STRATHCLYDE INSTITUTE OF PHARMACY & BIOMEDICAL SCIENCES

Decoding Protease Activated Receptor 4 (PAR4) Protein interaction Networks with Proteomics and Artificial Intelligence

A PhD thesis presented by:

Marco Bonfanti

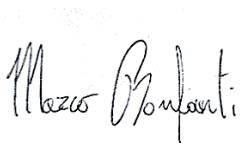
in fulfilment of the requirements for the degree of
Doctor of Philosophy

November 2024

Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow G4 0RE

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of degree.

The author has used ChatGPT 4o for assistance in formatting and data analysis purposes while developing the code for the proteomics analysis. All content is original and remains the intellectual property of the author.

Signed: 

Date: 22/11/2024

Technical abstract:

INTRODUCTION: Protease-activated receptor 4 (PAR4) is a G-protein coupled receptor (GPCR) activated by thrombin and cathepsin G, playing key roles in thrombosis and inflammation. As the least characterized member of the PAR family, elucidating PAR4's intracellular interactions and mutations, such as the Y157C mutation, is crucial for understanding its biological functions and interaction network. We hypothesised that PAR4 C-tail harbours a short linear motif (SLiM) resembling a class 1 PDZ binding sequence. To test this idea an array of technologies including SILAC proteomics, AI tools and super resolution microscopy was employed.

METHODS: This study utilized stable isotope labelling of amino acids in cell culture (SILAC) to analyse the proteome of HEK293 cells expressing wild-type PAR4-YFP, a PAR4 variant with a modified short linear motif (PAR4ΔSLiM-YFP), and a mutant PAR4-Y157C. Following GFP-trap affinity purification and LC MS/MS analysis, network analysis and gene ontology enrichment were conducted. Interactions with PDZ domain-containing proteins were predicted and validated computationally using AlphaFold Multimer. Additionally, super-resolution microscopy (STED and SIM) was employed to investigate the mitochondrial localization of PAR4.

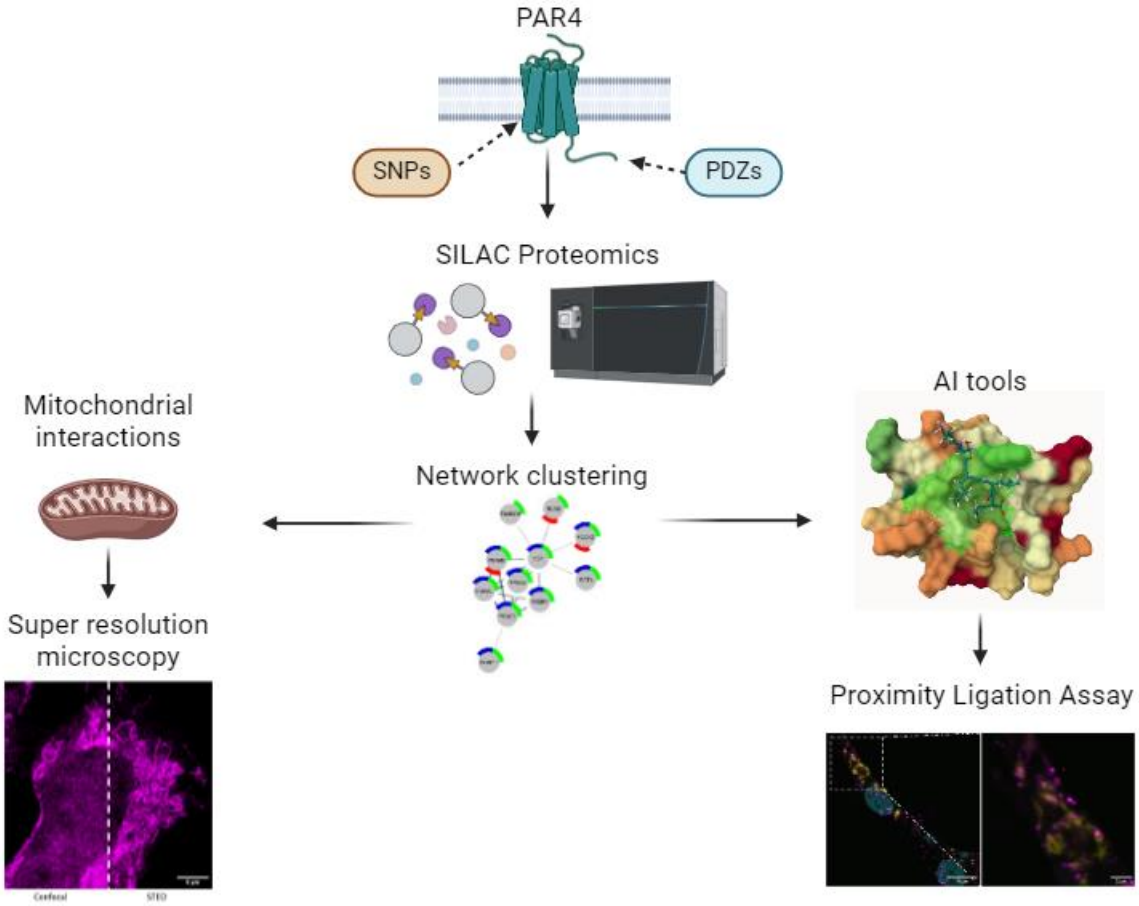
RESULTS: Across four datasets, 1056 unique proteins exceeded a 95% confidence threshold, with significant clusters of mitochondrial and ribosomal proteins highlighted. AlphaFold Multimer confirmed interactions between PAR4 variants and GIPC proteins, indicating specific PDZ domain involvements. Predicting protein-protein interactions (PrePPI) suggested five more interactions, which were also validated using AlphaFold multimer. Super-resolution microscopy provided insights into the subcellular localization of PAR4, particularly its association with mitochondria, though these results suggest further *in vitro* investigation.

CONCLUSION: This thesis significantly advances our understanding of PAR4 by detailing its interactions with mitochondrial components and PDZ domain-containing proteins, such as GIPCs. The integration of proteomic analyses, AI-driven predictions, and advanced microscopy techniques elucidates PAR4's complex interaction networks and supports its potential involvement in mitochondrial functions. These findings not only refine our understanding of PAR4's functional landscape but also highlight the utility of combining omics technologies with AI and super-resolution microscopy to enhance the precision and depth of biological studies. This study provides a proof of concept for the interaction of PAR4 with members of the GIPC family. Future developments of this project would include validation of these findings in more physiologically relevant models.

Lay abstract:

Protease-activated receptor 4 (PAR4) is a cell surface protein implicated in blood clotting and inflammation, activated by enzymes such as thrombin and cathepsin G, and stands as the least understood member of its protein family. The project explored how PAR4 interacts with downstream proteins and the effects of structural changes like the Y157C mutation by utilizing a technique known as SILAC. The analysis revealed over a thousand proteins interacting with PAR4, including significant groups found within mitochondria. Advanced AI tools were then employed to predict these interactions, with AI tools like AlphaFold confirming likely contacts with a family of proteins known as GIPCs. Another AI tool, PrePPI, suggested additional potential interacting proteins, enriching the understanding of PAR4's interaction network. Super-resolution microscopy techniques were then used to observe PAR4's potential association with mitochondrial proteins. Although microscopy indicated an association of PAR4 with mitochondria, further *in vitro* studies are needed for confirmation. This comprehensive approach has not only advanced understanding of PAR4's functionality, but also highlighted how combining various scientific techniques can provide detailed insights into complex networks of protein interactions, laying a foundation for future research in more complex biological models to verify these interactions.

Visual abstract:



Acknowledgments:

My deepest gratitude goes to my supervisor, Dr. Margaret Cunningham for giving me the opportunity to work on this project, and for her continuous guidance and support, both personally and professionally. From the bottom of my heart, thank you, Margaret. Your enthusiasm and love for science have been a true inspiration since the start of my Undergraduate studies. Whenever I had any issue, being an experiment not working or just a day with low motivation, I knew I could simply come into your office, and you would always give me a reason to go on and work harder than ever before. Thank you for constantly believing in me, even though I have not been the best wet scientist, and I kept on finding every reason to not run western blots. Thank you for your help with the experiments, especially PLAs and co-IPs. Above everything else, thank you for helping me finding my vocation, if I now dream of being an academic, I owe it all to you.

A huge thank you to Professor Gwyn Gould and his whole lab. Thank you, guys, for letting me share your tissue culture suite and your consumables, I would have not made it without you. A special mention goes to Angéline Geiser and Peter Tinning for your help with STED and the theory of super resolution microscopy. Justin Greig, you deserve a thank you on your own. You have been the best desk mate I could have asked for. Thank you for every time you battered me at chess, for every “are you winning son?” that kept me going even in the hardest day. In you I have found a great friend, and I wish you all the best for the future.

A special thank you to everyone from the Cunningham’s group, Ben and Nick thank you guys for introducing me to the experiments at the beginning of the PhD. Thank you, Graeme, for your help with confocal, and for making a dream come true by taking me to the old firm. Thank you, Zainab, Ryan, Maheen and Kat, for sharing this adventure with me. Kirsty, you deserve a thank you on your own. You are a great scientist, and an even greater friend. Thank you for your help with western blots, for answering every silly question I may have had even in my final year, but above all thank you for your continuous support, I do not deserve a friend like you.

A big thank you to all the students I have supervised throughout the years, thank you for your help in generating data for my thesis, but especially thank you for making me realise that teaching is what I like to do and the career I want to pursue.

Thank you to the Strathclyde Ultimate Frisbee Dark Horses. Being the president in the second year of my PhD has been one of the hardest, yet rewarding, experiences of my life. Thanks to you I have learned so much about managing projects and people, and these are skills I will carry on in my professional career too. Thank you for the laughs and the memories, my heart will always be full of pride whenever I will look at that gold medal we won together.

Thank you to Murray, Nikos, Cristina, Dave, Kiera, Kaileen, Charlie, Manuel, Erica, Callum and all the friends I have made along the way during my time in Scotland. People make Glasgow, and you guys are the living proof of it. Thank you for all the memories, I am so looking forward to making more with you.

Rui and Leah, you two came into my life like an earthquake and shook it down from its core. If I am the man I am today I owe so much to yous. Thank you for every moment we shared, for every new experience I had the opportunity to try with you. For every laugh, for every hug and for every tear. I love you from the bottom of my heart.

Luquitas, ni un libro entero de agradecimientos sería suficiente para ti. Llamarte mi mejor amigo sería reduccionista. Eres, sin lugar a dudas, la persona más influyente en mi vida. Gracias por todo. Lo bueno y lo malo. Nuestra amistad es una montaña rusa y sé que, aunque ahora estemos lejos, siempre nos respaldaremos mutuamente. Te quiero muchísimo, lamento todo el dolor que te he causado, pero te agradezco por seguir siendo un amigo tan valioso. Sabes cuánto valoro tu consideración y, aunque choquemos con frecuencia, no podría imaginar mi vida sin ti ahora. Por todo, para siempre - Gracias, gracias, gracias.

Davide, mio capitano. Nonostante la distanza ho sempre saputo di averti al mio fianco, spalla a spalla. Grazie di essere stato il mio ultras personale, non solo in questi anni di dottorato, ma da che ne ho memoria. Grazie di condividere i momenti più belli, ma soprattutto i momenti più duri. Nessuno sa motivarmi come te, nessuno mai saprà dirmi due paroline come me le dici tu. DAI NOI!

Ale, compagno di avventure accademiche. Non ci sono parole per descrivere la stima che provo nei tuoi confronti, come uomo e come accademico. Ti reputo una delle persone più intelligenti che conosca, e sono felice di condividere in parte questo pezzo della mia vita con te. I McDonalds a mezzanotte o le birre in riva al lago a parlare di politica, pene d'amore o quantum communications sono e resteranno uno dei miei piccoli piaceri della vita. Sei un grande scienziato, e sono sicuro che avrai una brillante carriera. Ti auguro veramente il meglio, amico mio.

Grazie a tutti i miei amici in Italia, che nonostante la distanza sia in termini di spazio che di tempo, mi sono sempre stati vicini. Un grazie particolare a Michela, per ogni sushi e per ogni gelato. Sei il mio spazio sicuro per condividere ogni gossip e ogni problema lavorativo e non. Ti voglio un mondo di bene e sono emozionatissimo per il tuo prossimo futuro!

Tak til Rachele, mit univers. Tak fordi du viste mig, hvordan kærlighed føles. Tak for din tillid til mig, for hver eneste oplevelse, vi har delt sammen. Du har løftet mig til utrolige højder, og jeg vil aldrig glemme dem. Jeg er ked af al den smerte, jeg har påført dig. Jeg håber, du vil finde din lykke både følelsesmæssigt og professionelt, du fortjener kun det bedste.

Ely e Kikkuz, siete da sempre il mio esempio. Vi amo più di quanto ogni parola possa spiegare, siete la mia forza e il mio porto sicuro. Scusate se ogni tanto vi ho fatto preoccupare e se non sempre ho avuto la testa sulle spalle. Se guardo al mio futuro, seguo sempre la vostra scia e so di essere in buone mani. Siete ciò che voglio diventare da grande, grazie per i vostri consigli ed il vostro appoggio, non sarei dove sono ora senza di voi.

Grazie Mamma. Di esserti sempre fidata di me, fin da quando me ne sono andato di casa. So di averti fatto preoccupare tanto in questi anni e ti chiedo scusa. Grazie di avermi sempre dato tutta la libertà del mondo, di non avermi detto mai di no, ma di avermi sempre incoraggiato a spiegare le mie ali e volare sempre più in alto. Sei la mia roccia e la mia forza. Grazie di avermi insegnato il valore dell'amore e della pazienza.

Grazie Papà. Se da piccolo non mi avessi portato in montagna, probabilmente non mi piacerebbe stare nella natura. Se non mi avessi insegnato a giocare a scacchi, probabilmente non avrei questa grande passione ora. Se non mi avessi portato in giro per farmacie, non sarei diventato farmacologo. Così tante delle cose che sono le devo a te. Mi spiace non esserci stato in questi anni di inferno per te. Scusami se andandomene da giovane non ho fatto altro che acuire il tuo dolore. Non ti meriti tutta questa sofferenza. Ricordati che sei amato e ti porterò per sempre dentro al mio cuore.

Publications:

Intracellular Signalling and Mitochondrial Dynamics of G Protein-Coupled Receptors – submitted for writing competition at *Open Biology* Open Question competition, November 2024

“Reporting evidence in an accurate way – the effects of COVID-19 on the brain” – Research the Headlines, April 2022

https://researchtheheadlines.org/2022/04/20/reporting-evidence-in-an-accurate-way-the-effects-of-covid-19-on-the-brain/

“Psyched for psychedelics” – Pharmacology Matters, August 2021
https://www.bps.ac.uk/publishing/pharmacology-matters/august-2021/psyched-for-psychedelics

Communications and conferences:

Protein signalling in space and time, Copenhagen (DK), November 2023 – Poster: Proteomic characterisation of PAR4 interacting proteins.

Genome bioinformatics: from short- to long-read sequencing 2023, Cambridge, November 2023 – Poster: Proteomic characterisation of PAR4 interacting proteins.

19th World Congress of Basic & Clinical Pharmacology (WCP2023), Glasgow, July 2023 - Poster: Exploring protease signalling within the cell, a focus on mitochondrial PAR4.

LEAPS Meets Life Sciencesences Conference, Elba Island (Italy), May 2023 - Poster: Exploring protease signalling within the cell utilising super resolution microscopy.

Glasgow Life Sciencesence ECR Symposium 2023, Glasgow, February 2023 - Organiser

Early Career Cardiovascular Scientists’ Symposium 2022, Manchester, November 2022 - Poster: Exploring the structural basis of Protease activated receptor 4 (PAR4) protein interactions by proteomic analysis.

Cardiovascular and Respiratory oral communication, Pharmacology 2022, Liverpool, September 2022 - Co-chair

Pharmacology 2022, Liverpool, September 2022 - Poster: Exploring the structural basis of Protease activated receptor 4 (PAR4) protein interactions by proteomic analysis.

5th Doctoral School Multidisciplinary Symposium, University of Strathclyde, June 2022 - Reviewer

6th European Research Network on Signal Transduction, Online, March 2022 - Poster: Exploring the structural basis of Protease activated receptor 4 (PAR4) protein interactions by proteomic analysis.

Bursary and Funding:

Novo Nordisk Foundation Science Cluster – November 2023

British Pharmacological Society Student Travel Grant – July 2023

European Research Network on Signal Transduction Student Grant – May 2023

British Pharmacological Society Student Society Fund – February 2023

British Pharmacological Society Student Travel Grant – September 2022

University Student Excellence Award (SEA) Studentship – June 2021 to June 2024

Training:

Computer-Aided Drug Design: Types, Uses and Opportunities – Biochemical Society, February 2024

Genome bioinformatics: from short- to long-read sequencing 2023 – November 2023

AI and computational drug discovery – British Pharmacological Society, June 2023

R for Biochemists 201 – Biochemical Society, June 2023

Standing up for Science 2022 – Voice of Young Science, April 2022

Best Practices for the qPCR Workflow – Thermo Fisher Scientific, March 2022

R for Biochemists 101 – Biochemical Society, March 2022

How to write a thesis – Margaret Watson, August 2021

Fear of the blank Page – The Scientific Editing Company, June 2021

Python for Biochemists – Biochemical Society, June 2021

Teaching qualifications:

Postgraduate Certificate in Researcher Professional Development, May 2024

Associate Fellow of Higher Education Academy (AFHEA), May 2023

Teaching experience:

BM110 – Biological chemistry: Visualisation of DNA – November 2023

BM110 – R studio video lectures – November 2023

BM110 – Tissues and Organs – February 2024

BM214 – Introduction to Immunology: Immune cells involved in immunity to *M. tuberculosis* infection – October 2023

BM214 – Introduction to Pharmacology: Smooth Muscle Receptor Pharmacology – February 2023

BM214 - Introduction to Biochemistry: Determination of the free energy change (ΔG0') for the conversion of glucose-6-phosphate to fructose-6-phosphate – November 2022

BM214 – Introduction to Pharmacology: *Lumbriculus variegatus* toxicity assay. – February 2022, February 2023, January 2024

BM327 – Fundamental Pharmacology: Functional analysis of smooth muscle relaxant responses in rat upper respiratory tract – March 2023

BM327 – Fundamental Immunology: Complement-mediated immunity: antibody-dependent lysis of target cells – February 2023

BM327 – Fundamental Pharmacology: Nitric oxide as a biomarker for rheumatoid arthritis disease progression – February 2023

BM327 – Fundamental Pharmacology: NF-κB linked luciferase reporter assay – February 2023

BM327 – Fundamental Pharmacology: Effect of dexamethasone on pro-inflammatory cytokine release from monocytes – January 2023

BM327 – Fundamental Pharmacology: Functional analysis of pre-synaptic and post-synaptic in rat vas deferens – November 2022

BM327 – Fundamental Pharmacology: Calculation of the affinity of atropine at muscarinic receptors in guinea-pig ileum – November 2022, November 2023

BM327 – Fundamental Pharmacology: Computational tools for Measuring GPCR activity – October 2022, October 2023

BM434 – Advanced Pharmacology: Clinical scenario online workshop – February 2022

MP224 – Interpretation of lab indices, liver function – March 2024

MP942 - Advanced Techniques in Biomedical Research 1: Basic skills, lab calculations and handling bacteria, November 2023

MP942 - Advanced Techniques in Biomedical Research 1: Recombinant DNA technology - Plasmid preparation, restriction analysis and transformation – November 2022, November 2023

Students' supervision:

Ross Stevenson – MSci Final Thesis "Characterization of Novel Interactions Between Protease-Activated Receptor 4 and GAIP-Interacting Protein 1 (GIPC1)", 2023/24

Tianshu Qu – BSc (Hons) CPU student Final Thesis "Preliminary study on protease-activated receptor 4 localization on mitochondria and how it is influenced by certain short linear motif", 2022/23

Reece Thompson – Summer Project, 2022

Lisa Van Den Driest – BSc (Hons) Final Thesis **"**Proteomic analysis reveals a potential role for Protease-Activated Receptor 4 in mitochondrial-derived cellular metabolism", 2021/22

List of abbreviations:

5-HT – 5-hydroxytriptamine (serotonin)

5HTR – Serotonin receptor

A2a – Adenosine receptor

AASDHPPT - L-aminoadipate-semialdehyde-dehydrogenase-phosphopantetheinyl-transferase

ACP – Acyl carrier protein

ACTN1 – Alpha actinin-1

ADGRV1 - Adhesion G Protein-Coupled Receptor V1

ADP – Adenosine diphosphate

ADR – Adrenergic receptor

ALP - Actinin-associated LIM protein

AngR – Angiotensin receptor

APC – Activated protein C

APPL1 - Adaptor Protein, Phosphotyrosine Interaction, PH Domain and Leucine Zipper Containing 1

ARF3 - ADP-ribosylation factor 3

ASPH - Aspartyl/asparaginyl beta-hydroxylase

ADP – Adenosine diphosphate

ATP - Adenosine triphosphate

ATP2A2 - Probable phospholipid-transporting ATPase IIA

ATP5D - ATP synthase subunit d, mitochondrial

ATP6V1F - V-type proton ATPase subunit F

BIND – Biomolecular Interaction Network Database

BioGRID - Biological General Repository for Interaction Datasets

BPAE - Bovine Pulmonary Artery Endothelial

CASP – Critical assessment of Protein Structure

CBR – Cannabinoid receptor

CCT3 - T-complex protein 1 subunit gamma

CD – cluster of differentiation

CETN2 - Centrin-2

CFTR - Cystic Fibrosis Transmembrane Conductance Regulator

CH – Calponin homologue

CHO – Chinese hamster ovary cell

CIS2 - CDGSH iron-sulphur domain containing protein 2

CKB - Creatine kinase B-type

CKMTIB - Creatine kinase U-type

CLIC1 - Chloride intracellular channel protein 1

COA3 - Cytochrome oxidase assembly factor 3

Co-IP – Co-immunoprecipitation

COX - Cytochrome c oxidase

CPPED1 - Serine/threonine-protein phosphatase CPPED1

CPSF6 - Cleavage and Polyadenylation Specificity Factor 6

CRC – Concentration response curve

CS – Citrate synthase

CTP – Cytidine Triphosphate

CVDs – Cardiovascular diseases

DAG – Diacylglycerol

DAPI - 4',6-diamidino-2-phenylindole

DDAH2 - N(G), N(G)-dimethylarginine dimethylaminohydrolase 2

DDX28 - Probable ATP-dependent RNA helicase

DIABLO - Second mitochondrial-derived activator of caspase

DIP – Database of Interacting Proteins

DNA-PAINT - DNA-Point Accumulation for Imaging in Nanoscale Topography DNM2 – Dynamin-2

DR – Dopamine receptor

Drp1 - Dynamin-related protein 1

ECH1 - ΔΔ-dienoyl-CoA isomerase

ECL – Extracellular loop

ECSIT - Evolutionarily conserved signalling intermediate in toll pathway

EGFR - Epidermal Growth Factor Receptor

ELAVL1 - ELAV Like RNA Binding Protein 1

ENDOG - Endonuclease G

ENH – Enigma homologue

EPDR1 - Ependymin-related protein 1

ER – Endoplasmic reticulum

ERAD - Endoplasmic Reticulum-Associated Degradation

EXOC1 - Exocyst complex component 1

FDR – False Discovery Ratio

FGF1 - Fibroblast growth factor 1

FH – Fumarate hydratase

FHL1 - Four and a Half LIM domains 1

Fox – Forkhead box protein

FSCN1 – Fascin

GAIP – G Alpha interacting protein

GART - Trifunctional purine biosynthetic protein adenosine-3

GDP – Guanosine Diphosphate

GEF – Guanine nucleotide exchange factor

GFM2 - Ribosome-releasing factor 2

GFP – Green Fluorescent Protein

GH domain – GIPC homology domain

GI – Gastrointestinal (tract)

GIPC – GAIP interacting protein, C-terminus

GLUT – Glucose transporter

GNG12 - Guanine nucleotide-binding protein

GO – Gene Ontology

GPCR – G protein coupled receptor

GPU – Graphic Processing Unit

GSDMD – Gasdermin D

GSTP1 - Glutathione S-transferase P

GTP – Guanosine Triphosphate

GTEx- Genotype tissue expression (database)

HACD3 - Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3

HAD - Trifunctional enzyme

HAGH - Hydrocyacylglutathione hydrolase

HARS1 - Histidine-tRNA ligase

HDL – High-density lipoprotein

HEK – Human embryonic kidney cell

HELM - Harmonic excitation light microscopy

HMMs – Hidden Markov Models

HNRNP - Heterogeneous Nuclear Ribonucleoprotein

HPA – Human Protein Atlas

HPRD - Human Protein Reference Database

HSP - Heat shock protein

HtrA - High-temperature requirement

HuR – Human antigen R (ELAV-like protein 1)

ICL – Intracellular loop

IDH3B - Isocitrate dehydrogenase subunit beta

IDH3G - Isocitrate dehydrogenase subunit gamma

IDP – Intrinsically disordered protein

IDR – Intrinsically disordered region

IGF1R - Insulin-like growth factor 1 receptor

IL – Interleukin

ILKAP - Integrin-linked kinase-associated serine/threonine phosphatase 2C

INVENT - International Network Against Venous Thrombosis

IP_3 – Inositol triphosphate

IPO – Importin

ipTM – interface predicted Template Modelling score

KEGG - Kyoto Encyclopedia of Genes and Genomes

KRT19 - Keratin type 1 cytoskeletal 19

KRTCAP2 - Keratinocyte-associated protein 2

LDL – Low-density lipoprotein

LHCGR - Luteinizing hormone/choriogonadotropin receptor

LIM - Lin-11, Isl-1, and Mec-3

LIMK - LIM domain kinases

LMEM - Laterally modulated excitation microscopy

LMO7 – LIM domain only 7

LPA1 - Lysophosphatidic acid receptor 1

LPS - Lipopolysaccharide

LRP - Low density lipoprotein receptor-related protein

LRRC59 - Leucine-rich repeat-containing protein 59

MAF – Minor allele frequency

MAGOHB - Protein mago nashi homolog 2

MAGUK - Membrane-associated guanylate kinase homologs

MAPK – Mitogen activated protein kinase

MCL – Markov Cluster Algorithm

MDN1 – Midasin

MEMS - Micro-electro-mechanical system

MI – Myocardial infarction

MIB - Mitochondrial intermembrane space bridging

MICOS - Mitochondrial contact site and cristae organizing system

MINT - Molecular INTeraction database

MLC – Myosin light chain

MMP – Matrix metalloprotease

MRPL – Large mitochondrial ribosomal protein

MRPS - Small mitochondrial ribosomal protein

MSA – Multiple Sequence Alignment

MSC – Microbiological safety cabinet

MTR – Melatonin receptor

MTX – Metaxin

MYH10 - Myosin-10

MYO6 – Myosin 6

NA - Numerical aperture

NADH - Nicotinamide Adenine Dinucleotide (Hydrogen)

NAXE - NAD(P)H-hydrate epimerase

NDUFA - NADH Dehydrogenase (Ubiquinone) 1 Alpha Subcomplex

NEAA – Non-essential amino acid

NGDN – Neuroguidin

NMDA - N-methyl-D-aspartate

NNT - NAD(P) transhydrogenase

NO – Nitric oxide

NOLC1 - Nucleolar and coiled-body phosphoprotein 1

NOS – Nitric oxide synthase

NRP - Neuropilin

NSUN2 - NOP2/Sun RNA methyltransferase family member 2

OAT - Ornithine aminotransferase

OMIM - Online Mendelian Inheritance in Man

OXA1L - Mitochondrial inner membrane protein OXA1L

OXPHOS – Oxidative phosphorylation

PAE – Predicted Aligned Error

PAF – Platelet activating factor

pAKT – (phospho) Protein kinase B

PALLD – Palladin

PAR – Protease activated receptor

PARS2 - Probable proline-tRNA ligase

PCC – Pearson's Correlation Coefficient

PCNT – Pericentrin

PDB – Protein database

PDE12 - 2',5'-phospohdiesterase 12

PDGF – Platelet derived growth factor

PDHB - Pyruvate dehydrogenase E1 subunit beta

PDLIM1 - PDZ and LIM domain 1

PDZ - PSD-95, Dlg, and ZO-1

pERK – (phospho) Extracellular signal-regulated kinase

Pfam – Protein families

PG – Prostaglandin

PGD - 6-phosphogluconate dehydrogenase

PGRMC2 - Membrane-associated progesterone receptor component 2

PI3K - Phosphoinositide 3-kinase

PINK1 - Mitochondrial Serine/threonine-protein kinase

PIP_2 - Phosphatidylinositol 4,5-bisphosphate

PLA – Proximity ligation assay

PLC – Phospholipase C

pLDDT – predicted Local Distance Difference Test

PMT - Photomultipliers tubes

POLDIP2 - Polymerase delta-interacting protein 2

PPIs – Protein-protein interactions

PPP2CA - Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform

PPT1 - Palmitoyl-protein thioesterase 1

PRDX3 - Thioredoxin-dependent peroxide reductase

PRDX6 - Peroxiredoxin 6

PRKN – Parkin

PRP – Platelet rich plasma

PRPSAP1 - Phosphoribosyl pyrophosphate synthase-associated protein 1

PSF - Point spread function

PSMB2 - Proteasome subunit beta type-2

PTGR3 - Prostaglandin reductase 3

pTM – predicted Template Modelling score

PTPC - Permeability transition pore complex

PXDN - Peroxidasin homolog

PYCR2 - Pyrroline-5-carboxylate reductase 2

r.m.s.d. – root mean square deviation

RABL3 – Rab-like protein 3

RAM – Random Access Memory

RBD – RNA binding domain

RGS - Regulator of G-protein signalling

RIL - Reversion-induced LIM protein

RNP – Ribonucleoprotein domain

ROCK – Rho associated protein kinase

ROS – Reactive oxygen species

RPS – Small ribosomal subunit protein

RRM – RNA recognition motif

SACS – Sacsin

SAM - Sorting assembly machinery

SAMM50 - Sorting and Assembly Machinery Component

SDC4 – Syndecan 4

SEMA4C – Semaphorin 4C

SETD3 - Actin-histidine N-methyltransferase

SFXN1 - Sideroflexin-1

SGD - Saccharomyces Genome Database

SH domain - Src Homology domain

SILAC – Stable Isotope Labelling of Amino acids in Cell culture

SIM – Structured Illumination Microscopy

SKEMPI - Structural Kinetic and Energetic database of Mutant Protein

SLC – Solute carrier family

SLC25A3 - Solute carrier family 25 member 3

SLiM – Short Linear Motif

SMC – Smooth muscle cell

SMU1 - WD40 repeat-containing protein SMU1

SNP – Single nucleotide polymorphism

SR-B1 - Scavenger receptor class B type 1

SRP – Signal recognition particle

SRP-RNC - SRP-ribosome-nascent chain complex

STAT – Signal transducer and activator of transcription

STED – Stimulated emission depletion microscopy

STOML2 - Stomatin-like protein 2

STRING - Search Tool for the Retrieval of Interacting Genes/Proteins

STZ - Streptozotocin

TBCC - Tubulin-specific chaperone C

TCA – Tricarboxylic acid cycle

TGFβ – Transforming growth factor β

TGFP – Turbo GFP

TGs - Triglycerides

THEM6 - Mesenchymal stem cell protein DSCD75

TIA – Transient ischaemic attack

TJ – Tight junction

TJP2 – Tight junction protein 2

TM – Transmembrane

TM score – Template modelling score

TMX3 - Protein disulfide-isomerase TMX3

TNFα – Tumour necrosis factor α

TOM - Translocase of the outer mitochondrial membrane

TPBG – trophoblast glycoprotein

TPI1 - Triosephosphate isomerase

TPR – Tetratricopeptide repeat

TrkA - Tropomyosin receptor kinase A

TUBB2A - Tubulin beta-2A chain

TX – Thromboxane

TYRP1 - Tyrosinase-related protein 1

UFC1 - Ubiquitin-fold modifier-conjugating enzyme 1

UQC - Ubiqinol-cytochrome c reductases

UTP – Uridine Triphosphate

V1a – Vasopressin 1a

VANGL2 – Vang-like 2

VDAC - Voltage-dependent anion-selective channel protein

VLGR1 – Very large G-protein coupled receptor 1

VPS26A - Vacuolar protein sorting-associated protein 26A

VSM – Vascular smooth muscle

vWf – Von Willebrand factor

WHO - World Health Organisation

WT – Wild type

YFP – Yellow Fluorescent Protein

ZO – Zona occludens

Table of Contents

1. INTRODUCTION

1.1. General Introduction

In 2021, 20.5 million deaths were owed to cardiovascular diseases globally. This shows an increase of 18.7% when compared to data from 2010[1]. According to the World Health Organisation (WHO), stroke was the second greatest killer after ischaemic heart disease[1]. Platelets are a main constituent of arterial thrombi and antiplatelet drugs are a first-line medication for the prevention of myocardial infarction and stroke[2]. However, these medicines are not void of side effects and may cause serious complications such as haemorrhages. There is still therefore a need for improved therapeutic options for regulating pathological thrombus formation. Protease activated receptors (PARs) are crucial membrane proteins involved in activating platelets, and thus aiding the formation and propagation of thrombi due to their ability to respond to the procoagulant thrombin. PAR4, has become an attractive target for drug development due to issues associated with the clinical use of drugs developed that target the original family member, PAR1. Technological advancement in fields such as 'omics', artificial intelligence and computational biology, are providing novel understanding of protein target, their pathways, and better ways to design molecules[3].

In this project, a proteomic approach was taken to investigate if mutations at discrete regions of PAR4 affect its protein-protein interactions. Research on PAR4 has mainly focused on its involvement on platelets and thrombosis[4]. Therefore, follow up validation of the proteomic analysis, will lead to further characterisation to identify the involvement of PAR4 and its interacting protein networks in novel aspects of cellular function.

1.2. Thrombin - function and structure:

Haemostasis is a crucial physiological process which prevents pathological bleeding but can lead to thrombosis when exaggerated. At the site of vascular injury, many enzymes, zymogens, and proteases are released and triggered in what is known as the coagulation cascade[5]. Thrombin is a serine protease synthesised in the liver and derived by the enzymatic amplification network of the inactive zymogen prothrombin. The precursor is encoded by the gene F2 located on chromosome 11p11-q12. Prothrombin is a glycoprotein with a molecular weight of 72kDa and it is composed of four domains. After its cleavage by the action of factor Xa and Va, thrombin maintains only the serine protease domain, with a molecular weight of 36kDa. The structure of human thrombin was first solved by Bode and colleagues in 1989[6] and since then different models have been developed which aided understanding the structure and the function of the protease.

Thrombin has both procoagulant and anticoagulant properties, as it can also shut down the coagulation cascade[5]. When interacting with thrombomodulin, found on endothelial cells, thrombin switches its activity from procoagulant to anticoagulant by interacting with protein C[5]. Activated protein C (APC) is an inhibitor of the coagulation cascade, with thrombin-thrombomodulin interaction serving as negative feedback to prevent excessive coagulation[7]. Owed to this, thrombin is a major clinical target in thrombosis. Regulating its activity is of crucial importance due to the dual nature of pro- and anti-coagulant activity of the enzyme. Being such a critical pharmacological target, different strategies have been employed to regulate this protease. Firstly, limiting the availability of prothrombin can be an avenue to indirectly downregulate thrombin. This is done by warfarin which limits the functional synthesis of the

zymogen. Another relevant inhibitor of thrombin is heparin which may come in two forms. Low-molecular weight heparin is useful in attenuating thrombin generation, whereas unfractionated heparin can antagonise fully formed thrombin. Albeit these strategies are widely employed clinically, they do not come without side effects. The main offset symptoms are excessive bleeding and heparin-induced thrombocytopenia[8].

The catalytic site of the enzyme is found in a deep pocket surrounded by loops, named 60-insertion and γ-loops, which restrict access to this active region. This improves the specificity for substrates and inhibitors. The 60-insertion loop is essential for co-factoring of thrombin, highlighting how allostery plays a vital role in thrombin activity[9]. For the recognition of physiologic substrates and cofactors, thrombin like other proteases, utilise exosites[10]. These regions of enzymes are spatially separated from the catalytic site, but communicate allosterically[11,12]. Thrombin possesses two anionic-binding exosites, named I and II, which are found close to the active moiety of the enzyme. The roles of each exosite have been determined by extensive biochemical, mutagenic and structural studies[13]. Thrombin exosite I shows a similar position to trypsin's Ca^{2+} binding loop, found 20-25Å from the active moiety. Exosite I is a binding region for thrombomodulin[14,15], fibrinogen[16,17], and for the protease activated receptors, PAR1[18,19] and PAR3[20,21]. A natural thrombin inhibitor, hirudin, also binds to exosite I due to its extended acidic C-terminal domain[22]. PAR1 and PAR3 bind thrombin because of a sequence of amino acids in their extracellular N-terminal which resembles hirudin, and it is therefore named hirudin-like domain[23].

1.3. Protease activated receptors (PARs):

The proteolytic activity of thrombin is not exclusively involved in the coagulation cascade, but it allows this enzyme to induce signal transduction in a variety of tissues, by targeting protease activated receptors (PARs) on the extracellular surface of cells. For many years it has been hypothesised that proteases such as thrombin and trypsin could act as hormone-like molecules to elicit signalling in cells[24]. However, it was not possible to assess the affinity of the enzymes binding to cells when classical radioligand-binding assays were performed. These failures were owed to the transient activity of proteases on the extracellular domain PARs. To elicit signalling, proteases would enzymatically alter the structure of the receptors through proteolytic cleavage rather than just binding as other molecules do. Moreover, the enzymatic activity of proteases would allow for the activation of several receptors by the action of one single molecule[25].

The involvement of thrombin in activating platelets to stimulate plug formation was recognised well before the discovery of a specific thrombin receptor, but the mechanisms that triggered the activation were still unclear. In 1991 Vu and colleagues elucidated the signalling transduction pathway that activated platelets by cloning the first thrombin receptor, protease activated receptor 1 (PAR1). Thrombin triggers signal transduction through PARs in a unique way. These are G protein coupled receptors (GPCRs) activated by proteolytic cleavage of their amino terminus. Four different PARs have been identified (PAR1 – PAR4). Proteases such as thrombin cleave a portion of the receptor's extracellular N terminal. This reveals a peptide sequence specific for each receptor called tethered ligand. The tethered ligand then binds to a conserved peptide sequence in the second extracellular loop (ECL2) of the receptor (I^{251} TTCHDV for PAR1-3, C^{228} HD for PAR4)[26]. Binding of the tethered ligand

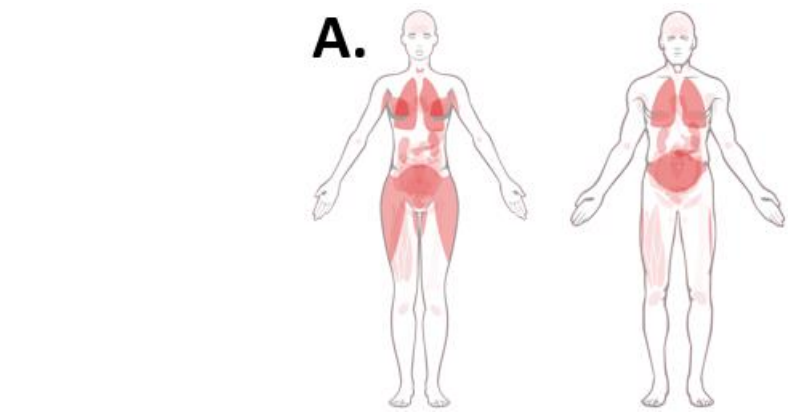
causes a conformational shape change in the receptor[27] which stimulates the guanine nucleotide exchange factor of the G protein's α-subunit, thus dissociating it from the βγ dimer and initiating downstream signalling. To mimic the activation of PARs, synthetic activating peptides specific for each receptor have been developed which bind directly to the receptor directly independent of cleavage[28].

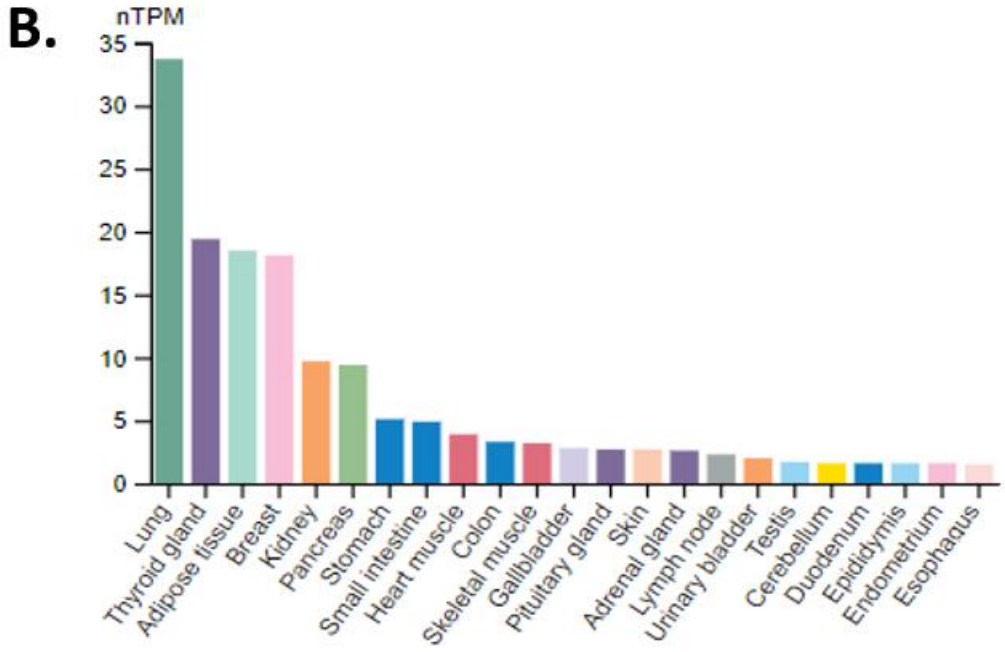
1.4. Other enzymes that activate PARs:

Thrombin is not the only enzyme activating PARs. Other proteases are able to cleave the receptors either at the same site, or at a different residue. This may cause biased signalling depending on the new tethered ligand exposed[29]. For instance, APC may cleave all four receptors, despite having lower catalytic efficiency compared to thrombin[30]. Matrix metalloproteases (MMPs) are another class of enzymes capable of activating PARs. MMP1, MMP2 and MMP13 in particular can cleave PAR1 at noncanonical sites, thus generating MMP-specific thethered ligands[31]. Other enzymes can activate PARs including protease 3, cathepsin G, plasmin, elastase, Factor VIIa, and Factor Xa[32]. Neutrophil-secreted cathepsin-G generates a novel functional tethered ligand when activating PAR4. The enzyme cleavage site was found at the Ser67-Arg68 bond, the tethered ligand thus created achieved greater platelet activation than thrombin-generated peptides. This aggregation was observed also in dog platelets, which have a conserved Ser-Arg binding site, however murine platelets were unresponsive to this peptide[33]. The ability of these additional enzymes to cleave PARs and regulate receptor activities adds to the additional complexity of understanding their physiological roles and designing drugs that effectively target the receptor family. The activity downstream of PARs and physiological impact will largely depend upon where the receptor is expressed and the presence of the relevant enzymes capable of cleaving the receptor. Cleavage specificity by different enzymes results in varied molecular and cellular responses. Enzymes such as plasmin, calpain and leukocyte elastase disable PAR1 by cleaving the COOH-terminal to the activation site, impeding the formation of a functional tethered ligand[34]. A variety of enzymes cleave PARs at distinct sites and still mediate signal transduction, however triggering pathways and causing cellular responses distinct from the canonical signalling. Zhao et al. extensively reviewed the array of enzymes causing biased signalling of each PAR receptor[35].

1.5. PAR4 expression

Despite being well-characterised in platelets, PAR4 is expressed in a variety of other cells and tissues[36]. Structural and functional studies are shedding new light on its implication in a variety of clinical contexts, not restricted to just thrombosis. Research on PAR4 mainly focused on its activity in platelets and its involvement in cardiovascular and inflammatory environments. However, when PAR4 was discovered, northern blot studies revealed high receptor expression in the lungs, pancreas, thyroid, testis and small intestine, while lower levels were found also in placental tissue, skeletal muscle, lymph nodes, adrenal gland, prostate, uterus and colon[26]. At the cellular level, the receptor is widely distributed in endothelial cells[37] and smooth muscle cells[38], but the highest prevalence of PAR4 is on platelets and leukocytes, where the receptor mediates platelet activation and inflammatory responses. Over time more evidence has been collected and an extensive list of all the tissues and cells where PAR4 is expressed is given in figure 1.1 according to the human protein atlas[39].


A.


B.
nTPM
35
30
25
20
15
10
5
0
Lung
Thyroid gland
Adipose tissue
Breast
Kidney
Pancreas
Stomach
Small intestine
Heart muscle
Colon
Skeletal muscle
Gallbladder
Pituitary gland
Skin
Adrenal gland
Lymph node
Urinary bladder
Testis
Cerebellum
Duodenum
Epididymis
Endometrium
Esophagus

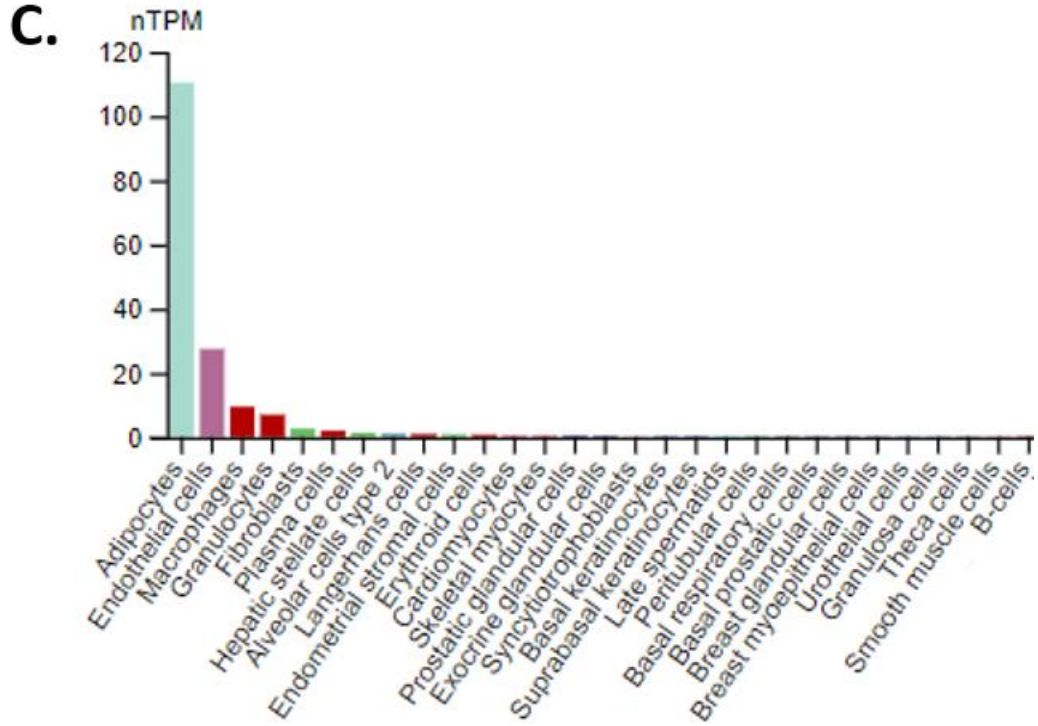


Figure 1.1 - Tissue and cellular distribution of F2RL3 RNA expression according to the consensus dataset. This dataset consists of normalized expression (nTPM) levels for 55 tissue types. Created by combining two datasets (HPA and GTEx). For both the HPA and GTEx transcriptomics datasets, the average TPM value of all individual samples for each human tissue or human cell type was used to estimate the gene expression level. To be able to combine the datasets into consensus transcript expression levels, a pipeline was set up to normalize the data for all samples. In brief, all TPM values per sample were scaled to a sum of 1 million TPM (denoted pTPM) to compensate for the non-coding transcripts that had been previously removed. Next, all TPM values of all samples within each data source (HPA + GTEx human tissues, HPA immune cell types, HPA cell lines) were normalized separately using Trimmed mean of M values (TMM) to allow for between-sample comparisons. The resulting normalized transcript expression values, denoted nTPM, were calculated for each gene in every sample. nTPM values below 0.1 are not visualized on the Atlas sections. **A**: diagram of the distribution of F2RL3 RNA in the human body. **B**: RNA tissue specificity. **C**: RNA single cell type specificity. Image and data available from proteinatlas.org[40].

1.6. Molecular structure of PAR4:

Although belonging to the same receptor family, PAR1 and PAR4 present substantial structural and pharmacological differences. Both receptors present the classical GPCR structure, being glycoproteins spanning through cellular membranes seven times. However, PAR4 shows little similarity with other PARs both at the amino and carboxy termini[26].

Artificial intelligence tools have become extremely important in predicting protein folding based on chemicals interactions. One of the most recent and powerful is Alphafold, a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm[41]. This allows to determine protein structure with unprecedented accuracy[42]. However, the reliability of AlphaFold2 has been questioned, since GPCR are flexible proteins that can change shape upon activation. AlphaFold predictions have been compared to GPCR structures solved empirically. Sometimes differences in structures were found, which impeded the use of predicted structures in functional assays[43]. When compared to other neuronal network-based software such as RoseTTAFold, AlphaFold shows better structure prediction, but with lower confidence[44].

To account for uncertainty, AlphaFold predictions are given a confidence metric called predicted local distance difference test (pLDDT) on a scale from 0 to 100, which is colour-coded on the protein structure. A very highly confident prediction (pLDDT>90) is shown in blue, confident (90 > pLDDT > 70) is depicted using light blue, a low confidence prediction (70 > pLDDT > 50) is highlighted in yellow and very low (pLDDT < 50) is orange. Figure 1.2 presents the Alphafold predicted structure of PAR4. Notably, the most flexible regions (carboxyl-terminal, amino-terminal and extracellular loop 3) also have the lowest confidence of prediction.

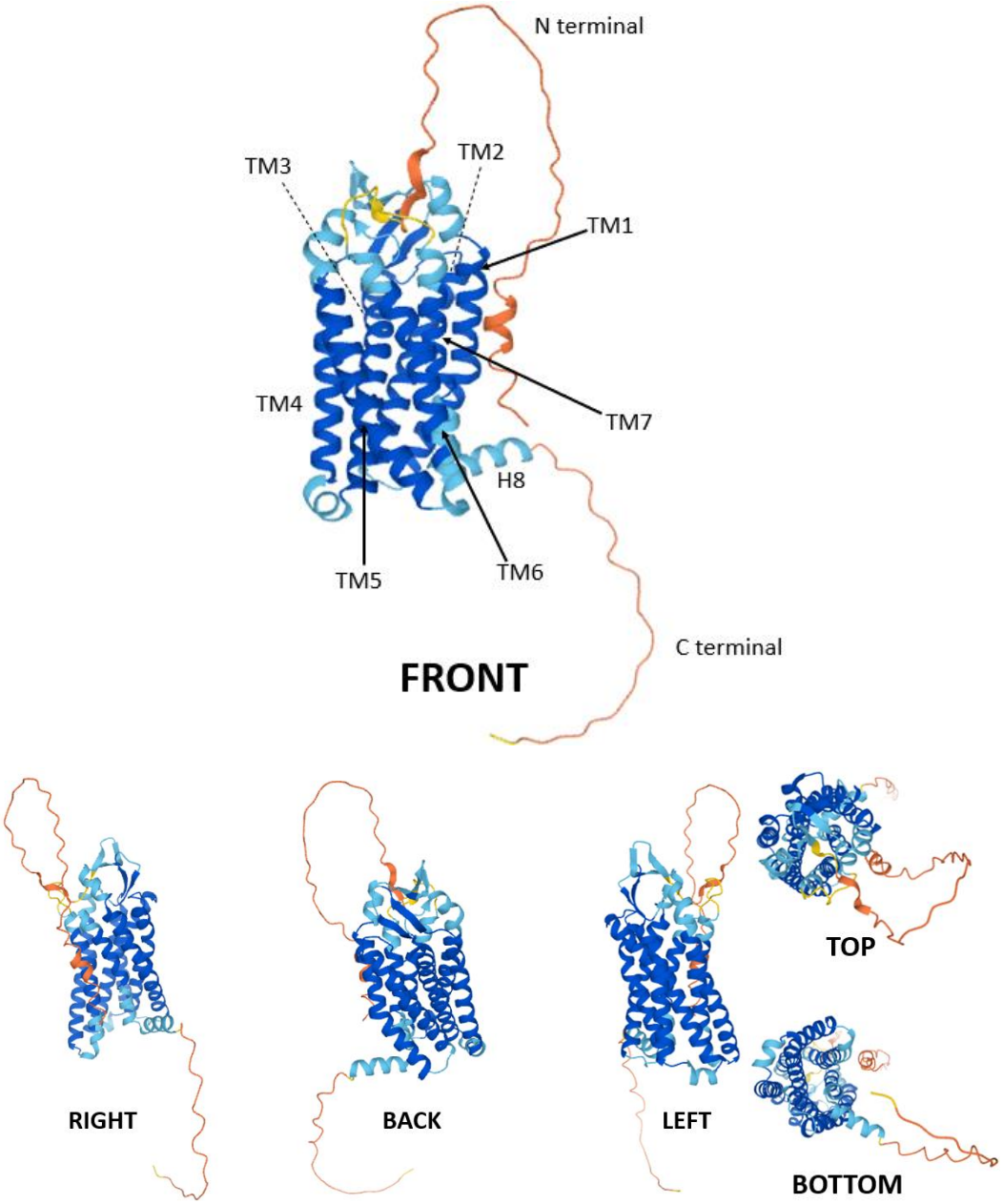


Figure 1.2 –AlphaFold prediction of PAR4 structure. The receptor view is shown from every plane. Different colours represent the prediction confidence. BLUE – Very high (pLDDT > 90), LIGHT BLUE – Confident (90 > pLDDT > 70), YELLOW – Low (70 > pLDDT > 50), ORANGE – Very low (pLDDT < 50). Structure retrieved from the AlphaFold protein structure database (https://alphafold.ebi.ac.uk/entry/Q96RI0).

The thrombin-PAR4 cleavage site (Arg47/Gly48) resides closer to the membrane compared to PAR1[45], which makes it more difficult for thrombin to reach, which explains the lower potency of the protease when compared to its activity on PAR1[45]. Mutations at this site (Arg->Ala) made the receptor completely unresponsive to cleavage and thus activation[26]. Another major structural difference between PAR1 and PAR4 is the lack of a hirudin-like binding domain in the extracellular terminus of PAR4. This region is particularly important in promoting a rapid association of thrombin [46]. PAR1 is a high-affinity receptor for thrombin (EC50 0.2nM) and it is rapidly stimulated by the protease. Instead, PAR4 presents a series of residues forming an anionic cluster (Asp57 - Asp59 - Glu62 - Asp65) which interact with cationic residues of proteases, thus slowing the dissociation rate[47]. Therefore, PAR4 is activated at much higher thrombin concentrations (EC50 5nM), typically found in pathological settings[48]. This activation is transient and more sustained over time, due to the lower dissociation rates of the enzymes[49].

Due to the lack of a hirudin-like domain, PAR4 shows minor interactions with exosite 1 of thrombin. PAR4 constructs with added hirudin-like acidic domains failed to engage with thrombin's exosite 1[50], suggesting that the tertiary structure of PAR4 is directly responsible for its reduced affinity with thrombin. Gamma (γ)-thrombin (a mutated form of the protease lacking a functional exosite 1) can activate PAR4 with the same potency of α-thrombin[26]. Compared to PAR1, PAR4 activation is significantly less affected by mutations of thrombin's exosite 1[47]. Thrombin's exosite 2 also plays a role, as a recent study demonstrated how targeting this site with oligosaccharides, resulted in the blockade of PAR1-PAR4 heterodimers activation[51]. PAR4 interacts with thrombin mainly via two prolines at position 44 and 46, right upstream of the cleavage site[50]. These two amino acids, together with a Leucine at position 43 help shape a three-dimensional structure which aids thrombin cleavage activity on the receptor[45]. Leu43 interacts with Leu99, Ile174 and Trp215 from thrombin, thus helping the association between the thrombin and PAR4[45]. The interactions between the protease and the receptor are also facilitated by the presence of a cluster of electronegative residues in the ECL2 (Asp224 – Asp230-Asp235)[52]. The crystal structure of murine thrombin has been resolved in complex with murine fragments of PAR3 and PAR4[21], and it is reported in figure 1.3.

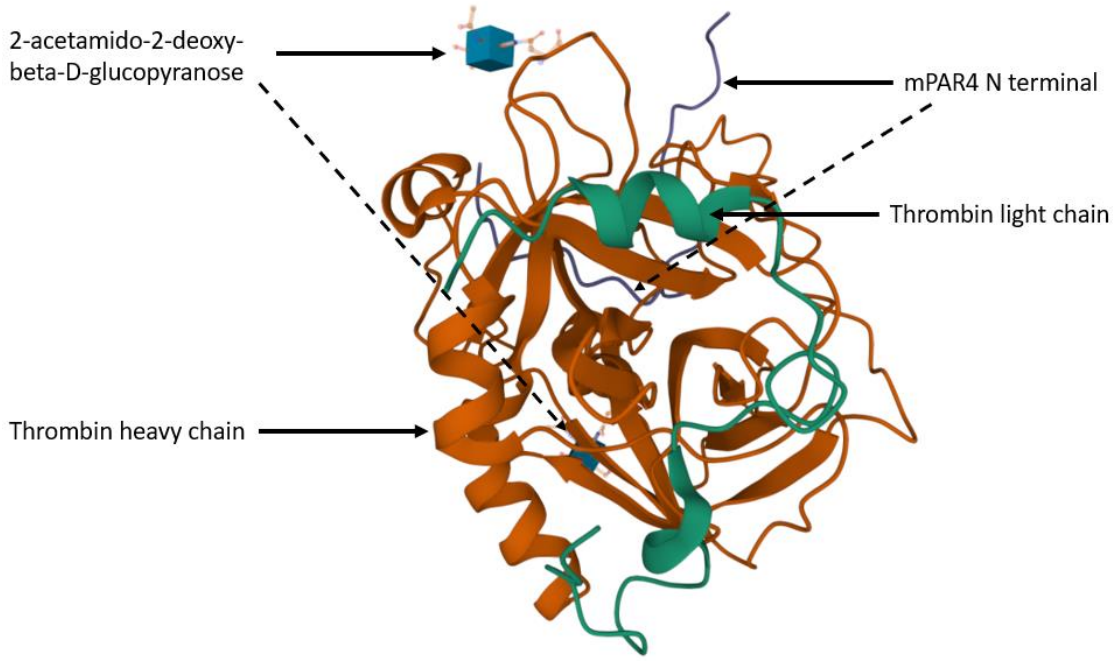


Figure 1.3 - Crystal structures of murine thrombin in complex with the extracellular fragments of murine protease-activated receptors PAR3 and PAR4. The structures reveal that when exosite I is accessible to bind cofactors, such as the cleaved version of PAR3, thrombin may activate PAR4. These cofactors function by modifying the conformation of the 60-loop through allosteric changes, which helps promote substrate diffusion into the active site. Structure retrieved from Research Collaboratory for Structural Bioinformatics Protein Database Bank (RCSB PDB - https://www.rcsb.org/structure/2pv9). This structure was achieved by X ray diffraction and added to the database by Bah and colleagues[21].

Homology modelling and *in silico* docking of PAR4 activating peptides has been useful to unveil the biochemical mechanism of the receptor activation. A series of amino acids were found to interact with the ligand with a high degree of predictability. The most frequently predicted interactions were with the residues His229, Asp230 and Gln242. These residues are highlighted in figure 1.4. Site-directed mutagenesis studies revealed how Asp230 is critical in receptor activation by either thrombin or PAR4 activating peptide[53] (PAR4-AP, AYPGKF-NH_2) and is also necessary to elicit β-arrestins recruitment[54]. Mice receptors possess a cysteine residue on the carboxyl terminal tail which is palmitoylated and it is targeted by allosteric modulators, whereas human PAR4 lacks any cysteine residue, making palmitoylation unlikely[55].

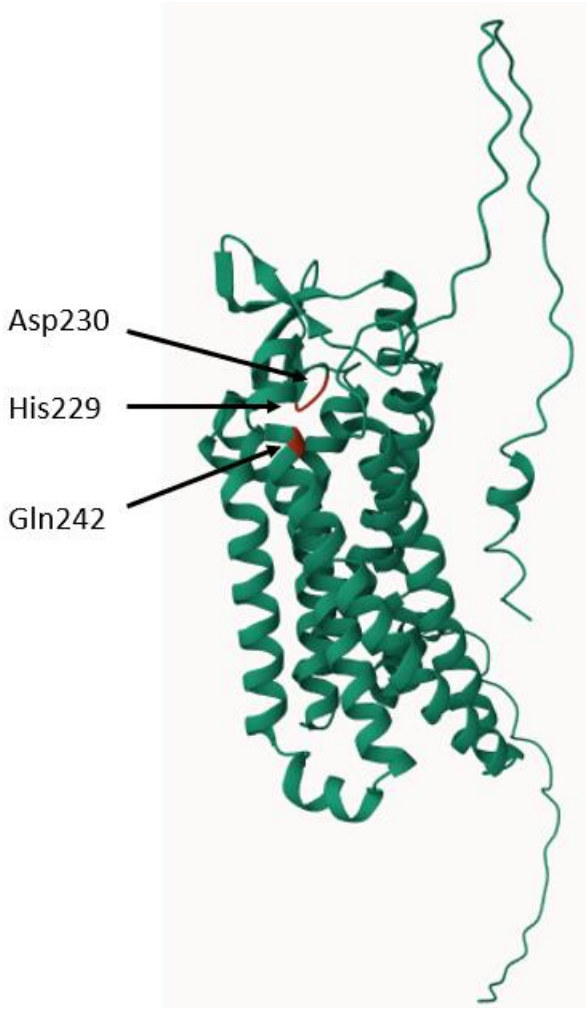


Figure 1.4 – Residues that most frequently interact with PAR4 activating peptide (AYPGKF-NH_2). The three most frequently interacting residues have been highlighted in red in this prediction of human PAR4. Visuals generated by retrieving data from Alphafold[41,42], and modifying amino acids colours in RCSB Protein Data Bank 3D viewer (https://www.rcsb.org/3d-view)[54,56].

1.7. Single nucleotide polymorphisms on PAR4:

Like all proteins, PAR4 is subject to genetic mutations which may alter the function of the receptor. Naturally occurring genetic variants may also help uncover the structural features of the receptor, as well as aid the development of new therapeutics.

To determine the impact of a single nucleotide polymorphism (SNP) on a protein, two distinct metrics have been developed. PolyPhen (Polymorphism Phenotyping) and SIFT (Sorting Intolerant From Tolerant) are two widely used bioinformatics tools for predicting the functional impact of genetic variations on protein structure and function[57,58]. Both methods aim to classify SNPs as either deleterious or benign based on their potential effects on protein structure and function. PolyPhen and SIFT are similar in that they both use computational algorithms to predict the impact of genetic variants on protein function. However, they differ in the specific features they consider and the way they assign scores.

PolyPhen is a computational tool that uses a combination of sequence-based and structure-based features to predict the functional impact of amino acid substitutions caused by SNPs. It assigns a PolyPhen score to each variant, indicating the likelihood of the variant being deleterious. The score ranges from 0 to 1, with higher scores indicating a higher probability of the variant being damaging. PolyPhen uses several features to make its predictions, including evolutionary conservation, physicochemical properties of the substituted amino acids, and structural information when available. It compares the properties of the wild-type (original) and mutant (variant) amino acids and calculates a score based on the observed differences. This score is then compared to a predefined threshold to determine the potential impact of the variant[59].

SIFT is another computational tool used for predicting the functional consequences of amino acid substitutions caused by SNPs. It focuses primarily on evolutionary conservation as a predictor of protein function. SIFT assigns a SIFT score to each variant, indicating the likelihood of the variant being tolerated or damaging. The score ranges from 0 to 1, with lower scores indicating a higher probability of the variant being deleterious. SIFT utilizes sequence homology across different species to assess the conservation of amino acids at specific positions in a protein. It calculates a score based on the degree of conservation at the substituted amino acid position. If the amino acid is highly conserved across species, it suggests that any change to that residue may have a functional impact[60]. While SIFT and PolyPhen can assist in prioritizing alterations that potentially result in protein function loss, it is still important to test experimentally the pathogenicity of the variants[61]. Another important parameter used to quantify the prevalence of a mutation in the human population is the minor allele frequency (MAF), which indicates how often the SNP can be found in a patient. A schematic of all the SNPs found in PAR4 according to the online database GPCRdb is given in figure 1.5.

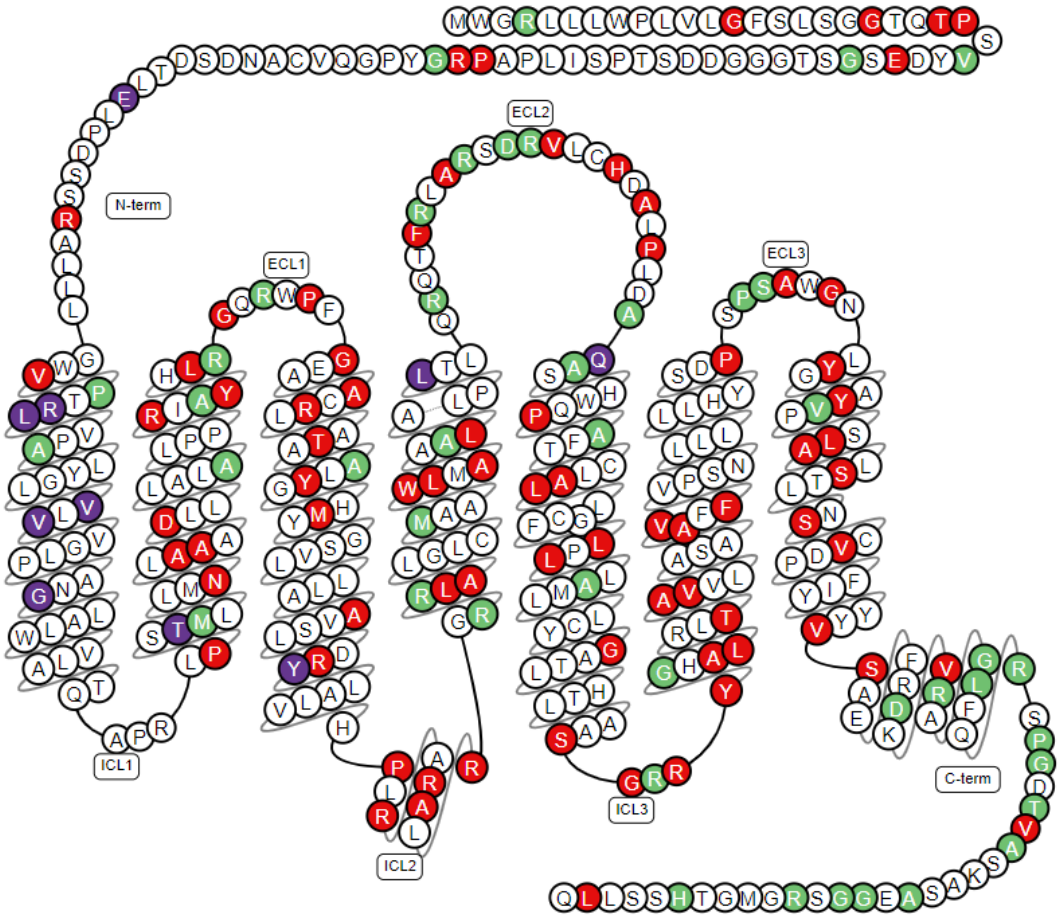


Figure 1.5 – PAR4 SNPs according to GPCRdb. Variants are colour coded based on the impact they have on PAR4. Green indicates missense mutations that are tolerated, red shows missense mutations that are annotated as deleterious under SIFT/PolyPhen. Purple shows either frameshift or stop gained mutations resulting in deleterious outcomes.

On the Ensembl database, 2734 variants are annotated; however, only four variants are reported in literature. These are shown in table 1.1 and are highlighted in figure 1.6 with different colours. They all happen in different regions of the receptor and affect PAR4 biochemical structure as well as its physiological responses in distinct ways.

Table 1.1 – SNPs that give rise to PAR4 variants reported in literature.

Variant	AA	MAF	SIFT	PolyPhen	Predicted outcome	Ref.
rs773902	A120T	0.342	0.24	0.35	Tolerated, Benign	[62]
rs2227376	P310L	0.015	0.57	0.838	Tolerated, Possibly damaging	[32]
rs767177635	Y157C	10^{-4}	0	1	Deleterious, Probably damaging	[63]
rs2227346	F296V	$3.95x10^{-3}$	0	0.626	Deleterious, Possibly damaging	[62]

Substitution of the Alanine with Threonine at position 120 (A120T) is the most common SNP, which causes hyperresponsive PAR4, leading to increased platelet activity. According to the 1000 genome project[64], this mutation is 61% more prevalent in patients with African ancestry, as opposed to 29% American, 24% eastern Asian, 26% southern Asian, and 21% European. This highlights the importance of pharmacogenomics and precision medicine, which aim to prescribe the right medicine, at the right dose, to the right patients. Platelets carrying this mutation are more susceptible to low doses of thrombin thus increasing the chances of thrombi formation. This was confirmed in *ex vivo* experiments. A120T polymorphism causes PAR4 to be more resistant to desensitization and platelets are more resistant to antithrombotic treatment, specifically clopidogrel, targeting P2Y12 receptor. PAR4-A120T is associated with increased risk of stroke, but not associated with venous thromboembolism in the International Network Against Venous Thrombosis (INVENT) database. This residue is in proximity of the tethered ligand binding site. Mutations of alanine to threonine may enhance binding of the tethered ligand, however this is yet to be tested[65]. The novel anti-thrombotic BMS-986120 has been tested considering this mutation, however no differences in calcium mobilisation assays were observed. The minimal differences in response were attributed to small differences in receptor expression[66]. A recent study found a strong association, but not causation, between this mutation and increased risk of placental vascular pathology and preterm birth[67]. Moreover, a murine model stably expressing humanised PAR4 found that the minor allele led to worse stroke outcomes. This was mediated partly by enhanced platelet activity, as well as platelet-neutrophil interactions. The study also found that mice carrying the mutation were less responsive to ticagrelor with or without aspirin, a common drug combination administered in humans. The study also included clinical data and found the AA genotype was a risk factor for incident ischemic stroke leading to worse functional outcomes within the black population[68].

Mutation of proline into leucine at residue 310 (P310L) reduces reactivity to both activating peptide and thrombin. According to INVENT this mutation is associated with a 15% reduction in risk of VTE. ECL3 acts as a gateway for PAR4 activation, blocking

the binding site in the inactive state and sliding away upon thrombin activation, thus revealing the binding site. Proline is necessary to maintain structural stability. For this reason, 310L makes ECL3 more flexible, impacting PAR4 reactivity[69]. Through CRISPR/Cas9 technology, this mutation was introduced in the PAR4 gene of mice, causing the corresponding mutation P322L, since the murine receptor is slightly longer. Platelet activation *ex vivo* as well as *in vivo*, were reduced in response to thrombin and PAR4-AP, but not ADP or convulxin[70].

The threonine to cysteine mutation at position 157 (Y157C) reduces PAR4 reactivity to activating peptide and thrombin. PAR4-157C platelets are also less responsive to thrombin compared to PAR4-157Y when pre-treated with the PAR1 antagonist vorapaxar. The cysteine mutation causes aberrant anterograde receptor trafficking, leading to unaltered PAR4 expression but reduced receptor numbers at the membrane[63]. According to Ensembl database this is the most deleterious variant.

The phenylalanine to valine variant at residue 296 (F296V) is found in a crucial region of PAR4. This is a Na^+ pocket microswitch, and a rotation of this site is necessary for receptor activation. Valine mutation makes the receptor less reactive, and this impacts G_q downstream signalling by generating less IP_3. Due to the rarity of this polymorphism, it is difficult to assess the physiological response at cellular level, but platelets carrying the mutation are thought to aggregate less compared to PAR4-296F[62].

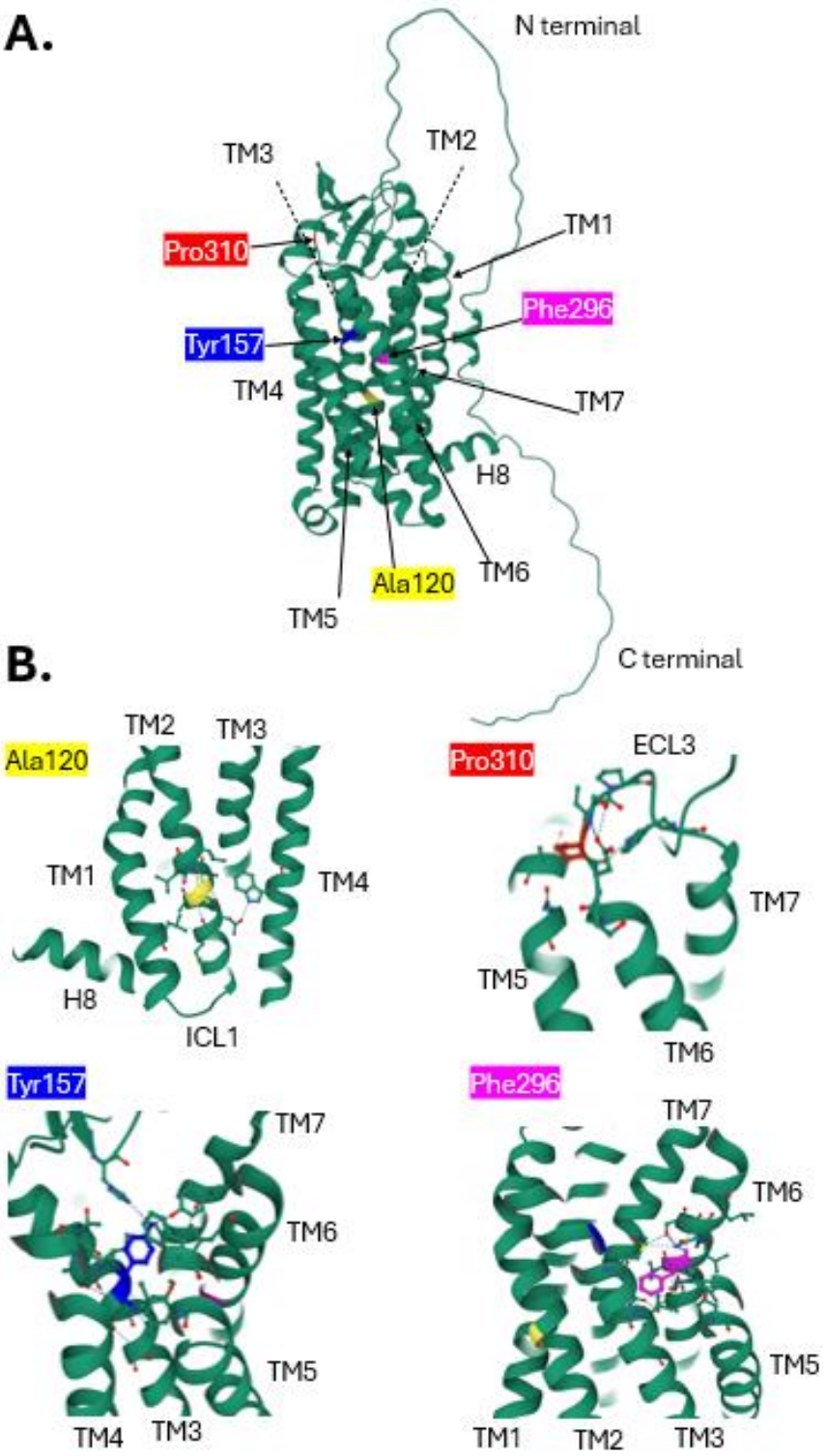


Figure 1.6 –Key variants presented within human PAR4 structure. A: Full receptor structure. B: Focus on each mutation. Overview of PAR4 and its deleterious variants: A120T - yellow, P310L - red, Y157C – blue, F296V - pink. Image generated by retrieving data from Alphafold[41,42], and modifying amino acids colours in RCSB Protein Data Bank 3D viewer (https://www.rcsb.org/3d-view)[54,56].

1.8. Signalling pathways involved in PAR activity:

Since thrombin was the first enzyme characterised with an action on PAR receptors[71], its activity was named “canonical signalling”. This refers to the cleavage of the N-terminal end of Arginine (R) residues on the extracellular end of the receptors (PAR1: R^{41}, PAR2: R^{36}, PAR3: R^{41}, PAR1: R^{47})[72], thus leading to the generation of canonical tethered ligands. However, the variety of proteases that can activate each PAR, with their unique biased signalling, demonstrated the involvement of the receptors in different biological functions. The canonical signalling of PAR1 causes G_q. protein to undergo a conformational change, which subsequentially triggers the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP), leading to its activation. Activated G_q then interacts with and activates the membrane-bound effector enzyme phospholipase C (PLC). Activated PLC cleaves phosphatidylinositol 4,5-bisphosphate (PIP_2) into two second messengers, inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). IP_3 then diffuses into the cytoplasm and binds to IP_3 receptors located on the endoplasmic reticulum (ER) membrane, causing them to release calcium ions (Ca^{2+}) from intracellular stores. DAG also interacts with downstream signalling molecules. The overall cellular effect of the canonical signalling of PAR1 are platelet activation and shape change[71]. Exposure of PAR1 to neutrophil-derived proteases such as elastase or protease-3 disarms the thrombin activated calcium signalling. Instead, it triggers G_i signalling which leads to regulation of endothelial cell barrier integrity[73]. This is just an example of the wide range of signallings which translate to different biological processes for each receptor. PARs transduce signalling through multiple intracellular mediators and are not restricted to one single biochemical pathway. This is owed to different factors: activating protease, different tethered ligands, tissues where the receptors are expressed. Another major factor that defines this promiscuity among receptors is their ability to couple to different G proteins. PAR1 couples to $G_{i/o}$, $G_{12/13}$ and $G_{q/11}$ subunits. PAR2 also initiates signalling through interactions with these subunits. PAR3 is the only receptor to bind only G_q since its main role is to act as a cofactor, rather than transduce signalling alone[74]. A comprehensive summary is reported in Table 1.2. PAR2 is the only receptor to not respond to thrombin, being activated by trypsin at concentrations as low as 300 pM instead[75], and it is mainly involved in inflammatory responses[76]. PAR3 is a second receptor for thrombin, but it does not transduce signalling per se. Instead, it is a cofactor that aids the binding of thrombin to PAR4, thus making the receptor respond to the protease at much lower concentrations. In murine platelets, PAR4 has been demonstrated to be activated by thrombin at 0.3 nM concentration. When PAR3 was co-expressed, this value dropped to 0.05 nM[77]. This is particularly relevant in murine models, as they lack PAR1[77]. PAR4 was the last receptor to be discovered[26]. It is also a receptor for thrombin, but due to structural differences it displays much lower affinity for the protease, and it is therefore often found dimerised with either PAR1 or PAR3[78]. Its canonical signalling involves the release of intracellular calcium stores by consequence of G_q activation[54]. PAR4 has mainly been studied in platelets, where it is reported to cause shape change and activation, albeit with different kinetics than PAR1[79].

Table 1.2 – Summary of PARs activation, signalling and function.

Receptor	Activating protease	Peptides	G protein	Non-G protein	Biological processes	Ref.
PAR1	Thrombin Activated protein C MMP1 MMP2 MMP13 Protease 3 Elastase Plasmin Cathepsin G Factor VIIa Factor Xa	SFLLR TFLLR	$G_{i/o}$ $G_{12/13}$ $G_{q/11}$	β-arrestin ERK1/2 MPAK Rac1 AKT p38	Platelet activation, Inflammation, Vasoconstriction, Profibrotic, Apoptosis, Cell proliferation, Angiogenesis, Cytoprotective, Proalgesic, analgesic	80–87
PAR2	Trypsin Tryptase Factor VIIa Factor Xa Elastase Cathepsin G	SLIGKV SLIGRL	$G_{12/13}$ G_q G_i	β-arrestin ERK1/2 MPAK Rac1 AKT Rho	Inflammation, Vasodilation, Hypersensitivity, Profibrotic, Apoptosis, Cell proliferation, Cell migration, Proalgesic	74,88–94
PAR3	Thrombin APC	Acts as a cofactor	G_q	ERK1/2	Blood coagulation, Platelet activation, PAR1 cofactor, PAR4 cofactor	74,93,94
PAR4	Thrombin Trypsin Cathepsin G Plasmin Factor Xa Gingipains-R Kallikrein 14	GYPGQV AYPGKF	$G_{12/13}$ G_q	NFκB p38 MAPK β-arrestin	Blood coagulation, Platelet activation, Apoptosis, Neutrophil recruitment, Proalgesic, Analgesic	95–98

1.9. Cellular signalling and receptor trafficking of PAR4:

PAR4 can couple to $G_{12/13}$ and G_q. The first subunit promotes the binding of guanine-nucleotide exchange factors that activate Rho (RhoGEFs), thus stimulating RhoA and Rho-associated protein kinase (ROCK). This kinase phosphorylates the myosin light chain (MLC) and inactivates MLC phosphatase. This results in cytoskeletal responses which vary depending on the cell type. In platelets activation of the $G_{12/13}$ pathway by PAR4 causing shape change [99], whereas in endothelial cells stimulation of this pathway increases vascular permeability[100]. PAR4 coupling to G_q leads to activation of PLCβ and phosphoinositide cleavage to inositol triphosphate (IP_3) and diacyl glycerol (DAG). Binding of IP_3 to its cognate ER receptor induces release of intracellular stores of calcium. Both these mechanisms are depicted in the schematic of figure 1.7. The increase in cytosolic calcium promotes the activation of several phosphatases and kinases including MAPKs, PKCs PLA_2 and calpain[99]. Two important hallmarks in the detection and quantification of G_q activation are phosphor-ERK (pERK) and phosphor-AKT (pAKT).

In platelets in particular, PAR4-dependent activation of phospholipase C (PLC) controls phosphorylation of AKT, independently of ADP activity on purinergic receptors or phosphatidylinositol-3-kinase (PI3K) activation. PAR4-dependent formation of pAKT is more sustained than PAR1 as the former receptor can stimulate the phosphorylation for up to 300 seconds, whereas the latter shows activity for a maximum of 60 seconds[101]. In platelets, activation of this pathways triggers integrin activation and secretion of storage granules, which finally result in platelet aggregation.

As previously stated, PAR4 has a slower activation kinetics (PAR1: 2.0±1.5 s – PAR4: 145±14 s), but leads to more sustained responses, as opposed to the transient activation of PAR1[79]. In platelets, the rapid desensitisation of PAR1 calcium-burst can be rescued by activation of PAR4. In smooth muscle and endothelial cells, prolonged signalling through G_q-coupled PAR4 stimulates transcriptional and metabolic responses which affect vascular remodelling, specifically in presence of high glucose levels[102].

Interestingly PAR4 acts as a receptor not only for proteases which reveal the tethered ligand, but also to peptide found endogenously in the human body. It is the case of the complement protein C4, which colocalizes with PAR1 and PAR4 in CHO-K1 cells expressing PAR1 or PAR4. Moreover, C4a induces ERK1/2 activation in human endothelial cells through Gai-independent signalling. The peptide triggers calcium mobilization through a $G_{\alpha q}$- coupled PLCβ-dependent signalling pathway and it increases endothelial permeability through the PAR1 signalling pathway[103].

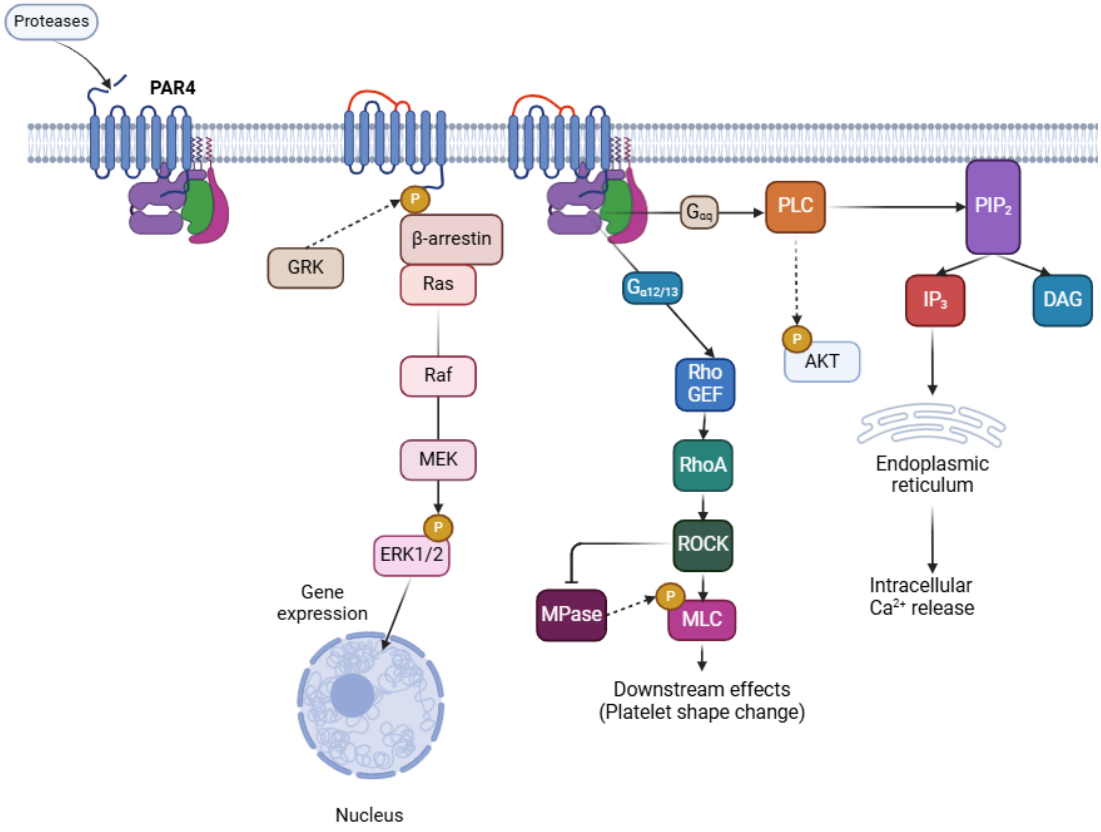


Figure 1.7 – Canonical signalling cascades following PAR4 activation. The receptor can couple to either $G_{12/13}$ or G_q. Stimulation of the first pathway causes cellular shape change through a Rho/ROCK dependent cascade, whereas the latter leads to an increase in cytosolic calcium levels by release of intracellular stores from mitochondria. PLC is also responsible for the phosphorylation of AKT, while phosphorylation of the receptor by GRKs leads to recruitment of β-arrestins which ultimately lead to phosphorylation of ERK1/2 through the Ras/Raf pathway. Both phosphorylations of AKT and ERK1/2 are hallmarks of receptor activation. Figure generated in BioRender.com.

PARs do not return to an inactive conformation after the ligand dissociates. Due to the irreversible proteolytic nature of the activation of the receptor, they are internalised while new receptors are trafficked to the membrane[36]. The receptors are either degraded in lysosomes for signal termination or recycled in endosomes if they were left inactivated. This process is illustrated in figure 1.8. Other members of the PAR family are rapidly internalised after phosphorylation of the C termini of the receptors[104]. However, PAR4 presents structural differences as it has a shorter C terminus compared to PAR1 and failed to undergo agonist-triggered phosphorylation[105]. This leads to a much longer internalisation time and desensitisation. After being stimulated with their respective activating peptides, PAR1 achieved 80% internalisation one hour post stimulation, whereas only 35% of PAR4 receptors were internalised in the same timeframe[105]. This provides another reason for the more sustained signalling following PAR4 activation. PAR4 trafficking is not initiated by β-arrestins, like the majority of other GPCRs[106]. Instead, interactions between the adaptor protein AP2 with a tyrosine motif found within the ICL3 facilitates clathrin-mediated endocytosis. Intriguingly, blockade of the internalisation results in increased ERK1/2 signalling, and reduced AKT signalling[106]. PAR4 trafficking involves transport from the endoplasmic reticulum via the Golgi for post-translational modifications. The receptor possesses an arginine-based ER retention motif in the ICL2 which regulates expression on the cellular membrane. Co-expression with PAR2 interfered with PAR4 ER retention and increased surface expression[107].

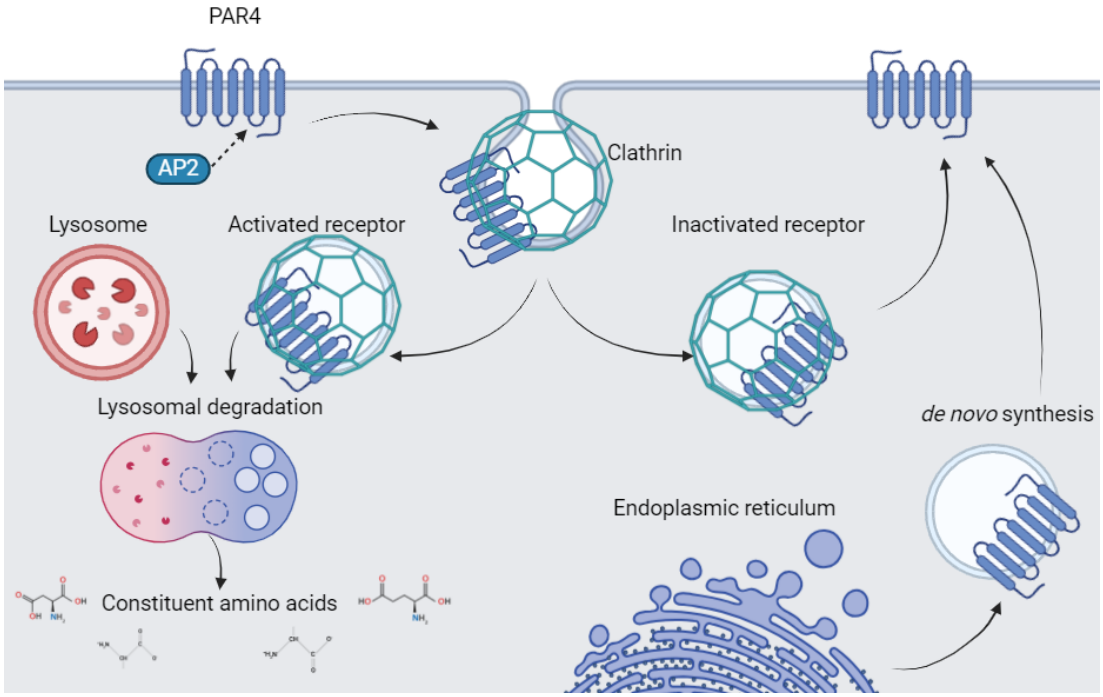


Figure 1.8 – Trafficking and recycling of PAR4 upon signalling termination. Interactions between AP2 and ICL3 of PAR4 aid the internalisation of the receptor through formation of clathrin pits. The newly formed endosomes are either directly recycled to the membrane if the receptor was left inactivated, or they fuse with lysosomes to degrade PAR4 if this had been previously activated. New receptors are then synthesised in the ER. Figure generated in BioRender.com.

On the C terminal side of the receptor, a sequence of amino acids ($R^{352}AGLFQRS^{359}$) regulates calcium signalling, and it is proposed to mediate β-arrestin binding. Ablation of this sequence impedes the activation of kinases such as RSK, GSK, JNK and AKT, as well as G_q dependent calcium signalling[108]. Although PAR4 trafficking is not regulated by β-arrestins, the receptor was still found to interact with this class of proteins by dimerization with ADP receptors on platelets. P2Y1 regulates arrestin-2 recruitment to PAR4 indirectly, through PKC activation. On the other hand, P2Y12 directly aids the recruitment of arrestin-2 by interaction with PAR4[109]. It is proposed that following the recruitment of β-arrestin 2, Lyn is incorporated in the complex. This results in the phosphorylation of AKT through a mechanism depending on PI3K. These interactions are shown in the schematic of figure 1.9. Studies conducted in murine models demonstrated how arrestin-2 might be implicated and contribute to thrombus formation *in vivo*[109]. A recent study also hypothesised how β-arrestins might be implicated in membrane blebbing after activation of PAR4, through a RhoA dependent mechanism[110].

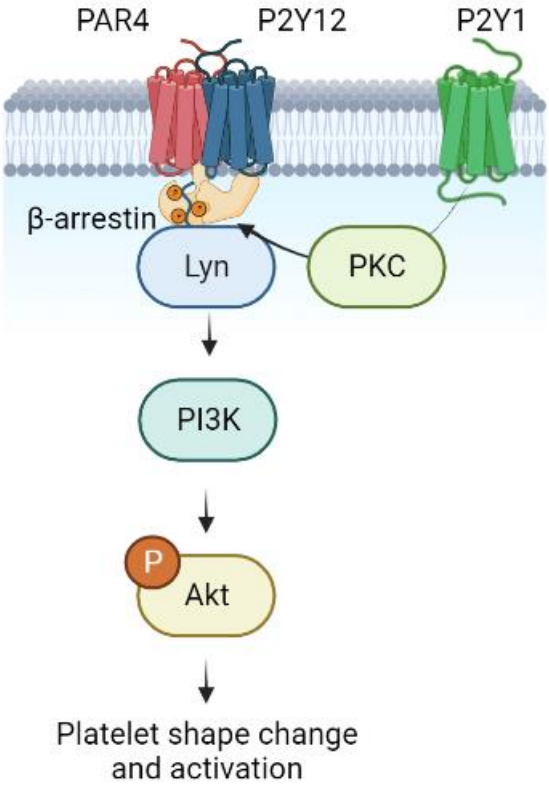


Figure 1.9 – Dimerisation with P2Y12 receptor can lead to recruitment of β-arrestin 2 to PAR4. PAR4 stimulates association of PI3K with β-arrestin 2 in a P2Y12 dependent manner, by incorporating Lyn in the complex. Activation of PI3K in turn results in phosphorylation of AKT. Interactions with P2Y12 are direct, whereas P2Y1 aids β-arrestin 2 recruitment indirectly through activation of PKC. Adapted from Li *et al.*[109].

Recent evidence found a link between PAR4 trafficking and the expression of the C-type lectin receptor CD93, a single pass type I transmembrane glycoprotein involved in inflammation, immunity, and angiogenesis. Platelet aggregation induced by stimulation of PAR4 was significantly reduced in the absence of CD93. Upon stimulation with a PAR4 activating peptide, a more pronounced clearance of PAR4 from the platelet surface was observed in CD93-deficient platelets compared with WT controls. PAR4 is found in cytosolic compartments of activated platelets lacking CD93. Therefore, platelet desensitization following PAR4 stimulation was more pronounced in CD93 KO platelets compared with WT controls[111].

1.10. Genetic differences across species:

When studying PARs, one of the primary limitations is the use of *in vivo* models, as they might not present the same physiological effects that would be exerted on a human being. Human platelets express PAR1 and PAR4, whereas murine models (mice, rats and rabbits) express PAR3 as a cofactor for PAR4[77,112]. Guinea pigs interestingly present both PAR1 and PAR3 together with PAR4[113]. Non-human primates show the same expression profile of PARs as humans[114], and they have been used to study PAR4 specific antagonists[115]. Moreover, the amino acid sequences may considerably vary between species. In figure 1.10 is reported the human homology model of PAR4 aligned with two murine models. Human PAR4 shares a 78% similarity with mice and 75% with rats. Notably, the human receptor lacks a series of amino acids at the beginning of the protein, which are present in both murine models. A recent study confirmed that these structural differences translate also on a functional level. A transgenic mouse line carrying human PAR4 was generated and compared to WT animals. Platelets expressing the receptor were more responsive to PAR-AP, but less to thrombin. These receptors were also irresponsive to GRK6, highlighting different signalling transductions between human and murine PAR4[116].

Figure 1.10 - Homology model of PAR4 sequence between human, mouse, and rat. Differences in amino acids sequences are shown on the lowest row. Percent identity: human - mouse= 77.75%, human – rat= 74.54%, mouse – rat= 92.15%. Homology models generated using Clustal Omega[117]; amino acids sequences retrieved from UniprotKB[118]. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties.

1.11. Studies in animal models:

To validate the effectiveness and safety of new drug candidates *in vivo*, it is necessary to conduct preclinical studies using animal models. However, developing drugs that target PAR4 is a challenging task due to the genetic differences across different species, which not only affect the structure of the receptor but also its expression, as outlined in the previous paragraph. For example, the receptor mediates thrombin signalling in mouse endothelial cells[37] and it drives endothelium-dependent relaxation of aorta in rats[119]. These responses are not observed in human models, unless additional proinflammatory mediators are present[38]. Nevertheless, animal models are still extensively used to understand the PAR4 at the systemic level. PAR4 knockout mice have been generated to study the physiological implication of the receptor activity. Murine platelets are completely unresponsive to thrombin in this scenario, and animals were protected against thrombosis impacting thrombus propagation but not fibrin generation[4]. Ablation of PAR4 from murine platelets reduced haemostatic plug stability in both arterial and venous thrombosis[120]. The same model has been

used in other studies, and deletion of PAR4 in mice showed to be protective also against thromboplastin induced pulmonary embolism[121], laser induced endothelial cell ablation in mesenteric arterioles[4], and injury of the carotid artery[122,123]. Bone marrow transplant studies confirmed protection against thrombosis was due to the lack of PAR4-dependent thrombin signalling in platelets[124]. Besides thrombosis, deletion of PAR4 proved to be beneficial also in murine models of stroke. PAR4 -/- mice showed lower cerebral infarct volume, improved neurological and motor function and reduced blood brain barrier disruption and cerebral oedema[125]. As well as mice, upregulation of neuronal PAR4 can be detrimental in rat models of ischaemia[126]. On the other hand, ablation of PAR4 causes impaired haemostasis, although there is still no evidence of spontaneous bleeding in PAR4 knockout models[127]. High fat diet upregulates cardiac PAR4 in C57Bl/6J mice. This did not correlate with a significant increase in TNFα production, nor it affected NLRP3 or ASC transcriptional priming. High fat diet did not trigger non canonical inflammasome pathways either. Instead, it was found that this specific diet augmented the activation of IL-1β, a proinflammatory cytokine, and gasdermin D, the substrate for caspase-1, in mouse left ventricles. Samples from PAR4 -/- mice showed blunted caspase-1 activation, even if fed with a high fat diet, highlighting the implication of the receptor in triggering this response. Moreover, diabetogenic diet also increased PAR4 expression and responsiveness in C57Bl6/J fibroblasts[128]. It was also found enhanced expression of vascular PAR4 in mouse models of systemic hyperglycaemia.

Conversely, PAR4 deficiency protected against neointimal hyperplasia[102]. A PAR4 inhibitory RAG8 pepducin reduced coronary artery atherosclerosis and myocardial fibrosis in SR-B1/LDLR double knockout mice fed a high-fat, high-cholesterol diet. Moreover, the pepducin reduced vascular cell adhesion molecule 1 (VCAM-1) levels in nonatherosclerotic coronary arteries and reduced leukocyte and platelet accumulation in atherosclerotic coronary arteries[129]. PAR4 was recently found to play a role also in mouse models of acute and chronic kidney injury, where it contributed to fibrosis and inflammation. Although expression of the receptor in this organ is low, unilateral ureter obstruction induced its expression over time[130].

Murine models of PAR4 have been employed also to study the physiological effects of implant surgery. One study investigated early-stage mucosal healing around titanium implants treated with either protease-activated receptor 4 agonist peptide (PAR4-AP) or perpendicularly protruded type I collagen (pCol) in rats. PAR4-AP enhanced laminin-5 expression at 3 days, while pCol treatment promoted collagen fiber formation and orientation. These findings suggest that PAR4-AP and pCol may improve mucosal sealing, potentially reducing complications such as peri-implantitis[131]. Using C57BL/6 mice models of partial hepatectomy, PAR4 inhibition reduced leukocyte recruitment, platelet-neutrophil interaction, microthrombosis, and liver function deterioration within 2 hours of surgery. PAR4 blockade significantly alleviated microcirculatory injury, suggesting that platelets and PAR4 play a critical role in early liver damage after extended resection, and targeting PAR4 could help mitigate these effects[132]. PAR4 protein expression increases in livers after brain death. Receptor blockade alleviates liver injury, reduces platelet activation and accumulation as well as inflammatory response and apoptosis. Moreover, PAR4 antagonism inhibits NF-κB and MAPK pathway activation induced by brain death[133]. A summary of the roles identified for PAR4 from animal studies is given in table 1.3.

Table 1.3 – Roles identified for PAR4 from animal studies.

Model	PAR4 role	Ref.
Isolated neurons from Male Wistar rats	PAR4 is an endogenous analgesic factor that modulates nociceptive responses in normal and inflammatory conditions.	134
Isolated neurons from mice	PAR4 plays a role in modulating visceral pain.	135
Macrophages harvested from female C57/BL6 mice	PAR4 counter-regulates the phagocytic activity of LPS-activated murine macrophages.	136
In vivo pharmacological or thermal stimulation of C57/BL6 N mice	PAR4-dependent pathways modulate the posttraumatic interplay of platelets and CD4+ Tregs.	137
Ola/C57Bl/6 or C57Bl/6 back-ground PAR KO mice	PAR4 is the only PAR responsible for Tissue Factor induced inflammation.	138
PAR4 KO mice	PAR4 contributes to antibacterial defence during murine pneumococcal pneumonia.	139
PAR1 and PAR4 KO C57BL/6 J mice	Lack of PAR4 increased tail bleed times and protection against thrombosis.	124
In vivo and *ex vivo* male C57BL/6 mice and male Wistar rats	PAR4 mediates oedema through recruitment of neutrophils and components of the kallikrein–kinin system	140
Isolated aorta from PAR1 and PAR4 KO mice	PAR1 is the major thrombin receptor in endothelial cells, but PAR4 is also involved.	37
Male C57BL/6J mice subjected to streptozotocin (STZ)-induced type 1 diabetes	PAR4 is expressed in cardiac fibroblasts and is regulated by extracellular glucose *in vitro* and diabetes *in vivo*.	128
PAR4 KO mice, mechanical induction of ischaemia	PAR4 regulates cardiomyocyte survival, and its inhibition can offer cardioprotection after acute IR injury.	141
PAR4 KO C57BL/6J mice, mechanical induction of MI	PAR4 is necessary for resolution of neutrophil-driven inflammation.	142
In vivo and *ex vivo* haemostasis of PAR4 KO C57Bl/6 mice	PAR antagonists used in combination with aspirin provide a potent yet safe antithrombotic strategy.	123
In vivo PAR4 KO C57Bl/6 mice	Genetic deletion of PAR4 impacts both arterial and venous thrombosis, as well as haemostatic plug stability.	120
Male PAR4 KO C57BL/6 mice	Deficiency of PAR4 protects from cerebral ischemia/reperfusion injury.	125
PAR4 KO C57BL/6 back-Ground mice STZ-induced diabetes	PAR4 drives extensive neointimal hyperplasia in diabetic animals.	102
PAR4 KO mice	PAR4-deficient mice platelets failed to change shape, mobilize calcium, secrete ATP or aggregate in response to thrombin.	127

Wistar rats, STZ-induced diabetes	Long-term thrombin inhibition increases vascular PAR4 expression, promotes atherosclerosis, and increases platelet aggregation in diabetic animals.	143
PAR4 KO mice	PAR4 protects animals from viral infections of the heart and lung.	144
PAR4 KO C57BL/6J mice and P-selectin null mice	PAR4 is required for platelet thrombus propagation.	4
Adult male PAR4 KO C57BL6 mice	PAR4 induces markers of inflammation and fibrosis in models of kidney injury	130
Sprague Dawley rats aged 6 weeks	PAR4-AP and pCol may improve mucosal sealing	131
Female C57BL/6 mice aged 10-14 weeks	PAR4 mediates acute hepatic microcirculatory injury after extended liver resection	132
Adult male Sprague-Dawley rats	Blocking protease-activated receptor 4 alleviates liver injury induced by brain death	133

1.12. Pharmacological targeting of PAR4:

The development of PAR4 agonist was relatively straightforward, as activating peptides are based on the amino acid sequences downstream of thrombin cleavage site and mimic the tethered ligand. However, these molecules are often required in prohibitively high concentrations (≈500μM), whereas the highly specific sequence AYPGKF-NH_2 shows an EC_{50} between 5 and 100μM and is often referred to as the standard PAR4 activating peptide (PAR4-AP)[53]. Recently, a more potent PAR4 agonist has been developed using an AYPG-based biased phage-display peptide library approach. Optimisation of the selected phage clones led to the discovery of A-Phe(4-F)- PGWLVKNG, a PAR4-AP with an EC_{50}= 3.4 μM, 16 times more potent than AYPGKF-NH_2[145].

Due to the irreversibility of PAR4 activating mechanisms by enzymatic cleavage, development of PAR4 antagonists is more challenging than the research on agonists. Nevertheless, antagonism of PAR4 has been achieved through different means some more specific of efficient than others. These include peptidomimetics, pepducins, antibodies and small molecules.

1.12.1. Peptide-based approaches:

Similar to the discovery of the agonists, the first approach to antagonism at PAR4 has been developing peptidomimetics based on the tethered ligand sequence which were unable to activate the receptor. Two types of peptidomimetics have been developed: *trans*-cinnamoyl-YPGKF-NH2 and *trans*-cinnamoyl-APGKF-NH2[146]. The compound tc-YPGKF-NH_2 abolished PAR4-AP induced platelet activation in murine models[147] and at high concentrations was effective in inhibiting thrombin-induced platelet aggregation. These results were later confirmed in human models[148], albeit the dependency of this response on PAR4 inhibition has been questioned[149].

Another class of peptide inhibitors is pepducins. The amino acid sequence of these molecules resembles the third intracellular loop of receptors, and the conjugation to a N-terminal palmitate allows the inhibitor to be linked to the cellular membrane. Thus, pepducins bind directly to the G protein and impede downstream signalling, rather

than blocking the interaction of the receptor with an agonist. Pepducins have been developed to target both PAR1 and PAR4[150], and an anti-PAR4 pepducin, P4pal-10 (N-pal-SGRRYGHALR-NH2) proved useful in inhibiting platelet aggregation with IC50=1μM and at 3μM caused increased tail bleeding in mice models[151]. However, the anti-PAR4 molecule shows high cross reactivity with PAR1-AP and other platelet activators such as collagen and Thromboxane A_2 [152]. This could be due to the similarity of the C-terminal sequence of the ICL3 between PAR4 and PAR1. This lack of specificity led to the development of a subsequent PAR4 pepducin, P4pal-i1 ((N-pal-ATGAPRLPST-NH2) which targets the first intracellular loop instead. This inhibitor proved effective in blocking PAR4-AP activity, without affecting PAR1-AP [153].P4pal-i1 inhibits platelet aggregation with an IC50=0.6μM and when administered *in vivo* at 0.13mg/kg dosages, it decreased occlusion time after thrombotic injuries in guinea pigs[153]. Still, a major problem of pepducins is the lack of specificity. The molecules do not discriminate particularly well between GPCRs, as they interfere with the interaction between receptor and G-protein at the membrane, rather than modulate the orthosteric binding site[154].

1.12.2. PAR4 blocking antibodies:

A novel means of blocking PAR4 activity is the use of non-lytic function-blocking antibodies. The first molecule belonging to this class was a rabbit polyclonal antibody developed by Kahn and colleagues[155]. This antibody inhibited thrombin-induced platelet aggregation in concert with PAR1 antagonists. More recent anti-PAR4 antibodies target the sequence C[54]ANDSDTLELPD of the anionic site downstream of thrombin cleavage site. This candidate antibody named CAN12 developed by the Mumaw laboratory inhibited platelet aggregation induced by thrombin, PAR4-AP, ADP and collagen, raising questions about its specificity [156]. The same group developed other monoclonal antibodies (14H6, 5F10) which blocked PAR4 activation and partially inhibited receptor activation by α-thrombin[157]. Together with high concentrations of PAR1 inhibition, anti-PAR4 antibodies showed efficacy in inhibiting calcium signalling in mouse lung fibroblasts and impaired thrombin-mediated human platelet aggregation[158]. Another polyclonal antibody was generated by the Wong laboratory in 2017 from rabbits. This antibody demonstrated antithrombotic effects comparable to clopidogrel in guinea pig models[115]. The most recent antibody developed, named mAb-RC3, is able to impair thrombin-dependent platelet aggregation also in hyperreactive PAR4 variants, and proved effective in reducing thrombi size in *ex vivo* assays[159]. Despite the wide variety of antibodies targeting PAR4, no candidate molecule has been tested yet in human *in vivo* studies.

1.12.3. PAR4 small molecules:

The first small molecule targeting PAR4 to be developed was the indazole derivative YD-3 by Lee and colleagues in 2001[160]. YD-3 showed promising results in inhibiting thrombin induced platelet aggregation in rabbits (IC_{50}= 28μM), and when administered at the concentration 10mg/kg, YD-3 inhibited neointima formation in rats[161]. In human models it showed inhibitory effects only when thrombin concentrations were lower than 0.5nM[162],and it was not able to inhibit PAR1-AP, collagen or U46619, but it selectively blocked PAR4-AP intracellular calcium release[163]. Moreover, *in vivo* studies showed how the high lipophilicity of the molecule can limit its applications[62]. Nevertheless, YD-3 served as the basis for the development of newer, more potent derivatives. Indazole derivatives from the Huang laboratory named compound 19, 25,

and 31 showed to have an IC_{50}>36μM when tested in human leukaemia cell lines[164]. More recent attempts at developing low molecular weight molecules targeting PAR4 include the compound ML354, which not only blocks PAR4-AP activity (IC_{50}=140nM), but was found to be effective also at antagonising PAR1-AP, albeit less potently (IC_{50}=10μM)[165].

In recent times, ethnopharmacological approaches have started to investigate how derivatives from vegetables, fruits and medicinal herbs might modulate PAR4. Resveratrol is a phenylpropanoid produces by grapes, peanuts and a variety of berries[166]. An analogue of this molecule, 3,5,2′,4′-Tetramethoxystilbene (TMS) selectively inhibits PAR4-mediated human platelet aggregation (IC_{50}=2.4 ± 0.5 μM), ATP secretion (complete inhibition at 5 μM), integrin activation, and reduced thrombus formation *in vitro*[167]. Another class of components widely found in food are flavonoids[168]. Using ligand-based virtual screening, the flavonoid 7, 4′-dimethoxy-3-hydroxyflavone was found to exert anti-PAR4 activity. The molecule inhibited human platelet aggregation and integrin activation (IC50=1.4 ± 0.0 μM), as well as blocking other downstream pathways such as Ca^{2+}/protein kinase C, AKT, ERK and p38. In CHO-K1 cells transfected with PAR4, the molecule suppressed β-arrestin recruitment to the receptor. This compound proved to reduce thrombus formation in a microfluidic system, as well as reducing carotid arterial occlusions in mice, without affecting tail bleeding time[169].

A major compound targeting PAR4 currently under clinical investigation is BMS-986120 developed by Bristol Myers-Squibb. BMS-986120 is an orally active, reversible inhibitor showing high selectivity for PAR4, showing a binding affinity K_d=0.098 nM[170]. This molecule was developed by screening more than 1.1 million compounds. In pre-clinical research, this antagonist showed significant antithrombotic activity without associated excessive bleeding. It decreased thrombus formation by 82% when 1mg/kg was administered to non-human primates. BMS-986120 showed a wider therapeutic window when compared to other standard antithrombotic medications such as the $P2Y_{12}$ inhibitor clopidogrel[115]. In a phase 1 clinical trial involving forty participants, the molecule was well tolerated by every patient. No significant bleeding or adverse effects were observed. Its half-life was approximately four hours, while *ex vivo* assays revealed how at 2 and 24 hours BMS-986120 persistently inhibited PAR4-mediated platelet activation[171]. A more recent phase 1 clinical trial assessed the pharmacokinetics, pharmacodynamics, and gene variant effects in humans. BMS-986120 was well tolerated over a wide dose range. No differences in platelet response were recorded between AA120 and TT120 receptor variants[66]. BMS-986120 reduced secondary brain injury in mice after traumatic brain injury (TBI). The drug inhibits thrombin-induced inflammation in astrocytes, via the Tab2/ERK/NF-κB signalling pathway[172].

An analogue of the molecule has been developed as well, BMS-986141. The molecule was tested in a primate model (cynomolgus monkey) and compared to clopidogrel, which is the standard care treatment for arterial thrombosis. BMS-986141 showed comparable antithrombotic efficacy with markedly reduced effects on bleeding. Moreover, BMS-986141 achieved nearly full efficacy at a 2-fold lower dose than BMS-986120[173]. A phase zero trial was conducted *ex vivo* on platelets of fifteen healthy volunteers. The molecule proved to be a selective inhibitor of PAR4-AP stimulated platelet aggregation, platelet-monocyte aggregates, and p-selectin

expression. Moreover, when administered in combination with the factor Xa inhibitor apixaban, BMS-986141 further reduced thrombus area[174].

This second PAR4 inhibitor was tested in a phase 1 clinical trial including 148 healthy participants to assess the safety of the drug with co-administered aspirin (NCT02341638). This study found that 15 and 150mg of the antagonist produced ≥80% inhibition of 25-100 μM PAR4-AP platelet aggregation, independently of PAR1-AP activity, for a period of over 24 hours. Moreover, this study also found that BMS986141 doses above 10 mg completely inhibited platelet aggregation induced by PAR4-AP at concentrations of 21.5 and 25 μM[175]. A follow up phase 2 study of 16 stroke patients were given either BMS-986141 or aspirin, but the trial was prematurely ended due to administrative reasons by the sponsor(NCT02671461). A list of the different classes of antagonists targeting PAR4 is given in table 1.4.

Very recently a series of quinoxaline-benzothiazole antagonists targeting PAR4 have been developed. These are structurally different from previous small molecules. Out of the high throughput screening (HTS) process, the lead hit is compound 48 which showed an IC_{50} as lows as 2 nM against PAR4 activation by γ-thrombin in platelet-rich plasma (PRP). This molecule also displayed high selectivity, blocking PAR4 with a greater than 2500-fold antagonistic activity over PAR1. The robustness of its antithrombotic efficacy and minimal bleeding was also confirmed in cynomolgus monkey models[176].

Due to the inherently difficult nature of pharmacological targeting of PAR4, owed to the tethered ligand activation dynamics, PAR4 served as the perfect model to develop computational pipelines for drug discovery. A novel method based off the structural generation of PAR4 model and virtual high throughput screening using DOCK, led to the identification of a one-hit lead compound. This was further optimised to develop a series of PAR4 antagonists, which potently inhibit thrombin-induced receptor activation, but interestingly do not affect PAR4-AP[177].

A recent study reported the discovery of novel 2,3-dihydro[1,4]dioxino[2,3-g]benzofuran compounds as PAR4 antagonists for arterial embolism treatment. Isomers 36 and 37 demonstrated potent *in vitro* antiplatelet activity y (IC50 = 26.13 nM for 36 and 14.26 nM for 37) and improved metabolic stability in human liver microsomes. Compound 36 showed favorable oral pharmacokinetics in mice and, along with 37, exhibited strong *ex vivo* antiplatelet effects without affecting coagulation or causing significant bleeding[178].

Table 1.4 – Different antagonists targeting PAR4.

Name	Class	IC_{50}	Target	Model	Ref.
tc-YPGKF-NH_2	Peptidomimetic	100 µM	Tethered ligand binding site	*Ex vivo* Rat platelets	[147]
P4pal-10	Pepducin	1 µM	ICL3	*In vivo, ex vivo* Mouse	[151]
P4pal-i1	Pepducin	5 µmol/L	ICL1	Guinea pig	[153]
CAN12	Blocking antibody	10 ng/mL	Anionic region (D^{57},D^{59},E^{62},D^{65})	*Ex vivo* Human platelets *In vivo* Guinea Pig	[157]
YD-3	Small molecule	28 µM	Tethered ligand binding site	*Ex vivo* Human and mice platelets	[162]
BMS-986120	Small molecule	7.3 nM	Unknown – thought to be competitive	*In vivo, ex vivo* Guinea pig Cynomolgus monkey	[115]
BMS-986141	Small molecule	2.2 nM	Unknown – thought to be competitive	*In vivo, ex vivo* Cynomolgus monkey	[173]
Compound 48	Small molecule	2.0 nM	Tethered ligand binding site	*In vivo, ex vivo* Cynomolgus monkey	[176]

1.13 Clinical significance of targeting PAR4:

PAR4 deficiency showed to be protective against stroke in murine models of transient middle cerebral artery occlusion. Mice showed reduced infarct volume and improved neuronal function after the injury when compared to PAR+/+ controls[125]. Moreover, the novel PAR4 antagonist BMS-986141 was tested in a small-scale phase 2 clinical trial (identifier: NCT02671461) in patients who suffered transient ischaemic attack (TIA)[179]. This highlights how targeting PAR4 might become an avenue for the prevention of the reoccurrence of such condition.

Kolpakov laboratory performed extensive research on the implication of PAR4 in cardiac inflammation. First, in 2016, they found that PAR4 expression is upregulated after myocardial ischemia reperfusion (IR) injury. Deletion of PAR4 resulted to be cardioprotective after myocardial IR, as well as reducing inflammation post-injury. PAR4 KO mice demonstrated increased tolerance to IR, since PAR4 stimulation led to cardiomyocyte apoptosis through JNK pathway[141]. However, in a follow up study from 2020, the same group showed how PAR4 is necessary for resolution of inflammation and its deletion results to be deleterious on a longer term. In fact, PAR4 is not only expressed in platelets, but also on neutrophils. Its stimulation induces the apoptosis of these granulocytes, a crucial step for the resolution of inflammation. PAR4 deficiency thus leads to cardiac haemorrhage and increases chances of cardiac rupture after myocardial infarction (MI). Interestingly, adoptive transfer of PAR4 expressing neutrophils can dampen inflammatory responses and improve cardiac remodelling after MI[142]. This is particularly relevant in a clinical context, as acute transient administration of PAR4 inhibitors can prevent early ischaemic stroke, however sustained administration of pharmacological agents may be deleterious for myocardial healing.

In acute coronary syndrome there is still a lack of therapies, as patients are not well managed with current treatments. The standard short- and long-term strategy is administration of anti-platelet agents[180], but concomitant use of aspirin and $P2Y_{12}$ inhibitors is also common. The major contraindication of this dual therapy is excessive bleeding[181]. Moreover, evidence shows that these therapies only partially antagonise the increased procoagulant activity of platelets in this cohort of patients[182]. Targeting PAR4 might therefore be a more efficient alternative method to treat acute coronary syndrome.

A rare event following percutaneous coronary intervention is acute stent thrombosis. This event is clinically managed with anticoagulants such as unfractionated heparin or enoxaparin during the acute phase of the disease, while the chronical management is achieved by administration of $P2Y_{12}$ inhibitors like clopidrogel or prasugrel. Albeit patients are already successfully managed with the aforementioned therapies, a small group is still not responding and could benefit from the specific targeting of PAR4, due to its clear role in thrombosis[183].

One of the most important, yet often underrated, complications of diabetes is its involvement in cardiovascular diseases. This condition appears to cause resistance to anti-platelet therapies in both animals and patients [184,185]. As mentioned previously, PAR4 is upregulated in this specific clinical setting and might therefore serve as a useful target to treat diabetes-related cardiovascular complications. Although just in murine models, evidence already shows the involvement of the receptor and its

increased expression, as well as enhanced PAR4 responses in diabetic mice[102,143]. The major clinical outcome of targeting PAR4 could be a reduction in vascular hypertrophy, which could in turn reduce inflammation and the risk of thrombosis. Even in cancer, after the malignancy itself, the leading cause of death is cancer associated thrombosis. Tumour progression correlates with platelet activation and coagulation. Increased levels of platelet-derived extracellular vesicles increase the occurrence of thrombosis in cancer patients[186]. PAR4 may therefore play a role in this scenario, as evidence shows that the receptor was upregulated in breast and colon cancer[186,187].

Recent evidence showed that F2LR3, the gene encoding for PAR4, can be epigenetically regulated by tobacco smoke. The association between tobacco smoke and DNA hypomethylation was already well acknowledged, but a recent study[188] demonstrated how the reduced regulation of F2LR3 results in increased expression of PAR4. The increased expression of the receptor on platelets would lead to complications at the systemic level, such as increased risk of atherosclerotic events and myocardial infarction. Moreover, they also demonstrated the implication of the rs773902 missense variant (A120T), and how this mutation may have pleiotropic effects on platelet function. Besides causing hyperreactivity in platelets, cells carrying this mutation would further increase receptor expression when exposed to cigarette smoke. Recently, F2LR3 mRNA was found to be elevated in Alzheimer's disease cases and was associated with worse retrospective longitudinal cognitive performance. This relationship was attenuated in patients without cognitive impairment[189].

Due to racial differences in cellular expression of PAR4, there is a need of personalised therapies according to the genome of the patient. Missense variants are therefore a crucial factor to take into consideration when developing new agents targeting PAR4. Dimorphism at amino acids A120T and F296V are particularly deleterious as they result in hyperresponsive platelets. Moreover, the A120T mutation already proved to be pharmacologically resistant to inhibition by YD-3, as well as not being affected by PAR1 antagonists[62].

Finally, anti-PAR4 agents may provide safe alternatives to drugs targeting PAR1. After the discovery of PAR receptors on platelets, many efforts have been invested in generating novel anti-platelet drugs. This led to the development of vorapaxar (trade name Zontivity), which has been approved by the U.S. Food and Drug Administration (FDA) for clinical use in prevention of myocardial infarction and peripheral artery disease. This novel PAR1 antagonist was tested in two large-scale phase III clinical trials. It was found that the molecule effectively reduced ischaemic output in patients with a history of MI or peripheral arterial disease, with concomitant use of clopidogrel and/or aspirin. However, one of the clinical trials was terminated early since administration of the drug was associated with a significant increase in intracranial haemorrhage[190]. Interestingly, it was later found that vorapaxar inhibited thrombin-induced platelet activation without affecting PAR4 responses[191]. As mentioned previously, PAR1 activation leads to a quick, transient activation of platelets necessary to stop excessive bleeding. On the other hand, PAR4 activation gives a slower, more transient stimulation, necessary to form the platelet plug and recruit more cells to the site of injury. For this reason, targeting PAR4 over PAR1 may be a better strategy to prevent thrombosis without excessive bleeding.

AIM AND HYPOTHESIS:

The aim of this thesis is to comprehensively investigate the interaction network and subcellular localization of PAR4. This is achieved by identifying and characterizing PAR4's interaction partners through SILAC proteomics, employing computational tools to predict and refine these interactions, and using advanced imaging techniques to validate key findings and explore PAR4's subcellular roles. Together, these approaches aim to expand our understanding of PAR4's functions, including its potential involvement in mitochondrial processes and signalling pathways mediated by PDZ domain-containing proteins.

We hypothesise that:

1. PAR4 engages in a diverse network of protein-protein interactions, including those involving PDZ domain-containing proteins and mitochondrial proteins, which play a role in its cellular signalling and localization. To investigate this, SILAC proteomics is used to identify and functionally characterize PAR4-interacting proteins, with a focus on PDZ domain-containing proteins and mitochondrial components.
2. Advanced AI tools, such as AlphaFold and PrePPI, can effectively predict and prioritize PAR4's interactions from proteomic data. These predictions refine the list of candidate proteins for experimental validation, particularly focusing on high-priority interactors like the GIPC protein family.
3. Selected interactions, particularly those with GIPC family proteins, can be validated *in vitro* using Proximity Ligation Assay (PLA), providing evidence of close molecular proximity between PAR4 and its interactors.
4. PAR4 localizes to mitochondria, as suggested by proteomic findings. Advanced imaging techniques, including confocal microscopy, Structured Illumination Microscopy (SIM), and Stimulated Emission Depletion (STED), are employed to confirm its subcellular localization and association with mitochondrial components.

2.MATERIALS AND METHODS

2.1. MATERIALS

2.1.1 Reagents:

Table 2.1 – List of reagents used to perform experiments.

REAGENT	SUPPLIER	FORM	CAT.NO.
Accutase	Sigma-Aldrich	Liquid	A6964-100mL
Acrylamide	Carl Roth GmbH	Liquid	3029.1
Ammonium persulfate (APS)	Thermo scientific	Powder	17874
Anti-HA (PLA)	Sigma-Aldrich	Liquid	A2095-1ML
Bovine Serum Albumin (BSA)	Fisher scientific	Powder	BP702-100
β-mercaptoethanol	Sigma-Aldrich	Liquid	60-24-2
Cal520 dye	Abcam	Powder	ab171868
Calcium Chloride	VWR BDH Chemicals	Liquid	190464K
Complete™ Protease Inhibitor	Roche	Tablet	04 693 124 001
4',6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich	Liquid	28718-90-3
Dulbecco's Modified Eagle Medium (4.5g/L Glucose, 0,11g/L Pyruvate, [-] Glutamine) (DMEM 1x)	Gibco®	Liquid	11500596
Dimethyl sulfoxide (DMSO)	Fisher bioreagents	Liquid	BP231-1
Dithiothreitol (DTT)	Chem Cruz	Powder	Sc-29089A
Dulbecco's modified Eagle's medium	Sigma-Aldrich	Liquid	D6434
Foetal calf serum (FCS)	Biosera	Liquid	S00FQ10001
Formaldehyde	Sigma-Aldrich	Liquid	47608-1L-F
Glucose	VWR BDH Chemicals	Powder	101174Y
Glycine	VWR BDH Chemicals	Powder	101196X
Hydrogen Peroxide	Chem Cruz	Liquid	Sc-203336
LDS sample Buffer	Novex	Liquid	NP0008
L-glutamine	Sigma-Aldrich	Liquid	G7513
Luminol	Sigma-Aldrich	Liquid	A8511-5G
Methanol	VWR BDH Chemicals	Liquid	20847.307
Mitochondria Isolation Kit for Cultured Cells	Thermo Fisher scientific	Kit (Liquids)	89874
MitoTracker™ Deep Red FM	Thermo Fisher scientific	Powder	M22426

Mowiol 4’888	Sigma-Aldrich	Powder	9002-89-5
NaveniFlex Cell MR RED	Navinci	Kit (Liquids)	39505
Phosphate buffer saline (PBS)	OXOID	Tablets	BR0014G
p-coumaric acid	Sigma-Aldrich	Powder	C9008-5G
Pen-Strip	Sigma-Aldrich	Liquid	P4458
Pierce™ BCA Protein Assay Kit	Thermo scientific	Kit (Liquids)	23225
Pluronic F-127	Thermo Fisher scientific	Liquid	P3000MP
Polyvinylidene fluoride (PVDF) membrane (0.45µm pore size)	GE Healthcare Life Sciencesences	Membrane	10600029
Protein G Plus / Protein A Agarose suspension	Calbiochem	Liquid	IP05-1.5ML
Protein ladder	Thermo scientific	Liquid	26619
Rhodamine Phalloidin	Thermo scientific	Powder	R415
Sodium dodecyl sulphate (SDS)	Fisher bioreagents	Powder	BP166-500
Sodium Chloride	Fisher scientific	Powder	S/3160/60
N,N,N’,N’-tetramethylenediamine (TEMED)	Sigma-Aldrich	Powder	110-18-9
TripLE™ Express	Gibco®	Liquid	12604021
Tris Base	Fisher bioreagents	Powder	BP152-1
Triton X-100	Sigma-Aldrich	Liquid	T-8787
Tween20	Sigma-Aldrich	Liquid	P9416
UltraCruz® Autoradiography film blue	Santa Cruz Technology	Film	SC-201696

2.1.2 Buffers:

Solution	Recipe
1x HEPES Buffer	Prepared fresh each time right before experiment 10X HEPES buffer – 100 mL ddH_2O – 900 mL Glucose – 1.98 g 1M $CaCl_2$ – 1.8 mL pH 7.3
10x HEPES buffer	ddH_2O – 1 L NaCl – 71.30 g KCl – 3.72 g HEPES – 23.83 g KH_2PO_4 – 0.68 g NaH_2PO_4 – 0.60 g 1 M $MgCl_2$ – 10 mL
10% Resolving gel	2x medium or 1x thick H_2O – 5 mL Buffer1 – 3 mL Acrylamide – 4 mL 10% APS – 100 mL TEMED – 10 mL
Buffer 1	Tris base – 90.75 g SDS – 2 g dH_2O – 400 mL pH – 8.8 Make up to 500 mL with dH_2O
Buffer 2	Tris base – 30.25 g SDS – 2 g dH_2O – 400 mL pH – 6.8 Make up to 500 mL with dH_2O
ECL1	dH_2O - 88.5 mL Luminol – 1 mL Coumaric acid – 480 mL 1 M Tris – 10 mL
ECL2	dH_2O – 90 mL H_2O_2 – 64 mL 1 M Tris– 10 mL
Lysis buffer base (100mL)	50 mM TRIS (pH7.4) + 0,5% v/v triton x100 in dH2O
Lysis buffer	10 mL base + 1 tablet from fridge in front of fume hood. Keep on ice/fridge
Mounting medium	0.2 M Tris (pH 8.5)– 12 mL Glycerol – 6 g Mowiol – 2.4 g (add slowly) dH_2O – 6 mL
Running buffer	Tris base -3 g Glycine – 14.4 g SDS – 1 g 1 L water (dH_2O)

Stacking gel	H_2O – 4.9 mL Buffer2 – 1.9 mL Acrylamide – 0.75 mL APS – 100 mL TEMED – 10 mL 10% APS = 0.1 g in 1 mL dH_2O 0.1% SDS = 0.05 g in 50 mL dH_2O
Stripping buffer	Tris base – 3.8 g SDS – 10 g dH_2O – 400 mL pH – 6.7 Make up to 500 mL with dH_2O
TBS-T	NaCl – 8.76 g Tris base – 2.4 g 1 L dH_2O Tween20 – 1 mL pH 7.5
Transfer buffer	Tris base – 3 g Glycine – 14.4 g Methanol – 200 mL dH_2O – 800 mL
Tris 1M for ECL	Dissolve 30.3 g in 150 mL dH_2O pH 8.5 Make up to 250 mL dH_2O

2.2. METHODS:

2.2.1. Cell culture safety note:

Work was performed in a sterile BioMAT^2 class II microbiological safety cabinet (MSC). All the surfaces of the MSC were sprayed with 70% ethanol before and after using it. All the instruments and reagents placed in the MSC were also sprayed with ethanol. Gloves were cleaned with ethanol as well. Work was performed lifting arms up in order not to interrupt the airflow inside the MSC. Cell morphology and growth was monitored using a Nikon Eclipse (TE300) inverted microscope (Nikon, Tokyo, Japan) and a Leica EC3 digital camera affixed to a Leica DM IL LED inverted microscope (Leica Biosystems, Wetzlar, Germany). Imaging was conducted with a 10- or 20-times magnification.

2.2.2. HEK293 cell culture:

HEK293 cells were cultured in DMEM medium supplemented with 10% (v/v) FBS, 1% (v/v) penicillin/streptomycin and 1% (v/v) and 1% (v/v) non-essential amino acids (NEAA). Cells were incubated in a humidified atmosphere at 37°C with 5% CO_2.The medium was changed every 48 hours and cells were split when around 70-90% confluence using 1x Saline sodium citrate (SSC).

2.2.3. Lysing cells:

Cells were first grown to very high confluence (≈100%), then they were washed twice with PBS carefully pipetting on the walls of the flask and not directly to the cells. 1 mL fresh (<1 week old) Lysis buffer was added, and cells were scraped from the bottom of the flask. The lysate was transferred to 1.5 mL Eppendorf tubes, rotated at 40 x G for 1 hour and then spun at 12.000 x G for 5 minutes. The pellet was discarded, supernatant was transferred to a new 1.5 mL Eppendorf. BCA assay was then performed to quantify protein content.

2.2.4. Freezing down cells:

Cryotubes were labelled with cell type, passage number, date, and initials. Cells were centrifuged at 300 G for 5 minutes, the supernatant was discarded, and pellet resuspended in 10% DMSO in FCS. 1mL was dispensed in each cryotube. Cryotubes were stored at -80 °C for brief period, in liquid nitrogen for longer terms.

2.2.5. Plasmids:

All protease-activated receptors (PAR) constructs were fluorescently tagged at the extreme C-terminal and have been previously published[192]. The following constructs were used in the study, CFP –N1 (Clontech), human PAR4-CFP, PAR4 'Y157C'-CFP. Y157C reflects the point mutation introduced to the PAR4 sequence that represents the recently published patient variant identified[193]. The mVenus-YFP vector was a gift from Steven Vogel[194]. The mVenus-YFP vector was used to insert PAR4 and PAR4ΔPDZ to create human PAR4-mVenus YFP, PAR4ΔPDZ –mVenus YFP plasmids. PAR4ΔPDZ represents removal of proposed PDZ ligand sequence from the extreme C-terminal of PAR4 ($S^{381}SLLQ^{385}$). The generation of the PAR4 mVenus constructs was carried out by Dr Roth Tate (University of Strathclyde). PAR4-HA plasmids were generated by Dr Margaret Rose Cunningham.

2.2.6. PAR4 proteomics:

SILAC was performed at Bristol Proteomics Facility in collaboration with Dr. Kate Heesom. The workflow of the laboratory techniques as well as the bioinformatic analysis is summarised in figure 2.1.

2.2.6.1. Introduction to Stable isotope labelling of amino acids in cell culture (SILAC):

Stable isotope labelling of amino acids in cell culture (SILAC) is a technique based on mass spectrometry (MS) to quantify protein abundance. This method was first described in 2002 by Ong and colleagues[195] and since then has been widely employed in proteomics to study the dynamics of posttranslational modifications, protein-protein interactions and protein turnover. In this method, cells are grown in media containing arginine and lysine labelled with stable isotopes of carbon and nitrogen. Plasmids tagged with fluorescent proteins can be transfected into the labelled cells. In our experiment we transfected YFP in the light medium, PAR4-YFP in the medium medium, and PAR4$_{\Delta SLiM}$-YFP The cells are then lysed, and samples are immunoprecipitated. These immunoprecipitations are mixed and submitted for MS analysis. The heavy and medium label causes a mass shift which can be detected and quantified relatively to the light labelled cells. SILAC immunoprecipitation allows to identify large numbers of both direct and indirect interactions with the transfected proteins from cultured cells.

2.2.6.2. Sample preparation:

All SILAC reagents were sourced from Thermo Fisher, except for dialyzed FBS, which was purchased from Sigma-Aldrich. HEK293 cells were metabolically labelled with light (R0K0), medium (R6K4), or heavy (R10K8) DMEM. SILAC media was prepared by supplementing arginine- and lysine-free DMEM with isotopically labelled L-arginine and L-lysine, along with 10% dialyzed FBS and 1% penicillin-streptomycin. Cells were cultured in SILAC DMEM for at least six passages to ensure complete labelling of proteins. Plasmids encoding YFP (R0K0), wild-type PAR4-YFP (R6K4), and PAR4-YFP ΔSLiM (R10K8) were transfected into labelled cells. Transfections were performed in media lacking antibiotics to avoid interference with transfection reagents. After 24 hours, cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.5), 0.5% Triton X-100, and protease inhibitor cocktail (Roche). Soluble lysates were affinity-purified using GFP-trap beads (Chromotek) following the manufacturer's protocol. The purified protein samples were mixed in a 1:1:1 ratio before gel separation, tryptic digestion, and LC-MS/MS analysis.

2.2.6.3. LC-MS/MS Analysis:

Samples were analysed using an Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). The Orbitrap was configured for survey scans at a resolution of 60,000, selecting the top six ions per duty cycle for MS/MS fragmentation in the LTQ linear ion trap. Data were acquired using Xcalibur v2.1 software and processed using Proteome Discoverer v1.2 (Thermo Fisher Scientific).

2.2.6.4. Peptide-spectrum matches – False Discovery Rate Rationale:

The raw proteomic data files were processed and quantified using Proteome Discoverer software v1.2 (Thermo Fisher Scientific) with searches performed against the UniProt human database by using the SEQUEST algorithm with the following

criteria: peptide tolerance at 10 ppm, trypsin as the enzyme. The reverse database search option was enabled, and all data were filtered to satisfy a false discovery rate (FDR) of <5%. Each spectrum that was generated by the MS was searched against a species protein database (DB) and also against the same DB but with all the protein sequences reversed (reverse decoy database). Each match of a spectra to a peptide in these databases was given a score, which depends upon how closely the spectra matches that predicted for the given peptide sequence. The better the match, the higher the score. The software looked at the distribution of the scores matched to both the genuine and the reversed decoy database and calculated the score cut-off at which there is only a 5% chance that a peptide matched to the reversed DB, rather than the real one. All peptides with a score below this calculated cut-off are low confidence peptides (i.e. Identified at >5% FDR) and excluded from the datasets. Similarly, data sets were filtered for High confidence level at <1% FDR. The software looked at the score distributions to identify the score value at which only 1 in every 100 peptides with that score matches to the decoy DB. Peptides with a score above this calculated cut-off were then regarded as High confidence peptides (ie. Identified at <1% FDR). Medium confidence peptides are those with scores that fall between these two cut-offs (i.e. Identified at >1%FDR, but <5%FDR). Therefore, the lists generated have at least 95% confidence that every peptide present was a genuine peptide from a protein in the species-specific database.

2.2.6.5. Data filtering:

Analysis was performed using R v. 4.2.3 running on R studio. Data were filtered according to the number of unique peptides identified. Proteins lacking SILAC ratios were removed as well. The remaining SILAC ratios were converted to $\log_2$ values to make them fit in a Gaussian distribution centred around the SILAC $\log_2$ value 0. For each SILAC ratio, 1.96 Standard Deviation was added to mean values. This generated threshold values with 95% confidence limits specific for every different SILAC ratio. The thresholds thus calculated were used to determine which proteins were interacting with PAR4 in each repeat, these data were then merged. A minimum of 3 out of 5 replicates, or 2 out of 3 replicates, was considered a hit as shown in previous publications[196].

2.2.6.6. Up- and down-regulation of common proteins visualised by volcano plotting:

SILAC ratios were compared across groups using paired t-test to generate p-values, which was then converted to $-\log_{10}$. Log_2 fold changes were then plotted against $-\log_{10}$ values to generate a volcano plot. To determine which proteins were upregulated and which ones were downregulated, threshold values on the $\log_2$ axis were set at -1 and +1, to represent the halving or doubling of the protein expression respectively.

2.2.6.7. Network analysis performed by STRING and Cytoscape:

Protein-protein interaction networks were studied using the STRING extension on Cytoscape. Gene ontology analysis was performed using the STRING plugin, collating information from different databases. The initial large network was then clustered, using an inflation parameter set at 4 (standard value provided by the app). The smaller clusters thus generated were then enriched again using the STRING plugin, and the nodes were colour-coded accordingly to their subcellular localisation.

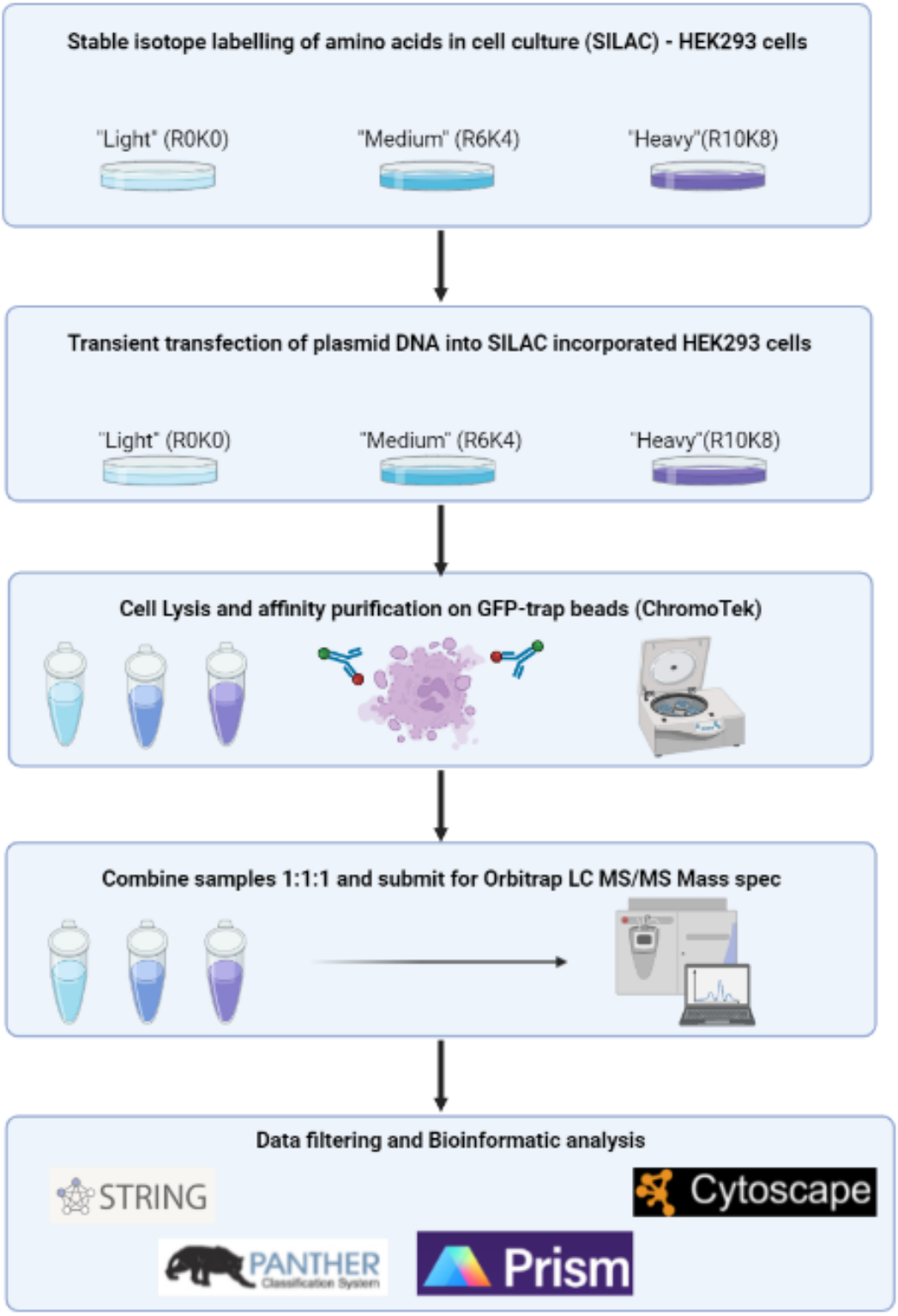


Figure 2.1 – Workflow diagram of the proteomics analysis. SILAC metabolic labelling, affinity purification and LC/MS MS was performed at Bristol Proteomics Facility by Dr. Kate Heesom laboratory. Bioinformatic analysis included Network analysis using STRING and Cytoscape, gene ontology enrichment analysis was performed on RStudio running R v.4.2.3.

2.2.7. AlphaFold:

Structure predictions were generated on ColabFold[197]. However, the Jupyter Notebook runtime did not allow to generate large predictions without running out of computational time. To overcome this issue, ColabFold (v. 1.5.2) was accessed through a science gateway platform, COSMIC2[198]. This platform is hosted on Expanse, a dedicated Advanced Cyberinfrastructure Coordination Ecosystem located at the University of San Diego CA and can be accessed only using an ORCID account. The greater computational capability of the cloud system, allowed to perform protein predictions in as little as two hours.

The first step in running a prediction was to generate a FASTA file containing the amino acid sequence of the proteins of interest. The sign “>” indicated the name of a new protein, and the actual sequence had to follow at the start of a new line. Sequences were retrieved from the database UniProt. Once all the FASTA files were prepared, they were uploaded to the COSMIC2 website, then a Task can be created. The first step when preparing a new Task is providing a description, or a title for the task. Then the input data is selected from the FASTA files uploaded previously. The next step is to select the tool to run the prediction. As COSMIC2 was a science gateway initially thought for Cryo-EM structures, most of these tools revolve around the processing and analysis of Cryo-EM and tomography data. However, since the release of protein prediction tools, the website now hosts tools such as AlphaFold2 and ColabFold. For the generation of the figures in this chapter, ColabFold (v. 1.5.2) was selected due to its faster predictions. Once ColabFold was selected as a tool, the last step was to specify the parameters for the prediction. The first parameter was the number of predictions to be generated, this was set as the standard number of 5. The second parameter was the number of recycles the algorithm had to repeat after obtaining a structure. The standard value was 3, but it was increased to 10 to obtain more confident predictions. This value is relative, as the next parameter was the stop score, meaning the algorithm would stop recompute the network after a certain threshold of pLDDT was reached. The standard value would have been 80, but this value was also increased to 90 to obtain more confident predictions. The web interface then asked if the predictions should be based off templates of published PDB structures, but this parameter did not affect the quality of the predictions at the cost of increasing the computing time, and it was therefore not selected. Then, Amber molecular dynamics relaxation[199] was applied to enforce exact peptide bond geometry post-prediction. This is done to remove distracting stereochemical violations without losing accuracy. Finally, the Max MSA depth was left to “auto”. Once all the parameters were set, the predictions were launched directly from the web interface. However, due to the computing costs for the cloud a max of two contemporary predictions were allowed, and a total max of 72 hours of computing time are given to each COSMIC2 user.

Each AlphaFold prediction created five models, ranked based on their mean pLDDT, pTM, and ipTM. The outputs were five different .PDB files, each with its own confidence metric, stored in a .JSON file. The 3D protein structures were visualised using the online software Mol* (https://molstar.org/viewer/). The interactions between the proposed SLiM motif and PDZ regions were assessed looking at the hydrogen bonds forming between the receptor and the scaffold protein, as well as the proximity of the SLiM with hydrophobic pockets and conserved binding residues.

2.2.8. HEK 293 transfection and colocalisation studies:

HEK293 cells were grown on coverslips in until 50% confluent. 1 µg total DNA was mixed with 5.7 µl PEI in 50 µl medium per well. In case of transfection of multiple fluorescent-tagged constructs, 0.5 µg PAR4-mCherry DNA and 0.5 µg GIPC-tGFP were mixed in the same tube. The mixture was gently mixed and incubated for 30 minutes, after which it was added to the cells together with 450 µL of fresh medium per well, and the plate was incubated at 37 °C 5% CO_2 for 48 hours.

The day after, the cells were fixed by adding 3.6% formaldehyde for 10 minutes. Due to the semi-adherent nature of the cells, the coverslips were carefully washed only once in PBS, before staining nuclei with DAPI (dilution 1:2000 incubated for 5 minutes) or mitochondria with MitoTracker (25 nM incubated for 15 minutes). The coverslips were carefully washed again once with PBS, before mounting them on glass slides. Images were captured on Leica TCS SP8 confocal microscope. Colocalisation analysis was performed using the JACoP plugin of ImageJ. Pearson's correlation coefficient was calculated to assess the degree of colocalisation of the receptors with proteins or organelles of interest.

2.2.9. Immunofluorescence staining:

Cells were seeded on coverslips placed in a 12 or 24 well plate and incubate at 37°C 5% CO_2 overnight or until confluent. The next day cells were carefully washed with PBS three times, before fixing them with 3.6% formaldehyde for 10 minutes. The formaldehyde was disposed in its special waste tank. The cells were again washed twice with PBS and then permeabilised with 0.25% Triton for ten minutes. After a consequent double wash with PBS, cells were blocked with 1% BSA in PBS for 30 minutes at room temperature. Antibodies dilutions were performed directly on parafilm due to the low volume required. The sample coverslips, but not the control coverslip, were laid with the cell side facing the antibody drops and incubated overnight in a humid box. The day after, the coverslips were washed three times with PBS and blocked again with 1% BSA for 15 minutes. The dilutions for secondary antibodies were performed directly on parafilm, similarly to the primary antibodies. This time also the control coverslip was exposed to the antibodies. Coverslips were incubated in the dark for two hours and then washed three times with PBS. 40 µL of 25 nM MitoTracker Deep Red FM were added to every coverslip, and then incubated in the dark at 37°C for 15 minutes, after which coverslips were washed twice in PBS. Coverslips were then stained with 1:2000 DAPI in PBS and incubated for 5 minutes in the dark. Cells were washed once with PBS and then stuck on glass slides by dispensing drops of mounting media.

2.2.10. Proximity Ligation assay (PLA)

Proximity Ligation Assay is a powerful tool to detect protein-protein interactions *in situ*. A signal is generated when interacting partners lie at a distance smaller than 40nm, which can be detected by classical microscopy methods. Cells were transfected with PAR4-HA and GIPC1/2/3-YFP. Selective primary antibodies directed against the HA epitope of PAR4-HA (anti-HA mouse-monoclonal antibody, BioLegend) and GIPC-GFP tagged proteins (anti-TurboGFP rabbit polyclonal antibody, Origene) were used. Oligonucleotide-conjugated secondary probes were added and if PAR4-GIPC1/2/3 were within the <40 nm proximity, a closed circle would form between the two oligonucleotides. Treatment with a ligase enzyme resulted in ligation of the oligonucleotides and subsequent polymerase treatment led to rolling circle amplification (RCA), forming a DNA structure that was detected by the labelled oligonucleotides. This signal was then visualised as fluorescent spots under the microscope. A schematic of this technique is illustrated in figure 2.2. PLA was performed following the NaveniFlex fluorescence *in situ* protocol. Briefly, cells are permeabilised and blocked in a similar manner to immunofluorescence. Primary antibodies are then incubated at 37°C for an hour or at 4°C overnight to target the proteins of interest (PAR4 and GIPC). The next day, slides are washed and Navenibodies M1 and R2 are added to the sample, followed by another incubation for an hour at 37°C. Navenibodies are washed in pre warmed TBS-t, followed by addition of Buffer1 and Enzyme1, which are incubated in a preheated humidity chamber for 30 min at 37 °C. Samples are then washed again using TBS-T. Buffer2 and Enzyme2 are then added and the samples incubated in a preheated humidity chamber for 90 min at 37 °C. Again, samples are washed, nuclei stained with DAPI and slides mounted on coverslips. PLA signal was detected using the Leica SP8 confocal microscope under the red channel (555nm wavelength).

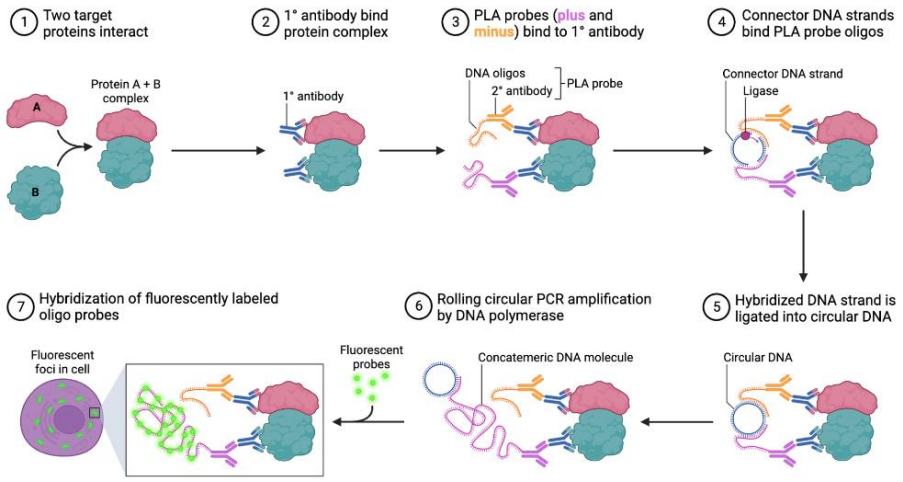


Figure 2.2 - Schematic of PLA procedure. The primary antibodies (HA and Turbo GFP) bound to PAR4 and GIPC1/2/3 tags respectively. The oligonucleotide conjugated secondary probes then bound to the HA and Turbo primary antibodies. Ligase and oligonucleotides hybridize the PLA probes and make a closed circle. This was used as a template for rolling circle amplification (RCA). Fluorescently labelled oligonucleotides hybridize with the RCA product to generate distinct spots that can be detected readily under any fluorescent microscope. Image reprinted from “Proximity Ligation Assay Protocol”, by Nashed and Mrovecova on Biorender (2024).

2.2.11. . Mitochondrial isolation:

Mitochondria have been isolated using the ThermoScientific™ Mitochondria Isolation Kit for Cultured cells. The extraction was performed following the kit protocol. Briefly, up to $20x10^6$ cells were pelleted and 800 µl Reagent A was added, and the tubes were incubated 2 minutes on ice. 10 µl Reagent B were added to the mixture, followed by 5 minutes incubation on ice, vortexing every minute. 800 µl Reagent C were added, and the mixture was carefully mixed by inverting the tubes. The samples were then centrifuged at 700 x G for 10 minutes at 4°C. The pellet was discarded, and the supernatant was centrifuged for 15 minutes at 4°C using 12,000 x G or 3,000 x G, to achieve a better mitochondrial yield. The supernatant was transferred and kept in new tubes, as this is the cytosolic fraction of the cells. The pellet containing the mitochondria was washed with 500 µl Reagent C and centrifuged at 12,000 x G for 5 minutes. The supernatant was discarded. Sample buffer containing 50mM DTT was added to both the cytosolic fraction and the mitochondrial pellet. The samples were boiled for 5 minutes at 95°C for gel electrophoresis. The tubes were stored at -20°C.

2.2.12. PierceTM BCA protein assay kits:

Standards and samples were prepared by diluting them in lysis buffer. Standards can be frozen at -20 °C for over a year. Standards (range: 25 – 2500 µg/µl) and samples were pipetted in duplicates. The working reagent was prepared by diluting solution B in solution A, ratio 1:50 (100 µl B in 5 mL A). Reagent B is light sensitive, so it was not left exposed to light long periods. 100µl working reagent were added to each plate, which was then incubated 30 minutes at 37 °C, covered in tin foil. The plate was read on FlexStation 3, and results acquired using SoftMax Pro 5.4.5.

2.2.13. SDS PAGE/Western blot:

SDS-PAGE:

15 µL of cell lysates were loaded onto 10% (v/v) acrylamide resolving gels, 1.5 mm in thickness. Gels were cast with running buffer and electrophoresis was performed at 120 V for 110 minutes at room temperature using a Bio-Rad Mini-PROTEAN™ tank. Protein separation was visualized using a PageRuler Plus Pre-stained protein ladder (Thermo Scientific) loaded alongside the samples.

Western Blotting:

Separated proteins were transferred to nitrocellulose membranes (0.45 µm pore size, GE Healthcare) using a Bio-Rad Mini Trans-Blot™ system at a constant current of 280 mA for 150 minutes at room temperature. Membranes were blocked for 2 hours in 3% (w/v) bovine serum albumin (BSA) prepared in TBS-T buffer to prevent non-specific binding. After blocking, membranes were incubated overnight at 4°C with the appropriate primary antibody (dilutions detailed in Table 2.2). The next day, membranes were washed three times with TBS-T (5 minutes each) and incubated with the corresponding HRP-conjugated secondary antibody for 2 hours at room temperature. Chemiluminescent detection was achieved using ECL-1 and ECL-2 substrates in a 1:1 ratio for 2 minutes. Membranes were exposed to UltraCruz® Blue X-ray film and developed using a JP-33 automatic film processor.

Membrane Stripping and Reprobing:

Membranes were stripped of antibodies by incubation in stripping buffer containing 14 mM β-mercaptoethanol at 60°C with gentle agitation (100 x G) for 15 minutes. Stripped membranes were washed three times with TBS-T (5 minutes each) to remove residual stripping solution and re-blocked with 3% (w/v) BSA in TBS-T for 2 hours. The membranes were subsequently re-incubated with primary antibodies specific to internal controls (e.g., GAPDH), and the same procedure was followed for secondary antibody incubation, detection, and development as described above.

Densitometry Analysis:

Developed autoradiography films were scanned at 600 dpi and saved as TIF files. Band intensities were analysed using ImageJ software. Bands of interest were quantified by selecting the region of interest (ROI) and subtracting the background signal from each measurement. The intensity of the bands for the target protein was normalized to the corresponding loading control, and the relative expression levels were calculated.

Table 2.2 – Antibodies used in western blotting analysis.

Antigen	Species	Dilution	Supplier	Cat.No.
PAR4	Rabbit	1:3000	Alomone	APR-034
Tubulin	Mouse	1:30K	Merck	05-829
HA	Mouse	1:10K	BioLegend	901516
GFP	Rat	1:5000	ChromoTek	3H9
Turbo GFP	Rabbit	1:20K	Origene	TA150071
RFP	Mouse	1:3000	ChromoTek	6G6
VDAC1/3	Mouse	1:10K	Abcam	ab14734
Anti-rabbit	Goat	1:7500	Jackson Lab	111-035-144
Anti-mouse	Goat	1:7500	Jackson Lab	115-035-003
Anti-rat	Goat	1:1000	R&D Systems	HAF005

2.2.14. Statistical analyses:

For SILAC proteomics, p-values were calculated using paired t-tests to compare protein abundance between experimental groups. For colocalization analyses, the intensity of pixel signals was quantified, and Pearson's correlation coefficient (r) was used to assess the degree of colocalization between fluorescent signals. To compare multiple groups, data were analysed using one-way analysis of variance (ANOVA), followed by Tukey's Honest Significant Difference (HSD) post hoc test. This approach allowed pairwise comparisons between all groups while controlling for type I error rates, ensuring statistically robust conclusions.

2.2.15. Software used:

Table 2.3 Software and application used throughout the thesis.

Software	**Application**
GraphPad Prism 8	Statistical analysis of colocalization
ImageJ v1.54	Western blot quantification Confocal image composition Colocalization analysis
Mendeley v.1.19.8	Reference Manager
Microsoft PowerPoint	Figure generation
R v.4.4.2 + RStudio	SILAC data analysis
Cytoscape v3.10.3 (STRING plugin)	Protein interaction network analysis
AlphaFold 2	Protein structure prediction
PrePPI	Database of predicted protein interactions
ChatGPT 4o	Formatting text Generating data analysis scripts

2.2.16. Data analysis:

Code scripts and Cytoscape files used to perform data filtering, network analyses and Gene Ontology Enrichments are publicly available at:

https://github.com/bonfamarco/SILAC-proteomics-data-analysis-pipeline/tree/main

3. CHARACTERISATION OF PAR4 INTERACTIONS

3.1. INTRODUCTION

3.1.1. PDZ proteins and recognition motifs on GPCRs:

PDZ proteins are a family of scaffold proteins that are widely distributed in various tissues and play important roles in cell signalling and organization[200]. They owe this name to the first three proteins characterised with this conserved motif (PSD-95, Dlg, and ZO-1)[201]. PDZ domains, which are approximately 90 amino acids in length, consist of six β-sheets situated between two α-helices and are highly conserved regions. Within PDZ domains, there is a centrally located globular binding groove positioned between the α2-helix and β2-sheet. This region contains a set of conserved amino acid residues collectively known as the "carboxylate-binding loop" and is characterized by the following sequence: R/L-X-X-X-G-L-G-F[202]. This carboxylate-binding loop enables PDZ domains to engage with the carboxyl tails of other proteins. Furthermore, PDZ domains can interact with internal motifs that form β-sheet extensions resembling carboxyl tail characteristics, and these extensions insert into the binding groove. The canonical PDZ domains display notable sequence diversity, which results in a broad spectrum of protein interactions. As a result, PDZ proteins are categorised in three classes, based on the carboxyl tail-binding motifs they recognize.

Table 3.1 – Classes of PDZ proteins based on their binding motifs[203].

Class / Type / Position(P)	I	II	III	IV
P^0		Φ/ ψ	Φ/ ψ	Negative residue/ #***
P^{-1}	S/T	Mostly Trp or Asp	Any	ψ
P^{-2}	X	Φ/ ψ	Negative residue	Any
P^{-3}	Φ	Mostly Glu	Gly or Glu	Any

This diversity allows PDZ proteins to interact with a variety of C-terminal ends, including the ones belonging to receptors, channels, and enzymes[204]. PDZ proteins function in the clustering, localisation, and trafficking of their binding partners, as well as in the assembly of signalling complexes and the regulation of signal transduction[205]. Dysregulation of PDZ proteins has been implicated in various diseases, such as cancer, neurological disorders, and infectious diseases, making them attractive targets for drug discovery[206]. Table S1 in the supplementary materials reports all the GPCR-PDZ interactions already known in the literature.

3.1.2. Impact of PAR4 Short Linear Motif (SLiM) and Mutations:

At the extreme end of the C-tail of PAR4, a short linear motif (SLiM – $S^{381}SLL^{384}$) resemble a PDZ-interacting domain found also in other GPCRs (Figure 3.1). Moreover, as outlined in chapter 1 paragraph 1.7, PAR4 harbours several mutations that may result in different clinical outcomes. The four most common mutations are A120T, Y157C, F296V and P310L, which are also highlighted in Figure 3.1. According to the Ensembl database Y157C is the most deleterious variant, yet it is not well characterized. The cysteine mutation causes aberrant anterograde receptor trafficking, reducing the number of receptors at the membrane, thus causing lower reactivity to activating peptides and thrombin[193]. Although the impact of this mutation

on PAR4 localisation has been determined, the effect on downstream signalling is still unknown. SILAC proteomics serves as an optimal tool to investigate whether any PDZ protein can be found at the proposed binding site, as well as elucidate how PAR4 signalling is affected by single point mutations.

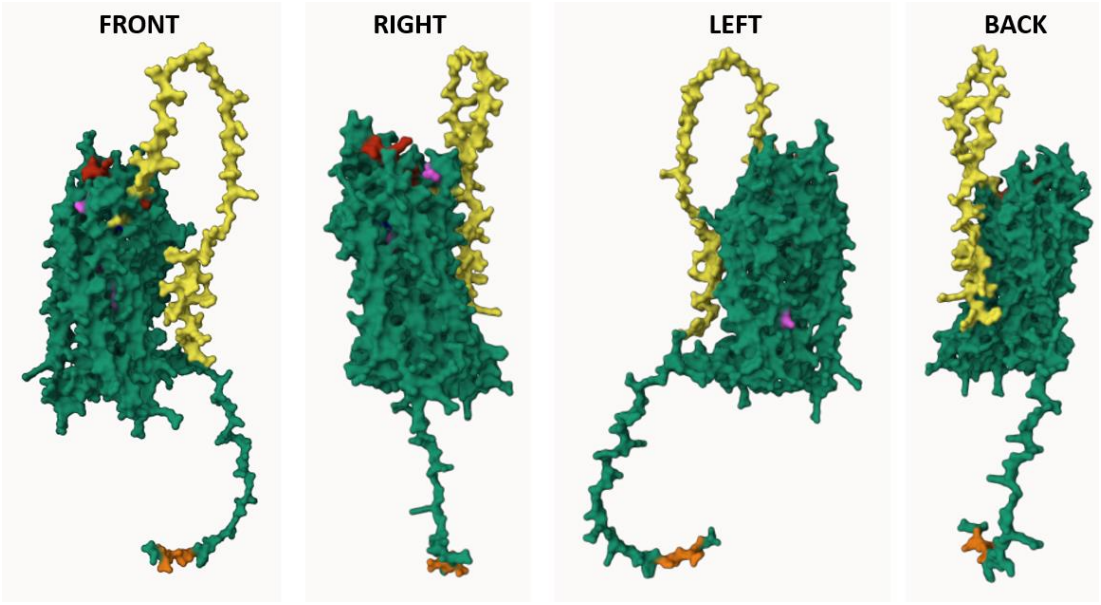


Figure 3.1 – Molecular surface animation of PAR4. Regions of interest are highlighted with different colours. The receptor presents the common GPCR structure, spanning the membrane seven times. The cleaved peptide on the N-terminal (**M^{1}…R^{47}**) is coloured in yellow. The tethered ligand (**G^{48}YPGQV^{53}**) is depicted in red. The binding region buried in the ECL2 (**C^{228}HD^{230}**) is blue. Deleterious mutations (**A^{120}, P^{310}, Y^{157}, F^{296}**) are depicted in pink. At the far end of its C-terminal a short linear motif (**S^{381} SLL**Q^{385}) is a proposed PDZ-binding motif, highlighted in orange. Structural prediction obtained from the AlphaFold database (https://alphafold.ebi.ac.uk/entry/Q96RI0) modified in Mol* (https://molstar.org/viewer/).

3.1.3. SILAC proteomics as a tool to study GPCRs:

SILAC stands for Stable Isotope Labelling of Amino acids in Cell culture and the principles of this technique have already been outlined in the Methods chapter 2.2.6. Briefly, cells are metabolically labelled by growing them in media containing different amino acids isotopes. The cells are then lysed, mixed between different growing conditions, and immunoprecipitated. Peptide pairs can then be detected and quantified based on their relative different isotope composition, which results in a mass difference on the spectrometer[207]. The advantages of SILAC, such as its straightforward implementation, quantitative accuracy, and reproducibility over chemical labelling or label-free quantification strategies, make it a favoured choice for proteomic research. Moreover, this method allows for flexibility depending on the overall aim of the experiment[208]. Before running the mass spectrometry, enrichment of post-translational modified peptides allows to investigate PTMomics, organelle prefractionation is used to investigate sub-proteomics, or in the case this experiment, immunoprecipitation is used to study interaction proteomics[208].

Affinity purification mass spectrometry is increasingly being used to elucidate receptor signalling and regulation, as well as allowing to characterise GPCR structures. In this regard, SILAC is an extremely powerful tool in quantitative proteomics, demonstrating a large range of applications. This method has already been widely employed in the GPCR field[209,210]. For example, SILAC allowed to differentiate agonist-selective phosphorylation and endocytosis of the β-2 adrenoceptor[211], which was also found to interact with the PDZ-protein sorting nexin 27 (SNX27) to regulate endocytic sorting of the receptor[212]. SILAC quantitative proteomics can have other applications besides interaction studies. For instance, it helped identifying key residues for the phosphorylation of the μ opioid receptor[213].

3.2. CHAPTER AIM:

The primary objective of this chapter is to employ Stable Isotope Labelling by Amino acids in Cell culture (SILAC) proteomics to identify and characterize the interaction partners of the protease-activated receptor 4 (PAR4). This study aims to elucidate whether the PDZ domain-containing proteins interact with PAR4 via its C-terminal PDZ-binding motif (ΔSLiM). Understanding these interactions is crucial as they can provide insights into the receptor's signalling mechanisms and its role in cellular functions.

Additionally, this chapter investigates the impact of a specific point mutation (Y157C) on the PAR4 interactome. By comparing the proteomic profiles of wild-type and Y157C mutant PAR4, this analysis seeks to uncover how this mutation alters PAR4's protein interactions and potentially its functional properties. This comparative approach is designed to reveal the molecular consequences of the Y157C mutation and its implications for PAR4's activity and regulatory mechanisms.

Objectives of this Chapter:

1. To employ SILAC proteomics to identify and characterize PAR4 interaction partners in a cellular context.
2. To determine whether PDZ domain-containing proteins interact with PAR4 via its C-terminal PDZ-binding motif (ΔSLiM).
3. To compare the proteomic profiles of wild-type and Y157C mutant PAR4 to assess the impact of the Y157C mutation on PAR4’s interactome.
4. To elucidate the potential functional consequences of the Y157C mutation on PAR4 signalling and regulatory mechanisms.

3.3. RESULTS:

The results presented in this section are derived from SILAC proteomics, a technique that enables the quantitative comparison of protein abundances across experimental conditions. SILAC ratios correspond to the relative abundance of proteins in each condition, normalized by isotopic labelling with light, medium, or heavy amino acids. A SILAC ratio greater than 1 indicates an enrichment of the protein in the experimental condition compared to the control, while a ratio less than 1 suggests reduced abundance. By analysing these ratios, potential interaction partners of PAR4, as well as proteins influenced by specific modifications such as the ΔSLiM deletion or Y157C mutation, can be identified and characterized. These results are interpreted through statistical analyses and visualized using volcano plots and protein-protein interaction networks, providing insights into PAR4's interactome and its functional implications.

3.3.1. Proteome filtering and characterisation:

The raw datasets obtained were uploaded on Pure and can be accessed at the following DOI: https://doi.org/10.15129/6fa95c9e-2188-46f4-92d4-705c0c96bdfd.

A description of the headlines of the file is presented in Table 3.2.

Table 3.2 – Standard column headings of the raw datasets

Column Heading	Description
Accession	Displays the accession number for the sequence
Description	The name of the protein
Score	The total score of a protein (which represents the sum of the individual peptide scores). The exact score required for significance will vary between experiments. A MS facility will usually apply a 5% false discovery rate cutoff.
Coverage	Proportion of the protein sequence covered by the identified peptides
♯unique peptides	Total number of unique peptides identified for a protein
♯peptides	Total number of peptides identified for a protein
♯PSM	Peptide spectral match
Ratio (Heavy/Medium/Light)	The relative intensity of peptides in a named labelled sample, compared to a second labelled sample
Ratio Count	The number of peptide ratios that were used to calculate a given protein ratio
Ratio variability (%)	The variability of the individual peptide ratios used to calculate a given protein ratio
♯AAs	The length of a protein in amino acids
MW (Da)	The molecular weight of a protein in Daltons. Excludes modifications
calc. PI	The theoretical isoelectric point of a protein

Proteome filtering was performed in accordance to previously described methods[196], using the tidyverse package on R studio running R 4.4.2. Briefly, entries lacking more than one peptide or unquantified were removed. SILAC ratios were then converted to log2 fold change values to make the interacting proteins fit into a normal distribution (Figure 3.2). The mean and standard deviation of the log2 values for each SILAC ratio was used to calculate the thresholds, which are reported in Tables 3.3 - 3.4. These values were used to filter only interactions with at least a 95% confidence limit.

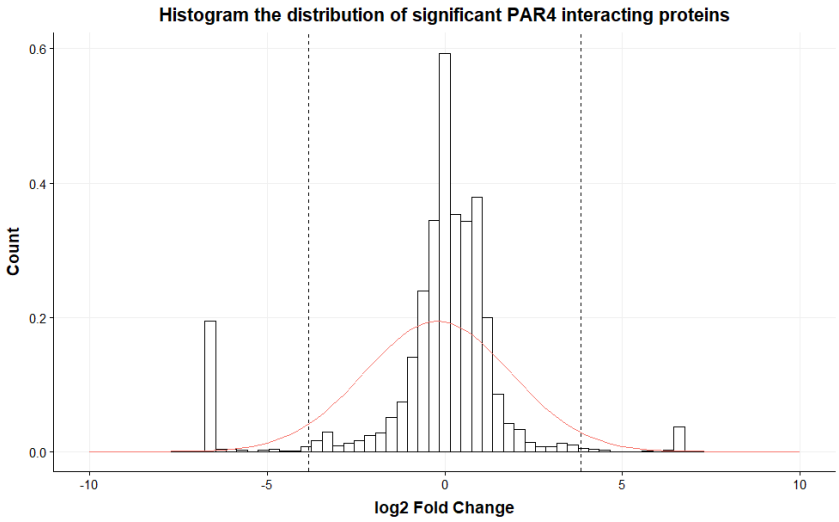


Figure 3.2 – Histogram of the distribution PAR4 interacting proteins. Histograms were generated for each SILAC ratio of each dataset. The experiment was repeated five times, and each run had three SILAC ratios (Heavy/Light, Medium/Light, Medium/Heavy), resulted in a total of 15 histograms. In the figure is represented the graph for the Heavy/Light ratio of dataset number 1.

Table 3.3 – Threshold for the SLiM proteomics

Dataset nr	Heavy/Light	Medium/Light	Medium/Heavy
1	3.827	5.039	3.588
2	3.272	5.613	4.601
3	3.241	4.665	3.688
4	3.964	4.415	2.598
5	3.841	5.173	3.463

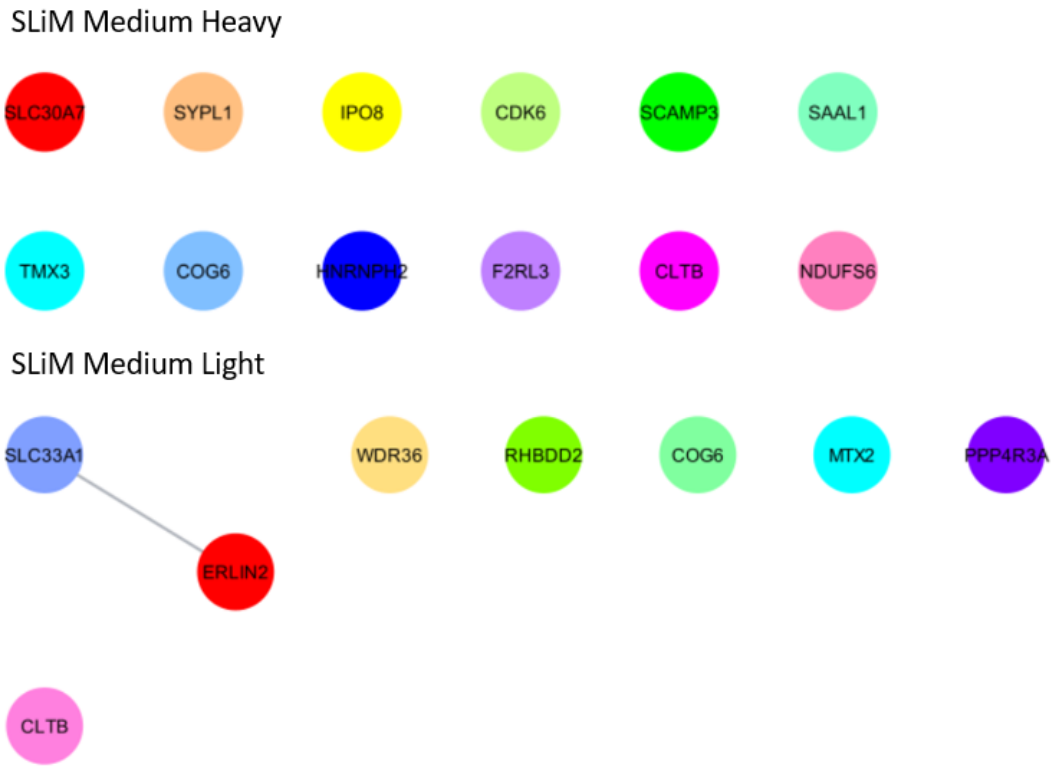
Table 3.4 – Threshold for the Y157C proteomics

Dataset nr	Heavy/Light	Medium/Light	Medium/Heavy
1	3.002	3.658	2.570
2	2.428	3.820	3.296
3	3.841	5.173	3.463

The datasets were then merged, and proteins present in triplicate (for SLiM proteomics) or duplicate (for Y157C) proteomics were kept. The SILAC ratios of the remaining hits were then averaged, and three datasets generated: Heavy/Light, Medium/Light, and Medium/Heavy. Due to the stringent criteria used to filter

significant hits, very few interactions were found. Each table is reported in the supplementary figures (Tables S2-S7).

The next step in characterising these confident hits would be to perform Functional Enrichment analysis by loading them on Cytoscape, with the STRING extension. However, the “Accession” entry is an identifying code in the UniProt database, which might refer to a peptide or just a protein fragment. For this reason, not every entry mapped to a protein on Cytoscape, this resulted in even fewer entries, thus making it impossible to generate a protein network and perform Functional Enrichment analysis. In the case of SLiM Heavy/Light dataset, the only two interactions left did not map to any protein on Cytoscape and therefore no network was generated. The remaining five protein networks for both the SLiM and Y157C proteome are reported in figure 3.3. Since these data could not be analysed, a different data analysis workflow was applied.



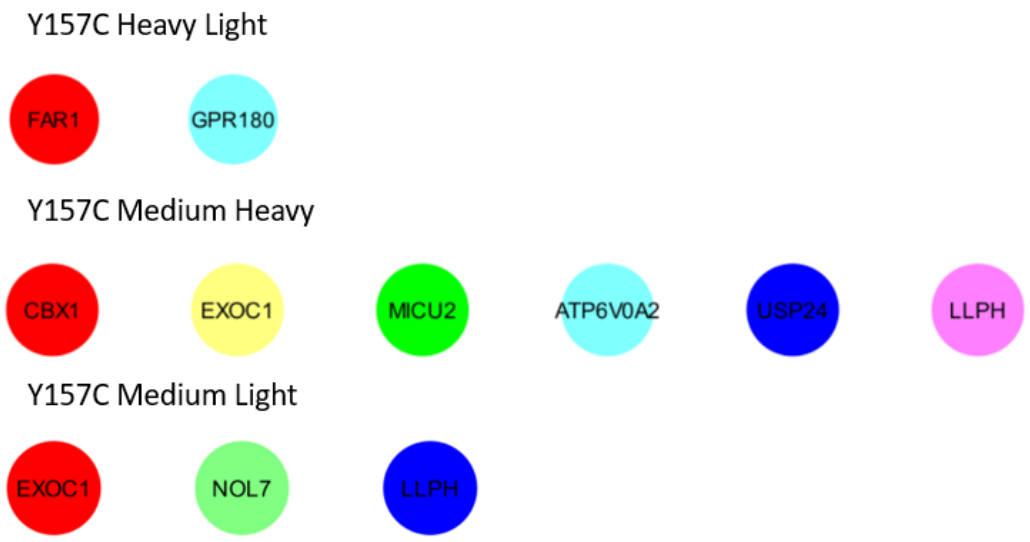


Figure 3.3 – The STRING app on Cytoscape does not map every Accession entry to a protein. It was therefore not possible to generate protein networks and perform functional enrichment analysis.

3.3.2. A novel approach to SILAC data analysis:

In the field of proteomics, the sensitivity and specificity of data analysis methods critically determine the comprehensiveness of detected protein interactions. While existing methods provide valuable insights, they often yield a limited number of hits, potentially overlooking significant interactions. This observation prompted the development of a novel data analysis method designed to enhance the detection capability and improve the accuracy of identifications. This new approach aimed at expanding the proteomic landscape accessible for study, thereby enabling a more detailed exploration of complex biological networks. First, the raw datasets were merged, proteins missing a SILAC ratio (unquantified) or lacking more than one unique peptide were removed. Then, entries present at least in triplicates (for SLiM proteomics) or duplicate (for Y157C proteomics) were kept, filtering out the rest. To obtain p-values to generate a volcano plot, paired t-test was performed on the SILAC ratios. To obtain the Heavy/Medium SILAC ratio, Medium/Heavy was simply multiplied by 1/ratio. Then Medium/Heavy was compared to Medium/Light, while Heavy/Medium was compared to Heavy/Light. This was done to investigate any differences between PAR4 (medium) or SLiM (heavy) when compared to YFP (light), which was used as a control. For each protein, the average SILAC ratio was then calculated, and any repetition was dropped. A diagram of the analysis workflow is provided in Figure 3.4.

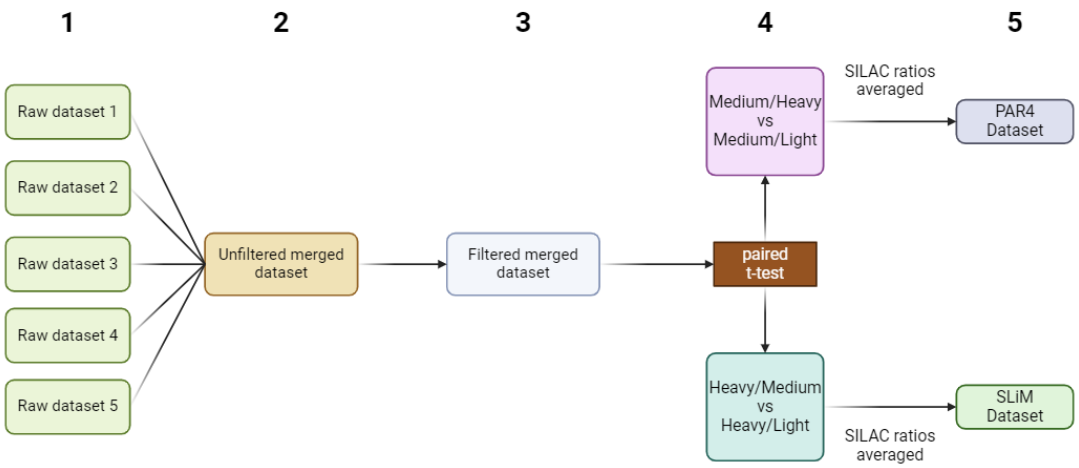


Figure 3.4 – Workflow of the data cleaning of the SILAC proteomics. This diagram illustrates the workflow for analysing proteomics datasets to enhance detection sensitivity and accuracy. **(1)** Raw datasets were merged into **(2)** an unfiltered dataset, which underwent **(3)** filtering to retain only proteins quantified in at least duplicate (Y157C proteomics) or triplicate (SLiM proteomics) with a SILAC ratio and more than one unique peptide. **(4)** SILAC ratios were subjected to paired t-tests, comparing Medium/Heavy versus Medium/Light for PAR4 analysis and Heavy/Medium versus Heavy/Light for SLiM analysis. **(5)** Average SILAC ratios were calculated, repetitions were removed, and datasets for PAR4 and SLiM were compiled. This workflow facilitates comprehensive identification of protein interactions across the proteomic landscape.

SILAC ratios then were converted to $\log_2$ values, and fitted a Gaussian distribution centred around a $\log_2$ SILAC ratio of 0. The distribution of the proteins for each experiment is shown in figure 3.5.

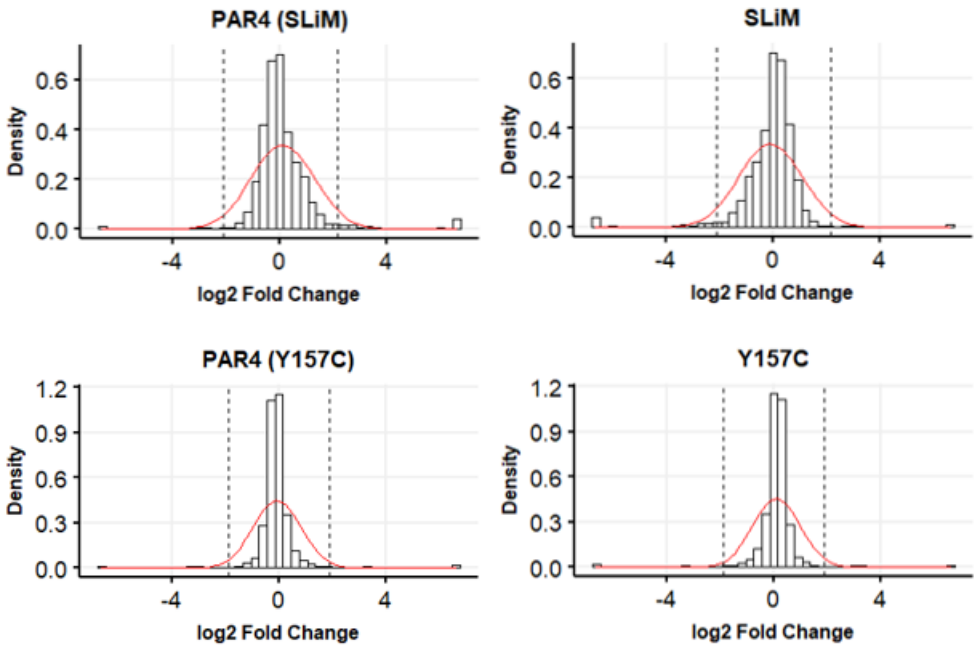


Figure 3.5 – Representative histograms of the unfiltered dataset. The data follow a gaussian distribution. The 1.96 standard deviation cut-off is marked with a dashed line.

The mean and standard deviation of the $\log_2$ SILAC ratio were used to calculate thresholds for significance, by adding 1.96 standard deviations to the mean. Table 3.5 shows the thresholds for each dataset. These values placed the threshold at a 95% confidence limit. Protein interactions were considered significant when achieving a p-value ≤ 0.05, which translated to a $-\log_{10}$ p-value ≥ 1.3. On the volcano plots (Figure 3.6), data points found above this line are considered significant, while down- and up-regulated proteins are found respectively on the left and on the right of the $\log_2$ threshold lines. A summary of these proteins and their role is given in Table 3.6.

Table 3.5 – Thresholds generated using the new data analysis workflow.

Dataset	Threshold
PAR4 (SLiM)	2.433
SLiM	2.131
PAR4 (Y157C)	1.751
Y157C	1.878

3.3.3. Venn diagram and Volcano Plots:

Across the five datasets 2,290 unique peptides were identified by LC MS/MS. After applying the p-value threshold, 472 peptides were found to be significantly interacting with PAR4, while 471 were significant interactions with SLiM. The two groups shared 224 interactions, leaving 248 proteins being unique to the PAR4 group and 247 unique SLiM interacting proteins. Concerning the Y157C database, 181 total proteins belonged to the PAR4 group and 312 to Y157C, however 52 were found to be in common, meaning 129 were unique to PAR4 and 260 to Y157C. This repartition is shown in the Venn diagrams of figure 3.6.

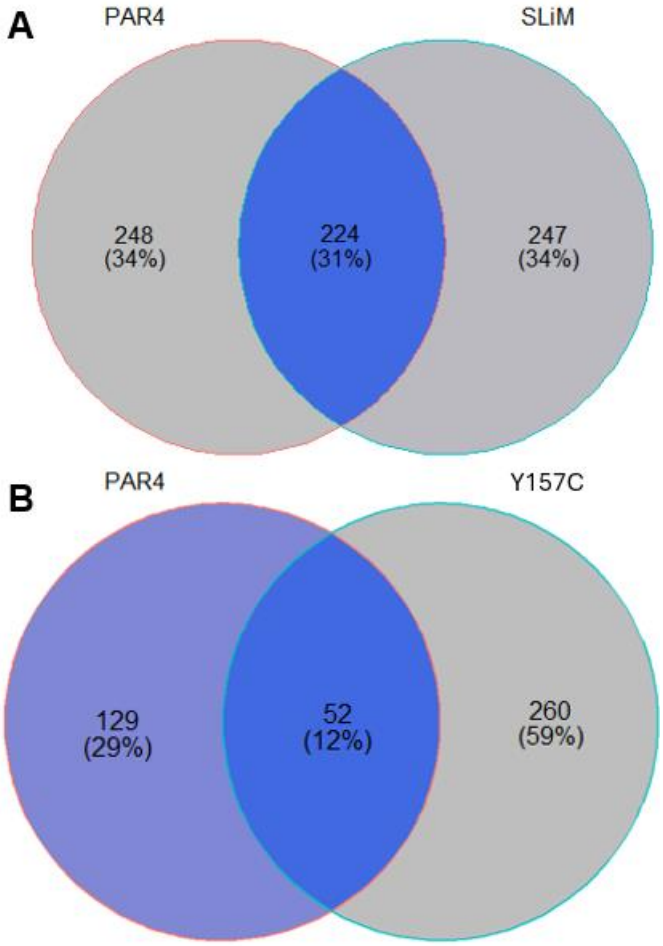


Figure 3.6 – Venn diagrams of the significant interacting proteins. A: After filtering the interacting proteins with a p-value > 0.05, 472 significant proteins were found in the PAR4 proteome and 471 in the SLiM proteome. 224 interactions were common between the two datasets. **B:** For the Y157C experiment, the PAR4 group was formed by 181 proteins in total while Y157C was made of 312 hits. 52 interactions were common to the two datasets, therefore 129 proteins were unique to PAR4 and 260 to Y157C.

The datasets utilized accession numbers referencing the UniProt database, leading to some peptides corresponding to protein fractions without an associated gene symbol. Consequently, not all datapoints in the volcano plots (Figure 3.7) display a gene name. A comprehensive summary of all proteins identified as up- or downregulated by the volcano plot analysis is detailed in Table S8 in the appendix.

In the SLiM experiment, a greater variability in protein regulation was observed, with a significantly higher number of both up- and downregulated proteins. This variability can be attributed to the inclusion of two additional replicates in this experiment, enhancing the detection of protein interactions. For example, in the PAR4 proteome of the SLiM experiment, 11 proteins were found upregulated and 6 downregulated. In contrast, in the Y157C experiment, only 2 proteins were upregulated and 1 downregulated for the same group. Similarly, the SLiM proteome showed 6 upregulated and 16 downregulated proteins, compared to the Y157C group's 1 upregulated and 3 downregulated proteins. Notably, among the proteins that could be mapped to a gene name, PDZ and LIM domain protein 1 was found upregulated in the PAR4 Y157C dataset. Distinctly, Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3 (HACD3) was upregulated in the PAR4 proteome but downregulated in the SLiM group, whereas Ubiquitin-fold modifier-conjugating enzyme 1 (UFC1) exhibited opposite regulation patterns, being downregulated in PAR4 and upregulated in SLiM. A similar pattern was observed with LAGE3, which was upregulated in the PAR4 proteome and downregulated in the Y157C group.

Out of the 16 downregulated proteins in the SLiM group, 6 were mitochondrial, including ATP5PD, COA3, HADHA, HADHB, SLC25A3, and TMX3. Interestingly, no mitochondrial proteins were found upregulated in the SLiM proteome or downregulated in the PAR4 group. However, OXA1L was significantly upregulated in the PAR4 proteome of the SLiM experiment. Although fewer hits were found in the Y157C experiment, those identified were not markedly dysregulated. Conversely, the SLiM experiment revealed significant proteins with more extreme changes, such as Prostaglandin reductase 3 (PTGR3), which was notably upregulated in the PAR4 proteome, and Aspartyl/asparaginyl beta-hydroxylase (ASPH), which was the most downregulated protein in the SLiM group.

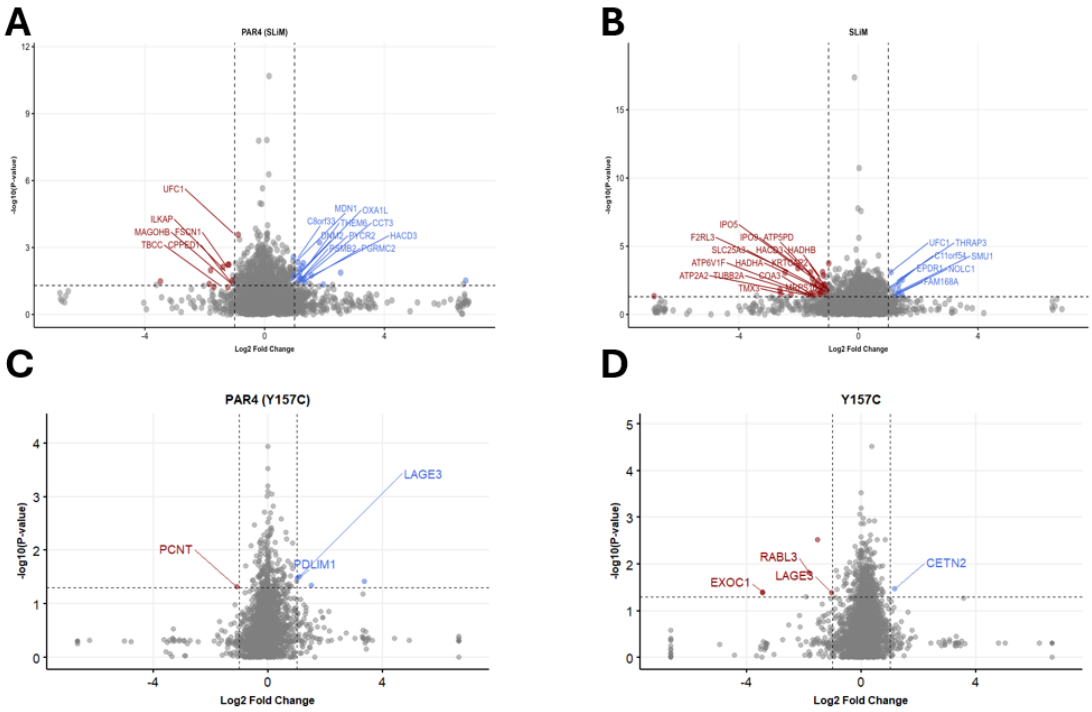


Figure 3.7 – Volcano plots of the proteins interacting with PAR4, SLiM or Y157C. Proteins found above the threshold on the p-value axis are considered significant interactions, while the thresholds on the $\log_2$ SILAC ratio define down- and up-regulated proteins. Gene names are shown for the top 20 most significant proteins, however not every UniProt accession number could be mapped to a gene name. **A:** PAR4 (SLiM proteome) **B:** SLiM **C:** PAR4 (Y157C proteome) **D:** Y157C.

3.3.4. Characterisation of PDZ interacting proteins

Initially, to investigate the presence of PDZ proteins within any proteome, the biomaRt library on R was used, to validate whether any Accession entry on the filtered databases would match the Pfam database. This database is a comprehensive collection of protein families and domains. Developed and maintained by the Wellcome Trust Sanger Institute, Pfam provides a systematic classification of protein sequences into families, which are groups of evolutionarily related proteins sharing a common ancestry[214]. The pfam database is however now out of date, and it is hosted by another database, InterPro. This is another resource that integrates information from multiple protein databases to provide a comprehensive and unified view of protein families, domains, and functional sites. It brings together data from well-known databases such as Pfam, PROSITE, PRINTS, SMART, and others. The primary goal of InterPro is to enhance the accuracy and reliability of protein functional annotations by combining evidence from different sources. InterPro uses a signature or profile-based approach to classify proteins into families and predict the presence of specific domains or motifs within sequences[215].

However, neither Pfam nor InterPro initially found any PDZ proteins within any of the four PAR4 proteomes. This can probably be owed to the databases being outdated and not constantly updated. To address this issue, the keyword “PDZ” was used in the UniProt database search toolbar, resulting in 1,227 entries for the Homo sapiens genome. These results were downloaded and compared to the accession entries of each filtered proteome. The analysis revealed several PDZ proteins across different proteomes. In the PAR4 proteome of the SLiM group, proteins such as Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12 (GNG12), Myosin-10 (MYH10), Tight junction protein 2 (TJP2), and LIM domain only protein 7 (LMO7) were identified. In the SLiM proteome, GNG12 and ADP-ribosylation factor 3 (ARF3) were found. For the Y157C mutant proteome, GIPC PDZ domain containing family member 1 (GIPC1) and PDZ and LIM domain protein 1 (PDLIM1) were identified, while in the PAR4 proteome of the Y157C experiment, GIPC1 and Vacuolar protein sorting-associated protein 26A (VPS26A) were detected. Interestingly, GNG12 and GIPC1 were flagged in different proteomes, indicating a high confidence of interaction with PAR4. However, since the UniProt database includes any entry that may interact with a PDZ protein but may not possess a PDZ domain itself, each entry was rechecked against the UniProt database to confirm the presence of a PDZ domain. After this verification, only TJP2, LMO7, GIPC1, and PDLIM1 were confirmed to have a PDZ domain in their structure and were therefore selected for further analysis. As expected, the only dataset to not have PDZ proteins was the SLiM one, as it lacked the necessary proposed binding domain. Every protein identified for each protein is presented alongside its corresponding p value and log2 fold change in tables 3.6-3.9.

Table 3.6 PDZ proteins identified in the PAR4 SLiM dataset.

Accession	Description	pvalue	log2fc	Gene
Q9UBI6	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12 OS=Homo sapiens GN=GNG12 PE=1 SV=3 - [GBG12_HUMAN]	0.036842	0.427402	GNG12
P35580	Myosin-10 OS=Homo sapiens GN=MYH10 PE=1 SV=3 - [MYH10_HUMAN]	0.007049	-0.64408	MYH10
B7Z2R3	cDNA FLJ59510, highly similar to Homo sapiens tight junction protein 2 (TJP2), transcript variant 2, mRNA OS=Homo sapiens PE=2 SV=1 - [B7Z2R3_HUMAN]	0.016947	-0.1641	TJP2
E9PMP7	LIM domain only protein 7 (Fragment) OS=Homo sapiens GN=LMO7 PE=1 SV=7 - [E9PMP7_HUMAN]	0.002317	-0.40789	LMO7

Table 3.7 PDZ proteins identified in the SLiM dataset.

Accession	Description	pvalue	log2fc	Gene
P61204	ADP-ribosylation factor 3 OS=Homo sapiens GN=ARF3 PE=1 SV=2 - [ARF3_HUMAN]	0.019063	0.221213	ARF3
Q9UBI6	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12 OS=Homo sapiens GN=GNG12 PE=1 SV=3 - [GBG12_HUMAN]	0.031291	-0.4274	GNG12

Table 3.8 PDZ proteins identified in the PAR4 Y157C dataset.

Accession	Description	pvalue	log2fc	Gene
A8K2I7	cDNA FLJ76072, highly similar to Homo sapiens GIPC PDZ domain containing family, member 1 (GIPC1), transcript variant 1, mRNA OS=Homo sapiens PE=2 SV=1 - [A8K2I7_HUMAN]	0.03469	0	GIPC1
O00151	PDZ and LIM domain protein 1 OS=Homo sapiens GN=PDLIM1 PE=1 SV=4 - [PDLI1_HUMAN]	0.031255	1.09611	PDLIM1

Table 3.9 PDZ proteins identified in the Y157C dataset.

Accession	Description	pvalue	log2fc	Gene
A8K2I7	cDNA FLJ76072, highly similar to Homo sapiens GIPC PDZ domain containing family, member 1 (GIPC1), transcript variant 1, mRNA OS=Homo sapiens PE=2 SV=1 - [A8K2I7_HUMAN]	0.048202	0	GIPC1
O75436	Vacuolar protein sorting-associated protein 26A OS=Homo sapiens GN=VPS26A PE=1 SV=2 - [VP26A_HUMAN]	0.020783	-0.05799	VPS26A

3.3.5. Functional enrichment using STRING:

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is an expansive, precomputed database designed to provide a comprehensive overview and analysis of protein-protein interactions (PPIs), both direct (physical) and indirect (functional) associations. STRING proves to be an extremely powerful tool in research not only for network visualisation, but as it comprehensively integrates data from other external sources. Experimental and biochemical data are extracted from the Biomolecular Interaction Network Database (BIND)[216], the Database of Interacting Proteins (DIP)[217], the Biological General Repository for Interaction Datasets (BioGRID)[218], the Human Protein Reference Database (HPRD)[219], IntAct[220], the Molecular INTeraction database (MINT)[221] and the Protein Database (PDB)[56]; while curated data are obtained from Biocarta[222], BioCyc[223], Gene Ontology (GO)[224], Kyoto Encyclopedia of Genes and Genomes (KEGG)[225] and Reactome[226]. Moreover, STRING is based off automated text mining algorithms, which allow it to scan the scientific literature present in databases such as full-text articles from the PMC Open Access Subset (up to April 2022), PubMed abstracts (up to August 2022), as well as summary texts from Online Mendelian Inheritance in Man (OMIM)[227] and Saccharomyces Genome Database (SGD)[228].

All the data obtained from these different sources are critically assessed and scored. The combined score is computed by combining the probabilities from the different evidence channels and corrected for the probability of randomly observing an interaction using a Bayesian approach[229]. The version used to perform the network analysis of the two proteomics datasets was STRING 12.0, which encompasses 12,535 organisms and 59.3 million proteins for a total of over 20 billion interactions[230]. STRING has a supporting app that works with Cytoscape[231], a popular software platform for visualizing complex networks. This app allows to get detailed information about the functions of proteins within the Cytoscape software. Furthermore, it visually displays these details on the network diagram. For example, it is possible to colour-code different points (or 'nodes') on the diagram to show whether they are involved in specific biological processes, pathway, or reaction. This annotated colour-coding is dependent on the database from which they are retrieved (e.g. KEGG, Reactome, GO, WikiPathways, UniProt Keywords). The STRING app running on Cytoscape is

therefore a convenient yet comprehensive tool to perform functional enrichment on the four proteomes generated.

The STRING database and Cytoscape app were therefore used to perform a comprehensive network analysis of the four proteomes generated in this study. The analysis revealed several enriched terms across the datasets, emphasizing key biological processes and molecular functions. The most consistently enriched term across all datasets was acetylation, indicating its widespread occurrence as a post-translational modification. For both the wild-type PAR4 and the SLiM variant proteomes, terms related to translation and gene expression, such as cytoplasmic translation, ribonucleoprotein complex biogenesis, and RNA binding, were prominently enriched. This underscores the significant involvement of these proteins in ribosomal and translational functions, as corroborated by various databases like GO, Reactome, KEGG, and UniProt. In the Y157C mutant proteome, there was a distinct shift towards terms associated with metabolic processes and energy production, including cellular metabolic process, primary metabolic process, and respiratory electron transport chain. This suggests a broader involvement in energy metabolism for the Y157C variant. Additionally, RNA metabolic processes and RNA splicing terms were notably enriched in the Y157C dataset, differentiating it further from the wild-type PAR4 proteome. Enrichment analysis also highlighted mitochondrial-related terms in the PAR4 proteome, including mitochondrial inner membrane and mitochondrial respirasome, while the Y157C proteome showed higher enrichment for nuclear and spliceosomal components. This indicates potential differential localization and functional roles for the wild-type and mutant forms of PAR4. Figures 3.8 - 3.9 visually represent the top 20 enriched terms across all categories, providing a detailed overview of the functional enrichment analysis.

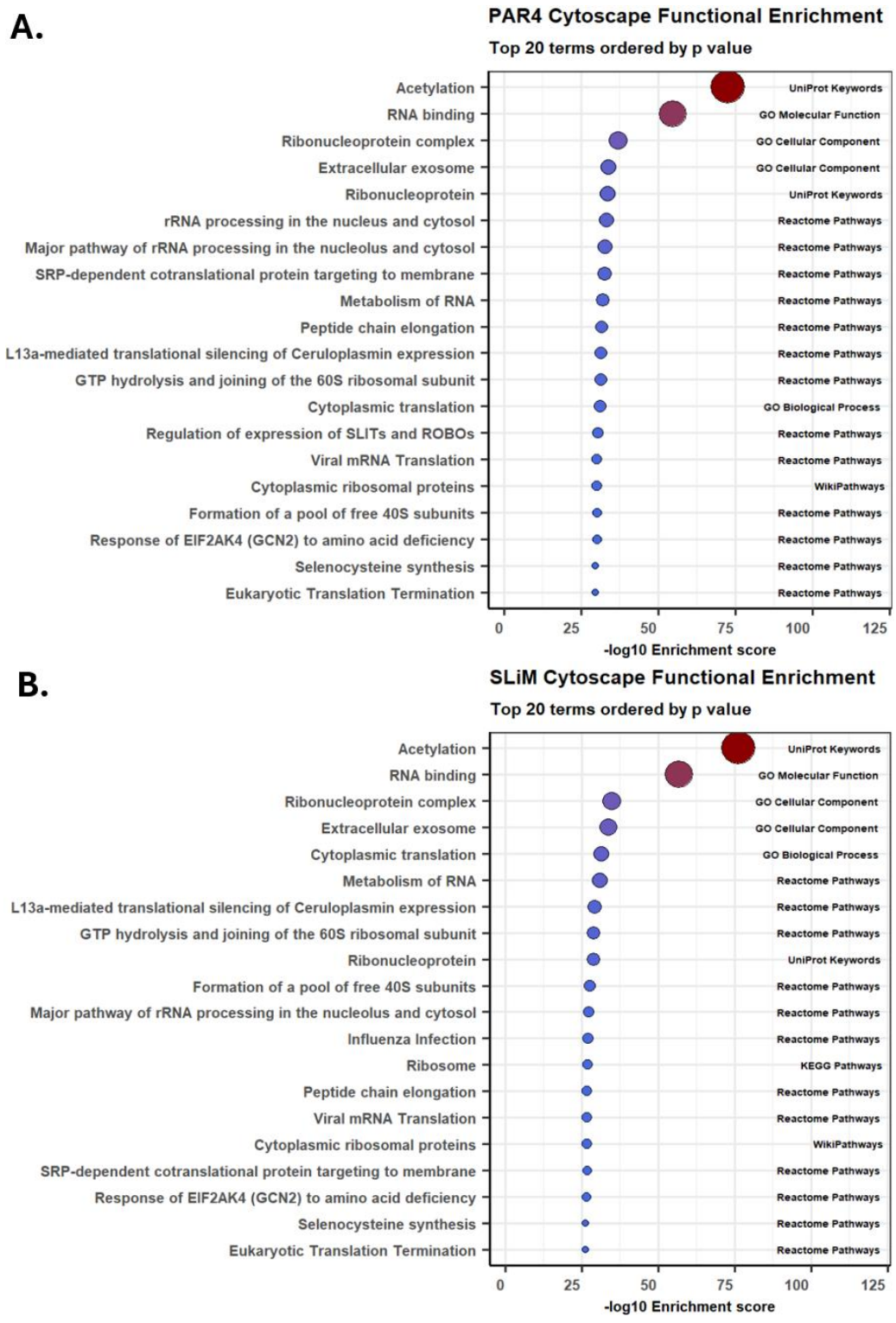


Figure 3.8 – Cytoscape functional enrichment for the SLiM proteomes. The enriched terms are shown on the left, the more enriched the bigger and redder would be the representation on the graph. On the right it is reported to which category each term belonged to. **A:** PAR4 proteome, **B:** SLiM proteome.

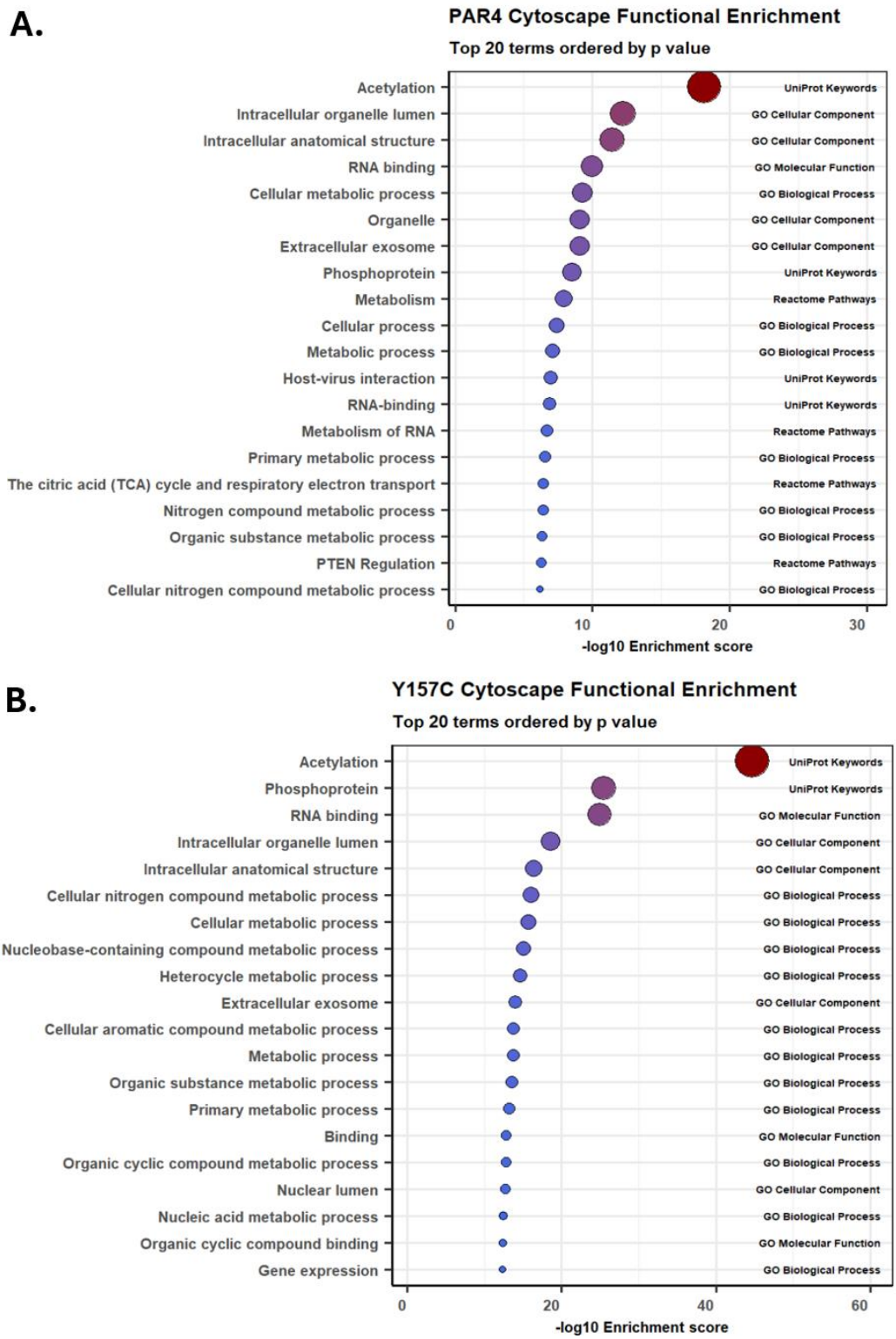


Figure 3.9 – Cytoscape functional enrichment for the Y157C proteomes. The enriched terms are shown on the left, the more enriched the bigger and redder would be the representation on the graph. On the right it is reported to which category each term belonged to. **A:** PAR4 proteome, **B:** Y157C proteome.

3.3.6. Clustering suggests interactions with mitochondrial proteins:

Importantly, the STRING extension in Cytoscape allows the user to perform network clustering using the Markov Clustering Algorithm[232]. The Markov Cluster Algorithm (MCL) is a popular algorithm to analyse protein-protein interaction (PPI) networks, where data points (such as proteins or genes) are represented as nodes, and the relationships between them (such as interactions or similarities) are represented as edges. First, the algorithm treats the input graph as a stochastic matrix - a mathematical representation of a graph where each edge is assigned a certain probability. If the graph is not weighted, all edges have an equal probability. The algorithm then expands, meaning that it raises the stochastic matrix to a power (usually 2), which has the effect of increasing the probability of the strong edges and weakening the weak ones. In terms of the graph, this corresponds to two nodes becoming more closely linked if they share many common neighbours. This specific process is called expansion. The next step is referred to as inflation, where the algorithm adjusts the probabilities in the stochastic matrix using the inflation parameter. Probabilities are raised to the power of the inflation parameter and then the matrix is rescaled to make sure it remains stochastic (i.e., the probabilities still add up to 1). A higher inflation value increases the difference between the high-probability and low-probability edges, which results in more distinct clusters. Conversely, a lower inflation value makes the clusters more like each other. The expansion and inflation steps are repeated until the matrix stops changing significantly from one iteration to the next, which means the algorithm has converged. The result is a stochastic matrix that represents the clusters in the original graph.

The large PPI networks were clustered using the built-in MCL algorithm of STRINGapp, the inflation parameter was left at 4, which was the preset value in the software. Singletons (nodes that were not connected to any other protein) were not included in the generation of the figures. Thus, an array of sub-networks was generated. In total, 69 clusters were found for PAR4 and 65 for SLiM. However, only 22 clusters for PAR4 and 17 clusters for SLiM had more than three nodes. while the Y157C proteomes had fewer protein, which resulted in fewer clusters as well: 28 for PAR4 and 47 for Y157C. Of these, only 7 for PAR4 and 18 for Y157C had three or more nodes.

Functional enrichment was then retrieved for each cluster using the STRINGapp. Each node was colour coded depending on its intracellular localisation to visualise any potential patterns within the sub-networks. For the SLiM experiment, ribosomal proteins were labelled in red, mitochondria yellow, nucleus blue, endoplasmic reticulum green and since a large deal of proteins were located in ribonucleoprotein complexes, this specific term was labelled in purple. These five regions were arbitrarily chosen as they represent five of the main components of the cell, but also because most of the nodes fell within at least one of these regions.

The major cluster in the PAR4 (Figure 3.10) group was composed by a ribonucleoprotein complex, more specifically **SRP-dependent cotranslational proteins targeting to membrane**, which was the most enriched term, belonging to the Reactome Pathways category and achieving a FDR value of 1E-61. As seen from the colour code, most of the proteins were found in ribosomes, nucleus, and ribonucleoprotein complex. A few nodes were highlighted in green, meaning they are present in the endoplasmic reticulum. These are: Signal recognition particle 9 kDa protein (SRP9), signal recognition particle 72 kDa protein (SRP72), signal recognition particle receptor (SRPR), small ribosomal subunit protein eS26 (RPS26), and small ribosomal subunit protein eS28 (RPS28). Interestingly three proteins were labelled also in yellow, highlighting their presence in mitochondria. These are: large mitochondrial ribosomal protein 33 (MRPL33), small mitochondrial ribosomal protein 11 (MRPS11), and NOP2/Sun RNA methyltransferase family member 2 (NSUN2). This last protein regulates epidermal cell growth and proliferation, and it is required for proper spindle assembly and chromosome segregation.

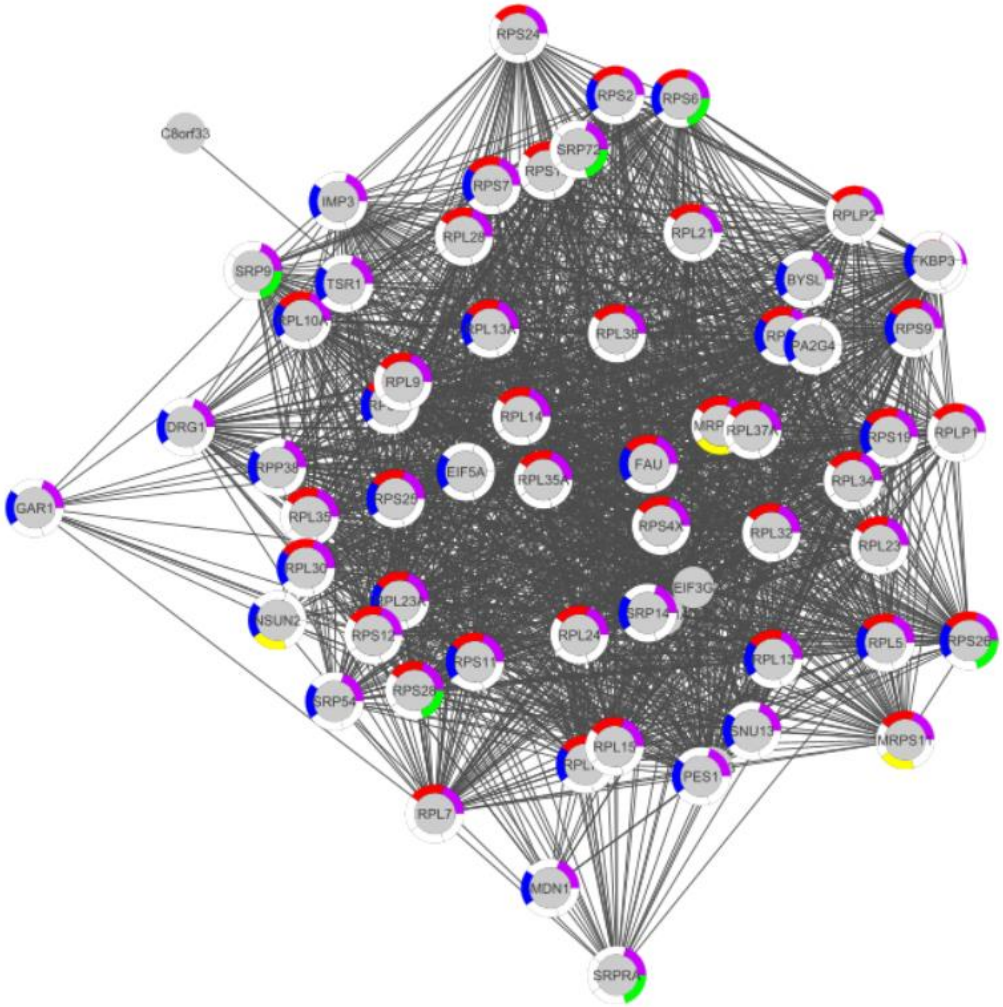


Figure 3.10 – The largest cluster in the PAR4 network is mainly composed of ribonucleoprotein complex nodes. Red: Ribosomal proteins, Blue: Nuclear proteins, Purple: Ribonucleoprotein complexes, Green: Endoplasmic reticulum proteins, Yellow: Mitochondrial proteins.

In the SLiM group, the largest cluster (Figure 3.11) was **Cytoplasmic translation** (FDR: 9.35E-62, GO Biological Process). Again, most of the proteins were found in ribosomes, nuclei, and ribonucleoprotein complex. Several ribosomal subunits and signal recognition particles present in this cluster are also found in the endoplasmic reticulum (RPS6, RPS28, SRP9, SRP72, SRPR). For this cluster, deletion of the SLiM motif completely ablated interactions with mitochondrial ribosomal proteins found in the PAR4 group.

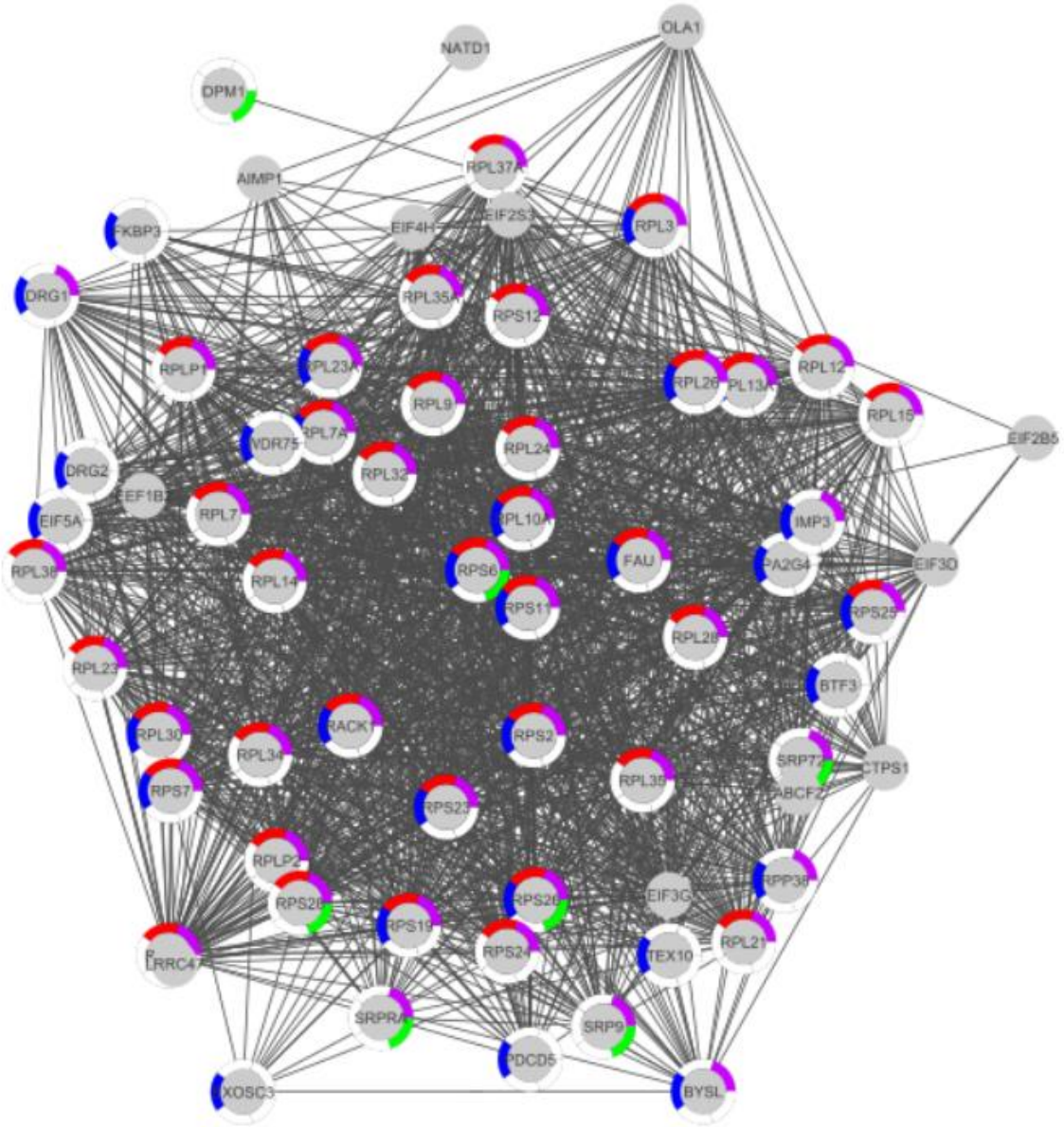


Figure 3.11 - The largest cluster in the SLiM network is composed of ribonucleoprotein complex nodes involved in cytoplasmic translation. Red: Ribosomal proteins, Blue: Nuclear proteins, Purple: Ribonucleoprotein complexes, Green: Endoplasmic reticulum proteins.

The second biggest sub-network in the PAR4 dataset (Figure 3.12) was composed mainly of **mRNA processing, and K homology domain, type 1** proteins, achieving a FDR score of 4.75E-27 in the String Clusters term, which was the most enriched. Most of these proteins were found in the nucleus, and a fraction of these also forming ribonucleoprotein complexes.

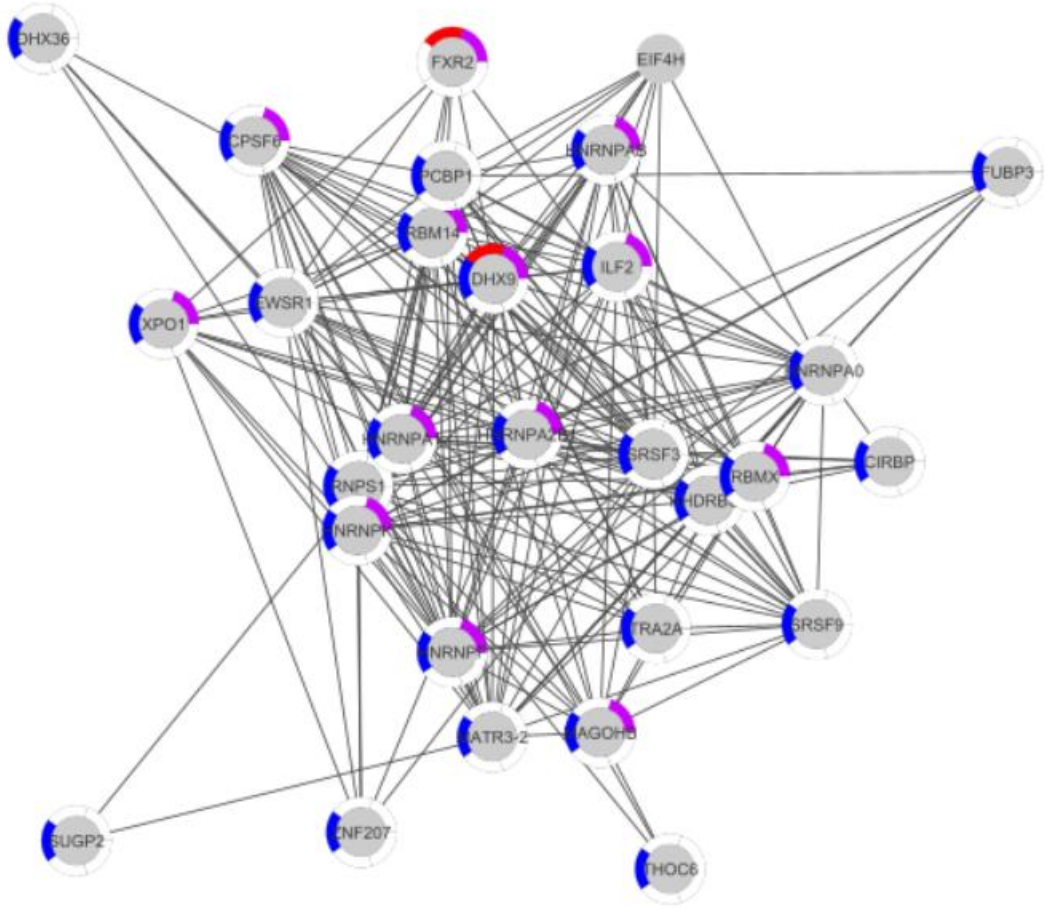


Figure 3.12 – The second largest cluster in the PAR4 network is mainly composed of mRNA processing proteins. Red: Ribosomal proteins, Blue: Nuclear proteins, Purple: Ribonucleoprotein.

The second largest cluster in the SLiM dataset (Figure 3.13) was also composed mainly of **mRNA processing and RNA recognition motif domain** proteins (FDR 1.37E-22, String Clusters), as well as **mRNA processing and CRD-mediated RNA stability complex** (7.6 E-22, String Clusters). The proteins present in this cluster were found in the nucleus and ribonucleoprotein complex, except for Calcium/calmodulin-dependent protein kinase type II subunit delta (CAMK2D) a kinase involved in the regulation of Calcium homeostasis, which is present in the endoplasmic reticulum.

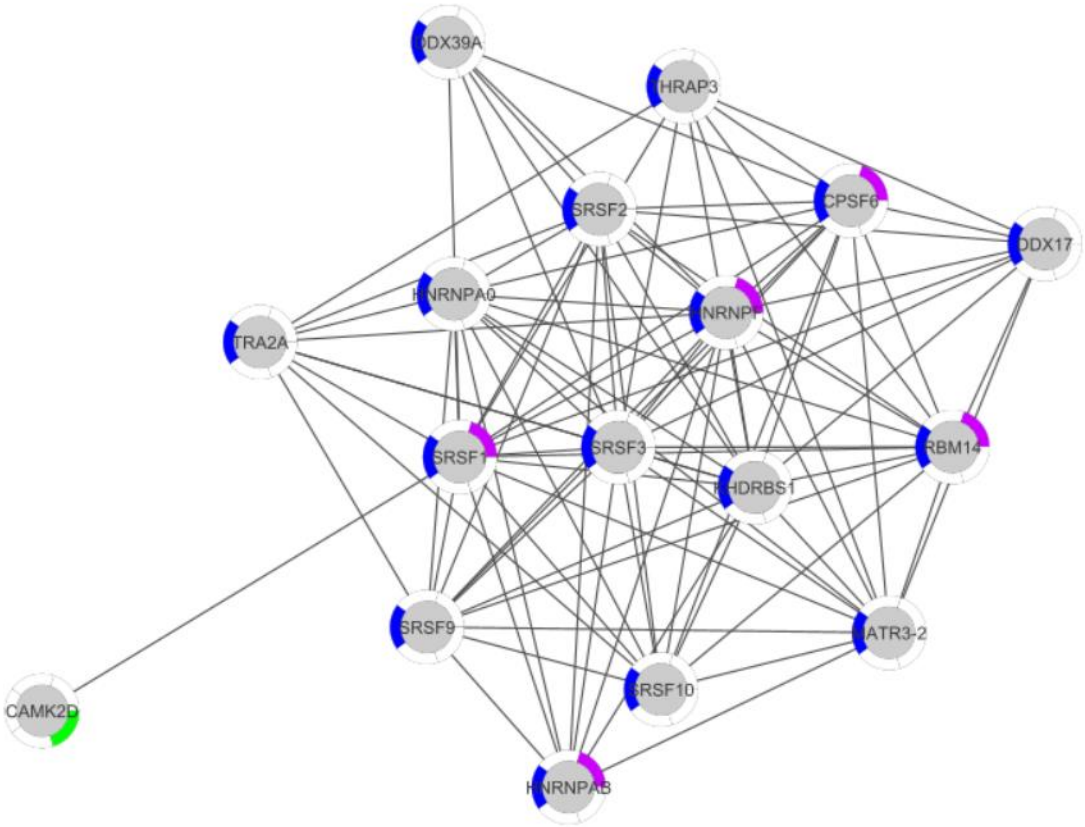


Figure 3.13 – The second largest cluster in the SLiM network is mainly composed of mRNA processing proteins. Red: Ribosomal proteins, Blue: Nuclear proteins, Purple: Ribonucleoprotein complexes, Green: Endoplasmic reticulum proteins.

The third biggest cluster of the PAR4 group (Figure 3.14) was formed of thirteen proteins, mostly mitochondrial, and the most enriched term was the String Cluster **Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling protein a Cytochrome complex** with an FDR score of 1.62E-15. According to the enriched GO cellular component, eleven of the thirteen proteins can be found in the mitochondria.

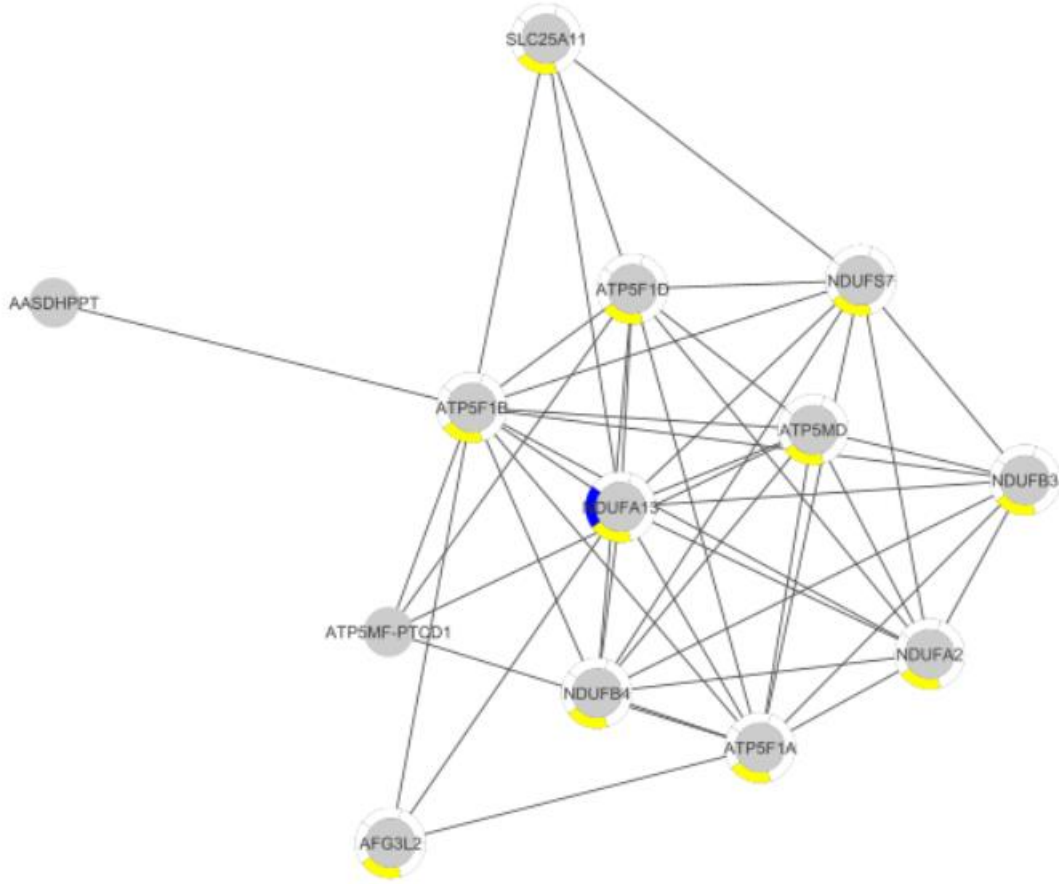


Figure 3.14 – The third largest cluster in the PAR4 network is mitochondrial proteins involved in respiratory electron transport. Blue: Nuclear proteins, Yellow: Mitochondrial proteins.

The third cluster in the SLiM group (Figure 3.15) was again formed by nuclear and ribonucleoprotein complex, thus the most enriched term was the String Cluster **U2-type spliceosomal complex, and Sm-like protein family complex**, with an FDR: 4.67E-17.

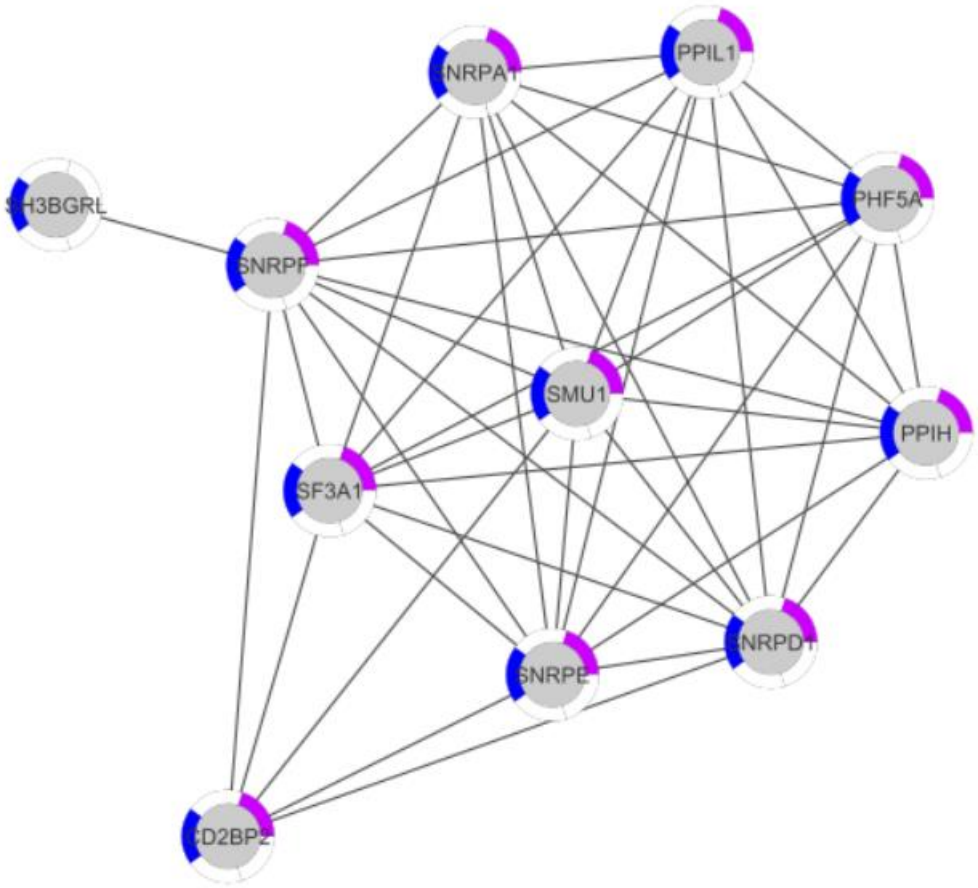


Figure 3.15 – The third cluster in the SLiM group is composed spliceosomal proteins. Blue: Nuclear proteins, Purple: Ribonucleoprotein complexes.

All the top 3 main clusters in both PAR4 and SLiM datasets were found in ribosome, nuclei, and ribonucleoprotein complexes. However, it is expected for a GPCR to interact with ribosomal proteins during the translation of mRNA to amino acids, therefore these results do not shed light on any new possible signalling pathway in the context of PAR4 pharmacology.

The STRINGapp allows to apply filters to select nodes present only in certain tissues or cellular compartments. Since a fraction of the clusters were colour coded in yellow (mitochondria), the mitochondrial filter was applied, to better visualise only the clusters found in these organelles. Figure 3.16 illustrates the comparison of both network clusters before and after applying the filter. PAR4 interacts with 51 mitochondrial proteins, while SLiM has 69 interactions in mitochondria.

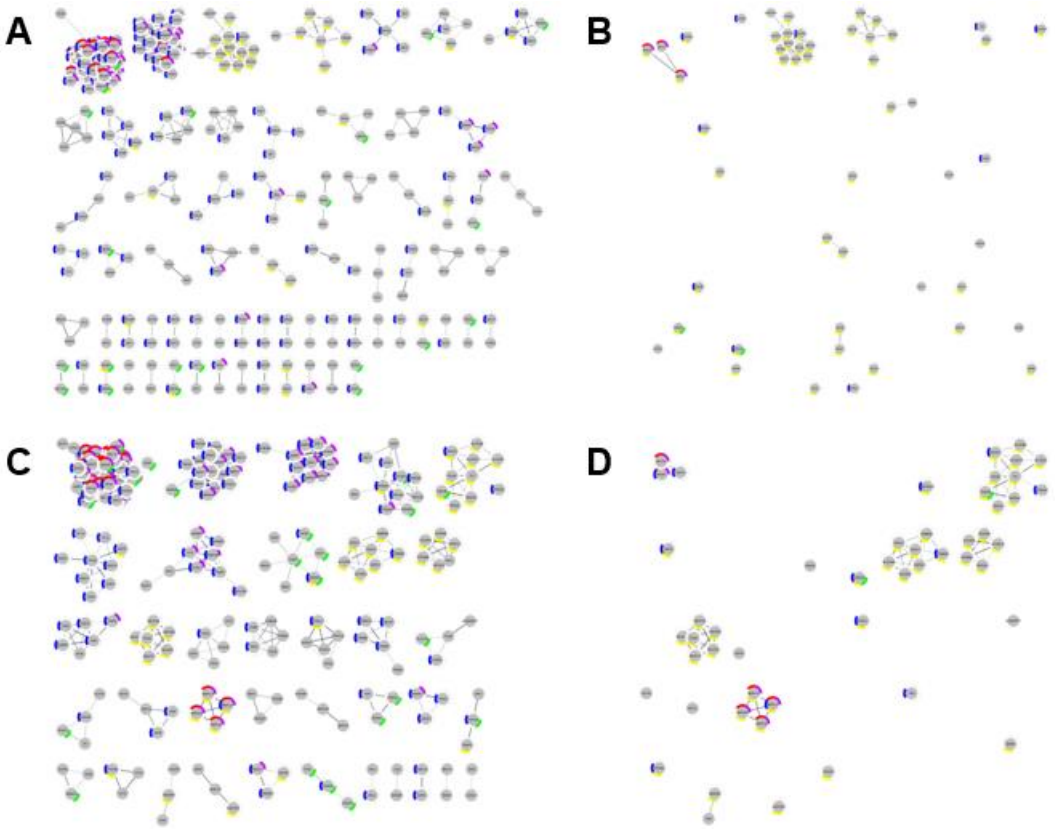


Figure 3.16 – Network clusters before and after applying the filters for mitochondrial proteins. A: PAR4 before filtering. **B:** PAR4 after filtering. **C:** SLiM before filtering. **D:** SLiM after filtering.

In the PAR4 group, only three mitochondrial clusters have more than 2 nodes. The most enriched term in the first cluster was **mitochondrial translation elongation** (FDR: 0.0108, STRING Clusters); all three proteins were mitochondrial ribosomes (MRPL33, MRPS11, RPLP2), thus explaining the enriched term. The second cluster was mostly enriched for **inner mitochondrial membrane protein complex** (FDR: 7.59E-14, COMPARTMENTS), the nodes included in this network were multiple ATP synthase proteins (ATP5F1A, ATP5F1B, ATP5F1D, ATP5MD), as well as NADH dehydrogenases (NDUFA2, NDUFA13, NDUFB3, NDUFB4, NDUFS7). The last mitochondrial cluster in the PAR4 network was composed by the following proteins: solute carrier family 25 member 5 (SLC25A5), voltage-dependent anion-selective channel protein 1 and 3 (VDAC1 and VDAC3), and second mitochondrial-derived activator of caspase (DIABLO-2), Creatine kinase, mitochondrial 1B (CKMT1B), and translocase of outer mitochondrial membrane 22 homolog (TOMM22). The most enriched term in this cluster was **Mitochondrial envelope** (FDR: 3.0E-5, COMPARTMENTS).

Since SLiM exhibited a higher number of interactions in the mitochondria, this led to SLiM having a greater number of clusters as well. The largest one was composed of solute carrier family 25 proteins (SLC25A1, SLC25A6, SLC25A11), Isocitrate dehydrogenase (IDH3G), Pyruvate dehydrogenase E1 (PDHB), Citrate synthase (CS), tridunctional enzyme subunits alpha and beta (HADHA and HADBH), thioredoxin-dependent peroxide reductase (PRDX3), and acidic calcium-independent

phospholipase A2 (PRDX6). The most enriched term in this cluster was **Mitochondrion** (FDR: 4.32E-8, GO Cellular Compartments). The second mitochondrial cluster was also mostly enriched for **Mitochondrion** (FDR: 8.01E-14, UniProt Keywords) and the proteins composing it were Sortin and assembly machinery component 50 (SAMM50), Translocase of inner mitochondrial membrane 50 (TIMM50), Heat shock protein mitochondrial (HSPD1), VDAC1/3, SLC25A3 and SLC25A5.

The third cluster in the SLiM group was mostly composed of ATP synthases (ATP5PB, ATP5MG, ATP5PD, ATP5F1C), and Paraplegin-like protein (AFG3L2), thus the most enriched term was **Formation of ATP by chemiosmotic coupling** (FDR: 1.33E-8; Reactome Pathways).

Another cluster was composed of NADH dehydrogenases (NDUFA9, NDUFAF3, NDUFB3, NDUFV1, NDUFA4) and cytochrome oxidase assembly factor 3 (COA3), which resulted in the most enriched term being **Thermogenesis** (FDR: 6.08E-6, KEGG Pathways). A small cluster of four mitochondrial ribosomes was also present (MRPL14, MRPS16, MRPL10, MRPL20), unsurprisingly the top term was **mitochondrial ribosome** (FDR: 7.43E-7, GO Cellular Component). Another cluster grouped mitochondrial proteins rather randomly, very few terms were present and the most enriched one was simply **mitochondrion inner membrane** (FDR:3.76E-5, UniProt Keywords).

The colour coding scheme for the Y157C experiments was slightly different. Ribonucleoprotein complex, mitochondria and nucleus retained the same colours (purple, yellow, and blue respectively), however since endoplasmic reticulum and ribosomal proteins were not found in these groups, cytosolic proteins were shown in green while cytoskeletal proteins for PAR4, and spliceosomal proteins for Y157C were highlighted in red.

The largest cluster in the PAR4 network, shown in figure 3.17, was mainly composed of members of the proteosome complex. The most enriched terms were three Reactome Pathways achieving the same FDR value of 7.11E-10. The terms were: **Hedgehog ligand biogenesis, Hh mutants are degraded by ERAD**, and **Defective CFTR causes cystic fibrosis**.

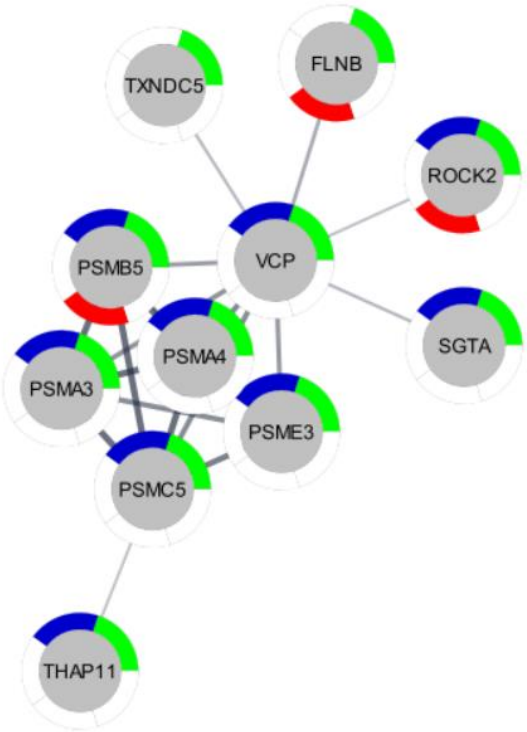


Figure 3.17 – The largest cluster in the PAR4 group is composed of proteosomal proteins. Red: Spliceosomal proteins, Blue: Nuclear proteins, Green: Cytosolic proteins.

The second cluster (Figure 3.18) was mainly composed of cytoplasmic proteins, and the top three enriched terms belonged to the STRING Clusters database. These were **Carbon metabolism and pyruvate metabolism** (FDR: 2.21E-6), **Citrate cycle (TCA cycle)**, **and Lactate dehydrogenase activity** (FDR:5.23E-6), and **Tricarboxylic acid cycle** (FDR: 4.85E-5). The involvement of this cluster in metabolism was highlighted by other databases as well, since three KEGG Pathways (Citrate cycle, metabolic pathways, and carbon metabolism) showed the same FDR value of 5.41E-5, whereas the most enriched GO Biological Process was **dicarboxylic acid metabolic process**.

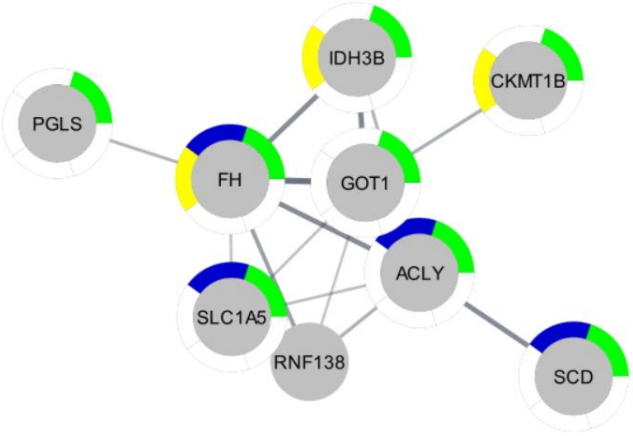


Figure 3.18 – The second cluster in the PAR4 group is formed by proteins involved in metabolism. Blue: Nuclear proteins, Green: Cytosolic proteins, Yellow: Mitochondrial proteins.

In the third cluster shown in figure 3.19, the main components were cytoplasmic and mitochondrial proteins. The most enriched term was **Respiratory electron transport** (Reactome Pathways, FDR: 2.53E-12), followed by **Respiratory chain complex, and complex I biogenesis** (STRING clusters, FDR: 3.38E-12), **Complex I biogenesis** (Reactome Pathways, FDR: 8.29E-9), **Mitochondrial complex I assembly model OXPHOS system** (WikiPathways, FDR: 1.04E-8), and **Thermogenesis** (KEGG Pathways, FDR:2.3E-8).

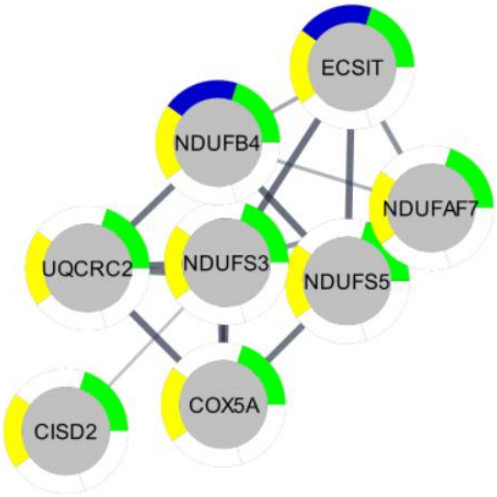


Figure 3.19 – The third cluster is formed by mitochondrial proteins found in the respiratory chain complex. Blue: Nuclear proteins, Green: Cytosolic proteins, Yellow: Mitochondrial proteins.

The first cluster in the Y157C group (Figure 3.24) was formed mainly by cytoplasmic proteins. The most enriched terms were **Proteasome** (KEGG Pathways, FDR: 1.71E-10), **UCH proteases** (Reactome Pathways, FDR:5.31E-10), **Proteasome** (UniProt Keywords, FDR: 9.85E-16).

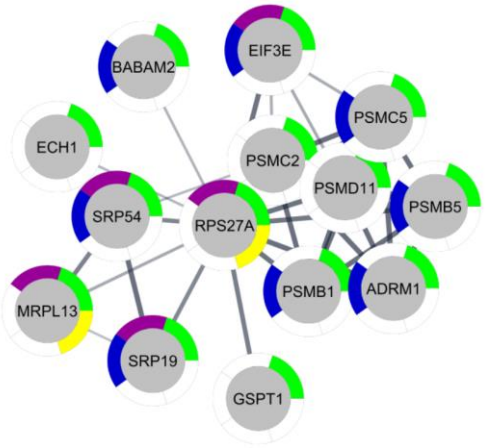


Figure 3.24 – The first cluster in the Y157C group was mainly formed by proteosomal proteins. Blue: Nuclear proteins, Purple: Ribonucleoprotein complexes, Green: Cytosolic proteins, Yellow: Mitochondrial proteins.

The second cluster in the Y157C group shown in figure 3.25 was mainly made by spliceosomal proteins. In fact, the most enriched terms for this group **were Spliceosomal snRNP complex** (Compartments, FDR: 1.65E-16), **U2-type spliceosomal complex, and mRNA Splicing – Major Pathway** (STRING clusters, FDR:2.42E-16), **precatalytic spliceosome, and Renpenning syndrome** (STRING clusters, FDR: 6.59E-15), **Spliceosome** (KEGG Pathways, FDR: 6.59E-15), **mRNA Splicing – Major Pathway** (Reactome Pathways, FDR: 1.11E-14), and many others.

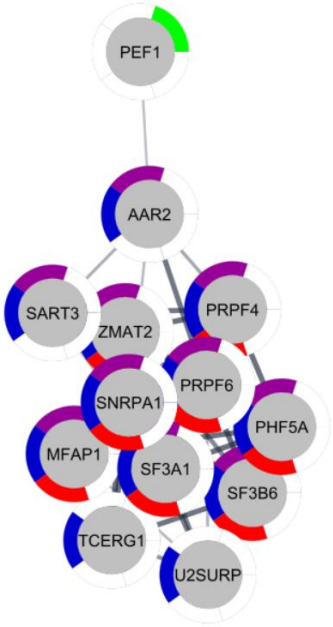


Figure 3.25 – The second cluster in the Y157C proteome is manly composed of spliceosomal proteins. Red: Spliceosomal proteins, Blue: Nuclear proteins, Purple: Ribonucleoprotein complexes, Green: Cytosolic proteins.

The third cluster (Figure 3.26) was composed of cytoplasmic and nuclear proteins, although two nodes were found in mitochondria: creatine kinase B-type (CKB) and hydrocyacylglutathione hydrolase (HAGH). The most enriched term in this cluster was the UniProt keyword **Glycolysis** (FDR: 7.89E-7), followed by the KEGG Pathways **Metabolic pathways** (FDR: 2.29E-6) and **Glycolysis/Gluconeogenesis** (FDR: 2.71E-6). **Glycolysis and gluconeogenesis** was reported also by WikiPathways with a FDR: 2.77E-6.

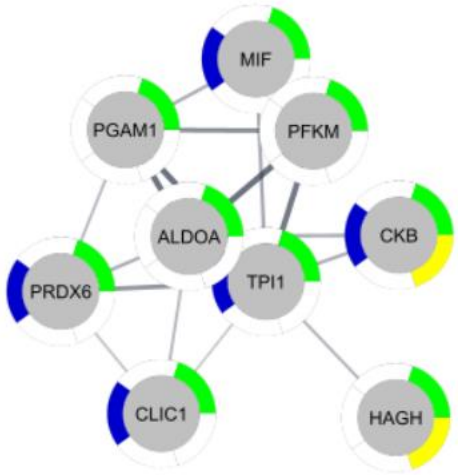


Figure 3.26 – The third cluster is formed by proteins involved in glycolysis and gluconeogenesis. Blue: Nuclear proteins, Green: Cytosolic proteins, Yellow: Mitochondrial proteins.

As with the SLiM proteome, mitochondrial clusters were filtered with the built-in function of the STRING app in Cytoscape. The cut-off was set at a value of 3, the same of the SLiM experiments. Figure 3.34 shows the clustered proteomes for both PAR4 and Y157C before and after the filtering was applied. The PAR4 proteome showed only two clusters with at least four nodes after applying the filter. These are clusters number three and five, which have been described previously. The larger one is cluster number three, showing the involvement of the proteins in this cluster in **Mitochondrial respiratory chain complex 1 assembly** (FDR: 3.15E-7, GO Biological Process). The majority of the proteins in this cluster belonged to the NADH dehydrogenase family (NDUFB4, NDUFS3, NDUFS5, NDUFAF7), while the remaining ones were cytochrome b-c1 complex subunit 2 (UQCRC2), cytochrome c oxidase subunit 5A (COX5A), CDGSH iron-sulfur domain containing protein 2 (CISD2), and evolutionarily conserved signalling intermediate in toll pathway (ECSIT). The other cluster left in the PAR4 proteome after applying the filter was composed by four nodes, three of which were mitochondrial ribosomes, therefore it is not surprising that the only enriched GO Biological Process term is **Mitochondrial translation** (FDR: 8.9E-5). The only node not belonging to the mitochondrial ribosome family is ribosome-releasing factor 2 (GFM2), a mitochondrial GTPase that mediates the disassembly of ribosomes from mRNA at the termination of mitochondrial protein biosynthesis. The other proteins that remained after applying the filter, but that did not cluster were: fumarate hydratase (FH), isocitrate dehydrogenase subunit beta (IDH3B), creatine kinase U-type (CKMTIB), serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform (PPP2CA), glutathione S-transferase P (GSTP1), chloride intracellular channel protein 1 (CLIC1), probable ATP-dependent RNA helicase (DDX28), keratin type I cytoskeletal 19 (KRT19), phosphoribosyl pyrophosphate synthase-associated protein 1(PRPSAP1), Stomatin-like protein 2 (STOML2), and polymerase delta-interacting protein 2 (POLDIP2).

On the other hand, the Y157C proteome is slightly larger and therefore it showed 3 clusters with more than 3 nodes after applying the filter. These are members of the third, fourth and ninth clusters previously described. The components of the third cluster, after applying the filter, were: peroxiredoxin 6 (PRDX6), chloride intracellular channel protein 1 (CLIC1), triosephosphate isomerase (TPI1), creatine kinase B-type (CKB), and hydroxyacylglutathione hydrolase (HAGH). The only terms enriched in this cluster were **methylglyoxal metabolic process** (GO Biological Process, FDR: 0.0291), **Nitration** (UniProt Keywords, FDR: 0.0405), and **Metabolic pathways** (KEGG Pathways, FDR: 0.0448). Functional enrichment of the fourth cluster after filtering did not return any enriched term. The nodes forming this cluster were: endonuclease G (ENDOG), sacsin (SACS), heat shock protein family A member 4 (HSPA4), peroxidasin homolog (PXDN), and keratin type 1 cytoskeletal 19 (KRT19). The last mitochondrial cluster was similar to the one found in the PAR4 proteome, being formed by NADH dehydrogenase proteins (NDUFA8, NDUFS3),), cytochrome c oxidase subunit 5A (COX5A), cytochrome b-c1 complex subunit Rieske (UQCRFS1), and ATP synthase subunit D (ATP5PD), therefore the main enriched terms revolved around oxidative phosphorylation, respiratory electron transport and ATP synthesis. Other proteins that remained in the Y157C proteome after applying the mitochondrial filter were: mitochondrial ribosomal proteins (MRPL13, MRPS2, MRPS5, MRPS34, RPS27A, MPV17L2, MRM3), delta(3,5)-delta(2,4)-dienoyl-CoA isomerase (ECH1), neuroguidin (NGDN), L-aminoadipate-semialdehyde-

dehydrogenase-phosphopantetheinyl-transferase (AASDHPPT), heat shock proteins (HSP90AA1, HSP90AB1), 2',5'-phospohdiesterase 12 (PDE12), actin-histidine N-methyltransferase (SETD3), sideroflexin-1 (SFXN1), 6-phosphogluconate dehydrogenase (PGD), NAD(P) transhydrogenase (NNT), ornithine aminotransferase (OAT), N(G), N(G)-dimethylarginine dimethylaminohydrolase 2 (DDAH2), histidine-tRNA ligase (HARS1), probable proline-tRNA ligase (PARS2), NAD(P)H-hydrate epimerase (NAXE), trifunctional purine biosynthetic protein adenosine-3 (GART). A summary of all the mitochondrial proteins identified by network clustering is given in the supplementary Table S9.

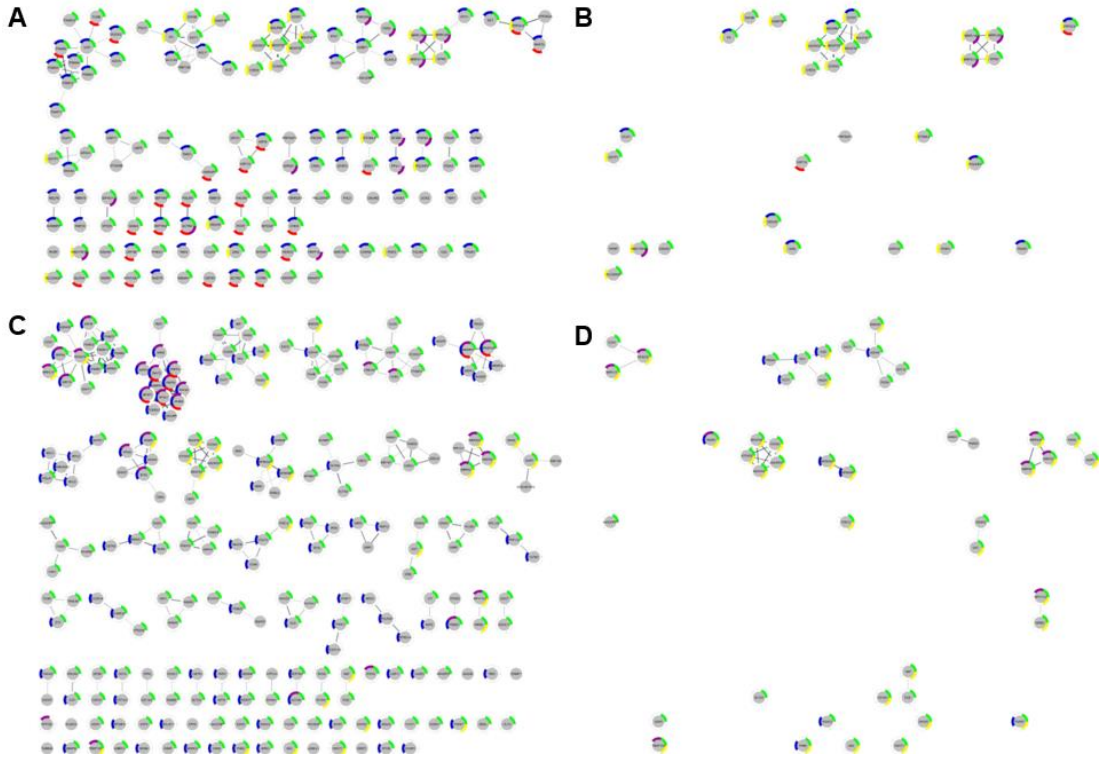


Figure 3.34 – Clustered proteomes before and after applying the filter for mitochondrial proteins. A: PAR4 before filtering. **B:** PAR4 after filtering. **C:** Y157C before filtering. **D:** Y157C after filtering.

3.4. DISCUSSION:

Using a proteomic approach, we characterised the network of PAR4-interacting partners. Most of these proteins were involved in translation and were found in nuclei and ribosomes. A small subsection of clusters was found in mitochondria.

3.4.1. PAR4 interactions with an array of PDZ proteins:

The rationale of the proteomics study was to explore how receptor-protein interactions were affected by the ablation of the PDZ-binding domain on the extreme C tail of PAR4. Known GPCR-PDZ interactions from previous literature[233] suggested that to date, no PDZ proteins were found interacting with the proposed binding site on PAR4. In this study, we identified a total of four proteins across three of the four datasets as potential PDZ interactions with PAR4. These are tight junction protein 2 (TJP2), Lim domain only 7 protein (LMO7), PDZ And LIM Domain 1 (PDLIM1), and GIPC PDZ Domain Containing Family Member 1 (GIPC1).

TJP2 is a crucial component of tight junction complexes that are integral for maintaining cell polarity and barrier functions in epithelial and endothelial cells. Genetic studies have linked mutations in the TJP2 gene to various health conditions, including progressive nonsyndromic hearing loss and liver disease. For instance, genomic duplications leading to the overexpression of TJP2 have been associated with altered expression of apoptosis genes in cases of progressive hearing loss (DFNA51)[234]. Moreover, mutations that result in truncated forms of the TJP2 protein disrupt tight junction integrity, leading to severe cholestatic liver disease[235]. Moreover, it has been reported to acts as a positive regulator of RANKL-induced osteoclast differentiation, potentially via mediating downstream transcriptional activity[236]. No interactions between its PDZ domains and any GPCRs have yet been reported.

LMO7 is a multifunctional protein implicated in various cellular processes. One significant role is its interaction with cell-cell adhesion molecules, where it facilitates the association of nectin and E-cadherin through afadin and α-actinin in epithelial cells[237]. This interaction is crucial for maintaining proper cell adhesion and integrity in epithelial tissues. Furthermore, LMO7 binds to emerin, a protein involved in the nuclear envelope and regulates the transcription of emerin and other muscle-relevant genes, highlighting its role in muscle function and development[238]. Despite being involved in multiple processes, LMO7 has never been documented to interact with GPCRs.

PDLIM1 is a cytoskeletal protein involved in various cellular mechanisms including signal transduction, organization of the cytoskeleton, and gene expression. This protein was also found to be upregulated in the PAR4 group of the Y157C experiment. Structurally, it contains both PDZ and LIM domains, facilitating interactions with multiple protein types to mediate its roles. It is involved in assembly, disassembly and directioning of stress fibers in fibroblasts. It is also required for the localisation of ACTN1 and PALLD to stress fibers, as well as being required for cell migration and in maintaining cell polarity of fibroblasts[239]. Moreover, PDLIM1 has been shown to inhibit NF-κB-mediated inflammatory signalling by sequestering the p65 subunit of NF-κB in the cytoplasm, suggesting its potential as a therapeutic target in inflammatory diseases[240]. Recent studies indicate that PDLIM1 has significant implications in cancer biology, where it influences cell migration, invasion, and metastasis,

particularly in breast cancer [241]. Despite being so well characterised in oncology, no interactions with GPCRs have been yet reported in literature.

GIPC1 was found across two proteomes, and it is therefore deemed as a highly confident interaction. This is also the only interaction already found to interact with other GPCRs, as in the case of the dopamine D2 receptor, which has been computationally modelled to interact with this protein at the proposed PDZ binding site[242]. Moreover, GIPC was found to interact with RGS-GAIP, a GTPase-activating protein (GAP) for $G_{\alpha i}$ subunits localized on clathrin-coated vesicles, suggesting that GIPC may be a component of a broader G protein-coupled signalling complex[243]. Interestingly, none of these interactions was found in the $PAR4_{\Delta SLiM}$ proteome, thus confirming that the proposed binding motif might be a potential PDZ-binding domain. However, to ascertain this hypothesis, more in depth structural studies are needed. Fortunately, the development of AI tools for structure prediction such as AlphaFold can be used to unravel whether the SLiM is really a binding site for PDZ proteins.

3.4.2. Involvement of PAR4 in RNA splicing:

RNA splicing is a crucial process in the maturation of mRNA. The genetic material within the nucleus, is initially transcribed into pre-mRNA, which undergoes various modifications before it becomes a mature mRNA, ready for ribosomal translation. RNA splicing is the removal of introns from the pre-mRNA, and the joining together of the exons. This process of removing introns and joining exons in RNA splicing contributes to the diversity of proteins that can be produced from a single gene. It allows for alternative splicing, where different combinations of exons can be included or excluded, resulting in multiple mRNA variants and, consequently, different protein isoforms. This mechanism greatly expands the coding potential of the genome and plays a crucial role in regulating gene expression and cellular function[244]. The precise excision of introns from pre-messenger RNA is performed by the spliceosome, a macromolecular machine containing five small nuclear RNAs and over 140 proteins, making the spliceosome one of the most complex cellular machines characterized[245]. Across the four datasets, only PAR4 in the Y157C experiment did not report any interaction with spliceosomal proteins, while for each of the other datasets at least 20 genes per group were found to be enriched in the term RNA splicing. This raises the question of whether PAR4 could be involved in this specific process. It is not the first time that affinity proteomics on GPCR targets finds interactions between receptors and spliceosomal proteins, as this was already documented with the very large G-protein-coupled receptor 1 (VLGR1/ADGRV1), a member of the adhesion G-protein-coupled receptor (ADGR) family[246]. In this case the authors, although showing strong evidence for the interaction of their receptor with transcriptional regulators and spliceosomal proteins, do not provide an explanation for the biological implications of such interactions. The notion of GPCRs being localised only at the cellular membrane is starting to be disregarded, as increasing evidence shows that these receptors can also associate with various intracellular membranes where they may couple to different signalling systems[247]. Since mRNA splicing happens in the nucleus[248], it can therefore be a possibility that PAR4 might localise on the nuclear membrane to regulate mRNA splicing. However, the direct involvement of the receptor in this process still needs to be elucidated. Rather than direct interaction with the spliceosome, it could be possible that PAR4 triggers a signalling cascade which leads to regulation of spliceosomal activity.

3.4.3. SRP-dependent cotranslational protein targeting pathway:

What emerged from analyses of SLiM datasets is that major clusters of PAR4 and $PAR4_{\Delta SLiM}$ interacting proteins were found in ribosomes and ribonucleoprotein complexes. Therefore, the most enriched terms were SRP-dependent cotranslational protein targeting to the membrane pathway, mRNA binding, or formation of spliceosomal complexes. This was true regardless of the deletion of the proposed PDZ-binding site. However, this does not provide any new knowledge regarding PAR4 or GPCRs in general. The SRP-dependent cotranslational protein targeting to membrane is a critical process in the cell, and it is involved in the targeting and translocation of proteins to the endoplasmic reticulum (ER) membrane[249]. Being GPCRs transmembrane proteins, their biogenesis involves this crucial cotranslational step[250].

The process begins in the cytoplasm, where ribosomes start to synthesize the proteins based on the mRNA sequence. As the nascent peptide chain emerges from the ribosome, if it has an N-terminal signal sequence (characteristic of membrane or secretory proteins), it's recognized and bound by the Signal Recognition Particle (SRP), a complex composed of a 7SL RNA and six different proteins. This signal sequence is usually hydrophobic. At this point, the binding of the SRP causes a temporary pause in the translation process, preventing the complete synthesis of the protein in the cytosol. The SRP-ribosome-nascent chain complex (SRP-RNC) is then targeted to the ER membrane via interaction with the SRP receptor (SR), which is an integral membrane protein on the ER. Once at the ER membrane, the SRP is released, and the ribosome-nascent chain complex is transferred to a protein-conducting channel in the ER membrane, known as the translocon (Sec61 complex)[251]. The presence of PAR4 at ER was already well documented, and retention in this organelle has been found to be due to interactions with β-COP1, while co-expression with PAR2 enhanced localisation to the plasma membrane, as well as PAR4 glycosylation and activation[107]. For these reasons, despite being the major clusters in both datasets, no further efforts were made towards validating interactions with proteins belonging to these networks.

3.4.4. Mitochondria in the PAR4 proteome:

Besides the ribonucleoprotein complexes involved in splicing and translation, a fraction of PAR4-interacting proteins was in mitochondria. Deletion of SLiM, increased the number of such interacting proteins. Volcano plot analysis highlighted how the majority of significantly downregulated proteins were mitochondrial. This raises the question to whether the PAR4 could be found in these organelles. Over the years, different GPCRs have been found located in mitochondria, such as $P2Y_1$ and $P2Y_2$[252], Ang II type 1 and 2[253], $5\text{-}HT_4$[254], CB_1[255] and MT_1[256]. These receptors and their function on mitochondria are described in more detail in Table 3.11.

Table 3.11 – List of GPCRs found in Mitochondria.

GPCR	Function	Ref.
$P2Y_1$	Stimulation of mitochondrial Ca^{2+} uptake	252
$P2Y_2$	Inhibition of mitochondrial Ca^{2+} uptake	252
AT_1	Regulation of superoxide production and increase in mitochondrial respiration	253
AT_2	Modulation of nitric oxide formation and decrease in mitochondrial respiration	253
$5\text{-}HT_3$	Stimulation of mitochondrial Ca^{2+} uptake in hypoxic condition	254
$5\text{-}HT_4$	Inhibition of mitochondrial Ca^{2+} uptake	254
CB_1	Inhibition of adenylyl cyclase, modulation of neuronal metabolism	255
MT_1	Inhibition of stress-mediated cytochrome C release and caspase activation	257
A_2B	Reducing infarct size following ischaemic reperfusion injury Reducing calcium-induced mitochondrial permeability transition	258
$VLGR_1$	Regulation of Ca^{2+} homeostasis in mitochondria-associated membranes (MAMs)	259
GPR_{35}	Modulates oxidative phosphorylation under stress conditions	260

Interactions with downstream proteins might lead to metabolic regulation, as suggested by the enrichment analysis reported in this study. Network analysis revealed the presence of many ATP synthase and calcium transporters, highlighting the involvement of these proteins in energy production and cell death. A potential transporter that could localise PAR4 to the outer mitochondrial membrane can be found in the initial analysis, using methods previously published in literature. One of the few hits in the SLiM Medium/Light group was the protein Metaxin 2 (MTX2). This protein, albeit not well characterised, it is known to interact with MTX1 on the cytosolic face of the outer mitochondrial membrane[261]. The mitochondrial contact site and cristae organizing system (MICOS) complex associates with mitochondrial outer membrane proteins SAMM50, MTX1 and MTX2 (together described as components of the mitochondrial outer membrane sorting assembly machinery (SAM) complex) and DNAJC11, mitochondrial inner membrane protein TMEM11 and with HSPA9. The MICOS and SAM complexes together with DNAJC11 are part of a large protein complex spanning both membranes termed the mitochondrial intermembrane space bridging (MIB) complex[262]. This complex also plays a crucial role in mitochondrial protein biogenesis. For example, the translocase of the outer mitochondrial membrane (TOM) complex plays a crucial role in importing the VDAC precursor. Deleting Sam50, the central component of the sorting and assembly machinery (SAM), resulted in a significant defect in VDAC assembly and a decrease in the

steady-state VDAC level. Mitochondria depleted of Metaxin 2 exhibited lower levels of metaxin 1, leading to deficiencies in the import and assembly of VDAC and Tom40. Interestingly, Sam50-depleted mitochondria also showed reduced levels of metaxin 1 and metaxin 2, suggesting a connection among these three proteins, even though Sam50 and metaxins appeared to be in different complexes[263].

3.4.5. VDACs as potential mediators of PAR4-induced apoptosis:

The Voltage-dependent anion-selective channel proteins VDAC1 and VDAC3 have been identified as being central in this proteome, as they were found in both PAR4 and SLiM datasets. The VDAC family is composed of three isoforms, however the most extensively characterised is VDAC1[264]. This is a beta barrel protein, encoded by the gene VDAC1 on chromosome 5 in humans[265]. It is an important regulator of mitochondrial function and plays a central role in controlling over energy sources and metabolism[266]. This protein forms a channel through the mitochondrial outer membrane or the plasma membrane[267]. The channel at the outer mitochondrial membrane allows diffusion of small hydrophilic molecules; in the plasma membrane it is involved in cell volume regulation and apoptosis[268]. VDAC1 adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV[269]. The open state has a weak anion selectivity whereas the closed state is cation-selective[270]. Molecules acting on the channel include the sphingolipid ceramide, the phospholipid phosphatidylcholine, and the sterol cholesterol[271]. In depolarized mitochondria, VDAC1 is ubiquitinated by the E3 ubiquitin-protein ligase parkin (PRKN) and the mitochondrial Serine/threonine-protein kinase (PINK1) to promote mitophagy or prevent apoptosis; polyubiquitination by PRKN promotes mitophagy, while monoubiquitination by PRKN decreases mitochondrial calcium influx which ultimately inhibits apoptosis[272]. However, VDAC1 participation in apoptosis is ambiguous, as it can participate in the formation of the permeability transition pore complex (PTPC) responsible for the release of mitochondrial products that triggers apoptosis[273,274]. This ion channel is therefore crucial in keeping Ca^{2+} homeostasis and regulating cell life and death[275]. VDAC1 also plays a metabolic role, as it may mediate ATP export from cells[276]. Although there is still no evidence of the existence of mitochondrial PAR4, the involvement of the receptor in apoptosis is already well documented. Previous research highlighted how PAR4 deficiency is cardioprotective after acute MI reperfusion injury, however in this scenario the receptor induced myocyte apoptosis through activation of the JNK pathway[141]. Another study performed in a model of human colon organoids, found that PAR4 antagonism inhibited apoptosis after stimulation with thrombin[277].

Here we report interactions of the receptor with VDAC1 and VDAC3, which are well recognised mediators of apoptosis[275]. Functional enrichment analysis of the clusters encompassing VDAC1 and VDAC3 in both the PAR4 and SLiM group, highlighted how interactions with these proteins may be involved in Vpr-mediated induction of apoptosis by mitochondrial outer membrane permeabilization. Viral protein R (Vpr) is a small accessory protein found in certain retroviruses such as HIV-1 and HIV-2. It is capable of inducing apoptosis by interactions with the permeability transition pore complex (PTPC)[278]. If the presence of PAR4 in mitochondria was to be confirmed, an interesting avenue to explore would be the involvement of the receptor in VDAC-mediated apoptosis. Therefore, VDAC1 and VDAC3 antibodies have been acquired to investigate, through immunofluorescence and western blotting, if VDACs are expressed and colocalise with PAR4. To this extent, super resolution microscopies

can be an extremely useful tool to validate interactions between the PAR4 and members of the VDAC family. By using these techniques, it would be possible to visualize the specific sites of interaction between PAR4 and VDACs, and potentially identify any changes or loss of interactions caused by the ΔSLiM mutation. These techniques can provide information about the spatial distribution and organization of PAR4 and VDACs in the cell, giving further insights into their potential functional roles and investigate any implication in PAR4-induced apoptosis. Overall, the use of super resolution microscopies in this context could provide a more detailed understanding of the molecular mechanisms underlying PAR4 signalling and inform the development of potential therapeutic interventions.

3.4.6. Involvement of PARs in metabolism and inflammation

It is well recognised that PAR4 is a crucial receptor in platelet pathophysiology and plays a major role in thrombosis and atherosclerosis. However, recent evidence is starting to support the role of PAR4 in other scenarios. Kolpakov group demonstrated the implications of PAR4 in cardiac inflammation[141,142], while Fender showed how activation of the receptor can drive canonical NLRP3 inflammasome signalling in the diabetic heart[279]. As seen in Figure 1.1 (Chapter 1 – Expression and pathophysiology paragraph) The Human Protein Atlas reports that adipocytes are the only cell type in which F2RL3 RNA is enriched. The evidence for this claim is based on three different sources: UniProt protein existence (UniProt evidence); a Human Protein Atlas antibody- or RNA based score (HPA evidence); and evidence based on PeptideAtlas (MS evidence). However, when literature was searched on databases such as PubMed, Google Scholar and ScienceDirect using the keywords “PAR4” and “Adipose cell” the only study that was flagged up was Strande and Phillips, 2009[280]. In this paper the authors demonstrated how exposing adipose cells to thrombin leads to the secretion of proinflammatory mediators, and how PAR1 and PAR4 were detected both by rtPCR and immunoblot analysis. In 2015, Amisten and colleagues comprehensively reviewed GPCR expression in adipose cells, and mentioned that only traces of PAR4 were present, but the effects of its activation on lipolysis and the secretion of adiponectin and leptin are unknown[281].

The connection between thrombosis, inflammation, and obesity is supported by clinical and epidemiological evidence[282], and it is already established the involvement of PAR1 and PAR2 in metabolic syndrome. Activation of PAR2 by tissue factor in adipocytes contributes to diet-induced obesity by decreasing metabolism and energy expenditure.TF acts on PAR2 also in hematopoietic cells, driving adipose tissue inflammation, hepatic steatosis, and insulin resistance[283]. Antagonism of the receptor has strongly attenuated adiposity, adipose tissue inflammation, infiltrated macrophages and mast cells, insulin resistance, and cardiac fibrosis and remodelling[284]. PAR1 was also found to contribute to hepatic steatosis and inflammation when mice were fed a methionine/choline–deficient diet[285]. The role of PAR4 in this context is however still not defined.

3.4.7. Validation of PAR4 proteomics:

This comprehensive SILAC proteomics analysis has provided valuable insights into the interactome of PAR4, revealing significant interactions with both PDZ domain-containing proteins and mitochondrial proteins (Figure 3.35). These findings underscore the intricate network of interactions that PAR4 engages in within the cellular environment, highlighting its potential roles in key biological processes, including translation and energy metabolism. The identification of critical interactors such as GIPC1, TJP2, LMO7, and PDLIM1, alongside evidence suggesting PAR4's involvement in mitochondrial functions, establishes a robust foundation for future studies.

However, a central question remains: how feasible are the interactions identified through proteomics, and how do they translate into biological relevance? Proteomic approaches like SILAC are invaluable for uncovering potential protein-protein interactions, but they inherently capture a snapshot of interactions that may occur under experimental conditions, some of which may not reflect physiological scenarios. This necessitates a systematic process of validation to confirm whether the observed interactions are not only feasible but also functionally relevant within the biological context. Biological relevance is contingent on multiple factors, such as protein expression levels, post-translational modifications, subcellular localization, and the dynamic conditions of the cellular environment.

Bridging the gap between proteomics data and biological function requires complementary approaches. Computational tools, particularly those leveraging artificial intelligence (AI), provide a sustainable and scalable pathway to refine the list of candidate interactors. Advanced AI tools like AlphaFold are particularly promising in this regard, offering highly accurate predictions of protein structures and their potential interfaces. By integrating these predictions with proteomic data, it is possible to prioritize interactions that are structurally plausible and biologically meaningful, greatly enhancing the efficiency of experimental validation efforts. The sustainability of this AI-driven approach cannot be overstated—it reduces the reliance on time- and resource-intensive laboratory experiments, minimizes waste, and enables high-throughput exploration of complex interactomes. This aligns with modern principles of resource-efficient research, which is becoming increasingly important in the face of limited funding and the need for sustainable scientific practices.

The translation of proteomics findings into biological relevance also requires precise experimental tools. For example, super-resolution microscopy can be used to confirm the subcellular localization of PAR4 and its interactors, providing direct visual evidence of their spatial relationship within the cell. Similarly, biochemical techniques such proximity labelling can validate the physical interactions between PAR4 and key proteins, bridging the gap between computational predictions and *in vivo* relevance. These validations are essential to determine whether identified interactions are functionally significant or represent transient, non-specific associations.

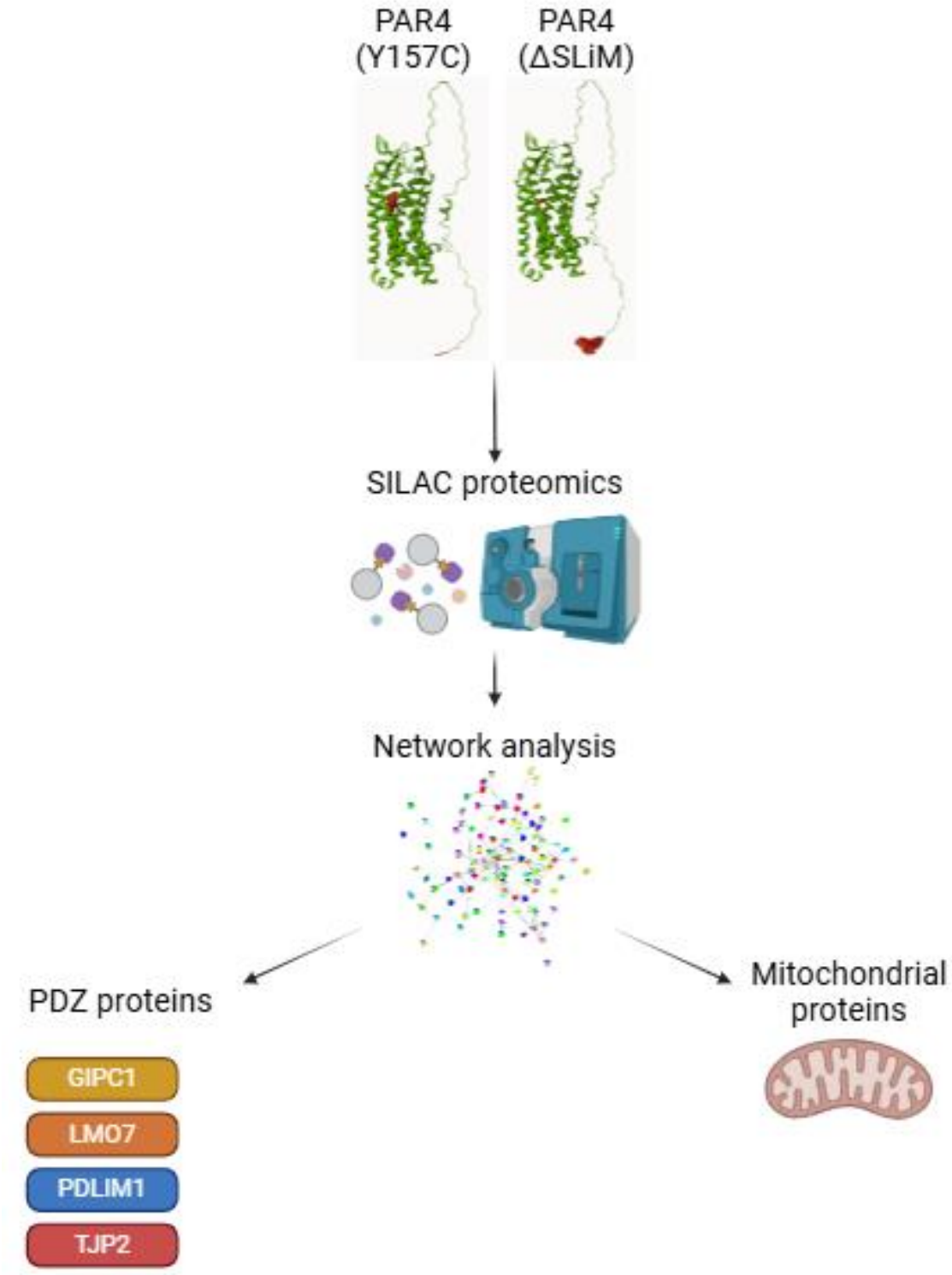


Figure 3.35 – Findings from Chapter 3. SILAC proteomics of PAR4 Y157C mutant and PAR4 ΔSLiM revealed interactions with 4 PDZ proteins (GIPC1, LMO7, PDLIM1, and TJP2) and clusters of mitochondrial proteins. These findings prompted questions for further validation.

4. AI AS A TOOL TO SCREEN PROTEIN INTERACTIONS

4.1. INTRODUCTION

4.1.1. AlphaFold and multiple sequence alignments (MSAs):

The fundamental basis for most biological processes involves the creation of both permanent and temporary protein complexes. Understanding the configuration of these complexes constitutes a crucial step in gaining insights into and potentially altering their functionality. In the absence of an experimental structure, computational methods are an invaluable tool to predict three-dimensional protein folding, as they have nearly reached the level of experimental precision[286]. The recent introduction of the AlphaFold structure prediction system has steered a significant improvement in prediction accuracy. AlphaFold is a state-of-the-art artificial intelligence tool developed by DeepMind, which can predict protein structures from their amino acid sequences. The developers of this tools won the Nobel Prize in Chemistry 2024[287]. AlphaFold employs a combination of information derived from the amino acid sequence, multiple sequence alignments (MSAs), and analogous structures to forecast the structure of individual protein chains[41]. MSAs are computational representations that align two or more sequences of biological macromolecules based on their shared evolutionary relationships[288]. These sequences are typically derived from different species or different members of a protein family. The primary goal of creating an MSA is to identify conserved regions (positions that have similar amino acids or nucleotides) and variable regions (positions with dissimilar elements) across the aligned sequences. MSAs are a cornerstone of homology modelling, in this approach, if a protein of interest lacks an experimentally determined structure, it can be predicted based on homologous proteins with known structures[289]. MSAs guide the alignment of the target sequence with the template, ensuring that conserved regions are aligned accurately, which improves the quality of the resulting model. When there are no close homologs with known structures, MSAs are used to identify distant homologs by recognizing common motifs in a database of known structures. This approach is known as fold recognition or threading[290]. MSAs help identify conserved features, even when sequence similarity is low. The central component of AlphaFold neuronal network, referred to as Evoformer, encompasses a neural representation of the MSAs and the pairwise relationships among amino acids in the protein. These two representations are merged and processed by an array of neural network modules. The pairwise representation carries information about the relative positions of amino acids within the chain and is utilized to anticipate the relative distances between these amino acids using a binned distance distribution (distogram)[41]. The first row of the MSA embedding, in conjunction with the pair embedding, is subsequently employed to make predictions about the final structure. The model undergoes end-to-end training, with gradients cascading from the predicted structure throughout the entire network[41].

4.1.2. Different metrics to assess protein structure prediction:

AlphaFold greatly outperformed other protein predicting algorithms at the 14th edition of the Critical Assessment of Protein Structure (CASP) in 2020. In this competition a specific metrics is used to assess the quality of a prediction, root mean square deviation (r.m.s.d). A lower r.m.s.d. corresponds to a more similar structure. AlphaFold2 predictions achieved an outstanding median backbone accuracy of 0.96 Å r.m.s.d.$_{95}$ (Cα root-mean-square deviation at 95% residue coverage) (95% confidence interval = 0.85–1.16 Å) whereas the next best performing method had a median backbone accuracy of 2.8 Å r.m.s.d.$_{95}$ (95% confidence interval = 2.7–4.0 Å).

Moreover, AlphaFold models achieved a high level of accuracy in predicting the position of residue side chains when the protein backbone prediction was accurate. These results meant that AlphaFold2 can predict protein structures with an accuracy within the width of a carbon atom (1.4 Å), making its predictions comparable to experimental data such as crystallography[41]. However, one issue with this metric is that all residue pairs are weighted evenly, and for larger proteins the r.m.s.d. can become sensitive to local structure deviation, rather than to the global topology of the structure. To overcome this issue, another method to evaluate protein similarity is the template modelling score (TM-score) which proves particularly useful when comparing predictions to resolved template structures. This different metrics gives a higher score to shorter distances between corresponding residues. This allows more sensitivity to global topology rather than local structure deviation[291]. The TM-score is always a value between 0 and 1. Large scale assessment of TM-score distribution found that it can be used as an approximate but quantitative criterion for protein topology classification. A value TM<0.17 translates to protein structures being indistinguishable from random pairs. However, a TM>0.5 signifies that the two structures are mostly in the same fold[292]. AlphaFold provides intrinsic model accuracy estimates of the predicted TM-score (pTM). In case of multi-domain or multi-structure prediction, AlphaFold Multimer can score interactions between residues belonging to different chains. This metric is called interface pTM or ipTM. The AlphaFold Multimer model uses a weighted combination of both metrics to assess a model confidence (0.8*ipTM + 0.2*pTM)[293]. ipTM scores gauge the precision of the predicted spatial arrangement of subunits within a complex. Scores over 0.8 indicate reliable, high-quality predictions, whereas scores below 0.6 often denote unsuccessful predictions. Scores between 0.6 and 0.8 fall into an uncertain range where predictions may be either correct or incorrect[294].

4.1.3. AlphaFold multimer can be used for protein complex prediction:

AlphaFold was initially trained on individual chains, including those found in complex with other proteins. However, it still demonstrated a remarkable ability to predict the structures of proteins bound with co-factors or stabilized by their interactions within multi-component complexes. Further research has indicated that inserting residue gaps or connecting chains with flexible linkers (referred to as pseudo-multimer input) into the single-chain AlphaFold model often yields successful predictions of multi-component interactions. The concept is that the neural network should recognize the linker segment as unstructured and then fold the single, multi-domain chain in a manner similar to multiple distinct chains[295]. AlphaFold source code and model parameters were made publicly available, so that it was possible to extend AlphaFold's capabilities to handle multiple chains both during training and inference. This upgraded system, known as AlphaFold-Multimer, offers native support for multi-chain featurization and symmetry handling, and it demonstrates superior performance when compared to existing methods, including pseudo-multimer inputs built on the foundation of AlphaFold. Several technical modifications were implemented, such as adapting the loss functions to account for permutation symmetry among identical chains, pairing the MSA alignments across individual chains to extract inter-chain genetic information, introducing a novel approach for selecting subsets of residues for training, and adjusting the model architecture[293]. Multimer templates were not implemented yet, which leaves room for potential accuracy improvements in the future. Interestingly, AlphaFold Multimer performance tends to be higher for

homomeric interfaces as opposed to heteromeric ones. This disparity arises because the MSA readily encodes evolutionary information about complex interfaces in the homomeric scenario, whereas such information is more limited and challenging to access in the case of heteromeric interfaces[293]. Recently, Lee and colleagues demonstrated how AlphaFold multimer largely fails to predict full length sequences of interacting proteins involving short linear motifs. However, protein fragmentation boosts AlphaFold sensitivity at the cost of specificity[296]. Since the idea behind the project was to investigate whether the extreme end of PAR4 ($S^{381}SLLQ^{385}$) would interact with the PDZ domain of different proteins, it was possible to model just this short linear motif with the domains of interest.

4.1.4. pLDDT and PAE:

To account for uncertainty, AlphaFold predictions are given a confidence metric called predicted local distance difference test (pLDDT) on a scale from 0 to 100, which is colour-coded on the protein structure. A very highly confident prediction (pLDDT>90) is shown in blue, confident (90 > pLDDT > 70) is depicted using light blue, a low confidence prediction (70 > pLDDT > 50) is highlighted in yellow and very low (pLDDT < 50) is orange. The 3D coordinates of regions with pLDDT < 50 often have a ribbon-like appearance and should not be interpreted. pLDDT < 50 is a reasonably strong predictor of disorder, suggesting that such a region is either unstructured in physiological conditions or only structured as part of a complex. Moreover, structured domains with many inter-residue contacts are likely to be more reliable than extended linkers or isolated long helices, while unphysical bond lengths and clashes do not usually appear in confident regions[42]. Figure 1.2 in Chapter 1 shows the predicted structure of the human PAR4, coloured according to the pLDDT of each residue.

Another metric that the AlphaFold provides is Predicted Aligned Error (PAE), which measures the confidence in the relative position of two residues within the predicted structure, providing insight into the reliability of relative position and orientations of different domains. It is usually colour coded: a darker colour corresponds to a good prediction (low error), whereas a light colour indicates poor prediction (high error). If a protein has multiple domains, or if multiple proteins are predicted together, these can often be seen on the PAE map as distinct squares. This is because AlphaFold is often confident of relative position of residues within, but not across, domains[297]. Figure 4.2 show the PAE for the same structure presented in figure 4.1.

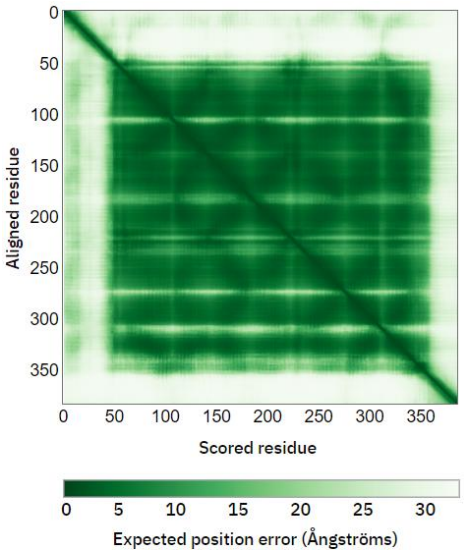


Figure 4.2 – Predicted Aligned Error (PAE) of human PAR4. Each residue is scored based on the alignment to every other amino acid in the protein. The higher the confidence of the relative position between two residues, the darker the colour on the plot.

4.1.5. AlphaFold database:

One of the major milestones in biology in recent years was the Human Genome Project, which ended in 2003 and uncovered new frontiers of protein-coding genes, sparking extensive efforts to connect these genes to their respective protein structures[298]. Since then, over 50,000 human protein structures have been documented, however only 35% of these have been successfully matched to a PDB entry. Furthermore, in many instances, these structures only encompass a portion of the protein sequence[299]. The advent of AlphaFold2 allowed to predict structures with a maximum length limit of 2,700 residues. The resulting dataset covered 98.5% of the human proteome with complete chain predictions. 35.7% of the total residues in the human proteome fell within the highest accuracy category. In total, 58.0% of residues received confident predictions (with pLDDT > 70), indicating significant added coverage for sequences lacking suitable templates in the PDB. On a per-protein basis, 43.8% of proteins exhibit confident predictions for at least three-quarters of their sequence, while 1,290 proteins contain substantial regions (comprising more than 200 residues) with pLDDT values of 70 or higher[300]. The predictions were made publicly available and can be accessed at the website alphafold.ebi.ac.uk/. The database is not restricted to the human genome and to date it contains over 214 million structures accounting for 48 complete proteomes.

4.1.6. Faster structure prediction with ColabFold:

A major drawback of AlphaFold is that protein prediction is extremely computationally demanding. To construct diverse Multiple Sequence Alignments (MSAs), extensive collections of protein sequences sourced from publicly available reference and environmental databases must be searched. These searches involve the use of highly sensitive homology detection methods, namely HMMer and HHblits[301], both of which rely on profile hidden Markov models (HMMs)[302]. These environmental databases contain billions of proteins obtained from metagenomic and transcriptomic experiments, often supplementing databases that primarily consist of isolated

genomes. Due to their considerable size, these searches can take several hours for a single protein and demand over 2 terabytes of storage space independently. Second, for the execution of deep neural networks, graphics processing units (GPUs) equipped with substantial GPU RAM (random access memory) are required, even for proteins of relatively common sizes, around ~1,000 residues. To address these high computational needs, an optimised version of AlphaFold has been developed. Since this new version runs on the Google Colaboratory Jupyter Notebook, it has been named ColabFold[197]. This new AI extends beyond AlphaFold2's initial capabilities by enhancing sequence searching, offering tools for modeling homo- and heteromer complexes, unveiling advanced features, broadening the range of environmental databases, and facilitating large-scale batch predictions of protein structures, all achieved with approximately 90-fold increase in processing speed compared to AlphaFold2. To expedite single predictions, ColabFold replaces AlphaFold2's homology search with the significantly faster MMseqs2 (Many-against-Many sequence searching)[303], resulting in a 40–60-fold acceleration. Additionally, batch predictions are accelerated by approximately 90-fold through the avoidance of recompilation and the introduction of an early termination criterion. ColabFold's performance matches AlphaFold2 in terms of prediction quality on CASP14 targets. Furthermore, it equals the prediction quality of AlphaFold-multimer on the ClusPro4 dataset. ColabFold network architecture comprises three key components. The first component is an MMseqs2-based homology search server designed to construct diverse Multiple Sequence Alignments (MSAs) and identify templates. This server efficiently aligns input sequences against databases such as UniRef100, PDB70, and an environmental sequence set. The second part involves a Python library that interacts with the MMseqs2 search server, prepares input features for structure inference (both single chains and complexes), and offers visualization capabilities. This library also includes a command-line interface. The final part consists of Jupyter notebooks for basic, advanced, and batch usage, leveraging a Python library interface[197]. MSA generation is optimized with three key objectives: speed, capturing diversity effectively, and maintaining a manageable size to run on computers with limited RAM. The first goal is achieved through the rapid MMseqs2 prefilter, while the second and third goals are addressed through a search workflow that maximizes sensitivity and a new filter that evenly samples the sequence space. While prediction quality is dependent on the input MSA, often, an MSA with only a few (~30) sufficiently diverse sequences is adequate for producing high-quality predictions.

4.1.7. AlphaFold struggles to predict amino acid mutations:

While ColabFold improved the initial algorithm in terms of speed and storage, AI protein prediction is still not flawless. A significant limitation of AlphaFold is its inability to predict the structural consequences of missense mutations, a fact acknowledged by its developers who stated that AlphaFold "has not been validated for predicting the effect of mutations" (source: https://alphafold.ebi.ac.uk/faq)[300]. Understanding how missense mutations affect protein structure is crucial for unraveling their biological implications. While AlphaFold2's structural prediction algorithm can accurately predict wild-type (WT) protein structures, it struggles to forecast the impact of missense mutations on the three-dimensional (3D) protein structures[304]. AlphaFold2 primarily relies on WT or homologous sequences for its predictions because there isn't a database of structure-disrupting mutations available, and thus, the network hasn't been trained on such data. This limitation is significant as missense mutations are often linked to human diseases, and even a single amino acid change can lead to protein aggregation, misfolding, and functional impairment[305]. Considering that the pLDDT score reflects the confidence in the residue's location within the structure, one might expect it to correlate with changes in protein stability free energy (ΔΔG) or protein function. However, a recent study found a very weak correlation between the difference in pLDDT scores and experimentally determined ΔΔG values (Pearson correlation coefficient, PCC = -0.17). The averaged pLDDT scores for all residues, both individually and when combined with the mutated residue's pLDDT score, showed no correlation either[306]. Similarly, these AlphaFold metrics exhibited a very weak correlation with the impact of single mutations on protein function, particularly in the context of GFP fluorescence. This lack of correlation underscores AlphaFold2's inability to predict the effects of point mutations on protein structure. This limitation likely arises from AlphaFold2's reliance on existing PDB structures rather than fundamental principles governing protein folding. To potentially overcome this limitation, integrating AlphaFold2 with modern advancements in molecular dynamics simulations of protein structure could be explored. Alternatively, establishing a database dedicated to storing structure-disrupting mutations may enable future iterations of AlphaFold2 or other artificial intelligence programs to incorporate this information into their protein-folding predictions. Since AlphaFold struggles to predict the impact of single point mutations, in this experiment the AI tool was employed to investigate the interactions between PDZ proteins and the WT form of PAR4.

4.1.8. Applying AlphaFold for GPCR research:

A crucial research focus in structural biology involves understanding the activation mechanisms of GPCRs (G protein-coupled receptors). Traditionally, progress in this field was hindered by the slow pace of X-ray-based methods. However, the introduction of new techniques like cryo-electron microscopy sparked a significant increase in the number of GPCR structures since 2017[307]. Nevertheless, due to the substantial workload and the inherent flexibility of internal receptors, over 700 GPCR structures have been made available to date[308]. The substantial cost and complexity associated with obtaining GPCR structures pose obstacles to both functional research and drug development related to GPCRs. This underscores the need for a precise and efficient structure prediction algorithm. He et al.[43] conducted a comparative analysis between experimentally determined GPCR structures and the corresponding structure models predicted using AlphaFold2. To ensure the integrity of the analysis, only experimental structures released after the emergence of the AlphaFold2

database were considered. It was found that AlphaFold achieves a high degree of backbone structure prediction, however it is unable to reliably predict residue sidechains conformations, which significantly affect the use of AlphaFold structures for drug design.

4.1.9. Structure prediction for intrinsically disordered proteins:

About 30% of randomly selected sequences, each containing 30 residues or more from the human proteome, are likely to be intrinsically disordered proteins or regions (IDPs/IDRs)[309]. On the AlphaFold Database approximately 30% of residues in predicted structures within the human proteome have pLDDT scores below 50, consistent with expectations for disorder prevalence in the human proteome[300]. While filtering based on pLDDT scores below 50 may overestimate disorder, these scores have been shown to be competitive predictors of disorder when compared to standard methods. In fact, many of the low and very low confidence regions in AlphaFold's predicted proteome coincide with areas expected to be IDRs. Unlike proteins that readily fold into specific three-dimensional structures, IDPs/IDRs are characterized by their conformational variability, which is encoded in their amino acid sequences. They are best described as diverse ensembles of conformations rather than having a single defined structure. This suggests that most of the regions with very low confidence are likely to be intrinsically disordered, rather than being well-defined three-dimensional structures that AlphaFold struggles to predict[310]. It's important to note that a single static "structure," even if classified as a low confidence prediction, cannot adequately represent the diverse ensemble of conformations in IDRs.

4.1.10. PDZ domains structures:

As previously mentioned, PDZ proteins show a conserved structure, being formed by six β-sheets enclosed by two α-helixes. PDZ domains have a central globular binding groove that is located between the α2-helix and β2-sheet. This region contains several conserved amino acid residues (R/K-XXX-G-Φ-G-Φ motif, where X is any amino acid residue and Φ is hydrophobic residues) that form a shallow cleft, also referred to as the “carboxylate-binding loop”[202]. The carboxylate-binding loop enables PDZ domains to interact with the carboxyl tails of other proteins, also referred to as the short linear motif (SLiM)[311]. Moreover, PDZ domains can interact with internal motifs that form β-sheet “finger”-like projections, which mimic carboxyl tail chemistry and insert into the binding groove in a similar manner[312].

4.1.11. Function and structure of potential PAR4 interactors:

The proteomic study highlighted how four PDZ-containing proteins – GIPC1, LMO7, PDLIM1, and TJP2 - could potentially interact with the SLiM domain at the end of PAR4 C-terminal. Before validating these results *in vitro* another layer of screening was adopted, by investigating if AlphaFold would be able to predict these same interactions.

The LIM domain only protein 7 (LMO7) is a large protein made of 1683 amino acids. Its large structure is composed of one Calponin-homolog domain (T^{64}…Q^{181}), four disordered regions (R^{312}…W^{343}, R^{749}…R^{901}, K^{947}…Q^{1037}, K^{1255}…R^{1604}), one PDZ domain (R^{1042}…G^{1128}), and one LIM zinc-binding domain (R^{1612}…G^{1678}). The structure of its PDZ domain and the associated carboxylate binding loop are shown in figure 4.3.

The LIM domain is another important protein-binding domain that is named after the first three proteins in which it was identified: Lin-11, Isl-1, and Mec-3. The LIM domain contains a distinct cysteine-rich motif, and it is composed of two zinc finger structures separated by a two-amino acid spacer which comprise the conserved sequence $CX_2CX_{16-23}HX_2CX_2CX_2CX_{16-21}CX_2(C/H/D)$[313]. All three domains identified within LMO7—namely, the LIM domain, the CH domain, and the PDZ domain—are recognized as protein interaction domains. LMO7 interacts with two F-actin binding proteins, α-actinin and afadin, through its PDZ and LIM domains, respectively[354]. Immunoprecipitation assays have validated these interactions, establishing connections between LMO7 and the afadin-nectin and E-cadherin/catenin systems. The CH domain is predicted to directly bind to actin[314]. LMO7 has also been implicated in regulating the transcription and protein function of emerin, a nuclear membrane protein. Furthermore, LMO7 appears to play a role in the transcriptional regulation of various muscle-relevant genes, including emerin, through a feedback mechanism[315]. Proper functioning of LMO7 is therefore necessary in skeletal and muscular development[316], dysregulation of this protein has been associated with cancer, where LMO7 interact with LRIG proteins leading to prognostic implications for early-stage disease[317]. No interactions with GPCRs have been documented to date.

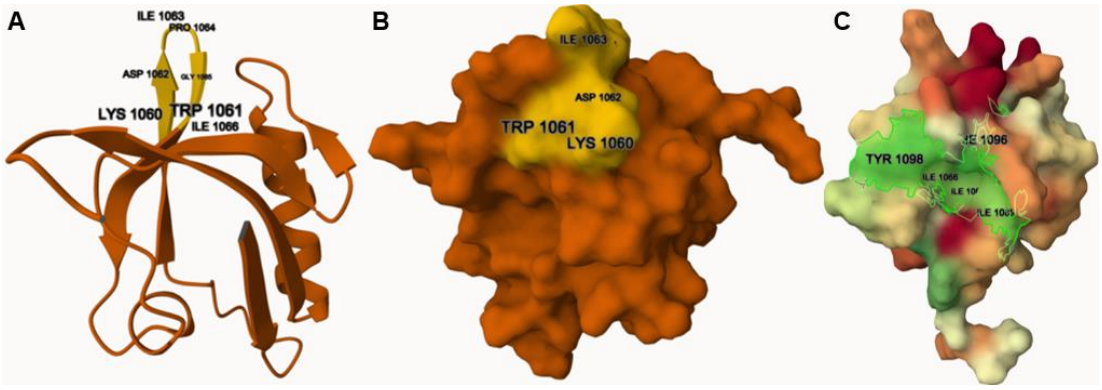


Figure 4.3 – PDZ domain of LMO7. A) Ribbon diagram. B) Molecular surface. C) Hydrophobicity of the molecular surface. Hydrophilic residues are redder, while hydrophobic residues are greener. The carboxylate binding groove is depicted in yellow, while the hydrophobic cleft is highlighted in green.

Another protein displaying a PDZ and LIM domain is PDLIM1, also known as CLP36, Elfin, or CLIM1. It plays a crucial role in cytoskeletal organization, neuronal signalling, and organ development by interacting with various proteins like α-actinin[318], paladin[319], FHL1[320], and EGFR[321]. It belongs to the actinin-associated LIM protein (ALP) subfamily within the PDZ-LIM protein family[322]. Members of the ALP subfamily, including ALP, reversion-induced LIM protein (RIL), Mystique, and PDLIM1, share a PDZ domain at the amino terminus and a LIM domain at the carboxyl terminus[355]. The PDZ domain of PDLIM1 exhibits 55% identity to human ALP and 66% to human RIL, while the LIM domain shows 67% identity to ALP and 57% to RIL[318]. Proteins reported to bind to PDLIM1's PDZ domain include α-actinin[324], p75NTR[325], and the β-catenin/E-cadherin complex[326]. Interaction with these proteins highlighted how PDLIM1 plays important roles in cell proliferation and metastasis during tumour initiation and progression[324–326].

Functioning as a cytoskeleton-associated protein, PDLIM1 plays a role in regulating actin cytoskeleton organization. The dynamic control of stress fibre assembly and disassembly is crucial for cytoskeleton-dependent functions, such as morphological changes and migration[327]. Inhibition of PDLIM1 has been shown to impair focal adhesion assembly and stress fibre formation, impacting the normal function of Sertoli cells, and ordered spermatid differentiation during spermiogenesis[328]. PDLIM1 may also influence neurite growth, as demonstrated by increased neurite cell growth rates in rat primary dorsal root ganglion neurons upon PDLIM1 inhibition[329]. Furthermore, PDLIM1 is implicated in modulating p65 nuclear translocation, participating in the regulation of the NF-κB pathway[356].

Structurally, the PDZ domain is present at the C-tail of the protein, between the residues T^{3}…E^{85} (Figure 4.4), while the LIM zinc-binding domain is present within the positions P^{285}…P^{317}.

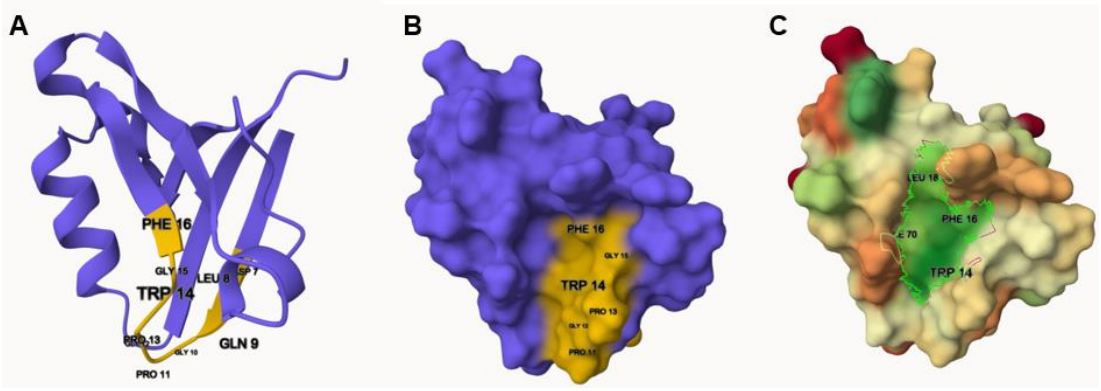


Figure 4.4 – PDZ domain of PDLIM1. A) Ribbon diagram. B) Molecular surface. C) Hydrophobicity of the molecular surface. Hydrophilic residues are redder, while hydrophobic residues are greener. The carboxylate binding groove is depicted in yellow, while the hydrophobic cleft is highlighted in green.

The next interaction found in the proteomic study was with tight junction protein ZO2 (TJP2). Tight junctions (TJs) are crucial components of cell-cell junctions in epithelial and endothelial tissues, playing a fundamental role in maintaining the integrity of barriers between cells. These junctions create selectively permeable barriers that regulate the passage of ions, molecules, and cells across epithelial and endothelial cell layers, thus contributing to tissue homeostasis. Tight junctions are essential for the proper functioning of various physiological processes, including the establishment of tissue barriers, maintenance of cell polarity, and regulation of paracellular transport[330]. At least 40 different proteins are responsible for the formation of TJs, which can be divided based on function and structural features[331]. Tight junction protein TJP2 belongs to the Membrane-associated guanylate kinase homologs (MAGUK) family, alongside with TJP1 and TJP3. The role of these proteins is to establish a connection between membrane proteins in tight junctions and the cytoskeleton[332]. Thus, they structurally facilitate the formation of multiprotein complexes on the cytoplasmic side of TJs. These proteins share the structural feature of possessing 3 PDZ domain, which help them link other TJ proteins, such as claudins, to assemble multicompetent TJ protein complexes[333].

All the proteins found by the proteomic study had only one PDZ domain, except for TJP2, which has a very large structure of 1190 amino acids and possesses three PDZ domains. This protein is formed by four disordered regions (R^{152}…I^{306}, I^{408}…Y^{506}, S^{920}…L^{1079}, N^{1105}…L^{1190}), three PDZ domains (T^{33}…R^{120}, G^{307}…S^{385}, N^{509}…S^{590}), one SH3 domain (G^{604}…A^{669}), one guanylate kinase-like domain (N^{678}…Q^{876}), and one interaction site with SCRIB ($T^{1188}EL^{1190}$). Each of the three PDZ domains is shown in depicted in figure 4.7. The second PDZ domain (PDZ2) of TJP-proteins is recognized for its ability to facilitate protein dimerization. This structural adjustment preserves the canonical peptide-binding groove in both subunits of the PDZ2 dimer, comprising elements from each monomer[357]. The resolved crystal structure of two homodimerized PDZ2, obtained from Chen *et al.*[335] (PDB entry: 3E17) is shown in figure 4.5.

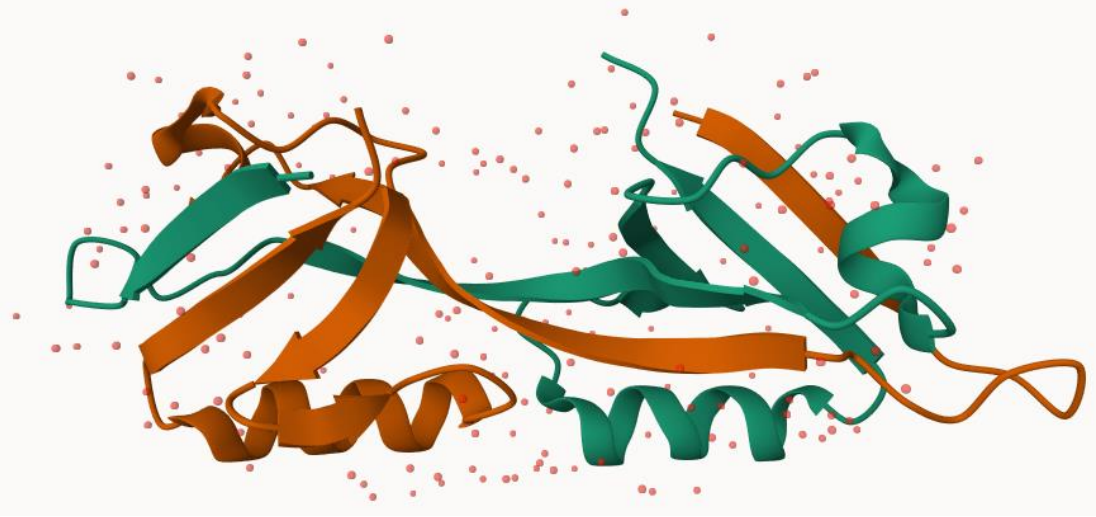


Figure 4.5 – Crystal structure of the homodimerised PDZ2 of TJP2. PDB entry: 3E17. Figure reported with permission of the author[335].

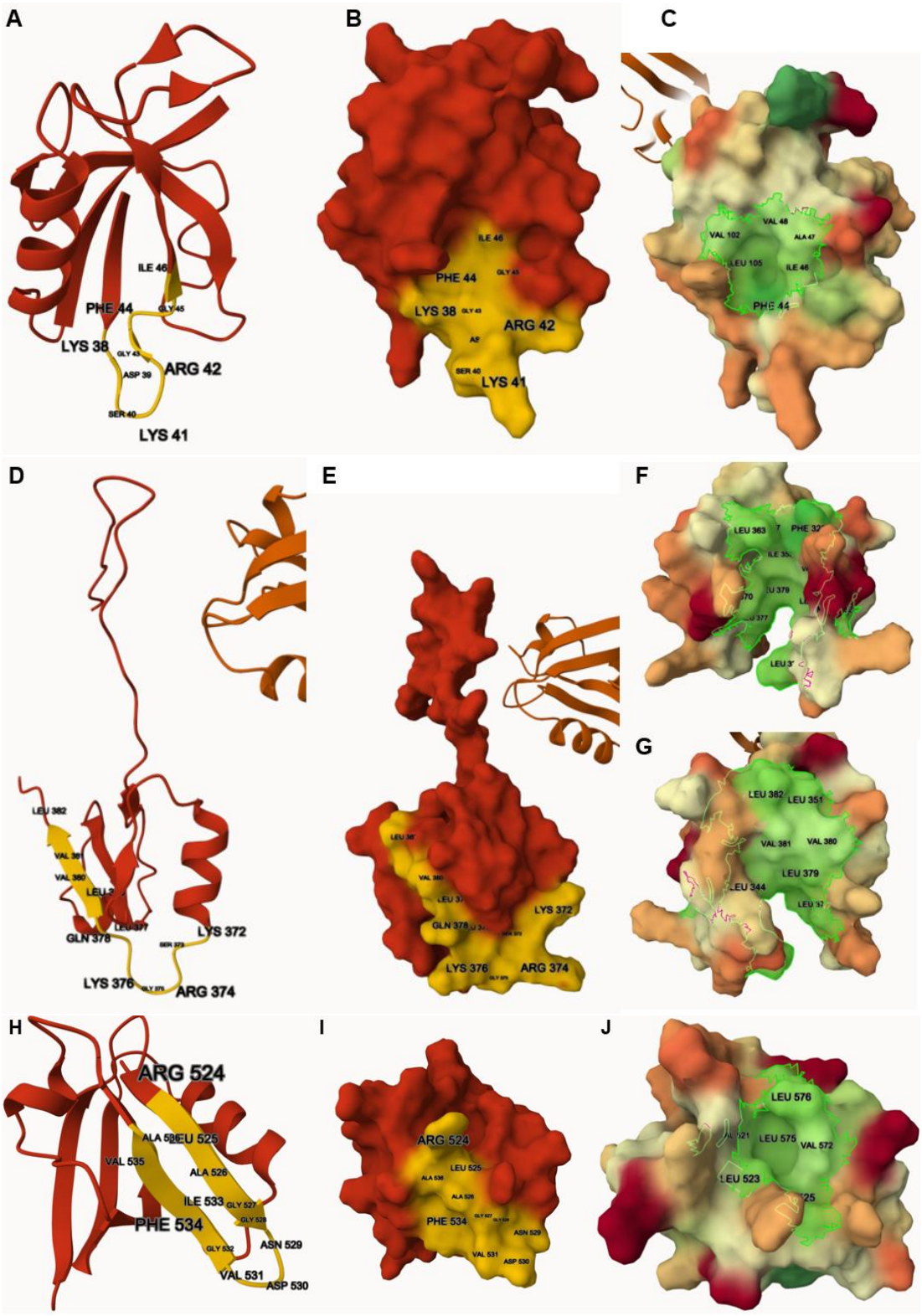


Figure 4.6 – PDZ domains of TJP2. This was the only potential interaction to display three PDZ domains. A-D-H are the respectively the ribbon diagram, molecular surface and hydrophobicity of domain 1. D-E-F-G depict domain 2, which has a belt of hydrophobic residues forming a hydrophobic cleft all around the domain. H-I-J show PDZ domain 3.

The GIPC family is constituted by the three proteins, GIPC1 (also known as GIPC, or RGS19IP1), GIPC2 and GIPC3.The three protein show a certain degree of similarity, as demonstrated by their total amino-acid-identity: GIPC1 and GIPC2 have 62%, GIPC1 and GIPC3 59.9%, while GIPC2 and GIPC3 show 55.3% identity. These similarities are shown also in the same domain organisation, as GIPCs proteins are composed by a central PDZ domain, surrounded by two GIPC homology domains (GH1 and GH2). Members of the GIPC family exhibit evolutionary conservation and functional similarity[336]. They play a role in the trafficking of diverse transmembrane proteins and oversee a range of cellular processes, including mitochondrial fission[358], proliferation[359], planar cell polarity[339], cytokinesis[340], and migration[359]. Pathologies linked to GIPCs dysfunction include hearing loss[341] and cancer[359].

Out of the three proteins, only GIPC1 was found to interact with PAR4 in the proteomic analysis. GIPC1 is 333 amino acids long, and presents two disordered domains ($M^{1}…L^{54}$, $Q^{223}…R^{244}$), and one acyl carrier protein (ACP) domain ($A^{264}…$ D^{320})[360]. Its PDZ domain is present in the region $E^{133}…E^{213}$, which is highlighted in figure 4.7.

The physiological roles of GIPC1 have been extensively characterized. GIPC1's N-terminal region plays a role in dimerization, while its C-terminal region is implicated in interacting with the retrograde motor protein, myosin VI (MYO6)[361]. Additionally, the PDZ domain of GIPC1 engages in direct interactions with various transmembrane proteins, including GPCRs. The interactions documented to date are: adrenergic receptor β1 (ADRB1)[362], APPL1[345], CD93 (C1QR1)[346], dopamine receptor D2 (DRD2)[348] and D3 (DRD3)[347], endoglin[363], Glucose transporter 1 (GLUT1)[350], Insulin-like growth factor 1 receptor (IGF1R)[351], integrin α5 and α6[352], luteinizing hormone/choriogonadotropin receptor (LHCGR)[353364]

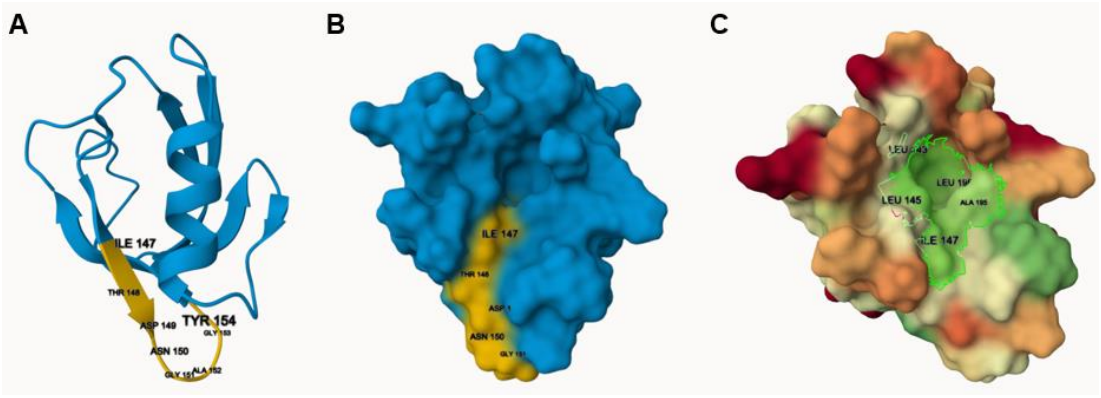


Figure 4.7 – PDZ domain of GIPC1. A) Ribbon diagram. B) Molecular surface. C) Hydrophobicity of the molecular surface. Hydrophilic residues are redder, while hydrophobic residues are greener. The carboxylate binding groove is depicted in yellow, while the hydrophobic cleft is highlighted in green.

4.2. CHAPTER AIM:

The aim of this chapter is to employ advanced computational tools, including AlphaFold Multimer and PrePPI, to predict and evaluate potential interactions between PAR4 and its proteomic partners (GIPC1, PDLIM1, LMO7, and TJP2). These tools are used to assess the structural feasibility of interactions at both the broader protein-protein interface and the specific interactions between the C-terminal short linear motif (SLiM) of PAR4 and the PDZ domains of the identified partners. This computational analysis aims to refine the list of high-confidence interaction candidates, guiding experimental efforts and advancing the understanding of PAR4's interactome.

We hypothesise that:

1. PAR4 interacts with specific proteomic partners (GIPC1, PDLIM1, LMO7, and TJP2) through its SLiM and PDZ domains. This hypothesis is investigated using AlphaFold Multimer to predict structural interactions and evaluate the molecular feasibility of these binding events.
2. PrePPI can identify additional novel interactors of PAR4, complementing findings from proteomics and AlphaFold. This analysis broadens the scope of potential interaction partners, providing a comprehensive overview of PAR4's binding capabilities.
3. Integrating predictions from AlphaFold Multimer and PrePPI allows for the prioritization of high-confidence candidates for future experimental validation. This computational screening serves as an efficient and reliable method to narrow down the list of interactors.

4.3. RESULTS:

As specified in the methods section, AlphaFold predictions were set to generate five models. As outlined by Lee *et al.*[296], it is possible to increase the accuracy of AlphaFold multimer predictions, by fragmenting the proteins and investigating if short linear motifs interact with specific domains on a target protein, rather than using full structures. This is perfectly suitable in this scenario, as the only SLiM to investigate was the terminal C-end of PAR4. For this reason, the predictions for the four proteins were performed twice, the first time the full structure of both PAR4 and the PDZ protein of interest, and then using only the extreme end of PAR4 ($S^{381}SLLQ^{385}$) and the PDZ domain of each protein.

4.3.1. LIM domain only protein 7 (LMO7):

Despite being found on the proteomics study, AlphaFold did not manage to predict any interaction between the PAR4 SLiM and the PDZ domains on LMO7 and PDLIM1. In the case of LMO7, the large structure of the protein, and the presence of many disordered regions, heavily impacted the reliability of the results, as across all five models this prediction achieved the lowest mean pLDDT scores, yet interestingly it achieved the highest ipTM scores when compared to the other proteins (table 4.1). Model 1 in particular achieved an ipTM = 0.50, meaning that the interface prediction would have mostly been in the same fold of a potential template model. Regardless, the PDZ domain was not found in proximity of PAR4 SLiM in any of the 5 models (Figure 4.8).

Table 4.1 – Prediction scores for interactions with LMO7.

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	35.10	0.32	0.50	N/A	N/A
2	36.69	0.31	0.37	N/A	N/A
3	35.79	0.31	0.31	N/A	N/A
4	38.29	0.31	0.27	N/A	N/A
5	37.52	0.31	0.24	N/A	N/A

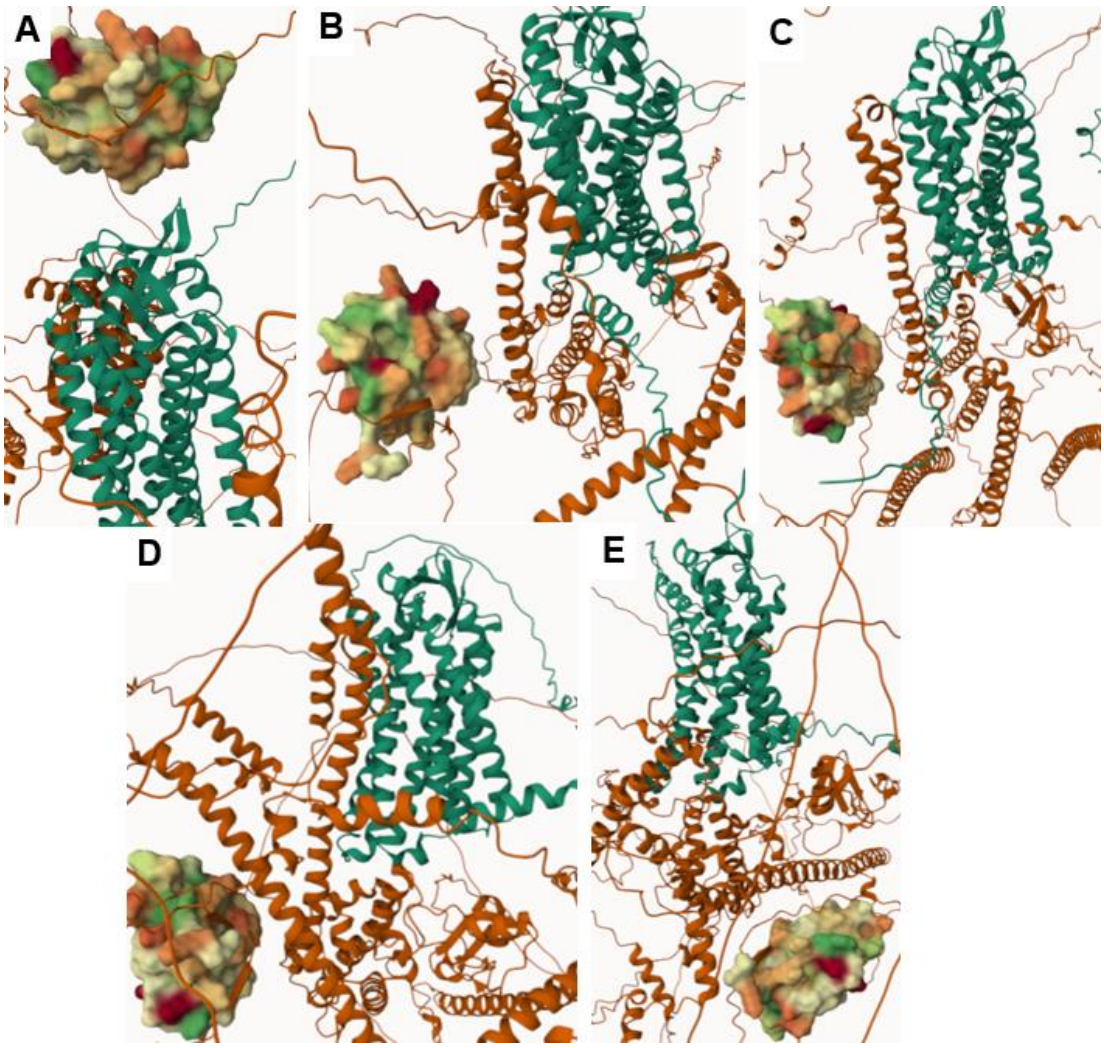


Figure 4.8 – No AlphaFold model predicted PAR4 SLiM to interact with the PDZ domain on LMO7. Ribbon diagram of the receptor bound to LMO7. PAR4 is shown in green, LMO7 in orange. The PDZ domain is shown as the molecular surface coloured based on the hydrophobicity. **A:** Model 1. **B**: Model 2. **C**: Model 3. **D:** Model 4. **E:** Model 5. Image generated using Mol*.

Due to its large structure, and the presence of many intrinsically disordered domains, predicting only LMO7 PDZ domain together with PAR4 SLiM greatly improved the metrics score for this protein. The mean pLDDT was particularly improved, highlighting higher accuracy of each amino acid's position within the protein structure. The pTM was also relatively high, meaning the model's overall fold matched with the topology of known structures. However, the ipTM was particularly low, therefore the accuracy of the interface between the two sequences in complex is not reliable (Table 4.2). This becomes apparent when looking at Figure 4.9, where the top ranked model did not even manage to predict PAR4 SLiM to bind on the correct site. Significant variability was observed in the predictive outcomes, with each model failing to identify identical hydrogen bonding patterns. Therefore, AlphaFold does not predict PAR4 SliM to bind with LMO7 PDZ domain.

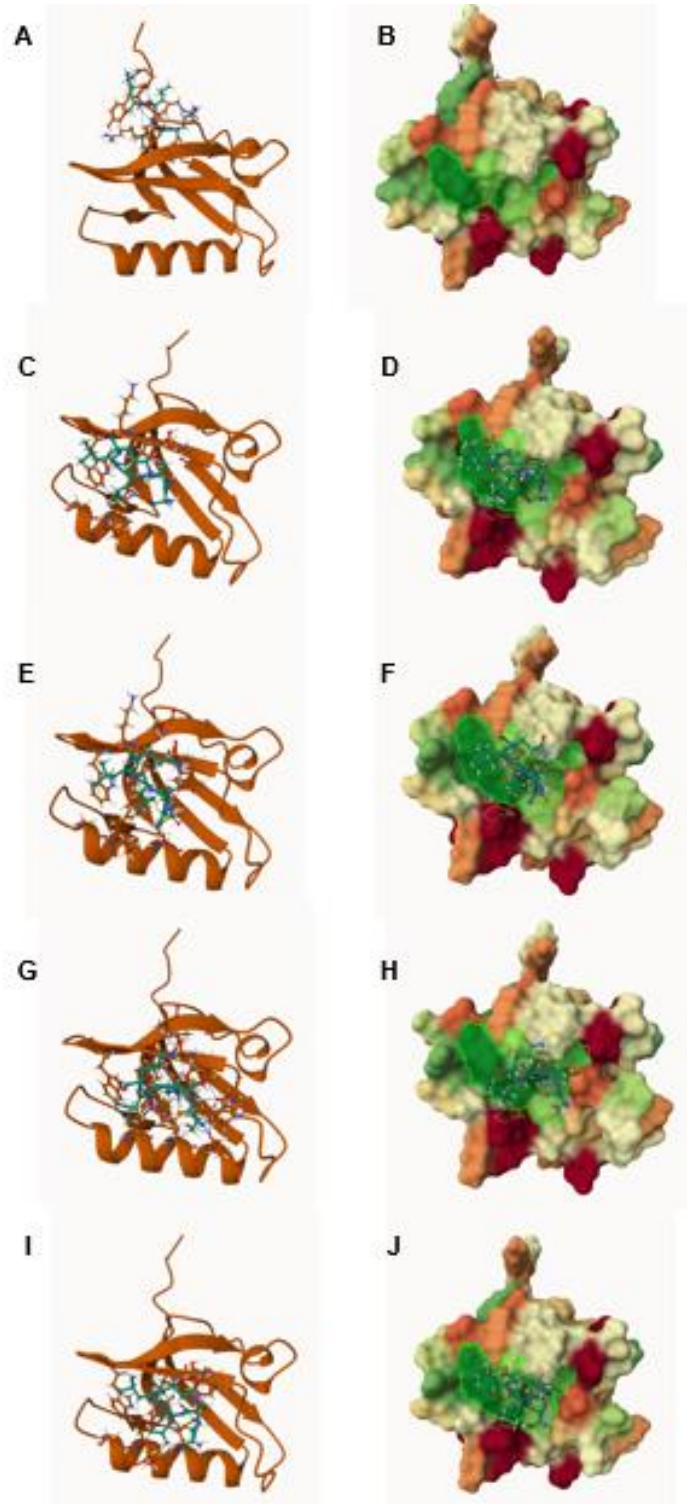


Figure 4.9 – PAR4 SLiM does not interact with LMO7 PDZ domain. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.2 - Prediction scores for PAR4 SLiM with LMO7 PDZ domain

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	84.57	0.76	0.29	S^{381} – G^{1125} S^{381} – G^{1125} L^{383} – R^{1123}	Y^{1124}
2	80.81	0.73	0.24	N/A	F^{1057}, I^{1059}, W^{1061}, W^{1104}, M^{1108}
3	83.65	0.77	0.23	L^{383} – I^{1059}	F^{1057}, I^{1059}, W^{1061}, W^{1104}, M^{1108}
4	84.71	0.78	0.21	S^{381} – Q^{1112} S^{381} – D^{1054} S^{382} – D^{1054}	F^{1057}, I^{1059}, W^{1061}, W^{1104}, M^{1108}
5	80.85	0.74	0.21	S^{381} – W^{1104} L^{383} – I^{1059}	F^{1057}, I^{1059}, W^{1061}, W^{1104}, M^{1108}

4.3.2. PDZ and LIM domain 1 (PDLIM1):

Concerning PDLIM1, the smaller structure and the absence of disordered domains translated to higher confidence metrics showing mean pLDDT scores in the range of 55.70 – 60.93, which are shown in table 4.3. Similarly, to GIPC1, the predictions achieved a pTM score near the cutoff for being in the same fold of a template structure, but the ipTM were drastically lower, remarking the low confidence of the AI in predicting the interfaces between the two proteins. Again, no interactions were found between PAR4 SLiM and the PDZ domain of PDLIM1 (figure 4.10). This was found either unbound to PAR4 (models 1 and 5) or bound to the extracellular N term of the receptor (models 2,3, and 4). Model four in particular is a clear example of an AlphaFold hallucination, as the PDZ domain was predicted to assume a different structure, with the hydrophobic pocket forming a tunnel for the N terminus of PAR4 slip through (Figure 4.10-D). This caused the formation of many hydrogen bonds and hydrophobic interactions which are reported in table 4.4.

Table 4.3 – Prediction scores for interactions with PDLIM1.

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	60.93	0.49	0.21	N/A	N/A
2	58.40	0.50	0.20	E^{28} – W^{14} S^{24} – Q^{67}	W^{14}, F^{16}, L^{18}, I^{70}
3	55.70	0.48	0.18	S^{31} – V^{53} L^{63} – Q^{26}	N/A
4	57.53	0.48	0.18	S^{24} – Q^{67} Y^{26} – L^{18} E^{28} – G^{15} G^{30} – S^{37} S^{31} – S^{37} S^{31} – P^{13} S^{31} – K^{38} T^{32} – K^{38} G^{34} – L^{44} G^{34} – R^{83} D^{36} – Q^{4} D^{36} – V^{81} S^{38} – M^{1}	L^{8}, W^{14}, F^{16}, L^{18}, L^{44}, V^{49}, I^{50}, I^{70}, V^{81}
5	59.67	0.48	0.14	N/A	N/A

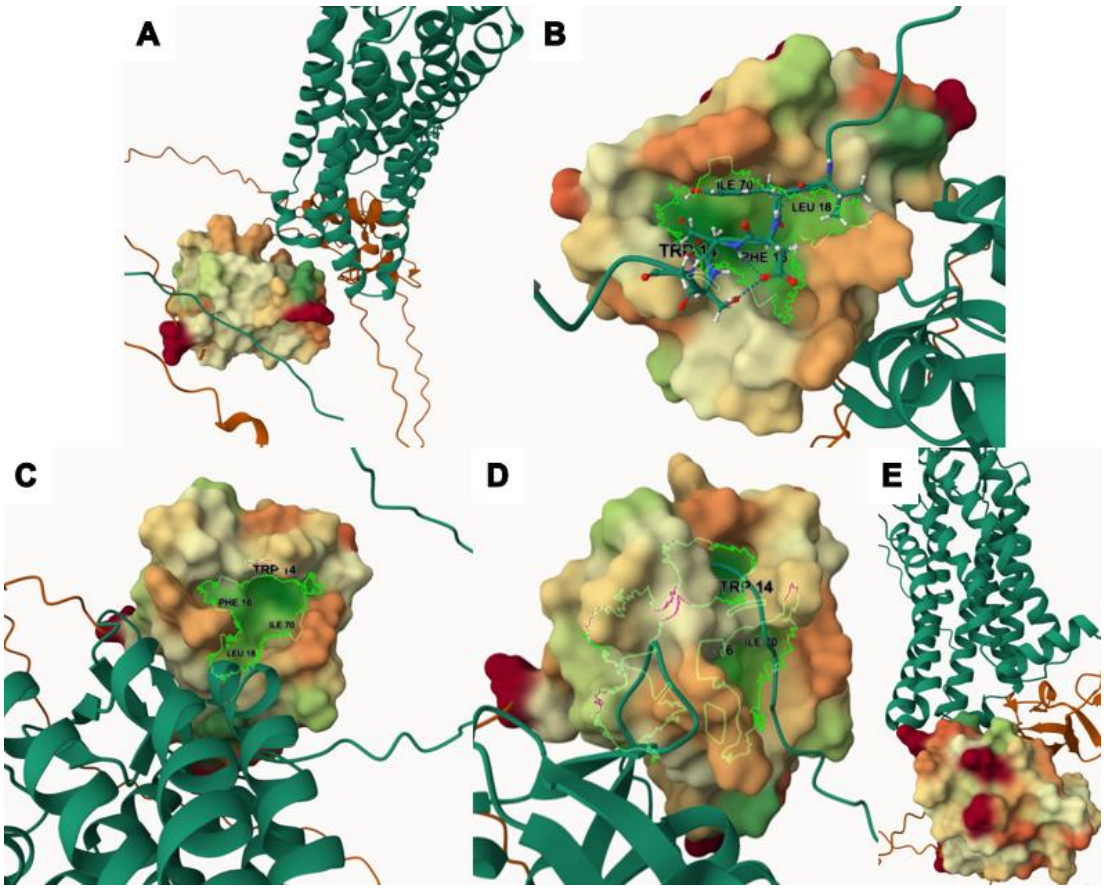


Figure 4.10 – No AlphaFold model predicted PAR4 SLiM to interact with the PDZ domain on PDLIM1. Ribbon diagram of the receptor bound to LMO7. PAR4 is shown in green, PDLIM1 in orange. The PDZ domain is shown as the molecular surface coloured based on the hydrophobicity. **A:** Model 1. **B**: Model 2. **C**: Model 3. **D:** Model 4. **E:** Model 5. Image generated using Mol*.

The prediction accuracy of the interaction with PDLIM1 was also increased, as with the previous cases. As shown in Table 4.4, the mean pLDDT and pTM scores were consistently found to be above 90 and 0.80 respectively, highlighting the quality of the predicted structures. However, as for LMO7, the ipTM score was significantly lower, despite the SLiM being found always in proximity of the hydrophobic cleft (Figure 4.11). This warrants caution in interpreting the interactions between the two proteins. No hydrogen bond was found consistently across models. S^{382} – K^{71} was present in the top 3 models, while Q^{385} – F^{16} was the only interaction common to the fourth and fifth ranked predictions. The low ipTM scores, together with the inconsistencies in determining hydrogen bonds, suggest that PAR4 SLiM might not bind the PDZ domain on PDLIM1.

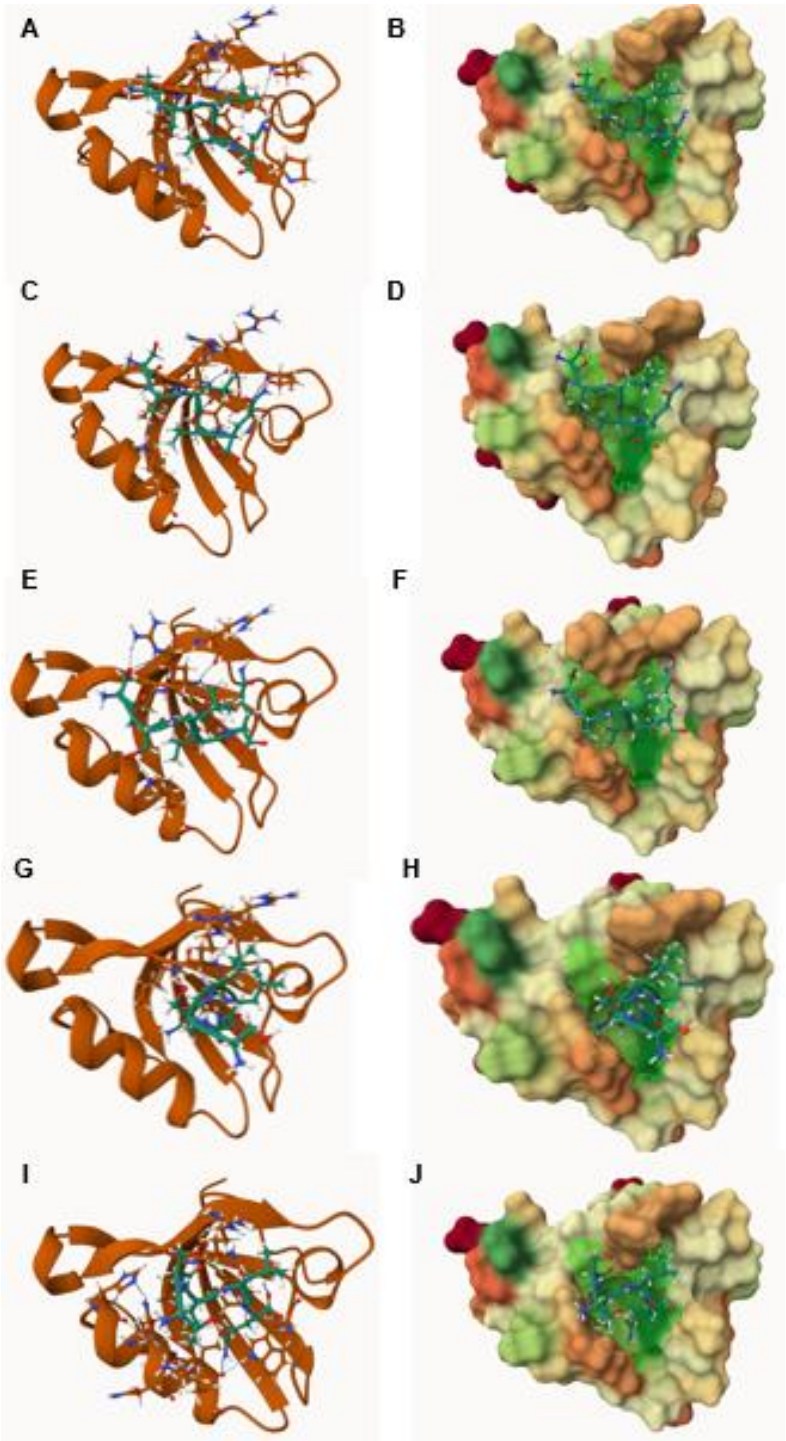


Figure 4.11 – AlphaFold did not find the correct fit for PAR4 SLiM in PDLIM1 PDZ domain. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.4 - Prediction scores for PAR4 SLiM with PDLIM1 PDZ domain

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	92.32	0.84	0.47	S^{381} – L^{18} S^{382} – K^{71} L^{384} – F^{16}	W^{14}, F^{16}, L^{18}, I^{70}
2	92.22	0.85	0.46	S^{382} – K^{71} L^{384} – F^{16}	W^{14}, F^{16}, L^{18}, I^{70}
3	92.89	0.85	0.46	S^{381} – R^{17} S^{381} – R^{17} S^{382} – K^{71} L^{384} – F^{16}	W^{14}, F^{16}, L^{18}, I^{70}
4	92.29	0.85	0.42	Q^{385} – F^{16}	W^{14}, F^{16}, L^{18}, I^{70}
5	92.13	0.85	0.39	Q^{385} – W^{14} Q^{385} – F^{16} Q^{385} – I^{70}	W^{14}, F^{16}, L^{18}, I^{70}

4.3.3. Tight junction protein 2 (TJP2):

Predictions are not consistent over different AlphaFold models of TJP2. As previously mentioned, the tight junction protein ZO-2 (TJP2) is a large structure composed of three distinct PDZ domains. When investigated for potential bindings with PAR4 using AlphaFold, only two models found weak interactions with the receptor's SLiM. Moreover, the bindings were found at different PDZ domains.

In this case, all five models achieved lower mean pLDDT scores compared to GIPC1. This is owed to the larger structure of TJP2 and the presence of multiple disordered regions. This also translated to slightly lower pTM values, but interestingly these predictions achieved higher ipTM scores. Regardless, the ipTM scores were still low, suggesting caution when interpreting the interactions between the two proteins. Confidence metrics and residues important for protein-protein interactions are reported in table 4.5.

Table 4.5 – Prediction scores for interactions with TJP2.

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	52.90	0.36	0.30	S^{382} – V^{48} L^{384} – I^{46}	F^{44}, I^{46}, A^{47}, V^{48}, V^{102}, L^{105}
2	52.67	0.37	0.26	N/A	N/A
3	50.64	0.35	0.21	L^{383} – T^{1188} Q^{385} – T^{1188} Q^{385} – E^{1189}	N/A
4	53.22	0.36	0.19	N/A	N/A
5	51.40	0.36	0.19	N/A	N/A

Model one (figure 4.12) was the best prediction, suggesting that PAR4 SliM might interact with TJP2 PDZ domain 1. In this model, only two hydrogen bonds were found (S^{382} – V^{48}, and L^{384} – I^{46}), which however resulted in the C-tail being slightly displaced from the expected hydrophobic pocket, as highlighted in figure 4.12-D. In this case, SLiM appears to sit on the edge of the cleft, with the residue L^{384} projecting towards the pocket, but not sitting tightly within it as happened in the case of GIPC1 model one. Regardless, this is still the best prediction of interactions between PAR4 SLiM and TJP2 PDZ domains, as models two, four and five did not find any binding across any PDZ domain (supplementary figure S6). Model three (supplementary figure S5) managed to locate the receptor's C tail in close proximity of PDZ 3. In this case however, hydrogen bonds were not formed with any residues of this domain. Instead, a series of interactions with the terminal side of a disordered region, prevented the binding on the PDZ domain.

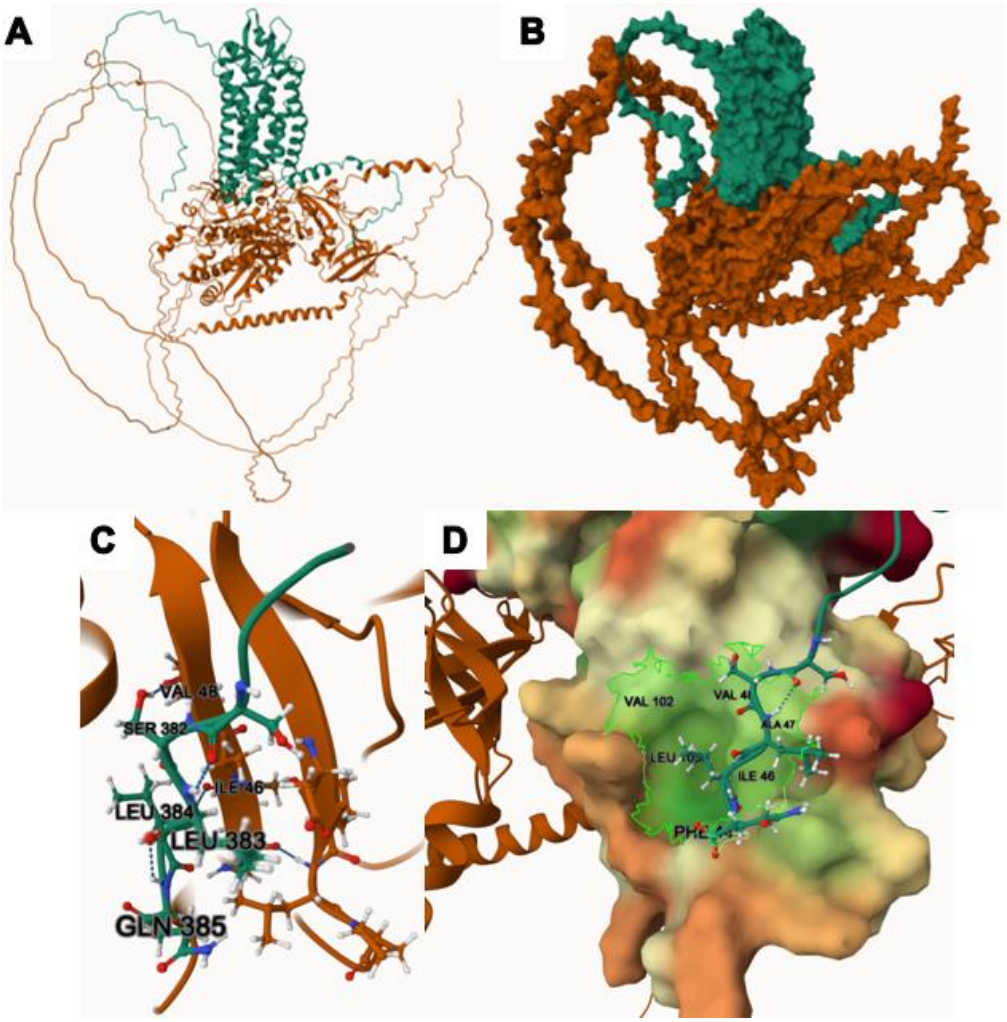


Figure 4.12 – The SLiM of Model 1 is slightly displaced from the hydrophobic pocket A: Ribbon diagram of the receptor bound to TJP2. PAR4 is shown in green, TJP2 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and residues on PDZ domain 1. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a the PDZ domain, however the binding does not occur in the hydrophobic pocket, which is highlighted in green. Image generated using Mol*.

TJP2 was the largest of the proteins studied, presenting multiple intrinsically disordered regions and three PDZ domains. To investigate whether PAR4 would bind this protein, each PDZ domain was considered individually. Again, excluding the large intrinsically disordered regions greatly improved the quality of the mean pLDDT and pTM scores for each PDZ domain, however the ipTM scores were significantly lower, especially in the case of PDZ domain 2 (Table 4.7) and PDZ domain 3 (Table 4.8). PDZ domain 1 achieved the highest ipTM scores, as model 1 showed a value of 0.66. Moreover models of this region consistently found the same hydrogen bonds across all five models. S^{382} – V^{48} was flagged in every prediction, while L^{384} – I^{36} also featured across the board except model two, where the isoleucine at position 36 was found to interact with the last glutamine residue instead (Q^{385} – I^{36}). As this large protein possesses 3 PDZ domains, 5 models for each domain were generated to better study the interactions. Figure 4.13 depicts domain 1, figure 4.14 shows domain 2, while figure 4.15 refers to domain 3.

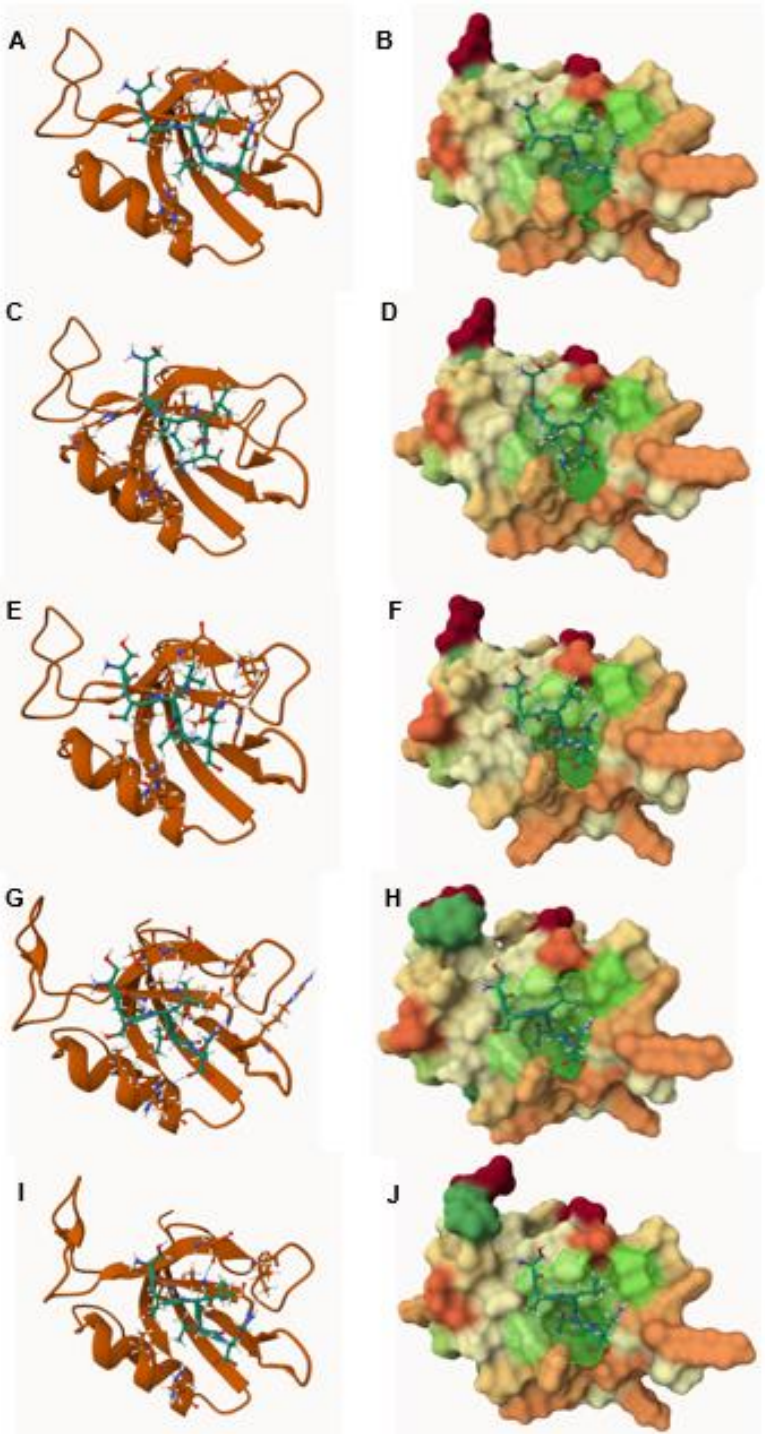


Figure 4.13 – AlphaFold finds different binding configurations for PAR4 SLiM and TJP PDZ domain 1. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.6 - Prediction scores for PAR4 SLiM with TJP2 PDZ domain 1

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	91.65	0.83	0.66	S^{382} – V^{48} S^{382} – V^{48} L^{384} – I^{36}	F^{44}, I^{46}, A^{47}, V^{48}, V^{102}, L^{105}
2	89.31	0.81	0.50	S^{382} – V^{48} Q^{385} – I^{36}	F^{44}, I^{46}, A^{47}, V^{48}, V^{102}, L^{105}
3	83.51	0.74	0.46	S^{382} – V^{48} L^{384} – I^{36}	F^{44}, I^{46}, A^{47}, V^{48}, V^{102}, L^{105}
4	83.72	0.74	0.45	S^{382} – V^{48} L^{384} – I^{36} Q^{385} – R^{106}	F^{44}, I^{46}, A^{47}, V^{48}, V^{102}, L^{105}
5	85.56	0.77	0.43	S^{382} – V^{48} S^{382} – V^{48} L^{384} – I^{36}	F^{44}, I^{46}, A^{47}, V^{48}, V^{102}, L^{105}

TJP2 PDZ domains 2 and 3 achieved much lower ipTM scores and the predicted hydrogen bonds were very inconsistent across models. In the case of PDZ domain 2, model 4 and 5 failed to find PAR4 SLiM in proximity of the binding site. PDZ domain 3 did not encounter this issue, however the PAR4 C-tail peptide was often found displaced, despite being next to the hydrophobic cleft. These notions suggest that PAR4 most likely does not interact with the PDZ regions present on TJP2.

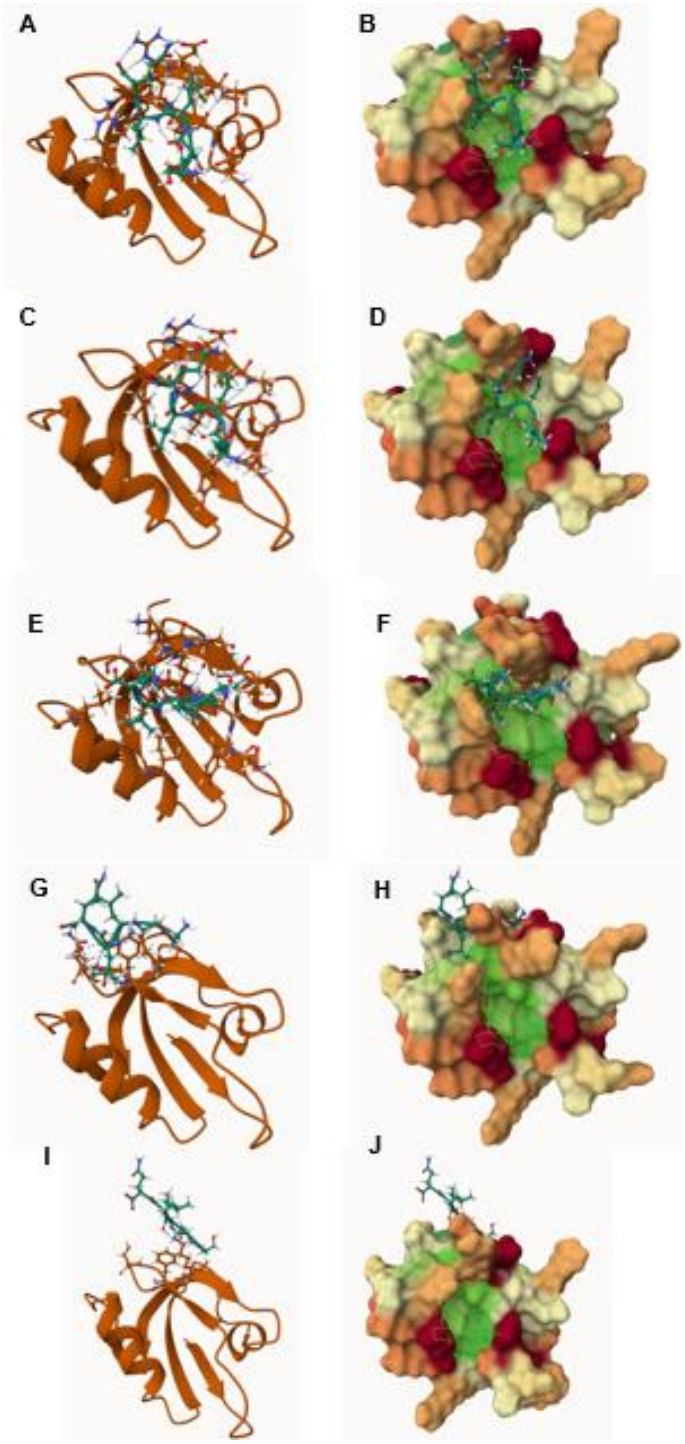


Figure 4.14 – AlphaFold does not predict the binding of PAR4 SLiM with TJP2 PDZ domain 2. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.7 - Prediction scores for PAR4 SLiM with TJP2 PDZ domain 2

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	87.25	0.77	0.27	S^{381} – E^{318} L^{384} – R^{367} Q^{385} – R^{322} Q^{385} – R^{322}	Y^{319}, L^{321}, L^{323}, I^{370}
2	88.35	0.80	0.23	S^{381} – E^{318}	Y^{319}, L^{321}, L^{323}, I^{370}
3	90.03	0.81	0.21	S^{381} – E^{331} L^{383} – L^{323} Q^{385} – R^{322}	Y^{319}, L^{321}, L^{323}, I^{370}
4	88.48	0.79	0.13	S^{382} – K^{330} S^{382} – K^{330} L^{383} – S^{325} L^{383} – S^{325} L^{384} – K^{330} Q^{385} – S^{325}	F^{328}
5	88.26	0.78	0.08	L^{383} – K^{330}	F^{328}

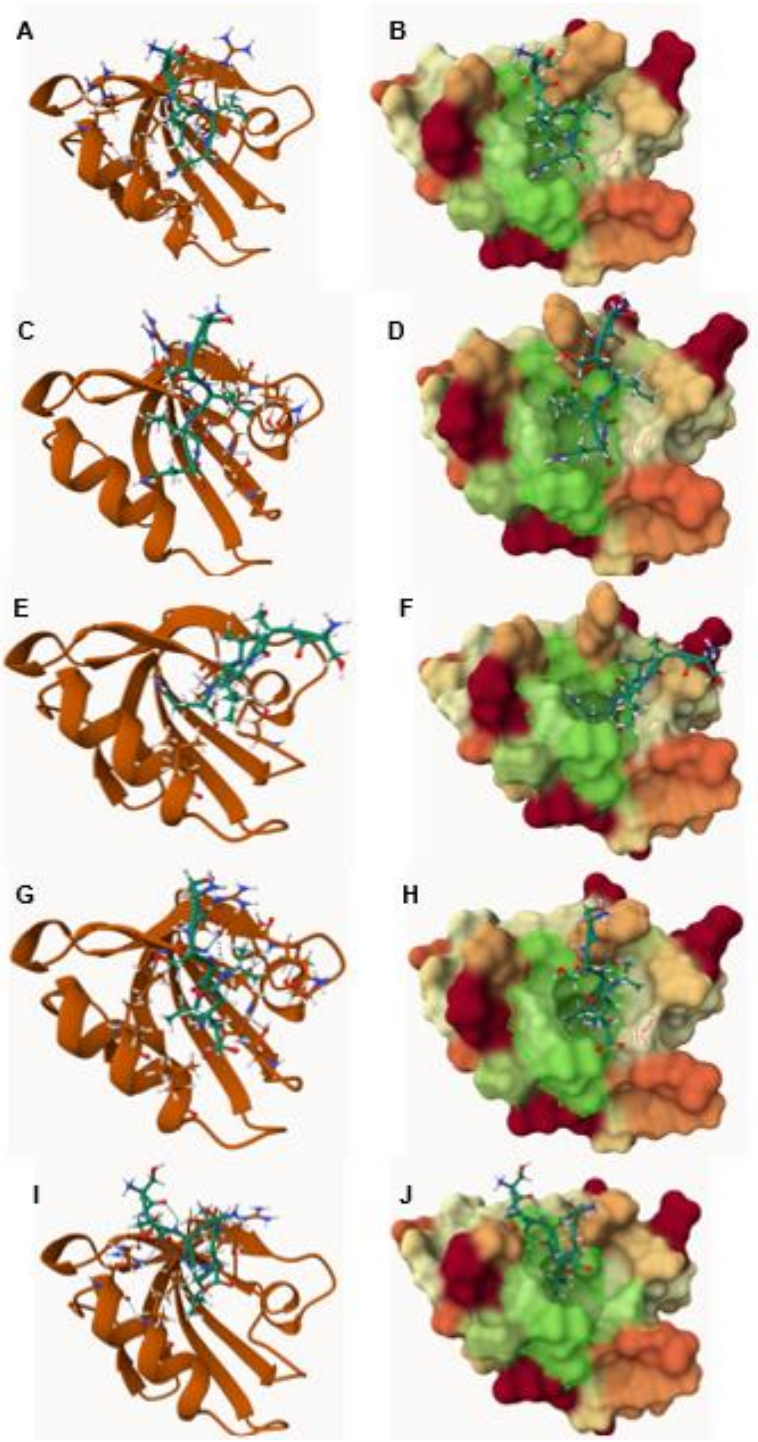


Figure 4.15 – AlphaFold does not find consistent binding between PAR4 SLiM and TJP2 PDZ domain 3. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.8 - Prediction scores for PAR4 SLiM with TJP2 PDZ domain 3

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	89.02	0.82	0.35	Q^{385} – L^{523}	V^{524}, L^{526}, L^{528}, V^{572}, L^{575}, L^{576}
2	92.32	0.85	0.34	S^{382} – R^{524} S^{382} – R^{524} L^{384} – L^{523} L^{384} – L^{523}	V^{524}, L^{526}, L^{528}, I^{538}, V^{572}, L^{575}, L^{576}
3	92.13	0.85	0.28	L^{384} – S^{520}	V^{524}, L^{526}, I^{538}, L^{575}
4	91.52	0.84	0.26	L^{384} – L^{523}	V^{524}, L^{526}, L^{528}, L^{575}
5	89.08	0.81	0.25	S^{382} – R^{568} L^{383} – R^{568}	L^{526}, L^{528}, L^{575}

4.3.4. GAIP interacting protein C-terminus (GIPC1):

Predictions for interactions with GIPC1 achieved the highest pLDDT across the five models when compared to other proteins. This can be owed to the relatively small structure of GIPC1 and the absence of disordered regions which negatively impact the overall confidence of the position of each residue. The 5 models have similar pTM and ipTM scores too. Interestingly, the pTM value was nearly 0.5 for each model, which is also the cut off score for the protein being in the same fold as the template model. However, the ipTM was drastically lower, achieving scores in the range of 0.15-0.24. As previously stated, 0.17 is indicated as the threshold for the structure being randomly paired to its template model. This means that AlphaFold is relatively confident of the folding of each protein and of the complex, but much less so for the interaction of the two molecules. The average pLDDT, pTM, ipTM scores as well as the hydrogen bonds and hydrophobic interactions for each model are reported in table 4.9.

Table 4.9 – Predictions scores for interactions with GIPC1.

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	65.96	0.48	0.24	S^{381} – I^{147} L^{384} – L^{145} Q^{385} - K^{199}	A^{142}, L^{143}, L^{145}, I^{147}, A^{195}, L^{198}
2	66.59	0.47	0.22	G^{18} - V^{79} S^{24} – H^{191} Y^{26} – I^{147} Q^{28} – K^{199} S^{29} -K^{161}	L^{143}, L^{145}, I^{147}, A^{195}, L^{198}
3	65.29	0.48	0.17	S^{382} - D^{141} Q^{385} – R^{203}	N/A
4	65.69	0.49	0.16	S^{382} – R^{203} L^{384} – R^{203}	N/A
5	63.87	0.47	0.15	N/A	N/A

Only the top ranked model managed to predict interactions between PAR4 SLiM ($S^{381}SLLQ^{385}$) with the hydrophobic pocket next to the carboxylate binding loop on GIPC1 PDZ domain. This was the only model to predict the formation of hydrogen bonds between the residues S^{381} – I^{147}, L^{384} – L^{145}, and Q^{385} - K^{199}. This last bond seems to hold the SLiM locked in place, while the adjacent L^{384} is oriented in such a way that sits deep within the hydrophobic cleft, as shown in Figure 4.16-D. This prediction is also the only one where PAR4 SLiM is well aligned with the globular groove positioned between the α2-helix and β2-strand of the PDZ domain (Figure 4.16-C), where the binding is expected to happen.

Model number 1 (figure 4.16) was the only prediction to find interactions at the expected binding site on the PDZ domain of GIPC1. Model 2 (supplementary Figure S1) also found interactions between the two proteins, however AlphaFold predicted the PDZ domain to bind on the extracellular region of the receptor. Two interactions were found within the binding groove between the α2-helix and β2-sheet of the PDZ domain, being S^{24} – H^{191} and Y^{26} – I^{147}. The interactions were also found aligned with the hydrophobic cleft on the PDZ domain of GIPC1. However, the interactions were not predicted to happen with the SLiM domain at the very end of the receptor C-terminal.

Models three and four (supplementary Figures S2 and S3) showed similar predictions. In this case, GIPC1 PDZ domain was found on the intracellular side of PAR4, and it also formed interactions with the SLiM region as expected. However, when looking at the hydrogen bonds and hydrophobic interactions, the binding appeared to be slightly displaced in both models. In model 3 two hydrogen bonds were formed between S^{382} - D^{141} and Q^{385} – R^{203}. This resulted in the SLiM binding on the side of the hydrophobic pocket. Something similar happened in the fourth model, where S^{382} and L^{384} were both predicted to form hydrogen bonds with the arginine residue in position 203 of the PDZ domain. This also resulted in binding distant from the hydrophobic pocket, where the interaction would be expected to happen.

Finally, in model 5 (supplementary Figure S4) the C-terminus of the receptor was found to project outwards away from GIPC1. Therefore, no interactions were found between the receptor and the PDZ protein.

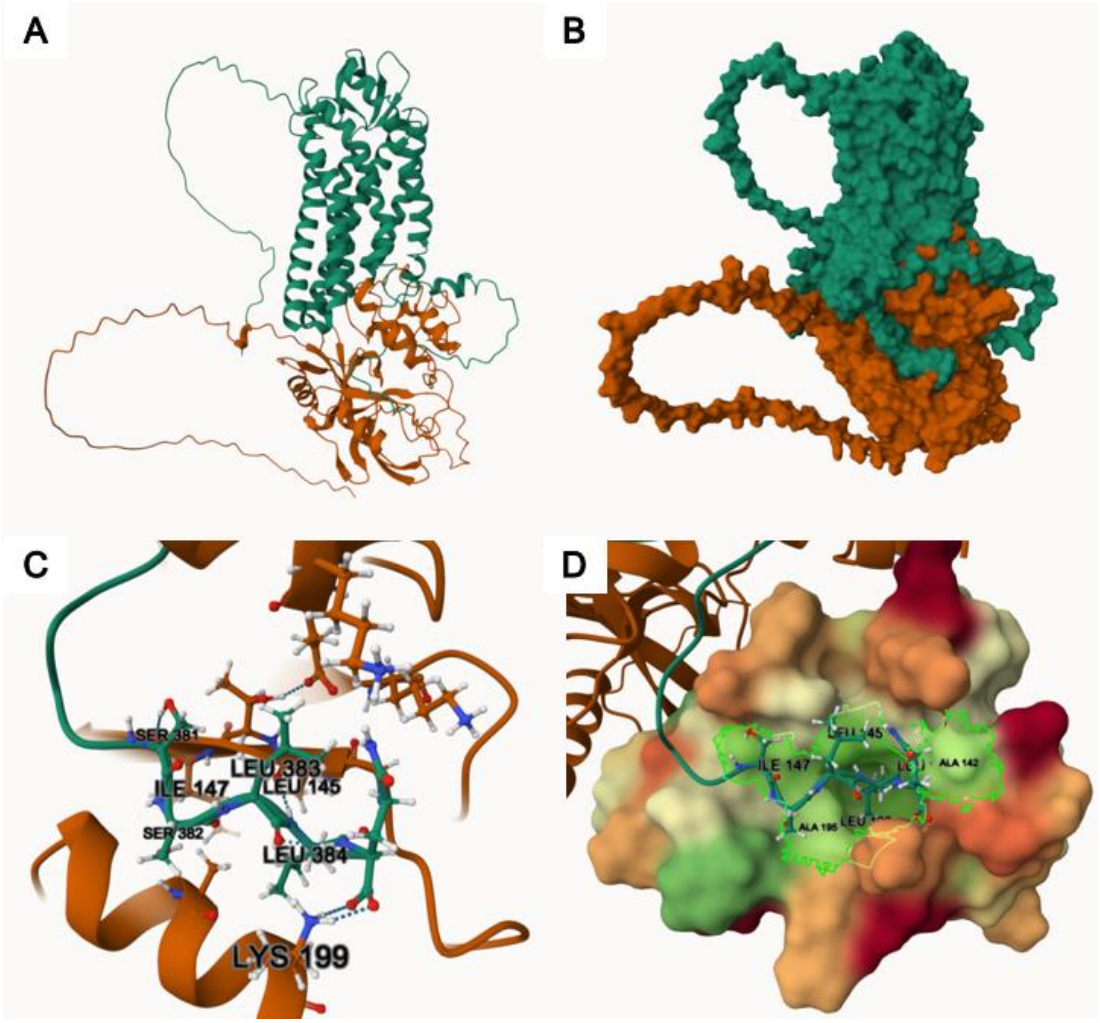


Figure 4.16 – First AlphaFold model of the prediction of PAR4 interacting with GIPC1. A: Ribbon diagram of the receptor bound to GIPC1. PAR4 is shown in green, GIPC1 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and a conserved series of residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a hydrophobic pocket on PDZ domain, residues are colour coded based on their hydrophobicity, from red being very hydrophilic to green being very hydrophobic. Image generated using Mol*.

In the case of GIPC1, this approach greatly improved the quality of the predictions, as every model achieved amean pLDDT greater than 94.5. The other metrics also testify the reliability of the predictions, especially in the case of the top two ranked models, as they displayed pTM and ipTM values above 0.80 (Table 4.10). Moreover, the predictions became much more consistent across different models, highlighting the formation of H bonds between the same amino acid residues. Particularly important was the hydrogen bond between Q^{385} and L^{135}, as this was the only one found in every prediction. We initially hypothesised that the last glutamine residue on PAR4 C-tail would prevent the receptor's SLiM to interact with PDZ proteins. These AlphaFold prediction argue the opposite, as this specific amino acid residue projects directly into the PDZ hydrophobic binding cleft and plays a crucial role in keeping the SLiM aligned in position. Another important interaction was the formation of a double hydrogen bond between the residues L^{383} – I^{147}. These were found in every model, except model 3. In this case PAR4 SLiM appeared to be slightly displaced from the binding pocket, and instead of the double bond between the aforementioned amino acids, a triple hydrogen bond between S^{382} – I^{147} was found instead. Model 3 (Figure 4.17 E-F) was also the only prediction to lack the interaction between S^{381} – T^{148}, which was present in every other model. Models 4 and 5 were the only ones to predict the formation of S^{381} – H^{191}. The formation of this hydrogen bond with a histidine residue present on the helix of the binding site is a common feature of PDZ binding. As Figure 4.17 shows, all 5 models displayed very similar features. PAR4 SLiM was always found in proximity of the hydrophobic binding pocket of GIPC1 PDZ domain, with the last glutamine residues projecting into the hydrophobic binding cleft and forming hydrogen bonds with the leucine residue found on the PDZ at position 135. Model 3 appears to be slightly different, as the SLiM is slightly displaced from the binding pocket and the final glutamine residue does not sit tight within the binding cleft.

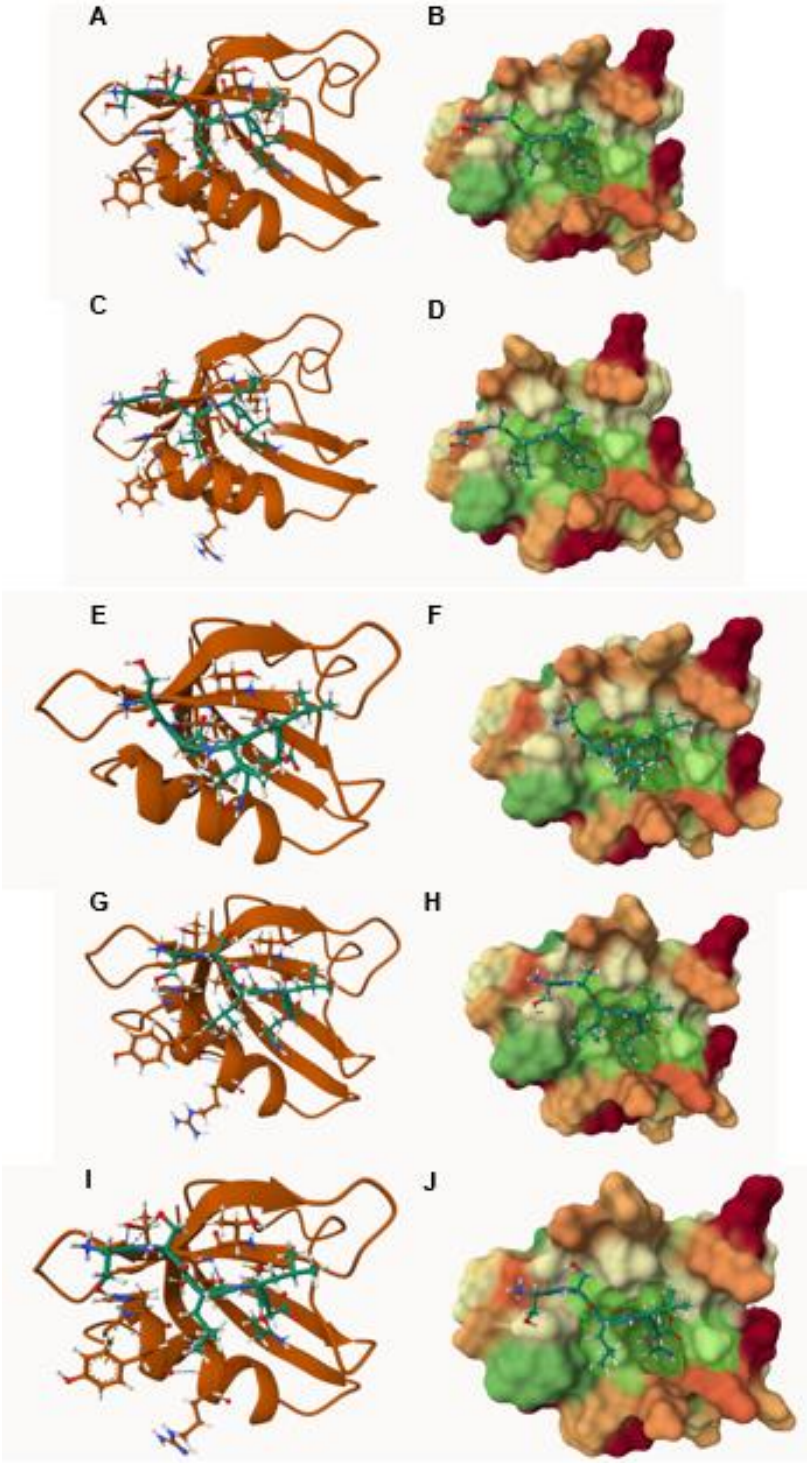


Figure 4.17 – AlphaFold consistently finds the same interactions between PAR4 SLiM and GIPC1 PDZ domain. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.10 – Prediction scores for PAR4 SLiM with GIPC1 PDZ domain

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	95.73	0.85	0.83	S^{381} – T^{148} L^{383} – I^{147} L^{383} – I^{147} Q^{385} – L^{135} Q^{385} – L^{135}	L^{143}, L^{145}, I^{147}, L^{196}
2	95.30	0.85	0.80	S^{381} – T^{148} L^{383} – I^{147} L^{383} – I^{147} Q^{385} – L^{135} Q^{385} – L^{135}	L^{143}, L^{145}, I^{147}, L^{196}
3	94.62	0.84	0.69	S^{382} – I^{147} S^{382} – I^{147} S^{382} – I^{147} Q^{385} – L^{135}	L^{143}, L^{145}, I^{147}, L^{196}
4	94.70	0.84	0.69	S^{381} – T^{148} S^{381} – H^{191} L^{383} – I^{147} L^{383} – I^{147} Q^{385} – L^{135}	L^{143}, L^{145}, I^{147}, L^{196}
5	94.58	0.84	0.66	S^{381} – T^{148} S^{381} – H^{191} L^{383} – I^{147} L^{383} – I^{147} Q^{385} – L^{135}	L^{143}, L^{145}, I^{147}, L^{196}

4.3.5. AlphaFold predicts binding with other members of the GIPC family:

The GIPC family is composed of three proteins, GIPC1, GIPC2 and GIPC3. Although only GIPC1 was flagged in the proteomics experiments, GIPC2 and GIPC3 binding of PAR4 SLiM was investigated using AlphaFold, since these proteins are structurally similar.

Interestingly, AlphaFold predicted binding of PAR4 C-tail on in both cases, finding interactions within the binding loop of the PDZ regions of the two proteins. Similarly to GIPC1, only one model out of five for each GIPC managed to correctly fold the proteins with hydrophobic interactions and hydrogen bonds in the correct region. However, when only the PAR4 SLiM and the PDZ regions were considered, the predictions became much more consistent and the interactions reliable.

In the case of GIPC2, model 2 correctly folded PAR4 SLiM alongside the α2-helix and β2-sheet, within the PDZ binding groove. This corresponds to a hydrophobic binding pocket composed of the residues L^{127}, L^{129}, I^{131}, A^{179}, L^{182}. Three hydrogen bonds were found between the SLiM and the PDZ, being S^{382} – I^{131}, L^{384} – L^{129}, and Q^{385} – K^{183}. In this prediction, the glutamine residue that was thought to prevent binding with PDZ proteins, is found to form a hydrogen bond that stabilises the SLiM in place and allows the leucine residue preceding it (L^{384}) to sit deep within the hydrophobic binding pocket (Figure 4.18). The prediction metrics, hydrogen bonds and hydrophobic interactions for each model are reported in Table 4.11.

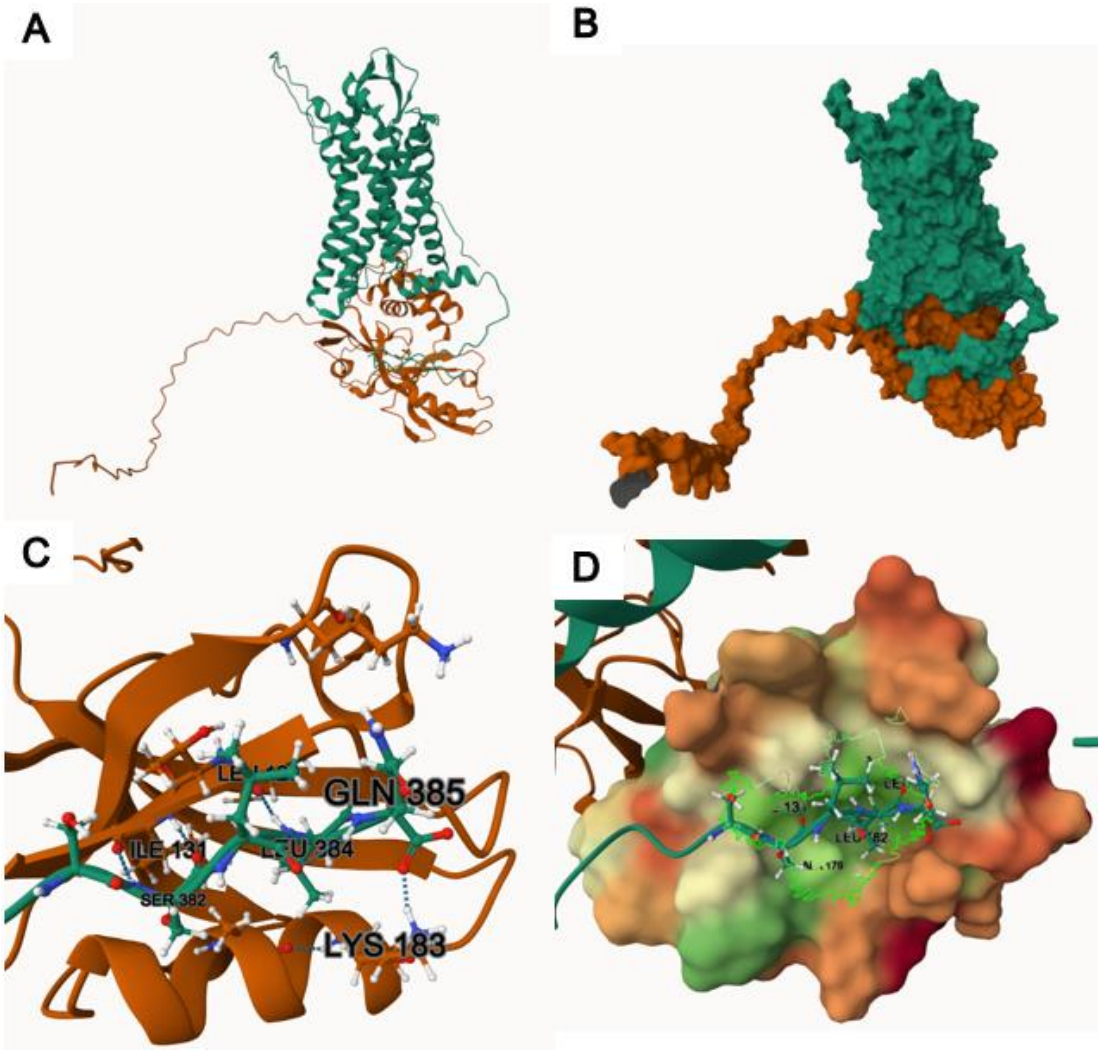


Figure 4.18 – AlphaFold predicts GIPC2 to bind PAR4 SLiM on its PDZ domain. **A**: Ribbon diagram of the receptor bound to GIPC2. PAR4 is shown in green, GIPC2 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and a conserved series of residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a hydrophobic pocket on PDZ domain, residues are colour coded based on their hydrophobicity, from red being very hydrophilic to green being very hydrophobic. Image generated using Mol*.

Table 4.11 - Prediction scores for the binding of PAR4 with GIPC2

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	69.79	0.50	0.30	D^{27} – T^{130}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
2	66.53	0.49	0.30	S^{382} – I^{131} S^{382} – I^{131} L^{384} – L^{129} Q^{385} – K^{183}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
3	67.61	0.47	0.23	E^{28} – K^{183}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
4	63.63	0.49	0.20	N/A	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
5	68.14	0.49	0.20	Y^{26} – I^{131} Y^{26} – I^{131} E^{28} – G^{128} E^{28} – G^{128} E^{28} – L^{129} S^{29} – K^{145}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}

When focusing solely on the SLiM and the PDZ domain, the predictions demonstrated significantly improved performance across all evaluated metrics. Enhancements were observed not only in the mean pLDDT and pTM scores but also in the ipTM scores (Table 4.12), meaning the prediction achieved high accuracy in the interface between the peptide and the domain. Moreover, the same hydrogen bonds were consistently found across models (Figure 4.19). These were similar to the ones observed with GIPC1. Due to structural similarities, the SLiM was found to form hydrogen bonds with the same residues found at the center of the PDZ binding groove, despite having different position numbers. Interestingly, model 3 seemed to differ in the same way it did for GIPC1. In the same fashion, S^{382} formed triple hydrogen bonds with the isoleucine residue central to the hydrophobic pocket. The final glutamine residue on the SLiM (Q^{385}) was consistently found to bind with a leucine residue deep within the binding cleft, often it was predicted to form double hydrogen bonds. Again, similarly to GIPC1, models 4 and 5 were the only ones to predict the hydrogen bonds between S^{381} – H^{175}, a common feature of PDZ binding.

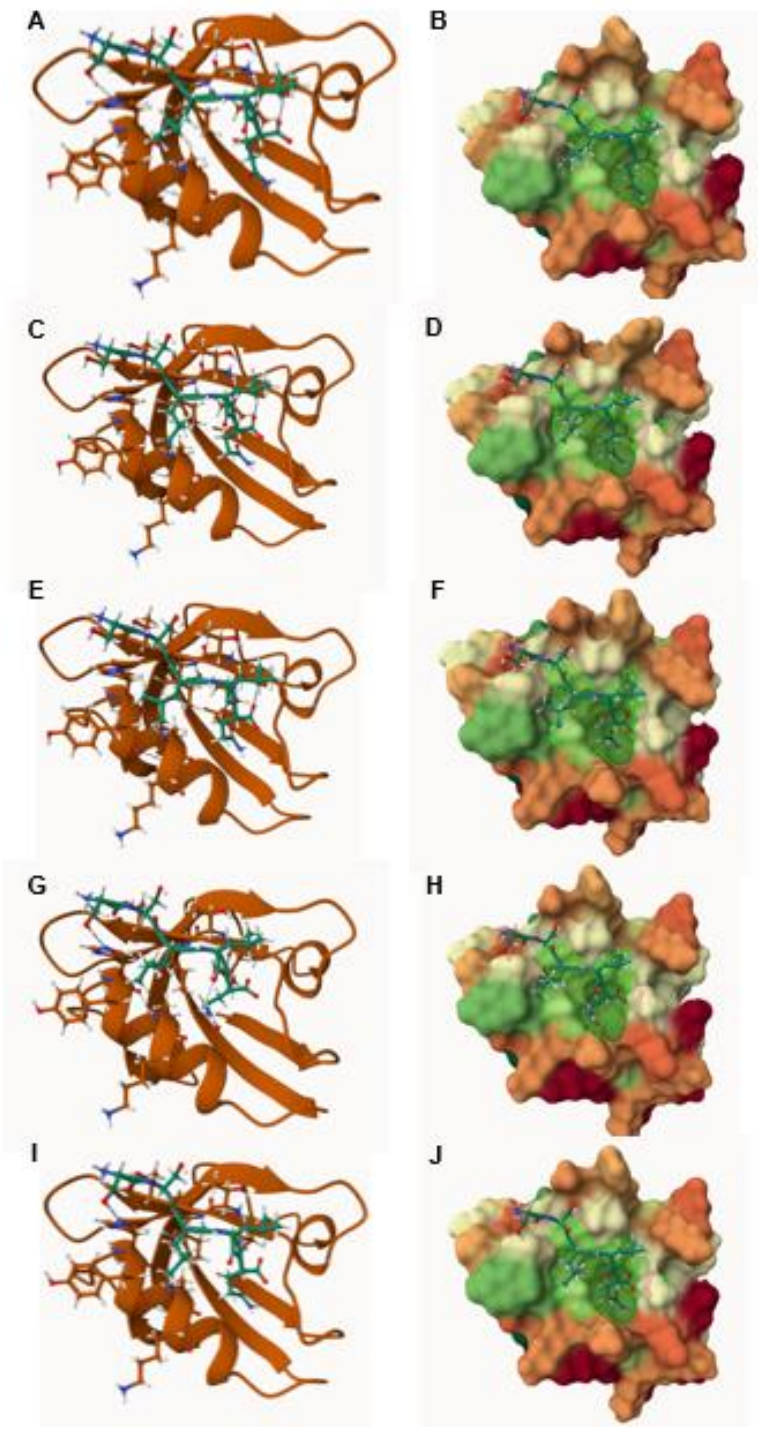


Figure 4.19 – AlphaFold consistently finds binding of PAR4 SLiM to GIPC2 PDZ domain. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.12 - Prediction scores for the binding of PAR4 SLiM with GIPC2 PDZ domain

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	94.79	0.84	0.85	S^{381} – T^{132} L^{383} – I^{131} L^{383} – I^{131} Q^{385} – L^{129} Q^{385} – L^{129}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
2	94.76	0.84	0.84	S^{381} – T^{132} L^{383} – I^{131} L^{383} – I^{131} Q^{385} – L^{129} Q^{385} – L^{129}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
3	95.33	0.85	0.80	S^{382} – I^{131} S^{382} – I^{131} S^{382} – I^{131} Q^{385} – L^{129}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
4	94.62	0.84	0.77	S^{381} – T^{132} S^{381} – H^{175} L^{383} – I^{131} L^{383} – I^{131} Q^{385} – L^{129} Q^{385} – L^{129}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}
5	94.61	0.84	0.77	S^{381} – T^{132} S^{381} – H^{175} L^{383} – I^{131} L^{383} – I^{131} Q^{385} – L^{129} Q^{385} – L^{129}	L^{127}, L^{129}, I^{131}, A^{179}, L^{182}

AlphaFold also predicted the binding of GIPC3 with PAR4 SLiM only in one model out of five when using the full protein sequences. In this case the hydrophobic pocket was composed by the residues: L^{122}, L^{124}, I^{126}, A^{174}, L^{177} on GIPC3 PDZ domain (Figure 4.20D). Only two hydrogen bonds were formed in this model, between S^{382} – I^{126} and L^{384} – L^{124}. However, both interactions present a double hydrogen bond between the amino acids. In this prediction Q^{385} does not form any hydrogen bond, but still it does not prevent L^{384} from accommodating deep in the hydrophobic cleft. The prediction metrics, hydrogen bonds and hydrophobic interactions for each model are presented in Table 4.13.

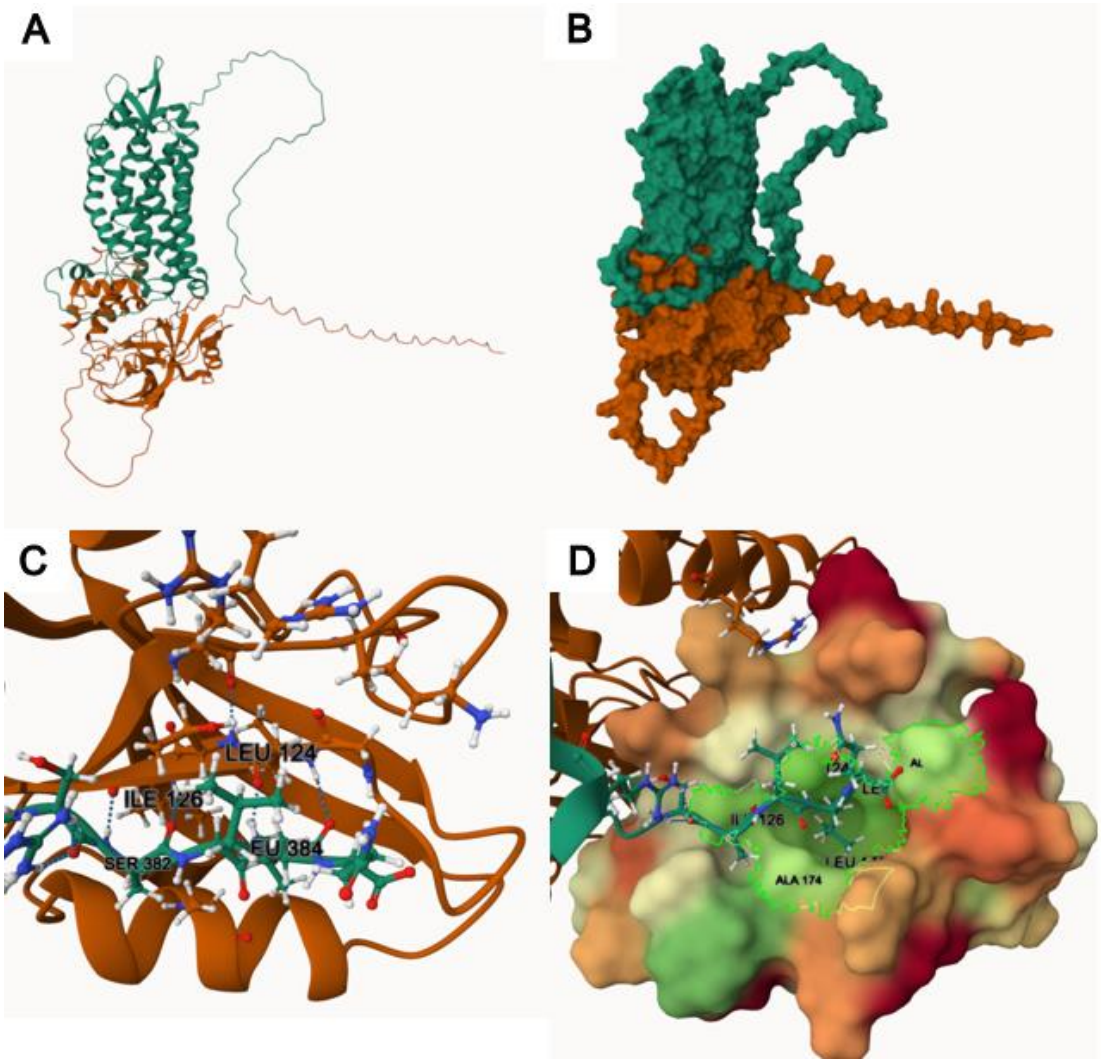


Figure 4.20 – AlphaFold consistently finds binding of PAR4 SLiM to GIPC3 PDZ domain. A: Ribbon diagram of the receptor bound to GIPC3. PAR4 is shown in green, GIPC3 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and a conserved series of residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a hydrophobic pocket on PDZ domain, residues are colour coded based on their hydrophobicity, from red being very hydrophilic to green being very hydrophobic. Image generated using Mol*.

Table 4.13 - Prediction scores for the binding of PAR4 with GIPC3

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	66.37	0.49	0.26	S^{382} - I^{126} S^{382} - I^{126} L^{384} – L^{124} L^{384} – L^{124}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
2	68.49	0.49	0.21	D^{27} – T^{125}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
3	67.45	0.48	0.19	S^{29} – K^{140} D^{37} – K^{140} D^{37} – K^{140}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
4	67.81	0.50	0.17	Y^{26} – T^{125} D^{27} – R^{138} E^{28} – G^{123} S^{29} – K^{140}	L^{122}, L^{124}, I^{126}
5	64.54	0.47	0.14	N/A	N/A

When AlphaFold was used on the protein fractions, the predictions metrics increased as expected. Just like GIPC1 and GIPC2, also GIPC3 achieved outstanding scores for mean pLDDT, pTM and ipTM, thus highlighting the quality of the models and the high probability of an interaction. As the previous proteins, hydrogen bonds were found between similar residues (Table 4.14). Although the residue numbers were slightly different, the first serine of PAR4 SLiM was found to interact with a threonine residue (S^{381} – T^{127}), and two hydrogen bonds were found between the third leucine in the peptide and the isoleucine of the hydrophobic cleft (L^{383} – I^{126}), while the last glutamine residue again played a crucial role by binding the leucine found deep within the binding pocket (Q^{385} – L^{124}). Three models out of five found interactions with the histidine at position 170, on the helix flanking the binding groove. Finally, compared to GIPC1 and GIPC2, no models predicted a triple hydrogen bond between the second serine (S^{382}) and the central isoleucine (I^{126}). AlphaFold predictions for PAR4 SLiM bound to GIPC3 PDZ domain are shown in figure 4.21.

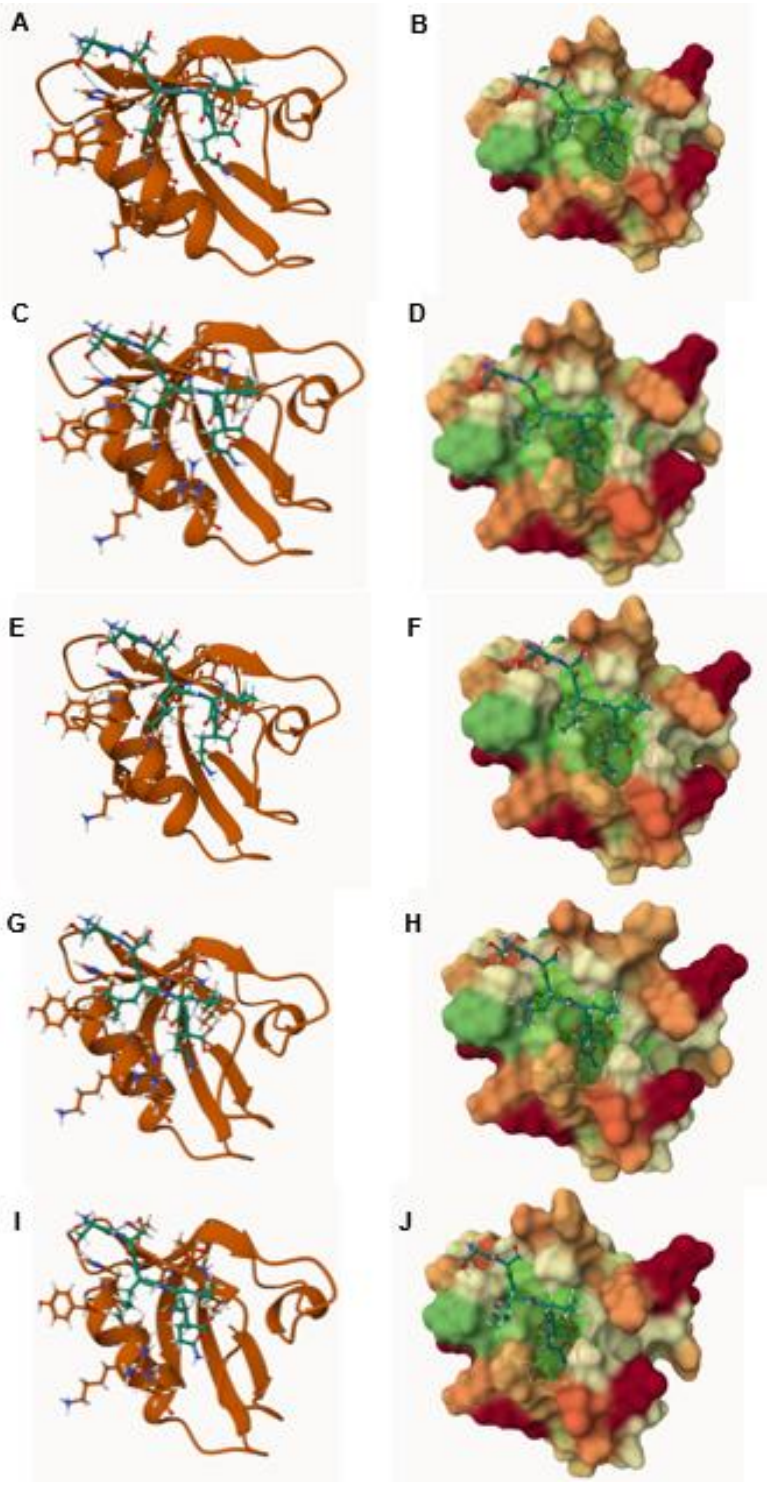


Figure 4.21 – AlphaFold predictions of PAR4 SLiM and GIPC3 PDZ domain only. To increase the accuracy of the predictions, AlphaFold multimer was used only on the fragments of the proteins. A-B: Ribbon diagram and molecular surface of model 1. C-D: Ribbon diagram and molecular surface of model 2. E-F: Ribbon diagram and molecular surface of model 3. A-B: Ribbon diagram and molecular surface of model 3. G-H: Ribbon diagram and molecular surface of model 4. I-J: Ribbon diagram and molecular surface of model 5. The hydrophobic binding cleft is highlighted in green.

Table 4.14 - Prediction scores for the binding of PAR4 SLiM with GIPC3

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	95.50	0.85	0.85	S^{381} – T^{127} S^{381} – H^{170} L^{383} – I^{126} L^{383} – I^{126} Q^{385} – L^{124}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
2	94.49	0.84	0.83	S^{381} – T^{127} S^{381} – H^{170} L^{383} – I^{126} L^{383} – I^{126} Q^{385} – L^{124} Q^{385} – L^{124}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
3	94.24	0.84	0.80	S^{381} – T^{127} L^{383} – I^{126} L^{383} – I^{126} Q^{385} – L^{124} Q^{385} – L^{124}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
4	94.17	0.83	0.79	S^{381} – T^{127} L^{383} – I^{126} L^{383} – I^{126} Q^{385} – L^{124} Q^{385} – L^{124}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}
5	94.18	0.83	0.72	S^{381} – T^{127} S^{381} – H^{170} S^{382} – T^{127} L^{383} – I^{126} L^{383} – I^{126} Q^{385} – L^{124} Q^{385} – L^{124}	L^{122}, L^{124}, I^{126}, A^{174}, L^{177}

4.3.6. PrePPI interactions:

Very recently, a novel webserver and database designed for predicting protein-protein interactions (PPIs) across the entire human proteome has been released. Its name is PrePPI (Predicting Protein-Protein Interactions), and it combines both structural and non-structural evidence within a Bayesian framework to compute a likelihood ratio for virtually every possible pair of proteins in a proteome. The system leverages structural modelling, utilizing template-based modelling and AlphaFold structures to predict how proteins might interact. The database holds approximately 1.3 million human PPIs and supports complex queries, which can be used to explore potential interactions, examine template complexes, and view 3D models of predicted complexes. PAR4 UniProt ID (Q96RI0) was queried directly on the webserver, this yielded a total of 189 potential interactions. The entire results of PrePPi predictions are in the appendix, in supplementary Table S10. PrePPI provides different metrics to assess the quality of a prediction, the most important being the Likelihood ratio (LR). Other metrics provided by the webserver include Structural modelling score (SM), Protein peptide LR (PrP), Partner redundancy LR (PR), Orthology LR (OR), Phylogenetic profile LR (PP), Expression profile LR (EP), Gene ontology LR (GO), with Total LR quantifying the total likelihood ratio of an interaction happening. The average Total LR for the 189 entries was 122.81, while the median was 536.3. The list of potential interactors identified by PrePPI was compared against the list of known PDZ proteins obtained from UniProt. Out of 189 entries, only 5 PDZ proteins were identified. These are nitric oxide synthase 1 (NOS1), Connector enhancer of kinase suppressor of ras 3 (CNKSR3), Na(+)/H(+) exchange regulatory cofactor NHE-RF3 (NHERF3), Protein Shroom 2 (SHROOM2), and Periaxin (PRX). Each protein with its corresponding likelihood ratios is reported in table 4.15. Only NOS1 displayed a Total LR above average, while the other proteins showed a Total LR similar to the median of the group.

Table 4.15 – PDZ proteins predicted to interact with PAR4 according to PrePPI.

UniProt	Gene	SM	PrP	PR	OR	PP	EP	GO	Total LR
P29475	NOS1	0.2	3.5			2.4	3.9	69.6	2315.5
Q6P9H4	CNKSR3	0.6	3.5			2.4	1.2	69.6	697.9
Q5T2W1	NHRF3		3.5			2.4	1.2	69.6	697.9
Q13796	SHROOM2	0.5	3.5			2.4	2.1	24.1	433.3
Q9BXM0	PRX		3.5			2.4	5.7	5.9	282.6

To investigate and validate the accuracy of PrePPI predictions, AlphaFold Multimer was applied on the PDZ domains of each protein together with the SLiM of PAR4. The results of each analysis are reported in the supplementary materials, Tables S11 to S17. Since NHERF3 possessed four distinct PDZ domains, each one was investigated for potential PAR4 SLiM binding using AlphaFold. The validation results revealed mixed support for the PrePPI predictions. Every model for every prediction achieved high values for mean pLDDT and pTM scores, highlighting the quality of the structural predictions. However, concerning ipTM and therefore the interface of the interactions, few models achieved an ipTM score >0.8, which is used as a cutoff for reliable interaction predictions. Every protein except PRX managed to position the

PAR4 SLiM roughly in the right position, aligned with the α-helix of the PDZ domain and right next to the carboxylate binding groove.

For NOS1 (Figure 4.22), the mean pLDDT ranged from 92.65 to 97.05, indicating high model confidence, while pTM scores were moderate to high (0.85-0.88). However, only models 1 and 2 had ipTM scores close to the reliability cutoff, with model 1 at 0.87 and model 2 at 0.74. Despite this, consistent hydrogen bonds and hydrophobic interactions suggested a strong interaction.

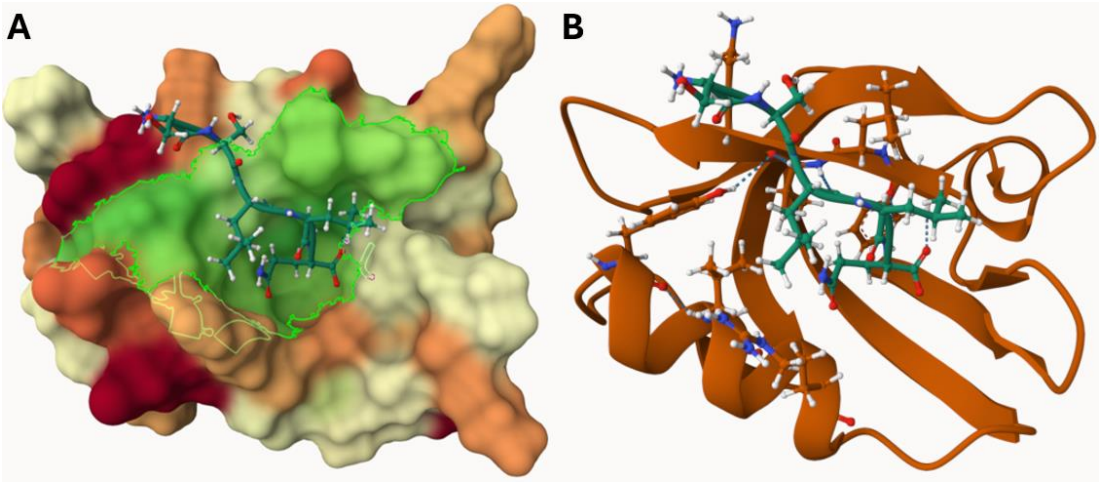


Figure 4.22 – Top ranked AlphaFold prediction for NOS1. A: Molecular surface of NOS1 PDZ domain. Hydrophobic residues are highlighted in green. **B:** Ribbon diagram of NOS1 PDZ domain.

Similarly, CNKSR3 (Figure 4.23) showed high model confidence with mean pLDDT ranging from 86.80 to 94.09 and pTM scores between 0.77 and 0.86. Model 1 achieved the highest ipTM score of 0.83, slightly below the reliability cutoff, yet consistent hydrogen bonds and hydrophobic interactions supported the interaction.

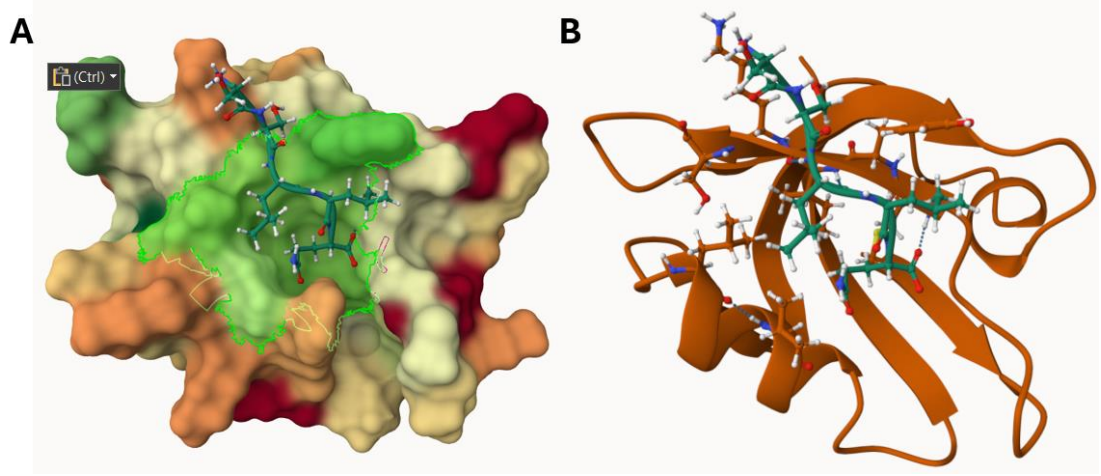


Figure 4.23 – Top ranked AlphaFold prediction for CNKSR3. A: Molecular surface of CNKSR3 PDZ domain. Hydrophobic residues are highlighted in green. **B:** Ribbon diagram of CNKSR3 PDZ domain.

The NHERF3 PDZ domains presented a different picture (Figure 4.24), with mean pLDDT values ranging from 89.68 to 92.17 and pTM scores around 0.81-0.85. However, the highest ipTM scores for its domains were significantly below the reliability cutoff, with domain 1 reaching only 0.48. This inconsistency in hydrogen bonds and hydrophobic interactions suggested weaker or less reliable interactions.

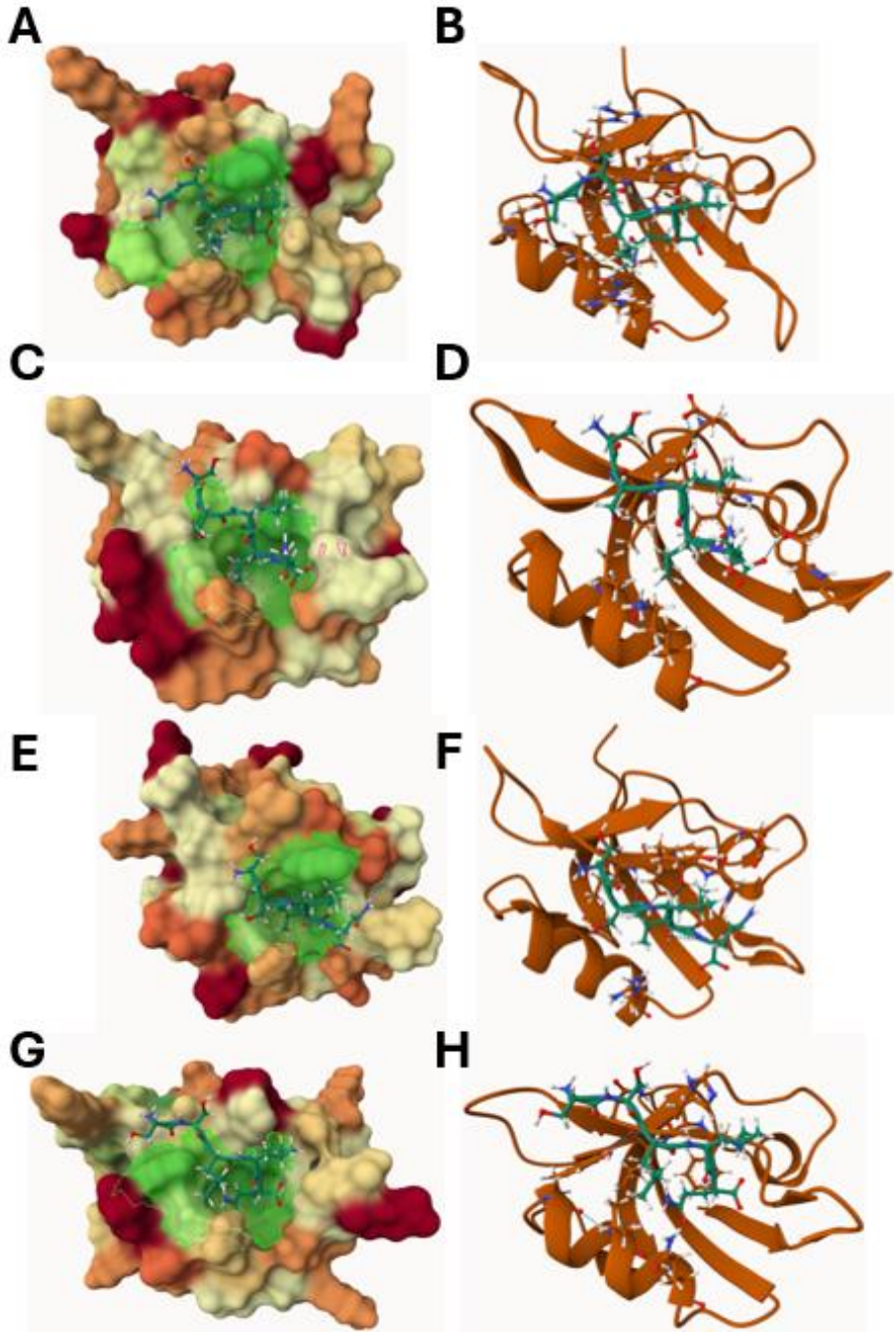


Figure 4.24 – Top ranked AlphaFold prediction for each NHERF3 PDZ domain. Hydrophobic residues are highlighted in green**. A-B:** domain 1, **C-D:** domain 2, **E-F:** domain 3, **G-H:** domain 4.

PRX (Figure 4.25) also showed lower model confidence, with mean pLDDT values between 67.88 and 85.32 and pTM scores from 0.53 to 0.71. Only model 1 displayed a moderate ipTM score of 0.46, and inconsistent interactions further indicated weak or unreliable binding.

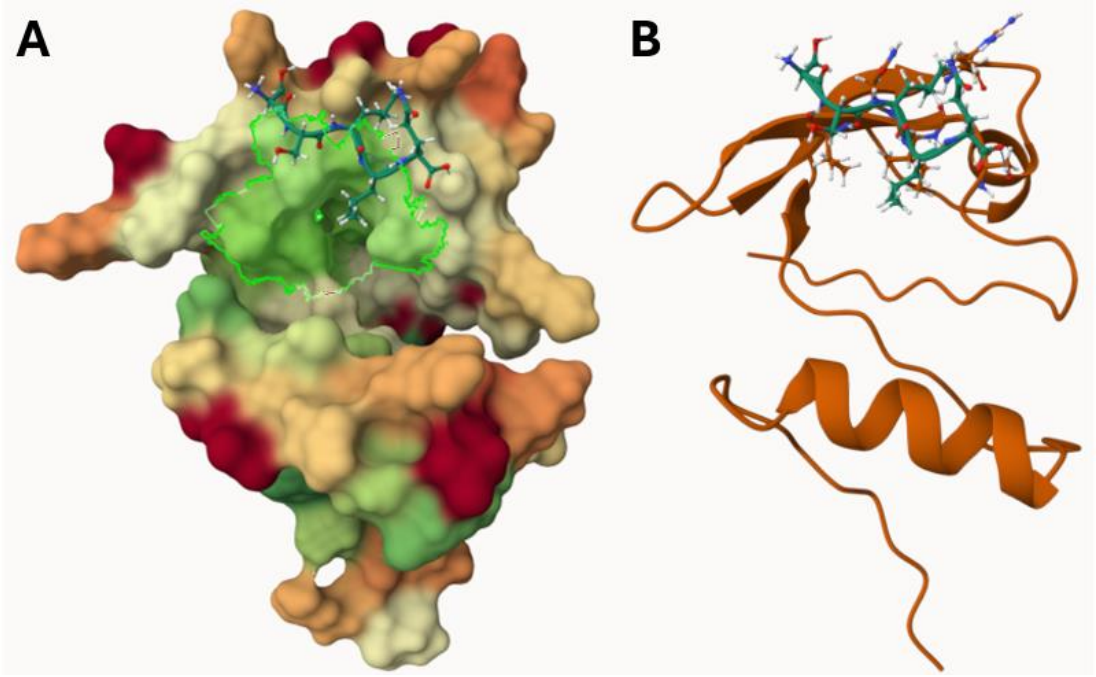


Figure 4.25 – Top ranked AlphaFold prediction for PRX. A: Molecular surface of PRX PDZ domain. Hydrophobic residues are highlighted in green. **B:** Ribbon diagram of PRX PDZ domain.

SHROOM2 (Figure 4.26) exhibited high model confidence with mean pLDDT values ranging from 87.46 to 94.22 and pTM scores around 0.78-0.83. Model 1 achieved an ipTM score of 0.84, surpassing the reliability cutoff, with consistent hydrogen bonds and hydrophobic interactions supporting a strong interaction prediction.

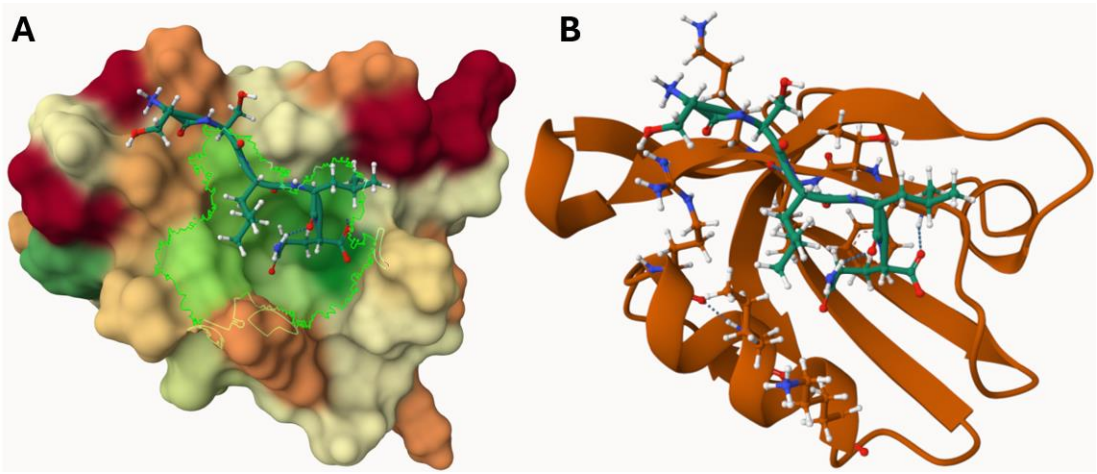


Figure 4.26 – Top ranked AlphaFold prediction for SHROOM2. A: Molecular surface of SHROOM2 PDZ domain. Hydrophobic residues are highlighted in green. **B:** Ribbon diagram of SHROOM2 PDZ domain.

4.4. DISCUSSION:

The aim of this study was to investigate whether AlphaFold would be able to predict the interactions found in the proteomic analysis. Four proteins were identified as potential interactions: GIPC1, TJP2, LMO7 and PDLIM1. The most probable interaction is GIPC1, as the other models did not find any realistic point of contact.

4.4.1. Roles of PDZ and LIM domains:

In the proteomics part of the project, two out of the four proteins shared the feature of possessing a LIM domain together with a PDZ domain. These proteins were LMO7 and PDLIM1. Interestingly, according to the Human Protein Atlas expression clustering and correlation, PDLIM1 was the 6th nearest neighbour to F2RL3 (PAR4 gene). This was based on single cell type RNA expression, PDLIM1 belonged to the same cluster as F2RL3 (cluster 57) with a correlation score of 0.627.

The PDZ and LIM domain-containing protein family is encoded by a diverse group of genes. In mammals, ten genes have been identified that encode both a PDZ and one or several LIM domains, including ALP, RIL, Elfin (CLP36), Mystique, Enigma (LMP-1), Enigma homologue (ENH), ZASP (Cypher, Oracle), LMO7, and the two LIM domain kinases (LIMK1 and LIMK2)[355]. Both domains exhibit a conserved structure. PDZ proteins are characterized by six β-sheets enclosed by two α-helixes, harbouring a central globular binding groove situated between the α2-helix and β2-sheet, and a series of hydrophobic amino acid residues forming a shallow cleft known as the "carboxylate-binding loop"[365]. On the other hand, the LIM domain, named after Lin-11, Isl-1, and Mec-3 where it was first identified, features a distinct cysteine-rich motif. Comprising two zinc finger structures separated by a two-amino acid spacer[313]. The coexistence of these domains in the identified genes underscores their multifaceted roles in protein-protein interactions, cellular signalling, and cytoskeletal regulation.

The phylogeny of PDZ and LIM containing proteins has been previously analysed in over 25 species[366]. A common ancestral gene was found, despite the proteins being so structurally different. All 10 PDZ/LIM genes have demonstrated the capability to interact with the actin cytoskeleton[354]. The presence of both LIM and PDZ domains within a single functional module suggests that a synergistic combination of these domains may be essential for specific interactions with both the actin cytoskeleton and other proteins, highlighting roles in scaffolding and transport. In this context, it would be plausible to speculate about the role of PDZ/LIM proteins in trafficking GPCRs. The PDZ domain could interact with the SLiM on the C-terminal side of the receptor, while the LIM domain would bind the cytoskeleton and lead to receptor translocation. However, no such cases have been yet documented, and AlphaFold predictions agree with this. No prediction for either LMO7 or PDLIM1 managed to find any meaningful interaction with the SLiM on PAR4 C-tail. Since AlphaFold was applied as a layer of screening for validation of proteomics results *in vitro*, LMO7 and PDLIM1 potential interactions with PAR4 will not be assessed in further assays.

4.4.2. TJP2 does not recognise GPCRs:

Despite having three PDZ domain, tight junction protein ZO2 does not recognise any SLiM on GPCRs. As mentioned previously, this is a tight junction (TJ) protein, whose role is to scaffold only other TJ proteins. Indeed, the majority of claudins localized at tight junctions possess conserved C-terminal tails that establish binding interactions with PDZ1 of TJP proteins[367]. Beyond claudin binding, PDZ1 facilitates associations with phosphoinositides[368]. Furthermore, the second PDZ domain (PDZ2) of TJP proteins is recognized for its role in promoting protein dimerization[334]. This explains its different structure when compared at classical PDZ domains and clarifies why AlphaFold predictions did not find any interaction at this domain. Two dimerised TJP2 proteins would have to be predicted to investigate the involvement of this domain. However, due to the large structure of the protein, it would be too computationally demanding to investigate such possibility. Moreover, NMR analysis has unequivocally demonstrated the potential for PDZ2 of TJP2 to dimerize through domain swapping. This phenomenon is supported by a crystal structure at 1.75 Å resolution, revealing a domain-swapped dimer in TJP2 PDZ2 with β-strands 1 and 2 exchanged[335]. Evidence suggests that PDZ2 promotes domain-swapped homodimer formation across all three TJP proteins. Given the high sequence similarity between TJP-1, -2, and -3 PDZ2 domains (66% sequence identity between TJP1 and TJP2, 50% between TJP1 and TJP3, 54% between TJP2 and TJP3), the proposal of heterodimer formation as a potential mechanism to create and stabilize the cytoplasmic plaque, a dense protein network located on the cytoplasmic side of the plasma membrane in tight junctions, has been put forth[369]. Finally, the third PDZ domain (PDZ3) of TJP proteins plays a crucial role in interacting with the C-termini of transmembrane junctional adhesion molecules[370].

Given that the scaffolding functions of TJP proteins is well characterised in the context of tight junction biology, it seems unlikely that TJP2 would interact with any GPCR. Moreover, PAR4 has never been documented to be found at TJs, for these reasons it appears questionable that any TJP2 PDZ domain could interact with PAR4 SLiM, and AlphaFold seems to confirm this hypothesis. Therefore, as with LMO7 and PDLIM1, no further validation will be applied *in vitro* to study any interaction between PAR4 and TJP2.

4.4.3. GIPC1 as a modulator of GPCR signalling:

When considering full sequence predictions, GIPC1 PDZ was the only domain to interact with PAR4 SLiM at the correct position of the carboxylate binding groove in model 1. It is important to note that AlphaFold was not set to use template models for protein folding, let alone for multimeric interactions. This means that model 1 is a genuine interaction found by AlphaFold and it is not based on previous similar structures. The consistency achieved in every model when testing the prediction of just the PAR4 SLiM peptide against GIPC1 PDZ domain, suggests these two proteins might interact *in vitro*.

GIPC1 has been previously documented in literature to recognise GPCR SLiMs, as it was found to interact with the C-terminals of the adrenergic receptor β1 (ADRB1)[362], dopamine receptor D2 (DRD2)[348] and D3 (DRD3)[347], luteinizing hormone/choriogonadotropin receptor (LHCGR)[353], and Lysophosphatidic acid receptor 1 (LPA1)[371]. Moreover, GIPC1 scaffolds the Regulator of G-protein signalling 1 (RGS19) also known as Gα-interacting protein (GAIP)[360], which is not a GPCR, but it is involved in GPCR signalling. This is the protein that gives GIPC its name, which stands for GAIP-interacting protein C terminus. The biological implication of each of these interactions is reported in table 4.16.

Table 4.16 – GIPC1 interactions with GPCRs and their biological roles.

GPCR	Short Linear Motif	PDZ Class	Role	Ref.
ADRB1	ESKV	I	Inhibits ERK activation	362
DRD2	LHC	III	Translocates receptor from membrane to vesicles	348
DRD3	LSC	III	Reduces the maximal inhibition of cAMP accumulation	347
LHCGR	YTEC	I	Mantain surface receptor levels Promotes hormone recycling	353
LPA1	SVV	I	Trafficking to endosomes Increased PKB	371
RGS19	LQUGPSQSSSEA		Enhance cell proliferation	372

On a structural point of view, GIPC1 seems to recognise a large pool of SLiMs, not restricted to Class I PDZ binding domains. The engagement of GIPC with both D2 receptor and D3 receptor appears to employ an unconventional form of PDZ recognition[373]. This peculiarity arises because both receptors exhibit a PDZ type III consensus motif X-X-C at their C-terminus, in contrast to the PDZ type I consensus motif S/T-X-V/A/L/I observed in most GIPC-interacting proteins' binding sites[360]. Notably, the interaction with D3 requires both the PDZ and acyl carrier protein (ACP) domains of GIPC[347]. In this context, the structural folding of the entire cytoplasmic tail needs to be considered, as evidenced by the lack of GIPC binding to the C-terminus of D4 receptor. Despite featuring the X-X-C motif, the D4 C-terminus is two amino acids longer than that of D2 and D3, leading to its non-interaction with GIPC. The plasticity of the PDZ-based recognition by GIPC, is highlighted also by its ability to bind to an internal motif of the TrkA receptor although interaction is less potent than with a C-terminal PDZ-binding consensus. The interactions of RGS19 (GAIP) and TrkA are based on two distinct sites in the PDZ domain of GIPC[374].

The promiscuity of GIPC recognition of patterns on GPCR tails is also highlighted by its interaction with the human lutropin receptor. The PDZ protein was found to bind the receptor, although the SLiM of the hLHR (YTEC) contains only one residue (the Thr in the -2 position) of the consensus sequence (X(S/T)X(V/L)) that appears to be recognised by type I PDZ domains[375] . Moreover, the cysteine residue at position 699 appears to be necessary for the binding of the hLHR to GIPC[353], despite this amino acid possessing a polar neutral side chain, rather than being a hydrophobic residue, which was thought to be necessary for PDZ class I binding[311]. This is particularly relevant in the context of potential PAR4 SLiM binding, where the glutamine residue at the end of the C tail (Q^{385}) also has a polar neutral side chain, which was thought to prevent the binding of the SLiM to the hydrophobic binding cleft on the PDZ domain. In model 1 of the full sequence prediction for GIPC1, Q^{385} appeared to play a role, stabilising PAR4 SliM in position by interacting with the lysine residue at position 199 on the PDZ domain. The importance of this residue became even more apparent when considering just the SLiM binding to the PDZ domain, where Q^{385} was consistently found buried within GIPC1 binding pocket and bound to L^{35}. The same hydrogen bonds were observed with residues L^{129} and L^{124} for GIPC2 and GIPC3 respectively.

Another common feature of class I PDZ domains is the formation of a hydrogen bond between the histidine projecting out from the α2 helix and a serine or threonine residue in the second position of the binding peptide[376]. This feature was observed across the fraction predictions of the GIPC family, although only in models 4 and 5 for GIPC1 and GIPC2, and models 1, 2 and 5 of GIPC3. However, the SLiM binding residue was slightly displaced, as it was always S^{381} to bind the histidine residues projecting out of the α2 helix.

In the β1 adrenergic receptor, a residue that appeared to be necessary for SLiM binding was the serine at position -2, as mutation of this amino acid to either alanine or aspartic acid resulted in complete ablation of the interaction. Mutating the last valine residue to an alanine had no effect on association between the PDZ protein and the receptor[362], thus suggesting again how the presence of a hydrophobic residue at the end of the SLiM is not a prerogative for GIPCs binding. Interestingly, the proposed SLiM on PAR4 (SSLLQ) has a much more similar structure to adrenergic β2 receptor SLiM (DSLL), than β1 (ESKV). However, when Hu and colleagues investigated any potential interactions with adrenergic receptors, they used both β1 and β2 C-terminal baits, but found interactions only with β1, highlighting the selectivity of GIPC binding[362]. It must be noted that every residue engaging with the PDZ domain binding cleft has the potential to influence peptide selectivity. Amacher *et al.* conceptualised PDZ binding sequences as barcodes, where the specific arrangement of residues at each position encodes an overall sequence that can be "read" by interacting PDZ domains[311]. This raises questions regarding the robustness of the AlphaFold prediction and warrants further validation through *in vitro* studies.

4.4.4. Validation of GIPCs binding *in vitro* and functional characterisation:

To validate *in vitro* the binding between PAR4 and GIPCs, multiple approaches can be applied. The two main avenues to study these interactions would be co-immunoprecipitations (co-Ips) and proximity ligation assays (PLAs). Co-IPs involve the selective precipitation of a target protein along with its interacting partners using specific antibodies. In the context of GIPCs and PAR4, Co-IPs can help confirm their physical association by capturing the protein complexes formed in cell lysates[377]. On the other hand, PLAs provide a sensitive and specific method for visualizing protein-protein interactions at the single-molecule level. This technique relies on the proximity of two proteins, detecting their interaction through the ligation of oligonucleotide-labelled secondary antibodies followed by rolling circle amplification. In this case the interaction would be validated through microscopy imaging[378]. Our laboratory has constructs readily available not just for GIPC1, but also for GIPC2 and GIPC3, thus allowing the investigation of the entire GIPC family gene, and its potential involvement in PAR4 regulation.

If GIPCs were able to bind PAR4, the immediately following question would be what their functional role is. Being able to bind a large variety of proteins, GIPC1 is involved in the modulation of a large pool of biological activities. Similar to many PDZ domain-containing proteins, GIPC1 has been shown to play an important role in regulating cell signalling[374], anchoring proteins in specific subcellular compartments[379], the clustering of signalling components and transmembrane receptors[374], and the regulation of signalling by membrane receptors[380]. In the context of GPCR biology, GIPC1 is involved in the modulation of the endosomal GPCR signalling. Through dimerization, GIPC may recruit either RGS or cytoskeleton-associated proteins[350] to uncouple GPCRs from their signalling cascade and to subsequently link the receptors to cargoes of the MYO6 motor protein in early endosomes[381]. As RGS19 (GAIP) functions as a GTPase-activating protein to inactivate the G protein, the dimerization of GIPC1 induces clustering of GPCR and RGS19 to the early endosomes for the attenuation of G protein signalling[348]. GIPC1 also regulates the expression of the TGF-β receptor at the cell surface and enhances the cellular response to TGF-β[382].

In this context, GIPC1 could mediate PAR4 trafficking and recycling, as well as opening new avenues towards PAR4 signalling from endosomes. As GAIP has been documented to attenuate Gq signalling as well[383],GIPCs may downregulate PAR4 activity. Since in the proteomic study GIPC1 was flagged up in the Y157C proteome as well as the PAR4 proteome, GIPC1 interaction might be responsible of the increased retention at the endoplasmic reticulum previously documented[193]. These are however just speculations, and the interactions between PAR4 and GIPCs first need to be verified *in vitro*.

4.4.5. PrePPI suggests low confidence interactions:

The field of Artificial Intelligence is constantly evolving at an extremely rapid pace. This holds true also for tools applied to biology and biochemistry, such as the Predicting Protein-Protein Interactions (PrePPI) webserver, which was released concomitantly to this project[384]. The underlying algorithms use structures from template modelling and AlphaFold predictions to compute the likelihood of interactions between proteins. However, this means that statistical models operate on the full sequence of the structures, which was reported to decrease specificity[296]. For this reason, we first selected all the PDZ proteins found by the webserver and then repeated the AlphaFold predictions using only the PDZ domains and the SLiM motif of PAR4. Six PDZ proteins were found among the 189 interactions suggested by PrePPI, however the predictions metrics for the interface between the template modelling of the structures were relatively low, and often did not reach the cutoff for considering a interaction reliable.

The protein with the highest likelihood of interaction was Nitric Oxide Synthase 1 (NOS1). Besides being involved in the formation of nitric oxide, this molecule carries nitrosylase activity and mediates cysteine S-nitrosylation (the covalent attachment of a nitric oxide moiety to specified cysteine thiol groups) of cytoplasmic target proteins. This of course includes GPCRs as well as their effector proteins, such as GRKs, β-arrestins, and dynamin[385]. This process has been shown to be particularly clinically relevant in the case of cardiac GPCRs, such as β-adrenoceptors, where S-nitrosylation of GRK2 and β-arrestins can contribute to pathological signalling in cardiac injury[386]. Moreover, S-nitrosylation of β-arrestins has been demonstrated to bias receptor signalling enabling ligand-independent β-arrestin function and exacerbating heart failure[387]. It can be therefore speculated that if PAR4 was interacting with NOS1, this would carry S-nitrosylation activity on the receptor. However, PAR4 does not present any cysteine residues on the intracellular side of the receptor where this post-translational modification could happen.

Another potential interaction found by PrePPi was the Connector enhancer of kinase suppressor of ras 3 (CNKSR3), which is involved in transepithelial sodium transport. It regulates aldosterone-induced and epithelial sodium channel (ENaC)-mediated sodium transport through regulation of ENaC cell surface expression. It also acts as a scaffold protein, coordinating the assembly of an ENaC-regulatory complex (ERC). To date, no direct association between this protein and any GPCR has been established. Only one study found inhibition of CNKSR3 in response to cyclic adenosine monophosphate (cAMP), which is often mediated by GPCRs[388]. This suggests a possible connection between CNKSR3 and GPCR pathways, however no direct structural association, such as interaction with its PDZ domain, has been documented.

The Na(+)/H(+) exchange regulatory cofactor NHERF3 acts as a scaffold protein, linking plasma membrane proteins with regulatory components to control their presence on the cell surface. It participates in coordinating a variety of regulatory processes related to ion transport and secondary messenger pathways. Members of the NHERF family are well known to recognise highly conserved motifs in GPCRs, and modulate their signalling, trafficking and function[389]. NHERF3 has been shown to interact with class I PDZ motifs, but most of its PDZ domain interactions remain largely uncharacterized. The second PDZ domain on NHERF3 was found to strongly bind

the carboxyl terminus of the β2-adrenergic receptor[390], which is structurally similar to PAR4 SLiM. NHERF3 also regulates phospholipase C-β3 (PLC-β3)-specific activation of somatostatin by forming a ternary complex with PLC-β3 and somatostatin receptors[391]. NHERF3 was found bind corticotropin-releasing hormone receptor 1 (CRFR1) in a PDZ-motif dependent manner, thus selectively increasing CRFR1-stimulated ERK1/2 phosphorylation. However, the same study found that NHERF3 also binds 5HT2AR, but independently of anzy PDZ-motif recognition. This interaction positively regulated the formation of inositol phosphate[392]. Only PDZ domain 1 of NHERF3 has been found to bind β-arrestin2[393], highlighting how this protein does not only interact with the receptors, but can also modulate downstream signalling by recognising effector molecules.

Protein SHROOM2 is involved in endothelial cell morphology changes during cell spreading. In the retinal pigment epithelium, it regulates the biogenesis of melanosomes and promote their association with the apical cell surface by inducing gamma-tubulin redistribution[394]. SHROOM2 shares genomic localisation and functional features with the intracellular GPCR ocular albinism type 1 protein (OA1)[395], although no direct structural interaction has been found.

The last protein found by PrePPI was Periaxin (PRX), which however displayed the worst AlphaFold predictions. No interactions with GPCRs have been documented in the literature. Periaxin is a scaffolding protein that functions as part of a dystroglycan complex in Schwann cells, and as part of EZR and AHNAK-containing complexes in eye lens fiber cells. It is required for the maintenance of the peripheral myelin sheath that is essential for normal transmission of nerve impulses and normal perception of sensory stimuli[396].

4.4.6. Concluding remarks:

The advent of artificial intelligence in biological research is profoundly transforming our approach to studying complex biological systems. In this chapter, we employed the PrePPI tool to identify novel potential interactors of PAR4, complementing this with validation of proteomics hits using AlphaFold. AlphaFold has proven exceptionally useful in predicting the three-dimensional structures of proteins, although the field of AI-driven protein structure prediction is still in its infancy and faces particular challenges with full sequence predictions of multimeric protein interactions. Nevertheless, AlphaFold has shown remarkable accuracy and reliability when analysing peptide-domain interactions, significantly enhancing the robustness of its models.

In this study, AlphaFold was utilized to screen potential interactions with PAR4, serving as a preliminary step before further *in vitro* validation. Among the four potential interactions identified in the proteomics analysis - LMO7, PDLIM1, TJP2, and GIPC1 - AlphaFold highlighted an interaction with GIPC1 as particularly plausible (Figure 4.27). Therefore, this protein was chosen for detailed *in vitro* validation. Furthermore, given the availability of constructs for other members of the GIPC family (GIPC2 and GIPC3), these will also be examined for any potential interactions with PAR4.

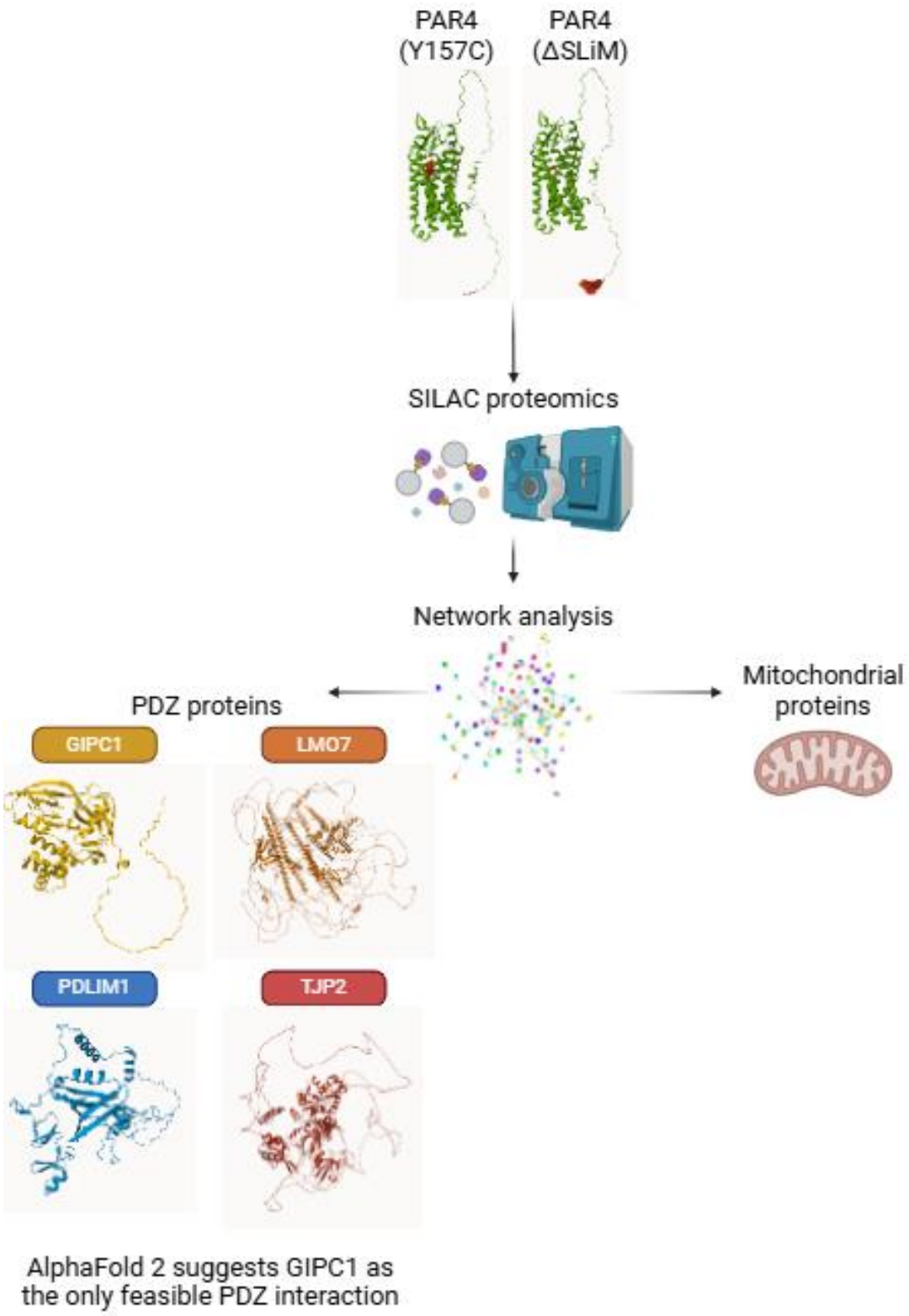


Figure 4.27 – Findings of Chapter 4. AlphaFold2 was used as a screening layer and found only GIPC1 as a feasible interaction with PAR4.

5. PRELIMINARY *IN VITRO* VALIDATION OF PROTEOMICS FINDINGS

5.1. INTRODUCTION:

The successful prediction and subsequent validation of GIPC1 binding to PAR4 underscores the effectiveness of using computational models as a preliminary screening step in the experimental workflow. Following the identification and computational prediction of protein-protein interactions between protease-activated receptor 4 (PAR4) and members of the GIPC family, this chapter focuses on the empirical validation of these interactions. Moreover, the results of the proteomic study indicated that a significant cluster of PAR4-interacting partners are localized to mitochondria with biochemical studies revealing for the first time that PAR4 is expressed on mitochondria. These novel findings open up a new avenue for PAR4 GPCR biology however they require further validation using different complimentary techniques to confirm expression and work towards understanding the role of PAR4 in the mitochondria.

5.1.1. GIPC family:

GIPC proteins are a family of PDZ domain-containing proteins that are pivotal in intracellular signalling pathways. The hallmark of GIPCs is the presence of a PDZ domain, which enables them to bind to the C-terminal motifs of various transmembrane proteins, thereby forming complexes with motor molecules such as myosin[397]. This is particularly relevant in the context of receptor pharmacology[398], as GIPC proteins can influence receptor internalization and recycling, affecting cellular responsiveness to growth factors, hormones, and other external stimuli. Besides the PDZ domain, GIPC proteins also contain other functional regions, including an N-terminal homology (GH) domain and a C-terminal acyl carrier protein (ACP) domain, though the functions of these regions are less understood.

These proteins serve as scaffolding molecules, facilitating the assembly of multi-protein complexes. Through interactions with different partners, including transmembrane receptors[399], kinases[400], and cytoskeletal proteins[401], GIPCs contribute to the fine-tuning of signalling cascades and cellular responses to external stimuli. Owed to the many interactions these proteins have, GIPCs are involved in the modulation of receptor signalling pathways, influencing multiple cellular processes such as cell adhesion[402], trafficking[339], and signal transduction[403], as well as cell migration[349], proliferation[404], and survival[405].

GIPC proteins are predominantly localised in the cytoplasm and are often associated with the plasma membrane or cytoskeletal components[342]. Their localisation is dynamic and can change in response to cell signalling events or during different phases of cell migration and adhesion. The localisation of GIPC proteins is crucial for their function, as it determines their ability to interact with specific partners and regulate signalling pathways effectively. GIPCs also play a role in the development and maintenance of cell polarity[406], are involved in vesicular trafficking processes[343], and their role in cancer progression and metastasis[338], as well as in angiogenesis[407], has been extensively studied, highlighting their potential as therapeutic targets.

5.1.2. Microscopy as a tool for probing protein localisation:

Classical microscopy is based on illuminating a specimen with a broadband light source and using optical elements to obtain a magnified image. It is one of the most widespread techniques in biological research, and the predominant imaging method[408]. This is because it is compatible with live cells, can be done noninvasively and at atmospheric conditions[408]. Far-field microscopy is a type of optical imaging technique that allows for the observation of microscopic samples with high resolution. Unlike near-field microscopy, which requires proximity between the sample and the optical detector, far-field microscopy can be used to observe samples that are further away. This technique utilizes the diffraction of light to produce a high-resolution image of the sample and is commonly used in various fields such as biology[409], materials science[410], and nanotechnology[411]. Far-field microscopy has several advantages, including its ability to image samples without damaging them and its compatibility with a wide range of samples, from tissues[412], to cells[413], to particles[410]. This makes it a valuable tool for various types of scientific and industrial applications[413]. Other microscopy methods such as scanning probe (near-field) or electron microscopy are also employed in research and can reach unparalleled resolution, since they are not limited by the diffraction of light[414]. However, these methods will not allow live imaging of specimens, are quite invasive and are time and energy expensive[415].

In fluorescence, absorption of a photon of energy from an excitation source, results in the emission of a photon of light with a longer wavelength. This process is called radiative decay and occurs when the energy absorbed by a molecule is re-emitted as light[416]. However, not all absorption events result in radiative decay. Some of the absorbed energy may be lost through non-radiative processes, such as internal conversion or intersystem crossing, which transfer the energy to vibrational or rotational modes of the molecule, instead of re-emitting it as light. Non-radiative processes compete with radiative decay and can reduce the overall efficiency of fluorescence[417].

Fluorescence was first observed in 1565[418], but it was not described until 1852 by George Gabriel Stokes. The change in excitation and emission wavelength was therefore named Stoke Shift[419]. It describes how excitation wavelengths always have higher energy than emission wavelengths. This is because phonon vibrations result in a thermal energy loss. Figure 5.1 describes this shift through a Jablonski diagram, which is a graphical representation of the electronic states and vibrational levels of a molecule[420].

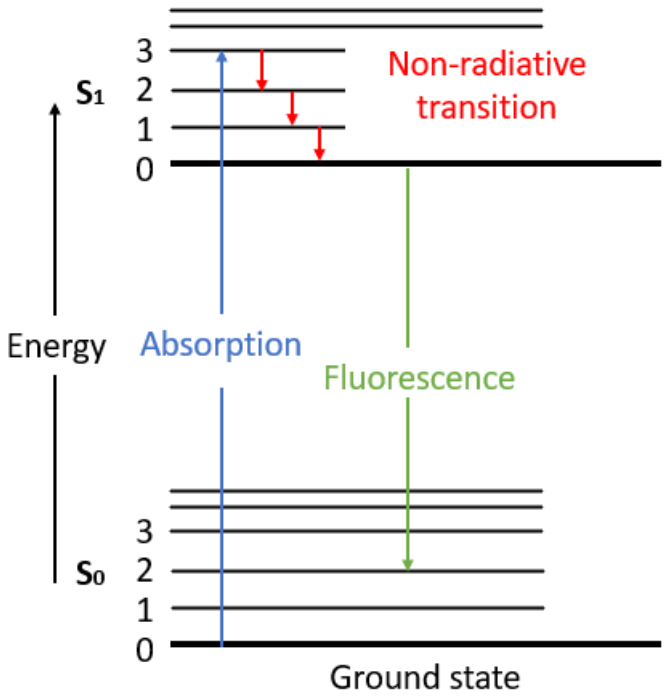


Figure 5.1 – Jablonski diagram illustrating the concept of fluorescence. Absorption of a high energy photon by a fluorescent molecule causes the emission of another photon with lower energy, thus causing fluorescence at a longer wavelength. The energy loss is due to phonon vibrations, which result in thermal energy loss[421].

In the past century, technical advances in physics, chemistry and optics allowed the development of fluorescence microscopy. In this method, specific molecules of interest are tagged with a fluorescent probe (named fluorophore) which absorbs a photon at a specific excitation wavelength and consequently emits another photon[422]. Fluorophores can be either dyes that chemically bind specific targets[423] (e.g.: Mitotracker with mitochondria), antibodies labelled with a fluorescent protein[424], or can be genetically expressed by transfecting cells with DNA containing a fluorescent protein[425] (e.g.: Green Fluorescent Protein, GFP). Since different fluorophores can be excited at different wavelengths, this technique allows the imaging of specific structures or features of cellular specimens with a high degree of specificity, provided that the excitation wavelengths do not overlap[426]. In multi-channel microscopy, samples tagged with different fluorophores can be imaged simultaneously, thus revealing different cellular features and potentially highlighting interactions between molecules[427]. As previously mentioned, this method is also non-invasive and compatible with live cells. This means that the specimen's biological activity can be recorded by taking multiple images at different timepoints, allowing to investigate the functional aspects of cellular biology[428].

However, common issues that occur with fluorescence microscopy are photobleaching and phototoxicity[429]. The excitation laser intensity needs to be optimised, as high-intensity illumination can cause the fluorophore to change its structure so that it can no longer fluoresce, this process is named photobleaching. This effect can be caused also by exposing the specimens too long. In a photobleached sample, the fluorophores are no longer excited, even when the

required light energy is provided[430]. As photobleaching, also phototoxicity is caused by high energy light or prolonged exposure times, however in this case are the cells directly to be damaged. Phototoxicity can cause formation of reactive oxygen species (ROS), cellular membrane blebbing, vacuole formation and even cell death[429]. These issues can limit the ability to observe biological samples over extended periods of time and can affect the accuracy of the data obtained through fluorescence microscopy. As a result, strategies must be employed to mitigate photobleaching and phototoxicity, such as reducing the intensity of the excitation light, using non-toxic fluorescent probes, or applying techniques such as super resolution microscopy that minimize light exposure.

5.1.3. The diffraction limit of light and the point spread function:

Resolution is defined as the smallest distance between two objects at which they can still be distinguished as separate[431]. Due to the dual particle-wave nature of light, eventually even the best microscopes have their resolution limited by the diffraction limit of light. In 1873 Abbe[432], demonstrated the need for optical lenses to be designed in such a way that the resolution was limited only by the diffraction limit of light, rather than other aberrations in the system. The diffraction limit of light can be calculated by using the wavelength of the light λ and the numerical aperture of the microscope NA. NA is the range of angles that the system can accept light from, and it is defined by the refractive index of the medium the light travels in n (1.00 for air, 1.33 for pure water, 1.52 for immersion oil), and the sine of the maximal half-angle of the cone of light that can enter the lens θ.

The diffraction limit d is thus described by the function:

$$1. \qquad d = \frac{\lambda}{2n \sin\theta} = \frac{\lambda}{2NA}$$

Therefore, microscopes with a higher NA value, can gather light over a larger set of angles, and therefore produce images with a higher resolution. Considering the instance of imaging a specimen emitting green light (λ=500 nm) using a modern instrument reaching a NA of ≈1.4, the diffraction limit would be roughly d≈ 200 nm depending on the instrument. This proves problematic when trying to image protein interactions as these kinds of molecules are in the 10 nm size range. This measure defines only lateral resolution perpendicular to the optical axis, but another important aspect is regarding the longitudinal resolution parallel to the optical axis (axial resolution). Abbe defined another formula to calculate axial resolution, as follows:

$$2. \qquad d = \frac{2\lambda}{NA^2}$$

Assuming the same values for wavelength and NA of the previous example, the axial resolution is ≈500nm, meaning that axial resolution is worse than lateral resolution.

Another definition of resolution commonly used is the Rayleigh criterion. It is a fundamental concept in optics that provides a measure of the minimum resolving power of an optical system. It states that two-point sources can just be resolved as separate entities if the centre of their diffraction patterns are separated by at least one-half of the wavelength of the light being used to observe them. In other words, the Rayleigh criterion sets a limit on the minimum distance between two points that can be distinguished as separate in an optical system. The Rayleigh criterion is used

to evaluate the performance of optical systems such as microscopes, telescopes, and cameras, and plays a critical role in determining the optimal design and performance of these instruments[431].

The Rayleigh criterion is described by the formula:

$$3. \qquad d = 1.22 \times \frac{\lambda}{NA}$$

The value of 1.22 in the Rayleigh criterion formula represents the first minimum in the Airy pattern, which is the diffraction pattern produced by an optical system when imaging a point source. The Airy pattern consists of a central bright spot, surrounded by alternating bright and dark rings, and the first minimum occurs at a certain radial distance from the centre of the pattern. The value of 1.22 in the Rayleigh criterion formula represents the radial distance from the centre of the Airy pattern to the first minimum, expressed in units of the wavelength of light being used[431].

A valid measure of the quality of an optical system is its point spread function (PSF). This is defined as a three-dimensional diffraction pattern of light emitted from an infinitely small point source. This 3D spot must not be considered uniform, as previously explained microscopes achieve better resolutions laterally than axially. The PSF can be considered as the fundamental unit of an image that can be detected by a system, and it is useful to determine how points are blurred in the image. As molecules emitting light lying within the PSF are impossible to distinguish, the PSF therefore gives a precise definition of the diffraction limit of a microscope[433].

To overcome this issue, and to image specimens under the diffraction limit of light, an array of super-resolution techniques have been developed in recent decades[409]. Over the course of this project two super-resolution microscopy have been employed, alongside a conventional confocal fluorescence microscopy. A brief overview of the rationale behind each method is briefly presented.

5.1.4. Confocal microscopy:

In widefield microscopy, the specimen is illuminated with a broad and uniform field of light, thereby enabling the collection of information from multiple focal planes. Some of these are in-focus, but some out of focus fluorophores can still be detected[434]. In confocal microscopy, these fluorophores still get excited, but a pinhole blocks most of the out of focus light and only fluorophores in one focal plane are detected[435]. A schematic of the difference between widefield and confocal microscopy is given in figure 5.2. Confocal microscopy not only improves the image quality, but has proven to be an excellent tool for investigating the subcellular localisation and trafficking of GPCRs[436]. The better quality of confocal imaging however comes at the cost of being more complex and time consuming than widefield[437]. This is because the image acquisition is based on scanning the specimen. Moreover, higher exposure to light can cause photobleaching or phototoxicity, as discussed previously.

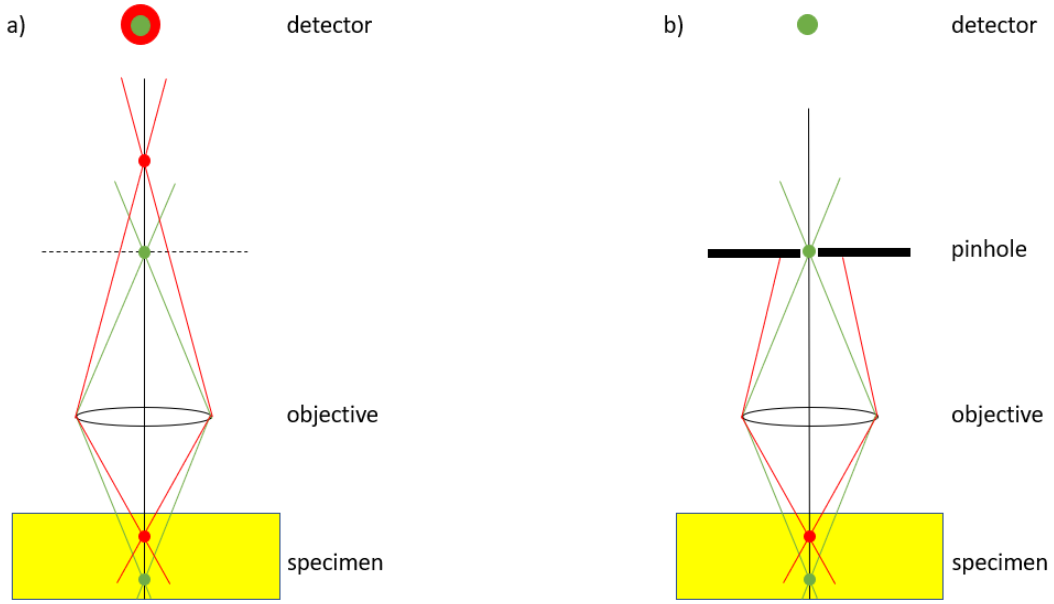


Figure 5.2 – Schematic of a confocal microscope. An objective lens in the microscope creates an image from the specimen which includes both focal (green) and extrafocal (red) signals. When a single spot is imaged, the detector records both a sharp spot-shape from the focal plane and blurred discs from other regions. By inserting a pinhole most of the extrafocal signal is removed, allowing only the focal plane emission to reach the detector[438].

The first confocal microscopy system was developed in Japan in 1943[439], but the first scanning microscope was built in 1955[440], whereby scanning was achieved by moving the stage. In the 1960s Czech researchers developed the Tandem-Scanning-Microscope, the first confocal to be commercialised[441]. It was not until 1969 that the first paper describing a confocal laser scanning microscope was published[442]. The first prototypes of the modern confocal microscope were developed by Amos in the 1980s[437]. Over the decades these instruments and their applications became more advanced. For example, by using fluorescent markers[443], or by implementing electronic autofocus, photomultiplier tubes (PMT) as detectors, stage scanning along the z-axis[444], laser point scanners with beam scanning[437].

Compared to widefield, confocal microscopy allows to improve the resolution of an image by closing the confocal aperture to eliminate higher orders of the diffraction pattern. This increase in sharpness, comes at the cost of decreased brightness[435]. Since only a small number of photons is available in fluorescence microscopy, the signal-to-noise ratio of confocal microscopy can often be a limiting factor. This issue can be addressed by employing more sensitive detectors, or by increasing the laser power. However, this latter strategy may lead to photobleaching and phototoxicity as discussed previously[445].

Confocal microscopy has been extensively employed to investigate GPCR biology, especially in the context of spatiotemporal investigation, subcellular localisation, trafficking and colocalization studies to investigate dimerization as well as signalling pathways[446,447]. This microscopy has been useful also in drug discovery efforts, allowing for high throughput screening of GPCRs and associated messengers such as β-arrestins[448,449].

5.1.5. Stimulated emission depletion (STED) microscopy:

The super resolution microscopy employed to investigate whether PAR4 was present in mitochondria was Stimulated emission depletion (STED) microscopy, which achieves an increase in diffraction resolution by reducing the effective PSF. The idea of STED was first patented by Victor Okhonin in 1986[450], but the microscope was developed in 1994 by Stefan W. Hell and Jan Wichmann who were unaware of the patent[451]. The first experimental demonstration was in 1999 by Thomas Klar and Hell again[452], who was then awarded the Nobel Prize in Chemistry in 2014[453].

As mentioned previously, fluorescence is based on the excitation of an electron from the ground state (S0) to an electronic state of different fundamental energy (S1). A photon is released after the relaxation from S1 back to S0[426]. STED blocks this process before the photon is emitted. The excited electron is forced into a higher vibration state than the one that would be achieved by fluorescence; thus the photon is redshifted[454]. Since the electron goes to a higher vibrational state, the energy difference of the two states is lower compared to fluorescence difference. Lower energy translates to a longer wavelength, thus shifting the photon into the far red of the spectrum (Figure 5.3).

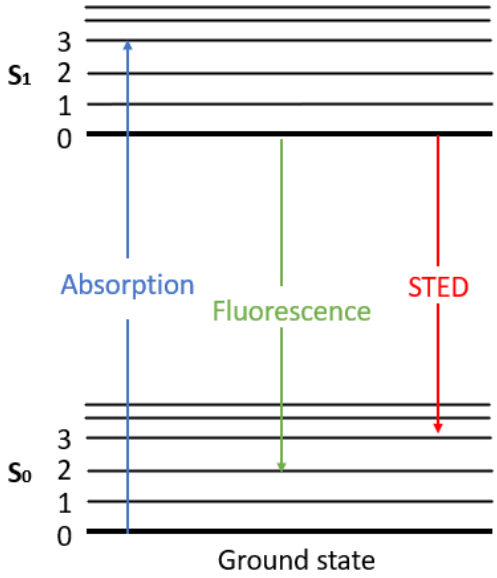


Figure 5.3 – Jablonski diagram of STED microscopy. The STED beam depletes excited fluorophores to a higher vibrational level. The lower difference in energy translates to higher wavelength, which shift the light emission towards the deep red spectrum (red shifting).

This alternative emission is referred to as depletion, and it is forced by striking the fluorophore with an incident photon. The number of incident photons determines the efficiency of the depletion (Figure 5.4); fluorescence can be completely suppressed if enough incident photons are employed[455]. More photons translate to higher laser intensity. As discussed previously, too high intensity laser might cause photobleaching the fluorophore.

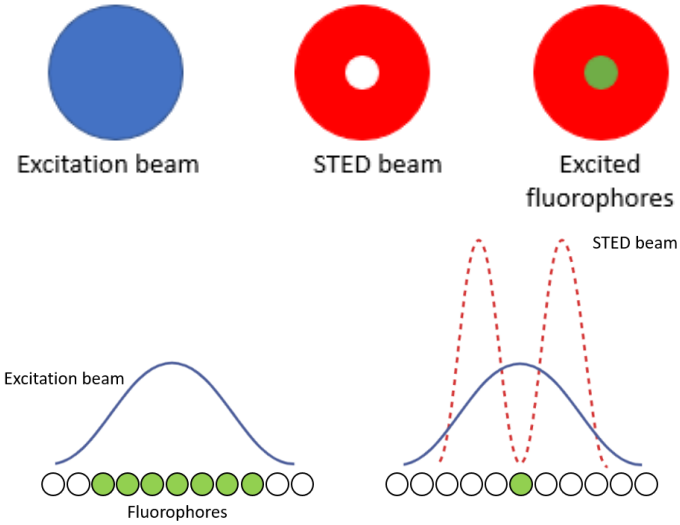


Figure 5.4 – Schematic of excitation and depletion wavelengths. Higher intensity in the STED beam allows to increase resolution by reducing the size of the donut, thus depleting more fluorophores, and leaving a smaller portion in the excited state.

STED is based on a laser scanning confocal microscopy; however, a second hollow donut-shaped beam follows the initial excitation laser (Figure 5.4). The ring of the

donut overlaps with the excitation beam, thus quenching the excited fluorophores and bringing them back to the ground state. Therefore, only the fluorophores at the centre of the donut emit fluorescence, achieving a sub-diffraction emission area[451]. To obtain the donut-shaped STED beam, light is polarised using a circular polariser. This, way the STED laser has a very low, ideally zero, intensity at the centre. Another crucial aspect of this microscopy is the time difference between the excitation and the STED beams. This is due to the emission depletion, as the molecules in the excited state need to be depleted before a photon is emitted. Properly calibrated instruments can achieve a time difference down to the femtoseconds, which greatly improves the signal-to-noise ratio[456]. STED microscopy theoretically offers unlimited resolution power, which is dependent on the STED beam intensity. However, since resolution enhancement is related to the intensity of the beams, higher intensity translates to higher photodamage to the specimen. This means that higher intensity can be used of fixed cells, while live cell imaging requires lower laser intensities. Therefore, a STED microscope can usually image structures down to 40 nm in live cells[457], and 20 nm in fixed cells[458]. In our laboratory we used a Picoquant MicroTime 200 STED. This microscope has multiple excitation line. The lasers compatible with STED have excitation wavelengths at 594 and 640 nm, while the STED beam wavelength is 765 nm. Moreover, the instrument gives the opportunity to manually change the filters setup, depending on the emission light that needs to be isolated. The hardware setup of the instrument is shown in figure 5.5.

The application of STED to study GPCR biology is relatively new, although this super resolution microscopy has been employed to visualise the clustering of cannabinoid receptor 1 in neurons[459], as well as investigate the distribution of endogenous glucagon-like peptide 1 (GLP1)[460]. A comprehensive review on microscopy and spectroscopy approaches to study GPCR structure and function was given by Fessl, Majellaro and Bondar in 2023[461].

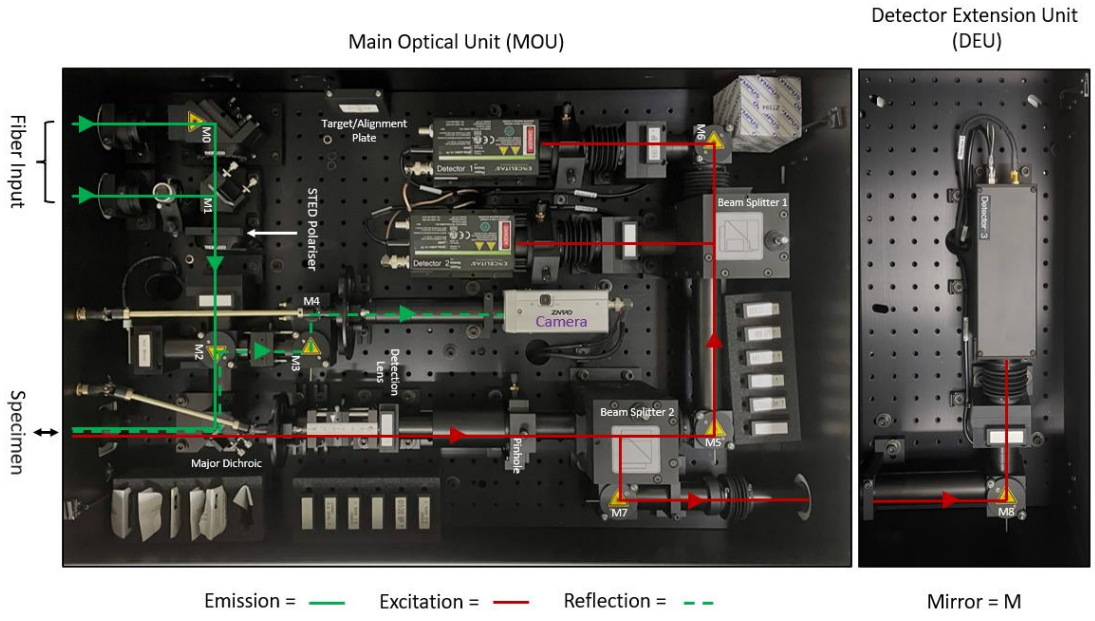


Figure 5.5 – Top-down view of the main optical unit of Picoquant MicroTime 200. The excitation pathway and reflected excitation light from the specimen cover glass

boundary is shown in green. The optical path of the resulting fluorescence emission is shown in red. Graph created by PhD student Angéline Geiser.

5.1.6. Structured illumination microscopy (SIM):

Another super resolution imaging technique that can be utilised is structured illumination microscopy (SIM)[462]. SIM uses a specific pattern of illumination to effectively surpass the diffraction limit and reveal details that would otherwise be hidden in confocal microscopy. This gives rise to a well-known mathematical concept called moiré effect, where superimposing two high-frequency fringes (in this case the object and the illuminating light), gives rise to low frequency patterns[463]. For this specific interference to appear, the two overlaid fringe patterns need to be slightly displaced, for instance by rotating them.

The first practical application of structured illumination was in 1993 when standing wave fluorescence microscopy (SWFM) was first introduced[464]. Concurrently with the development of this technique, 4Pi and I5M microscopies were also developed. 4Pi microscopy, also known as dual-objective 4Pi microscopy, uses two objectives to illuminate the sample with a pair of orthogonal patterns. This results in a significant improvement in axial resolution compared to traditional SIM techniques[465]. i5M (interference-based 5D microscopy) is a newer variant of SIM that combines the principles of 4Pi with confocal microscopy. This allows for high-resolution imaging in all three dimensions, with greatly improved axial resolution compared to standard confocal microscopy[466]. In 1997, structured illumination was applied to a conventional widefield microscope, in a method called optical sectioning structured illumination microscopy (OS-SIM)[467]. In the new millennium multiple techniques have been developed to achieve super resolution using structural illumination, including laterally modulated excitation microscopy (LMEM)[468], harmonic excitation light microscopy (HELM)[469], but one of the most relevant discoveries was super resolution structured illumination microscopy described by Gustafsson in 2000, where he demonstrated double lateral resolution[470]. In 2008, using a three-beam interference pattern, Gustafsson modified his initial SR-SIM setup to double also axial resolution[471].

As previously stated, the object is illuminated with patterns containing high spatial frequencies and the image obtained is mapped to a function using Fourier transform. This is a mathematical transform which maps a function to its frequency components. The high frequencies of the pattern and object are changed to low frequencies in the image transform using a convolution operation (a mathematical operation that expresses how the shape of a function is modified by another function). The low frequencies image from this transform is finally reconstructed using an algorithm[472].

Figure 5.6 helps understand how SIM generates a resolution increase greater than the diffraction limit. The graph is represented in Fourier space, which is a space where Fourier transform maps frequencies to function. Towards the centre of the graph, is where the low spatial frequency information is. Towards the edge is mapped the high spatial frequency, which are very fine features of the image. Since a microscope objective cannot be infinitely large, the instrument has some finite size of back aperture. This means that low spatial frequencies within the size of the back aperture, will be able to go through the objective and be detected. On the other hand, high spatial frequencies outside of the back aperture will be cut out. To obtain sharper images with better detail resolution, high frequency information from the edges of the

Fourier space needs to be detected. These information fringes can be observed by SIM microscopy as the moiré fringes, which have lower frequency. Rotating the grid pattern allows to obtain moiré fringes in different directions, thus collecting more high frequency information, which finally results in an increase of detail resolution[473].

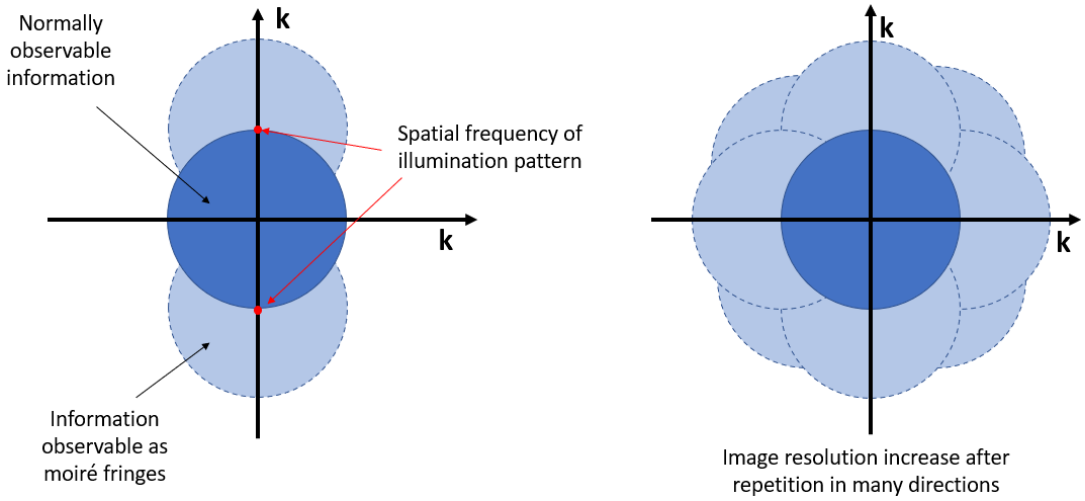


Figure 5.6 – Resolution extension by Structured illumination. The graph is represented in Fourier space and both axes represent the frequency of the image. Information at lower frequencies is at the centre of the graph and can be observed. Higher frequency information is towards the periphery of the graph and gives higher resolution to the image details. Moiré fringes lower the frequency of the image and allow to detect more information.

The observable region of reciprocal space produced by a microscope objective (which is analogous to its diffraction pattern) is limited at the edges by the highest spatial frequencies that the objective can transmit ($2NA / \lambda$)[473]. That is because this value will yield the most information in the moiré fringes. By rotating the grid several times, more information will be generated, and thus greater imaging resolution achieved. A limitation of structured illumination is the best resolution increase is a factor of two. This is because the illumination pattern frequency can be the diffraction limit value at most. It is impossible to create a pattern with higher frequencies than the one defined by the diffraction limit of the objective lens. Phase changes are used to extract higher frequency information, therefore multiple images are required[472]. Multiple images are taken using a different pattern for each focus plane, and the image its then reconstructed computationally, removing the effect of the structure, thus obtaining a resolution enhancement[462]. For a 2D image, a minimum of 9 images is required[470]. Different patterns can be applied, with the most common being parallel lines that are shifted and rotated to obtain a grid. However, it must be noted that these angle changes to obtain isotropic resolution enhancement, result to improved resolution only in the axis of the grating (X and Y), while no resolution enhancement is obtained along the Z-axis[471]. In the context of GPCR research, SIM has mainly been employed to study β-arrestins, and β-arrestin mediated receptor trafficking. One study found β-arrestin-2 in the membrane inside and outside of clathrin-coated pits[474], while another study detected the differential segregation of activated β_2-adrenoceptor molecules phosphorylated by GRKs[475].

5.1.7. Proximity ligation assay (PLA) as tools for PPI detection:

Proximity Ligation Assay (PLA) is a molecular technique that allows the detection of proteins, protein interactions, and post-translational modifications at very low concentrations. The fundamental theory behind PLA involves the use of two probes—each attached to an antibody specific to the target protein or protein complex. These probes are conjugated with oligonucleotides which, when brought into proximity by the binding of the antibodies to their target or targets, can be ligated to form a circular DNA molecule. The ligation event only occurs if the probes are in close proximity, typically within 40 nm of each other, which is indicative of a molecular interaction between the target proteins. This proximity-dependent ligation is then followed by the amplification of the circular DNA by rolling circle amplification (RCA), a process that creates a concatemer of repeated DNA sequences that can be easily detected through fluorescently labeled complementary DNA probes. The combination of proximity-dependent ligation and RCA allows for highly sensitive and specific detection of proteins and their interactions, which can be visualized and quantified using standard fluorescence microscopy techniques or other methods depending on the configuration of the assay[476].

PLA has been extensively applied in the study of GPCRs to understand their dimerization and interaction with other proteins, which are crucial aspects of their function and regulation. For example, Paek and colleagues combined peroxidase-catalysed proximity labelling with mass spectrometry to track GPCR signalling dynamics in living cells, offering insights into the spatial and temporal aspects of GPCR function[477]. Another group used in situ PLA to validate the existence and characterize the distribution of GPCR homo- and heteroreceptor complexes in the brain, providing a tool to study the molecular basis of neuronal communication and its alterations in disorders[478]. This technique has not been applied just to GPCRs, but also to their downstream signalling proteins, such as β-arrestins, which revealed a wide array of β-arrestin 1 interacting partners[479].

5.2. CHAPTER AIM:

This chapter will conduct *in vitro* validation of the interaction between PAR4 and GIPC1, utilizing confocal microscopy to visualize co-localization within cells, Proximity Ligation Assay (PLA) to confirm close-proximity interactions indicative of molecular binding between these proteins. The potential interactions between PAR4 and other members of the GIPC family, specifically GIPC2 and GIPC3, will be explored. This extension aims to assess the generality of the GIPC-PAR4 interaction across the GIPC family, which may indicate a broader functional role for these interactions within the signalling pathways mediated by PAR4.

The primary aim of this chapter is to investigate the subcellular localization and protein interactions of PAR4, focusing on its potential localization within mitochondria and its interactions with the GIPC protein family. This dual approach stems from proteomic data suggesting interactions between PAR4 and mitochondrial proteins, as well as predicted interactions with GIPC1 identified through AlphaFold Multimer.

To corroborate the hypothesis that PAR4 localizes within mitochondria and interacts with key mitochondrial components, advanced microscopy techniques, including confocal microscopy, Structured Illumination Microscopy (SIM), and Stimulated Emission Depletion (STED) microscopy, will be employed. These techniques will provide high-resolution imaging to confirm the mitochondrial localization of PAR4.

The hypotheses of this chapter are that:

1. PAR4 localizes to mitochondria, as suggested by proteomic data. This hypothesis is tested using advanced imaging techniques, including confocal microscopy, Structured Illumination Microscopy (SIM), and Stimulated Emission Depletion (STED) microscopy, to confirm its subcellular localization and assess its association with mitochondrial components.
2. PAR4 interacts with members of the GIPC family (GIPC1, GIPC2, and GIPC3) via molecular binding, as predicted by computational models and proteomic findings. This hypothesis is investigated using Proximity Ligation Assay (PLA) to validate close-proximity interactions and confocal microscopy to visualize potential co-localization of PAR4 with GIPC proteins within cells.
3. The interaction between PAR4 and GIPC proteins may extend across the GIPC family, suggesting a broader functional role for these interactions within PAR4-mediated signalling pathways. This objective involves testing GIPC2 and GIPC3 for potential interactions, complementing findings on GIPC1.

5.3. RESULTS:

Since a large subset of PAR4-interacting proteins was found in mitochondria, or involved in mitochondrial functions, the presence of PAR4 in this organelle was questioned. To validate this hypothesis, cells were transfected with the same plasmids used for the proteomic study (untransfected control, YFP, PAR4-YFP, $PAR4_{\Delta SLiM}$-YFP), and the presence of the receptor in mitochondria was investigated by both biochemical and microscopy techniques.

5.3.1. Optimisation of mitochondrial dyes:

Since the aim of the project was to determine if PAR4 may be localised in mitochondria, a crucial aspect was using a viable method to image these organelles. The excitation lasers compatible with STED are only at wavelengths 594 and 640 nm. This means that although the instrument is perfect to image specimens in the red and deep red wavelengths, it is not suitable to image proteins tagged with YFP, since these molecules are excited at 513 nm, and possess an emission peak at 530 nm. Since in previous experiments PAR4 was tagged with YFP, a different fluorescent protein with a different excitation and emission spectrum had to be used. Fortunately, PAR4 constructs were available tagged with mCherry as well. This fluorescent protein has an excitation maximum of 587 nm, and an emission maximum of 610 nm, lying within the PicoQuant light filters, and allowing to be detected together with MitoTracker Deep Red. Previous research has employed immunofluorescence using antibodies specifically targeting mitochondrial proteins such as translocase of the outer membrane (TOM)[480], translocase of the inner membrane (TIM)[481] and voltage dependent anion channel (VDAC)[482]. These methods can reveal the presence and localisation of mitochondria; however, they are not suitable for this specific project. Proteins such as VDAC, for example, may not localise only in these organelles, as they are found also at the cellular membrane[274], and colocalisation with PAR4 would not necessarily mean that the complexes would be found in mitochondria. For this reason, mitochondria were stained with the selective dye MitoTracker[483]. These dyes come in a range of colours and possess a chloromethyl moiety that acts as a thiol-reactive site, thus anchoring the probes to the mitochondrial membrane[484]. The MitoTrackers employed in this project were MitoTracker Red FM for SIM microscopy and MitoTracker Deep Red FM for STED. As discussed previously, PAR4-GFP tagged could not be used for STED microscopy, therefore mCherry tagged proteins had to be employed. As shown in supplementary figure S7, the fluorescent protein's excitation and emission profile lied quite closely to the MitoTracker dyes, however a slightly better separation was achieved with Deep Red. This dye was therefore selected for STED microscopy.

The dyes had to be optimised before investigating for the presence of PAR4. In the supplier's protocol, it was suggested an initial working concentration range of 25-500nM with an incubation time of 15-45 minutes. However, since the cells were going to be fixed, the concentration range was reduced to 100-500 nM. Moreover, to reduce potential artifacts and mitochondrial toxicity, the concentration was always kept as low as possible. Therefore, the initial imaging was performed at 100 nM MitoTracker Deep Red FM, with an incubation time of 15 minutes. Four different cell types were dyed to check whether any differences would occur in mitochondrial staining. The cells were cardiac fibroblasts (CFs), smooth muscle cells (SMCs), cardiac myocytes (AC16s) and human embryonic kidney transfected with PAR4 (HEK-PAR4). The working concentration was too high and the cells were stained all over, making impossible the visualisation of mitochondria, moreover the images were obtained using a minimal laser intensity on the confocal microscope, which is good to avoid phototoxicity, but highlights again how the dye concentration was too high. HEK293 cells were selected to optimise the working concentration, since they were the cells used for transfection in the proteomic study. As the cells were stained with a working concentration lower than the recommended value, the incubation time was increased from 15 minutes to 30 minutes. Together with the working concentration, the laser profile on the microscope settings was optimised as well. The working concentrations shown in figure 5.7 were: 10 nM, 25 nM, 50 nM, and a specimen of 100 nM was imaged again with the new laser settings to show how the cells were overloaded with the dye previously.

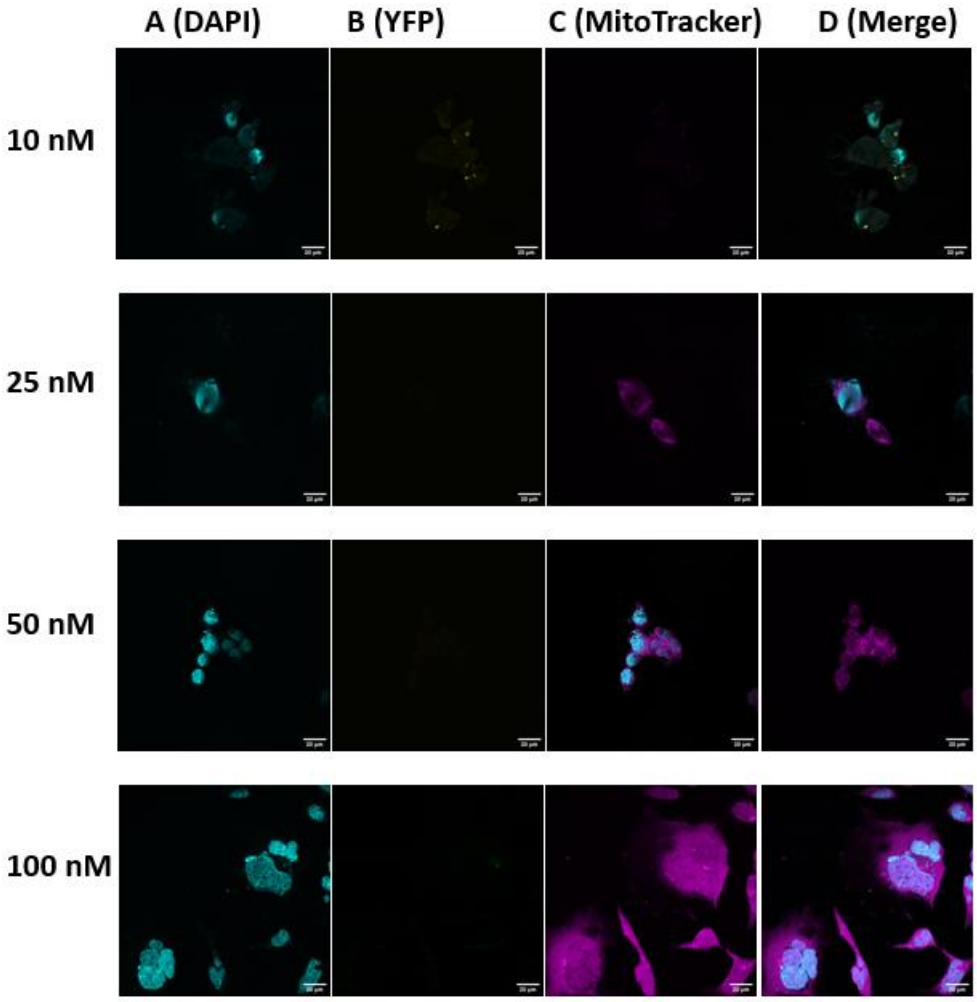


Figure 5.7 – Optimisation of MitoTracker Deep Red FM working concentration. HEK293 cells were incubated with different concentrations of the dye for 30 minutes. Already at a concentration of 50 nM it becomes difficult to distinguish Mitochondrial structures. **A:** Cyan: DAPI, **B:** yellow channel (empty), **C:** magenta: MitoTracker Deep Red, **D:** merge.

5.3.2. Resolution improvement of mitochondrial imaging:

Confocal microscopy is a widely used imaging technique that provides high resolution images of biological specimens. However, it is still limited by the diffraction limit, which limits the smallest details that can be resolved. This means that some fine structures within the sample remain unresolved, hindering our ability to gain a full understanding of the specimen. Since MitoTracker staining resulted in very strong signals all across the cell, confocal microscopy was not sufficient to discern pixels displaying genuine mitochondrial staining. Therefore, stimulated emission depletion (STED) microscopy was applied to achieve a better resolution. Unfortunately, the STED laser allows for a narrow set of wavelengths in the far-red end of the colour spectrum to be excited. Therefore, it was not possible to image YFP constructs since their excitation wavelength is too distant from the depletion laser. mCherry was lightly excited, but not even with lasers at full power was it possible to image the receptor. Nevertheless, the comparison between confocal microscopy and STED is given in figure 5.80.

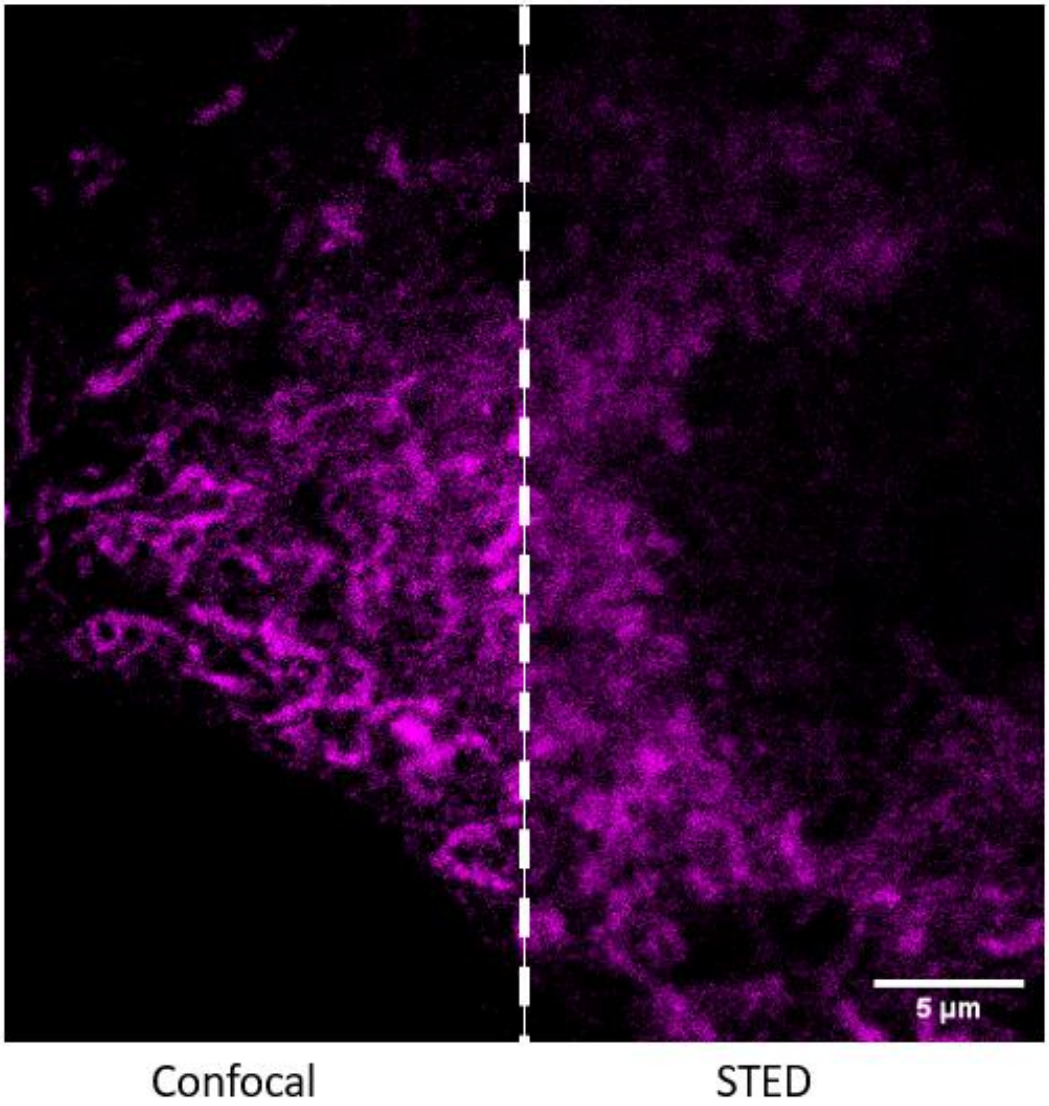


Figure 5.8 – Comparison in resolution between confocal and STED microscopy on the PicoQuant Microtime 200. HEK293 cells were stained with 25 nM MitoTracker Deep Red for 15 minutes. Images were taken consequentially first on the confocal, then gated STED was applied.

At the Centre for Microsystems and Photonics of the University of Strathclyde, a custom-made SIM microscope has been developed. The peculiarity of this instrument is that the grating pattern is generated using two optical beams controlled via 2 micro-electro-mechanical system (MEMS) three-axis scanning micromirrors. This allows a precise position of the optical beams, resulting in a resolution improvement between 1.3 and 1.8 the diffraction limit of light[485]. The microscope has been tested to image mitochondria stained with Mitotracker. Initially, the instrument did not possess excitation lasers in the deep red spectrum, therefore a different dye was applied. In supplementary figure S8 the imaging capability of the microscope was tested by buying a premade slide of Bovine Pulmonary Artery Endothelial (BPAE) cells, stained with MitoTracker Red. 18 images were taken with the custom-made imaging system,

and the images were reconstructed using the fairSIM plugin of ImageJ. The reconstructed image resulted in a clear definition of mitochondrial networks, showing the organelles in a much higher resolution when compared to the Leica SP8 confocal microscope. Together with STED, this custom-made SIM microscope will be used to investigate the existence of mitochondrial PAR4.

5.3.3. HEK293 cells transfection using different constructs:

After optimizing the MitoTracker concentration to 25 nM and the incubation period to 15 minutes, cells were transfected for 48 hours with YFP, PAR4-YFP, and $PAR4_{\Delta SLiM}$-YFP constructs. This experimental setup aimed to determine whether PAR4 localizes to mitochondria and if the deletion of SLiM affects this localization. The presence of YFP samples was to ascertain that the eventual colocalization with mitochondria was due to the receptor and not the fluorescent tag. The transfections and the staining were successful, as evidenced by the strong signal detected for all constructs in both the green (YFP) and red (MitoTracker) channels. However, the signal intensity was so high that fluorescence was observed throughout the entire cell, even at very low laser power and gain settings. This resulted in non-specific staining, as illustrated in Figures 5.9, where not only mitochondria but also other cytosolic structures were dyed. In the green channel, a similar phenomenon was observed, with fluorescence detected across the entire cell. Consequently, this led to a high degree of pixel colocalization in the green and red channels for all constructs, rendering it impossible to accurately determine and evaluate the extent of PAR4's colocalization with mitochondria. For this reason, the three constructs achieved very similar Pearson's r coefficients: YFP= 0.641±0.178, PAR4= 0.596±0.188, ΔSLiM= 0.614±0.211 (Figure 5.10). These findings suggest that the strong fluorescence signals may obscure finer details of protein localization, necessitating further optimization of imaging conditions or alternative methodologies to reduce non-specific staining and improve the specificity of mitochondrial localization assessments.

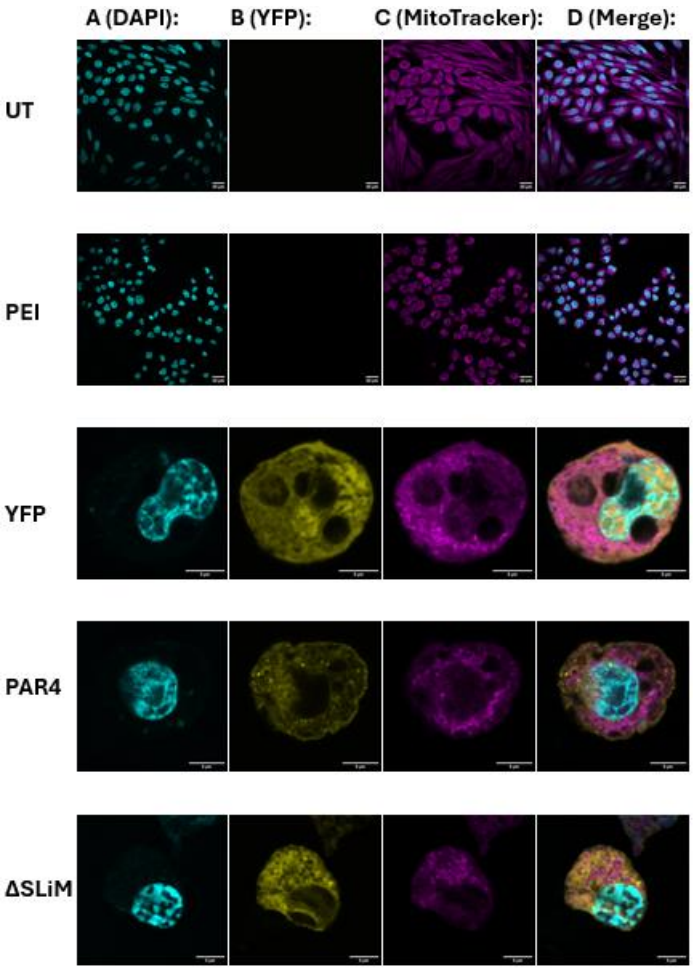


Figure 5.9 – Colocalisation of PAR4 and YFP constructs with MitoTracker Deep Red FM. Cells were transfected with different constructs for 48 hours, followed by mitochondrial staining with 25 nM MitoTracker Deep Red FM for 15 minutes. **A:** DAPI, **B:** YFP, **C:** MitoTracker, **D:** Merge.

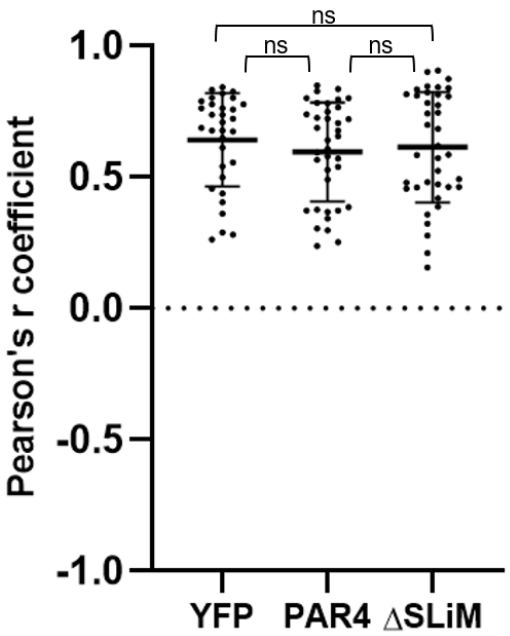


Figure 5.10 – Different YFP constructs achieve the same correlation coefficient. YFP, PAR4-YFP, and PAR4$_{\Delta SLiM}$-YFP were transfected in HEK293 for 48 hours, followed by 25 nM MitoTracker staining for 15 minutes. Colocalization between the transfected proteins and mitochondria were investigated by confocal microscopy. Colocalization of 35 measurements was quantified using the Pearson's r coefficient, resulting in YFP= 0.641±0.178, PAR4= 0.596±0.188, ΔSLiM= 0.614±0.211.

5.3.4. Colocalisation of PAR4-mCherry with MitoTracker Deep Red:

To investigate whether the colocalization was dependent on the YFP tag, cells were also transfected with mCherry constructs. Due to the unavailability of ΔSLiM-mCherry constructs, only cells transfected with PAR4-mCherry were imaged. The results suggested no significant colocalization of PAR4 with mitochondria, as shown as the example of Figure 5.11. This observation was further supported by quantitative analysis, where 18 measurements were grouped, resulting in an average colocalization coefficient of r=0.383±0.173, as shown in Figure 5.12. These findings imply that the use of the mCherry tag does not support significant colocalization of PAR4 with mitochondria under the given experimental conditions, contrasting with the results observed with YFP-tagged constructs. This discrepancy highlights the importance of tag selection in colocalization studies and suggests that further investigation is needed to reconcile these differences and understand the underlying mechanisms.

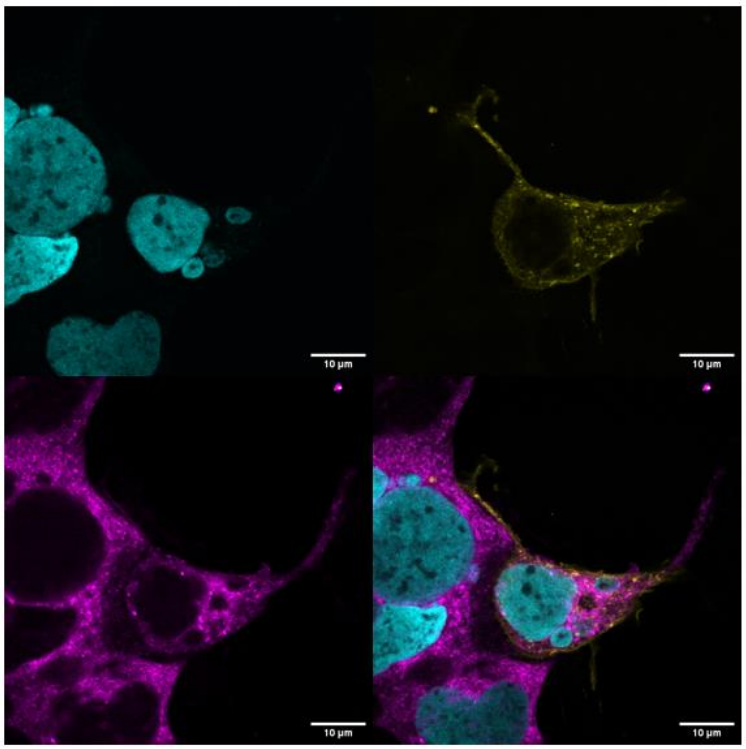


Figure 5.11 – PAR4-mCherry constructs do not colocalise with MitoTracker Deep Red. HEK293 cells were transfected with PAR4-mCherry constructs and stained with DAPI (nucleus) and MitoTracker Deep Red (mitochondria). mCherry is a fluorescent protein that emits light in the wavelength range of 550-650 nm.

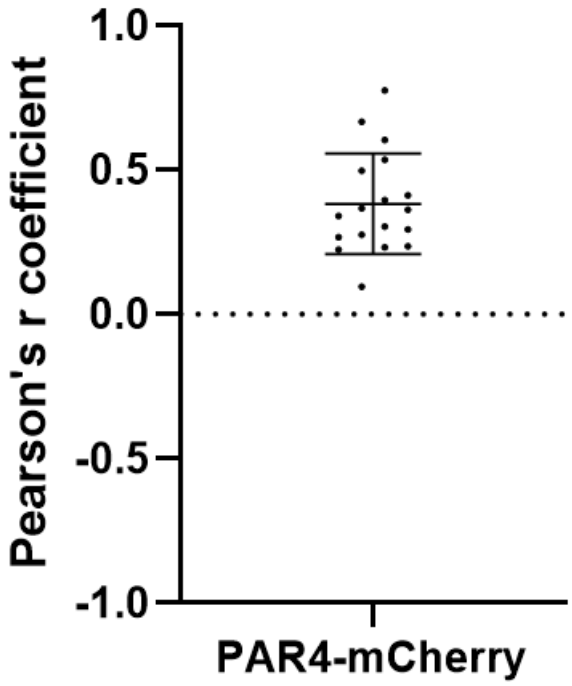


Figure 5.12 – PAR4-mCherry does not colocalise with MitoTracker Deep Red. 18 measurements were taken and resulted in an average Pearson's coefficient r=0.383±0.173.

5.3.5. Western blots suggest PAR4 can be found in mitochondria:

Mitochondria were extracted using the Mitochondria Isolation Kit for Cultured Cells (Thermo Fisher). This method yields three fractions: the first lysate contains various contaminants from the membrane and other organelles (therefore named total), the second is a cytosolic fraction, while the third and last one contains purified mitochondria. Lysates were collected from each transfected sample, and the fractions were analysed by western blot using an anti-YFP antibody. This antibody was used instead of an anti-PAR4 in order to validate that the localisation of the receptor to mitochondria was not owed to the presence of the fluorescent protein. VDAC1/3 were used as internal controls for total and mitochondrial fractions. Since VDAC was not found in the cytosol, Tubulin was used as a cytosolic marker. Figure 5.13 shows how the receptor was found in mitochondrial fractions, and deletion of the short linear motif (SLiM) reduced its levels in mitochondria. The presence of PAR4 in these organelles was not due to the fluorescent tag, as transfection with just YFP confirmed the presence of the fluorescent protein in the total and cytosolic fractions, but not in the mitochondrial one.

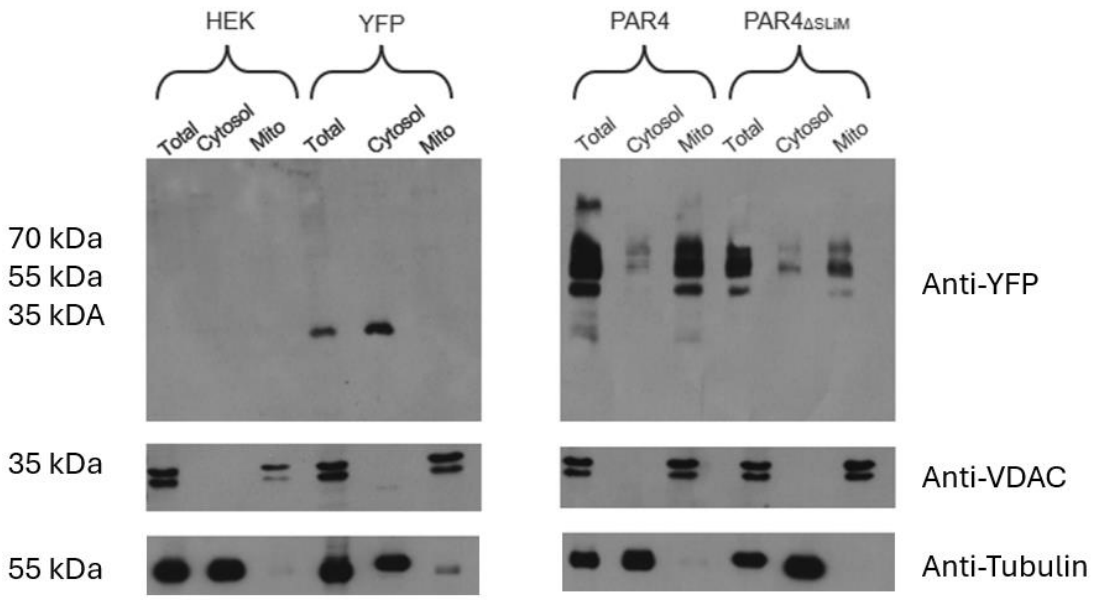


Figure 5.13 – Intracellular localisation of PAR4. A: Western blot analysis confirms the presence of PAR4-YFP in mitochondria. Cells were transfected for 48 hours with YFP, PAR4-YFP, $PAR4_{\Delta SLiM}$-YFP. Mitochondria were isolated using commercially available kits and the three fractions were analysed by western blot using anti-YFP antibodies. Only PAR4-YFP and $PAR4_{\Delta SLiM}$-YFP were detected in mitochondrial fractions. **B: Schematic of potential localisations of PAR4 within the cell.** Image generated in BioRender.

5.3.6. PLA confirms interactions between PAR4 and GIPCs:

Figure 5.14 illustrates the expression and localization of PAR4-mCherry and GIPC1-GFP, GIPC2-GFP, and GIPC3-GFP in HEK293 cells. Co-transfection of PAR4 and the three GIPC isoforms was confirmed by Western blotting, showing distinct bands corresponding to the expected molecular weights of the constructs, normalized to Tubulin as a loading control (Figure 5.14 A). Confocal microscopy further highlights the cellular localization of GIPC1-GFP, with fluorescence observed in close proximity to the nucleus, as indicated by DAPI staining (Figure 5.14B). These results validate the expression of PAR4 and GIPC isoforms.

As illustrated in Figure 5.15, GIPC1 demonstrated pronounced interactions with PAR4. Numerous red fluorescent spots, indicative of close proximity interactions, were observed adjacent to areas expressing transfected GIPC1. Notably, some red spots appeared isolated from the green regions, which can be attributed to interactions occurring on different focal planes in the Z-axis, not captured in the single plane displayed. Interestingly, GIPC1 frequently formed ring-like structures around the nuclei, suggesting a potential perinuclear localization of these interactions.

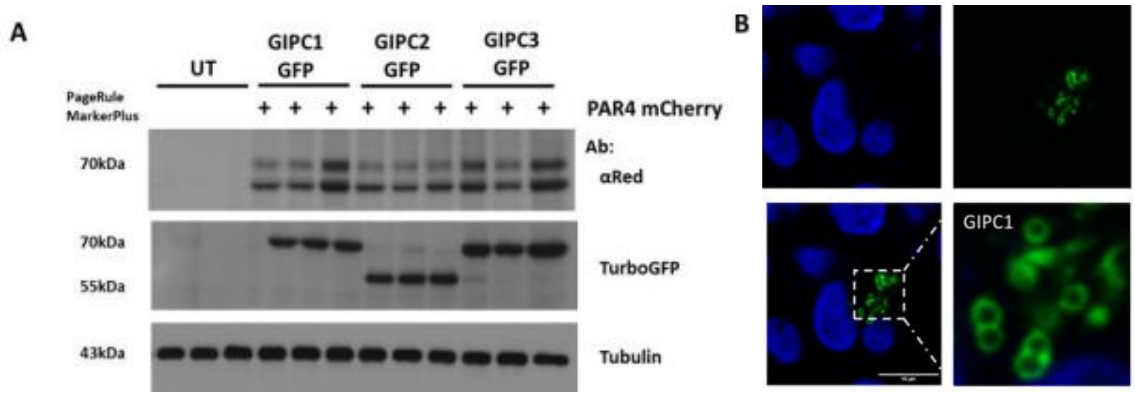


Figure 5.14 - Expression of PAR4-mCherry and GIPC-GFP Isoforms in HEK293 Cells. Protein expression levels of PAR4 and GIPC1-3 isoforms in co-transfected HEK293 cells. Cells were cultured and co-transfected with PAR4-mCherry and GIPC1-GFP, GIPC2-GFP, or GIPC3-GFP for 48 hours, followed by lysis and protein extraction. Western blot analysis was performed to resolve the samples, with membranes probed using antibodies specific to mCherry and Turbo GFP. **A**: Detection of PAR4-mCherry and GIPC-GFP constructs on nitrocellulose membranes alongside untransfected (UT) control samples. Tubulin is shown as the housekeeping protein. **B**: Confocal microscopy images showing the localization of GIPC1-GFP in HEK293 cells. The images include DAPI staining for the nucleus, GIPC1 fluorescence, merged DAPI/GIPC1 channels, and an enhanced view of the merged image.

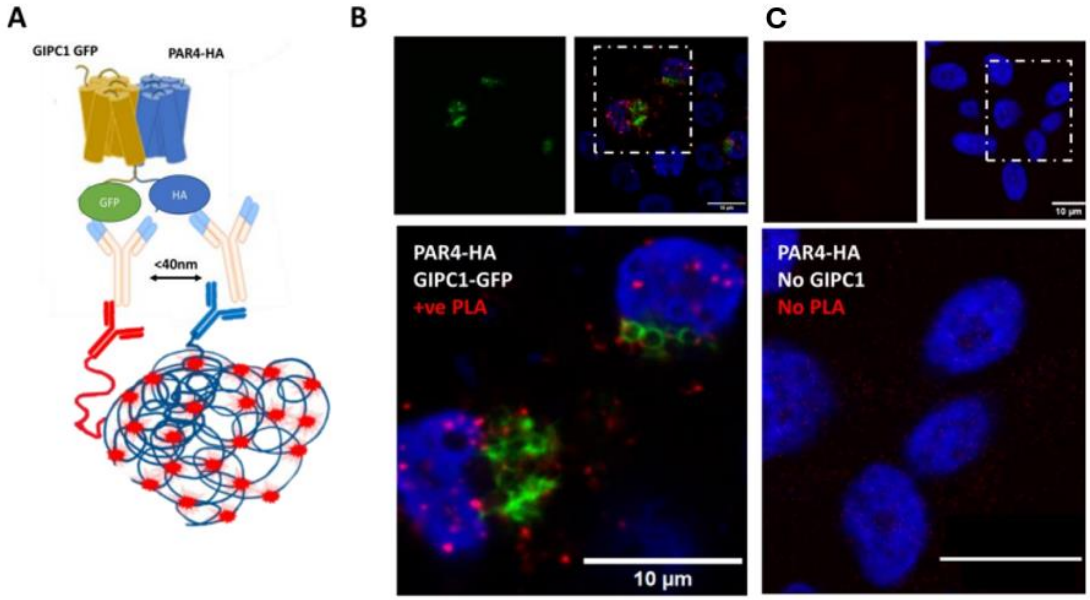


Figure 5.15 – PLA confirms interactions between PAR4-HA and GIPC1-YFP. HEK293 cells were transfected for 48 hours with 0.5 µg PAR4-HA and 0.5 µg GIPC1-YFP (green). Nuclei were stained with DAPI (blue. **A:** Schematic of PLA mechanism of detection for PAR4-GIPC1 interaction. **B:** Confocal image of cells transfected with both PAR4-HA and GIPC1-YFP. Interactions between the two proteins were detected by PLA and are visible as red spots. **C:** Negative control, transfected with PAR4-HA, but not GIPC1, does not display any red dot.

5.4. DISCUSSION:

Proteomics is a rapidly evolving field that has the potential to revolutionize the understanding of cellular processes and disease pathogenesis. However, the large-scale nature of proteomic studies can often generate a significant amount of data, which can be challenging to validate and interpret. Therefore, the validation of proteomic results is crucial to ensure the reliability and reproducibility of the findings. In this chapter, we applied confocal and super resolution microscopies (STED and SIM) to validate the hypothesis that PAR4 could interact with mitochondrial proteins and localise to these organelles.

5.4.1. YFP constructs colocalize with MitoTracker:

The proteomics study revealed interactions with mitochondrial proteins, and Western blot analysis indicated the presence of the transfected receptor within these organelles. However, the existence of mitochondrial PAR4 (mito-PAR4) remains to be conclusively confirmed. Imaging using the Leica SP8 microscope demonstrated that both PAR4-YFP and PAR4ΔSLiM-YFP constructs were localized within mitochondria. Despite optimizing the protocol for MitoTracker Deep Red staining, the dye exhibited a high degree of non-specificity, resulting in fluorescence throughout the entire cell rather than being restricted to the mitochondria. This non-specific staining effect has been previously reported in the literature, where MitoTracker Deep Red was observed to label non-mitochondrial structures in HeLa cells[486]. The high fluorescence intensity impeded the ability to determine whether colocalization was attributable to the receptor or the fluorescent tag, as YFP alone also showed a high degree of colocalization with the mitochondrial dye. This observation stands in contrast to the findings from the Western blot analysis, where PAR4-YFP and PAR4ΔSLiM-YFP were present in mitochondrial fractions, whereas YFP alone appeared to be confined to the cytosolic fraction.

The discrepancy between imaging and Western blotting results highlights the complexity and potential pitfalls in interpreting colocalization studies using different biomolecular tools. The high degree of non-specific staining in fluorescence imaging suggests that the observed colocalization might be an artifact of the experimental conditions rather than a true representation of PAR4 localization. Therefore, it is crucial to approach these findings with caution and acknowledge the need for further validation through additional experimental approaches. Future studies should aim to refine the imaging protocols and employ complementary techniques, such as immuno-electron microscopy to provide more definitive evidence of PAR4's mitochondrial localization. Additionally, alternative fluorescent tags with reduced non-specific binding should be considered to improve the accuracy of colocalization assessments.

5.4.2. Application of Super Resolution Microscopies:

By looking at images of the MitoTracker dye alone, it appears clear how the STED microscope can achieve a much better resolution when compared to the Leica SP8. Interestingly, even the confocal microscope of the Picoquant MicroTime 200 can image specimens with a better resolution than the Leica SP8. When comparing confocal images taken on the Picoquant MicroTime 200 with images taken with the depletion laser on, the resolution enhancement is minimal, even after gating the signal.

As previously mentioned, YFP tags cannot be used on the STED microscope as their excitation lies too far away from the depletion laser. However, even the mCherry tag (which has an excitation profile much closer to the wavelength of the STED laser) was not imaged. This is because the excitation laser had to be set to the full power to obtain only an extremely faint signal. This is not feasible since exposing cells to such a high laser intensity for prolonged periods rapidly caused photobleaching and cytotoxicity. This raises the question to whether it is possible to perform dual colour imaging using only one STED laser. This same problem has been tackled in a previous publication by Tønnesen et al.[487]. The authors used two different fluorescent dyes, one emitting at a shorter wavelength (cyan) and the other emitting at a longer wavelength (red). They then used a single laser source to excite both dyes simultaneously. The excitation laser was tuned to a wavelength that could excite both dyes, and the emission signals were separated using a dichroic mirror. The excitation laser pulse was followed by a single STED pulse that was tuned to deplete only one of the dyes, leaving the other dye in an excited state. By tuning the STED pulse to different wavelengths, the authors were able to selectively deplete either the cyan or the red dye. The remaining fluorescence signal from the undepleted dye was then detected and imaged. This innovative technique could be tested on our microscope, however different dyes may have to be used.

Another avenue that could be exploited would be changing the filters that are currently being used for STED microscopy. In microscopy, band pass filters and long pass filters are two different types of optical filters that are used to selectively transmit certain wavelengths of light while blocking others. A band pass filter allows light to pass through only within a specific wavelength range, while blocking light outside that range. They are commonly used in fluorescence microscopy to isolate the fluorescence emission of a specific fluorophore from other unwanted background signals[488]. By selecting an appropriate band pass filter, the emitted light from the fluorophore can be selectively passed through while blocking the excitation light and other undesired emissions. A long pass filter, on the other hand, transmits light that has longer wavelengths than a certain cut-off value, while blocking shorter wavelengths. This type of filter is often used to remove unwanted shorter wavelength light or to isolate longer wavelength light for imaging. Long pass filters can be used to remove scattered or reflected light, which can interfere with the imaging of the sample, but they can also be used to isolate the emission from a fluorophore with a long emission wavelength, such as red or far-red fluorophores.

When the images were taken, the instrument had installed band-pass filters, which allowed only light with wavelengths between 655 nm and 725 nm to pass through. This narrow window caused mCherry signal to be extremely faint even when the excitation laser was set to the maximal intensity. On the Picoquant MicroTime 200

microscope, an array of different filters is available, including long pass ones. The most suitable for this experiment would be the 594 nm long pass filter, which should allow to visualise both mCherry and MitoTracker Deep Red on different detectors. On this note, the microscope's detector 2 was not always working properly, which also hampered the efforts of trying to utilise the STED microscope.

Super-resolution microscopy techniques have been extensively employed to characterise mitochondrial biology. Many groups applied both STED and SIM to study these organelles. Singh *et al.* were among the first one to employ STED to visualise and quantify mitochondrial protein clusters from murine hearts, achieving a resolution of ~30 nm, they managed to image clusters of cytochrome C oxidase subunit 2 and voltage dependent anion channel[489]. The following year, another group employed STED to study protein clusters in mitochondria, more specifically the mitochondrial inner membrane organizing system (MINOS), which is a conserved large hetero-oligomeric protein complex in the mitochondrial inner membrane, crucial for the maintenance of cristae morphology. Super-resolution microscopy revealed that MINOS is more abundant in mitochondria around the nucleus than in peripheral mitochondria. Moreover, at the submitochondrial level, core MINOS subunits are preferentially localized at cristae junctions[490]. Through the years, methods have been developed to allow mitochondrial visualisation also in live cells, such as combining STED with tetramethylrhodamine methyl ester (TMRM)[491]. These advancements consequently led to imaging submitochondrial structures such as cristae in live HeLa cells. Relying on a cell line stably expressing the mitochondrial protein COX8A fused to the SNAP-tag, STED unveiled lamellar cristae arranged in groups separated by voids that were generally occupied by mitochondrial nucleoids[492]. More recently, another enhanced squaraine variant dye, with low saturation intensity and high photostability which allowed for long-term, high-resolution STED nanoscopy. This novel dye allowed to image the forms of the cristae during mitochondrial fusion and fission over time[493]. Recently, a novel probe characterized by exceptional photostability, fluorogenicity within lipid membranes, and low saturation power was developed. This tool allowed for the visualisation of inner membrane mitochondrial dynamics with a resolution of 40 nm, more specifically dual colour imaging revealed that mitochondrial DNA mtDNA tends to habitat at mitochondrial tips or branch points, exhibiting an overall spatially uniform distribution. However, in conditions of apoptosis and ferroptosis where the cristae structure is compromised, mtDNA distribution becomes irregular[494].

Through the years SIM has also been extensively employed to investigate mitochondrial biology. Opstad *et al.* were among the first ones to combine multi-colour imaging with SIM to image sub-mitochondrial structures, developing a protocol and thus paving the way for further research[495]. A great focus for many groups has been the employment of SIM to study interactions of mitochondria with other vesicles and organelles, such as lysosomes and lipid droplets. Initially, quantitative parametrization of interactions between mitochondria and lysosomes under super-resolution optical microscopy was unavailable, therefore an M-value to quantitatively investigate mitochondria and lysosome contact mitophagy was introduced[496]. SIM proved effective in capturing the production of mitochondria-derived vesicles (MDVs) over time in living or fixed H9c2 cardiomyoblasts, showing correlative imaging of lysosomes[497,498]. More recently, SIM was also used to image interactions between mitochondria and lipid droplets in hepatocytes[499].

5.4.3. PLA confirms PAR4-GIPC1 interaction:

This study provided a novel approach to investigating PAR4 interactions by employing PLA to visualize the binding of GIPC1 to the receptor, marking the first time this technique has been applied to study these interactions. This innovation not only demonstrates the feasibility of using PLA to confirm molecular interactions *in situ* but also offers a valuable tool for probing the spatial and functional relationships of PAR4 with its interactors. The identification of GIPC1 as a binding partner has significant implications for PAR4 biology, potentially linking the receptor's signalling pathways to mitochondrial function. These findings open new avenues for further validation, emphasizing the need to explore the broader functional roles of PAR4-GIPC1 interactions.

An intriguing connection emerges between GIPC1 binding and mitochondrial biology. Ramonett et al. demonstrated that GIPC proteins recognize the atypical PDZ-binding domain of Drp1, mediating its actin-based retrograde transport toward perinuclear mitochondria to promote mitochondrial fission[337]. Similarly, PAR4 contains an atypical PDZ-binding motif, suggesting a possible parallel mechanism. It could be hypothesized that GIPC1 facilitates the localization of PAR4 to mitochondria, potentially influencing mitochondrial dynamics or signalling. However, this hypothesis remains to be validated experimentally.

Future studies could address this by employing complementary approaches. High-resolution microscopy, such as Structured Illumination Microscopy (SIM) or Stimulated Emission Depletion (STED), could directly visualize PAR4 and GIPC1 localization relative to mitochondria. Proximity-dependent biotinylation methods, such as BioID[500] or APEX[501], could provide insights into the mitochondrial protein microenvironment surrounding PAR4. Functional assays exploring mitochondrial dynamics in cells co-expressing PAR4 and GIPC1 could further delineate their roles. Together, these methods would provide a robust framework to validate the hypothesis and uncover the potential interplay between PAR4 and mitochondrial biology mediated by GIPC1.

5.4.4. Use of DNA-PAINT for mitochondrial imaging:

To address the issues that faced with super resolution imaging of PAR4, another avenue would be to employ DNA-Point Accumulation for Imaging in Nanoscale Topography (DNA-PAINT). This is a super-resolution imaging technique that utilizes reversible binding of short DNA strands to achieve a resolution beyond the diffraction limit of light microscopy. In DNA-PAINT, two types of DNA strands are used: the target strands and the imaging strands. The protein of interest is targeted using biotin-labelled antibodies, which are then conjugated with avidin. A DNA docking strand is also labelled with biotin which binds avidin as well. The cognate imaging strand is tagged with a fluorophore and binds the docking strand, resulting in the accumulation of fluorescent signal. By precisely controlling the binding and unbinding kinetics of the imaging strands, high-resolution images can be obtained with sub-10 nm spatial resolution. Additionally, it can be combined with other imaging techniques such as fluorescent labelling and STED to provide information about the sample's structure at even higher resolution. Schnitzbauer et al.[502] outlined a protocol for DNA-PAINT where the imaging strands are introduced to the sample in a buffer containing an oxygen scavenger and an imaging buffer. The imaging strands bind transiently to their complementary target molecules, resulting in the accumulation of fluorescent signals. The binding and unbinding kinetics of the imaging strands are controlled by adjusting the concentration of the imaging buffer and the temperature of the sample. Moreover, Jungmann et al.[503] managed to perform multicolour imaging using exchange DNA-PAINT. They demonstrated the multiplexed imaging approach *in vitro* on synthetic DNA structures, achieving up to ten-colour super-resolution imaging with sub-10-nm resolution. Using this method, they successfully imaged up to four colours in two-dimensional imaging and up to three colours in three-dimensional imaging of proteins. The Exchange-PAINT approach enabled sequential imaging of multiple targets with a single dye and laser source by using a "replenishment" step to replace the imaging strands and avoid signal degradation. In the paper by Unterauer et al.[504], the authors first optimized the DNA-PAINT protocol for imaging the synaptic protein Synapsin I, which is involved in neurotransmitter release. They used a combination of antibodies and DNA-conjugated oligonucleotides to label Synapsin I with a fluorophore and immobilize it on the coverslip surface. They then used a complementary DNA strand labelled with another fluorophore to achieve super-resolution imaging with a resolution of ~13 nm. Next, the authors used DNA-PAINT to study the distribution of Synapsin I in cultured hippocampal neurons. They found that Synapsin I puncta were distributed in a clustered pattern, and the density of clusters increased with the degree of synaptic maturation. The authors also used DNA-PAINT to study the interaction between Synapsin I and the PDZ protein PSD-95. They found that these two proteins were often co-localized in clusters, suggesting that they may interact with each other. Recently, DNA-PAINT has been used to quantitatively analyse GPCR oligomerisation. By employing qPAINT, the purinergic receptor $P2Y_2$ was found to oligomerise in cancer AsPC-1 cells. Antagonistic treatment reduced the percentage of oligomers, while agonists did not affect the number of complexes[505]. The same group also demonstrated that $P2Y_2$ interacts with αV-integrins, and using DNA-PAINT they found that $P2Y_2$ regulates the amount and distribution of integrins at the plasma membrane. Receptor-integrin interactions were required for effective signalling, leading to cancer cell evasion. Overall, this demonstrates what a powerful technique DNA-PAINT can be, and how it has been applied to study GPCR pharmacology.

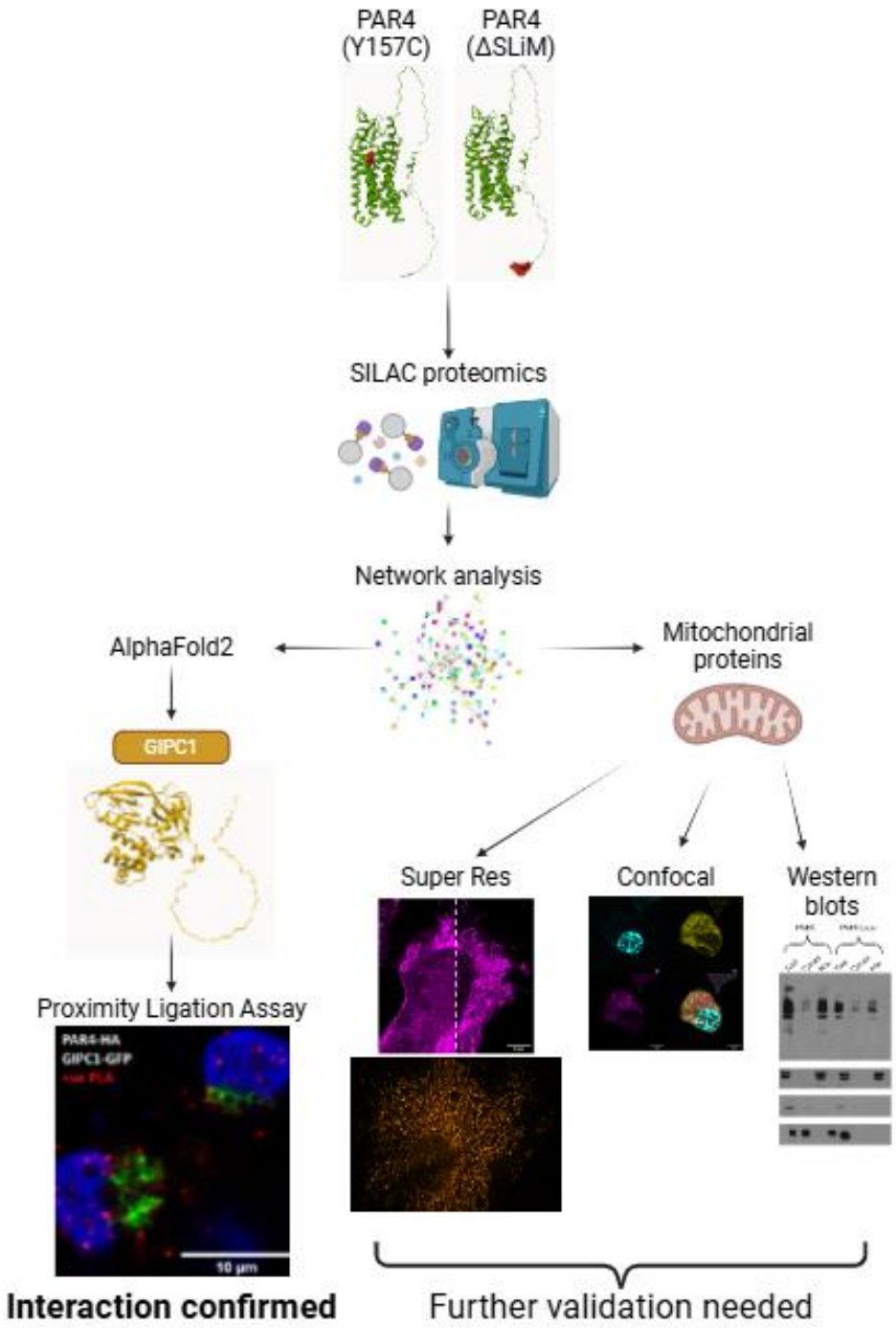


Figure 5.16 – Findings of Chapter 5. Proximity ligation assay confirmed interactions between PAR4 and GIPC1, however different methods provided contradictory results regarding the mitochondrial localisation of the receptor. Further validation is thus needed to ascertain the existence of mito-PAR4.

6. GENERAL DISCUSSION

6.1. Overview of findings:

The initial aim of this thesis was to explore the interactome of protease-activated receptor 4 (PAR4), beginning with SILAC proteomics as the foundational technique. The study focused on the wild-type PAR4 proteome along with two structural modifications. The first modification involved deleting the last five residues ($S^{381}SLLQ^{385}$) of the receptor, a sequence resembling a PDZ binding motif (SLiM). This was intended to determine if PAR4 binds to PDZ domain-containing proteins and to assess the impact of removing this SLiM on such interactions. Additionally, the investigation encompassed a single nucleotide polymorphism (SNP) resulting in the Y157C mutant receptor. Previous research identified that this mutation affects receptor trafficking, retaining PAR4 partially in the endoplasmic reticulum and leading to diminished surface expression and platelet functionality[193]. The effects of this mutation on downstream signalling were unclear, thus necessitating the SILAC experiments conducted. The findings indicated that the receptor lacking the SLiM did not interact with any PDZ proteins, whereas the wild-type PAR4 bound notably to TJP2 and LMO7. Furthermore, the Y157C mutant consistently highlighted interactions with GIPC1 and an upregulation of PDLIM1, another PDZ protein, suggesting a mutation-induced alteration in interaction patterns.

Given the range of potential interactors identified, AlphaFold Multimer was employed—an artificial intelligence tool designed for predicting protein folding and binding[506]. This tool identified GIPC1 as the most plausible interaction partner, which was subsequently confirmed through proximity ligation assays, confirming interactions with this scaffold protein.

Initial proteomic analyses identified clusters of mitochondrial proteins in all four datasets, prompting an investigation into the intracellular localization of PAR4. Western blot analyses of mitochondrial isolations confirmed the presence of PAR4 in these fractions, supporting the hypothesis of the existence of the receptor in these organelles. Although attempts to obtain definitive imaging of mitochondrial PAR4 using super-resolution microscopy were inconclusive, confocal microscopy provided evidence of colocalization between PAR4 and mitochondria. This suggests that the receptor may indeed localize within these organelles, although further studies are needed to clarify this observation.

This research is the first to report interactions between PAR4 and the GIPC proteins, as well as the mitochondrial localization of PAR4. These findings, however, may reflect overexpression artifacts associated with the highly efficient transfection methods used in HEK293 cells, a factor that should be considered when interpreting these results.

6.2. Different approaches to metabolic labelling proteomics:

Stable Isotope Labelling by Amino acids in Cell culture (SILAC) is a robust and widely used quantitative proteomics method, distinguished by its accuracy in measuring differential protein expression across samples[207]. A key strength of SILAC lies in its simplicity and reliability, allowing for direct incorporation of labelled amino acids into proteins during cell culture, which facilitates precise quantitation through mass spectrometry. However, SILAC has limitations, including its restriction to cell culture systems, which makes it unsuitable for use with tissues or organisms that cannot be cultured, such as platelets. Additionally, SILAC can be costly and time-consuming due to the need for extensive labelling periods and the requirement of special culture conditions.

To address these drawbacks, recent advancements have led to the development of novel methods. One of these is the Culture-Derived Isotope Tags (CDITs) method, which builds on the SILAC technique[507]. In this case a labelled internal standard is created for proteome quantification of specific tissues. This standard involves cultivating cells from the target tissue in a medium enriched with stable isotopes, similar to SILAC procedures. For instance, cells from mouse neuroblastoma are integrated with each brain sample from mice. For each sample, the isotopic peptide distribution ratio (tissue versus cultured cells) is calculated, and the relative protein abundance changes between two brain samples are determined by comparing these ratios. Notably, proteins present in a brain sample but absent in cultured cells are also quantifiable by comparing the target peptide's peak ratio to a differently sequenced, isotopically labelled peptide from the standard cells that shares the same LC/MS retention time.

The super-SILAC method enhances this approach by adding a mix of lysates from various SILAC-labelled cell lines as an internal standard to the proteomic samples under analysis. This advancement facilitates the characterization and comparison of cell lines, particularly in Cancer Researchearch by providing a broader dynamic range of quantitation[508]. This adaptation enhances the versatility and practicality of SILAC, making it more comparable to other cutting-edge proteomic techniques like TMT (Tandem Mass Tag) labelling and label-free quantitation, which offer advantages in scalability and ease of use in diverse experimental setups[509]. Additionally, SILAC has been adapted for *in vivo* studies in animals, termed Stable Isotope Labelling of Mammals (SILAM)[510]. In SILAM, animals are fed a diet exclusively composed of stable isotope-enriched proteins, ensuring comprehensive metabolic labelling of proteins across all cells. Despite the extensive time and cost due to the need for multiple generations, the labelled animals remain healthy and phenotypically consistent with their unlabelled counterparts. These metabolic labelling strategies also enable the identification and quantification of post-translational modifications (PTMs), such as methylation[511]. In the heavy methyl SILAC method, cells are cultured in media containing [13CD3]-methionine, which, following several cell divisions, labels all *in vivo* methylation sites.

Initially, it was observed that traditional data analysis methods applied to SILAC proteomics often resulted in the identification of a limited number of protein interactions, making impossible to perform cluster analyses of the proteomes. To enhance the detection and analysis of more subtle but potentially significant interactions, a novel data analysis pipeline was developed. This new approach was

designed to improve sensitivity and specificity, incorporating advanced statistical models and data processing techniques to discern genuine interactions more effectively from background noise. Rigorous validation of this pipeline demonstrated its ability to capture the interactome of PAR4 more comprehensively, as shown by the *in vitro* confirmations of GIPCs interactions, as well as the colocalization of PAR4 in mitochondria. This novel data analysis pipeline therefore expands the potential of employing SILAC proteomics for novel biological insights without compromising data integrity.

6.3. AI to confirm interactions:

Proteins exert their functions through the formation of specific three-dimensional structures or interactions, yet only a minor fraction of these characteristics within a proteome are determined through experimental methods. The use of artificial intelligence to accurately model the structures of proteins and their complexes is transforming molecular biology. Experimental data facilitate a candidate-based approach for systematically modelling novel protein assemblies. Therefore, this project integrated AlphaFold as a screening tool to discriminate which interactions to validate *in vitro*. This approach held multiple benefits including huge savings in terms of time and costs; as well as scalability of the approach, since multiple proteins could be studied at the same time. However, this AI tool is still far from perfect and can struggle to predict certain interactions, especially in the case of large proteins, or in the presence of intrinsically disordered regions. These issues regarding AlphaFold accuracy warrant caution when interpreting its results, and validation is still always needed. A great advancement in the use of AlphaFold was the development of AlphaFold Multimer[506], and the discovery that using shorter sequences, investigating peptide-domain interactions, rather than full protein-protein interactions, greatly enhances the accuracy of the predictions and the consistency of the models[296]. This experiment, in the case of GIPCs, confirms just that. Initially only one model per protein would find meaningful interactions between PAR4 SLiM and the binding site on the GIPC PDZ domain. However, using just these regions of the two proteins achieved much greater consistency across models, and significantly increased every prediction metric, especially the ipTM, which measures the quality of the interface between the two predicted structures.

This similar approach of using shorter sequences was proposed also for the multiple sequence alignments that form the basis of AlphaFold prediction algorithms. The evolutionary constraints imposed by the binding interfaces of interacting proteins are effectively used to predict PPIs from multiple sequence alignments (MSAs). Constructing MSAs requires crucial decisions regarding the reliable identification of orthologs and achieving a balance between the size of alignments and their quality. Recently, Fang and colleagues developed an innovative approach to MSA construction: rather than aligning full sequences, multiple distinct alignments for each protein are created within different branches of the tree of life. Coevolutionary signals are initially explored within these groups and subsequently integrated using machine learning techniques, significantly enhancing prediction performance and alignment quality. Due to the novelty of the finding, this MSA strategy has not been implemented yet in the AlphaFold algorithms, however it is easy to imagine how this approach could offer a rapid and precise technique for genome-wide pre-screening, directing only promising interaction candidates to the final prediction, thereby minimizing false

positives and computational demands[512]. Other groups have begun using AlphaFold as a guide for mapping protein-protein interactions. Most notably, Bartolec and Vàzquez-Campos deployed this tool massively to confirm and evaluate protein structures and interactions identified through cross-linking mass spectrometry, covering a significant portion of the human proteome. This serves as an empirically validated tool, which enables the experimental mapping of spatially constrained pairs of amino acids within proteins or across interaction interfaces. These cross-links are significant as they encompass proteins with native sequences, posttranslational modifications, subcellular niches, and cofactors. The utility of this resource is demonstrated in the mapping, assessment, and contextualization of the recently expanded structural proteome by the breAKThrough of AlphaFold[513]. A similar approach was utilised by another group, who employed a combination of in-cell crosslinking mass spectrometry and co-fractionation mass spectrometry (CoFrac-MS) to detect protein–protein interactions in Gram-positive bacterium *Bacillus subtilis*. Crosslinking interactions prior to cell lysis preserved protein interactions, which are often disrupted upon cell lysis. The structures of these interactions, along with others in the SubtiWiki database, were predicted using AlphaFold-Multimer. After addressing the false-positive rate of these predictions, novel structural models for 153 dimeric and 14 trimeric protein assemblies were proposed. Once again, crosslinking MS data independently confirms the validity of the AlphaFold predictions and scoring[514]. Employment of machine learning methods is improving the drug discovery pipeline, as demonstrated by Trepte et. al, who formulated a methodology to prioritize interactions by analysing quantitative data from binary PPI assays or predictions from AlphaFold-Multimer. This pipeline significantly aids in prioritizing PPI targets, thereby accelerating the discovery of early-stage drug candidates that target protein complexes and pathways. By employing the quantitative assay LuTHy alongside this machine learning algorithm, they identified high-confidence interactions among SARS-CoV-2 proteins, and their three-dimensional structures have been predicted using AlphaFold-Multimer. Thus, this process led to the identification of a compound that disrupts SARS-CoV-2 replication[515]. Another AI tool employed in this project to study PAR4 interactions was the webserver Predicting Protein-Protein Interactions (PrePPI), which predicts PPIs on a proteome-wide scale. PrePPI utilizes both structural and non-structural evidence within a Bayesian framework to compute a likelihood ratio (LR) for virtually every possible pair of proteins within a proteome; the current database focuses on the human interactome. The structural modelling (SM) component, derived from template-based modelling, is supported by a unique scoring function for evaluating assumed complexes. The enhanced version of PrePPI incorporates AlphaFold structures, divided into individual domains. To date, the database contains approximately 1.3 million human PPIs, and it displays multiple functionalities such as examining query proteins, template complexes, 3D models of predicted complexes, and related features[384].

This AI tool was therefore tested by loading PAR4 UniProt identifier on the webserver. This yielded 189 predicted interactions, but only 5 of those were PDZ proteins. These interactions were computationally validated on AlphaFold using only the PDZ domains and the SLiM motif of PAR4. Unfortunately, none of these proteins achieved an acceptable consistency or quality of the prediction metrics, suggesting that the interactions given by PrePPI are not highly reliable yet.

6.4. AlphaFold 3:

Artificial intelligence is an extremely rapidly evolving field. This holds true especially for tools applied to biology as novel tools are constantly released. As previously mentioned, one of the most important breAKThroughs in the field was the release of AlphaFold in 2020, which completely reshaped the landscape of structural biology, allowing to resolve protein structures in hours. During the course of this project, protein-protein interactions were investigated using ColabFold, which runs on algorithms similar to AlphaFold2. In May 2024 a novel version of this tool was released, AlphaFold3[294].

This improved version shows a substantially updated diffusion-based architecture, thus allowing the structure prediction not just of protein complexes, but also nucleic acids, small molecules, ions and modified residues. AlphaFold3 greatly outperformed AlphaFold-Multimer v2.3. The improvements are due to the different MSA algorithms, which replaced Evoformer with the simpler Pairformer module, which has a smaller and simpler MSA embedding block thus improving the speed and computational expenses. Moreover, the novel system uses a Diffusion Module to predict the exact positions of atoms, instead of the AlphaFold 2 Structure Module which focused on specific amino acid structures and the angles of side chains. This diffusion process works at multiple levels; at lower noise levels, it helps refine the details of local structures. This approach also reduces the need for dealing with complex bonding patterns and chemical properties in the model, making it easier to handle different chemical components. Importantly, the presence of a diffusion module technically qualifies AlphaFold 3 as a generative machine learning model. This is because it creates new data similar to the training it had learnt from, rather than identifying patterns in existing data, like AlphaFold 2. On a practical level, besides being more reliable, AlphaFold3 predictions are substantially faster, requiring minutes as opposed to AlphaFold2 which took hours and was often left running overnight.

A major advancement from the previous version is the possibility of integrate post-translational modifications (pTMs) and ligands from a preset list. One of such ligands is oleic acids, and it was found that by providing as an input over 50 of these ligands, the AI recognises them as a membrane-like structure and correctly recognises intracellular and extracellular regions. AlphaFold3 still shows some drawbacks compared to its previous version. One of the major ones is the inability to resolve intrinsically disordered structures, just like AlphaFold2. However, the newer version “forces” these regions to fold in ribbon and α-helixes structures, even when these are simple strands of amino acids. To highlight that these are disordered regions, the AI colours them with very low pLDDT levels. This is clearly shown in supplementary figure S9 where PAR4 and GIPC2 full sequence structures were predicted together with 100 oleic acids. Notably the AI predicted the membrane-like structure and the disordered regions were folded as helixes, but with low pLDDT scores.

Despite being released very recently and outperforming previous tools, the quality of the predictions of AlphaFold 3 has already been put to the test in various preprints. One study examined AlphaFold 3's ability to predict metal-protein interactions and compared its performance to RoseTTAfold-AllAtom, concluding that AlphaFold 3 provides realistic predictions for metal ions[516]. Another study focused on human T cell receptors and their recognition of NRAS cancer neoantigens. While AlphaFold 3 showed strong performance, it didn’t quite match the accuracy of a specialised

AlphaFold 2 version called TCRmodel2[517]. A third study explored how well deep learning systems capture the physics of protein-ligand interactions. They discovered significant divergence from expected physical behaviours, as AlphaFold 3 placed small molecules like ATP into their natural binding sites even when the binding residues had been mutated. The authors concluded that AlphaFold 3 predicts small molecule binding based on patterns in distant protein regions or overall structure, rather than molecular interactions[518]. AlphaFold 3's performance was also assessed in predicting binding energy landscapes using the Structural Kinetic and Energetic database of Mutant Protein (SKEMPI) dataset, a commonly used binding energy dataset. The authors noted that AlphaFold 3 learns unique features beneficial for estimating binding free energy and can improve initial predictions made by other methods[519]. In the context of molecular docking, the structural prediction of the type 2 taste receptor (TAS2R) was compared with Cryo-EM structures bound to the single-chain variable fragment scFv16[520]. AF3 incorrectly placed the fragment and displayed a 2.0 Å r.m.s.d. This underscores how, despite great and constant improvements, computational investigations using AI tools still need to be coupled with structural studies and empirical validations. For this reason, although AlphaFold 2 suggested GIPC1 as a potential interactor of PAR4, experimental proof was still needed to confirm the interaction.

6.5. Potential significance of PAR4-GIPC interactions:

The initial goal of this study was to investigate whether PAR4 would bind to any PDZ protein. This idea emerged from the observation that the C-terminal tail of the receptor ($S^{381}SLLQ^{385}$) resembles a class I PDZ binding motif, known as a Short Linear Motif (SLiM). The class I PDZ domain consensus sequence is typically S/T-X-Φ, where S/T is serine or threonine, X is any amino acid, and Φ is a hydrophobic residue such as leucine, isoleucine, valine, or methionine. In the PAR4 sequence, the first serine (S^{381}) meets the serine/threonine requirement, the second serine serves as the variable amino acid, and the two leucine residues fit well into the hydrophobic slot required by class I PDZ domains. Although the glutamine (Q^{385}) at the C-terminus is not a typical hydrophobic residue, we initially hypothesized that it would prevent the receptor's C-tail from interacting with PDZ domains. However, AlphaFold predictions indicated that this glutamine forms key hydrogen bonds that stabilize the SLiM within the binding cleft of GAIP-interacting protein, C-terminus (GIPC) proteins. GIPC proteins are known to bind sequences that fit the S/T-X-Φ pattern, confirming their classification as class I PDZ proteins. Given the PAR4 sequence $S^{381}SLLQ^{385}$, the PDZ domain of GIPC1 is likely to recognize the hydrophobic leucine residues, despite the less typical presence of glutamine at the C-terminus. The flexibility in binding specificity allows for such variations. Notably, using just the PAR4 SLiM and GIPC PDZ domain, rather than the full protein sequences, significantly improved the accuracy and consistency of predictions across AlphaFold models. This consistency demonstrated that Q^{385} does not hinder the binding of the two proteins; instead, it is essential for forming crucial hydrogen bonds with L^{135}, a hydrophobic residue deep within the carboxylate binding groove of GIPC1. This same interaction was observed with L^{129} in GIPC2 and L^{124} in GIPC3, highlighting not only consistency across models for a single protein but also across different proteins within the same family.

The binding between PAR4 and GIPC1 was confirmed *in vitro* by proximity ligation assay (PLA). This raised multiple questions concerning the functional context of this

interaction, such as potential mechanism of sorting the receptor. For example, PAR4 might be transported to mitochondria by GIPCs, which have already been documented to play a role in mitochondrial transport. Ramonett et al. found how GIPC1 mediates the retrograde transport of Drp1 to enhance mitochondrial fission. Interestingly, the interaction between Drp1 and GIPC also happens through its atypical C-terminal PDZ-binding motif[337]. Drp1 extreme C-terminal residues are conserved across species and consist of THLW, which partially matches the consensus PDZ-binding motif sequence ending with a hydrophobic residue. Just like PAR4, this motif is unique and distinct from the typical PDZ-binding motifs found in other proteins, underscoring how despite belonging to class 1 PDZ proteins, member of the GIPC family are capable of recognising atypical SLiMs. Another study extensively investigated the role of Drp1 SLiM in the context of GIPC1 binding and mitochondrial fission. They found how the SLiM in Drp1 is a critical regulatory site that governs the protein's structure and function. These residues play a pivotal role in controlling Drp1's conformational dynamics, oligomerization propensity, self-assembly geometry, and cooperative GTPase activity. This SLiM also influences Drp1's ability to remodel target membranes, essential for its role in membrane fission. The SLiM functions as a spacer, setting the register and geometry of inter-subunit interactions during nucleotide-dependent helical self-assembly. It also acts as an auto-inhibitory motif that restricts high membrane curvature generation, ensuring controlled and partner protein-regulated membrane fission. When Drp1 interacts with GIPC1 via the SLiM, it results in controlled disassembly of the Drp1 scaffold, crucial for effective membrane constriction and fission. GIPC1 binding disrupts the helical propagation of Drp1, affecting its self-assembly into rings and spirals and thus regulating the disassembly necessary for high membrane curvature generation. The interaction between Drp1 and GIPC1 via the SLiM steers and directs Drp1's inter-subunit spacing, oligomerization geometry, nucleotide-sensitive conformational rearrangements, and assembly-disassembly dynamics[521]. Recent evidence found that the C-terminus of the prototypical M2 muscarinic receptor localizes to the mitochondria and regulates cell respiration under stress conditions[522]. Interestingly, this event happened independently of interactions with any PDZ or scaffold protein. The expression of the C-terminal domain of a GPCR, which can regulate mitochondrial function, represents a previously unknown mechanism distinct from its traditional signalling role at the plasma membrane.

Recently, another role for GIPC1-mediated protein transport to mitochondria was discovered. In dilated cardiomyopathy (DCM), GIPC1 remodels lipid metabolism by transporting DECR1 into these organelles, thereby protecting cardiac myocytes against ferroptosis. GIPC1 regulates the actin-based transport of DECR1 into mitochondria, which promotes mitochondrial fatty acid β-oxidation (FAO) and reduces the accumulation of polyunsaturated fatty acids (PUFAs). This process ultimately protects cardiac myocytes from ferroptosis in DCM.GIPC1 remodels lipid metabolism by transporting DECR1 into mitochondria against ferroptosis during dilated cardiomyopathy (DCM)[523]. These examples underscore the versatile roles of GIPCs in mitochondrial transport and function, suggesting that PAR4 might similarly be trafficked to mitochondria by GIPC1, potentially influencing cellular metabolism or stress responses.

In addition to GIPCs, another PDZ protein identified in the proteomics study was PDLIM1, which was found to be upregulated in the PAR4 proteome. While PDLIM1

did not achieve the same consistency across models and prediction metrics as GIPC1, particularly in terms of interface predicted template modelling (ipTM) scores, it still represents an intriguing subject for future research. The upregulation of PDLIM1 in the context of PAR4 suggests a potential functional interaction that warrants further investigation, despite the initial predictive models being less robust. Exploring the interaction between PAR4 and PDLIM1 could provide new insights into the broader network of PDZ domain-mediated protein interactions and their roles in various cellular processes. Given the established role of PDLIM1 in cytoskeletal organization and signal transduction, its interaction with PAR4 might uncover novel pathways or mechanisms that influence receptor sorting, trafficking, or signalling. For instance, PDLIM1 is known to play a crucial role in organizing the actin cytoskeleton and facilitating signal transduction pathways. If PDLIM1 interacts with PAR4, it might affect the localization and function of PAR4 within the cell, potentially altering how the receptor responds to external signals or how it is internalized and trafficked within the cell. This interaction could also impact cellular processes such as migration, adhesion, and intracellular signalling cascades that depend on the dynamic reorganization of the cytoskeleton. Furthermore, the potential interaction between PAR4 and PDLIM1 could have implications for understanding disease mechanisms where these proteins are involved. For example, alterations in PAR4 or PDLIM1 expression or function have been linked to various pathologies, including cancer[241] and cardiovascular diseases[524]. Investigating their interaction might reveal new therapeutic targets or strategies for modulating these pathways in disease contexts. Overall, the study of PAR4 and PDLIM1 interactions opens up a promising avenue for research, with the potential to enhance our understanding of PDZ domain-mediated protein networks and their impact on cell physiology and pathology. By delving into these interactions, researchers can uncover novel mechanisms of cellular regulation and identify new points of intervention for therapeutic development.

6.6. Future investigation into mito-PAR4:

Since the gene ontology analysis suggested that most enriched biological process and cellular component terms were revolving around metabolism and mitochondria, we hypothesised that PAR4 could be directly found in these organelles. To verify this hypothesis biochemical and microscopy techniques were employed. Western blot analysis confirmed the presence of the receptor in mitochondria. Using antibody targeting YFP, it was confirmed that the mitochondrial localisation was due to the receptor itself and not to the fluorescent tag. The low number of repetitions and the presence of outliers caused a high standard deviation. If the presence of PAR4 was to be confirmed, this would raise other questions such as what its role in these organelles is and how can it be activated intracellularly. If PAR4 was to be found on the outer membrane of mitochondria, its activation could be dependent on cytosolic proteases, without the need for the activating enzymes to be located within the organelles. However, if the receptor localised in any other mitochondrial compartment, such as the intermembrane space, the inner membrane, cristae or matrix, PAR4 would have to be activated by mitochondrial proteases. Considering the irreversible nature of PAR4 activation, mitochondrial proteases would also be essential in the recycling of the receptors.

Mitochondria are crucial metabolic hubs that adapt dynamically to physiological demands. These organelles contain over 1000 distinct proteins, with more than 40

mitoproteases residing in different compartments, which preserve mitochondrial proteostasis and serve as central regulators of mitochondrial plasticity[525]. This complex system of mitochondrial proteases performs limited and terminal proteolysis to build the mitochondrial proteome, regulate its functions, and eliminate mitochondrial proteins and peptides. Pre-sequence proteases are responsible for cleaving and breaking down mitochondrial targeting signals during protein biogenesis to produce fully functional proteins. Protease processing also regulates mitochondrial functions and quality control enzymes remove misfolded or excess proteins. Various signalling cascades coordinate the activity of mitoproteases to maintain mitochondrial homeostasis and ensure cell survival. Loss of mitoproteases severely compromises the functional integrity of mitochondria, is linked to aging, and causes pleiotropic diseases[526]. Among these proteases, enzymes belonging to the high-temperature requirement (HtrA) family of oligomeric serine proteases are particularly interesting. PAR4 is activated by serine proteases such as trypsin and thrombin, and members of the HtrA family combine a trypsin-like domain with one or two PDZ domains. The PDZ domains act as a protein-protein interaction site that binds HtrAs to their target, thus allowing the enzymatic activity of the trypsin-like domain[527]. HtrA2, also known as Omi, is the most well characterised enzyme. Its biological activity has been shown to be pro-apoptotic[528], whereas loss of function is associated with Parkinson's disease in humans[529]. However, the PDZ binding domain on PAR4 is found at the very end of its C-terminus, therefore on the opposing side of the membrane as the activation site cleaved by serine proteases, thus meaning that HtrA proteins might interact with the receptor in two distinct ways.

Similarly to GIPCs, proximity ligation assay (PLA) can be employed to ascertain PAR4 interactions with VDACs and other mitochondrial proteins. PLA could also be combined with super-resolution microscopies, thus allowing for more detailed and precise visualization of protein-protein interactions. This combination of techniques has been already proposed to investigate mitochondrial dynamics, specifically interactions with the ER[530,531], and has been recently successfully employed in fixed optic nerve sections[532]. Due to the nature of this project being borne out of an interactome study, the combination of these two methods could provide valuable insights into the organization and dynamics of protein complexes, and it has the potential to enhance the understanding of complex cellular processes involving mitochondrial PAR4 signalling and regulation.

6.7. Limitations of the project:

While this project has provided valuable insights into the interactome of PAR4, several limitations must be acknowledged. Transfection of PAR4 and GIPC proteins into HEK293 cells was a fundamental technique used in this study to elucidate protein interactions. However, this method inherently leads to overexpression of the transfected proteins, which can result in artifacts. Overexpression may cause non-physiological interactions that do not occur under normal cellular conditions, thereby potentially skewing the results. This limitation underscores the importance of validating key findings in more physiologically relevant systems or using endogenous levels of protein expression where possible. Future studies should aim to use more native systems or inducible expression systems to mitigate these effects.

The analysis pipeline developed for the SILAC proteomics in this project is novel and tailored to the specific needs of this study. While it provided significant insights and

identified a wide array of interacting proteins, it diverges from canonical methods traditionally used in SILAC proteomics. This deviation may introduce biases or overlook certain interactions typically identified by standard approaches. The novelty of the pipeline, while advantageous in some respects, requires thorough validation against established methods to ensure its reliability and reproducibility. Nevertheless, the identification of interactions with GIPC1 using the novel analysis pipeline, which were subsequently confirmed *in vitro*, testifies to the validity of this new method. Comparative studies using canonical pipelines could help in benchmarking the new approach and in identifying any potential shortcomings.

The project heavily relied on artificial intelligence tools like AlphaFold and PrePPI to predict protein-protein interactions. These tools represent a significant advancement in computational biology, providing predictions that were previously unattainable. However, their predictions still require rigorous experimental validation. In this study, although many interactions were predicted, only the interaction with GIPC1 was partially validated *in vitro*. This highlights a critical limitation: computational predictions are only as useful as their experimental corroboration. The inability to validate many predicted interactions can be attributed to the inherent challenges and variability of experimental biology. Future work should focus on systematically validating these predictions to strengthen the confidence in AI-derived interactions.

The field of AI is rapidly evolving, with new models and algorithms being released frequently. Very recently, AlphaFold 3 was released, showing differences in predicted interactions compared to AlphaFold 2, which was used in this study. This constant evolution presents both an opportunity and a challenge. On one hand, newer models may provide more accurate predictions; on the other hand, they may also lead to inconsistencies in data interpretation if different versions are used. This highlights the need for a standardized approach to AI tool usage in research and the importance of staying updated with the latest advancements. Retrospective validation of findings with updated models can help ensure the robustness of the conclusions drawn.

The investigation of mitochondrial PAR4, despite yielding valuable insights, is subject to several limitations that warrant careful consideration. Firstly, the high degree of non-specific fluorescence observed with MitoTracker Deep Red staining significantly impeded the ability to accurately assess the colocalization of PAR4 constructs with mitochondria. This issue was compounded by the strong fluorescence signals from the YFP tag, which further obscured the distinction between genuine receptor localization and background staining. While Western blot analysis suggested the presence of PAR4-YFP and PAR4ΔSLiM-YFP in mitochondrial fractions, this method cannot definitively confirm mitochondrial localization without complementary imaging techniques. The discrepancy between imaging and Western blotting results underscores the need for more robust validation methods. The relatively low transfection efficiency observed with mCherry constructs limited the sample size and may have affected the reliability of the colocalization analysis. Finally, the absence of ΔSLiM-mCherry constructs in the study prevented a comprehensive comparison across all constructs. These limitations highlight the necessity for further optimization of staining protocols, the use of alternative fluorescent tags, and the application of additional validation techniques to ensure the accuracy and reliability of the findings regarding PAR4's mitochondrial localization.

6.8. Concluding remarks and future outlook:

This thesis has provided a comprehensive exploration of protease-activated receptor 4 (PAR4), significantly advancing the understanding of its intracellular interactions and functional roles. By employing a combination of SILAC proteomics, advanced AI tools, and high-resolution microscopy, the study has unveiled a complex network of protein interactions, particularly highlighting the roles of mitochondrial and PDZ domain-containing proteins such as GIPCs. The novel analysis pipeline developed for SILAC proteomics has demonstrated its efficacy, particularly through the identification and *in vitro* validation of the PAR4-GIPC1 interaction, underscoring the potential of innovative analytical methods in proteomics. Despite inherent challenges such as overexpression artifacts and the evolving nature of AI tools, the integration of computational predictions with experimental validation has proven to be a powerful approach in elucidating protein interaction networks.

Looking forward, the findings of this thesis open several promising avenues for future research. Investigating the interactions of PAR4 in more physiologically relevant models, such as primary cells or animal models, will be crucial to validate and extend these results. Further exploration of PAR4's potential roles in mitochondrial functions could provide deeper insights into its involvement in cellular metabolism and signalling pathways. Another PDZ protein flagged in the proteomics experiments, PDLIM1, was found to be upregulated in the PAR4 group of the Y157C experiment. Although PDLIM1 did not achieve the same consistency as GIPC1 in the AlphaFold predictions, it remains a promising candidate for future study. Investigating the interaction between PAR4 and PDLIM1 could reveal additional aspects of PAR4's functional landscape. As AI tools continue to evolve, revisiting these interactions with newer models could refine predictions and uncover novel insights. Overall, this work not only enriches the current understanding of PAR4 but also exemplifies the synergy between cutting-edge technologies and traditional experimental techniques, paving the way for future research in the dynamic field of proteomics and cell signalling. Through the careful consideration of these limitations and the application of emerging methodologies, future studies can build upon this foundation to uncover deeper insights into the biological significance of PAR4 and its interactions.

7. REFERENCES:

1. The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

2. Hankey, G. J. & Eikelboom, J. W. Antiplatelet drugs. *Medical Journal of Australia* **178**, 568–574 (2003).

3. Tiwari, A. & Singh, S. Computational approaches in drug designing. *Bioinformatics: Methods and Applications* 207–217 (2022)

4. Vandendries, E. R., Hamilton, J. R., Coughlin, S. R., Furie, B. & Furie, B. C. Par4 is required for platelet thrombus propagation but not fibrin generation in a mouse model of thrombosis. *Proceedings of the National Academy of Sciences***104**, 288–292 (2006).

5. Green, D. Coagulation cascade. *Hemodialysis International***10 Suppl 2**, (2006).

6. Bode, W. *et al.* The refined 1.9 Å crystal structure of human α-thrombin: Interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO Journal* **8**, 3467–3475 (1989).

7. Esmon, C. T. Thrombomodulin as a model of molecular mechanisms that modulate protease specificity and function at the vessel surface. *The FASEB JOURNALournal* **9**, 946–955 (1995).

8. Huntington, J. A. & Baglin, T. P. Targeting thrombin - Rational drug design from natural mechanisms. *Trends in Pharmacological Sciences***24**, 589–595 (2003).

9. Huntington, J. A. & Esmon, C. T. The Molecular Basis of Thrombin Allostery Revealed by a 1.8 Å Structure of the “Slow” Form. *Structure* **11**, 469–479 (2003).

10. Bock, P. E., Panizzi, P. & Verhamme, I. M. A. Exosites in the substrate specificity of blood coagulation reactions. *Journal of Thrombosis and Haemostasis* **5**, 81–94 (2007).

11. Liu, L. W., Vu, T. K. H., Esmon, C. T. & Coughlin, S. R. The region of the thrombin receptor resembling hirudin binds to thrombin and alters enzyme specificity. *Journal of Biological Chemistry* **266**, 16977–16980 (1991).

12. Verhamme, I. M., Olson, S. T., Tollefsen, D. M. & Bock, P. E. Binding of exosite ligands to human thrombin. Re-evaluation of allosteric linkage between thrombin exosites I and II. *Journal of Biological Chemistry* **277**, 6788–6798 (2002).

13. Stubbs, M. T. & Bode, W. The clot thickens: clues provided by thrombin structure. *Trends in Biochemical Sciences***20**, 23–28 (1995).

14. Xu, H., Bush, L. A., Pineda, A. O., Caccia, S. & Di Cera, E. Thrombomodulin Changes the Molecular Surface of Interaction and the Rate of Complex Formation between Thrombin and Protein C *. *Journal of Biological Chemistry* **280**, 7956–7961 (2005).

15. Fuentes-Prior, P. *et al.* Structural basis for the anticoagulant activity of the thrombin–thrombomodulin complex. *Nature* **404**, 518–525 (2000).

16. Marino, F., Pelc, L. A., Vogt, A., Gandhi, P. S. & Di Cera, E. Engineering Thrombin for Selective Specificity toward Protein C and PAR1. *Journal of Biological Chemistry* **285**, 19145 (2010).

17. Tsiang, M. *et al.* Functional Mapping of the Surface Residues of Human Thrombin. *Journal of Biological Chemistry* **270**, 16854–16863 (1995).

18. Vu, T. K. H., Wheaton, V. I., Hung, D. T., Charo, I. & Coughlin, S. R. Domains specifying thrombin–receptor interaction. *Nature* **353**, 674–677 (1991).

19. Gandhi, P. S., Chen, Z., Mathews, F. S. & Di Cera, E. Structural identification of the pathway of long-range communication in an allosteric enzyme. *Proceedings of the National Academy of Sciences* **105**, 1832–1837 (2008).

20. Ishihara, H. *et al.* Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* **386**, 502–506 (1997).

21. Bah, A., Chen, Z., Bush-Pelc, L. A., Mathews, F. S. & Di Cera, E. Crystal structures of murine thrombin in complex with the extracellular fragments of murine protease-activated receptors PAR3 and PAR4. *Proceedings of the National Academy of Sciences* **104**, 11603–11608 (2007).

22. Rydel, T. J., Tulinsky, A., Bode, W. & Huber, R. Refined structure of the Hirudin-thrombin complex. *Journal of Molecular Biology* **221**, 583–601 (1991).

23. Kahn, M. L. *et al.* A dual thrombin receptor system for platelet activation. *Nature* **394**, 690–694 (1998).

24. Bar-Shavit, R. Hormone-like activity of human thrombin. *Annals of the New York Academy of Sciences* **485**, 335–348 (1986).

25. Mackie, E. J. *et al.* Protease-Activated Receptors: A Means of Converting Extracellular Proteolysis into Intracellular Signals. *IUBMB Life* **53**, 277–281 (2002).

26. Xu, W. F. *et al.* Cloning and characterization of human protease-activated receptor 4. *Proceedings of the National Academy of Sciences U S A* **95**, 6642–6646 (1998).

27. Chandrabalan, A. & Ramachandran, R. Molecular mechanisms regulating Proteinase-Activated Receptors (PARs). *FEBS Journal*, **288,** 2697–2726 (2021).

28. Hollenberg, M. D., Saifeddine, M., Al-Ani, B. & Kawabata, A. Proteinase-activated receptors: Structural requirements for activity, receptor cross-reactivity, and receptor selectivity of receptor-activating peptides. *Journal of Physiology and Pharmacology* **75**, 832–841 (1997).

29. Zhao, P., Metcalf, M. & Bunnett, N. W. Biased signalling of protease-activated receptors. *Frontiers in Endocrinology*, **5** (2014).

30. Ludeman, M. J. *et al.* PAR1 cleavage and signalling in response to activated protein C and thrombin. *Journal of Biological Chemistry* **280**, 13122–13128 (2005).

31. Austin, K. M., Covic, L. & Kuliopulos, A. Matrix metalloproteases and PAR1 activation. *Blood* **121**, 431–439 (2013).

32. Han, X. & Nieman, M. T. The domino effect triggered by the tethered ligand of the protease activated receptors. *Thrombosis Research* **196**, 87–98 (2020).

33. Stoller, M. L. *et al.* Neutrophil cathepsin G proteolysis of protease-activated receptor 4 generates a novel, functional tethered ligand. *Blood Advances* **6**, 2303–2308 (2022).

34. Loew, D. *et al.* Proteolysis of the exodomain of recombinant protease-activated receptors: Prediction of receptor activation or inactivation by MALDI mass spectrometry. *Biochemistry* **39**, 10812–10822 (2000).

35. Zhao, P., Metcalf, M. & Bunnett, N. W. Biased signalling of protease-activated receptors. *Frontiers in Endocrinology (Lausanne)* **5**, 1–16 (2014).

36. French, S. L. & Hamilton, J. R. Protease-activated receptor 4: from structure to function and back again. *British Journal of Pharmacology* ,**173,** 2952–2965 (2016).

37. Kataoka, H. *et al.* Protease-activated receptors 1 and 4 mediate thrombin signalling in endothelial cells. *Blood* **102**, 3224–3231 (2003).

38. Hamilton, J. R., Frauman, A. G. & Cocks, T. M. Increased expression of protease-activated receptor-2 (PAR2) and PAR4 in human coronary artery by inflammatory stimuli unveils endothelium-dependent relaxations to PAR2 and PAR4 agonists. *Circulation Research* **89**, 92–98 (2001).

39. Tissue expression of F2RL3 - Summary - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000127533-F2RL3/tissue.

40. Karlsson, M. *et al.* A single–cell type transcriptomics map of human tissues. *Science Advances* **7**, (2021).

41. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

42. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acid Research* **50**, 439–444 (2022).

43. He, X. *et al.* AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors. *Acta Pharmacologica Sinica* **44**,1–7 (2022)

44. Lee, C., Su, B.-H. & Jane Tseng Corresponding author Tseng, Y. Y. Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Briefings in Bioinformatics* **23** (2022)

45. Nieman, M. T. & Schmaier, A. H. Interaction of Thrombin with PAR1 and PAR4 at the Thrombin Cleavage Site. *Biochemistry* **46**, 8603–8610 (2007).

46. Jacques, S. L., LeMasurier, M., Sheridan, P. J., Seeley, S. K. & Kuliopulos, A. Substrate-assisted catalysis of the PAR1 thrombin receptor: Enhancement of macromolecular association and cleavage. *Journal of Biological Chemistry* **275**, 40671–40678 (2000).

47. Ayala, Y. M. *et al.* Molecular mapping of thrombin-receptor interactions. *Proteins: Structure, Function and Genetics* **45**, 107–116 (2001).

48. Wolberg, A. S. & Campbell, R. A. Thrombin Generation, Fibrin Clot Formation and Hemostasis. *Transfusion and Apheresis Science* **38**, 15 (2008).

49. Covic, L., Singh, C., Smith, H. & Kuliopulos, A. Role of the PAR4 thrombin receptor in stabilizing platelet-platelet aggregates as revealed by a patient with Hermansky-Pudlak syndrome. *undefined* **87**, 722–727 (2002).

50. Jacques, S. L. & Kuliopulos, A. Protease-activated receptor-4 uses dual prolines and an anionic retention motif for thrombin recognition and cleavage. *Biochemical Journal* **376**, 733–740 (2003).

51. Li, S. *et al.* Oligosaccharide Blocks PAR1 (Proteinase-Activated Receptor 1)-PAR4-Mediated Platelet Activation by Binding to Thrombin Exosite II and Impairs Thrombosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **43**, 253-266 (2022).

52. Sánchez Centellas, D., Gudlur, S., Vicente-Carrillo, A., Ramström, S. & Lindahl, T. L. A cluster of aspartic residues in the extracellular loop II of PAR 4 is important for thrombin interaction and activation of platelets. *Thrombosis Research* **154**, 84–92 (2017).

53. Faruqi, T. R., Weiss, E. J., Shapiro, M. J., Huang, W. & Coughlin, S. R. Structure-function analysis of protease-activated receptor 4 thetered Ligand peptides. Determinants of specificity and utility in assays of receptor function. *Journal of Biological Chemistry* **275**, 19728–19734 (2000).

54. Thibeault, P. E. *et al.* Molecular basis for activation and biased signalling at the thrombin-activated GPCR proteinase activated receptor-4 (PAR4). *Journal of Biological Chemistry* **295**, 2520–2540 (2020).

55. Canto, I., J.K. Soh, U. & Trejo, J. Allosteric Modulation of Protease-Activated Receptor Signalling. *Mini-Reviews in Medicinal Chemistry* **12**, 804–811 (2012).

56. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acid Research* **28**, 235–242 (2000).

57. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073–1081 (2009).

58. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics* **76**, 7.20.1-7.20.41 (2013).

59. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).

60. Ng, P. C. & Henikoff, S. Predicting Deleterious Amino Acid Substitutions. *Genome Research* **11**, 863–874 (2001).

61. Flanagan, S. E., Patch, A. M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic Testing and Molecular Biomarkers* **14**, 533–537 (2010).

62. Edelstein, L. C. *et al.* Common variants in the human platelet PAR4 thrombin receptor alter platelet function and differ by race. *Blood* **124**, 3450–3458 (2014).

63. Norman, J. E. *et al.* Protease-activated receptor 4 variant p.Tyr157Cys reduces platelet functional responses and alters receptor trafficking. *Arteriosclerosis, Thrombosis, and Vascular Biology* **36**, 952–960 (2016).

64. Clarke, L. *et al.* The 1000 Genomes Project: data management and community access. *Nature Methods* **9**, 459–462 (2012).

65. Tourdot, B. E. *et al.* Genetic variant in human PAR (Protease-Activated Receptor) 4 enhances thrombus formation resulting in resistance to antiplatelet therapeutics. *Arteriosclerosis, Thrombosis, and Vascular Biology* **38**, 1632–1643 (2018).

66. Merali, S. *et al.* New oral protease-activated receptor 4 antagonist BMS-986120: tolerability, pharmacokinetics, pharmacodynamics, and gene variant effects in humans. *Platelets* **33**, 969-978(2022)

67. Boelig, R. C. *et al.* Platelet protease activated receptor 4 (PAR 4) receptor genotype is associated with an increased risk of preterm birth. *Journal of Thrombosis and Haemostasis* **20**, 2419–2428 (2022).

68. Denorme, F. *et al.* The predominant PAR4 variant in individuals of African ancestry worsens murine and human stroke outcomes. *Journal of Clinical Investigation* **133** (2023)

69. Orban, T. PAR4 activation involves extracellular loop-3 and transmembrane residue Thr 153. *Blood* **136**, 2217-2228 (2021).

70. Han, X. *et al.* A mouse model of the protease-activated receptor 4 Pro310Leu variant has reduced platelet reactivity. *Journal of Thrombosis and Haemostasis* **22**, 1715–1726 (2024).

71. Vu, T., Hung, D., Wheaton, V. & Coughlin, S. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* **64**, 1057–1068 (1991).

72. Han, X., Nieman, M. T. & Kerlin, B. A. Protease-activated receptors: An illustrated review. *Research and Practice in Thrombosis and Haemostasis* **5**, 17–26 (2021).

73. Mihara, K., Ramachandran, R., Renaux, B., Saifeddine, M. & Hollenberg, M. D. Neutrophil Elastase and Proteinase-3 Trigger G Protein-biased Signalling through Proteinase-activated Receptor-1 (PAR1). *Journal of Biological Chemistry* **288**, 32979–32990 (2013).

74. Hakanishi-Matsui, M. *et al.* PAR3 is a cofactor for PAR4 activation by thrombin. *Nature* **404**, 609–613 (2000).

75. Nystedt, S., Emilsson, K., Wahlestedt, C. & Sundelin, J. Molecular cloning of a potential proteinase activated receptor. *Proceedings of the National Academy of Sciences U S A* **91**, 9208–9212 (1994).

76. Miike, S., McWilliam, A. S. & Kita, H. Trypsin Induces Activation and Inflammatory Mediator Release from Human Eosinophils Through Protease-Activated Receptor-2. *The Journal of Immunology* **167**, 6615–6622 (2001).

77. Hakanishi-Matsui, M. *et al.* PAR3 is a cofactor for PAR4 activation by thrombin. *Nature* **404**, 609–613 (2000).

78. Lin, H., Liu, A. P., Smith, T. H. & Trejo, J. A. Cofactoring and dimerization of proteinase-activated receptors. *Pharmacological Reviews* , **65,** 1198–1213 (2013).

79. Covic, L., Gresser, A. L. & Kuliopulos, A. Biphasic kinetics of activation and signalling for PAR1 and PAR4 thrombin receptors in platelets. *Biochemistry* **39**, 5458–5467 (2000).

80. Han, X. & Nieman, M. T. The domino effect triggered by the tethered ligand of the protease activated receptors. *Thrombosis Research* **196**, 87–98 (2020).

81. Han, X., Nieman, M. T. & Kerlin, B. A. Protease-activated receptors: An illustrated review. *Research and Practice in Thrombosis and Haemostasis* **5**, 17–26 (2021).

82. Nieman, M. T. & Schmaier, A. H. Interaction of Thrombin with PAR1 and PAR4 at the Thrombin Cleavage Site. *Biochemistry* **46**, 8603–8610 (2007).

83. Jacques, S. L., LeMasurier, M., Sheridan, P. J., Seeley, S. K. & Kuliopulos, A. Substrate-assisted catalysis of the PAR1 thrombin receptor: Enhancement of macromolecular association and cleavage. *Journal of Biological Chemistry* **275**, 40671–40678 (2000).

84. Ayala, Y. M. *et al.* Molecular mapping of thrombin-receptor interactions. *Proteins: Structure, Function, and Bioinformatics* **45**, 107–116 (2001).

85. Wolberg, A. S. & Campbell, R. A. Thrombin Generation, Fibrin Clot Formation and Hemostasis. *Transfusion and Apheresis Science* **38**, 15 (2008).

86. Covic, L., Singh, C., Smith, H. & Kuliopulos, A. Role of the PAR4 thrombin receptor in stabilizing platelet-platelet aggregates as revealed by a patient

with Hermansky-Pudlak syndrome. *Thrombosis and Haemostasis* **87**, 722–727 (2002).

87. Jacques, S. L. & Kuliopulos, A. Protease-activated receptor-4 uses dual prolines and an anionic retention motif for thrombin recognition and cleavage. *Biochemical Journal* **376**, 733–740 (2003).

88. Hamilton, J. R., Frauman, A. G. & Cocks, T. M. Increased expression of protease-activated receptor-2 (PAR2) and PAR4 in human coronary artery by inflammatory stimuli unveils endothelium-dependent relaxations to PAR2 and PAR4 agonists. *Circulation Research* **89**, 92–98 (2001).

89. Sánchez Centellas, D., Gudlur, S., Vicente-Carrillo, A., Ramström, S. & Lindahl, T. L. A cluster of aspartic residues in the extracellular loop II of PAR 4 is important for thrombin interaction and activation of platelets. *Thrombosis Research* **154**, 84–92 (2017).

90. Thibeault, P. E. *et al.* Molecular basis for activation and biased signalling at the thrombin-activated GPCR proteinase activated receptor-4 (PAR4). *Journal of Biological Chemistry* **295**, 2520–2540 (2020).

91. Lim, J. *et al.* Diet-induced obesity, adipose inflammation, and metabolic dysfunction correlating with PAR2 expression are attenuated by PAR2 antagonism. *FASEB JOURNALournal* **27**, 4757–4767 (2013).

92. Hollenberg, M. D., Saifeddine, M., Al-Ani, B. & Gui, Y. Proteinase-activated receptor 4 (PAR4): action of PAR4-activating peptides in vascular and gastric tissue and lack of cross-reactivity with PAR1 and PAR2. *Can Journal of Physiology and Pharmacology* **77**, 458–464 (1999).

93. Weiss, E. J., Hamilton, J. R., Lease, K. E. & Coughlin, S. R. Protection against thrombosis in mice lacking PAR3. *Blood* **100**, 3240–3244 (2002).

94. Bah, A., Chen, Z., Bush-Pelc, L. A., Mathews, F. S. & Di Cera, E. Crystal structures of murine thrombin in complex with the extracellular fragments of murine protease-activated receptors PAR3 and PAR4. *Proceedings of the National Academy of Sciences* **104**, 11603–11608 (2007).

95. Li, S., Tarlac, V. & Hamilton, J. R. Using par4 inhibition as an anti-thrombotic approach: Why, how, and when? *International Journal of Molecular Science* **20**, (2019).

96. Lee, R. H. *et al.* Genetic deletion of platelet PAR4 results in reduced thrombosis and impaired hemostatic plug stability. *Journal of Thrombosis and Haemostasis* **20**, 422–433 (2022).

97. Covic, L., Gresser, A. L. & Kuliopulos, A. Biphasic kinetics of activation and signalling for PAR1 and PAR4 thrombin receptors in platelets. *Biochemistry* **39**, 5458–5467 (2000).

98. Vandendries, E. R., Hamilton, J. R., Coughlin, S. R., Furie, B. & Furie, B. C. Par4 is required for platelet thrombus propagation but not fibrin generation in a mouse model of thrombosis. *Proceedings of the National Academy of Sciences U S A* **104**, 288–292 (2006).

99. Coughlin, S. R. Thrombin signalling and protease-activated receptors. *Nature* , **407,** 258–264 (2000).

100. Dabek, M. *et al.* Luminal Cathepsin G and Protease-Activated Receptor 4: A Duet Involved in Alterations of the Colonic Epithelial Barrier in Ulcerative Colitis. *The American Journal of Pathology* **175**, 207–214 (2009).

101. Reséndiz, J. C., Kroll, M. H. & Lassila, R. Protease-activated receptor-induced Akt activation – regulation and possible function. *Journal of Thrombosis and Haemostasis* **5**, 2484–2493 (2007).

102. Pavic, G. *et al.* Thrombin receptor protease-activated receptor 4 is a key regulator of exaggerated intimal thickening in diabetes mellitus. *Circulation* **130**, 1700–1711 (2014).

103. Wang, H., Ricklin, D. & Lambris, J. Complement-activation fragment C4a mediates effector functions by binding as untethered agonist to protease-activated receptors 1 and 4. *Proceedings of the National Academy of Sciences U S A* **114**, 10948–10953 (2017).

104. Shapiro, M. J., Trejo, J., Zeng, D. & Coughlin, S. R. Role of the thrombin receptor's cytoplasmic tail in intracellular trafficking: Distinct determinants for agonist-triggered versus tonic internalization and intracellular localization. *Journal of Biological Chemistry* **271**, 32874–32880 (1996).

105. Shapiro, M. J., Weiss, E. J., Faruqi, T. R. & Coughlin, S. R. Protease-activated receptors 1 and 4 are shut off with distinct kinetics after activation by thrombin. *Journal of Biological Chemistry* **275**, 25216–25221 (2000).

106. Smith, T. H. *et al.* Protease-activated Receptor-4 Signalling and Trafficking Is Regulated by the Clathrin Adaptor Protein Complex-2 Independent of β-Arrestins. *Journal of Biological Chemistry* **291**, 18453–18464 (2016).

107. Cunningham, M. R. *et al.* Novel role for proteinase-activated receptor 2 (PAR2) in membrane trafficking of proteinase-activated receptor 4 (PAR4). *Journal of Biological Chemistry* **287**, 16656–16669 (2012).

108. Ramachandran, R. *et al.* Targeting a proteinase-activated receptor 4 (PAR4) carboxyl terminal motif to regulate platelet function. *Molecular Pharmacology* **91**, 287–295 (2017).

109. Li, D., D'Angelo, L., Chavez, M. & Woulfe, D. S. Arrestin-2 differentially regulates PAR4 and ADP receptor signalling in platelets. *Journal of Biological Chemistry* **286**, 3805–3814 (2011).

110. Vanderboor, C. M. G. *et al.* Proteinase-Activated Receptor 4 Activation Triggers Cell Membrane Blebbing through RhoA and β-Arrestin. *Molecular Pharmacology* **97**, 365–376 (2020).

111. Trivigno, S. M. G. *et al.* The C-Type Lectin Receptor CD93 Regulates Platelet Activation and Surface Expression of the Protease Activated Receptor 4. *Thrombosis and Haemostasis* **124**, 122–134 (2023).

112. Derian, C. K., Santulli, R. J., Tomko, K. A., Haertlein, B. J. & Andrade-Gordon, P. Species differences in platelet responses to thrombin and SFLLRN. Receptor-mediated calcium mobilization and aggregation, and regulation by protein kinases. *Thrombosis Research* **78**, 505–519 (1995).

113. Andrade-Gordon, P. *et al.* Administration of a potent antagonist of protease-activated receptor-1 (PAR-1) attenuates vascular restenosis following balloon angioplasty in rats. *Journal of Pharmacology and Experimental Therapeutics* **298**, 34–42 (2001).

114. Derian, C. K. *et al.* Blockade of the thrombin receptor protease-activated receptor-1 with a small-molecule antagonist prevents thrombus formation and vascular occlusion in nonhuman primates. *Journal of Pharmacology and Experimental Therapeutics* **304**, 855–861 (2003).

115. Wong, P. C. *et al.* Blockade of protease-activated receptor- 4(PAR4) provides robust antithrombotic activity with low bleeding. *Science Translational Medicine* **9**, (2017).

116. Renna, S. A. *et al.* Human and mouse PAR4 are functionally distinct receptors: studies in novel humanized mice. *Journal of Thrombosis and Haemostasis* **20**, 1236-1247 (2022)

117. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular System Biology* **7**, 539 (2011).

118. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acid Research* **49**, D480–D489 (2021).

119. Hollenberg, M. D., Saifeddine, M., Al-Ani, B. & Gui, Y. Proteinase-activated receptor 4 (PAR4): action of PAR4-activating peptides in vascular and gastric tissue and lack of cross-reactivity with PAR1 and PAR2. *Journal of Physiology and Pharmacology* **77**, 458–464 (1999).

120. Lee, R. H. *et al.* Genetic deletion of platelet PAR4 results in reduced thrombosis and impaired hemostatic plug stability. *Journal of Thrombosis and Haemostasis* **20**, 422–433 (2022).

121. Weiss, E. J., Hamilton, J. R., Lease, K. E. & Coughlin, S. R. Protection against thrombosis in mice lacking PAR3. *Blood* **100**, 3240–3244 (2002).

122. Lee, H., Sturgeon, S. A., Jackson, S. P. & Hamilton, J. R. The contribution of thrombin-induced platelet activation to thrombus growth is diminished under pathological blood shear conditions. *Thrombosis and Haemostasis* **107**, 328–337 (2012).

123. Lee, H., Sturgeon, S. A., Mountford, J. K., Jackson, S. P. & Hamilton, J. R. Safety and efficacy of targeting platelet proteinase-activated receptors in combination with existing anti-platelet drugs as antithrombotics in mice. *British Journal of Pharmacology* **166**, 2188–2197 (2012).

124. Hamilton, J. R., Cornelissen, I. & Coughlin, S. R. Impaired hemostasis and protection against thrombosis in protease-activated receptor 4-deficient mice

is due to lack of thrombin signalling in platelets. *Journal of Thrombosis and Haemostasis* **2**, 1429–1435 (2004).

125. Mao, Y., Zhang, M., Tuma, R. F. & Kunapuli, S. P. Deficiency of PAR4 attenuates cerebral ischemia/reperfusion injury in mice. *Journal of Cerebral Blood Flow & Metabolism* **30**, 1044–1052 (2010).

126. Adams, M. N. *et al.* Structure, function and pathophysiology of protease activated receptors. *Pharmacology and Therapeutics* **130**, 248–282 (2011).

127. Sambrano, G. R., Weiss, E. J., Zheng, Y. W., Huang, W. & Coughlin, S. R. Role of thrombin signalling in platelets in haemostasis and thrombosis. *Nature* **413**, 74–78 (2001).

128. Kleeschulte, S., Jerrentrup, J., Gorski, D., Schmitt, J. & Fender, A. Evidence for functional PAR-4 thrombin receptor expression in cardiac fibroblasts and its regulation by high glucose: PAR-4 in cardiac fibroblasts. *International Journal of Cardiology* **252**, 163–166 (2018).

129. Lee, S. K. *et al.* PAR4 Inhibition Reduces Coronary Artery Atherosclerosis and Myocardial Fibrosis in SR-B1/LDLR Double Knockout Mice. *Arteriosclerosis, Thrombosis, and Vascular Biology* **43**, 0–00 (2023).

130. Erreger, K. *et al.* Role of protease-activated receptor 4 in mouse models of acute and chronic kidney injury. *American Journal of Physiology-Renal Physiology***326**, F219–F226 (2024).

131. Song, Y. W. *et al.* Microscopic in-situ analysis of the mucosal healing around implants treated by protease activated receptor 4-agonist peptide or perpendicularly protruded type I collagen in rats. *Journal of Biomedical Materials Research Part B* **112**, e35330 (2024).

132. Zhang, Y. *et al.* Platelets mediate acute hepatic microcirculatory injury in a protease-activated-receptor-4-dependent manner after extended liver resection. *Transplant Immunology* **77**, 101795 (2023).

133. Fang, H. *et al.* Blocking protease-activated receptor 4 alleviates liver injury induced by brain death. *Biochemical and Biophysical Research Communications* **595**, 47–53 (2022).

134. Asfaha, S. *et al.* Protease-activated receptor-4: a novel mechanism of inflammatory pain modulation. *British Journal of Pharmacology* **150**, 176–185 (2006).

135. Augé, C., Balz-Hara, D., Steinhoff, M., Vergnolle, N. & Cenac, N. S. Protease-activated receptor-4 (PAR4): a role as inhibitor of visceral pain and hypersensitivity. *Neurogastroenterology & Motility* **21**, 1189-e107 (2009).

136. Barra, A. *et al.* Protease-activated receptor 4 plays a role in lipopolysaccharide-induced inflammatory mechanisms in murine macrophages. *Naunyn-Schmiedeberg's Archives of Pharmacology* **394**, 853–862 (2021).

137. Bock, M. *et al.* Platelets differentially modulate CD4 + Treg activation via GPIIa/IIIb-, fibrinogen-, and PAR4-dependent pathways. *Immunologic Research* **70**, 185-196 (2021).

138. Busso, N. *et al.* Essential role of platelet activation via protease activated receptor 4 in tissue factor-initiated inflammation. *Arthritis Research and Therapies* **10**, 1–8 (2008).

139. de Stoppelaar, S. F. *et al.* Protease activated receptor 4 limits bacterial growth and lung pathology during late stage Streptococcus pneumoniae induced pneumonia in mice. *Thrombosis and Haemostasis* **110**, 582–592 (2013).

140. Houle, S., Papez, M. D., Ferazzini, M., Hollenberg, M. D. & Vergnolle, N. Neutrophils and the kallikrein–kinin system in proteinase-activated receptor 4-mediated inflammation in rodents. *British Journal of Pharmacology* **146**, 670 (2005).

141. Kolpakov, M. *et al.* Protease-activated receptor 4 deficiency offers cardioprotection after acute ischemia reperfusion injury. *Journal of Molecular and Cellular Cardiology* **90**, 21–29 (2016).

142. Kolpakov, M. A. *et al.* Loss of Protease-Activated Receptor 4 Prevents Inflammation Resolution and Predisposes the Heart to Cardiac Rupture after Myocardial Infarction. *Circulation* **142**, 758–775 (2020).

143. Scridon, A. *et al.* Vascular protease-activated receptor 4 upregulation, increased platelet aggregation, and coronary lipid deposits induced by long-term dabigatran administration – results from a diabetes animal model. *Journal of Thrombosis and Haemostasis* **17**, 538–550 (2019).

144. Tatsumi, K. *et al.* Protease-activated receptor 4 protects mice from Coxsackievirus B3 and H1N1 influenza A virus infection. *Cellular Immunology* **344**, 103949 (2019).

145. Yang, J. *et al.* An optimized agonist peptide of protease-activated receptor 4 and its use in a validated platelet-aggregation assay. *Platelets* **33**, 979-986 (2022).

146. Moschonas, I. C. *et al.* Molecular requirements involving the human platelet protease-activated receptor-4 mechanism of activation by peptide analogues of its tethered-ligand. *Platelets* **28**, 812–821 (2017).

147. Hollenberg, M. D. & Saifeddine, M. Proteinase-activated receptor 4 (PAR4): activation and inhibition of rat platelet aggregation by PAR4-derived peptides. *Can Journal of Physiology and Pharmacology* **79**, 439–442 (2001).

148. Ma, L. *et al.* Proteinase-activated receptors 1 and 4 counter-regulate endostatin and VEGF release from human platelets. *Proceedings of the National Academy of Sciences U S A* **102**, 216 (2005).

149. Hollenberg, M. D., Saifeddine, M., Sandhu, S., Houle, S. & Vergnolle, N. Proteinase-activated receptor-4: evaluation of tethered ligand-derived

peptides as probes for receptor function and as inflammatory agonists in vivo. *British Journal of Pharmacology* **143**, 443–454 (2004).

150. Covic, L., Gresser, A. L., Talavera, J., Swift, S. & Kuliopulos, A. Activation and inhibition of G protein-coupled receptors by cell-penetrating membrane-tethered peptides. *Proceedings of the National Academy of Sciences* **99**, 643–648 (2002).

151. Covic, L., Misra, M., Badar, J., Singh, C. & Kuliopulos, A. Pepducin-based intervention of thrombin-receptor signalling and systemic platelet activation. *Nature Medicine* **8**, 1161–1165 (2002).

152. Stampfuss, J. J., Schrör, K. & Weber, A.-A. Inhibition of platelet thromboxane receptor function by a thrombin receptor–targeted pepducin. *Nature Medicine* **9**, 1447–1447 (2003).

153. Leger, A. *et al.* Blocking the protease-activated receptor 1-4 heterodimer in platelet-mediated thrombosis. *Circulation* **113**, 1244–1254 (2006).

154. O'Callaghan, K., Kuliopulos, A. & Covic, L. Turning receptors on and off with intracellular pepducins: New insights into G-protein-coupled receptor drug development. *Journal of Biological Chemistry* **287,** 12787–12796 (2012).

155. Kahn, M., Nakanishi-Matsui, M., Shapiro, M., Ishihara, H. & Coughlin, S. Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *Journal of Clinical Investigation* **103**, 879–887 (1999).

156. Mumaw, M. M., Fuente, M. de la, Noble, D. N. & Nieman, M. T. Targeting the anionic region of human protease-activated receptor 4 inhibits platelet aggregation and thrombosis without interfering with hemostasis. *Journal of Thrombosis and Haemostasis* **12**, 1331–1341 (2014).

157. Mumaw, M. M., De La Fuente, M., Arachiche, A., Wahl, J. K. & Nieman, M. T. Development and characterization of monoclonal antibodies against Protease Activated Receptor 4 (PAR4). *Thrombosis Research* **135**, 1165–1171 (2015).

158. Hamilton, J. R. Protease-activated receptors as targets for antiplatelet therapy. *Blood Reviews* **23**, 61–65 (2009).

159. French, S. L. *et al.* A function-blocking PAR4 antibody is markedly antithrombotic in the face of a hyperreactive PAR4 variant. *Blood Advances* **2**, 1283–1293 (2018).

160. Fang-Yu Lee, *et al.* Synthesis of 1-Benzyl-3-(5'-hydroxymethyl-2'-furyl)indazole Analogues as Novel Antiplatelet Agents. *Journal of Medicinal Chemistry* **44**, 3746–3749 (2001).

161. Powell, J. R., Reeves, R. A., Marino, M. R., Cazaubon, C. & Nisato, D. A Review of the New Angiotensin II-Receptor Antagonist Irbesartan. *Cardiovascular Drug Reviews* **16**, 169–194 (1998).

162. Wu, C. C. *et al.* Selective inhibition of protease-activated receptor 4-dependent platelet activation by YD-3. *Thrombosis and Haemostasis* **87**, 1026–1033 (2002).

163. Wu, C. C. & Teng, C. M. Comparison of the effects of PAR1 antagonists, PAR4 antagonists, and their combinations on thrombin-induced human platelet activation. *European Journal of Pharmacology* **546**, 142–147 (2006).

164. Huang, L. J. *et al.* Synthesis of N2-(substituted benzyl)-3-(4-methylphenyl)indazoles as novel anti-angiogenic agents. *Bioorganic & Medicinal Chemistry* **14**, 528–536 (2006).

165. Young, S., Duvernay, M., Schulte, M., Lindsley, C. & Hamm, H. Synthesis of indole derived protease-activated receptor 4 antagonists and characterization in human platelets. *PLoS One* **8**, (2013).

166. Frémont, L. Biological effects of resveratrol. *Life Sciences* **66**, 663–673 (2000).

167. Chiang, Y. C. *et al.* 3,5,2',4'-Tetramethoxystilbene, a fully methylated resveratrol analog, prevents platelet aggregation and thrombus formation by targeting the protease-activated receptor 4 pathway. *Chemico-Biological Interactions* **357**, (2022).

168. Wen, K. *et al.* Recent Research on Flavonoids and their Biomedical Applications. *Current Medicinal Chemistry* **28**, 1042–1066 (2021).

169. Lin, Y.-T. *et al.* Discovery of 7, 4'-dimethoxy-3-hydroxyflavone as a protease-activated receptor 4 antagonist with antithrombotic activity and less bleeding tendency in mice. *Biochemical Pharmacology* **202**, 115152 (2022).

170. Merali, S. *et al.* New oral protease-activated receptor 4 antagonist BMS-986120: tolerability, pharmacokinetics, pharmacodynamics, and gene variant effects in humans. *Platelets* **33**, 969-978 (2022)

171. Wilson, S. J. *et al.* PAR4 (Protease-Activated Receptor 4) Antagonism with BMS-986120 Inhibits Human Ex Vivo Thrombus Formation. *Arteriosclerosis, Thrombosis, and Vascular Biology* **38**, 448–456 (2018).

172. Luo, J. *et al.* Antagonism of Protease-Activated Receptor 4 Protects Against Traumatic Brain Injury by Suppressing Neuroinflammation via Inhibition of Tab2/NF-κB Signalling. *Neuroscience Bulletin* **37**, 242–254 (2021).

173. Priestley, E. S. *et al.* Discovery of Two Novel Antiplatelet Clinical Candidates (BMS-986120 and BMS-986141) That Antagonize Protease-Activated Receptor 4. *Journal of Medicinal Chemistry* **65**, 8843-8854 (2022)

174. Meah, M. N. *et al.* Antithrombotic effects of combined PAR (protease-activated receptor)-4 antagonism and factor Xa inhibition. *Arteriosclerosis, Thrombosis, and Vascular Biology* **40**, 2678–2685 (2020).

175. Merali, S. *et al.* First-in-human study to assess the safety, pharmacokinetics, and pharmacodynamics of BMS-986141, a novel, reversible, small-molecule, PAR4 agonist in non-Japanese and Japanese healthy participants. *Platelets* **34**, 2222846 (2023).

176. Zhang, X. *et al.* Discovery of Potent and Selective Quinoxaline-Based Protease-Activated Receptor 4 (PAR4) Antagonists for the Prevention of Arterial Thrombosis. *Journal of Medicinal Chemistry* **67**, 3571–3589 (2024).

177. Smith, S. T. *et al.* Discovery of Protease-Activated Receptor 4 (PAR4)-Tethered Ligand Antagonists Using Ultralarge Virtual Screening. *ACS Pharmacology & Translational Science Journal* **7**, 1086–1100 (2024).

178. Chen, P. *et al.* Discovery of 2,3-Dihydro[1,4]dioxino[2,3-g]benzofuran Derivatives as Protease Activated Receptor 4 (PAR4) Antagonists with Potent Antiplatelet Aggregation Activity and Low Bleeding Tendency. *Journal of Medicinal Chemistry* **67**, 5502–5537 (2024).

179. French, S. L. & Hamilton, J. R. Drugs targeting protease-activated receptor-4 improve the antithrombotic therapeutic window. *Annals of Translational Medicine* **5**, 10–10 (2017).

180. McCarthy, C. P., Steg, G. P. & Bhatt, D. L. The management of antiplatelet therapy in acute coronary syndrome patients with thrombocytopenia: a clinical conundrum. *European Heart Journal* **38**, 3488–3492 (2017).

181. Jneid, H. *et al.* Aspirin and Clopidogrel in Acute Coronary Syndromes: Therapeutic Insights From the CURE Study. *Archives of Internal Medicine* **163**, 1145–1153 (2003).

182. Pasalic, L. *et al.* Novel assay demonstrates that coronary artery disease patients have heightened procoagulant platelet response. *Journal of Thrombosis and Haemostasis* **16**, 1198–1210 (2018).

183. Li, S., Tarlac, V. & Hamilton, J. R. Using par4 inhibition as an anti-thrombotic approach: Why, how, and when? *International Journal of Molecular Science* **20**, (2019).

184. Ogawa, H. *et al.* Low-Dose Aspirin for Primary Prevention of Atherosclerotic Events in Patients With Type 2 Diabetes: A Randomized Controlled Trial. *Journal of the American Medical Association* **300**, 2134–2141 (2008).

185. Mehilli, J. *et al.* Randomized clinical trial of abciximab in diabetic patients undergoing elective percutaneous coronary interventions after treatment with a high loading dose of clopidogrel. *Circulation* **110**, 3627–3635 (2004).

186. Mitrugno, A. *et al.* The role of coagulation and platelets in colon cancer-associated thrombosis. *American Journal of Physiology-Cell Physiology* **316**, 264–273 (2019).

187. Yu, G. *et al.* Increased Expression of Protease-Activated Receptor 4 and Trefoil Factor 2 in Human Colorectal Cancer. *PLoS One* **10**, (2015).

188. Corbin, L. J. *et al.* Epigenetic Regulation of F2RL3 Associates With Myocardial Infarction and Platelet Function. *Circulation Research* **130**, 384–400 (2022).

189. Winfree, R. L. *et al.* Elevated protease-activated receptor 4 (PAR4) gene expression in Alzheimer's disease predicts cognitive decline. *Neurobiology of Aging* **140**, 93–101 (2024)

190. Gupta, R. *et al.* Protease-Activated Receptor Antagonist for Reducing Cardiovascular Events - A Review on Vorapaxar. *Current Problems in Cardiology* **48** (2021)

191. Judge, H. M. *et al.* PAR1 antagonists inhibit thrombin-induced platelet activation whilst leaving the PAR4-mediated response intact. *Platelets* **26**, 236–242 (2015).

192. Cunningham, M. R. *et al.* Novel role for proteinase-activated receptor 2 (PAR2) in membrane trafficking of proteinase-activated receptor 4 (PAR4). *Journal of Biological Chemistry* **287**, 16656–16669 (2012).

193. Norman, J. E. *et al.* Protease-activated receptor 4 variant p.Tyr157Cys reduces platelet functional responses and alters receptor trafficking. *Arteriosclerosis, Thrombosis, and Vascular Biology* **36**, 952–960 (2016).

194. Koushik, S. V., Chen, H., Thaler, C., Puhl, H. L. & Vogel, S. S. Cerulean, venus, and venusY67C FRET reference standards. *Biophysical Journal* **91**, L99–L101 (2006).

195. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**, 376–386 (2002).

196. Emmott, E. & Goodfellow, I. Identification of Protein Interaction Partners in Mammalian Cells Using SILAC-immunoprecipitation Quantitative Proteomics. *JoVE (Journal of Visualized Experiments)* **6** (2014)

197. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679–682 (2022).

198. Cianfrocco, M. A., Wong-Barnum, M., Youn, C., Wagner, R. & Leschziner, A. COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination. *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact* **Part F128771**, (2017).

199. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**, 712–725 (2006).

200. Luck, K., Charbonnier, S. & Travé, G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Letters* **586**, 2648–2661 (2012).

201. Kennedy, M. B. Origin of PDZ (DHR, GLGF) domains. *Trends in Biochemical Sciences***20**, 350 (1995).

202. Lee, H. J. & Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Communication and Signalling* **8**, 1–18 (2010).

203. Romero, G., Von Zastrow, M. & Friedman, P. A. Role of PDZ Proteins in Regulating Trafficking, Signalling, and Function of GPCRs: Means, Motif, and Opportunity. *Advances in Pharmacology* **62**, 279–314 (2011).

204. Morais Cabral, J. H. *et al.* Crystal structure of a PDZ domain. *Nature* **382**, 649–652 (1996).

205. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signalling complex assembly. *Journal of Cell Science* **114**, 3219–3231 (2001).

206. Wang, N. X., Lee, H. J. & Zheng, J. J. Therapeutic use of PDZ protein-protein interaction antagonism. *Drug News & Perspectives* **21**, 137 (2008).

207. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**, 376–386 (2002).

208. Chen, X., Wei, S., Ji, Y., Guo, X. & Yang, F. Quantitative proteomics using SILAC : Principles , applications , and developments. *Proteomics* **15**, 3175–3192 (2015)

209. Williams, G. R. *et al.* Exploring G protein-coupled receptor signalling networks using SILAC-based phosphoproteomics. *Methods* **92**, 36–50 (2016).

210. Xiao, K. & Sun, J. Elucidating structural and molecular mechanisms of β-arrestin-biased agonism at GPCRs via MS-based proteomics. *Cell Signal* **41**, 56–64 (2018).

211. Trester-Zedlitz, M., Burlingame, A., Kobilka, B. & Von Zastrow, M. Mass spectrometric analysis of agonist effects on posttranslational modifications of the β-2 adrenoceptor in mammalian cells. *Biochemistry* **44**, 6133–6143 (2005).

212. Cao, T. T., Deacon, H. W., Reczek, D., Bretscher, A. & Von Zastrow, M. A kinase-regulated PDZ-domain interaction controls endocytic sorting of the β2-adrenergic receptor. *Nature* **401**, 286–290 (1999).

213. Lau, E. K. *et al.* Quantitative encoding of the effect of a partial agonist on individual opioid receptors by multisite phosphorylation and threshold detection. *Science Signalling* **4**, (2011).

214. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acid Research* **49**, 412–419 (2021).

215. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acid Research* **51**, 418–427 (2023).

216. Bader, G. D., Betel, D. & Hogue, C. W. V. BIND: the Biomolecular Interaction Network Database. *Nucleic Acid Research* **31**, 248 (2003).

217. Xenarios, I. *et al.* DIP: the Database of Interacting Proteins. *Nucleic Acid Research* **28**, 289–291 (2000).

218. Breitkreutz, B.-J., Stark, C. & Tyers, M. The GRID: The General Repository for Interaction Datasets. *Genome Biology* **4**, 1–3 (2003).

219. Peri, S. *et al.* Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. *Genome Research* **13**, 2363–2371 (2003).

220. Hermjakob, H. *et al.* IntAct: an open source molecular interaction database. *Nucleic Acid Research* **32**, 452–455 (2004).

221. Zanzoni, A. *et al.* MINT: a Molecular INTeraction database. *FEBS Letters* **513**, 135–140 (2002).

222. Nishimura, D. BioCarta. *Biotech Software & Internet Report***2**, 117–120 (2004).

223. Karp, P. D. *et al.* Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acid Research* **33**, 6083–6089 (2005).

224. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).

225. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acid Research* **28**, 27–30 (2000).

226. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acid Research* **33**, 428–432 (2005).

227. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acid Research* **33**, 514–517 (2005).

228. Cherry, J. M. *et al.* SGD: Saccharomyces Genome Database. *Nucleic Acid Research* **26**, 73–79 (1998).

229. von Mering, C. *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acid Research* **33**, (2005).

230. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acid Research* **51**, 638–646 (2023).

231. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).

232. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acid Research* **30**, 1575–1584 (2002).

233. Romero, G., Von Zastrow, M. & Friedman, P. A. Role of PDZ Proteins in Regulating Trafficking, Signalling, and Function of GPCRs: Means, Motif, and Opportunity. *Advances in Pharmacology* **62**, 279–314 (2011).

234. Walsh, T. *et al.* Genomic Duplication and Overexpression of TJP2/ZO-2 Leads to Altered Expression of Apoptosis Genes in Progressive Nonsyndromic Hearing Loss DFNA51. *The American Journal of Human Genetics* **87**, 101 (2010).

235. Sambrotta, M. *et al.* Mutations in TJP2 cause progressive cholestatic liver disease. *Nature Genetics* **46**, 326–328 (2014).

236. Chlenski, A. *et al.* Organization and expression of the human zo-2 gene (tjp-2) in normal and neoplastic tissues. *Biochimica et Biophysica Acta* **1493**, 319–324 (2000).

237. Ooshio, T. *et al.* Involvement of LMO7 in the association of two cell-cell adhesion molecules, nectin and E-cadherin, through afadin and α-actinin in epithelial cells. *Journal of Biological Chemistry* **279**, 31365–31373 (2004).

238. Melcon, G. *et al.* Lmo7 is an emerin-binding protein that regulates the transcription of emerin and many other muscle-relevant genes. *Human Molecular Genetics* **15**, 3459–3472 (2006).

239. Kotaka, M. *et al.* Interaction of hCLIM1 , An Enigma Family Protein , With Actinin 2. *Journal of Cellular Biochemistry* **565**, 558–565 (2000).

240. Ono, R., Kaisho, T. & Tanaka, T. PDLIM1 inhibits NF-κB-mediated inflammatory signalling by sequestering the p65 subunit of NF-κB in the cytoplasm. *Scientific Reports 2015 5:1* **5**, 1–11 (2015).

241. Zhou, J. K., Fan, X., Cheng, J., Liu, W. & Peng, Y. PDLIM1: Structure, function and implication in cancer. *Cell Stress* **5**, 119 (2021).

242. Sensoy, O. & Weinstein, H. A mechanistic role of Helix 8 in GPCRs: Computational modeling of the dopamine D2 receptor interaction with the GIPC1–PDZ-domain. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1848**, 976–983 (2015).

243. De Vries, L., Lou, X., Zhao, G., Zheng, B. & Farquhar, M. G. GIPC, a PDZ domain containing protein, interacts specifically with the C terminus of RGS-GAIP. *Proceedings of the National Academy of Sciences U S A* **95**, 12340–12345 (1998).

244. Stamm, S. *et al.* Function of alternative splicing. *Gene* **344**, 1–20 (2005).

245. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185 (2002).

246. Knapp, B. *et al.* Affinity Proteomics Identifies Interaction Partners and Defines Novel Insights into the Function of the Adhesion GPCR VLGR1/ADGRV1. *Molecules* **27**, 3108 (2022).

247. Jong, Y. J. I., Harmon, S. K. & O'Malley, K. L. GPCR signalling from within the cell. *British Journal of Pharmacology* **175**, 4026–4035 (2018).

248. Palacios, I. M. RNA Processing: Splicing and the Cytoplasmic Localisation of mRNA Dispatch. *Current Biology* **12**, 50–52 (2002).

249. Kellogg, M. K., Miller, S. C., Tikhonova, E. B. & Karamyshev, A. L. SRPassing Co-translational Targeting: The Role of the Signal Recognition Particle in Protein Targeting and mRNA Protection. *International Journal of Molecular Science* **22**, (2021).

250. Hegde, R. S. & Keenan, R. J. The mechanisms of integral membrane protein biogenesis. *Nature Reviews Molecular Cell Biology* **23**, 107–124 (2021).

251. Saraogi, I. & Shan, S. ou. Molecular Mechanism of Co-translational Protein Targeting by the Signal Recognition Particle. *Traffic* **12**, 535–542 (2011).

252. Belous, A. *et al.* Mitochondrial P2Y-Like receptors link cytosolic adenosine nucleotides to mitochondrial calcium uptake. *Journal of Cellular Biochemistry* **92**, 1062–1073 (2004).

253. Abadir, P. M. *et al.* Identification and characterization of a functional mitochondrial angiotensin system. *Proceedings of the National Academy of Sciences U S A* **108**, 14849–14854 (2011).

254. Wang, Q. *et al.* 5-HTR3 and 5-HTR4 located on the mitochondrial membrane and functionally regulated mitochondrial functions. *Scientific Reports* **6**, (2016).

255. Bénard, G. *et al.* Mitochondrial CB_1 receptors regulate neuronal energy metabolism. *Nature Neuroscience* **15**, 558–564 (2012).

256. Gbahou, F. *et al.* Design and validation of the first cell-impermeant melatonin receptor agonist. *British Journal of Pharmacology* **174**, 2409–2421 (2017).

257. Suofu, Y. *et al.* Dual role of mitochondria in producing melatonin and driving GPCR signalling to block cytochrome c release. *Proceedings of the National Academy of Sciences U S A* **114**, E7997–E8006 (2017).

258. Grube, K. *et al.* Evidence for an intracellular localization of the adenosine A 2B receptor in rat cardiomyocytes. *Basic Research in Cardiology* **106**, 385–396 (2011).

259. Krzysko, J. *et al.* The Adhesion GPCR VLGR1/ADGRV1 Regulates the Ca2+ Homeostasis at Mitochondria-Associated ER Membranes. *Cells* **11**, 2790 (2022).

260. Wyant, G. A. *et al.* Mitochondrial remodeling and ischemic protection by G protein–coupled receptor 35 agonists. *Science* **377**, 621–629 (2022).

261. Armstrong, L. C., Saenz, A. J. & Bornstein, P. Metaxin 1 interacts with metaxin 2, a novel related protein associated with the mammalian mitochondrial outer membrane. *Journal of Cellular Biochemistry* **74**, 11–22 (1999).

262. Armstrong, L. C., Komiya, T., Bergman, B. E., Mihara, K. & Bornstein, P. Metaxin is a component of a preprotein import complex in the outer

membrane of the mammalian mitochondrion. *Journal of Biological Chemistry* **272**, 6510–6518 (1997).

263. Kozjak-Pavlovic, V. *et al.* Conserved roles of Sam50 and metaxins in VDAC biogenesis. *EMBO Rep* **8**, 576–582 (2007).

264. McCommis, K. S. & Baines, C. P. The role of VDAC in cell death: Friend or foe? *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1818**, 1444–1450 (2012).

265. Blachly-Dyson, E., Baldini, A., Litt, M., McCabe, E. R. B. & Forte, M. Human Genes Encoding the Voltage-Dependent Anion Channel (VDAC) of the Outer Mitochondrial Membrane: Mapping and Identification of Two New Isoforms. *Genomics* **20**, 62–67 (1994).

266. Shoshan-Barmatz, V., Shteinfer-Kuzmine, A. & Verma, A. VDAC1 at the Intersection of Cell Metabolism, Apoptosis, and Diseases. *Biomolecules* **10**, 1–40 (2020).

267. Thinnes, F. *et al.* Gadolinium as an opener of the outwardly rectifying Cl(-) channel (ORCC). Is there relevance for cystic fibrosis therapy? *Pflugers Archiv* **443 Suppl 1**, (2001).

268. Hiller, S. *et al.* Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321**, 1206–1210 (2008).

269. Chen, Y., Gaczynska, M., Osmulski, P., Polci, R. & Riley, D. J. Phosphorylation by Nek1 regulates opening and closing of voltage dependent anion channel 1. *Biochemical and Biophysical Research Communications* **394**, 798–803 (2010).

270. Blachly-Dyson, E. *et al.* Cloning and functional expression in yeast of two human isoforms of the outer mitochondrial membrane channel, the voltage-dependent anion channel. *Journal of Biological Chemistry* **268**, 1835–1841 (1993).

271. Dadsena, S. *et al.* Ceramides bind VDAC2 to trigger mitochondrial apoptosis. *Nature Communications* **10**, (2019).

272. Ham, S. J. *et al.* Decision between mitophagy and apoptosis by Parkin via VDAC1 ubiquitination. *Proceedings of the National Academy of Sciences U S A* **117**, 4281–4291 (2020).

273. Verrier, F., Mignotte, B., Jan, G. & Brenner, C. Study of PTPC composition during apoptosis for identification of viral protein target. *Annals of the New York Academy of Sciences* **1010**, 126–142 (2003).

274. Li, L. *et al.* Plasminogen kringle 5 induces endothelial cell apoptosis by triggering a voltage-dependent anion channel 1 (VDAC1) positive feedback loop. *Journal of Biological Chemistry* **289**, 32628–32638 (2014).

275. Shoshan-Barmatz, V., Krelin, Y. & Shteinfer-Kuzmine, A. VDAC1 functions in Ca 2+ homeostasis and cell life and death in health and disease. *Cell Calcium* **69**, 81–100 (2018).

276. Marginedas-Freixa, I. *et al.* Human erythrocytes release ATP by a novel pathway involving VDAC oligomerization independent of pannexin-1. *Scientific Reports* **8**, (2018).

277. Sébert, M. *et al.* Thrombin modifies growth, proliferation and apoptosis of human colon organoids: a protease-activated receptor 1- and protease-activated receptor 4-dependent mechanism. *British Journal of Pharmacology* **175**, 3656–3668 (2018).

278. Deniaud, A., Brenner, C. & Kroemer, G. Mitochondrial membrane permeabilization by HIV-1 Vpr. *Mitochondrion* **4**, 223–233 (2004).

279. Fender, A. C. *et al.* Thrombin receptor PAR4 drives canonical NLRP3 inflammasome signalling in the heart. *Basic Research in Cardiology* **115**, 3 (2020).

280. Strande, J. L. & Phillips, S. A. Thrombin increases inflammatory cytokine and angiogenic growth factor secretion in human adipose cells in vitro. *Journal of Inflammation* **6**, (2009).

281. Amisten, S. *et al.* An atlas of G-protein coupled receptor expression and function in human subcutaneous adipose tissue. *Pharmacology and Therapeutics* **146**, 61–93 (2015).

282. Samad, F. & Ruf, W. Inflammation, obesity, and thrombosis. *Blood* **122**, 3415–3422 (2013).

283. Badeanlou, L., Furlan-Freguia, C., Yang, G., Ruf, W. & Samad, F. Tissue factor-protease-activated receptor 2 signalling promotes diet-induced obesity and adipose inflammation. *Nature Medicine* **17**, 1490–1497 (2011).

284. Lim, J. *et al.* Diet-induced obesity, adipose inflammation, and metabolic dysfunction correlating with PAR2 expression are attenuated by PAR2 antagonism. *FASEB JOURNAL* **27**, 4757–4767 (2013).

285. Kassel, K. M. *et al.* Protease-activated receptor 1 and hematopoietic cell tissue factor are required for hepatic steatosis in mice fed a Western diet. *The American Journal of Pathology* **179**, 2278–2289 (2011).

286. Marx, V. Method of the Year: protein structure prediction. *Nature Methods* **19**, 5–10 (2022).

287. The Nobel Prize in Chemistry 2024 - NobelPrize.org. https://www.nobelprize.org/prizes/chemistry/2024/summary/.

288. Edgar, R. C. & Batzoglou, S. Multiple sequence alignment. *Current Opinions in Structural Biology* **16**, 368–373 (2006).

289. Agnihotry, S., Pathak, R. K., Singh, D. B., Tiwari, A. & Hussain, I. Protein structure prediction. *Bioinformatics: Methods and Applications* 177–188 (2022).

290. Jones, D. T., Taylort, W. R. & Thornton, J. M. A new approach to protein fold recognition. *Nature* **358**, 86–89 (1992).

291. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).

292. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).

293. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022).

294. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).

295. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications* **13**, 1–11 (2022).

296. Lee, C. Y. *et al.* Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. *Molecular System Biology* **20**, 75–97 (2024).

297. Elfmann, C. & Stülke, J. PAE viewer: a webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks. *Nucleic Acid Research* **51**, 404–410 (2023).

298. Gibbs, R. A. The Human Genome Project changed everything. *Nature Reviews Genetics* **21**, 575–576 (2020).

299. Bateman, A. *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acid Research* **51**, 523–531 (2023).

300. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

301. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 1–15 (2019).

302. Eddy, S. R. What is a hidden Markov model? *Nature Biotechnology 2004 22:10* **22**, 1315–1316 (2004).

303. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**, 1026–1028 (2017).

304. Buel, G. R. & Walters, K. J. Can AlphaFold2 predict the impact of missense mutations on structure? *Nature Structural & Molecular Biology* **29**, 1–2 (2022).

305. Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R. & Alexov, E. Molecular Mechanisms of Disease-Causing Missense Mutations. *Journal of Molecular Biology* **425**, 3919–3936 (2013).

306. Pak, M. A. *et al.* Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLoS One* **18**, 1–9 (2023).

307. Safdari, H. A., Pandey, S., Shukla, A. K. & Dutta, S. Illuminating GPCR Signalling by Cryo-EM. *Trends in Cell Biology* **28**, 591–594 (2018).

308. Caroli, J., Mamyrbekov, A. & Kermani, A. A. GPCRdb in 2023 : state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Research* **51**395–402 (2023).

309. Necci, M. *et al.* Critical assessment of protein intrinsic disorder prediction. *Nature Methods* **18**, 472–481 (2021).

310. Ruff, K. M. & Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology* **433**, 167208 (2021).

311. Amacher, J. F., Brooks, L., Hampton, T. H. & Madden, D. R. Specificity in PDZ-peptide interaction networks: Computational analysis and review. *Journal of Structural Biology X* **4**, 100022 (2020).

312. Broadbent, D. *et al.* Roles of NHERF Family of PDZ-Binding Proteins in Regulating GPCR Functions. *Advances in Immunology* **136**, 353–385 (2017).

313. Kontaxis, G., Bister, K. & Konrat, R. LIM domain proteins. *Encyclopedia of Inorganic and Bioinorganic Chemistry* **318**, 295-306 (2004).

314. Lehman, W., Craig, R., Kendrick-Jones, J. & Sutherland-Smith, A. J. An open or closed case for the conformation of calponin homology domains on F-actin? *Journal of Muscle Research and Cell Motility* **25**, 351–358 (2004).

315. Holaska, J. M., Rais-Bahrami, S. & Wilson, K. L. Lmo7 is an emerin-binding protein that regulates the transcription of emerin and many other muscle-relevant genes. *Human Molecular Genetics* **15**, 3459–3472 (2006).

316. Ott, E. B. *et al.* The lim domain only protein 7 is important in zebrafish heart development. *Developmental Dynamics* **237**, 3940–3952 (2008).

317. Karlsson, T. *et al.* LMO7 and LIMCH1 interact with LRIG proteins in lung cancer, with prognostic implications for early-stage disease. *Lung Cancer* **125**, 174–184 (2018).

318. Bauer, K. *et al.* Human CLP36, a PDZ-domain and LIM-domain protein, binds to α-actinin-1 and associates with actin filaments and stress fibers in activated platelets and endothelial cells. *Blood* **96**, 4236–4245 (2000).

319. Hasegawa, T. *et al.* CLP36 interacts with palladin in dorsal root ganglion neurons. *Neuroscience Letters* **476**, 53–57 (2010).

320. Sharma, P., Shathasivam, T., Ignatchenko, V., Kislinger, T. & Gramolini, A. O. Identification of an FHL1 protein complex containing ACTN1, ACTN4, and PDLIM1 using affinity purifications and MS-based protein-protein interaction analysis. *Molecular Biosystems* **7**, 1185–1196 (2011).

321. Naegle, K. M., White, F. M., Lauffenburger, D. A. & Yaffe, M. B. Robust co-regulation of tyrosine phosphorylation sites on proteins reveals novel protein interactions. *Molecular Biosystems* **8**, 2771 (2012).

322. Zheng, M., Cheng, H., Banerjee, I. & Chen, J. ALP/Enigma PDZ-LIM domain proteins in the heart. *Journal of Molecular Cell Biology* **2**, 96–102 (2010).

323. Te Velthuis, A. J. W. & Bagowski, C. P. PDZ and LIM Domain-Encoding Genes: Molecular Interactions and their Role in Development. *The Scientific World Journal* **7**, 1470–1492 (2007).

324. Liu, Z. *et al.* PDZ and LIM domain protein 1(PDLIM1)/CLP36 promotes breast cancer cell migration, invasion and metastasis through interaction with α-actinin. *Oncogene* **34**, 1300–1311 (2015).

325. Ahn, B. Y. *et al.* Glioma invasion mediated by the p75 neurotrophin receptor (p75(NTR)/CD271) requires regulated interaction with PDLIM1. *Oncogene* **35**, 1411–1422 (2016).

326. Chen, H. N. *et al.* PDLIM1 Stabilizes the E-Cadherin/β-Catenin Complex to Prevent Epithelial-Mesenchymal Transition and Metastatic Potential of Colorectal Cancer Cells. *Cancer Research* **76**, 1122–1134 (2016).

327. Tojkander, S., Gateva, G. & Lappalainen, P. Actin stress fibers--assembly, dynamics and biological roles. *Journal of Cell Science* **125**, 1855–1864 (2012).

328. Shang, Y. *et al.* Autophagy regulates spermatid differentiation via degradation of PDLIM1. *Autophagy* **12**, 1575–1592 (2016).

329. Ohno, K., Kato, H., Funahashi, S., Hasegawa, T. & Sato, K. Characterization of CLP36/Elfin/PDLIM1 in the nervous system. *Journal of Neurochemistry* **111**, 790–800 (2009).

330. González-Mariscal, L., Betanzos, A., Nava, P. & Jaramillo, B. E. Tight junction proteins. *Progress in Biophysics and Molecular Biology* **81**, 1–44 (2003).

331. Heinemann, U. & Schuetz, A. Structural Features of Tight-Junction Proteins. *International Journal of Molecular Sciences* **20**, 6020 (2019).

332. Bauer, H., Zweimueller-Mayer, J., Steinbacher, P., Lametschwandtner, A. & Bauer, H. C. The dual role of zonula occludens (ZO) proteins. *Journal of Biomedicine and Biotechnology* **2010**, (2010).

333. González-Mariscal, L., Betanzos, A. & Ávila-Flores, A. MAGUK proteins: structure and role in the tight junction. *Seminars in Cell & Developmental Biology* **11**, 315–324 (2000).

334. Itoh, M., Morita, K. & Tsukita, S. Characterization of ZO-2 as a MAGUK family member associated with tight as well as adherens junctions with a binding affinity to occludin and α catenin. *Journal of Biological Chemistry* **274**, 5981–5986 (1999).

335. Chen, H. *et al.* Structure of the second PDZ domain from human zonula occludens 2. *Acta Crystallographica* **65**, 327–330 (2009).

336. Katoh, M. GIPC gene family (Review). *International Journal of Molecular Medicine* **9**, 585–589 (2002).

337. Ramonett, A. *et al.* Regulation of mitochondrial fission by GIPC-mediated Drp1 retrograde transport. *Molecular Biology of the Cell* **33**, (2022).

338. Ahmed, T., Mythreye, K. & Lee, N. Y. Strength and duration of GIPC-dependent signalling networks as determinants in cancer. *Neoplasia* **23**, 181–188 (2021).

339. Giese, A. P. *et al.* Gipc1 has a dual role in Vangl2 trafficking and hair bundle integrity in the inner ear. *Development* **139**, 3775–3785 (2012).

340. de Jonge, J. J., Batters, C., O'Loughlin, T., Arden, S. D. & Buss, F. The MYO6 interactome: selective motor-cargo complexes for diverse cellular processes. *FEBS Letters* **593**, 1494–1507 (2019).

341. Charizopoulou, N. *et al.* Gipc3 mutations associated with audiogenic seizures and sensorineural hearing loss in mouse and human. *Nature Communications* **2**, 1–12 (2011).

342. De Vries, L., Lou, X., Zhao, G., Zheng, B. & Farquhar, M. G. GIPC, a PDZ domain containing protein, interacts specifically with the C terminus of RGS-GAIP. *Proceedings of the National Academy of Sciences U S A* **95**, 12340–12345 (1998).

343. Naccache, S. N., Hasson, T. & Horowitz, A. Binding of internalized receptors to the PDZ domain of GIPC/synectin recruits myosin VI to endocytic vesicles. *Proceedings of the National Academy of Sciences U S A* **103**, 12735–12740 (2006).

344. Hu, L. A. *et al.* GIPC interacts with the β1-adrenergic receptor and regulates β1-adrenergic receptor-mediated ERK activation. *Journal of Biological Chemistry* **278**, 26295–26301 (2003).

345. Varsano, T. *et al.* GIPC is recruited by APPL to peripheral TrkA endosomes and regulates TrkA trafficking and signalling. *Molecular and Cell Biology* **26**, 8942–8952 (2006).

346. Bohlson, S. S., Zhang, M., Ortiz, C. E. & Tenne, A. J. CD93 interacts with the PDZ domain-containing adaptor protein GIPC: implications in the modulation of phagocytosis. *Journal of Leukocyte Biology* **77**, 80–89 (2005).

347. Jeanneteau, F., Diaz, J., Sokoloff, P. & Griffon, N. Interactions of GIPC with Dopamine D2, D3 but not D4 Receptors Define a Novel Mode of Regulation of G Protein-coupled Receptors. *Molecular Biology of the Cell* **15**, 696 (2004).

348. Jeanneteau, F., Guillin, O., Diaz, J., Griffon, N. & Sokoloff, P. GIPC recruits GAIP (RGS19) to attenuate dopamine D2 receptor signalling. *Molecular Biology of the Cell* **15**, 4926–4937 (2004).

349. Lee, N. Y., Ray, B., How, T. & Blobe, G. C. Endoglin promotes transforming growth factor β-mediated Smad 1/5/8 signalling and inhibits endothelial cell migration through its association with GIPC. *Journal of Biological Chemistry* **283**, 32527–32533 (2008).

350. Bunn, R. C., Jensen, M. A. & Reed, B. C. Protein interactions with the glucose transporter binding protein GLUT1CBP that provide a link between GLUT1 and the cytoskeleton. *Molecular Biology of the Cell* **10**, 819–832 (1999).

351. Ligensa, T. *et al.* A PDZ domain protein interacts with the C-terminal tail of the insulin-like growth factor-1 receptor but not with the insulin receptor. *Journal of Biological Chemistry* **276**, 33419–33427 (2001).

352. Tani, T. T. & Mercurio, A. M. PDZ interaction sites in integrin alpha subunits. T14853, TIP/GIPC binds to a type I recognition sequence in alpha 6A/alpha 5 and a novel sequence in alpha 6B. *Journal of Biological Chemistry* **276**, 36535–36542 (2001).

353. Hirakawa, T., Galet, C., Kishi, M. & Ascoli, M. GIPC binds to the human lutropin receptor (hLHR) through an unusual PDZ domain binding motif, and it regulates the sorting of the internalized human choriogonadotropin and the density of cell surface hLHR. *Journal of Biological Chemistry* **278**, 49348–49357 (2003).

354. Ooshio, T. *et al.* Involvement of LMO7 in the Association of Two Cell-Cell Adhesion Molecules, Nectin and E-cadherin, through Afadin and α-Actinin in Epithelial Cells. *Journal of Biological Chemistry* **279**, 31365–31373 (2004).

355. Te Velthuis, A. J. W. & Bagowski, C. P. PDZ and LIM domain-encoding genes: molecular interactions and their role in development. *Scientific World Journal* **7**, 1470–1492 (2007).

356. Ono, R., Kaisho, T. & Tanaka, T. PDLIM1 inhibits NF-κB-mediated inflammatory signalling by sequestering the p65 subunit of NF-κB in the cytoplasm. *Scientific Reports* **5**, (2015).

357. Itoh, M., Morita, K. & Tsukita, S. Characterization of ZO-2 as a MAGUK family member associated with tight as well as adherens junctions with a binding affinity to occludin and α catenin. *Journal of Biological Chemistry* **274**, 5981–5986 (1999).

358. Ramonett, A. *et al.* Regulation of mitochondrial fission by GIPC-mediated Drp1 retrograde transport. *Molecular Biology of the Cell* **33**, (2022).

359. Ahmed, T., Mythreye, K. & Lee, N. Y. Strength and duration of GIPC-dependent signalling networks as determinants in cancer. *Neoplasia* **23**, 181–188 (2021).

360. De Vries, L., Lou, X., Zhao, G., Zheng, B. & Farquhar, M. G. GIPC, a PDZ domain containing protein, interacts specifically with the C terminus of RGS-GAIP. *Proceedings of the National Academy of Sciences U S A* **95**, 12340–12345 (1998).

361. Naccache, S. N., Hasson, T. & Horowitz, A. Binding of internalized receptors to the PDZ domain of GIPC/synectin recruits myosin VI to endocytic vesicles. *Proceedings of the National Academy of Sciences U S A* **103**, 12735–12740 (2006).

362. Hu, L. A. *et al.* GIPC interacts with the beta1-adrenergic receptor and regulates beta1-adrenergic receptor-mediated ERK activation. *Journal of Biological Chemistry* **278**, 26295–26301 (2003).

363. Lee, N. Y., Ray, B., How, T. & Blobe, G. C. Endoglin promotes transforming growth factor beta-mediated Smad 1/5/8 signalling and inhibits endothelial cell migration through its association with GIPC. *Journal of Biological Chemistry* **283**, 32527–32533 (2008).

364. Tsao, C. W. *et al.* Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. *Circulation* **145**, 153-639 (2022).

365. Lee, H. J. & Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Communication and Signalling* **8**, 1–18 (2010).

366. te Velthuis, A. J. W., Isogai, T., Gerrits, L. & Bagowski, C. P. Insights into the Molecular Evolution of the PDZ/LIM Family and Identification of a Novel Conserved Protein Motif. *PLoS One* **2**, 189 (2007).

367. Appleton, B. A. *et al.* Comparative Structural Analysis of the Erbin PDZ Domain and the First PDZ Domain of ZO-1. *Journal of Biological Chemistry* **281**, 22312–22320 (2006).

368. Hiroaki, H. *et al.* Spatial Overlap of Claudin- and Phosphatidylinositol Phosphate-Binding Sites on the First PDZ Domain of Zonula Occludens 1 Studied by NMR. *Molecules* **23**, (2018).

369. Wu, J. *et al.* Domain-swapped dimerization of the second PDZ domain of ZO2 may provide a structural basis for the polymerization of claudins. *Journal of Biological Chemistry* **282**, 35988–35999 (2007).

370. Ernst, A. *et al.* A Structural Portrait of the PDZ Domain Family. *Journal of Molecular Biology* **426**, 3509–3519 (2014).

371. Varsano, T., Taupin, V., Guo, L., Baterina, O. Y. & Farquhar, M. G. The PDZ Protein GIPC Regulates Trafficking of the LPA1 Receptor from APPL Signalling Endosomes and Attenuates the Cell's Response to LPA. *PLoS One* **7**, 49227 (2012).

372. Tso, P. H. *et al.* RGS19 enhances cell proliferation through its C-terminal PDZ motif. *Cell Signal* **22**, 1700–1707 (2010).

373. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signalling complex assembly. *Journal of Cell Science* **114**, 3219–3231 (2001).

374. Lou, X., Yano, H., Lee, F., Chao, M. V. & Farquhar, M. G. GIPC and GAIP form a complex with TrkA: a putative link between G protein and receptor tyrosine kinase pathways. *Molecular Biology of the Cell* **12**, 615–627 (2001).

375. Sheng, M. & Sala, C. PDZ domains and the organization of supramolecular complexes. *Annual Reviews in Neuroscience* **24**, 1–29 (2001).

376. Amacher, J. F., Brooks, L., Hampton, T. H. & Madden, D. R. Specificity in PDZ-peptide interaction networks: Computational analysis and review. *Journal of Structural Biology X* **4**, 100022 (2020).

377. Lin, J. S. & Lai, E. M. Protein–protein interactions: Co-immunoprecipitation. *Methods in Molecular Biology* **1615**, 211–219 (2017).

378. Alam, M. S. Proximity Ligation Assay (PLA). *Current Protocols in Immunology* **123**, e58 (2018).

379. Awan, A. *et al.* 5T4 interacts with TIP-2/GIPC, a PDZ protein, with implications for metastasis. *Biochemical and Biophysical Research Communications* **290**, 1030–1036 (2002).

380. Booth, R. A., Cummings, C., Tiberi, M. & Johné Liu, X. GIPC participates in G protein signalling downstream of insulin-like growth factor 1 receptor. *Journal of Biological Chemistry* **277**, 6719–6725 (2002).

381. Reed, B. C. *et al.* GLUT1CBP(TIP2/GIPC1) interactions with GLUT1 and myosin VI: evidence supporting an adapter function for GLUT1CBP. *Molecular Biology of the Cell* **16**, 4183–4201 (2005).

382. Blobe, G. C., Liu, X., Fang, S. J., How, T. & Lodish, H. F. A novel mechanism for regulating transforming growth factor beta (TGF-beta) signalling. Functional modulation of type III TGF-beta receptor expression through interaction with the PDZ domain protein, GIPC. *Journal of Biological Chemistry* **276**, 39608–39617 (2001).

383. Huang, C., Hepler, J. R., Gilman, A. G. & Mumby, S. M. Attenuation of Gi- and Gq-mediated signalling by expression of RGS4 or GAIP in mammalian cells. *Proceedings of the National Academy of Sciences* **94**, 6159–6163 (1997).

384. Petrey, D., Zhao, H., Trudeau, S. J., Murray, D. & Honig, B. PrePPI: A Structure Informed Proteome-wide Database of Protein–Protein Interactions. *Journal of Molecular Biology* **435**, 168052 (2023).

385. Daaka, Y. S-nitrosylation-regulated GPCR signalling. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 743–751 (2012).

386. Kayki-Mutlu, G. & Koch, W. J. Nitric Oxide and S-Nitrosylation in Cardiac Regulation: G Protein-Coupled Receptor Kinase-2 and β-Arrestins as Targets. *International Journal of Molecular Sciences* **22**, 521 (2021).

387. Hayashi, H. *et al.* S-Nitrosylation of β-Arrestins Biases Receptor Signalling and Confers Ligand Independence. *Molecular Cell* **70**, 473–487 (2018).

388. Kimura, T. E. *et al.* Inhibition of Egr1 expression underlies the anti-mitogenic effects of cAMP in vascular smooth muscle cells. *Journal of Molecular and Cellular Cardiology* **72**, 9–19 (2014).

389. Ardura, J. A. & Friedman, P. A. Regulation of G Protein-Coupled Receptor Function by Na+/H+ Exchange Regulatory Factors. *Pharmacological Reviews* **63**, 882–900 (2011).

390. He, J. *et al.* Proteomic analysis of β1-adrenergic receptor interactions with PDZ scaffold proteins. *Journal of Biological Chemistry* **281**, 2820–2827 (2006).

391. Kim, J. K. *et al.* PDZ Domain-containing 1 (PDZK1) Protein Regulates Phospholipase C-β3 (PLC-β3)-specific Activation of Somatostatin by Forming a Ternary Complex with PLC-β3 and Somatostatin Receptors. *Journal of Biological Chemistry* **287**, 21012–21024 (2012).

392. Walther, C., Caetano, F. A., Dunn, H. A. & Ferguson, S. S. G. PDZK1/NHERF3 Differentially Regulates Corticotropin-releasing Factor Receptor 1 and Serotonin 2A Receptor Signalling and Endocytosis. *Cell Signal* **27**, 519–531 (2015).

393. Gupta, S., Abd-Elrahman, K. S., Albaker, A., Dunn, H. A. & Ferguson, S. S. G. Structural determinants governing β-arrestin2 interaction with PDZ proteins and recruitment to CRFR1. *Cell Signal* **63**, 109361 (2019).

394. Fairbank, P. D. *et al.* Shroom2 (APXL) regulates melanosome biogenesis and localization in the retinal pigment epithelium. *Development* **133**, 4109–4118 (2006).

395. Palmisano, I. *et al.* The ocular albinism type 1 protein, an intracellular G protein-coupled receptor, regulates melanosome transport in pigment cells. *Human Molecular Genetics* **17**, 3487–3501 (2008).

396. Scherer, S. S., Xu, Y. T., Bannerman, P. G. C., Sherman, D. L. & Brophy, P. J. Periaxin expression in myelinating Schwann cells: modulation by axon-glial interactions and polarized localization during development. *Development* **121**, 4265–4273 (1995).

397. Shang, G. *et al.* Structure analyses reveal a regulated oligomerization mechanism of the PlexinD1/GIPC/myosin VI complex. *Elife* **6**, (2017).

398. Kunselman, J. M., Lott, J. & Puthenveedu, M. A. Mechanisms of selective G protein–coupled receptor localization and trafficking. *Current Opinions in Cell Biology* **71**, 158–165 (2021).

399. Dunn, H. A. & Ferguson, S. S. G. PDZ Protein Regulation of G Protein-Coupled Receptor Trafficking and Signalling Pathways. *Molecular Pharmacology* **88**, 624–629 (2015).

400. Lou, X., Yano, H., Lee, F., Chao, M. V. & Farquhar, M. G. GIPC and GAIP form a complex with TrkA: A putative link between G protein and receptor tyrosine kinase pathways. *Molecular Biology of the Cell* **12**, 615–627 (2001).

401. Lou, X., McQuistan, T., Orlando, R. A. & Farquhar, M. G. GAIP, GIPC and Gαi3 are Concentrated in Endocytic Compartments of Proximal Tubule Cells. *Journal of the American Society of Nephrology* **13**, 918–927 (2002).

402. Valdembri, D. *et al.* Neuropilin-1/GIPC1 Signalling Regulates α5β1 Integrin Traffic and Function in Endothelial Cells. *PLoS Biol* **7**, e1000025 (2009).

403. Yi, Z. *et al.* The Role of the PDZ Protein GIPC in Regulating NMDA Receptor Trafficking. *Journal of Neuroscience* **27**, 11663–11675 (2007).

404. Choi, J. S., Paek, A. R., Kim, S. Y. & You, H. J. GIPC mediates the generation of reactive oxygen species and the regulation of cancer cell

proliferation by insulin-like growth factor-1/IGF-1R signalling. *Cancer Letters* **294**, 254–263 (2010).

405. Muders, M. H. *et al.* Targeting GIPC/synectin in pancreatic cancer inhibits tumor growth. *Clinical Cancer Research* **15**, 4095–4103 (2009).

406. Sebbagh, M. & Borg, J. P. Insight into planar cell polarity. *Experimental Cell Research* **328**, 284–295 (2014).

407. Wang, L. *et al.* C terminus of RGS-GAIP-interacting protein conveys neuropilin-1-mediated signalling during angiogenesis. *The FASEB Journal* **20**, 1513–1515 (2006).

408. Davidson, M. W. & Abramowitz, M. Optical Microscopy. *Polymer Science: A Comprehensive Reference, 10 Volume Set* **2**, 465–478 (2012).

409. Jacquemet, G., Carisey, A. F., Hamidi, H., Henriques, R. & Leterrier, C. The cell biologist's guide to super-resolution microscopy. *Journal of Cell Science* **133**, (2020).

410. Van Dijk, M. A., Lippitz, M. & Orrit, M. Far-field optical microscopy of single metal nanoparticles. *Accounts of Chemical Research* **38**, 594–601 (2005).

411. Balzarotti, F. & Stefani, F. D. Plasmonics meets far-field optical nanoscopy. *ACS Nano* **6**, 4580–4584 (2012).

412. Chang, J. & Wetzstein, G. Single-shot speckle correlation fluorescence microscopy in thick scattering tissue with image reconstruction priors. *Journal of Biophotonics* **11**, e201700224 (2018).

413. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).

414. Schmidt, D. A., Kopf, I. & Bründermann, E. A matter of scale: from far-field microscopy to near-field nanoscopy. *Laser Photon Reviews* **6**, 296–332 (2012).

415. Smith, K. C. A. & Oatley, C. W. The scanning electron microscope and its fields of application. *British Journal of Applied Physics* **6**, 391 (1955).

416. Lakowicz, J. R. Radiative Decay Engineering: Biophysical and Biomedical Applications. *Analytical Biochemistry* **298**, 1–24 (2001).

417. Carminati, R., Greffet, J. J., Henkel, C. & Vigoureux, J. M. Radiative and non-radiative decay of a single molecule close to a metallic nanoparticle. *Optics Communications* **261**, 368–375 (2006).

418. Valeur, B. & Berberan-Santos, M. N. A brief history of fluorescence and phosphorescence before the emergence of quantum theory. *Journal of Chemical Education* **88**, 731–738 (2011).

419. Stokes, G. G. XXX. On the change of refrangibility of light. *Philosophical Transactions of the Royal Society* **142**, 463–562 (1852).

420. Jabłoństski, A. Efficiency of Anti-Stokes Fluorescence in Dyes. *Nature* **131**, 839–840 (1933).

421. Frackowiak, D. The Jablonski diagram. *Journal of Photochemical Photobiology B* **2**, 399 (1988).

422. Renz, M. Fluorescence microscopy—A historical and technical perspective. *Cytometry Part A* **83**, 767–779 (2013).

423. Jelley, E. E. Spectral Absorption and Fluorescence of Dyes in the Molecular State. *Nature* **138**, 1009–1010 (1936).

424. Coons, A. H., Creech, H. J., Jones, R. N. & Berliner, E. The Demonstration of Pneumococcal Antigen in Tissues by the Use of Fluorescent Antibody. *The Journal of Immunology* **45**, 159–170 (1942).

425. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111**, 229–233 (1992).

426. Lichtman, J. W. & Conchello, J. A. Fluorescence microscopy. *Nature Methods* **2**, 910–919 (2005).

427. Herce, H. D., Casas-Delucchi, C. S. & Cardoso, M. C. New image colocalization coefficient for fluorescence microscopy to quantify (bio-)molecular interactions. *Journal of Microscopy* **249**, 184–194 (2013).

428. Galdeen, S. A. & North, A. J. Live cell fluorescence microscopy techniques. *Methods in Molecular Biology* **769**, 205–222 (2011).

429. Icha, J., Weber, M., Waters, J. C. & Norden, C. Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* **39**, 1700003 (2017).

430. Diaspro, A., Chirico, G., Usai, C., Ramoino, P. & Dobrucki, J. Photobleaching. *Handbook of Biological Confocal Microscopy* **3**, 690–702 (2006).

431. Schermelleh, L., Heintzmann, R. & Leonhardt, H. A guide to super-resolution fluorescence microscopy. *Journal of Cell Biology* **190**, 165–175 (2010).

432. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung: I. Die Construction von Mikroskopen auf Grund der Theorie. *Archiv für mikroskopische Anatomie* **9**, 413–418 (1873).

433. Rossmann, K. Point spread-function, line spread-function, and modulation transfer function. Tools for the study of imaging systems. *Radiology* **93**, 257–272 (1969).

434. Sanderson, M. J., Smith, I., Parker, I. & Bootman, M. D. Fluorescence Microscopy. *Cold Spring Harb Protocols* **2014**, pdb.top071795 (2014).

435. Elliot, A. D. Confocal Microscopy: Principles and Modern Practices. *Current Protocols in Cytometry* **92**, 1–11 (2020).

436. Hislop, J. N. & Von Zastrow, M. Analysis of GPCR localization and trafficking. *Methods in Molecular Biology* **746**, 425–440 (2011).

437. White, J. G., Amos, W. B. & Fordham, M. An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy. *Journal of Cell Biology* **105**, 41–48 (1987).

438. Borlinghaus, R. T. The White Confocal. *The White Confocal* (2017)

439. Naora, H. Microspectrophotometry and cytochemical analysis of nucleic acids. *Science (1979)* **114**, 279–280 (1951).

440. Minsky, M. Memoir on inventing the confocal scanning microscope. *Scanning* **10**, 128–138 (1988).

441. David Egger, M. & Petran, M. New Reflected-Light Microscope for Viewing Unstained Brain and Ganglion Cells. *Science* **157**, 305–307 (1967).

442. Davidovits, P. & Egger, M. D. Scanning Laser Microscope. *Nature* **223**, 831–831 (1969).

443. Cremer, C. & Cremer, T. Considerations on a laser-scanning-microscope with high resolution and depth of field. *Microscopica Acta* **81**, 1–600 (1978).

444. Inoué, S. 'Chapter 1: Foundations of Confocal Scanned Imaging in Light Microscopy'. In James Pawley (ed.). Handbook of Biological Confocal Microscopy. *Springer Science and Business Media LLC.* 232 (2006).

445. Denis Semwogerere, E. R. W. Confocal Microscopy. *Encyclopedia of Biomaterials and Biomedical Engineering* **1**, 737–746 (2008).

446. Nooh, M. M. & Bahouth, S. W. Visualization and quantification of GPCR trafficking in mammalian cells by confocal microscopy. *Methods in Cellular Biol*ogy**142**, 67–78 (2017).

447. Castillo-Badillo, J. A., Cabrera-Wrooman, A. & García-Sáinz, J. A. Visualizing G Protein-coupled Receptors in Action through Confocal Microscopy Techniques. *Archives of Medical Research* **45**, 283–293 (2014).

448. Heilker, R., Zemanova, L., Valler, M. J. & Nienhaus, G. U. Confocal Fluorescence Microscopy for High-Throughput Screening of G-Protein Coupled Receptors. *Current Medicinal Chemistry* **12**, 2551–2559 (2005).

449. Garippa, R. J., Hoffman, A. F., Gradl, G. & Kirsch, A. High-Throughput Confocal Microscopy for β-Arrestin–Green Fluorescent Protein Translocation G Protein-Coupled Receptor Assays Using the Evotec Opera. *Methods Enzymology* **414**, 99–120 (2006).

450. US5394268A - Field synthesis and optical subsectioning for standing wave microscopy - Google Patents. https://patents.google.com/patent/US5394268.

451. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letterst* **19**, 780 (1994).

452. Klar, T. A. & Hell, S. W. Subdiffraction resolution in far-field fluorescence microscopy. *Optics Letters, Vol. 24, Issue 14, pp. 954-956* **24**, 954–956 (1999).

453. The Nobel Prize in Chemistry 2014. https://www.nobelprize.org/prizes/chemistry/2014/summary/.

454. Müller, T., Schumann, C. & Kraegeloh, A. STED Microscopy and its Applications: New Insights into Cellular Processes on the Nanoscale. *ChemPhysChem* **13**, 1986–2000 (2012).

455. Dyba, M. & Hell, S. W. Photostability of a fluorescent marker under pulsed excited-state depletion through stimulated emission. *Applied Optics***42**, 5123–5129 (2003).

456. Vicidomini, G., Bianchini, P. & Diaspro, A. STED super-resolved microscopy. *Nature Methods* **15**, 173–182 (2018).

457. Bottanelli, F. *et al.* Two-colour live-cell nanoscale imaging of intracellular targets. *Nature Communications* **7**, 1–5 (2016).

458. Göttfert, F. *et al.* Coaligned dual-channel STED nanoscopy and molecular diffusion analysis at 20 nm resolution. *Biophysical Journal* **105**, L01–L03 (2013).

459. Li, H. *et al.* Organized cannabinoid receptor distribution in neurons revealed by super-resolution fluorescence imaging. *Nature Communications* **11**, (2020).

460. Ast, J. *et al.* Super-resolution microscopy compatible fluorescent probes reveal endogenous glucagon-like peptide-1 receptor distribution and dynamics. *Nature Communications* **11**, (2020).

461. Fessl, T., Majellaro, M. & Bondar, A. Microscopy and spectroscopy approaches to study GPCR structure and function. *British Journal of Pharmacology* **12** (2023)

462. Saxena, M., Eluru, G. & Gorthi, S. S. Structured illumination microscopy. *Advances in Optics and Photonics* **7**, 241–275 (2015).

463. Lipson, S. G. Is There a Fundamental Limit to Spatial Resolution in Phase Microscopy? *Biomedical Optical Phase Microscopy and Nanoscopy* 341–351 (2013)

464. Bailey, B., Farkas, D. L., Taylor, D. L. & Lanni, F. Enhancement of axial resolution in fluorescence microscopy by standing-wave excitation. *Nature* **366**, 44–48 (1993).

465. Stelzer, E. H. K. & Hell, S. Properties of a 4Pi confocal fluorescence microscope. *JOSA A***9**, 2159–2166 (1992).

466. Bewersdorf, J., Schmidt, R. & Hell, S. W. Comparison of I5M and 4Pi-microscopy. *Journal of Microscopy* **222**, 105–117 (2006).

467. Wilson, T., Juškaitis, R. & Neil, M. A. A. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optics Letters,* **22**, 1905–1907 (1997).

468. Heintzmann, R., Cremer Rainer Heintzmann, C. & Cremer, C. G. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. *Optical Biopsies and Microscopic Techniques* **3568**, 185–196 (1999).

469. Frohn, J. T., Knapp, H. F. & Stemmer, A. True optical resolution beyond the Rayleigh limit achieved by standing wave illumination. *Proceedings of the National Academy of Sciences* **97**, 7232–7236 (2000).

470. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy* **198**, 82–87 (2000).

471. Gustafsson, M. G. L. *et al.* Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical Journal* **94**, 4957–4970 (2008).

472. Lal, A., Shan, C. & Xi, P. Structured illumination microscopy image reconstruction algorithm. *IEEE Journal of Selected Topics in Quantum Electronics* **22**, 50–63 (2016).

473. Allen, J. R., Ross, S. T. & Davidson, M. W. Structured Illumination Microscopy for Superresolution. *ChemPhysChem* **15**, 566–576 (2014).

474. Eichel, K., Jullié, D. & Von Zastrow, M. β-Arrestin drives MAP kinase signalling from clathrin-coated structures after GPCR dissociation. *Nature Cell Biology* **18**, 303–310 (2016).

475. Shen, A. *et al.* Functionally distinct and selectively phosphorylated GPCR subpopulations co-exist in a single cell. *Nature Communications* **9**, (2018).

476. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology 2002* **20**, 473–477 (2002).

477. Paek, J. *et al.* Multidimensional Tracking of GPCR Signalling via Peroxidase-Catalyzed Proximity Labeling In Brief Using APEX proximity labeling to monitor GPCR signalling provides both spatial and temporal insight into receptor signalling and internalization in response to both balanced and biased GPCR ligands. *Cell* **169**, 338-349 (2017).

478. Borroto-Escuela, D. O. *et al.* In situ proximity ligation assay to study and understand the distribution and balance of GPCR homo- and heteroreceptor complexes in the brain. *Neuromethods* **110**, 109–124 (2016).

479. Zhuo, Y., Robleto, V. L. & Marchese, A. Proximity Labeling to Identify β-Arrestin1 Binding Partners Downstream of Ligand-Activated G Protein-Coupled Receptors. *International Journal of Molecular Science* **24**, 3285 (2023).

480. Käser, S. *et al.* Outer membrane protein functions as integrator of protein import and DNA inheritance in mitochondria. *Proceedings of the National Academy of Sciences U S A* **113**, E4467–E4475 (2016).

481. Ishihara, N. & Mihara, K. Identification of the Protein Import Components of the Rat Mitochondrial Inner Membrane, rTIM17, rTIM23, and rTIM44. *The Journal of Biochemistry* **123**, 722–732 (1998).

482. Massa, R. *et al.* Intracellular localization and isoform expression of the voltage-dependent anion channel (VDAC) in normal and dystrophic skeletal muscle. *Journal of Muscle Research and Cell Motility* **21**, 433–442 (2000).

483. Clutton, G., Mollan, K., Hudgens, M. & Goonetilleke, N. A Reproducible, Objective Method Using MitoTracker® Fluorescent Dyes to Assess Mitochondrial Mass in T Cells by Flow Cytometry. *Cytometry* **95**, 450 (2019).

484. Samanta, S. *et al.* Fluorescent Probes for Nanoscopic Imaging of Mitochondria. *Chemistry* **5**, 1697–1726 (2019).

485. Tinning, P. *et al.* Miniaturized structured illumination microscopy using two 3-axis MEMS micromirrors. *Biomedical Optics Express,* **13**, 6443–6456 (2022).

486. Dan, L. *et al.* Comparison Between Different Mitochondrial Staining Methods in Cell and Tissue Samples.*Journal of Sichuan University,* **51**, 388–392 (2020).

487. Tønnesen, J., Nadrigny, F., Willig, K. I., Wedlich-Söldner, R. & Nägerl, U. V. Two-Color STED Microscopy of Living Synapses Using A Single Laser-Beam Pair. *Biophysical Journal* **101**, 2545–2552 (2011).

488. Herman, B. Fluorescence Microscopy. *Current Protocols in Cell Biology* **00**, 4.2.1-4.2.10 (1998).

489. Singh, H. *et al.* Visualization and quantification of cardiac mitochondrial protein clusters with STED microscopy. *Mitochondrion* **12**, 230–236 (2012).

490. Jans, D. C. *et al.* STED super-resolution microscopy reveals an array of MINOS clusters along human mitochondria. *Proceedings of the National Academy of Sciences U S A* **110**, 8936–8941 (2013).

491. Ishigaki, M. *et al.* STED super-resolution imaging of mitochondria labeled with TMRM in living cells. *Mitochondrion* **28**, 79–87 (2016).

492. Stephan, T., Roesch, A., Riedel, D. & Jakobs, S. Live-cell STED nanoscopy of mitochondrial cristae. *Scientific Reports 2019 9:1* **9**, 1–6 (2019).

493. Yang, X. *et al.* Mitochondrial dynamics quantitatively revealed by STED nanoscopy with an enhanced squaraine variant probe. *Nature Communications* **11**, 1–9 (2020).

494. Ren, W. *et al.* Visualization of cristae and mtDNA interactions via STED nanoscopy using a low saturation power probe. *Light: Science & Applications* **13**, 1–16 (2024).

495. Opstad, I. S., Wolfson, D. L., Øie, C. I. & Ahluwalia, B. S. Multi-color imaging of sub-mitochondrial structures in living cells using structured illumination microscopy. *Nanophotonics* **7**, 935–947 (2018).

496. Chen, Q. *et al.* Quantitative analysis of interactive behavior of mitochondria and lysosomes using structured illumination microscopy. *Biomaterials* **250**, 120059 (2020).

497. Opstad, I. S. *et al.* Three-dimensional structured illumination microscopy data of mitochondria and lysosomes in cardiomyoblasts under normal and galactose-adapted conditions. *Scientific Data 2022* **9**, 1–7 (2022).

498. Opstad, I. S. *et al.* Mitochondrial dynamics and quantification of mitochondria-derived vesicles in cardiomyoblasts using structured illumination microscopy. *Journal of Biophotonics* **15**, 202100305 (2022).

499. He, T. *et al.* Super-Resolution Structured Illumination Microscopy for the Visualization of Interactions between Mitochondria and Lipid Droplets. *Photonics* **10**, 313 (2023).

500. Kim, D. I. *et al.* Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences U S A* **111**, 2453–2461 (2014).

501. Nguyen, T. M. T., Kim, J., Doan, T. T., Lee, M. W. & Lee, M. APEX Proximity Labeling as a Versatile Tool for Biological Research. *Biochemistry* **59**, 260–269 (2020).

502. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nature Protocols* **12**, 1198–1228 (2017).

503. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature Methods* **11**, 313–318 (2014).

504. Unterauer, E. M. & Jungmann, R. Quantitative Imaging With DNA-PAINT for Applications in Synaptic Neuroscience. *Frontiers in Synaptic Neuroscience* **13**, 71 (2022).

505. Joseph, M. D., Bort, E. T., Grose, R. P., McCormick, P. J. & Simoncelli, S. Quantitative Super-Resolution Imaging for the Analysis of GPCR Oligomerization. *Biomolecules* **11**, (2021).

506. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022).

507. Ishihama, Y. *et al.* Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nature Biotechnology* **23**, 617–621 (2005).

508. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* **7**, 383–385 (2010).

509. Thompson, A. *et al.* Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry* **75**, 1895–1904 (2003).

510. McClatchy, D. B., Dong, M. Q., Wu, C. C., Venable, J. D. & Yates, J. R. 15N metabolic labeling of mammalian tissue with slow protein turnover. *Journal ofProteome Research* **6**, 2005–2010 (2007).

511. Ong, S. E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nature Methods* **1**, 119–126 (2004).

512. Fang, T., Szklarczyk, D., Hachilif, R. & von Mering, C. Enhancing coevolutionary signals in protein–protein interaction prediction through clade-wise alignment integration. *Scientific Reports* **14**, 1–17 (2024).

513. Bartolec, T. K. *et al.* Cross-linking mass spectrometry discovers, evaluates, and corroborates structures and protein–protein interactions in the human cell. *Proceedings of the National Academy of Sciences U S A* **120**, e2219418120 (2023).

514. O'Reilly, F. J. *et al.* Protein complexes in cells by AI -assisted structural proteomics . *Molecular System Biology* **19**, 11544 (2023).

515. Trepte, P. *et al.* AI-guided pipeline for protein–protein interaction drug discovery identifies a SARS-CoV-2 inhibitor. *Molecular System Biology* **20**, 428–457 (2024).

516. Dürr, S. L. & Rothlisberger, U. Predicting metal-protein interactions using cofolding methods: Status quo. *bioRxiv* 2024.05.28.596236 (2024).

517. Wu, D. *et al.* Structural characterization and AlphaFold modeling of human T cell receptor recognition of NRAS cancer neoantigens. *bioRxiv* 2024.05.21.595215 (2024).

518. Masters, M. R., Amr, Mahmoud, H. & Lill, M. A. Do Deep Learning Models for Co-Folding Learn the Physics of Protein-Ligand Interactions? *bioRxiv* 2024.06.03.597219 (2024) .

519. Lu, W., Zhang, J., Rao, J., Zhang, Z. & Zheng, S. AlphaFold3, a secret sauce for predicting mutational effects on protein-protein interactions. *bioRxiv* 2024.05.25.595871 (2024).

520. Kim, Y. *et al.* Bitter taste receptor activation by cholesterol and an intracellular tastant. *Nature* **628**, 664–671 (2024).

521. Pérez-Jover, I. *et al.* Allosteric control of dynamin-related protein 1 through a disordered C-terminal Short Linear Motif. *Nature Communications* **15**, 1–17 (2024).

522. Fasciani, I. *et al.* The C-terminus of the prototypical M2 muscarinic receptor localizes to the mitochondria and regulates cell respiration under stress conditions. *PLoS Biol* **22**, e3002582 (2024).

523. Liu, Y. *et al.* GIPC1 remodels lipid metabolism by transporting DECR1 into mitochondria against ferroptosis during dilated cardiomyopathy (DCM). (2024).

524. Fender, A. C. *et al.* Thrombin receptor PAR4 drives canonical NLRP3 inflammasome signalling in the heart. *Basic Research in Cardiology* **115**, 3 (2020).

525. Deshwal, S., Fiedler, K. U., Langer, T., Langer, T. & Langer, T. Mitochondrial Proteases: Multifaceted Regulators of Mitochondrial Plasticity. *Annual Revies in Biochemistry***89**, 501–528 (2020).

526. Gomez-Fabra Gala, M. & Vögtle, F. N. Mitochondrial proteases in human diseases. *FEBS Letters* **595**, 1205–1222 (2021).

527. Vande Walle, L., Lamkanfi, M. & Vandenabeele, P. The mitochondrial serine protease HtrA2/Omi: an overview. *Cell Death & Differentiation 2008* **15**, 453–460 (2008).

528. Li, W. *et al.* Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. *Nature Structural Biology* **9**, 436–441 (2002).

529. Strauss, K. M. *et al.* Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Human Molecular Genetics* **14**, 2099–2111 (2005).

530. Giamogante, F., Barazzuol, L., Brini, M. & Calì, T. ER-Mitochondria Contact Sites Reporters: Strengths and Weaknesses of the Available Approaches. *International Journal of Molecular Science* **21**, 1–18 (2020).

531. Benhammouda, S., Vishwakarma, A., Gatti, P. & Germain, M. Mitochondria Endoplasmic Reticulum Contact Sites (MERCs): Proximity Ligation Assay as a Tool to Study Organelle Interaction. *Frontiers in Cell Devevlopment Biology* **9**, 3489 (2021).

532. Ching, J., Osborne, A., Eva, R., Prudent, J. & Yu-Wai-Man, P. Quantifying inter-organelle membrane contact sites using proximity ligation assay in fixed optic nerve sections. *Expert Eye Research* **213**, 108793 (2021).

533. Dunn, H. A. & Ferguson, S. S. G. PDZ Protein Regulation of G Protein-Coupled Receptor Trafficking and Signalling Pathways. *Molecular Pharmacology* **88**, 624–629 (2015).

534. Thurner, P. *et al.* A Two-state Model for the Diffusion of the A2A Adenosine Receptor in Hippocampal Neurons: Agonist induced switch to slow mobility is modified by Synapse-Associated Protein 102 (SAP102)*. *Journal of Biological Chemistry* **289**, 9263 (2014).

535. Lyssand, J. S., Lee, K. S., DeFino, M., Adams, M. E. & Hague, C. Syntrophin isoforms play specific functional roles in the α1D-adrenergic receptor/DAPC signalosome. *Biochemical and Biophysical Research Communications* **412**, 596–601 (2011).

536. Lyssand, J. S. *et al.* Blood Pressure Is Regulated by an α1D-Adrenergic Receptor/Dystrophin Signalosome. *Journal of Biological Chemistry* **283**, 18792 (2008).

537. Wang, Q. & Limbird, L. E. Regulated interactions of the alpha 2A adrenergic receptor with spinophilin, 14-3-3zeta, and arrestin 3. *Journal of Biological Chemistry* **277**, 50589–50596 (2002).

538. Lu, R. *et al.* Enhanced hypotensive, bradycardic, and hypnotic responses to alpha2-adrenergic agonists in spinophilin-null mice are accompanied by increased G protein coupling to the alpha2A-adrenergic receptor. *Molecular Pharmacology* **78**, 279–286 (2010).

539. He, J. *et al.* Proteomic analysis of beta1-adrenergic receptor interactions with PDZ scaffold proteins. *Journal of Biological Chemistry* **281**, 2820–2827 (2006).

540. Gardner, L. A., Naren, A. P. & Bahouth, S. W. Assembly of an SAP97-AKAP79-cAMP-dependent protein kinase scaffold at the type 1 PSD-95/DLG/ZO1 motif of the human beta(1)-adrenergic receptor generates a receptosome involved in receptor recycling and networking. *Journal of Biological Chemistry* **282**, 5085–5099 (2007).

541. Hall, R. A. *et al.* The beta2-adrenergic receptor interacts with the Na+/H+-exchanger regulatory factor to control Na+/H+ exchange. *Nature* **392**, 626–630 (1998).

542. Lauffer, B. E. L. *et al.* SNX27 mediates PDZ-directed sorting from endosomes to the plasma membrane. *Journal of Cellular Biology* **190**, 565–574 (2010).

543. Chakir, K. *et al.* Galphas-biased beta2-adrenergic receptor signalling from restoring synchronous contraction in the failing heart. *Science Translational Medicine* **3**, (2011).

544. Booden, M. A., Siderovski, D. P. & Der, C. J. Leukemia-associated Rho guanine nucleotide exchange factor promotes G alpha q-coupled activation of RhoA. *Molecular and Cell Biology* **22**, 4053–4061 (2002).

545. Ying, Z., Jin, L., Palmer, T. & Webb, R. C. Angiotensin II up-regulates the leukemia-associated Rho guanine nucleotide exchange factor (RhoGEF), a regulator of G protein signalling domain-containing RhoGEF, in vascular smooth muscle cells. *Molecular Pharmacology* **69**, 932–940 (2006).

546. Chiu, W. C. *et al.* Angiotensin II regulates the LARG/RhoA/MYPT1 axis in rat vascular smooth muscle in vitro. *Acta Pharmacologica Sinica* **33**, 1502–1510 (2012).

547. Choi, J. W. *et al.* Subtype-specific role of phospholipase C-beta in bradykinin and LPA signalling through differential binding of different PDZ scaffold proteins. *Cell Signal* **22**, 1153–1161 (2010).

548. Sánchez-Blázquez, P., Rodríguez-Muñoz, M., Bailón, C. & Garzón, J. GPCRs promote the release of zinc ions mediated by nNOS/NO and the redox transducer RGSZ2 protein. *Antioxidant Redox Signal* **17**, 1163–1177 (2012).

549. Hammad, M. M., Kuang, Y. Q., Yan, R., Allen, H. & Dupre, D. J. Na+/H+ exchanger regulatory factor-1 is involved in chemokine receptor homodimer CCR5 internalization and signal transduction but does not affect CXCR4

homodimer or CXCR4-CCR5 heterodimer. *Journal of Biological Chemistry* **285**, 34653–34664 (2010).

550. Kuang, Y. Q., Pang, W., Zheng, Y. T. & Dupré, D. J. NHERF1 regulates gp120-induced internalization and signalling by CCR5, and HIV-1 production. *European Journal of Immunology* **42**, 299–310 (2012).

551. Walther, C., Caetano, F. A., Dunn, H. A. & Ferguson, S. S. G. PDZK1/NHERF3 Differentially Regulates Corticotropin-releasing Factor Receptor 1 and Serotonin 2A Receptor Signalling and Endocytosis. *Cell Signal* **27**, 519–531 (2015).

552. Dunn, H. A., Walther, C., Godin, C. M., Hall, R. A. & Ferguson, S. S. G. Role of SAP97 Protein in the Regulation of Corticotropin-releasing Factor Receptor 1 Endocytosis and Extracellular Signal-regulated Kinase 1/2 Signalling. *Journal of Biological Chemistry* **288**, 15023 (2013).

553. Zhang, J. *et al.* Inhibition of the dopamine D1 receptor signalling by PSD-95. *Journal of Biological Chemistry* **282**, 15778–15789 (2007).

554. Sun, P. *et al.* PSD-95 regulates D1 dopamine receptor resensitization, but not receptor-mediated Gs-protein activation. *Cell Research* **19**, 612 (2009).

555. Jeanneteau, F., Diaz, J., Sokoloff, P. & Griffon, N. Interactions of GIPC with Dopamine D2, D3 but not D4 Receptors Define a Novel Mode of Regulation of G Protein-coupled Receptors. *Molecular Biology of the Cell* **15**, 696 (2004).

556. Artamonov, M. V. *et al.* Agonist-induced Ca2+ Sensitization in Smooth Muscle: Redundancy of Rho guanine nucleotide exchange factors (RhoGEFs) and response kinetics a caged compound study*. *Journal of Biological Chemistry* **288**, 34030 (2013).

557. Dulin, N. O. *et al.* RGS3 inhibits G protein-mediated signalling via translocation to the membrane and binding to Galpha11. *Molecular and Cell Biology* **19**, 714–723 (1999).

558. Castro-Fernández, C., Maya-Núñez, G. & Méndez, J. P. Regulation of follicle-stimulating and luteinizing hormone receptor signalling by. *Endocrine* **25**, 49–54 (2004).

559. Wheeler, D. S. *et al.* Direct interaction between NHERF1 and Frizzled regulates β-catenin signalling. *Oncogene* **30**, 32–42 (2011).

560. Luyten, A. *et al.* The Postsynaptic Density 95/Disc-Large/Zona Occludens Protein Syntenin Directly Interacts with Frizzled 7 and Supports Noncanonical Wnt Signalling. *Molecular Biology of the Cell* **19**, 1594 (2008).

561. Balasubramanian, S., Fam, S. R. & Hall, R. A. GABAB receptor association with the PDZ scaffold Mupp1 alters receptor stability and function. *Journal of Biological Chemistry* **282**, 4162–4171 (2007).

562. Patel, M. *et al.* Gα13/PDZ-RhoGEF/RhoA signalling is essential for gastrin-releasing peptide receptor-mediated colon cancer cell migration. *Molecular Pharmacology* **86**, 252–262 (2014).

563. Neill, J. D. *et al.* Potential role for a regulator of G protein signalling (RGS3) in gonadotropin-releasing hormone (GnRH) stimulated desensitization. *Endocrinology* **138**, 843–846 (1997).

564. Neill, J. D., Duck, L. W., Sellers, J. C., Musgrove, L. C. & Kehrl, J. H. A regulator of G protein signalling, RGS3, inhibits gonadotropin-releasing hormone (GnRH)-stimulated luteinizing hormone (LH) secretion. *BMC Cell Biology* **2**, (2001).

565. Castro-Fernández, C. & Conn, P. M. Regulation of the gonadotropin-releasing hormone receptor (GnRHR) by RGS proteins: Role of the GnRHR carboxyl-terminus. *Molecular Cell Endocrinology* **191**, 149–156 (2002).

566. Karakoula, A., Tovey, S. C., Brighton, P. J. & Willars, G. B. Lack of receptor-selective effects of either RGS2, RGS3 or RGS4 on muscarinic M3- and gonadotropin-releasing hormone receptor-mediated signalling through G alpha q/11. *European Journal of Pharmacology* **587**, 16–24 (2008).

567. Katsushima, Y. *et al.* Interaction of PICK1 with C-Terminus of Growth Hormone-Releasing Hormone Receptor (GHRHR) modulates trafficking and signal transduction of human GHRHR. *Journal of Pharmacolical Science* **122**, 193–204 (2013).

568. Pfreimer, M. *et al.* LARG links histamine-H1-receptor-activated Gq to Rho-GTPase-dependent signalling pathways. *Cell Signal* **24**, 652–663 (2012).

569. Turner, E. C., Mulvaney, E. P., Reid, H. M. & Kinsella, B. T. Interaction of the human prostacyclin receptor with the PDZ adapter protein PDZK1: role in endothelial cell migration and angiogenesis. *Molecular Biology of the Cell* **22**, 2664–2679 (2011).

570. Hirakawa, T., Galet, C., Kishi, M. & Ascoli, M. GIPC binds to the human lutropin receptor (hLHR) through an unusual PDZ domain binding motif, and it regulates the sorting of the internalized human choriogonadotropin and the density of cell surface hLHR. *Journal of Biological Chemistry***278**, 49348–49357 (2003).

571. Bhaskaran, R. S. & Ascoli, M. The post-endocytotic fate of the gonadotropin receptors is an important determinant of the desensitization of gonadotropin responses. *Journal of Molecular Endocrinology* **34**, 447–457 (2005).

572. Varsano, T., Taupin, V., Guo, L., Baterina, O. Y. & Farquhar, M. G. The PDZ Protein GIPC Regulates Trafficking of the LPA1 Receptor from APPL Signalling Endosomes and Attenuates the Cell's Response to LPA. *PLoS One* **7**, (2012).

573. Oh, Y.-S. *et al.* NHERF2 specifically interacts with LPA2 receptor and defines the specificity and efficiency of receptor-mediated phospholipase C-beta3 activation. *Molecular and Cell Biology* **24**, 5069–5079 (2004).

574. Yamada, T., Ohoka, Y., Kogo, M. & Inagaki, S. Physical and functional interactions of the lysophosphatidic acid receptors with PDZ domain-

containing Rho guanine nucleotide exchange factors (RhoGEFs). *Journal of Biological Chemistry* **280**, 19358–19363 (2005).

575. Zhang, H., Wang, D., Sun, H., Hall, R. A. & Yun, C. C. MAGI-3 regulates LPA-induced activation of Erk and RhoA. *Cell Signal* **19**, 261–268 (2007).

576. Lee, S. *et al.* MAGI-3 competes with NHERF-2 to negatively regulate LPA2 receptor signalling in colon cancer cells. *Gastroenterology* **140**, 924–934 (2011).

577. Guillaume, J. L. *et al.* The PDZ Protein Mupp1 Promotes Gi Coupling and Signalling of the Mt1 Melatonin Receptor. *Journal of Biological Chemistry* **283**, 16762–16771 (2008).

578. Paquet, M. *et al.* The PDZ scaffold NHERF-2 interacts with mGluR5 and regulates receptor activity. *Journal of Biological Chemistry* **281**, 29949–29961 (2006).

579. Kudo, N. *et al.* Molecular cloning and cell cycle-dependent expression of mammalian CRM1, a protein involved in nuclear export of proteins. *Journal of Biological Chemistry* **272**, 29742–29751 (1997).

580. Perroy, J. *et al.* PICK1 is required for the control of synaptic transmission by the metabotropic glutamate receptor 7. *EMBO Journal* **21**, 2990–2999 (2002).

581. Suh, Y. H. *et al.* Corequirement of PICK1 binding and PKC phosphorylation for stable surface expression of the metabotropic glutamate receptor mGluR7. *Neuron* **58**, 736–748 (2008).

582. Dooley, R., Baumgart, S., Rasche, S., Hatt, H. & Neuhaus, E. M. Olfactory receptor signalling is regulated by the post-synaptic density 95, Drosophila discs large, zona-occludens 1 (PDZ) scaffold multi-PDZ domain protein 1. *FEBS Journal* **276**, 7279–7290 (2009).

583. Charlton, J. J. *et al.* Multiple Actions of Spinophilin Regulate Mu Opioid Receptor Function. *Neuron* **58**, 238 (2008).

584. Fourla, D. D., Papakonstantinou, M. P., Vrana, S. M. & Georgoussi, Z. Selective interactions of spinophilin with the C-terminal domains of the δ- and μ-opioid receptors and G proteins differentially modulate opioid receptor signalling. *Cell Signal* **24**, 2315–2328 (2012).

585. Fam, S. R. *et al.* P2Y1 receptor signalling is controlled by interaction with the PDZ scaffold NHERF-2. *Proceedings of the National Academy of Sciences U S A* **102**, 8042–8047 (2005).

586. Mahon, M. J., Donowitz, M., Yun, C. C. & Segre, G. V. Na(+)/H(+) exchanger regulatory factor 2 directs parathyroid hormone 1 receptor signalling. *Nature* **417**, 858–861 (2002).

587. Sneddon, W. B. *et al.* Activation-independent parathyroid hormone receptor internalization is regulated by NHERF1 (EBP50). *Journal of Biological Chemistry* **278**, 43787–43796 (2003).

588. Wang, B., Bisello, A., Yang, Y., Romero, G. G. & Friedman, P. A. NHERF1 regulates parathyroid hormone receptor membrane retention without affecting recycling. *Journal of Biological Chemistry* **282**, 36214–36222 (2007).

589. Wang, B. *et al.* Na/H exchanger regulatory factors control parathyroid hormone receptor signalling by facilitating differential activation of G(alpha) protein subunits. *Journal of Biological Chemistry* **285**, 26976–26986 (2010).

590. Wang, B., Yang, Y., Abou-Samra, A. B. & Friedman, P. A. NHERF1 regulates parathyroid hormone receptor desensitization: interference with beta-arrestin binding. *Molecular Pharmacology* **75**, 1189–1197 (2009).

591. Ladds, G., Goddard, A., Hill, C., Thornton, S. & Davey, J. Differential effects of RGS proteins on G alpha(q) and G alpha(11) activity. *Cell Signal* **19**, 103–113 (2007).

592. Emery, A. C., Eiden, M. V., Mustafa, T. & Eiden, L. E. Rapgef2 Connects GPCR-Mediated cAMP Signals to ERK Activation in Neuronal and Endocrine Cells. *Science Signalling* **6**, 51 (2013).

593. Chen, B., Siderovski, D. P., Neubig, R. R., Lawson, M. A. & Trejo, J. Regulation of protease-activated receptor 1 signalling by the adaptor protein complex 2 and R4 subfamily of regulator of G protein signalling proteins. *Journal of Biological Chemistry* **289**, 1580–1591 (2014).

594. Dunn, H. A. *et al.* Role of SAP97 in the regulation of 5-HT2AR endocytosis and signalling. *Molecular Pharmacology* **86**, 275–283 (2014).

595. Backstrom, J. R., Price, R. D., Reasoner, D. T. & Sanders-Bush, E. Deletion of the serotonin 5-HT2C receptor PDZ recognition motif prevents receptor phosphorylation and delays resensitization of receptor responses. *Journal of Biological Chemistry* **275**, 23620–23626 (2000).

596. Gavarini, S. *et al.* Opposite effects of PSD-95 and MPP3 PDZ proteins on serotonin 5-hydroxytryptamine2C receptor desensitization and membrane stability. *Molecular Biology of the Cell* **17**, 4619–4631 (2006).

597. Medlin, M. D., Staus, D. P., Dubash, A. D., Taylor, J. M. & MacK, C. P. Sphingosine 1-phosphate receptor 2 signals through leukemia-associated RhoGEF (LARG), to promote smooth muscle cell differentiation. *Arteriosclerosis, Thrombosis, and Vascular Biology* **30**, 1779–1786 (2010).

598. Del Galdo, S., Vettel, C., Heringdorf, D. M. zu & Wieland, T. The activation of RhoC in vascular endothelial cells is required for the S1P receptor type 2-induced inhibition of angiogenesis. *Cell Signal* **25**, 2478–2484 (2013).

599. Gee, H. Y. *et al.* Synaptic Scaffolding Molecule Binds to and Regulates Vasoactive Intestinal Polypeptide Type-1 Receptor in Epithelial Cells. *Gastroenterology* **137**, 607-617 (2009).

600. Ota, T. *et al.* Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genetics* **36**, 40–45 (2004).

601. Seo, S. *et al.* BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. *Proceedings of the National Academy of Sciences U S A* **107**, 1488–1493 (2010).

602. Freund, A. *et al.* Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1. *Cell* **159**, 1389–1403 (2014).

603. Bitoun, M. *et al.* Dynamin 2 mutations associated with human diseases impair clathrin-mediated receptor endocytosis. *Human Mutations* **30**, 1419–1427 (2009).

604. Züchner, S. *et al.* Mutations in the pleckstrin homology domain of dynamin 2 cause dominant intermediate Charcot-Marie-Tooth disease. *Nature Genetics* **37**, 289–294 (2005).

605. Boutchueng-Djidjou, M. *et al.* The last enzyme of the de novo purine synthesis pathway 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC) plays a central role in insulin signalling and the golgi/endosomes protein network. *Molecular and Cellular Proteomics* **14**, 1079–1092 (2015).

606. Raman, N., Weir, E. & Müller, S. The AAA ATPase MDN1 Acts as a SUMO-Targeted Regulator in Mammalian Pre-ribosome Remodeling. *Moecularl Cell* **64**, 607–615 (2016).

607. Rötig, A. *et al.* Sequence and structure of the human OXA1L gene and its upstream elements. *Biochimica et Biophysica Acta Molecular Basis of Disease* **1361**, 6–10 (1997).

608. Bonnefoy, N. *et al.* Cloning of a human gene involved in cytochrome oxidase assembly by functional complementation of an oxal- mutation in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences U S A* **91**, 11978–11982 (1994).

609. Ryu, C. S., Klein, K. & Zanger, U. M. Membrane Associated Progesterone Receptors: Promiscuous Proteins with Pleiotropic Functions - Focus on Interactions with Cytochromes P450. *Frontiers in Pharmacology* **8**, (2017).

610. Groettrup, M. *et al.* A role for the proteasome regulator PA28alpha in antigen presentation. *Nature* **381**, 166–168 (1996).

611. Yu, Y. H. *et al.* Prostaglandin reductase-3 negatively modulates adipogenesis through regulation of PPARγ activity. *Journal of Lipid Research* **54**, 2391–2399 (2013).

612. Merrill, M. J., Yeh, G. C. & Phang, J. M. Purified Human Erythrocyte Pyrroline-5-carboxylate Reductase. *Journal of Biological Chemistry* **264**, 9352–9358 (1989).

613. Yeh, G. C., Harris, S. C. & Phang, J. M. Pyrroline-5-carboxylate reductase in human erythrocytes. A comparison of differential regulation. *Journal of Clinical Investigation* **67**, 1042–1046 (1981).

614. Vaittinen, M. *et al.* Downregulation of CPPED1 expression improves glucose metabolism in vitro in adipocytes. *Diabetes* **62**, 3747–3750 (2013).

615. Yang, S. *et al.* Molecular mechanism of fascin function in filopodial formation. *Journal of Biological Chemistry* **288**, 274–284 (2013).

616. Jansen, S. *et al.* Mechanism of actin filament bundling by fascin. *Journal of Biological Chemistry* **286**, 30087–30096 (2011).

617. Chen, L., Yang, S., Jakoncic, J., Zhang, J. J. & Huang, X. Y. Migrastatin analogues target fascin to block tumour metastasis. *Nature* **464**, 1062–1066 (2010).

618. Kumar, A. S., Naruszewicz, I., Wang, P., Leung-Hagesteijn, C. & Hannigan, G. E. ILKAP regulates ILK signalling and inhibits anchorage-independent growth. *Oncogene* **23**, 3454–3461 (2004).

619. Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363**, 710–714 (2019).

620. Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* **359**, 537–545 (2018).

621. Zhang, X. *et al.* An Atomic Structure of the Human Spliceosome. *Cell* **169**, 918-929 (2017).

622. Bartolini, F. *et al.* Functional overlap between retinitis pigmentosa 2 protein and the tubulin-specific chaperone cofactor C. *Journal of Biological Chemistry* **277**, 14629–14634 (2002).

623. Liang, J. R. *et al.* A Genome-wide ER-phagy Screen Highlights Key Roles of Mitochondrial Metabolism and ER-Resident UFMylation. *Cell* **180**, 1160-1177.e20 (2020).

624. Nahorski, M. S. *et al.* Biallelic UFM1 and UFC1 mutations expand the essential role of ufmylation in brain development. *Brain* **141**, 1934–1945 (2018).

625. Komatsu, M. *et al.* A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO Journal* **23**, 1977–1986 (2004).

626. Manjasetty, B. A. *et al.* Crystal structure of Homo sapiens PTD012 reveals a zinc-containing hydrolase fold. *Protein Science* **15**, 914–920 (2006).

627. Wei, Y. *et al.* Crystal structures of human lysosomal EPDR1 reveal homology with the superfamily of bacterial lipoprotein transporters. *Communications in Biology* **2**, (2019).

628. Gu, Y. *et al.* TCRP1 promotes radioresistance of oral squamous cell carcinoma cells via Akt signal pathway. *Molecular and Cellular Biochemistry* **357**, 107–113 (2011).

629. Chen, H. K. & Yeh, N. H. The nucleolar phosphoprotein P130 is a GTPase/ATPase with intrinsic property to form large complexes triggered by

F- and Mg2+. *Biochemical and Biophysical Research Communications* **230**, 370–375 (1997).

630. Werner, A. *et al.* Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature* **525**, 523–527 (2015).

631. Chen, H.-K., Pai, C.-Y., Huang, J.-Y. & Yeh, N.-H. Human Nopp140, Which Interacts with RNA Polymerase I: Implications for rRNA Gene Transcription and Nucleolar Structural Organization. *Molecular and Cell Biology* **19**, 8536–8546 (1999).

632. Bertram, K. *et al.* Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* **170**, 701-713 (2017).

633. Fournier, G. *et al.* Recruitment of RED-SMU1 Complex by Influenza A Virus RNA Polymerase to Control Ciral mRNA Splicing. *PLoS Pathogens* **10**, (2014).

634. Srikanth, S. *et al.* Junctate is a Ca2+-sensing structural component of Orai1 and stromal interaction molecule 1 (STIM1). *Proceedings of the National Academy of Sciences U S A* **109**, 8682–8687 (2012).

635. Dinchuk, J. E. *et al.* Absence of post-translational aspartyl beta-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia. *Journal of Biological Chemistry* **277**, 12970–12977 (2002).

636. Naik, J. *et al.* The P4-ATPase ATP9A is a novel determinant of exosome release. *PLoS One* **14**, (2019).

637. McGough, I. J. *et al.* SNX3-retromer requires an evolutionary conserved MON2:DOPEY2:ATP9A complex to mediate Wntless sorting and Wnt secretion. *Nature Communications* **9**, (2018).

638. Tanaka, Y. *et al.* The phospholipid flippase ATP9A is required for the recycling pathway from the endosomes to the plasma membrane. *Molecular Biology of the Cell* **27**, 3883–3893 (2016).

639. Castora, F. J., Kerns, K. A., Pflanzer, H. K., Shelton, M. G. & Coleman, R. A. Differential expression of mitochondrial energy producing genes in AD brains compromise ATP Synthesis and neuronal viability and stimulate inflammatory pathways and ROS synthesis. *Alzheimer's & Dementia* **19**, 077104 (2023).

640. Wang, L., Wu, D., Robinson, C. V., Wu, H. & Fu, T. M. Structures of a Complete Human V-ATPase Reveal Mechanisms of Its Assembly. *Molecular Cell* **80**, 501-511.e3 (2020).

641. Mick, D. U. *et al.* MITRAC links mitochondrial protein translocation to respiratory-chain assembly and translational regulation. *Cell* **151**, 1528–1541 (2012).

642. Kahn, M., Nakanishi-Matsui, M., Shapiro, M., Ishihara, H. & Coughlin, S. Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *Journal of Clinical Investigation* **103**, 879–887 (1999).

643. Kamijo, T., Aoyama, T., Komiyama, A. & Hashimoto, T. Structural analysis of cDNAS for subunits of human mitochondrial fatty acid β-oxidation trifunctional protein. *Biochemical and Biophysical Research Communications* **199**, 818–825 (1994).

644. Liang, K. *et al.* Cryo-EM structure of human mitochondrial trifunctional protein. *Proceedings of the National Academy of Sciences U S A* **115**, 7039–7044 (2018).

645. Xia, C., Fu, Z., Battaile, K. P. & Kim, J. J. P. Crystal structure of human mitochondrial trifunctional protein, a fatty acid β-oxidation metabolon. *Proceedings of the National Academy of Sciences U S A* **116**, 6069–6074 (2019).

646. Taylor, W. A. *et al.* Human Trifunctional Protein Alpha Links Cardiolipin Remodeling to Beta-Oxidation. *PLoS One* **7**, (2012).

647. Jäkel, S. & Görlich, D. Importin β, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO Journal* **17**, 4491–4502 (1998).

648. Dean, K. A., Von Ahsen, O., Görlich, D. & Fried, H. M. Signal recognition particle protein 19 is imported into the nucleus by importin 8 (RanBP8) and transportin. *Journal of Cell Science* **114**, 3479–3485 (2001).

649. Padavannil, A. *et al.* Importin-9 wraps around the H2A-H2B core to act as nuclear importer and histone chaperone. *Elife* **8**, (2019).

650. de Almeida, M. *et al.* AKIRIN2 controls the nuclear import of proteasomes in vertebrates. *Nature* **599**, 491–496 (2021).

651. Jakel, S. Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *EMBO Journal* **21**, 377–386 (2002).

652. Wilson, C. M., Magnaudeix, A., Yardin, C. & Terro, F. DC2 and keratinocyte-associated protein 2 (KCP2), subunits of the oligosaccharyltransferase complex, are regulators of the γ-secretase- directed processing of amyloid precursor protein (APP). *Journal of Biological Chemistry* **286**, 31080–31091 (2011).

653. Roboti, P. & High, S. The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation. *Journal of Cell Science* **125**, 3474–3484 (2012).

654. Boulet, A. *et al.* The mammalian phosphate carrier SLC25A3 is a mitochondrial copper transporter required for cytochrome c oxidase biogenesis. *Journal of Biological Chemistry* **293**, 1887–1896 (2018).

655. Mayr, J. A. *et al.* Mitochondrial phosphate-carrier deficiency: A novel disorder of oxidative phosphorylation. *American Journal of Human Genetics* **80**, 478–484 (2007).

656. Zhang, X. *et al.* Redox signals at the ER –mitochondria interface control melanoma progression *EMBO Journal* **38**, (2019).

657. Haugstetter, J., Blicher, T. & Ellgaard, L. Identification and characterization of a novel thioredoxin-related transmembrane protein of the endoplasmic reticulum. *Journal of Biological Chemistry* **280**, 8371–8380 (2005).

658. Brock, S. *et al.* Defining the phenotypical spectrum associated with variants in TUBB2A. *Journal of Medical Genetics* **58**, 33–40 (2021).

659. Wan, L. C. K. *et al.* Proteomic analysis of the human KEOPS complex identifies C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic Acid Research* **45**, 805–817 (2017).

660. Costessi, A. *et al.* The human EKC/KEOPS complex is recruited to cullin2 ubiquitin ligases by the human tumour antigen PRAME. *PLoS One* **7**, (2012).

661. Masayo, K. *et al.* Interaction of hCLIM1, an enigma family protein, with α-actinin 2. *Journal of Cellular Biochemistry* **78**, 558–565 (2000).

662. Sillibourne, J. E., Delaval, B., Redick, S., Sinha, M. & Doxsey, S. J. Chromatin remodeling proteins interact with pericentrin to regulate centrosome integrity. *Molecular Biology of the Cell* **18**, 3667–3680 (2007).

663. Jani, D. *et al.* Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export. *Nucleic Acid Research* **40**, 4562–4573 (2012).

664. Bhuvanakantham, R., Li, J., Terence Tan, T. T. & Ng, M. L. Human Sec3 protein is a novel transcriptional and translational repressor of flavivirus. *Cell Microbiology* **12**, 453–472 (2010).

665. Nissim, S. *et al.* Mutations in RABL3 alter KRAS prenylation and are associated with hereditary pancreatic cancer. *Nature Genetics* **51**, 1308–1314 (2019).

8. APPENDIX

STED microscope alignment:

This procedure is a check of the optical hardware and evaluates the overall system performance. It was performed at least once a week, prior to every new measurement session, or after changing a key element of the optical layout, for example the major dichroic (e.g. changing the excitation wavelength) or the microscope objective. First, it was checked that the correct dichroic and filter were mounted (60 X water dipping objective, Atto655 = 646 + 740 dichroic and a 690/70 BP). Measurements were taken on the test mode workspace on the computer. Clear glass coverslip of standard thickness #1 or #1.5 were mounted in the sample holder. 200 µL of a 1µM concentrated solution of fluorescent dye suitable for the laser wavelength were added. The OD3 neutral density filter in the detection filter wheel was selected, this is done for safety reason being the strongest density filter. Detectors 1 and 2 were turned on, detector 3 can also be switched on if required. The camera filter wheel was opened, followed by selection of the laser of choice and opening of the excitation shutter. After pressing start, a light ray was visible on the camera image (if this does not happen, it may be required to increase the laser power). The objective was slowly raised towards the sample using the microscope focusing block, until a beam spot was found. The excitation intensity was adjusted, and suitable OD filters were applied to avoid over illumination. The actual alignment consisted in adjusting the beam using the beam displacer unit found directly on the main optical unit. A well aligned beam has symmetric intensity distribution - pattern should be symmetric in the horizontal and in the vertical direction. However, the exact pattern depends on the excitation wavelength, the type of objective, coverside, as well as on laser polarization. Once beam spot is symmetric, the top reflection was targeted, then the objective was raised so that the focus is 20 µm inside the dye solution. The laser power was lowered suitable fluorescence filter in front of detector were installed if necessary. The computer settings were then changed to time trace, and the pinhole was aligned by setting the values to mean and moving the pinhole in both perpendicular directions with the knobs on its holder. The detected count rate was thus adjusted for the maximum count rate in both directions. To align the detectors, TCPSC was selected, and the integration time was set to 0.1. The beginning of the curve was selected, and edges were set. Using the pinholes in front of the detectors, the curve was adjusted by moving the curve to the left.

After that, the STED timing alignment was performed by adjusting the power of the 645-laser line to get 100k intensity counts from the specimen and set to mean. All lids were closed so that all the interlocks could be active, then the STED laser was slowly increased until the intensity counts were roughly 50 % depleted and set this to the mean. The “doughnut” was then aligned by turning off the MCL box at the back of the controller whilst changing the objective. Then objective was changed to 100X. The MCL box was turned back on, and the bandpass filter was removed from in front of the detector. In the drop-down panel the 100x objective was selected and the gold bead slides were mounted. In Time Trace, the excitation laser power was increased to approximately 2000 arb. Units. The top reflection was found again and, the viewing range was selected within the green square displayed on the screen. The intensity was then fixed to max on the image display. A single bead was selected within the FOV and the piezo stage was used to achieve the best focus. The excitation laser was blocked by setting the intensity selection between two values. The AUX shutter

was set to auto, the STED laser intensity shutter was closed completely and OD3 shutter filter was placed in the illumination path. Imaging began by slowly opening the STED laser intensity shutter until a clear intensity spot was observed. The beam displacer was used to form this into a doughnut shape. Typically, this displacer should be placed on the red dot which appears optimum.

The Gatta slide specimen was mounted and the band pass filter was placed back in front of the detector. Focus onto the test specimen was applied using only the excitation laser and the piezo stage. On the computer, time trace was selected again, the AUX shutter was opened and set the STED intensity to approx. 500,000 arb units. On the software, measurement was selected, and a STED image was taken with the AUX EXC (out of 2 position) and DET 1 shutters open. Analysis was performed by selecting data then open the analysis panel. Gated STED analysis was performed by changing the TCSPC curve to just after the left most peak. After calculation, Lorentzian fit was selected

Table S1 – Summary of the GPCRs known to interact with PDZ proteins and their effects. Adapted from Romero *et al.,* 2011[233] and Dunn *et al.*, 2015[533].

GPCR	**PDZ motif**	**PDZ protein**	**Trafficking effect**	**Signalling effect**	**Ref.**
Adenosine A2 receptor	DVELL	SAP102	Increased residence at plasma membrane	Increases ERK activity	[534]
A-1D adrenergic receptor	ETDI	Synotrophins	Receptor recyclying (SNT-2A) Receptor internalisation and degradation (SNT-2B)	Increased Inositol 1,4,5-trisphosphate, Ca^{2+} and ERK; decreased cAMP	[535,536]
α-2 adrenergic receptor	Folded surface of non-contiguous regions in the third intracellular loop	Spinophilin	NA	Decreases Ca^{2+} and G_i coupling	[537,538]
β-1 adrenergic receptor	ESKV	PSD95, SAP97, GIPC, CAL, MAGI2, MAGI3	Promotes recycling (SAP97)	Resensitisation of cAMP signalling (SAP97)	[539,540]
β-2 adrenergic receptor	DSLL	NHERF1, NHERF2, PDZK1, SNX27 RGS3	Promotes recycling (SNX27, NHERF2)	Signalling via NHE3 sodium-protein exchanger (NHERF1)	[212,541–543]

				Increased G_s mediated signalling	
Angiotensin II receptor 1	Not reported	LARG	Recruitment of Rho to cell surface	Increased Rho	[544–546]
Bradykinin 2 receptor	Not reported	Par3	NA	Increases PLC interaction	[547]
Cannabinoid receptor 1	Not reported	nNOS	NA	Increased PKC interaction	[548]
C-C chemokine receptor type 5	SVGL	NHERF1, NHERF 2	Increases cellular adhesion (NHERF2)	Increases ERK activity (NHERF1) and Rho activity (NHERF2)	[549,550]
Corticotropin-releasing factor receptor 1	TAV	PDZK1 SAP97	Receptor endocytosis (SAP97)	Increases ERK activity	[551,552]
Dopamine-1 Receptor	N^{334}…T^{446}	PSD95	Reduce receptor surface expression	Decreases cAMP	[553,554]
Dopamine-2 Receptor	Not reported	nNOS RGS3	NA	Increased PKC interactions (nNOS) Decreased G_i mediated signalling (RGS3)	[548]
Dopamine-3 Receptor	LSC	GIPC	Colocalisation in plasma membrane	Decreased G_i coupling	[555]
Endothelin 1 receptor	Not reported	LARG RGS3	NA	Increased Rho (LARG) Decreased Ca^{2+} (RGS3)	[556,557]
Follicle-stimulating hormone receptor	Not reported	RGS3	NA	Decreased inositol 1,4,5-trisphosphate	[558]
Frizzled receptor 2	ETTV	NHERF2	NA	Decreases β-catenin	[559]
Frizzled receptor 7	ETAV	Syntenin-1	NA	Increased c-Jun, CDC42, and PKCα	[560]

GABA-B receptor	VSGL	MUPP1	NA	Increased Ca^{2+} Prolonged signalling duration	[561]
Gastrin-releasing peptide receptor	Not reported	PDZ-RhoGEF	NA	Increased Rho	[562]
Gonadotropin-releasing hormone receptor	Not reported	RGS3	NA	Decreased Inositol 1,4,5-triphosphate G-protein signalling inhibition Decreased DAG	[563–566]
Growth hormone–releasing hormone receptor	$I^{387}…C^{423}$	PICK1	NA	Decreased cAMP	[567]
Histamine 1 Receptor	TFKRILHIRS	LARG	NA	Increased Rho	[568]
Human prostacyclin receptor	CSLC	PDZK1 PDZK2	Increased functional expression at cell surface	Increase cAMP formation	[569]
Lutropin-choriogonadotropic hormone receptor	YTEC	GIPC RGS3	Promotes hormone recycling (GIPC)	Mantains surface receptor levels (GIPC) Decreased Inositol 1,4,5-trisphosphate and cAMP(RGS3)	[570,571]
Lysophosphatidic acid receptor 1	SVV	GIPC	Trafficking to endosomes	Increased PKB	[572]
Lysophosphatidic acid receptor 2	DSTL	NHERF2, PDZ-RhoGEF, MAGI3	None reported	Potentiates LPA-induced activation of PLC-β (NHERF2), rewuired for LPA-induced RhoA activation, (PDZ-RhoGEF),	[573–576]

				promote receptor coupling to $G_{\alpha 12}$ and Erk activation (MAGI3)	
Melatonin 1 receptor	DSV	MUPP1	NA	Promotes G_i coupling	577
Metabotropic glutamate receptor 5	SSSL	NHERF2		Prolong duration of the receptor mediated Ca^{2+} response	578
Metabotropic glutamate receptor 7	NLVI	PICK1	Stabilises receptors at plasma membrane	Required for inhibition receptor-mediated inhibition of P/Q-type Ca^{2+} channels	579–581
Muscarinic acetylcholine receptor 3	Not reported	RGS3	NA	Decreased Inositol 1,4,5-trisphosphate, DAG, Ca^{2+}, ERK, and PKB	566
Muscarinic acetylcholine receptor 4	Not reported	nNOS	NA	Increased PKC interactions	548
Olfactory receptor 2AG1	ESHS	MUPP1	NA	Increased Ca^{2+} decay	582
μ-opioid receptor	Intracellular loop 3	Spinophilin nNOS RGS3	Modulates endocytosis (Spinophilin)	Increased G_i coupling (Spinophilin) Increased PKC interaction (nNOS) Decreased G_i mediated signalling (RGS3)	548,583,584

δ-opioid receptor	Intracellular loop 3	Spinophilin	NA	Increased G_i coupling Decreased ERK activity	[584]
P2Y purinoceptor 1	DSTL	NHERF2	NA	Prolong duration of the receptor mediated Ca^{2+} response	[585]
Parathyroid hormone/parathyroid hormone-related peptide receptor	ETVM	NHERF1, NHERF2	Tethers receptor at cell membrane	Switches G protein signalling, regulates ERK signalling, imparts ligand bias, regulates desensitisation	[586–590]
Pheromone p-factor receptor	Not reported	RGS3	NA	Decreased G_q and G_{11} activation	[591]
Pituitary adenylate cyclase–activating polypeptide 1 receptor	Not reported	PDZ-GEF	NA	Increased ERK	[592]
Protease activated receptor 1	YKKAA	AP-2	Modulate surface expression	Decreased inositol 1,4,5-triphosphate	[593]
Serotonin 1A receptor	Not reported	nNOS	NA	Increased PKC interaction	[548]
Serotonin 2A receptor	VSCV	SAP97 PDZK1 nNOS	Regulates endocytosis (SAP97)	Increases inositol 1,4,5-triphosphate (SAP97) Increased ERK activity (PDZK1) Increased PKC interaction (nNOS)	[548,594]
Serotonin 2C receptor	ISSV	PSD95, MPP3	Promotes endocytosis (PSD95, inhibits endocytosis (MPP3)	Promotes desensitisation (PSD95), inhibits desensitisation (MPP3)	[595,596]

Sphingosine-1-phosphate receptor 2	Not reported	LARG	NA	Increased Rho	597,598
Thromboxane A2 receptor	Not reported	LARG	NA	Increased Rho	556
Vasoactive intestinal polypeptide receptor 1	SLV	MAGI-2	Recruitment to junctional area	Decreases cAMP formation	599

Table S2 – Heavy/Light dataset for the SLiM proteome

Accession	Description	Heavy/Light
A8K905	cDNA FLJ77615, highly similar to Homo sapiens nucleolar complex associated 3 homolog (S. cerevisiae) (NOC3L), mRNA OS=Homo sapiens PE=2 SV=1 - [A8K905_HUMAN]	4.040958
B4DHX2	cDNA FLJ58181, highly similar to Homo sapiens ATP/GTP binding protein 1 (AGTPBP1), mRNA OS=Homo sapiens PE=2 SV=1 - [B4DHX2_HUMAN]	100

Table S3 – Medium/Light dataset for the SLiM proteome

Accession	Description	Medium/Light
A0A024QZW7	Nucleoporin 153kDa, isoform CRA_a OS=Homo sapiens GN=NUP153 PE=4 SV=1 - [A0A024QZW7_HUMAN]	100
A0A0A0MTB8	WD repeat-containing protein 36 OS=Homo sapiens GN=WDR36 PE=1 SV=1 - [A0A0A0MTB8_HUMAN]	11.10851
A8K905	cDNA FLJ77615, highly similar to Homo sapiens nucleolar complex associated 3 homolog (S. cerevisiae) (NOC3L), mRNA OS=Homo sapiens PE=2 SV=1 - [A8K905_HUMAN]	8.030817
A8K964	cDNA FLJ75071, highly similar to Homo sapiens pinin, desmosome associated protein (PNN), mRNA OS=Homo sapiens PE=2 SV=1 - [A8K964_HUMAN]	100
B4DHX2	cDNA FLJ58181, highly similar to Homo sapiens ATP/GTP binding protein 1 (AGTPBP1), mRNA OS=Homo sapiens PE=2 SV=1 - [B4DHX2_HUMAN]	100
B7Z5N7	cDNA FLJ58612, highly similar to Sec1 family domain-containing protein 1 OS=Homo sapiens PE=2 SV=1 - [B7Z5N7_HUMAN]	100

C9JNK6	Metaxin-2 OS=Homo sapiens GN=MTX2 PE=1 SV=1 - [C9JNK6_HUMAN]	100
D6RJD1	Clathrin light chain B (Fragment) OS=Homo sapiens GN=CLTB PE=1 SV=1 - [D6RJD1_HUMAN]	100
E5RHW4	Erlin-2 (Fragment) OS=Homo sapiens GN=ERLIN2 PE=1 SV=1 - [E5RHW4_HUMAN]	7.932748
G3V5Z3	Serine/threonine-protein phosphatase 4 regulatory subunit 3A OS=Homo sapiens GN=PPP4R3A PE=1 SV=1 - [G3V5Z3_HUMAN]	10.76131
O00400	Acetyl-coenzyme A transporter 1 OS=Homo sapiens GN=SLC33A1 PE=1 SV=1 - [ACATN_HUMAN]	6.27442
Q05D80	RNMT protein (Fragment) OS=Homo sapiens GN=RNMT PE=2 SV=1 - [Q05D80_HUMAN]	100
Q53SY1	Reticulon (Fragment) OS=Homo sapiens GN=RTN4 PE=4 SV=1 - [Q53SY1_HUMAN]	9.869413
Q59FU4	Diacylglycerol kinase (Fragment) OS=Homo sapiens PE=2 SV=1 - [Q59FU4_HUMAN]	7.3947
Q8IVQ8	XPO4 protein (Fragment) OS=Homo sapiens GN=XPO4 PE=2 SV=1 - [Q8IVQ8_HUMAN]	10.27355
Q9UDT1	Rhomboid domain-containing protein 2 OS=Homo sapiens GN=WUGSC:H_RG122E10.2a PE=1 SV=1 - [Q9UDT1_HUMAN]	6.557967
Q9Y2V7	Conserved oligomeric Golgi complex subunit 6 OS=Homo sapiens GN=COG6 PE=1 SV=2 - [COG6_HUMAN]	100

Table S4 – Medium/Heavy dataset for the SLiM proteome

Accession	Description	Medium/Heavy
A0A0U1RQT9	Synaptophysin-like protein 1 (Fragment) OS=Homo sapiens GN=SYPL1 PE=1 SV=1 - [A0A0U1RQT9_HUMAN]	100
B2R8A2	cDNA, FLJ93804, highly similar to Homo sapiens gp25L2 protein (HSGP25L2G), mRNA OS=Homo sapiens PE=2 SV=1 - [B2R8A2_HUMAN]	100
B4DDC8	cDNA FLJ57252, highly similar to Protein phosphatase 2C isoform gamma (EC 3.1.3.16) OS=Homo sapiens PE=2 SV=1 - [B4DDC8_HUMAN]	15.14589
B4E2Z3	cDNA FLJ54090, highly similar to 4F2 cell-surface antigen heavy chain OS=Homo sapiens PE=2 SV=1 - [B4E2Z3_HUMAN]	3.718703
B7ZM73	MON2 protein OS=Homo sapiens GN=MON2 PE=2 SV=1 - [B7ZM73_HUMAN]	4.521683
D6RJD1	Clathrin light chain B (Fragment) OS=Homo sapiens GN=CLTB PE=1 SV=1 - [D6RJD1_HUMAN]	100
O14828	Secretory carrier-associated membrane protein 3 OS=Homo sapiens GN=SCAMP3 PE=1 SV=3 - [SCAM3_HUMAN]	4.159701
O15397	Importin-8 OS=Homo sapiens GN=IPO8 PE=1 SV=2 - [IPO8_HUMAN]	5.084975

O75380	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial OS=Homo sapiens GN=NDUFS6 PE=1 SV=1 - [NDUS6_HUMAN]	11.31014
P55795	Heterogeneous nuclear ribonucleoprotein H2 OS=Homo sapiens GN=HNRNPH2 PE=1 SV=1 - [HNRH2_HUMAN]	4.732366
Q00534	Cyclin-dependent kinase 6 OS=Homo sapiens GN=CDK6 PE=1 SV=1 - [CDK6_HUMAN]	7.179366
Q59FU4	Diacylglycerol kinase (Fragment) OS=Homo sapiens PE=2 SV=1 - [Q59FU4_HUMAN]	6.612119
Q8NEW0	Zinc transporter 7 OS=Homo sapiens GN=SLC30A7 PE=2 SV=1 - [ZNT7_HUMAN]	7.054742
Q96ER3	Protein SAAL1 OS=Homo sapiens GN=SAAL1 PE=1 SV=2 - [SAAL1_HUMAN]	4.00743
Q96JJ7	Protein disulfide-isomerase TMX3 OS=Homo sapiens GN=TMX3 PE=1 SV=2 - [TMX3_HUMAN]	5.415393
Q96RI0	Protease-activated receptor 4 OS=Homo sapiens GN=F2RL3 PE=1 SV=3 - [PAR4_HUMAN]	5.048625
Q9BSY0	PTDSS1 protein (Fragment) OS=Homo sapiens GN=PTDSS1 PE=2 SV=2 - [Q9BSY0_HUMAN]	3.783625
Q9Y2V7	Conserved oligomeric Golgi complex subunit 6 OS=Homo sapiens GN=COG6 PE=1 SV=2 - [COG6_HUMAN]	100

Table S5 –Heavy/Light dataset for the Y157C proteome

Accession	Description	Heavy/Light
E9PNW8	Fatty acyl-CoA reductase 1 (Fragment) OS=Homo sapiens GN=FAR1 PE=1 SV=1 - [E9PNW8_HUMAN]	3.607271
Q86V85	Integral membrane protein GPR180 OS=Homo sapiens GN=GPR180 PE=2 SV=1 - [GP180_HUMAN]	3.904533

Table S6 –Medium/Light dataset for the Y157C proteome

Accession	Description	Medium/Light
A0A024RDA1	Exocyst complex component 1, isoform CRA_a OS=Homo sapiens GN=EXOC1 PE=4 SV=1 - [A0A024RDA1_HUMAN]	100
B3KS18	cDNA FLJ35285 fis, clone PROST2008079, highly similar to Golgi phosphoprotein 3 OS=Homo sapiens PE=2 SV=1 - [B3KS18_HUMAN]	100
B3KW34	Protein YIPF OS=Homo sapiens PE=2 SV=1 - [B3KW34_HUMAN]	5.127925
B4DXM0	cDNA FLJ58639, highly similar to Sterile alpha and TIR motif-containing protein 1 OS=Homo sapiens PE=2 SV=1 - [B4DXM0_HUMAN]	100
B4E2E5	cDNA FLJ56289, highly similar to Homo sapiens sperm specific antigen 2 (SSFA2), mRNA OS=Homo sapiens PE=2 SV=1 - [B4E2E5_HUMAN]	100

Q49AG2	TMED5 protein OS=Homo sapiens GN=TMED5 PE=2 SV=1 - [Q49AG2_HUMAN]	10.34607
Q9BRT6	Protein LLP homolog OS=Homo sapiens GN=LLPH PE=2 SV=1 - [LLPH_HUMAN]	8.086
Q9UMY1	Nucleolar protein 7 OS=Homo sapiens GN=NOL7 PE=1 SV=2 - [NOL7_HUMAN]	100

Table S7 –Medium/Heavy dataset for the Y157C proteome

Accession	Description	Medium/Heavy
A0A024RDA1	Exocyst complex component 1, isoform CRA_a OS=Homo sapiens GN=EXOC1 PE=4 SV=1 - [A0A024RDA1_HUMAN]	10.78327
B4DSL6	cDNA FLJ57190, highly similar to Actin-binding protein anillin OS=Homo sapiens PE=2 SV=1 - [B4DSL6_HUMAN]	2.85302
B4E2E5	cDNA FLJ56289, highly similar to Homo sapiens sperm specific antigen 2 (SSFA2), mRNA OS=Homo sapiens PE=2 SV=1 - [B4E2E5_HUMAN]	100
B5MD17	Chromobox protein homolog 1 (Fragment) OS=Homo sapiens GN=CBX1 PE=1 SV=1 - [B5MD17_HUMAN]	100
Q05BU6	SFRS11 protein (Fragment) OS=Homo sapiens GN=SFRS11 PE=2 SV=1 - [Q05BU6_HUMAN]	4.317
Q49AG2	TMED5 protein OS=Homo sapiens GN=TMED5 PE=2 SV=1 - [Q49AG2_HUMAN]	10.74034
Q8IYU8	Calcium uptake protein 2, mitochondrial OS=Homo sapiens GN=MICU2 PE=1 SV=2 - [MICU2_HUMAN]	3.777632
Q9BRT6	Protein LLP homolog OS=Homo sapiens GN=LLPH PE=2 SV=1 - [LLPH_HUMAN]	100
Q9UPU5	Ubiquitin carboxyl-terminal hydrolase 24 OS=Homo sapiens GN=USP24 PE=1 SV=3 - [UBP24_HUMAN]	10
Q9Y487	V-type proton ATPase 116 kDa subunit a isoform 2 OS=Homo sapiens GN=ATP6V0A2 PE=1 SV=2 - [VPP2_HUMAN]	100

Table S8 – Summary of the down- and up-regulated proteins found by volcano plots.

Gene name	log_2 fc	p value	Description	Change	Proteome	Ref.
C8orf33	1.11	0.002	UPF0488 protein C8orf33	Up	PAR4 (SLiM)	[600]
CCT3	1.10	0.038	T-complex protein 1 subunit gamma. Component of the chaperonin-containing	Up	PAR4 (SLiM)	[601,602]

			T-complex (TRiC), a molecular chaperone complex that assists the folding of proteins upon ATP hydrolysis. The TRiC complex mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance. As part of the TRiC complex may play a role in the assembly of BBSome, a complex involved in ciliogenesis regulating transports vesicles to the cilia. The TRiC complex plays a role in the folding of actin and tubulin.			
DNM2	1.07	0.017	Dynamin-2. Catalyzes the hydrolysis of GTP to mediate vesicle scission at plasma membrane during endocytosis and filament remodeling at many actin structures during organization of the actin cytoskeleton. Plays an important role in clathrin-mediated endocytosis (CME), exocytic and clathrin-coated vesicle from the trans-Golgi network, and PDGF stimulated macropinocytosis. During vesicular trafficking process, associates to the membrane, through lipid binding, and self-assembles into ring-like structure through oligomerization to form a helical polymer around the vesicle membrane and leading to vesicle scission.	Up	PAR4 (SLiM)	603,604

HACD3	1.43	0.023	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3. Catalyzes the third reaction of the long-chain fatty acids elongation cycle. This allows the addition of two carbons to the chain of long- and very long-chain fatty acids/VLCFAs per cycle. This enzyme catalyzes the dehydration of the 3-hydroxyacyl-CoA intermediate into trans-2,3-enoyl-CoA, within each cycle of fatty acid elongation. Thereby, it participates in the production of VLCFAs of different chain lengths that are involved in multiple biological processes as precursors of membrane lipids and lipid mediators. May be involved in Rac1-signalling pathways leading to the modulation of gene expression. Promotes insulin receptor/INSR autophosphorylation and is involved in INSR internalization.	Up	PAR4 (SLiM)	[605]
MDN1	1.06	0.004	Midasin. Nuclear chaperone required for maturation and nuclear export of pre-60S ribosome subunits.Subsequently recruited to the nucleoplasmic particles through interaction with SUMO-conjugated PELP1 complex.	Up	PAR4 (SLiM)	[606]
OXA1L	1.21	0.013	Mitochondrial inner membrane protein	Up	PAR4 (SLiM)	[607,608]

			OXA1L. Required for the insertion of integral membrane proteins into the mitochondrial inner membrane. Essential for the activity and assembly of cytochrome oxidase. Required for the correct biogenesis of ATP synthase and complex I in mitochondria.			
PGRMC2	1.37	0.030	Membrane-associated progesterone receptor component 2. Required for the maintenance of uterine histoarchitecture and normal female reproductive lifespan. May serve as a universal non-classical progesterone receptor in the uterus. Intracellular heme chaperone required for delivery of labile, or signalling heme, to the nucleus. Plays a role in adipocyte function and systemic glucose homeostasis. In brown fat, which has a high demand for heme, delivery of labile heme in the nucleus regulates the activity of heme-responsive transcriptional repressors such as NR1D1 and BACH1.	Up	PAR4 (SLiM)	[609]
PSMB2	1.21	0.025	Proteasome subunit beta type-2. Non-catalytic component of the 20S core proteasome complex. Associated with two 19S regulatory particles, forms the 26S proteasome and thus participates in the	Up	PAR4 (SLiM)	[610]

			ATP-dependent degradation of ubiquitinated proteins. Associated with the PA200 or PA28, the 20S proteasome mediates ubiquitin-independent protein degradation. This type of proteolysis is required in several pathways including spermatogenesis (20S-PA200 complex) or generation of a subset of MHC class I-presented antigenic peptides (20S-PA28 complex).			
PTGR3	6.64	0.034	Prostaglandin reductase 3. Functions as 15-oxo-prostaglandin 13-reductase and acts on 15-keto-PGE1, 15-keto-PGE2, 15-keto-PGE1-alpha and 15-keto-PGE2-alpha with highest efficiency towards 15-keto-PGE2-alpha. Overexpression represses transcriptional activity of PPARG and inhibits adipocyte differentiation.	Up	PAR4 (SLiM)	611
PYCR2	1.06	0.039	Pyrroline-5-carboxylate reductase 2. Housekeeping enzyme that catalyzes the last step in proline biosynthesis. In some cell types, such as erythrocytes, its primary function may be the generation of NADP+. Can utilize both NAD and NADP. Has higher affinity for NADP, but higher	Up	PAR4 (SLiM)	612,613

			catalytic efficiency with NADH.			
THEM6	1.38	0.007	Mesenchymal stem cell protein DSCD75. Thioesterase superfamily member 6.	Up	PAR4 (SLiM)	[600]
CPPED1	-1.14	0.033	Serine/threonine-protein phosphatase CPPED1. Protein phosphatase that dephosphorylates AKT family kinase specifically at 'Ser-473', blocking cell cycle progression and promoting cell apoptosis. May play an inhibitory role in glucose uptake by adipocytes.	Down	PAR4 (SLiM)	[614]
FSCN1	-1.24	0.008	Fascin. Actin-binding protein that contains 2 major actin binding sites. Organizes filamentous actin into parallel bundles. Plays a role in the organization of actin filament bundles and the formation of microspikes, membrane ruffles, and stress fibers. Important for the formation of a diverse set of cell protrusions, such as filopodia, and for cell motility and migration. Mediates reorganization of the actin cytoskeleton and axon growth cone collapse in response to NGF.	Down	PAR4 (SLiM)	[615–617]
ILKAP	-1.30	0.011	Integrin-linked kinase-associated serine/threonine phosphatase 2C. Protein phosphatase that may play a role in regulation of cell cycle progression via	Down	PAR4 (SLiM)	[618]

			dephosphorylation of its substrates whose appropriate phosphorylation states might be crucial for cell proliferation. Selectively associates with integrin linked kinase (ILK), to modulate cell adhesion and growth factor signalling. Inhibits the ILK-GSK3B signalling axis and may play an important role in inhibiting oncogenic transformation.			
MAGOHB	-1.09	0.044	Protein mago nashi homolog 2. Required for pre-mRNA splicing as component of the spliceosome. Plays a redundant role with MAGOH in the exon junction complex and in the nonsense-mediated decay (NMD) pathway.	Down	PAR4 (SLiM)	[619–621]
TBCC	-1.50	0.048	Tubulin-specific chaperone C. Tubulin-folding protein; involved in the final step of the tubulin folding pathway.	Down	PAR4 (SLiM)	[622]
UFC1	-1.04	0.0002	Ubiquitin-fold modifier-conjugating enzyme 1. E1-like enzyme which specifically catalyzes the second step in ufmylation. Accepts the ubiquitin-like modifier UFM1 from the E1 enzyme UBA5 and forms an intermediate with UFM1 via a thioester linkage. Ufmylation is involved in reticulophagy (also called ER-phagy) induced in response to	Down	PAR4 (SLiM)	[623–625]

			endoplasmic reticulum stress.			
C11orf54	1.01	0.011	Ester hydrolase C11orf54. Exhibits ester hydrolase activity on the substrate p-nitrophenyl acetate.	Up	SLiM	[626]
EPDR1	1.20	0.008	Mammalian ependymin-related protein 1. Binds anionic lipids and gangliosides at acidic pH.	Up	SLiM	[627]
FAM168A	1.22	0.039	Protein FAM168A. In cancer context, protects cells from induced-DNA damage and apoptosis. Acts, at least in part, through PI3K/AKT/NFKB signalling pathway and by preventing POLB degradation. Decreases POLB ubiquitation and stabilizes its protein levels.	Up	SLiM	[628]
NOLC1	1.14	0.041	Nucleolar and coiled-body phosphoprotein 1. Nucleolar protein that acts as a regulator of RNA polymerase I by connecting RNA polymerase I with enzymes responsible for ribosomal processing and modification. Required for neural crest specification: following monoubiquitination by the BCR(KBTBD8) complex, associates with TCOF1 and acts as a platform to connect RNA polymerase I with enzymes responsible for ribosomal processing and modification, leading to remodel the	Up	SLiM	[629–631]

			translational program of differentiating cells in favor of neural crest specification. Involved in nucleologenesis, possibly by playing a role in the maintenance of the fundamental structure of the fibrillar center and dense fibrillar component in the nucleolus			
SMU1	1.26	0.025	WD40 repeat-containing protein SMU1. Involved in pre-mRNA splicing as a component of the spliceosome. Regulates alternative splicing of the HSPG2 pre-mRNA. Required for normal accumulation of IK. Required for normal mitotic spindle assembly and normal progress through mitosis.	Up	SLiM	632,633
UFC1	1.04	0.0009	See above	Up	SLiM	623–625
ASPH	-6.64	0.035	Aspartyl/asparaginyl beta-hydroxylase. Isoform 1 Specifically hydroxylates an Asp or Asn residue in certain epidermal growth factor-like (EGF) domains of a number of proteins. Isoform 8 Membrane-bound Ca^{2+}-sensing protein, which is a structural component of the ER-plasma membrane junctions. Isoform 8 regulates the activity of Ca(+2) released-activated Ca(+2)	Down	SLiM	634,635

			(CRAC) channels in T-cells.			
ATP2A2	-1.74	0.038	Probable phospholipid-transporting ATPase IIA. Plays a role in regulating membrane trafficking of cargo proteins, namely endosome to plasma membrane recycling, probably acting through RAB5 and RAB11 activation. Also involved in endosome to trans-Golgi network retrograde transport. In complex with MON2 and DOP1B, regulates SNX3 retromer-mediated endosomal sorting of WLS, a transporter of Wnt morphogens in developing tissues. Participates in the formation of endosomal carriers that direct WLS trafficking back to Golgi, away from lysosomal degradation. Appears to be implicated in intercellular communication by negatively regulating the release of exosomes. The flippase activity towards membrane lipids and its role in membrane asymmetry remains to be proved. Required for the maintenance of neurite morphology and synaptic transmission.	Down	SLiM	[636–638]
ATP5PD	-1.57	0.0012	ATP synthase subunit d, mitochondrial. Mitochondrial membrane ATP synthase (F1F0 ATP	Down	SLiM	[639]

			synthase or Complex V) produces ATP from ADP in the presence of a proton gradient across the membrane which is generated by electron transport complexes of the respiratory chain.			
ATP6V1F	-1.58	0.040	V-type proton ATPase subunit F. Subunit of the V1 complex of vacuolar(H^+)-ATPase (V-ATPase), a multisubunit enzyme composed of a peripheral complex (V1) that hydrolyzes ATP and a membrane integral complex (V0) that translocates protons. V-ATPase is responsible for acidifying and maintaining the pH of intracellular compartments and in some cell types, is targeted to the plasma membrane, where it is responsible for acidifying the extracellular environment.	Down	SLiM	640
COA3	-1.55	0.038	Cytochrome c oxidase assembly factor 3 homolog, mitochondrial. Core component of the MITRAC (mitochondrial translation regulation assembly intermediate of cytochrome c oxidase complex) complex, that regulates cytochrome c oxidase assembly. MITRAC complexes regulate both translation of mitochondrial encoded	Down	SLiM	641

			components and assembly of nuclear-encoded components imported in mitochondrion. Required for efficient translation of MT-CO1 and mitochondrial respiratory chain complex IV assembly.			
F2RL3	-2.34	0.0006	Protease-activated receptor 4. Receptor for activated thrombin or trypsin coupled to G proteins that stimulate phosphoinositide hydrolysis. Plays a role in platelets activation.	Down	SLiM	[642]
HACD3	-1.43	0.0026	See above	Down	SLiM	[605]
HADHA	-1.17	0.017	Trifunctional enzyme subunit alpha, mitochondrial. It catalyzes the last three reactions of the mitochondrial beta-oxidation pathway, which is the major energy-producing process in tissues, breaking down fatty acids into acetyl-CoA. Mitochondrial trifunctional enzyme is a heterotetrameric complex composed of two proteins, the trifunctional enzyme subunit alpha/HADHA the 2,3-enoyl-CoA hydratase and the 3-hydroxyacyl-CoA dehydrogenase activities while the trifunctional enzyme subunit beta/HADHB bears the 3-ketoacyl-CoA thiolase activity. Independently of the subunit beta, the trifunctional enzyme subunit alpha/HADHA	Down	SLiM	[643–646]

			also has a monolysocardiolipin acyltransferase activity. It acylates monolysocardiolipin into cardiolipin, a major mitochondrial membrane phospholipid which plays a key role in apoptosis and supports mitochondrial respiratory chain complexes in the generation of ATP. HADHA allows the acylation of monolysocardiolipin with different acyl-CoA substrates including oleoyl-CoA for which it displays the highest activity.			
HADHB	-1.28	0.018	Same as HADHA, but subunit beta HADHB bears the 3-ketoacyl-CoA thiolase activity.	Down	SLiM	[643–646]
IPO5	-1.06	0.011	Importin-5. Functions in nuclear protein import as nuclear transport receptor. Serves as receptor for nuclear localization signals (NLS) in cargo substrates. Mediates docking of the importin/substrate complex to the nuclear pore complex (NPC) through binding to nucleoporin and the complex is subsequently translocated through the pore by a Ran-dependent mechanism. The directionality of nuclear import is thought to be conferred by an asymmetric distribution of the GTP- and GDP-	Down	SLiM	[647,648]

			bound forms of Ran between the cytoplasm and nucleus. Mediates the nuclear import of ribosomal proteins RPL23A, RPS7 and RPL5.			
IPO9	-1.83	0.006	Importin-9. Same as IPO5 but mediates the import of pre-assembled proteasomes into the nucleus; AKIRIN2 acts as a molecular bridge between IPO9 and the proteasome complex. Mediates the nuclear import of histones H2A, H2B, H4 and H4. In addition to nuclear import, also acts as a chaperone for histones by preventing inappropriate non-nucleosomal interactions.	Down	SLiM	[649–651]
KRTCAP2	-1.17	0.006	Keratinocyte-associated protein 2. Subunit of the oligosaccharyl transferase (OST) complex that catalyzes the first step in protein N-glycosylation. This occurs cotranslationally and the complex associates with the Sec61 complex at the channel-forming translocon complex that mediates protein translocation across the endoplasmic reticulum (ER). All subunits are required for a maximal enzyme activity. May be involved in N-glycosylation of APP (amyloid-beta precursor protein).	Down	SLiM	[652,653]

			Can modulate gamma-secretase cleavage of APP by enhancing endoprotelysis of PSEN1.			
MRPS16	-1.30	0.027	Small ribosomal subunit protein bS16m	Down	SLiM	600
SLC25A3	-1.07	0.002	Solute carrier family 25 member 3. Inorganic ion transporter that transports phosphate or copper ions across the mitochondrial inner membrane into the matrix compartment. Mediates proton-coupled symport of phosphate ions necessary for mitochondrial oxidative phosphorylation of ADP to ATP. Transports copper ions probably in the form of anionic copper complexes to maintain mitochondrial matrix copper pool and to supply copper for cytochrome C oxidase complex assembly. May also play a role in regulation of the mitochondrial permeability transition pore (mPTP).	Down	SLiM	654,655
TMX3	-2.43	0.034	Protein disulfide-isomerase TMX3. Probable disulfide isomerase, which participates in the folding of proteins containing disulfide bonds. May act as a dithiol oxidase. Acts as a regulator of endoplasmic reticulum-mitochondria contact sites via its ability to regulate redox signals.	Down	SLiM	656,657

TUBB2A	-1.48	0.048	Tubulin beta-2A chain. Tubulin is the major constituent of microtubules, a cylinder consisting of laterally associated linear protofilaments composed of alpha- and beta-tubulin heterodimers. Microtubules grow by the addition of GTP-tubulin dimers to the microtubule end, where a stabilizing cap forms. Below the cap, tubulin dimers are in GDP-bound state, owing to GTPase activity of alpha-tubulin.	Down	SLiM	[658]
LAGE3	1.03	0.033	Component of the EKC/KEOPS complex that is required for the formation of a threonylcarbamoyl group on adenosine at position 37 (t6A37) in tRNAs that read codons beginning with adenine. The complex is probably involved in the transfer of the threonylcarbamoyl moiety of threonylcarbamoyl-AMP (TC-AMP) to the N6 group of A37. LAGE3 functions as a dimerization module for the complex.	Up	PAR4 (Y157C)	[659,660]
PDLIM1	1.10	0.031	PDZ and LIM domain protein 1. Cytoskeletal protein that may act as an adapter that brings other proteins (like kinases) to the cytoskeleton. Involved in assembly, disassembly and directioning of stress fibers in fibroblasts.	Up	PAR4 (Y157C)	[661]

			Required for the localization of ACTN1 and PALLD to stress fibers. Required for cell migration and in maintaining cell polarity of fibroblasts.			
PCNT	-1.09	0.048	Pericentrin. Integral component of the filamentous matrix of the centrosome involved in the initial establishment of organized microtubule arrays in both mitosis and meiosis. Plays a role, together with DISC1, in the microtubule network formation. Is an integral component of the pericentriolar material (PCM). May play an important role in preventing premature centrosome splitting during interphase by inhibiting NEK2 kinase activity at the centrosome.	Down	PAR4 (Y157C)	[662]
CETN2	1.16	0.033	Centrin-2. Plays a fundamental role in microtubule organizing center structure and function. Required for centriole duplication and correct spindle formation. Has a role in regulating cytokinesis and genome stability via cooperation with CALM1 and CCP110. Involved in global genome nucleotide excision repair (GG-NER) by acting as component of the XPC complex. Cooperatively with RAD23B appears to stabilize XPC.	Up	Y157C	[663]

			Stimulates DNA binding of the XPC:RAD23B dimer. As a component of the TREX-2 complex, involved in the export of mRNAs to the cytoplasm through the nuclear pores.			
EXOC1	-3.43	0.040	Exocyst complex component 1. Component of the exocyst complex involved in the docking of exocytic vesicles with fusion sites on the plasma membrane. Has an antiviral effect against flaviviruses by affecting viral RNA transcription and translation through the sequestration of elongation factor 1-alpha (EEF1A1). This results in decreased viral RNA synthesis and decreased viral protein translation.	Down	Y157C	[664]
LAGE3	-1.03	0.041	See above	Down	Y157C	[659,660]
RABL3	-1.83	0.015	Rab-like protein 3. Required for KRAS signalling regulation and modulation of cell proliferation. Regulator of KRAS prenylation, and probably prenylation of other small GTPases.	Down	Y157C	[665]

Table S9 – Mitochondrial proteins identified by network clustering on STRING

Proteome	Term	FDR	Group	Proteins
PAR4 (SLiM)	Mitochondrial translation elongation	0.0108	STRING Clusters	MRPL33, MRPS11, RPLP2
PAR4 (SLiM)	Inner mitochondrial membrane protein complex	7.59E-14	COMPARTMENTS	ATP5F1A, ATP5F1B, ATP5F1D, ATP5MD, NDUFA2,

				NDUFA13, NDUFB3, NDUFB4, NDUFS7
PAR4 (SLiM)	Mitochondrial Envelope	3.0E-5	COMPARTMENTS	SLC25A5 VDAC1 VDAC3 DIABLO-2 CKMT1B TOMM22
SLiM	Mitochondrion	8.01E-14	GO Cellular Compartments	SLC25A1, SLC25A6, SLC25A11, IDH3G, PDHB, CS, HADHA, HADBH, PRDX6, SAMM50, TIMM50, HSPD1, VDAC1/3, SLC25A3, SLC25A5
SLiM	Formation of ATP by chemiosmotic coupling	1.33E-8	Reactome Pathways	ATP5PB, ATP5MG, ATP5PD, ATP5F1C, AFG3L2
SLiM	Thermogenesis	6.08E-6	KEGG Pathways	NDUFA9, NDUFAF3, NDUFB3, NDUFV1, NDUFA4, COA3
SLiM	Mitochondrial ribosome	7.43E-7	GO Cellular Components	MRPL14, MRPS16, MRPL10, MRPL20
PAR4 (Y157C)	Mitochondrial respiratory chain complex 1 assembly	3.15E-7	GO Biological Process	NDUFB4, NDUFS3, NDUFS5, NDUFAF7, UQCRC2, COX5A, CISD2 ECSIT
PAR4 (Y157C)	Mitochondrial translation	8.9E-5	GO Biological Process	GFM2, FH, IDH3B, CKMTIB, PPP2CA, GSTP1,

				CLIC1, DDX28, KRT19, PRPSAP1, STOML2, POLDIP2
Y157C	Methylglyoxal metabolic process	0.0291	GO Biological Process	PRDX6, CLIC1, TPI1, CKB, HAGH
Y157C	Other mitochondrial proteins	N/A	N/A	ENDOG, SACS, HSPA4, PXDN, KRT19. NDUFA8, NDUFS3, COX5A, UQCRFS1, ATP5PD, MRPL13, MRPS2, MRPS5, MRPS34, RPS27A, MPV17L2, MRM3, ECH1, NGDN, AASDHPPT, HSP90AA1, HSP90AB1, PDE12, SETD3, SFXN1, PGD, NNT, OAT, DDAH2, HARS1, PARS2, NAXE, GART

Table S10 – PrePPI predictions.

LR: Likelihood ratio. SM: Structural modelling score. PrP: Protein peptide LR. PR: Partner redundancy LR. OR: Orthology LR. PP: Phylogenetic profile LR. EP: Expression profile LR. GO: Gene ontology LR.

UniProt	Gene	SM	PrP	PR	OR	PP	EP	GO	Total LR

Q99705	MCHR1 GPR24 SLC1	21.6				2.4	3.9	69.6	14163.6
P06239	LCK	0.4	20.7			2.4	1.6	105.2	8332.9
P47900	P2RY1	21.6				2.4	2.1	69.6	7653
P37288	AVPR1A AVPR1	21.6				2.4	2.1	69.6	7653
P00734	F2		7.4			2.4	2.1	181	6825.7
P22888	LHCGR LCGR LGR2 LHRHR	6.9				2.4	2.1	181	6341.6
P11229	CHRM1	9				2.4	3.9	69.6	5891.1
O43614	HCRTR2	9				2.4	3.9	69.6	5891.1
P16885	PLCG2	0.7	20.7			2.4	1.6	69.6	5515.4
P46094	XCR1 CCXCR1 GPR5	6.9				2.4	2.8	105.2	4844.8
P28336	NMBR	7.4				2.4	3.9	69.6	4841.5
P25116	F2R CF2R PAR1 TR	6.9				2.4	1.6	181	4773.1
P47901	AVPR1B AVPR3 VPR3	6.9				2.4	3.9	69.6	4513.5
P14416	DRD2	6.9					5.7	105.2	4127.3
P32239	CCKBR CCKRB	4.1				2.4	5.7	69.6	3950.4
Q9NSD7	RXFP3 GPCR135 RLN3R1 SALPR	4.1				2.4	5.7	69.6	3950.4
P41143	OPRD1 OPRD	27.3					5.7	24.1	3743.5
P25103	TACR1 NK1R TAC1R	6.9				2.4	2.8	69.6	3206.7
P21728	DRD1	3.2				2.4	2.1	181	2912.7
P41146	OPRL1 OOR ORL1	27.3				2.4	3.9	10.8	2778.7
P51582	P2RY4 NRU	6.9					5.7	69.6	2723.8

Q9HB89	NMUR1 GPR66	4.1				2.4	3.9	69.6	2709.4
P41231	P2RY2 P2RU1	6.9				2.4	2.1	69.6	2438.7
Q9UBY5	LPAR3 EDG7 LPA3	9				2.4	1.6	69.6	2388.9
P29475	NOS1	0.2	3.5			2.4	3.9	69.6	2315.5
P28335	HTR2C HTR1C	3.2				2.4	2.8	105.2	2225.2
P35372	OPRM1 MOR1	21.6				2.4	3.9	10.8	2197.5
Q6DWJ6	GPR139 GPRG1 PGR3	3.2				2.4	3.9	69.6	2073.1
P32238	CCKAR CCKRA	6.9				2	2.1	69.6	2032
O14842	FFAR1 GPR40	1.2				2.4	5.7	105.2	1763.1
P28223	HTR2A HTR2	3.2				2.4	2.1	105.2	1697.7
Q9UKP6	UTS2R GPR14	4.1					5.7	69.6	1635.1
O95665	NTSR2	4.1					5.7	69.6	1635.1
P31947	SFN HME1	0.3	72.8			2.4	1.6	5.9	1630.5
Q96CA5	BIRC7 KIAP LIVIN MLIAP RNF50 UNQ5800/PRO19607/PRO21344	0.2	57.2			2.4	5.7	2.1	1627
P35462	DRD3	6.9				2.4	3.9	24.1	1563.3
P12931	SRC SRC1	0.2	20.7			2.4	2.8	10.8	1499.3
P08913	ADRA2A ADRA2R ADRAR	3.2				2.4	2.8	69.6	1472.8
Q9UKW4	VAV3	0.5	20.7			2.4	1.2	24.1	1419.9
O60674	JAK2	0.2	20.7			2.4	1.2	24.1	1419.9
Q9H244	P2RY12 HORK3	6.9				2.4	1.2	69.6	1364.3
P04216	THY1	0.3	2.6			2.4	1.2	181	1337.8

P31749	AKT1 PKB RAC	0.2	5			2.4	1.6	69.6	1326.1
Q15554	TERF2 TRBF2 TRF2	0.2	5			2.4	1.6	69.6	1326.1
P21918	DRD5 DRD1B DRD1L2	3.2					5.7	69.6	1255
P46663	BDKRB1 BRADYB1	4.1					1.6	181	1182.5
Q92835	INPP5D SHIP SHIP1		20.7			2.4	2.1	10.8	1140.2
P21554	CNR1 CNR	3.2				2.4	2.1	69.6	1123.7
Q15722	LTB4R BLT BLT1 BLTR CMKRL1 GPR16 P2RY7	3.2				2.4	2.1	69.6	1120.1
P08172	CHRM2	3.2				2.4	2.1	69.6	1120.1
P31751	AKT2		4.2			2.4	1.6	69.6	1104.7
Q6QNY0	BLOC1S3 BLOS3						5.7	181	1029.5
P46095	GPR6	6.9				2.4	5.7	10.8	1024
P56373	P2RX3					2.4	3.9	105.2	991.2
P30556	AGTR1 AGTR1A AGTR1B AT2R1 AT2R1B	9					1.6	69.6	988.8
P01178	OXT OT					2.4	5.7	69.6	956.5
P01185	AVP ARVP VP					2.4	5.7	69.6	956.5
P35367	HRH1	3.4				2.4	1.6	69.6	915.1

Q96F15	GIMAP5 IAN4L1 IAN5 IMAP3	0.3	2.6			2	1.6	105.2	871.3
P51692	STAT5B	0.4	20.7			2.4	1.6	10.8	855.7
Q9HBW0	LPAR2 EDG4 LPA2	6.9				2.4	2.1	24.1	847.1
Q9H6Q3	SLA2 C20orf156 SLAP2	0.2	20.7			2.4	2.8	5.9	813.5
P25025	CXCR2 IL8RB	7.4					1.6	69.6	812.6
Q9P2W3	GNG13					2	5.7	69.6	797
O15392	BIRC5 API4 IAP4	0.2	57.2			2	1.2	5.9	793.3
O00329	PIK3CD	0.4	6.3			2.4	2.1	24.1	778.9
P30518	AVPR2 ADHR DIR DIR3 V2R	7.4				2.4	3.9	10.8	751.2
P24530	EDNRB ETRB	9					1.2	69.6	734.9
Q6P9H4	CNKSR3 MAGI1	0.6	3.5			2.4	1.2	69.6	697.9
Q5T2W1	PDZK1 CAP70 NHERF3 PDZD1		3.5			2.4	1.2	69.6	697.9
P32418	SLC8A1 CNC NCX1	0.6	2.6			2.4	1.6	69.6	692.2
O95477	ABCA1 ABC1 CERP					2.4	1.6	181	691.8
Q9H267	VPS33B		2.6			2	0.7	181	672.5
P25101	EDNRA ETA ETRA	3.2					1.2	181	672.5
Q09013	DMPK DM1PK MDPK		4.2			2.4	2.8	24.1	670.4
P06307	CCK					2.4	3.9	69.6	656

Q15077	P2RY6 PP2891	3.3					2.8	69.6	644.3
P42224	STAT1	0.6	20.7			2.4	1.2	10.8	636
P06241	FYN	0.2	20.7			2.4	1.2	10.8	636
P42229	STAT5A STAT5	0.7	20.7			2.4	1.2	10.8	636
P25105	PTAFR PAFR	6.9				2.4	1.6	24.1	635.8
P43403	ZAP70 SRK	0.2	20.7			2.4	2.1	5.9	618.7
P35348	ADRA1A ADRA1C	3.2					2.8	69.6	611.5
Q9GZQ4	NMUR2 NMU2R TGR1	4.1					2.1	69.6	605.9
Q9UBL9	P2RX2 P2X2						5.7	105.2	598.2
Q9Y5Y4	PTGDR2 CRTH2 DL1R GPR44	7.4				2.4	5.7	5.9	594.3
Q14184	DOC2B DOC2BL	0.5	2.6			2.4	3.9	24.1	591.2
O75747	PIK3C2G	0.5	6.3			2.4	1.6	24.1	584.6
O00443	PIK3C2A	0.5	6.3			2.4	1.6	24.1	584.6
P00748	F12	0.4	7.4	0.5		2.4	5.7	10.8	560.5
P28222	HTR1B HTR1DB	6.9				2.4	5.7	5.9	555.6
O43603	GALR2 GALNR2	9					5.7	10.8	551.6
Q8NDV2	GPR26	16.5					5.7	5.9	550.9
P35225	IL13 NC30					2	3.9	69.6	546.6
P41235	HNF4A HNF4 NR2A1 TCF14	0.2	7.4			2.4	2.8	10.8	536.3
P04054	PLA2G1B PLA2 PLA2A PPLA2					2.4	2.1	105.2	535.5

P51684	CCR6 CKRL3 CMKBR6 GPR29 STRL22	7.4				2.4	2.8	10.8	533.7
P51679	CCR4 CMKBR4	7.4				2.4	2.8	10.8	533.7
P17787	CHRNB2	0.3	3.5			2.4	5.7	10.8	523.8
P43681	CHRNA4 NACRA4	0.3	3.5			2.4	5.7	10.8	523.8
P35346	SSTR5	9.3				2.4	3.9	5.9	515.3
P10124	SRGN PRG PRG1					2.4	1.2	181	514.2
Q05513	PRKCZ PKC2	0.2	4.2			2.4	2.1	24.1	509.8
P55055	NR1H2 LXRB NER UNR		7.4			2.4	1.2	24.1	507.9
Q13133	NR1H3 LXRA	0.2	7.4			2.4	1.2	24.1	507.9
P55211	CASP9 MCH6		12.2			2.4	1.6	10.8	503.5
P34981	TRHR	6.9				2.4	2.8	10.8	497.5
P30874	SSTR2	9				2.4	3.9	5.9	495.9
Q8NEB9	PIK3C3 VPS34	0.5	6.3			2	1.6	24.1	487
O00305	CACNB4 CACNLB4	0.5	5.2			2.4	1.6	24.1	481
P30679	GNA15 GNA16	0.5	2.6		0.6	2	2.1	69.6	470.7
O14788	TNFSF11 OPGL RANKL TRANCE					2.4	2.8	69.6	466.1
P18054	ALOX12 12LO LOG12					2.4	2.8	69.6	466.1

P07550	ADRB2 ADRB2R B2AR	3.2					2.1	69.6	463.6
Q969F8	KISS1R AXOR12 GPR54	6.9				2	5.7	5.9	461.6
P00533	EGFR ERBB ERBB1 HER1	0.2	3.7			2.4	2.1	24.1	451.2
Q99698	LYST CHS CHS1	1.5	1.3				1.6	181	435.2
P48736	PIK3CG	0.5	6.3			2.4	1.2	24.1	434.5
P42336	PIK3CA	0.2	6.3			2.4	1.2	24.1	434.5
Q13796	SHROOM2 APXL	0.5	3.5			2.4	2.1	24.1	433.3
P21917	DRD4	3.2				1	5.7	24.1	429.4
Q13554	CAMK2B CAM2 CAMK2 CAMKB		5			2	3.9	10.8	422.7
Q9UEW8	STK39 SPAK	0.5	4.2			2	0.7	69.6	414.7
P29992	GNA11 GA11	3	2.6		0.6	2	1.6	69.6	411.9
P32745	SSTR3	7.4					9.5	5.9	409.5
Q03181	PPARD NR1C2 PPARB		7.4			2.4	2.1	10.8	407.8
P03951	F11		7.4			2.4	2.1	10.8	407.3
O14492	SH2B2 APS	0.3	20.7			2.4	3.9	2.1	404.5
P41181	AQP2						5.7	69.6	395.9
Q8IYX4	DND1 RBMS4	0.2	42			2.4	3.9		395.4
P41595	HTR2B	3.2					1.2	105.2	390.7
Q92633	LPAR1 EDG2 LPA1	5.7				2.4	1.2	24.1	386.9
Q06124	PTPN11 PTP2C SHPTP2	0.2	20.7			2	1.6	5.9	386.9
Q9UPR5	SLC8A2 KIAA1087 NCX2	0.4	2.6			2.4	5.7	10.8	386.1

P21730	C5AR1 C5AR C5R1	4.1				2.4	1.6	24.1	380.5
P51685	CCR8 CKRL1 CMKBR8 CMKBRL2	6.9				2.4	2.1	10.8	378.4
Q99500	S1PR3 C9orf108 C9orf47 EDG3	6.9				2.4	2.1	10.8	378.4
P28566	HTR1E	16.5					3.9	5.9	377.9
P19235	EPOR	0.7	3.7			2.4	3.9	10.8	372.5
P42356	PI4KA PIK4 PIK4CA		6.3			2	1.2	24.1	362
Q9GZZ6	CHRNA10 NACHRA10	0.3	3.5			2.4	3.9	10.8	359.2
P61981	YWHAG	0.2	72.8			2	1.2	2.1	356.7
Q9BSW2	CRACR2A EFCAB4B RAB46					2.4	2.1	69.6	354.5
P10415	BCL2					2.4	2.1	69.6	354.5
P41594	GRM5 GPRC1E MGLUR5	0.5				2.4	3.9	69.6	354.3
Q96P47	AGAP3 CENTG3	9	2.6			2.4	2.8	5.9	352.3
Q99490	AGAP2 CENTG1 KIAA0167	9	2.6			2.4	2.8	5.9	352.3

Q6ZUM4	ARHGAP27 CAMGAP1 SH3D20 PP905	0.4	6.4			2.4	2.1	10.8	351.3
P35243	RCVRN RCV1	0.2	2.6			2	2.8	24.1	350
P50052	AGTR2	6.9					2.1	24.1	349.6
P32245	MC4R	3.2					1.6	69.6	349
Q06187	BTK AGMX1 ATK BPK	0.5	20.7			2.4	1.2	5.9	345.1
P43405	SYK	0.2	20.7			2.4	1.2	5.9	345.1
P27986	PIK3R1 GRB1	0.2	20.7			2.4	1.2	5.9	345.1
P15498	VAV1 VAV	0.4	20.7			2.4	1.2	5.9	345.1
P09471	GNAO1	6.9	2.6		0.6	1.6	2.1	24.1	343.7
P46721	SLCO1A2 OATP OATP1 OATP1A2 SLC21A3	0.4	42.2			2.4	1.6	2.1	333.6
O00308	WWP2	0.5	42.2			2.4	1.6	2.1	333.6
O60391	GRIN3B					2.4	5.7	24.1	331.3
O00750	PIK3C2B	0.6	6.3			2.4	0.9	24.1	328.5
Q9NZU5	LMCD1	0.2	2.6			2.4	2.1	24.1	319.5
Q8IV63	VRK3		4.2			2	1.6	24.1	318.8
P48995	TRPC1 TRP1					2.4	0.7	181	311.6
P61006	RAB8A MEL RAB8	6.9	2.6			1.6	1.2	24.1	311.6
Q7Z3S7	CACNA2D4	3				2.4	3.9	10.8	307.4
P11473	VDR NR1I1		7.4			2.4	1.6	10.8	306.1
P51677	CCR3 CMKBR3	7.4				2.4	1.6	10.8	304.6

O60502	OGA HEXC KIAA0679 MEA5 MGEA5					2.4	1.2	105.2	298.7
Q99572	P2RX7					2.4	1.2	105.2	298.7
Q93086	P2RX5 P2X5					2.4	1.2	105.2	298.7
P22736	NR4A1 GFRP1 HMR NAK1		7.4			2.4	2.8	5.9	291
Q15466	NR0B2 SHP		7.4			2.4	2.8	5.9	291
Q13936	CACNA1C CACH2 CACN2 CACNL1A1 CCHL1A1	0.6				2.4	2.8	69.6	289
Q9UBF8	PI4KB PIK4CB		6.3			1.6	1.2	24.1	287.3
P00747	PLG	0.2	5.2			2.4	2.1	10.8	287.1
Q02641	CACNB1 CACNLB1		5.2			2.4	2.1	10.8	287.1
P37840	SNCA NACP PARK1						1.6	181	286.3
Q5SQS7	SH2D4B		20.7			2.4	5.7		284.9
O14965	AURKA AIK AIRK1 ARK1 AURA AYK1 BTAK IAK1 STK15 STK6		4.2			2.4	1.2	24.1	284.4
O60566	BUB1B BUBR1 MAD3L SSK1	0.2	4.2			2.4	1.2	24.1	284.4

Q07001	CHRND ACHRD	0.3	3.5			2.4	5.7	5.9	284.2
Q9BXM0	PRX KIAA1620		3.5			2.4	5.7	5.9	282.6
Q5VY43	PEAR1 MEGF12	0.2	42.2			2.4	2.8		282.4
Q5VZ18	SHE		42.2			2.4	2.8		282.4
Q96PZ7	CSMD1 KIAA1890 UNQ5952/PRO19863	0.4	42			2.4	2.8		280.9
Q9UJD0	RIMS3 KIAA0237	0.5	42			2.4	2.8		280.9
Q7Z408	CSMD2 KIAA1884	0.4	42			2.4	2.8		280.9
P30542	ADORA1	1					3.9	69.6	277
Q8IV16	GPIHBP1 HBP1					2	5.7	24.1	276

Table S11 – Prediction scores for the binding of PAR4 SLiM with NOS1

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	97.05	0.88	0.87	$L^{383} – V^{32}$ $L^{383} – V^{32}$ $Q^{385} – F^{30}$ $Q^{385} – F^{30}$	L^{28}, F^{30}, L^{31}, V^{32}, I^{47}, Y^{77}, L^{81}, L^{84}
2	95.40	0.86	0.74	$S^{381} – Y^{77}$ $L^{383} – V^{32}$ $L^{383} – V^{32}$ $Q^{385} – F^{30}$ $Q^{385} – F^{30}$	L^{28}, F^{30}, L^{31}, V^{32}, I^{47}, Y^{77}, L^{81}, L^{84}, I^{87}
3	92.65	0.85	0.47	$L^{383} – V^{32}$ $L^{383} – V^{32}$ $L^{384} – F^{30}$	L^{28}, F^{30}, L^{31}, V^{32}, I^{47}, Y^{77}, L^{81}, L^{84}
4	93.20	0.85	0.47	$L^{383} – V^{32}$ $L^{383} – V^{32}$ $L^{384} – F^{30}$	L^{28}, F^{30}, L^{31}, V^{32}, I^{47}, Y^{77}, L^{81}, L^{84}
5	93.23	0.86	0.45	$S^{381} – Y^{77}$ $L^{383} – V^{32}$ $L^{383} – V^{32}$ $Q^{385} – F^{30}$	L^{28}, F^{30}, L^{31}, V^{32}, I^{47}, Y^{77}, L^{81}, L^{84}

				Q^{385} – F^{30}	

Table S12 – Prediction scores for the binding of PAR4 SLiM with CNKSR3

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	94.09	0.86	0.83	L^{383} – I^{227} L^{383} – I^{227} Q^{385} – M^{225} Q^{385} – M^{225}	L^{223}, M^{225}, Y^{226}, I^{227}, L^{271}, L^{274}, V^{275}, L^{278}
2	87.00	0.78	0.62	S^{382} – I^{227} S^{382} – I^{227} L^{384} – M^{225}	L^{223}, M^{225}, Y^{226}, I^{227}, V^{258}, L^{271}, L^{274}, V^{275}, L^{278}, L^{287}
3	87.06	0.77	0.54	S^{382} – I^{227} S^{382} – I^{227} L^{384} – M^{225}	I^{217}, L^{223}, M^{225}, Y^{226}, I^{227}, L^{271}, L^{274}, V^{275}, L^{278}
4	90.75	0.83	0.49	S^{382} – I^{227} S^{382} – I^{227} L^{384} – M^{225}	I^{217}, L^{223}, M^{225}, Y^{226}, I^{227}, L^{271}, L^{274}, V^{275}, L^{278}, V^{285}
5	86.80	0.78	0.43	S^{382} – I^{227} L^{384} – M^{225}	I^{217}, L^{223}, M^{225}, Y^{226}, I^{227}, V^{258}, L^{271}, L^{274}, V^{275}, L^{278}, L^{287}

Table S13 – Prediction scores for the binding of PAR4 SLiM with NHERF3 PDZ domain 1

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	91.81	0.83	0.48	S^{381} – H^{68} S^{381} – H^{68} L^{383} – L^{24} Q^{385} – F^{22} Q^{385} – F^{22} Q^{385} – V^{72}	Y^{20}, F^{22}, F^{23}, L^{24}, M^{69}, V^{72}, V^{75}
2	91.73	0.83	0.47	S^{382} – L^{24}	Y^{20}, F^{22}, F^{23}, L^{24}, M^{69}, V^{72}, V^{75}, L^{84}
3	90.30	0.81	0.34	S^{382} – H^{68} L^{384} – L^{24} Q^{385} – R^{68} Q^{385} – R^{68}	Y^{20}, F^{22}, F^{23}, L^{24}, M^{69}, V^{72}, V^{75}
4	89.68	0.81	0.33	S^{382} – H^{68} L^{384} – L^{24} L^{384} – L^{24} Q^{385} – R^{68} Q^{385} – R^{68}	Y^{20}, F^{22}, F^{23}, L^{24}, M^{69}, V^{72}, V^{75}
5	89.89	0.82	0.29	S^{382} – H^{68} S^{382} – H^{68}	Y^{20}, F^{22}, F^{23}, L^{24}, M^{69}, V^{72}, V^{75}

				Q^{385} – R^{68} Q^{385} – R^{68}	

Table S14 – Prediction scores for the binding of PAR4 SLiM with NHRF3 PDZ domain 2

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	92.17	0.84	0.38	S^{382} – L^{149} Q^{385} – S^{144}	Y^{145}, F^{147}, L^{149}, V^{197}, V^{200}
2	91.69	0.85	0.37	Q^{385} – F^{147} Q^{385} – F^{147}	Y^{145}, F^{147}, L^{149}, V^{197}, V^{200}
3	90.88	0.84	0.31	S^{381} – D^{162} S^{382} – L^{149}	Y^{145}, F^{147}, L^{149}, V^{197}, V^{200}
4	90.65	0.82	0.28	S^{381} – S^{148} S^{382} – H^{193}	Y^{145}, F^{147}, L^{149}, V^{197}, V^{200}
5	90.56	0.83	0.25	S^{382} – L^{149} S^{382} – S^{148} L^{384} – H^{193} L^{384} – L^{149}	Y^{145}, F^{147}, L^{149}, V^{197}, V^{200}

Table S15 – Prediction scores for the binding of PAR4 SLiM with NHRF3 PDZ domain 3

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	90.04	0.82	0.47	S^{382} – L^{257} S^{382} – L^{257} L^{384} – F^{255}	Y^{253}, F^{255}, Y^{256}, L^{257}, V^{305}, I^{308}
2	89.05	0.82	0.45	S^{382} – H^{301} S^{382} – H^{301} S^{382} – L^{257} Q^{385} – Y^{253}	Y^{253}, F^{255}, Y^{256}, L^{257}, V^{305}, I^{308}
3	89.54	0.82	0.43	S^{382} – L^{257} L^{384} – Y^{253} L^{384} – F^{255}	Y^{253}, F^{255}, Y^{256}, L^{257}, V^{305}, I^{308}
4	89.71	0.82	0.41	N/A	Y^{253}, F^{255}, Y^{256}, L^{257}, V^{305}, I^{308}
5	90.11	0.83	0.40	S^{382} – L^{257} L^{384} – F^{255}	Y^{253}, F^{255}, Y^{256}, L^{257}, V^{305}, I^{308}

Table S16 – Prediction scores for the binding of PAR4 SLiM with NHRF3 PDZ domain 4

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	93.18	0.85	0.73	L^{383} – L^{392} L^{383} – L^{392} Q^{385} – F^{390}	Y^{388}, F^{390}, L^{392}, Y^{436}, V^{440}, I^{443}

2	92.81	0.85	0.62	S^{382} – L^{392} S^{382} – L^{392} L^{394} – F^{390}	Y^{388}, F^{390}, L^{392}, Y^{436}, V^{440}, I^{443}
3	89.81	0.83	0.41	L^{383} – L^{392} L^{383} – L^{392} Q^{385} – F^{390} Q^{385} – V^{440}	Y^{388}, F^{390}, L^{392}, Y^{436}, V^{440}, I^{443}
4	90.99	0.84	0.40	S^{381} – L^{392} S^{382} – L^{392} L^{384} – Y^{388} L^{384} – F^{390}	Y^{388}, F^{390}, L^{392}, Y^{436}, V^{440}, I^{443}
5	89.58	0.82	0.33	S^{382} – Y^{436} S^{382} – Y^{436} S^{382} – Y^{436} S^{382} – L^{392}	Y^{388}, F^{390}, L^{392}, Y^{436}, V^{440}, I^{443}

Table S17 – Prediction scores for the binding of PAR4 SLiM with PRX

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	84.88	0.70	0.46	S^{382} – V^{35} S^{382} – V^{35} L^{384} – I^{32} Q^{385} – S^{30}	I^{19}, V^{21}, V^{29}, I^{32}, V^{34}, I^{42}, V^{44}, L^{57}, L^{59}, L^{65}
2	82.08	0.67	0.30	N/A	N/A
3	85.32	0.71	0.28	N/A	N/A
4	76.29	0.64	0.16	L^{383} – L^{84}	L^{81}, L^{84}
5	67.88	0.53	0.16	N/A	N/A

Table S18 – Prediction scores for the binding of PAR4 SLiM with SHROOM2

Model nr	Mean pLDDT	pTM	ipTM	H bonds	Hydrophobic interactions
1	94.22	0.83	0.84	L^{383} – L^{41} L^{383} – L^{41} Q^{385} – F^{39} Q^{385} – F^{39}	W^{37}, F^{39}, L^{41}, I^{90}, V^{93}
2	92.03	0.83	0.55	L^{383} – L^{41} L^{383} – L^{41} Q^{385} – F^{39} Q^{385} – F^{39}	W^{37}, F^{39}, L^{41}, I^{90}, V^{93}
3	89.94	0.81	0.45	S^{382} – R^{86} S^{382} – R^{86} L^{384} – F^{39} Q^{385} – F^{39}	W^{37}, F^{39}, L^{41}, I^{90}, V^{93}
4	90.98	0.82	0.39	S^{382} – L^{41} L^{384} – F^{39}	W^{37}, F^{39}, L^{41}, I^{64}, I^{77}, A^{84}, I^{90}, V^{93}
5	87.46	0.78	0.39	S^{382} – T^{40} Q^{385} – K^{94} Q^{385} – K^{94}	W^{37}, F^{39}, L^{41}, I^{77}, A^{84}, I^{90}, V^{93}

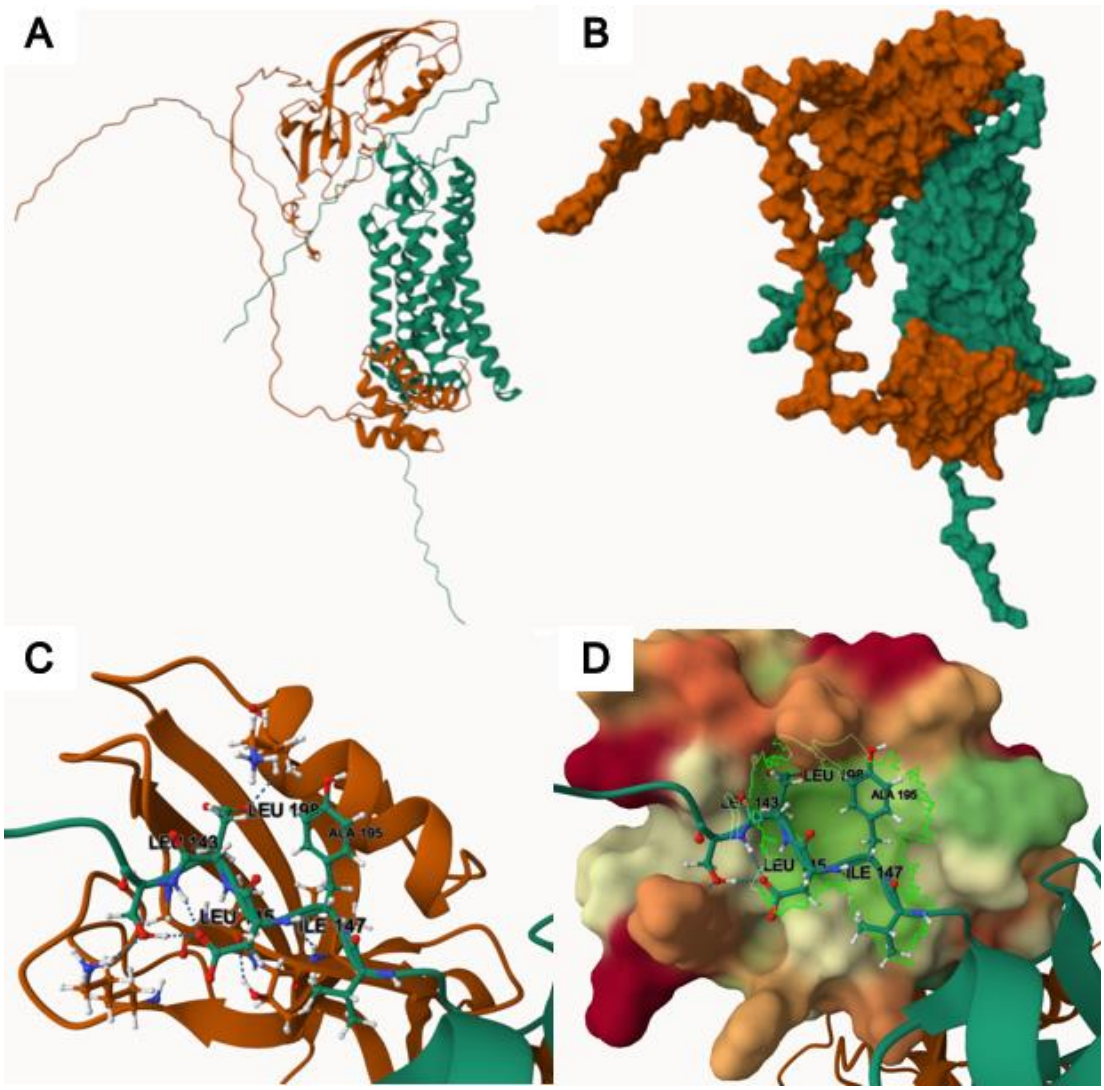


Figure S1 – Second AlphaFold model of the prediction of PAR4 interacting with GIPC1. A: Ribbon diagram of the receptor bound to GIPC1. PAR4 is shown in green, GIPC1 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 N-terminus and a conserved series of residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: PAR4 binds a hydrophobic pocket on PDZ domain, residues are colour coded based on their hydrophobicity, from red being very hydrophilic to green being very hydrophobic. Image generated using Mol*.

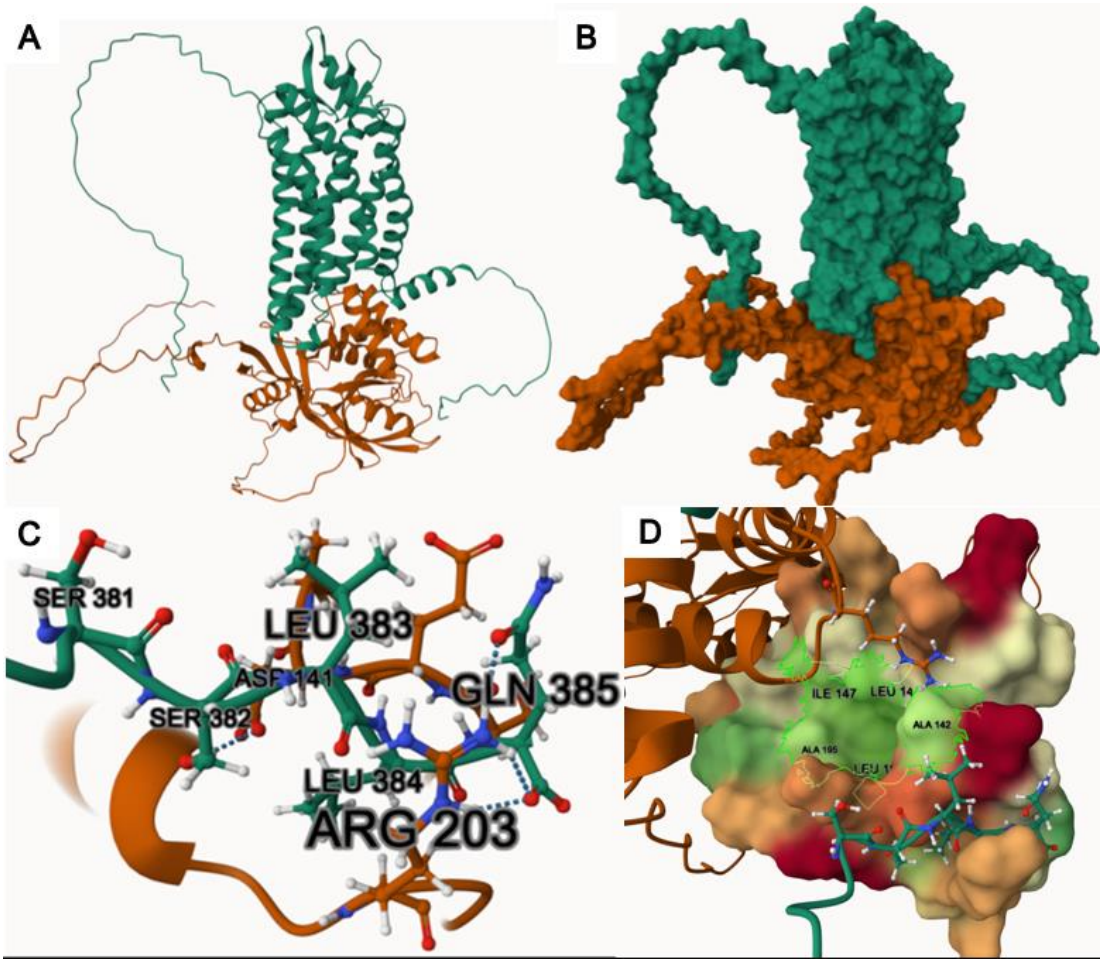


Figure S2 – Third AlphaFold model of the prediction of PAR4 interacting with GIPC1. A: Ribbon diagram of the receptor bound to GIPC1. PAR4 is shown in green, GIPC1 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a the PDZ domain, however the binding does not occur in the hydrophobic pocket, which is highlighted in green. Image generated using Mol*.

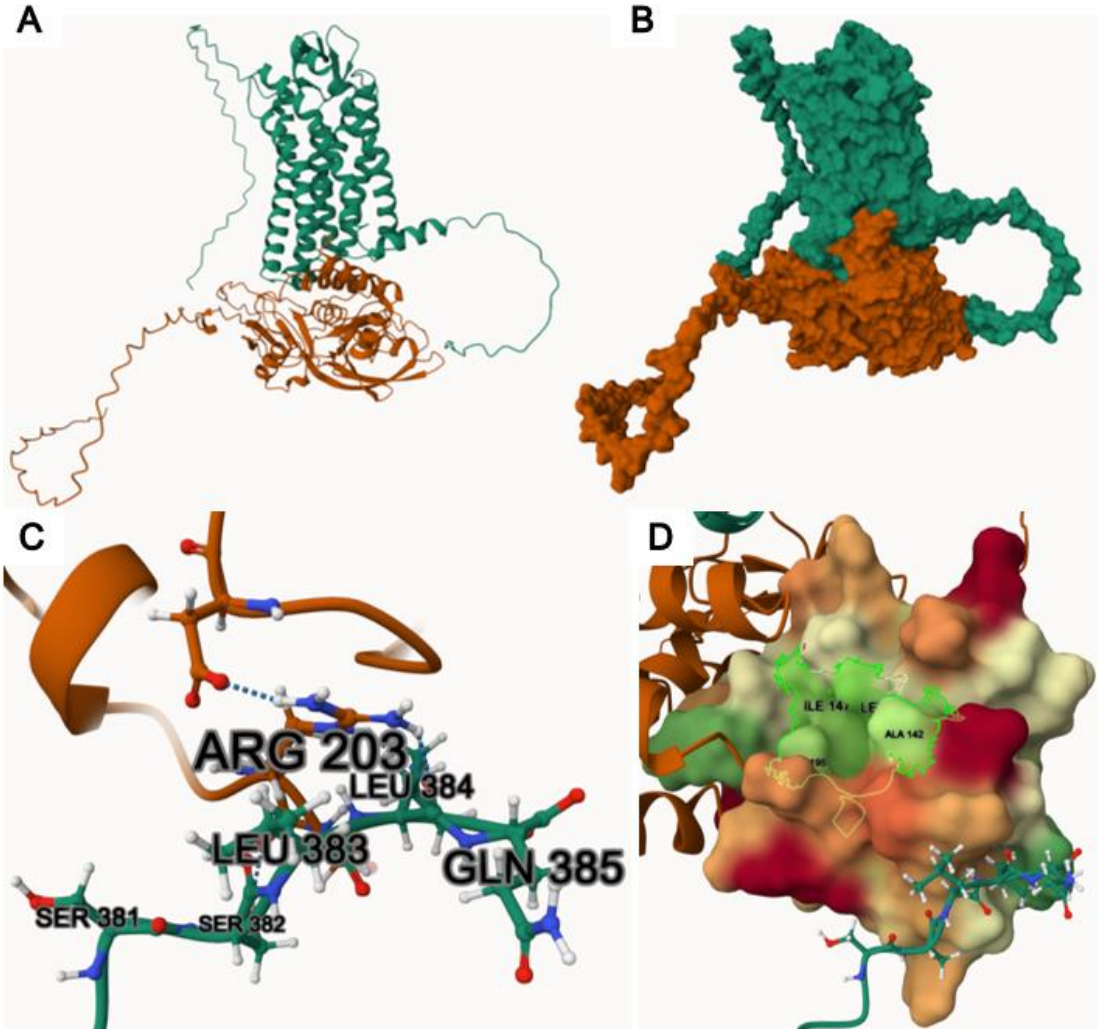


Figure S3 – Fourth AlphaFold model of the prediction of PAR4 interacting with GIPC1. A: Ribbon diagram of the receptor bound to GIPC1. PAR4 is shown in green, GIPC1 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a the PDZ domain, however the binding does not occur in the hydrophobic pocket, which is highlighted in green. Image generated using Mol*.

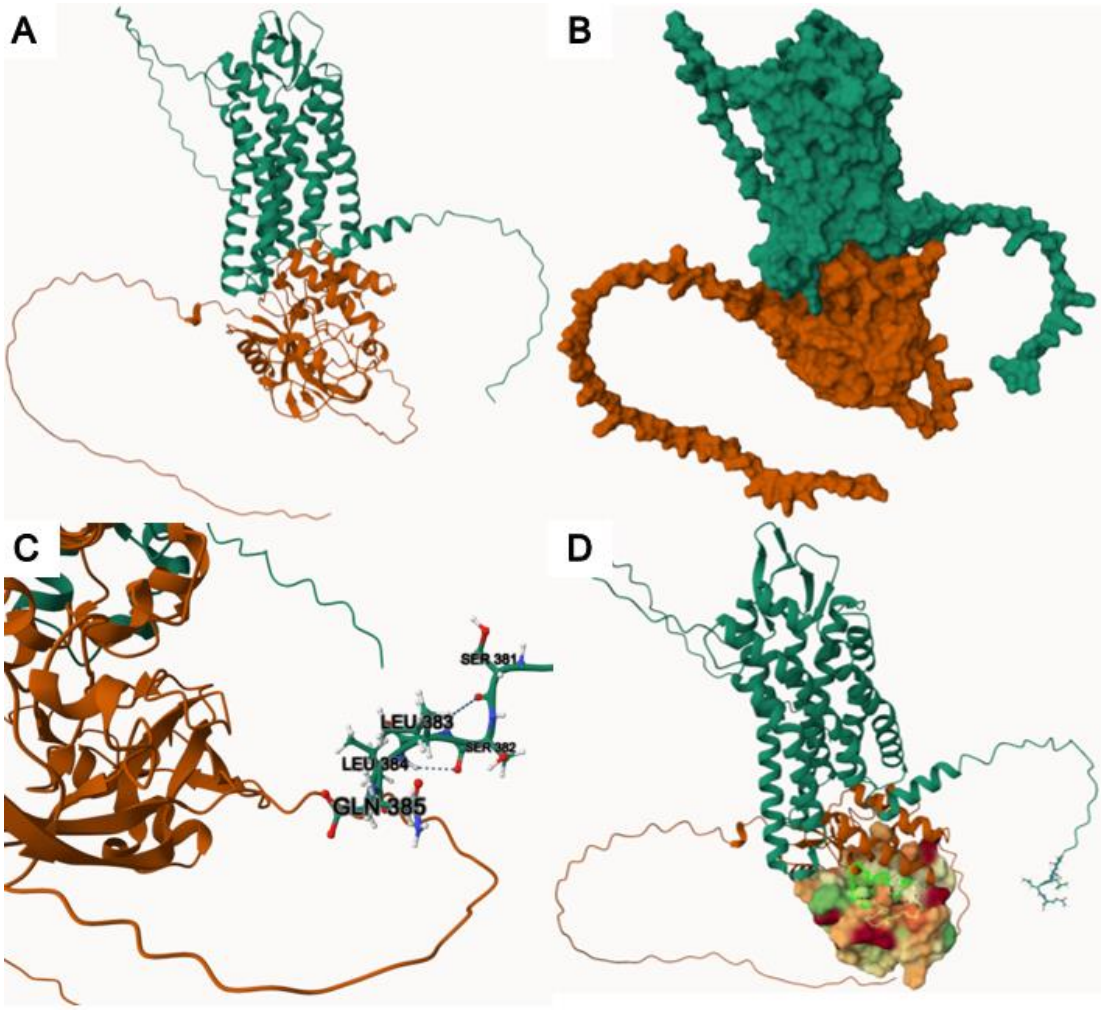


Figure S4 – Fifth AlphaFold model of the prediction of PAR4 interacting with GIPC1. A: Ribbon diagram of the receptor bound to GIPC1. PAR4 is shown in green, GIPC1 in orange. **B**: Molecular surface of the same interaction. **C**: In this model no interactions were detected between PAR4 SLiM and GIPC1 PDZ domain. **D**: SLiM did not bind the hydrophobic pocket, which is highlighted in green. Image generated using Mol*.

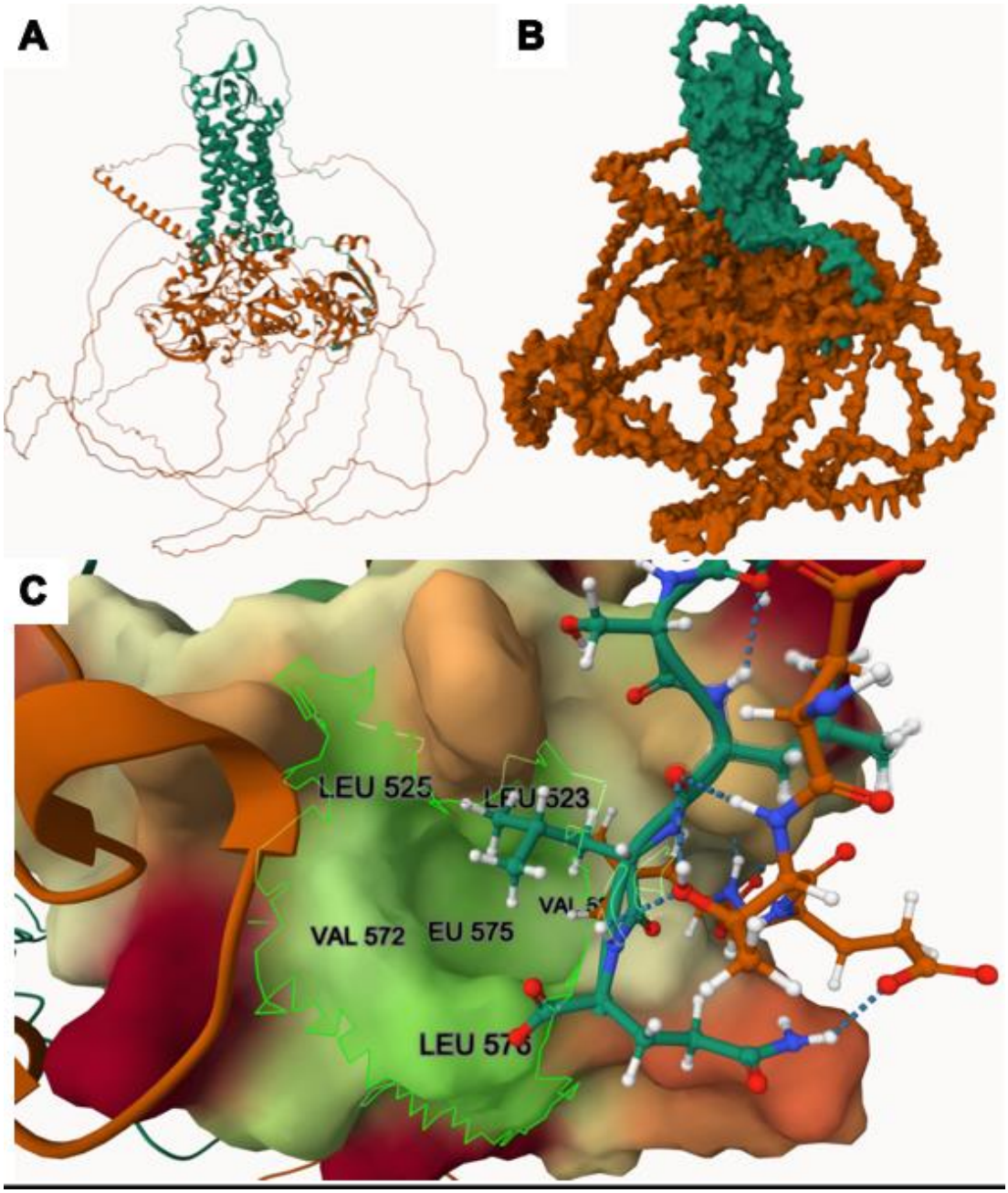


Figure S5 – Interactions with a disordered region prevent binding on the PDZ domain 3 in model 3. A: Ribbon diagram of the receptor bound to TJP2. PAR4 is shown in green, TJP2 in orange. **B**: Molecular surface of the same interaction. **C**: SLiM does not interact with any hydrophobic residues on PDZ domain 3, since it formed hydrogen bonds with the end of a disordered region. Image generated using Mol*.

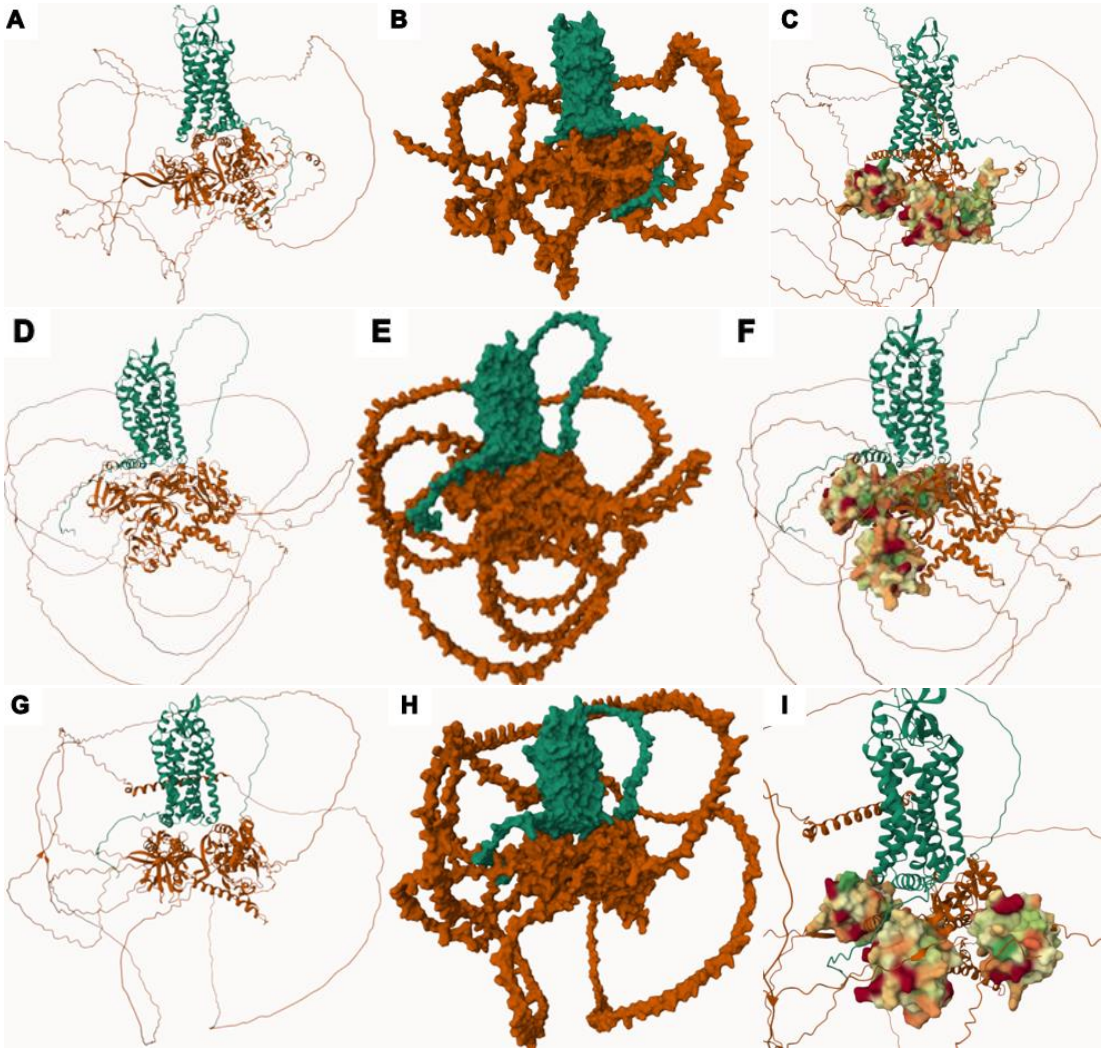


Figure S6 – Models two, four and five did not predict interactions between PAR4 SliM and any PDZ domain on TJP2. A: Ribbon diagram of the receptor bound to GIPC1. PAR4 is shown in green, GIPC1 in orange. **B**: Molecular surface of the same interaction. **C**: Ribbon diagram of the interactions between PAR4 SLiM and residues on PDZ domain. Hydrogen bonds are highlighted with a dotted line. **D**: SLiM binds a the PDZ domain, however the binding does not occur in the hydrophobic pocket, which is highlighted in green. Image generated using Mol*.

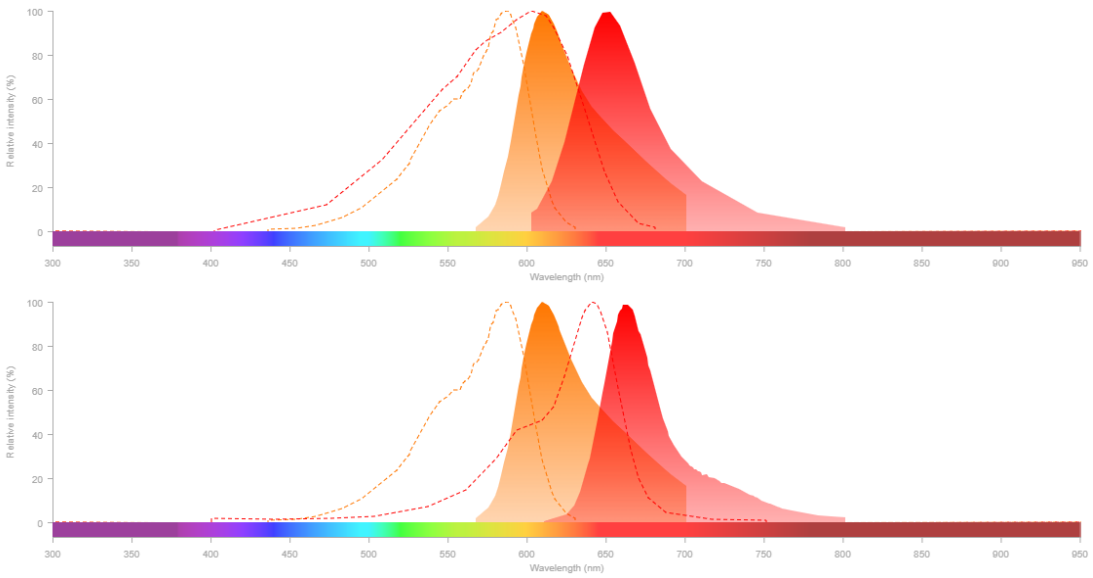


Figure S7 – Excitation and emission profiles for mCherry fluorescent protein and Mitotracker Red FM (top) and Mitotracker Deep Red FM (bottom). Image generated using Fluorescence SpectraViewer by ThermoFisher Scientific.

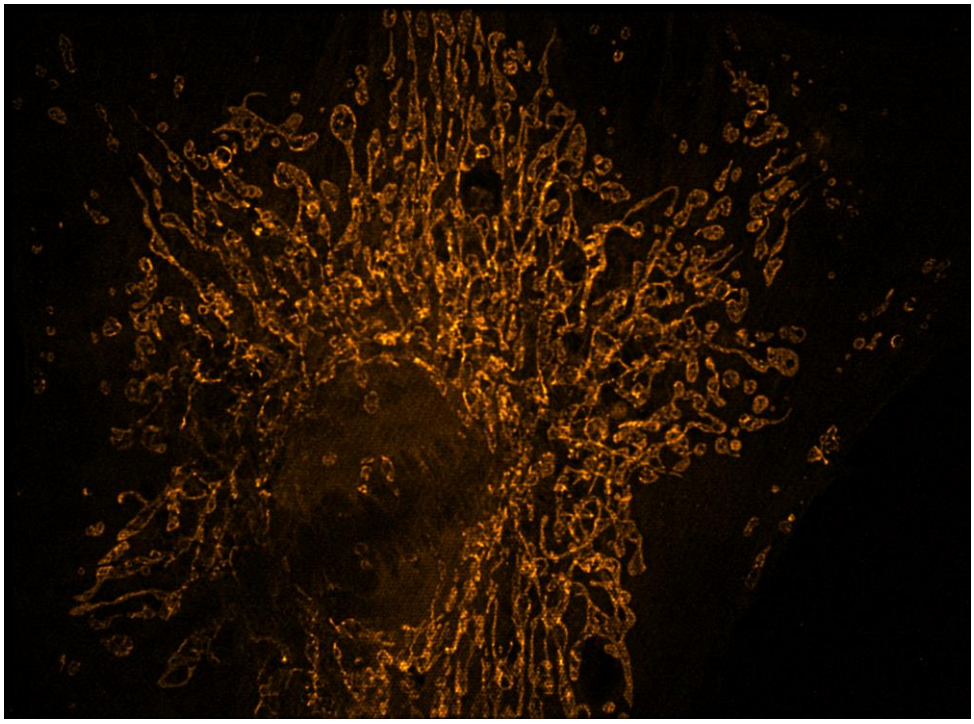


Figure S8 – SIM imaging of test slide with Bovine Pulmonary Artery Endothelial (BPAE) cells stained with MitoTracker Red. Cells were acquired already stained and used to demonstrate the increase in resolution with SIM microscopy.

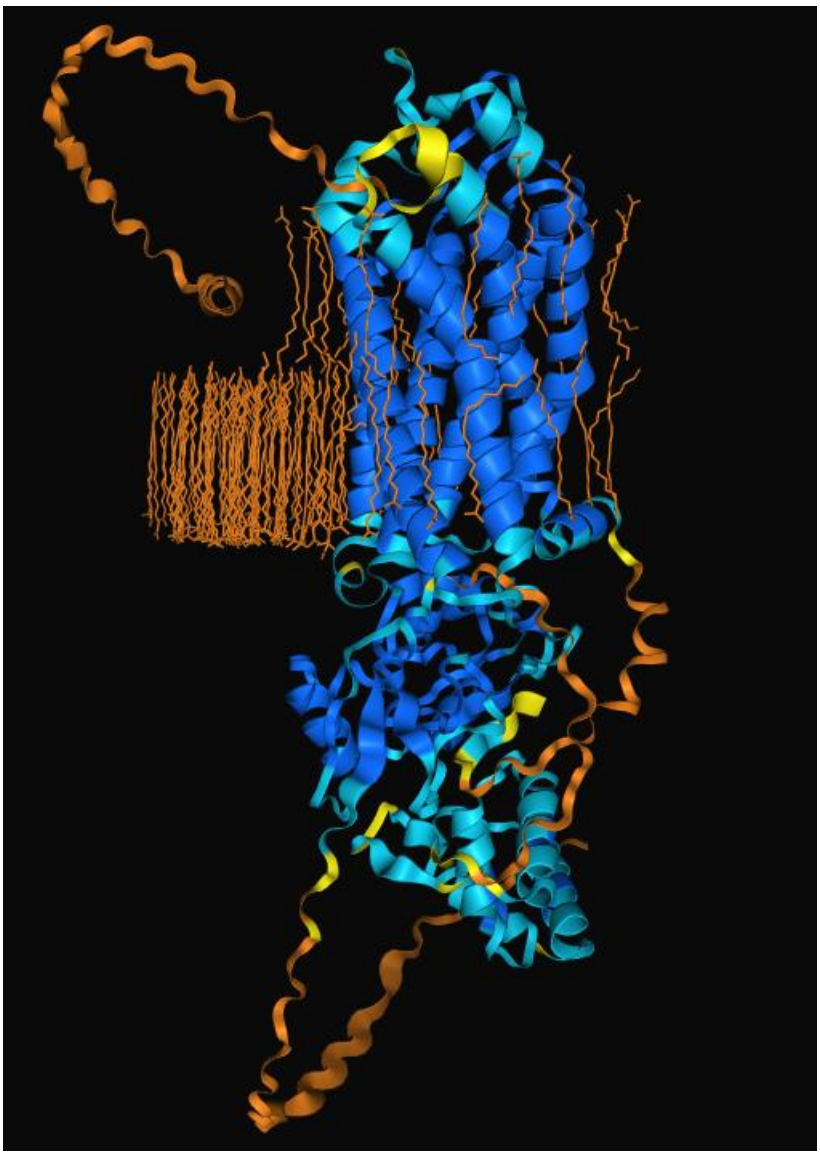


Figure S9 - AlphaFold 3 prediction of the full structure of PAR4 with GIPC2. AlphaFold3 allows to predict structures with a set of ligands. By providing >50 oleic acids the AI folds them in a membrane-like structure and it is able to differentiate intracellular and extracellular regions of proteins. Disordered regions were folded as helixes, but with low pLDDT scores.