Towards a Structural Understanding of the iRhom/ADAM17 Complex

Archna Rajesh Shah

The Queen's College

University of Oxford



A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2020

To my parents

Towards a Structural Understanding of the iRhom/ADAM17 Complex

Archna Rajesh Shah

Submitted for the degree of Doctor of Philosophy, Trinity Term 2020

Sir William Dunn School of Pathology, Medical Sciences Doctoral Training Centre, The Queen's College

Abstract

The rhomboid-like superfamily consists of the active rhomboid proteases and the inactive rhomboid pseudoproteases. The latter group includes the iRhoms, a family of polytopic transmembrane proteins composed of a long N-terminal cytoplasmic domain followed by seven transmembrane helices (TMH), with a highly conserved loop between TMH1 and TMH2 called the iRhom Homology Domain (IRHD). There are two mammalian iRhoms; iRhom1, which is broadly expressed, and iRhom2, which is predominantly expressed in immune cells and skin. A key function of iRhoms is regulating A disintegrin and metalloproteinase 17 (ADAM17), a metalloprotease responsible for the cleavage of tethered ligands on the plasma membrane, including the primary inflammatory cytokine tumour necrosis factor α. iRhoms are responsible for the trafficking of the catalytically inactive and immature form of ADAM17 from the endoplasmic reticulum (ER) to the Golgi apparatus where the inhibitory pro-domain of ADAM17 is cleaved and removed. The iRhom/ADAM17 complex traffics to the cell surface where ADAM17 is now mature. Mature ADAM17 is activated by the phosphorylation of iRhom and therefore able to cleave its substrates. iRhoms have an indirect but important role in ADAM17-dependent cleavage of substrates and are increasingly viewed as regulatory subunits of ADAM17. The complex is stabilised at the cell surface as a result of the prevention of its degradation by FRMD8.

Before this thesis, little to no structural studies have been performed on the iRhom/ADAM17 complex. Understanding the complex at a structural level will enable further insight to the mechanisms by which these two proteins interact. Therefore, the aim of this project was to express and purify iRhoms alone and in complex with ADAM17 for structure determination.

The project initially started by characterising the IRHD, of human iRhom1 (IRHD1), with the aim of crystallisation because it is the only major ER luminal domain of iRhom that may interact with the extracellular domain of ADAM17. However, after insect cell expression and bacterial refolding of IRHD1 it became apparent that the IRHD may be more stable as part of the full length protein. Human cell lines were generated to stably overexpress mammalian and non-mammalian iRhoms with their corresponding ADAM17 and a chimeric form of human iRhom2 fused to FRMD8 (*Hs*FRMD8iRhom2). Expression and sub-cellular localisation of the generated cell lines was verified and optimal solubilising conditions were determined for maximal protein extraction out of the membrane and for minimal complex dissociation. *Hs*FRMD8iRhom2 chimera was successfully purified and preliminary cryo electron microscopy data was generated. Therefore, the results reported in this thesis provide a foundation towards solving the structure of iRhom and iRhom/ADAM17 complex.

Acknowledgements

There are many people I would like to thank for enabling me to reach this point of my academic journey. First and foremost, I would like to thank my supervisors Professor Susan Lea and Professor Matthew Freeman for the great privilege to work in their laboratories on this project and for their support both academically and personally. It was a wonderful opportunity to learn both structural biology and cell biology for which I am grateful. Certainly this thesis would not be possible without them.

Being a member of two groups, current and past members of the Lea and Freeman groups made me feel like I was a whole member of each rather than just the 'official' 50 %. Therefore my thanks go to past and present members of both groups, over thirty individuals, for making my DPhil an enjoyable and enriching time. The Lea group: Alice Ahn, Nandita Bodra, Katharina Braunger, Joseph Caesar, Justin Deme, Arun Dhillon, Ivan Fong, Emily Furlong, Andreas Hänsele, Rory Hennell James, Steve Johnson, Lucas Kuhlen, Camilla Lundgren, Nada Mohamad, Martin Reichhardt, Jacob Robson-Tull and Molly Tang. The Freeman group: Shaked Ashkenazi, Iqbal Dulloo, Adam Grieve, Nina Jajčanin Jozić, Lucie Kafkova, Ulrike Künzel, Clémence Levet, Fangfang Lu, Ángela Moncada Pazos, Sonia Muliyil, Jo Peel, Boris Sieber, Lucrezia Zarcone and Boyan Zhang.

There are a few individuals in the laboratory I must give special thanks to. I want to thank Justin for his supervision and mentorship on the structural biology side. His experience and knowledge in membrane protein biochemistry and cryo electron microscopy (EM) are second to none. Once I had finally reached that stage of my project, his insight and experience was invaluable. Thanks to Steve for advice during this project and proofreading various reports and to Joseph C. for his supervision on bacterial refolding project. On the cell biology side, I thank Nina, Clémence and Adam for their cell biological insights and making me think about problems in a different way. Many thanks to Nina for her supervision whilst I was implementing a new expression system in the Lea group, for scientific and non-scientific discussions and for proofreading various reports. I thank my cell culture buddy, Clémence, for her help and support in tissue culture on the numerous days when I was not able to tend to cells. The help from Clémence certainly accelerated progress on this project and I even learnt some French along the way. I would also like to thank her and Sonia for all the cell imaging. I thank Adam G. for scientific discussions, feedback on seminars and academic reports.

I thank Michal Maj, Line Eriksen and Robert Hedley from the Dunn School Flow Cytometry Facility for enriching my numerous cell lines. Also, thanks to Errin Johnson, Adam Costin and Raman Dhaliwal from The Central Oxford Structural Molecular Imaging Centre and the Dunn School Bioimaging Facility for the various training and support of everything EM related. I felt your optimism during the progression of this project! I would also like to acknowledge members of the School of Life Sciences and Warwick Medical School at the University of Warwick particularly Dr. Robert Spooner, Prof. David Roper and Dr. Graham Ladds for their supervision and training during multiple projects during my time there.

It would be remiss of me not to acknowledge these unsettled times. This thesis was written during the COVID-19 lockdown from my Oxford home. I thank Emily and Katharina for their help, support and encouragement. During writing, I heavily relied on Emily who ran around the Dunn School (safely) to give me any technical information that I did not make note of!

Finally, to end my academic acknowledgments I am very grateful for the generous support from the Wellcome Trust for funding this project.

Of course, there are a number of individuals outside of the laboratory who helped me through the tougher times of my DPhil. I thank my housemates Hamza Alawiye, Tristan Giron and Avi Mayorcas for introducing me to their world of mathematics and for their care, support and encouragement through the hard times. Whilst I regret being unable to take advantage of Oxford's numerous libraries whilst writing this thesis, there is not a group of individuals that I would have rather spent lockdown working from home with. I also thank Alexandra Toivenen, who became a member of our bubble, for adding more fun in last six months of this project. Spending these last four years with them is something I will treasure.

My heartfelt thanks go to my partner, Joseph Hogg, who also became a member of the working from home bubble to further add to the mix of characters. I thank him for putting up with my long days, helping me gain sight of what was truly important and being there for me at times that felt difficult. He was, and still is, an infinite source of love, encouragement and motivation. Further, I also thank him academically, for helping and tutoring me on all of my coding problems; I have a new found appreciation for programming.

My thanks go to my dearest friends outside of Oxford, Lauren Aris, Louise Balloch, Olivia Batchelor, Laura Darley, Megan Devonald, Elle Stott, Lauren Webb and Holly Williams for their continued support and curiosity about science throughout my academic career.

There are three individuals that I must thank because without them I would not be where I am today. My sincerest thanks go to Angela Godfrey, Simone Newman and Helen Brown. I will be forever grateful to them for helping me through a life defining period.

Finally I thank my family, particularly my parents, Rajesh and Sheelpa Shah, who have supported and encouraged me throughout my life. Their continued love and support enabled me to attend university and to end up studying at one of the top institutes in the world. Without them, I would not be the person and the scientist I have become.

Table of Contents

Abstract I

Acknowledgements III

Table of Contents V

List of Figures X

List of Tables XII

List of Abbreviations XIII

Chapter 1: Introduction 1

1.1 iRhoms are members of the rhomboid-like superfamily 1

1.1.1 The rhomboid-like superfamily 1

1.1.2 iRhoms 4

1.1.2.1 iRhom expression 4

1.1.2.2 iRhom topological structure 5

1.2 ADAM17 (A disintegrin and metalloproteinase 17) 7

1.2.1 Intercellular signalling 7

1.2.2 ADAM17 function and topological structure 8

1.2.3 ADAM17 regulation 11

1.3 iRhoms and their functions 12

1.3.1 iRhoms are regulatory subunits of ADAM17 12

1.3.2 Other iRhom binding partners 16

1.3.3 Other functions of iRhoms 16

1.3.4 Pathological functions of iRhoms 17

1.4 Rhomboid structure and their client/substrate recognition 21

1.4.1 Known rhomboid structures 21

1.4.2 Rhomboid substrate/client protein recognition 26

1.5 Structural determination by cryogenic electron microscopy 27

1.5.1 Sample preparation 29

1.5.2 Data collection and processing 30

1.6 Aims of thesis 32

Chapter 2: Material and Methods 34

2.1 Materials 34

2.2 Molecular Biology 34

2.2.1 Genes and plasmids 34
2.2.2 DNA oligonucleotides 35
2.2.3 DNA amplification 35
2.2.4 Amplicon visualisation and purification 36
2.2.5 Cloning 36
2.2.6 Bacterial cell transformation by heat shock 37
2.2.7 Colony PCR 38
2.2.8 Plasmid amplification and purification 38
2.2.9 DNA sequencing 39
2.3 Cell culture 39
2.3.1 Bacterial cell culture 39
2.3.2 Eukaryotic cell culture 40
2.3.2.1 Schneider 2 cell culture and transfection 40
2.3.2.2 Adherent HEK cell culture 40
2.3.2.3 Suspension HEK cell culture 42
2.3.2.4 DNA transfection and lentiviral transduction 43
2.4 Biochemistry 43
2.4.1 Standard procedure for purifying recombinant soluble proteins 43
2.4.1.1 Purifying from Schneider 2 cell media 43
2.4.1.2 Purifying from Schneider 2 cells 43
2.4.1.3 Purifying from bacterial cells 44
2.4.2 Standard procedure for refolding and purifying from bacterial inclusion bodies 45
2.4.3 Standard procedure for purifying recombinant membrane proteins 46
2.4.4 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis 47
2.4.5 Western blotting 48
2.5 Biophysical techniques 50
2.5.1 Mass spectrometry 50
2.5.2 Circular dichroism 50
2.6 Structural techniques 50
2.6.2 Negative stain electron microscopy 50
2.6.3 Cryogenic electron microscopy 51
2.6.3.1 Data collection 51
2.6.3.2 Single particle analysis reconstruction 51
2.7 Cell biology 52

2.7.1 Cell lysis for expression testing 52
2.7.2 Immunofluorescence 52
2.7.3 Cell imaging 53
2.7.3.1 Live cell imaging 53
2.7.3.2 Confocal imaging 53
2.7.3.3 Image analysis 54
2.8 Flow cytometry 54
2.8.1 Flow cytometry analysers 54
2.8.2 Fluorescence activated cell sorting 54
2.8.3 Data analysis 55
Chapter 3: Expression and purification of the iRhom Homology Domain 56
3.1 Introduction 56
3.1.1 iRhom Homology Domain as a structural target 56
3.2 Bioinformatics 57
3.3 Schneider 2 cell expression of *Hs*IRHD1 60
3.3.1 Stable cell line generation expressing *Hs*IRHD1 60
3.3.2 Purification of *Hs*IRHD1 from Schneider 2 cell media 61
3.3.3 Characterisation of *Hs*IRHD1 64
3.3.4 *Hs*IRHD1 is poorly secreted into media 66
3.4 Bacterial expression and purification of *Hs*IRHD1 70
3.4.1 Small-scale soluble protein expression of *Hs*IRHD1 in bacterial cells 70
3.4.2 Bacterial protein expression into inclusion bodies and preparation of inclusion bodies 72
3.5 Discussion 77
3.5.1 Further strategies for purifying IRHD 77
3.5.2 IRHD may be more stable as part of the full length protein 79
Chapter 4: Cell line generation for expression of full length iRhom and ADAM17 80
4.1 Introduction 80
4.1.1 Stable iRhom target for structural studies 80
4.2 Expression and purification on previously generated reagents 81
4.2.1 Testing expression with previously generated reagents 81
4.2.2 Small scale expansion and purification of overexpressed *Hs*iRhom2 82
4.2.3 Large scale expansion and purification of overexpressed *Hs*iRhom2 84
4.3 Redesigning eukaryotic expression of iRhoms and ADAM17 for protein purification 86

4.3.1 Construct design 86
4.3.2 Protocol for generating stable cell lines by lentiviral transduction 88
4.3.2.1 Plasmid preparation 88
4.3.2.2 Transfection 89
4.3.2.3 Transduction 90
4.3.2.4 Determining lentiviral titre 90
4.3.2.5 Recovery 91
4.3.3 Expression of mammalian iRhoms in FreeStyle™ 293F cell lines 91
4.3.4 Expression of mammalian iRhoms with the addition of fluorescent proteins 94
4.3.4.1 Flow cytometry of generated cell lines 95
4.3.4.2 Population distribution of generated cell lines 95
4.3.4.3 Cell sorting enrichment 97
4.3.5 Small scale purification of iRhom constructs 99
4.3.6 Coexpression of iRhom and ADAM17 in generated cell lines 101
4.3.7 Sub cellular localisation of generated HEK293F cell lines 103
4.3.7.1 Live cell imaging of iRhom alone cell lines 103
4.3.7.2 Immunofluorescence staining of mammalian iRhom and ADAM17 cell lines 103
4.4 Discussion 104
4.4.1 Initial experiments using the *Hs*iRhom2-3xHA cell lines 104
4.4.2 Fluorescent tags aid in cell enrichment and poor transduction efficiency 105
Chapter 5: Detergent screening and purification of cell lines 108
5.1 Introduction 108
5.2 Membrane isolation 110
5.3 *Hs*FRMD8iRhom2Δ300-mVenus-TwS 112
5.3.1 Detergent screening 112
5.3.2 Purification 115
5.3.2.1 Initial large scale purification 115
5.3.2.2 Further purifications 118
5.3.3 HCV 3C protease cleavage testing 123
5.4 iRhom and ADAM complex 124
5.4.1 Detergent screening 124
5.4.1.1 *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His 125
5.4.1.2 Detergent screen of *Hs*iRhom1 or *Hs*iRhom2 in complex with *Hs*ADAM17 128
5.4.2 iRhom/ADAM17 complex purification 134

5.4.2.1 Two-step purification of *Mm*iRhom-mVn-TwS/*Mm*ADAM17-mSc-His 134
5.4.2.2 Three-step purification of *Mm*iRhom-mVn-TwS/*Mm*ADAM17-mSc-His 137
5.5 Discussion 140
5.5.1 *Hs*FRMDiRhom2 chimera 140
5.5.2 iRhom and ADAM17 complex 140
5.5.3 Future work 141
Chapter 6: Preliminary structural determination of *Hs*FRMD8iRhom2Δ300 chimera and thesis summary 143
6.1 Introduction 143
6.2 Cryo-EM of *Hs*FRMD8iRhom2Δ300 152
6.2.1 2D class averages 154
6.2.2 3D class averages 157
6.3 Discussion 159
6.4 Thesis summary 159
6.4.1 Conclusions 159
6.4.2 Future directions 160
Appendices 162
Appendix A: Primers used in this thesis 162
Appendix B: Automated immunofluorescence image processing script 168
Appendix C: Multiple sequence alignment of iRhoms 169
Appendix D: Derivation of physical lentiviral titre 178
References 179

List of Figures

Figure 1.1. iRhoms are members of the rhomboid-like superfamily. 3
Figure 1.2. Schematic of iRhom topology. 6
Figure 1.3. Schematic of ADAM17 structure. 10
Figure 1.4. iRhoms are regulatory subunits of ADAM17. 13
Figure 1.5. FRMD8 stabilises the iRhom/ADAM17 complex by preventing the degradation of the complex. 15
Figure 1.6. Schematic of iRhom2 with their phosphorylation sites and interacting partners. 20
Figure 1.7. The crystal structure of *Escherichia coli* active rhomboid protease, GlpG. 21
Figure 1.8. Overlay of the cartoon representation of Derlin-1 solved by cryo-EM and the crystal structure of GlpG. 25
Figure 1.9. Structural targets for this thesis. 33
Figure 3.1. The iRhom homology domain (IRHD) is highly conserved with little secondary structure. 60
Figure 3.2. *Hs*IRHD1 is expressed in the media of S2 cells. 61
Figure 3.3. Purification of *Hs*IRHD1. 64
Figure 3.4. Glycoprotein staining and circular dichroism of *Hs*IRHD1 purified from S2 cell media. 66
Figure 3.5. *Hs*IRHD1 remains in S2 cells. 70
Figure 3.6. Bacterial expression trials of *Hs*IRHD1. 71
Figure 3.7. *Hs*IRHD1 refolded from bacterial inclusion bodies. 75
Figure 3.8. Anion exchange of purified refolded *Hs*IRHD1. 76
Figure 3.9. The refolded species of *Hs*IRHD1 was not resolvable by high-performance liquid chromatography. 77
Figure 4.1. Expression of iRhom2 from HEK293T. 82
Figure 4.2. Immunoprecipitation of overexpressed *Hs*iRhom2-3xHA. 84
Figure 4.3. Western blot of large scale HA purification. 86
Figure 4.4. iRhom and ADAM17 construct designs for mammalian expression. 88
Figure 4.5. Stable lentiviral cell line generation. 93
Figure 4.6. Interpolations of lentiviral physical titre. 95
Figure 4.7. Population distributions of generated cell lines. 99
Figure 4.8. Small scale purification of polyclonal cell lines transduced with iRhom genes. 100
Figure 4.9. Expression testing of iRhom and ADAM17 in double transduced cell lines. 102

Figure 4.10. Immunofluorescence staining of double transduced HEK293F cells with iRhom and ADAM17. 104
Figure 5.1. Chemical structures of commonly used membrane protein detergents. 109
Figure 5.2. Stages of membrane isolation. 111
Figure 5.3. *Hs*FRMD8iRhom2-mVenus-TwS detergent screen. 114
Figure 5.4 *Hs*FRMD8iRhom2-mVenus-TwS purification. 117
Figure 5.5 Larger scale *Hs*FRMD8iRhom2-mVenus-TwS purification. 119
Figure 5.6. Superdex 200 purification of *Hs*iFRMD8iRhom2-mVenus-TwS. 121
Figure 5.7. *Hs*iFRMD8iRhom2-mVenus-TwS stability. 122
Figure 5.8. Repeated size-exclusion of Superdex 200 peak fractions. 123
Figure 5.9. HCV-3C protease cleavage of *Hs*FRMD8iRhomΔ300-mVenus-TwS. 124
Figure 5.10. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His detergent screen. 127
Figure 5.11. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His solubilised in 1 % LMNG and 0.1 % CHS. 128
Figure 5.12. *Hs*iRhom1-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His detergent screen. 131
Figure 5.13. *Hs*iRhom2-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His detergent screen. 132
Figure 5.14. Solubilising conditions of *Hs*iRhom1-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His and *Hs*iRhom2-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His visualised by western blot. 134
Figure 5.15. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His two-step purification. 136
Figure 5.16. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His three-step purification. 139
Figure 6.1. *Hs*FRMD8iRhom2Δ300-mVn-TwS grid screening. 154
Figure 6.3. 2D class averages of *Hs*FRMD8iRhom2Δ300-mVn-TwS. 156
Figure 6.2. Cleanup2D class averages of *Hs*FRMD8iRhom2Δ300-mVn-TwS. 156
Figure 6.4. Overview of data process to produce 3-dimensional classes of *Hs*FRMD8iRhom2Δ300-mVn-TwS 158

List of Tables

Table 2.1. Genes used in this thesis. 34
Table 2.2. Plasmids used in this thesis. 35
Table 2.3. Solutions for DNA visualisation. 36
Table 2.4. Protocol for Gibson Assembly. 37
Table 2.5. Solutions and buffers for SDS-PAGE analysis. 47
Table 2.6. Antibodies used for western blots in this thesis and protocol used. 49
Table 2.7. Antibodies used for immunofluorescence staining. 53
Table 2.8. Parameters used for fluorescence-activated cell sorting. 55
Table 3.1. List of species used in alignments 58
Table 4.1. Parameters used for flow cytometry analysers. 96
Table 4.2. Predicted molecular weights of iRhoms and ADAM17 101
Table 5.1. Solubilising conditions for the *Hs*FRMD8iRhom2Δ300-mVn-TwS detergent screen 113
Table 5.2. Excitation and emission wavelengths of mVenus and mScarlet-I. 114
Table 5.3. Solubilising conditions for the mammalian iRhom/ADAM complex detergent screen. 125
Table 6.1. List of protein structures solved by cryogenic electron microscopy single particle analysis. 144
Table 6.2. Parameters of the Vitrobot Mark IV of prepared grids of *Hs*FRMD8iRhom2Δ300-mVn-TwS. 153
Table 6.3. Cryo-EM data collection parameters. 153
Table A1. Primers used in this thesis. 162

List of Abbreviations

2D	2-dimensional
3D	3-dimensional
ADAM17	A disintegrin and metalloproteinase 17
Adh	Adherent
Am	*Alligator mississippiensis*
APS	Ammonium persulfate
BCA	Bicinchoninic acid
BiP	Binding protein signal peptide
BSA	Bovine serum albumin
CANDIS	conserved adam seventeen dynamic interaction sequence
CCD	Charged coupled device
CD	Circular dichroism
CHS	Cholesteryl Hemisuccinate
cmc	Critical micelle concentration
CMV	Cytomegalovirus
ConA	Concanavalin A
Cryo	Cryogenic
Cryo-EM	Cyogenic electron microscopy
cub	*curly-bare* mutant
CV	Column volume
DAPI	4',6-diamidino-2-phenylindole
DARPins	Designed Ankyrin Repeat Proteins
DDM	n-Dodecyl-β-D-Maltopyranoside
DEDs	Direct electron detectors
Der-1	Derlin 1
DKO	Double knock out
DM	n-Decyl-β-D-Maltopyranoside
DMEM	Dulbecco's Modified Eagle Medium
DMEM/F-12	Dulbecco's Modified Eagle Medium:Nutrient Mixture F-12
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase I	Bovine Pancreas Deoxyribonuclease I
DQE	Detective quantum efficiency
DTT	Dithiothreitol
ECD	Extracellular domain
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EMDB	The Electron Microscopy Data Bank
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated protein degradation

Fabs	Antigen binding-fragments
FBS	Fetal bovine serum
FRMD8	FERM domain-containing protein 8
FSEC	Fluorescent detection size-exclusion chromatography
FSG	Fish skin gelatine
GDN	Glyco-diosgenin
Gg	*Gallus gallus*
GPCR	G protein coupled receptor
HA	Human influenza hemagglutinin
Hd	*Hypsibius dujardini*
HDAC	Histone deacetylase
HEK	Human embryonic kidney
HEK293F	FreeStyle™ 293-F
His	Histidine tag
HPLC	High-performance Liquid Chromatography
Hs	*Homo sapiens*
HSV-1	Herpes simplex virus 1
IF	Immunofluorescence
IMAC	Immobilised metal affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRHD	iRhom homology domain
IRHD1	iRhom homology domain of iRhom1
IRHD2	iRhom homology domain of iRhom2
JMD	Juxtamembrane domain
KO	Knock out
LB	Lysogeny broth
LDS	Lithium dodecyl sulphate
LMNG	Lauryl Maltose Neopentyl Glycol
LPS	Lipopolysaccharide
MAVS	Mitochondrial antiviral signalling protein
MEF	Mouse embryonic fibroblast
MEM NEAA	Minimum Essential Medium Non-Essential Amino Acids
Mm	*Mus musculus*
MPD	Membrane proximal domain
MSA	Multiple sequence alignment
mSc	mScarlet-I
mVn	mVenus
MWCO	molecular weightc cut off
MX	Macromolecular X-ray crystallography
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine

PFA	Paraformaldehyde
PM	Plasma membrane
PVDF	Polyvinylidene difluoride
RMSD	Root-mean-square deviation
Rn	Rat, *Rattus norvegicus*
S2	Schneider 2
S200	Superdex 200
S75	Superdex 75
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SN	Supernatant
SNR	Signal to noise ratio
SPA	Single particle analysis
SSM	Secondary Structure Matching
STING	Stimulator of interferon genes
Sus	Suspension
TACE	Tumour necrosis factor-α converting enzyme
TEM	Transmission electron microscopy
TFA	Trifluoroacetic acid
TGFα	Transforming growth factor alpha
TMH	Transmembrane helix
TNFR	TNFα receptor
TNFα	Tumour necrosis factor alpha
TOC	Tylosis with oesophageal cancer
TU	Transducing units
TwS	Twin Strep
uncv	*uncovered mutant*
VSV	Vesicular stomatitis virus
WT	Wild type

Chapter 1: Introduction

1.1 iRhoms are members of the rhomboid-like superfamily

1.1.1 The rhomboid-like superfamily

The rhomboid-like superfamily is a conserved family of polytopic membrane proteins present in all kingdoms of life (Freeman, 2014). The superfamily is named after the founding protein member, Rhomboid-1 in *Drosophila melanogaster* which, when mutated, causes a pointed head skeleton phenotype in *Drosophila* embryos (Jürgens et al., 1984). Rhomboid-1 was the first intramembrane serine protease to be discovered. It was shown to cleave its substrate Spitz, an epidermal growth factor (EGF) receptor (EGFR) ligand, within the transmembrane helix (TMH) (Urban et al., 2001). Spitz is synthesised as an inactive form comprised of an extracellular EGF domain bound to a single transmembrane helix (Rutledge et al., 1992). The cleavage of Spitz enables the release of soluble Spitz followed by the subsequent activation of EGFRs (Lee et al., 2001, Urban et al., 2001).

Rhomboid homologues are present in all eukaryotes and most prokaryotes with varying functions including (reviewed in (Düsterhöft et al., 2017, Ticha et al., 2018)): quorum sensing in bacteria (Stevenson et al., 2007), host cell invasion by apicomplexan parasites (Baker et al., 2006, O'Donnell et al., 2006, Buguliskis et al., 2010), regulating mitochondrial dynamics (McQuibban et al., 2006), wound healing (Cheng et al., 2011) and endoplasmic reticulum associated protein degradation (ERAD) (Fleig et al., 2012).

Extensive bioinformatic analyses of the rhomboid proteases revealed rhomboid homologues in eukaryotes that lacked the catalytic residues (Koonin et al., 2003, Lemberg and Freeman, 2007), rendering these proteins inactive. These inactive proteases were termed pseudoproteases and form part of the rhomboid-like superfamily. In mammals, the superfamily comprises of fourteen

members, of which nine are pseudoproteases (Figure 1.1). It is likely that the pseudoproteases evolved by gene duplication of an ancient active rhomboid followed by the loss of catalytic activity, then diversified to gain new functions. Each member has distinct cellular functions but share some predicted structural similarities (Adrain and Freeman, 2012) including the same six core conserved TMHs known as the rhomboid fold. In some members, the core six TMHs are extended either at the N- or C- terminus by an additional TMH or larger protein domains in loops or at the termini. The active site of rhomboid proteases consists of a conserved catalytic dyad composed of serine and histidine, located on the fourth and sixth TMH of the rhomboid core TMHs respectively. The other conserved features of the rhomboid-like superfamily are the GxxxG motif and tryptophan and arginine (WR) motif in the L1 Loop between TMH1 and TMH2 of the rhomboid core (Figure 1.1B). The GxxxG motif is a stabilisation motif and the WR motif is involved in stabilisation of the L1 Loop, these motifs are discussed further in Section 1.4.1. However the iRhoms do not have an embedded L1 Loop like other members of the superfamily whereas the WR motif varies across the other inactive proteases: Derlins (WR), RHBDD2-3 (HR), TMEM115 (WT), UBAC2 (WR).

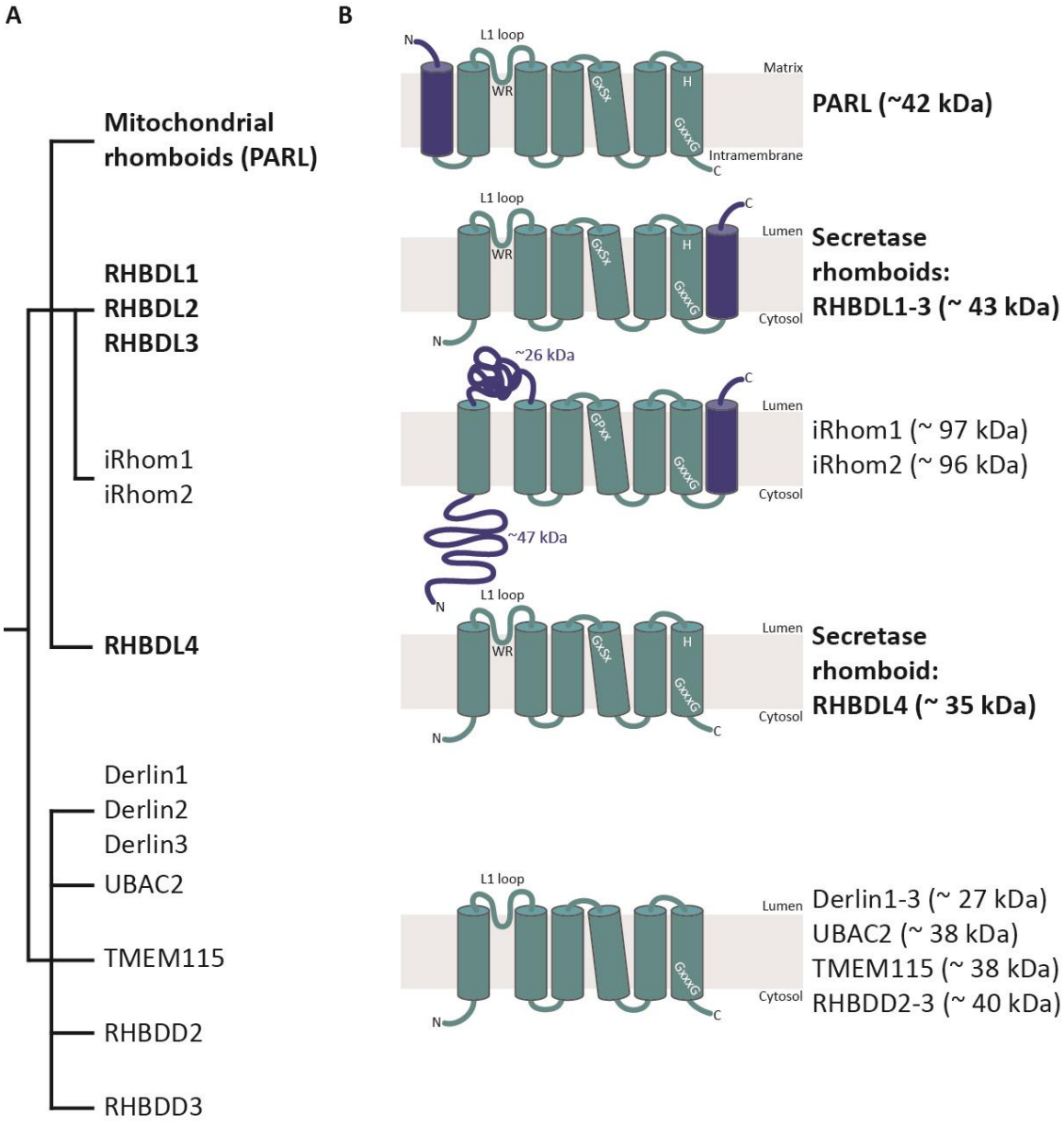


Figure 1.1. iRhoms are members of the rhomboid-like superfamily. An overview of the rhomboid-like superfamily in mammals. The tree shown is representation of their classification **(A)**. The superfamily is divided in two branches according to their conservation in sequence and function: the top branch comprises of the rhomboid active proteases (bold) and the closely related iRhoms and the bottom branch includes the remaining inactive proteases. Schematics of different members are shown in **(B)**. The six transmembrane helix (TMH) core rhomboid topology is shown in teal and additional domains are shown in dark blue. The catalytic serine (GxSx) and histidine (H) dyad is shown on TMH4 and TMH6, respectively, of the core rhomboid fold and the interaction motif (GxxxS) on TMH6. The WR motif is present in all the rhomboids that have L1 loop embedded in the membrane whereas this motif is different in the other inactive rhomboids: Derlins (WR), RHBDD2-3 (HR), TMEM115 (WT), UBAC2 (WR). iRhoms contain a GPxx motif on TMH4 that impairs catalytic activity. The classification is adapted from (Freeman, 2014) (copyright licence number: 1051769-1) and the schematic of the superfamily members are adapted from (Ticha et al., 2018) (copyright licence number: 4877680608671).

1.1.2 iRhoms

iRhoms are inactive, or pseudoprotease, members of the rhomboid-like superfamily and are the pseudoproteases most phylogenetically related to the active secretase rhomboids (RHBDL1-3) (Bergbold and Lemberg, 2013, Lemberg and Freeman, 2007). iRhoms lack either one or both of the catalytic residues, with the exception of the *Caenorhabditis elegans* iRhom which has an intact catalytic dyad (Lemberg and Freeman, 2007). Instead of the catalytic serine present within the GxSx motif found within TMH4 of rhomboid proteases, iRhoms have a conserved proline (GPxx motif) thought to abolish catalytic activity (Lemberg and Freeman, 2007). In order to support this hypothesis, the amino acid residue immediately before the catalytic serine, was mutated to proline in *Drosophila* Rhomboid-I and mouse RHBDL2 (Zettl et al., 2011). The mutation abolished enzymatic activity in Rhomboid-I and strongly reduced activity in RHBDL2 (Zettl et al., 2011). iRhom was also tested against known rhomboid substrates at the time, but was not able to cleave them (Zettl et al., 2011). Therefore the presence of the GPxx motif in *C. elegans* iRhom abolishes activity despite having an intact catalytic dyad.

1.1.2.1 iRhom expression

iRhoms are conserved in metazoans but are not present in prokaryotes or unicellular eukaryotes. Most vertebrates and *C. elegans* have two paralogues, *RHBDF1* coding for iRhom1 and *RHBDF2* coding for iRhom2 (Ensembl Genome Browser (Yates et al., 2019)), whereas *D. melanogaster* and some fish species, such as *Danio rerio*, only encode a singular iRhom gene. Mammalian iRhom1 and iRhom2 differ in tissue expression. Mammalian iRhom1 is broadly expressed (Christova et al., 2013, Li et al., 2015) and mammalian iRhom2 is predominantly expressed in immune cells and skin (Christova et al., 2013). iRhoms are predominately localised in the endoplasmic reticulum (ER) (Zou et al., 2009, McIlwain et al., 2012, Hosur et al., 2014, Luo et al., 2016, Kunzel et al., 2018). iRhoms are also localised, in smaller pools, in the Golgi apparatus (Zou et al., 2009, McIlwain et al., 2012, Maney et al., 2015, Luo et al., 2016), plasma membrane

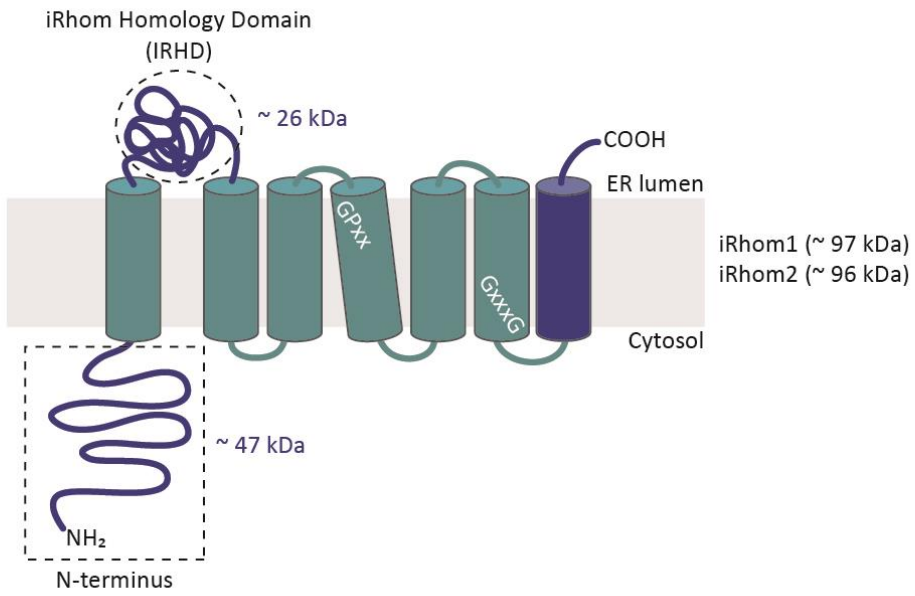
(Maney et al., 2015, Grieve et al., 2017, Kunzel et al., 2018) and lysosomes (Luo et al., 2016, Kunzel et al., 2018).

1.1.2.2 iRhom topological structure

In contrast to the secretase rhomboids, iRhoms consist of a long, extensive cytoplasmic N-terminal tail that makes up almost half the amino acids of the protein, and a highly conserved cysteine-rich loop between TMH1 and TMH2 named the iRhom homology domain (IRHD) inserted in the L1 Loop (Figure 1.2A). iRhoms also have an additional TMH at the C-terminal end of the rhomboid core that extends the C-terminus of the protein to the ER lumen/extracellular space. Although the overall amino acid sequence identity between human iRhom1 and iRhom2 is 61 %, there are variations when comparing domains within iRhoms: the IRHD has the highest similarity (Lemberg and Freeman, 2007) whereas the N-terminal cytoplasmic tail is the most divergent (Figure 1.2B) (Dulloo et al., 2019).

The cytoplasmic N-terminal tail of iRhoms has important regulatory functions but is not well defined or characterised, as is discussed further in Section 1.3.4. The IRHD contain sixteen cysteines (Lemberg and Freeman, 2007), except for *Anopheles gambiae* iRhom which only has fifteen. These cysteines are likely to form disulphide bonds because the loop resides in an oxidising environment of secretory pathway and extracellular space. However, the function of this loop is not fully characterised but has been suggested to have an important role in the binding and maturation of a disintegrin and metalloproteinase 17 (ADAM17) (Grieve et al., 2017). The role of iRhom in ADAM17 trafficking and processing is described in more detail in Section 1.3.1.

A



B

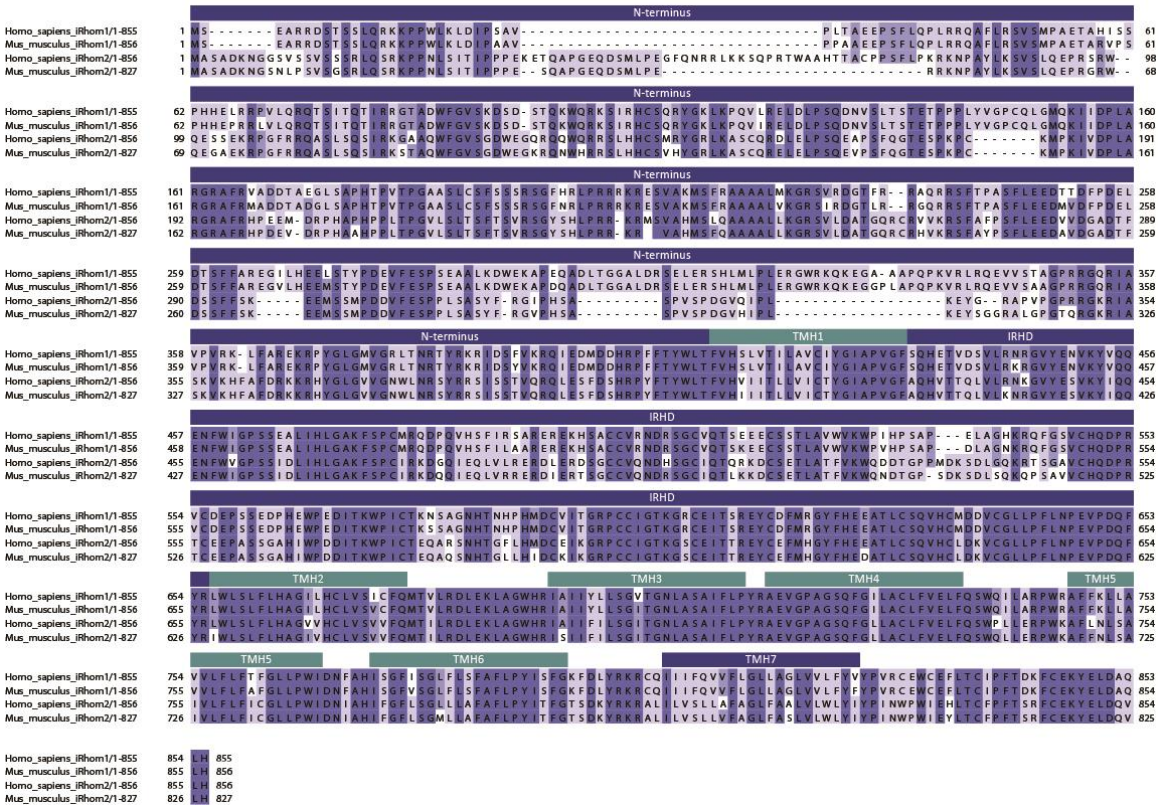


Figure 1.2. Schematic of iRhom topology. iRhoms consist of the six core rhomboid transmembrane helices (TMH) with an additional TMH at the C-terminus and long N-terminus **(A)**. The iRhom homology domain (IRHD) is inserted in the L1 loop between TMH1 and TMH2. The interaction motif (GxxxG) is conserved from the wider rhomboid-like superfamily and the GPxx on TMH4 abrogates catalytic activity. **(B)** Multiple sequence alignment of human (*Homo sapiens*) and mouse (*Mus musculus*) iRhom1 and iRhom2 with the N-terminus, IRHD and THMs shown. Dark purple represents high conservation.

1.2 ADAM17 (A disintegrin and metalloproteinase 17)

1.2.1 Intercellular signalling

Cell to cell, or intercellular, communication is essential for multicellular organisms to optimise cellular function in the context of other cells. Intercellular signalling is not limited to metazoans as unicellular eukaryotes and prokaryotes are able to communicate with their counterparts to coordinate their functions (Alberts et al., 2008). Cells communicate by releasing signalling factors or molecules into the extracellular space whereby these signalling factors are received by the target cell. Once these signals reach their target cell, they bind onto cell surface receptors triggering an intercellular signalling cascade to produce the appropriate/desired response of the target cell. Many signalling factors are synthesised as membrane-tethered precursors that need to be cleaved at the plasma membrane in order to liberate the signalling factor from the cell. This process of proteolytic cleavage of extracellular domains at the cell surface is called ectodomain shedding (Harris et al., 2003, Hayashida et al., 2010).

EGFR ligands are one of the most studied signalling factors (Harris et al., 2003, Hayashida et al., 2010, Adrain and Freeman, 2014). EGFR ligands are composed of one TMH, in the Type I orientation, with one or more EGF domains in the extracellular N-terminus. Shedding of the EGFR ligand, from the signalling cell and its subsequent release into the extracellular space, binds and activates EGFRs (ERBB1-4). The activation of EGFRs triggers intracellular signalling pathways which eventually cause cell differentiation, proliferation or migration (Adrain and Freeman, 2014).

In mammals, the ADAM (A disintegrin and metalloproteinase) family of proteins are responsible for ectodomain shedding of a number of EGFR ligands (Blobel et al., 2009). ADAM10 and ADAM17 are the key proteases for shedding EGFR ligands (Blobel, 2005) where ADAM10 cleaves EGF and betacellulin and ADAM17 is the main enzyme for amphiregulin, epigin, epiregulin,

heparin-binding EGF, transforming growth factor alpha (TGFα) (Sahin et al., 2004, Sahin and Blobel, 2007).

1.2.2 ADAM17 function and topological structure

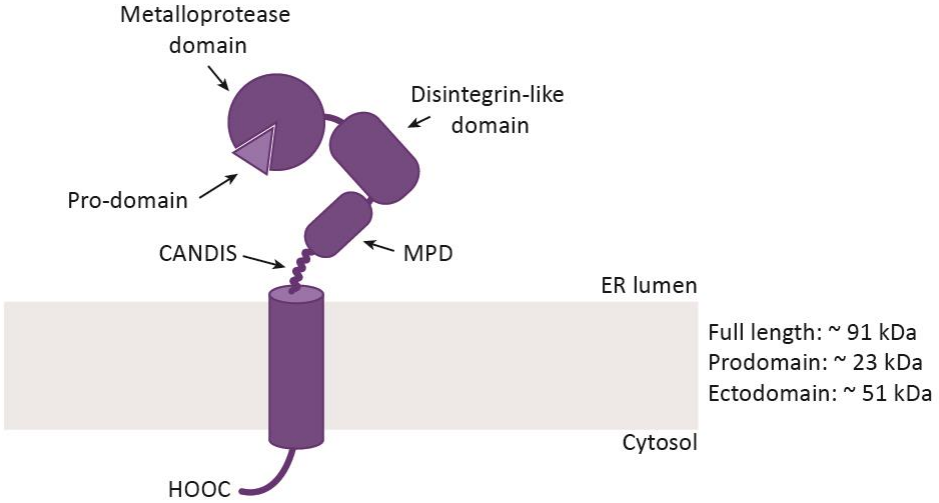
ADAM17 was initially discovered as the sheddase responsible for the release of soluble tumour necrosis factor alpha (TNFα) from its membrane tether (Black et al., 1997, Moss et al., 1997). Therefore ADAM17 is also known as the TNFα converting enzyme (TACE). Soluble TNFα activates two receptors, TNFα receptor (TNFR) 1 and TNFR2, and activation of the TNFRs leads to the transcriptional activation of pro-inflammatory cytokines (Lewis et al., 1991). TNFα initiates the innate immune response to bacterial infection and therefore its hyperactivation leads to inflammatory diseases such as rheumatoid arthritis (Kalliolias and Ivashkiv, 2016).

Currently, ADAM17 in mammals has over 90 substrates identified *in vitro* (Düsterhöft et al., 2019) but not all of these have been verified *in vivo* or are physiologically relevant. The best characterised substrates include the EGFR ligands listed in Section 1.2.1, TNFα, the primary inflammatory cytokine, interleukin 6 receptor and L-selectin (Düsterhöft et al., 2019). An ADAM17 homologue is found in *Drosophila* but up until recently the function of ADAM17 in *Drosophila* was largely unknown. The loss of ADAM17 in *Drosophila* retinal cells results in the abnormal accumulation of lipid droplets in the glial-like cells. The accumulation of lipid droplets initially is protective against reactive oxygen species but later in life the lipid droplets become peroxidated which then degenerate the *Drosophila* retina (Muliyil et al., 2020).

ADAM17 is a single pass membrane protein in the Type I orientation and consists of multiple protein domains in the luminal/extracellular space (Figure 1.3) (Zunke and Rose-John, 2017, Düsterhöft et al., 2019). The multiple protein domains are the inhibitory pro-domain, metalloprotease domain, disintegrin-like domain, membrane proximal domain (MPD) and

CANDIS (conserved adam seventeen dynamic interaction sequence) (reviewed in (Grötzinger et al., 2017)). The catalytic site is located in the metalloprotease domain where the amino acid motif HExGHxxGxxHD coordinates a zinc ion (Edwards et al., 2008). The structure of the metalloprotease domain was first solved by X-ray crystallography in 1998 which first illustrated the catalytic site (Maskos et al., 1998) and since then multiple crystal structures of the metalloprotease domain have been solved. The only other additional structural information of ADAM17 that has been solved is the MPD by nuclear magnetic resonance (NMR) spectroscopy in open and closed conformations (Düsterhöft et al., 2013) leading to the hypothesis that protein disulphide isomerase at the cell surface reshuffle the disulphide bridges within the MPD. This reshuffling may act as a molecular switch and potentially regulating shedding activity of ADAM17 (Düsterhöft et al., 2013). But to date there is no structure of the full ectodomain or full length ADAM17. The only structural insight of ADAM17 can be inferred from the crystal structure of the ectodomain of ADAM10 (Seegar et al., 2017) which may suggest that the ectodomain of ADAM17 may conform to a C-shape (Figure 1.3) but only a full length structure will be able to confirm this because the sequence identity between ADAM17 and ADAM10 is only ~ 30 %.

A



B

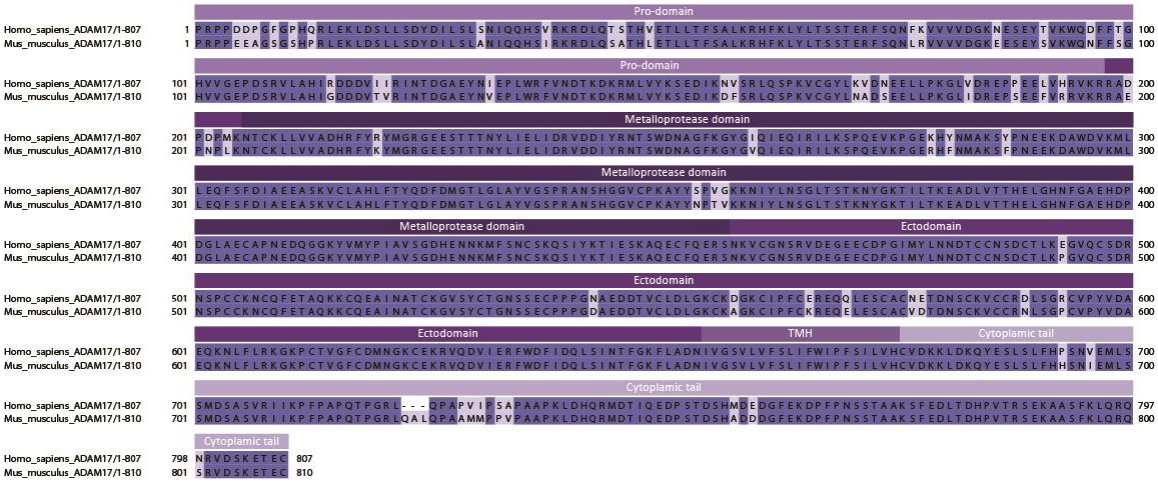


Figure 1.3. Schematic of ADAM17 structure. (A) The proposed C-shape of the ectodomain of ADAM17 adapted from (Düsterhöft et al., 2019) (copyright licence number: 4867710162870) with domains labelled. The N-terminal pro-domain is followed by the metalloprotease domain, disintegrin-like domain, membrane proximal domain (MPD), conserved adam seventeen dynamic interaction sequence (CANDIS), transmembrane helix (TMH) and cytoplasmic C-terminus. **(B)** Multiple sequence alignment of human (*Homo sapiens*) and mouse (*Mus musculus*) ADAM17 with the pro-domain, metalloprotease domain within the ectodomain, TMH and cytoplasmic tail shown. Dark purple represents high conservation.

1.2.3 ADAM17 regulation

ADAM17-dependent proteolysis is a post-translational modification that irreversibly changes the target substrate, and therefore must be tightly controlled and regulated from synthesis to degradation. ADAM17 is first synthesised in the ER as a zymogen, whereby the catalytic site in the metalloprotease domain is inhibited by the pro-domain. Once proADAM17, also referred to as immature ADAM17, migrates to the trans-Golgi network, the pro-domain is removed by furin-like convertases to form mature ADAM17 (Endres et al., 2003). Mature ADAM17 traffics to the plasma membrane where is it activated to cleave its substrates. One of the most characterised regulators of ADAM17 are the iRhoms which is discussed further in Section 1.3.1.

1.3 iRhoms and their functions

1.3.1 iRhoms are regulatory subunits of ADAM17

One of the most characterised function of iRhoms is their regulation of ADAM17 throughout the secretory pathway (Adrain et al., 2012, McIlwain et al., 2012) (reviewed in (Düsterhöft et al., 2019)). Both iRhoms bind to pro- and mature ADAM17 (Adrain et al., 2012, McIlwain et al., 2012, Christova et al., 2013) but do not bind to the closely related ADAM10 (Adrain et al., 2012). iRhoms bind to proADAM17 in the ER and transport the complex to the Golgi apparatus; there the inhibitory pro-domain of ADAM17 is cleaved by the Golgi resident enzyme furin. Then the complex traffics to the cell surface where ADAM17, now mature, is activated and able to shed its substrates. An overview of iRhoms role in trafficking and maturation of ADAM17 shown in Figure 1.4. The loss of both iRhoms results in the complete loss of mature ADAM17 trafficking to the cell surface and ADAM17-dependent shedding of TNFα (Christova et al., 2013, Li et al., 2015). There is some redundancy between iRhom1 and iRhom2; if one of them is knocked out in mice the other can partially rescue the loss in most tissues (Christova et al., 2013, Li et al., 2015). However, the rescue does not occur when only one iRhom is expressed, for example iRhom2 in macrophages, where all of ADAM17 maturation and activity is controlled iRhom2 (Issuree et al., 2013). The removal of IRHD in iRhom2 impaired the ability of iRhom2 to bind to proADAM17 but not much more is known about the role of the IRHD in ADAM17 binding and maturation (Grieve et al., 2017).

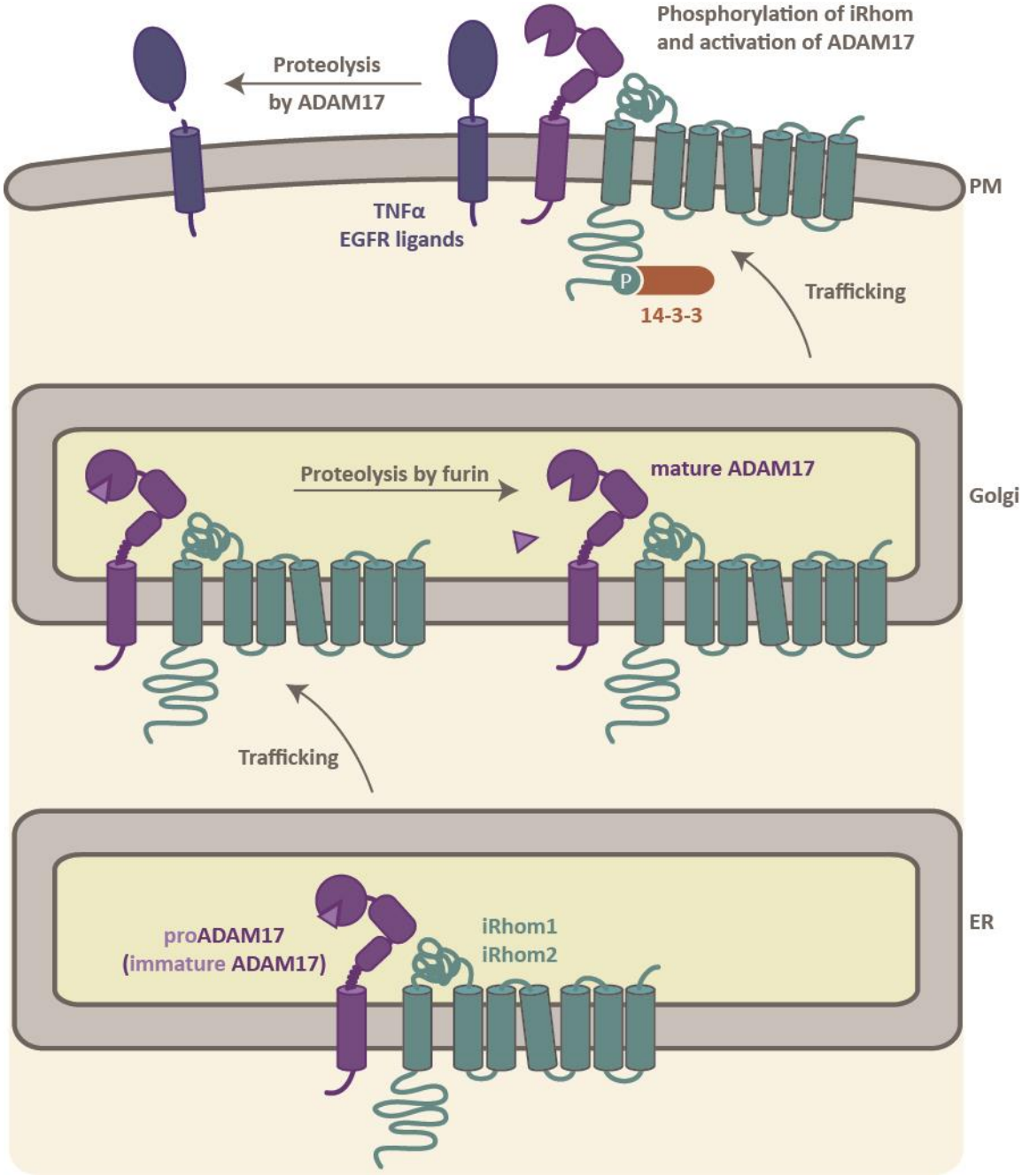


Figure 1.4. iRhoms are regulatory subunits of ADAM17. Mammalian iRhoms (iRhom1, iRhom2) traffic proADAM17, or immature ADAM17, from the endoplasmic reticulum (ER) to the Golgi apparatus (Adrain et al., 2012, McIlwain et al., 2012). In the Golgi, the pro-domain of ADAM17 is cleaved by furin. The mature ADAM17 and iRhom complex traffics to the plasma membrane (PM) where by the N-terminus of iRhom becomes phosphorylated and recruits 14-3-3 proteins (Grieve et al., 2017). The phosphorylation of iRhom activates ADAM17 enabling the cleavage of ADAM17 dependent substrates, TNFα and EGFR ligands.

iRhoms also play a role in ADAM17 substrate selectivity: which iRhom binds to ADAM17 influences the substrate repertoire (Maretzky et al., 2013). Maretzky and colleagues reported that iRhom2 knock out (KO) mouse embryonic fibroblast (MEF) cells cannot shed ADMA17 substrates heparin-binding EGF and kit ligand 2, but iRhom2 KO MEF cells can shed TGFα. The shedding of TGFα was reduced when iRhom1 was knocked down in the iRhom2 KO MEFs (Maretzky et al., 2013). Furthermore, there is evidence that the regions immediately before or after the TMH, termed juxtamembrane domains (JMD), flanking the IRHD of iRhom2 and ADAM17 regulate the stimulation and substrate selectivity of ADAM17 (Tang et al., 2020). A JMD mutant of ADAM17 reduced the shedding of iRhom2 selective substrates (heparin-binding EGF and kit ligand 2). A JMD deletion on the luminal side of TMH1 and TMH2 of iRhom2 abolished kit ligand 2 shedding and reduced shedding of TGFα.

It has been shown that the N-terminus of iRhom is required for the rapid activation of ADAM17 rather than the cytoplasmic C-terminus of ADAM17 (Grieve et al., 2017). Three distinct sites within the N-terminus of iRhom2 are phosphorylated in an MAPK1/3-dependent manner at the plasma membrane, the phosphorylation sites are conserved in iRhom1. Upon phosphorylation, 14-3-3 proteins are recruited and bind to the phosphorylation sites (Grieve et al., 2017). The binding then alters the interaction of mature ADAM17 and iRhom at the cell surface resulting in the upregulation of ADAM17-dependent shedding. Increasingly, the function of the N-terminus of iRhoms is becoming a signal-receiving hub to enable its functions. A mass spectrometry screen identified FERM domain-containing protein 8 (FRMD8) as a binding partner of iRhom2 (Kunzel et al., 2018, Oikonomidi et al., 2018). The binding site of FRMD8 was identified between amino acids 200 to 300 on the N-terminus of iRhom2. FRMD8, also known as iTAP, was shown to stabilise the iRhom2/ADAM17 complex at the plasma membrane by protecting the complex from lysosomal degradation (Kunzel et al., 2018, Oikonomidi et al., 2018) (Figure 1.5). However, it appears that FRMD8 binding to the N-terminus is independent from 14-3-3 binding. It is not

known if there is a relationship between the phosphorylation-dependent binding of 14-3-3 and stabilisation of the complex by FRMD8. A fusion of human FRMD8 to a 300 amino acid truncation of the N-termination of iRhom show enhanced stabilisation of the chimeric protein at the cell surface (Kunzel et al., 2018). In addition, ADAM17 has shown to be required for the stabilisation of iRhom2 but not for iRhom1 (Weskamp et al., 2020).

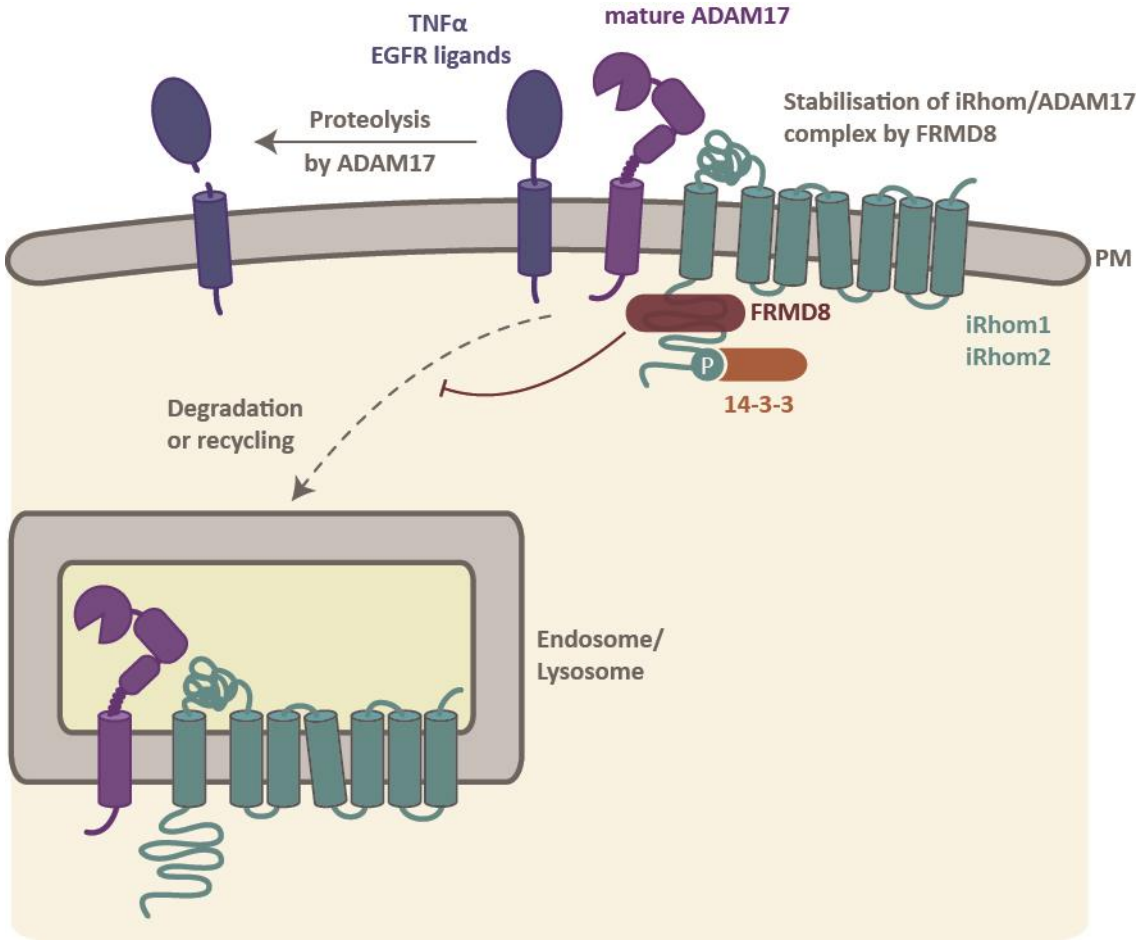


Figure 1.5. FRMD8 stabilises the iRhom/ADAM17 complex by preventing the degradation of the complex. FRMD8 binds to the N-terminus of iRhom, between amino acids 200-300, and stabilises the complex at the plasma membrane (PM) by preventing the degradation of the complex via the endosomal pathway (Kunzel et al., 2018).

1.3.2 Other iRhom binding partners

iRhom2 plays an important role in the innate immunity in response to viral infections. Recently, iRhom2 has been shown to bind to and positively regulate innate immunity adapters, stimulator of interferon genes (STING) (Luo et al., 2016) and mitochondrial antiviral signalling protein (MAVS) (also referred to as VISA) (Luo et al., 2017), in response to infection of DNA and RNA viruses respectively. iRhom2 KO mice infected with herpes simplex virus 1 (HSV-1), a DNA virus, developed lethargy and ataxia three days post infection and died one after displaying symptoms (Luo et al., 2016). Similarly iRhom2 KO mice infected with vesicular stomatitis virus (VSV), a RNA virus, presented similar symptoms when infected with HSV-1 and died within four days of presenting symptoms (Luo et al., 2017). In addition, VSV viral titres were higher in spleen and liver of iRhom2 KO mice compared to iRhom2 WT mice (Luo et al., 2017). However it was reported that STING is not required for the stability of iRhom2 (Weskamp et al., 2020).

1.3.3 Other functions of iRhoms

In contrast to iRhoms' role in the ADAM17-dependent shedding of EGFR ligands, iRhom also acts as a negative regulator of EGFR ligands in *Drosophila*. iRhom KO *Drosophila* exhibit a sleep-like phenotype and undergo periods of inactivity (Zettl et al., 2011). In addition, activation of EGFR signalling in *Drosophila* also exhibited a similar phenotype to the iRhom KO flies (Foltenyi et al., 2007), therefore suggesting that the iRhom2 KO phenotype is due to the hyperactivation of EGFR. The sleep-like phenotype was rescued by the knock down of EGFR or EGFR ligand, Spitz, in iRhom KO *Drosophila* which provided more evidence to support the hypothesis that iRhom in flies is a negative regulator of EGFR ligands (Zettl et al., 2011). Cell culture experiments where iRhom was overexpressed caused a decrease in Spitz and the secretion of soluble Spitz or cleavage fragments were not detected in cell lysates (Zettl et al., 2011). Taking advantage of the rough eye phenotype as a readout of EGFR dysregulation (Bai et al., 2001), knock down of iRhom expression suppressed the rough eye phenotype caused by inhibition of EGFR (Zettl et al., 2011).

The mechanism to help facilitate iRhom in *Drosophila* decreasing EGFR ligands was proposed to involve components of the ER-associated degradation (ERAD). Evidence for this hypothesis was supported by rough eye phenotype in *Drosophila* was partially rescued of Hrd1 or EDEM2 (Zettl et al., 2011). In conclusion, *Drosophila* iRhom was discovered as a novel negative regulator of EGFR ligands, therefore decreasing EGFR signalling, perhaps by an ERAD-dependent manner (Zettl et al., 2011). Overexpressed iRhoms in COS-7 cells strongly decreased the level of EGF. Therefore the role of iRhom in inhibiting EGFR ligands in *Drosophila* appears to be conserved in mammals. Both in mammals and *Drosophila*, the inhibitory effect of iRhoms on EGFR ligands is independent from active rhomboid proteases cleaving EGFR ligands.

1.3.4 Pathological functions of iRhoms

iRhom1 KO mice, generated by the definitive-null KO strategy, failed to gain weight, suffered from multi-organ defects and died within six weeks (Christova et al., 2013). The definitive-null KO strategy used by Christova and colleagues was deleting exons 2 to 18, whole gene deletion. They also found the iRhom1 and iRhom2 double knock out (DKO) resulted in embryonic lethality. Whereas iRhom1 KO mice generated by the targeted KO-first allele approach appeared normal with normal heart and brain functions (Li et al., 2015). The targeted KO-first allele approached used by Li and colleagues was deleting exons 4 to 11. And their iRhom1 and iRhom2 DKO mice resulted in perinatal lethality with eye-lids open phenotype. This caused confusion in the field to why one approach of generating iRhom1 KO and iRhom1/2 DKO mice would differ drastically in phenotype to the other.

It was recently reported by Hosur and colleagues that, using the iRhom1 KO mice as a case study, genes respond differently to different KO strategies (Hosur et al., 2020). In this study, the authors generated iRhom1 KO mice that were analogous to the definitive-null KO by deleting exons 3-17 and the targeted KO-first allele by deleting exons 2-3, which removed the start

codon. They found that their definitive-null iRhom1 KO mice failed to gain weight, suffered from brain haemorrhages, cardiac fibrosis and died within 4 weeks (Hosur et al., 2020). And their KO-first allele iRhom1 mice were viable and fertile. Their analogous iRhom1 KO mice phenotypes were similar to the reported iRhom1 KO mice from Christova et al. and Li et al. Hosur and colleagues also found that the gene product of the KO-first allele of *RHBDF1* was a truncated isoform of iRhom1 that was able rescue the pathology of the definitive-null KO iRhom1 mouse. They concluded that different KO strategies can result in different phenotypes and genes are able to circumvent certain strategies to produce an isoform gene product (Hosur et al., 2020).

The iRhom2 KO mice present no obvious deficiencies, are fertile and viable (Adrain et al., 2012, McIlwain et al., 2012, Christova et al., 2013, Siggs et al., 2014); but show defects when in their innate immunity when immunologically challenged by lipopolysaccharide (LPS), a bacterial cell wall derivative, (McIlwain et al., 2012, Luo et al., 2016, Luo et al., 2017) and skin defects (Maruthappu et al., 2017). The absence of iRhom2 in macrophages, ADAM17 is retained in the ER and therefore does not undergo its maturation through the Golgi apparatus and onto the plasma membrane. The consequence of the lack of ADAM17 on the plasma membrane results in no secretion of TNFα (Adrain and Freeman, 2012, McIlwain et al., 2012, Siggs et al., 2012) and so iRhom2 KO mice were susceptible to infection from *Listeria monocytogenes* (McIlwain et al., 2012). However, iRhom2 KO mice were also protected against septic shock from a lethal dose of LPS (McIlwain et al., 2012) and rheumatoid arthritis (Issuree et al., 2013). The protection from rheumatoid arthritis lead to an idea that targeting iRhom2 could be cell-specific drug target against TNFα related diseases (Issuree et al., 2013, Lichtenthaler, 2013).

Mutations in the N-terminus of iRhom2 result in skin and hair follicle defects in both mice and humans. Single point mutations in a highly conserved area of the N-terminus cause an inherited cancer syndrome called tylosis with oesophageal cancer (TOC) (Saarinen et al., 2012, Blaydon et

al., 2012, Mokoena et al., 2018). TOC is a dominant autosomal disease causing the thickening of skin of the palms of hands and soles of feet with the high risk of developing oesophageal squamous cell carcinoma. Mouse models mimicking human TOC mutation in mouse iRhom2 show a loss in hair and increased wound healing, likely caused by the increase in amphiregulin (Hosur et al., 2017). The TOC mutations lead to enhanced activity of ADAM17 and the constitutive shedding of EGFR ligands however the mechanism by which these mutations cause this phenotype are unknown (Maney et al., 2015). Two additional mouse models, comprising of two different genomic deletions, both showed skin and hair follicle defects. These mouse models were: the *curly-bare* (*cub*) mutation where most of the N-terminus was removed (amino acids 1 to 268) (Hosur et al., 2014, Siggs et al., 2014) and the *uncovered* (*uncv*) mutation where amino acids 118 to 191 were deleted (Leilei et al., 2014, Liu et al., 2015, Yang et al., 2016). In both *cub* and *uncv* mutants, the highly conserved tylotic region was removed providing further evidence that this region in iRhom2 is important in skin homeostasis. However iRhom2 KO mice display a normal hair coat (Hosur et al., 2014, Leilei et al., 2014, Siggs et al., 2014) implying that a mutation in tylotic region is a gain-of-function mutation. Additionally the N-terminal domain mutations of iRhom2 can still bind to ADAM17 to aid its maturation and trafficking through the secretory pathway but ADAM17 is quickly degraded via the lysosome (Siggs et al., 2014, Grieve et al., 2017). Here, the N-terminus of iRhom2 provides a stabilising role to ADAM17 which was later shown to be contributed by the binding of FRMD8 to the N-terminus (Kunzel et al., 2018, Oikonomidi et al., 2018). iRhom1 and iRhom2 both bind to FRMD8 and share the phosphorylation site where 14-3-3 protein (Kunzel et al., 2018) bind but no TOC-like disease mutations have been reported on iRhom1. This may be a consequence that iRhom1 is widely expressed and therefore likely make these mutations lethal. But as the N-terminus is the most divergent region between iRhom1 and iRhom2, it is possible that there may be iRhom paralogue specific regulatory functions. Overall the N-terminus of iRhoms (Figure 1.6) is increasingly becoming seen as a signalling hub to bind with other proteins to enable iRhoms functions: 14-3-

3 in activating ADAM17 at the plasma membrane, FRMD8 in stabilising the iRhom/ADAM17 complex at the cell surface and the tylotic region in skin homeostasis, though more is yet to be understood about the mechanism of this region.

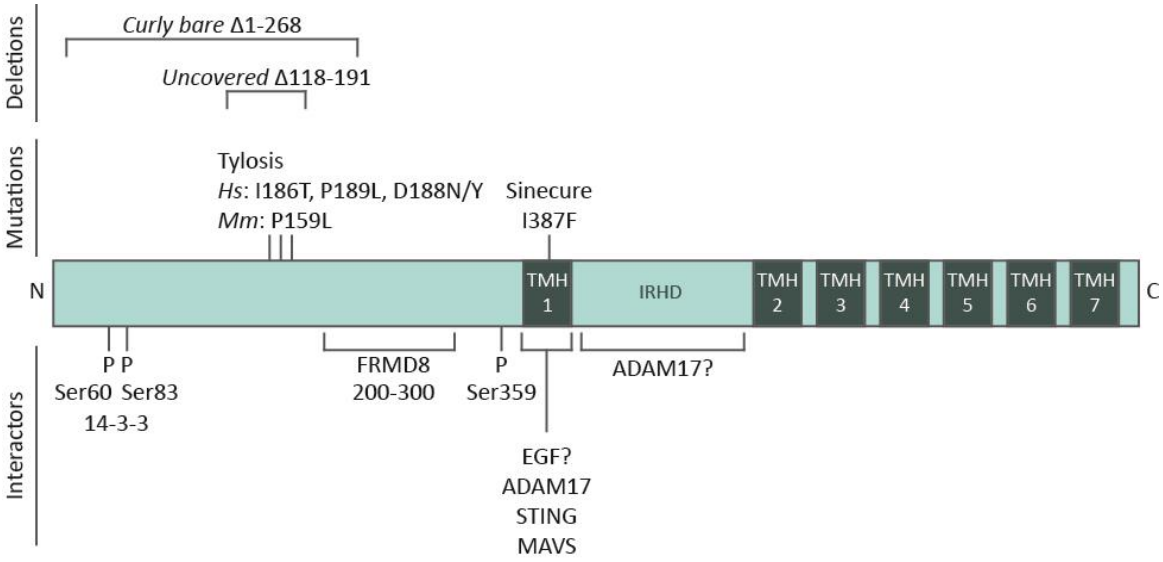


Figure 1.6. Schematic of iRhom2 with their phosphorylation sites and interacting partners. The transmembrane helices (TMH) are shown in teal and the iRhom homology domain (IRHD) between TMH1 and TMH2. The N-terminus harbours the phosphorylation sites at which 14-3-3 binds, the FRMD8 binding site, mutations implicated in tylosis oesophageal cancer and mouse model deletions. The sinecure mutation has shown to impair interactions with ADAM17 and the IRHD is suggested to bind and aid maturation of ADAM17. The figure is adapted from (Dulloo et al., 2019).

1.4 Rhomboid structure and their client/substrate recognition

1.4.1 Known rhomboid structures

To date there are only two protein members of the rhomboid-like superfamily that have their protein structures solved: the active intramembrane protease GlpG from *Escherichia coli* (Wang et al., 2006) and, more recently, the inactive rhomboid Derlin-1 from *Saccharomyces cerevisiae* (Wu et al., 2020b).

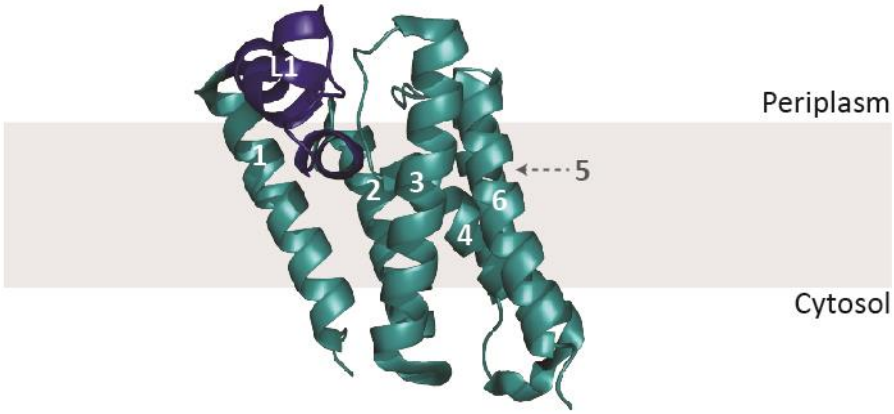


Figure 1.7. The crystal structure of *Escherichia coli* active rhomboid protease, GlpG. The crystal structure of GlpG (PDB ID: 2IC8) was first solved to 2.10 Å (Wang et al., 2006) and illustrates the core rhomboid fold of the six transmembrane helices. The L1 loop is highlighted in dark blue.

The crystal structure of *E. coli* GlpG (Figure 1.7) was a turning point in rhomboid protein biology and intramembrane proteases. The structure first illustrated the core rhomboid fold formed by five TMH surrounding the central TMH4 and the hairpin loop, called the L1 loop, which is laterally embedded in the membrane between TMH1 and TMH2 (Wang et al., 2006, Vinothkumar, 2011, Baker and Urban, 2012). After the first structure of GlpG was reported, it was quickly followed by other structures of *E. coli* GlpG (Wu et al., 2006, Ben-Shem et al., 2007) and the *Haemophilus influenzae* GlpG (Lemieux et al., 2007). The L1 loop is shown to be involved in substrate binding (Zoll et al., 2014) and can act as a 'flotation device' to ensure that the

rhomboid TMH bundle is in the correct orientation in the membrane (Bondar et al., 2009). Interestingly, the active site of GlpG is located within the TMH bundle and exposed to water thereby allowing the enzyme to cleave membrane helices within GlpG rhomboid core (Wang et al., 2006). It was proposed that the active site is protected on the periplasmic side by the L5 loop, between TMH5 and TMH6, and gated by TMH5 (Baker et al., 2007). Therefore the active site is protected from the hydrophobic membrane to ensure hydrolysis of the peptide bond.

Mutagenesis experiments showed that GlpG is stabilised by GxxxG interaction motifs in TMH4 and TMH6 and hydrogen bonding (Baker and Urban, 2012). Stabilising hydrogen bonds were in two particular regions, the cytoplasmic side bridging the apex of TMH1, TMH2 and TMH3 and the hydrogen bonding within the L1 loop involving the highly conserved WR motif (Baker and Urban, 2012). Mutagenesis of the tryptophan, to alanine, of the WR motif in the L1 loop showed reduced proteolytic activity of GlpG without changing the overall structure of the protein (Wang et al., 2007). Instead, the polarity of the protein at the site of the mutation changed whereby more water molecules were present. Structural studies showed that the WR motif is at the distal tip of L1 loop, near the lipid:water interface, and the mutation no longer supports the buried nature of the wild type L1 loop (Wang et al., 2007). Therefore, the tryptophan is important in ensuring that the L1 loop is embedded in the membrane. The GxxxG interaction motif in GlpG is conserved across the wider rhomboid-like superfamily and the WR motif in the L1 loop conserved in all rhomboids except for iRhoms where the IRHD is inserted in the L1 loop. It is likely that these stabilising features of GlpG are preserved in the wider rhomboid-like superfamily (Ticha et al., 2018). Multiple crystal structures have been solved of GlpG in complex with inhibitors (reviewed in (Wolf and Verhelst, 2016)) which have provided insights to how GlpG may recognise its substrates. The proposed mechanisms of substrate access to GlpG are further discussed in Section 1.4.2.

The first crystal structure of GlpG was solved as a trimer, however, at the time of the study it was unclear if the oligomeric state was native or due to crystal packing (Wang et al., 2006). Subsequent studies have shown that rhomboid proteases are dimeric (Sampathkumar et al., 2012, Sherratt et al., 2012, Lazareno-Saez et al., 2013, Cournia et al., 2015) and that dimeric state is important for stability (Sampathkumar et al., 2012, Cournia et al., 2015) and substrate specificity (Lazareno-Saez et al., 2013, Cournia et al., 2015). Therefore, the trimer observed in the crystal structure (Wang et al., 2006) was likely due to crystal packing. The studies reporting that GlpG is oligomeric were *in vitro* because detergents were used to extract GlpG out of the membrane; hence, the true oligomeric state of GlpG *in vivo* was uncertain. Another study determined the oligomeric state of prokaryotic and eukaryotic members (protease and pseudoprotease) of the rhomboid-like superfamily utilising single-molecule photobleaching of rhomboids fused with a halo tag in whole cells, planar supported lipid bilayers and detergent micelles (Kreutzberger and Urban, 2018). The authors found that the all of the rhomboids studied were monomeric in whole cells and planar supported lipid bilayers and dimeric in detergent micelles (Kreutzberger and Urban, 2018). Therefore, GlpG forms unphysiological dimers in detergent micelles and the rhomboid superfamily maintains a monomeric state *in vivo* (Kreutzberger and Urban, 2018).

Recently, the structure of Derlin-1 (Der1) was solved by cryogenic electron microscopy (cryo-EM) as a component of the *S. cerevisiae* ERAD multimeric complex, Hrd1~Usa1-Der1-Hrd3 (Wu et al., 2020b) and is currently the only structure of a full length eukaryotic inactive rhomboid. The Hrd1 ERAD complex is responsible for the retrotranslocation of proteins into the cytosol for protein degradation. Here the structure of the complex was solved by fusing Hrd1 and Usa1 to form a chimeric, protein labelled Hrd1~Usa1 in complex with Derlin-1 and Hrd3. The structure of Derlin-1 is similar to the *E. coli* GlpG structure (Figure 1.8A-C) where the helices and loops align. However, the authors did not report the root-mean-square deviation (RMSD) between the two

structures. The coordinates of Derlin-1 were aligned with GlpG (Figure 1.8D and E) in Coot 0.8.9.1 (Winn et al., 2011) using the Secondary Structure Matching (SSM) Superimpose function (Krissinel and Henrick, 2004, Mitchell et al., 1990). Core RMSD achieved was 3.1014 Å with 139 residues aligned. Therefore the two structures align but are different, which is expected because the L5 loop is shorter in Derlin-1 than in GlpG and the L1 loop in both structures do not align well.

Derlin-1 in this multi-protein complex is in the closed state where TMH5 is in contact with TMH2 which resembles the closed state of GlpG. In this closed conformation of Derlin-1, the luminal side has a shallow hydrophilic cavity, similar in some respects to the active site of GlpG. However, the active site in GlpG is protected by the L5 loop between TMH5 and TMH6 and moves when the lateral gate opens by the displacement of TMH5. Whereas the L5 loop in Derlin-1 is much shorter and hence the luminal cavity is permanently open. Derlin-1 only binds to Hrd1 on the ER side of the membrane to form a 'half channel' and contributes to the thinning of the ER membrane. The role of Derlin-1 is to receive the hairpin loop from the unfolded protein substrate and, in conjunction with Hrd1 aids the retrotranslocation of the substrate across the membrane.

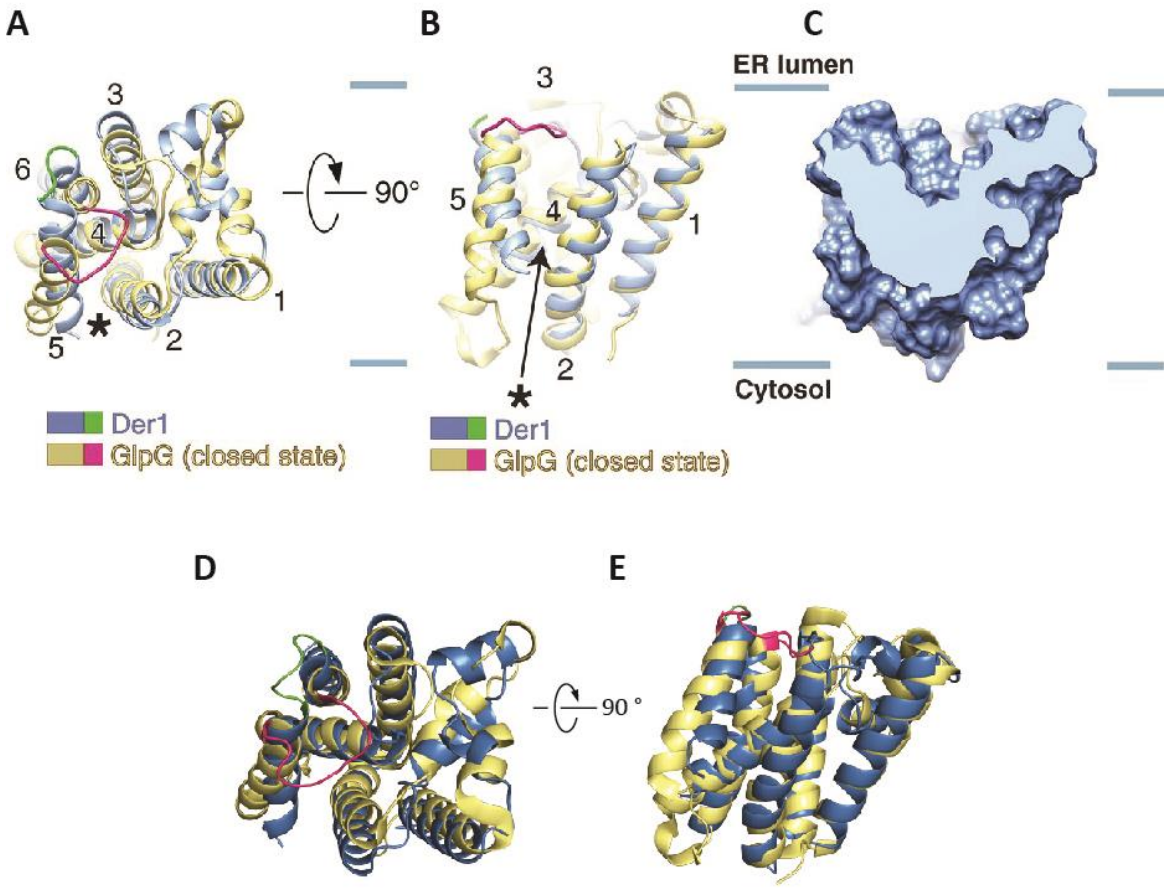


Figure 1.8. Overlay of the cartoon representation of Derlin-1 solved by cryo-EM and the crystal structure of GlpG. (A) Overlay of *Saccharomyces cerevisiae* Derlin-1 (Der-1) (blue) with *Escherichia coli* GlpG in the closed state (yellow) (PDB ID: 2IC8) shown from the lumen. The L5 loop (between TMH5 and TMH6) of Derlin-1 and GlpG is shown in green and pink, respectively and the lateral gate highlighted by the asterisk (*). **(B)** As shown in **(A)** but viewed in the plane of the membrane (blue lines) with the lateral gate at the front. **(C)** The cross section of the side view space-filling model of Derlin-1. **(D), (E)** Crystal structure of GlpG (PDB ID: 2IC8) and model coordinates of Derlin-1 (PBD ID: 6VJZ), extracted from the Hrd1~Usa1-Der1-Hrd3 complex, were aligned in Coot 0.8.9.1 (Winn et al., 2011) using the Secondary Structure Matching (SSM) Superimpose function (Krissinel and Henrick, 2004, Mitchell et al., 1990). Core root-mean-square deviation (RMSD) achieved was 3.1014 Å with 139 residues aligned. **(D)** is similar to **(A)** and **(E)** is similar to **(B)**. Panel **(A)**, **(B)** and **(C)** from this figure was copied from (Wu et al., 2020b) under copyright licence number: 4864190236459.

1.4.2 Rhomboid substrate/client protein recognition

It is proposed that rhomboid proteases recognise their substrates by scanning the stability of the substrate TMH. Rhomboids are able to recognise unstable helical features in their substrates thereby cleaving them (Urban and Freeman, 2003, Strisovsky et al., 2009, Moin and Urban, 2012).

The multiple crystal structures of GlpG in complex with and without inhibitors has provided insights to how substrates may bind and the active site, within the TMH bundle, becomes hydrated to hydrolyse its substrates. Multiple mechanisms have been proposed: the lateral gate between TMH2 and TMH5 opening by the displacement of the THM5 (Baker et al., 2007) or the opening of the L5 loop cap over the active site (Wang and Ha, 2007). The lateral gate model is supported by the mutagenesis of the TMH2 and TMH5 interface (Baker et al., 2007, Urban and Baker, 2008) but was opposed by another study where the authors found that crosslinking of TMH2 and TMH5 did not have a significant effect on hydrolysis (Xue and Ha, 2013). A recent study reported multiple crystal structures demonstrating the proteolytic cleavage in a number of stages which has provided a structure based mechanism of substrate binding and cleavage (Cho et al., 2019). The authors show the requirement of the gate to open for substrates, but this opening is not required for inhibitors as substrates and inhibitors have different paths to the active site. Additionally, they report that the L5 loop has multiple conformations throughout the reaction, first the loop retracts to allow substrate access to the active site whereby the substrate binds to the active through stabilising interactions. Then the L5 loop returns to 'clamp down' the substrate holding the substrate in the active site. The stabilisation of the substrate in the active site supports the hypothesis that rhomboid proteases scan their substrate helices for instability where unstable helices unravel will be cut and helices that maintain their secondary will not be cut. Remarkably the authors also report the crystal structure of the catalytic tetrahedral

intermediate of the substrate, suggesting that the tetrahedral intermediate is a stalling point rather than a transition state because this state was crystallised (Cho et al., 2019).

The lack of structural information on iRhoms and their interactions with ADAM17 lead to questions on how iRhoms may recognise their client proteins and interact with them. It is reasonable to draw some insight from current structural information gained from GlpG. It was first showed that iRhom2 binds with ADAM17 and not to the closely the related ADAM10 (Adrain et al., 2012) and the TMH swapping of ADAM17 to ADAM10 showed binding of a chimeric ADAM10 with iRhom2 (Cavadas et al., 2017). The main difference between the TMHs of ADAM17 and ADAM10 is that ADAM17 contains a proline. It is highly likely that this proline is destabilising the helix and possible that the iRhoms are able to recognise this. Molecular dynamic (MD) simulations of iRhom2 and ADAM17 show that the TMH1 of iRhom2 interact with the TMH of ADAM17 (Li et al., 2017) which is further supported by mutagenesis experiments. Li et al. investigated a mutation in TMH1 of iRhom2 termed sinecure mutation, I387F, which reduces iRhom2 function to facilitate ADAM17 maturation and shedding (Li et al., 2017). The MD simulations of ADAM17 show a kink in the TMH caused by the proline which is close to the location of the sinecure mutation. Current data suggests that the TMH1 of iRhom2 interacts with the TMH of ADAM17 rather than TMH2 and TMH5, mimicking the substrate binding site of GlpG (Dr. Iqbal Dulloo, personal communication). Instead, TMH2 and TMH5 of iRhom2 may recruit substrates for ADMA17 to shed (Ticha et al., 2018), however, this hypothesis based on known interactions from GlpG. Thus, more structural data is required to elucidate the roles of the other TMHs in iRhom2.

1.5 Structural determination by cryogenic electron microscopy

Structural biology is a field that aims to elucidate the 3-dimensional (3D) structure of biological macromolecules to help our understanding of the function of these molecules. Structural biology

in the 20th century was revolutionised by macromolecular X-ray crystallography (MX) a broadly used technique to solve the architecture of biological macromolecules in atomic detail (reviewed in (Maveyraud and Mourey, 2020)). However, this technique has limitations: a large amount of purified sample (milligram quantities) is required and the technique is dependent on the sample crystallising which is typically an empirical process. Even if the sample was to crystallise there are potential pitfalls because the crystal formed may not be well ordered and therefore diffract poorly which would necessitate optimising crystallisation conditions. An alternative technique, NMR spectroscopy, does not require sample to crystallise but does require large amounts of protein. For NMR, the purified protein needs to isotopically enriched, which can be technically challenging and expensive. Furthermore, this technique is generally limited to small, soluble, proteins (~ 40 kDa) and single domains (reviewed in (Sugiki et al., 2017)). What these 'classical' structural determination techniques have in common is that their limitations impose difficulties in solving structures of large proteins and RNA, multimeric protein complexes and integral membrane proteins.

Single particle analysis (SPA) electron microscopy (EM) is an alternative structural technique that can be applied to solving 3D structures of challenging protein complexes whilst only needing a fraction of protein amounts required for MX or NMR. In EM, biological samples are directly imaged using transmission electron microscopy (TEM) which generates a 2-dimensional (2D) projection of the sample in the direction of the electron beam, similar to a shadow cast by an object in a beam of light. A 3D reconstruction of the object is obtained, *in silico*, by combining these 2D projections of the object in different orientations, or views. EM as a technique for protein structural determination, generally for proteins > 200 kDa, is further discussed in this section, but can also be applied to nucleic acids.

1.5.1 Sample preparation

Biological samples, such as proteins, need to be purified to homogeneity because several thousands of protein molecules, or particles, will be imaged and computationally averaged together. Sample heterogeneity can stem from inherent protein flexibility, conformational variations, or compositional differences where there is an absence or presence of subunits or binding partners. In principle sample heterogeneity can be resolved computationally as shown by Lauber and colleagues (Lauber et al., 2018). The authors were able to solve two unique protein Type 9 secretion complexes (SprA + PPI + PorV and SprA + PPI + plug) from a single dataset (Lauber et al., 2018) however this is not common. For challenging samples, increasing the homogeneity aids structure determination. There are three main obstacles that need to be overcome when imaging using TEM. First, protein samples need to withstand the high vacuum within the microscope. Next, to minimise the visible radiation damage to protein samples caused by the electron beam, which can increase sample heterogeneity. And finally, to overcome the low contrast caused by the protein sample scattering electrons similarly to its water/buffer background during imaging.

There are two grid preparation methods for imaging using TEM: proteins can be stained using a low concentration of heavy metal salts and dried before being inserted in the microscope, or imaged in a frozen, hydrated state after vitrification. The first method is called negative staining whereby a heavy metal salt solution, typically 1-2 % uranyl acetate, forms a cast around the protein (Brenner and Horne, 1959). Negative stain for 3D reconstruction was first used in 1968 by De Rosier and Klug to report the helical reconstruction of the T4 bacteriophage (De Rosier and Klug, 1968). Negative stain overcomes the drawback of radiation damage because it is the cast formed by stain that is imaged and not the protein. Additionally, negative stain micrographs generate high-contrast images because of the heavy metal in the staining solution. However, the stain may not be homogenous across the grid and not penetrate into small crevices of the

protein. Furthermore, the drying of the protein sample may result in the collapse of the structure. But for actual 3D structure determination, negative stain EM structure reconstructions are limited in resolution to the grain size of the dried heavy metal (to 15-20 Å). Negative stain EM is, however, a quick way to verify sample homogeneity before proceeding to vitrification (Passmore and Russo, 2016). The second grid preparation method by vitrification (Adrian et al., 1984, Dubochet et al., 1988) is commonly used in cryogenic (cryo) EM for high-resolution structural determination. Here sample, in the buffer it was purified in, is applied onto a grid, blotted, and then plunge frozen in liquid ethane. The rapid freezing rate caused by liquid ethane (approximately 10^6 °C/s) freezes water in a vitreous state (Dubochet et al., 1988). Proteins are preserved in a hydrated, native-like state to be imaged.

1.5.2 Data collection and processing

After cryo-EM grids have been prepared, diagnostic cryo-EM, or screening, is performed to maximize grid quality. First, the grid is evaluated for amorphous ice and ice thickness is to ensure uniform thickness across the grid. Then particle density and overall heterogeneity is assessed to ensure particles are abundant and stable. The contrast between protein particles and vitreous ice is low because both protein and water are made up of light elements and therefore scatter electrons similarly (Nogales, 2016, Nogales and Scheres, 2015). To maximize signal-to-noise, ice needs to thin enough so that it supports embedded protein but does not exclude protein. Ice that is too thin is apparent during screening because there is a circular region of ice in the centre of the hole where particles are excluded (Passmore and Russo, 2016). During this stage, grids may require further optimisation to ensure optimal particle density and ice thickness. Once protein concentration, ice thickness and quality have been optimised an initial data set should be collected. The aim of this initial data collection is to align enough particles together to generate 2D class averages that contain features of secondary structure and generate initial 3D reconstruction classes. However at this stage if promising 2D classes are not obtained then

protein sample needs to be re-evaluated by changing purification, buffers and detergents, grid preparation. If the initial 3D reconstruction reaches a reasonable resolution then more data should be collected, ideally on a high-end TEM (e.g. Titan Krios), to generate a high resolution structure. 2D classification of particles is the clustering of similar particles, similar in structure and orientation, followed by calculating the average of the particles collected within each class average (Sigworth, 2016). Secondary structures, such as α-helices, are visible in the 2D class averages from high resolution data. The process starts with a set of random reference images, and then the probabilities are calculated *in silico* of each extracted particle image with respect to its rotation, translation and the degree of matching to each reference. The rotated and translated images are combined to form the overall class average images (Sigworth, 2016). The generated averages then become the new set of references and the entire procedure is iterated several times to produce representative averages, or 2D class averages, of the collected dataset. After averages are generated, particles belonging to the highest resolution class averages are selected for further data processing for 3D initial model reconstruction and 3D classification. 3D initial model reconstruction involves back projecting selected 2D class averages or particles to produce a low resolution model that is then used as a reference. The critical step of the 3D initial model is working out the relative orientations of the class averages/particles with respect to each other to allow their combination to generate a 3D volume. Similar to 2D classification, 3D classification involves classifying large differences in the structure of the collected dataset which further aids in identifying sample heterogeneity (Scheres et al., 2007, Serna, 2019). After several refinement iterations, the assignment of each particle to each 3D class is normally straightforward, but can be computationally demanding, and additional reconstructions can be generated from the separately assigned particle stacks (Sigworth, 2016). It is routine to use repeated rounds of 3D classification with increasingly more exhaustive angular searches, which may reveal smaller structural differences between the classes to generate a more homogenous particle dataset for the final high-resolution reconstruction (Scheres, 2016).

In the past decade, SPA cryo-EM has undergone, and still is in, a 'resolution revolution' (Kühlbrandt, 2014, Callaway, 2015, Egelman, 2016, Mitra, 2019). Since the first high-resolution structure of bacteriorhodopsin at 3.5 Å reported in 1990 by Henderson and colleagues (Henderson et al., 1990) an exponential increase in high-resolution structures (< 4 Å) have been deposited in the Electron Microscopy Data Bank (EMDB) (EMBL-EBI, 2020b) within the past 5 years. Now near atomic resolution can be achieved from SPA cryo-EM as 1.25 Å resolution of apoferritin has been reported (Yip et al., 2020).

1.6 Aims of thesis

iRhoms are regarded as the regulatory subunit of ADAM17 and hence iRhoms have an indirect, yet essential, role in ADAM17-dependent shedding. At the start of this project, little to no structural work had been performed on the iRhom/ADAM17 complex. Therefore, the aim of this project was to express and purify iRhoms alone and in complex with ADAM17, for structural determination, to help further understand and underpin the molecular mechanisms by which these two proteins interact (Figure 1.9). Understanding the interaction between iRhom and ADAM17 could be utilised as a tool as an ADAM17 inhibitor.

Chapter 3 will illustrate how the project initially started by attempting to characterise the IRHD of human iRhom1 with the aim of crystallisation (Figure 1.9A, dashed circle). However the IRHD purification lead to poor yields and it became apparent that the IRHD may be more stable as part of the full length protein. In Chapter 4, I report generating the cell lines that over express mammalian and non-mammalian iRhoms with their corresponding ADAM17, as well as their expression and sub-cellular localisation, for their eventual expansion and purification. In addition to the physiological structural target of the native iRhom/ADAM17 complex (Figure 1.9B) I also generated a cell line expressing the stable human FRMD8iRhom2 chimera (Figure 1.9A). The

expansion and purification of these cell lines is reported in Chapter 5 where I first screened to optimise solubilising conditions for protein extraction out of the membrane and for the presence of the intact iRhom/ADAM17 complex. I finish in Chapter 6 with preliminary SPA cyro-EM data of the human FRMD8iRhom2 chimera and my conclusions of how the results reported in this thesis should suggest further work.

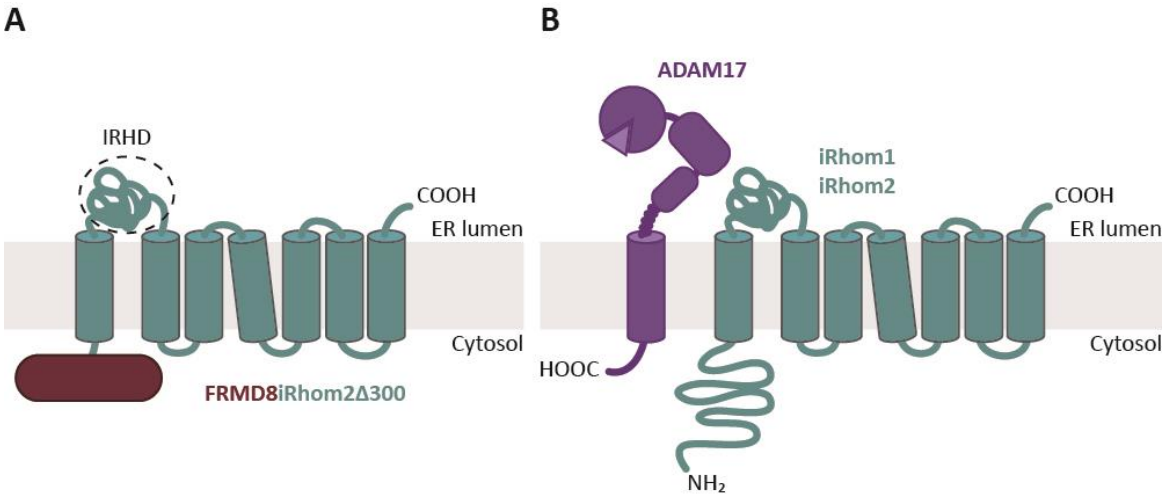


Figure 1.9. Structural targets for this thesis. The overview of the structural targets aimed in this thesis: the iRhom homology domain (IRHD) of human iRhom1 (dotted line), the full length stable target **(A)** of the human FRMDiRhom2Δ300 and physiological target **(B)** of the iRhom/ADAM17 complex.

Chapter 2: Material and Methods

2.1 Materials

All chemicals were purchased from Sigma-Aldrich unless specified and all detergents were purchased from Anatrace. Enzymes were either purchased from New England BioLabs (NEB) or Thermo Fisher Scientific.

2.2 Molecular Biology

2.2.1 Genes and plasmids

Genes listed in Table 2.1 were used in this thesis. Genes purchased from GeneArt (Thermo Fisher Scientific) were codon optimised for their intended expression system.

Table 2.1. Genes used in this thesis. (*) Denotes iRhom homology domain of human iRhom1

Gene	Protein	Species	Expression system	GenBank/ Ensembl ID	Source
RHBDF1	*Hs*IRHD1*	*Homo sapiens*	*E. coli*	-	GeneArt
RHBDF1	*Hs*iRhom1	*Homo sapiens*	Human	-	Freeman Group
RHBDF2	*Hs*iRhom2	*Homo sapiens*	Human	-	Freeman Group
ADAM17	*Hs*ADAM17	*Homo sapiens*	Human	-	Addgene (Table 2.2)
RHBDF2	*Mm*iRhom2	*Mus musculus*	Human	-	Freeman Group
ADAM17	*Mm*ADAM17	*Mus musculus*	Human	-	Freeman Group
RHBDF	*Hd*iRhom	*Hypsibius dujardini*	Human	MTYJ01000065.1	GeneArt

The following plasmids listed in Table 2.2 were used in this thesis.

Table 2.2. Plasmids used in this thesis.

Plasmid name	Antibiotic resistance	Source	Additional information
pExpreS2-2	Kanamycin/G418	ExpreS^{2}ion Biotechnologies	
pETM14	Kanamycin	Lea Group	
pHR-CMV-TetO2_3C-Twin-Strep	Ampicillin	Addgene	Addgene plasmid #113883
pHR-CMV-TetO2_3C-mVenus-Twin-Strep	Ampicillin	Addgene	Addgene plasmid #113891
pHR-CMV-TetO2_3C-Avi-His6	Ampicillin	Addgene	Addgene plasmid #113887
pHR-CMV-TetO2_3C-Twin-Strep_IRES-EmGFP	Ampicillin	Addgene	Addgene plasmid #113884
psPAX2	Ampicillin	Addgene	Addgene plasmid #12260
pMD2.G	Ampicillin	Addgene	Addgene plasmid #12259
pcDNA3-ADAM17-HA	Ampicillin	Addgene	Addgene plasmid #65105

2.2.2 DNA oligonucleotides

Primers were designed using the PCR or NEBuilder® HiFI DNA Assembly simulations in SnapGene 4.3.11. Primers were purchased from Invitrogen and reconstituted in dH_2O to a final concentration of 100 µM. The primers for cloning in this study are listed in Appendix A.

2.2.3 DNA amplification

DNA inserts were amplified by polymerase chain reaction (PCR) using Q5® Hot Start High-Fidelity 2X Master Mix (NEB) following the manufacturer's protocol for reaction assembly and thermocycling. Plasmid linearization for cloning was also performed by PCR using either PCRBIO VeriFi™ Polymerase (PCR Biosystems) or Platinum™ SuperFi™ PCR Master Mix (Invitrogen) using the manufacturer's protocol for reaction assembly, with the addition of GC enhancers or buffers, and thermocyling. Thermocycling for PCRs were performed using a T100™ Thermal Cycler (Bio-Rad). PCR products, whereby DNA templates were plasmid DNA, was treated with FastDigest

DpnI (Thermo Fisher Scientific) following the manufacturer's recommendations to cut methylated DNA.

2.2.4 Amplicon visualisation and purification

DNA amplicons were prepared with Gel Loading Dye, Purple (6X) (NEB) and loaded onto 1 % (w/v) agarose gels, as outlined in Table 2.3, and electrophoresed in TAE buffer (Table 2.3) at 100 V for 20-30 minutes. DNA samples were run alongside a GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific) and imaged using a Molecular Imager® ChemiDoc™ XRS+ Imaging System (Bio-Rad). DNA gels were visualised on a Safe Imager™ 2.0 Blue Light Transilluminator (Invitrogen) and appropriate bands excised using FastGene Agarose Gel Band Cutter (Geneflow). Excised DNA bands or PCR products were purified using either a QIAquick® Gel Extraction Kit (Qiagen) or a QIAquick® PCR Purification Kit (Qiagen), respectively, following the manufacturer's protocol. Purified DNA concentration was measured using a NanoDrop Spectrophotometer (Thermo Fisher Scientific).

Table 2.3. Solutions for DNA visualisation.

Solution	Component
50X TAE buffer	242 g Tris base 57.1 mL glacial acetic acid 100 mL 500 mM EDTA dH_2O to 1 L
1 % (w/v) agarose gel	1 % (w/v) of agarose in 1X TAE buffer 1:10,000 dilution of SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, Invitrogen)

2.2.5 Cloning

Constructs were generated by the assembly of PCR products containing overlapping regions, of 20-30 nucleotides at both 5' and 3' ends, matching the insert regions of their designated plasmid. The assembly was performed using the NEBuilder® HiFi DNA Assembly Master Mix

(NEB) with the reaction composition described in Table 2.4 and incubated at 50 °C for 1 hour using a thermocycler. All of the assembled products were transformed into chemically competent *E. coli* DH5α cells following the protocol for bacterial cell transformation by heat shock in Section 2.2.6. Transformed cells were plated out onto lysogeny broth (LB) agar (per 1 L: 10 g typtone, 5 g yeast extract, 10 g NaCl, 20 g agar) supplemented with the appropriate concentration of antibiotic, depending on the antibiotic resistance of the plasmid. Plates were incubated overnight (between 14 to 16 hours) at 37 °C.

Table 2.4. Protocol for Gibson Assembly. Gibson Assembly was performed using the NEBuilder® HiFi DNA Assembly Master Mix.

Component	Volume (µL)
NEBuilder Master Mix	5
Plasmid DNA	1
Insert DNA	3
dH_2O	1

2.2.6 Bacterial cell transformation by heat shock

Assembled DNA products or 100 ng of purified plasmid was added to 50 µL of chemically competent *E. coli* cells in a sterile 1.5 mL microfuge tube and incubated on ice for 30 minutes. Cells were heat shocked in a water bath at 42 °C for 30 seconds then incubated on ice for 5 minutes. 900 µl of 2x YT medium (per 1 L: 16 g typtone, 10 g yeast extract, 5 g NaCl, adjusted to pH 7.0 with 5M NaOH) was added to the cells and allowed to recover at 37 °C with shaking for 30-45 minutes. Cells that were transformed with HiFi assembled products were first centrifuged at 5,000 rpm using a benchtop microfuge for 1 minute, supernatant discarded, cell pellet resuspended in 100 µl of 2x YT and the entire cell resuspension plated out onto LB agar plates supplemented with working concentration of appropriate antibiotic. For cells that were transformed with purified plasmids, 50-100 µl of cells were plated out onto LB agar plates

supplemented with the appropriate antibiotic. Plates were incubated overnight (between 14 to 16 hours) at 37 °C.

2.2.7 Colony PCR

Transformants were subject to colony PCR to screen for gene insertion into the plasmid. Isolated colonies were picked using a sterile pipette tip, streaked onto a referenced LB agar plate, with antibiotic and then inoculated into a tube containing colony PCR mix. The streaked colonies were incubated at 37 °C for 6-8 hours. The colony PCR reagent used was the OneTaq® Quick-Load® 2X Master Mix with Standard Buffer (NEB) following the manufacturer's protocol for reaction preparation, but adapted for a 10 µL reaction, and thermocycling conditions. The HiFi assembly primers used for DNA insert amplification for cloning were used as primers for the colony PCR. PCR products were immediately loaded onto 1 % (w/v) agarose gels as described in Section 2.2.4 and imaged. Colonies that produced positive result were picked from the reference LB agar pate to inoculate LB media, with antibiotic, and grown overnight (14-16 hours) at 37 °C with shaking for plasmid amplification and purification.

2.2.8 Plasmid amplification and purification

For plasmid amplification and purification, a single isolated bacterial colony was used to inoculate LB media supplemented with working concentration of appropriate antibiotic, and incubated overnight (14-16 hours) at 37 °C with shaking. Plasmids were extracted and purified from overnight bacterial cultures using the QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's protocol. For large scale plasmid amplification for concentrated and large volumes of plasmid DNA, plasmids were extracted and purified using the HiSpeed® Plasmid Maxi Kit (Qiagen) according to the manufacturer's protocol.

2.2.9 DNA sequencing

All DNA sequencing was outsourced to Source BioScience Sanger Sequencing Services following their recommendations for sample preparation and requirements. Primers used for sequencing were either provided by the company or sent with the DNA samples.

2.3 Cell culture

2.3.1 Bacterial cell culture

Bacterial cultures were grown in Miller's LB (per 1 L: 10 g typtone, 5 g yeast extract, 10 g NaCl) medium and autoclaved. LB was supplemented with 1.5 % (w/v) agar to prepare solid LB agar pates, after autoclaving and cooling before solidification, the solution was supplemented with antibiotic at working concentration. Working concentrations of ampicillin and kanamycin were 100 µg/mL and 50 µg/mL respectively. Antibiotics were prepared as 1,000-fold concentrated stocks in dH_2O, sterile-filtered, aliquoted and stored at -20 °C until use. *E. coli* cells were grown in LB media or LB agar at 37 °C, unless stated otherwise. Bacterial overnight cultures, for plasmid purification or protein expression, were prepared by inoculating a single isolated colony in 10 mL of Miller's LB media supplemented with antibiotic at working concentration. Cultures were incubated overnight in an orbital shaker at 37 °C. Bacterial culture growth was monitored by measuring the optical density at 600 nm (OD_{600}). Bacterial glycerol stocks were prepared by mixing 580 µL of overnight culture with 420 µL of sterile 60 % (v/v) glycerol in 1.8 mL Cryogenic Vial (Star Lab). Bacterial glycerol stocks were stored at -80 °C and when needed were kept on dry ice and streaked onto LB agar plates using a metal inoculating loop.

2.3.2 Eukaryotic cell culture

All media used for eukaryotic cell maintenance was pre-warmed to room temperature for insect cells or to 37 °C for mammalian cells. Cells were regularly checked for *Mycoplasma* contamination.

2.3.2.1 Schneider 2 cell culture and transfection

For insect cell protein expression, *Drosophila melanogaster* Schneider 2 (S2) suspension cells (ExpreS^{2}ion Biotechnologies) were grown in EX-CELL® 420 Serum-Free media (Sigma-Aldrich) in an orbital shaker at 100 rpm at 25 °C. Cells were passaged to cell density of 8 x 10^6 cells/mL in serum-free media or to 5 x 10^6 cells/mL in EX-CELL® 420 medium supplemented with 10 % (v/v) fetal bovine serum (FBS) (Thermo Fisher Scientific, Gibco). Transfections to generate stable cell lines were performed as per manufacturer's protocol (ExpreS^{2}ion Biotechnologies). Cells were frozen in ice-cold 40 % (v/v) EX-CELL 420 medium, 50 % (v/v) FBS, 10 % (v/v) cell-culture grade dimethyl sulfoxide (DMSO) (Sigma-Aldrich), in 1.8 mL Cryogenic Vials, using Corning® CoolCell™ LX Cell Freezing Container (Corning) then transferred to -80 °C. The cell freezing medium was sterile-filtered. After cells were frozen, vials were transferred to liquid nitrogen long term storage.

For cell expansion, cells were passaged to cell density of 8 x 10^6 cells/mL, in serum-free media, to a volume desired for purification.

2.3.2.2 Adherent HEK cell culture

Human embryotic kidney 293T (HEK293T), HEK293T Lenti-X (Takara/Clontech), FreeStyle™ 293-F (HEK293F) (Thermo Fisher Scientific, Gibco) and Expi293F™ GnTI- (Thermo Fisher Scientific, Gibco) cells were maintained in adherent cell culture. Cells were grown in Dulbecco's Modified Eagle Medium:Nutrient Mixture F-12 (DMEM/F-12) supplemented with 10 % (v/v) FBS (Thermo

Fisher Scientific, Gibco) and 1X of Minimum Essential Medium Non-Essential Amino Acids (MEM NEAA) (Thermo Fisher Scientific, Gibco). Cells that were originally in suspension culture (HEK293F and Expi293F™ GnTI-) were transferred to adherent culture by centrifuging 1 x 10^6 cells at 500 *x g* for 3 minutes, resuspending the cell pellet in 12 mL DMEM/F-12/10 % FBS/MEM NEAA and transferring to a T75 flask. Cells were maintained in vented, cell culture treated flasks and incubated at 37 °C and 5 % CO_2. Cells were passaged by trypsin detachment. Monolayer of cells in a T75 flask, was washed in 10 mL of Dulbecco's Phosphate Buffered Saline (PBS) (Sigma-Aldrich) then incubated with 3 mL of TrypLE Express Enzyme (Thermo Fisher Scientific, Gibco) diluted 1:5 in PBS. Detached cells were resuspended in 9 mL of fresh media and 1 mL of resuspended cells was used to sub-culture 11 mL of media in a fresh flask. Cells were passaged at least 70 % confluency. Cells were frozen in ice-cold 90 % FBS, 10 % (v/v) cell-culture grade DMSO, in 1.8 mL Cyrogenic Vials, using Corning® CoolCell™ LX Cell Freezing Container (Corning) then transferred to -80 °C. The cell freezing medium was sterile-filtered. After cells were frozen, vials were transferred to liquid nitrogen for long term storage.

For frozen cells frozen in 90 % FBS and 10 % DMSO, vials were incubated in a 37 °C water bath for 1-1.5 minutes, until approximately 90 % of the cell suspension thawed, then transferred to a tube containing 5 mL with pre-warmed DMEM/F-12/10 % FBS/MEM NEAA. Cells were centrifuged (500 *x g*, 3 minutes), cell pellet resuspended in 12 mL pre-warmed DMEM/F-12/10 % FBS/MEM NEAA then transferred to a T75 flask.

For cell expansion, cells were seeded in Cell Culture Roller Bottle, 2125 cm^2 (Greiner Bio-One) in DMEM/F-12/10 % FBS/MEM NEAA, gassed and incubated at 37 °C on a roller rack for 3-4 days. Cells were harvested by detachment using PBS supplemented with 5 mM ethylenediaminetetraacetic acid (EDTA). Conditioned medium was discarded and replaced with a volume of PBS, 5 mM EDTA enough to cover the surface of the bottle laid horizontally. Bottles

were incubated on the roller rack at 37 °C for 5 minutes, cells pelleted by centrifugation (500 *x g*, 30 minutes) and stored at -20 °C until use.

2.3.2.3 Suspension HEK cell culture

HEK293F and Expi293F™ GnTI- cells were grown in FreeStyle™ 293 Expression Medium (Thermo Fisher Scientific, Gibco) and Expi293™ Expression Medium (Thermo Fisher Scientific, Gibco), respectively. Cells were maintained in Polycarbonate Erlenmeyer Flask, Flat Cap (Corning), gassed with 5 % CO_2 Balance Air (BOC) and incubated at 37 °C on platform shaker at 135 rpm. HEK293F and Expi293F™ GnTI- cells were passaged to a cell density of 0.1 x 10^6 cell/mL and 0.2 x 10^6 cell/mL, respectively. Cells were frozen following the manufacturer's recommendations. The freezing medium was 45 % (v/v) condition medium, 45 % (v/v) cell line appropriate cell medium, 10 % (v/v) cell-culture grade DMSO and sterile-filtered. Controlled cell freezing was achieved using the Corning® CoolCell™ LX Cell Freezing Container transferred to -80 °C. Frozen vials were transferred to liquid nitrogen for long term storage.

For cells frozen in suspension freezing medium, vials were incubated in a 37 °C water bath for 1-1.5 minutes, until approximately 90 % of the cell suspension thawed, immediately transferred to a 125 mL Polycarbonate Erlenmeyer Flask, Flat Cap (Corning) containing 30 mL of cell line appropriate medium then gassed and incubated as described above. Cell viability was monitored by Trypan Blue (Sigma-Aldrich) exclusion 24 hours after thawing.

For cell expansion of HEK293F cell lines, cells were seeded at a final density of 0.1 x 10^6 cells/mL and incubated at 37 °C with shaking or fast rolling for 4-5 days. Cells were harvested by centrifugation (500 *x g*, 30 minutes) and stored at -20 °C until use.

2.3.2.4 DNA transfection and lentiviral transduction

Lentiviral particles were generated using second-generation envelope plasmid, pMD2.G, second-generation packaging plasmid, psPAX2, and the pHR-CMV-TetO2 transfer plasmids (Table 2.2). Following an adapted form of the protocol from Elegheert et al. (Elegheert et al., 2018). Method of stable cell line generation, transfection and transduction, and deviations from the protocol is described in Chapter 4.

After cells were passaged for the first time post-transduction, some cells were frozen as described in Section 2.3.2.2.

2.4 Biochemistry

2.4.1 Standard procedure for purifying recombinant soluble proteins

2.4.1.1 Purifying from Schneider 2 cell media

For insect cell protein expression, recombinant proteins were expected to be secreted into the media due to incorporation of the signal peptide. S2 media was harvested by centrifugation (1000 *x g*, 30 minutes, 4 °C) and filtered through 0.22 µm filtering unit (Millipore). The media was incubated with cOmplete™ His-Tag Purification free resin (Sigma-Aldrich), pre-equilibrated in PBS, with gentle rolling. The media/resin was applied to a gravity column for the media to flow through. The resin was washed in PBS, 5 mM imidazole, pH 7.5 and bound proteins eluted in PBS, 250 mM imidazole, pH 7.5. The His-tag eluate was dialysed against 100 mM Tris HCl, 150 mM NaCl, pH 8.0 and then applied to an equilibrated (100 mM Tris HCl, 150 mM NaCl, pH 8.0) Strep-Tactin®XT Superflow® (IBA Lifesciences) column. The column was washed and eluted as per manufacturer's instructions. Elution fractions containing protein were pooled, concentrated and applied to a S200 10/300 column. Details of purifications are described in the Chapter 3.

2.4.1.2 Purifying from Schneider 2 cells

S2 cells were harvested by centrifugation described above in Section 2.4.1.1. The cell pellet was washed in PBS then frozen at -20 °C until use. Cell pellet was thawed and resuspended in PBS with SIGMAFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free (Sigma-Aldrich) and Benzonase® Nuclease (Sigma-Aldrich). Cells were lysed using an EmulsiFlex-C5 homogeniser (Avestin) at approximately 10,000 psi by passing the cell lysate through the EmulsiFlex at least 5 times. Lysates were centrifuged at 18,000 x *g* for 30 minutes at 4 °C using a JA25.50 rotor. Supernatant was applied to a pre-equilibrated Ni-NTA column (QIAGEN) following the manufacturer's instructions. Eluate was applied to Strep-Tactin®XT Superflow® (IBA Lifesciences) column. The column was washed and eluted as per manufacturer's instructions. Elution fractions containing protein was pooled, concentrated and applied to a S200 10/300 column.

2.4.1.3 Purifying from bacterial cells

10 mL of overnight culture was used to inoculate 1 L of LB media supplemented with the appropriate antibiotic. Cells were incubated at 37 °C with shaking at 180 rpm until the culture reached an OD_{600} of 0.6-0.8. At this point, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cell culture to induce protein expression. Cells were incubated at 21 °C with 180 rpm shaking overnight (14-16 hours). Cells were harvested by centrifugation (JLA8.100 rotor, 4500 rpm, 20 minutes, 4 °C). At this stage, cell pellets were stored at -20 °C until use. Thawed cell pellets were resuspended in 40 mL (per 1 L of culture) of PBS supplemented with 50 µg/mL Bovine Pancreas Deoxyribonuclease I (DNase I) (Sigma-Aldrich) and 1 mg/mL Lysozyme from chicken egg white (Sigma-Aldrich). Resuspension of the cell pellet was aided by a hand-held Dounce homogeniser. Resuspended cells were lysed by EmulsiFlex-C5 homogeniser (Avestin) at 15,000 psi by passing the cell lysate through the EmulsiFlex at least 5 times. Soluble and insoluble fractions were separated by centrifugation using the JA25.50 rotor (Beckman) at 20,000 rpm for 30 minutes at 4 °C. Supernatant was used for purification.

2.4.2 Standard procedure for refolding and purifying from bacterial inclusion bodies

For expression and purification from inclusion bodies, *E. coli* BL21 cells were induced as described in Section 2.4.1.3. After induction, cells were incubated at 37 °C with shaking at 180 rpm for 4 hours. Cells were harvest by centrifugation (JLA8.100 rotor, 4500 rpm, 20 minutes, 4 °C) and stored at -20 °C until use. Cell pellet, derived from 1 L of culture, was resuspended in 40 mL of homogenisation buffer (50 mM Tris, 1 mM EDTA, 150 mM NaCl, pH 8.0) with DNase I and lysozyme to a final concentration of 50 µg/mL and 1 mg/mL respectively. Cells were lysed and centrifuged as described in Section 2.4.1.3 but with the addition of Tween 20 to the cell lysate at a final concentration of 0.05 % (v/v). The pellet was resuspended in homogenisation buffer using a Dounce homogeniser and then Tween 20 was added to a final concentration of 0.05 % (v/v). Resuspended pellet was slowly rotated on a rotary wheel at 4 °C for 1 hour. Then centrifuged using a JA25.50 rotor at 20,000 rpm for 20 minutes at 4 °C. The pellet from this spin was resupended in homogenisation buffer, incubated with Tween 20 and centrifuged as described previously once more. The pellet was solubilised in 40 mL of 8 M urea, 1 mM EDTA, 100 mM Tris, 25 mM dithiothreitol (DTT), pH 8.0 and incubated on a rotary wheel, with gentle spinning, for 2 hours at 4 °C. The solubilised pellet was acidified with 37 % HCl until the pH was 3-4 then centrifuged using a JA25.50 rotor at 20,000 rpm for 20 minutes at 4 °C. 40 mL of dialysis buffer (6 M urea, pH 3-4) was added to the supernatant and diluted supernatant was dialysed against 1 L of 6 M urea with three more 1 L buffer exchanges. Dialysed sample was centrifuged using a JA25.50 rotor at 20,000 rpm for 40 minutes at 4 °C, the supernatant filtered using a 0.22 µM syringe end filter (Millipore) and added dropwise to 2 L of refold buffer, using a peristaltic pump, overnight. The refold buffer consisted of 500 mM arginine, 1 mM cysteine, 2 mM cystine, 20 mM ethanolamine, 1 mM EDTA, pH 11.0, filtered and degassed. Refolded protein in refold buffer was concentrated to approximately 30 mL using Vivaflow 200 10,000 MWCO (Sartorius) and 1 M Tris HCl, pH 7.0 was added to a final concentration of 50 mM. The sample was further centrifuged using a JA25.50 rotor at 20,000 rpm for 20 minutes at 4 °C followed by dialysis of the

supernatant against 1 L 50 mM Tris HCl, 150 mM NaCl, pH 7.5 with two 1 L buffer exchanges. Dialysed sample was centrifuged (20,000 rpm, 40 minutes, 4 °C), supernatant filtered using a syringe end filter, concentrated down to 5 mL using a 10,000 MWCO spin concentrator (Millipore) and applied the concentrate to Superdex 75 16/600 column (GE Healthcare) on an ÄKTA Pure (GE Healthcare) system, using 50 mM Tris HCl, 150 mM NaCl, pH 7.5 as the running buffer.

2.4.3 Standard procedure for purifying recombinant membrane proteins

HEK cells expressing recombinant membrane proteins were harvest by centrifugation (500 *x g*, 30 minutes, 4 °C) as described in Section 2.3.2.3. Cells were resuspended in chilled PBS, pH 7.5 supplemented SIGMAFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free (Sigma-Aldrich) in a volume 2.5 times of the mass of the cell pellet. Benzonase (Sigma-Aldrich) was added to the cell suspension to a final concentration of 0.02 % (v/v). Cell suspension was lysed by EmulsiFlex-C5 homogeniser at 10,000-15,000 psi passing the cell lysate through the EmulsiFlex at least 5 times. Insoluble and unlysed material was pelleted by centrifugation (JA25.50 rotor, 25,000 *x g*, 30 minutes, 4 °C) and the supernatant was subjected to an ultracentrifugation spin at 210,000 *x g* for 1 hour and 30 minutes at 4 °C. The membrane pellet was resupended in buffer, containing 100 mM Tris HCl, 150 mM NaCl, pH 8.0, supplemented with SIGMAFAST™ EDTA-Free Protease Inhibitor Cocktail Tablet, in a volume 2.5 times of the mass of the membrane pellet using a Dounce homogeniser. Resuspended membranes were aliquoted in 5 mL aliquots, flash frozen in liquid nitrogen and stored at -80 °C until use. Details of how the resuspended membranes were used and purified are outlined in Chapter 5. Total protein concentration of resuspended membranes was estimated by serially diluting the resuspended membranes and using the Pierce™ Bicinchoninic acid (BCA) Protein Assay (Thermo Fisher Scientific) following the manufacturer's recommendations.

2.4.4 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

Protein samples were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were mixed with 5X SDS Loading Dye to final concentration of 1X. Samples from soluble protein purifications were heated at 95 °C for 10 minutes. Samples from membrane protein purifications were not heat treated. Samples were loaded onto either prepared 10 or 15 % Tris/Glycine gels or precast Novex™ WedgeWell™ 8-16 %, Tris-Glycine (Thermo Fisher Scientific, Invitrogen). Gels were run in SDS Tris/Glycine running buffer at 200 V until completion. Prepared gels and precast gels were ran in the Mini-PROTEAN Tetra Cell (Bio-Rad) and XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher Scientific, Invitrogen). Gels were stained in InstantBlue (Expedeon) and gel band molecular masses were estimated by PageRuler™ Plus Prestained Protein Ladder (Thermo Fisher Scientific). All buffers and solutions used for SDS-PAGE are shown in Table 2.5. Gels were imaged using Molecular Imager® ChemiDoc™ XRS+ Imaging System (Bio-Rad).

Table 2.5. Solutions and buffers for SDS-PAGE analysis.

Solution	Component	Concentration
5X SDS loading Dye	Tris HCl (pH 6.8)	250 mM
	SDS	10 % (w/v)
	Glycerol	50 % (v/v)
	Bromophenol blue	0.3 % (v/v)
	Dithiothreitol	100 mM

SDS-PAGE gel	*15 % resolving gel:*	
	Acrylamide Bis-Acrylamide (19:1)	15 % (w/v)
	1.5 M Tris HCl, pH 8.8	375 mM
	10 % SDS	0.1 % (w/v)
	Ammonium persulfate (APS)	0.1 % (w/v)
	N,N,N′,N′-Tetramethylethylenediamine (TEMED)	0.01 % (v/v)
	10 % resolving gel:	15 % (w/v)
	Acrylamide Bis-Acrylamide (19:1)	375 mM
	1.5 M Tris HCl, pH 8.8	0.1 % (w/v)
	10 % SDS	0.1 % (w/v)
	APS	0.01 % (v/v)
	TEMED	
	Stacking gel:	5 %
	Acrylamide Bis-Acrylamide (19:1)	125 mM
	1 M Tris HCl, pH 6.8	0.1 % (w/v)
	10 % SDS	0.1 % (w/v)
	APS	0.01 % (v/v)
	TEMED	
10X Tris/Glycine running buffer	Tris base	0.25 M
	Glycine	1.92 M
	SDS	0.1 % (w/v)

2.4.5 Western blotting

After proteins were separated by SDS-PAGE, proteins were transferred onto methanol activated Amersham Hybond P 0.45 µM Blotting Membrane (GE Healthcare) using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) at 22 V for 1 hour. The transfer buffer used was 48 mM Tris, 39 mM Glycine, 0.0375 % (w/v) SDS, 20 % (v/v) methanol. After protein transfer, membranes were blocked for 1 hour at room temperature or overnight at 4 °C in either 5 % (w/v) of skimmed milk in PBS-T (PBS, 0.1 % (v/v) Tween 20) or 3 % (w/v) Bovine Serum Albumin (BSA) in 10 mM Tris HCl, 150 mM NaCl, pH 7.5, depending on the primary antibody used. Membranes were incubated with primary antibody at room temperature or overnight at 4 °C. Membranes were washed three times for 15 minutes with PBS-T, and then incubated with HRP-conjugated secondary antibody for 1 hour at room temperature. For antibodies used refer to Table 2.6. Membranes were washed three times for 15 minutes with PBS-T and then incubated with ECL Western

Blotting Substrate (Promega), following the manufacturer's recommendations. Membranes were developed, in a dark room, on Amersham Hyperfilm™ ECL (GE Healthcare) using a SRX-101A Medical Film Processor (Konica).

Table 2.6. Antibodies used for western blots in this thesis and protocol used.

Probing	Blocking buffer	Primary antibody dilution and buffer	First wash and buffer	Secondary Antibody and buffer	Second wash and buffer
Twin Strep Tag	5 % (w/v) skimmed milk, PBS, 0.1 % (v/v) Tween 20	StrepMAB Classic (IBA Lifesciences) 1:2,000 in milk blocking buffer	PBS, 0.1 % (v/v) Tween 20 (3x 15 minute washes)	α-Mouse IgG HRP (H+L) Conjugate (Promega) 1:10,000 in milk blocking buffer	PBS, 0.1 % (v/v) Tween 20 (3x 15 minute washes)
Hexahistine Tag	3 % (w/v) BSA in 10 mM Tris HCl, 150 mM NaCl, pH 7.5	Tetra·His Antibody (Qiagen) 1:1,000 in BSA blocking buffer	10 mM Tris HCl, 500 mM NaCl, 0.05 % (v/v) Tween 20, 0.2 % Triton X-100, pH 7.5 (2x 15 minute washes) 10 mM Tris HCl, 150 mM NaCl, pH 7.5 (1x 15 minute wash)	α-Mouse IgG HRP (H+L) Conjugate (Promega) 1:5,000 in 10 % (w/v) skimmed milk, 10 mM Tris HCl, 150 mM NaCl, pH 7.5	10 mM Tris HCl, 500 mM NaCl, 0.05 % (v/v) Tween 20, 0.2 % Triton X-100, pH 7.5 (2x 15 minute washes) 10 mM Tris HCl, 150 mM NaCl, pH 7.5 (1x 15 minute wash)
HA Tag	5 % (w/v) skimmed milk, PBS, 0.1 % (v/v) Tween 20	HA mouse monoclonal (Enzo) in milk blocking buffer	PBS, 0.1 % (v/v) Tween 20 (3x 15 minute washes)	α-Mouse IgG HRP Conjugate (Cell Signalling) 1:2,000 in milk blocking buffer	PBS, 0.1 % (v/v) Tween 20 (3x 15 minute washes)
HA Tag	5 % (w/v) skimmed milk, PBS, 0.1 % (v/v) Tween 20	Anti-HA-Peroxidase (Roche) in milk blocking buffer	PBS, 0.1 % (v/v) Tween 20 (3x 15 minute washes)	-	-

2.5 Biophysical techniques

2.5.1 Mass spectrometry

For protein band identification, gel samples were run on keratin-free precast gels and stained with InstantBlue. Gel bands were excised for in-gel digestion and analysed either by The Advanced Proteomic Facility, University of Oxford or The BSRC Mass Spectrometry Facility, University of St. Andrews.

2.5.2 Circular dichroism

Circular dichroism (CD) data were collected on a Jasco CD spectrophotometer (J-815) in the wavelength range of 190 nm to 260 nm. Protein was dialysed in 0.1 M sodium phosphate buffer, pH 8.0 and diluted in the same buffer to 0.15 mg/mL. CD measurements were performed by the Department of Biochemistry Molecular Biophysics Facility, University of Oxford with a data pitch of 0.1 nm, data integration time of 1 second, bandwidth of 1.00 nm, and scanning speed of 50 nm/min.

2.6 Structural techniques

2.6.2 Negative stain electron microscopy

Purified protein was diluted, in appropriate buffer, in a range between 70 µg/mL and 10 µg/mL. 8 µL of diluted protein was adsorbed on the carbon side of glow discharged (EM ACE200 Vacuum Coater (Leica); 15 mA, 20 seconds) 300 mesh carbon coated copper grids (TAAB Laboratories Equipment Ltd.) for 2 minutes and 30 seconds. Protein droplet on the grid was blotted away with filter paper and the carbon side was washed twice in 20 µL of dH_2O with blotting between wash steps. The carbon side was washed on a 20 µL drop of 2 % (v/v) uranyl acetate, blotted then incubated on a 20 µL drop of 2 % (v/v) uranyl acetate for 30 seconds. After the incubation, the grid was blotted and allowed to air dry carbon side up. Grids were imaged on a Tecnai T12

microscope (FEI) operated at 120 kV equipped with a OneView CMOS camera (Gatan). Imaging of negative stained grids on the Talos 200c (FEI) was performed by Dr. Justin Deme.

2.6.3 Cryogenic electron microscopy

Gold Quantifoil® Holey Carbon R1.2/1.3 grids (300 mesh; spacing 2.5 µM hole to hole) were glow discharged (15 mA, 30 seconds) using an EM ACE200 Vacuum Coater (Leica). Grids were loaded onto a Vitrobot Mark IV (FEI) at 100 % humidity at 8 °C. 4 µL of purified protein was loaded onto the glow discharged side and plunged into liquid ethane cooled by liquid nitrogen. Imaging parameters are described in Chapter 6. Grids were screened using a Talos Artica (FEI) equipped with a Falcon 3 (FEI) direct electron detector and data acquired on a Titan Krios (FEI) equipped with a K3 (Gatan) direct electron detector. All cryogenic electron microscopy imaging was performed by either Emily Furlong or Justin Deme.

2.6.3.1 Data collection

Data were collected in superresolution counting mode on a Titan Krios G3 (FEI) operating at 300 kV with a 20 eV GIF energy filter (Gatan) and K3 direct electron detector (Gatan) using a pixel size 0.669 Å in super resolution mode (1.34 Å real pixel), a dose rate of 7.9 $e^-/Å^2/s$, and an exposure of 6 s, corresponding to a total dose of 47.4 $e^-/Å^2$. Movies were collected over 40 fractions. Data collected on the Titan Krios was performed by Justin Deme.

2.6.3.2 Single particle analysis reconstruction

Motion correction, dose weighting and patched CTF estimation were performed within SIMPLE. Particles were autopicked using reference 2D averages produced from a small subset of manually picked particles. Picked particles (7,376,621) were subsequently extracted into 134 Å by 134 Å boxes and subjected to a 2D cleanup routine using 300 classes in SIMPLE to eliminate junk particles. 2,310,833 particles were recovered from this process and subjected to 2D

clustering using the cluster2D program in SIMPLE (150 classes). 1,016,804 particles were recovered and exported to RELION3.0. Initial 3D model was reconstructed in RELION by applying a mask diameter of 140 Å, 15 ° angular sampling, 50 iterations, starting resolution and final resolution of 35 Å and 15 Å, respectively in 3 classes. 3D classification (3 classes) was performed on the intial 3D model: intial low-pass filter 40 Å, 25 iterations, 150 Å mask with 3.7 ° angular sampling. The data processing in RELION was performed by Susan Lea.

2.7 Cell biology

2.7.1 Cell lysis for expression testing

Cells cultured in 6 cm dishes, were washed three times in ice-cold PBS and lysed in 1 mL 50 mM Tris HCl, 150 mM NaCl, 1 % Triton X-100, pH 7.5 supplemented with cOmplete™ EDTA-free protease inhibitor mix (Roche) and 10 mM 1,10 phenanthroline for 10 minutes. Cells were scraped using a cell scraper and cell debris was pelleted using a bench top microfuge (20,000 *x g*, 10 minutes, 4 °C). Supernatant was mixed with 2X NuPAGE™ lithium dodecyl sulphate (LDS) Sample Buffer (Thermo Fisher Scientific, Invitrogen) to a final concentration of 1X and loaded onto NuPAGE™ 4-12 % Bis-Tris gel (Thermo Fisher Scientific, Invitrogen) and transferred onto methanol activated Amersham Hybond P 0.45 µM Blotting Membrane (GE Healthcare) by the Mini Trans-Blot Cell Assembly (Bio-Rad).

2.7.2 Immunofluorescence

For immunofluorescence staining, cells were cultured in adherent cell culture. Poly-L-lysine coated round coverslips were placed in the wells of 24-well plates. 5 x 10^4 cells were seeded per well and plates incubated in a standing 37 °C incubator with 5 % CO_2 for 48 hours. Cells were washed in PBS at room temperature and fixated using 4 % paraformaldehyde (PFA) in the dark at room temperature with gentle shaking for 20 minutes. PFA was removed, cells were washed

twice in PBS at room temperature and PFA quenched by 50 mM NH_4Cl for 5 minutes. Cells were permeabilised using PBS, 0.2 % Triton-X at room temperature for 20 minutes then washed once in PBS. Cells were blocked in 1 % fish skin gelatine (FSG) at room temperature for 30 minutes then washed three times in PBS. Wells were incubated with primary antibody solution (Table 2.7) for 1 hour at room temperature then washed three times in PBS. Wells were incubated in secondary antibody solution for 1 hour in the dark. Wells were washed three times in PBS followed by three times in dH_2O. Coverslips were inverted and mounted onto glass slide using VECTORSHEILD® mounting medium (Vector Laboratories), excess mounting media blotted away and edges of the coverslip sealed with clear nail varnish. Slides were stored at 4 °C in the dark until imaging.

Table 2.7. Antibodies used for immunofluorescence staining.

Antibody/dye	Species	Dilution
Primary:		
α-BAP31 (Alexis Biochemicals)	Mouse	1 in 250
α-GFP (Thermo Fisher Scientific, Invitrogen)	Rabbit	1 in 500
Secondary:		
Goat α-mouse Alexa Fluor 405 (Thermo Fisher Scientific, Invitrogen)	Goat	1 in 200
	Donkey	1 in 400
Donkey α-rabbit 488 (Life technologies, Thermo Fisher Scientific)	-	1 in 100
Alexa Fluor 647 Phalloidin (Cell Signalling Technologies)		

2.7.3 Cell imaging

2.7.3.1 Live cell imaging

Cells were cultured in adherent cell culture. 2 x 10^5 cells were seeded onto poly-L-lysine coated 35 mm uncoated glass bottom microwell dishes (MatTek Life Sciences) and incubated in a standing 37 °C incubator with 5 % CO_2 for 72 hours. Dishes were imaged by Sonia Muliyil on an Olympus Fluoview FV1200 using a 60X objective.

2.7.3.2 Confocal imaging

Slides from immunofluorescence staining were imaged by Clémence Levet on an Olympus Fluoview FV1000 using a 60X objective.

2.7.3.3 Image analysis

Images generated of the immunofluorescence cell staining were analysed in Fiji ImageJ using a script that exports the individual channels in greyscale and generate merge images in user specified colours. Script was written in Python by Joseph Hogg, refer to Appendix B for script.

2.8 Flow cytometry

2.8.1 Flow cytometry analysers

Cells expressing fluorescent proteins were analysed using either Cytek DxP8, Cytoflex LX (Beckman) or BD Fortessa X20. Viability dye, 4',6-diamidino-2-phenylindole (DAPI), was diluted from a stock to 10 µg/mL. 5 µl of diluted DAPI was added per 1 mL of cell suspension and incubated at room temperature for 5 minutes. Wild type, untransfected or untransduced cells were used as a negative control.

2.8.2 Fluorescence activated cell sorting

Single positive cells for cells only expressing mVenus and double positive cells expressing mVenus and mScarlet-I were enriched by fluorescence-activated cell sorting (FACS) performed by the Dunn School Flow Cytometry Facility using a Beckman Coulter MoFlo AstriosEQ (Beckman) under the 'purify' sorting mode with 1.0 drop envelope. The Summit Software (version: 6.2.7.16492) was used. The optical configuration used is show in Table 2.8.

Table 2.8. Parameters used for fluorescence-activated cell sorting. Cell sorting was performed by the Dunn School Flow Cytometry Facility.

Fluorescent protein or dye	Wavelength (nm)	Band pass filter	Laser power (mW)
mScarlet-I	561	614/20	200
DAPI	405	488/59	50
mVenus/GFP	488	513/26	200

2.8.3 Data analysis

Flow cytometry data were analysed using FlowJo Software (version: 10.6.2). Data was gated and displayed by the following strategy: cells were gated on side scatter area vs forward scatter area, singlets were gated on forward scatter width vs forward scatter area. Live cells were gated on laser for DAPI detection vs forward scatter area. Live single cells were plotted on a histogram against fluorescence. Where two fluorescent proteins were analysed live single cells were plotted red (mScarlet-I) fluorescence vs green fluorescence (mVenus).

Chapter 3: Expression and purification of the iRhom Homology Domain

3.1 Introduction

3.1.1 iRhom Homology Domain as a structural target

As described in Chapter 1, the iRhom Homology Domain (IRHD) is an interesting target for structural biology. The IRHD represents approximately a quarter of the total protein and it is specific to iRhoms, no other member of the mammalian rhomboid-like superfamily contains such a domain. The IRHD is more than just a loop linking TMH1 and TMH2 because it required for the trafficking of proADAM17 to Golgi (Grieve et al., 2017). Cysteines within the IRHD likely form disulphide bonds because the domain is in an oxidising environment in the ER lumen, but it is not known for certain. The IRHD is the only major portion of iRhoms that is present in the ER lumen and therefore likely to be interacting with the ectodomain of ADAM17. Recently, juxtamembrane regions of the IRHD of iRhom2 (IRHD2), regions next to the TMHs, have been shown to stimulate and substrate activity of ADAM17 (Tang et al., 2020). However, there is little else known about the IRHD and its potential interactions with known iRhom1/iRhom2 interactors other than ADAM17.

Structures of extensive extracellular domains (ECD) between TMHs have previously been solved. Such examples are the ECD, or 'loop', of the mouse (*Mm*) PepT1 and rat (*Rn*) PepT2 transporters solved by X-ray crystallography (Beale et al., 2015). These transporters are proton-coupled symporters that are responsible for uptake of dietary peptides. The structure of the *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} revealed that both ECDs consisted of two β-sandwich immunoglobulin-like folds (Beale et al., 2015). Trypsin also localises to a conserved di-aspartate motif within the tertiary structure *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} and interaction was abolished when the motif was removed. Therefore, trypsin localising to the *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} suggests that local concentration of peptides may be higher in close proximity to the transporter hence improving

the transport of the peptides into the cell. In principle, large ECDs between TMHs have been structurally solved and their interactions with other proteins determined. However, in these examples, the ECDs of the *Mm*PepT1 and the *Rn*PepT2 consisted of a high proportion secondary structure and a stable immunoglobulin-like fold. But the examples of the *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} form the basis that extensive ECDs between TMHs can be solved and investigate interacting partners to improve overall understanding.

This result chapter will present and discuss the expression and purification of the IRHD of human iRhom1 (*Hs*IRHD1) with the aim to solve to the structure of *Hs*IRHD1. Solving the structure of the IRHD would be the first domain structure of iRhom and provide insight to how such a cysteine rich domain folds. The structure of the IRHD may also aid in understanding its function in binding to ADAM17.

3.2 Bioinformatics

iRhom amino acid sequences from species listed in Table 3.1 were aligned using Clustal Omega Multiple Sequence Alignment Tool (EMBL-EBI) (Sievers et al., 2011, Madeira et al., 2019) and viewed and analysed in Jalview (Waterhouse et al., 2009). Within Jalview, the predicted IRHD was excised from the alignment generated and the consensus sequence of the IRHD was used to predict the secondary structure and disorder using JPred4 (Drozdetskiy et al., 2015) and Jronn (Yang et al., 2005, Troshin et al., 2011), respectively, within Jalview. The IRHD alignment (Figure 3.1) of different species showed that the IRHD is conserved. The sixteen cysteine residues are conserved across all aligned species. However, there was little secondary structure predicted for *Hs*IRHD1 as only ~ 40 % of residues are predicted to have secondary structure (26 % α-helical, 16 % β-sheet and 28 % disordered). For a good crystallographic target a high proportion of predicted secondary structure is ideal but this is not the case for the IRHD. But the uniqueness of the IRHD is an interesting and worthwhile target to solve. The IRHD alignment corroborates

previously published data (Lemberg and Freeman, 2007). Alignment of the full length iRhom sequences is in Appendix C and visualised using ESPript (Robert and Gouet, 2014).

Table 3.1. List of species used in alignments.

Species	Organism
Alligator mississippiensis	Alligator
Anolis carolinensis	Green anole
Boleophthalmus pectinirostris	Bluespotted mudhopper
Bos taurus	Cattle
Callithrix jacchus	Common marmoset
Canis lupus familiaris	Domestic dog
Danio rerio	Zebrafish
Drosophila melanogaster	Fruit fly
Felis catus	Domestic cat
Gallus gallus	Chicken
Homo sapiens	Human
Hypsibius dujardini	Tardigrade
Larimichthyscrocea	Yellow croaker
Meriones unguiculatus	Mongolian gerbil
Mus caroli	Ryukyu mouse
Mus musculus	Mouse
Odocoileus virginianus texanus	White-tailed deer
Oryctolagus cuniculus	European rabbit
Papio anubis	Olive baboon
Plecturocebus moloch	Red-bellied titi
Ramazzottius varieornatus	Tardigrade
Rattus norvegicus	Rat
Xenopus tropicalis	Western clawed frog

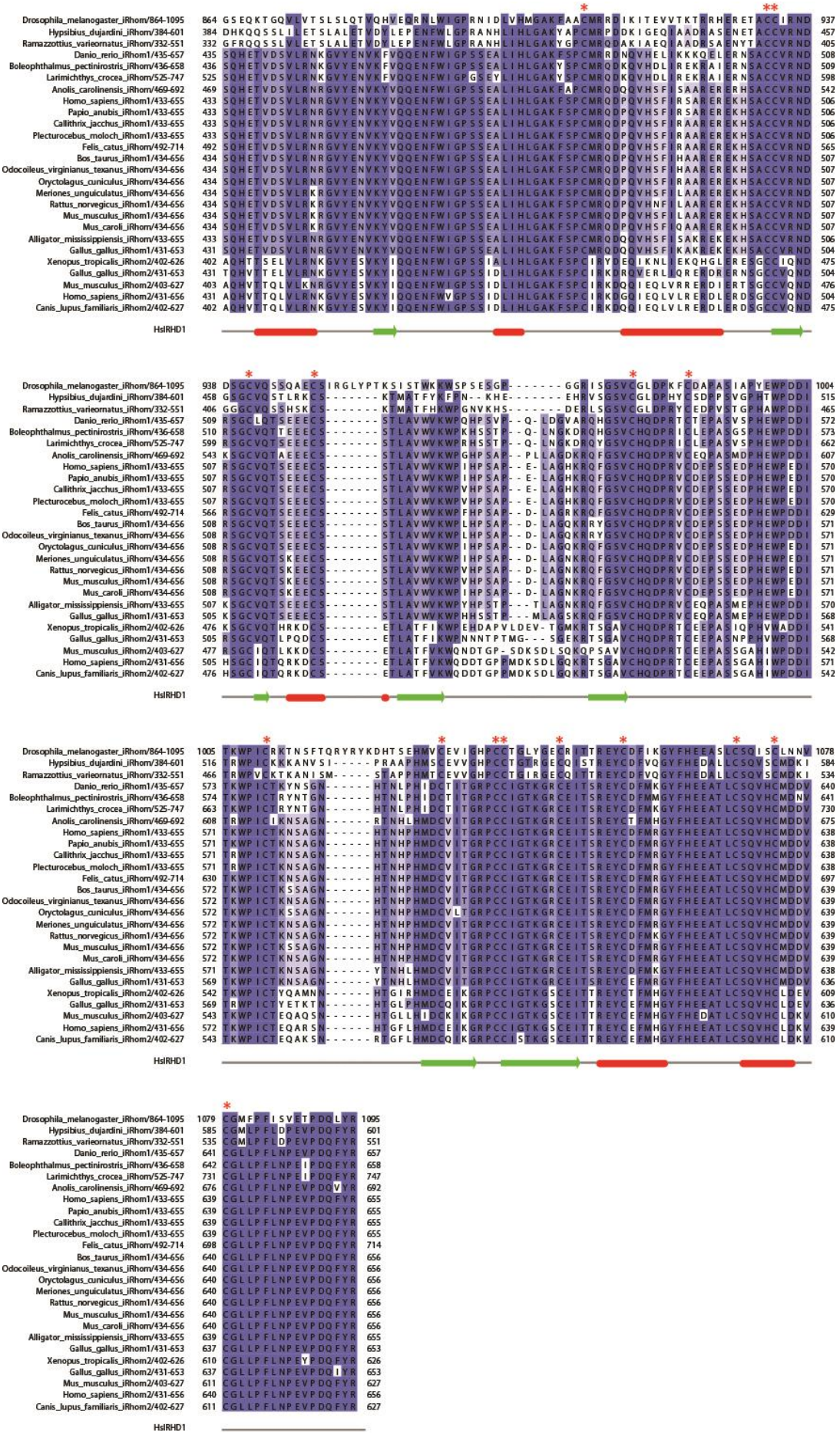

Drosophila_melanogaster_iRhom/864-1095
Hypsibius_dujardini_iRhom/384-601
Ramazzottius_varieornatus_iRhom/332-551
Danio_rerio_iRhom1/435-657
Boleophthalmus_pectinirostris_iRhom/436-658
Larimichthys_crocea_iRhom/525-747
Anolis_carolinensis_iRhom/469-692
Homo_sapiens_iRhom1/433-655
Papio_anubis_iRhom1/433-655
Callithrix_jacchus_iRhom1/433-655
Plecturocebus_moloch_iRhom1/433-655
Felis_catus_iRhom/492-714
Bos_taurus_iRhom1/434-656
Odocoileus_virginianus_texanus_iRhom/434-656
Oryctolagus_cuniculus_iRhom/434-656
Meriones_unguiculatus_iRhom/434-656
Rattus_norvegicus_iRhom1/434-656
Mus_musculus_iRhom1/434-656
Mus_caroli_iRhom1/434-656
Alligator_mississippiensis_iRhom/433-655
Gallus_gallus_iRhom1/431-653
Xenopus_tropicalis_iRhom2/402-626
Gallus_gallus_iRhom2/431-653
Mus_musculus_iRhom2/403-627
Homo_sapiens_iRhom2/431-656
Canis_lupus_familiaris_iRhom2/402-627
HsIRHD1

Figure 3.1. The iRhom homology domain (IRHD) is highly conserved with little secondary structure. Multiple sequence alignment of the IRHD from various species was performed using Clustal Omega and visualised in Jalview. Conserved cysteines are highlighted by the red asterisk (*). Predicted secondary structure of the IRHD of *Homo sapiens* iRhom1 is shown and was generated using the JPred Secondary Structure Prediction within Jalview. Dark purple represents identical residues. Red tubes represent helices and green arrows represent beta sheets.

3.3 Schneider 2 cell expression of *Hs*IRHD1

3.3.1 Stable cell line generation expressing *Hs*IRHD1

*Hs*IRHD1 was cloned into pExpreS2-2 vector, for insect cell expression, including a BiP signal peptide, cleavable N-terminal penta-Histidine tag and C-terminal Twin Strep tag by Dr. Stefan Düsterhöft as shown in Figure 3.2A. Insect cell expression was chosen because it is a eukaryotic cell line that can express secreted proteins and can be maintained at room temperature. The media used to culture S2 cells contains biotin therefore two affinity tags were cloned so that secreted, due to the signal peptide, *Hs*IRHD1 could be purified from media by the penta-histidine tag. Previous experience from the lab using this expression system found that using the penta-histidine tag also bound to non-specific proteins (personal communication, Dr. Stefan Düsterhöft). Therefore the Twin Strep tag was added to anticipate further purification required of *Hs*IRHD1 from the His-tag eluate and that the Twin Strep tag is specific affinity tag. The *Hs*IRHD1:pExpreS2-2 construct was transfected in *D. melanogaster* S2 suspension cells (ExpreS2ion Biotechnologies) to generate a stable cell-line following the manufacturer's protocol. During the selection period, western blot was performed on the supernatant (media) of the transfected and non-transfected S2 cells using the StrepMAB Classic antibody (IBA Lifesciences) to confirm *Hs*IRHD1 expression. Shown in Figure 3.2B, the transfected cell line expressed recombinant *Hs*IRHD1 migrated at a slightly higher molecular weight (~ 35 kDa) expected molecular weight of 30 kDa. In the transfected cell line there was a smaller StrepMAB reactive band ~ 18 kDa which was likely due to degradation of the *Hs*IRHD1. The ~ 35 kDa band

shown in the media control is likely to be overspill from loading or non-specific binding of the antibody of a similar molecular weight of *Hs*IRHD1.

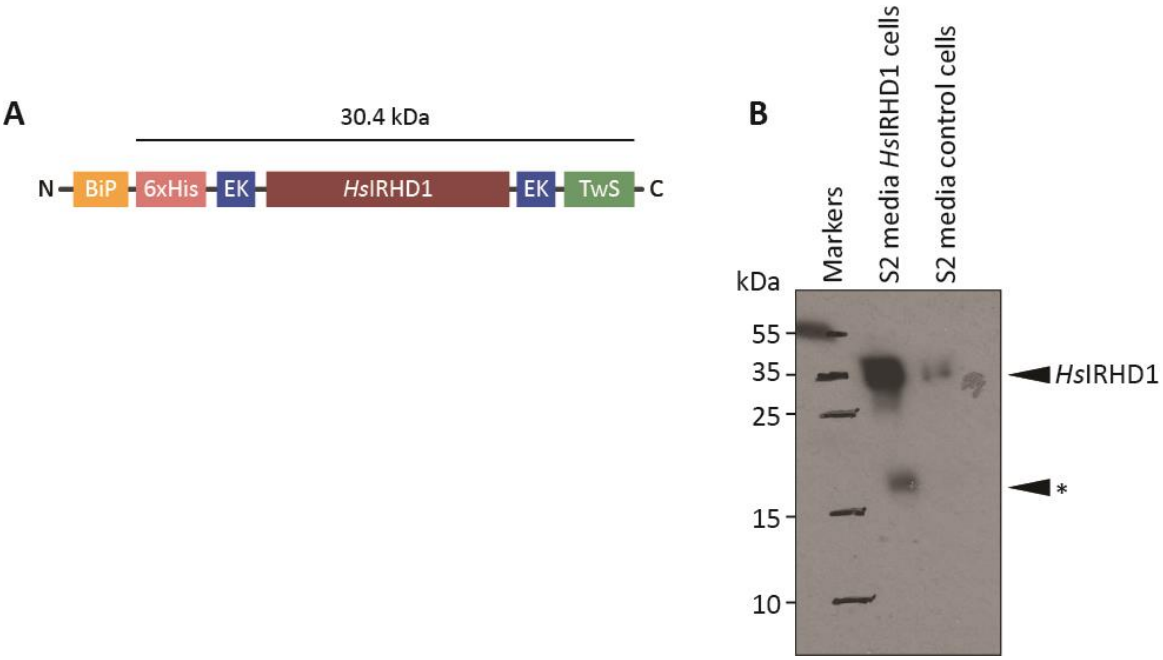


Figure 3.2. *Hs*IRHD1 is expressed in the media of S2 cells. (A) Construct design of *Hs*IRHD1 expressed in S2 cells. *Hs*IRHD1 was recombinantly expressed with an N-terminal hexahistidine (6xHis) tag and a C-terminal Twin Strep tag (TwS). Both affinity tags are enterokinase-cleavable (EK). *Hs*IRHD1 is expressed and secreted into the media **(B)** due to the *Drosophila melanogaster* binding protein signal peptide (BiP). *Hs*IRHD1 (black arrow) was immunoblotted against the Twin Strep tag in the media of transfected (S2 media *Hs*IRHD1 cells) and non-transfected cells (S2 media control cells). A strep antibody reactive band is highlighted by the asterisk (*).

3.3.2 Purification of *Hs*IRHD1 from Schneider 2 cell media

After confirmation of *Hs*IRHD1 expression, S2 cells expressing *Hs*IRHD1 were expanded, in serum-free EX-CELL® 420 media (Sigma-Aldrich), to a volume of 1-4 L and harvested at a density of 15-30 x 10^6 cell/mL. S2 insect cell media was harvested by centrifugation at 1000 *x g* for 30 minutes at 4 °C. The supernatant was filtered using a 0.22 µm filtering unit (Millipore). *Hs*IRHD1 in the clarified, conditioned media was purified using tandem affinity chromatography using its His-tag and Twin Strep tag, followed by size-exclusion chromatography (SEC). Due to high biotin

content present in the cell culture media which precludes Strep-tag affinity purification, purification by immobilized metal affinity chromatography (IMAC) was performed first using cOmplete™ His-Tag Purification resin (Roche) which is further compatible with levels of EDTA present in media. This first purification step was optimised from prepacked columns to batch purification to decrease the total protein preparation time, from harvesting insect cell media to SEC, without affecting the yield. 1 L of *Hs*IRHD1 conditioned media was incubated with 5 mL of cOmplete™ His-Tag Purification free resin with gentle rolling, for at least 4 hours at room temperature. The media/resin was then transferred to an Econo-Column® Chromatography Column (Bio-Rad) for the media to flow through. The resin was washed in 50 column volumes (CVs) phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) supplemented with 5 mM imidazole and a final pH of 7.5 and bound proteins eluted in 6 CVs PBS supplemented with 250 mM imidazole with a final pH 7.5. Eluate was dialysed against 1 L of 100 mM Tris HCl, 150 mM NaCl, pH 8.0 overnight, with stirring at 4 °C using a 10,000 molecular weight cut-off (MWCO). Dialysate was applied to 1 mL Strep-Tactin®XT Superflow® (IBA Lifesciences) column using a peristaltic pump at a flow rate approximately 1 mL/min. The column was washed in 20 CVs of 100 mM Tris HCl, 150 mM NaCl, pH 8.0 and eluted in 8 CVs of 50 mM Biotin, 100 mM Tris HCl, 150 mM NaCl, pH 8.0 collected in 1 mL fractions. Elution fractions containing protein were pooled, concentrated to 500 µL using a 10,000 MWCO centrifugal concentrator (Millipore) and injected to a pre-equilibrated Superdex 200 (S200) 10/300 column (GE Healthcare) in 20 mM Tris HCl, 150 mM NaCl, pH 7.5 at a flow rate of 0.5 mL/min. A representative purification of the optimised three-step purification of *Hs*IRHD1 is shown in Figure 3.3A.

Despite a symmetrical, monomeric elution peak based on the size-exclusion chromatogram (Figure 3.3B), purified *Hs*IRHD1 appeared as multiple bands by SDS-PAGE, possibly due to proteolysis, (Figure 3.3 A and C) the elution peak is that of an expected protein mass of ~ 30 kDa

based on published calibration standards from the manufacturer. The peak fractions were also visualised on a non-reducing SDS-PAGE (Figure 3.3C).

Gel band profiles were similar in both reducing and non-reducing SDS-PAGE (Figure 3.3C) suggesting that no disulphide bonds were formed. This was unexpected because the cysteines are located in an oxidising environment but it may be that cysteines are involved in metal binding (Giles et al., 2003) but was not tested in this thesis. The yield of *Hs*IRHD1 after SEC was approximately 500 µg per litre of S2 cell media unfortunately, expanding up the S2 culture to 4 L did not increase the yield. This may be due to the protein instability during the cell culture expansion and therefore reduced secretion.

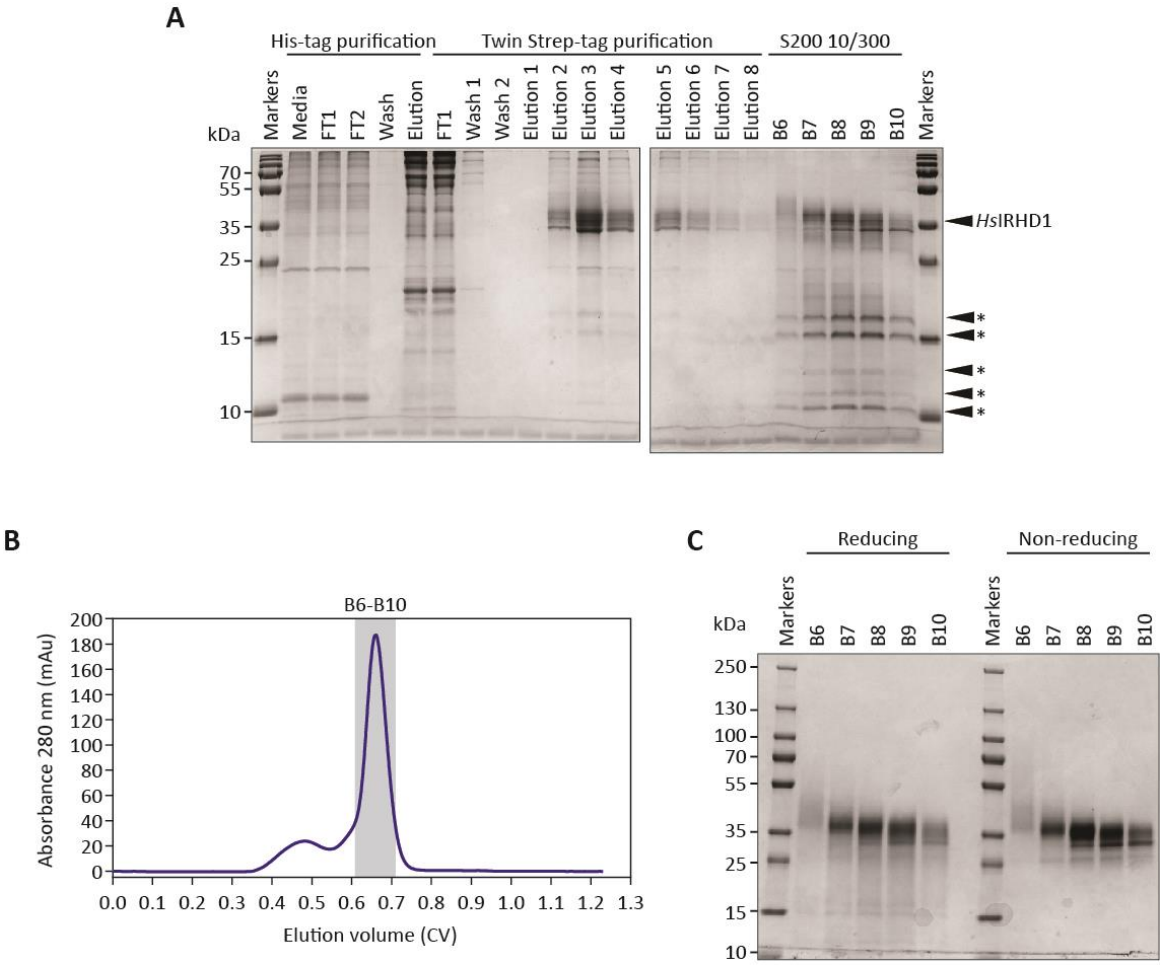


Figure 3.3. Purification of *Hs*IRHD1. *Hs*IRHD1 was purified from S2 cell media using tandem affinity chromatography by first purifying using the hexahistidine tag followed by the Twin Strep tag. **(A)** Each purification stage was visualised by SDS-PAGE. Affinity purified *Hs*IRHD1 was further purified by size-exclusion chromatography (SEC) using the Superdex200 (S200) 10/300 column **(B)** and degradation products shown by the asterisk (*). Grey highlighted region in **(B)** was visualised on SDS-PAGE and in reducing and non-reducing conditions **(C)**. Abbreviations used: FT, flow through; B6-B10, SEC fractions.

3.3.3 Characterisation of *Hs*IRHD1

Purified *Hs*IRHD1 and its apparent degradation products were confirmed by gel band trypsin digest mass spectrometry. *Hs*IRHD1 is predicted to have one N-linked glycosylation site but did not appear glycosylated after glycoprotein staining (Figure 3.4A). For glycoprotein detection, after SDS-PAGE gels were stained for glycoproteins using the Pierce™ Glycoprotein Staining Kit (Thermo Fisher Scientific) following the manufacturer's protocol. The kit detects glycoproteins

by oxidising the cis-diol sugar groups by treating the gel with a periodate solution. The resulting aldehyde groups are detected through the formation of Schiff-base bonds with the kits Glycoprotein Stain that produces magenta bands. The lack of glycoprotein staining may be due to that *Hs*IRHD1 only has one predicted N-linked glycosylation site whereas the positive control, horseradish peroxidase, contains nine N-linked glycans. Therefore the kit may be limited depending on the extent of glycosylation.

*Hs*IRHD1 was subject to circular dichroism to test for secondary structure. Circular dichroism data were collected on a Jasco CD spectrophotometer (Molecular Biophysics Facility, Department of Biochemistry, University of Oxford) in the wavelength range of 190 nm to 260 nm. Protein was dialysed in 0.1 M sodium phosphate buffer, pH 8.0 and diluted in the same buffer to 0.15 mg/mL. The CD measurements were performed on a J-815 CD instrument with a Data Pitch = 0.1 nm, Data Integration Time = 1 second, Bandwidth = 1.00 nm, and scanning speed = 50 nm/min. Circular dichroism shows α-helical structure and random coil (Figure 3.4B) which was consistent with predicted secondary structure (26 % α-helix). An example of expected CD spectra of 'pure' secondary structure protein elements is shown in Figure 3.4C.

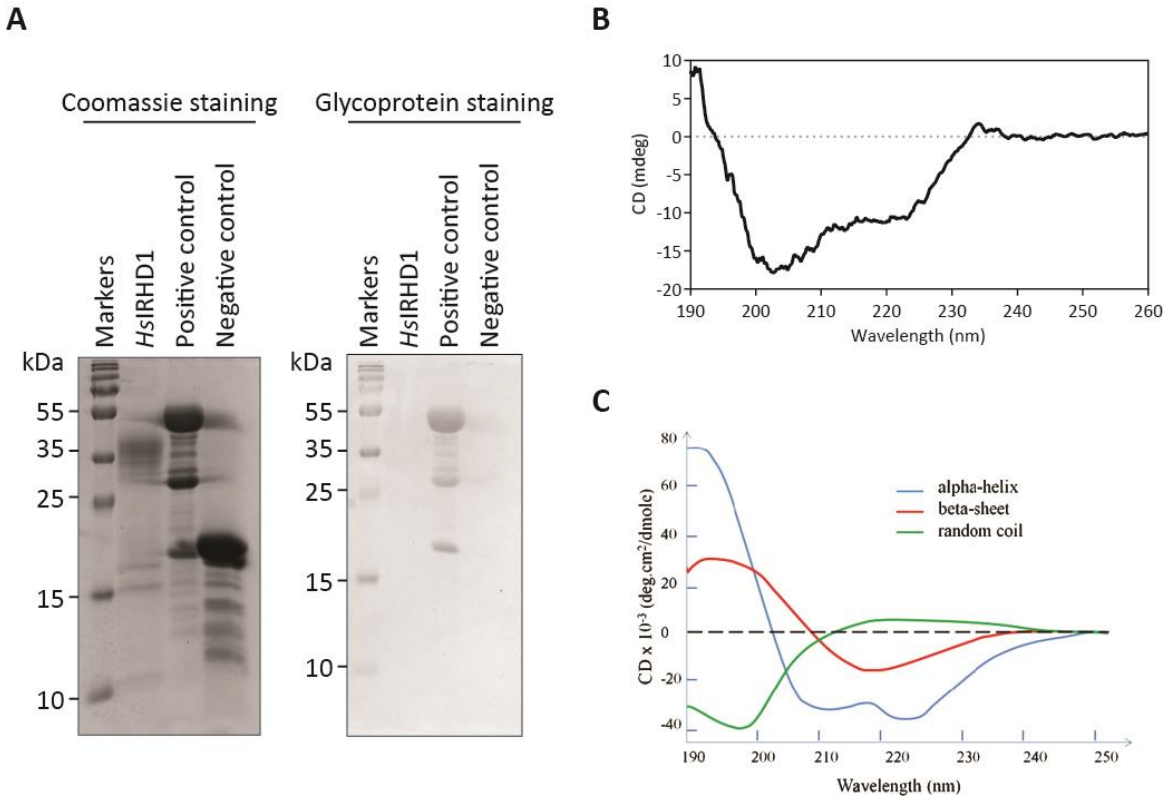


Figure 3.4. Glycoprotein staining and circular dichroism of *Hs*IRHD1 purified from S2 cell media. (A) After SDS-PAGE, each gel was stained separately using Coomassie for all proteins and Pierce™ Glycoprotein Staining Kit (Thermo Fisher Scientific) for glycoproteins. Positive and negative controls were horseradish peroxidase (~ 44 kDa) and soybean trypsin inhibitor (~ 20 kDa), respectively. **(B)** Circular dichroism (CD) of *Hs*IRHD1 purified from S2 cell media. *Hs*IRHD1 was dialysed into 10 mM phosphate buffer (pH 7.4) and diluted to 0.1 mg/ml. **(C)** Expected CD spectra of the three secondary structures of proteins (α-helix, blue trace; β-sheet, red trace; random coil, green trace) and was taken from (Wei et al., 2014) under copyright licence number 4881851315835.

3.3.4 *Hs*IRHD1 is poorly secreted into media

Due to the poor yields of secreted *Hs*IRHD1 in S2 cell media shown Section 3.3.2, a western blot was performed on the S2 cells to check for expression using the StrepMAB Classic antibody (Figure 3.5A). Surprisingly, *Hs*IRHD1 was detected in the cells but not media, suggesting it was being expressed and trapped inside the cell resulting in poor secretion. *Hs*IRHD1 was purified directly from S2 cells using the three-step purification protocol (Figure 3.5B and 3.5C) to test whether the purifying from cells may increase yields.

S2 cells were harvested as described in Section 3.3.2 and the cell pellet was washed in PBS and stored at -20 °C until use. The cell pellet derived from 1 L of cell culture expanding as described in Section 2.3.2.1 would normally yield between 15-20 g of cells. Cell pellet, derived from 1 L of cell culture, was thawed and resuspended in 50 mL of PBS with SIGMA*FAST*™ protease inhibitor cocktail (Sigma) and Benzonase® Nuclease (Sigma). Cell pellet resuspension was aided by a Dounce homogeniser. Cells were lysed by passage through an EmulsiFlex-C5 (Avestin) five times at 10,000 psi. Lysates were centrifuged at 18,000 x *g* for 30 minutes at 4 °C. Supernatant was applied to a 5 mL prepacked Ni-NTA column (Qiagen) following the manufacturer's instructions using a peristaltic pump at a flow rate approximately 5 mL/min. Column was washed in 20 CVs of PBS, 20 mM imidazole at a final pH of 7.5 and bound proteins eluted in 5 CVs of PBS, 250 mM imidazole at a final pH of 7.5. Ni-NTA eluate was then applied to a 1 mL Strep-Tactin®XT Superflow® (IBA Lifesciences) column, following the manufacturer's recommendations, using a peristaltic pump at a flow rate of 1 mL/min. The flow through was re-circulated through the column overnight at 4 °C. The column was washed in 20 CVs of 100 mM Tris HCl, 150 mM NaCl, pH 8.0 and bound proteins eluted in 7 CVs of 100 mM Tris HCl, 150 mM NaCl, 50 mM biotin, pH 8.0, collected in 1 mL fractions. Elution fractions containing protein were pooled, concentrated to 500 µL using a 10,000 MWCO centrifugal concentrator (Millipore) and injected to a pre-equilibrated S200 10/300 column (GE Healthcare) in 20 mM Tris HCl, 150 mM NaCl, pH 7.5 at a flow rate of 0.5 mL/min. However the SEC elution profile was different to *Hs*IRHD1 purified from the media (Figure 3.5D). The eluting peaks before 0.6 CV that were not present when purifying from the media (Figure 3.3B), are consistent with oligomerisation or extended conformations. Therefore, *Hs*IRHD1 purified from S2 cells was likely to be mis-/unfolded or aggregated because the disulphide bonds would have been reduced in the cytoplasm. The level of misfolding could be evaluated by comparing gel-band shifts on reducing and non-reducing SDS-PAGE. Proteins that are disulphide bonded will migrate faster through gel in non-reducing conditions compared to reducing conditions because the protein is not fully linearised as the disulphides are not

reduced. Because of the elution profile of *Hs*IRHD1 purified from S2 cells and that amount of protein at 0.65 CV (Figure 3.5D) is almost an order of magnitude less than *Hs*IRHD1 purified from the media (0.65 CV, Figure 3.3B). It became apparent that purifying from cells did not increase yield as most of it was misfolded and aggregated.

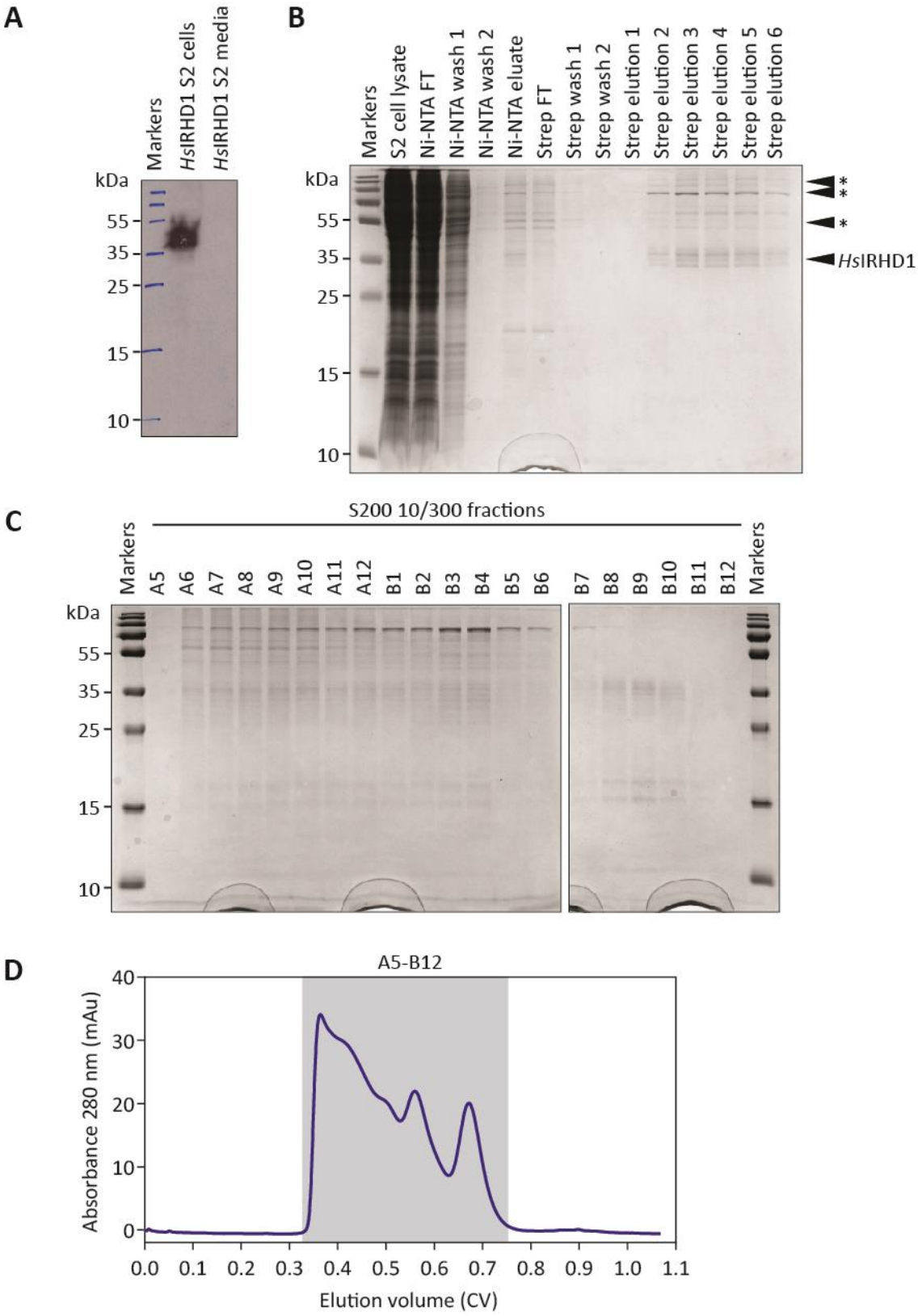

A
Markers
HsIRHD1 S2 cells
HsIRHD1 S2 media
kDa
55
35
25
15
10
B
Markers
S2 cell lysate
Ni-NTA FT
Ni-NTA wash 1
Ni-NTA wash 2
Ni-NTA eluate
Strep FT
Strep wash 1
Strep wash 2
Strep elution 1
Strep elution 2
Strep elution 3
Strep elution 4
Strep elution 5
Strep elution 6
kDa
55
35
25
15
10
**
*
HsIRHD1
C
S200 10/300 fractions
Markers
A5
A6
A7
A8
A9
A10
A11
A12
B1
B2
B3
B4
B5
B6
B7
B8
B9
B10
B11
B12
Markers
kDa
55
35
25
15
10
D
A5-B12
Absorbance 280 nm (mAu)
40
30
20
10
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
Elution volume (CV)

Figure 3.5. *Hs*IRHD1 remains in S2 cells. (A) Western blot of *Hs*IRHD1 in S2 cell lysates and S2 conditioned media. *Hs*IRHD1 was detected by the Twin Strep tag. SDS-PAGE of dual affinity (His tag, Ni-NTA and Twin Strep tag, Strep) chromatography **(B)** and size-exclusion chromatography (SEC fractions: A5-B12) **(C)** using the Superdex200 10/300 column **(D)**. Eluted proteins from the Strep purification are highlighted with the black arrows. Abbreviation used: FT, flow through.

3.4 Bacterial expression and purification of *Hs*IRHD1

3.4.1 Small-scale soluble protein expression of *Hs*IRHD1 in bacterial cells

To alleviate poor protein expression and yields from S2 cells, *Hs*IRHD1 was codon optimised for bacterial expression by GeneArt (Thermo Fisher Scientific) and sub-cloned into pETM14b to include a cleavable N-terminal hexa-histidine (Figure 3.6A) by Gibson Assembly as described in Section 2.2.5. The main caveat of using bacterial cells to express *Hs*IRHD1 is that bacteria are limited in post-translational modifications. Bacteria do not have an oxidising subcellular compartment for forming disulphide bonds, if they are formed in *Hs*IRHD1. *Hs*IRHD1 expression was tested in two *E. coli* expression strains: BL21 (DE3) and SHuffle T7 cells (NEB). *E. coli* SHuffle T7 cells are engineered to constitutively express DsbC, a disulphide bond isomerase, in the cytosol to promote disulphide bond formation and therefore better express the cysteine-rich *Hs*IRHD1. *E. coli* expression cells were grown in 1 L LB media in an orbital shaker at 37 °C for BL21 (DE3) cells or 30 °C for SHuffle T7. Cells treated for induced protein expression, cell harvest and lysis as described in Section 2.4.1.3. The supernatant and pellet after cell lysis was visualised on a SDS-PAGE (Figure 3.6B and C). There was better induction of *Hs*IRHD1 (~ 28 kDa) in BL21 than SHuffle cells. However in both cases, *Hs*IRHD1 was found in the pellet following cell lysis and centrifugation, suggestive of incorporation into inclusion bodies. This was not surprising because the *Hs*IRHD1 expressed was not able to fold properly with the number of cysteines and as result the protein was shuttled into inclusion bodies.

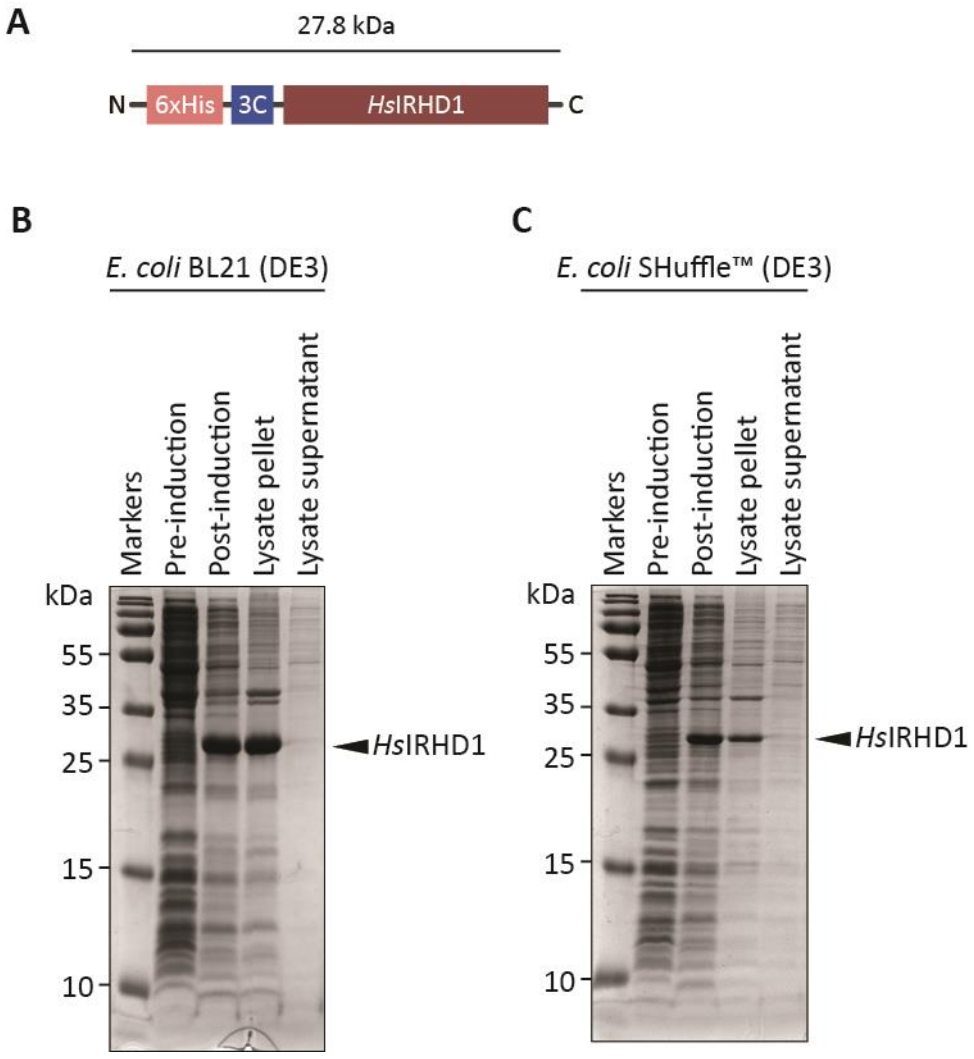


Figure 3.6. Bacterial expression trials of *Hs*IRHD1. Expression of *Hs*IRHD1 with 3C cleavable N-terminal hexa-histidine tag **(A)** (27.8 kDa) was induced by 1 mM IPTG in **(B)** *E. coli* BL21 (DE3) and **(C)** *E. coli* SHuffle T7 cells. Expressed *Hs*IRHD1 was insoluble and therefore present in the pellet and not the supernatant after cell lysis.

3.4.2 Bacterial protein expression into inclusion bodies and preparation of inclusion bodies

Because induced expression of *Hs*IRHD1 was suggested to be incorporated into inclusion bodies, and the large quantity of protein, attempts were made to solubilise and refold from inclusion bodies. *Hs*IRHD1 was expressed in BL21 cells rather than SHuffle cells because there was better induction (Figure 3.6B and C, post-induction lanes) as BL21 cells were induced at 37 °C and expressed at 21 °C whereas SHuffle cells were induced at 30 °C and expressed at 21 °C. SHuffle cells are a *lacIq* strain that produce more lac repressor for a tighter control of T7-based plasmid expression, whereas BL21 lacks this repressor and therefore the BL21 strain produces more protein. Because refold experiments were trialled, a large amount of protein was required to begin thus the BL21 strain was the appropriate strain to use. *Hs*IRHD1 was expressed and purified from inclusion bodies described in Section 2.4.2. After centrifugation of the cell lysate, the pellet was resuspended in 50 mM Tris HCl, 1 mM EDTA, 150 mM NaCl, pH 8.0 and Tween 20 was added to the resuspended pellet to wash the inclusion bodies. Washed inclusion bodies were solubilised in a buffer containing 8 M urea and 25 mM DTT to denature proteins and reduce any disulphide bonds formed. The solubilised inclusion bodies were acidified to pH 3-4 to protonate reduced cysteines to prevent spontaneous disulphide formation and dialysed to remove the DTT from solubilised *Hs*IRHD1 solution. Solubilised *Hs*IRHD1 was refolded in an oxido-reduction buffer containing 1 mM cysteine, 2 mM cystine and with or without the addition of 500 mM arginine. Arginine protects the hydrophobic core of proteins during the refold to prevent aggregation. Refolded *Hs*IRHD1 was purified by SEC rather than affinity chromatography because the isolation of inclusion bodies was a purification step because majority of the protein was *Hs*IRHD1. Purification, solubilisation and SEC of *Hs*IRHD1 from inclusion bodies in the refold condition containing 500 mM arginine are shown in Figure 3.7. Disulphide bonds were formed during the refold because of the gel band shift between a reducing and non-reducing SDS-PAGE of the monomeric peak which eluted at ~ 0.55 CV (highlighted in grey in Figure 3.7C). A reducing/non-reducing SDS-PAGE was also performed (Figure 3.7E and Figure 3.7F) on earlier

eluted fractions (0.4 CV) which show aggregates and/or multimers of *Hs*IRHD1 whereby protein did not enter the resolving gel. Of the two refold conditions tested (± 500 mM arginine), the refold buffer containing 500 mM arginine (Figure 3.7C**,** solid line) yielded more monomeric *Hs*IRHD1 than without (Figure 3.7C dashed line). However the non-reducing SDS-PAGE of the elution peak (Figure 3.7B) shows one refolded state and a smear of other refolded species. Circular dichroism of refolded *Hs*IRHD1 in the presence of 500 mM arginine (Figure 3.7D) show no secondary structure elements and resembles the spectra of a random coil. Therefore the refolded protein has a different structure to the *Hs*IRHD1 purified from S2 cell culture (Figure 4.3B).

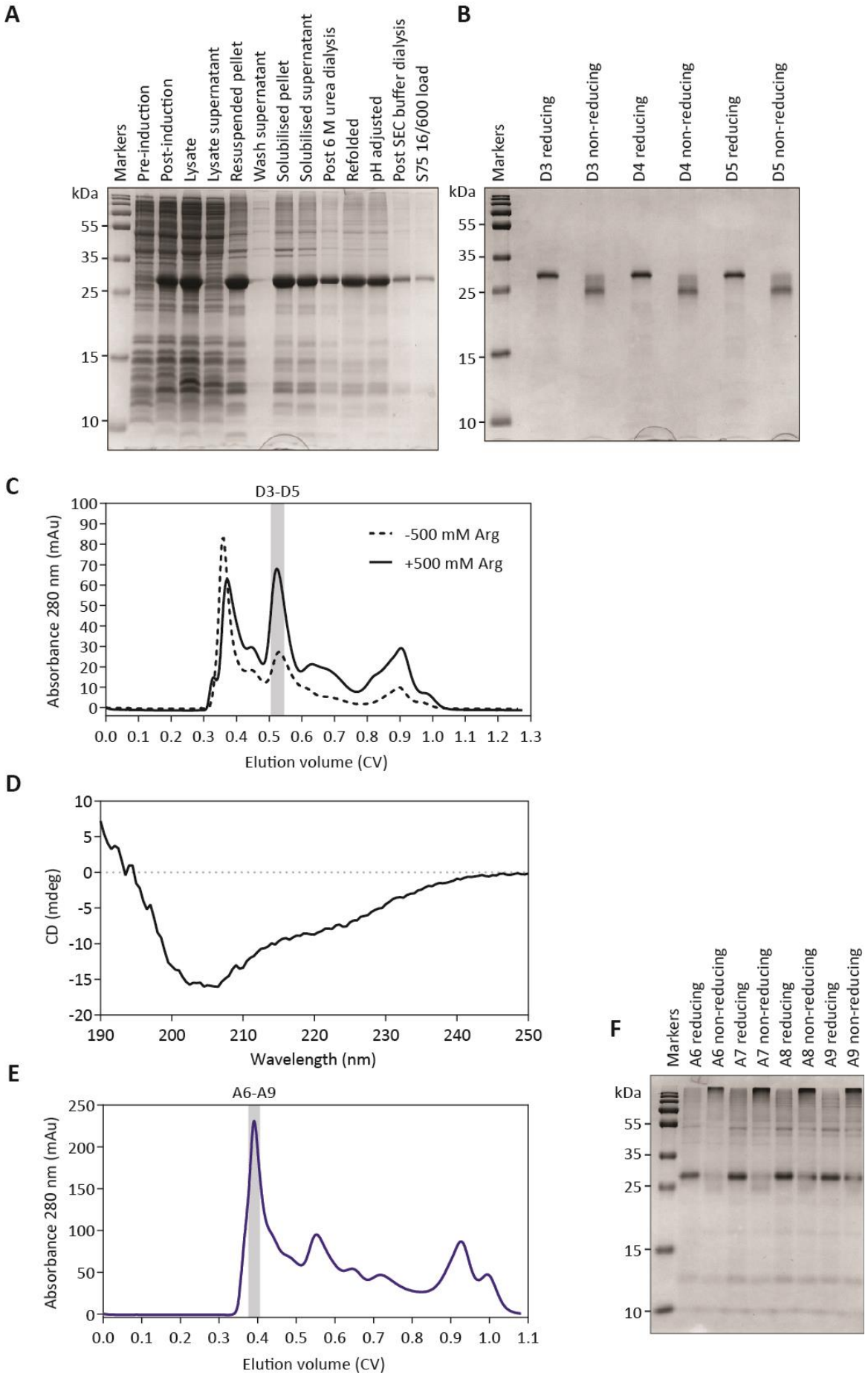

A
Markers
Pre-induction
Post-induction
Lysate
Lysate supernatant
Resuspended pellet
Wash supernatant
Solubilised pellet
Solubilised supernatant
Post 6 M urea dialysis
Refolded
pH adjusted
Post SEC buffer dialysis
S75 16/600 load
kDa
55
35
25
15
10
B
Markers
D3 reducing
D3 non-reducing
D4 reducing
D4 non-reducing
D5 reducing
D5 non-reducing
kDa
55
35
25
15
10
C
D3-D5
-500 mM Arg
+500 mM Arg
Absorbance 280 nm (mAu)
Elution volume (CV)
D
CD (mdeg)
Wavelength (nm)
E
A6-A9
Absorbance 280 nm (mAu)
Elution volume (CV)
F
Markers
A6 reducing
A6 non-reducing
A7 reducing
A7 non-reducing
A8 reducing
A8 non-reducing
A9 reducing
A9 non-reducing
kDa
55
35
25
15
10

Figure 3.7. *Hs*IRHD1 refolded from bacterial inclusion bodies. (A) SDS-PAGE of inclusion body purification, refold steps of *Hs*IRHD1 in the +500 mM arginine refold condition and size-exclusion chromatography (SEC) **(B)** using the Superdex75 (S75) 16/600 column **(C)**. Fractions of the +500 mM arginine refold condition (solid black line) in the grey highlighted region in **(C)** were visualised by SDS-PAGE under reducing and non-reducing conditions. SEC elution profile of *Hs*IRHD1 refolded in the -500 mM arginine refolded condition is shown by the dashed black line. **(D)** Circular dichroism spectra of refolded *Hs*IRHD1 in the presence of 500 mM arginine. **(E, F)** Multimeric *Hs*IRHD1 was present in the first eluting peak and the fractions in grey highlighted area in **(E)** was visualised by SDS-PAGE in reducing and non-reducing conditions **(F)**. The trace shown in **(E)** was from another *Hs*IRHD1 refold in the +500 mM arginine condition.

Anion exchange was attempted to resolve the refolded states. *Hs*IRHD1 was dialysed overnight at 4 °C against 50 mM Tris HCl, 20 mM NaCl, pH 8.5. A MonoQ 5/50 column (GE Healthcare) was equilibrated in Buffer A (50 mM Tris HCl, pH 8.5) and 2 % Buffer B (50 mM Tris HCl, 1 M NaCl, pH 8.5) prior to sample application. Protein was eluted from the column with a continuous gradients up to 40 % Buffer B over 40 CVs and final wash of a gradient up to 100 % Buffer B over 10 CVs. These refolded states were unsuccessfully separated by anion exchange (Figure 3.8) because different refolded states were present in the elution peak, due to the smear. It appeared that one major refold species was present and a smear of other species.

As a final attempt to resolve the refolded *Hs*IRHD1 species to form sharp protein band, high performance liquid chromatography (HPLC) was performed. Concentrated and purified *Hs*IRHD1 was applied to a 5 µm Discovery BIO Wide Pore C5 (10 cm x 4.6 mm) (Sigma-Aldrich) column equilibrated in 0.1 % (v/v) trifluoroacetic acid (TFA), 25 % (v/v) acetonitrile. Proteins eluted off the column in a gradient up to 0.1 % TFA, 75 % acetonitrile over 15 CVs. *Hs*IRHD1 eluted after 40 % concentration of 75 % acetonitrile. There was a poorly resolved double peak shown in Figure 3.9 and refolded species were present in this peak because the non-reducing SDS-PAGE shows a smear.

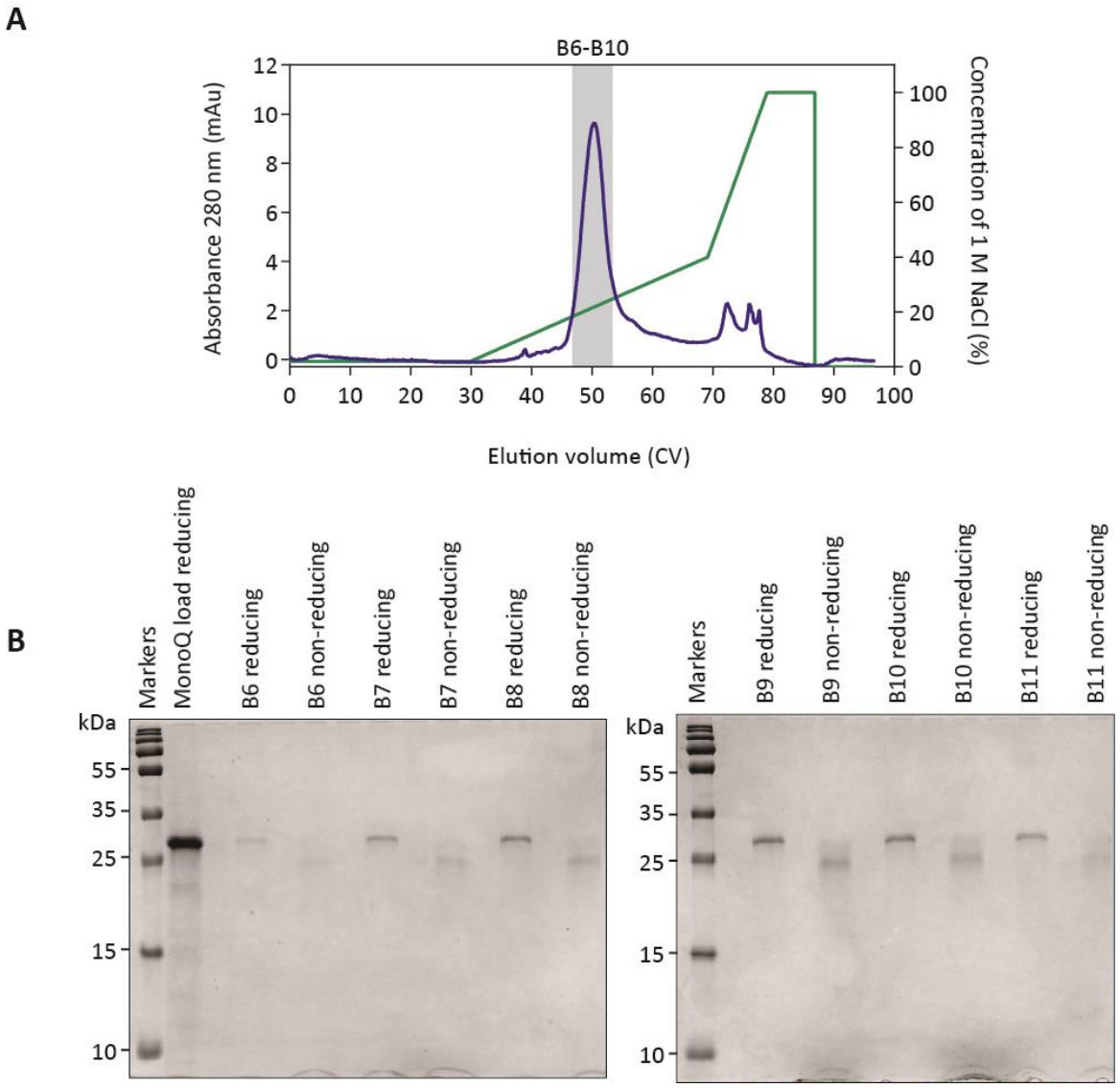


Figure 3.8. Anion exchange of purified refolded *Hs*IRHD1. After size-exclusion chromatography, *Hs*IRHD1 was applied to a MonoQ 5/50 **(A)** column in attempt to separate the refolded states. *Hs*IRHD1 was eluted via a gradient elution (green trace) from 2 % of 1 M NaCl (20 mM) to 100 % of 1 M NaCl. *Hs*IRHD1 eluted (blue trace) at ~25 % concentration of 1 M NaCl (250 mM NaCl). Fractions from the elution peak highlighted in grey **(A)** were visualised on SDS-PAGE in reducing and non-reducing conditions **(B)**.

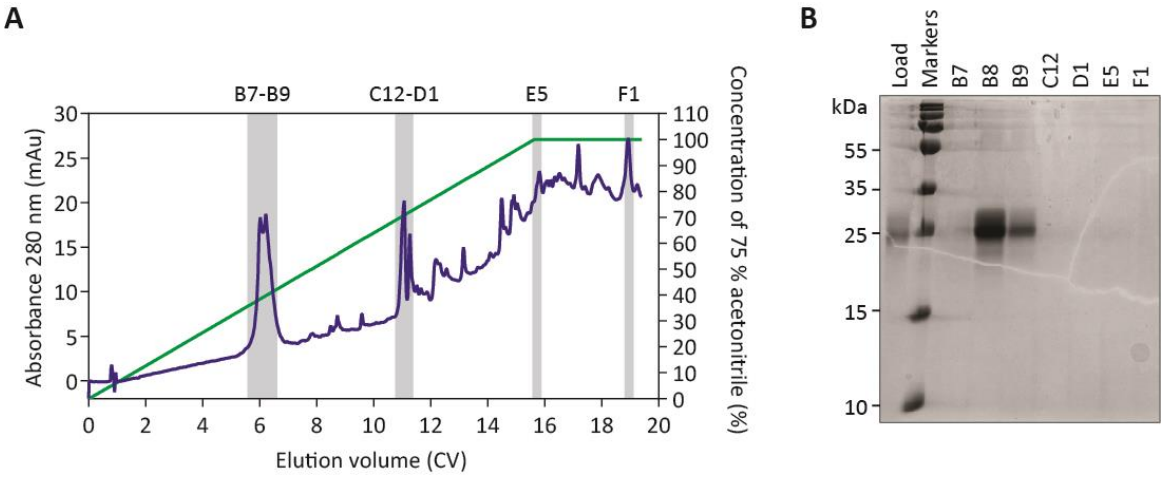


Figure 3.9. The refolded species of *Hs*IRHD1 was not resolvable by high-performance liquid chromatography. Refolded purified *Hs*IRHD1 was injected into a high-performance liquid chromatography (HPLC) column (5 µm Discovery BIO Wide Pore C5 (10 cm x 4.6 mm)) in attempt to separate the refolded species of *Hs*IRHD1 (A). Fractions on the HPLC trace **(A)** highlighted in grey were visualised by non-reducing SDS-PAGE **(B)**.

3.5 Discussion

3.5.1 Further strategies for purifying IRHD

At the start of the project, probing IRHD seemed to be an attractive and interesting target for structural studies because not much is known about its function and architecture. However, given the high cysteine content, the possibility that the IRHD is glycosylated and the lack of secondary structure proved to be difficult to express and purify in S2 cells. This is unsurprising because glycosylated and disulphide-rich mammalian proteins tend to be more difficult to express. Optimising insect expression yielded approximately 500 µg of pure *Hs*IRHD1 per litre of insect culture medium after reducing the total purification time.

The gel band profiles were similar in both reducing and non-reducing SDS-PAGE (Figure 3.3C) suggesting that no disulphide bonds were formed in the purified protein. It has not been previously confirmed whether the cysteines do in fact form disulphide bridges, it has only been inferred because the cysteines are located in an oxidising environment of the ER. But it may be

that cysteines are involved in metal binding and that cysteines can act as 'redox switches' or a storage facility for excess metal ions ($Fe^{2+}/^{3+}$, Zn^{2+}, Cd^{2+}, and Cu^{+}) (Giles et al., 2003). The metal binding capacity of the cysteines may be relevant for ADAM17, because the catalytic site of ADAM17, in the metalloprotease domains, coordinates a zinc ion. The yield of *Hs*IRHD1 after SEC was approximately 500 µg per litre of S2 cell media unfortunately, expanding up the S2 culture to 4 L did not increase the yield. This may be due to the protein instability during the cell culture expansion and therefore reduced secretion. The approach used, in this thesis, to purify *Hs*IRHD1 from S2 cell culture media may be improved by adding metal ions to media and/or during the purification process.

Despite successful purification, *Hs*IRHD1 protein yields generated from insect culture were not high enough to attempt crystallisation trials because milligram quantities were required. In an attempt to improve yields, *Hs*IRHD1 was expressed into inclusion bodies of *E. coli* BL21 cells for refolding. Refolding conditions that incorporated 500 mM arginine in the refold buffer were 'successful' in generating apparently monomeric IRHD because of the SEC elution peak at ~ 0.55 CV (Figure 3.7C). After non-reducing SDS-PAGE analysis of the SEC monomeric peak there were multiple refold species present (Figure 3.7B). These species could not be separated after ion exchange chromatography or HPLC. The different refold species were likely the products of the uncontrolled nature of the refolding process. The sixteen cysteines within *Hs*IRHD1 are conserved which may mean that up to eight disulphide bonds can potentially form, which we do not know for certain. Therefore, during the *in vitro* refolding experiments up to eight disulphides can form and the purification products likely reflect this. Native IRHD folding would likely involve protein disulphide isomerases and lectin chaperones that bind to the N-glycosylation site within the protein to aid in folding. Because a homogenous *Hs*IRHD1 purification product was not achieved, it was not suitable for structural studies.

3.5.2 IRHD may be more stable as part of the full length protein

The logic behind purifying *Hs*IRHD1 from a eukaryotic expression system to maximize disulphide bond formation in near-native conditions was sound. Another approach that may have improved *Hs*IRHD1 expression in S2 cells was to design the construct so that *Hs*IRHD1 was flanked by TMH1 on the N-terminus and TMH2 on the C-terminus. But this would mean that *Hs*IRHD1 would be purified as a membrane protein rather than a soluble one and purifying membrane proteins have their additional obstacles because the requirement of extracting the TMHs out of the membrane without aggregation.

If this was the case that *Hs*IRHD1 was stable with its flanking transmembrane helices then expressing the full length protein, iRhom, would be the best approach. By expressing the full length iRhom for structural studies we would be able to understand the IRHD within the context of whole protein and work towards a mammalian rhomboid-like structure by itself and with one of its binding partners, ADAM17. The next results chapter will discuss the expressing full length iRhom and ADAM17 in a mammalian expression system.

Chapter 4: Cell line generation for expression of full length iRhom and ADAM17

4.1 Introduction

4.1.1 Stable iRhom target for structural studies

The previous Chapter discussed the challenges associated with expressing *Hs*IRHD1 and that the IRHD may be more stable as part of the full-length protein. This Chapter will show and discuss the approaches used to express full-length iRhom alone and with its binding partner, ADAM17. There were two main approaches for probing iRhom for structural studies in this thesis: the *Hs*FRMD8iRhom2Δ300 fusion, a 'stable' target, and the iRhom/ADAM17 complex, a 'physiological target'.

A former Freeman Group member, Dr. Ulrike Künzel, designed a construct where the first 300 N-terminal amino acids of iRhom were replaced by FRMD8, a cytosolic binding partner of iRhom. The *Hs*FRMD8iRhomΔ300 construct, also referred to as the *Hs*FRMD8iRhom2 chimera in this thesis, showed enhanced and stable expression at the cell surface of HEK293T cells (Kunzel et al., 2018). The stability of the *Hs*FRMD8iRhom2 chimera in HEK cells is an advantage for the expression and purification, particularly for membrane proteins. A structure of the stabilised *Hs*FRMD8iRhom2 chimera would be valuable tool in helping to understand the organisation of the rhomboid core transmembrane helices (TMHs) in conjunction with the additional TMH7 of iRhom1/2.

The structure of iRhoms bound to one of its physiological binding partners, ADAM17, is important in order to understand the mechanisms that initiate the inflammatory response. Therefore a structure of iRhom1 or iRhom2 in complex with ADAM17 would achieve this goal.

Structure determination of eukaryotic membrane proteins is inherently challenging; to improve search space and enhance chances of structure solution, iRhoms from various species were expressed. Additionally, a 'native' expression system in HEK293T whereby iRhom2 is expressed endogenously with a triple HA tag at native expression levels, from its natural promoter, was tested.

4.2 Expression and purification on previously generated reagents

4.2.1 Testing expression with previously generated reagents

Stable HEK293T cells overexpressing *Hs*iRhom2 under the human cytomegalovirus (CMV) promoter, and HEK293T cells with chromosomally-tagged iRhom2 were constructed by Dr. Ulrike Künzel and Dr. Iqbal Dulloo respectively. In both cell lines iRhom2 contained a triple human influenza hemagglutinin (HA) tag at the C-terminus, with an expected molecular weight of 100.5 kDa (Figure 4.1A). The advantage of this endogenous cell line is iRhom2 will bind its native binding partners in the correct stoichiometric ratios which may result in improved expression and stabilisation. Both cell lines were tested for expression by expansion to ~ 1.5 x 10^7 cells and the cell pellets immunoblotted, as described in Section 2.3.5, against the HA tag (Figure 4.1B). Endogenous *Hs*iRhom2 was not detectable by western blot, likely due to low native expression in HEK cells. In the overexpression cell line, exogenous *Hs*iRhom2 was detectable by western blot at the expected molecular weight of ~ 100 kDa however, along with a strong HA-reactive band at ~ 50 kDa. This degradation product was consistent in molecular weight with the transmembrane helix bundle of *Hs*iRhom2 without its N-terminus. Due to the lack of signal in the endogenous *Hs*iRhom2 cell line, further experiments were performed on the overexpressed *Hs*iRhom2 cell line.

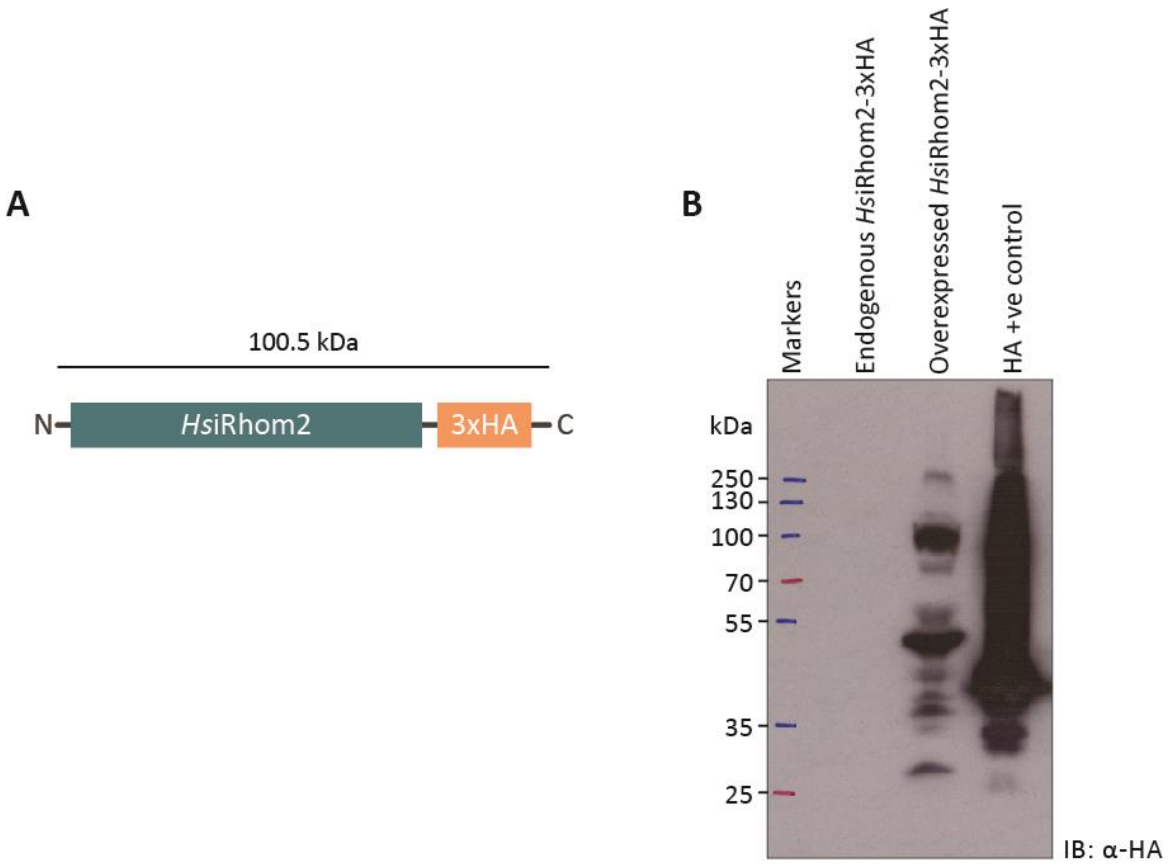


Figure 4.1. Expression of iRhom2 from HEK293T. Endogenous *Hs*iRhom2-3xHA was a knock-in of a triple HA tag on the iRhom2 gene, *RHBDF2*, in HEK393T cells. Overexpressed *Hs*iRhom2-3xHA was a stable cell line generated using a plasmid containing *Hs*iRhom2-3xHA. **(A)** Both endogenously tagged human iRhom2 and overexpressed human iRhom2 were predicted to be the expected size of 100.5 kDa. **(B)** Cell lysates from previously generated human iRhom2 cell lines were immunoblotted for iRhom2 using the HA tag. HA +ve control was a cell lysate that contained a HA tag to test for antibody.

4.2.2 Small scale expansion and purification of overexpressed *Hs*iRhom2

Overexpressed *Hs*iRhom2 HEK293T cells were expanded, at a small scale, to wet cell pellet weight approximately 2 g for immunoprecipitation (IP). Cell pellets were resuspended in 15 mL of PBS supplemented with SIGMAFAST protease inhibitor cocktail, treated with Benzonase and lysed by sonication on ice (10 second bursts at 10 microns with 15 seconds pauses, repeated five times). Cell debris was pelleted by centrifugation at 4000 *x g* and removed. Supernatant was incubated with n-Dodecyl-β-D-Maltopyranoside (DDM) to final concentration of 1 %, for two hours on a rotary wheel at 4 °C, to extract membrane proteins from lipid micelles. 200 µL of

Pierce™ Anti-HA magnetic beads (Thermo Fisher Scientific) were washed in twice in PBS, 0.1 % (v/v) Tween 20 followed by two washes in PBS using a magnetic stand to separate out the beads. Solubilised supernatant from the cell lysate was centrifuged at 20,000 *x g* for ten minutes and the supernatant was retained. Washed beads were applied to the supernatant and incubated at room temperature on a rotary wheel for 30 minutes. After incubation, tubes were applied to a magnetic stand to separate the beads from the supernatant. Beads were washed three times in PBS, 0.02 % (v/v) DDM to remove unbound proteins. Proteins were eluted by the following strategy: beads were incubated with 1 mg/mL HA Peptide (Thermo Fisher Scientific) for 15 minutes at 37 °C with gentle mixing, beads were separated by a magnetic stand and supernatant retained. A second elution fraction was generated by repeating the procedure of incubating the beads with 1 mg/mL HA peptide for 15 minutes at 37 °C. A final elution fraction was generated by incubating the beads in 1 mg/mL HA peptide overnight at 4 °C on a rotary wheel. Each stage was analysed by SDS-PAGE (Figure 4.2A) and western blot (Figure 4.2B) against the HA tag of *Hs*iRhom2. Eluted *Hs*iRhom2 was not directly detectable by Coomassie staining. Protein bands at ~ 100 and ~ 55 kDa in the Boiled HA beads lane of Figure 4.2A could be *Hs*iRhom2 and its degradation product because of the strong HA reactivity in same well of the western blot. The majority of *Hs*iRhom2 did not elute off the beads indicating that elution was not efficient. Elution fractions were incubated with 6 µL StrataClean resin, protein binding beads, to concentrate elution fractions and the beads loaded onto a SDS-PAGE and the gel stained by Coomassie (Figure 4.2C). However, the bands shown in Figure 4.2C were lower than that of the HA reactive band in Figure 4.2B (Elution lanes) at ~ 100 kDa suggesting that other contaminant proteins were eluting. In the Elution 1 lane of Figure 4.2B, there is a protein band below 100 kDa that is consistent with the predicted molecular weight of *Hs*iRhom2.

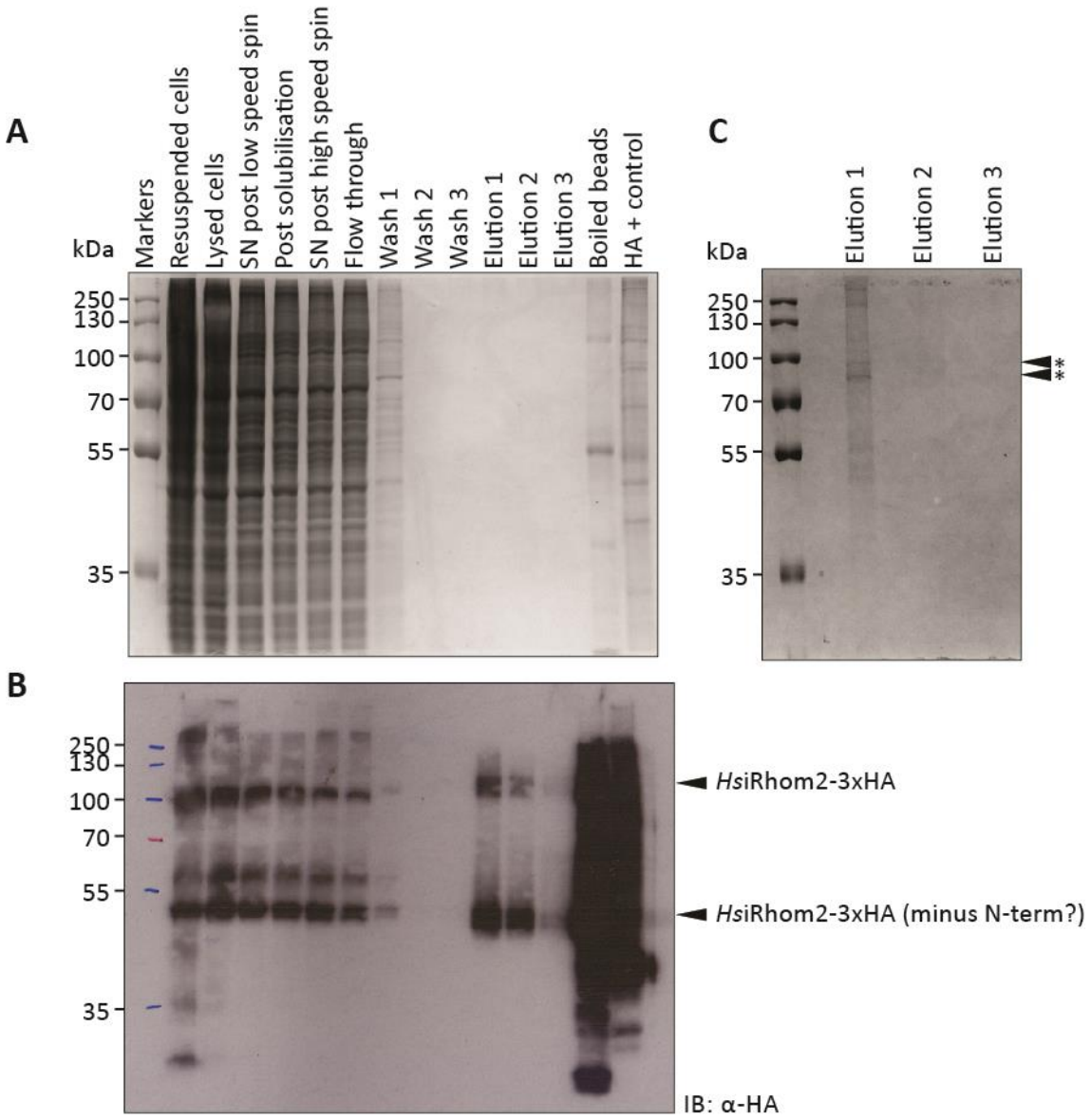


Figure 4.2. Immunoprecipitation of overexpressed *Hs*iRhom2-3xHA. Stages of the immunoprecipitation were visualised using Coomassie stained SDS-PAGE **(A)** and western blot against the HA tag **(B)**. Elution fractions were treated with StrataClean beads to concentrate the elution fractions and visualised by SDS-PAGE **(C)**. Bands highlighted in **(C)** by the asterisk (*) are of a lower size compared to **(B)**. Proteins were eluted using 1 mg/mL HA peptide. The shorter band (~ 50 kDa) HA reactive band was consistent of transmembrane bundle of *Hs*iRhom2 without the N-terminus. HA +ve control was a sample that contained a HA tag. Abbreviation used: SN, supernatant.

4.2.3 Large scale expansion and purification of overexpressed *Hs*iRhom2

HEK293T *Hs*iRhom2 cells were expanded, harvested and stored as described in Section 2.3.2.2.

Once at least 40 g of wet cell pellet accumulated, membranes were isolated as described in

Section 2.4.3 and total protein content in membranes estimated by the Pierce™ BCA Protein Assay Kit as described in Section 2.4.3. Membranes were thawed and diluted to 7.5 mg/mL in PBS supplemented with SIGMAFAST protease inhibitor cocktail tablet. DDM and cholesteryl hemisuccinate tris salt (CHS) was added to the diluted membranes to a final concentration DDM concentration of 1 % (v/v) and CHS concentration of 0.1 % (v/v) from a 10X stock. DDM was chosen as detergent for the purification of HA tagged *Hs*iRhom2, in particular, because it was used successfully for the IP for this cell line (Künzel, 2017). CHS was added because it is a cholesterol analogue that mimics the eukaryotic membrane (Seddon et al., 2004, Kulig et al., 2014), the choice of detergents for other constructs is further discussed in Chapter 5. Membranes were solubilised with gentle mixing for two hours at 4 °C. For large scale purification of HA-tagged *Hs*iRhom2, Pierce™ HA agarose beads were used due to the higher binding capacity compared to magnetic beads. 0.5 mL of HA agarose beads (50 % slurry) were washed five times in PBS. Insoluble material was pelleted by ultracentrifugation at 100,000 *x g* for 30 minutes. Washed beads were added to the solubilised supernatant and incubated with gentle spinning on a rotary wheel overnight at 4 °C. The supernatant/agarose bead suspension was applied to a 1 ml Polypropylene Column (Qiagen) to allow the supernatant to flow through by gravity flow and the beads captured by the column. The beads were washed in 10 column volumes (CVs) of PBS with 0.02 % DDM, 0.002 % CHS. Proteins were eluted off the beads by PBS with 1.2 mg/mL HA peptide, 0.037 % DDM, 0.0037 % CHS in six 0.5 CV fractions. A sample of each stage was kept for western blot analysis (Figure 4.3). This revealed that full length *Hs*iRhom2 (~ 100 kDa) was not present in the supernatant following solubilisation and ultracentrifugation. Suggesting that full length *Hs*iRhom2 may have aggregated at the solubilisation stage and therefore pelleted with the insoluble material. However the 50 kDa degradation product could be solubilised and purified. Other detergents may have improved solubilisation of full length *Hs*iRhom2 but due to the poor yields from the HA purification trials it became evident that these constructs required redesigning.

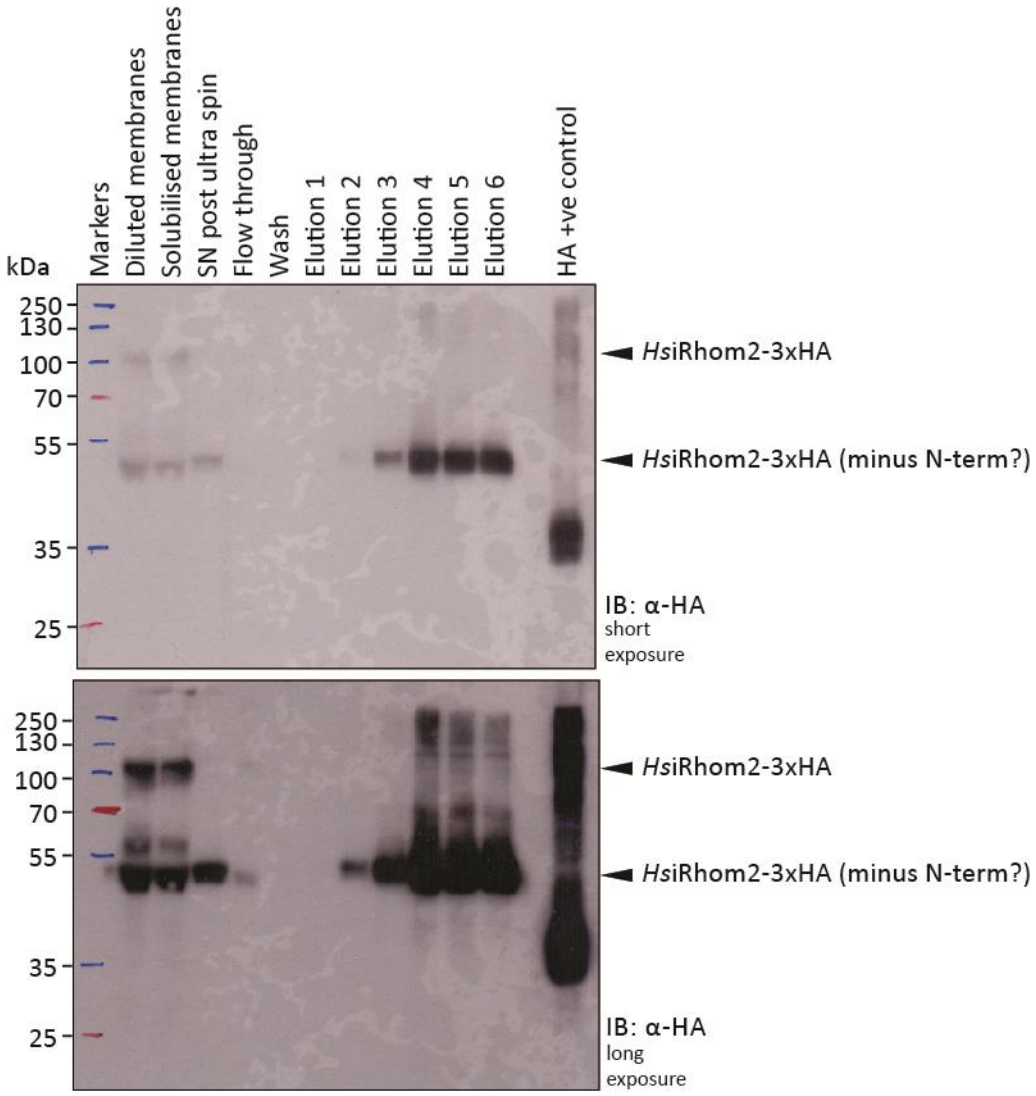


Figure 4.3. Western blot of large scale HA purification. Stages of the purification of *Hs*iRhom2-3xHA using HA agarose beads was visualised by western blot against the HA tag. Full length *Hs*iRhom2-3xHA was not successfully solubilised because it was not present in the elution fractions. Proteins were eluted using 1 mg/mL HA peptide. The shorter band (~ 50 kDa) HA reactive band was consistent of transmembrane bundle of *Hs*iRhom2 without the N-terminus. HA +ve control was a sample that contained a HA tag. Abbreviation used: SN, supernatant.

4.3 Redesigning eukaryotic expression of iRhoms and ADAM17 for protein purification

4.3.1 Construct design

The triple HA tag was not an ideal affinity tag for large scale protein purification because of the poor binding of the HA tag with anti-HA beads and the poor yields from HA peptide elution. The endogenous levels of *Hs*iRhom2 in HEK293T were not high enough for structural studies. The

constructs were redesigned so that all iRhom constructs, including mammalian, non-mammalian, and the *Hs*FRMDiRhom2 chimera contained a C-terminal 3C cleavable Twin-Strep tag. The ADAM17s were further redesigned to incorporate a C-terminal 3C cleavable hexahistidine tag (Figure 4.4). Tags were added to the C-terminus to not disrupt the folding and insertion of iRhoms or ADAM17 in the ER for secretion. Two different affinity tags were added to iRhoms and ADAM17 to allow tandem affinity chromatography for complex purification. The Twin-Strep tag was added to iRhoms because of the higher specificity of the resin (Strep-Tactin® XT) used to purify the tag (Dr. Justin Deme, personal communication). Both iRhoms and ADAM17 were placed under the CMV promoter. These constructs were cloned into second generation lentiviral vectors for lentivirus production to generate stable cell lines. The lentiviral system was a good choice to generate stable cell lines because it has been successfully used for structural studies of eukaryotic membrane proteins (Chang et al., 2016).

The source of mammalian and non-mammalian ADAM17 and iRhom cDNA is listed in Table 2.1 of Section 2.2.1. The pHR-CMV-TetO2 plasmids were a gift from A. Radu Aricescu (Elegheert et al., 2018) Table 2.2 of Section 2.2.1. PCRs and cloning was performed as described in Section 2.2. Mammalian and non-mammalian iRhoms were first subcloned into pHR-CMV-TetO2_3C-Twin-Strep for iRhom genes to contain a 3C cleavable Twin Strep (TwS) tag at the C-terminus. *Hs*ADAM17 was subcloned from pcDNA3-ADAM17-HA which was a gift from Axel Ullrich (Gschwind et al., 2003) Table 2.2 of Section 2.2.1. ADAM17 genes were subcloned into pHR-CMV-TetO2_3C-Avi-His6 to generate 3C-cleavable hexahistidine fusions at the C-terminus.

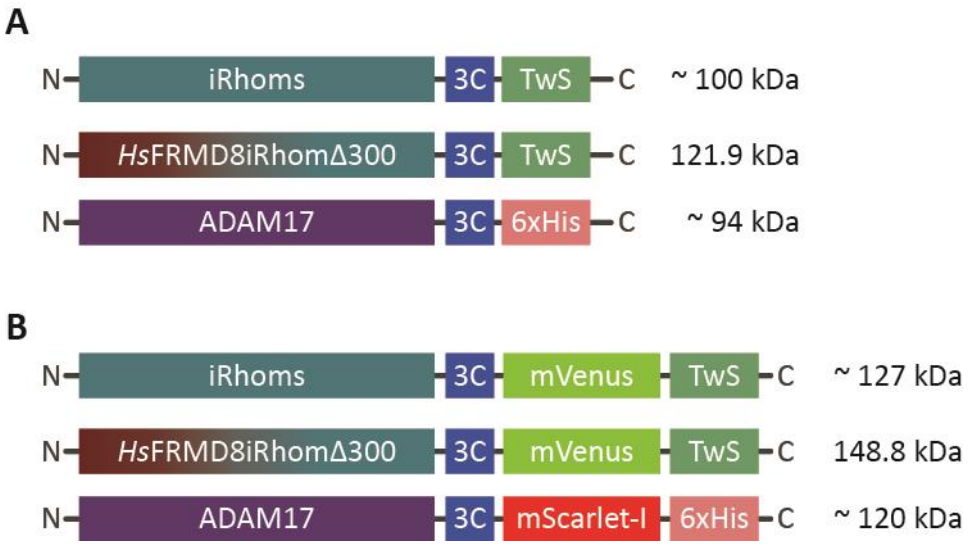


Figure 4.4. iRhom and ADAM17 construct designs for mammalian expression. (A) The iRhom constructs, mammalian and non-mammalian iRhoms, and the *Hs*FRMD8iRhom2 chimera were cloned in the pHR-TetO2-CMV vector to contain C-terminal cleavable Twin Strep tag. The ADAM17 constructs were cloned into the pHR-TetO2-CMV vector to contain a C-terminal cleavable hexahistidine tag. **(B)** The constructs shown in **(A)** were subcloned to contain cleavable fluorescent protein with an affinity tag. iRhoms and *Hs*FRMD8iRhom2 chimera constructs were fused with mVenus and ADAM17 constructs were fused with mScarlet-I.

4.3.2 Protocol for generating stable cell lines by lentiviral transduction

4.3.2.1 Plasmid preparation

Second-generation envelope plasmid, pMD2.G, and second-generation packaging plasmid, psPAX2, were used with the pHR-CMV-TetO2 transfer plasmids to generate lentivirus particles. The envelope and packaging plasmids were a gift from Didier Trono Table 2.2 of Section 2.2.1. Envelope and packaging plasmids were amplified and purified using HiSpeed Plasmid Maxi Kit (Qiagen) following the manufacturer's protocol. Plasmid yields were at least 500 ng/µL. The pHR-CMV-TetO2 transfer plasmids were amplified and purified using QiaPrep Spin Miniprep kit and yields generated were at least 500 ng/µL. Stable cell lines were generated using an adapted protocol from (Elegheert et al., 2018).

4.3.2.2 Transfection

Lentivirus producer cells, HEK293T Lenti-X (Takara/Clontech), were maintained as described in Section 2.3.2.2 and up until passage 12. HEK293T Lenti-X cells were seeded into T75 flask at an appropriate cell number to ensure that cells were at >90 % confluency on the day of transfection. Transfection plasmid mix was prepared at a 1:1:1 (wt/wt/wt) of transfer/packaging/envelope plasmids with a total of 30 μg of DNA in sterile 1.5 mL tubes. Plasmid DNA mix was topped up with DMEM/F-12/serum-free media (SFM)/MEM NEAA final volume of 0.25 mL. For the co-transfection of two different transfer plasmids the total amount of transfer plasmid DNA was divided into equal parts, 0.5:0.5:1:1 (transfer 1/transfer 2/packaging/envelope), so that total DNA did not exceed 30 μg. Transfection reagent solution was prepared by adding 75 μL of 1 mg/mL polyethylenimine (PEI) to sterile 1.5 mL tubes and topped up with DMEM/F-12/SFM to a total volume of 0.25 mL. The plasmid DNA mix and transfection reagent solution were combined to produce a final volume of 0.5 mL. Final DNA:PEI ratio was 1:2.5 (wt/wt). PEI/DNA mix was briefly vortexed to mix and briefly centrifuged to collect the liquid to the bottom of the tube. PEI/DNA mix was incubated in the biological safety cabinet for 20 minutes. During the incubation time, media of the >90 % confluent HEK293T Lenti-X cells was removed, washed once with PBS and replaced with 11.5 mL of DMEM/F-12/2 % FBS. Immediately before transfection, HEK293T Lenti-X flasks were turned upright and the DNA/PEI mix was added drop by drop to the cell medium to ensure that there were no concentrated areas of the mix. The flask was gently mixed in this upright position then gently tilted to cover the cells. Final volume of transfection was 12 mL. Flasks were incubated at 37 °C at 5 % CO2. One to two days after transfection, cells that were transfected with transfer plasmids containing a fluorescent protein were monitored by a bench epifluorescence microscope.

4.3.2.3 Transduction

Recipient cells for the infection, or transduction, of lentivirus were the FreeStyle™ 293-F cells (Thermo Fisher Scientific, Gibco) (HEK293F) that were maintained as adherent cell culture as described in Section 2.3.2.2. Recipient cells were seeded into T75 flask at an appropriate cell number to ensure that cells were 50-60 % confluent on the day that lentivirus conditioned medium was collected. The confluency of the recipient cells was lower than that recommended in (Elegheert et al., 2018) as the authors recommended a confluency of >90 %. Three days after transfection or when cell medium became acidic, the conditioned medium (~ 12 mL) from the HEK293T Lenti-X cells were transferred into sterile 50 mL tubes and centrifuged for 3 minutes at 500 *x g* to pellet detached cells and cell debris. Separate sterile 50 mL tubes were prepared with 6 mL of DMEM/F-12/10 % FBS medium and 18 µL of 10 mg/mL polybrene (Merck). Conditioned supernatant was filtered through 0.45 µm cellulose acetate syringe end filter (Millipore) into the DMEM/F-12/10 % FBS and polybrene mix to a total volume of 18 mL and a final concentration of 0.01 mg/mL polybrene. A small aliquot (~ 100 µL) of the filtered lentivirus and polybrene containing medium was retained for a p24 ELISA. Media of recipient cells was removed, cells washed in cell culture grade PBS and replaced with the filtered lentivirus and polybrene containing medium. Flasks were placed in a lentivirus dedicated incubator at 37 °C at 5 % CO_2. Two days after infection, cells that were transduced with fluorescent protein containing lentivirus were monitored by a bench epifluorescence microscope.

4.3.2.4 Determining lentiviral titre

The aliquot kept after the lentivirus conditioned medium was filtered into the polybrene containing medium was used to determine lentiviral titre by enzyme-linked immunosorbent assay (ELISA). Titre was determined using the HIV-1 p24 SimpleStep ELISA® Kit (Abcam) by following the manufacturer's protocol for the Reagent Preparation, Standard Preparation and Assay Procedure for an endpoint reading. The manufacturer's protocol for Sample Preparation

for cell culture supernatants was not followed because the sample was already centrifuged and filtered as part of the transduction protocol. Sample was serially diluted to 1 x 10^{-6} and 1 x 10^{-3}, 1 x 10^{-4}, 1 x 10^{-5} and 1 x 10^{-6} samples were assayed. Each sample and standard was assayed in duplicate. The absorbance was recorded at 450 nm using the Spectramax M5 plate reader (Molecular Devices), blanking the plate reader on the blank control. Absorbance of the standards was plotted using GraphPad Prism 7 and standard curve was generated using the linear regression model within the software. Samples were interpolated from the standard curve to determine the concentration of p24 in pg/mL. The physical titre of each sample was determined by this equation: Physical titre (TU/mL) = interpolated result (pg/mL) x 100 x dilution factor. Refer to Appendix D for the derivation of lentiviral particles. The interpolations that gave a positive result were used in calculating the physical titre. Dilutions that came from the same sample were averaged to give the final physical lentiviral titre.

4.3.2.5 Recovery

Three days after the target/recipient cell line was infected with lentivirus, flasks were washed at least three times in PBS and split into two T75 flasks. Cells were incubated in a normal 37 °C standing incubator at 5 % CO_2 until confluent. This stage took approximately 7 days for cells to become confluent after their first split. Cells that began to detach during the PBS washing stage were not split into two flasks. The cells that remained attached were trypsinised as described in Section 2.3.2.2 and all cells were reseeded into T75 with fresh DMEM/F-12/10 % FBS. After these cells reached confluency, cells were split as described in Section 2.3.2.2 into two T75. One of the two T75 flasks generated for each cell line was frozen as described in Section 2.3.2.2.

4.3.3 Expression of mammalian iRhoms in FreeStyle™ 293F cell lines

FreeStyle™ 293F (HEK293F) suspension cells were used for stable cell line generation rather than HEK293T because HEK293F cells can reach higher cell densities than the adherent HEK293T cells.

The higher cell density is important for membrane proteins because more cells will result in more membrane. HEK293F cells were initially transduced with lentivirus as described above to generate three stable cell lines overexpressing *Hs*iRhom2, *Hs*FRMD8iRhom2Δ300 (*Hs*FRMD8iRhom2 chimera) and Emerald Green Fluorescent Protein (EmGFP) as a transfection and transduction control. EmGFP control cell line was generated in parallel with the other cell lines to track transfection and transduction efficiency by monitoring EmGFP fluorescence.

Expression of *Hs*iRhom2 and *Hs*FRMD8iRhom2 in the transduced HEK293F cell lines were analysed by western blot (Figure 4.5A) by detection of the Twin-Strep tag. There was no signal for the *Hs*iRhom2 cell line and signal for *Hs*FRM8iRhom2 chimera. Immunofluorescence was performed on the *Hs*iRhom2 and *Hs*FRMD8iRhom2 chimera (Figure 4.5B) whereby cells were stained for the Twin-Strep tag, ER marker BAP31 and DAPI (performed by Dr. Nina Jajčanin Jozić). Membrane isolation of the *Hs*FRM8iRhom2 chimera was attempted; however, no detectable levels of protein were observed by western blotting (data not shown). The lack of protein in the membrane isolation steps was likely due to the low proportion of cells expressing the transgene.

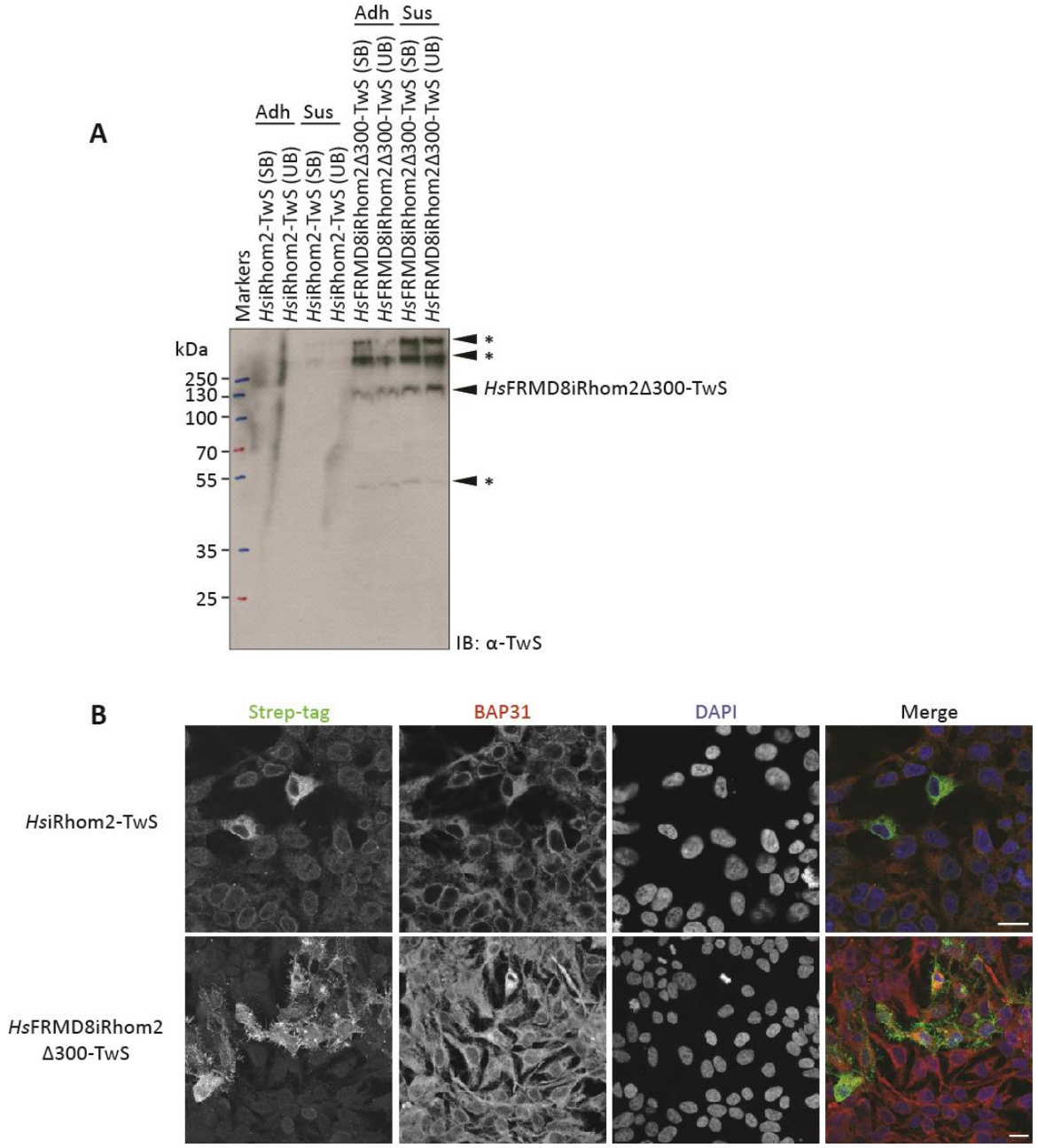


Figure 4.5. Stable lentiviral cell line generation. In the first trial of stable cell line generation in HEK293F cells, *Hs*iRhom2 and *Hs*FRMD8iRhom chimera were tested. After lentiviral transduction whole cell lysates were analysed by western blot probing for the Twin Strep tag **(A)**. Only the *Hs*FRMD8iRhom chimera appeared to be expressed in the western blot and larger, aggregated strep reactive bands (*) that did not enter the resolving gel. Generated HEK293F cell lines were cultured in both adherent (Adh) and suspension (Sus) culture. Cell pellets were lysed in loading sample buffer (SB) or urea buffer (UB). **(B)** *Hs*iRhom2 and *Hs*FRMDiRhom2 chimera HEK293F cells were stained for the Twin Strep tag (green channel), BAP31 (red channel) and nucleus staining by DAPI (blue channel). Scale bar represents 20 µm.

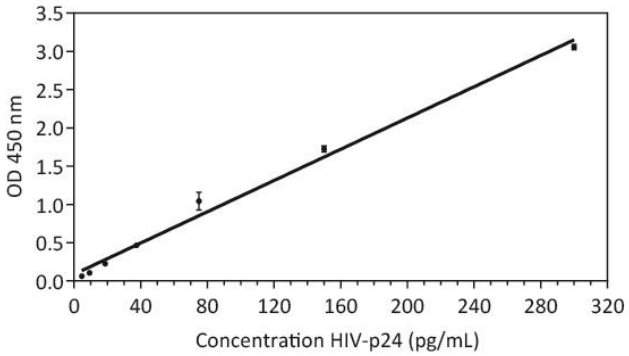
4.3.4 Expression of mammalian iRhoms with the addition of fluorescent proteins

Due to low transduction efficiency, constructs were subcloned with a cleavable C-terminal fluorescent protein and affinity tags (Figure 4.4B). The reasoning for fusing the fluorescent protein and tags at the C-terminus was for similar reasons explained above (Section 4.3.1), to avoid disrupting the folding and insertion of membrane proteins in the ER for secretion. This was particularly important because the fluorescent proteins are large folded tags which may impair correct folding of the protein of interest if the tags were located at the N-terminus. Fortunately the C-terminus of iRhom and ADAM17 are in the ER and cytosol respectively therefore the two fluorescent tags should not interact with one another. Fluorescent proteins are useful for three reasons: first, they aid in the cell generation procedure because stages can be monitored by simple epifluorescent microscopes and cell lines expressing fluorescent protein can be enriched by fluorescence-activated cell sorting (FACS). Finally, fluorescent proteins are utilised for determining optimal solubilising conditions for membrane proteins, in a technique called fluorescent detection size-exclusion chromatography which is further discussed in Chapter 5. Green/yellow fluorescent protein mVenus (mVn) was fused to the iRhom constructs and the red fluorescent protein mScarlet-I (mSc) was fused to the ADAM17 constructs. The iRhom genes were subcloned in the pHR-CMV-TetO2_3C-mVenus-Twin-Strep plasmid and the Avi Tag in the ADAM17 plasmids was replaced by mScarlet-I using Gibson Assembly. Gene insertion into plasmids was confirmed by colony PCR using the appropriate Gibson Assembly primers. These colonies were inoculated in LB media for plasmid amplification, purified and sequenced as described in Section 2.2.

Due to the poor transduction of the target cell lines, a p24 ELISA was performed on the lentiviral conditioned media that was used to treat the cells. Physical lentiviral titre, transducing units/mL (TU/mL) was interpolated using a standard curve that was generated at the time of the ELISA. An example of standard curve and the interpolations are shown in Figure 4.6. Cells were

sequentially transduced and enriched to produce cell lines that were expressing both iRhom-mVenus and ADAM17-mScarlet-I.

A



B

Lentivirus	Titre (TU/mL)
*Hs*iRhom1-mV	1.72E+06
*Hs*iRhom2-mV	4.19E+06
*Am*iRhom1-mV	1.62E+06
*Dr*iRhom1-mV	1.58E+06
EmGFP	5.52E+07

Figure 4.6. Interpolations of lentiviral physical titre. (A) A representative of the standard curve generated from the HIV-1 p24 SimpleStep ELISA® Kit (Abcam). Optical density (OD) at 450 nm was plotted against the concentration of HIV-p24 in pg/mL in GraphPad Prism. Standard curve was generated by linear regression. OD 450 nm of samples were recorded and concentrations were interpolated using the standard that was generated at the time of the ELISA. **(B)** Physical lentiviral titre (Transducing Units (TU)/mL) was determined using this formula: interpolated result (pg/mL) x 100 x dilution factor.

4.3.4.1 Flow cytometry of generated cell lines

After cells recovered from transduction, the population distributions of transduced and untransduced cells were measured using cell flow cytometry analysers.

4.3.4.2 Population distribution of generated cell lines

After cells were split post recovery from transduction, the remaining cell suspension in DMEM/F-12/10 % FBS was used for the flow cytometry analysers. Wild type cells that were not infected with lentivirus were used as negative controls for analysis. For cell lines expressing iRhom-mVenus the Cytek DxP analyser (Cytek) was used whereas cell lines coexpressing iRhom-mVenus and ADAM17-mScarlet-I the Cytoflex LX (Beckman Coulter) analyser was used. The parameters used for flow cytometry analysers are shown in Table 4.1.

Table 4.1. Parameters used for flow cytometry analysers.

Fluorescent protein or dye	Wavelength (nm)	Band pass filter
Cytek DxP:		
mVenus	488	530/30
DAPI	405	445/50
Cytoflex LX:		
mVenus	488	525/40
mScarlet-I	561	610/20
DAPI	405	450/45

Negative control cells were first run without 4',6-diamidino-2-phenylindole (DAPI) treatment followed by DAPI treatment. DAPI was diluted from a stock to 10 µg/mL, 5 µl of diluted DAPI was added per 1 mL of cell suspension and incubated at room temperature for 5 minutes. After running the negative control cells with and without DAPI, the following gating strategies were used: cells were gated on side scatter area vs forward scatter area, singlets were gated on forward scatter width vs forward scatter area. Live cells were gated on 405 nm vs forward scatter area. Live single cells were plotted on a histogram against fluorescence. Where two fluorescent proteins were analysed live single cells were plotted red (mScarlet-I) fluorescence vs green fluorescence (mVenus).

Examples of population distributions are shown in Figure 4.7 of single (Figure 4.7A) and double transduced (Figure 4.7B-E). Varying population differences existed between cells positive, double positive or negative for fluorescence. The difference in populations highlighted the need to enrich the single positive for iRhom alone and double positive cells for iRhom and ADAM17. Transduced cells were sorted by FACS to produce enriched polyclonal cell lines for single positive cells for wild type iRhoms-mVn and *Hs*FRMD8iRhom2-mVn chimera and double positive cells for iRhom and ADAM17. Cells were sorted based on fluorescence levels: medium, high and very high in the case of *Hs*iRhom1.

4.3.4.3 Cell sorting enrichment

Single positive cells for cells only expressing iRhom-mVenus and double positive cells expressing iRhom-mVenus and ADAM17-mScarlet-I were enriched by fluorescence-activated cell sorting (FACS) performed by the Dunn School Flow Cytometry Facility as described in Section 2.7.2.

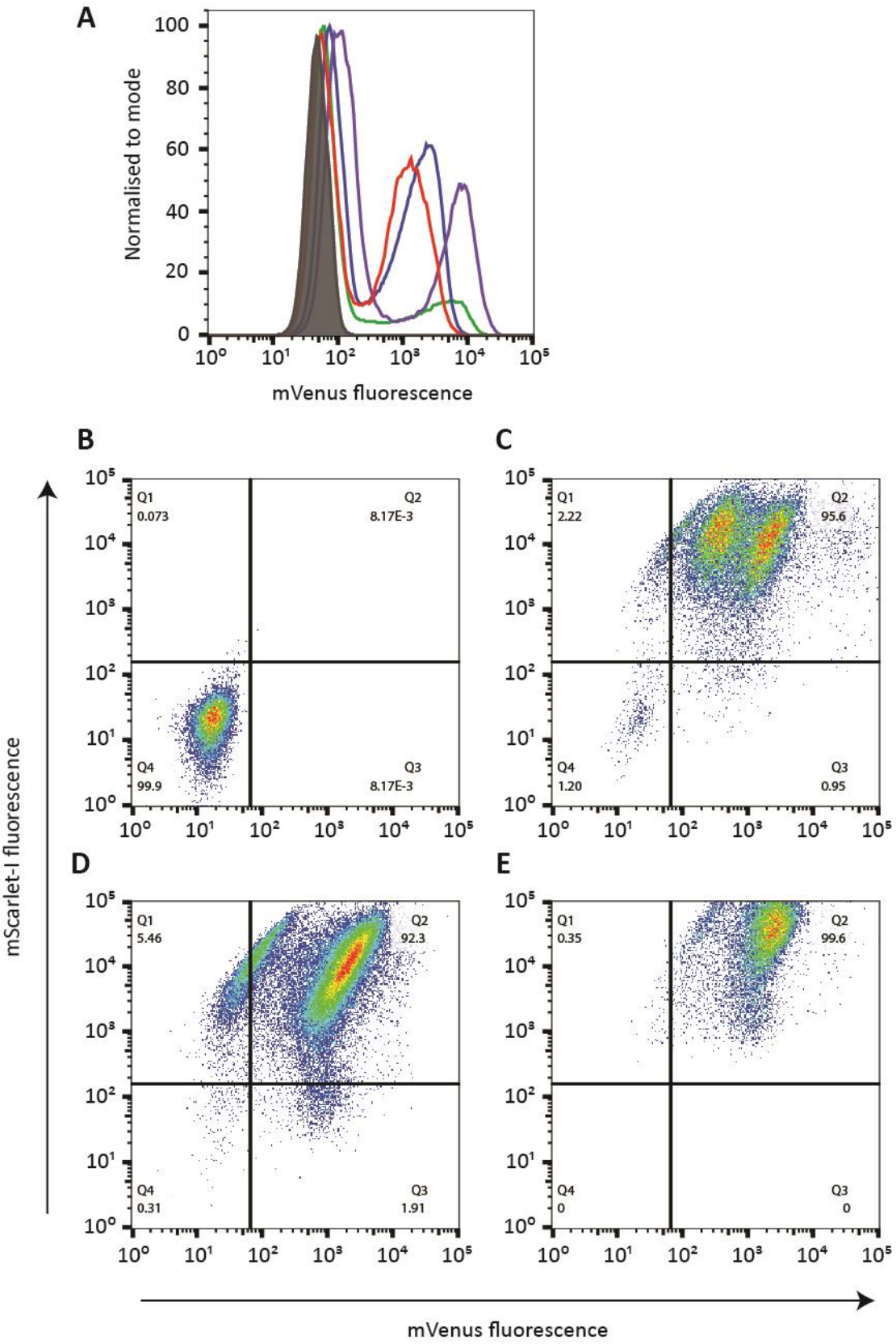

A
Normalised to mode
mVenus fluorescence
B
C
D
E
mScarlet-I fluorescence
Q1
0.073
Q2
8.17E-3
Q4
99.9
Q3
8.17E-3
Q1
2.22
Q2
95.6
Q4
1.20
Q3
0.95
Q1
5.46
Q2
92.3
Q4
0.31
Q3
1.91
Q1
0.35
Q2
99.6
Q4
0
Q3
0
mVenus fluorescence

Figure 4.7. Population distributions of generated cell lines. (A) Population distribution HEK293F cells transduced with iRhom-mVn constructs. Cells were gated on single, live cells and the distribution of fluorescence of these cells is displayed as a histogram. The negative control was untransduced HEK293F cells (grey trace). Cell lines shown were *Hs*iRhom2-mVn (red trace), *Hs*FRMDiRhom2-mVn chimera (purple trace), *Mm*iRhom2-mVn (blue trace) and *Hd*iRhom-mVn (green trace). **(B-E)** Population distributions of HEK293F iRhom cells transduced with ADAM17-mSc. Cells were gated on single, live cells and the distribution of fluorescence of these cells is displayed as mScarlet-I fluorescence against mVenus fluorescence. The negative control **(B)** was untransduced HEK293F cells. Cell lines shown *Hs*iRhom1-mVn/*Hs*ADAM17-mS **(C)**, *Hs*iRhom2-mVn/*Hs*ADAM17-mSc **(D)** and *Mm*iRhom2-mVn/*Mm*ADAM17-mSc **(E)**.

4.3.5 Small scale purification of iRhom constructs

Before FACS enrichment of transduced HEK293F cells, with *Hs*iRhom2-mVn, *Hs*FRMD8iRhom2Δ300-mVn, *Hd*iRhom-mVn and *Mm*iRhom2-mVn, were tested for transgene expression. Remaining cell suspension from splitting the mixed positive/negative polyclonal cell line by trypsinisation, from a confluent T75 flask, were harvested by centrifugation at 500 *x g* for 3 minutes. Cells were lysed in Solubilisation Buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 1 % DDM, 0.1 % CHS, pH 8.0) to a total volume of approximately 5 times the cell pellet weight. Cell lysates were incubated in Solubilisation Buffer for three hours at 4 °C on a rotary wheel. Cell lysates were centrifuged for 10 minutes at 21,000 *x g* using a refrigerated benchtop microfuge cooled to 4 °C. Supernatants were centrifuged at 100,000 *x g* for 30 minutes at 4 °C to pellet insoluble material. MagStrep “type3” XT beads (IBA Life Sciences) were washed three times in Wash Buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 0.03 % DDM, 0.003 % CHS, pH 8.0) and beads separated out using a magnetic stand. Supernatant was added to the washed beads and incubated at 4 °C on a rotary wheel for 40 minutes. Supernatant was removed and beads washed three times in Wash Buffer. Proteins were eluted in Elution Buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 0.03 % DDM, 0.003 % CHS, 50 mM biotin, pH 8.0). All stages were analysed by western blot as described in Section 2.3.5 using the StepMAB classic antibody (Figure 4.8). As expected, there was no detectable protein in any of the lanes except for elution lanes because of the mixed population of untransduced and transduced cells and that the

positive cells must have been below the detection limit of the western blot. These results further highlight the need of enriching for the positively transduced cells. The Twin Strep reactive bands at ~ 100 kDa in the elution lanes of *Hs*iRhom2, *Hs*FRMD8iRhom2 chimera and *Mm*iRhom2 (Figure 4.8) were slightly lower than their predicted molecular weight (Table 4.2). The lower apparent molecular weight may due to small cleavage at the N-terminal side. Whereas the Twin Strep reactive band at ~ 200 kDa in elution lane of *Hd*iRhom was higher than the predicted molecular weight of *Hd*iRhom (119.8 kDa). The higher molecular weight of *Hd*iRhom on the immunoblot may be due to posttranslational modifications resulting in different migration or the protein running slower in the SDS-PAGE due to the SDS binding more to the TMHs, which is a common artefact for membrane proteins (Dr. Justin Deme, personal communication).

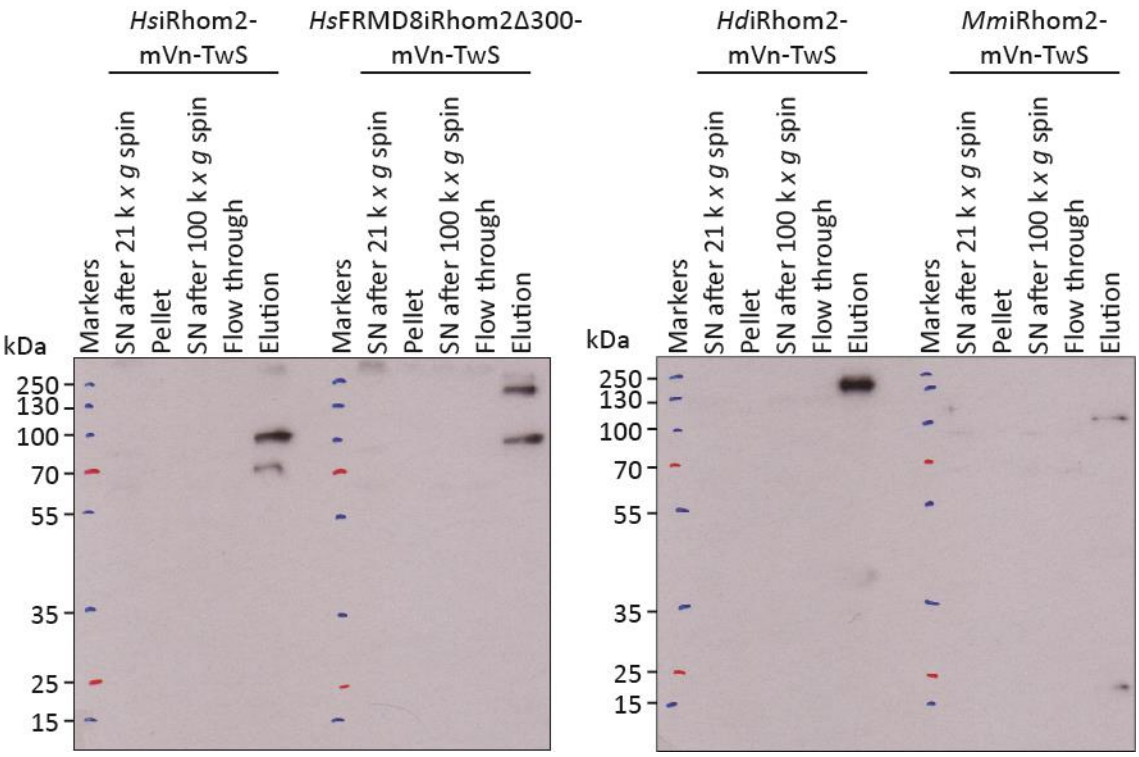


Figure 4.8. Small scale purification of polyclonal cell lines transduced with iRhom genes. Cells, prior to cell sorting enrichment, were tested for expression by affinity purifying cell lysates utilising the Twin Strep tag and eluting with biotin. Stages of purification were analysed by western blot by detection of the Twin Strep tag. Abbreviation used: SN, supernatant.

Table 4.2. Predicted molecular weights of iRhoms and ADAM17.

Protein	Molecular weight (kDa)
*Hs*FRMD8iRhom2Δ300-mVn-TwS	145.8
*Hs*iRhom1-mVn-TwS	128.5
*Hs*iRhom2-mVn-TwS	127.8
*Mm*iRhom2-mVn-TwS	124.5
*Hd*iRhom-mVn-TwS	119.8
*Hs*ADAM17-mSc-His:	
Immature	119.8
Mature	96.9
*Mm*ADAM17-mSc-His:	
Immature	120.9
Mature	98.2

4.3.6 Coexpression of iRhom and ADAM17 in generated cell lines

Expression of iRhom and ADAM17 in the enriched double transduced cell lines were confirmed by western blot (Figure 4.9) by detection of the Twin-Strep tag on the iRhoms (Figure 4.9A) and of the hexahistidine tag on ADAM17s (Figure 4.9B). 2 x 10^6 cells were seeded onto poly-L-lysine coated 100 mm dishes, for adherent cell culture, and incubated in a standing 37 °C incubator with 5 % CO_2 for 48 hours. On ice, the cells were gently washed in PBS three times then lysed in 1 mL Cell Lysis Buffer (50 mM Tris HCl, 150 mM NaCl, 1 % Triton X-100, pH 7.5). Cells were lysed on ice for 10 minutes then scraped using a cell scraper, to ensure maximal recovery, and transferred to 1.5 mL tubes. Cell lysates were centrifuged for 10 minutes at 20,000 *x g* using a refrigerated benchtop microfuge cooled to 4°C. Supernatants were transferred into fresh 1.5 mL tubes. Gel samples were prepared by adding 2X NuPAGE™LDS Sample Buffer (Thermo Fisher Scientific, Invitrogen) to a final concentration of 1X and heated at 65 °C for 10 minutes. Samples were loaded onto NuPAGE™ Bis-Tris 4-12 % gels and proteins separated by electrophoresis using Bis-Tris running buffer. Proteins were transferred onto methanol-activated PVDF membranes using Bis-Tris transfer buffer. Transferred proteins were verified by Ponceau S staining of the membranes and washed off before blocking. Western blots were performed and developed as described in Section 2.3.5 using the StrepMAB classic (IBA Lifesciences) and Tetra·His antibody

(Qiagen) to detect iRhom-mVenus-TwS and ADAM17-mScarlet-I-His respectively. In the case of the very highly expressing *Hs*iRhom1 cell line there was no *Hs*iRhom1, just free mVenus (~ 27 kDa) with a Twin-Strep tag (Figure 4.9A). *Hs*iRhom1/*Hs*ADAM17 and *Hs*iRhom2/*Hs*ADAM17 cells expressed both iRhom and immature ADAM17 whereas in the *Mm*iRhom2/*Mm*ADAM17 cell line, *Mm*ADAM17 is processed because (Figure 4.9B) both immature (~ 120 kDa) and mature (~ 90 kDa) *Mm*ADAM17 is shown. Non-specific bands are highlighted by the asterisk (*) in Figure 4.9 because untransduced HEK293F cells (HEK293F wt, Figure 4.9) served as a control for antibody binding.

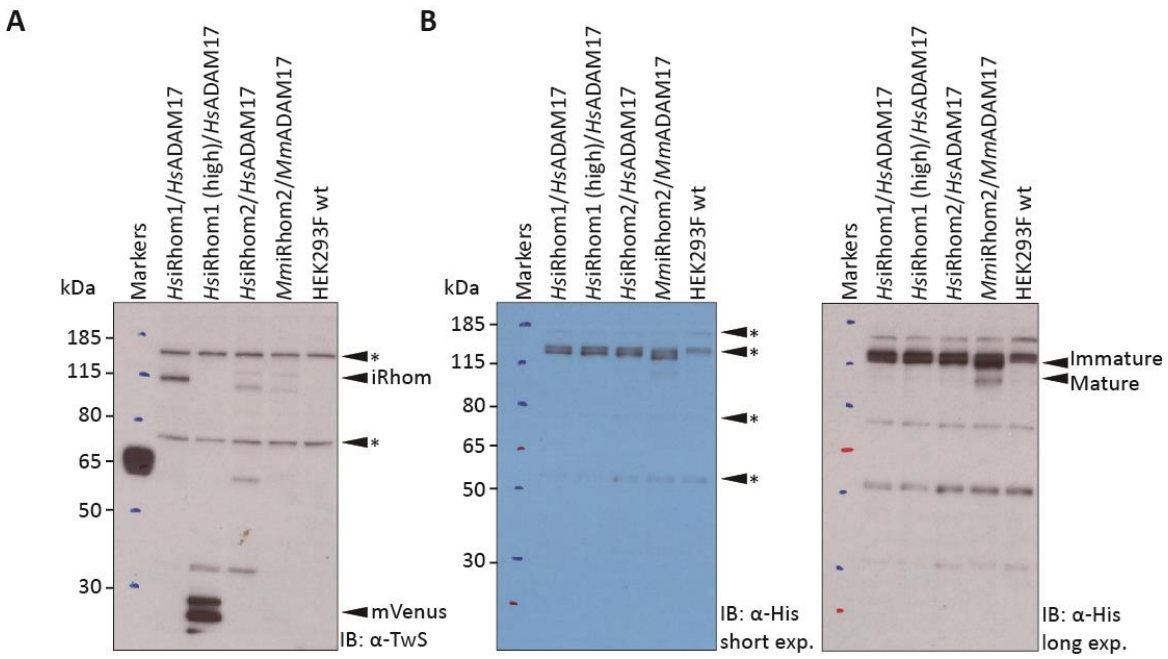


Figure 4.9. Expression testing of iRhom and ADAM17 in double transduced cell lines. Whole cell lysates of *Hs*iRhom1-mVn/*Hs*ADAM17-mSc, *Hs*iRhom1-mVn(high)/*Hs*ADAM17-mSc, *Hs*iRhom2-mV/*Hs*ADAM17-mSc, *Mm*iRhom2-mVn/*Hs*ADAM17-mSc and HEK293F parental wt cells were analysed by western blot. iRhom was probed against the Twin Strep tag **(A)** and ADAM17 was probed against the His tag **(B)**. Long exposure of the left panel of **(B)** is shown on the right hand side. Non-specific bands are highlighted by the asterisk (*).

4.3.7 Sub cellular localisation of generated HEK293F cell lines

Sub cellular localisation of exogenous, overexpressed iRhom and ADAM17 were demonstrated by either live cell imaging or immunofluorescence staining. Cells only expressing iRhoms were subject to live cell imaging (data not shown) whereas the double transduced cells were stained (Figure 4.10).

4.3.7.1 Live cell imaging of iRhom alone cell lines

Cells were seeded and imaged as described in Section 2.6.3.1. HEK293F cells expressing wildtype iRhoms: *Hs*iRhom2, *Mm*iRhom2, *Hd*iRhom (data not shown) were localised in the ER where as *Hs*FRMD8iRhom2 chimera was at the cell surface (data not shown). The live imaging data of the *Hs*FRMD8iRhom2 chimera corresponds to published data of this construct being expressed on the cell surface of HEK cells (Kunzel et al., 2018).

4.3.7.2 Immunofluorescence staining of mammalian iRhom and ADAM17 cell lines

Immunofluorescence staining of iRhom/ADAM17 cells is shown in Figure 4.11. Cells were stained and imaged as described in Sections 2.6.2 and 2.6.3.2 respectively. For *Hs*iRhom1/*Hs*ADAM17 and *Hs*iRhom2/*Hs*ADAM17 cells both iRhoms and ADAM17 were expressed in the ER and co-localised. ADAM17 was only seen in the ER therefore remains in its immature form, consistent with our immunoblotting results (Figure 4.9) However the staining of *Mm*iRhom2/*Mm*ADAM17 show that iRhom2 and ADAM17 were at the cell surface which corroborates with western blot data in Figure 4.9B where mature ADAM17 is present.

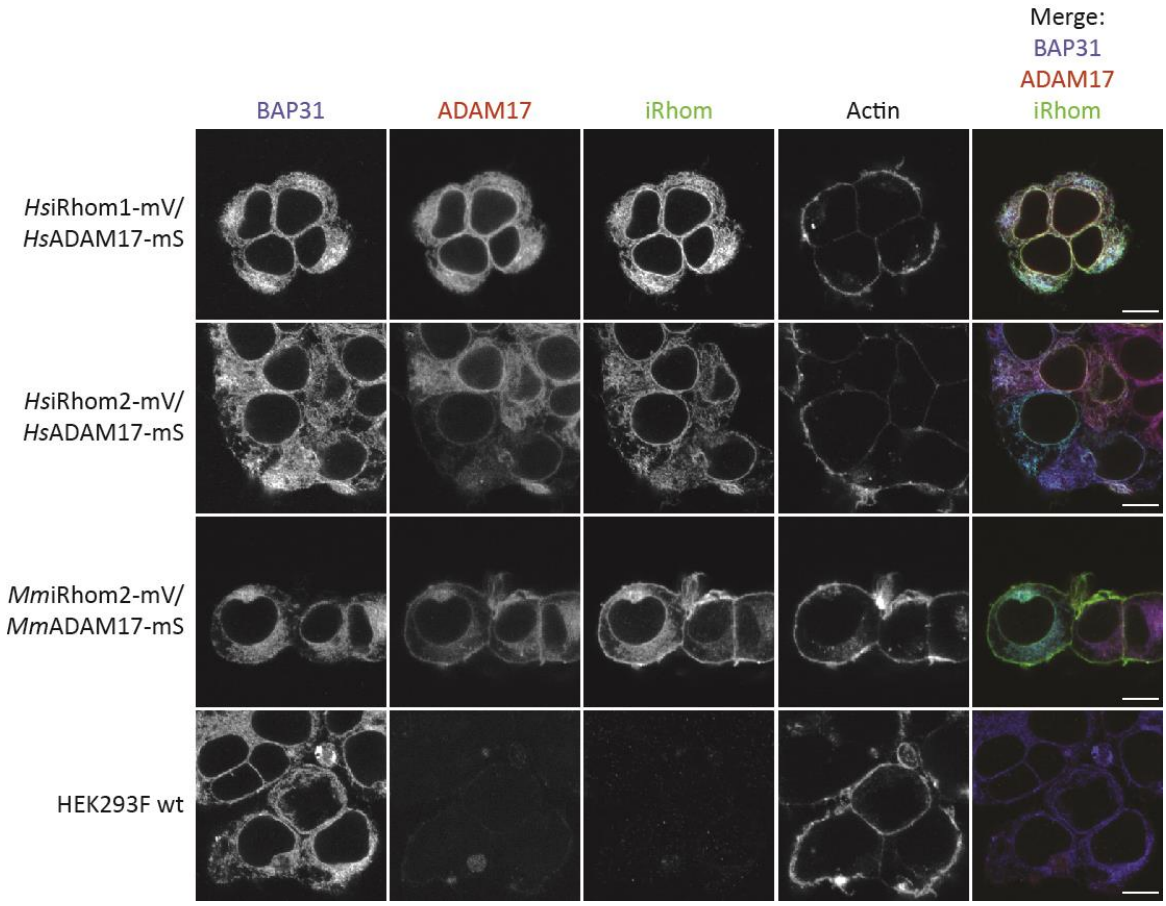


Figure 4.10. Immunofluorescence staining of double transduced HEK293F cells with iRhom and ADAM17. Cells were fixed and stained for BAP31, an endoplasmic reticulum marker, (blue channel), iRhom using α-GFP antibody against the mVenus (green channel), and actin, a cell edge/surface marker, (far red channel). Endogenous fluorescence of mScarlet-I on ADAM17 was imaged (red channel). Merge images are of the blue (BAP31), red (ADAM17) and green (mVenus on iRhom) channels. Colocalisation of all three channels is represented by white/grey, blue and green channel colocalisation is represented by cyan, blue and red channel colocalisation is represented by magenta and green and red channel colocalisation is represented by yellow. Scale bar represents 10 µm.

4.4 Discussion

4.4.1 Initial experiments using the *Hs*iRhom2-3xHA cell lines

Endogenously HA tagged *Hs*iRhom2 was not detectable in whole cell lysates. Therefore iRhom2 is not natively expressed at a high level in HEK293T cells. However, endogenously tagging iRhom1 in HEK293T cells may have higher expression levels because iRhom1 is broadly expressed in mammalian tissues. Overexpressed *Hs*iRhom2-3xHA could be detectable by

western blotting following pulldown on the HA tag however the number of cells required for harvest would not have been feasible. In addition, the HA tag was not a good candidate for affinity purifications and therefore difficult to purify a high enough yield of *Hs*iRhom2 for structural studies. The HA tag was poor for large scale purifications because of poor elution efficiency by HA peptide in membranes isolated from cells overexpressing *Hs*iRhom2. In the *Hs*iRhom2 overexpression cell line, full length iRhom was not solubilised in DDM/CHS from purified membrane whereas a smaller (~ 50 kDa) degradation product was solubilised. Therefore, the degradation product appears to be more stable than full length *Hs*iRhom2 in the DDM/CHS solubilising condition. Because both full length *Hs*iRhom2 and the degradation band are present in the whole cell lysates and isolated membranes, full length *Hs*iRhom2 is not being degraded *in vitro* post cell harvesting. Generating large volumes of HEK293T cells for membrane isolation/protein purification in adherent culture was time consuming and impractical; therefore expanding cells in this way was not a feasible strategy.

Due to the difficulty of using the HA tag for affinity purifications and the difficulty culturing adherent cells for large scale membrane preparations, the constructs were redesigned and a different mammalian expression method adopted.

4.4.2 Fluorescent tags aid in cell enrichment and poor transduction efficiency

Mammalian and non-mammalian iRhoms and ADAM17 genes were cloned into second generation lentiviral vectors. Mammalian constructs were still used because of the aim to understand iRhom and ADAM17 in a physiologically relevant context. The non-mammalian homologues were chosen to broaden the possibility that these constructs may express and or purify better. The tardigrade is the only eukaryotic extremophile and proteins from extremophiles tend to be more stable. Further, chickens have body temperatures of 42 °C, suggesting increased protein thermo-stabilisation compared to humans or mice.

The first lentiviral expression trial concluded that the transduction efficiency was low because a low proportion of cells were expressing *Hs*iRhom2 or *Hs*FRMD8iRhom2Δ300 based on immunofluorescence (Figure 4.5). Therefore, these constructs were subcloned to include a fluorescent protein which enabled enrichment via FACS. The cell sorting also showed the population distributions of each cell line after lentiviral transductions. Cells lines had varying population distributions where certain cell lines had more positive cells for fluorescence than negative. From the example shown in Figure 4.7A, mammalian iRhom cell lines had a higher proportion of positive cells than *Hd*iRhom-mVn. The variations may be due to a number of reasons: poor lentiviral titre due to poor packaging of the virus and/or low transduction efficiency where the gene was not integrated into the genome efficiently.

Lentiviral titre was determined using a HIV-p24 ELISA. As part of the protocol, the entire conditioned medium from the HEK293T Lenti-X cells was applied to the target cell line. But a small aliquot of the conditioned media was assayed for p24 just as a crude check to approximate the physical titre. A draw back of the HIV-p24 ELISA is that there is detection of free p24 in the medium and not just p24 associated with the lentiviral particles. Therefore, the physical titre determined is an over estimation but still is informative to refer back to in the case of poor gene integration into the genome of the recipient cell line.

Poor packaging of the transgene into lentiviral particles may be due to the length between the two flanking long terminal repeat (LTR) sequences. The shorter the length of the transgene the higher the viral titre because the ease of packaging shorter genetic material into lentiviral particles (Kumar et al., 2001). In the example shown in Figure 4.6B, shows the titre of the EmGFP cell line was 10 fold higher than iRhom cell lines. The average length of iRhom fused with mVenus and a Twin-Strep tag across the different species was 3.5 kb and the total average

length of DNA between the LTR sequences was 7.4 kb. Whereas length of EmGFP alone and the total length between the LTR sequence was 0.7 kb and 5.5 kb respectively. The size between the transgenes provides an explanation of the difference in titre.

An alternative to using fluorescent proteins was to use antibiotic resistant genes and to select transduced cells with antibiotic. The replacement of fluorescent proteins with antibiotic genes would have kept the DNA length between the LTR sequences similar and thereby would be unlikely to improve the lentiviral titre. There were several reasons to using fluorescent proteins. To start, second generation transfer vectors used already contained mVenus with a Twin-Strep tag therefore it was relatively easy to subclone constructs into this vector and the ease of using FACS to enrich the cell lines. Another reason was recombinant proteins fused with a fluorescent protein is particularly useful for membrane proteins. This is because the fluorescent protein is utilised in fluorescent detection size-exclusion chromatography, which is discussed in Chapter 5.

Chapter 5: Detergent screening and purification of cell lines

5.1 Introduction

In the previous Chapter, mammalian iRhom and ADAM cell lines were generated, expression was verified, and subcellular localisation established. Because these constructs had not been previously purified before, it was necessary to establish the optimal detergent for protein extraction out of the membrane and, for the iRhom/ADAM17 complex, to keep the complex intact. This Chapter will describe and discuss the detergent optimisation screening and subsequent purifications performed for these constructs.

The detergents used for extracting membrane proteins used in this Chapter for structural determination were: n-decyl-β-D-maltopyranoside (DM), n-dodecyl-β-D-maltoside (DDM), glyco-diosgenin (GDN) and lauryl maltose neopentyl glycol (LMNG). The structure of these detergents are shown in Figure 5.1A-D and classed as non-ionic detergents because they contain uncharged hydrophilic maltose head groups. Because of the uncharged head groups, non-ionic detergents are considered as mild solubilisers, non-denaturing and tend to break lipid-lipid and lipid-protein interactions rather than protein-protein interactions (le Maire et al., 2000, Seddon et al., 2004). Not disrupting protein-protein interactions is important when purifying protein complexes. The head groups of DDM is similar to DM but the tail consists of a longer alkyl chain by two carbons compared to DM (Seddon et al., 2004). LMNG is made up of two linked hydrophobic alkyl chains of equal length to two maltose groups (Chae et al., 2010, Breyton et al., 2019) and has been shown to increase protein stability (Chae et al., 2010). GDN is a synthetic drop-in substitute for Digitonin where, instead of saturated alkyl chains, the hydrophobic group consists of a rigid steroid-based group (Chae et al., 2012). These non-ionic detergents were chosen because of their general properties and that DDM (Jungnickel et al., 2018, Parker et al., 2019, Wright and Lee, 2019), LMNG (Rollauer et al., 2012, Kuhlen et al., 2018, Gao et al., 2019) and GDN (Wojnowska et al., 2018) have successfully been used to purify membrane proteins. However, in

the case of mammalian membrane proteins, such as GPCRs (Deshpande et al., 2019, Kato et al., 2019, Dong et al., 2020, Lin et al., 2020), transporters (Bräuer et al., 2019, Chi et al., 2020), GABA receptors (Masiulis et al., 2019, Park et al., 2020, Papasergi-Scott et al., 2020, Shaye et al., 2020) and glutamate receptors (Kumari et al., 2019) were purified in DDM or LMNG with cholesteryl hemisuccinate (CHS) (Figure 5.1E) which is a more water soluble analogue of cholesterol. Cholesterol is an important component of mammalian cell membranes and is important for the stability of some membrane proteins, particularly GPCRs (Zocher et al., 2012). CHS is therefore used with detergents to produce mixed micelles that better mimic mammalian cell membranes and to aid in membrane protein stability (Seddon et al., 2004, Kulig et al., 2014).

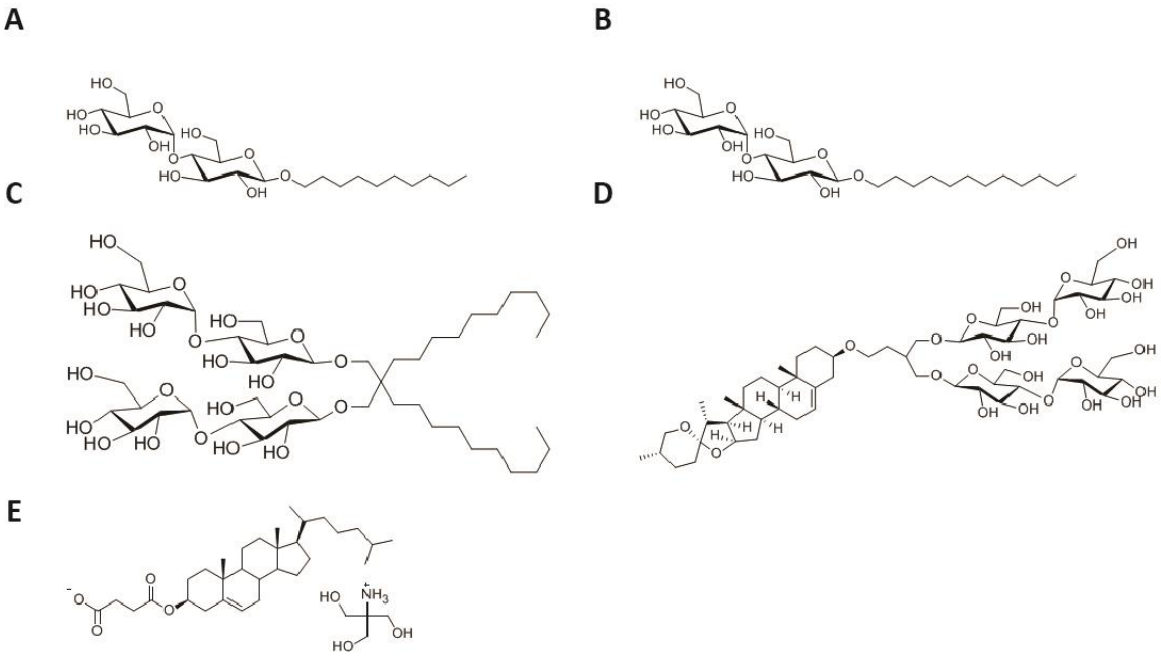


Figure 5.1. Chemical structures of commonly used membrane protein detergents. (A) n-decyl-β-D-maltopyranoside (DM), **(B)** n-dodecyl-β-D-maltoside (DDM), **(C)** lauryl maltose neopentyl glycol (LMNG), **(D)** glyco-diosgenin (GDN) and **(E)** cholesteryl hemisuccinate (CHS).

5.2 Membrane isolation

Expanded cell lines were subject to membrane isolation prior to membrane protein purification as described in Section 2.4.3. Isolating membranes is a purification step because it removes soluble proteins. Each step of the membrane isolation was visualised by immunoblotting when isolating the membranes for the first time. Stages of membrane isolation from three different cells lines, *Hs*iRhom2-mVn/*Hs*ADAM17-mSc, *Mm*iRhom2-mVn/*Mm*ADAM17-mSc and *Hs*FRMD8iRhom2Δ300-mV, are shown in Figure 5.2. For the doubly transduced cell lines (Figure 5.2A and B), iRhom was probed against the Twin-Strep tag and ADAM17 was probed against the hexahistidine tag. In each of the examples shown in Figure 5.2 (lanes 6), expected signal was obtained from resuspended membranes. *Hs*iRhom2, *Mm*iRhom2 and *Hs*FRMD8iRhom2 chimera, western blot bands were ~120 kDa, ~120 kDa and ~130 kDa respectively, similar to their predicted molecular weights (Table 4.2 in Chapter 4). For the ADAM17 blots, in both cases, the developed signal was higher than their predicted molecular weights (Table 4.2 in Chapter 4) where *Hs*ADAM17 and *Mm*ADAM17 were between the 130 and 250 kDa markers. Only one strong band can be seen in lane 6 of each cell line of the ADAM17 blots therefore only the immature form of ADAM17 is present. Only seeing immature ADAM17 for *Hs*iRhom2/*Hs*ADAM17 cell was to be expected because only the immature form was present in the expression testing and immunofluorescence data in Chapter 4. The expression testing and immunofluorescence data of the *Mm*iRhom2/*Mm*ADAM17 cell line in Chapter 4, both immature (predicted molecular weight: 120.0 kDa) and mature (predicted molecular weight: 98.2 kDa) ADAM17 were both present. Only the immature form of *Mm*ADAM17 is seen in Figure 5.2 therefore it appears that the mature form is unstable when the cell are processed for membrane isolation. Whereas the whole cell lysates shown in Figure 4.9, were lysed and analysed fresh straight from culture. This may be due the expanded cell pellets being stored at -20 °C after flash freezing with liquid nitrogen. The loss of mature *Mm*ADAM17 means that the pool of *Mm*ADAM17 is homogenous because only the immature form is present.

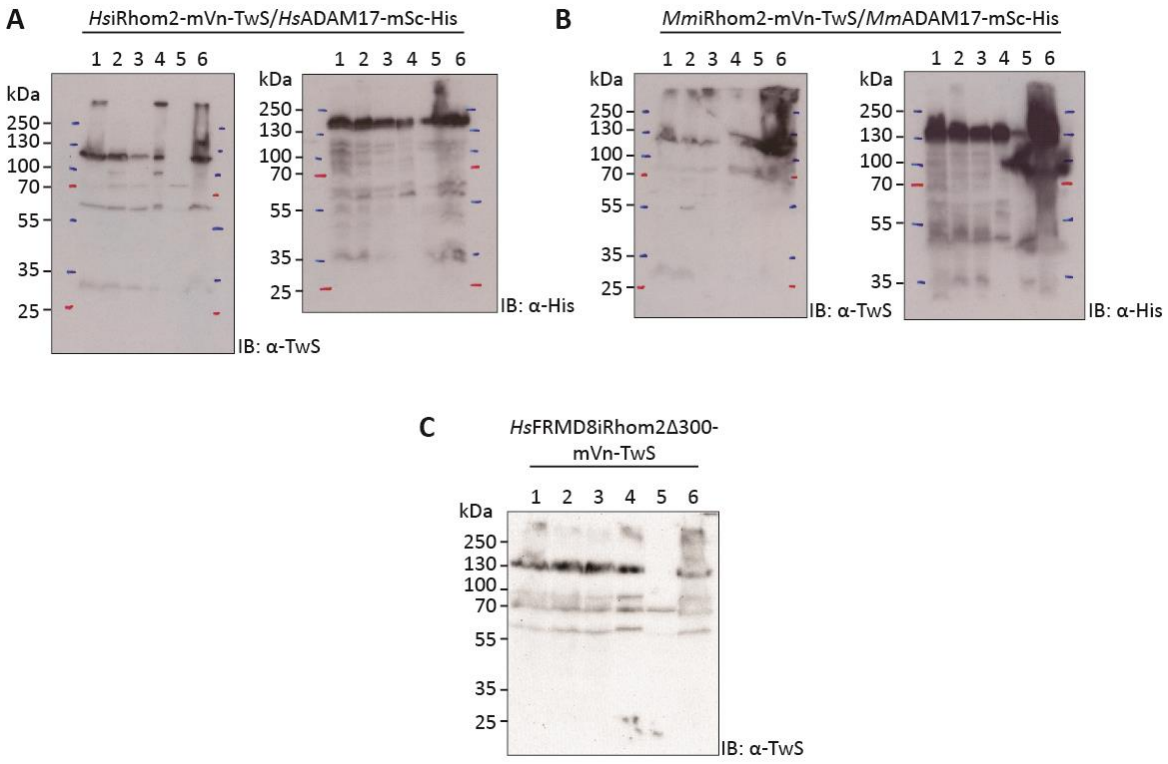


Figure 5.2. Stages of membrane isolation. Western blots visualising the stages of membrane isolation from HEK293F cells over co-expressing **(A)** *Hs*iRhom2-mVenus-Tws/*Hs*ADAM17-mScarlet-I-His, **(B)** *Mm*iRhom2-mVenus-Tws/*Mm*ADAM17-mScarlet-I-His and **(C)** HEK293F cells over-expressing *Hs*FRMD8iRhomΔ300-mVenus-TwS. iRhom was detected for the Twin-Strep tag using α-TwS antibody and ADAM17 was detected for the hexahisting tag using α-His antibody. Lane numbers represent each stage of the membrane isolation: 1, resuspended cells; 2, lysed cells; 3, supernatant after 27,000 *x g*; 4, resuspended pellet; 5, supernatant after 210,000 *x g*; 6, resuspended membranes.

Isolated membranes from the generated cell lines were resuspended in the appropriate buffer supplemented with protease inhibitor cocktail, aliquoted, flash frozen in liquid nitrogen and stored at -80 °C until use. Membranes harvested from ultracentrifugation containing *Hs*FRMD8iRhom2 chimera were resuspended in 100 mM Tris HCl, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 and membranes containing the iRhom/ADAM17 complex were resuspended in 100 mM Tris HCl, 150 mM NaCl, pH 8.0.

5.3 *Hs*FRMD8iRhom2Δ300-mVenus-TwS

5.3.1 Detergent screening

The *Hs*FRMD8iRhom2 chimera was screened for the best solubilising conditions using fluorescent detection size-exclusion chromatography (FSEC). FSEC is a technique that was originally developed for pre-crystallisation screening of detergents to solubilise membrane proteins (Kawate and Gouaux, 2006). The protein of interest is fused to a fluorescent protein, typically GFP, and expressed in either bacteria or eukaryotic cells. Cells are harvested and lysed by whole-cell detergent solubilisation or mechanically to isolate membranes to then solubilise crude membranes in detergent. Solubilised sample is then applied to a size exclusion column equipped with a fluorescence detector and fluorescence elution profile is monitored rather than UV absorbance. Because fluorescence is more sensitive than absorbance, only small amount of sample is required therefore FSEC is a useful technique to screen for optimal detergent for membrane protein extraction and purification (Rollauer et al., 2012, Miller and Aricescu, 2014, Wright and Lee, 2019). Resuspended membranes were diluted to 10 mg/mL (based on BCA assay), supplemented with protease inhibitor cocktail, and solubilised in the conditions shown in Table 5.1 for 90 minutes on a rotary wheel at 4 °C. Solubilised material was subject to two rounds of centrifugation at 4 °C, the first at 20,000 *x g* using a benchtop microfuge and the supernatant centrifuged at 100,000 *x g*. The supernatant from the last centrifuge spin was filtered using a 0.22 µm 0.5 mL filter (Millipore) and applied to a Superose 6 Increase 5/150 column (GE Healthcare). The size exclusion column was connected to an ÄKTA Pure (GE Healthcare) with inline fluorescence detector (Shimadzu). Fluorescence from the mVenus tag was monitored rather than A_{280nm} absorbance. Excitation and emission wavelengths of the fluorophores used are shown in Table 5.2. All solubilising conditions are shown in Figure 5.3. DDM, GDN and LMNG were chosen as part of the screen because they have been used successfully in the past to extract membrane proteins out of membrane (Section 5.1) and DM was also part of the screen because it has a shorter alkyl chain than DDM. Each detergent was

screened without CHS and DM, DDM and LMNG was screened with two different CHS ratios (detergent:CHS): 10:1 and 5:1, to test whether *Hs*FRMD8iRhom2 chimera requires CHS. GDN was not screened with CHS because the hydrophobic tail of GDN is a steroid-based group and therefore already mimics CHS. In every condition there were two main peaks present (Figures 5.3A-C), except for membranes solubilised in GDN (Figure 5.3D) where only one peak was resolved. The first eluting peak (0.40 CV) in each condition corresponded to the void peak whereby the proteins did not interact with the column and eluted first. Material in the void peak was likely to be aggregated protein of interest. Therefore GDN was a poor solubiliser of the *Hs*FRMD8iRhom2 chimera because most the protein was in the void. The second eluting peak for each condition in Figures 5.3A-C was the peak of interest. Here, the *Hs*FRMD8iRhom2 chimera was solubilised and interacted with the pores within the column. The increase in CHS caused a slight shift to the left of the second eluting peak, likely the result of the detergent micelle around the protein changing shape and morphology to elute sooner (O'Malley et al., 2011). Out the conditions screened lauryl maltose neopentyl glycol (LMNG) with 0.1 % CHS (Figure 5.3C, green trace) was carried forward for purification because it was the condition that gave the highest fluorescence reading of the peaks of interest.

Table 5.1. Solubilising conditions for the *Hs*FRMD8iRhom2Δ300-mVn-TwS detergent screen.

Detergent	Detergent concentration (%) (w/v)	CHS (%) (w/v)
DM	1.0	0.0
DM	1.0	0.1
DM	1.0	0.2
DDM	1.0	0.0
DDM	1.0	0.1
DDM	1.0	0.2
GDN	1.5	0.0
LMNG	1.0	0.0
LMNG	1.0	0.1
LMNG	1.0	0.2

Table 5.2. Excitation and emission wavelengths of mVenus and mScarlet-I.

Fluorophore	Excitation λ (nm)	Emission λ (nm)
mVenus	515	527
mScarlet-I	569	593

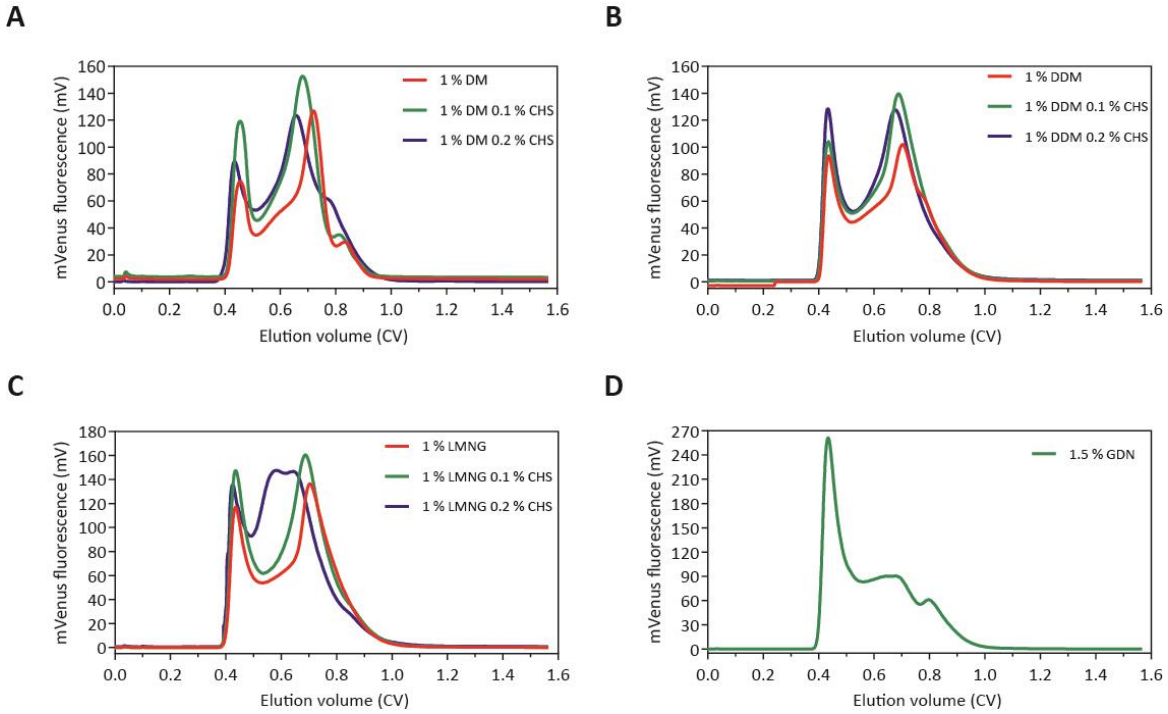


Figure 5.3. *Hs*FRMD8iRhom2-mVenus-TwS detergent screen. *Hs*FRMD8iRhomΔ300-mVenus-TwS was screened for the best solubilising conditions using fluorescent detection size-exclusion chromatography. Membranes were solubilised in **(A)** 1 % n-decyl-β-D-maltopyranoside (DM), **(B)** 1 % n-dodecyl-β-D-maltoside (DDM), **(C)** 1 % lauryl maltose neopentyl glycol (LMNG) and **(D)** 1.5 % glyco-diosgenin (GDN). In (A-C) the detergent used was to solubilise membranes alone (red trace), with 0.1 % cholesteryl hemisuccinate (CHS) (green trace) or 0.2 % CHS (blue trace). The column used was Superose 6 Increase 5/150.

5.3.2 Purification

5.3.2.1 Initial large scale purification

For large scale purification, resuspended *Hs*FRMD8iRhom2 chimera membranes were diluted to 10 mg/mL in 100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0 supplemented with EDTA-free protease inhibitor cocktail. Membranes were solubilised in a final concentration of 1 % LMNG and 0.1 % CHS where LMNG:CHS was prepared a 10:1 ratio on a rotary wheel for 90 minutes at 4 °C. Insolubilised material was pelleted by ultracentrifugation (100,000 *x g*, 30 minutes, 4 °C). Supernatant was incubated with 1 mL (1 CV) of equilibrated Strep-Tactin XT Superflow beads with gentle spinning for 2 hours at 4 °C. The supernatant/bead solution was applied to a disposable gravity flow column and the flow through collected. Beads were washed in 10 CV of 100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 0.02 % LMNG, 0.002 % CHS, pH 8.0 and proteins eluted in 3 CV using 100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 0.02 % LMNG, 0.002 % CHS, 50 mM biotin, pH 8.0. Proteins were eluted by the following strategy: elution 1, 0.6 CV; elution 2, 1.6 CV; elution 3, 0.8 CV for a more concentrated sample in the second elution fraction. Elution fractions were pooled and concentrated using a 100,000 MWCO (GE Healthcare) spin concentrator and applied to a Superose 6 Increase 10/300 column. The size-exclusion running buffer was 100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 0.02 % LMNG, 0.002 % CHS, pH 8.0. SDS-PAGE of the affinity purification and size-exclusion chromatography for the first large scale purification is shown in Figure 5.4A and Figure 5.4C respectively. There are fewer protein bands in the elution fractions that the flow through because the Strep-tactin resin has as higher specificity for the Twin-Strep tag therefore there was less non-specific binding. *Hs*FRMD8iRhom2 chimera eluted from the size-exclusion column as a roughly symmetrical peak with shoulder at approximately 0.64 CV (Figure 5.4B, highlighted). Size-exclusion fractions D1-D3 were pooled, concentrated and negative stain grids prepared, as described in Section 2.5.2, to assess sample quality (Figure 5.4D). Negative stain of the concentrated SEC peak fractions in Figure 5.4D showed dispersed, monomeric particles with little aggregation, hence it was worthwhile to

visualise *Hs*FRMD8iRhom2 chimera by cryo-electron microscopy. This particular purification, the final yield was 16 µg from 1.8 g *Hs*FRMD8iRhom2 chimera HEK293F membranes (2 L), which was enough to prepare negative stained grids but not enough to prepare cryo grids. To produce multiple cryo grids at different concentrations and different grid preparation parameters, which is required to screen for the optimal grid, the amount of protein required is in the range of 200-400 µg. Therefore the amount of protein after negative grid preparation was not sufficient to prepare a range cryo grids required for the grid screening stage or for other protein characterisation experiments.

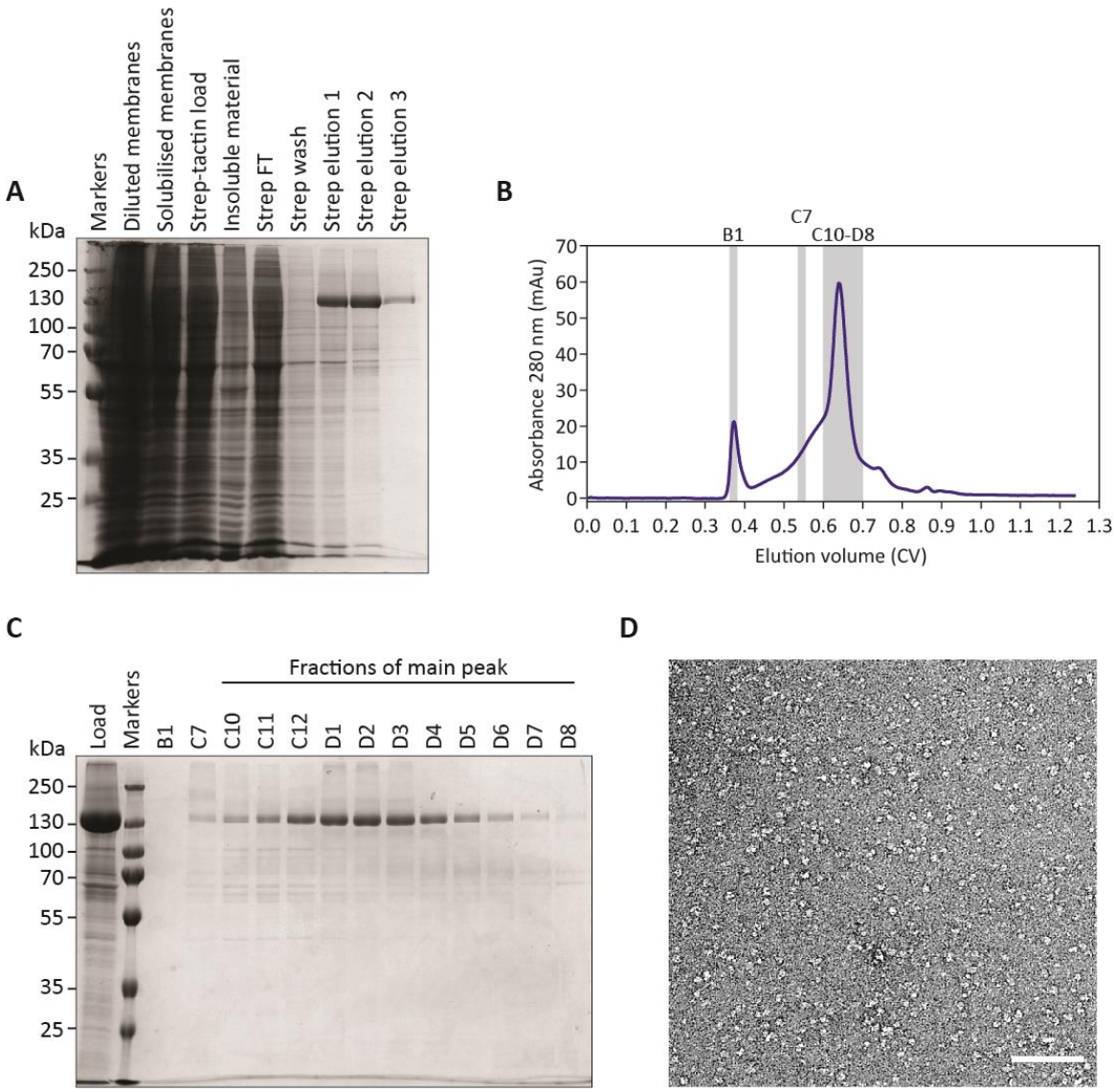


Figure 5.4 *Hs*FRMD8iRhom2-mVenus-TwS purification. *Hs*FRMD8iRhomΔ300-mVenus-TwS was purified by affinity chromatography of the Twin Strep tag **(A)** followed by size-exclusion chromatography **(B and C)**. Pooled and concentrated fractions D1-D2 of **(C)** were visualised by negative stain **(D)** as a sample quality control. Scale bar represents 100 nm.

5.3.2.2 Further purifications

In order to produce enough purified *Hs*FRMD8iRhom2 chimera for cryo-EM analyses, *Hs*FRMD8iRhom2 chimera HEK293F cells were scaled up to 9 L to yield 13 g of membranes. These membranes were solubilised and purified as described in Section 5.3.2.1 but volumes scaled appropriately, using 5 mL of Strep-Tactin XT Superflow High Capacity beads and the elution strategy of the batch purification had an additional elution fraction of 1.2 CV (Figure 5.5A). The elution fractions were pooled and applied to the Superose 6 Increase 10/300 column with a fluorescence detector (Figure 5.5B). Fluorescence of the mVenus tag was monitored (Figure 5.5B, green trace) and had the same elution pattern as the A280 trace (Figure 5.5B, blue trace). The main eluting SEC peak was not as symmetric as the previous purification (Figure 5.4B) as there was an additional late shoulder. The fractions in the highlighted region in Figure 5.5B were visualised by SDS-PAGE (Figure 5.5C). The shoulder that was eluting later mainly consisted of protein band at 80 kDa and a small proportion of the full length protein. The 70 kDa protein is consistent with the molecular weight of transmembrane bundle and mVenus without FRMD8 which was confirmed by gel-band trypsin digest mass spectrometry because human iRhom2 and GFP were identified. Whereas the protein shoulder eluting earlier is the same size of the full length *Hs*FRMD8iRhom2 chimera. Fractions D4 to D6 of the main peak were pooled and concentrated to 1.17 mg/mL. The total amount of protein in the pooled fractions was 398 µg. 100 µg of purified *Hs*FRMD8iRhom2 chimera was dedicated for 3C cleavage testing (Section 5.3.3) and the remaining protein was concentrated further to 4.2 mg/mL to prepare cryo-EM grids. Purified *Hs*FRMD8iRhom2 chimera was confirmed by gel-band trypsin digest mass spectrometry. Preparation and imaging of cryo-EM grids is described in Chapter 6.

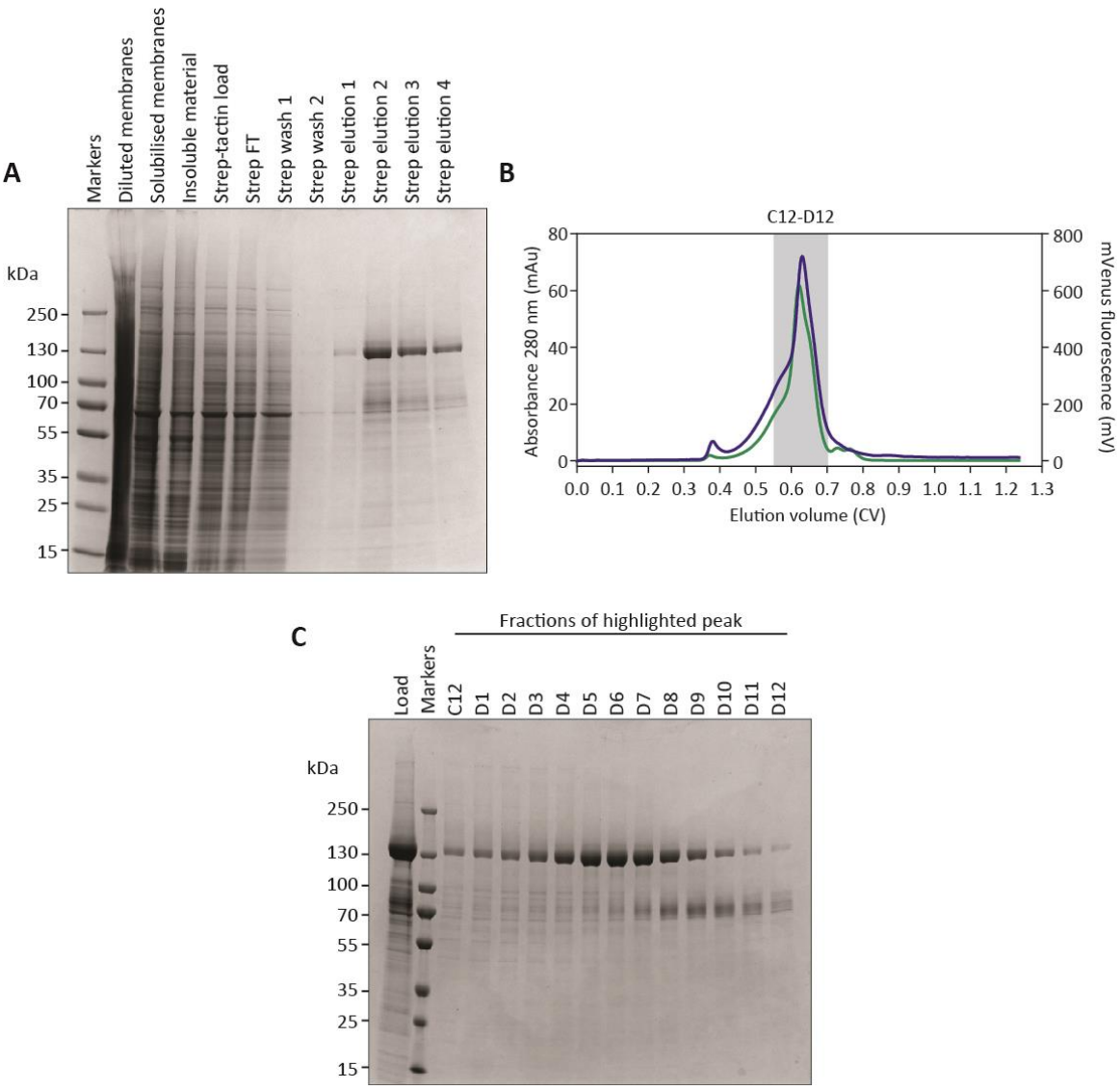


Figure 5.5 Larger scale *Hs*FRMD8iRhom2-mVenus-TwS purification. *Hs*FRMD8iRhomΔ300-mVenus-TwS was purified by affinity chromatography of the Twin Strep tag **(A)** followed by size-exclusion chromatography using a Superose 6 Increase 10/300 column **(B)**. Absorbance at 280 nm (blue trace) and fluorescence (green trace) of the mVenus tag were measured simultaneously and fractions of grey highlighted region in **(B)** was visualised by SDS-PAGE **(C)**.

The remaining fractions (D1-D3 and D7-D12) were pooled and concentrated (Figure 5.6B) and applied to Superdex 200 Increase 10/300 with the aim to better to separate the full length protein from the 80 kDa protein (Figure 5.6A). The Superdex 200 matrix has smaller pore size than that of Superose 6 which may provide a better resolution between peaks. The main eluting peak could be resolved into three features: a shoulder and two overlapping peaks. The fractions

within the highlighted region was visualised by SDS-PAGE (Figure 5.6C). The later fractions (C5-C9) that made up of the second peak contained a higher proportion of the 80 kDa protein than the full length protein. However, the early eluting shoulder contained the full length protein. Therefore it is likely that the protein in the early-eluting shoulder is *Hs*FRMD8iRhom2 in a different confirmation, different oligomeric state, or has a different size detergent micelle that would affect the overall shape of the protein-detergent complex.

The same fractions shown in Figure 5.6C were visualised again on SDS-PAGE (Figure 5.7A) after 14 days stored at 4 °C and tested for in-gel fluorescence (Figure 5.7B) of the mVenus tag on the chimera. In-gel fluorescence revealed that the 80 kDa protein contains mVenus which suggests that the 80 kDa protein is a clipped form without FRMD8. However comparing the SDS-PAGE of Figure 5.6C and Figure 5.7A, there was no visible difference suggesting that *Hs*FRMD8iRhom2 is stable at 4 °C for a short period of time. The central peak made up of fractions C1-C2 was pooled, concentration and applied to an additional Superdex 200 Increase 10/300 column (Figure 5.8A) and fractions visualised by SDS-PAGE (Figure 5.8B). The later eluting peak seen in Figure 5.6A was reduced to a shoulder but the earlier eluting peak was present despite only using the main peak fractions from the previous Superdex 200 run. Again the main peak fractions, C2-C3, shown in Figure 5.8A-B were pooled and concentrated and applied again to the same column (Figure 5.8C) with the aim to achieving a symmetrical monomeric peak. This was achieved as the peak shown in Figure 5.8C was a single peak but was broad at the base of the peak. However the repeated size-exclusion purifications of peaks reduced the protein yield. Using the Superdex 200 columns may be a better choice of size-exclusion column for purifying the *Hs*FRMD8iRhom2 chimera.

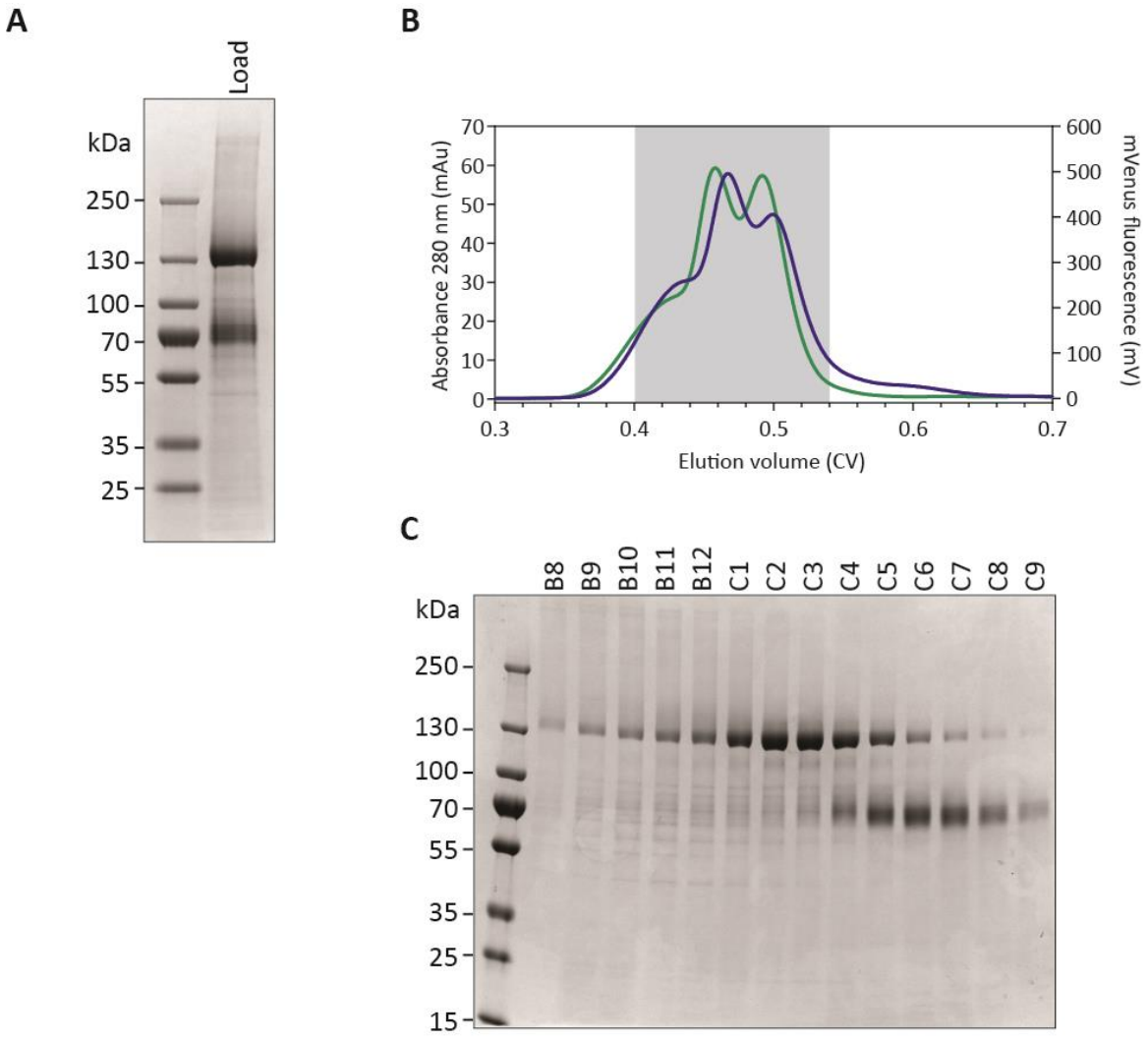


Figure 5.6. Superdex 200 purification of *Hs*iFRMD8iRhom2-mVenus-TwS. *Hs*FRMD8iRhomΔ300-mVenus-TwS purified from a Superose 6 column **(A)** was applied to a Superdex 200 Increase 10/300 column **(B)** for better peak resolution. Absorbance at 280 nm (blue trace) and fluorescence (green trace) of the mVenus tag were measured simultaneously and fractions of grey highlighted region in **(A)** was visualised by SDS-PAGE **(C)**.

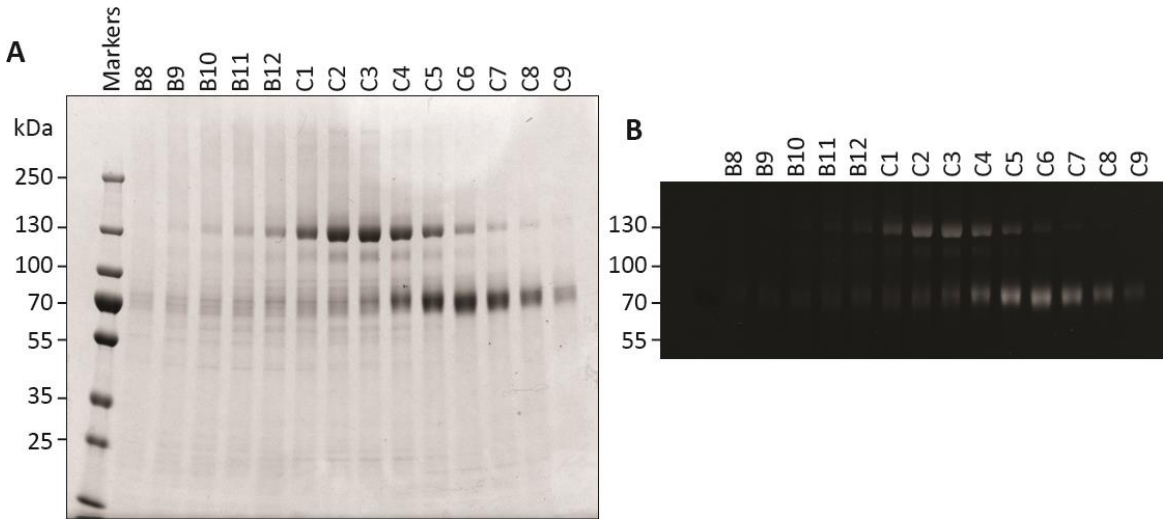


Figure 5.7. *Hs*iFRMD8iRhom2-mVenus-TwS stability. Fractions from the Superdex 200 purification were stored at 4 °C for 14 days and visualised by SDS-PAGE followed by Coomassie staining **(A)** and in-gel fluorescence of the mVenus tag **(C)**.

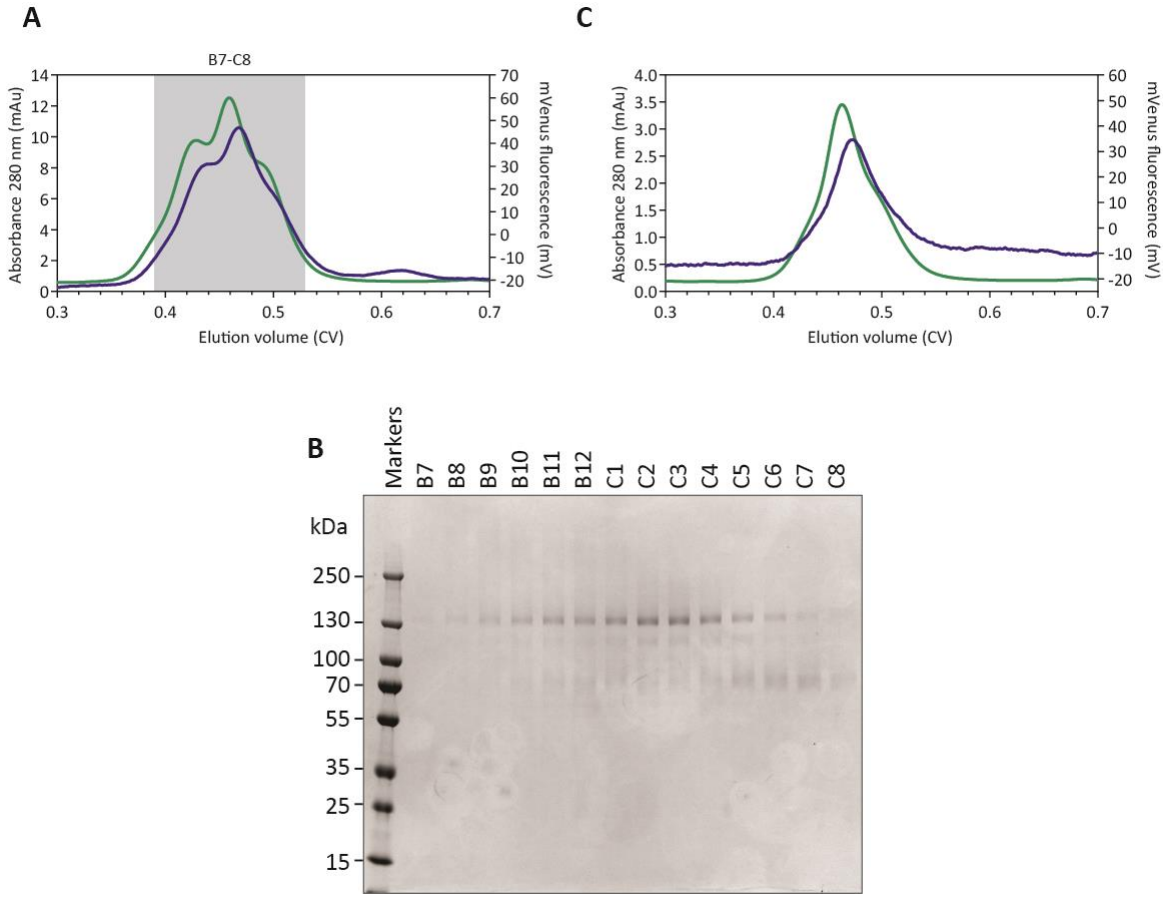


Figure 5.8. Repeated size-exclusion of Superdex 200 peak fractions. Superdex 200 purified *Hs*FRMD8iRhomΔ300-mVenus-TwS was subjected another round of Superdex 200 **(A)** and the grey highlighted peak was visualised by SDS-PAGE **(B)**. Peak fraction C2 and C3 from **(B)** was pooled and concentrated and applied to Superdex 200 10/300 column **(C)**. Absorbance at 280 nm (blue trace) and fluorescence (green trace) of the mVenus tag were measured simultaneously.

5.3.3 HCV 3C protease cleavage testing

Removal of the C-terminal mVenus and Twin-Strep tag on *Hs*FRMD8iRhom2 was tested. Because the mVenus and Twin-Strep tag are linked to *Hs*FRMD8iRhom2 by the flexible 3C linker, the tags are not rigid with respect to the rest of the protein and therefore not ideal for particle alignment during 2D classifications. Purified *Hs*FRMD8iRhom2 chimera was dialysed against 50 mM Tris HCl, 150 mM NaCl, pH 8.0 to reduce the concentration of Tris and to remove EDTA. The removal of EDTA was required because of the reverse affinity step after cleavage. *Hs*FRMD8iRhom2 was incubated with purified his-tagged 3C protease at a 2:1 ratio (protein:3C) overnight at 4 °C. Post 3C treatment, the protein solution was subject to reverse affinity purification. The protein solution was incubated with a mixed bead solution of Strep-Tactin and Ni-NTA, and the flow through collected. Samples from flow through, beads, pre- and post 3C treatment were visualised by SDS-PAGE (Figure 5.9). Tag cleavage was efficient because there was a gel band shift from 130 kDa (*Hs*FRMD8iRhom2Δ300-3C-mVenus-TwS) in the pre 3C treatment lane to 120 kDa (*Hs*FRMD8iRhom2Δ300) in the post 3C treatment lane. The protein band at 45 kDa was the 3C protease and the band at 25 kDa was the mVenus-TwS. A low ratio of protein to 3C protease was used to ensure complete cleavage of the tag, as there was a possibility of incomplete cleavage due to the inaccessibility of the 3C cleavage site due to the presence of bound detergent micelles. If there was an excess of purified protein, different ratios of protein to 3C would have been tested to determine the minimal amount of 3C to reduce the amount of 3C protein added to the purified *Hs*FRMD8iRhom2.

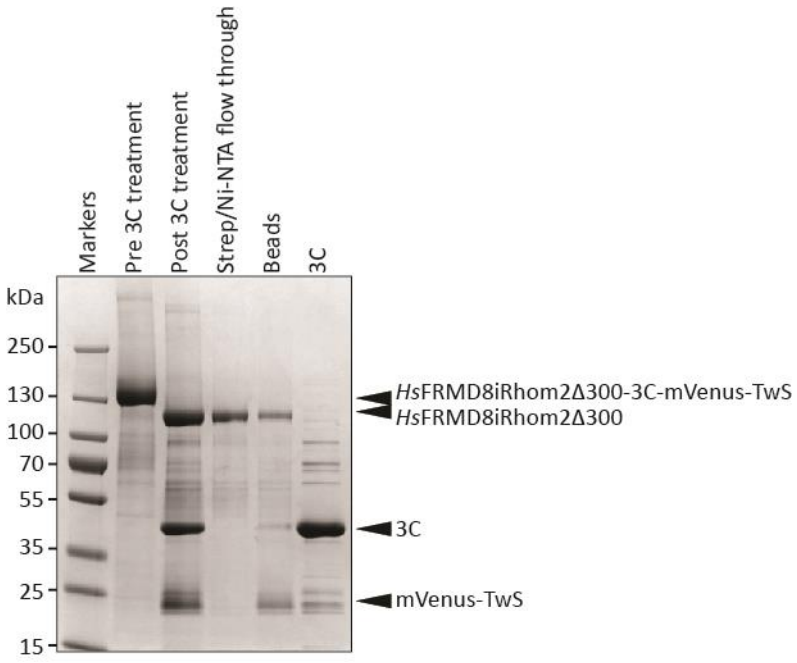


Figure 5.9. HCV-3C protease cleavage of *Hs*FRMD8iRhomΔ300-mVenus-TwS. Purified *Hs*FRMD8iRhomΔ300-mVenus-TwS was treated with Hepatitis C virus 3C protease overnight at 4 °C to remove the mVenus-TwS C-terminal tag. Cleaved *Hs*FRMD8iRhomΔ300 was collected in the flow through of the reverse affinity purification. The 3C protease and mVenus-TwS was captured by Ni-NTA and Strep-Tactin beads respectively.

5.4 iRhom and ADAM complex

5.4.1 Detergent screening

Membranes isolated where iRhom and ADAM17 were overexpressed were screened for optimal solubilising conditions and co-elution of the mVenus and mScarlet-I fluorescence by FSEC. Resuspended membranes were treated as described in Section 5.3.1 but the detergents used specifically for the cell line is shown in Table 5.3 for similar reasons specified in Section 5.3.1. However, detergents with CHS were only tested at 10:1 ratio (detergent:CHS) because the increase in CHS concentration did not increase the fluorescence of the *Hs*FRMD8iRhom2 chimera detergent screen.

Table 5.3. Solubilising conditions for the mammalian iRhom/ADAM complex detergent screen.

***Mm*iRhom2-mVn-TwS/*Mm*ADAM17-mSc-His**

Detergent	Detergent concentration (%) (w/v)	CHS (%) (w/v)
DM	1.0	0.0
DM	1.0	0.1
DDM	1.0	0.0
DDM	1.0	0.1
GDN	1.5	0.0
LMNG	1.0	0.0
LMNG	1.0	0.1

***Hs*iRhom1/2-mVn-TwS/*Hs*ADAM17-mSc-His**

Detergent	Detergent concentration (%) (w/v)	CHS (%) (w/v)
DDM	1.0	0.0
DDM	1.0	0.1
GDN	1.5	0.0
LMNG	1.0	0.0
LMNG	1.0	0.1

5.4.1.1 *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His

Solubilised membranes were applied twice to a Superose 6 Increase 5/150 column to measure the fluorescence of mVenus and mScarlet-I separately. The fluorophores were measured in two separate but identical runs then compiled manually due to technical limitations of the fluorescence detector. The detergent screen of the *Mm*iRhom2-mVn-TwS/*Mm*ADAM17-mSc-His coexpression is shown in Figure 5.10. Similar to the *Hs*FRMD8iRhom2 chimera detergent screen, in all conditions there was a first eluting peak that corresponded to the void volume of the column (0.4 CV).The only condition where mVenus and mScarlet-I demonstrated distinct but co-eluting peaks not in the void volume was in the 1 % LMNG and 0.1 % CHS condition (Figure 5.10G). Whereas in all the other solubilising conditions the fluorescence peaks did not overlap or were not well resolved. As seen in previous results, GDN was a poor solubiliser because most of the fluorescence signal appeared in the void volume. The mVenus traces where a straight line was recorded at 600 mV (Figure 5.10E and G) at 0.45 CV was due to the signal reaching the detection limit of the fluorescence detector. Selected fractions, highlighted in grey, from the FSEC trace in 1 % LMNG and 0.1 % CHS (Figure 5.11A) were visualised by western blot against

the Twin Strep tag of *Mm*iRhom (Figure 5.11B) and His tag of *Mm*ADAM17 (Figure 5.11C). The trace shown in Figure 5.10G is the same as in Figure 5.11A. For the mVenus trace, only fractions A7 and A11-B1 were analysed and for the mScarlet-I trace, all fractions highlighted in Figure 5.11A were analysed. Fraction A7 was the void peak where both fluorophores eluted, fractions A11-B1 contained the resolved peak where both fluorophores co-eluted and fractions B4-B8 contained additional peaks eluted from mScarlet-I trace. All fractions tested for *Mm*ADAM17 (Figure 5.11C) show a band at approximately 130 kDa that is consistent with the predicted molecular weight of the mature form of *Mm*ADAM17-mSc-His (120.9 kDa). The elution profile of *Mm*ADAM17 consisted of a void peak and two eluting peaks. The void peak likely contained aggregated *Mm*ADAM17. The first eluting peak after the void peak corresponded to *Mm*ADAM17 in complex with *Mm*iRhom and the later eluting peak consisted of *Mm*ADAM17 alone. Based on peak fluorescent height and intensity of detected band by immunoblotting, most of *Mm*ADAM17 eluted complexed with *Mm*iRhom2. The western blot probing for the Twin Strep tag (Figure 5.11B) was not well developed however a band can be seen at approximately 120 kDa in fractions A11-B1 of the mScarlet-I trace, consistent with the predicted molecular weight of *Mm*iRhom2-mVn-TwS (124.5 kDa). The bands developed in the same fractions when probing for *Mm*iRhom2 and *Mm*ADAM17 in the column interacting peak and co-elution of the mVenus and mScarlet-I peaks suggest complex formation. Therefore, LMNG with CHS was selected for solubilising further *Mm*iRhom/*Mm*ADAM17 complex preparations.

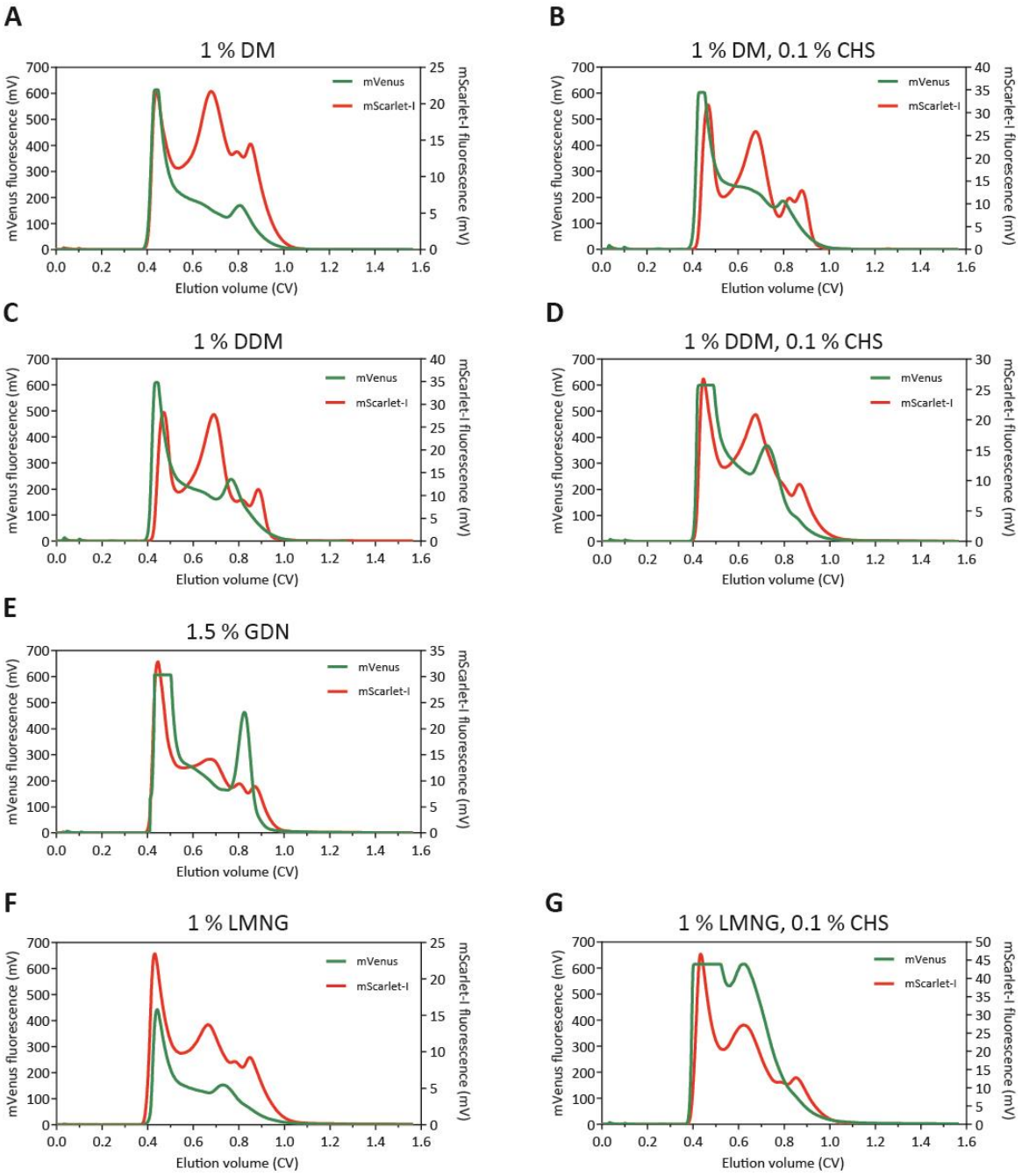


Figure 5.10. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His detergent screen. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His was screened for the best solubilising conditions and co-elution of the mVenus (green trace) and mScarlet-I (red trace) using fluorescent detection size-exclusion chromatography. Membranes were solubilised in the conditions shown within the figure and applied to a Superose 6 Increase 5/150 was twice measuring the mVenus and mScarlet-I subsequently.

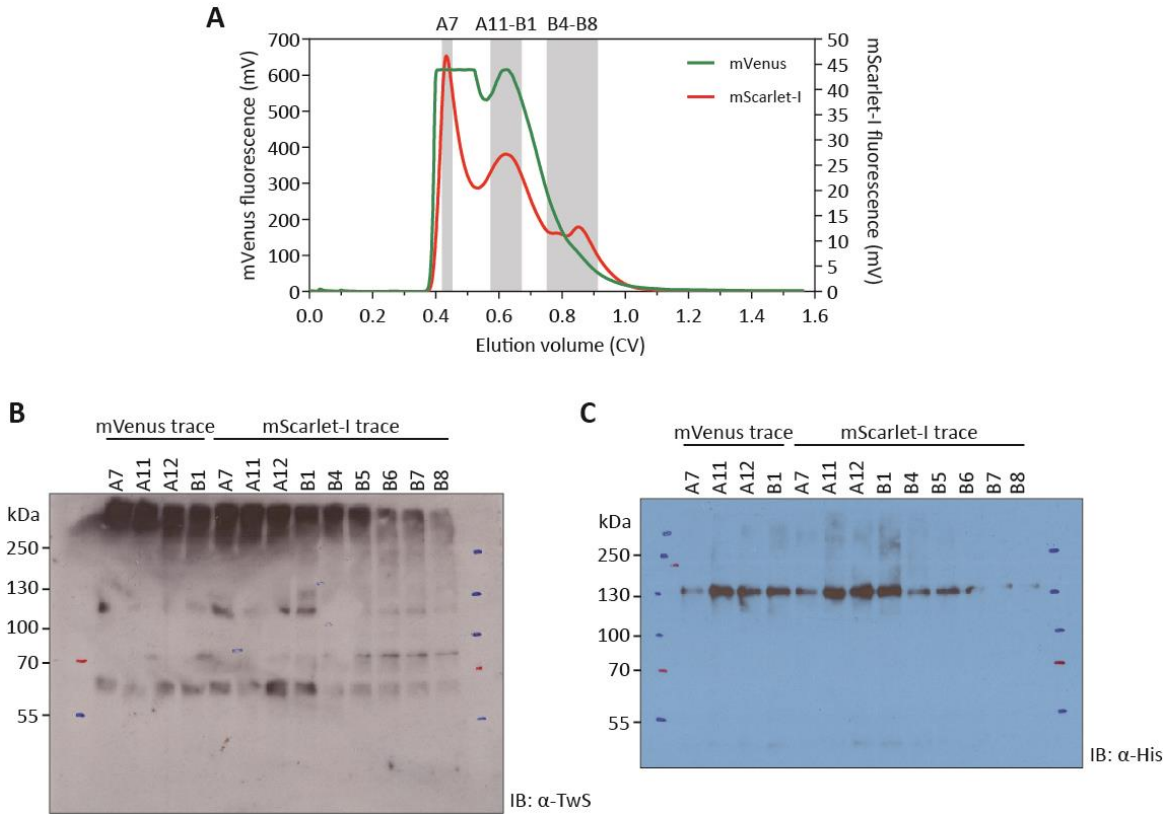


Figure 5.11. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His solubilised in 1 % LMNG and 0.1 % CHS. The only solubilising condition of *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His where there was co-elution of the mVenus (green trace) and mScarlet-I (red trace) was in 1 % LMNG and 0.1 % CHS **(A)**, as shown in Figure 5.10. The solubilised membranes were applied twice to the Superose 6 Increase 5/150 to measure mVenus and mScarlet-I fluorescence separately. Peaks highlighted in grey were visualised by western blot probing for the Twin-Strep tag on iRhom **(B)** and the his tag on ADAM17 **(C).**

5.4.1.2 Detergent screen of *Hs*iRhom1 or *Hs*iRhom2 in complex with *Hs*ADAM17

Membranes solubilised in each detergent condition (Table 5.3) were applied once to a Superose 6 Increase 5/150 where both mVenus and mScarlet-I fluorophores were measured simultaneously because the fluorescence detector was upgraded to detect two wavelengths. DM was not part of the detergent screen because it was a poor at solubilising *Mm*iRhom2 in the *Mm*iRhom2/*Mm*ADAM17 detergent screen (Figure 5.10A and B). The detergent screen of *Hs*iRhom1-mVn-TwS/*Hs*ADAM17-mSc-His is shown in Figure 5.12 and of *Hs*iRhom2-mVn-TwS/*Hs*ADAM17-mSc-His is shown in Figure 5.13. Similar to previous detergent screens, in all

conditions there was a first eluting peak that eluted in the void volume of the column (0.4 CV) followed by at least peak resolved within the column. Where the mVenus peak appears to plateau at 600 mV was due to the signal measured was over the detection limit of the detector.

There were three solubilising conditions where both complexes screened produced co-eluting peaks of the two fluorophores. These conditions were: 1 % DDM, 1 % LMNG and 1 % LMNG with 0.1 % CHS, the traces of the conditions listed are shown again in Figure 5.14. Co-eluting peak fractions (Figure 5.14A-F, grey highlights) were analysed by western blot, probing for *Hs*iRhom1 and *Hs*iRhom2 by the Twin Strep tag (Figure 5.14G) and for *Hs*ADAM17 by the His tag (Figure 5.14H). There was very little signal from the *Hs*iRhom1-mVn-TwS blot against the Twin-Strep tag (Figure 5.14G), however bands were detected at 55, 70 and 100 kDa in the LMNG and LMNG with CHS fractions of the *Hs*iRhom2-mVn-TwS runs. The lower molecular weight bands are expected to be either non-specific binding of antibody or degraded forms of *Hs*iRhom2-mVn-TwS. The molecular weight of the 100 kDa band was lower than the expected molecular weight of *Hs*iRhom2-mVn-TwS of 127.8 kDa. This deviation in protein mass may have been due to degradation of the protein or the aberrant migration pattern observed for membrane proteins when analysed by western blot (Dr. Justin Deme, personal communication). The same fractions, B3 and B1 of *Hs*iRhom2-mVn-TwS/*Hs*ADAM17-mSc-His, in Figure 5.14H showed a single band at approximately 130 kDa for *Hs*ADAM17, consistent with the expected molecular weight of the mature form of *Hs*ADAM17 (119.8 kDa). The slightly larger observed molecular weight may have been due to post transitional modifications to *Hs*ADAM17. *Hs*ADAM17 has nine predicted N-glycosylation sites which may be responsible for the increase in the predicted molecular weight. The bands developed in fractions B3 of LMNG and B1 of LMNG with CHS of *Hs*iRhom2-mVn-TwS/*Hs*ADAM17-mSc-His when probing for *Hs*iRhom2 and *Hs*ADAM17 were indicative of complex formation. Complex formation was further supported by the co-elution of the mVenus and mScarlet-I in the FSEC traces (Figure 5.14D and F). However the immunoblots in Figure

5.14G of *Hs*iRhom1-mVn-TwS could not confirm the presence of *Hs*iRhom1 in the co-eluting peaks from the DDM, LMNG and LMNG with CHS conditions. The best detergent to solubilise membranes of *Hs*iRhom1/*Hs*ADAM17 and *Hs*iRhom2/*Hs*ADAM17 for complex purification based on the data generated was LMNG. The addition of CHS did not appear to reduce the amount of ADAM17 (Figure 5.14H) and it also better resolved the eluting peaks of the mScarlet-I traces than without CHS (for *Hs*iRhom1, Figure 5.14C and E; for *Hs*iRhom2, Figure 5.14D and F). Therefore, the best solubilising conditions for the human iRhom/ADAM17 complexes for purification was LMNG with CHS, the same conditions that of the mouse iRhom/ADAM17 complex.

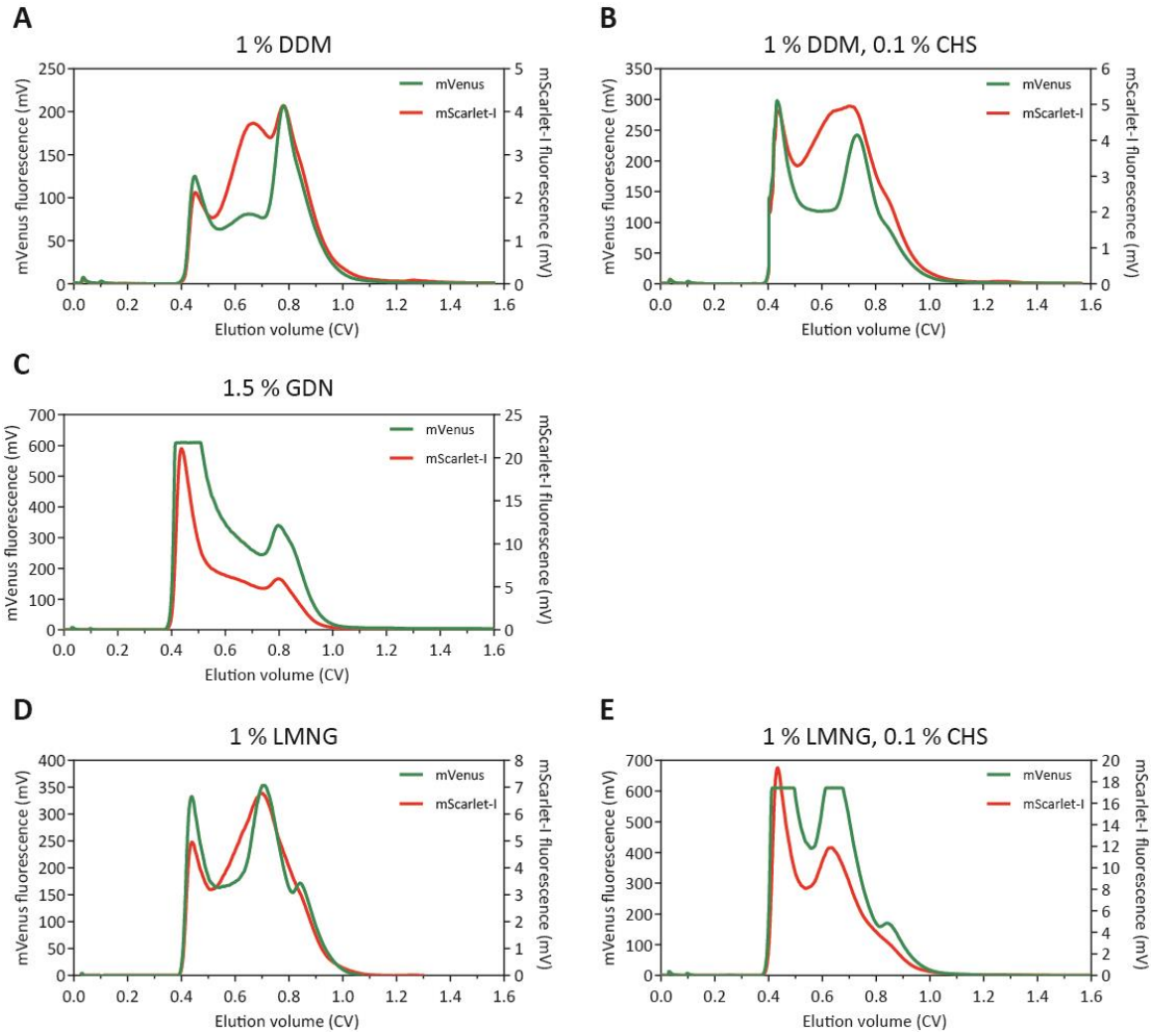


Figure 5.12. *Hs*iRhom1-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His detergent screen. *Hs*iRhom1-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His was screened for the best solubilising conditions and co-elution of the mVenus (green trace) and mScarlet-I (red trace) using fluorescent detection size-exclusion chromatography. Membranes were solubilised in the conditions shown within the figure and applied to a Superose 6 Increase 5/150 and mVenus and mScarlet-I fluorescence signal was measured simultaneously.

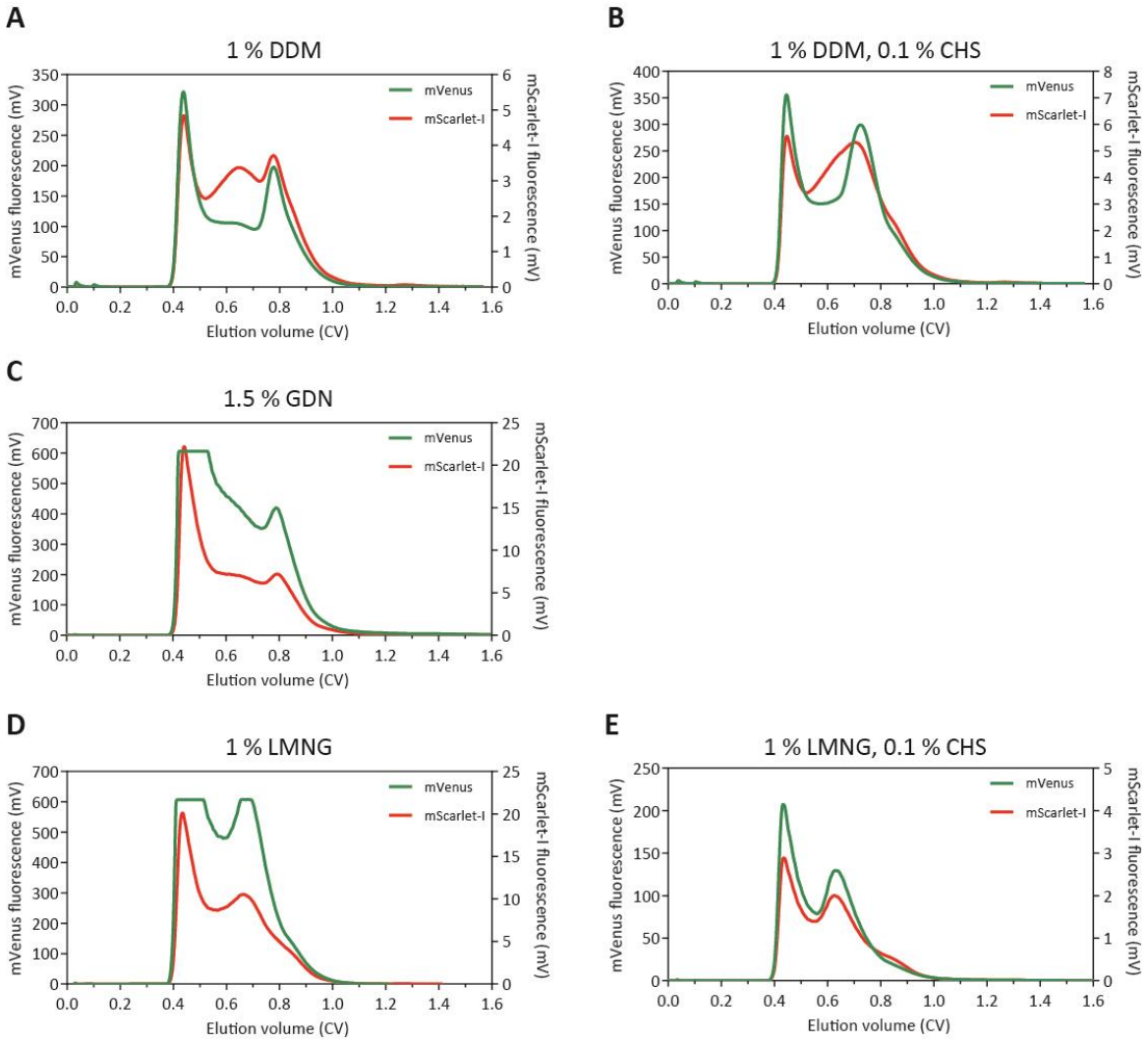


Figure 5.13. *Hs*iRhom2-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His detergent screen. *Hs*iRhom2-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His was screened for the best solubilising conditions and co-elution of the mVenus (green trace) and mScarlet-I (red trace) using fluorescent detection size-exclusion chromatography. Membranes were solubilised in the conditions shown within the figure and applied to a Superose 6 Increase 5/150 once were mVenus and mScarlet-I fluorescence was measured simultaneously.

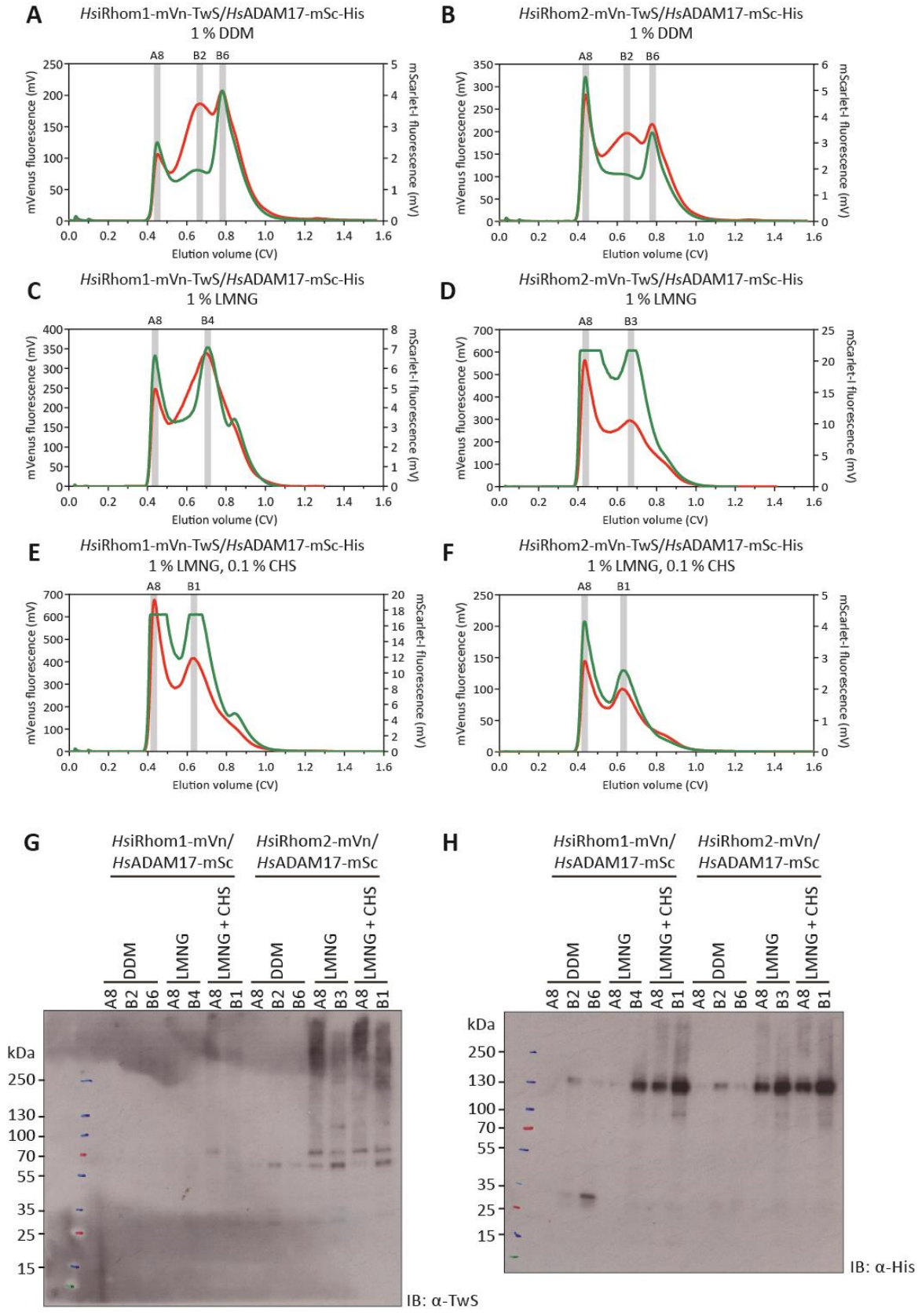

A
HsiRhom1-mVn-TwS/HsADAM17-mSc-His
1 % DDM
B
HsiRhom2-mVn-TwS/HsADAM17-mSc-His
1 % DDM
C
HsiRhom1-mVn-TwS/HsADAM17-mSc-His
1 % LMNG
D
HsiRhom2-mVn-TwS/HsADAM17-mSc-His
1 % LMNG
E
HsiRhom1-mVn-TwS/HsADAM17-mSc-His
1 % LMNG, 0.1 % CHS
F
HsiRhom2-mVn-TwS/HsADAM17-mSc-His
1 % LMNG, 0.1 % CHS
mVenus fluorescence (mV)
mScarlet-I fluorescence (mV)
Elution volume (CV)
G
HsiRhom1-mVn/ HsADAM17-mSc
HsiRhom2-mVn/ HsADAM17-mSc
DDM
LMNG
LMNG + CHS
kDa
IB: α-TwS
H
IB: α-His

Figure 5.14. Solubilising conditions of *Hs*iRhom1-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His and *Hs*iRhom2-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His visualised by western blot. The solubilising conditions where the mVenus and mScarlet-I peaks co-eluted for *Hs*iRhom1-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His **(A, C and E)** and *Hs*iRhom2-mVenus-TwS/*Hs*ADAM17-mScarlet-I-His **(B, D and F)**, as shown in Figure 5.13 and Figure 5.14, were verified by western blot. Peaks highlighted in grey were visualised by western blot probing for the Twin-Strep tag on *Hs*iRhom1/2 **(G)** and the His tag on *Hs*ADAM17 **(H)**.

5.4.2 iRhom/ADAM17 complex purification

5.4.2.1 Two-step purification of *Mm*iRhom-mVn-TwS/*Mm*ADAM17-mSc-His

Resuspended membranes overexpressing *Mm*iRhom2-mVn-TwS and *Mm*ADAM17-mSc-His were diluted to 10 mg/mL in 100 mM Tris HCl, 150 mM NaCl, pH 8.0 supplemented with EDTA-free protease inhibitor cocktail. Membranes were solubilised in a final concentration of 1 % LMNG and 0.1 % CHS, where LMNG:CHS was prepared a 10:1 ratio, on a rotary wheel for 2 hours at 4 °C. Insoluble material was pelleted by ultracentrifugation (100,000 *x g*, 30 minutes, 4 °C). Supernatant was applied to an equilibrated 1 mL Strep-Tactin XT Superflow High Capacity cartridge using a peristaltic pump at flow rate approximately 1 mL/min. The column was equilibrated in 100 mM Tris HCl, 150 mM NaCl, 0.02 % LMNG, 0.002% CHS, pH 8.0. The flow through was reapplied to the column twice to maximise the binding of the Twin-Strep tag to the column. The column was washed in 6 mL (6 CV) of 100 mM Tris HCl, 150 mM NaCl, 0.02 % LMNG, 0.002% CHS, pH 8.0 and proteins eluted in 6 CV using 100 mM Tris HCl, 150 mM NaCl, 0.02 % LMNG, 0.002 % CHS, 50 mM biotin, pH 8.0. Proteins were eluted by the following strategy: elution 1, 0.6 CV; elution 2, 1.6 CV; elution 3, 0.8 CV; elution 4, 1 CV; elution 5, 1 CV; elution 6, 1 CV for a more concentrated sample in the second elution fraction. Elution fractions 4-6 were collected to ensure no protein was lost in the dead volume of the peristaltic pump. The stages of the Strep-Tactin purification were visualised by SDS-PAGE and stained by Coomassie (Figure 5.15A). A faint band can be seen in Elution fractions 3 and 4 approximately at 130 kDa however more bands are revealed when all the elution fractions were pooled and concentrated

(Figure 5.15B). Two bands are shown at approximately 130 kDa (Figure 5.15B), a higher band and lower band. The higher band was likely to be *Mm*ADAM17 and the lower was likely to be iRhom however this could not be confirmed by proteomics because too little protein was present.

Pooled elution fractions were concentrated using a 100,000 kDa MWCO concentrator (Figure 5.15B) and applied to a Superose 6 Increase 10/300 column with a fluorescence detector, measuring absorbance at 280 nm, mVenus fluorescence and mScarlet-I fluoresce simultaneously. The size-exclusion running buffer was 100 mM Tris HCl, 150 mM NaCl, 0.02 % LMNG, 0.002 % CHS, pH 8.0. The size-exclusion traces are shown in Figure 5.15 where absorbance (Figure 5.15C) and fluorescence of both fluorophores, mVenus (Figure 5.15D) and mScarlet-I (Figure 5.15E), measured against elution volume. The peak eluting at approximately 0.60 CV was the *Mm*iRhom/*Mm*ADAM17 complex because it was the only peak where the mVenus and mScarlet-I fluorescence co-eluted. There was an excess of free *Mm*iRhom2 compared to *Mm*iRhom2 complexed with *Mm*ADAM17, as demonstrated by the peak eluting at approximately 0.65 CV in the mVenus trace (Figure 5.15D). This was due to the affinity purification of the Twin Strep tag present on *Mm*iRhom2 which permitted *Mm*iRhom2 in complex with *Mm*ADAM17 to be purified in addition to free, uncomplexed *Mm*iRhom2. An additional affinity step was required for affinity purification of the complex. This was achieved by Ni-NTA purification of the Strep-Tactin elution fractions as described and discussed in the next section. The maximum A_{280nm} reading of the complex peak was 2.0 mAu (Figure 5.15C). 2.0 mAu roughly equates to approximately 4 µg of complex which was too little protein to pool and concentrate for further experiments. This yield was the result of preparation from 2.19 g of resuspended membranes from 13.87 g of HEK293F cells (1 L). To achieve the yield required for structural experiments it was necessary to increase culture volumes significantly, approximately 50-fold if the yield was to scale linearly to the volume of cells cultured.

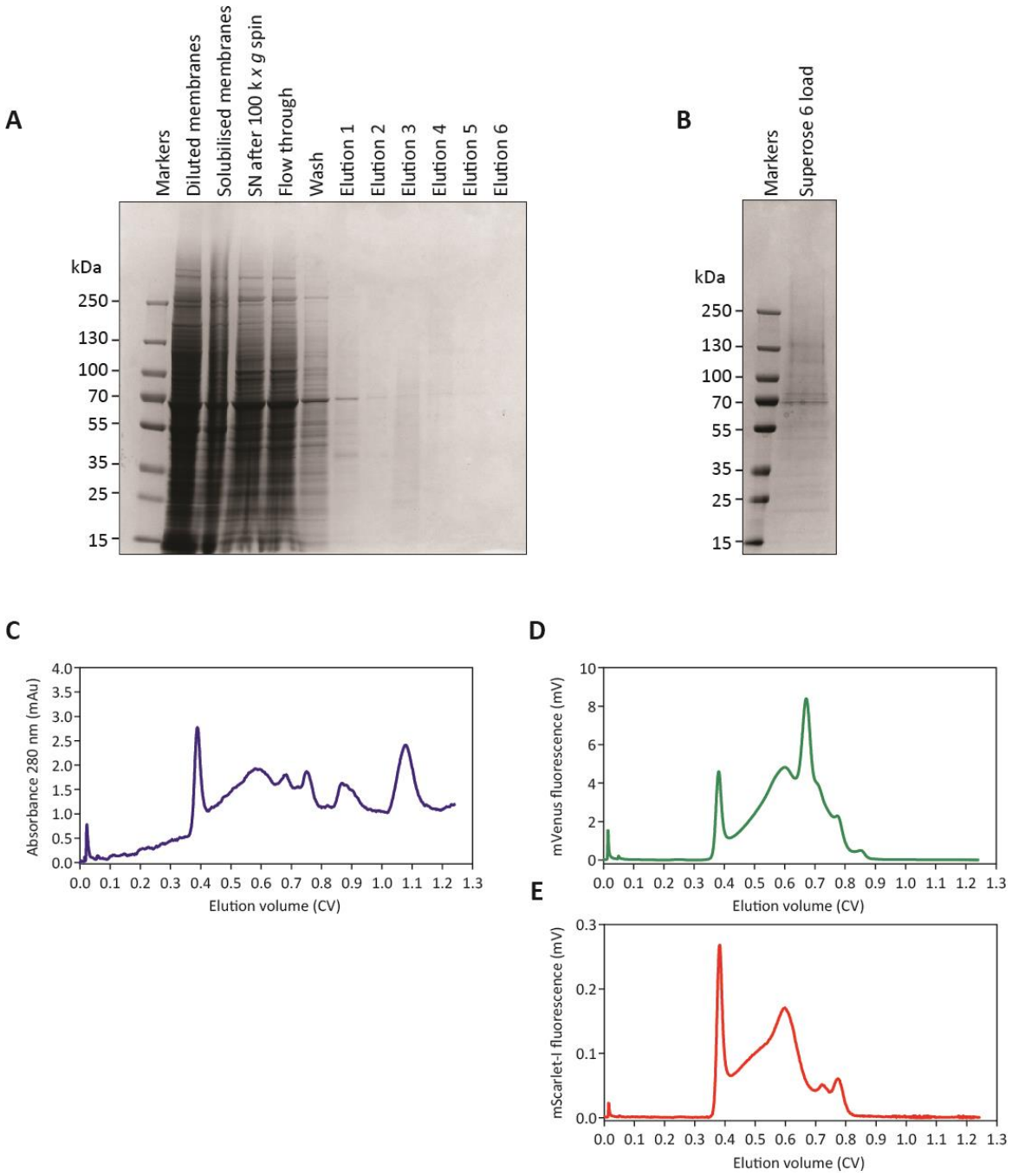


Figure 5.15. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His two-step purification. *Mm*iRhom and *Mm*ADAM17 complex was purified by a two-step process, Strep-Tactin followed by size-exclusion. The stages of the Strep-tactin purification was visualised by SDS-PAGE followed by Coomassie staining **(A)**. The strep eluate from the Strep-Tactin elution was pooled and concentrated **(B)** and applied to a Superose 6 Increase 10/300 column. Absorbance at 280 nm **(C)** (blue trace), mVenus **(D)** (green trace) and mScarlet-I **(E)** (red trace) fluorescence was measured simultaneously against elution column volume.

5.4.2.2 Three-step purification of *Mm*iRhom-mVn-TwS/*Mm*ADAM17-mSc-His

*Mm*iRhom2-mVn-TwS/*Mm*ADAM17-mSc-His HEK293F cells were scaled up further (4 L) to yield 5.40 g of membranes from 34.7 g of HEK293F cells. These membranes were solubilised, clarified by ultracentrifugation and purified using the 1 mL Strep-Tactin XT Superflow High Capacity cartridge as described in Section 5.4.2.1. The elution strategy used was to collect one fraction of 6 CV of the Strep-Tactin elution buffer, supplemented with 0.02 % LMNG and 0.002 % CHS. The Strep-Tactin eluate was diluted with 150 mM NaCl, 0.02 % LMNG, 0.002 % CHS to reduce the concentration of Tris from 100 mM to 50 mM so that Tris concentration was below the manufacturer's maximum recommended concentration for immobilised metal affinity chromatography (Ni-NTA). The diluted Strep-Tactin eluate was applied to a 1 mL Ni-NTA cartridge (Qiagen) preequilibrated in 50 mM Tris HCl, 150 mM NaCl, pH 8.0 at flow rate of 1 mL/min. The cartridge and pump were equilibrated in 50 mM Tris HCl, 150 mM NaCl, pH 8.0. The flow through was reapplied to the column twice to maximise the binding of the complex to the column. The Ni-NTA cartridge was washed in 12 CV of 50 mM Tris HCl, 150 mM NaCl, 20 mM imidazole, 0.02 % LMNG, 0.002 % CHS, pH 8.0 and proteins eluted in 8 CV of 50 mM Tris HCl, 150 mM NaCl, 250 mM imidazole, 0.02 % LMNG, 0.002 % CHS, pH 8.0. The Ni-NTA eluate was dialysed against 50 mM Tris HCl, 150 mM NaCl, 0.02 % LMNG, 0.002 % CHS, pH 8.0 overnight to remove excess imidazole from the eluate. All stages of the purification were visualised by western blot where *Mm*iRhom2 (Figure 5.16A) was detected using StrepMAB classic antibody and *Mm*ADAM17 (Figure 5.16B) was detected using Tetra·His antibody. In the Ni-NTA load lanes, which is also the diluted Strep-Tactin eluate, there was one band present in the western blot against *Mm*ADAM 17 and multiple bands in the western blot against *Mm*iRhom2.

The dialysed Ni-NTA eluate was concentrated to 500 µL using a 100,000 kDa MWCO concentrator and applied to a Superose 6 Increase 10/300 column with fluorescence detector, measuring absorbance at 280 nm, mVenus fluorescence and mScarlet-I fluorescence

simultaneously. The size-exclusion running buffer was 50 mM Tris HCl, 150 mM NaCl, 0.02 % LMNG, 0.002 % CHS, pH 8.0. The size-exclusion traces are shown in Figure 5.16 where absorbance (Figure 5.16C) and fluorescence of both fluorophores, mVenus (Figure 5.16D) and mScarlet-I (Figure 5.16E), measured against elution volume. The peak eluting at approximately 0.60 CV, with a similar elution volume to the previous purification, was solely the complex because the mVenus and mScarlet-I elution traces perfectly superimpose. However the maximum A_{280nm} reading of the *Mm*iRhom2/*Mm*ADAM17 complex peak was 1.6 mAu (Figure 5.16C) equating to approximately 2 µg of complex. Therefore despite increasing the mass of HEK293F cells cultured more than 40 g of cells are needed to purify this complex in order to gain a high enough yield to perform structural experiments. A cause of the low yield despite the increase in cell mass was the dual affinity purifications because each purification stage would have resulted in a loss of protein.

Unfortunately due to the COVID-19 pandemic all experimental work was halted therefore no more work progressed on the *Mm*iRhom2/*Mm*ADAM17, *Hs*iRhom1/*Hs*ADAM17 and *Hs*iRhom1/*Hs*ADAM17 complex.

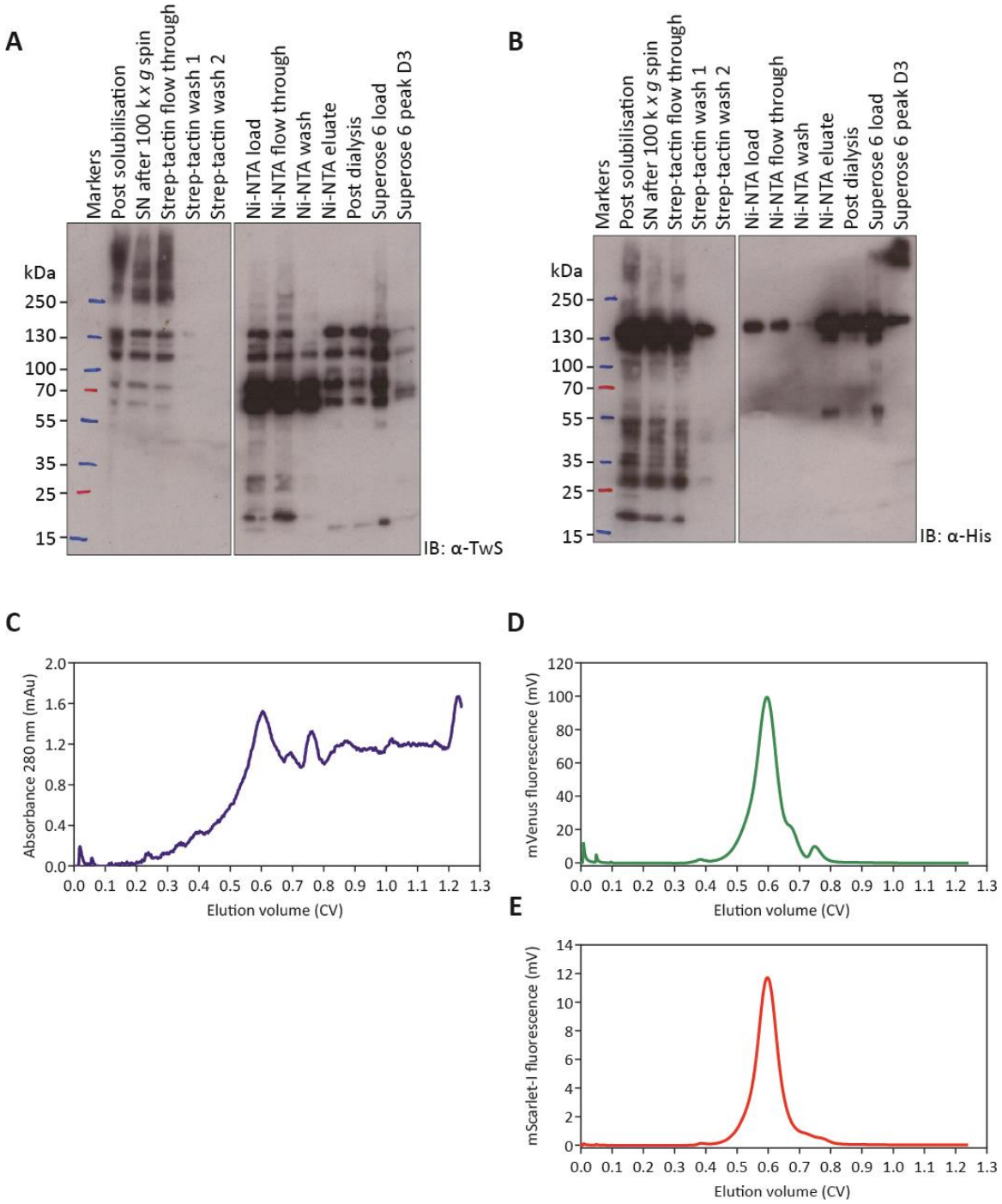


Figure 5.16. *Mm*iRhom2-mVenus-TwS/*Mm*ADAM17-mScarlet-I-His three-step purification. The iRhom and ADAM17 complex was purified by a three-step process, affinity purification by Strep-Tactin, then Ni-NTA followed by size-exclusion. Stages of the three step purification was visualised by western blot **(A and B)** whereby the iRhom was probed against its Twin-Strep tag **(A)** and ADAM17 was probed against its His tag **(B)**. The eluate of from Ni-NTA purification was applied to a Superose 6 Increase 10/300 column. Absorbance at 280 nm **(C)** (blue trace), mVenus **(D)** (green trace) and mScarlet-I **(E)** (red trace) fluorescence was measured simultaneously against elution column volume.

5.5 Discussion

5.5.1 *Hs*FRMDiRhom2 chimera

A protocol for purifying the *Hs*FRMD8iRhom2 chimera, the stable structural target, was successfully developed and improved upon during the course of this project. Future purifications of this particular construct should use the Superdex 200 rather than the Superose 6 size-exclusion columns because there was slightly better peak resolution using the Superdex 200. When the HEK293F cells were expanded further to increase the yield of protein, the size-exclusion trace showed more shouldering (Figure 5.5) than the previous purification (Figure 5.4). This may have been due to the harvested cell pellets stored at -20 °C longer than previously or the cells cultured were at a high passage number and therefore not expressing the construct properly. Despite the volumes of the buffers were scaled appropriately with the increase in cell pellet mass, the volume of the Strep-Tactin beads was not scaled up by the same factor therefore may have caused aggregation when *Hs*FRM8iRhom2 chimera was binding to the beads.

Cryo grids were prepared using the purified *Hs*FRMD8iRhom2 chimera and the outcome of that is described and discussed in Chapter 6

5.5.2 iRhom and ADAM17 complex

For all of the complexes screened in the detergent screen the solubilising condition LMNG with CHS, at a 10:1 ratio, was the best condition because there was co-elution of the mVenus and mScarlet-I elution traces. LMNG consists of two linked hydrophobic chains of equal length to two maltoside groups and CHS is a cholesterol analogue, which can replicate the native mammalian membrane (Kulig et al., 2015). Despite the detergents tested were non-ionic, therefore unlikely to disrupt protein-protein interaction, LMNG with CHS was the only condition across of all the complexes tested to co-elute complex. This may be due to the reduced conformational flexibility

of the quaternary carbon in LMNG and therefore better stabilise the complex than the other detergents (Chae et al., 2010).

The purification of the *Mm*iRhom2 and *Mm*ADAM17 complex benefitted from a double affinity chromatography with size-exclusion rather than just a single affinity chromatography step. The double affinity chromatography steps ensured that only *Mm*iRhom2 and *Mm*ADAM17 in complex with one another were purified therefore reducing the amount of monomeric *Mm*iRhom2.

5.5.3 Future work

Due to the COVID-19 pandemic no further experiments were performed on the iRhom/ADAM17 complex project. However the final experiments, before the United Kingdom's lockdown restrictions were enforced, were promising and just a matter of scaling up the cell mass to purify enough protein to prepare cryo-EM grids. However due to the low yields as a result from expression and dual affinity chromatography followed by size-exclusion chromatography other mammalian expression systems were explored. A higher protein expressing cell line called the Expi293F (Thermo Fisher Scientific) was reported by the manufacturer to produce higher yields of over-expressed soluble proteins by transient transfection than their older FreeStyle 293F cell line due to the Expi293F cells have been adapted to grow in denser cultures. Therefore generating the same constructs that were used in the FreeStyle 293F cells in the newer Expi293F cells may produce higher yields of protein than their FreeStyle counterparts. Weeks before the lockdown, the Expi293F GnTI- cell lines were successfully transduced, enriched for fluorescence and frozen down but no more experiments were performed. First, these cell lines would be revived and expanded and their expression for iRhom and ADAM17 verified followed by immunofluorescence to establish the subcellular localisation of overexpressed iRhom and ADAM17 in a similar process to Chapter 4. Moving to these Expi293F cells, in theory, should

increase the yield and therefore provide more iRhom/ADAM17 complex to perform several experiments with. Another approach to enhance protein expression in HEK cells is the use of histone deacetylase (HDAC) inhibitors such as valproic acid or sodium butyrate (Backliwal et al., 2008, Masiulis et al., 2019). HDAC inhibitors inhibit the actions of histone deacetylase to condense and transcriptionally silence chromatin. Therefore resulting in the hyperacetylation of histones causing chromatin to be in an open conformation to upregulate the expression of genes (Thiagalingam et al., 2003, Dokmanovic et al., 2007). The combination of the Expi293 cell line, reaching higher cell densities, and HDAC inhibitors, to enhance transcription, may result in more protein expression which may increase yield of purified protein.

Alongside purifying the iRhom/ADAM17 in detergents for structural studies, we don't know how the complex will look in a LMNG/CHS detergent micelle. Therefore it may be prudent to first purify the complex in detergent then to insert the complex into nanodiscs with either synthetic lipids or native lipid extracts so that the complex is within a more native environment. However, this may be difficult if the complex was purified in LMNG due to its low critical micelle concentration (0.001 % (w/v)) (cmc). The cmc is defined as the minimal concentration required for detergent molecules form micelles; below the cmc the detergent molecules exist as monomers (Seddon et al., 2004). Because the cmc of LMNG is low, dialysis is not suitable to remove LMNG from protein complexes but hydrophobic adsorption beads, that non-ionic detergents bind to, can be used to remove detergents (Seddon et al., 2004).

In addition to the mammalian iRhom/ADAM cell lines generated, non-mammalian iRhom/ADAM17 cell lines were also generated. Because in the detergent screen of the mammalian cell lines resulted in LMNG being the best detergent for solubilising and keeping the complex intact, these cell lines should be tested for their fluorescence of complex formation. As these non-mammalian cell lines may express better than the mammalian homologues.

Chapter 6: Preliminary structural determination of *Hs*FRMD8iRhom2Δ300 chimera and thesis summary

6.1 Introduction

Chapter 5 described and discussed the purification strategy used for the *Hs*FRMD8iRhom2Δ300-mVn-TwS chimera. This Chapter will discuss the preliminary steps towards solving the structure of the chimera by cryogenic electron microscopy (cryo-EM) single particle analysis (SPA).

Until recently, solving protein structures of a molecular weight lower than 200 kDa by cryo-EM SPA was extremely challenging. Small proteins have fewer scattering electrons to contribute to signal and therefore inherently generate a poor signal to noise ratio (SNR). Since biological molecules are sensitive to the ionising radiation of the electron beam, sample imaging is a comprise between maximising the SNR whilst minimising the radiation damage to the sample. Merk and colleagues (Merk et al., 2016) were the first to report cryo-EM maps of ≤ 150 kDa proteins, solving structures of the 145 kDa lactase dehydrogenase and 93 kDa isocitrate dehydrogenase. At the time of writing this thesis, there are currently 4776 map entries of protein samples in the Electron Microscopy Data Bank (EMDB) (EMBL-EBI, 2020a) that were solved by SPA at a resolution of 4 Å or higher. Of these, only 1.1 % are cryo-EM maps of proteins ≤ 150 kDa, excluding maps of the same protein but in a different conformation or cofactor/ligand bound (Table 6.1). Furthermore, 31 proteins from the list in Table 6.1 are membrane proteins.

Table 6.1. List of protein structures solved by cryogenic electron microscopy single particle analysis. List of proteins, molecular weight (MW) ≤ 150 kDa, that have been solved by single particle analysis at a resolution of ≤ 4 Å. This table was compiled from reviews (Wu and Lander, 2020, Nygaard et al., 2020) and from the EMBL-EBI EMDB (EMBL-EBI, 2020a) using their database search page. Note that structures that were solved with Antigen binding-fragments (Fabs), the MW contributed by the Fab is not reflected in the MW of the protein.
* Denotes additional EM maps but the highest resolution is listed
† Denotes a membrane protein

EMDB ID	Structure Title	Resolution (Å)	MW (kDa)	Sample	Symmetry	Reference
20755	Apo SAM-IV Riboswitch	3.7	39	RNA	C1	(Zhang et al., 2019)
20907	2.6 Angstrom reconstruction of apo-state streptavidin	2.6	50	Protein	D2	(Han et al., 2020)
0689*	Biotin-bound streptavidin	3.2	51	Protein	D2	(Fan et al., 2019)
0407*	Human methemoglobin state 1 determined using single-particle cryo-EM at 200 keV	2.8	63	Protein	C2	(Herzik et al., 2019)
3488	CryoEM structure of haemoglobin at 3.2 A determined with the Volta phase plate	3.2	64	Protein	C2	(Khoshouei et al., 2017)
0406	Horse liver alcohol dehydrogenase determined using single-particle cryo-EM at 200 keV	2.9	81	Protein	C2	(Herzik et al., 2019)
8943†	Cryo-EM Structure of the wild-type human serotonin transporter in complex with ibogaine and 15B8 Fab in the inward conformation	3.6	87	Protein, Fab	C1	(Coleman et al., 2019)
10676	Human Coronavirus HKU1 Haemagglutinin-Esterase	3.4	88	Protein	C2	(Hurdiss et al., 2020)
8193*	Cryo-EM structure of isocitrate dehydrogenase (IDH1) in complex with ML309 inhibitor	3.8	93	Protein	C2	(Merk et al., 2016)
21387†	Structure of G-alpha-q bound to its chaperone Ric-8A	3.5	96	Protein	C1	(Seven et al., 2020)
20806†	Single-Particle Cryo-EM Structure of Plasmodium falciparum Chloroquine Resistance Transporter (PfCRT) 7G8 Isoform	3.3	103	Protein, Fab	C1	(Kim et al., 2019)
20532†	Toll receptor EM map from Topaz picks	3.7	105	Protein	C1	(Bepler et al., 2019)
4908†	Cryo-EM structure of the E. coli cytochrome bd-I oxidase at 2.68 A resolution	2.7	108	Protein, nanodisc	C1	(Safarian et al., 2019)

10049	The structure of BD Oxidase from Escherichia Coli	3.3	110	Protein	C1	(Theßeling et al., 2019)
21603†*	Human open state TMEM175 in KCl	2.6	112	Protein	C2	(Oh et al., 2020)
4358†	Coupling specificity of heterotrimeric Go to the serotonin 5-HT1B receptor	3.8	112	Protein	C1	(García-Nafría et al., 2018)
9108†	Structural basis of coreceptor recognition by HIV-1 envelope spike	3.9	114	Protein	C1	(Shaik et al., 2019)
10258†*	Cryo-EM structure of nanodisc reconstituted yeast ALG6 in complex with 6AG9 Fab	3.0	118	Protein, nanodisc, Fab	C1	(Bloch et al., 2020)
9790†	Cryo-EM structure and transport mechanism of a wall teichoic acid ABC transporter	3.9	127	Protein	C2	(Chen et al., 2020)
22136†	Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs	2.9	128	Protein, nanodisc	C2	(Kern et al., 2020)
0902†	Cryo-EM structure of GPR52-miniGs-NB35	3.3	129	Protein, nanobody	C1	(Lin et al., 2020)
0498†	CryoEM structure of Helicobacter pylori urea channel in closed state	2.7	130	Protein	C6	(Cui et al., 2019)
0521	SARS-Coronavirus NSP12 bound to NSP8 co-factor	3.5	131	Protein	C1	(Kirchdoerfer and Ward, 2019)
9721†*	Human LAT1-4F2hc complex incubated with JPH203	3.3	131	Protein	C1	(Yan et al., 2019)
7835†	Cryo-EM structure of the human adenosine A1 receptor-Gi2-protein complex bound to its endogenous agonist	3.6	131	Protein	C1	(Draper-Joyce et al., 2018)
20264†*	Structure of YbtPQ importer with substrate Ybt-Fe bound	3.4	133	Protein	C1	(Wang et al., 2020b)
4775†	Heterodimeric ABC exporter TmrAB in ATP-bound outward-facing open conformation	2.8	133	Protein, nanobody	C1	(Hofmann et al., 2019)
10092†	KimA from Bacillus subtilis in inward-facing, occluded state	3.7	134	Protein	C2	(Tascón et al., 2020)
9360†*	Structure of zebrafish Otop1 in nanodiscs	3.0	135	Protein, nanodisc	C2	(Saotome et al., 2019)
21367	Cryo-EM structure of Cas13(crRNA)	3.2	135	Protein	C1	(Meeske et al., 2020)
9014*	EsCas13d-crRNA-target RNA ternary complex	3.3	136	Protein, RNA	C1	(Zhang et al., 2018a)
10337	XPF-ERCC1 Cryo-EM Structure, Apo-form	3.6	137	Protein	C1	(Jones et al., 2020)

10636†*	Structure of glutamate transporter homologue GltTk in sodium only condition	3.2	137	Protein, nanodisc	C1	(Arkhipova et al., 2020)
0683*	Ligand-triggered allosteric ADP release primes a plant NLR complex	3.7	138	Protein	C1	(Wang et al., 2019)
9847*	Cryo-EM structure of glutamate dehydrogenase from Thermococcus profundus	3.7	141	Protein	D3	(Oide et al., 2020)
4616	Structural basis of cohesin ring opening	3.3	141	Protein	C1	(Muir et al., 2020)
4997†	Structure of murine Solute Carrier 26 family member A9 (Slc26a9) anion transporter in the inward-facing state	4.0	141	Protein	C2	(Walter et al., 2019)
20223†*	Structure of the Rhodopsin-Transducin-Nanobody Complex	3.3	142	Protein, nanobody	C1	(Gao et al., 2019)
10550†	The structure of ABC transporter Rv1819c in AMP-PNP bound state	3.5	143	Protein	C2	(Rempel et al., 2020)
21373*	Single particle reconstruction of HemQ from Geobacillus based on data acquired in the presence of substantial aberrations	2.3	144	Protein	C5	(Bromberg et al., 2020).
7325†	Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation	3.4	144	Protein	C1	(Kim and Chen, 2018)
10111†	Nucleotide bound ABCB4	3.2	145	Protein	C1	(Olsen et al., 2020)
8191	Cryo-EM structure of lactate dehydrogenase (LDH) in complex with GSK2837808A	2.8	145	Protein	D2	(Merk et al., 2016)
9111†	Structural basis for cholesterol transport-like activity of the Hedgehog receptor Patched	3.6	146	Protein	C1	(Zhang et al., 2018b)
0190†	Cryo-EM structure of the ABCG2 E211Q mutant bound to ATP and Magnesium	3.1	146	Protein, nanodisc	C2	(Manolaridis et al., 2018)
0956†	Structure of human bile salt exporter ABCB11	3.5	147	Protein	C1	(Wang et al., 2020a)
9791†	Cryo-EM structure of human lysosomal cobalamin exporter ABCD4	3.6	147	Protein	C2	(Xu et al., 2019)
0650†	Cryo-EM structure of OTOP3 from xenopus tropicalis	3.9	147	Protein	C2	(Chen et al., 2019)

4256†	Structure of an inhibitor-bound ABC transporter	3.6	148	Protein	C2	(Jackson et al., 2018)
21023	Rabbit muscle aldolase determined using single-particle cryo-EM at 200 keV	2.1	149	Protein	D2	(Wu et al., 2020a)
9885*	Structure of human soluble guanylate cyclase in the NO activated state	3.8	149	Protein	C1	(Kang et al., 2019)
20442*	Cryo-EM structure of human NatE complex (NatA/Naa50)	3.0	150	Protein	C1	(Deng et al., 2020)
0476	Cryo-EM structure of NLRP3 bound to NEK7	3.8	150	Protein	C1	(Sharif et al., 2019)

Small membrane proteins present an additional problem if there are no defining features outside of the membrane. The lack of these features makes particle alignment difficult; the addition of proteins outside of the membrane can help with this. Examples include nanobodies (Table 5.1: (Gao et al., 2019, Lin et al., 2020, Hofmann et al., 2019)), megabodies (Laverty et al., 2019, Masiulis et al., 2019) and protein scaffolds (Liu et al., 2018) and Antigen binding-fragments (Fabs) (Table 5.1: (Coleman et al., 2019, Kim et al., 2019, Bloch et al., 2020)). Small membrane proteins that are asymmetric (Table 5.1, C1 symmetry) pose a problem because symmetry is a favourable for SPA. During SPA reconstructions, high symmetry can produce good results from fewer particles because each particle provides views from different directions (Penczek, 2010, Drulyte et al., 2018). Therefore small, asymmetric proteins are more challenging to solve by SPA.

Nanobodies are antibody fragments that consist of only variable heavy-chain domains (~15 kDa) that occur naturally in camelids and encoded by single gene fragments (Hamers-Casterman et al., 1993, Muyldermans, 2013). Nanobodies are raised *in vivo* against the purified protein of interest which then are used stabilise proteins and aid in crystallisation. For cryo-EM SPA, nanobodies can stabilise proteins but because of their small size do not always assist in particle alignment (Table 5.1: (Gao et al., 2019, Lin et al., 2020, Hofmann et al., 2019)). Therefore, nanobodies have been inserted into or fused onto protein scaffolds called megabodies (Uchański

et al., 2019) although finding a sufficiently rigid assembly is challenging. Protein scaffolds is a strategy used to address the size limitations imaging of small proteins; where the protein of interest, the 'cargo', is attached to large macromolecular structure that is amenable to imaging, the 'scaffold'. A number of considerations are needed when choosing a protein scaffold. The protein scaffold should be well ordered with little flexibility and, ideally, with a high proportion of secondary structure. Instead of using flexible linkers to fuse the cargo proteins to scaffolds, α-helical linkers are used to promote rigidity (Padilla et al., 2001, Liu et al., 2018, Yeates et al., 2020) thereby aiding in protein alignments because the scaffold does not average out during data processing. If the scaffold is well ordered with a high proportion of secondary structure elements there will be low intrinsic disorder within the scaffold therefore reducing heterogeneity and aiding alignments (Thompson et al., 2016). In addition, scaffold protein with symmetry can aid in SPA because usually fewer particles are needed (Penczek, 2010, Drulyte et al., 2018, Yeates et al., 2020). Near atomic resolution of GFP was achieved using a scaffold of Designed Ankyrin Repeat Proteins (DARPin) fused to an engineered tetrahedral symmetric protein complex (Liu et al., 2018, Yeates et al., 2020). Megabodies bind specifically to their target, increase the particle size and overcome the preferential orientation at the air-water interface of the grid (discussed further below) (Naydenova and Russo, 2017, Passmore and Russo, 2016, Uchański et al., 2020). Recently, the first high-resolution structures, utilising megabodies, of human synaptic $GABA_A$ receptor bound to the channel-blocker picrotoxin, the antagonist bicuculline, the agonist GABA, and the classical benzodiazepine drugs: alprazolam and diazepam (Laverty et al., 2019, Masiulis et al., 2019). The authors used a nanobody fused to a circularised form of an extracellular adhesion domain of *Helicobactor pylori*, HopQ (Laverty et al., 2019, Masiulis et al., 2019, Uchański et al., 2019).

Fabs have been used to promote crystal contacts and increase the probability of high resolution diffraction patterns of membrane proteins in X-ray crystallography (Hunte and Michel, 2002,

Carpenter et al., 2008). The use of Fabs has been transferred to cryo-EM SPA to successfully increase particle size which resulted in high resolution structures (Table 6.1) (Coleman et al., 2019, Kim et al., 2019, Bloch et al., 2020). Not only Fabs increase particle size but also aid in particle alignment, thus Fabs are a valuable tool for imaging small membrane proteins, particularly membrane proteins that lack significant protein mass to align on outside the membrane.

Recent advancements of hardware and software of cryo-EM SPA have enabled structural determination of sub 200 kDa proteins (Wu and Lander, 2020, Nygaard et al., 2020). These advancements include increasing accessibility of transmission electron microscopes (TEMs) operating at 300 kV, improvements in direct electron detectors (DEDs) and TEMs equipped with phase plates and/or energy filters (Thompson et al., 2016). In TEM high energy electrons are recorded using a detector, these include charged-coupled device (CCD) cameras or DEDs. Cryo-EM micrographs of biological samples are noisy because of the low electron doses to reduce radiation damage to the sample. Ideally, a perfect detector would not add additional noise but in reality all detectors do. The additional noise caused by the detector can be expressed at the detective quantum efficiency (DQE), which is calculated by the square of the output signal to noise ratio over the square of the input signal to noise ratio (McMullan et al., 2014). Therefore a perfect detector will have a DQE of 1, in practise detectors will have a DQE of less than 1 (Bai et al., 2015). Historically, film was used to collect TEM data however collecting large datasets by this method was inconvenient and time consuming. Additionally, film also had to be developed and digitised accurately in order to perform the data processing, which was expensive (Thompson et al., 2016). CCD cameras were developed as a solution to film whereby CCDs uses a phosphor scintillator that emits photons when electrons strike the scintillator. The CCD camera transformed the emitted photon into electrical signals that accumulate in photosensitive elements of the CCD. This accumulation results in the increase of electrical charge which is then

transferred to a register, amplified and digitised to form an image (Thompson et al., 2016). The DQE of a CCD camera is lower than that of film, at the electron energy level, due to the scintillation layer causing electron and photon scattering which results in the loss of sensitivity. CCD camera data recording has enabled automated microscope alignment and automated image acquisition thus easing housekeeping of the microscopes (Thompson et al., 2016). DEDs have a high DQE value and produce an electrical output which are now the detector of choice for high-resolution cryo-EM (Bai et al., 2015). DEDs detect incident electrons as they pass through a semiconductor membrane (~10 μm) and deposit energy that is detected by electronics present on the membrane. But backscattered electrons are also detected, as incident electrons pass through the membrane they backscatter from the support matrix which then pass through the semiconductor membrane again (Bai et al., 2015, Thompson et al., 2016). As a result, backscattered electrons contribute to noise and lowering the DQE. The DQE of DEDs can be improved by removing the support matrix as much as possible, < 50 μm, in a process called back-thinning (McMullan et al., 2009) whereby electrons pass through without backscattering (Bai et al., 2015). DEDs are able to spread the dose across a multiple frames generating a movie to the exposure which was made possible by the high frame rates of DEDs. This is a valuable benefit because each frame can be aligned relative to each other, allowing the motion correction of sample movement caused by mechanical drift or beam induced movement (Li et al., 2013, Rubinstein and Brubaker, 2015, Scheres, 2014, Thompson et al., 2016). Recent hardware advancements of DEDs, particularly Gatan's K3™ Camera, have improved the DQE, increased the field of view and frames per second generating high-resolution structures (Billesbølle et al., 2020).

When imaging biological samples in ice, there is little contrast (Orlova and Saibil, 2011) which can be compensated by defocussing the microscope (Thompson et al., 2016) to result in phase contrast of the micrograph but at the cost losing high resolution information. Phase plates have

developed to improve this issue, they function by adding an additional phase shift to more strongly scattered electrons of the beam. Therefore, phase plates allow imaging close to focus, or at focus, retaining high resolution information whilst producing enough contrast to visualise the sample on the micrograph (Thompson et al., 2016). Imaging samples using an electron beam, the electrons that interact with the sample become inelastically scattered as they lose some energy. The lower energy and hence longer wavelength of inelastic electrons means that they are focused in a different plane to the elastically scattered electrons (Orlova and Saibil, 2011) therefore the inelastically scattered electrons introduce noise to reduce the SNR. Energy filters can prevent these inelastically scattered electrons contributing to the micrograph by deflecting electrons of different energies/wavelengths along different paths, so that only elastically scattered (or electrons that did not scatter) electrons contribute to the image which improves the SNR (Thompson et al., 2016). The hardware advancements described above have allowed high-resolution maps of SPA reconstructions of small proteins over and are increasingly becoming routine for structural determination over the long standing X-ray crystallography. But no matter how advanced the hardware used to image small proteins, high-resolution structures is ultimately protein dependent on how well-ordered and well-behaved they are during purification and imaging.

The amount of protein required for grid preparations is in the microgram range, up to three orders magnitude lower than the amounts of protein needed for crystallisation. This makes cryo-EM even more appealing for membrane proteins, as membrane proteins tend to express more poorly compared to soluble proteins. Further, mammalian expression systems are longer, more costly, and higher maintenance than bacterial expression systems therefore protein yields from mammalian systems tend to be lower.

Plunge freezing of grids in liquid ethane produces vitreous ice thereby preserving the conformational state of protein in solution. Whereas protein structures solved by X-ray crystallography tend to represent only a static conformation which may or may not be biologically native. Generally, membrane proteins are dynamic molecules, a property required for their function. Examples of these include G-protein coupled receptors (GPCRs) that trigger a cellular second messenger response upon binding of their extracellular ligand, protein transporters that transport molecules across membranes, and inactive rhomboids that bind to their client proteins. The main drawback of preparing grids by plunge freezing is that proteins can denature at the air-water interface. Proteins, particularly membrane proteins, can adsorb to the hydrophobic air-water interface artificially concentrating them. This can be overcome by adding modified graphene supports. Graphene alone is hydrophobic there for not appropriate to solve the air-water interface problem. However, graphene can be rendered hydrophilic where graphene is hydrogenated by low-energy hydrogen plasma (Russo and Passmore, 2014, Passmore and Russo, 2016).

6.2 Cryo-EM of *Hs*FRMD8iRhom2Δ300

Purified *Hs*FRMD8iRhom2Δ300-mVn-TwS chimera was blotted onto glow discharged gold Quantifoil® R1.2/1.3 Holey Carbon grids as described in Section 2.5.3. The parameters of the plunge freezing, using the Vitrobot Mark IV, are shown in Table 6.2 and were performed by Dr. Justin Deme. These parameters were empirically determined with similar small membrane protein samples in same detergent to achieve optimal ice thickness (Dr. Justin Deme, personal communication).

The grids shown in Table 6.2 were prepared and screened on a 200 kV Talos Arctica with Falcon 3 direct electron detector, as part of The Central Oxford Structural Molecular Imaging Centre (COSMIC), by Dr. Justin Deme. Screening was performed to assess ice thickness and particle

density. Example micrographs from the grids screening is shown in Figure 6.1. Ideal particle density was achieved between 2 and 3 mg/mL (Figure 6.1B and Figure 6.1C, respectively). This highlights the importance of screening and preparing grids at several concentrations and with different freezing parameters because grids need to be verified before performing a full data collection. In addition screening will show how the particles behave in ice and in the examples shown in Figure 6.1 there was no protein aggregation.

Table 6.2. Parameters of the Vitrobot Mark IV of prepared grids of *Hs*FRMD8iRhom2Δ300-mVn-TwS. The grids below were prepared by Dr. Justin Deme. The asterisk (*) denotes the grid used for a full data collection on the Titan Krios.

Protein concentration (mg/mL)	Volume loaded (µL)	Wait time (seconds)	Blot time (seconds)	Blot force	Humidity (%)	Temperature (°C)
1	4	10	2	-15	100	8
2	4	10	2	-15	100	8
3	4	10	2	-15	100	8
4	4	10	2	-15	100	8
1	4	10	2	-10	100	8
2*	4	10	2	-10	100	8
3	4	10	2	-10	100	8
4	4	10	2	-10	100	8

The condition where a full data set was collected on the Titan Krios was the 2 mg/mL sample prepared using a blot force of -10 (Table 6.2, asterisk). Microscope imaging and data collection were performed by Dr. Justin Deme, using the parameters outlined in Table 6.3 and described in Section 2.5.3.1.

Table 6.3. Cryo-EM data collection parameters. Data collection was performed by Justin Deme.

Parameter	Value
Voltage (kV)	300
Magnification	64,000X
Micrographs collected	6913
Electron dose rate ($e^-/Å^2/s$)	7.9
Total dose ($e^-/Å^2$)	47.232
Particles extracted (initial no.)	7,376,621
Defocus range (µm)	-3.0 to -1.0
Pixel size (Å)	0.669 in super resolution; 1.34 real pixel

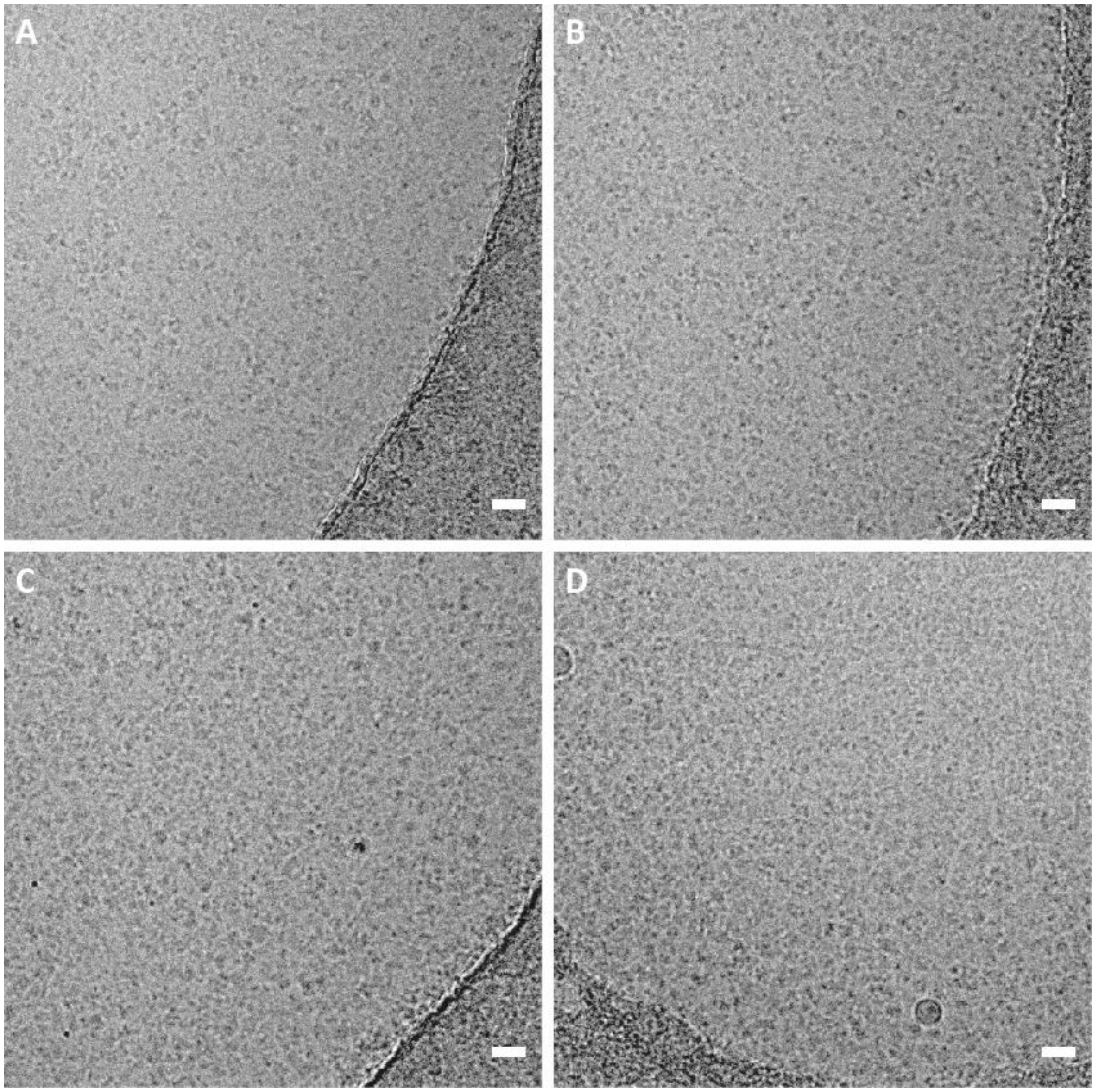


Figure 6.1. *Hs*FRMD8iRhom2Δ300-mVn-TwS grid screening. Prepared cryo-EM grids at different protein concentrations, **(A)** 1 mg/mL at blot force -10, **(B)** 2 mg/mL at blot force -10, **(C)** 3 mg/mL at blot force -10 and **(D)** 4 mg/mL at blot force -15, were imaged on a Talos Arctica by Justin Deme. Scale bar 20 nm, focus -3.0 μm.

6.2.1 2D class averages

After movies (in super-resolution) were collected, they were imported and pre-processed in the SIMPLE 3.0 pipeline (Reboul et al., 2018, Caesar et al., 2020). During this step super-resolution movies were motion-corrected and downsampled to generate micrographs with a real pixel size of 1.34 Å. These micrographs were then used for patched CTF estimation followed by 2D classes

generated from a subset of manually picked particles that were used as references for the particle auto-picking, and particles extracted into 134 Å by 134 Å boxes. Auto-picked, extracted particles were subject to cleanup2D and clusters (300 classes), generated from the cleanup2D, were manually selected. The cleanup2D stage is a fast and lower resolution method that separates the 'good' particles from the 'junk' particles. Examples of 'good' and 'junk' clusters are shown in Figure 6.2. Selected 2-dimersional (2D) clusters were subject to cluster2D (a higher resolution 2D cluster generation step), clusters were selected stringently, and selected clusters were re-extracted (to minimise stack size by inclusion only of selected particles) and exported to RELION (Zivanov et al., 2018, Zivanov et al., 2020). The cluster2D stage generates the best 2D clusters from selected classes from the cleanup2D. Examples of 2D clusters, or class averages, exported to RELION are shown in Figure 6.3A. In all views, the detergent micelle can be seen, and in side views, density within the detergent micelle and density outside of the micelle (Figure 6.3B). Examples of 2D clusters that were not exported to RELION are shown in Figure 6.3C. It is likely that the density outside the micelle was the FRMD8 portion of the chimera however this is only speculative. Another density, on the opposite side of the detergent micelle, representing the mVenus protein was expected however this is not seen. The lack of a second density outside of the detergent membrane could be due a flexible 3C protease site linker between the protein and the mVenus and therefore averaged out. 2D classification produces an early impression of the collected dataset because if the 2D was heterogeneous or there were few plausibly protein classes then the data processing can stopped at this stage. Data processing of large datasets, for example in Table 6.3, requires large computational capacity.

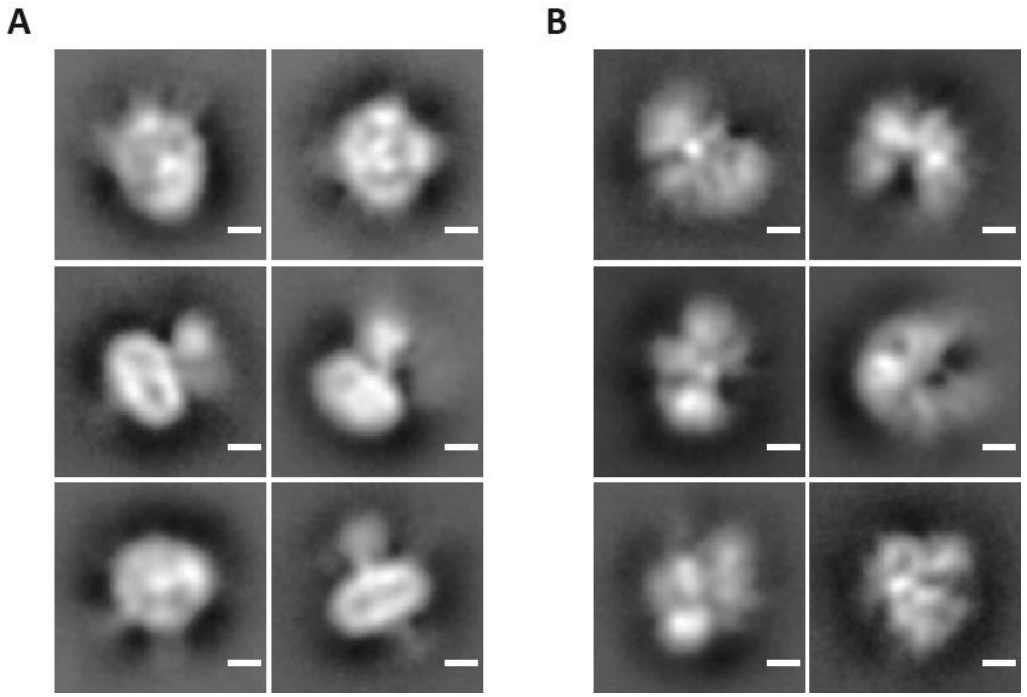


Figure 6.2. Cleanup2D class averages of *Hs*FRMD8iRhom2Δ300-mVn-TwS. Examples of the 'good' **(A)** and 'junk' **(B)** cleanup 2-dimensional (2D) clusters (or averages) generated from the SIMPLE cluster2D. Scale bar represents 30 Å.

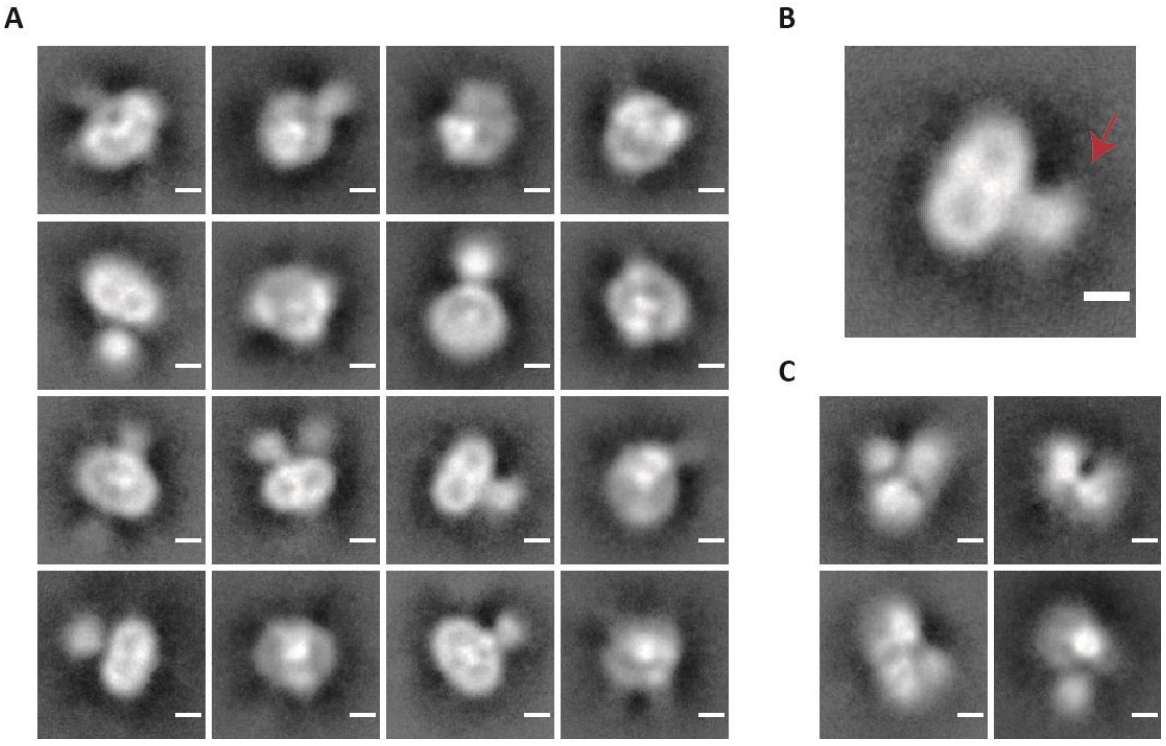


Figure 6.3. 2D class averages of *Hs*FRMD8iRhom2Δ300-mVn-TwS. (A) Examples of the 2-dimensional (2D) class clusters (or averages) generated from the SIMPLE cluster2D that were exported to RELION. Some views show density outside of the detergent micelle, and an enlarged example is shown in **(B)** and highlighted by the red arrow. Examples of poor 2D clusters that was not exported is shown in **(C)**. 2D data processing was performed by Archna Shah and Susan Lea. Scale bar represents 30 Å.

6.2.2 3D class averages

Particles belonging to high resolution 2D clusters were exported to RELION and processed for initial 3-dimensional (3D) reconstruction in RELION 3.1 under the guidance of Prof. Susan Lea as described in Section 2.5.3.2. An overview of the data processing workflow is shown in Figure 6.4A. Three 3D class volumes were produced at a estimated resolution of ~ 9.4 Å and shown in Figure 6.4B-D. Across the three classes, the detergent micelle varied in size and shape. This variability may due to different protein conformations and CHS in detergent micelles are inherently heterogeneous due to their effect on changing the shape of micelles (O'Malley et al., 2011). The molecular weight of *Hs*FRMD8iRhom2 chimera (145.8 kDa) is considered on the smaller scale for SPA. As the *Hs*FRMD8iRhom2 chimera is small, a significant proportion of the mass of the object is the detergent micelle. The 3D reconstruction was aligning based on the detergent micelle which resulted in poor alignments of the embedded protein. Attempts to perform subtraction of the density from the micelled or alignments focussed on the protein by use of a tight protein mask failed to generate better volumes. Despite the limited resolution of the volumes generated, this is an encouraging step forwards for this project as this is the first time a mammalian inactive rhomboid has been imaged by cryo-EM. The varying shape of the detergent micelles corroborates with the size-exclusion data during the purification (Section 5.3.2) whereby *Hs*FRMD8iRhom2 chimera was eluting across a peak and two shoulders.

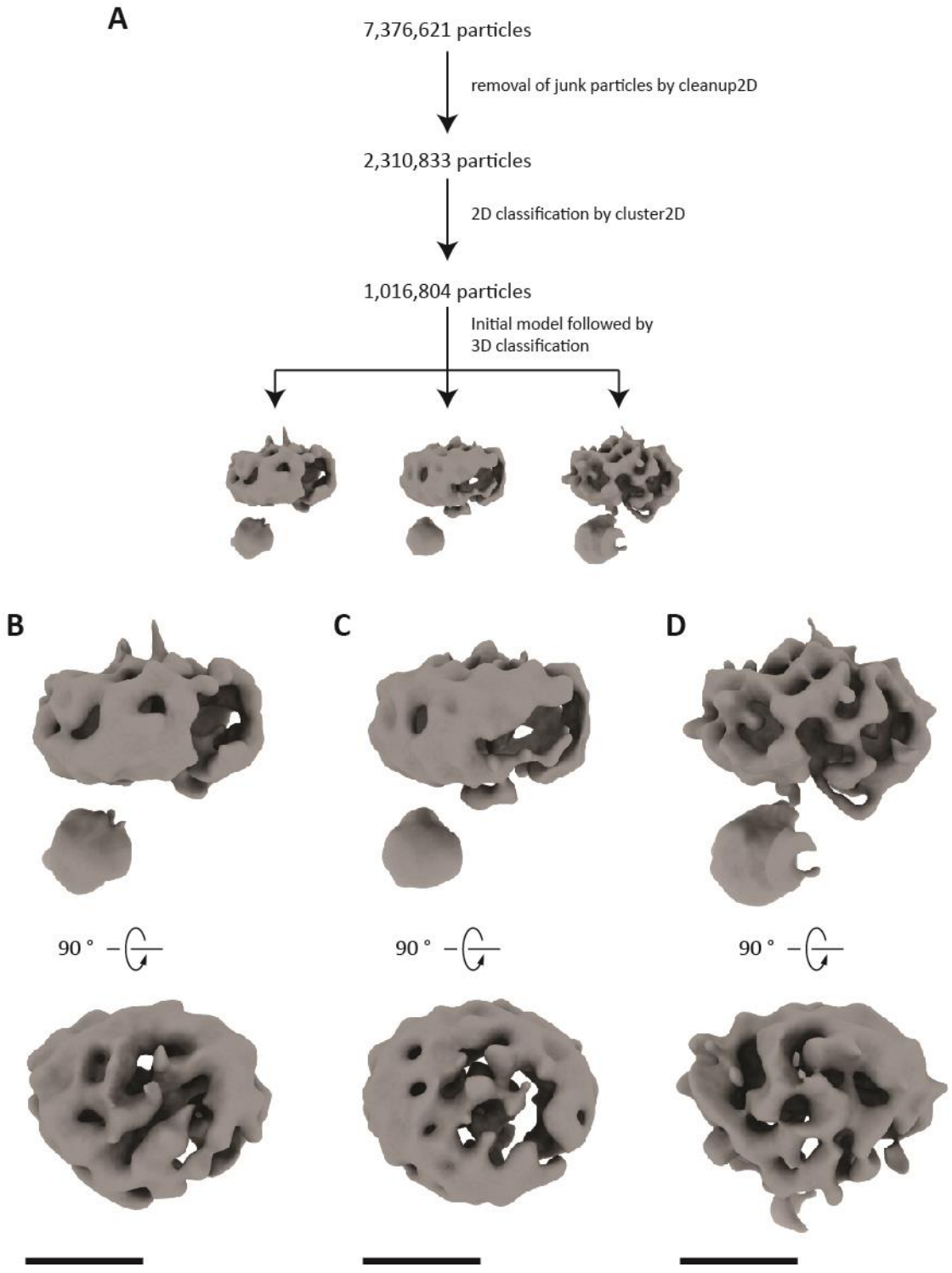


Figure 6.4. Overview of data process to produce 3-dimensional classes of *Hs*FRMD8iRhom2Δ300-mVn-TwS. (A) An overview of the data processing that resulted in the three 3-dimensional (3D) initial classes **(B-D)** were generated in RELION 3.1 by Susan Lea and visualised in ChimeraX. Scale bar represents ~ 44 Å.

6.3 Discussion

Despite the number of particles collected a high resolution structure was not attainable. This is likely to be a combination of sample quality and the size of the protein. The chimera, in future, could be prepared in LMNG alone rather than with CHS and/or redesign of the construct to remove amino acids at the C-terminus of the *Hs*FRMD8iRhom2 before the 3C cleavage site. The *Hs*FRMD8iRhom2Δ300 was chosen as a target because of its stability, its higher yields from purification relative to the purification trials of the iRhom/ADAM17 complex, and was the construct to produce results faster than the complex. Given the molecular weight of the chimera, there are very few protein structures by SPA that have been solved that are less than 150 kDa therefore it will be prudent to re-evaluate the structural target for an inactive rhomboid protease. This is further discussed in the next Section 6.4.

6.4 Thesis summary

6.4.1 Conclusions

In this thesis numerous HEK cell lines were generated either expressing iRhom homologues alone or with their ADAM17 with the goal to express and purify for structural determination. And with the progression of the *Hs*FRMD8iRhom2 chimera faster than that of the complex, the chimera was imaged in ice and low resolution 3D reconstructions were generated. These maps are a proof of principle that a mammalian inactive rhomboid protease can be imaged in ice but sample and sample preparation needs to be reassessed. However the ground work of generating the cell lines and finding the best detergent, LMNG, for protein and complex extraction was completed.

When analysing the cryo-EM SPA protein maps in the EMDB (EMBL-EBI, 2020a), there is an explosion of structures solved above 200 kDa. Therefore adding a binding partner will increase the size of the object imaged which may increase the quality of the data collected. Progressing

with the iRhom/ADAM17 will be the best way forward for solving the structure of iRhom. Shown in Chapter 5 were promising fluorescence size-exclusion data showing the mouse iRhom2/ADAM17 complex, expressed in HEK293F cells, co-eluted and the complex confirmed by Western blot. However the complex co-eluted after two affinity and one size-exclusion chromatography steps and resulted in an unmeasurably low yield. To address the low yield, a large amount of isolated membranes is needed to begin with or to use a higher expressing cell line.

A candidate high expressing mammalian cell line is the Expi293F GnTI- (Thermo Fisher Scientific). A month prior to the United Kingdom's Lockdown Restrictions in March 2020, the same lentiviral constructs used in the HEK293F cell line generation were successfully transduced in the Expi293F GnTI- cells and enriched by FACS. However, these cells have not been verified for the expression and sub cellular localisation of the complex.

6.4.2 Future directions

Experiments to follow this thesis should be first to verify expression of the iRhom and ADAM17 in these newly generated Expi293F cell lines and to show co-localisation/subcellular localisation of the complex. If complex is present and co-localised, the cell lines should be expanded in excess of 10-15 L of cell culture for membrane isolation. The complex should be extracted with LMNG/CHS and purified by the three-step purification method, Step-Tactin, Ni-NTA and SEC. Purified complex should be first imaged on negative stain grids to confirm little to no protein aggregation before preparing cryo grids. Ultimately the results from the screening and data collection of the complex in ice will determine the progression to a high resolution structure. In addition to purifying the complex in detergent, the complex should also imaged in nanodiscs. In nanodiscs, the protein complex will be in a more native environment and therefore may be more stable than in detergent. Further, imaging the complex in nanodiscs may improve the alignments

in the data processing. With larger yields of protein, exogenous binders such as Fabs and megabodies could be generated to aid protein alignments as discussed in Section 6.1. In addition, nanobodies generated against the iRhom/ADAM17 complex may aid in its stabilisation particularly when plunge frozen on cryo grids.

If a high resolution structure is obtained, it will reveal the biding sites between iRhom and ADAM17 and therefore perform mutagenesis of these to binding sites to abrogate binding. Further, the structure of the complex will enable to develop a mechanism to inhibit the interaction by designing short hydrophobic peptides based on the structure determined. As part of the large scale membrane preparations for protein purifications, isolated membranes consisting of over expressed iRhom1 or iRhom2 may be used for native mass spectrometry of whole membranes to identify unknown binding partners of iRhom. The results that have been achieved in this thesis have formed the foundation of expressing and purifying the iRhom/ADAM17 complex and will hopefully lead to a high resolution structure.

Appendices

Appendix A: Primers used in this thesis

Table A1. Primers used in this thesis. Primers that were used in this thesis for cloning and sequencing are listed with their description, orientation (5'-3', F and 3'-5' R), annealing temperature (Ta), number of base pairs (bp) and guanine/cytosine content (GC).

Primer no and description	Primer name	Sequence	F/R	Ta (°C)	bp	GC (%)
Gibson primers to insert Hs*IRHD1 into pETM14 between NcoI and NotI via PCR*						
1	1_F1_HsIRHD1_NcoI	ttctgttccaggggcccatgGGCAGCCAGCATGAAACC	F	68	38	61
2	2_R2_HsIRHD1_NotI	tggtgctcgagtgcggccgcTTAACGATAAAACTGATCCGGCAC	R	65	44	57
3	3_F_pETM14_NotI	GCGGCCGCACTCGAGCAC	F	76	18	78
4	4_R_pETM14_NcoI	CATGGGCCCCTGGAACAG	R	69	18	67
Gibson primers to insert Hs*IRHD1 into pET15b without his tag via PCR*						
5	5_F5_HsIRHD1	ctttgttagcagccggatccTTAACGATAAAACTGATCCGGC	F	62	42	48
6	6_R6_HsIRHD1	gaaggagatataccatgggcAGCCAGCATGAAACCGTTG	R	66	39	51
7	7_F7_pET15b	GCCCATGGTATATCTCCTTC	F	62	20	50
8	8_R8_pET15b	GGATCCGGCTGCTAACAAAG	R	66	20	55
Sequencing primers Dr*iRhom*						
24	24_DriR_1	ctctggctaactagagaaccc	F	56	21	52
25	25_DriR_2	cagaggcagagccttcagaatg	F	59	22	55
26	26_DriR_3	tgaagcaagaggtggtgtcc	F	57	20	55
27	27_DriR_4	agacagcatggctctgtgtg	F	60	20	55
28	28_DriR_5	cctgtctgttcgtggaactg	F	57	20	55
Sequencing primers Hd*iRhom*						
29	29_HdiR_1	ggctaactagagaacccactg	F	56	21	52
30	30_HdiR_2	ccgagatggcttggagaaag	F	57	20	55
31	31_HdiR_3	tcatgagcagattctttggcgg	F	59	22	50
32	32_HdiR_4	tcacacctggcctgacgatatcac	F	61	24	54
33	33_HdiR_5	ctggatgcctgtgctgagaatg	F	60	22	55
Primers for sequencing UK_96 FRMD8-iRhom2 fusion						
49	49_pLEX_F	TCCATAGAAGACACCGACTCTAC	F	56	23	48
50	50_pLEX_R	ACACCGGCCTTATTCCAAGCG	R	62	21	57
Sequencing primers Am*iRhom*						
63	63_AmiR_1	CTCTGGCTAACTAGAGAACCC	F	56	21	52
64	64_AmiR_2	CATGCAGAAGATCATCGACC	F	55	20	50
65	65_AmiR_3	AAGTGCGGCTGAGACAAGAG	F	59	20	55
66	66_AmiR_4	CCTACCATCCTAGCACACCTAC	F	57	22	55
67	67_AmiR_5	CCGGCTCTCAGTTTGGAATC	F	57	20	55

Sequencing primers Hs*iRhom1*						
68	68_HsiR1_1	CTCTGGCTAACTAGAGAACCC	F	56	21	52
69	69_HsiR1_2	TGATGAAAGGCCGCTCCGTTAG	F	60	22	55
70	70_HsiR1_3	AAGCGCATCGACAGCTTCGTCAAG	F	64	24	54
71	71_HsiR1_4	GGGCCACAAGAGACAGTTTG	F	56	20	55
72	72_HsiR1_5	GTGCCTTCTTCAAGCTGCTG	F	57	20	55
Sequencing primers Hs*iRhom2*						
73	73_HsiR2_1	CTCTGGCTAACTAGAGAACCC	F	56	21	52
74	74_HsiR2_2	AAAGCCCTGCAAGATGCCCAAG	F	62	22	55
75	75_HsiR2_3	TTGATCGGAAGAAGCGGCACTACG	F	62	24	54
76	76_HsiR2_4	ACATCTGGCCCGATGACATCAC	F	61	22	55
77	77_HsiR2_5	GCCATCGTGCTCTTCCTGTTCATC	F	62	24	54
Sequencing primers Mm*iRhom2*						
78	78_MmiR2_1	CTCTGGCTAACTAGAGAACCC	F	56	21	52
79	79_MmiR2_2	GGTCCTGTCTCTCACATCCTTC	F	58	22	55
80	80_MmiR2_3	TGTCGTGGGTAACTGGCTCAAC	F	60	22	55
81	81_MmiR2_4	CCACATCTGGCCTGATGACATTAC	F	59	24	50
82	82_MmiR2_5	TTCGTGGAGCTGTTCCAGAG	F	57	20	55
Sequencing primers for Hs*FRMD8iRhom2 chimera*						
87	87_FRMDhiR2_1	CTCTGGCTAACTAGAGAACCC	F	56	21	52
88	88_FRMDhiR2_2	GTATACCTAGCGGATGACACG	F	56	21	52
89	89_FRMDhiR2_3	CATGGATGAGCCTTTCCTGC	F	57	20	55
90	90_FRMDhiR2_4	CCTATCTCCTCAAGTGCCAC	F	55	20	55
91	91_FRMDhiR2_5	AAGCTCCTCAAGATCTACTCC	F	55	21	48
92	92_FRMDhiR2_6	TTGATCGGAAGAAGCGGCACTACG	F	62	24	54
93	93_FRMDhiR2_7	ACTCAGGCTGCTGTGTCCAGAATG	F	63	24	54
94	94_FRMDhiR2_8	CACTGCTTGGACAAGGTGTG	F	58	20	55
95	95_FRMDhiR2_9	TGCCCTGGATCGACAACATC	F	57	20	55
Linearising pHR-CMV-TetO2-3C-Avi-His6						
98	98_pHR-Avi-His6_F	ctggaggtgctgttccagg	F	59	19	63
99	99_pHR-Avi-His6_R	ggtggcaagcttgaattctctagac	R	60	25	48
Inserting FRMD8 into pHR-CMV-TetO2-3C-Avi-His6						
100	100_FRMD8_F	tgatctcaacaagctgtctagagaattcaagcttgccaccATGGACGGGACAGAAGGC	F	59	58	50
101	101_FRMD8_R	aaccaccggaacctccggggccctggaacagcacctccagGCCCTGCTCCAGGCTGtc	R	61	58	69

Linearising pHR-CMV-TetO2-3C-TwS

102	102_pHR-TwS_F	CTTGAGGTGCTGTTTCAGGGAC	F	59	22	55
103	103_pHR-TwS_R	GGTGGCAAGCTTGAATTCTCTAGAC	R	60	25	48

Inserting iRhoms into pHR-CMV-TetO2-3C-TwS and pHR-CMV-TetO2-3C-mVenus-TwS

104	104_AmiRhom_F	GAGAATTCAAGCTTGCCACCATGAGCGAGGCCAGACAC	F	60	38	55
105	105_AmiRhom_R	CCCTGAAACAGCACCTCAAGGTGCAGCTGGGC	R	64	32	63
106	106_DriRhom_F	GAGAATTCAAGCTTGCCACCATGGCCGAGCTGAGAAGAG	F	60	39	54
107	107_DriRhom_R	CCCTGAAACAGCACCTCAAGATGGAGGTGGGC	R	62	32	59
108	108_HdiRhom_F	GAGAATTCAAGCTTGCCACCATGGCCGACGGCGAT	F	62	35	57
109	109_HdiRhom_R	CCCTGAAACAGCACCTCAAGGTCGGTGGGCTT	R	63	32	59
110	110_HsiRhom1_F	GAGAATTCAAGCTTGCCACCATGAGTGAGGCC	F	61	21	53
111	111_HsiRhom1_R	CCCTGAAACAGCACCTCAAGGTGGAGCTGAGC	R	62	32	59
112	112_HsiRhom2_F	GAGAATTCAAGCTTGCCACCATGGCCTCTGCTGA	F	62	34	53
113	113_HsiRhom2_R	CCCTGAAACAGCACCTCAAGGTGCAGCACCTGgtc	R	65	35	60
114	114_MmiRhom2_F	GAGAATTCAAGCTTGCCACCATGGCCTCAGCTG	F	63	33	55
115	115_MmiRhom2_R	CCCTGAAACAGCACCTCAAGGTGTAGCACCTG	R	60	32	56

Sequencing primers for pHR vectors for gene insertion

116	116_CMV_F	CGCAAATGGGCGGTAGGCGTG	F	65	21	67
117	117_pHR_R	gatatcaagcttgcatgcctg	R	55	21	48
118	118_WPRE_R_seq	caaattttgtaatccagagg	R	49	20	39
119	119_WPRE_R	CATAGCGTAAAAGGAGCAACA	R	55	21	43
120	120_EGFP/mVenus_N_R	CGTCGCCGTCCAGCTCGACCAG	R	67	22	73
121	121_IRES_R_seq	gtggaaaataacatatagac	R	45	20	30
122	122_IRES_rev	GCATTCCTTTGGCGAGAG	R	56	18	56
123	123_IRES_F	TGGCTCTCCTCAAGCGTATT	F	57	20	50
124	124_IRES_F_seq	cacatgctttacatgtgtttag	F	53	22	36

Sequencing primers for pHR vectors

197	197_Amp-R	ATAATACCGCGCCACATAGC	R	58	20	50
198	198_EBV-rev	GTGGTTTGTCCAAACTCATC	F	54	20	45
199	199_EGFP-C	CATGGTCCTGCTGGAGTTCGTG	F	54	22	59
200	200_EXFP-R	GTCTTGTAGTTGCCGTCGTC	R	57	20	55

201	201_L4440	AGCGAGTCAGTGAGCGAG	F	58	18	61
202	202_pBABE-3prime	ACCCTAACTGACACACATTCC	R	56	21	48
203	203_pBR322ori-F	GCATTCCTTTGGCGAGAG	F	56	20	50
204	204_pBRforEco	AATAGGCGTATCACGAGGC	R	57	19	53
205	205_pBRrevBam	GGTGATGTCGGCGATATAGG	R	56	20	55
206	206_pGEX-3prime	CCGGGAGCTGCATGTGTCAGAGG	F	66	23	65
207	207_pRS-marker	CGGCATCAGAGCAGATTGTA	R	54	20	50
208	208_SP6	ATTTAGGTGACACTATAG	R	44	18	33
209	209_SV40pA-R	GAAATTTGTGATGCTATTGC	R	51	20	35
210	210_SV40pro-F	TATTTATGCAGAGGCCGAGG	F	56	20	50

Sequencing primers for pcDNA-Hs*ADAM17-HA*

213	213_HsADAM17_F	CACTGCTTACTGGCTTATCG				
214	214_HsADAM17_R	TGGCAACTAGAAGGCACAG				

Sequencing primers Hs*ADAM17*

215	215_A17_1	ATGAGGCAGTCTCTCCTATTC	F	55	21	48
216	216_A17_2	CAAGAATGTTTCACGTTTGCAGTC	F	58	24	42
217	217_A17_3	TGATATGGGAACTCTTGGATTAGC	F	56	24	42
218	218_A17_4	TATCTGAACAACGACACCTGCTG	F	58	23	48
219	219_A17_5	TTTAGCAGACAACATCGTTGG	F	55	21	43

Inserting ADAM17 into pHR-CMV-TetO2-3C-Avi-His6
For linarising the vector, use primers 98, 99

224	224_pHRA17_F	aagctgtctagagaattcaagcttgccaccATGAGGCAGTCTCTCCTATTCCT	F	59	53	47
225	225_pHRA17_R	gaacctccggggccctggaacagcacctccagGCACTCTGTTTCTTTGCTGTCAACACGATTC	R	63	63	57

Inserting Gg*iRhoms into pHR-CMV-TetO2-3C-mVenus-TwS*

238	238_GgiRhom1_F	tgatctcaacaagctgtctagagaattcaagcttgccaccATGTCTGAGGCCAGACACGATAGCACAAG	F	65	69	48
239	239_GgiRhom1_R	tcaccatactacctcctggtccctgaaacagcacctcaagATGCAGCTGGGCGTCCAGGTCGTACTTCTC	R	70	70	56
240	240_GgiRhom2_F	tgatctcaacaagctgtctagagaattcaagcttgccaccATGGCCAGCAGCAACAAGAATGGCAGCAG	F	69	69	49
241	241_GgiRhom2_F	tcaccatactacctcctggtccctgaaacagcacctcaagGTGCAGCACCTGTTCCAGCTCGTACTTCTC	R	66	70	54

ADAM17s into pHR-CMV-TetO2-3C-Avi-His6

242	242_GgADAM17_F	tgatctcaacaagctgtctagagaattcaagcttgccaccATGAGACTGAGACTGTGGCTGGTGTCTG	F	65	68	49
243	243_GgADAM17_R	aaccaccggaacctccggggccctggaacagcacctccagGCACTCGGTTTCTTTGCTGTCCACTCTATTCTG	R	64	73	59
244	244_AmADAM17_F	tgatctcaacaagctgtctagagaattcaagcttgccaccATGAGGCCTAGACTGTCTCTGTGTGCTCTG	F	66	70	48
245	245_AmADAM	aaccaccggaacctccggggccctggaacagcacctccag	R	68	71	62

	17_R	GCACTCGGTTTCTTTGCTGTCCACGCGATTC				
246	246_MmADAM17_F	tgatctcaacaagctgtctagagaattcaagcttgccaccATGAGGCGGCGTCTCCTATTCCTGACCACTTTG	F	68	73	49
247	247_MmADAM17_R	aaccaccggaacctccggggccctggaacagcacctccagGCACTCTGTCTCTTTGCTGTCAACTCGGCTCTG	R	67	73	62
Inserting mScarlet-I in place of Avi tag						
250	250_mScarletI_F	ggaggtgctgttccagggccccggaggttccggtggttccatgagtaaaggagaagctgtgattaaagagtTCATG	F	61	76	54
251	251_mScarletI_R	gtgatcattagtgatggtgatggtggtgcttggtacgtccttttgtatagttcatccatgccaccggtAGAATGACG	R	64	76	47
Sequencing primers Mm*ADAM17*						
252	252_MmADAM17_1	CGTCTCCTATTCCTGACCAC	F	56	20	55
253	253_MmADAM17_2	ACATAGAGCCACTTTGGAGGTTTG	F	60	24	46
254	254_MmADAM17_3	AAACCTGGTGAAAGACACTTC	F	55	21	43
255	255_MmADAM17_4	AAGTGCCAGGAGGCTATTAATGC	F	59	23	48
256	256_MmADAM17_5	TGGATTCCTTTCAGCATTCTTG	F	56	22	41
Sequencing primers Gg*iRhom1*						
257	257_GgiRhom1_1	TCTGAGGCCAGACACGATAG	F	58	20	55
258	258_GgiRhom1_2	CTACAGCATCCCTCACACAC	F	56	20	55
259	259_GgiRhom1_3	CCCAGAAGAGAGGCCAGAGAATTG	F	60	24	54
260	260_GgiRhom1_4	GAGGAAGAGTGTAGCTCCACACTG	F	60	24	54
261	261_GgiRhom1_5	ATGGCACAGAATCAGCATCATC	F	58	22	45
Sequencing primers Gg*iRhom2*						
262	262_GgiRhom2_1	GGCCAGCAGCAACAAGAATG	F	60	20	55
263	263_GgiRhom2_2	ATAGACCTCACACACCTCAC	F	55	20	50
264	264_GgiRhom2_3	TGCTCCTCAGAGAGGTAGAC	F	56	20	55
265	265_GgiRhom2_4	TGCCACAGGATTGCTCTGAG	F	59	20	55
266	266_GgiRhom2_5	ATGGCTGCGGATCAGCATCATC	F	62	22	55
Sequencing primers GgADAM17						
267	267_GgADAM17_1	ATGAGACTGAGACTGTGGCTGGTG	F	62	24	54
268	268_GgADAM17_2	GTGTACAGAAGCGAGGACATCAAG	F	59	24	50
269	269_GgADAM17_3	TGGCTCACCTGTTCACCTAC	F	58	20	55
270	270_GgADAM17_4	GCCGATTGCAAGCTGAAGGATG	F	61	22	55

271	271_GgADAM17_5	CAAGTTCCTGGCCGACAACATC	F	60	22	55
Sequencing primers Am*ADAM17*						
272	272_AmADAM17_1	GCCTAGACTGTCTCTGTGTG	F	56	20	55
273	273_AmADAM17_2	TGCTGGTGTACAGAAGCGAG	F	59	20	55
274	274_AmADAM17_3	CCACCTGTTCACCTACCAAG	F	56	20	55
275	275_AmADAM17_4	TGCACCAGCACGATGATCCTTG	F	62	22	55
276	276_AmADAM17_5	AGCATCAATACCTTCGGCAAG	F	58	21	48
Sequencing primers mScarlet-I						
280	280_mScarlet-I_1	AGCTGTGATTAAAGAGTTCATGCG	F	58	24	42
281	281_mScarlet-I_2	AAAGACGATGGGTTGGGAGG	F	59	20	55
282	282_mScarlet-I_3	TTTAGTCACTTTCAGTTTGGCG	R	56	22	41

Appendix B: Automated immunofluorescence image processing script

Script was written by Joseph Hogg for processing immunofluorescence data.

```
#@ File    (label = "Input directory", style = "directory") srcFile
#@ File    (label = "Output directory", style = "directory") dstFile
#@ String  (label = "File extension", value=".tif") ext
#@ String  (label = "File name contains", value = "") containString
#@ String (choices={"None","1","2","3","4"}, label = "Red channel") red_c
#@ String (choices={"None","1","2","3","4"}, label = "Green channel") green_c
#@ String (choices={"None","1","2","3","4"}, label = "Blue channel") blue_c
#@ String (choices={"None","1","2","3","4"}, label = "Grey channel") grey_c
#@ boolean (label = "Keep directory structure when saving", value = true) keepDirectories
#@ boolean (label = "Output individual channels as 8-bit grayscale", value = false) outputChannels

# See also Process_Folder.ijm for a version of this code
# in the ImageJ 1.x macro language.

import os
from ij import IJ, ImagePlus
from ij.plugin import ChannelSplitter, RGBStackMerge, RGBStackConverter
from ij.process import ImageConverter
from loci.plugins import BF
from loci.plugins.in import ImporterOptions

def run():
  srcDir = srcFile.getAbsolutePath()
  dstDir = dstFile.getAbsolutePath()
  for root, directories, filenames in os.walk(srcDir):
    filenames.sort();
    for filename in filenames:
      # Check for file extension
      if not filename.endswith(ext):
        continue
      # Check for file name pattern
      if containString not in filename:
        continue
      process(srcDir, dstDir, root, filename, keepDirectories)

def process(srcDir, dstDir, currentDir, fileName, keepDirectories):

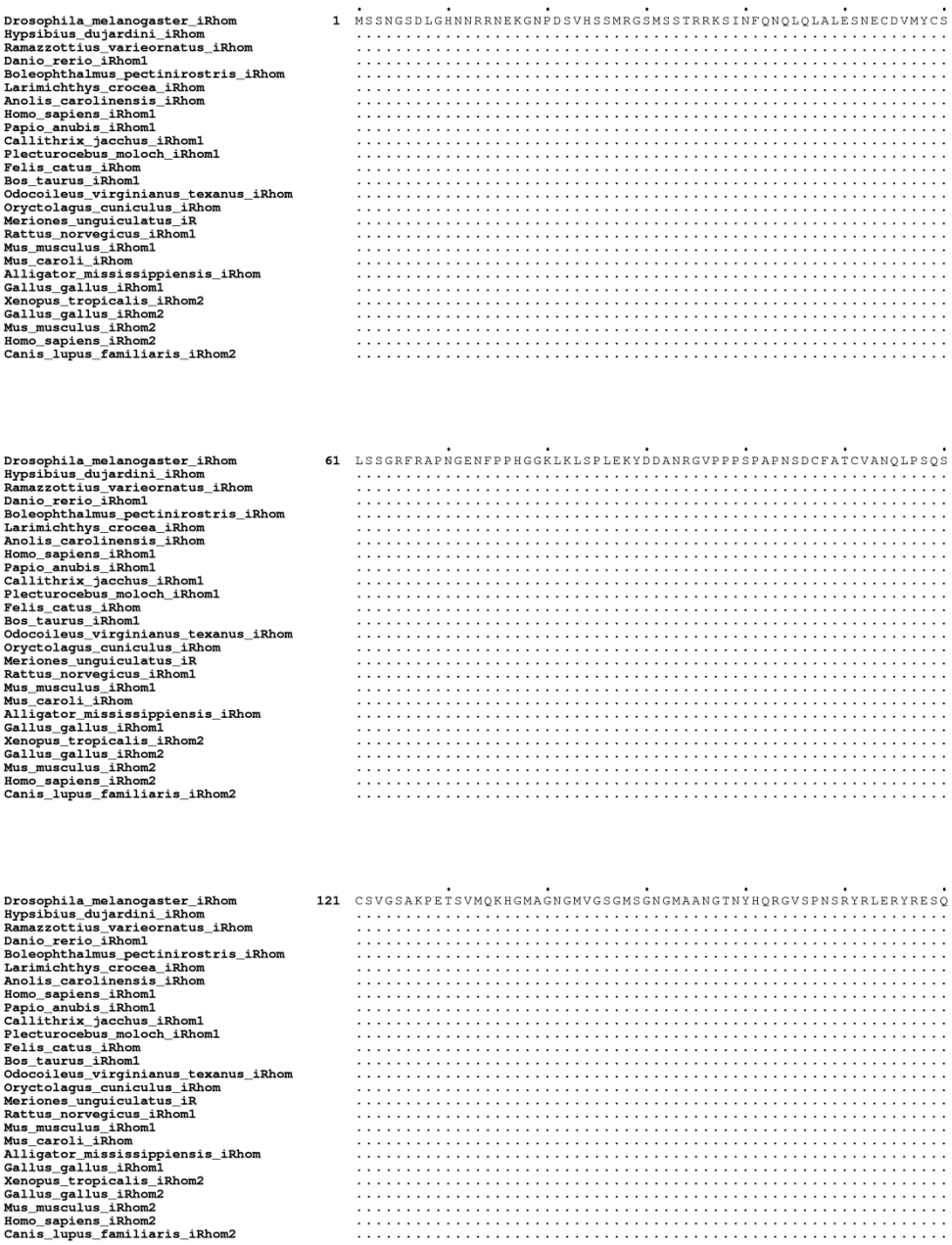


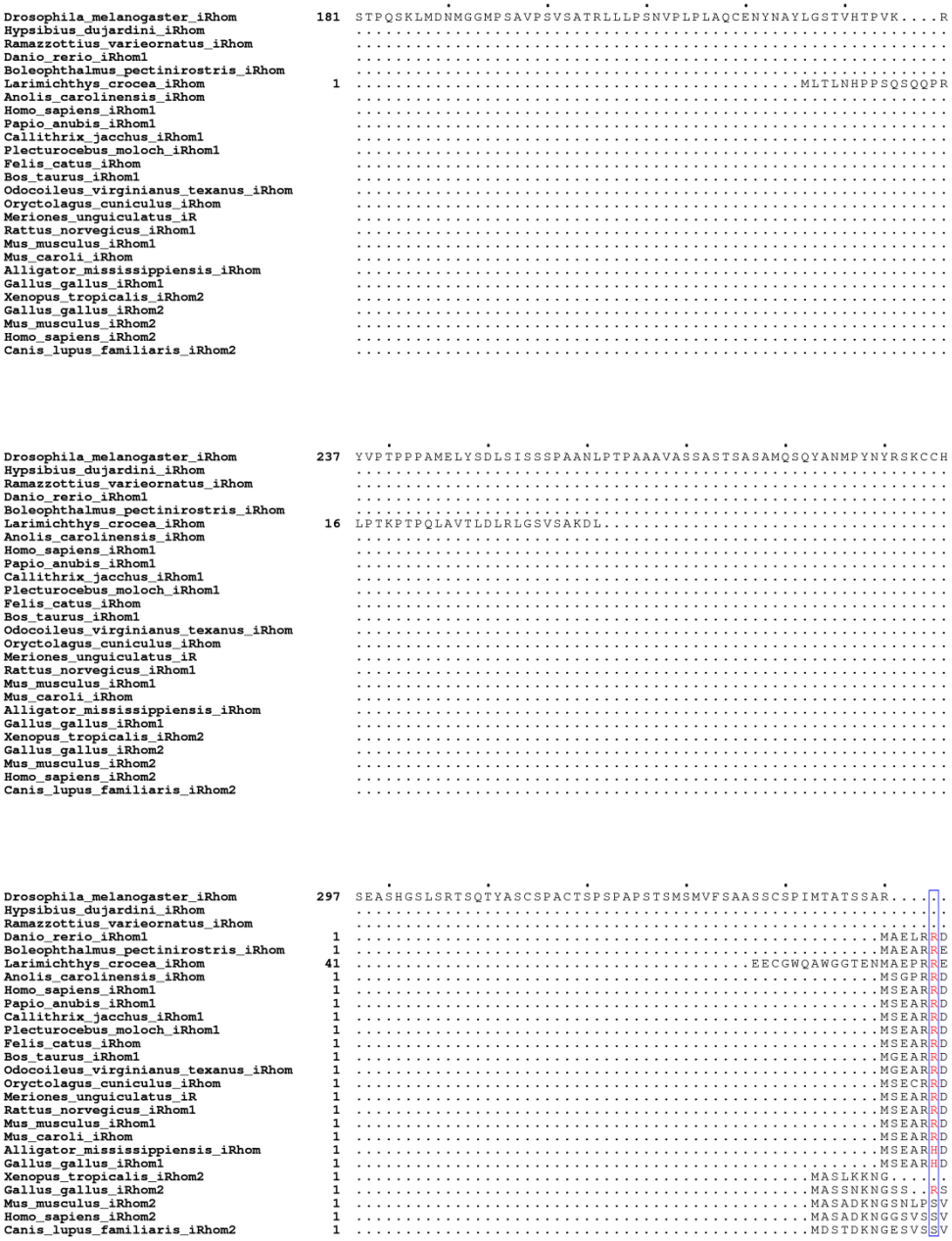
  print "Processing:"

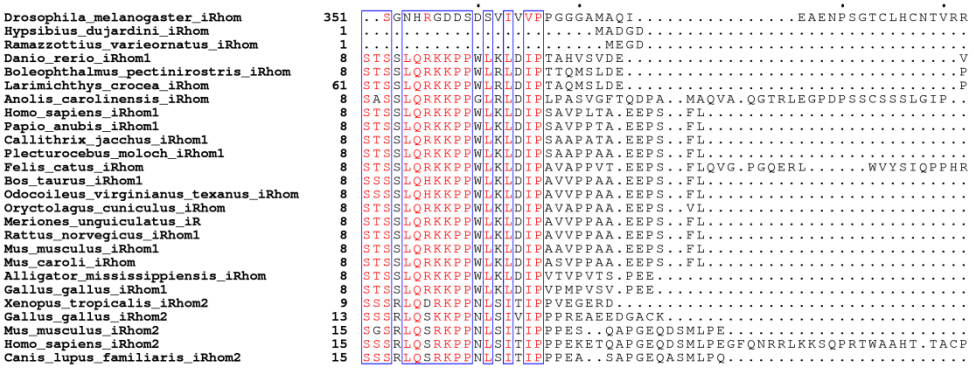
  # Opening the image
  print "Open image file", fileName
  options=ImporterOptions()
  options.setColorMode(ImporterOptions.COLOR_MODE_GRAYSCALE)
  options.setId(os.path.join(currentDir, fileName))
  imp = BF.openImagePlus(options)[0]
```

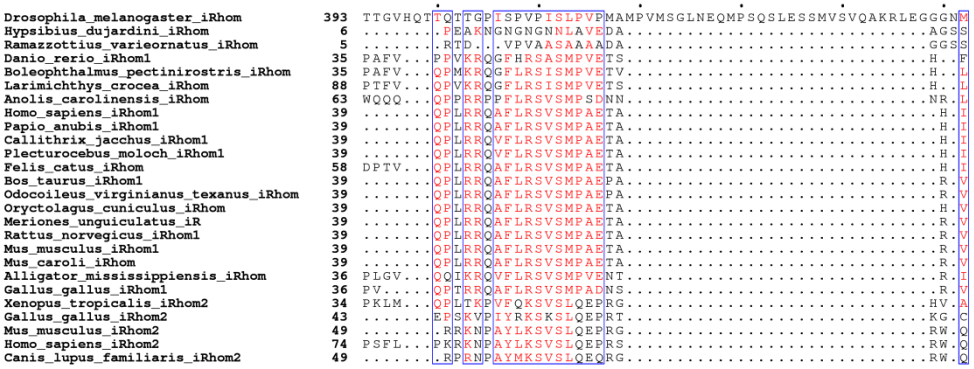
Appendix C: Multiple sequence alignment of iRhoms

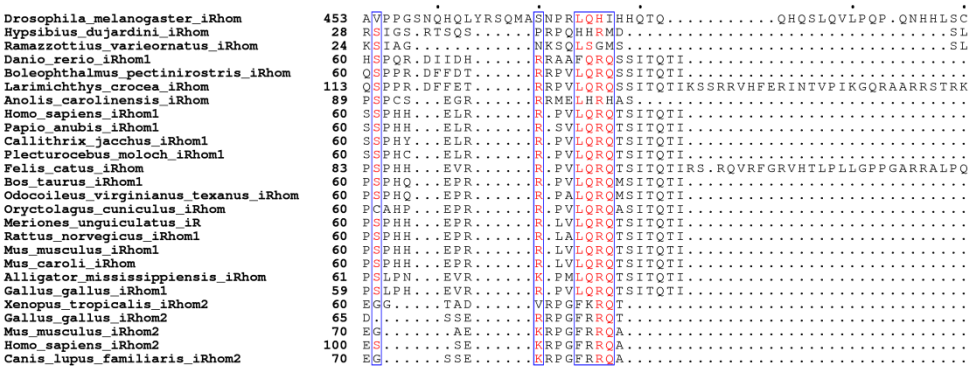
Full length multiple sequence alignment (MSA) of iRhom protein sequences shown in Table 3.1 in Chapter 3. The MSA was performed used Clustal Omega (Sievers et al., 2011, Madeira et al., 2019). ESPript Server was used to visualise the full length MSA (Robert and Gouet, 2014).

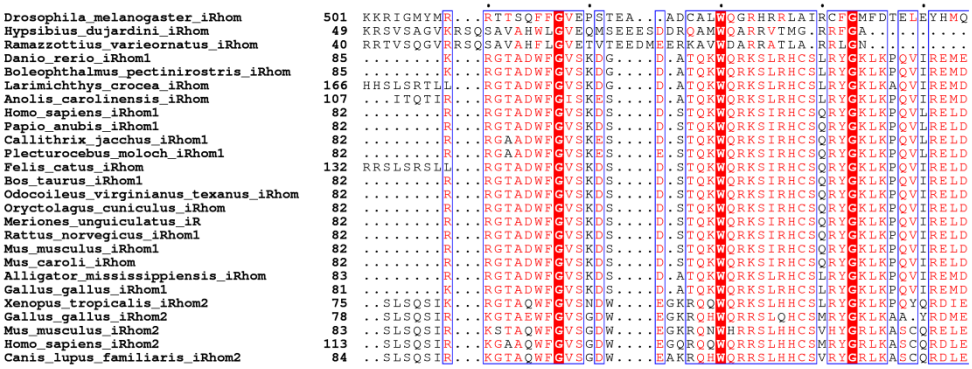


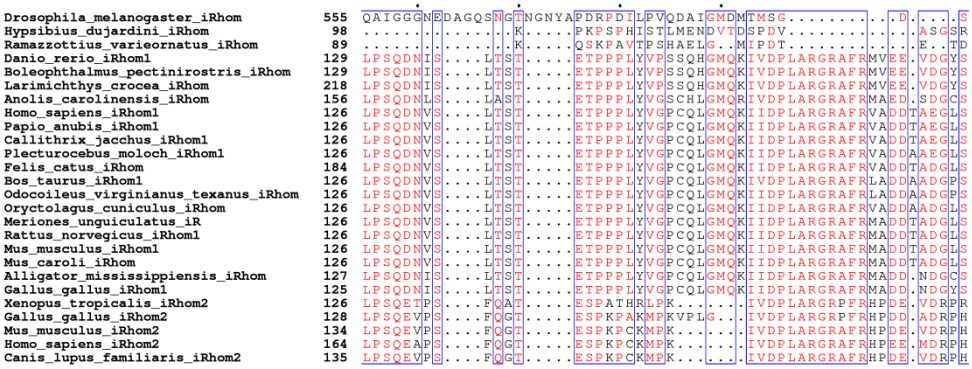

Drosophila_melanogaster_iRhom 181 STPQSKLMDNMGGMPSAVPSVSATRLLLPSNVPLPLAQCENYNAYLGSTVHTPVK....R
Hypsibius_dujardini_iRhom
Ramazzottius_varieornatus_iRhom
Danio_rerio_iRhom1
Boleophthalmus_pectinirostris_iRhom
Larimichthys_crocea_iRhom 1 .....MLTLNHPPSQSQQPR
Anolis_carolinensis_iRhom
Homo_sapiens_iRhom1
Papio_anubis_iRhom1
Callithrix_jacchus_iRhom1
Plecturocebus_moloch_iRhom1
Felis_catus_iRhom
Bos_taurus_iRhom1
Odocoileus_virginianus_texanus_iRhom
Oryctolagus_cuniculus_iRhom
Meriones_unguiculatus_iR
Rattus_norvegicus_iRhom1
Mus_musculus_iRhom1
Mus_caroli_iRhom
Alligator_mississippiensis_iRhom
Gallus_gallus_iRhom1
Xenopus_tropicalis_iRhom2
Gallus_gallus_iRhom2
Mus_musculus_iRhom2
Homo_sapiens_iRhom2
Canis_lupus_familiaris_iRhom2
Drosophila_melanogaster_iRhom 237 YVPTPPPAMELYSDLSISSSPAANLPTPAAAVASSASTSASAMQSQYANMPYNYRSKCCH
Hypsibius_dujardini_iRhom
Ramazzottius_varieornatus_iRhom
Danio_rerio_iRhom1
Boleophthalmus_pectinirostris_iRhom
Larimichthys_crocea_iRhom 16 LPTKPTPQLAVTLDLRLGSVSAKDL.....
Anolis_carolinensis_iRhom
Homo_sapiens_iRhom1
Papio_anubis_iRhom1
Callithrix_jacchus_iRhom1
Plecturocebus_moloch_iRhom1
Felis_catus_iRhom
Bos_taurus_iRhom1
Odocoileus_virginianus_texanus_iRhom
Oryctolagus_cuniculus_iRhom
Meriones_unguiculatus_iR
Rattus_norvegicus_iRhom1
Mus_musculus_iRhom1
Mus_caroli_iRhom
Alligator_mississippiensis_iRhom
Gallus_gallus_iRhom1
Xenopus_tropicalis_iRhom2
Gallus_gallus_iRhom2
Mus_musculus_iRhom2
Homo_sapiens_iRhom2
Canis_lupus_familiaris_iRhom2
Drosophila_melanogaster_iRhom 297 SEASHGSLSRTSQTYASCSPACTSPSPAPSTSMSMVFSAASSCSPIMTATSSAR....
Hypsibius_dujardini_iRhom
Ramazzottius_varieornatus_iRhom
Danio_rerio_iRhom1 1 .....MAELRRD
Boleophthalmus_pectinirostris_iRhom 1 .....MAEARRE
Larimichthys_crocea_iRhom 41 .....EECGWQAWGGTENMAEPRRE
Anolis_carolinensis_iRhom 1 .....MSGPRRD
Homo_sapiens_iRhom1 1 .....MSEARRD
Papio_anubis_iRhom1 1 .....MSEARRD
Callithrix_jacchus_iRhom1 1 .....MSEARRD
Plecturocebus_moloch_iRhom1 1 .....MSEARRD
Felis_catus_iRhom 1 .....MSEARRD
Bos_taurus_iRhom1 1 .....MGEARRD
Odocoileus_virginianus_texanus_iRhom 1 .....MGEARRD
Oryctolagus_cuniculus_iRhom 1 .....MSECRRD
Meriones_unguiculatus_iR 1 .....MSEARRD
Rattus_norvegicus_iRhom1 1 .....MSEARRD
Mus_musculus_iRhom1 1 .....MSEARRD
Mus_caroli_iRhom 1 .....MSEARRD
Alligator_mississippiensis_iRhom 1 .....MSEARHD
Gallus_gallus_iRhom1 1 .....MSEARHD
Xenopus_tropicalis_iRhom2 1 .....MASLKKNG....
Gallus_gallus_iRhom2 1 .....MASSNKNGSS...RS
Mus_musculus_iRhom2 1 .....MASADKNGSNLPSV
Homo_sapiens_iRhom2 1 .....MASADKNGGSVSSV
Canis_lupus_familiaris_iRhom2 1 .....MDSTDKNGESVSSV

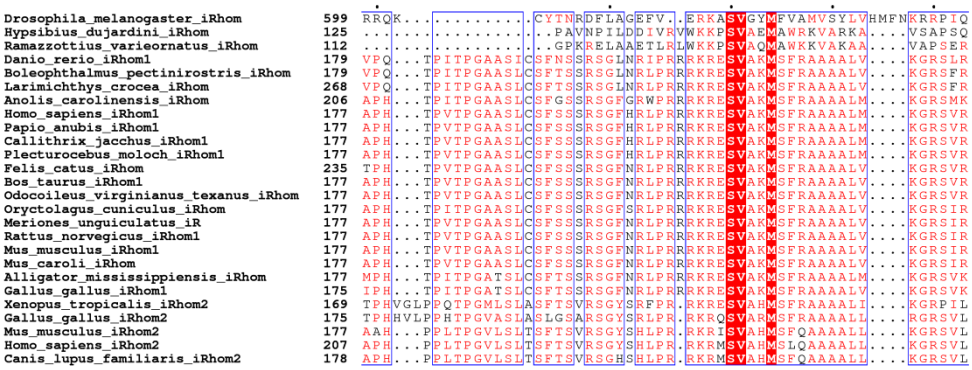

Drosophila_melanogaster_iRhom 351 ..SGNHRGDDSDSVIVVPPGGGAMAQI.....EAENPSGTCLHCNTVRR
Hypsibius_dujardini_iRhom 1 .....MADGD.....
Ramazzottius_varieornatus_iRhom 1 .....MEGD.....
Danio_rerio_iRhom1 8 STSSLQRKKPPWLKLDIPTAHVSVDE.....V
Boleophthalmus_pectinirostris_iRhom 8 STSSLQRKKPPWLRLDIPTTQMSLDE.....P
Larimichthys_crocea_iRhom 61 STSSLQRKKPPWLRLDIPTAQMSLDE.....P
Anolis_carolinensis_iRhom 8 SASSLQRKKPPGLRLDIPLPASVGFTQDPA..MAQVA.QGTRLEGPDPSSCSSSLGIP..
Homo_sapiens_iRhom1 8 STSSLQRKKPPWLKLDIPSAVPLTA.EEPS..FL.....
Papio_anubis_iRhom1 8 STSSLQRKKPPWLKLDIPSAVPPTA.EEPS..FL.....
Callithrix_jacchus_iRhom1 8 STSSLQRKKPPWLKLDIPSAAPATA.EEPS..FL.....
Plecturocebus_moloch_iRhom1 8 STSSLQRKKPPWLKLDIPSAAPPAA.EEPS..FL.....
Felis_catus_iRhom 8 STSSLQRKKPPWLKLDIPAVAPPVT.EEPS..FLQVG.PGQERL.....WVYSIQPPHR
Bos_taurus_iRhom1 8 SSSSLQHKKPPWLKLDIPAVVPPAA.EEPS..FL.....
Odocoileus_virginianus_texanus_iRhom 8 SSSSLQHKKPPWLKLDIPAVVPPAA.EEPS..FL.....
Oryctolagus_cuniculus_iRhom 8 STSSLQRKKPPWLKLDIPAVAPPAA.EEPS..VL.....
Meriones_unguiculatus_iR 8 STSSLQRKKPPWLKLDIPAVVPPAA.EEPS..FL.....
Rattus_norvegicus_iRhom1 8 STSSLQRKKPPWLKLDIPAVVPPAA.EEPS..FL.....
Mus_musculus_iRhom1 8 STSSLQRKKPPWLKLDIPAAVPPAA.EEPS..FL.....
Mus_caroli_iRhom 8 STSSLQRKKPPWLKLDIPASVPPAA.EEPS..FL.....
Alligator_mississippiensis_iRhom 8 STSSLQRKKPPWLKLDIPVTVPVTS.PEE.....
Gallus_gallus_iRhom1 8 STSSLQRKKPPWLKLDIPVPMPVSV.PEE.....
Xenopus_tropicalis_iRhom2 9 SSSRLQDRKPPNLSITIPPVEGERD.....
Gallus_gallus_iRhom2 13 SSSRLQSKKPPNLSIVIPPPREAEEDGACK.....
Mus_musculus_iRhom2 15 SGSRLQSRKPPNLSITIPPPES..QAPGEQDSMLPE.....
Homo_sapiens_iRhom2 15 SSSRLQSRKPPNLSITIPPPEKETQAPGEQDSMLPEGFQNRRLKKSQPRTWAAHT.TACP
Canis_lupus_familiaris_iRhom2 15 SSSRLQSRKPPNLSITIPPPEA..SAPGEQASMLPQ.....

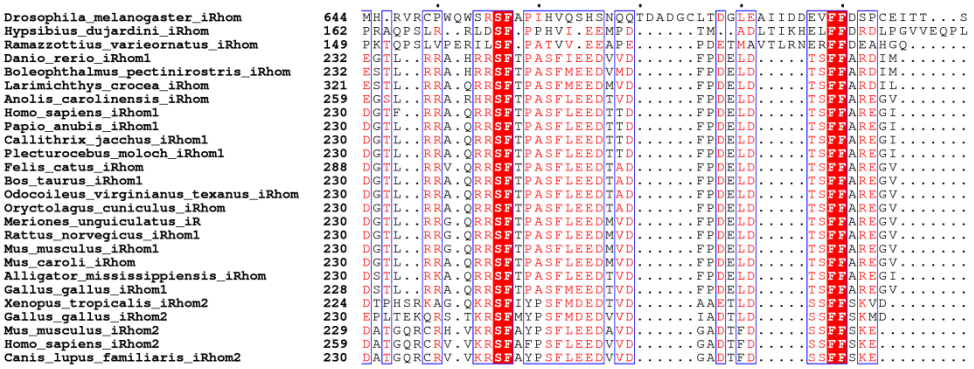

Drosophila_melanogaster_iRhom 393 TTGVHQTTQTTGPISPVPISLPVPMAMPVMSGLNEQMPSQSLESSMVSVQAKRLEGGGNM
Hypsibius_dujardini_iRhom 6 .....PEAKNGNGNGNNLAVEDA.....AGSS
Ramazzottius_varieornatus_iRhom 5 .....RTD..VPVAASAAAADA.....GGSS
Danio_rerio_iRhom1 35 PAFV...PPVKRQGFHRSASMPVETS.....H..F
Boleophthalmus_pectinirostris_iRhom 35 PAFV...QPMKRQGFLRSISMPVETV.....H..L
Larimichthys_crocea_iRhom 88 PTFV...QPVKRQGFLRSISMPVETS.....H..L
Anolis_carolinensis_iRhom 63 WQQQ...QPPRRPPFLRSVSMPSDNN.....NR.L
Homo_sapiens_iRhom1 39 .....QPLRRQAFLRSVSMPAETA.....H.I
Papio_anubis_iRhom1 39 .....QPLRRQAFLRSVSMPAETA.....H.I
Callithrix_jacchus_iRhom1 39 .....QPLRRQVFLRSVSMPAETA.....H.I
Plecturocebus_moloch_iRhom1 39 .....QPLRRQVFLRSVSMPAETA.....H.I
Felis_catus_iRhom 58 DPTV...QPLRRQAFLRSVSMPAETA.....H.I
Bos_taurus_iRhom1 39 .....QPLRRQAFLRSVSMPAEPA.....R.V
Odocoileus_virginianus_texanus_iRhom 39 .....QPLRRQAFLRSVSMPAEPA.....R.V
Oryctolagus_cuniculus_iRhom 39 .....QPLRRQAFLRSVSMPAETA.....H.V
Meriones_unguiculatus_iR 39 .....QPLRRQAFLRSVSMPAETA.....R.V
Rattus_norvegicus_iRhom1 39 .....QPLRRQAFLRSVSMPAETA.....R.V
Mus_musculus_iRhom1 39 .....QPLRRQAFLRSVSMPAETA.....R.V
Mus_caroli_iRhom 39 .....QPLRRQAFLRSVSMPAETA.....R.V
Alligator_mississippiensis_iRhom 36 PLGV...QQIKRQVFLRSVSMPVENT.....R.I
Gallus_gallus_iRhom1 36 PV.....QPTRRQAFLRSVSMPADNS.....R.V
Xenopus_tropicalis_iRhom2 34 PKLM...QPLTKPVFQKSVSLQEPRG.....HV.A
Gallus_gallus_iRhom2 43 .....EPSKVPIYRKSKSLQEPRT.....KG.C
Mus_musculus_iRhom2 49 .....RRKNPAYLKSVSLQEPRG.....RW.Q
Homo_sapiens_iRhom2 74 PSFL...PKRKNPAYLKSVSLQEPRS.....RW.Q
Canis_lupus_familiaris_iRhom2 49 .....RPRNPAYMKSVSLQEQRG.....RW.Q

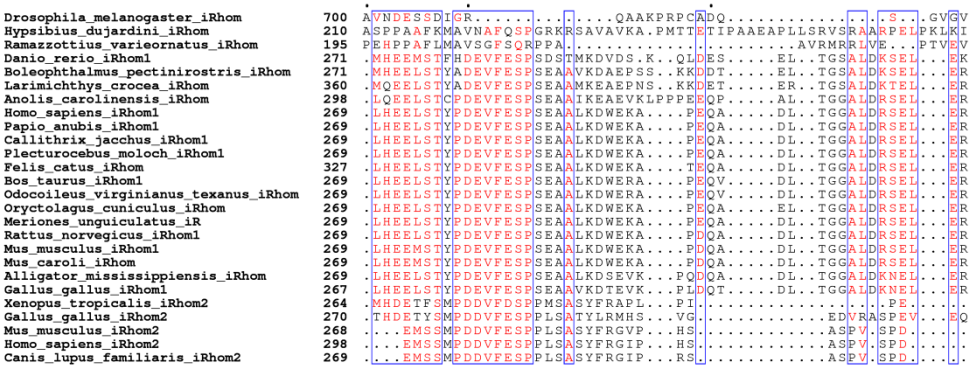

Drosophila_melanogaster_iRhom 453 AVPPGSNQHQLYRSQMASNPRLQHIHHQTQ.....QHQSLQVLPQP.QNHHLSC
Hypsibius_dujardini_iRhom 28 RSIGS.RTSQS.....PRPQHHRMD.....SL
Ramazzottius_varieornatus_iRhom 24 KSIAG.....NKSQLSGMS.....SL
Danio_rerio_iRhom1 60 HSPQR.DIIDH.....RRAAFQRQSSITQTI.....
Boleophthalmus_pectinirostris_iRhom 60 QSPPR.DFFDT.....RRPVLQRQSSITQTI.....
Larimichthys_crocea_iRhom 113 QSPPR.DFFET.....RRPVLQRQSSITQTIKSSRRVHFERINTVPIKGQRAARRSTRK
Anolis_carolinensis_iRhom 89 PSPCS...EGR.....RRMELHRHAS.....
Homo_sapiens_iRhom1 60 SSPHH...ELR.....R.PVLQRQTSITQTI.....
Papio_anubis_iRhom1 60 SSPHH...ELR.....R.SVLQRQTSITQTI.....
Callithrix_jacchus_iRhom1 60 SSPHY...ELR.....R.PVLQRQTSITQTI.....
Plecturocebus_moloch_iRhom1 60 SSPHC...ELR.....R.PVLQRQTSITQTI.....
Felis_catus_iRhom 83 PSPHH...EVR.....R.PVLQRQTSITQTIRS.RQVRFGRVHTLPLLGPPGARRALPQ
Bos_taurus_iRhom1 60 PSPHQ...EPR.....R.PVLQRQMSITQTI.....
Odocoileus_virginianus_texanus_iRhom 60 PSPHQ...EPR.....R.PALQRQMSITQTI.....
Oryctolagus_cuniculus_iRhom 60 PCAHP...EPR.....R.PVLQRQASITQTI.....
Meriones_unguiculatus_iR 60 PSPHH...EPR.....R.LVLQRQTSITQTI.....
Rattus_norvegicus_iRhom1 60 PSPHH...EPR.....R.LALQRQTSITQTI.....
Mus_musculus_iRhom1 60 PSPHH...EPR.....R.LVLQRQTSITQTI.....
Mus_caroli_iRhom 60 PSPHH...EPR.....R.LVLQRQTSITQTI.....
Alligator_mississippiensis_iRhom 61 PSLPN...EVR.....K.PMLQRQTSITQTI.....
Gallus_gallus_iRhom1 59 PSLPH...EVR.....R.PVLQRQTSITQTI.....
Xenopus_tropicalis_iRhom2 60 EGG.....TAD.....VRPGFKRQT.....
Gallus_gallus_iRhom2 65 D.....SSE.....RRPGFRRQT.....
Mus_musculus_iRhom2 70 EG.....AE.....KRPGFRRQA.....
Homo_sapiens_iRhom2 100 ES.....SE.....KRPGFRRQA.....
Canis_lupus_familiaris_iRhom2 70 EG.....SSE.....KRPGFRRQA.....

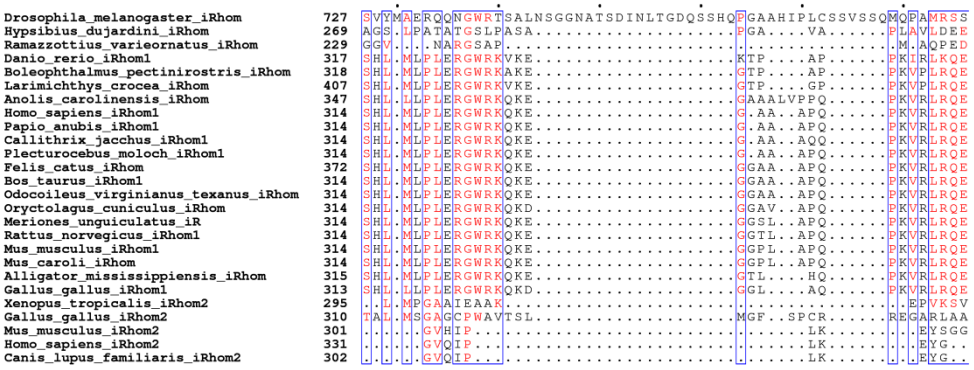

Drosophila_melanogaster_iRhom 501 KKRIGMYMR...RTTSQFFGVEPSTEA..ADCALWQGRHRRLAIRCFGMFDTELEYHMQ
Hypsibius_dujardini_iRhom 49 KRSVSAGVKRSQSAVAHWLGVEQMSEEESDDRQAMWQARRVTMG.RRFGA.....
Ramazzottius_varieornatus_iRhom 40 RRTVSQGVRRSQSAVAHFLGVETVTEEDMEERKAVWDARRATLA.RRLGN.....
Danio_rerio_iRhom1 85 ........K...RGTADWFGVSKDG....D.ATQKWQRKSLRHCSLRYGKLKPQVIREME
Boleophthalmus_pectinirostris_iRhom 85 ........K...RGTADWFGVSKDG....D.ATQKWQRKSLRHCSLRYGKLKPQVIREMD
Larimichthys_crocea_iRhom 166 HHSLSRTLL...RGTADWFGVSKDG....D.ATQKWQRKSLRHCSLRYGKLKAQVIREMD
Anolis_carolinensis_iRhom 107 ...ITQTIR...RGTADWFGISKES....D.ATQKWQRKSLRHCSLRYGKLKPQVIREMD
Homo_sapiens_iRhom1 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVLRELD
Papio_anubis_iRhom1 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVLRELD
Callithrix_jacchus_iRhom1 82 ........R...RGAADWFGVSKES....D.STQKWQRKSIRHCSQRYGKLKPQVLRELD
Plecturocebus_moloch_iRhom1 82 ........R...RGAADWFGVSKES....E.STQKWQRKSIRHCSQRYGKLKPQVLRELD
Felis_catus_iRhom 132 RRSLSRSLL...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Bos_taurus_iRhom1 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Odocoileus_virginianus_texanus_iRhom 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Oryctolagus_cuniculus_iRhom 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Meriones_unguiculatus_iR 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Rattus_norvegicus_iRhom1 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Mus_musculus_iRhom1 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Mus_caroli_iRhom 82 ........R...RGTADWFGVSKDS....D.STQKWQRKSIRHCSQRYGKLKPQVIRELD
Alligator_mississippiensis_iRhom 83 ........R...RGTADWFGVSKDS....D.ATQKWQRKSLRHCSLRYGKLKPQVIREMD
Gallus_gallus_iRhom1 81 ........K...RGTADWFGVSKDS....D.ATQKWQRKSLRHCSLRYGKLKPQVIREMD
Xenopus_tropicalis_iRhom2 75 ..SLSQSIK...RGTAQWFGVSNDW....EGKRQQWQRKSLHHCSLRYGKLKPQYQRDIE
Gallus_gallus_iRhom2 78 ..SLSQSIR...KGTAEWFGVSGDW....EGKRQHWQRRSLQHCSMRYGKLKAA.YRDME
Mus_musculus_iRhom2 83 ..SLSQSIR...KSTAQWFGVSGDW....EGKRQNWHRRSLHHCSVHYGRLKASCQRELE
Homo_sapiens_iRhom2 113 ..SLSQSIR...KGAAQWFGVSGDW....EGQRQQWQRRSLHHCSMRYGRLKASCQRDLE
Canis_lupus_familiaris_iRhom2 84 ..SLSQSIR...KGTAQWFGVSGDW....EAKRQHWQRRSLHHCSVRYGRLKASCQRDLE

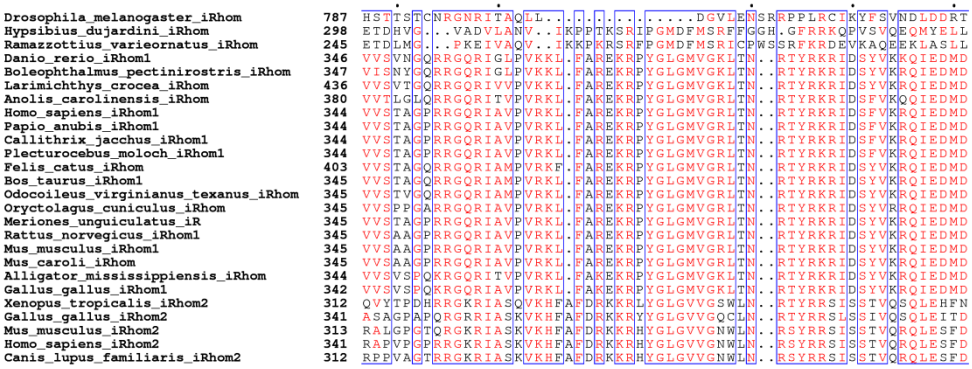

Drosophila_melanogaster_iRhom 555 QAIGGGNEDAGQSNGTNGNYAPDRPDILPVQDAIGMDMTMSG.....D....S
Hypsibius_dujardini_iRhom 98 .....K.....PKPSPHISTLMENDVTDSPDV.....ASGSR
Ramazzottius_varieornatus_iRhom 89 .....K.....QSKPAVTPSHAELG..MIPDT.....E..TD
Danio_rerio_iRhom1 129 LPSQDNISLTST.....ETPPPLYVPSSQHGMQKIVDPLARGRAFRMVEE.VDGYS
Boleophthalmus_pectinirostris_iRhom 129 LPSQDNISLTST.....ETPPPLYVPSSQHGMQKIVDPLARGRAFRMVEE.VDGYS
Larimichthys_crocea_iRhom 218 LPSQDNISLTST.....ETPPPLYVPSSQHGMQKIVDPLARGRAFRMVEE.VDGYS
Anolis_carolinensis_iRhom 156 LPSQDNLSLAST.....ETPPPLYVGSCHLGMQRIVDPLARGRAFRMAED.SDGCS
Homo_sapiens_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRVADDTAEGLS
Papio_anubis_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRVADDTAEGLS
Callithrix_jacchus_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRVADDTAEGLS
Plecturocebus_moloch_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRVADDAAEGLS
Felis_catus_iRhom 184 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRVADDTADGLS
Bos_taurus_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIVDPLARGRAFRLADDAADGPS
Odocoileus_virginianus_texanus_iRhom 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRLADDAADGPS
Oryctolagus_cuniculus_iRhom 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRVADDAADGLS
Meriones_unguiculatus_iR 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRMADDTADGLS
Rattus_norvegicus_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRMADDTADGLS
Mus_musculus_iRhom1 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRMADDTADGLS
Mus_caroli_iRhom 126 LPSQDNVSLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRMADDTADGLS
Alligator_mississippiensis_iRhom 127 LPSQDNISLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRMADD.NDGCS
Gallus_gallus_iRhom1 125 LPSQDNISLTST.....ETPPPLYVGPCQLGMQKIIDPLARGRAFRMADD.NDGYS
Xenopus_tropicalis_iRhom2 126 LPSQETPSLFQAT.....ESPATHRLPK.....IVDPLARGRPFRHPDE.VDRPR
Gallus_gallus_iRhom2 128 LPSQEVPSLFQGT.....ESPKPAKMPKVPLG..IVDPLARGRPFRHPDD.ADRPH
Mus_musculus_iRhom2 134 LPSQEVPSLFQGT.....ESPKPCKMPK.....IVDPLARGRAFRHPDE.VDRPH
Homo_sapiens_iRhom2 164 LPSQEAPSLFQGT.....ESPKPCKMPK.....IVDPLARGRAFRHPEE.MDRPH
Canis_lupus_familiaris_iRhom2 135 LPSQEVPSLFQGT.....ESPKPCKMPK.....IVDPLARGRAFRHPDE.VDRPH

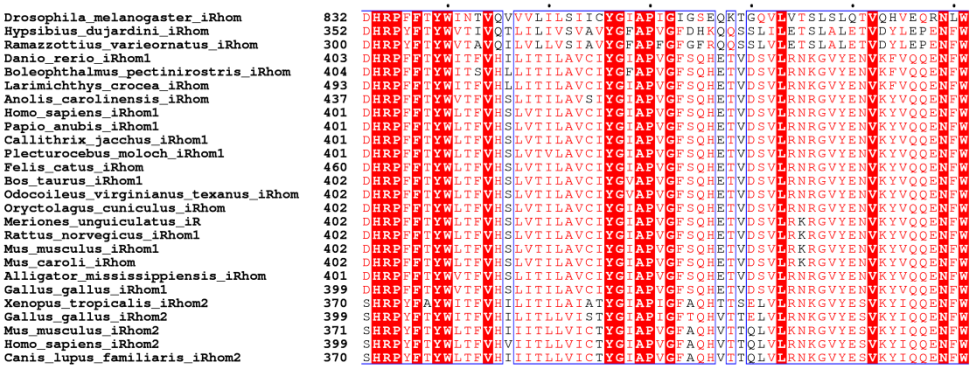

Drosophila_melanogaster_iRhom 599 RRQK.....CYTNRDFLAGEFV..ERKASVGYMFVAMVSYLVHMFNKRRPIQ
Hypsibius_dujardini_iRhom 125 .....PAVNPILDDIVRVWKKPSVAEMAWRKVARKA.....VSAPSQ
Ramazzottius_varieornatus_iRhom 112 .....GPKRELAAETLRLWKKPSVAQMAWKKVAKAA.....VAPSER
Danio_rerio_iRhom1 179 VPQ...TPITPGAASICSFNSSRSGLNRIPRRRKRESVAKMSFRAAAALV.....KGRSLR
Boleophthalmus_pectinirostris_iRhom 179 VPQ...TPITPGAASLCSFTSSRSGLNRLPRRRKRESVAKMSFRAAAALV.....KGRSFR
Larimichthys_crocea_iRhom 268 VPQ...TPITPGAASLCSFTSSRSGLNRLPRRRKRESVAKMSFRAAAALV.....KGRSFR
Anolis_carolinensis_iRhom 206 APH...TPITPGAASLCSFGSSRSGFGRWPRRRKRESVAKMSFRAAAALM.....KGRSMK
Homo_sapiens_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFHRLPRRRKRESVAKMSFRAAAALM.....KGRSVR
Papio_anubis_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFHRLPRRRKRESVAKMSFRAAAALM.....KGRSVR
Callithrix_jacchus_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFHRLPRRRKRESVAKMSFRAAAALM.....KGRSVR
Plecturocebus_moloch_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFHRLPRRRKRESVAKMSFRAAAALM.....KGRSVR
Felis_catus_iRhom 235 TPH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSVR
Bos_taurus_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSVR
Odocoileus_virginianus_texanus_iRhom 177 APH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSVR
Oryctolagus_cuniculus_iRhom 177 APH...TPITPGAASLCSFSSSRSGFSRLPRRRKRESVAKMSFRAAAALV.....KGRSIR
Meriones_unguiculatus_iR 177 APH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSIR
Rattus_norvegicus_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSIR
Mus_musculus_iRhom1 177 APH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSIR
Mus_caroli_iRhom 177 APH...TPVTPGAASLCSFSSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSIR
Alligator_mississippiensis_iRhom 177 MPH...TPITPGATSLCSFTSSRSGFNRLPRRRKRESVAKMSFRAAAALM.....KGRSVK
Gallus_gallus_iRhom1 175 IPH...TPITPGATSLCSFTSSRSGFNRLPRRRKRESVAKMSFRAAAALV.....KGRSVK
Xenopus_tropicalis_iRhom2 169 TPHVGLPPQTPGMLSLASFTSVRSGYSRFPR.RKRESVAHMSFRAAAALI.....KGRPIL
Gallus_gallus_iRhom2 175 TPHHVLPPHTPGVASLASLGSARSGYSRLPR.RKRQSVARMSFRAAAALL.....RGRSVL
Mus_musculus_iRhom2 177 AAHL...PPLTPGVLSLTSFTSVRSGYSHLPR.RKRISVAHMSFQAAAALL.....KGRSVL
Homo_sapiens_iRhom2 207 APH...PPLTPGVLSLTSFTSVRSGYSHLPR.RKRMSVAHMSLQAAAALL.....KGRVSL
Canis_lupus_familiaris_iRhom2 178 APH...PPLTPGVLSLTSFTSVRSGHSHLPR.RKRMSVAHMSFQAAAALL.....KGRSVL

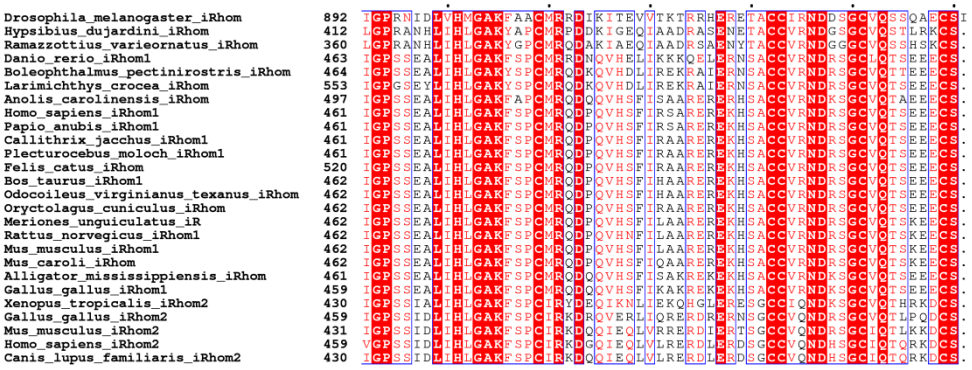

Drosophila_melanogaster_iRhom 644 MH.RVRCPWQWSRSFAPIHVQSHSNQQTDADGCLTDGLEAIIDDEVFFDSFCEITT...S
Hypsibius_dujardini_iRhom 162 PRAQPSLR..RLDSF.PPHVI.EEMPD.....TM..ADLTIKHELFFDRDLPGVVEQPL
Ramazzottius_varieornatus_iRhom 149 PKTQPSLVPERILSF.PATVV.EEAPE.....PDETMAVTLRNERFFDEAHGQ.....
Danio_rerio_iRhom1 232 EGTL..RRA.HRRSFTPASFIEEDVVD.....FPDELD.....TSFFARDIM.....
Boleophthalmus_pectinirostris_iRhom 232 ESTL..RRA.HRRSFTPASFMEEDVMD.....FPDELD.....TSFFARDIM.....
Larimichthys_crocea_iRhom 321 ESTL..RRA.QRRSFTPASFMEEDMVD.....FPDELD.....TSFFARDIL.....
Anolis_carolinensis_iRhom 259 EGSL..RRA.RHRSFTPASFLEEDTVD.....FPDELD.....TSFFAREGV.....
Homo_sapiens_iRhom1 230 DGTF..RRA.QRRSFTPASFLEEDTTD.....FPDELD.....TSFFAREGI.....
Papio_anubis_iRhom1 230 DGTL..RRA.QRRSFTPASFLEEDTTD.....FPDELD.....TSFFAREGI.....
Callithrix_jacchus_iRhom1 230 DGTL..RRA.QRRSFTPASFLEEDTTD.....FPDELD.....TSFFAREGI.....
Plecturocebus_moloch_iRhom1 230 DGTL..RRA.QRRSFTPASFLEEDTTD.....FPDELD.....TSFFAREGI.....
Felis_catus_iRhom 288 DGTL..RRV.QRRSFTPASFLEEDTAD.....FPDELD.....TSFFAREGV.....
Bos_taurus_iRhom1 230 DGTL..RRA.QRRSFTPASFLEEDTAD.....FPDELD.....TSFFAREGV.....
Odocoileus_virginianus_texanus_iRhom 230 DGTL..RRA.QRRSFTPASFLEEDTAD.....FPDELD.....TSFFAREGV.....
Oryctolagus_cuniculus_iRhom 230 DGTL..RRA.QRRSFTPASFLEEDTAD.....FPDELD.....TSFFAREGV.....
Meriones_unguiculatus_iR 230 DGTL..RRG.QRRSFTPASFLEEDMVD.....FPDELD.....TSFFAREGV.....
Rattus_norvegicus_iRhom1 230 DGTL..RRG.QRRSFTPASFLEEDMVD.....FPDELD.....TSFFAREGV.....
Mus_musculus_iRhom1 230 DGTL..RRG.QRRSFTPASFLEEDMVD.....FPDELD.....TSFFAREGV.....
Mus_caroli_iRhom 230 DGTL..RRA.QRRSFTPASFLEEDMVD.....FPDELD.....TSFFAREGV.....
Alligator_mississippiensis_iRhom 230 DSTL..RRA.QRRSFTPASFLEEDTVD.....FPDELD.....TSFFAREGI.....
Gallus_gallus_iRhom1 228 DSTL..RRA.QRRSFTPASFMEEDTVD.....FPDELD.....TSFFAREGV.....
Xenopus_tropicalis_iRhom2 224 DTPHSRKAG.QKRSFIYPSFMDEDTVD.....AAETLD.....SSFFSKVD.....
Gallus_gallus_iRhom2 230 EPLTEKQRS.TKRSFMYPSFMDEDVVD.....IADTLD.....SSFFSKMD.....
Mus_musculus_iRhom2 229 DATGQRCRH.VKRSFAYPSFLEEDAVD.....GADTFD.....SSFFSKE.....
Homo_sapiens_iRhom2 259 DATGQRCRV.VKRSFAFPSFLEEDVVD.....GADTFD.....SSFFSKE.....
Canis_lupus_familiaris_iRhom2 230 DATGQRCRV.VKRSFAYPSFLEEDVVD.....GADTFD.....SSFFSKE.....

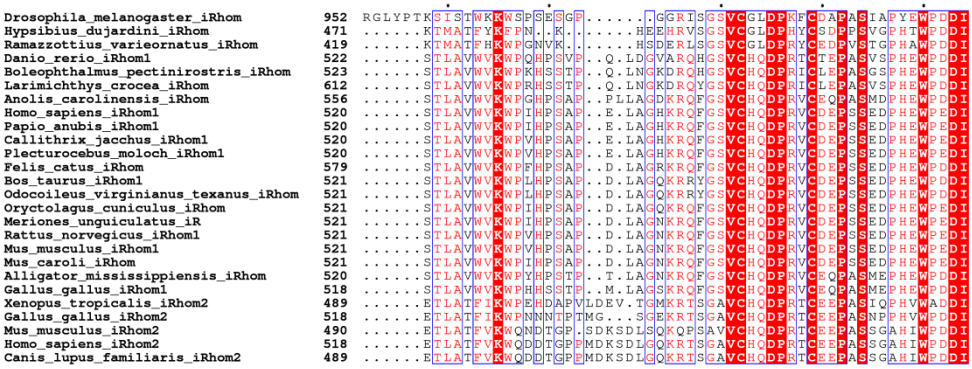

Drosophila_melanogaster_iRhom 700 AVNDESSDIGR.....QAAKPRPCADQ.....S..GVGV
Hypsibius_dujardini_iRhom 210 ASPPAAFKMAVNAFQSPGRKRSAVAVKA.PMTTETIPAAEAPLLSRVSRAARPELPKLKI
Ramazzottius_varieornatus_iRhom 195 PEHPPAFLMAVSGFSQRPPA.....AVRMRRLVE...PTVEV
Danio_rerio_iRhom1 271 .MHEEMSTFHDEVFESPSDSTMKDVDS.K..QLDES.....EL..TGSALDKSEL...EK
Boleophthalmus_pectinirostris_iRhom 271 .MHEELSTYADEVFESPSEAAVKDAEPSS..KKDDT.....EL..TGSALDKSEL...ER
Larimichthys_crocea_iRhom 360 .MQEELSTYADEVFESPSEAAMKEAEPNS..KKDET.....ER..TGSALDKTEL...ER
Anolis_carolinensis_iRhom 298 .LQEELSTCPDEVFESPSEAAIKEAEVKLPPPEEQP.....AL..TGGALDRSEL...ER
Homo_sapiens_iRhom1 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PEQA.....DL..TGGALDRSEL...ER
Papio_anubis_iRhom1 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PEQA.....DL..TGGALDRSEL...ER
Callithrix_jacchus_iRhom1 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PEQA.....DL..TGGALDRSEL...ER
Plecturocebus_moloch_iRhom1 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PEQA.....DL..TGGALDRSEL...ER
Felis_catus_iRhom 327 .LHEELSTYPDEVFESPSEAALKDWEKA....TEQA.....DL..TGGALDRSEL...ER
Bos_taurus_iRhom1 269 .LHEELSTYPDEVFESPSEAALKDWERA....PEQV.....DL..TGGALDRSEL...ER
Odocoileus_virginianus_texanus_iRhom 269 .LHEELSTYPDEVFESPSEAALKDWERA....PEQV.....DL..TGGALDRSEL...ER
Oryctolagus_cuniculus_iRhom 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PEQA.....EL..TGGALDRSEL...ER
Meriones_unguiculatus_iR 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PEQA.....DL..TGGALDRSEL...ER
Rattus_norvegicus_iRhom1 269 .LHEELSTYPDEVFESPSEAALKDWEKA....PDQA.....DL..TGGALDRSEL...ER
Mus_musculus_iRhom1 269 .LHEEMSTYPDEVFESPSEAALKDWEKA....PDQA.....DL..TGGALDRSEL...ER
Mus_caroli_iRhom 269 .LHEEMSTYPDEVFESPSEAALKDWEKA....PDQA.....DL..TGGALDRSEL...ER
Alligator_mississippiensis_iRhom 269 .LHEELSTYPDEVFESPSEAALKDSEVK...PQDQA.....DL..TGGALDKNEL...ER
Gallus_gallus_iRhom1 267 .LHEELSTYPDEVFESPSEAAVKDTEVK...PLDQT.....DL..TGGALDKNEL...ER
Xenopus_tropicalis_iRhom2 264 .MHDETFSMPDDVFDSPPMSASYFRAPL...PI.....PE.....
Gallus_gallus_iRhom2 270 .THDETYSMPDDVFESPPLSATYLRMHS...VG.....EDVRASPEV...EQ
Mus_musculus_iRhom2 268 .....EMSSMPDDVFESPPLSASYFRGVP...HS.....ASPV.SPD.....
Homo_sapiens_iRhom2 298 .....EMSSMPDDVFESPPLSASYFRGIP...HS.....ASPV.SPD.....
Canis_lupus_familiaris_iRhom2 269 .....EMSSMPDDVFESPPLSASYFRGIP...RS.....ASPV.SPD.....

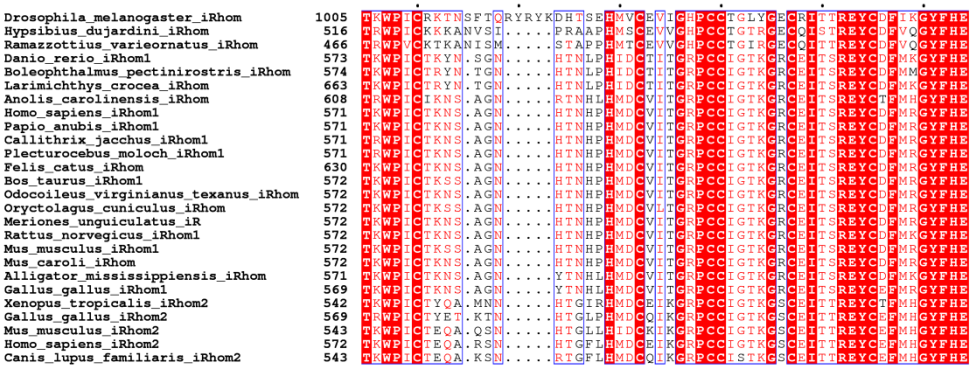

Drosophila_melanogaster_iRhom 727 SVYMAERQQNGWRTSALNSGGNATSDINLTGDQSSHQPGAAHIPLCSSVSSQMQPAMRSS
Hypsibius_dujardini_iRhom 269 AGS.LPATATGSLPASA.....PGA....VA.....PLAVLDEE
Ramazzottius_varieornatus_iRhom 229 GGV....NARGSAP.....M.AQPED
Danio_rerio_iRhom1 317 SHL.MLPLERGWRKVKE.....KTP....AP.....PKIRLKQE
Boleophthalmus_pectinirostris_iRhom 318 SHL.MLPLERGWRKAKE.....GTP....AP.....PKVPLRQE
Larimichthys_crocea_iRhom 407 SHL.MLPLERGWRKVKE.....GTP....GP.....PKVPLRQE
Anolis_carolinensis_iRhom 347 SHL.LLPLERGWRKQKE.....GAAALVPPQ.....PKVRLRQE
Homo_sapiens_iRhom1 314 SHL.MLPLERGWRKQKE.....G.AA..APQ.....PKVRLRQE
Papio_anubis_iRhom1 314 SHL.MLPLERGWRKQKE.....G.AA..APQ.....PKVRLRQE
Callithrix_jacchus_iRhom1 314 SHL.MLPLERGWRKQKE.....G.AA..AQQ.....PKVRLRQE
Plecturocebus_moloch_iRhom1 314 SHL.MLPLERGWRKQKE.....G.AA..APQ.....PKVRLRQE
Felis_catus_iRhom 372 SHL.MLPLERGWRKQKE.....GGAA..APQ.....PKVRLRQE
Bos_taurus_iRhom1 314 SHL.MLPLERGWRKQKE.....GGAA..APQ.....PKVRLRQE
Odocoileus_virginianus_texanus_iRhom 314 SHL.MLPLERGWRKQKE.....GGAA..APQ.....PKVRLRQE
Oryctolagus_cuniculus_iRhom 314 SHL.MLPLERGWRKQKD.....GGAV..APQ.....PKVRLRQE
Meriones_unguiculatus_iR 314 SHL.MLPLERGWRKQKE.....GGSL..APQ.....PKVRLRQE
Rattus_norvegicus_iRhom1 314 SHL.MLPLERGWRKQKE.....GGTL..APQ.....PKVRLRQE
Mus_musculus_iRhom1 314 SHL.MLPLERGWRKQKE.....GGPL..APQ.....PKVRLRQE
Mus_caroli_iRhom 314 SHL.MLPLERGWRKQKE.....GGPL..APQ.....PKVRLRQE
Alligator_mississippiensis_iRhom 315 SHL.MLPLERGWRKQKE.....GTL....HQ.....PKVRLRQE
Gallus_gallus_iRhom1 313 SHL.LLPLERGWRKQKD.....GGL....AQ.....PKVRLRQE
Xenopus_tropicalis_iRhom2 295 ..L.MPGAAIEAAK.....EPVKSV
Gallus_gallus_iRhom2 310 TAL.MSGAGCPWAVTSL.....MGF..SPCR.....REGARLAA
Mus_musculus_iRhom2 301 .....GVHIP...LK.....EYSGG
Homo_sapiens_iRhom2 331 .....GVQIP...LK.....EYG..
Canis_lupus_familiaris_iRhom2 302 .....GVQIP...LK.....EYG..

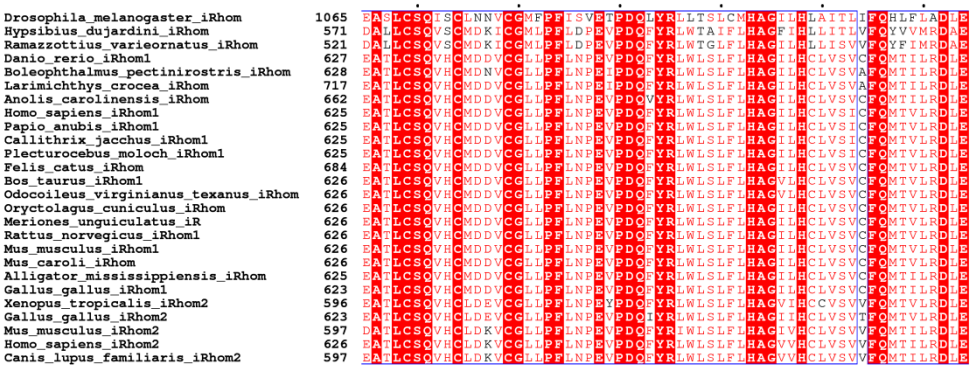

Drosophila_melanogaster_iRhom 787 HSTTSTCNRGNRITAQLL.....DGVLENSRRPPLRCIKYFSVNDLDDRT
Hypsibius_dujardini_iRhom 298 ETDHVG...VADVLANV..IKPPTKSRIPGMDFMSRFFGGH.GFRRKQPVSVQEQMYELL
Ramazzottius_varieornatus_iRhom 245 ETDLMG...PKEIVAQV..IKKPKRSRFPGMDFMSRICPWSSRFKRDEVKAQEEKLASLL
Danio_rerio_iRhom1 346 VVSVNGQRRGQRIGLPVKKL.FAREKRPYGLGMVGKLTN..RTYRKRIDSYVKKQIEDMD
Boleophthalmus_pectinirostris_iRhom 347 VISNYGQRRGQRIGLPVKKL.FAREKRPYGLGMVGKLTN..RTYRKRIDSYVKRQIEDMD
Larimichthys_crocea_iRhom 436 VVSVTGQRRGQRIVVPVKKL.FAREKRPYGLGMVGKLTN..RTYRKRIDSYVKRQIEDMD
Anolis_carolinensis_iRhom 380 VVTLGLQRRGQRITVPVRKL.FAKEKRPYGLGMVGRLTN..RTYRKRIDSFVKQQIEDMD
Homo_sapiens_iRhom1 344 VVSTAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSFVKRQIEDMD
Papio_anubis_iRhom1 344 VVSTAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSFVKRQIEDMD
Callithrix_jacchus_iRhom1 344 VVSTAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSFVKRQIEDMD
Plecturocebus_moloch_iRhom1 344 VVSTAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSFVKRQIEDMD
Felis_catus_iRhom 345 VVSTAGQRRGQRIAMPVRKF.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Bos_taurus_iRhom1 345 VVSTAGQRRGQRIAMPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Odocoileus_virginianus_texanus_iRhom 345 VVSTVGQRRGQRIAMPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Oryctolagus_cuniculus_iRhom 345 VVSPPGARRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVRRQIEDMD
Meriones_unguiculatus_iR 345 VVSTAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Rattus_norvegicus_iRhom1 345 VVSAAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Mus_musculus_iRhom1 345 VVSAAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Mus_caroli_iRhom 345 VVSAAGPRRGQRIAVPVRKL.FAREKRPYGLGMVGRLTN..RTYRKRIDSYVKRQIEDMD
Alligator_mississippiensis_iRhom 344 VVSVSPQKRGQRITVPVRKL.FAKEKRPYGLGMVGKLTN..RTYRKRIDSYVKRQIEDMD
Gallus_gallus_iRhom1 342 VVSVSPQKRGQRIAVPVRKL.FAKEKRPYGLGMVGKLTN..RTYRKRIDSYVKRQIEDMD
Xenopus_tropicalis_iRhom2 312 QVYTPDHRRGKRIASQVKHFAFDRKKRLYGLGVVGSWLN..RTYRRSISSTVQSQLEHFN
Gallus_gallus_iRhom2 341 ASAGPAPQRGRRIASKVKHFAFDRKKRYYGLGVVGQCLN..RTYRRSLSSIVQSQLEITD
Mus_musculus_iRhom2 313 RALGPGTQRGKRIASKVKHFAFDRKKRHYGLGVVGNWLN..RSYRRSISSTVQRQLESFD
Homo_sapiens_iRhom2 341 RAPVPGPRRGKRIASKVKHFAFDRKKRHYGLGVVGNWLN..RSYRRSISSTVQRQLESFD
Canis_lupus_familiaris_iRhom2 312 RPPVAGTRRGKRIASKVKHFAFDRKKRHYGLGVVGNWLN..RSYRRSISSTVQRQLESFD

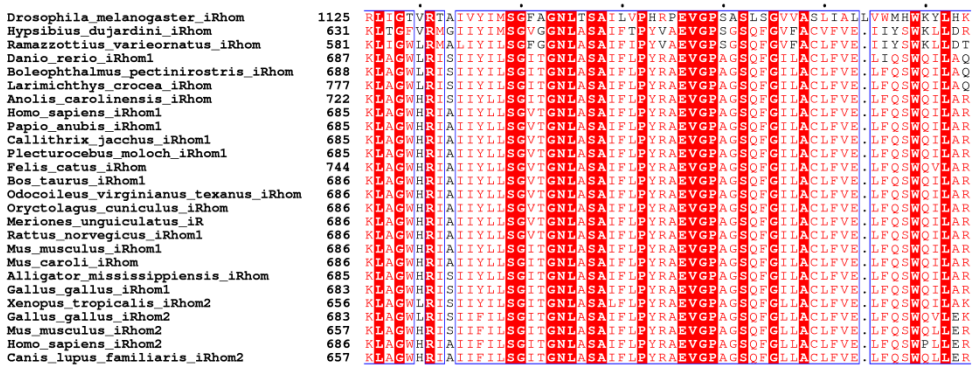

Drosophila_melanogaster_iRhom 832 DHRPFFTYWINTVQVVLILSIICYGIAPIGIGSEQKTGQVLVTSLSLQTVQHVEQRNLW
Hypsibius_dujardini_iRhom 352 DHRPYFTYWVTIVQTLILIVSVAVYGFAPVGFDHKQQSSLILETSLALETVDYLEPEMFW
Ramazzottius_varieornatus_iRhom 300 DHRPYFTYWVTAVQILVLLVSIAVYGFAPFGFGFRQQSSLVLETSLALETVDYLEPEMFW
Danio_rerio_iRhom1 403 DHRPFFTYWITFVHILITILAVCIYGIAPVGFSQHETVDSVLRNKGVYENVKFVQQEMFW
Boleophthalmus_pectinirostris_iRhom 404 DHRPFFTYWITSVHLLITILAVCIYGFAPVGFSQHETVDSVLRNKGVYENVKFVQQEMFW
Larimichthys_crocea_iRhom 493 DHRPFFTYWITFVHLLITILAVCIYGIAPVGFSQHETVDSVLRNKGVYENVKFVQQEMFW
Anolis_carolinensis_iRhom 437 DHRPFFTYWVTFVHSLITILAVSIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Homo_sapiens_iRhom1 401 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Papio_anubis_iRhom1 401 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Callithrix_jacchus_iRhom1 401 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Plecturocebus_moloch_iRhom1 401 DHRPFFTYWLTFVHSLVTVLAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Felis_catus_iRhom 460 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Bos_taurus_iRhom1 402 DHRPFFTYWLTFVHSLVTILAVCIYGVAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Odocoileus_virginianus_texanus_iRhom 402 DHRPFFTYWLTFVHSLVTILAVCIYGVAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Oryctolagus_cuniculus_iRhom 402 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Meriones_unguiculatus_iR 402 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Rattus_norvegicus_iRhom1 402 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRKRGVYENVKYVQQEMFW
Mus_musculus_iRhom1 402 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRKRGVYENVKYVQQEMFW
Mus_caroli_iRhom 402 DHRPFFTYWLTFVHSLVTILAVCIYGIAPVGFSQHETVDSVLRKRGVYENVKYVQQEMFW
Alligator_mississippiensis_iRhom 401 DHRPFFTYWVTFVHSLITILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Gallus_gallus_iRhom1 399 DHRPFFTYWVTFVHSLITILAVCIYGIAPVGFSQHETVDSVLRNRGVYENVKYVQQEMFW
Xenopus_tropicalis_iRhom2 370 SHRPYFAYWITFVHIITILAIATYGIAPIGFAQHTTSELVLRNKGVYESVKYIQQENFW
Gallus_gallus_iRhom2 399 SHRPYFTYWITFVHILITLLVISTYGIAPIGFTQHVTTELVLRNKGVYESVKYIQQENFW
Mus_musculus_iRhom2 371 SHRPYFTYWLTFVHIIITLLVICTYGIAPVGFAQHVTTQLVLKNRGVYESVKYIQQENFW
Homo_sapiens_iRhom2 399 SHRPYFTYWLTFVHVIITLLVICTYGIAPVGFAQHVTTQLVLRNKGVYESVKYIQQENFW
Canis_lupus_familiaris_iRhom2 370 SHRPYFTYWLTFVHIIITLLVICTYGIAPVGFAQHVTTQLVLRNKGVYESVKYIQQENFW

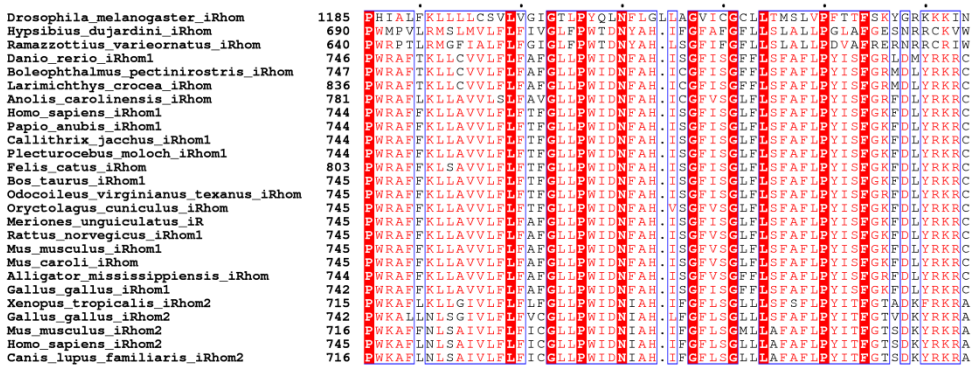

Drosophila_melanogaster_iRhom 892 IGPRNIDLVHKGAKFAAQMRDIKITEVVTKTRRHPRETACCIRNDDSGCVQSSQAEGSI
Hypsibius_dujardini_iRhom 412 LGPRANHLIHHGAKYAPQMRPDDKIGEQIAADRASENETACCVRNDGSGCVQSTLRKGS.
Ramazzottius_varieornatus_iRhom 360 LGPRANHLIHHGAKYGPQMRQDAKIAEQIAADRSAENYTACCVRNDGGGCVQSSHSKGS.
Danio_rerio_iRhom1 463 IGPSSEALIHHGAKFSPQMRRDNQVHELIKKKQELRRNSACCVRNDRSGCVQTSEEEGS.
Boleophthalmus_pectinirostris_iRhom 464 IGPSSEALIHHGAKYSPQMRQDKQVHDLIREKRAIRRNSACCVRNDRSGCVQTTEEEGS.
Larimichthys_crocea_iRhom 553 IGPGSEYLIHHGAKYSPQMRQDKQVHDLIREKRAIRRNSACCVRNDRSGCVQTSEEEGS.
Anolis_carolinensis_iRhom 497 IGPSSEALIHHGAKFAPQMRQDQQVHSFISAARERRKHSACCVRNDKSGCVQTAEEEGS.
Homo_sapiens_iRhom1 461 IGPSSEALIHHGAKFSPQMRQDPQVHSFIRSARERRKHSACCVRNDRSGCVQTSEEEGS.
Papio_anubis_iRhom1 461 IGPSSEALIHHGAKFSPQMRQDPQVHSFIRSARERRKHSACCVRNDRSGCVQTSEEEGS.
Callithrix_jacchus_iRhom1 461 IGPSSEALIHHGAKFSPQMRQDPQVHSFIRAARERRKHSACCVRNDRSGCVQTSEEEGS.
Plecturocebus_moloch_iRhom1 461 IGPSSEALIHHGAKFSPQMRQDPQVHSFIRAARERRKHSACCVRNDRSGCVQTSEEEGS.
Felis_catus_iRhom 520 IGPSSEALIHHGAKFSPQMRQDPQVHSFIRAARERRKHSACCVRNDRSGCVQTSEEEGS.
Bos_taurus_iRhom1 462 IGPSSEALIHHGAKFSPQMRQDPQVHSFIHAARERRKHSACCVRNDRSGCVQTSEEEGS.
Odocoileus_virginianus_texanus_iRhom 462 IGPSSEALIHHGAKFSPQMRQDPQVHSFIHAARERRKHSACCVRNDRSGCVQTSEEEGS.
Oryctolagus_cuniculus_iRhom 462 IGPSSEALIHHGAKFSPQMRQDPQVHSFIRAARERRKHSACCVRNDRSGCVQTSEEEGS.
Meriones_unguiculatus_iR 462 IGPSSEALIHHGAKFSPQMRQDPQVHSFILAARERRKHSACCVRNDRSGCVQTSKEEGS.
Rattus_norvegicus_iRhom1 462 IGPSSEALIHHGAKFSPQMRQDPQVHNFILAARERRKHSACCVRNDRSGCVQTSKEEGS.
Mus_musculus_iRhom1 462 IGPSSEALIHHGAKFSPQMRQDPQVHSFILAARERRKHSACCVRNDRSGCVQTSKEEGS.
Mus_caroli_iRhom 462 IGPSSEALIHHGAKFSPQMRQDPQVHSFIQAARERRKHSACCVRNDRSGCVQTSKEEGS.
Alligator_mississippiensis_iRhom 461 IGPSSEALIHHGAKFSPQMRQDQQVHSFISAKREKRKHSACCVRNDRSGCVQTSEEEGS.
Gallus_gallus_iRhom1 459 IGPSSEALIHHGAKFSPQMRQDQQVHSFIKAKREKRKHSACCVRNDKSGCVQTSEEEGS.
Xenopus_tropicalis_iRhom2 430 IGPSSIALIHHGAKFSPQIRYDEQIKNLIEKQKGLRESGCCIQNDKSGCVQTHRKQGS.
Gallus_gallus_iRhom2 459 IGPSSIDLIHHGAKFSPQIRKDRQVERLIQRERDRRNSGCCVQNDRSGCVQTLPQQGS.
Mus_musculus_iRhom2 431 IGPSSIDLIHHGAKFSPQIRKDQQIEQLVRRERDLERTSGCCVQNDRSGCIQTLKKQGS.
Homo_sapiens_iRhom2 459 VGPSSIDLIHHGAKFSPQIRKDGQIEQLVLRERDLERDSGCCVQNDHSGCIQTQRKQGS.
Canis_lupus_familiaris_iRhom2 430 IGPSSIDLIHHGAKFSPQIRKDQQIEQLVLRERDLERDSGCCVQNDHSGCIQTQRKQGS.

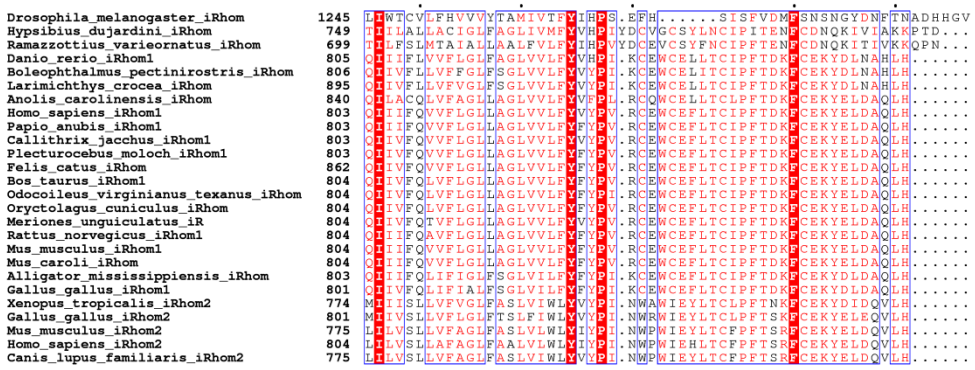

Drosophila_melanogaster_iRhom 952 RGLYPTKSISTWKKWSPSESGP.....GGRISGSVCGIDPKFKCDAPASIAPYEWPDDI
Hypsibius_dujardini_iRhom 471 .....KTMATFYKFPN..K.....HEEHRVSGSVCGIDPHYCSDPFSVGPHTWPDDI
Ramazzottius_varieornatus_iRhom 419 .....KTMATFHKWPGNVK.....HSDERLSGSVCGIDPRYCEDPVSTGPHAWPDDI
Danio_rerio_iRhom1 522 .....STLAVWVKWPQHPSVP..Q.LDGVARQHGSVCHQDPRTCTEPASVSPHEWPDDI
Boleophthalmus_pectinirostris_iRhom 523 .....STLAVWVKWPKHSSTP..Q.LNGKDRQHGSVCHQDPRICLEPASGSPHEWPDDI
Larimichthys_crocea_iRhom 612 .....STLAVWVKWPRHSSTP..Q.LNGKDRQYGSVCHQDPRICLEPASVSPHEWPDDI
Anolis_carolinensis_iRhom 556 .....STLAVWVKWPGHPSAP..PLLAGDKRQFGSVCHQDPRVCEQPASMDPHEWPDDI
Homo_sapiens_iRhom1 520 .....STLAVWVKWPIHPSAP..E.LAGHKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Papio_anubis_iRhom1 520 .....STLAVWVKWPIHPSAP..E.LAGHKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Callithrix_jacchus_iRhom1 520 .....STLAVWVKWPVHPSAP..E.LAGHKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Plecturocebus_moloch_iRhom1 520 .....STLAVWVKWPVHPSAP..E.LAGHKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Felis_catus_iRhom 579 .....STLAVWVKWPFHPSAP..D.LAGRKRQFGSVCHQDPRVCDEPSSEDPHEWPDDI
Bos_taurus_iRhom1 521 .....STLAVWVKWPLHPSAP..D.LAGQKRRYGSVCHQDPRVCDEPSSEDPHEWPDDI
Odocoileus_virginianus_texanus_iRhom 521 .....STLAVWVKWPLHPSAP..D.LAGQKRRYGSVCHQDPRVCDEPSSEDPHEWPDDI
Oryctolagus_cuniculus_iRhom 521 .....STLAVWVKWPIHPSAP..D.LAGQKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Meriones_unguiculatus_iR 521 .....STLAVWVKWPIHPSAP..D.LAGNKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Rattus_norvegicus_iRhom1 521 .....STLAVWVKWPVHPSAP..D.LAGNKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Mus_musculus_iRhom1 521 .....STLAVWVKWPVHPSAP..D.LAGNKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Mus_caroli_iRhom 521 .....STLAVWVKWPIHPSAP..D.LAGNKRQFGSVCHQDPRVCDEPSSEDPHEWPEDI
Alligator_mississippiensis_iRhom 520 .....STLAVWVKWPYHPSTP..T.LAGNKRQFGSVCHQDPRVCEQPASMEPHEWPDDI
Gallus_gallus_iRhom1 518 .....STLAVWVKWPHHSSTP..M.LAGSKRQFGSVCHQDPRVCEQPASMEPHEWPDDI
Xenopus_tropicalis_iRhom2 489 .....ETLATFIKWPEHDAPVLDEV.TGMKRTSGAVCHQDPRTCEEPASIQPHVWADDI
Gallus_gallus_iRhom2 518 .....ETLATFIKWPNNNNTPTMG...SGEKRTSGAVCHQDPRTCEEPASNPPHVWPDDI
Mus_musculus_iRhom2 490 .....ETLATFVKWQNDTGP.SDKSDLSQKQPSAVVCHQDPRTCEEPASSGAHIWPDDI
Homo_sapiens_iRhom2 518 .....ETLATFVKWQDDTGPPMDKSDLGQKRTSGAVCHQDPRTCEEPASSGAHIWPDDI
Canis_lupus_familiaris_iRhom2 489 .....ETLATFVKWQDDTGPPMDKSDLGQKRTSGAVCHQDPRTCEEPASSGAHIWPDDI


Drosophila_melanogaster_iRhom 1005 TKWPICRKTNSFTQRYRYKDHTSEHMVCEVIGHPCCTGLYGEGHITTREYCDFIKGYFHE
Hypsibius_dujardini_iRhom 516 TRWPICKKKANVSI.....PRAAPHMSCEVVGHPCCTGTRGEGQISTREYCDFVQGYFHE
Ramazzottius_varieornatus_iRhom 466 TRWPVCKTKANISM.....STAPPHMTCEVVGHPCCTGIRGEGQITTREYCDFVQGYFHE
Danio_rerio_iRhom1 573 TKWPICTKYN.SGN.....HTNLPHIDCTITGHPCCIGTKGRGHITSREYCDFMKGYFHE
Boleophthalmus_pectinirostris_iRhom 574 TKWPICTRYN.TGN.....HTNLPHIDCTITGHPCCIGTKGRGHITSREYCDFMMGYFHE
Larimichthys_crocea_iRhom 663 TKWPICTRYN.TGN.....HTNLPHIDCTITGHPCCIGTKGRGHITSREYCDFMKGYFHE
Anolis_carolinensis_iRhom 608 TRWPICIKNS.AGN.....RTNHLHMDCVITGHPCCITGKGRGHITSREYCTFMHGYFHE
Homo_sapiens_iRhom1 571 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Papio_anubis_iRhom1 571 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Callithrix_jacchus_iRhom1 571 TRWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Plecturocebus_moloch_iRhom1 571 TRWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Felis_catus_iRhom 630 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Bos_taurus_iRhom1 572 TKWPICTKSS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Odocoileus_virginianus_texanus_iRhom 572 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Oryctolagus_cuniculus_iRhom 572 TKWPICTKSS.AGN.....HTNHPHMDCVLTGHPCCITGKGRGHITSREYCDFMRGYFHE
Meriones_unguiculatus_iR 572 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Rattus_norvegicus_iRhom1 572 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMKGYFHE
Mus_musculus_iRhom1 572 TKWPICTKSS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Mus_caroli_iRhom 572 TKWPICTKNS.AGN.....HTNHPHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Alligator_mississippiensis_iRhom 571 TKWPICTKNS.AGN.....YTNHLHMDCVITGHPCCITGKGRGHITSREYCDFMKGYFHE
Gallus_gallus_iRhom1 569 TKWPICTKNS.AGN.....YTNHLHMDCVITGHPCCITGKGRGHITSREYCDFMRGYFHE
Xenopus_tropicalis_iRhom2 542 TKWPICTYQA.MNN.....HTGIRHMDCEIKGHPCCITGKGSGHITTREYCTFMHGYFHE
Gallus_gallus_iRhom2 569 TRWPICTYET.KTN.....HTGLPHMDCQIKGHPCCITGKGSGHITTREYCEFMHGYFHE
Mus_musculus_iRhom2 543 TKWPICTEQA.QSN.....HTGLLHIDCKIKGHPCCITGKGSGHITTREYCEFMHGYFHE
Homo_sapiens_iRhom2 572 TKWPICTEQA.RSN.....HTGFLHMDCEIKGHPCCITGKGSGHITTREYCEFMHGYFHE
Canis_lupus_familiaris_iRhom2 543 TKWPICTEQA.KSN.....RTGFLHMDCQIKGHPCCISTKGSGHITTREYCEFMHGYFHE


Drosophila_melanogaster_iRhom 1065 EASLCSQISCLNNVCGMPPFISVETPDQEYRLLTSLCMHAGILHLAITLIFQHLFLADLE
Hypsibius_dujardini_iRhom 571 DALLCSQVSCMDKICGMPPLLDPEVPDQEYRLWTAIFLHAGFIHLLITLVFQYVVMRDAE
Ramazzottius_varieornatus_iRhom 521 DALLCSQVSCMDKICGMPPLLDPEVPDQEYRLWTGLFLHAGILHLLISVVFQYFIMRDAE
Danio_rerio_iRhom1 627 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTILRDLE
Boleophthalmus_pectinirostris_iRhom 628 EALLCSQVHCMDNVCGLLPFLNPEIPDQEYRLWLSLFLHAGILHLCLVSVAFQMTILRDLE
Larimichthys_crocea_iRhom 717 EALLCSQVHCMDDVCGLLPFLNPEIPDQEYRLWLSLFLHAGILHLCLVSVAFQMTILRDLE
Anolis_carolinensis_iRhom 662 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTILRDLE
Homo_sapiens_iRhom1 625 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSICFQMTVLRDLE
Papio_anubis_iRhom1 625 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSICFQMTVLRDLE
Callithrix_jacchus_iRhom1 625 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSICFQMTVLRDLE
Plecturocebus_moloch_iRhom1 625 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSICFQMTVLRDLE
Felis_catus_iRhom 684 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTVLRDLE
Bos_taurus_iRhom1 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGVLHLCLVSVCFQMTVLRDLE
Odocoileus_virginianus_texanus_iRhom 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGVLHLCLVSVCFQMTVLRDLE
Oryctolagus_cuniculus_iRhom 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTVLRDLE
Meriones_unguiculatus_iR 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTVLRDLE
Rattus_norvegicus_iRhom1 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTVLRDLE
Mus_musculus_iRhom1 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTVLRDLE
Mus_caroli_iRhom 626 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTVLRDLE
Alligator_mississippiensis_iRhom 625 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTILRDLE
Gallus_gallus_iRhom1 623 EALLCSQVHCMDDVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVCFQMTILRDLE
Xenopus_tropicalis_iRhom2 596 EALLCSQVHCLDEVCGLLPFLNPEVPDQEYRLWLSLFLHAGVLHLCCVSVVFQMTVLRDLE
Gallus_gallus_iRhom2 623 EALLCSQVHCLDEVCGLLPFLNPEVPDQEYRLWLSLFLHAGILHLCLVSVTFQMTVLRDLE
Mus_musculus_iRhom2 597 DALLCSQVHCLDKVCGLLPFLNPEVPDQEYRIWLSLFLHAGIVHLCLVSVVFQMTILRDLE
Homo_sapiens_iRhom2 626 EALLCSQVHCLDKVCGLLPFLNPEVPDQEYRLWLSLFLHAGVVHLCLVSVVFQMTILRDLE
Canis_lupus_familiaris_iRhom2 597 EALLCSQVHCLDKVCGLLPFLNPEVPDQEYRLWLSLFLHAGVVHLCLVSVVFQMTILRDLE


Drosophila_melanogaster_iRhom 1125
Hypsibius_dujardini_iRhom 631
Ramazzottius_varieornatus_iRhom 581
Danio_rerio_iRhom1 687
Boleophthalmus_pectinirostris_iRhom 688
Larimichthys_crocea_iRhom 777
Anolis_carolinensis_iRhom 722
Homo_sapiens_iRhom1 685
Papio_anubis_iRhom1 685
Callithrix_jacchus_iRhom1 685
Plecturocebus_moloch_iRhom1 685
Felis_catus_iRhom 744
Bos_taurus_iRhom1 686
Odocoileus_virginianus_texanus_iRhom 686
Oryctolagus_cuniculus_iRhom 686
Meriones_unguiculatus_iR 686
Rattus_norvegicus_iRhom1 686
Mus_musculus_iRhom1 686
Mus_caroli_iRhom 686
Alligator_mississippiensis_iRhom 685
Gallus_gallus_iRhom1 683
Xenopus_tropicalis_iRhom2 656
Gallus_gallus_iRhom2 683
Mus_musculus_iRhom2 657
Homo_sapiens_iRhom2 686
Canis_lupus_familiaris_iRhom2 657


Drosophila_melanogaster_iRhom 1185
Hypsibius_dujardini_iRhom 690
Ramazzottius_varieornatus_iRhom 640
Danio_rerio_iRhom1 746
Boleophthalmus_pectinirostris_iRhom 747
Larimichthys_crocea_iRhom 836
Anolis_carolinensis_iRhom 781
Homo_sapiens_iRhom1 744
Papio_anubis_iRhom1 744
Callithrix_jacchus_iRhom1 744
Plecturocebus_moloch_iRhom1 744
Felis_catus_iRhom 803
Bos_taurus_iRhom1 745
Odocoileus_virginianus_texanus_iRhom 745
Oryctolagus_cuniculus_iRhom 745
Meriones_unguiculatus_iR 745
Rattus_norvegicus_iRhom1 745
Mus_musculus_iRhom1 745
Mus_caroli_iRhom 745
Alligator_mississippiensis_iRhom 744
Gallus_gallus_iRhom1 742
Xenopus_tropicalis_iRhom2 715
Gallus_gallus_iRhom2 742
Mus_musculus_iRhom2 716
Homo_sapiens_iRhom2 745
Canis_lupus_familiaris_iRhom2 716


Drosophila_melanogaster_iRhom 1245
Hypsibius_dujardini_iRhom 749
Ramazzottius_varieornatus_iRhom 699
Danio_rerio_iRhom1 805
Boleophthalmus_pectinirostris_iRhom 806
Larimichthys_crocea_iRhom 895
Anolis_carolinensis_iRhom 840
Homo_sapiens_iRhom1 803
Papio_anubis_iRhom1 803
Callithrix_jacchus_iRhom1 803
Plecturocebus_moloch_iRhom1 803
Felis_catus_iRhom 862
Bos_taurus_iRhom1 804
Odocoileus_virginianus_texanus_iRhom 804
Oryctolagus_cuniculus_iRhom 804
Meriones_unguiculatus_iR 804
Rattus_norvegicus_iRhom1 804
Mus_musculus_iRhom1 804
Mus_caroli_iRhom 804
Alligator_mississippiensis_iRhom 803
Gallus_gallus_iRhom1 801
Xenopus_tropicalis_iRhom2 774
Gallus_gallus_iRhom2 801
Mus_musculus_iRhom2 775
Homo_sapiens_iRhom2 804
Canis_lupus_familiaris_iRhom2 775

```
                                                  .         .         .         .         .         .
Drosophila_melanogaster_iRhom          1298  DVVSSNTRYSQTQNSQYYYHHHSDDIIRKSVTFTEKALVSHILYPTAPRKTSAQQWQEVE
Hypsibius_dujardini_iRhom                    ............................................................
Ramazzottius_varieornatus_iRhom              ............................................................
Danio_rerio_iRhom1                           ............................................................
Boleophthalmus_pectinirostris_iRhom          ............................................................
Larimichthys_crocea_iRhom                    ............................................................
Anolis_carolinensis_iRhom                    ............................................................
Homo_sapiens_iRhom1                          ............................................................
Papio_anubis_iRhom1                          ............................................................
Callithrix_jacchus_iRhom1                    ............................................................
Plecturocebus_moloch_iRhom1                  ............................................................
Felis_catus_iRhom                            ............................................................
Bos_taurus_iRhom1                            ............................................................
Odocoileus_virginianus_texanus_iRhom         ............................................................
Oryctolagus_cuniculus_iRhom                  ............................................................
Meriones_unguiculatus_iR                     ............................................................
Rattus_norvegicus_iRhom1                     ............................................................
Mus_musculus_iRhom1                          ............................................................
Mus_caroli_iRhom                             ............................................................
Alligator_mississippiensis_iRhom             ............................................................
Gallus_gallus_iRhom1                         ............................................................
Xenopus_tropicalis_iRhom2                    ............................................................
Gallus_gallus_iRhom2                         ............................................................
Mus_musculus_iRhom2                          ............................................................
Homo_sapiens_iRhom2                          ............................................................
Canis_lupus_familiaris_iRhom2                ............................................................

                                                  .         .         .         .         .         .
Drosophila_melanogaster_iRhom          1358  YSRSFNHLSNYSDRIKKSIGNISKLKQVFTSPIRFSNKNNHSNLMTELTSVHSENKQKYL
Hypsibius_dujardini_iRhom                    ............................................................
Ramazzottius_varieornatus_iRhom              ............................................................
Danio_rerio_iRhom1                           ............................................................
Boleophthalmus_pectinirostris_iRhom          ............................................................
Larimichthys_crocea_iRhom                    ............................................................
Anolis_carolinensis_iRhom                    ............................................................
Homo_sapiens_iRhom1                          ............................................................
Papio_anubis_iRhom1                          ............................................................
Callithrix_jacchus_iRhom1                    ............................................................
Plecturocebus_moloch_iRhom1                  ............................................................
Felis_catus_iRhom                            ............................................................
Bos_taurus_iRhom1                            ............................................................
Odocoileus_virginianus_texanus_iRhom         ............................................................
Oryctolagus_cuniculus_iRhom                  ............................................................
Meriones_unguiculatus_iR                     ............................................................
Rattus_norvegicus_iRhom1                     ............................................................
Mus_musculus_iRhom1                          ............................................................
Mus_caroli_iRhom                             ............................................................
Alligator_mississippiensis_iRhom             ............................................................
Gallus_gallus_iRhom1                         ............................................................
Xenopus_tropicalis_iRhom2                    ............................................................
Gallus_gallus_iRhom2                         ............................................................
Mus_musculus_iRhom2                          ............................................................
Homo_sapiens_iRhom2                          ............................................................
Canis_lupus_familiaris_iRhom2                ............................................................

                                                  .
Drosophila_melanogaster_iRhom          1418  GYINNNTEFNVL
Hypsibius_dujardini_iRhom                    ............
Ramazzottius_varieornatus_iRhom              ............
Danio_rerio_iRhom1                           ............
Boleophthalmus_pectinirostris_iRhom          ............
Larimichthys_crocea_iRhom                    ............
Anolis_carolinensis_iRhom                    ............
Homo_sapiens_iRhom1                          ............
Papio_anubis_iRhom1                          ............
Callithrix_jacchus_iRhom1                    ............
Plecturocebus_moloch_iRhom1                  ............
Felis_catus_iRhom                            ............
Bos_taurus_iRhom1                            ............
Odocoileus_virginianus_texanus_iRhom         ............
Oryctolagus_cuniculus_iRhom                  ............
Meriones_unguiculatus_iR                     ............
Rattus_norvegicus_iRhom1                     ............
Mus_musculus_iRhom1                          ............
Mus_caroli_iRhom                             ............
Alligator_mississippiensis_iRhom             ............
Gallus_gallus_iRhom1                         ............
Xenopus_tropicalis_iRhom2                    ............
Gallus_gallus_iRhom2                         ............
Mus_musculus_iRhom2                          ............
Homo_sapiens_iRhom2                          ............
Canis_lupus_familiaris_iRhom2                ............
```

Appendix D: Derivation of physical lentiviral titre

Physical titre (TU/mL) was calculated using this formula:

interpolated result (pg/mL) x 100 (TU/pg) x dilution factor

1 lentiviral particle (LP) contains approximately 2,000 molecules of p24 (24 kDa)

Therefore:

1 LP = (2,000 x 24,000 $g.mol^{-1}$) / 6 x 10^{23} mol^{-1}

1 LP = 8 x 10^{-17} g of p24

There is approximately 1 LP per 1 x 10^{-16} g of p24 (1 LP per 1 x 10^{-4} pg of p24) so 1,000 LP in 1 pg of p24

There are approximately 100 transducing units (TU) per 1,000 LP

Therefore 100 TU per 1 pg of p24

References

ADRAIN, C. & FREEMAN, M. 2012. New lives for old: evolution of pseudoenzyme function illustrated by iRhoms. *Nat Rev Mol Cell Biol,* 13**,** 489-98.

ADRAIN, C. & FREEMAN, M. 2014. Regulation of receptor tyrosine kinase ligand processing. *Cold Spring Harb Perspect Biol,* 6.

ADRAIN, C., ZETTL, M., CHRISTOVA, Y., TAYLOR, N. & FREEMAN, M. 2012. Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science,* 335**,** 225-8.

ADRIAN, M., DUBOCHET, J., LEPAULT, J. & MCDOWALL, A. W. 1984. Cryo-electron microscopy of viruses. *Nature,* 308**,** 32-6.

ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. & WALTER, P. 2008. General Principles of Cell Communication. *In:* ANDERSON, M. & GRANUM, S. (eds.) *Molecular Biology of the Cell.* 5th ed. New York: Garland Science.

ARKHIPOVA, V., GUSKOV, A. & SLOTBOOM, D. J. 2020. Structural ensemble of a glutamate transporter homologue in lipid nanodisc environment. *Nature communications,* 11**,** 998.

BACKLIWAL, G., HILDINGER, M., KUETTEL, I., DELEGRANGE, F., HACKER, D. L. & WURM, F. M. 2008. Valproic acid: a viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. *Biotechnol Bioeng,* 101**,** 182-9.

BAI, J., CHIU, W., WANG, J., TZENG, T., PERRIMON, N. & HSU, J. 2001. The cell adhesion molecule Echinoid defines a new pathway that antagonizes the Drosophila EGF receptor signaling pathway. *Development,* 128**,** 591-601.

BAI, X. C., MCMULLAN, G. & SCHERES, S. H. 2015. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci,* 40**,** 49-57.

BAKER, R. P. & URBAN, S. 2012. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol,* 8**,** 759-68.

BAKER, R. P., WIJETILAKA, R. & URBAN, S. 2006. Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS Pathog,* 2**,** e113.

BAKER, R. P., YOUNG, K., FENG, L., SHI, Y. & URBAN, S. 2007. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci U S A,* 104**,** 8257-62.

BEALE, J. H., PARKER, J. L., SAMSUDIN, F., BARRETT, A. L., SENAN, A., BIRD, L. E., SCOTT, D., OWENS, R. J., SANSOM, M. S. P., TUCKER, S. J., MEREDITH, D., FOWLER, P. W. & NEWSTEAD, S. 2015. Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport. *Structure,* 23**,** 1889-1899.

BEN-SHEM, A., FASS, D. & BIBI, E. 2007. Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A,* 104**,** 462-6.

BEPLER, T., MORIN, A., RAPP, M., BRASCH, J., SHAPIRO, L., NOBLE, A. J. & BERGER, B. 2019. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods,* 16**,** 1153-1160.

BERGBOLD, N. & LEMBERG, M. K. 2013. Emerging role of rhomboid family proteins in mammalian biology and disease. *Biochim Biophys Acta,* 1828**,** 2840-8.

BILLESBØLLE, C. B., AZUMAYA, C. M., KRETSCH, R. C., POWERS, A. S., GONEN, S., SCHNEIDER, S., ARVEDSON, T., DROR, R. O., CHENG, Y. & MANGLIK, A. 2020. Structure of hepcidin-bound ferroportin reveals iron homeostatic mechanisms. *Nature*.

BLACK, R. A., RAUCH, C. T., KOZLOSKY, C. J., PESCHON, J. J., SLACK, J. L., WOLFSON, M. F., CASTNER, B. J., STOCKING, K. L., REDDY, P., SRINIVASAN, S., NELSON, N., BOIANI, N., SCHOOLEY, K. A., GERHART, M., DAVIS, R., FITZNER, J. N., JOHNSON, R. S., PAXTON, R. J.,

MARCH, C. J. & CERRETTI, D. P. 1997. A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. *Nature,* 385**,** 729-33.
BLAYDON, D. C., ETHERIDGE, S. L., RISK, J. M., HENNIES, H. C., GAY, L. J., CARROLL, R., PLAGNOL, V., MCRONALD, F. E., STEVENS, H. P., SPURR, N. K., BISHOP, D. T., ELLIS, A., JANKOWSKI, J., FIELD, J. K., LEIGH, I. M., SOUTH, A. P. & KELSELL, D. P. 2012. RHBDF2 mutations are associated with tylosis, a familial esophageal cancer syndrome. *Am J Hum Genet,* 90**,** 340-6.
BLOBEL, C. P. 2005. ADAMs: key components in EGFR signalling and development. *Nat Rev Mol Cell Biol,* 6**,** 32-43.
BLOBEL, C. P., CARPENTER, G. & FREEMAN, M. 2009. The role of protease activity in ErbB biology. *Exp Cell Res,* 315**,** 671-82.
BLOCH, J. S., PESCIULLESI, G., BOILEVIN, J., NOSOL, K., IROBALIEVA, R. N., DARBRE, T., AEBI, M., KOSSIAKOFF, A. A., REYMOND, J. L. & LOCHER, K. P. 2020. Structure and mechanism of the ER-based glucosyltransferase ALG6. *Nature,* 579**,** 443-447.
BONDAR, A. N., DEL VAL, C. & WHITE, S. H. 2009. Rhomboid protease dynamics and lipid interactions. *Structure,* 17**,** 395-405.
BRÄUER, P., PARKER, J. L., GERONDOPOULOS, A., ZIMMERMANN, I., SEEGER, M. A., BARR, F. A. & NEWSTEAD, S. 2019. Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science,* 363**,** 1103-1107.
BRENNER, S. & HORNE, R. W. 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochim Biophys Acta,* 34**,** 103-10.
BREYTON, C., JAVED, W., VERMOT, A., ARNAUD, C. A., HAJJAR, C., DUPUY, J., PETIT-HARTLEIN, I., LE ROY, A., MARTEL, A., THÉPAUT, M., ORELLE, C., JAULT, J. M., FIESCHI, F., PORCAR, L. & EBEL, C. 2019. Assemblies of lauryl maltose neopentyl glycol (LMNG) and LMNG-solubilized membrane proteins. *Biochim Biophys Acta Biomembr,* 1861**,** 939-957.
BROMBERG, R., GUO, Y., BOREK, D. & OTWINOWSKI, Z. 2020. High-resolution cryo-EM reconstructions in the presence of substantial aberrations. *IUCrJ,* 7**,** 445-452.
BUGULISKIS, J. S., BROSSIER, F., SHUMAN, J. & SIBLEY, L. D. 2010. Rhomboid 4 (ROM4) affects the processing of surface adhesins and facilitates host cell invasion by Toxoplasma gondii. *PLoS Pathog,* 6**,** e1000858.
CAESAR, J., REBOUL, C. F., MACHELLO, C., KIESEWETTER, S., TANG, M. L., DEME, J. C., JOHNSON, S., ELMLUND, D., LEA, S. M. & ELMLUND, H. 2020. SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *J Struct Biol***,** 107635.
CALLAWAY, E. 2015. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature,* 525**,** 172-4.
CARPENTER, E. P., BEIS, K., CAMERON, A. D. & IWATA, S. 2008. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol,* 18**,** 581-6.
CAVADAS, M., OIKONOMIDI, I., GASPAR, C. J., BURBRIDGE, E., BADENES, M., FÉLIX, I., BOLADO, A., HU, T., BILECK, A., GERNER, C., DOMINGOS, P. M., VON KRIEGSHEIM, A. & ADRAIN, C. 2017. Phosphorylation of iRhom2 Controls Stimulated Proteolytic Shedding by the Metalloprotease ADAM17/TACE. *Cell Rep,* 21**,** 745-757.
CHAE, P. S., RASMUSSEN, S. G., RANA, R. R., GOTFRYD, K., CHANDRA, R., GOREN, M. A., KRUSE, A. C., NURVA, S., LOLAND, C. J., PIERRE, Y., DREW, D., POPOT, J. L., PICOT, D., FOX, B. G., GUAN, L., GETHER, U., BYRNE, B., KOBILKA, B. & GELLMAN, S. H. 2010. Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature methods,* 7**,** 1003-8.
CHAE, P. S., RASMUSSEN, S. G., RANA, R. R., GOTFRYD, K., KRUSE, A. C., MANGLIK, A., CHO, K. H., NURVA, S., GETHER, U., GUAN, L., LOLAND, C. J., BYRNE, B., KOBILKA, B. K. & GELLMAN, S. H. 2012. A new class of amphiphiles bearing rigid hydrophobic groups for solubilization and stabilization of membrane proteins. *Chemistry,* 18**,** 9485-90.

CHANG, V. T., FERNANDES, R. A., GANZINGER, K. A., LEE, S. F., SIEBOLD, C., MCCOLL, J., JÖNSSON, P., PALAYRET, M., HARLOS, K., COLES, C. H., JONES, E. Y., LUI, Y., HUANG, E., GILBERT, R. J. C., KLENERMAN, D., ARICESCU, A. R. & DAVIS, S. J. 2016. Initiation of T cell signaling by CD45 segregation at 'close contacts'. *Nat Immunol,* 17**,** 574-582.

CHEN, L., HOU, W. T., FAN, T., LIU, B., PAN, T., LI, Y. H., JIANG, Y. L., WEN, W., CHEN, Z. P., SUN, L., ZHOU, C. Z. & CHEN, Y. 2020. Cryo-electron Microscopy Structure and Transport Mechanism of a Wall Teichoic Acid ABC Transporter. *mBio,* 11.

CHEN, Q., ZENG, W., SHE, J., BAI, X. C. & JIANG, Y. 2019. Structural and functional characterization of an otopetrin family proton channel. *Elife,* 8.

CHENG, T. L., WU, Y. T., LIN, H. Y., HSU, F. C., LIU, S. K., CHANG, B. I., CHEN, W. S., LAI, C. H., SHI, G. Y. & WU, H. L. 2011. Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing. *J Invest Dermatol,* 131**,** 2486-94.

CHI, X., JIN, X., CHEN, Y., LU, X., TU, X., LI, X., ZHANG, Y., LEI, J., HUANG, J., HUANG, Z., ZHOU, Q. & PAN, X. 2020. Structural insights into the gating mechanism of human SLC26A9 mediated by its C-terminal sequence. *Cell Discov,* 6**,** 55.

CHO, S., BAKER, R. P., JI, M. & URBAN, S. 2019. Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat Struct Mol Biol,* 26**,** 910-918.

CHRISTOVA, Y., ADRAIN, C., BAMBROUGH, P., IBRAHIM, A. & FREEMAN, M. 2013. Mammalian iRhoms have distinct physiological functions including an essential role in TACE regulation. *EMBO Rep,* 14**,** 884-90.

COLEMAN, J. A., YANG, D., ZHAO, Z., WEN, P. C., YOSHIOKA, C., TAJKHORSHID, E. & GOUAUX, E. 2019. Serotonin transporter-ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature,* 569**,** 141-145.

COURNIA, Z., ALLEN, T. W., ANDRICIOAEI, I., ANTONNY, B., BAUM, D., BRANNIGAN, G., BUCHETE, N. V., DECKMAN, J. T., DELEMOTTE, L., DEL VAL, C., FRIEDMAN, R., GKEKA, P., HEGE, H. C., HÉNIN, J., KASIMOVA, M. A., KOLOCOURIS, A., KLEIN, M. L., KHALID, S., LEMIEUX, M. J., LINDOW, N., ROY, M., SELENT, J., TAREK, M., TOFOLEANU, F., VANNI, S., URBAN, S., WALES, D. J., SMITH, J. C. & BONDAR, A. N. 2015. Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *J Membr Biol,* 248**,** 611-40.

CUI, Y., ZHOU, K., STRUGATSKY, D., WEN, Y., SACHS, G., ZHOU, Z. H. & MUNSON, K. 2019. pH-dependent gating mechanism of the Helicobacter pylori urea channel revealed by cryo-EM. *Sci Adv,* 5**,** eaav8423.

DE ROSIER, D. J. & KLUG, A. 1968. Reconstruction of three dimensional structures from electron micrographs. *Nature,* 217**,** 130-4.

DENG, S., MCTIERNAN, N., WEI, X., ARNESEN, T. & MARMORSTEIN, R. 2020. Molecular basis for N-terminal acetylation by human NatE and its modulation by HYPK. *Nature communications,* 11**,** 818.

DESHPANDE, I., LIANG, J., HEDEEN, D., ROBERTS, K. J., ZHANG, Y., HA, B., LATORRACA, N. R., FAUST, B., DROR, R. O., BEACHY, P. A., MYERS, B. R. & MANGLIK, A. 2019. Smoothened stimulation by membrane sterols drives Hedgehog pathway activity. *Nature,* 571**,** 284-288.

DOKMANOVIC, M., CLARKE, C. & MARKS, P. A. 2007. Histone deacetylase inhibitors: overview and perspectives. *Mol Cancer Res,* 5**,** 981-9.

DONG, M., DEGANUTTI, G., PIPER, S. J., LIANG, Y. L., KHOSHOUEI, M., BELOUSOFF, M. J., HARIKUMAR, K. G., REYNOLDS, C. A., GLUKHOVA, A., FURNESS, S. G. B., CHRISTOPOULOS, A., DANEV, R., WOOTTEN, D., SEXTON, P. M. & MILLER, L. J. 2020. Structure and dynamics of the active Gs-coupled human secretin receptor. *Nature communications,* 11**,** 4137.

DRAPER-JOYCE, C. J., KHOSHOUEI, M., THAL, D. M., LIANG, Y. L., NGUYEN, A. T. N., FURNESS, S. G. B., VENUGOPAL, H., BALTOS, J. A., PLITZKO, J. M., DANEV, R., BAUMEISTER, W., MAY, L.

T., WOOTTEN, D., SEXTON, P. M., GLUKHOVA, A. & CHRISTOPOULOS, A. 2018. Structure of the adenosine-bound human adenosine A(1) receptor-G(i) complex. *Nature,* 558**,** 559-563.

DROZDETSKIY, A., COLE, C., PROCTER, J. & BARTON, G. J. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res,* 43**,** W389-94.

DRULYTE, I., JOHNSON, R. M., HESKETH, E. L., HURDISS, D. L., SCARFF, C. A., PORAV, S. A., RANSON, N. A., MUENCH, S. P. & THOMPSON, R. F. 2018. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr D Struct Biol,* 74**,** 560-571.

DUBOCHET, J., ADRIAN, M., CHANG, J. J., HOMO, J. C., LEPAULT, J., MCDOWALL, A. W. & SCHULTZ, P. 1988. Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys,* 21**,** 129-228.

DULLOO, I., MULIYIL, S. & FREEMAN, M. 2019. The molecular, cellular and pathophysiological roles of iRhom pseudoproteases. *Open Biol,* 9**,** 190003.

DÜSTERHÖFT, S., BABENDREYER, A., GIESE, A. A., FLASSHOVE, C. & LUDWIG, A. 2019. Status update on iRhom and ADAM17: It's still complicated. *Biochim Biophys Acta Mol Cell Res,* 1866**,** 1567-1583.

DÜSTERHÖFT, S., JUNG, S., HUNG, C. W., THOLEY, A., SÖNNICHSEN, F. D., GRÖTZINGER, J. & LORENZEN, I. 2013. Membrane-proximal domain of a disintegrin and metalloprotease-17 represents the putative molecular switch of its shedding activity operated by protein-disulfide isomerase. *J Am Chem Soc,* 135**,** 5776-81.

DÜSTERHÖFT, S., KÜNZEL, U. & FREEMAN, M. 2017. Rhomboid proteases in human disease: Mechanisms and future prospects. *Biochim Biophys Acta Mol Cell Res,* 1864**,** 2200-2209.

EDWARDS, D. R., HANDSLEY, M. M. & PENNINGTON, C. J. 2008. The ADAM metalloproteinases. *Mol Aspects Med,* 29**,** 258-89.

EGELMAN, E. H. 2016. The Current Revolution in Cryo-EM. *Biophys J,* 110**,** 1008-12.

ELEGHEERT, J., BEHIELS, E., BISHOP, B., SCOTT, S., WOOLLEY, R. E., GRIFFITHS, S. C., BYRNE, E. F. X., CHANG, V. T., STUART, D. I., JONES, E. Y., SIEBOLD, C. & ARICESCU, A. R. 2018. Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat Protoc,* 13**,** 2991-3017.

EMBL-EBI. 2020a. Available: https://www.ebi.ac.uk/pdbe/emdb/searchForm.html/ [Accessed 3 July 2020].

EMBL-EBI. 2020b. *Number of released maps achieving given resolution levels* [Online]. Available: https://www.ebi.ac.uk/pdbe/emdb/statistics_num_res.html/ [Accessed 24 July 2020].

ENDRES, K., ANDERS, A., KOJRO, E., GILBERT, S., FAHRENHOLZ, F. & POSTINA, R. 2003. Tumor necrosis factor-alpha converting enzyme is processed by proprotein-convertases to its mature form which is degraded upon phorbol ester stimulation. *Eur J Biochem,* 270**,** 2386-93.

FAN, X., WANG, J., ZHANG, X., YANG, Z., ZHANG, J. C., ZHAO, L., PENG, H. L., LEI, J. & WANG, H. W. 2019. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nature communications,* 10**,** 2386.

FLEIG, L., BERGBOLD, N., SAHASRABUDHE, P., GEIGER, B., KALTAK, L. & LEMBERG, M. K. 2012. Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol Cell,* 47**,** 558-69.

FOLTENYI, K., GREENSPAN, R. J. & NEWPORT, J. W. 2007. Activation of EGFR and ERK by rhomboid signaling regulates the consolidation and maintenance of sleep in Drosophila. *Nat Neurosci,* 10**,** 1160-7.

FREEMAN, M. 2014. The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu Rev Cell Dev Biol,* 30**,** 235-54.

GAO, Y., HU, H., RAMACHANDRAN, S., ERICKSON, J. W., CERIONE, R. A. & SKINIOTIS, G. 2019. Structures of the Rhodopsin-Transducin Complex: Insights into G-Protein Activation. *Mol Cell,* 75**,** 781-790 e3.

GARCÍA-NAFRÍA, J., NEHMÉ, R., EDWARDS, P. C. & TATE, C. G. 2018. Cryo-EM structure of the serotonin 5-HT(1B) receptor coupled to heterotrimeric G(o). *Nature,* 558**,** 620-623.

GILES, N. M., WATTS, A. B., GILES, G. I., FRY, F. H., LITTLECHILD, J. A. & JACOB, C. 2003. Metal and redox modulation of cysteine protein function. *Chem Biol,* 10**,** 677-93.

GRIEVE, A. G., XU, H., KUNZEL, U., BAMBROUGH, P., SIEBER, B. & FREEMAN, M. 2017. Phosphorylation of iRhom2 at the plasma membrane controls mammalian TACE-dependent inflammatory and growth factor signalling. *Elife,* 6.

GRÖTZINGER, J., LORENZEN, I. & DÜSTERHÖFT, S. 2017. Molecular insights into the multilayered regulation of ADAM17: The role of the extracellular region. *Biochim Biophys Acta Mol Cell Res,* 1864**,** 2088-2095.

GSCHWIND, A., HART, S., FISCHER, O. M. & ULLRICH, A. 2003. TACE cleavage of proamphiregulin regulates GPCR-induced proliferation and motility of cancer cells. *EMBO J,* 22**,** 2411-21.

HAMERS-CASTERMAN, C., ATARHOUCH, T., MUYLDERMANS, S., ROBINSON, G., HAMERS, C., SONGA, E. B., BENDAHMAN, N. & HAMERS, R. 1993. Naturally occurring antibodies devoid of light chains. *Nature,* 363**,** 446-8.

HAN, Y., FAN, X., WANG, H., ZHAO, F., TULLY, C. G., KONG, J., YAO, N. & YAN, N. 2020. High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. *Proc Natl Acad Sci U S A,* 117**,** 1009-1014.

HARRIS, R. C., CHUNG, E. & COFFEY, R. J. 2003. EGF receptor ligands. *Exp Cell Res,* 284**,** 2-13.

HAYASHIDA, K., BARTLETT, A. H., CHEN, Y. & PARK, P. W. 2010. Molecular and cellular mechanisms of ectodomain shedding. *Anat Rec (Hoboken),* 293**,** 925-37.

HENDERSON, R., BALDWIN, J. M., CESKA, T. A., ZEMLIN, F., BECKMANN, E. & DOWNING, K. H. 1990. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol,* 213**,** 899-929.

HERZIK, M. A., JR., WU, M. & LANDER, G. C. 2019. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nature communications,* 10**,** 1032.

HOFMANN, S., JANULIENE, D., MEHDIPOUR, A. R., THOMAS, C., STEFAN, E., BRÜCHERT, S., KUHN, B. T., GEERTSMA, E. R., HUMMER, G., TAMPÉ, R. & MOELLER, A. 2019. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature,* 571**,** 580-583.

HOSUR, V., JOHNSON, K. R., BURZENSKI, L. M., STEARNS, T. M., MASER, R. S. & SHULTZ, L. D. 2014. Rhbdf2 mutations increase its protein stability and drive EGFR hyperactivation through enhanced secretion of amphiregulin. *Proc Natl Acad Sci U S A,* 111**,** E2200-9.

HOSUR, V., LOW, B. E., LI, D., STAFFORD, G. A., KOHAR, V., SHULTZ, L. D. & WILES, M. V. 2020. Genes adapt to outsmart gene-targeting strategies in mutant mouse strains by skipping exons to reinitiate transcription and translation. *Genome Biol,* 21**,** 168.

HOSUR, V., LOW, B. E., SHULTZ, L. D. & WILES, M. V. 2017. Genetic deletion of amphiregulin restores the normal skin phenotype in a mouse model of the human skin disease tylosis. *Biol Open,* 6**,** 1174-1179.

HUNTE, C. & MICHEL, H. 2002. Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol,* 12**,** 503-8.

HURDISS, D. L., DRULYTE, I., LANG, Y., SHAMORKINA, T. M., PRONKER, M. F., VAN KUPPEVELD, F. J. M., SNIJDER, J. & DE GROOT, R. J. 2020. Cryo-EM structure of coronavirus-HKU1 haemagglutinin esterase reveals architectural changes arising from prolonged circulation in humans. *bioRxiv***,** 2020.03.25.998963.

ISSUREE, P. D., MARETZKY, T., MCILWAIN, D. R., MONETTE, S., QING, X., LANG, P. A., SWENDEMAN, S. L., PARK-MIN, K. H., BINDER, N., KALLIOLIAS, G. D., YARILINA, A., HORIUCHI, K., IVASHKIV, L. B., MAK, T. W., SALMON, J. E. & BLOBEL, C. P. 2013. iRHOM2 is a critical pathogenic mediator of inflammatory arthritis. *J Clin Invest,* 123**,** 928-32.

JACKSON, S. M., MANOLARIDIS, I., KOWAL, J., ZECHNER, M., TAYLOR, N. M. I., BAUSE, M., BAUER, S., BARTHOLOMAEUS, R., BERNHARDT, G., KOENIG, B., BUSCHAUER, A., STAHLBERG, H., ALTMANN, K. H. & LOCHER, K. P. 2018. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol,* 25**,** 333-340.

JONES, M., BEURON, F., BORG, A., NANS, A., EARL, C. P., BRIGGS, D. C., SNIJDERS, A. P., BOWLES, M., MORRIS, E. P., LINCH, M. & MCDONALD, N. Q. 2020. Cryo-EM structures of the XPF-ERCC1 endonuclease reveal how DNA-junction engagement disrupts an auto-inhibited conformation. *Nature communications,* 11**,** 1120.

JUNGNICKEL, K. E. J., PARKER, J. L. & NEWSTEAD, S. 2018. Structural basis for amino acid transport by the CAT family of SLC7 transporters. *Nature communications,* 9**,** 550.

JÜRGENS, G., WIESCHAUS, E., NÜSSLEIN-VOLHARD, C. & KLUDING, H. 1984. Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : II. Zygotic loci on the third chromosome. *Wilehm Roux Arch Dev Biol,* 193**,** 283-295.

KALLIOLIAS, G. D. & IVASHKIV, L. B. 2016. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. *Nat Rev Rheumatol,* 12**,** 49-62.

KANG, Y., LIU, R., WU, J. X. & CHEN, L. 2019. Structural insights into the mechanism of human soluble guanylate cyclase. *Nature,* 574**,** 206-210.

KATO, H. E., ZHANG, Y., HU, H., SUOMIVUORI, C. M., KADJI, F. M. N., AOKI, J., KRISHNA KUMAR, K., FONSECA, R., HILGER, D., HUANG, W., LATORRACA, N. R., INOUE, A., DROR, R. O., KOBILKA, B. K. & SKINIOTIS, G. 2019. Conformational transitions of a neurotensin receptor 1-G(i1) complex. *Nature,* 572**,** 80-85.

KAWATE, T. & GOUAUX, E. 2006. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure,* 14**,** 673-81.

KERN, D. M., SORUM, B., HOEL, C. M., SRIDHARAN, S., REMIS, J. P., TOSO, D. B. & BROHAWN, S. G. 2020. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv*.

KHOSHOUEI, M., RADJAINIA, M., BAUMEISTER, W. & DANEV, R. 2017. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nature communications,* 8**,** 16099.

KIM, J., TAN, Y. Z., WICHT, K. J., ERRAMILLI, S. K., DHINGRA, S. K., OKOMBO, J., VENDOME, J., HAGENAH, L. M., GIACOMETTI, S. I., WARREN, A. L., NOSOL, K., ROEPE, P. D., POTTER, C. S., CARRAGHER, B., KOSSIAKOFF, A. A., QUICK, M., FIDOCK, D. A. & MANCIA, F. 2019. Structure and drug resistance of the Plasmodium falciparum transporter PfCRT. *Nature,* 576**,** 315-320.

KIM, Y. & CHEN, J. 2018. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. *Science,* 359**,** 915-919.

KIRCHDOERFER, R. N. & WARD, A. B. 2019. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications,* 10**,** 2342.

KOONIN, E. V., MAKAROVA, K. S., ROGOZIN, I. B., DAVIDOVIC, L., LETELLIER, M. C. & PELLEGRINI, L. 2003. The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol,* 4**,** R19.

KREUTZBERGER, A. J. B. & URBAN, S. 2018. Single-Molecule Analyses Reveal Rhomboid Proteins Are Strict and Functional Monomers in the Membrane. *Biophys J,* 115**,** 1755-1761.

KRISSINEL, E. & HENRICK, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2256-68.

KÜHLBRANDT, W. 2014. Biochemistry. The resolution revolution. *Science,* 343**,** 1443-4.

KUHLEN, L., ABRUSCI, P., JOHNSON, S., GAULT, J., DEME, J., CAESAR, J., DIETSCHE, T., MEBRHATU, M. T., GANIEF, T., MACEK, B., WAGNER, S., ROBINSON, C. V. & LEA, S. M. 2018. Structure of the core of the type III secretion system export apparatus. *Nat Struct Mol Biol,* 25**,** 583-590.

KULIG, W., JURKIEWICZ, P., OLŻYŃSKA, A., TYNKKYNEN, J., JAVANAINEN, M., MANNA, M., ROG, T., HOF, M., VATTULAINEN, I. & JUNGWIRTH, P. 2015. Experimental determination and computational interpretation of biophysical properties of lipid bilayers enriched by cholesteryl hemisuccinate. *Biochim Biophys Acta,* 1848**,** 422-32.

KULIG, W., TYNKKYNEN, J., JAVANAINEN, M., MANNA, M., ROG, T., VATTULAINEN, I. & JUNGWIRTH, P. 2014. How well does cholesteryl hemisuccinate mimic cholesterol in saturated phospholipid bilayers? *J Mol Model,* 20**,** 2121.

KUMAR, M., KELLER, B., MAKALOU, N. & SUTTON, R. E. 2001. Systematic determination of the packaging limit of lentiviral vectors. *Hum Gene Ther,* 12**,** 1893-905.

KUMARI, J., VINNAKOTA, R. & KUMAR, J. 2019. Structural and Functional Insights into GluK3-kainate Receptor Desensitization and Recovery. *Sci Rep,* 9**,** 10254.

KÜNZEL, U. 2017. *FRMD8 is a novel regulator of iRhom-dependent ADAM17 activity.*

KUNZEL, U., GRIEVE, A. G., MENG, Y., SIEBER, B., COWLEY, S. A. & FREEMAN, M. 2018. FRMD8 promotes inflammatory and growth factor signalling by stabilising the iRhom/ADAM17 sheddase complex. *Elife,* 7.

LAUBER, F., DEME, J. C., LEA, S. M. & BERKS, B. C. 2018. Type 9 secretion system structures reveal a new protein transport mechanism. *Nature,* 564**,** 77-82.

LAVERTY, D., DESAI, R., UCHAŃSKI, T., MASIULIS, S., STEC, W. J., MALINAUSKAS, T., ZIVANOV, J., PARDON, E., STEYAERT, J., MILLER, K. W. & ARICESCU, A. R. 2019. Cryo-EM structure of the human α1β3γ2 GABA(A) receptor in a lipid bilayer. *Nature,* 565**,** 516-520.

LAZARENO-SAEZ, C., ARUTYUNOVA, E., COQUELLE, N. & LEMIEUX, M. J. 2013. Domain swapping in the cytoplasmic domain of the Escherichia coli rhomboid protease. *J Mol Biol,* 425**,** 1127-42.

LE MAIRE, M., CHAMPEIL, P. & MOLLER, J. V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta,* 1508**,** 86-111.

LEE, J. R., URBAN, S., GARVEY, C. F. & FREEMAN, M. 2001. Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell,* 107**,** 161-71.

LEILEI, Y., BING, L., YANG, L., SHAOXIA, W., YUAN, X., DONGPING, W., HUAHU, Y., SHICHEN, S., GUANGZHOU, Z., RUIYUN, P., LIN, Z. & WENLONG, L. 2014. iRhom2 mutation leads to aberrant hair follicle differentiation in mice. *PLoS One,* 9**,** e115114.

LEMBERG, M. K. & FREEMAN, M. 2007. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res,* 17**,** 1634-46.

LEMIEUX, M. J., FISCHER, S. J., CHERNEY, M. M., BATEMAN, K. S. & JAMES, M. N. 2007. The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. *Proc Natl Acad Sci U S A,* 104**,** 750-4.

LEWIS, M., TARTAGLIA, L. A., LEE, A., BENNETT, G. L., RICE, G. C., WONG, G. H., CHEN, E. Y. & GOEDDEL, D. V. 1991. Cloning and expression of cDNAs for two distinct murine tumor necrosis factor receptors demonstrate one receptor is species specific. *Proc Natl Acad Sci U S A,* 88**,** 2830-4.

LI, X., MARETZKY, T., PEREZ-AGUILAR, J. M., MONETTE, S., WESKAMP, G., LE GALL, S., BEUTLER, B., WEINSTEIN, H. & BLOBEL, C. P. 2017. Structural modeling defines transmembrane residues in ADAM17 that are crucial for Rhbdf2-ADAM17-dependent proteolysis. *J Cell Sci,* 130**,** 868-878.

LI, X., MARETZKY, T., WESKAMP, G., MONETTE, S., QING, X., ISSUREE, P. D., CRAWFORD, H. C., MCILWAIN, D. R., MAK, T. W., SALMON, J. E. & BLOBEL, C. P. 2015. iRhoms 1 and 2 are essential upstream regulators of ADAM17-dependent EGFR signaling. *Proc Natl Acad Sci U S A,* 112**,** 6080-5.

LI, X., MOONEY, P., ZHENG, S., BOOTH, C. R., BRAUNFELD, M. B., GUBBENS, S., AGARD, D. A. & CHENG, Y. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods,* 10**,** 584-90.

LICHTENTHALER, S. F. 2013. iRHOM2 takes control of rheumatoid arthritis. *J Clin Invest,* 123**,** 560-2.

LIN, X., LI, M., WANG, N., WU, Y., LUO, Z., GUO, S., HAN, G. W., LI, S., YUE, Y., WEI, X., XIE, X., CHEN, Y., ZHAO, S., WU, J., LEI, M. & XU, F. 2020. Structural basis of ligand recognition and self-activation of orphan GPR52. *Nature,* 579**,** 152-157.

LIU, B., XU, Y., LI, W. L. & ZENG, L. 2015. Proteomic analysis of differentially expressed skin proteins in iRhom2(Uncv) mice. *BMB Rep,* 48**,** 19-24.

LIU, Y., GONEN, S., GONEN, T. & YEATES, T. O. 2018. Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc Natl Acad Sci U S A,* 115**,** 3362-3367.

LUO, W. W., LI, S., LI, C., LIAN, H., YANG, Q., ZHONG, B. & SHU, H. B. 2016. iRhom2 is essential for innate immunity to DNA viruses by mediating trafficking and stability of the adaptor STING. *Nat Immunol,* 17**,** 1057-66.

LUO, W. W., LI, S., LI, C., ZHENG, Z. Q., CAO, P., TONG, Z., LIAN, H., WANG, S. Y., SHU, H. B. & WANG, Y. Y. 2017. iRhom2 is essential for innate immunity to RNA virus by antagonizing ER- and mitochondria-associated degradation of VISA. *PLoS Pathog,* 13**,** e1006693.

MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R. N., POTTER, S. C., FINN, R. D. & LOPEZ, R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research,* 47**,** W636-W641.

MANEY, S. K., MCILWAIN, D. R., POLZ, R., PANDYRA, A. A., SUNDARAM, B., WOLFF, D., OHISHI, K., MARETZKY, T., BROOKE, M. A., EVERS, A., VASUDEVAN, A. A., AGHAEEPOUR, N., SCHELLER, J., MÜNK, C., HÄUSSINGER, D., MAK, T. W., NOLAN, G. P., KELSELL, D. P., BLOBEL, C. P., LANG, K. S. & LANG, P. A. 2015. Deletions in the cytoplasmic domain of iRhom1 and iRhom2 promote shedding of the TNF receptor by the protease ADAM17. *Sci Signal,* 8**,** ra109.

MANOLARIDIS, I., JACKSON, S. M., TAYLOR, N. M. I., KOWAL, J., STAHLBERG, H. & LOCHER, K. P. 2018. Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. *Nature,* 563**,** 426-430.

MARETZKY, T., MCILWAIN, D. R., ISSUREE, P. D., LI, X., MALAPEIRA, J., AMIN, S., LANG, P. A., MAK, T. W. & BLOBEL, C. P. 2013. iRhom2 controls the substrate selectivity of stimulated ADAM17-dependent ectodomain shedding. *Proc Natl Acad Sci U S A,* 110**,** 11433-8.

MARUTHAPPU, T., CHIKH, A., FELL, B., DELANEY, P. J., BROOKE, M. A., LEVET, C., MONCADA-PAZOS, A., ISHIDA-YAMAMOTO, A., BLAYDON, D., WASEEM, A., LEIGH, I. M., FREEMAN, M. & KELSELL, D. P. 2017. Rhomboid family member 2 regulates cytoskeletal stress-associated Keratin 16. *Nature communications,* 8**,** 14174.

MASIULIS, S., DESAI, R., UCHAŃSKI, T., SERNA MARTIN, I., LAVERTY, D., KARIA, D., MALINAUSKAS, T., ZIVANOV, J., PARDON, E., KOTECHA, A., STEYAERT, J., MILLER, K. W. & ARICESCU, A. R. 2019. GABA(A) receptor signalling mechanisms revealed by structural pharmacology. *Nature,* 565**,** 454-459.

MASKOS, K., FERNANDEZ-CATALAN, C., HUBER, R., BOURENKOV, G. P., BARTUNIK, H., ELLESTAD, G. A., REDDY, P., WOLFSON, M. F., RAUCH, C. T., CASTNER, B. J., DAVIS, R., CLARKE, H. R., PETERSEN, M., FITZNER, J. N., CERRETTI, D. P., MARCH, C. J., PAXTON, R. J., BLACK, R. A. & BODE, W. 1998. Crystal structure of the catalytic domain of human tumor necrosis factor-alpha-converting enzyme. *Proc Natl Acad Sci U S A,* 95**,** 3408-12.

MAVEYRAUD, L. & MOUREY, L. 2020. Protein X-ray Crystallography and Drug Discovery. *Molecules,* 25.

MCILWAIN, D. R., LANG, P. A., MARETZKY, T., HAMADA, K., OHISHI, K., MANEY, S. K., BERGER, T., MURTHY, A., DUNCAN, G., XU, H. C., LANG, K. S., HÄUSSINGER, D., WAKEHAM, A., ITIE-YOUTEN, A., KHOKHA, R., OHASHI, P. S., BLOBEL, C. P. & MAK, T. W. 2012. iRhom2 regulation of TACE controls TNF-mediated protection against Listeria and responses to LPS. *Science,* 335**,** 229-32.

MCMULLAN, G., FARUQI, A. R., CLARE, D. & HENDERSON, R. 2014. Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy,* 147**,** 156-63.

MCMULLAN, G., FARUQI, A. R., HENDERSON, R., GUERRINI, N., TURCHETTA, R., JACOBS, A. & VAN HOFTEN, G. 2009. Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. *Ultramicroscopy,* 109**,** 1144-7.

MCQUIBBAN, G. A., LEE, J. R., ZHENG, L., JUUSOLA, M. & FREEMAN, M. 2006. Normal mitochondrial dynamics requires rhomboid-7 and affects Drosophila lifespan and neuronal function. *Curr Biol,* 16**,** 982-9.

MEESKE, A. J., JIA, N., CASSEL, A. K., KOZLOVA, A., LIAO, J., WIEDMANN, M., PATEL, D. J. & MARRAFFINI, L. A. 2020. A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. *Science,* 369**,** 54-59.

MERK, A., BARTESAGHI, A., BANERJEE, S., FALCONIERI, V., RAO, P., DAVIS, M. I., PRAGANI, R., BOXER, M. B., EARL, L. A., MILNE, J. L. S. & SUBRAMANIAM, S. 2016. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell,* 165**,** 1698-1707.

MILLER, P. S. & ARICESCU, A. R. 2014. Crystal structure of a human GABAA receptor. *Nature,* 512**,** 270-5.

MITCHELL, E. M., ARTYMIUK, P. J., RICE, D. W. & WILLETT, P. 1990. Use of techniques derived from graph theory to compare secondary structure motifs in proteins. *J Mol Biol,* 212**,** 151-66.

MITRA, A. K. 2019. Visualization of biological macromolecules at near-atomic resolution: cryo-electron microscopy comes of age. *Acta Crystallogr F Struct Biol Commun,* 75**,** 3-11.

MOIN, S. M. & URBAN, S. 2012. Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife,* 1**,** e00173.

MOKOENA, T., SMIT, J. G. M., KARUSSEIT, V. O., DORFLING, C. M. & VAN RENSBURG, E. J. 2018. Tylosis associated with squamous cell carcinoma of the oesophagus (TOC): Report of an African family with a novel RHBDF2 variant. *Clin Genet,* 93**,** 1114-1116.

MOSS, M. L., JIN, S. L., MILLA, M. E., BICKETT, D. M., BURKHART, W., CARTER, H. L., CHEN, W. J., CLAY, W. C., DIDSBURY, J. R., HASSLER, D., HOFFMAN, C. R., KOST, T. A., LAMBERT, M. H., LEESNITZER, M. A., MCCAULEY, P., MCGEEHAN, G., MITCHELL, J., MOYER, M., PAHEL, G., ROCQUE, W., OVERTON, L. K., SCHOENEN, F., SEATON, T., SU, J. L., BECHERER, J. D. & ET AL. 1997. Cloning of a disintegrin metalloproteinase that processes precursor tumour-necrosis factor-alpha. *Nature,* 385**,** 733-6.

MUIR, K. W., LI, Y., WEIS, F. & PANNE, D. 2020. The structure of the cohesin ATPase elucidates the mechanism of SMC-kleisin ring opening. *Nat Struct Mol Biol,* 27**,** 233-239.

MULIYIL, S., LEVET, C., DÜSTERHÖFT, S., DULLOO, I., COWLEY, S. A. & FREEMAN, M. 2020. ADAM17-triggered TNF signalling protects the ageing Drosophila retina from lipid droplet-mediated degeneration. *EMBO J***,** e104415.

MUYLDERMANS, S. 2013. Nanobodies: natural single-domain antibodies. *Annu Rev Biochem,* 82**,** 775-97.

NAYDENOVA, K. & RUSSO, C. J. 2017. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature communications,* 8**,** 629.

NOGALES, E. 2016. The development of cryo-EM into a mainstream structural biology technique. *Nature methods,* 13**,** 24-7.

NOGALES, E. & SCHERES, S. H. 2015. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol Cell,* 58**,** 677-89.

NYGAARD, R., KIM, J. & MANCIA, F. 2020. Cryo-electron microscopy analysis of small membrane proteins. *Curr Opin Struct Biol,* 64**,** 26-33.

O'DONNELL, R. A., HACKETT, F., HOWELL, S. A., TREECK, M., STRUCK, N., KRNAJSKI, Z., WITHERS-MARTINEZ, C., GILBERGER, T. W. & BLACKMAN, M. J. 2006. Intramembrane proteolysis

mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J Cell Biol,* 174**,** 1023-33.
O'MALLEY, M. A., HELGESON, M. E., WAGNER, N. J. & ROBINSON, A. S. 2011. The morphology and composition of cholesterol-rich micellar nanostructures determine transmembrane protein (GPCR) activity. *Biophys J,* 100**,** L11-3.
OH, S., PAKNEJAD, N. & HITE, R. K. 2020. Gating and selectivity mechanisms for the lysosomal K(+) channel TMEM175. *Elife,* 9.
OIDE, M., KATO, T., OROGUCHI, T. & NAKASAKO, M. 2020. Energy landscape of domain motion in glutamate dehydrogenase deduced from cryo-electron microscopy. *FEBS J*.
OIKONOMIDI, I., BURBRIDGE, E., CAVADAS, M., SULLIVAN, G., COLLIS, B., NAEGELE, H., CLANCY, D., BREZINOVA, J., HU, T., BILECK, A., GERNER, C., BOLADO, A., VON KRIEGSHEIM, A., MARTIN, S. J., STEINBERG, F., STRISOVSKY, K. & ADRAIN, C. 2018. iTAP, a novel iRhom interactor, controls TNF secretion by policing the stability of iRhom/TACE. *Elife,* 7.
OLSEN, J. A., ALAM, A., KOWAL, J., STIEGER, B. & LOCHER, K. P. 2020. Structure of the human lipid exporter ABCB4 in a lipid environment. *Nat Struct Mol Biol,* 27**,** 62-70.
ORLOVA, E. V. & SAIBIL, H. R. 2011. Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev,* 111**,** 7710-48.
PADILLA, J. E., COLOVOS, C. & YEATES, T. O. 2001. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A,* 98**,** 2217-21.
PAPASERGI-SCOTT, M. M., ROBERTSON, M. J., SEVEN, A. B., PANOVA, O., MATHIESEN, J. M. & SKINIOTIS, G. 2020. Structures of metabotropic GABA(B) receptor. *Nature,* 584**,** 310-314.
PARK, J., FU, Z., FRANGAJ, A., LIU, J., MOSYAK, L., SHEN, T., SLAVKOVICH, V. N., RAY, K. M., TAURA, J., CAO, B., GENG, Y., ZUO, H., KOU, Y., GRASSUCCI, R., CHEN, S., LIU, Z., LIN, X., WILLIAMS, J. P., RICE, W. J., ENG, E. T., HUANG, R. K., SONI, R. K., KLOSS, B., YU, Z., JAVITCH, J. A., HENDRICKSON, W. A., SLESINGER, P. A., QUICK, M., GRAZIANO, J., YU, H., FIEHN, O., CLARKE, O. B., FRANK, J. & FAN, Q. R. 2020. Structure of human GABA(B) receptor in an inactive state. *Nature,* 584**,** 304-309.
PARKER, J. L., COREY, R. A., STANSFELD, P. J. & NEWSTEAD, S. 2019. Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nature communications,* 10**,** 4657.
PASSMORE, L. A. & RUSSO, C. J. 2016. Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzymol,* 579**,** 51-86.
PENCZEK, P. A. 2010. Fundamentals of three-dimensional reconstruction from projections. *Methods Enzymol,* 482**,** 1-33.
REBOUL, C. F., KIESEWETTER, S., EAGER, M., BELOUSOFF, M., CUI, T., DE STERCK, H., ELMLUND, D. & ELMLUND, H. 2018. Rapid near-atomic resolution single-particle 3D reconstruction with SIMPLE. *J Struct Biol,* 204**,** 172-181.
REMPEL, S., GATI, C., NIJLAND, M., THANGARATNARAJAH, C., KARYOLAIMOS, A., DE GIER, J. W., GUSKOV, A. & SLOTBOOM, D. J. 2020. A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature,* 580**,** 409-412.
ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res,* 42**,** W320-4.
ROLLAUER, S. E., TARRY, M. J., GRAHAM, J. E., JÄÄSKELÄINEN, M., JÄGER, F., JOHNSON, S., KREHENBRINK, M., LIU, S. M., LUKEY, M. J., MARCOUX, J., MCDOWELL, M. A., RODRIGUEZ, F., ROVERSI, P., STANSFELD, P. J., ROBINSON, C. V., SANSOM, M. S., PALMER, T., HÖGBOM, M., BERKS, B. C. & LEA, S. M. 2012. Structure of the TatC core of the twin-arginine protein transport system. *Nature,* 492**,** 210-4.
RUBINSTEIN, J. L. & BRUBAKER, M. A. 2015. Alignment of cryo-EM movies of individual particles by optimization of image translations. *J Struct Biol,* 192**,** 188-95.

RUSSO, C. J. & PASSMORE, L. A. 2014. Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. *Nature methods,* 11**,** 649-52.
RUTLEDGE, B. J., ZHANG, K., BIER, E., JAN, Y. N. & PERRIMON, N. 1992. The Drosophila spitz gene encodes a putative EGF-like growth factor involved in dorsal-ventral axis formation and neurogenesis. *Genes Dev,* 6**,** 1503-17.
SAARINEN, S., VAHTERISTO, P., LEHTONEN, R., AITTOMÄKI, K., LAUNONEN, V., KIVILUOTO, T. & AALTONEN, L. A. 2012. Analysis of a Finnish family confirms RHBDF2 mutations as the underlying factor in tylosis with esophageal cancer. *Fam Cancer,* 11**,** 525-8.
SAFARIAN, S., HAHN, A., MILLS, D. J., RADLOFF, M., EISINGER, M. L., NIKOLAEV, A., MEIER-CREDO, J., MELIN, F., MIYOSHI, H., GENNIS, R. B., SAKAMOTO, J., LANGER, J. D., HELLWIG, P., KÜHLBRANDT, W. & MICHEL, H. 2019. Active site rearrangement and structural divergence in prokaryotic respiratory oxidases. *Science,* 366**,** 100-104.
SAHIN, U. & BLOBEL, C. P. 2007. Ectodomain shedding of the EGF-receptor ligand epigen is mediated by ADAM17. *FEBS Lett,* 581**,** 41-4.
SAHIN, U., WESKAMP, G., KELLY, K., ZHOU, H. M., HIGASHIYAMA, S., PESCHON, J., HARTMANN, D., SAFTIG, P. & BLOBEL, C. P. 2004. Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands. *J Cell Biol,* 164**,** 769-79.
SAMPATHKUMAR, P., MAK, M. W., FISCHER-WITHOLT, S. J., GUIGARD, E., KAY, C. M. & LEMIEUX, M. J. 2012. Oligomeric state study of prokaryotic rhomboid proteases. *Biochim Biophys Acta,* 1818**,** 3090-7.
SAOTOME, K., TENG, B., TSUI, C. C. A., LEE, W. H., TU, Y. H., KAPLAN, J. P., SANSOM, M. S. P., LIMAN, E. R. & WARD, A. B. 2019. Structures of the otopetrin proton channels Otop1 and Otop3. *Nat Struct Mol Biol,* 26**,** 518-525.
SCHERES, S. H. 2014. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife,* 3**,** e03665.
SCHERES, S. H. 2016. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol,* 579**,** 125-57.
SCHERES, S. H., GAO, H., VALLE, M., HERMAN, G. T., EGGERMONT, P. P., FRANK, J. & CARAZO, J. M. 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature methods,* 4**,** 27-9.
SEDDON, A. M., CURNOW, P. & BOOTH, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta,* 1666**,** 105-17.
SEEGAR, T. C. M., KILLINGSWORTH, L. B., SAHA, N., MEYER, P. A., PATRA, D., ZIMMERMAN, B., JANES, P. W., RUBINSTEIN, E., NIKOLOV, D. B., SKINIOTIS, G., KRUSE, A. C. & BLACKLOW, S. C. 2017. Structural Basis for Regulated Proteolysis by the α-Secretase ADAM10. *Cell,* 171**,** 1638-1648 e7.
SERNA, M. 2019. Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes. *Front Mol Biosci,* 6**,** 33.
SEVEN, A. B., HILGER, D., PAPASERGI-SCOTT, M. M., ZHANG, L., QU, Q., KOBILKA, B. K., TALL, G. G. & SKINIOTIS, G. 2020. Structures of Gα Proteins in Complex with Their Chaperone Reveal Quality Control Mechanisms. *Cell Rep,* 30**,** 3699-3709 e6.
SHAIK, M. M., PENG, H., LU, J., RITS-VOLLOCH, S., XU, C., LIAO, M. & CHEN, B. 2019. Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature,* 565**,** 318-323.
SHARIF, H., WANG, L., WANG, W. L., MAGUPALLI, V. G., ANDREEVA, L., QIAO, Q., HAUENSTEIN, A. V., WU, Z., NÚÑEZ, G., MAO, Y. & WU, H. 2019. Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome. *Nature,* 570**,** 338-343.
SHAYE, H., ISHCHENKO, A., LAM, J. H., HAN, G. W., XUE, L., RONDARD, P., PIN, J. P., KATRITCH, V., GATI, C. & CHEREZOV, V. 2020. Structural basis of the activation of a metabotropic GABA receptor. *Nature,* 584**,** 298-303.

SHERRATT, A. R., BLAIS, D. R., GHASRIANI, H., PEZACKI, J. P. & GOTO, N. K. 2012. Activity-based protein profiling of the Escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry,* 51**,** 7794-803.

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. *Molecular systems biology,* 7**,** 539.

SIGGS, O. M., GRIEVE, A., XU, H., BAMBROUGH, P., CHRISTOVA, Y. & FREEMAN, M. 2014. Genetic interaction implicates iRhom2 in the regulation of EGF receptor signalling in mice. *Biol Open,* 3**,** 1151-7.

SIGGS, O. M., XIAO, N., WANG, Y., SHI, H., TOMISATO, W., LI, X., XIA, Y. & BEUTLER, B. 2012. iRhom2 is required for the secretion of mouse TNFα. *Blood,* 119**,** 5769-71.

SIGWORTH, F. J. 2016. Principles of cryo-EM single-particle image processing. *Microscopy (Oxf),* 65**,** 57-67.

STEVENSON, L. G., STRISOVSKY, K., CLEMMER, K. M., BHATT, S., FREEMAN, M. & RATHER, P. N. 2007. Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci U S A,* 104**,** 1003-8.

STRISOVSKY, K., SHARPE, H. J. & FREEMAN, M. 2009. Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell,* 36**,** 1048-59.

SUGIKI, T., KOBAYASHI, N. & FUJIWARA, T. 2017. Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. *Comput Struct Biotechnol J,* 15**,** 328-339.

TANG, B., LI, X., MARETZKY, T., PEREZ-AGUILAR, J. M., MCILWAIN, D., XIE, Y., ZHENG, Y., MAK, T. W., WEINSTEIN, H. & BLOBEL, C. P. 2020. Substrate-selective protein ectodomain shedding by ADAM17 and iRhom2 depends on their juxtamembrane and transmembrane domains. *FASEB J,* 34**,** 4956-4969.

TASCÓN, I., SOUSA, J. S., COREY, R. A., MILLS, D. J., GRIWATZ, D., AUMÜLLER, N., MIKUSEVIC, V., STANSFELD, P. J., VONCK, J. & HÄNELT, I. 2020. Structural basis of proton-coupled potassium transport in the KUP family. *Nature communications,* 11**,** 626.

THEßELING, A., RASMUSSEN, T., BURSCHEL, S., WOHLWEND, D., KÄGI, J., MÜLLER, R., BÖTTCHER, B. & FRIEDRICH, T. 2019. Homologous bd oxidases share the same architecture but differ in mechanism. *Nature communications,* 10**,** 5138.

THIAGALINGAM, S., CHENG, K. H., LEE, H. J., MINEVA, N., THIAGALINGAM, A. & PONTE, J. F. 2003. Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann N Y Acad Sci,* 983**,** 84-100.

THOMPSON, R. F., WALKER, M., SIEBERT, C. A., MUENCH, S. P. & RANSON, N. A. 2016. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods,* 100**,** 3-15.

TICHA, A., COLLIS, B. & STRISOVSKY, K. 2018. The Rhomboid Superfamily: Structural Mechanisms and Chemical Biology Opportunities. *Trends Biochem Sci,* 43**,** 726-739.

TROSHIN, P. V., PROCTER, J. B. & BARTON, G. J. 2011. Java bioinformatics analysis web services for multiple sequence alignment--JABAWS:MSA. *Bioinformatics,* 27**,** 2001-2.

UCHAŃSKI, T., MASIULIS, S., FISCHER, B., KALICHUK, V., WOHLKÖNIG, A., ZÖGG, T., REMAUT, H., VRANKEN, W., ARICESCU, A. R., PARDON, E. & STEYAERT, J. 2019. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *bioRxiv***,** 812230.

UCHAŃSKI, T., PARDON, E. & STEYAERT, J. 2020. Nanobodies to study protein conformational states. *Curr Opin Struct Biol,* 60**,** 117-123.

URBAN, S. & BAKER, R. P. 2008. In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol Chem,* 389**,** 1107-15.
URBAN, S. & FREEMAN, M. 2003. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell,* 11**,** 1425-34.
URBAN, S., LEE, J. R. & FREEMAN, M. 2001. Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell,* 107**,** 173-82.
VINOTHKUMAR, K. R. 2011. Structure of rhomboid protease in a lipid environment. *J Mol Biol,* 407**,** 232-47.
WALTER, J. D., SAWICKA, M. & DUTZLER, R. 2019. Cryo-EM structures and functional characterization of murine Slc26a9 reveal mechanism of uncoupled chloride transport. *Elife,* 8.
WANG, J., HU, M., WU, S., QI, J., WANG, G., HAN, Z., QI, Y., GAO, N., WANG, H. W., ZHOU, J. M. & CHAI, J. 2019. Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science,* 364.
WANG, L., HOU, W. T., CHEN, L., JIANG, Y. L., XU, D., SUN, L., ZHOU, C. Z. & CHEN, Y. 2020a. Cryo-EM structure of human bile salts exporter ABCB11. *Cell Res,* 30**,** 623-625.
WANG, Y. & HA, Y. 2007. Open-cap conformation of intramembrane protease GlpG. *Proc Natl Acad Sci U S A,* 104**,** 2098-102.
WANG, Y., MAEGAWA, S., AKIYAMA, Y. & HA, Y. 2007. The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J Mol Biol,* 374**,** 1104-13.
WANG, Y., ZHANG, Y. & HA, Y. 2006. Crystal structure of a rhomboid family intramembrane protease. *Nature,* 444**,** 179-80.
WANG, Z., HU, W. & ZHENG, H. 2020b. Pathogenic siderophore ABC importer YbtPQ adopts a surprising fold of exporter. *Sci Adv,* 6**,** eaay7997.
WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M., CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* 25**,** 1189-91.
WEI, Y., THYPARAMBIL, A. A. & LATOUR, R. A. 2014. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim Biophys Acta,* 1844**,** 2331-7.
WESKAMP, G., TÜSHAUS, J., LI, D., FEEDERLE, R., MARETZKY, T., SWENDEMANN, S., FALCK-PEDERSEN, E., MCILWAIN, D. R., MAK, T. W., SALMON, J. E., LICHTENTHALER, S. F. & BLOBEL, C. P. 2020. ADAM17 stabilizes its interacting partner inactive Rhomboid 2 (iRhom2) but not inactive Rhomboid 1 (iRhom1). *J Biol Chem,* 295**,** 4350-4358.
WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67**,** 235-42.
WOJNOWSKA, M., GAULT, J., YONG, S. C., ROBINSON, C. V. & BERKS, B. C. 2018. Precursor-Receptor Interactions in the Twin Arginine Protein Transport Pathway Probed with a New Receptor Complex Preparation. *Biochemistry,* 57**,** 1663-1671.
WOLF, E. V. & VERHELST, S. H. 2016. Inhibitors of rhomboid proteases. *Biochimie,* 122**,** 38-47.
WRIGHT, N. J. & LEE, S. Y. 2019. Structures of human ENT1 in complex with adenosine reuptake inhibitors. *Nat Struct Mol Biol,* 26**,** 599-606.
WU, M. & LANDER, G. C. 2020. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Curr Opin Struct Biol,* 64**,** 9-16.
WU, M., LANDER, G. C. & HERZIK, M. A., JR. 2020a. Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *J Struct Biol X,* 4**,** 100020.

WU, X., SIGGEL, M., OVCHINNIKOV, S., MI, W., SVETLOV, V., NUDLER, E., LIAO, M., HUMMER, G. & RAPOPORT, T. A. 2020b. Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. *Science,* 368.

WU, Z., YAN, N., FENG, L., OBERSTEIN, A., YAN, H., BAKER, R. P., GU, L., JEFFREY, P. D., URBAN, S. & SHI, Y. 2006. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol,* 13**,** 1084-91.

XU, D., FENG, Z., HOU, W. T., JIANG, Y. L., WANG, L., SUN, L., ZHOU, C. Z. & CHEN, Y. 2019. Cryo-EM structure of human lysosomal cobalamin exporter ABCD4. *Cell Res,* 29**,** 1039-1041.

XUE, Y. & HA, Y. 2013. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J Biol Chem,* 288**,** 16645-54.

YAN, R., ZHAO, X., LEI, J. & ZHOU, Q. 2019. Structure of the human LAT1-4F2hc heteromeric amino acid transporter complex. *Nature,* 568**,** 127-130.

YANG, L., LI, W., LIU, B., WANG, S., ZENG, L., ZHANG, C. & LI, Y. 2016. iRhom2 (Uncv) mutation blocks bulge stem cells assuming the fate of hair follicle. *Arch Dermatol Res,* 308**,** 503-10.

YANG, Z. R., THOMSON, R., MCNEIL, P. & ESNOUF, R. M. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics,* 21**,** 3369-76.

YATES, A. D., ACHUTHAN, P., AKANNI, W., ALLEN, J., ALLEN, J., ALVAREZ-JARRETA, J., AMODE, M. R., ARMEAN, I. M., AZOV, A. G., BENNETT, R., BHAI, J., BILLIS, K., BODDU, S., MARUGÁN, J. C., CUMMINS, C., DAVIDSON, C., DODIYA, K., FATIMA, R., GALL, A., GIRON, C. G., GIL, L., GREGO, T., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., KAY, M., LAVIDAS, I., LE, T., LEMOS, D., MARTINEZ, J. G., MAUREL, T., MCDOWALL, M., MCMAHON, A., MOHANAN, S., MOORE, B., NUHN, M., OHEH, D. N., PARKER, A., PARTON, A., PATRICIO, M., SAKTHIVEL, M. P., ABDUL SALAM, A. I., SCHMITT, B. M., SCHUILENBURG, H., SHEPPARD, D., SYCHEVA, M., SZUBA, M., TAYLOR, K., THORMANN, A., THREADGOLD, G., VULLO, A., WALTS, B., WINTERBOTTOM, A., ZADISSA, A., CHAKIACHVILI, M., FLINT, B., FRANKISH, A., HUNT, S. E., IISLEY, G., KOSTADIMA, M., LANGRIDGE, N., LOVELAND, J. E., MARTIN, F. J., MORALES, J., MUDGE, J. M., MUFFATO, M., PERRY, E., RUFFIER, M., TREVANION, S. J., CUNNINGHAM, F., HOWE, K. L., ZERBINO, D. R. & FLICEK, P. 2019. Ensembl 2020. *Nucleic Acids Research,* 48**,** D682-D688.

YEATES, T. O., AGDANOWSKI, M. P. & LIU, Y. 2020. Development of imaging scaffolds for cryo-electron microscopy. *Curr Opin Struct Biol,* 60**,** 142-149.

YIP, K. M., FISCHER, N., PAKNIA, E., CHARI, A. & STARK, H. 2020. Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! *bioRxiv***,** 2020.05.21.106740.

ZETTL, M., ADRAIN, C., STRISOVSKY, K., LASTUN, V. & FREEMAN, M. 2011. Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. *Cell,* 145**,** 79-91.

ZHANG, C., KONERMANN, S., BRIDEAU, N. J., LOTFY, P., WU, X., NOVICK, S. J., STRUTZENBERG, T., GRIFFIN, P. R., HSU, P. D. & LYUMKIS, D. 2018a. Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. *Cell,* 175**,** 212-223 e17.

ZHANG, K., LI, S., KAPPEL, K., PINTILIE, G., SU, Z., MOU, T. C., SCHMID, M. F., DAS, R. & CHIU, W. 2019. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature communications,* 10**,** 5511.

ZHANG, Y., BULKLEY, D. P., XIN, Y., ROBERTS, K. J., ASARNOW, D. E., SHARMA, A., MYERS, B. R., CHO, W., CHENG, Y. & BEACHY, P. A. 2018b. Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched. *Cell,* 175**,** 1352-1364 e14.

ZIVANOV, J., NAKANE, T., FORSBERG, B. O., KIMANIUS, D., HAGEN, W. J., LINDAHL, E. & SCHERES, S. H. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife,* 7.

ZIVANOV, J., NAKANE, T. & SCHERES, S. H. W. 2020. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ,* 7**,** 253-267.
ZOCHER, M., ZHANG, C., RASMUSSEN, S. G., KOBILKA, B. K. & MÜLLER, D. J. 2012. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor. *Proc Natl Acad Sci U S A,* 109**,** E3463-72.
ZOLL, S., STANCHEV, S., BEGAN, J., SKERLE, J., LEPSIK, M., PECLINOVSKA, L., MAJER, P. & STRISOVSKY, K. 2014. Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J,* 33**,** 2408-21.
ZOU, H., THOMAS, S. M., YAN, Z. W., GRANDIS, J. R., VOGT, A. & LI, L. Y. 2009. Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells. *FASEB J,* 23**,** 425-32.
ZUNKE, F. & ROSE-JOHN, S. 2017. The shedding protease ADAM17: Physiology and pathophysiology. *Biochim Biophys Acta,* 1864**,** 2059-2070.