLCITVMCILINQILGMAFSMTL
TSPAN10 LAASGGYATAVLLQSAELL LAARL
TSPAN11 LLLMGAVGIGVACIQICGMVLTCCCL
TSPAN12 LQVLRFLGISIGVTQILAMILTITL
TSPAN13 LRFVGGIGLFFSFTEILGVWLTTRY
TSPAN14 IYIVAGVFIAISLQIFGIFLARTL
TSPAN15 YTIMAGILLGILLEQFLGVLLTLLY
TSPAN16 SFTLSGSSLGAAVIRWGSRYVAQA
TSPAN17 LIIVAGVFMGIALIQIFGICLAQNL
TSPAN18 VYLAGALAGVLAIEELFAMIFAMCL
TSPAN19 VLTLLIGINFGLTSEVFQVSLTVCF
TSPAN31 LKILGGVGLFFSFTEILGVWLAMRF
TSPAN32 SIGLALTVSALLFSSFLWF A IRCGC
TSPAN33 LFLGGVALGLAIEQVVGILLSQIL
UP1a  TWGISWFGF AILMWTLPVMLIAMYF
UP1b  AWGVAVWFGF AILCWTFWVLLGTMFY
ROM1  LAGTLLGSMLAVTFILQALVLLGLRY
PRPH2 MGVVLLIWLFEVILITIGLRYLQTS

```

Figure 3.15. Sequence alignment and residues with >70% conservation of the TM4 domain in human tetraspanins. The TM4 domains of all known human tetraspanins were determined using TMHMM (Moller et al., 2001) were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >70% conserved for a single amino acid (highlighted in black) or similar amino acids (highlighted in grey). Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro. The 75% conserved Glu/Gln residue in the 15th position is highlighted in the black box.

Its conservation of 75% is higher than the 67% for the highest single amino acid conservation of TM4 which is Gly in the 18th position (Figure 3.16A) making it a good candidate as TM4 anchor residue. The locally aligned TM4 regions of CD81, CD9 and CD53 have an average RMSD of 0.59 Å indicating structural similarity. The Gln/Glu position, occupied by Met and Glu in CD81, CD9 and CD53, are in a near identical position structurally in TM4 (Figure 3.16B, black circle).

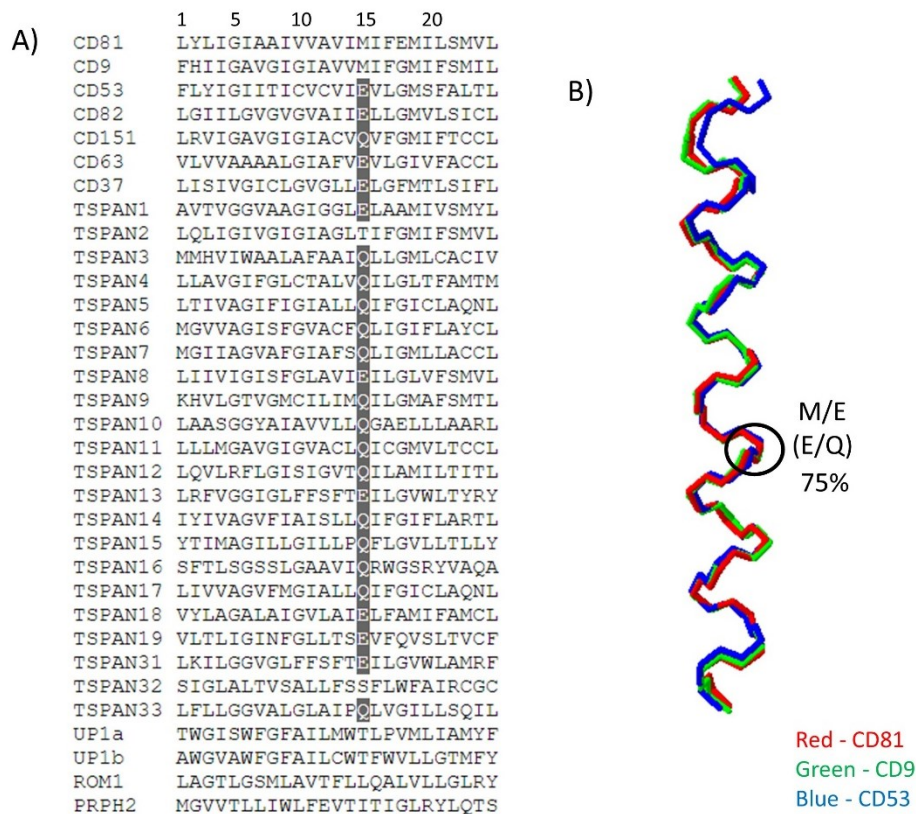


Figure 3.16. Sequence and structural alignment of the TM4 domain of human tetraspanins. (A) The sequence alignment of TM4 showing residue with >70% sequence conservation of Glu/Gln highlighted in grey. The 75% conserved Glu/Gln residue in the 15th position is highlighted in grey. (B) The 3D structures of the TM4 domains of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 0.59 Å between the three structures was calculated. As the anchor residue, Glu/Gln (75% conserved) in the 15th position is highlighted in the black circle.

3.5 Anchor residue in the 4 transmembrane regions of tetraspanins

Figure 3.17 shows the location of all four TM anchor residues within the closed conformation of CD81. The closed CD81 structure was chosen to display the anchor residues because it was the first high resolution tetraspanin structure to be solved. The 82% conserved Gly residue in TM1 is Gly25 of CD81 (Figure 3.17B), the 97% conserved Gly residue in TM2 is Gly69 (Figure 3.167C) and the 94% conserved Glu/Gln residue in TM3 is Glu105 (Figure 3.167D). CD81 does not have a Glu or Gln residue in the position that contains 75% conservation of Glu/Gln in TM4. Instead, CD81 has a Met residue at this position which can be seen in TM4 of the full-length protein (Figure 3.17E).

In CD81, we propose that Gly25 is labelled Gly1.50; 1 represents TM1 and 50 indicates that Gly25 is the TM anchor residue. Its position within the complete primary sequence of CD81 may also be added, such that the notation is Gly1.50⁽²⁵⁾.

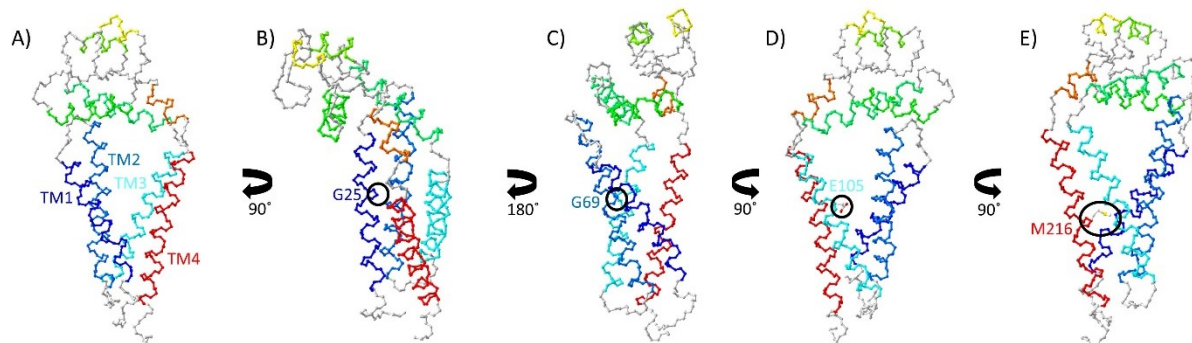


Figure 3.17. Ball and stick representation of CD81 showing the position of the anchor residues in the TMs of human tetraspanins. (A) Ball and stick representation of CD81 in its closed conformation (5tcx) with TM1 (dark blue), TM2 (light blue), TM3 (cyan) and TM4 (red) labelled. (B) Location of the 82% conserved glycine residue (G25 in CD81) in full length CD81 viewed 90° clockwise compared to (A). (C) Location of the 97% conserved glycine residue (G69) in full length CD81 viewed 180° clockwise compared to (B). (D) Location of the 94% conserved glutamic acid/glutamine residue (E105 in CD81) in full length CD81 viewed 90° anti-clockwise compared to (C). (E) Location of the 75% conserved glutamic acid/glutamine residue (M216 in CD81) in full length CD81 viewed 90° anti-clockwise compared to (D).

Anchor residues have been determined for the four TM domains in human tetraspanins. However, in order to compile a comprehensive numbering system for tetraspanins other conserved regions will need to be included. Analysis, therefore, will be carried out on α , β and ϵ helices in the LEL to determine anchor residues in each of these regions.

3.6 Sequence and structural alignments of the α helix of the large extracellular loop of human tetraspanins

The LEL α helix is located on the TM2-TM3 side of the protein and rests parallel to the membrane when it is in the closed conformation (Figure 3.1). The α helix ranges in length from approximately 12 to 20 amino acids (Figure 3.18A). Sequence alignment using Clustal Omega (Sievers et al., 2011) produced no position with >40% conservation of any one single amino acid.

Grouping chemically and structurally similar amino acids together in the following groups results in position five for the majority of tetraspanins having >80% sequence conservation. The groups are: Group 1: Gly, Ala, Val, Leu and Ile, Group 2: Phe, Tyr and Trp, Group 3: Cys and Met, Group 4: Lys, Arg and His, Group 5: Asp, Glu, Asn, Gln, Group 6: Ser and Thr and Group 7: Pro. This position in the sequence is largely hydrophobic with 87% of the amino acids at this position either Ile, Val, Leu or Ala (Figure 3.18B).

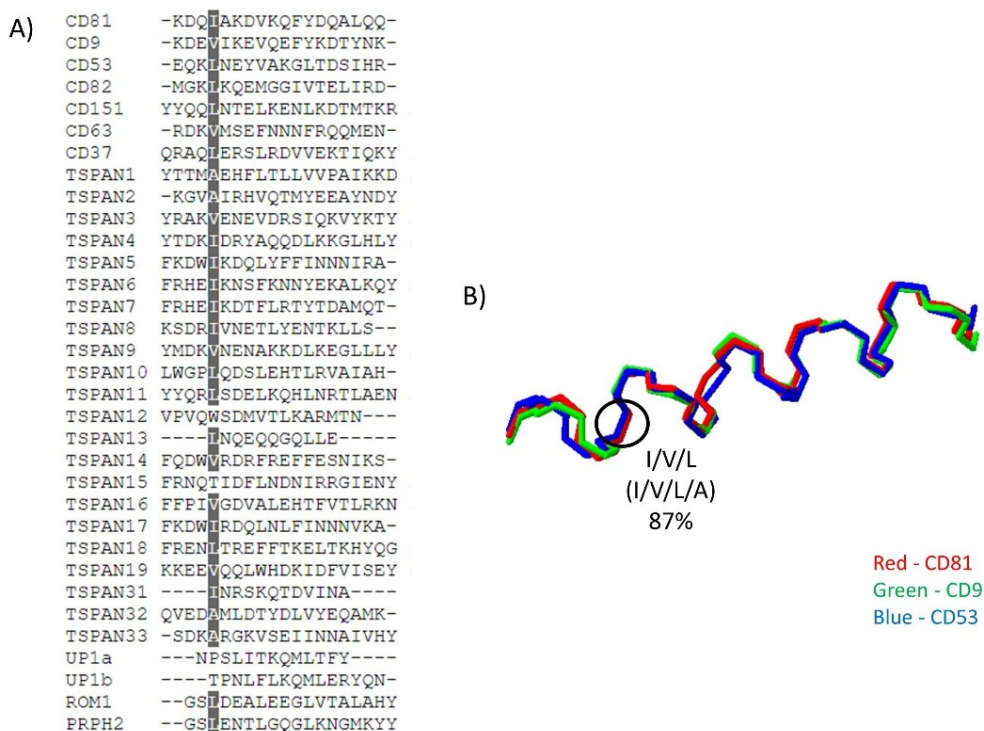


Figure 3.18. Sequence and structural alignment of the α helix in the large extracellular loop of human tetraspanins. (A) The sequence alignment of the α helix within the large extracellular loop. Residues that are >80% conserved for similar amino acids are highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro. The 87% conserved Ile/Val/Leu/Ala residue in the 4th/5th position is highlighted in grey. (B) The 3D structures of the α helix of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 0.57 Å between the three structures was calculated. As the anchor residue, Ile/Val/Leu/Ala (87% conserved) in the 4th/5th position is highlighted in the black circle.

The α helices of the solved structures, CD81, CD9 and CD53 were locally aligned and the average RMSD between the three structures was calculated to be 0.57 Å (Figure 3.18B). The low RMSD between the structures highlights the structural similarity despite a lack of sequence conservation in the α helices of tetraspanins. The 87% conserved hydrophobic residue in the 5th position of the majority of tetraspanins α helices (it is in the 4th position in CD81, CD9 and CD53) is situated in the same position structurally, therefore it is an ideal position for an anchor residue for the universal numbering system (Figure 3.18B, black circle).

3.7 Sequence and structural alignments of the β helix of the large extracellular loop of human tetraspanins

The β helix is located on the TM1-TM4 side of human tetraspanins. It forms part of the large extracellular loop, resting perpendicular to the membrane in the closed conformation and varies in length. The longest belongs to Tspan31 which is 15 residues long, while the shortest

is in Tspan2 which is just 9 residues long (Figure 3.19A). The vast majority, however, are 13 residues long. Only six (CD37, Tspan2, Tspan8, Tspan16, Tspan31 and PRPH2) of the 33 human tetraspanins possess a β helix that is not 13 residues long, therefore any reference hereafter to residue positions in the β helix will be with regards to a structure that is 13 residues long.

Only two of the 13 positions in the β helix have >50% sequence conservation. Asp in the 8th position (61% sequence conservation) and Gln in the 11th position (73%). Neither of these positions has >80% sequence conservation for any one specific amino acid. Gln in the 11th position, however, is situated just two residues away from the C-terminal end of the β helix which means that it would not make an ideal candidate for an anchor residue.

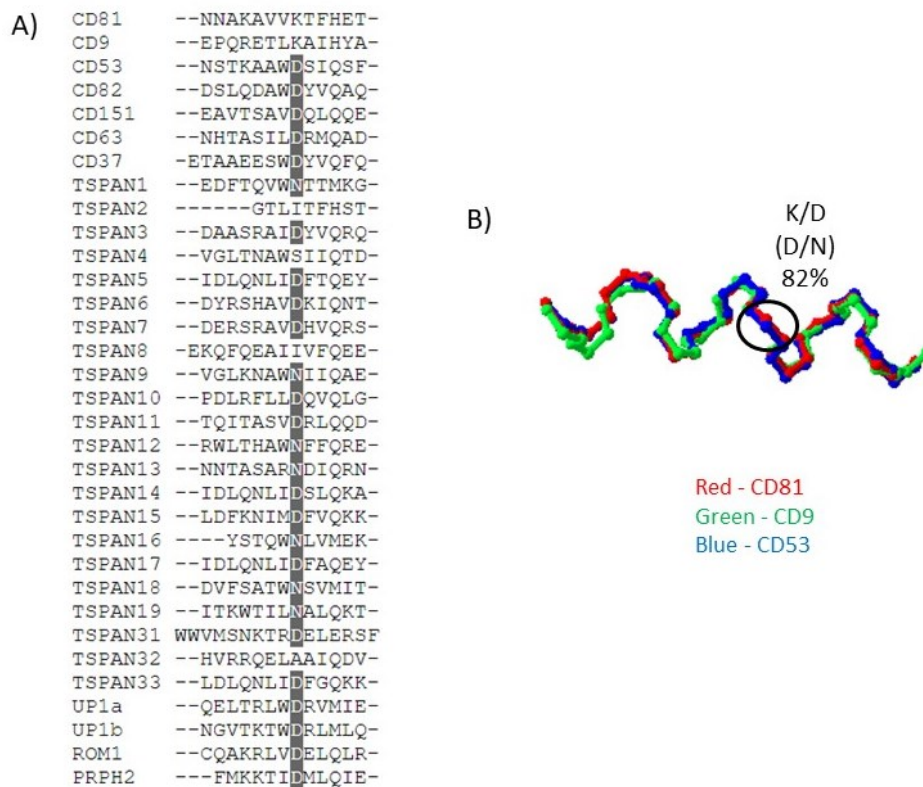


Figure 3.19. Sequence and structural alignment of the β helix in the large extracellular loop of human tetraspanins. (A) The sequence alignment of the β helix within the large extracellular loop. Residues that are >80% conserved for a single amino acid are highlighted in black and those for similar amino acids are highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro. The 82% conserved Asp/Asn residue in the 8th position is highlighted in grey. (B) The 3D structures of the β helix of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 0.55 Å between the three structures was calculated. As the anchor residue, Asp/Asn (82% conserved) in the 8th position is highlighted in the black circle.

Asp in the 8th position, on the other hand, is situated closer to the centre of the β helix and when combined with the structurally similar residue, Asn, there is 82% sequence conservation for Asp/Asn (Figure 3.19A). The positive score of 1 in the BLOSUM62 score matrix highlights the chemical similarity between Asp and Asn with the most significant difference being that Asp is negatively charged (Henikoff & Henikoff, 1992).

The locally aligned β helix structures of CD81, CD9 and CD53 have an average RMSD of 0.55 Å demonstrating the structural similarity of the region despite little primary sequence conservation. The Asp/Asn position, occupied by Lys and Asp in CD81, CD9 and CD53, are in a similar position structurally making it ideal to act as anchor residue in the universal numbering system (Figure 3.19B, black circle).

3.8 Sequence and structural alignments of the ϵ helix of the large extracellular loop of human tetraspanins

The ϵ helix is located on the TM1-TM4 side of tetraspanins in the large extracellular loop immediately before TM4. The longest belongs to Tspan32 which is 20 residues long, while the shortest is in ROM1 which is 11 residues long (Fig. 3.20). All but four (Tspan12, Tspan32, ROM1 and PRPH2) of the 33 human tetraspanins possess an ϵ helix that is 12 residues in length, therefore any reference hereafter to residue positions in the ϵ helix will be with regards to a structure that is 12 residues long.

Only one of the 12 positions in the ϵ helix has >50% sequence conservation, which is the first Cys at its N-terminal end. It has 100% sequence conservation and forms a disulfide bond with the first Cys residue in the CCG motif situated immediately after the β helix. No other position has >50 sequence conservation for a single amino acid (Figure 3.20). The closest to reaching this threshold is Asn in the 12th, and final, position which is 48% conserved.

Neither the Cys residue in the 1st position, nor the Asn residue in the final position would make a good anchor residue because their position at the extreme ends of the ϵ helix would make a numbering system redundant.

```

CD81  CHQKIDDLFSGK-----
CD9   CPDAIKEVFDNK-----
CD53  CYAKARLWFHNS-----
CD82  CMEKVQAWLQEN-----
CD151 CITKLETFIQEH-----
CD63  CVEKIGGWLRRN-----
CD37  CAQGLQKWLHNN-----
TSPAN1 CFNQLLYDIRTN-----
TSPAN2 CIDEIETIISVK-----
TSPAN3 CEALVVKKLQEI-----
TSPAN4 CYETVKVWLQEN-----
TSPAN5 CVPQFEKWLQDN-----
TSPAN6 CFIKVTIIESE-----
TSPAN7 CYDLVTSFMETN-----
TSPAN8 CISFIKDFLAKN-----
TSPAN9 CYEKVKMWFDDN-----
TSPAN10 CGPPLRRWLRAN-----
TSPAN11 CLTKLEQFLADH-----
TSPAN12 CGKKMYSFLRGTKQ-----
TSPAN13 CAPIIGEYAGEV-----
TSPAN14 CIQALESWLPRN-----
TSPAN15 CTNAVIWFMDN-----
TSPAN16 CFHKLLKITKTQ-----
TSPAN17 CVGQFEKWLQDN-----
TSPAN18 CYTVILNTFETY-----
TSPAN19 CENKISAWYNVN-----
TSPAN31 CGEKFLKHSDEA-----
TSPAN32 CLQGIRSFRLRTHQQVASSLT
TSPAN33 CIDKLVNWIHSN-----
UP1a  CFEHIGHAIDSY-----
UP1b  CYELISGPMNRH-----
ROM1  CHEVLLLEHLQD-----
PRPH2 CRAALLSYSSLMNS-----

```

Figure 3.20. Sequence alignment and conservation of the ϵ helix in the large extracellular loop of human tetraspanins. The sequence alignment of the ϵ helix. Residues that are >80% conserved for a single amino acid are highlighted in black. This highlights the 100% conserved Cys residue in the 1st position.

The BLOSUM62 scores for the ϵ helix can help locate an anchor residue. Two of the 12 positions have a BLOSUM62 score >150. The 1st position has a score of 215.65 and the 5th position has a score of 173.04 (Figure 3.21). With the 1st position not offering a viable position for an anchor residue, the 5th position with the second highest BLOSUM62 score is a suitable position. It is located near the centre of the ϵ helix and is a largely hydrophobic position. When grouping the hydrophobic Gly, Ala, Val, Leu and Ile residues together because of their similar chemical properties (and using the default groupings on the Sequence Manipulation Suite highlighted previously) the conservation seen in the 5th position is 84% (Figure 3.22A).

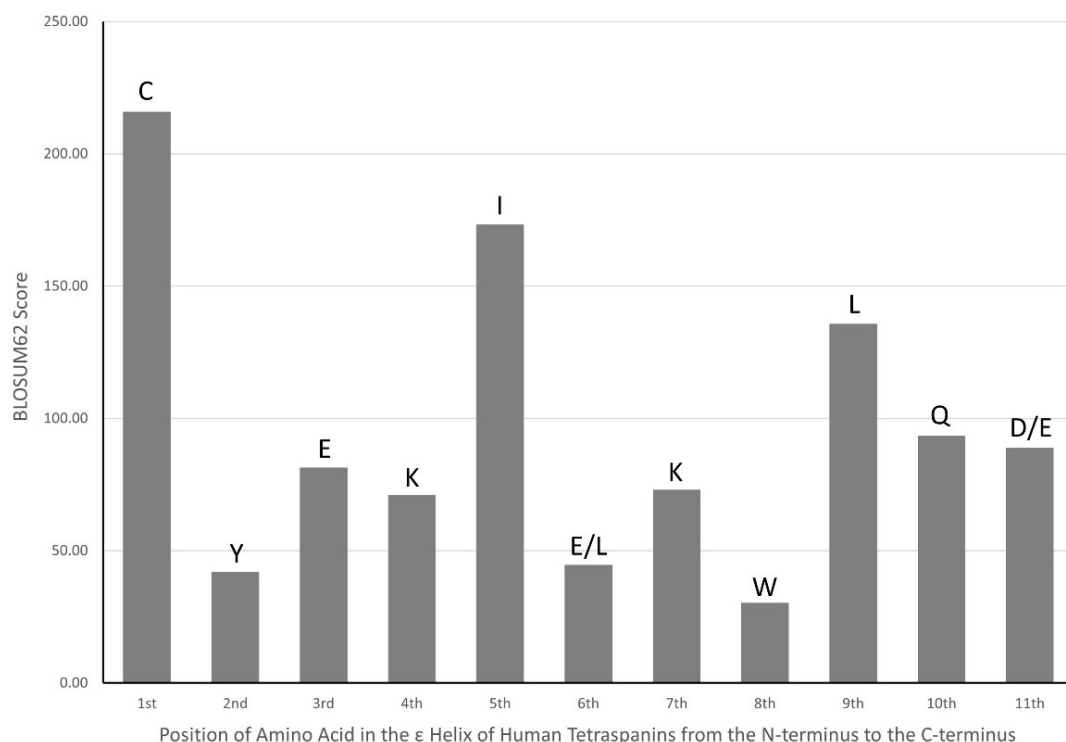


Figure 3.21. BLOSUM62 scores for each amino acid position in the ϵ helix in the large extracellular loop of human tetraspanins. The aligned ϵ helices were analysed in JalView (Waterhouse et al., 2009) to obtain BLOSUM62 scores. BLOSUM62 scores were plotted against each position located in TM3 from the N-terminal end of TM3 to its C-terminus. The most conserved residue in each position is written above the corresponding bar on the graph.

The locally aligned ϵ helix structures of CD81, CD9 and CD53 have an average RMSD of 1.16 Å (Figure 3.22B). This relatively high RMSD can be attributed to the structural difference seen at the C-terminal end of the ϵ helix where it meets TM4. The average RMSD for the final four residues at the C-terminal end is 2.12 Å, whereas the average RMSD for the first eight residues is 0.73Å.

Despite being made up largely of three amino acids (Ile, Leu and Ala) the 5th position with its high BLOSUM62 score and the similarity in structural position in CD81, CD9 and CD53 makes it a good candidate for an anchor residue (Figure 3.22B, black circle).

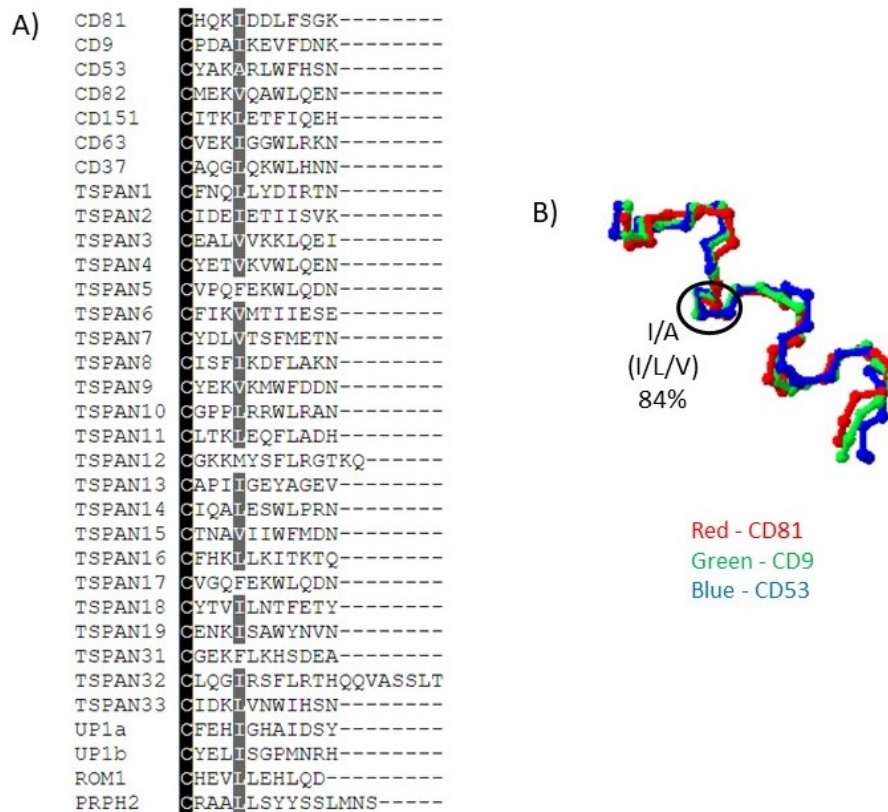


Figure 3.22. Sequence and structural alignment of the ϵ helix in the large extracellular loop of human tetraspanins. (A) The sequence alignment of the ϵ helix. Residues with >80% sequence conservation for a single amino acid are highlighted in black and for similar amino acids are highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro. The 84% conserved Ile/Leu/Val residue in the 5th position is highlighted in grey. (B) The 3D structures of the ϵ helix of the solved human tetraspanin structures CD81 (PDB: 5tcx, in red), CD9 (PDB: 6k4j, in green) and CD53 (PDB: 6wvg, in blue) were locally superimposed in Swiss-PdbViewer and the average RMSD of 1.16 Å between the three structures was calculated. As the anchor residue, Ile/Leu/Val (84% conserved) in the 5th position is highlighted in the black circle.

3.9 A universal tetraspanin numbering system

Anchor residues have been determined for the main conserved structural region of human tetraspanins (Figure 3.23). The conserved structural regions are the four transmembrane domains (TM1, TM2, TM3 and TM4) and the three common regions in the large extracellular loop (the α helix, the β helix and the ϵ helix).

The basic principle is that each anchor residue serves as a reference point. For instance, the anchor residue of TM1 is 1.50. Moving towards the N-terminus results in the number decreasing, while moving towards the C-terminus results in the number increasing. For example, the 22nd residue in TM1 of CD81 is Trp and is three residues closer to the N-terminus than the anchor residue, meaning it is labelled Trp1.47 or Trp1.47⁽²²⁾.

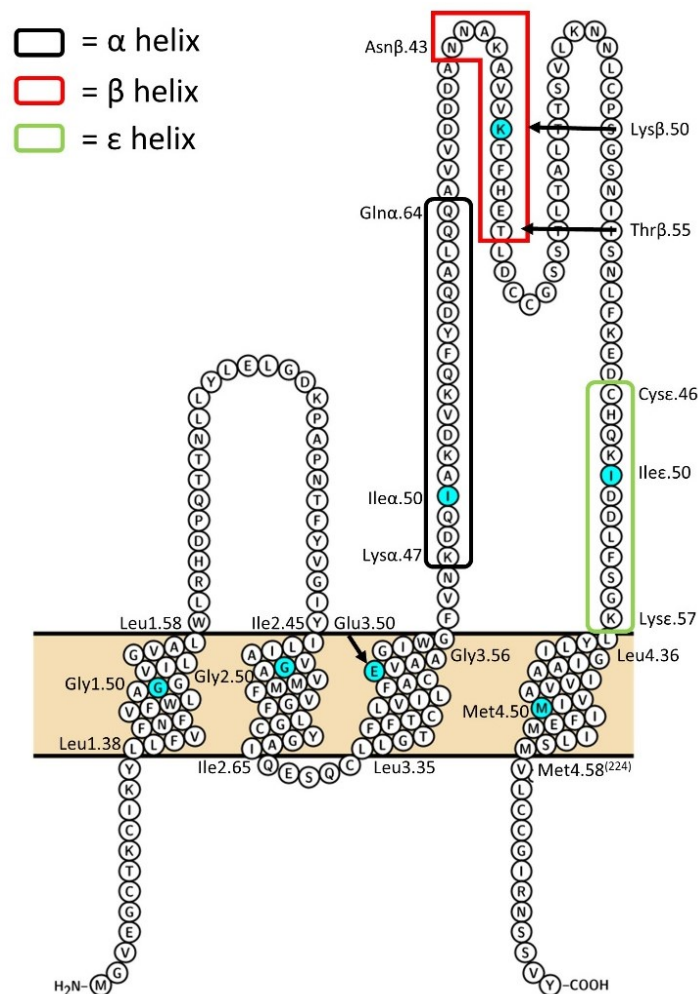


Figure 3.23. A universal tetraspanin residue numbering system. A snake plot of CD81 produced using Protter (Omasits et al., 2014) is used to exemplify the numbering system. In TM1 the reference residue is Gly1.50. Gly refers to the amino acid present in that position, 1 refers to the TM and 50 is the most conserved (anchor) residue. Moving towards the N-terminus, the last number in this notation decreases, so the adjacent upstream residue is Ala1.49 and the first residue in TM1 is Leu1.38. Moving towards the C-terminus, the next residue in CD81 is Gly1.51. The final residue in TM1 is Leu1.58. The anchor residues, as well the first and last residues in TM2, TM3 and TM4 of CD81 are highlighted. It is also possible to denote the position of any residue in the primary sequence, as exemplified by Met4.58⁽²²⁴⁾.

3.10 Conclusion

In conclusion, a universal tetraspanin numbering system has been proposed that categorises conserved regions in human tetraspanins. A reference point, or anchor residue, has been highlighted in each conserved region allowing for a simple way for tetraspanin researchers to communicate the position of a specific amino acid. The variable loop region between the β helix and the ϵ helix in the LEL is less well conserved, both structurally and with regards to primary sequence, therefore the next chapter will look at these variable loop sequences and structures.

Chapter 4 – Characterisation of the loops in the LELs of tetraspanins

Defining a comprehensive universal numbering system for tetraspanins proved challenging. Numbering the membrane domains was feasible (Chapter 3) but for the loops situated in the LEL between the β helix and the ϵ helix (Figure 4.1), and the SEL, robust numbering was less straightforward, especially as these regions have significantly different lengths within the family. To analyse these loop regions, they were divided into sub-categories based on the arrangement of the Cys residues within them. Some of the LELs from these subfamilies were modelled using trRosetta (Yang et al., 2020; Du et al., 2021).

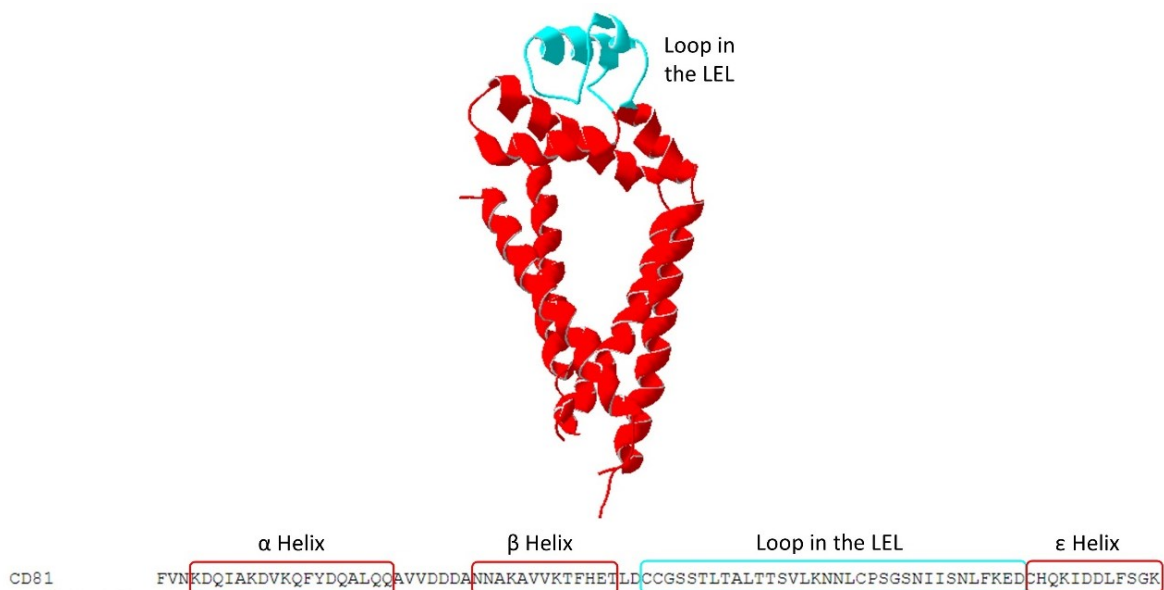


Figure 4.1. Location of the loop in the LEL of closed CD81. The position of the loop (light blue) in the LEL of closed CD81 (PDB: 5tcx) between the β helix and the ϵ helix. Underneath the structure is the CD81 LEL sequence with the LEL helices and the loop highlighted.

4.1 Sequence Alignment of the loops in the LELs of tetraspanins

The sequences of the loops in the LEL, which are the regions between the β helix and the ϵ helix, were taken from all known 33 human tetraspanins and aligned using Clustal Omega (Sievers et al., 2011). The start of the sequence was taken as the first residue of the CCG motif which is ubiquitous across all tetraspanins. The alignments were then analysed in the Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) with residues grouped together using the default setting because of their chemical and structural similarities so that sequence conservation could be analysed. Amino acids were grouped together in the following groups: Group 1: Gly, Ala, Val, Leu and Ile, Group 2: Phe, Tyr and Trp, Group 3: Cys

and Met, Group 4: Lys, Arg and His, Group 5: Asp, Glu, Asn and Gln, Group 6: Ser and Thr and Group 7: Pro. Using the default amino acid groups, any part of the sequence that had >80% conservation but <100% conservation is shown in light grey, while any part of the sequence with 100% conservation is shown in black.

There is little conservation in the loop regions of tetraspanins with only one residue that is 100% conserved, namely a Cys residue in the middle of the loop in the LEL (Figure 4.1). The only other notable position in loop in the LEL is the (largely) Trp residue found seven residues from the start of the loop in most tetraspanins. This position is 73% Trp but rises to >80% conservation when grouped together with other amino acids that have an aromatic ring, such as Phe and Tyr. When including these residues every tetraspanin that has a residue at that position on the alignment has one of those three residues except for Tspan18 which has a Val residue (Figure 4.2).

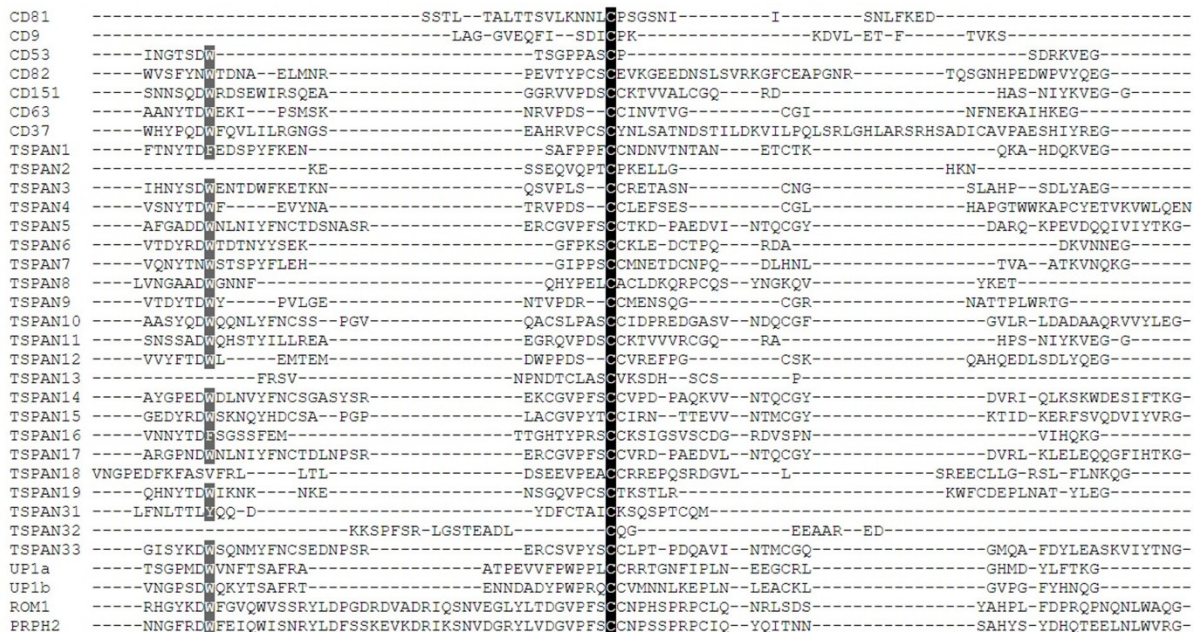


Figure 4.2. Sequence alignment and conservation of the loops between the β helix and the ϵ helix in the LELs of human tetraspanins. The loops in the LELs of all human tetraspanins were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that were >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

Despite the ideal location of a 100% conserved Cys residue in the middle of the loop in the LEL in every human tetraspanin, it would be challenging to devise a numbering system for the loops in the LEL because of the large differences in sequence length within the family, which ranges from 20 residues in Tspan2 to 82 residues in PRPH2 (Figure 4.2).

4.2 Cysteine arrangement in the LEL of human tetraspanins

In order to look at LEL sequence conservation, the constituent loops were divided into subfamilies that had previously been suggested in the literature (Huang et al., 2005; De Salle et al., 2010). These subfamilies are defined by the number and arrangement of Cys residues in the LEL. All known human tetraspanin loop sequences were divided into these subfamilies and the sequences were aligned using Clustal Omega (Sievers et al., 2011). The subfamilies are named as follows: TspanC4 (CD81, CD9, CD53, Tspan2, Tspan32), TspanC6-CC (CD151, CD63, Tspan1, Tspan3, Tspan4, Tspan6, Tspan7, Tspan9, Tspan11, Tspan12, Tspan16, Tspan18, Tspan19, UP1a, UP1b), TspanC6-CxC (CD82, CD37, Tspan8, Tspan19), TspanC6-CxxxC (Tspan13, Tspan31), TspanC6 β (ROM1, PRPH2) and TspanC8 (Tspan5, Tspan10, Tspan14, Tspan15, Tspan17, Tspan33) (Figure 4.3).

In Huang et al. (2005) and De Salle et al. (2010) each subfamily was given a generic name, such as 6a for the TspanC6-CC family. Names like TspanC4 and TspanC8 are in common usage, therefore the same nomenclature was used to name the tetraspanins that have six or seven Cys residues in the LEL. These families have been named to specifically indicate the arrangement of the middle two Cys residues in the LEL. TspanC6-CCs, for instance, contain six Cys residues in the LEL, as indicated by “C6”, and the two middle Cys residues are consecutive in the sequence. An “x” in between the “CC” is representative of any amino acid. TspanC6 β s have the same Cys arrangement as TspanC6-CCs except for the distinguishing feature of an extra Cys residue in the β helix, giving rise to its name. The new names for tetraspanin subfamilies proposed in this thesis are: TspanC6-CC, TspanC6-CxC, TspanC6-CxxxC and TspanC6 β .

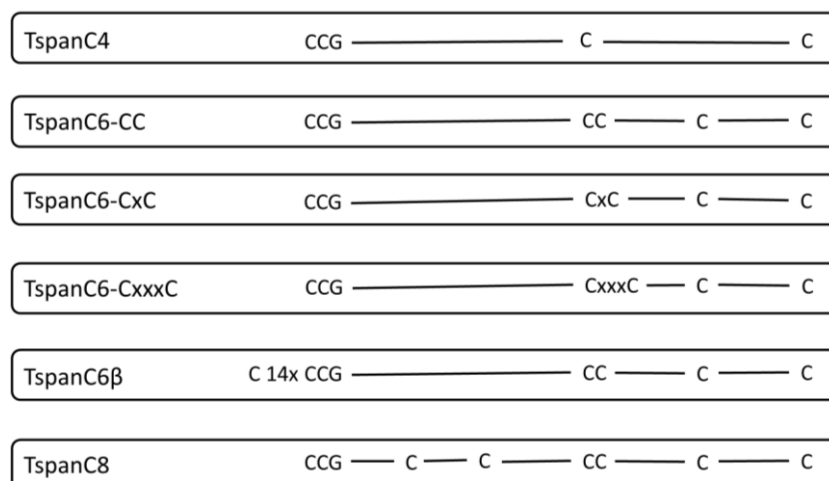


Figure 4.3. The arrangement of cysteines in the LELs of human tetraspanins. The Cys patterns found in the LELs of human tetraspanins with “C4”, “C6” and “C8” representing the number of Cys residues in the loops of the LELs. In the “C6” families this is followed by the arrangement of the two Cys residues situated near the middle of the loops. An “x” between the “CC” represents any amino acids. The “β” in TspanC6β indicates that another Cys residue is found those LELs, 15 amino acids before the CCG-motif in the β helix.

4.3 Disulfide bond formation in the LEL of CD81

With the arrangement of Cys residues in the loop being a crucial factor in the structural classification of tetraspanins, the likelihood of disulfide bond formation was analysed using SSBondPre (Gao et al., 2020). The primary function of SSBondPre is to find residues that can be mutated to Cys to introduce new disulfide bonds into a protein, giving a probability from 0-1 on disulfide bond formation. The output highlights native Cys residues that are likely to form disulfide bonds with an energy score of 0 since no mutations need to be done to the protein.

To ensure that SSBondPre recognises and predicts disulfide bonds that are known to form in tetraspanins, the solved open and closed structures of CD81 were uploaded to the SSBondPre web server for analysis. CD81 has four Cys residues in its loop and is known to form two disulfide bonds with one bond between the first Cys residue in the CCG motif and the final Cys residue in the loop located at the start of the ε helix. The second disulfide bond exists between the second Cys residue in the CCG motif and a Cys residue situated near the centre of the loop. SSBondPre was able to detect the two disulfide bonds that form in the loop of the closed structure of CD81 with a probability of 0.995 for Cys156-Cys190 bond and 0.985 for the Cys157-Cys175 bond (Figure 4.4B). In the solved closed CD81 structure Cys156 and Cys190 are 2.02 Å apart and form a disulfide bond, however, Cys157 and Cys175 are 4.85 Å apart which is too far to form a disulfide bond (Figure 4.4A). Despite the lack of disulfide bond between Cys157 and Cys175 in the structure, SSBondPre predicted that a disulfide bond

between these two residues is highly likely, giving a high degree of confidence in SSBondPre to be able to predict the presence of disulfide bonds.

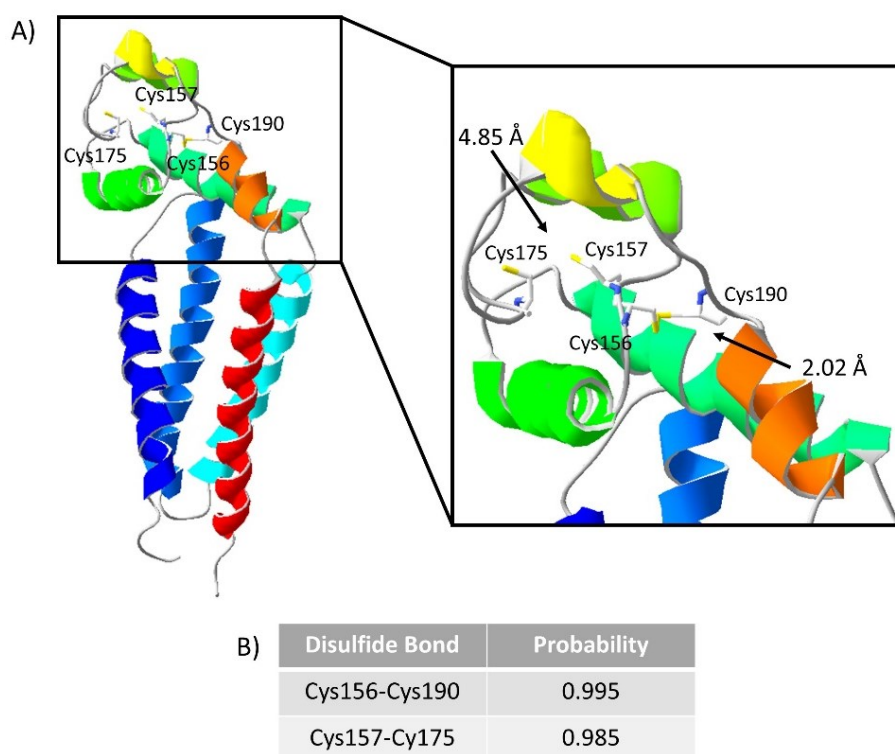


Figure 4.4. Disulfide bond pattern in the loop of the LEL in the closed structure of CD81. (A) The solved closed structure of CD81 (PDB: 5TCX) with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of closed CD81 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of closed CD81 as predicted by SSBondPre (Gao et al., 2020).

In the solved open structure of CD81, the Cys156-Cys190 disulfide bond remains largely unchanged with 2.01 Å between the residues (Figure 4.5A), just 0.01 Å less than the distance in the closed structure. The probability predicted by SSBondPre is only 0.002 lower than that found in the closed structure at 0.993 (Figure 4.5B). Unlike in the closed structure, Cys157-Cys175 do form a disulfide bond in the solved open structure of CD81 with 2.03 Å between the residues (Figure 4.5A). The probability of these residues forming a disulfide bond, according to SSBondPre, increases slightly by 0.002 to 0.987 (Figure 4.5B).

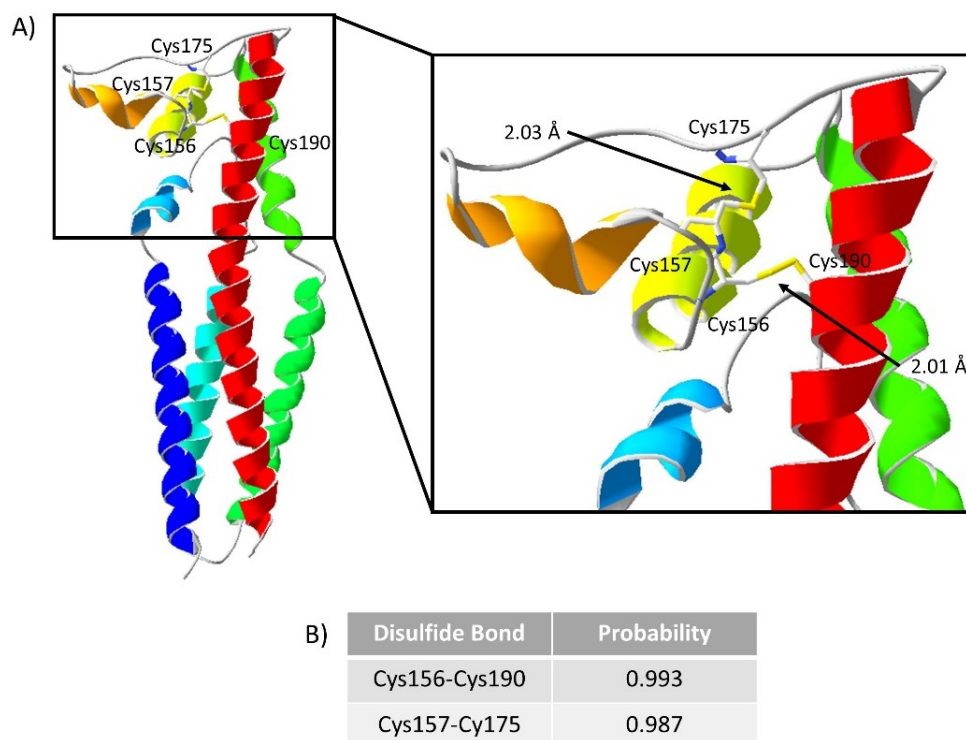


Figure 4.5. Disulfide bond pattern in the loop of the LEL in the open structure of CD81. (A) The solved open structure of CD81 (PDB: 7JIC) with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of open CD81 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of open CD81 as predicted by SSBondPre (Gao et al., 2020).

Given that there is little sequence conservation in the loops of tetraspanins (Figure 4.1), dividing them into their subfamilies and performing a sequence alignment might provide an insight into each subfamily. The first subfamily to have its loops aligned was TspanC4.

4.4 Sequence conservation in the loops of the LEL of TspanC4s

The TspanC4 subfamily consisting of CD81, CD9, CD53, Tspan2 and Tspan32 are one of the most widely studied subfamilies. All three of the full length solved structures of tetraspanins come from this family but there is little sequence conservation between the members in their loops.

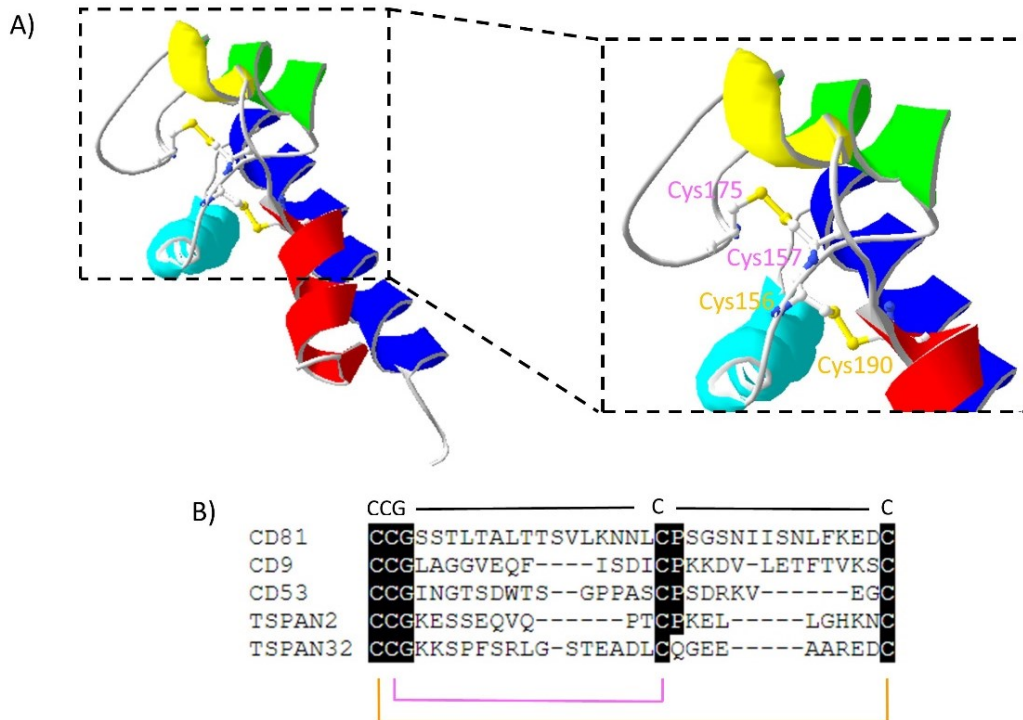


Figure 4.6. Disulfide bond pattern in the loop of the LEL in the closed structure of CD81 and sequence alignment of the loops of TspanC4s. (A) The solved closed structure of CD81 (PDB: 5TCX) with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of closed CD81 with the Cys residues and their disulfide bonds. (B) Sequence alignment of the loops in the LELs of TspanC4s aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

Other than the four Cys residues and the Gly in the CCG motif at the start of the loop the only other position that shows significant conservation is an 80% conserved Pro residue immediately after the middle Cys residue in all the proteins except Tspan32 which has a Gln residue (Figure 4.6B).

4.5 Structural characterisation and sequence conservation of the LEL in TspanC6-CCs

To try to further characterise the LELs, at least one loop from each subfamily was modelled by uploading the LEL sequence to trRosetta (Yang et al., 2020, Du et al., 2021). Doing this provided an insight into the structural characteristics of the tetraspanins which is beneficial because, to date, the only full-length structures of human tetraspanins all belong to the TspanC4 subfamily. Secondary structure prediction was performed by uploading the full-length sequences to PSIPRED (McGuffin et al., 2000) as a comparison with the 3D structure produced by trRosetta.

The LEL region of CD151, a member of the TspanC6-CC subfamily, was modelled using trRosetta to produce a typical looking tetraspanin LEL with an α helix, β helix and ϵ helix which are believed to be consistent through all tetraspanins. Between the β helix and the ϵ helix, there are two more, short, helical structures known as the γ helix and the δ helix (Figure 4.7A).

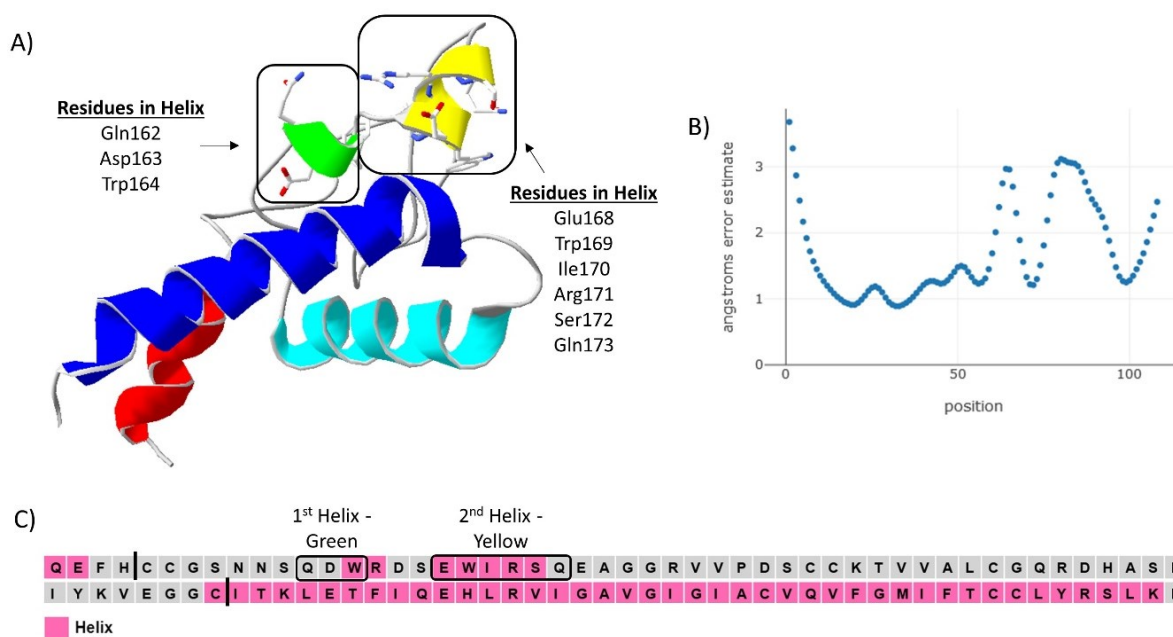


Figure 4.7. Modelled structure of the LEL of CD151 with secondary structure prediction. (A) Structure of the LEL of CD151, a member of the TspanC6-CC subfamily, modelled using trRosetta (Yang et al., 2020). Structured regions with the loop connecting the β helix to the ϵ helix are highlighted in black boxes. (B) Plot of the estimated error in Å for each amino acid in the modelled structure in (A). (C) Secondary structure prediction of the loop in the LEL of CD151 using PSIPRED (McGuffin et al., 2000). The start and end of the loop are situated between the two black lines. Amino acids forming part of a predicted helical structure are coloured in a pink background. Amino acids that make up secondary structures predicted by the trRosetta model of CD151 (A) are highlighted in the black boxes.

The γ helix is a short helical structure consisting of three amino acids (Gln162, Asp163 and Trp164) which is partly predicted by PSIPRED. PSIPRED predicts that a helical structure is formed by Trp164 and Arg165 (Figure 4.7C). The secondary structure prediction by PSIPRED is closer to the structure predicted using trRosetta for the γ helix because it predicts a five-residue helical structure comprising Glu168, Trp169, Ile170, Arg171 and Ser 172 (Figure 4.7C). This prediction matches the trRosetta structure except for the inclusion of Gln173 in the helical structure in the predicted model (Figure 4.7A). There is little error predicted in the first half of the CD151 LEL, which is likely because the first half is well structured consisting of four helical structure, whereas there is an estimated 3 Å error for some of the residues in the second, largely unstructured, half of the LEL before the ϵ helix (Figure 4.7B).

The disulfide bond patterns, using the modelled CD151, and sequence conservation in the loops of the LEL of TspanC6s were investigated. CD151 modelled with trRosetta has three disulfide bonds (Figure 4.8A). The first disulfide bond is formed between Cys145, which is the first Cys residue in the CCG motif, and Cys191 and the probability of a disulfide bond between these residues is 0.994, as predicted by SSBondPre (Figure 4.8B). The second disulfide bond, Cys146-Cys169, exists between the second Cys residue in the CCG motif and the first of the two Cys residues that are located next to each other in the middle of the loop (Figure 4.8A). SSBondPre predicts that the probability of a disulfide bond between Cys146 and Cys169 is 0.983 (Figure 4.8B). The disulfide bond with the highest probability, at 0.996 (Figure 4.8B), is between the second Cys residue of the two situated next to each other in the middle of the loop, Cys170, and the fifth Cys of the six in the loop, Cys177 (Figure 4.8A).

Aligning the sequences of the TspanC6-CC loops highlights some positions other than the Cys residues that have significant sequence conservation (Figure 4.8C). There is a 100% conserved Pro residue three amino acids before the two middle Cys residues, as well as a Gly residue one amino acid before the last Cys residue in all but one protein. The exception is Tspan4 which has a Pro residue immediately preceding the final Cys residue (Figure 4.8C). TspanC6-CCs have a position entirely occupied by aromatic amino acids near the beginning of the loop. The 10th position consists of 79% Trp residues and 21% Phe residues, while the 9th position is 93% Asp and 7% Asn (Figure 4.8C).

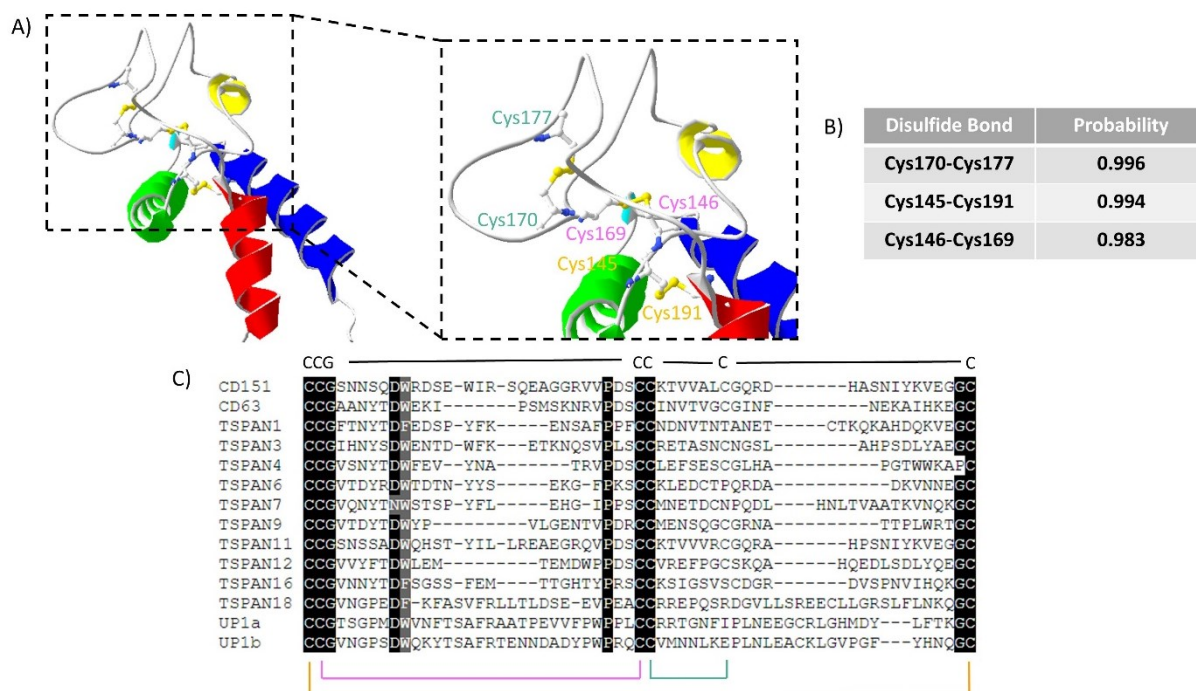


Figure 4.8. Disulfide bond pattern in the loop of the LEL in the modelled structure of CD151 and sequence alignment of the loops of TspanC6-CCs. (A) The modelled structure of CD151 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled CD151 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of modelled CD151 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (C) Sequence alignment of the loops in the LELs of TspanC6-CCs aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

4.6 Structural characterisation and sequence conservation of the LEL in TspanC6-CxCs

CD37 is a member of the TspanC6-CxC subfamily and has the third longest loop of all human tetraspanins at a length of 79 residues. CD37's LEL was modelled using trRosetta and the result was a more complex structure than the CD81 and CD151 tetraspanin LELs. Near the start of the loop, just after the β helix is a short, four residue, helical structure which has been seen in other models and would be called the γ helix. This helical structure consists of Gln159, Asp160, Trp161 and Phe162 (Figure 4.6A). Trp161 and Phe162 are predicted to be the start of a three-residue helix using PSIPRED, which also predicts another three residue helix two residues after this one (Figure 4.6C).

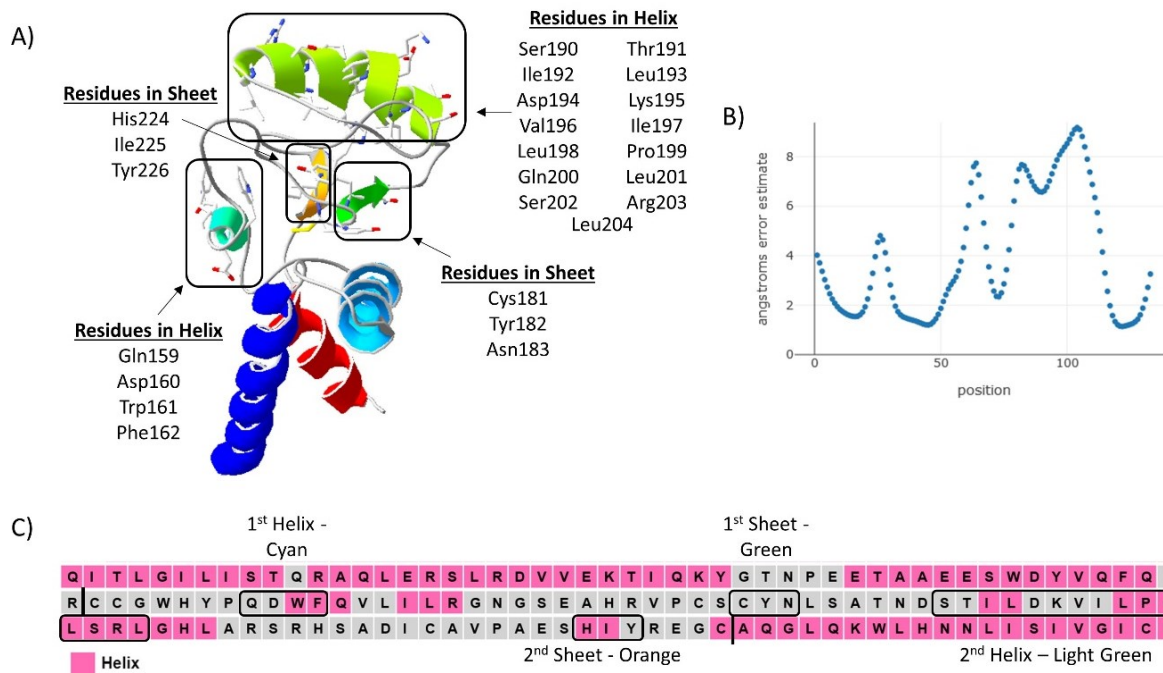


Figure 4.9. Modelled structure of the LEL of CD37 with secondary structure prediction. (A) Structure of the LEL of CD37, a member of the TspanC6-CxC subfamily, modelled using trRosetta (Yang et al., 2020). Structured regions with the loop connecting the β helix to the ϵ helix are highlighted in black boxes. (B) Plot of the estimated error in Å for each amino acid in the modelled structure in (A). (C) Secondary structure prediction of the loop in the LEL of CD37 using PSIPRED (McGuffin et al., 2000). The start and end of the loop are situated between the two black lines. Amino acids forming part of a predicted helical structure are coloured in a pink background. Amino acids that make up secondary structures predicted by the trRosetta model of CD37 (A) are highlighted in the black boxes.

The second predicted helix by PSIPRED is not present in the trRosetta structure. The next structured region in the model is a short sheet comprising Cys181, Tyr182 and Asn183 (Figure 4.9A) which was not predicted by PSIPRED (Figure 4.9C). Following this sheet is a long, 15 residue, helix that is partially predicted by PSIPRED. PSIPRED predicts a Ile192 and Leu193 helix followed by four unstructured residues and then a ten-residue helix (Figure 4.9C). Combining these two predicted helices together means that nine of the residues predicted to be in a helical structure by PSIPRED are in the 15-residue helix modelled by trRosetta. The modelled helix consists of Ser190, Thr191, Ile192, Leu193, Asp194, Lys195, Val196, Ile197, Leu198, Pro199, Gln200, Leu 201, Ser 202, Arg203 and Leu204 (Figure 4.9A).

The final structured region in the modelled CD37 loop has good agreement between the trRosetta model and the PSIPRED prediction with regards to the residues involved but not the structure. In the model a sheet comprising His224, Ile225 and Tyr226 is situated just before the end of the loop and the ϵ helix (Figure 4.9A). PSIPRED predicts that His224 and Ile225 form a helix, not a sheet (Figure 4.9C). The first half of the modelled CD37 LEL structure has small levels of estimated error with much of this region having an estimated error of >2 Å

(Figure 4.9B). The greater estimated error in the model occurs in the second half of the structure with significant parts between 6-8 Å estimated error. The 19-residue strand between the 15-residue helix and the second sheet has particularly high levels of estimated error which go above 8 Å, nearly reaching 10 Å (Figure 4.9B).

The disulfide bond pattern of the CD37 model that was modelled using trRosetta demonstrated a disulfide bond pattern not seen before in any of the other tetraspanin. In other tetraspanins the second Cys residue in the CCG motif forms a disulfide bond with the first Cys of the middle two. In CD37, however, the second Cys in the CCG motif forms a disulfide bond with the second Cys of the middle two, leaving the first Cys of the middle two to form a disulfide bond with the fifth Cys residue in the loop (Figure 4.10A). This finding is backed up by the SSBondPre prediction because the only possible disulfide bonds predicted are the ones that are formed in the model. They are the Cys162-Cys230, Cys163-Cys181 and Cys179-Cys217 bonds which have probabilities of 0.990, 0.979 and 0.961 respectively (Figure 4.10B).

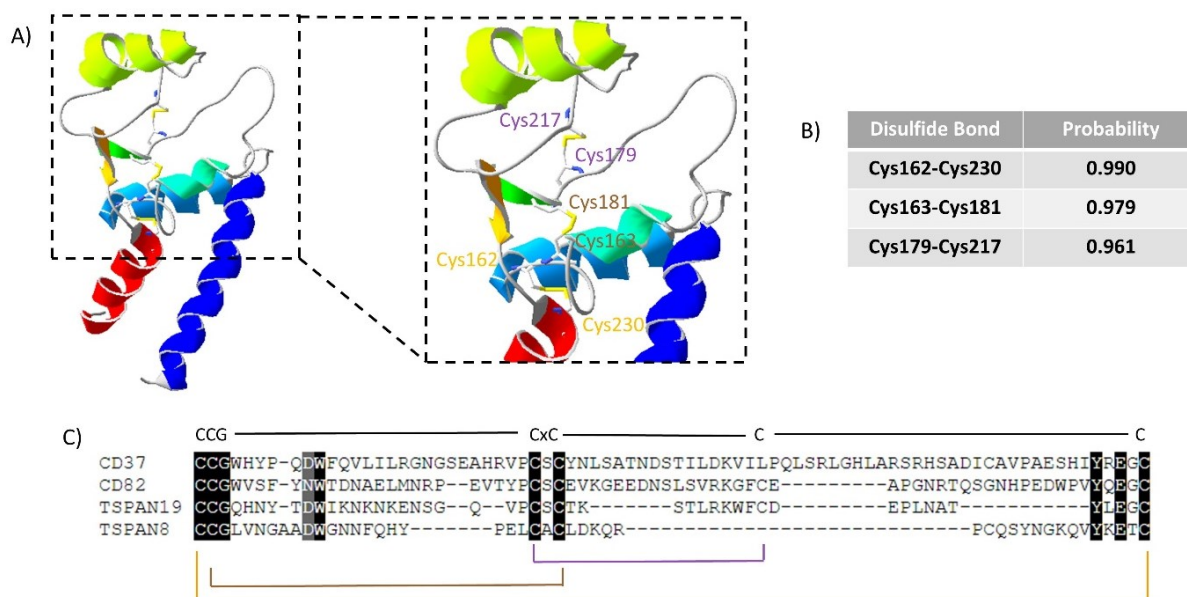


Figure 4.10. Disulfide bond pattern in the loop of the LEL in the modelled structure of CD37 and sequence alignment of the loops of TspanC6-CxCs. (A) The modelled structure of CD37 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled Cd37 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of modelled CD37 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (C) Sequence alignment of the loops in the LELs of TspanC6-CxCs aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

Other than a (D/N)-W region near the start of the loop and a YxEx region immediately before the final Cys residue there is no sequence conservation (Figure 4.10C). TspanC6-CxC loops also have vastly different lengths ranging from the shorter Tspan8 and Tspan19 loops with 44 and 48 amino acids respectively to CD82 and CD37 with 68 and 79 amino acids respectively (Figure 4.10C).

Given that the CD37 LEL model had an unusual and unexpected disulfide bond pattern the other three TspanC6-CxCs, CD82, Tspan19 and Tspan8, LELs were modelled using trRosetta to see if this disulfide bond pattern is a characteristic of this subfamily or just CD37. The CD82 LEL model has the normal disulfide bond between the first and last Cys in the loop between Cys149-Cys216 (Figure 4.11A) with a probability of 0.991 predicted by SSBondPre (Figure 4.11B). Except for CD37, all other modelled LELs have a disulfide bond between the second and third Cys residues, however, CD82, like CD37, has a disulfide bond between the second and fourth Cys residues (Figure 4.11A). The Cys150-Cys176 bond has a probability of 0.986 (Figure 4.11B). There is no third disulfide bond formed in the CD82 LEL model because Cys174 and Cys193 are 9.14 Å apart which is too great a distance for disulfide bond formation (Figure 4.11A). SSBondPre does not predict a bond between these two residues either. It does, however, predict with a high probability of 0.982 a disulfide bond between Cys176 and Cys193 (Figure 4.11B). In the event of this bond forming, it would free Cys150, which is bound to Cys176 in the CD82 LEL model, to bind to Cys174 which would be the usual disulfide bond between the second and third Cys residues in the loop.

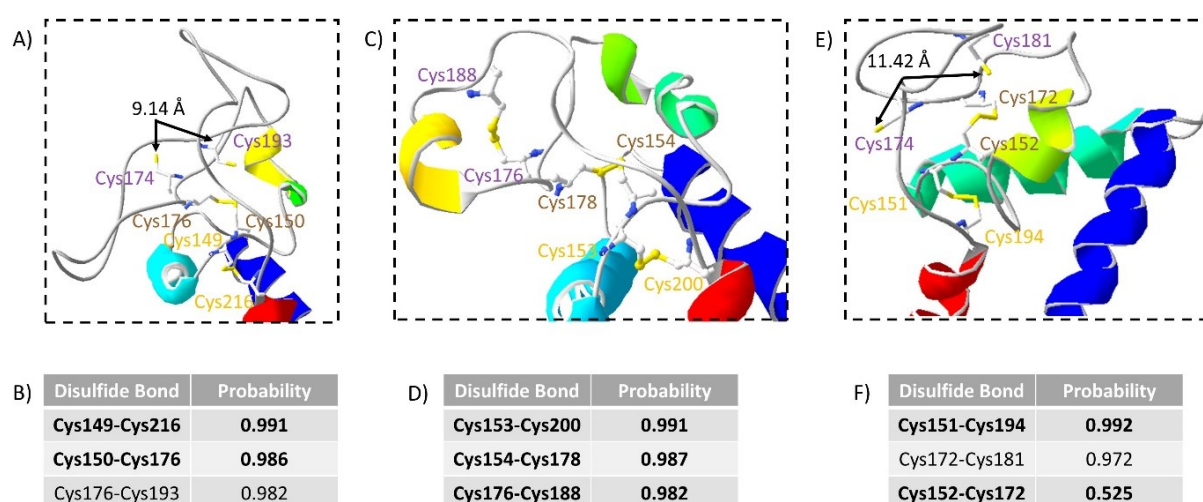


Figure 4.11. Disulfide bond pattern in the loop of the LEL in the modelled structure of CD82, Tspan19 and Tspan8. (A) The modelled structure of CD82 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled CD82 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loop of the LEL of modelled CD82 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (C) The modelled structure of Tspan19 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled Tspan19 with the Cys residues and their

disulfide bonds. (D) The probability of disulfide bond formation between two named Cys residues in the loop of the LEL of modelled Tspan19 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (E) The modelled structure of Tspan8 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled Tspan8 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loop of the LEL of modelled Tspan8 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold.

The Tspan19 LEL model exhibits the same disulfide bond arrangement as that seen in the CD37 LEL model (Figure 4.11C). Cys153, the first Cys residue, forms a bond with Cys200, the last Cys residue and the probability of this is 0.991, as predicted by SSBondPre (Figure 4.11D). The second Cys residue, Cys154, forms a disulfide bond with the fourth Cys residue, Cys178, not the third third Cys residue as seen in other tetraspanin models. As a result, the third Cys residue, Cys176, and the fifth Cys residue, Cys188, interact with each other (Figure 4.11C). The probability of the formation of these disulfide bonds according to SSBondPre is 0.987 and 0.982 respectively (Figure 4.11D).

Tspan8, however, does not have a disulfide bond between the second Cys residue and the fourth Cys residue, a phenomenon seen in the other member of the TspanC6-CxC subfamily. It has the more common second Cys and third Cys disulfide bond formed between Cys152-Cys172 (Figure 4.11E). Interestingly, however, despite this bond existing the Tspan8 LEL model the probability predicted by SSBondPre is only 0.525 (Figure 4.11F). The probability of Cys172, the third Cys residue, forming a disulfide bond with Cys181, the fifth Cys residue, is much higher at 0.972 (Figure 4.11F). If this bond did form, then the possibility exists for the second and fourth Cys to form a disulfide bond like in the other TspanC6-CxC LEL models. However, there is no predicted bond between these cysteines by SSBondPre. The only other predicted disulfide bond is the usual bond between the first Cys residue, Cys151, and the sixth Cys residue, Cys194, with a probability of 0.992 (Figure 4.11F). No bond is predicted or present in the Tspan8 LEL model between Cys174-Cys181, the fourth and fifth Cys residues, because there is an 11.42 Å distance between them which is too big for disulfide bond formation (Figure 4.11E).

4.7 Structural characterisation and sequence conservation of the LEL in TspanC6-CxxxCs

Tspan13 and Tspan31 make up a small subfamily called TspanC6-CxxxC, so Tspan13 was modelled using trRosetta. The predicted secondary structure using PSIPRED did not predict any structured region between the β helix and the ϵ helix (Figure 4.12C), however, the trRosetta predicted structure predicts a helical structure in the middle of the loop between the β helix and the ϵ helix (Figure 4.12A). The helical structure comprises Cys140, Leu141,

Ala142, Ser143 and Cys144, so this structure contains the two middle Cys residues, separated by three amino acids, that form disulfide bonds with other Cys residues in the loop. The estimated error in the model follows a similar pattern to that seen the CD151 LEL model with less error in the first half of the structure and more error in the second half, especially between the β helix and the ϵ helix (Figure 4.12B).

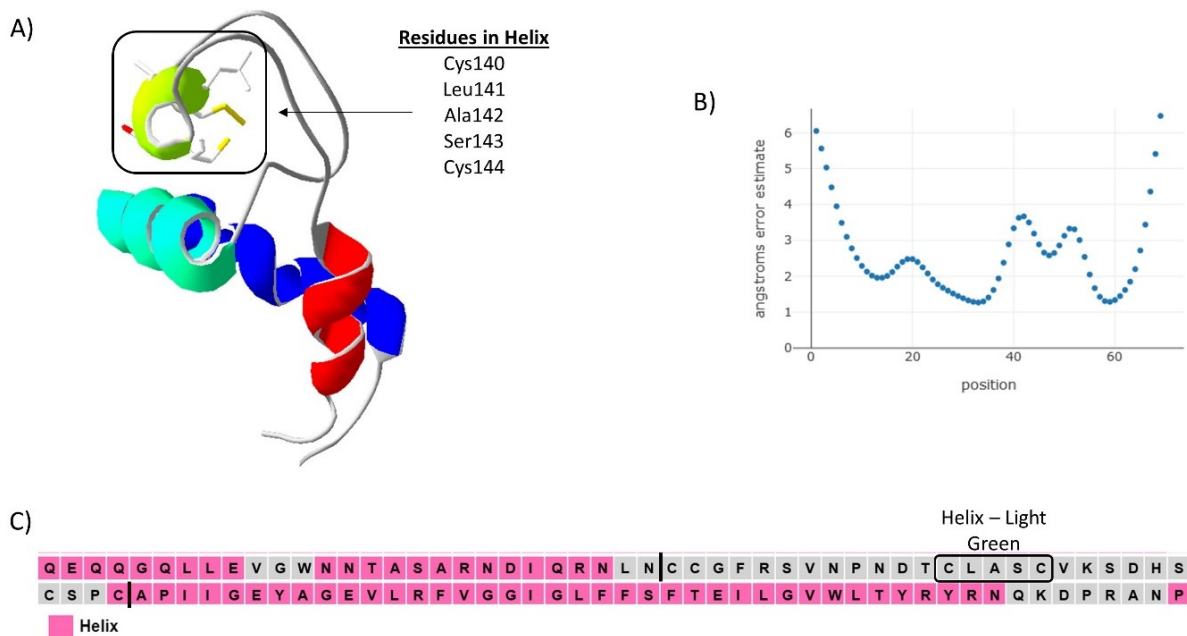


Figure 4.12. Modelled structure of the LEL of Tspan13 with secondary structure prediction. (A) Structure of the LEL of Tspan13, a member of the TspanC6-CxxxC subfamily, modelled using trRosetta (Yang et al., 2020). Structured regions with the loop connecting the β helix to the ϵ helix are highlighted in black boxes. (B) Plot of the estimated error in Å for each amino acid in the modelled structure in (A). (C) Secondary structure prediction of the loop in the LEL of Tspan13 using PSIPRED (McGuffin et al., 2000). The start and end of the loop are situated between the two black lines. Amino acids forming part of a predicted helical structure are coloured in a pink background. Amino acids that make up secondary structures predicted by the trRosetta model of Tspan13 (A) are highlighted in the black boxes.

TspanC6-CxxxCs have a similar cysteine arrangement to TspanC6-CCs with the notable difference of three amino acids between the two middle Cys residues, hence the nomenclature TspanC6-CxxxC. Tspan13 was modelled using trRosetta and the resulting structure has three disulfide bonds between the six Cys residues (Figure 4.13A).

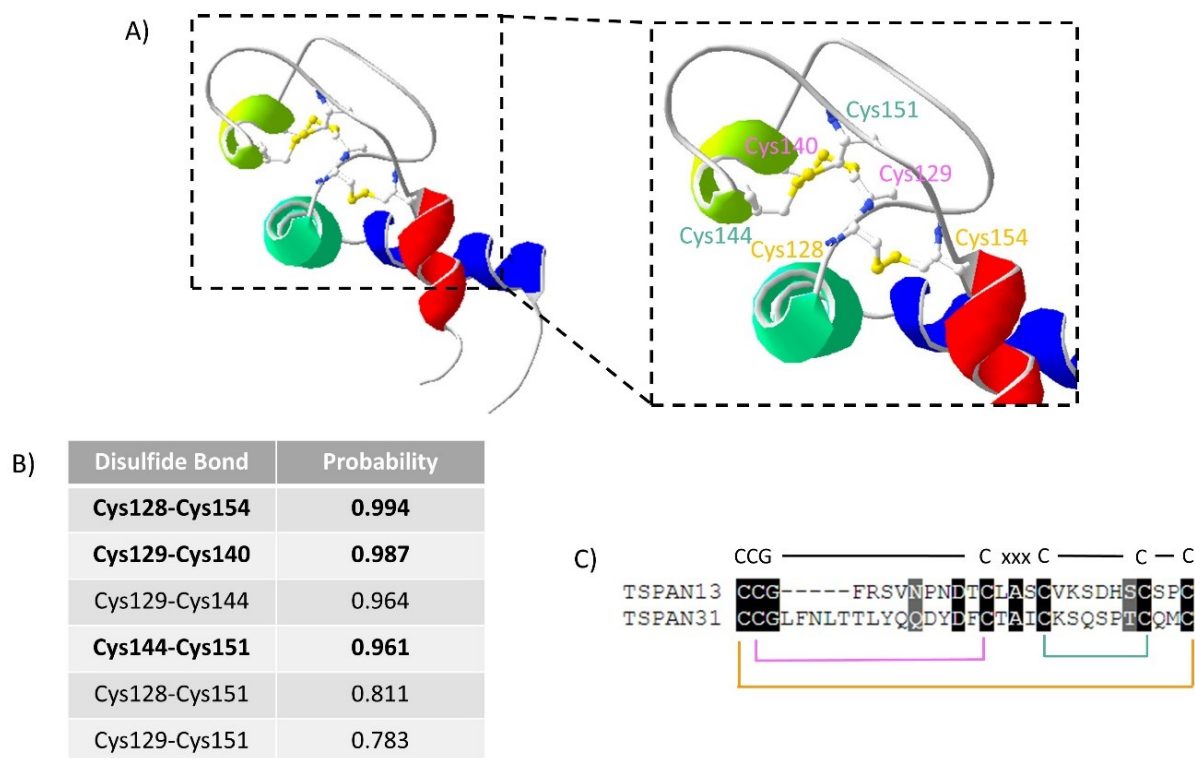


Figure 4.13. Disulfide bond pattern in the loop of the LEL in the modelled structure of Tspan13 and sequence alignment of the loops of TspanC6-CxxxCs. (A) The modelled structure of Tspan13 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled Tspan13 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of modelled Tspan13 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (C) Sequence alignment of the loops in the LELs of TspanC6-CxxxCs aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

The three disulfide bonds are formed between Cys128-Cys154, Cys129-Cys140 and Cys144-Cys151, which have probabilities of forming disulfide bonds predicted by SSBondPre of 0.994, 0.987 and 0.961 respectively (Figure 4.13B). SSBondPre, however, highlights three other potential disulfide bonds between these six Cys residues. Two of them, Cys128-Cys151 and Cys129-Cys151, have a lower probability than three disulfide bonds that are formed in the Tspan13 model. Cys128-Cys151 has a probability of 0.811 and Cys129-Cys151 has a probability of 0.783 (Figure 4.13B). A predicted bond between Cys129 and Cys144, on the other hand, has a higher probability than the Cys144-Cys151 bond found in the Tspan13 model with a probability of 0.964, which is 0.003 higher than Cys144-Cys151 (Figure 4.13B). Besides the six conserved Cys residues, TspanC6-CxxxCs have a conserved Ala residue in

the middle of the CxxxC arrangement in the centre of the loop (Figure 4.13C). Two residues before the CxxxC sequence is a conserved Asp residue.

4.7 Structural characterisation and sequence conservation of the LEL in TspanC6 β s

ROM1 is one of two members of the TspanC6 β subfamily and its structure, modelled by trRosetta, has similar levels of complexity to CD37, due in large part because it has the joint longest loop in human tetraspanins with 86 residues. The other human tetraspanin with 86 residues is the other member of the TspanC6 β subfamily, PRPH2. Of all the LELs modelled ROM1 has the biggest discrepancy between the trRosetta model and the PSIPRED secondary structure prediction. PSIPRED only predicts the presence of one helical structure in the loop (Figure 4.14C) but the trRosetta model has four helices in the loop (Figure 4.14A).

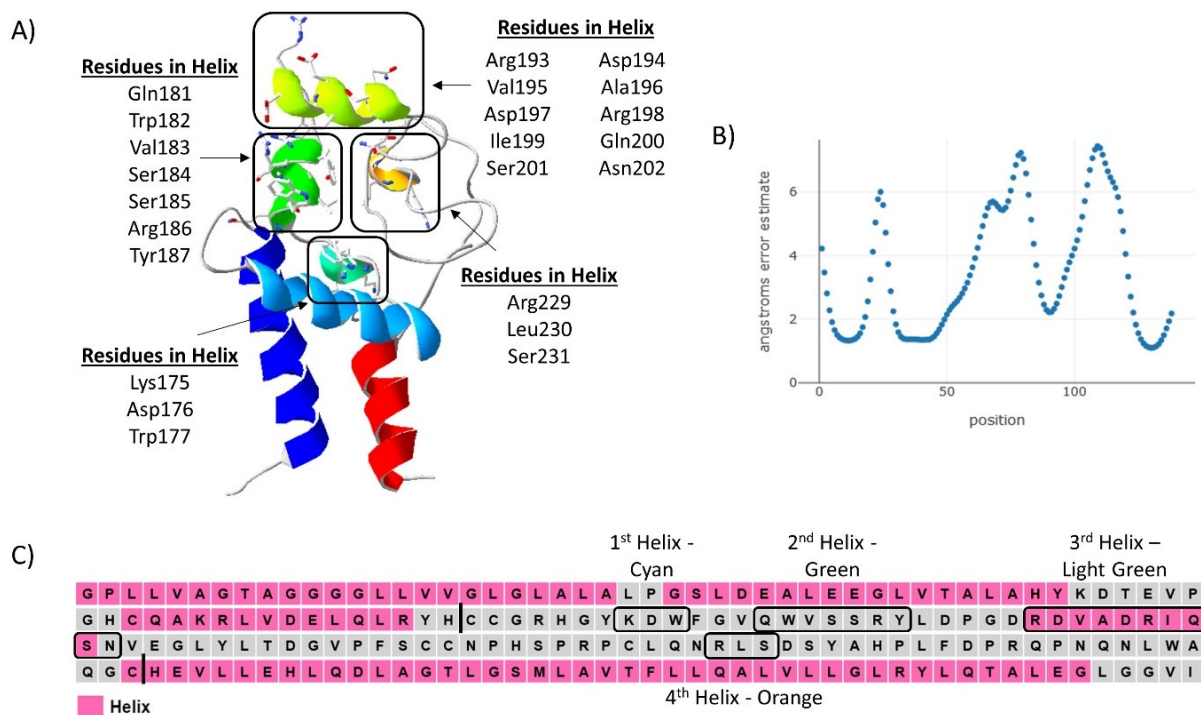


Figure 4.14. Modelled structure of the LEL of ROM1 with secondary structure prediction. (A) Structure of the LEL of ROM1, a member of the TspanC6 β subfamily, modelled using trRosetta (Yang et al., 2020). Structured regions with the loop connecting the β helix to the ϵ helix are highlighted in black boxes. (B) Plot of the estimated error in Å for each amino acid in the modelled structure in (A). (C) Secondary structure prediction of the loop in the LEL of ROM1 using PSIPRED (McGuffin et al., 2000). The start and end of the loop are situated between the two black lines. Amino acids forming part of a predicted helical structure are coloured in a pink background. Amino acids that make up secondary structures predicted by the trRosetta model of ROM1 (A) are highlighted in the black boxes.

The first two helices in the modelled loop are not predicted by PSIPRED (Figure 4.14C). The first modelled loop is near the beginning of the loop and is made up of Lys175, Asp176 and Trp177, while the second helix starts four residues after the end of the first one and consists

of Gln181, Trp182, Val183, Ser184, Ser185, Arg186 and Tyr187 (Figure 4.14A). Although PSIPRED only predicts one helix is in the loop the one it does predict is almost entirely in agreement with the third helix in the modelled LEL. In the modelled ROM1 the next helix contains Arg193, Asp194, Val195, Ala196, Asp197, Arg198, Ile199, Gln200, Ser201 and Asn202 (Figure 4.14A), which is the same as the helix predicted by PSIPRED except for the omission of Asn202 (Figure 4.14C). A final short helix consisting of Arg229, Leu230 and Ser231 is situated in the middle of a large unstructured strand after the third helix in the model (Figure 4.14A).

Like the other modelled LELs most of the first half of the structure of ROM1 has error estimate for each position of $<2 \text{ \AA}$, while the second half having a higher error estimate of between 4-8 \AA (Figure 4.14B).

The first two helices in the modelled loop are not predicted by PSIPRED (Figure 4.14C). The first modelled loop is near the beginning of the loop and is made up of Lys175, Asp176 and Trp177, while the second helix starts four residues after the end of the first one and consists of Gln181, Trp182, Val183, Ser184, Ser185, Arg186 and Tyr187 (Figure 4.14A). Although PSIPRED only predicts one helix is in the loop the one it does predict is almost entirely in agreement with the third helix in the modelled LEL. In the modelled ROM1 the next helix contains Arg193, Asp194, Val195, Ala196, Asp197, Arg198, Ile199, Gln200, Ser201 and Asn202 (Figure 4.14A), which is the same as the helix predicted by PSIPRED except for the omission of Asn202 (Figure 4.14C). A final short helix consisting of Arg229, Leu230 and Ser231 is situated in the middle of a large unstructured strand after the third helix in the model (Figure 4.14A). Like the other modelled LELs most of the first half of the structure of ROM1 has error estimate for each position of $<2 \text{ \AA}$, while the second half having a higher error estimate of between 4-8 \AA (Figure 4.14B).

The TspanC6 β subfamily is a unique subfamily of human tetraspanins. They not only have six Cys residues in their loops but they also have a seventh Cys residue in the LEL which is located 15 residues before the CCG motif in the β helix (Figure 4.15C). TspanC6 β proteins have the same cysteine arrangement as TspanC6-CCs with the two middle Cys residues situated next to each other. Like the other subfamilies previously investigated with six Cys residues in the loop the cysteine arrangement is a disulfide bond between the first Cys residue in the CCG motif and the sixth Cys residue in the loop. In ROM1, which was modelled using trRosetta (Figure 4.15A), this disulfide bond is formed between Cys168 and Cys253 with a 0.994 probability predicted by SSBondPre (Figure 4.15B). The second Cys residue in the CCG motif, Cys169 in ROM1, forms a disulfide bond with the first Cys residue of the middle two, Cys216 in ROM1 (Figure 4.15A), with a probability of 0.978 (Figure 4.15B). The third disulfide

bond in TspanC6βs is between the fourth and fifth Cys residue in the loop. Cys217, the second Cys of the middle two, forms a disulfide bond with the fifth Cys residue in the loop, Cys225 in ROM1 (Figure 4.15A) with a probability of 0.989 (Figure 4.15B). SSBondPre also predicts the possibility of a disulfide bond between Cys216 and Cys225, although the probability of this bond is much lower than the three formed in the ROM1 model with a probability of 0.605 (Figure 4.15B).

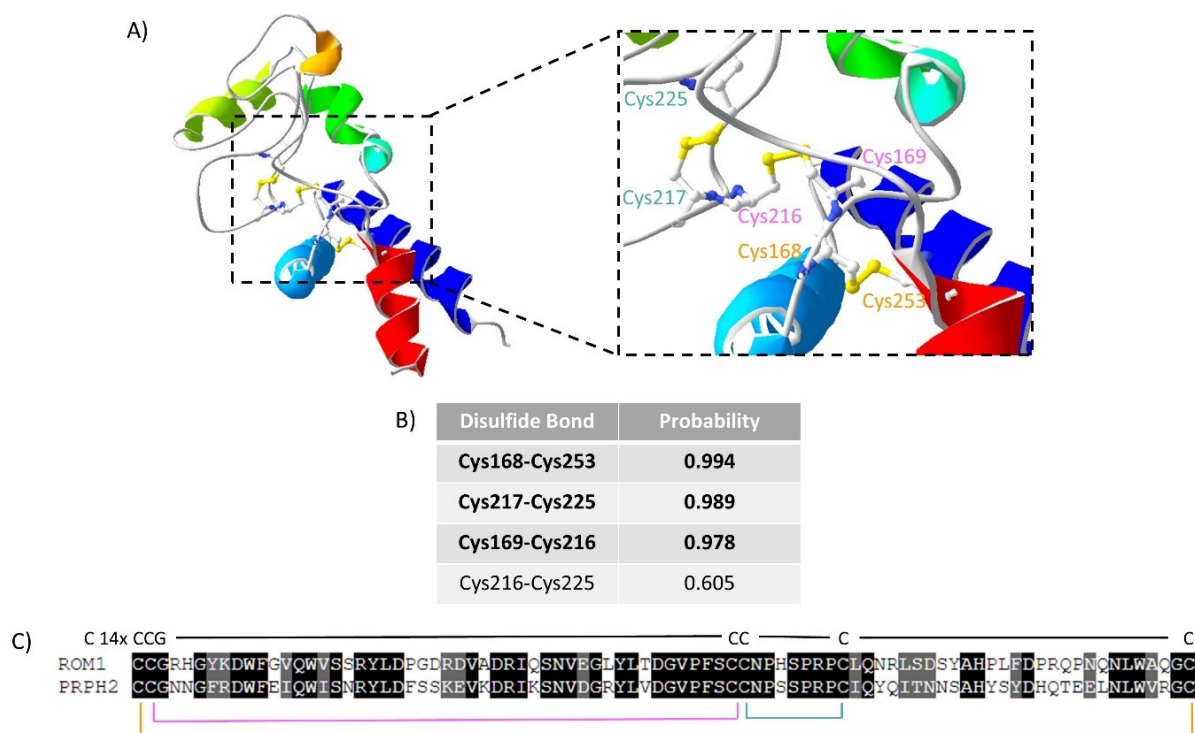


Figure 4.15. Disulfide bond pattern in the loop of the LEL in the modelled structure of ROM1 and sequence alignment of the loops of TspanC6βs. (A) The modelled structure of ROM1 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled ROM1 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of modelled ROM1 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (C) Sequence alignment of the loops in the LELs of TspanC6βs aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

Sequence alignment of TspanC6βs shows large amounts of sequence conservation (Figure 4.15C). The loops of ROM1 and PRPH2 are identical lengths, both containing 86 amino acids, 48 of which are conserved and a further 14 are similar amino acids. Most of the conserved residues are in the first two thirds of the loop sequence with just nine of the 48 conserved residues located in the final third (Figure 4.15C).

4.7 Structural characterisation and sequence conservation of the LEL in TspanC8s

Tspan15 belongs to the TspanC8 subfamily and its LEL region was modelled using trRosetta, resulting in a similar structure to Tspan13 (Figure 4.12A), albeit with a much longer unstructured region between the β helix and the ϵ helix (Figure 4.16A). Like Tspan13, there is a small helical structure between the β helix and the ϵ helix comprised of Arg161, Asp162, Trp163 and Ser164 (Figure 4.16A). PSIPRED does predict that there is one helical structure in the loop of Tspan15 but it is predicted to be just two amino acids in length (Figure 4.16C). Despite this, there is some overlap between the PSIPRED prediction and the trRosetta structure. The helical structure predicted by PSIPRED includes the last residue in the helical structure predicted by trRosetta, Ser164, and includes the following residue, Lys165.

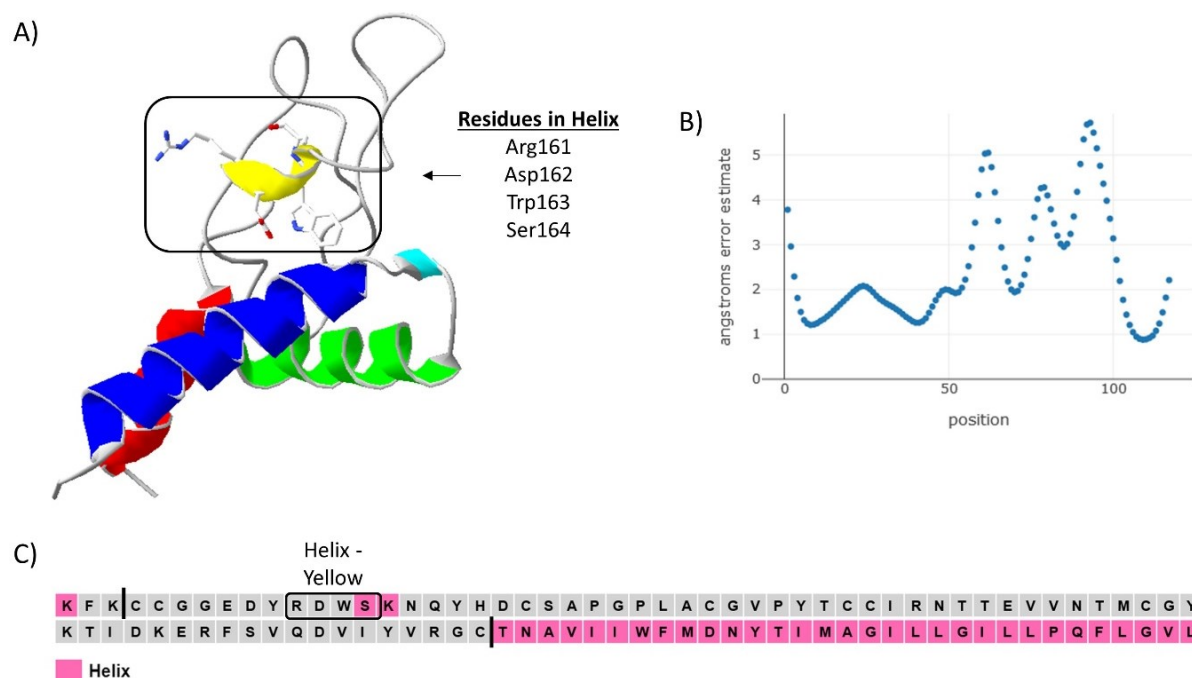


Figure 4.16. Modelled structure of the LEL of Tspan15 with secondary structure prediction. (A) Structure of the LEL of Tspan15, a member of the TspanC8 subfamily, modelled using trRosetta (Yang et al., 2020). Structured regions with the loop connecting the β helix to the ϵ helix are highlighted in black boxes. (B) Plot of the estimated error in Å for each amino acid in the modelled structure in (A). (C) Secondary structure prediction of the loop in the LEL of Tspan15 using PSIPRED (McGuffin et al., 2000). The start and end of the loop are situated between the two black lines. Amino acids forming part of a predicted helical structure are coloured in a pink background. Amino acids that make up secondary structures predicted by the trRosetta model of Tspan15 (A) are highlighted in the black boxes.

Due to the well-defined α and β helices there is little error estimated in the trRosetta model in the first half of the structure but there are more pronounced error estimates in the second half of the structure which is mostly a long unstructured strand before the ϵ helix (Figure 4.16B).

As a result of this large unstructured region the estimated error for some of the amino acids rises as high as 5 Å.

Since producing the Tspan15 LEL model a solved structure of this region has been published (Lipper et al., 2022). The solved structure of the Tspan15 LEL is like the one modelled by trRosetta except for the inclusion of two short sheets and another short helical structure between the predicted helical structure in the loop and the ϵ helix (Figure 4.17B). The first helical structure in the solved structure is identical to the modelled structure because it contains the same four residues. Overlaying the two structures (Figure 4.17C) demonstrates the similarity with regards to the first half of the LEL and the core structures, the α , β and ϵ helices. The RMSD between the two molecules is 4.09 Å, highlighting both the strengths of using computational methods to predict protein structure and its limitations because neither trRosetta or PSIPRED predicted the two sheets and extra helix in the loop.

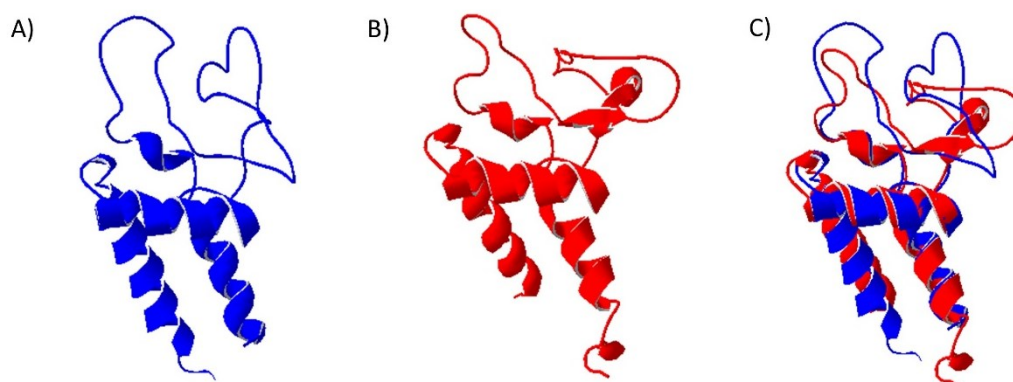


Figure 4.17. Comparison of the modelled Tspan15 LEL and the solved structure of Tspan15. (A) The modelled structure of the LEL of Tspan15 using trRosetta (Yang et al., 2020). (B) The solved structure of the LEL of Tspan15 (PDB: 7RDB). (C) The modelled structure of the LEL of Tspan15 overlaid with the solved structure of Tspan15 with an RMSD between the two molecules of 4.09 Å.

TspanC8s, named because they are the only subfamily of tetraspanins to have eight Cys residues in their loops, consist of Tspan5, Tspan10, Tspan14, Tspan 15, Tspan17 and Tspan33 (Figure 4.18C). Despite having two more Cys residues than other known human tetraspanins the Cys arrangement is the same as that found in TspanC6-CCs except for two Cys residues between the CCG motif and the middle two Cys residues which form a disulfide bond with each other, as seen in the Tspan15 LEL model (Figure 4.18A). The disulfide bonding pattern between the eight Cys residues is 1-8, 2-5, 3-4 and 6-7.

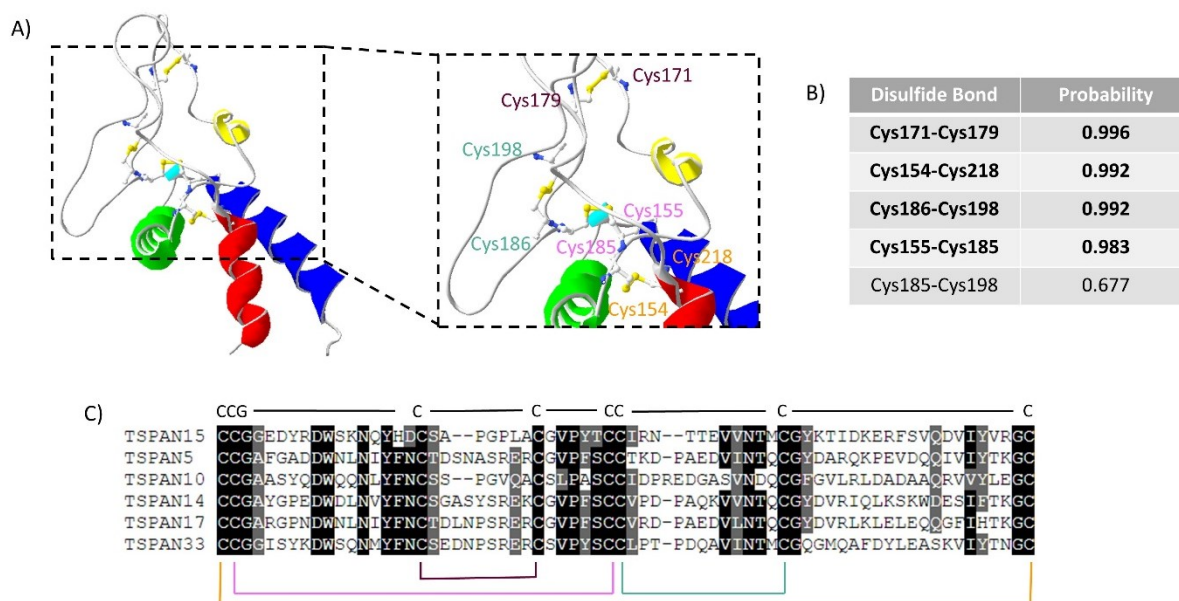


Figure 4.18. Disulfide bond pattern in the loop of the LEL in the modelled structure of Tspan15 and sequence alignment of the loops of TspanC8s. (A) The modelled structure of Tspan15 with the Cys residues in the loop of the LEL highlighted. A close-up of the LEL of modelled Tspan15 with the Cys residues and their disulfide bonds. (B) The probability of disulfide bond formation between two named Cys residues in the loops of the LEL of modelled Tspan15 as predicted by SSBondPre (Gao et al., 2020). Disulfide bonds formed in the modelled are indicated in bold. (C) Sequence alignment of the loops in the LELs of TspanC8s aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >80% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

The 1-8 bond between Cys154-Cys218 in Tspan15 has a probability of 0.992, as predicted by SSBondPre (Figure 4.18B). The predicted probability of a disulfide bond between the 2-5 Cys residues, Cys155-Cys185, is lower at 0.983, while the 6-7 bond between Cys186-Cys198 has a probability of 0.992, the same as the Cys154-Cys218 bond. The unique 3-4 bond is formed between Cys171-Cys179 in Tspan15 and has the highest probability of 0.996 (Figure 4.18B). SSBondPre also predicts the potential for a disulfide bond between Cys185-Cys198, which would be a 5-7 bond, although the probability of this disulfide bond is much lower at 0.677.

TspanC8s have a high degree of sequence conservation in the loop (Figure 4.18C). Six amino acids after the CCG motif at the start of the loop is a conserved DWxxNxY region which ends three amino acids before the third Cys residue. Except for Tspan15 this conserved region extends up to the third Cys residue for Tspan5, Tspan10, Tspan14, Tspan17 and Tspan33 with all of them having a DWxxNxYFN conserved region immediately before the third Cys residue in the loop (Figure 4.18C). Between the fourth and fifth Cys residue is another region with high conservation. The four residues prior to the fifth Cys residue are (V/L)-Px-(S/T) with

Val 80% conserved at the start and Ser 80% conserved at the end. Five amino acids before the seventh Cys residue is a V-(V/I/L)-NT region in all but Tspan10 which has a Ser residue in place of Val at the beginning and an Asp residue instead of the Thr at the end (Figure 4.18C). Like the TspanC6 β subfamily most sequence conservation in the loops of TspanC8s is in the first two thirds of the sequence.

4.8 Conservation in the SEL of tetraspanins by subfamily

Having characterised the TMs, the α , β and ϵ helices in the LEL (Chapter 3) and the loop in region in the LEL in this chapter the only significant region of tetraspanins yet to be explored is the small extracellular loop which connects TM1 and TM2. Conservation in this region was initially explored by aligning all 33 human tetraspanin SELs with Clustal Omega (Sievers et al., 2011) (Figure 4.19).

```

CD81      WLRHDPQTNNLLYLELGDKPAPNT--FYVGIY
CD9       WRFDSQTKSIFEQETNNNN--SS--FYTGVY
CD53      YLLI-HNNFGVLFHNLPSLTLGN-----
CD82      WILADKSSFISVLQTSSSSLRMGA-Y-----
CD151     WTLALKSDYISLLASGT--YLATA-Y-----
CD63      -----GAQLVLSQTIIQGATPGS-LLPV---
CD37      WILIDKTSFVSVFVGLAFVPLQIWS-K-----
TSPAN1    WVSIDGASFLKIFGPLSSAMQFV----NVGY
TSPAN2    WERFGGAIKELSED----KSPEY--FYVGLY
TSPAN3    YVFITYDDYDHFEDVYTLIPAV-----
TSPAN4    WLAATQGSFATLSSSFPSLSAAN-----
TSPAN5    WAWNEKGVLSNISSITDLGGFDPV-----W--
TSPAN6    WCKVSLENYFSLLEKATNVFF-----
TSPAN7    WCKLTLGTYISLIAENSTNAPY-----
TSPAN8    WVRVSNDSQAIFGSEDVG----SS--SYVAVD
TSPAN9    WLSVSQGNFATFSPSPSLSAAN-----
TSPAN10   WCLAVKGSGLGSDLGGPLPTDP--M-----L--
TSPAN11   WTLVEKSGYLSVLASST--FAASA-Y-----
TSPAN12   WMRDYLNNVLTTLTAETRVEEAVILTYFPVHP
TSPAN13   WCI GFGLISSLRVVG-----
TSPAN14   WAWSEKGVLSDLTKVTRMHGIDPV-----V--
TSPAN15   YAEVERQKYKTLES AFLAPAI-----
TSPAN16   GKCGGASLTNVLGLSS-----AY--LLHVG
TSPAN17   WAWGEGVLSNISALTDLGGLDPV-----W--
TSPAN18   WVMVDPTGFREIVAANPL-LLTGA-Y-----
TSPAN19   WLLDRNNFLTAFDENNHFIVPISQ-----
TSPAN31   WCKGLGLVSSIHIIIG-----
TSPAN32   LTYFGAHFAVIRRASLEKNPYQAV-HQWAFSA
TSPAN33   YARLMKHAEAAALACLAVDPAI-----
UP1a     WVTADQYRVYPLMGVSGKDDVFAGAW-----
UP1b     EFVSDQHSLYPLLEATDNDDIYGAAW-----
ROM1     HDLVQLRHLGTFLAPSCQ-----FPVLPQ
PRPH2    FLKIELRKRSDVMNNSE-----SH--FVPN--

```

Figure 4.19 Sequence alignment and conservation of the SELs of human tetraspanins. The SELs of human tetraspanins were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >70% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

The SEL is one region of tetraspanins that will almost no conservation when all 33 human tetraspanins are aligned. The SELs range in length from as little as 15 residues to 32 residues, making aligning the sequences difficult. Once the sequences were aligned with Clustal Omega (Sievers et al., 2011) only the first two residues in the SEL exhibit significant conservation (Figure 4.19).

Neither of these positions have one amino acid that is >70% conserved. The first two positions are 70% conserved when the amino acids are grouped together in the following groups: Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro. The first position in the SEL of human tetraspanins is 88% a residue from Group 2, with Trp making up the most significant portion (Figure 4.19), while the second position consists of 75% of residues from Group 1.

Separating the alignments of the SEL up into groups based on their Cys arrangement in the loop provides some insight into sequence conservation for some of the subfamilies. Nevertheless, TspanC4 (Figure 4.20A) and TspanC6-CC (Figure 4.20E) do not have much sequence conservation. There is not one amino acid anywhere in the TspanC4 SEL that is >70% conserved (Figure 4.20A).

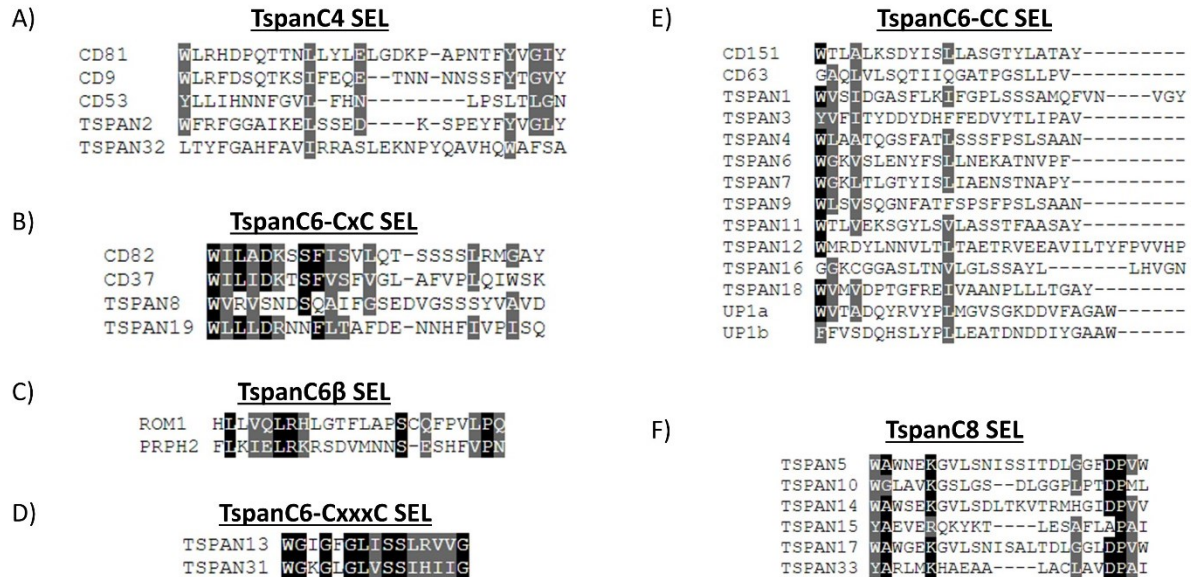


Figure 4.20 Sequence alignment and conservation of the SELs by subfamily of human tetraspanins. The SELs of human tetraspanins were separated into their subfamilies based on Cys arrangement in their loops: (A) TspanC4 (B) TspanC6-CxC (C) TspanC6 β (D) TspanC6-CxxxC (E) TspanC6-CC (F) TspanC8. They were aligned using Clustal Omega (Sievers et al., 2011). Sequence conservation was determined using Sequence Manipulation Suite: Color Align Conservation (Stothard, 2000) by looking for residues that are >70% conserved with identical amino acids highlighted in black and similar amino acids highlighted in grey. Similar amino acids were grouped based on chemical and structural similarity. Group 1: Gly, Ala, Val, Leu and Ile. Group 2: Phe, Tyr and Trp. Group 3: Cys and Met. Group 4: Lys, Arg and His. Group 5: Asp, Glu, Asn, Gln. Group 6: Ser and Thr. Group 7: Pro.

TspanC6-CxC SELs are highly conserved in the first half of their sequences (Figure 4.20B). All four of them start with a Trp residue followed by three small hydrophobic residues (Ala, Leu, Ile or Val), except for the third residue in the Tspan8 SEL which is an Arg residue. The next residue is a 75% conserved Asp, except for Tspan8 which has a Ser instead. CD82, CD37 and Tspan19 have a positively charged residue that follows the Asp residue, while Tspan8 differs again with an Asn residue (Figure 4.18B). Two positions after the positively charged residue is a 75% conserved Ser which appears in all but Tspan19 which has an Asn, while the next residue is a 75% conserved Phe except for Gln in Tspan8. The second half of the TspanC6-CxC sequences have little conservation (Figure 4.20B).

The two TspanC6 β s, ROM1 and PRPH2, have conserved residues near the start of the SEL (Figure 4.20C). The second residue is a conserved Leu, as is the sixth residue which is followed by a conserved Arg. Immediately preceding the second conserved Leu residue is a small hydrophobic amino acid followed by the chemical and structurally similar amino acids Gln/Gln (Figure 4.20C). The residue after the conserved Arg in ROM1 is a His and a Lys in PRPH2, both of which are positively charged. Other than a conserved Ser in the 16th position and a conserved Pro in the penultimate position of the SEL, there is not any more conservation in TspanC6 β SELs (Figure 4.20C).

Tspan13 and Tspan31, members of the TspanC6-Cxxx subfamily, are the most conserved subfamily when looking at the SEL (Figure 4.20D). They both have the SEL length at 15 residues which are also the shortest SELs of any human tetraspanin. Only two of the 15 residues have no similarity and eight of them are identical. TspanC6-CxxxCs begin with a Trp residue and five of the next seven residues are small hydrophobic residues. One area of particular interest could be the next two amino acids which are two Ser residues in both proteins (Figure 4.20D).

The TspanC8 SEL has little conservation in the middle of the sequence but some conservation at the ends of the sequence (Figure 4.20F). The first residue in all TspanC8 SELs is an aromatic amino acid (67% Trp and 33% Tyr), followed by an 83% Ala residue. An 83% conserved Lys residue is situated in the sixth position. The only TspanC8 that does not have a Lys residue here is Tspan15 which has another positively charged residue, Arg (Figure 4.20F). Four positions from the end of the SEL is an 83% conserved Asp residue with Tspan15 being the only exception again because it has an Ala residue. The next position, three residues from the end of the SEL is a 100% conserved Pro (Figure 4.20F).

4.8 Conclusion

It was difficult to align the sequences of the loops in the LEL so that a numbering system could be devised but through analysis of tetraspanins by subfamily based on Cys residues this thesis

has shed some light on disulfide bond arrangement. Tetraspanins were divided into their subfamilies and disulfide bonds were predicted with surprising results for the TspanC6-CxC family which in three of the four members forms a disulfide bond between the second Cys residue in the CCG-motif and the second Cys residue as the two middle Cys residues in the loop. This provides some insight into tetraspanin disulfide bonds and raises the question as to whether other tetraspanins may have a similar or whether disulfide bond arrangement is flexible but certain subfamilies of tetraspanin have a preference for one disulfide bond over another.

This work has also shown that while sequence conservation in the loop across all human tetraspanins is scarce it is much higher when dividing them into their subfamilies which could pave the way for researchers to look at specific conserved residues, particularly in under-researched tetraspanins, such as the TspanC6-CxxxC.

Separating the human tetraspanins into subfamilies also demonstrates that the SELs of the same family members are more likely to exhibit conservation than human tetraspanins as a whole family. Further analysis of some of these conserved residues could shed some light on some of the functions of the SEL.

This thesis also proposes new, clearer names for each subfamily which gives a clear description of how many Cys residues are in the LEL and the sequence of the two middle Cys residues.

Finally, it has been shown through modelling of tetraspanin LEL structures that some of them are more complex with more structural regions in the loop between the β helix and the ϵ helix than had previously been thought.

Chapter 5 – Expression and purification of the tetraspanin, CD81

The structures of three tetraspanins, CD81, CD9 and CD53, have been elucidated in recent years. CD81 and CD9 have been captured in a closed conformation and CD81 has also been captured in an open conformation, while CD53 was in a semi-open conformation (Zimmerman et al., 2016; Umeda et al., 2020; Susa et al., 2021). The discovery of CD81 in a closed and open conformation raises the question of what causes this change in conformation. Molecular modelling suggests that the binding of cholesterol to the TM region cavity is a determining factor with the closed conformation preferred when cholesterol is bound and the open conformation preferred when cholesterol is not bound. To see whether CD81 changes conformation and if so, whether the binding of cholesterol is a factor, the aim of this chapter was to purify a mutated CD81 protein to probe conformational change using EPR.

To study the conformational change of CD81 using EPR requires sufficient amounts of pure protein. To do this CD81-p-null, a human CD81 protein with all palmitoylation sites removed to prevent unwanted protein-protein interactions, has previously been cloned in *Pichia pastoris* for overexpression. *P. pastoris* cells are grown and the membrane isolated to be able solubilise CD81-p-null (Jamshad et al., 2008). Being able to solubilise a membrane protein, such as CD81-p-null, with surrounding lipids is advantageous because it can increase stability of the protein and lipids are known to have a functional role for some membrane proteins. To capture the surrounding lipids SMA2000 can be used and a recently developed method whereby SMA2000 is hydrolysed by autoclaving was investigated to see if it offered similar solubilisation efficiency to that of refluxed SMA2000, which is the traditional method of hydrolysis. Furthermore, given that pure protein was required for further experimentation to be done, the purification protocol for CD81-p-null solubilised with SMA2000 was optimised.

5.1 Solubilisation Efficiency of Autoclaved SMA2000

The hydrolysis of styrene-maleic anhydride polymer to the membrane active styrene-maleic acid form has usually been carried out using reflux in NaOH for several hours (Lee et al., 2016; Rothnie, 2016), which required some specialist equipment. However, it was reported in 2019 that autoclaving could substitute for the reflux and that full hydrolysis of the SMA2000 occurs after two autoclave cycles (Kopf et al., 2019). Therefore, solubilisation of CD81-p-null was tested using SMA2000 that had been autoclaved for two or three cycles for a greater comparison. To see if there was any difference between the solubility of SMA2000 when solubilising CD81-p-null in *P. pastoris* membranes the autoclaved SMA2000 was compared to refluxed SMA2000.

CD81-p-null is sufficiently solubilised after one hour (Figure 5.1A), so solubilisation efficiency was checked after one hour and two hours. After one hour the SMA2000 prepared using the

reflux method had a solubilisation efficiency of $65.82\% \pm 13.46$, whereas the autoclave method had a solubilisation efficiency of $52.93\% \pm 10.82$ and $52.31\% \pm 12.06$ for two autoclave cycles and three autoclave cycles respectively with a P value determined by one-way ANOVA >0.05 at 0.09 indicating that the results are not statistically significant (Figure 5.1B). The difference between the solubilisation efficiency of the reflux method and the autoclave method is 12.89% when compared to the SMA2000 autoclaved for two cycles and 13.51% when compared to the SMA2000 autoclaved for three cycles. There is just a small difference of 0.62% in solubilisation efficiency between the SMA2000 autoclaved for two and three cycles.

After two hours the SMA2000 prepared using the reflux method had a solubilisation efficiency of $69.91\% \pm 9.44$, whereas the autoclave method has a solubilisation efficiency of $56.53\% \pm 7.95$ and $57.58\% \pm 12.75$ for two autoclave cycles and three autoclave cycles respectively with a P value determined by one-way ANOVA >0.05 at 0.07 indicating that the results are not statistically significant (Figure 5.1B). The difference between the solubilisation efficiency of the reflux method and the autoclave method is 13.38% when compared to the SMA2000 autoclaved for two cycles and 12.33% when compared to the SMA2000 autoclaved for three cycles. The difference between the solubilisation efficiency of SMA2000 autoclaved for two and three cycles is 1.05%.

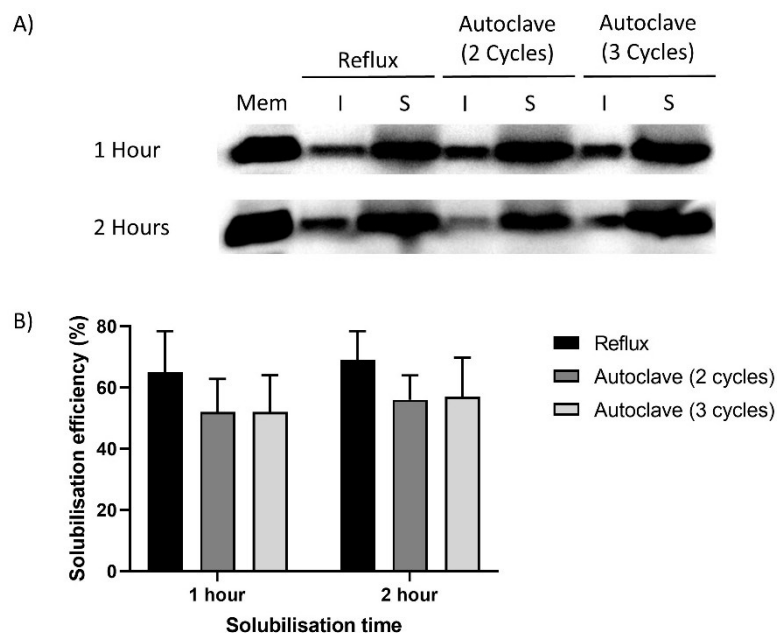


Figure 5.1. Comparison of the solubilisation efficiency of autoclaved SMA2000 compared with refluxed SMA2000. (A) Western blot of the solubilisation of CD81-p-null at 1 hour and 2 hours with refluxed SMA2000 and SMA2000 autoclaved for 2 cycles and 3 cycles. 20 μ l of sample was loaded into each well and the nitrocellulose membrane was incubated with a 1:5000 dilution of 6x His monoclonal antibody (Takara) and then incubated with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signaling Technology). The membrane was then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene). Mem = Membrane fraction, I = Insoluble fraction and S = Soluble fraction (B) Bar chart of the average solubilisation efficiency at 1 hour and 2 hours of refluxed SMA2000 and SMA2000 autoclaved for 2 cycles and 3 cycles. Error bars = SD (n = 3).

Despite a decrease in solubilisation efficiency of between 12.33%-13.51% when using the autoclaved SMA2000 when compared to the reflux method, the autoclaved method still produced a solubilisation efficiency >50%. Given the solubilisation efficiency is >50% and the small difference between the SMA2000 autoclaved for two cycles and three cycles all future SMA2000 used was prepared by autoclaving for two cycles.

When preparing the autoclaved SMA2000 it was noticeable that transferring the SMA2000 that had been left to dissolve in 1 M NaOH overnight into a glass bottle and leaving it to rest for approximately 15-30 minutes resulted in an accumulation of solid SMA2000 stuck to the bottom of the bottle after the first autoclave cycle (Figure 5.2A). To overcome this problem the SMA2000 dissolved in 1 M NaOH was autoclaved immediately after being transferred to the bottle which resulted in a clear bottom on the base of the bottle (Figure 5.2B).

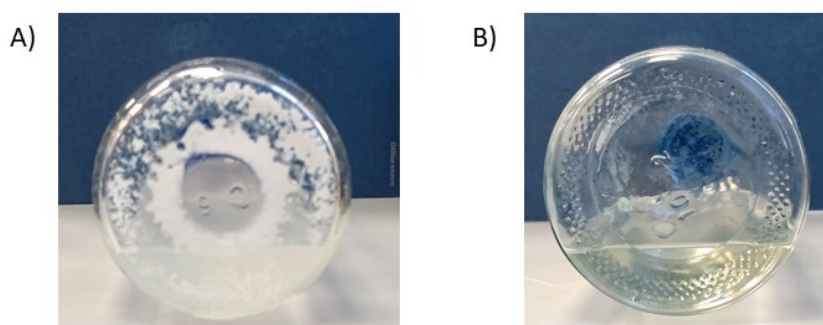


Figure 5.2. Comparison of SMA2000 stuck to the bottom of the bottle after first the autoclave cycle. (A) The bottom of the bottle containing dissolved SMA2000 in 1 M NaOH that was left static for 30 minutes before autoclaving. After the first autoclave cycle an accumulation of SMA2000 stuck to the bottom of the bottle can be seen. (B) The bottom of the bottle containing dissolved SMA2000 in 1 M NaOH that was autoclaved immediately after being transferred to the bottle. After the first autoclave cycle a clear bottom of the bottle can be seen.

5.2 Optimisation of the Purification of CD81-p-null Solubilised with SMA2000

To be able to perform experiments, such as EPR, to shed light on the mechanism of action with regards to CD81 adopting the open and closed conformation purified protein is required. In the first instance, it was shown that CD81-p-null could be expressed in *P. pastoris* and purified before trying to express and purify any mutants of CD81-p-null that would be required for EPR experiments.

A small scale purification of CD81-p-null was performed by solubilising 3 ml of *P. pastoris* membrane expressing CD81-p-null (Jamshad et al., 2008; Ayub et al., 2020) with 2.5% (w/v) SMA2000. The solubilised protein was incubated with Ni-NTA resin overnight at 4°C. The resin was then washed with 20 mM and 40 mM imidazole and the protein was eluted with 300 mM imidazole.

Under these conditions a clear, dense band can be seen in the 3rd and 4th 300 mM imidazole elution (Figure 5.3) which is similar to what has been shown previously (Ayub et al., 2020). CD81 has a mass of 25,809 Da and these clear bands are situated just below the 25 kDa marker on the protein ladder.

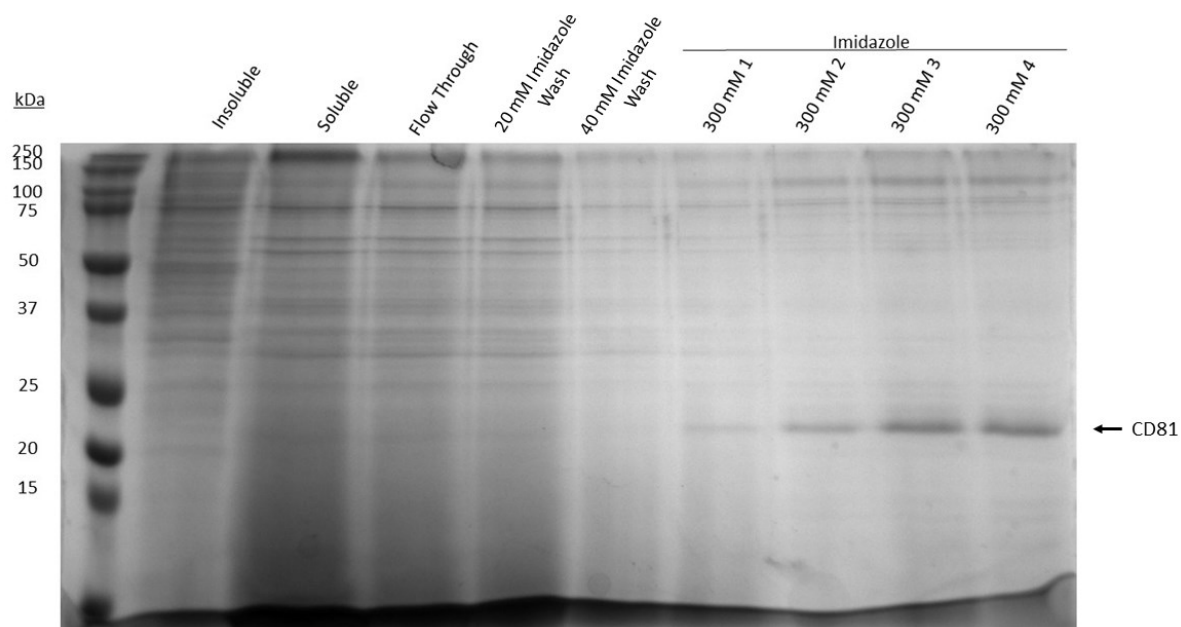


Figure 5.3. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null with 20 mM and 40 mM wash steps followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

The SDS-PAGE gel (Figure 5.3) shows a relatively pure CD81-p-null, but with many other proteins present in the elution fractions besides CD81-p-null. In particular, there is an abundance of protein at approximately 100-130 kDa. To try and improve the purity of CD81-p-null in the elution fractions an optimisation strategy was devised to include the doubling of the volume of both wash steps, an increase in imidazole concentration in the wash steps and the introduction of imidazole into the binding buffer when incubating the protein with the Ni-NTA resin.

The first strategy employed was to keep the concentration of imidazole in the wash steps at 20 mM and 40 mM but to double the volume of each wash step. In the first purification (Figure 5.3) the wash steps were 10x the bed volume of Ni-NTA resin used, therefore when the volume was doubled 20x the bed volume was used. Doubling the volume of the wash steps had little effect on the purity of CD81-p-null with bands at around the 100-130 kDa region still clearly visible (Figure 5.4). Far less protein is visible in the second wash step, however, which

indicates that more of the proteins that are binding to the Ni-NTA resin are eluting off the resin in the first wash step.

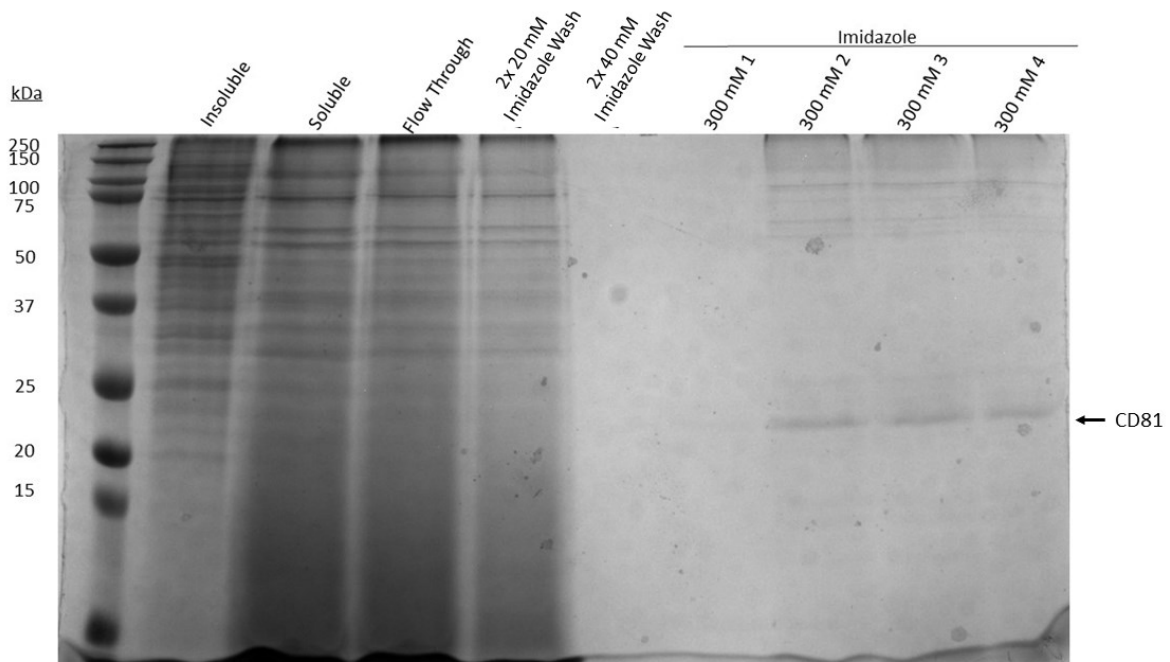


Figure 5.4. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in *P. pastoris* with the volume of the wash steps doubled. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null with 20x Ni-NTA resin bed volume in the wash steps followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

Keeping the wash steps at 10x the bed volume of Ni-NTA resin and increasing the concentration of imidazole in the wash steps from 20 mM and 40 mM to 30 mM and 60 mM had a more significant effect on the purity of CD81-p-null (Figure 5.5). Increasing the imidazole concentration had a similar effect to doubling the volume of the wash steps with regards to a lack of protein in the second wash step. The significant difference, however, is in the 300 mM imidazole elution steps with less protein visible at around the 100-130 kDa mark.

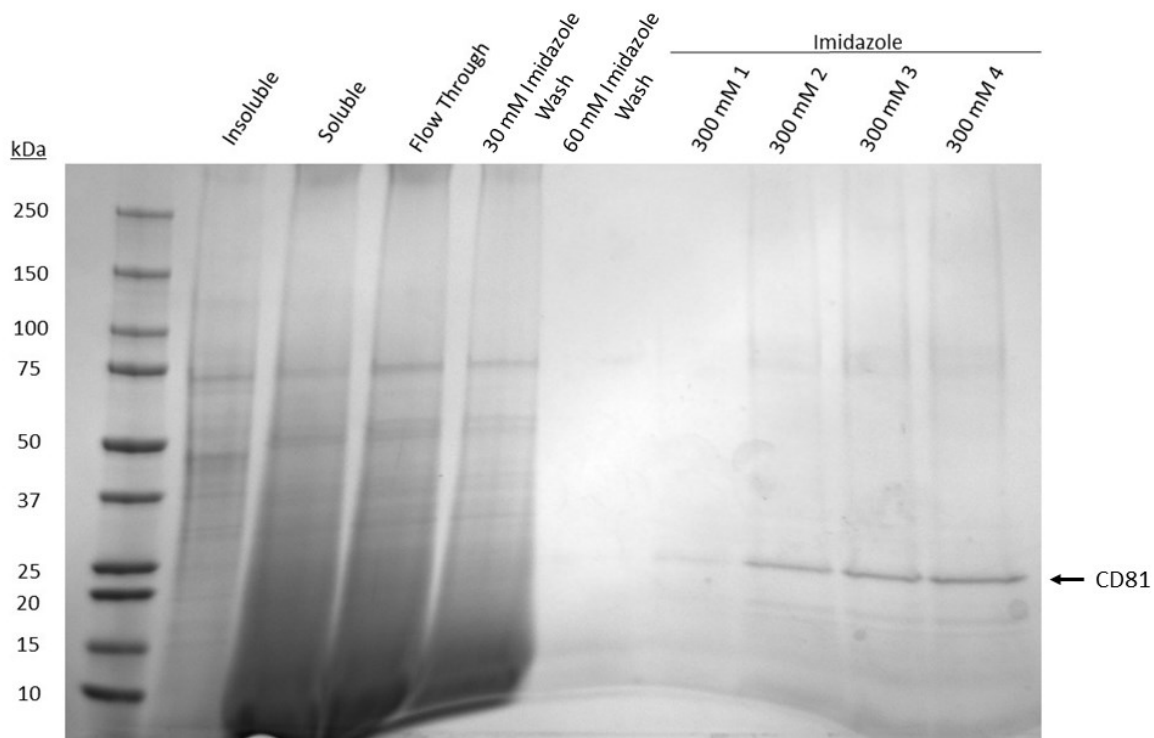


Figure 5.5. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in *P. pastoris* with the concentration of imidazole in the wash steps increased. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null with the concentration of imidazole in the wash steps increased from 20 mM and 40 mM to 30 mM and 60 mM followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

To try to improve the purity of CD81-p-null further, the addition of 20 mM imidazole to the binding buffer when incubating the solubilised protein with the Ni-NTA resin was investigated. The inclusion of imidazole in the binding buffer is known to reduce non-specific binding of proteins that are not the protein of interest due to imidazole binding to the Ni-NTA resin instead.

The same conditions used in the two previous purifications of doubling the volume of the wash steps and increasing the concentration of imidazole in the wash steps were used again. In both instances, however, 20 mM imidazole was included in the binding buffer. Combining a small concentration of imidazole in the binding buffer had a significant effect on the purity of CD81-p-null in the case of both purifications with almost no contaminating proteins visible in any of the elution steps.

When combining 20 mM imidazole in the binding buffer with doubling the volume of the wash steps, CD81-p-null purity is higher (Figure 5.6). There are no contaminating bands visible at the higher molecular weight of 100-130 kDa.

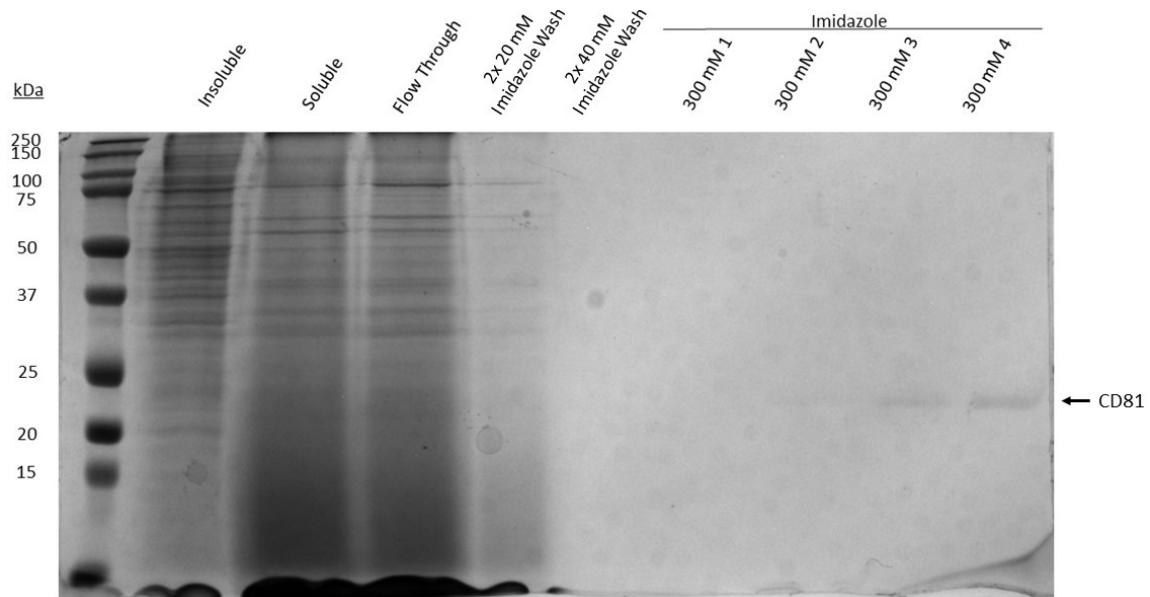


Figure 5.6. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in *P. pastoris* with 20 mM imidazole in the binding buffer and the volume of the wash steps doubled. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null with 20 mM imidazole included in the binding buffer. The Ni-NTA resin was washed with 20x bed volume followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

Adding 20 mM imidazole to the binding buffer and increasing the concentration of imidazole in the wash steps to 30 mM and 60 mM has a similar effect on the purity of CD81-p-null. The contaminating proteins at approximately 100-130 kDa are no longer visible (Figure 5.7). The one notable difference, however, is that using this combination of purification conditions yields more CD81-p-null in the elution fractions which can be seen by the darker, more dense bands corresponding with the molecular weight of CD81 (Figure 5.7).

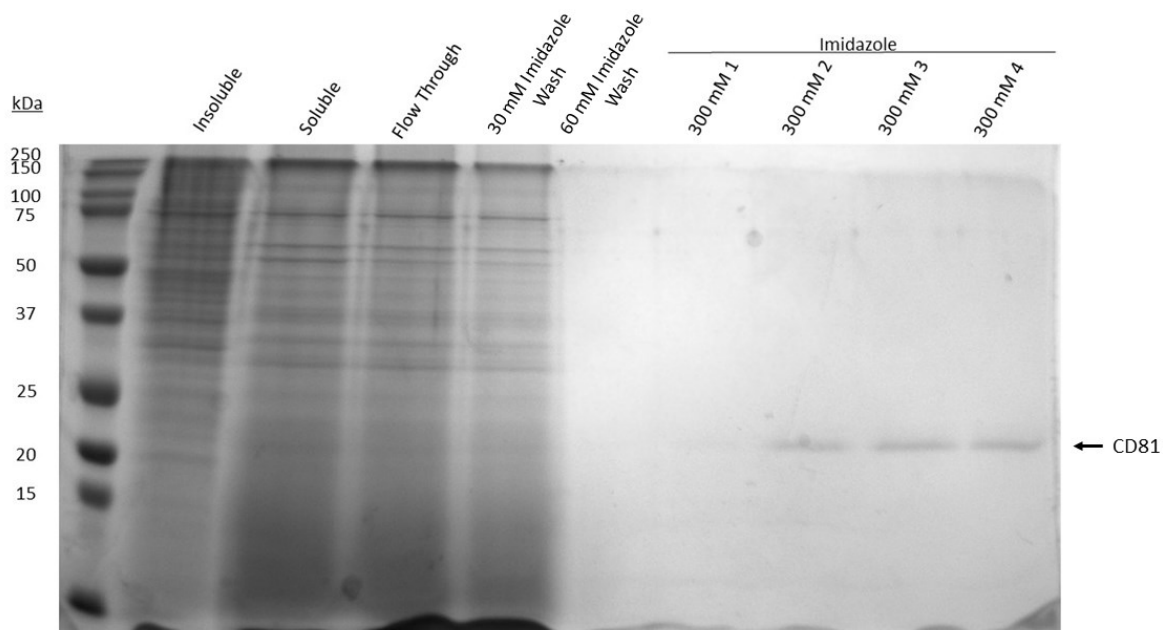


Figure 5.7. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null expressed in *P. pastoris* with 20 mM imidazole in the binding buffer and the increased imidazole in the wash steps. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null with 20 mM imidazole included in the binding buffer. The Ni-NTA resin was washed with 30 mM and 60 mM imidazole followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

Given the improved purity of the purification method that includes 20 mM imidazole in the binding buffer and an increase of imidazole in the wash steps from 20 mM and 40 mM to 30 mM and 60 mM these conditions were chosen for future purifications. To further validate this new optimised purification method for CD81-p-null a comparison between the new method (Figure 5.7) and the original purification method (Figure 5.3) was required.

To compare between these two methods CD81-p-null was purified using both methods and all the relevant fractions collected compared on a Western blot. With the introduction of 20 mM imidazole to the binding buffer to reduce non-specific binding it also introduces the possibility of reducing the amount of the protein of interest, CD81-p-null, that binds to the Ni-NTA resin. The Western blots, therefore, were done to check that there was not an excessive amount of CD81-p-null being lost in the flow through as a result of not being able to bind to the Ni-NTA resin because of the 20 mM imidazole in the binding buffer.

The Western blot of the original purification method shows minimal loss of CD81-p-null in the flow through (Figure 5.8A) which is to be expected due to a lack of imidazole in the binding buffer. In the optimised purification a marked increase in CD81-p-null present in the flow through could be seen (Figure 5.8B). Also, the intensity of the CD81 bands in the elution

fraction (compared to the membrane starting material) are a little lower. However, it was decided this small loss in protein was acceptable given the improvements in purity.

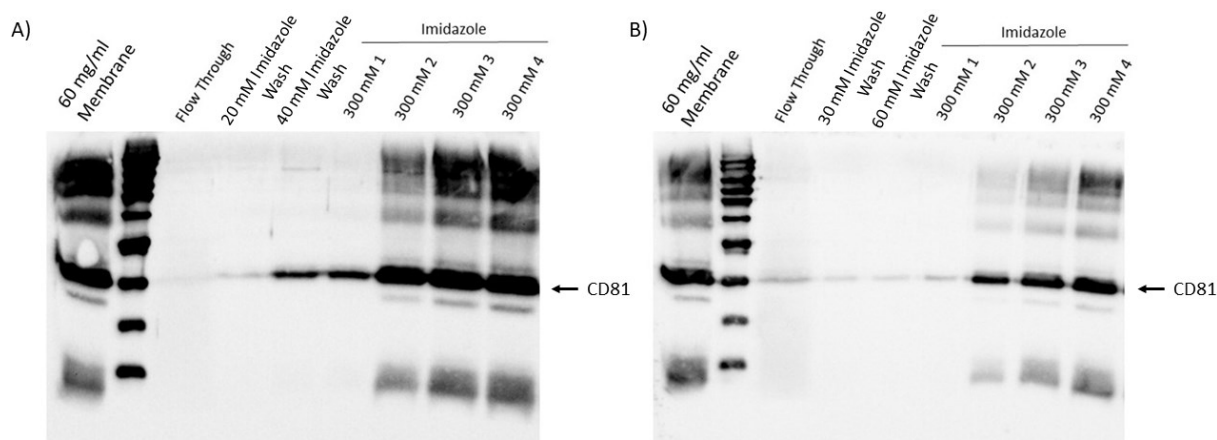


Figure 5.8. Western blot analysis of the optimised CD81-p-null purification. Small scale (3 ml membrane preparation at 60 mg/ml) solubilisation of CD81-p-null. Western blot of the original method (A) and the optimised method (B). 20 μ l of sample was loaded into each well and the nitrocellulose membrane was incubated with a 1:5000 dilution of 6x His monoclonal antibody (Takara) and then incubated with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signaling Technology). The membrane was then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene).

The final step in optimising the purification protocol for CD81-p-null was to ensure that all the protein was being eluted off the Ni-NTA resin. Until now CD81-p-null has been eluted with 300 mM imidazole but it is possible that a higher concentration of imidazole is required to elute all of the protein.

To investigate this the optimised purification method was performed but instead of just eluting the protein with 300 mM imidazole a further 500 mM imidazole step was added. An SDS-PAGE gel was run which included five 300 mM fractions and four 500 mM fractions (Figure 5.9). More protein was eluted in the 500 mM fractions, however, the densest band appeared in the fifth 300 mM fraction which means that a higher concentration than 300 mM imidazole is not required to elute all the protein. Increasing the volume of the 300 mM elution steps instead ensured that all the CD81-p-null that bound to the Ni-NTA resin was eluted.

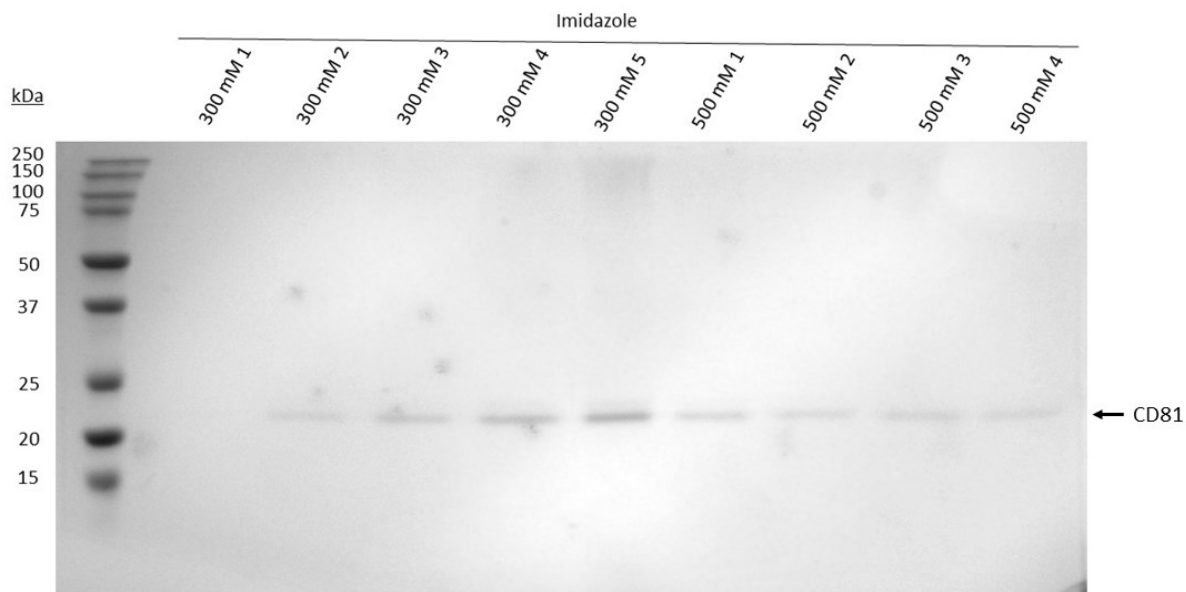


Figure 5.9. SDS-PAGE analysis of a small scale SMA2000 purification of CD81 using the optimised method with increased elution steps. Small scale (3 ml membrane preparation at 60 mg/ml) solubilisation of CD81-p-null. After using the optimised solubilisation and purification method elution steps were continued after the 300 mM imidazole elution to a concentration of 500 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

It has been shown that modifying the protocol for the purification of CD81-p-null to include 20 mM imidazole in the binding buffer and to increase the imidazole concentration to 30 mM and 60 mM in the two wash steps stopped non-specific binding to the Ni-NTA resin and yielded purer protein. Confirming that only small quantities of CD81-p-null were lost in the flow through means that any future purification was carried out using the modified protocol.

5.3 CD81-p-null Mutations for PELDOR Experiments

To investigate the change in CD81 conformation if cholesterol is present using PELDOR required the introduction of two spin labels in the protein so that the distance between the two could be measured. The spin labels are bonded to Cys residues. CD81 has six intracellular Cys residues but they have been mutated out in the CD81-p-null construct used to avoid palmitoylation (Jamshad et al., 2008). CD81 has four more Cys residues in the LEL which form disulfide bonds with each other (Kitadokoro et al., 2001), so they are not free Cys residues, therefore they will not be labelled with the spin label. CD81-p-null possesses two further native Cys residues, located in TM3 which would have to be mutated out to introduce a Cys residue elsewhere for spin labelling. To avoid needlessly mutating both Cys, if one is mutated then the other one can be spin labelled. With one spin label in the TM region, an additional Cys residue would have to be mutated into the LEL region so that the distance between the two spin labels can be measured using PELDOR.

Initial candidates for the introduction of a Cys residue in the LEL were Asp128, Val136 and Asp137 because these are all part of the LEL that have no known interaction with hepatitis C virus E2 protein. They are also part of the α helix which lies perpendicular to the membrane in the closed conformation and is expected to extend upwards away from the membrane when in the open conformation. Advice from Dr Christos Pliotas, an expert in EPR, was to mutate Asp128.

To produce the desired mutants of CD81 site directed mutagenesis was performed. In the first instance mutants with Cys97 and Cys104 mutated to alanine were created. Cys residues are typically mutated to either Ser or Ala. Mutations to alanine were done because serines may undergo unexpected post-translational modifications, such as phosphorylation or glycosylation. Ala, on the other hand, is only post-translationally modified when it is situated at the N-terminus.

Two separate mutants of CD81-p-null were created by mutating Cys97Ala and Cys104Ala separately. After site-directed-mutagenesis, the plasmids were run on an agarose gel (Figure 5.10) which shows a PCR product at approximately 4.0 kb. This corresponds with the size of the 3.3 kb pPICZB plasmid and the 708 bp CD81-p-null gene.

Mock mutagenesis experiments were also run and digested with Dpn1 and, as expected, no PCR product is visible on the agarose gel (Figure 5.10). The presence of two bands for each of the mutagenesis experiments is surprising because the expectation is that any templates that have not been mutated will be removed by Dpn1. Nevertheless, with bands visible at the appropriate size the next step was the transformation of the mutated plasmids into competent *E. coli* cells.

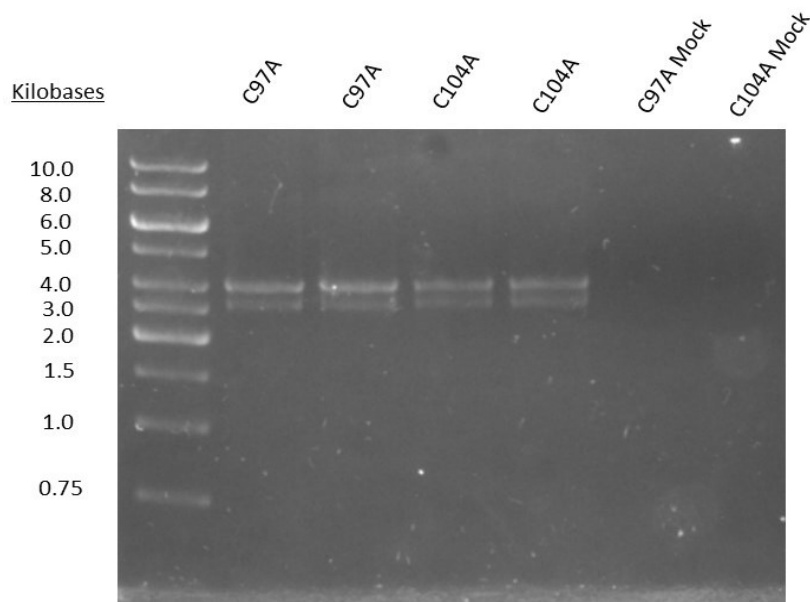


Figure 5.10. Agarose gel of PCR product after site-directed mutagenesis of CD81-p-null to create single cysteine mutants. pPICZB plasmids containing CD81-p-null were mutated by site-directed mutagenesis to produce single cysteine mutants with just Cys97 or Cys104 remaining, except the cysteines in the LEL. 6 μ l of each sample loaded onto the agarose gel, along with mock mutations as controls to ensure that Dpn1 digestion removes non-mutated plasmids.

The mutants that had been created were transformed into TOP10 *E. coli* cells, followed by a mini-prep so that a sample could be sent to Eurofins Genomics for sequencing. Sequencing would ensure that the correct mutation had been introduced. The nucleotide sequence was analysed using ExPasy's Translate tool so that the protein sequence could be determined. The frame containing the desired sequence could easily be located in six frames because of the 6x His tag at the end of CD81-p-null at the N-terminal (Jamshad et al., 2008). Once the correct amino acid sequence was found the sequence was compared to the wild-type CD81 sequence to ensure that the desired mutation had been introduced. Using Clustal Omega to align the sequence it was shown that C97A (Figure 5.11A) and C104A (Figure 5.11B) mutants had been created.

A)

CD81-p-null-C97A	MSGVEGATKAIKYLLFVFNFWLGGVILGVALWLRHDPQTNNLLYLELGDKPAPNTFY	60
sp P60033 CD81_HUMAN	-MGVEGCTKCIKYLFLVFNFWLGGVILGVALWLRHDPQTNNLLYLELGDKPAPNTFY	59
	**** * ..	
CD81-p-null-C97A	VGIYILIAVGAVMMFVGFGLGAYGAIQESQALLGTFFTALVILFACEVAAGINGFVNKDQI	120
sp P60033 CD81_HUMAN	VGIYILIAVGAVMMFVGFGLGAYGAIQESQALLGTFFTCLVILFACEVAAGINGFVNKDQI	119
	***** . ***** . *****	
CD81-p-null-C97A	AKDVKQFYDQALQQAVVDDANNKAVVKTFFHETLDCCGSSTLTALTTSVLKNNLCPSGS	180
sp P60033 CD81_HUMAN	AKDVKQFYDQALQQAVVDDANNKAVVKTFFHETLDCCGSSTLTALTTSVLKNNLCPSGS	179

CD81-p-null-C97A	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLAAGIRNSSVYGGG	240
sp P60033 CD81_HUMAN	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLCGIRNSSVY---	236
	***** .. *****	
CD81-p-null-C97A	HHHHHH	246
sp P60033 CD81_HUMAN	-----	236

B)

CD81-p-null-C104A	MSGVEGATKAIKYLLFVFNFWLGGVILGVALWLRHDPQTNNLLYLELGDKPAPNTFY	60
sp P60033 CD81_HUMAN	-MGVEGCTKCIKYLFLVFNFWLGGVILGVALWLRHDPQTNNLLYLELGDKPAPNTFY	59
	**** * ..	
CD81-p-null-C104A	VGIYILIAVGAVMMFVGFGLGAYGAIQESQALLGTFFTCLVILFAAEVAAGINGFVNKDQI	120
sp P60033 CD81_HUMAN	VGIYILIAVGAVMMFVGFGLGAYGAIQESQALLGTFFTCLVILFACEVAAGINGFVNKDQI	119
	***** . ***** . *****	
CD81-p-null-C104A	AKDVKQFYDQALQQAVVDDANNKAVVKTFFHETLDCCGSSTLTALTTSVLKNNLCPSGS	180
sp P60033 CD81_HUMAN	AKDVKQFYDQALQQAVVDDANNKAVVKTFFHETLDCCGSSTLTALTTSVLKNNLCPSGS	179

CD81-p-null-C104A	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLAAGIRNSSVYGGG	240
sp P60033 CD81_HUMAN	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLCGIRNSSVY---	236
	***** .. *****	
CD81-p-null-C104A	HHHHHH	246
sp P60033 CD81_HUMAN	-----	236

Figure 5.11. Single cysteine mutant sequences compared to human CD81. Transformed mutants were sequenced by Sanger sequencing with a sample of 15 µl (50 ng/µl) with 2 µl of forward AOX1 primer (10 pmol/µl). The sequences were translated using ExPasy's translate tool and the sequences were aligned using Clustal Omega. The desired mutations, (A) C97A and (B) C104A are shown in the boxes.

With the single mutants successfully created these mutants were taken and mutated further to create the double mutants that include the D128C mutation so that the protein can be purified and used in PELDOR experiments. After site-directed mutagenesis the transformations of the double mutants were successful (Figure 5.12) when they were transformed into E. coli TOP10 cells on LB agar plates which contained a 100 µg/ml zeocin concentration. For further validation there was no growth on the mock transformation plates.

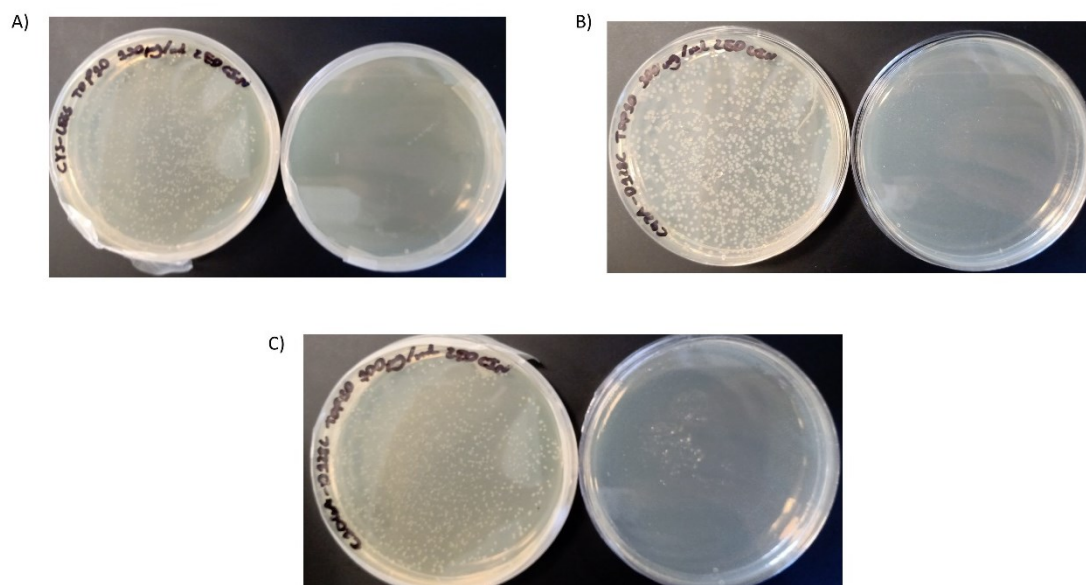


Figure 5.12. Low salt LB agar plates with transformed cells after site-directed mutagenesis. pPICZB plasmids containing CD81-p-null were mutated by site-directed mutagenesis to produce (A) Cys-less, (B) C97A-D128C and (C) C104A-D128C mutants. 150 μ l of cells after transformation and recovery step were plated. Mock transformations were performed.

The three mutated proteins represented a double C97A-C104A mutant referred to as Cys-less because it has all free Cys residues removed. The only remaining Cys residues all form disulfide bonds in the LEL and are not relevant for EPR studies. The other two double mutants were the C97A-D128C and C104A-D128C mutants which can be used to investigate conformational change in CD81. All three mutants were confirmed to be successful when they were mini-prepped and sent for sequencing at Eurofins Genomics (Figure 5.13).

A)	CD81-p-null-C97A-C104A sp P60033 CD81_HUMAN	MSGVEGATKAIKYL L FVFN FVFW LAGGVILGVALWLRHDPQT TNL L YLELGDKPAPNTFY -MGVEGCTKCIKYL L FVFN FVFW LAGGVILGVALWLRHDPQT TNL L YLELGDKPAPNTFY *****	60 59
	CD81-p-null-C97A-C104A sp P60033 CD81_HUMAN	VGIYILIAVGAVMMFVGLGAYGAIQESQALLGTFFTALVILFAEVAAGINGFVNKDQI VGIYILIAVGAVMMFVGLGCGYAIQESQCLLGTFFTCVLILFAEVAAGINGFVNKDQI *****	120 119
	CD81-p-null-C97A-C104A sp P60033 CD81_HUMAN	AKDVKQFYDQALQQAVVDDANNKAVVKT FHETLDCCGSSTLTAL TTSVLKNNLCPSGS AKDVKQFYDQALQQAVVDDANNKAVVKT FHETLDCCGSSTLTAL TTSVLKNNLCPSGS *****	180 179
	CD81-p-null-C97A-C104A sp P60033 CD81_HUMAN	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIWAVIMIFEMILSHVLAAGIRNSSVYGGG NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIWAVIMIFEMILSHVLCGIRNSSVY--- *****	240 236
	CD81-p-null-C97A-C104A sp P60033 CD81_HUMAN	HHHHHH -----	246 236

B)	CD81-p-null-C97A-D128C	MSGVEGATKAIKYL L F V F N F V F W L A G G V I L G V A L W L R H D P Q T T N L L Y L E L G D K P A P N T F Y	60
	sp P60033 CD81_HUMAN	-MGEVGTCKIKYL L F V F N F V F W L A G G V I L G V A L W L R H D P Q T T N L L Y L E L G D K P A P N T F Y	59

	CD81-p-null-C97A-D128C	VGIYILIAVGAVMMFVGF L G Y G A I Q E S Q A L L G T F F T A L V I L F A C E V A A G I N G F V N K D Q I	120
	sp P60033 CD81_HUMAN	VGIYILIAVGAVMMFVGF L G C Y G A I Q E S Q C L L G T F F T C L V I L F A C E V A A G I N G F V N K D Q I	119

CD81-p-null-C97A-D128C	AKDVKQFYD Q A L Q Q A V V D D A N N A K A V V K T F H E T L D C C G S S T L T A L T T S V L K N N L C P S G S	180	
sp P60033 CD81_HUMAN	AKDVKQFYD Q A L Q Q A V V D D A N N A K A V V K T F H E T L D C C G S S T L T A L T T S V L K N N L C P S G S	179	

CD81-p-null-C97A-D128C	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLAAGIRNSSVYGGG	240	
sp P60033 CD81_HUMAN	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLCGIRNSSVY---	236	

CD81-p-null-C97A-D128C	HHHHHH	246	
sp P60033 CD81_HUMAN	-----	236	
C)	CD81-p-null-C104A-D128C	MSGVEGATKAIKYL L F V F N F V F W L A G G V I L G V A L W L R H D P Q T T N L L Y L E L G D K P A P N T F Y	60
	sp P60033 CD81_HUMAN	-MGEVGTCKIKYL L F V F N F V F W L A G G V I L G V A L W L R H D P Q T T N L L Y L E L G D K P A P N T F Y	59

	CD81-p-null-C104A-D128C	VGIYILIAVGAVMMFVGF L G Y G A I Q E S Q A L L G T F F T C L V I L F A E V A A G I N G F V N K D Q I	120
	sp P60033 CD81_HUMAN	VGIYILIAVGAVMMFVGF L G C Y G A I Q E S Q C L L G T F F T C L V I L F A C E V A A G I N G F V N K D Q I	119

CD81-p-null-C104A-D128C	AKDVKQFYD Q A L Q Q A V V D D A N N A K A V V K T F H E T L D C C G S S T L T A L T T S V L K N N L C P S G S	180	
sp P60033 CD81_HUMAN	AKDVKQFYD Q A L Q Q A V V D D A N N A K A V V K T F H E T L D C C G S S T L T A L T T S V L K N N L C P S G S	179	

CD81-p-null-C104A-D128C	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLAAGIRNSSVYGGG	240	
sp P60033 CD81_HUMAN	NIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLCGIRNSSVY---	236	

CD81-p-null-C104A-D128C	HHHHHH	246	
sp P60033 CD81_HUMAN	-----	236	

Figure 5.13. Double cysteine mutant sequences compared to human CD81. Transformed mutants were sequenced by Sanger sequencing with a sample of 15 μ l (50 ng/ μ l) with 2 μ l of forward AOX1 primer (10 pmol/ μ l). The sequences were translated using ExPasy's translate tool and the sequences were aligned using Clustal Omega. The desired mutations are shown in the boxes for (A) Cys-less (B) C97A-D128C and (C) C104A-D128C.

All five mutations were chemically transformed into *P. pastoris* X33 cells and grown on YPD agar plates with varying concentration of zeocin (50 μ g/ml, 100 μ g/ml, 250 μ g/ml and 500 μ g/ml). There were 3-5 colonies for each mutation that grew on the 250 μ g/ml and 500 μ g/ml zeocin plates, so these colonies were picked, grown in YPD overnight and a glycerol stock of each was made. They were then tested for protein expression.

To test expression of the mutants created the *P. pastoris* cells that had been transformed with each of the mutants were grown on a small scale. Cells were streaked on YPD agar containing 100 μ g/ml of zeocin and colonies were grown in 5 ml of BMGY overnight at 30°C and 200 rpm,

followed by centrifugation to remove all media. The cells were resuspended in 5 ml of BMMY and induced at 30°C and 200 rpm.

All the mutants tested (C97A, C97A-D128C, C104A and C104A-D128C) were expressing CD81. The C97A and the C97A-D128C mutants looked similar when analysing them on a western blot. Both of them had their densest band at approximately 100 kDa with a fainter band at approximately 40-50 kDa. CD81 has a molecular mass of 25 kDa and it is known to form oligomers, therefore it is likely that the bands visible at approximately 40-50 kDa are dimers and the bands visible at approximately 100 kDa are tetramers (Figure 5.14). CD81-p-null that had not had neither C97 nor D128 mutated was run on the same Western as a control. It was seen that CD81-p-null produces an intense band at around 20 kDa. This band is more intense than that seen for the C97A (Fig. 5.14A) and C97A-D128C mutants (Figure 5.14B).

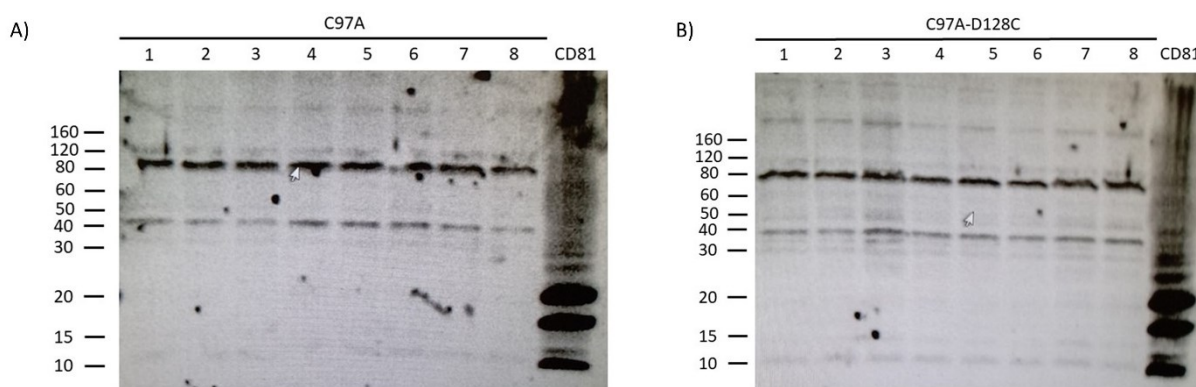


Figure 5.14. Western blot analysis of CD81-p-null-C97A and CD81-p-null-C97A-D128C small scale expression. Small scale expression of the (A) CD81-p-null-C97A and (B) CD81-p-null-C97A-D128C in 5 ml of BMGY, followed by induction in 5 ml BMMY. 20 μ l of sample was loaded into each well and the nitrocellulose membrane was incubated with a 1:5000 dilution of 6x His monoclonal antibody (Takara) and then incubated with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signaling Technology). The membrane was then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene).

The C104A and C104A-D128C mutants exhibited much better expression than that seen for C97A and C97A-D128C. For C104A an intense band, comparable with the intensity of the Cd81-p-null control, was visible at approximately 80-100 kDa in five (lane 1-5) of the eight colonies tested. Three of the colonies (lane 6-8) has little to no expression. The molecular mass of the bands suggests that the C104A mutant is forming tetramers (Figure 5.15A).

The C104A-D128C mutant also exhibited similar intensity to CD81-p-null. The dense visible band, however, was seen at approximately 50 kDa in seven of the colonies tested (lane 1-7)

with no 50 kDa band visible in lane 8. An intense band at approximately 50kDa suggests that the C104A-D128C mutant is forming dimers (Figure 5.15B).

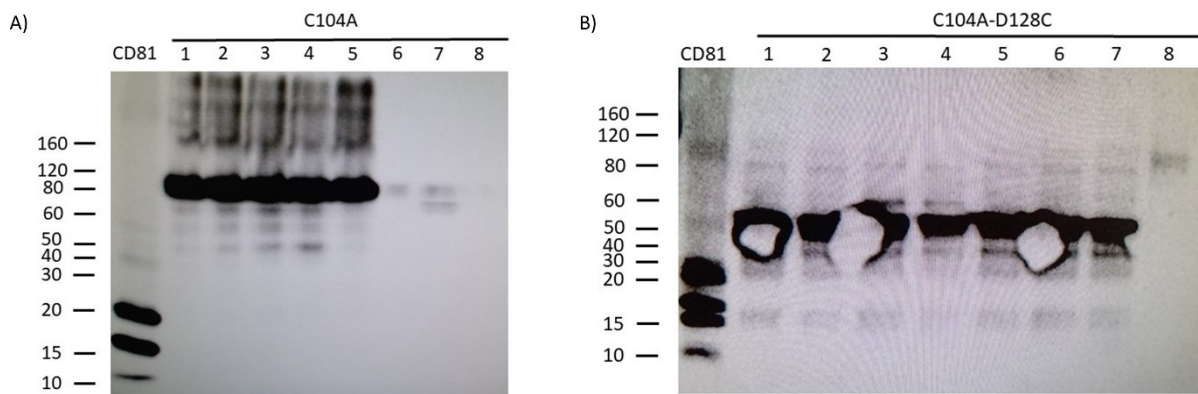


Figure 5.15. Western blot analysis of CD81-p-null-C104A and CD81-p-null-C104A-D128C small scale expression. Small scale expression of the (A) CD81-p-null-C104A and (B) CD81-p-null-C104A-D128C in 5 ml of BMGY, followed by induction in 5 ml BMMY. 20 μ l of sample was loaded into each well and the nitrocellulose membrane was incubated with a 1:5000 dilution of 6x His monoclonal antibody (Takara) and then incubated with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signaling Technology). The membrane was then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene).

Given that the double mutants are required for PELDOR experimentation because two Cys residues need to be spin labelled, they were taken forward and grown on a larger scale to test expression in medium scale growth. Colonies of C97A-D128C and C104A-D128C were grown in 50 ml BMGY overnight at 30°C and 200 rpm, followed by centrifugation to remove all the media. The cell pellet was resuspended in 200 ml of BMMY and grown overnight at 30°C and 200 rpm, followed by cell harvesting.

Medium scale growth exacerbated the differences previously seen between the C97A-D128C and C104A-D128C during small scale growth. The C97A-D128C mutants expressed less than the C104A-D128C mutants when they were analysed on a Western blot with an anti-His antibody. The C97A-D128C mutants were either not expressing or not expressing to the same level as CD81-p-null (Figure 5.16A). Similar to small scale expression, when the C97A-D128C mutant was expressing the band was visible at approximately 80-100 kDa.

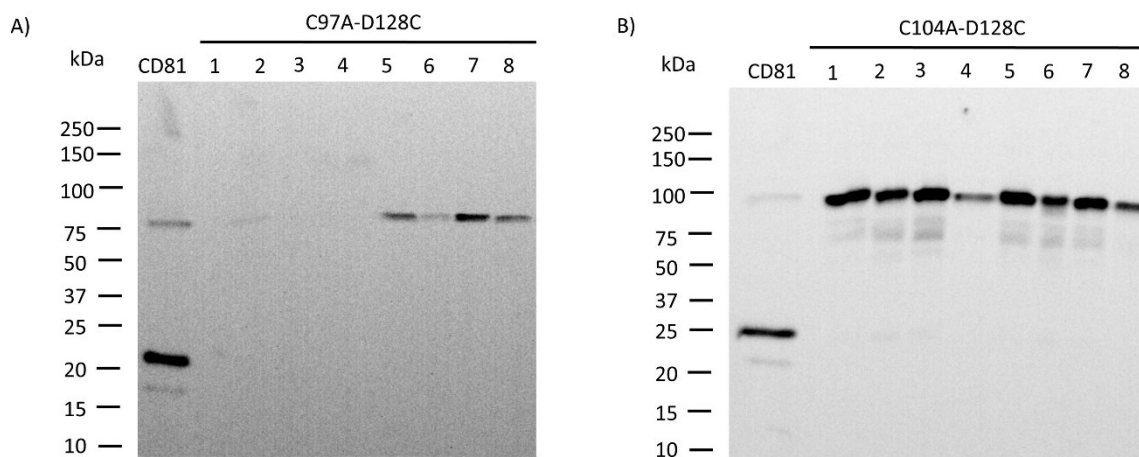


Figure 5.16. Western blot analysis of CD81-p-null-C97A-D128C and CD81-p-null-C104A-D128C medium scale expression. Medium scale expression of the (A) CD81-p-null-C97A-D128C and (B) CD81-p-null-C104A-D128C in 50 ml of BMGY, followed by induction in 200 ml BMMY. 20 μ l of sample was loaded into each well and the nitrocellulose membrane was incubated with a 1:5000 dilution of 6x His monoclonal antibody (Takara) and then incubated with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signaling Technology). The membrane was then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene).

Expression of the C104A-D128C was better than C97A-D128C expression. All the C104A-D12C mutants were expressing the protein with a dense band visible at approximately 100 kDa which means that this mutant may be favouring the formation of tetramers (Figure 5.16B). Nevertheless, this C104A-D128C mutant was taken forward for further experiments.

5.4 Purification of CD81-p-null-C104A-D128C Double Mutant

With both single mutants (C97A and C104A) and both double mutants (C97A-D128C and C104A-D128C) expressing, albeit with highest expression at weights that would suggest oligomeric formations, the focus was on purifying the C104A-D128C double mutant. The aim was to purify enough of one of the double mutants to be able to attach spin labels to it and perform EPR to see whether CD81 does open and close depending on whether cholesterol is bound to the cavity in the TM region as suggested by molecular modelling (Zimmermann et al., 2016). The focus was on the C104A-D128C double mutants because expression of this mutant was better when grown on a medium scale than expression of C97A-D128C. Another benefit of choosing the C104A mutation is that it is further away from the Asn18 and Glu219 cholesterol binding site, therefore this mutation is less likely to have an impact on cholesterol binding which is vital because it is the effect of cholesterol binding on CD81 conformation that will be investigated. Furthermore, the expression levels of the C104A-D128C mutant were

much higher in both small scale and medium scale expression when compared to the C97A-D128C mutant.

The first attempt at purifying the C104A-D128C double mutant was performed using the same conditions that were optimised for CD81-p-null. It was solubilised in 2.5% (w/v) SMA2000 and incubated with Ni-NTA resin in binding buffer containing 20 mM imidazole overnight at 4°C. The next day the resin was washed with 30 mM imidazole and 60 mM imidazole before eluting the protein with 300 mM imidazole.

The attempt to purify the C104A-D128C double mutants under these conditions proved unsuccessful (Figure 5.17). There are no bands visible on the SDS-PAGE gel in the lanes containing the 300 mM imidazole elutions and only faint bands can be seen in the lane containing the 30 mM imidazole wash step. This is in stark contrast to the dense bands in the first wash steps of all the purifications performed when trying to optimise the purification conditions of CD81-p-null. The bands in the first wash step are likely to be an indication the double mutant is not binding sufficiently to the Ni-NTA resin.

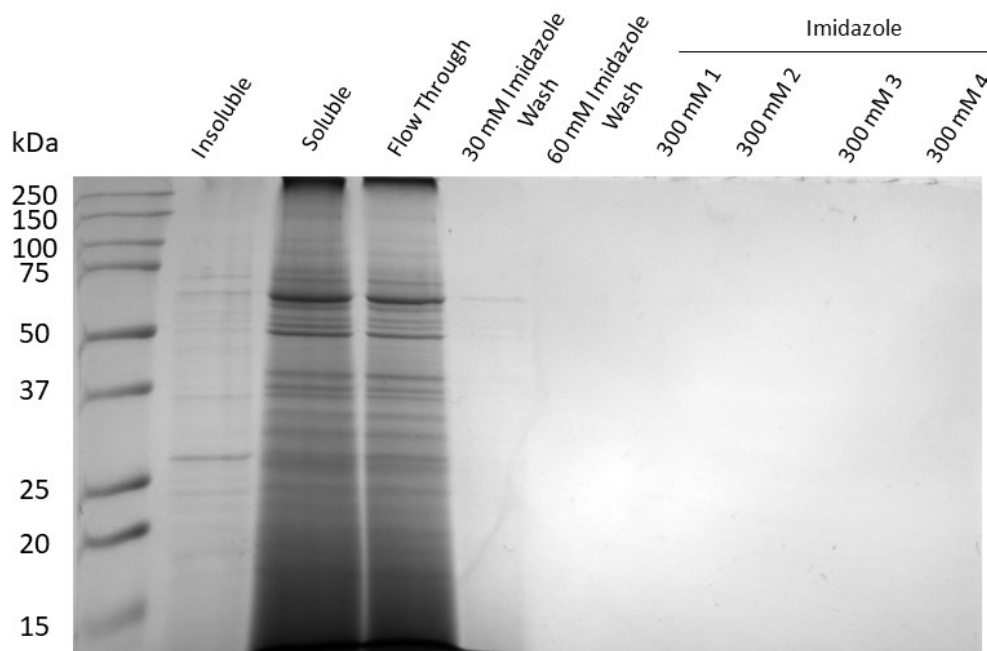


Figure 5.17. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null-C104A-D128C expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using the optimised purification protocol for CD81-p-null with 30 mM and 60 mM imidazole in the wash steps followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

One of the problems that can arise when trying to purify membrane proteins with SMA2000 is that any excess SMA2000 can interfere with the binding of the protein to the Ni-NTA resin. Previously the concentration of the membrane preparation when solubilising with 2.5% (w/v) SMA2000 was 60 mg/ml. To try and reduce the amount of excess SMA2000 the membrane preparation concentration was doubled to 120 mg/ml so that there will be more protein for any SMA2000 to solubilise, thus reducing the amount of excess SMA2000.

The C104A-D128C double mutants was purified under the same conditions as the first attempted purification except for the higher membrane preparation concentration. The SDS-PAGE gel looks like the first attempted purification with no significant amounts of protein visible in the 300 mM imidazole elutions (Figure 5.18). Furthermore, the same faint bands can be seen in the 30 mM imidazole wash step which again may be because the protein is not binding to the Ni-NTA resin.

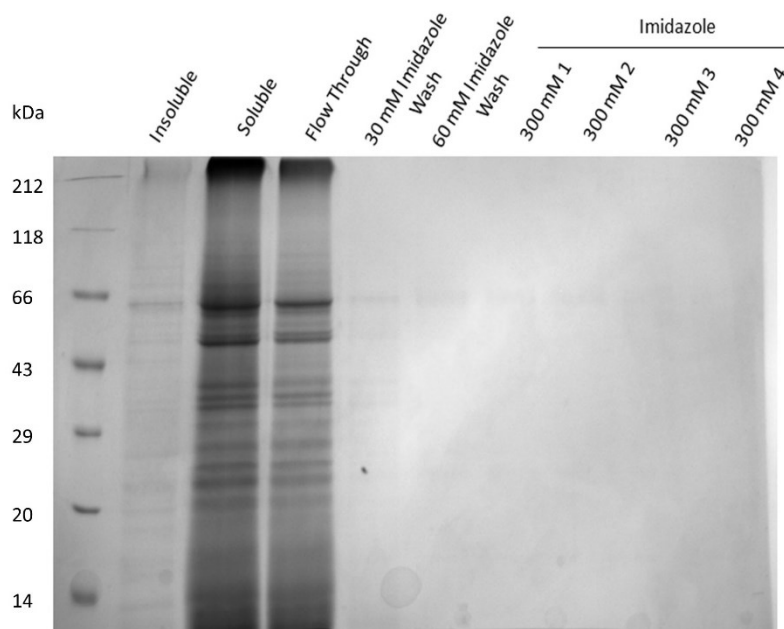


Figure 5.18. SDS-PAGE analysis of a small scale SMA2000 purification of CD81-p-null-C104A-D128C expressed in *P. pastoris* solubilised at 120 mg/ml membrane concentration. A small scale (3 ml of membrane preparations at 120 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using the optimised purification protocol for CD81-p-null with 30 mM and 60 mM imidazole in the wash steps followed by elution of the protein with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

Further attempts to try and purify the C104A-D128C double mutant were done with a different brand of Ni-NTA resin. A Qiagen and a Generson resin were both used with 2.5% (w/v) SMA2000 and 1.25% (w/v) SMA2000 used for solubilisation. The 2.5% (w/v) SMA2000 concentration is the same as previously used for all other purification attempts, while the

1.25% (w/v) SMA2000 concentration is another way to try and limit the amount of excess SMA2000 that may be an impact on the binding of the proteins His-tag to the Ni-NTA resin.

Neither the 2.5% (w/v) SMA2000 (Figure 5.19A), nor the 1.25% (w/v) SMA2000 (Figure 5.19B) Qiagen purifications did anything to improve the purification. No bands are visible in any of the 300 mM imidazole elutions for either of the purifications and, like before, there are only faint bands visible in the 30 mM imidazole wash steps. Although, a faint band can be seen at the correct weight of approximately 25 kDa in the 30 mM imidazole wash step of the purification that was solubilised using 1.25% (w/v) SMA2000.

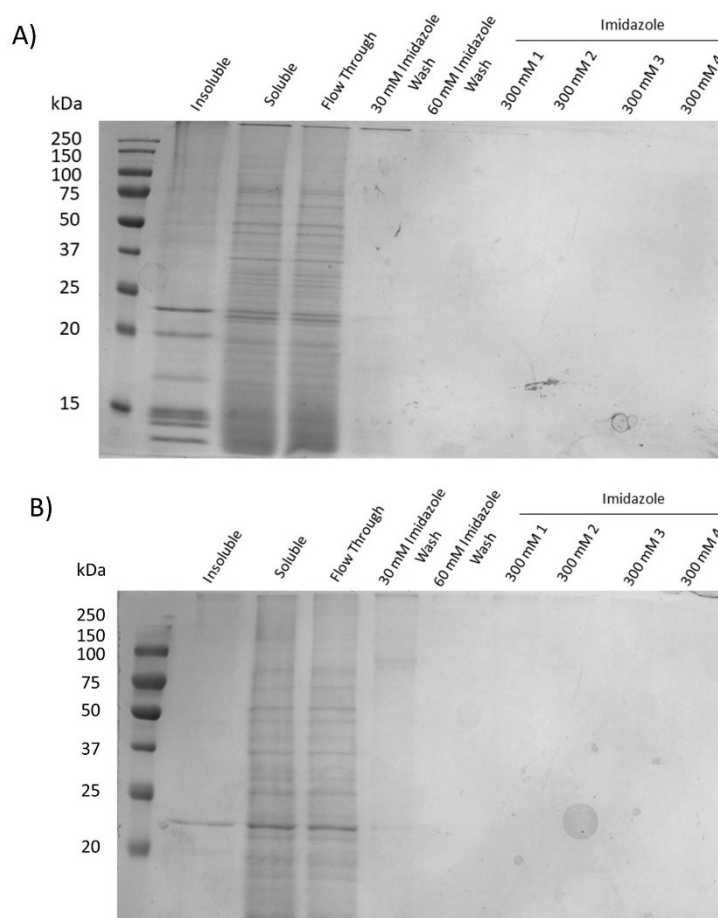


Figure 5.19. SDS-PAGE analysis of a small scale SMA2000 purification with Qiagen Ni-NTA resin of CD81-p-null-C104A-D128C expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using the optimised purification protocol for CD81-p-null with 30 mM and 60 mM imidazole in the wash steps followed by elution of the protein with 300 mM imidazole. (A) CD81-p-null-C104A-D128C solubilised with 2.5% (w/v) SMA2000 and incubated with Qiagen Ni-NTA resin. (B) CD81-p-null-C104A-D128C solubilised with 1.25% (w/v) SMA2000 and incubated with Qiagen Ni-NTA resin. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

The Generon Ni-NTA resin did not perform better than the Qiagen Ni-NTA resin. The Generon Ni-NTA resin in conjunction with 2.5% (w/v) SMA2000 (Figure 5.20A) for solubilisation produced the same results that had been seen throughout the purification process. When solubilising with 1.25% (w/v) SMA2000 (Figure 5.20B), however, there were slightly darker bands visible in the 30 mM imidazole wash but still no bands in the elution steps.

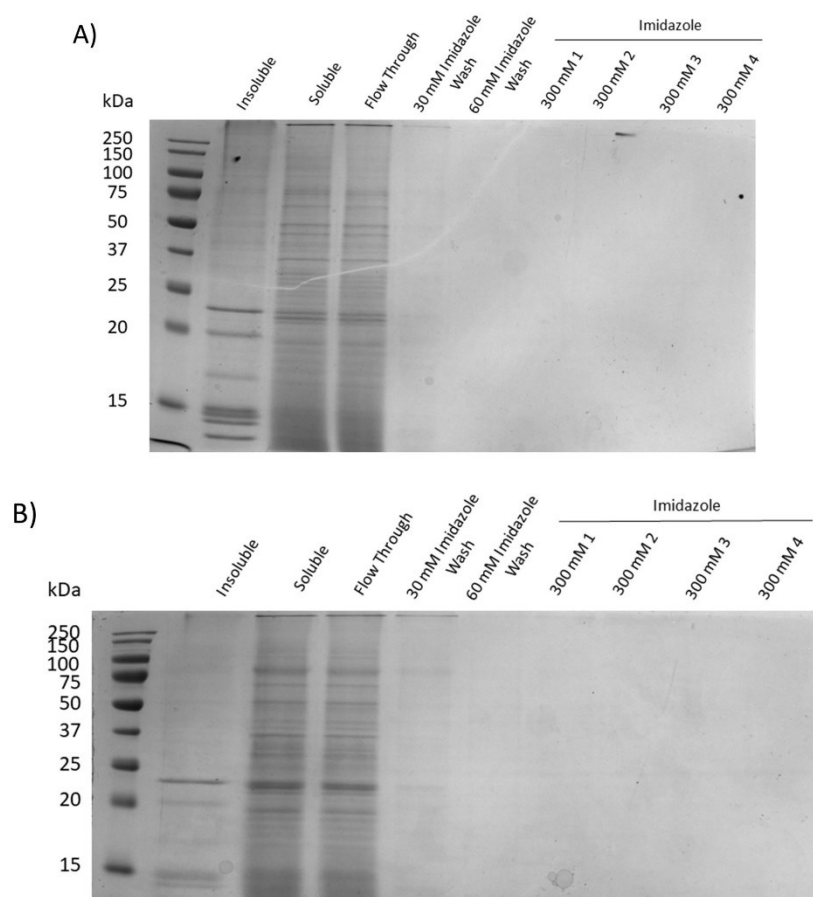


Figure 5.20. SDS-PAGE analysis of a small scale SMA2000 purification with Generon Ni-NTA resin of CD81-p-null-C104A-D128C expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using the optimised purification protocol for CD81-p-null with 30 mM and 60 mM imidazole in the wash steps followed by elution of the protein with 300 mM imidazole. (A) CD81-p-null-C104A-D128C solubilised with 2.5% (w/v) SMA2000 and incubated with Generon Ni-NTA resin. (B) CD81-p-null-C104A-D128C solubilised with 1.25% (w/v) SMA2000 and incubated with Generon Ni-NTA resin. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

The aim had been to purify C104A-D128C double mutant after solubilising the protein in SMA2000 so that some lipids, albeit from *P. pastoris*, would be surrounding the protein to give a better model of how the protein would be *in vivo*. Attempts to solubilise the protein with SMA, however, did not work, despite trying a variety of conditions.

The next step was to try and purify the C104A-D128C double mutant by solubilising in DDM. DDM solubilisation was done by adding the membrane preparation to solubilisation buffer containing 1% (w/v) DDM.

The first purification attempt using DDM was better than previous attempts using SMA2000 (Figure 5.21). There were still no dense bands in the elution steps, nor in the first wash step. There were, however, slightly darker and a greater number of bands visible in the first wash step which is an indication that the protein is binding to the resin better after being solubilised with DDM. Despite a lack of dense bands in the elution steps there are some faint bands visible which indicates that it would be better to pursue solubilising the C104A-D128C double with DDM.

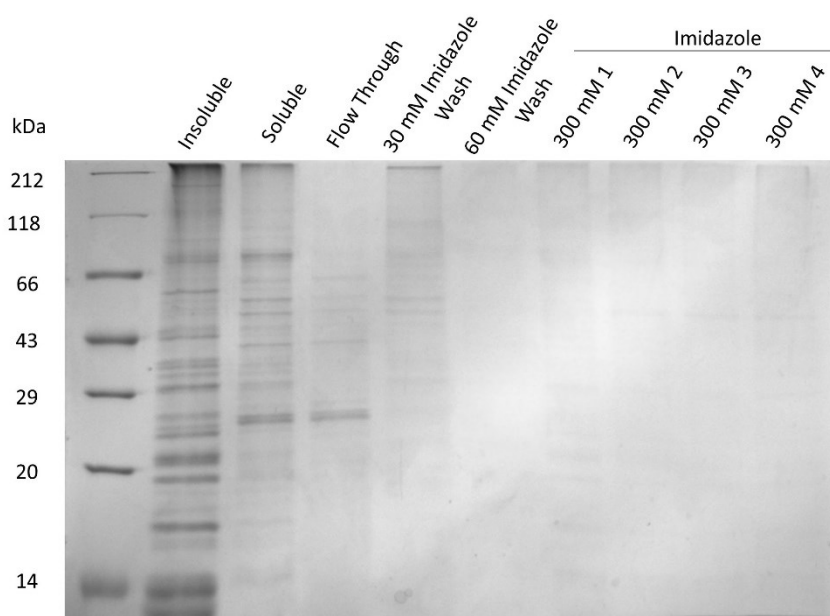


Figure 5.21. SDS-PAGE analysis of a small scale DDM purification of CD81-p-null-C104A-D128C expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using 1% (w/v) DDM to solubilise the protein. The Ni-NTA resin was washed with 30 mM and 60 mM imidazole followed by elution with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

When solubilising with SMA2000 the bands on the SDS-PAGE gels in the soluble fraction were normally darker and more numerous. Solubilising with DDM had the reverse effect with much of the protein content appearing in the insoluble fraction (Figure 5.21). To try and overcome this the volume of the solubilisation buffer containing 1% (w/v) DDM was doubled so that more protein could be solubilised.

The result is an insoluble fraction with fewer and less dense bands in it (Figure 5.22). There is little visible in the soluble fraction but this can be attributed to the fact that the volume of the

solubilisation buffer has been doubled, therefore, when taking a small 20 µl sample from a 30 ml solubilisation step the sample will be more dilute than when the solubilisation step was done in a smaller volume.

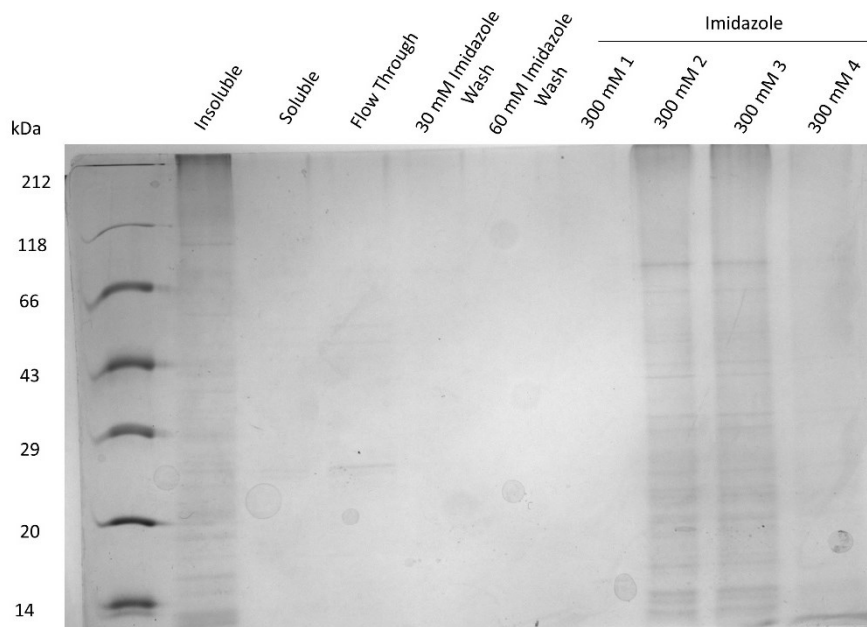


Figure 5.22. SDS-PAGE analysis of a small scale DDM purification with double the volume of solubilisation buffer of CD81-p-null-C104A-D128C expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using 1% (w/v) DDM to solubilise the protein. Double the amount of solubilisation buffer was used compared with Fig. 5.16. The Ni-NTA resin was washed with 30 mM and 60 mM imidazole followed by elution with 300 mM imidazole. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

There is more protein in the elution steps than any other attempt to purify the C104A-D128C double mutant (Figure 5.22). There are clear visible bands in the second and third elution fractions but those bands have almost completely disappeared in the fourth elution fraction. The presence of bands in the second and third fraction but not the fourth fraction is probably because there was not an abundance of protein binding to the Ni-NTA resin with almost all of it coming off the resin after just three elution steps.

5.5 Small Scale Size Exclusion Chromatography of CD81-C104A-D128C

Despite switching the solubilisation method from SMA2000 to DDM it still proved difficult to purify enough protein. Given the problems faced using Ni-NTA resin to exploit the His-tag on the protein a change in purification strategy was needed. It is possible that the His-tag is being obscured in some way, either by the excess SMA2000 and DDM, although it would be unlikely in the case of DDM, or the folding of the protein has buried the His-tag so it is not as accessible.

The next step was to try to purify the protein without the use of the His-tag because of the possible problems highlighted with its use. One possible method to use is to go straight to small scale size exclusion chromatography. Using a 1 ml size exclusion column all the DDM solubilised membrane preparation was loaded onto the column and eluted with 3 ml of solubilisation buffer.

In the first instance the solubilised membrane was concentrated to 500 μ l to be loaded onto the size exclusion column. It was eluted with 3 ml of solubilisation buffer and collected in six 500 μ l fractions. The chromatogram of the purification shows a jagged line with protein starting to be detected at just after the 2.25 ml mark because the first 2 ml of the chromatogram is the equilibration step (Figure 5.23). A jagged line with sharp ups and downs is not unexpected because all solubilised protein was loaded onto the column since no purification step had been done before size exclusion chromatography.

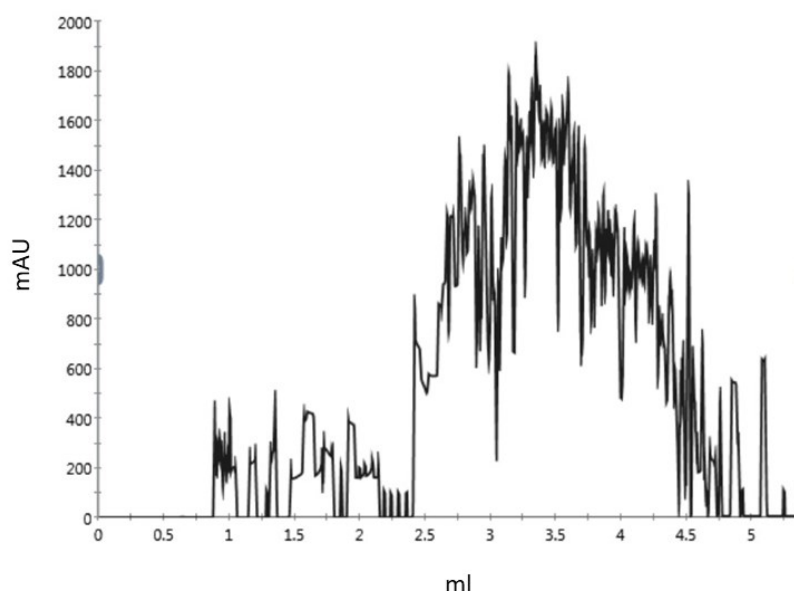


Figure 5.23. Size exclusion chromatogram of CD81-p-null-C104A-D128C concentrated to 500 μ l. A small scale (3 ml membrane preparation at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using 1% (w/v) DDM to solubilise the protein. After solubilisation the protein was concentrated to 500 μ l before loading onto the SEC column.

All six 500 μ l fractions were loaded onto an SDS-PAGE gel and in the second and third fraction faint bands can be seen (Figure 5.24). These faint bands and the lack of bands in the other fractions despite a clear signal on the chromatogram throughout the whole elution step could be because only a small amount of solubilised protein was loaded onto the column and the fraction size of 500 μ l may be too large and diluting the protein too much.

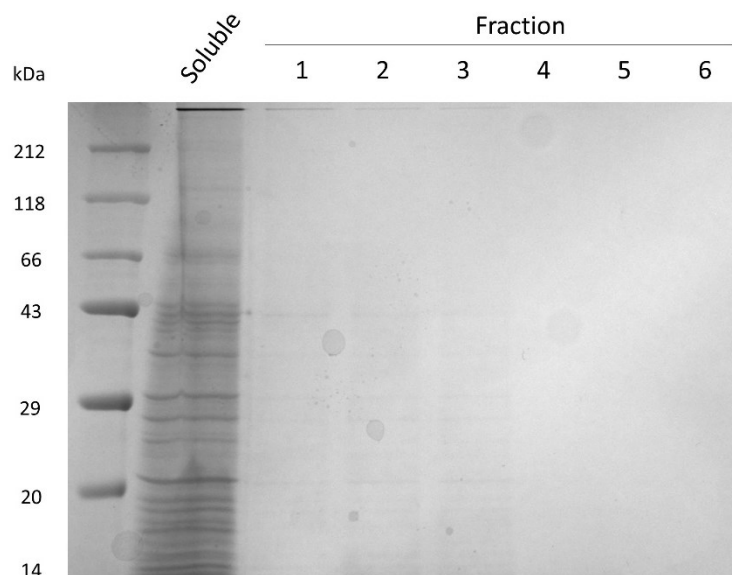


Figure 5.24. SDS-PAGE analysis of a small scale size exclusion chromatography of CD81-p-null-C104A-D128C expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using 1% (w/v) DDM to solubilise the protein. The protein was concentrated to 500 μ l before loading on the SEC column. The protein was eluted in 500 μ l fractions and run on the SDS-PAGE gel. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

To see if the protein samples are being diluted too much in a 500 μ l fraction the experiment was repeated with a smaller fraction volume of 200 μ l. In addition to this change the solubilised membrane was concentrated to 200 μ l before being loaded onto the column to try to improve the resolution on the chromatogram and then eluted with 3 ml of solubilisation buffer.

Concentrating the solubilised membrane into a smaller volume to make it more concentrated had a significant effect on the chromatogram. The protein began eluting off the column at the same place as before at just after the 2.25 ml mark with a sharp peak visible at this point (Figure 5.25). After that the chromatogram takes on a different shape, almost like it is one continuous peak with slight deviations up and down with less of a jagged look which is in stark contrast to the previous chromatogram.

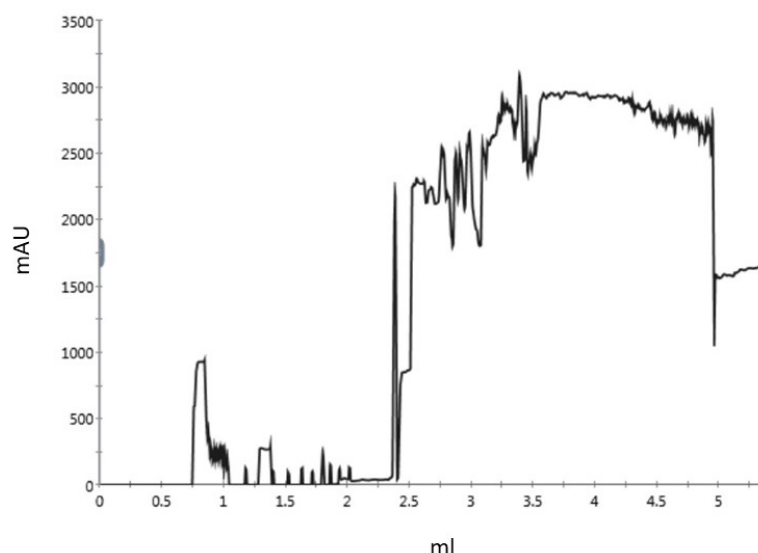


Figure 5.25. Size exclusion chromatogram of CD81-p-null-C104A-D128C concentrated to 200 μ l. A small scale (3 ml membrane preparation at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using 1% (w/v) DDM to solubilise the protein. After solubilisation the protein was concentrated to 200 μ l before loading onto the SEC column and eluted off the column in 200 μ l fractions.

The protein was eluted off the column in 200 μ l fractions and 16 fractions were run on an SDS-PAGE gel. There are clear visible bands that can be seen, particularly in fractions 3-6 with fraction 3 containing a darker band at approximately 250 kDa and this band fades as the fractions increase (Figure 5.26). At the same time fractions 4-6 see an increase in the visibility of proteins at lower molecular weights. The darkest band visible in fractions 4-6 is located at approximately 25 kDa which corresponds with the molecular weight of CD81, therefore it is likely that this band represents the protein of interest.

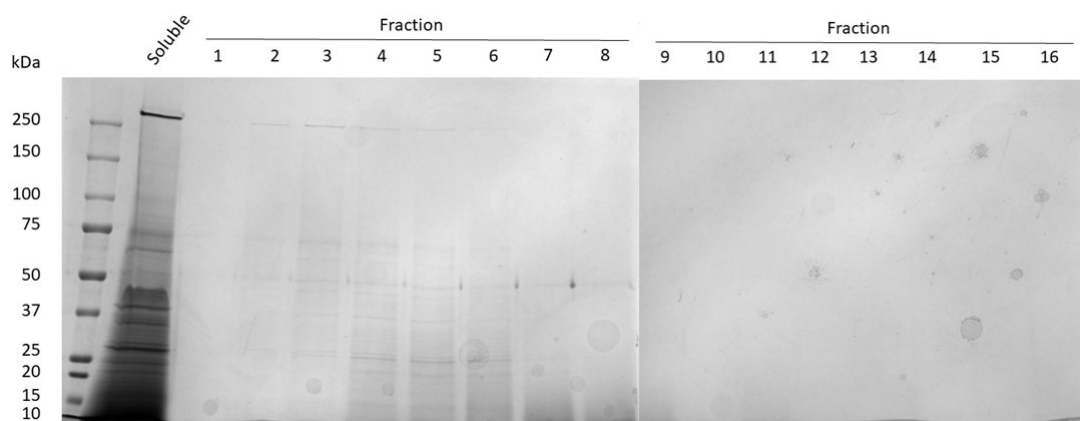


Figure 5.26. SDS-PAGE analysis of a small scale size exclusion chromatography of CD81-p-null-C104A-D128C concentrated to 200 μ l expressed in *P. pastoris*. A small scale (3 ml of membrane preparations at 60 mg/ml) solubilisation and purification of CD81-p-null-C104A-D128C using 1% (w/v) DDM to solubilise the protein. The protein was concentrated to 200 μ l before loading on the SEC column. The protein was eluted in 200 μ l fractions and run on the SDS-PAGE gel. The gel was stained with InstantBlue Coomassie protein stain (Expedeon).

Having tried different methods, such as solubilising with different concentrations of SMA2000 and DDM and attempting to start the purification with size exclusion, purifying the CD81-p-null-C104A-D128C mutant proved too difficult to complete in a reasonable timeframe. Some results, such as the use of DDM and doing SEC after concentrating the sample to 200 μ l, offered some promise and could benefit from further exploration in the future.

5.6 Conclusion

It has been shown that the purification process when purifying CD81-P-null expressed in *P. pastoris* can be improved by adding a small concentration of imidazole to the binding buffer when incubated the solubilised protein with the Ni-NTA resin. Increasing the imidazole concentration to 30 mM and 60 mM in the two wash steps during purification also improves the purity of the protein. This will be useful in the future when pure protein is required for further downstream analysis.

Mutants of CD81-p-null were successfully transformed into *P. pastoris* with the aim of purifying them and using them in PELDOR experiments to investigate the opening and closing mechanism of CD81. Attempts were made to purify the protein in SMA and DDM. DDM purification performed better but this will require further optimisation in the future. So, at the moment it was not possible to produce sufficient purified protein of the mutants to carry out PELDOR studies.

Chapter 6 – Interaction between ganglioside sugars and tetraspanins

It is known that tetraspanins and gangliosides are present together in TEMs along with cholesterol. Deciphering the specific interactions between these three molecules would provide a greater understanding of how TEMs are formed. Modelling on CD81 in the open conformation suggests that gangliosides interact with the SEL, TM1, TM2 and to a lesser extent the δ -loop in the LEL (Schmidt et al. 2016). Since then, a closed CD81 structure has been discovered (Zimmerman et al., 2016), therefore this thesis will look at ganglioside-tetraspanin interactions when tetraspanins are in their closed state.

6.1 Docking ganglioside sugars to closed CD81

To understand the interactions between gangliosides and tetraspanins, galactose, glucose and lactose molecules were docked onto the closed CD81 structure using Galaxy7TM (Seok & Lee, 2016). Galactose and glucose are sugar residues in the headgroup of gangliosides and the main interactors with membrane proteins when they form hydrogen bonds with the protein. Given that they are the main interactors they were docked without any other part of the ganglioside and results of each molecule were compared to find galactose and glucose molecules that overlapped with the docked lactose molecule.

To find a region of closed CD81 where the sugar residues might bind the closed CD81 structure (PDB: 5tcx) was uploaded to DeepSite (Jimenez et al., 2017) to find potential binding pockets. A promising potential binding pocket discovered using DeepSite was situated on the TM2-TM3 side of the protein around the apex of where TM3 and the α helix in the LEL meet (Figure 6.1). The amino acids highlighted by DeepSite in the binding pocket were Ala108 and Trp111 in TM3 and Ile119, Asp122 and Val123 in the α helix.

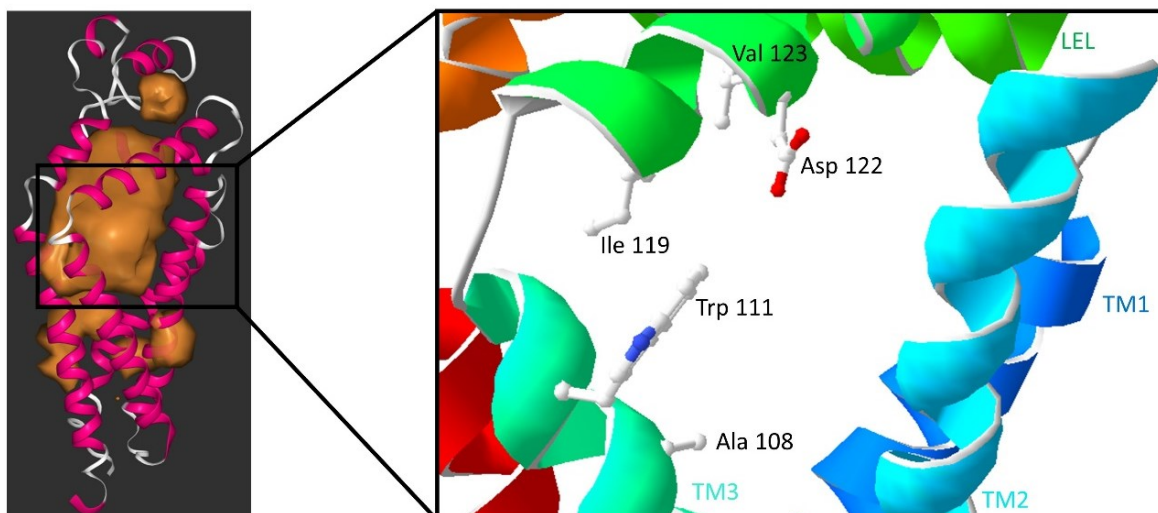


Figure 6.1. Binding pocket of closed CD81. The binding pocket of CD81 in the closed conformation determined using DeepSite (Jimenez et al., 2017). A zoomed in view highlights Ala108 and Trp111, residues located in TM3, and Ile119, Asp122 and Val123, residues located in the α helix in the LEL as a potential binding pocket.

Using the residues identified in the binding pocket the galactose, glucose and lactose molecules were docked to closed CD81 using Galaxy7TM (Seok & Lee, 2016). A galactose molecule docked above where the membrane would be located near the LEL on the TM2-TM3 side and the SEL and TM2 with a docking energy of -6.462 kcal/mol (Figure 6.2A). Thr57 and Gly61, located at the top of TM2, both interact with the galactose molecule via hydrophobic contacts, while Asp122, located in the LEL, which can act as a hydrogen acceptor forms two hydrogen bonds with the galactose molecule. One hydrogen bond is formed between the OH group at the C3 position of galactose and OD1 on aspartic acid and the other hydrogen bond is formed between the OH group at the C4 position of galactose and the oxygen of aspartic acid that forms the protein backbone. The distances of these hydrogen bonds are 3.21 Å and 3.26 Å respectively (Figure 6.2B).

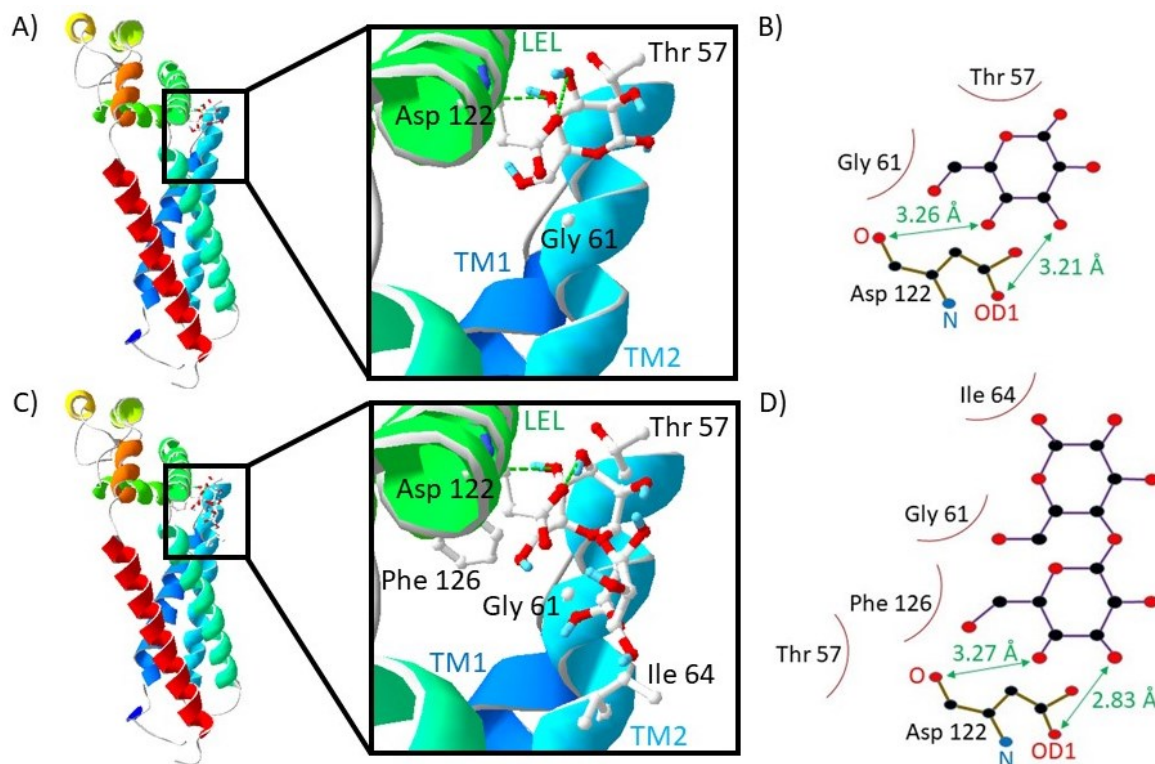


Figure 6.2. Interaction of ganglioside sugars with closed CD81. Cartoon representation of closed CD81 interacting with β -D-galactose and lactose. Galaxy7TM (Seok & Lee, 2016) was used to determine the interactions with binding residues Ala108, Trp111, Ile119, Asp122 and Val123. (A) The interaction of β -D-galactose with the LEL and TM2 of closed CD81 (B) 2D representation of the interaction between closed CD81 and β -D-galactose (C) The interaction of lactose with the LEL and TM2 of closed CD81 (D) 2D representation of the interaction between closed CD81 and lactose. (B) and (D) were adapted from the LIGPLOT (Wallace et al., 1996) output on Galaxy7TM.

Lactose was docked to see if the galactose molecule that docked with closed CD81 is in a similar position when it is docked as part of a lactose molecule with glucose with a docking energy of -8.730 kcal/mol. One of models did line up well with the same OH groups on the galactose forming hydrogen bonds with the same parts of Asp122 (Figure 6.2C). There is minimal change in the hydrogen bond distance between the OH group at the C4 position of galactose and the oxygen that forms part of the protein backbone with an increase of 0.01 Å to 3.27 Å. There is, however, a more pronounced difference between the hydrogen bond distance between the OH group at the C3 position of galactose and OD1 on aspartic acid with a decrease of 0.38 Å to 2.83 Å (Figure 6.2D). Thr57 and Gly61 interact with lactose via hydrophobic contacts, like when just galactose is docked, but Ile64, located in TM2 and Phe126 also interact with lactose in the same way.

To demonstrate the similar position of galactose on its own and as part of lactose with regards to its interaction with closed CD81 the two molecules were superimposed onto each other (Figure 6.3). Glucose on its own was docked and while many of the models resulted in glucose interacting with closed CD81 in the same region of the protein as galactose and lactose, none of them could be overlaid with lactose to demonstrate a similar position in both.

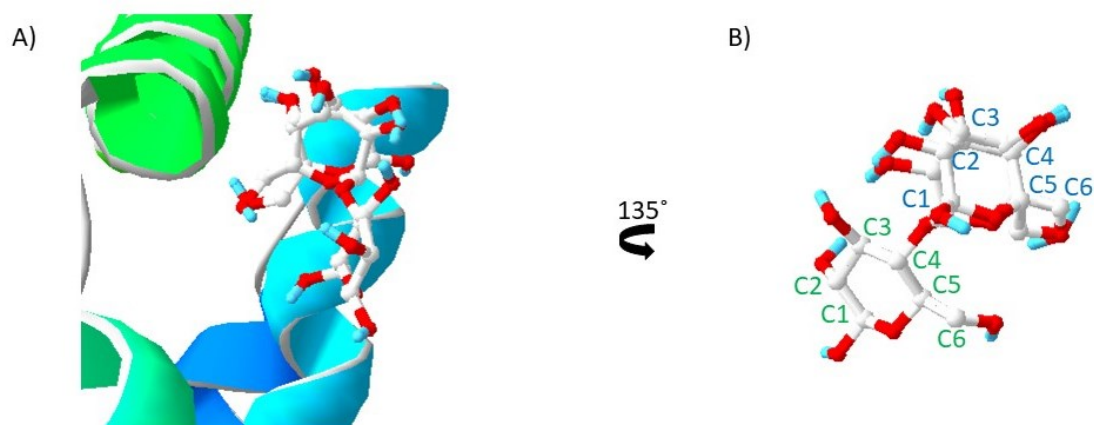


Figure 6.3. Overlaying β -D-galactose and lactose with closed CD81. β -D-galactose and lactose superimposed on the same image to show the similar position and orientation of β -D-galactose on its own and as part of a disaccharide with β -D-glucose to form lactose. (A) The superimposed β -D-galactose and lactose in the presence of the ribbon structure of closed CD81 (B) The superimposed β -D-galactose and lactose with the ribbon structure of closed CD81 removed. Carbons are labelled (green on β -D-glucose and blue on β -D-galactose) and the image has been rotated 135° along the horizontal plane.

The effect of Asp122 on galactose and lactose interactions with closed CD81 was further investigated by mutating it to Ala, an amino acid which is not able act as a hydrogen acceptor or donor. The maximum, minimum and average docking energy of galactose and lactose binding to closed CD81 increased when Asp122 is mutated to Ala (Table 6.1). While the average docking energy for galactose increases by 0.434 kcal/mol from -6.249 kcal/mol to -5.815 kcal/mol when Asp122 is mutated the increase is less pronounced for lactose. Docking lactose resulted in a 0.095 kcal/mol increase from -8.414 kcal/mol to -8.319 kcal/mol when Asp122 was mutated to Ala.

Thr57 is another amino acid involved in the interaction of galactose and lactose by providing hydrophobic contacts (Figure 6.2). In some of the other docked models Thr57 was also involved in hydrogen bonding, alongside Asp122. The effect of mutating Thr57 to alanine was investigated in a similar manner to that was Asp122. Mutating Thr57 to alanine leads to an

increase in the maximum, minimum and average docking energy when galactose is docked (Table 6.1). However, when lactose is docked the maximum and average docking energy decrease when compared to wild-type CD81, despite an increase in the minimum docking energy.

Table 6.1. Docking energy of galactose and lactose with closed CD81. Galactose and lactose docked with wild-type closed CD81 using Galaxy7TM (Seok & Lee, 2016) and a series of mutants to investigate the role of Asp122 and Thr57 in ganglioside binding.

CD81 Protein	Sugar	Average Docking Energy (kcal/mol)	Minimum Docking Energy (kcal/mol)	Maximum Docking Energy (kcal/mol)
Wild-type	Galactose	-6.249	-7.370	-5.605
D122A	Galactose	-5.815	-6.697	-5.102
T57A	Galactose	-6.003	-6.538	-5.574
D122A-T57A	Galactose	-5.861	-6.849	-5.188
Wild-type	Lactose	-8.414	-10.059	-7.456
D122A	Lactose	-8.319	-8.514	-7.444
T57A	Lactose	-8.899	-9.608	-8.113
D122A-T57A	Lactose	-8.258	-8.551	-7.960

To further investigate the effect of Asp122 and Thr57 both amino acids were mutated to Ala in the same CD81 protein. The average docking energy for both galactose and lactose increased when they were docked on the double mutant. The average docking energy for galactose increased by 0.388 kcal/mol from -6.249 kcal/mol to -5.861 kcal/mol and by 0.156 kcal/mol for lactose with an increase from -8.414 kcal/mol to -8.258 kcal/mol.

6.2 Mutating Asp122 in CD81 in *P. pastoris*

Having performed docking studies on sugar interactions with CD81, Asp122 was selected as an amino acid which requires further research. Thr57 has some involvement in the sugar interaction when docking the lactose molecule binding between lactose and CD81 improves when Thr57 is mutated to Ala (Table 6.1). Mutating Asp122 disrupts the interaction, as does mutating both. In the first instance, however, only Asp122 was mutated because of its propensity to form hydrogen bonds with the sugar molecules.

To investigate ganglioside-tetraspanin interactions gangliosides can be fixed to nitrocellulose or polyvinylidene fluoride (PVDF) membranes followed by incubation with purified CD81 and CD82. Furthermore, gangliosides can be coated onto ELISA plates to test the same thing by doing an ELISA. In both instances, CD81-p-null binding can be compared to another version

of the protein that has had Asp122 mutated to Ala to see the effect of Asp122 on ganglioside binding.

Asp122 mutants were produced by site-directed mutagenesis. Asp122 was mutated to Ala, followed by Dpn1 incubation. Mock mutations were run where no primers were added and after Dpn1 incubation 6 µl of each sample was run on a 1% agarose gel. At first, mutagenesis did not work as no PCR product was visible on agarose gels that were run after mutagenesis (Figure 6.4A). The pPICZB plasmid that the CD81-p-null gene had been transformed into is 3.3 kb in length and the CD81-p-null gene is 708 kb. Combined the plasmid is approximately 4 kb, so the first mutation was performed with a 4-minute cycle on the annealing temperature of 68°C because the time should be 1 minute per 1 kb of plasmid. The time was increased to 5 minutes because the annealing time may not be high enough because the size of the plasmid is only just 4 kb. After increasing the time to 5 minutes clear bands were seen at approximately 4.0 kb and in line pPICZB-CD81-p-null plasmid which had not been mutated (Figure 6.4B). This indicated that the plasmid was at the correct size. Some fainter bands were visible in the mock mutations, but the mutated plasmids were taken forward for transformation into TOP10 *E. coli* cells.

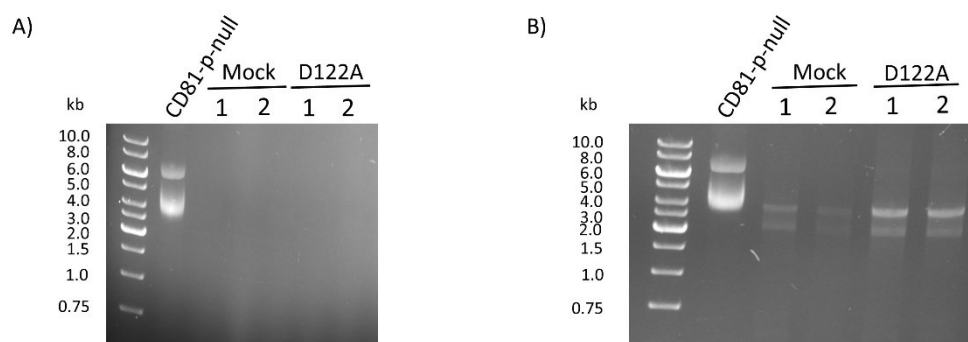


Figure 6.4. Agarose gels of the PCR product after site-directed mutagenesis of CD81-p-null to mutate Asp122 to Ala. pPICZB plasmids containing CD81-p-null were mutated by site-directed mutagenesis to produce a Asp122Ala mutant and run on 1% agarose gels. (A) The first attempt where the annealing time was 4 minutes (B) The second attempt after optimising the annealing time from 4 minutes to 5 minutes. 6 µl of sample was loaded into each lane, consisting of 5 µl plasmid and 1 µl loading dye.

After transformation into TOP10 *E. coli* cells the cells were mini-prepped and sequenced at Eurofins Genomics. Sequencing data was analysed by taking the nucleotide sequence from the sequencing and using ExPasy Translate. Once the frame containing the correct sequence for CD81-p-null had been located its sequence was compared to wild-type human CD81 in

Clustal Omega to check that the desired Asp122Ala mutation had been introduced. Comparing the two sequences showed that Asp122 had been mutated to Ala (Figure 6.5, black box).

```

CD81_WT      -MGVEGCTKCIKYL L F V F N F V F W L A G G V I L G V A L W L R H D P Q T T N L L Y L E L G D K P A P N T F Y   59
CD81_D122A   M S G V E G A T K A I K Y L L F V F N F V F W L A G G V I L G V A L W L R H D P Q T T N L L Y L E L G D K P A P N T F Y   60
              **** * . *****

CD81_WT      V G I Y I L I A V G A M M F V G F L G C Y G A I Q E S Q C L L G T F F T C L V I L F A C E V A A G I W G F V N K D Q I   119
CD81_D122A   V G I Y I L I A V G A M M F V G F L G A Y G A I Q E S Q A L L G T F F T C L V I L F A C E V A A G I W G F V N K D Q I   120
              ***** . *****

CD81_WT      A K D V K Q F Y D Q A L Q Q A V V D D A N N A K A V V K T F H E T L D C C G S S T L T A L T T S V L K N N L C P S G S   179
CD81_D122A   A N A V K Q F Y D Q A L Q Q A V V D D A N N A K A V V K T F H E T L D C C G S S T L T A L T T S V L K N N L C P S G S   180
              *|*****

CD81_WT      N I I S N L F K E D C H Q K I D D L F S G K L Y L I G I A A I V V A V I M I F E M I L S M V L C C G I R N S S V Y ---   236
CD81_D122A   N I I S N L F K E D C H Q K I D D L F S G K L Y L I G I A A I V V A V I M I F E M I L S M V L A A G I R N S S V Y G G G   240
              ***** . *****

CD81_WT      ----- 236
CD81_D122A   H H H H H H 246

```

Figure 6.5. Asp122Ala mutant sequence compared to wild-type human CD81. Asp122Ala mutants transformed into TOP10 *E. coli* cells were sequenced by Sanger sequencing with a 15 µl sample of plasmid (50 ng/µl) with 2 µl of forward AOX1 primer (10 pmol/µl). Sequences were translated using ExPasy Translate and aligned with Clustal Omega. Asp122Ala is shown in a black box.

To be able to produce enough pure protein the mutated plasmid was linearised and chemically transformed into *P. pastoris* X33 cells so that the protein could be expressed. Having enough pure protein is needed so that the dot blot and ELISA experiments could be conducted. After chemical transformation cells were plated on YPD agar plates with different concentrations of zeocin. The concentrations of zeocin used were: (100 µg/ml, 250 µg/ml and 500 µg/ml). There were few colonies that grew on the plates with just one colony growing on the plates containing 250 µg/ml and 500 µg/ml, while 1 large colony and numerous smaller colonies grew on the 100 µg/ml plate (Figure 6.6).

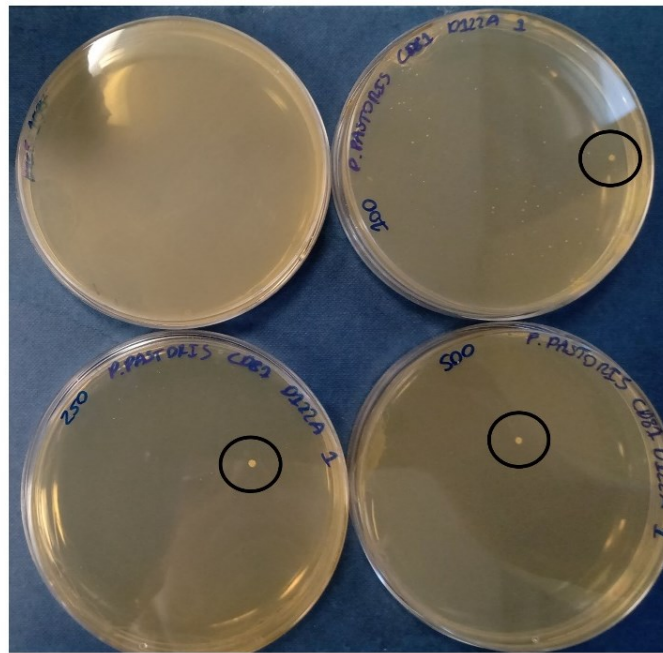


Figure 6.6. Transformed CD81-p-null-D122A mutants on YPD agar plates with varying concentrations of zeocin. pPICZB plasmids with the Asp122Ala mutant were chemically transformed into *P. pastoris* X33 cells on YPD plates containing 100 µg/ml (top right), 250 µg/ml (bottom left) and 500 µg/ml (bottom right). A mock transformation was performed with no pPICZB plasmid added to the transformation (top left). The large, transformed colonies are highlighted in black circles.

The large colonies that grew on each one of the plates, as well as one of the smaller colonies on the 100 µg/ml, were taken and grown on a small scale for small scale protein expression. The cells were grown in 5 ml of BMGY before centrifuging the cells and resuspending in 5 ml of BMMY to express the protein. A Western blot was run and an anti-His antibody used to detect expression of the mutant protein (Figure 6.7).

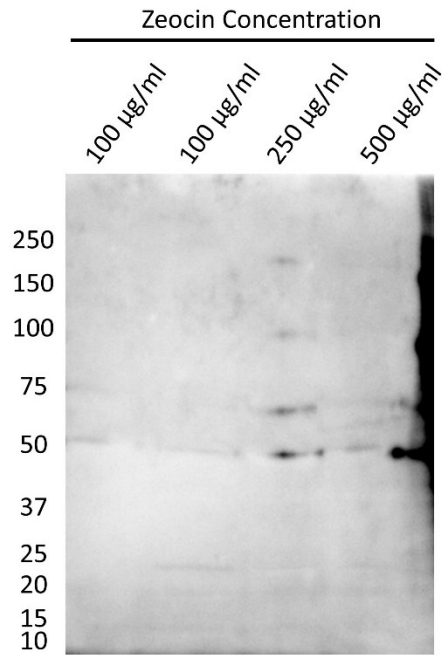


Figure 6.7 Expression of CD81-p-null-D122A mutants. Small scale expression (5 ml) of CD81-p-null-D122A mutants grown on YPD agar plates containing 100 µg/ml, 250 µg/ml and 500 µg/ml of zeocin. Western blot containing 20 µl of sample in each lane transferred to a nitrocellulose membrane. The membrane was incubated with a 1:5000 dilution of 6x His monoclonal antibody (Takara) and then incubated with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signalling Technology). The membrane was then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene).

Expression of the CD81-p-null-D122A mutants was not high (Figure 6.7). There were no dense bands that were usually seen when working with CD81-p-null. A faint band is visible in the second 100 µg/ml lane at approximately 20-25 kDa which would correspond with the size of CD81. Bands are visible in all lanes at approximately 50 kDa which could be dimers of CD81 since CD81 is known to form dimers and higher order oligomers. This could explain the presence of faint bands at approximately 75 kDa and a 100 kDa band in the 250 µg/ml lane. The mutant that grew on the YPD agar plate with 250 µg/ml of zeocin has the densest bands and may require further exploration. CD81-p-null-D122A mutants could not be expressed to a satisfactory level, therefore expression of these mutants will need to be optimised.

6.3 Expression of CD81 in HEK293T cells

Expressing CD81 in *P. pastoris* cells is not ideal when looking at its interaction with gangliosides because gangliosides are not expressed in *P. pastoris*. It is more beneficial to express the protein in mammalian cells. CD81 is endogenously expressed in HEK293 cells

and so are the gangliosides GM1, GM2, GM3 and GD1a making them ideal cells to express CD81 and a mutated CD81 where Asp122 is mutated to Ala.

Wild-type CD81 was expressed in HEK293T cells rather than the CD81-p-null expressed in *P. pastoris*. This decision was made because the p-null mutant is not palmitoylated. Palmitoylation of tetraspanins is needed to interact with other constituents of TEMs. Since the integrity of TEMs, and specifically CD81's interaction with gangliosides, will be investigated it would be better to use a version of CD81 that can interact with gangliosides in its native state. When the Asp122Ala mutation is introduced, there will be no other mutations that could interfere with the CD81-ganglioside interaction which will give clearer information of the role of Asp122 in CD81-ganglioside interactions.

Wild-type CD81 with a FLAG-tag in the pEF6.A plasmid was transiently transfected using 2 µg of plasmid DNA and polyethylenimine (PEI) in HEK293T cells at different cell densities followed by incubation at 37°C for 24 hours. There were either 500,000 or 800,000 cells in 6-well plates at the time of transfection. Mock transfections were done where no DNA was added to the cells. Whole cell lysates were analysed on Western blot using an anti-FLAG and anti-CD81 antibody to detect expression of the protein. Cells were lysed after scraping using radioimmunoprecipitation assay (RIPA) buffer.

Expression of wild-type CD81 was high in HEK293T that were transfected with DNA. When analysing on a Western with anti-CD81 a clear dense band was seen at approximately 25 kDa which corresponds with the correct weight of CD81 (Figure 6.8A). The bands are more intense when CD81 DNA was transfected in 800,000 HEK293T cells rather than 500,000 cells. Despite no DNA being transfected in the mock transfections there are faint bands visible in the mock transfection lanes, especially in the mock transfection with 800,000 cells.

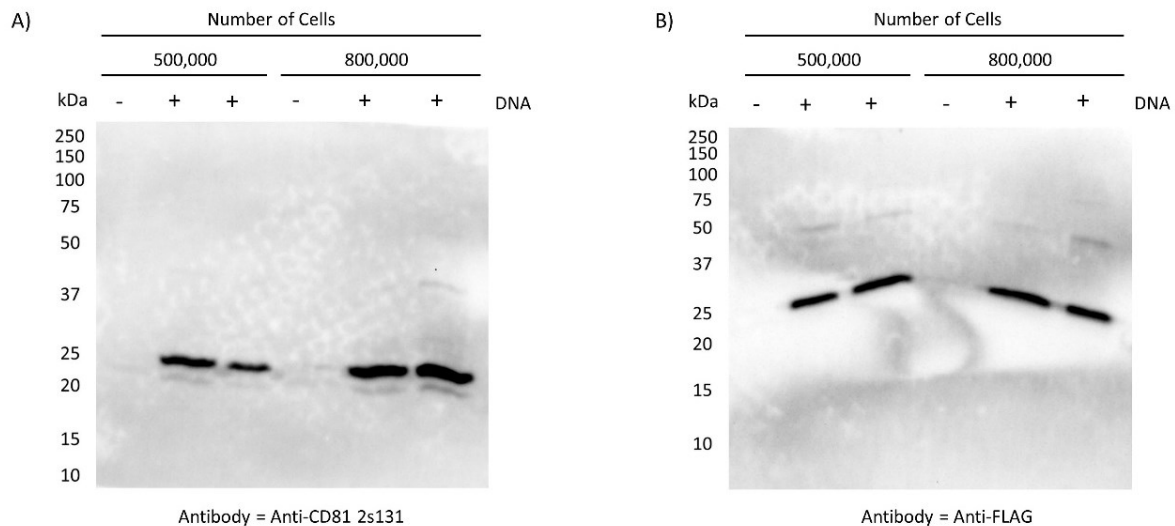


Figure 6.8. Expression of wild-type CD81 in HEK293T cells. Wild-type CD81 was transfected into HEK293T cells and grown at 37°C for 24 hours. After which cells were lysed and analysed on a Western blot. (A) The membrane was incubated with a 1:100 dilution of anti CD81 2s131 (Grove et al., 2017) followed by incubation with a 1:10,000 of HRP conjugated anti-mouse IgG antibody (Cell Signalling Technology) (B) The membrane was incubated with a 1:5000 dilution of anti-FLAG antibody (Sigma-Aldrich) followed by incubation with a 1:2000 dilution of HRP conjugated anti-rabbit IgG antibody (Cell Signalling Technology). The membranes were then incubated with ECL Western blotting substrate (Pierce) and imaged using a G:BOX Chemi XRQ (Syngene).

Analysis with the anti-FLAG antibody showed a similar result to that seen with the anti-CD81 antibody (Figure 6.8B). There were intense bands visible at approximately 25 kDa in all HEK293T cells that were transfected with DNA. Unlike when analysing with anti-CD81, there was no increase in density in DNA transfected in 800,000 cells compared to 500,000. Another difference between the different antibodies was that fainter bands at approximately 50 kDa were more visible when analysing with anti-FLAG which is an indication of dimerization of CD81.

6.4 Docking ganglioside sugars to closed CD82

Gangliosides are known to interact with tetraspanins and the main focus of this thesis has been looking at how CD81 interacts with gangliosides. Little is known about the CD81-ganglioside interaction. Most of what is known about tetraspanin-ganglioside interactions comes from experimental research which involved CD82 (Todeschini et al., 2007; Todeschini et al., 2008). No work has been done on the specific interactions of CD82 with gangliosides, therefore docking of ganglioside sugars was performed on CD82. Probing this interaction and continuing the work *in vitro* would also provide a control for the CD81 work since CD82 is known to interact with so many gangliosides.

No structure of CD82 exists, therefore CD82 was modelled using CD9 as a template. CD82 shares greater sequence similarity with CD9 than any other tetraspanin and a crystal structure of CD9 was discovered at 2.70 Å by X-ray crystallography (Umeda et al., 2020). Homology modelling was performed using SWISS-MODEL (Waterhouse et al., 2018) to generate a model of CD82 (Figure 6.9) that can be used to dock with sugar molecules in a similar manner to that done with CD81.

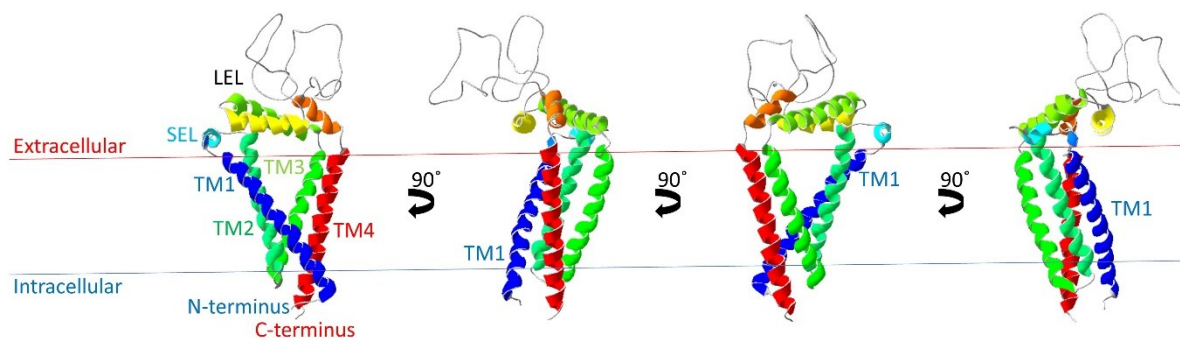


Figure 6.9. Homology model of closed CD82 in the membrane. Cartoon representation of CD82 in the closed conformation modelled using SWISS-MODEL (Waterhouse et al., 2018) with the crystal structure of CD9 (PDB: 6k4J) as a template viewed at 90° intervals parallel to the membrane plane. Four TM domains traverse the membrane in a cone-like formation with TM1 (dark blue) and TM2 (dark green) linked by the SEL (light blue) forming one side of the cone. The other side of the cone comprises TM3 (light green) and TM4 (red) which are linked by the LEL. The N-terminus and C-terminus are situated in the intracellular region, alongside a short SIL connecting TM2 to TM3. TM1 has been labelled in each image.

When analysing the CD82 homology model the overall quality of the structure would be bad. The Quality Model Energy Analysis (QMEAN) score is an overall assessment of the quality of a model (Benkert et al., 2011). It essentially gives a score with regards to how the model deviates from the mean of all known experimentally determined structures. It is based on four factors, which are: the torsion angle potential across three join residues, the interaction potential at the level of C β atoms and all atoms and, finally, the solvation energy which determines how accessible water is to the protein. A score below -4.00 is considered a bad model.

The CD82 homology with CD9 (PDB: 6k4j) as a template has a QMEAN score of -4.98 which is an indication of a bad model (Figure 6.10B). The Ramachandran plot (Ramachandran et al. 1963) showed a majority amino acids at the top of the bottom left quadrant. The accumulation of amino acids in this region means that these residues are likely to be part of a right-handed α helix structure (Figure 6.10A). A lot of the other amino acids are scattered in various parts of the plot which is likely to mean that these residues are part of an unstructured random coil

region. Two amino acids are almost certain to be modelled wrong and they are Ser154 and Gln201 because they both appear in the bottom right quadrant of the plot.

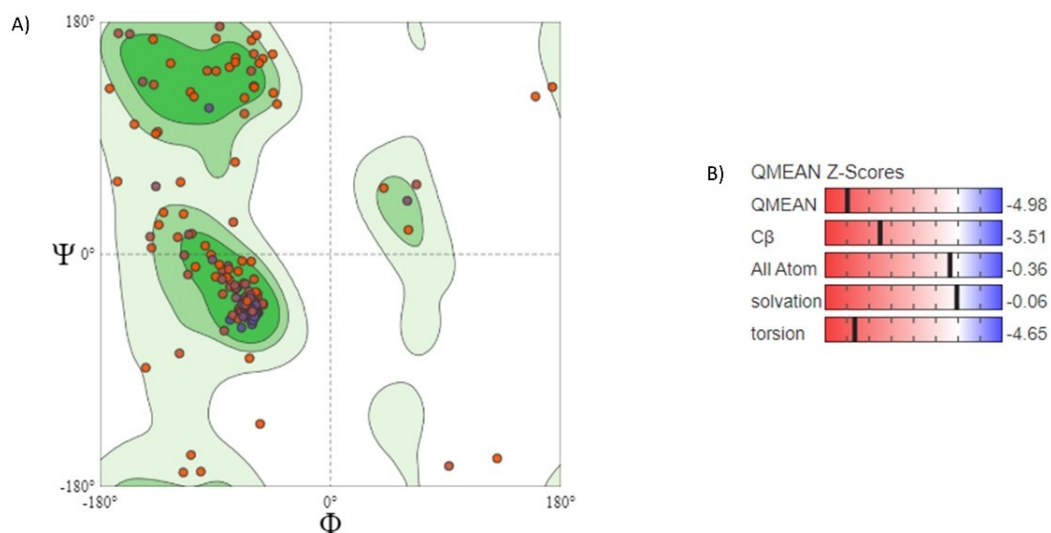


Figure 6.10. Ramachandran plot and QMEAN Z-scores of CD82 homology model. (A) Ramachandran plot showing the angle of amino acids in the CD81 homology model based on CD9 (PDB: 6k4j) and the (B) accompanying QMEAN Z-scores.

The poor quality of the model is largely down to the loop in the LEL between the β helix and the ϵ helix. This is the most variable region in human tetraspanins and the CD9 structure that it was modelled on has a loop region of 26 amino acids, whereas CD82 has a loop region of 52 amino acids. It is the fourth longest loop region in human tetraspanins. All three full-length tetraspanin structures, CD81, CD9 and CD53, have short loop regions. Modelling the CD82 loop region, therefore, is difficult and much many of the problems with the structure are in that loop region (Figure 6.11A, green box).

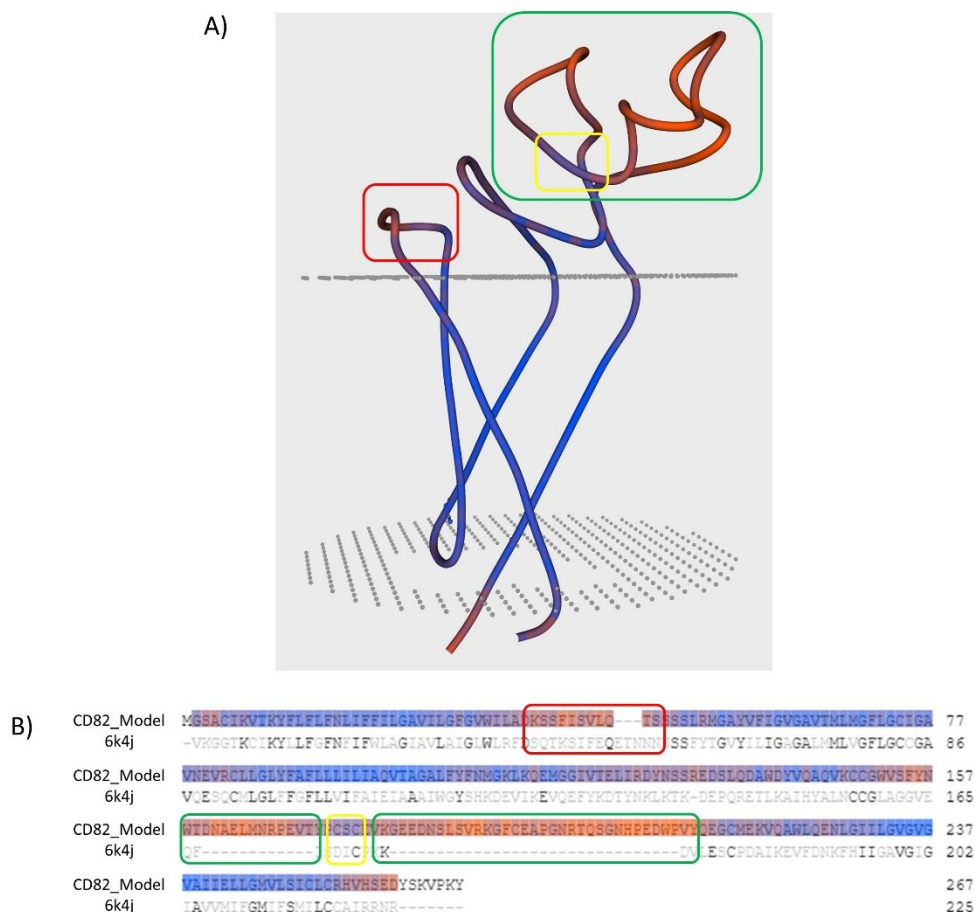


Figure 6.11. QMEANBrane analysis of the CD82 homology model. (A) A rope structure of the CD82 homology model which was modelled on CD9 (PDB: 6k4j) in the membrane showing the location of QMEANBrane (Studer et al., 2014) scores with orange regions having a score below 0.50 and as the colour becomes bluer in the structure the score increases. (B) Model-template alignment showing the corresponding colours for QMEANBrane score in the “CD82_Model” sequence. The unstructured region in the loop in the LEL is in the green boxes. The smaller low QMEANBrane region in the SEL is in the red boxes and the relatively high scoring “CSC” region in the middle of the loop in the loop is in the yellow boxes.

The QMEANBrane score reflects statistical potentials which aims to estimate the local quality of models of membrane proteins (Studer et al., 2014). Low scores, below 0.50, are an indication of lower quality and they are coloured orange in the structure (Figure 6.11A) and the model-template alignment (Figure 6.11B). The colour changes to a purple and changes further to a blue as the quality increases.

The TM regions and the α , β and ϵ helices in the LEL are almost exclusively purple and blue with a QMEANBrane scored of >0.70 . The SEL (Figures 6.11A & 6.11B, red boxes) has a lower QMEANBrane score with the amino acids that comprise the SEL having score of approximately 0.50. The lowest quality region in the CD82 homology model is the region between the β helix and the ϵ in the LEL with many of the amino acids having a QMEANBrane

score > 0.50 (Figures 6.11A & 6.11B, green boxes). In the middle of this region is a three amino acid run of Cys-Ser-Cys (Figures 6.11A & 6.11B, yellow boxes) which have QMEANBrane scores of 0.72-076. The increase in quality in the middle of a region that lacks quality is because of the stability provided by the disulfide bond formed between the second Cys residue in the Cys-Ser-Cys run and the second Cys in the CCG-motif at the start of the loop. The rest of the loop is unstructured which leads to its poor model quality. The rest of the model is satisfactory and will serve as a good model for ganglioside sugar docking.

To determine where the sugars, glucose, galactose and lactose bind to CD82 it was first submitted to the online program DeepSite (Jimenez et al., 2017) to find a potential binding pocket on the CD82 molecule (Figure 6.12). The potential binding pocket differed from that discovered on CD81. The binding pocket on CD81 was located on the TM2-TM3 side of the protein, whereas the binding pocket of CD82 is situated on the TM1-TM4 side of the protein. The amino acids that constitute the potential binding pocket are Ile27, Phe30, Trp33 and Ile34 which are located at the top of TM1, along with Ala139, Tyr142 and Val143 located on the LEL on the TM1-TM4 and, finally, Phe59 situated near the top of TM2 (Figure 6.12).

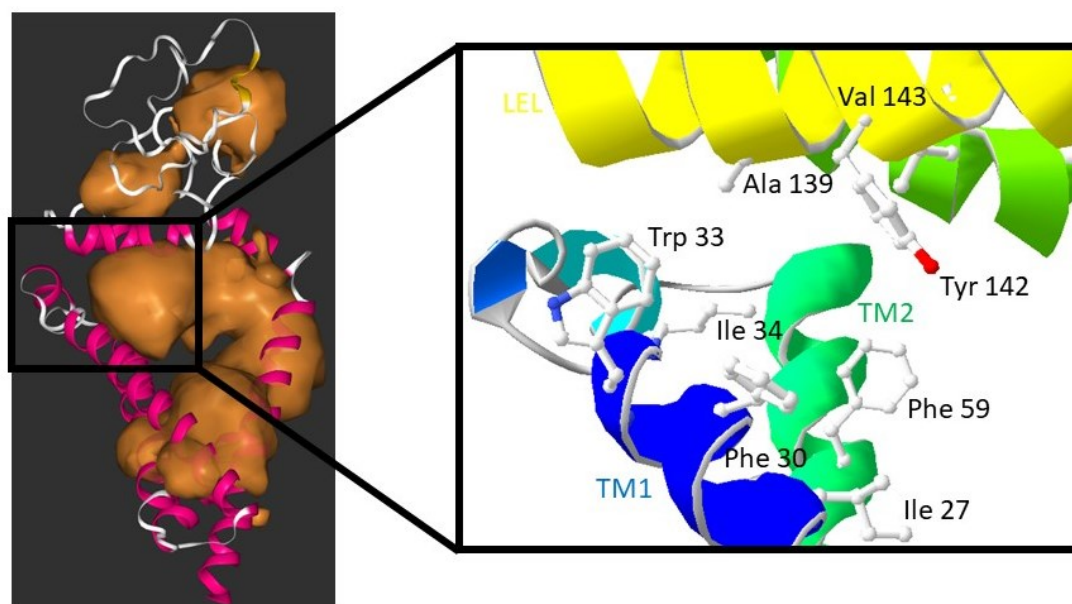


Figure. 6.12. Binding pocket of CD82. The binding pocket of CD82 in the closed conformation was determined using DeepSite (Jimenez et al., 2017). A zoomed in view highlights Ile27, Phe30, Trp33 and Ile34, residues located in TM1, Phe59, a residue located in TM2, and a Ala139, Tyr142 and Val143, residues located in the LEL, as a potential binding pocket.

Galaxy7TM was used to dock glucose, galactose and lactose on CD82 using the amino acids in the potential binding pocket as a guide for where to dock the sugars. Galactose docked with CD82 at the top of TM1 where TM1 and the SEL meet the B helix of the LEL (Figure 6.13A). Four hydrogen bonds are formed between galactose and CD82 with one hydrogen bond at a distance of 3.33Å to Trp33, located at the top of TM1. This interaction occurs between the OH group at the C2 position of galactose and the Trp33 oxygen that forms part of the protein backbone (Figure 6.13B). The other three hydrogen bonds are between galactose and the sidechains of the amino acids it interacts with. For instance, Asp37 is situated in the SEL and forms a hydrogen bond with the OH group at the C3 position at a distance of 2.89Å. The other two interactions are between the OH group at the C6 position of galactose and the sidechains of Ser135 and Asp138, both of which form part of the B helix in the LEL. The hydrogen bond distances are 3.31Å and 3.24Å respectively (Figure 6.13B). Further to the hydrogen bonds forms there are minor hydrophobic contacts between galactose and Leu136 and Ala139 which are positioned near Ser135 and Asp138 in the LEL.

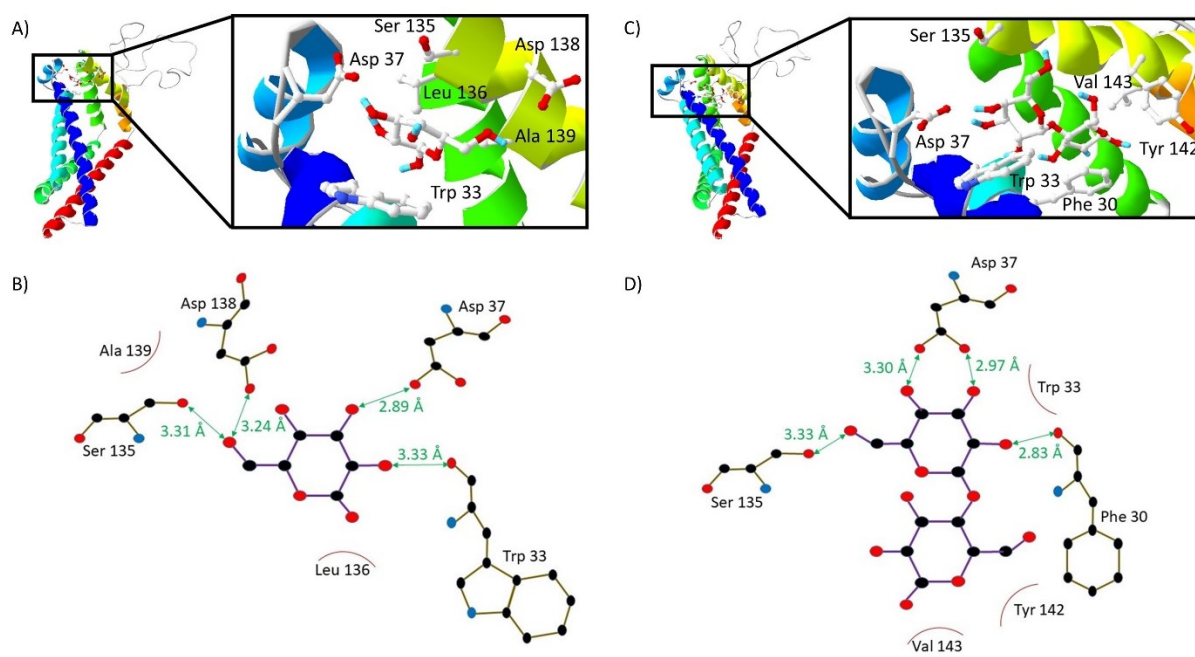


Figure 6.13. Interaction of ganglioside sugars with closed CD82. Cartoon representation of closed CD82 interacting with β -D-galactose and lactose. Galaxy7TM (Seok & Lee, 2016) was used to determine the interactions with binding pocket residues Ile27, Phe30, Trp33, Phe59, Ala139, Tyr142 and Val143. (A) The interaction of β -D-galactose with the LEL and TM1 of closed CD82 (B) 2D representation of the interaction between closed CD82 and β -D-galactose (C) The interaction of lactose with the LEL and TM1 of closed CD82 (D) 2D representation of the interaction between closed CD82 and lactose. (B) and (D) were adapted from the LIGPLOT (Wallace et al., 1996) output on Galaxy7TM.

Docking lactose to CD82 results in the sugar interacting at the same TM1-SEL-LEL region of CD82, much like that seen when galactose is docked (Figure 6.13C). The interaction includes four hydrogen bonds, similar to when galactose is docked. While all four hydrogen bonds are formed between the galactose part of lactose, they are not all with the same amino acids as those seen when galactose was docked on its own.

It can be seen that lactose docks in a position slightly below that of galactose (Figure 6.14A) which brings the galactose part of lactose closer to Phe30 rather than Trp33. While the OH group at the C2 position of galactose forms a hydrogen bond with the Trp33 oxygen that is part of the protein backbone, when lactose is docked the same OH group forms a hydrogen bond with the Phe30 oxygen that is part of the protein backbone (Figure 6.13D). There is also a decrease in hydrogen bond distance between these two interactions of 0.50Å to 2.83Å giving an indication of the slight change in position of galactose when it is part of lactose.

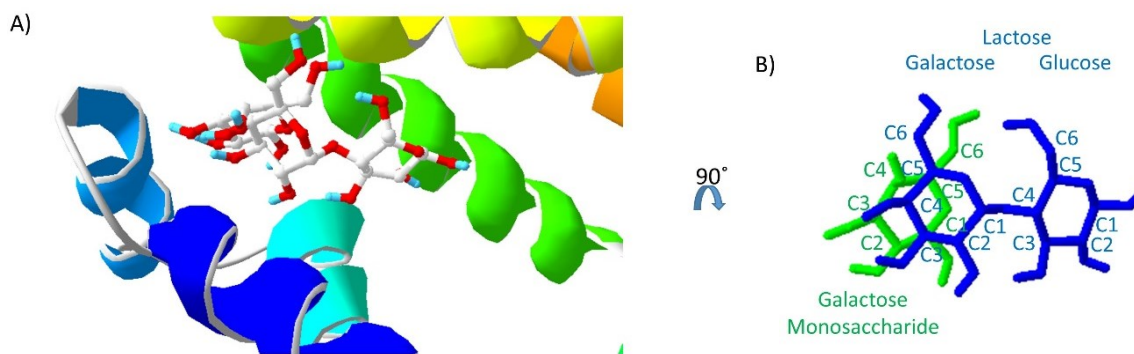


Figure 6.14. Overlaying β -D-galactose and lactose interacting with closed CD82. β -D-galactose and lactose superimposed on the same image to show the similar position and orientation of β -D-galactose on its own and as part of the disaccharide with β -D-glucose to form lactose. (A) The superimposed β -D-galactose and lactose in the presence of the ribbon structure of closed CD82 (B) The superimposed β -D-galactose (green) and lactose (blue) with the ribbon structure of closed CD82 removed. The carbons are labelled and the image has been rotated 90° along the vertical plane.

When galactose on its own is docked it forms one hydrogen bond with the sidechain of Asp37 but as part of lactose there are two hydrogen bonds formed with Asp37. One of these is with the OH group at the C3 position of galactose and the other is with the OH group at the C4 position of galactose. The C3 hydrogen bond is similar to that seen when galactose on its own is docked, despite a small 0.08Å increase in distance to 2.97Å, while the C4 hydrogen bond is at a distance of 3.30Å (Figure 6.13D). The final hydrogen bond occurs between the OH group at the C6 position and the sidechain of Ser135, like that seen when just galactose is docked. The hydrogen bond seen is almost identical with just a 0.02Å increase in the hydrogen bond distance to 3.33Å (Figure 6.13D).

The galactose and lactose molecule were superimposed on top of each other with the interacting CD82 molecule to demonstrate that they are docked in a similar position (Figure 6.14A). The fit between the two molecules is not as good as the fit when docked to CD81 (Figure 6.3A). The galactose monosaccharide C2 carbon is closer to the C3 carbon in the galactose that forms part of the lactose molecule. Glucose was docked and did bind in the same region as galactose and lactose but none of the glucose molecules were close to overlapping the glucose molecule in lactose when they were superimposed.

6.5 Conclusion

Tetraspanins interact with gangliosides in the membrane (Todeschini et al., 2008) as part of TEMs. Research using the Seigneuret model (2006) highlighted some regions of CD81 that interact with gangliosides (Schmidt et al., 2016). Using that model, however, means that the only analysis of tetraspanin-ganglioside interaction was conducted using CD81 in the open conformation.

In this chapter the interaction between CD81 in the closed conformation and the sugar residues in gangliosides was investigated. It highlighted that Asp122 in the β helix forms a hydrogen bond with the galactose residue as a monosaccharide and as part of a lactose molecule. A homology model of CD82 was created in the closed conformation and ganglioside sugars were docked to the model. Asp37, in the SEL, and Ser135, in the β helix, were found to form hydrogen bonds with the galactose.

It is hoped that these findings can be further investigated using molecular modelling techniques and wet laboratory methods, such as thin layer chromatography, dot blots and ELISAs.

Chapter 7 – Discussion

The main aims of this thesis were to characterise the structures and sequence of human tetraspanins and analyse disulfide bond formation in the LEL. Using a combination of sequence and structural analysis, tetraspanins were aligned to devise a universal amino acid numbering system. The numbering system will allow tetraspanin researchers to accurately describe the position of an amino acid. Some tetraspanin LELs were modelled to provide an insight into the structure of their loop regions between the β helix and the ϵ helix. In doing so, disulfide bond patterns were used to further analyse sequence conservation of tetraspanin subfamilies. There were also surprising disulfide bond arrangements discovered in the TspanC6-CxC subfamily which may provide an insight into the conformational flexibility of the LEL.

Further aims of this thesis were to improve the purification process of CD81 expressed in *P. pastoris* by solubilising the protein with SMA copolymer that had been hydrolysed by autoclaving rather than refluxing. Autoclaved SMA proved to be comparable to refluxed SMA when solubilising CD81.

The final aims of this thesis were to investigate the interaction between tetraspanins and cholesterol, as well as their interaction with gangliosides. The CD81-cholesterol interaction and its role in CD81 conformational change were to be investigated with PELDOR. As yet, a sufficient amount of pure CD81 protein has not been obtained, so further optimisation will be required. Molecular docking of ganglioside sugars with CD81 and CD82 was performed to investigate tetraspanin-ganglioside interactions. In CD81, Asp122, and in CD82, Asp37 and Ser135, were identified as amino acids that interact with ganglioside sugar residues via hydrogen bonds.

7.1 Universal tetraspanin numbering system

This research established a universal numbering system to underpin the structural and functional characterisation of human tetraspanins. To do so, structural and sequence conservation were first examined in the TM domains using data from three known crystal structures. TM1 has two positions with 82% conservation: the 6th position is Asn and the 13th position is Gly, which form the first two residues in the NGG7 motif (Figure 3.5) (Kovalenko et al., 2005). Both play significant functional roles in tetraspanins. Asn1.43⁽¹⁸⁾ in CD81 forms a hydrogen bond with the hydroxyl group of the cholesterol molecule when cholesterol is bound inside the cavity (Zimmerman et al., 2016). Gly1.50⁽²⁵⁾ in CD9 is believed to play a crucial role in intramolecular packing between TM1 and TM2 by interacting with Met2.47⁽⁷⁰⁾ in TM2 to aid the correct folding of the protein (Kovalenko et al., 2005). Notably, Gly is the 13th residue of 21 residues in TM1, whereas Asn is the 6th residue. Studies on GWALP23, a well-researched

WALP-like peptide, has shown that the ends of the TM are frayed (Vostrikov et al., 2010; Mortazavi et al., 2016). In contrast, the stability of more central residues is aided by the fact that backbone hydrogen bonds are strengthened by a more apolar environment (Cao et al., 2017). The Gly in the 13th position has a higher BLOSUM62 score than the Asn in the 6th position by 31.97 (Figure 3.4). The 82% conserved Gly residue in the centre of TM1 was therefore chosen as the anchor residue for the universal residue numbering system and denoted Gly1.50.

The Gly at the 6th position in TM2 is 97% conserved. One limitation in making it the anchor residue is that it is close to the end of the TM. However, the next most conserved residue in TM2 is a Gly residue at the 16th position which is also 6 residues from the end of the TM. The 97% conserved Gly at the 6th position makes up part of the TM2 GGA7 motif (Figure 3.8). The corresponding Gly2.50⁽⁶⁷⁾ in CD9 is an important residue because it interacts with Leu1.54⁽²⁹⁾ in TM1. Mutating Gly2.50⁽⁶⁷⁾ to Leu perturbs this Gly2.50⁽⁶⁷⁾-Leu1.54⁽²⁹⁾ interaction which has a detrimental effect on CD9 folding, leading to aggregation (Kovalenko et al., 2005).

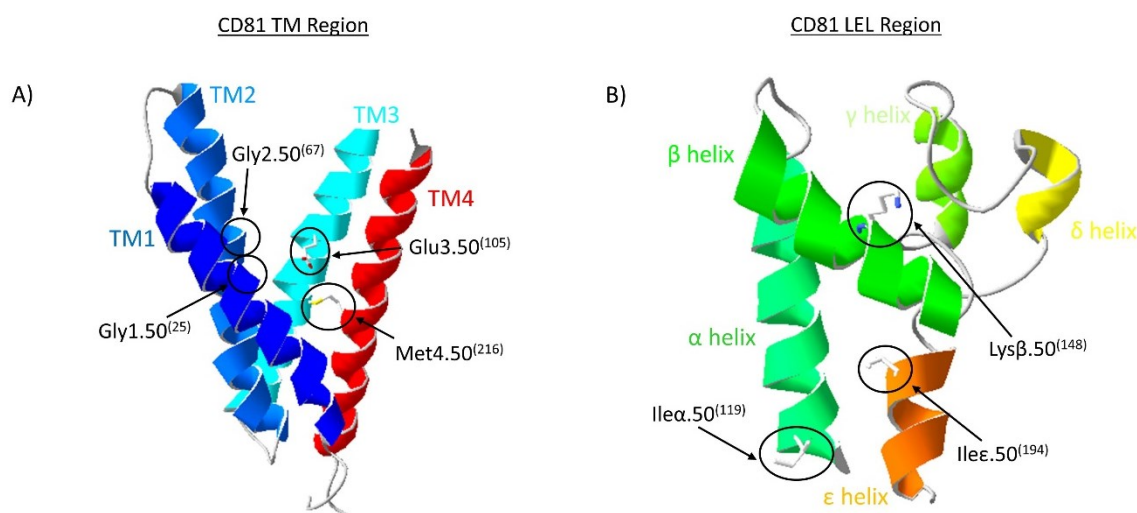


Figure 7.1. Position of anchor residues in the TM and LEL regions of CD81. (A) The position of the anchor residues (Gly1.50⁽²⁵⁾, Gly2.50⁽⁶⁷⁾, Glu3.50⁽¹⁰⁵⁾ and Met4.50⁽²¹⁶⁾) in the TM domains of CD81 (PDB: 5tcx) (B) The position of the anchor residues (Ile.alpha.50⁽¹¹⁹⁾, Lys.beta.50⁽¹⁴⁸⁾ and Ile.epsilon.50⁽¹⁹⁴⁾) in the LEL region of CD81 (PDB: 5tcx).

In TM3, Glu/Gln in the 18th position are 94% conserved (Figure 3.12). Glu and Gln are chemically similar, a fact exemplified by the positive BLOSUM62 score of 2 assigned to the likelihood of this substitution occurring (Henikoff & Henikoff, 1992). The pKa of the Glu residue in TM3 is approximately 6.9 which would suggest that in that context it is protonated and uncharged. UP1b, a tetraspanin expressed in the bladder, has been shown to stay in the endoplasmic reticulum (ER) when Glu3.50⁽¹⁰²⁾ is mutated to Ala (Tu et al., 2006), indicating

that this residue may be required for correct folding of the protein. To demonstrate the chemical similarities of Glu and Gln, the authors of this study mutated Glu3.50⁽¹⁰²⁾ to Gln, which had no effect on the protein's ability to exit the ER. This finding gives further confidence in choosing Glu/Gln 3.50 as an anchor residue.

Glu/Gln 4.50 is 75% conserved. There is a Gly residue in the 18th position that shows 67% conservation (Figure 3.16). In CD81 it is a Glu residue, Glu4.58⁽²¹⁹⁾, which has been shown to form a hydrogen bond with the hydroxyl group of cholesterol when it is bound inside the cavity (Zimmerman et al., 2016). Glu4.58⁽²¹⁹⁾ is one helical turn closer to the C-terminal than 4.50. Other tetraspanins, such as CD9 and CD82, are known to interact with cholesterol (Charrin et al., 2003; Huang et al., 2020), although it is unclear whether this interaction occurs inside the cavity as for CD81 (Zimmerman et al., 2016). If this interaction does occur inside the cavity then it is possible that a hydrogen bond could be formed with the Glu/Gln residue at 4.50 in a similar manner to the bond formed between the hydroxyl group of cholesterol and Glu4.58⁽²¹⁹⁾.

One problem with regards to the alignment of the TM sequences and structures is that the only full-length structures of human tetraspanins that we have are CD81, CD9 and CD53. All three of these belong to the TspanC4 sub-family. It is possible, therefore, that there could be structural differences amongst the tetraspanins. For instance, it is possible that the length of some of the TMs could be different from those reported in this thesis. The ends of the TMs were decided based on a combination of TMHMM 2.0 (Moller et al., 2001), structural analysis and sequence conservation. In terms of sequence conservation particular emphasis was placed on aromatic residues like Trp and Tyr. Trp and Tyr are often found in the interfacial region forming an "aromatic belt" (Braun & von Heijne, 1999). Another aspect considered when trying to define the ends of TMs was the presence of "snorkelling" residues, such as Lys and Arg which interact with phospholipid heads and are present at the water-membrane interface (Kim et al., 2012). It may be that some of the TMs in the tetraspanins in this thesis are longer but, in most cases, the aromatic residues were used as the cut-off point. The key to designing the universal numbering system is that main core of the TMs is included since this where an anchor residue would ideally be located because of uncertainty surrounding the ends of TMs.

Another assumption made is that there are no kinks in the TMs. Approximately half of TMs have a kink in them with one of the most common causes for a kink being the presence of Pro residue because it is not compatible with a helical structure (Hall et al., 2009; Langelaan et al., 2010). CD81, CD9 nor CD53 contain a Pro residue in any of their TMs but there is a Pro residue present in TM1 of Tspan10, TM3 of PRPH2, although it is the first residue of the TM, and TM4 of Tspan15, UP1a and Tspan33. The presence of these Pro residues raises the possibility of kinks within these TMs that would have to be considered when using this generic

tetraspanin residue numbering system. However, even if kinks do arise the generic numbering system presented in this thesis can accommodate them, just like the system developed for GPCRs (Isberg et al., 2015).

Within the LEL, the α helix is perpendicular to the membrane with a slight incline away from the membrane as the α helix moves further away from TM3. Four hydrophobic residues are situated on the underside facing inwards towards the cavity formed in the centre of the protein by the arrangement of the TMs. Two of them, Ile α .50⁽¹¹⁹⁾ and Val α .54⁽¹²³⁾, are situated on the underside of the α helix in the closed conformation of CD81. Of the four hydrophobic residues, the best choice for an anchor residue in a universal residue numbering system appears to be the first largely hydrophobic position (Figure 3.18).

The β helix is a short helical structure situated immediately before the CCG motif and it contains two positions of relevance for designing a universal residue numbering system. The 8th position is 61% Asp, while the 11th residue is 73% Gln (Figure 3.19). The higher conservation of Gln is compelling but this residue is situated further away from the middle of the region than the conserved Asp residue. At the 8th position, Asp and Asn can be grouped together because they are chemically similar, highlighted by the positive score of 1 attributed to a substitution from one to the other in the BLOSUM62 score matrix (Henikoff & Henikoff, 1992). The structural significance of the 8th position is unclear with little work being conducted on residues at this position. In CD81, the position is occupied by a Lys rather than Asp/Asn and *in silico* research suggested that this Lys β .50⁽¹⁴⁸⁾ was involved in CD81's interaction with claudin 1, although experimental data demonstrated the effect of mutating this residue had little impact on this interaction (Davis et al., 2012).

The 100% conserved Cys residue at the start of the ϵ helix is not ideal for a universal residue numbering system. For example, the closed structure of CD81 (Zimmerman et al., 2016) has a shorter ϵ helix than the open structure of CD81 (Susa et al., 2021). In neither structure does the helical structure include this 100% Cys residue but the prediction of PSIPRED indicates that this Cys is the start of the ϵ helix. For the purposes of this work, the Cys residue is highlighted as the start of the ϵ helix because it is conserved across all tetraspanins. By doing this we see an ϵ helix of 12 amino acids for the majority of tetraspanins, allowing for the placement of an anchor residue near the centre of the structure. One such position is the 5th residue in the sequence which has a BLOSUM62 score of 173.04, which is significantly higher than the next highest of 135.48, excluding the position occupied by the 100% conserved Cys residue (Figure 3.21). The 5th position is 84% hydrophobic, so it is possible that the same phenomenon is seen as in the α helix, whereby hydrophobic residues are situated on the

underside of the structure because they are located within the cavity formed by the arrangement of the TMs.

The lengths of the α , β and ϵ helices were determined by secondary structure prediction which was analysed using PSIPRED. The level of confusion between helices and coils when using secondary structure prediction tools is 8-9%, indicating that the ends of helices are not as easily predicted as internal regions of helices (Yang et al., 2016). It is highly likely, therefore that the ends of some of the helices are wrong, particularly when looking at the α helix which has a greater variation in length. Also, the α helix doesn't have a well-defined start or end like the CCG-motif at the end of the β helix/start of the loop in the LEL or the conserved Cys residue that starts the ϵ helix. Problems with secondary structure prediction add to the need to find anchor residues located as close to the centre of a structural region as possible.

The universal tetraspanin numbering system will be beneficial for tetraspanin researchers to describe the position of an amino acid accurately and efficiently in a simple manner. The basic principle which has been introduced here is that the anchor residues that have been presented will serve as a reference point in each region of the protein to work from. For instance, the anchor of TM1 will be given the generic number 1.50. The 1 denotes that the amino acid is in TM1 and 50 represents the designated anchor residue. Moving towards the N-terminal results in the number 50 decreasing, while moving towards the C-terminal results in the number increasing. As an example, the 22nd residue in TM1 of CD81 is Trp and is three residues closer to the N-terminal than the anchor residue, therefore it would be labelled 1.47. Further information can be added to give a full picture by including the name of the amino acid first and the residue number in after the generic number. For example, Trp22 would be written as Trp1.47⁽²²⁾.

To conclude, a universal tetraspanin numbering system has been proposed that categorises known family members by their TM domains and features of their LEL. It's anticipated that future categorisation of yet-to-be-discovered tetraspanins using our system will provide insight into their function.

7.2 Greater sequence conservation in loops of the LEL is apparent when categorised into subfamilies based on cysteine arrangement

The significant difference in lengths of the variable loop in the LEL which connects the β and ϵ helices makes finding an anchor residue almost impossible. Instead, loops were categorised by sub-family (Huang et al., 2005; De Salle et al., 2010). Here, the variable loops were aligned based on these sub-families and the arrangement of the in-loop Cys residues. It is already common to refer to Tspan5, Tspan10, Tspan14, Tspan15, Tspan17 and Tspan33, all of which have eight Cys residues in their variable loop, as members of the TspanC8 sub-family

(Matthews et al., 2017). Terms such as TspanC4 and TspanC6 are also in use but here this thesis has introduced nomenclature for all sub-families based on how many Cys residues are in the LEL and the arrangement of the two Cys residues located near the centre of the variable loop.

Tetraspanins with four Cys residues in the loop are called TspanC4 and tetraspanins with six residues in the loop are divided based on the arrangement of the middle Cys residues. These sub-families are TspanC6-CC, TspanC6-CxC and TspanC6-CxxxC. The C6 refers to the six Cys residues in the loop, while the -CxC refers to the two middle Cys residues being separated by one amino acid. There is another family known as TspanC6 β which also has six Cys residues in the variable loop but they also possess a further Cys residue in the LEL, in their β helix.

There is little sequence conservation across the human tetraspanins in the variable loop of the LEL other than the CCG-motif and other cysteine residues (Figure 4.2). Greater sequence conservation can be seen, however, when separating the tetraspanins based on their cysteine arrangement.

Tetraspanins in the same subfamily have been shown to have similar functions. TspanC8 proteins, for instance, are known interactors with ADAM10 (Matthews et al., 2017) and TspanC6 β proteins are only expressed in the eye and interact with each other via the extra cysteine in the β helix which forms an intermolecular disulfide bond (Goldberg & Molday, 1996). This may provide an insight into the function of other, largely unexplored, tetraspanins. Many of the TspanC6-CCs have not been researched extensively but taking what is known about well-researched ones, like CD151, may guide research into the other tetraspanins in the same family.

7.3 Disulfide bond pattern differs by tetraspanin subfamily based on cysteine arrangement in the LEL

The arrangement of cysteines in the LEL of tetraspanins aids structural stability via the formation of disulfide bonds (Kitadokoro et al., 2001; Seigneuret, 2006). Modelling the 3D structures of tetraspanin LELs shows that the preferred disulfide bond formation for most tetraspanins is with a disulfide bond formed between the second Cys residue of the CCG-motif and the first Cys residue in the middle of the loop. Three of the four tetraspanins belonging to the TspanC6-CxC, however, see the formation of a disulfide bond between the second Cys of the CCG-motif and the second Cys of the two Cys residues found in the middle of the loop (Figures 4.10 & 4.11). Despite this modelling it may be that due to large parts of the loops being unstructured the disulfide bond arrangement may be more fluid than is expected. Crystal structures of the CD81 LEL showed different LEL conformations based on disulfide bond

formation with some structures showing a loss of disulfide bond between the second Cys residue of the CCG-motif and the only Cys residue in the middle of the loop (Cunha et al., 2017). Molecular modelling by Cunha et al. (2017) then showed that low pH conditions could lead to the loss of this disulfide bond and that the resulting LEL conformation could be exploited by the HCV under endosomal conditions. The loss of this disulfide bond leads to a more open and flexible LEL and it is possible that a similar phenomenon could occur in other tetraspanins. If it did then it is possible that the the second Cys residue of the CCG-motif could form a disulfide bond with either the first or second of the middle two Cys residues due to their proximity and the flexibility of the unstructured regions of the LEL.

The classification of secondary structures within the LEL are called into question based on the structures predicted here and recent experimentally discoveries. Some of the PSIPRED secondary structure predictions and trRosetta structure predictions here highlight the possible existence of numerous secondary structures in the loop of the LEL. The current naming system is based on CD81 since it is the most researched tetraspanin. The naming of the structures in the LEL is predicated on a maximum of five structural regions because it is named from α - ϵ .

The predicted structures of CD37 (Figure 4.9) and ROM1 (Figure 4.14) question this assumption because the predicted trRosetta structure of CD37 has two helical structures and two sheets between the β helix and the ϵ helix. The PSIPRED prediction also predicts that there are four structured regions in the loop, although they are all predicted to be helices rather than any sheets. Recent experimental discoveries have confirmed that some tetraspanins have a greater complexity of structures in the loop in the LEL. The Tspan15 LEL structure has two helices and two sheets in the loop (Figure 4.17) (Lipper et al., 2022).

The ROM1 trRosetta structure also has more than two helices in the loop because it is predicted to have four helices in the loop, although the PSIPRED prediction only predicts one helix (Figure 4.14). Secondary structure prediction is not perfect with many of the latest secondary structure prediction tools having an accuracy in the region of 80-85% (Yang et al., 2016). The latest version of PSIPRED, PSIPRED 4, falls within this range with 84.2% accuracy (Buchan & Jones, 2019). Helices are predicted with greater accuracy than sheets and coils which may explain why there is a discrepancy in the secondary structure prediction of Tspan15 and the Tspan15 LEL model (Lipper et al., 2022).

These predictions of secondary structures, as well as the Tspan15 LEL structure, mean that the classification of secondary structures in tetraspanin LELs may need to be renamed. The current naming of helices from α to ϵ assumes that five helices is the maximum found in the

LEL which has been proven to be wrong and it neglects the possibility of the presence of sheets rather than helices which has been predicted and shown experimentally.

Given that the three structures common to all tetraspanins are well known as the α , β and ϵ helices it would be best to keep these names the same. One possible way to name the structures in the loop between the β and ϵ helices would be to label the first helix in the loop of CD81 H1^(1/2) (Fig 7.2A). This naming system works with either an H or S to indicate the secondary structure as either a helix or a sheet. The number which follows the letter indicates which number of that specific structure (either helix or sheet) it is in the loop. That number is followed by a number in brackets which indicates which position the structure in question is in relation to the total number of structured regions that there are in the loop. In this example, therefore, it would be a helix, indicated by the “H”. It is the first helix in the loop, indicated by the “1” and the number “1/2” in brackets shows that it is first structured region out of two in the loop. Another example can be seen in Tspan15 which has a more complex loop region with four structured regions in total (Lipper et al., 2022). For instance, the second sheet in the Tspan15 LEL is the fourth structured region in the loop, therefore this sheet would be named S2^(4/4) (Fig. 7.2B).

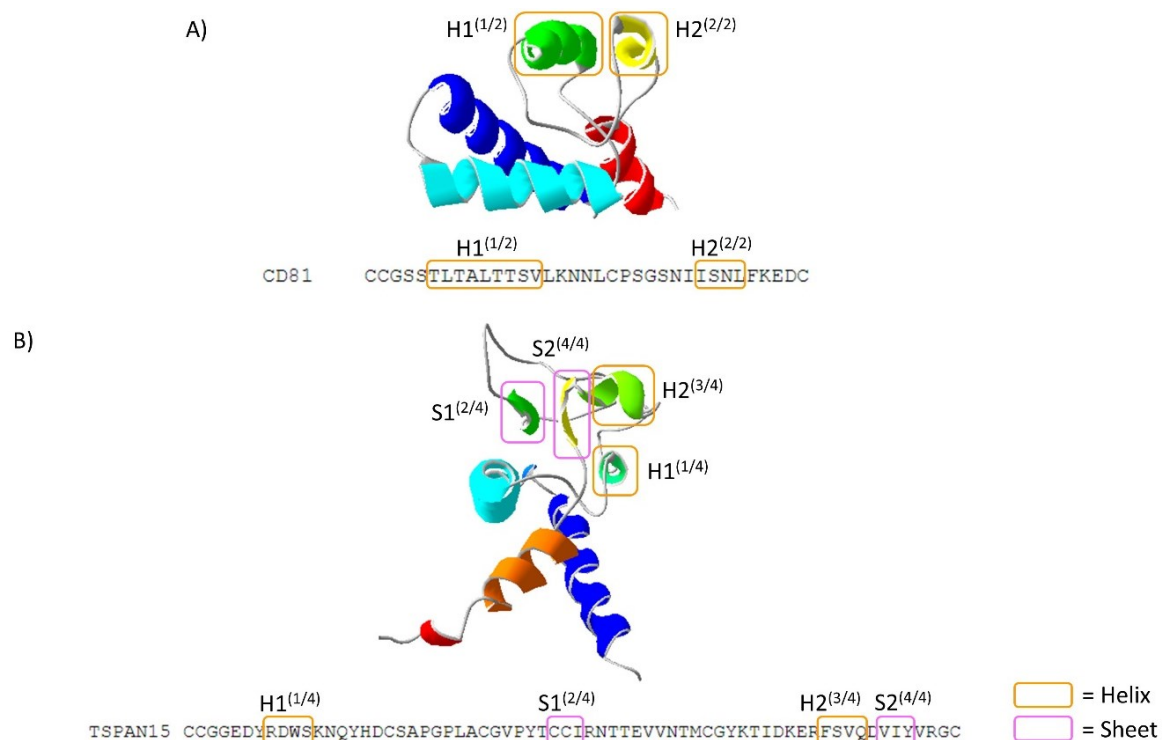


Figure 7.2. The proposed new naming system of the structures in the loop region in the LEL of human tetraspanins. (A) The new names (H1^(1/2) and H2^(2/2)) for the two helical structures in the loop region in the LEL of CD81 (PDB: 5tcx) (B) The new names (H1^(1/4), S1^(2/4), H2^(3/4) and S2^(4/4)) in the loop region in the LEL of Tspan15 (PDB: 7rdb). Helices and sheets are labelled in the structures and at their respective position in the CD81 and Tspan15 loop sequences. Helices are located in an orange box and sheets are located in a pink box.

Sequence and structural conservation methods have provided an insight into the structure of tetraspanin LELs and their disulfide bond arrangement but there are still structural questions with regards to tetraspanins that need to be answered. The CD81 LEL has been proposed to move upwards when cholesterol is not present in its central cavity (Zimmerman et al., 2016), therefore this mechanism will be investigated with PELDOR. To investigate this it is first necessary to solubilise and purify CD81, therefore CD81 was solubilised and purified using SMA copolymer.

7.4 Solubilisation of CD81 with autoclaved SMA2000 is comparable to solubilisation with refluxed SMA2000

SMA is a copolymer capable of solubilising membrane proteins in their native lipid environment. To be able to solubilise membrane proteins the hydrophobic SMA_{nh} needs to be hydrolysed to produce SMA which is amphiphilic and capable of inserting itself into the membrane. The traditional method of hydrolysis of SMA2000 is refluxing but a new method whereby the SMA2000 is autoclaved two times (Kopf et al., 2019) has been described. Kopf et al. (2019) describes the method and chemically characterises the SMA200 that is hydrolysed using an autoclave but they did not test the solubilisation efficiency of the autoclaved SMA2000, therefore the solubilisation efficiency of autoclaved SMA2000 was compared to refluxed SMA2000.

25 g of SMA2000 was dissolved overnight in 250 ml 1 M NaOH and autoclaved the following day two and three times. Hydrolysis is complete after two autoclave cycles (Kopf et al., 2019) but in that paper they did three cycles as well, therefore a comparison between two and three cycles of the autoclave was done, as well as comparing to the refluxed method.

After one hour of solubilisation of CD81 in *P. pastoris* membranes at room temperature the refluxed method performed better with a solubilisation efficiency of 65.82%. This is better than the SMA2000 autoclaved twice which was 12.89 percentage points lower and the SMA2000 autoclaved three times which was 13.51 percentage points lower. Solubilising for two hours results in a similar difference in solubilisation efficiency. The refluxed method has a 69.91% solubilisation efficiency which is 13.38 percentage points higher than SMA2000 autoclaved twice and 12.33 percentage points higher than SMA2000 autoclaved three times (Figure 5.1). The P value determined by one-way ANOVA for both the one hour and two-hour solubilisations were >0.05 which is an indication of a lack of statistical significance between the results at each time point. Further experiments to increase the n number from 3 could lead to there being statistical significance between the results. Nevertheless, the results obtained showed a small difference in solubilisation efficiency between the reflux and autoclave methods, as well

subsequent solubilisations using the autoclave method provided sufficient solubilisation of CD81.

These are relatively small differences in solubilisation efficiency and considering the standard deviation of each of the methods the possibility exists that the difference could be closer. Furthermore, the autoclaved method is easier and quicker because all it requires is placing the dissolved SMA2000 into the autoclave, whereas when refluxing it must be watched in the initial stage to ensure that the solution does not boil over. Despite performing slightly worse than the refluxed method both autoclave methods have solubilisation efficiencies >50% which is adequate for solubilising with SMA2000, therefore the choice was made to hydrolyse SMA2000 by autoclaving rather than refluxing.

Given the even smaller difference in solubilisation efficiency between SMA2000 autoclaved twice and three times it was decided that there was no need to do the third autoclave cycle. This agrees with Kopf et al. (2019) which indicated that hydrolysis of SMA2000 is complete after just two autoclave cycles.

When preparing SMA2000 using the reflux method the dissolved SMA2000 is transferred to a round bottomed glass and refluxed with a stirrer (Rothnie, 2016). Preparing the SMA2000 using the autoclave method requires placing either the conical flask that the SMA2000 was dissolved in into the autoclave or transferring to a bottle to be autoclaved. In this study the dissolved SMA2000 was transferred to a glass bottle and autoclaved. It was discovered that not autoclaving immediately after transferring the dissolved SMA2000 to the glass bottle resulted in an accumulation of SMA2000 stuck to the bottom of the bottle after the first cycle. SMA2000 would stick to the bottom of the bottle when the solution was allowed to rest for 30-60 minutes before autoclaving (Figure 5.2).

When dissolving 10% (w/v) of SMA2000 in 1 M NaOH overnight not all the SMA2000 is dissolved. Consequently, it is likely that leaving the solution to rest for 30-60 minutes without stirring leads to some of the undissolved SMA2000 sinking to the bottom of the bottle and then the high temperature reached during autoclaving causes it to become stuck to the bottom of the glass bottle. This is not a problem encountered when using the reflux method because the solution undergoes constant stirring with a stir bar (Rothnie, 2016). It should be noted, therefore, that when choosing the autoclave method it should be autoclaved as soon as possible after dissolving to limit the amount of SMA2000 that could stick to the bottom of the glass bottle.

In conclusion, the solubilisation efficiency when solubilising CD81 expressed in *P. pastoris* cells with autoclaved SMA is almost as good as when solubilising with refluxed SMA.

Preparation of autoclaved SMA is more simple, therefore subsequent work which requires SMA solubilisation was carried out using autoclaved SMA.

7.5 Purification of CD81-SMALPs with Ni-NTA resin can be improved by increasing imidazole in the wash steps and adding imidazole to the binding buffer

To do biochemical and/or structural analysis on a recombinant protein it is necessary to be able to produce enough pure protein. The purity of recombinant protein can be improved by changing the conditions in which the protein is purified, therefore steps were taken to optimise the purification of CD81 solubilised with SMA during the Ni-NTA chromatography stage.

Four different conditions were tested to try and improve the purity of CD81. The four conditions were: increasing the concentration of imidazole in the wash steps, doubling the volume of the wash steps and adding imidazole to the binding buffer in combination with either increasing the concentration of imidazole in the wash steps or doubling the volume of the wash steps. All four conditions tested helped to reduce non-specific binding to the Ni-NTA resin with fewer contaminating bands visible on an SDS-PAGE gel. The best performing condition, however, was when 20 mM imidazole was included in the binding buffer and the concentration of imidazole in the wash steps was increased from 20 mM and 40 mM to 30 mM and 60 mM (Figure 5.7). Adding 20 mM imidazole to the binding buffer and doubling the volume of the 20 mM and 40 mM wash steps had a similar effect on reducing the contaminant proteins found on the SDS-PAGE gel but it also led to a reduction in the amount of CD81 that was captured in the elution steps (Figure 5.6).

It is clear, therefore, that adding a small concentration of imidazole (20 mM) to the binding buffer can have a significant impact on reducing contaminant proteins from being eluted in the elution steps, presumably because most are not binding to the Ni-NTA resin in the first place due to the presence of imidazole in the binding buffer. The addition of imidazole to the binding buffer may seem counter-intuitive when working with SMALPs because it has been shown that membrane proteins encapsulated in SMALPs can have reduced binding to Ni-NTA resin (Pollock et al., 2018). This does not appear to be a problem working with CD81-SMALPs, however, because only small amounts of CD81 monomer can be seen in the flow through of a Western blot of the original purification conditions that did not include imidazole in the binding buffer (Figure 5.8).

It is sometimes necessary to increase the concentration of imidazole used in the elution steps to ensure that all the target protein is eluted but that is not necessary when working with CD81. One optimisation step tried was to increase the concentration of imidazole in the elution step from 300 mM to 500 mM but there was no increase in CD81 eluted. In fact, the densest band was the fifth 300 mM elution which shows that all CD81 can be eluted with 300 mM imidazole

but it may require a larger volume of elution buffer to make sure that all CD81 is eluted (Figure 5.9).

7.6 Purification of CD81-C104A-D128C improves when solubilised with DDM instead of SMA2000

Mutants to probe the conformational change of CD81 were created using site-directed mutagenesis. All the mutants showed high levels of expression when expressed on a small scale. None of the mutants were showing high levels of expression at 25 kDa, the molecular weight of CD81. Instead, high levels of expression were at approximately 50 kDa and 100 kDa, suggesting that oligomers of CD81 were being formed. It is known that CD81 forms oligomers both *in vivo* and when expressing the protein in *P. pastoris* and solubilising with SMA2000 (Ayub et al., 2020).

The CD81-C104A-D128C mutant was chosen to go forward with because the Cys104 mutation is further away from the crucial Asn18 and Glu219 residues involved in cholesterol binding than Cys97 and it expressed better than the CD81-C97A-D128C mutant (Figure 5.16). Attempts to purify this mutant in SMALPs were unsuccessful. Only faint bands were visible on SDS-PAGE gels in the wash and elution steps. The SDS-PAGE gels show that protein is being sufficiently solubilised, but it appears that almost all of it is in the flow-through fraction, which means that CD81-C104A-D128C is not binding sufficiently to the Ni-NTA resin (Figures 5.17-5.20).

One problem when working with SMA solubilised membrane protein is that excess SMA can interfere with and reduce the binding of the protein of interest to the Ni-NTA resin (Pollock et al., 2018). Solubilisation was done with 2.5% (w/v) SMA2000 but given that excess SMA can interfere with binding to the Ni-NTA resin the concentration was lowered to 1.25% but this did not yield any better results.

Due to these problems encountered when trying to purify CD81-C104A-D128C the solubilisation method was changed from using SMALPs to using the detergent, DDM, which has been used to solubilise and purify CD81-p-null from *P. pastoris* (Jamshad et al., 2008; Ayub et al., 2020). Solubilising in 1% (w/v) DDM resulted in an improvement with some bands visible in the elution steps, particularly when doubling the volume of solubilisation buffer (Figure 5.22). The amount of protein being eluted, however, was not comparable to that which is eluted when solubilising and purifying CD81-p-null.

With expression levels of the mutant CD81 being good it is possible that mutating a Cys residue in the TM region has had an impact on the structure of the protein. Cys residues are often important for protein structural integrity or interactions, exemplified by the fact that Cys

residues rarely mutate over evolutionary time when compared to other more mutable residues, such as Ala (Pearson, 2013). The possibility exists that mutating either of the Cys residues in TM3 to Ala has had a detrimental effect on the structural integrity of CD81. However, mutating both, Cys97 and Cys104, to Ala has no observable effect on the ability of CD81 to bind to partner proteins (Zhu et al., 2012). In another study the Cys residues were mutated to Ser rather than mutating both and there was no observable impact on CD81 interaction (Delandre et al., 2009). In the event of an impact on the structural integrity of CD81 it is possible that the protein is not folding as expected and the His-tag which is normally accessible for Ni-NTA chromatography has become buried and inaccessible.

Size exclusion chromatography is a purification step ordinarily done after an initial purification step, such as Ni-NTA chromatography or ion-exchange chromatography. Rather than performing the initial purification step a small volume of concentrated DDM solubilised CD81-C104A-D128C was loaded straight onto the size exclusion column to see if the recombinant protein could be seen. Initial attempts did not work because the sample that was loaded onto the column was only concentrated to 500 µl and the elutions were collected in 500 µl fractions (Figures 5.23 & 5.24). The 500 µl fractions were most likely too large, meaning that the protein was too dilute to see when it was run on an SDS-PAGE gel.

Concentrating the DDM solubilised CD81-C104A-D128C to a smaller volume of 200 µl and collecting the elutions in 200 µl fractions had a positive effect. There are some bands visible in the first eight 200 µl fractions following a pattern expected when using size exclusion chromatography (Figures 5.25 & 5.26). The larger proteins at approximately 250 kDa appear as denser bands in the earlier fractions, while the smaller proteins at approximately 25 kDa appear as denser bands in the fractions after the larger proteins. This is an indication that the size exclusion chromatography works with regards to separating the solubilised *P. pastoris* membrane by size.

Starting the purification process with size exclusion chromatography could be used going forward but it does present the problem of scalability. In these experiments 3 ml of *P. pastoris* membrane preparation at a concentration of 60 mg/ml was used and it was loaded onto a 1 ml size exclusion column, so these purifications were done on a small scale. Solubilising more membrane and, crucially, concentrating to a small enough volume to load onto a bigger size exclusion column could be problematic.

7.7 Future work on probing the interaction of cholesterol with tetraspanins and its significance for conformational change

There may be other ways to investigate the relationship between the presence of cholesterol in association with CD81 and whether the proteins changes conformation based on the

presence of cholesterol. *P. pastoris* cells synthesise ergosterol rather than cholesterol (Grillitsch et al., 2014) and it is not known whether ergosterol binds to CD81 in a similar manner to that proposed for the binding of cholesterol. The possibility exists to deplete *P. pastoris* membranes of ergosterol such as work done on the mu opioid receptor (hMOR) (Talmont et al., 2020).

In that study the authors used methyl- β -cyclodextrin (M β CD) to deplete ergosterol from the *P. pastoris* membranes, followed by loading the membranes with cholesterol using M β CD-cholesterol complexes. Binding of an agonist to hMOR was investigated in membranes with depleted ergosterol, depleted ergosterol with cholesterol loaded, ergosterol and ergosterol with cholesterol loaded (Talmont et al., 2020). The same principle could be applied to *P. pastoris* membranes overexpressing CD81 using the binding of sE2 of HCV. It is known that sE2 binds to human CD81 (Pileri et al., 1998) and infectivity of HCV is higher when CD81 is said to adopt the open conformation, suggesting that binding of E2 may be more likely in the open conformation (Palor et al., 2020). The expectation, therefore, would be that sE2 binding to CD81 would be higher in *P. pastoris* cells with cholesterol loaded, compared to ergosterol depleted membranes, and possibly membranes with ergosterol.

This could raise the question of whether ergosterol would have the same effect as cholesterol. If there were no difference in binding of sE2 to CD81 in *P. pastoris* membranes with ergosterol or cholesterol then it would pave the way to look at other sterols, such as testosterone that could bind to CD81. It is known that estradiol, another sterol, does not bind to CD81 (Zimmerman et al., 2016) but it would be beneficial to gain a fuller understanding of which sterols do bind to the cavity of CD81. *P. pastoris* membranes overexpressing CD81 can be attached to an ELISA plate (Ayub et al., 2020) allowing for the investigation of sE2 binding after preparing the *P. pastoris* membranes with cholesterol or ergosterol depletion.

One benefit of conducting such an experiment is that it would answer whether sE2 binding to CD81 is dependent on cholesterol. It is known that HCV infectivity is lower when CD81 is said to be in the closed state (Palor et al., 2020) but these experiments were conducted using HCV pseudo particles (HCVpp) in human cell culture that express other proteins, such as SRB1, CLDN1, OCLN that are involved in HCV and could compensate for any deficiency in E2 binding to CD81 to help increase infectivity. Looking at the dependence of cholesterol on sE2 binding in *P. pastoris* membranes has the benefit of not having the other proteins known to be involved in HCV infection present.

Another possibility could be to utilise a *P. pastoris* strain genetically modified to produce cholesterol (Hirz et al., 2013). If ergosterol does not interact with CD81 in a similar manner to cholesterol then it is possible that CD81 expressed in *P. pastoris* could be problematic since

the overall effects of cholesterol on structure and function has yet to be determined experimentally. So, if ergosterol does not bind to CD81 then it is possible that some aspects of CD81 biology are being missed in *P. pastoris* cells that are not modified to produce cholesterol. This is especially important with the possibility of cholesterol providing a stabilising role in the folding of the TM domains (Yang et al., 2020).

If, as discussed earlier, the Cys mutations are having a detrimental effect on the structural integrity of CD81 then the Cys residues could be mutated to something other than Ala. Future work could be done by mutating the Cys residues to Ser rather than Ala since Delandre et al. (2009) chose Ser when mutating just one of the two Cys residues. Ser is more like Cys in terms of its structure given that they are identical except for an -SH sidechain in Cys and an -OH sidechain in Ser. The Cys residues being mutated, however, are in the TM3 domain, a hydrophobic part of the protein and Cys is more hydrophobic than Ser, so Ala could be more suitable since it is also more hydrophobic than Ser (Kyte & Doolittle, 1982).

Due to having to remove a native Cys residue in TM3 of CD81 another solution could be to change which tetraspanin is used to investigate conformational change. CD53, for instance, is well characterised and a structure in the semi-open conformation exists (Yang et al., 2020). CD53 also does not have any Cys residues in the TM domains that may have to be mutated, allowing for the introduction of Cys residues in other regions of the protein, such as the intracellular or extracellular tail.

7.8 Asp122, in CD81, and Asp37 & Ser135, in CD82, bind ganglioside sugar residues

It is known that tetraspanins interact with gangliosides (Odintsova et al., 2006) but the specific nature of those interactions are unclear. Molecular dynamics performed on CD81 showed that gangliosides bind to CD81 and can aid oligomerisation with contact between CD81 and the gangliosides occurring at the SEL and some regions of the LEL, such as the α , δ and ϵ helical regions (Schmidt et al., 2016).

This work is aimed at discovering the specific amino acids involved in the interaction between tetraspanins and gangliosides. The work done by Schmidt et al. (2016) was done using a CD81 protein that would now be considered to be in the open state because their work was done using the model of CD81 before the crystal structure was obtained (Seigneuret, 2006).

Docking the sugars of gangliosides on the closed CD81 crystal structure highlighted the importance of Asp122 in the interaction. Two hydrogen bonds are formed between Asp122 and both galactose and lactose, although only one of those hydrogen bonds is with its side chain because one is with the protein backbone (Figure 6.2). This finding is not surprising because aspartic acid can act as a hydrogen acceptor and modelling of other proteins has

demonstrated its role in interactions with gangliosides. Modelling of the interaction between Aquaporin-1 (AQP1) and GM1 suggests that the interaction occurs at a polar rich extracellular region of AQP1 (Gu et al., 2017). Amino acids, such as Asp, Asn, Ser and Thr are found in that region which can all act as hydrogen donors or acceptors and therefore form hydrogen bonds with gangliosides.

One surprising result is that there is a decrease in the average docking energy when lactose is docked on closed CD81 containing the T57A mutation when compared to when lactose is docked on wild-type closed CD81 (Table 6.1). One possible explanation for this, and other results such as a decrease in the minimum docking energy of lactose on the double mutant, is that the docked sugars dock at another site on the protein. For instance, a region at the apex of TM3 and the LEL contains Gln118, Asn115 and Trp111 which are all amino acids capable of forming hydrogen bonds. It may be the case, therefore, that the Asp122-Thr57 region is the preferential region for ganglioside binding but there are further regions where gangliosides bind if they are unable to bind in the Asp122-Thr57 region.

The next step would be to carry out experiments in the laboratory to test the docking results. This work was begun in this thesis by mutating the Asp122 residue to Ala in CD81-p-null expressed in *P.pastoris*. The work progressed as far as transforming the mutant into *P.pastoris* X33 cells and trying to express the protein (Figures 6.6 & 6.7). The next steps will be to test the interaction between specific gangliosides and CD81-p-null and compare this to the interaction between specific gangliosides and CD81-p-null-D122A. Both proteins will be purified and ganglioside binding will be tested by using dot blots and ELISAs.

Protein-lipid interactions can be investigated using a dot blot assay (Munnik & Wierchowicka, 2013) where serial dilutions of lipids are spotted onto a PVDF or nitrocellulose membrane and left to soak into the membrane for approximately an hour. So, in this instance the gangliosides will be spotted onto a membrane and left for one hour, after which the purified protein will be incubated for one hour with the ganglioside spotted membrane. The purified protein possesses a His-tag which can then be used in a similar manner in a Western blot since the protein should be bound to the ganglioside.

Another method that will be used is a ganglioside ELISA similar to one that is routinely used in toxin studies. *E. coli* heat-labile enterotoxin and cholera toxin have both been used in GM1-ELISA experiments (Ristaino et al., 1983; Dawson, 2015). GM1 is coated to an ELISA plate and the toxin then incubated with the GM1-coated ELISA plate to test for binding. The same experiments will be conducted with a range of gangliosides, such as GM1, GM2 and GM3, and purified tetraspanins.

Work in this thesis was also started on expressing wild-type CD81 in HEK293T cells. Wild-type CD81 was successfully expressed in HEK293T cells 24 hours after transient transfection (Figure 6.8). In the future this system will be used to analyse the native lipid composition surrounding CD81. After expression, cells will be lysed and CD81 will be purified for thin layer chromatography (TLC) analysis. After running the samples on a TLC plate a resorcinol reagent is sprayed onto the plate to be able to visualise any gangliosides present in the samples (Cho et al., 2010). This will provide information of which gangliosides are favoured in binding to CD81. In a similar manner to the *P. pastoris* work, mutants can be created to see the effect this has on ganglioside composition in a native lipid environment.

The sugar residues were docked on a homology model of CD82 to test which amino acids were involved in CD82-ganglioside interactions. The homology model had some problems with it because of its large loop region in the LEL between the β helix and the ϵ helix (Figures 6.10 & 6.11). The majority of the model was a good quality as shown by the high QMEANBrane scores for all of the structured regions. The TM regions, in particular, scored highly. It has been shown by sequence alignment of all the human TM regions that they are similar and there are three full-length tetraspanin structures (CD81, CD9 and CD53) which have provided a good structural basis for the TM region. Consequently, modelling the structured regions of tetraspanins should not prove difficult because of their similarity. The problematic region is the loop in the LEL between the β helix and the ϵ helix. This is the sequentially variable region with a range of different lengths and amino acid compositions seen. CD82, for instance, has the fourth longest loop region with 52 amino acids which makes it difficult to model, especially when it is a largely unstructured region.

Docking galactose and lactose with the CD82 model showed a propensity for the sugar molecules to form a bond with CD82 at the N-terminal end of the β helix and the SEL. Ser135 in the β helix and Asp37 in the SEL form hydrogen bonds with the galactose residue when it is docked on its own and when it is docked as part of the lactose molecule (Figure 6.13). The interaction between Ser135 and galactose remains similar when both molecules are docked. In both instances it forms a hydrogen bond with the OH group on the C6 carbon with only a 0.02 Å difference between the bond length between the galactose docking and the lactose docking.

Further research into the CD82-ganglioside interaction will be conducted using the same methods as those discussed earlier for the CD81-ganglioside interaction by utilising dot blots and ELISAs. Additional research into the interaction can be done by improving the CD82 model. Recently, the structure of the Tspan15 LEL has been published (Lipper et al., 2022). Tspan15 is another tetraspanin with a long loop region. Its loop region is 48 amino acids long,

just four residues shorter than the loop region in CD82. It may be beneficial when moving forward with this research to use the Tspan15 LEL structure as a model for the CD82 LEL. Having a slightly poor quality loop region in CD82 was not detrimental in the experiments conducted in this thesis because the focus was always going to be in the structured regions close to the membrane interface.

The work in this thesis only looked at one galactose or one lactose molecule docking to the proteins. Due to this the loop region didn't necessarily need to be modelled to a high quality. Obtaining a better CD82 structure model by using Tspan15 as a guide would allow for a more sophisticated look at CD82-ganglioside interactions. Similar molecular modelling experiments to those performed by Schmidt et al. (2016) and those performed on AQP1 (Gu et al., 2017) could be performed to look at the CD82-ganglioside interaction in an environment with more gangliosides and other lipids.

7.9 Overall Conclusion

This thesis has made an advancement in the understanding of structure and sequence conservation in human tetraspanins. By combining methods of sequence conservation and structural conservation this thesis has introduced a universal amino acid numbering system based on the most conserved residue in each structured region that is present in all human tetraspanins. An advancement in our understanding of disulfide bond arrangement has been made in this thesis. It has mostly been assumed that all human tetraspanins follow the same disulfide bond arrangement with the second Cys residue in the CCG-motif at the end of the β helix forming a disulfide bond with the first Cys residue of the middle two in the middle of the loop between the β helix and the ϵ helix. The modelled structures of the TspanC6-CxC LELs presented in this thesis have shown that the second Cys in the CCG-motif can form, and may prefer, to form a disulfide bond with the second Cys residue of the middle two Cys residues in the loop region in the LEL.

This thesis has improved the purification process of CD81 solubilised with SMA. The initial stages of SMA hydrolysis have been improved after demonstrating that SMA that has been hydrolysed by autoclaving is comparable when solubilising CD81 when compared to SMA prepared by refluxing. This process is easier and less hands-on which is an improvement. Once CD81 has been solubilised with autoclaved SMA an improvement has been with regards to increasing the purity CD81 by adding imidazole to the binding buffer and increasing the concentration of imidazole in each of the wash steps when purifying the protein.

Progress has been made producing CD81 mutants that can be used in PELDOR experiments to investigate the opening and closing mechanism of the protein as it switches between conformational states. Difficulties arose when trying to purify the mutants but progress was

made by switching from SMA solubilisation to DDM. Further work will be required to obtain protein for PELDOR experiments.

Amino acids involved in ganglioside interactions with CD81 and CD82 have been identified. Asp122, in CD81, and Asp37 and Ser135, in CD82, were identified when docking ganglioside sugar residues with CD81 and CD82. Further work can be conducted to investigate these residues further with the goal of finding regions of tetraspanins that can be targeted by therapeutics to disrupt the formation of TEMs, which would have benefits for overcoming numerous illnesses and infectious diseases.

Chapter 8 – References

- Altenbach, C., Marti, T., Khorana, H.G., Hubbell, W.L., 1990. Transmembrane Protein Structure: Spin Labeling of Bacteriorhodopsin Mutants. *Science* 248, 1088–1092. <https://doi.org/10.1126/science.2160734>
- Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Andersen, O.S., Koeppe, R.E., 2007. Bilayer Thickness and Membrane Protein Function: An Energetic Perspective. *Annu. Rev. Biophys. Biomol. Struct.* 36, 107–130. <https://doi.org/10.1146/annurev.biophys.36.040306.132643>
- Ayub, H., Clare, M., Milic, I., Chmel, N.P., Böning, H., Devitt, A., Krey, T., Bill, R.M., Rothnie, A.J., 2020. CD81 extracted in SMALP nanodiscs comprises two distinct protein populations within a lipid environment enriched with negatively charged headgroups. *Biochim. Biophys. Acta BBA - Biomembr.* 1862, 183419. <https://doi.org/10.1016/j.bbamem.2020.183419>
- Baier, C.J., Fantini, J., Barrantes, F.J., 2011. Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor. *Sci. Rep.* 1, 69. <https://doi.org/10.1038/srep00069>
- Baldwin, J.M., Schertler, G.F.X., Unger, V.M., 1997. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors 1 Edited by R. Huber. *J. Mol. Biol.* 272, 144–164. <https://doi.org/10.1006/jmbi.1997.1240>
- Ballesteros, J.A., Weinstein, H., 1995. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. pp. 366–428. [https://doi.org/10.1016/S1043-9471\(05\)80049-7](https://doi.org/10.1016/S1043-9471(05)80049-7)
- Bari, R., Guo, Q., Xia, B., Zhang, Y.H., Giesert, E.E., Levy, S., Zheng, J.J., Zhang, X.A., 2011. Tetraspanins regulate the protrusive activities of cell membrane. *Biochem. Biophys. Res. Commun.* 415, 619–626. <https://doi.org/10.1016/j.bbrc.2011.10.121>
- Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A., Cosset, F.-L., 2003. Cell Entry of Hepatitis C Virus Requires a Set of Co-receptors That Include the CD81 Tetraspanin and the SR-B1 Scavenger Receptor. *J. Biol. Chem.* 278, 41624–41630. <https://doi.org/10.1074/jbc.M305289200>
- Basith, S., Cui, M., Macalino, S.J.Y., Park, J., Clavio, N.A.B., Kang, S., Choi, S., 2018. Exploring G Protein-Coupled Receptors (GPCRs) Ligand Space via Cheminformatics Approaches: Impact on Rational Drug Design. *Front. Pharmacol.* 9. <https://doi.org/10.3389/fphar.2018.00128>
- Benedicto, I., Molina-Jiménez, F., Bartosch, B., Cosset, F.-L., Lavillette, D., Prieto, J., Moreno-Otero, R., Valenzuela-Fernández, A., Aldabe, R., López-Cabrera, M., Majano, P.L., 2009. The Tight Junction-Associated Protein Occludin Is Required for a Postbinding Step in Hepatitis C Virus Entry and Infection. *J. Virol.* 83, 8012–8020. <https://doi.org/10.1128/JVI.00038-09>
- Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27, 343–350. <https://doi.org/10.1093/bioinformatics/btq662>
- Bao, H., Dalal, K., Wang, V., Rouiller, I., Duong, F., 2013. The maltose ABC transporter: Action of membrane lipids on the transporter stability, coupling and ATPase activity. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1828, 1723–1730. <https://doi.org/10.1016/j.bbamem.2013.03.024>
- Barniol-Xicotá, M., Verhelst, S.H.L., 2018. Stable and Functional Rhomboid Proteases in Lipid Nanodiscs by Using Diisobutylene/Maleic Acid Copolymers. *Journal of the American Chemical Society* 140, 14557–14561. <https://doi.org/10.1021/jacs.8b08441>
- Berdichevski, F., Odintsova, E., 1999. Characterization of Integrin–Tetraspanin Adhesion Complexes. *J. Cell Biol.* 146, 477–492. <https://doi.org/10.1083/jcb.146.2.477>

- Berditchevski, F., Odintsova, E., Sawada, S., Gilbert, E., 2002. Expression of the Palmitoylation-deficient CD151 Weakens the Association of $\alpha\beta 1$ Integrin with the Tetraspanin-enriched Microdomains and Affects Integrin-dependent Signaling. *J. Biol. Chem.* 277, 36991–37000. <https://doi.org/10.1074/jbc.M205265200>
- Braun, P., von Heijne, G., 1999. The Aromatic Residues Trp and Phe Have Different Effects on the Positioning of a Transmembrane Helix in the Microsomal Membrane. *Biochemistry* 38, 9778–9782. <https://doi.org/10.1021/bi990923a>
- Brazzoli, M., Bianchi, A., Filippini, S., Weiner, A., Zhu, Q., Pizza, M., Crotta, S., 2008. CD81 Is a Central Regulator of Cellular Events Required for Hepatitis C Virus Infection of Human Hepatocytes. *J. Virol.* 82, 8316–8329. <https://doi.org/10.1128/JVI.00665-08>
- Brown, D.A., London, E., 2000. Structure and Function of Sphingolipid- and Cholesterol-rich Membrane Rafts. *J. Biol. Chem.* 275, 17221–17224. <https://doi.org/10.1074/jbc.R000005200>
- Buchan, D.W.A., Jones, D.T., 2019. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* 47, W402–W407. <https://doi.org/10.1093/nar/gkz297>
- Butera, D., Cook, K.M., Chiu, J., Wong, J.W.H., Hogg, P.J., 2014. Control of blood proteins by functional disulfide bonds. *Blood* 123, 2000–7. <https://doi.org/10.1182/blood-2014-01-549816>
- Cantor, R.S., 1999. Lipid Composition and the Lateral Pressure Profile in Bilayers. *Biophys. J.* 76, 2625–2639. [https://doi.org/10.1016/S0006-3495\(99\)77415-1](https://doi.org/10.1016/S0006-3495(99)77415-1)
- Cevik, S.I., Keskin, N., Belkaya, S., Ozlu, M.I., Deniz, E., Tazebay, U.H., Erman, B., 2012. CD81 interacts with the T cell receptor to suppress signaling. *PloS One* 7, e50396. <https://doi.org/10.1371/journal.pone.0050396>
- Charrin, S., Jouannet, S., Boucheix, C., Rubinstein, E., 2014. Tetraspanins at a glance. *Journal of Cell Science.* <https://doi.org/10.1242/jcs.154906>
- Charrin, S., Latil, M., Soave, S., Poleskaya, A., Chrétien, F., Boucheix, C., Rubinstein, E., 2013. Normal muscle regeneration requires tight control of muscle cell fusion by tetraspanins CD9 and CD81. *Nature Communications* 4, 1674. <https://doi.org/10.1038/ncomms2675>
- Charrin, S., Manié, S., Oualid, M., Billard, M., Boucheix, C., Rubinstein, E., 2002. Differential stability of tetraspanin/tetraspanin interactions: role of palmitoylation. *FEBS Lett.* 516, 139–144. [https://doi.org/10.1016/S0014-5793\(02\)02522-X](https://doi.org/10.1016/S0014-5793(02)02522-X)
- Charrin, S., Manié, S., Thiele, C., Billard, M., Gerlier, D., Boucheix, C., Rubinstein, E., 2003. A physical and functional link between cholesterol and tetraspanins. *Eur. J. Immunol.* 33, 2479–2489. <https://doi.org/10.1002/eji.200323884>
- Chernomordik, L., Chanturiya, A., Green, J., Zimmerberg, J., 1995. The hemifusion intermediate and its conversion to complete fusion: regulation by membrane composition. *Biophys. J.* 69, 922–929. [https://doi.org/10.1016/S0006-3495\(95\)79966-0](https://doi.org/10.1016/S0006-3495(95)79966-0)
- Cho, J.H., Kim, J.S., Lee, Y.C., Oh, K.B., Kwak, D. hoon, Kim, W.S., Hwang, S.S., Ko, K., Chang, K.T., Choo, Y.-K., 2010. Differential expression patterns of gangliosides in the tissues and cells of NIH-mini pig kidneys. *Anim. Cells Syst.* 14, 83–89. <https://doi.org/10.1080/19768354.2010.486935>
- Chowdhury, B., Garai, G., 2017. A review on multiple sequence alignment from the perspective of genetic algorithm. *Genomics* 109, 419–431. <https://doi.org/10.1016/j.ygeno.2017.06.007>
- Claxton, D.P., Kazmier, K., Mishra, S., Mchaourab, H.S., 2015. Navigating Membrane Protein Structure, Dynamics, and Energy Landscapes Using Spin Labeling and EPR Spectroscopy. *Methods Enzymol.* 564, 349–87. <https://doi.org/10.1016/bs.mie.2015.07.026>
- Columbus, L., Hubbell, W.L., 2002. A new spin on protein dynamics. *Trends Biochem. Sci.* 27, 288–295. [https://doi.org/10.1016/S0968-0004\(02\)02095-9](https://doi.org/10.1016/S0968-0004(02)02095-9)
- Craig, A.F., Clark, E.E., Sahu, I.D., Zhang, R., Frantz, N.D., Al-Abdul-Wahid, M.S., Dabney-Smith, C., Konkolewicz, D., Lorigan, G.A., 2016. Tuning the size of styrene-maleic acid copolymer-lipid nanoparticles (SMALPs) using RAFT polymerization for biophysical

- studies. *Biochim. Biophys. Acta BBA - Biomembr.* 1858, 2931–2939. <https://doi.org/10.1016/j.bbamem.2016.08.004>
- Cummins, P.M., 2012. Occludin: One Protein, Many Forms. *Mol. Cell. Biol.* 32, 242–250. <https://doi.org/10.1128/MCB.06029-11>
- Cunha, E.S., Sfriso, P., Rojas, A.L., Roversi, P., Hospital, A., Orozco, M., Abrescia, N.G.A., 2017. Mechanism of Structural Tuning of the Hepatitis C Virus Human Cellular Receptor CD81 Large Extracellular Loop. *Structure* 25, 53–65. <https://doi.org/10.1016/j.str.2016.11.003>
- Cuttillo, G., Saariaho, A.-H., Meri, S., 2020. Physiology of gangliosides and the role of antiganglioside antibodies in human diseases. *Cellular & Molecular Immunology* 17, 313–322. <https://doi.org/10.1038/s41423-020-0388-9>
- Daly, R., Hearn, M.T.W., 2005. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* 18, 119–138. <https://doi.org/10.1002/jmr.687>
- Daugelaite, J., O' Driscoll, A., Sleator, R.D., 2013. An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomath.* 2013, 1–14. <https://doi.org/10.1155/2013/615630>
- Davis, C., Harris, H.J., Hu, K., Drummer, H.E., McKeating, J.A., Mullins, J.G.L., Balfe, P., 2012. *In silico* directed mutagenesis identifies the <scp>CD</scp> 81/ Claudin-1 hepatitis <scp>C</scp> virus receptor interface. *Cell. Microbiol.* 14, 1892–1903. <https://doi.org/10.1111/cmi.12008>
- Dawson, R.M., 2005. Characterization of the binding of cholera toxin to ganglioside GM1 immobilized onto microtitre plates. *J. Appl. Toxicol.* 25, 30–38. <https://doi.org/10.1002/jat.1015>
- Delandre, C., Penabaz, T.R., Passarelli, A.L., Chapes, S.K., Clem, R.J., 2009. Mutation of juxtamembrane cysteines in the tetraspanin CD81 affects palmitoylation and alters interaction with other proteins at the cell surface. *Exp. Cell Res.* 315, 1953–1963. <https://doi.org/10.1016/j.yexcr.2009.03.013>
- Deng, H., Jia, Y., Zhang, Y., 2018. Protein structure prediction. *Int. J. Mod. Phys. B* 32. <https://doi.org/10.1142/S021797921840009X>
- DeSalle, R., Mares, R., Garcia-España, A., 2010. Evolution of cysteine patterns in the large extracellular loop of tetraspanins from animals, fungi, plants and single-celled eukaryotes. *Mol. Phylogenet. Evol.* 56, 486–491. <https://doi.org/10.1016/j.ympev.2010.02.015>
- Do, C.B., Katoh, K., 2008. Protein Multiple Sequence Alignment. pp. 379–413. https://doi.org/10.1007/978-1-59745-398-1_25
- Drew, D., Boudker, O., 2016. Shared Molecular Mechanisms of Membrane Transporters. *Annu. Rev. Biochem.* 85, 543–572. <https://doi.org/10.1146/annurev-biochem-060815-014520>
- Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., Anishchenko, I., Baker, D., Yang, J., 2021. The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* 16, 5634–5651. <https://doi.org/10.1038/s41596-021-00628-9>
- Engelman, D.M., 2005. Membranes are more mosaic than fluid. *Nature* 438, 578–580. <https://doi.org/10.1038/nature04394>
- Fantini, J., Barrantes, F.J., 2013. How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front. Physiol.* 4, 31. <https://doi.org/10.3389/fphys.2013.00031>
- Fantini, J., Yahi, N., Garmy, N., 2013. Cholesterol accelerates the binding of Alzheimer's β -amyloid peptide to ganglioside GM1 through a universal hydrogen-bond-dependent sterol tuning of glycolipid conformation. *Front. Physiol.* 4. <https://doi.org/10.3389/fphys.2013.00120>
- Felmlee, D., Hafirassou, M., Lefevre, M., Baumert, T., Schuster, C., 2013. Hepatitis C Virus, Cholesterol and Lipoproteins — Impact for the Viral Life Cycle and Pathogenesis of Liver Disease. *Viruses* 5, 1292–1324. <https://doi.org/10.3390/v5051292>

- Fénéant, L., Levy, S., Cocquerel, L., 2014. CD81 and Hepatitis C Virus (HCV) Infection. *Viruses* 6, 535–572. <https://doi.org/10.3390/v6020535>
- Flannery, A.R., Czibener, C., Andrews, N.W., 2010. Palmitoylation-dependent association with CD63 targets the Ca²⁺ sensor synaptotagmin VII to lysosomes. *Journal of Cell Biology* 191, 599–613. <https://doi.org/10.1083/jcb.201003021>
- Florin, L., Lang, T., 2018. Tetraspanin Assemblies in Virus Infection. *Front. Immunol.* 9. <https://doi.org/10.3389/fimmu.2018.01140>
- Forstner, M., Leder, L., Mayr, L.M., 2007. Optimization of protein expression systems for modern drug discovery. *Expert Rev. Proteomics* 4, 67–78. <https://doi.org/10.1586/14789450.4.1.67>
- Frick, M., Schmidt, C., 2019. Mass spectrometry—A versatile tool for characterising the lipid environment of membrane protein assemblies. *Chem. Phys. Lipids* 221, 145–157. <https://doi.org/10.1016/j.chemphyslip.2019.04.001>
- Fus-Kujawa, A., Prus, P., Bajdak-Rusinek, K., Teper, P., Gawron, K., Kowalczyk, A., Sieron, A.L., 2021. An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front. Bioeng. Biotechnol.* 9. <https://doi.org/10.3389/fbioe.2021.701031>
- Gao, X., Dong, X., Li, X., Liu, Z., Liu, H., 2020. Prediction of disulfide bond engineering sites using a machine learning method. *Sci. Rep.* 10, 10330. <https://doi.org/10.1038/s41598-020-67230-z>
- Glaenger, J., Peter, M.F., Hagelueken, G., 2018. Studying structure and function of membrane proteins with PELDOR/DEER spectroscopy – The crystallographers’ perspective. *Methods* 147, 163–175. <https://doi.org/10.1016/j.ymeth.2018.03.002>
- Goldberg, A.F.X., Molday, R.S., 1996. Defective subunit assembly underlies a digenic form of retinitis pigmentosa linked to mutations in peripherin/rds and rom-1. *Proc. Natl. Acad. Sci.* 93, 13726–13730. <https://doi.org/10.1073/pnas.93.24.13726>
- Grillitsch, K., Tarazona, P., Klug, L., Wriessnegger, T., Zellnig, G., Leitner, E., Feussner, I., Daum, G., 2014. Isolation and characterization of the plasma membrane from the yeast *Pichia pastoris*. *Biochim. Biophys. Acta BBA - Biomembr.* 1838, 1889–1897. <https://doi.org/10.1016/j.bbamem.2014.03.012>
- Gronnier, J., Germain, V., Gouguet, P., Cacas, J.-L., Mongrand, S., 2016. GIPC: Glycosyl Inositol Phospho Ceramides, the major sphingolipids on earth. *Plant Signal. Behav.* 11, e1152438. <https://doi.org/10.1080/15592324.2016.1152438>
- Gu, R.-X., Ingólfsson, H.I., de Vries, A.H., Marrink, S.J., Tieleman, D.P., 2017. Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. *J. Phys. Chem. B* 121, 3262–3275. <https://doi.org/10.1021/acs.jpcc.6b07142>
- Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. <https://doi.org/10.1002/elps.1150181505>
- Gulamhussein, A.A., Meah, D., Soja, D.D., Fenner, S., Saidani, Z., Akram, A., Lallie, S., Mathews, A., Painter, C., Liddar, M.K., Mohammed, Z., Chiu, L.K., Sumar, S.S., Healy, H., Hussain, N., Patel, J.H., Hall, S.C.L., Dafforn, T.R., Rothnie, A.J., 2019. Examining the stability of membrane proteins within SMALPs. *Eur. Polym. J.* 112, 120–125. <https://doi.org/10.1016/j.eurpolymj.2018.12.008>
- Gullingsrud, J., Schulten, K., 2004. Lipid Bilayer Pressure Profiles and Mechanosensitive Channel Gating. *Biophysical Journal* 86, 3496–3509. <https://doi.org/10.1529/biophysj.103.034322>
- Hall, S.C.L., Tognoloni, C., Charlton, J., Bragginton, É.C., Rothnie, A.J., Sridhar, P., Wheatley, M., Knowles, T.J., Arnold, T., Edler, K.J., Dafforn, T.R., 2018. An acid-compatible copolymer for the solubilization of membranes and proteins into lipid bilayer-containing nanoparticles. *Nanoscale* 10, 10609–10619. <https://doi.org/10.1039/C8NR01322E>
- Hall, S.E., Roberts, K., Vaidehi, N., 2009. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. *J. Mol. Graph. Model.* 27, 944–950. <https://doi.org/10.1016/j.jmgm.2009.02.004>

- Hama, H., 2010. Fatty acid 2-Hydroxylation in mammalian sphingolipid biology. *Biochim. Biophys. Acta* 1801, 405–14. <https://doi.org/10.1016/j.bbalip.2009.12.004>
- Hammes, G.G., 2002. Multiple Conformational Changes in Enzyme Catalysis. *Biochemistry* 41, 8221–8228. <https://doi.org/10.1021/bi0260839>
- Hemler, M.E., 2008. Targeting of tetraspanin proteins — potential benefits and strategies. *Nat. Rev. Drug Discov.* 7, 747–758. <https://doi.org/10.1038/nrd2659>
- Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* 89, 10915–10919. <https://doi.org/10.1073/pnas.89.22.10915>
- Hentschel, A., Zahedi, R.P., Ahrends, R., 2016. Protein lipid modifications—More than just a greasy ballast. *PROTEOMICS* 16, 759–782. <https://doi.org/10.1002/pmic.201500353>
- Heo, L., Park, H., Seok, C., 2013. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, W384–W388. <https://doi.org/10.1093/nar/gkt458>
- Hirz, M., Richter, G., Leitner, E., Wriessnegger, T., Pichler, H., 2013. A novel cholesterol-producing *Pichia pastoris* strain is an ideal host for functional expression of human Na,K-ATPase $\alpha 3\beta 1$ isoform. *Appl. Microbiol. Biotechnol.* 97, 9465–9478. <https://doi.org/10.1007/s00253-013-5156-7>
- Huang, C., Hays, F.A., Tomasek, J.J., Benyajati, S., Zhang, X.A., 2020. Tetraspanin CD82 interaction with cholesterol promotes extracellular vesicle-mediated release of ezrin to inhibit tumour cell movement. *J. Extracell. Vesicles* 9, 1692417. <https://doi.org/10.1080/20013078.2019.1692417>
- Huang, S., Yuan, S., Dong, M., Su, J., Yu, C., Shen, Y., Xie, X., Yu, Y., Yu, X., Chen, S., Zhang, S., Pontarotti, P., Xu, A., 2005. The phylogenetic analysis of tetraspanins projects the evolution of cell–cell interactions from unicellular to multicellular organisms. *Genomics* 86, 674–684. <https://doi.org/10.1016/j.ygeno.2005.08.004>
- Hubbell, W.L., Cafiso, D.S., Altenbach, C., 2000. Identifying conformational changes with site-directed spin labeling. *Nat. Struct. Biol.* 7, 735–739. <https://doi.org/10.1038/78956>
- Huh, S.-H., Do, H.-J., Lim, H.-Y., Kim, D.-K., Choi, S.-J., Song, H., Kim, N.-H., Park, J.-K., Chang, W.-K., Chung, H.-M., Kim, J.-H., 2007. Optimization of 25kDa linear polyethylenimine for efficient gene delivery. *Biologicals* 35, 165–171. <https://doi.org/10.1016/j.biologicals.2006.08.004>
- Imberty, A., Varrot, A., 2008. Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* 18, 567–576. <https://doi.org/10.1016/j.sbi.2008.08.001>
- Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F.H., Mordalski, S., Pin, J.-P., Stevens, R.C., Vriend, G., Gloriam, D.E., 2015. Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* 36, 22–31. <https://doi.org/10.1016/j.tips.2014.11.001>
- Israelachvili, J.N., Mitchell, D.J., Ninham, B.W., 1977. Theory of self-assembly of lipid bilayers and vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 470, 185–201. [https://doi.org/10.1016/0005-2736\(77\)90099-2](https://doi.org/10.1016/0005-2736(77)90099-2)
- Jamshad, M., Rajesh, S., Stamataki, Z., McKeating, J.A., Dafforn, T., Overduin, M., Bill, R.M., 2008. Structural characterization of recombinant human CD81 produced in *Pichia pastoris*. *Protein Expr. Purif.* 57, 206–216. <https://doi.org/10.1016/j.pep.2007.10.013>
- Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A.S., De Fabritiis, G., 2017. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* 33, 3036–3042. <https://doi.org/10.1093/bioinformatics/btx350>
- Kalipatnapu, S., Chattopadhyay, A., 2005. Membrane Protein Solubilization: Recent Advances and Challenges in Solubilization of Serotonin1A Receptors. *IUBMB Life Int. Union Biochem. Mol. Biol. Life* 57, 505–512. <https://doi.org/10.1080/15216540500167237>
- Kapadia, S.B., Barth, H., Baumert, T., McKeating, J.A., Chisari, F.V., 2007. Initiation of Hepatitis C Virus Infection Is Dependent on Cholesterol and Cooperativity between CD81 and Scavenger Receptor B Type I. *J. Virol.* 81, 374–383. <https://doi.org/10.1128/JVI.01134-06>

- Kawakami, Y., Kawakami, K., Steelant, W.F.A., Ono, M., Baek, R.C., Handa, K., Withers, D.A., Hakomori, S., 2002. Tetraspanin CD9 Is a "Proteolipid," and Its Interaction with $\alpha 3$ Integrin in Microdomain Is Promoted by GM3 Ganglioside, Leading to Inhibition of Laminin-5-dependent Cell Motility. *J. Biol. Chem.* 277, 34349–34358. <https://doi.org/10.1074/jbc.M200771200>
- Khan, K.H., 2013. Gene expression in Mammalian cells and its applications. *Adv. Pharm. Bull.* 3, 257–63. <https://doi.org/10.5681/apb.2013.042>
- Khutornenko, A.A., Roudko, V.V., Chernyak, B.V., Vartapetian, A.B., Chumakov, P.M., Evstafieva, A.G., 2010. Pyrimidine biosynthesis links mitochondrial respiration to the p53 pathway. *Proc. Natl. Acad. Sci.* 107, 12828–12833. <https://doi.org/10.1073/pnas.0910885107>
- Kim, C., Schmidt, T., Cho, E.-G., Ye, F., Ulmer, T.S., Ginsberg, M.H., 2011. Basic amino-acid side chains regulate transmembrane integrin signalling. *Nature* 481, 209–13. <https://doi.org/10.1038/nature10697>
- Kitadokoro, K., 2001. CD81 extracellular domain 3D structure: insight into the tetraspanin superfamily structural motifs. *EMBO J.* 20, 12–18. <https://doi.org/10.1093/emboj/20.1.12>
- Kopf, A.H., Koorengevel, M.C., van Walree, C.A., Dafforn, T.R., Killian, J.A., 2019. A simple and convenient method for the hydrolysis of styrene-maleic anhydride copolymers to styrene-maleic acid copolymers. *Chem. Phys. Lipids* 218, 85–90. <https://doi.org/10.1016/j.chemphyslip.2018.11.011>
- Kovalenko, O.V., Metcalf, D.G., DeGrado, W.F., Hemler, M.E., 2005. Structural organization and interactions of transmembrane domains in tetraspanin proteins. *BMC Struct. Biol.* 5, 11. <https://doi.org/10.1186/1472-6807-5-11>
- KOVALENKO, O.V., YANG, X., KOLESNIKOVA, T.V., HEMLER, M.E., 2004. Evidence for specific tetraspanin homodimers: inhibition of palmitoylation makes cysteine residues available for cross-linking. *Biochem. J.* 377, 407–417. <https://doi.org/10.1042/bj20031037>
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹Edited by F. Cohen. *J. Mol. Biol.* 305, 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Kuhlman, B., Bradley, P., 2019. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* 20, 681–697. <https://doi.org/10.1038/s41580-019-0163-x>
- Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Langelaan, D.N., Wieczorek, M., Blouin, C., Rainey, J.K., 2010. Improved Helix and Kink Characterization in Membrane Proteins Allows Evaluation of Kink Sequence Predictors. *J. Chem. Inf. Model.* 50, 2213–2220. <https://doi.org/10.1021/ci100324n>
- Lau, L.-M., Wee, J.L., Wright, M.D., Moseley, G.W., Hogarth, P.M., Ashman, L.K., Jackson, D.E., 2004. The tetraspanin superfamily member CD151 regulates outside-in integrin $\alpha \text{IIb}\beta 3$ signaling and platelet function. *Blood* 104, 2368–2375. <https://doi.org/10.1182/blood-2003-12-4430>
- Lee, G.R., Seok, C., 2016. Galaxy7TM: flexible GPCR–ligand docking by structure refinement. *Nucleic Acids Res.* 44, W502–W506. <https://doi.org/10.1093/nar/gkw360>
- Lee, S.C., Knowles, T.J., Postis, V.L.G., Jamshad, M., Parslow, R.A., Lin, Y., Goldman, A., Sridhar, P., Overduin, M., Muench, S.P., Dafforn, T.R., 2016. A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat. Protoc.* 11, 1149–1162. <https://doi.org/10.1038/nprot.2016.070>
- Lemmon, M.A., Schlessinger, J., 2010. Cell Signaling by Receptor Tyrosine Kinases. *Cell* 141, 1117–1134. <https://doi.org/10.1016/j.cell.2010.06.011>
- Li, H., Papadopoulos, V., 1998. Peripheral-Type Benzodiazepine Receptor Function in Cholesterol Transport. Identification of a Putative Cholesterol Recognition/Interaction Amino Acid Sequence and Consensus Pattern ¹. *Endocrinology* 139, 4991–4997. <https://doi.org/10.1210/endo.139.12.6390>

- Linder, M.E., Deschenes, R.J., 2003. New Insights into the Mechanisms of Protein Palmitoylation. *Biochemistry* 42, 4311–4320. <https://doi.org/10.1021/bi034159a>
- Lingwood, C.A., 2011. Glycosphingolipid Functions. *Cold Spring Harb. Perspect. Biol.* 3, a004788–a004788. <https://doi.org/10.1101/cshperspect.a004788>
- Lipper, C.H., Gabriel, K.-H., Seegar, T.C.M., Dürr, K.L., Tomlinson, M.G., Blacklow, S.C., 2022. Crystal structure of the Tspan15 LEL domain reveals a conserved ADAM10 binding site. *Structure* 30, 206-214.e4. <https://doi.org/10.1016/j.str.2021.10.007>
- Liu, N.-J., Wang, N., Bao, J.-J., Zhu, H.-X., Wang, L.-J., Chen, X.-Y., 2020. Lipidomic Analysis Reveals the Importance of GIPCs in Arabidopsis Leaf Extracellular Vesicles. *Mol. Plant* 13, 1523–1532. <https://doi.org/10.1016/j.molp.2020.07.016>
- Lopes, P.E.M., Guvench, O., MacKerell, A.D., 2015. Current Status of Protein Force Fields for Molecular Dynamics Simulations. pp. 47–71. https://doi.org/10.1007/978-1-4939-1465-4_3
- Lopez, P.H., Schnaar, R.L., 2009. Gangliosides in cell recognition and membrane protein regulation. *Curr. Opin. Struct. Biol.* 19, 549–557. <https://doi.org/10.1016/j.sbi.2009.06.001>
- Malhotra, R., 2012. Membrane Glycolipids: Functional Heterogeneity: A Review. *Biochem. Anal. Biochem.* 1. <https://doi.org/10.4172/2161-1009.1000108>
- Marjon, K.D., Termini, C.M., Karlen, K.L., Saito-Reis, C., Soria, C.E., Lidke, K.A., Gillette, J.M., 2016. Tetraspanin CD82 regulates bone marrow homing of acute myeloid leukemia by modulating the molecular organization of N-cadherin. *Oncogene* 35, 4132–4140. <https://doi.org/10.1038/onc.2015.449>
- Matsumoto, K., Kusaka, J., Nishibori, A., Hara, H., 2006. Lipid domains in bacterial membranes. *Mol. Microbiol.* 61, 1110–1117. <https://doi.org/10.1111/j.1365-2958.2006.05317.x>
- Matthews, A.L., Szyroka, J., Collier, R., Noy, P.J., Tomlinson, M.G., 2017. Scissor sisters: regulation of ADAM10 by the TspanC8 tetraspanins. *Biochem. Soc. Trans.* 45, 719–730. <https://doi.org/10.1042/BST20160290>
- McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405. <https://doi.org/10.1093/bioinformatics/16.4.404>
- Merritt, E.A., Kuhn, P., Sarfaty, S., Erbe, J.L., Holmes, R.K., Hol, W.G.J., 1998. The 1.25 Å resolution refinement of the cholera toxin B-pentamer: evidence of peptide backbone strain at the receptor-binding site. *J. Mol. Biol.* 282, 1043–1059. <https://doi.org/10.1006/jmbi.1998.2076>
- Mezei, M., 2003. A new method for mapping macromolecular topography. *J. Mol. Graph. Model.* 21, 463–472. [https://doi.org/10.1016/S1093-3263\(02\)00203-6](https://doi.org/10.1016/S1093-3263(02)00203-6)
- Miljan, E.A., Meillet, E.J., Mania-Farnell, B., George, D., Yamamoto, H., Simon, H.-G., Bremer, E.G., 2002. Interaction of the Extracellular Domain of the Epidermal Growth Factor Receptor with Gangliosides. *J. Biol. Chem.* 277, 10108–10113. <https://doi.org/10.1074/jbc.M111669200>
- Min, G., Wang, H., Sun, T.-T., Kong, X.-P., 2006. Structural basis for tetraspanin functions as revealed by the cryo-EM structure of uroplakin complexes at 6-Å resolution. *J. Cell Biol.* 173, 975–983. <https://doi.org/10.1083/jcb.200602086>
- Miyado, K., Yamada, G., Yamada, S., Hasuwa, H., Nakamura, Y., Ryu, F., Suzuki, K., Kosai, K., Inoue, K., Ogura, A., Okabe, M., Mekada, E., 2000. Requirement of CD9 on the Egg Plasma Membrane for Fertilization. *Science* 287, 321–324. <https://doi.org/10.1126/science.287.5451.321>
- Moller, S., Croning, M.D.R., Apweiler, R., 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17, 646–653. <https://doi.org/10.1093/bioinformatics/17.7.646>
- Montenegro, M.F., Moral-Naranjo, M.T., Campoy, F.J., Muñoz-Delgado, E., Vidal, C.J., 2014. The lipid raft-bound alkaline phosphatase activity increases and the level of transcripts remains unaffected in liver of merosin-deficient LAMA2dy mouse. *Chem. Biol. Interact.* 216, 1–8. <https://doi.org/10.1016/j.cbi.2014.03.008>

- Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., Stewart, P.D.S., 2014. Membrane protein structure determination — The next generation. *Biochim. Biophys. Acta BBA - Biomembr.* 1838, 78–87. <https://doi.org/10.1016/j.bbamem.2013.07.010>
- Mori, K., Mahmood, Md.I., Neya, S., Matsuzaki, K., Hoshino, T., 2012. Formation of GM1 Ganglioside Clusters on the Lipid Membrane Containing Sphingomyeline and Cholesterol. *J. Phys. Chem. B* 116, 5111–5121. <https://doi.org/10.1021/jp207881k>
- Morrison, K.A., Akram, A., Mathews, A., Khan, Z.A., Patel, J.H., Zhou, C., Hardy, D.J., Moore-Kelly, C., Patel, R., Odiba, V., Knowles, T.J., Javed, M.-H., Chmel, N.P., Dafforn, T.R., Rothnie, A.J., 2016. Membrane protein extraction and purification using styrene–maleic acid (SMA) copolymer: effect of variations in polymer structure. *Biochem. J.* 473, 4349–4360. <https://doi.org/10.1042/BCJ20160723>
- Morrow, M.R., Singh, D., Lu, D., Grant, C.W., 1995. Glycosphingolipid fatty acid arrangement in phospholipid bilayers: cholesterol effects. *Biophys. J.* 68, 179–186. [https://doi.org/10.1016/S0006-3495\(95\)80173-6](https://doi.org/10.1016/S0006-3495(95)80173-6)
- Mortazavi, A., Rajagopalan, V., Sparks, K.A., Greathouse, D.V., Koeppe, R.E., 2016. Juxta-terminal Helix Unwinding as a Stabilizing Factor to Modulate the Dynamics of Transmembrane Helices. *ChemBioChem* 17, 462–465. <https://doi.org/10.1002/cbic.201500656>
- Munnik, T., Wierzchowicka, M., 2013. Lipid-Binding Analysis Using a Fat Blot Assay. pp. 253–259. https://doi.org/10.1007/978-1-62703-401-2_23
- Odintsova, E., Butters, T.D., Monti, E., Sprong, H., van Meer, G., Berditchevski, F., 2006. Gangliosides play an important role in the organization of CD82-enriched microdomains. *Biochem. J.* 400, 315–25. <https://doi.org/10.1042/BJ20060259>
- Oliveira, L., Paiva, A.C.M., Vriend, G., 1993. A common motif in G-protein-coupled seven transmembrane helix receptors. *J. Comput. Aided Mol. Des.* 7, 649–658. <https://doi.org/10.1007/BF00125323>
- Oliver, R.C., Lipfert, J., Fox, D.A., Lo, R.H., Kim, J.J., Doniach, S., Columbus, L., 2014. Tuning Micelle Dimensions and Properties with Binary Surfactant Mixtures. *Langmuir* 30, 13353–13361. <https://doi.org/10.1021/la503458n>
- Oluwole, A.O., Danielczak, B., Meister, A., Babalola, J.O., Vargas, C., Keller, S., 2017. Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angew. Chem. Int. Ed.* 56, 1919–1924. <https://doi.org/10.1002/anie.201610778>
- Omasits, U., Ahrens, C.H., Müller, S., Wollscheid, B., 2014. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* 30, 884–886. <https://doi.org/10.1093/bioinformatics/btt607>
- Orekhov, P.S., Bozdaganyan, M.E., Voskoboynikova, N., Mulkidjanian, A.Y., Steinhoff, H.-J., Shaitan, K.V., 2019. Styrene/Maleic Acid Copolymers Form SMALPs by Pulling Lipid Patches out of the Lipid Bilayer. *Langmuir* 35, 3748–3758. <https://doi.org/10.1021/acs.langmuir.8b03978>
- Overduin, M., Esmaili, M., 2019. Structures and Interactions of Transmembrane Targets in Native Nanodiscs. *SLAS Discov.* 24, 943–952. <https://doi.org/10.1177/2472555219857691>
- Overington, J.P., Al-Lazikani, B., Hopkins, A.L., 2006. How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996. <https://doi.org/10.1038/nrd2199>
- Pagadala, N.S., Syed, K., Tuszynski, J., 2017. Software for molecular docking: a review. *Biophys. Rev.* 9, 91–102. <https://doi.org/10.1007/s12551-016-0247-1>
- Palor, M., Stejskal, L., Mandal, P., Lenman, A., Alberione, M.P., Kirui, J., Moeller, R., Ebner, S., Meissner, F., Gerold, G., Shepherd, A.J., Grove, J., 2020. Cholesterol sensing by CD81 is important for hepatitis C virus entry. *J. Biol. Chem.* 295, 16931–16948. <https://doi.org/10.1074/jbc.RA120.014761>
- Pearson, W.R., 2013. Selecting the Right Similarity-Scoring Matrix. *Curr. Protoc. Bioinforma.* 43, 3.5.1-3.5.9. <https://doi.org/10.1002/0471250953.bi0305s43>

- Perozo, E., Cortes, D.M., Sompornpisut, P., Kloda, A., Martinac, B., 2002. Open channel structure of MscL and the gating mechanism of mechanosensitive channels. *Nature* 418, 942–948. <https://doi.org/10.1038/nature00992>
- Pieper, R., Huang, S.-T., Clark, D.J., Robinson, J.M., Alami, H., Parmar, P.P., Suh, M.-J., Kuntumalla, S., Bunai, C.L., Perry, R.D., Fleischmann, R.D., Peterson, S.N., 2009. Integral and peripheral association of proteins and protein complexes with *Yersinia pestis* inner and outer membranes. *Proteome Sci.* 7, 5. <https://doi.org/10.1186/1477-5956-7-5>
- Pike, L.J., 2009. The challenge of lipid rafts. *J. Lipid Res.* 50 Suppl, S323-8. <https://doi.org/10.1194/jlr.R800040-JLR200>
- Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G., Abrignani, S., 1998. Binding of Hepatitis C Virus to CD81. *Science* 282, 938–941. <https://doi.org/10.1126/science.282.5390.938>
- Pollock, N.L., Lee, S.C., Patel, J.H., Gulamhussein, A.A., Rothnie, A.J., 2018. Structure and function of membrane proteins encapsulated in a polymer-bound lipid bilayer. *Biochim. Biophys. Acta BBA - Biomembr.* 1860, 809–817. <https://doi.org/10.1016/j.bbamem.2017.08.012>
- Prieto-Martínez, F.D., Arciniega, M., Medina-Franco, J.L., 2018. Acoplamiento Molecular: Avances Recientes y Retos. *TIP Rev. Espec. En Cienc. Quím.-Biológicas* 21. <https://doi.org/10.22201/fesz.23958723e.2018.0.143>
- Privé, G.G., 2007. Detergents for the stabilization and crystallization of membrane proteins. *Methods* 41, 388–397. <https://doi.org/10.1016/j.ymeth.2007.01.007>
- Prokazova, N.V., Samoilova, N.N., Gracheva, E.V., Golovanova, N.K., 2009. Ganglioside GM3 and its biological functions. *Biochem. Mosc.* 74, 235–249. <https://doi.org/10.1134/S0006297909030018>
- Puklin-Faucher, E., Sheetz, M.P., 2009. The mechanical integrin cycle. *Journal of Cell Science* 122, 179–186. <https://doi.org/10.1242/jcs.042127>
- Purushothaman, G., Thiruvengatam, V., 2019. High Yield Expression of Recombinant CD151 in *E. coli* and a Structural Insight into Cholesterol Binding Domain. *Mol. Biotechnol.* 61, 905–915. <https://doi.org/10.1007/s12033-019-00212-3>
- Qiu, W., Fu, Z., Xu, G.G., Grassucci, R.A., Zhang, Y., Frank, J., Hendrickson, W.A., Guo, Y., 2018. Structure and activity of lipid bilayer within a membrane-protein transporter. *Proceedings of the National Academy of Sciences* 115, 12985–12990. <https://doi.org/10.1073/pnas.1812526115>
- Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99. [https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)
- Remmert, M., Biegert, A., Hauser, A., Söding, J., 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9, 173–175. <https://doi.org/10.1038/nmeth.1818>
- Ristaino, P.A., Levine, M.M., Young, C.R., 1983. Improved GM1-enzyme-linked immunosorbent assay for detection of *Escherichia coli* heat-labile enterotoxin. *J. Clin. Microbiol.* 18, 808–15. <https://doi.org/10.1128/jcm.18.4.808-815.1983>
- Rosenbaum, D.M., Rasmussen, S.G.F., Kobilka, B.K., 2009. The structure and function of G-protein-coupled receptors. *Nature* 459, 356–363. <https://doi.org/10.1038/nature08144>
- Rothnie, A.J., 2016. Detergent-Free Membrane Protein Purification. pp. 261–267. https://doi.org/10.1007/978-1-4939-3637-3_16
- Routledge, S.J., Mikaliunaite, L., Patel, A., Clare, M., Cartwright, S.P., Bawa, Z., Wilks, M.D.B., Low, F., Hardy, D., Rothnie, A.J., Bill, R.M., 2016. The synthesis of recombinant membrane proteins in yeast for structural studies. *Methods* 95, 26–37. <https://doi.org/10.1016/j.ymeth.2015.09.027>
- Runge, K.E., Evans, J.E., He, Z.-Y., Gupta, S., McDonald, K.L., Stahlberg, H., Primakoff, P., Myles, D.G., 2007. Oocyte CD9 is enriched on the microvillar membrane and required for normal microvillar shape and distribution. *Developmental Biology* 304, 317–325. <https://doi.org/10.1016/j.ydbio.2006.12.041>

- Sahu, I.D., Lorigan, G.A., 2018. Site-Directed Spin Labeling EPR for Studying Membrane Proteins. *BioMed Res. Int.* 2018, 1–13. <https://doi.org/10.1155/2018/3248289>
- Schmidt, T.H., Homsy, Y., Lang, T., 2016. Oligomerization of the Tetraspanin CD81 via the Flexibility of Its δ -Loop. *Biophys. J.* 110, 2463–2474. <https://doi.org/10.1016/j.bpj.2016.05.003>
- Schnaar, R.L., 2004. Glycolipid-mediated cell–cell recognition in inflammation and nerve regeneration. *Archives of Biochemistry and Biophysics* 426, 163–172. <https://doi.org/10.1016/j.abb.2004.02.019>
- Schnaar, R.L., Gerardy-Schahn, R., Hildebrandt, H., 2014. Sialic Acids in the Brain: Gangliosides and Polysialic Acid in Nervous System Development, Stability, Disease, and Regeneration. *Physiol. Rev.* 94, 461–518. <https://doi.org/10.1152/physrev.00033.2013>
- Schwartz, T.W., 1994. Locating ligand-binding sites in 7tm receptors by protein engineering. *Curr. Opin. Biotechnol.* 5, 434–444. [https://doi.org/10.1016/0958-1669\(94\)90054-X](https://doi.org/10.1016/0958-1669(94)90054-X)
- Seddon, A.M., Curnow, P., Booth, P.J., 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta BBA - Biomembr.* 1666, 105–117. <https://doi.org/10.1016/j.bbamem.2004.04.011>
- Seelig, A., Seelig, J., 1974. Dynamic structure of fatty acyl chains in a phospholipid bilayer measured by deuterium magnetic resonance. *Biochemistry* 13, 4839–4845. <https://doi.org/10.1021/bi00720a024>
- Seigneuret, M., 2006. Complete Predicted Three-Dimensional Structure of the Facilitator Transmembrane Protein and Hepatitis C Virus Receptor CD81: Conserved and Variable Structural Domains in the Tetraspanin Superfamily. *Biophys. J.* 90, 212–227. <https://doi.org/10.1529/biophysj.105.069666>
- Shaul, P.W., Anderson, R.G.W., 1998. Role of plasmalemmal caveolae in signal transduction. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* 275, L843–L851. <https://doi.org/10.1152/ajplung.1998.275.5.L843>
- Shin, W.-H., Kim, J.-K., Kim, D.-S., Seok, C., 2013. GalaxyDock2: Protein-ligand docking using beta-complex and global optimization. *J. Comput. Chem.* 34, 2647–2656. <https://doi.org/10.1002/jcc.23438>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. <https://doi.org/10.1038/msb.2011.75>
- Silvie, O., Charrin, S., Billard, M., Franetich, J.-F., Clark, K.L., van Gemert, G.-J., Sauerwein, R.W., Dautry, F., Boucheix, C., Mazier, D., Rubinstein, E., 2006. Cholesterol contributes to the organization of tetraspanin-enriched microdomains and to CD81-dependent infection by malaria sporozoites. *J. Cell Sci.* 119, 1992–2002. <https://doi.org/10.1242/jcs.02911>
- Simons, K., Toomre, D., 2000. Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.* 1, 31–39. <https://doi.org/10.1038/35036052>
- Singer, S.J., Nicolson, G.L., 1972. The Fluid Mosaic Model of the Structure of Cell Membranes. *Science* 175, 720–731. <https://doi.org/10.1126/science.175.4023.720>
- Sipione, S., Monyror, J., Galleguillos, D., Steinberg, N., Kadam, V., 2020. Gangliosides in the Brain: Physiology, Pathophysiology and Therapeutic Applications. *Front. Neurosci.* 14. <https://doi.org/10.3389/fnins.2020.572965>
- Spangler, B.D., 1992. Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin. *Microbiological reviews* 56, 622–47. <https://doi.org/10.1128/mr.56.4.622-647.1992>
- Stothard, P., 2000. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques* 28, 1102–1104. <https://doi.org/10.2144/00286ir01>
- Studer, G., Biasini, M., Schwede, T., 2014. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* 30, i505–i511. <https://doi.org/10.1093/bioinformatics/btu457>

- Sun, C., Benlekbir, S., Venkatakrishnan, P., Wang, Y., Hong, S., Hosler, J., Tajkhorshid, E., Rubinstein, J.L., Gennis, R.B., 2018. Structure of the alternative complex III in a supercomplex with cytochrome oxidase. *Nature* 557, 123–126. <https://doi.org/10.1038/s41586-018-0061-y>
- Susa, K.J., Rawson, S., Kruse, A.C., Blacklow, S.C., 2021. Cryo-EM structure of the B cell co-receptor CD19 bound to the tetraspanin CD81. *Science* 371, 300–305. <https://doi.org/10.1126/science.abd9836>
- Talmont, F., Lebrun, C., Zajac, J.-M., 2020. Agonist binding of human mu opioid receptors expressed in the yeast *Pichia pastoris*: Effect of cholesterol complementation. *Neurochem. Int.* 132, 104588. <https://doi.org/10.1016/j.neuint.2019.104588>
- Tettamanti, G., Bonali, F., Marchesini, S., Zambotti, V., 1973. A new procedure for the extraction, purification and fractionation of brain gangliosides. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* 296, 160–170. [https://doi.org/10.1016/0005-2760\(73\)90055-6](https://doi.org/10.1016/0005-2760(73)90055-6)
- Timpe, J.M., Stamataki, Z., Jennings, A., Hu, K., Farquhar, M.J., Harris, H.J., Schwarz, A., Desombere, I., Roels, G.L., Balfe, P., McKeating, J.A., 2007. Hepatitis C virus cell-cell transmission in hepatoma cells in the presence of neutralizing antibodies. *Hepatology* 47, 17–24. <https://doi.org/10.1002/hep.21959>
- Todeschini, A.R., Dos Santos, J.N., Handa, K., Hakomori, S., 2008. Ganglioside GM2/GM3 complex affixed on silica nanospheres strongly inhibits cell motility through CD82/cMet-mediated pathway. *Proc. Natl. Acad. Sci.* 105, 1925–1930. <https://doi.org/10.1073/pnas.0709619104>
- Todeschini, A.R., Dos Santos, J.N., Handa, K., Hakomori, S., 2007. Ganglioside GM2-Tetraspanin CD82 Complex Inhibits Met and Its Cross-talk with Integrins, Providing a Basis for Control of Cell Motility through Glycosynapse. *J. Biol. Chem.* 282, 8123–8133. <https://doi.org/10.1074/jbc.M611407200>
- Toledo, M.S., Suzuki, E., Handa, K., Hakomori, S., 2004. Cell Growth Regulation through GM3-enriched Microdomain (Glycosynapse) in Human Lung Embryonal Fibroblast WI38 and Its Oncogenic Transformant VA13. *J. Biol. Chem.* 279, 34655–34664. <https://doi.org/10.1074/jbc.M403857200>
- Totrov, M., Abagyan, R., 1994. Detailed ab initio prediction of lysozyme–antibody complex with 1.6 Å accuracy. *Nat. Struct. Mol. Biol.* 1, 259–263. <https://doi.org/10.1038/nsb0494-259>
- Tu, L., Kong, X.-P., Sun, T.-T., Kreibich, G., 2006. Integrity of all four transmembrane domains of the tetraspanin uroplakin Ib is required for its exit from the ER. *J. Cell Sci.* 119, 5077–5086. <https://doi.org/10.1242/jcs.03285>
- Umeda, R., Satouh, Y., Takemoto, M., Nakada-Nakura, Y., Liu, K., Yokoyama, T., Shirouzu, M., Iwata, S., Nomura, N., Sato, K., Ikawa, M., Nishizawa, T., Nureki, O., 2020. Structural insights into tetraspanin CD9 function. *Nat. Commun.* 11, 1606. <https://doi.org/10.1038/s41467-020-15459-7>
- van Meer, G., de Kroon, A.I.P.M., 2011. Lipid map of the mammalian cell. *J. Cell Sci.* 124, 5–8. <https://doi.org/10.1242/jcs.071233>
- van Niel, G., Charrin, S., Simoes, S., Romao, M., Rochin, L., Saftig, P., Marks, M.S., Rubinstein, E., Raposo, G., 2011. The Tetraspanin CD63 Regulates ESCRT-Independent and -Dependent Endosomal Sorting during Melanogenesis. *Developmental Cell* 21, 708–721. <https://doi.org/10.1016/j.devcel.2011.08.019>
- van Sriel, A.B., de Keijzer, S., van der Schaaf, A., Gartlan, K.H., Sofi, M., Light, A., Linszen, P.C., Boezeman, J.B., Zuidschewoude, M., Reinieren-Beeren, I., Cambi, A., Mackay, F., Tarlinton, D.M., Figdor, C.G., Wright, M.D., 2012. The Tetraspanin CD37 Orchestrates the $\alpha_4\beta_1$ Integrin–Akt Signaling Axis and Supports Long-Lived Plasma Cell Survival. *Science Signaling* 5. <https://doi.org/10.1126/scisignal.2003113>
- Varki, A., 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 3, 97–130. <https://doi.org/10.1093/glycob/3.2.97>
- Vereb, G., Szöllösi, J., Matkó, J., Nagy, P., Farkas, T., Vigh, L., Mátyus, L., Waldmann, T.A., Damjanovich, S., 2003. Dynamic, yet structured: The cell membrane three decades

- after the Singer–Nicolson model. *Proc. Natl. Acad. Sci.* 100, 8053–8058. <https://doi.org/10.1073/pnas.1332550100>
- Verma, R., Boleti, E., George, A.J.T., 1998. Antibody engineering: Comparison of bacterial, yeast, insect and mammalian expression systems. *J. Immunol. Methods* 216, 165–181. [https://doi.org/10.1016/S0022-1759\(98\)00077-5](https://doi.org/10.1016/S0022-1759(98)00077-5)
- Vostrikov, V.V., Daily, A.E., Greathouse, D.V., Koeppe, R.E., 2010. Charged or Aromatic Anchor Residue Dependence of Transmembrane Peptide Tilt. *J. Biol. Chem.* 285, 31723–31730. <https://doi.org/10.1074/jbc.M110.152470>
- Wallace, A.C., Laskowski, R.A., Thornton, J.M., 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* 8, 127–134. <https://doi.org/10.1093/protein/8.2.127>
- Wallin, E., Heijne, G.V., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* 7, 1029–1038. <https://doi.org/10.1002/pro.5560070420>
- Wang, X., Yan, Q., Sun, P., Liu, J.-W., Go, L., McDaniel, S.M., Paller, A.S., 2007. Suppression of Epidermal Growth Factor Receptor Signaling by Protein Kinase C- α Activation Requires CD82, Caveolin-1, and Ganglioside. *Cancer Res.* 67, 9986–9995. <https://doi.org/10.1158/0008-5472.CAN-07-1300>
- Wang, Y., Tong, X., Omoregie, E.S., Liu, W., Meng, S., Ye, X., 2012. Tetraspanin 6 (TSPAN6) Negatively Regulates Retinoic Acid-inducible Gene I-like Receptor-mediated Immune Signaling in a Ubiquitination-dependent Manner. *J. Biol. Chem.* 287, 34626–34634. <https://doi.org/10.1074/jbc.M112.390401>
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wiedemann, C., Kumar, A., Lang, A., Ohlenschläger, O., 2020. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Front. Chem.* 8. <https://doi.org/10.3389/fchem.2020.00280>
- Xia, Y., Fischer, A.W., Teixeira, P., Weiner, B., Meiler, J., 2018. Integrated Structural Biology for α -Helical Membrane Protein Structure Determination. *Structure* 26, 657-666.e2. <https://doi.org/10.1016/j.str.2018.02.006>
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., Baker, D., 2020. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci.* 117, 1496–1503. <https://doi.org/10.1073/pnas.1914677117>
- Yang, Y., Gao, J., Wang, J., Heffernan, R., Hanson, J., Paliwal, K., Zhou, Y., 2016. Sixty-five years of the long march in protein secondary structure prediction: the final stretch? *Brief. Bioinform.* bbw129. <https://doi.org/10.1093/bib/bbw129>
- Yang, Y., Liu, X.R., Greenberg, Z.J., Zhou, F., He, P., Fan, L., Liu, S., Shen, G., Egawa, T., Gross, M.L., Schuettelpelz, L.G., Li, W., 2020. Open conformation of tetraspanins shapes interaction partner networks on cell membranes. *EMBO J.* 39. <https://doi.org/10.15252/emj.2020105246>
- Yazdani, K., Dolguikh, K., Zhang, W., Shayegi-Nik, S., Ly, J., Cooper, S., Trigg, J., Bartlett, S., Barrios, R., Montaner, J.S.G., Salters, K., 2022. Knowledge of hepatitis C and awareness of reinfection risk among people who successfully completed direct acting antiviral therapy. *PLOS ONE* 17, e0265811. <https://doi.org/10.1371/journal.pone.0265811>
- Yu, R.K., Tsai, Y.-T., Ariga, T., Yanagisawa, M., 2011. Structures, Biosynthesis, and Functions of Gangliosides—an Overview. *J. Oleo Sci.* 60, 537–544. <https://doi.org/10.5650/jos.60.537>
- Zhang, Y., Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302–9. <https://doi.org/10.1093/nar/gki524>
- Zhu, Y.-Z., Luo, Y., Cao, M.-M., Liu, Y., Liu, X.-Q., Wang, W., Wu, D.-G., Guan, M., Xu, Q.-Q., Ren, H., Zhao, P., Qi, Z.-T., 2012. Significance of palmitoylation of CD81 on its

- association with tetraspanin-enriched microdomains and mediating hepatitis C virus cell entry. *Virology* 429, 112–123. <https://doi.org/10.1016/j.virol.2012.03.002>
- Zimmerman, B., Kelly, B., McMillan, B.J., Seegar, T.C.M., Dror, R.O., Kruse, A.C., Blacklow, S.C., 2016. Crystal Structure of a Full-Length Human Tetraspanin Reveals a Cholesterol-Binding Pocket. *Cell* 167, 1041-1051.e11. <https://doi.org/10.1016/j.cell.2016.09.056>
- Zinser, E., Paltauf, F., Daum, G., 1993. Sterol composition of yeast organelle membranes and subcellular distribution of enzymes involved in sterol metabolism. *J. Bacteriol.* 175, 2853–8. <https://doi.org/10.1128/jb.175.10.2853-2858.1993>
- Zuidscherwoude, M., Göttfert, F., Dunlock, V.M.E., Figdor, C.G., van den Bogaart, G., Sriel, A.B. van, 2015. The tetraspanin web revisited by super-resolution microscopy. *Sci. Rep.* 5, 12201. <https://doi.org/10.1038/srep12201>