Chemical Probes for GPCRs: Investigating GPR119

KATE SHIRIN HADAVIZADEH

Imperial College London

Department of Chemistry

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and the Diploma of Imperial College London

November 2019

Declaration of Originality

This Thesis is my own work and reports the results of my original research. Any work that is the result of collaboration or the input of others has been specifically stated within the text. Information derived from the work of others has been acknowledged and referenced.

In accordance with the Department of Chemistry guidelines, this thesis does not exceed 100,000 words.

Kate Hadavizadeh

September 2019

Copyright Declaration

Abstract

GPCRs are one of the most important classes of proteins within the arena of drug discovery and therapeutics. They are involved in transmitting external stimuli across the cell membrane to elicit intracellular responses that can alter gene activation, protein synthesis and messenger signalling that can be intrinsic to health and disease. GPR119 is a GPCR shown to be involved in glucose-dependent insulin secretion with expression in the pancreatic β-cells and intestinal K- and L-cells, implicating the protein as a potential therapeutic target for type 2 diabetes. There have been many agonists published for GPR119, however none have progressed further than phase II clinical trials, largely due to lack of efficacy.

This thesis focuses on the synthesis and evaluation of chemical probes for GPR119 based on agonists developed by AstraZeneca, with differing pharmacophores exhibiting both on-target GPR119 mediated blood glucose lowering, and off-target GPR119 independent blood glucose lowering. Modifying the pharmacophores with diazirine photoreactive groups and alkyne click chemistry handles, a library of probes was synthesised. Probes exhibited a range of potencies for GPR119, with the most potent having an EC_{50} in the nanomolar range.

Within this thesis, the challenges of working with GPCRs are addressed: expression and identification of GPR119 protein is a difficulty encountered both within the literature and during this study. By exploring the use of epitope tags and fusion proteins, GPR119 expression was detectable using fluorescence and immunoblotting techniques. The generation of a HaloTag-GPR119 construct enabled the identification of GPR119 using proteomic methods, which has not previously been reported.

By utilising the biological tools for GPR119, evaluation of the chemical photocrosslinking probes for GPR119 was possible in cell lines over expressing HaloTag-GPR119. Dual fluorescent labelling of the protein and probe, alongside immunoblotting identified a chemical probe showing response to competition with the parent compound, presenting a potential chemical probe for GPR119.

Acknowledgements

Firstly, I would like to thank Ed Tate for allowing me to be a part of a fantastic group with a supportive, enthusiastic, collaborative and friendly culture. Thank you for giving me the courage and space to develop a research project in, what I now understand as, a very challenging but rewarding area of chemical biology. Thanks for the encouragement and continued support even after departing the group to pursue a career in healthcare sciences – there were many times where I did not think I'd make it to a final thesis.

During this project, I was fortunate enough to have support form AstraZeneca. Thank you to Jamie Scott for your direction and assistance with this project and thank you for providing key intermediates and final compounds. Thank you to Daniel Hothersall (AZ) for providing GLUTag and HEK293-GPR119 cells, as well as advice for the culture. Similary, thank you to George Holz and Oleg Chepurny for providing the FLAG-GPR119 construct used in this thesis.

This work would not have been possible without my fellow Tate Group members. Even through the toughest of times, the many generations of Tater's have kept morale high with laughter, kindness and comradery. I have been fortunate enough to form lifelong friendships during this PhD, mediated through cake, science, office relocations and a collective enjoyment of the outdoor lunch. Theo, Charlie, Scott (LanthaWHAT lives on), Chiara, Jasmine, Ryan, Cass and the rest of you – here's to many more years of friendship. Thank you to my proofreaders for your constructive comments, advice and amazingly quick turnaround: Theo, Charlie, Scott, Chiara, Masha, Jennie and Ryan. I owe you all a drink.

On a personal note, thank you Toby for getting me to the end. Your promises of tea and food helped me get back to the laptop. Thank you for celebrating all the mini milestones along the way, your enthusiasm reminded me to be proud of where I am, the achievements and the journey I have made. I'm not sure I'd have made it without you.

To my family, thank you for your lifelong support. From the very beginning you have pushed me further than I thought I could get, I have no doubt that I wouldn't have achieved this without you. Tom – you are far more the intellect than I, and I am continually impressed and proud of your achievements, can't wait for your Nobel prize in Physics. Thanks for not gloating when you beat me to the PhD. Dad – your journey is a constant inspiration to me, proof that if you want something you can achieve anything you set your mind to and there are no barriers to success. Never stop being curious. Mum – you may be the only one of us without a PhD, but you are by far the smartest person I know. Your continued pursuit for knowledge about the subjects you love is by far the most important form of academia. To Grandma and Grampa, Martin and Michiko, and my family back in Iran thank you for avoiding the dreaded thesis question but most importantly for continuing to support me during my many years of education. Love you all.

Presentations and Publications

Poster Presentations

Kate Hadavizadeh *et al.* A chemical biology approach to target identification for drugs with unknown modes-of-action, Joint Warwick Imperial CDT Conference (London), June 2015

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, MaxQuant Summer School (Munich), June 2015 (Royal Society of Chemistry Travel Grant awarded)

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, MedImmune Symposium (Cambridge), September 2015

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, NCCR International Chemical Biology Symposium (Geneva), January 2016 (Conference Fee Waiver & Travel Grant awarded by conference committee)

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, Postgraduate Symposium (Imperial College London), July 2016.

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, RICT Interfacing Chemical Biology and Drug Discovery (Caen), July 2016

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, EMBL Chemical Biology Conference (Heidelberg), August 2016

Oral Presentations

Kate Hadavizadeh *et al.* A chemical biology approach to target identification for drugs with unknown modes-of-action, MRes Conference (Imperial College London), September 2014

Kate Hadavizadeh *et al*. A chemical biology approach to the target identification of bioactive molecules with unknown modes of action, Joint Warwick Imperial CDT Conference (Warwick), June 2016*

Kate Hadavizadeh *et al.* A chemical biology approach to target identification for drugs with unknown mode-of-action**,** AstraZeneca & ICB Symposium (Macclesfield), May 2015

Kate Hadavizadeh *et al. Design, synthesis and application of photoaffinity probes to enable target profiling of bioactive compounds against GPR119*, Postgraduate Symposium (Imperial College London), July 2017

*Prize for best presentation awarded.

Publications

Dominic J. Pollard, Cedric N. Berger, Ernest C. So, Lu Yu, Kate Hadavizadeh, Patricia Jennings, Edward W. Tate, Jyoti S. Choudhary, Gad Frankel, **Broad-Spectrum Regulation of Nonreceptor Tyrosine Kinases by the Bacterial ADP-Ribosyltransferase EspJ**, mBio Apr 2018, 9 (2) e00170-18; DOI: 10.1128/mBio.00170-18.

Alexandra S. Lubin, Ainoa Rueda-Zubiaurre, Holly Matthews, Hella Baumann, Fabio R. Fisher, Julia Morales-Sanfrutos, Kate S. Hadavizadeh, Flore Nardella, Edward W. Tate, Jake Baum, Artur Scherf, and Matthew J. Fuchter, **Development of a Photo-Cross-Linkable Diaminoquinazoline Inhibitor for Target Identification in Plasmodium falciparum**, ACS Infectious Diseases 2018 4 (4), 523-530, DOI: 10.1021/acsinfecdis.7b00228.

Wouter W. Kallemeijn, Gregor A. Lueg, Monica Faronato, Kate Hadavizadeh, Andrea Goya Grocin, Ok-Ryul Song, Michael Howell, Dinis P. Calado, and Edward W. Tate, **Validation and Invalidation of Chemical Probes for the Human N-myristoyltransferases**, Cell Chemical Biology 26, 892–900, June 20, 2019.

Abbreviations

2AG	2-arachidonulglycerol
2OG	2-oleoylglycerol
ABPP	Activity based protein profiling
Ac	Acetate
AEA	Anandamide
AfBPP	Affinity based protein profiling
AIBN	Azobisisobutyronitrile
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
B2AR	β-adrenergic receptor 2
bp	Base pairs
BSA	Bovine serum albumin
Bu	Butyl
CAA	Chloroacetamide
cAMP	Cyclic adenosine monophosphate
Cat.	Catalytic
CB1	Cannabinoid receptor 1
CB2	Cannabinoid receptor 2
cDNA	complementary deoxyribonucleic acid
CI	Confidence interval
CI	Chemical? ionisation
CMC	Critical micelle concentration
CRE	cAMP response-element binding protein
Ct	Cycle threshold
CuAAC	Copper catalysed azide alkyne cycloaddition
d	doublet
DCM	Dichloromethane
DDM	*n*-Dodecyl β-D-maltopyranoside
DIPEA	Diisopropylethylamine

DMEM	Dulbecco's modified Eagle medium
DMF	Dimethylformamide
DMSO	dimethyl sulfoxide
DMTMM	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride
dNTP	Deoxyribonucleotide triphosphate
DPP-4	Dipeptidyl peptidase 4
EC_{50}	Half maximal effective concentration
ECL	Extracellular loop
Epac2	cAMP-regulated guanine nucleotide exchange factor II
ESI	Electrospray ionisation
Et	Ethyl
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
FFA2	Free fatty acid receptor 2
FLAG	AspTyrLysAspAspAspAspLys
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
GDIR	Glucose dependent insulinotropic receptor
GFP	Green fluorescence protein
GIP	Glucose-dependent insulinotropic peptide
GLP-1	Glucagon-like peptide 1
GPCR	G-Coupled Protein Receptor
GPR119	G-protein coupled receptor 119
GPR120	G-protein coupled receptor 120
GPR40	G-protein coupled receptor 40
GTP	Guanidine triphosphate
h	Human
HBBS	Hands Balanced Buffer Solution
HEK293	Human embryonic kidney 293 cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonicacid
hERG	Human ether-*à-go-go*-related gene

HRMS	High resolution mass spectrometry
HRP	Horseradish peroxidase
HTS	High throughput screening
IA	Intrinsic Activity
IBMX	(3-isobutyl-1-methylxanthine
ICL	Intracellular loop
IF	Immunofluorescence
KO	Knockout
LC	Liquid chromatography
LDA	Lithium diisopropyl amine
LLE	Lipophilic ligand efficiency
LPC	Oleoyl-lysophosphatidylcholine
LPI	Lysophosphatidylinositol
m	Mouse
Me	Methyl
MECA	Melanocortin, endothelial, cannabinoid and adenosine
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NBS	*N*-bromosuccinide
NHS	National Health Service
NHS-ester	N-Hydroxysuccinimide-ester
NMR	Nuclear magnetic resonance
NP40	Nonyl phenoxypolyethoxylethanol
OEA	Ethanolamide olelylethanolamide
OG	*n*-Octyl-β-D-glucopyranoside
OGTT	Oral glucose tolerance test
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDC	Pyridinium dichromate

PEG	Polyethylene glycol
PFA	Parafformaldehyde
PHE	Public Health England
PKA	Protein kinase A
PPAR-γ	Peroxisome proliferator-activated receptor gamma
ppm	Parts per million
PTM	Post translational modification
PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
R_f	Retardation factor
rpm	Revolutions per minute
RT	Room temperature
rT	Reverse transcription
s	Singlet
SAR	Structure Activity Relationship
SDS	Sodium dodecyl sulfate
siRNA	Small interfering ribonucleic acid
SPA	Scintillation proximity assay
STR	Short tandem repeats
t	Triplet
T2D	Type 2 diabetes
TAE	Tris Acetate EDTA buffer
TBS	Tris buffered saline
TBST	Tris buffered saline with 0.1% Tween
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
THC	trans-Δ^9-tetrahydrocannabinol
THF	Tetrahydrofuran
T_m	Melting temperature
ToF	Time of flight

TR	Time resolved
Ts	Tosyl
TST	Twin Strep Tag
TZD	Thiazolidineiones
UV	Ulraviolet
WB	Western blot
WHO	Wold Health Organisation
WT	Wild type
δ	Chemical shift (ppm)

Contents

1. Introduction

1.1. GPCRs as therapeutic targets

1.1.1. The Structure and Function of GPCRs

With over 800 protein members, G-Coupled Protein Receptors (GPCRs) make up the largest class of cell surface receptors in the proteome. Within this superfamily are five distinct classes of GPCR based on sequence similarities: Class A (Rhodopsin), Class B (Adhesion), Class B (Secretion), Class C (Glutamate) and Class F (Frizzled/Taste2).[1,2] The superfamily share the structural feature of seven transmembrane spanning domains consisting of α-helical structures.[3] Within this structure, there are several distinct domains: the extra-cellular N-terminal domain, extracellular loops 1 to 3, seven transmembrane α-helices, intracellular loops 1 to 3 and the intracellular C-terminal domain (**Figure 1**).

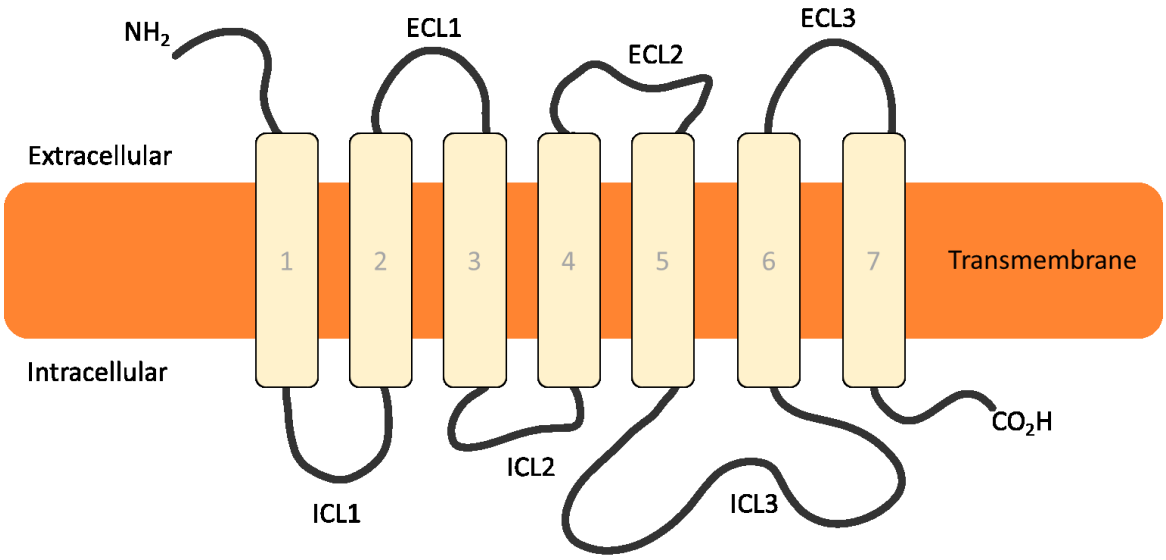


Figure 1: The generic structure of a GPCR showing several distinct domains: the extracellular N-terminal domain, transmembrane α-helices (yellow rectangles 1-7), extracellular loops (ECL) 1-3, intracellular loops (ICL) 1-3 and the C-terminal intracellular domain.

As these proteins straddle the cell membrane, they play a very important role in signal transduction from extracellular domains to the intracellular domain and thus elicit a biological response from an external stimulus (endogenous or exogenous). Interactions are not confined to extracellular small molecules as GPCRs have been shown to be activated by basal level signalling in about 7.5% of family members,

including both the adenosine and cannabinoid receptors.[4,5] GPCRs can also be activated through an interaction with other receptors, as is observed with the taste receptors.[6–8]

GPCRs are in constant equilibrium between active and inactive states. As such, ligands for GPCRs can be classified by their effect on the equilibrium. Agonists increase the signalling effect of the GPCR by binding to the protein and inducing a conformational change to an active state. Conversely, inverse agonists achieve the opposite effect of suppressing signalling by moving the receptor conformation to an inactive state. Antagonists act independently of the equilibrium by preventing binding of either an agonist or inverse agonist by blocking the binding site.[9] Binding of a ligand induces a conformational change within the transmembrane domain of the GPCR. In turn, this results in changes of conformation of the intracellular loop and modulates the interaction with effector proteins (such as G-proteins) in the cytoplasmic region of the cell (**Figure 2**).

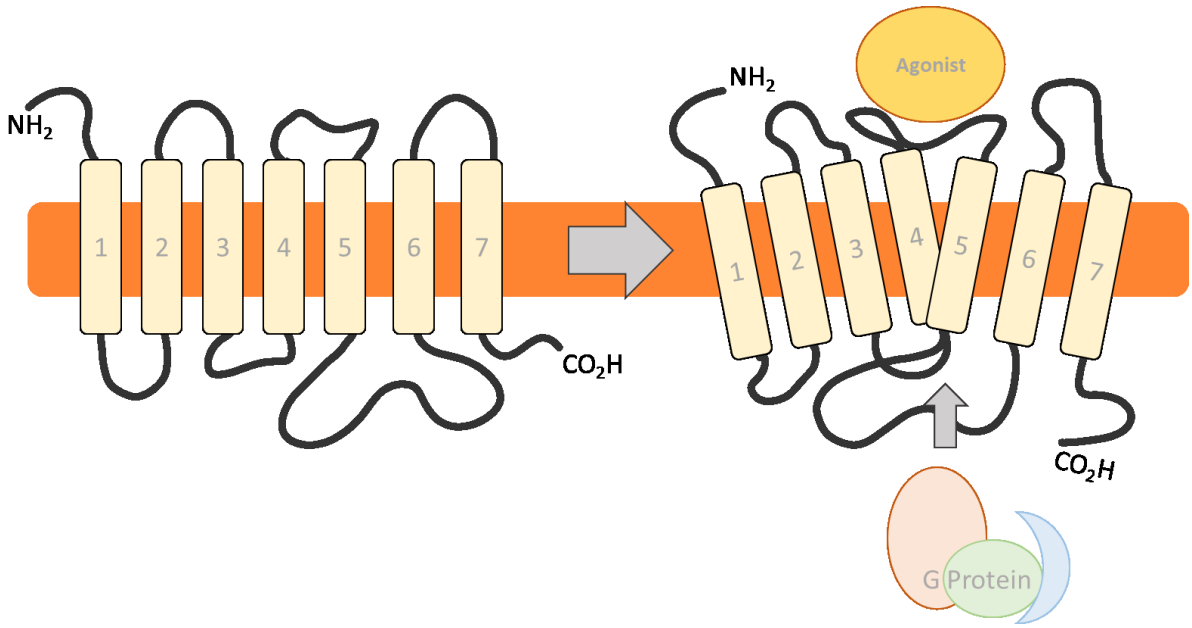


Figure 2: Binding of an agonist to a GPCR changes the conformation of the transmembrane domains, thus transmitting signal through the cell membrane where the intracellular loops respond by altering conformation to recruit effector proteins, such as the G Protein complex.

On the intracellular side of the GPCR, function is determined by the G-protein complex made up of 3 subunits: α, β and γ. The β and γ subunits are tightly associated and can be considered a heterodimer. The α subunit defines the identity of the GPCR into one of four major classes based on sequence identity: $G\alpha_s$ (stimulation), $G\alpha_i$ (inhibition), $G\alpha_q$ and $G\alpha_{12}$.[9] Each α subunit triggers a different downstream signalling pathway,

giving a vast range of responses intracellularly. The understanding of regulation and signalling of GPCRs is an ever growing and debated field; there are a number of reviews available that discuss this topic.[10–12]

1.1.2. GPCRs in Health & Disease

As GPCRs represent such a large class of proteins, it is perhaps unsurprising that they play an important role in modern medicine and drug discovery. Membrane proteins are said to be the target for around 60% of the current therapeutic medicines landscape.[13,14] Within the membrane protein domain, approximately half of these drugs interact with GPCRs.[15]

GPCR signalling is diverse, and has been shown to be involved in a range of important of health and disease processes such as cardiovascular function,[16] cognitive responses[17] and neurodegenerative diseases with vastly unmet needs such as Alzheimer's[18] and Huntingdon's disease.[18] Oncology is currently a growing area of GPCR drug discovery, with many novel GPCRs being shown to be involved in cancer growth and development.[19] At the time of writing in 2017, Hauser *et al*. identified 21 approved drugs with neoplastic indications.[15] More recently, an accelerating area of research and development is in the treatment of diabetes and obesity[20]; approximately 9% of clinical trials active in 2017 for diabetes and obesity were for agents targeting GPCRs.[15]

1.1.3. Challenges of Working with GPCRs

Despite the significant contribution to challenges in health and disease that GPCRs may be able to address, they are notoriously troublesome to work with. This is evidenced by the fact that the first crystal structure for a GPCR, rhodopsin, was only published in 2000 and is currently the only GPCR to be purified and structurally characterised from its endogenous source.[21,22] It then took a further 7 years for the structure of β_2-adrenergic receptor to be elucidated through expression, purification and crystallisation of the GPCR.[23,24] This will be discussed in more detail in a later chapter.

As GPCRs are transmembrane receptors, the proteins typically have a high degree of hydrophobicity especially within the transmembrane helical regions,[25,26] and a vast

array of post translational modifications. Upon solubilisation of the protein and removal from its endogenous membrane environment, the proteins are prone to aggregation and conformational change. For chemical probes, this can lead to a loss of interaction with the receptor binding site of interest, unless it is covalently modified. As the state of a GPCR is not static during its function, interactions with ligands are similarly transient even within the membrane as the GPCR cycles through active and non-active conformations.[27] Upon prolonged interaction with a ligand, GPCRs are prone to mechanisms such as internalization of the receptor and desensitization of the receptor as a means of modulating signalling.[28] A common feature across all classes of GPCRs is their low endogenous expression levels within tissues, enhancing the challenges compared to more highly expressed proteins[29,30]

1.2. Diabetes

1.2.1. Prevalence and Consequence

Diabetes is a chronic disease that is growing in prevalence. Historically it is a condition associated with affluence, however global diabetes statistics show the growth of diagnosed and undiagnosed cases globally, especially in low- and middle-income countries.[31] The World Health Organisation (WHO) state that as of 2015 the global prevalence of diabetes has risen to 8.5%. In 2016, it estimates that 1.6 million deaths can be directly attributed to diabetes, with a further 2.2 million deaths being linked to high blood glucose levels. In 2015, diabetes was the sixth greatest cause of death globally.[32]

Within the UK, the rise in obesity and the concurrent rise in type 2 diabetes has been recognised as one of the modern public health epidemics leading to the launch of many campaigns by Public Health England (PHE), such as the Health Matters: Preventing Type 2 Diabetes resource.[33] In England, type 2 diabetes affects 3.8 million people, with a further 5 million people identified as at risk of developing the disease. With the current rate of growth, PHE estimate that 1 in 10 people will develop the disease by 2034.[33]

Whilst there is a lot of public awareness of diabetes, the extremity of consequences if not diagnosed and properly treated are less common knowledge. Those who are

diagnosed with type 2 diabetes are at a greater risk of developing cardiovascular heart disease, including heart attack and stroke, diabetic retinopathy leading to blindness, kidney failure, peripheral neuropathy, diabetic foot disease and ultimately premature death.[33] Given the array of complications associated with diabetes, there is a large financial impact with approximately 9% of the annual National Health Service (NHS) budget going towards diabetes care.[33] Therefore, access to effective therapeutic intervention is essential for the growing population of those living with the condition.

1.2.2. Normal Blood Glucose Regulation

Blood glucose levels are normally regulated by the means of a negative feedback loop comprising of insulin, glucagon and glycogen. Insulin and glucagon are the key hormones that ensure homeostasis is achieved, while glycogen is the polysaccharide storage form of glucose found within the muscle and liver as an energy reserve.[34] If blood glucose levels drop, signalling to the pancreas initiates the production of glucagon from the α-cells in the pancreatic islets of Langerhans, which in turn triggers the conversion of glycogen to glucose in the liver. Conversely, if blood glucose levels rise excessively, insulin is secreted from the β-cells in the pancreas, and through cell-surface insulin receptors mediates glucose transport into the cell in order to be stored as glycogen.

1.2.3. Abnormal Blood Glucose Regulation: Diabetes

In people with diabetes, this feedback loop is dysregulated due to the body's inability to produce the endocrine hormones or an acquired resistance/insensitivity to insulin. Thus, blood sugar levels are not maintained at the normal levels leading to episodes of hypoglycaemia (low blood sugar) or hyperglycaemia (high blood sugar).[35]

Diabetes can be subclassified into different disease states depending on the source of dysregulation. Type 1 diabetes (also referred to as insulin dependent diabetes) is the consequence of the lack of production of insulin, commonly due to an autoimmune response towards the pancreatic β-cells.[35] Due to the physiology of type 1 diabetes, it is most commonly diagnosed during childhood and subsequent disease management requiring daily bloody glucose monitoring and life-long self-medication with insulin injections.

Type 2 diabetes (also known as non-insulin dependent diabetes) is a progressive form of the disease, often being diagnosed during adulthood as a result of acquired insulin resistance as a consequence of environmental factors and some genetic predisposition.[36–38] There have been substantial links made between diet and the development of type 2 diabetes, with links being drawn between the obesity epidemic and growing type 2 diabetes prevalence.[39–44] Type 2 diabetes develops over time as abnormal function of the β-cells increases, presenting as an increased resistance to insulin.[43,45] As this function declines, and resistance increases, a reduction in insulin secretion is observed. However, due to the gradual progression of the disease, it is often not diagnosed until years after the initial decline in secretion capacity.[37]

1.2.4. Current Therapies for Type 2 Diabetes

Approximately 90% of all diabetes diagnoses are type 2. Given the different physiology of disease to type 1, daily insulin injections are not the best form of disease management as typically these patients can still produce insulin, just not to the degree that elicits a complete biological response. Within the clinic, there are many options to treating type 2 diabetes; interestingly, the majority of patients can control their diabetes with good diet and adequate exercise, however the remaining 49% require oral medication.[46]

There are a number of classes of medication available for type 2 diabetes, varying in the mechanisms of action by targeting different stages of the glucose regulation cycle:

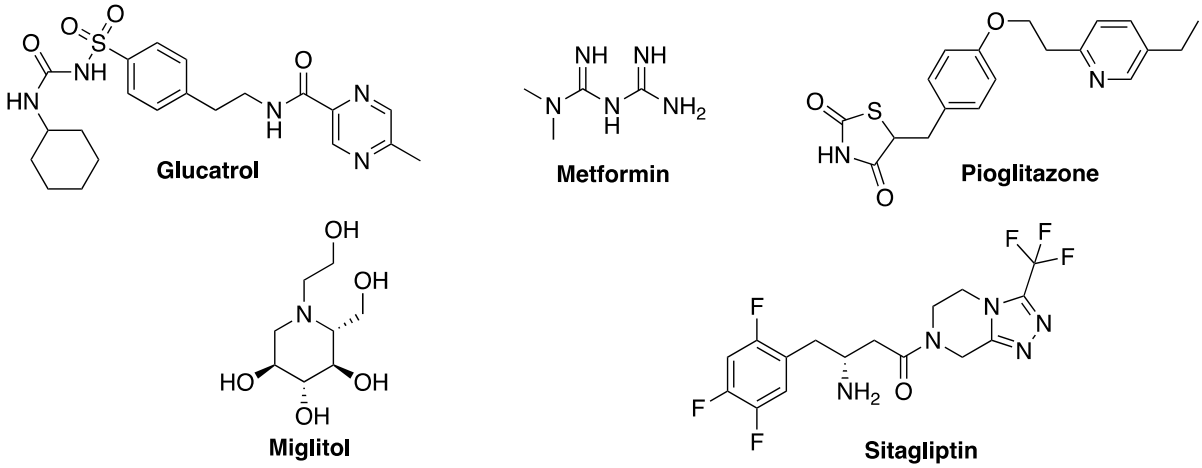


Figure 3: Examples of current therapeutics for type 2 diabetes from each class: Insulin secretagogues (glucatrol), Insulin sensitizers (metformin and pioglitazone), glucose gut-absorption modulators (miglitol) and glucose-lowering effectors (sitagliptin).

insulin secretagogues, insulin sensitizers, glucose gut-absorption modulators and glucose-lowering effectors (examples shown in **Figure 3**).[47]

Insulin Secretagogues act upon receptors on the surface of pancreatic β-cells, triggering a depolarization through the closure of potassium channels, hence stimulating insulin secretion.[47] This class of drugs are one of the longest standing prescribed drugs to treat type 2 diabetes in the US in the form of sulfonylureas (such as **Glucatrol**). However, sulfonylureas and their analogues commonly cause unfavourable side effects such as hypoglycaemia (due to excessive insulin secretion) and weight gain.[48]

Insulin Sensitizers act by increasing the response to insulin in the tissue.[37] Different classes of sensitizer have different mechanisms of action. The biguanide **Metformin** (first described in 1957),[49] is the current first in line treatment for type 2 diabetes after lifestyle changes.[50,51] It has a dual mode of action for reducing hyperglycaemia: by reducing hepatic glucose output and by improving insulin sensitivity in skeletal muscle tissue.[52] Aside from its effective blood glucose effect, the drug also exhibits benefits over the sulfonylurea class of compounds as it can act as an appetite suppressor, leading to weight maintenance (and in some cases, slight weight loss).[47] Amongst the available type 2 diabetes therapeutics, it is the only one to have clinically demonstrated a reduction in mortality rates.[53] However, side effects such as gastrointestinal intolerances have been reported but these can be managed by meal time dosing.[54]

Thiazolidineiones (TZD) such as **pioglitazone** are another class of insulin sensitizers that bind to the peroxisome proliferator-activated receptor gamma (PPAR-γ) and alter gene transcription in adipose tissue.[46] This results in a lowering of free fatty acids in the target cells, increasing insulin sensitivity in skeletal muscle tissue, and also increasing the function of β-cells by reducing lipotoxicity.[47] However, compared to Metformin, the list of side effects are much less desirable with reported increased risk of myocardial infarction, weight gain and edema.[37,55]

Glucose Gut-absorption Modulators are a class of α-glucoside inhibitors, such as **miglitol**. Inhibition of this pathway results in the decreased breakdown of dietary complex carbohydrates, leading to delayed glucose absorption in the gastrointestinal tract.[47] Compared to other available agents, these inhibitors are not as clinically

effective for glycaemic control, and can cause gastrointestinal side effects.[47,56]. However, the lack of weight gain and hypoglycaemia offers a clinical alternative.[37]

Glucose-lowering Effectors offer an alternative to small molecule drug mediated responses, by employing the endogenous peptide hormone effectors, incretins, that are involved in glucose regulation. Incretins are advantageous as they can offer excellent glycaemic control without side effects such as hypoglycaemia and weight gain.[57] Glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic peptide (GIP) are both hormone peptides that are secreted in response to ingestion of complex carbohydrates, leading to a glucose-dependent insulin response.[58] As these are small peptides, it is possible to supplement the body's natural response to improve glycaemic control. However, the use of the native sequences is limited by their degradation by the enzyme dipeptidyl peptidase 4 (DPP-4). Therefore possible clinical strategies are to use GLP-1 mimetics that resist the DPP-4 degradation, such as the glutide liraglutide. Given the structure of incretins and glutides, administration is limited to injectables which is less preferable for the patient. DDP-4 inhibitors (such as gliptins, for example **sitagliptin**) can be used therapeutically to prevent the degradation of native GLP-1 and therefore prolong its insulin-dependent blood glucose lowering activity, however, this therapeutic strategy is limited by the endogenous rate of GLP-1 production by the body.[59,60]

1.3. GPR119 as a Potential Therapeutic Target

1.3.1. Expression of GPR119

Following on from the promising action of incretins on blood glucose levels, G-protein coupled receptor 119 (GPR119) was identified as a protein that is capable of stimulating the secretion of these incretins, thus offering as a potential therapeutic target for an oral drug for type 2 diabetes.[61] Initial biochemical characterisation supports the role of GPR119 in insulin secretion with the receptor being identified in the pancreatic β- cells[62], and reduction of insulin secretion upon siRNA knock down of GPR119.[63] Activation of glucose-dependent insulin secretion was also shown using a published GPR119 agonist.[64] It is worth noting that GPR119 is not the only GPCR associated with insulin secretion to be identified within pancreatic β- cells; GPR40 was the first GPCR to be associated with insulin secretion,[65] and overexpression of GPR120 has been identified within intestinal L-cells suggesting there may well be a more complex network of glucose related signalling within the pancreas and gut.[66]

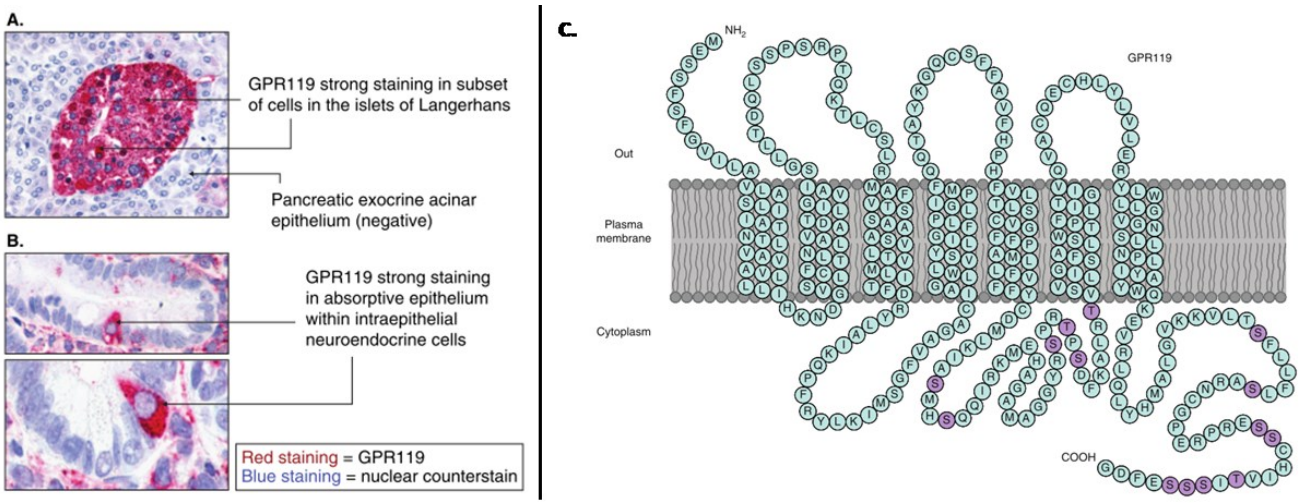


Figure 4: Immunocytochemical staining showing GPR119 localisation (red) within the islets of Langerhans (**A**) and small intestine (**B**). **C** Amino acid sequence of human GPR119 within the membrane, showing predicted extracellular loops, transmembrane regions, intracellular loops and potential phosphorylation sites (purple). Figures taken from Fyfe *et al.* **2008**.

GPR119 expression levels have been investigated across a range of tissue types, with the majority of tissues demonstrating an absence of GPR119. However, expression was identified within the pancreas and foetal liver, as well as high expression levels being identified within the gastrointestinal tract.[67–69] Within these specific tissues, expression was further identified within the β-cells of the islets of Langerhans (as mentioned above)[63,64] and the K- and L-cells in the small intestine (**Figure 4**).[70,71] This

expression profile strengthens the potential for GPR119 to act as a glucose-dependent insulin secretion receptor, as K- and L-cells have been shown to be responsible for secretion of GIP and GLP-1 respectively.[70,71]

There are a number of synonyms for GPR119 within the literature: SNORF25,[68,69] RUP3,[72] GPCR2,[73] 19AJ,[74] and glucose-dependent insulinotropic receptor (GDIR).[67] The GPCR belongs to the A-type rhodopsin family, with some reports classifying it as a sub-group of the bioamine receptors,[61] part of the free fatty acid receptors cluster,[62] while others have classified it as a member of the more general MECA (melanocortin, endothelial, cannabinoid and adenosine) class.[75,76] More recently, it is now accepted to be an endocannabinoid receptor.[77,78] The ambiguous classification is due to the

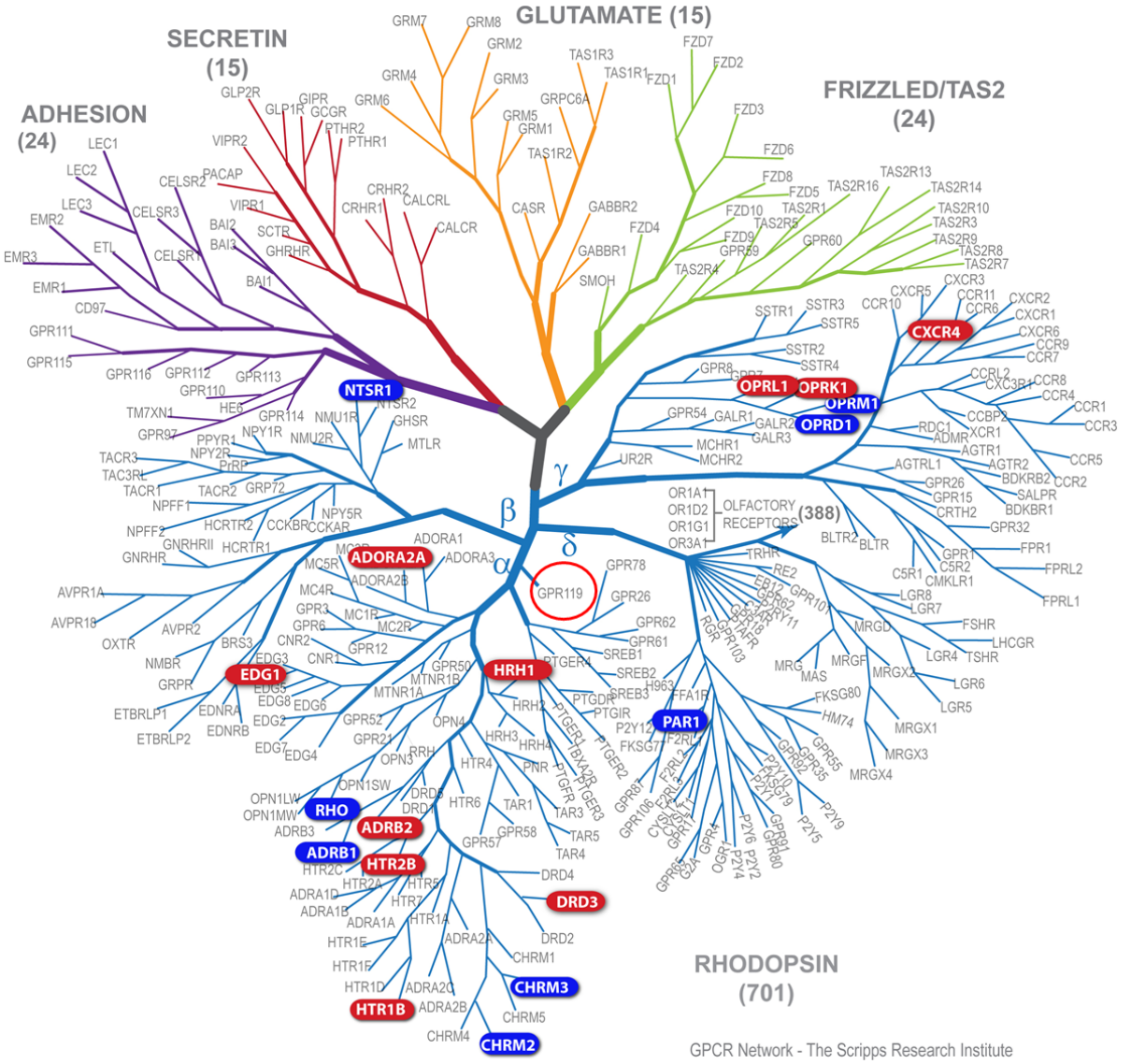


Figure 5: GPCR phylogenetic tree featuring all GPCRs grouped into the classifications. GPR119 is circled in red. Image downloaded from GPCR Network – The Scripps Research Institute community page.

differences in structural comparisons, and comparisons of endogenous ligands. Structurally, the closest relative in terms of sequence identity are the adenosine (A_1 and A_3) receptors with a transmembrane sequence identity of 28%,[75] whereas there are little sequential similarities with the cannabinoid receptors.[78] When GPR119 is visualised within the global GPCR family, it can be seen that it does sit on its own with no close neighbours (**Figure 5**). This isolation does mean that structural information is limited by analogy to close neighbours but may make GPR119 an attractive target with respect to the potential selectively and specificity of agonists for the receptor.

Several mammalian isoforms have been identified in mouse, rat, dog and human, with the amino acid sequence being generally conserved with some variation in length (human GPR119 2D amino acid chain shown in **Figure 4**).[67]

1.3.2. Endogenous Ligands for GPR119

The deorphanization of GPR119 was first described by Overton *et al.* in 2006 as the outcome of a screening of orphan GPCRs as potential receptors for the endogenous fatty acid ethanolamide oleoylethanolamide (OEA) (**Figure 6**).[62]

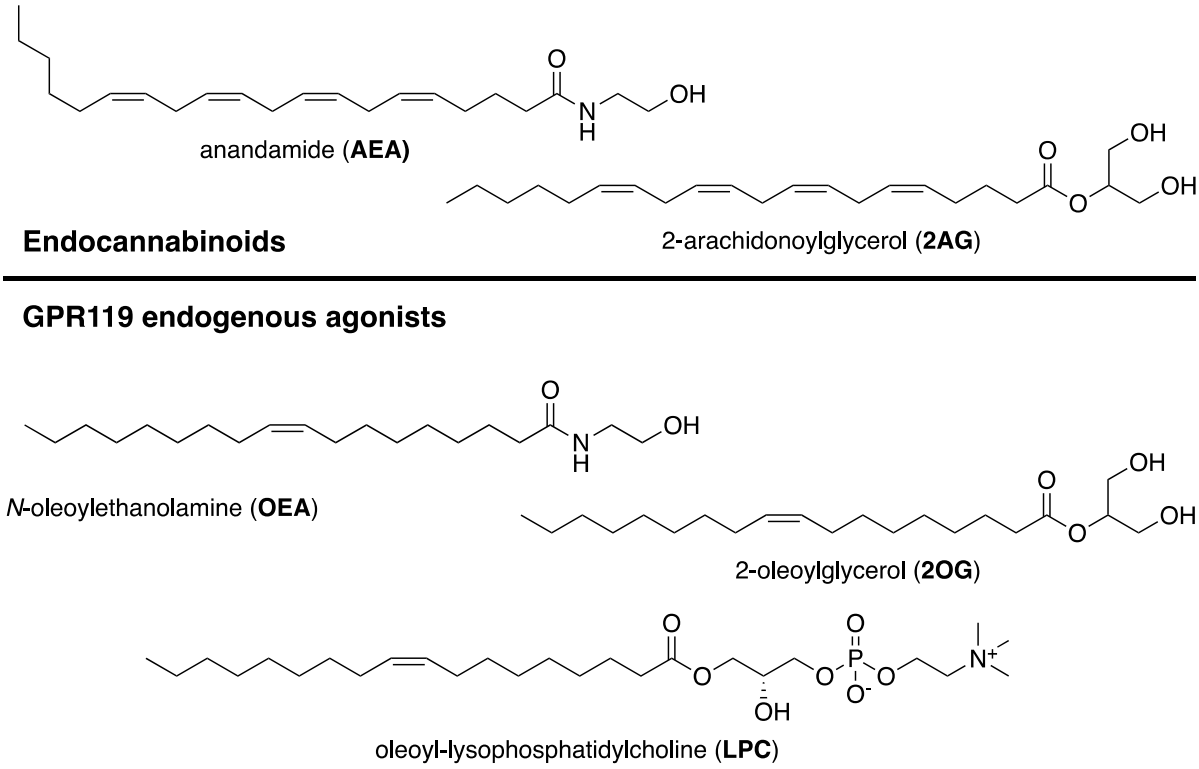


Figure 6: Chemical structures of the endocannabinoids and the structurally related GPR119 agonists.

OEA peaked interest as it is the monounsaturated analogue of the endocannabinoid anandamide (AEA), however it did not stimulate cannabinoid receptors, and presented with an unknown biological function.[79] A number of studies later linked OEA to satiety regulation and reduction in weight gain in rodent feeding models.[79–83] Overton *et al*. also demonstrated the GPR119 specific response to OEA in a yeast model system, and a reduction in food intake in rats upon OEA administration.[62] Several further studies have also found evidence that OEA as an agonist of GPR119.[84–86]

OEA is not the only endocannabinoid identified as an agonist of GPR119: 2-oleoylglycerol (a monounsaturated analogue of endocannabinoid 2-arachidonoylglycerol, 2AG) showed agonism of the receptor.[87] The phospholipid oleoyl lysophophatidylcholine (LPC) was also identified as another endogenous ligand.[63] A number of lipid amides were screened against GPR119 by Chu *et al.* in order to better understand the scope of ligands for the receptor (**Figure 6**).[84] A recent publication provides evidence for oleoyl-lysophosphatidylinositol (oleoyl-LPI) stimulating GLP-1 secretion in a GPR119 dependent manner.[88] Attempted agonism of GPR119 with small molecule cannabinoid ligands such as trans-Δ^9-tetrahydrocannabinol (THC) showed no activity.[78]

1.3.3. Signalling Cascade & Mechanism of Action

Initial cellular work with GPR119 overexpression systems led to an increase in cyclic adenosine monophosphate (cAMP) levels, as observed by multiple groups.[64,69] This is indicative of a GPCR coupled to $G\alpha_s$ subunit. Extracellular agonism leads to intracellular recruitment of this G protein ($G\alpha_s$). Downstream activation of adenyl cyclase (a catalyst for the conversion of ATP to cAMP and pyrophosphate) then leads to the observed increase in intracellular cAMP levels. These results were correlated with direct agonism of GPR119 with OEA and LPC leading to an increase in intracellular cAMP levels.[62] Downstream, this cAMP secondary messenger activates secondary effector proteins (such as protein kinase A (PKA)) leading to protein phosphorylation and de/activating events, alongside gene regulation.

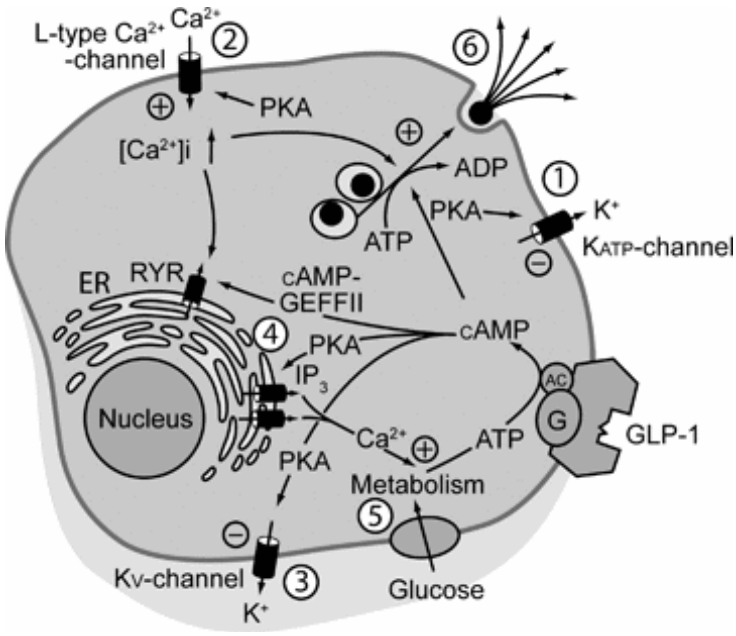


Figure 7: Intracellular signalling cascade upon agonism of the GLP-1 receptor on the surface of a β-cell. Events **1** and **3** represent changes in membrane polarisation through the closure of K-ATP channels. Stages **2** and **4** show the increase in Ca^{2+} concentration within the cells as release from the ER and Ca^{2+} channel influx. **5** shows the uptake of glucose, because of changes in membrane polarisation. **6** represents the exocytosis of insulin containing vesicles as a result of the above. Reproduced with permission from Vilsbøll, **2009.**

The effect of GLP-1 on β-cells has been studied and reviewed in depth and gives an insight into the subcellular mechanisms that take place upon the initial agonism, that ultimately lead to insulin secretion from these cells (overview in **Figure 7**).[89,90] As mentioned above, agonism of the receptor in the presence of elevated glucose activates adenylyl cyclase, resulting in the accumulation of cAMP and activation of PKA.[90] cAMP has also been stated to activate the cAMP-regulated guanine nucleotide exchange factor II (cAMP-GEFII or Epac2), which together with PKA act as secondary effector proteins for a number of signalling pathways that control a range of β-cell functions.[91,92] In order for insulin to be secreted, essential phosphorylation events are carried out by PKA while Epac2 activates members of the Ras superfamily. Together, PKA and Epac2 shift ATP-sensitive potassium channels (KATP) to a closed configuration, which causes a change in membrane potential that sensitises the cells to glucose. Epac2 has also been shown to be involved with Ca^{2+} release from the endoplasmic reticulum, increasing the intracellular concentrations, which in turn increases the availability of insulin secretory vesicles that are eventually released in an exocytosis event.[90,93]

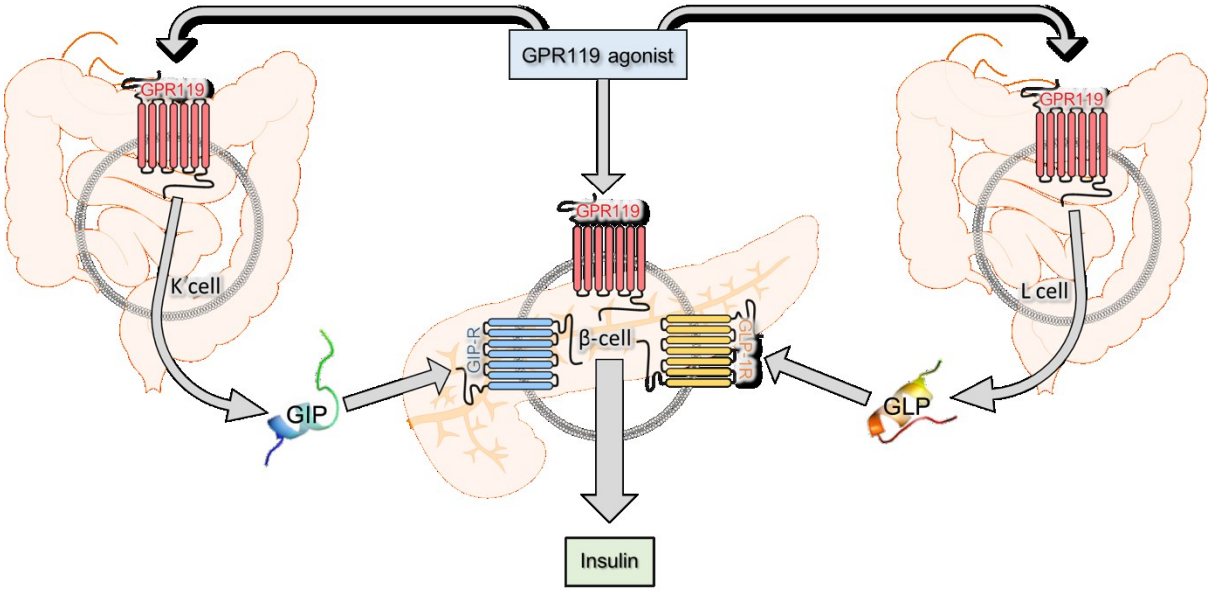


Figure 8: GPR119 signalling in K- and L-cells leading to incretin release. GPR119, GLP-1R and GIPR agonism in the β-cells leading to insulin secretion.

For GPR119, the cellular response to this increase in cAMP is dependent on cell type. Within intestinal L-cells, it has been shown that GPR119 agonism leads to a cAMP dependent release of the enteroendocrine hormone GLP-1.[94] Similarly, in intestinal K-cells, GIP is secreted in response to GPR119 agonism. Both these incretins can then signal via their specific receptors on the β-cell surface to illicit an insulin secretion response. GPR119 agonism also leads to a direct insulin secretion response within the β-cells thus leading to an increase in insulin production from the β-cells directly as a result of both mechanisms (**Figure 8**).[70,95] As GLP-1 can improve β-cell regeneration and proliferation, the overall viability of the β-cells may also improve (indirectly) upon GPR119 agonism.[93] This multimodal approach to improve glucose-dependent insulin secretion has proven GPR119 to be an attractive target for medicine chemistry efforts to identify a new therapeutic agent for GPR119.

1.3.4. Synthetic Agonists for GPR119

The first reported synthetic agonist of GPR119, **AR231453,** was developed by Chu *et al.* in conjunction with Arena Pharmaceuticals as the result of a high throughput screen against the receptor. The agonist was shown to be potent with an EC_{50} = 5.7 ± 1.6 nM in transfected cells and selective for GPR119. They also demonstrated efficacy for glucose-dependant insulin secretion in an isolated mouse islet model.[64,96] However, this original agonist was limited by toxicity observed in mice receiving multiple doses. Subsequent development led to the alteration of the pharmacophore by replacing the potentially toxic aniline and nitro moieties with a bicyclic heterocycle to produce **APD668** (**Figure 9**).[97] This compound then entered phase 1 clinical trials as a first-in-class candidate for GPR119, however the trial did not progress beyond phase 1 due to the off target inhibition of CYP2C9 and unexpected accumulation of a hydroxylated metabolite with a long half-life that was considered too great a liability to continue the study.[94]

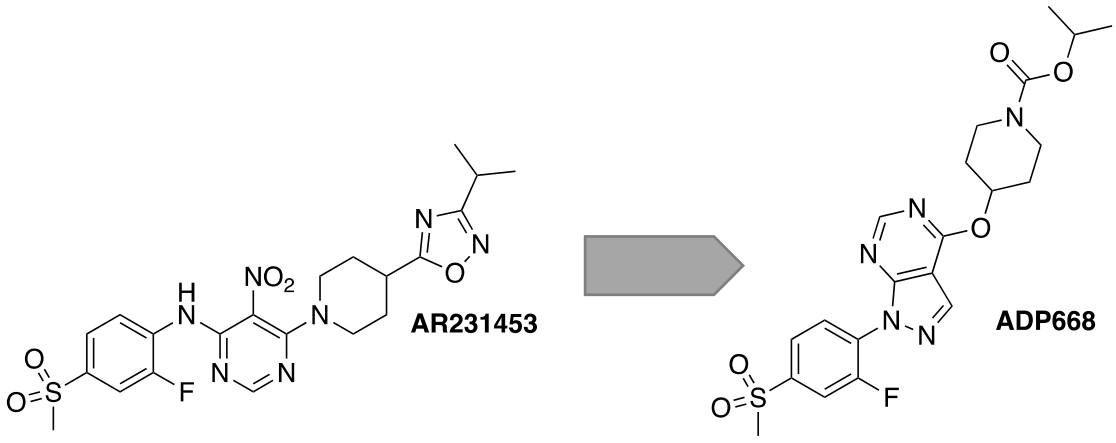
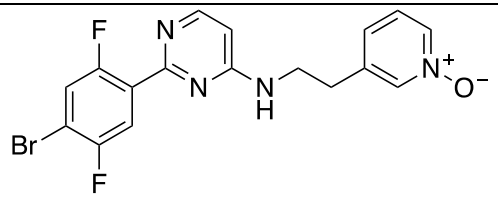
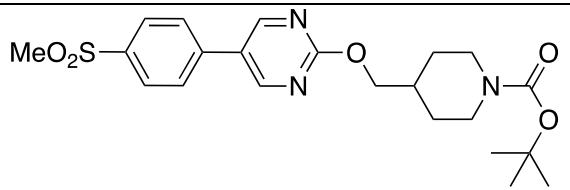
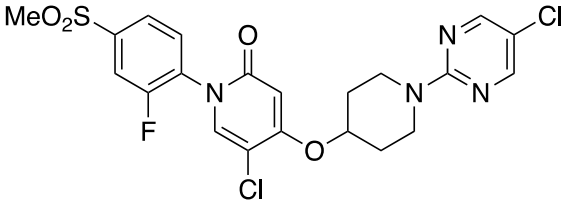
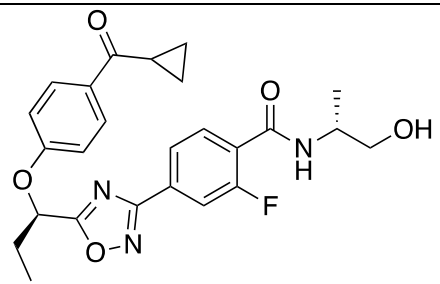


Figure 9:The development of the first synthetic agonist for GPR119 (**AR231453**) into the first-in-class candidate for GPR119 agonism in phase I clinical trials (**ADP668**).

Since the publication of this work, there have been many patents, clinical trials and research papers for the development of GPR119 agonists with novel features, however at the time of writing a GPR119 agonist has yet to make it to through a clinical trial to become a viable approved medicine. There are a number of reviews available on the development of agonists for GPR119.[46,57,67,98–102] **Table 1** gives a snapshot of the current GPR119 agonist landscape.

Compound *Company*	**Structure**	**EC_{50}**	**Clinical Trial**	**Comments**	**Ref.**
AS166958 *Astella*		111 nM	-	Improved blood glucose levels and pancreatic insulin levels in *db/db* mice	[103]
Biovitrum		14 nM	-	Patents explore interchanging various heterocycles at the core	[104]
BMS-903452 *Bristol-Myers Squibb*		14 nM	Phase I	Safe and tolerable but did not alter GLP-1 levels in subjects	[105]
DS-8500a *Daiichi Sankyo Co.*		51.5 nM	Phase II	Add on therapy with sitagliptin in Japanese T2D patients. Showed significant glycaemic benefits as a combination therapy.	[106–108]

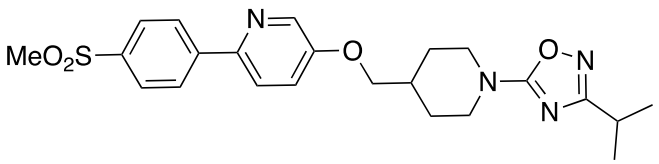
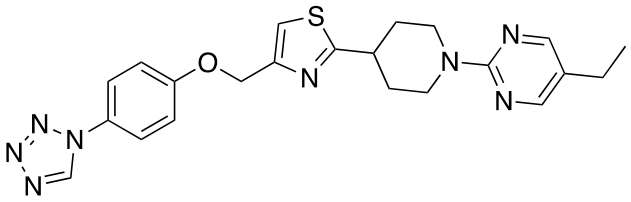
GSK1292263 *GSK*		6.9 nM	Phase II	Tolerated well in Phase I, but no clinical efficacy observed in Phase II in subjects with T2D	[109,110]
MBX-2982 *Metabolex*		3.9 nM	Phase II	Drug was safe and well tolerated in phase I and produced statistically significant increase in insulin secretion. Phase II carried out but not carried forward after	[46,99,102]
LEZ73 *Novartis*	Structural data not disclosed	N/A	Phase I/II	Phase II discontinued in 2014	[101]
PSN821 *Prosidion Ltd./Astellas*	Structural data not disclosed	N/A	Phase II	Phase II discontinued in 2012.	[46,102]

Table 1: A snapshot of the GPR119 agonist landscape. Many potent agonists of GPR119 have been discovered and have entered clinical trials. However, there is yet to be a clinical candidate for an oral GPR119 agonist as a treatment for diabetes. All EC_{50} data reported is for agonism of the human GPR119 receptor.

Despite the vast array of potent GPR119 agonists with favourable pharmacokinetic and lack of toxicity, so far, all clinical trials have been discontinued at phase II or earlier. The underlying failure of the trails is not often disclosed, but the lack of safety issues suggests a lack of efficacy. Given robust results in animal models, but lack of efficacy in human studies, there may be variation in cross-species pharmacodynamics.[100]

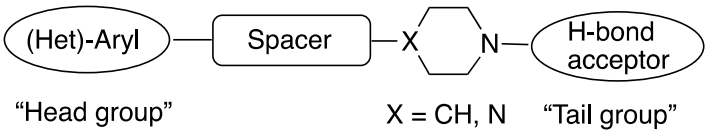


Figure 10: Example of model architecture for GPR119 agonists

The vast amount of literature available for GPR119 agonists allows a comparison of the medicinal chemistry and has allowed the creation of "model architecture" for synthetic agonists of GPR119 by a number of groups (example in **Figure 10**).[101,111] Interestingly, since this model was developed, a second generation of GPR119 agonist architecture has been explored to introduce more conformational rigidity by the inclusion of bridged piperidines. As an example, Pfizer demonstrated that derivatizing one of Arena Pharmaceutical's original scaffolds (**APD597**, a partial agonist) to an oxabicyclic froze the equatorial and axial conformations of the piperidine ring to give an agonist and an antagonist, demonstrating the potential of agonist-antagonist relationships of isomeric GPR119 agonists (**Figure 11**).[112]

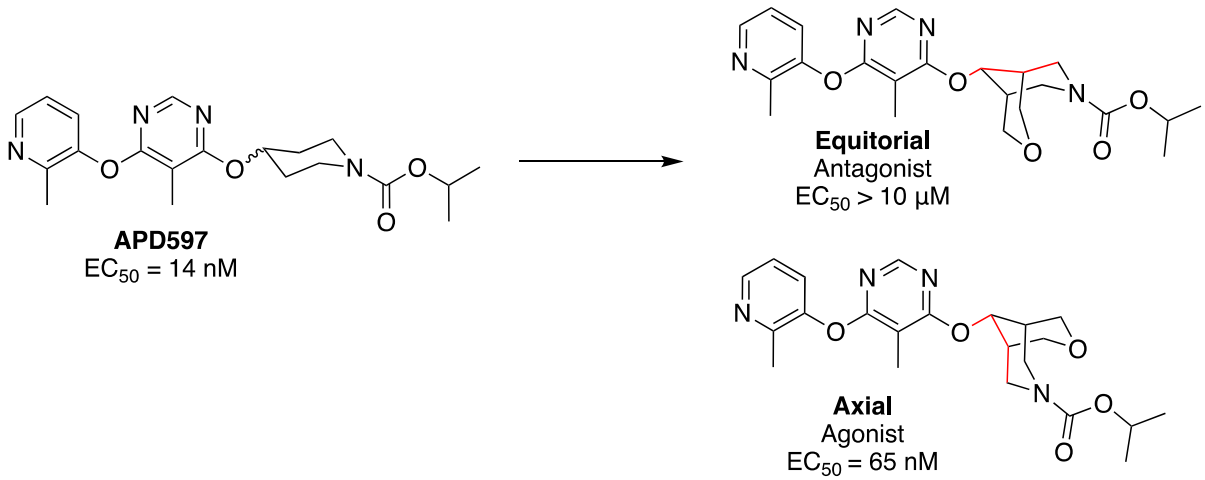


Figure 11: The effect of rigidifying GPR119 agonist architecture on unveiling the potential agonist-antagonist relationship within isomeric agonists

1.3.5. Structural Characterisation of GPR119

Despite the vast quantity of ligands now available for GPR119, there is little in the literature by way of structural characterisation and molecular pharmacology of the GPCR itself. Bar a handful of examples, characterisation of protein expression in tissues and cell lines is carried out predominately by mRNA detection methods. This will be discussed in more detail later in this thesis.

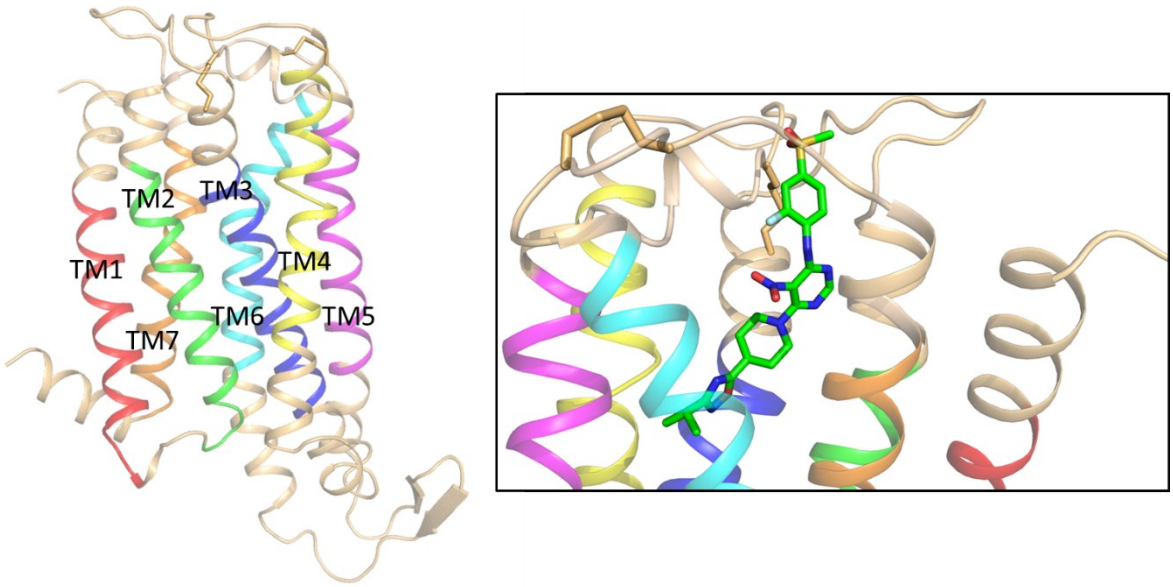


Figure 12: Model of GPR119 based on mutation-guided docking of **AR213453** published by Norn *et al*.. Image created in PyMol to show transmembrane region of each helix and the best docking model of the agonist with sulfonyl moiety interacting with the surface loop region.

A solid-state structure for GPR119 has yet to be obtained, however Norn *et al.* have published a mutation-guided model of GPR119 (using a multi-template approach with the adenosine A2a receptor, CXCR4 receptor and the dopamine D3 receptor) with **AR213453** docked into the proposed binding site.[113] While this is not a crystal structure, the unbiased mutation guided model that has been published is a useful resource to guide drug discovery in combination with the GPR119 model architecture discussed above.[113,114] Interestingly, their model did highlight some binding preferences for GPR119 agonists and suggestions of areas that were more amenable to modification; the sulfonyl moiety common to many GPR119 agonists appears to sit in the loop region of the model, with room for larger substitutions (**Figure 12**).[113]

1.4. Studying GPCRs: A Chemical Biology Approach

1.4.1. Chemical probes: Affinity Based Protein Profiling

The development of chemical probes provides useful tools to explore the biological properties of proteins; such as their activity, relationships and interactions with other proteins and cellular machinery. These properties can often be elucidated, using chemical probes, regardless of the target protein's expression level or abundance. Depending on the interaction of interest, chemical probes can be designed to include an activity dependant specificity element (Activity Based Protein Profiling, **ABPP**) or an affinity element designed to explore non-catalytic regions of interest (Affinity Based Protein Profiling, **AfBPP**).[115] As this thesis focusses on the design and synthesis of photoaffinity probes, discussed below, therefore ABPP will not be discussed further.

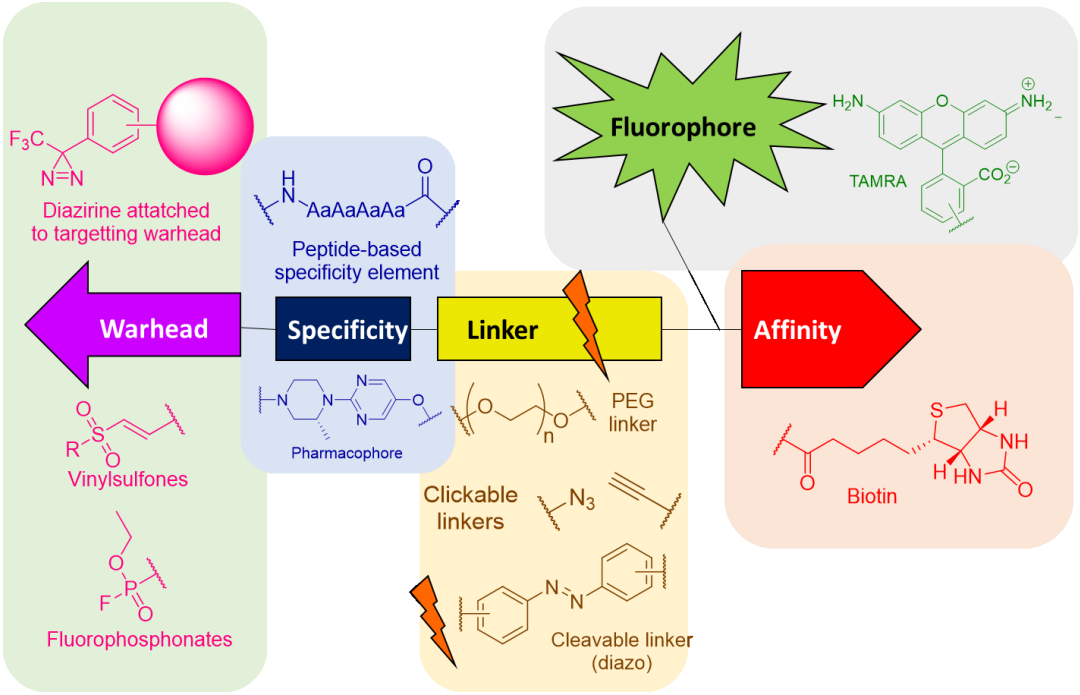


Figure 13: The components of an Activity Based Probe or Affinity Based Probe: containing a warhead, specifitiy element such as a pharmacophore to target the interaction of interest, and a linker that can contain clickable groups, affinity groups and reporter tags to enable characterisation of probe intereactions.

AfBPP relies on the design of robust affinity based chemical probes. The general structure of an affinity probe contains three key functional groups: the pharmacophore to target the interaction of interest, a capture group to allow covalent modification and a reporter tag to provide information about the interactions between probe and protein (**Figure 13**). These elements are discussed more thoroughly in future sections.

The affinity element of the probe acts as bait to capture the protein or interaction of interest. Depending on the biological interaction or process being studied, this bait can take many forms; for example it could be based on a drug of interest (for target profiling), the peptide sequence of an interface (for profiling protein-protein interactions) or a modified molecule for a post translational modification (PTM) (for tracking processing and PTM interactions in pathways) as a subset of examples.[116–121] The nature of the probe means the interactions are intrinsically reversible (in contrast to ABPP), unless the interaction is trapped by an activatable event to form a permanent linkage between probe and target. Typically, photochemistry is used to unmask a reactive group, enabling covalent bond formation and protein capture. Reporter tags can be incorporated into the probe directly or appended after capture using two-step labelling, utilising bio-orthogonal ligation.

Affinity based probes are advantageous for targeting proteins without nucleophilic residues in the active site, where the active site is not understood or where the mechanism of action is unknown all which precludes the use of activity-based probes.[122] Thus, the application of affinity-based probes greatly widens the field of exploration in the field of proteomics.[123] Examples of photoaffinity probes used to investigate GPCRs will be discussed later in this section.

1.4.2. Photoaffinity Labelling

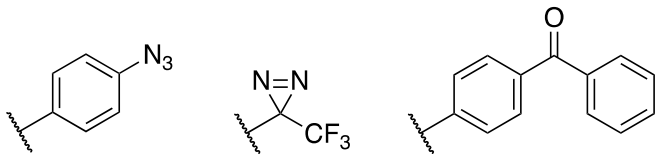
Photoaffinity labelling is an invaluable tool used in the field of chemical biology and was first reported in the 1960s by Singh, Thornton and Westheimer.[124] The application of photoaffinity labelling is broad: identifying potential drug targets, tracking signal transduction and clarifying transport processes.[125,126] Photoaffinity labelling can also be combined with photocaging chemistry to allow optical control of receptors and signalling.[127] There are a number of excellent reviews discussing the utility, application and challenges of photoaffinity labelling.[115,116,127–131]

In order for a photoreactive group to be successfully incorporated into the probe and used in subsequent experiments, it must be stable under ambient light conditions and normal biological conditions (pH, temperature etc). The activation of the group must only occur under specific UV wavelengths, to allow experimental control of the

formation of the protein-ligand complex by exposure to a UV light source. With this in mind, a set of criteria that a photoreactive group should meet have been devised to ensure labelling is as specific and efficient as possible.[127,132] Briefly, the moiety should:

- Avoid altering binding characteristics of the pharmacophore, the photoreactive group should be as small as possible.
- Be relatively stable under ambient light conditions and in the dark to avoid impractical handling and ensure covalent bonding formation only occurs under at the desired wavelength of activation.
- Have a wavelength of activation suitable to minimise photolytic damage to biological material of interest (greater than 300 nm).[133–135]
- Insert into C-H bonds and X-H bonds, giving a stable protein-ligand adduct, thus ensuring full coverage, through a non-specific covalent bond forming event.
- Should have a very short half-life, in its activated form in order to minimise the risk of non-specific complex formation. This activated form should also not be prone to intramolecular rearrangements, to avoid diffusion of the photoaffinity ligand away from the interaction of interest.
- Form a stable covalently bonded adduct that can withstand downstream denaturation, processing and rigorous analysis.

With that in mind, there are a number of potential photoreactive groups that could be incorporated into an affinity probe. Intrinsically photoreactive groups may already exist within pharmacophores, such as sulphides, enones, dienones and halogens that can form radical entities under UV irradiation. However these moieties are not considered biorthogonal as they are also prone to other chemical reactions, such as nucleophilic



azido **diazirine** **benzophenone**

Figure 14: The three most frequently used photoreactive groups: aryl azide, diazirine and benzophenone moieties.

attack.[136] The three most frequently utilised chemical functionalities (**Figure 14**) are the **azido** probes,[137] **diazirine** probes[128] and **benzophenone** probes.[138]

The first reported use of aryl azides in a photoaffinity probe was in 1969.[137] Since then, they have remained a popular choice of photoreactive group due to their relatively small size and ease of synthesis. Upon irradiation with UV light, nitrogen is liberated leaving a highly reactive singlet nitrene species with a life time of approximately 0.1 ms.[139,140] This reactive intermediate can insert into proximal bonds to generate a covalent adduct, however there are also a vast number of possible rearrangements and side reactions to be aware of when utilising aryl azide photochemistry (**Figure 15**).[115]

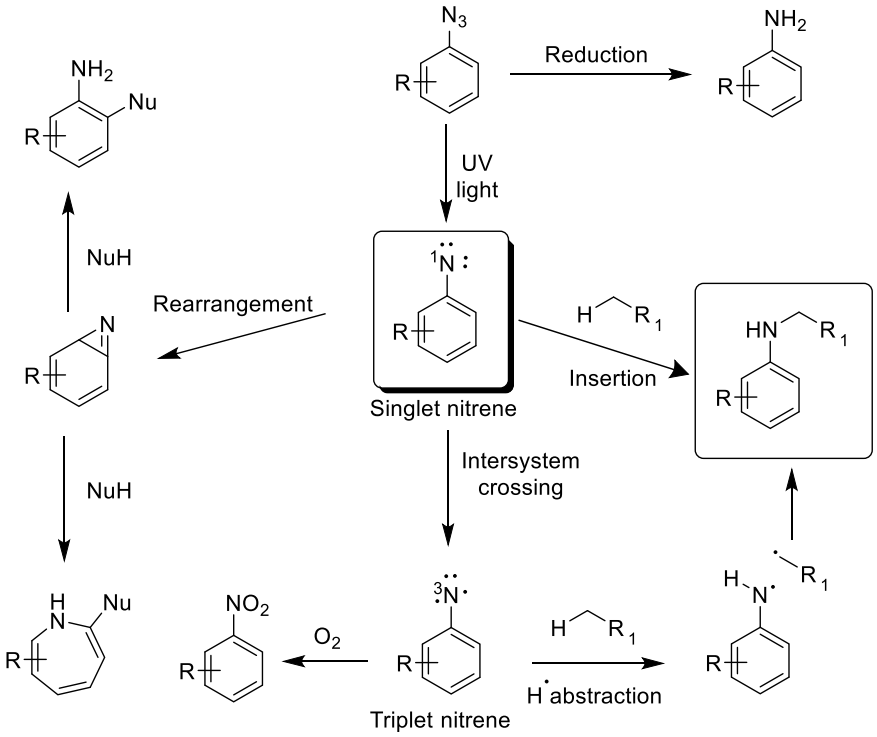


Figure 15: Array of possible rearrangements and side reactions possible after generating a nitrene from an aryl azide. Scheme adapted from Geurink et al. 2012.

The multiple fates of the nitrene mean that labelling efficiency will be limited, and unwanted side products may be detected as a result of rearrangement to long lived electrophilic species. It has also been demonstrated that environmental conditions affect the labelling yield, through both the reduction of the aryl azide to the aniline[141] and the oxidation of the triplet nitrene species to the corresponding nitro species.[142] However, one of the greater limitations of the aryl azide is the requirement to use short

wavelengths of light (less than 300 nm) to unmask the nitrene, which may lead to degradation of biological material.[135] It has also been suggested that nitrenes are less reactive than carbenes.[130]

Carbenes can be generated from diazirine species, first used in 1973.[143] They have been reported to offer a number of advantages over aryl azides: the maximum absorption wavelength is in the range 350-380 nm, they have been shown to be stable under a number of reaction conditions including oxidation, reduction, acidic and basic conditions and they are relatively small.[116] Similar to nitrenes, the chemistry of the carbene can be complex (**Figure 16**).[115]

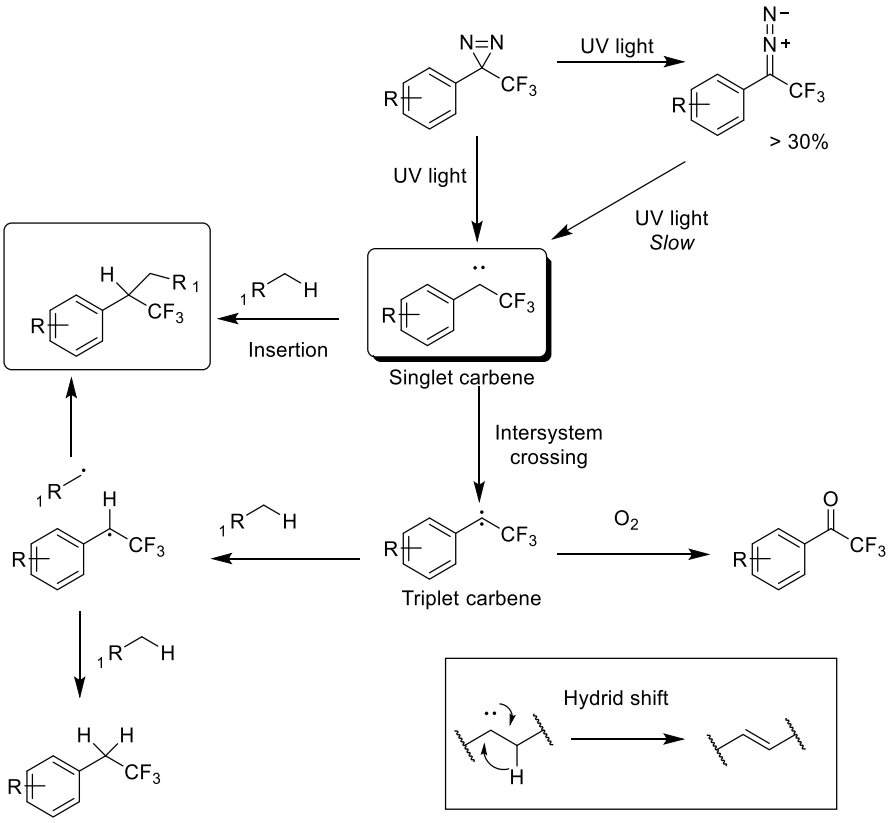


Figure 16: Potential fate of the carbene generated from UV activation of a diazirine. Inset is the possible hydride shift that can occur in unsubstituted alkyl diazirines. Scheme adapted from Geurink et al., 2012.

UV illumination at 360 nm affords the desired carbene with liberation of nitrogen however it also gives the diazo species as the product of a rearrangement in a considerable yield (>30%). This diazo is also capable of liberating nitrogen to generate the same carbene intermediate. However, at 360 nm this transformation occurs at a slow rate, allowing ample time for the probe to diffuse away from the site of interest.

To minimise the effect of this unwarranted UV product, Brunner *et al*. showed that an alpha trifluoromethyl group provided the electron withdrawing effect to stabilise the diazoisomer, limiting its photolabeling abilities.[144] The half-life of the singlet carbene is extremely short (1 ns) and will readily insert into a bond (acting as an electrophile, nucleophile or ambiphile) or undergo intersystem crossing to the corresponding triplet carbene that can also react with proximal bonds (via a diradical mechanism). Carbenes exhibit little preference for the bonds into which they insert, meaning that water and solvent often quench any carbene not interacting with a protein.[145] Unlike aryl azides, the synthesis of diazirine containing compounds is challenging and often be the bottle neck of a probe synthesis. A recent review gives an overview of the challenges and advances in the field.[146] There are also many examples of aliphatic diazirine containing probes.[147]

Benzophenones present a different class of photoreactive group, as they react purely via a diradical mechanism.[148] Like diazirines, the excitation wavelength (in the region of 350-360 nm) is favourable for biological systems, with respect to the limitation of damage to DNA and protein. Unlike the photoactiviation of azido or diazirine moieties, which lead to the elimination of nitrogen, benzophenone activation is reversible meaning unreacted diradical (with a relatively long lifetime of 120 μs) will fall back to the benzophenone. The rate of subsequent reaction is then governed by the type of bond the diradical is reacting with in the initial hydrogen abstraction (favouring the generation of stabilised radicals).[138] The chemistry of the photoactivated species is comparatively less complex than that of a carbene or nitrene. There is also the added advantage that any side reactions occur often fall back to the starting material (**Figure 17**).[115]

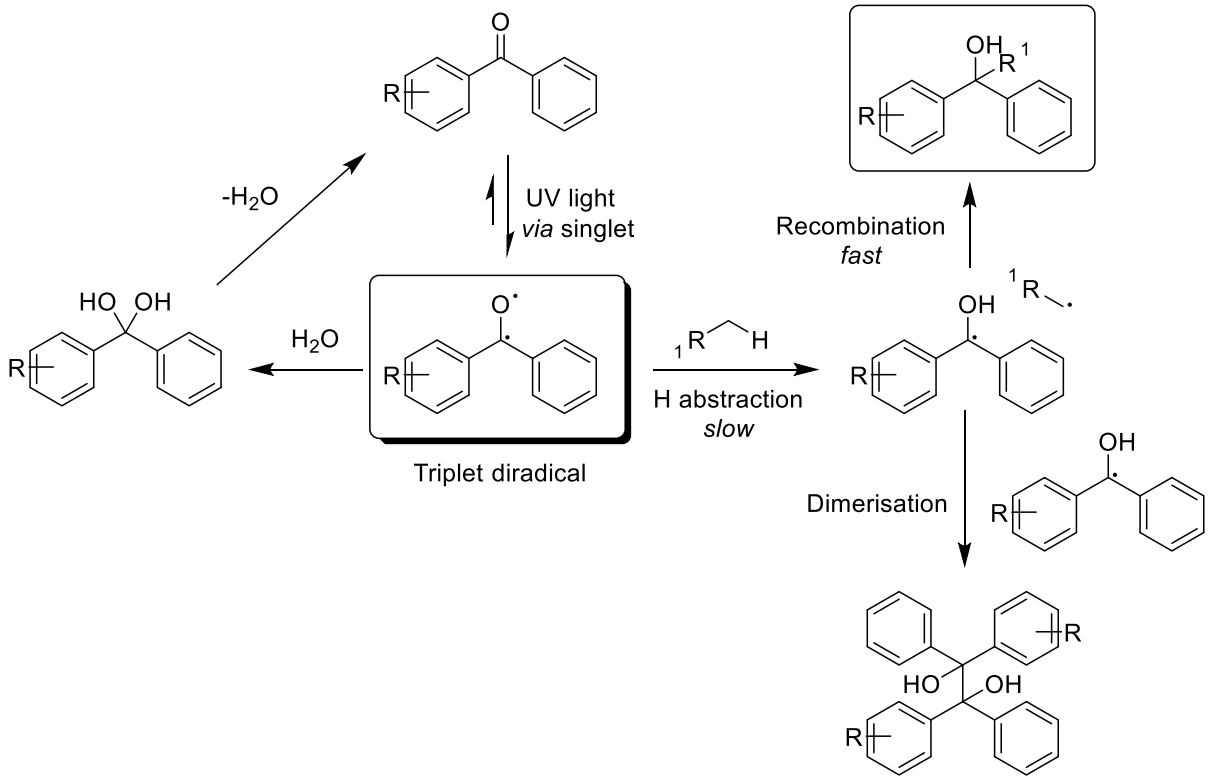


Figure 17: Chemistry of a benzophenone photoreactive group. Unreacted diradical will eventually fall back to the starting material. Scheme adapted from Geurink et al., 2012.[115]

Unlike aryl azides and diazirines, the starting materials for the synthesis of benzophenone are readily commercially available making the synthetic chemistry more facile and versatile. However, the large size of these groups makes their incorporation into a compound potentially less well tolerated. Unless there is already a similar moiety present in the structure, the addition will vastly alter the properties of the molecule, for example by increasing the steric hinderance and thus favouring certain reaction sites over others or by greatly reducing the overall binding affinity of the probe with respect to the parent molecule. However they can be of use for targeting hydrophobic domains or large targets.[149] Due to the reversible nature of the diradical, long irradiation times are required (often over 30 minutes), which increases the probability of non-specific binding and denaturation of biological samples.[150]

1.4.3. Bio-orthogonal Ligation

To detect and analyse the covalently bound probe-protein complex, “reporter” groups must be appended to the probe structure. However, it is often not practical or reliable to synthesise the chemical probe with these groups in place before protein incubation,

due to the potential disruption of the protein-ligand interactions of interest. Instead, chemical probes can be synthesised containing a bio-orthogonal ligation group that can be later used to append the desired reporter group. Much like the photoreactive groups, there are a few guidelines for the selection of the ideal bio-orthogonal ligation handle:

- The group should be small enough to limit any perturbations in binding,
- The chemistry of the ligation should be orthogonal; not one that is found in biological systems,
- The reaction should be highly selective and reactive to produce the ligated product in high yield with a vast array of potential substrates,
- The chemistry should be compatible with biological systems such as the aqueous physiological conditions of the cell.[151]

Given these criteria, there are a limited number of substrates and reactions suitable. The azide is a useful functional group as it satisfies many of the criteria listed above; it is small, uncharged, absent within biological systems and reacts in several different ways.[151,152] The Staudinger Ligation[153] involves the reaction of an azide with phosphorus compounds and is driven by formation of a strong phosphorus-oxygen bond, as well as the loss of nitrogen from the aza-ylide intermediate.[154–156] The reaction was developed further by the Bertozzi group to optimise reagents and conditions for bio-orthogonal ligation in an aqueous environment.[157]

Similarly, in conjunction with another unsaturated component such as an alkyne, azides can undergo a 1,3-dipolar cycloaddition to form a 5-membered 1,2,3-triazole in a chemoselective reaction.[158,159] Optimisation of this reaction simultaneously by Sharpless and Meldal led to the important development of the copper catalysed azide alkyne cycloaddition (CuAAC); now one of the key tools used in chemical proteomics.[160–164] The use of a Cu(I) catalyst negates the need for heat to drive the reaction; improving the biological compatibility and the regioselectivity of the CuAAC reaction.[160,164] The reaction is substantially quicker (up to 20 times the rate) than the Staudinger Ligation[123], remains efficient when water is a solvent, tolerates a wide pH range and exhibits no functional group bias.[160] The reaction mechanism is hypothesised to occur in a stepwise manner with the formation of a copper complexed

intermediate, rather an a concerted 1,3-dipolar addition (**Figure 18**).[165] The reaction can be considered an example of click chemistry, whereby the reaction is "spring loaded" to produce the desired product in a high yield with a diverse range of possible substrates, without requiring purification of any intermediates or the formation of undesired side products.[166]

One of the disadvantages with copper catalysed click chemistry is the potential toxicity of Cu(I) to cells that may lead to changes in protein expression and function.[123,152,167] The use of Cu(I) can be avoided if another rate accelerating moiety can be introduced: Bertozzi *et al*. describe the use of strained alkynes to drive the click reaction.[167] Using cyclooctyne, the smallest cyclic cycloalkyne, which possess a distorted alkyne bond angle (180° to 163°).[168] This strain leads to the destabilisation of the ground state and

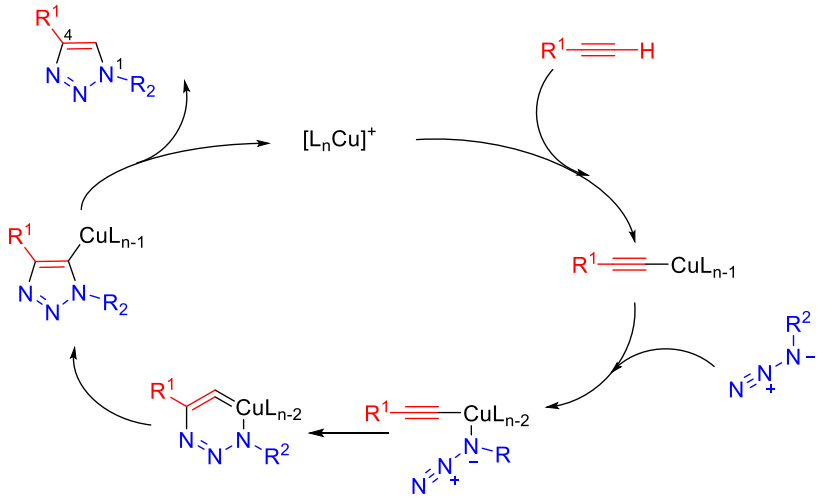


Figure 18: Proposed stepwise and regioselective mechanism for the formation of 1,4-disubstituted-1,2,3-triazoles.

hence a much more reactive towards azide species.[169] This can enable [3+2] cycloaddition click chemistry to occur *in-vivo* and in real time studies.

1.4.4. Reporter Tags

Another advantage of the use of bio-orthogonal ligations with chemical probes is the range of functionalities available to be "clicked" on, by attachment to an azide/alkyne species, depending on the desired outcome of the probe investigation. For example, when enrichment of protein is desired, an affinity tag can be used. Alternatively, for detection of proteins, for example on a SDS-PAGE gel or western blot, a range of fluorescent, visible or radioactive tags can be used. Affinity and detection elements can also be combined into multifunctional capture reagents.

One of the most common affinity tags utilised in chemical biology is biotin, due to its ability to form very strong non-covalent bonds with avidin and streptavidin.[170] Streptavidin can be immobilised onto beads or plates to enable the capture and purification of the proteins of interest.[171] Visualisation of probe labelled proteins can be achieved with fluorophores. There are many fluorophores available which offer orthogonal excitation and emission wavelength profiles; this allows the multimodal imaging and tracking of processes. The most frequently used fluorophore classes offer various advantages and disadvantages to be considered upon selection:

- Fluorescein[172] fluorophores offer high absorptivity, good quantum yield and are water soluble. However, they are prone to fast photobleaching, broad emission spectra and lack cell permeability as well as being sensitive to pH below 7.
- Rhodamine[172] fluorophores are more photostable than fluorescein and are readily excitable. However, they only have a quarter of the quantum yield of fluorescein and are still found lacking in terms of cell permeability.
- BODIPY[173] dyes offer high photostability, extinction coefficient and quantum yield as well as a narrow emission band. Additionally, they are insensitive to pH and are readily cell permeable. However, they are prohibitively expensive, in the quantities required for many synthetic chemistry routes.[174]

Other classes of fluorophores also exist such as Cy-dyes[175] and the proprietary Alexa Fluor dyes. These fluorophores offer superior qualities to dyes within each fluorophore class and are commercially available across the electromagnetic spectrum, albeit at a much greater price.

1.4.5. Chemical Probes for GPCRs

The first bifunctional photocrosslinking probe for a GPCR was reported by Westheimer in the 1960s to illustrate the potential for mapping binding sites of GPCRs.[176] Since then, there has been a plethora of chemical probes targeting GPCR binding sites using endogenous ligands, synthetic small molecules and synthetic peptides.[177] For example, 11-*cis*-retinal, modified to contain a diazirine containing photoactivatable group, has been used to show the binding orientation to opsin.[178,179] Rhodopsin is generally found in high abundance within the endogenous cells and thus presented a

less complex system to work with. Within the GPCR family, there are a number of receptors that contain endogenously photoactivatable ligands such as dopamine[180] and isoprenaline,[180] an a β-adrenergic receptor agonist.

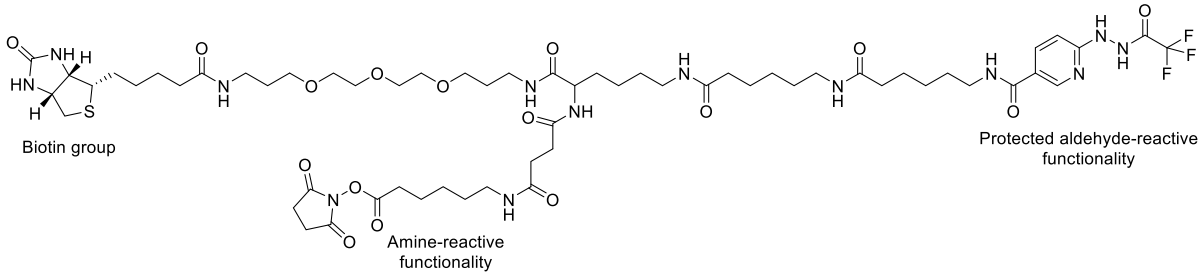


Figure 19: Structure of the TRICEPs universal probe as reported by Frei *et al.* containing an amine reactive NHS-ester, glycosylation capture group and a biotin group for affinity.

One method aiming to explore and identify GPCRs, by modifying ligands of interest with a universal probe called TRICEPS (**Figure 19**). This multifunctional probe contained a NHS-ester that coupled to the ligand of interest; this produced a chemical probe that was able to capture glycosylated membrane proteins.[181] This allows the complex to be enriched, purified and identified and has resulted in a number of successful results.[182–184] However, a limitation of this method is the requirement for the membrane proteins to undergo *N*-glycosylation. Muskens *et al*. describe a novel trifunctional linker that retains the NHS-ester component for ligand modification but utilises photoaffinity labelling to capture receptors of interest; thus, removing any bias or modification requirements. They successfully showed the capture of the tachykinin neurokinin 1 receptor by its ligand Substance P.[29]

Examples of chemical probes for GPCRs that utilise the affinity-based protein profiling two-step methodology are much more limited. One of the first examples in the literature, published by Gregory *et al.,* involved the synthesis of $mGlu_5$ allosteric ligand derivatives containing clickable and photoactivatable groups. They successfully demonstrated the chemical probes retained tight binding affinity with the receptor as well as maintaining biological allosteric behaviour. Using purified receptor, the authors were able to show binding site competition, indicating specificity of the chemical probe, after the preincubation of target protein with an alternative photoprobe. However, within this report photoaffinity labelling in HEK293 cells, that overexpressed the receptor, led to high non-specific binding and was not pursued further.[185]

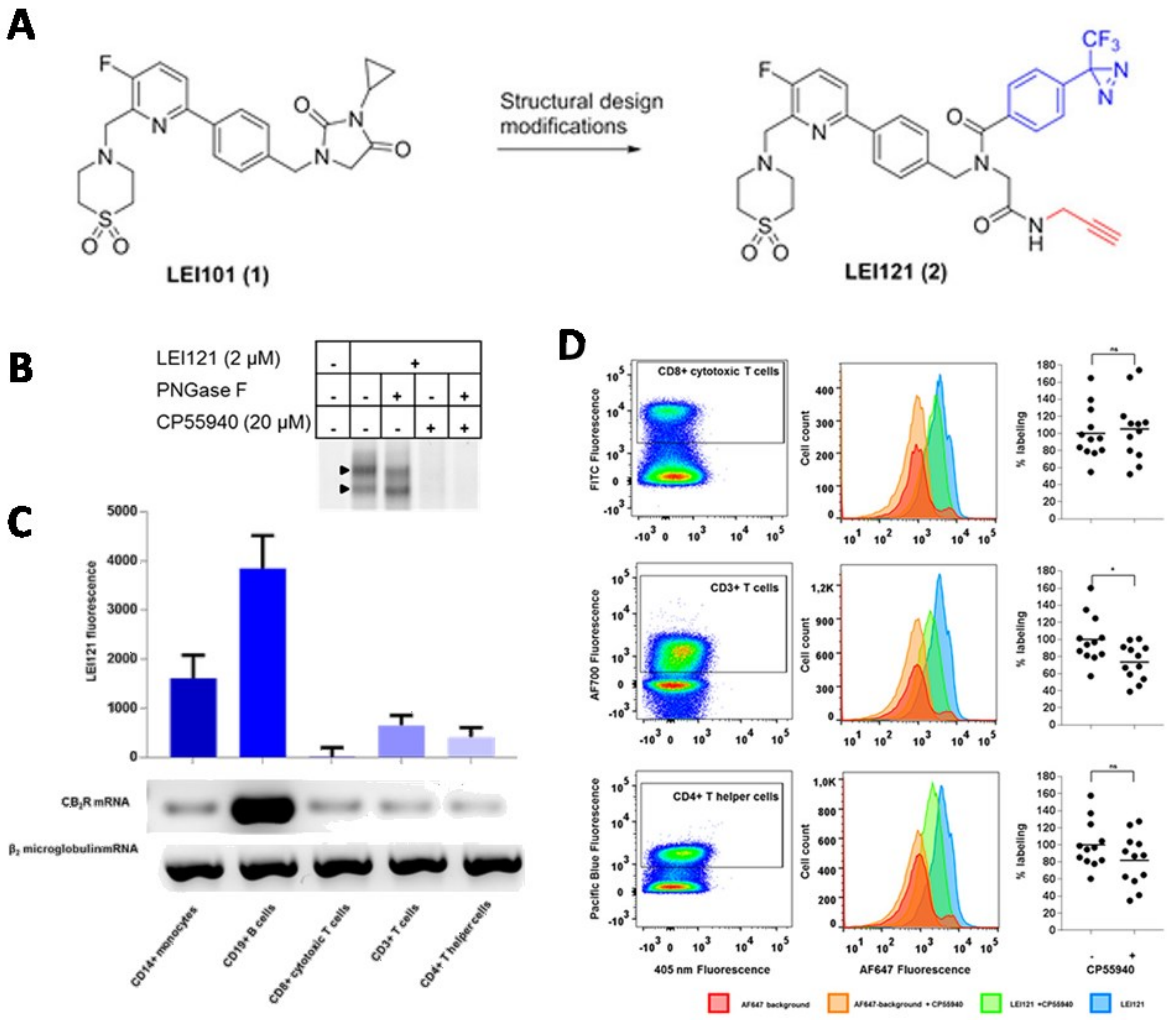


Figure 20: **A** Derivatization of CB2 agonist LEI101 to photoaffinity probe LEI121 containing a diazirine and alkyne group. **B** Visualisation of CB2 using in-gel fluorescence in CHO-CB2 overexpressing cells and competition of the bands with another agonist. **C** Quantification of endogenous levels of CB2 in several human primary cells. **D** use of FACS as a method for quantifying CB2 levels in primary cells. Adapted from Soethoubt *et al.*

More recently, a tandem chemical probe for cannabinoid receptor 2 (CB2) has been reported by Soethoudt *et al.* (**Figure 20**).[186] This work demonstrates the first example of a tandem photoaffinity clickable chemical probe capable of visualisation and demonstration of target engagement at the endogenous GPCR level, in primary human cells. Using the chemical probe, modified with a diazirine and alkyne handle, the authors visualised CB2 isolated from overexpressing membranes using in-gel fluorescence and mass spectrometry (MS). Fluorescence-activated cell sorting (FACS) was used to identify a selection of primary cells that endogenously express the receptor, using labelling with CB2.[186] The combination of tandem photoaffinity probes and FACS presents a powerful new approach to identifying GPCRs even at low endogenous expression levels.

1.4.6. Evaluating Protein Enrichment Using Proteomics

The techniques described thus far allow the identification of probe-protein products, using qualitative techniques such as in-gel fluorescence and western blot that don't give certain protein identification. Incorporating MS into the workflow can give more accurate protein identification and allow global *de novo* protein analysis. The use of soft ionisation techniques such as electrospray ionisation (ESI) enables the analysis of complex mixtures biomolecules such as proteins from cell lysate, without requiring the destructive fragmentation of the sample.[187,188] Tandem mass spectrometry (MS/MS) is a powerful technique that can be applied to biomolecules. MS/MS allows the collection of richer data sets through subsequent rounds of fragmentation and mass analysis of the polypeptides, leading to faster and more accurate protein identifications.[187]

There are multiple approaches to the analysis of a complex mixture of proteins: top-down, bottom-up and shotgun analysis. Top-down proteomics involves an initial separation and identification of labelled proteins using in-gel fluorescence or other visualisation techniques.[189] The protein of interest is extracted from the gel and analysed by MS. This method is useful for identifying a band of interest but restricts analysis to high abundance proteins, and those without extreme isoelectric points or molecular weights, required for good visualisation and separation on SDS-PAGE gel.[189–192] Whole proteome analysis using this technique would be slow and would neglect large groups of proteins such as hydrophobic membrane proteins, of particular interest to this thesis.[173]

Bottom-up proteomics also involves an initial separation of proteins using gel-based techniques. The resulting protein bands are extracted and subjected to a protease digest, producing a mixture of peptide fragments that can be analysed by MS. The results are then analysed against available databases of sequenced proteins to correlate the peptide mixture to the parent protein by matching peptide fingerprints.[191]

Shotgun proteomics is a bottom-up analysis workflow without any initial separation of the protein mixture. The complex protein mixture is digested, and the peptide pool is separated by liquid chromatography in line with mass analysis and tandem MS/MS.

The peptide fragments produced can then be correlated to the parent peptides and proteins using databases.[191,193,194] Compared to the previous workflows, shotgun proteomics is fast and comparatively high throughput in nature. It also removes any bias or loss of interactions that pre-separation may give, as well as producing a global snapshot of the proteome. However, hydrophobic membrane bound proteins are still disadvantaged during trypsin digestion and purification by liquid chromatography.

There have been attempts to address the bias against membrane proteins that conventional proteomic workflows have.[195–200] However, the major challenge is the recovery efficiency of the hydrophobic peptides generated upon digestion. Inclusion of co-solvents or detergents to increase their recovery can lead to incompatibilities with conventional mass spectrometry columns, solvents and detection.[196] There are a number of processes in the proteomic workflow that should be considered for optimisation for membrane protein analysis: enrichment, solubilisation, digestion, fractionation and removal of MS-incompatible reagents.[196] The use of chemical probes for GPCRs increases the enrichment of the protein and thus by optimising the other parameters, the chances of identifying a protein of interest increase.

1.5. Thesis Aims

Drug discovery for GPCRs is an enormous sector within the pharmaceutical industry and GPCRs are known to be integral within many processes in health and disease. Yet, the biology of GPCRs is complex and the tools available to study GPCRs and aid drug discovery are limited. Often drug discovery campaigns rely on high throughput screening to find hits, without robust structural information about the protein target.

GPR119 is one such GPCR that has been heavily targeted by drug discovery campaigns for many years, but a clinical candidate has yet to be identified. More information about GPR119 and the interactions of drug candidates with the GPCR may improve understanding of the receptor and potential agonists. This thesis is focused on the development of chemical biology tools to identify GPR119 using agonists for the receptor and engineered cell line systems. Increasing the understanding of these systems may lead to more focused drug discovery campaigns and expand the tool kit for GPCRs.

The aim of this Thesis is to discuss:

- The design, synthesis and biological evaluation of photocrosslinking chemical probes for GPR119 based on a series of compounds synthesised by AstraZeneca as part of their GPR119 drug discovery campaign,
- The development of appropriate biological systems to allow the identification and characterisation of GPR119 protein expression in complex protein systems,
- Photocrosslinking experiments to evaluate the chemical probes in appropriate systems to successfully identify and enrich GPR119.

2. Chemical Probes for GPR119

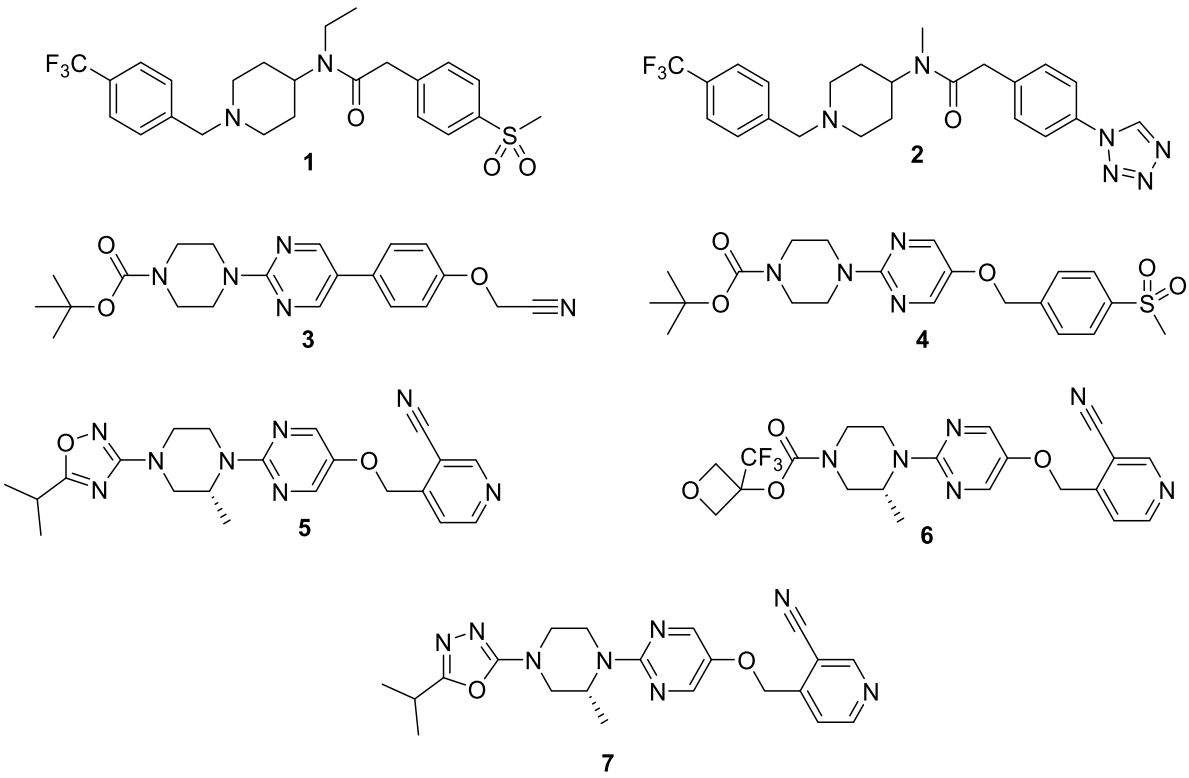
2.1. Introduction

2.1.1. AstraZeneca's Approach to GPR119

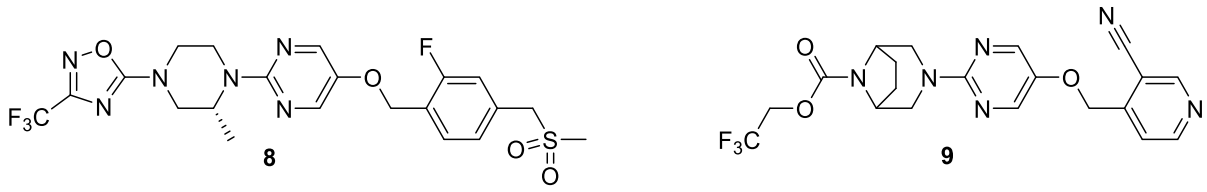
GPR119 has long been an attractive target for pharmaceutical companies. For many years AstraZeneca had an active GPR119 agonist discovery programme, publishing the first paper detailing their approach in 2011.[201] As there is no crystal structure for GPR119, efforts to find a suitable starting point for agonist development were centred on a high throughput screening (HTS) approach using AstraZeneca's compound libraries. The compounds were evaluated against a number of criteria including potency (EC_{50}) and selectivity towards GPR119 and Intrinsic Activity (IA). IA is used as a metric to compare response of an agonist compared to a reference agonist and expressed as a percentage.

From this screen, compound **1** was chosen as a starting point for agonist development, as it showed moderate potency towards GPR119 (EC_{50} = 538 nM) and an IA of 42% (compared to OEA, normalised to 100%). However, limitations in the relative affinity towards the hERG (**h**uman ***E**ther-à-go-go*-**R**elated **G**ene) ion channel (IC_{50} = 2.5 μM) required further development to enhance the selectivity of the compounds for GPR119 over potential off-target receptors.[201]

As the GPR119 drug development campaign progressed, AstraZeneca published multiple papers on their efforts highlighting advances and potential pitfalls of the agonists. Their first paper highlighted the issue of unknown off-target proteins being responsible for the apparent desirable phenotype of their optimised agonist **2** - this will be discussed in more detail below.[201] The follow up paper discussed structural optimisation of **3** to overcome solubility issues, and develop an alternative pharmacophore displaying specific GPR119 agonism. By including rotatable bonds within the pharmacophore, compound **4** was taken forward as the initial lead hit for a novel series of agonists with improved solubility properties.[202] Further optimisation of the structures to improve solubility led to compounds **5, 6** and **7**.[202,203]



Unfortunately, upon testing **5** *in vivo,* tonic-clonic seizures were observed in several mice, highlighting the potentially unacceptable toxicity of these compounds. By evaluating a few of the developed agonists for seizureogenic activity using an *ex vivo* hippocampus brain slice assay, the 3-cyano pyridial moiety was identified as the potential source of convulsions. This was confirmed by lack of seizureogenic activity in matched pairs of agonists lacking that group. Further optimisation led to compound **8** that elicited no convulsions in mice and showed positive on-target blood glucose lowering phenotypes in the mice models.[204]



Interestingly, a series of agonists developed with conformationally restricted piperazine moieties, such as compound **9,** demonstrated bias towards *h*GPR119 rather than *m*GPR119 in terms of intrinsic activity, despite being strong binders.[205] This observation may go some way to explain the lack of efficacy of GPR119 agonists translated from mouse models to human clinical trials and highlights the need to truly understand the pharmacology of the receptor of interest within the model of study.

2.1.1.1. On-target Series

As mentioned above, AstraZeneca developed a series of agonists based around a cluster of compounds (such as **3**) from their HTS against GPR119. The general architecture of these compounds consists of a series of three contiguous rings forming a lipophilic, rigid, linear pharmacophore with poor pharmacokinetic properties. **3** showed good potency (EC_{50} = 78 nM) and IA (81%) however the high lipophilicity (logD = 4.3) led to poor solubility and high plasma protein binding (0.1% free). In order to improve these properties, an extra methylene group and heteroatom was added within the contiguous rings to disrupt the lipophilic structure. This led to compound **4** that showed comparable potency and IA (EC_{50} = 65 nM, IA = 83%) but with a considerably improved logD = 3.2, which was reflected in the pharmacokinetic properties (plasma protein binding was improved to 3.3% free). This improved pharmacophore was retained for further medicinal chemistry optimisations, while R groups were varied.

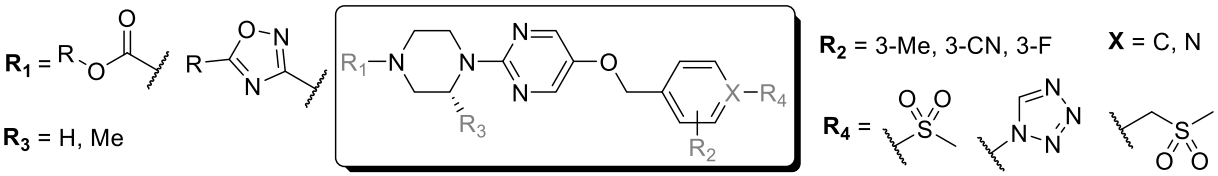


Figure 21: The optimised pharmacophore for the on-target series of GPR119 agonists developed by AstraZeneca.

The addition of a (*R*)-Me group increased potency and IA (EC_{50} = 19 nM, IA = 150%) compared to H in the R_3 position. In the R_2 position, 3-Me improved potency and IA (EC_{50} = 75 nM, 101%), whereas 2-Me resulted in a loss of activity (EC_{50} >1.9 μM, 19%) compared to R_2 = H (EC_{50} = 621 nM, 71%). Similarly, 3-CN (EC_{50} = 21 nM, 155%) and 3-F (EC_{50} = 63 nM, 137%) also showed favourable improvements. The combined effect of R_2 = 3-CN and R_3 = (*R*)-Me led to a highly potent agonist (EC_{50} = 5 nM, IA = 171). In the R_1 position, carbamates generally show favourable properties with potencies in the single digit nanomolar range, however they possess a lack of stability over a large pH range due to the susceptibility of carbamates to hydrolysis in acidic conditions. To increase stability, 5-membered heterocycles were trialled as an isostere of the carbamate group with some success – maintaining similar potency and IA compared to carbamate containing agonists, while showing hydrolytic stability throughout the pH rage 1 – 10. Within the field of GPR119 agonist development, heterocycles have been used as isosteres for carbamates previously.[96,97] At the opposite end of the structure, a fairly large range of different chemical entities are

tolerated in R_4 ranging from a tetrazole through to a sulphone moiety either with or without a linking methylene group.[202–204] The SAR is summarised in **Figure 21**.

Once agonists were optimised for potency, intrinsic activity and pharmacokinetic properties a candidate was taken forward for *in vivo* tolerance tests in mice. In these models, the agonists were evaluated in an oral glucose tolerance test (OGTT) to assess the ability of the compounds to control the levels of glucose within the organisms. In order to verify agonist specificity for GPR119 OGTTs were carried out in wild type (WT) mice and GPR119 knockout (KO) mice. For specific GPR119 agonists, blood glucose lowering would be expected to be observed in the WT mice, but not the KO mice. Indeed, this is what AstraZeneca observed for a number of agonists they profiled. A dose-dependent blood-glucose lowering response was also observed with GPR119 agonists in WT mice. In **Figure 22** are representative graphs demonstrating the *in vivo* blood glucose lowering effects for compound **5** and **8**.

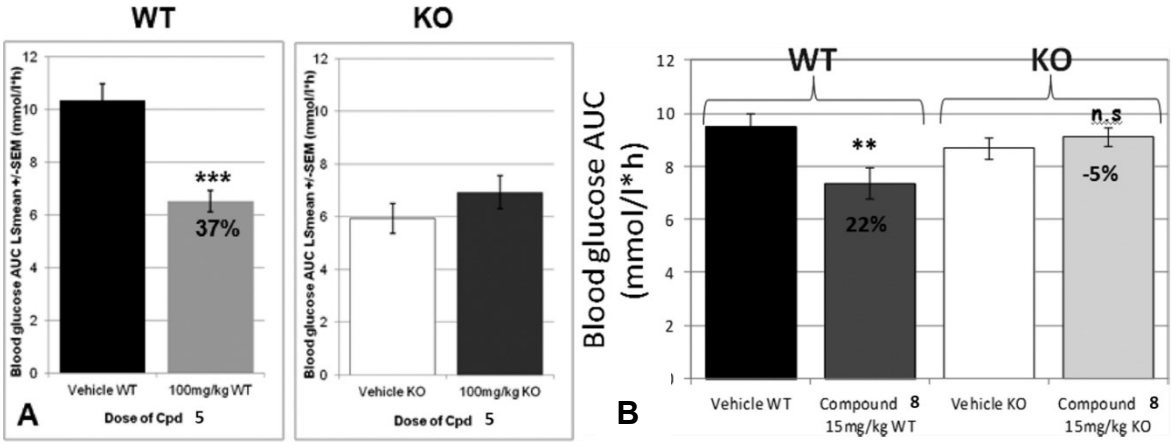
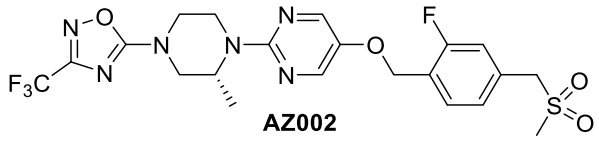


Figure 22: Representative graphs showing the results from *in vivo* OGTT in both WT and GPR119 KO mice demonstrating the on-target agonism of GPR119 with agonists (**A**) **5**[202] and (**B**) **8**[204].

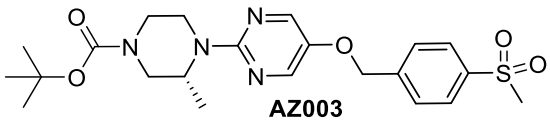
Given these results, two compounds were chosen as GPR119 specific tool compounds for use in this study, considering the described properties and on advice from the scientists responsible for the original GPR119 study. They will be referred to as **AZ002** and **AZ003** within this study, as depicted in **Figure 23**.



AZ002

EC_{50} (hGPR119) = 11 nM

IA = 84%



AZ003

EC_{50} (hGPR119) = 19 nM

IA = 150%

Figure 23: GPR119 specific compounds selected as tools for this study.

2.1.1.2. Off-target Series

The first hit to emerge from the HTS screen against GPR119 was compound **1.** As mentioned previously, the initial favourable potency and activity was compounded by the lack of selectivity over the hERG channel. Initial medicinal chemistry efforts focused around optimising the selectivity of these agonists for GPR119 over hERG by either increasing potency towards GPR119 or decreasing affinity for hERG (**Figure 24**).[201] Previously, AstraZeneca have shown that basic lipophilic compounds are more likely to exhibit hERG activity, so medicinal chemistry efforts to alter both of these properties were conducted.[206]

By replacing the R_4 ethyl group with a truncated methyl group, the metabolic profile and hERG selectivity was improved while maintaining potency for GPR119. The substitution of hydrogen for fluorine on the aromatic ring in the pharmacophore improved the Lipophilic Ligand Efficiency (LLE = pIC50 - cLogP) score as well as increasing the hERG selectivity by an improving the GPR119 EC_{50}. It was also shown that aromaticity of this moiety is essential for activity. Amides were shown to be tolerated at R_1, however the greatest optimisation at this position was the use of an *N*-linked tetrazole, leading to a potent compound (**2**), which has an EC_{50} of 74 nM, IA of 132% and an improved LLE.[201]

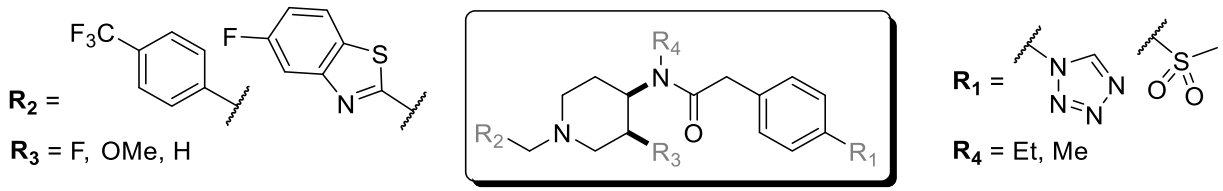
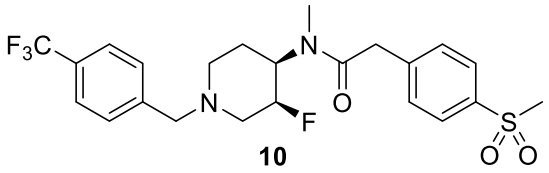


Figure 24: SAR around the "off-target" pharmacophore developed by AstraZeneca.

At R_2, little improvement was found compared to the original hit of an aryl-CF_3 unit. Interestingly, a benzothiazole had a similar profile to the aryl-CF_3. Saturation of the aryl ring led to drastic increase in rates of metabolic turn over (8 times the rate of the unsaturated analogue) as well as a decrease in LLE value with a decrease in potency towards GPR119. The final point of variation was the piperidine core. In order to reduce the basicity of the nitrogen, electron withdrawing groups were substituted onto the ring at the R_3 position in enantiomeric pairs with the aim of reducing the affinity for the hERG channel. Indeed, the (3*S*,4*R*)-F substitution led to a drastic increase in GPR119 potency (EC_{50} = 7 nM) with only moderate hERG affinity (IC_{50} = 2 uM) leading to <250 times selectivity for GPR119 over hERG. Analogues containing -OMe

demonstrated significant metabolic instability compared to the unsubstituted analogue (10x increase in clearance rates). Through a combination of the building blocks of optimisation, compound **10** was the most potent agonist developed (EC_{50} = 4 nM) with an IA of 85%, LLE of 5.7 (an LLE around 5-7 or greater is considered to be a high quality drug-like compound)[207] and decreased hERG affinity (IC_{50} = 4.2 uM) giving a 1000-fold selectivity for GPR119 over the hERG channel. The agonist also showed good levels of metabolic stability.[201]



10

At the start of the lead optimisation campaign, compound **2** was taken into *in vivo* mouse models to assess its capacity to stimulate blood glucose lowering in an OGTT following promising *in vivo* pharmacokinetic profiling. As described above, compound **2** was dosed in both WT mice and GPR119 KO mice, however unlike previously, significant blood glucose lowering was observed in the GPR119 KO mice (**Figure 25**). Even upon reductions in dosage, blood glucose lowering was observed, but at a similar level in both WT and KO mice. This observation is indicative of another target being responsible for the phenotype observed, aside from GPR119. Further compounds from the series were investigated, and it was shown that this off-target phenotype was characteristic of the series rather than a specific artefact of compound **2**.[201]

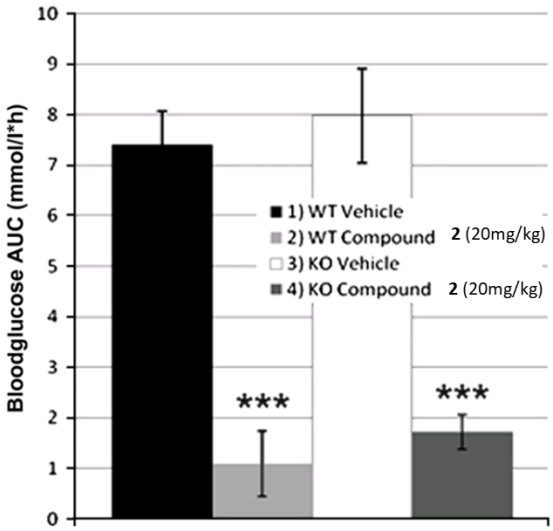
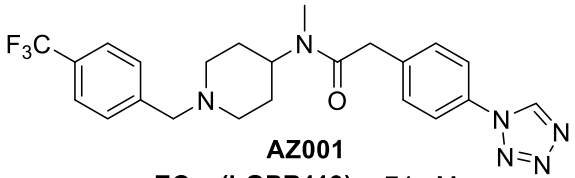


Figure 25: Compound **2** exhibiting off-target blood glucose lowering in GPR119-KO mice.[201]

AstraZeneca conducted a screen against a library of 100 potential targets of interest in order to identify the target(s) responsible for blood glucose lowering. However, amongst those screened, only three showed significant activity with compound **2**. All GPCRs, the receptors identified were 5-HT_{2A} (a serotonin receptor),[208] D4.2 (a dopamine receptor)[209] and CCR5 (a chemokine receptor)[210] none of which have a biological function known to be related to insulin signalling. However, these were eliminated after profiling with other agonists from both the off-target and on-target series of agonists: D4.2 and CCR5 activity was not observed for other off-target agonists, however 5-HT_{2A} activity was maintained at a similar level for both series of agonists implying the target could not be responsible for the response seen in KO mice.[201]



AZ001
EC_{50} (hGPR119) = 74 nM
IA = 132%

Given the extensive testing both *in vitro* and using mouse models, compound **2** was chosen as the tool compound for the off-target series of agonists and will be referred to as **AZ001** throughout this work.

2.2. Design of chemical probes

2.2.1. Justification of the Choice of Chemical Handles

For the development of chemical probes for GPR119 based around the pharmacophores described above, chemical handles were to be introduced to the scaffold in-line with SAR and without altering the properties of the compounds significantly (especially for the pharmacophores exhibiting off-target activity). There are a wide range of chemical handles available and for the purpose of this project photocrosslinking probes containing an affinity handle were desired.

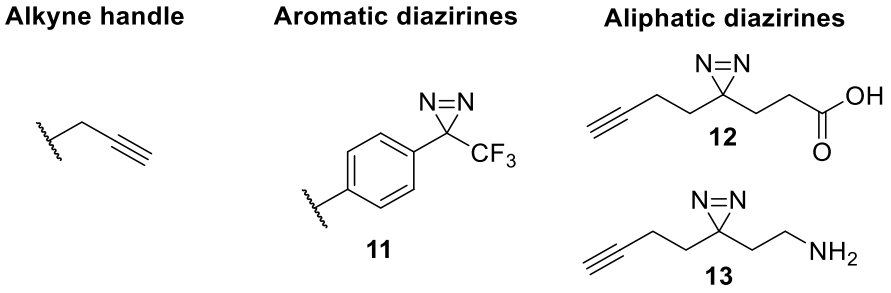


Figure 26: The selection of chemical handles chosen for incorporation into probe structures.

An alkyne handle was chosen as an affinity handle for the chemical probes to allow the attachment of diverse chemical groups via CuAAC, as optimised and heavily utilised within the Tate Group. For the photocrosslinking aspect of the probe, a diazirine was selected as the photoreactive group of choice given its small size, high reactivity and relatively broad tolerance of chemistries and reaction conditions. Given the apparent requirement for aromatic groups within the GPR119 agonist series, it was proposed that the incorporation of an aromatic diazirine, such as **11** would be facile without altering the properties of the compounds too dramatically. The synthesis and considerations of this kind of group have been published and discussed previously.[119,146,211,212] This would also require the attachment of an alkyne group at another point on the structure. For agonists without an easily accessible aromatic ring to modify, it was envisaged to modify the structures by the addition of a minimal photoaffinity crosslinking handle as first described by Shao Yao, examples of which are given in **Figure 26**.[213] This has the advantage of allowing late stage modification of intermediates to form chemical probes, as well as only requiring one point of attachment with diverse available chemistries depending on the linker synthesised.

2.2.2. Derivatisation to Probes

After examination of the SAR around these pharmacophores, optimal positioning of the chemical handles was determined based on both relative ease of synthesis and minimal disruption to existing agonist-receptor interactions. For the off-target series, it was envisioned that the aromatic diazirine **11** could be incorporated as a good surrogate for the trifluoromethyl-benzyl group already present in the structure. While only methyl and ethyl alkyl chains were tested at the amide nitrogen, given the tolerance of both species, it was hypothesised that a propyne group may also be tolerated here and would be synthetically simple to incorporate, leading to potential probe **14**. It was also shown in the original study that amides were tolerated (albeit with a loss of potency towards GPR119)[201] at the R_1 position, which provides another useful point of attachment for the alkyne handle as an extension from the amide nitrogen, leading to probe **15**, as depicted in **Figure 27**.

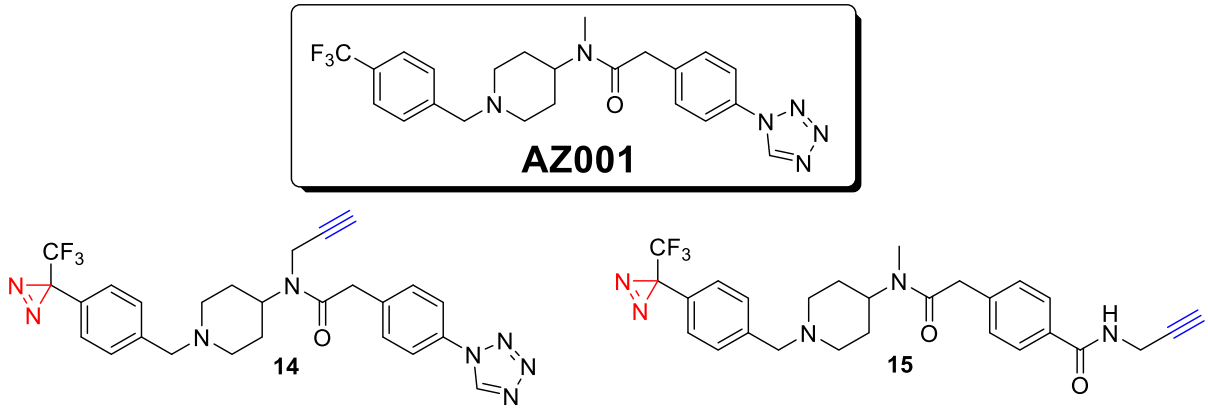


Figure 27: Proposed chemical probes based around the structure of **AZ001** incorporating a diazirine (red) and an alkyne (blue).

For the on-target series, **AZ003** was used as a template structure. Being a planar molecule, there were fewer obvious points of modification compared to **AZ001**. Based on the SAR, there was a moderately large scope for modification at the carbamate however longer alkyl chains had not been rigorously assessed. None-the-less, it was envisioned that synthesis of an amide with a minimal photoaffinity crosslinking handle to form probe **16** would be possible. At the opposite end of the agonist, it had been shown (but not published) that amides were also tolerated off the aromatic ring, and thus attachment of a photoaffinity crosslinking handle here would give probe **17**. Without knowing the preferred binding modes of these agonists, by utilising a matched

set of probes with chemical handles at either end it was anticipated that activity should be retained with one of the probes (**Figure 28**).

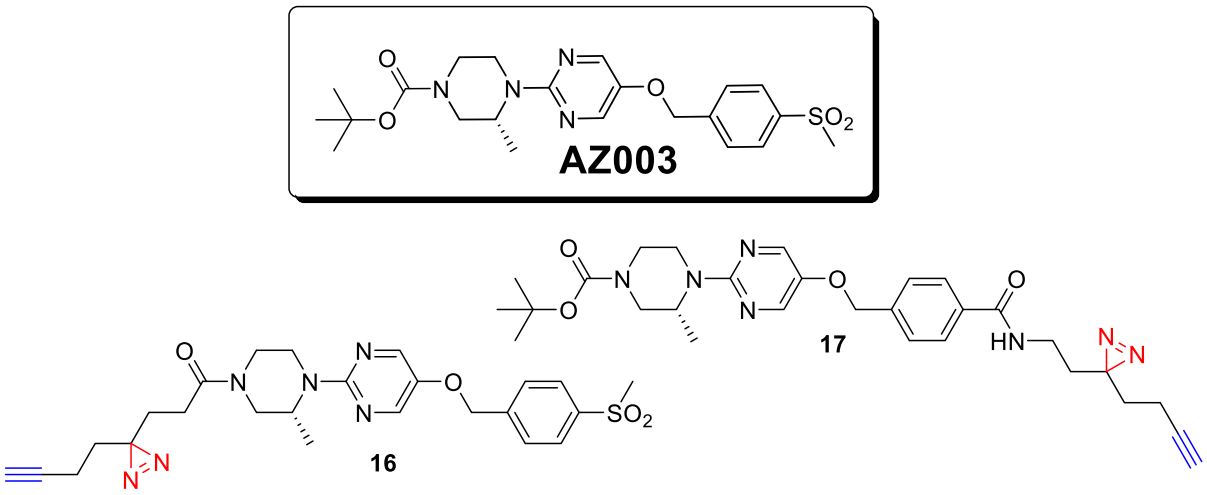


Figure 28: Potential chemical probes for GPR119 based around the structure of **AZ003** incorporating a diazirine (red) and an alkyne (blue).

2.2.3. Proposed Synthetic Routes

In the original synthesis of these compounds, AstraZeneca employed a block-wise synthetic strategy allowing easy diversification of the scaffolds. Following such a strategy, the retrosynthetic approach to the probes is shown in **Figure 29**. The building blocks shown are either reported in the synthetic literature, available from AstraZeneca or commercially available.

For probes **14** and **15** simple disconnection from the final probe gives the bromobenzyl aromatic diazirine **11**[214] as a building block that can be inserted via alkylation of the secondary amine. Further disconnection at the amide bond gives an amine-piperidine moiety **18** and **19** that can be prepared by reductive amination from the respective primary amines and piperidin-4-one and a carboxylic acid moiety. The synthesis of **19** has been reported by AstraZeneca as part of the original medicinal chemistry efforts.[201] For probe **15** disconnection at the second amide bond gives propargylamine and 2-(4-(methoxycarbonyl)phenyl)acetic acid **21**, of which the selective hydrolysis from the double methyl ester precursor has been reported.[215]

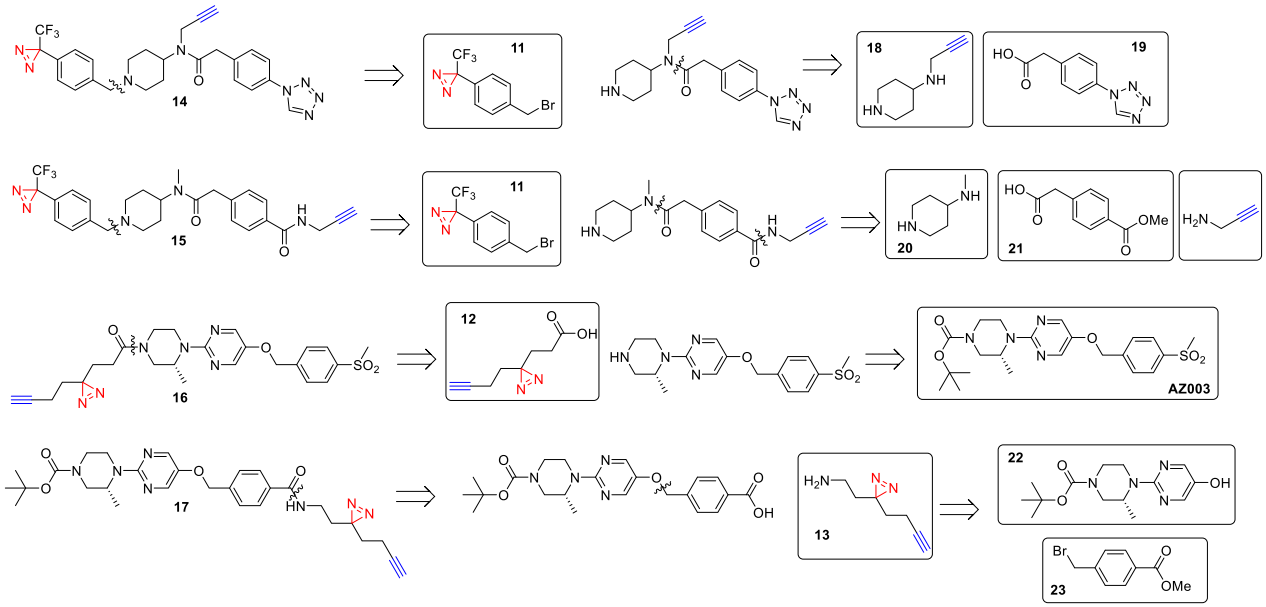


Figure 29: The disconnect approach towards the proposed probes **14**, **15**, **16** and **17** back to synthetically reported or commercially available building blocks.

The retrosynthetic analysis for probe **16** is fairly simple: disconnecting the amide bond to give the photoreactive crosslinking group **12**[213,216] leaves a secondary amine component that is the boc-deprotected **AZ003** parent compound. A similar approach can be taken for probe **17**; disconnection via the amide bond gives photoreactive crosslinking group **13**[213] and a carboxylic acid moiety that can be further disconnected via *O*-alkylation to **22**, an intermediate that was available from AstraZeneca, and **23** that is commercially available.

2.3. Synthesis of a Chemical Probe Library

2.3.1. Synthesis of On-target Chemical Probes

Given the availability of probe precursors for the on-target pharmacophores, initial efforts were based around the synthesis of suitable minimal photoaffinity crosslinking moieties, as outlined in **Figure 30**. Based on the publication by Li *et al.* (2013),[213] starting from ethyl acetoacetate the alkyne handle was introduced using two equivalents of lithium diisoproyl amine (LDA) generated in situ from fresh BuLi and dry diisopropyl amine to form intermediate **24**. Two equivalents of LDA are necessary in this case to selectively alkylate on the terminal carbon as opposed to the more basic central carbon and freshly generated LDA is needed to ensure the reaction goes to completion as chromatographic separation is not possible due to no change in Rf.

Protection of the ketone using acetal chemistry is also required to go to completion for the same reasons. As such, several conditions were tested as outlined in **Table 2**. Acetal formation is a reversible reaction that produces water as a by-product. As the reaction proceeds, more water is formed, and this can drive the equilibria back to the starting materials. In order to manipulate the equilibrium, Le Chatelier's principle states that the removal of water from the reaction mixture should drive the equilibrium towards the desired product. In practice, such acetal formation reactions are frequently carried out in solvents that are known to azeotrope water (such as toluene and benzene) using a Dean-Stark apparatus to capture and measure the volume of water produced to monitor reaction progress.

Ref	Conditions	Yield and Observations
a	Ethylene glycol, cat. TsOH, Toluene, reflux, overnight. Dean-stark apparatus.	13%, major side product? Possible decomposition?
b	Ethylene glycol, cat. TsOH, Toluene, RT, 4 hours. 4 Å molecular sieves.	No reaction
c	Ethylene glycol, cat. TsOH, $CH(OEt)_3$, benzene, 95 °C, 3 hours. Dean-stark apparatus.	6%, hard to separate from starting material
d	Ethylene glycol, cat. TsOH, $CH(OEt)_3$, RT, 36 hours.	77% and no chromatography required

Table 2: Reaction conditions trialled for acetal protection of intermediate **24**.

Initial attempts to recreate dean-stark conditions using toluene at reflux were unsuccessful (**Table 2**, **condition (a)**) and resulted in a poor yield of 13% with a major unidentified side product which could be the result of decomposition due to refluxing at a high temperature for a prolonged amount of time. The reaction was repeated employing molecular sieves as the means to remove water from the reaction mixture (**Table 2**, **condition (b)**) and avoiding heating the reaction mixture to high temperatures, however, this reaction didn't progress and returned the starting materials. **Condition (a)** was repeated employing benzene as a lower boiling point solvent to reduce the possibility of decomposition in **Table 2**, **condition (c)**. After 3 hours, there was no observable side/decomposition product, however there was also a poor yield of the desired product. Upon work up, only 6% isolated yield could be achieved due to the reaction not going to completion. To drive the reaction forward,

Table 2, **condition (d)** utilised ethylene glycol as the solvent, as well as increasing the equivalents of the nucleophile and concentration of the reagents thus forcing the reaction equilibrium towards the desired products. This also allowed the reaction to proceed at room temperature and conversion was observed after 24 hours however leaving for longer ensured full conversation, eliminating the need for chromatography and resulting in an isolated yield after work-up of 77% of intermediate **25**.

Reduction of **25** using $LiAlH_4$ resulted in the conversion of the ester to an alcohol to give **26** and subsequent acetal deprotection using tosyl alcohol in acetone afforded ketone **27**, both in good yield.

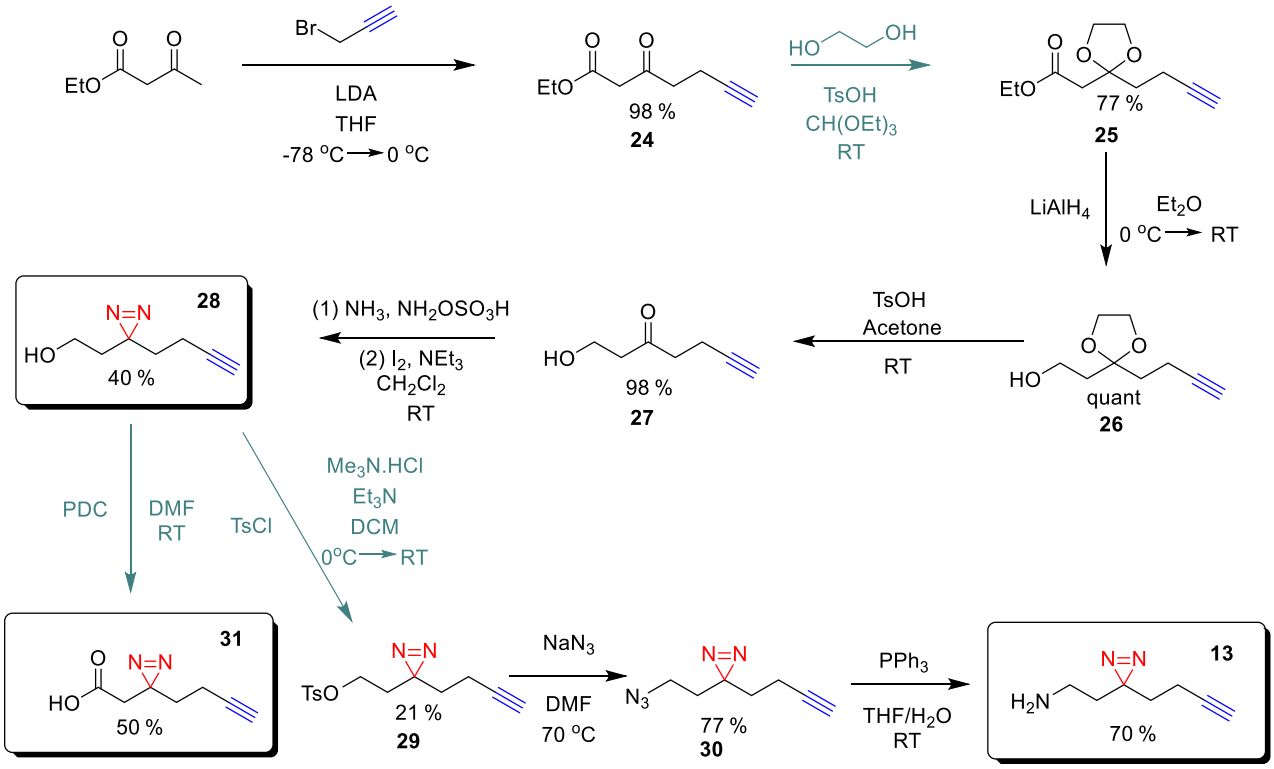


Figure 30: Synthetic routes towards diazirine containing crosslinking linkers **13** and **31**. The synthesis is based on the publish synthesis by Li *et al.* (2013).[213] Deviations from the published protocol are highlighted in teal.

The conversion of the ketone group to the photoreactive diazirine was carried out using a frequently reported method. Briefly, liquid ammonia was condensed into a dry flask with the use of a dry-ice/acetone bath and a dry-ice/acetone cold finger condenser. To this, the ketone material **27** was added and stirred for at least 5 hours at -78 ºC before a solution of hydroxylamine-*O*-sulfonic acid in anhydrous MeOH was added dropwise to form the diazidirine intermediate *in situ*. Oxidation using trimethylamine and iodine afforded diazirine intermediate **28** in a moderate yield of 40% after column chromatography.

Due to the number of synthetic steps and loss of yield along the way, it was decided to attempt oxidation of intermediate **28** directly to carboxylic acid **31** to increase the amount of material available to couple to the pharmacophores. Oxidation using pyridinium dichromate (PDC) indeed afforded a more minimalist linker **31**, one methylene group shorter than that in the original publication. This synthetic strategy has since been published by Kleiner *et al.* (2017).[217]

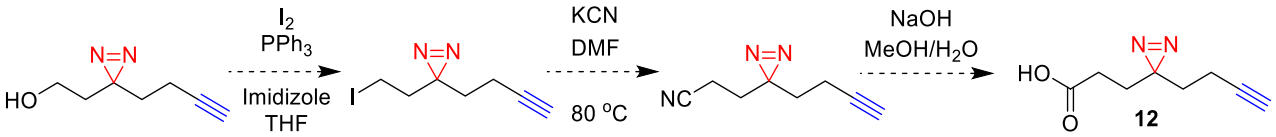


Figure 31: The remaining steps required to form species **12** as published by Li *et al* (2013).[213]

The original 2013 publication required two further steps to achieve the conversion to the carboxylic acid **12** as outlined in **Figure 31** giving a total number of 8 steps to the synthesis. More recently, Parker *et al*. (2017) have published a shorter route of 4 steps to achieve the same diazirine containing crosslinker **12** (**Figure 32**).[216] In order to validate that the shorter linker **31** did not limit the photocrosslinking capacity of the diazirine, the longer linker **12** was also synthesised following the published route.

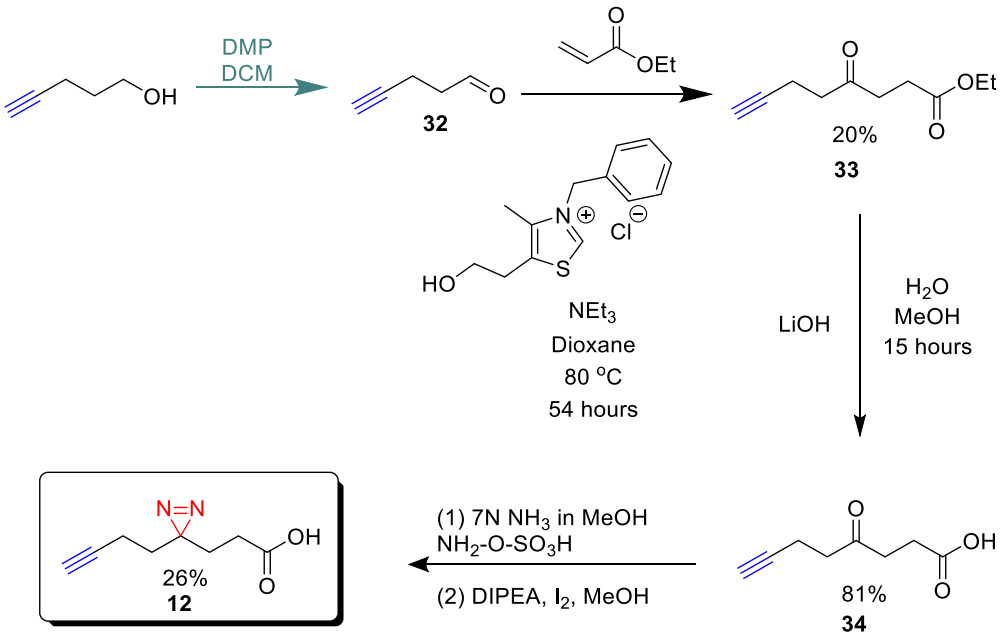


Figure 32: Alternative synthetic route towards carboxylic acid functionalised minimalist linker **12** as published by Parker *et al.* (2017).[216] Deviations from the published route are highlighted in teal.

The route was largely followed as published, however aldehyde **32** was synthesised using Dess-Martin periodinane oxidation, as opposed to swern oxidation chemistry as most commonly reported, due to the milder reaction conditions, shorter reactions times

and relative ease of work up given the limited stability and volatility of the aldehyde product.[218] Used as crude, the aldehyde was coupled to ethyl acrylate by employing Umpolung chemistry to invert the reactivity of the aldehyde. Subsequent ester hydrolysis and diazirine oxidation as described previously afforded diazirine **12**.

In order to introduce amine functionality into the minimalist crosslinker, the alcohol in intermediate **28** was to be converted into a good leaving group. In the original publication, a variant of the Appel reaction was used, employing I_2 as the electrophilic source of iodine. However, the production of triphenylphosphine oxide as a side product is a major organic contaminant, potentially making purification more troublesome and decreasing the potential yield. Given the small quantity of intermediate available, it was decided to go via a tosylate intermediate to give a good shift in Rf and produce a product that was more UV active to facilitate easier clean-up of the reaction mixture. The first conditions attempted using just tosyl chloride in pyridine did not progress and afforded starting material. By using trimethylamine as a nucleophilic catalyst and triethylamine as a base in DCM, conversion as achieved, but only a modest yield of **29** was isolated. Simple nucleophilic substitution of the tosyl leaving group with sodium azide afforded intermediate **30** in a good yield. Subsequent Staudinger reduction with triphenyl phosphine gave the desired amine product **13** ready for further coupling to GPR119 pharmacophores.

As previously mentioned, agonist intermediate **22** and **AZ003** were made available from AstraZeneca for facile assembly of the final probe. After boc deprotection of **AZ003** to **35**, amide coupling conditions using freshly prepared 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM, **36**) with either carboxylic acid linker **31** or **12** afforded the final diazirine containing probes **37** or **16** respectively (**Figure 33**).

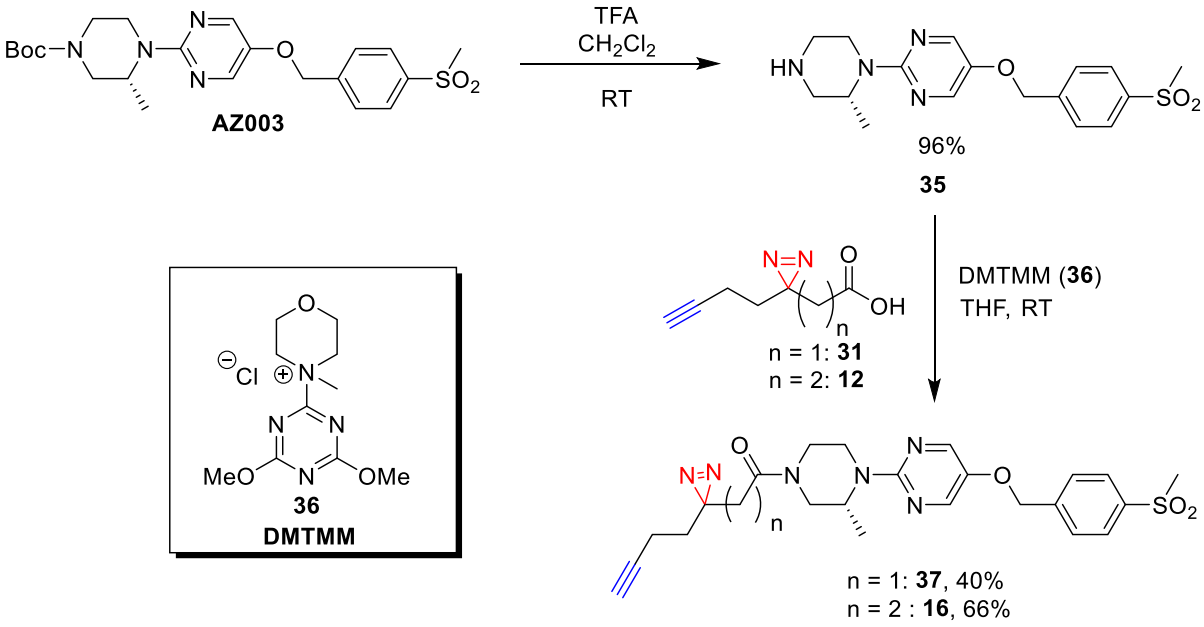


Figure 33: The synthetic route towards probes **37** and **16** from parent compound **AZ003** utilising DMTMM (**36**) as an amide coupling reagent.

For assembly of probe **17**, *O*-alkylation of **22** with commercially available methyl 4-(bromomethyl)benzoate afforded intermediate **38**. The methyl ester was hydrolysed and using the same coupling conditions as before, the amine diazirine linker **13** was appended to give the final probe in a low yield (**Figure 34**). A significant side product was formed during the coupling reaction identified as lacking the amine diazirine chain in the NMR, but with 6 additional protons around ~4 ppm in the ^{1}H NMR that show correlation to a carbon at ~ 55 ppm and a quaternary carbonyl at ~174 ppm in the ^{13}C NMR. A pair of aromatic protons in the benzene ring also show a 0.4 ppm shift downfield. This characterisation is indicative of an alternative coupling occurring at the benzylic acid.

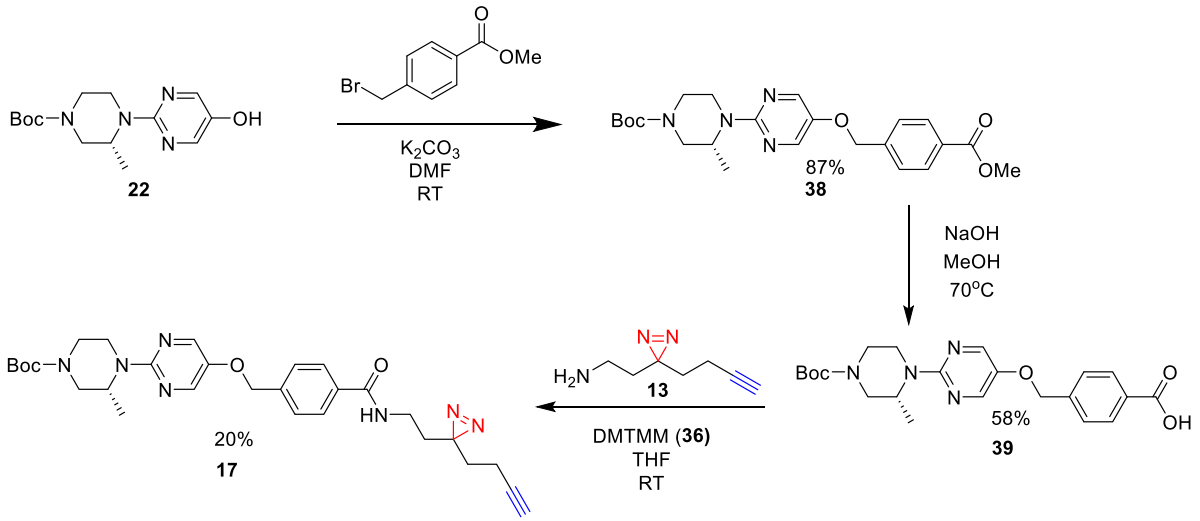


Figure 34: The synthetic route taken to probe **17** from starting material **22**.

2.3.2. Synthesis of off-target chemical probes

Unlike the synthesis of the probes described above, for the chemical probes based around the off-target pharmacophore, the chemical handles were designed to be integrated into the scaffold of the compounds. Given this, the chemical probes were built bottom-up from the modified building blocks required (**Figure 35**).

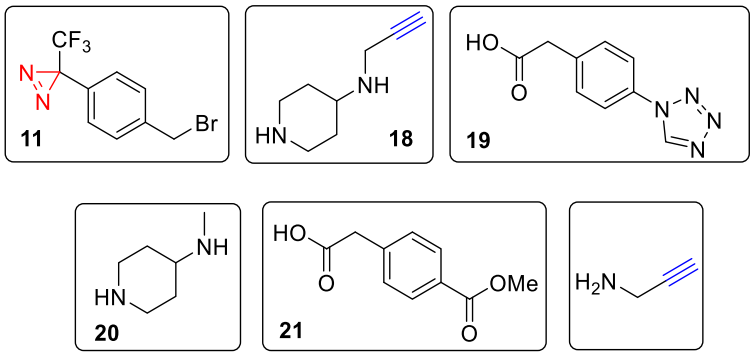


Figure 35: Building blocks required for the synthesis of the off-target chemical probes.

Common to both off-target probes was the diazirine containing aromatic building block, **11**. The synthesis, challenges and optimisations have been extensively described in an earlier report (*Kate Hadavizadeh*, *MRes Report, 2014*).[219] The synthetic scheme is shown in **Figure 36**. Briefly, imine formation with hydroxylamine (**40**) and subsequent tosylation (**41**) provides the precursor to diazidirine formation. Displacement of the tosyl group with ammonia affords the diazidirine three membered ring (**42**) as a stable intermediate in the diazirine forming reaction. Oxidation with iodine and trimethylamine gives the diazirine **43** before final functionalisation with the installation of a benzyl

bromide in a radical reaction employing *N*-bromosuccinimide (NBS) and azobisisobutyronitrile (AIBN) to provide the diazirine containing building block **11**.

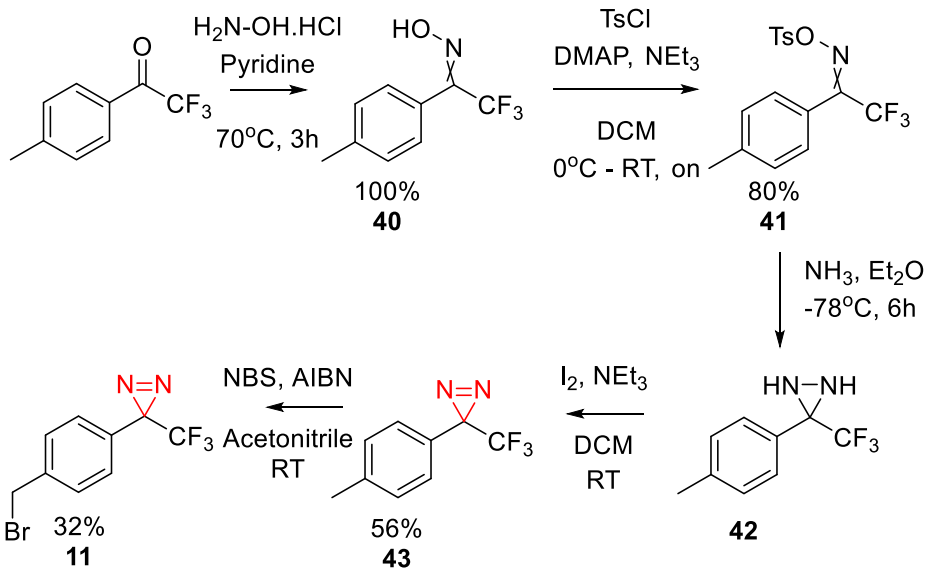


Figure 36: The synthetic route towards diazirine containing building block **11**.

Synthesis of probe **14** has been reported previously (*Kate Hadavizadeh, MRes Thesis, 2014*),[219] however the synthetic scheme is outlined in **Figure 37**. Incorporation of the alkyne handle to give intermediate **18** was achieved by reductive amination of boc-piperidone, while the tetrazole containing fragment **19** was synthesised as previously reported by AstraZeneca.[201] Again, assembly of the building blocks was achieved by DMTMM (**36**) amide coupling and *N*-alkylation.

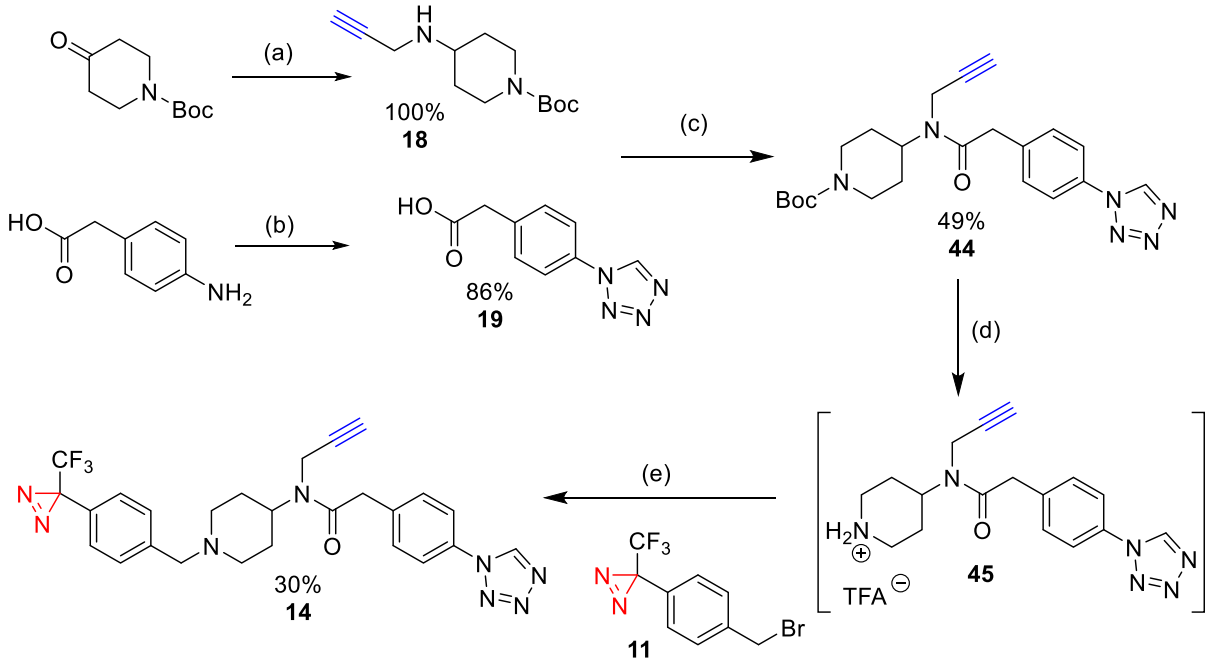


Figure 37: Synthesis of probe **14**. Reagents and conditions: (a) propargyl amine (1.1 eq.), $NaBH(OAc)_3$ (1.5 eq.), CH_2Cl_2, 16h, RT; (b) NaN_3 (1.1 eq.), triethylorthoformate (3 eq.), AcOH (8 eq.), 4h, 80°C; (c) amine **18** (1.1 eq.), acid **19** (1.0 eq.), DMTMM **36** (1.1 eq.), THF, 3h, RT; (d) TFA, CH_2Cl_2, 2h, RT; (e) **11** (1.0 eq.), DIPEA (6.5 eq.), MeCN, 16h, RT. Previously reported in *Kate Hadavizadeh, MRes Thesis, 2014*.[219]

Probe **15** was synthesised with a similar methodology as outlined in **Figure 38**. For this probe, the alkyne handle was placed at the opposite end of the pharmacophore to the diazirine group. Given this, the piperidine building block was synthesised as per the parent compound to contain a methyl amine, again this was installed using reductive amination chemistry to give intermediate **20**. Selective ester hydrolysis afforded **21** enabling amide coupling of the two building blocks to give intermediate **46**. In order to install the alkyne handle, more forceful ester hydrolysis conditions were employed to give intermediate **47** to allow amide coupling chemistry to occur with propargylamine to produce intermediate **48.** Boc deprotection with TFA, and *N*-alkylation with diazirine containing **11** afforded the final probe **15**.

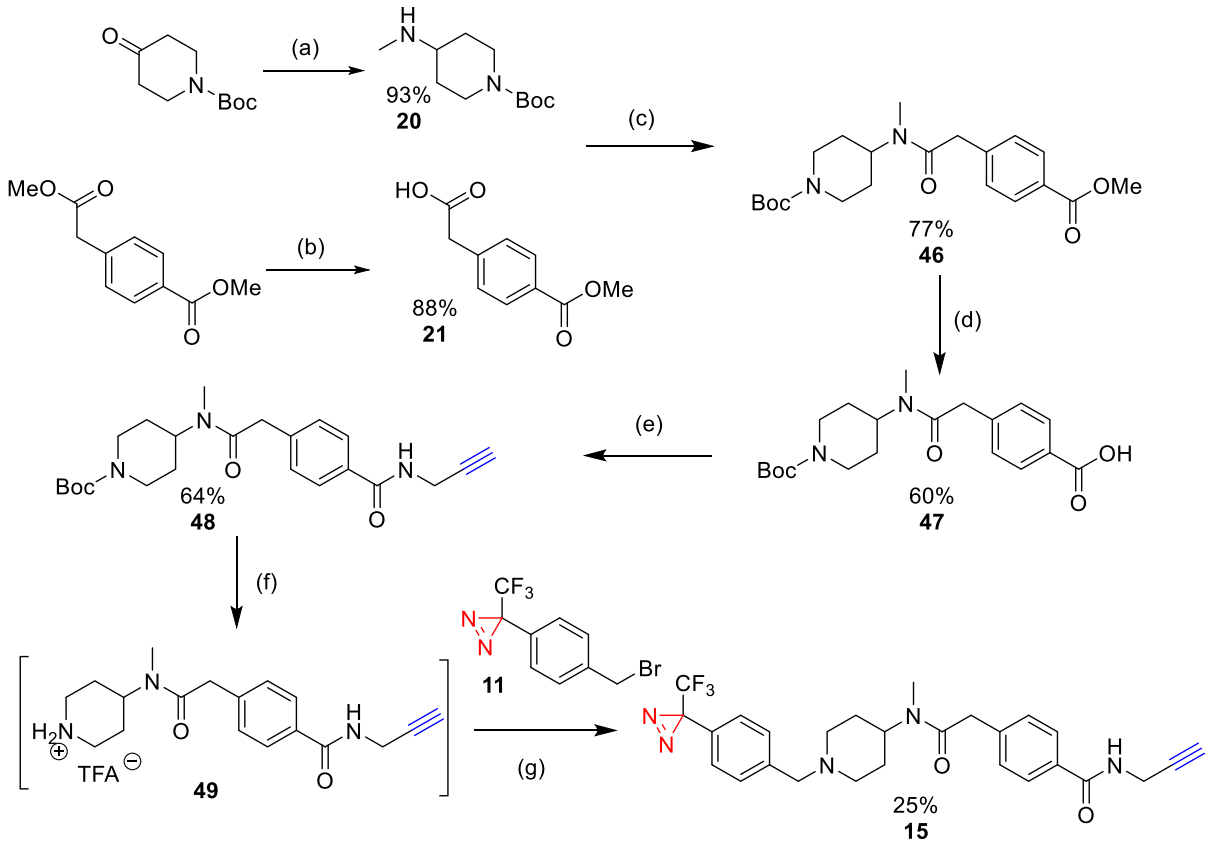


Figure 38: Synthesis of probe **15**. Reagents and conditions: (a) methyl amine hydrochloride (1.1 eq.), $NaBH(OAc)_3$ (1.5 eq.), CH_2Cl_2, 16h, RT; (b) NaOH (1.0 eq.), MeOH, 4h, 50 °C; (c) amine **20** (1.1 eq.), acid **21** (1.0 eq.), DMTMM (**36**) (1.1 eq.), THF, 3h, RT; (d) NaOH (2.0 eq.), MeOH, 16h, 70 °C; (e) propargyl amine (1.1 eq.), DMTMM (**36**) (1.1 eq.), THF, 16h, RT; (f) TFA, CH_2Cl_2, 2h, RT; (g) **11** (1.0 eq.), DIPEA (6.5 eq.), MeCN, 16h, RT.

2.3.3. Additional chemical probes

In addition to the photocrosslinking probes, several control chemical probes lacking the diazirine moiety were synthesised to enable comparisons in activity assays and probe engagement experiments. The synthetic strategy was as described above, but using moieties lacking a diazirine functional group as shown in **Figure 39**.

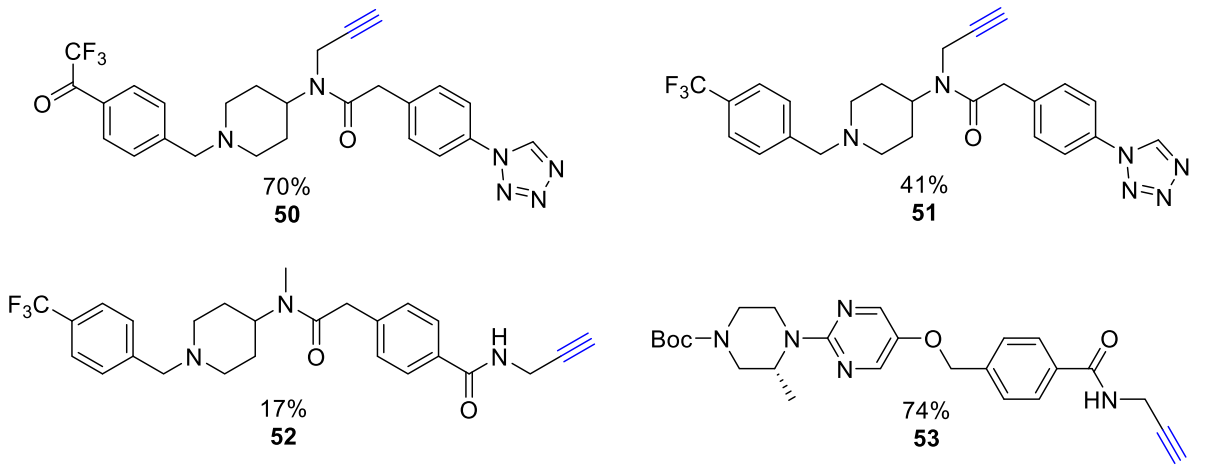


Figure 39: Additional probes synthesised lacking the photocrosslinking moiety.

2.4. Biological evaluation of chemical probes

2.4.1. Commonly used assays for GPCRs

In order to validate the chemical probes against GPR119, a suitable activity assay is required. Within GPCR drug discovery, a number of assays are available to probe GPCR ligand binding interactions. The current available assays can be separated into those that are independent of G-protein binding and those that rely on secondary messenger signalling to assess receptor activity (**Figure 40**).

In 1970, the first radioligand binding assay was described for the adrenocorticotropic hormone receptor using an ^{125}I labelled ligand.[220] Radioligand binding assays have been, and still are, commonly used to determine ligand binding to a receptor of interest by utilising a radiolabelled ligand (either a library of ligands of interest, or a known binder that is competed out) and using radiometric methods to quantify binding. However, by solely assessing ligand-receptor interaction, information regarding the nature of the ligand (agonist, antagonist or inverse agonist) is not available.[221]

An alternative assay format that has potential for a much higher throughput is the Scintillation Proximity Assay (SPA).[222] Briefly, scintillation is an energy transfer phenomenon that occurs between decaying isotopes whereby the particles of ionizing radiation collide with electrons in conjugated systems, such as aromatic systems. Upon decay, the high energy particle excites the electrons within the scintillation material such that upon relaxation to ground state, photons are emitted. These photons can be captured by the inclusion of fluorophores within the scintillation materials resulting in the emission at a wavelength that is captured by the detector.

The proximity element of the assay is achieved by functionalising the scintillation material (beads or plates) to include an affinity element for either the receptor of interest or a radiolabelled reaction product. Thus, when the radiolabelled element of the assay is within proximity, a fluorescent read out can be observed. The advantages of this method over the radioligand binding assay are primarily the ability to scale the reaction down. It has utility in measuring enzyme kinetics[223] and investigating equilibrium dynamics of a receptor-ligand system.[224] However, non-specific proximity effects can cause false positive results (the use of high energy isotopes can exacerbate this problem).[222]

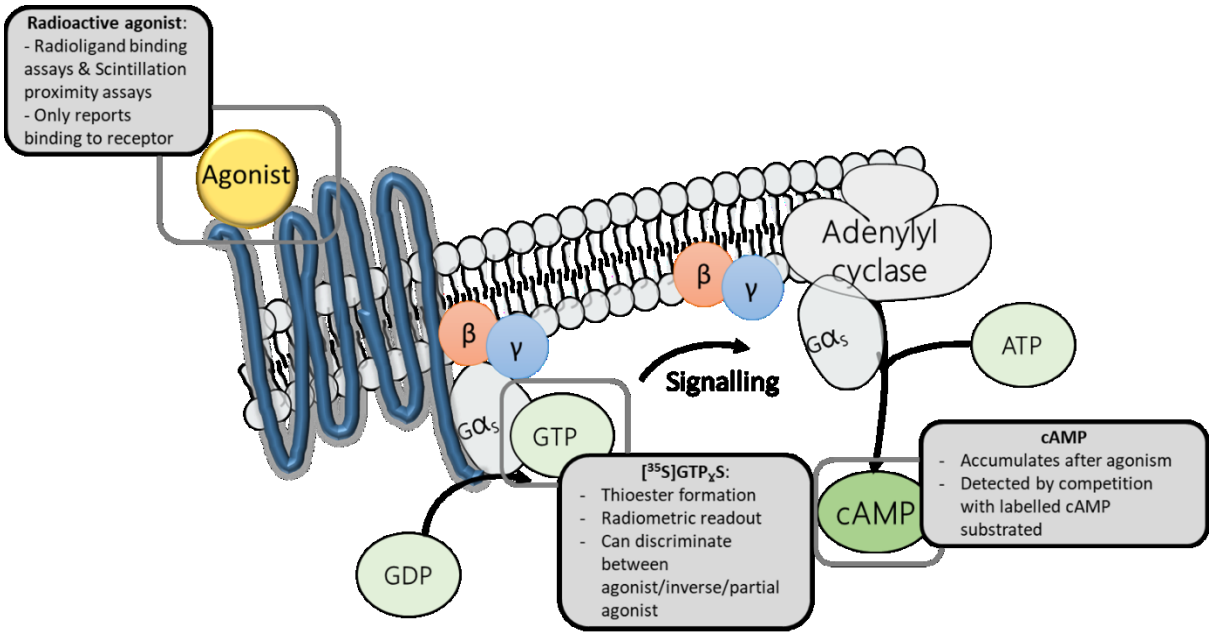


Figure 40: A summative diagram of the possible methods for assaying the binding and intereactions of an agonist towards a GPCR: Radioactive ligand binding, radiolabelled thio-GTP and secondary messangers (such as cAMP).

A derivative of the SPA has been developed to enable the analysis of GPCR function by employing a radiolabelled thio-GTP moiety ([^{35}S]GTPγS), first reported in 1983.[225] The thio-GTP moiety forms a thio-ester bond that resists hydrolysis with the G-protein upon activation and GDP cycling, thus for each activation event, one molecule of radioactive GTP is covalently bound to the receptor allowing detection. This method is particularly advantageous for receptors coupling to $G_{\alpha i}$ due to the low basal levels of GTP cycling and an alternative to measuring the low cAMP levels, that can make secondary messenger assays problematic.[226] However, the mechanism of labelling does mean that per activation event only one radioactive moiety is appended per receptor, potentially leading to a low signal that is dependent on receptor density on

the membrane. The assay is less suitable for Gαs coupled receptors due to the high background due to the high level of basal GTP cycling.

Functional G-protein dependent assays can give more information about the nature of the ligand being tested. With the ever-increasing range of photostable, funtionalisable fluorophores becoming available commercially, fluorescence read out is now employed more often than radiometric assays. With knowledge of the signalling cascade of the GPCR of interest, the secondary messenger can be probed as an indicator of receptor engagement and ligand binding mode in a sensitive and high throughput manner.

For GPR119, receptor activation leads to a signalling cascade that elevates internal cAMP levels (as detailed in Chapter 1), thus cAMP levels can be used to quantify receptor engagement. The general premise of these methods relies on competition of the endogenous cellular cAMP and a labelled cAMP that is used as the assay readout. In order to prevent hydrolysis of cAMP to AMP by phosphodiesterases an inhibitor such as IBMX is added to the cells prior to stimulation to allow the accumulation of cAMP.[227,228]

A number of cAMP detection kits have been commercialised.[229,230] Fluorescence Polarisation (FP) can be used to monitor the production of cellular cAMP: a fluorescently labelled cAMP analogue is introduced into the assay alongside an anti-cAMP antibody. Given the relatively large size of this complex compared to the free cAMP, the tumbling rate of the complex is slower. Thus, when the sample is irradiated with plane polarised light, the rate of depolarization of light is slower for the antibody-cAMP complex. When cellular cAMP is introduced, the labelled cAMP is competed out of the complex, increasing the rate of tumbling and also the rate of depolarization. Therefore, loss of polarization indicates the release of cellular cAMP. However, this technique has been shown to suffer from compound interference at higher concentrations.[231,232]

By modifying the labelled cAMP and antibody to contain a fluorescence resonance energy transfer (FRET) pair, a homogenous time resolved fluorescence (HTRF) assay was created. Within this set up, much like before, competition of the cellular cAMP for the fluorescently labelled anti-cAMP antibody releases one component of the FRET pair, thus leading to a reduction in FRET signal. This method is less susceptible to

compound interference as emission is measured at two wavelengths allowing a ratiometric normalisation of samples to remove background from ligands or changes in buffer pH/ionic strength under assay conditions.[233]

There are many more options when considering measuring cAMP as a readout of GPCR activity, such as Perkin Elmer's AlphaScreen technology that employs chemiluminescence, G-protein independent assays such as β-arrestin recruitment and methods that probe more distal events such as cAMP response-element binding protein (CRE) activation and downstream gene-transcription.[234]

2.4.2. cAMP Assay for the Evaluation of Chemical Probes for GPR119

For the purpose of assessing the activity of the synthesised chemical probes against GPR119, Perkin Elmer's LANCE *Ultra* cAMP assay was chosen. This assay employs time-resolved FRET (TR-FRET) as described previously: the cAMP tracer is labelled with europium cryptate while the anti-cAMP antibody is conjugated with a fluorophore, U*Light* (a proprietary fluorophore, but typically allophycocyanins are used as a FRET pair for Eu^{3+}). When bound to the antibody, excitation of the tracer at 320 nm leads to FRET energy transfer and fluorescence emission at 665 nm from the U*Light* fluorophore, as well as fluorescence emission at 615 nm from the Eu tracer regardless of FRET energy transfer. When cellular cAMP is bound to the antibody, the FRET pair is separated and thus no emission at 665 nm is observed hence the signal at 665 nm is inversely related to the concentration of cAMP within the samples (**Figure 41**).

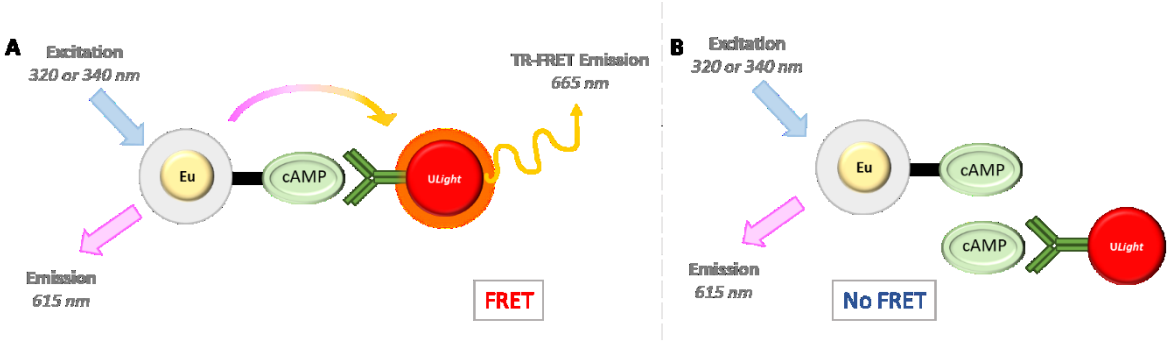


Figure 41: The mechanism of FRET emission using the LANCE *Ultra* cAMP assay kit. **A** in the absence of cAMP, FRET emission occurs, however **B** in the presence of cAMP, the *U*Light antibody binds to free cAMP, breaking the FRET pair. Hence, no FRET emission is observed in the presence of cAMP. The level of FRET emission is inversely proportional to the quantity of free cAMP in the system.

As an initial indicator into the binding of the probes to GPR119, the series of synthesised compounds were profiled alongside the AstraZeneca parent compounds (**AZ001**, **AZ002** and **AZ003**) all at 5 μM, **DMSO** (as a negative control) and 50 μM **OEA** (in line with GPR119 literature for a 100% activity control). To determine the

degree of GPR119 mediated cellular cAMP, the compounds were profiled in a HEK293 cell line stably expressing GPR119 alongside the wild type HEK293 cell line. Using a cAMP standard curve, the fluorescence data was interpolated to the give the cellular concentration of cAMP (**Figure 42**).

At this concentration, the AstraZeneca compounds all showed an activity comparable or greater than the activity of 50 μM OEA, in line with the literature values of intrinsic activity. When evaluating the probe compounds, the "off-target" pharmacophore does not tolerate much modification; probe **14** and the non-diazirine control probe **50** show no cellular cAMP relative to the wild type HEK293 cell line suggesting the compounds are not agonists of GPR119. However, by extending the amide to contain the alkyne handle in probe **15** and **52**, rather than extending from the amine we regain some activity. This is in line with the many structures in the GPR119 literature that possess a linear pharmacophore and demonstrate GPR119 agonism.

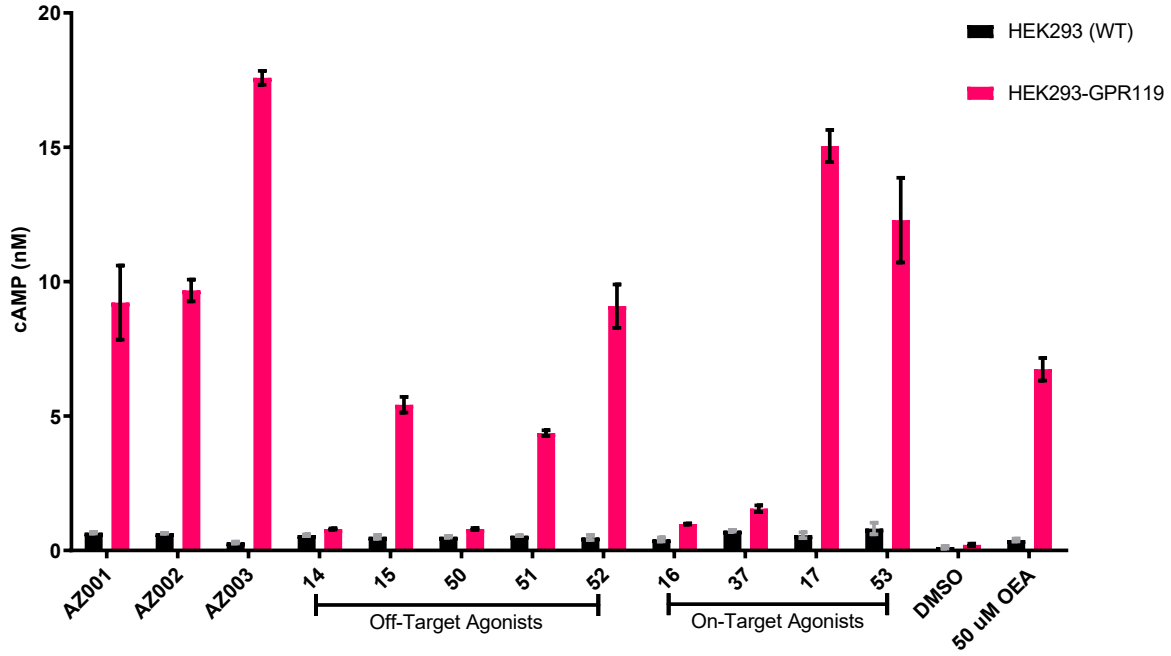


Figure 42: Agonist profiles of the chemical probes at 5 μM in both wild type HEK293 cells and HEK293 cells over expressing GPR119. Data has been interpolated from a cAMP standard curve to convert fluorescence readings into cellular cAMP concentration (nM).

Within the "on-target" series of probes, similarly appending the photocrosslinking handle on the terminal aromatic ring as in probe **17** retains activity for GPR119 whereas appending the same handle on the terminal cyclic secondary amine in probes **16** and **37** abolishes this activity. With the non-diazirine containing probe **53** activity

was also retained. Again, this preference for a linear pharmacophore has been documented and agrees with the homology model discussed in chapter 1.

With those compounds showing activity, full EC_{50} data was sought. Initial attempts to generate EC_{50} curves were problematic; data varied significantly between runs and issues with compound solubility at higher concentrations meant EC_{50} curves couldn't be plotted and compound concentrations weren't certain.

While the assay format as provided within the kit allows the incubation of the compounds with GPCRs in a "normal" cellular environment within intact cells as opposed to purified membranes, it also has several disadvantages. Using the suggested protocol, cells are resuspended in stimulation buffer (5 mM HEPES pH 7.5, 0.5 mM IBMX, 0.1% BSA, Hanks Balanced Buffer Solution) before stimulation with agonists. This buffer varies significantly from culture medium used to maintain the cells and requires stimulating the cells while not adherent; considerably different conditions to the normal behaviour of cells and other signalling assays. What's more, the assay relies on the data acquired falling within the linear region of the cAMP standard curve in order to be valid and comparable to prior assay runs. If the cellular concentration of cAMP falls outside of this region, for reasons such as differing number of cells, receptor density or stimulation propensity, the data for that specific run will be of no use. The current format is a "one shot" assay that requires substantial optimisation at every stage.

This is an issue that has been noted by other researchers and as such an adaptation of the suggested protocol has been published by *Hunter & Glass, 2015* in which the cells are stimulated in culture and then lysed using the buffer provided in the kit.[235] This allows multiple runs of the assay to occur with the same lysate, adjusting the concentration to allow the data to fall within the correct ranges. It also means the cells are stimulated in a more native environment and will be subject to less environmental variation. The lysate produced has been shown to be stable once frozen allowing multiple experiments to be run within the same assay run.[235] This method increases the flexibility and reliability of the kit.

In order to utilise this adaptation to improve the results for GPR119, cells were plated in 96 well plates to achieve around 80% confluency on the day of stimulation. Pre-incubation in stimulation media (DMEM, 0.5 mM IBMX and 5 mg/mL BSA) for 30

minutes ensured a stable base line of cAMP was established prior to inclusion of agonist. BSA was included as a lipophilic aid to increase the solubility of the ligands. Cells were then stimulated for 30 minutes in the presence of agonists before being

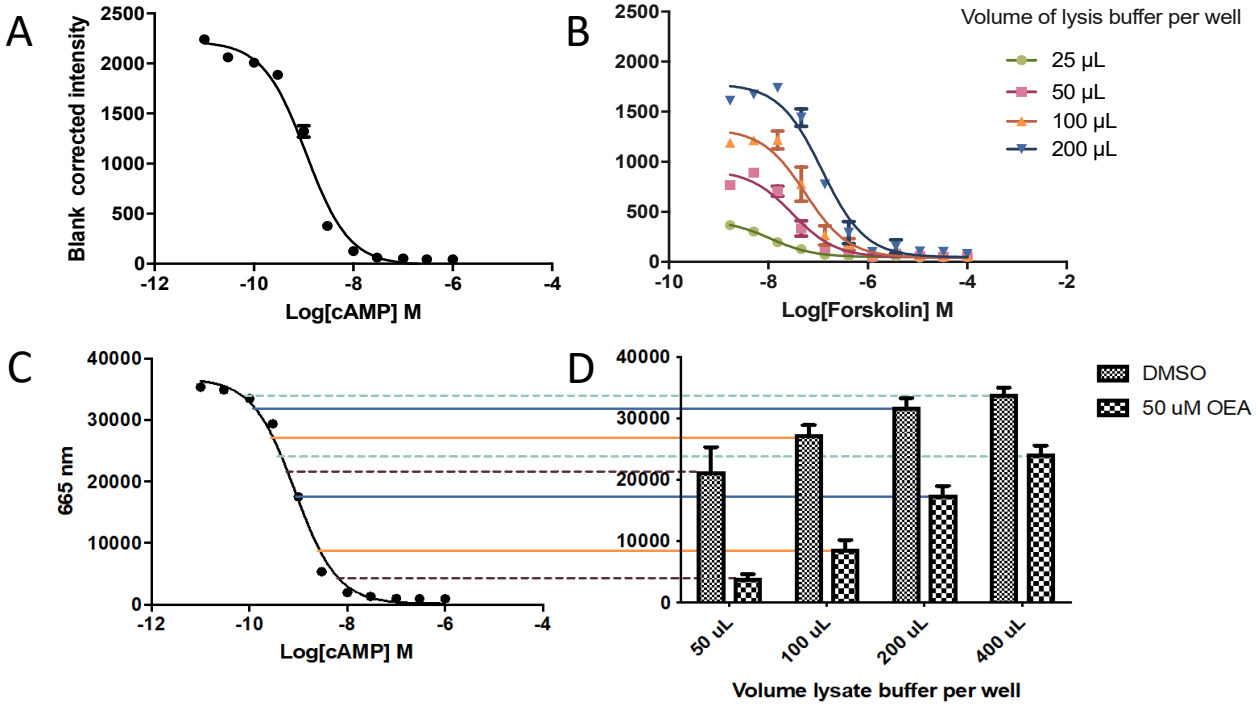
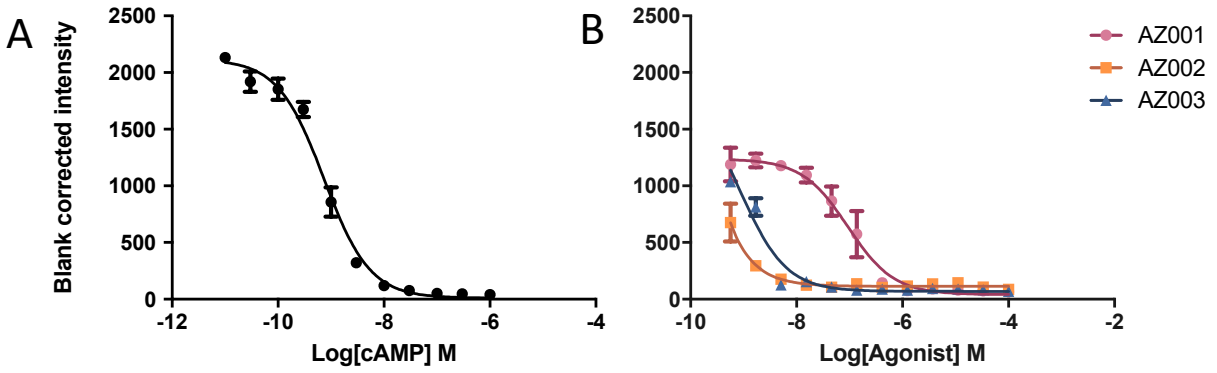


Figure 43: Optimisation of the cAMP workflow to enable analysis of treated cell lysates: **A** standard curve generated with a cAMP standard solution. **B** Dilution of forskolin stimulated cell lysates to obtain cAMP responses within a similar range to the standard curve. **C** cAMP standard curve comparison for **D** cell lysates stimulated with DMSO (0% activity) and 50 µM OEA (100% activity) as a measure of the dynamic range of the assay. Solid blue and orange lines represent conditions that fall within suitable pointed on the standard curve.



C

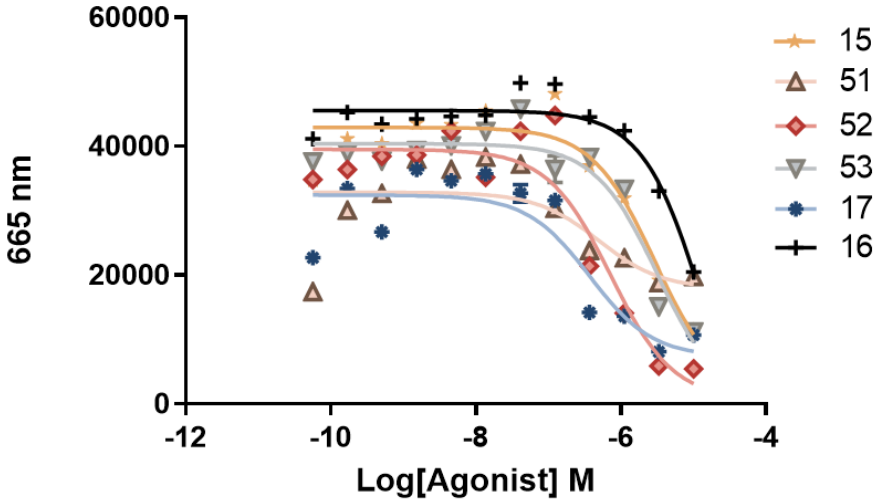
	Literature EC_{50} (nM)	Calculated EC_{50} (nM)
AZ001	74	93
AZ002	11	<1
AZ003	19	<1

Figure 44: Profiling the AstraZeneca tool GPR119 agonists. **A** cAMP standard curve using the optimised lysate method **B** Agonist profiles of the three GPR119 agonists. **C** EC_{50} as calculated compared to the reported literature values.

quenched by removal of stimulation media and placement of the plates on ice. Cells were lysed using the detection buffer included within the kit.

Dilutions of the lysates stimulated with forskolin, 50 μM OEA ("100% activity" control) and DMSO ("0% activity" control) were profiled against the cAMP standard curve to establish the most suitable volume of lysis buffer per well to ensure data fell within the linear region of the curve (**Figure 43**). For both, it was observed that 100 μL and 200 μL sat within the linear portion of the standard curve, and thus for future experiments samples were diluted to 150 μL total lysis buffer per 96 well for most data points to be within the linear region of the standard curve.

As an initial test of this adapted method, the three AstraZeneca compounds were used to stimulate the cells to evaluate their EC_{50}s relative to the literature values (**Figure 44**). The relative order of potency of the three compounds agrees with that reported in the literature. The actual EC_{50} values are not directly comparable as a different assay kit and format was used for the two, however they do appear to be in agreement with what has been previously reported.



	15	51	52	53	17	16
EC_{50} (μM)	3.922	0.5414	0.8422	2.085	0.3917	17.73
95% CI	2.785 to 5.523	0.191 to 1.532	0.614 to 1.154	1.284 to 3.385	0.235 to 0.654	8.054 to 39.03

Figure 45: Profiling GPR119 probes within the cAMP lysate adapted assay to calculate EC_{50} of the agonist probes.

With this optimised protocol, EC_{50} plots and calculations were obtained for a selection of chemical probes that had previously shown a response in the screening assay.

Previous issues with compound solubility appeared to be resolved in this format, with higher concentrations of probe tolerated (**Figure 45**). Probe potencies span the μM to nM range with the most potent probe **17** giving an EC_{50} of 392 nM and 95% confidence interval (CI) between 235 nM to 654 nM. While this is approximately a factor of 100 less potent than the parent compound, the covalent nature of the probes upon irradiation should ensure sufficient engagement to investigate the binding of these compounds.

2.4.3. Comments on Agonist Profiles

The implementation of an assay for GPR119 provides an essential read out for probe agonism as an indication of the effect on the probe-protein relationship upon modification. However, challenges around assay reliability and probe solubility required the optimisation of published and commercial assay protocols. While this study has achieved the development of an assay that allows the analysis of the agonist capabilities of probes within the nM to μM range, there remain intrinsic issues with the assay format.

Due to the nature of GPCRs, isolating recombinant protein to use in protein assays is a challenging task, thus the assays rely on the use of cells expressing the protein of interest. However, for most GPCRs the endogenous expression is low, giving inadequate responses and reliability for reproducible assays. The use of engineered cell lines stably expressing the protein of interest is therefore a prerequisite of these assays, however protein expression levels within cells vary within the same population and considerably between different passages of the cell line. The results generated are therefore only comparable within the same assay run, as levels of protein expression will vary in an undetectable manner, due to the lack of reliable antibodies for GPR119. The work described in chapter 3 may provide a means to identifying GPR119 expression and producing homogenous cell lines.

2.4.4. Conclusions

Within this chapter, the synthesis of a selection of chemical probes for GPR119 has been described based on pharmacophores generated by AstraZeneca. The pharmacophores selected demonstrate GPR119 mediated blood glucose lowering, and blood glucose lowering by an unknown mechanism of action. This is the first report of chemical probes for GPR119. This thesis will explore the use of these probes for

characterising GPR119 expression. The synthesis of “off-target” GPR119 chemical probes may allow the elucidation of a novel target for anti-diabetic drugs. Optimisation of a commercial cAMP kit has been described to allow the characterisation of the GPR119 chemical probes and provide further insight into the SAR of these agonists and select chemical probes to carry forward to biological experiments.

3. Development of biological tools to enable the study of GPR119

3.1. Introduction

3.1.1. Conventional tools and their disadvantages

Within the field of membrane proteins, GPCRs are a notoriously troublesome class of protein receptors to study. The fundamental structure of these receptors, with large areas of buried, hydrophobic residues, means their properties are highly sensitive to the environment they are handled in; often leading to aggregation and solubility issues when extracted from the cell membrane. What's more, GPCRs have varied levels of expression across relevant tissues, commonly with native levels of protein expression being very low.[28] This issue can be exacerbated by the dynamic processes involved in GPCR signalling. Upon various physiological stimuli or drug-mediated activation, receptors are often translocated to endosomes via clathrin-mediated endocytosis[236,237] for further processing by either degradation by lysosomes or recycling back to the membrane.[28,238,239] Given this, it is apparent that GPCR mRNA levels are not a sufficient indicator alone for the presence and abundance of GPCRs in various tissue types.[240–242]

To understand more about the roles of GPCRs of interest, detection at the protein level across various tissues, cell types and subcellular compartments is vital. Commonly, this is achieved using antibodies raised against the protein of interest. However, in the case of GPCRs the use of antibodies is often frustrating and unreliable; lack of selectivity within GPCR families due to high homology often leads to high background and false positive/false negatives.[243] Given the hydrophobic nature of the receptors, protein bands often run at unexpected molecular weights under gel electrophoresis conditions,[244] meaning western blots show shifts in molecular weight or the presence of dimers/multimers, reducing the confidence in specificity of the antibody.

Validation of the specificity of antibodies for GPCRs requires stringent screening as suggested by Michel *et al.* (2009): 1. Antibody staining disappears from tissue genetically engineered to lack the receptor (knock-out), 2. Treatment with genetic tools to knock-down expression significantly reduces antibody staining, 3. Transfection with

a screen of related subtypes of the receptor only yields antibody staining for the subtype of interest and 4. Multiple antibodies are raised against different epitopes of the receptor, all conferring the same immunoblot/immunohistological staining.[240] This method of validation has been applied to antibodies raised for CB2, showing the need for positive controls for antibodies with excellent sensitivity that lack specificity.[245,246] However, this is particularly challenging to achieve and has led to a number of examples in the literature of dubious western blot data, and even retractions of publications due to insufficient characterisation of antibodies.[247]

3.1.2. Methods used to study GPCRs

An increasing number of GPCR crystal structures have recently been published, to able the structural evaluation, enhance the understanding of their roles and lead more targeted medicinal chemistry campaigns.[243,248] Since the first high resolution 3D crystal structure for Rhodopsin was published in 2000,[21] 46 unique structures have now been found and characterised.[249] This growth is largely down to the development of new tools to alleviate the common issues surrounding the challenging crystallisation of GPCRs.

In order to produce high quality crystal structures, a large quantity of stable, uniform protein must be available.[243] Advances in protein expression in a number of organisms has aided this; while mammalian cells express protein with the correct post-translational modifications, they can be challenging and costly to culture in a large volume to produce adequate yields of protein, due to low expression levels.[250,251] Bacteria such as *E. coli* are routinely used to express recombinant proteins due to the ease and scalability of culture. However, bacterial expression is less well suited for GPCRs as the cells tend to lack the machinery and suitable lipid membrane composition required to stabilise the expressed mammalian proteins in bacterial membranes, and those proteins that are stabilised lack the correct post-translation modifications.[252] There have been several examples of GPCRs being expressed in large quantity in bacteria as insoluble inclusion bodies; however, this method requires very complex subsequent refolding of receptors to generate an active GPCR, the success of which varies from receptor to receptor.[253,254] Recently, more high-throughput methods for screening and solubilising GPCRs have enabled the expression of receptors in *E. coli* and led to several new GPCR structures being obtained.[255,256]

Alternative expression systems with increasing popularity include insect cell expression, and expression in yeast. Yeast strains, such as *Pichia pastoris* and *Saccharomyces cerevisiae* provide alternatives to mammalian cells as, like *E. coli*, they are known to rapidly expand, be economic to culture and produce large quantities of biomass.[257] Yeast are considered to be similar to higher eukaryotes in their ability to perform post-translational modifications, such as glycosylation in a simplified form.[258] Proteins may also retain stability when expressed in yeast and inserted into the membrane, as it has been shown that the lipid composition of yeast membranes is more comparable to that of higher eukaryotes.[259] However, the presence of a cell wall in yeast makes the lysis and isolation of the membrane, and proteins contained within, challenging.[260]

Similarly, insect cell expression in *Spodoptera frugiperda* cell lines has shown increasing popularity and success producing many of the recently published crystal structures via baculovirus-mediated protein expression.[250,260] Insect cell expression is cost-effective, and allows large volumes of cells to be cultured using biotechnologies such as WAVE bags and bioreactors.[250] Insect cells also possess machinery to post-translationally modify proteins in a similar manner to mammalian systems. As with yeast cells, *N*-glycosylation is less well represented by these organisms; however, cell lines have been engineered to replicate mammalian glycosylation patterns.[261,262]

Despite the various advantages and disadvantages, the successful expression of the protein in a specific system is still very dependent on the GPCR of interest. For example, each GPCR will have unique requirements regarding membrane lipid composition, post-translational modifications, protein trafficking and ultimately protein synthesis. Codon optimisation can go some way to enhance expression in the desired system.[263]

Given the intrinsic flexibility of GPCRs, crystals often suffer from poor packing and heterogenous conformations of the receptor,[264] leading to poor quality crystals and low resolution diffraction data.[243,260] Attempts to alleviate this issue involve locking GPCRs into one conformation: protein engineering,[24] stabilisation by point mutation[265,266] and interaction with a ligand or antibody[23] in order to drive the equilibrium to favour one conformation have all shown success in increasing the likelihood of crystal formation.[243] The engineering of thermostable GPCRs is a strategy that has proved particularly successful, allowing the concept to be commercialised by

Heptares under the trademarked production of stabilised receptors (StaRs™).[267] This technique has enabled many drug discovery campaigns towards GPCRs, including structure-based drug discovery and fragment-based drug discovery utilising previously unknown binding modes and conformations of GPCRs.[268,269]

After successfully expressing a GPCR, depending on the required use, more optimisation might be needed to stabilise it outside of the membrane, as, when removed, GPCRs will quickly degrade and denature.[270] For studies on activity and binding, membrane preparations can be obtained directly from the expression system without requiring extraction of the GPCR from the membrane. However, for crystallisation and processes requiring recombinant or solubilised protein, extraction from the membrane into either detergent or artificial membrane is necessary. The choice of detergent is critical to ensure complete solubilisation and compatibility with downstream processing.

There are four general classes of detergents based upon the various headgroups; non-ionic, anionic, cationic and zwitterionic detergents.[271] The hydrophobic component of the detergent varies considerably from unbranched saturated alkane chains, to aromatic or steroid moieties, and as such there is a huge variety of detergents to select from.[271] When choosing a detergent to use, general considerations include whether the receptor should remain active/correctly folded, if it needs to be solubilised or crystallised, or processed using a technique that may be particularly sensitive to the presence of detergent (such as mass spectrometry or nuclear magnetic resonance). As a general rule of thumb, detergents with larger, neutral headgroups, and longer hydrophobic tails will have milder properties, while conversely protein complexes will be disrupted or denatured in the presence of ionic short chained detergents (**Figure 46**).[272] This can be rationalised by considering the ability of the detergent to mimic the natural lipid bilayer; alkyl chains of 7-12 carbons in length correspond to the approximate thickness of a membrane.[270] In order to have a monodispersed protein-detergent complex in solution, approximately one micelle per protein is required – this can lead to the use of detergents significantly over their critical micelle concentration (CMC). Exceeding the CMC can cause problems in further processing of samples, as dialysis or detergent removal may be required for crystallisation or analysis of the protein.

By comparing a few commonly used detergents in protein science, it can be shown how detergent choice is critical to success. Sodium dodecyl sulphate (SDS) is a common detergent used across many applications within biology from SDS-PAGE electrophoresis, to cell lysis. It is also found in cosmetic products. The small anionic sulphate head group makes SDS a harsh detergent, and indeed it is known to denature many soluble and membrane bound proteins,[272] which makes it less well suited for stabilising membrane proteins. Triton X-100 is another commonly used laboratory detergent able to permeabilise and lyse cells. It is mild, with a short branched aromatic hydrophobic tail and a long hydrophilic polyethylene oxide chain and has been used to solubilise membrane proteins. However, the low CMC value of Triton X-100 makes it challenging to remove via dialysis, which, given its high UV absorption at 280 nm can be a large disadvantage for protein purification.[273]

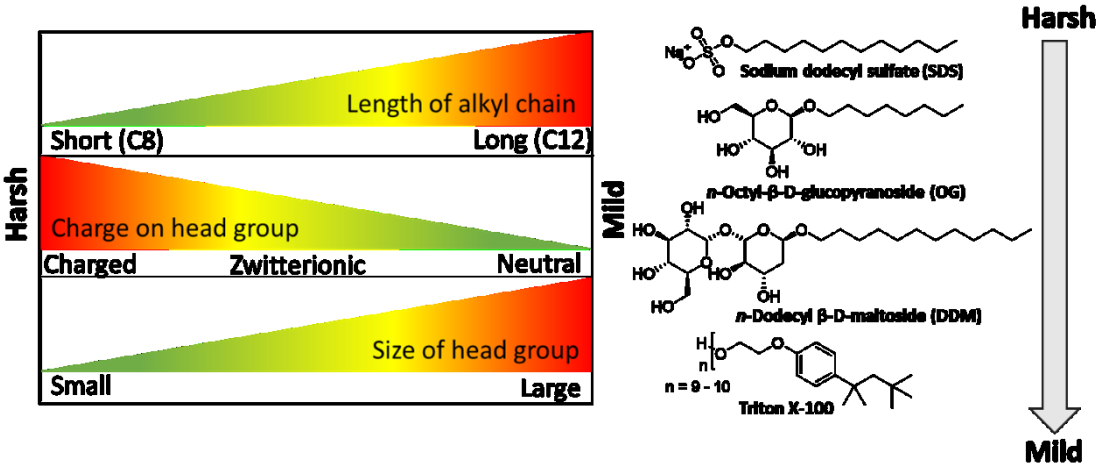


Figure 46: Variation within the detergent families: Left: characteristics of detergents that contribute to their relative harshness. Adapted from *Privé, 2007*. Right: A selection of commonly used detergents with a variety of headgroups & hydrophobic tails.

Another class of detergent frequently seen in membrane protein biology are the glycoside-based compounds. These non-ionic detergents have sugar based hydrophilic headgroups, and straight chain alkyl tails as a hydrophobic component. *n*-Octyl-β-D-glucopyranoside (OG) has a short alkyl tail (C8), with one sugar headgroup making it capable of forming small micelles and as such is considered a moderately harsh detergent. However, this detergent is one of the most successful for membrane protein crystallisation studies as it doesn't cause protein denaturation and can be removed with ease (due to a high CMC).[274] *n*-Dodecyl β-D-maltopyranoside (DDM) varies structurally from OG by the addition of a second sugar in the head group, and

a longer (C12) hydrophobic tail making this one of the more mild choices. DDM is very commonly used within the field of GPCRs and membrane proteins as the larger micelles produced shield more of the protein, increasing the probability of maintaining function once solubilised. However, the larger micelles and lower CMC make DDM a challenging detergent to use for protein crystallisation.[272]

An alternative to detergent solubilisation is the use of nanodiscs as artificial membranes. Nanodiscs are composed of lipids assembled into lipid bilayer discs encased in solubilising membrane scaffold proteins.[275,276] The lipid composition and stoichiometry determines the properties of the self-assembled discs and can be optimised for various membrane proteins. This membrane mimicry can be particularly useful for membrane proteins that otherwise aggregate or are unstable to detergent-based protein purification, via the direct assembly of the protein of interest into nanodiscs directly from detergent solubilised cell membranes, allowing the purification process to be avoided entirely.[277,278] Nanodiscs have shown to successfully stabilise membrane proteins for a number of structural biological techniques such as X-ray crystallography[279] and cryo-electron microscopy (cryo-EM).[280]

3.1.3. Use of Protein Tags and Labels

An additional tool within the protein expression toolbox is the use of affinity tags to aid expression, localisation, purification, solubilisation and identification of proteins of interest. The first uses of fusion proteins were used to increase protein expression and solubilisation in order to make the expression system more efficient, such as the use of Staphylococcal protein A.[281] Typically, solubility is enhanced by the use of fusion proteins (as opposed to small peptide tags). Not all fusion proteins carry the same solubility enhancing ability, however: tags such as Protein A, Maltose binding protein (MBP)[282,283] and N-utilization substance A (NusA)[284] have been shown to significantly enhance solubility upon protein expression.

Epitope tags are short peptide sequences that, when engineered into a protein sequence, ease the detection and allow purification of the protein by using a surrogate biorthogonal sequence that can be recognised by antibodies or affinity beads. Short tags, such as HisTag (6xHistidine residues) and FLAG (DYKDDDDK) are frequently favoured due to their low metabolic burden, availability of biochemical regents (such as specific antibodies and resin) and mild, native elution conditions.[285–287] Similarly,

tags with varying sequences and specificities are available, such as StrepTag, HA tag and c-myc.

Similarly, fluorescent tags are now available in an array of functionalities. The discovery of Green Fluorescent Protein (GFP) led to the use of fluorescent protein adducts to illuminate several biological processes.[288] By taking advantage of the unique excitation and emission profile of each fluorophore, it is also possible to use fluorescent proteins as biosensors through fluorescence resonance energy transfer (FRET).[289,290] This field has developed further to introduce engineered proteins that can be made fluorescent on the addition of small molecule ligands, to improve the fluorescent properties and quantum yield, allowing more advanced microscopic methods to be used. Such tags as SNAP-tag and CLIP-tag consist of a relatively small protein (20 kDa) that forms a covalent fluorescent adduct on the addition of O^6-benzylguanine or O^2-benzylcystosine derivatives respectively.[291,292] Fluorophores have also been developed to enable site-specific labelling of amino acid sequences incorporated into proteins. For example, fluorescein arsenical hairpin binder-ethanedithiol (FlAsH-EDT_2) is a derivative of fluorescein containing arsenic dithiols that, upon incubation with the recognition sequence (Cys-Cys-Xxx-Xxx-Cys-Cys), form strong bonds to cysteine pairs (the same chemistry responsible for arsenic toxicity). Fluorescence can then be switched off by removal of the FlAsH molecule using 1,2-ethanedithiol, a strong sequester of arsenic.[293,294]

The growing array of protein tags available in the academic and commercial domains means previously challenging proteins, such as GPCRs, are becoming approachable through protein engineering. The choice of tag is dependent on the system of choice; while small epitope tags give minimal disruption to protein sequence and thus binding and folding, they may not always be the best choice. Large protein tags, while potentially hindering correct folding, can offer increased solubility and detectability of the protein of interest. Balancing the requirements of the protein of interest with the outcome desired will enable the selection of a suitable tag.

3.2. Attempts to ID GPR119

3.2.1. The trouble with antibodies

As mentioned previously in this chapter, the use of antibodies to infer protein expression of GPCRs is limited. Within the literature to date, only a limited number of publications report data relating to GPR119 utilising an antibody for GPR119.[295–297] The extent of this has been observed by Cornall: "However, due to the unavailability of a reliable commercially available antibody raised against the mouse GPR119 protein, we are unable to report with confidence the protein product in this cell type. This finding is consistent with other recent papers reporting only GPR119 mRNA abundance" (Cornall 2013).

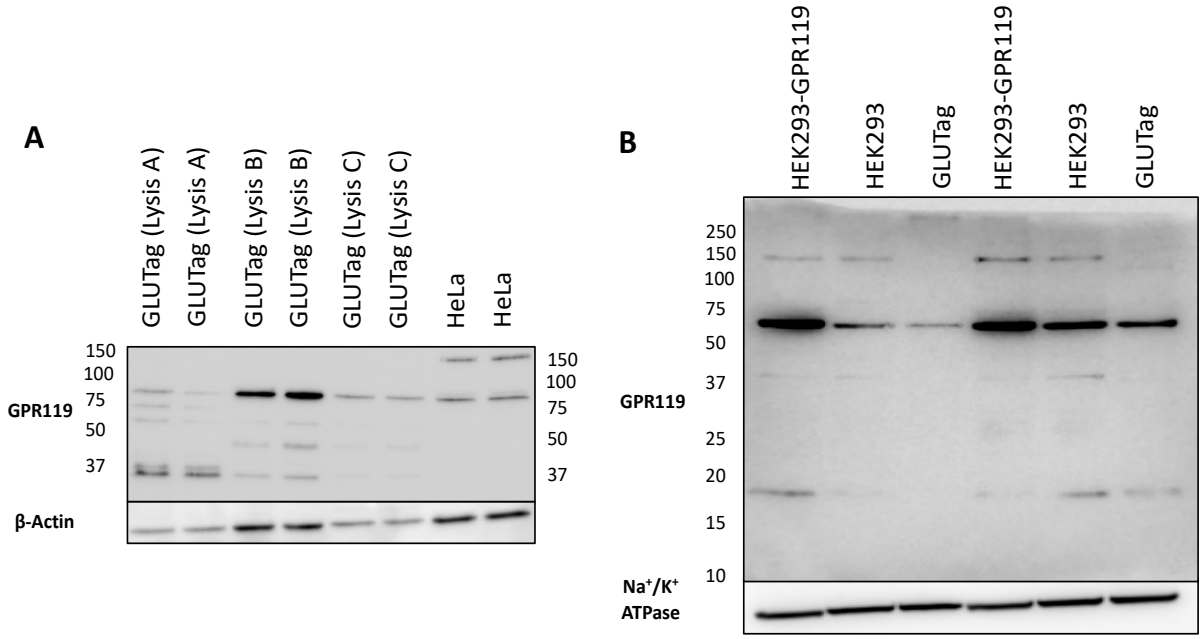


Figure 47: Immunoblot with various GPR119 antibodies: (**A**) GLUTag cells (GPR119 positive) or HeLa cells (GPR119 negative) were lysed using either **A** 0.1% SDS, 1% Triton X-100 in PBS, **B** 3x freeze-thaw in PBS or **C** 1% Triton X-100, 150 mM NaCl in 50 mM Tris pH 8 all showing a variety of non-specific bands with a GPR119 antibody (sc-99103). Replicates of each lysate have been run. (**B**) Comparing HEK293-GPR119 cells (overexpressing GPR119), HEK293 cells (negative for GPR119 expression) and GLUTag cells (endogenous levels of GPR119) for GPR119 expression using a GPR119 antibody (ab75312), no bands correlate to the expected relative levels of GPR119. Cells were lysed in RIPA buffer and replicates of each lysate have been run. Expected MW of GPR119 is 37 kDa.

There is a plethora of commercial antibodies available from various suppliers claiming to be specific for GPR119. However, upon evaluating a number of these antibodies in relevant cell lines, no specific immunoblotting was observed (**Figure 47**). By comparing the bands across various cell lines and lysis conditions, it was hypothesised that a band specific for GPR119 could be observed. For the Santa-Cruz antibody, bands correlating to the expected molecular weight of GPR119 at 37 kDa were present

in GLUTag cells (a cell line expressing endogenous levels of GPR119), as were many non-specific bands (**Figure 47A**). Upon changing lysis conditions, the relative intensity of bands changed, potentially suggesting the presence of aggregates. This is especially apparent when using detergent-free freeze-thaw lysis conditions; lack of detergent promotes the formation of aggregates within solution. However, bands of a similar molecular weight were also observed in a HeLa cell lysate, suggesting the bands were not specific for GPR119.

Similarly, using an Abcam antibody for GPR119, an analogous non-specific band pattern was observed across cell lines with theoretically varying levels of GPR119 (**Figure 47B**). HEK293 cells overexpressing GPR119 were compared to the endogenous GLUTag cell line, as well as negative control HEK293 cells. Given the failure of commercial antibodies for GPR119, an alternative strategy, such as engineering a recognition tag into the protein sequence, is required to allow the detection of GPR119.

3.2.2. Limitations of Mass Spectrometry

An alternative widely used and more global method of monitoring protein levels in a tissue of interest is mass spectrometry-based proteomics. Soluble, highly abundant proteins are readily identified by the standard protocols available. However, hydrophobic proteins of a lower abundance are perceived as major challenges within the field.[196]

In order to expand the coverage of proteins from the cell lines of interest, HEK293-GPR119 cells were fractionated to enrich cytosolic, nuclear and membrane proteins. The fractionation process utilises mechanical shearing and hypotonic buffers to swell and break the cellular membrane, releasing cytosolic proteins and leaving cellular debris, cell membranes and nuclei as a suspension in solution. Gentle centrifugation allows the removal of larger organelles and cellular debris (referred to as the nuclear fraction) leaving membranes and cytosolic proteins in solution. Ultracentrifugation then clarifies the solution to provide the soluble proteins (referred to as the cytosolic fraction) and insoluble fraction (consisting of the membrane fraction). Fractions were created in triplicate and after isolation, the fractions were subjected to reduction and alkylation, followed by trypsin digestion overnight. The peptide samples were then analysed using LC-MS/MS. However, despite the vastly increased coverage with the

identification of over 5,000 proteins, no GPR119 was identified despite the increased expression levels of the engineered cell line.

3.3. Design of Constructs for GPR119

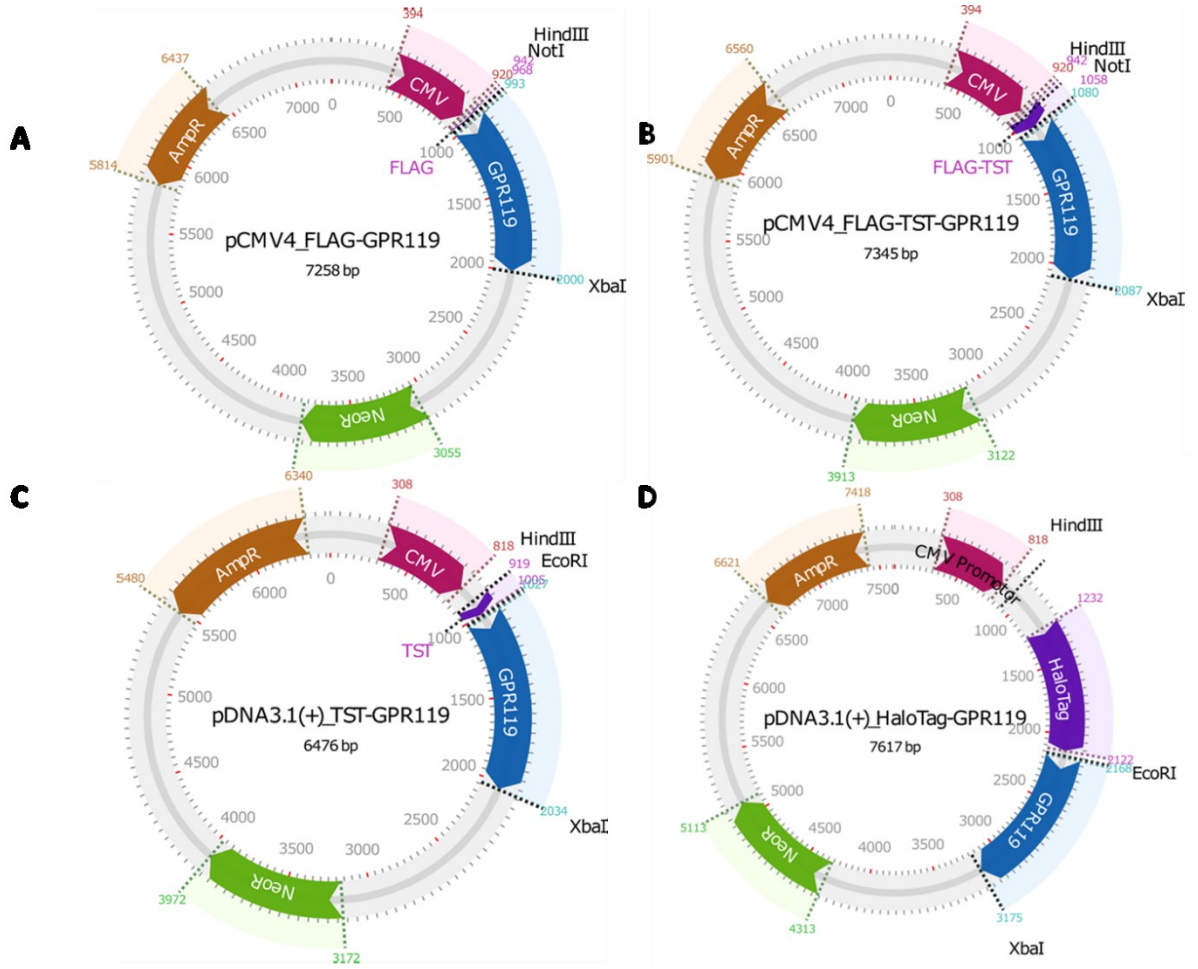


Figure 48: Plasmid maps of plasmids synthesised for tagged-GPR119 expression. (**A**) FLAG-GPR119 plasmid, as provided by Chepurny *et al.*; (**B**) FLAG-TST-GPR119 plasmid, constructed by inserting TST sequence into plasmid A; (**C**) TST-GPR119 plasmid, constructed by inserted TST-GPR119 sequence from B into a pDNA3.1 vector backbone; (**D**) HaloTag-GPR119 plasmid, constructed by inserting HaloTag sequence into plasmid C using appropriate restriction site manipulations.

An array of constructs was designed to contain a variety of tags in order to optimise chances of detection. Within the GPR119 literature, FLAG-tagged constructs have been used with apparent success.[298] Chepurny *et al.* demonstrated the identification of FLAG-tagged GPR119 in cell lysates, and kindly provided a sample of their plasmid for use in this study (**Figure 48A**). The protein constructs were designed such that the tag was expressed on the *N*-terminal of the protein. This domain is expressed extracellularly within the GPCR family and thus should allow probing and profiling of intact cells without the need to permeabilise the membrane. This is particularly useful

to confirm protein expression during cell culture if a fluorescent tag is fused to the protein, or to confirm surface protein expression on mounted cells using immunofluorescence. Where larger tags or proteins were expected to be expressed before the protein, short spacers were included to prevent loss of function of the engineered construct.

3.4. Mammalian cell expression

3.4.1. FLAG-GPR119 transfection limitations

HEK293 cells were transfected with the FLAG-GPR119 construct described above using a liposomal transfection method. Uptake of the plasmid and successful expression of the labelled protein were assessed initially using western blot, taking advantage of the FLAG epitope. Titrating increasing amounts of plasmid cDNA into the cells gave rise to an increase of band density at around 30 kDa and 50 kDa, indicative of protein expression related to the successful uptake of the plasmid, as shown in **Figure 49**. Controls for protein loading and protein expression remain constant, however a significant number of non-specific bands in the western blot are also observed at similar molecular weights to the potential bands of interest. The identity of the two plasmid-dependent bands could be assigned as the monomeric protein and a dimer at the higher molecular weight, as this is not uncommon for membrane proteins. The aggregate may be observed due to a large volume of protein

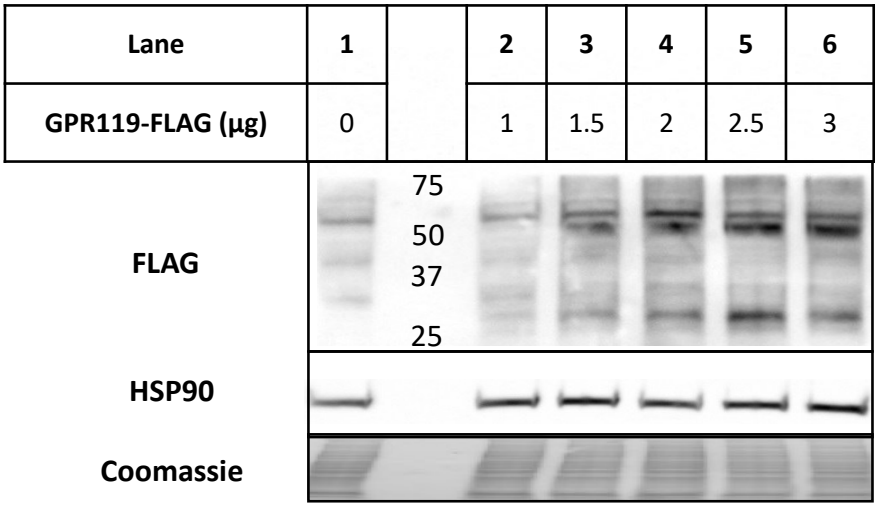


Figure 49: Immunoblot with anti-FLAG and anti-HSP showing the titration of GPR119-FLAG plasmid DNA into HEK293 cells using Lipofectamine LTX. Increasing density of bands can be observed around 30 kDa and 50 kDa relative to the loading control of HSP90 and Coomassie staining. Non-specific bands for the FLAG antibody can also be observed.

expressed by the cell being incorrectly trafficked to the membrane leading to incorrect folding or aggregation within the cytosol.

The FLAG epitope was visualised using immunofluorescence in order to establish the location of protein expression across the cell. For a GPCR, evidence of membrane localisation would be expected as an indication of successful processing. As a control for transfection, expression and localisation, a plasmid containing FLAG tagged β-adrenergic receptor 2 (FLAG-B2AR) was used as this is a highly studied system within the field of GPCR research.[239,299–301]

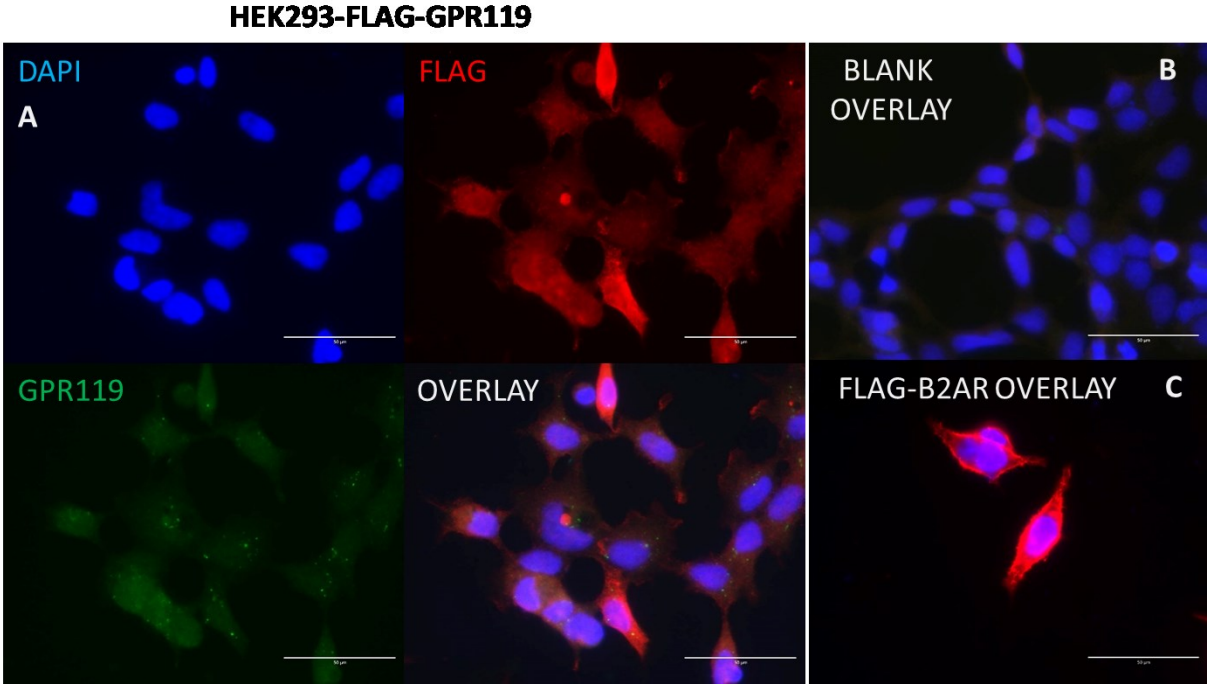


Figure 50: Immunofluorescence imaging of HEK293 cells transfected with FLAG-GPR119. **A** The nucleus is stained with DAPI, with GPR119 fluorescence detected in the green channel and FLAG in the red. The merged image shows both FLAG and GPR119 being present in the cytosolic domain with limited co-localisation. **B** The blank image represents untransfected cells treated with both antibodies and DAPI stained. **C** Conversely, the FLAG-B2AR control shows good FLAG membrane localisation. 40 x objective; scale bar = 50 µm.

Indeed, FLAG-B2AR shows a strong immunofluorescent signal at the membrane of FLAG-B2AR transfected HEK293 cells (**Figure 50C**) compared to wildtype HEK293 cells (**Figure 50B**). For FLAG-GPR119 transfected HEK293 cells, the picture is more complex. Interestingly, an antibody for GPR119 shows a specific immunofluorescent signal to cells transfected with GPR119, however, expression appears to be low and predominantly as small islands within the cytoplasmic domain (**Figure 50A**). This may be indicative of the antibody having limited immunoreactivity to epitopes present during early processing or unavailable in fully formed GPCR, or it may be indicative of insufficient protein expression and membrane localisation. The lack of membrane

localisation is also apparent from the FLAG immunofluorescence response, with localisation predominantly within the cytosol and the nucleus.

To further investigate the distribution of FLAG-GPR119 expression within the transfected HEK293 cells, a comparison between detergent (0.2% Triton X-100) permeabilised cells and cells incubated in the absence of detergent was evaluated. In the presence of detergent, the cell membrane will be permeabilised allowing reagents (in this case, antibodies) to cross the membrane and react with any protein inside the cell. Conversely, in the absence of detergent, there will be no immunoreaction within the cell, allowing the assessment and comparison of surface expression of the protein of interest.

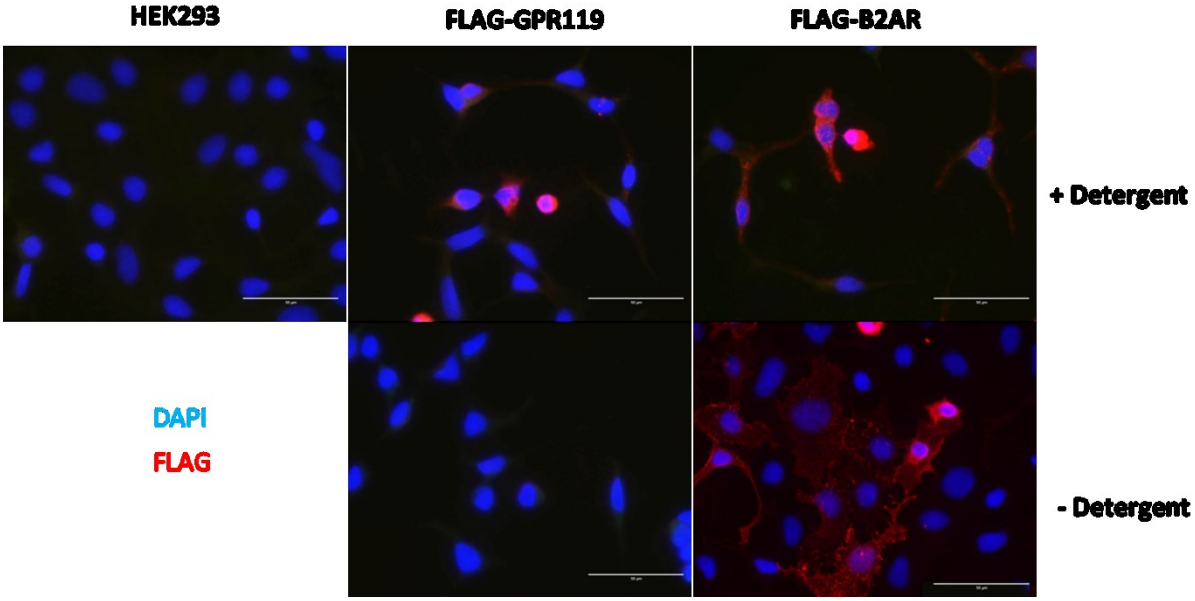


Figure 51: Immunofluorescence images of HEK293 cells transfected with either FLAG-GPR119 or FLAG-B2AR. No transfection control pictured on the left. Cells were incubated with PBS containing 0.2% Triton X-100 (+ Detergent row) or PBS alone (- Detergent row). Detergent incubation permeabilises cells and thus allows comparison of FLAG expression on the surface (- detergent) with internalised FLAG expression (+ detergent). FLAG-GPR119 appears to have no surface expression compared to FLAG-B2AR which exhibits a typical surface expression pattern in the absence of detergent. 40 x objective; scale bar = 50 µm.

In cells transfected with FLAG-B2AR and incubated in the absence of detergent, cell membrane localisation is observed, with the cell membrane boundaries easily identifiable (**Figure 51**). The membrane localisation remains present with the addition of detergent, but the internal space of the cell also shows a fluorescent signal, perhaps representing protein that has yet to traffic to the membrane or is located within internal cellular membranes. However, when FLAG-GPR119 is assessed with the same criteria, no fluorescent signal is observed in the absence of detergent, suggesting that there is no FLAG-tagged protein localised to the external membrane of the cells.

Due to the large quantities of DNA often used to successfully transfect cells with the protein of interest, the cells may express an abundance of protein that may lead to aggregation and improper processing, which could explain the unsuccessful results observed so far. Within multiple cell cycles, protein expression would be regulated to a level that is non-toxic to the cell, to enable correct processing, localisation and function of the desired protein. However, without a method to select for cells expressing the protein of interest, wild type cells thrive, and the expression is lost over a series of cell passages.

In order to retain expression, antibiotic resistance can be introduced along with the protein of interest either by incorporating both genes into a plasmid, or by co-transfection of a plasmid containing the protein of interest and a plasmid containing a resistance gene. In theory, any cells expressing the protein of interest should also be expressing the antibiotic resistance and can thus be selected by including the antibiotic in the cell culture medium to kill off untransfected cells. In this case, the FLAG-GPR119 construct (**Figure 48A**) used included a neomycin resistance (NeoR) gene to enable the transfected cells to be selected using geneticin (G418), an aminoglycoside peptide synthesis inhibitor. G418 concentration was optimised to give a moderate level of cell death, in order to encourage the transfected cells to thrive, while killing wild type cells. A G418 concentration of 0.4 mg/mL was chosen to select a cell population over 3-4 passages. However, immunofluorescence of the surviving cells showed no reactivity towards a FLAG antibody, suggesting cell selection was not successful.

3.4.2. FLAG-GPR119 Lentiviral Transfections

As an alternative route towards a stable FLAG-GPR119 expressing cell line, a lentivirus containing FLAG-GPR119 DNA was produced by harvesting the media from HEK293T cells transfected with a pLenti-FLAG-GPR119 plasmid (synthesised by GeneArt® ThermoFisher), along with viral envelope and packaging plasmids, pMDG and p8.9 respectively. The harvested virus was filtered and added to the media of HEK293 cells, in order to inoculate the culture and initiate viral transduction of FLAG-GPR119. Selection of successfully transduced cells was possible by the addition of puromycin (a potent protein synthesis inhibitor), thanks to the relevant resistance gene inclusion in the plasmid. Once the selected cells achieved confluency, they were analysed as before to probe for GPR119 expression *via* the FLAG tag. However, as depicted in **Figure 52**, no expression was observed by immunoblotting or by immunofluorescence.

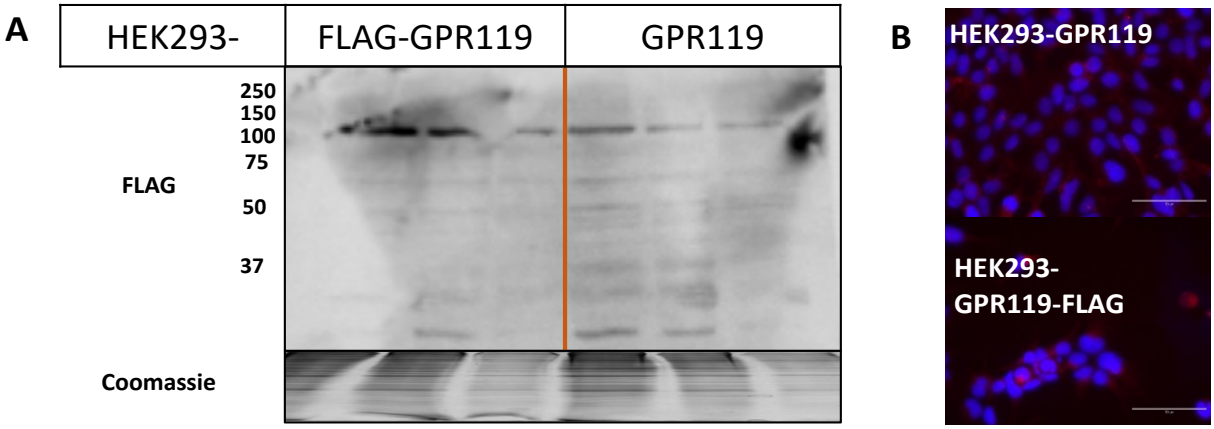


Figure 52: **A** Immunoblot and **B** immunofluorescence using anti-FLAG antibody in HEK293 cells overexpressiong GPR119 and HEK293 cells over expressiong FLAG-GPR119. No detection of FLAG signals is apparent.

There are multiple possibilities for this negative result; (1) at the genetic level, the cells have rejected the gene for the protein of interest while adapting to maintain puromycin resistance; (2) at the protein level, the cells have incorporated the desired genes but are not successfully expressing the protein or (3) the cells have incorporated the desired genes and are successfully expressing GPR119 protein, but FLAG is not a suitable epitope tag for this system.

Initial investigation focused on whether the transduction was successful in incorporating FLAG-GPR119 DNA into the genome of the stabilised HEK293 cells. qPCR using RNA extracted from wild type HEK293 cells, HEK293-GPR119 cells obtained from AZ and the transduced HEK293-FLAG-GPR119 cells with quantification using a housekeeping gene (GAPDH) allowed the comparison of GPR119 mRNA across the cell lines. Interestingly, by calculating the $2^{\Delta\Delta Ct}$ value, it was shown that GPR119 was indeed incorporated into the genome of the transduced cell line and with three times the abundance of the stabilised GPR119 cell line obtained from AZ (**Figure 53**).

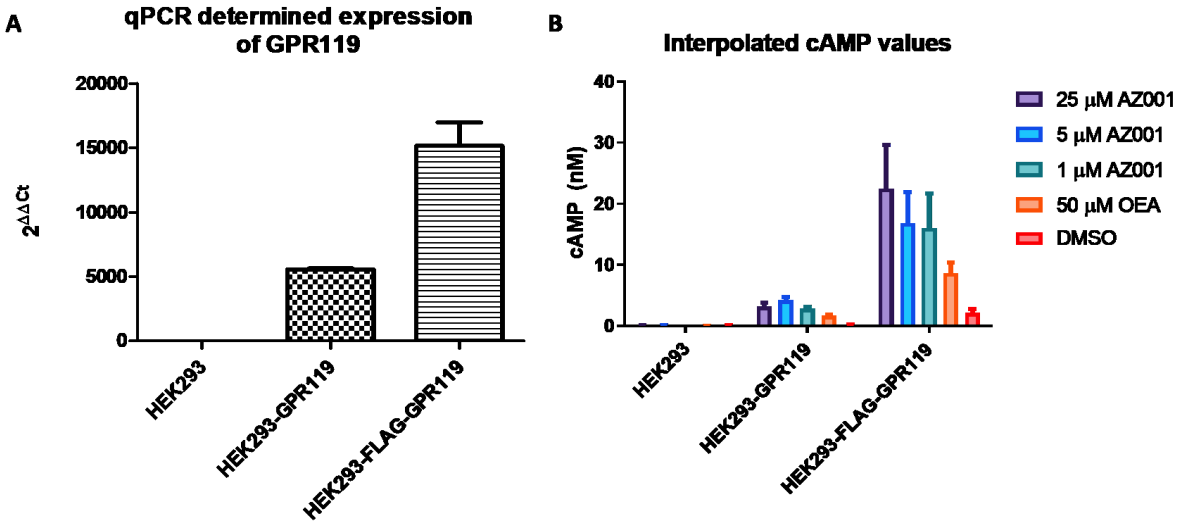


Figure 53: **A** Expression levels of GPR119 across wild type cell line, and both GPR119 and FLAG-GPR119 stabilised cell lines as quantified by mRNA levels relative to GAPDH. **B** Expression of functional GPR119 levels across the three cell lines determined by cAMP assay with GPR119 agonist AZ001, OEA and DMSO.

Following on from this result, a cAMP assay was carried out using the same three cell lines, to establish whether GPR119 protein was successfully expressed and functional. Using the AZ GPR119 agonist **AZ001** across a range of concentrations largely in excess of the EC_{50}, GPR119 specific cAMP production was interpreted, in comparison to the endogenous ligand **OEA** as a positive control and **DMSO** as a negative control. As shown in **Figure 53** the transduced cell line is indeed capable of expressing functional GPR119, with considerably more active protein than the AZ cell line.

Given these results, it can be concluded that the FLAG epitope is not ideal for detecting correctly expressed GPR119 at the membrane. These results correlate with the earlier immunofluorescence data that showed a lack of membrane FLAG localisation,

suggesting that the protein may have indeed been present at the membrane but undetectable. As detection using antibodies is so critical to this project, a variety of tags were explored for their properties, as described in the following sections.

Constructs containing N-terminal TST and TST-FLAG sequences (**Figure 48B and C**) were also transfected into HEK293 cells with no successful identification of modified GPR119. Baculoviral generation and subsequent transduction within insect cells also failed to generate TST-GPR119 in insect cell membranes.

3.4.3. HaloTag-GPR119

In order to combat the apparent failure of short epitope tags (perhaps due to interactions of the tags with the extracellular loops) to aid GPR119 identification, a fusion protein construct was synthesised comprising of an N-terminal HaloTag fused to GPR119 (**Figure 48D**). HaloTag first appeared in the literature in 2008, published by scientists at Promega.[302] The protein tag is a modified bacterial protein that has no endogenous function in mammalian cells. In its normal function, the protein is a haloalkane dehydrogenase operating by a nucleophilic displacement mechanism to remove halogens from aliphatic hydrocarbons (**Figure 54**).[303]

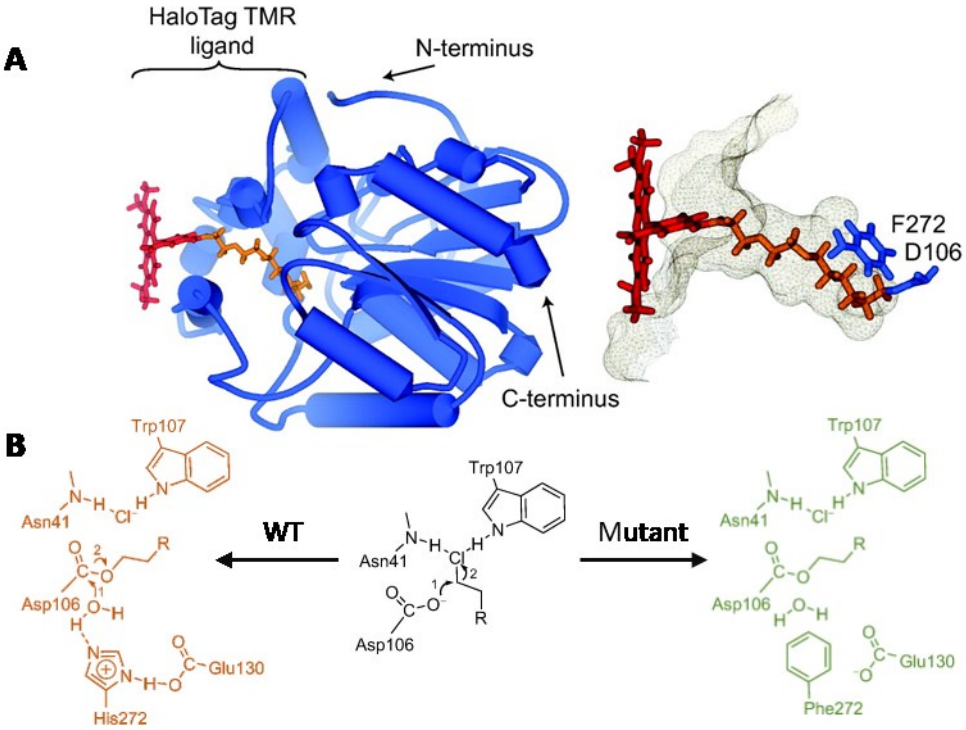


Figure 54: **A** HaloTag protein structure with HaloTag TMR (TAMRA) ligand covalent modification. **B** Mechanism of action of the mutant dehydrogenase protein. Mutation of Histidine to a non-basic residue prevents hydrolysis of the protein-substrate complex, and hence results in covalent modification. Adapted and reprinted with permission from *HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis*, Los *et al*. Copyright 2008 American Chemical Society.

The active site of the protein houses typical catalytic residues that facilitate the specific base-catalysed hydrolysis of haloalkane bonds (**Figure 54**): Aspartate forms a covalent adduct with the substrate, while histidine acts as the nucleophilic base that catalyses the hydrolysis of the protein-substrate adduct. It was shown that by mutating this conserved histidine residue to a non-basic amino acid, hydrolysis of the adduct was unable to proceed, leaving a covalently modified protein. By using haloalkane-based substrate, the protein can be modified with an array of functionalities such as fluorophores for visualisation or affinity tags (such as biotin) for enrichment.[302]

3.5. HaloTag-GPR119 as a viable tool to study GPR119

3.5.1. Evaluation using in-gel fluorescence & western blot

A panel of transfection conditions were screened in order to determine the optimal quantity of DNA, ratio of DNA to lipofectamine and duration of transfection to achieve maximal protein expression with minimal impact on cell viability. HEK293 cells were simultaneously transfected with a free fatty acid receptor 2 HaloTag (FFA2-HaloTag) construct previously validated within the group as a positive control (lane **10**, **Figure 55**, data unpublished). Transfected cells demonstrated significant cell death when

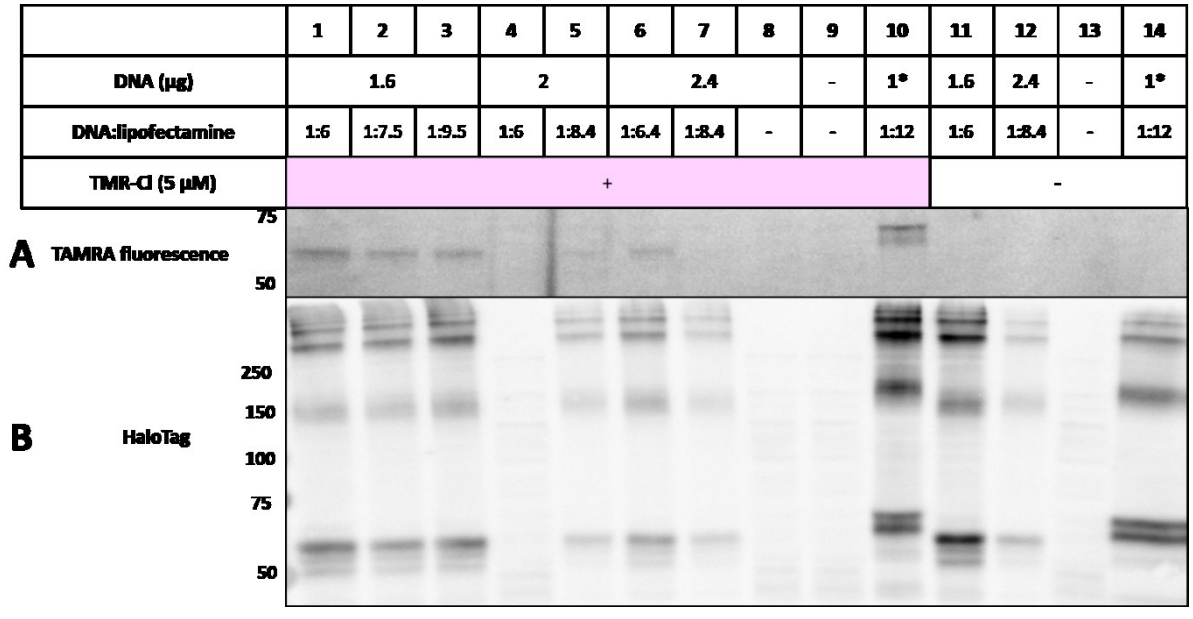


Figure 55: Transfection optimisation in HEK293 cells plated in 12 well plates, lysed 24 hours after transfection. Control samples (*) with positive control HaloTag-FFA2 (10 and 14), no DNA (9 and 13) and no lipofectamine (8) were also prepared. Transfection conditions used in 1 were used for future studies. **A** Samples 1-10 were incubated with TMR-Cl (5 µM) for 1 hour in the dark post lysis before being visualised using in-gel fluorescence. Samples 11-14 were incubated with DMSO. **B** After fluorescence visualisation, the gel was transferred to nitrocellulose, blocked and visualised with HaloTag antibody.

transfected for over 24 hours, with large quantities of DNA and with higher ratios of lipofectamine, therefore conditions used in lane **1** from **Figure 55** were chosen for future experiments. To verify the reactivity of the HaloTag with chloroalkane ligands, samples were incubated with, or in the absence of, a TAMRA conjugated chloroalkane ligand TMR-Cl synthesised by a member of the group (**Figure 56**). Samples incubated with the ligand show a clear band around the expected molecular weight, and a double band observed in the FFA2-HaloTag samples (FFA2 is known to be glycosylated and annotated as such on UniProt),[304] the expected molecular weight of both HaloTag-GPR119 and HaloTag-FFA2 is 72 kDa (**Figure 55A**).

To orthogonally validate that the bands being visualised were in-fact the HaloTag-GPR119 protein, the gel was transferred to a nitrocellulose membrane and immunoblotted using a HaloTag antibody (**Figure 55B**). The same bands were visualised using this methodology, as well as probable multimers and aggregates at higher molecular weights, strengthening the confidence that the HaloTag modification allows the visualisation of GPR119 using in-gel fluorescence and immunoblotting techniques. Given the previous difficulties isolating and identifying GPR119, this result provided a significant advance in the methodologies available to study this GPCR.

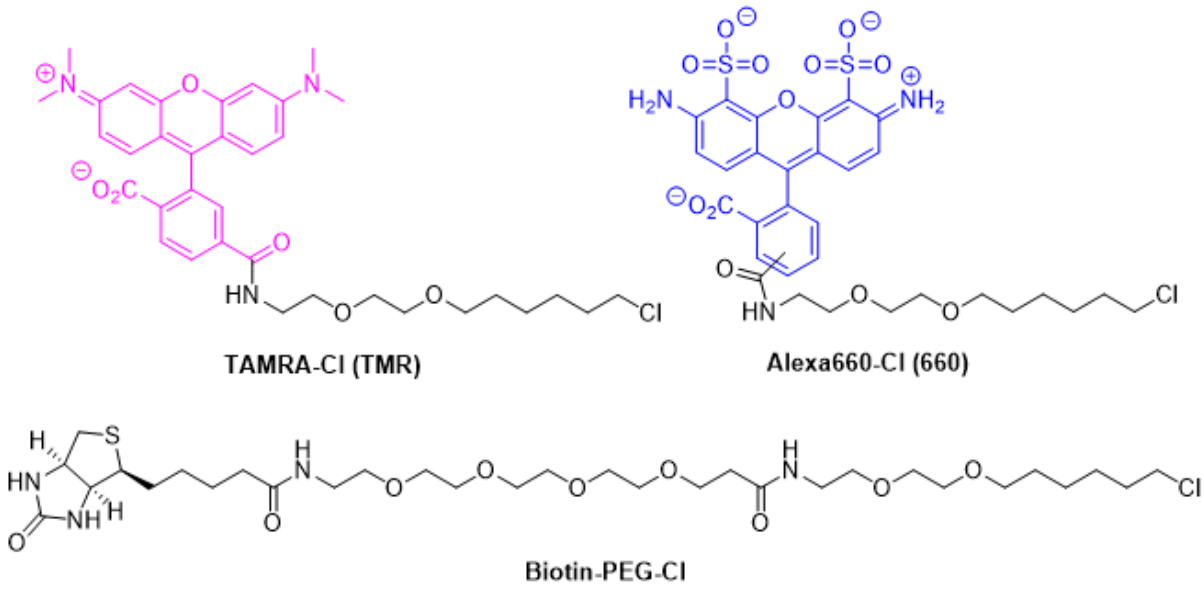


Figure 56: The range of HaloTag ligands used in this study. **TAMRA-Cl** (**TMR**) contains a Tamra fluorophore conjugated to a chloroalkane and is cell permeable. **Alexa660-Cl** (**660**) contains a charged AlexaFluor660 fluorophore conjugated to a chloroalkane linker and is unable to permeate the cell membrane. **Biotin-PEG-Cl** selectively biotinylates HaloTag proteins at the cell surface, or all HaloTag proteins in cell lysate.

The variety of HaloTag ligands available, both commercially and through chemical synthesis, facilitates the labelling of the protein construct under a range of conditions.

HEK293 cells transfected with HaloTag-GPR119 or HaloTag-FFA2 as a positive control were labelled either in lysate or in culture with two different ligands. **TAMRA-Cl** (TMR) contains a Tamra fluorescent reporter group coupled *via* amide bond to a PEG alkyl chloro linker, to specifically label HaloTag (**Figure 56**). The ligand can cross the cell membrane and is thus assigned the property of a "permeable" HaloTag ligand. Conversely, **Alexa660-Cl** is a HaloTag ligand designed to probe the cell surface expression of HaloTag protein due to the charged nature of the AlexaFluor dye used in this ligand. Both ligands have suitable excitation and emission spectra to enable the fluorophores to be used and visualised simultaneously, without bleed-through.

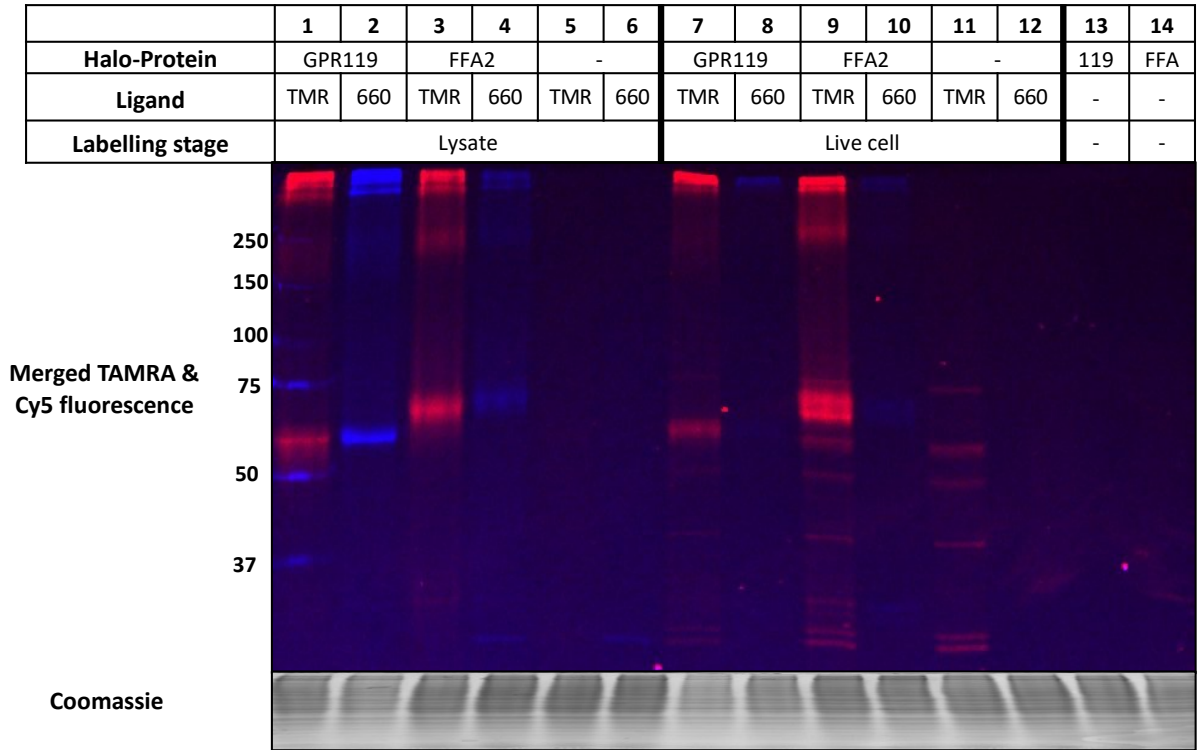


Figure 57: Merged in-gel fluorescence data showing the successful labelling of HaloTag-GPR119 (and HaloTag-FFA2 as a positive control) both in lysates and in live cells in culture. Red fluroescence image represents labelling with TMR-Cl ligand, blue fluorescence image represents labelling with 660-Cl ligand. Negative controls of wild-type HEK293 cells **(lanes 5, 6, 11 and 12)** and cells transfected but not labelled **(13 and 14)** are shown. Coomassie staining of the gel is used as a loading control.

Under these conditions, HaloTag-GPR119 was successfully labelled in lysate with both ligands, showing excellent selectivity towards HaloTag (**Figure 57**). Labelling cells in culture was also possible with both ligands, however less specificity was observed with the TMR-Cl ligand, with multiple bands being observed in the transfected cells and wild type cells (as seen in lanes 7, 9 and 11). Promisingly, HaloTag-GPR119 was also labelled in culture using Alexa660-Cl, visualising the

expression of the protein correctly orientated within the membrane of HEK293 cells (fluorescence image for each channel independently is available in appendix **S1**).

To further explore the increased utility of the HaloTag-GPR119 protein construct, a selection of lysis conditions and methods were screened. Classically, GPCRs are highly sensitive to lysis methods (as discussed earlier), often forming multimers or insoluble aggregates in the absence of suitable detergents. A range of detergents were screened: SDS, Triton X-100, NP40 and DDM, as well as detergent free conditions. Sonication was also carried out using a microtipped sonicator probe for several samples to establish whether a mechanical sheering of the membrane could solubilise a greater proportion of the protein.

The results of this study are incredibly interesting given the known troubles with GPCRs. As a standard lysis buffer for a GPCR, 1% DDM, 10% glycerol, 150 mM NaCl in 50 mM HEPES (condition **E** in **Figure 58,** lane **5**) gave the benchmark for the proportion of soluble to aggregated HaloTag-GPR119. In both the in-gel fluorescence data and western blot, a large band represented aggregated protein can be seen at the front of the resolving gel. With the addition of sonication to this lysis method, there is a slight improvement to the proportion of recovered soluble protein at around 60 kDa (lane 6). Similar levels of soluble and insoluble protein are also observed with other commonly used lysis buffers (both containing non-ionic detergents): 1% Triton X-100, 150 mM NaCl in 50 mM HEPES (condition **B** in **Figure 58**) and 1% NP40, 150 mM NaCl in 50 mM HEPES.

	1	2	3	4	5	6	7	8	9	10	11
Lysis buffer	A	B	C	D	E	E	F	F	D	E	F
Sonication	-					+	-	+	+	-	
Cells	HaloTag-GPR119								WT		

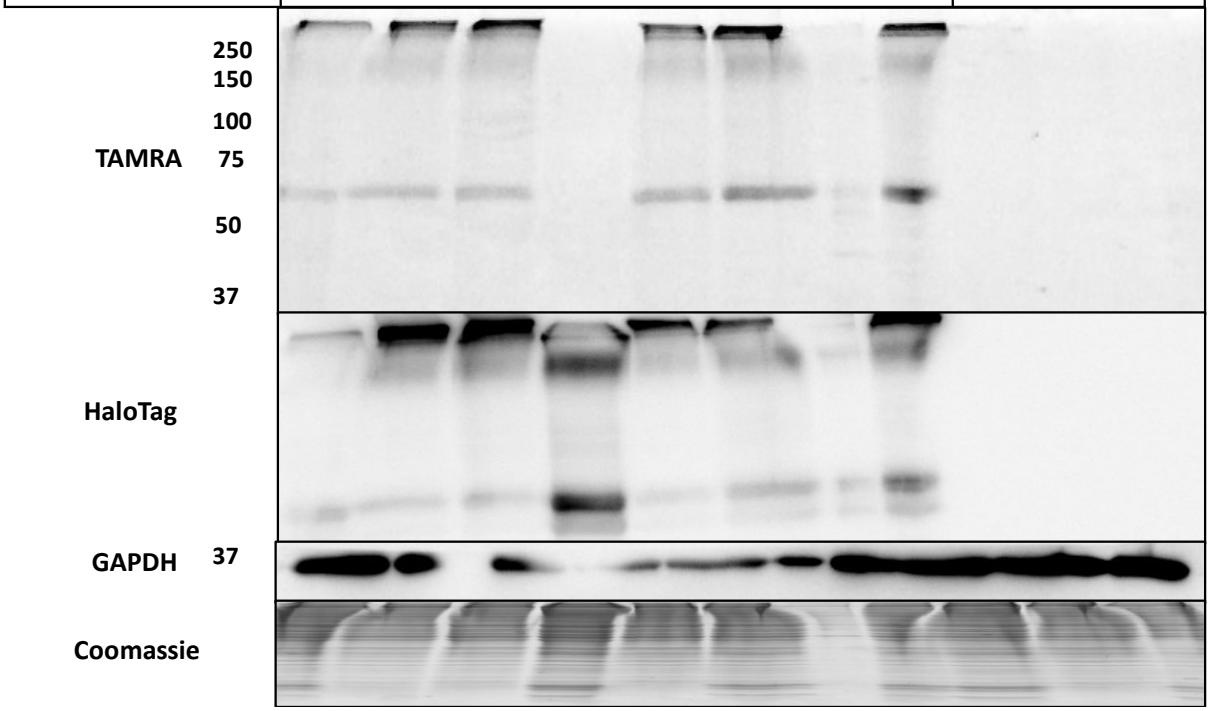


Figure 58: Screening of lysis conditions in HaloTag-GPR119 transfected cells and wild type cells. Lysis conditions: **A** 0.1% SDS, 1% Triton X-100, PBS; **B** 1% Triton, 150 mM NaCl, 50 mM HEPES; **C** 1% NP40, 150 mM NaCl, 50 mM HEPES; **D** 1% SDS, 150 mM NaCl, 50 mM HEPES; **E** 1% DDM, 10% glycerol, 150 mM NaCl, 50 mM HEPES; **F** PBS. Sonication of cells was also included as a method of lysis to determine whether an enhanced proportion of soluble protein could be obtained. Protein solubility and retention of structure was determined by 1. in-gel fluorescence to confirm correctly folded HaloTag-GPR119 protein, 2. HaloTag western blot to determine total levels of HaloTag-GPR119 solublised, 3. GAPDH and Coomassie as an indicator of total soluble protien and as a loading control.

Switching the detergent for a stronger, ionic detergent such as SDS led to harsh denaturing conditions for the HaloTag-GPR119 protein. Using 0.1% SDS, 1% Triton X-100 in PBS (condition **A** in **Figure 58**) resulted in a reduction in fluorescent labelling but also a reduction in protein aggregate. When the concentration of SDS was increased to 1% SDS, 150 mM NaCl in 50 mM HEPES (condition **D** in **Figure 58**) labelling of HaloTag-GPR119 was unable to occur due to denaturation of the protein, however there was a vast increase in the proportion of protein solubilised, with trace amounts of aggregate observed when compared to the other conditions. Perhaps the most interesting result, with greatest potential utility, was the comparison of cells lysed in PBS alone with and without sonication (condition **F** in **Figure 58**). In the absence of sonication, labelled protein is not observed in the in-gel fluorescence, with very low level observed in the HaloTag western blot, however with sonication good labelling

and protein solubilisation is observed. This suggests that if the lysis method allows the GPCR to remain within the cell-membrane or membrane mimicking detergents, the protein will remain correctly folded without a significant degree of aggregation. This is particularly useful for applications that may require the removal of detergent.

ID	1	2	3	4	5	6	7	8	9	10	11
Lysis buffer	A	B	C	D	E	E	F	F	D	E	F
Sonication	-	-	-	-	-	+	-	+	+	-	-
Cells	HaloTag-GPR119								WT		

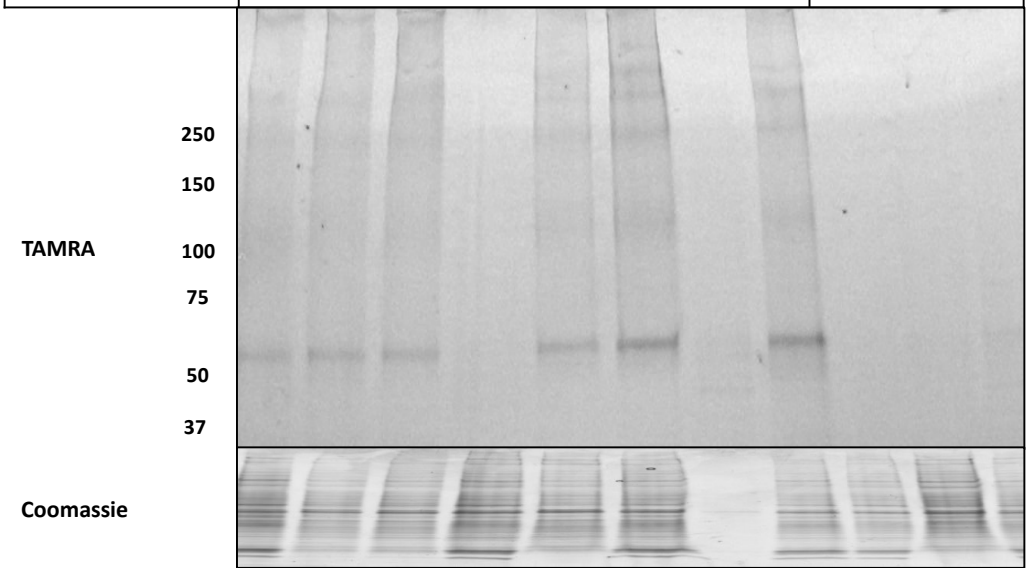


Figure 59: Resolution of aggregate bands using a 4-20% gradient gel previously observed when using a 4% stacking with 12% resolving gel. Fluoresence imaging shows the majority of HaloGPR119 is present as the monomer, but with evidence of multimers at higher molecular weight.

In order to further analyse the aggregate band observed at the gel front, samples were run on a 4-20% gradient gel allowing the resolution of the multimers from the monomer, validating that the large majority of protein in solution exists as the monomer using in-gel fluorescence (**Figure 59**).

3.5.2. Evaluation Using Mass Spectrometry

Alongside fluorescent tagging of GPR119, the addition of HaloTag allows a vast array of functionalisation, particularly for enriching GPR119. Using commercially available reagents, HaloTag-GPR119 can be enriched by functionalisation of the protein with a biotin handle and subsequent incubation with neutravidin beads, or directly onto resin functionalised with chloroalkane handles (HaloTag Resin). Both methods offer advantages and disadvantages: using biotin is a two-step process with labelling and enrichment, requiring both steps to be efficient to get a satisfactory enrichment however it does allow recovery of intact protein due to the nature of the non-covalent interaction. Conversely, enriching directly onto HaloTag resin is more likely to produce a higher yield of protein, however the covalent bond formed between the resin and the protein means the tag is destroyed as GPR119 can only be recovered from the resin via proteolysis of the protein (either at the TEV cleavage site, or by digesting the protein into peptides). Biotinylation has the additional benefit that labelling can occur under a range of conditions as it is not the terminal step in the enrichment processes, allowing interrogation of the protein under a variety of treatments or pathways.

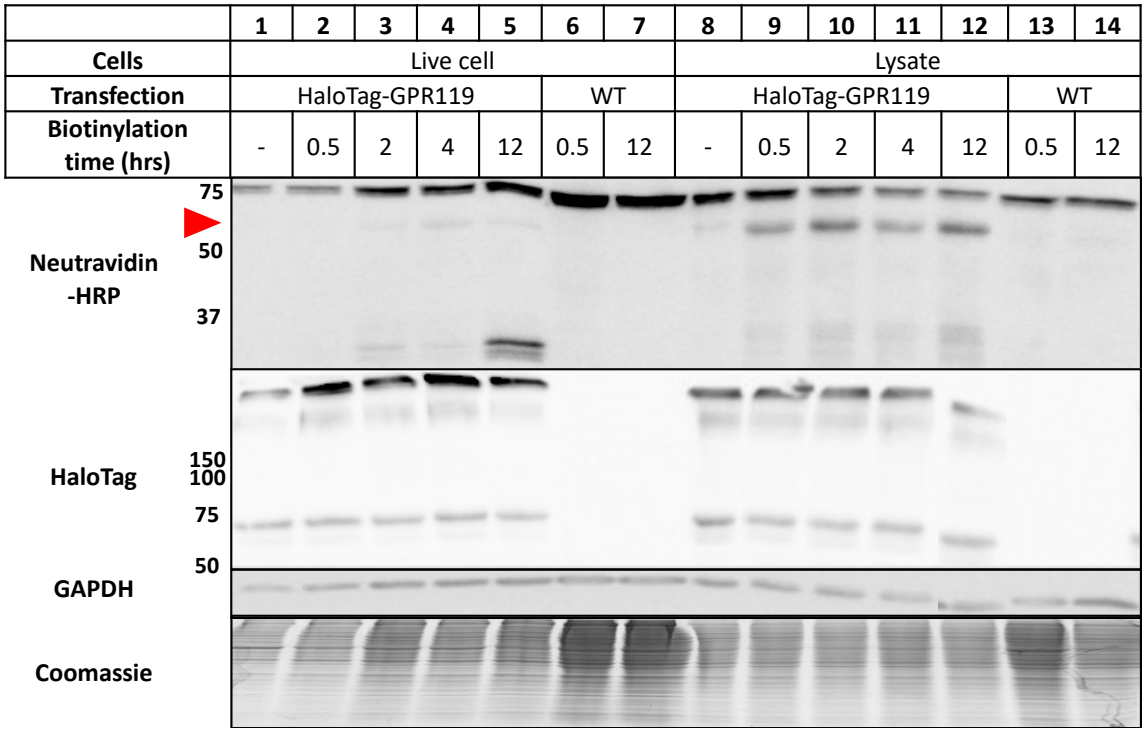


Figure 60: Optimisation of biotinlyation condition using Biotin-PEG-Cl HaloTag ligand. Biotin ligand was incubated in cells in culture, and in cell lysate. Labelling was analysed over a time course using a neutravidin-HRP chemical probe. Lysate labelling was more efficent and specific. Red arrow denotes HaloTag-GPR119 expected molecular weight.

To establish the labelling potential with biotin, transfected cells were labelled at various time points within cells in culture and in cell lysate with Biotin-PEG-Cl (**Figure 60**). Successful labelling was identified using a neutravidin-HRP conjugate. Labelling in cells in culture was apparent after a 2-hour labelling period, however after extending labelling an unexpected lower molecular weight band appeared (lanes **3**, **4** and **5**), potentially the result of a cleavage product or degradation the protein in cells under stress. This band was not present in wild type cells incubated with the labelling reagent for the same amount of time. The labelling reaction in cell lysate appeared to go to completion within half an hour of incubation, as the intensity of labelling did not increase further over a 12-hour period and was more efficient than labelling cells in culture as evident by the increased relative density of the HaloTag-GPR119 band in the neutravidin-HRP immunoblot. From these results, labelling in lysate for 30 minutes was taken forward as the conditions to trial for enrichment of the protein.

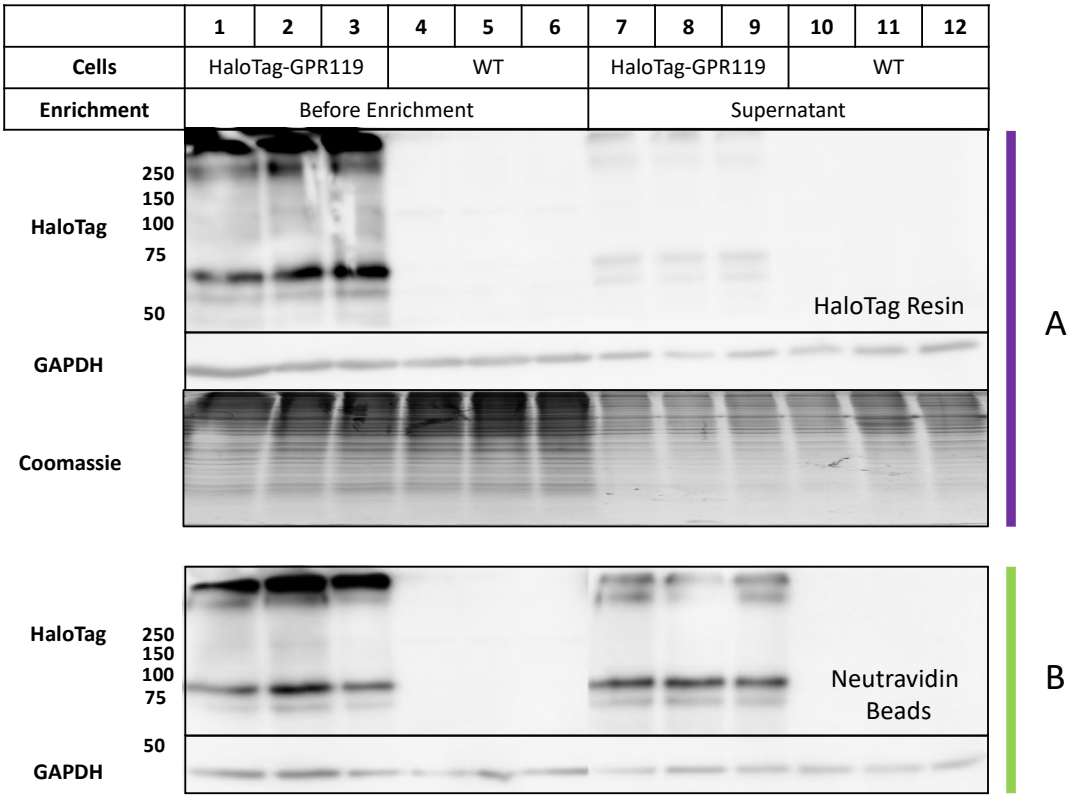


Figure 61: Enrichment of HaloTag-GPR119 from cell lysate using **A** HaloTag Resin or **B** labelling with Biotin-PEG-Cl and subsequent enrichment on Neutravidin beads. Comparison of the presence of HaloTag-GPR119 in the lysate before enrichment and the supernatent after enrichment shows direct enrichment using HaloTag resin is the more efficent approach.

HaloTag-GPR119 containing cell lysates were prepared and either enriched directly onto HaloLink resin (a bead coated in a haloalkane linker to specifically capture HaloTag proteins) or labelled with Biotin-PEG-Cl and enriched on neutravidin resin.

The samples were prepared in triplicate and enriched samples processed for proteomic using on bead trypsin digest supplemented with ProteaseMax, stage-tipped and resuspended before injection on the Q-Exactive.

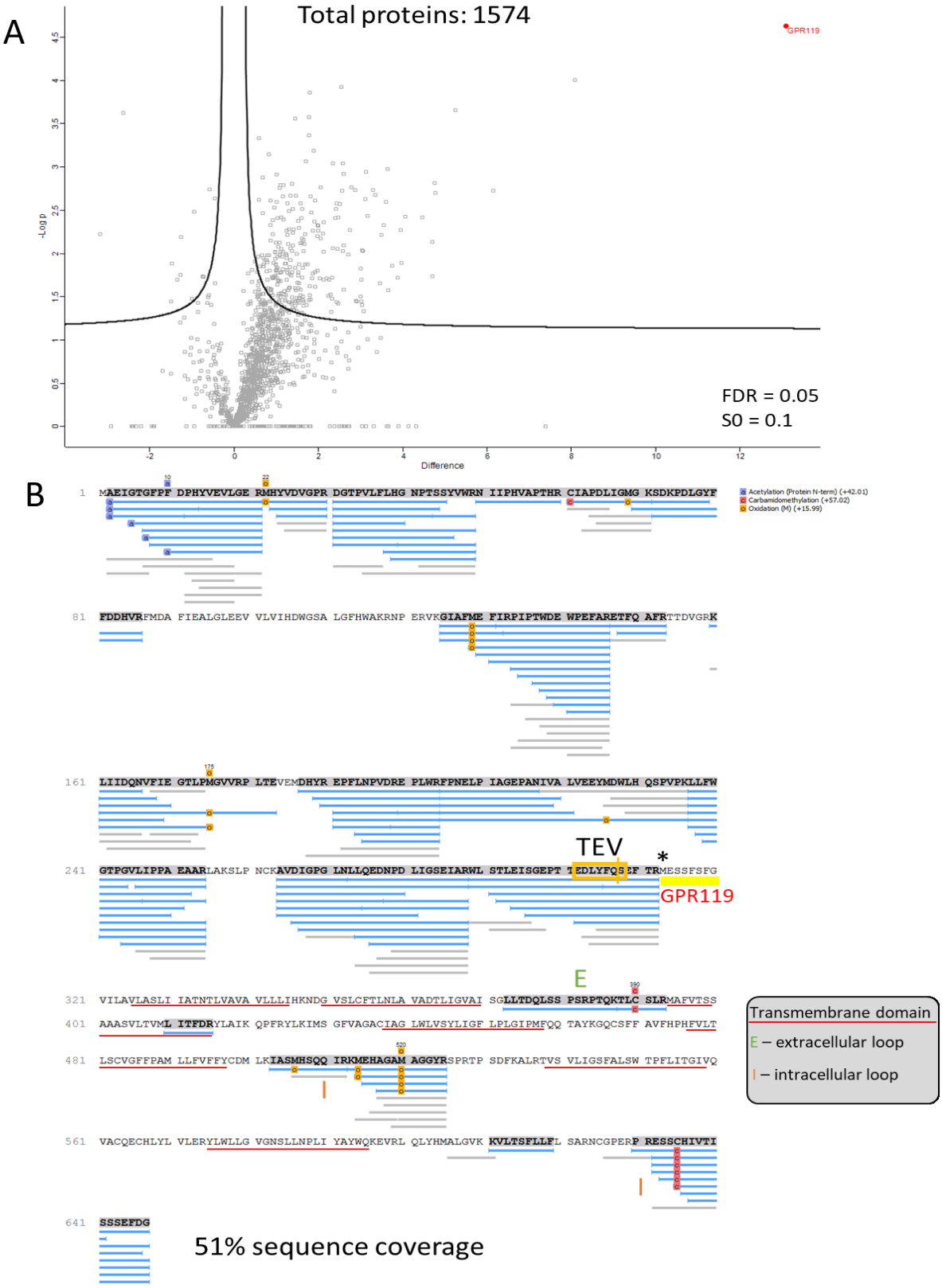


Figure 62: **A** A volcano plot representing the level of significant enrichment of all the proteins identified from HaloTag-GPR119 lysate enriched onto HaloLink Resin. Y-axis represents level of significance (-logP value) and x-axis represents difference in enrichment between control (wild type) and experimental samples. HaloTag-GPR119 was identified and displayed on the plot as the red dot. **B** For the same experiment, the peptides identified are shown has highlighted in grey. The start of GPR119 protein sequence(*) is underlined in yellow. Transmembrane domains are underlined in red, and peptides identified from GPR119 sequence have been annotated as either extracellular (E) or intracellular (I). For HaloTag-GPR119, 51% sequence coverage was seen.

The results were processed using MaxQuant and Perseus to identify proteins significantly enriched. From the HaloLink enriched samples, 1574 proteins were identified but GPR119 was the most significantly enriched protein compared to wild type cells (**Figure 62**). This result presents the first time GPR119 has been successfully identified using a proteomic approach, and selectively enriched from a whole cell lysate. A similar result was obtained with samples prepared from biotin-neutravidin enrichment (data not shown).

The data files were also processed with PEAKS using the inbuilt algorithms to build a picture of the protein landscape with *de novo* peptide sequencing. Using the software to specifically look at HaloTag-GPR119, the peptides fragmented and identified with the sample preparation method used can be analysed. In **Figure 62**, 51% sequence coverage was obtained of HaloTag-GPR119. Perhaps unsurprisingly, most of the peptides identified were located in the HaloTag domain of the fusion protein, however peptides were found across intracellular, extracellular and even transmembrane domains of GPR119, with the HaloTag “amplifying” the proteomic signal.

In order to increase the coverage of GPR119 peptides, it was hypothesised that optimising the proteolysis step may lead to the identification of more hydrophobic peptides. As trypsin selectively digests at arginine and lysine, hydrophobic areas of the protein will be neglected. Chymotrypsin preferentially cleaves at aromatic side chains, such as tryptophan, tyrosine and phenylalanine leaving a different subset of peptides behind. Elastase (Ala, Val, Ser, Gly, Leu and Ile cleavage) and Proteinase K (broad digestion – aliphatic and aromatic amino acids) present additional tools to digest a range of peptide bonds.

While time restraints meant a full study with optimisation was not possible, the initial test presented some promising results. Based on the manufacturer’s digestion conditions (**Table 3**), trypsin, chymotrypsin, elastase and proteinase K were applied to HaloTag-GPR119 enriched lysate on resin (and supplemented with 0.05% ProteaseMax) as follows:

Protease	Buffer	Concentration	Time	Temperature (°C)
Trypsin	50 mM Tris.HCl pH 8	1 µg/100 µg protein	Overnight	37
Chymotrypsin	50 mM Tris.HCl pH 8	1 µg/75 µg protein	Overnight	25
Elastase	50 mM Tris.HCl pH 9	1 µg/50 µg protein	Overnight	37
Proteinase K	50 mM Tris.HCl pH 8	1 µg/100 µg protein	3 hours	37

Table 3: Digestion conditions initially attempted for a protease screen. Overnight represents approximately 16 hours of digestion. All samples were left shaking for the stated duration.

Collating the results for each protease against the peptides identified (**Table 4**), there is a large degree of complementarity within the results: peptides identified within the tryptic digest are not seen with the other proteases, and likewise for the other samples. Of interest, the number of peptides identified with chymotrypsin is equal to that of trypsin, suggesting a stepwise, or simultaneous, double digestion may yield greater coverage of GPR119. However, this experiment requires further optimisation before conclusions can be made. The experiment yielded fewer peptides than the previous experiment conducted with trypsin that could be a result of protein or peptide loss on sample preparation or loss of yield somewhere within the sample preparation workflow. Preparing fresh samples and further optimisation of all protease methods are required to provide direction for future proteomic sample preparation of GPCR containing mixtures, however, this initial study provides a positive direction for future work.

Sequence	Domain	Length	Mass (Da)	Trypsin			Chymotrypsin			Elastase			Proteinase K		
AEIGTGFPFDPHYVEVLGER	HaloTag	20	2232.0851	1	2	2									
AVDIGPGLNLLQEDNPDLIGSEIAR	HaloTag	25	2618.3552	2		2									
CIAPDLIGMGK	HaloTag	11	1173.5886	2											
EISGEPTTEDLY	HaloTag	12	1352.5984				1	1	1						
EISGEPTTEDLYFQSEF	HaloTag	17	1990.8684				1	1	1						
ESSCHIVTISSSEFDG	GPR119	16	1753.7465	1	1	1									
GEPTTEDLYFQS	HaloTag	12	1385.5987							1		1	1	1	
IAPDLIGM	HaloTag	8	828.44153											1	1
IGVAISGLLTDQL	GPR119	13	1298.7446					1	1						
IIDQNVF	HaloTag	7	847.44397					1	1						
IIDQNVFIEGTLPM	HaloTag	14	1588.8171				2	1	3						
KMEHAGAMAGGYR	GPR119	13	1377.6282	1											
LHGNPTSSY	HaloTag	9	974.44576				1	1	1						
LSTLEISGEPTTEDLY	HaloTag	16	1766.8462				1	1	1						
MEHAGAMAGGYR	GPR119	12	1249.5332	2											
MHYVDVGPR	HaloTag	9	1072.5124	3											
NIIPHVAPTHR	HaloTag	11	1253.6993	1											
SGEPTTEDLYFQS	HaloTag	13	1472.6307							1		1			
STLEISGEPTTEDLY	HaloTag	15	1653.7621				1	1	1						
VLIPPAEAA	HaloTag	9	879.50657										1	1	1
VTSSAAASVL	GPR119	10	904.48656				1	1	1						
WLSTLEISGEPTTEDLYFQSEFTR	HaloTag	24	2848.3443			2									

Table 4: Peptides identified using proteomics from HaloTag-GPR119 samples enriched on HaloLink Resin and digested with either trypsin, chymotrypsin, elastase or proteinase K. Each condition was performed in triplicate. For each peptide identified, the number identified in each sample is given in the table. In this experiment, chymotrypsin gives the greatest reproducible coverage of complementary peptides.

3.6. Conclusions and Future Directions

Throughout this work, and previously published work, the inability to reliably identify GPR119 in endogenously expressing tissue as well as engineered overexpressing tissue limits the exploration of the protein. Western blot and proteomics failed to identify GPR119 when overexpressed in HEK293 cells, though cAMP activity assays confirmed expression of functional protein. The use of fusion proteins and epitope tags is a common method to work around these issues for hard to identify proteins in the literature. By synthesising a library of plasmids containing various tagged forms of GPR119, an attempt to optimise the expression and identification of GPR119 was made.

There is literature precedent for using FLAG-GPR119 and this route was explored in the first instance. Using a gifted plasmid previously used in the literature[298], identification of GPR119 was not successful using either transfection or viral transduction of the gene. However, using PCR to determine the mRNA levels within the samples prove that the genetic material had been successfully incorporated into the genome. Expression of the tagged protein was then investigated using protein function as a marker, by comparing the cAMP response of wild type and transfected cells (both GPR119 and FLAG-GPR119) with a range of GPR119 agonists. Surprisingly, the response from FLAG-GPR119 was greater than that obtained with the previous tool cell line overexpressing GPR119, suggesting the protein was expressed but the FLAG epitope tag was not ideal for detection.

Given the hydrophobic nature of GPCRs, they are prone to aggregation to hide hydrophobic surfaces in solution. The FLAG-tag peptide sequence was added to the extracellular N-terminal of the GPR119 sequence, however, the extracellular N-terminal domain of GPR119 is only 12 residues long and with the addition of the small FLAG-tag, this sequence may not be accessible when in solution or even while in the membrane if the peptide sequence has any interaction with the proximal extracellular loops. Thus, a larger, more hydrophilic and functional protein modification was made.

HaloTag is a mutated protein that covalently binds haloalkanes in a selective manner, allowing a range of protein modifications to be made. Protein engineering to append HaloTag to the N-terminal of GPR119 with a spacer resulted in the first repeatable

detection of the GPCR using in-gel fluorescence and western blot. The functionality of HaloTag subsequently allowed the enrichment of GPR119 from a lysate mixture using direct enrichment onto HaloLink resin, and indirect enrichment by labelling the fusion protein with a biotin handle and then enriching onto neutravidin beads. Proteomic analysis of these samples showed significant enrichment of HaloTag-GPR119 specifically from the lysate mixture. Initial results show that utilising a double digest approach with complementary protease may lead to increased coverage of GPR119 and thus improved sensitivity.

While time dictated the scope of work carried out with HaloTag-GPR119, there are several avenues that would be interesting to pursue in the future. Firstly, continuation of the protease optimisation may lead to enhanced sequence coverage and increase the proportion of GPR119 identified. Secondly, the variation between proteomic runs and number of peptides identified may be a result of suboptimal sample preparation, but it could also be an indicator of variability in protein expression from each transfection experiment. Towards the end of this project, initial work was carried out on the design and synthesis of a plasmid for the generation of a HaloTag-GPR119 Flp-In T-REx inducible stable cell line. The Flp-In system allows targeted homogenous integration of the gene of interest into a specific locus of each cell, giving much greater results than transfection and antibiotic based stabilisation, which has been used with success in the literature.[305–307]

4. Evaluation of chemical probes for GPR119

4.1. Introduction

The synthesis and biochemical evaluation of a series of chemical probes for GPR119 has been described in chapter 2 of this thesis. In this chapter, the labelling ability of a selection of promising probes (**Figure 63**) in cell lines overexpressing GPR119 is reported. The chemical probes were selected based upon their agonist properties as assayed in a cAMP assay, as well as including a range of control probes lacking a photocrosslinking element but retaining a clickable component.

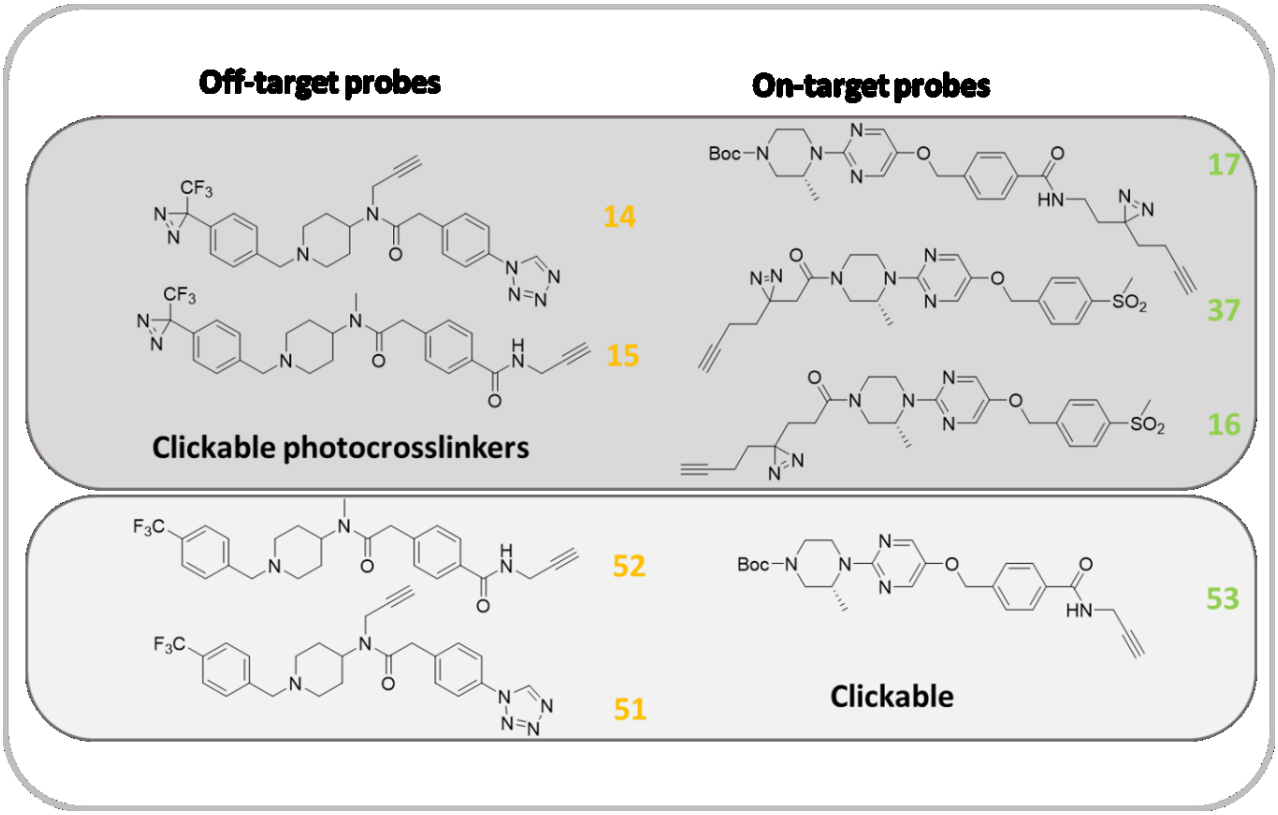


Figure 63: The library of chemical probes for GPR119 initially tested in cell lysates

4.2. Validation of Photocrosslinking in cells

4.2.1. Chemical Probes Taken Forward

The labelling capacity of each chemical probe was tested in HEK293-GPR119 cells by incubating at 10 uM for 1 hour. Cells were then either irradiated with UV light for 10 minutes or left on ice in the dark before being lysed. The cell lysates were labelled

using click chemistry with AzT (TAMRA conjugated azide click reagent) and the resulting labelled proteins analysed by in-gel fluorescence.

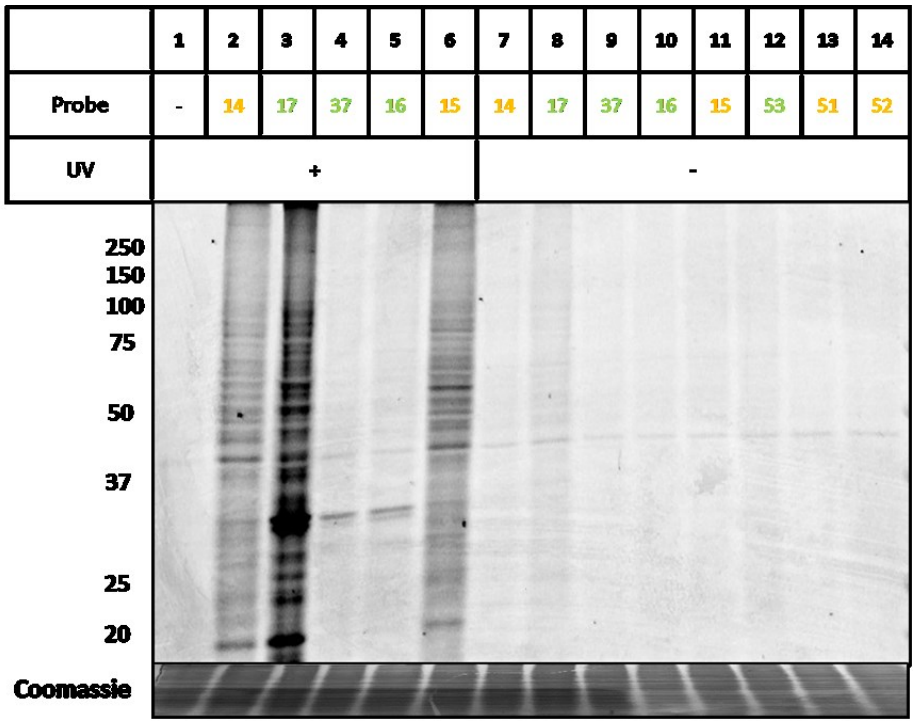


Figure 64: Probe labelling patterns at 10 μM with and without UV irradiation.

Without UV irradiation, all lanes display the same distinct pattern of background non-specific fluorescence as a by-product of the click labelling reaction. With UV, the chemical probes display vastly different labelling patterns. Probes **14** and **15** possess the pharmacophore of the "off-target" AstraZeneca compound **AZ001** and both show comparable labelling patterns, with additional bands present in the sample labelled with probe **15**. Interestingly, probe **15** possessed some agonist quality in the cAMP assay, whereas probe **14** was inactive thus this additional labelling is promising. Probes **16** and **37** again possess a very similar structure, with probe **16** containing an additional methylene spacer in the diazirine handle and appear to produce the same labelling pattern: very little background labelling and a strong band around 30 kDa. Probe **17** contains the photoaffinity group at the opposite end of the pharmacophore and at 10 μM produces a heavy labelling pattern. These characteristics will be explored in more depth in the following sections.

4.2.2. Optimisation & Evaluation of Labelling

The variables to photoaffinity labelling consist of concentration of probe and time of irradiation. Specific probe labelling can be evaluated by using a competitor compound,

in this case, the original template pharmacophore **AZ001**, **AZ002** or **AZ003** depending on the probe. Increasing the concentration of competitor with fixed concentration of probe should give a concentration dependant decrease of band intensity for protein targets that bind both the original compound and the chemical probe, showing the varied probe profiles.

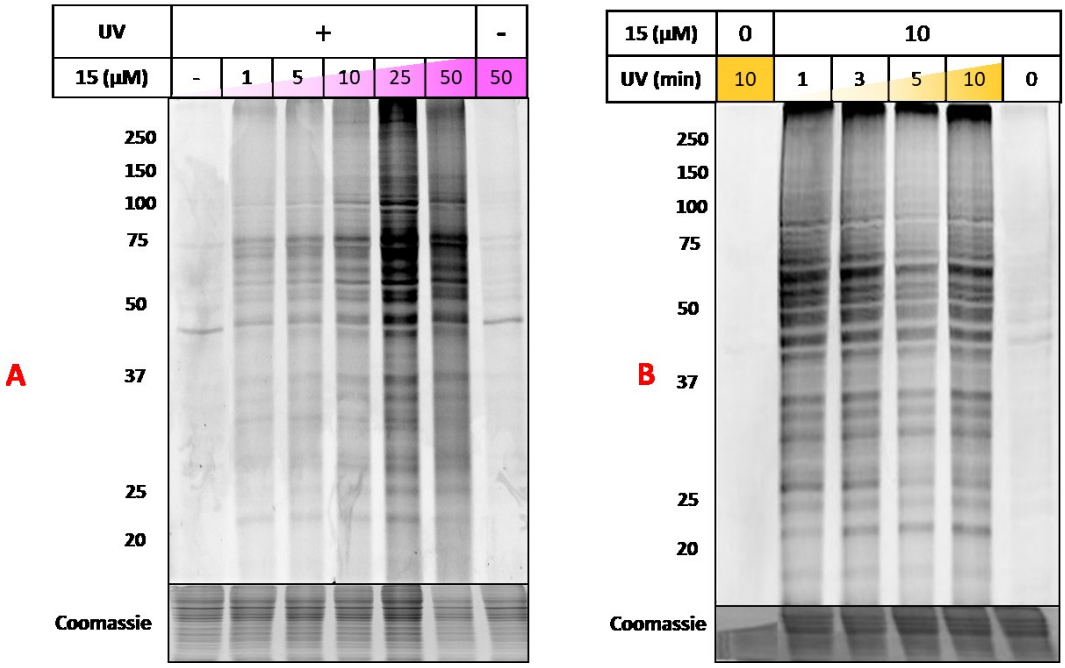


Figure 66: Labelling profile of off-target probe **15**. **A** concentration gradient in the presence and absence of UV irradiation. **B** Irradiation time of 1 minute provides equivalent labelling to longer irradiation times.

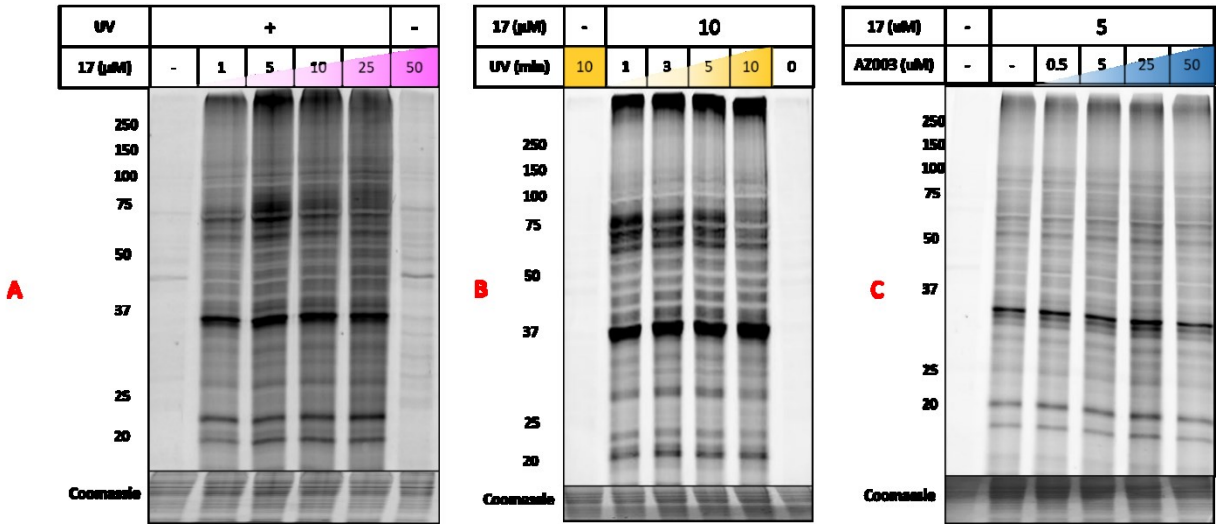


Figure 65: Labelling profile of on-target probe **17**. **A** Strong labelling with probe observed at 1 µM of probe. **B** Irradiation time of 1 minute is sufficient for labelling. **C** Competition with parent compound **AZ003** does not reveal any visible decreases in band intensity suggesting substantial background labelling.

Off-target chemical probe **15** showed a strong concentration dependant labelling pattern, with heavy labelling at 50 µM and retaining the distinct labelling pattern down to 1µM. Controls without chemical probe, and without UV irradiation show no additional strong labelling (**Figure 66**, **A**). Faint bands can be seen in the sample with 50 µM probe and no UV irradiation, and this may be due to exposure to ambient levels of UV light during the sample preparation, as lysates were not protected from ambient light. At 10 µM, an irradiation time of 1 minute provides as much labelling as 10 minutes exposure, emphasising the fast rate of diazirine crosslinking reactions (**Figure 66**, **B**).

Interestingly, on-target probes **16** and **17** showed vastly different labelling patterns. In **Figure 65**, probe **17** gives strong labelling even at 1 µM and shows little concentration dependence. This could be due to either high levels of non-specific labelling or the potency of the chemical probe being sub micromolar in terms of labelling capacity. Similarly, competition of the probe with **AZ003** up to 50 µM (10 times the concentration of probe) does not show any decrease in band intensity, suggesting high levels of background labelling are masking any specific labelling.

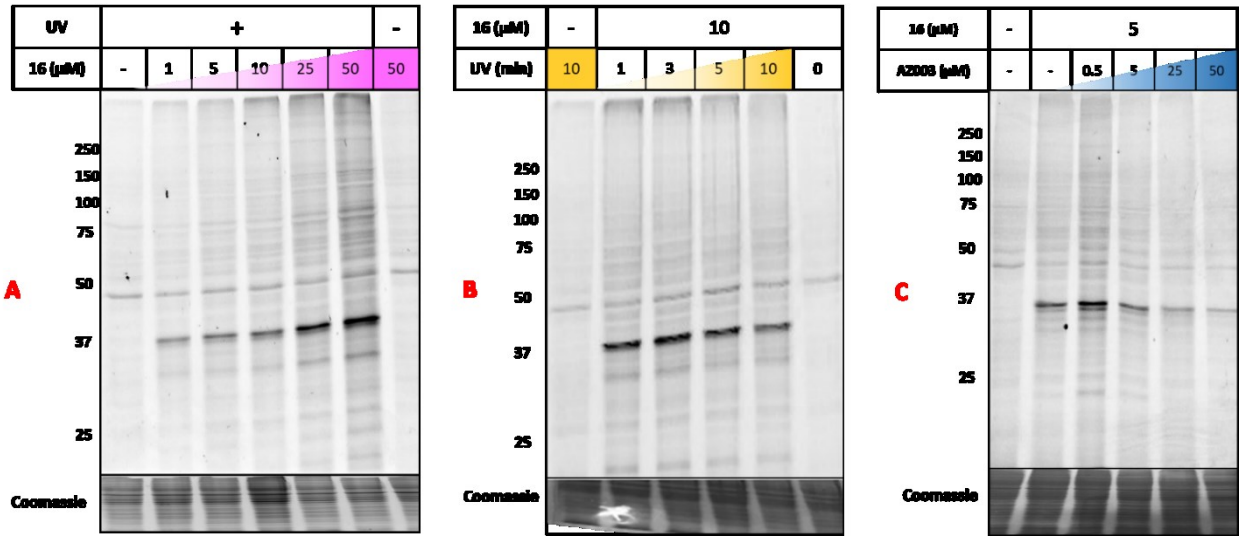


Figure 67: Labelling pattern for on-target probe **16**. **A** Concentration dependant labelling is observed, however there are considerably fewer background bands than probe **17**. **B** UV irradiation of 1 minute is sufficient to label all bands observed after 10 minutes irradiation. **C** Competition with parent compound **AZ003** does appear to reduce the intensity of probe specific bands around 37 kDa.

However, probe **16** shows little non-specific labelling in comparison. A concentration gradient of the probe gives a strong response with bands at 37 kDa and 30 kDa increasing in intensity along the gradient (**Figure 67**). Similarly, when the probe labelling is subjected to competition with **AZ003** a decrease in band intensity is observed with the primary band at 37 kDa. What's more, the secondary band at 30

kDa is not observable at the highest concentration of competitor. Both probes exhibit the same fast labelling after one minute of UV irradiation.

As GPR119 resides within the cell membrane, membranes were isolated from the cell lysate and the level of membrane-specific labelling was assessed (**Figure 68**). Compared to the previous experiments using the whole cell lysate, the labelling patterns differed. What's more, probe **16** that previously had low background labelling, appears to label proteins within the cell membrane more prolifically. Again, the labelling patterns for all three classes of probe appears distinct and the ability to isolate labelled cell membrane proteins is promising for further development of these methods.

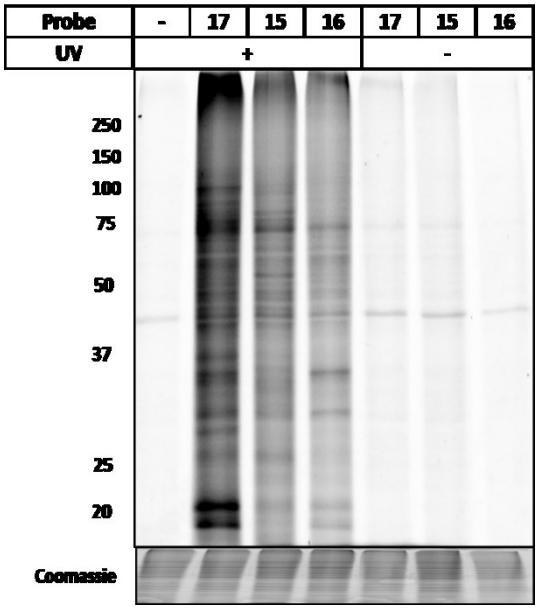


Figure 68: Probe labelling (10 μM) patterns on the membrane fraction of cell lysate.

4.3. Photocrosslinking and Enrichment of Halo-GPR119

With the developed HEK293 HaloTag-GPR119 cells (described in chapter 3) in hand, the protein labelling capacity of the chemical probes could be determined with a marker for GPR119. The additional functionality of the HaloTag fusion protein enables GPR119 to be fluorescently labelled with a complimentary fluorophore to the click reagent. This allows the samples to be dual labelled both with a marker for GPR119 expression and successful probe labelling, and any overlap in fluorescence bands may be indicative of probe-labelled GPR119. For the experiments described in this section, on-target chemical probes **16** and **17** were used in order to attempt labelling of GPR119.

4.3.1. Labelling in HEK293 HaloTag-GPR119 cells

Following incubation of probes **16** and **17** with the transfected cell line, the lysates were subjected to dual labelling of the HaloTag and the probe alkyne handle utilising their biorthogonal chemistries. Probe labelling was imaged with AzTB (TAMRA labelling denoted in pink in **Figure 69**) while HaloTag-GPR119 labelling was imaged with HaloAlexa660 (Cy5 labelling shown as blue in **Figure 69**). Concentration gradients of both probes were once again assessed by in-gel fluorescence but with the presence of GPR119 co-labelled. Despite the clear GPR119 band, the presence of many bands in the probe labelling hinders identification of any overlap, but fluorescence does appear in similar regions.

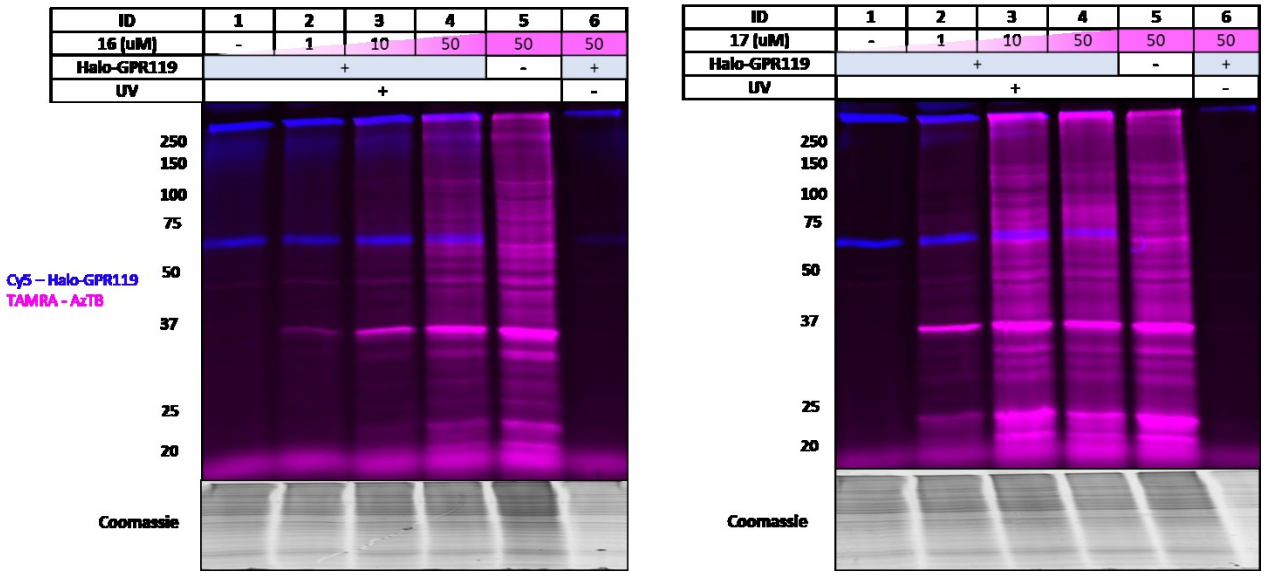


Figure 69: Probe **16** and **17** incubated with HEK293 HaloTag-GPR119 cells. Dual fluorophore labelling of HaloTag-GPR119 (blue) and the chemical probe (pink) show potential regions of co-labelling. See **S2** for each channel independently.

4.3.2. Enrichment of Halo-GPR119 with a Chemical Probe

To evaluate whether GPR119 was successfully labelled by either chemical probe samples were clicked with AzTB (an azide capture reagent containing a TAMRA fluorophore and biotin handle) and probe-labelled proteins were enriched onto streptavidin beads. The protein population obtained was compared to the complete

(before pull-down) sample, as well as the supernatant of the enrichment using in-gel fluorescence and subsequent western blot.

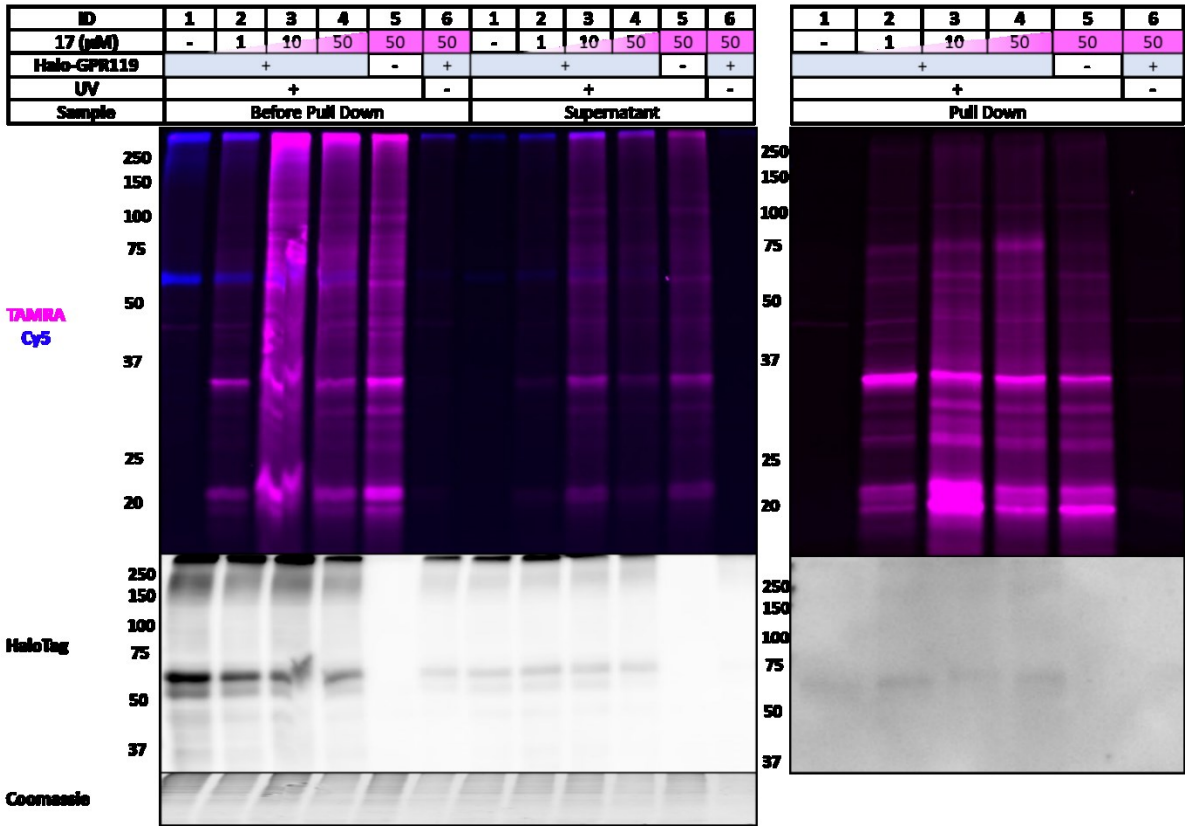


Figure 70: Probe **17** incubation in cells followed by labelling the lysates with AzTB and HaloAlexa660. A band in the HaloTag western blot does appear at the correct molecular weight in the probe containing lanes. See **S3** for each channel independently.

As before, a concentration dependant labelling of cell lysate with probe **17** is observed alongside HaloTag-GPR119 expression (**Figure 70**). Enrichment of the probe-labelled proteins appears successful with a diminished population of proteins being labelled in the supernatant lanes (as assessed by eye comparing Coomassie protein loading and in-gel fluorescence), and significant in-gel fluorescence observed in the pull-down lanes. Similarly, diminished HaloTag-GPR119 is observed in the supernatant lanes, however, a corresponding band is not visible in the pull-down lanes using in-gel fluorescence.

For increased sensitivity, HaloTag-GPR119 enrichment was evaluated using western blot taking advantage of the HaloTag protein for antibody recognition. HaloTag-GPR119 protein is visible across the before pull-down samples and the supernatant (as well as HaloTag-GPR119 aggregates at higher molecular weights). Within the pull-down samples, there are bands visible in probe containing lanes treated with UV, suggesting HaloTag-GPR119 is enriched using chemical probe **17**. However, poor

band resolution on the western blot suggests the quantity of protein enriched is minimal and verifying this enrichment as a true hit requires further validation.

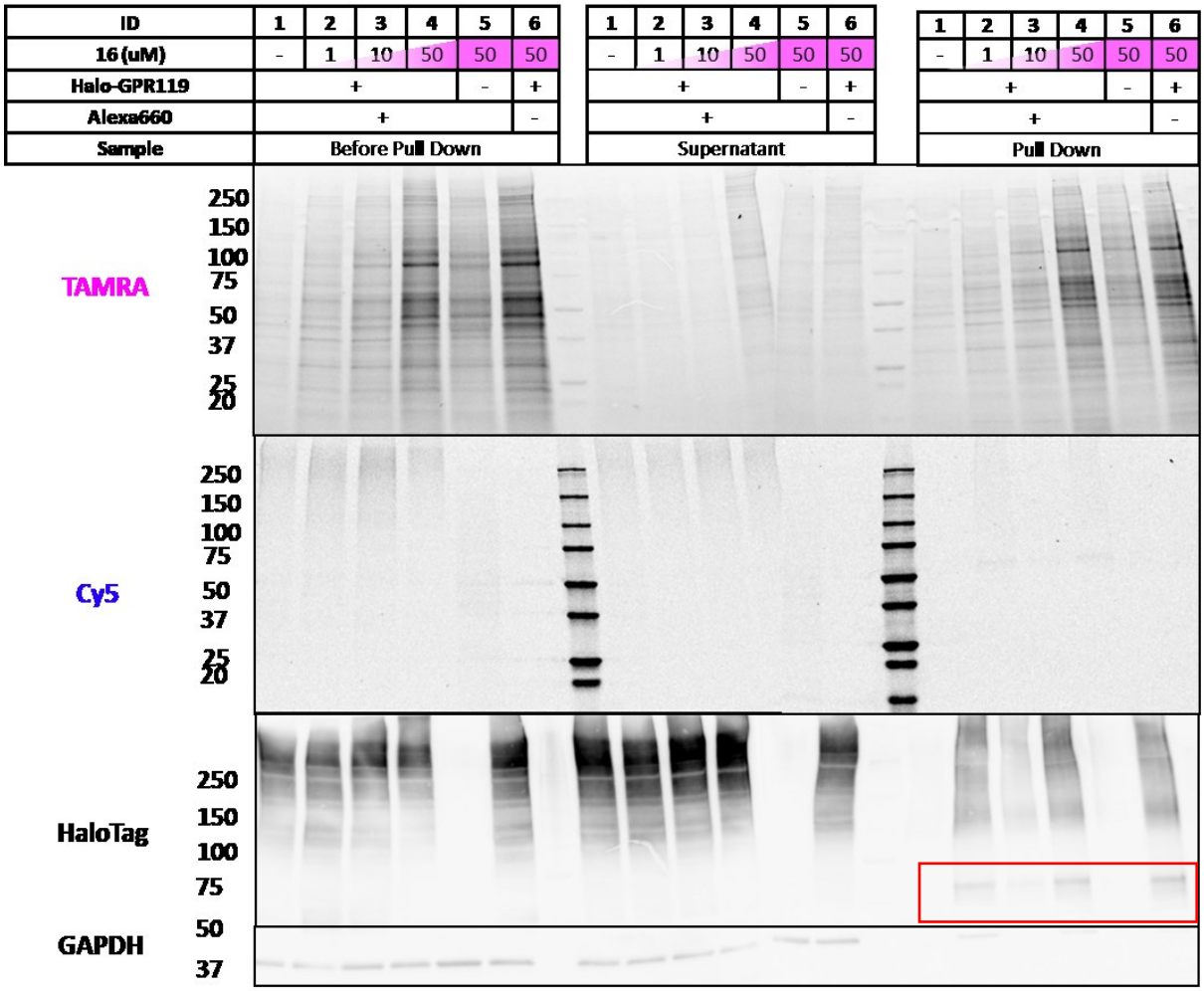


Figure 71: HaloTag-GPR119 HEK293 cell lysates were incubated with probe **16**, cross linked and ligated with AzTB. HaloTag-GPR119 was labelled with Alexa660 Cy5 fluorophore. Before pull-down samples, the supernatant and bead enriched samples were evaluated using in-gel fluorescence and western blot using a HaloTag antibody and GAPDH as a loading control. Enrichment of HaloTag-GPR119 is observed in the expected lanes (as indicated by the red box).

With probe **16**, labelling, cross linking and ligation was carried out in cell lysates before enrichment was evaluated in a similar way. In **Figure 71** the TAMRA and Cy5 fluorescence images have been shown individually rather than overlaid, due to the weak Cy5 fluorescent signal, to enable both parameters to be evaluated. A comparable pattern of enrichment is observed with probe **16** as was seen with probe **17**. For HaloTag-GPR119, the signal appears weaker due to the appearance of higher molecular weight multimers, however, in the enriched pull-down lane, a clear band is observed for HaloTag-GPR119 using in-gel fluorescence in lanes representing increasing concentration of chemical probe, whereas no signal is observed in control lanes. As before, the enrichment of HaloTag-GPR119 was further analysed using

HaloTag-GPR119 western blot. The signal with probe **16** is significantly stronger than that observed with probe **17**. Enrichment of HaloTag-GPR119 monomer is clear in lanes representing 1 μM, 10 μM, 50 μM and 50 μM probe without HaloTag cy5 labelling, suggesting that the enrichment is dependent on probe **16** and presence of HaloTag-GPR119. This significant labelling requires further validation to show the labelling is indeed a specific interaction.

4.3.3. Competition with the Parent Compound

To establish whether labelling with probe **16** is specific, competition experiments with parent compound **AZ003** were conducted. Protein lysate was pre-incubated with **AZ003** for one hour before probe **16** was added to the lysate. Thus, **AZ003** should block the specific proteins from subsequent chemical probe labelling, giving a

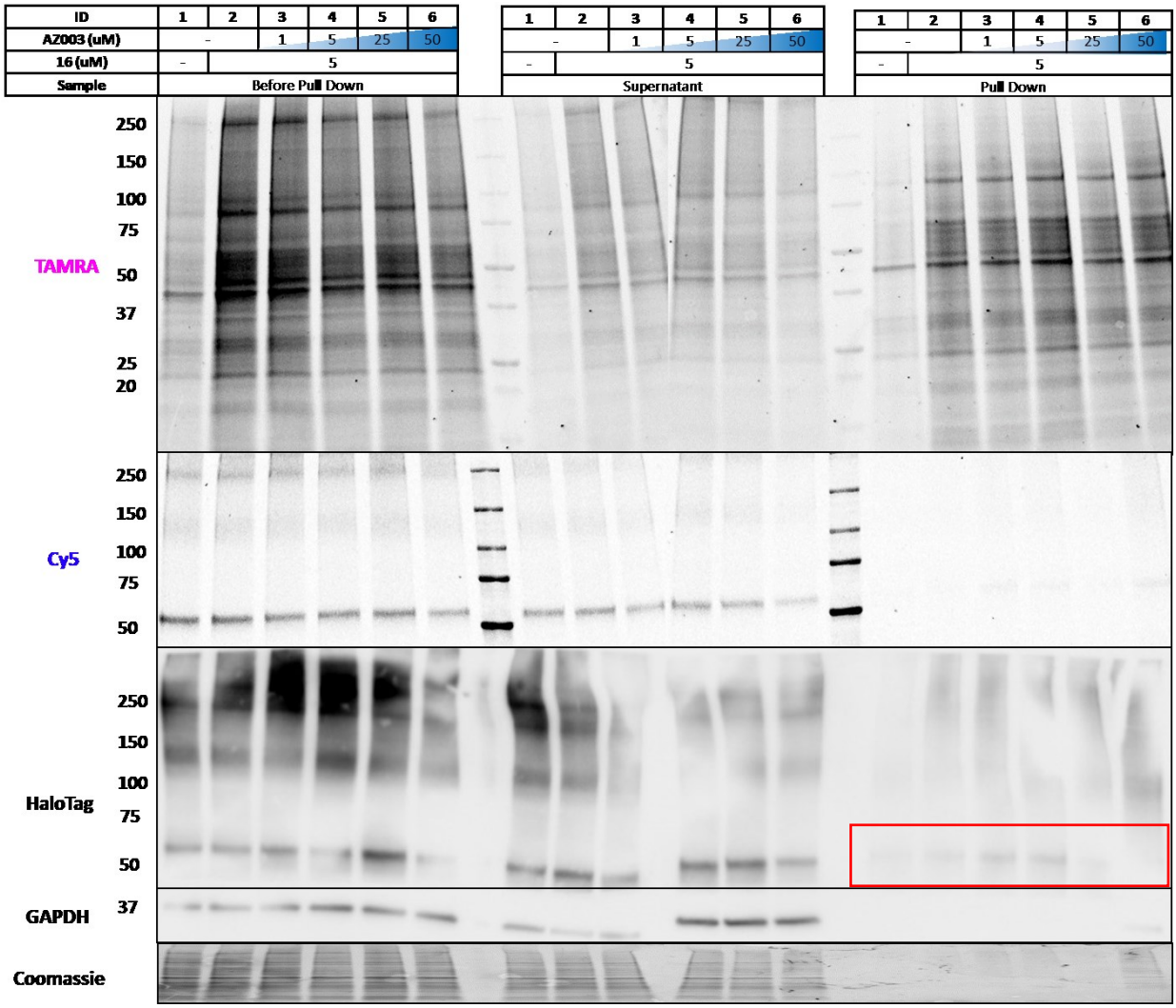


Figure 72: Competition of probe **16** labelling with the parent compound **AZ003**. Pre-incubation with **AZ003** blocks specific sites of interaction. Within the pull-down samples, a concentration dependent decrease in labelling is observed by HaloTag western blot as the concentration of competitor increases (as indicated by the red box), indicative of specific probe labelling.

concentration dependent labelling and enrichment profile. As before, most sensitivity is seen with the HaloTag western blot.

In the pull-down lanes, a band for HaloTag-GPR119 is seen at the expected molecular weight. As the concentration of competitor increases, the density of the band decreases relative to the band before until 50 μM where no band is apparent at 50 kDa. This concentration dependant band is reassuringly suggestive of specific photocrosslinking of probe **16** with HaloTag-GPR119.

4.4. Conclusions

This chapter presents the development work around the chemical biological evaluation of chemical probes previously synthesised. Through optimisation of labelling conditions, a series of chemical probes were shown to have varied labelling patterns in cells overexpressing GPR119. Utilising the biological tools developed in an earlier chapter, a more thorough investigation into the labelling capacity of the chemical probes towards GPR119 could be carried out indirectly by utilising the HaloTag fusion protein.

Enrichment of the probe-labelled proteins appears to show HaloTag-GPR119 being captured. Further assessment of the interaction by competition experiments with parent compound **AZ003** suggests that probe **16** may show HaloTag-GPR119 specific labelling. This represents the first reported chemical probe for GPR119 and potentially provides an alternative to antibodies for detection of the protein.

While time did not allow further experimentation with these chemical probes, significant follow up, optimisation and validation of the chemical probes is still required. Initially, the competition experiments with probes **16** and **17** should be repeated to confirm the result. Translating the experiment to a proteomic readout would provide vast amounts of quantitative data. With the optimisation of the HaloTag-GPR119 proteomic experiments described in chapter 3, a conclusive read out of HaloTag-GPR119 labelling could be achieved. What's more, the enrichment of the lysate using HaloLink resin may provide enough protein to carry out peptide level analysis to identify probe modified peptides as evidence of probe labelling, and to identify the binding site of the probes.

5. Conclusions and Future Outlooks

The aims of this thesis were to design and evaluate photocrosslinking probes to help understand the interactions of agonists for GPR119 and aid the identification and chemical biological evaluation of GPR119. The design, synthesis and biochemical evaluation of the photoprobes is described in Chapter 2. While cell lines overexpressing GPR119 are available, there are limited tools to validate expression levels of the protein. Chapter 3 discusses the development of fusion proteins for GPR119 to enable detectable cell line expression of the protein using fluorescence, immunoblotting and proteomics. Finally, the photocrosslinking probes were used in photocrosslinking experiments in cell lines overexpressing GPR119 and HaloTag-GPR119 to determine whether these probes do interact with and enrich GPR119, as described in Chapter 4. A more detailed conclusion for each chapter is detailed below, alongside suggestions for future work.

5.1. Probe Design, Synthesis and Evaluation

The synthesis of a selection of chemical probes for GPR119 has been described based on pharmacophores generated by an AstraZeneca drug discovery campaign. Two different classes of pharmacophores were selected: one demonstrating GPR119 mediated blood glucose lowering, and the other with GPR119 agonist activity but blood glucose lowering by an unknown mechanism of action. Chemical probes were designed to incorporate diazirine and alkyne chemical handles within the pharmacophores. Given the planar nature of the pharmacophores, the structures were deconstructed and reassembled to include the chemical probe moieties as either an aromatic diazirine or an aliphatic handle containing both the diazirine and alkyne functional groups at various points on the structure. The synthesis of diazirine alkyne containing linkers has been published,[213,216,217] however, within this thesis the route was further optimised to improve the yield and purification of the intermediates.

Potency of the chemical probes were evaluated in an assay for cAMP as an indicator of GPR119 agonism. The assay kit was optimised to allow sampling of stimulated lysates to provide more repeatable results with the scope for further assay iterations by dilution of stimulated lysate. The validity of the assay was shown by obtaining EC_{50}

values for AstraZeneca agonists that were in the same order of magnitude and order of potency as published previously.[201,202,204] The chemical probes were subjected to the assay, and revealed probe **17** to have the greatest potency for GPR119 with an EC_{50} of 392 nM. The results correlated to previous literature observations that GPR119 agonists tend towards an agonist architecture with little branching off the main pharmacophore. [101,111]

Overall, the development of these chemical probes was challenging due to the difficulties encountered synthesising diazirine building blocks in a great enough yield for further synthesis. However, a library of probes was successfully synthesised and isolated. Moreover, the methods of assessing GPR119 receptor agonism required much optimisation. Despite repeated attempts, significant variation in assay results, as a result of GPR119 expression in cells and the relative insolubility of the chemical probes in the buffers required, led to the requirement to alter the assay. The optimisation of the assay to enable cells in culture to be stimulated and the lysates used for analysis allowed the lysates to be diluted to account for changes in activity on different passages of the cell line. By stimulating the cells in culture, the chemical probes experienced greater levels of solubilisation at high concentrations due to the inclusion of BSA as a hydrophobic carrier in the optimised stimulation buffer. Future work may be to validate these results in an orthogonal binding assay, examples of which have been described in chapter 2.

5.2. Development of Biological Constructs for GPR119

The inability to reliably identify GPR119 in endogenously expressing tissue and engineered overexpressing tissue limits the exploration of the protein. Western blot analysis and proteomic experiments failed to identify GPR119 when overexpressed in HEK293 cells, though cAMP activity assays confirmed expression of functional protein. A library of plasmids containing various tagged forms of GPR119 and fusion proteins was synthesised to find the optimal expression construct.

There is literature precedent for using FLAG-GPR119[298], but transfections with the published plasmid did not lead to identification of GPR119. Attempts at viral transduction of the FLAG-GPR119 did not produce FLAG signal by immunoblot or immunofluorescence. However, PCR confirmed that the genetic material had been

successfully incorporated into the genome. Expression of the tagged protein was then confirmed using protein function as a marker by comparing the cAMP response of wild type and FLAG-GPR119 transfected HEK293 cells with a range of GPR119 agonists. These results suggest that the protein was expressed but the FLAG epitope tag was not ideal for detection.

Protein engineering to append a larger fusion protein, HaloTag, to the N-terminal of GPR119 with a spacer resulted in the first repeatable detection of the GPCR using in-gel fluorescence and western blot. The biorthogonal labelling of HaloTag subsequently allowed the enrichment of GPR119 from a lysate mixture using direct enrichment onto HaloLink resin, and indirect enrichment by labelling the fusion protein with a biotin handle onto neutravidin beads. Proteomic analysis of these samples showed significant enrichment of HaloTag-GPR119 specifically from the lysate mixture. Initial results show that utilising a double digest approach with complementary protease may lead to increased coverage of GPR119 and thus improved sensitivity.

HaloTag-GPR119 represents a significant step within this thesis to understanding GPR119 and there are several avenues that would be interesting to pursue in the future. Firstly, continuation of the protease optimisation may lead to enhanced sequence coverage and increase the proportion of GPR119 identified using mass spectrometry. Using trypsin, sequence coverage of around 50% was achieved for HaloTag-GPR119. However, most of the peptides identified originated from the HaloTag fusion protein. Optimisation of the digestion and sample preparation may lead to the increased coverage of the GPR119 portion of the protein, ultimately leading to a method for identifying wild type GPR119.

Secondly, the variation between proteomic runs and number of peptides identified may be a result of suboptimal sample preparation, but it could also be an indicator of variability in protein expression from each transfection experiment. Towards the end of this project, initial work was carried out on the design and synthesis of a plasmid for the generation of a HaloTag-GPR119 Flp-In T-REx inducible stable cell line. The Flp-In system allows targeted homogenous integration of the gene of interest into a specific locus of each cell, giving much greater results than transfection and antibiotic based stabilisation, and has been used with success in the literature.[305–307]

5.3. Chemical Biology Evaluation of Chemical Probes for GPR119

Through optimisation of labelling conditions, a series of chemical probes were shown to have varied labelling patterns in cells overexpressing GPR119. Utilising the biological tools developed in this thesis, a more thorough investigation into the labelling capacity of the chemical probes towards GPR119 was carried out indirectly by utilising the HaloTag fusion protein. Enrichment of the probe-labelled proteins appears to show HaloTag-GPR119 being captured. Further assessment of the interaction by competition experiments with parent compound **AZ003** suggests that probe **16** shows HaloTag-GPR119 labelling. This represents the first reported chemical probe for GPR119 and potentially provides an alternative to antibodies for detection of the protein.

Significant follow up, optimisation and validation of the chemical probes is still required. The competition experiments with probes **16** and **17** should be repeated to confirm the result. Translating the experiment to a proteomic readout with the optimised HaloTag-GPR119 MS conditions could provide vast amounts of quantitative data. What's more, the enrichment of the lysate using HaloLink resin may provide enough protein to carry out peptide level analysis to identify probe modified peptides as evidence of probe labelling, and to identify the binding site of the probes.

As the HaloTag functionality allows dual labelling of the probe labelled samples, it would be interesting to explore the use of flow cytometry to determine the cell populations expressing HaloTag-GPR119 and those also labelled with the GPR119 chemical probes. A nice example by Soethoudt *et al*. of this technique exists in the literature for chemical probes derived from a cannabinoid receptor 1 (CB1) agonist. The authors demonstrate changes in cell populations with the introduction of the chemical probe and illustrates displacement of the populations with pre-treatment of an agonist.[186] By employing flow cytometry whole cells could be routinely used instead of lysates, thus negating issues with solubilisation and aggregation.

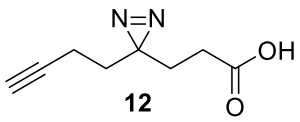
Additionally, fluorescence may be exploited using microscopy techniques to verify that HaloTag-GPR119 expression is observed at the membrane of proteins, and to identify any co-localisation of the chemical probes at the membrane. This method was used by Müskens *et al.* to demonstrate probe-specific labelling at the cell membrane of the neurokinin1 (NK_1) receptor, alongside proteomics, in-gel fluorescence and western

blotting as complementary techniques.[29] GPCRs present a challenge for the development of chemical probes, but by using as many of the orthogonal methods of validation that are accessible as possible, more information can be obtained and a greater picture of the GPCR chemical probe landscape can be developed.

6. Materials and methods

6.1. Chemical synthesis

6.1.1. 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanoic acid[213] (12)

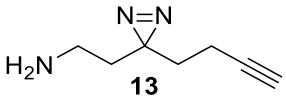


12

4-Oxooct-7-ynoic acid (240 mg, 1.57 mmol) was dissolved in 7N NH_3 in MeOH (20 mL) at 0 ºC and stirred under nitrogen for 3 hours. A solution of hydroxylamine-*O*-sulfonic acid (249 mg, 2.20 mmol) in anhydrous MeOH (5 mL) was added dropwise at 0 ºC and allowed to stir for an additional hour before being warmed to RT and stirred overnight. The suspension was filtered under vacuum, washed with MeOH (2 x 10 mL) and concentrated *in vacuo*. The residue was resuspended in anhydrous MeOH (15 mL) and cooled to 0 ºC. Diisopropylethylamine (0.61 mL) was added, followed by iodine portion-wise until a dark brown colour persisted for more than 30 minutes, indicating complete oxidation of the diaziridine. The solution was diluted with EtOAc (10 mL) and washed with 1M HCl (10 mL), sat. $Na_2S_2O_3$ (3 x 10 mL) and brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (CH_2Cl_2:MeOH 99:1 gradient to 96:4) afforded diazirine **12** as a brown oil (68 mg, 26%).[216]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.69 (t, *J* = 7.3 Hz, 2H), 1.85 (t, *J* = 7.8 Hz, 2H), 2.00 – 2.08 (m, 3H), 2.15 – 2.25 (m, 2H); *m/z* (ES-ToF$^-$) 165 ([M-H]$^-$); HRMS (ESI$^-$) $C_8H_9N_2O_2^-$ requires 165.0644, found 165.0659.

6.1.2. 2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)ethan-1-amine (13)[213]



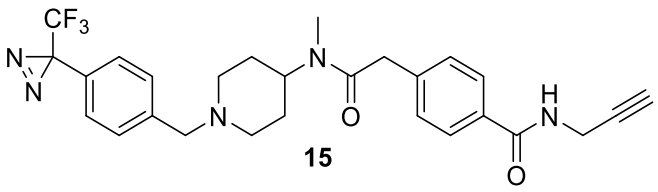
13

Azide **30** (25 mg, 0.15 mmol) was dissolved in a mixture of THF (1 mL) and H_2O (100 µL). Triphenylphosphine (44 mg, 0.17 mmol) was added and the reaction mixture was stirred at RT for 24 hours. The reaction mixture was diluted into Et_2O (10 mL) and washed with 1M HCl (10 mL). The organic layer was discarded and the aqueous layer was basified with 1M NaOH and extracted with Et_2O (2 x 10 mL). The combined

organic layers were washed with H_2O (10 mL) and brine (10 mL), dried and concentrated *in vacuo* to afford amine **13** as a yellow oil (10 mg, 50%).[213]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.48 – 1.78 (m, 6H), 1.97 – 2.07 (m, 3H), 2.43 – 2.65 (m, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.3, 27.0, 32.6, 36.1, 36.6, 69.2, 82.7; *m/z* unable to obtain mass spectra.

6.1.3. 4-(2-(methyl(1-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzyl)piperidin-4-yl)amino)-2-oxoethyl)-N-(prop-2-yn-1-yl)benzamide (15)

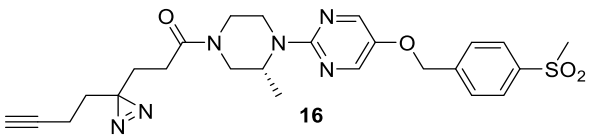


15

To a solution of **48** (35 mg, 0.085 mmol) dissolved in CH_2Cl_2 (2.5 mL) was added TFA (0.5 mL). The reaction mixture was stirred at RT until TLC indicated the reaction was complete. The solvent was removed *in vacuo* and crude mixture resuspended in MeCN (2 mL). DIPEA (96 µL, 0.55 mmol) and diazirine **11** (26 mg, 0.094 mmol) were added and the reaction mixture stirred at RT overnight. The solvent was removed *in vacuo* and the resulting residue resuspended in CH_2Cl_2 (10 mL) and washed with H_2O (2 x 10 mL) and brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (CH_2Cl_2:MeOH 95:5) afforded **15** as a yellow oil (11 mg, 25 %).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.20 – 2.20 (m, 8H), 2.22 – 2.39 (m, 1H), 2.85 (s, 3H), 3.38 – 3.64 (m, 2H), 3.70 – 3.88 (m, 2H), 4.27 (dd, *J* = 5.3, 2.8 Hz, 2H), 4.53 (s, 1H), 6.24 – 6.46 (m, 1H), 7.16 (d, *J* = 7.7 Hz, 2H), 7.30 – 7.47 (m, 4H), 7.66 – 7.90 (m, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 27.6, 29.8, 30.0, 41.5, 41.5, 51.0, 52.9, 62.2, 71.9, 79.5, 126.4, 127.4, 127.5, 128.9, 129.1, 129.3, 132.3, 139.5, 166.8, 169.9; *m/z* (ESI^+) 484 ($[M-N_2+H]^+$), 512 ($[M+H]^+$); HRMS (ESI^+) $C_{27}H_{29}N_5O_2F_3^+$ requires 512.2273, found 512.2278.

6.1.4. (R)-3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)-1-(3-methyl-4-(5-((4-(methylsulfonyl)benzyl)oxy)pyrimidin-2-yl)piperazin-1-yl)propan-1-one (16)



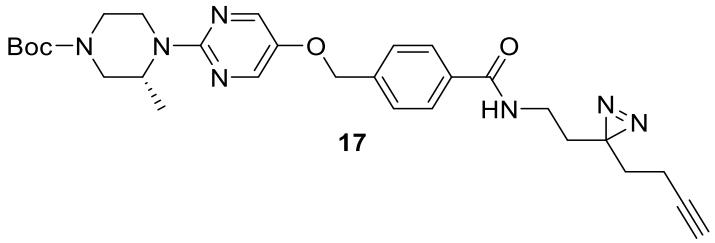
16

Acid **12** (5 mg, 0.033) was dissolved in THF (1 mL) and amine **35** (12 mg, 0.03 mmol) was added and the solution stirred for 10 minutes. DMTMM **36** (9 mg, 0.033 mmol) was added and stirring continued at RT overnight. The reaction mixture was diluted with H_2O (10 mL) and extracted into CH_2Cl_2 (2 x 10 mL). The combined organic layers were washed sequentially with sat. $NaHCO_3$ (10 mL), H_2O (10 mL), 1M HCl (10 mL), H_2O (10 mL), brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (eluent CH_2Cl_2:MeOH 99:1 to 98:2) afforded **16** as a yellow oil (10 mg, 66%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 0.99 – 1.52 (m, 4H), 1.60 (s, 3H), 1.71 (t, *J* = 7.5 Hz, 1H), 1.91 – 2.25 (m, 4H), 3.09 (s, 3H), 3.17 – 4.08 (m, 3H), 4.35 – 4.50 (m, 2H), 4.51 – 4.68 (m, 1H), 4.84 (s, 1H), 5.14 (s, 2H), 7.64 (d, *J* = 8.1 Hz, 2H), 8.00 (d, *J* = 8.2 Hz, 2H), 8.16 (s, 2H);* *m/z* (ESI$^+$) 511 ([M+H]$^+$); HRMS (ESI$^+$) $C_{25}H_{31}N_6O_4S^+$ requires 511.2128, found 511.2133.

*Unable to obtain ^{13}C NMR due to small quantity of material obtained and photosensitive nature of compound.

6.1.5. tert-butyl (R)-4-(5-((4-((2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)ethyl)carbamoyl)benzyl)oxy)pyrimidin-2-yl)-3-methylpiperazine-1-carboxylate (17)



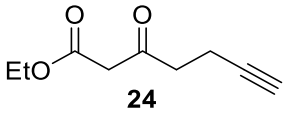
17

Acid **39** (28 mg, 0.065 mmol) and amine **13** (9.0 mg, 0.065 mmol) were stirred together in THF (1 mL) for 10 minutes. DMTMM **36** (22 mg, 0.080 mmol) was added and the solution stirred at RT overnight. The resulting mixture was diluted with H_2O (10 mL) and extracted into EtOAc (2 x 10 mL). The organic phase was washed sequentially with sat. $NaHCO_3$ (10 mL), H_2O (10 mL), 1M HCl (10 mL), H_2O (10 mL) and brine (10 mL). The organic layer was dried and concentrated *in vacuo*. Column chromatography (EtOAc:hexane 30:70 – 100:0) afforded diazirine **17** was a pale yellow solid (4 mg, 11%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.15 (d, *J* = 6.7 Hz, 3H), 1.48 (s, 9H), 1.69 (t, *J* = 7.2 Hz, 2H), 1.84 (t, *J* = 6.6 Hz, 2H), 1.99 (t, *J* = 2.7 Hz, 1H), 2.05 (td, *J* = 7.1, 2.6 Hz, 2H),

2.91 (s, 1H), 3.15 (td, *J* = 12.7, 3.7 Hz, 2H), 3.33 (q, *J* = 6.4 Hz, 2H), 3.93 (s, 1H), 4.13 (m, 1H), 4.31 (d, *J* = 13.1 Hz, 1H), 4.75 (s, 1H), 5.07 (s, 2H), 6.25 (t, *J* = 5.9 Hz, 1H), 7.44 – 7.52 (m, 2H), 7.77 – 7.85 (m, 2H), 8.12 (s, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.20, 14.21, 26.93, 28.45, 32.14, 32.48, 34.94, 38.70, 43.28, 47.07, 69.49, 71.60, 76.71, 77.02, 77.34, 83.52, 127.35, 127.60, 134.33, 140.02, 145.10, 146.36, 157.33, 167.02. One quaternary carbon was not observed.; *m/z* (ESI$^+$) 548 ([M+H]$^+$), 448 ([M-Boc+H]$^+$). HRMS (ESI$^+$) $C_{29}H_{38}N_7O_4$ requires 548.2985, found 548.2975.

6.1.6. Ethyl 3-oxo-6-heptynoate (24)[308]

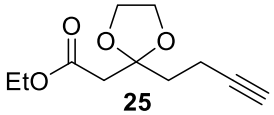


24

A solution LDA was generated by dropwise addition of *n*BuLi (2M in hexanes, 20.2 mL, 50.6 mmol) to a solution of diisopropylamine (7.10 mL, 50.6 mmol) in THF (50 mL) at -78^{o}C under an inert atmosphere. The solution was then allowed to warm to 0 ^{o}C. A solution of ethyl acetoacetate (3.22 mL, 25.3 mmol) in dry THF (20 mL) was added dropwise at 0 ^{o}C and the resulting solution was stirred for 30 minutes. Propargyl bromide (2.40 mL, 25.3 mmol) was added and the mixture stirred for 1 hour followed by the addition of acetic acid (2.90 mL, 50.6 mmol). The resulting solution was extracted into Et_2O (2 x 20 mL), washed with brine (2 x 10 mL), dried over Na_2SO_4 and concentrated in vacuo. The crude residue was purified by column chromatography (ethyl acetate:hexane 10:90) to afford **24** as a pale yellow oil (4.17 g, 98%).[308]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.31 (t, *J* = 7.1 Hz, 3H), 1.95 – 2.09 (m, 1H), 2.41 – 2.57 (m, 2H), 2.84 (t, *J* = 7.2 Hz, 2H), 3.49 (s, 2H), 4.23 (q, *J* = 7.3 Hz, 2H); *m/z* (TOF MS ES$^+$) 169 ([M+H]$^+$).

6.1.7. Ethyl 2-[2-(but-3-yn-1-yl)-1,3-dioxolan-2-yl]acetate (25)[308]



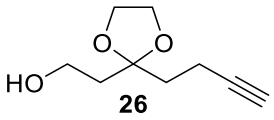
25

To Ketone **24** (4.17 g, 24.2 mmol) was added ethylene glycol (2.81 mL, 50.5 mmol), *p*-toluenesulphonic acid monohydrate (431 mg, 2.27 mmol) and triethylorthoformate (6.28 mL, 27.9 mmol). The resulting viscous solution was stirred at RT for 36 hours. The reaction mixture was diluted in Et_2O (10 mL) and washed with $NaHCO_3$ (10 mL),

H_2O (10 mL) and brine (10 mL), dried ($MgSO_4$) and concentrated *in vacuo* to give acetal **25** as a pale yellow oil (4.12 g, 77%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.26 (t, *J* = 7.2 Hz, 3H), 1.92 (t, *J* = 2.7 Hz, 1H), 2.07 – 2.16 (m, 2H), 2.30 (ddd, *J* = 8.7, 6.5, 2.7 Hz, 2H), 2.64 (s, 2H), 3.92 – 4.03 (m, 4H), 4.14 (q, *J* = 7.1 Hz, 2H); *m/z* (ESI$^+$) 213 ([M+H]$^+$).

6.1.8. 2-[2-(but-3-yn-1-yl)-1,3-dioxolan-2-yl]ethanol (26)[308]

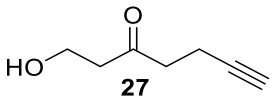


26

$LiAlH_4$ (515 mg, 13.6 mmol) was added to stirring solution of dry Et_2O (100 mL) under argon at 0^o C. Acetal **25** (2.40g, 11.3 mmol) was dissolved in a minimal amount of Et_2O and added slowly to the stirring suspension. The reaction mixture was stirred at RT for 30 minutes. To quench the reaction the following were added slowly with stirring at 0 ^{o}C: H_2O (515 µL), 15% NaOH (515 µL) and H_2O (1.54 µL). The solution was stirred for 30 minutes at RT, dried with $MgSO_4$, filtered and concentrated *in vacuo* to afford alcohol **26** as a clear, colourless oil which was used without further purification (1.95 g, quant.).[308]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.84 – 2.01 (m, 5H), 2.22 – 2.36 (m, 2H), 3.73 – 3.82 (m, 2H), 3.94 – 4.06 (m, 4H); *m/z* (TOF MS ES$^+$) 171 ([M+H]$^+$).

6.1.9. 1-hydroxyhept-6-yn-3-one (27)[213]

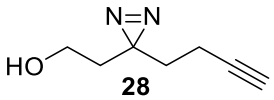


27

p-Toluenesulphonic acid monohydrate (515 mg, 2.95 mmol) was added alcohol **26** (1.90g, 11.2 mmol) in acetone (50 mL) and the resulting solution stirred for 4 hours. The reaction mixture was diluted with EtOAc (20 mL) and washed with $NaHCO_3$ (2 x 20 mL), water (20 mL) and brine (20 mL), dried and concentrated *in vacuo* to give **27** as a light yellow oil which was used without further purification (1.15 g, 82%).[213]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.98 (t, *J* = 2.7 Hz, 1H), 2.44 – 2.54 (m, 2H), 2.68 – 2.77 (m, 4H), 3.89 (t, *J* = 5.4 Hz, 2H); *m/z* (ESI$^+$) 127 ([M+H]$^+$).

6.1.10. 2-[3-(but-3-yn-1-yl)-3H-diazirin-3-yl]ethan-1-ol (28)[213]

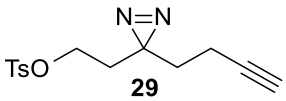


28

Ammonia (20 mL) was condensed into a 3-neck flask containing a stirrer bar with a dry ice condenser at -78 °C, under nitrogen. Ketone **27** (800 mg, 6.35 mmol) was dissolved in a minimal amount of dry MeOH (2 mL) and added to the ammonia and stirred for 5 hours at 0 °C. Hydroxylamine-*O*-sulfonic acid (932 mg, 8.25 mmol) was dissolved in dry MeOH (8 mL) and added to the reaction mixture over 30 minutes. The reaction was left stirring for 1 hour at -78 °C and the remaining ammonia was allowed to evaporate overnight. The resulting suspension was filtered, and the filter cake washed with MeOH, the filtrate was concentrated *in vacuo* and dissolved in CH_2Cl_2 (10 mL). Triethylamine (1.10 mL, 7.94 mmol) was added, and a solution of I_2 (2.31 g, 9.14 mmol) in CH_2Cl_2 (10 mL) was added slowly until a persistent brown colour remained. The mixture was washed with 1M HCl (10 mL), saturated sodium thiosulfate (2 x 10 mL) and brine (10 mL), before being concentrated in vacuo and purified via column chromatography (DCM:hexane 4:1 to DCM:diethyl ether 95:5) to give diazirine **28** as a yellow oil (420 mg, 48%).[213]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.64 – 1.83 (m, 4H), 2.03 (t, *J* = 2.6 Hz, 1H), 2.07 (td, *J* = 7.5, 2.7 Hz, 2H), 3.52 (t, *J* = 6.1 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.2, 26.6, 32.7, 35.5, 57.4, 69.2, 82.8; *m/z* (CI) 138 (M).

6.1.11. 2-[3-(but-3-yn-1-yl)-3H-diazirin-3-yl]ethyl 4-methylbenzene-1-sulfonate (29)

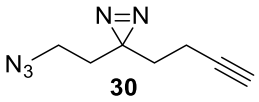


29

To a stirring solution of alcohol **28** (100 mg, 0.72mmol) in CH_2Cl_2 (2 mL) under nitrogen at 0 °C was added a solution of TsCl (205 mg, 1.08 mmol) in CH_2Cl_2 (2 mL) followed by solid trimethylamine hydrochloride (4.5 mg, 0.1 mmol) and trimethylamine (200 µL, 1.44 mmol). The reaction was stirred for 5 mins at this temperature after which it was allowed to warm to RT and stirred for 2 h. The reaction mixture was diluted with CH_2Cl_2 (5 mL) and washed with saturated sodium bicarbonate solution (5 mL), brine (5 mL), dried and concentrated *in vacuo*. Column chromatography (petroleum ether:Et_2O 3:1) afforded tosylate **29** as a yellow oil (65 mg, 21%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.63 (t, *J* = 7.3 Hz, 2H), 1.80 (t, *J* = 6.5 Hz, 2H), 1.94 – 2.03 (m, 3H), 2.48 (s, 3H), 3.93 (t, *J* = 6.4 Hz, 2H), 7.39 (d, *J* = 8.0 Hz, 2H), 7.84 (d, *J* = 8.3 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.2, 21.7, 25.8, 32.1, 32.8, 64.7, 69.4, 82.4, 128.0, 129.9, 132.8, 145.1; *m/z* 293 ($[M+H]^+$); HRMS (ESI^+) $C_{14}H_{17}N_2O_3S$ requires 293.0960, found 293.0964.

6.1.12. 2-[3-(but-3-yn-1-yl)-3H-diazirin-3-yl]ethyl 4-methylbenzene-1-sulfonate (30)[213]

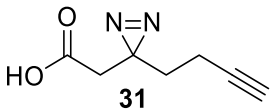


30

To a stirring solution of tosylate **29** (65 mg, 0.22 mmol) in DMF (1 mL) was added sodium azide (17 mg, 0.26 mmol). The reaction was heated to 70 °C and stirred for 5 hours. Upon completion, the reaction mixture was diluted with H_2O (5 mL) and extracted into EtOAc (2 x 5 mL). The combined organic layers were washed with brine (5 mL), dried and concentrated *in vacuo* to afford azide **30** as a yellow oil (27 mg, 77%) which was used without further purification.[213]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.72 (m, 4H), 1.99 – 2.09 (m, 3H), 3.19 (t, *J* = 6.8 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.3, 26.4, 32.3, 32.5, 46.0, 69.4, 82.5; *m/z* 164 ($[M+H]^+$).

6.1.13. 2-[3-(but-3-yn-1-yl)-3H-diazirin-3-yl]acetic acid (31)[217]

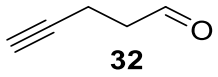


31

To a stirring solution of alcohol **28** (100 mg, 0.72 mmol) in DMF (2.5 mL) at 0 °C was added pyridinium dichromate (1.09 g, 2.90 mmol) slowly. The reaction mixture was stirred at RT for 3 hours. The resulting solution was diluted with H_2O (10 mL) and extracted into Et_2O (2 x 5 mL). The combined organic layers were washed with sat. NH_4Cl (5 mL), H_2O (5 mL), brine (5 mL), dried and concentrated *in vacuo*. Column chromatography (hexane:EtOAc 4:1 +1% AcOH) afforded acid **31** as an orange oil (28 mg, 26%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.82 (t, *J* = 7.3 Hz, 2H), 2.03 (t, *J* = 2.6 Hz, 1H), 2.10 (td, *J* = 7.2, 2.6 Hz, 2H), 2.44 (s, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.2, 25.3, 31.9, 39.4, 69.5, 82.4, 174.5; *m/z* unable to obtain mass spectra.

6.1.14. pent-4-ynal[309] (32)

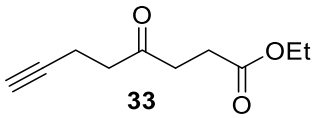


32

To a solution of pent-4-yn-1-ol (0.933 mL, 10 mmol) in CH_2Cl_2 (80 mL) was added $NaHCO_3$ (1.68 g, 20 mmol) and Dess-Martin periodinane (8.48 g, 20 mmol) in one portion at 0 °C. The resulting solution was stirred at RT for 2 hours. The reaction was quenched by the addition of $Na_2S_2O_3$ (50 mL) and $NaHCO_3$ (25 mL) and allowed to stir for 1 hour. The mixture was diluted with Et_2O (20 mL) and partitioned. The aqueous layer was washed with Et_2O (2 x 20 mL). The combined organic layers were washed with H_2O (2 x 10 mL), $NaHCO_3$ (10 mL) and brine (10 mL), dried and concentrated *in vacuo*. The suspension was filtered under vacuum, the filter cake was washed with Et_2O (2 x 10 mL) and the solution concentrated again *in vacuo* at 30 °C to afford crude aldehyde **32** as a dark yellow oil that was used without further purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 2.00 (t, *J* = 2.7 Hz, 1H), 2.33 (td, *J* = 7.0, 2.6 Hz, 2H), 2.52 (td, *J* = 7.2, 2.7 Hz, 2H), 9.81 (s, 1H).

6.1.15. ethyl 4-oxooct-7-ynoate[216] (33)

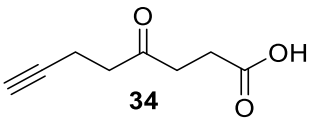


33

A solution of crude, freshly synthesised pent-4-ynal **32** (10 mmol) and ethyl acrylate (2.13 mL, 20.0 mmol) in dioxane (12 mL) was added dropwise over a period of 4 hours to a suspension of 3-benzyl-5-(2-hydroxyethyl)-4-methylthiaolium chloride (540 mg, 0.20 mmol), triethylamine (0.97 mL, 7.0 mmol) and ethyl acrylate (2.13 mL, 20.0 mmol) in dioxane (14 mL) at 80 °C under nitrogen. The mixture was heated at 80 °C for 54 hours. Once cooled, the mixture was diluted with CH_2Cl_2 (20 mL) and washed with 10% H_2SO_4 (10 mL), sat. $NaHCO_3$ (20 mL) and brine (20 mL), dried and concentrated *in vacuo*. Column chromatography (EtOAc:hexane 0:100 gradient to 20:80) afforded ketone **33** as a pale brown oil (369 mg, 20%).[216]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.27 (t, *J* = 7.1 Hz, 4H), 1.93 – 1.98 (m, 1H), 2.49 (td, *J* = 7.4, 2.6 Hz, 2H), 2.62 (t, *J* = 6.6 Hz, 2H), 2.75 (q, *J* = 7.5, 6.9 Hz, 4H), 4.15 (q, *J* = 7.2 Hz, 2H).; HRMS (APCI$^+$) $C_{10}H_{15}O_3^+$ $[M+H]^+$ requires 183.1016, found 183.1017.

6.1.16. 4-oxooct-7-ynoic acid[216] (34)

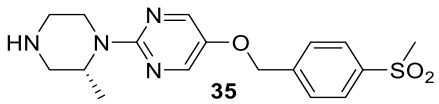


34

To a solution of ester **33** (360 mg, 1.97 mmol) in MeOH (16 mL) was added LiOH (411 mg, 9.80 mmol) and H_2O (180 µL, 10.0 mmol). The solution was stirred at RT overnight. The reaction mixture was acidified using 6M HCl until a pH of 3 was achieved. The resulting solution was then extracted into CH_2Cl_2 (2 x 10 mL), dried and concentrated *in vacuo* to afford acid **34** as a brown solid (246 mg, 81 %) which was used without further purification.[216]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.98 (t, *J* = 2.7 Hz, 1H), 2.50 (td, *J* = 7.3, 2.7 Hz, 2H), 2.67 – 2.82 (m, 6H); HRMS (ESI$^-$) $C_8H_9O_3^-$ [M-H]$^-$, requires 153.0552, found 153.0546.

6.1.17. (R)-2-(2-methylpiperazin-1-yl)-5-((4-(methylsulfonyl)benzyl)oxy)pyrimidine (35)

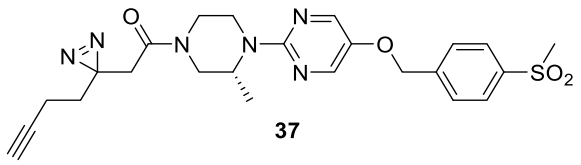


35

To a stirring solution of **AZ003** (100 mg, 0.216 mmol) in CH_2Cl_2 (3 mL) was added trifluoroacetic acid (1 mL). The reaction mixture was stirred at RT overnight. 1M NaOH was added to the resulting solution until a pH of 9 was obtained and subsequently extracted into CH_2Cl_2 (2 x 5 mL). The combined organic layers were washed with brine (10 mL), dried and concentrated *in vacuo* to give amine **35** as an off-white solid (75 mg, 96%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.26 (d, *J* = 6.8 Hz, 3H), 2.81 (td, *J* = 13.1, 3.6 Hz, 1H), 2.92 (dt, *J* = 12.3, 1.5 Hz, 1H), 3.00 – 3.06 (m, 1H), 3.09 (s, 3H), 3.06 – 3.16 (m, 2H), 4.27 – 4.38 (m, 1H), 4.66 – 4.77 (m, 1H), 5.13 (s, 2H), 7.60 – 7.68 (m, 2H), 7.96 – 8.04 (m, 2H), 8.15 (s, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.4, 39.7, 44.5, 46.0, 46.5, 50.4, 71.2, 77.3, 127.9, 128.0, 140.4, 144.7, 146.4, 157.9; *m/z* (ESI$^+$) 363 ([M+H]$^+$); HRMS (ESI$^+$) $C_{12}H_{23}N_4O_3S$ requires 363.1491, found 363.1505.

6.1.18. (R)-2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)-1-(3-methyl-4-(5-((4-(methylsulfonyl)benzyl)oxy)pyrimidin-2-yl)piperazin-1-yl)ethan-1-one (37)

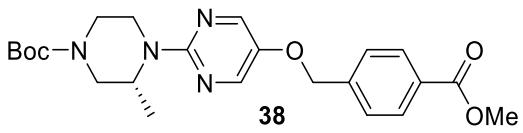


37

Amine **35** (35 mg, 0.10 mmol) and carboxylic acid **31** (13 mg, 0.09 mmol) were dissolved in THF (1 mL) and stirred for 10 minutes at RT. 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM, **36**) (28 mg, 0.10 mmol) was added and the reaction mixture stirred for 3 hours. The suspension was diluted with H_2O (5 mL) and extracted into EtOAc (2 x 5 mL). The organic layer was washed sequentially with saturated sodium bicarbonate (5 mL), H_2O (5 mL), 1M HCl (5 mL), H_2O (5 mL) and brine (5 mL). The organic layer was dried and concentrated *in vacuo*. Column chromatography (CH_2Cl_2:MeOH 97:3) afforded final probe **37** as a pale yellow solid (20 mg, 40%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.12 – 1.18 (m, 3H), 1.62 – 1.95 (m, 2H), 1.98 – 2.18 (m, 2H), 2.43 – 2.66 (m, 2H), 2.73 – 3.03 (m, 1H), 3.09 (s, 3H), 3.13 – 3.34 (m, 2H), 3.44 (d, *J* = 3.0 Hz, 1H), 3.49 – 3.68 (m, 1H), 3.99 (d, *J* = 5.8 Hz, 1H), 4.35 – 4.50 (m, 1H), 4.51 – 4.69 (m, 1H), 4.78 – 4.93 (m, 1H), 5.14 (s, 2H), 7.64 (d, *J* = 8.0 Hz, 2H), 7.96 – 8.03 (m, 2H), 8.13 – 8.19 (m, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.2, 14.1, 26.2, 32.0, 38.5, 39.1, 41.8, 44.5, 45.9, 46.9, 47.1, 50.2, 53.4, 54.6, 69.3, 71.0, 127.9, 128.0, 140.4, 142.6, 146.2, 157.2, 167.4, 172.4; *m/z* (ESI$^+$) 497 ([M+H]$^+$), 502 ([M-N_2+MeOH]$^+$); HRMS (ESI$^+$) $C_{24}H_{29}N_6O_4S$ requires 497.1971, found 497.1971).

6.1.19. tert-butyl (R)-4-(5-((4-(methoxycarbonyl)benzyl)oxy)pyrimidin-2-yl)-3-methylpiperazine-1-carboxylate (38)



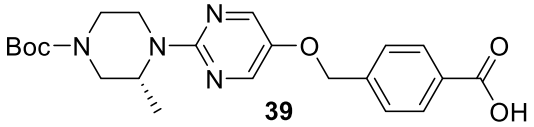
38

To a stirring solution of intermediate **22** (200 mg, 0.68 mmol) in DMF (2.5 mL) was added methyl 4-(bromomethyl)benzoate (155 mg, 0.68 mmol) and potassium carbonate (140 mg, 1.02 mmol). The resulting suspension was stirred at RT for 16 hours. The reaction mixture was partitioned between H_2O (10 mL) and EtOAc (5 mL).

The organic layer was washed with brine (5 mL), dried and concentrated *in vacuo*. Column chromatography (hexane:EtOAc 80:20 to 50:50) afforded ester **38** as an off-white solid (266 mg, 87%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.16 (d, *J* = 6.7 Hz, 3H), 1.50 (s, 9H), 2.93 (s, 1H), 3.16 (m, 2H), 3.94 (s, 3H), 4.09 – 4.25 (m, 1H), 4.32 (d, *J* = 13.3 Hz, 1H), 4.78 (d, *J* = 8.5 Hz, 1H), 5.09 (s, 2H), 7.44 – 7.54 (m, 2H), 8.04 – 8.11 (m, 2H), 8.14 (s, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.9, 28.4, 38.7, 43.0, 47.0, 48.5, 52.2, 71.6, 79.8, 127.2, 130.0, 130.1, 141.4, 145.1, 146.4, 155.3, 157.5, 166.7; *m/z* (ESI$^+$) 443 ([M+H]$^+$); HRMS (ESI$^+$) $C_{23}H_{31}N_4O_5$ requires 443.2294, found 443.2284.

6.1.20. (R)-4-(((2-(4-(tert-butoxycarbonyl)-2-methylpiperazin-1-yl)pyrimidin-5-yl)oxy)methyl)benzoic acid (39)

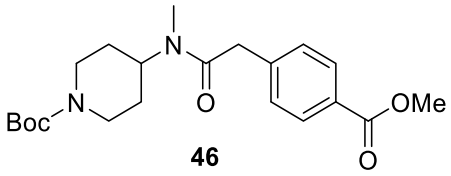


39

Ester **38** (100 mg, 0.23 mmol) was dissolved in MeOH (2 mL). Freshly powdered NaOH (14 mg, 0.34 mmol) was added. The reaction mixture was stirred at 70 ^{o}C overnight. The resulting mixture was partitioned between H_2O (10 mL) and EtOAc (10 mL). The aqueous layer was acidified using 1M HCl (5 mL) and the product extracted into EtOAc (3 x 10 mL). The combined organic layers were washed with H_2O (10 mL) and brine (10 mL), dried and concentrated *in vacuo* to afford acid **39** as a white solid (56 mg, 58%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.17 (d, *J* = 6.6 Hz, 3H), 1.51 (s, 9H), 2.94 (s, 1H), 3.17 (ddd, *J* = 13.4, 12.1, 3.6 Hz, 2H), 3.83 – 4.24 (m, 2H), 4.33 (d, *J* = 13.2 Hz, 1H), 4.77 (s, 1H), 5.12 (s, 2H), 7.50 – 7.59 (m, 2H), 8.13 – 8.19 (m, 4H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 13.9, 28.4, 38.7, 43.8, 47.1, 48.6, 71.6, 79.8, 127.3, 129.1, 130.6, 142.3, 145.1, 146.4, 155.3, 170.2; *m/z* (ESI$^+$) 429 ([M+H]$^+$); HRMS (ESI$^+$) $C_{22}H_{29}N_4O_5$ requires 429.2138, found 429.2138.

6.1.21. *tert*-Butyl 4-{2-[4-(methoxycarbonyl)phenyl]-N-methylacetamido}piperidine-1-carboxylate (46)

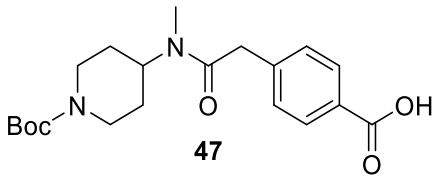


46

A solution of acid **21** (500 mg, 2.58 mmol) and amine **20** (607 mg, 2.84 mmol) in THF (20 mL) was stirred for 10 minutes. DMTMM **36** (785 mg, 2.84 mmol) was then added and the solution stirred for a further 3h. The reaction mixture was diluted with water (10 mL) and extracted into CH_2Cl_2 (2 x 10 mL). The combined organic layers were washed sequentially with saturated sodium carbonate (10 mL), water (10 mL), 1M HCl (10 mL), water (10 mL) and brine (10 mL), dried and concentrated *in vacuo*. Purification via column chromatography (methanol:DCM 1:99) afforded **46** as a white solid (770 mg, 77%).[310]

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.46 (s, 9H), 1.49 – 1.83 (m, 4H), 2.43 – 2.81 (m, 2H), 2.82 (s, 3H), 3.78 (s, 2H), 3.92 (s, 3H), 4.11 – 4.31 (m, 2H), 4.67 (ddt, *J* = 11.9, 7.5, 4.2 Hz, 1H), 7.31 – 7.39 (m, 2H), 8.01 (dd, *J* = 8.3, 2.7 Hz, 2H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 28.4, 28.7, 30.0, 41.7, 43.2, 51.0, 52.1, 79.8, 128.6, 128.8, 130.0, 140.3, 166.8, 169.7, 170.1.; *m/z* (ESI$^+$) 291 [M-Boc+H$^+$]$^+$, 65%), 335 ([M-^{t}Bu+H$^+$]$^+$,85%), 391 ([M+H$^+$]$^+$, 25%), 413 ([M+Na$^+$]$^+$, 100%); HRMS (ESI$^+$) $C_{21}H_{31}N_2O_5^+$ ([M+H$^+$]$^+$) requires 391.2230, found 391.2233.

6.1.22. 4-(2-((1-(*tert*-butoxycarbonyl)piperidin-4-yl)(methyl)amino)-2-oxoethyl)benzoic acid (47)



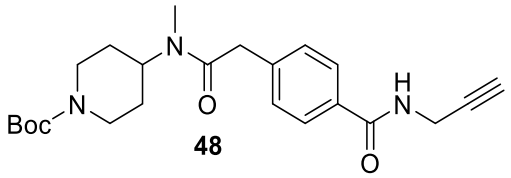
47

Ester **46** (240 mg, 0.62 mmol) was dissolved in MeOH (10 mL). Freshly powdered NaOH (49 mg, 1.20 mmol) was added and the reaction mixture stirred at 70 ºC overnight. The resulting mixture was partitioned between EtOAc and H_2O. The aqueous layer was acidified with 1M HCl and the product extracted with EtOAc (3 x 10 mL). The combined organic layers were washed with H_2O (2 x 10 mL), brine (10

mL), dried and concentrated *in vacuo* to afford acid **47** as a pale yellow solid (140 mg, 60%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.2 – 1.8 (m, 13H), 2.5 – 2.8 (m, 2H), 2.8 (s, 3H), 3.75 – 3.97 (m, 2H), 4.1 – 4.3 (m, 2H), 4.7 (ddt, *J* = 11.8, 8.5, 4.3 Hz, 1H), 7.31 – 7.49 (m, 2H), 8.1 (dd, *J* = 8.3, 2.9 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 28.4, 28.7, 30.1, 41.7, 43.1, 51.1, 79.8, 128.1, 129.0, 130.6, 141.0, 154.7, 170.3, 170.7; *m/z* (ESI^+) 377 ($[M+H]^+$), 321 ($[M-^tBu+H]^+$); HRMS (ESI^+) $C_{20}H_{29}N_2O_5^+$ requires 377.2076, found 377.2077.

6.1.23. *tert*-butyl 4-(N-methyl-2-(4-(prop-2-yn-1-ylcarbamoyl)phenyl)acetamido)piperidine-1-carboxylate (48)

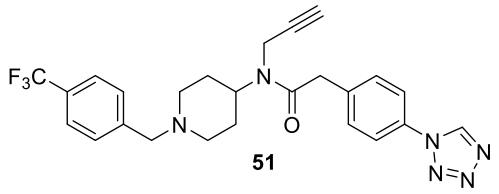


48

Acid **47** (100 mg, 0.26 mmol) was dissolved in THF (5 mL) and propargylamine (18 µL, 0.29 mmol) was added and the solution stirred for 10 minutes. DMTMM **36** (80 mg, 0.29 mmol) was added and stirring continued at RT overnight. The reaction mixture was diluted with H_2O (10 mL) and extracted into CH_2Cl_2 (2 x 10 mL). The combined organic layers were washed sequentially with sat. $NaHCO_3$ (10 mL), H_2O (10 mL), 1M HCl (10 mL), H_2O (10 mL), brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (EtOAc:hexane 70:30 gradient to 100:0) afforded **48** as a yellow oil (69 mg, 64%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.27 – 1.74 (m, 13H), 2.27 (t, *J* = 2.4 Hz, 1H), 2.50 – 2.88 (m, 5H), 3.70 – 3.88 (m, 2H), 3.95 – 4.23 (m, 2H), 4.23 (dd, *J* = 5.3, 2.6 Hz, 2H), 4.64 (tt, *J* = 11.5, 4.5 Hz, 1H), 6.66 – 6.79 (m, 1H), 7.16 – 7.39 (m, 2H), 7.64 – 7.84 (m, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 28.4, 28.7, 29.7, 30.0, 41.5, 43.2, 51.0, 71.7, 79.7, 80.0, 127.5, 129.0, 132.4, 138.9, 154.6, 166.8, 170.2; *m/z* (ESI^+) 314 ($[M-Boc+H]^+$), 414 ($[M+H]^+$), 436 ($[M+Na]^+$); HRMS (ESI^+) $C_{23}H_{32}N_3O_4^+$ requires 414.2392, found 414.2389.

6.1.24. 2-(4-(1H-tetrazol-1-yl)phenyl)-N-(prop-2-yn-1-yl)-N-(1-(4-(trifluoromethyl)benzyl)piperidin-4-yl)acetamide (51)

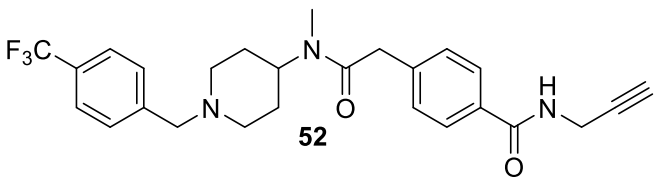


51

Tetrazole **44** (1.77g, 0.417 mmol) was dissolved in CH_2Cl_2 (5 mL) and TFA (1 mL) was added. The reaction mixture was stirred at RT until TLC indicated the reaction was complete. The solvent was removed *in vacuo* and the crude mixture resuspended in MeCN (5 mL). DIPEA (0.47 mL, 2.71 mmol) and 1-(bromomethyl)-4-(trifluoromethyl)benzene (109 mg, 0.46 mmol) were added and the reaction mixture stirred at RT overnight. The solvent was removed and the resulting residue resuspended in CH_2Cl_2 (10 mL) and washed with H_2O (2 x 10 mL) and brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (EtOAc:hexane 70:30 gradient to 100:0) afforded **51** as a pale yellow solid (83 mg, 41%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.22 – 1.63 (m, 2H), 1.66 – 1.89 (m, 2H), 1.93 – 2.20 (m, 2H), 2.31 – 2.46 (m, 1H), 2.96 (d, *J* = 8.5 Hz, 2H), 3.57 (d, *J* = 9.6 Hz, 2H), 3.85 – 3.97 (m, 2H), 3.99 – 4.19 (m, 2H), 4.42 – 4.66 (m, 1H), 7.49 (dd, *J* = 18.0, 10.0 Hz, 5H), 7.59 (d, *J* = 7.9 Hz, 2H), 7.69 (dd, *J* = 8.3, 5.9 Hz, 2H), 9.00 (d, *J* = 4.0 Hz, 1H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 30.6, 32.8, 40.4, 52.0, 53.0, 62.2, 72.6, 79.9, 121.4, 125.3, 129.0, 129.2, 131.2, 132.7, 137.4, 140.5, 142.4, 170.2; *m/z* (ESI$^+$) 455 ([M-N_2+H]$^+$), 483 ([M+H]$^+$); HRMS (ESI$^+$) $C_{25}H_{26}N_6OF_3^+$ requires 483.2120, found 483.2121.

6.1.25. 4-(2-(methyl(1-(4-(trifluoromethyl)benzyl)piperidin-4-yl)amino)-2-oxoethyl)-N-(prop-2-yn-1-yl)benzamide (52)



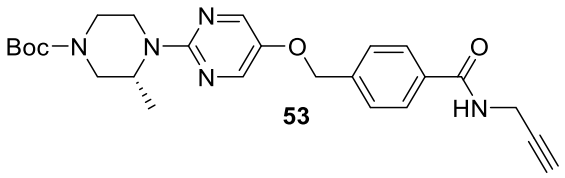
52

Boc-protected amine **48** (35 mg, 0.085 mmol) was dissolved in CH_2Cl_2 (2.5 mL) and TFA (0.5 mL) added. The mixture was stirred until complete by TLC. The solvent was removed and the crude residue resuspended in MeCN (2 mL). DIPEA (96 µL, 0.55

mmol) and 1-(bromomethyl)-4-(trifluoromethyl)benzene (22 mg, 0.094 mmol) was added and the mixture stirred at RT overnight. The solvent was removed and the residue resuspended in CH_2Cl_2 (10 mL) and washed with H_2O (10 mL) and brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (MeOH:DCM 1:99 gradient to 10:90) afforded **52** as a yellow oil (7 mg, 17%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.46 – 2.22 (m, 6H), 2.30 (q, *J* = 2.6 Hz, 1H), 2.87 (s, 3H), 2.88 – 3.15 (m, 2H), 3.43 – 3.69 (m, 2H), 3.73 – 3.83 (m, 2H), 4.27 (dd, *J* = 5.0, 2.5 Hz, 2H), 4.55 (s, 1H), 6.25 – 6.42 (m, 1H), 7.31 – 7.40 (m, 2H), 7.40 – 7.52 (m, 2H), 7.55 – 7.64 (m, 2H), 7.70 – 7.84 (m, 2H); *m/z* (ESI$^+$) 472 ([M+H]$^+$); HRMS (ESI$^+$) $C_{26}H_{29}N_3O_2F_3^+$ requires 472.2212, found 472.2212.

6.1.26. *tert*-butyl (R)-3-methyl-4-(5-((4-(prop-2-yn-1-ylcarbamoyl)benzyl)oxy)pyrimidin-2-yl)piperazine-1-carboxylate (53)



53

Acid **39** (100 mg, 0.23 mmol) was dissolved in THF (5 mL) and propargylamine (17 µL, 0.26 mmol) was added and the solution stirred for 10 minutes. DMTMM **36** (71 mg, 0.26 mmol) was added and stirring continued at RT overnight. The reaction mixture was diluted with H_2O (10 mL) and extracted into CH_2Cl_2 (2 x 10 mL). The combined organic layers were washed sequentially with sat. $NaHCO_3$ (10 mL), H_2O (10 mL), 1M HCl (10 mL), H_2O (10 mL), brine (10 mL), dried and concentrated *in vacuo*. Column chromatography (EtOAc:hexane 70:30) afforded **53** as a yellow oil (81 mg, 74%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.15 (d, *J* = 6.6 Hz, 3H), 1.50 (s, 9H), 2.31 (t, *J* = 2.6 Hz, 1H), 2.74 – 3.27 (m, 2H), 4.06 (s, 2H), 4.14 – 4.39 (m, 3H), 4.76 (s, 1H), 5.08 (s, 2H), 6.44 (t, *J* = 5.3 Hz, 1H), 7.49 (d, *J* = 8.1 Hz, 2H), 7.73 – 7.88 (m, 2H), 8.12 (s, 2H); *m/z* (ESI$^+$) 466 ([M+H]$^+$); HRMS (ESI$^+$) $C_{25}H_{32}N_5O_4^+$ requires 466.2454, found 466.2470.

6.2. Molecular biology

6.2.1. General Methods

For molecular biology polymerase chain reaction (PCR) reactions, ultra-pure water certified to be free of RNA and DNA was used, otherwise ultra-pure MilliQ® water

(obtained using a Millipore water purification system) was used to prepare buffers. Primers were ordered from Invitrogen and stored as a 100 µM stock (UltraPure H_2O) at -20 ºC. Restriction enzymes, polymerases, ligases and dNTPs were obtained from New England Biolabs. DH5α and DH10Bac competent bacteria were obtained from Invitrogen. For bacterial work, all media, flasks and tips were autoclaved prior to use. 0.8-2% agarose gels were prepared from high grade agarose powder dissolved in TAE buffer (Tris-acetate-EDTA buffer) in the microwave. GelRed stain was then added before casting the gel. Gels were run at 80 V for 80 minutes, or until sufficient migration was achieved. Gels were imaged using either a UV trans-illuminator or on a Typhoon FLA 9500 imager (GE Healthcare). DNA concentration was determined using a Nanodrop (Thermo Fisher Scientific) using 1.5 µL of material.

6.2.2. Primer Design

Primers were designed such that melting temperatures (T_m) of primer pairs were within 5 ºC of each other and not exceeding 60 ºC. Primers were checked for hairpin structures using OligoAnalyzer[311] and sequences altered if stable hairpin structures were present. The general structure for primers (5' to 3') consisted of 2-6 base pairs (bp) prior to annealing region, appropriate restriction site, followed by 12-18 bp to anneal to template DNA. Primers used in this study were used to amplify either gene or tag of interest in order to produce a tagged GPR119 variant, as listed in **Table 5**.

	Insert	Primer (5' tp 3')
1	SgfI-**GPR119**-PmeI	F: TAATGCGCGATCGC**CATGGAATCATCTTTCTC**
		R: TGCTGCGTTTAAAC**TTAGCCATCAAACTCTG**
2	HindIII-**TST**-NotI	F: TAAGCAAAGCTT **ATGTGGTCCCATCCTCAG**
		R: ATTATCGCGGCCGC**CTTTTCAAACTGTG**
3	HindIII-**TST**-**GPR119**-XbaI	R: CGCTGCTCTAGA**GTCTTAGCCATCAAACTC**
4	BamHI-**TST**-**GPR119**-XbaI	F: TAAAGCGGATCC**ATGTGGTCCCATCCTC**
5	HindIII-**Halo**-EcoRI	F: TGAACGAAGCTT**TATTGCGGTAGTTTA**
		R: TTAGAATTC**GCTCTGAAAGTACAG**

6	Halo Site directed mutagenesis	F: *CACGAGAATGGAATCA*TCTTTCTCATTTGGAGTGATCC
		R: *TGATTCCATTCTCGTG*AATTCGCTCTGAAAGTACAGATC
7	Halo NEB site directed mutagenesis	F: ATGGAATCATCTTTCTCATTTG
		R: TCTCGTGAATTCGCTCTG

Table 5: Primer pairs designed & used in this study.

6.2.3. PCR Amplification of Inserts

PCR mixtures were prepared in 0.2 mL PCR tubes on ice as follows:

Reagent	**Final concentration**	**Volume (µL)**
5× Q5 buffer	1x	10
dNTP (10 mM)	200 µM	1
10 µM Forward primer	0.5 µM	2.5
10 µM Reverse primer	0.5 µM	2.5
Template DNA	10 ng	X
H_2O		33.5-X
Q5 polymerase enzyme		0.5
	Total Volume	50

For each reaction, control samples were also prepared with H_2O in place of template DNA and in place of primers. The tubes were then placed in a PCR thermocycler (Techne TC-412 Thermal Cycler) with a heated lid and with the following programme:

Stage	**Temperature (°C)**	**Time**	
1. Initial denaturation	98	30 seconds	
2. Denature	98	10 seconds	30 cycles
3. Anneal	50-72*	30 seconds	
4. Extend DNA	72	30 seconds/kb	
5. Final extension	72	5 minutes	
6. Hold	10	∞	

*The annealing temperature was calculated using the primer pair T_m & NEB calculator[312] for the polymerase.

Once the PCR cycles were complete, the PCR products were checked using agarose gel to confirm successful and specific amplification. Briefly, 1 µL of purple loading buffer was added to 5 µL PCR reaction mixture and the samples run at 80 V for 80 minutes. For clean PCR products, the remaining 45 µL reaction mixture was purified using a Qiagen QIAquick PCR purification column following the kit instructions. DNA was eluted using 30 µL H_2O and DNA concentration determined. PCR products were stored at 4 °C until use, or at -20 °C for longer term storage.

6.2.4. Digestion and Ligation

2 µg of vector DNA and 500 ng of insert DNA were digested independently. The reaction mixtures were prepared in 1.5 mL Eppendorf tubes on ice as follows:

Component	**Volume (µL)**
10X Cut smart buffer	5
Restriction Enzyme 1	1
Restriction Enzyme 2	1
DNA (500 ng/2 µg)	X
H_2O	43-X
Total	50

The mixture was incubated without agitation for 1 hour at 37 ºC. For the insert sample, 1 µL of purple loading buffer was added to 5 µL of digestion mixture. For the vector sample, 10 µL of purple loading buffer was added to the 50 µL reaction. Both samples were run on a 0.8% agarose gel at 80 V for 80 minutes. The gel was imaged using a UV trans-illuminator and following successful digestion, the digested vector sample was excised from the agarose gel using a clean scalpel, using the UV to ensure all digested DNA was removed. Vector digestion samples were purified using Qiagen QIAquick Gel Purification Kit following the manufacturer's instructions, eluted in 30 µL H_2O and DNA concentration determined by Nanodrop. Insert digestion samples were purified using QIAquick PCR purification columns as before, eluted in 30 µL of H_2O and DNA concentration determined.

For ligation, 50 ng vector was ligated with the insert in a 5:1 (insert:vector) ratio. Quantity of insert was calculated[313] and mixtures assembled as follows:

Reagent	**Final concentration**	**Volume (µL)**
10x ligase buffer	1x	2
Digested vector	50 ng	X
Digested insert	? ng (5:1 insert:vector)	Y
3 U/µl T4 DNA ligase	~1 U	0.3
H_2O		17.7-X-Y
	Total Volume	20

A control ligation was also prepared without insert to check background levels of undigested vector. The samples were incubated without agitation for 20 minutes at RT. The reaction was then chilled on ice for 5 minutes.

6.2.5. Transformation and Colony Validation

For this work, ligated plasmids were transformed into DH5α competent cells. Stock cells were thawed on ice for 20 minutes. 50 µl aliquots were prepared in 1.5 mL Eppendorfs and those that were not required were stored at -80 °C. 5 µl of pre-chilled ligation reaction was added to 50 µl cells, mixed by gentle tapping and incubated on ice for 20 minutes. The cell mixture was then heat shocked using a preheated heat block at 42 °C for 45 seconds and placed immediately back on ice for 2 minutes. 450 µl of Super Optimal broth with Catabolite repression (SOC) media was added and the cells incubated with shaking for 1 hour at 37 °C to recover. 10% of the mixture was spread on an ampicillin (100 µg/mg) LB agar plate and the remaining 90% pelleted at 2000 rpm and resuspended in 50 µl SOC media and spread on another ampicillin plate. Colonies were grown overnight at 37 °C.

Following visual inspection of the colonies, colony PCR was used to confirm successful ligation of the insert into the vector. Briefly, 6-10 colonies per plate were circled and numbered and the original vector was used as a control. Primers were selected that either flank the insert site (sequencing primers) or are only present in the insert (original amplification primers) or a mix of a primer flanking the insert and one only found within the insert that would result in a detectable bp shift or presence/absence of a PCR product. The PCR master mix was prepared as follows and dispensed into 0.2 mL PCR tubes:

Component	**Final concentration**	**Volume (µL)**
5x Green GoTaq buffer	1x	4
10 mM dNTPs	200 µM	0.4
10 µM forward primer	0.3 µM	0.6
10 µM reverse primer	0.3 µM	0.6
GoTaq polymerase	0.5 U	0.1
H_2O		14.3
	Total Volume	20

Colonies were picked with autoclaved pipette tips and added to each PCR tube. 0.1 µL of original plasmid was added to the control reaction tubes. PCR was carried out in a thermocycler using the following conditions:

Stage	**Temperature (°C)**	**Time**	
1. Initial denaturation	95	2 minutes	
2. Denature	95	20 seconds	30 cycles
3. Anneal	55	30 seconds	
4. Extend DNA	72	1 minute/kb	

5. Final extension	72	5 minutes
6. Hold	10	∞

5 µL of each reaction mixture was directly loaded onto an agarose gel and run at 80 V for 80 minutes, or until sufficient migration was achieved as visually inspected by the yellow and blue migration markers present in the GoTag Green buffer. The gel was imaged and positive clones identified by those with the expected base pair shift.

6.2.6. Growth of Bacteria and DNA Purification

Remaining positive colonies were picked with a sterile tip and grown in 5 mL of ampicillin (100 µg/mg) containing LB broth with gentle shaking over an 8 hour period or overnight at 37 °C. 1 mL of culture was taken and centrifuged at 4000 rpm for 20 minutes. The remaining culture was stored at 4 °C for the short term. The bacterial pellet was resuspended and processed using a Qiagen QIAprep Spin Miniprep Kit following the manufacturer's instructions. The final DNA was eluted using 30 µL H_2O, DNA concentration calculated and ~700 ng of DNA was sent for sequencing with GeneWiz using an appropriate universal sequencing primer.

Following confirmation of correct DNA sequence, 150 mL ampicillin containing LB broth was inoculated with 500 µL of the original culture and grown overnight with gentle shaking at 37 °C. Glycerol stocks were created by diluting 500 µL of bacterial culture in 500 µL of autoclaved 50% glycerol solution in cryovials. These stocks were kept at -80 °C for long term storage. The remaining culture was pelleted by centrifugation at 4000 rpm for 20 minutes. The bacterial pellet was resuspended, lysed and processed using a Qiagen QIAprep Spin Maxiprep kit following the manufacturer's instructions. The final DNA was eluted using 750 µL of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) buffer, DNA concentration calculated and aliquoted for storage at -20 °C. 700 ng of DNA was sent for sequencing to confirm the correct DNA product using both forward and reverse sequencing primers.

6.2.7. Bacmid production & verification

A pFastBac vector containing a TST-GPR119 insert was constructed following the methods described above. DH10Bac competent bacmid producing cells were thawed on ice. 1 ng of pFastBac_TST-GPR119 plasmid DNA was added, the cells gentle mixed by tapping and incubated on ice for 30 minutes. The cell mixture was then heat shocked using a preheated heat block at 42 °C for 45 seconds and placed immediately

back on ice for 2 minutes. 900 µl of SOC media was added and the cells incubated with shaking for 4 hours at 37 ºC to recover. Three 10-fold dilutions of the bacterial growth were prepared in SOC media and 100 µl of each dilution was spread on a LB agar plate containing the following additives: Kanamycin (50 µg/mL), Gentamicin (7 µg/mL), Tetracycline (10 µg/mL), Xgal (100 µg/mL) and IPTG (40 µg/mL). The plates were incubated at 37 ºC for 48 hours to allow colonies to grow large enough to allow blue/white selection. Upon visual inspection, twelve colonies presenting a white phenotype (those that had successfully produced the desired bacmid) were picked and restreaked onto fresh LB agar plates containing the full additives list to confirm white phenotype. Two positive colonies were picked using sterile pipette tips and grown in 250 mL antibiotic (Kanamycin (50 µg/mL), Gentamicin (7 µg/mL) and Tetracycline (10 µg/mL)) containing LB broth overnight with gentle shaking at 37 ºC.

Glycerol stocks were created by diluting 500 µL of bacterial culture in 500 µL of autoclaved 50% glycerol solution in cryovials. These stocks were kept at -80 ºC for long term storage. The remaining culture was pelleted by centrifugation at 4000 rpm for 20 minutes. The bacterial pellet was resuspended, lysed and processed using a PureLink HiPure Midiprep kit following the manufacturer's instructions. Purified bacmid DNA was resuspended in 400 µL of TE buffer, DNA concentration calculated and stored in 10 µL aliquots at -20 ºC for long term storage.

Successful transposition of the desired sequence into the bacmid was verified using PCR as previously described using 100 ng of Bacmid DNA and the following primers:

Primer	Sequence
pUC/M13 Forward	5′-CCCAGTCACGACGTTGTAAAACG-3′
pUC/M13 Reverse	5′-AGCGGATAACAATTTCACACAGG-3′

The PCR was run using a thermocycler programmed as before (6.2.3) a 55 ºC annealing temperature and 120 second DNA extension time. 5 µL of PCR mixture was analysed by agarose gel as before. Successful transposition was indicated by the appearance of a 3300 base pair band.

6.3. Biochemical method

6.3.1. General Methods

For biological experiments, ultra-pure MilliQ® water was obtained using a Millipore water purification system. SDS-PAGE was carried out using either 12%

polyacrylamide tris/glycine gels (cast in-house before use) or 4-20% Mini-PROTEAN® TGX ™ gradient gels (BioRad) with Precision plus all blue protein ladder standard. In-gel fluorescence was imaged using a Typhoon FLA 9500 imager (GE Healthcare). Chemiluminescence was imaged using a LAS-4000 imager (GE Healthcare). Protein concentration was measured using the BioRad DC Protein Assay following the manufacturer's protocol, using BSA as a protein standard and measuring absorbance at 750 nm. Absorbance & TR-FRET were recorded with an EnVision Xcite plate reader (Perkin Elmer). Tissue culture media and reagents were obtained from Sigma- Aldrich. Insect cell work was carried out in the Imperial Baculovirus Lab by Eloise Morecroft. AzTB, AzT and AzRB were synthesised in-house as previously reported.[314,315] All agonists and probes were prepared as 10 mM stocks in DMSO and stored at 4°C or -20°C unless otherwise stated.

6.3.2. Mammalian Cell Culture

Human embryonic kidney 293 (HEK293) cells were obtained from the Francis Crick Institute and identity confirmed via Short Tandem Repeat (STR) analysis. HEK393T cells were kindly gifted by Eloise Morecroft (Imperial College London) and identity confirmed via STR analysis. GLUTag and HEK293-GPR119 cells were obtained from AstraZeneca. All cell lines were maintained in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum (FBS) in a humidified 5% CO_2 atmosphere at 37 °C. Cells were passaged once 90% confluency was obtained by treatment with 0.25% trypsin and resplitting at a ratio of 1:3-1:6. Prior to treatment, cells were grown to 90% confluency.

6.3.2.1. Poly D-Lysine Coating of Plates or Coverslips

For experiments requiring multiple washes or incubations, plates or coverslips were coated with Poly D-Lysine prior to cell seeding. A 0.1 mg/mL solution of Poly D-Lysine was prepared by solubilising 5 mg Poly D-Lysine (HBr salt, M_r >300,000 g/mol, P7405) in 50 mL of sterile H_2O. 1 mL (or enough to completely cover the surface) of this solution was added per well and incubated for 10 minutes at RT. The solution was aspirated and the surface washed with sterile water (×2). The surface was left to dry completely for 1 hour at RT within the biological safety cabinet. Plates were either used immediately or stored at 4 °C for future use.

6.3.3. Mammalian cell transfections

Prior to transfection, cells were seeded to achieve 95% confluency upon transfection. Unless otherwise stated, Lipofectamine 2000 (ThermoFisher) was used for transfections in a 1:6 (DNA:lipofectamine) ratio. Lipofectamine 2000 was added to OptiMEM Reduced Serum Media (ThermoFisher) in a sterile Eppendorf and incubated for 5 minutes at RT. DNA was added to OptiMEM in another sterile Eppendoff (unless otherwise stated, DNA quantities used per plate size as described in Table 6). The diluted DNA media mixture was then added to the Lipofectamine mixture and incubated at RT for 20 minutes to create DNA containing liposomes. The mixture was then added dropwise to the plate with swirling. The cells were incubated for 5 h at 5% CO_2 and 37 ^oC. The media was exchanged for fresh media and returned to the incubator for 19 h. Transfected cells were then ready for use.

Plate size	HaloTag-GPR119 DNA (µg)
24-well	0.8
12-well	1.6
6-well	4
10cm dish	13.92

Table 6: Concentration of plasmid DNA used per well in transfection experiments (unless otherwise stated)

6.3.4. Mammalian cell lentiviral transduction

6.3.4.1. Generation of the Lentivirus

HEK293T cells were plated in DMEM supplemented with 10% FBS 24 hours prior to transfection in a 10 cm dish to achieve a confluency of 90-95%. 12 µg pLenti-FLAG-GPR119 (synthesised by GeneArt, Invitrogen), 10.2 µg p8.9 packaging plasmid (gifted from Eloise Morecroft) and 1.8 µg pMDG VSV-G envelope plasmid (gifted from Eloise Morecroft) were mixed and diluted up to 1500 µL with OptiMEM, vortexed and incubated at RT for 5 minutes. 60 µL Lipofectamine 2000 was diluted up to 1500 µL with OptiMEM, vortexted and incubated at RT for 5 minutes. Both mixtures were combined, vortexed and incubated at RT for 20 minutes. The media was replaced on the HEK293T cells with OptiMEM. The DNA/Lipofectamine mixture was then added dropwise to the cells with swirling. The cells were incubated for 5 h at 5% CO_2 at 37 ^oC. The media was then exchanged for fresh media. The cells were incubated for a further 48 h before the virus was collected from the cell medium, centrifuged for 4

minutes at 1500 rpm and filtered (0.45 µm filter). The virus was then aliquoted and stored at -80 °C until use.

6.3.4.2. Lentiviral Transduction and Selection

HEK293 cells were plated in a 24 well plate, to achieve 90-95% confluency the following day. The cell media was exchanged to media containing 8 µg/mL polybrene (Sigma-Aldrich) and the cells incubated for 30 minutes at 37 °C prior to incubation. The virus was then added to the wells in various dilutions (6:4 -> 1:9 virus:media) to total 1 mL total media added and then the cells returned to the incubator for 24 hours. The media was then exchanged to media containing 10 puromycin µg/mL for selection of positively transduced cells. Once selected cells were confluent, they were expanded into larger flasks to produce stocks of the transduced cell line.

6.3.5. Insect cell culture

Insect cell culture was carried out by Eloise Morecroft in the Baculovirus Lab in the Structural Biology department, Imperial College London. Briefly, Sf9 cells and High5 cells were maintained in Lonza InsectXpress media supplemented with sodium benzylpenicillin at 30 µg/mL and streptomycin sulphate at 50 µg/mL in a suspension culture. Cells were grown to mid-log phase of approximately 1-2 x 10^6 cells/mL and split by dilution in culture media to approximately 2 x 10^5 cells/mL in 2 L roller bottles and incubated at 27 °C with shaking at 130 rpm.

6.3.6. Bacmid transfection & viral transduction

Bacmid transfection & viral transduction was carried out by Eloise Morecroft. Cells were counted using a 1:1 dilution of cells in Trypan Blue stain and read in a Countess cell counting chamber slide. Sf9 cells were diluted to 0.5 x 10^6 cells/mL and 2 mL added to the wells of a 6-well plate. The cells were allowed to adhere to the plate for 1 hour at room temperature. The medium was replaced with 2.5 mL Lonza InsectXpress without antibiotics.

5 µL of recombinant bacmid was added to 100 µL of Lonza InsectXpress media (without antibiotics) in a sterile tube. In a separate tube, 10 µL of Cellfectin II (Invitrogen) was mixed with 100 µL of media. The two solutions were then mixed together with gentle pipetting and left at room temperature for 30 minutes.

The transfection mixture was then added dropwise to each well and the plate placed inside a sterile boxed lined with damp paper towel in an incubator for 5 hours at 27 °C. The media was removed and replaced with 2 mL of Lonza InsectXpress media (containing antibiotics) in each well. The plate was inspected after 72 hours to establish whether successful transfection and virus production had occured by observation of a rounding up phenotype and stalling of cell division. The media containing virus (P1) was harvested and centrifuged at 3000 rpm for 5 minutes to pellet any cell debris. 100 mL of Sf9 cells were then infected with the P1 virus over 72 hours to obtain a P2 stock of virus.

100 mL High5 cells were then transduced with 5 mL of P2 virus stock and aliquots taken at appropriate time points to track expression of tagged protein. After 72 hours, the cells were harvested.

6.3.7. Cell Lysis

The protein concentration of all lysates was calculated using the BioRad DC Protein concentration assay following the manufacturer's protocol. BSA was used to produce a standard curve in each different lysis buffer for more accurate determination of protein concentration. Unless stated, EDTA-free Complete protease inhibitors (Roche) were added to all lysis buffers.

6.3.7.1. Whole Cell Lysis

For cell lysis, this protocol was followed unless stated otherwise. Once desired confluency was achieved, or after treatment, the culture media was aspirated and the cells were washed with PBS (×2). The cells were placed on ice and enough lysis buffer (see **Table 7**) was added to the plate to achieve 1-3 mg/mL protein concentration. The cells were then scraped into an Eppendorf tube and left on ice to lyse for at least 30 minutes. For lysates with a large insoluble fraction, the cells were additionally sheared by passing the lysates through a 25 gauge needle. The lysates were clarified by centrifugation (17,000 × g, 10 minutes), the protein concentration of the supernatant determined, and stored at -80 °C until further use.

Lysis buffer	Recipe
Standard	1% Triton X-100, 0.1% SDS, PBS
Triton X-100	1% Triton X-100, 150 mM NaCl, 50 mM HEPES pH 7.5
NP-40	1% NP-40, 150 mM NaCl, 50 mM HEPES pH 7.5

SDS	1% SDS, 150 mM NaCl, 50 mM HEPES pH 7.5
2% SDS	2% SDS, PBS
DDM	1% DDM, 10% glycerol, 150 mM NaCl, 50 mM HEPES pH 7.5
HaloLink	1% Triton X-100, 0.1% sodium deoxycholate, 150 mM NaCl, 50 mM Tris-HCl pH 7.5

Table 7: **Lysis buffers used in this study**

6.3.7.2. Cell Fractionation

For experiments requiring enriched membrane fractions, the following lysis protocol was used. Once desired confluency was achieved, or after treatment, the culture media was aspirated and the cells were washed with ice-cold PBS (X2). The cells were detached using Non-enzymatic Cell Dissociation Buffer (Sigma) and the plates washed with 2 mL ice-cold HEPES buffer. The plate was washed again with 1 mL ice-cold HEPES buffer and the washings combined. The cells were pelleted at 1000 rpm for 4 minutes, the pellet resuspended in 1 mL hypotonic buffer and left on ice for 20 minutes.

Buffer	**Recipe (plus protease inhibitors)**
HEPES Buffer	150 mM NaCl, 50 mM HEPES pH 7.4
Hypotonic Buffer	10 mM NaCl, 1.5 mM $MgCl_2$, 10 mM HEPES pH 7.4

The cells were sheared by passing the cell solution through needles of various sizes: 21G needle for 20 strokes, 25G needle for 20 strokes and a 30G needle for 20 strokes. The cell lysate was checked under the microscope for complete cell disruption and inspected for the presence of membranes and nuclei and absence of whole cells.

The nuclear fraction and remaining cell debris was collected by centrifugation (800 g, 5 minutes, 4 ^{o}C). The supernatant (containing the membrane and soluble proteins) was transferred to 1.5 mL ultracentifuge tubes (Beckman) and balanced (± 0.2 mg). The membrane was pelleted by ultracentrifugation (100,000 g, 1 hour, 4 ^{o}C) using a Beckman Optima TL-100 ultracentrifuge and TLA-55 fixed angle rotor.

The cytosolic fraction was obtained by removing the supernatant. The nuclear fraction was obtained by washing the nuclear pellet with 500 µL HEPES buffer and shearing the cells with a 21G needle (30 strokes). The nuclear fraction was pelleted by centrifugation (800 g, 5 minutes) and washed a further two times with HEPES buffer before being pelleted once more (1000 g, 5 minutes). The supernatant was removed and the pellet air dried before being solubilised in 100 µL 2% SDS in PBS. For the membrane fraction, the membrane pellet was washed with HEPES buffer (X2)

collecting the pellet by centrifugation (17,000 g, 5 minutes). The membranes were then solubilised in 100 µL 2% SDS in PBS. All fractions were stored at -80 ºC until future use.

6.3.7.3. Sonication

For lysates requiring sonication, a microtipped sonicator probe (Fisher Scientific) was used. Samples were sonicated on ice for a total duration of 20 seconds with 2 seconds on, 3 seconds off at an amplitude of 20%.

6.3.8. In-cell Probe Labelling

For labelling experiments, cells were grown to 90-95% confluency prior to treatment, media aspirated and cells washed with PBS. For competition experiments, the probe competitor was diluted in DMEM (without supplemented FBS), added to the cells and returned to the incubator for 1 hour. The competitor containing media was aspirated and the cells washed with PBS. Probe containing media was prepared by diluting the probe in DMEM (without supplemented FBS) and vortexed. The probe containing media was added to the cells and returned to the incubator for 2 hours (unless otherwise stated). No probe controls were prepared with DMSO containing media.

6.3.9. In-lysate Probe Labelling

Lysates were prepared as described previously. Probe dilutions were prepared in DMSO before being added to the lysate. The lysates were incubated for 1 hour with shaking at room temperature before being photocrosslinked.

6.3.10. Photocrosslinking

Photocrosslinking was performed using a LED light box with wavelength 365 nm constructed in-house by Charlie Saunders (Imperial College/WaveyTech). Samples were irradiated either in a well plate or in Eppendorfs for 1 minute unless otherwise stated.

6.3.11. Cu(I)-catalysed Alkyne-Azide Cycloaddition (CuAAC) click chemistry

A click chemistry cocktail was prepared fresh as follows:

Reagent	Stock	Final Concentration	Volume (for 100 µL of lysate)
Capture reagent	10 mM in DMSO	100 µM	1 µL

(AzT, AzTB or AzRB)			
$CuSO_4$	50 mM in H_2O	1 mM	2 µL
TCEP	50 mM in H_2O	1 mM	2 µL
TBTA	10 mM in DMSO	100 µM	1 µL

6 µL of the mixture was added to 94 µL of lysate with a protein concentration between 1-2 mg/mL. The reactions were shaken at room temperature for 1 hour before being quenched with 5 mM EDTA. The proteins were precipitated using a MeOH/$CHCl_3$ regime. Briefly, 2 volumes of MeOH, 0.5 volumes $CHCl_3$ and 1 volume H_2O was added to the click mixture and vortexed before being centrifuged (10,000 g, 5 minutes) to produce a pellet suspended between the aqueous and organic phase. Carefully, the solvents were removed by decanting or pipetting and 1 mL MeOH was added. The pellets were broken up by vortexing and sonication before being pelleted by centrifugation (17,000 g, 5 minutes). The MeOH was removed and the process repeated once more. The pellet was air-dried and dissolved in 0.1 volumes of 2% SDS in PBS with vortexing and sonication. The solution was diluted stepwise to 0.2% SDS with 0.9 volumes of PBS. A final centrifugation step (17,000 g, 5 minutes) was performed to remove any insoluble material. Clicked lysates were stored at -20 °C until future use.

6.3.12. HaloTag labelling

TAMRA-Cl HaloTag ligand was synthesised by Maria Shchepinova as previously reported.[302] All other reported HaloTag reagents, antibodies and resins were purchased from Promega. In order to label HaloTag proteins, cells were transfected as described before (*6.3.3*).

6.3.12.1. Fluorescent Labelling in Lysate

Lysate was prepared as previously described. A working stock of fluorophore was prepared by dilution of the stock in DMSO (TAMRA-Cl 100 µM, Alexa660-HaloTag-Cl 35 µM) and diluted 1:20 into the lysate to give the final fluorophore concentration desired. Samples were labelled for 1 hour with gentle shaking at room temperature. Samples were either precipitated to remove excess dye or mixed with SDS Sample Loading buffer and run directly on a gel for imaging.

6.3.12.2. Fluorescent Labelling in Cells in Culture

Prior to transfection, cells were seeded on Poly D-Lysine coated plates. Dilutions of the desired fluorophore in OptiMEM media (1:200) were prepared from the provided DMSO stocks to create a 5x fluorophore solution. The culture media was aspirated from the cells and replaced with 0.8 volumes of OptiMEM media. 0.2 volumes of fluorophore containing media was then carefully added and the plates returned to the incubator for 60 minutes. The cells were then washed with DMEM supplemented with 10% FBS (X3) before being lysed.

6.3.12.3. Biotin-PEG-Cl Labelling in Lysates

Cells were grown, transfected and lysed as described previously. Lysates were pre-cleared on streptavidin beads to remove any endogenously biotinylated proteins for 1 hour. The supernatant was then incubated with 1 μM Biotin-PEG-Cl (1:100 dilution from a 100 μM DMSO working stock) with gentle shaking for 2 hours at room temperature. The protein was precipitated to remove excess reagent and the protein pellet solubilised in 0.2% SDS in PBS for further analysis.

6.3.12.4. Biotin-PEG-Cl Labelling in Cells in Culture

Prior to transfection, cells were seeded on Poly D-Lysine coated plates. A 1:1000 dilution of Biotin-PEG-Cl (5 mM stock) reagent was prepared in DMEM supplemented with 10% FBS. 200 μL was added to the desired well and incubated for the required amount of time. The cells were then washed with PBS (x2), fresh culture media was added to the cells and the plate returned to the incubator for 1 hour. The media was aspirated, cells washed with PBS and lysed as before.

6.3.13. Affinity Enrichment

Prior to enrichment, lysates were prepared as previously described from cells either expressing a tagged protein or from cells/lysates modified chemically to include an affinity tag.

6.3.13.1. Streptavidin-Biotin

For 100 μg of lysate, 10 μL of Dynabeads® MyOne™ Streptavidin C1 magnetic beads were used. The beads were first washed with 0.2% SDS in PBS (x3) by vortexing for 1 minute and recovering the beads with a magnetic rack. 90% of the biotinylated cell lysate was then added to the beads and gently vortex for 2 hours at RT. The remaining

10% of lysate was retained as a before pull down sample to be analysed on gel alongside the enriched fraction. The supernatant was also saved and the beads washed with 0.2% SDS in PBS (x3). The final wash was removed, and 20 µL of 2% SDS in PBS and 5 µL of 4x SDS sample loading buffer (containing 5% (v/v) β-mercaptoethanol final concentration). To elute from the beads, the samples were gently vortexed and then incubated at 95 ºC for 10 minutes, and the supernatant analysed by gel.

6.3.13.2. HaloLink

200 µL of lysate (1-2 mg/mL) was diluted with 800 TBS µL. For each sample, 200 µL of HaloLink Resin suspension (corresponding to 50 µL of resin) was dispensed into a 1.5 mL Eppendorf, centrifuged for 1 minute and the suspension buffer discarded. The resin was washed with wash buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.005% IGEPAL CA-630) three times, each time vortexing for 1 minute and then pelleting the resin by centrifugation. The diluted lysates were added to the resin and the samples incubated with inversion at 4 ºC for 1 hour.

6.3.14. In-gel Fluorescence

For samples labelled with a fluorophore, samples were mixed with a 4x solution of SDS sample loading buffer (containing 5% (v/v) β-mercaptoethanol final concentration) and the mixtures either incubated at room temperature for 10 minutes with agitation, heated to 50 ºC with agitation for 10 minutes or boiled at 95 ºC for 5 minutes. Samples were resolved by SDS-PAGE and visualised by in-gel fluorescence. Gel loading was visualised using Coomassie Blue stain.

6.3.15. Western Blot Analysis

After SDS-PAGE (and in-gel fluorescence analysis if necessary), gels were transferred to nitrocellulose membrane using a wet-transfer system. Transfer cassettes were prepared in fresh transfer buffer (25 mM Tris, 192 mM Glycine and 20% MeOH) and proteins transferred to the membrane at 100 V for 1 hour. After transfer, membranes were briefly checked for successful transfer using Panceau S stain before being rinsed with TSBT (50 mM Tris, 150 mM NaCl, 0.1% Tween-20, pH 7.5) and blocked with blocking buffer (**Table 8**) for 1 hour at room temperature or 4 ºC overnight. The blocking solution was decanted and replaced with the primary antibody solution and incubated for 1 hour at room temperature or 4 ºC overnight. The membrane was then

washed (3 x 10 minutes, TBST) before being imaged (for HRP conjugated antibodies) or incubated with the appropriate secondary antibody for 1 hour at RT. The membranes were then washed (3 x 10 minutes, TBST) and the chemiluminescence imaged using Luminata crescendo Western HRP substrate (Millipore).

Antibody	Source	Blocking buffer	Dilution	Secondary
Primary				
GPR119	Santa Cruz Sc-99103	5% skimmed milk in TBS-T	1:200	Rabbit
GPR119	Abcam Ab75312	5% skimmed milk in TBS-T	1:250 (IF) 1:500 (WB)	Rabbit
FLAG-M2	Sigma	10% BSA in PBS	1:50 (IF)	Mouse
HaloTag	Promega G921A	5% BSA in TBS-T	1:1000	Mouse
GAPDH	Abcam Ab9485	5% skimmed milk in TBS-T	1:5000	Rabbit
β-Actin	Abcam Ab6276	5% skimmed milk in TBS-T	1:20000	Mouse
Na^+/K^+ ATPase	Abcam Ab7671	5% skimmed milk in TBS-T	1:5000	Mouse
HSP90	Santa Cruz Sc-69703	5% skimmed milk in TBS-T	1:500	Mouse
HRP-Conjugate				
FLAG-M2-HRP	Sigma A8592	5% skimmed milk in TBS-T	1:500	
NeautrAvidin-HRP	Alpha Diagnostics 20366	3% BSA in TBS-T	1:2000	
Streptactin-HRP	BioRad	3% BSA in TBS-T	1:5000	
Secondary				
Mouse-HRP	Advansta R-05071-500	5% skimmed milk in TBS-T	1:10000	
Rabbit-HRP	Advansta R-05072-500	5% skimmed milk in TBS-T	1:10000	
Mouse-AF594	Abcam	10% BSA in PBS	1:800	
Rabbit-AF488	Abcam	10% BSA in PBS	1:500	

Table 8: Antibodies used in this study.

6.3.16. Immunofluorescence

To investigate transfected cells: the day prior to transfection, cells were seeded onto Poly- D-Lysine coverslips within a 6-well plate at 2-3 x 10^5 cells per well. Cells were

transfected using methods previously described. 3 hours post transfection, media was exchanged for DMEM +10% FBS.

After 48 hours, the media was removed from the wells and the cells were carefully washed with PBS (×3). The cells were fixed using 4% paraformaldehyde (PFA) for 15 minutes at room temperature using approximately 1mL per coverslip. The coverslips were carefully rinsed with PBS twice before being neutralised with 50 mM ammonium chloride in PBS for 20 minutes using 2 mL per coverslip. The coverslips were rinsed twice with PBS before either being stored overnight (at 4 ^{o}C) or continuing with the protocol.

For cells requiring permeabilisation, coverslips were coated with 0.2% Triton X-100 in PBS for 5 minutes using 2 mL per coverslip. The coverslips were then blocked with 10% BSA in PBS for 30 minutes using 1 mL per coverslip. The first antibody was prepared in 5% BSA in PBS and 100 µL was carefully pipetted onto each coverslip before being covered with parafilm and incubated at 4 ^{o}C overnight. The coverslips were washed with PBS for 4 minutes (×3) before incubation with the secondary antibody. The antibody dilution was prepared in 5% BSA in PBS before pipetting 100 µL onto the coverslip and covering with parafilm. The coverslip was incubated for 20 minutes before rinsing with PBS for 4 minutes (×3). Using tweezers, the coverslip was rinsed with PBS, dipped into a pool of PBS, then MilliQ water before the excess water was dried using filter paper. Coverslips were mounted onto slides using 50 µL of moviol (containing nuclear stain) before being covered with filter paper and allowed to dry overnight. Coverslips were imaged using an iRiS™ Digital Cell Imaging System fluorescent microscope.

6.3.17. qRT-PCR

Cell culture, disruption and homogenisation

The day prior to the experiment, cells were plated in a 6-well plate and left overnight. The following day, the media was removed and the cells washed with 1 mL PBS. The cells were disrupted by adding 350 µL Buffer RLT (RNease minikit from QIAGEN) with 1% β-mercaptoethanol to each well. The cells were pipetted to mix and transferred to a QIAshredder spin column placed in a 2 mL collection tube before being centrifuged for 2 minutes at full speed. 340 µL of 70% EtOH (on HyPure Water) was added to the lysate and mixed well by pipetting before being transferred to a RNeasy spin column

placed in a 2 mL collection tube. The lid was closed the column centrifuged for 30 seconds at 9000 x g. The flow through was discarded. The column was washed with the following buffers at 9000 x g, discarding flow through each time: 700 µL buffer RW1 for 30 seconds, 500 µL buffer RPE for 30 seconds and 500 µL buffer RPE for 2 minutes. The column was placed in a fresh 2 mL collection tube before centrifuging at full speed for 1 minute. The column was placed in a fresh 1.5 mL collection tube and 40 µL of RNase-free water was added. The column was centrifuged for 1 minute at 9000 x g to eluate the RNA. The concentration of RNA was calculated using a Nanodrop.

Reverse transcription (rT) to cDNA

500 ng of RNA was diluted to 10 µL in a 0.2 mL PCR tube. A 2× rT master mix was prepared on ice:

Component	**µL**
10x rT Reaction buffer	14
25x dNTP mix	5.6
10x rT random primers	14
MultiScribe™ Reverse Transcriptase	7
HyPure H_2O	29.4
Total	70

10 µL of 2× rT mix was added to each RNA sample, mixed by pipetting and briefly centrifuged before being subjected to the following thermo cycle:

Temperature (°C)	**Time (minutes)**
25	10
37	120
85	5
10	hold

The resulting cDNA was diluted 1:5 with HyPure water.

rtPCR

Gene	**Primer sequence**	
GPR119	Forward	TCTCGGCCCACACAGAAGA
	Reverse	GCTGCGGAGGAAGTGACAAA
GAPDH	Forward	CAACTTTGGTATCGTGGAAGGAC
	Reverse	ACAGTCTTCTGGGTGGCAGTG

Primer pairs to be used were diluted with H_2O to 5 µM of each in the solution. PCR mix was prepared with the following ratios per well: 1 µL of primer mix, 5 µL SybrGreen. 6 µL of this mix was added to each well followed by 4 µL of cDNA (or water) and the well plate sealed with an adhesive cover before being centrifuged at 1000 rpm for 5 minutes.

The plate was subjected to the following qPCR protocol on a Mx3000P QPCR system using MxPro rtPCR software:

Stage	Temperature (°C)	Time	
1.	95	10 minutes	
2.	95	15 seconds	40 cycles
	60	30 seconds	
	72	30 seconds	
3.	95	1 minute	
4.	55	30 seconds	
5.	95	30 seconds	

Raw data was exported and the $2^{\Delta\Delta Ct}$ value calculated using the following equation:

$$2^{\Delta\Delta Ct} = 2^{Ct(WT[GPR119-GAPDH]-OE[GPR119-GAPDH])}$$

Where:

Ct is the cycle threshold value

WT represents counts obtained from wild type cells

OE represents counts obtained from overexpressing cells

GAPDH is acting as a housekeeping gene

6.4. Biochemical assays

6.4.1. General methods

Reagents and materials for the cAMP assay were obtained from Perkin Elmer. Either the LANCE cAMP kit or the LANCE Ultra cAMP kit was used, following the manufacturer's instructions. For each assay run, a standard curve using the provided cAMP standard solution was prepared using 12 serial dilutions from 1 µM to 0. Stimulation buffer was prepared fresh for each assay run: 5 mM HEPES. 0.5 mM IBMX, 0.1% BSA in Hanks Balanced Buffer Solution (HBBS) at pH 7.4. Plates were read using an EnVision plate reader with a UV2 320 excitation filter and two emission

filters (203 Eu 615 and 205 APC 665). Assays were conducted in OptiPlate-384 wells, with a total volume of 20 µL.

6.4.2. cAMP assay lysate adaptation

Cells were plated in 96-well plates to achieve 70-80% confluence upon stimulation (approximately 25,000 cells per well). On the day of stimulation, the media was replaced with 50 µL per well of stimulation media (0.5 mM IBMX, 5mg/mL BSA in FBS free DMEM) and the plate incubated for 30 minutes. Ligand dilutions (2x desired stimulation concentration) were prepared in stimulation buffer and kept in water bath until use. 50 µL of the 2x ligand containing solution was added carefully to the each well and the cells stimulated for 30 minutes. The plate was quenched by placing on ice, before removing the stimulation media. The cells were lysed with 50 µL of detection buffer (as provided in the LANCE Ultra cAMP kit) for 30 minutes at 4 ^{o}C with gentle agitation. The plates were then centrifuged briefly to obtain the cAMP containing supernatant. Adaptation based on previously published protocol.[235]

The assay was then carried out as described in the manual for cell suspension stimulation. Briefly, 5 µL of lysate was added to each well followed by 5 µL of 4x U*Light* solution and 10 µL of 2x Eu-cAMP tracer solution. The plate was incubated for 60 minutes and then read on the EnVision plate reader.

6.5. Chemical proteomics

6.5.1. General methods

For proteomic sample preparation, LC-MS grade solvent, MilliQ water and proteomic grade water for final suspension were used. All buffers were prepared fresh and filtered prior to use. Dedicated solutions and fresh tips were used to minimise contamination of the samples. Washes were carried out by shaking samples for 2 minutes before pelleting beads by centrifugation. Lyophilised samples were stored at -80^{o}C until the time of analysis. Digestion was performed using sequencing grade modified proteases: trypsin (Promega), Proteinase K (New England Biolabs), elastase (Promega) or chymotrypsin (Promega) and 0.05% (w/v) ProteaseMax (Promega).

6.5.2. Neutravidin Enrichment

For biotinylated samples, lysates were prepared as described previously and suspended in 0.2% SDS in HEPES 50 mM. For each sample, 8 µL of Neutravidin agarose beads (Invitrogen) and 22 µL of blank agarose beads were combined and washed with 0.2% SDS in HEPES 50 mM pH 8.0, centrifuged to pellet the beads and washed a further two times. The beads were resuspended in 50 µL of 0.2% SDS in HEPES 50 mM pH 8.0. The previously prepared samples were centrifuged at maximum speed for 5 minutes, before the top 90% of the supernatant was added to the beads. With agitation, the beads were incubated at room temperature for 2 hours. The beads were pelleted using centrifugation at 7000 rpm for 4 minutes and the supernatant removed and retained. The beads were washed with 400 µL 0.2% SDS in HEPES 50 mM pH 8.0 two times, followed by four washes using 400 µL of 50 mM HEPES pH 8.0.

6.5.3. HaloLink Enrichment

Resin was prepared and lysates incubated as previously described. After 1 hour at 4 ^{o}C, the resin was pelleted by centrifugation and the supernatant removed and retained. 1 mL of resin wash buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.005% IGEPAL CA-630) was added to each Eppendorf and thoroughly inverted to wash. The pellet was obtained, and the supernatant discarded, and the wash repeated a further three times. A final extended wash was carried out with resin wash buffer for 5 minutes with constant agitation. The resin was then washed three times with 50 mM HEPES pH 8.0.

6.5.4. Reduction, alkylation and digestion

After washing, beads were resuspended in 50 µL 50 mM HEPES pH 8.0 containing 10 mM chloroacetamide (CAA) and 5 mM tris(2-carboxyethyl)phosphine (TCEP), sonicated for 5 minutes and incubated for 30 minutes with gentle shaking. Trypsin (or other stated protease) and ProteaseMax were added and the samples incubated overnight at 37 ^{o}C (unless stated otherwise). The samples were pelleted and the supernatant transferred to a LoBind Eppendorf. The beads were washed with 70 µL of 50 mM HEPES pH 8.0 and the supernatants combined. TFA was added to a final concentration of 0.5% (v/v).

6.5.5. Stage-tipping and sample preparation

Samples were desalted on polystyrene-divinylbenzene copolymer (SDB-XC, from 3M) disks following the Stage Tip protocol.[316] The peptides were eluted from the stage tip using 79% (v/v) acetonitrile in proteomic grade water and dried using speed-vac assisted solvent removal. Dried peptides were stored at -80^{o}C until reconstitution. Samples were resuspended in 0.5% (v/v) TFA and 2% (v/v) MeOH in proteomic grade water and solubilised using vortex shaking and sonication. Any remaining insoluble material was removed by stage-tip filtration through 3-layer PVDF Durapore filter (0.1 µM) using centrifugal force at 2000 rpm for 3 minutes into the final LC-MS vial.

6.5.6. LC-MS/MS analysis

Analysis was performed using an EASY-Spray Acclaim PepMap C_{18} column (50 cm x 75 µm inner diameter, Thermo Fisher Scientific) using a 2-hour linear acetonitrile gradient in 0.1% aqueous formic acid at a flow rate of 250 nL/min. The easy nLC-1000 was coupled to a QExactive mass spectrometer via an easy spray source (Thermo Fisher Scientific). The QExactive was operated in a data-dependent mode with survey scans acquired at a resolution of 75.000 at *m/z* 200 (transient time 256 ms). Up to 10 of the most abundant isotope patterns with a charge of 2+ or higher from the survey scan were selected with an isolation window of 3.0 *m/z* and fragmented by higher collision dissociation (HCD) with normalised collision energies of 25. The maximum ion injection times for the survey scan and the MS/MS scans (acquired with a resolution of 17,500 at *m/z* 200) were 20 and 120 ms respectively. The ion target value for MS was set to 10^6, and the intensity threshold was set to 8.3 x 10^2.

6.5.7. Data analysis

Experimental data was processed using MaxQuant software v1.6.0.1. The peptides were identified from the MS/MS spectra searched against the complete reviewed human proteome (UniProt) and the complete human proteome modified to include HaloTag-GPR119 using the in-built Andromeda search engine. Cysteine carbamidomethylation was selected as a fixed modification and methionine oxidation and N-terminal acetylation as variable modifications. The digestion enzyme used for the experiment was selected to dictate allowed cleavage sites. Up to two missed cleavages were allowed.

The data was analysed using Perseus v.1.5.0.9 and filtered to remove proteins categorised as ‘Only identified by site’, ‘Reverse’ and ‘Potential contaminant’. Replicates were grouped per sample and the intensities were log_2-transformed. Two-sample t-tests were performed.

For the identification of peptides, Peaks Studio 8.5 was used. The complete reviewed human proteome (Uniprot) with the addition of HaloTag-GPR119 were employed for the peptide identification search. Cysteine carbamidomethylation was selected as a fixed modification and methionine oxidation was selected as a variable modification. Precursor mass tolerance was set to 5 ppm, fragment ion mass tolerance was set to 0.01 Da with maximum allowable variable PTMs set to 5. Maximum number of missed cleavages was set to 5.

7. References

1. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–72 (2003).
2. Lagerström, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7**, 339–357 (2008).
3. Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. Seven-transmembrane receptors. *Nature Reviews Molecular Cell Biology* **3**, 639–650 (2002).
4. Seifert, R. & Wenzel-Seifert, K. Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **366**, 381–416 (2002).
5. Gether, U. *et al.* Structural instability of a constitutively active G protein-coupled receptor. Agonist-independent activation due to conformational flexibility. *J. Biol. Chem.* **272**, 2587–90 (1997).
6. Milligan, G. G protein-coupled receptor dimerization: function and ligand pharmacology. *Mol. Pharmacol.* **66**, 1–7 (2004).
7. Nelson, G. *et al.* An amino-acid taste receptor. *Nature* **416**, 199–202 (2002).
8. Nelson, G. *et al.* Mammalian sweet taste receptors. *Cell* **106**, 381–90 (2001).
9. Syrovatkina, V., Alegre, K. O., Dey, R. & Huang, X.-Y. Regulation, Signaling, and Physiological Functions of G-Proteins. *J. Mol. Biol.* **428**, 3850–3868 (2016).
10. Reed, R. R. G protein diversity and the regulation of signaling pathways. *New Biol.* **2**, 957–60 (1990).
11. Wang, W., Qiao, Y. & Li, Z. New Insights into Modes of GPCR Activation. *Trends Pharmacol. Sci.* **39**, 367–386 (2018).
12. Duc, N. M., Kim, H. R. & Chung, K. Y. Structural mechanism of G protein activation by G protein-coupled receptor. *Eur. J. Pharmacol.* **763**, 214–222 (2015).
13. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
14. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
15. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
16. Yang, J., Villar, V. A. M., Armando, I., Jose, P. A. & Zeng, C. G Protein–Coupled Receptor Kinases: Crucial Regulators of Blood Pressure. *J. Am. Heart Assoc.* **5**, (2016).
17. Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17–27 (2012).
18. Thathiah, A. & De Strooper, B. The role of G protein-coupled receptors in the pathology of Alzheimer's disease. *Nat. Rev. Neurosci.* **12**, 73–87 (2011).
19. Bar-Shavit, R. *et al.* G Protein-Coupled Receptors in Cancer. *Int. J. Mol. Sci.* **17**, 1320 (2016).
20. Hampp, C., Borders-Hemphill, V., Moeny, D. G. & Wysowski, D. K. Use of antidiabetic drugs in the U.S., 2003-2012. *Diabetes Care* **37**, 1367–74 (2014).

21. Palczewski, K. *et al.* Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science (80-.).* **289**, 739–745 (2000).

22. B. Gacasan, S., L. Baker, D. & L. Parrill, A. G protein-coupled receptors: the evolution of structural insight. *AIMS Biophys.* **4**, 491–527 (2017).

23. Rasmussen, S. G. F. *et al.* Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).

24. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science (80-.).* **318**, 1258–1265 (2007).

25. Crasto, C. Hydrophobicity profiles in G protein-coupled receptor transmembrane helical domains. *J. Receptor. Ligand Channel Res.* **2010**, 123 (2010).

26. Gómez Tamayo, J. C. *et al.* GPCR-SAS: A web application for statistical analyses on G protein-coupled receptors sequences. *PLoS One* **13**, e0199843 (2018).

27. Wright, G. J. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol. Biosyst.* **5**, 1405 (2009).

28. Böhme, I. & Beck-Sickinger, A. G. Illuminating the life of GPCRs. *Cell Communication and Signaling* **7**, 16 (2009).

29. Müskens, F. M. *et al.* Design, Synthesis, and Evaluation of a Diazirine Photoaffinity Probe for Ligand-Based Receptor Capture Targeting G Protein–Coupled Receptors. *Mol. Pharmacol.* **95**, 196–209 (2019).

30. Helbig, A. O., Heck, A. J. R. & Slijper, M. Exploring the membrane proteome--challenges and analytical strategies. *J. Proteomics* **73**, 868–78 (2010).

31. International Diabetes Foundation. IDF Diabetes Atlas. (2017).

32. World Health Organisation. Diabetes Fact Sheet. (2018). Available at: https://www.who.int/health-topics/news-room/fact-sheets/detail/diabetes. (Accessed: 8th April 2019)

33. Public Health England. Health Matters: Preventing Type 2 diabetes. *Public Health Matters* Available at: https://publichealthmatters.blog.gov.uk/2018/05/24/health-matters-preventing-type-2-diabetes/. (Accessed: 8th April 2019)

34. Berg, J. M. (Jeremy M., Tymoczko, J. L., Stryer, L. & Stryer, L. *Biochemistry*. (W.H. Freeman, 2002).

35. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **26**, s5–s20 (2003).

36. Hunter, S. J. & Garvey, W. T. T. Insulin action and insulin resistance: diseases involving defects in insulin receptors, signal transduction, and the glucose transport effector system. *Am. J. Med.* **105**, 331–345 (1998).

37. Cefalu, W. T., Waldman, S. & Ryder, S. Pharmacotherapy for the Treatment of Patients with Type 2 Diabetes Mellitus: Rationale and Specific Agents. *Clin Pharmacol Ther* **81**, 636–649 (2007).

38. Himsworth, H. P. Diabetes mellitus: Its differentiation into insulin-sensitive and insulin-insensitive types. *Diabet. Med.* **28**, 1440–1444 (2011).

39. Neeland, I. J. *et al.* Dysfunctional adiposity and the risk of prediabetes and type 2 diabetes in obese adults. *J. Am. Med. Assoc.* **308**, 1150–1159 (2012).

40. Eckel, R. H. *et al.* Obesity and type 2 diabetes: what can be unified and what needs to be individualized? *Diabetes Care* **34**, 1424–30 (2011).

41. Abdullah, A., Peeters, A., de Courten, M. & Stoelwinder, J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. *Diabetes Res. Clin. Pract.* **89**, 309–19 (2010).

42. Sami, W., Ansari, T., Butt, N. S. & Hamid, M. R. A. Effect of diet on type 2 diabetes mellitus: A review. *Int. J. Health Sci. (Qassim).* **11**, 65 (2017).

43. Stumvoll, M., Goldstein, B. J. & van Haeften, T. W. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* **365**, 1333–46 (2010).

44. Bogardus, C., Lillioja, S., Mott, D. M., Hollenbeck, C. & Reaven, G. Relationship between degree of obesity and in vivo insulin action in man. *Am. J. Physiol.* **248**, E286-91 (1985).

45. Kahn, S. E. The importance of the β-cell in the pathogenesis of type 2 diabetes mellitus. *Am. J. Med.* **108**, 2–8 (2000).

46. Jones, R. M., Leonard, J. N., Buzard, D. J. & Lehmann, J. GPR119 agonists for the treatment of type 2 diabetes. *Expert Opin. Ther. Pat.* **19**, 1339–59 (2009).

47. Sheehan, M. T. Current Therapeutic Options in Type 2 Diabetes Mellitus: A Practical Approach. *Clin. Med. Res.* **1**, 189–200 (2003).

48. Ripsin, C. M., Kang, H. & Urban, R. J. Management of blood glucose in type 2 diabetes mellitus. *Am. Fam. Physician* **79**, 29–36 (2009).

49. Ungar, G., Freedman, L. & Shapiro, S. L. Pharmacological Studies of a New Oral Hypoglycemic Drug. *Proc. Soc. Exp. Biol. Med.* **95**, 190–192 (1957).

50. Holman, R. Metformin as first choice in oral diabetes treatment: the UKPDS experience. *Journ. Annu. Diabetol. Hotel. Dieu.* 13–20 (2007).

51. Palmer, S. C. & Strippoli, G. F. M. Metformin as first-line treatment for type 2 diabetes. *Lancet (London, England)* **392**, 120 (2018).

52. Nathan, D. M. *et al.* Management of Hyperglycemia in Type 2 Diabetes: A Consensus Algorithm for the Initiation and Adjustment of Therapy: A consensus statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **29**, 1963–1972 (2006).

53. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* **352**, 854–865 (1998).

54. Garber Md, P. A. J., Duncan Md, T. G., Goodman Md, A. M., Mills Rn, B. S. N. D. J. & Rohlf Md, J. L. Efficacy of Metformin in Type II Diabetes: Results of a Double-Blind, Placebo-controlled, Dose-Response Trial. *Am. J. Med.* **103**, 491–497 (1997).

55. Nissen, S. E. & Wolski, K. Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *N. Engl. J. Med.* **356**, 2457–2471 (2007).

56. van de Laar, F. A. *et al.* α-Glucosidase Inhibitors for Patients With Type 2 Diabetes: Results from a Cochrane systematic review and meta-analysis. *Diabetes Care* **28**, 154–163 (2005).

57. Shah, U., Edmondson, S. & Szewczyk, J. W. Chapter 7. Recent Advances in the Discovery of GPR119 Agonists. in *New Therapeutic Strategies for Type 2 Diabetes: Small Molecule Approaches* 177–214 (2012).

58. Baggio, L. L. & Drucker, D. J. Biology of Incretins: GLP-1 and GIP. *Gastroenterology* **132**, 2131–2157 (2007).

59. Drucker, D. J. & Nauck, M. A. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet* **368**, 1696–

1705 (2006).

60. Goldstein, B. J., Feinglos, M. N., Lunceford, J. K., Johnson, J. & Williams-Herman, D. E. Effect of Initial Combination Therapy With Sitagliptin, a Dipeptidyl Peptidase-4 Inhibitor, and Metformin on Glycemic Control in Patients With Type 2 Diabetes. *Diabetes Care* **30**, 1979–1987 (2007).

61. Fredriksson, R. *et al.* Seven evolutionarily conserved human rhodopsin G protein-coupled receptors lacking close relatives. *FEBS Lett.* **554**, 381–388 (2003).

62. Overton, H. A. *et al.* Deorphanization of a G protein-coupled receptor for oleoylethanolamide and its use in the discovery of small-molecule hypophagic agents. *Cell Metab.* **3**, 167–75 (2006).

63. Soga, T. *et al.* Lysophosphatidylcholine enhances glucose-dependent insulin secretion via an orphan G-protein-coupled receptor. *Biochem. Biophys. Res. Commun.* **326**, 744–51 (2005).

64. Chu, Z.-L. L. *et al.* A role for beta-cell-expressed G protein-coupled receptor 119 in glycemic control by enhancing glucose-dependent insulin release. *Endocrinology* **148**, 2601–9 (2007).

65. Itoh, Y. *et al.* Free fatty acids regulate insulin secretion from pancreatic [beta] cells through GPR40. *Nature* **422**, 173–176 (2003).

66. Hirasawa, A. *et al.* Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through GPR120. *Nat Med* **11**, 90–94 (2005).

67. Overton, H. A., Fyfe, M. C. T. & Reynet, C. GPR119, a novel G protein-coupled receptor target for the treatment of type 2 diabetes and obesity. *Br. J. Pharmacol.* **153 Suppl**, S76-81 (2008).

68. Bonini, J. A., Borowsky, B. E., Adham, N., Boyle, N. & Thompson, T. O. DNA encoding SNORF25 receptor. **09387699**, (2001).

69. Bonini, J. A., Borowsky, B. E., Adham, N., Boyle, N. & Thompson, T. O. Methods of identifying compounds that bind to SNORF25 receptors. **09641259**, (2002).

70. Chu, Z.-L. L. *et al.* A role for intestinal endocrine cell-expressed g protein-coupled receptor 119 in glycemic control by enhancing glucagon-like Peptide-1 and glucose-dependent insulinotropic Peptide release. *Endocrinology* **149**, 2038–47 (2008).

71. Parker, H. E., Habib, A. M., Rogers, G. J., Gribble, F. M. & Reimann, F. Nutrient-dependent secretion of glucose-dependent insulinotropic polypeptide from primary murine K cells. *Diabetologia* **52**, 289–298 (2009).

72. Jones, R. M. *et al.* 1,2,3-Trisubstituted aryl and heteroaryl derivatives as modulators of metabolism and the prophylaxis and treatment of disorders related thereto such as diabetes and hyperglycemia. **US20040012**, (2004).

73. Takeda, S., Kadowaki, S., Haga, T., Takaesu, H. & Mitaku, S. Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett.* **520**, 97–101 (2002).

74. Davey, J. G-Protein-Coupled Receptors: New Approaches to Maximise the Impact of GPCRs in Drug Discovery. *Expert Opin. Ther. Targets* **8**, 165–170 (2004).

75. Costanzi, S., Neumann, S. & Gershengorn, M. C. Seven Transmembrane-spanning Receptors for Free Fatty Acids as Therapeutic Targets for Diabetes Mellitus: Pharmacological, Phylogenetic, and Drug Discovery Aspects. *J. Biol. Chem.* **283**, 16269–16273 (2008).

76. Brown, A. J. Novel cannabinoid receptors. *Br. J. Pharmacol.* **152**, 567–575 (2007).

77. Cohen, L. J. *et al.* Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nat. Publ. Gr.* **549**, (2017).

78. Irving, A. *et al.* Cannabinoid Receptor-Related Orphan G Protein-Coupled Receptors. *Adv. Pharmacol.* **80**, 223–247 (2017).

79. Rodríguez de Fonseca, F. *et al.* An anorexic lipid mediator regulated by feeding. *Nature* **414**, 209–212 (2001).

80. Nielsen, M. J., Petersen, G., Astrup, A. & Hansen, H. S. Food intake is inhibited by oral oleoylethanolamide. *J. Lipid Res.* **45**, 1027–1029 (2004).

81. Oveisi, F., Gaetani, S., Eng, K. T.-P. & Piomelli, D. Oleoylethanolamide inhibits food intake in free-feeding rats after oral administration. *Pharmacol. Res.* **49**, 461–466 (2004).

82. Gaetani, S., Oveisi, F. & Piomelli, D. Modulation of Meal Pattern in the Rat by the Anorexic Lipid Mediator Oleoylethanolamide. *Neuropsychopharmacology* **28**, 1311–1316 (2003).

83. Fu, J. *et al.* Oleylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR-α. *Nature* **425**, 90–93 (2003).

84. Chu, Z.-L. L. *et al.* N-oleoyldopamine enhances glucose homeostasis through the activation of GPR119. *Mol. Endocrinol.* **24**, 161–70 (2010).

85. Lauffer, L. M., Iakoubov, R. & Brubaker, P. L. GPR119 is essential for oleoylethanolamide-induced glucagon-like peptide-1 secretion from the intestinal enteroendocrine L-cell. *Diabetes* **58**, 1058–66 (2009).

86. Yore, M. M. *et al.* Discovery of a Class of Endogenous Mammalian Lipids with Anti-Diabetic and Anti-inflammatory Effects. *Cell* **159**, 318–332 (2014).

87. Hansen, K. B. *et al.* 2-Oleoyl Glycerol Is a GPR119 Agonist and Signals GLP-1 Release in Humans. *J. Clin. Endocrinol. Metab.* **96**, E1409–E1417 (2011).

88. Arifin, S. A. *et al.* Oleoyl-lysophosphatidylinositol enhances glucagon-like peptide-1 secretion from enteroendocrine L-cells through GPR119. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1863**, 1132–1141 (2018).

89. Doyle, M. E. & Egan, J. M. Mechanisms of action of glucagon-like peptide 1 in the pancreas. *Pharmacol. Ther.* **113**, 546–593 (2007).

90. Holst, J. J. & Gromada, J. Role of incretin hormones in the regulation of insulin secretion in diabetic and nondiabetic humans. *Am. J. Physiol. Metab.* **287**, E199–E206 (2004).

91. Shibasaki, T. *et al.* Essential role of Epac2/Rap1 signaling in regulation of insulin granule dynamics by cAMP. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19333–8 (2007).

92. Renström, E., Eliasson, L. & Rorsman, P. Protein kinase A-dependent and -independent stimulation of exocytosis by cAMP in mouse pancreatic B-cells. *J. Physiol.* **502**, 105–118 (1997).

93. Vilsbøll, T. The effects of glucagon-like peptide-1 on the beta cell. *Diabetes, Obes. Metab.* **11**, 11–18 (2009).

94. Shah, U. & Kowalski, T. J. Chapter Sixteen - GPR119 Agonists for the Potential Treatment of Type 2 Diabetes and Related Metabolic Disorders. in *Vitamins & Hormones* (ed. Gerald, L.) **Volume 84**, 415–448 (Academic Press, 2010).

95. Dhayal, S. & Morgan, N. G. The significance of GPR119 agonists as a future treatment for type 2 diabetes. *Drug News Perspect.* **23**, 418 (2010).

96. Semple, G. *et al.* Discovery of the first potent and orally efficacious agonist of the orphan G-protein coupled receptor 119. *J. Med. Chem.* **51**, 5172–5 (2008).

97. Semple, G. *et al.* Discovery of fused bicyclic agonists of the orphan G-protein coupled receptor GPR119 with in vivo activity in rodent models of glucose control. *Bioorg. Med. Chem. Lett.* **21**, 3134–41 (2011).

98. Fyfe, M. C., McCormack, J. G., Overton, H. A., Procter, M. J. & Reynet, C. GPR119 agonists as potential new oral agents for the treatment of type 2 diabetes and obesity. *Expert Opin. Drug Discov.* **3**, 403–13 (2008).

99. Zhu, X., Huang, W. & Qian, H. GPR119 Agonists : A Novel Strategy for Type 2 Diabetes Treatment. **119**, (2013).

100. Yang, J. W., Kim, H. S., Choi, Y. W., Kim, Y. M. & Kang, K. W. Therapeutic application of GPR119 ligands in metabolic disorders. *Diabetes, Obes. Metab.* **20**, 257–269 (2018).

101. Ritter, K., Buning, C., Halland, N., Pöverlein, C. & Schwink, L. G Protein-Coupled Receptor 119 (GPR119) Agonists for the Treatment of Diabetes: Recent Progress and Prevailing Challenges. *J. Med. Chem.* **59**, 3579–3592 (2016).

102. Kang, S.-U. GPR119 agonists: a promising approach for T2DM treatment? A SWOT analysis of GPR119. *Drug Discov. Today* **18**, 1309–1315 (2013).

103. Oshima, H. *et al.* Novel GPR119 agonist AS1669058 potentiates insulin secretion from rat islets and has potent anti-diabetic effects in ICR and diabetic db/db mice. *Life Sci.* **92**, 167–173 (2013).

104. Brandt, P. & G, J. Pyrimidine compounds for treating GPR119 related disorders. WO2008025800 (2008).

105. Wacker, D. A. *et al.* Discovery of 5-chloro-4-((1-(5-chloropyrimidin-2-yl)piperidin-4-yl)oxy)-1-(2-fluoro-4-(methylsulfonyl)phenyl)pyridin-2(1 H)-one (BMS-903452), an antidiabetic clinical candidate targeting GPR119. *J. Med. Chem.* **57**, 7499–7508 (2014).

106. Matsumoto, K. *et al.* DS-8500a, an Orally Available G Protein-Coupled Receptor 119 Agonist, Upregulates Glucagon-Like Peptide-1 and Enhances Glucose-Dependent Insulin Secretion and Improves Glucose Homeostasis in Type 2 Diabetic Rats. *J. Pharmacol. Exp. Ther.* **367**, 509–517 (2018).

107. Kato, M. *et al.* Safety and Pharmacokinetics of DS-8500a, a Novel GPR119 Agonist, After Multiple Oral Doses in Healthy Japanese Males. *Clin. Drug Investig.* 1–7 doi:10.1007/s40261-018-0633-5

108. Terauchi, Y. *et al.* Efficacy and safety of the G protein-coupled receptor 119 agonist DS-8500a in Japanese type 2 diabetes mellitus patients with inadequate glycemic control on sitagliptin: A phase 2 randomized placebo-controlled study. *J. Diabetes Investig.* **9**, 1333–1341 (2018).

109. Nunez, D. J. *et al.* Gut Hormone Pharmacology of a Novel GPR119 Agonist (GSK1292263), Metformin, and Sitagliptin in Type 2 Diabetes Mellitus: Results from Two Randomized Studies. *PLoS One* **9**, e92494 (2014).

110. Polli, J. W. *et al.* Evaluation of drug interactions of GSK1292263 (a GPR119 agonist) with statins: from *in vitro* data to clinical study design. *Xenobiotica* **43**, 498–508 (2013).

111. Zhu, X. *et al.* The first pharmacophore model for potent G protein-coupled receptor 119 agonist. *Eur. J. Med. Chem.* **46**, 2901–2907 (2011).

112. McClure, K. F. *et al.* Activation of the G-Protein-Coupled Receptor 119: A Conformation-Based Hypothesis for Understanding Agonist Response. *J. Med. Chem.* **54**, 1948–1952 (2011).

113. Norn, C. *et al.* Mutation-Guided Unbiased Modeling of the Fat Sensor GPR119 for High-Yield Agonist Screening. *Structure* **23**, 2377–2386 (2015).

114. Engelstoft, M. S. *et al.* Structural basis for constitutive activity and agonist-induced

activation of the enteroendocrine fat sensor GPR119. *Br. J. Pharmacol.* **171**, 5774–5789 (2014).

115. Geurink, P. P., Prely, L. M., Van Der Marel, G. A., Bischoff, R. & Overkleeft, H. S. Photoaffinity labeling in activity-based protein profiling. *Topics in Current Chemistry* **324**, 85–113 (2012).
116. Lapinsky, D. J. Tandem photoaffinity labeling-bioorthogonal conjugation in medicinal chemistry. *Bioorg. Med. Chem.* **20**, 6237–47 (2012).
117. Kaiser, N. *et al.* Photoactivatable Myristic Acid Probes for UNC119-Cargo Interactions. *ChemBioChem* **20**, 134–139 (2019).
118. Yang, T., Li, X. M., Bao, X., Fung, Y. M. E. & Li, X. D. Photo-lysine captures proteins that bind lysine post-translational modifications. *Nat. Chem. Biol.* **12**, 70–72 (2016).
119. Murale, D. P., Hong, S. C., Haque, M. M. & Lee, J.-S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **15**, 14 (2016).
120. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nature Chemical Biology* **5**, 616–624 (2009).
121. Liang, J. *et al.* Chemical Synthesis of Diubiquitin-Based Photoaffinity Probes for Selectively Profiling Ubiquitin-Binding Proteins. *Angew. Chemie Int. Ed.* **56**, 2744–2748 (2017).
122. Shi, H., Liu, K., Xu, A. & Yao, S. Q. Small molecule microarray-facilitated screening of affinity-based probes (AfBPs) for gamma-secretase. *Chem. Commun. (Camb).* 5030–2 (2009). doi:10.1039/b910611a
123. Kalesh, K. a, Shi, H., Ge, J. & Yao, S. Q. The use of click chemistry in the emerging field of catalomics. *Org. Biomol. Chem.* **8**, 1749–62 (2010).
124. Singh, A., Thornton, E. R. & Westheimer, F. H. The photolysis of diazoacetylchymotrypsin. *J Biol Chem* **237**, 3006–3008 (1962).
125. Wright, M. H. & Sieber, S. A. Chemical proteomics approaches for identifying the cellular targets of natural products. *Nat. Prod. Rep.* **33**, 681–708 (2016).
126. Dormán, G. Photoaffinity Labeling in Biological Signal Transduction. in 169–225 (Springer, Berlin, Heidelberg, 2000). doi:10.1007/3-540-45035-1_6
127. Dormán, G. & Prestwich, G. D. Using photolabile ligands in drug discovery and development. *Trends Biotechnol.* **18**, 64–77 (2000).
128. Dubinsky, L., Krom, B. P. & Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg. Med. Chem.* **20**, 554–70 (2012).
129. Robinette, D., Neamati, N., Tomer, K. B. & Borchers, C. H. Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. *Expert Rev. Proteomics* **3**, 399–408 (2006).
130. Sadakane, Y. & Hatanaka, Y. Photochemical fishing approaches for identifying target proteins and elucidating the structure of a ligand-binding region using carbene-generating photoreactive probes. *Anal. Sci.* **22**, 209–18 (2006).
131. Sumranjit, J., Chung, S., Sumranjit, J. & Chung, S. J. Recent Advances in Target Characterization and Identification by Photoaffinity Probes. *Molecules* **18**, 10425–10451 (2013).
132. Brunner, J. New photolabeling and crosslinking methods. *Annu Rev Biochem* **62**, 483–514 (1993).
133. Panikkanvalappil, S. R., Hira, S. M. & El-Sayed, M. A. Elucidation of ultraviolet radiation-

induced cell responses and intracellular biomolecular dynamics in mammalian cells using surface-enhanced Raman spectroscopy. *Chem. Sci.* **7**, 1133–1141 (2016).

134. Santos, A. L. *et al.* Effects of UV Radiation on the Lipids and Proteins of Bacteria Studied by Mid-Infrared Spectroscopy. *Environ. Sci. Technol.* **47**, 6306–6315 (2013).

135. Pattison, D. I. & Davies, M. J. Actions of ultraviolet light on cellular structures. in *Cancer: Cell Structures, Carcinogens and Genomic Instability* 131–157 (Birkhäuser-Verlag, 2006). doi:10.1007/3-7643-7378-4_6

136. Fleming, S. A. & A. Fleming, S. Chemical reagents in photoaffinity labeling. *Tetrahedron* **51**, 12479–12520 (1995).

137. Fleet, G. W. J., Porter, R. R. & Knowles, J. R. Affinity labelling of antibodies with aryl nitrene as reactive group [8]. *Nature* **224**, 511–512 (1969).

138. Dorman, G. & Prestwich, G. D. Benzophenone photophores in biochemistry. *Biochemistry* **33**, 5661–5673 (1994).

139. Kotzyba-Hibert, F., Kapfer, I. & Goeldner, M. Recent Trends in Photoaffinity Labeling. *Angew. Chemie Int. Ed. English* **34**, 1296–1312 (1995).

140. Buchmueller, K. L., Hill, B. T., Platz, M. S. & Weeks, K. M. RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile. *J. Am. Chem. Soc.* **125**, 10850–10861 (2003).

141. Staros, J. V., Bayley, H., Standring, D. N. & Knowles, J. R. Reduction of aryl azides by thiols: Implications for the use of photoaffinity reagents. *Biochem. Biophys. Res. Commun.* **80**, 568–572 (1978).

142. Bergeron, R. J., Dionis, J. B. & Ingeno, M. J. Synthesis of a parabactin photoaffinity label. *J. Org. Chem.* **52**, 144–149 (1987).

143. Smith, R. A. & Knowles, J. R. Letter: Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *J. Am. Chem. Soc.* **95**, 5072–3 (1973).

144. Brunner, J., Senn, H. & Richards, F. M. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J Biol Chem* **255**, 3313–3318 (1980).

145. Platz, M. *et al.* Photolysis of 3-aryl-3-(trifluoromethyl)diazirines: a caveat regarding their use in photoaffinity probes. *Bioconjug. Chem.* **2**, 337–41

146. Hill, J. R. & Robertson, A. A. B. Fishing for Drug Targets: A Focus on Diazirine Photoaffinity Probe Synthesis. *J. Med. Chem.* **61**, 6945–6963 (2018).

147. Das, J. Aliphatic diazirines as photoaffinity probes for proteins: recent developments. *Chem. Rev.* **111**, 4405–17 (2011).

148. Galardy, R. E., Craig, L. C. & Printz, M. P. Benzophenone triplet: a new photochemical probe of biological ligand-receptor interactions. *Nat. New Biol.* **242**, 127–8 (1973).

149. Xia, Y. & Peng, L. Photoactivatable lipid probes for studying biomembranes by photoaffinity labeling. *Chemical Reviews* **113**, 7880–7929 (2013).

150. Vodovozova, E. L. Photoaffinity Labelling and Its Application in Structural Biology. *Biochem.* **72**, 1–20 (2007).

151. Kohn, M., Breinbauer, R., Köhn, M. & Breinbauer, R. The Staudinger ligation-a gift to chemical biology. *Angew. Chem. Int. Ed. Engl.* **43**, 3106–16 (2004).

152. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A Comparative Study of Bioorthogonal Reactions with Azides. *ACS Chem. Biol.* **1**, 644–648 (2006).

153. Staudinger, H. & Meyer, J. Über neue organische Phosphorverbindungen III.

Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **2**, 635–646 (1919).

154. Gololobov, Y. Y. G. & Kasukhin, L. F. L. Recent advances in the Staudinger reaction. *Tetrahedron* **48**, 1353–1406 (1992).

155. Gololobov, Y. G. Y., Zhmurova, I. N. & Kasukhin, L. F. Sixty years of Staudinger reaction. *Tetrahedron* **37**, 437–472 (1981).

156. Lin, F. L. *et al.* Mechanistic Investigation of the Staudinger Ligation Investigation of the Kinetic Parameters of the Staudinger. 2686–2695 (2005).

157. Saxon, E. & Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science (80-.).* **287**, 2007–2010 (2000).

158. Gothelf, K. V & Jørgensen, K. A. Asymmetric 1,3-Dipolar Cycloaddition Reactions. *Chem. Rev.* **98**, 863–910 (1998).

159. Wittig, G. & Krebs, A. Zur Existenz niedergliedriger Cycloalkine, I. *Chem. Ber.* **94**, 3260–3275 (1961).

160. Rostovtsev, V. V, Green, L. G., Fokin, V. V & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **41**, 2596–9 (2002).

161. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128–37 (2003).

162. Speers, A. E., Cravatt, B. F. & Jolla, L. Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chem. Biol.* **11**, 535–546 (2004).

163. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686–7 (2003).

164. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–64 (2002).

165. Himo, F. *et al.* Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *J. Am. Chem. Soc.* **127**, 210–6 (2005).

166. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed. Engl.* **40**, 2004–2021 (2001).

167. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–7 (2004).

168. Meier, H., Petersen, H. & Kolshorn, H. Die Ringspannung von Cycloalkinen und ihre spektroskopischen Auswirkungen. *Chem. Ber.* **113**, 2398–2409 (1980).

169. Shea, K. J. & Kim, J. S. Influence of Strain on Chemical-Reactivity - Relative Reactivity of Torsionally Strained Double-Bonds in 1,3-Dipolar Cycloadditions. *J. Am. Chem. Soc.* **114**, 4846–4855 (1992).

170. Green, N. M. & Michael Green, N. [5] Avidin and streptavidin. in *Methods in Enzymology* (eds. Meir, W. & Edward, A. B.) **Volume 184**, 51–67 (Academic Press, 1990).

171. Budin, G. *et al.* Nondenaturing chemical proteomics for protein complex isolation and identification. *Chembiochem* **11**, 2359–61 (2010).

172. Patricelli, M. P., Giang, D. K., Stamp, L. M. & Burbaum, J. J. Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. *Proteomics* **1**, 1067–1071 (2001).

173. Greenbaum, D. *et al.* Chemical Approaches for Functionally Probing the Proteome. *Mol. Cell. Proteomics* **1**, 60–68 (2002).

174. Sadaghiani, A. M., Verhelst, S. H. & Bogyo, M. Tagging and detection strategies for activity-based proteomics. *Curr. Opin. Chem. Biol.* **11**, 20–8 (2007).

175. Chan, E. W. S., Chattopadhaya, S., Panicker, R. C., Huang, X. & Yao, S. Q. Developing Photoactive Affinity Probes for Proteomic Profiling: Hydroxamate-based Probes for Metalloproteases. *J. Am. Chem. Soc.* **126**, 14435–14446 (2004).

176. Singh, A., Thornton, E. R. & Westheimer, F. H. Preliminary Communications: The Photolysis of Diazoacetylchymotrypsin. *J Biol Chem* **237**, PC3006–PC3008 (1962).

177. Grunbeck, A. & Sakmar, T. P. Probing G Protein-Coupled Receptor—Ligand Interactions with Targeted Photoactivatable Cross-Linkers. *Biochemistry* **52**, 8625–8632 (2013).

178. Nakayama, T. A. & Khorana, H. G. Orientation of retinal in bovine rhodopsin determined by cross-linking using a photoactivatable analog of 11-cis-retinal. *J. Biol. Chem.* **265**, 15762–9 (1990).

179. Nakayama, T. A. & Khorana, H. G. Synthesis of a new photoactivatable analog of 11-cis-retinal. *J. Org. Chem.* **55**, 4953–4956 (1990).

180. Nishikori, K., Noshiro, O., Sano, K. & Maeno, H. Characterization, solubilization, and separation of two distinct dopamine receptors in canine caudate nucleus. *J. Biol. Chem.* **255**, 10909–15 (1980).

181. Frei, A. P. *et al.* Direct identification of ligand-receptor interactions on living cells and tissues. *Nat. Biotechnol.* **30**, 997–1001 (2012).

182. Ponda, M. P. & Breslow, J. L. Serum stimulation of CCR7 chemotaxis due to coagulation factor XIIa-dependent production of high-molecular-weight kininogen domain 5. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E7059–E7068 (2016).

183. Li, Z. *et al.* Staphylococcal Superantigens Use LAMA2 as a Coreceptor To Activate T Cells. *J. Immunol.* **200**, 1471–1479 (2018).

184. Li, Y., Ozment, T., Wright, G. L. & Peterson, J. M. Identification of Putative Receptors for the Novel Adipokine CTRP3 Using Ligand-Receptor Capture Technology. *PLoS One* **11**, e0164593 (2016).

185. Gregory, K. J. *et al.* Clickable Photoaffinity Ligands for Metabotropic Glutamate Receptor 5 Based on Select Acetylenic Negative Allosteric Modulators. *ACS Chem. Biol.* **11**, 1870–1879 (2016).

186. Soethoudt, M. *et al.* Selective Photoaffinity Probe That Enables Assessment of Cannabinoid CB2 Receptor Expression and Ligand Engagement in Human Cells. *J. Am. Chem. Soc.* **140**, 6067–6075 (2018).

187. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–7 (2006).

188. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).

189. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–7 (2001).

190. Oh-Ishi, M., Satoh, M. & Maeda, T. Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins. *Electrophoresis* **21**, 1653–1669 (2000).

191. Cantin, G. T., Yates, J. R. & Yates Iii, J. R. Strategies for shotgun identification of post-translational modifications by mass spectrometry. *J. Chromatogr. A* **1053**, 7–14 (2004).

192. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9390–5 (2000).

193. Eng, J. K., McCormack, A. L. & Yates Iii, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

194. McDonald, W. H. & Yates, J. R. Shotgun proteomics: integrating technologies to answer biological questions. *Curr. Opin. Mol. Ther.* **5**, 302–9 (2003).

195. Savas, J. N., Stein, B. D., Wu, C. C. & Yates, J. R. Mass spectrometry accelerates membrane protein analysis. *Trends in Biochemical Sciences* **36**, 388–96 (2011).

196. Vuckovic, D., Dagley, L. F., Purcell, A. W. & Emili, A. Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges. *Proteomics* **13**, 404–423 (2013).

197. Blex, C. *et al.* Targeting G protein-coupled receptors by Capture Compound Mass Spectrometry: A case study with sertindole. *ChemBioChem* (2017). doi:10.1002/cbic.201700152

198. Lai, X. A reproducible method to enrich membrane proteins with high-purity and high-yield for an LC-MS/MS approach in quantitative membrane proteomics. *Electrophoresis* **34**, 809 (2013).

199. Schey, K. L., Grey, A. C. & Nicklay, J. J. Mass spectrometry of membrane proteins: A focus on aquaporins. *Biochemistry* **52**, 3807–3817 (2013).

200. Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314 (2015).

201. Brocklehurst, K. J. *et al.* Discovery, optimisation and in vivo evaluation of novel GPR119 agonists. *Bioorg. Med. Chem. Lett.* **21**, 7310–7316 (2011).

202. Scott, J. S. *et al.* Use of small-molecule crystal structures to address solubility in a novel series of G protein coupled receptor 119 agonists: optimization of a lead and in vivo evaluation. *J. Med. Chem.* **55**, 5361–79 (2012).

203. Scott, J. S. *et al.* Optimisation of aqueous solubility in a series of G protein coupled receptor 119 (GPR119) agonists. *Medchemcomm* **4**, 95 (2013).

204. Scott, J. S. *et al.* Circumventing Seizure Activity in a Series of G Protein Coupled Receptor 119 (GPR119) Agonists. *J. Med. Chem.* **119**, 8984–8998 (2014).

205. Scott, J. S. *et al.* Conformational restriction in a series of GPR119 agonists: differences in pharmacology between mouse and human. *Bioorg. Med. Chem. Lett.* **23**, 3175–9 (2013).

206. Waring, M. J. & Johnstone, C. A quantitative assessment of hERG liability as a function of lipophilicity. *Bioorg. Med. Chem. Lett.* **17**, 1759–1764 (2007).

207. Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **6**, 881 (2007).

208. Cook, E. H. *et al.* Primary Structure of the Human Platelet Serotonin 5-HT2A Receptor: Identity with Frontal Cortex Serotonin 5-HT2A Receptor. *J. Neurochem.* **63**, 465–469 (1994).

209. Bouvier, C. *et al.* Functional characterization of the human dopamine D4.2 receptor using vaccinia virus as an expression system. *Eur. J. Pharmacol. Mol. Pharmacol.* **290**,

11–17 (1995).

210. Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. Molecular Cloning and Functional Expression of a New Human CC-Chemokine Receptor Gene. *Biochemistry* **35**, 3362–3367 (1996).

211. Wang, L. *et al.* Alternative One-Pot Synthesis of (Trifluoromethyl)phenyldiazirines from Tosyloxime Derivatives: Application for New Synthesis of Optically Pure Diazirinylphenylalanines for Photoaffinity Labeling. *Org. Lett.* **17**, 616–619 (2015).

212. Hashimoto, M. & Hatanaka, Y. Recent Progress in Diazirine-Based Photoaffinity Labeling. (2008). doi:10.1002/ejoc.200701069

213. Li, Z. *et al.* Design and Synthesis of Minimalist Terminal Alkyne-Containing Diazirine Photo-Crosslinkers and Their Incorporation into Kinase Inhibitors for Cell- and Tissue-Based Proteome Profiling. *Angew. Chemie Int. Ed.* **52**, 8551–8556 (2013).

214. Strømgaard, K. *et al.* Ginkgolide derivatives for photolabeling studies: Preparation and pharmacological evaluation. *J. Med. Chem.* **45**, 4038–4046 (2002).

215. Grell, W. *et al.* Repaglinide and Related Hypoglycemic Benzoic Acid Derivatives. *J. Med. Chem.* **41**, 5219–5246 (1998).

216. Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527-541.e29 (2017).

217. Kleiner, P., Heydenreuter, W., Stahl, M., Korotkov, V. S. & Sieber, S. A. A Whole Proteome Inventory of Background Photocrosslinker Binding. *Angew. Chemie Int. Ed.* **56**, 1396–1401 (2017).

218. Dess, D. B. & Martin, J. C. A useful 12- I -5 triacetoxyperiodinane (the Dess-Martin Periodinane) for the selective oxidation of primary or secondary alcohols and a variety of related 12- I -5 species. *J. Am. Chem. Soc.* **113**, 7277–7280 (1991).

219. Hadavizadeh, K. A chemical biology approach to target identification for drugs with an unknown mode-of-action. *Institute of Chemical Biology* (Imperial College London, 2014).

220. Lefkowitz, R. J., Roth, J. & Pastan, I. Radioreceptor Assay of Adrenocorticotropic Hormone: New Approach to Assay of Polypeptide Hormones in Plasma. *Science (80-.).* **170**, 633 (1970).

221. Zhang, R. & Xie, X. Tools for GPCR drug discovery. *Acta Pharmacol. Sin.* **33**, 372 (2012).

222. Glickman, J. F., Schmid, A. & Ferrand, S. Scintillation Proximity Assays in High-Throughput Screening. *Assay Drug Dev. Technol.* **6**, 433–455 (2008).

223. Glickman, J. F. & Schmid, A. Farnesyl Pyrophosphate Synthase: Real-Time Kinetics and Inhibition by Nitrogen-Containing Bisphosphonates in a Scintillation Assay. *Assay Drug Dev. Technol.* **5**, 205–214 (2007).

224. Carrick, T. A., Bingham, B., Eppler, C. M., Baumbach, W. R. & Zysk, J. R. A rapid and sensitive binding assay for growth hormone releasing factor. *Endocrinology* **136**, 4701–4704 (1995).

225. Brandt, D. R., Asano, T., Pedersen, S. E. & Ross, E. M. Reconstitution of catecholamine-stimulated guanosine triphosphatase activity. *Biochemistry* **22**, 4357–4362 (1983).

226. Asano, T., Katada, T., Gilman, A. G. & Ross, E. M. Activation of the inhibitory GTP-binding protein of adenylate cyclase, Gi, by beta-adrenergic receptors in reconstituted phospholipid vesicles. *J. Biol. Chem.* **259**, 9351–9354 (1984).

227. Beavo, J. A. *et al.* Effects of Xanthine Derivatives on Lipolysis and on Adenosine 3',5'-Monophosphate Phosphodiesterase Activity. *Mol. Pharmacol.* **6**, 597 (1970).

228. Essayan, D. M. Cyclic nucleotide phosphodiesterases. *J. Allergy Clin. Immunol.* **108**, 671–680 (2001).

229. Gabriel, D. *et al.* High Throughput Screening Technologies for Direct Cyclic AMP Measurement. *Assay Drug Dev. Technol.* **1**, 291–303 (2003).

230. Williams, C. cAMP detection methods in HTS: selecting the best from the rest. *Nat. Rev. Drug Discov.* **3**, 125 (2004).

231. Allen, M., Hall, D., Collins, B. & Moore, K. A Homogeneous High Throughput Nonradioactive Method for Measurement of Functional Activity of Gs-Coupled Receptors in Membranes. *J. Biomol. Screen.* **7**, 35–44 (2002).

232. Prystay, L., Gagne, A., Kasila, P., Yeh, L.-A. & Banks, P. Homogeneous Cell-Based Fluorescence Polarization Assay for the Direct Detection of cAMP. *J. Biomol. Screen.* **6**, 75–82 (2001).

233. Degorce, F. *et al.* HTRF: A Technology Tailored for Drug Discovery –A Review of Theoretical Aspects and Recent Applications. *Curr. Chem. Genomics* **3**, 22–32 (2009).

234. Hill, S. J., Williams, C. & May, L. T. Insights into GPCR pharmacology from the measurement of changes in intracellular cyclic AMP; advantages and pitfalls of differing methodologies. *Br. J. Pharmacol.* **161**, 1266–1275 (2010).

235. Hunter, M. R. & Glass, M. Increasing the flexibility of the LANCE cAMP detection kit. *J. Pharmacol. Toxicol. Methods* **71**, 42–45 (2015).

236. Ferguson, S. S., Zhang, J., Barak, L. S. & Caron, M. G. G-protein-coupled receptor kinases and arrestins: regulators of G-protein-coupled receptor sequestration. *Biochem. Soc. Trans.* **24**, 953–9 (1996).

237. Goodman, O. B. *et al.* β-Arrestin acts as a clathrin adaptor in endocytosis of the β2-adrenergic receptor. *Nature* **383**, 447–450 (1996).

238. Magalhaes, A. C., Dunn, H. & Ferguson, S. S. Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins. *Br. J. Pharmacol.* **165**, 1717–1736 (2012).

239. Bowman, S. L., Shiwarski, D. J. & Puthenveedu, M. A. Distinct G protein-coupled receptor recycling pathways allow spatial control of downstream G protein signaling. *J. Cell Biol.* **214**, 797–806 (2016).

240. Michel, M. C., Wieland, T. & Tsujimoto, G. How reliable are G-protein-coupled receptor antibodies? *Naunyn. Schmiedebergs. Arch. Pharmacol.* **379**, 385–388 (2009).

241. de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5**, 1512–1526 (2009).

242. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Letters* **583**, 3966–3973 (2009).

243. Congreve, M. & Marshall, F. The impact of GPCR structures on pharmacology and structure-based drug design. *Br. J. Pharmacol.* **159**, 986–996 (2010).

244. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci.* **106**, 1760–1765 (2009).

245. Marchalant, Y., Brownjohn, P. W., Bonnet, A., Kleffmann, T. & Ashton, J. C. Validating Antibodies to the Cannabinoid CB2 Receptor: Antibody Sensitivity Is Not Evidence of Antibody Specificity. *J. Histochem. Cytochem.* **62**, 395–404 (2014).

246. Cécyre, B., Thomas, S., Ptito, M., Casanova, C. & Bouchard, J.-F. Evaluation of the specificity of antibodies raised against cannabinoid receptor type 2 in the mouse retina. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **387**, 175–184 (2014).

247. Hawes, JJ; Picciotto, M. Corrigendum: Hawes JJ, Picciotto MR 2004. Characterization of GalR1, GalR2, and GalR3 immunoreactivity in catecholaminergic nuclei of the mouse brain. J Comp Neurol 479:410-423. *J. Comp. Neurol.* **490**, 98–100 (2005).

248. Shonberg, J., Kling, R. C., Gmeiner, P. & Löber, S. GPCR crystal structures: Medicinal chemistry in the pocket. *Bioorganic Med. Chem.* **23**, 3880–3906 (2015).

249. Pándy-Szekeres, G. *et al.* GPCRdb in 2018: Adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).

250. Saarenpää, T., Jaakola, V.-P. P. & Goldman, A. Baculovirus-mediated expression of GPCRs in insect cells. *Methods Enzymol.* **556**, 185–218 (2015).

251. Zhu, J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* **30**, 1158–1170 (2012).

252. Lundstrom, K. *et al.* Structural genomics on membrane proteins: Comparison of more than 100 GPCRs in 3 expression systems. *J. Struct. Funct. Genomics* **7**, 77–91 (2006).

253. López de Maturana, R., Willshaw, A., Kuntzsch, A., Rudolph, R. & Donnelly, D. The isolated N-terminal domain of the glucagon-like peptide-1 (GLP-1) receptor binds exendin peptides with much higher affinity than GLP-1. *J. Biol. Chem.* **278**, 10195–200 (2003).

254. Baneres, J.-L. *et al.* Structure-based Analysis of GPCR Function: Conformational Adaptation of both Agonist and Receptor upon Leukotriene B4 Binding to Recombinant BLT1. *J. Mol. Biol.* **329**, 801–814 (2003).

255. Scott, D. J. & Plückthun, A. Direct Molecular Evolution of Detergent-Stable G Protein-Coupled Receptors Using Polymer Encapsulated Cells. *J. Mol. Biol.* **425**, 662–677 (2013).

256. Egloff, P. *et al.* Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E655-62 (2014).

257. Bill, R. M. Playing catch-up with escherichia coli: Using yeast to increase success rates in recombinant protein production experiments. *Front. Microbiol.* **5**, 85 (2014).

258. Bertheleme, N., Singh, S., Dowell, S. & Byrne, B. Heterologous expression of G-protein-coupled receptors in yeast. in *Methods in Enzymology* **556**, 141–164 (Academic Press, 2015).

259. Opekarová, M. & Tanner, W. Specific lipid requirements of membrane proteins - A putative bottleneck in heterologous expression. *Biochim. Biophys. Acta - Biomembr.* **1610**, 11–22 (2003).

260. Maeda, S. & Schertler, G. F. Production of GPCR and GPCR complexes for structure determination. *Curr. Opin. Struct. Biol.* **23**, 381–392 (2013).

261. Jarvis, D. L. & Finn, E. E. Biochemical Analysis of the N-Glycosylation Pathway in Baculovirus-Infected Lepidopteran Insect Cells. *Virology* **212**, 500–511 (1995).

262. Jarvis, D. L., Howe, D. & Aumiller, J. J. Novel baculovirus expression vectors that provide sialylation of recombinant glycoproteins in lepidopteran insect cells. *J. Virol.* **75**, 6223–7 (2001).

263. Grisshammer, R. Chapter 36 Purification of Recombinant G-Protein-Coupled Receptors. in *Methods in Enzymology* **463**, 631–645 (NIH Public Access, 2009).

264. Kenakin, T. Inverse, protean, and ligand-selective agonism: matters of receptor conformation. *FASEB J.* **15**, 598–611 (2001).

265. Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 877–882 (2008).

266. Warne, T. *et al.* Structure of a β1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).

267. Robertson, N. *et al.* The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* **60**, 36–44 (2011).

268. Tehan, B. G. & Christopher, J. A. The use of conformationally thermostabilised GPCRs in drug discovery: application to fragment, structure and biophysical techniques. *Curr. Opin. Pharmacol.* **30**, 8–13 (2016).

269. Cooke, R. M., Brown, A. J. H., Marshall, F. H. & Mason, J. S. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. *Drug Discovery Today* **20**, 1355–1364 (2015).

270. Wiener, M. C. A pedestrian guide to membrane protein crystallization. *Methods* **34**, 364–372 (2004).

271. Linke, D. Chapter 34 Detergents: An Overview. *Methods Enzymol.* **463**, 603–617 (2009).

272. Privé, G. G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**, 388–397 (2007).

273. Kalipatnapu, S. & Chattopadhyay, A. Membrane Protein Solubilization: Recent Advances and Challenges in Solubilization of Serotonin1A Receptors. *International Union of Biochemistry and Molecular Biology: Life* **57**, 505–512 (2005).

274. Morandat, S. & El Kirat, K. Solubilization of supported lipid membranes by octyl glucoside observed by time-lapse atomic force microscopy. *Colloids Surfaces B Biointerfaces* **55**, 179–184 (2007).

275. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Lett.* **2**, 853–856 (2002).

276. Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **126**, 3477–3487 (2004).

277. Denisov, I. G. & Sligar, S. G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 481–486 (2016).

278. Wilcox, K. C. *et al.* Nanoscale Synaptic Membrane Mimetic Allows Unbiased High Throughput Screen That Targets Binding Sites for Alzheimer's-Associated Aβ Oligomers. *PLoS One* **10**, e0125263 (2015).

279. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* **469**, 175–181 (2011).

280. Efremov, R. G., Leitner, A., Aebersold, R. & Raunser, S. Architecture and conformational switch mechanism of the ryanodine receptor. *Nature* **517**, 39–43 (2015).

281. Kimple, M. E. & Sondek, J. Overview of Affinity Tags for Protein Purification. in *Current Protocols in Protein Science* **36**, 9.9.1-9.9.19 (John Wiley & Sons, Inc., 2004).

282. Kapust, R. B. & Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).

283. Maina, C. V. *et al.* An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74**, 365–373 (1988).

284. Davis, G. D., Elisee, C., Newham, D. M. & Harrison, R. G. New fusion protein systems designed to give soluble expression in Escherichia coli. *Biotechnol. Bioeng.* **65**, 382–8 (1999).

285. Hochuli, E., Döbeli, H. & Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr. A* **411**, 177–184 (1987).

286. Hopp, T. P. *et al.* A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **6**, 1204–1210 (1988).

287. Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49**, 455–465 (2001).

288. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).

289. Hutter, H. Fluorescent Protein Methods: Strategies and Applications. *Methods Cell Biol.* **107**, 67–92 (2012).

290. Ohba, Y., Fujioka, Y., Nakada, S. & Tsuda, M. Fluorescent Protein-Based Biosensors and Their Clinical Applications. *Prog. Mol. Biol. Transl. Sci.* **113**, 313–348 (2013).

291. Mollwitz, B. *et al.* Directed evolution of the suicide protein O 6-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **51**, 986–994 (2012).

292. Gautier, A. *et al.* An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15**, 128–136 (2008).

293. Kalef, E. & Gitler, C. Purification of vicinal dithiol-containing proteins by arsenical-based affinity chromatography. *Methods Enzymol.* **233**, 395–403 (1994).

294. Adams, S. R. & Tsien, R. Y. Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins. *Nat. Protoc.* **3**, 1527–1534 (2008).

295. Cornall, L. M., Hryciw, D. H., Mathai, M. L. & McAinch, A. J. Direct activation of the proposed anti-diabetic receptor, GPR119 in cardiomyoblasts decreases markers of muscle metabolic activity. *Mol. Cell. Endocrinol.* **402**, 72–85 (2015).

296. Cornall, L. M. *et al.* GPR119 regulates genetic markers of fatty acid oxidation in cultured skeletal muscle myotubes. *Mol. Cell. Endocrinol.* **365**, 108–118 (2013).

297. Hu, Y.-W. Y.-R. *et al.* A lincRNA-DYNLRB2-2/GPR119/GLP-1R/ABCA1-dependent signal transduction pathway is essential for the regulation of cholesterol homeostasis. *J. Lipid Res.* **55**, 681–697 (2014).

298. Chepurny, O. G. *et al.* Stimulation of proglucagon gene expression by human GPR119 in enteroendocrine L-cell line GLUTag. *Mol. Endocrinol.* **27**, 1267–82 (2013).

299. Hanyaloglu, A. C., McCullagh, E. & von Zastrow, M. Essential role of Hrs in a recycling mechanism mediating functional resensitization of cell signaling. *EMBO J.* **24**, 2265–2283 (2005).

300. Tang, Y. *et al.* Identification of the endophilins (SH3p4/p8/p13) as novel binding partners for the beta 1-adrenergic receptor. *Proc. Natl. Acad. Sci.* **96**, 12559–12564 (1999).

301. Ronik Lachance, V. *et al.* Ubiquitylation and activation of a Rab GTPase is promoted by a b 2 AR-HACE1 complex. *J. Cell Sci.* doi:10.1242/jcs.132944

302. Los, G. V. *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).

303. Janssen, D. B. Evolving haloalkane dehalogenases. *Curr. Opin. Chem. Biol.* **8**, 150–159 (2004).

304. Stoddart, L. A., Smith, N. J. & Milligan, G. International union of pharmacology. LXXI. Free fatty acid receptors FFA1, -2, and -3: Pharmacology and pathophysiological functions. *Pharmacological Reviews* **60**, 405–417 (2008).

305. Köse, M. *et al.* Development of [3H]2-carboxy-4,6-dichloro-1 H -indole-3-propionic acid ([3H]PSB-12150): A useful tool for studying GPR17. *ACS Med. Chem. Lett.* **5**, 326–330 (2014).

306. Al Mahmud, Z. *et al.* Three classes of ligands each bind to distinct sites on the orphan G protein-coupled receptor GPR84. *Sci. Rep.* **7**, 17953 (2017).

307. Ward, R. J., Alvarez-Curto, E. & Milligan, G. Using the Flp-In™ T-Rex™ System to Regulate GPCR Expression. in *Methods in Molecular Biology* **746**, 21–37 (Humana Press, Totowa, NJ, 2011).

308. Hayakawa, K., Yodo, M., Ohsuki, S. & Kanematsu, K. Novel bicycloannulation via tandem vinylation and intramolecular Diels-Alder reaction of five-membered heterocycles: a new approach to construction of psoralen and azapsoralen. *J. Am. Chem. Soc.* **106**, 6735–6740 (1984).

309. Feng, Y., Liu, J., Carrasco, Y. P., MacMillan, J. B. & De Brabander, J. K. Rifamycin Biosynthetic Congeners: Isolation and Total Synthesis of Rifsaliniketal and Total Synthesis of Salinisporamycin and Saliniketals A and B. *J. Am. Chem. Soc.* **138**, 7130–7142 (2016).

310. Kunishima, M. *et al.* 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride: an efficient condensing agent leading to the formation of amides and esters. *Tetrahedron* **55**, 13159–13170 (1999).

311. OligoAnalyzer. Available at: https://www.idtdna.com/calc/analyser.

312. NEB Tm Calculator. Available at: https://tmcalculator.neb.com/#!/main.

313. NEB Ligation Calculator. Available at: https://nebiocalculator.neb.com/#!/ligation.

314. Heal, W. P. *et al.* Bioorthogonal chemical tagging of protein cholesterylation in living cells. *Chem. Commun.* **47**, 4081–4083 (2011).

315. Broncel, M. *et al.* Multifunctional Reagents for Quantitative Proteome-Wide Analysis of Protein Modification in Human Cells and Dynamic Profiling of Protein Lipidation During Vertebrate Development. *Angew. Chemie Int. Ed.* **54**, 5948–5951 (2015).

316. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

317. Basarab, G., Ni, H., Sherer, B. & Zhou, F. Antibacterial pyrrolopyridines, pyrrolopyrimidines, and pyrroloazepines. **WO 2007/07**, (2007).

318. Aridoss, G. & Laali, K. K. Building Heterocyclic Systems with RC(OR)2+ Carbocations in Recyclable Brønsted Acidic Ionic Liquids: Facile Synthesis of 1-Substituted 1H-1,2,3,4-Tetrazoles, Benzazoles and Other Ring Systems with CH(OEt)3 and EtC(OEt)3 in [EtNH3][NO3] and [PMIM(SO3H)][O. *European J. Org. Chem.* **2011**, 2827–2835 (2011).

319. Nh, L., Satoh, Y. & Marcopulos, N. Application of 5-lithiotetrazoles in organic synthesis. *Tetrahedron Lett.* **36**, 1759–1762 (1995).

320. Sasmal, S. *et al.* Design and optimization of quinazoline derivatives as melanin concentrating hormone receptor 1 (MCHR1) antagonists: Part 2. *Bioorg. Med. Chem. Lett.* **22**, 3163–3167 (2012).

321. Kunishima, M., Kawachi, C., Iwasaki, F., Terao, K. & Tani, S. Synthesis and characterization of 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride. *Tetrahedron Lett.* **40**, 5327–5330 (1999).

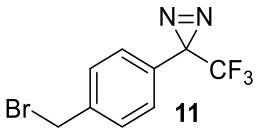
322. Yip, G. M. *et al.* A propofol binding site on mammalian GABA receptors identified by photolabeling. *Nat Chem Biol* (2013). doi:10.1038/nchembio.1340

323. Reiner, R. & Rossmann, K. Nukleophile Substanzen zur Entgiftung von Phosphorestern. *Chem. Mon.* **113**, 223–231 (1982).

8. Appendix

8.1. Appendix 1: Additional Chemical Synthesis

8.1.1. 3-[4-(bromomethyl)phenyl]-3-(trifluoromethyl)-3H-diazirine[214] (11)

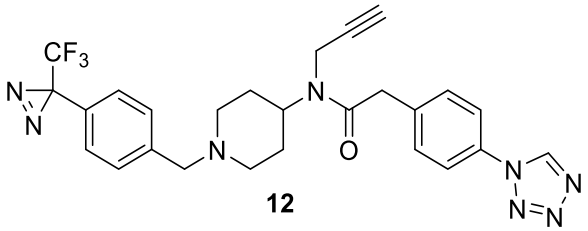


11

Diazirine **43** (100 mg, 0.500 mmol) was dissolved in acetonitrile (1 mL). *N*-bromosuccimide (98 mg, 0.550 mmol) and azobisisobutyronitrile (AIBN) (8 mg, 0.050 mmol) were added, and the solution stirred at 90 °C for 2h. The resulting solution was diluted with CH_2Cl_2 (10 mL), washed with water (2 x 10 mL) and brine (1 x 10 mL), dried and concentrated *in vacuo*. Purification via column chromatography (hexane:DCM 90:10) afforded product **11** as a clear oil (90 mg, 65%).

^{1}H NMR (400 MHz, $CDCl_3$) δ = 4.47 (s, 2H), 7.17 (d, *J*=8.2, 2H), 7.38 – 7.47 (m, 2H); ^{19}F NMR (377 MHz, $CDCl_3$) δ = -65.17 (s); ^{13}C NMR could not be obtained due to decomposition in the NMR tube; *m/z* data not available due to decomposition.

8.1.2. N-(1-{[4-(3-methyl-3H-diazirin-3-yl)phenyl]methyl}piperidin-4-yl)-N-(prop-2-yn-1-yl)-2-[4-(1H-1,2,3,4-tetrazol-1-yl)phenyl]acetamide (14)

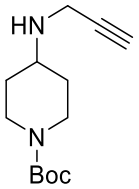


12

Carbamate **44** (150 mg, 0.36 mmol) was dissolved in CH_2Cl_2 (7.5 mL), TFA (1.5 mL) was added dropwise and stirred for 2h. The resulting reaction mixture was concentrated *in vacuo* and the residue was redissolved in MeCN (3 mL). Bromo **11** (70 mg, 0.25 mmol) and DIPEA (0.28 mL, 1.63 mmol) were added and the resulting solution stirred for 16h. The reaction mixture was diluted with water (5 mL), extracted into CH_2Cl_2 (5 mL) and washed with water (3 x 5 mL). The organic layer was dried and concentrated *in vacuo*. Subsequent purification via column chromatography (methanol:DCM 1:99 – 5:95) afforded **14** as a pale yellow oil (47 mg, 35%).

ν_{max} (neat) 3294, 2921, 2852, 1639, 1520, 1458, 1150, 994, 734, 665; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 1.67 – 1.77 (m, 2H), 1.92 – 1.99 (m, 2H), 2.09 – 2.24 (m, 2H), 2.36 (d, *J*=2.5, 1H), 2.93 (d, *J*=8.7, 2H), 3.51 (d, *J*=3.9, 2H), 3.91 (d, *J*=30.6, 2H), 4.07 (dd, *J*=41.3, 2.7, 2H), 4.53 (tt, *J*=11.6, 4.5, 1H), 7.16 (d, *J*=7.8, 2H), 7.36 (t, *J*=8.0, 2H), 7.51 (dd, *J*=8.7, 2.7, 2H), 7.62 – 7.80 (m, 2H), 9.00 (d, *J*=4.2, 1H); ^{19}F NMR (377 MHz, $CDCl_3$) δ = -65.25 (s); ^{13}C NMR (101 MHz, $CDCl_3$) δ = 29.5, 32.8, 40.4, 52.1, 53.0, 56.5, 62.2, 72.6, 79.9, 121.3, 122.2 (q, *J*=276.0), 126.4, 129.4, 131.1, 132.7, 137.4, 140.5, 170.1; *m/z* (ESI$^+$) 467 ([M-2N_2+H$^+$]$^+$, 20%), 495 ([M-N_2+H$^+$]$^+$, 30%), 523 ([M+H]$^+$, 100%); HRMS (ESI$^+$) $C_{26}H_{26}N_8OF_3^+$ ([M+H$^+$]$^+$) requires 523.2172, found 523.2182.

8.1.3. *tert*-butyl 4-[(prop-2-yn-1-yl)amino]piperidine-1-carboxylate[317] (18)

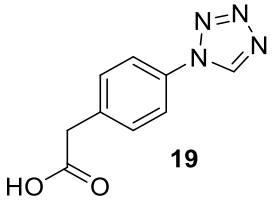


18

A solution of 1-boc-4-piperidone (1.50 g, 7.50 mmol), propargylamine (0.530 mL, 8.25 mmol) and sodium triacetoxyborohydride (2.396 g, 11.30 mmol) in CH_2Cl_2 was stirred at room temperature overnight. The resulting reaction mixture was quenched with 1M HCl (10 mL) and basified with saturated aqueous sodium carbonate (15 mL) before being extracted into CH_2Cl_2 (2 x 20 mL). The combined organic extracts were washed with brine (20 mL), dried ($MgSO_4$) and concentrated *in vacuo* to afford **18** as a yellow oil (1.80 g, quant.).[317]

^{1}H NMR (400 MHz, DMSO) δ 0.99 – 1.19 (m, 2H), 1.39 (s, 9H), 1.68 – 1.77 (m, 2H), 2.67 – 2.98 (m, 3H), 3.02 (t, *J* = 2.4 Hz, 1H), 3.33 (d, *J* = 2.5 Hz, 2H), 3.70 – 3.88 (m, 2H); *m/z* 183 ([M-^{t}Bu+H$^+$]$^+$, 100%), 239 ([M+H$^+$]$^+$, 20%).

8.1.4. 2-[4-(1H-1,2,3,4-tetrazol-1-yl)phenyl]acetic acid[318] (19)

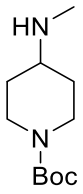


19

A solution of 4-aminoacetic acid (400 mg, 2.60 mmol), sodium azide (190 mg, 2.90 mmol) and triethyl-orthoformate (1.30 mL, 3.00 mmol) in acetic acid (1.20 mL, 8.00 mmol) was stirred for 6 hours at 80 °C. The reaction mixture was allowed to cool and diluted with water (10 mL). The product was extracted into EtOAC (2 x 10 mL). The aqueous layer was acidified using 1M HCl, and extracted into EtOAc (2 x 10 mL). Saturated aqueous sodium bicarbonate was used to neutralise the aqueous extracts before disposal. The combined organic layers were washed with water (2 x 10 mL) and brine (10 mL), dried ($MgSO_4$) and concentrated *in vacuo* to afford **19** as a light yellow solid (455 mg, 86%).[319]

^{1}H NMR (400 MHz, DMSO) δ 3.72 (s, 2H), 7.54 (d, *J* = 8.2 Hz, 2H), 7.85 (d, *J* = 8.2 Hz, 2H), 10.08 (s, 1H); *m/z* (ESI$^+$) 131, 177 [(M-N_2+H$^+$]$^+$, 95%), 205 ([M+H$^+$]$^+$, 100%).

8.1.5. *tert*-Butyl 4-(methylamino)piperidine-1-carboxylate[320] (20)

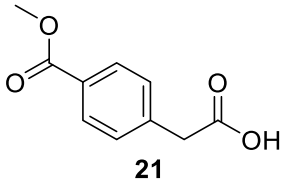


20

A solution of 1-boc-4-piperidone (500 mg, 2.50 mmol), methylamine (40% in H_2O) (0.238 mL, 2.75 mmol) and sodium triacetoxyborohydride (795 mg, 3.75 mmol) in THF was stirred for 16h. The resulting reaction mixture was quenched with 1M HCl (10 mL) and basified with 1M NaOH (pH = 11) before being extracted into CH_2Cl_2 (3 x 20 mL). The combined organic extracts were washed with water, brine, dried ($MgSO_4$) and concentrated *in vacuo* to afford **20** as an off white solid (497 mg, 93 %).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.12 – 1.34 (m, 2H), 1.45 (s, 9H), 1.82 – 1.90 (m, 2H), 1.95 (s, 1H), 2.44 (s, 3H), 2.52 (tt, *J* = 10.4, 3.9 Hz, 1H), 2.79 (t, *J* = 12.6 Hz, 2H), 4.04 (d, *J* = 13.0 Hz, 2H); *m/z* (ESI$^+$) 159 ([M-^{t}Bu+H$^+$]$^+$, 100%), 215 ([M+H$^+$]$^+$, 60%).

8.1.6. 2-[4-(methoxycarbonyl)phenyl]acetic acid[215] (21)

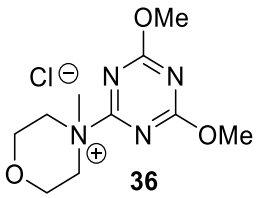


21

A solution of the diester methyl 4-(2-methoxy-2-oxoethyl)benzoate (1.00 g, 4.81 mmol) and NaOH (0.192 g, 4.81 mmol) in MeOH (20 mL) was stirred at 50 °C for 4 h. The reaction mixture was concentrated *in vacuo*, the residue taken up in Et_2O (10 mL) and washed with water (2 x 10 mL), the resulting organic layer containing unreacted start material was kept separate. The aqueous layer was acidified with c. HCl and the product **21** extracted into Et_2O (3 x 10 mL) and these combined organic layers washed with brine (2 x 10 mL), dried ($MgSO_4$) and concentrated *in vacuo* to afford **21** as a white solid (0.823 g, 88%).[215]

^{1}H NMR (400 MHz, $CDCl_3$) δ 3.74 (s, 2H), 3.94 (s, 3H), 7.39 (d, *J* = 8.0 Hz, 2H), 8.03 (d, *J* = 8.3 Hz, 2H); *m/z* (ESI$^+$) 195 ([M+H$^+$]$^+$, 100%), 217 ([M+Na$^+$]$^+$, 30%).

8.1.7. 4-(dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholin-4-ium chloride[321] (36)



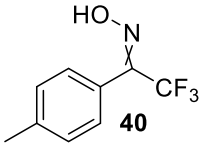
36

4-Methylmorpholine (0.285 mL, 2.590 mmol) was added to a solution of 2-chloro-4,6-dimethoxy-1,3,5-triazine (500 mg, 2.850 mmol) in THF (6 mL) and stirred for 30 minutes. The precipitate was collected via suction filtration and washed with THF to give the dry product **36** (770 mg, 100%). The product was used without further purification.[310]

^{1}H NMR (400 MHz, DMSO) δ 3.49 (s, 3H), 3.76 – 3.83 (m, 2H), 3.93 (ddd, *J* = 12.7, 10.2, 2.8 Hz, 2H), 4.01 (dt, *J* = 14.4, 3.2 Hz, 2H), 4.11 (s, 6H), 4.36 (d, *J* = 12.3 Hz, 2H).*

*Peaks corresponding to the decomposition of **36** via demethylation were also observed in the NMR spectrum (~25% of the sample); this is most likely due to the sample being in solvent for around 3 hours prior to running. ^{1}H NMR (400 MHz, DMSO) δ 3.49 (s, 3H, chloromethane), 3.76 – 3.83 (m, 2H), 3.93 (ddd, *J* = 12.7, 10.2, 2.8 Hz, 2H), 4.01 (dt, *J* = 14.4, 3.2 Hz, 2H), 4.11 (s, 6H), 4.36 (d, *J* = 12.3 Hz, 2H). These values match the values given by Kunishima et al.[321]

8.1.8. N-[2,2,2-trifluoro-1-(4-methylphenyl)ethylidene]hydroxylamine[214] (40)



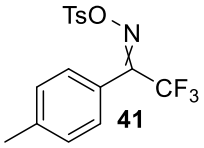
40

2,2,2-trifluoro-1-(4-methylphenyl)ethan-1-one (0.660 mL, 4.30 mmol) was dissolved in pyridine (10 mL), and hydroxylamine hydrochloride (890 mg, 12.90 mmol) added. The resulting solution was stirred at 70 °C for 3h. The crude reaction mixture was concentrated *in vacuo*, and the resulting residue taken up in Et_2O (10 mL), washed with 0.01M HCl (2 x 5 mL), water (3 x 5 mL) then dried and concentrated *in vacuo* to afford **40** as a colourless oil in a quantitative yield.[214]

^{1}H NMR (400 MHz, $CDCl_3$) δ = 2.39 (d, *J*=6.7, 3H), 7.21 (d, *J*=7.9, 1H), 7.28 (d, *J*=8.1, 1H), 7.43 (dd, *J*=12.9, 7.9, 2H);* ^{19}F NMR (377 MHz, $CDCl_3$) δ = -66.32 (s), -62.41 (s); ^{13}C NMR (101 MHz, $CDCl_3$) δ = 21.4, 120.9 (q, *J*=274.0), 123.4, 1285, 129.2, 140.6, 147.4 (q, *J*=30.2);* *m/z* (ESI^-) 202 ($[M-H^+]^-$, 100%); *m/z* (ESI^+) 204 ($[M+H^+]^+$, 100%).

*Multiple splitting patterns observed due to the presence of anti/syn configurations of the oxime, average value stated in ^{13}C NMR.

8.1.9. [2,2,2-trifluoro-1-(4-methylphenyl)ethylidene]amino 4-methylbenzene-1-sulfonate[214] (41)



41

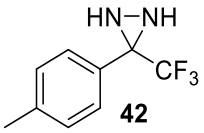
To a stirring solution of oxime **40** (870 mg, 4.29 mmol) in CH_2Cl_2 (30 mL) at 0°C, 4-dimethylaminopyridine (157 mg, 1.29 mmol) and triethylamine (0.895 mL, 6.44 mmol) were added. *p*-Toluenesulfonyl chloride (1.628 g, 6.44 mmol) was then added portion-wise, the solution warmed to RT and left to stir for a further 3h. The resulting solution was washed with water (3 x 10 mL), dried and concentrated *in vacuo*. Purification via column chromatography (hexane:DCM 70:30) afforded tosylate **41** as a white solid (1.230 g, 80%).[322]

^{1}H NMR (400 MHz, $CDCl_3$) δ = 2.40 (d, *J*=5.9, 3H), 2.47 (d, *J*=7.8, 3H), 7.20 – 7.30 (m, 2H), 7.30 – 7.41 (m, 4H), 7.85 – 7.91 (m, 2H);* ^{19}F NMR (377 MHz, $CDCl_3$) δ = -66.56 (s), -61.51 (s);* ^{13}C NMR (101 MHz, $CDCl_3$) δ = 21.5, 21.8, 119.7 (q, *J*=277.5),

121.7, 128.8, 129.2, 129.5, 129.9, 131.4, 142.4, 146.0, 154.0 (q, *J*=32.5); *m/z* (ESI^+) 358 ($[M+H^+]^+$, 70%), 380 ($[M+Na^+]^+$, 100%).

*Multiple splitting patterns observed due to the presence of anti/syn configurations of the oxime, average value stated in ^{13}C NMR.

8.1.10. 3-(4-methylphenyl)-3-(trifluoromethyl)diaziridine[214] (42)

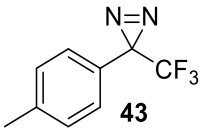


42

$NH_{3(l)}$ (~2 mL) was condensed into a round bottomed flask attached to a dry-ice cold finger condenser. Tosyl **41** (1 g, 2.80 mmol) was then dissolved in Et_2O (5 mL) and added to the reaction vessel and stirred at -78 °C for 6h. The reaction was allowed to continue stirring overnight at RT to allow the ammonia to evaporate. The residue was taken up in Et_2O (10 mL), washed with water (2 x 10 mL), dried and concentrated *in vacuo*. Purification via column chromatography (hexane:DCM 50:50) afforded diaziridine **42** as a white solid (449 mg, 79%).[214]

^1H NMR (400 MHz, $CDCl_3$) δ = 2.38 (s, 3H), 7.22 (d, *J*=5.7, 2H), 7.50 (d, *J*=7.9, 2H); ^{19}F NMR (377 MHz, $CDCl_3$) δ = -75.68 (s); ^{13}C NMR (101 MHz, $CDCl_3$) δ 21.3, 123.6 (q, *J* = 279.3 Hz), 128.0, 128.8, 129.4, 140.3. One quaternary carbon not observed; *m/z* $(ESI^+)^+$ 203 ($[M+H^+]^+$, 100%).

8.1.11. 3-(4-methylphenyl)-3-(trifluoromethyl)-3H-diazirine[214] (43)

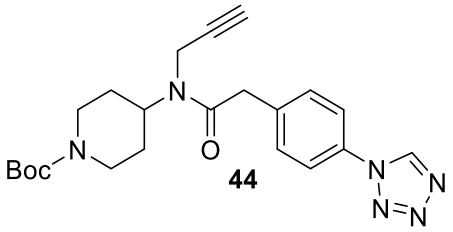


43

Diaziridine **42** (449 mg, 2.22 mmol) was dissolved in CH_2Cl_2 (9 mL) and triethylamine (0.927 mL, 6.66 mmol) was added. The solution was cooled to 0 °C in the dark, and Iodine (617 mg, 2.44 mmol) was added in small portions until a persistent brown colour remained for longer than 5 minutes. The reaction mixture was washed with 1M NaOH (2 x 5 mL), water (2 x 5 mL) and brine (5 mL), then dried and concentrated *in vacuo* at 20 °C. Purification *via* column chromatography (hexane:DCM 90:10) afforded **43** as a clear oil (93 mg, 21%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 2.4 (s, 3H), 7.1 (d, *J* = 7.9 Hz, 2H), 7.2 (d, *J* = 8.0 Hz, 2H); ^{19}F NMR (377 MHz, $CDCl_3$) δ = -65.40 (s); ^{13}C NMR (101 MHz, $CDCl_3$) δ 21.2, 28.4 (q, *J* = 45.6 Hz), 122.2 (q, *J* = 275.0 Hz), 126.1, 126.4, 129.5, 139.8; ***m/z*** data not available due to decomposition.

8.1.12. *tert*-Butyl 4-[N-(prop-2-yn-1-yl)-2-[4-(1H-1,2,3,4-tetrazol-1-yl)phenyl]acetamido]piperidine-1-carboxylate (44)

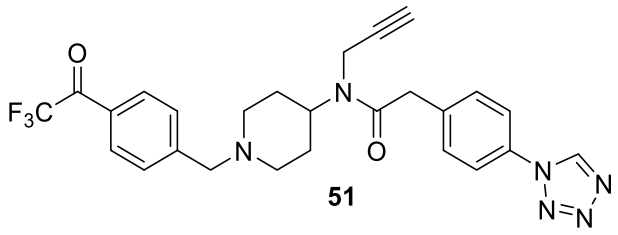


44

Acid **19** (300 mg, 1.47 mmol) and amine **18** (385 mg, 1.62 mmol) were dissolved in THF (9 mL) and stirred for 10 minutes at RT. Coupling reagent **36** (446 mg, 1.62 mmol) was then added and the reaction mixture stirred for 3h. The reaction mixture was diluted with water (10 mL) and extracted into CH_2Cl_2 (2 x 10 mL). The combined organic layers were washed sequentially with sat. sodium carbonate (10 mL), water (10 mL), 1M HCl (10 mL), water (10 mL) and brine (10 mL), dried and concentrated *in vacuo*. Purification via column chromatography (methanol:DCM 1:99) afford **44** as an off white solid (331 mg, 49%).[310]

^{1}H NMR (400 MHz, $CDCl_3$) δ = 1.52 (s, 9H), 1.59 – 1.79 (m, 4H), 2.39 (s, 1H), 2.73 (dt, *J*=47.5, 13.7, 2H), 3.95 (t, *J*=18.0, 4H), 4.24 (s, 2H), 4.64 (tt, *J*=12.3, 3.8, 1H), 7.52 (d, *J*=8.2, 2H), 7.70 (d, *J*=7.7, 2H), 9.02 (d, *J*=5.5, 1H); ^{13}C NMR (101 MHz, $CDCl_3$) δ = 28.4, 29.5, 32.9, 40.4, 43.3, 52.3, 72.8, 79.8, 121.4, 130.8, 131.1, 132.7, 137.3, 140.5, 154.6, 170.1; ***m/z*** (ESI$^+$) 297 ([M-Boc-N_2+H$^+$]$^+$, 85%), 325 ([M-Boc+H$^+$]$^+$, 100%) 369 ([M-^{t}Bu+H$^+$]$^+$, 60%), 397 ([M-N_2+H$^+$]$^+$, 30%), 425 ([M+H$^+$]$^+$,20%), 447 ([M+Na$^+$]$^+$, 85%). HRMS (ESI$^+$) $C_{22}H_{29}N_6O_3^+$ ([M+H$^+$]$^+$) requires 425.2313, found 425.2301.

8.1.13. Difluoromethyl 4-({4-[N-(prop-2-yn-1-yl)-2-[4-(1H-1,2,3,4-tetrazol-1-yl)phenyl]acetamido]piperidin-1-yl}methyl)benzoyl fluoride (51)

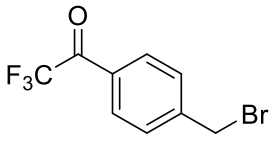


51

Carbamate **44** (150 mg, 0.36 mmol) was dissolved in CH_2Cl_2 (5 mL), TFA (1 mL) was added dropwise and stirred for 2h. The resulting reaction mixture was concentrated *in vacuo* and the residue was redissolved in MeCN (3 mL). Bromo **54** (95 mg, 0.36 mmol) and DIPEA (0.37 mL, 2.10 mmol) were added and the resulting solution stirred for 16h. The reaction mixture was concentrated *in vacuo*, extracted into CH_2Cl_2 (5 mL) and washed with water (3 x 5 mL). The organic layer was dried and concentrated *in vacuo*. Subsequent purification via column chromatography (methanol:DCM 1:99 – 5:95) afforded **51** as a pale yellow solid (114 mg, 70%).

ν_{max} (neat) 2922, 1714, 1632, 1520, 1164, 1091, 808, 666; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 1.69 – 1.90 (m, 4H), 1.98 – 2.18 (m, 2H), 2.37 (t, *J*=2.4, 1H), 2.95 (d, *J*=9.8, 2H), 3.60 (d, *J*=2.9, 2H), 3.91 (d, *J*=29.4, 2H), 4.09 (dd, *J*=41.4, 2.7, 2H), 4.54 (tt, *J*=11.7, 4.2, 1H), 7.47 – 7.61 (m, 4H), 7.63 – 7.74 (m, 2H), 8.05 (d, *J*=7.9, 2H), 9.00 (d, *J*=3.9, 1H); ^{19}F NMR (377 MHz, $CDCl_3$) δ = -71.36 (d, *J*=6.6); ^{13}C NMR (101 MHz, $CDCl_3$) δ = 29.5, 32.8, 40.4, 52.0, 53.1, 62.3, 72.7, 79.9, 116.7 (q, *J*=291.5), 121.3, 129.4, 130.3, 130.7, 131.1, 132.7, 140.5, 170.2, 180.1 (q, *J*=35.1); *m/z* (ESI$^+$) 529 ([M+H_2O+H$^+$]$^+$, 100%), 543 ([M+MeOH+H$^+$]$^+$, 100%); HRMS (ESI$^+$) $C_{26}H_{26}N_6O_2F_3^+$ ([M+H$^+$]$^+$) requires 511.2051, found 511.2069.

8.1.14. 1-[4-(bromomethyl)phenyl]-2,2,2-trifluoroethan-1-one[323] (54)



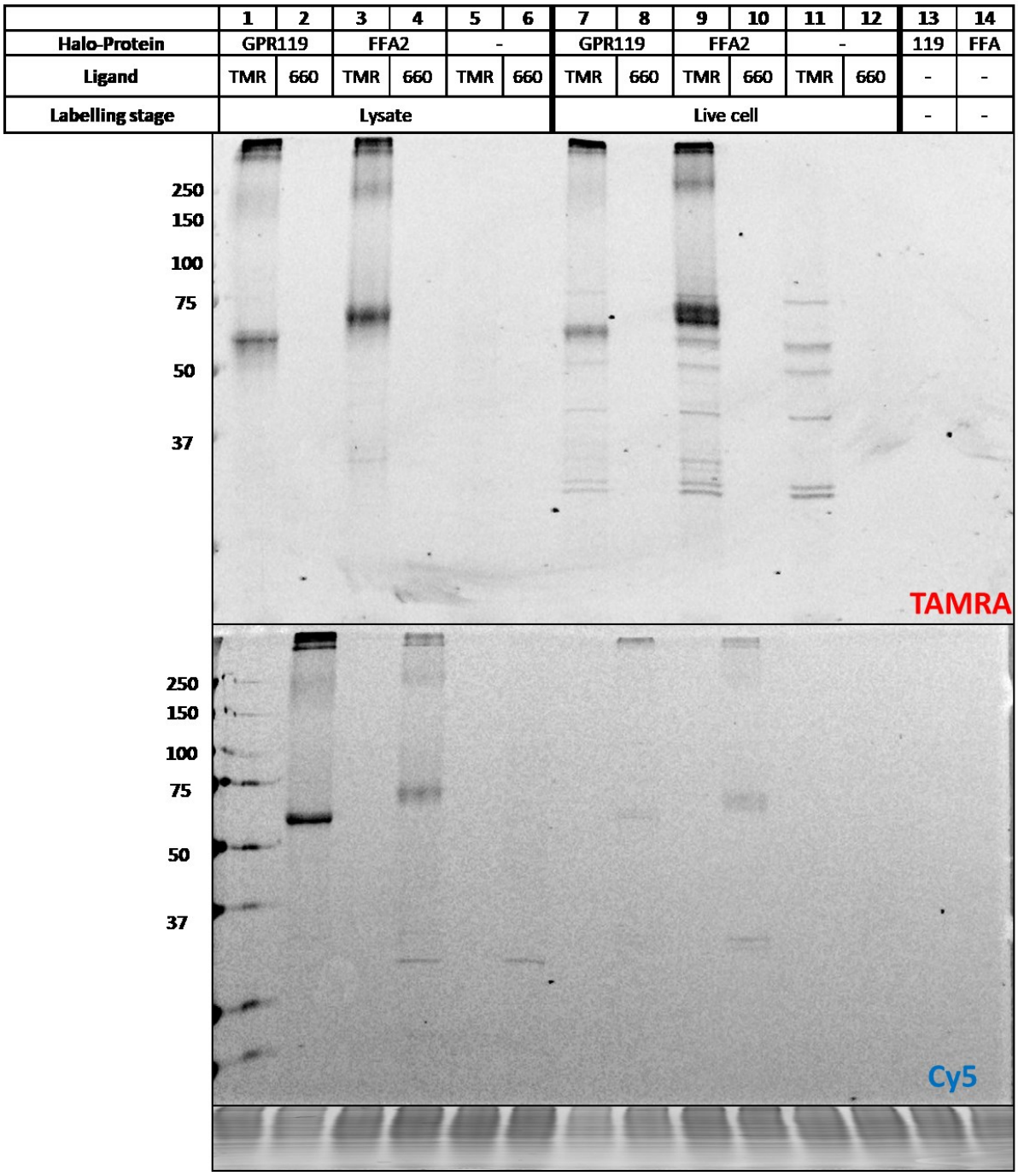
54

4'-Methyl-2,2,2-trifluoroacetophenone (376 mg, 2.000 mmol) was dissolved in acetonitrile (4 mL) under an inert atmosphere. *N*-bromosuccimide (392 mg, 2.200 mmol) and azobisisobutyronitrile (32 mg, 0.200 mmol) were added, and the solution

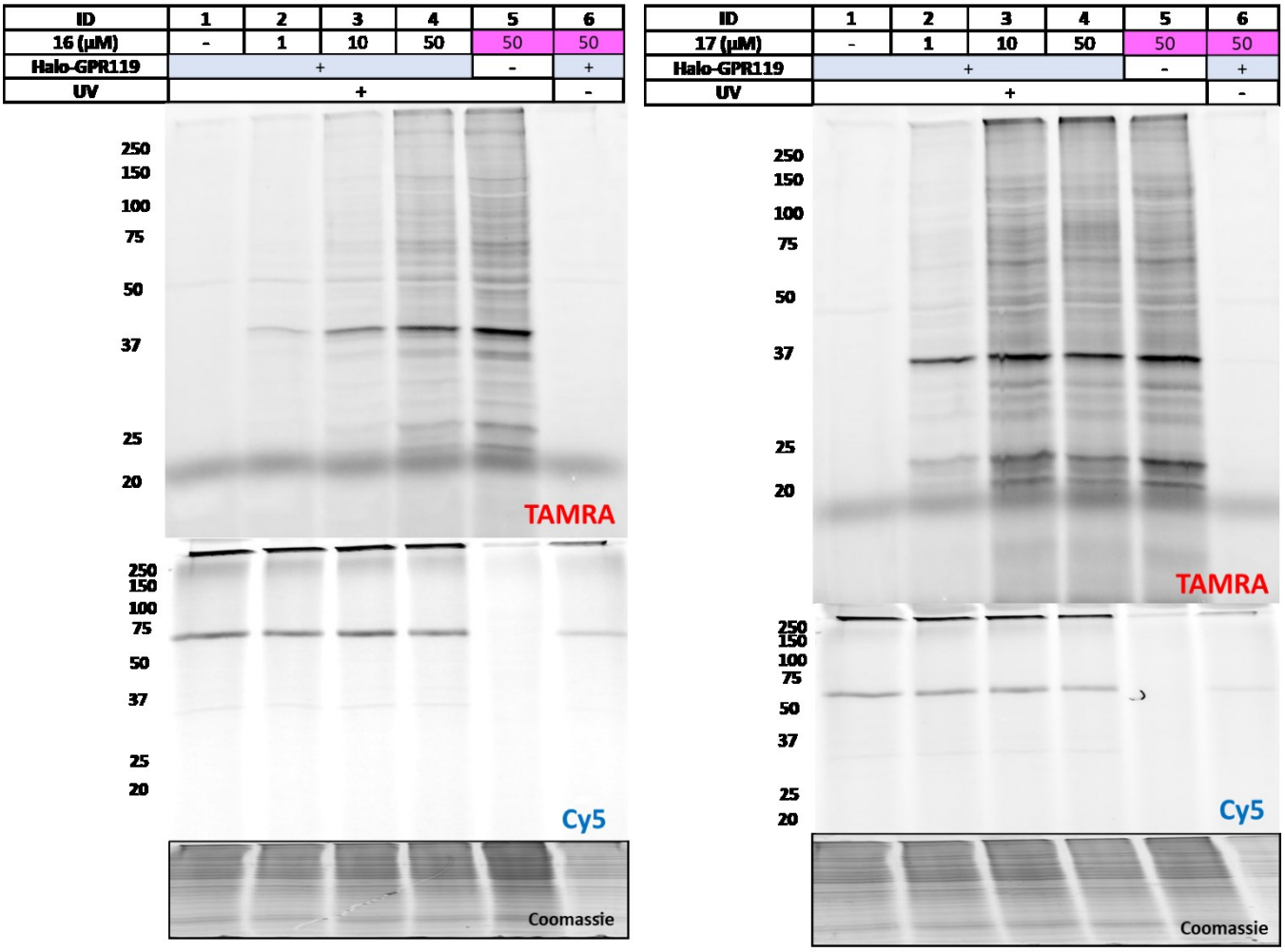
stirred for 4h at 90 °C. The solvent was removed *in vacuo* and the residue taken up in EtOAc (10 mL), washed with brine (2 x 10 mL), dried and concentrated *in vacuo*. Purification via column chromatography (eluent ethyl acetate:hexane 10:90) afforded product **54** as a clear oil (178 mg, 35%).

^{1}H NMR (400 MHz, $CDCl_3$) δ = 4.54 (s, 2H), 7.56 – 7.63 (m, 2H), 8.08 (d, *J*=7.9, 2H); *m/z* (EI^+) 188 ([M^+-Br], 100%), 266 [($M(Br^{79})^+$, 10%)], 268 [$M(Br^{81})^+$, 10%)].

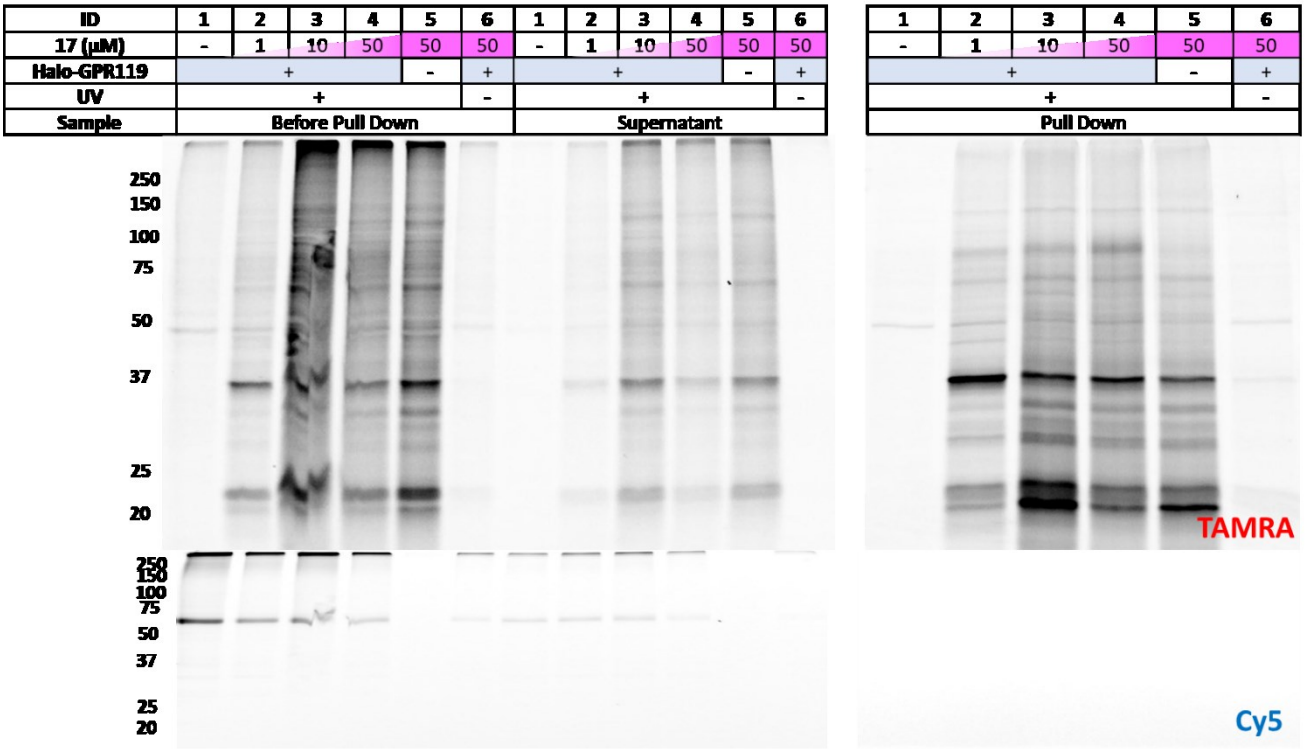
8.2. Appendix 2: Black and White Dual Fluorescence Images



S1:Black and white fluorescence image of the dual fluorescence as shown in **Figure 57**. Fluorescence was imaged in the TAMRA filter for HaloTag protein labelled with TMR-Cl ligand, and with the cy5 filter for HaloTag protein labelled with Alexa660-Cl ligand.



S2: Black and white fluorescence image of the dual fluorescence as shown in **Figure 69**. Dual fluorophore labelling of HaloTag-GPR119 (Cy5) and probe **16** [left] and **17** [right] (TAMRA).



S3: Black and white fluorescence image of dual fluorescence as shown in **Figure 70**. Dual fluorophore labelling of HaloTag-GPR119 (Cy5) and probe **17** (TAMRA) and enrichment using streptavidin.