Elucidating the activation mechanisms of adhesion GPCRs

Rebecca Fern Holdich

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

School of Molecular and Cellular Biology

September 2018

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First of all, I would like to thank the Stacey group for the past 4 years. Tom and Joe, you have made every day a pleasure. Thank you for letting me rant on about failed experiments, for providing advice when I needed it, reminding me things were never as bad as they seemed, and for many great evenings of laughs and Be At One cocktails. Thank you to Chi for persisting with teaching me crystallography. I'm sorry that it will always go over my head! Thank you for organising our lab social life, for always trying to make me see the upside, and for providing me a peaceful environment to write the second half of this thesis. I would've gone crazy without it.

Martin, thank you for highlighting the fact that I am taking away from this experience a face full of wrinkles and a much more pessimistic outlook on life. Despite the constant sarcasm, if it wasn't for you teaching me how to "roll it in glitter", this thesis would've been even more depressing to read. Most of all, thank you for being a wonderful first boss, and for your encouragement, help, patience, and relentless insistence that "something has to work". I may just have given up a long time ago without it.

I could not have competed this in one piece without you, Joe. You've been my person every day in the lab, ready with the supply of biscuits, Marmite toast, supportive encouragement and Friday night drinks. Jess, Lauren, Cait, Kathy, Viv, Helen, Polly, Sophie and Danielle, I'm lucky enough to call you my best friends and you have been my rocks. If I wrote all I wanted to about each of you, it would take up half the thesis! I cannot thank you enough for providing me with so many incredible memories and being at my side every step of the way.

Lauren, our evenings on the sofa, Saturday morning brunches, and constant supply of prosecco and Cointreau for the past 5 years has been the most wonderful thing. And for letting me be a part-time lodger at your house whilst writing this book. I wouldn't have done this alongside anyone else. I miss you already.

To the Bonhams and the Ramsdens, the kind of aunties/uncles/cousins that are one in a million. Together we've tackled some unbearable losses over the last two years but have made it through intact and still smiling. The support you have provided over the years is the reason I am here writing this today. I could go on and on, but please know how much you mean to me.

Will, thank you for coming into my life with a Martinez in hand, and the best culinary skills around outside of a Michelin restaurant. You know nothing makes me happier

than your ability to calm my hanger! Your seemingly endless supply of love and patience has helped me through so many tough times. Thank you for sticking with me long enough to create countless amazing memories.

Mum and Niki; there for me every single step of the way. Your faith, love, support and encouragement mean more than you could know. Thank you for being my constants, for teaching me to be strong and resilient in the face of tough times, and always believing I could do it, especially when I didn't.

And finally, I would like to dedicate the past four years to my extraordinarily wonderful Nana and Grandad. I may not have had the sought after "Eureka!" moment, but I could not have done this without your wise words, love and encouragement. Nana, I wish you were here with me to read this now.

Abstract

The sensing of mechanical forces by cells is an essential process in physiological systems. If perturbed, it can result in various pathologies including cancer, vascular disease, and deafness. Despite its obvious importance, only a limited number of cell receptors are currently implicated in mediating responses to mechanical stimuli. One group of receptors routinely exposed to physical forces within their microenvironments, including the circulation, inner ear, and epididymis, and with great potential in mechanosensing, are the adhesion GPCR family. Adhesion GPCRs comprise 33 human transmembrane signalling receptors with particularly large extracellular domains, composed of several protein repeats known to facilitate adhesive interactions. However, despite their involvement in key physiological and pathological processes the exact mechanisms of signal transduction for these receptors remains unspecified.

The work undertaken for this research therefore aimed to provide mechanistic insights into the signalling of adhesion GPCRs using the following experimental approaches: 1) elucidation of the atomic structure of a crucial signalling domain, 2) characterisation of signal induction mechanisms and intracellular signalling pathways and 3) interrogation of the mechanical sensing properties of the extracellular domain.

In summary, the results of this work have generated a 2.2Å resolution native X-ray diffraction dataset of the extracellular GAIN domain of the adhesion GPCR, CD97. In addition, using luciferase reporter assays this research has also generated robust readouts for the dissection of adhesion GPCR signalling. These readouts have for the first time established $G\alpha_{12/13}$ as a major signalling pathway for GPR97, shown vibration-induced signalling of GPR56, and have demonstrated that small protein scaffolds known as Affimers can be used as specific modulators of adhesion GPCR receptor signalling. Finally, atomic force microscopy was utilised to dissect the force required to induce the removal of the extracellular domains of adhesion-GPCR, a key potential mechanism involved in mechanosensing. Together, these results provide structural information essential for the future atomic resolution of CD97 GAIN domain, demonstrate that members of the adhesion GPCR family are indeed mechanosensors, and that modulation of receptor signalling via Affimers can be a useful potential strategy for future mechanistic and pharmacological studies.

Table of Contents

Acknowledgements ii

Abstract iii

Table of Contents iv

Table of Figures x

List of Tables xii

Abbreviations xiii

Chapter 1. Introduction 1

1.1 The G Protein-Coupled Receptor Superfamily 2

1.1.1 GPCR structure 3

1.1.2 GPCR activation 4

1.1.3 Heterotrimeric G proteins 5

1.1.4 Heterotrimeric G protein signalling pathways 6

1.2 Adhesion GPCRs 8

1.2.1 General structure 8

1.2.2 The GPS motif and the GAIN domain 11

1.2.3 Mechanisms of activation and signalling 13

1.2.4 Heterotrimeric G protein-dependent signalling 15

1.2.5 EGF-TM7/Group II Subfamily 16

1.2.6 Group VIII Subfamily 17

1.2.7 Adhesion GPCR binding partners 18

1.2.8 Physiological and pathological function of adhesion GPCRs 19

1.2.8.1 Adhesion GPCRs in migration, adhesion and cytoskeletal regulation 20

1.2.8.2 Adhesion GPCRs in cancer 20

1.2.8.3 Adhesion GPCRs in the nervous system 21

1.2.8.4 Adhesion GPCRs in the immune system 22

1.2.9 Adhesion GPCRs as mechanosensors 23

1.3 Mechanical force in physiological systems 25

1.3.1 AFM: investigating mechanical protein unfolding 28

1.3.2 Protein structure and force resistance 31
1.4 Protein characterisation with non-Ig binding proteins 32
1.4.1 Affimer molecules 34
1.5 Thesis aims 35
Chapter 2. Materials and Methods 37
2.1 Materials 38
2.1.1 Buffers 38
2.2 Molecular Biology 39
2.2.1 Agarose electrophoresis 39
2.2.2 PCR 39
2.2.3 Overlap extension PCR 39
2.2.4 Gibson assembly PCR 39
2.2.5 Site directed mutagenesis 39
2.2.6 Restriction digest 40
2.2.7 Ligation 40
2.2.8 Transformation 40
2.2.9 Competent Cell Production 40
2.2.10 SDS-Polyacrylamide gel electrophoresis (PAGE) 41
2.2.11 Western and dot blot 41
2.2.12 Protein Expression in *E. coli* 41
2.2.13 Protein Purification 42
2.2.14 Ion Exchange Chromatography 43
2.2.15 Affimer Protein Preparation 43
2.2.16 Biotinylation 43
2.2.17 Dialysis 43
2.2.18 Mass Spectrometry 43
2.3 General Methods 44
2.3.1 Mammalian cell culture 44

2.3.2 Cell Transfection 44
2.3.3 Insect cell culture 45
2.3.4 Insect Cell Protein Expression 45
2.3.5 Determining viral titre by quantitative real-time PCR (qRT-PCR) 45
2.3.6 Dual Luciferase reporter assays 46
2.3.7 Enzyme-Linked Immunosorbent Assay (ELISA) 46
2.3.8 Flow cytometry 46
2.3.9 Fixing and staining cells for immunofluorescence 47
2.4 Protein Crystallographic Trials 48
2.4.1 Sparse matrix screening 48
2.4.2 Crystal optimisation 48
2.4.3 Crystal streak- and microbead seeding 48
2.4.4 Crystal harvesting and data collection 49
2.4.5 Data processing 49
2.5 Atomic Force Microscopy 50
2.5.1 Functionalisation of AFM surface and probe 50
2.5.1.1 Oxidisation of surface and silicon nitride probe 50
2.5.1.2 Aminosilanisation of silicon nitride 50
2.5.1.3 Attachment of NHS-PEG24-maleimide linkers 51
2.5.1.5 Attachment of HaloTag® ligand to surface 51
2.5.1.6 Attachment of Tris-NTA to probe 51
2.5.1.7 Protein attachment 51
2.5.2 Polyprotein Force Microscopy 52
2.5.2.1 AFM Calibration and set up 52
2.5.2.2 Data collection 53
2.5.2.3 Data processing 53
Chapter 3. Structural determination of EGF-TM7 GAIN domains 54
3.1 Introduction 55
3.1.1 Structural determination of proteins by crystallography 56
3.2 CD97 GAIN domain expression trials 59

3.2.1 Baculovirus expression system....59
3.2.1.1 Recombinant bacmid production....59
3.2.1.2 Expression of recombinant CD97 GAIN EGF-1,2,5 bacmid....60
3.2.1.3 Optimisation of CD97 GAIN EGF-1,2,5 expression....64
3.2.2 Bacterial Expression System....67
3.2.2.1 Expression trial of CD97 GAIN domain in *E. coli*....68
3.2.2.2 Vector optimisation for CD97 GAIN expression....74
3.2.2.3 Optimisation of *E. coli* expression strain....75
3.2.3 Protein purification....77
3.2.3.1 Initial CD97-GAIN purification....77
3.2.3.2 Purification optimisation....79
3.2.3.2 TEV protease cleavage for His-Tag removal....80
3.3 Crystallographic study of CD97 GAIN domain....82
3.3.1 Initial screens for crystallisation conditions....82
3.3.2 Crystal growth optimisation....84
3.3.3 Crystal harvesting and data processing....89
3.3.4 Structural determination data with the use of Affimers....91
3.3.4.1 Generation of Affimer scaffold proteins against CD97 GAIN....91
3.3.4.2 Expression and purification of Affimers....92
3.3.4.3 Confirmation of Affimer specificity for CD97-GAIN....94
3.4 Discussion and future perspectives....95
3.4.1 Protein expression....95
3.4.2 Crystallographic investigation of CD97-GAIN....100
Chapter 4. Interrogating adhesion GPCR signalling pathways and their activation with surrogate ligands....105
4.1 Introduction....106
4.2 Validation of luciferase reporter assay output....107
4.3 Signalling activity of GPR97 and GPR56....113
4.3.1 Cloning, expression and purification of receptor-specific Affimers....114
4.3.2 Verification of Affimer binding....117
4.3.3 Investigation of GPR56 and GPR97-activating Affimers....119

4.3.4 Adhesion GPCR signalling under mechanical stress....121
4.3.5 Functional assessment of naturally occurring mutants....123
4.4 Discussion and future work....130
Chapter 5. Interrogating the mechanical stability of adhesion GPCR GAIN domains....139
5.1 Introduction....140
5.2 Generation of GAIN-I_{27} polypeptide constructs for force spectroscopy....140
5.2.1 Gibson assembly and overlap extension cloning....141
5.2.2 Site directed mutagenesis for cleavage deficient GAIN domains....143
5.2.3 Expression and purification of GAIN-I_{27}-HaloTag constructs....145
5.2.3.1 Protein expression in mammalian cells....145
5.2.3.2 GAIN-I_{27}-HaloTag protein purification....147
5.3 Single molecule force spectroscopy of adhesion GPCR GAIN domains....152
5.3.1 Establishing SMFS experimental set-up....152
5.3.2 Optimisation of SMFS experiments....159
5.4 Discussion and future perspectives....164
Chapter 6. General discussion....170
6.1 Adhesion GPCRs as mechanosensors....171
6.2 Structural basis of adhesion GPCR activity....173
6.3 Characterising adhesion GPCR signalling....175
6.4 Conclusion....177
References....178
Appendix....195

Table of Figures

Figure 1.1 Phylogenetic representation of the G protein-coupled receptor superfamily.. 3
Figure 1.2 GPCR secondary structure. 4
Figure 1.3 Schematic illustration of GPCR activation.. 5
Figure 1.4 Overview of G-protein signalling. 7
Figure 1.5 Schematic depiction of the N-termini of adhesion GPCRs and their subgroups as predicted by RPS-BLAST. 10
Figure 1.6 Adhesion GPCR domain structure.. 12
Figure 1.7 Proposed mechanisms of adhesion GPCR activation. 15
Figure 1.8 EGF-TM7 subfamily characteristics overview. 17
Figure 1.9 Mechanical stimulation experienced by plasma membrane proteins.. 26
Figure 1.10 Single molecule force spectroscopy (SMFS) for studying the mechanical unfolding of single domains in a polyprotein. 29
Figure 1.11 Structural characteristics of a selection non-IG binding proteins. 33
Figure 3.1. Principle and techniques for crystal growth.. 57
Figure 3.2. Overview of the baculovirus expression system. 60
Figure 3.3. Baculoviral expression of GFP recombinant bacmid. 61
Figure 3.3. Time course infection of Sf9s expressing CD97-GAIN EGF-1,2,5. 63
Figure 3.4. Ni^{2+} affinity purification of CD97 GAIN EGF-1,2,5 from Sf9 whole cell lysate. 64
Figure 3.5. Baculorviral expression of secreted GPR114 GAIN and CD97 GAIN EGF-1,2,5 expression constructs. 66
Figure 3.6. Solubilisation of CD97 GAIN EGF-1,2,5 from Sf9 cell pellets. 67
Figure 3.7. Basic schematic of IPTG induction of protein expression under lac operon control. 68
Figure 3.8. Basic schematic of leaky expression. 71
Figure 3.9. His-CD97 GAIN expression trial in BL21-CodonPlus (DE3) E. coli. 73
Figure 3.10. Expression of His-CD97 GAIN cloned into pKK223-3. 75
Figure 3.11. Ni^{2+} affinity purification of His-CD97 GAIN expressed in SHuffle® T7 E. coli.. 76
Figure 3.12. Ion exchange purification of CD97 GAIN domain, resolved on 12% SDS PAGE.. 78
Figure 3.13. Purification of His-CD97 GAIN by Ni^{2+} NTA affinity chromatography. 79
Figure 3.14. Native atomic mass spectrum of CD97 GAIN. 80
Figure 3.15. TEV protease cleavage of N' terminal His-tag from CD97-GAIN. 81
Figure 3.16. Crystal growth after His-tag cleavage. 83

Figure 3.17. Crystal growth as a result of sparse matrix screening. 84
Figure 3.18. Micro-seeding using the streak seeding technique 85
Figure 3.19. Further optimisation of CD97-GAIN crystallisation conditions with MIB buffer. 86
Figure 3.20. Streak seeding of His-tag cleaved CD97-GAIN.. 87
Figure 3.21. CD97-GAIN additive screening for crystallisation.. 88
Figure 3.22 Diffraction pattern CD97-GAIN. 89
Figure 3.23. Cloning, expression and confirmation of Affimers against CD97-GAIN. 93
Figure 3.24 Relative binding affinities for anti-CD97 GAIN domain Affimers. 94
Figure 4.1 Schematic of GPCR-stimulated luciferase reporter assays. 107
Figure 4.2 Schematic representation of luciferase assay set up. 109
Figure 4.3 Confirmation of CD97-dependent SRF-RE and SRE activation. 111
Figure 4.4 Affimer-induced CD97 SRF-RE signalling. 112
Figure 4.5 GPR56 and GPR97-dependent SRF-RE activation 114
Figure 4.6 Identification of receptor-specific Affimers. 115
Figure 4.7 Sequence alignment and purification of GPR56- and GPR97-specific Affimers. 116
Figure 4.8 Confirmation of α-GAIN Affimers binding to surface expressed GPR56 118
Figure 4.9 Confirmation of α-GAIN Affimers binding to GPR97-GFP expressing cells. 119
Figure 4.10 GPR97 and GPR56 activate SRF-RE signalling.. 120
Figure 4.11 Vibration induced signalling of CD97. 121
Figure 4.12 Analysis of vibration-induced signalling of GPR97 and GPR56. 122
Figure 4.13 Signalling activity of adhesion GPCR variants identified from the Born in Bradford cohort. 125
Figure 4.14 Visualisation of GPR56M487T expression at the cell membrane. 126
Figure 4.15 A single nucleotide polymorphism within GPR56 TM3 does not alter surface expression 127
Figure 4.16 Detection of GPR97G437S expression in HEK293Ts. 128
Figure 4.17 GPR97G437S expression is reduced in comparison to GPR97wt. 129
Figure 5.1 Schematic representation of Gibson assembly cloning. 142
Figure 5.2 Agarose gel electrophoresis analysis of site directed mutagenesis 144
Figure 5.3 Schematic of mammalian polyproteins generated for single molecule force spectroscopy experiments. 145
Figure 5.4 Transient transfection of AFM polyproteins. 147

Figure 5.5 Purification of I_{27}(4) and EMR2 GAINwt by Ni^{2+} affinity chromatography and batch method.....149
Figure 5.6 Optimised purification of AFM constructs confirmed by western blot...150
Figure 5.7 Confirmation of LAT1, GPR97 and EMR1 polyprotein expression and purification. 151
Figure 5.8 Representation of example GAIN domain I_{27}-HaloTag constructs for atomic force microscopy.....153
Figure 5.9 Example force-extension traces from SMFS. 155
Figure 5.10 Pulling experiment of I_{27}(4)-HaloTag at 700 nms^{-1}.. 156
Figure 5.11 SMFS of cleavage deficient EMR2 GAIN domain.. 158
Figure 5.12 Example of cleavage-deficient EMR2 GAIN domain unfolding in two-steps 159
Figure 5.13 I_{27}(4)-HaloTag SMFS after cantilever functionalisation.. 160
Figure 5.14 SMFS of EMR1 cleavage-deficient GAIN domain.. 161
Figure 5.15 SMFS of cleavage-deficient Lphn1 GAIN domain.. 163

List of Tables

Table 1.1 G protein classification of α-subunits 6
Table 1.2 Overview of mechanical forces experienced within physiological systems 28
Table 2.1 Cell lines used to transfect DNA. 44
Table 3.1. Characteristics of virally infected cells. Adapted from Bac-2-Bac manual, Invitrogen. 61
Table 3.2. Viral titre of recombinant bacmids calculated by plaque assay and qPCR. 62
Table 3.3 65
Table 3.4. Summary of CD97 GAIN expression trials. 69
Table 3.2 Data collection and processing for CD97 GAIN domain 90
Table 5.1 Predicted lengths of unfolded GAIN domains from structural information. 158

Abbreviations

AC - adenylyl cyclase

ADHD - attention deficit hyperactivity disorder

AFM - atomic force microscopy

AKT - protein kinase B

ANOVA - analysis of variance

BAI - brain-specific angiogenesis inhibitor

BFPP - bilateral frontoparietal polymicrogyria

cAMP - cyclic adenosine monophosphate

CNS - central nervous system

CRE - cAMP response element

CREB - cAMP response element-binding

CTF - C-terminal fragment

DAF - decay accelerating factor

DAG – diacylglycerol

DARPin - designed ankaryin repeat protein

DNA - deoxyribonucleic acid

EC – extracellular

ECD - extracellular domain

ECM - extracellular matrix

EGF - epidermal growth factor

ELISA - enzyme-linked immunosorbent assay

EMR - EGF module-containing mucin-like hormone receptor

EMT - epithelial-to-mesenchymal transition

EV - empty vector

FACS - fluorescence-activated cell sorting

FBS - foetal bovine serum

FDA - food and Drug Administration

FL - full length

FLRT - fibronectin leucine-rich transmembrane

FN - fibronectin

Fr - rupture force

GAIN - GPCR auto proteolysis-inducing

GDP - guanosine diphosphate

GEFs - G protein exchange factors

GFP - green fluorescent protein

GPR - G protein receptor

GPCR - G protein-coupled receptor

GPS - GPCR proteolysis site

GRAFS - glutamate, rhodopsin, adhesion, frizzled/taste, and secretin

GRKs - GPCR kinases

GTP - guanosine trisphosphate

HaloTag - haloalkane dehydrogenase tag

HBM - honeybee melittin

HEK - human embryonic kidney

HisTag - histidine tag

IC - intracellular

ICD - intracellular domain

IG – immunoglobulin

I_{27} - immunoglobulin module 27

IPTG - isopropyl β-D-1-thiogalactopyranoside

IP3 - inositol triphosphate

JCSG - joint centre for structural genomics

Lat1 - Latrophilin *C. elegans* gene

Lc - contour length

LPAR - lysophosphatidic acid receptor

LPHN - Latrophilin human protein

LPS - lipopolysaccharide

LRRs - leucine rich repeats

MAPK - mitogen-associated protein kinase

MBP - maltose-binding protein

MOI - multiplicity of infection

MR - molecular replacement

NFAT - nuclear factor of activated T-cells

NF-κB - nuclear factor-kappa B

NHS-biotin - *N*-hydroxysuccinimidobiotin

NK - natural killer

NOMPC - no mechanoreceptor potential C

NTA - nitrilotriacetate

NTF - N-terminal fragment

OD - optical density

PAR - protease-activated receptors

PAGE - polyacrylamide gel electrophoresis

PBS - phosphate buffered saline

PCR - polymerase chain reaction

PDB - protein databank

PEG - polyethylene glycol

PI3K - phosphatidylinositol-3-kinase

PIP2 - phosphatidylinositol-4,5-bisphosphate

PKA - protein kinase A

PKD - polycystic kidney disease

PMNs - polymorphonuclear cells

pN - picoNewton

PPI - protein-protein interaction

RE - response element

RGD - arginine glycine, and aspartate

RGS - Regulators of G-protein Signalling

RNA - ribonucleic acid

SDM - site directed mutagenesis

SDS - sodium dodecyl sulphate

SEC - size exclusion chromatography

SMFS - single molecule force spectroscopy

SNP - single nucleotide polymorphism

SOB - super optimal broth

SR - sarcoplasmic reticulum

SRE - serum response element

SRF - serum response factor

SUMO - small ubiquitin-like modifier

TCF - ternary complex factor

TEV - tobacco etch virus

TG2 - transglutaminase 2

TM - transmembrane

VLGR1 - very large GPCR receptor

WLC - worm-like chain

Chapter 1. Introduction

1.1 The G Protein-Coupled Receptor Superfamily

G protein-coupled receptors (GPCRs) constitute one of the largest protein superfamilies in the human genome with over 800 members. They play critical roles in the nearly all conceivable physiological processes, ranging from the sensing of photons and odorants to metabolic homeostasis, cardiac muscle contraction and migration of leukocytes. GPCRs are characterised by an extracellular N-terminus which facilitates sensing of a variety of external signals, a seven transmembrane (7TM) α-helical domain, and an intracellular C-terminus. Upon receptor activation, the C-terminus associates with and activates heterotrimeric G proteins, enabling the transduction of extracellular signals into physiological responses, via the stimulation of intracellular signalling pathways and production of second messengers (1).

GPCRs are classified into five distinct phylogenetic sub-groups according to the Glutamate, Rhodopsin, Adhesion, Frizzled/Taste, and Secretin (GRAFS) system (2) (Figure 1.1). Despite structural similarities, the subgroups exhibit little sequence homology, and their diversity is reflected by a large repertoire of ligands. These include monoamines, neurotransmitters, chemokines, lipids, nucleotides, and other naturally occurring molecules, for example calcium and chloride ions (3). This diversity is echoed by their widespread distribution in human tissue and involvement in fundamental processes such as vision and taste, reproductive, cardiovascular and endocrine functions. GPCRs are implicated in a plethora of human diseases, including hyperparathyroidism, in which the parathyroid glands produce excess parathyroid hormone; retinitis pigmentosa, a degenerative eye disease; and various forms of cancer. Currently, approximately 34% of Food and Drug Administration (FDA)-approved drugs target GPCRs, with over 90% targeting the rhodopsin class (4). This demonstrates the therapeutic significance of well characterised GPCRs, highlighting therapeutic potential and the need for a greater mechanistic understanding of less well characterised non-rhodopsin classes of GPCRs (5).

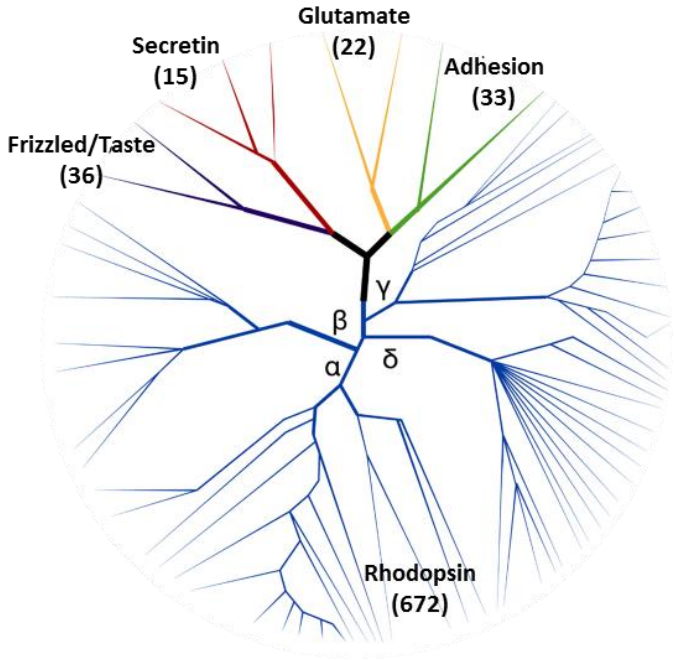


Figure 1.1 ***Phylogenetic representation of the G protein-coupled receptor superfamily.*** The superfamily is separated into 5 distinct families based on the GRAFS classification. The Rhodopsin class is the largest, being further divided into 4 subgroups, and is the most well characterised, containing classical and olfactory GPCRs. Adapted from (Cardoso J et. al, 2012)(6).

1.1.1 GPCR structure

The structural study of GPCRs remains a constant challenge due to their inherent dynamic behaviour, making them notoriously difficult to crystallise for macromolecular crystallography (7). Rhodopsin, being easier than most GPCRs to express and purify at a high yield, paved the way for structural determination of the rhodopsin subfamily. Initial structural insights were gained from circular dichroism. Such studies indicated the presence of a largely α-helical structure, consisting of 7 α-helices, with its N- and C-termini at opposing sides of the membrane. However, the first major breakthrough was gained in 1993, when the transmembrane bundle of Rhodopsin was first visualised by cryo-electron microscopy, providing confirmation of basic GPCR structure (Figure 1.2).

More recently, the three-dimensional atomic structures of an increasing number of Rhodopsin class GPCRs have been solved in both active and inactive states. This has provided mechanistic insights into the conformational changes within the transmembrane domain upon ligand interaction with the extracellular domain which results in receptor activation (8). This led to the identification of an “ionic lock”; one of a number of functionally relevant non-covalent interactions between the

transmembrane (TM) helices. The ionic lock forms between the cytosolic tails of TM3 and TM6 (9), which is modified upon ligand binding and alters the interface between TM3, TM5 and TM6, allowing for interaction with and activation of heterotrimeric G proteins (10). This structural knowledge has facilitated the identification of particular residues implicated in GPCR-associated disease states, and the development of high throughput compound screening, aiding identification of small molecule inhibitors for therapeutic exploitation (11).

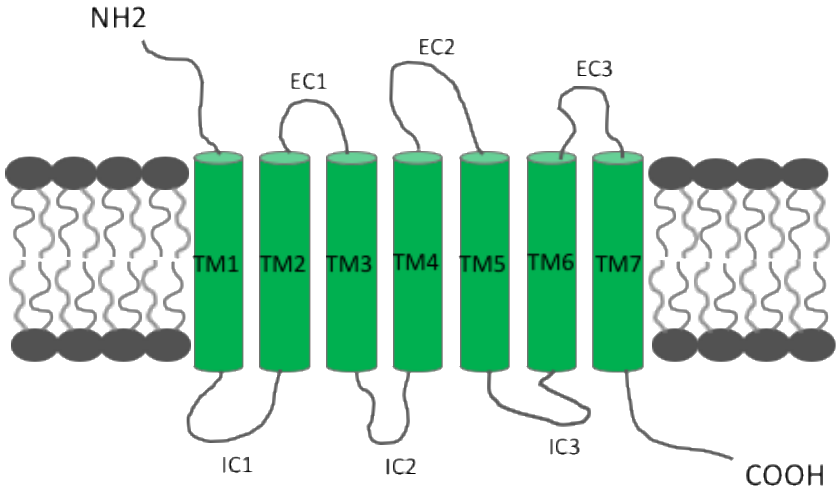


Figure 1.2 ***GPCR secondary structure.*** All GPCRs share the 7 transmembrane (TM) topology, with an extracellular (EC) N-terminus and intracellular (IC) C-terminus, and 3-interhelix loops on both sides of the plasma membrane.

1.1.2 GPCR activation

Despite their diversity, the majority of GPCRs are believed to signal though a similar mechanism; that is the transduction of external stimuli into intracellular secondary messengers via associated heterotrimeric G protein complexes (12). These complexes consist of Gα, and Gβγ subunits that prior to GPCR-ligand binding, exist in an inactive GDP bound state. However, upon ligand binding, conformational changes through the transmembrane helices facilitate GDP-GTP exchange within the α-subunit, resulting in the dissociation of Gβγ complex from GTP-bound Gα (Figure 1.3)(13). Both subunits can then regulate various effectors, influencing second messenger production and transcription factor activation, therefore mediating concomitant alterations in gene expression and cellular phenotype. GPCRs are

deactivated by either ligand dissociation, or receptor desensitisation and internalisation, and cycling of GTP to GDP, to mediate association of the Gα and Gβγ subunits.

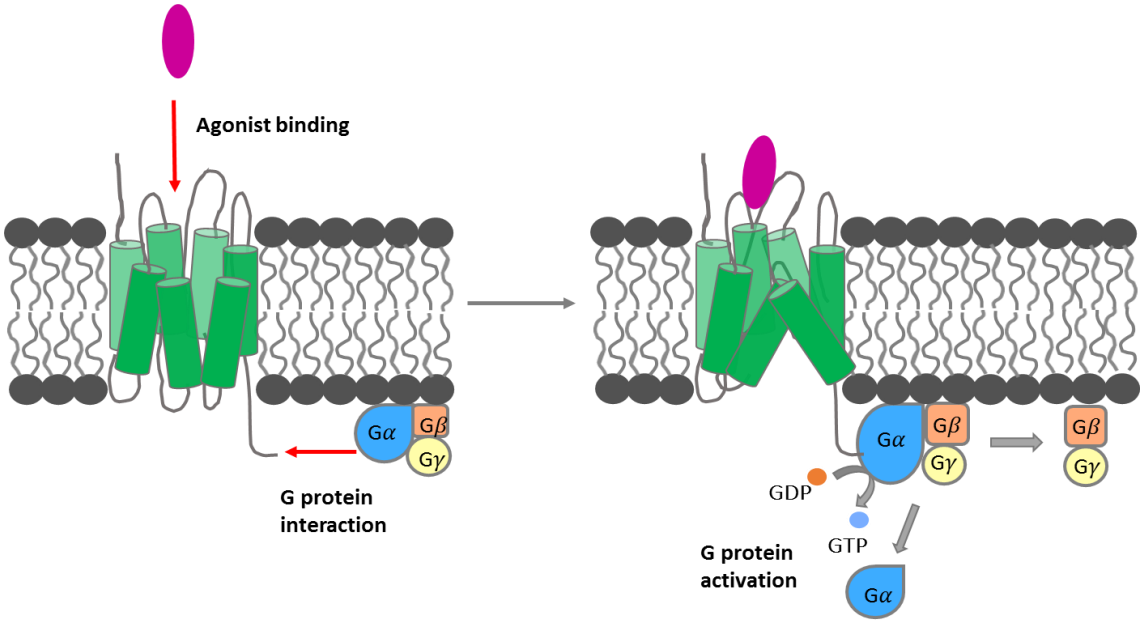


Figure 1.3 ***Schematic illustration of GPCR activation*****.** Ligand interaction with the extracellular-facing TM helices and interhelical loops stimulates movement of TM helices to activate the receptor via the action of a number of molecular switches (9). This promotes interaction of the cytosolic tail with the α-subunit of a heterotrimeric G protein, which stimulates the subunit's GTPase potential to cycle GDP to GTP, resulting in separation of the α- and βγ-subunits.

1.1.3 Heterotrimeric G proteins

Heterotrimeric G proteins are common eukaryotic signalling proteins which transduce signals from the plasma membrane in response to GPCR activation, and are responsible for the specificity of cellular action in response to extracellular stimuli (14). There are 4 main classes of heterotrimeric G proteins, which are defined based on sequence similarity of the α-subunit; $G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_q$, and $G\alpha_{12/13}$ (Table 1.1), whilst there are 21, 6 and 12 potential subclasses of α-, β- and γ-subunits respectively (15). The Gα subunit is comprised of a Ras domain and an α-helical domain which form a pocket for guanine nucleotide binding. Gβγ is able to associate with the α-subunit in its GDP-bound, inactive state (16). The proteins remain localised to the plasma membrane after undergoing post-translational myristylation and palmitoylation.

G protein class	Isotypes within the class
Gs	Gαs, Gαolf
Gi	Gαi1, Gαi2, Gαi3, Gαo, Gαt1, Gαt2, Gαgust, Gαz
Gq	Gαq, Gα11, Gα14, Gα16
G12	Gα12, Gα13

Table 1.1 ***G protein classification of α-subunits.*** Taken from (Simon, M *et.al* 1991)(17).

Signalling specificity is mediated by both the α and βγ subunits which engage different receptors and effectors, and through regulation of effector activity. G protein activity is controlled by numerous protein families including guanine nucleotide exchange factors (GEFs), and Regulators of G-protein Signalling (RGS), which modulate α-subunit GDP-GTP exchange and GTPase activity to activate and deactivate G protein signalling respectively. Upon ligand binding, GPCRs function as GEFs, catalysing the dissociation of GDP from the G protein α-subunit which acts as a GTPase, allowing GTP to bind in its place. The rate of GTP hydrolysis can be accelerated by RGS, which ultimately leads to GTPase, hence α-subunit, deactivation and consequently termination of signalling (18). Further regulation of signalling is mediated via receptor desensitisation by GPCR kinases (GRKs) which phosphorylate the intracellular domains of GPCRs, which acts as a binding site for arrestin proteins. Arrestins prevent re-association of GPCRs with their G proteins and hence prevent reactivation of signalling pathways. Both GRKs and arrestins possess regulated and restricted expression profiles (19).

1.1.4 Heterotrimeric G protein signalling pathways

Initially, the Gβγ subunit was regarded as a purely negative regulator of Gα. However, Gα and Gβγ are both able to interact with and activate various cellular effectors upon agonist activation of GPCRs. The 4 classes of Gα have well established targets to activate specific signalling pathways (20)(21). $G\alpha_{12/13}$ mediate cytoskeletal remodelling, cell migration, proliferation and oncogenic transformation, by activating RhoGEFs, and subsequently the GTPase activity of RhoA, Ras and Cdc42, for example (Figure 1.4)(22). The GTPases themselves then stimulate effector proteins responsible for remodelling, which is crucial for cell migration and adhesion. $G\alpha_s$ trigger adenylyl cyclase (AC) activity to generate the second messenger cAMP from ATP. cAMP binds to the cAMP-dependent regulatory subunits of protein kinase A (PKA) to mediate their dissociation, freeing PKA to phosphorylate target serine and threonine residues (23). cAMP also stimulates

calcium channels. Both calcium and cAMP are prolific second messengers involved in a myriad of physiological functions. $G\alpha_i$ inhibits AC activity and closes calcium channels but activates mitogen-associated protein kinase (MAPK) and phosphatidylinositol-3-kinase (PI3K). This stimulates pathways that influence cell proliferation, survival and growth. Finally, activation of $G\alpha_q$ results in protein kinase C (PKC)-mediated cleavage of phosphatidylinositol-4,5-bisphosphate (PIP2) into diacylglycerol (DAG) and inositol triphosphate (IP3), stimulating cellular calcium mobilisation. The Gβγ subunit has been described to specifically mediate activation of effectors Ras and Rac, and initiate AKT-related signalling (20).

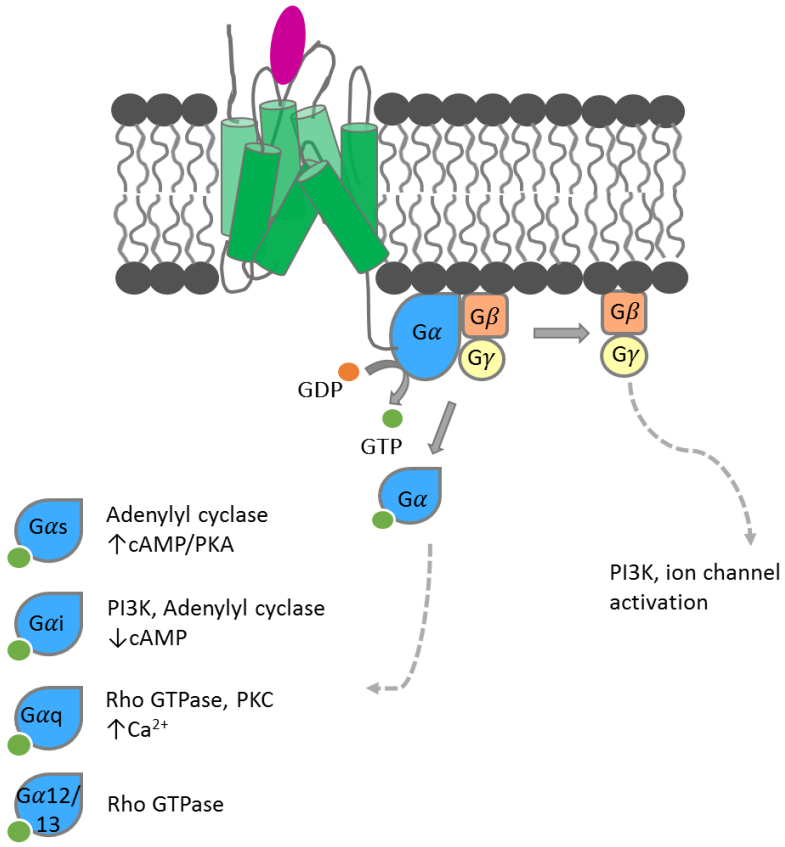


Figure 1.4 ***Overview of G-protein signalling.*** Active, GTP-bound Gα subunits dissociate from the Gβγ subunit to stimulate a range of downstream signalling cascades, depending on their specificity. Gαs activates cAMP-dependent pathways to mediate responses to a range of hormones, whilst Gαi inhibits the cAMP signalling by preventing the production of cAMP. Gαq activates phospholipase C kinase, resulting in the production of second messengers diacylglycerol and inositol trisphosphate. Gα12/13 subgroup is responsible for facilitating cytoskeletal remodelling which enables cell migration, adhesion, and proliferation.

As with most receptors, GPCR signalling is regulated to prevent overstimulation which may result in cell toxicity or uncontrolled growth. GPCR signalling is desensitised by a few mechanisms, enabling both short and long-term control. Short term desensitisation occurs via receptor phosphorylation by GPCR kinases (GRKs) which enables recruited β-arrestins to bind to the phosphorylation site, as well as a region at the base of the TM domains (19). Arrestins regulate a network of effectors, such as clathrin, to promote internalisation of GPCRs in clathrin coated pits for long term downregulation to fully terminate signalling, and recycling of GPCRs back to the cell surface. Arrestins recruit other proteins such as phosphodiesterases and diacylglycerol kinases to degrade second messengers and terminate further signalling (24)(25).

1.2 Adhesion GPCRs

Adhesion GPCRs constitute the second largest subgroup of GPCRs in humans, consisting of 33 proteins. Evolutionary studies have demonstrated they predate many key signalling proteins, and have been found in fungi and unicellular organisms, as well as metazoans (26)(27). Adhesion GPCRs have pivotal functions in processes such as male fertility, neuronal cell myelination, cancer metastasis, and cell remodelling and adhesion (28), yet remain the least understood of the subfamilies with the majority of members described as orphan receptors. The mode of activation and signalling of adhesion GPCRs has been a major focus within the field, but a definitive signalling mechanism for the family is yet to be fully elucidated (29)(30). Although a handful of receptors are becoming well defined in the biological roles they play, the lack of insight into their activation and functional regulation has hampered progression in fundamental research. Further understanding of adhesion GPCR physiological mechanisms of action is required to identify druggable targets and unlock their potential for therapeutic exploitation.

1.2.1 General structure

The adhesion subgroup is separated from the wider GPCR superfamily by their unusual structural features. These unique proteins exist as heterodimers at the cell surface because of protein processing at the endoplasmic reticulum, and often have particularly large extracellular regions. The aptly named very large GPCR receptor (VLGR1) for example, has 5800 amino acids in its N-terminus (31)(32). The

extracellular regions can be composed of an array of protein domains, including epidermal growth factor (EGF)-like repeats, immunoglobulin (Ig) domains, leucine rich repeats (LRRs), and cadherin repeats (29)(33). These domains are thought to facilitate cell-cell and cell-matrix interactions, adhesion and migration, and are likely important for cell function. Adhesion GPCRs are further subdivided into 9 groups based on sequence similarity of the transmembrane domain and by their extracellular protein domains (Figure 1.5). In some cases, receptors within subfamilies share similar tissue expression patterns, and therefore may have related functions.

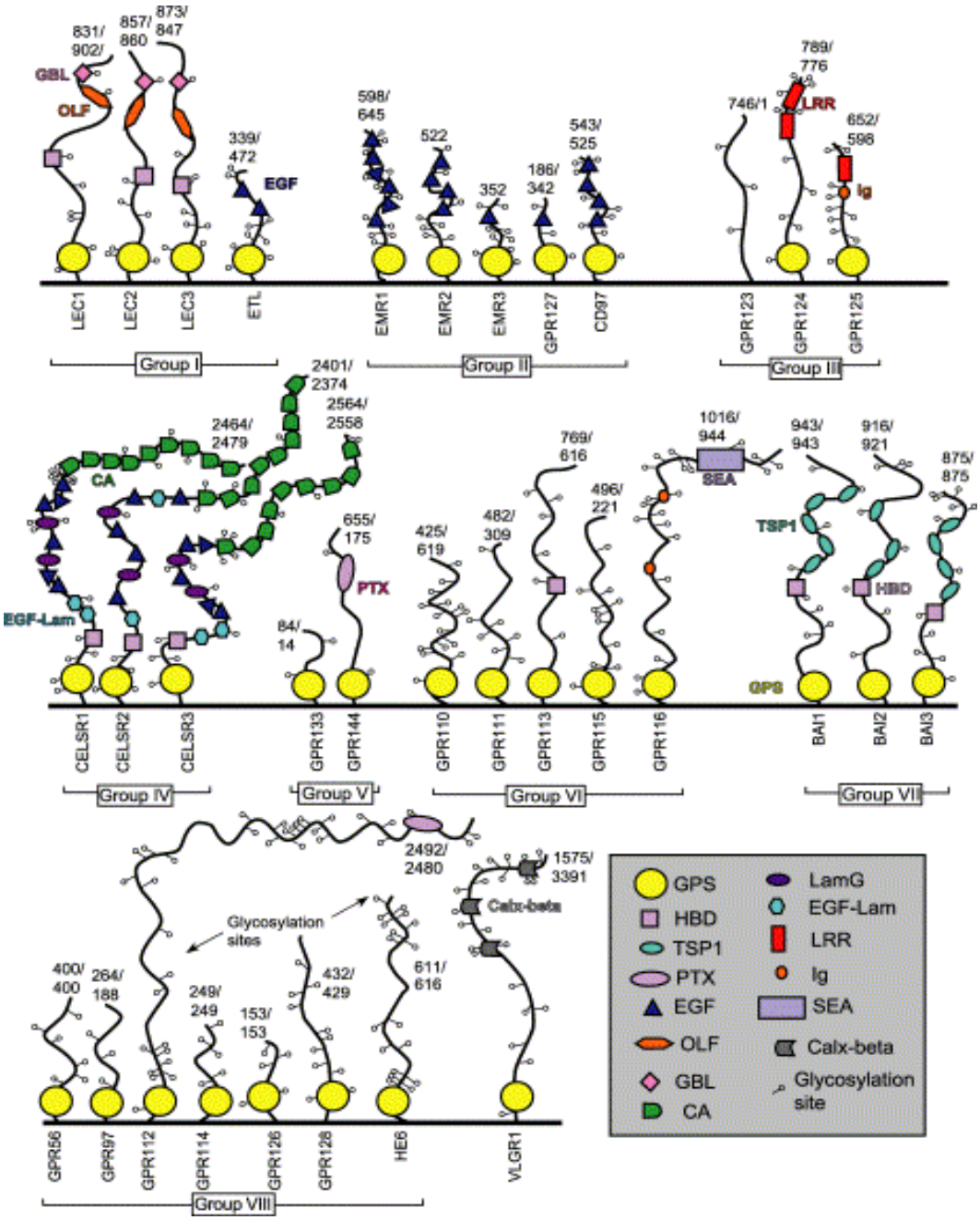


Figure 1.5 ***Schematic depiction of the N-termini of adhesion GPCRs and their subgroups as predicted by RPS-BLAST*.** Receptors are grouped based on TM-sequence similarity. The GPS motif, shown in yellow, is part of a wider, evolutionarily conserved domain, which lies immediately N-terminal to the transmembrane domain (not represented). The number at the top of each extracellular domain indicates the predicted number of residues from the start of the first TM helix. The presence of >1 number indicates splice variant differences. Each subgroup has a unique array of extracellular N-terminal adhesive protein domains, and varying sized intracellular C-terminals. Taken from (Bjarnadóttir et al. 2004)(34).

1.2.2 The GPS motif and the GAIN domain

Adhesion GPCRs have a unique hybrid structure that is likely pivotal to receptor function. Within their large extracellular domain lies the GPCR proteolysis site (GPS) motif; a highly conserved cysteine-rich stretch of approximately 40 amino acids (35). It forms a molecular hallmark of adhesion GPCRs, yet is also found in all 5 non-GPCR polycystic kidney disease (PKD) proteins (36). The consensus motif for cleavage is similar in all adhesion GPCRs, occurring between an aliphatic leucine and either a threonine, serine or cysteine; $His_{-2}Leu_{1}\downarrow Thr/Ser_{+1}$ (33)(37). These residues have been demonstrated to be an absolute requirement for self-catalysed proteolysis, but not necessarily protein expression. Experimental site directional mutagenesis of this consensus motif results in improper processing and a lack of trafficking to cell surface (38). Usually, this post-translational autoproteolytic cleavage event at the GPS generates a heterodimeric protein product that is expressed at the cell membrane, consisting of an N-terminal fragment (NTF) and a C-terminal fragment (CTF), also termed the α- and β-subunits, respectively (Figure 1.6A) (29)(39).

Although the region preceding the GPS motif was initially thought to be functionally redundant, Araç *et al*. (2012) characterised an approximate 320-residue domain, termed the GPCR auto proteolysis-inducing (GAIN) domain. This evolutionarily conserved domain, which encompasses the GPS, is the only extracellular domain common to all adhesion GPCRs and PKD proteins (26). The successful determination of the GAIN domain crystal structure for adhesion GPCRs Latrophilin-1 and brain-specific angiogenesis inhibitor 3 (BAI3) facilitated a deeper understanding of the unique characteristics of the GAIN domain. Although sequence alignment demonstrates low homology within the family, the secondary structure is well conserved and can be subdivided into different regions (Figure 1.6B). Subdomain A is composed of 6 α-helices, and subdomain B composed of 2 α-helices and 13 β-strands. The point of cleavage within the GPS lies at a tight turn caused by two disulphide bonds between strands β12 and β13 (Figure 1.6C)(40). The NTF therefore encompasses all extracellular adhesive protein domains, and the GAIN domain up to β12, whilst the CTF is composed of the final β-strand of the GAIN domain, the proximal transmembrane domain and intracellular tail.

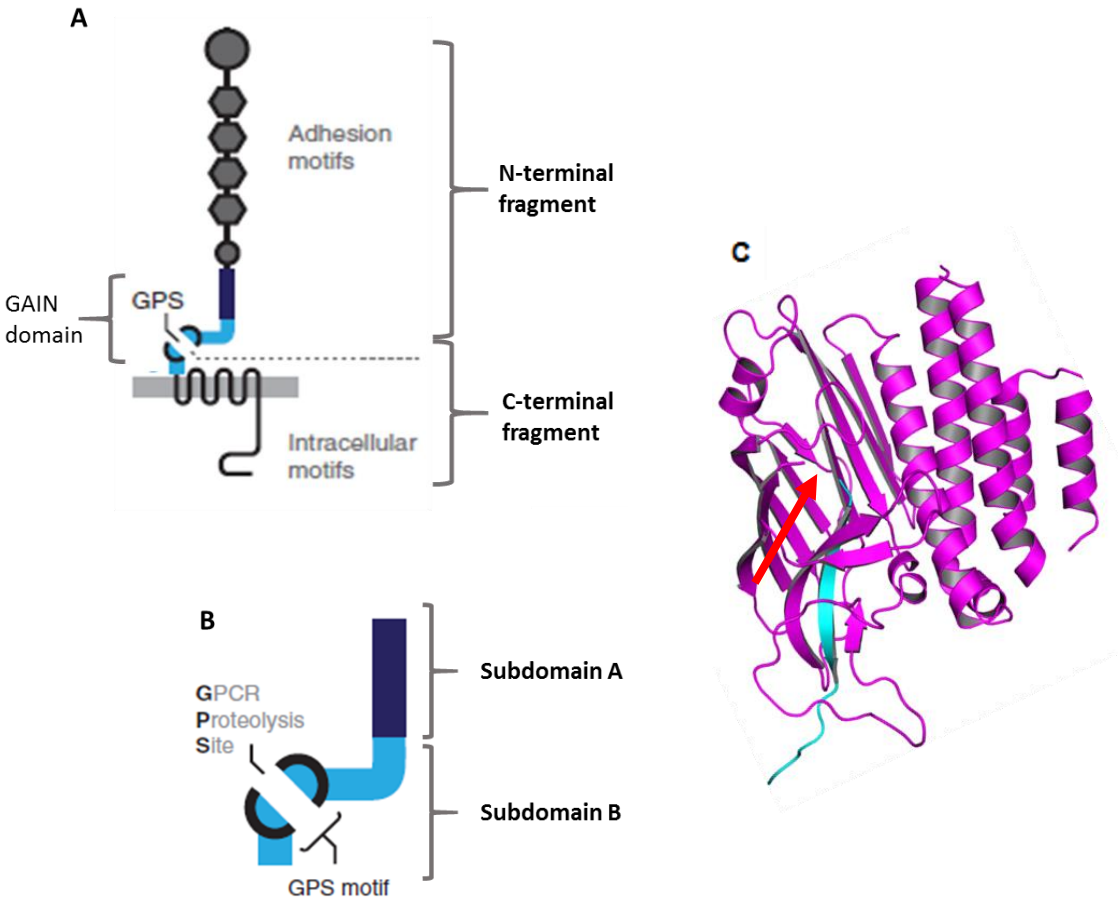


Figure 1.6 ***Adhesion GPCR domain structure.*** (**A**) The basic structure of adhesion GPCRs, illustrating the location of the GPS and wider GAIN domain proximal to the transmembrane region. (**B**) The GAIN domain is separated into two subdomains, based on its secondary structure. Subdomain B contains the GPS motif. The GAIN domain is necessary and sufficient for complete protein cleavage. (**C**) Cartoon representation of Lphn GAIN domain (PDB 4DLQ). The final β-strand immediately preceding the TM is highlighted in blue, and the cleavage site indicated by the red arrow. (**A+B**) Adapted from (T. Langenhan et al. 2013)(41).

Studies of Latrophilin 1 and BAI3 determined that the GPS motif itself is not enough to mediate cleavage. Receptors expressed in the absence of the majority of the GAIN domain, truncated to the GPS, did not undergo auto-proteolysis, indicating that the entire GAIN domain is both necessary and sufficient for cleavage (26). Post-translational cleavage however, does not necessarily mediate separation of the receptor fragments. The subunits remain associated via a network of hydrogen bonds and hydrophobic interfaces, hence complete dissociation is likely energetically unfavourable. Additionally, as the final β-strand of the GAIN domain is located within a central core of the wider GAIN domain, separation of the fragments is expected to

cause the entire structure to destabilise (26)(27). Despite this, investigation of the receptor EMR2 proteolysis indicated the presence of two distinct receptor complexes after trafficking to the membrane. Alongside the non-covalently associated heterodimer, EMR2 NTF and CTF were also able to exist separately in membrane raft domains (28). The detection of two receptor complexes demonstrates the diversity in structure and function of the enigmatic adhesion GPCR family.

Aside from mediating receptor cleavage, the GAIN domain can be a site for ligand binding for various adhesion GPCRs as shown by one member, CD97. It has also been identified as a "mutational hotspot" for both adhesion GPCRs and PKD proteins. Indeed, mutations affecting the domain's ability to mediate cleavage alters both expression and signalling activity of receptors. Point mutations introduced at the GPS of PKD1 have been shown to prevent receptor cleavage in murine models, resulting in either abnormal kidney development (42), or death in the first month of life (43). Adhesion GPCR GAIN domain mutations are implicated in a variety of diseases, including Usher's syndrome, bilateral frontoparietal polymicrogyria (BFPP), vibratory urticaria, and Severe Arthrogryposis Multiplex Congenita, for VLGR1, GPR56, EMR2 and GPR126 respectively (44)(45)(46)(47). Furthermore, GAIN domain mutations have been implicated in cancer. As post-translational processing remains functional despite these mutations (48), it is likely that either receptor signalling or ligand binding via the GAIN domain is altered to mediate tumorigenic effects.

1.2.3 Mechanisms of activation and signalling

With very few instances of ligand-induced signalling outlined, the molecular mechanisms by which adhesion GPCRs are activated have been one of the foremost focuses of the field for many years. Two possible modes of activation have now been proposed. In 2014, Liebscher *et al*. presented evidence towards the "tethered agonist" model (Figure 1.7), a mechanism that is analogous to the activation of the protease-activated receptors (PARs). PARs are GPCRs that undergo N-terminal cleavage by serine proteases thrombin and trypsin (49). The group identified a short peptide sequence proximal to the transmembrane domain which had activation properties, and was thereafter termed *Stachel*, German for "stinger". This agonist region was defined as the final β-strand of the GAIN domain (Figure 1.6C), which would be unveiled upon conformational changes within, or complete removal of, the NTF. GPR126 and GPR133 were initially exploited to test the tethered agonist hypothesis. Residues immediately downstream of the GPS were sequentially

deleted, which subsequently pinpointed a tethered agonist region of approximately 13 amino acids. G protein coupling upon exposure to soluble peptides mimicking the agonistic region was confirmed via analysis of cyclic-AMP accumulation (50). Indeed, the activity of receptors stimulated with soluble peptides derived from the post-cleavage sequence was largely increased; an effect that has since been replicated in a handful of other adhesion GPCRs (51)(52). Interestingly, the inclusion of just one extra amino acid in the peptide sequence can have inverse agonist effects on receptor activity (53). Furthermore, although a number of receptors can be activated by exposure of the peptide, BAI1, truncated to remove the *Stachel,* does not exhibit signalling deficits in comparison to a construct containing the *Stachel* (54). However, further investigation of this mode of activation for a few other receptors in the family is hampered by the hydrophobic, and therefore insoluble, nature of the peptides relating to their specific *Stachel* sequence.

The idea of NTF removal similarly accompanies the “disinhibition model” (Figure 1.7), whereby the NTF acts to inhibit the intrinsic signalling activity of the CTF. Adhesion GPCRs truncated to lack the entire NTF are shown to have highly upregulated signalling activity (55)(56)(57). This suggests that the large extracellular domain acts to hold the receptor in an inactive conformation. Upon stimulation by ligand binding for example, the NTF may undergo a significant conformational change to mediate removal of the wider GAIN domain from the final β-strand. Indeed, binding of receptor specific antibodies, but not complete NTF dissociation, causes receptor activation in the case of GPR56 and EMR2 (58)(59). However, the lack of ligand-induced signalling examples in comparison to the number of characterised ligands suggests another factor, such as mechanical force, may be required to mediate receptor activation. Furthermore, not all adhesion GPCRs undergo autocatalytic cleavage, such as GPR111 and GPR115, yet can still be activated (60). It may be possible that force, such as shear flow, mediates conformational changes internally within the GAIN domain to allow the *Stachel* to interact with the transmembrane domain, rather than uncovering it completely by domain separation (53). This notion is reinforced by an increasing number of adhesion GPCRs being implicated in mechanosensation (61)(62).

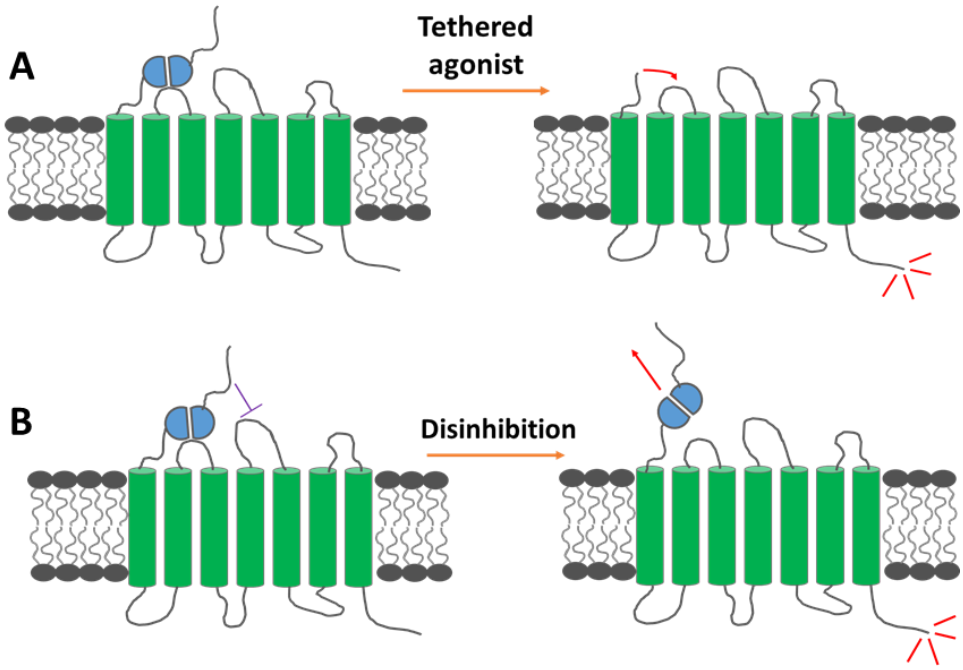


Figure 1.7 ***Proposed mechanisms of adhesion GPCR activation*****.** (**A**) Tethered agonist; the NTF masks the tethered agonist peptide and upon removal, by ligand binding or force, frees the Stachel to interact with the extracellular portion of the transmembrane domain to activate the receptor. (**B**) Disinhibition; the NTF interacts with the transmembrane domain to inhibit constitutive activity. Structural rearrangement, again by ligand binding or other force, causes the NTF to move away from the transmembrane domain, releasing inhibitory action.

1.2.4 Heterotrimeric G protein-dependent signalling

Due to a lack of endogenous stimulus, the downstream signalling pathways of many adhesion GPCRs remain undefined. However, G protein-mediated signal transduction has been demonstrated for a small number of receptors in the subgroup. Collagen IV interaction with GPR126 was one of the first examples of ligand-induced G protein signalling, regulating Schwann cell myelination via activation of $G\alpha_s$ to induce cAMP-dependent pathways (63). Other examples of ligand-induced G protein signalling include GPR56-Collagen III which stimulates $G\alpha_{12/13}$ (64), and α-latrotoxin binding of Latrophilin 1, resulting in activation of $G\alpha_o$ and $G\alpha_q$ (65). BAI1 couples with $G\alpha_{12/13}$, consequently stimulating Rho signalling, whilst truncation of the N-terminus causes a dramatic increase in receptor signalling (66). CD97 and ERM2 couple with $G\alpha_{12/13}$ and $G\alpha_{15}$ respectively, with CD97 mediating prostate cancer cell migration (67)(68). GPR64 can trigger downstream signalling via all 4 G protein subgroups

(69), stimulating SRE and NFκB pathways, and also cAMP accumulation when stimulated with a soluble mimic of its *Stachel* peptide (51).

Although not the focus of this thesis, adhesion GPCRs can also signal independently of G proteins via PDZ-binding domains. For example, GPR124 activates β-catenin by interaction of the soluble signalling molecule WNT7, stimulating angiogenesis within the CNS (70). Interestingly this mechanism absolutely requires GPR124 to have an intact NTF, whilst the final 4 amino acids of the CTF facilitate optimal β-catenin signalling, eluding to the complexity of signalling mechanisms for adhesion GPCRs (70).

1.2.5 EGF-TM7/Group II Subfamily

The group II family epidermal growth factor TM7 (EGF-TM7) adhesion GPCRs consists of 5 receptors; CD97 and EGF-like module-containing mucin-like receptors 1-4 (EMR1-4) (Figure 1.8). The group are predominantly limited in expression to cells of the immune system, with the exception of CD97 which is also expressed by haematopoietic, progenitor, muscle and epithelial cells (25)(26)(71). Therefore, unsurprisingly the family is functionally diverse, involved in immunological processes such as host defence, T-cell activation, inflammation, and myeloid cell differentiation and migration, alongside angiogenesis and carcinogenesis (29)(37)(72).

The extended N-terminal region is distinguished by a number of EGF-like tandem repeats which are responsible for mediating interactions with a variety of extracellular matrix and cell membrane components (27). Alternative RNA splicing is a common feature within the adhesion GPCR family and generates various receptor isoforms thought to increase functional diversity and regulation (73). EGF-TM7 isoforms differ in the number and arrangement of EGF-like repeats. CD97 has 3 isoforms, CD97-EGF 1-5, CD97-EGF 1,2,3,5, and CD97-EGF 1,2,5 whilst both EMR1 and EMR2 have 4 isoforms (25). Investigation of the extracellular region demonstrates extremely high sequence similarity between CD97 and EMR2, which differ by only 6 amino acids between the 5 EGF-domains (28). Similarly, the transmembrane domain of EMR2 is largely homologous to that of EMR3, and thus genes encoding CD97 and EMR3 are thought to have given rise to EMR2 by duplication events (74). EMR3 expression is very similar to that of EMR2, but is more highly expressed by granulocytes and therefore is a mature polymorphonuclear cell marker such as neutrophils and basophils (75). EMR1 expression is restricted to eosinophils, but the mouse homolog F4/80 is a well-defined marker for macrophage populations (76).

EMR4 is seemingly silent in humans, but upregulated in activated murine macrophages and expressed at much lower levels by bone marrow-derived dendritic cells and treated neutrophils (77). Interestingly, EMR1/ F4/80 is the only member of the subfamily that does not undergo post-translational cleavage (Figure 1.7).

	EMR1	EMR2	EMR3	EMR4	CD97
					
Expression					
Human	Eosinophils	Monocytes, Macrophages, Dendritic cells, Granulocytes	Monocytes, Macrophages, Dendritic cells, Granulocytes	Inactive	All leukocytes, Some non-immune
Mouse	Monocytes, Macrophages, Myeloid dendritic cells, Eosinophils	No	No	Macrophages, Monocytes, CD8+ dendritic cells	All leukocytes, Some non-immune
Splice variants	Yes	Yes	Yes	No	Yes
EGF domains					
Human	4-6	2-5	2	/	3-5
Mouse	4-7	/	/	2	3-4

Figure 1.8 ***EGF-TM7 subfamily characteristics overview.*** Schematic representation of example isoforms of each receptor in the subfamily. Cleavage is indicated by a strand break immediately before the membrane, which is absent in EMR1. Adapted from (M. Stacey, et.al. 2011)(78).

1.2.6 Group VIII Subfamily

Group VIII adhesion GPCRs consists of 7 receptors; GPR56, GPR64, GPR97, GPR112, GPR114, GPR126 and GPR128. The expression pattern of this group is extensive, and includes liver, gastric, immune, neural and muscle cells, and the epididymis (79). GPR56 is one of the most well characterised adhesion GPCRs and has a broad expression profile. Whilst the receptor is pivotal in development in the CNS (Section 1.2.8.3), it is also implicated in the immune system and is a distinct

marker for mature natural killer (NK) cells and T-cells, demonstrated by antibody studies (80). It acts to negatively regulate the immune effector functions of NK cells by suppressing production of inflammatory cytokines and degranulation (81). Like the EGF-TM7 receptors, GPR97 is largely expressed within the immune system, as well as vascular endothelial cells. The receptor remains orphaned but has some characterised signalling properties, including activating cAMP response element-binding protein (CREB) and NF-κB, alongside Rho GTPase (68).

1.2.7 Adhesion GPCR binding partners

Most adhesion GPCRs are considered orphan receptors, due to a lack of known ligands. However, a handful have characterised binding partners. CD97 interacts with 5 proteins; the complement regulatory protein decay accelerating factor (DAF)/CD55, extracellular matrix protein chondroitin sulphate, Thy1/CD90, and integrins α5β1 and αvβ3 via an RGD motif (31)(82)(83). This binding promiscuity is facilitated by the expression of multiple CD97 isoforms (Section 1.2.5); for example, CD97 EGF-1,2,3,4,5 binds chondroitin sulphate, whilst CD55 binds CD97 EGF-1,2,5. The GAIN domain also serves as a ligand binding domain; it is here that it interacts with Thy1/CD90. However, signalling has not been characterised as a result of this interaction. EMR2 also binds chondroitin sulphate via its EGF-repeats, which are almost identical to those of the CD97 EGF-1-5 isoform. Neither CD97 and EMR2 are activated by interaction with their ligands, however ligation of EMR2 with antibody 2A1 has been shown to stimulate receptor signalling (59).

GPR56 has two binding partners, interacting with type III collagen and tissue transglutaminase 2 (TG2), both of which are components of the extracellular matrix (84). GPR56 acts to inhibit melanoma growth and migration by interacting with and internalising TG2, preventing ECM cross-linking. Association with type III collagen stimulates intracellular RhoA signalling cascades by activation of $G\alpha_{12/13}$ and enables proper lamination of the cerebral cortex (85)(86). GPR126, an adhesion GPCR implicated in myelination of peripheral nerves, also binds ECM components, specifically type IV collagen and laminin-211, which acts to antagonise receptor function and increase cAMP-dependent signalling via activation of $G\alpha_s$ (56)(87).

Latrophilin proteins are highly expressed within the central nervous system and implicated in neuronal developmental processes. Latrophilin 1 interacts with a number of proteins, including teneurin, post-synaptic fibronectin leucine-rich transmembrane (FLRT) proteins, and pre-synaptic neurexins which are implicated in

the formation of synapse adhesion complexes (88). Meanwhile, Latrophilin 3 has recently been demonstrated to dimerise upon interaction with proteins Unc5 and FLRT2 as part of a ternary complex (89).

Finally BAI1, is expressed by macrophages, amongst other cell types, and binds phosphatidylserine to facilitate internalisation of apoptotic cells (90). The receptor enables macrophage recognition of Gram-negative bacteria via interaction of its extracellular thrombospondin repeats with lipopolysaccharide (LPS) (91). BAI1 also has an RGD motif within its NTF which mediates integrin αvβ5 binding, inhibiting proliferation of endothelial cells to possibly regulate brain angiogenesis (92). It is important to note that most receptor-ligand binding instances outlined here do not mediate receptor activation, emphasising the functional complexity of the group. Elucidation of the signalling mechanisms of adhesion GPCRs upon ligand binding will aid further understanding of their functional roles in key biological processes and how they themselves are regulated, subsequently increasing their potential for therapeutic exploitation.

1.2.8 Physiological and pathological function of adhesion GPCRs

The functional diversity of the adhesion GPCRs is exemplified by their expression within human physiological systems. The receptors are implicated in molecular and cellular functions such as cell polarity, motility, and differentiation, and key physiological systems such as the cardiovascular system, central nervous system, urinary and endocrine systems, and the immune system. They are also now being associated with sensing of mechanical stimuli, a crucial influence in various physiological systems. Furthermore, adhesion GPCRs are frequently being implicated in disease states such as development defects and tumorigenesis, yet they remain an unexploited therapeutic target. Outlined below are key example physiological processes in which adhesion GPCRs are pivotal, demonstrating their potential as future drugs targets.

1.2.8.1 Adhesion GPCRs in migration, adhesion and cytoskeletal regulation

The characteristic, long extracellular domains of adhesion GPCRs were originally defined due to their role in the formation of cell-cell or cell-matrix adhesions. Domains which facilitate this interaction include EGF-like repeats, leucine-rich repeats, and cadherin-like repeats. GPR56 facilitates cell-matrix interactions in both neuronal cells and haematopoietic stem cells. Loss of expression of GPR56 results in the downregulation of adhesive interactions with extracellular matrix components laminin and fibronectin (58). Extensive study has shown that adhesion and migration of cells are closely linked, although the extracellular signalling of adhesion-GPCR associated migration has not been clarified.

Cellular adhesion via CD97 is mediated by high affinity binding to integrins $\alpha_5\beta_1$ and $\alpha_v\beta_3$. Interaction with these transmembrane proteins stimulates increased endothelial cell motility and angiogenesis in an isoform dependent manner (72). The interaction of CD97 with CD55 under shear stress in the circulatory system results in the removal of the NTF from the CTF. Consequently, expression of the CTF is downregulated by internalisation, which may serve to prevent cell aggregation in the circulatory system (93). Interestingly this response is not seen with CD55-CD97 binding in the absence of shear flow. Experimental upregulation of CD97 has been demonstrated to activate matrix metalloproteinases responsible for the degradation of the extracellular matrix, in turn facilitating cell migration (94). Finally, expression of CD97 within the sarcoplasmic reticulum (SR) of skeletal muscle influences SR structure but does not appear to affect muscle function (95).

1.2.8.2 Adhesion GPCRs in cancer

Adhesion GPCR-regulated processes such as cell migration, adhesion and angiogenesis are all pivotal to tumour development. It is therefore unsurprising that members of the adhesion GPCR family are some of the most mutated GPCRs in human cancers (65). CD97, BAI1, GPR56, Latrophilin, VLGR1, GPR116 and GPR133 are all implicated in anti- and pro-tumorigenesis. Examples of adhesion GPCRs expressed at the invasive front of metastatic tumours are frequent, with metastatic potential linked to their upregulation.

Changes in the dynamic extracellular environment, such as increased ECM stiffness, are sensed by cell surface receptors. Alongside epigenetic changes, these mechanical cues act to promote cancer progression by stimulating reorganisation of

the cytoskeleton and subsequent signalling pathways (96). The PKD proteins, which also contain the GPS motif, have been identified as novel mechanosensitive receptors (Section 1.2.9) that are implicated in aggressive colorectal cancer phenotypes (96). With mounting evidence suggesting a role for adhesion GPCRs as mechanosensors and transducers (46)(61)(97), there is potential for the action of adhesion GPCRs in tumour progression to be force-dependent.

CD97 expression has been shown to be induced or upregulated in a plethora of malignancies including colon, stomach, brain, and pancreatic carcinomas, and correlates with overall prognosis and tumour grade (98). CD97 expressed on thyroid tumour cells promotes angiogenesis upon binding mechanosensitive protein integrin α5β1. When bound in tandem with chondroitin sulphate, endothelial cell invasion is initiated (72), alluding to a role of CD97 in mechanosensing. In pancreatic cancer, binding to the complement regulating protein DAF further increases cancer invasion and metastasis (83).

Despite sharing structural homology with CD97, EMR2 is detected in very few tumour cell lines in which CD97 has been implicated (99). Instead, it is a marker of breast cancer, being upregulated in the breast carcinoma cell line SK-BR3 (78), which also expresses the EMR2 binding partner chondroitin sulphate. Increased expression of EMR2 correlates with relatively poor prognosis and high tumour grade.

Contrasting CD97, GPR56 influences anti-metastatic processes by reducing cell adhesion and fibronectin deposits in the ECM which limits melanoma growth and metastatic potential (100). Expression of full length GPR56 also negatively regulates VEGF production during melanoma angiogenesis. Association of the NTF with CTF limits the ability of melanoma angiogenesis and aggressiveness, processes that are largely upregulated when the CTF is expressed alone (101). The contrasting roles of the receptor subunits provides another example of NTF regulation of cell activity.

1.2.8.3 Adhesion GPCRs in the nervous system

Adhesion-GPCRs are highly expressed in neuronal tissue and are associated with neuronal development and disease. A total of 22 mutations mapped to the GPR56 gene have been shown to be the cause of the recessive cortical development disease BFPP (102), in which the brain cortex displays an excessive number of folds and abnormal lamination (103). These mutations cause BFPP by a number of mechanisms due to their location within the protein; an inability to undergo post-

translation cleavage, interact with ligands, shedding from the cell surface, and a loss of surface expression have all been implicated in disease (45). Aside from BFPP, GPR56 expression in oligodendrocytes regulates myelination during development, and also influences migration of neural progenitor cells, a lack of which is associated with cortical malformation (104)(105).

BAI1 and BAI3 are critical for neuronal dendrite development, maturation and stability (106). Single nucleotide polymorphisms within these genes have been linked with both schizophrenia and addiction-predisposition disorders (107). The Latrophilin proteins are located at synapses, expressed both pre- and post-synaptically, where Latrophilin 1 in particular interacts with structural proteins via its intracellular C-terminus (108). Latrophilin 3 is exclusively expressed in the brain and is found at its highest levels immediately after birth, and has been associated with attention deficit hyperactivity disorder (ADHD) as a genetic risk factor for different types of the phenotype (109). Finally, murine models for VLGR1 deficiency demonstrated the development of epilepsy, corresponding with expression studies showing VLGR1 presence on oligodendrocytes, therefore suggesting involvement of VLGR1 in myelination (30).

1.2.8.4 Adhesion GPCRs in the immune system

The EGF-TM7 sub-family of adhesion GPCRs are largely leukocyte restricted, with a diverse involvement in immunological processes, appearing to be pivotal in myeloid maturation and activation. CD97-ligand binding influences the migration pattern of continually circulating myeloid cells. CD97 is vastly upregulated during lymphocyte activation (31), causing increased migration to and adhesion at sites of inflammation. CD97 binding to Thy-1/CD90, via a site within the GAIN domain, on the surface of endothelial cells in psoriatic lesions causes a firm adhesion of polymorphonuclear cells (PMNs) during inflammation (110).

Closely related EMR2, found expressed on monocytes, macrophages, neutrophils and dendritic cells, is implicated at sites of inflammation, where expression is upregulated by myeloid cells and foamy macrophages (111). Studies by Chen, T.Y. *et al* (2011) revealed EMR2 potentiation of the neutrophil response to inflammatory stimuli; both adhesion and migration of the immune cell were upregulated, as well as the production of reactive oxygen species (ROS) and degranulation. This manipulation of the neutrophil phenotype is essential for rapid elimination of invading pathogens. Binding of EMR2 by its monoclonal antibody 2A1 stimulates an enhanced

neutrophil response upon exposure to inflammatory stimuli, and a modulation to the cellular cytokine expression profile. Furthermore, EMR2 ligation also inhibits the effect of LPS of delaying neutrophil apoptosis (59), causing a shorter cellular lifespan but a more rapid resolution of inflammation. Recently, expression of a missense varient of EMR2, expressed by mast cells, has been implicated in a rare, hereditary, vibration-induced urticaria. Exacerbated mast cells were found to be desensitised to IgE-independent degranulation, induced instead by vibration, alluding to mechanosenstive activtiy of EMR2 (46). Currently this is the only gain of function mutant known amongst adhesion GPCRs.

Other adhesion GPCRs expressed in immune tissue include the closely related GPR97 and GPR56, which have been identified in both lymphocytes and granulocyte cell populations. GPR56 influences granulocyte migration, despite having no known ligand binding and signalling functionality in the immune system (80). GPR97 on the other hand, has been associated with determining follicular B-cell fate, and inflammation in the gut caused by high fat diet (112)(113). To date, GPR97 remains an orphan without a characterised binding partner, yet small molecule compound screens have identified beclomethasone dipropionate as an agonist of GPR97 (68). Regardless of the clear involvement of adhesion GPCRs in the immune system, signalling because of adhesion and migration in many cases remains ambiguous, therefore small molecular screening approaches as such provides an invaluble tool in interrogating receptor function and signalling.

1.2.9 Adhesion GPCRs as mechanosensors

As already outlined, adhesion GPCRs are pivotal to cellular functions such as cell migration, polarity and adhesion, as well as being associated with tumorigenesis, processes that all involve changes in mechanical environment. Furthermore, as the NTF of many adhesion GPCRs contains multiple adhesive protein domains that facilitate interactions between cells and the ECM, it is plausible that these receptors are involved in mechanical regulation of cellular processes.

In recent years, adhesion GPCRs have indeed been associated with mechanosensitive functions. For example, expression of GPR56 in skeletal muscle mediates signalling in response to mechanical overload, inducing muscular hypertrophy upon binding to its ligand collagen III (114). Furthermore, GPR56 expression was found to be increased, alongside expression of $G\alpha_{12/13}$ subunits, when skeletal muscle was subject to pressure in the form of resistance training (114).

Meanwhile, Scholz et al. (2015) identified the *Drosophila* Latrophilin homolog, *dCirl,* as a key receptor in shaping responses by neuronal sensitisation to auditory, tactile and proprioceptive cues (61). Intriguingly, although not absolutely required for development, *dCirl* was also found to regulate movement; knockout *Drosophila* larvae appeared to have thwarted forward movement in comparison to control larvae (61). It would appear that *dCirl* also co-localises with the mechanosensory channel NOMPC and modulates the response of neurons to mechanical stimuli by adjusting NOMPC activity (97).

Expression of VLGR1 by hair cells of the inner ear has also been implicated in mechanotransduction, in response to soundwave-mediated hair cell movement. The receptor is a part of the ankle-link complex between stereocilia, and is therefore a key protein in the normal development in mechanosensitive hair bundles (33). Mutation of the *VLRG1* gene causing faulty expression or activity is the cause of type 2 Usher's syndrome, the most common hereditary form of deaf-blindness (115). Its aberrant expression prevents proper formation of hair bundles through the loss of fibrous links between stereocilia, although the molecular basis behind disease development remain unclear (79).

Finally, a missense mutation within the NTF of EMR2, upstream of the GPS, has been proven to be the cause of a rare vibratory urticaria. The mutation at position 492 from cysteine to tyrosine (C492Y), causes reduced mechanical stability of the EMR2 GAIN domain. Upon exposure to vibration, the NTF and CTF dissociate more-readily than those of wild type EMR2, which mediates mast cell degranulation, resulting in patients presenting with hives (46). These results indicate a role for adhesion GPCRs as novel mechanoceptive proteins, although the field remains largely unexplored.

A definitive mechanism of activation remains to be determined for the entire class. It certainly seems unlikely that a single process will apply to all 33 receptors, with examples of activation proven when the receptor has not undergone cleavage. Although mounting evidence signifies the *Stachel* peptide as a key mediator of activation, the way in which it is exposed is still undefined. Conformational changes upon exposure to force provides a potential mechanism for uncovering. For example, fibronectin contains a number of cryptic binding sites that are revealed when the protein is stretched, consequently inducing fibril formation (116). In the case of the GAIN domain, the secondary structure implies that a high force would be required to induce separation of the final β-strand from the wider GAIN domain; a process that is likely thermodynamically unfavourable. However, experiments have determined

that high forces are responsible for mediating biological activity. Indeed, vascular endothelial cells under shear flow can be exposed to over 1 nano-Newton (nN) (116). Understanding the mechanical properties of the GAIN domain may therefore provide a new paradigm for the molecular activation and function of the adhesion GPCRs.

1.3 Mechanical force in physiological systems

Physiological communication is typically associated with receptors identifying ligands and mediating an intracellular response. However, it is crucial that physical forces are also recognised. Mechanosensing describes the ability of cells and tissues to respond to mechanical cues from their microenvironment. The conversion of an extracellular force into an intracellular signal is termed mechanotransduction, which is essential for responding and adapting to the mechanical stimuli (117). Mechanical force in the extracellular matrix plays a fundamental role in physiological development, homeostasis, and number of pathologies including cardiovascular and pulmonary disease, and cancer (118). Perturbation of the cell membrane is also thought to mediate activation of cytoplasmic proteins, like G-proteins, to mediate signal transduction (119). For example, $G\alpha_{13}$ is activated upon cardiomyocyte sensing of high blood pressure by laminin, actin and myosin (120) resulting in pressure overload-induced cardiac remodelling and hypertrophy (121). Despite this, the molecular basis of mechanotransduction is uncertain with questions surrounding why certain cells are more sensitive to the pressures exerted by the ECM, for example (122). Gaining a more comprehensive picture of the role mechanosensing molecules play will generate possibilities for therapeutic intervention.

Cellular adhesion such as cell-cell and cell-matrix interactions between ECM components like collagens, laminins and fibronectin (123), and cell receptors provide a mechanical connection between cells and their external environment (Figure 1.9). Furthermore, intracellular forces mediated by actin polymerisation and myosin-dependent contraction of the cytoskeleton are also applied to the cell surface. Therefore force is commonly applied directly to transmembrane receptors, often being the first molecules responsible for mechanosensing (117). Critical receptors in tension-sensing, and therefore maintenance of tissue structural integrity and functionality, are integrins. Integrins are the core component of multi-protein focal adhesions that are highly dynamic to facilitate responses to changing mechanical stress. For example, integrins αvβ3 and α5β1 bind to fibronectin and fibrinogen in the ECM alongside cell surface growth factor receptors, and intracellular cytoskeletal proteins (124). This binding activates integrins to mediate bidirectional signalling

between the ECM and cytoskeleton, resulting in cell motility and cycle progression in wound repair (125).

Examples of physiological forces include tissue stiffness and rigidity, circulatory shear flow, and blood pressure, which influence cell migration, vascular remodelling and cell stretching respectively. Membrane fluidity can be increased by shear flow which in turn increases the lateral mobility of αvβ3, for example. This mediates a conformational change to activate and augment the affinity of the integrin for its ligands in the ECM, which is essential for activating downstream signalling molecules such as RhoA (124). Moreover, changes in shear flow can mediate neutrophil activation, inducing changes in membrane fluidity and therefore cell morphology (126).

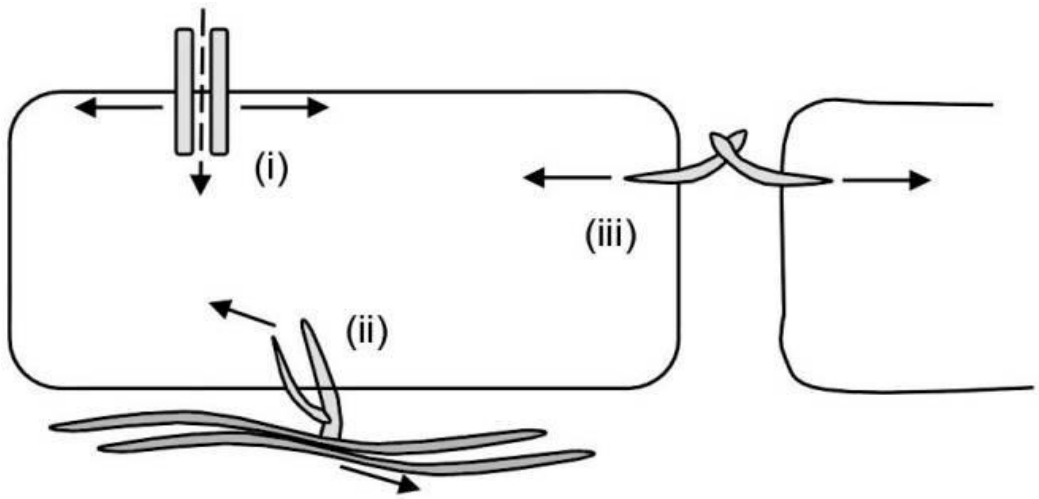


Figure 1.9 ***Mechanical stimulation experienced by plasma membrane proteins.*** The three ways in which cells sense mechanical change in their microenvironment include (i) mechanosensitive ion channels, (ii), mechanosensitive interactions between the ECM and transmembrane proteins, (iii), cell-cell interactions. Arrows are representative of the direction of stretch/movement/force experienced by these mechanosensors. Taken from (I. Muhamed et al. 2017)(127).

Mechanical force such as pressure-loading influences bone strength and renewal, and cardiac tissue regulation. Osteocytes detect mechanical load through currently undefined mechanosensors, though integrins, GPCRs and mechanosensitive channels are likely candidates (128). This consequently activates a chain of mechano-dependent events. Osteocytes signal to osteoblasts and osteoclasts via paracrine factors such as prostaglandin E2 and insulin-like growth factor (128). They also release the glycoprotein sclerostin to regulate the Wnt signalling pathway in

osteoblasts via modulation of transmembrane receptor Lrp5 (129). This signalling pathway is responsible for regulating bone formation (130). Meanwhile, detection of stretch and pressure overload by cardiomyocytes regulates heart function via activation of αvβ1 which stimulates the extracellular-regulated kinase (ERK) signalling pathway. This pathway regulates expression of α-actin and when perturbed, can culminate in mechanical-stress induced hypertrophy (131).

The mechanical properties of the biological microenvironment are increasingly recognised for their role in promoting tumour development, metastasis and invasion, and are therefore becoming a focus for potential therapeutic intervention. Mechanosensitive integrin α6β4, for example, binds to ECM protein laminin and is activated in the presence of P-cadherin, the expression of which is strongly associated with a metastatic and invasive phenotype for breast cancer (132)(133). Activation of downstream signalling pathways by focal adhesion-sensing of ECM stiffness eventually activates tumour-associated transcription-pathways that regulate proliferation, survival and metastasis (96). Other cell surface receptors involved in mechanosensing and tumour development are the PKD proteins polycystin (PC) 1 and 2, outlined in Section 1.2.6.2. Their upregulated expression has been detected within focal adhesions and are involved in the promotion of epithelial-to-mesenchymal transition (EMT). Furthermore, the C-terminus of PC1 has been shown to translocate to the nucleus under mechanical stress to influence cell differentiation (96).

Various factors influence the forces and stresses experienced *in vivo* by cells, including cell size, morphology, and the local mechanical environment. With that in mind, physiologically significant forces have a vast range, from picoNewton (pN) to kiloNewton (kN) forces in the inner ear and musculoskeletal system respectively (134). Examples of the forces applied to tissues in the body are outlined in Table 1.2. The implications of mechanical force in key developmental, homeostatic and pathophysiological processes exemplifies why defining the molecular mechanisms of mechanotransduction is of the utmost importance for understanding pathological progression and the development of therapeutics.

Physiological Environment	Force Experienced / Exerted
Vascular endothelium	0.5-5 μN
Arteries	0.5-5 μN
Cartilage	30 kN
Achilles / Patella Tendons	10 kN
Auditory hair cell transducer channel	300-500 fN
Adherent cells	10nN

Table 1.2 ***Overview of mechanical forces experienced within physiological systems.*** *In vitro* studies have facilitated the determination of force values in large anatomical structures by studying traction forces generated by single cells. kN; kiloNewtons, fN; femtoNewtons, nN; nanoNewtons, μN; microNewtons. Values taken from (Addae-Mensah KA, Wikswo JP) (2008)(134).

1.3.1 AFM: investigating mechanical protein unfolding

Mechanical forces are largely responsible for protein remodelling, influencing enzymatic activity, ion channel opening, cell signalling, and cell adhesion (119). Understanding the mechanical stability of proteins and complexes *in vitro* at a molecular level helps decipher biological functionality and whether they are likely to be able to withstand or be vulnerable to force *in vivo*. Techniques used to investigate the implications of force on biological molecules include optical/magnetic tweezers, patch clamping and atomic force microscopy (AFM). These techniques expose single molecules to an applied force in the range of 0.001 pN-10 nN; typically proteins are exposed to forces up to 1 nN (135).

Observing single protein domain unfolding and interrupting protein interactions via disruption of non-covalent interactions using AFM has become a mainstay technique. It has dynamic force range of 1-10,000 pN, with high spatial accuracy and the ability to investigate unfolding under physiologically relevant conditions (135)(136). AFM instruments classically consist of an optical head into which a thin, flexible, gold coated silicon nitride cantilever, is inserted. This is used as a probe to interact with the molecule under investigation. A laser is positioned onto the end of the probe, and laser deflection upon application of force to the molecule is detected by a photo-diode (Figure 1.10). The optical head is lowered to bring the cantilever into contact with the sample before being withdrawn. Both approach and retraction occur at a defined constant velocity. The deflection observed equates to a measurement of force, which is calculated according to Hooke's law using the known spring constant of the cantilever, usually in the range of 10-40 pN/nm.

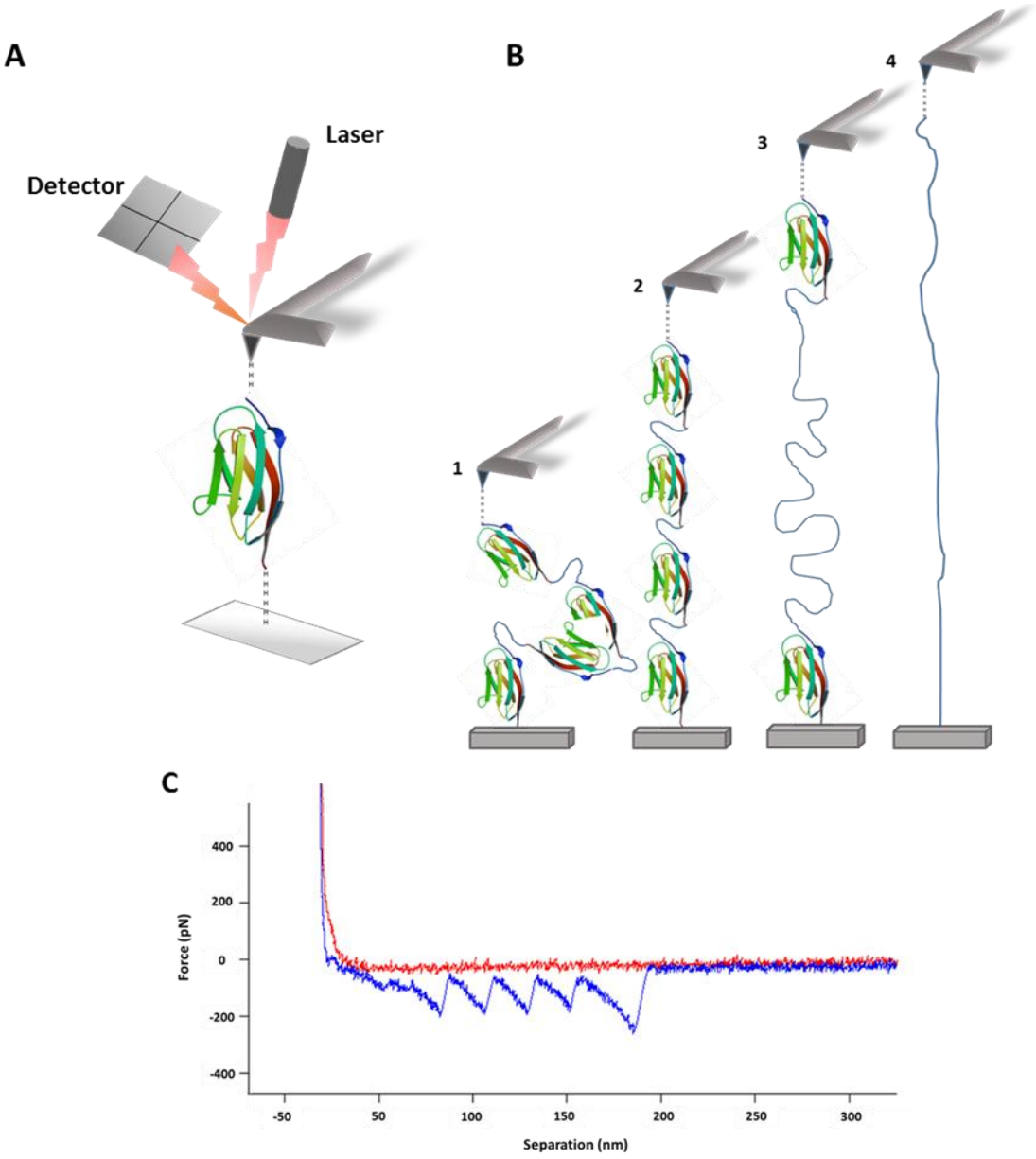


Figure 1.10 ***Single molecule force spectroscopy (SMFS) for studying the mechanical unfolding of single domains in a polyprotein*****.** (**A**) Schematic representation of a typical SMFS experiment. The protein of interest picked up by a silicon nitride probe with a gold-coated cantilever by either adsorption, or linkers to attach to a single molecule. A laser is positioned over the cantilever, which detects protein deformation upon movement of the cantilever. (**B**) Cartoon of $(4)I_{27}$ domain repeats unfolding under pulling force. When the tip approaches the surface, it picks up the protein chain (1). The chain fully extends when the tip is retracted (2) and as more force is applied, domains begin to unfold (3), until the entire chain is unfolded (4). Eventually enough force is applied to facilitate protein unbinding from the tip. (**C**) A typical force-extension profile gained from unfolding of a polyprotein consisting of 4 I_{27} repeats; the first 4 peaks represent each repeat unfolding, whilst the final demonstrates cantilever unbinding from the unfolded chain.

When bringing the cantilever tip into contact with the protein of interest, very rarely will the folded molecule be correctly suspended between the surface and tip. Moreover, the tip is likely to come into contact with the protein at a random point and therefore not encompass the entire domain. Instead, studying the mechanical strength and unfolding of protein domains via engineered polyprotein constructs is advantageous in comparison to a single protein.

Another method employed to circumvent this problem is functionalisation of cantilevers. This encourages the tip to interact specifically with the molecule, ensuring tethering of the sample at each end of the protein chain and further reducing the chances of high levels of non-specific interactions. After cantilever-protein interaction, the cantilever is retracted and pulls the molecule until a force is applied that results in unfolding of a domain (Figure 1.10B), represented by a single peak on the resultant force-extension profile (Figure 1.10C). Ultimately, the protein unbinds from the cantilever, and the force rapidly returns to zero before the process is repeated. The dissociation force, or rupture force (Fr), is represented by the final peak on the profile. Pulling of polyprotein constructs rather than single domains, allows for a high signal to noise ratio (137), removing the molecule from the region of non-specific interactions which would otherwise be seen with a single molecule. Additionally, having constructs with repeats of the same protein domain means a characteristic unfolding “fingerprint” can be gained (Figure 1.10C), aiding the determination of mechanical properties.

From single molecule force spectroscopy (SMFS) experiments the rupture force of proteins domains and protein complexes can be estimated using Hooke’s law, and the length of the domains before dissociation can be calculated using the Worm-like chain (WLC) model. Briefly, Hooke’s law (Section 2.5.2.1) describes the amount of force applied to a spring or molecule that causes it to stretch, which is linearly proportional to the extension or stiffness of the spring/molecule (138). The WLC describes the behaviour of semi-flexible polymers, calculating sample contour length (Lc) which represents the maximum length of the fully unfolded molecule, and uses persistence length as a measure of local stiffness, which equates to the length between two α-carbons of two amino acids (0.4 nm)(139). The WLC has been used extensively to estimate the length of polymers by fitting to force-extension profiles (Figure 1.10C) gained in SMFS experiments for both protein and DNA molecules (140)(141). However, its application may be limited under very high force where overstretching of bonds occurs (142). Lc is important to be able to identify whether the event observed is specific, and to be able to distinguish between which domain

has unfolded. The expected Lc can be calculated from structural knowledge. As many proteins have a multi-domain structure, it is likely that they unfold in a multi-step process, therefore contour length can also be used to determine how much protein has unfolded by comparing the observed and the expected Lc. Another factor to consider in observing the unfolded length of molecules is the presence of disulphide bonds within a molecule. Unfolding of residues between the cysteines mediating disulphide bond formation is often prevented at lower forces, therefore reducing the length of the total observed Lc. Prior knowledge of their position is beneficial when comparing calculated and observed Lc.

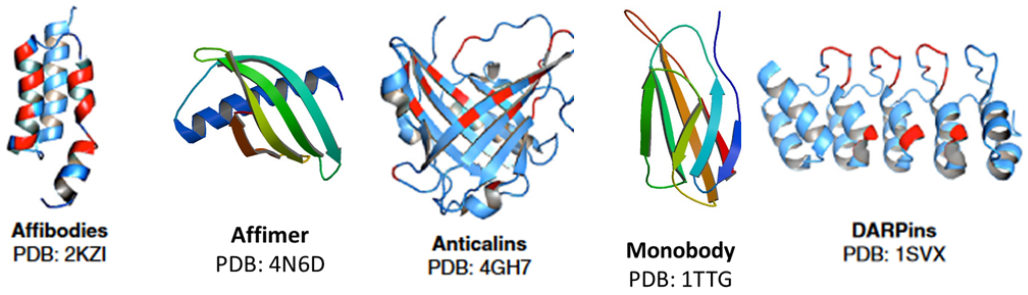
1.3.2 Protein structure and force resistance

An important indicator of mechanical strength is the secondary structure of proteins, which has evolved to facilitate protein-specific activity. AFM has been used in various instances to test the mechanical strength of α-helical and β-sheet structures. SMFS experiments have investigated the unfolding of proteins such as human titin immunoglobulin-like domain 27 (I27), fibronectin and spectrin. These experiments have ultimately demonstrated that α-helical structures are much less force-resistant, unfolding in a single step, in comparison to β-stranded structures which provide greater resistance to force due to the presence of hydrogen bonds between strands, or mixed topology proteins which offer varied unfolding resistance (143)(144). Further, the topology of β-strands confers mechanical strength, demonstrated by the amount of force required to break multiple hydrogen bonds simultaneously as part of an antiparallel β-sheet, compared to that of breaking them successively in “zipper” formation in a β-barrel (145). The direction of the applied force also influences the force resistance of proteins, exhibited by bacterial haloalkane dehalogenase (HaloTag) in SMFS pulling experiments. When anchored covalently to a surface and pulled from its N-terminus as part of a polyprotein, the HaloTag unfolds under the force of 491 pN, a force 4 times greater than that required when pulling from the C-terminus, at 131 pN (146). Likewise, an investigation by Brockwell et al. (2003) of the mechanical unfolding properties of bacterial pyruvate dehydrogenase enzyme E2lip3 demonstrated that when a force of ~180 pN was applied in parallel to β-strands, mechanical distortion was resisted. However, when as little as 15 pN was applied perpendicular to the strands, protein unfolding could not obviously be detected by SMFS (147).

1.4 Protein characterisation with non-Ig binding proteins

Antibodies provide a diverse resource for interrogating protein expression, activation, function, and interactions. However, despite still being the most commonly used class of binding protein, several issues surround their use including high production costs, instability, target affinity, and reproducibility. Synthetic binding proteins, or non-Ig scaffolds, present a versatile alternative. They are composed of either single domains or repeats generated from proteins of plant, human, bacterial or insect origin (148) and are characteristically less than 200 amino acids in size. Binding proteins have previously been selected due to having a distinct fold with desirable function in existing physiological systems (149). Large scale libraries are generated by phage, yeast or ribosome display, and typically the protein scaffolds do not contain disulphide bonds so they can be expressed at high yield in bacterial systems (150). They often also completely lack cysteines to enable the incorporation of free cysteine residues for biotinylation or addition of other tags for functional use.

A growing number of synthetic proteins have been generated with various structures; some are purely α-helical or β-sheets, a combination of α-helices and β-strands, whilst others form β-barrels. Each is diversified by the presence of surface patches or loops of randomized amino acids. These act as molecular recognition interfaces for a vast array of targets without increasing the size of the protein (150). Amongst the most well characterized are Affibodies, DARPins (designed ankaryin repeat proteins), Affimers, Anticalins, and Monobodies. Their basic structural characteristics are outlined in Figure 1.11. Anticalins are based on the lipocalins, proteins that transport hydrophobic molecules such as steroids around the body. It is this ability to bind small molecules that made them an attractive scaffold (151). DARPins are larger than other non-IG binding proteins, with repetitive ankaryin structural units generating a larger binding site. Monobodies have a fibronectin type III (FN3) scaffold that lacks disulphide bonds despite having an immunoglobulin fold (151). Mimics of these proteins have been generated commercially for use in therapeutics and are termed Adnectins (152).



Protein Scaffold	Scaffold Origin	Structure	Variation
Affibody	Protein A Z-domain	α-helical	Randomised helix
Affimer	Cystatin	β-sheet with single α-helix	Two randomised loops
Anticalin	Lipocalin	β-sheet with α-helical terminus	Randomised loops and β-strand
DARPin	Ankyrin repeats	α-helical and β-turn	Randomised helix and β-turn
Monobody	Fibronectin III	β-sheet	Randomised loops and β-strand

Figure 1.11 ***Structural characteristics of a selection non-IG binding proteins*****.** Adapted from (R Vazquez-Lombardi et al. (2015))

Today, a number of non-Ig binding proteins are being exploited as therapeutic alternatives to antibodies to circumvent issues with molecule size, post-translational processing and stability, and delivery (153). DARPins are currently utilised in breast cancer and macular degeneration treatment (154). Affibodies, generated from Protein A Z domains, are currently in clinical trials to investigate their potential for targeting inflammation and autoimmune responses, as well already being used for imaging of metastases via PET scans (155). Aside from their clinical applications, non-Ig binding proteins have become a valuable tool in molecular biology. Many studies have outlined roles such as crystallisation chaperones and protein interaction inhibitors. The ability to block specific protein-protein interactions indicates that these small proteins likely bind functional sites, and therefore potentially act as modulators of biological functions. This makes them an attractive tool, for example, in deciphering the function of orphan receptors.

1.4.1 Affimer molecules

Affimers, originally termed Adhirons, are particularly stable non-Ig proteins with a melting temperature of 101°C. These monomeric protein scaffolds are based on the plant protein phytocystatin, and contain two randomised 9-amino acid loop regions that facilitate target recognition with affinity in the nanomolar range, mimicking that of antibody binding (156)(157). This ability to bind targets with high affinity makes them a viable and advantageous alternative to both small molecule inhibitors and antibodies, which can act via allosteric modulation and stimulate receptor activation, as demonstrated by GPR56 antibody-stimulated signalling (158). This characteristic of Affimers is therefore attractive in deorphanizing signalling pathways.

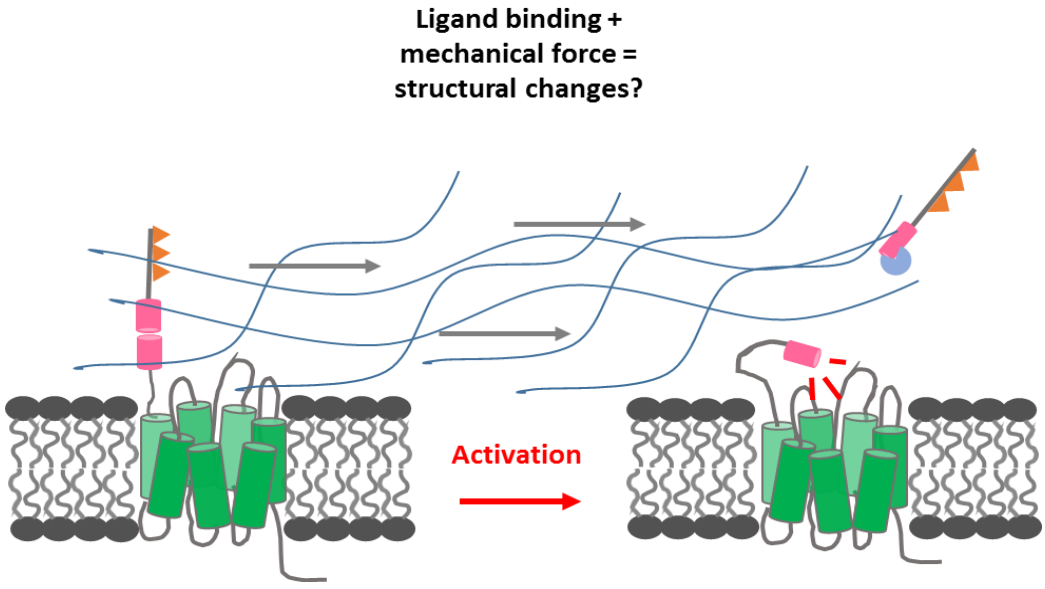
Affimers are generated by ‘phage-display, utilising filamentous bacteriophage. Briefly, the protein of interest is inserted into a tail coat protein gene which “displays” on the surface of the phage. Phage are then screened against an immobilised target protein, incubated to allow time for binding, and subsequently washed to remove phage that are not expressing a specific binding partner on their tail surface. The specific phage are then selected and amplified in a suitable bacterial host, which reduces the number of non-specific phage. This is repeated in a process known as panning to steadily increase the population of phage expressing the positively binding coat protein (159). After isolating the desired phage, a host is again infected and the resulting phagemid can be purified, and the DNA excised to confirm the binding sequence.

Thus far, Affimers have proven to be highly versatile molecular biology tools, in both research and therapeutics. The technology has been used by a number of groups to inhibit protein-protein interactions, interrogate protein function, and as an antibody replacement (160)(161). A study carried out by D. Hughes et al. (2017) demonstrated their use as effective inhibitors of SUMO2-dependent protein-protein interactions in cellular processes such as transcription regulation, apoptosis and protein stability (162). More recently, Affimers have successfully been generated as a tool for visualising F-actin in live cell imaging, proving to be a suitable alternative to eGFP fusion proteins (163). Affimers are also being used in clinical settings, and have been utilised as novel biosensors in detection of the cancer biomarkers Her4, detecting expression between 1 pM and 100 nM (164) and Glypican-3, a marker for hepatocellular cancer (165) illustrating their versatility in both academia and healthcare.

1.5 Thesis aims

The adhesion GPCRs are rapidly becoming established in a plethora of critical physiological and pathological processes, and more recently, mechanotransduction. Interestingly, despite several adhesion GPCRs having characterised binding partners, and evidence towards coupling to heterotrimeric G proteins for several receptors, few examples of ligand-induced signalling have been determined. Although identification of a tethered agonist peptide within the extracellular GAIN domain has facilitated substantial progress in characterising downstream signalling, it is unclear how this region is exposed to mediate activation. As discussed in this chapter, the precise mechanism by which this receptor family is activated and subsequently regulated remains elusive. However, it is clear that the GAIN domain is pivotal and appears to prevent augmented receptor activity.

We suspect that this conserved domain acts as a sensor of soluble and mechanical cues and undergoes conformational changes as a result, liberating the *Stachel* peptide to an extent that facilitates receptor activation, stimulating intracellular signalling pathways to mediate a functional response to an extracellular signal.



The GPCR superfamily have proven to be highly druggable targets once characterised. With an expanding repertoire of associated pathologies, it is crucial to understand the molecular mechanisms governing the action of adhesion GPCRs in order to exploit their therapeutic potential. Therefore, the primary aims of this thesis are as follows:

- Chapter 3; to elucidate the atomic resolution of a GAIN domain to identify key residues implicated in structural stability and receptor activation.
- Chapter 4; to decrypt the signalling pathways of adhesion GPCRs, interrogating mechanical stimulus and non-Ig surrogate ligand-induced signalling.
- Chapter 5; to interrogate the mechanical stability and properties of the GAIN domain in response to applied force using AFM.

Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Buffers

Bacterial lysis buffer: 20 mM Tris pH 8, 10% Glycerol, 5 mM $MgSO_4$

2YT bacterial media: 1 litre preparation; 16g Tryptone soya broth, 10g Yeast extract, 5g NaCl (Oxoid, Thermo Scientific)

SOC bacterial media: 2% Tryptone, 05% Yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 20mM Glucose

Cell freezing media: Foetal Calf Serum, 10% DMSO

Ni^{2+} column buffer: 500 mM NaCl, 20 mM Tris pH 7.4, 10% Glycerol, 20- 500 mM Imidazole

Size exclusion buffer: 150 mM NaCl, 20 mM TRIS pH 7.4

Western blot transfer buffer: 90 mM glycine, 25 mM Tris 20% Methanol

Western blot blocking buffer: Tween-20 0.1% in PBS + 5% milk

SDS-PAGE running buffer: 25 mM Tris, 90 mM glycine, 0.1% SDS

SDS-PAGE loading buffer: 62.5 mM Tris pH 6.8, 10% Glycerol, 2% SDS, 0.001% bromophenol blue +/- 5% β2-mercaptoethanol

Coomassie stain: 40% Methanol, 10% Acetic Acid, 0.05% Brilliant Blue R- 250, in dH2O

Coomassie de-stain: 10% Acetic Acid, 40% Methanol, in dH2O

TBE: 0.09 M Tris, 0.09 M Boric acid, 2 mM EDTA

6x DNA loading dye: 0.05% Orange G, 30% glycerol

ELISA blocking buffer: 2% BSA, 0.1% Tween-20 in PBS

Piranha Solution: 3:1 0.5 M (>95 %) H_2SO_4 to 30 % (v/v) H_2O_2

2.2 Molecular Biology

2.2.1 Agarose electrophoresis

Gels were prepared with electrophoresis grade agarose dissolved in 1 x TBE containing 0.2 µg/ml ethidium bromide. Samples were mixed with 6x DNA loading dye and run alongside a 1 kb or 10 kb DNA ladder depending on sample size, at 100 V. DNA was observed and photographed under short-wave UV light for observation, or long-wave UV light if further cloning was required.

2.2.2 PCR

The proteins of interests' sequences were amplified from template DNA using sequence-specific primers in conventional PCR using Q5® High-Fidelity DNA Polymerase. Samples of amplified DNA were run on 1% agarose gel and compared to the appropriate DNA ladder to verify product size. Positive bands were excised from the gel and purified using Qiagen QIAquick Gel Extraction Kit according to the manufacturer's instructions.

2.2.3 Overlap extension PCR

For polyprotein constructs used in AFM experiments, 5X Phusion® High Fidelity DNA polymerase protocol was followed according to the New England Biolabs (NEB) manufacturer's instruction. For example, I_{27}(1+2) and CD97-GAIN domain DNA fragments were joined, whilst I_{27}(3+4) and Halo domain DNA fragments were joined. Forward I_{27}(1+2) and I_{27}(3+4), and reverse CD97-GAIN and Halo-domain primers were added to the master mix.

2.2.4 Gibson assembly PCR

Following initial overlap extension PCR, I_{27}(1+2) - CD97-GAIN domain, and I_{27}(3+4) - Halo domain DNA fragments were mixed with 2X Gibson Assembly® Master Mix (NEB) at a 3-molar excess to the 100 ng pSectag2B vector DNA. Samples were incubated for 1 hour at 50˚C and stored at -20°C for subsequent transformation. The same protocol was followed for all polyproteins generated for AFM.

2.2.5 Site directed mutagenesis

QuikChange™ mutagenesis was carried out according to manufacturer's protocol (Agilent Technologies). 125ng of each appropriate forward and reverse mutant primers, 2.5 units *PfuUltra* High Fidelity polymerase were added to the reaction master mix and dilutions of template ranging from 20-100ng were made, before addition of 1µl QuikChange Lightning Enzyme. Subsequent reaction mixtures were

digested with Dpn I enzyme for 5 minutes at 37°C. The resultant reaction was used to transform XL-10 Gold supercompetent cells according to standard protocol. Cells were grown on 1% agar plates containing selection antibiotic and incubated overnight at 37˚C. Colonies were grown up in 2YT media and plasmid DNA was extracted by QIAprep Spin Miniprep Kit (Qiagen) before being sent to GATC-Biotech, Germany for Sanger-sequencing to confirm sequence mutagenesis was successful.

2.2.6 Restriction digest

Gel purified PCR products and vectors were digested using the appropriate restriction enzymes with 1% BSA in optimal buffers (NEB) at 37°C for 3 hours. Alkaline phosphatase was added to the vector for the final hour of incubation to prevent vector re-ligation. The digest products were run on a 1% agarose gel then excised and purified using QIAquick gel extraction kit (Qiagen) following manufacturer's instructions.

2.2.7 Ligation

Digested PCR products and vectors were ligated at a molar ratio of 1:3, with 1 ng of insert and 3 ng of vector. T4 ligase was added to the manufacturer's instruction (Invitrogen, USA), and the reaction was incubated at 16°C overnight.

2.2.8 Transformation

Ligation mix was added to thawed competent bacteria on ice for 30 minutes. The bacteria/DNA mix was heat shocked at 42°C for 30 seconds and returned to ice for a further 2-minute incubation. SOC medium was added to the mixture to a total volume of 1 ml and incubated for 1 hour at 37°C. Transformed bacteria were then plated on 1% agar containing the appropriate selective antibiotics and incubated at 37°C overnight.

To confirm successful ligation and transformation, a colony PCR was performed using Taq polymerase (Promega, USA) and appropriate primers. Each reaction was analysed by 1% agarose gel electrophoresis, and positive colonies were grown in 2YT media at 37°C overnight. Plasmid DNA was extracted by QIAprep Spin Miniprep Kit (Qiagen) before 80-100 ng of insert DNA sequence was confirmed by Sanger sequencing at GATC-Biotech, Germany.

2.2.9 Competent Cell Production

A single colony of MAX Efficiency® DH10Bac (Invitrogen) was grown overnight at 37°C with shaking (250rpm), in 5ml LB medium containing 10 µg/ml tetracycline and

50µg/ml kanamycin. 4ml of overnight culture was inoculated in 400ml LB medium containing aforementioned antibiotics and grown at 37°C with shaking to OD_{590} 0.375. Cells were (aliquot) into 8 pre-chilled tubes and incubated on ice before centrifugation at 4000g for 15 minutes. Pellets were re-suspended in 5 ml ice cold $CaCl_2$ solution and centrifuged for 15 minutes. Again, pellets were re-suspended in 5 ml $CaCl_2$, incubated on ice for 30 minutes, and centrifuged for 15 minutes at 4000g. After re-suspension in 2ml $CaCl_2$ solution, 250µl aliquots were flash frozen and stored at -80°C.

2.2.10 SDS-Polyacrylamide gel electrophoresis (PAGE)

Protein samples were mixed with loading buffer and boiled for 5 minutes. Following brief centrifugation, samples were loaded onto a 4% polyacrylamide stacking gel (pH 6.8) and resolved on a 10-15% acrylamide gel (pH 8.8) alongside molecular weight markers (New England Biolabs). 1.5 mm gels were cast and run using a Mini PROTEAN III gel system (Biorad). Gels were run in SDS-PAGE running buffer at 30-40 mA per gel until good separation was achieved. Gels were either analysed by Coomassie staining followed by de-stain or were used in Western blots.

2.2.11 Western and dot blot

For western blotting, electrophoresis resolved polypeptides were transferred to HybondC+ nitrocellulose (Amersham biosciences) membranes in transfer buffer by applying 100V for 1 hour. In dot blotting, samples were dotted directly onto HybondC+ nitrocellulose membranes applied through a dot blot apparatus via suction. Membranes were blocked with 5% milk, 0.1% Tween-20 and PBS for 1 hour with agitation, before addition of the appropriate primary antibody diluted to 1µg/ml in 5% blocking buffer. The membrane was incubated for 1 hour with agitation and then washed 3 times with PBS and 0.1% Tween-20. The appropriate secondary horse radish peroxidase (HRP) conjugated antibody was added for 1 hour and incubation was carried out at room temperature. Following subsequent washing, antibodies were detected with combined ECL I and II as per the manufacturer's instructions (GE Healthcare). Emitted light was detected on X-ray film (GE Healthcare) following incubation of varying times ranging from 10 seconds to 2 minutes.

2.2.12 Protein Expression in *E. coli*

Sequence-verified constructs were transformed into the desired expression strain of *E. coli*. Starter cultures of $1/50^{th}$ of the final culture volume were inoculated with competent *E. coli* expressing the protein of interest with appropriate antibiotics and grown overnight at 37°C with shaking. Cultures were then used to inoculate the

desired volume of 2YT media containing the same antibiotics and grown at 37°C until an optical density of 0.6-0.8 at 600 nm wavelength was reached. Expression was induced with addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at the desired final concentration at 16°C, 18°C or 25°C either overnight or for 65 hours depending on the expressed recombinant protein.

Cells were centrifuged at 5,000 g for 10 minutes to pellet the bacteria. The supernatant broth was discarded, and cells re-suspended in bacterial lysis buffer in the presence of 0.35mg/ml lysozyme and 1µg/ml DNaseI. Cells were sonicated on ice over 10 cycles of 20 seconds on 40 seconds off. Resultant lysate was centrifuged at 20,000 g for 50 minutes to separate the soluble and insoluble fractions. The soluble fractions containing the expressed recombinant proteins were filtered through 0.45 nm syringe filters prior to purification by Ni^{2+} affinity chromatography.

2.2.13 Protein Purification

Recombinant proteins were purified from filtered bacterial soluble fractions by Ni^{2+}-affinity chromatography with a GE Healthcare HisTrap column, as all protein constructs had N-terminal His-tags. The column was prepared with 5 column volumes of equilibration buffer containing 20mM Tris-HCL pH7.4, 500mM NaCl, 5% glycerol and 20mM imidazole. The filtered protein sample was equilibrated with the same buffer containing 500mM imidazole to a final concentration of 20mM imidazole and slowly passed through the column. The protein of interest was eluted in increasing imidazole concentration. The flow through was collected and stored at 4°C, and 5 column volume samples of at 20mM, 50mM, 100mM and 500mM imidazole elutions were collected.

For more stringent Ni^{2+}-affinity chromatography, the column was attached to an ÄKTAprime chromatography system and equilibrated as above. The HisTrap column was extensively washed with 20mM imidazole buffer until the UV trace returned to the baseline, indicating loosely bound protein had been eluted. This was repeated with 50mM imidazole buffer. The recombinant protein was eluted with a gradient of 500mM imidazole buffer, starting at 80mM imidazole.

Ni^{2+}-affinity purified proteins were further purified by size exclusion chromatography (SEC). Proteins were passed through a GE Healthcare Superdex 75 or Superdex 200 column depending on the size of the protein, attached to an ÄKTAprime. Prior to use, the columns were washed through with 1 column volume of de-ionized, filtered and degassed water then equilibrated in SEC buffer (Section 2.1.1). Before purification, samples were filtered and concentrated to a suitable volume.

2.2.14 Ion Exchange Chromatography

If contaminant proteins remained after initial purification steps, samples were further purified using a HiTrap Q ion exchange column (GE Healthcare). The column was equilibrated with 50mM Tris pH 7.4. The same buffer was used to dilute the protein sample to reduce salt concentration, before protein was applied to the column. Protein was eluted using a 20 ml gradient of Tris pH 7.4 1 M NaCl.

2.2.15 Affimer Protein Preparation

Affimer were generated against the GAIN domain of adhesion GPCRs via phage display screening by the Tomlinson Group, University of Leeds, as described by Tiede, C. et al. 2014 (156). Those selected were amplified by PCR and digested and ligated as described in sections 2.2.2, 2.2.6 and 2.2.7 respectively. Protein expression was carried out as described in section 2.2.12 unless stated otherwise. Cells were grown for 5 hours at 25°C following IPTG induction.

2.2.16 Biotinylation

Affimers engineered to contain an additional single free cysteine residue were dialysed into PBS and then incubated with immobilised TCEP disulphide reducing gel according to the manufacturer's protocol (Thermo Scientific™). Immediately after, proteins were incubated with EZ-Link Maleimide-PEG2-Biotin (Thermo Scientific™) following the same protocol. Biotinylation was confirmed by mass spectrometry.

2.2.17 Dialysis

Proteins dialysed were contained in 10 kDa cut-off dialysis membrane (Pierce) and suspended in 500x volume of the desired buffer with stirring for 5 hours at 4°C. Dialysis was then repeated using fresh dialysis buffer.

2.2.18 Mass Spectrometry

Non-denaturing mass spectrometry was performed in-house by the Mass Spectrometry Facility of the Faculty of Biological Sciences, University of Leeds.

2.3 General Methods

2.3.1 Mammalian cell culture

Cell Line	Derivation	Cell Type
HEK 293T/293TT	Human embryonic kidney	Fibroblast
Cos7	African green monkey kidney	Fibroblast
CHO	Chinese hamster ovary	Epithelial

Table 2.1 ***Cell lines used to transfect DNA*****.**

HEK and Cos7 cell lines were grown and maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) heat-inactivated foetal bovine serum (FBS), 0.5% L-Glutamine, 0.05 mg/ml streptomycin and 50 U/ml penicillin (all Sigma-Aldrich, UK). CHO cell line was cultured in Ham's F10 media (Lonza, Slough, UK) supplemented in the same way but without FBS. All cell lines were passaged once 80-90% confluent, by washing the cells in sterile PBS and dissociating by incubation with trypsin (170 U/ml)-EDTA (200 mg/ml) (Lonza, Slough, UK) for 2 to 5 minutes. Equal quantity of serum- containing DMEM was added to neutralise the trypsin before washing and re-seeding. CHO cells were centrifuged at 1000g for 5 minutes and the pellet resuspended in Ham's F10. Cell lines were frozen by suspension in FCS and 10% DMSO and placed in a Mr. Frosty^{TM} (Thermo Scientific) at -80℃ for 24 hours before transferring to liquid nitrogen for long term storage.

2.3.2 Cell Transfection

Flasks or plates were seeded 1-2 days prior to transfection with the appropriate cell volume to ensure 80% confluency at transfection. On the day of transfection, DMEM was removed and cells were gently washed with sterile PBS to ensure FBS was not present in the transfection media. Plasmids encoding the gene/s of interest were transfected into cells using Lipofectamine™ 2000 (Invitrogen), TurboFect (Thermo Scientific, UK) or JetPEI® (Polyplus Transfection SA) in Opti-MEM (Gibco, Life Technologies, UK) following manufacturer's instruction. The final flask/well volume was made up with OptiMEM.

2.3.3 Insect cell culture

Spodoptera frugiperda (Sf9) cells were grown in suspension culture in SF-900 III serum free media (Gibco) at 27°C with agitation and maintained in mid-log growth phase at 1 x 10^6 cells/ml with 95% viability.

2.3.4 Insect Cell Protein Expression

The production of recombinant bacmid and baculovirus was carried out according to the Bac-to-Bac® Baculovirus Expression System manual (Invitrogen). 6 well plates were seeded with 8 x 10^5 cells and transfected with 1µg bacmid DNA using Cellfectin™ II according to manufacturer's guidelines. 72 hours post transfection, cultures were centrifuged at 1000g for 5 minutes, the supernatant harvested and stored as viral stock in the presence of 2% FBS.

The stock was amplified to generate high titre virus and used to infect 1 x 10^6 cells/ml in 15ml cultures. 72 hours post transfection, cultures were pelleted at 1000g for 5 minutes and the soluble fraction was collected and stored at 4°C. 15µl samples of soluble protein were mixed with 10µl SDS-loading buffer and resolved on 12% SDS-PAGE gel and stained with Coomassie Brilliant Blue, followed by de-staining in 10% acetic acid and 40% methanol solution. Supernatant samples were also analysed by western blot.

In the case of insoluble protein expression, cell pellets were lysed in buffer consisting of 0.2-0.5% NP40, 20 mM Tris pH 7.4, 300 mM NaCl and supplemented with Complete EDTA-free protease inhibitors (Roche). Cell debris was removed by centrifugation for 10 minutes at 4000g, and the supernatant collected and diluted 100% in the buffer without NP40 detergent. The supernatant was filtered through 0.45 nm syringe filters prior to purification by Ni^{2+} affinity chromatography as described in section 2.2.13. Both soluble and insoluble samples were detected by western blot.

2.3.5 Determining viral titre by quantitative real-time PCR (qRT-PCR)

Quantitative real-time PCR was carried out by a RotorGen (Qiagen, Hilton, Germany) and a ΔΔCT-analysis formed from the generation of standard curves for the house keeping genes and genes of interest. RNA isolation 46 was carried out using the Qiagen RNeasy Mini Kit. cDNA was generated by a first strand cDNA synthesis kit (Thermo Fischer Scientific). Qiagen QuantiFast SYBR green PCR was used to carry out the qRT-PCR according to the manufacturer's protocol.

2.3.6 Dual Luciferase reporter assays

2 days prior to transfection, HEK293T cells were seeded in the desired number of wells in a 96-well plate, and transfected using Lipofectamine 2000 (Thermo Scientific, Waltham, MA) with 50 ng DNA total per well. The receptor of interest and luciferase reporter gene specific to the signalling pathway of interest were co-transfected with pRL-TK plasmid for normalising transfection at a ratio of 3.5:3.5:1 respectively. The empty vector pcDNA 3.1 was used as a negative control plasmid in each reporter assay. 24 hours post-transfection, cells were stimulated with 5µg/ml of Affimer or antibody specific to the receptor. Cells were stimulated with PBS for a negative control. 10% FBS was included as a crude positive control for investigation of SRE and SRF-RE, and carbachol for investigation of NFAT and NFκB signalling pathways. After 6 hours, plates were equilibrated to room temperature for 30 minutes, 0.5X Dual-Glo® luciferase reagent was added and incubated for 15 minutes. After firefly luminescence was measured, 0.5X Dual-Glo® Stop and Glo® was added to cells and incubated for a further 15 minutes, and *Renilla* luminescence was measured. Assay data was first normalised to *Renilla* luciferase activity and then to either pcDNA 3.1 or signalling deficient constructs stimulated with PBS. Luminescence was measured using a BMG LABTECH FLUOstar OPTIMA microplate reader and data collected using OPTIMA software.

2.3.7 Enzyme-Linked Immunosorbent Assay (ELISA)

Affimer specificity for the relevant GAIN domain was confirmed by ELISA. Maxisorp™ 96-well plates were coated with 200ng per well of purified CD97 GAIN domain and incubated at 4°C overnight. Plates were washed thoroughly in PBS 0.05% Tween-20, and blocked in 2% BSA, PBS 0.05% Tween-20 for 1 hour at room temperature. Affimers were serially diluted across the plate with a starting concentration of 2 µg/ml and incubated for a further hour. Streptavidin HRP conjugate (Biolegend) was added to a final concentration of 1 µg/ml, incubated for 1 hour, and plates washed thoroughly. TMB substrate was added to each well, and the reaction quenched by addition of 1M sulphuric acid, before reading at 450nm.

2.3.8 Flow cytometry

Transiently transfected HEK293T cells were washed in PBS and incubated on ice with 10 mM EDTA for subsequent removal from flasks. Cells were harvested at 600 g for 5 minutes, re-suspended in 2% FBS in PBS, and incubated on ice for a further 30 minutes. $5x10^5$ transfected and un-transfected cells were split appropriately into 50µl aliquots. Biotinylated Affimers were mixed in the appropriate tubes at 10 µg/ml

and incubated for 1 hour on ice. Following washing with 800 µl 2% BSA, cells were harvested at 600 g for 5 minutes. Washing was repeated 3 times before avidin Alexafluor-488® conjugate (Life technologies) was added to a final concentration of 1 µg/ml and incubated on ice for 30 minutes. Cells were again washed 3 times and kept in 2% BSA on ice before analysis.

Transfection efficiency was measured by addition of protein-specific antibody FITC conjugate (MEM180) to a final concentration of 1 µg/ml, incubated on ice, and subsequently washed as above. Affimers that were not biotinylated were instead assessed using α-polyhistadine (Sigma). Anti-mouse Alexafluor 488® raised in donkey (Molecular Probes) was used to detect α-polyhistadine antibody binding to Affimers.

Flow cytometry was performed on a BD LSR Fortessa machine and data collected using BD FACSDiva software. 5000–10,000 events were collected, and data was analysed with FlowJo software.

2.3.9 Fixing and staining cells for immunofluorescence

Following culture of cells on a 13mm glass cover slip in a 24 well plate cells were washed 3 times with PBS. 4% paraformaldehyde was then added to cells and incubated at room temperature for 15 minutes. Cells were washed with PBS 3 times then permeabilised with 0.1% saponin for 10 minutes. Following another wash cells were blocked with 2% BSA in PBS for 30 minutes at room temperature with agitation. Cells were incubated with 1 µg/ml of primary antibodies diluted in PBS containing 2% BSA and 0.05% saponin for 1 hour at room temperature with agitation. After 4 washes in PBS, cells were incubated in the dark with 1 µg/ml of secondary antibody diluted in PBS containing 2% BSA and 0.05% saponin for 1 hour. The cells were thoroughly washed with 4 x 10 minute washes in PBS before being mounted onto microscope slides using ProLong® Gold antifade containing 4',6-diamidino-2- phenylindole (DAPI) (Life Technologies). Once mounted, slides were kept at 4°C before imaging using an inverted LSM700 confocal microscope coupled to a LSM Image Browser.

2.4 Protein Crystallographic Trials

2.4.1 Sparse matrix screening

Sparse matrix screening of CD97 GAIN-His Tag domain was carried out using in-house screens which tested a range of PEG polymer lengths at different % v/v, and a pH range of 4.5-7 using MIB buffer (Molecular Dimensions) or citric acid. Sitting drops were set up in 96-well 3 drop Swissci plates (Molecular Dimensions) using a Formulatrix® NT8 crystallisation robot with 30 µl reservoir solution. A range of protein concentrations were tested to identify a point of supersaturation where crystal growth was possible. CD97 GAIN-His Tag was concentrated to 8 mg/ml, 6 mg/ml, 4 mg/ml or 2 mg/ml and dispensed as 0.3-0.6 µl droplets containing protein and reservoir solution at ratios of 1:1, 1:2, 2:1. Plates were sealed with Viewseal pressure adhesive clear seals (Grenier Bio-One) and incubated at 20°C.

2.4.2 Crystal optimisation

Conditions yielding any form of crystals were identified and further optimised. 24-well plates were set up with the hanging drop method on cover slips and sealed with vacuum grease, then monitored for growth using a light microscope for up to one month.

Additional sparse matrix screening was performed using the JCSG core suite screens I-IV (Qiagen) of commercially available crystallisation conditions. Sitting drops were set up in 96-well 3 drop Swissci plates using a Formulatrix® NT8 crystallisation robot. CD97 GAIN was concentrated to 4 mg/ml, and dispensed in 0.05, 0.1, or 0.2 µl drops, with reservoir solution. Plates were sealed with Viewseal pressure adhesive clear seals (Grenier Bio-One) and incubated at 20 °C. Additive screen from Hampton Research Corp. was also used during crystal optimisation trials. Drops were examined using a light microscope for crystal growth 24 hours after seeding, and then periodically for 1 month using a Rockimager® 1000 (Formulatrix®).

2.4.3 Crystal streak- and microbead seeding

To obtain large single diffracting quality crystals, streak seeding was employed to introduce pre-formed crystal nuclei into a hanging-drop to control nucleation in the aim of improving crystal growth. Poor quality crystals were aspirated from sitting-drop vapour diffusion plates and crushed into micro-seeds using either a cat whisker to produce a concentrated seed stock. The seed sock was subsequently diluted 1:5 and

1:10 in Gel filtration buffer (Section 2.1.1) and streaked using a cat whisker into drops on individual coverslips comprised of 2 µl freshly purified protein and 2 µl of well solution. Coverslips were inverted over a single well of a 24-well plate, sealed with vacuum grease, and incubated at 25 °C. Drops were monitored using a light microscope for signs of crystal growth for one month. Microseed beads (Molecular dimensions) were also used to generate a concentrated seed stock.

2.4.4 Crystal harvesting and data collection

Radiation damage of macromolecular crystals occurs when crystals are exposed to X-rays at room temperature, and subsequently multiple crystals are required to be able to collect sufficient data to generate experimental electron density maps. Therefore, the use of cryo-protectants is typically essential, with cryo-genic temperature data collection being the principle method used. Crystals are transferred to conditions usually containing the mother liquor and a cryo-protectant, prior to flash cool in liquid nitrogen. A trial and error approach can be used to determine the most suitable cryo-protectant for individual proteins. Screening was performed using CD97 GAIN crystals and transferred into mother liquor containing different cryo-protectant including; 25% (v/v) glycerol, 25% (v/v) ethylene glycol, 25% (v/v) PEG 400 and 25% (v/v) methyl-2 4-pentanediol. From the screening, 25% (v/v) glycerol provided the best conditions and was used for cryo-cooling of all collected crystals.

Frozen crystals were sent to the crystallography beamline station I02 at the Diamond Light Source for diffraction data screening and subsequent data collection at 100 K.

2.4.5 Data processing

Autoindexing, integration, scaling and reduction of CD97 GAIN diffraction images was performed using different sets of programs implemented as part of the CCP4 and Xia2 program suites; XDS, AIMLESS, and CTRUNCATE or Xia2 (166)(167)(168). Five percent of the reflections were selected at random and excluded from the refinement using the program FREERFLAG and constituted the Rfree set (166)(167). In order to determine the structure of CD97 GAIN domain, an electron density map must be generated to outline the atomic positions of the protein residues. Two sets of information are required to generate the electron density map; the experimental observed amplitude derived from the diffraction data, and the atomic phase information. The phase information can be ‘derived’ from a homologous protein structure or from experimental phases. For the CD97 GAIN domain, initial structural determination was attempted by molecular replacement (MR). MR is a phasing technique that is widely used to solve the crystal structure of the target

proteins by relying upon the known structure of a protein, referred to as the search model. The underlying principle of MR relies on having a search model that is homologous to the target, and a set of measured diffraction intensities derived from the recorded diffraction data. A search is then performed using different orientations and positions of the search model in the unknown crystal to find where the predicted diffraction base on the search model best matches the observed diffraction. The search model at this position can then be assumed to represent the target structure. The phases calculated from the search model in this best position are used with the experimentally observed amplitudes to generate the electron density map for the target crystal structure.

In the absence of a MR solution, experimental phases could be gained by either heavy atom soaks or single-wavelength anomalous diffraction. Additionally selenomethionine substitution is another common technique employed for obtaining experimental phases and involves incorporating selenomethionine into the recombinant protein for subsequent multi- and single-wavelength anomalous diffraction data collection.

2.5 Atomic Force Microscopy

2.5.1 Functionalisation of AFM surface and probe

2.5.1.1 Oxidisation of surface and silicon nitride probe

To oxidise the surfaces of the Si_3N_4 AFM probe (MLCT with reflective gold, Bruker) and surface (1 cm^2 cut from a Si_3N_4 disc, Rockwood electronic material), piranha solution was prepared on ice and surfaces were submerged in the solution for 30 minutes at room temperature. The piranha solution was removed from the surfaces by washing with ddH_2O and dried with N_2. The AFM probes were submerged in the piranha solution for 30 seconds and washed with ddH_2O then dried with N_2. The surfaces and AFM probes were then placed on a microscope slide inside a petri dish with a hole in the lid and placed under a UV lamp (UVIlite, UVItec) set to 254 nm for 30 minutes (ozone cleaning).

2.5.1.2 Aminosilanisation of silicon nitride

Following oxidisation, AFM probes and surfaces were placed into a desiccator along with 80 µl of (3-aminopropyl)triethoxysilane (APTES) and 20 µl of N,-N-diisopropylethylamine (DIPEA) held in separate 1.5 ml Eppendorf tube lids on

opposite sides of the chamber. The desiccator was evacuated using a vacuum pump and left to incubate at room temperature for 2 hours. After the incubation, the APTES and DIPEA solutions were removed and the desiccator was flooded with N_2 and left to cure for 48 hours.

2.5.1.3 Attachment of NHS-PEG24-maleimide linkers

Amino-silanised AFM probes and surfaces were immersed in 1 ml chloroform containing 20 µl of 250 mM NHS-PEG_{24}-maleimide (in DMSO) and left to incubate at room temperature for 1 hour. Both AFM probe and surface were then washed with chloroform and dried with N_2.

2.5.1.5 Attachment of HaloTag® ligand to surface

The surfaces were incubated in 1 ml 7.5 mM Thiol-PEG_4-Chloroalkane ligand (HaloTag® Thiol O4 ligand, Promega), dissolved in 1 ml 50 mM Tris pH 8 buffer for 1 hour at room temperature. The surfaces were then washed and dried with ddH_20 and N_2 respectively. Surfaces not used immediately for AFM were kept in a dark, humid chamber at 4°C for up to 1 week.

2.5.1.6 Attachment of Tris-NTA to probe

Disulphide-tris-NTA was diluted in 50 µL water to 1 mM and mixed with 2 µl EDTA (50 mM, pH 7.5), 5 µl Hepes (500 mM, pH 7.5), 2 µl TCEP hydrochloride (50 mM), 2.5 µl Hepes (500 mM, pH 9.6). The solution was pipette as droplets on to a microscope slide, the AFM probes placed within and covered with a petri dish. Probes were incubated for 1 hour at room temperature and washed in 6 droplets of 150 mM Tris pH8. 96 µl of the same buffer was mixed with 4 µl 5 mM $NiCl_2$, and the probes submerged in a droplet placed in the original tip box and covered for at least 1 hour until the AFM surface had been incubated with protein.

2.5.1.7 Protein attachment

Protein (0.2 mg/mL) was deposited over the surface and left to incubate in a covered container for 30 minutes at room temperature. Unreacted protein was then washed from the surface with the reaction buffer (25 mM Tris-HCl, 128 mM NaCl), before creating a surface tension droplet of buffer over the surface.

2.5.2 Polyprotein Force Microscopy

2.5.2.1 AFM Calibration and set up

The AFM probe, either only aminosilanised or further functionalised with Tris-NTA, was inserted into a cantilever holder and secured. The protein derivatised surface was attached to a microscope slide with double-sided tape and secured to the XY scanner with magnetic bars. A droplet of 25 mM Tris-HCl pH 7.5, 128 mM NaCl reaction buffer was applied onto the surface and was held by surface tension.

The AFM probe in the holder was mounted to the MFP-3D head and approached towards the surface until the probe was fully submerged in the buffer droplet. Using the microscope optics, the laser was positioned to the tip of cantilever D (spring constant: 30 pN/nm) and the deflection was set to zero using the photodetector (PD) disc. The cantilever was engaged using the Asylum Research software and the Z-piezo voltage maximised (+150), indicating that the Z-piezo was fully extended and had zero contact with the surface. Using the thumb wheel on the MFP-3D head (Asylum), the cantilever was approached to the surface. If the z-voltage fell, it was adjusted to 70 z-voltage, assuring that the piezo was in the middle of its z range at 7.5 μm.

Calibration of the spring constant was carried out by determining the slope of contact from a force curve to find the sensitivity of the cantilever (nm/V). Non-destructive thermal tuning was then performed to determine resonant frequency of the cantilever. A single force-extension plot was recorded with the trigger (amount of deflection the cantilever undergoes before retraction) set to 20-40 nm, which gives a quantifiable deflection slope. The slope of the contact region (inverse optical lever sensitivity (InvOLS)) was measured by a linear fit. The cantilever was withdrawn from the surface and the deflection was set to 0. A thermal tune was carried out to detect the natural thermal fluctuation of the cantilever by performing ~50 frequency sweeps (0-1 MHz). The natural frequency (first major resonance peak) was selected and the area of thermal fluctuation (*P*) was used to find the spring constant (*K*).

$$K = \frac{K_B T}{P}$$

The spring constant was always within error of the manufacturer's guidelines before data collection. Hooke's Law (Equation 2.13) allows the calculation of the force applied (*F*) to the cantilever with a known spring constant by the deflection (or extension (x~~x~~)).

$$F = kx$$

2.5.2.2 Data collection

The trigger point was set to 10 nm, force distance set to 1 µm, approach velocity set to 2 μms^{-1} during data collection and kept constant using the various retraction velocities and sample rate (10 kHz per µm velocity). The buffer on the sample surface was replenished every hour to prevent evaporation, which would change the NaCl concentration and dry the sample. Force maps of 20 μm^2 with 250-2000 force-extensions were taken depending on protein distribution and events seen in that area, and to ensure a good coverage of the surface. Between force maps the sample was repositioned using the XY scanner. The retraction velocities used were 700 and 1000 nms^{-1}.

2.5.2.3 Data processing

All collected data was analysed using IGOR pro 6.32A with an Asylum Research extension (MFP3DXop v30). The hard contact (0 nm) and baseline (0 pN) of all the force-extension retraction traces were manually set, by taking a section of the hard contact or the retraction with no events and setting the zero to the centre of the data. The worm-like chain (WLC) model that represents the semi-flexible behaviour of a polymer when stretched by force was used to find the contour length (Lc) and extension (Ex) using a fixed persistence length of 0.4 nm (p). The WLC model was fit to unbinding events that displayed 'worm-like' curves using the Asylum software by manually inserting locks at the apex and the base of the curve and recording the Lc, Fr and Ex.

Force-extension profiles were binned for analysis if: 1) the data fitted to the WLC model i.e. single molecule events displayed WLC like behaviour where the force-distance profile is not linear, and 2) the expected unfolding profile was present for the number I_{27} domains in the polyprotein construct, at the very least. Single Gaussian distributions were fit to unbinding force and contour length histograms in order to determine the most probable force and contour length at unbinding.

Chapter 3. Structural determination of EGF-TM7 GAIN domains

3.1 Introduction

Adhesion GPCRs constitute the second largest sub-family of the GPCR superfamily yet remain the most poorly understood. A lack of known signalling characteristics, binding partners and biological function for many members makes them a challenging research target. In recent years, the study of the extracellular GAIN domain has been a major focus in the field as increasing evidence suggests it plays a major role in the activation of these receptors (27). Despite the *Stachel,* or tethered agonist region, being defined as a small peptide sequence immediately C-terminal to the cleavage site, it is unclear exactly how this region is exposed to allow it to interact with the transmembrane alpha helices to induce receptor activation. The N-terminal portion of the GAIN domain seemingly acts to mask to this agonistic peptide, preventing constitutive receptor activity. By gaining a deeper understanding of the structure of this conserved protein domain, key residues responsible for maintaining the non-covalent interactions that hold the EC domain and IC domain together, and for mediating ligand binding can be identified, and thus provide important information about the mode of activation of adhesion GPCRs.

During the course of this research, the crystal structures of adhesion GPCRs rat Latrophilin 1 (Lat1/Lphn1) (PDB, 4DLQ), GPR56 (PDB, 5KVN) and Brain Angiogenesis Inhibitor 3 (BAI3) (PDB, 4DLO) have been determined, providing insights as to how the GAIN domain interacts with the 7-transmembrane signalling domain. The crystal structure of an example GAIN domain is illustrated in Figure 1.6C and described in Section 1.2.2. Interestingly, subsequent sequence alignment of all known GAIN domains indicates that the EGF-TM7 subgroup of adhesion GPCRs differ somewhat from the wider family. They lack at least 3 N-terminal alpha helices within the GAIN domain, but still have all 13 conserved beta strands. This significant difference may be key to understanding how these mostly myeloid restricted receptors influence responses in the immune system. Gaining sound structural understanding of these domains will provide insight as to the importance of the tertiary protein structure in relation to the function of the receptor subgroup. Furthermore, although a handful of adhesion GPCRs have known binding partners, there are very few examples of ligand-induced activation.

Whilst an increasing number of protein structures have been determined by electron microscopy (EM) and nuclear magnetic resonance (NMR), X-ray crystallography was selected in this instance for an example EGF-TM7 GAIN domain. The domain alone is too small for investigation by EM, and due to notorious protein expression difficulties in deuterated media, NMR was also deemed unsuitable. This chapter

therefore primarily aimed to report the work untaken to solve the crystal structure of CD97 GAIN domain, a model EGF-TM7 receptor, both in the presence and absence of the extracellular adhesive EGF repeats. This was done with a view to further our understanding of ligand selectivity and the potential conformational rearrangements required for receptor activation.

3.1.1 Structural determination of proteins by crystallography

X-ray crystallography is a staple technique in determining protein structure, having been used to obtain the macromolecular structure of approximately 80% of entries in the Protein Databank (PDB). Despite this, structural determination is not without its difficulties which range from the fundamental, such as proteins having inherently flexible regions rather than being rigid and ordered, to the practical, such as crystal growth and subsequent diffraction data collection (RCSB PDB, 2018). Regardless of its limitations, X-ray crystallography remains the preferred option for high-resolution structure solution of most macromolecular molecules with the ability to achieve atomic resolutions of 1-3Å.

Crystal production is required as they contain multiple protein copies in an ordered lattice that is utilised to amplify the signal of diffracted X-rays to enable their detection. Well-ordered and diffracting crystals are necessary for recording experimental diffraction spots from these crystals. This ultimately allows the generation of an electron density map into which the atomic model of the protein can be fitted. However, even with automated robotic systems for establishing crystal screens and making buffers to increase the reproducibility of crystallisation conditions, crystal growth remains one of trial and error and a constant hindrance in structural determination.

Obtaining homogenous, supersaturated protein solution without inducing precipitation or aggregation is crucial for producing suitable crystals. Crystal nucleation can be achieved by the inclusion of precipitation agents such as salt or polymers, for example polyethylene glycol (PEG), and by manipulating parameters such as incubation temperature and solution pH (170). Vapour diffusion and batch crystallisation are techniques that can be used to achieve crystal formation. The former involves sealing a drop of precipitant, called the mother liquor, and a drop of homogenous protein in a chamber containing a much larger reservoir of the mother liquor. This facilitates the diffusion of water molecules until the osmolarity of both solutions are the same and the nucleation zone is reached over time. Batch

crystallisation brings the protein immediately to nucleation by directly mixing with the precipitant (171). Vapour diffusion is performed by “sitting drop” or “hanging drop” as shown in Figure 3.1. The resultant dehydration of the drop causes a slow increase in the concentration of both the protein and precipitant solution. If the mother liquor contains the appropriate reagents to produce a favourable chemical environment to promote lattice formation, crystals will grow in the protein drop. Extensive trials are often required to determine the optimal conditions for good crystal growth, taking anywhere from hours and months to occur.

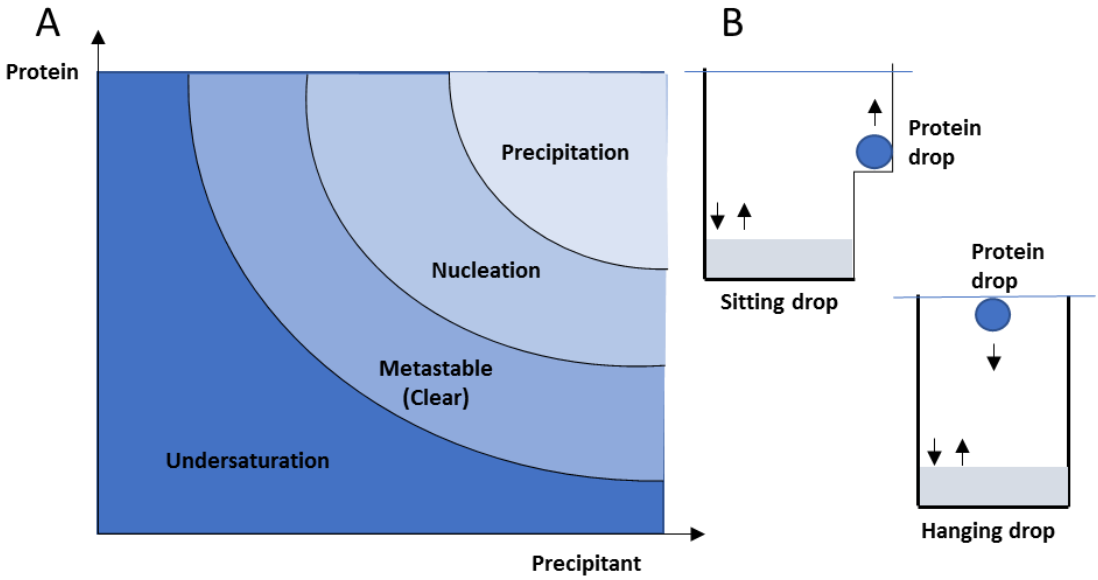


Figure 3.1. ***Principle and techniques for crystal growth.*** **(A)** Nucleation describes the initial steps in crystal formation, whereby protein molecules arrange in a pattern characteristic of crystalline solids, which forms a site for deposition of other molecules as the crystal forms. The clear, or metastable, zone describes conditions where crystal growth is supported but nuclei cannot form. **(B)** Methods used for crystal formation; arrows represent vapour diffusion. A drop of pure protein mixed with a drop of mother liquor containing precipitant is placed in a chamber next to or hanging above a larger reservoir of the mother liquor. The chamber is sealed, and diffusion occurs changing osmolarity of the drop, increasing the concentration of precipitant to enable nucleation formation and subsequent crystal growth.

The composition of the mother liquor is a crucial determinant in crystal formation and growth. Small molecule additives are often included to increase macromolecular heterogeneity, by perturbing their environment and providing additional interactions in the lattice to enhance crystal nucleation and thus growth (172). Additives may influence biological activity and induce conformational changes to a more favourable

or stable arrangement for nucleation, act as chemical protectants, or be solubilising agents or detergents. Inclusion of these molecules in the mother liquor have proven to be vital in the crystallisation of difficult target proteins, increasing their quality and reproducibility, and their diffraction characteristics (173)(174).

Various methods can be employed to aid nucleation. Streak seeding is a technique used to introduce existing crystal nuclei into a fresh protein:mother liquor drop to control nucleation, encouraging formation of larger, singular protein crystals within the drop, and increasing the likelihood that crystals will grow in the metastable zone. In this zone, crystals cannot spontaneously nucleate, as this requires a higher concentration of precipitant than crystal growth does. By providing a seed or seed stock, crystal growth is encouraged in this zone in a controlled and reproducible manner. Introduction of such nuclei is achieved by mechanical breakage of existing crystals, using either animal whisker or hairs, or by vortexing with seed-beads (175).

Due to the range of difficulties often encountered during crystallography experiments, various mechanisms for stabilising proteins to aid the growth of well-diffracting crystals have been developed, such as protein engineering or complexing with endogenous binding partners. However, this proves problematic when the protein of interest has no known ligands, or if engineering modifies the target in such a way that alters its natural conformation and therefore activity. One way to circumvent these issues is by the use of small protein scaffolds (Section 1.4) that act as crystallisation “helpers” or chaperones. DARPins and nanobodies, for example, have been successfully used to aid structural determination of challenging targets (176)(177), whilst the adhesion GPR56 extracellular domain, including the GAIN domain, was crystallised in complex with a monobody (178). These synthetic binding proteins are constructed using existing protein domains, such as ankyrin and fibronectin type III for DARPins and monobodies respectively, as molecular scaffolds. They are increasingly utilised as an alternative technology to antibodies. Affimers are a similar technology proven to be versatile molecular tools in both research and therapeutics. These highly stable proteins are generated by phage display and consist of a cystatin-based scaffold containing two randomised loop regions of nine residues to facilitate highly specific target recognition (156). Affimers are currently the most stable protein scaffold, and due to their comprehensive use at the University of Leeds, were exploited within this study.

3.2 CD97 GAIN domain expression trials

3.2.1 Baculovirus expression system

3.2.1.1 Recombinant bacmid production

The previously crystallised GAIN domains have been expressed to a high level in insect Sf9 cells, based on infection with the lytic baculovirus as this system facilitates proper post-translational processing and increases the chance of gaining soluble, correctly folded protein. The Bac-to-Bac® Expression System, outlined in Figure 3.2, rapidly generates recombinant baculovirus particles that can be used to infect insect cells and enable efficient protein expression. The crystal structures of three different adhesion GPCR GAIN domains have recently been determined having expressed the proteins in a similar manner. Lpn1 and BAI3 were solved in the absence of almost their entire extracellular adhesive domains, but inclusive of the hormone binding domain (26) whilst GPR56 was solved with its entire extracellular domain, facilitating characterisation of a previously unknown domain (178). Previous attempts at expressing EGF-TM7 adhesion GPCRs with intact EGF repeats using a bacterial expression system by the Stacey group, Leeds, have produced low yield or insoluble protein. Therefore, with the successful production of other GPCRs and indeed adhesion GPCRs in insect systems, this approach was exploited.

CD97 GAIN domain EGF-1,2,5 cDNA was cloned into the pFastBac™HT-A vector, which includes a polyhedron promoter to facilitate high yield protein expression and an N-terminal 6xHis tag for subsequent protein purification. The recombinant plasmid was transformed into DH5α *E. coli,* analysed by colony PCR and verified by Sanger sequencing, before being transformed into competent DH10Bac™ *E. coli* containing bacmid DNA, a baculovirus shuttle vector encoding a target site for bacterial transposition and the Tn7 transposon, and a helper plasmid. Successful recombinant bacmids containing the gene of interest were identified by blue-white selection and further colony PCR.

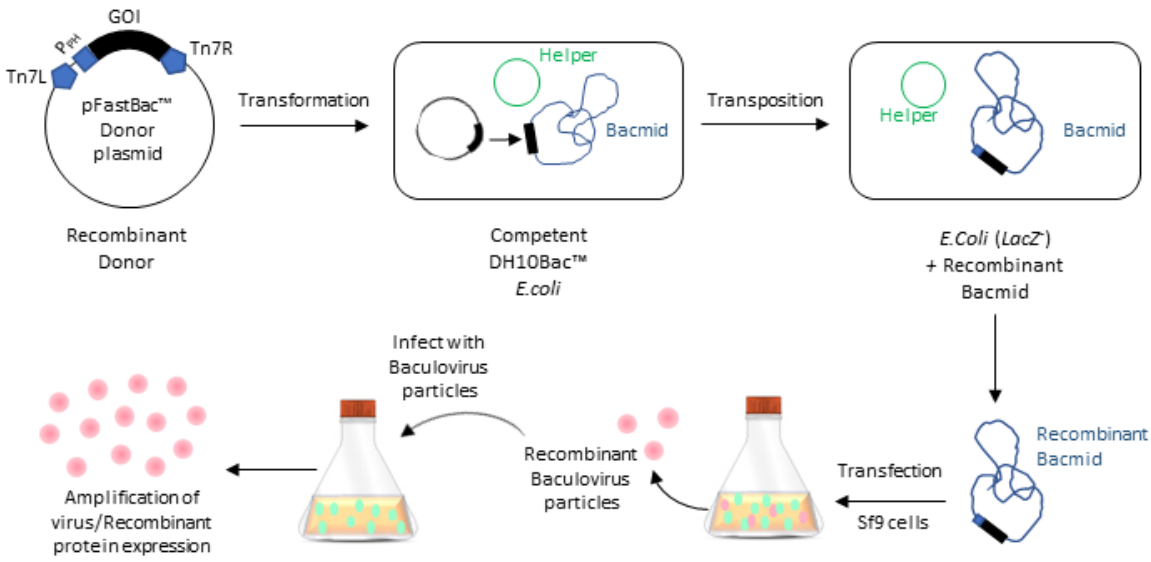


Figure 3.2. *Overview of the baculovirus expression system*. The gene encoding the protein of interest is cloned into the pFastBac™ donor plasmid by conventional PCR and transformed into DH10Bac™ *E. coli* which contains a baculovirus shuttle vector and a helper plasmid. Transposition occurs between a mini-Tn7 element on the pFastBac™ vector and a mini-*att*Tn7 target site on the bacmid to generate a recombinant bacmid, facilitated by proteins encoded by the helper plasmid. The purified bacmid is then transfected into Sf9 cells to generate infectious baculovirus particles that are subsequently used to re-infect a new plate of Sf9s and amplify the virus titre. Once the desired titre has been achieved, the viral stock is used to infect cells again and facilitate recombinant protein expression. The protein can then be harvested and purified.

3.2.1.2 Expression of recombinant CD97 GAIN EGF-1,2,5 bacmid

To establish bacmid transfection and viral amplification, the baculovirus system was optimised using a GFP-expressing bacmid alongside CD97 GAIN EGF-1,2,5, previously generated by the Stacey group, Leeds. Purified recombinant bacmid was used for immediate transfection of Sf9 cells. Cells were confirmed to be in logarithmic phase at a density of $1.5x10^6$–$2.5x10^6$ cells/mL and a 6-well plate was seeded with $8x10^5$ cells/well and left to adhere for 30 minutes. Transfection was carried out using Cellfectin® II according to the manufacturer's protocol (Invitrogen). Growth medium was changed 3-5 hours post-transfection, and the cells were incubated for 72 hours and monitored for evidence of viral infection (Figure 3.3, Table 3.1). The supernatant containing budded virus (P1 stock) was harvested and stored at 4°C with 2% FBS, protected from light.

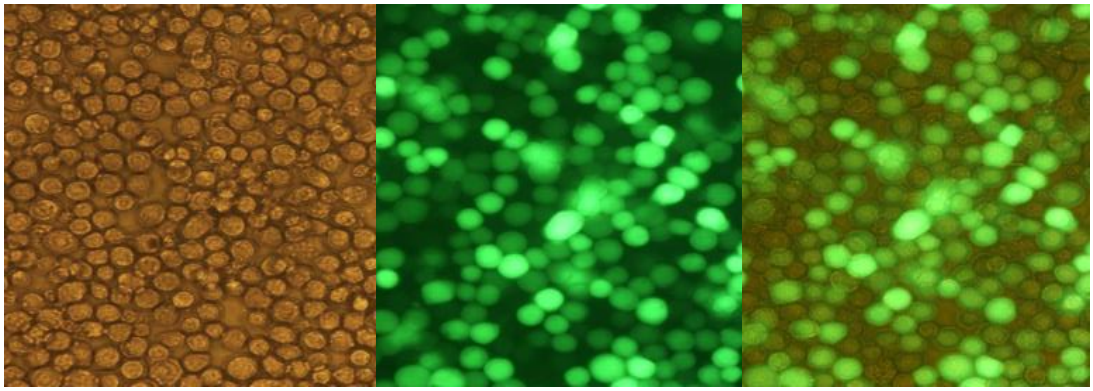


Figure 3.3. ***Baculoviral expression of GFP recombinant bacmid*** **Sf9 cells 72 hours post-transfection with recombinant bacmid.** 1 µg of purified bacmid DNA was diluted in Grace's growth medium and added dropwise to cells. Supernatant containing baculovirus particles was harvested for further amplification of viral titre.

Sign of Infection	Characteristic
Early stage	25-50% increase in cell size, and nuclei may appear to fill the cell.
Late stage	Cells may stop growing in comparison to uninfected cells and detach easily from the plate. Cells become granular in appearance.
Very late stage	Cells appear lysed.

Table 3.1. ***Characteristics of virally infected cells*****. Adapted from Bac-2-Bac manual, Invitrogen.**

P1 viral stocks are typically of low-titre and small scale. Therefore, to amplify both the volume and viral titre and thus generate a P2 stock suitable for protein expression trials, P1 was used to infect Sf9s at 95% viability at a multiplicity of infection (MOI) of 0.1. Before performing a plaque assay, it can be assumed that viral titre of the P1 stock will be between $1x10^6$ and $1x10^7$ pfu/ml (Invitrogen manual). 15 ml suspension cultures were incubated for 72 hours and subsequently centrifuged for 5 minutes at 300g. The supernatant was harvested and stored as previously stated as the P2 viral stock.

To calculate the viral titre of the P1 and P2 stocks of CD97 GAIN EGF-1,2,5 and GFP, plaque assays were carried out. Infected cells were plated on plaquing medium to immobilise cells and incubated for 10 days before analysis. Example titres for CD97 GAIN EGF-1,2,5 and GFP are displayed in Table 3.2. However, as plaque assays are a laborious process a Real-Time SYBR® green quantitative PCR (qPCR) based calculation of viral titre was established (179). As each virus particle contains one copy of the recombinant gene, determining copy number provides a value for virus titre and thus plaque forming units. qPCR has become a reliable, gold standard tool for determining viral titre in molecular diagnostics (180), using serially diluted standards of known concentrations to generate a standard curve to calculate viral load of the experimental sample (181). Calculated titres determined by qPCR are shown in Table 3.2.

Recombinant Bacmid	Plaque Assay	qPCR
GFP	$2.4x10^6$ pfu/ml	$2.5x10^6$ pfu/ml
CD97 GAIN EGF-1,2,5	$2.7x10^6$ pfu/ml	$2.6x10^6$ pfu/ml

Table 3.2. Viral titre of recombinant bacmids calculated by plaque assay and qPCR.

Protein expression trials were performed using P2 stocks to determine both the optimum MOI and time point for harvesting high yield protein. 15 ml cultures of Sf9 cells were infected at logarithmic phase at a density of $1x10^6$/ml with P2 stock at a MOI of 1, 3, 5 or 10. Infections were done in triplicate to be able to harvest at 48, 72 and 96 hours post transfection, as signs of viral infection were previously evident from 48 hours onwards. Cultures were pelleted for 5 minutes at 1000g, and lysed in 20 mM Tris pH8, 200mM NaCl and 1X protease inhibitors without EDTA (Roche), in the absence of detergent to prevent disruption of the α- and β- subunits of the GAIN domain (Figure 1.6). A sample equivalent to 100 µl of cells was resolved on 12% SDS-PAGE gel and further analysed by western blot (Figure 3.4).

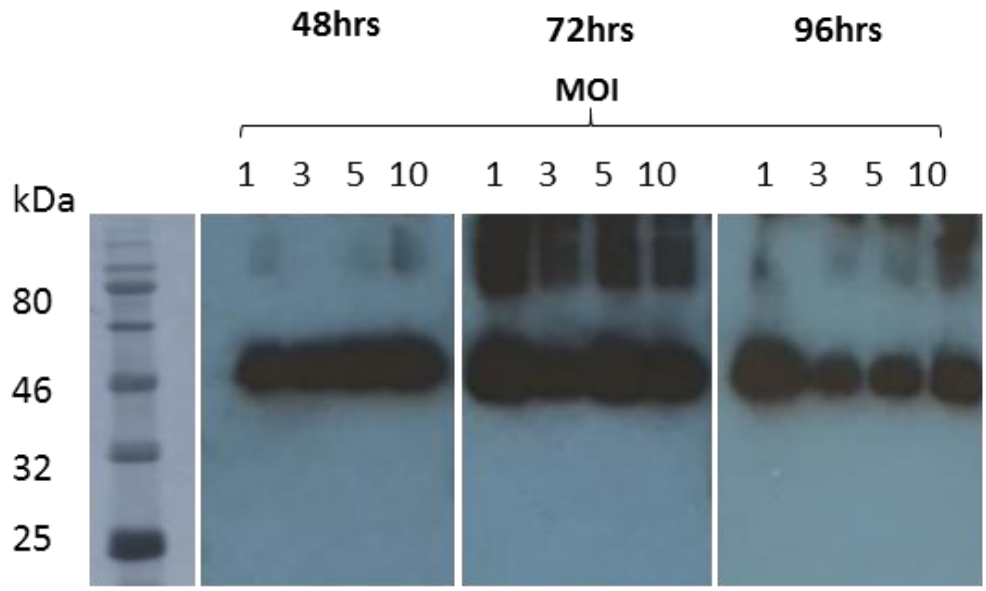


Figure 3.4. ***Time course infection of Sf9s expressing CD97-GAIN EGF-1,2,5.*** Western blot analysis was performed for whole cell lysates at 48, 72 and 96 hours infection with CD97-GAIN EGF-1,2,5 at various MOIs. The membrane was probed with anti-His HRP. Molecular weight of CD97 GAIN EGF-1,2,5; 52kDa**.**

Western blot analysis demonstrated little difference in protein expression between each MOI tested. Overall, the time course infection indicated the optimal time point to harvest protein was 72 hours post infection, although higher levels of contaminants were also visible at this time. Subsequent infection of Sf9 cells was carried out with a MOI of 1 and cultures harvested 72 hours post-infection.

CD97 GAIN EGF-1,2,5 expression was scaled up to 50 ml cultures, and cells were subject to multiple freeze thaw cycles and sonication to ensure optimal cell lysis post protein expression. The protein was purified from whole cell lysates by Ni^{2+} affinity chromatography, as outlined in section 2.2.13. Figure 3.5 clearly indicates minimal protein purification, whilst both the pre-purification sample and insoluble fraction contain an abundance of protein demonstrating the majority of the recombinant protein to be insoluble. Western blotting confirmed CD97 GAIN EGF-1,2,5 to be within the insoluble fraction, suggestive of expression within inclusion bodies or poor cell lysis. The insoluble fraction was subject to further harsh lysis with 0.5% non-denaturing detergent IGEPAL-630, 1X protease inhibitors and 150mM NaCl, yet CD97 GAIN remained insoluble.

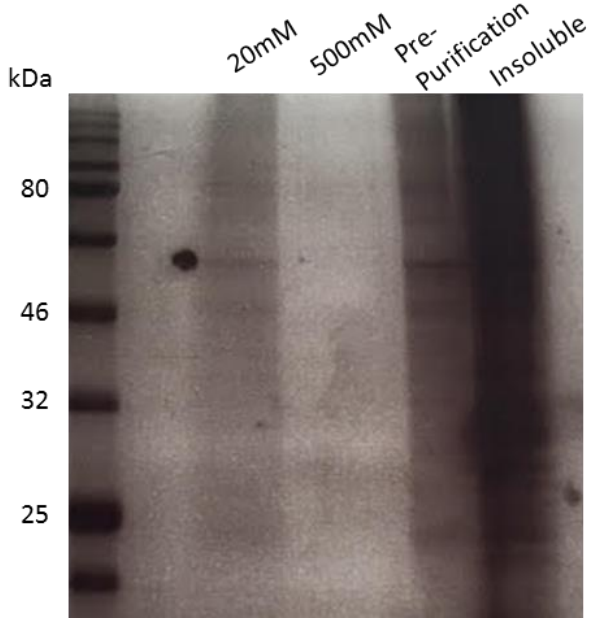


Figure 3.5 *Ni^{2+} affinity purification of CD97 GAIN EGF-1,2,5 from Sf9 whole cell lysate.* Purification fractions were resolved on 12% SDS-PAGE post purification and stained with Coomassie Blue. No soluble protein was purified.

3.2.1.3 Optimisation of CD97 GAIN EGF-1,2,5 expression

Although insect cell systems are capable of post-translational modifications like those of mammalian expression systems, the lack of soluble CD97 GAIN EGF-1,2,5 suggests improper protein processing. To overcome this and allow secretion of the CD97 GAIN EGF-1,2,5, constructs possessing the honeybee melittin (HBM) leader peptide upstream of the N-terminal 6xHis tag were generated by Gibson assembly (182). In addition, the GAIN domains of EGF-TM7 member EMR2, and myeloid expressed GPR114 were cloned into the pFastBac vector, and subsequent recombinant bacmid generation was carried out as previously stated. Truncated GAIN domains constructs with the absence of potential flexible N and C-terminal regions were also generated (Table 3.3). Furthermore, as the proteins are smaller, they may be likely to be expressed at higher levels and be more amenable to crystallisation due to reduced flexibility.

	pFastBac Constructs	Recombinant DH10Bac
CD97-EGF 1,2,5 GAIN	✓	✓
CD97 GAIN short	✓	✓
CD97 GAIN long	✓	✓
EMR2 GAIN short	✓	✓
EMR2 GAIN long	✓	✓
EMR3 GAIN	✓	✓
GPR114	✓	✓

Table 3.3. GAIN domains of different adhesion GPCRs cloned into pFastBac vector, inclusive of HBM leader peptide, and subsequently transformed into competent DH10Bac *E. coli*.

Initially, CD97 GAIN EGF-1,2,5 and GPR114 were selected for expression trials. Fresh viral stocks were generated as stated previously, amplified to a P2 stock and the viral titre determined by qPCR. 50 ml cultures of Sf9s were infected with recombinant baculovirus at an MOI of 1 and incubated for 72-96 hours, depending on signs of viral infection and percentage of Sf9 cell death. The expression of secreted and cellular recombinant protein was analysed via western blotting. Proteins were detected by anti-His-HRP. Despite the inclusion of the HBM leader peptide for secretion of target protein into the supernatant, both CD97 and GPR114 GAIN domains were not found to be secreted. Figure 3.6 demonstrates an abundance of CD97 GAIN domain within the cell pellet, whilst a much lower level of GPR114 expression was detected.

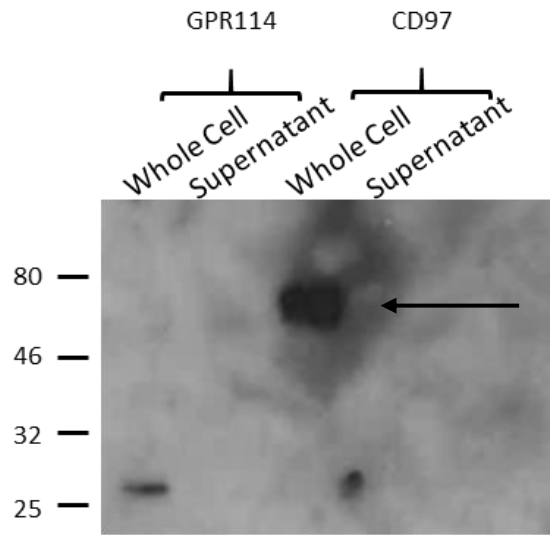


Figure 3.6. ***Baculorviral expression of secreted GPR114 GAIN and CD97 GAIN EGF-1,2,5 expression constructs.*** Western blot analysis of Sf9 cell conditioned media and cell pellet. Membrane probed with anti-His HRP. The expected molecular weight of CD97 GAIN EGF-1,2,5 and GPR114 were 52kDa and 29kDa, respectively.

The whole cell fractions of each CD97 GAIN EGF-1,2,5 and GPR114 GAIN were subsequently subject to multiple freeze-thaw cycles and further lysed by resuspension in 20mM Tris pH8, 20mM NaCl, 1X protease inhibitors, and either 0.2% or 0.5% NP-40 detergent, to liberate the protein from the cell pellets. Samples were incubated on ice for 1 hour before being pelleted at 5000g for 10 minutes. The resulting supernatant was passed through a 0.22 µm pore filter, and protein was purified by Ni^{2+} affinity chromatography. Protein in purification fractions was detected by western blot. GPR114 remained in the insoluble fraction (data not shown). CD97 however, was partially purified after cell pellet treatment with detergent. Figure 3.7 demonstrates the majority of CD97 GAIN EGF-1,2,5 remained insoluble, or may have been degraded after treatment with 0.5% NP40, with a very small smear visible in the 100mM imidazole fraction. However, a distinct band is evident in the same fraction after treatment with 0.2% NP40. Although a large proportion appears to still be insoluble, CD97 GAIN EGF-1,2,5 was present in all fractions after purification at the correct size of 52kDa. The presence of the GAIN domain in both the flow through and 20mM imidazole wash fractions suggests potentially weak binding of the 6xHis tag to the Ni^{2+} column resin therefore purification required further optimisation to maximise yield. However, subsequent trials were hindered by continued variability in the level of protein expression and stability during purification, regardless of the MOI,

incubation time and cell lysis technique, resulting in minimal soluble protein production. Therefore, a different strategy was used to express CD97 GAIN domain.

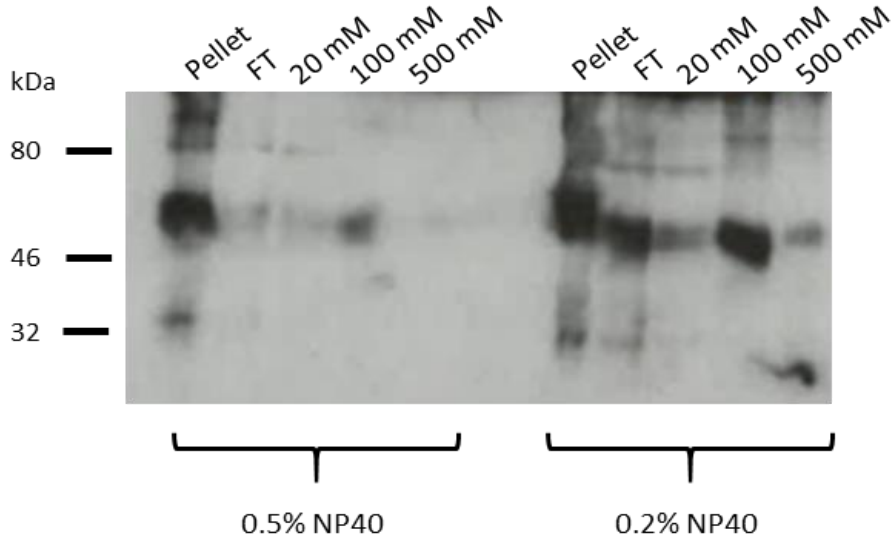


Figure 3.7. ***Solubilisation of CD97 GAIN EGF-1,2,5 from Sf9 cell pellets.*** Western blot of Ni^{2+} affinity chromatography purification of CD97 GAIN EGF-1,2,5 from Sf9 cells, after cell lysis with NP40 buffer. Treatment with both 0.5% and 0.2% NP40 was unable to fully release recombinant protein from cell pellets evident from the large band in both pellet fractions. However, CD97 GAIN was strongly detected in the 100mM imidazole fraction after 0.2% NP40 treatment, suggesting successful release from cell pellet.

3.2.2 Bacterial Expression System

Escherichia coli is the most commonly employed prokaryotic expression host used for production of recombinant protein. The system is simplistic, rapid and inexpensive (183), and is well characterised and understood. An extensive number of molecular tools are available for use in bacterial expression, such as expression plasmids and engineered mutant strains, as well as various cultivation strategies for system optimisation (184). It is often the chosen method for protein production when studying molecular structure, investigating protein-protein interactions, and for interrogating biological activity of in functional assays due to the scalable nature of the system allowing production of milligram quantities of homogenous pure protein per litre.

3.2.2.1 Expression trial of CD97 GAIN domain in *E. coli*

The GAIN domain of CD97 was cloned into the pET-28a vector, incorporating an N-terminal 8xHis tag (His-CD97 GAIN). Utilising the pET system, expression of the protein of interest is driven by bacteriophage T7 RNA polymerase which binds to the specific promoter region upstream of the target CD97-GAIN ORF for expression. pET systems can therefore be introduced into strains of *E. coli* which contain the DE3 lysogen encoding T7 RNA polymerase. Expression of target protein encoded within the pET28a vector is regulated by the *lac* operon, and expression is induced upon addition of IPTG to the growth medium upon cells reaching exponential growth. Figure 3.8 demonstrates a basic schematic of this system.

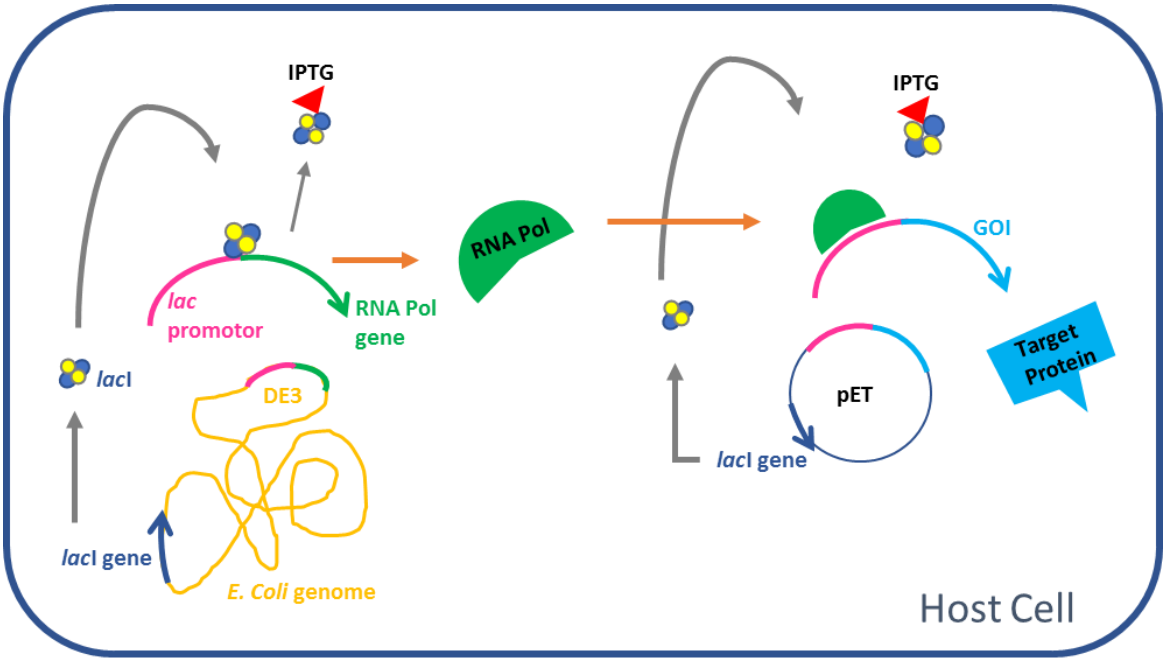


Figure 3.8 ***Basic schematic of IPTG induction of protein expression under lac operon control.*** The lac repressor, lacI, binds the lac promotor preventing gene transcription. In the presence of IPTG, a molecular mimic of the lactose metabolite allolactose, lacI is bound enabling *E. coli* polymerase to transcribe the T7 RNA polymerase gene. Similarly, T7 RNA polymerase is then free to initiate transcription of the gene of interest, resulting in expression of the target protein, but only in the presence of IPTG. In the absence of IPTG, lacI inhibits transcription of the gene of interest (GOI).

Initial protein expression was performed by transforming several competent *E. coli* expression strains with the GAIN domain-containing pET28a vector. Small scale expression trials were carried out, in which the following conditions were tested; incubation temperature post induction, IPTG concentration, growth media, and presence or absence of glucose in growth media. An overview of these conditions is shown in Table 3.4. Results indicated low level protein expression (not shown), yet subsequent preparations showed completely abolished expression, and little bacterial growth.

		SHuffle® T7	***SHuffle® Xpress T7***	***BL21 Gold***	***BL21 Codon Plus***	***Rosetta pLysS***	***Arctic Express***
Temperature (°C)	**37**	✓	✓	✓	✓	✓	✓
	25	✓	✓	✓	✓	✓	✓
	16	✓	✓	✓	✓	✓	✓
IPTG (M)	**0.01**	✓	✓	✓	✓	✓	✓
	0.1	✓	✓	✓	✓	✓	✓
	0.5	✓	✓	✓	✓	✓	✓
+ Glucose	**1%**	✓	x	✓	✓	x	x
Growth media	**2YT**	✓	✓	✓	✓	✓	✓
	TB	✓	x	x	x	x	x
	SOB	✓	x	x	x	x	x

Table 3.4. *Summary of CD97 GAIN expression trials*. His-CD97 GAIN cloned into pET28a was transformed into multiple *E. coli* expression strains. Protein expression was carried out, testing the above experimental parameters. All expression strains, except Shuffle® T7, ultimately did not yield soluble protein under any condition, therefore were unable to be taken forward in expression trials.

A possible explanation for the loss of GAIN domain is a lack of regulated expression. The *lac* operon is associated with being “leaky”, meaning low-level expression of the target gene occurs before induction (185). This may result in plasmid instability or a total loss of the plasmid containing the target gene, hence loss of expression of the protein of interest. Figure 3.9 outlines the basis of leaky expression of proteins under control of the *lac* operon. If lactose is present in growth media it is converted to allolactose which binds to the *lac* repressor, releasing it from the operator and enabling RNA polymerase binding to facilitate gene transcription. However, glucose is a more favourable energy source for *E. coli* and therefore preferentially used over lactose. As a result, lactose is not metabolised to allolactose in its presence, and thus the repressor is not released from the operator to allow for gene expression.

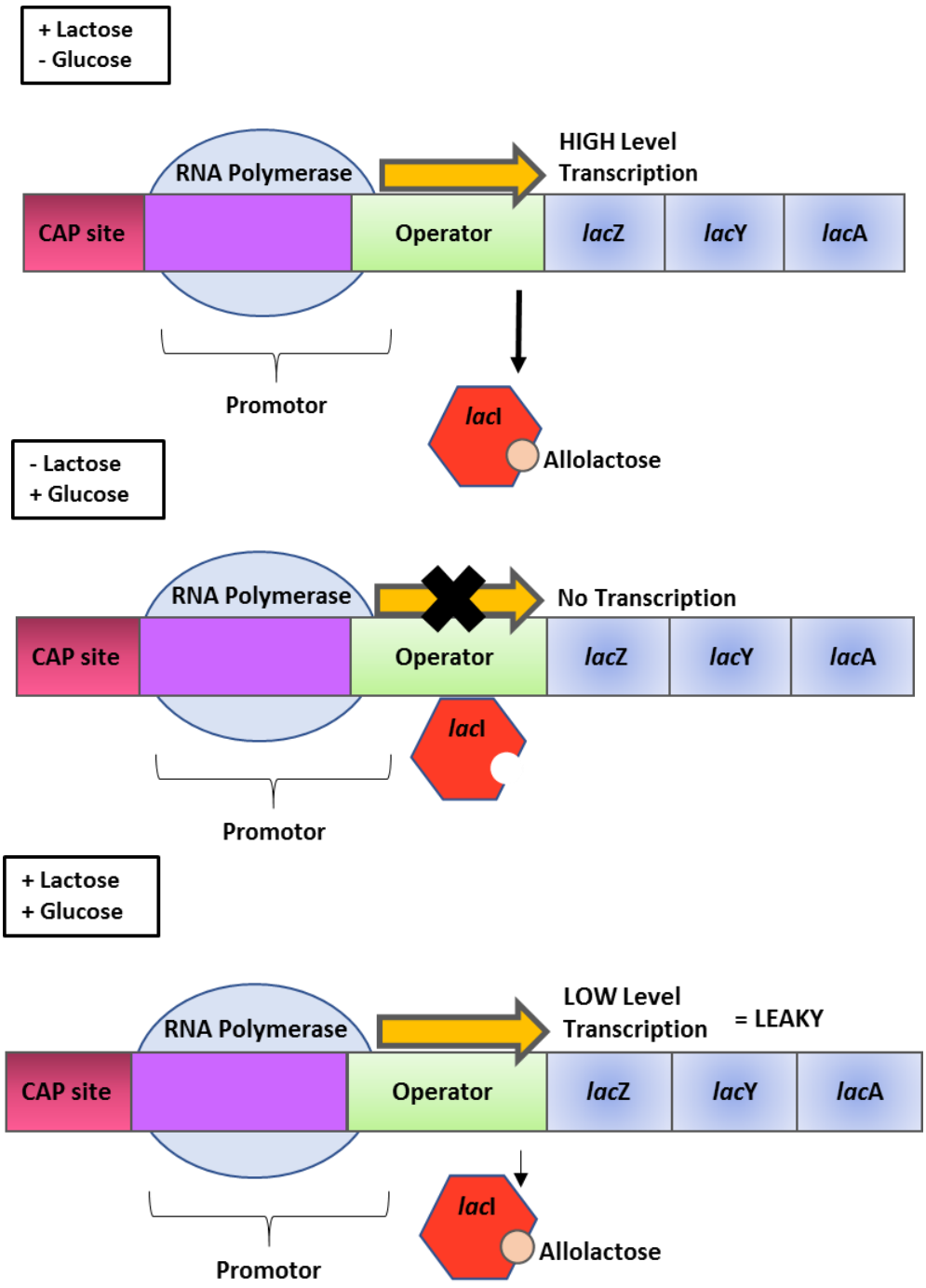


Figure 3.9. ***Basic schematic of leaky expression.*** In the presence of lactose but absence of glucose, lactose is metabolised to allolactose, which binds to the *lac* repressor, releasing it from the operator and allowing RNA polymerase to transcribe the *lac* operon. This cannot happen when there is no lactose, so the repressor remains bound, preventing transcription initiation. If both lactose and glucose are present in similar concentrations, "leaky" low level expression occurs due to small amounts of lactose being converted to allolactose. This may be overcome by vastly increasing the concentration of glucose, ensuring *E. coli* do not resort to lactose as an energy source and thus preventing production of allolactose.

To investigate whether expression of His-CD97 GAIN was leaky, two 30 ml flasks of 2YT media, one containing 1% glucose and the other without, were inoculated with transformed BL21-CodonPlus (DE3) at 1/100 of the final volume. Cultures with and without glucose, induced with different IPTG concentrations, were incubated at different temperatures overnight. To ensure cells were thoroughly lysed, pellets were resuspended in lysis buffer and incubated at room temperature for 2 hours, before being resolved on 12% SDS-PAGE gel. Expression of His-CD97 GAIN under all conditions was largely insoluble. Figure 3.10 indicates strong insoluble expression, whilst more faint bands corresponding to the expected size of His-CD97 GAIN are arguably evident in the following soluble fractions; 16 °C with glucose and all IPTG concentrations, 25 °C with glucose and 0.1 M IPTG, 25 °C without glucose and 0.5 M IPTG, and 37 °C without glucose at all IPTG concentrations. The low molecular weight band visible in all soluble fractions corresponds to lysozyme. The strong bands for His-CD97 GAIN in the insoluble fractions suggest that expression is being regulated, however proper folding and processing of His-CD97 GAIN is not occurring, resulting in formation of inclusion body aggregates. The results depicted in Figure 3.10 are also representative of protein expression in Rosetta pLysS, ArcticExpress and BL21-Gold.

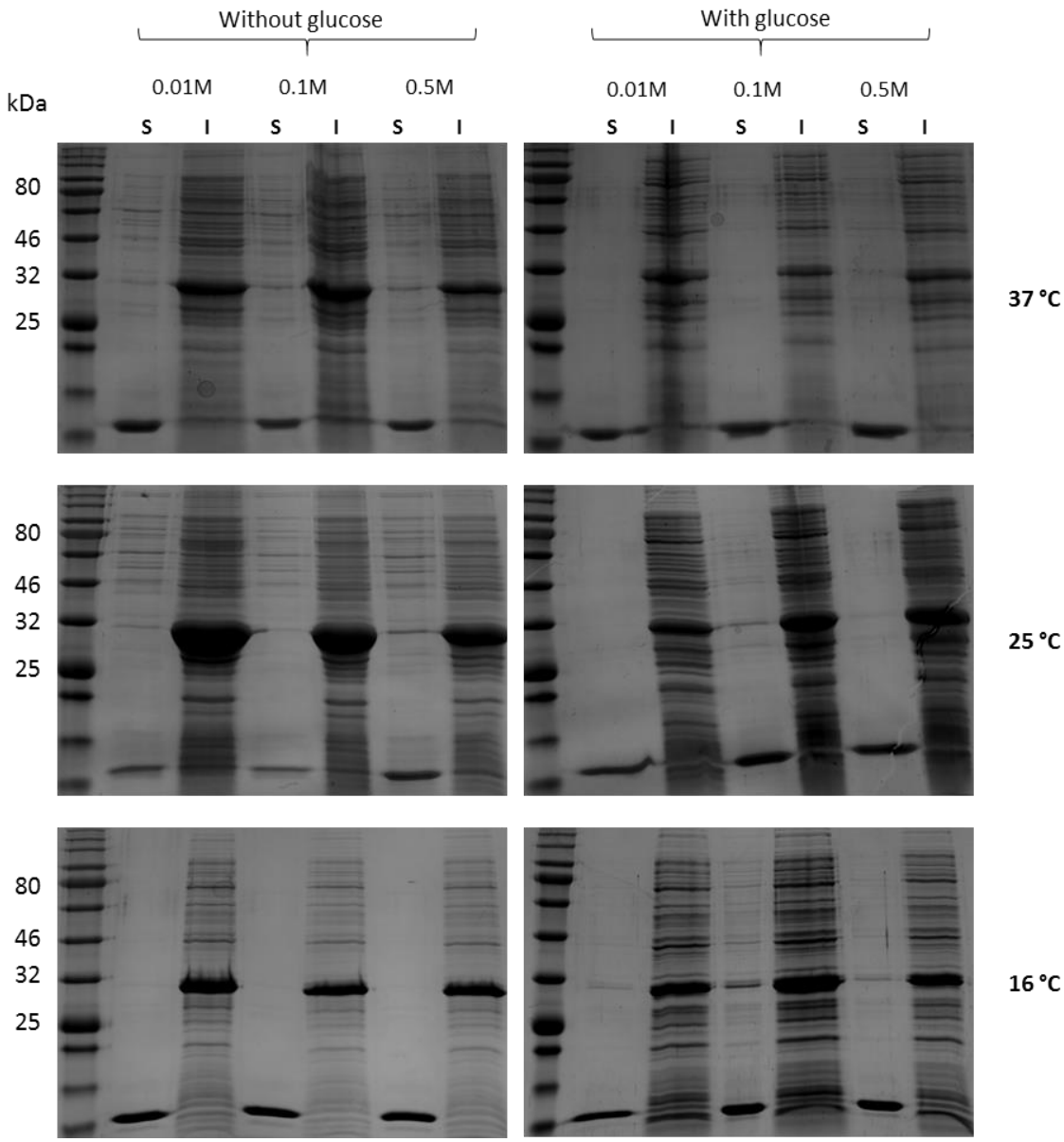


Figure 3.10. His-*CD97 GAIN expression trial in BL21-CodonPlus (DE3)* E. coli. Small scale trials investigated the impact of temperature, IPTG concentration and the presence of glucose on the level of soluble GAIN domain expression. Strong bands present at approximately 32kDa corresponding to CD97 at visible in all insoluble (I) lanes, whilst minimal soluble expression can be seen (S). Lysozyme at approximately 14kDa is evident in all soluble fractions.

3.2.2.2 Vector optimisation for CD97 GAIN expression

Repeated His-CD97 GAIN expression yielded inconsistent amounts of protein, almost to a point of abolished protein expression, despite following the same protocol. Therefore, to optimise consistent protein expression, His-CD97 GAIN was cloned into high-expression vector pKK223-3. pKK223-3 contains the strong tac promoter upstream of the gene of interest, and a strong ribosomal terminator downstream for control of protein expression. Tac promotor is a synthetically generated hybrid of the *E. coli* promotors *lac* and *trp*. Its activity is inhibited by the *lac* repressor, therefore expression, like the pET system, can be induced by addition of IPTG (186). After cloning, His-CD97 GAIN pKK223-3 was transformed in SHuffle® T7 *E. coli* and an expression trial testing both growth medium and IPTG concentration was carried out. The transformed cells were incubated overnight in 2YT, SOB and TB growth media at 37 °C in the presence of ampicillin and used to inoculate 2 litres of the respective medium in duplicate. Protein expression was induced at a final concentration of 0.1 mM or 0.5 mM IPTG and incubated for 68 hours as stated above. 75 µl samples of each culture were pelleted and cells lysed in 30 µl SDS-loading buffer. Cell debris was subsequently pelleted and 10 µl of the resulting supernatant was analysed by SDS-PAGE. His-CD97 GAIN was not expressed under any of the experimental conditions, all of which appear to have similar protein expression to that displayed by the uninduced sample, depicted in Figure 3.11.

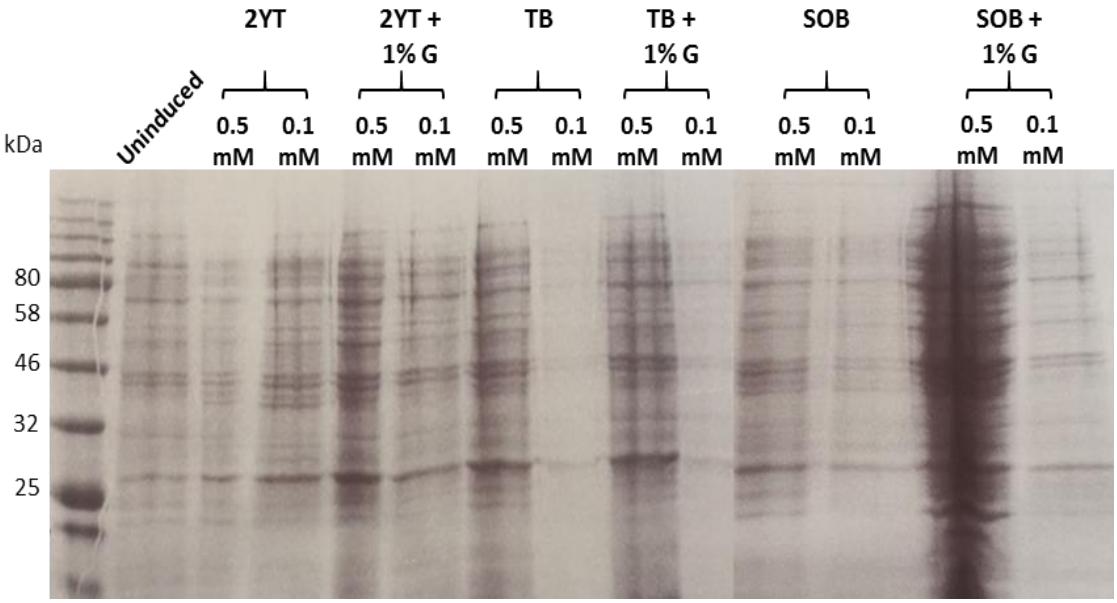


Figure 3.11. ***Expression of His-CD97 GAIN cloned into pKK223-3.*** SHuffle® T7 cells transformed with recombinant pKK223-3 were grown in 3 different growth media, with or without 1% glucose, and induced with 0.1mM or 0.5mM IPTG. A clear lack of protein band at approximately 32 kDa indicates protein expression was unsuccessful.

3.2.2.3 Optimisation of *E. coli* expression strain

As expression of CD97 GAIN cloned into pKK223-3 vector was unsuccessful, the strain which yielded the most His-CD97 GAIN in expression trials was selected for further optimisation. SHuffle® T7 competent *E. coli* are engineered to promote formation of disulphide bonds in the cytoplasm, therefore likely enable proper processing of the GAIN domain due to the presence of two conserved disulphide bonds between the last two beta strands. An initial protein expression trial was carried out by inoculating 5 litres of 2YT media with cultures grown overnight at 37°C, at 1/100 of the final volume. Cultures were incubated at 37 °C until protein expression was induced by addition of IPTG at a final concentration of 0.5mM, and the temperature reduced to 16 °C, predetermined by the Stacey lab. Cultures were grown overnight with agitation and pelleted at 5000g for 5 minutes. The supernatant was discarded, and cells were resuspended in 20mM Tris pH8, 5mM $MgCl_2$ and 10% glycerol, and lysed upon addition of lysozyme and DNaseI. To ensure complete lysis, cells were subject to sonication for 10 rounds at 20 seconds on, 40 seconds off. Cell lysates were then centrifuged for 50 minutes at 18000g, and the supernatant collected as the soluble fraction. After filtration, the protein was purified from the supernatant by Ni^{2+} affinity chromatography in section 2.2.13. The resultant fractions

were resolved on 12% SDS PAGE gel (data not shown). There seemed to be no expression of His-CD97 GAIN in both the insoluble and pre-purification sample, and subsequently only bacterial protein bands were visible in the following elution fractions.

As there was no visible expression in the cell pellet, the same expression conditions were repeated, however incubation time post IPTG induction was increased. Cultures were incubated at 16 °C for 68 hours, over a weekend, and purified as above. SDS-PAGE demonstrated very low-level expression of His-CD97 GAIN, mostly within the pre-purification soluble fraction, and in the 500mM imidazole elution (data not shown). In an attempt to further increase the yield of His-CD97 GAIN, the above modified expression trial was carried out, but the culture volume was scaled up to 16 litres. Figure 3.12 shows an increase in His-CD97 GAIN expression and successful purification in lanes corresponding to the 500mM imidazole elution. It is unsurprising that contaminants still remain, as a number of native *E. coli* proteins show affinity towards metal ions like nickel, and therefore can be difficult to remove (187). However, crystallographic studies require largely pure protein sample to ensure contaminating species are not crystallised, therefore further purification was required to remove both high and low molecular weight contaminants.

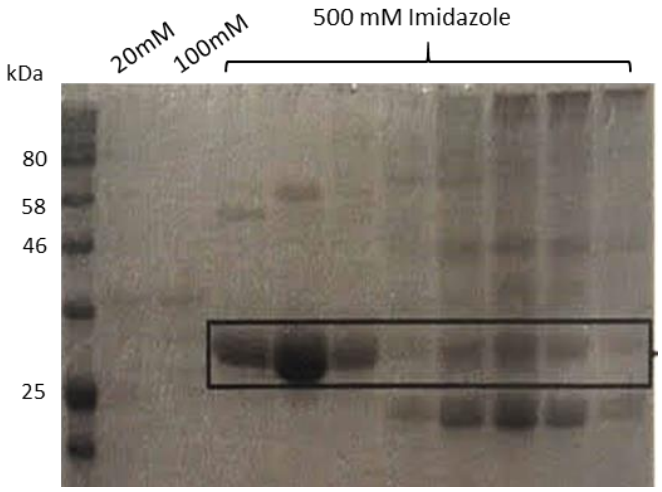


Figure 3.12. ***Ni^{2+} affinity purification of His-CD97 GAIN expressed in SHuffle® T7 E. coli.*** The soluble fraction of His-CD97 GAIN bacterial culture was purified using Ni^{2+}-NTA HisTrap affinity column. Although not shown fully, extensive column washing with 20 mM imidazole removed various contaminants from the column before His-CD97 GAIN, at the expected weight of 32kDa was eluted in 1 ml fractions of 500mM imidazole buffer.

The fractions highlighted above were pooled and concentrated to a volume of 5 ml and subject to further purification by size exclusion chromatography as outlined in section 2.2.13. However, the resultant ÄKTA trace showed only one peak representative of contaminant proteins eluting from the column in the void volume. The lack of a peak representative of His-CD97 GAIN elution suggests either the protein remained attached to the resin, or that the protein was completely degraded by the time it was applied to the column. Subsequently, protein expression was repeated, purified by Ni^{2+} affinity and stored at 4 °C, and monitored for degradation over several days to test protein stability during storage. A sample was taken every 24 hours and analysed by western blot (data not shown). His-CD97 GAIN remained stable in 500mM imidazole buffer at 4 °C for up to 7 days, with minimal evidence of protein degradation demonstrated by precipitation.

3.2.3 Protein purification

3.2.3.1 Initial CD97-GAIN purification

To optimise the yield of soluble His-CD97 GAIN, recombinant pET28a vector was re-transformed into SHuffle® T7 *E. coli* to generate a fresh glycerol stock, and conditions were re-trialled. Expression was induced at an $O.D_{600}$ of 0.6 with 0.5 mM IPTG and incubated with agitation at 16°C for 68 hours. Protein expression was confirmed by SDS-PAGE, and Ni^{2+} affinity chromatography followed by size exclusion chromatography was used to purify His-CD97 GAIN (data not shown). Once again, most protein was lost during size exclusion, therefore to achieve homogenous protein samples, ion exchange was attempted. Protein was eluted from the HiTrap Q HP anion exchange column using a gradient of increasing NaCl concentration.

The chromatogram demonstrated protein elution in 3 peaks, therefore fractions corresponding to these peaks were resolved on SDS-PAGE. His-CD97 GAIN was eluted across the final two peaks alongside a number of contaminants, demonstrated in Figure 3.13A. Little purification was achieved, suggesting that the contaminating proteins had a similar pI to CD97 GAIN. Subsequent rounds of purification gave similar results. The concentration of contaminating proteins was much lower in Figure 3.13B, as is the apparent concentration of His-CD97 GAIN.

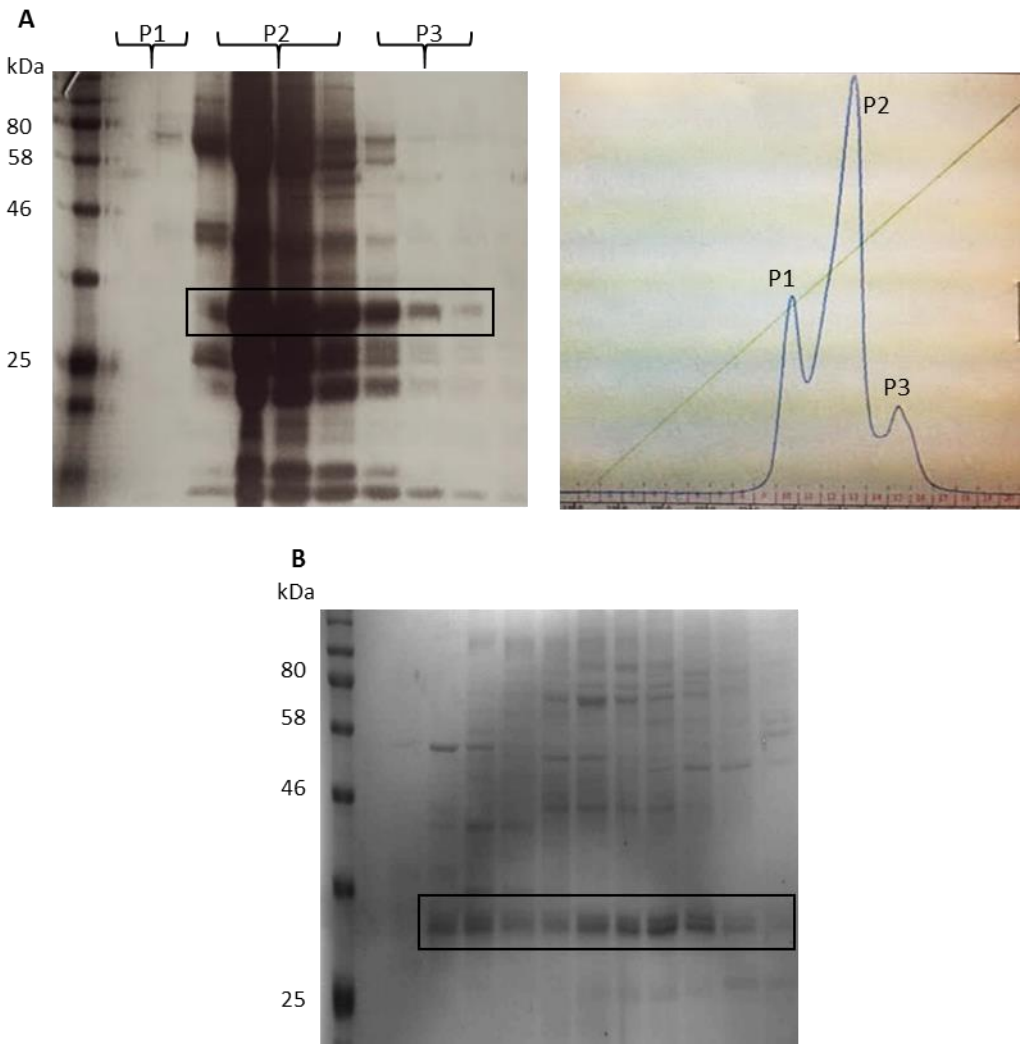


Figure 3.13. ***Ion exchange purification of CD97 GAIN domain, resolved on 12% SDS PAGE.*** (**A**)His-CD97 GAIN was unsuccessfully purified by ion exchange following Ni^{2+} NTA affinity chromatography. All fractions contain a high concentration of contaminating protein products, demonstrated by the thick, smeared bands**.** Repeated ion exchange purification demonstrated little improvement, with comparatively similar levels of contaminants present in elution fractions as CD97 GAIN (**B**)**.**

3.2.3.2 Purification optimisation

As both size exclusion and ion exchange were unable to successfully purify His-CD97 GAIN to give a sample suitable for crystallographic trials, Ni^{2+} affinity purification was optimised. ÄKTAprime apparatus and 1ml HisTrap column (GE Healthcare) were used to employ an imidazole gradient elution protocol (20 column volumes of 20 mM and 50 mM imidazole buffer, 50 column volumes of 80-500mM imidazole). A single clean peak was seen almost immediately after the gradient was set, at approximately 150 mM imidazole. The fractions collected corresponding to the peak were analysed on SDS-PAGE, alongside samples from each wash step. His-CD97 GAIN was successfully purified by shallow gradient elution of increasing imidazole concentration (Figure 3.14). The fractions containing CD97 were pooled and concentrated before dialysis in 1X PBS and 5% glycerol. Prior to crystallisation trials protein identity was confirmed by LC-mass spectroscopy. Figure 3.15 displays the atomic spectra of the sample confirms the mass to be 32611Da, which corresponds to His tagged-CD97 GAIN domain.

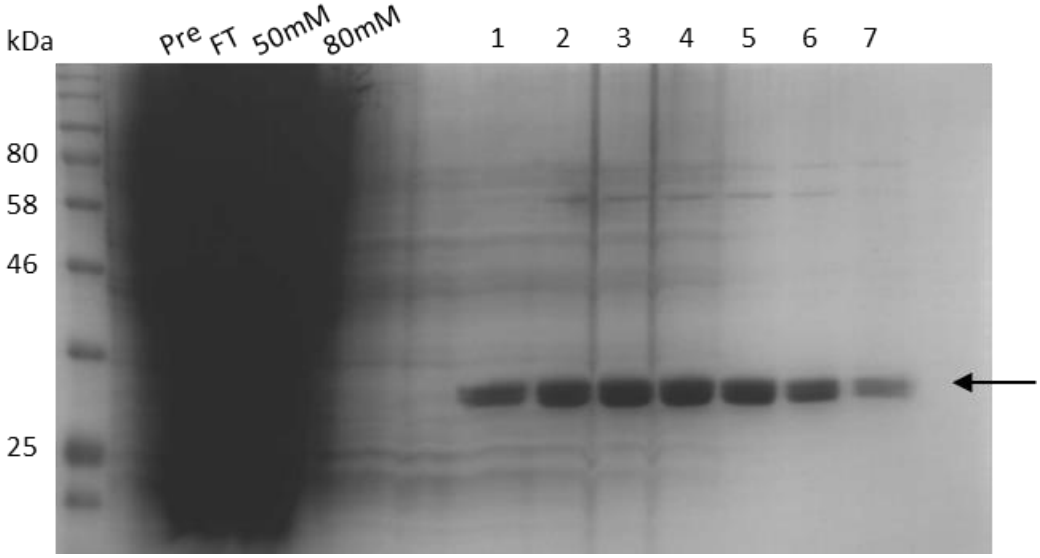


Figure 3.14. ***Purification of His-CD97 GAIN by Ni^{2+} NTA affinity chromatography.*** His-CD97 GAIN was eluted from the HisTrap column by a shallow gradient of increasing imidazole concentration. Protein was eluted at a relatively low concentration without contaminants, which were eluted from the column in the flow through (FT), 50mM and 80mM washes, evident from the thick dark smear of protein in lanes corresponding to these fractions. Numbers correspond to fractions collected from the increasing imidazole gradient.

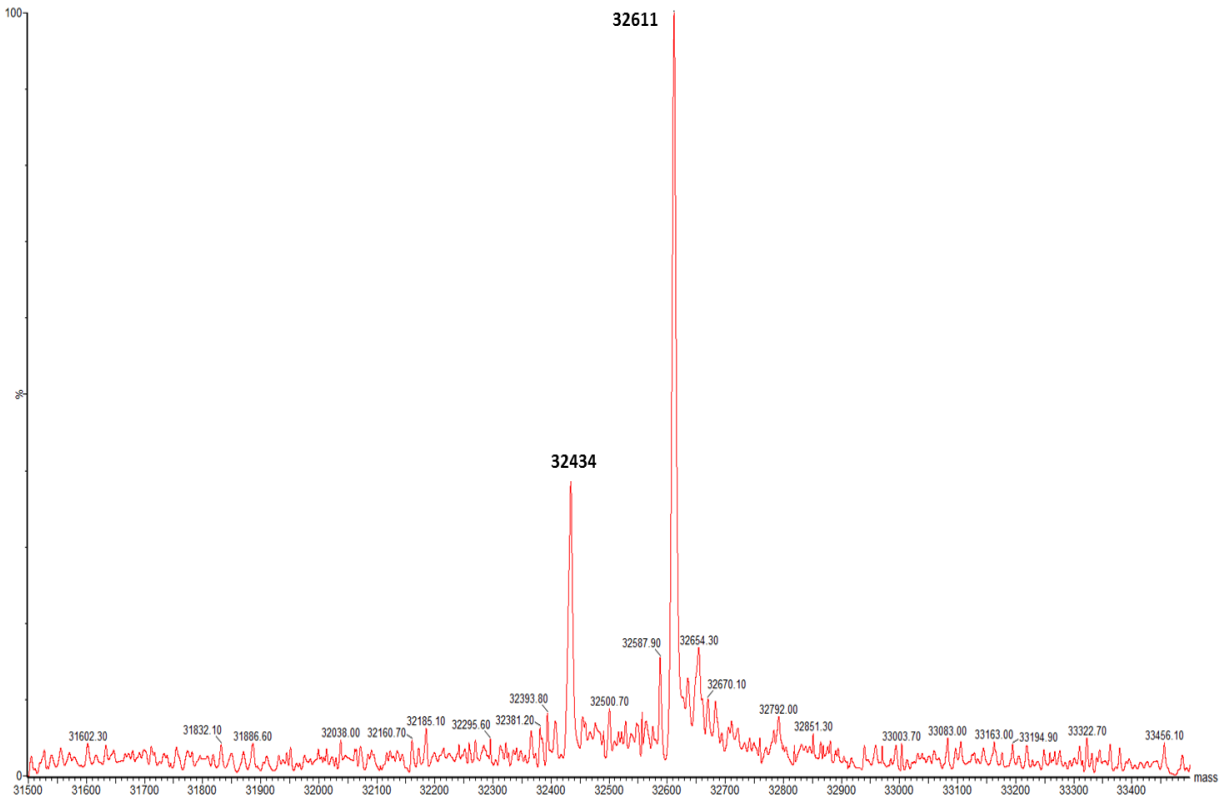


Figure 3.15. ***Native atomic mass spectrum of CD97 GAIN*****.** Purified CD97 was analysed by LC-mass spectrometry. The major protein species in the sample was confirmed to be 32611 Da. Numbers by peaks denote the mass of the protein.

3.2.3.2 TEV protease cleavage for His-Tag removal

A substantial challenge in crystallography is posed by the structural stability of the protein. Areas of protein structure that are flexible or disordered reduce the ability to form crystalline asymmetric units that are of the same shape and size. Therefore, conformational stability is of the utmost importance for increasing the likelihood of producing well-formed crystals.

The presence of a purification tag can introduce a flexible region and reduce the chance of lattice formation. However, removal of such tags can also reduce the stability of the protein of interest, causing it to precipitate before reaching crystal trials. Therefore, the stability of CD97 GAIN without an N' terminal His-Tag was investigated. Tobacco etch virus (TEV) protease is a highly sequence-specific protease used for strict removal of fusion-tags, recognising a 7-amino acid sequence. The His-tag was removed via TEV protease cleavage, and a sample of un-cleaved

(His-CD97 GAIN) and cleaved (CD97 GAIN), alongside TEV protease alone, was resolved by 12% SDS-PAGE (Figure 3.16). The His-tag was successfully cleaved from the GAIN domain.

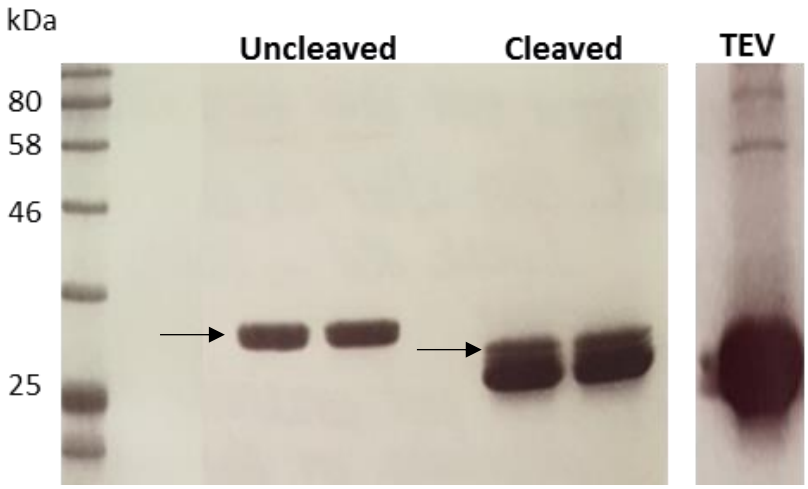


Figure 3.16. ***TEV protease cleavage of N' terminal His-tag from CD97-GAIN.*** Duplicate samples were resolved for clarity as only a 2kDa reduction in size was expected. Pure TEV protease was included as a reference for the cleaved sample.

After cleavage, CD97 GAIN was stored at 4°C overnight for use in crystal trials. However, the protein was seemingly unstable without the N' terminal tag and largely precipitated, substantially reducing the concentration of the sample. CD97 GAIN continued to precipitate almost immediately after treatment with TEV protease in subsequent rounds of expression, and consequently could not be taken forward for crystal trials.

3.3 Crystallographic study of CD97 GAIN domain

3.3.1 Initial screens for crystallisation conditions

To obtain atomic resolution of the CD97 GAIN domain using X-ray crystallography, diffracting crystals were first required. His-CD97 GAIN was purified as outlined in section 3.2.2.4 and screened for crystallisation conditions using multi-factorial 96-well plates. A large number of commercially available crystallisation screens were used during the course of this study including; Joint Centre for Structural Genomics Core (JCSG) suite screens (I-IV), Hampton Research Crystal screens 1 and 2, SaltRX and Index screens, Rigaku Reagents Wizard screens 1-4 and Molecular Dimensions PACT, MIDAS and Morpheus screens. These screens enabled the sampling of approximately 1,200 different conditions covering a wide-range of precipitants, salts, buffers, pH, polymers and organic molecules in a high-throughput 96 well format. Screening was performed using the Formulatrix automated NT8 crystallisation robot system to efficiently and reproducibly set up sitting-drops with various protein:mother liquor ratios. Protein was concentrated to 4 mg/ml and screens were established with three protein:mother liquor drops at 1:1, 1:2 and 2:1 ratios, with protein volumes of 0.1 µl, 0.2 µl and 0.05 µl. Crystallisation trials were automatically imaged using the Formulatrix Rockimager 1000 system with visible and UV imaging technology. Drops were monitored for one month for signs of crystal growth. However, only very small crystals, protein precipitation or clear drops were observed across all screen conditions

His-tag cleavage via TEV protease was attempted again to reduce the likelihood of flexible regions hindering crystal formation. Contrasting previous efforts (section 3.2.3.2), CD97 GAIN remained stable after removal of the N' terminal tag for up to one week at 4 °C as confirmed by native mass spectrometry analysis. In addition to crystallisation trials conducted using the commercially available screens, a narrower set of conditions were screened. A range of PEG polymer sizes were tested as the precipitating agent, across a set pH range established using citric acid or MIB (sodium malonate, imidazole and boric acid) buffer. These conditions were identified previously as those yielding very small crystals. The four screens were as follows; PEG400 MIB pH 4.5-6.5, PEG1500 MIB pH 5-7, PEG2000 MIB pH 4.5-6.5 and mixed PEG sizes with critic acid pH 4.6-5.8. CD97 GAIN crystallised in a few conditions, with the most promising conditions being 30%, 23% and 15% v/v PEG1500, at pH 5.8 and pH5 as shown in Figure 3.17. However, these crystals were very thin (less

than 10 µm) and were difficult to distinguish under visible light. However, they were confirmed as protein crystals by UV imaging.

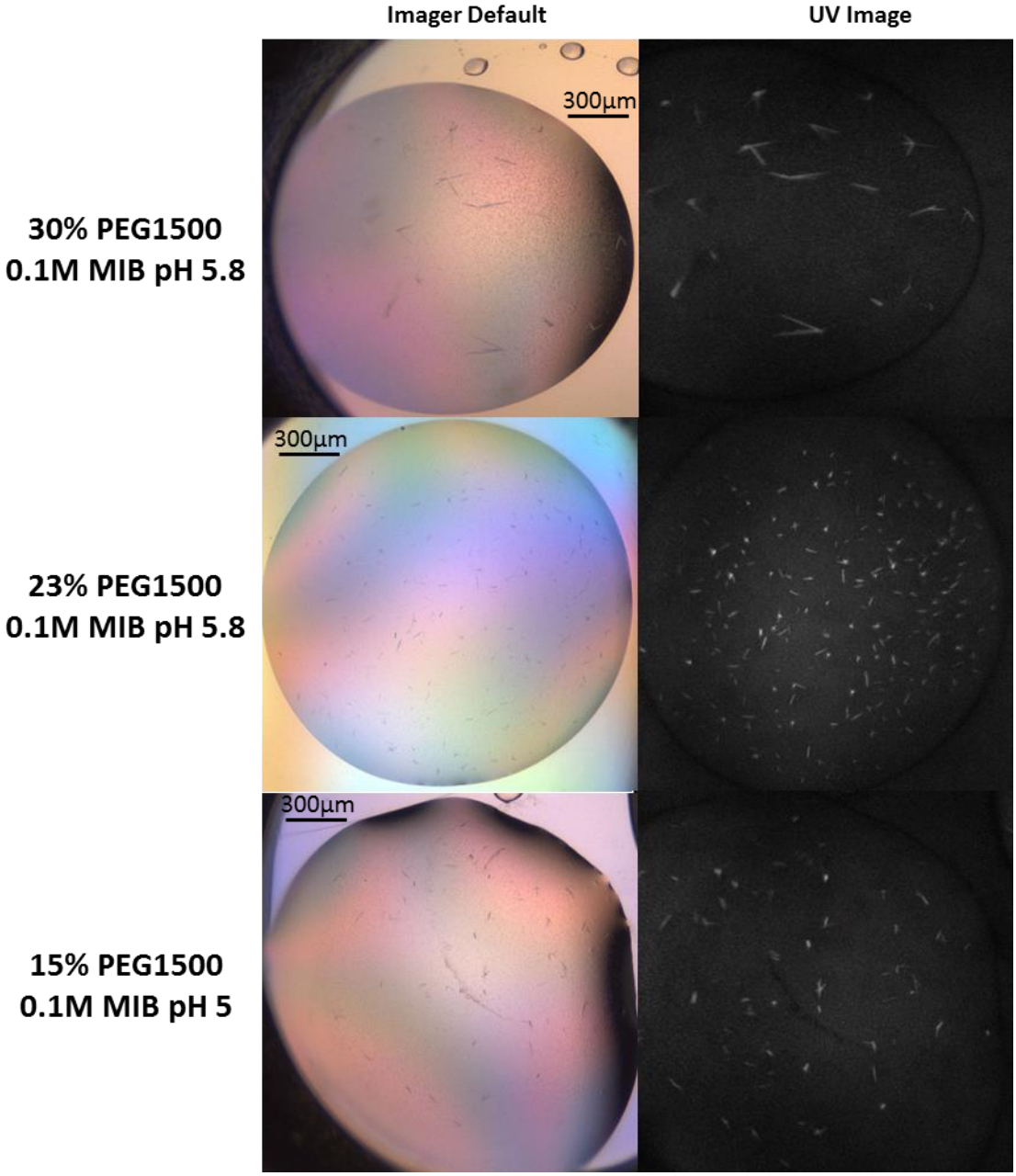


Figure 3.17. ***Crystal growth after His-tag cleavage*****.** CD97 GAIN protein formed thin crystals of less than 10 µm in thickness over a period of two weeks, using 15-30% PEG 1500 and 0.1M MIB buffer pH 5-5.8 at 20 °C.

3.3.2 Crystal growth optimisation

In attempt to gain large singular, diffraction quality crystals, the same focussed screens trialled for crystallisation of CD97 GAIN (section 3.3.1) (Figure 3.17) were tested for His-CD97 GAIN. Each plate was screened in duplicate, investigating the effect of different protein concentrations; 8 mg/ml or 2-6 mg/ml. Initial "hits" were identified after one week with crystals formed in drops without excess protein precipitation in drops containing PEG1500 and MIB buffer pH5, at different protein concentrations. A selection of crystals that were produced under these conditions are illustrated in Figure 3.18.

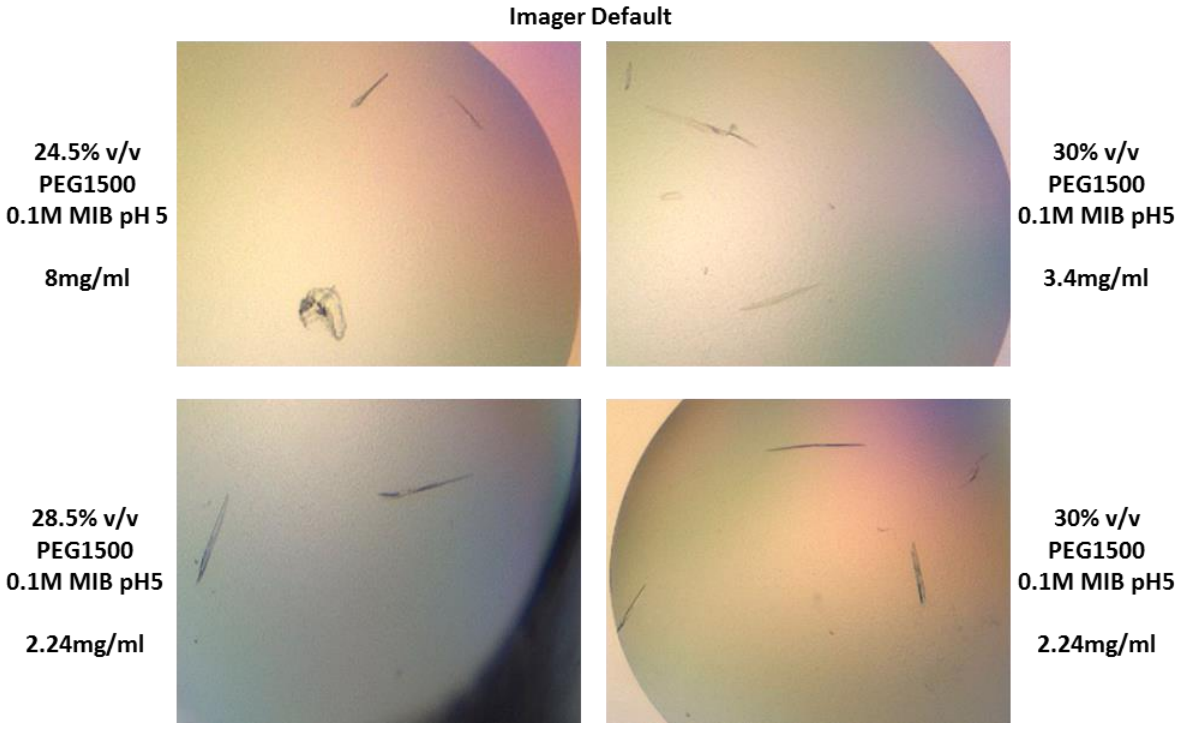


Figure 3.18. ***Crystal growth as a result of sparse matrix screening.*** His-CD97 GAIN protein formed thin crystals of less than 10 μm in thickness after one week using a 20-30% PEG 1500 and 0.1M MIB buffer pH 5 at 20 °C.

As the crystals were less than 10 μm for high resolution X-ray diffraction data collection, further optimisation around these crystallisation screens was performed using the hanging drop vapour diffusion technique and streak seeding (Section 3.1.1) (188). A seed stock using typically small crystals, not suitable for data collection, was generated using a cat whisker to crush the crystals. A drop of freshly purified protein, at a final concentration of 4 mg/ml, and mother liquor was mixed at 1:1 ratio. The

whisker was streaked through a first drop, and then through subsequent drops to generate a serial dilution, moving across the plate. Figure 3.19 demonstrates the 24-well plate layout and basic mother liquor composition at 500 µl per well.

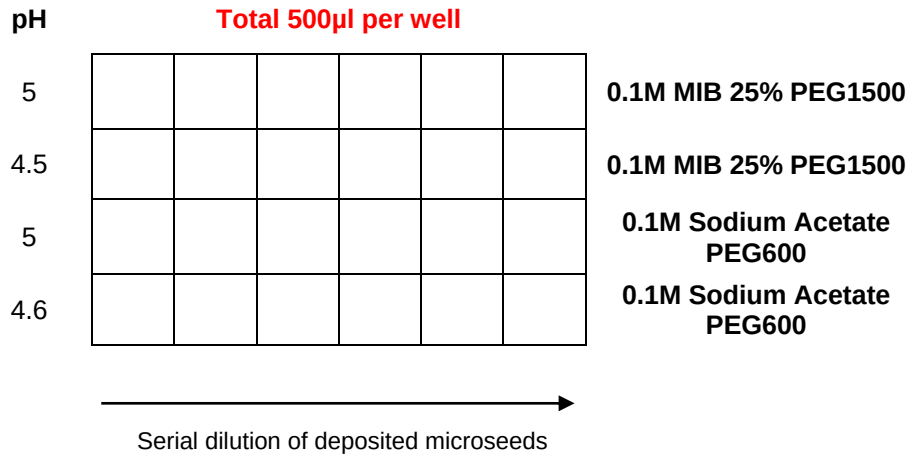


Figure 3.19. ***Micro-seeding using the streak seeding technique***. A 24-well hanging drop crystallisation plate was set up based on conditions that previously produced crystals. Each drop contained 1 µl of protein (4 mg/ml) along with 1 µl of mother liquor for that respective well. A cat whisker was used to dislodge seeds from a CD97 GAIN crystal before streaking through each drop in a row, creating a serial dilution of the deposited seeds.

After two weeks, there was no observed crystal formation or precipitation in drops containing sodium acetate buffer, however small crystals had again grown in drops with MIB buffer. It was also noted that low-level precipitate formed between pH 4.8 and 5.2, and that drops with less than 15% PEG were under-saturated. Therefore, further optimisation was carried after streak seeding, around the conditions outlined in Figure 3.19. A new screen was established in 24-well plate format, in triplicate, for incubation at 25 °C, 18°C and 4°C (Figure 3.20).

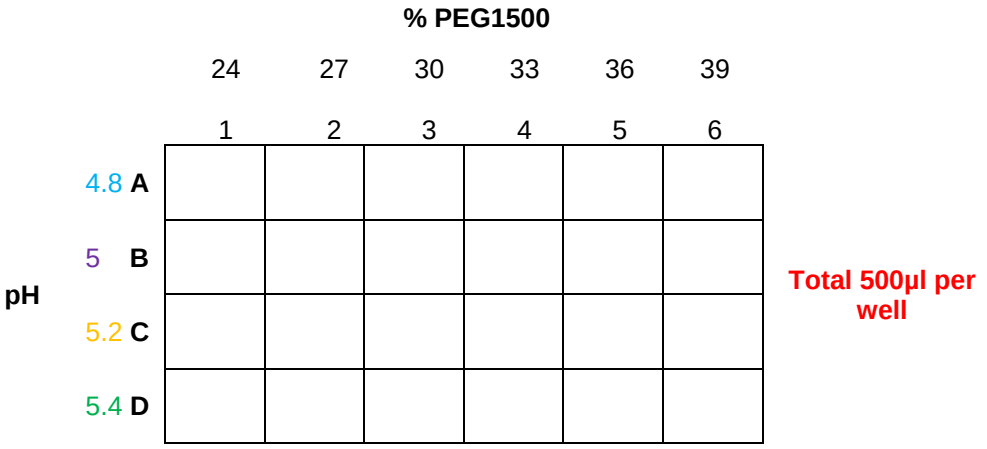


Figure 3.20. ***Further optimisation of CD97-GAIN crystallisation conditions with MIB buffer.*** 1 µl of purified CD97-GAIN at 2 mg/ml was mixed with 1 µl of mother liquor and set up as a hanging drop above the respective well buffer. Plates were established in triplicate to test the implication of different incubation temperatures.

Drops were monitored for crystal growth for one week, after which spindle crystals along with precipitation had formed in the plate incubated at 18 °C. In contrast, neither precipitate or crystals were observed at 4 °C. Multiple conditions facilitated crystal growth when incubated at 25 °C, and many formed branched spindles which are unsuitable for X-ray crystallographic experiments. Consequently, in a further attempt to increase crystal size and thickness, the spindle crystals were used to generate a seed-stock. Conditions B1-4, C1-4 and D1-4, as in Figure 3.20, were replicated in a 96-well plate, detailed in Figure 3.21. A 6 µl drop of purified protein, at a final concentration of 2 mg/ml, and mother liquor was added per well at a 1:1 ratio. The cat whisker was streaked through the 6 drops to serially dilute crushed crystal, and the plate incubated at 25 °C for up to one month, with constant observation for crystal growth. Growth was observed after 2 weeks in the direction of streaking, across the full range of PEG1500 concentrations, with crystals ranging in size and thickness, although once again, only spindle-shaped crystals were obtained.

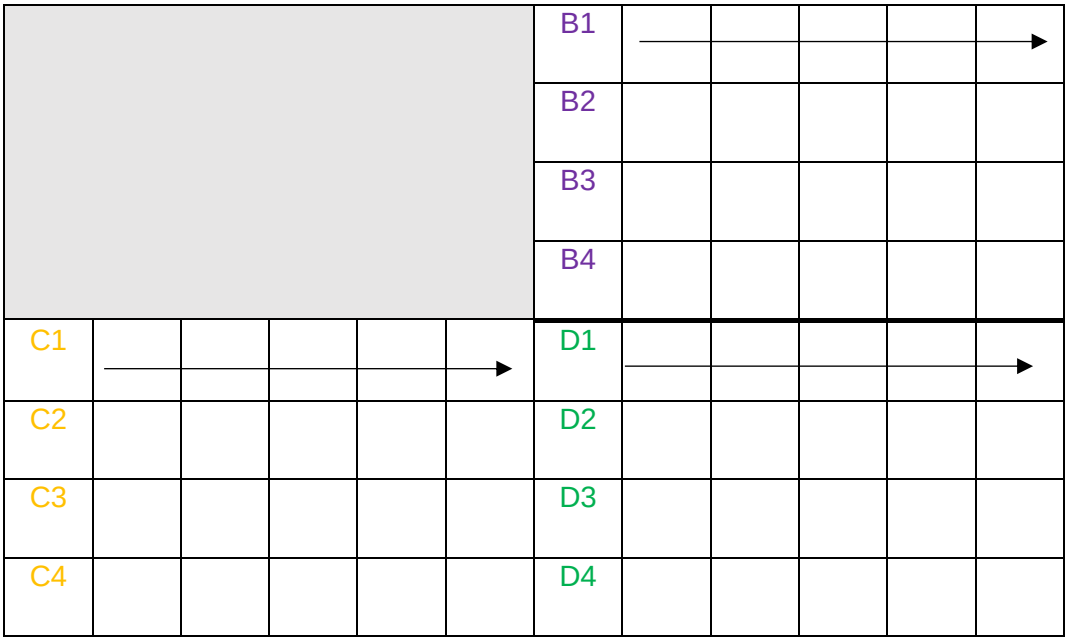


Figure 3.21. ***Streak seeding of His-tag cleaved CD97-GAIN.*** Mother liquor components were selected from wells in which crystal growth was visible in a previous screen. Streak seeding was again used to improve crystal quality. Arrows depict the direction of dilution. MIB buffer: **B**; pH 5, **C**; pH 5.2, **D**; pH5.4. PEG1500 was included as the precipitant: **1**; 24%, **2**; 27%, **3**; 30%, **4**; 33%.

Alongside focussed PEG screening, an additive screen was also performed. CD97-GAIN, including the N' terminal His-tag, was freshly purified and concentrated to 4 mg/ml. The additive screen protein conditions replicated those of the JCSG core suit screen stated previously, and trays were monitored for one month for signs of crystal growth. Additive trays with wells containing 0.05µl 4mg/ml CD97-His had no crystal formation, with only precipitation evident, or were clear suggestive of under-saturation. Three wells containing 0.1µl 4mg/ml CD97-His had 2D shard-like crystal growth, whilst the remaining wells had precipitation of varying amounts. Seventeen wells containing 0.2µl 4mg/ml CD97-His formed shard-like crystals of varying sizes, a selection of which are presented in Figure 3.22. In this instance, PEG2000 rather than PEG1500 yielded the best quality crystals.

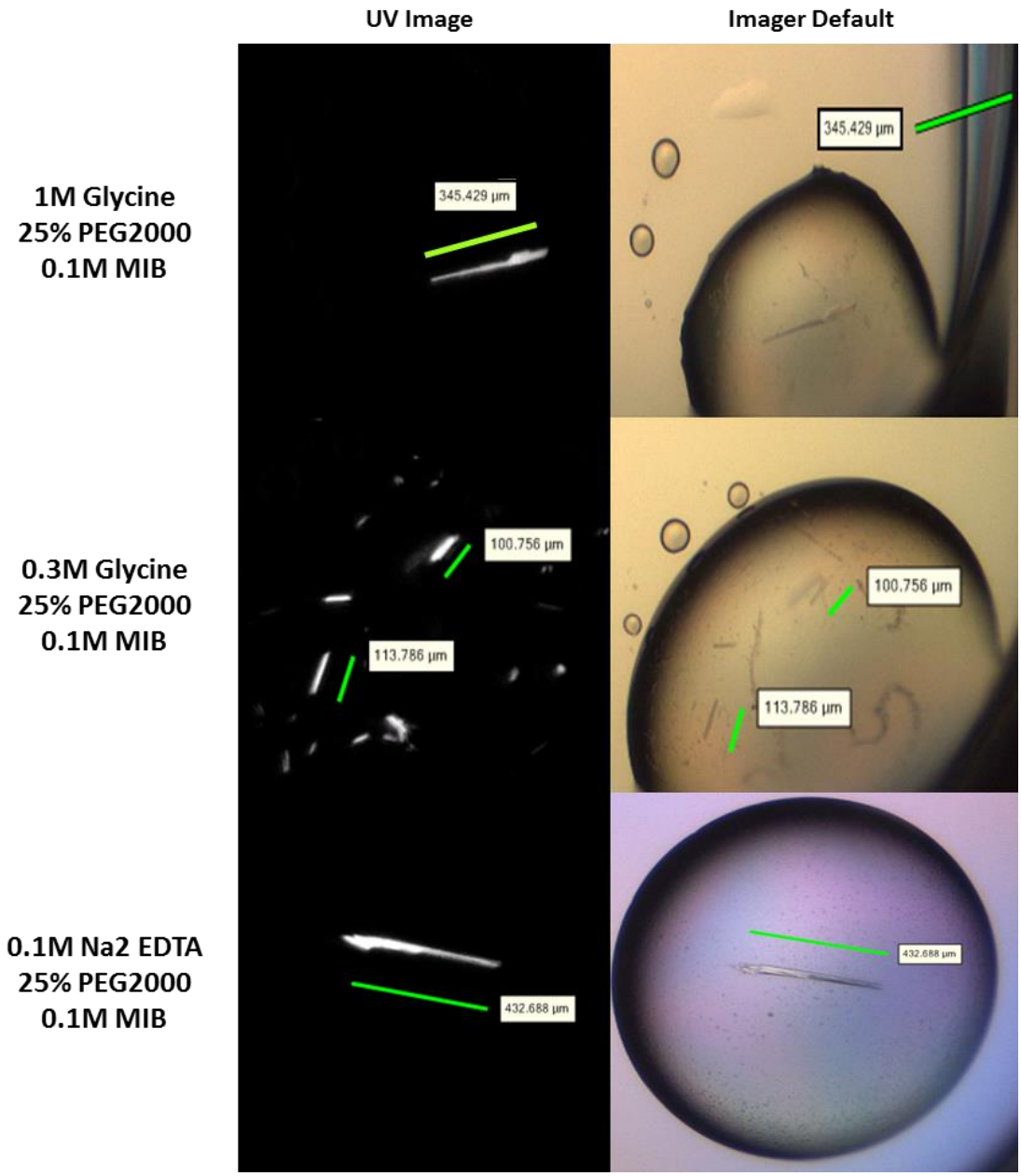


Figure 3.22. ***CD97-GAIN additive screening for crystallisation*****. Longer and thicker spindle** crystals than those previously obtained were formed in the presence of a range of additives, including glycine, EDTA, methanol, acetone and galactose.

3.3.3 Crystal harvesting and data processing

Crystals obtained from both the additive screen and focussed PEG trays, were transferred by Dr Chi Trinh to a cryoprotectant solution consisting of 25% glycerol with the mother liquor before mounting in loops and flash-cooled directly into liquid nitrogen. A single crystal of approximately 100 µm x 20 µm x 20 µm diffracted to 2.2 Å on beamline I02 at the Diamond Light Source. A representative image of the diffraction pattern recorded is shown in Figure 3.23. Subsequent data processing was carried out by Dr Chi Trinh. The diffraction images were integrated, scaled and truncated using the XIA2 suite of programmes (168), by Dr Chi Trinh. The data belong to space group $P2_1$, with unit-cell parameters a = 53.8Å, b = 62.1Å, c = 89.2Å, α=90.0°, β=93.3°, γ=90.0°. The processing and crystallographic statistics from XIA2 are summarized in Table 3.2. From the data a Matthew coefficient was calculated pertaining to 42 % solvent content with the most likelihood of one molecule of CD97 GAIN in the asymmetric unit cell.

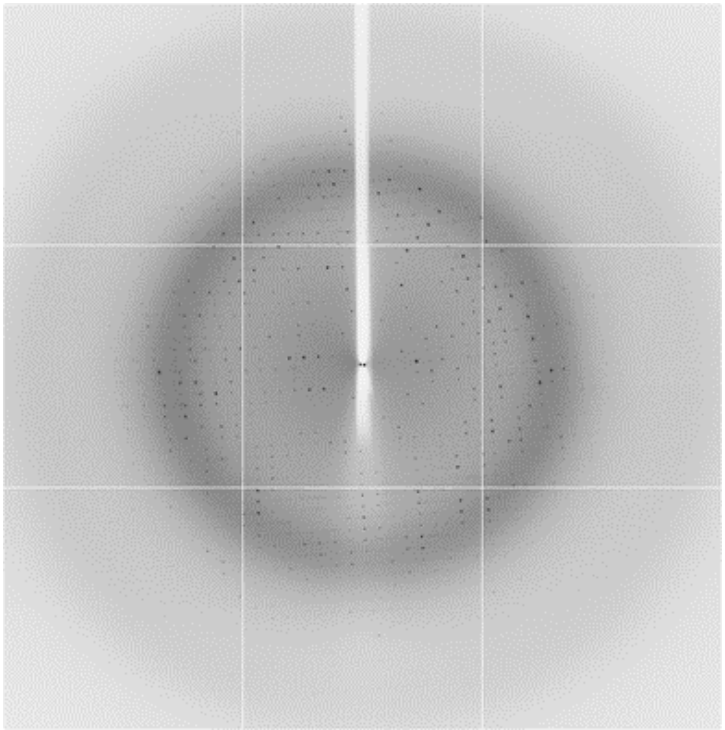


Figure 3.23 ***Diffraction pattern CD97-GAIN*****.** The order and intensities of the reflection spots provides the information needed to determine the X-ray structure of CD97-GAIN. However, the phase information cannot be obtained experimentally from the diffraction pattern.

Dataset	CD97 GAIN domain
Source	Diamond Beamline i02
Wavelength (Å)	0.9795
Resolution range (Å) *	37.59–2.16 (2.22-2.16)
Space group	$P2_1$
Unit-cell parameters (Å)	a=53.8, b=62.13, c=89.22
	α=90.0, β=93.3, γ=90.0
No. of observed reflections	114823 (8579)
No. of unique reflections	31439 (2329)
Redundancy	3.7 (3.7)
Completeness (%) *	99.0 (99.4)
$< I/\sigma(I) >$*	15.8 (2.6)
R_{merge} (%)§*	5.3 (49.7)

*Values given in parentheses correspond to those in the outermost shell of the resolution range.

§ $$R_{merge} = \sum_{hkl} \sum_{i} \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum I_i(hkl)$$

Table 3.2 ***Data collection and processing for CD97 GAIN domain***

In order to determine the structure of the CD97 GAIN domain, molecular replacement (MR) (189) was performed with search models comprising of the crystal structures of the adhesion-GPCRs Lat1/Lphn1 (PDB 4DLQ (26)), BAI3 (PDB 4DLO (26)) and GPR56 (PDB 5KVN (178)).

Whilst using the crystal structures of the adhesion-GPCRs Lphn1, BAI1 and GPR56 as search models, it was not possible to determine the structure CD97 GAIN domain. Exhaustive searches were performed using numerous edited constructs of these search models. These models included; truncated loop regions, purely the N-terminal helical region of the GAIN domain, purely the conserved C-terminal β-sheet domain,

and sequence and structurally-aligned regions of the Lat1, BAI3 and GPR56 to CD97 GAIN domain. Different molecular replacement programmes were used to carry out these searches; PHASER (190), AMPLE (191) and SIMBAD (192).

Heavy atom (HA) soaking was conducted using small needle crystals of CD97 GAIN domain with a solution consisting of 4-Chloromercuribenzenesulfonic acid, sodium salt and hexa-chloroplatinate (IV), at concentrations ranging between 0.1 – 5 mM in the mother liquor. Different soak times were also trialled. However, at the highest concentration of HA solution, crystals cracked and dissolved. Subsequently, crystals soaked in both the heavy atom solution for 30 minutes were harvested and flash frozen in liquid nitrogen. However, these crystals gave rise to a poorly diffused and disorder diffraction pattern below 10 Å resolution.

Finally, sulphur single-wavelength anomalous diffraction data collection was attempted on some CD97 GAIN domain needle crystals. This aimed to collect the anomalous signals from the sulphur atoms of the methionine and non-disulphide bonded cysteine residues. Three datasets were collected to resolution of approximately 3.5 Å. However, no sulphur anomalous signal was found during subsequent data reduction.

3.3.4 Structural determination data with the use of Affimers

Binding proteins are routinely used to stabilise protein conformation in structural studies. Indeed, adhesion GPCR GPR56 crystal structure was recently solved in complex with a monobody scaffold protein, therefore we investigated the potential of Affimers as crystallisation chaperones. Furthermore, as the Affimer crystal structure has previously been solved, we hoped the co-crystallisation of the GAIN domain in complex with an Affimer would aid in solving the experimental phases after acquiring an electron diffraction data set.

3.3.4.1 Generation of Affimer scaffold proteins against CD97 GAIN

For generation of stable, target specific affimers, the CD97 GAIN domain was expressed and purified as previously stated. To identify binding molecules, an *in vitro* selection approach was used, described by Tiede, C. et al., 2014. Purified CD97 GAIN was screened against a phage library, consisting of $1.3x10^{10}$ filamentous phage expressing affimer clones, each with randomised loop regions. A 1000-fold amplification of eluted phage was achieved after three rounds of panning against the

GAIN domain. Following DNA sequencing of specific-binding clones, many were found to be replicates and were consequently removed from the pool. 10 positive clones were selected (Figure 3.24A) and a 6xHis-tag was incorporated at the C-terminus of each via sub-cloning in to pET11a. Further, an additional copy of each affimer was engineered to possess a single free C-terminal cysteine residue for application in future experiments generating a total of 20 proteins.

3.3.4.2 Expression and purification of Affimers

Plasmids for each affimer were transformed into BL21 (DE3) *E. coli* and expressed at a final concentration of 8-12 mg/l. Purification samples were resolved by SDS-PAGE, which demonstrated soluble and stable expression of each scaffold protein without an additional cysteine, at the expected size of 12kDa. However, after nickel affinity chromatography, Affimers 3, 4 and 8, containing an extra cysteine, were found to be insoluble. To optimise the purification process, these were dialysed in 1X PBS, 5% glycerol, and 20mM Tris, 300mM NaCl 5% glycerol. This however did not rescue protein solubility, and hence affimers 3, 4 and 8 were disregarded. Mass spectrometry further confirmed the molecular mass of each protein and indicated samples contained no obvious contaminants after purification (Figure 3.24D).

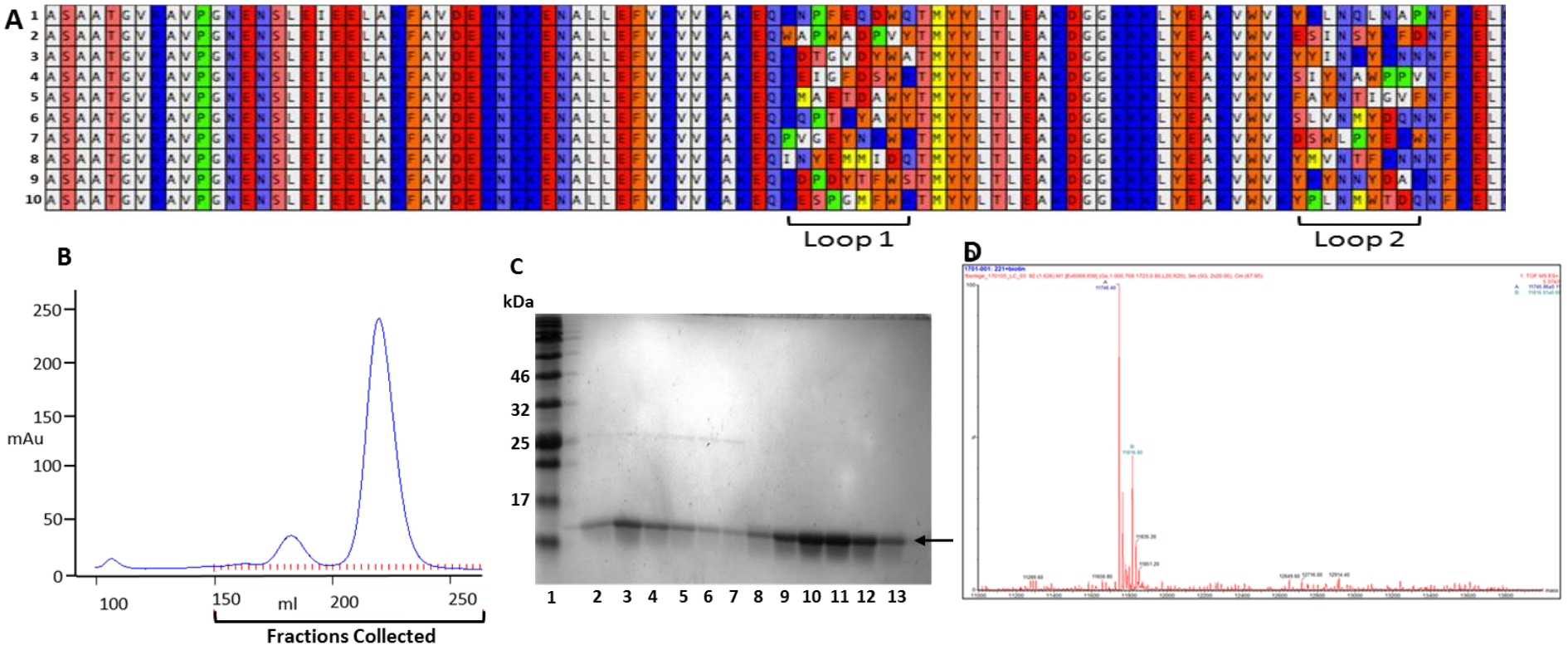


Figure 3.24. ***Cloning, expression and confirmation of Affimers against CD97-GAIN.*** **(A)** Sequence alignment of 10 affimers positive for binding to CD97 GAIN domain, showing variable sequence regions 1 and 2. (**B**)+(**C**) Following Ni^{2+} affinity purification of Affimers, contaminants were removed by size-exclusion chromatography and resolved on 15% SDS-PAGE gel. (**D**) Example mass spectrometry analysis and verification of biotinylated affimer 1, representative of all produced.

3.3.4.3 Confirmation of Affimer specificity for CD97-GAIN

To clarify the specificity of each Affimer for CD97 GAIN, those engineered to contain an extra cysteine residue were biotinylated using EZ-link Sulfo-NHS-Biotin labelling kit and used in ELISA analysis. Affimers were added to both CD97-GAIN and control protein interleukin-36α coated plates in a two-fold serial dilution. Relative binding affinity was normalised against Affimer binding to interleukin-36α (Figure 3.25). Affimers 3 and 5 bound specifically with the most affinity after serial dilution at approximately 1 µg/ml. Interaction with almost all affimers was detectable at dilutions considerably higher than that determined as background binding by interaction with IL-36α. Non-specific or weak binding to CD97 GAIN was demonstrated by Affimer 4 and somewhat Affimer 1. ELISA analysis revealed little difference between the dilutions at which binding to CD97 GAIN and interleukin-36α was detected and therefore the two Affimers were not carried forward for future experiments. Due to their specific interaction with the GAIN domain, the remaining 5 will be taken forward for trials as stabilising proteins in crystal trials, although have not yet been tested.

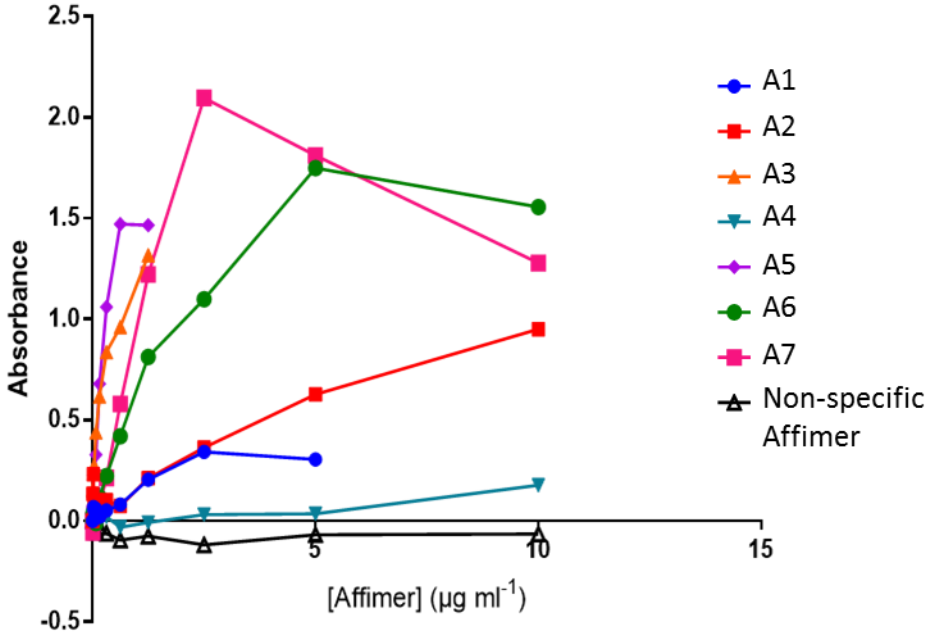


Figure 3.25 ***Relative binding affinities for anti-CD97 GAIN domain Affimers*****.** 96-well plates were coated with CD97 GAIN domain or interleukin-36α, and Affimer binding measured by ELISA. OD was normalised to non-specific binding of affimers against interleukin-36α.

3.4 Discussion and future perspectives

Adhesion GPCRs play critical roles in various biological systems although as outlined in the introduction, the mode by which the tethered agonistic peptide is exposed remains ambiguous. While the atomic structure of three adhesion GPCR GAIN domains have already been determined, more insight is required to confirm which residues are responsible for mediating possible conformational changes which facilitate uncovering of the *Stachel* sequence.

Whilst the GAIN domain is an evolutionarily conserved protein domain, the EGF-TM7 subfamily GAIN domain displays distinct structural variation. The predicted secondary sequence alignment of all known adhesion GPCRs suggests that CD97 possibly lacks the first 3 α-helices. The implications, if any, of these differences on receptor activity are currently unclear. This chapter aimed to express CD97 GAIN domain in association with its EGF repeats and solve the crystal structure as an example EGF-TM7 family member, providing molecular insights into the mode of action of these largely myeloid-restricted receptors.

3.4.1 Protein expression

Molecular and structural characterisation of proteins requires successful expression and purification in yields high enough for the chosen method. Crystallographic experiments require milligram quantities of homogenous protein to test an array of crystallisation conditions. *E. coli* systems are highly popular due to low cultivation costs, rapid growth and often very high protein yield. However, expression of native mammalian proteins can prove problematic, as the absence of post-translational modification machinery and chaperones can result in improper protein processing and the generation of insoluble clumps of protein within inclusion bodies (193). On the other hand, whilst mammalian cell culture enables the correct post-translational modifications to take place, often too little protein is produced for utilisation in structural studies. This is particularly challenging for smaller research groups where large-scale expression is expensive, time consuming and difficult, and therefore unattainable.

The insect cell protein expression system, based on infection with lytic baculovirus, provides a favourable alternative, due to relatively low cost, large cloning capacity and the extensive protein modifications that can be executed, such as phosphorylation and acetylation, whilst generating a relatively protein yield

(194)(195). Furthermore, unlike the reducing environment of the E. coli cytosol, folding and disulphide bond formation is also performed, which is desirable in the case of adhesion GPCR expression due to the presence of two disulphide bonds close to the cleavage site within the GAIN domain. Although both O- and N-glycosylation can be achieved, it should be noted that N-glycosylation does not necessarily occur as it would do natively (196)(197). Many GPCRs including adenosine A_{2A}, bradykinin BK_2, various cannabinoids and opioids have been expressed to a high level in Sf9 insect cells via baculovirus infection, and indeed the majority of previously crystallised GPCRs were successfully expressed for structural determination in this manner (198)(199). Similarly, the GAIN domains of Lpn1, BAI3 and GPR56 were expressed the same way, demonstrating it to be a suitable method for adhesion GPCR expression.

Therefore, baculovirus was prepared with CD97 GAIN-EGF 1,2,5 by site-specific transposition of an expression cassette from a donor plasmid to the bacmid vector in DH10Bac *E. coli*, before transfecting Sf9s with recombinant bacmid. To increase protein yield, virus particles were harvested and used to infect Sf9s for multiple rounds of viral amplification. Viral titre was determined by qPCR, measuring the accumulation of amplified product by use of a fluorescent reporter, in this case SYBR® green, which binds to double stranded DNA molecules as they form. This enables to rapid quantification of gene copy number in a sample in real time. After determining the titre, expression trials were conducted to determine the optimal MOI and harvest time after infection. Protein expression was primarily confirmed by western blot. CD97 GAIN-EGF 1,2,5 was initially expressed as insoluble protein within inclusion bodies. Subsequent solubilisation of the cell pellet was carried out using non-ionic detergent NP-40 to ensure the GAIN domain retained biological activity by preventing disruption of the non-covalent interactions responsible for association of the subunits (200). This however, was unsuccessful in releasing the protein from inclusion bodies. GAIN domain expression in insoluble pellets is likely to have resulted from improper receptor processing, preventing trafficking to the cell membrane.

In an attempt to circumvent this issue and target CD97 for secretion, Gibson assembly PCR was used to incorporate the honeybee melittin leader peptide upstream of the CD97-GAIN EGF 1,2,5 gene in the pFastBac vector. The HBM leader sequence facilitates receptor secretion, thus aiding purification of the target protein, and has been previously used to markedly increase the yield of secreted protein (182). Whilst cloning CD97-GAIN EGF 1,2,5, the GPR114 GAIN domain was also

cloned into a pFastBac vector with the HBM leader sequence. Though GPR114 is not a member of the EGF-TM7 subfamily, the receptor mRNA has been found to be specifically expressed by various immune cells like the EGF-TM7s, including human eosinophils, and mouse monocytes, macrophages, lymphocytes and dendritic cells (79).

Despite inclusion of a secretion sequence, both CD97-GAIN EGF 1,2,5 and GPR114 remained insoluble. As protein aggregates arise from the production of proteins in structurally altered states, the fact that N-glycosylation is known to be hindered in insect cells may be a contributing factor to the insolubility of CD97 in this case (201). The requirement of post-translational N-glycosylation of the receptor N-terminus has previously been demonstrated as an important factor in proteolysis at the GPS, and is therefore essential for receptor trafficking to the membrane (202). Additionally, protein expression is driven by the polyhedrin promoter in the baculovirus expression system, and optimum expression levels are reached close to the time of cell death due to lytic virus infection. Consequently, post-translational machinery and the secretory pathway are likely to be compromised, reducing the folding and trafficking of proteins directed to the cell membrane (203).

The amplification and titering of viral stocks, and subsequent expression optimisation is a lengthy process taking a minimum of one month, before needing to reamplify, re-titre and optimise the MOI of a fresh viral stock. Furthermore, substantial loss of viral titre after freezing also greatly limits the “shelf-life” of viral stocks. Therefore, due to inefficiency and time constraints, the baculovirus system was not taken further for expression and purification of CD97-GAIN EGF 1,2,5, or GPR114.

Bacterial *E. coli* expression systems remain the most popular choice for large-scale expression of recombinant protein for structural studies due to the ease of culture manipulation, low cost, and rapid production of high yield protein. However, there are various drawbacks. The reduced ability to form disulphide bonds due to the reducing environment of the cytosol and the inability to perform post-translational modifications such as glycosylation often results of the production of heterologous protein as insoluble clumps. These inclusion bodies consist of aggregated polypeptide chains and partially folded intermediates (193) which are commonly observed as a result of overexpression in *E. coli* (204).

CD97-GAIN was cloned into the pET28a vector in the absence of EGF repeats and transformed into several strains (Table 3.4). Despite BL21-CodonPlus (DE3) being the typical strain for expression of proteins under control of the T7 promotor,

expression trials, both small and scaled, indicated insoluble expression, which was also seen by expression in improved efficiency strain BL21-Gold (DE3). These strains were selected for expression as they deficient in *omp*T and *lon* proteases and thus reduce the likelihood of proteolytic degradation and are engineered to contain extra gene copies encoding rare *E. coli* tRNAs *argU, ileY*, and *leuW* that are found in abundance in mammalian cells. The difference in tRNA levels is reflective of codon usage of the particular organism, and if the usage differs greatly between the heterologous protein and host proteins, decreased mRNA stability, early termination of transcription, mutations and inhibition of growth are commonly experienced (205). Furthermore, codon bias has been shown to impact on protein folding therefore increasing the chances of aggregate formation (206).

Although bacterial cultures were induced during log-phase and incubated for extended periods of time at low temperature to enable slow growth and hence protein production, very little bacterial pellet was harvested upon centrifugation. The small cell pellet indicated a lack of *E. coli* growth and replication, suggesting a level of toxicity was experienced either before, or immediately upon induction of CD97-GAIN, resulting in cell death. To try and offset the toxicity, recombinant pET28a was subsequently transformed into competent Rosetta (DE3) pLysS, a derivative strain of BL21. In addition to the features already described for BL21 (DE3), Rosetta pLysS are engineered to express T7 lysozyme which binds to and suppresses basal expression of T7 RNA polymerase pre-induction, therefore stabilising expression of the potentially toxic heterologous protein of interest. Moreover, inclusion of the pLysS plasmid in *E. coli* has greatly increased expression of a number of target proteins (207)(208). Expression of CD97-GAIN however, was unsuccessful with no protein evident in either the soluble or insoluble fraction.

Continued BL-21(DE3) expression trials yielded consistently insoluble protein. Attempts to purify CD97 from insect inclusion bodies were unsuccessful, possibly because the use of non-ionic detergent NP-40 was not harsh enough to solubilise the aggregated pellets. Furthermore, the harsh conditions required to solubilise inclusion bodies results in the need for protein refolding. However, in the case of the GAIN domain, as such treatment would irreversibly disrupt the non-covalent bonds that mediate association of the ECD with the *Stachel* peptide and transmembrane domain, rendering the GAIN domain incomplete. To prevent the expression of the target protein in IBs, research indicates that expressing the target with a fusion protein can increase its solubility (209). Highly soluble *E. coli* thermostable proteins such as trigger factor, thioredoxin and glutathione S-transferase have all been proven

to increase the level of solubility of the target protein due to their high expression and stability (210). Maltose-binding protein (MBP) is common fusion protein that lacks cysteine residues and therefore is stable regardless of the presence of reducing agents (211). Reporter gene GFP has also been implicated in improving target solubility with relatively high yield when expressed in the same BL21(DE3)-pET system used in this study (212). Martin Stacey's lab have previously tried to improve solubility by expression with SUMO fusion protein and MBP, however these attempts were unsuccessful.

An additional hurdle commonly encountered whilst expressing mammalian proteins in prokaryotic systems is the formation of disulphide bonds. Although bacteria are capable of producing disulphide-bonded proteins, the enzymes that mediate bond formation are located within the periplasmic space (213)(214). As protein processing occurs within the reducing environment of the bacterial cytoplasm, the oxidation required for the formation of disulphide bonds is prevented (204)(215). Adhesion GPCR GAIN domains contain two disulphide bonds that stabilise the native structure and thus in their absence, unfolded and unstructured polypeptide chains are likely to be produced, which are prone to aggregation and IB formation (216), evidenced by insoluble expression of CD97-GAIN in BL21 (DE3) strains. To overcome the formation of insoluble protein, CD97-GAIN pET28a was transformed into SHuffle® T7 *E. coli.* SHuffle® cells have been engineered to have diminished cytoplasmic reductive pathways by mutating cytoplasmic enzymes thioredoxin reductase and glutathione reductase, whilst the periplasmic disulphide bond isomerase *dsbC* is overexpressed within the cytoplasm via deletion of its native signal peptide (217). Multi-disulphide bonded proteins have been successfully expressed to high yield by these cells (218)(219). Soluble expression of CD97-GAIN by SHuffle® T7 was gained from initial trials, albeit at a low yield. It has been suggested that auto-induced cultures yield a much higher quantity of heterologous proteins than standard IPTG induction (220), however there was very little CD97-GAIN expression detected when this method was employed (data not shown). Induction temperature has previously been shown to be protein-specific (217), therefore expression was induced at 16 °C, 25 °C and 37 °C. Cells incubated at 16 °C for 68 hours produced the highest yield of soluble protein. After optimisation of induction conditions, expression was scaled up to 10 litres and subsequently purified for use in crystal trials.

The stability of the resulting protein was relatively low, with substantial formation of precipitate when stored at 4 °C. One group have demonstrated increased protein stability during protein concentration by the addition of amino acids Arg and Glu, in

millimolar quantities without interfering with protein-protein interactions or protein structure. Furthermore, inclusion of these amino acids in buffers increased long term stability of proteins during storage, an attractive quality whilst carrying out crystal trials, whereby different preparations of the same protein can yield quite different quality crystals (221).

CD97-GAIN identity was verified by native (non-ionising) mass spectrometry, in which proteins that associate non-covalently are maintained in their native folded state prior to ionisation (222)(223), and therefore both cleavage and subunit association could be confirmed. The major species detected indicated a mass of 32611 Da, the expected size of the entire GAIN domain, suggesting correct processing and folding of the protein had occurred. This was crucial in order to investigate the importance of the alpha helical region and specific residues in both the cleavage and processing of the conserved GAIN domain.

3.4.2 Crystallographic investigation of CD97-GAIN

Crystallisation of proteins occurs in a two-phase process; screening and optimisation. Firstly, chemical physical conditions are ascertained under which the protein of interest has the potential to form singular crystals, and these conditions are subsequently refined to increase the likelihood of gaining diffraction quality crystals (224). CD97-GAIN was subject to crystal trials after successful purification. Multi-factorial and sparse-matrix screening was carried out to identify crystallisation conditions. Due to the vast array of chemicals previously shown to crystallise protein and difficulties in predicting those that will yield crystals, screening as such allows for a broad overview of numerous combinations and conditions that would be completely unfeasible to test individually (225)(226). Identification of even one or two crystalline outcomes provides a starting point for further optimisation.

Screening identified PEG1500 as the most favourable precipitating agent, at pH 4-5 mediated by MIB buffer from Molecular Dimensions, with small yet consistent formation of crystalline spindles. Optimisation around these conditions however, did not improve their size or morphology. Although the shape of a crystal does not necessarily determine its ability to diffract well and obtain a diffraction data set of high quality (227), needle crystals are often too poor quality to collect x-ray diffraction data due to their width and size, meaning they easily become damaged by radiation.

Various approaches can be taken to improve crystal quality. Easy non-covalent modifications can be made to crystal-yielding conditions, such as changing temperature and pH, including various additives in the mother liquor, and the inclusion of protein binding partners, due to being most simplistic and cost-effective, and the existing protein sample can still be used. These parameters have all immensely aided crystal formation and structural determination (172)(177). Despite their notoriety for being an extremely challenging target, the structure of the GPCR β2AR was solved by exploiting antibodies as stabilising binding proteins (228).

More drastic methods can also be used to improve experimental data, although crystal formation may end up quite different than those obtained from the original protein sample due to such structural changes. Protein surface mutagenesis of loops that are large flexible regions to smaller loops, and mutation of surface residues to those that entropically favour and promote crystallisation, have been successful in improving structural resolution by up to 1Å (229)(230).

Despite the thin, spindly nature of the CD97-GAIN crystals, a native data set of 2.2 Å resolution was obtained. However multiple molecular replacement searches did not generate a solution to enable the structure determination for the CD97 GAIN domain. The most likely explanation for the lack of success was perhaps due to the low sequence homology between the CD97 GAIN domain and the search models used. Molecular replacement requires approximately 50% of the model structure to cover the structure to be solved, whilst sharing ≤35% sequence identity. Sequence similarity of no less than 20% may result in relatively good fold conservation between the two proteins, but generally the differences are too large to be solved by molecular replacement (231)(232). CD97-GAIN shares approximately 25-30% homology with the solved Lat1/BAI3 structures and is therefore close to the lower limit of the homology threshold, meaning obtaining a solution may not be possible (232). Trials to explore this were conducting using the Lat1 and BAI3 crystal structures and their respective experimentally observed amplitudes. Using Lat1 as the search model with the BAI3 experimentally observed amplitudes as the target molecule, it was possible to show that successive truncation of the Lat1 search PDB model eventually changed from a very clear MR solution, to no solution. Similarly, the use of BAI3 as the search model with Lat1 experimentally observed amplitudes was unsuccessful. The successive truncation of the PDB model represented the high homology between Lat1 and BAI3 to the low homology between Lat1/BAI3 and CD97 GAIN domain.

Other approaches can be used to obtain data for phasing if molecular replacement is unsuccessful or if the data obtained is of low resolution. Methods for obtaining phase information experimentally include the attachment of heavy atoms to the protein, and the incorporation of selenium in place of sulphur within methionine residues. The latter is a commonly tried and tested option, which benefits from being able to identify the location of the residue from the known sequence (233). Although it is easily achieved by *E. coli* expression system, by replacing methionine in growth media with selenomethionine, the results of protein expression can vary therefore is not always a simple solution (234). Heavy atom soaking assists in gaining a better diffraction pattern; X-rays are diffracted by electrons, and therefore atoms that have a higher atomic number have more electrons, which creates a noticeable change to the diffraction pattern in comparison to that gained with the protein alone (235). Soaking allows heavy atoms to interact with proteins via covalent or electrostatic interactions, forming an ordered arrangement in the asymmetric unit. However, although many of the first crystal structures were solved using this method, the trial and error approach requires multiple crystals for soaking, and multiple data collections, and is therefore laborious and time consuming (236). Poor expression of CD97-GAIN severely hindered further optimisation of crystal growth conditions, and subsequent attempts at structural determination by solving the phase problem were ineffective.

The published atomic structures of Lat 1 and BAI3 detailed for the first time that the GPS cleavage motif and the previously termed “stalk” region, were in fact an integral part of the wider GAIN domain, comprising the last 5 β-strands of the structure (26). The GAIN domain itself is composed of 8 α-helices and 13 β-strands; the first 6 helices make up subdomain A, whilst the β-strands and two small α-helices equate to subdomain B. The cleavage product remains firmly intact with the rest of the domain, facilitated by a network of hydrogen bonds and largely hydrophobic side chain interactions. Interestingly, the structure demonstrated that BAI3 had not undergone autoproteolytic cleavage yet was still successfully expressed. There were no obvious structural differences as a result of the lack of cleavage, except a slight tilt in the positioning of the final β-strand and a distorted bond at the autocatalytic site (26).

More recently, the structure of GPR56 GAIN domain associated with a previously undefined N’ terminal domain was elucidated to, in complex with a fibronectin type III scaffold monobody (178)(237). The GAIN domain, however, was smaller than that of Lat 1 and BAI3, with only 3 of the 6 α-helices in subdomain A present. It is possible

that the presence of N-terminal protein domains acts to maintain the structural integrity of the GAIN domain. Indeed, the broader structure of GPR56 was seemingly stabilised by the presence of an interdomain disulphide bond (178), although a free-cysteine is not available to mediate such an interaction in CD97 for example. On the other hand, myeloid expressed GPR114 completely lacks N-terminal protein domains except the GAIN, indicating that they are not always necessary for receptor function. However, subdomain B was highly conserved across the 3 crystallised GAIN domains, strengthening the notion that this domain is crucial to receptor activity.

In the absence of a crystal structure for the CD97 GAIN domain, it is not clear for certain how many α-helices are present in the protein. Secondary structure prediction based on the amino acid sequence of CD97 GAIN domain, using programs including I-TASSER (238), HHpred (239) and JPred4 (240) indicated the presence of 3 α-helices only, lacking the most N-terminal in the solved GAIN domain structure. The difference in the number of α-helices and the relative orientation of the helices between CD97 GAIN domain and Lat1, BAI3 and GPR56 likely contributes to the lack of success in obtaining a molecular replacement solution. However, the truncation of these helices in the search models would not facilitate a clear solution.

Despite the publication of the 3D structure of these GAIN domains, clarification of an example EGF-TM7 GAIN domain remains an important target. A lack of the majority of subdomain A in the EGF-TM7 GAIN domain is a point of interest as to whether the structural conservation of subdomain B is maintained in the absence of A. Moreover, the fact the EGF-TM7 subfamily are functional in the absence of subdomain A, suggests that only subdomain B is necessary for Adhesion GPCR activity. Indeed, a number of studies have investigated the signalling potential of truncation mutants, and demonstrated upregulated activity and defined the tethered agonistic, or *Stachel*, peptide (50)(54)(57). Therefore, further structural characterisation of this region is required if we are to fully understand the activation mechanism of these receptors, and how this domain may interact with ligands and the transmembrane domain. An endogenous ligand of CD97, CD90/Thy1, binds to the GAIN domain itself rather than with the N-terminal EGF repeats (110), but is not shown to stimulate G-protein signalling. Instead, a firm adhesive interaction is established, possibly indicating that ligand binding in this instance stabilises the GAIN domain, preventing either a conformational change or liberation of the *Stachel* to mediate receptor activation.

Future work should also aim to characterise the structural differences, if any, of known disease-mutant GAIN domains. Indeed, both GPR56 and BAI3 have known

disease-causing mutations at the first cysteine residue responsible for formation of one of the two disulphide bonds that seemingly stabilise the cleaved product within the wider GAIN domain (26). Moreover, whilst the mutation of cysteine to tyrosine at residue 492 of the EMR2 GAIN domain is not a disulphide bond-forming residue, it does reduce domain stability, resulting in upregulated dissociation of the wider GAIN domain from the final β-strand and transmembrane domain, stimulating increased mast cell degranulation (46). Understanding the potential structural rearrangements and implications of such mutations will provide invaluable insight required for design of pharmacological therapeutics.

Chapter 4. Interrogating adhesion GPCR signalling pathways and their activation with surrogate ligands

4.1 Introduction

A primary focus of the adhesion GPCR field has been the mode of activation and subsequent signalling properties of these enigmatic receptors. It has now been well established that adhesion GPCRs have a tethered agonist within the GAIN domain immediately downstream of the autocatalytic cleavage site (50). However, the lack of ligands and heterotrimeric G-protein coupling, hence effective signalling readouts, for many adhesion GPCRs has greatly hampered their functional characterisation (33)(241). Due to their involvement in a variety of disease states, such as BFPP, cancer progression, male infertility and Usher's syndrome, fathoming the mode of signalling for adhesion GPCRs is imperative for the future progression of therapeutic intervention.

Therefore, using GPR56 as a well characterised comparator, this chapter sought to dissect the potential G-protein coupling of the orphan adhesion GPCR GPR97, and generate small protein scaffolds known as Affimers to act as surrogate ligands with a view to aid the characterisation of GPR97. The establishment of robust signalling readouts allowed for the interrogation of mechanically induced signalling and the testing of rare naturally occurring variants of GPR56 and GPR97, identified from a cohort of consanguineous individuals.

4.2 Validation of luciferase reporter assay output

As the majority of GPCR mediated signalling pathways culminate in the activation of a limited number of transcription factors (Figure 4.1), GPCR activation and signalling can be interrogated via the use of transcription reporter assays. Using CD97 to validate the system, a luciferase activity strategy was therefore established to elucidate the potential signalling pathway of GPR97.

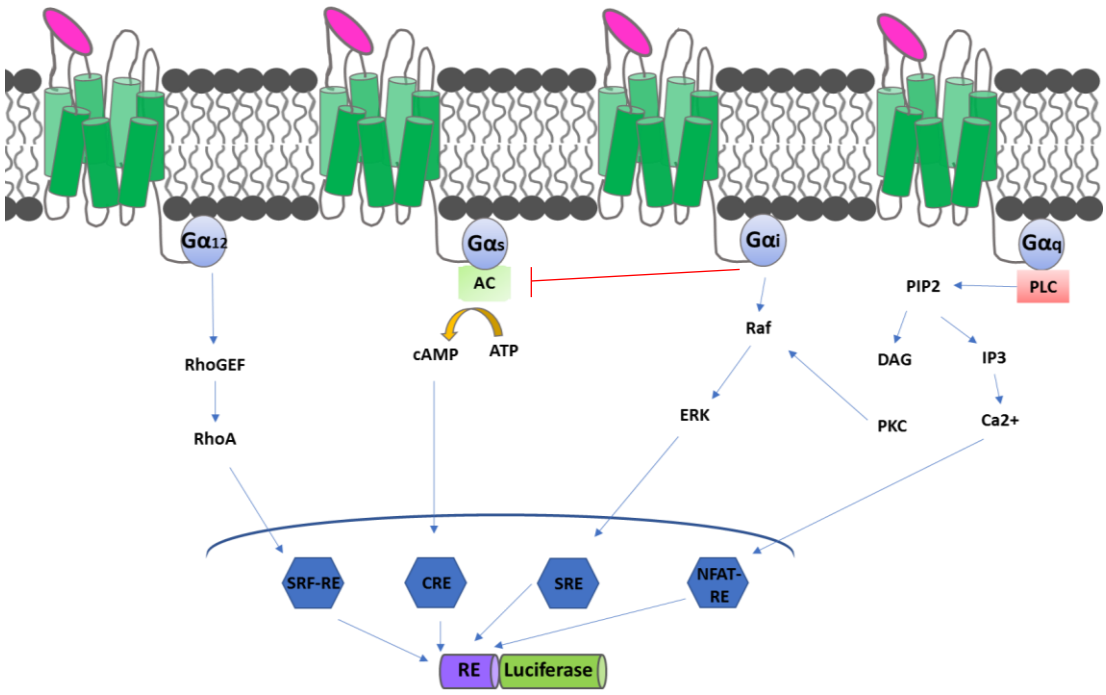


Figure 4.1 ***Schematic of GPCR-stimulated luciferase reporter assays*****.** Different Gα subunits stimulate different downstream effectors. In turn, pathway-specific response elements (RE) are activated to drive transcription of luciferase gene. SRF-RE; serum response factor response element, CRE; cAMP response element, SRE; serum response element, NFAT-RE; nuclear factor of activated T-cells response element. (Adapted from Z. Cheng *et.al* 2009)(242)**.**

Previous studies have demonstrated that CD97 signals through $G\alpha_{12/13}$ to stimulate increased levels of Rho-GTP, and heterodimerises with the lysophosphatidic acid receptor 1 (LPAR1) to further enhance Rho via LPA-dependent mechanisms (67). Rho family GTPases regulate the activity of serum response factor (SRF) and its binding to its response element (RE). CD97 has also previously been shown to signal via MAP kinase pathway in promoting gastric cancer proliferation (243). Activation of

this pathway can be measured via the activity of the serum response element (SRE). Therefore, to test the luciferase assay set up and ultimately further explore the ability of adhesion GPCRs to interact with G-proteins via surrogate ligand-dependent mechanisms, the signalling output of CD97 was assessed using pSRF-RE and pSRE. Expression of a firefly luciferase reporter gene was driven in response to both SRF-RE and SRE pathway activation. SRF-RE is a mutant form of SRE lacking the ternary complex factor (TCF) binding domain which would otherwise facilitate interaction with the MAP-kinase pathway. SFR-RE is designed to respond to SRF-dependent and TCF-independent pathways, like Rho activation (244).

The Dual-Glo® luciferase assay system (Promega) was employed whereby both firefly and *Renilla* luciferase reporter genes are co-transfected in a single experimental sample. The *Renilla* luciferase plasmid (pRL-TK) acts as an internal control to normalise data for transfection efficiency. As 293T cells express an almost full complement of G-protein subunits, these cells were chosen for investigating adhesion GPCR signalling (245). Cells were seeded in 96-well plates and grown to 80% confluency, before transfection with one of the following receptor constructs; full length CD97 EGF-1,2,5 (CD97 FL), CD97 EGF-1,2,5 truncated to have only the first transmembrane helix (CD97 TM1), or CD97 truncated to the natural cleavage site and therefore lacking the NTF (CD97 ΔNTF). TM1 receptors are typically unable to stimulate G protein-mediated signalling due to lacking the transmembrane helices required to interact with intracellular G proteins. Adhesion GPCR ΔNTF constructs on the other hand, are confirmed by a number of reports to exhibit an enhanced signalling response in comparison to the full length receptor and are constitutively active (50)(54).

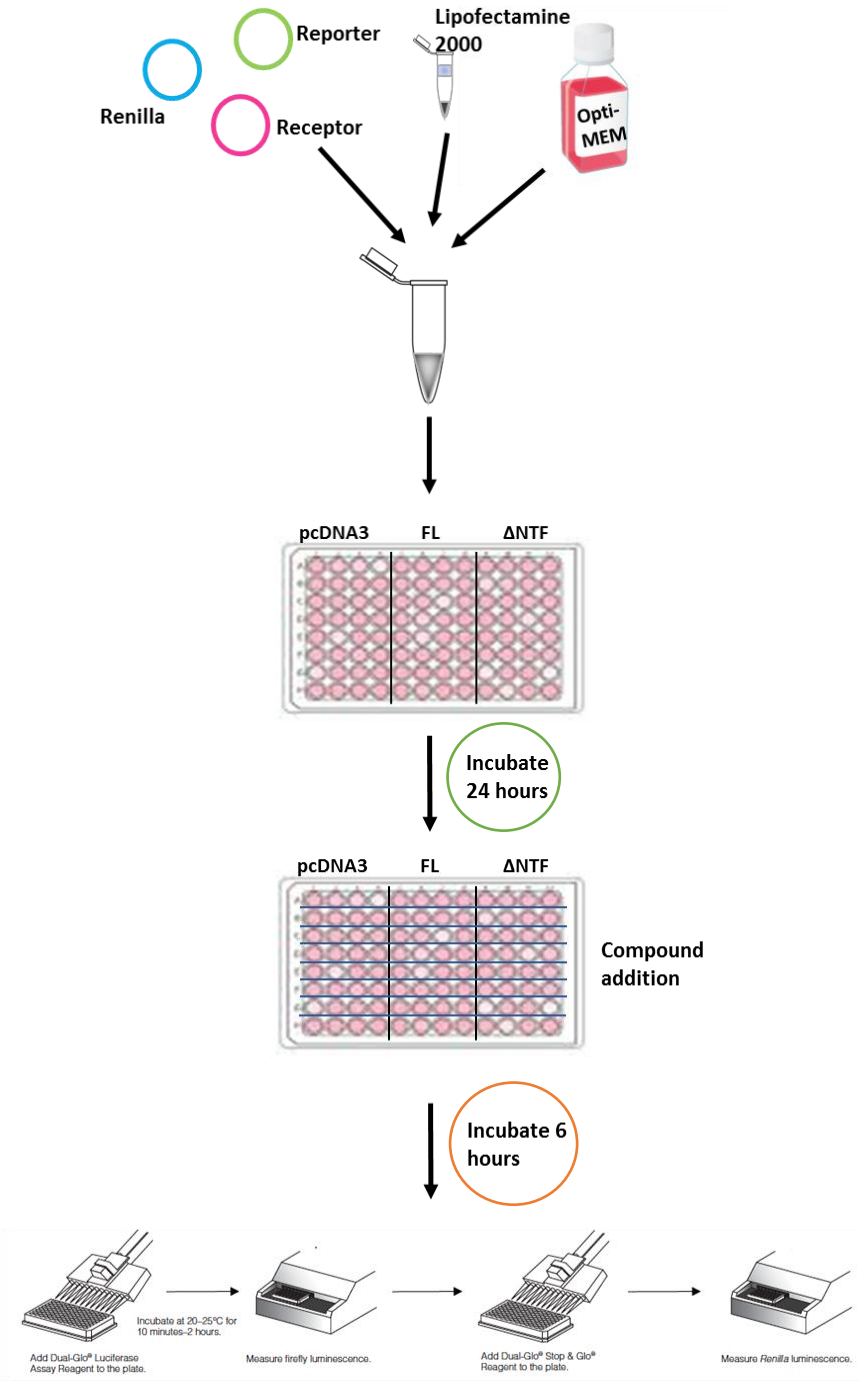


Figure 4.2 ***Schematic representation of luciferase assay set up.*** Assays were carried out in 96-well plates, divided into 3 sections for each receptor construct. Compound addition was carried out across the plate. Figure adapted from Dual-Glo (Promega) user manual.

24 hours post-transfection cells were incubated with PBS or 10% serum for 6 hours, a time point previously determined by the manufacturer to be optimum for luciferase expression. All conditions were tested in quadruple. SRF-RE was not activated in cells transfected with CD97 TM1, except in those incubated with 10% serum which induced low level signalling activity (Figure 4.3). The basal activity of CD97 FL seen when stimulated with PBS was not significantly increased upon addition of 10% serum. An approximate 80-fold increase in activity was exhibited by CD97ΔNTF in comparison to that of CD97 FL, confirming that truncation of the receptor to the natural cleavage point does induce increased constitutive activity. Furthermore, this robust signal alongside the basal activity exhibited by CD97 FL indicates that CD97 couples with $G\alpha_{12/13}$ to activate SRF-RE-driven firefly luciferase expression.

Despite unpublished data previously indicating strong activation of SRE (Aust, Liebscher), induction of the SRE pathway in our assay was much less robust than the SRF-RE pathway. Constitutively active CD97ΔNTF-driven firefly luciferase expression of SRE was lower than that of the full length CD97 stimulated with PBS for SRF-RE. Moreover, SRE-stimulated activity by CD97 FL was comparative to that by CD97 TM1 (Figure 4.3). Therefore, the SRE pathway was not considered for further study.

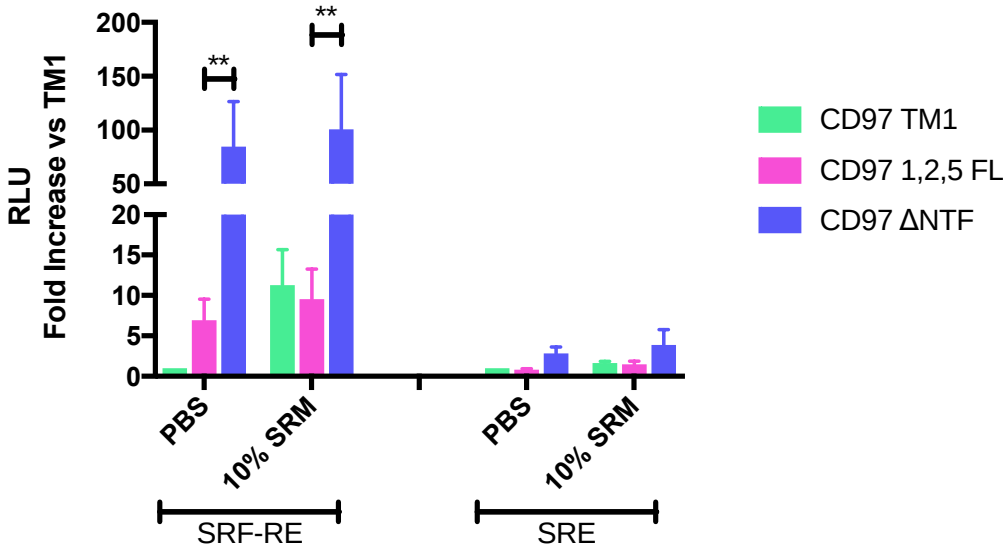


Figure 4.3 ***Confirmation of CD97-dependent SRF-RE and SRE activation.*** HEK293Ts were transiently transfected with different CD97 constructs; signalling deficient TM1, constitutively active ΔNTF, and full length 1,2,5 isoform. An equal concentration of reporter to receptor was also transfected alongside pRL-TK at one eighth of the concentration of the reporter plasmid. 24 hours post-transfection, cells were stimulated with PBS, or 10% serum as a positive control. Cells were incubated for a further 6 hours before luminescent output was determined. The ratio of experimental reporter to control reporter luciferase was calculated, and subsequent relative response rations were calculated. Averaged data for each experimental condition was normalised to the average read for CD97 TM1 stimulated with PBS. RLU; relative light units. Significance analysed by two-way AVONA. Mean data ± SE. N=3.

After successfully establishing set up of SRF-RE luciferase assay, Affimers that were previously designed and confirmed by ELISA to bind specifically to CD97 GAIN domain (Section 3.3.4.3) for crystal trials were exploited as surrogate ligands. Although CD97 has numerous known binding partners including CD55, chondroitin sulphate, and several integrins, downstream signalling has not been associated with ligand binding. Therefore, an alternative approach is required to further characterise both the activation and signalling properties of CD97.

As done previously, cells were stimulated 24 hours post-transfection with control compounds or Affimers at a final concentration of 5 µg/ml. CD97 signalling was not induced by Affimer binding (Figure 4.4). There was no significant change in the level of detected luciferase expression for cells expressing full-length CD97 in comparison

to those stimulated with buffer. Similarly, although signalling activity appeared to be slightly upregulated for CD97ΔNTF when stimulated with many Affimers, differences were not significant. No signalling output was detected for CD97 TM1, reflecting that of buffer stimulated cells.

Due to a lack of up- or downregulation of full length receptor signalling, further verification of Affimer specificity for CD97 GAIN domain was undertaken. The binding capability of each small protein was examined by FACS analysis and was detected by a primary α-polyhistidine and α-mouse Alexafluor 488® (ThermoFischer Scientific) secondary antibody. Binding was confirmed to be negative for all Affimers (data not shown), contrasting the results gained by ELISA. As a result, CD97 Affimers were not further utilised in signalling assays.

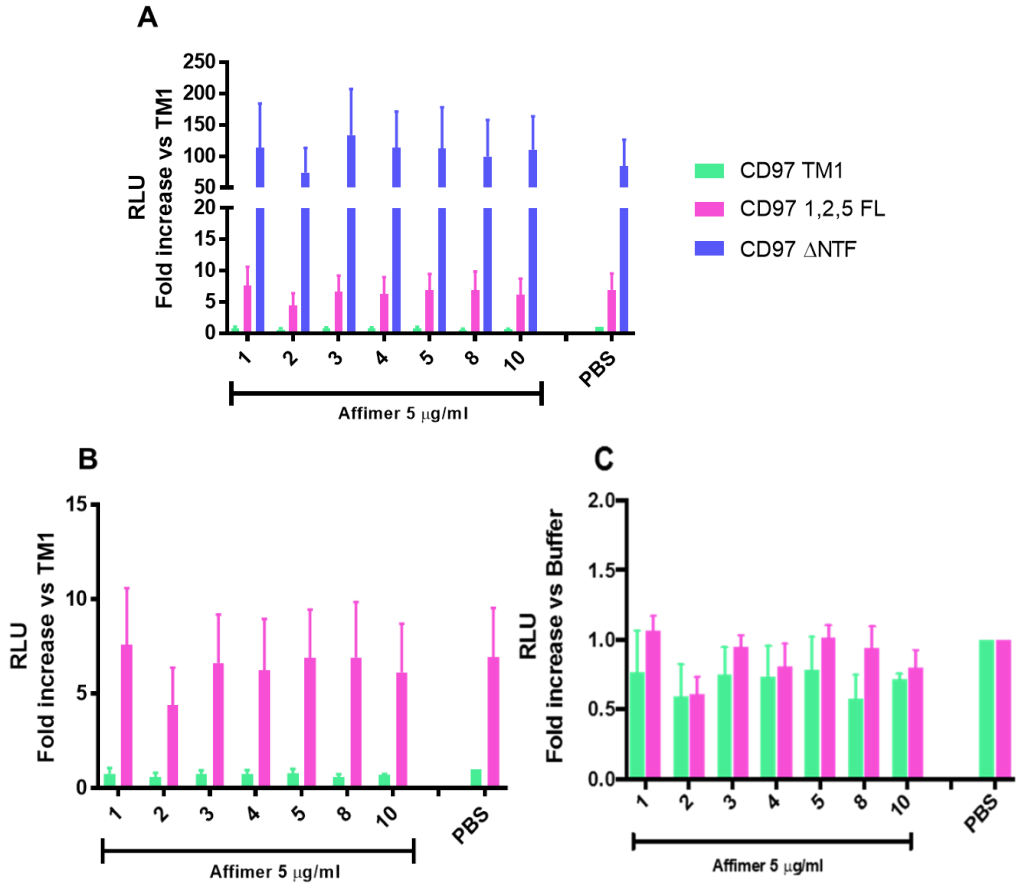


Figure 4.4 ***Affimer-induced CD97 SRF-RE signalling.*** Cells were stimulated with 5 µg/ml Affimer, and luciferase expression was detected after 6 hours. (**A+B**) Data normalised to buffer stimulated CD97 TM1, or the respective receptor (**C**). Affimers had no significant effect on CD97 full length signalling activity. Data is mean ± SE. N=3.

4.3 Signalling activity of GPR97 and GPR56

GPR97 is expressed by a range of immune cells and is implicated in macrophage polarisation in adipose tissue and B-lymphocyte maintenance and development in splenic follicles (112)(113). However, it remains an orphan receptor as it is yet to have a defined binding partner. Chimeric G-protein screening assays have previously established GPR97 coupling with $G\alpha_o$ but not to others (68).

Similarly to CD97, GPR56 has been shown to couple with $G\alpha_{12/13}$ and mediate Rho-dependent transcription to regulate neuronal cell migration (105). GPR97 influences lymphatic remodelling and endothelial cell migration, and when knocked out, alters RhoA activity (246). However, the specific mechanism by which this occurs has not yet been determined. Both receptors were therefore investigated for their ability to activate SRF-RE luciferase expression, with the aim of determining $G\alpha_{12/13}$ coupling for GPR97. GAIN domain-specific Affimers were also generated with a view to exploit the small binding proteins as surrogate ligands. During the course of this research studies confirmed nanobody-dependent modulation of GPR56 signalling (58), therefore the alternative technology offered by Affimers was explored as to whether they could mediate a similar effect. SRF-RE assays were established as with the CD97 signalling stated above. Cells were transfected with full-length receptors, constitutively active ΔNTF constructs, and empty vector pcDNA3 as a signalling deficient control.

Using GPR56 as a control for $G\alpha_{12/13}$ coupling, Figure 4.5 demonstrates GPR97 coupling to $G\alpha_{12/13}$ to activate the SRF-RE signalling pathway. Each full-length receptor demonstrated basal signalling activity, determined by readout after addition of PBS to the growth medium, with a near 15-fold and 14-fold increase compared to that of empty vector pcDNA3 for GPR56 and GPR97, respectively. Unexpectedly, both GPR56 and GPR97 constitutively active ΔNTF constructs exhibited a lower level of signalling than their full-length counterparts, contrasting existing data for GPR56, which indicates much higher SRF-RE activation by the ΔNTF construct (54). The luciferase output of cells transfected with pcDNA3 was increased upon addition of 10% serum indicating successful transfection of SRF-RE reporter. Excitingly, this data demonstrates for the first time GPR97 coupling to $G\alpha_{12/13}$ and confirms SRF-RE luciferase assay to be a suitable method to measure signalling activity of both GPR56 and GPR97.

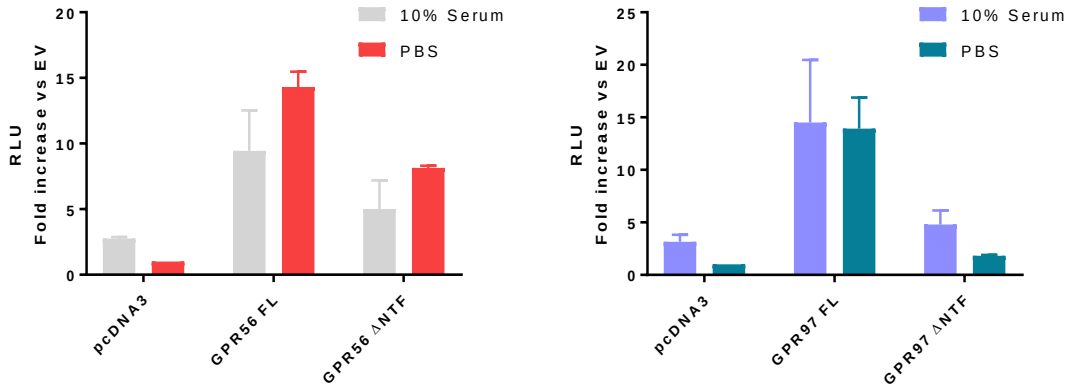


Figure 4.5 ***GPR56 and GPR97-dependent SRF-RE activation*****.** SRF-luciferase activity measured in empty vector, GPR56 full length and GPR56 ΔNTF transfected cells 30 hours post-transfection. Data shown is mean ±SE. N=3

4.3.1 Cloning, expression and purification of receptor-specific Affimers

Having successfully demonstrated $G\alpha_{12/13}$ coupling by both GPR97 and GPR56, ligand induced signalling was tested using the same luciferase assay. Affimers were raised against the GAIN domains of GPR56 and GPR97, expressed in a mammalian system. HEK293T cells were transfected in T75 flasks, as outlined in Section 2.3.2, and incubated for 72 hours. Culture supernatant was harvested, and cell debris was removed by centrifugation at low speed, purified by Ni^{2+} affinity chromatography as previously stated. Binding proteins were identified and amplified as stated in Section 3.3.4.1. The binding specificity of a pool of Affimers was confirmed by ELISA for binding to GPR56 and GPR97, and negative control Fc-EMR3 (Figure 4.6). Five Affimers displaying receptor-specific selectivity were identified and sequenced (Figure 4.7A). Each was subsequently cloned into a pET11c vector by Q5 polymerase standard PCR, with a 6xHis-tag incorporated at the C-terminus and then transformed into DH5α *E. coli.* Affimers were expressed in BL21 (DE3) *E. coli* and purified as previously outlined (Section 3.3.4.2). Purified products were resolved by SDS-PAGE, which confirmed soluble and stable expression at the expected size of 12kDa (Figure 4.7B). Fractions were pooled and concentrated before being dialysed in PBS and 10% glycerol and finally verified by mass spectrometry.

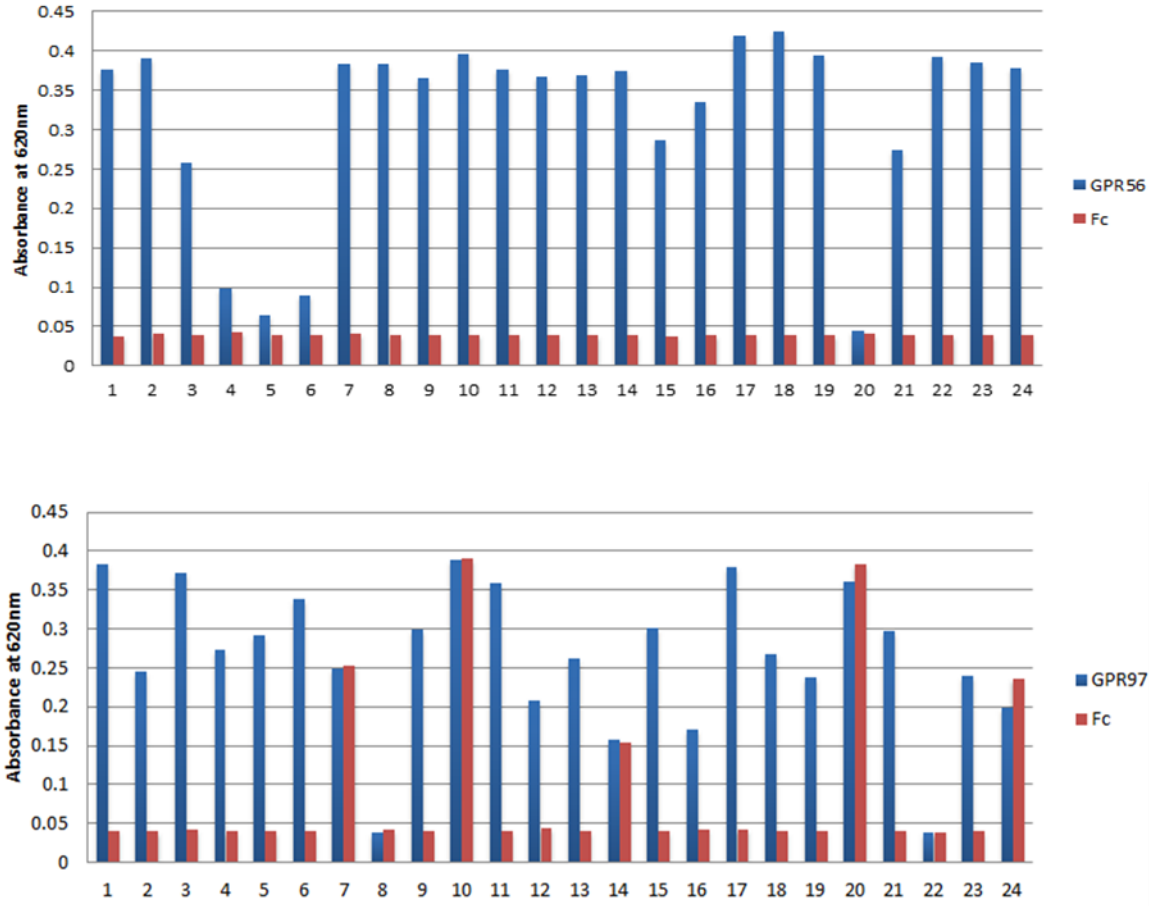


Figure 4.6 ***Identification of receptor-specific Affimers*****.** 96-well plates were coated at 200ng per well with one of GPR97, GPR56 or Fc-EMR3, and Affimers serially diluted across the plate. HRP-conjugated anti-His was used to detect Affimer-receptor binding. Absorbance was measured and recorded relative to non-specific binding of Fc-EMR3.

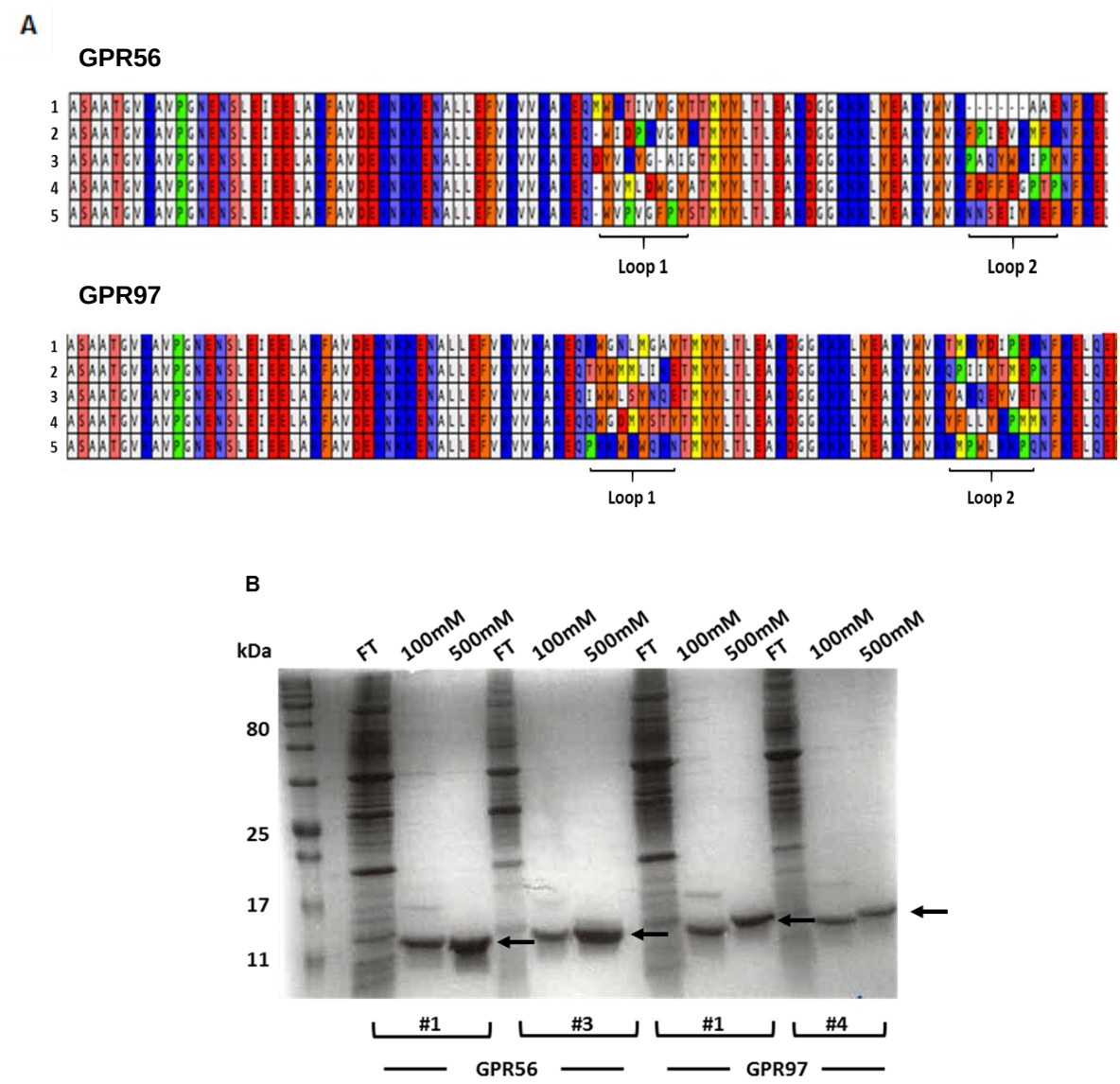


Figure 4.7 ***Sequence alignment and purification of GPR56- and GPR97-specific Affimers.*** **(A)** Sequence alignment of Affimers positive for binding to GPR56/GPR97 GAIN domain, showing variable sequence regions 1 and 2. **(B)** Representative examples of Affimers; expression and purification samples were resolved by 15% SDS-PAGE electrophoresis. Arrows indicate Affimers in purification fractions.

4.3.2 Verification of Affimer binding

To validate Affimer specificity and thus confirm their suitability for interrogating receptor activation, their binding capabilities to the GAIN domains of GPR56 and GPR97 were investigated by flow cytometry. HEK 293Ts were transfected at 80% confluency with full length GPR56 and GPR97. GPR97 was expressed with a C-terminal GFP-tag. After 48 hours, cells were incubated with 10 µg/ml Affimer. Affimer binding was detected by labelling with primary α-polyhistidine and secondary α-mouse Alexafluor 647 (ThermoFischer Scientific), alongside appropriate controls for transfection and Affimer specificity. Cell-only controls were included to determine viable cell population.

As detailed in Figure 4.8, cells positive for GPR56 expression were detected with α-GPR56 clone CG4 monoclonal antibody, which bound to 9.79% of the gated single cell population. This population was not detected for GPR97-transfected cells. Affimers 1 and 3 were deemed to be positive for GPR56 GAIN domain, binding to 12.34% and 5.34% respectively of the GPR56 expressing cells. Cell populations were observed in the same quadrant as those indicative of receptor-antibody binding, with Affimer 1 binding greater than that of the anit-GPR56 antibody. Binding of Affimers 2, 4 and 5 was detected at a similar percentage to that of the non-specific GPR97 Affimer, and therefore were deemed negative for GPR56 interaction.

As the mammalian GPR97 construct used contained a GFP tag, GPR97 expression was inferred via GFP fluorescence. GFP expression was indicated by a population shift to the right for cells transfected with GPR97 (Figure 4.9). Affimers were deemed to be positive for GAIN domain binding if a double positive population was present (Figure 4.9, quadrant B). GFP expression is indicated at >25% for all populations transfected with GPR97, indicative of a low expression. Binding of Affimers 1 and 4 was observed on cells with higher GFP expression. The percentages displayed in quadrants in Figure 4.9 represent those of the total cell population. However, of the GFP positive total population, Affimers 1 and 4 bound to 15.25% and 9.43% respectively. Affimers 2, 3 and 5 bound to 1.4% or less, reflecting that of GPR56-Affimer binding to GPR97, and were therefore deemed negative for GPR97-specificity. These results for both GPR56 and GPR97 show for the first time Affimers are applicable for the binding of cells surface protein via FACS analysis.

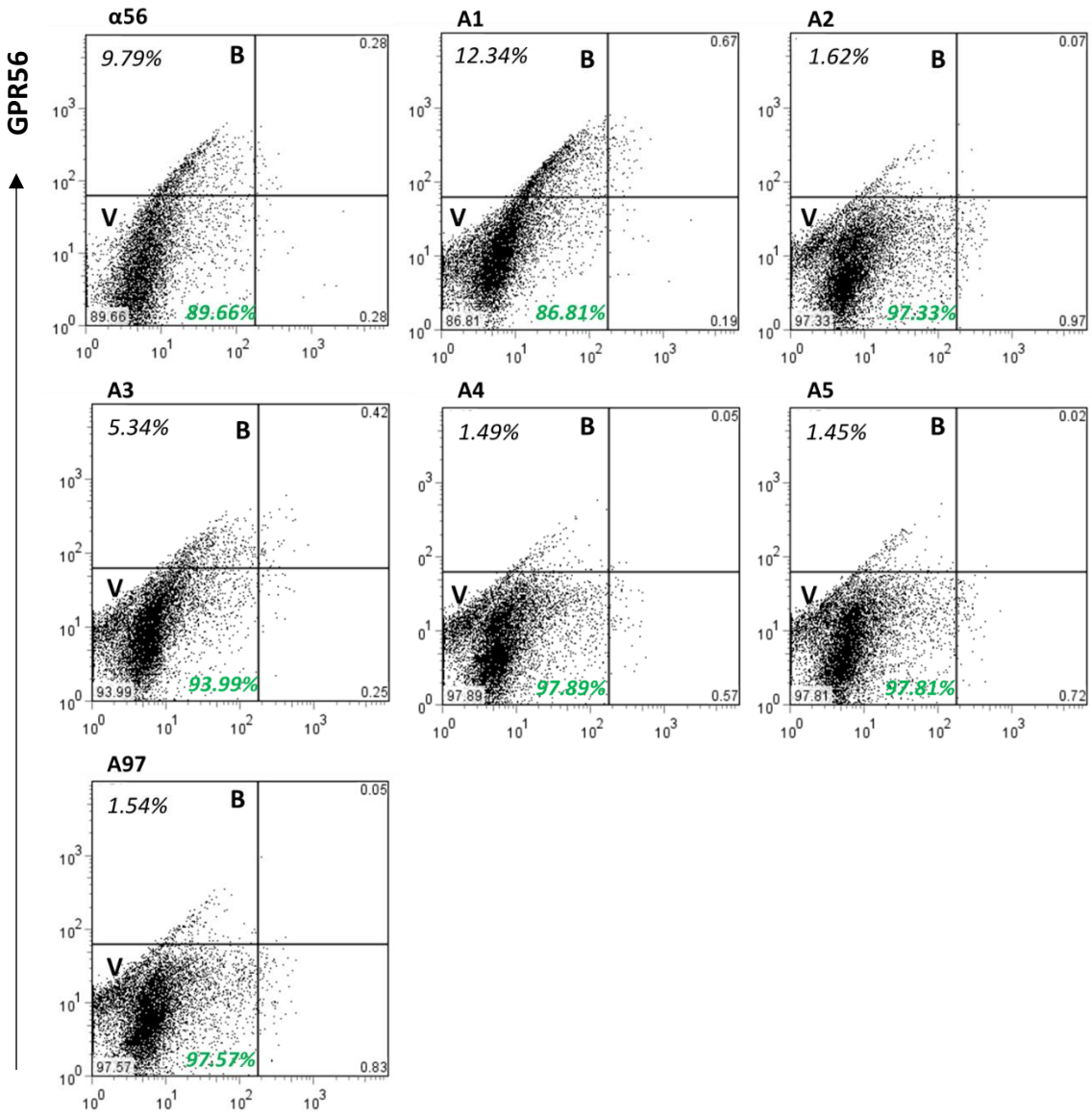


Figure 4.8 ***Confirmation of α-GAIN Affimers binding to surface expressed GPR56.*** 293Ts expressing surface GPR56 were analysed by flow cytometry for binding of anti-GPR56 monoclonal antibody (**α56**), receptor-specific Affimers (**A1-5**) and GPR97-specific Affimer (**A97**). Binding was detected on single cell gated populations using an anti-mouse Alexafluor 647 secondary antibody. **V** indicates viable cell population, **B** indicates bound cell population. Percentages for each population are indicated in the respective quadrant.

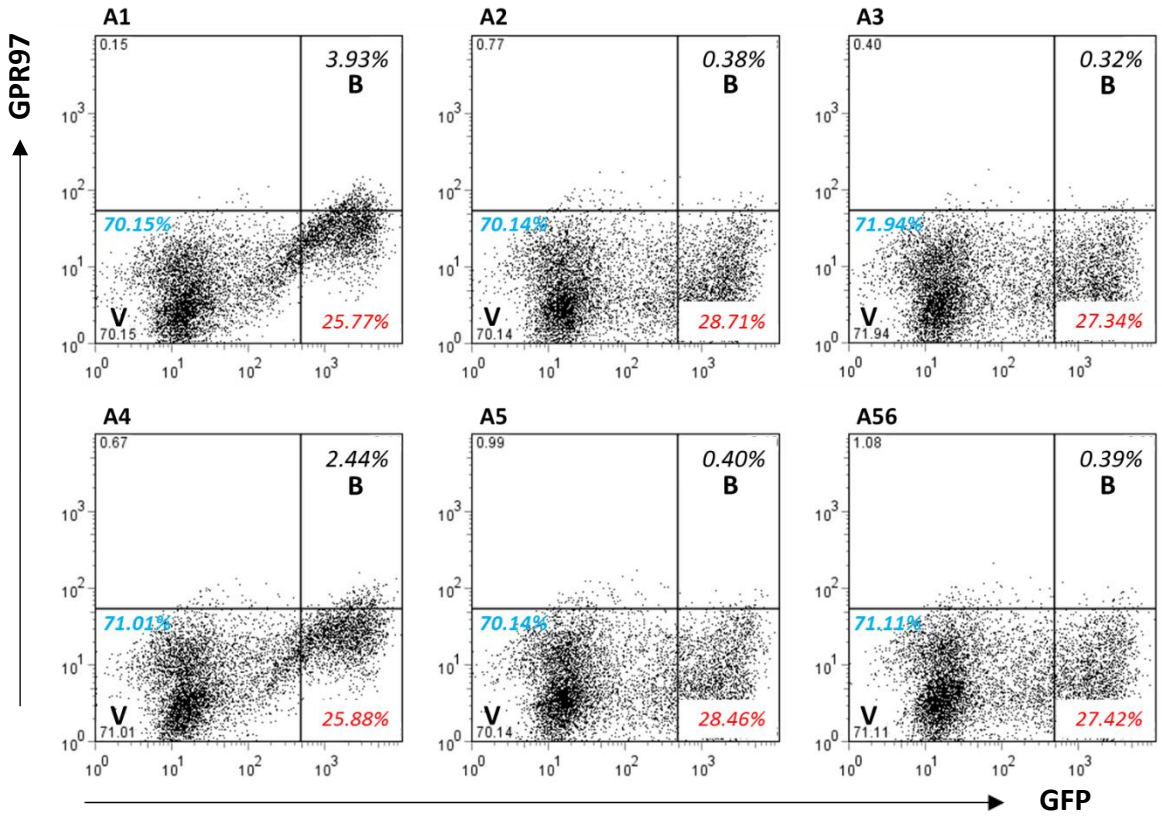


Figure 4.9 ***Confirmation of α-GAIN Affimers binding to GPR97-GFP expressing cells.*** 293T cells expressing GPR97-GFP were incubated with GPR97-specific Affimers (**A1-A5**), or a GPR56-specific Affimer (**A56**). Affimer binding was detected on single cell gated populations using an anti-mouse Alexafluor 647 secondary antibody. **V** indicates viable population, **B** indicates Affimer-bound population. Percentages of each population are displayed in the quadrant.

4.3.3 Investigation of GPR56 and GPR97-activating Affimers

Following confirmation of Affimer binding specificity, their capacity to modulate receptor function was investigated. Alongside PBS and 10% serum, GPR56 Affimers 1 and 3, and GPR97 Affimers 1 and 4 were added to cells to stimulate receptors at a final concentration of 5 µg/ml. Plates were incubated for 6 hours before luciferase expression was detected. Each assay condition was tested in triplicate, and the average calculated before normalising for transfection, and further normalising to empty vector (Figure 4.10A+C) and to PBS stimulation (Figure 4.10B+D).

No significant effect was mediated by binding of Affimers to both GPR56- and GPR97 ΔNTF (Figure 4.10A+C), confirming the binding site to be in the GAIN domain which

lies N-terminal to the cleavage site. Affimer 1 was shown to downregulate GPR56 FL signalling. When data was normalised to buffer stimulation, a similar response was also apparent in empty vector transfected cells when stimulated with Affimer 1 (Figure 4.10B). However, this decrease was not significant. These results were mirrored in the signalling activity of GPR97 FL and pcDNA3 incubation with Affimer 4, which also caused downregulated SRF-RE activity, though was not found to be statistically significant (Figure 4.10D).

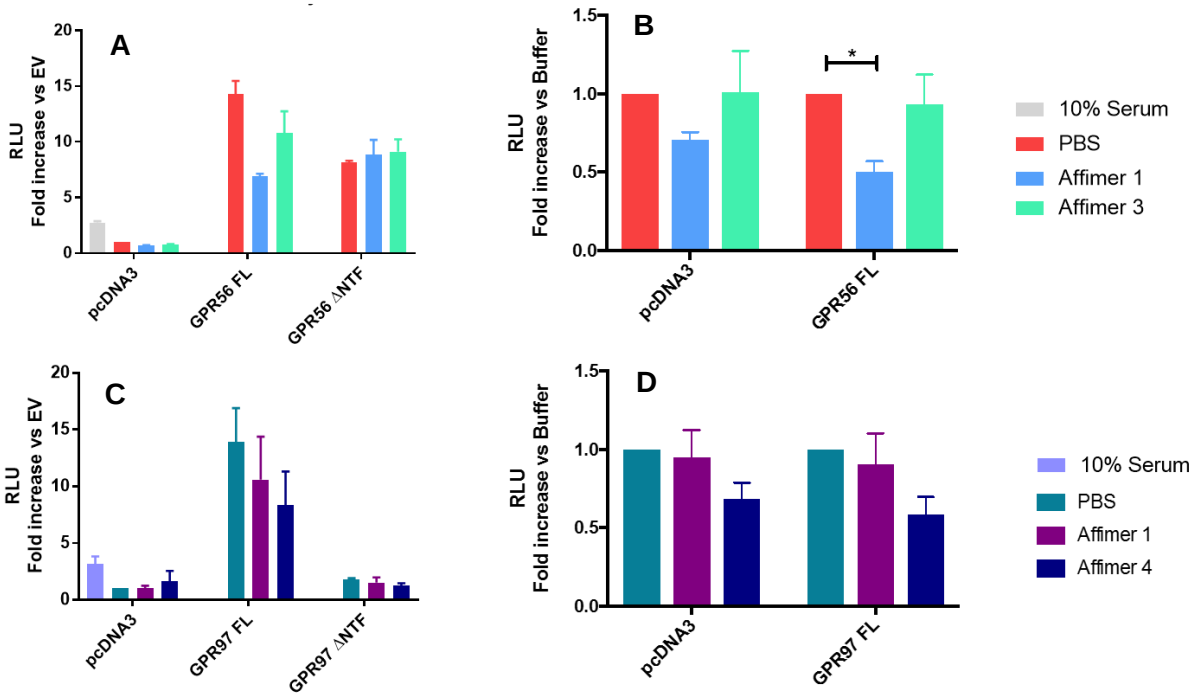


Figure 4.10 ***GPR97 and GPR56 activate SRF-RE signalling.*** SRF-luciferase activity measured in empty vector, GPR56 full length and GPR56 ΔNTF transfected cells (**A-B)** and empty vector, GPR97 full length and GPR97 ΔNTF transfected cells (**C+D**). **A+C** show fold increase of SRF activity relative to pcDNA3, **B+D** indicates the effect of Affimers normalised to PBS stimulation. Data shown is mean ±SE. A two-way ANOVA was performed to determine statistical significance of differences in signalling with each stimulus. *, $p < 0.05$. N=3.

4.3.4 Adhesion GPCR signalling under mechanical stress

With increasing evidence towards adhesion GPCRs acting as mechanosensory receptors, it has been hypothesised that mechanical force/stress likely plays a role in their activation. To investigate the effects of ligand-GAIN domain interaction under mechanical stimulation, SRF-RE luciferase assays were performed and subject to shaking at 1000rpm for 5 minutes. The signalling activity of CD97, GPR56 and GPR97 was tested in this manner. Before stimulation of GPR56 and GPR97, Affimers were incubated with nickel sepharose beads (GE Healthcare) for protein capture and immobilisation via the C-terminal 6xHisTag. These were subsequently added to cells for 30 minutes, acting to mimic receptor binding to an immobilised endogenous ligand. CD97 was not incubated with Affimers due to their previously determined lack of binding specificity. Plates were further incubated for 6 hours after vibration before the luciferase signal was read. The speed and intensity of mechanical shaking was replicative of that capable of inducing mast cell degranulation upon EMR2 subunit separation (46). Duplicate assays were also performed without vibration.

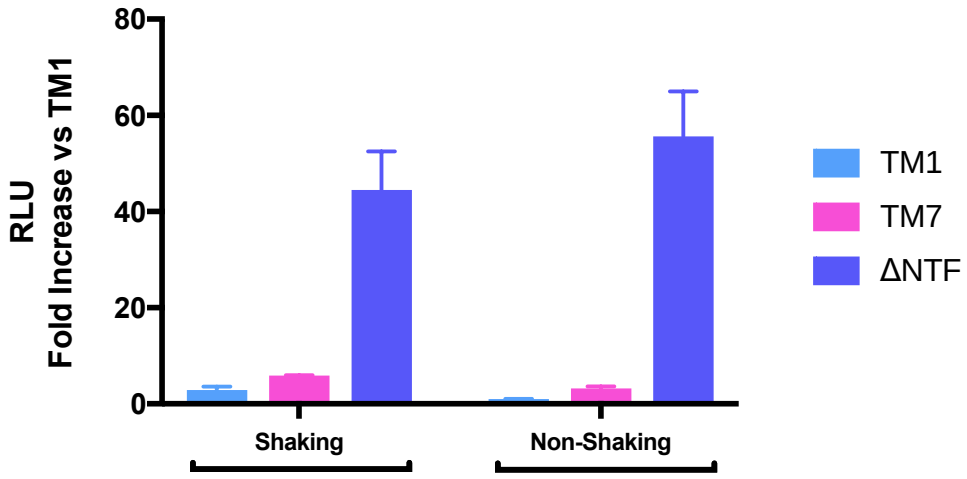


Figure 4.11 ***Vibration induced signalling of CD97.*** SRF-RE activity was detected after HEK 293Ts expressing full length, TM1 and ΔNTF CD97 were subject to vibration for 5 minutes at 1000 rpm. Data shows the comparison of SRF- stimulation in the absence and presence of mechanical stimulation. RLU; relative light units. All data is indicative of fold increase over buffer stimulated TM1 signalling. Data plotted represents mean 3 biological repeats ± SE

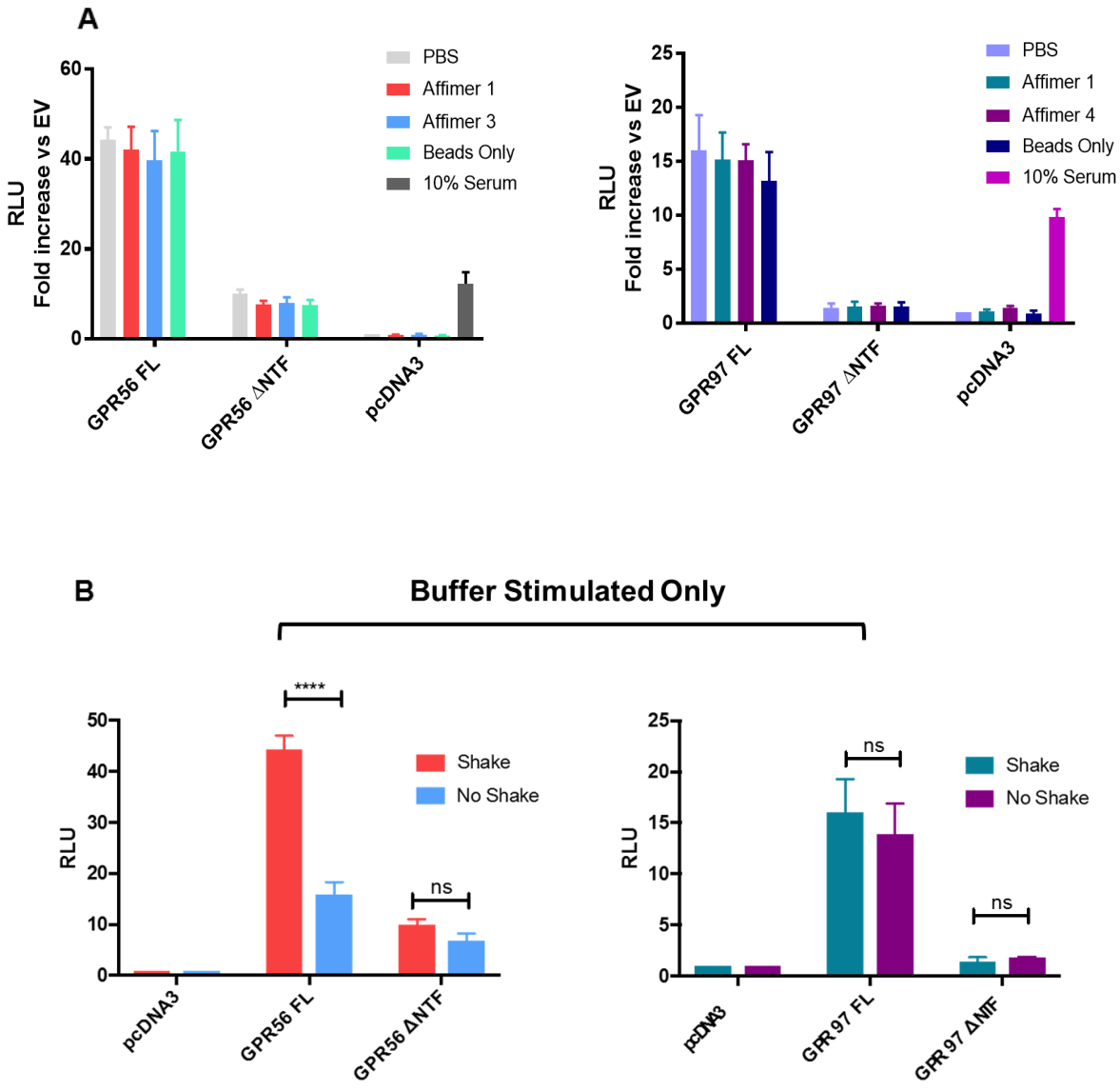


Figure 4.12 ***Analysis of vibration-induced signalling of GPR97 and GPR56.*** SRF-RE luciferase assays of full length and ΔNTF GPR56, and full length and ΔNTF GPR97 transfected cells. (**A**) Cells were vibrated for 5 minutes at 1000 rpm. Luciferase signal was read 6 hours post stimulation. Cells were stimulated with receptor-specific Affimers or nickel sepharose beads alone. Affimers were added at 5 µg/ml. (**B**) Comparison of SRF- stimulation in the absence and presence of mechanical stimulation. RLU; relative light units. All data is indicative of fold increase over buffer stimulated empty vector signalling. Data plotted represents mean 3 biological repeats ± SE. A two-way ANOVA was performed to determine statistical significance of receptor response with and without vibration. ****, p <0.0001. ns, not significant. N=3.

The addition of Affimers to pcDNA3-transfected cells did not stimulate SRF-RE activity in comparison to buffer stimulated cells. Signalling also remained unchanged in cells expressing the ΔNTF construct for both GPR56 and GPR97 in response to Affimer binding under mechanical stress (Figure 4.12A). 10% serum induced a SRF response in EV-transfected cells, again confirming successful transfection and expression of the reporter.

Independent of Affimer binding, vibration alone caused increased SRF activity in cells expressing full length GPR56. A ~40-fold increase in luciferase signal compared to that of EV, and a 3-fold increase compared to full length GPR56 under static conditions was observed (Figure 4.12B). However, this effect was not observed with GPR56 ΔNTF, or indeed full length or ΔNTF GPR97. Moreover, serum-induced SRF activity was not increased upon shaking. Taken together, this implies the response is dependent on vibration of GPR56 and its extracellular domain, and not an experimental artefact (Figure 4.12B). A bead-only control was included to ensure their addition would not induce non-specific signalling activity, confirmed by a low luciferase signal from by cells transfected with pcDNA3 (Figure 4.12A). Interestingly, there is no significant decrease in signalling mediated by Affimer 1 binding to full-length GPR56, unlike binding under static conditions demonstrated in Figure 4.11. Here, SRF-RE activity is returned to almost the same the level exhibited by PBS-stimulated cells, suggesting the potential downregulatory influence of Affimer 1 on signalling is alleviated under vibration.

4.3.5 Functional assessment of naturally occurring mutants

Epidemiological studies enable the identification of disease risk factors and quantification of their significance. Indeed, the use of disease cohort studies and “phenotype to genotype” approaches have led to the discovery of various disease-causing mutations in adhesion GPCRs. These include identifying a previously unknown variant of EMR2 in patients with vibratory urticaria (46), GPR56 in BFPP (247), and VLGR1 in Usher’s syndrome (248). However, with the advent of large genomic datasets, “genotype to phenotype” strategies are now potentially viable for dissecting the function of adhesion GPCRs.

The Born in Bradford birth cohort study was established in 2007 and recruited 12,500 pregnant women over a period of 3 years. The lives of their children have been tracked to link patients’ longitudinal questionnaire data and electronic health records with individual exome sequencing data and genomic SNP-chip analysis. Moreover,

the high levels of consanguinity within this cohort mean there is an increased likelihood of the occurrence of rare deleterious homozygous genetic variants. The robust reporter assays described previously permit interrogation of the function of naturally occurring variants from the Born in Bradford study or other cohorts.

Whole genome sequencing data from chip array analysis (Illumina) was performed for the cohort and was screened for common single nucleotide polymorphisms (SNPs). Variants occurring in >5% of the general population were removed, and the remaining candidate SNPs were subject to SIFT analysis, predicting whether an amino acid substitution would affect protein function (249). Homozygous variants for both GPR56 and GPR97 were identified in the Born in Bradford cohort, each found to be within the third transmembrane helix (TM3).

As outlined, the SRF-RE luciferase assays provide a suitable method to assess the signalling activity of these adhesion GPCRs. Therefore, for preliminary experiments into the functional implications of these SNPs, GPR56*M487T* and GPR97*G437S* were generated by site directed mutagenesis and transfected in 96-well plate format alongside the wild-type constructs. Receptors were stimulated with only PBS or 10% serum to determine whether the variants caused up- or downregulation of signalling, or had no effect at all, in comparison to wild-type receptor activity.

Figure 4.13 reveals significant upregulation of SRF-RE activity in both GPR56*wt* and GPR56*M487T* transfected cells in comparison to the empty vector control. However, there is no significant difference between wild type and variant signalling output. Conversely, GPR97*G437S* shows almost completely abolished signalling, with a 10-fold decrease in activity compared to that of GPR97*wt*. To explore whether this was a *bona fide* defect in receptor signalling or an effect of poor receptor expression, further analysis was performed to investigate surface expression of each receptor.

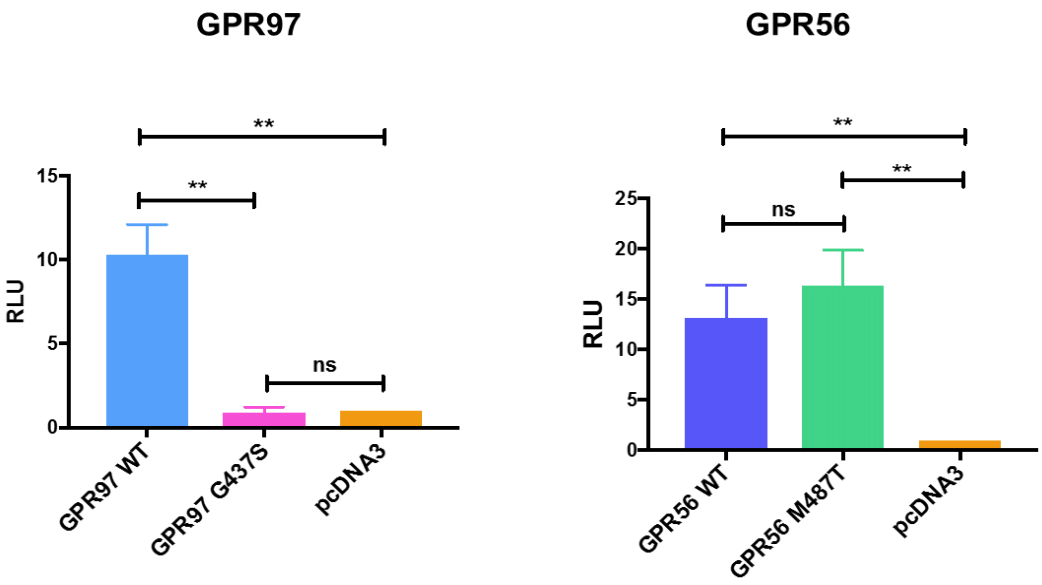


Figure 4.13 ***Signalling activity of adhesion GPCR variants identified from the Born in Bradford cohort***. HEK293T were transfected at 90% confluency with GPR97*wt*, GPR97*G437S*, GPR56*wt* or GPR56*M492T*. Cells were buffer stimulated and luciferase expression detected after 6 hours. Data represents fold increase over empty vector. Data show are mean ± SE. Two-way ANOVA analysis was performed to compare the significant of signalling response. **, $p < 0.01$. ns, not significant. N=3.

Mutations in adhesion GPCRs Lat1 (250) and GPR56 (45) have been shown to influence protein stability and reduce trafficking to the cell surface, consequently altering receptor activity. To investigate the cell surface expression of the variants identified in the Born in Bradford cohort, immunofluorescent and flow cytometric analysis was performed. 48 hours post-transfection, 293T cells were incubated with a receptor-specific primary monoclonal antibody, provided by Dr Hsi-Hsien Lin and binding was detected by anti-mouse Alexafluor 647 secondary antibody (ThermoFischer Scientific). Cells were also counterstained with DAPI. As illustrated in Figure 4.14, GPR56*wt* and GPR56*M487T* were visualised at the plasma cell membrane after staining with primary anti-GPR56 and secondary anti-mouse Alexafluor 647, confirming successful transfection and expression of each receptor. This finding was further supported by FACS analysis (Figure 4.15) which confirmed near equal surface expression of both GPR56*wt* and GPR56*M487T*, evidenced by the large population shift when compared to unstained GPR56-transfected cells. These results indicate the mutation within TM3 does not alter receptor stability and subsequent trafficking to the cell surface and therefore provides validation of the signalling activity of each receptor detected in SRF-RE assays (Figure 4.13).

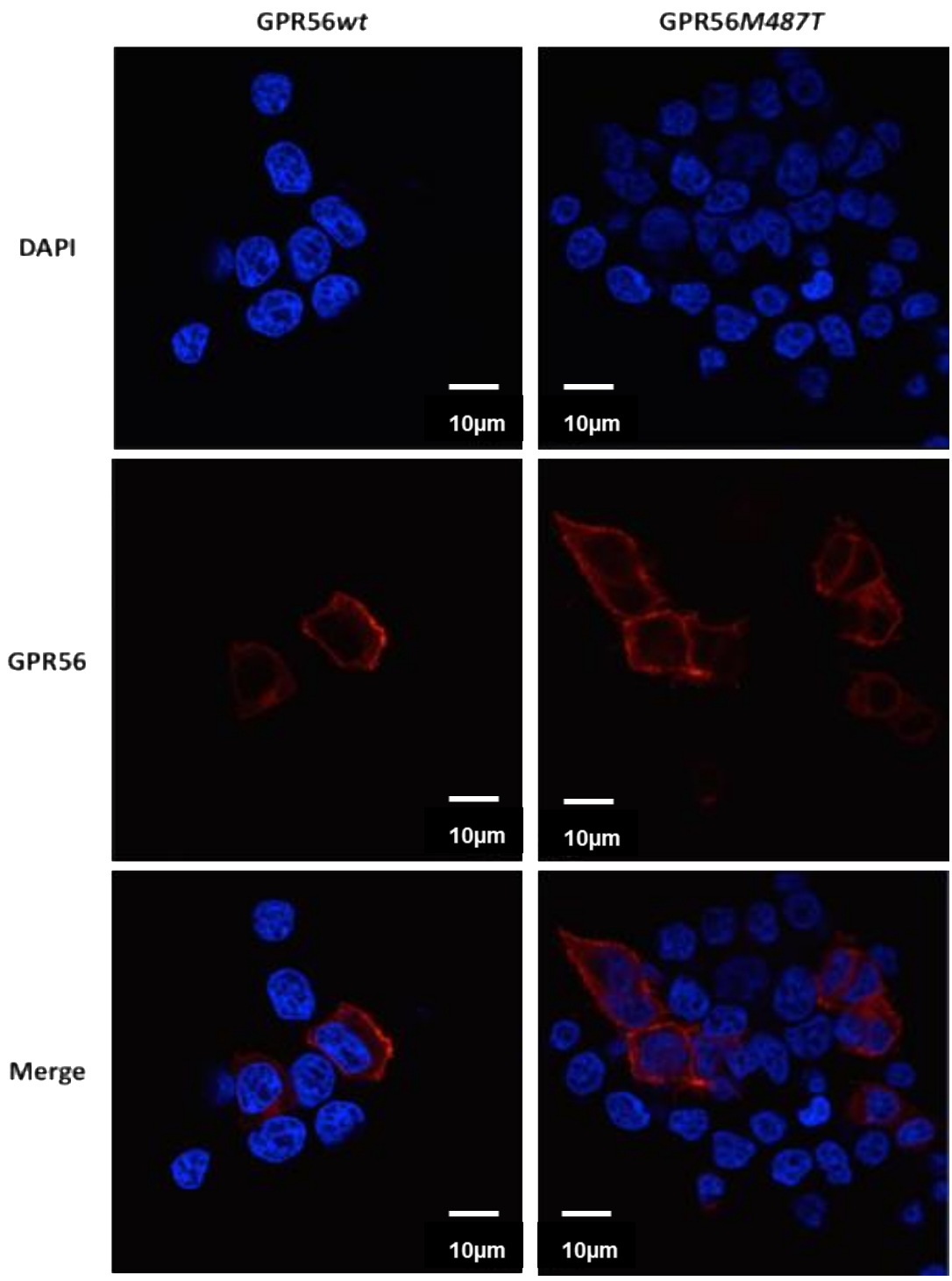


Figure 4.14 ***Visualisation of GPR56M487T expression at the cell membrane.*** HEK293Ts were transfected at 90% confluency with GPR56*wt* or GPR56*M487T*. After 24 hours, cells were fixed in 4% paraformaldehyde and staining with primary αGPR56 and AlexaFluor 647 conjugated anti-mouse secondary. Cells were counterstained with DAPI**.** Slides were imaged by an inverted LSM700 confocal microscope coupled to a LSM Image Browser.

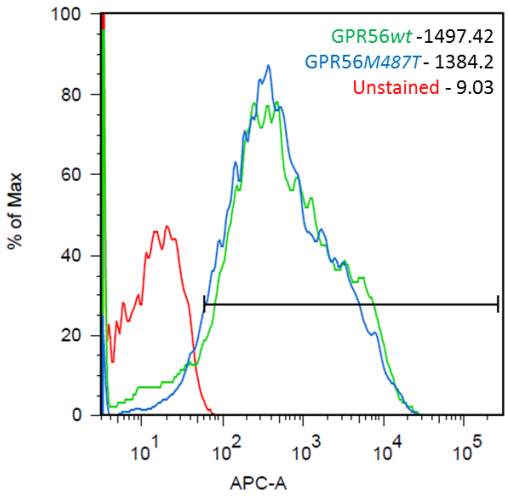


Figure 4.15 A *single nucleotide polymorphism within GPR56 TM3 does not alter surface expression*. 293Ts transfected with wild type or variant GPR56 were incubated with primary αGPR56, which was detected by AlexaFluor 647 (APC) anti-mouse secondary. Mean value for APC detection is stated.

Immunofluorescent detection of GPR97*wt* and GPR97*G437S* using primary anti-GPR97 and AlexaFluor 647 conjugated α-mouse secondary was poor. However, visualisation of the C-terminal GFP tag confirmed low level expression of each receptor, demonstrated in Figure 4.16. From observing GFP localisation, it appears that GPR97 is expressed within the cytoplasm. Flow cytometry verified receptor expression by GFP detection, signified by an upward population shift in comparison with the untransfected cell sample. The mean value of detected GFP for wild type GPR97 was ~10-fold greater than that of GPR97*G437S* (Figure 4.17). Conversely, receptor expression was not detected by primary αGPR97 and AlexaFluor 647 conjugated secondary. This agrees with the immunofluorescence results indicating that GPR97 is not being expressed at the cell surface. Furthermore, this also supports the lack of SRF-RE activation by GPR97*G437S* (Figure 4.13). However, GPR97*wt* signalling was found to be significant in comparison to both variant GPR97 and empty vector transfected cells, consistent with the large difference in expression between the wild type and variant (Figure 4.17). Taken together, it is possible that the SNP within GPR97 TM3 results in receptor instability and hence is not properly expressed and trafficked to the plasma membrane. These results do indicate expression of the wild type receptor, which therefore suggests that the αGPR97 antibody being used is not suitable for application in flow cytometry or immunofluorescent microscopy.

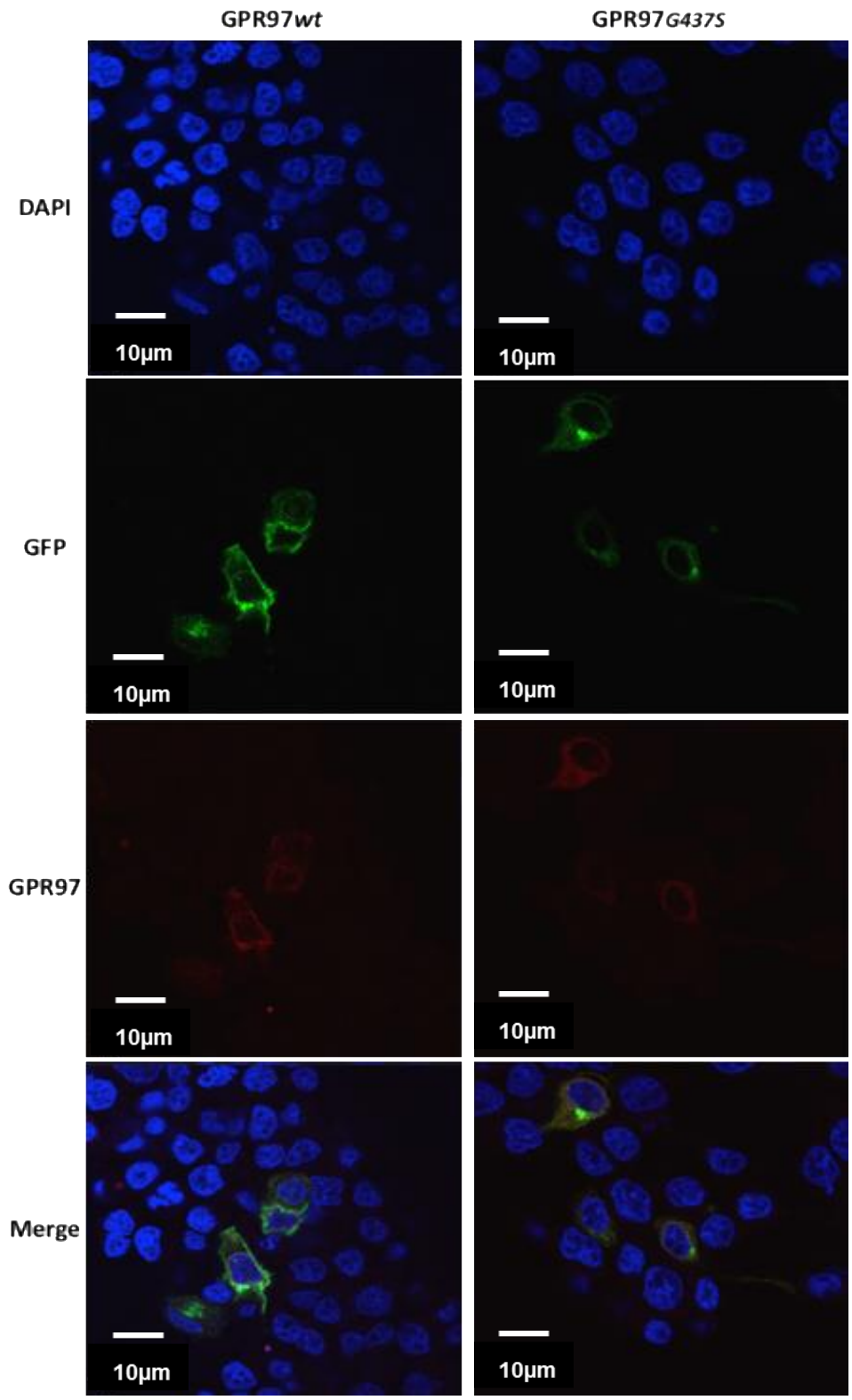


Figure 4.16 ***Detection of GPR97G437S expression in HEK293Ts.*** Cells were fixed in 4% paraformaldehyde 24 hours after transfection and stained with primary αGPR97 and AlexaFluor™ 647 (ThermoFischer Scientific) conjugated anti-mouse secondary. Cells were counterstained with DAPI and slides were imaged by an inverted LSM700 confocal microscope coupled to a LSM Image Browser.

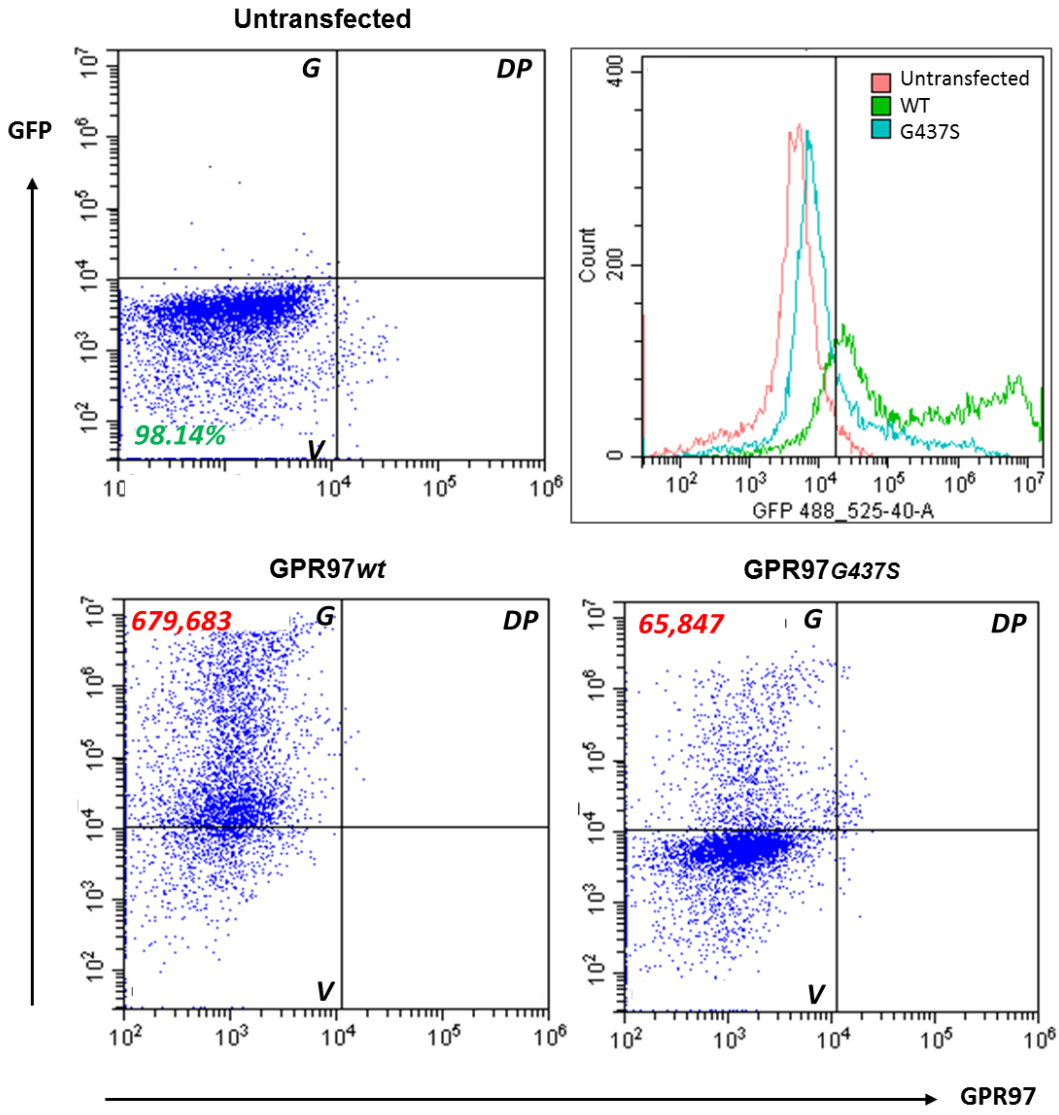


Figure 4.17 *GPR97G437S expression is reduced in comparison to GPR97wt*. 293Ts transfected with wild type or variant GPR97 were incubated with primary αGPR97, and AlexaFluor 647 anti-mouse secondary. Both scatter plot and histogram FACS data is displayed. V indicates viable cell population, G indicates GPR97 expression. DP indicates cell populations positive for both GFP and GPR97. Mean detected GFP is displayed in red for wild-type and variant GPR97.

4.4 Discussion and future work

The results presented in this chapter facilitate the interrogation of orphan adhesion GPCR signalling by establishing robust reporter assays for CD97, GPR56 and GPR97, whilst evaluating the use of small, highly stable scaffold protein Affimers as surrogate ligands. The reporter assays are also established as an appropriate way to analyse the functional implications of natural variants of receptors identified in cohort studies. Finally, this chapter also strongly indicates GPR56 as a mechanosensitive receptor, demonstrated by a 3-fold increase in signalling activity upon mechanical stimulation in comparison to unstimulated cells.

Adhesion GPCRs CD97 and GPR56 are two of the more well-characterised receptors in the family, and are known to act via $G\alpha_{12/13}$ and subsequently activate RhoA to influence cell migration (94)(158). RhoA GTPase potentiates both SRE- and SRF-driven transcription in response to heterotrimeric G-protein activation (251). Therefore, luciferase assays were established utilising SRE and SRF-RE to drive firefly luciferase expression in response to adhesion GPCR activation. HEK293Ts were first transfected with various CD97 constructs with different signalling capacities, and experimental set-up was confirmed as demonstrated in Figure 4.3. Constitutively active CD97ΔNTF generated a robust luciferase signal by activating SRF-RE but not SRE, whilst low level constitutive activity of CD97 1,2,5-FL was detected for both reporters. As expected, SRF-RE was not activated in cells expressing signalling deficient CD97-TM1. However, SRE-signalling remained very low even in the presence of serum, used as a positive control to confirm reporter activity, therefore SRF-RE only was used to examine receptor activity in subsequent assays.

CD97 has a number of characterised endogenous ligands such as chondroitin sulphate, CD55, CD90, and integrins α5β1 and αvβ3, interacting via various regions of its N terminus (72)(82)(110)(252). EMR2 also interacts with chondroitin sulphate via its EGF-domains, which are highly analogous to those of CD97 (253). Despite this, G protein coupling as a result of these interactions has not been shown. This may insinuate an extra factor is required, like mechanical stimulation, which was not included in these previous studies. Interestingly however, studies have previously confirmed that binding of the monoclonal antibody 2A1 to EMR2 NTF causes neutrophil enhancement in response to inflammatory stimuli (59). Furthermore, binding of an anti-GPR56 antibody to the NTF of GPR56 augments receptor signalling (58). These studies indicate the importance of N-terminal interactions for receptor activation, which may need to be of high affinity. However, many receptor-

ligand complexes are often transient in nature in order to regulate aspects of cellular function. CD97 binding to CD55, for example, has been characterised as low affinity with a rapid off-rate, at 86 μM and ~0.6s^{-1} respectively (254). Additionally, EMR2 and CD97 interaction with chondroitin sulphate is low affinity, characteristic of leukocyte cell surface receptor interactions (253). With such transient interactions, investigating the molecular mechanisms of signalling activation remains a challenge.

Alternative approaches such as the design of surrogate ligands that are capable of inducing signal transduction are required to examine receptor activity. Although antibodies remain a highly popular molecular tool for probing protein-protein interactions and protein activity, they are often difficult to generate and validate. Affimers however, provide an alternative technology that are easily produced and screened for utilising phage display and *in vitro* binding selection (Section 1.4.1), and can be easily modified to include affinity tags for experimental application (161). They are an attractive tool in modulating protein-protein interactions, and mediate very high affinity interactions, capable of binding to target proteins in nanomolar quantities (156), mirroring that of antibodies (157). We therefore hypothesised that Affimers could be used as synthetic ligands for adhesion GPCRs in signalling assays. Subsequently the potential of Affimers raised against CD97 GAIN domain for use as stabilising proteins for crystal formation (Chapter 3) were tested in SRF-RE luciferase assays. Furthermore, we successfully generated 5 Affimers against both GPR56 and GPR97 GAIN domains, expressed in a mammalian system, to examine their ability to act as synthetic ligands and ultimately decrypt GPR97-G protein coupling. Strengthening this hypothesis, Salzman et. al (2017) published data during this investigation validating the use of monobodies as synthetic ligands, providing evidence towards a *Stachel*- independent mode of adhesion GPCR signalling, and revealing the potential of small binding proteins to act as both allosteric agonists and inverse agonists (255).

The results shown in Figure 4.4 confirm that the Affimers generated against CD97 had minimal influence on full-length receptor signalling (Figure 4.4C). There is no significant difference observed in luciferase signal between buffer and Affimer-stimulated cells. Any effects observed were likely to be non-specific, confirmed by the apparent increase in ΔNTF signalling as this construct lacked the Affimer binding site. This notion is further supported by the fact that Affimer binding was not detected by FACS analysis. As the Affimers were raised against bacterially expressed CD97 GAIN domain, it is possible that the glycosylation pattern of CD97 GAIN domain when

expressed by mammalian cells alters the binding sites to prevent GAIN domain-Affimer interaction.

The SRF-RE assay was validated for GPR97. As illustrated in Figure 4.5, we show for the first time that GPR97 activates SRF-RE driven transcription, indicating coupling to and activation of $G\alpha_{12/13}$. Unexpectedly, signalling output by constitutively active GPR97 ΔNTF was considerably lower than that of GPR97 FL, which may indicate poor trafficking to the cell membrane in comparison to the full-length receptor. The same result was also gained by expression of GPR56 ΔNTF, contradicting previous studies detailing significantly higher signalling activity by ΔNTF constructs in comparison to full length GPR56, GPR126 and BAI2 (55)(58)(63). However, signalling of the full-length receptors demonstrated in Figure 4.5 confirmed SRF-RE reporter assays to be an appropriate method to investigate signalling of both adhesion GPCRs. Future experiments could include mutants of GPR97 that inhibit cleavage, and alter key residues within the *Stachel* peptide, to determine the importance of the NTF in mediating signalling.

Prior to use in luciferase assays, the binding specificity of Affimers raised against GPR56 and GPR97 were confirmed by flow cytometry. The data depicted in Figures 4.8 and 4.9 indicate specific binding of two Affimers for each receptor. Monoclonal primary αGPR56 antibody was included as a positive control, confirming receptor expression and binding to 9.79% of the viable cell population. Affimer 1 bound with high specificity at 12.34%, indicating stronger receptor recognition and binding than that of the antibody. Affimer 3 was also deemed positive, with a cell population of 5.34%.

The expression of GPR97 with a C-terminal GFP tag served as a positive control for receptor expression. GFP-positive populations accounted for 25-28% of the viable cell population (Figure 4.9). Affimers 1 and 4 were considered specific, binding 3.93% and 2.44% of the whole population respectively. The remaining Affimers for both receptors bound similarly to the negative controls. Although small population shifts, representative of GPR97 Affimer binding, were seen in comparison to the whole viable population, when compared to the population of cells expressing GFP, the percentage binding was 15.25% and 9.43% for GPR97 A1 and A4 respectively. It has also been reported that GPR97 surface expression is significantly lower than transfection levels (personal communication with Dr HH Lin, Taiwan). Therefore, in practice, Affimer binding is likely to be higher than shown in Figure 4.9. Furthermore, the αHis antibody used in this experiment to detect Affimer binding has since been shown to be unsuitable for application in FACS analysis (Stacey group, Leeds).

Consequently, future experiments would directly label the Affimer to determine binding specificity.

While Affimers have been documented to effectively bind in other systems such as *in vitro* intracellular protein visualisation, ion channel modulation, transmembrane receptor modulation, and particle tracking (161), this is the first documented use of Affimers in staining of cell surface molecules for analysis by flow cytometry. This demonstrates exciting potential for Affimers to be exploited to define both cis and trans protein-protein interactions (PPIs) that influence cell receptor activity. Certainly, GPR56 has previously been shown to interact in trans with other GPR56 NTFs. This interaction alleviated the inhibitory influence of the NTF, and therefore potentiated RhoA signalling (58), although it is not known whether this response is exhibited by all adhesion GPCRs. The likely high affinity and specific interactions mediated by Affimers, alongside their ease of production, makes them an attractive alternative to antibodies (256) for identifying and quantifying previously unknown PPIs that govern signal transduction of adhesion GPCRs.

The Affimers identified by flow cytometry to bind specifically to GPR56 and GPR97 were applied in reporter assays to investigate their ability to modulate receptor signalling. A small decrease in GPR97-mediating signalling was evident upon interaction with Affimer 4 (Figure 4.10D). However, despite the previous confirmation of Affimer-GAIN domain interaction specificity, these Affimers also seemed to mediate a small decrease in signalling in EV-transfected cells, though it was not deemed significant by ANOVA.

The results presented in Figure 4.10B identify GPR56 Affimer 1 as an inhibitor of GPR56-mediated signalling. SRF-RE activity was significantly decreased in cells transfected with GPR56 FL in comparison to that of buffer stimulation, showing approximately 50% inhibition at a concentration of 0.4 µM / 5 µg/ml. This result is comparative to those in the study by Salzman *et.al* (2017), investigating the potential of monobodies as activity-modulating proteins (255). Different monobodies were found to agonise or antagonise signalling at a final concentration of 0.7 µM. These effects were shown to be independent of both receptor cleavage and the *Stachel.* Therefore, to define the molecular mechanism of Affimer-induced signalling inhibition, future experiments may include Affimer treatment of the following constructs; GPR56 mutated within the *Stachel* region at conserved residues, and cleavage-deficient GPR56.

Excitingly, this is the first recorded instance of Affimers modulating GPCR signalling, an observation that indicates the further potential for Affimers to be used as a tool in blocking receptor-specific biological function and therapeutics. As this particular GPR56 Affimer was selected from a panel containing only one randomised loop (Figure 4.7A), it may therefore be possible to generate cyclic peptides based on this region. Cyclic peptides are already proven to be excellent therapeutic alternatives to small molecules and larger biologics (257). Therefore, the rationale presented here in targeting receptor-specific signalling using Affimers could be exploited in the future to target disease-variants of adhesion GPCRs.

With a lack of observed ligand-induced signalling within the adhesion GPCR family, another factor is likely required to influence receptor activity. The significant vibration-induced signalling of GPR56 outlined in this chapter provides exciting support for the hypothesis that adhesion GPCRs are indeed mechanosensors. Figure 4.12B illustrates for the first time, the sensitivity of GPR56 to vibration. Stimulation for only 5 minutes increased the GPR56 FL-mediated signalling activity 3-fold in comparison to those under static conditions. The increase was not observed in cells expressing GPR56 ΔNTF, with the slight difference in signalling deemed insignificant, implying that the NTF is crucial in regulating GPR56 signalling. This notion is in agreement with data previously published by A. Kishore *et.al* (54). Like EV-transfected cells, CD97/GPR97 FL and CD97/GPR97 ΔNTF did not exhibit upregulated signalling in response to vibration (Figures 4.11 and 4.12), confirming that the response to mechanical stimulation was GPR56-dependent. The results for GPR56 FL mimic those of the urticaria-associated EMR2 variant, with activity largely upregulated in response to vibration alone. These observations strongly link GPR56 with mechanosensitive properties, concurring with the findings of White *et.al* (2014) and the role of GPR56-mediated $G\alpha_{12/13}$ signalling in mechanical overload-induced muscle hypertrophy (114). Moreover, the activation of GPR56, by removal of the NTF, in melanoma cells contributes to cell migration (158). The need for *Stachel* exposure in this process indicates that GPR56 is potentially regulated by changes in the mechano-microenvironment, which are usually experienced by metastatic tissues (Section 1.3).

Indeed, several other adhesion GPCRs are routinely exposed to mechanical stimuli within their microenvironment and involved in its recognition. For example, VLGR1 is associated with the perception of sound waves in the cochlear, whilst those adhesion GPCRs that are expressed by leukocytes are constantly exposed to shear flow in the circulation (93)(248). Further, Latrophilin 1 has been proven to mediate cellular

responses to proprioceptive and auditory stimuli in *C. elegans*. In one particular experiment, N. Scholz *et.al* used a piezoelectrical probe to apply a range of frequencies to the organ responsible for locomotion control. They demonstrated that *dCirl* (Latrophilin homologue) is necessary for a relative response of chordotonal neurons and subsequent crawling as a result of mechanical stimulation (61).

As previously discussed, structural determination of Latrophilin 1 and BAI3 illustrates the tethered agonist to be deeply embedded within the wider GAIN domain, and therefore notable structural changes would likely be required to liberate the *Stachel* or mediate domain rearrangement at the very least. *In vivo* models have demonstrated that arterial shear stress mediates subunit separation of CD97 in combination with binding to ligand CD55. The presence of CD97 NTF was not detected in the serum of CD55-knockout (KO) mice under flow conditions inferring that shear flow alone is not sufficient for separation. However, when CD55-KO leukocytes were transferred to wild type mice, levels of CD97 NTF in serum were upregulated indicating shedding/removal of the NTF from the cell surface after ligation with CD55. Simultaneously, the CTF of CD97 was internalised from the surface of cells, an outcome that was not detected in the absence of shear flow (93). Although the signalling consequences were not assessed in this case, these results demonstrate at least a regulatory role for mechanical force in adhesion GPCR biology.

EMR2 is implicated as a mechanosensor in mast cells, confirmed as the cause of a rare, vibration-induced urticaria. A novel variant of EMR2 was identified in families presenting with hives after repetitive mechanical stimulation caused by towel rubbing for example. The missense mutation was found upstream of the GPS (C492Y). This seemingly destabilised the GAIN domain making the receptor more susceptible to subunit separation, and alleviating the inhibitory interaction mediated by the NTF. The variant confers a pathological gain of function. Mast cells expressing the receptor were sensitised to vibration-induced degranulation, experimentally mediated by vortexing the forearm of patients, and upregulated NTF shedding was also detected (46). This response, along with the high expression of the EMR2 ligand dermatan sulphate in skin support a role for the adhesion GPCR in perception and response to mechanical stimuli.

SRF activation is significant in stimulating transcription in response to mechanical stimuli. Transmembrane receptor integrins are largely responsible for mediating mechanotransduction. They connect the ECM to the actin cytoskeleton and detect changes in environmental stiffness, adding pressure to the focal adhesions that

mediate the interaction. Mechanical stretching of integrins results in RhoA activation, and RhoA is an upstream regulator SRF-mediated signalling (258), the targets of which include genes associated with mechanotransduction and cancer metastasis (259). Taking this into consideration fuels the idea that mechanical force can influence the activation of receptors coupled to $G\alpha_{12/13}$ which activate Rho-family GTPases. Therefore, to test the effect of ligand-binding under mechanical stress, the response of GPR56 and GPR97 GAIN domain-Affimer binding was tested under vibration. HEK293Ts expressing EV, FL or ΔNTF receptors were incubated with Affimers immobilised to nickel sepharose beads for 30 minutes and exposed to vibration mimicking that previously shown to induce mast cell degranulation and EMR2 NTF dissociation. A bead only control was included to confirm that any alterations seen in signalling were due to Affimer binding under mechanical force, Affimer binding alone, or mechanical force alone. Figure 4.10A confirms the beads did not influence signalling, with the luciferase signal mirroring that of buffer stimulation. Incubating cells with receptor-specific Affimers did not provoke receptor-mediated signalling and luciferase output imitated that of buffer and bead-only stimulated cells.

These results demonstrate a varied sensitivity to mechanical force within the adhesion GPCR family. It may be that whilst GPR56, like EMR2, dissociates in response to vibration alone, GPR97 like CD97 (93) may require simultaneous ligand binding and vibration/shear flow to mediate NTF dissociation. However, the data described is preliminary and a number of variables should be further tested, such as vibration time and intensity. This will help to provide a more conclusive picture of the effect of mechanical force and ligand binding on receptor activation. To advance our understanding of GPR97 subunit dissociation, flow chamber experiments could be established to mimic circulatory shear flow, as previously described for CD97. Moreover, the levels of soluble GPR56 NTF could be assessed both before and after vibration to confirm shedding, or to see if vibration is responsible for inducing structural rearrangement within the GAIN domain to facilitate *Stachel* exposure. Furthermore, future experiments using cleavage-deficient and *Stachel* mutants, and ligand-binding domain mutants will help dissect whether ligand binding contributes to *Stachel*-dependent signalling, and if total NTF dissociation is required to expose the tethered agonist or whether the disinhibition model of activation is correct (Section 1.2.3, Figure 1.7).

Disease-causing mutations from classical genetic cohort studies have provided functional insights into a limited number of adhesion GPCRs (260)(261), however the lack of pharmacological agents and functional readouts have greatly hindered the understanding of the vast majority. Having established readouts for both GPR56 and GPR97, we were able to exploit the extremely novel “genotype to phenotype” approach in humans, which has rarely been used. Birth cohort studies employ extensive patient genomic data, which may result in identifying gene knockouts in patients who have non-specific symptoms and therefore do not display an obvious phenotype.

Using Whole genome sequencing data from chip array analysis (Illumina-HumanCoreExome-24-v1-0) of the Born in Bradford cohort patient data, we identified a homozygous mutation within TM3 of both GPR56 (13 individuals from 13000) and GPR97 (5 individuals from 13000). TM3 is documented to be pivotal in GPCR activation through interaction with TM6. Upon ligand binding, the latter swings in an outward movement to provide an intracellular binding site for the α-subunit of heterotrimeric G proteins (262). TM3 has recently been shown to contribute to the formation of a hydrophobic core between TM2,TM3, TM6 and TM7, stabilising TM6 to prevent constitutive receptor activity (263). The effect of these TM3 variants on GPR56/967 signalling was therefore investigated via SRF-RE luciferase assays.

Basal signalling activity of GPR56*M487T* was slightly increased in comparison to the wild type receptor, though not significantly (Figure 4.11). Further investigation by immunofluorescence and flow cytometry, illustrated in Figures 4.12 and 4.13, detected similar levels of expression at the cell surface for GPR56*wt* and GPR56*M487T* suggesting the mutation within TM3 does not alter receptor function. On the other hand, activation of GPR97-mediated signalling was abolished in the variant receptor. The level of luciferase mirrored that of EV-transfected cells, with a ~10-fold reduction in comparison to GPR97*wt* (Figure 4.11). The point mutation may have altered the interaction between TM3 and TM6, preventing the movement of TM6 which would otherwise enable basal signalling activity seen by unstimulated GPR97*wt*. The mutation could also impair receptor processing and trafficking to the cell membrane, therefore eradicating receptor function. Indeed, the data presented in Figures 4.14 and 4.15 suggest that is the most plausible explanation for abolished signalling activity of GPR97*G437S*, with a lack of GPR97 visualised and detected at the cell surface by immunofluorescence and flow cytometry respectively. Our identification of naturally occurring receptor variants and their physiological

implications showcases a technique that can be applied to the wider adhesion GPCR family and GPCR superfamily alike, to help decipher ambiguous clinical symptoms.

To conclude, the experiments performed in this chapter have identified GPR97-mediated SRF activation via coupling to $G\alpha_{12/13}$, revealing a signalling mechanism for this orphan receptor. The assays established offer a simple and effective method for interrogating the functional implications of homozygous variants in the adhesion GPCR family. These can be utilised for the wider family which are increasingly implicated in disease. This work has also demonstrated the effective design and use of highly stable and specific binding protein Affimers as surrogate ligands, with GPR56-specific Affimer 1 negatively regulating receptor-mediated SRF activation. The design of such proteins may allow for future elucidation of orphan receptor activity and prove highly valuable in the development of therapeutics by modulating disease-associated signalling pathways. Finally, mechanical sensitivity of GPR56 in response to vibration has been determined, displaying a near 3-fold increase in activity. This result not only shows that GPR56 may act as a mechanosensor but may also provide a paradigm for all adhesion GPCRs as mechanosensitive proteins. At the very least, these data show that mechanically-induced signalling, and the mechanical structural properties, of all other adhesion GPCRs, are worthy of investigation.

Chapter 5. Interrogating the mechanical stability of adhesion GPCR GAIN domains

5.1 Introduction

Cell adhesion proteins are frequently subjected to mechanical forces that can instigate conformational changes to reveal cryptic binding sites or induce signalling. This subsequently enables receptors to influence cellular phenotypes by mechanotransduction (Section 1.3). As previously indicated, adhesion GPCRs are routinely exposed to mechanical stimuli within their microenvironment and hence have great potential for being involved in mechanosensing. It has also been proposed that mechanical force is likely required in tandem with ligand binding to the extracellular domain (ECD) of adhesion GPCRs to mediate uncovering of the tethered agonist.

To gain deeper understanding of their mechanical properties and how these novel receptors are potentially activated, this chapter describes experiments that utilised atomic force microscopy (AFM) to dissect the force required to disrupt the non-covalent interactions holding together the GAIN domain (Section 1.3.1). The work presented here aims to measure the mechanical stability of the GAIN domain and confirm whether the force required to induce separation is within the range observed under physiological settings. These measurements will ultimately determine the feasibility of N-terminal removal for the tethered agonist model of signal induction (Section 1.2.3).

5.2 Generation of GAIN-I_{27} polypeptide constructs for force spectroscopy

GAIN domains of the EGF-TM7 subgroup were primarily selected for investigation due to their predicted secondary structures differing from the wider adhesion GPCR family (Section 3.1). As the force-extension profile (Section 1.3.2, Figure 1.10) for any adhesion GPCR GAIN domain is completely unknown, the inclusion of domains with characterised unfolding force-extension profiles is recommended. These can be used to facilitate the identification of unique peaks seen on traces from SMFS pulling experiments, which can then be attributed to the unfolding of the unknown domain. Constructs containing multiple repeats of the Ig module 27 (I_{27}) of the human titin protein have been used to this end in multiple experiments, and as a result, have a very well defined characteristic sawtooth profile (264)(265).

To be able to measure the mechanical strength of protein domains, single molecule force spectroscopy experiments require molecules to be tethered at each end, one to a solid surface and the other to the AFM cantilever. Specific covalent protein attachment can be mediated by functionalising either the solid surface, AFM cantilever, or both, with PEG-linkers that are derived with a maleimide at one end, which interacts with free-cysteine residues in proteins. The use of PEG linkers is beneficial for preventing non-specific protein interactions, enabling protein flexibility, and they have a signature force when stretched (266). To further increase the specificity of interaction and ensure the protein is pulled from the desired end, i.e. the C- or N-terminus, protein tags can be used. One such tag that has been used in SMFS experiments is the mutant haloalkane dehydrogenase tag (HaloTag) (141)(146), which forms a covalent, irreversible interaction with its ligand. When brought into contact, an ester bond forms between the ligand and HaloTag, and therefore ensures precise polyprotein tethering. This approach is particularly useful for this investigation due to the free cysteines in the GAIN domain which may otherwise interact with the functionalised surface, preventing complete domain unfolding from a known position. We therefore designed constructs to include two I_{27} repeats either side of a GAIN domain as internal force-extension profile controls, and a HaloTag at the C' terminus for specific polyprotein tethering and immobilisation.

5.2.1 Gibson assembly and overlap extension cloning

Due to the size and complexity of the constructs generated for AFM investigation, Gibson Assembly was used in place of classic PCR for cloning. Gibson Assembly facilitates cost-effective and rapid joining of up to 5 DNA fragments in a single reaction, mediated by an enzymatic master mix containing DNA polymerase, a 5' exonuclease to generate 3' overhangs for annealing to the following fragment, and DNA ligase, whilst being ideal for manipulation of large constructs (Figure 5.1). I_{27} ($_{1+2}$), GAIN domain, I_{27} ($_{3+4}$) and the HaloTag domain were previously amplified by PCR by the Stacey Lab. The amplified fragments were mixed with pSecTag2b vector and NEB Gibson Assembly master mix, according to the manufacturer's protocol as outlined in section 2.2.4.

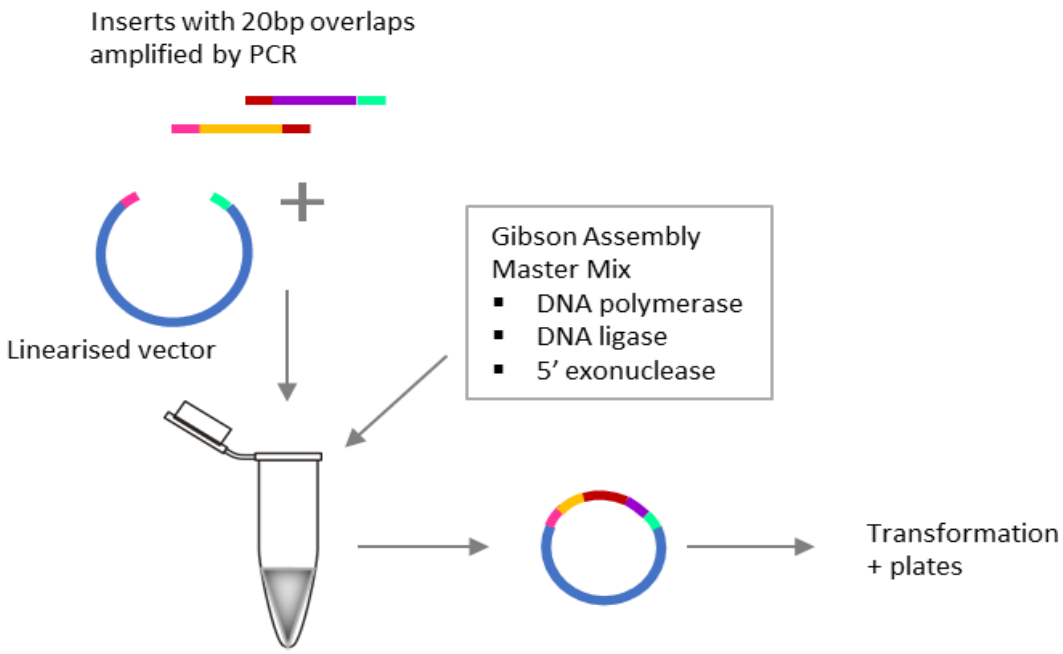


Figure 5.1 ***Schematic representation of Gibson assembly cloning.***

Initial attempts yielded no colony growth after transformation of DH5α *E. coli* with the reaction mixture, suggesting unsuccessful joining of the fragments and vector. In order to increase the efficiency of Gibson Assembly, $I_{27\ (1+2)}$ was joined with either wild type EMR2- or CD97 GAIN, and $I_{27\ (3+4)}$ joined with the HaloTag, by extension overlap PCR (EO-PCR), a method employed to generate "oversized primers" for the vector DNA in a subsequent PCR reaction (267). EO-PCR was performed using Phusion DNA polymerase as outlined in Section 2.2.3. Agarose gel electrophoresis suggested the reaction had been successful (data not included). The DNA was purified before being used for Gibson Assembly PCR. DNA was subsequently used to transform *E. coli* (DH5α). Constructs containing CD97*wt* and EMR2*wt* GAIN domains were successfully generated and confirmed by Sanger sequencing with the following primers; T7, CD97/EMR2 GAIN and bovine growth hormone (BGH). However, the sequencing results indicated that only 1 of 2 I_{27} domains was incorporated upstream of CD97 GAIN domain, whilst an additional I_{27} was incorporated downstream of EMR2 GAIN domain.

5.2.2 Site directed mutagenesis for cleavage deficient GAIN domains

Cleavage deficient adhesion GPCR GAIN domains are associated with phenotypic changes and disease, therefore it was appropriate to assess the implication of a mutated GPS motif on the force required to separate the GAIN domain. A mutation at the GPS site previously shown to prevent post-translational autocatalytic cleavage (38), but still enable receptor trafficking to the plasma membrane, was incorporated into CD97 and EMR2 GAIN domains. Site directed mutagenesis (SDM) was employed to introduce the point mutation of serine, at the GPS cleavage point (HL↓S), to alanine (HL↓A) within the GAIN domain of the already assembled CD97 GAIN*wt*-I_{27}-Halo, and EMR2*wt*-I_{27}-Halo constructs, generating CD97*S531A* and EMR2 GAIN*S518A*.

SDM was initially performed following a protocol outlined by Laible M., Boonrod K. (2009). Primers were designed to have >10 nucleotides either side of the mutation site, a restriction site, and a melting temperature of >60°C. 50ng of vector DNA and 150ng of each forward and reverse primer DNA was added to the PCR reaction with Phusion high fidelity polymerase. Reaction mixtures were digested with *Dpn* I for 1 hour at 37°C to ensure complete degradation of remaining bacterial DNA. Agarose electrophoresis indicated that PCR and digestion was unsuccessful for both CD97*S531A* and EMR2 GAIN*S518A* proteins with no visible band at the expected size of 8.5kb. Thereafter, the mutagenesis reaction was optimised with changes to both template and primer concentration, and to the melting temperature and extension time in the PCR cycle. Reaction samples were subject to analysis by 0.8% agarose electrophoresis. Figure 5.2 indicates high DNA concentration roughly between 8-10kb. Therefore, 10µl of each PCR reaction was subject to *Dpn* I digestion and transformed into *E. coli* supercompetent strain XL1-Blue cells. Colonies were randomly selected and grown overnight at 37°C for DNA purification and verification. However, Sanger sequencing established that the point mutation had not been incorporated within the GAIN domain of each construct.

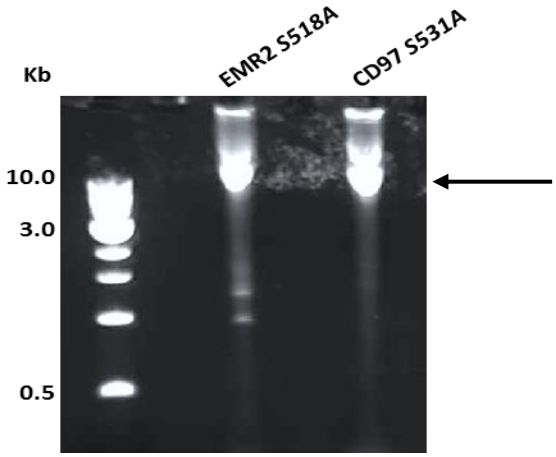


Figure 5.2 ***Agarose gel electrophoresis analysis of site directed mutagenesis*****.** Samples of both EMR2 and CD97 PCR reactions were resolved by electrophoresis, and subsequently treated with *Dpn*I. PCR products were transformed into *E. coli* XL1-Blue cells for further confirmation.

Mutagenesis was subsequently carried out using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). Immediately following PCR, *Dpn* I was added to reaction mixtures, before products were transformed into supercompetent *E. coli* strain XL1-Blue cells. To clarify the identity of the positive clones, colonies were selected and grown in ampicillin containing media. The DNA was isolated and validated by Sanger sequencing, which confirmed the point mutation at the GPS motif for both EMR2 GAIN*S518A* and CD97 GAIN*S531A* respectively. The constructs generated are schematically represented in Figure 5.3.

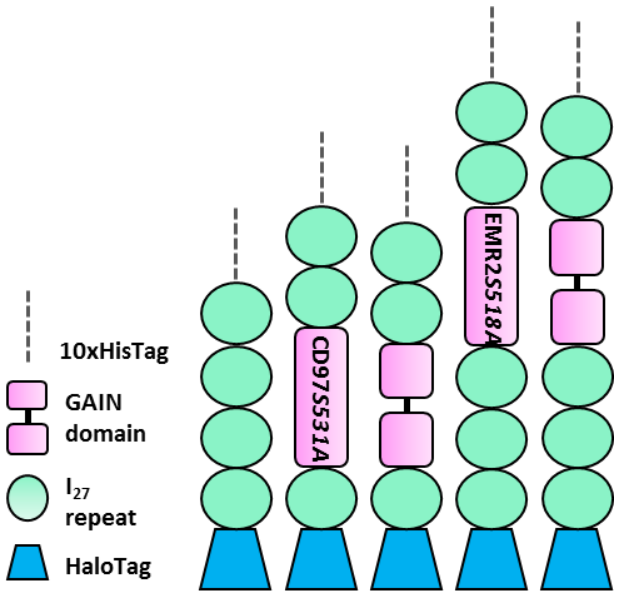


Figure 5.3 ***Schematic of mammalian polyproteins generated for single molecule force spectroscopy experiments*****.** A control protein of four I_{27} repeats and a HaloTag was generated alongside wild type (wt) and cleavage deficient GAIN domains of CD97 and EMR2. An N-terminal HisTag was incorporated for purification. GAINwt domain is represented here as two joined units to highlight the non-covalent association post cleavage, whilst cleavage mutants are represented as a single unit**.**

5.2.3 Expression and purification of GAIN-I_{27}-HaloTag constructs

5.2.3.1 Protein expression in mammalian cells

As GAIN domains possess conserved disulphide bonds and are highly glycosylated, and only microgram quantities of protein were required for AFM pulling experiments, a mammalian expression system was chosen for the generation of protein. The DNA for each CD97 GAIN*wt*, CD97 GAIN*S531A,* EMR2 GAIN*wt,* EMR2 GAIN*S518A*, and I_{27}(x4) HaloTag construct was transfected into HEK293T in a 6-well plate, using TurboFect reagent, as outlined in section 2.3.2. A sample of supernatant taken from the well of each transfected construct was resolved on SDS-PAGE and analysed by western blot. Figure 5.4A shows the successful small-scale expression of all but one construct. The expected weight of each wildtype GAIN domain is much less than that of the cleavage deficient mutant, due to separation of the subdomains. Proteins were detected with primary anti-HaloTag antibody and secondary anti-mouse HRP, meaning only the protein domains downstream of the cleavage site, in tandem with the HaloTag, would be detected. A band at approximately 100 kDa in the lane

corresponding to CD97 GAIN*wt*-I_{27} indicated that the GAIN domain had not undergone cleavage. Expression was scaled up to larger volumes, however very little of each construct was detected by western blot, indicating poor transfection efficiency.

Subsequent expression trials included a GFP control to visually assess transfection efficiency, which was consequently estimated to be 10%. To improve efficiency, expression was optimised by assessing transfection media, transfection reagent, harvest time post-transfection and cell density at seeding. Supernatants were harvested at 48, 72 and 96 hours post-transfection, resolved on SDS-PAGE and protein detected by western blot. Transfecting in either DMEM or OptiMEM yielded very similar levels of protein expression, whilst harvesting after 48 hours was deemed optimal. The use of transfection reagents TurboFect (ThermoFisher) and Lipofectamine® 2000 (ThermoFisher) presented vastly different results. As shown in Figure 5.4B, transfection of all polyprotein constructs except CD97 GAIN*wt* using Lipofectamine® 2000 resulted in good protein expression. This contrasted the results gained after use of TurboFect; only I_{27}(4) and EMR2 GAIN*wt* constructs were detected by western blot at the expected molecular weights of 78 kDa and 68 kDa respectively. Future transfection thereafter was accordingly carried out using Lipofectamine® 2000.

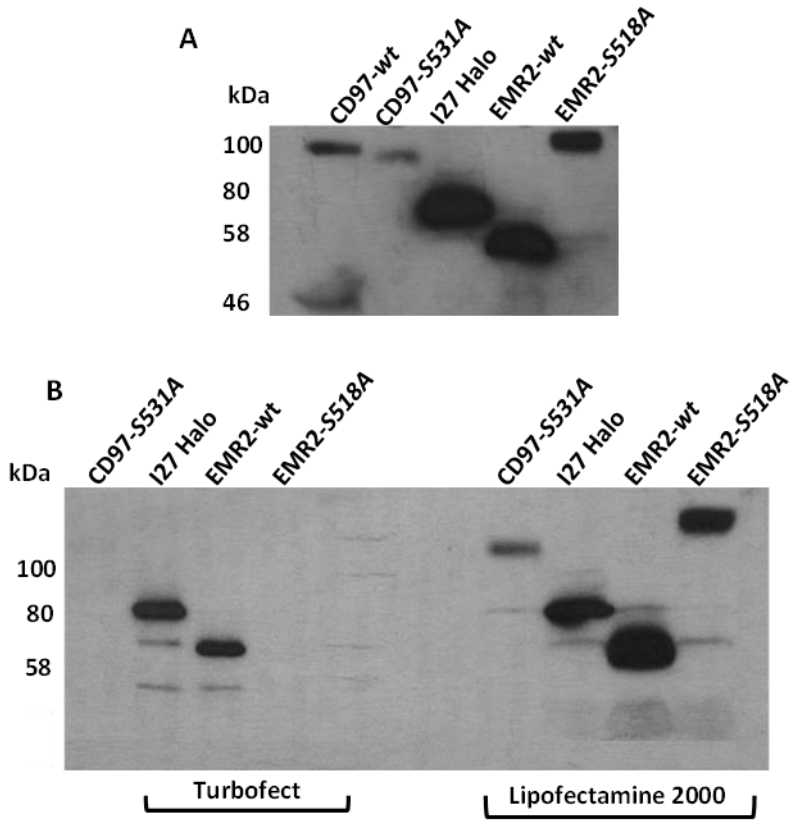


Figure 5.4 ***Transient transfection of AFM polyproteins.*** HEK293T cells were transfected with 1 µg DNA in 6-well plates, and supernatant harvested 72 hours post-transfection. Proteins were detected with primary anti-Halo (Promega) and secondary anti-mouse HRP. The expected molecular weight of each protein; CD97 GAIN*wt* 57 kDa, CD97 GAIN*S531A* 100 kDa, EMR2 GAIN*wt* 68 kDa, EMR2 GAIN*S518A* 110 kDa and I_{27}(4) 78 kDa. (**A**) Initial expression of all constructs except CD97 GAIN*wt* which was consequently removed from future expression trials. (**B**) Transfection was optimised using different reagents.

5.2.3.2 GAIN-I_{27}-HaloTag protein purification

Polyproteins were purified by Ni^{2+} affinity chromatography as outlined in Section 2.2.13. Purification fractions were resolved by 12% SDS-PAGE and detected by western blot. Figure 5.5A illustrates a western blot of example constructs I_{27}(4) and EMR2 GAIN*wt*. Bands representative of each protein are evident in the eluted fractions of increasing stepwise concentration of imidazole. Additionally, the largest band representative of the target protein is visible in lanes representative of the flow through of each construct, suggestive of a low affinity interaction between protein and column resin. 12% SDS-PAGE analysis further demonstrated a high level of

contamination in each fraction (data not shown). Both histidine and glycine are found in mammalian cell culture media, and therefore may compete with the relatively low levels of the His-tagged proteins. Therefore, to remove low molecular weight contaminants, supernatants were dialysed into 300mM NaCl and 20mM Tris pH 8 using a membrane with a 15 kDa exclusion limit. Despite this, both proteins were still largely eluted in the flow through and wash fractions (Figure 5.5B), suggesting buffer exchange was perhaps ineffective at removing contaminants.

Polyproteins were subject to batch method purification to optimise the yield of purified protein. Supernatants were concentrated to a final volume of 3 ml, equilibrated with 20mM imidazole, and incubated with nickel resin overnight at 4 °C with agitation. The resin and bound protein were pelleted at 5000g for 5 minutes. Supernatant was collected as the flow through and the pellet was re-suspended in 1 ml of 20mM imidazole and incubated on ice for 2 minutes before centrifugation. This was repeated for 100mM and 500mM imidazole and fractions were resolved on 12% SDS-PAGE and further analysed by western blot. Figure 5.5Ci shows the results of the SDS-PAGE electrophoresis. Contaminating protein, represented by a large band at approximately 50 kDa, was eluted in both the flow through and subsequent 20mM and 100mM imidazole washes. There is a lack of I_{27} Halo and EMR2-GAIN*wt* at 78kDa and 68kDa respectively in any lanes resolved by SDS-PAGE, signifying low protein concentration.

Although the western blot of batch purification (Figure 5.5Cii) shows both I_{27} Halo and EMR2 GAIN*wt* were eluted in the flow through, very little protein was detected by primary anti-Halo antibody (Promega) in the 20mM fraction. This suggests a higher proportion of EMR2 GAIN*wt* and I_{27}-Halo bound to the nickel resin with higher affinity than previously. This may be due to the extended incubation time in which each construct could interact with and bind to the nickel resin. Additionally, in contrast to the preceding purification, a distinctive band is representative of both I_{27} Halo and EMR2 GAIN*wt* is evident in each 500mM imidazole elution at their respective expected molecular weights (Figure 5.5Cii).

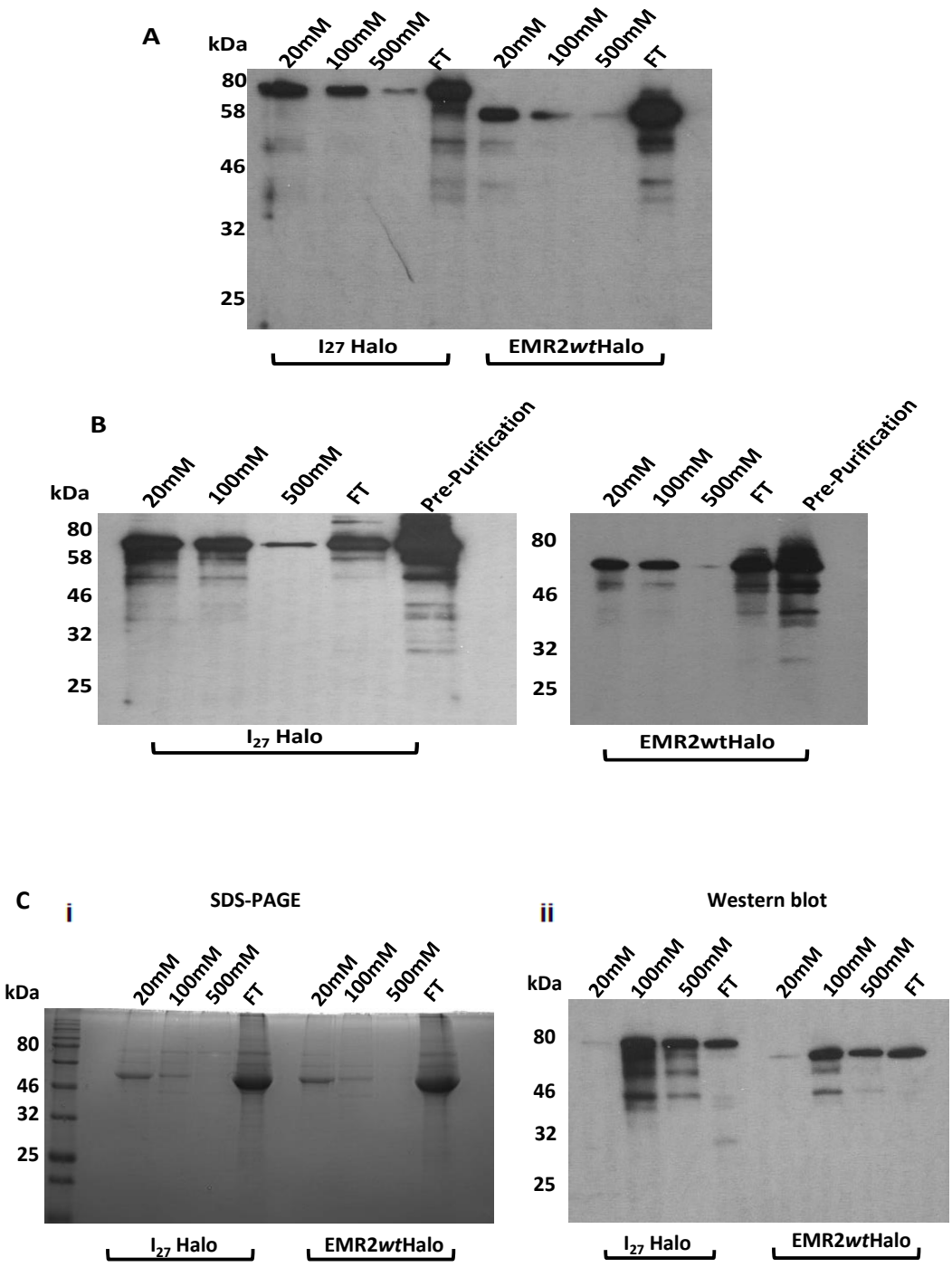


Figure 5.5 ***Purification of I_{27}(4) and EMR2 GAINwt by Ni^{2+} affinity chromatography and batch method.*** (**A**) Western blot of Ni^{2+} affinity purification fractions. (**B**) Western blot of Ni^{2+} affinity purification fractions following dialysis. (**C**) SDS-PAGE and western blot of batch-purification fractions. Proteins were detected with primary anti-HaloTag antibody (Promega) and secondary anti-mouseHRP (Invitrogen).

As a large proportion of protein was present in the flow through during purification, all constructs were re-cloned and engineered to have an extended 10-residue HisTag. This aimed to circumvent the potential masking of the affinity tag during protein folding, thus enabling a stronger interaction of the tag with the nickel resin.

Simultaneously, polyproteins including the cleavage deficient GAIN domain of adhesion GPCRs Latrophilin 1 (Lphn1), EMR1 and GPR97, flanked by a single I_{27} domain and a C-terminal HaloTag were cloned as stated in Section 5.2.1. These GAIN domains were selected as Lphn1 has previously been proven to influence cellular response to mechanical stimuli (61), whilst the EMR1 cleavage site is known to be inactive (35), and GPR97 expression is immune cell restricted, and therefore routinely exposed to shear stress.

Expression of all 10xHis-tag constructs was subsequently scaled up; 200 ml suspension cultures of HEK293T were transiently transfected using Gemini reagent, gifted by GlaxoSmithKline. Cultures were monitored for cell viability daily and upon reaching 40% viability, cells were pelleted and the supernatant harvested. Supernatants were concentrated to 10 ml and purified by Ni^{2+} affinity as previously outlined in Section 2.2.13 (Figures 5.6 and 5.7).

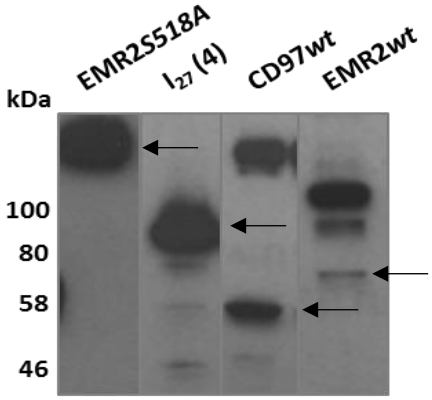


Figure 5.6 ***Optimised purification of AFM constructs confirmed by western blot.*** Fractions representing elution in 500 mM imidazole. EMR2 GAIN*S518A* and I_{27}(4) were the predominant products in their particular elutes at 120 kDa and 78 kDa respectively. Two distinct bands at 110 kDa and 57 kDa for CD97*wt* suggests that both cleaved and un-cleaved GAIN domains have been expressed. The expected weight of EMR2*wt* was 67 kDa, yet the predominant product appears to be at approximately 110 kDa.

Wild-type GAIN domains of CD97 and EMR2 were largely expressed as un-cleaved product, with the predominant protein species detected at approximately twice the expected molecular weight of 57 kDa and 67 kDa respectively (Figure 5.6). Successfully expressed and purified proteins were further concentrated, flash frozen, and stored at -80 °C in 100 µl aliquots at a final concentration of 0.2 mg/ml.

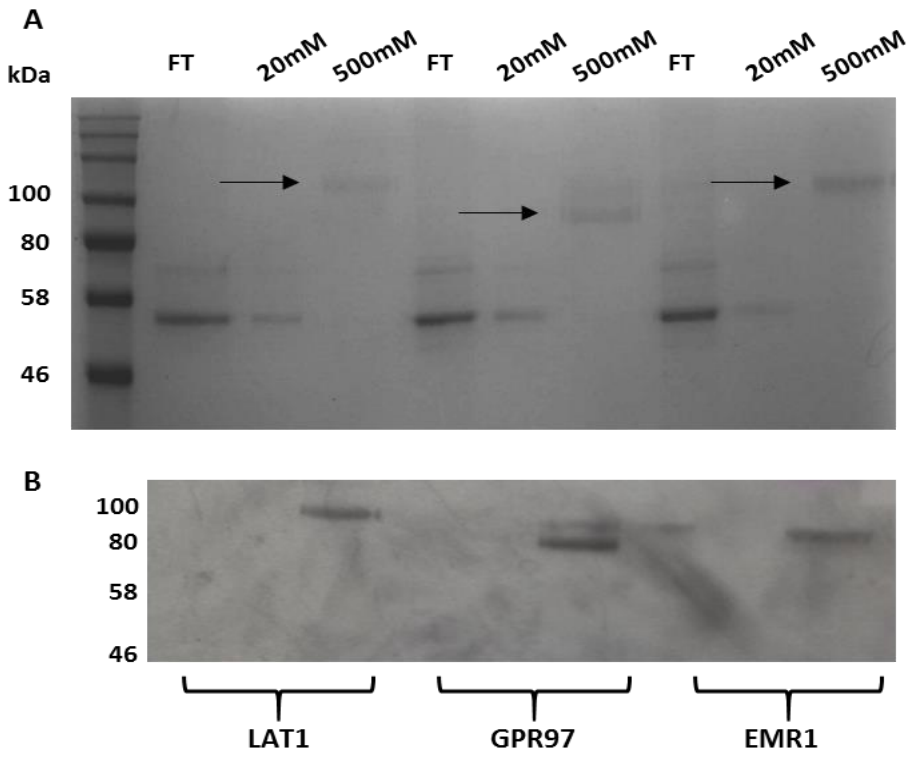


Figure 5.7 ***Confirmation of LAT1, GPR97 and EMR1 polyprotein expression and purification.*** (**A**) Fractions representing protein elution at different imidazole concentrations. Fractions resolved on 12% SDS-PAGE indicate pure, albeit low concentration, protein in each 500 mM fraction at the expected molecular weight of 95 kDa, 84 kDa and 90 kDa for LPHN1, GPR97 and EMR1 respectively. (**B**) Proteins detected with primary anti-Halo in 500 mM fractions.

5.3 Single molecule force spectroscopy of adhesion GPCR GAIN domains

5.3.1 Establishing SMFS experimental set-up

Prior knowledge of protein structure can provide important insights into the potential unfolding threshold under mechanical strain. As discussed in Section 1.3.3, α-helical secondary structures are typically less resistant to mechanical force than those of β-sheets. Structural determination of the GAIN domain revealed a two-subdomain arrangement, with subdomain A being composed of 6 α-helices and subdomain B consisting of a twisted β-sandwich with 13 β-strands and 2 α-helices. Therefore, it is possible that the GAIN domain unfolds in a two-step process, with subdomain A unfolding at a lower force than subdomain B.

If complete separation of the wild-type GAIN domain subunits at the natural cleavage site were to occur at a force lower than that required to unfold the control I_{27} domains, the characteristic saw-tooth pattern of I_{27}-domain unfolding would not be seen on any resulting force-extension profiles. This would be problematic for the interpretation of the force-extension profile, therefore GAIN domain cleavage mutants were used for the preliminary experiments. These proteins do not possess a break in the covalent bonds of the polypeptide, so would ensure the measurement of the I_{27}-domain unfolding. With this in mind, cleavage-deficient GAIN domains were first utilised in AFM experiments. This aimed to define the mechanical strength of the domain in comparison to that of the I_{27} domain, whilst being able to identify the characteristic I_{27} repeat sawtooth unfolding. The experimental set-up (Figure 5.8) was first optimised using the control polyprotein construct, I_{27}(4)-HaloTag.

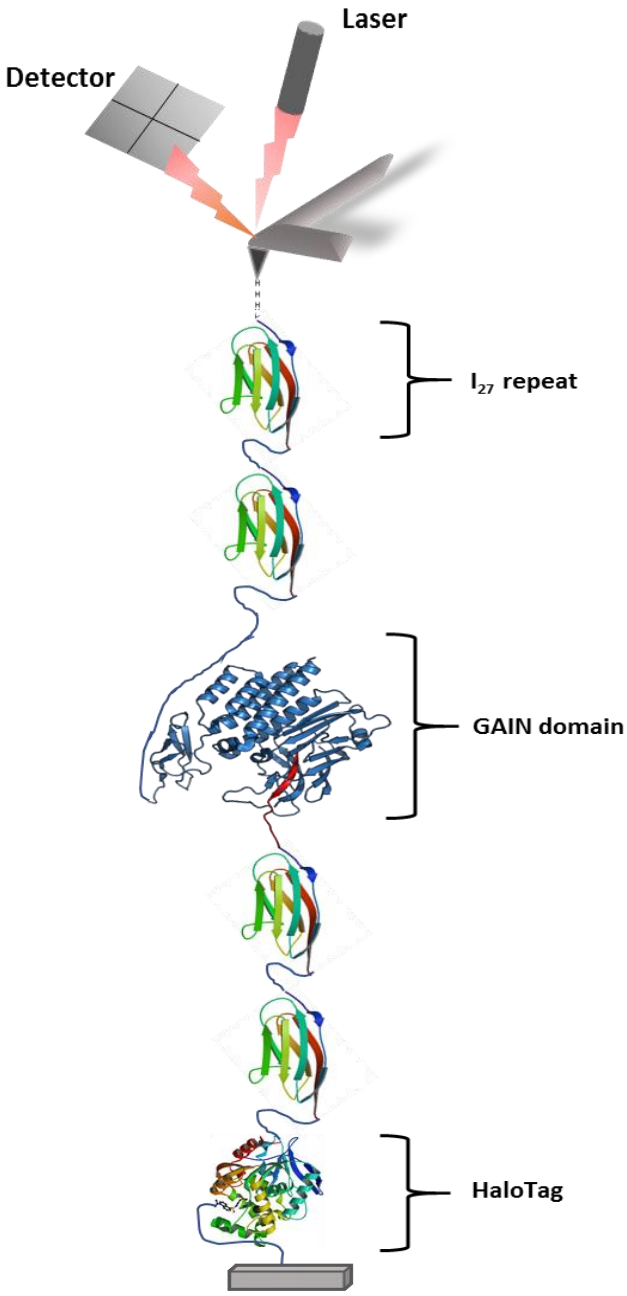


Figure 5.8 ***Representation of example GAIN domain I_{27}-HaloTag constructs for atomic force microscopy*.** Two N' terminal I_{27} domains were incorporated into the pSecTag2b vector downstream of a His-tag utilised for purification, followed by the GAIN-domain of various adhesion GPCRs. Two further I_{27} domains and a HaloTag were also included downstream of the GAIN domain. A control construct consisting only of the 4 I_{27} domains and a HaloTag was also generated.

The polyprotein was immobilised on a silicon nitride surface via SM-PEG linkers and covalent HaloTag-HaloTag ligand interaction (Section 2.5.1). The AFM cantilever and polyprotein were brought into contact in buffered solution (Section 2.5.1.7), mimicking physiological salt concentration, and attaching by adsorption. The interaction was then pulled apart by withdrawing the cantilever at 700 nms^{-1}. Force maps of 20 μm^2 were probed, with between 500-2000 approach and retract cycles to ensure sufficient surface coverage. The cantilever was repositioned after each map to maximise the number of protein unfolding events seen across the surface, as equal distribution of protein attachment across the silicon nitride surface was unlikely. SMFS experiments were carried out in triplicate. The hit rate was low with approximately 1 in 50 approach-retract cycles generating a protein unfolding event. Detected protein events were manually fit to the worm-like chain (WLC) model with a fixed persistence length of 0.4 nm (peptide-unit length), by placing locks on the apex and base of each peak (139). The zero distance was established by the point of hard contact, whilst the retraction baseline was used to zero the applied force (Figure 5.9). From this, the rupture force of each domain could be calculated, alongside the contour length, which is described as the end-to-end distance of a protein/domain when stretched under force (268).

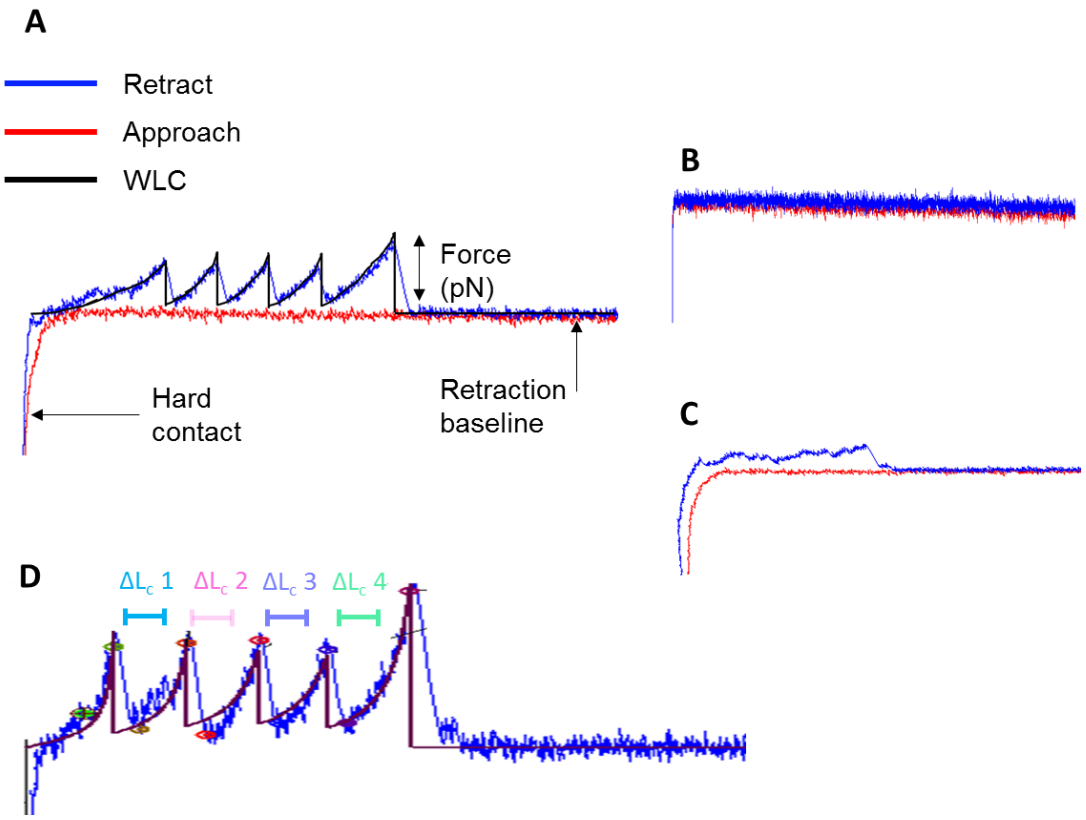


Figure 5.9 ***Example force-extension traces from SMFS*****. (A**) Annotated example force-extension trace from I_{27}(4)-HaloTag experiments, with the worm-like chain (WLC) fit to the protein unfolding peaks, with a fixed persistence length of 0.4 nm. Locks are fit to the base and peak of each curve of the peaks. By doing this, the contour length (Lc) and rupture force for each event can be calculated. Fitting was done manually, using Igor Pro 6.32 (wavemetrics) with an MFP3A AFM software add on (Asylum research), and the calculation performed by the software. (**B**) Example trace of no protein events. (**C**) Example trace of "noise", occurring upon non-specific cantilever interaction with the surface. (**D**) The difference in length (delta contour length (dLc/ΔLc)) between each peak equates to the length of one specific domain unfolding; $\Delta Lc_2 - \Delta Lc_1$ = length of the first domain.

Force-extension traces with 5 peaks were selected for analysis if their profile displayed the sawtooth pattern of I_{27} domain unfolding and fit the WLC model with the fixed persistence length. The 5 peaks were representative of the unfolding of 4 I_{27} domain repeats and the unbinding event from the cantilever (269). The HaloTag domain was not unfolded in this instance. When pulled from its N-terminus, as done in this particular experimental set up, the mechanical strength of the HaloTag domain increases 4-fold in comparison to being pulled from its C-terminus (146).

In a study carried out by I Popa et al. 2013, it was demonstrated that the HaloTag unfolding occurred at a similar force as unbinding from the tip when the polyprotein

containing the HaloTag was covalently attached to the tip (146). Therefore, in the instance outlined here in which the HaloTag is tethered to the surface, polyprotein unbinding is likely to occur before HaloTag unfolding, due to not being covalently attached to the tip.

Manual analysis of all measured force-extension traces revealed a low hit rate of approximately 1-5% of traces with the appropriate number of peaks for unfolding of the entire construct. This however is a typical occurrence of reported SMFS experiments, with hit rates ranging between 1 and 15% (270). After fitting locks to the force-extension profile for fitting the WLC model, the rupture force and contour length (Lc) for each peak, and therefore the entire polyprotein, was recorded. The individual contour length of each unfolding event was then calculated (ΔLc), as the difference between each peak e.g. Lc_2 – Lc_1 (Figure 5.9d). The data was combined in a scatterplot to allow for the specificity of domain unfolding to be identified by the presence or absence of a "hot-spot" indicative of recurring forces and distances. The data represented in Figure 5.10 indicates a mean rupture force of 175.5 pN, and a mean contour length of 28 nm for I_{27} domain, consistent with previously published data for I_{27} length (271).

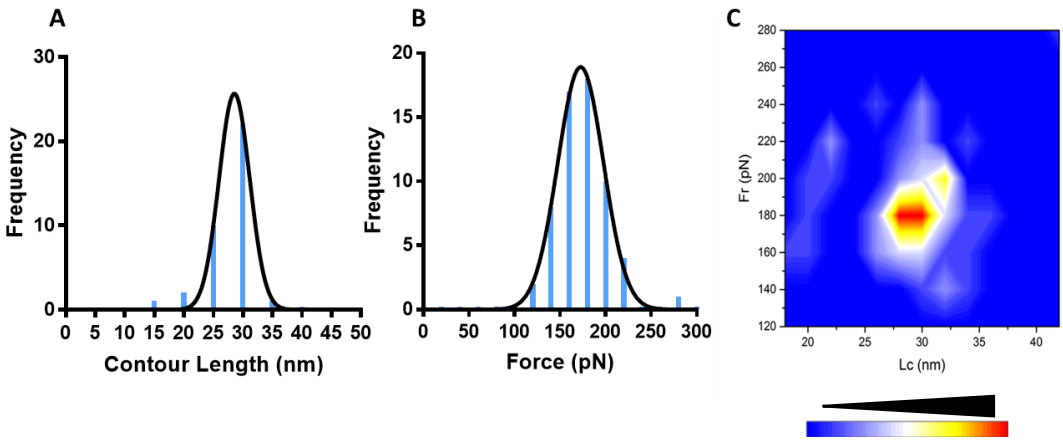


Figure 5.10 ***Pulling experiment of I_{27}(4)-HaloTag at 700 nms^{-1}.*** Protein-cantilever interaction was mediated by adsorption. Contour length (**A**) and force-frequency distribution (**B**) for the dissociation events of the reference polyprotein, each fit to a Gaussian distribution. Contour length is plotted as ΔLc. Histograms contain all data for traces with 5 peaks fit to the WLC on the force-extension trace. (**C**) Combined rupture force and contour length as a scatterplot, with the "hotspot" indicating repeat domain-specific events.

Despite the low hit rate, the data acquired and outlined in Figure 5.10 indicates the expected contour lengths and rupture force observed for the I_{27} polyprotein construct. Therefore, SMFS of the EMR2*S518A* polyprotein was carried out. As the construct contained 5 I_{27} domains, force-extension traces with a minimum of 6 peaks were selected (1 for cantilever unbinding), which would indicate no GAIN domain unfolding. On the other hand, as previously outlined the GAIN domain has the potential to unfold in a one or two-step process, therefore traces with 7 and 8 peaks were also selected. A very poor hit rate was observed, with a total of 48,500 approach-retract cycles across repeated experiments yielding >0.1% hit rate for 6, 7 and 8-peak traces combined. In contrast, a higher percentage of 2-4 peak traces was observed, suggesting the experimental set-up was not sufficient for specifically picking up the protein at the N-terminus.

Figure 5.11 demonstrates some consistency in the protein unfolding events observed, with 1 hotspot evident in the scatterplots for each 5, 6 and 7-event traces. However, the domain lengths outlined in Table 5.1 suggest that if the GAIN domain was to unfold in either a one- or two-step process, two “hotspots” should be evident in a scatterplot. A two-step unfolding would see the length of Subdomain A closely related to the length of unfolded I_{27}, with another spot representative of Subdomain B unfolding at 60.8 nm. Likewise, one-step unfolding of the GAIN domain would be evident with “hotspots” at 28 nm and 108 nm for I_{27} and GAIN respectively. Figure 5.12 shows an example force-extension trace that suggests unfolding may occur in two steps, though this was not replicated frequently.

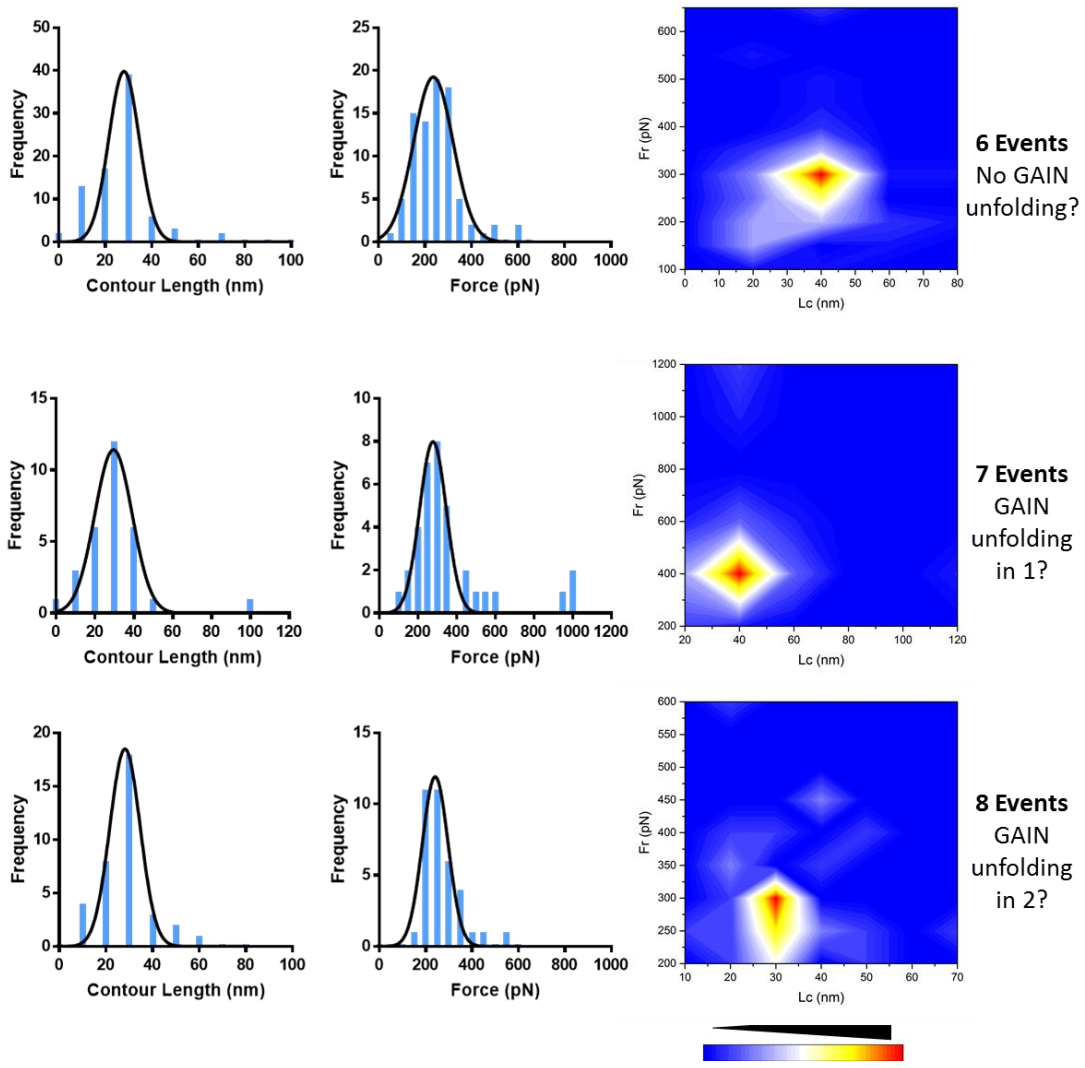


Figure 5.11 ***SMFS of cleavage deficient EMR2 GAIN domain*****.** Contour length and rupture force data was recorded from force-extension traces with 6, 7 and 8 events. ΔLc was calculated, and both length and force were plotted on histograms and fit to a Gaussian distribution. SMFS data was then combined on a scatterplot for each group of data. Data represents 4 experimental repeats.

Protein	Total (nm)	Total – Disulphides (nm)	Subdomain A (nm)	Subdomain B (nm)
EMR1	113.2	102.0	42.8	59.2
Lat1	126.8	114.8	50.8	64.0
EMR2	108.0	96.4	35.6	60.8

Table 5.1 ***Predicted lengths of unfolded GAIN domains from structural information*****.**

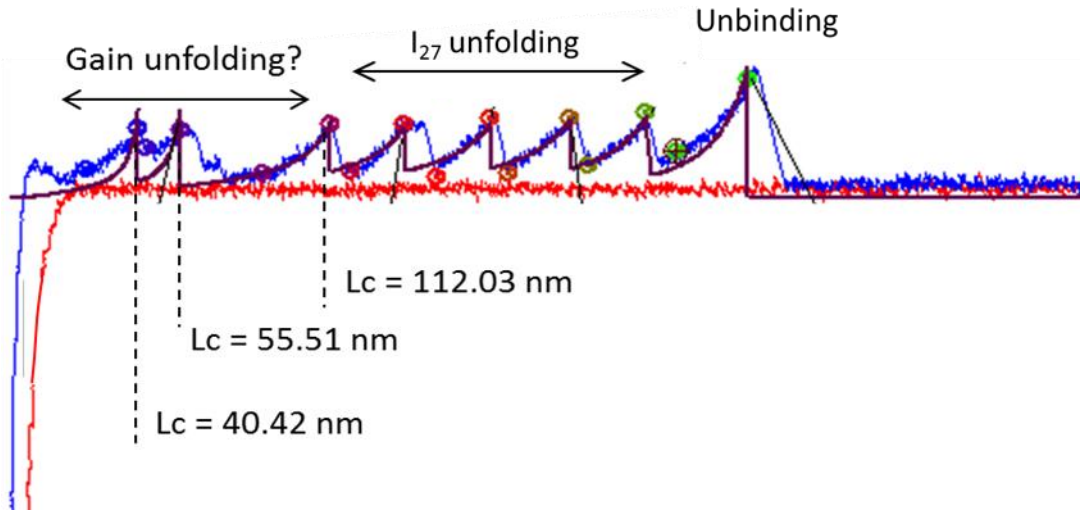


Figure 5.12 Example of *cleavage-deficient EMR2 GAIN domain unfolding in two-steps*. An example force-extension trace of EMR2*S518A* unfolding at 700 nms^{-1}. The first two peaks correspond approximately to the expected lengths of the GAIN subdomains. The subsequent 5 peaks represent the 5 I_{27} domains incorporated into the polypeptide, with the final peak representing unbinding of the protein from the AFM cantilever.

5.3.2 Optimisation of SMFS experiments

To increase the hit rate for SMFS of adhesion GPCR GAIN domains, the AFM cantilever was functionalised with Tris-nitrilotriacetate (NTA) (Section 2.5.1.6), to mediate specific binding to the N-terminal 10xHis tag of the polyproteins via Ni^{2+} ions (272). The protocol followed for AFM cantilever Tris-NTA functionalisation was developed by H J Gruber (Johannes Kepler University), and has been successfully used in SMFS for the study of nuclear-pore protein transport (273). To reconfirm correct experimental set up, SMFS of the I_{27} control polyprotein was performed at the same pulling speed of 700 nms^{-1}.

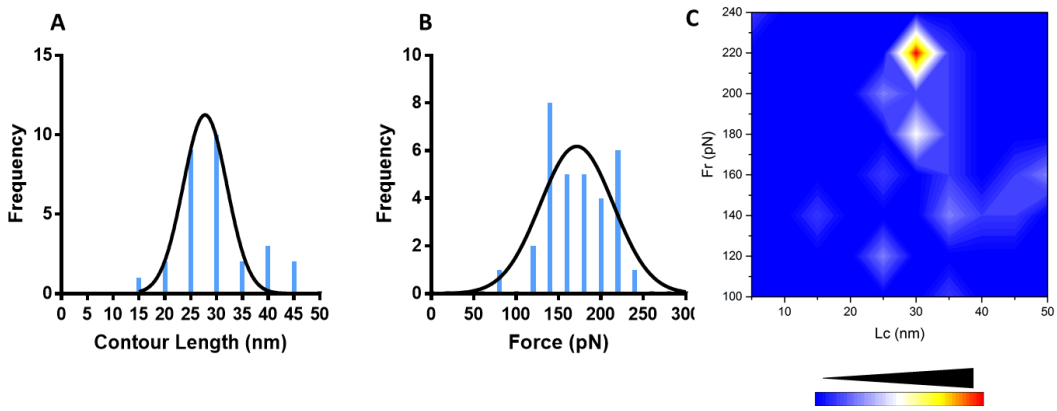


Figure 5.13 ***I_{27}(4)-HaloTag SMFS after cantilever functionalisation*.** Frequency distribution histograms for contour length and force (**A+B**). SMFS data combined in a scatterplot (**C**). Graphed data is n=2 for 5-peak force-extension profiles.

Data was collected for force-extension profiles with 5 events. Figure 5.13 indicates the expected unfolded I_{27} domain length of 28.9 nm. Although the rupture force was found to be more variable, the average force was calculated at 171.4 pN. indicating the experimental set-up was correct and consistent with published literature. Investigation of EMR2*S518A* was subsequently carried out following cantilever functionalisation. However, the majority of approach-retract cycles resulted in no protein contact (Figure 5.9B) or a non-specific interaction with the surface, illustrated as “noise” (Figure 5.9C). Therefore, the mechanical properties of alternative GAIN domains were investigated. Polyproteins with EMR1 and Lphn1 GAIN domains (Section 5.2.3.2) were pulled at 700 nms^{-1} and experiments were carried out in triplicate. As the constructs in this instance contained two I_{27} domain repeats, one either side of the GAIN domain, traces with 3, 4 and 5 events were selected, for the possibilities of no GAIN unfolding, or unfolding in 1 or 2 events respectively.

The force-extension profiles for cleavage-deficient EMR1 indicate an unfolding event occurring at a lower force than that of I_{27} domain (Figure 5.13), an example of which is outlined in Figure 5.14A. Events at a lower force are apparent in traces with both 4 and 5 events, indicated by the “hotspot” areas in the combined data scatter plots (Figure 5.14B). The contour length histograms indicate approximately 30 nm as the most common unfolding length, presumably attributed to the I_{27} domain unfolding. The expected lengths for each subdomain of the EMR1 GAIN domain are 42.8 nm 59.2 nm for A and B respectively (Table 5.1). Therefore, if GAIN domain unfolding was occurring in a single step, one would expect to see a hotspot shift to the right towards 100 nm on the scatterplot for 4-event traces.

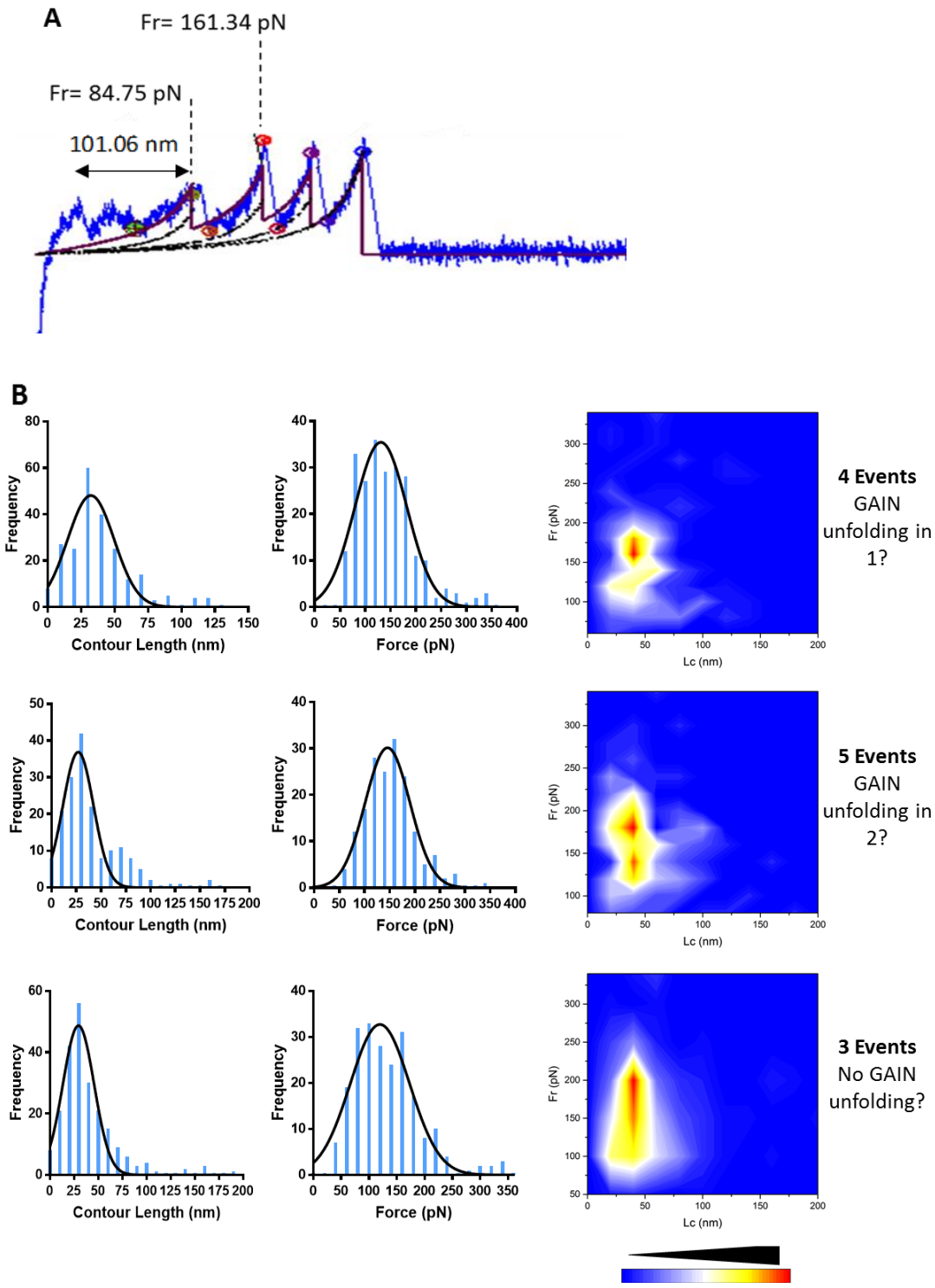


Figure 5.14 ***SMFS of EMR1 cleavage-deficient GAIN domain*****.** (**A**) A representative force-extension profile for EMR1 AFM experiments. (B) Contour length and rupture force data was collected and plotted on histograms, then fit to a Gaussian-distribution. Contour length values were calculated as ΔLc before plotting. N=3.

Initial SMFS AFM of cleavage-deficient Lphn1 demonstrated a relatively low rate of protein pick up. Force-extension profiles with 4 events were most commonly observed. Like EMR1, a domain unfolding event was observed at a lower force than that previously demonstrated for I_{27} but again the contour length does not equate to the full-length GAIN domain of 114.8 nm (Figure 5.15). Unexpectedly, the length-frequency histogram indicates a wide range of domain lengths, despite the rupture force distribution having a lesser spread. Furthermore, a distinct hotspot is not clear for 3 event traces, which would otherwise be expected for unfolding of the two I_{27} reference domains.

Overall, although the analysis of the force-extension traces recorded for EMR1, Lphn1 and EMR2 showed some traces consistent with complete or partial unfolding of the GAIN domain, the low frequency of these observed events and the lack of clear hotspots in the scatterplots indicates it is not possible to make a conclusive interpretation of the data. Experiments will need to be repeated to be able to define the mechanical properties of adhesion GPCR GAIN domains.

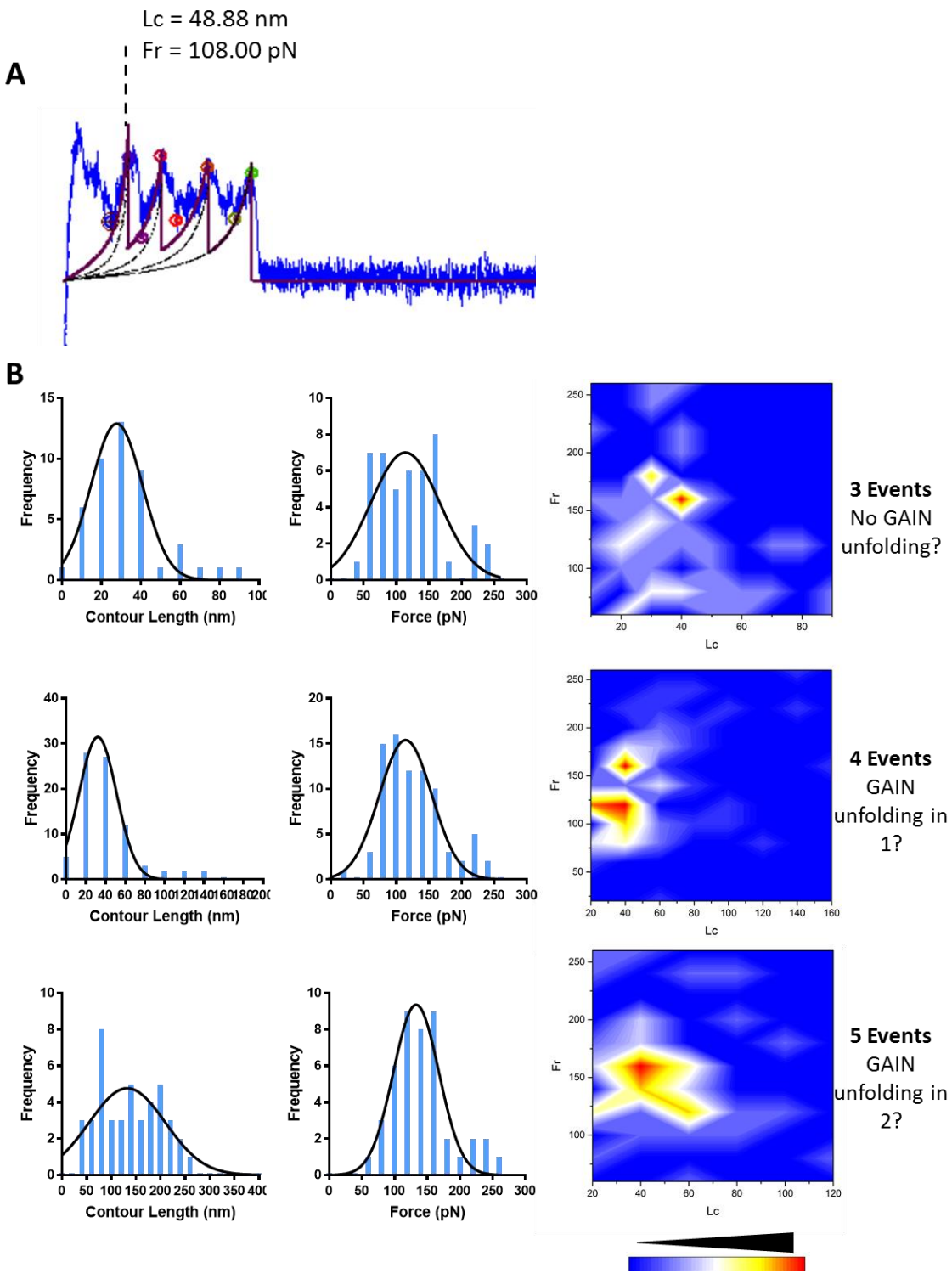


Figure 5.15 ***SMFS of cleavage-deficient Lphn1 GAIN domain*****.** (**A**) A representative force-extension profile for Lphn1 AFM experiments. (**B**)Contour length and rupture force data was collected and plotted on histograms, then fit to a Gaussian-distribution. Contour length values were calculated as ΔLc before plotting. The data represents that from 3 repeat experiments.

5.4 Discussion and future perspectives

Mechanical forces play a fundamental role in a plethora of biological processes. They influence changes in gene transcription, cell morphology and ECM remodelling upon cell-sensing of mechanical cues in their microenvironment. Mechanically sensitive proteins, such as ion channels, facilitate the transduction of mechanical signals across the membrane (274). Ionic currents perceived by ionotropic receptors have been shown to be moderated by the *Drosophila* homolog of adhesion GPCR Latrophilin, *dCirl,* impacting on action potential initiation by neurons (97). Other members of the adhesion GPCR family are increasingly implicated in other mechanosensory processes, such as transduction of sound waves by VLGR1, hypertrophic response to pressure in skeletal muscle by GPR56, and regulation of mast cell degranulation by EMR2. Perturbation of these processes cause pathophysiological states including Usher's syndrome and vibratory urticaria for VLGR1 and EMR2 respectively. Measuring the forces both experienced and exerted by biological molecules provides insights on intra- and intermolecular interactions. Therefore, understanding the mechanical regulation of adhesion GPCRs and their interactions is critical for understanding their activity, and in the design of novel therapeutics.

Due to a significant lack of exemplified ligand-induced activation, it is thought that mechanical force may be required to mediate structural rearrangements within the extracellular domain. This would serve to expose the so-called "*Stachel*" agonistic peptide for receptor activation. An example of another protein undergoing such structural changes is fibronectin. Stretching of this protein, which is pivotal in mechanotransduction, exposes multiple cryptic binding sites. This activates fibronectin to stimulate fibril formation when subject to cell contractile forces, modulating its activity (275). Consequently, the experiments carried out in this chapter aimed to elucidate the mechanical characteristics of adhesion GPCRs to provide molecular insights as to their means of activity.

Investigation of the mechanical properties of biological molecules can be approached by various techniques. AFM, optical and magnetic tweezers, and patch clamping are commonly chosen to observe the effect of an applied force or deformation of single molecules (276). AFM was chosen here, being a versatile tool that has been largely developed to probe the mechanical behaviour of multi-domain proteins, providing measurements of both resistance and elasticity in response to mechanical influence.

Without previous knowledge of a GAIN domain unfolding profile determined by AFM, an I_{27}(4)-HaloTag polyprotein was included in the investigation as a control. As this I_{27} polyprotein construct has a previously determined mechanical fingerprint (141)(146), identification of unknown domain unfolding is simplified. The data collected in this chapter from replicate experiments showed unfolding of I_{27} repeats in agreement with the literature; a single domain length of 28 nm at an average force of 175.5 pN. Consequently, these results could be applied for the identification of GAIN domain unfolding by incorporating a GAIN domain within the I_{27}(4)-HaloTag polyprotein.

Adhesion GPCR GAIN domains were subject to SMFS to quantify the inter- and intramolecular interactions responsible for structural and functional properties. The GAIN domains of adhesion GPCRs EMR2 and Lphn1 were investigated due to the involvement of these receptors in mechanosensitive processes. The wild type EMR1 GAIN domain was also investigated because it does not undergo cleavage at the GPS, and therefore may possess different mechanical properties to those that are naturally cleaved. As outlined in Section 5.2.2, it was predicted that the force-extension profile for the cleavage-deficient GAIN domain could appear as 0, 1 or 2 peaks, representative of no unfolding, or 1 or 2 step unfolding. With no recorded examples of adhesion GPCR GAIN domain force-extension profiles, traces were selected for analysis if they fit the WLC and included at least the appropriate number of I_{27} reference peaks. This allowed for the identification of peaks that could be attributed to the GAIN domain, rather than assumed as “noise”.

Experiments for EMR2*S518A* (Figure 5.11) yielded very little data that could be attributed to polyprotein unfolding. A large proportion of traces demonstrated non-specific interactions of the tip with the hard surface, exemplified in Figure 5.9C. Other traces showed 2 or 3 force peaks (data not shown), indicating cantilever interaction had occurred part way down the polypeptide chain instead of at the N-terminus. A small percentage of EMR2*S518A* traces displayed the unfolding of all I_{27} domains, with 2 preceding events occurring at a lower force (Figure 5.12). The observed contour lengths were approximately those expected for extension of each subdomain of the EMR2 GAIN domain (Table 5.1). However, the reproducibility of these traces was poor and therefore it was difficult to draw a clear conclusion.

In initial experiments, the cantilever-protein interaction was mediated by adsorption, which is not necessarily specific in its point of contact, meaning the protein can be pulled from any point along the polypeptide chain (145). To circumvent this issue, the cantilever was functionalised with Tris-NTA to facilitate specific interaction with the

N-terminal 10xHisTag, providing control over the pick-up point and increasing the likelihood of end-to-end extension (277). Despite this, similar results were gained with <0.1% of traces demonstrating I_{27} domain unfolding. Most approach-retract cycles resulted in no protein interaction at all. As experiments performed using the control polyprotein I_{27}(4)-HaloTag were deemed successful, the lack of protein unfolding events may be a consequence of inconsistent protein distribution across the functionalised solid surface.

Another possibility for the lack of GAIN domain unfolding events may be that the force required for domain unfolding is higher than the strength of interaction between Tris-NTA and 10xHisTag. However, whilst previously performed experiments by Tang et al. (2009) indicated an unbinding force of 160 pN (278), our results contrast this. The data presented in Figure 5.13, shows the force distribution curve for unfolding of all 4 I_{27} repeats, before tip unbinding. It indicates that the unfolding force for the I_{27} domain was often more than 160 pN, with an average unfolding force of 171.4 pN. This suggests that the interaction between the 10xHis-Tag and Tris-NTA is able to withstand larger force, and therefore unlikely that unbinding of the polyprotein and AFM cantilever occurred before protein unfolding.

Subsequent experiments investigated the unfolding of EMR1 and Lphn1 GAIN domains, with a single I_{27} domain either side of the specific GAIN domain and the C-terminal HaloTag. A similar frequency of 3, 4 and 5-peak force-extension profiles fitting to the WLC were observed for EMR1. A predominant force hotspot was found to be 150-200 pN and a delta contour length of 25-50 nm (Figure 5.14). Despite this, the observed contour length before the first of the I_{27} unfolding events did not match the expected value of 113.2 nm (data not shown). Therefore, it was not possible to attribute this to specific domain unfolding. If the GAIN domain was indeed unfolding before the I_{27} repeats, one would expect to see a constant contour length before the first I_{27}.

Data acquired for Lphn1 GAIN domain unfolding with two I_{27} domains also revealed a hotspot with a rupture force between 125-175 pN and a ΔLc of approximately 40 nm (Figure 5.15). However, the contour length preceding the first I_{27} unfolding event was again varied, and thus attributing the event to a specific domain unfolding was not possible. Therefore, the data presented in Figures 5.14 and 5.15 cannot be unequivocally assigned to unfolding of EMR1 or Lat1 GAIN domain respectively.

A limitation of AFM-SMFS is the time-consuming recording of thousands of force-extension traces, often with very little conclusive data gained for describing the

mechanical properties or behaviours of single molecules. Typically, a high volume of traces with significant events are acquired, and traces can be overlaid to indicate the reproducibility of the observed unfolding events. Despite this, AFM-based SMFS has been proven a suitable technique to investigate the interactions and properties of a multitude of membrane proteins, including class A GPCRs. GPCRs have been probed within membrane sections, rather than as part of a polyprotein as done in this chapter, and stabilising regions of both bovine rhodopsin and β2-adrenergic receptor have been identified by AFM (279). Receptor stability was found to be mediated by covalent interactions between transmembrane helices, amino acid loops, and by post-translational modifications like palmitoylation (279). Furthermore, the presence of a highly conserved disulphide bond in both receptors was highlighted as a crucial structural element, and in its absence, intramolecular interactions were shifted to stabilise different structural segments. This could be further explored to investigate receptor dysfunction in its absences and demonstrates that AFM-SMFS may be able to determine the stabilising effect of small molecule therapeutics for other GPCRs (280).

SMFS has also been exploited in combination with high-resolution AFM imaging to characterise and interrogate the surface structure and molecular interactions that mediate ligand binding and protein activation of the transmembrane ion transporter, NhaA. Experiments were able to identify molecular interactions responsible for receptor activation (281). Therefore, despite the need for high throughput instrumentation and data processing, the future investigation of adhesion GPCRs expressed and probed within intact membranes may yield more biologically relevant and insightful information.

The secondary structure of proteins has been shown to be a key determinant of the mechanical stability of proteins; α-helical structures have little mechanical strength, whilst a number of studies have demonstrated significant mechanical strength by β-sheet proteins. These include E2lip3, a lipoyl domain of pyruvate dehydrogenase (147), and immunoglobulin domains found in muscle (I_{27}) and fibronectin III (282), mediated by the presence of hydrogen bonds between strands (276). The secondary structure of the GAIN domain indicates it is composed of two-subdomains, with the tethered agonist deeply embedded within subdomain B (Figure 1.6C). This domain is composed of a twisted β-sandwich (26), characterised by two opposing antiparallel β-sheets, and therefore has the potential to be mechanically robust. Moreover, the presence of two disulphide bonds in the GAIN domain further implies protein stability. As several adhesion GPCRs are exposed to shear flow, this promotes the idea that

the GAIN domain likely withstands substantial mechanical force to prevent constant *Stachel* exposure. For example, vascular endothelial cells have been shown to be exposed to up to a total of 1 nN of fluid shear stress (116). In the results presented in this chapter, unfolding of I_{27} within the polyprotein consistently occurred between 150-200 pN (271), whilst final unbinding of the polyprotein from the cantilever occurred on average at 220 pN, 172 pN and 156 pN for EMR2, EMR1 and Lphn1 respectively (data not shown). It therefore may be possible that unfolding of the GAIN domain requires a force higher than that mediating unbinding in this instance. Certainly, in the SMFS study of bovine rhodopsin, applied forces ranging from 2-4 nN were required to rupture covalent bonds between the extracellular side of TM3 and the proximal extracellular loop (279).

Further studies are required to delineate the mechanical properties of the GAIN domain and further understand its role in mechanosensory processes. S Boyden et al. (2016) revealed that wild type EMR2, as well as the disease causing variant C492Y, was susceptible to vibration and able to mediate mast cell degranulation (46). Intriguingly, double mutant S518A-C492Y was unable to illicit degranulation. This indicates receptor cleavage is imperative and suggests NTF removal is required to promote degranulation. AFM has previously been exploited to interrogate the effects of pathogenic amino acid substitutions and quantify receptor stability as a result (279). Therefore, if fully optimised, AFM could be exploited to provide insights into the mechanical properties of wild type and variant adhesion GPCRs, and the subsequent functional implications. The work outlined here is preliminary, and therefore future studies should look at repeating and optimising the experiments to determine a force-extension profile for adhesion GPCR GAIN domains. In order to streamline the analysis of SMFS data collection, coding should be employed to circumvent the need for manual inspection and analysis of each individual force-extension profile.

Other biophysical techniques could also be employed to manipulate adhesion GPCRs and investigate their mechanical properties. Optical tweezers, also referred to as optical trapping, involves “trapping” a bead between two lasers, which interacts with the molecule of interest, and the amount of force required to manipulate the molecule can be measured by displacement of the bead in the optical trap. The use of magnetic tweezers is very similar, instead relying on external magnets to apply a force to a single molecule. Optical and magnetic tweezers can be employed in the study of small forces, such as the force applied by movement of a polymerase along a strand of DNA (283), and also in the study of larger forces including the mechanical

properties of human titin (284), the interaction strength between integrin-fibrinogen pairs on cell surfaces (285), and the movement of kinesin down microtubules (286). However, a notable advantage of AFM in comparison to these techniques is the potential to carry out measurements under near physiological conditions (287).

Further development of the single molecule experiments of adhesion GPCRs is required to define and understand their mechanical properties, such as performing pulling experiments at different speeds. This will allow a dynamic force spectrum to be determined based on rupture force versus loading rate, therefore helping to understand how adhesion GPCRs may function under different physiological forces. AFM can also be performed on live cells that are adhered to the glass surface, and on segments of intact membranes containing receptors. Single cells expressing receptors of interest can be manipulated by force spectroscopy to measure adhesion forces, characterise ligand-receptor binding, and study the unfolding and possible refolding of single proteins (287). This increases the physiological relevance as the conditions typically experienced by membrane proteins can be replicated and performed in physiological buffers with controlled pH. Aside from biophysical techniques, others could be employed to investigate the mechanical stability of the GAIN domain. With the expression of so many adhesion GPCRs being on the surface of immune cells, a fluid-flow chamber may be a very appropriate technique, as has been done previously for CD97 (93). The experiments presented in this chapter lay the foundations of a much larger body of work required to determine mechanical properties and the energy landscapes required for GAIN domain unfolding. This will perhaps aid in the further understand the structural rearrangements occurring to mediate *Stachel*-dependent receptor activation.

Chapter 6. General discussion

The adhesion GPCRs have now become well established in fundamental physiological processes, including CNS development, cell migration and immune regulation. Members of this novel receptor family are also implicated in a spectrum of pathologies, such as cancer metastasis, Usher's syndrome and bilateral frontoparietal microgyria. However, only a small proportion of these receptors have characterised binding partners, and there are fewer known instances of ligand-induced signalling. This has greatly hindered clarification of a definitive mechanism of action and the subsequent signalling pathways of these noteworthy and intriguing receptors.

Despite physiological communication being typically associated with receptors identifying ligands in order to mediate appropriate biological responses, the sensing of intra- and extracellular mechanical forces is imperative for proper tissue development and homeostasis. Interestingly, a growing number of studies have recently described a role for adhesion GPCRs in mechanosensitive processes, highlighting the possibility that some form of mechanical stimulation is an important factor in adhesion GPCR activation.

6.1 Adhesion GPCRs as mechanosensors

The work presented in Chapter 4 validates previous studies implicating GPR56 as a mechanosensitive receptor (114). The results show GPR56 activation and coupling to $G\alpha_{12/13}$ under vibration for the first time. This excitingly reflects published data revealing the sensitivity of another adhesion GPCR, EMR2, to vibration, a variant of which has been determined as the cause of a rare genetic urticaria by inducing mast cell degranulation (46). EMR2 NTF removal was largely detected after vibration under experimental conditions, contrasting low level dissociation for the wild type receptor. This suggests that receptor activity and subsequent signalling is mediated by exposure of the *Stachel* peptide, which complements the proposed tethered agonist mode of activation outlined in Section 1.2.3.

The significant 3-fold increase in GPR56 FL signalling under vibration in comparison to that of the unstimulated receptor (Section 4.12), suggests NTF dissociation has occurred, facilitating exposure of the tethered agonist. This result is consistent with published data by Nien-Yi Chiang et al. (2017), in which NTF removal activates $G\alpha_{12/13}$ and RhoA signalling to induce melanoma cell migration (158). Conversely, signalling by GPR56 ΔNTF in the results presented here was consistently lower than that of the full-length receptor, which could demonstrate different signalling modalities

of the receptor. However, this greatly contrasts a number of published reports that GPR56 ΔNTF, and other adhesion GPCR ΔNTF constructs, exhibit much higher signalling potential (50)(53)(54). It is therefore possible that these constructs were not properly trafficked to the cell surface, and hence had less surface expression than the full-length receptor, thus reducing their apparent signalling potential. Consequently, future experiments in this instance should look to confirm the level of surface expression of both full-length and ΔNTF GPR56.

S Boyden et al. (2016) demonstrated that EMR2 NTF dissociation was both vibration and cleavage dependent. The wild type receptor demonstrated reduced NTF removal in comparison to the C492Y-disease causing variant, but increased removal compared to that of cleavage-deficient S518A. Furthermore, the level of NTF dissociation of the double mutant C492Y-S518A mirrored that of the empty vector, confirming that cleavage is necessary for subunit dissociation (46). Similar experiments should be carried out to determine whether the nature of vibration-induced GPR56 signalling is dependent on receptor cleavage and/or the *Stachel* sequence. Future experiments would include cleavage-deficient mutants, and mutants of key residues within the *Stachel* peptide. This would show whether cleavage and the *Stachel* peptide is required for signalling and determine whether the complete removal of the NTF is required for signalling to occur.

Ligand binding may also be necessary to mediate dissociation of the NTF. In the instance of EMR2, incubating mast cells expressing EMR2 on plates coated with its endogenous ligand dermatan sulphate, or the monoclonal antibody 2A1, stimulated significant degranulation under vibration (46). The level of degranulation was notably less in the absence of ligand binding, though still upregulated compared to EMR2 signalling under static conditions. As dermatan sulphate is the major glycan present in skin (288), it is likely this binding contributes to dissociation in the physiological setting.

GPR56 is well characterised as a neuronal cell receptor, and binds to type III collagen which is highly expressed in the meninges and pial basement membrane in the CNS (86). Interestingly, GPR56 is implicated in neuronal cell migration (102), and as such is routinely exposed to shear stress. It is therefore possible that binding of this receptor-ligand pair occurs in tandem with mechanical stress. The collagen III-GPR56 interaction has been demonstrated to inhibit neuronal migration via activation of RhoA, but the role of mechanical stimuli was not considered in the study (64). On the other hand, the results demonstrated in Figure 4.12B confirm that $G\alpha_{12/13}$, the upstream activator of RhoA, is activated by GPR56 in response to vibration. Taken

with the collagen III binding study, it seems likely that ligand binding in combination with mechanical stress acts to potentiate downstream signalling cascades. Future experiments interrogating GPR56 mutants lacking the collagen-binding domain or chimeric receptors of other adhesion GPCRs containing the GPR56 collagen-binding domain, may shed light on the ligand-dependency of vibration-induced signalling.

Figures 4.11 and 4.12 illustrate minimal change in signalling under vibration by CD97 and GPR97. However, the results of Figure 4.12 do show for the first time a robust signalling output for GPR56, and therefore provide the basis for future ligand-induced or vibration studies. Whilst CD97 NTF shedding has been shown as a result of binding with CD55 under shear conditions (93), the signalling implications were not been assessed. Recent published work however, shows that CD97 undergoes mechano-dependent phosphorylation at its intracellular PDZ-motif, resulting in loss of contact with cytoskeletal proteins (289). Therefore, also evaluating transient changes in GPR56 phosphorylation could outline a mechanism for vibration-dependent signalling. Future experiments should build on the experiments outlined in Chapter 4 to investigate the necessity of mechanical stimulation and ligand binding in receptor activation. Expression of both wild-type and variants of GPR56 with altered binding sites will highlight the importance of ligand binding, whilst detection of the level of NTF in culture supernatant will help determine the mode of activation, be it either by subunit shedding to expose the cryptic *Stachel* peptide, or structural rearrangements reflecting disinhibition.

Whilst GPR56 has previously been implicated as a mechanosensor in skeletal muscle (Section 1.2.9)(114), the finding in Chapter 4 outlines a role for GPR56 as a mechano-transducer in a previously undefined mechanical setting, with a robust molecular readout. As members of the adhesion GPCRs, such as Latrophilin 1 and EMR2, are also defined as mechanoreceptors, it is highly likely others have currently unknown roles to play in mechanobiology. The assay developed here provides a simple and quick assessment of this potential for all adhesion GPCRs and will therefore help further our understanding of this enigmatic receptor family.

6.2 Structural basis of adhesion GPCR activity

Over recent years, our understanding surrounding the structural characteristics of adhesion GPCRs has greatly developed, particularly with the advent of the GAIN domain crystal structures of Latrophilin 1, BAI3 and GPR56. As a result, the *Stachel* peptide was identified as a key regulator of adhesion GPCR activation. However, a

complete elucidation of protein activation remains a major challenge. Furthermore, with the increasing implication of force-sensing, it seems much more work is required to fully understand the role of the highly conserved GAIN domain in the regulation of adhesion GPCR function.

Alignment of the predicted secondary structures based on sequence alignment have highlighted notable differences in structure between a number of adhesion GPCR GAIN domains, particularly between members of the EGF-TM7 subfamily and the wider family. It is predicted that they lack the majority, if not all, of the N-terminal α-helices within the domain, suggesting that the subdomains of the GAIN domain infer different functional properties. This body of work therefore endeavoured to solve the atomic structure of an example EGF-TM7 GAIN domain. However, a multitude of attempts at expression, purification, crystallisation and subsequent efforts towards obtaining the experimental phases of CD97 GAIN were unsuccessful. It was not possible to determine the atomic structure, and therefore efforts could not be translated to the understanding of the functional roles of the GAIN subdomains. The generation of 2.2 Å native dataset for the GAIN domain outlined in Chapter 3 does however provide a key step toward the future elucidation of the crystal structure of the CD97 GAIN domain. The structure of this domain will not only provide information on the first GAIN domain with a known ligand, Thy1/CD90 (110), but also provide structural information on the homologous domains of the EGF-TM7 subfamily of adhesion GPCRs for which there is, as of yet, no information.

As it has been reiterated throughout this work, these multi-faceted receptors likely also function as mechano-transducers. Therefore, the mechanical properties of the GAIN domain were investigated using AFM, a mainstay technique in decrypting protein mechanical strength. Following a considered and relevant approach, using highly characterised protein reference domains to identify the presence of novel protein unfolding events, a small number of force-extension traces were obtained showing GAIN domain unfolding in one or two steps. However, the low frequency of these traces amongst the large number of experimentally collected traces indicates much more data will be required alongside this preliminary data in order to gain insight into the mechanical structural properties of the GAIN domain.

The physiological environments in which the adhesion GPCRs are expressed imply these receptors are mechanically stable. Indeed, when the extracellular GAIN domain is perturbed, as demonstrated by the EMR2 C492Y variant, typical receptor function is altered, resulting in the development of disease. The mutation of cysteine to tyrosine at position 492 located within the GAIN domain seemingly destabilises the

non-covalent interactions between the NTF and CTF. This undoubtedly advocates a role of mechanical stability for regulating the activity of these receptors. Therefore, future work towards determining the atomic structure of other GAIN domains will facilitate the identification of key residues for receptor activity and ligand binding sites, the implications of which could be interrogated in signalling assays. Increased knowledge of the potential binding pockets will help further the design of pharmaceutical small molecules. Furthermore, such findings alongside currently known disease-causing mutations within the GAIN domain could be taken forward to SMFS studies and investigated with a view to defining their mechanical influence and significance. Fathoming the molecular mechanism of adhesion GPCR activation under mechanical influence will undeniably provide insights into the pathophysiological roles of adhesion GPCRs.

6.3 Characterising adhesion GPCR signalling

The work carried out in this thesis has successfully shown GPR97 coupling to $G\alpha_{12/13}$ via SRF-RE luciferase reporter assays. GPR56 was used as a suitable control due to already having a characterised signalling output via coupling with $G\alpha_{12/13}$. Full length GPR97 demonstrated basal signalling activity that is a typical characteristic of GPCRs (290). Although GPR97 has previously been identified as a modulator of RhoA GTPase, it has not been documented to couple to $G\alpha_{12/13}$. However, it is not entirely unexpected due to the influence GPR97 has on lymphatic endothelial cell migration, controlled by RhoA and Cdc42 (246). $G\alpha_{12/13}$ can bind and activate Rho-specific GTPases to stimulate RhoA and its downstream response (291). GPR97 is also expressed on inflammatory macrophages (113), thus the expression pattern of this receptor indicates, like GPR56, it is constantly exposed to shear stress. Therefore, despite not having a documented binding partner, it appears that GPR97 may also be a mechanosensitive receptor, with sensing of mechanical forces enabling the regulation of endothelial cell migration, as we hypothesise for GPR56 and neuronal cells.

A constant hinderance in understanding receptor action is a lack of known ligands. As a large proportion of the adhesion GPCR family are orphan receptors like GPR97, this thesis aimed to circumvent this issue by the design of highly specific binding proteins. Small molecules or protein binding partners that can interact with the extracellular or transmembrane domain of adhesion GPCRs, to mimic ligand binding, provide an ideal alternative. Affimers, being scaffold proteins that are incredibly stable and simple to produce against specific protein domains (Section 1.4), have

been previously exploited to modulate the activity of cell surface vascular endothelial growth factor receptors (161). Therefore, after having characterised a signalling output for GPR97, Affimers were generated against the GAIN domains of GPR97 and GPR56 to test their validity as surrogate ligands. In addition, Affimers generated primarily for use as crystallisation chaperones for CD97 GAIN domain (Section 3.3.4) were also investigated.

Of the Affimers demonstrated to bind the GPR97 GAIN domain, signalling was not significantly modified. Similarly, those raised against CD97 had little effect on receptor signalling. In this case however, Affimers were raised against a bacterially expressed GAIN domain. Mammalian CD97 GAIN domain is highly glycosylated, a post-translational modification that bacteria are unable to perform. It is therefore possible that the original Affimer binding sites were subsequently masked in the GAIN domain expressed by mammalian cells.

However, GPR56 signalling was significantly perturbed by treatment with Affimer 1 at a final concentration of 0.4 µM, indicating a more potent action than that of published monobody action against GPR56 at 0.7 µM (255). This exciting result validates the use of Affimers as surrogate ligands, which can be readily used to decipher the signalling modalities of other receptors in the adhesion GPCR family. Interestingly, Affimer 1 was derived from the “single loop” phage library. Therefore, there is scope to produce small molecule mimetics or cyclic peptides corresponding to the 9-amino acid binding loop within Affimer 1. If effective, these smaller molecules will be more easily produced, have better understood pharmacokinetics and may have better tissue penetrance than the protein based parental Affimer.

Probing receptor activation of signalling pathways is crucial for determining the influence of receptors on disease states. CD97 stimulation of the serum response signalling pathway for example, is central to the receptor’s pivotal role in cancer invasion of healthy tissues. CD97 influences metastatic properties and is upregulated at the invasive front of tumours, therefore understanding the mechanistic activation of SRE and SRF by CD97 will prove highly valuable in the development of therapeutics. As small binding proteins have now been documented by this work and G Salzman et al. (2017)(151) to both exacerbate and inhibit signalling, they may provide an invaluable tool in directing mechanistic and pharmacological studies of adhesion GPCRs and could be used to target known pathological activities. Furthermore, the establishment of simple yet effective receptor signalling reporter assays, as demonstrated in this work, offers the opportunity for dissecting functional

outputs of natural receptor variants, and their implications in the presentation of clinical symptoms.

6.4 Conclusion

To conclude, the work undertaken in this thesis has generated a simple and robust functional readout for assessing receptor activity. Importantly, it has enabled us to describe the coupling of orphan adhesion GPCR, GPR97, to the heterotrimeric G protein $G\alpha_{12/13}$, confirming classical GPCR activity. It has also generated a small binding protein, capable of downregulating GPR56 signalling in a receptor-specific manner, suggesting the adhesion GPCRs are likely druggable targets. Most excitingly, this study has contributed to the growing observation of adhesion GPCRs as mechanosensors, outlining GPR56-mediated signalling as a mechanosensitive process.

Currently, a high proportion of FDA-approved drugs target only a handful of well-defined members of the GPCR superfamily, highlighting the importance of receptor characterisation and the opportunities it presents in therapeutic development. The Affimer technology used in this thesis ensures high specificity targeting. As seven-transmembrane domain proteins are the most abundant receptor family in humans, this is an ideal attribute for pharmacological design of small molecules to ensure cross-reactivity and non-specific effects are not encountered.

It is possible that the methodologies optimised in this current study will be useful for the entire family of adhesion GPCRs. Critically, this would significantly increase our knowledge at the molecular level for the mechanism of adhesion GPCR activation, ultimately paving the way for the design of potential therapeutics against this physiologically and pathologically relevant group of receptors.

References

1. Kroeze WK, Sheffler DJ, Roth BL. G-protein-coupled receptors at a glance. J Cell Sci. The Company of Biologists Ltd; 2003 Dec 15;116(Pt 24):4867–9.
2. Schiöth HB, Fredriksson R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. Gen Comp Endocrinol. 2005;142:94–101.
3. Civelli O, Reinscheid RK, Zhang Y, Wang Z, Fredriksson R, Schiöth HB. G Protein–Coupled Receptor Deorphanizations. Annu Rev Pharmacol Toxicol. 2012;53:121002130630009.
4. Hauser AS, Chavali S, Masuho I, Jahn LJ, Martemyanov KA, Gloriam DE, et al. Pharmacogenomics of GPCR Drug Targets. Cell. Elsevier; 2018 Jan 11;172(1–2):41–54.e19.
5. Tang X, Wang Y, Li D, Luo J, Liu M. Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets. Acta Pharmacol Sin. 2012;33:363–71.
6. Cardoso JCR, Félix RC, Fonseca VG, Power DM. Feeding and the Rhodopsin Family G-Protein Coupled Receptors in Nematodes and Arthropods. Front Endocrinol (Lausanne). Frontiers; 2012 Dec 18;3:157.
7. Kobilka BK, Deupi X. Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci. Elsevier Current Trends; 2007 Aug 1;28(8):397–406.
8. Latorraca NR, Venkatakrishnan AJ, Dror RO. GPCR Dynamics: Structures in Motion. Chem Rev. American Chemical Society; 2017 Jan 11;117(1):139–55.
9. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S. Action of molecular switches in GPCRs--theoretical and experimental studies. Curr Med Chem. Bentham Science Publishers; 2012;19(8):1090–109.
10. Deupi X, Standfuss J. Structural insights into agonist-induced activation of G-protein-coupled receptors. Curr Opin Struct Biol. Elsevier Current Trends; 2011 Aug 1;21(4):541–51.
11. Shoichet BK, Kobilka BK. Structure-based drug screening for G-protein-coupled receptors. Trends Pharmacol Sci. Elsevier Ltd; 2012;33(5):268–72.
12. Ritter SL, Hall R a. Fine-tuning of GPCR activity by receptor-interacting proteins. Mol Cell. 2010;10(12):819–30.
13. Katritch V. Structure-function of the G protein-coupled receptor superfamily. Annu Rev …. 2013;(7):531–56.
14. Koelle MR. Heterotrimeric G Protein Signaling: Getting inside the Cell. Cell. Elsevier; 2006 Jul 14;126(1):25–7.
15. Oldham WM, Hamm HE. Heterotrimeric G protein activation by G-protein-coupled receptors. Nat Rev Mol Cell Biol. Nature Publishing Group; 2008 Jan 1;9(1):60–71.
16. Duc NM, Kim HR, Chung KY. Recent Progress in Understanding the Conformational Mechanism of Heterotrimeric G Protein Activation. Biomol Ther (Seoul). Korean Society of Applied Pharmacology; 2017 Jan 1;25(1):4–11.
17. Simon MI, Strathmann MP, Gautam N. Diversity of G proteins in signal transduction. Science. 1991 May 10;252(5007):802–8.
18. McCudden CR, Hains MD, Kimple RJ, Siderovski DP, Willard FS. G-protein signaling: back to the future. Cell Mol Life Sci. Springer; 2005 Mar;62(5):551–77.
19. Rajagopal S, Shenoy SK. GPCR desensitization: Acute and prolonged phases. Cell Signal. Pergamon; 2018 Jan 1;41:9–16.

20. Lappano R, Maggiolini M. G protein-coupled receptors: novel targets for drug discovery in cancer. Nat Rev Drug Discov. 2011;10(1):47–60.
21. Sánchez-Fernández G, Cabezudo S, García-Hoz C, Benincá C, Aragay AM, Mayor F, et al. Gαq signalling: The new and the old. Cell Signal. Pergamon; 2014 May 1;26(5):833–48.
22. Shan D, Chen L, Wang D, Tan Y-C, Gu JL, Huang X-Y. The G Protein Gα13 Is Required for Growth Factor-Induced Cell Migration. Dev Cell. Cell Press; 2006 Jun 1;10(6):707–18.
23. Hildebrandt JD. Role of subunit diversity in signaling by heterotrimeric G proteins. Biochem Pharmacol. Elsevier; 1997 Aug 1;54(3):325–39.
24. Shenoy SK, Lefkowitz RJ. β-arrestin-mediated receptor trafficking and signal transduction. Trends Pharmacol Sci. Elsevier Current Trends; 2011 Sep 1;32(9):521–33.
25. Goodman OB, Krupnick JG, Santini F, Gurevich V V., Penn RB, Gagnon AW, et al. β-Arrestin acts as a clathrin adaptor in endocytosis of the β2-adrenergic receptor. Nature. Nature Publishing Group; 1996 Oct 3;383(6599):447–50.
26. Araç D, Boucard A a, Bolliger MF, Nguyen J, Soltis SM, Südhof TC, et al. A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. EMBO J. 2012 Mar 21;31(6):1364–78.
27. Prömel S, Langenhan T, Araç D. Matching structure with function: the GAIN domain of adhesion-GPCR and PKD1-like proteins. Trends Pharmacol Sci. 2013 Aug;34(8):470–8.
28. Huang Y-S, Chiang N-Y, Hu C-H, Hsiao C-C, Cheng K-F, Tsai W-P, et al. Activation of myeloid cell-specific adhesion class G protein-coupled receptor EMR2 via ligation-induced translocation and interaction of receptor subunits in lipid raft microdomains. Mol Cell Biol. 2012 Apr;32(8):1408–20.
29. Yona S, Lin H-H, Siu WO, Gordon S, Stacey M. Adhesion-GPCRs: emerging roles for novel receptors. Trends Biochem Sci. 2008 Oct;33(10):491–500.
30. Liebscher I, Ackley B, Araç D, Ariestanti DM, Aust G, Bae B, et al. New functions and signaling mechanisms for the class of adhesion G protein-coupled receptors. Ann N Y Acad Sci. NIH Public Access; 2014 Dec;1333(1):43–64.
31. Safaee M, Clark AJ, Ivan ME, Oh MC, Bloch O, Sun MZ, et al. CD97 is a multifunctional leukocyte receptor with distinct roles in human cancers (Review). Int J Oncol. 2013;43(21):1343–50.
32. Swietnicki W. Folding aggregated proteins into functionally active forms. Curr Opin Biotechnol. 2006 Aug;17(4):367–72.
33. Araç D, Aust G, Calebiro D, Engel FB, Formstone C, Goffinet A, et al. Dissecting signaling and functions of adhesion G protein-coupled receptors. Ann N Y Acad Sci. 2012 Dec;1276:1–25.
34. Bjarnadóttir TK, Fredriksson R, Höglund PJ, Gloriam DE, Lagerström MC, Schiöth HB. The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. Genomics. Academic Press; 2004 Jul 1;84(1):23–33.
35. Liao Y, Pei J, Cheng H, Grishin N V. An Ancient Autoproteolytic Domain Found in GAIN, ZU5 and Nucleoporin98. J Mol Biol. Elsevier Ltd; 2014 Dec 12;426(24):3935–45.
36. Wei W, Hackmann K, Xu H, Germino G, Qian F. Characterization of cis-autoproteolysis of polycystin-1, the product of human polycystic kidney disease 1 gene. J Biol Chem. 2007;282(30):21729–37.
37. Chang G-W, Stacey M, Kwakkenbos MJ, Hamann J, Gordon S, Lin H-H. Proteolytic cleavage of the EMR2 receptor requires both the extracellular stalk and the GPS motif. FEBS Lett. 2003 Jul;547(1–3):145–50.
38. Lin H-H, Chang G-W, Davies JQ, Stacey M, Harris J, Gordon S. Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. J Biol Chem. 2004 Jul 23;279(30):31823–32.

39. Ward Y, Lake R, Martin PL, Killian K, Salerno P, Wang T, et al. CD97 amplifies LPA receptor signaling and promotes thyroid cancer progression in a mouse model. Oncogene. Nature Publishing Group; 2013 May 30;32(22):2726–38.
40. Tesmer JJG. A GAIN in understanding autoproteolytic G protein-coupled receptors and polycystic kidney disease proteins. EMBO J. European Molecular Biology Organization; 2012 Mar 21;31(6):1334–5.
41. Langenhan T, Aust G, Hamann J. Sticky signaling--adhesion class G protein-coupled receptors take the stage. Sci Signal. 2013 May 21;6(276):re3.
42. Torres VE, Harris PC. Autosomal dominant polycystic kidney disease: the last 3 years. Kidney Int. Nature Publishing Group; 2009;76(2):149–68.
43. Kurbegovic A, Kim H, Xu H, Yu S, Cruanès J, Maser RL, et al. Novel functional complexity of polycystin-1 by GPS cleavage in vivo: role in polycystic kidney disease. Mol Cell Biol. American Society for Microbiology (ASM); 2014 Sep;34(17):3341–53.
44. Ebermann I, Wiesen MHJ, Zrenner E, Lopez I, Pigeon R, Kohl S, et al. GPR98 mutations cause Usher syndrome type 2 in males. J Med Genet. BMJ Publishing Group Ltd; 2009 Apr 1;46(4):277–80.
45. Chiang N-Y, Hsiao C-C, Huang Y-S, Chen H-Y, Hsieh I-J, Chang G-W, et al. Disease-associated GPR56 mutations cause bilateral frontoparietal polymicrogyria via multiple mechanisms. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2011 Apr 22;286(16):14215–25.
46. Boyden SE, Desai A, Cruse G, Young ML, Bolan HC, Scott LM, et al. Vibratory Urticaria Associated with a Missense Variant in ADGRE2. N Engl J Med. NIH Public Access; 2016 Feb 18;374(7):656–63.
47. Ravenscroft G, Nolent F, Rajagopalan S, Meireles AM, Paavola KJ, Gaillard D, et al. Mutations of GPR126 are responsible for severe arthrogryposis multiplex congenita. Am J Hum Genet. Elsevier; 2015 Jun 4;96(6):955–61.
48. Aust G, Zhu D, Van Meir EG, Xu L. Adhesion GPCRs in Tumorigenesis. Handb Exp Pharmacol. NIH Public Access; 2016;234:369–96.
49. Soh UJK, Dores MR, Chen B, Trejo J. Signal transduction by protease-activated receptors. Br J Pharmacol. Wiley-Blackwell; 2010 May;160(2):191–203.
50. Liebscher I, Scho J, Monk KR, Scho T. A Tethered Agonist within the Ectodomain Activates the Adhesion G Protein-Coupled Receptors GPR126 Report A Tethered Agonist within the Ectodomain Activates the Adhesion G Protein-Coupled. Cell. 2014;2018–26.
51. Demberg LM, Rothemund S, Schöneberg T, Liebscher I. Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor GPR64/ADGRG2. Biochem Biophys Res Commun. Academic Press; 2015 Aug 28;464(3):743–7.
52. Demberg LM, Winkler J, Wilde C, Simon K-U, Schön J, Rothemund S, et al. Activation of Adhesion G Protein-coupled Receptors: Agonist specificty of Stchel sequence-derived peptides. J Biol Chem. 2017 Mar 17;292(11):4383–94.
53. Stoveken HM, Hajduczok AG, Xu L, Tall GG. Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. Proc Natl Acad Sci U S A. National Academy of Sciences; 2015 May 12;112(19):6194–9.
54. Kishore A, Purcell RH, Nassiri-Toosi Z, Hall RA. Stalk-dependent and Stalk-independent Signaling by the Adhesion G Protein-coupled Receptors GPR56 (ADGRG1) and BAI1 (ADGRB1) *. 2015;
55. Okajima D, Kudo G, Yokota H. Brain-specific angiogenesis inhibitor 2 (BAI2) may be activated by proteolytic processing. J Recept Signal Transduct Res. 2010;30(November 2009):143–53.
56. Paavola KJ, Sidik H, Zuchero JB, Eckart M, Talbot WS. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. Sci

Signal. 2014;7(338):ra76-ra76.
57. Hu Q-X, Dong J-H, Du H-B, Zhang D-L, Ren H-Z, Ma M-L, et al. Constitutive Gαi Coupling Activity of VLGR1 and its Regulation by PDZD7. J Biol Chem. 2014;289(35):24215–25.
58. Paavola KJ, Stephenson JR, Ritter SL, Alter SP, Hall R a. The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. J Biol Chem. 2011;286(33):28914–21.
59. Chen T-Y, Hwang T-L, Lin C-Y, Lin T-N, Lai H-Y, Tsai W-P, et al. EMR2 receptor ligation modulates cytokine secretion profiles and cell survival of lipopolysaccharide-treated neutrophils. Chang Gung Med J. 2011;34:468–77.
60. Prömel S, Waller-Evans H, Dixon J, Zahn D, Colledge WH, Doran J, et al. Characterization and functional study of a cluster of four highly conserved orphan adhesion-GPCR in mouse. Dev Dyn. Wiley-Blackwell; 2012 Oct;241(10):1591–602.
61. Scholz N, Gehring J, Kittel RJ, Langenhan T, Guan C, Ljaschenko D, et al. The Adhesion GPCR Latrophilin/CIRL Shapes Mechanosensation. Cell Rep. 2015;11:866–74.
62. Scholz N. Cancer Cell Mechanics: Adhesion G Protein-coupled Receptors in Action? Front Oncol. Frontiers Media SA; 2018;8:59.
63. Paavola KJ, Sidik H, Zuchero JB, Eckart M, Talbot WS. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. Sci Signal. NIH Public Access; 2014 Aug 12;7(338):ra76.
64. Luo R, Jin Z, Deng Y, Strokes N, Piao X. Disease-Associated Mutations Prevent GPR56-Collagen III Interaction. Mei L, editor. PLoS One. Public Library of Science; 2012 Jan 4;7(1):e29818.
65. Purcell RH, Hall RA. Adhesion G Protein–Coupled Receptors as Drug Targets. Annu Rev Pharmacol Toxicol. Annual Reviews ; 2018 Jan 6;58(1):429–49.
66. Stephenson JR, Paavola KJ, Schaefer S a., Kaur B, Van Meir EG, Hall R a. Brain-specific angiogenesis inhibitor-1 signaling, regulation, and enrichment in the postsynaptic density. J Biol Chem. 2013;288(31):22248–56.
67. Ward Y, Lake R, Yin JJ, Heger CD, Raffeld M, Goldsmith PK, et al. LPA receptor heterodimerizes with CD97 to amplify LPA-initiated RHO-dependent signaling and invasion in prostate cancer cells. Cancer Res. 2011;71:7301–11.
68. Gupte J, Swaminath G, Danao J, Tian H, Li Y, Wu X. Signaling property study of adhesion G-protein-coupled receptors. FEBS Lett. Federation of European Biochemical Societies; 2012 Apr 24;586(8):1214–9.
69. Peeters MC, Fokkelman M, Boogaard B, Egerod KL, van de Water B, IJzerman AP, et al. The adhesion G protein-coupled receptor G2 (ADGRG2/GPR64) constitutively activates SRE and NFκB and is involved in cell adhesion and migration. Cell Signal. Pergamon; 2015 Dec 1;27(12):2579–88.
70. Posokhova E, Shukla A, Seaman S, Volate S, Hilton MB, Wu B, et al. GPR124 Functions as a WNT7-Specific Coactivator of Canonical β-Catenin Signaling. Cell Rep. Cell Press; 2015 Jan 13;10(2):123–30.
71. van Pel M, Hagoort H, Hamann J, Fibbe WE. CD97 is differentially expressed on murine hematopoietic stem-and progenitor-cells. Haematologica. Haematologica; 2008 Aug 1;93(8):1137–44.
72. Wang T, Ward Y, Tian L, Lake R, Guedez L, Stetler-Stevenson WG, et al. CD97, an adhesion receptor on inflammatory cells, stimulates angiogenesis through binding integrin counterreceptors on endothelial cells. Blood. 2005;105(7):2836–44.
73. Bjarnadóttir TK, Geirardsdóttir K, Ingemansson M, Mirza MAI, Fredriksson R, Schiöth HB. Identification of novel splice variants of Adhesion G protein-coupled receptors. Gene. Elsevier; 2007 Jan 31;387(1–2):38–48.
74. Kwakkenbos MJ, Kop EN, Stacey M, Matmati M, Gordon S, Lin H-H, et al. The

EGF-TM7 family: a postgenomic view. Immunogenetics. 2004 Jan;55(10):655–66.
75. Matmati M, Pouwels W, van Bruggen R, Jansen M, Hoek RM, Verhoeven AJ, et al. The human EGF-TM7 receptor EMR3 is a marker for mature granulocytes. J Leukoc Biol. 2007 Feb;81(2):440–8.
76. Gordon S, Hamann J, Lin H-H, Stacey M. F4/80 and the related adhesion-GPCRs. Eur J Immunol. 2011 Sep;41(9):2472–6.
77. Stacey M, Chang G-W, Sanos SL, Chittenden LR, Stubbs L, Gordon S, et al. EMR4, a novel epidermal growth factor (EGF)-TM7 molecule up-regulated in activated mouse macrophages, binds to a putative cellular ligand on B lymphoma cell line A20. J Biol Chem. 2002 Aug 9;277(32):29283–93.
78. Davies JQ, Lin H-H, Stacey M, Yona S, Chang G-W, Gordon S, et al. Leukocyte adhesion-GPCR EMR2 is aberrantly expressed in human breast carcinomas and is associated with patient survival. Oncol Rep. 2011 Mar;25(3):619–27.
79. Hamann J, Aust G, Araç D, Engel FB, Formstone C, Fredriksson R, et al. International Union of Basic and Clinical Pharmacology . XCIV . Adhesion G Protein – Coupled Receptors. 2015;(April):338–67.
80. Peng Y-M, van de Garde MDB, Cheng K-F, Baars PA, Remmerswaal EBM, van Lier RAW, et al. Specific expression of GPR56 by human cytotoxic lymphocytes. J Leukoc Biol. Wiley-Blackwell; 2011 Oct 1;90(4):735–40.
81. Chang G-W, Hsiao C-C, Peng Y-M, Vieira Braga FA, Kragten NAM, Remmerswaal EBM, et al. The Adhesion G Protein-Coupled Receptor GPR56/ADGRG1 Is an Inhibitory Receptor on Human NK Cells. Cell Rep. Elsevier; 2016 May 24;15(8):1757–70.
82. Kwakkenbos MJ, Pouwels W, Matmati M, Stacey M, Lin H, Gordon S, et al. Expression of the largest CD97 and EMR2 isoforms on leukocytes facilitates a specific interaction with chondroitin sulfate on B cells role in the interaction of activated T cells. J Leukoc Biol. 2004;77:112–9.
83. He Z, Wu HUI, Jiao Y, Zheng JUN. Expression and prognostic value of CD97 and its ligand CD55 in pancreatic cancer. Oncol Lett. 2015;(9):793–7.
84. Xu L, Begum S, Hearn JD, Hynes RO. GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. Proc Natl Acad Sci U S A. National Academy of Sciences; 2006 Jun 13;103(24):9023–8.
85. Luo R, Jeong S-J, Yang A, Wen M, Saslowsky DE, Lencer WI, et al. Mechanism for Adhesion G Protein-Coupled Receptor GPR56-Mediated RhoA Activation Induced By Collagen III Stimulation. Zhang R, editor. PLoS One. Public Library of Science; 2014 Jun 20;9(6):e100043.
86. Luo R, Jeong S-J, Jin Z, Strokes N, Li S, Piao X. G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. Proc Natl Acad Sci U S A. National Academy of Sciences; 2011 Aug 2;108(31):12925–30.
87. Petersen SC, Luo R, Liebscher I, Giera S, Jeong S-J, Mogha A, et al. The Adhesion GPCR GPR126 Has Distinct, Domain-Dependent Functions in Schwann Cell Development Mediated by Interaction with Laminin-211. Neuron. 2015;85(4):755–69.
88. Jackson VA, del Toro D, Carrasquero M, Roversi P, Harlos K, Klein R, et al. Structural Basis of Latrophilin-FLRT Interaction. Structure. 2015;1–8.
89. Jackson VA, Mehmood S, Chavent M, Roversi P, Carrasquero M, del Toro D, et al. Super-complexes of adhesion GPCRs and neural guidance receptors. Nat Commun. Nature Publishing Group; 2016 Apr 19;7:11184.
90. Park D, Tosello-Trampont A-C, Elliott MR, Lu M, Haney LB, Ma Z, et al. BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. Nature. 2007;450(November):430–4.
91. Billings EA, Lee CS, Owen KA, D'Souza RS, Ravichandran KS, Casanova JE.

The adhesion GPCR BAI1 mediates macrophage ROS production and microbicidal activity against Gram-negative bacteria. Sci Signal. 2016 Feb 2;9(413):ra14.
92. Koh JT, Kook H, Kee HJ, Seo Y-W, Jeong BC, Lee JH, et al. Extracellular fragment of brain-specific angiogenesis inhibitor 1 suppresses endothelial cell proliferation by blocking αvβ5 integrin. Exp Cell Res. Academic Press; 2004 Mar 10;294(1):172–84.
93. Karpus ON, Veninga H, Hoek RM, Flierman D, van Buul JD, Vandenakker CC, et al. Shear stress-dependent downregulation of the adhesion-G protein-coupled receptor CD97 on circulating leukocytes upon contact with its ligand CD55. J Immunol. 2013 Apr 1;190(7):3740–8.
94. Hsiao C-C, Wang W-C, Kuo W-L, Chen H-Y, Chen T-C, Hamann J, et al. CD97 inhibits cell migration in human fibrosarcoma cells by modulating TIMP-2/MT1-MMP/MMP-2 activity--role of GPS autoproteolysis and functional cooperation between the N- and C-terminal fragments. FEBS J. 2014 Nov;281(21):4878–91.
95. Zyryanova T, Schneider R, Adams V, Sittig D, Kerner C, Gebhardt C, et al. Skeletal Muscle Expression of the Adhesion-GPCR CD97: CD97 Deletion Induces an Abnormal Structure of the Sarcoplasmatic Reticulum but Does Not Impair Skeletal Muscle Function. Csernoch L, editor. PLoS One. Public Library of Science; 2014 Jun 20;9(6):e100513.
96. Gargalionis AN, Basdra EK, Papavassiliou AG. Tumor mechanosensing and its therapeutic potential. J Cell Biochem. Wiley-Blackwell; 2018 Jun 1;119(6):4304–8.
97. Scholz N, Guan C, Nieberler M, Grotemeyer A, Maiellaro I, Gao S, et al. Mechano-dependent signaling by Latrophilin/CIRL quenches cAMP in proprioceptive neurons. Elife. eLife Sciences Publications Limited; 2017 Aug 8;6:e28360.
98. Liu D, Trojanowicz B, Ye L, Li C, Zhang L, Li X, et al. The invasion and metastasis promotion role of CD97 small isoform in gastric carcinoma. PLoS One. 2012 Jan;7(6):e39989.
99. Lin HH, Stacey M, Hamann J, Gordon S MA. Human EMR2, a novel EGF-TM7 molecule on chromosome 19p13.1, is closely related to CD97. Genomics. 2000;67(2):188–200.
100. Millar MW, Corson N, Xu L. The Adhesion G-Protein-Coupled Receptor, GPR56/ADGRG1, Inhibits Cell-Extracellular Matrix Signaling to Prevent Metastatic Melanoma Growth. Front Oncol. Frontiers Media SA; 2018;8:8.
101. Yang L, Chen G, Mohanty S, Scott G, Fazal F, Rahman A, et al. GPR56 Regulates VEGF production and angiogenesis during melanoma progression. Cancer Res. American Association for Cancer Research; 2011 Aug 15;71(16):5558–68.
102. Singer K, Luo R, Jeong S-J, Piao X. GPR56 and the developing cerebral cortex: cells, matrix, and neuronal migration. Mol Neurobiol. NIH Public Access; 2013 Feb;47(1):186–96.
103. Chang BS, Piao X, Bodell A, Basel-Vanagaite L, Straussberg R, Dobyns WB, et al. Bilateral frontoparietal polymicrogyria: Clinical and radiological features in 10 families with linkage to chromosome 16. Ann Neurol. Wiley-Blackwell; 2003 May 1;53(5):596–606.
104. Giera S, Deng Y, Luo R, Ackerman SD, Mogha A, Monk KR, et al. The adhesion G protein-coupled receptor GPR56 is a cell-autonomous regulator of oligodendrocyte development. Nat Commun. Nature Publishing Group; 2015 Dec 21;6(1):6121.
105. Iguchi T, Sakata K, Yoshizaki K, Tago K, Mizuno N, Itoh H. Orphan G protein-coupled receptor GPR56 regulates neural progenitor cell migration via a G alpha 12/13 and Rho pathway. J Biol Chem. 2008 May 23;283(21):14469–78.
106. Stephenson JR, Purcell RH, Hall R a. The BAI subfamily of adhesion GPCRs:

synaptic regulation and beyond. Trends Pharmacol Sci. Elsevier Ltd; 2014 Apr;35(4):208–15.
107. DeRosse P, Lencz T, Burdick KE, Siris SG, Kane JM, Malhotra AK. The genetics of symptom-based phenotypes: Toward a molecular classification of schizophrenia. Schizophr Bull. 2008;34(6):1047–53.
108. Kreienkamp H-J, Soltau M, Richter D, Böckers T. Interaction of G-protein-coupled receptors with synaptic scaffolding proteins. Biochem Soc Trans. Portland Press Limited; 2002 Aug 1;30(4):464–8.
109. Ribasés M, Ramos-Quiroga JA, Sánchez-Mora C, Bosch R, Richarte V, Palomar G, et al. Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: a replication study. Genes, Brain Behav. Wiley/Blackwell (10.1111); 2011 Mar 1;10(2):149–57.
110. Wandel E, Saalbach A, Sittig D, Gebhardt C, Aust G. Thy-1 (CD90) is an interacting partner for CD97 on activated endothelial cells. J Immunol. 2012 Feb 1;188(3):1442–50.
111. Yona S, Lin H-H, Dri P, Davies JQ, Hayhoe RPG, Lewis SM, et al. Ligation of the adhesion-GPCR EMR2 regulates human neutrophil function. FASEB J. 2008;22:741–51.
112. Wang J, Zhang L, Zhang H, Shen C, Lu S, Kuang Y, et al. Gpr97 is essential for the follicular versus marginal zone B-lymphocyte fate decision. Cell Death Dis. Nature Publishing Group; 2013 Oct 10;4(10):e853–e853.
113. Shi J, Zhang X, Wang S, Wang J, Du B, Wang Z, et al. Gpr97 is dispensable for metabolic syndrome but is involved in macrophage inflammation in high-fat diet-induced obesity in mice. Sci Rep. Nature Publishing Group; 2016 Jul 19;6(1):24649.
114. White JP, Wrann CD, Rao RR, Nair SK, Jedrychowski MP, You J-S, et al. G protein-coupled receptor 56 regulates mechanical overload-induced muscle hypertrophy. Proc Natl Acad Sci U S A. National Academy of Sciences; 2014 Nov 4;111(44):15756–61.
115. Reiners J, Nagel-Wolfrum K, Jürgens K, Märker T, Wolfrum U. Molecular basis of human Usher syndrome: Deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. Exp Eye Res. 2006 Jul;83(1):97–119.
116. Huang H, Kamm RD, Lee RT. Cell mechanics and mechanotransduction: pathways, probes, and physiology. Am J Physiol Physiol. American Physiological Society; 2004 Jul;287(1):C1–11.
117. Chen Y, Ju L, Rushdi M, Ge C, Zhu C. Receptor-mediated cell mechanosensing. Mol Biol Cell. American Society for Cell Biology; 2017 Nov 7;28(23):3134–55.
118. Ingber DE. Mechanobiology and diseases of mechanotransduction. Ann Med. 2003;35:1–14.
119. Zhu C, Bao G, Wang N. Cell Mechanics: Mechanical Response, Cell Adhesion, and Molecular Deformation. Annu Rev Biomed Eng. Annual Reviews; 2000 Aug 28;2(1):189–226.
120. Sequeira V, Nijenkamp LLA., Regan JA, van der Velden J. The physiological role of cardiac cytoskeleton and its alterations in heart failure. Biochim Biophys Acta - Biomembr. Elsevier; 2014 Feb 1;1838(2):700–22.
121. Takefuji M, Wirth A, Lukasova M, Takefuji S, Boettger T, Braun T, et al. G13-Mediated Signaling Pathway Is Required for Pressure Overload-Induced Cardiac Remodeling and Heart Failure. Circulation. 2012 Oct 16;126(16):1972–82.
122. Wang N. Review of cellular mechanotransduction. J Phys D Appl Phys. IOP Publishing; 2017 Jun 14;50(23):233002.
123. Jansen KA, Atherton P, Ballestrem C. Mechanotransduction at the cell-matrix interface. Semin Cell Dev Biol. Academic Press; 2017 Nov 1;71:75–83.
124. Shyy JY-J, Chien S. Role of integrins in endothelial mechanosensing of shear

stress. Circ Res. American Heart Association, Inc.; 2002 Nov 1;91(9):769–75.
125. Jansen KA, Donato DM, Balcioglu HE, Schmidt T, Danen EHJ, Koenderink GH. A guide to mechanobiology: Where biology and physics meet. Biochim Biophys Acta - Mol Cell Res. Elsevier; 2015 Nov 1;1853(11):3043–52.
126. Makino A, Prossnitz ER, Bünemann M, Wang JM, Yao W, Schmid-Schönbein GW. G protein-coupled receptors serve as mechanosensors for fluid shear stress in neutrophils. Am J Physiol Physiol. American Physiological Society; 2006 Jun;290(6):C1633–9.
127. Muhamed I, Chowdhury F, Maruthamuthu V. Biophysical Tools to Study Cellular Mechanotransduction. Bioeng (Basel, Switzerland). Multidisciplinary Digital Publishing Institute (MDPI); 2017 Feb 7;4(1).
128. Xiao Z, Quarles LD. Physiological mechanisms and therapeutic potential of bone mechanosensing. Rev Endocr Metab Disord. NIH Public Access; 2015 Jun;16(2):115–29.
129. Robling AG, Turner CH. Mechanical signaling for bone modeling and remodeling. Crit Rev Eukaryot Gene Expr. NIH Public Access; 2009;19(4):319–38.
130. Kim JH, Liu X, Wang J, Chen X, Zhang H, Kim SH, et al. Wnt signaling in bone formation and its therapeutic potential for bone diseases. Ther Adv Musculoskelet Dis. SAGE Publications; 2013 Feb;5(1):13–31.
131. Ruwhof C, van der Laarse A. Mechanical stress-induced cardiac hypertrophy: mechanisms and signal transduction pathways. Cardiovasc Res. Oxford University Press; 2000 Jul 1;47(1):23–37.
132. Ribeiro AS, Paredes J. P-Cadherin Linking Breast Cancer Stem Cells and Invasion: A Promising Marker to Identify an Intermediate/Metastable EMT State. Front Oncol. Frontiers Media SA; 2014;4:371.
133. Vieira AF, Ribeiro AS, Dionísio MR, Sousa B, Nobre AR, Albergaria A, et al. P-cadherin signals through the laminin receptor α6β4 integrin to induce stem cell and invasive properties in basal-like breast cancer cells. Oncotarget. Impact Journals, LLC; 2014 Feb 15;5(3):679–92.
134. Addae-Mensah KA, Wikswo JP. Measurement Techniques for Cellular Biomechanics In Vitro. Exp Biol Med. SAGE PublicationsSage UK: London, England; 2008 Jul 1;233(7):792–809.
135. Bao G, Kamm RD, Thomas W, Hwang W, Fletcher DA, Grodzinsky AJ, et al. Molecular Biomechanics: The Molecular Basis of How Forces Regulate Cellular Function. Mol Cell Biomech. NIH Public Access; 2010 Mar 2;3(2):91–105.
136. Moretti M, Canale C, Canale C, Francardi M, Dante S, De Angelis F, et al. AFM characterization of biomolecules in physiological environment by an advanced nanofabricated probe. Microsc Res Tech. Wiley-Blackwell; 2012 Dec;75(12):1723–31.
137. Whited AM, Park PS-H. Atomic force microscopy: a multifaceted tool to study membrane proteins and their interactions with ligands. Biochim Biophys Acta. NIH Public Access; 2014 Jan;1838(1 Pt A):56–68.
138. Paluch EK, Nelson CM, Biais N, Fabry B, Moeller J, Pruitt BL, et al. Mechanotransduction: use the force(s). BMC Biol. BioMed Central; 2015 Jul 4;13:47.
139. Stirnemann G, Giganti D, Fernandez JM, Berne BJ. Elasticity, structure, and relaxation of extended proteins under force. Proc Natl Acad Sci U S A. National Academy of Sciences; 2013 Mar 5;110(10):3847–52.
140. Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, et al. Mechanical and chemical unfolding of a single protein: A comparison. Proc Natl Acad Sci. 1999 Mar 30;96(7):3694–9.
141. Taniguchi Y, Kawakami M. Application of Halotag protein to covalent immobilization of recombinant proteins for single molecule force spectroscopy. Langmuir. 2010;26(14):10433–6.

142. Hughes ML, Dougan L. The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. Reports Prog Phys. 2016 Jul 1;79(7):076601.
143. Rief M, Pascual J, Saraste M, Gaub HE. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. J Mol Biol. Academic Press; 1999 Feb 19;286(2):553–61.
144. Carrion-Vazquez M. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. Prog Biophys Mol Biol. 2000 Aug;74(1–2):63–91.
145. Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. Prog Biophys Mol Biol. Pergamon; 2000 Jul 1;74(1–2):63–91.
146. Popa I, Berkovich R, Alegre-cebollada J, Badilla CL, Andre J, Taniguchi Y, et al. Nanomechanics of HaloTag Tethers. J Am Chem Socitey. 2013;
147. Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, et al. Pulling geometry defines the mechanical resistance of a β-sheetprotein. Nat Struct Mol Biol. Nature Publishing Group; 2003 Sep 17;10(9):731–7.
148. Richter A, Eggenstein E, Skerra A. Anticalins: Exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. FEBS Lett. No longer published by Elsevier; 2014 Jan 21;588(2):213–8.
149. Binz HK, Amstutz P, Plückthun A. Engineering novel binding proteins from nonimmunoglobulin domains. Nat Biotechnol. Nature Publishing Group; 2005 Oct 1;23(10):1257–68.
150. Skrlec K, Strukelj B, Berlec A. Non-immunoglobulin scaffolds: a focus on their targets. Trends Biotechnol. 2015;33(7):408–18.
151. Sha F, Salzman G, Gupta A, Koide S. Monobodies and other synthetic binding proteins for expanding protein science. Protein Sci. Wiley-Blackwell; 2017 May;26(5):910–24.
152. Lipovsek D. Adnectins: engineered target-binding protein therapeutics. Protein Eng Des Sel. Oxford University Press; 2011 Jan;24(1–2):3–9.
153. Vazquez-Lombardi R, Phan TG, Zimmermann C, Lowe D, Jermutus L, Christ D. Challenges and opportunities for non-antibody scaffold drugs. Drug Discov Today. Elsevier Current Trends; 2015 Oct 1;20(10):1271–83.
154. Zahnd C, Wyler E, Schwenk JM, Steiner D, Lawrence MC, McKern NM, et al. A Designed Ankyrin Repeat Protein Evolved to Picomolar Affinity to Her2. J Mol Biol. Academic Press; 2007 Jun 15;369(4):1015–28.
155. Frejd FY, Kim K-T. Affibody molecules as engineered protein drugs. Exp Mol Med. Korean Society for Biochemistry and Molecular Biology; 2017;49(3):e306.
156. Tiede C, Tang AAS, Deacon SE, Mandal U, Nettleship JE, Owen RL, et al. Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. Protein Eng Des Sel. 2014;27:145–55.
157. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. Structure. 2010 Oct;18(10):1233–43.
158. Chiang N-Y, Peng Y-M, Juang H-H, Chen T-C, Pan H-L, Chang G-W, et al. GPR56/ADGRG1 Activation Promotes Melanoma Cell Migration via NTF Dissociation and CTF-Mediated Gα12/13/RhoA Signaling. J Invest Dermatol. Elsevier; 2017 Mar 1;137(3):727–36.
159. Hammers CM, Stanley JR. Antibody phage display: technique and applications. J Invest Dermatol. NIH Public Access; 2014 Feb;134(2):1–5.
160. Barnard A, Long K, Martin HL, Miles JA, Edwards TA, Tomlinson DC, et al. Selective and Potent Proteomimetic Inhibitors of Intracellular Protein-Protein Interactions. Angew Chemie Int Ed. WILEY-VCH Verlag; 2015 Mar 2;54(10):2960–5.

161. Tiede C, Bedford R, Heseltine SJ, Smith G, Wijetunga I, Ross R, et al. Affimer proteins are versatile and renewable affinity reagents. Elife. 2017;6(c):1–35.
162. Hughes DJ, Tiede C, Penswick N, Tang AA-S, Trinh CH, Mandal U, et al. Generation of specific inhibitors of SUMO-1- and SUMO-2/3-mediated protein-protein interactions using Affimer (Adhiron) technology. Sci Signal. 2017 Nov 14;10(505):eaaj2005.
163. Lopata A, Hughes R, Tiede C, Heissler SM, Sellers JR, Knight PJ, et al. Affimer proteins for F-actin: novel affinity reagents that label F-actin in live and fixed cells. Sci Rep. Nature Publishing Group; 2018 Dec 26;8(1):6572.
164. Zhurauski P, Arya SK, Jolly P, Tiede C, Tomlinson DC, Ko Ferrigno P, et al. Sensitive and selective Affimer-functionalised interdigitated electrode-based capacitive biosensor for Her4 protein tumour biomarker detection. Biosens Bioelectron. Elsevier; 2018 Jun 15;108:1–8.
165. Xie C, Tiede C, Zhang X, Wang C, Li Z, Xu X, et al. Development of an Affimer-antibody combined immunological diagnosis kit for glypican-3. Sci Rep. Nature Publishing Group; 2017 Dec 29;7(1):9608.
166. Potterton E, Briggs P, Turkenburg M, Dodson E, IUCr. A graphical user interface to the CCP4 program suite. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2003 Jul 1;59(7):1131–7.
167. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2011 Apr 1;67(4):235–42.
168. Winter G. xia2 : an expert system for macromolecular crystallography data reduction. J Appl Crystallogr. International Union of Crystallography; 2010 Feb 1;43(1):186–90.
169. RCSB PDB. PDB-101: Learn: Guide to Understanding PDB Data: Methods for Determining Structure [Internet]. 2018 [cited 2018 Jun 4]. Available from: http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure
170. Dessau MA, Modis Y. Protein crystallization for X-ray crystallography. J Vis Exp. MyJoVE Corporation; 2011 Jan 16;(47).
171. McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallogr Sect F, Struct Biol Commun. International Union of Crystallography; 2014 Jan;70(Pt 1):2–20.
172. Chruszcz M, Zimmerman MD, Wang S, Koclega KD, Zheng H, Evdokimova E, et al. Function-biased choice of additives for optimization of protein crystallization - the case of the putative thioesterase PA5185 from Pseudomonas aeruginosa PAO1. Cryst Growth Des. NIH Public Access; 2008 Nov 5;8(11):4054–61.
173. Sauter C, Ng JD, Lorber B, Keith G, Philippe Brion, Wais Hosseini M, et al. Additives for the crystallization of proteins and nucleic acids. J Cryst Growth. North-Holland; 1999 Jan 15;196(2–4):365–76.
174. Guan R-J, Wang M, Liu X-Q, Wang D-C. Optimization of soluble protein crystallization with detergents. J Cryst Growth. North-Holland; 2001 Sep 1;231(1–2):273–9.
175. Bergfors T. Seeds to crystals. J Struct Biol. Academic Press; 2003 Apr 1;142(1):66–76.
176. Sennhauser G, Grütter MG. Chaperone-assisted crystallography with DARPins. Structure. Elsevier; 2008 Oct 8;16(10):1443–53.
177. Griffin L, Lawson A. Antibody fragments as tools in crystallography. Clin Exp Immunol. Wiley-Blackwell; 2011 Sep;165(3):285–91.
178. Salzman GS, Ackerman SD, Ding C, Koide A, Leon K, Luo R, et al. Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. Neuron. Elsevier; 2016 Sep 21;91(6):1292–304.
179. Deepak S, Kottapalli K, Rakwal R, Oros G, Rangappa K, Iwahashi H, et al.

Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes. Curr Genomics. Bentham Science Publishers; 2007 Jun;8(4):234–51.
180. Mackay IM, Arden KE, Nitsche A. Real-time PCR in virology. Nucleic Acids Res. Oxford University Press; 2002 Mar 15;30(6):1292–305.
181. Watzinger F, Suda M, Preuner S, Baumgartinger R, Ebner K, Baskova L, et al. Real-time quantitative PCR assays for detection and monitoring of pathogenic human viruses in immunosuppressed pediatric patients. J Clin Microbiol. American Society for Microbiology (ASM); 2004 Nov;42(11):5189–98.
182. Tessier DC, Thomas DY, Khouri HE, Laliberié F, Vernet T. Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. Gene. Elsevier; 1991 Feb 15;98(2):177–83.
183. San-Miguel T, Pérez-Bermúdez P, Gavidia I. Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. Springerplus. 2013 Dec;2(1):89.
184. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. Frontiers; 2014 Apr 17;5:172.
185. Baneyx F. Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol. Elsevier Current Trends; 1999 Oct 1;10(5):411–21.
186. de Boer HA, Comstock LJ, Vasser M. The tac promoter: a functional hybrid derived from the trp and lac promoters. Proc Natl Acad Sci U S A. National Academy of Sciences; 1983 Jan 1;80(1):21–5.
187. Niedzialkowska E, Gasiorowska O, Handing KB, Majorek KA, Porebski PJ, Shabalin IG, et al. Protein purification and crystallization artifacts: The tale usually not told. Protein Sci. Wiley-Blackwell; 2016 Mar;25(3):720–33.
188. Stura EA, Wilson IA. Applications of the streak seeding technique in protein crystallization. J Cryst Growth. North-Holland; 1991 Mar 1;110(1–2):270–82.
189. Rossmann M. The molecular replacement method. Acta Crystallogr Sect D Biol Crystallogr. 1990;(A46):73–82.
190. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., & Read RJ. Phaser Crystallographic Software. J Appl Crystallogr. 2007;(40):658–74.
191. Bibby J, Keegan RM, Mayans O, Winn MD, Rigden DJ. AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. Acta Crystallogr Sect D, Struct Biol. 2013;(68):1622–31.
192. Simpkin A, Simkovic F, Thomas J, Savko M, Lebedev A, Uski V, et al. SIMBAD: a sequence-independent molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr. 2018;(7):595–605.
193. Singh SM, Panda AK. Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng. 2005 Apr;99(4):303–10.
194. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv. Elsevier; 2009 May 1;27(3):297–306.
195. Luckow VA, Lee SC, Barry GF, Olins PO. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. J Virol. American Society for Microbiology (ASM); 1993 Aug;67(8):4566–79.
196. Kost TA, Condreay JP, Jarvis DL. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nat Biotechnol. NIH Public Access; 2005 May;23(5):567–75.
197. Bouvier M, Ménard L, Dennis M, Marullo S. Expression and recovery of functional G-protein-coupled receptors using baculovirus expression systems. Curr Opin Biotechnol. 1998;9(5):522–7.
198. Milić D, Veprintsev DB. Large-scale production and protein engineering of G protein-coupled receptors for structural studies. Front Pharmacol. 2015;6(March):1–24.

199. Akermoun M, Koglin M, Zvalova-Iooss D, Folschweiller N, Dowell SJ, Gearing KL. Characterization of 16 human G protein-coupled receptors expressed in baculovirus-infected insect cells. Protein Expr Purif. 2005;44(1):65–74.
200. Santos H de L, Ciancaglini P. A practical approach to the choice of a suitable detergent and optimal conditions for solubilizing a membrane protein. Biochem Educ. Wiley-Blackwell; 2000 May;28(3):178–82.
201. Lebendiker M, Danieli T. Production of prone-to-aggregate proteins. FEBS Lett. Wiley-Blackwell; 2013 Nov 6;588(2):236–46.
202. Hsiao C-C, Cheng K-F, Chen H-Y, Chou Y-H, Stacey M, Chang G-W, et al. Site-specific N-glycosylation regulates the GPS auto-proteolysis of CD97. FEBS Lett. Federation of European Biochemical Societies; 2009 Oct 6;583(19):3285–90.
203. HU Y. Baculovirus as a highly efficient expression vector in insect and mammalian cells. Acta Pharmacol Sin. Nature Publishing Group; 2005 Apr 1;26(4):405–16.
204. Mayer M, Buchner J. Refolding of inclusion body proteins. Methods Mol Med. 2004 Jan;94(3):239–54.
205. Zhou Z, Dang Y, Zhou M, Li L, Yu C-H, Fu J, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc Natl Acad Sci U S A. National Academy of Sciences; 2016 Oct 11;113(41):E6117–25.
206. Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Mol Cell. NIH Public Access; 2015 Jul 16;59(2):149–61.
207. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol. Elsevier Current Trends; 1995 Jan 1;6(5):494–500.
208. Kopanic JL, Al-Mugotir M, Zach S, Das S, Grosely R, Sorgen PL. An Escherichia coli strain for expression of the connexin45 carboxyl terminus attached to the 4th transmembrane domain. Front Pharmacol. Frontiers Media SA; 2013;4:106.
209. Uhlén M, Nilsson B, Guss B, Lindberg M, Gatenbeck S, Philipson L. Gene fusion vectors based on the gene for staphylococcal protein A. Gene. 1983 Sep;23(3):369–78.
210. Thapa A, Shahnawaz M, Karki P, Raj Dahal G, Sharoar MG, Yub Shin S, et al. Purification of inclusion body-forming peptides and proteins in soluble form by fusion to Escherichia coli thermostable proteins. Biotechniques. 2008 May;44(6):787–96.
211. Pryor KD, Leiting B. High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. Protein Expr Purif. 1997 Aug;10(3):309–19.
212. Drew D, Lerch M, Kunji E, Slotboom D-J, de Gier J-W. Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods. 2006 Apr;3(4):303–13.
213. Shevchik VE, Condemine G, Robert-Baudouy J. Characterization of DsbC, a periplasmic protein of Erwinia chrysanthemi and Escherichia coli with disulfide isomerase activity. EMBO J. European Molecular Biology Organization; 1994 Apr 15;13(8):2007–12.
214. Bessette PH, Aslund F, Beckwith J, Georgiou G. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proc Natl Acad Sci U S A. National Academy of Sciences; 1999 Nov 23;96(24):13703–8.
215. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol. 2004 Nov;22(11):1399–408.
216. Vellejo L., Rinas U. Strategies for the recovery of active proteins through resolving of bacterial inclusion body proteins. Microb Cell Fact. 2004;3(11).
217. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a

novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact. BioMed Central; 2012 May 8;11:56.
218. Robinson M-P, Ke N, Lobstein J, Peterson C, Szkodny A, Mansell TJ, et al. Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. Nat Commun. Nature Publishing Group; 2015 Dec 27;6(1):8072.
219. Kong B, Guo GL. Soluble Expression of Disulfide Bond Containing Proteins FGF15 and FGF19 in the Cytoplasm of Escherichia coli. Riggs PD, editor. PLoS One. Public Library of Science; 2014 Jan 20;9(1):e85890.
220. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif. 2005 May;41(1):207–34.
221. Alexander P. Golovanov, Guillaume M. Hautbergue, Stuart A. Wilson and, Lu-Yun Lian. A Simple Method for Improving Protein Solubility and Long-Term Stability. American Chemical Society ; 2004;
222. Leney AC, Heck AJR. Native Mass Spectrometry: What is in the Name? J Am Soc Mass Spectrom. Springer; 2017 Jan;28(1):5–13.
223. Boeri Erba E, Petosa C. The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. Protein Sci. Wiley-Blackwell; 2015 Aug;24(8):1176–92.
224. Luft JR, Wolfley JR, Said MI, Nagel RM, Lauricella AM, Smith JL, et al. Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. Protein Sci. Wiley-Blackwell; 2007 Apr;16(4):715–22.
225. Jancarik J, Kim SH. Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Crystallogr. International Union of Crystallography (IUCr); 1991 Aug 1;24(4):409–11.
226. McPherson A. Two approaches to the rapid screening of crystallization conditions. J Cryst Growth. North-Holland; 1992 Aug 2;122(1–4):161–7.
227. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J. NIH Public Access; 2013 Nov;280(22):5705–36.
228. Day PW, Rasmussen SGF, Parnot C, Fung JJ, Masood A, Kobilka TS, et al. A monoclonal antibody for G protein–coupled receptor crystallography. Nat Methods. Nature Publishing Group; 2007 Nov 21;4(11):927–9.
229. Longenecker KL, Garrard SM, Sheffield PJ, Derewenda ZS, IUCr. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2001 May 1;57(5):679–88.
230. Derewenda ZS. Rational Protein Crystallization by Mutational Surface Engineering. Structure. Cell Press; 2004 Apr 1;12(4):529–35.
231. DiMaio F, Terwilliger TC, Read RJ, Wlodawer A, Oberdorfer G, Wagner U, et al. Improved molecular replacement by density- and energy-guided protein structure optimization. Nature. Nature Publishing Group; 2011 May 1;473(7348):540–3.
232. Abergel C. Molecular replacement: tricks and treats. Acta Crystallogr D Biol Crystallogr. International Union of Crystallography; 2013 Nov;69(Pt 11):2167–73.
233. Walden H. Selenium incorporation using recombinant techniques. Acta Crystallogr D Biol Crystallogr. International Union of Crystallography; 2010 Apr;66(Pt 4):352–7.
234. Pike ACW, Garman EF, Krojer T, von Delft F, Carpenter EP. An overview of heavy-atom derivatization of protein crystals. Acta Crystallogr Sect D, Struct Biol. International Union of Crystallography; 2016 Mar;72(Pt 3):303–18.
235. Garman E, Murray JW. Heavy-atom derivatization. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2003 Nov 1;59(11):1903–13.

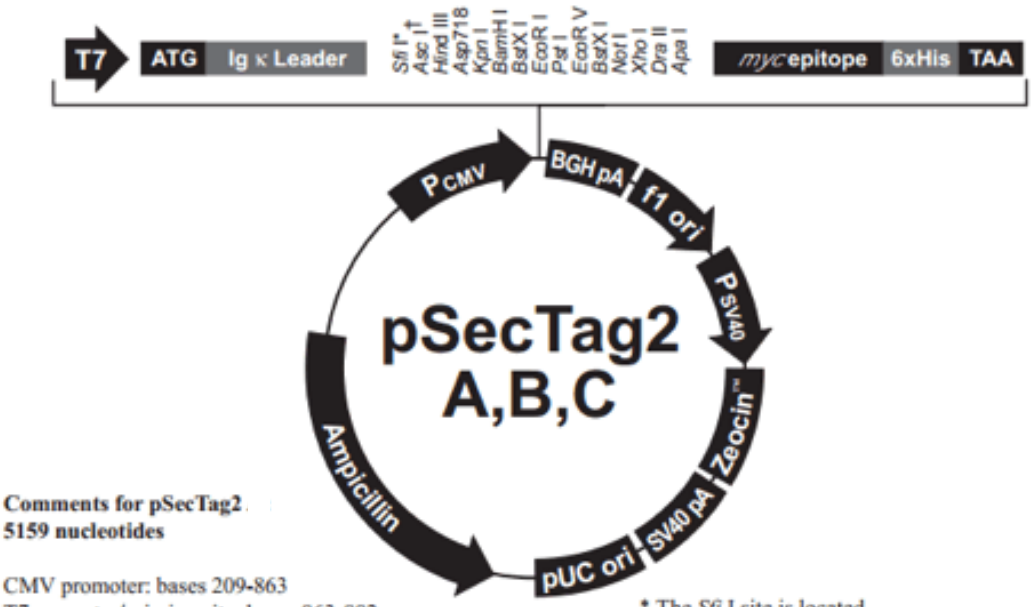
236. Joyce MG, Radaev S, Sun PD. A rational approach to heavy-atom derivative screening. Acta Crystallogr D Biol Crystallogr. International Union of Crystallography; 2010 Apr;66(Pt 4):358–65.
237. Koide A, Bailey CW, Huang X, Koide S. The fibronectin type III domain as a scaffold for novel binding proteins. J Mol Biol. Academic Press; 1998 Dec 11;284(4):1141–51.
238. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. BioMed Central; 2008 Jan 23;9(1):40.
239. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. Oxford University Press; 2005 Jul 1;33(Web Server issue):W244-8.
240. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. Oxford University Press; 2015 Jul 1;43(W1):W389-94.
241. Wilde C, Fischer L, Lede V, Kirchberger J, Rothemund S, Schöneberg T, et al. The constitutive activity of the adhesion GPCR GPR114/ADGRG5 is mediated by its tethered agonist. FASEB J. Federation of American Societies for Experimental BiologyBethesda, MD, USA; 2016 Feb 23;30(2):666–73.
242. Cheng, Z., Paguio, A., Garvin, D., Stecha, P, Wood, K. and Fan F. Luciferase Reporter Assay for Deciphering GPCR Pathways [Internet]. 2009 [cited 2018 Aug 14].
243. Li C. CD97 promotes gastric cancer cell proliferation and invasion through exosome-mediated MAPK signaling pathway. World J Gastroenterol. 2015;21(20):6215.
244. Hill CS, Treisman R. Differential activation of c-fos promoter elements by serum, lysophosphatidic acid, G proteins and polypeptide growth factors. EMBO J. European Molecular Biology Organization; 1995 Oct 16;14(20):5037–47.
245. Atwood BK, Lopez J, Wager-Miller J, Mackie K, Straiker A. Expression of G protein-coupled receptors and related proteins in HEK293, AtT20, BV2, and N18 cell lines as revealed by microarray analysis. BMC Genomics. BioMed Central Ltd; 2011 Jan 7;12(1):14.
246. Valtcheva N, Primorac A, Jurisic G, Hollmén M, Detmar M. The orphan adhesion G protein-coupled receptor GPR97 regulates migration of lymphatic endothelial cells via the small GTPases RhoA and Cdc42. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2013 Dec 13;288(50):35736–48.
247. Piao X, Hill RS, Bodell A, Chang BS, Basel-Vanagaite L, Straussberg R, et al. G protein-coupled receptor-dependent development of human frontal cortex. Science. American Association for the Advancement of Science; 2004 Mar 26;303(5666):2033–6.
248. Weston MD, Luijendijk MWJ, Humphrey KD, Möller C, Kimberling WJ. Mutations in the VLGR1 gene implicate G-protein signaling in the pathogenesis of Usher syndrome type II. Am J Hum Genet. 2004;74:357–66.
249. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. Nature Publishing Group; 2009 Jul 25;4(7):1073–81.
250. Nazarko O, Kibrom A, Winkler J, Leon K, Stoveken H, Salzman G, et al. A Comprehensive Mutagenesis Screen of the Adhesion GPCR Latrophilin-1/ADGRL1. iScience. Elsevier; 2018 May 25;3:264–78.
251. Hill CS, Wynne J, Treisman R. The Rho family GTPases RhoA, Racl, and CDC42Hsregulate transcriptional activation by SRF. Cell. 1995;81(7):1159–70.
252. Qian Y, Haino M, Kelly K. Structural characterization of mouse CD97 and study of its specific interaction with the murine decay-accelerating factor (DAF , CD55). Immunology. 1999;98:303–11.

253. Stacey M, Chang G, Davies JQ, Kwakkenbos MJ, Sanderson RD, Gordon S, et al. The epidermal growth factor – like domains of the human EMR2 receptor mediate cell attachment through chondroitin sulfate glycosaminoglycans. Blood. 2003;102(8):2916–24.
254. Lin HH, Stacey M, Saxby C, Knott V, Chaudhry Y, Evans D, et al. Molecular analysis of the epidermal growth factor-like short consensus repeat domain-mediated protein-protein interactions: dissection of the CD97-CD55 complex. J Biol Chem. 2001 Jun 29;276(26):24160–9.
255. Salzman GS, Zhang S, Gupta A, Koide A, Koide S, Araç D. Stachel-independent modulation of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. Proc Natl Acad Sci U S A. National Academy of Sciences; 2017 Sep 19;114(38):10095–100.
256. Smith SEP, Bida AT, Davis TR, Sicotte H, Patterson SE, Gil D, et al. IP-FCM Measures Physiologic Protein-Protein Interactions Modulated by Signal Transduction and Small-Molecule Drug Inhibition. Kanellopoulos J, editor. PLoS One. Public Library of Science; 2012 Sep 21;7(9):e45722.
257. Joo SH. Cyclic peptides as therapeutic agents and biochemical tools. Biomol Ther (Seoul). Korean Society of Applied Pharmacology; 2012 Jan;20(1):19–26.
258. Liu HW, Halayko AJ, Fernandes DJ, Harmon GS, McCauley JA, Kocieniewski P, et al. The RhoA/Rho Kinase Pathway Regulates Nuclear Localization of Serum Response Factor. Am J Respir Cell Mol Biol. American Thoracic Society; 2003 Jul 20;29(1):39–47.
259. Sun Z, Guo SS, Fässler R. Integrin-mediated mechanotransduction. J Cell Biol. Rockefeller University Press; 2016 Nov 21;215(4):445–56.
260. Arcos-Burgos M, Muenke M. Toward a better understanding of ADHD: LPHN3 gene variants and the susceptibility to develop ADHD. Atten Defic Hyperact Disord. NIH Public Access; 2010 Nov;2(3):139–47.
261. Piao X, Basel-Vanagaite L, Straussberg R, Grant PE, Pugh EW, Doheny K, et al. An autosomal recessive form of bilateral frontoparietal polymicrogyria maps to chromosome 16q12.2-21. Am J Hum Genet. Elsevier; 2002 Apr;70(4):1028–33.
262. Kobilka BK. G protein coupled receptor structure and activation. Biochim Biophys Acta. NIH Public Access; 2007 Apr;1768(4):794–807.
263. Yin Y, de Waal PW, He Y, Zhao L-H, Yang D, Cai X, et al. Rearrangement of a polar core provides a conserved mechanism for constitutive activation of class B G protein-coupled receptors. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2017 Jun 16;292(24):9865–81.
264. Carrion-Vazquez M, Marszalek PE, Oberhauser AF, Fernandez JM. Atomic force microscopy captures length phenotypes in single proteins. Proc Natl Acad Sci U S A. National Academy of Sciences; 1999 Sep 28;96(20):11288–92.
265. Oroz J, Bruix M, Laurents D V, Galera-Prat A, Schönfelder J, Cañada FJ, et al. The Y9P Variant of the Titin I27 Module: Structural Determinants of Its Revisited Nanomechanics. Structure. Elsevier; 2016 Apr 5;24(4):606–16.
266. Riener CK, Stroh CM, Ebner A, Klampfl C, Gall AA, Romanin C, et al. Simple test system for single molecule recognition force microscopy. Anal Chim Acta. Elsevier; 2003 Mar 5;479(1):59–75.
267. Bryksin A V., Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques. 2010;48(6):463–5.
268. Javadi Y, Fernandez JM, Perez-Jimenez R. Protein folding under mechanical forces: a physiological view. Physiology (Bethesda). 2013;28:9–17.
269. Zinober RC, Brockwell DJ, Beddard GS, Blake AW, Olmsted PD, Radford SE, et al. Mechanically unfolding proteins: the effect of unfolding history and the supramolecular scaffold. Protein Sci. Wiley-Blackwell; 2002 Dec;11(12):2759–

65.
270. Ott W, Jobst MA, Schoeler C, Gaub HE, Nash MA. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. J Struct Biol. Academic Press; 2017 Jan 1;197(1):3–12.
271. Linke WA, Grützner A. Pulling single molecules of titin by AFM—recent advances and physiological implications. Pflügers Arch - Eur J Physiol. Springer-Verlag; 2008 Apr 6;456(1):101–15.
272. Block H, Maertens B, Spriestersbach A, Brinker N, Kubicek J, Fabis R, et al. Immobilized-Metal Affinity Chromatography (IMAC): A Review. Methods Enzymol. Academic Press; 2009 Jan 1;463:439–73.
273. Rangl M, Ebner A, Yamada J, Rankl C, Tampé R, Gruber HJ, et al. Single-molecule analysis of the recognition forces underlying nucleo-cytoplasmic transport. Angew Chem Int Ed Engl. NIH Public Access; 2013 Sep 23;52(39):10356–9.
274. Schoeler C, Bernardi RC, Malinowska KH, Durner E, Ott W, Bayer EA, et al. Mapping Mechanical Force Propagation through Biomolecular Complexes. Nano Lett. NIH Public Access; 2015 Nov 11;15(11):7370–6.
275. Lemmon CA, Weinberg SH. Multiple Cryptic Binding Sites are Necessary for Robust Fibronectin Assembly: An In Silico Study. Sci Rep. Nature Publishing Group; 2017 Dec 22;7(1):18061.
276. Brockwell DJ. Probing the mechanical stability of proteins using the atomic force microscope. Biochem Soc Trans. 2007;35(6).
277. Puchner EM, Gaub HE. Force and function: probing proteins with AFM-based force spectroscopy. Curr Opin Struct Biol. Elsevier Current Trends; 2009 Oct 1;19(5):605–14.
278. Tang J, Ebner A, Kraxberger B, Leitner M, Hykollari A, Kepplinger C, et al. Detection of metal binding sites on functional S-layer nanoarrays using single molecule force spectroscopy. J Struct Biol. Academic Press; 2009 Oct 1;168(1):217–22.
279. Zocher M, Bippes CA, Zhang C, Müller DJ. Single-molecule force spectroscopy of G-protein-coupled receptors. Chem Soc Rev. Royal Society of Chemistry; 2013 Sep 9;42(19):7801.
280. Müller DJ, Sapra KT, Scheuring S, Kedrov A, Frederix PL, Fotiadis D, et al. Single-molecule studies of membrane proteins. Curr Opin Struct Biol. Elsevier Current Trends; 2006 Aug 1;16(4):489–95.
281. Kedrov A, Krieg M, Ziegler C, Kuhlbrandt W, Muller DJ. Locating ligand binding and activation of a single antiporter. EMBO Rep. EMBO Press; 2005 Jul 1;6(7):668–74.
282. Brockwell DJ, Beddard GS, Paci E, West DK, Olmsted PD, Smith DA, et al. Mechanically unfolding the small, topologically simple protein L. Biophys J. Elsevier; 2005 Jul 1;89(1):506–19.
283. Kim S, Blainey PC, Schroeder CM, Xie XS. Multiplexed single-molecule assay for enzymatic activity on flow-stretched DNA. Nat Methods. Nature Publishing Group; 2007 Apr 15;4(5):397.
284. Wang K, Forbes JG, Jin AJ. Single molecule measurements of titin elasticity. Prog Biophys Mol Biol. Pergamon; 2001 Jan 1;77(1):1–44.
285. Litvinov RI, Shuman H, Bennett JS, Weisel JW. Binding strength and activation state of single fibrinogen-integrin pairs on living cells. Proc Natl Acad Sci U S A. National Academy of Sciences; 2002 May 28;99(11):7426–31.
286. Block SM, Goldstein LSB, Schnapp BJ. Bead movement by single kinesin molecules studied with optical tweezers. Nature. Nature Publishing Group; 1990 Nov 22;348(6299):348–52.
287. Neuman KC, Nagy A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods. NIH Public Access; 2008 Jun;5(6):491–505.
288. Trowbridge JM, Gallo RL. Dermatan sulfate: new functions from an old

glycosaminoglycan. Glycobiology. Oxford University Press; 2002 Sep 1;12(9):117R–125R.
289. Hilbig D, Sittig D, Hoffmann F, Rothemund S, Warmt E, Quaas M, et al. Mechano-Dependent Phosphorylation of the PDZ-Binding Motif of CD97/ADGRE5 Modulates Cellular Detachment. Cell Rep. Cell Press; 2018 Aug 21;24(8):1986–95.
290. Marinissen MJ, Gutkind JS. G-protein-coupled receptors and signaling networks: emerging paradigms. Trends Pharmacol Sci. Elsevier Current Trends; 2001 Jul 1;22(7):368–76.
291. Sah VP, Seasholtz TM, Sagi SA, Brown JH. The Role of Rho in G Protein-Coupled Receptor Signal Transduction. Annu Rev Pharmacol Toxicol. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA ; 2000 Apr 28;40(1):459–89.

Appendix



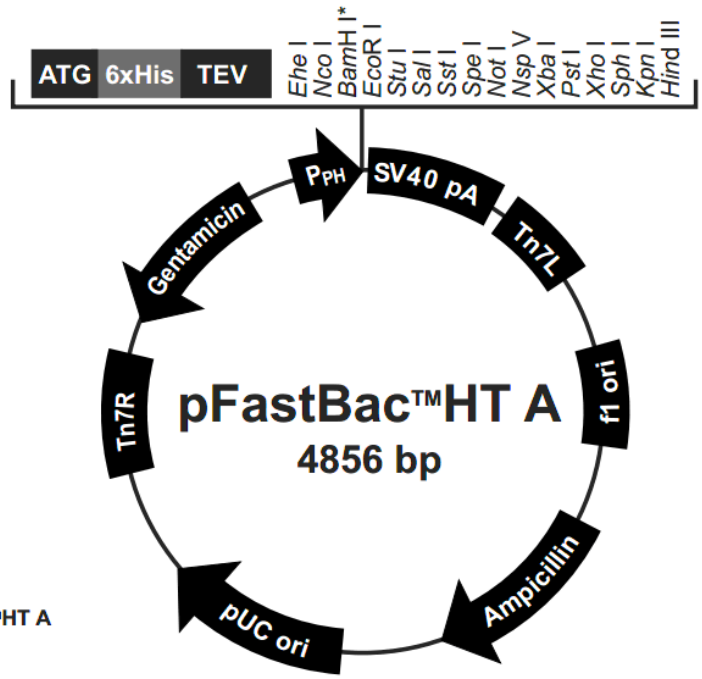
Comments for pSecTag2
5159 nucleotides

CMV promoter: bases 209-863
T7 promoter/priming site: bases 863-882
Murine Ig κ-chain V-J2-C signal peptide: bases 905-967
Multiple cloning site: bases 970-1081
c-myc epitope: bases 1082-1114
Polyhistidine tag: bases 1127-1144
BGH reverse priming site: bases 1167-1184
BGH polyadenylation sequence: bases 1166-1380
f1 origin: bases 1443-1856
SV40 promoter and origin: bases 1924-2245
EM-7 promoter: bases 2261-2327
Zeocin™ resistance gene: bases 2328-2702
SV40 polyadenylation sequence: bases 2832-2961
pUC origin: bases 3345-4018
Ampicillin resistance gene: bases 4163-5023 (C)

* The *Sfi* I site is located before the variable region.

†Frame-dependent variations: *Asc* I is only found in version A.

Appendix Figure 1. *pSecTag2B mammalian expression vector*. Used for mammalian expression of AFM constructs.



Comments for pFastBac™HT A
4856 nucleotides

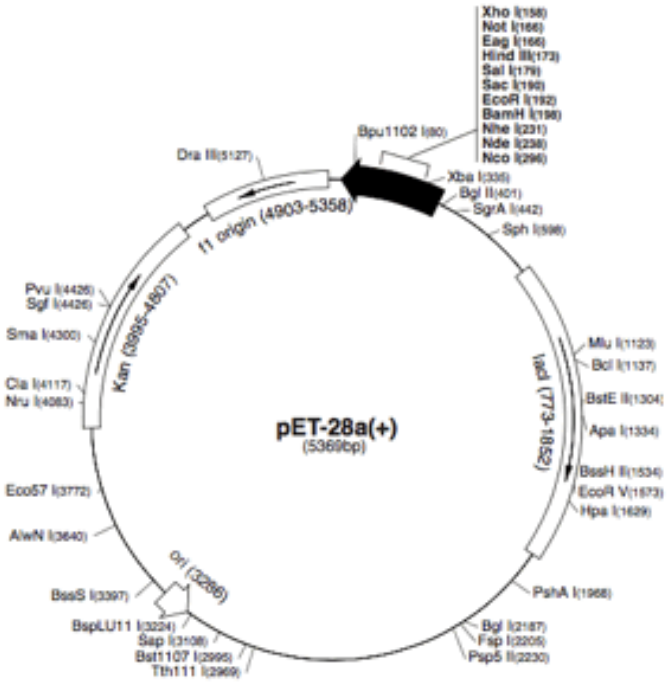
f1 origin: bases 2-457
Ampicillin resistance gene: bases 589-1449
pUC origin: bases 1594-2267
Tn7R: bases 2511-2735
Gentamicin resistance gene: bases 2802-3335 (complementary strand)
Polyhedrin promoter (P_{PH}): bases 3904-4032
Initiation ATG: bases 4050-4052
6xHis tag: bases 4062-4079
TEV recognition site: bases 4101-4121
Multiple cloning site: bases 4119-4222
SV40 polyadenylation signal: bases 4240-4480
Tn7L: bases 4509-4674

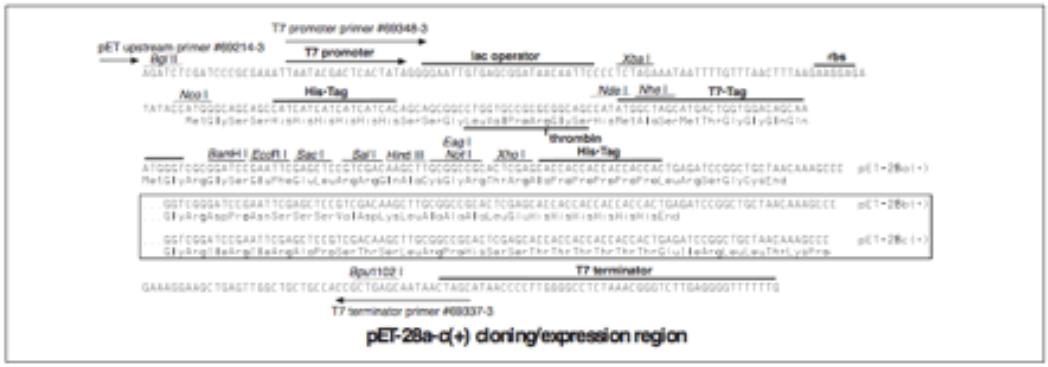
*Frameshift occurs at the *Bam*H I site in each vector

Appendix Figure 2. *pFastBacHT A bacterial expression vector*. Used for bacterial expression of recombinant bacmid, for subsequent insect expression.

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites	
(*Bam* H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam* H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam* H I at 198.

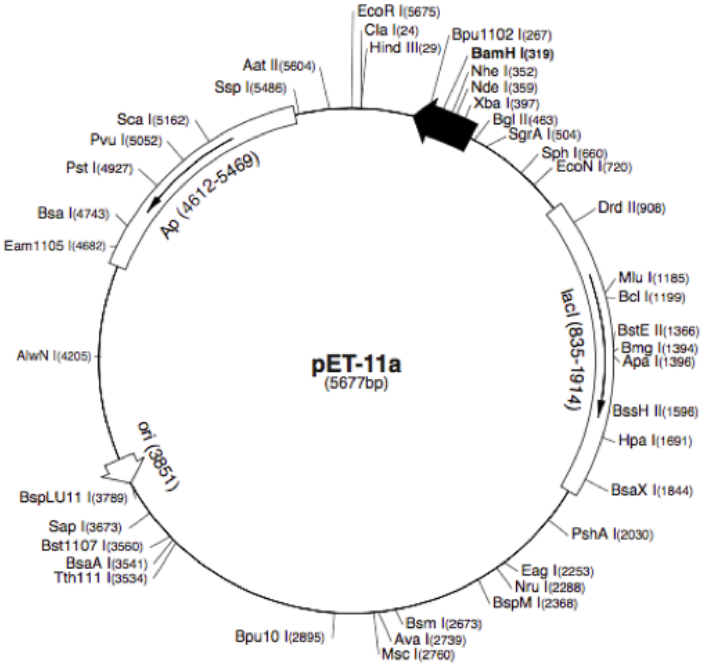


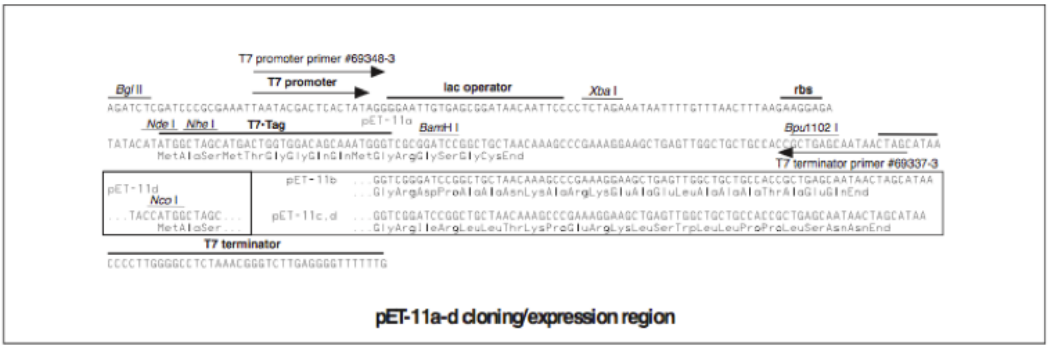


Appendix Figure 3. ***pET-28A bacterial expression vector*.** Used for bacterial expression of CD97 GAIN domain.

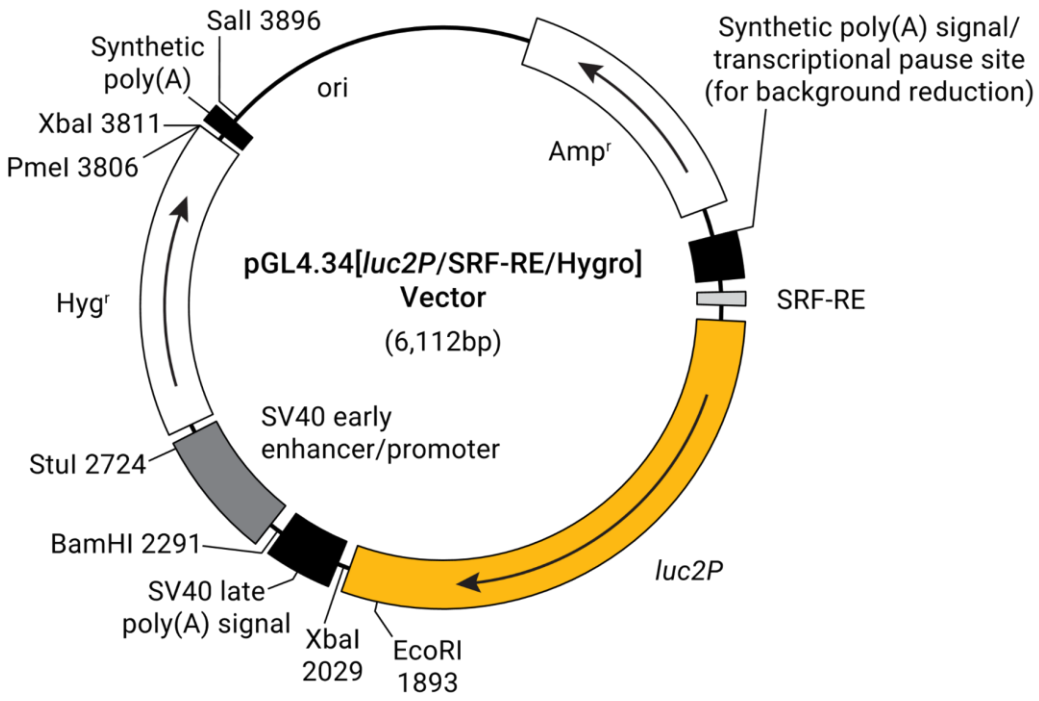
pET-11a sequence landmarks	
T7 promoter	432-448
T7 transcription start	431
T7•Tag coding sequence	328-360
T7 terminator	213-259
lacI coding sequence	835-1914
pBR322 origin	3851
bla coding sequence	4612-5469

The maps for pET-11b, pET-11c and pET-11d are the same as pET-11a (shown) with the following exceptions: pET-11b is a 5676bp plasmid; subtract 1bp from each site beyond *Bam* H I at 319. pET-11c is a 5675bp plasmid; subtract 2bp from each site beyond *Bam* H I at 319. pET-11d is a 5674bp plasmid; the *Bam* H I site is in the same reading frame as in pET-11c. An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 359 of pET-11c. As a result, *Nco* I cuts pET-11d at 355. For the rest of the sites, subtract 3bp from each site beyond position 360 in pET-11a. *Nde* I does not cut pET-11d.





Appendix Figure 4. *pET11a bacterial expression vector*. Used for bacterial expression of Affimers.



Appendix Figure 5. ***pGL4.34 SRF-RE mammalian expression vector*****.** Used for mammalian expression of SRF-RE in luciferase signalling assays. Representative of SRE expression vector.