DOCTOR OF PHILOSOPHY

Recombinant production and investigation of neurotransmitter channels and transporters

Davis, Samuel

Award date:
2022

Recombinant production and investigation of neurotransmitter channels and transporters

Samuel Davis

PhD

University of Dundee

2022

Declaration

I hereby declare that the following thesis is based upon the results of investigations conducted by myself, and this thesis is of my own composition. This thesis has not in whole or part, been previously presented for a higher degree. Work other than my own is clearly indicated in the text by reference to the relevant researchers or their publications. Those people, who allowed me to refer to their unpublished observations or data, are acknowledged individually where they occur.

Samuel Davis

22/08/2022

Summary

Channels and transporters are essential proteins for the control, mediation and termination of neurotransmission. These are implicated in numerous pathological conditions, including epilepsy, Parkinson's disease, neuropathic pain and nicotine addiction. However, the structural and ligand binding aspects of many of these channel and transporter proteins are poorly defined, which limits being able to design new molecules that can effectively target these conditions.

It was aimed to investigate structure and ligand binding at neurotransmitter channels and transporters. The first aims involved elucidating the binding modes and structure-activity relationships of novel ligands at nicotinic acetylcholine receptors (nAChRs), using the surrogate protein acetylcholine binding protein (AChBP). The ligands being characterised were of interest as potential anti-smoking agents and as research tools for studying nAChRs. Binding data and protein complex crystal structures were obtained for several of these novel ligands and it was possible to identify residues which were potentially responsible for their modes of action and affinity to AChBP, and henceforth likely to nAChRs. Knowledge of these interactions could assist in the future design of other ligands targeting nAChRs.

The second set of aims were associated with attempting to establish methodologies for the efficient recombinant production of complex eukaryotic ion channels and neurotransmitter sodium symporters. The initial objective was the insect ligand gated ion channel resistance to dieldrin (RDL), which is a target for insecticides. *Sf*9 insect cells proved unsuitable for production as only a small amount of the total protein could be extracted with non-ionic detergents and it was implied that the majority of the protein was likely in an un-folded state. Mammalian HEK293 cells were more successful as the protein could be efficiently solubilised, but ultimately the yields of purified protein were too low for this to be a feasible approach.

There was more success with producing the human GABA transporter 1 (GAT1). This terminates the actions of the inhibitory neurotransmitter GABA by removing it from the synapse and is a therapeutic target for the control of epilepsy. Using *Sf*9 cells and a conventional baculovirus system showed initial success, but there

were ultimately problems with aggregation. Use of a recently described baculovirus system with an early *Drosophila* Hsp70 promoter however resolved these problems and led to high yields of purified GAT1. The obtained protein was suggested to be potentially suitable for future structural studies by single particle cryogenic electron microscopy (cryo-EM). Purified GAT1 was also used as a target to isolate recombinant nanobodies from a yeast library and these may be of assistance for increasing the size of the protein for cryo-EM studies.

Publications

Original Research Articles (Reproduced in Section S2)

Bueno, R. V.*, **Davis, S.***, Dawson, A., Ondachi, P.W., Carroll, F.I., and Hunter, W.N. (2022). Interactions between 2′-fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues and acetylcholine binding protein inform on potent antagonist activity against nicotinic receptors. Acta Crystallogr. Sect. D Struct. Biol. *78*: 353–362. (* Joint first authors).

Davis, S., Rego Campello, H., Gallagher, T., and Hunter, W.N. (2020). The thermodynamic profile and molecular interactions of a C(9)-cytisine derivative-binding acetylcholine-binding protein from Aplysia californica. Acta Crystallogr. Sect. F Struct. Biol. Commun. *76*: 74–80.

Poster Presentations

Davis, S., Bueno, R. V., Dawson, A., and Hunter, W.N. (2018). Mode of action of novel epibatidine analogues. Young Life Scientists Symposium 2018, University of Strathclyde, Glasgow, UK.

Acknowledgements

I would firstly like to thank my two supervisors, Bill Hunter and Tim Hales, for all their invaluable support and advice during the ultimately four and half years that it took to submit this thesis, which was just slightly derailed by an unexpected global pandemic.

I would like to thank everyone, past and present, from within my supervisors respective lab groups. From Bill's group, thanks to Mat Jones, Genady Pankovs, Renata Bueno and especially thanks to Carla Gottschald-Chiodi for guiding me through the first few weeks in the lab (and for being a great friend in the years since!). From Tim's group, I thank Daniel Baptista-Hon and Dave Walker for directing me around the lab and assisting with the functional assays. Thanks also to Alice Dawson, Paul Fyfe, Pierre Petit, Joel McMillan, Greg Stewart, Avinash Punekar, Sharon Shepherd and Alison Scott for all your many varied contributions over the years.

I would also like to extend an enormous amount of gratitude toward Tom Eadsforth for his assistance with so many of the experiments which I ended up conducting. I do miss us sharing an office together, with the Nespresso machine permanently on the go.

I would additionally like to thank Paul Wyatt and Doreen Cantrell on my thesis monitoring committee for all of their guidance and advice throughout my PhD. Their expertise, comments, critique and personal support were invaluable towards me ultimately completing my PhD.

Thanks also goes to the many experts who helped me with various techniques. To Paul Appleton and Alan Prescott for their help with the confocal microscopy and negative stain EM, and to Dougie Lamont and the rest of the mass spec facility for their assistance on numerous occasions. This is especially for when I was unsure of what exactly it was that I wanted them to do for me…

To anybody else who I've neglected to mention that provided materials, reagents, advice and other resources or support, thank you very much!

I would also like to raise an especially big thanks and as much gratitude as possible to Lucy Barnsby-Greer for being such a wonderful friend to me. For all

of her support and personal encouragement and especially for taking on proofreading this thesis. I've deeply enjoyed discussing about each other's research and you've given me some great ideas over the years. Plus, I'm so very grateful for the many game nights, lunches, walks and all the countless other things that we've done together in the years after we met right on the first day of the PhD. These went a long way towards keeping me sane and staying on course with my PhD during the past few years and I shall never forget them.

Thanks also to Jordana Freemantle for all of her invaluable help with the confocal microscopy. Plus, also for all of our chats on many varied topics, both science related and otherwise. I am glad that the cake which I made and brought along each time always went down so well!

Lastly, thanks to my Mum and Dad for everything. You are really missed Dad and I know that you would proud of me for getting to the end after so many difficult times during the past four and half years.

COVID-19 Statement

The research contained in this thesis was impacted by the unexpected COVID-19 global pandemic. In March 2020, practical lab research had to be abruptly ceased and it was not possible to restart such work for approximately six months. Even at this point, occupancy limits in laboratory space and shortages of consumables amongst other factors imposed by the pandemic imparted severe constraints on the work which could be conducted. The rise of the Alpha COVID-19 variant in December 2020 later led to a further period of being unable to conduct lab work until May 2021 and pandemic related issues continued to impose limitations up until submission of this thesis. Ultimately, these factors led to a reduction in the scope of the work which could be completed in the time available.

Table of Contents

List of Abbreviations

5-HT	5-hydroytryptophan
ABC	ATP binding cassette
Ac	*Aplysia californica*
AChBP	Acetylcholine binding protein
Am	*Apis mellifera*
ATP	Adenosine Triphosphate
AUC	Analytical ultracentrifugation
BCA	Bicinchoninic acid
BGT	Betaine-GABA transporter
BN-PAGE	Blue native polyacrylamide gel electrophoresis
BSA	Bovine serum albumin
BSC	Biological safety cabinet
CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
CHO	Chinese Hamster Ovary
CHS	Cholesteryl hemisucinate
CMC	Critical micellar concentration
CNS	Central nervous system
DAT	Dopamine transporter
ddH_2O	Double distilled water
DDM	n-Dodecyl-b-D-Maltopyranoside
DDT	Dichlorodiphenyltrichloroethane
Dm	*Drosophila melanogaster*
DM	n-Decyl-β-D-Maltoside
DMEM	Dulbeccos modified Eagles medium
DMNG	Decyl Maltose Neopentyl Glycol
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbeccos phosphate buffered saline
DTT	Dithiothreitol
EB	EarlyBac
ECD	Extra-cellular domain
ECL	Enhanced chemiluminescence
EDO	Ethane-1,2-diol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation

FBS	Foetal bovine serum
FC	Fos-choline
FMP	FLIPR membrane potential
FSEC	Fluorescent size exclusion chromatography
GABA	γ-aminobutyric acid
GAT	GABA transporter
GFP	Green fluorescent protein
GluCl	Glutamate-gated chloride channel
GlyT	Glycine transporter
GPCR	G-protein coupled receptor
GRD	GABA/glycine-like receptor of *Drosophila*
HBSS	Hanks balanced salt solution
HCH	Hexachlorocyclohexane
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HisCl	Histamine-gated chloride channel
HRP	Horseradish peroxidase
Hs	*Homo sapiens*
ICD	Intra-cellular domain
ITC	Isothermal titration calorimetry
IUPHAR	International Union of Basic and Clinical Pharmacology
LB	Lysogeny Broth
LCCH3	Ligand-gated chloride channel homologue 3
LDAO	Lauryldimethylamine oxide
LeuT	Leucine transporter
LGIC	Ligand gated ion channel
Ls	*Lymnaea stagnalis*
mEm	mEmerald
ms	Milliseconds
nACh	Nicotinic acetylcholine
NAG	N-acetyl-D-glucosamine
NET	Noradrenaline transporter
NMR	Nuclear magnetic resonance
NSS	Neurotransmitter sodium symporter
OG	n-Octyl-β-D-Glucopyranoside
PBE	PBS plus 0.5% BSA and 2 mM EDTA
PBS	Phosphate buffered saline

PBST	Phosphate buffered saline plus tween 20
PCR	Polymerase chain reaction
PDB	Protein data bank
PDZ	Postsynaptic density 95/discs large/zona occludens 1
PEI	Polyethylenimine
PFA	Paraformaldehyde
PFB	pFastBac 1
PLGIC	Pentameric ligand gated ion channel
PTM	Post-translational modification
RDL	Resistance to Dieldrin
RMMS	Random matrix micro-seeding
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography with multi angle light scattering
SERT	5-HT transporter
Sf	*Spodoptera frugiperda*
SLC	Solute carrier
SMA	Styrene and maleic acid
SNRI	Selective noradrenaline re-uptake inhibitor
SOC	Super Optimal broth with Catabolite
SSRI	Selective 5-HT re-uptake inhibitor
TAE	Tris-acetate-EDTA
TCA	Tricyclic antidepressant
TEV	Tobacco etch virus
T_i	Inflection temperature
TMH	Transmembrane helix
TOPO	Topoisomerase
UDM	n-Undecyl-β-D-Maltoside
Vd	*Varroa destructor*
VDW	van der Waals
ZAC	Zinc activated channel

List of Amino Acids

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

List of Figures, Tables and Equations

Chapter 1: Introduction

Chapter 2: General Methods

Chapter 3: Nicotinic acetylcholine receptor ligand structure-activity relationships

Appendices

Chapter 1: Introduction

1.1.0.0: Transport across biological membranes

To maintain physiological function and homeostasis, it is essential that molecules can be passed in and out of cells. While small, highly lipid soluble molecules (e.g oxygen, CO_2) can cross cell membranes simply by diffusion, for charged or polar molecules diffusion is slow or impossible and therefore these require other methods.

Transporters are membrane bound proteins which facilitate the passage of molecules across these membranes. As depicted in Figure 1.1, there are two main types of transporter proteins, namely channels and carriers. Channels span the membrane and open in response to a stimulus such as ligand binding to allow rapid passage of molecules down a gradient. Carrier proteins (also referred to as carriers, permeases or as transporters in general) on the other hand bind the substrate and undergo a conformational change to release the substrate on the opposing side of the membrane. For substrates where the concentration on the other side is lower, the transport can then proceed passively down a concentration or ion gradient as appropriate. This is therefore referred to as "facilitated" diffusion or transport.

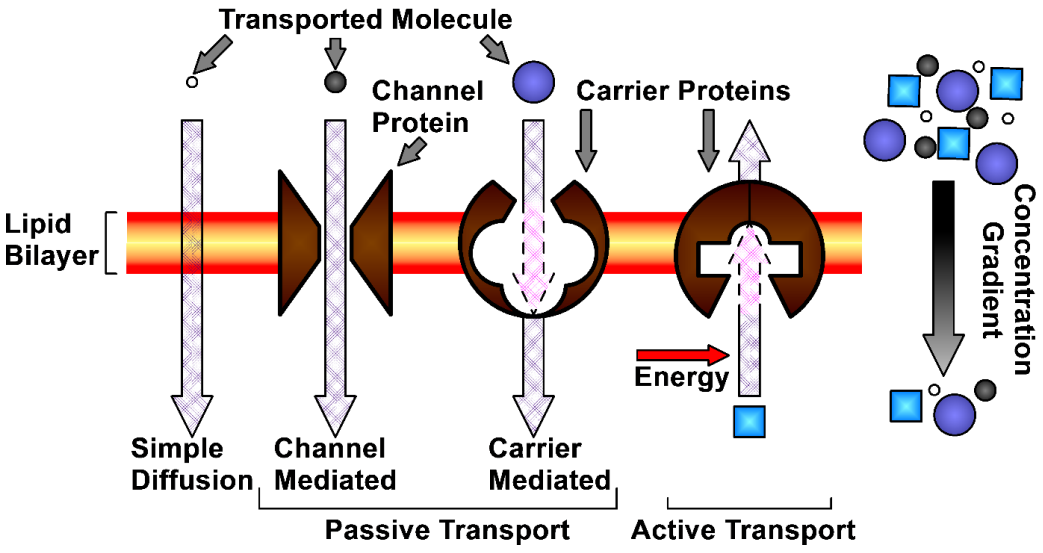


Figure 1.1: Simplified graphical depiction of the basic methods through which molecules can be transported across cell membranes.

Adapted from Stillwell (2016a).

For other substrates where it is necessary for the carrier protein to transport "uphill" against the gradient, active transport is needed. There are two main

mechanisms that can drive active transport. Primary active transport, as typified by the ATP binding cassette (ABC) transporter super-family and pumps, utilises hydrolysis of adenosine triphosphate (ATP) to drive transport against the gradient. Secondary active transport conversely couples the translocation of the substrate to the electrochemical gradient formed by the simultaneous co-transport of a partner ion, typically Na^+ or H^+, although there are multiple other possibilities. Secondary active transporters can operate in a symporter or antiporter mode, in the former the ions travel in the same direction as the substrate, but for the latter in an opposing direction (Alberts *et al.*, 2002; Guyton and Hall, 2006; Stillwell, 2016a).

1.2.0.0: Regulation of neurotransmission by transporter proteins

Neurons in the central and peripheral nervous system communicate via the process of neurotransmission. Neurotransmission is essential for maintaining body processes such as muscle contraction, vision, memory and homeostasis in general. Neurons are bridged via synapses, of which there are two types, namely electrical and chemical. While a minority of synapses directly connect adjacent neurons electrically using gap junctions, most synapses are chemically mediated via the release of neurotransmitters which bind to receptors on the subsequent neuron. Amongst many other diverse types, neurotransmitters can be monoamines, amino acids, purines, lipids or peptides. Examples of prominent neurotransmitters include adrenaline, acetylcholine, glycine, γ-aminobutyric acid (GABA) and adenosine triphosphate the chemical

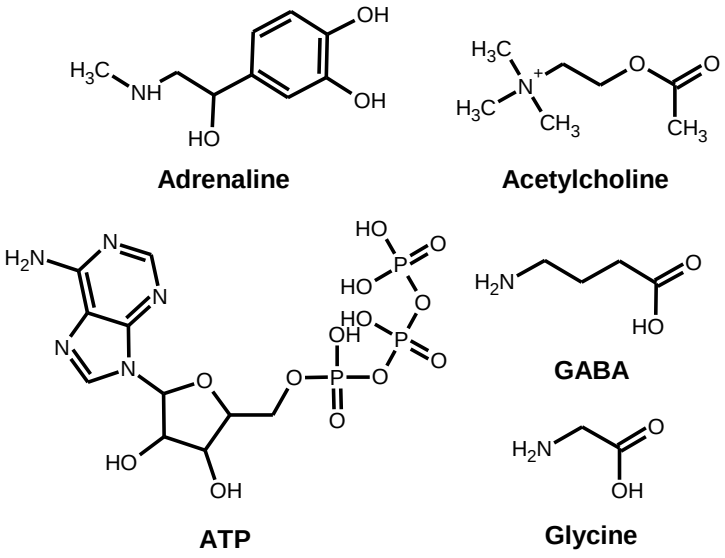


Figure 1.2: Chemical structures of neurotransmitters which are prominent examples.

ATP, adenosine triphosphate; GABA, γ-aminobutyric acid.

structures for which are depicted in Figure 1.2 (Purves and Williams, 2001; Hyman, 2005) .

An action potential is conducted along the neuronal axon until it reaches the pre-synaptic terminal. The depolarisation of the pre-synaptic membrane induces the opening of plasma membrane voltage gated Ca^{2+} channels, leading to an influx of the divalent cation. This triggers the exocytosis of small vesicles loaded with a neurotransmitter into the synaptic cleft. The neurotransmitters can then act on post-synaptic receptors, of which there are two main families. Metabotropic G-protein coupled receptors (GPCR) act to modulate ion channels and other downstream targets via G-protein complexes and second messenger systems on a timescale of hundreds of milliseconds (ms) to minutes or longer; this is hence referred to as slow synaptic transmission. With ionotropic ligand gated ion channels however, neurotransmitters directly act on them to mediate fast synaptic transmission with a timescale of <100 ms. Following post-synaptic receptor interaction, the neurotransmitters are either degraded or taken back up from the synaptic cleft into the pre-synaptic terminal by transporter proteins (Holz and Fisher, 1999; Purves and Williams, 2001; Hyman, 2005; Stillwell, 2016b).

1.3.0.0: Ligand gated ion channel mediated fast synaptic transmission

In mammals, fast synaptic transmission is mediated by three families of ligand gated ion channels, LGICs. These are the pentameric ligand gated ion channels (PLGIC), the trimeric P2X receptors and the ionotropic glutamate receptors (Alexander *et al.*, 2019b). A general overview of the PLGIC family is presented first, with specific information for individual proteins in subsequent sections.

1.3.1.0: Overview of pentameric ligand gated ion channels

There are four families of PLGICs in mammals. Additional members present in invertebrates will not be further described at this point. The $GABA_A$, glycine, 5-HT_3 and nicotinic acetylcholine (nACh) receptors are respectfully endogenously gated by GABA, glycine, 5-hydroytryptophan (5-HT [serotonin]) and acetylcholine. The $GABA_A$ and glycine gated receptors are anion selective inhibitory channels which pass Cl^-, while the nAChR and 5-HT_3 receptors are cationic selective excitatory channels. These families were formerly referred to

as Cys-loop receptors due to a conserved loop incorporating a disulphide bond, but under the current International Union of Basic and Clinical Pharmacology (IUPHAR) nomenclature they are denoted as PLGICs because this loop is absent in prokaryotic precursors (Alexander *et al.*, 2019a).

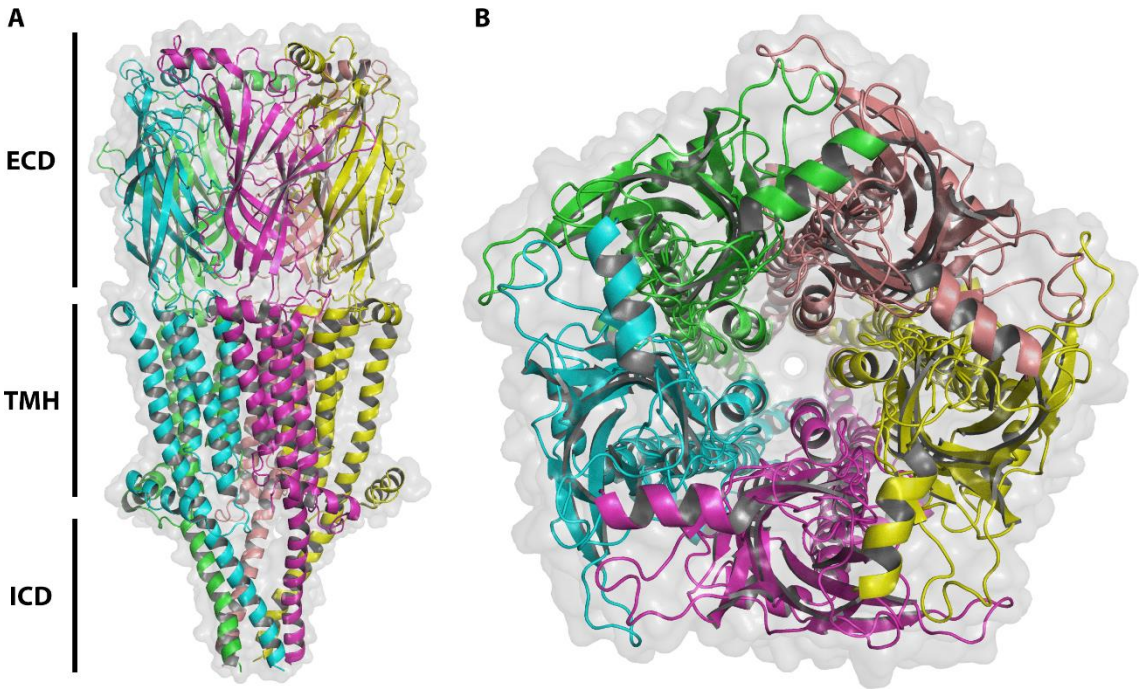


Figure 1.3: The characteristic structural arrangement of pentameric ligand gated ion channels as depicted with the homomeric human α_7 nicotinic acetylcholine receptor.

PDB ID 7EKI. **A**) Side view, **B**) Top view. Cartoon representation, coloured by sub-unit, with an overlaid surface representation. Only the immediate start and end of the intra-cellular domain (ICD) is depicted as the remainder was disordered in the sample and could not be modelled. ECD, extracellular domain; TMH, transmembrane helices.

PLGICs are characterised by a quaternary structure of five sub-units around a central ion pore, as shown with the human nicotinic acetylcholine α_7 receptor (Figure 1.3). Each sub-unit is comprised of an extra-cellular domain (ECD) formed mainly of β-strands, followed by four transmembrane α-helices (TMH). An unstructured, long intracellular loop is also present between TMH3 and TMH4. TMH2 from each subunit lines the ion pore and three rings of charged/polar residues comprise the ion selectivity filter. TMH1 and TMH3 shield TMH2 from surrounding lipids, while TMH4 faces and interacts with the lipid bilayer (daCosta and Baenziger, 2013; Thompson and Baenziger, 2020).The TMH3-TMH4 intracellular loop interacts with the cytoskeleton and has roles in single channel conductance, desensitisation, trafficking and localisation. The loop is frequently truncated in PLGIC structures as it is mainly disordered and removal does not

typically appear to grossly affect channel functioning (O'Toole and Jenkins, 2011; Baptista-Hon *et al.*, 2013; Rudell *et al.*, 2020)

Each PLGIC family contains several different subunits and in a pentameric complex these can be identical, forming homomeric receptors, or assorted, creating heteromeric receptors (Walsh *et al.*, 2018; Masiulis *et al.*, 2019; Howard, 2021). The large variety of subunits in each family potentially enables a great deal of receptor diversity and *in-vitro* many different combinations can be recombinantly created. Conversely, only a select few for each family have been conclusively identified *in-vivo* (Olsen and Sieghart, 2009; Alexander *et al.*, 2019a).

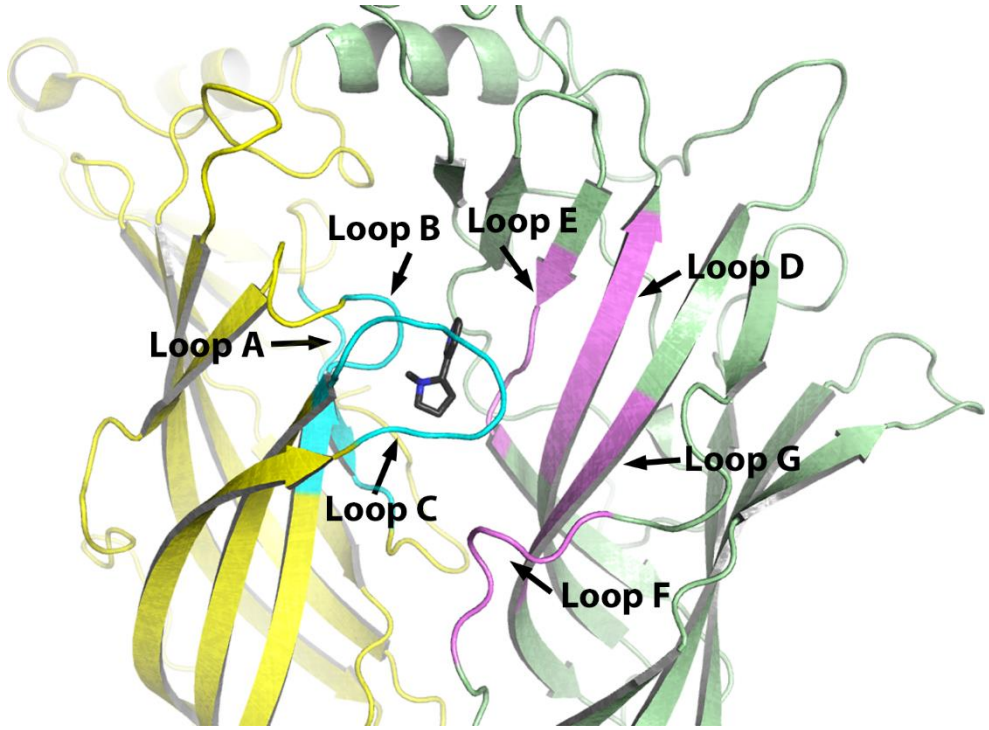


Figure 1.4: Loops comprising a representative PLGIC orthosteric ligand binding site.

A: Cartoon view of the 2 α_4:3 β_2 human $\alpha_4\beta_2$ nicotinic acetylcholine receptor (PDB ID 6CNJ) in the desensitised state with nicotine in the orthosteric site, highlighting the key loops contributed by each subunit. The (+) face α_4 subunit is coloured in yellow, with residues comprising the loops shaded cyan. The β_2 subunit contributing the (-) side is coloured light green with residues comprising the loops coloured magenta. Nicotine is shown as black sticks.

Orthosteric ligands bind at the interface between subunits, while allosteric ligands and other molecules such as neurosteroids bind at other distinct sites between heteromeric subunits or on the TMHs. Channel blocker molecules such as picrotoxin furthermore bind between the TMHs in the ion pore to inhibit passage of ions (Laverty *et al.*, 2017; Masiulis *et al.*, 2019). One subunit contributes the

principal (+) face and is formed of three loops known as A, B and C (Figure 1.4). On the adjacent subunit, loops D, E and F and G comprise the complementary (-) side of the binding site (Corringer *et al.*, 2012; Walsh *et al.*, 2018). Many of the key details of the orthosteric ligand binding site were originally elucidated from studies using acetylcholine binding protein (AChBP), a soluble homologue of the PLGIC ECD in various species which can be used as a simpler surrogate for ligand binding assays and structural studies (Hansen *et al.*, 2005; Rucktooa *et al.*, 2009). AChBP is described in further detail in section 1.3.2.0.

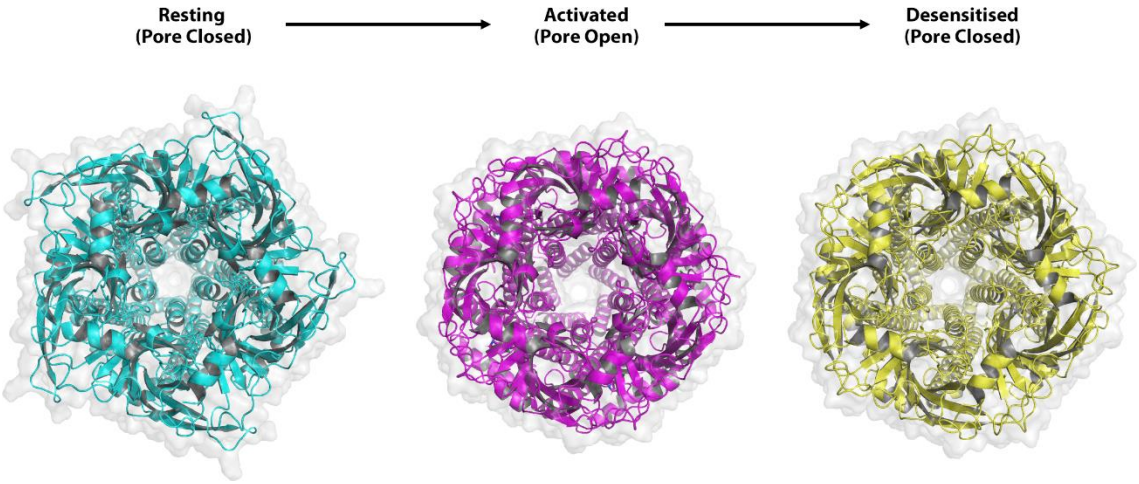


Figure 1.5: Conformations of a representative PLGIC in the resting, activated and desensitised states.

Top-down cartoon views with overlaid surface representations of the α7 nicotinic acetylcholine receptor in the resting state bound to the antagonist α-bungatoxin (PDB ID 7KOO), the activated state bound to the agonist epibatidine and a positive allosteric modulator (PDB ID 7KOX) and the desensitised state bound to epibatidine (PDB ID 7KOO).

Despite the differences between the subunits in differing PLGIC families, the key residues for ligand binding comprising the orthosteric binding site are well conserved. An “aromatic box” on the loops is formed by five key aromatic residues, with four on the principal face and one on the complementary face (Shahsavar *et al.*, 2016; Pless and Sivilotti, 2018). In the absence of bound agonist, the channel is in a low agonist affinity “resting” state, with the ion pore closed. Agonist binding in the orthosteric site leads to a conformational change (Figure 1.5), passing the receptor into a high affinity “activated” state. Loop C closes in on the agonist and movement of the binding site loops translates to the surrounding β-sheets and subsequently the TMHs. Movement of the TMH2-TMH3 linker leads to the pore lining TMH2 helices twisting and tilting outward, leading to the channel opening and ions diffusing down the electrochemical

gradient into the cell. The channel subsequently passes into a “desensitised” state, where the channel does not open despite having high affinity for agonists (Colquhoun and Lape, 2012; daCosta and Baenziger, 2013; Kumar *et al.*, 2020; Thompson and Baenziger, 2020). In nicotinic and glycine receptors, two additional intermediate states during activation designated as “flipped” and “primed” have also been identified. In these, an agonist is bound and affinity is increased, but the channel remains closed (Colquhoun and Lape, 2012).

Dysfunctions within the PLGICs are implicated with a wide variety of pathological conditions, such as epilepsy (Galanopoulou, 2008), nicotine addiction (Brunzell *et al.*, 2015), schizophrenia (Caton *et al.*, 2020) and Alzheimer's disease (Lombardo and Maskos, 2015). Correspondingly, they are the target of a range of clinically relevant drugs. The $GABA_A$ receptors for instance are targets for the control of anxiety and seizures via benzodiazepines such as diazepam (Rowlett *et al.*, 2005; Greenfield, 2013) and the $5\text{-}HT_3$ receptors are the site of action for anti-nausea -setron class drugs such as ondansetron (Basak *et al.*, 2019).

Also of note is the zinc activated channel (ZAC), which is included in its own PLGIC sub-group by the IUPHAR because of the low sequence homology (~15% maximum) to other human PLGIC subunits. It is thought to be a homopentamer which is gated by Cu^{2+}, Zn^{2+} and H^+ in ascending order of sensitivity. *In-vitro*, the channel is cation selective and permeable to Na^+, K^+ and Cs^+, but not Ca^{2+} or Mg^{2+}. Very little is known about the ZAC in comparison to other PLGICs as it is a pseudogene in the rat and mouse genomes, which has greatly limited research into its physiology *in-vivo* (Davies *et al.*, 2003; Trattnig *et al.*, 2016; Alexander *et al.*, 2019a).

1.3.2.0: Nicotinic Acetylcholine Receptors

Nicotinic acetylcholine receptors are cation selective pentameric ligand gated ion channels which are gated *in-vivo* by the neurotransmitters acetylcholine and choline (Alexander *et al.*, 2019b). They additionally comprise the binding site of multiple other pharmaceutical and toxin molecules, perhaps the most notorious being the drug of abuse nicotine (Kudryavtsev *et al.*, 2015). There are two types of nAChR, namely the muscle and the neuronal. Muscle nAChRs on the neuromuscular junction are comprised of 5 subunits which assemble from a pool

of α_1, β_1, δ, and γ subunits, although only two combinations form. These are comprised of α_1, β_1, δ and γ or α_1, β_1, δ, and ε subunits, with stoichiometry being 2:1:1:1 (Kalamida *et al.*, 2007). The nine alpha ($\alpha_{2\text{-}7,\ 9,10}$) and three beta ($\beta_{2\text{-}4}$) subunits expressed in the central nervous system however enable neuronal nAChRs to assemble in in a number of diverse possible combinations, which can be homo- or heteropentameric (Gotti *et al.*, 2009; Alexander *et al.*, 2019b). The most common combinations in the mammalian CNS are the $\alpha_4\beta_2$ heteromer, which has high affinity for nicotine, and the α_7 homomer, which comprise the majority of binding sites for the venom α-bungatoxin (Gotti *et al.*, 2006; Wu and Lukas, 2011).

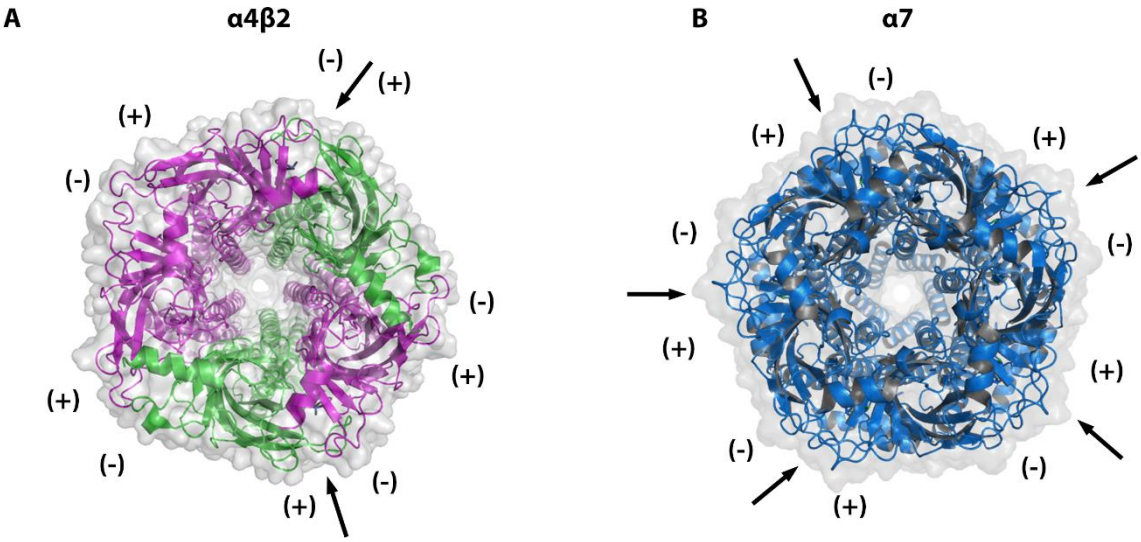


Figure 1.6: Structural arrangement and orthosteric binding sites of the heteromeric human $\alpha_4\beta_2$ and α_7 nicotinic acetylcholine receptors

A: Top-down cartoon with overlaid surface representation of the 2 α_4:3 β_2 human $\alpha_4\beta_2$ nicotinic receptor (PDB ID 6CNJ). α_4 subunits are coloured in green and β_2 in magenta. The principal (+) and complementary (-) sides of each subunit are indicated, with the locations of the orthosteric binding sites at the α_4(+)-β_2(-) interface indicated by arrows. **B:** Top-down cartoon with overlaid surface representation of the homomeric α_7 human nicotinic receptor (PDB ID 7KOX), with each subunit coloured blue. The principal (+) and complementary (-) sides of each subunit are indicated, with the locations of the orthosteric binding sites at the subunit interfaces indicated by arrows.

nAChRs follow the standard structural arrangement of other PLGICs as previously described, with five subunits around a central ion pore. In the case of the $\alpha_4\beta_2$ nAChR heteromer (Figure 1.6A), the orthosteric binding site is formed between the principal face of the α_4 subunit and the complementary face of the β_2 subunit. This therefore giving a total of two orthosteric sites per pentamer with a stoichiometry of either 2 α_4:3 β_2 subunits or 3 α_4:2 β_2. Both combinations form in the CNS, with the 2 α_4:3 β_2 having a higher affinity for nicotine (Moroni *et al.*, 2006;

Wilkerson *et al.*, 2020). For the α_7 homomer however (Figure 1.6B), a total of five orthosteric binding sites are formed, with a complex relationship between site occupancy and channel activation. Rayes *et al.* (2009) proposed that occupation of two to three sites is most efficacious, but occupation of more than three sites promotes desensitisation. Using a different assay, Andersen *et al.* (2013) however suggest that occupation of a single site is sufficient for maximal channel activation, with additional sites enhancing agonist sensitivity. In addition to the orthosteric site, it is important to note that potential “unorthodox” binding sites can also form at other subunit interfaces in a sub-population of nAChRs, which may act as sites for allosteric modulation (Wang and Lindstrom, 2018). For instance, a site which acts as a target for positive allosteric modulators forms between the α4(+):α4(-) interface in the 3 α4:2 β2 combination (Wang *et al.*, 2017).

It has only been within the past approximately ten years that full length structures of mammalian PLGICs have begun to be elucidated. Both protein production and structural studies represent considerable challenges even to this day. Advancements in the understanding of the structural architecture of PLGICs were previously imparted by a variety of methods. These included electron crystallography studies of the *Torpedo marmorata* muscle nAChR (Unwin, 2005), X-ray crystal structures of prokaryotic PLGICs (Hilf and Dutzler, 2008, 2009) and the *Caenorhabditis elegans* glutamate-gated chloride (GluCl) channel (Hibbs and Gouaux, 2011) and additionally nuclear magnetic resonance (NMR) studies of isolated TMHs in detergent (Bondarenko *et al.*, 2010). These were pre-dated however by studies on acetylcholine binding protein, which imparted the first high resolution information on the ECD and ligand interactions in the orthosteric site (Brejc *et al.*, 2001).

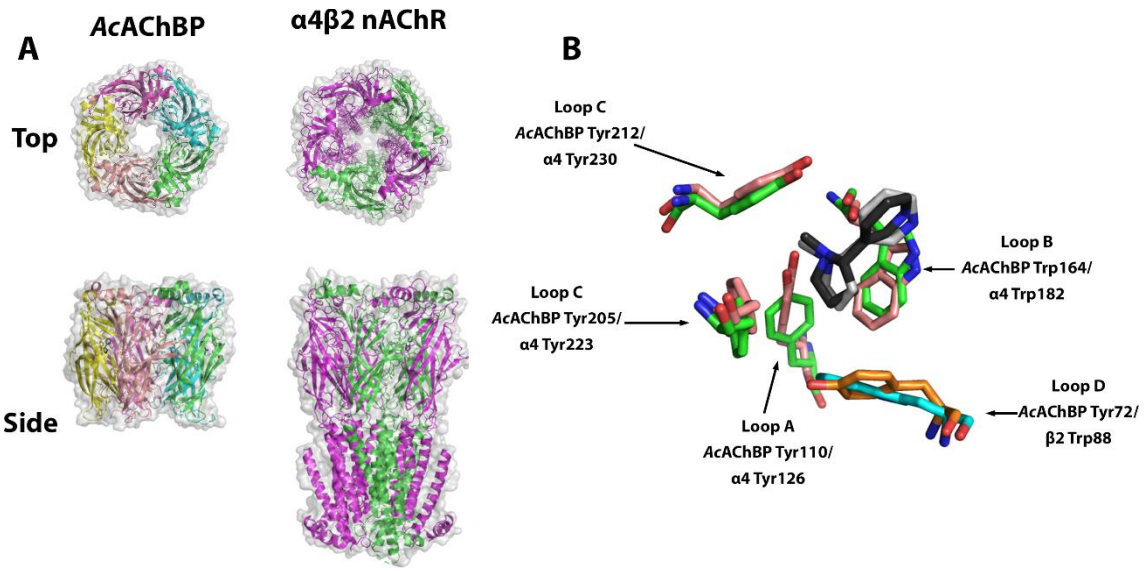


Figure 1.7: Comparison of *Ac*AChBP to the human $\alpha_4\beta_2$ nAChR

A: Cartoon with overlaid surface representations of the top and side of *Ac*AChBP (PDB ID 5O87) and the human $\alpha_4\beta_2$ (2 α_4:3 β_2 combination) nAChR (PDB ID 6CNJ). Each individual subunit of the homomeric *Ac*AChBP is coloured differently, while the heteromeric human $\alpha_4\beta_2$ nAChR is coloured with the α_4 subunits in green and the β_2 in magenta. **B:** Overlay of the key aromatic residues comprising a representative orthosteric binding site in AcAChBP and the human $\alpha_4\beta_2$ nAChR. Principal side residues are in green for the $\alpha_4\beta_2$ nAChR and salmon for AcAChBP, with the complementary side residues being in cyan and orange respectively. Nicotine is in the binding site, represented as grey sticks for *Ac*AChBP and black sticks for $\alpha_4\beta_2$ nAChR.

AChBP is a soluble, homomeric homologue of the ECD of nAChRs (Figure 1.7) which is expressed in various species, including *Lymnaea stagnalis* (*Ls*) and *Aplysia californica* (*Ac*); it is proposed to act as a signalling molecule. Despite having a sequence identity of only 20-24% with the ECD of nAChR subunits and 15-18% for other pLGIC families, the critical residues comprising the nAChR orthosteric binding site are highly conserved and the complex responds with comparable pharmacology to the homomeric α_7 nAChR. *Ls*AChBP directly conserves the five key aromatic residues in the nAChR orthosteric site, while *Ac*AChBP differs slightly by substituting a Trp for a Tyr, like in $GABA_A$ receptor subunits. The differing isoforms show some variations in pharmacology, with *Ls*AchBP having greater affinity for nicotine and lesser for certain other ligands compared to *Ac*AChBP (Brejc *et al.*, 2001; Smit *et al.*, 2003; Hansen *et al.*, 2005; Shahsavar *et al.*, 2016)

Because of these features, AChBP forms can be used as a “surrogate” for investigating the interactions of ligands at the orthosteric site of nAChRs. There are some differences however, especially for residues on the (-) side. To compensate, one strategy can be to engineer the loops comprising the AChBP

orthosteric binding site to be more like those of the relevant nAChR homo- or heteromeric subunit interface. Because of the highly conserved architecture of the PLGIC orthosteric binding site, AChBP can additionally act as a surrogate for other PLGIC family members if the binding site residues are modified to simulate the template of interest (Kesters *et al.*, 2013; Shahsavar *et al.*, 2015, 2016; Gottschald Chiodi, 2017; Dawson *et al.*, 2019; Jones, 2021).

Despite the recent releases of membrane PLGIC structures, including the key $\alpha_4\beta_2$ (Morales-Perez *et al.*, 2016; Walsh *et al.*, 2018), $\alpha_3\beta_4$ (Gharpure *et al.*, 2019) and α_7 (Noviello *et al.*, 2021; Zhao *et al.*, 2021) nAChRs, AChBP continues to represent a viable approach to studying binding in these proteins. Being a soluble, stable protein for which efficient production protocols exist, AChBP is relatively simple to produce and crystallise in comparison to membrane bound full length PLGIC proteins (Shahsavar *et al.*, 2016). This is especially prevalent as the structure-activity relationships for ligands acting at many nAChR subtypes are still poorly understood.

1.3.3.0: Insect pentameric ligand gated ion channels

Vertebrate PLGICs are now relatively well characterised. Invertebrate species however also possess their own, in some cases unique, PLGIC families. Focusing on insects, homologues to the mammalian $GABA_A$ and nAChR subunits have been found in native tissue and a total of 21-26 genes typically encode for PLGICs in insect species genomes sequenced so far. These include receptors which have no equivalent in vertebrates, such as the inhibitory GluCl and histamine-gated chloride (HisCl) channels (Jones, 2018).

For insect $GABA_A$ orthologs, the first identified and best characterised receptor is Resistance to Dieldrin (RDL), the gene for which (*Rdl*) was originally cloned from wild type *Drosophila melanogaster (Dm*) fruit flies that appeared resistant to cyclodine insecticides (Ffrench-Constant *et al.*, 1991). Expression of this cDNA in *Xenopus* for electrophysiology revealed an apparently homomeric functional chloride channel which appeared to respond to GABA and related agonists with a similar profile to mammalian $GABA_A$ receptors. Contrastingly, it was insensitive to the antagonist bicuculline, at the time considered the “hallmark” of a $GABA_A$ type receptor. The channel blocker picrotoxin, the cyclodiene insecticide dieldrin

and other non-competitive antagonists however blocked currents similarly to with mammalian $GABA_A$ receptors. Benzodiazepines showed no effects. (Ffrench-Constant *et al.*, 1993a; Buckingham *et al.*, 1994). *Dm* RDL has subsequently shown similar characteristics when expressed with various systems, including *Sf*21 insect cells (Lee *et al.*, 1993) and *Drosophila* S2 cells (Millar *et al.*, 1994).

In *Dm*, *Rdl* is an essential gene and *Rdl* null flies die in the early embryonic stage (Ffrench-Constant *et al.*, 1991), but overexpression is also typically lethal (Liu *et al.*, 2007). *Dm* RDL has been implicated with roles in visual and olfactory learning, sleep and aggression (Taylor-Wells and Jones, 2017). From immunocytochemistry and *in-situ* hybridisation studies, wide expression of RDL has been identified throughout the developing embryotic and adult *Dm* CNS. Generally this overlaps with tissue that stains for GABA, glutamic acid decarboxylase or for the synaptic vesicle protein synaptotagmin (Aronstein and Ffrench-Constant, 1995; Harrison *et al.*, 1996). Despite earlier studies suggesting that RDL was only expressed in the CNS, transcriptional analysis has also implicated a presence in the peripheral nervous system (Stilwell and Ffrench-Constant, 1998).

Genes homologous to *Dm Rdl* are present in diverse insect species that are widely apart phylogenetically with a high amino acid sequence identity, typically between 70-90%. *Dm* RDL is therefore viewed as being a model for RDL in general (Taylor-Wells and Jones, 2017). While most species only possess a single *Rdl* gene, some possess multiple copies. The pea aphid, *Acyrthosiphon pisum,* for instance possesses two *Rdl* genes which encode proteins with slightly differing sensitivities to GABA and non-competitive antagonists such as fipronil (del Villar and Jones, 2018).

The apparent homomeric organisation of RDL differs from that mammalian $GABA_A$ receptors, the majority of which are heteromeric with the most combination being $(2)\alpha_1(2)\beta_2(1)\gamma_2$ (Alexander *et al.*, 2019b). Compared with the major human $GABA_A$ subunits (Figure 1.8), *Dm* RDL is most similar to those of the β class, with 30.1% average sequence identity overall. It is then most similar the α class subunits at 26.3% and the γ subunits at 24.8%. The lack of sensitivity to bicuculline and benzodiazepines, but the expected response to picrotoxin and the homomeric organisation is pharmacologically similar to the

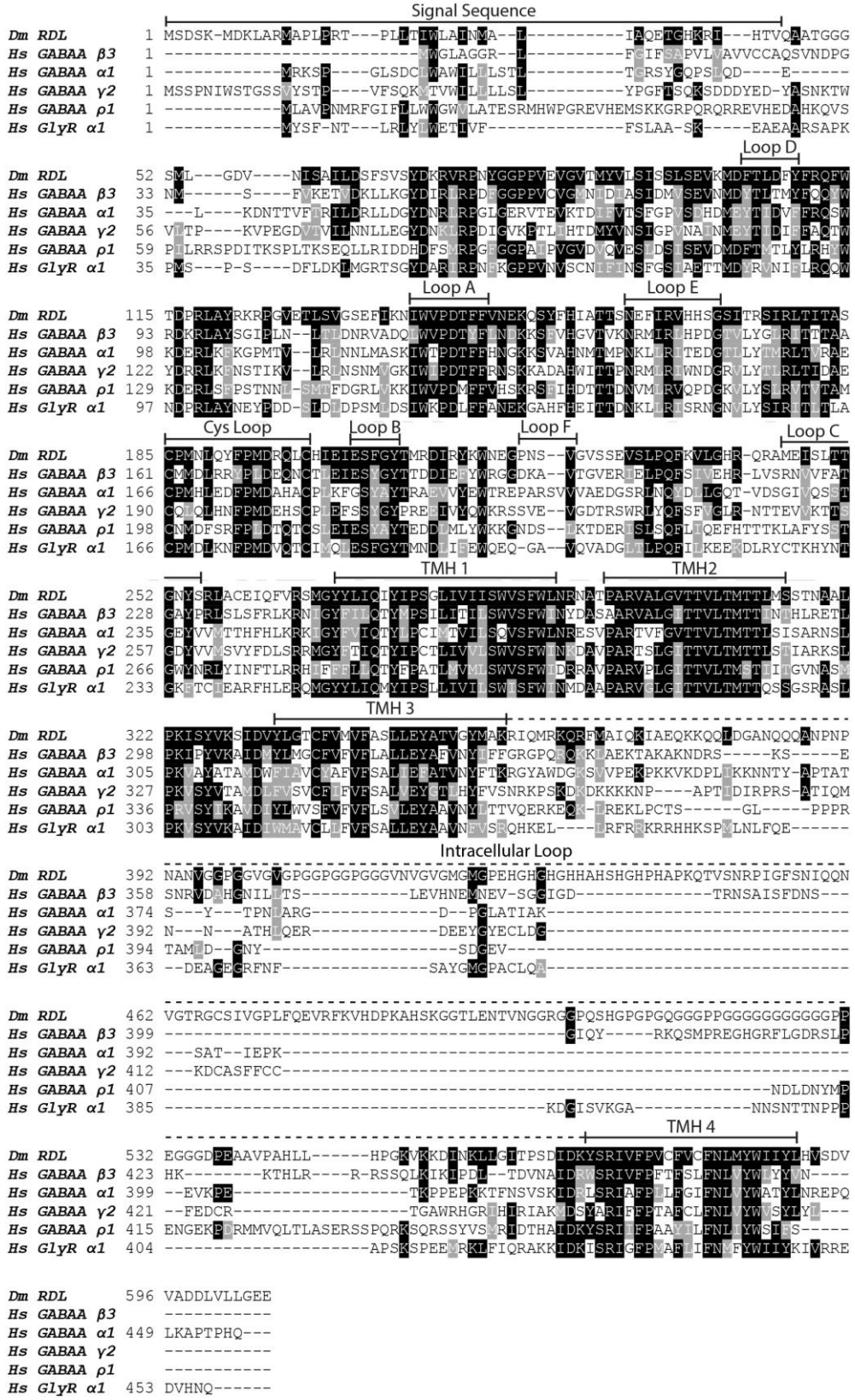


Figure 1.8: Clustal Omega sequence alignment of Drosophila melanogaster RDL with representative classes of human $GABA_A$ and glycine receptor subunits.

Shaded with BoxShade server (https://embnet.vital-it.ch/software/BOX_form.html) to indicate consensus to the RDL sequence (Uniprot ID P25123). Features of the RDL structural arrangement are indicated above the sequence. TMH, transmembrane helices. *Hs*, *Homo sapiens; Dm*, *Drosophila melanogaster.*

human $GABA_A$ ρ subunits (Naffaa *et al.*, 2017), although the average sequence identity is similar to that of other subunit families at 25.4%. When tested *in-vitro,* RDL responds to muscimol and the $GABA_A$ ρ selective ligand *cis*-4-aminocrotonic acid with full rather than partial agonist profiles (Buckingham *et al.*, 1994). RDL therefore does not fit into any established category of vertebrate $GABA_A$ subunits based on sequence identity and pharmacology. There is also considerable sequence homology (29.9% average) with human glycine receptor subunits, especially in the TMHs. Glycine however fails to activate RDL and strychnine has no effect (Chen *et al.*, 1994; McGonigle and Lummis, 2010). It has only recently been suggested that equivalents to glycine receptors may be present in selected insect species, although so far not in *D. melanogaster* (Jones, 2018).

From sequence homology and pharmacological similarities, RDL is considered as structurally similar to vertebrate $GABA_A$ receptors and other PLGICs, with five identical subunits forming a pentamer with 5-fold pseudosymmetry. The binding sites are similarly located, with the orthosteric site on the ECD between subunits. Other allosteric agents such as neurosteroids bind on the TMHs, while non-competitive channel blockers such as picrotoxin and dieldrin bind inside the ion pore. The RDL intracellular loop between TMH3 and TMH4 is of interest as it is longer compared to any mammalian PLGIC and is also where the greatest degree of dissimilarity occurs with vertebrate $GABA_A$ receptors and between insect species. The loop has been identified to contain phosphorylation sites which can affect receptor assembly, desensitisation and sensitivity to channel blockers (McGonigle and Lummis, 2010; Ashby *et al.*, 2012; Taylor-Wells *et al.*, 2017).

1.3.3.1: Diversity of RDL receptors

RDL is typically viewed as a homomeric receptor because when heterologously expressed alone it forms an active chloride channel that is gated by GABA, unlike the majority of vertebrate homomeric $GABA_A$ receptors (Wooltorton *et al.*, 1997; Gottschald Chiodi *et al.*, 2018). However, this may be an overly simplistic conclusion, as *Rdl* uses several post-transcriptional methods to increase diversity (Figure 1.9), including alternative and differential splicing and RNA editing (Taylor-Wells and Jones, 2017). In *Dm*, RDL consists of 9 exons, with exons 3 and 6 undergoing alternative splicing. This therefore results in four possible combinations, all of which are transcribed *in-vivo*. The variants on exon 3 are

referred to as isoforms a and b and those on exon 6 as c and d (ffrench-Constant and Rocheleau, 1993). Similar alternatively spliced variants have been described for *Rdl* in many insect species, although some species maintain their own unique variants (Taylor-Wells and Jones, 2017).

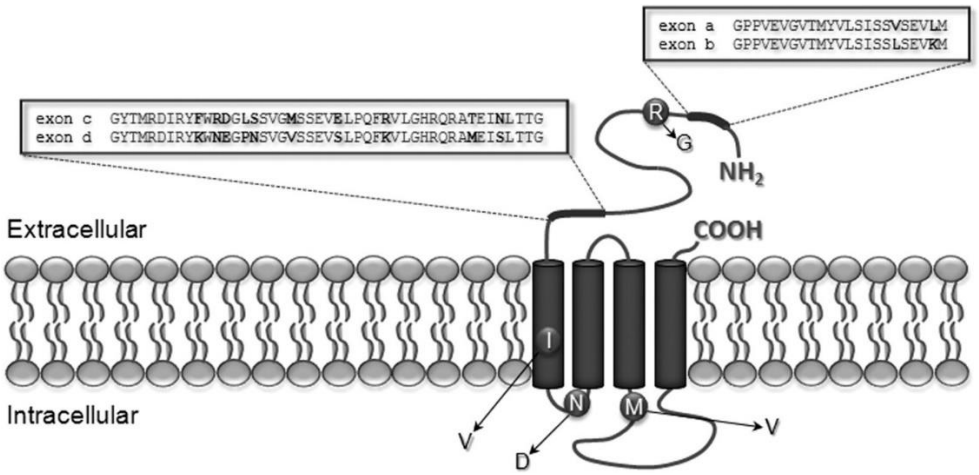


Figure 1.9: Graphical representation of post-transcriptional methods which can diversify *Dm* RDL subunits.

Alternative splicing forms two variants on exons 3 (a and b) and 6 (c and d), represented as amino acid sequences. Sequence differences between the isoforms are indicated in bold. RNA editing furthermore creates the R122G substitution on the ECD, I283V on the first TMH and N294D and M360V on the intracellular loops. Reproduced from Jones, A.K., Buckingham, S.D., Papadaki, M., Yokota, M., Sattelle, B.M., Matsuda, K., *et al.* (2009). Splice-Variant- and Stage-Specific RNA Editing of the *Drosophila* GABA Receptor Modulates Agonist Potency. J. Neurosci. 29: 4287–4292. Copyright 2009 Society for Neuroscience.

RDL_{bd} is the most common isoform in wild *Dm* in the adult stage, but in embryonic forms the RDL_{bc} variant is most prevalent, suggesting that certain isoforms are more important in different stages of development. Despite being the least common isoform in wild type *Dm*, RDL_{ac} is most frequently used in laboratory studies. These isoforms show varied pharmacology from each other, likely as the alternative splices change residues in loops C and F which form part of the orthosteric binding site. In terms of GABA sensitivity, the order is bc>ac>ad>bd (ffrench-Constant and Rocheleau, 1993; Buckingham *et al.*, 2005; Jones *et al.*, 2009; Lees *et al.*, 2014). It has also been identified that differential splicing can yield insertions in the intracellular loop, which can affect the sensitivity to channel blockers by potentially disrupting a protein kinase C phosphorylation site in RDL of various species (Taylor-Wells *et al.*, 2017).

Further diversity is induced through RNA editing, although in contrast to differential/alternative splicing, editing sites are highly species specific. Four RNA

editing sites have been identified on *Dm* RDL, with one each on the ECD (R122G), first TMH (I283V) and the M1-2 (N294D) and M3-4 (M360V) loops, leading to 16 possible combinations. As these are in structurally important locations, unsurprisingly they are associated with affecting ligand pharmacology (Taylor-Wells and Jones, 2017). R122G is between loops A and D of the orthosteric binding site for instance and causes an increase in the EC_{50} for GABA and the channel blocker fipronil (Es-Salah *et al.*, 2008).

The presence of these alternatively spliced isoforms and the identification of multiple genes encoding RDL in certain species has raised the suggestion that *in-vivo* RDL may assemble heteromeric complexes between different forms of RDL to increase receptor diversity (Taylor-Wells and Jones, 2017). Direct evidence in native tissue is currently limited though due to the difficulties of differentiating complexes containing such similar individual subunits. Overexpression studies potentially support this theory though, as heterologous co-expression of different RDL subunits from the honeybee parasite *Varroa destructor* (*Vd*) appeared to induce formation of heteromeric complexes that could be co-immunoprecipitated and showed altered pharmacology compared to individually expressed receptors. Of note is that the *Vd*RDL4 subunit appeared to be unable to form functional receptors except when co-expressed with *Vd*RDL1. Therefore, for some forms heteromeric assembly may be necessary (Ménard *et al.*, 2018).

The potential ability of RDL to form heteromers is further expanded by two genes transcribing other insect GABA subunits being identified originally in *Dm* and designated as *Lcch3* (ligand-gated chloride channel homologue 3) and G*rd* (GABA/glycine-like receptor of *Drosophila*) (Henderson *et al.*, 1993; Harvey *et al.*, 1994). When *Dm, Vd* or *Apis mellifera (Am)* European honeybee LCCH3 or GRD subunits are expressed alone *in-vitro*, they demonstrate no response to GABA or other ligands, suggesting that they do not form homomeric channels. This is consistent with them lacking key residues of the GABA binding site identified in both RDL and mammalian $GABA_A$ receptors. Co-expression of LCCH3 and/or GRD subunits in various combinations though, with or without RDL, appears to induce formation of functional channels in many cases and these can also be immunoprecipitated (Zhang *et al.*, 1995; Gisselmann *et al.*, 2004; Ménard *et al.*,

2018; Henry *et al.*, 2020). The stoichiometry of any of these combinations is unclear however and the data to date has shown large intra-species inconsistencies. For instance, *Dm* RDL co-expressed with LCCH3 appeared to form a chloride channel which intriguingly restored bicuculine sensitivity, but reduced sensitivity to picrotoxin (Zhang *et al.*, 1995). In contrast, when this combination was tested with *Am* RDL and LCCH3, the results did not show bicuculline sensitivity (Ménard *et al.*, 2018; Henry *et al.*, 2020)

These *in-vitro* studies could suggest that insects can use these additional subunits to further increase receptor diversity, in a similar way to vertebrates. Immunohistochemistry studies have so far not supported the *in-vitro* data though, as the spatial patterns of the subunits in the CNS are different (Aronstein *et al.*, 1996; Kita *et al.*, 2013). Currently there is no technique which can reliably differentiate heteromeric channels from homomeric ones in native tissue, so it is unknown if the subunits form heteromeric complexes as suggested from *in-vitro* studies. Electrophysiology studies using insect brain tissue have suggested pharmacological responses though that cannot be attributed to RDL alone based on the known pharmacology *in-vitro*, for example sensitivity to benzodiazepines (Hosie *et al.*, 1997; Buckingham *et al.*, 2005).

1.3.3.2: RDL as a target for insecticides

It may be questioned as to why RDL is seen as a target of interest. Much of the attention lies in how RDL is now known to be the binding site of certain insecticides, especially a sub-family of organochlorine agents that are referred to as cyclodienes and includes examples such as dieldrin, endrin, heptachlor and chlordane (Figure 1.10). These act as channel blockers at RDL by forcing a closed conformation, therefore stopping passage of chloride ions and preventing inhibitory transmission. This increases the likelihood of an action potential, which leads to convulsions and eventually death from over-stimulation.

Other non-cyclodiene agents acting at RDL include the organochlorine insecticides lindane (γ-hexachlorocyclohexane [HCH] and a mixture of its isomers) and toxaphene, the phenylpyrazole insecticide fipronil and the anti-parasitic drug ivermectin (Casida and Durkin, 2013, 2015; Lees *et al.*, 2014; Masiulis *et al.*, 2019). Because of the high degree of sequence homology for RDL between many insect species, agents acting on the receptor are therefore broadly

effective. Another sub-family of organochlorine insecticides, the dichlorodiphenylethanes, which most notably includes DDT (dichlorodiphenyltrichloroethane), do not act on RDL but rather target insect voltage gated sodium channels (Davies *et al.*, 2007).

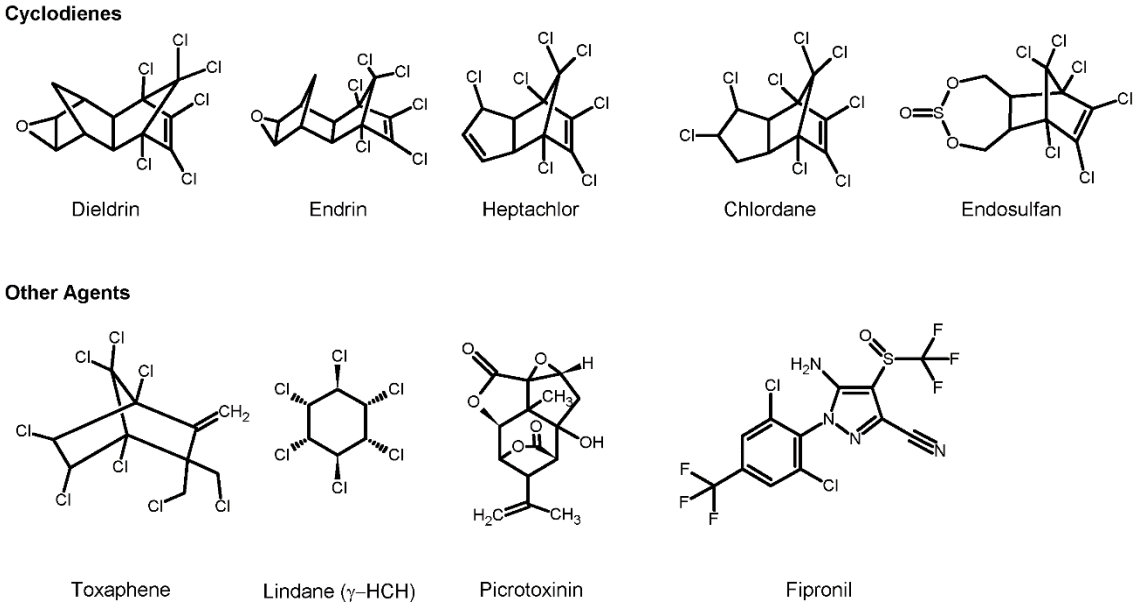


Figure 1.10: Chemical structures of representative insecticides which act at RDL.

HCH, hexachlorocyclohexane.

Worldwide, between the 1940's and 1970's organochlorine insecticides were extensively used in agriculture to kill "pest" insects that threatened crops and livestock. Additional uses included control of invasive insects (e.g. fire ants) and public health disease vectors, such as mosquitos. These agents were highly effective when introduced, but began to fall out of favour in the early 1970's when their potent negative environmental effects and high toxicity began to be recognised (Jorgenson, 2001; Jayaraj *et al.*, 2016). In particular, public attention was drawn by the publication of the book Silent Spring (Carson, 1962). This warned on the dangers of mass use of insecticide agents and highlighted their deleterious, long-term effects on the environment and other animals indiscriminately exposed.

Organochlorines have long half-lives, typically of multiple years, and a tendency to exert long-term toxic effects through bio-accumulating in animal tissues and being passed along the food chain (Ritter *et al.*, 1995; Jayaraj *et al.*, 2016). Eventually, restrictions on production and uses of organochlorines were imposed in many countries. Ultimately, this resulted in the Stockholm Convention on

Persistent Organic Pollutants, which near completely bans production and use in states party to the convention. There are limited exceptions permitted however, with DDT for instance still being utilised for vector control under restricted circumstances (UN Environment Programme, 2002).

1.3.3.3: The role of RDL in insecticide resistance

An additional problem of resistance rapidly developing to cyclodiene insecticides and organochlorines in general began to be recognised early on. Despite the extraordinary effectiveness of DDT against mosquitos upon its introduction in 1944, within five years reports of resistance emerged and subsequently mosquitos in many areas where DDT was used rapidly became resistant as well. In many cases, dieldrin or a related cyclodiene agent replaced DDT or vice versa. However, resistance rapidly developed to these as well and a similar pattern emerged in agricultural pests and many other insect species (Carson, 1962; Corbel and N'Guessan, 2013; Zare *et al.*, 2016).

In the case of cyclodienes, it was not until after their withdrawal in many countries that the cause of the insect's resistance to them was identified, when *Rdl* was cloned from *Dm* flies that were resistant to dieldrin. It was subsequently found that resistant flies universally have a mutation that results in an Ala to Ser or Gly substitution at residue 301 on the second TMH of RDL, which lines the ion pore. Either of these substitutions leads to a 10-10,000 fold resistance to cyclodienes and other similar channel blocker insecticides, with the glycine substitution appearing to generate greater resistance (Ffrench-Constant *et al.*, 1991, 1993b; Remnant *et al.*, 2014).

Identical Ala to Ser or Gly (or more rarely Asn) substitutions at the equivalent residue have since been widely found in RDL of wild type insects of diverse species and it is therefore commonly referred to as the A2' mutation (Taylor-Wells and Jones, 2017). Homology modelling of the mosquito *Anopheles gambiae* RDL using a picrotoxin bound human $GABA_A$ $\alpha_1\beta_3\gamma_2$ structure as the template suggests that these substitutions reduce sensitivity to dieldrin by altering the channel pore dimensions, which disrupts dieldrin-TMH interactions (Grau-Bové *et al.*, 2020). Despite inducing some negative fitness costs in the absence of insecticides, for example temperature sensitivity in *Dm* (Remnant *et al.*, 2013) and reducing mating success in *A. gambiae* (Platt *et al.*, 2015), the A2' mutation

is frequently found in areas where cyclodienes have not been used for decades and there is no advantage in retaining it. This could be explained by the long half-life of these agents, bio-accumulation still exerting a selection pressure and cross-tolerance to agents currently in use (Aronstein *et al.*, 1995; Wondji *et al.*, 2011; Domingues *et al.*, 2013). The presence of the A2' mutation has also been associated with other further resistance generating mutations appearing in parallel in diverse locations across the protein, some of which are thought to offset the A2' negative fitness costs (Nakao, 2017; Taylor-Wells and Jones, 2017).

1.4.0.0: Transporter mediated termination of neurotransmission

Following the interaction of neurotransmitters with post synaptic receptors, a critical subsequent stage is the rapid removal of these neurotransmitters from the synaptic cleft to prevent continued signalling. Some neurotransmitters such as acetylcholine are broken down by specific enzymes which are released following post synaptic receptor interaction (Hyman, 2005). For many neurotransmitters though, their removal is dependent on a further set of transporter proteins which take them back into the pre-synaptic terminal for recycling and reuse. There are two main families of transporter proteins involved in neurotransmitter re-uptake, namely the excitatory amino acid transporters and, of relevance to this work, the neurotransmitter sodium symporters (NSS). Both of these families are part of the solute carrier super-family (Focke *et al.*, 2013).

The solute carrier (SLC) super-family comprises the second largest group of transporters in the human genome, representing over 400 proteins that can be ordered into 65 families based on sequence homology. Solute carriers transport a wide variety of substrates, including but not limited to, neurotransmitters, amino acids, vitamins, ions, toxins and pharmaceutical drugs. While most solute carriers utilise facilitated diffusion along concentration or electrochemical gradients, others rely on secondary active transport (Hediger *et al.*, 2013; Schlessinger *et al.*, 2013).

1.4.1.0: Solute carrier 6 family

From the solute carrier super-family, the SLC6 family of Na^+/Cl^- dependent transporters is of particular interest for the roles of certain members in removing selected neurotransmitters from the synaptic cleft. These are commonly termed as the NSSs and are the main focus of this section. Additional transporters in the SLC6 family not involved in neurotransmission have functions such as regulation of nutrient uptake (Bröer and Gether, 2012; Focke *et al.*, 2013). Work on SLC6 transporters presented in this thesis focuses on the human SCL6A1 GABA transporter 1 (GAT1) and this section will therefore only go into broad details of the NSSs. Specific information on GAT1 is presented subsequently.

The NSSs and other SLC6 transporters function as secondary active co-transporters that operate in a symporter mode, utilising a Na^+ electrochemical gradient to uptake between one and three Na^+ ions and, for all the NSSs, also one Cl^- for one molecule of a substrate from outside the cell. Under certain conditions it is also possible for reverse transport to occur where the substrate is moved in the opposing direction (Pramod *et al.*, 2013; Alexander *et al.*, 2019b).

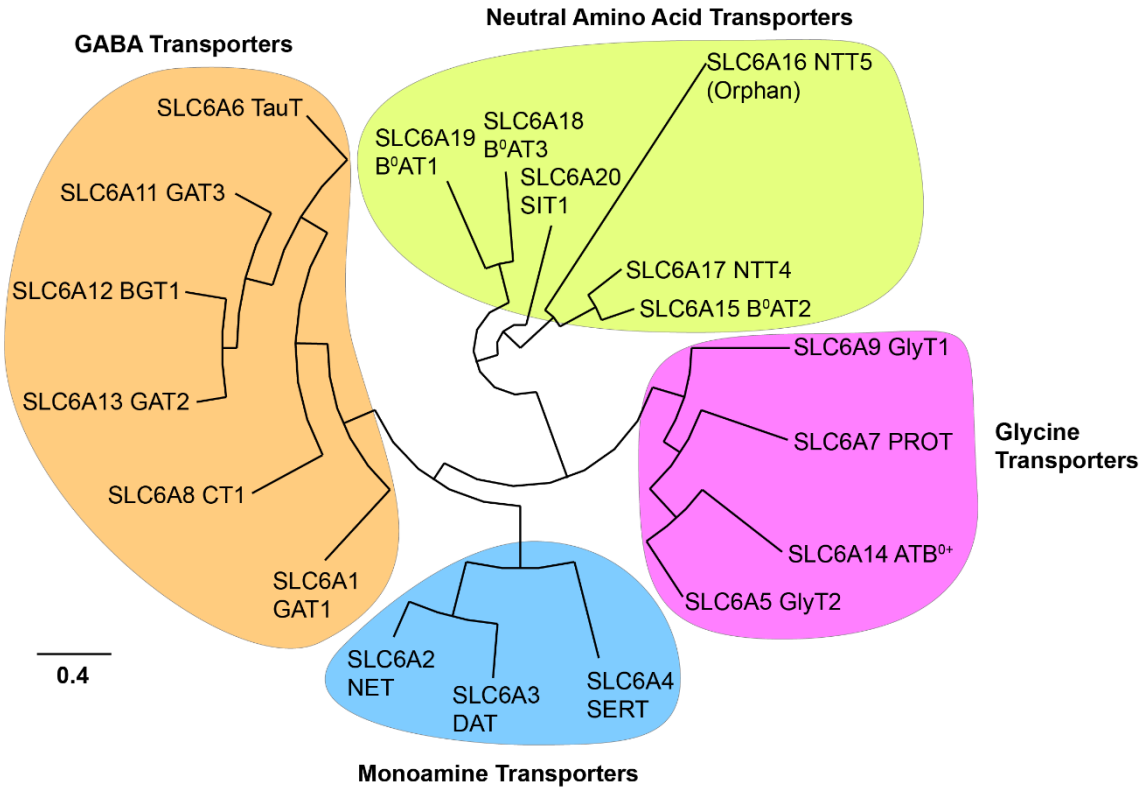


Figure 1.11: Phylogenetic tree of the solute carrier 6 transporter family.

Generated with the http://phylogeny.fr server. After sequence alignment with Clustal Omega, phylogenetitc reconstruction was undertaken with MrBayes and the tree rendered with TreeDyn. SLC6A10 is absent as it is a psuedogene of SLC6A8. The transporters are designated with the systematic nomenclature for solute carrier superfamily genes as assigned by the HUGO Gene Nomenclature Committee (e.g SLC6A4) and also their names as assigned under the IUPHAR nomenclature (correspondingly SERT for SLC6A4).

In the human genome, the SLC6 family is comprised of 19 genes (Figure 1.11), as well as a pseudogene (SCL6A10/CT2). By sequence similarity these can be further split into four sub-families, respectfully being designated by the IUPHAR as the GABA transporters, the monoamine transporters, the glycine transporters and the neutral amino acid transporters. It is important to note though that members of each group may carry vastly different substrates despite the naming convention, as this is based purely on sequence similarity. The NSSs are comprised of the GABA transporters GAT1, GAT2, GAT3 and the betaine-GABA transporter 1 (BGT1) of the GABA sub-family, the dopamine (DAT), noradrenaline (NET) and 5-HT (SERT) transporters of the monoamine sub-family and the glycine transporter (GlyT) 1 and GlyT2 of the glycine sub-family (Pramod *et al.*, 2013; Alexander *et al.*, 2019b). Properties of the NSS members in terms of their substrates, transport stoichiometries and selected drugs are depicted in Table 1.1.

Table 1.1: Properties of the neurotransmitter sodium symporters

Transporter*	Sub-family	Endogenous substrate(s)	Transport stoichiometry	Selected relevant drugs
GAT1	GABA	GABA	2 Na^+:1 Cl^-:1 GABA	Tiagabine
GAT2	GABA	GABA	2 Na^+:1 Cl^-:1 GABA	-
GAT3	GABA	GABA	≥2 Na^+:1 Cl^-:1 GABA	-
BGT1	GABA	Betaine GABA	3 Na^+:1 Cl^-:1 betaine or GABA	-
DAT	Monoamine	Dopamine	1-2 Na^+:1 Cl^-:1 dopamine	Amphetamines Cocaine
NET	Monoamine	Noradrenaline	1 Na^+:1 Cl^-:1 noradrenaline	TCAs SNRIs
SERT	Monoamine	5-HT	1 Na^+:1 Cl^-:1 5-HT + 1 K^+ in reverse direction	TCAs SSRIs
GlyT1	Glycine	Glycine	2 Na^+:1 Cl^-:1 glycine	-
GlyT2	Glycine	Glycine	3 Na^+:1 Cl^-:1 glycine	-

*These are as per the IUPHAR nomenclature and apply to the human/rat SLC6 transporters. The mouse SLC6 transporters differ in GAT2 being equivalent to the human/rat BGT1 and respectfully GAT3 to GAT2 and GAT4 to GAT3. GABA, gamma aminobutyric acid; 5-HT, 5-hydroxytryptophan; TCA, tricyclic antidepressant; SNRI, selective noradrenaline re-uptake inhibitor; SSRI, selective 5-HT re-uptake inhibitor. Adapted from Rang *et al.* (2011), Pramod *et al.* (2013) and Alexander *et al.* (2019b).

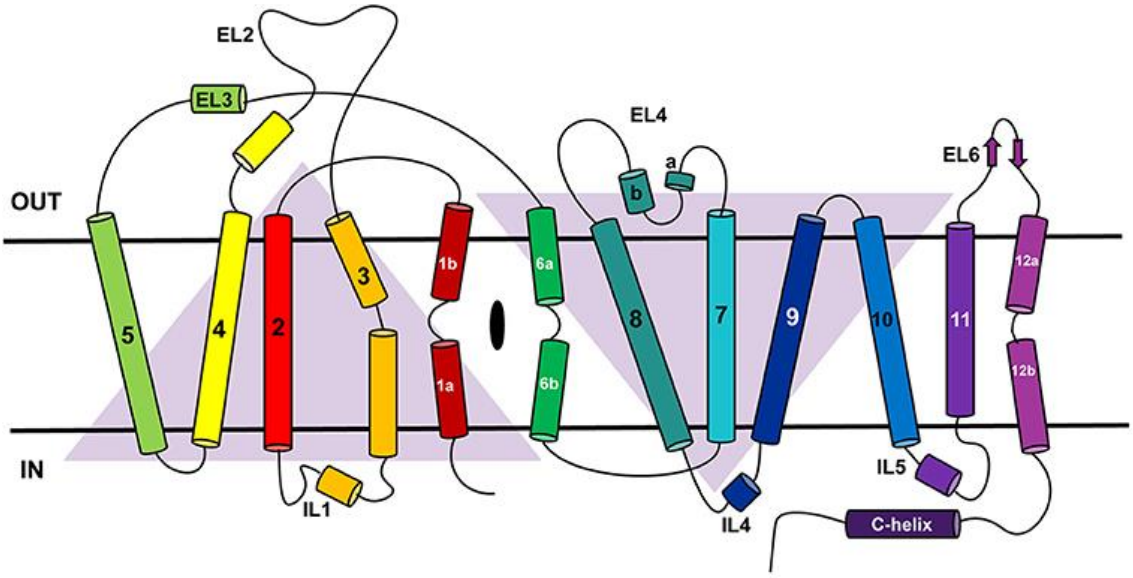


Figure 1.12: Topology of Solute Carrier 6 Transporters showing the 5+5 inverted repeat of the TMHs.

ECL, extracellular loop; ICL, intracellular loop. Reproduced from Joseph *et al.* (2019) under a Creative Commons CC-BY licence. Copyright Frontiers Media SA.

The SLC6 family have a common structural arrangement of twelve TMHs with a long extracellular loop between TMH3 and 4 and intra-cellular N- and C-termini (Figure 1.12) (Pramod *et al.*, 2013). This arrangement had been inferred for many years from techniques such as hydrophobicity analysis, site directed mutagenesis and alanine scanning, but confirmation came in 2005 when a crystal structure of the bacterial leucine transporter LeuT was solved at 1.65 Å resolution (Yamashita *et al.*, 2005). Although LeuT only has an average sequence identify of ~20% overall to SLC6 transporters, it increases to greater than 50% close to where substrates bind within the “core” of the protein. The LeuT fold consists of ten TMHs arranged into inverted repeats of five TMHs each (TMHs 1-5 and 6-10) with two axis pseudo symmetry on the membrane plane (Yamashita *et al.*, 2005). This fold has since been commonly found across the SLC super-family (Bai *et al.*, 2017) and later structures of *Drosophila* DAT (Penmatsa *et al.*, 2013, 2015), human SERT (Coleman *et al.*, 2016, 2019) and human GlyT1 (Shahsavar *et al.*, 2021) have shown an analogous arrangement.

In LeuT and the NSSs, substrates bind in the core of the transporter between TMHs 1, 3, 6, 8 and 10 within three sub-pockets formed by TMHs 1/6, 6/10 and 3/8 respectfully. Residues in these sub-pockets appear to dictate the substrate selectivity for each transporter sub-family. The co-transported Na^+ ions bind at two sites flanking the substrate and these are formed between TMHs 1/6/7 (Na1

site) and TMHs 1/8 (Na2 site) respectively. Although not present in LeuT, the co-transported Cl^- binds in the vicinity of the Na1 site in the NSSs (Joseph *et al.*, 2019; Colas, 2020).

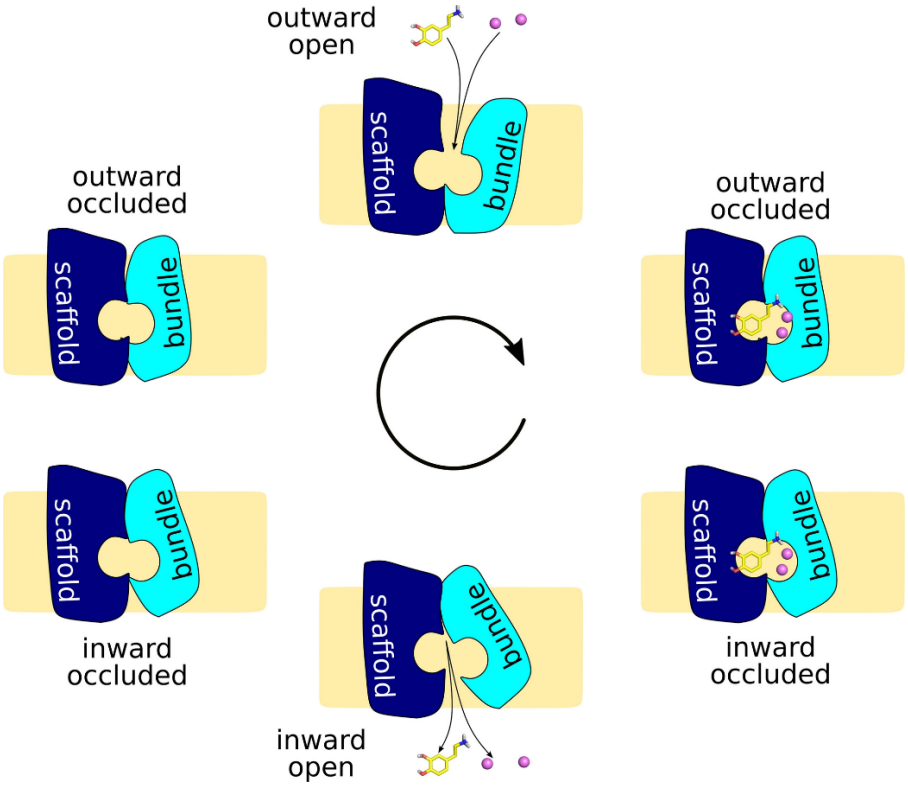


Figure 1.13: Graphical representation of the stages of the alternating access model.

The membrane, coloured in straw, is by spanned a transporter protein comprised of a rigid scaffold domain (in navy) and a "rocking" bundle domain (in cyan). The substrate is depicted as yellow sticks and the two ions as magenta spheres.

Based on the conformational states adopted by LeuT and other transporters showing the same or similar fold, the alternating access model provides a template for how secondary active transporters such as the NSSs may operate to transport a substrate while avoiding exposing it to both sides of the membrane during transport. In the model, depicted in Figure 1.13, the transporter possesses a rigid scaffold domain and a "rocking" bundle domain which moves and rotates to alternately block access to the outside and inside of the transporter and the cell. In the resting state, the transporter is bound with Na^+ and in the outward open conformation where the interior of the transporter can be accessed from outside the cell. Upon the substrate entering and binding inside the protein core, a conformational change is induced where the bundle domain shifts to crowd

around the substrate and occlude access to the outside. The transporter is hence in the outward occluded state. The bundle domain then moves the transporter to the inward occluded and subsequently the inward open state, where the substrate and ions are released into the cytoplasm. The transporter subsequently cycles through the inward and outward occluded states until the restring outward open state is reached again (Focke *et al.*, 2013; Jayaraman *et al.*, 2020).

The N- and C-termini of SLC6 transporters represent important sites for protein-protein interactions that regulate transporter trafficking and function (Pramod *et al.*, 2013). The T-SNARE protein Syntaxin-1A for instance interacts with the N-terminus to mediate the surface localisation and functioning of multiple NSSs (Quick, 2002; Carvelli *et al.*, 2008). Post-translational modifications are also frequently applied to multiple sites across the transporters, with the extracellular loop 2 between TMH3 and 4 for instance typically being a location for 1-4 N-linked glycosylation sites which are essential for transport activity (Cai *et al.*, 2005; Scott and Panin, 2014). NSSs may also form homodimers or higher order oligomers and it is suggested that this oligomerisation may be essential for some members to exit from the endoplasmic reticulum (ER) and translocate to the cell surface (Jayaraman *et al.*, 2020).

1.4.2.1: Neurotransmitter sodium symporters as drug targets

The role played by the NSSs makes them a rich area for drug targeting, as their activity can be enhanced or inhibited to reduce or prolong the time that a neurotransmitter is in the synaptic cleft after vesicle exocytosis. This modulates the degree of interaction with post synaptic receptors. There are however only a few examples of drug classes in current clinical use which target the NSSs. Aside from the GAT1 re-uptake inhibitor tiagabine (described in the subsequent section), these are near exclusively for the monoamine NSSs. Agents which inhibit re-uptake by NET and/or SERT are widely used for the control of depression. Examples include the tricyclic antidepressants (e.g amitriptyline) which inhibit both transporters, and the selective noradrenaline (e.g nortriptyline) and selective 5-HT (e.g fluoxetine) re-uptake inhibitors (Rang *et al.*, 2011). DAT represents the binding site of clinically relevant and abused stimulants such as amphetamines and cocaine. The former both competitively inhibits DAT and

induces reverse transport, while the latter only acts as an inhibitor. Either way, dopamine in the synaptic cleft is increased above normal levels (Zhu and Reith, 2008).

1.4.3.0: Human GABA transporter 1

The GABA transporter sub-family represents six of the 20 SLC6 transporters. GAT1, GAT2 and GAT3 are high affinity GABA transporters which are common in exchanging 2 Na^+:1 Cl^-:1 GABA. By sequence similarity, GAT1 has approximately 50% homology to GAT2 and GAT3, but these are respectively more similar to each other at ~65%. BGT1 differs in primarily transporting the osmolyte betaine, although it will also carry GABA with reduced affinity compared to the other GABA transporters. It also contrasts in exchanging 3 Na^+:1 Cl^-:1 GABA. By sequence homology BGT1 is more similar to GAT2 (68.4%) and GAT3 (63.1%) than it is to GAT1 (46.6%) (Pramod *et al.*, 2013; Zhou and Danbolt, 2013; Alexander *et al.*, 2019a). The remaining two transporters (CT1 and TauT) carry the nutrient creatine and the osmolyte taurine respectively, but these shall not be further described (Pramod *et al.*, 2013; Zhou and Danbolt, 2013).

GAT1 is the main GABA transporter in the human brain and is the focus of this section. It was the first SLC6 transporter to be isolated and purified - from rat brain tissue (Radian *et al.*, 1986). It was additionally the first to be cloned (Guastella *et al.*, 1990), which enabled the subsequent identification and cloning of other SLC6 transporters. Expression of GAT1 is common throughout the brain, but it is mainly present in neurons at axon presynaptic terminals and additionally in smaller quantities in ganglia. Expression patterns strongly follow those GABAergic neurons, being especially high in the basal ganglia, cerebellum, olfactory bulb, interpeduncular nucleus and retina. There are also smaller amounts of GAT1 present in non-GABAergic neurons and the glia, the functions of which are less clearly understood (Madsen *et al.*, 2007; Jin *et al.*, 2011; Pramod *et al.*, 2013).

1.4.3.1: GAT1 structure and trafficking

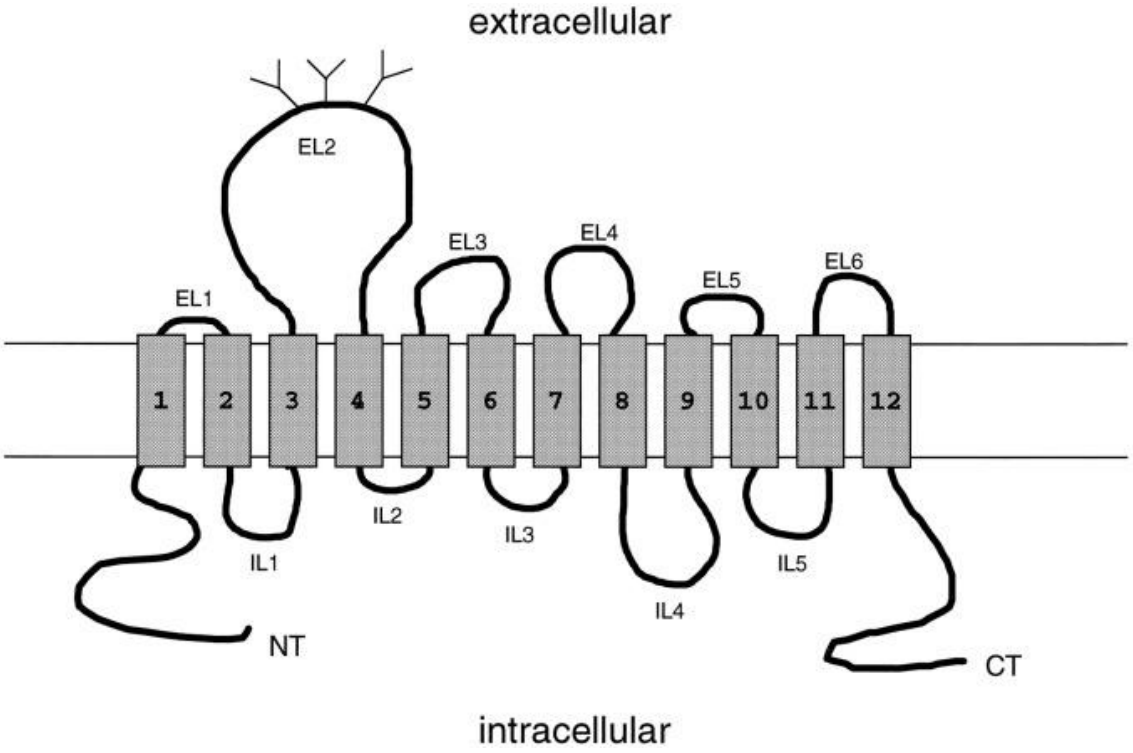


Figure 1.14: Topological depiction of GAT1.

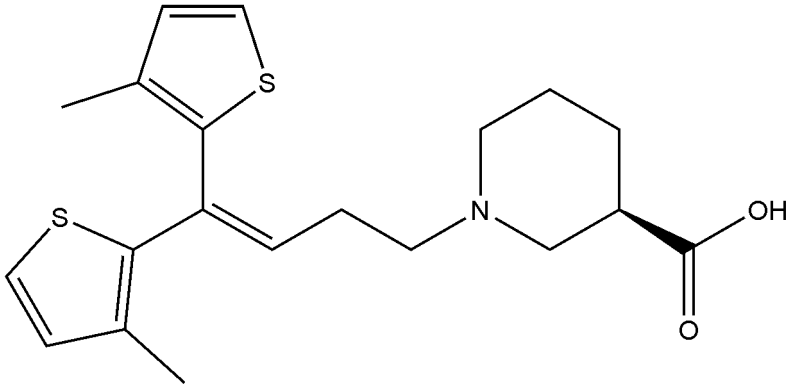
The transmembrane helices are numbered and the intracellular loops (IL) and extracellular loops (EL) are indicated, plus the N-terminus (NT) and C-terminus (CT). The three identified N-linked glycosylation sites are shown on the long EL2 as Y-shaped symbols.

Sequence homology suggests that GAT1 follows the structural arrangement of the bacterial homolog LeuT and other eukaryotic SLC6 transporters (Figure 1.14), with 12 TMHs and intracellular N- and C-termini as previously described previously. Prior to the elucidation of the LeuT structure, the key membrane topology of GAT1 had been suggested for many years through hydrophobicity analysis and N-linked glycosylation scanning mutagenesis. There is a long extracellular loop between TMH3 and TMH4, which is the location of three N-linked glycosylation sites. Removal of all three of these glycosylation sites on this extracellular loop is associated with a drastic reduction in transport activity (Bennett and Kanner, 1997). The recent cryo-EM structure of GAT1 in complex with tiagabine at 3.8 Å overall resolution confirmed the long postulated structural arrangement and provided transporter details of the substrate and ion binding pockets.

It has been suggested through multiple methods, including resonance energy transfer techniques (Schmid *et al.*, 2001; Moss *et al.*, 2009) and freeze-fracture EM (Gonzales *et al.*, 2007), that GAT1 forms oligomers of at least dimeric order. It has been found to be essential that GAT1 oligomerises in order to exit the ER (Schmid *et al.*, 2001; Moss *et al.*, 2009), but the protein appears to be able to function independently as a monomer (Soragna *et al.*, 2005). Although for many years it was postulated that a leucine heptad “zipper” of four repeated leucine residues on TMH2 formed the oligomerisation interface based on mutagenesis studies (Scholze *et al.*, 2002; Korkhov *et al.*, 2004), the LeuT structure eventually showed that this was not possible. The interface is now suggested to be formed with sections of TMH9 and TMH12 (Yamashita *et al.*, 2005).

The activity and trafficking of GAT1 is regulated by a complex interplay of tightly regulated protein-protein interactions and signalling events. Many of these are similar between SLC6 transporters, although some are exclusive to GAT1. On the N-terminus, the SNARE complex protein syntaxin 1A interacts to traffic GAT1 to the cell surface, although paradoxically it also downregulates GAT1 activity (Deken *et al.*, 2000; Wang *et al.*, 2003). Toward the C-terminus, SEC24D coat protein complex II-coated vesicles interact with an Arg-Leu motif which is conserved in SLC6 transporters. If this motif is truncated or altered in any fashion to prevent interaction with SEC24D then GAT1 does not exit the ER (Farhan *et al.*, 2007). An Ala-Tyr-Ile residue sequence on the C-terminus encodes a PDZ (postsynaptic density 95/discs large/zona occludens 1) binding motif. This has been identified to bind to the protein PALS1, which then tethers the protein to the cytoskeleton via ezrin; recombinant co-expression of GAT1 and PALS1 increases the uptake of GABA (McHugh *et al.*, 2004). Although this motif can be disrupted or removed without affecting the function of GAT1, surface expression is impaired and consequentially reduced *in-vitro* in neuronal cells and also *in-vivo.* HEK293 cells however do not show this deficient, even though they lack endogenous PALS1 (Imoukhuede *et al.*, 2009; Moss *et al.*, 2009).

1.4.3.2: GAT1 as a pharmacological target



Tiagabine

Figure 1.15: Chemical structure of tiagabine.

GAT1 is a target for control of epilepsy and neuropathic pain, although the inhibitor tiagabine (Figure 1.15) is the only therapeutic drug currently marketed which specifically targets GAT1. Tiagabine potently and selectively inhibits GAT1, leading to increased concentrations of GABA in the synapse and therefore amplified interaction with inhibitory post-synaptic GABA receptors. Tiagabine is seen as ineffective in most patients for preventing epileptic seizures when used as a monotherapy, but it is licensed for use as add-on treatment for partial epilepsy. Side effects limit patient tolerability of tiagabine and include central nervous system effects (dizziness, headache, decreased concentration, speech problems and confusion), sedation and most concerningly, non-status epilepticus. Therefore, it is only prescribed to a limited population of patients. Tiagabine is also prescribed for other indications such as anxiety disorders, although it is not licenced for these purposes and use is discouraged due to the rate of side effects (Uthman *et al.*, 1998; Novak *et al.*, 2001; Jette *et al.*, 2006; Bauer and Cooper-Mahkorn, 2008).

Thesis Aims

This thesis reports and presents on research carried out to investigate a central theme of structure and ligand binding at neurotransmitter channels and transporters. The first aims related to illuminating the binding modes and structure-activity relationships of novel ligands at nAChRs using both binding assays and X-ray crystallography with the surrogate protein AChBP. The ligands being characterised were of interest as potential smoking cessation agents and as research tools for studying nAChRs. The second aims were associated with attempting to establish methodologies for the efficient recombinant production of complex eukaryotic ion channels and neurotransmitter solute symporters. These proteins potentially represent targets for the development of new insecticides and the control of epilepsy.

Chapter 2: General Methods

2.1.0.0 Design and generation of expression constructs

Expression constructs were designed from protein sequences deposited in the Uniprot database (Bateman, 2019) and are detailed in Table 2.1. Where modifications were made to the base protein sequence, to remove domains for example, these were informed by both sequence alignments with similar proteins using Clustal Omega (Sievers *et al.*, 2011) and by generation of 3D homology models with the Phyre2 (Kelley *et al.*, 2015) or SwissModel (Waterhouse *et al.*, 2018) servers. Amino acid sequences in this thesis are numbered from Met1, the respective initiating methionine.

Constructs were either ordered as synthesised genes from Genscript (Piscataway, NJ, USA) or Twist BioScience (San Francisco, CA, USA) with codon optimisation for the relevant expression system or were constructed from existing sources by overlapping extension polymerase chain reaction (PCR). The appropriate restriction sites were designed in for sub-cloning into the desired plasmid vector.

Table 2.1: Details of expression constructs and outcomes

Protein	Modifications	Unipro t ID	Species	Tags	Expression system	Result
Secreted soluble proteins						
Acetylcholine binding protein (AChBP)		Q8WS F8	*Ac*	C- TEV-His_6	Insect cells (*Sf*9/High Five™)	Purified and used for crystallography and ITC.
Membrane Proteins						
Intracellular domain truncated Resistance to Dieldrin ($RDL_{ICD\ Trun}$)	Residues 357-565 replaced with SQPARAA linker	P25123	*Dm*	C- TEV-His_6	Insect cells (*Sf*9)	Present on membrane, but soluble protein could only be detected using harsh zwitterionic detergents.
GFPuv-$RDL_{ac(INS)}$	GFPuv fused between residues V44 and Q45 of the canonical protein sequence following assumed signal peptide.	P25123 -1	*Dm*	N-GFPuv-TEV, C-TwinStrep	Insect cells (*Sf*9)	Present on membrane, but only minimal solubilisation (<10%) except with harsh zwitterionic detergents.
mGFPuv-$RDL_{ac(MAM)}$	mGFPuv fused between residues G50 and G51 of the canonical protein sequence	P25123 -1	*Dm*	N-mGFPuv-TEV, C-TwinStrep	Mammalian cells (Adherent HEK293, suspension Expi293)	Present on cell surface. Successfully solubilised in non-ionic and zwitterionic detergents. Micro-scale amount purified after zwitterionic detergent extraction

GABA Transporter 1-mEmerald in pFastBac 1 vector (GAT1-$mEm_{(PFB)}$)	P30531-1	*Hs*	C-TEV-mEmerald-TwinStrep	Insect cells (*Sf*9)	Successfully solubilised and purified in detergent and SMA polymer. Aggregation issues following TEV cleavage, so only small amounts of GAT1 purified.
GABA Transporter 1-mEmerald in EarlyBac PFP10 vector (GAT1-$mEm_{(EB)}$)	P30531-1	*Hs*	C-TEV-mEmerald-TwinStrep	Insect cells (*Sf*9)	Successfully solubilised and purified in detergent and TEV cleaved to yield GAT1. Functional assays unsuccessfully attempted with a fluorescent membrane potential assay.
GABA Transporter 1-mEmerald in pTwist CMV WPRE Neo vector (GAT1-$mEm_{(HEK293)}$)	P30531-1	*Hs*	C-TEV-mEmerald-TwinStrep	Mammalian cells (adherent HEK293)	Successfully solubilised in detergent. Functionality demonstrated with a fluorescent membrane potential assay.

Ac, Aplysia californica; Dm, Drosophila melanogaster; GFP, Green fluorescent protein; *Hs, Homo sapiens;* mEm, mEmerald*; Sf, Spodoptera frugiperda;* SMA, styrene and maleic acid*;* TEV, Tobacco etch virus protease.

2.1.1.0: Plasmid growth and purification

Plasmids were routinely grown up by transforming into XL1-blue super-competent *E. coli* cells (Agilent Technologies, Santa Clara, CA, USA). The plasmid (~0.1-1 ng) was incubated on ice with 20 µl of cells for at least 20 minutes, before the mixture was heat shocked at 42 °C for 45 seconds, placed on ice for 5 minutes and 0.2 ml Super Optimal broth with Catabolite media (SOC:

2% [w/v] tryptone, 0.5% [w/v] yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose – in-house media kitchen, School of Life Sciences, University of Dundee, Dundee, UK) added. The cells were incubated in a 37 °C water bath for 1 hr for recovery and antibiotic resistance outgrowth and then all plated on Lysogeny Broth (LB: 1% [w/v] Tryptone, 0.5% [w/v] tryptone, 10 g/L NaCl) agar plates (in-house media kitchen) plus the relevant selection antibiotic for the plasmid (50 µg/ml kanamycin or carbenicillin). The plates were incubated at 37 °C overnight then single colonies picked to inoculate 5 ml liquid cultures of LB in which the selection antibiotic was present. These cultures were incubated for 12-16 hrs at 37 °C with shaking at 200 RPM and the plasmid DNA subsequently extracted and purified using a Qiagen (Venlo, Netherlands) QIAprep spin miniprep kit as per the manufacturers instructions.

2.1.2.0: Restriction digest cloning

Thermo-Fisher Scientific (Waltham, MA, USA) fast-digest restriction enzymes were used for by restriction digest analysis and sub-cloning of plasmids. In a 20 µl restriction-digest reaction, 1 µl of each restriction enzyme was added per 1 µg of plasmid, plus 2 µl of 10x fast-digest green buffer and the remainder double distilled water (ddH_2O). Controls consisted of no enzyme and single enzyme digests. The digests were incubated at 37 °C for 10 minutes, then loaded on a 0.75-1% agarose gel in 1x Tris-acetate-EDTA (TAE) buffer plus 0.2 µg/ml ethidium bromide. New England Biolabs (Ipswich, MA, USA) 100 bp and/or 1 kb DNA MW markers were also loaded in at least one lane. The running buffer consisted of 1x TAE plus 0.2 µg/ml ethidium bromide. The gel was run at 4-5 V/cm until the tracking dye ran off the bottom and the gel photographed using the UV tray of a GelDoc EZ (Bio-Rad Laboratories, Hercules, CA, USA) The relevant bands for the products of interest were cut out on a UV transilluminator and purified using a Qiagen QIAquick gel extraction kit as per the recommended protocol. The purified vector and insert were ligated using Clonables 2x ligation pre-mix (Merck KGaA, Darmstadt, Germany) or Anza T4 4x ligase mix (Thermo-Fisher Scientific) and the respective reaction buffer on a 1:3 vector:insert molar ratio (calculated using http://www.insilico.uni-duesseldorf.de/Lig_Input.html) and 1 µl transformed into NovaBlue Singles (Merck KGaA, Darmstadt, Germany) or XL-10 Gold (Agilent Technologies) ultra-competent cells. In the case of XL10 gold

cells the heat shock stage was shortened to 30 seconds as recommended by the manufacturer. After setting up liquid cultures from a single colony and purifying the plasmid DNA after overnight incubation at 37 °C, the presence of the insert was confirmed by Sanger sequencing using an in-house facility (www.dnaseq.co.uk).

2.1.3.0: Assembly of mGFPuv-RDL$_{(MAM)}$

mGFPuv-RDL$_{(MAM)}$ was assembled from several existing plasmids using overlapping extension PCR as depicted in Figure 2.1. A Kozak sequence was added before the signal peptide to ensure a strong initiation of protein translation (Kozak, 1987) and because of uncertainties regarding the signal peptide cleavage site, the GFPuv was moved to be between residues 50 and 51 (the 6th and 7th residues following the predicted cleavage site) of RDL based on sequence alignments with other GABA$_A$ related subunits. The GFPuv also had a substitution subsequently inserted to ensure exclusive formation of monomers (see section 2.1.4.0). XbaI and EcoRV restriction sites were added to the 5' and 3' of the gene respectively.

Phusion Green Hot Start II 2x PCR master-mix (Thermo-Fisher Scientific) was used for amplification, with each 50 µl reaction containing 10 µl of 2x master-mix, 0.5 µm each of forward and reverse primers and ~1 pg–10 ng of each template, with the remainder made up in ddH_2O. The thermocycler conditions were 30 seconds of initial denaturation at 98 °C, followed by 30 cycles of denaturation at 98 °C for 10 seconds, 30 seconds of annealing at the optimal temperature for the primers (calculated via www.thermofisher.com/tmcalculator) and 30 seconds/kb of plasmid length of extension at 72 °C. There was then a final extension for 10 minutes at 72 °C, before the reactions were held at 4 °C and subsequently gel purified to isolate the products of interest.

Primer 1: 5' TCTAGAGCCGCCATGAGCGACTCCAAAATGG 3'

Primer 2: 5' GCCACCGGTAGCGGCCTGCACAGTGTGGATGCGTTTATGG 3'

Primer 3: 5' AGTTCCTCTCCCTTGGACATGCCACCGGTAGCGGCCTGCA 3'

Primer 4: 5’ TGCAGGCCGCTACCGGTGGCATGTCCAAGGGAGAGGAACT 3’

Primer 5: 5’ ACGTCACCCAGCATGCTTCCAGATTGGAAGTACAGGTTTT 3’

Primer 6: 5’ AAAACCTGTACTTCCAATCTGGAAGCATGCTGGGTGACGT 3’

Primer 7: 5’ GATATCTTATTACTTCTCGAACTGGGGATGA 3’

Primer 8: 5’ TCTAGAGCCGCCATGAG 3’

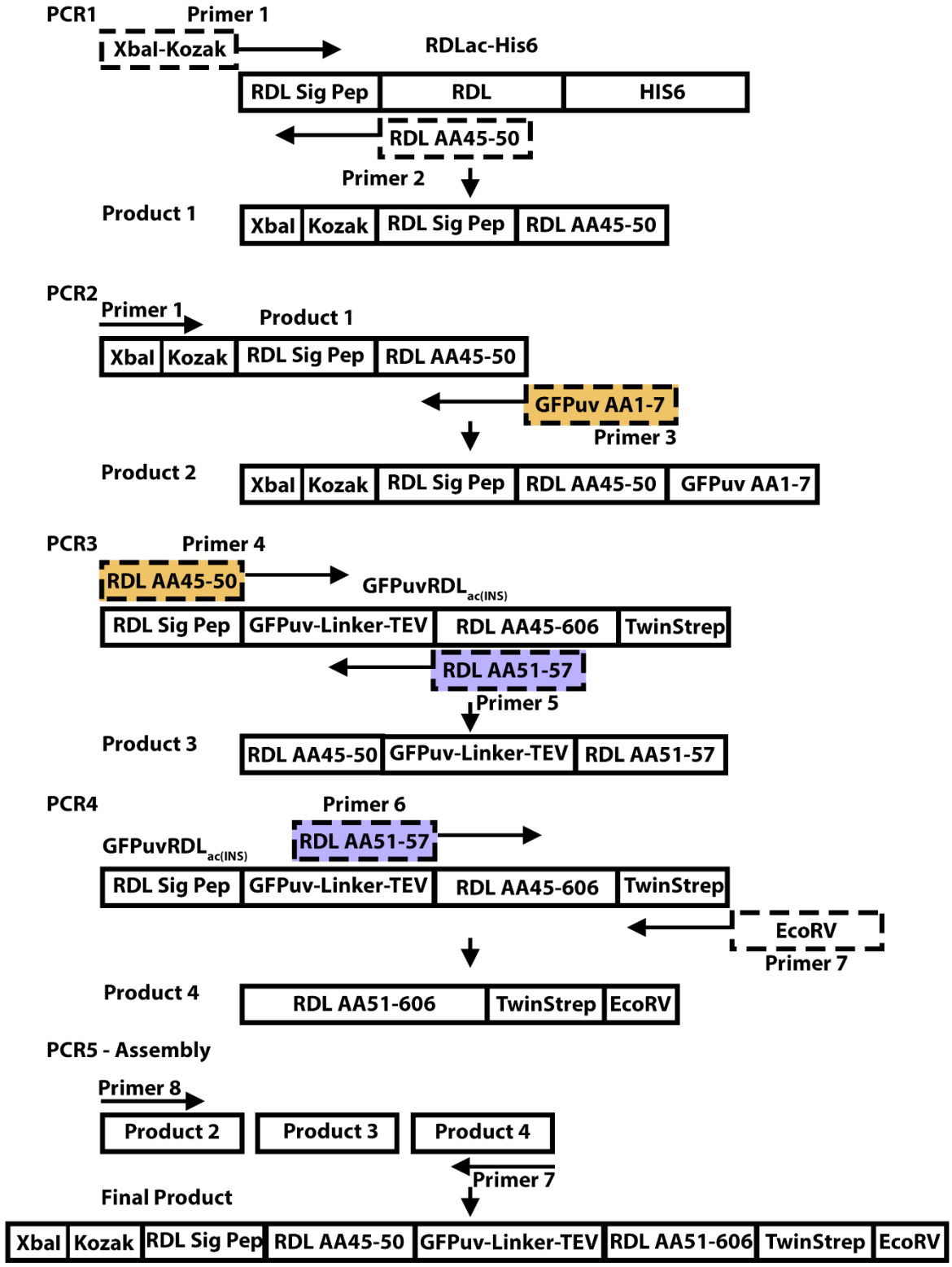


Figure 2.1: Overlapping extension PCR cloning of mGFPuv-RDL$_{(MAM)}$.

Arrows represent primers and any overhangs generated are shown attached in dash-lined boxes, which are colour co-ordinated for each over-lapping pair. The RDL$_{ac}$-His6 template was received from Dr. Alice Dawson.

The final PCR assembled the fragments and amplified the end product, which after gel purification was topoisomerase (TOPO) cloned into the pBlunt2 TOPO cloning vector (Thermo-Fisher Scientific) as per the manufacturers instructions. The sequence of the insert between the plasmids M13 forward and reverse motifs was read by Sanger sequencing to check for PCR induced errors, with Primer 6 also used as an internal primer to ensure coverage of the full sequence. The insert was subsequently cut out between the XBaI and EcoRV restriction sites, gel purified and ligated into pcDNA3.4 TOPO expression vector digested with the respective enzymes.

2.1.4.0: Introduction of A206K to mGFPuv-RDL$_{(MAM)}$ by site directed mutagenesis

Site directed mutagenesis used an adaption of the “QuikChange” method (Papworth *et al.*, 1996). Mutagenic primers were designed using the Quikchange primer design tool (https://www.chem.agilent.com/store/primerDesignProgram.jsp). 50 µl reactions were setup containing 5, 10, 20 or 50 ng of template, plus 1 µl of PfuTurbo polymerase (2.5 U/µl) (Agilent Technologies), 125 ng each of forward and reverse mutagenic primers, 5 µl PfuTurbo buffer, 200 µM dNTPs and the remainder ddH_2O. The thermocycler conditions were 30 seconds of initial denaturation at 95 °C, followed by 16 cycles of denaturation at 95 °C for 30 seconds, 60 seconds of annealing at 55 °C and 60 seconds/kb of plasmid length of extension at 72 °C. The reactions were then held at 4C, before 1 µl of DpnI enzyme (New England Biolabs) was added, the mixture briefly centrifuged and then incubated at 37 °C for 1 hour to digest un-mutated supercoiled double stranded DNA. 1 µl of the DNA was subsequently transformed into XL10 Gold ultra-competent cells (Agilent Technologies) and plated onto LB agar plates containing the appropriate selection antibiotic for the plasmid, with liquid cultures subsequently setup and the DNA purified as previously described. Sanger sequencing was used to check for the presence of the mutation and any PCR induced errors.

To prevent the linked GFPuv from forming dimers and therefore potentially inducing mis-leading results, after assembly of the final product for mGFPuv-RDL$_{(MAM)}$ (see section 2.1.3.0), the equivalent numbered residue to Ala206 in the

GFPuv was substituted to Lys using the following mutagenic primers. This substitution favours almost exclusive formation of monomers by sterically blocking dimerisation (Zacharias *et al.*, 2002).

Forward: 5’ CTACCTGTCAACTCAGTCCAAGCTGAGCAAGGACCCAAACG 3’

Reverse: 5’ CGTTTGGGTCCTTGCTCAGCTTGGACTGAGTTGACAGGTAG 3’

2.2.0.0: Protein production in eukaryotic systems

All recombinant proteins in this thesis were produced using insect or mammalian cells. All eukaryotic cell culture was carried out in a class II biological safety cabinet (BSC) under sterile conditions. Suspension cell density/viability was measured using a Cellometer Auto 100 automated cell counter (Nexcelom Bioscience LLC, Lawrence, MA, USA) with trypan blue screening.

2.2.1.0: Insect cell expression

Spodoptera frugiperda (*Sf*) 9 and High Five™ insect cell lines were maintained in polycarbonate non-baffled Erlenmeyer shaker flasks in a 27 °C atmosphere with ambient CO_2 and shaking at 135 RPM. *Sf*9 cells were grown in *Sf*900 II or *Sf*900III (Thermo-Fisher Scientific) serum free mediums supplemented with 2 mM L-glutamine and 100 U ml^{-1} penicillin/streptomycin, while High Five™ cells were grown in Insect XPRESS (Lonza Group AG, Basel, Switzerland) or Express Five (Thermo-Fisher Scientific) serum free mediums supplemented with 2 mM L-glutamine and 100 U ml^{-1} penicillin/streptomycin. Foetal bovine serum (FBS) was not added to the medium for routine culture. Cultures were passaged every 2-3 days to a density of 0.8 to 2.0x10^6 cells/ml, to ensure they were in the log phase when split.

2.2.1.1: Bacmid DNA isolation

The Bac-to-Bac system (Thermo-Fisher Scientific) was used for production of proteins in insect cells. Genes were sub-cloned into the pFastBac 1 or EarlyBac PFP10 vector using previously detailed methods and the plasmids transformed into DH10Bac cells (Thermo-Fisher Scientific) to produce recombinant bacmid DNA. The transformed cells were streaked on insect cell agar plates (1% Peptone 140, 0.5% yeast extract, 10 g/L NaCl, 50 µg/ml kanamycin, 7 µg/ml gentamicin,

10 µg/ml tetracycline, 40 µg/ml IPTG, 100 µg/ml X-Gal – in-house media kitchen) and blue/white screening used to determine positive clones after 48 hours at 37 °C. Single white colonies were re-streaked on fresh insect cell agar plates with a sterile loop and checked after 48 hours to confirm the presence of the plasmid.

Bacmid DNA was isolated using either commercial anion-exchange resin maxi-prep column kits or isopropanol precipitation. For maxi-preps 500 ml of insect cell media was inoculated with a single positive clone and after 12-16 hrs at 37 °C with shaking was spun down and the pellet submitted to an in-house facility (www.dnaseq.co.uk) for purification.

For isopropanol precipitation, 5 ml of insect cell media was inoculated with a single white colony and grown for 12-16 hrs at 37 °C with shaking. A 1.5 ml sample of the culture was transferred to a microfuge tube, the cells collected at 14,000 *g* and the supernatant removed. The pellet was re-suspended in 0.3 ml ice cold Qiagen buffer P1 (obtained from a Qiagen miniprep kit), then lysed by addition of Qiagen buffer P2 and left for 5 minutes at room temperature. 0.3 ml of 3 M potassium acetate was gently mixed in and the solution left on ice for 5 minutes, before centrifuging at 14,000 *g* for 10 minutes to pellet the formed precipitate. The supernatant was transferred to a new microfuge tube and 0.8 ml ice-cold isopropanol added. After 15 minutes on ice, the bacmid DNA was pelleted at 14,000 *g* for 10 minutes, the supernatant removed and the pellet washed with ice-cold 70% ethanol, then left to air dry. The dried bacmid DNA was re-suspended in 40 µl Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0). Bacmid DNA was stored at 4 °C prior to transfection and frozen at -20 °C for longer-term storage. The number of freeze-thaw cycles was minimised to avoid shearing.

2.2.1.2: Baculovirus production

Bacmid DNA was transfected into *Sf*9 cells using a liposome-based reagent to generate the initial baculovirus. An hour prior to transfection, *Sf*9 cells were diluted to 0.8×10^6 cells/ml in antibiotic free media and 1 ml of cells seeded per sample into a 6 well tissue culture plate. Per sample, 500 ng of bacmid DNA was diluted into 1 ml of antibiotic free media plus 5 µl of Insect Gene Juice (Merck Millipore) and the mixture left to complex at room temperature for 30 minutes. Negative and positive controls were setup for each set of transfections,

respectively comprising of transfection reagent only and bacmid DNA previously known to work successfully. The media was removed from the cells and the mixtures gently added. The 6 well plate was placed in a box containing damp paper towels to maintain an appropriate level of humidity and prevent evaporation, then stored at 27 °C. After 24 hours, 1 ml of additional antibiotic containing media was added to each well.

The wells were monitored each day for signs of successful transfection - e.g rounding, increased cell size, sloughing from the culture surface – and compared to the positive/negative controls. For fluorescently tagged constructs the cells were additionally monitored for fluorescence using a ZOE™ fluorescent cell imager (Bio-Rad). The supernatant – containing the P0 baculovirus – was harvested after 7 days and stored at 4 °C, with a backup sample typically snap frozen and stored at -80 °C. The viral titre was amplified by infecting 50 ml of *Sf*9 suspension cells at $1.5x10^6$ cells/ml with 125 µl of P0 virus and incubating for 3-5 days. The cells were collected by centrifugation (1000 *g*, 10 minutes, 4 °C) and the supernatant containing the P1 virus retained. This was passed through a 0.2 µm filter, FBS added to 5% and then stored at 4 °C, with samples also snap frozen in liquid N_2 and stored at -80 °C as a long-term backup.

When larger volumes of virus were required, the titre was re-amplified by infecting 200 ml of *Sf*9 cells at $1.5x10^6$ cells/ml with 500 µl of P1 virus and incubating for 3 days at 27 °C before collecting the P2 virus containing supernatant and storing at 4 °C.

2.2.1.3: Production of secreted AChBP in insect cells

*Sf*9 or High Five™ cells at $1.5x10^6$ cells/ml were infected with AChBP baculovirus at 5% and incubated at 27 °C in shaking flasks for 48 hours, with the protein being secreted into the culture media. The cells were pelleted by centrifugation (1500 *g*, 10 minutes, 12°C followed by 4000 *g*, 10 minutes, 12°C), with protein purified from the media as described subsequently in section 2.4.1.0.

2.2.1.4: Production of membrane proteins in insect cells

*Sf*9 cells were used for production of membrane proteins in insect cells. Cells were sub-cultured to 1.5 – $3.0x10^6$ cells/ml, the baculovirus added to 1 % and for membrane proteins FBS to 0.5% to reduce proteolytic degradation as detailed in

the Bac-to-Bac product literature. Except where noted, cultures were incubated for 24 hours at 27 °C before being shifted to 18 °C for a further 48 hours and harvesting. For some membrane proteins, optimal temperature conditions and harvest time were determined using procedures detailed in section 2.3.1.0. For this, typically two cultures for each baculovirus were setup, with both infected at 1% and one shifted to 18 °C after 24 hours. Small cell samples were taken and frozen every 24 hours after infection until ~72 hours.

2.2.2.0: Mammalian cell expression

Genes encoding proteins destined for mammalian cell expression were ordered in the pTwist CMV WPRE Neo expression vector or sub-cloned into the pcDNA3.4 TOPO vector. Large quantities of pure plasmid DNA were obtained by midi- or maxi-prep.

2.2.2.1: Adherent HEK293 cells

HEK293 cells were cultured in an adherent format in T-25 or T-75 flasks in Dulbeccos modified Eagles medium (DMEM) supplemented with 10% FBS and penicillin/streptomycin (all from Thermo-Fisher Scientific). The cells were incubated in a 95% humidified 37 °C atmosphere plus 5% CO_2. The media was changed every 2-3 days and the cells sub-cultured when at 70-90% confluence. After washing with Ca^{2+} and Mg^{2+} free Dulbeccos phosphate buffered saline (DPBS – Thermo-Fisher Scientific), the cells were detached by incubating with Tryspin-EDTA solution (Thermo-Fisher Scientific) in the 37 °C incubator for 5 minutes. The trypsin was inactivated by addition of serum containing media and the cells spun down at 1000 g for 5 minutes, then resuspended in fresh medium. For routine sub-culturing, cells were generally split between 1:5-1:15 in a fresh flask. For cells destined for transfection however, density was read using the Countess II FL cell counter (Thermo-Fisher Scientific).

Transfection of adherent HEK293 cells with Lipofectamine 2000

For routine transfection of HEK293 cells, cells from sub-culturing were seeded in T-25 or T-75 flasks at a density of ~66-80x10^3 cells/cm^2 and grown overnight. Once the confluence had reached 70-90% the cells were transfected with the desired plasmid using Lipofectamine 2000 (Thermo-Fisher Scientific). When

growing cells on 13 mm glass coverslips in a 24 well plate however, 100 µl of cells at ~0.8x10^6 cells/ml cells were plated per well and left to adhere for 1 hour in the 37 °C incubator. 0.5 ml of media was subsequently added and the cells were transfected the next day when ~50% confluent.

0.75, 10 and 30 µg of plasmid DNA was respectively used for transfection of 13 mm coverslips, T-25 and T-75 flasks. Lipofectamine 2000 was used on a 1 µg:2 µg DNA:lipofectamine ratio. Plasmid DNA and lipofectamine-2000 were separately diluted in equal volumes of OptiMem media (Thermo-Fisher Scientific) and left for 5 minutes to mix. The plasmid DNA was then added to the lipofectamine-2000, gently mixed and left to incubate in the BSC for 20 minutes, before being slowly added to the cells in a drop-wise fashion. The medium was not changed for transfection. The media was changed after 24 hours and the cells harvested between 24 and 48 hours post transfection.

2.2.2.2: Suspension Expi293 cells

Expi293 cells (Thermo-Fisher Scientific), a commercial derivative of HEK293S suspension cells adapted for high density culture, were cultured in Expi293 serum free media (Thermo-Fisher Scientific) supplemented with 1% FBS to reduce proteolysis and act as a lipid source. Cultures were maintained in polycarbonate non-baffled Erlenmeyer shaker flasks incubated in a 37 °C atmosphere plus 8% CO^2, with shaking at 125 RPM. The cultures were sub-cultured to 0.3-0.4x10^6 cells/ml when the density reached ≥3.0x10^6 cells/ml, generally every 2-3 days.

Transfection of expi293 cells with linear polyethylenimine

The day prior to transfection, expi293 cells were sub-cultured to 1.5-2.0x10^6 cells/ml so that the density would be approximately 3.0-4.0x10^6 cells/ml on the subsequent day. 40K linear polyethylenimine (PEI - obtained as Transporter 5 pre-prepared transfection reagent, Polysciences, Warrington, PA, USA) was used for transfection, with 1 µg:2.5 µg and 1 µg:4 µg DNA:PEI ratios initially tested; 1 µg of DNA was transfected per 1 ml of culture. Plasmid DNA and PEI were separately diluted in warm sterile PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4 – in-house media kitchen) to a total volume 1/20th of the suspension culture size, combined together, briefly vortexed and left to complex at room temperature in the BSC for 20 minutes. The DNA:PEI

complexes were gently mixed and slowly added to the cells, which were returned to the incubator. After 12 hours sodium butyrate (Sigma-Aldrich, 1M dissolved in PBS and filter sterilised) was added to 10 mM to boost protein expression. Optimal harvest time was determined using the procedures in section 2.3.1.0

2.3.0.0: Processing of membrane protein expressing cell pellets

All procedures in this section were performed on ice or, preferentially, in a 4 °C cold room.

2.3.1.0: Assessment of membrane protein production efficiency

Small samples – approximately 10^7 cells - of membrane protein expressing cells were taken at 24-hour intervals to determine the optimal culture transfection/infection parameters, temperature and harvest time for each protein. The samples were centrifuged at 1500 *g* and the cell pellets gently re-suspended in ice cold PBS before re-centrifuging to collect the washed cells, which were frozen at -20 °C until needed. The pellets were thawed on ice and re-suspended at approximately 10^7 cells/ml in the respective buffer for the protein of interest, plus protease inhibitors (cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail, Roche Diagnostics, Basel, Switzerland or Protease Inhibitor Cocktail EDTA Free (ab270055), Abcam, Cambridge, UK) and DNAse I or benzonase. n-Dodecyl-b-D-Maltopyranoside (DDM – Anatrace Products, Maumee, OH, USA) detergent was added to 2% final concentration and the cells left to lyse at 4°C for 2 hours with gentle end-over-end rotation. Lysates were obtained by retaining the supernatant after centrifugation at 16,000 *g* for 30 minutes and their concentrations were quantified by Rapid Gold bicinchoninic acid assay (BCA – Thermo-Fisher Scientific) in a microplate format.

Equal amounts of each protein sample (between 10 – 30 µg, depending on the protein) were subsequently loaded on an SDS-PAGE gel and run as described in section 2.5.1.0. The level of protein in the samples was assessed by in-gel fluorescence of an attached fluorophore and/or by western blot or by reading the total fluorescence in a microplate format.

2.3.2.0: Sub-cellular fractionation of membrane protein producing cells by differential centrifugation

Membrane protein expressing cells were harvested by centrifugation (1500 *g*, 4 °C) and the pellet gently re-suspended in ice cold PBS before re-centrifuging to collect the washed cells. At approximately 10^8 cells/ml, the pellet was re-suspended in the respective buffer for the protein of interest, plus protease inhibitors and DNAse I or benzonase and gently passed through 19 G and 25 G needles to break up clumps. For volumes of ≤3 ml, the cells were lysed by passing through a 27 G syringe 10 times, while samples of 3 to 20 ml were lysed with 10-15 strokes of a Dounce homogeniser with a tight-fitting pestle, followed by passing through a 27 G syringe 2-3 times. Samples greater than 20 ml were passed through a CF1 Cell Disruptor (Constant Systems, Daventry, UK) at 30,000 PSI with cooling. The lysed cells were centrifuged at 16,000g for 30 minutes to pellet un-lysed cells, nuclei, nucleic acids, mitochondria and peroxisomes. A sample of the supernatant (lysate) was retained for later analysis and the remainder ultra-centrifuged at 200,000 *g* for 1.5 hours to pellet the plasma membranes.

A sample of the obtained supernatant - containing the crude cytosolic fraction - was subsequently re-ultracentrifuged at 300,000 *g* for 2 hours to pellet the ribosomes and any remaining contamination from the previous ultracentrifugation, leaving the isolated cytosolic fraction.

The membrane pellet was re-suspended in buffer plus protease inhibitors at initially approximately 1 ml per 100 ml of initial culture and homogenised into solution with 10-20 strokes of a chilled Dounce homogeniser on ice, followed by passing through 25 g and 27 g needles. Smaller sample volumes (<3 ml) were homogenised by passing through a 21 g needle followed by 25 and 27 g needles. The total protein concentration of the membranes was quantified using a Rapid Gold BCA kit in a micro-plate format and adjusted to 3-10 mg/ml. The membranes were either used immediately or snap frozen in liquid N_2 and stored at -80 °C.

Membranes were adjusted to a final concentration of 1.0-2.0 mg/ml and a detergent added to a concentration of 1.0 to 5.0%, ensuring to be at least 10x the critical micellar concentration (CMC) and at a detergent:protein ratio of at least

5:1 (w/w), but preferentially at >10:1. The membranes were left to solubilise in the detergent for 2-3 hours at 4°C with gentle end over end rotation, before being ultra-centrifuged at 200,000 *g* for 45 minutes to remove insoluble material.

2.3.3.0: Screening for membrane protein detergent solubility

For membrane proteins it is critical to find the most appropriate detergent which efficiently extracts the protein from the plasma membrane, but also maintains the function and desired oligomeric state. As standard, proteins were screened against a selection of the following detergents in Table 2.2, taking into account the sample properties and previously published data. In most cases cholesteryl hemisucinate (CHS) as the tris salt was added to approximately 1/5th-10th the detergent concentration. Detergents were obtained from Sigma-Aldrich or Anatrace Inc. (Maumee, OH, USA).

Table 2.2: Detergents used for screening

Detergent	Type	Concentration (%)
Digitonin	Non-ionic	1.0
n-Dodecyl-β-D-Maltoside (DDM)	Non-ionic	1.0-2.0
n-Decyl-β-D-Maltoside (DM)	Non-ionic	1.5
n-Undecyl-β-D-Maltoside (UDM)	Non-ionic	1.5
C12E9	Non-ionic	1.5
Cymal-5	Non-ionic	1.5
Triton-X-100	Non-ionic	1.5
Decyl Maltose Neopentyl Glycol (DMNG)	Non-ionic	1.5
n-Octyl-β-D-Glucopyranoside (OG)	Non-ionic	5.0
Lauryldimethylamine oxide (LDAO)	Zwitterionic	1.5
Fos-choline-13	Zwitterionic	1.0-1.5
Fos-choline-14	Zwitterionic	1.5

Screening was performed by solubilising either extracted membranes or whole cells. These were respectively suspended to a final concentration of ~1.0-2.0 mg/ml and approximately 10^7 cells/ml in buffer plus protease inhibitors, DNAse I or benzonase and the detergent of interest. The samples were left to solubilise for 2-3 hrs with gentle rotation at 4 °C. Insoluble material was collected by ultra-centrifugation at 200,000 *g* for 45 minutes.

After obtaining the soluble material, the detergent solubilisation efficiency was determined by SDS-PAGE with Western blotting and/or in-gel fluorescence by loading equal amounts of protein as determined by BCA assay. Fluorescently tagged proteins were also analysed by comparing the total fluorescence with an equal volume sample of un-solubilised cells or membranes in a microplate format. In the case of whole cells, these were mechanically lysed by needle shearing. To correct for differing sample concentrations, the fluorescence values were divided by the sample protein concentration in µg/ml to yield the fluorescence:[protein] ratio. Fluorescent size exclusion chromatography (FSEC – see section 2.5.6.0) was used to determine the effect of the detergents on the oligomeric state and dispersity of the protein without having to purify the mixture.

2.4.0.0: Protein purification

2.4.1.0: Purification of secreted AChBP

For purification of AChBP secreted into culture media, the Sartojet system with a 10 kDa cut-off Sartocon Slice filter (Sartorius) was first used to exchange the media for Buffer A (50 mM Tris-HCl, 250 mM NaCl pH 7.5) and concentrate the protein solution. The solution was applied to a charged 5 ml Ni^{2+} HisTrap column (Cytivia Life Sciences, Marlborough, MA, USA) equilibrated in Buffer A (50 mM Tris-HCl, 250 mM NaCl pH 7.5) at 5 ml/min for immobilized metal anion chromatography. The column was washed with 10-15 column volumes of buffer A plus 7.5% 800 mM imidazole The product was eluted over 10-30 column volumes using a stepped and/or linear gradient of buffer A plus 800 mM imidazole. Fractions were analysed by stain free SDS-PAGE gels (Bio-Rad) and those containing the desired protein pooled then exchanged into buffer A and concentrated using 10 kDa cut-off centrifugal concentrators (Pall).

2.4.2.0: Co^{2+} resin purification of His6 tagged membrane proteins

1 ml of 50% Co^{2+} resin suspension (obtained from in-house reagents service) per litre of initial culture was washed with buffer to remove the storage buffer, then added to the detergent soluble sample. The resin was left to bind for at least 1 hr with gentle end-over-end rotation or stirring before being collected in disposable gravity flow columns (Bio-Rad). The resin was washed with 5-10 bed volumes of

buffer and the protein subsequently eluted using 10 bed volumes of buffer plus 250 mM imidazole. Samples of the flow through and washes were retained for analysis.

2.4.2.0: Strep-tactin resin purification of twin-strep tagged membrane proteins

Twin-strep tagged membrane proteins were purified using either Strep-tactin XT resin suspension (IBA Lifesciences GmbH, Göttingen, Germany) in a batch format or a 1 ml Strep-trap column (GE) connected to an Äkta purification system. All purification steps were carried out at 4 °C. For Strep-trap column purification, the detergent solubilised sample was loaded onto the column at 0.3 ml/min and the column washed with 6 CVs of buffer before the protein was eluted off the column over 6 CVs using buffer plus 7.5 mM desthiobiotin, taking 1 ml fractions and monitoring the absorbance at 280 nm (A280). The column was subsequently washed and re-generated with 500 mM NaOH and stored in 20% ethanol. Samples of the flow through, wash and elution fractions were analysed by SDS-PAGE with Western blotting to determine which contained the protein of interest.

For batch purifications, 2 ml of resin suspension per litre of initial culture was washed with 3 bed volumes of buffer to remove the storage buffer, then added to the detergent soluble sample. The resin was left to bind for at least 1 hr with gentle end-over-end rotation or stirring. Small volumes (<2 ml) were passed through spin column filters by either centrifugation (800 *g*) or gravity flow, while larger volumes were collected in 10 ml or 25 ml disposable gravity flow columns (Bio-Rad). The resin was washed with 5-10 bed volumes of buffer and the protein subsequently eluted using at least one addition of 3 bed volumes of buffer plus 50 mM *d*-biotin. Samples of the flow through and washes were retained for analysis. The resin was subsequently washed with buffer, re-generated with 10 bed volumes of 10 mM NaOH or 3 M MgCl, re-washed and stored in 20% ethanol at 4 °C.

2.4.3.0: Size-exclusion chromatography

Commercial 15 and 30 cm columns pre-packed with Superdex 200 or Superose 6 resin were used for size-exclusion chromatography (SEC) to analyse protein

purity and oligomeric state, as well as for polishing after the first purification step if needed. The column was attached to an Äkta automated purification system and equilibrated in the appropriate buffer. Protein samples were loaded onto the column with a loop appropriate for the size of the column and sample volume. For preparative SEC the sample volume was less than 50% of the loop volume to maximise recovery, while for analytical SEC the sample volume was preferentially at least twice that of the loop to maximise resolution. To minimise sample loss with smaller (< 500 µl) samples, a 250 µl gas-tight Hamilton syringe (Hamilton Bonaduz AG, Bonaduz, GR, Switzerland) was used for loading. The proteins were then eluted in 0.2 - 2 mL fractions over 1.2 column volumes with the A280 recorded. The columns were calibrated by passing Bio-rad protein molecular weight standards (Thyroglobulin, 670 kDa; γ-globulin, 158 kDa; Ovalbumin, 44 kDa; Myoglobin, 17 kDa; vitamin B12, 1.35 kDa) through, determining the retention volumes from the A280 peaks and plotting a standard curve.

2.5.0.0: Detection and analysis of proteins

2.5.1.0: SDS-PAGE

For separation of proteins by denaturing PAGE and to determine purity, the Bio-Rad Mini-Protean Tetra system was used with Laemmli like precast Bio-Rad mini TGX 10% or AnyKd™ gradient gels. These were typically of the stain-free type which contain a tri-halo compound that fluoresces under UV light. The running buffer (25 mM Tris-HCl, 19 mM glycine, 0.1% SDS ~pH 8.3) was freshly prepared each time from a 10x concentrated stock. Samples were loaded on the gel using thin gel-loading type pipette tips, with at least one lane containing Bio-Rad Precision Plus un-stained or all-blue MW standards as appropriate for the gels downstream applications.

For secreted soluble proteins, equal volumes of sample were prepared in 2x SDS loading buffer (125 mM Tris-HCl, 4% SDS, 20% glycerol, 0.004% bromophenol blue, 20 mM DTT pH 6.8) and heated for 5 minutes at 95 °C to denature before being loaded. The gel was then run at 300 V (44.8 V/cm) for approximately 17 minutes.

For membrane proteins, samples were prepared in 2X SDS loading buffer or 4x NuPage LDS loading buffer (424 mM Tris-HCl, 564 mM Tris base, 8% LDS, 40% glycerol, 2.4 mM EDTA, 0.88 mM SERVA Blue G250, 0.7 phenol red pH 8.5 – Thermo-Fisher Scientific) plus protease inhibitors and incubated at 37 °C for 30 minutes to 1 hour to denature, with the lower temperature preventing aggregation and destruction of the fluorescent properties of any attached fluorophores. Equal amounts of sample (between 10-30 µg as determined by BCA assay of the input protein) were loaded on the gel to allow quantitative comparisons. To limit gel heating, the running buffer was chilled and the gels run at 200 V (29.9 V/cm) for approximately 45 minutes. For some experiments where higher MW samples needed to be separated, 3-8% NuPage Tris-Acetate gels (Thermo-Fisher Scientific) were used. These were run for ~1 hour at 150 V (18.8 V/cm) with chilled tris-acetate-SDS running buffer (50 mM Tricine, 50 mM Tris, 0.1% SDS pH 8.24).

After electrophoresis, stain free gels were imaged on the stain free tray of a Bio-Rad Gel-Doc EZ or Chemi-Doc MP, prior to any subsequent techniques. Non-stain free gels were blotted or stained with InstantBlue® coomassie protein stain (Abcam) overnight at room temperature on a rocking shaker and, after rinsing excess stain away, imaged on the Gel-Doc EZ or Chemi-Doc MP white tray. Alternatively, the Typhoon 5 plus the digitisation plate and a 570 nm laser or the fluorescence tray with the 685 nm laser and the short infra-red filter set were used. For higher sensitivity, Zn^{2+}-imidazole negative staining or silver staining were utilised if necessary.

2.5.2.0: Zn^{2+}-imidazole staining

Zn^{2+}-imidazole staining is a negative staining technique that stains the gel rather than the protein bands, which become visible upon viewing over a dark background. Sensitivity is comparable to silver staining (~1-10 ng), but has the advantages of being reversible, mass spectrometry compatible and applicable to previously Coomassie brilliant blue stained gels as a counter-stain. The staining was implemented as described in Castellanos-Serra and Hardy (2006). Briefly, after electrophoresis SDS-PAGE gels were washed twice in water, before incubation in 0.2 M imidazole plus 0.1% SDS on a rocking platform for 15

minutes. SDS-PAGE gels that had been previously Coomassie stained however were extensively washed in water to remove residual acetic acid, then incubated in 0.2 M imidazole plus 0.3% SDS for 15 minutes with rocking. The gels were briefly washed again in water, before addition of 0.2 M $ZnSO_4$ and being gently shaken manually for ~60 seconds over a dark background to allow observation of the staining progress. Once satisfactory, the stain development was stopped by washing with water.

Stained gels were stored in water and imaged on the Gel-Doc EZ UV tray (taking advantage of the dark background) or by enclosing in a polyester document wallet and scanning on a Canon flatbed scanner with black card behind the gel. If reversal of the staining was required (e.g. for the gel to be blotted), the gel was incubated in PBS plus 100 mM EDTA for 5 minutes or until the gel turned transparent, followed by a water wash.

2.5.2.0: Silver staining

Silver staining is an extremely sensitive protein staining method, with approximate sensitivity of 1 ng for silver nitrate based methods. Downsides however include reduced compatibility with mass spectrometry and greatly increased complexity compared to other staining methods. SDS-PAGE gels were stained using a method described by Alphalyse Inc. (https://www.alphalyse.com/customer-support/protocol/silver-staining-mass-spectrometry-analysis/). Briefly, after electrophoresis gels were incubated for at least 30 minutes in fixative (40% ethanol, 10% acetic acid, 50% water), then washed for at least 30 minutes in water. The gel was sensitised in 0.02% $Na_2S_2O_3$ for 1 minute and then washed 3x in water for 20 seconds, before being incubated in cold 0.1% silver nitrate plus 0.002% formaldehyde for 20 minutes. An additional 3x 20 second washes in water were followed by development in 3% sodium carbonate plus 0.05% formaldehyde. Once staining was sufficient, development was terminated by washing in water, followed by incubation in 5% acetic acid for 5 minutes. The stained gel was then photographed using the white tray of the Gel-doc EZ.

2.5.2.0: In-gel fluorescence of XFP tagged proteins

Green fluorescent protein (GFP) and its various derivatives (XFP) can be excited in denaturing or native gels due to the strong stability of the fluorophore core, therefore allowing rapid selective viewing of XFP tagged proteins tagged without further processing. For SDS-PAGE gels the samples in loading buffer were not boiled to avoid destroying the GFP fluorescence and were instead incubated at 37 °C for 1 hr with protease inhibitors as described in section 2.5.1.0. After electrophoresis gels were mounted on the fluorescence tray of an Amersham Typhoon 5 laser imager (Cytiva Life Sciences). The 488 nm laser line with a Cy2 filter (525/20BP nm) was used to selectively excite the tag with a photomultiplier tube gain automatically determined during the pre-scan to avoid saturation; the gel image was recorded using the 50, 25 or 10 µm mode. An additional pass with the 635 nm laser line and a Cy5 filter (670/30BP nm) imaged the Bio-Rad Precision Plus all blue molecular weight standards.

2.5.3.0: Western blotting SDS-PAGE gels

Bio-Rad TGX gels were blotted using pre-packed and activated polyvinylidene difluoride (PVDF) membrane stacks with the Bio-Rad Trans-Blot Turbo semi-dry system. The lower part of the transfer stack was placed in the transfer module and the gel positioned on top of the membrane, followed by the top part of the stack; bubbles were gently pushed out with a roller before the lid electrode was locked in place and the module inserted into the system. For transferring smaller proteins of interest (<60 kDa) the standard 7-minute method (1.8 A constant, 25 V max for mini membranes; 2.5 A constant, 25 V max for midi or two mini membranes) was used, but for larger proteins the time was increased to 8 minutes to improve transfer efficiency. After electrophoresis the membrane was briefly washed in water. To fix the proteins strongly to the surface, the membrane was then either air dried at 37 °C, or incubated in ice cold acetone with gentle shaking at 4 °C for 30 minutes before being air dried at 50 °C, as described in Xu *et al.* (2019). The membrane was then re-activated in methanol, which was washed out with water. For non-stain-free gels the transfer was confirmed by soaking in Ponceau S stain (0.1% [w/v] in 1% acetic acid [v/v]), with a brief water wash to remove exogenous dye; the staining was then removed by repeated

washing in water. For stain free gels, transfer was confirmed using the stain free tray of the Bio-rad Gel-Doc EZ or Chemi-Doc MP.

2.5.4.0: Immuno-detection of blotted proteins

Blotted PVDF membranes were blocked for 2 hours at room temperature on an orbital shaker. As standard 5% (w/v) skimmed milk in PBS plus tween 20 (PBST: 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, 0.1% tween 20) was used. For membranes to be probed with anti-strep II primary antibodies however, 5% (w/v) bovine serum albumin (BSA) in PBST was used instead to avoid cross-reactivity with the biotin naturally present in milk. The primary antibody (details of antibodies used in Table 2.3) was applied at the appropriate concentration in 5% BSA and the membranes left on a rocking shaker at 4 °C for 12 – 24 hours. The membranes were subsequently washed in PBST for at least 25 minutes with a minimum of 5 changes of buffer. The horseradish peroxidase (HRP) conjugated secondary antibody was then applied in 5% skimmed milk in PBST for 1 hr at room temperature (anti-mouse 1:80,000-100,000; anti-rabbit 1:50,000) before the membranes were re-washed as before.

Table 2.3: Primary antibodies used for western blot immunodetection

Antibody	Clonality	Concentration	Host	Supplier	Supplier ID
Anti-His_6	Monoclonal	1:1000-2000	Mouse	Abcam	ab18184
Anti-StrepII	Polyclonal	1:2000	Rabbit	Abcam	ab76950
Anti-StrepII	Monoclonal	1:1000-1500	Mouse	Abcam	ab184224
Anti-GFP	Polyclonal	1:2000 –1:5000	Rabbit	Abcam	ab6556

Enhanced chemiluminescence (ECL) substrate (Pierce ECL Plus, Thermo-Fisher Scientific) was applied to the membranes for 5 minutes to interact with the HRP conjugated secondary antibody. The membranes were placed in an exposure cassette under polythene wrap or in a polyester document wallet and the signals recorded on X-ray film (Roche Diagnostics Lumi-Film Chemiluminescent Detection Film). Alternatively, the blots were imaged on a Typhoon 5 laser imager with the 488 nm laser and a Cy2 filter to exploit the additional chemifluorescence properties of the ECL substrate.

2.5.5.0: Stripping and re-probing of PVDF membranes

PVDF membranes to be re-probed were stripped of antibodies using a modification of a gentle acid-glycine method originally described by Legocki and Verma (1981). The membrane was incubated with the stripping buffer (200 mM glycine, 0.1% SDS, 1% tween-20 pH 2.2) twice for 10 minutes at room temperature with shaking, before being washed twice for 10 minutes in PBS, followed by two 5-minute washes with PBST. The stripping efficiency was checked by incubation with ECL and exposing to film for 5 minutes. If no signal was visible after development, then blocking and re-probing was proceeded to. If not, the stripping procedure was repeated.

2.5.6.0: Fluorescent size exclusion chromatography

For FSEC analysis, a Superdex 200 10/300GL or Superose 6 10/300GL size exclusion column was equilibrated with the appropriate buffer for the protein of interest, plus 0.05/0.005% DDM/CHS on an Äkta Pure system at 4 °C. The flow path length was minimised and the column mounted directly in or as close to the UV detector as possible to minimise sample dilution. Approximately 800-1000 µl of the detergent soluble fraction for each sample was injected on to the column using a 500 ul or 1 ml loop with a flow rate of 0.3 ml/ml and after 4 ml (Superdex 200 10/300GL) or 6 ml (Superose 6 10/300GL) of elution, 200 µl fractions were directly collected in black 96 well plates, with either opaque or clear bottoms. Fluorescence of the fractions was read on a BMG OptimaStar plate reader with the appropriate filter set for the fluorophore (405/510 nm for GFPuv and 485/510 nm for mEmerald), using a gain calculated to reach 70% of the plate readers dynamic range from a pre-scan of the plate. For quantitative comparisons between detergents, the plates were re-read at the lowest respective gain value. The fraction numbers were converted to volumes and plotted against the fluorescence to determine the dispersity of the fluorescence trace.

2.5.7.0: Live cell fluorescence microscopy

For live cell fluorescence microscopy of cells, 1 ml of suspension cells expressing the fluorophore tagged protein of interest were plated at a density of $1x10^6$ cells/ml onto #1.5 glass bottomed tissue culture treated 35 mm dishes and left to

adhere in the incubator for 24 hours. Samples were viewed under a Leica SP8 microscope in an atmosphere-controlled environment suited to the cell type: 27 °C with ambient CO_2 for *Sf*9 cells and 30 °C with 5% CO_2 for expi293 cells. The cells were imaged with a laser line and filter set suitable to excite the fluorophore of interest and images were recorded using Leica LAS X software. Recorded images were analysed using ImageJ (Schneider *et al.*, 2012).

2.5.8.0: Fixed cell fluorescence microscopy

For fixed cell fluorescence microscopy, HEK293 and *Sf*9 cells were grown on #1.5 borosilicate glass coverslips (VWR). For HEK293 cells, poly-L-lysine coating was used on the coverslips to enhance adherence. In a class II BSC, coverslips in a 24 well plate were coated with 100 µl of pre-diluted 0.01% poly-L-lysine (Sigma-Aldrich) using one of two methods. In the first, 0.01% poly-L-lysine was added to coverslips for 10 minutes before being aspirated and the excess washed out twice with DPBS. The coated coverslips were then incubated in 100% ethanol for 20 minutes to sterilise. After aspirating the ethanol, they were left in the BSCto air dry and then stored in the 24 well plate at 4 °C until needed. Alternatively, in a BSC the coverslips were first sterilised by dipping in 100% ethanol and passing through a blue Bunsen flame, before adding the 0.01% poly-L-lysine (Sigma) for 10 minutes in a 24 well plate. After aspirating and washing three times with DPBS, the coverslips were left to air dry for 30 minutes. For un-coated coverslips as used with *Sf*9 cells, the coverslips were sterilised in a 24 well plate with 70% ethanol for 10 minutes, followed by 10 minutes in 100% ethanol and two brief washes in DPBS.

After aspirating the culture medium, cells on coverslips were fixed in 4% (HEK293 cells) or 3% (*Sf*9 cells) paraformaldehyde in PBS for 10-20 minutes, before being washed twice in DPBS and stored at 4 °C in DPBS or PBS until needed. For samples being subjected to indirect immunofluorescence, the coverslips were incubated in 1% BSA in PBS at room temperature for approximately 1-1.5 hrs to block. For samples to be permeabilised, 1% BSA in PBS plus 0.1% TX-100 was then added for 10 minutes at room temperature. After washing twice in PBS to remove residual TX-100, the primary antibody was added in 1% BSA in PBS at the dilution listed in Table 2.4 and the coverslips incubated overnight at 4 °C with

gentle rocking. The next day they were washed 3 times in PBS for 5 minutes with gentle rocking. The secondary antibody (Alexa Fluor™ 594 conjugated anti-mouse [A-11032] or anti-rabbit [A-21207] or Alexa Fluor™ 680 conjugated anti-mouse [A-21057] from Thermo-Fisher Scientific) was then added at 1:1000 in 1% BSA in PBS for either 1 hour at room temperature or overnight at 4 °C with gentle rocking. After a further 3 rounds of washing in PBS, the coverslips were mounted on glass slides (VWR) using Vectashield with DAPI anti-fade mounting fluid (Vector Laboratories, Burlingham CA, USA). The slides were then imaged on a Zeiss LSM Xenon 710 inverted confocal microscope with a 60 x 1.5 NA objective and immersion oil. Appropriate laser lines and filter sets were used for the fluorophores of interest.

Table 2.4: Primary antibodies used for indirect immunofluorescence

Antibody	Clonality	Concentration	Host	Supplier	Supplier ID
Anti-His_6	Monoclonal	1:200	Mouse	Abcam	ab18184
Anti-StrepII	Monoclonal	1:350-400	Mouse	Abcam	ab184224
Anti-Na^+K^+ATPase	Monoclonal	1:500	Rabbit	Abcam	ab76020

2.6.0.0: Characterisation of purified proteins

2.6.1.0: Thermal shift assays

Thermal-shift assays determine the stability and folding state of a protein by using an increasing temperature ramp and measuring an output such as heat or fluorescence as the protein un-folds and aggregates. This is a useful method for the initial determination of buffer conditions or ligand binding that may increase the thermo-stability of a protein. Fluorescence based thermal-shift assays (generically referred to “thermofluor”) offer rapid determination of thermostability and use micro-plate or capillary approaches to reduce the amount of protein required and improve the throughput compared to traditional techniques such as differential scanning calorimetry (Huynh and Partch, 2015; Gao *et al.*, 2020).

The Nanotemper Tycho instrument (NanoTemper Technologies GmbH, München, Germany) uses a small (~ 10 µl) amount of sample drawn up into a single use glass capillary and utilises the intrinsic tryptophan fluorescence of a protein as a reporter. The instrument applies a rapid temperature ramp while

recording the fluorescence at 330 and 350 nm. Protein un-folding induces a red shift in the fluorescence, so taking the first derivative of the 330/350 nm ratio against temperature the inflection temperature (T_i) can be calculated, which gives a simple indication of protein quality and stability (Gao *et al.*, 2020).

2.6.3.0: Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a technique which quantifies the energy taken in or given out by a binding reaction by measuring the heat change upon injection of a binding partner. This can be used with protein-protein or protein-ligand interactions to quantify the thermodynamic binding parameters, from which the affinity and number of binding sites can also be calculated. These interactions can be exothermic or endothermic.

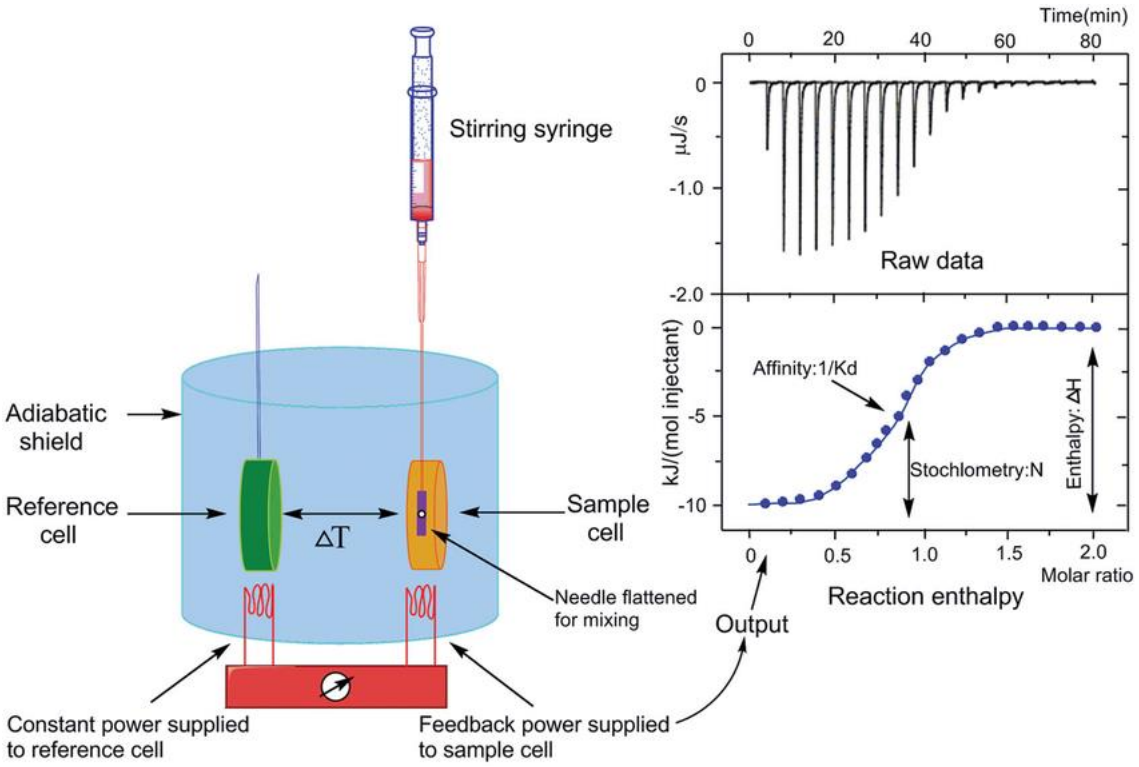


Figure 2.2: Graphical depiction of an isothermal titration calorimetry experiment.

Reproduced from Song *et al.*, (2015) under a creative commons attribution licence (CC BY 4.0).

In an ITC experiment, a protein or ligand sample is placed in a temperature-controlled cell (Figure 2.2), which is matched to a reference cell that has a constant power supplied by a heating element. By measuring the amount of feedback power required to maintain the sample cell at the temperature of the

reference cell, the net input or output of energy - the enthalpy - can be measured upon each injection of the binding partner protein or ligand into the sample cell.

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S$$

Equation 2.1: Calculation of the total energy (ΔG) and the entropic (ΔS) contribution to binding in an ITC experiment

Enthalpy (ΔH) and the association constant (K_a) are obtained from a binding model fitted to normalised injection heats. *R* is the universal gas constant and *T* the absolute temperature.

The change in enthalpy (ΔH) is measured under varying molar ratios of the cell and titrant samples and once the response saturates, all available binding sites are therefore occupied. The ΔH, number of sites (N) and the association constant (K_a, equal to 1/dissociation constant [K_d]) can then be calculated by fitting a binding model to a normalised injection heat trace. As depicted in Equation 2.1, from the fitted K_a the total energy (ΔG) can be calculated. Subsequently, from the ΔG and the previously fitted ΔH the entropic contribution (ΔS) to binding can be calculated. ITC is seen as the “gold standard” for quantification of binding parameters and is a truly label free technique. However, ITC is limited by the high amount of protein required, low throughput and the extensive trial and error required to obtain good quality data (Saboury, 2014).

ITC experiments were performed on a Microcal PEAQ-ITC instrument (Malvern Panalytical Ltd, Malvern, UK). Ligands were dissolved in dimethyl sulfoxide (DMSO) as 100 mM stocks, with the concentration of DMSO in the titrant and cell solutions matched to minimise heat changes from buffer mismatch. Multiple injection experiments utilised 12 x 3.0 µl or 15 x 2.5 µl injections protocols, while single continuous injection experiments injected 37.5 µl of titrant over 75 seconds. All experiments were performed at a cell temperature of 25 °C, with a reference power of 10 kcal/mol and a stirring speed of 750 RPM. The data were fitted with a one-site binding model in the manufacturers software and for multiple injection experiments the reference response of ligand titrated into buffer was subtracted.

2.6.4.0: Mass photometry

Mass photometry is a label free single molecule technique which utilises interferometric scattering of light off a glass-water interface to measure the mass of molecules (Young *et al.*, 2017). Being a single molecule technique, it is essential to remove surface contaminants from consumables used for mass photometry, as these interfere with analysis and produce false signals. 24x50 mm #1.5H glass coverslips (Paul Marienfeld GmbH & Co. KG, Lauda-Königshofen, Germany) were therefore washed in ultra-pure water followed by 100% isopropanol alcohol a total of 3 times, with a final wash in water to remove any remaining alcohol. These were then dried in a stream of compressed air to eliminate moisture. CultureWell™ re-usable 50 mm cell gaskets with 3 mm x 1 mm wells (Sigma-Aldrich) were cut into 4x4 well sets and washed and dried as with the coverslips. After cleaning the lens of the Refeyn One^{MP} instrument (Refeyn Ltd, Oxford, UK) extensively with 100% isopropanol and lint free cotton buds, immersion oil was applied to the lens and a coverslip gently mounted on the stage and oil drop. This was then secured with magnets. A prepared cut 4x4 well gasket was gently pressed onto the coverslip with forceps and the stage positioned so that the lens was under one of the gasket wells. 10 µl of buffer was applied to the well, the instrument closed and the AcquireMP software used to find the focus. If any adhered contaminating molecules were found in the field of view, the stage was re-positioned and the instrument re-focused until optimal. Native-mark standards (Thermo-Fisher Scientific) were used for calibration. 3 µl of standards diluted 1:1000 was applied to the prepared well, the focus re-adjusted and the data recorded. In the DiscovererMP software, the data was processed and binned. After fitting the peaks to the contrast values, the standards MW values were applied to create a calibration file. Samples were analysed using analogous procedures and the calibration file used to fit mass values to fitted gaussians.

Analysis of membrane proteins with mass photometry can be difficult due to the presence of empty detergent micelles, nanodiscs etc. A methodology described by Olerinyova *et al.* (2021) was therefore used, where the samples were rapidly diluted to beneath the detergent CMC to collapse empty micelles. Taking

advantage of the delay in protein-detergent micelles collapsing, these were quickly measured on the instrument.

2.7.0.0: Fluorescent membrane potential assay

The FLIPR membrane potential (FMP) assay blue kit from Molecular Devices (San Jose, CA, USA) was used to assess the function of GAT1 in a cell based 96 well format. For HEK293 cells, the plasmid of interest was transfected in with as previously described in section 2.2.2.1 and after 24 hours the cells were split into black walled, clear bottomed 96 well microplates with $5x10^4$ cells per well in 100 µl of media. After ~18-24 hours, when an 80-90% confluent layer had formed, the media was removed and the cells washed with Hanks balanced salt solution (1.26 mM $CaCl_2$, 0.49 mM $MgCl_2$-$6H_2O$, 0.41 mM $MgSO_4$-7H2O, 5.33 mM KCL, 0.44 mM KH_2PO_4, 4.17 mM $NaHCO_3$, 137 mM NaCl, 0.34 mM Na_2HPO_4, 5.56 mM D-glucose) with HEPES (20 mM) buffer (HBSS+HEPES), which was made from a 10x stock provided in the kit or from a self-prepared 10x stock. 100 µl of HBSS+HEPES buffer was added to each well, except for when the cells were going to be pre-incubated with the GAT1 antagonist tiagabine, in which case only 50 µl of buffer was added. Tiagabine (Sigma-Aldrich) was diluted from frozen 100 mM in water aliquots and made up in HBSS+HEPES buffer at 4x the final concentration. 50 µl of diluted tiagabine was then added to the wells of interest.

For *Sf*9 cells, suspension cultures were infected with the baculovirus of interest for ~36 hours. The cells were pelleted at 200 *g* and washed in a physiological buffer (128 mM NaCl, 2 mM KCl, 4 mM $MgCl_2$, 1.8 mM $CaCl_2$, 35.5 mM sucrose, 5 mM HEPES, pH 7.1, [Lee *et al.*, 1993]) before being re-suspended at $0.8x10^6$ cells/ml. 100 µl of cells were plated per well of a black walled, clear bottomed 96 well plate and left in the hood for 30 minutes to settle. The plate was then centrifuged at 1000 RPM for four minutes with the brake off to tightly adhere the cells to the plate.

An aliquot of FMP blue dye was re-suspended in 10 ml HBSS+HEPES or *Sf*9 physiological buffer as appropriate and 100 µl applied to the 100 µl of buffer in each well of the 96 well plate with a multichannel pipette. The plate was then incubated at 37 °C or 27 °C respectively for HEK293 and *Sf*9 cells, before being loaded into the read chamber of a Molecular Devices FlexStation 3 robotic plate

reader. The automated dispensing system was loaded with the compounds of interest made up in HBSS+HEPES or *Sf*9 physiological buffer at five times the final concentration, along with a buffer only control well for each lane. The plate reader was programmed to apply the loaded concentrations of each compound to two or three adjacent lanes of the plate for duplicate/triplicate reads. The software (SoftMax Pro 5) was set with the excitation and emission frequencies as 530 nm and 565 nm and the photomultiplier tube sensitivity as “high”. Averaged reads for each well were reported every 1.52 seconds, with the compounds being dispensed at the maximum rate into each lane of wells after recording 60 seconds of baseline. Recording stopped after 180 seconds and the data was automatically transformed in SoftMax Pro to present as maximum – minimum, then analysed in GraphPad Prism 9. The response to buffer only was subtracted from each lane and a variable slope Hill equation fitted, with the bottom constrained to zero.

2.8.0.0: Screening a recombinant nanobody yeast library

Nanobodies were raised again GAT1 by screening a synthetic yeast library, pYDS649, using procedures modified from Mcmahon *et al.*, (2018). To briefly describe, the nanobody selection procedure used rounds of screening the yeast against biotinylated protein. Fluorescently labelled streptavidin was used to label yeast which had bound to biotinylated GAT1 and after addition of anti-fluorophore conjugated streptavidin microbeads beads, magnetic cell sorting could selectively isolate these yeast for use in subsequent screening. Decreasing concentrations of biotinylated protein in each round should enable the isolation of binders with a steadily increasing affinity to the protein.

A frozen aliquot of pYDS649 yeast was defrosted and used to inoculate 500 ml of YGLC media + 40% glucose + penicillin/streptomycin (YGLC-glu). The culture was expanded overnight at 30 °C/190 RPM, then $1x10^{10}$ cells were used to inoculate a further 500 ml culture of YGLC+40% galactose + penicillin/streptomycin (YGLC-gal) media. The library is under the control of the GAL1 promoter and nanobody expression was induced for 48-72 hours at 20 °C/190 RPM. Nanobody library expression was confirmed using anti-Alex-Fluor™

488 labelled anti-HA antibody and flow cytometry. These initial procedures were carried out by Dr. Tom Eadsforth.

The library was subsequently used for selection rounds against purified GAT1, which was biotinylated with an EZ-Link™ Sulfo-NHS-LC-LC-Biotin kit (Thermo-Fisher Scientific). PBE (PBS plus 0.5% BSA and 2 mM EDTA) plus 0.05/0.01 % DDM/CHS buffer was used for all procedures during the screening rounds. Efforts were made to maintain a sterile work environment at all times to prevent the yeast cultures from becoming contaminated with exogenous yeasts or bacteria.

Selection worked on a basis of each round being comprised of negative selection to remove binders that did not bind streptavidin, followed by positive selection against GAT1. For the negative selection rounds, $5x10^9$ cells (for the first round) or $1x10^8$ cells (for subsequent rounds) were pelleted at 3500 *g* for five minutes and washed in PBE, before being resuspended in PBE plus 100 µl of streptavidin microbeads during the first round or Alex-Fluor™ 647 conjugated streptavidin microbeads (Miltenyi Biotech, Bergisch Gladbach, Germany) for subsequent rounds. This was left to rotate for 40 minutes at 4 °C before the yeast were pelleted at 1000 *g*, washed and re-suspended in PBE and passed down an equilibrated LD column (Miltenyi Biotech) on a magnetic rack at 4 °C to deplete yeast which non-specifically bound streptavidin.

For the subsequent positive selection, during the first round 400 nM of biotinylated GAT1 was pre-incubated for 5 minutes at room temperature with 250 µl of streptavidin microbeads before being added to the washed yeast and left to rotate for 1 hour at 4 °C before proceeding to magnetic sorting. In the second round though, 400 nM of GAT1 was pre-incubated with 100 nM of AF647 labelled streptavidin (Thermo-Fisher Scientific) and left to rotate for 2 hours at 4 °C before being washed and 50 µl of anti-AF647 conjugated streptavidin microbeads added in 950 µl of PBE. This was left to rotate for 20 minutes at 4 °C before the yeast were pelleted and resuspended for sorting. For these first two rounds biotin-streptavidin tetramers were formed during the pre-incubation to encourage enrichment for avidity rather than affinity with the yeast. From the third round onward though, biotinylated GAT1 was added directly to the yeast for monomeric antigen selection to encourage affinity. This was left to rotate at 4 °C for 1-2 hours before being washed out and the yeast labelled with 50 nM of AF647 conjugated

streptavidin for 15 minutes. After washing, 50 µl of anti-AF647 conjugated streptavidin microbeads in 450 µl of PBE were for 15 minutes at 4 °C with gentle rotation.

The cells were magnetically sorted to retain yeast which had bound biotinylated GAT1 and were hence labelled with the microbeads of interest for the selection round. Prior to sorting a 10 µl sample of yeast were retained and diluted in 200 µl of PBE (stained pre-sort sample) for later analysis. An LS column (Miltenyi Biotech) was equilibrated with PBE on a magnetic rack at 4 °C and the yeast passed through. A 10 µl sample of flow through yeast was retained. After washing the column with PBE, the bound yeast were eluted with a plunger in PBE and a 10 µl sample retained (post sort sample). The remainder were pelleted and resuspended in YGLC+glu medium and grown overnight at 30°C/190 RPM. The next day the OD600 of the yeast was measured and 4 ml of YGLC-gal +penicillin/streptomycin culture inoculated at an OD600 of 1. This was incubated at 20 °C/190 RPM for 48-72 hours. The retained samples were analysed in the APC channel of a Beckman-Coulter Cytoflex 5 flow cytometer to quantify AF647 positive yeast over the total number. The gates were set automatically from the FSC-A vs SSC-A plot.

After the final round, Dr. Tom Eadsforth and Mr. Greg Stewart plated the yeast on YPD media at different dilutions to isolate single colonies. 24 of these were subsequently grown up in liquid culture, before being lysed with zymolase and plasmid DNA extracted and purified using standard methods. The plasmid DNA was heat shocked into DH5α cells, which were plated and subsequently used to setup overnight liquid cultures. Plasmid DNA was then extracted and purified with standard methods and submitted for Sanger sequencing using primers flanking the nanobody sequence in the plasmid.

2.9.0.0: Acetylcholine Binding Protein X-Ray crystallography

2.9.1.0: Crystallisation of the AChBP:BS82 complex

Initial crystallisation trials were carried out mixing 0.1 µl of protein solution with 0.1 µl reservoir in two well MRC sitting-drop vapour diffusion plates. Drops were dispensed with a Mosquito crystallisation robot (SPT Labtech Ltd, Melbourn, UK),

exploiting the JCSG+ and Morpheus crystal screens (Molecular Dimensions, Rotherham, UK) with protein solutions of 4 and 10 mg/ml AChBP incubated with cytisine (3 mM, Sigma-Aldrich) or ligand BS82 ({(1*R*,5*S*)-9-(3-hydroxypropyl)-1,2,3,4,5,6-hexahydro-1,5-methano-8*H*-pyrido[1,2a][1,5]diazocin-8-one}, 5 mM) for 1-hour at room temperature. No crystals formed in either screen when co-incubated with ligand BS82. For cytisine however, 10 mg/ml protein plus 3 mM cytisine with a reservoir containing 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 0.1 M HEPES pH 7.5 produced small crystals. With the objective of inducing formation of AChBP:BS82 crystals, these initial AChBP:Cytisine crystals were crushed by pulse vortexing with a Seed Bead™ (Hampton Research, CA, USA) and applied to screening plates for random matrix micro-seeding (RMMS) as described by Shaw *et al.* (2011); remains of this seed stock was frozen at -80 °C for later use. RMMS applies an initial seed stock of the protein of interest (or a close homologue) to the screening conditions, which can increase the number of hits and/or crystal size by inducing nucleation. Seeded plates used a protein solution of 12.5 mg/ml plus 6 mM ligand BS82 in 0.6 µl drops (0.3 µl protein solution + 0.2 µl reservoir + 0.1 µl seed stock mixture); larger crystals grew in the same reservoir as the initial seed stock. 24 well hanging drop plates with 2 µl drops (all including 0.3 µl seed stock) were subsequently setup to optimise conditions, testing drop ratios, precipitant concentration, pH and additives. Suitable crystals with maximal dimensions of ~400x300 µm grew in a reservoir of 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 10% glycerol, 0.1 M HEPES pH 7.0 with the drop containing 1.5 µl protein and 0.2 µl reservoir.

2.9.2.0: Crystallisation of the AChBP:(1) complex

Crystals of AChBP bound with ligand (1) (4-[5-[(1~{R},2~{R},4~{S})-7-azabicyclo[2.2.1]heptan-2-yl]-2-fluoranyl-pyridin-3-yl]benzamide) were previously grown by Dr. Renatta Bueno. AChBP at 4 mg/ml in buffer A was incubated with 2 mM of ligand (1) for 1-hour at room temperature. Initial crystals of AChBP bound with (1) were obtained using hanging drops of 2 µL protein solution plus 1 µL of reservoir (0.2 M NaCl, 0.1 M of phosphate/citrate pH 4.2, 12% PEG 8000) equilibrated against 800 µL of reservoir for 24 hours at 18 °C. A well-formed crystal was transferred to a microfuge tube containing reservoir solution, 40% glycerol and a Seed Bead™. The sample was vortexed for 1 minute

to crush the crystal, then a cleaned eyelash was dipped into the seed stock and passed through newly assembled crystallisation drops. Crystals with maximum dimensions of approximately 20 µm appeared after four days.

2.9.3.0: Diffraction data processing and model building

Crystals of the AChBP complexes were harvested with a nylon loop and cryo-protected in 20% glycerol for AChBP:BS82) and ethane-1,2-diol (EDO) for AChBP:(1) made up in the reservoir solution. The crystals were then snap-frozen in liquid N_2. Diffraction data were recorded at Diamond Light Source on beamline I03 for AChBP:BS82 and beamline I24 for AChBP:(1). The diffraction data were indexed and integrated with *DIALS* (Winter *et al.*, 2018) for AChBP:BS82 and XDS (Kabsch, 2010) for AChBP:(1), then scaled and merged in Aimless (Evans and Murshudov, 2013). The structures were solved by molecular replacement with Phaser (McCoy *et al.*, 2007), using the 1.90 Å resolution structure of *Ac*AChBP in complex with strychnine (PDB ID 2XYS, Brams *et al.*, 2011) as the search model for the AChBP:(1) data. The finished 2.20 Å resolution model of AChBP:(1) (PDB ID 6QKK; Bueno *et al.*, 2022) was subsequently used as the search model for the AChBP:BS82 data. REFMAC 5 (Murshudov *et al.*, 2011) was used for multiple rounds of automated restrained refinement, with manual refinement and model building in Coot (Emsley and Cowtan, 2004). *MolProbity* (Chen *et al.*, 2010) was used for Ramachandran analysis. Restraints and models for ligands were generated using the GRADE server (Global Phasing - http://grade.globalphasing.org/cgi-bin/grade/server.cgi) and the PyMol Molecular Graphics System (Schrödinger) used to create 3D graphics. AChBP is glycosylated and N-acetyl-D-glucosamine (NAG) was modelled onto Asn91 in every subunit. Strict local non-crystallographic symmetry restraints were applied initially but then relaxed during the course of the refinement. Where multiple conformations were visible for certain residues, the occupancies were manually assigned and tested until the difference density maps suggested appropriate modelling. OMIT maps were generated by re-calculating the F_O-F_C maps after removal of the feature of interest from the model, with Polder OMIT maps (Liebschner *et al.*, 2017) having the bulk solvent correction turned off during re-calculation.

Chapter 3: Nicotinic Acetylcholine Receptor Ligand Structure-activity Relationships

This chapter focuses on two series of derivatives of the natural nAChR ligands epibatidine and cytisine. The nature of how the differing modifications made to the originating molecules affected their molecular interactions and affinity to nAChRs was unclear. Therefore, AChBP was used as a surrogate of the nAChR orthosteric binding site, with the objective of elucidating structure-activity relationships within the two series of derivatives.

3.1.0.0: Aims:

- Quantify the ligand binding affinity for the binding of novel epibatidine and cytisine derivates to AChBP.

- Crystalise the ligands with AChBP and elucidate the molecular interactions.

3.2.0.0: Epibatidine derivatives overview

Originally extracted from the skin of the Equadorian poison-arrow frog *Epipedobates tricolor* (Traynor, 1998), epibatidine (Figure 3.1) is a potent - K_d 0.018 nM at recombinant human $\alpha_4\beta_2$ receptors (Ondachi *et al.*, 2016) - nAChR agonist, with powerful non-opiate mediated anti-nociceptive effects. Due to toxicity and a narrow therapeutic window, epibatidine has never been studied *in-vivo* in humans, but it has found use as a basis for developing nicotinic receptor targeting compounds (Spande *et al.*, 1992; Traynor, 1998; Mu *et al.*, 2006).

Compounds (1), (2) and (3), (4), (5) and (6) (Figure 3.1) are 2′-Fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues developed at the Research Triangle Institute. These have been previously characterised as having high (K_d <1 nM) affinity at $\alpha_4\beta_2$ nAChRs, but differing degrees of subtype selectivity and novel pharmacological effects compared to epibatidine. The synthesis of these analogues has been previously reported in Ondachi *et al.* (2016).

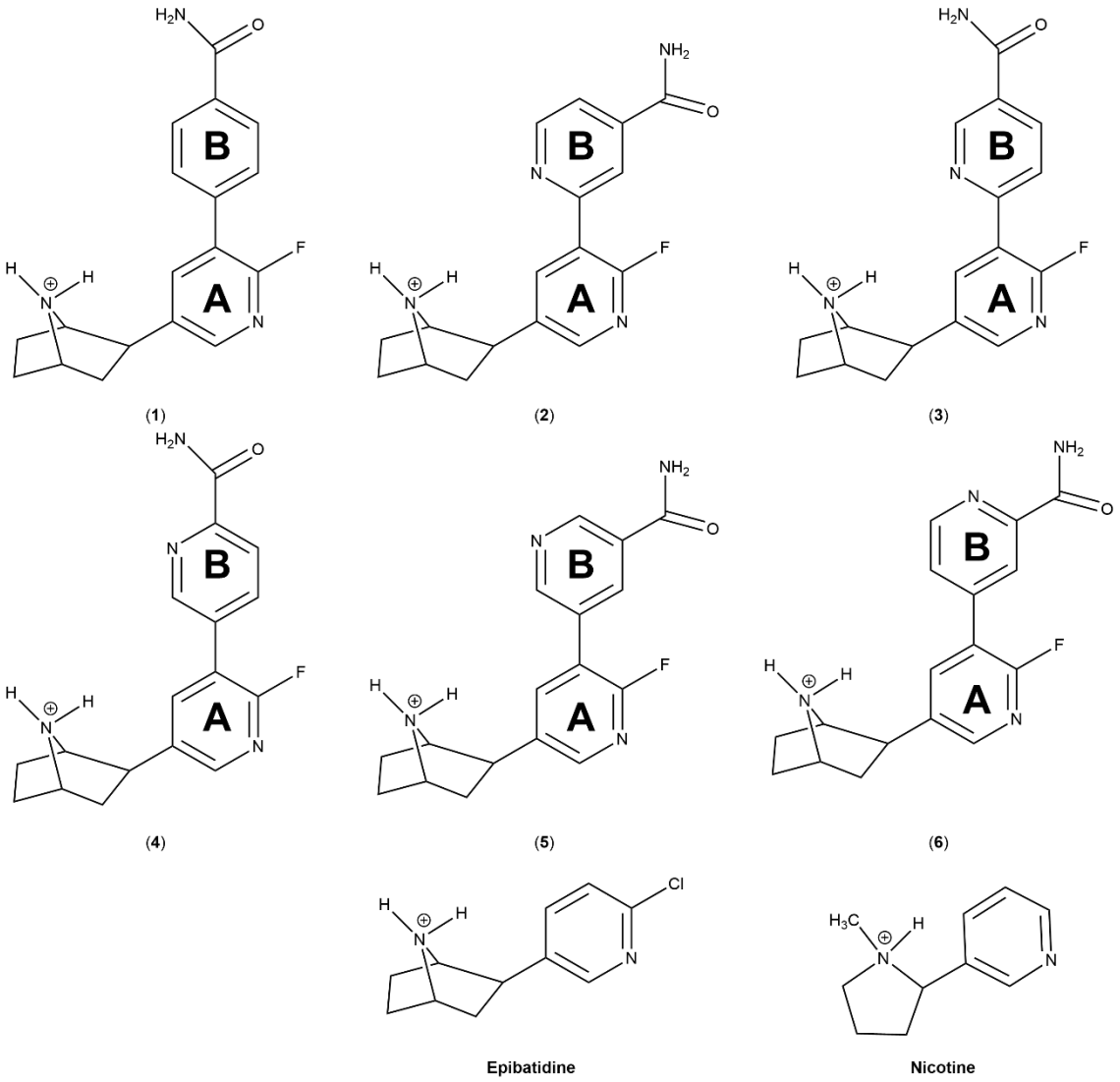


Figure 3.1: Structures of epibatidine, nicotine and epibatidine derivatives (1)-(6).

The protonated forms most likely under physiological conditions are shown. The first fluoropyridine ring is labelled as **A** and the second benzylcarboxamide (1) or pyridinecarboxamide (2)-(6) ring is labelled as **B**.

Xenopus oocyte electrophysiology previously characterised all of the analogues as demonstrating almost exclusively antagonist activity at recombinant $\alpha_4\beta_2$ and $\alpha_3\beta_4$ nAChRs, in contrast to the full agonist profile of epibatidine. However, at α_7 receptors differing effects were observed, with (3) acting as a partial agonist with no antagonist properties, (1) as a partial agonist with additional antagonist effects and (2), (4), (5) and (6) as purely antagonists. (4) and (5) were furthermore suggested to be more selective for the α_7 receptor over the $\alpha_4\beta_2$. Studies *in-vivo* have furthermore suggested that these analogues potently antagonise the anti-nociceptive action of nicotine, with (2) approaching the potency of the $\alpha_4\beta_2$ partial agonist and α_7 full agonist varenicline (Ondachi *et al.*, 2016). While past

epibatidine analogues focusing on the position of the chloropyridine ring nitrogen have also been associated with a conversion to an antagonist profile (Spang *et al.*, 2000), the reasons for the derivatives showing differing actions at nAChR subtypes are unclear.

*Ac*AChBP was therefore utilised as a surrogate of the nAChR binding site to try and illuminate why the derivatives showed differing modes of action at nAChR sub-types. Binding assays and co-crystallography studies were used to respectfully elucidate the affinity and the molecular interactions for the derivatives. These structure-activity relationships may assist in the design of future nAChR ligands which can act as research tools or as potential therapeutic molecules.

3.3.0.0: Epibatidine derivatives results and discussion

This section has been published in Bueno *et al.* (2022), a copy of which is included in Appendix S2. Contributions by those other than the author are defined in the text. Selected figures are reproduced or modified from those in the paper.

Table 3.1: Epibatidine analogues affinity properties at *Ac*AChBP

Ligand	*Ac*AChBp K_d (nM) [±SE]	$\alpha_4\beta_2$ nAChR K_i (nM)* [±SE]
(1)	9.8 [0.03]	0.12 [0.02]
(2)	60 [0.1]	0.28 [0.01]
(3)	29 [0.1]	0.94 [0.07]
(4)	33 [0.1]	0.07 [0.002]
(5)	30 [0.1]	0.28 [0.03]
(6)	9.8 [0.01]	0.67 [0.28]
Epibatidine	14.0[&]	0.02 [0.001]
Varenicline	342.0[$]	0.12 [0.002]
Nicotine	835.0[$]	0.95[^]

[&] Stopped flow Trp fluorescence quenching assay (Hansen *et al.*, 2005). * [^{3}H]-Epibatidine competition assay, K_d = 0.02 nM (Ondachi *et al.*, 2016). [$] Isothermal titration calorimetry (Rucktooa *et al.*, 2012). [^] [3H]-nictotine competition assay (Coe *et al.*, 2005)

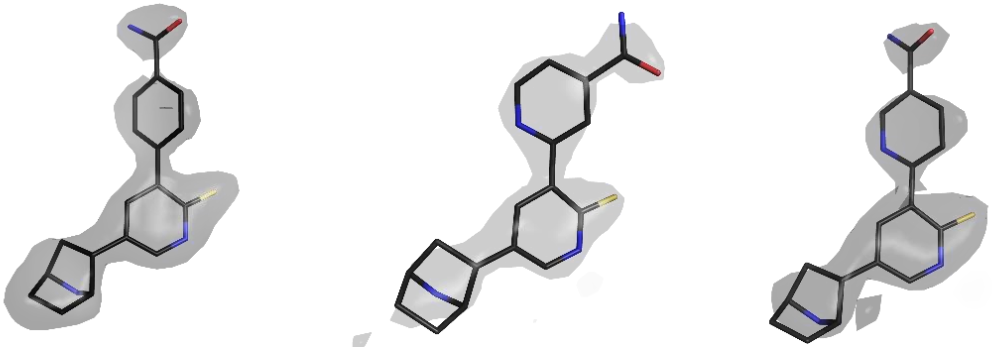
3.3.1.0: Epibatidine derivatives binding studies

*Ac*AChBP was produced by Dr. Renata Bueno, using a construct previously designed by Dr. Alice Dawson (Dawson *et al.*, 2019). Biolayer interferometry was used by Dr. Renata Bueno to characterise the affinity of the six epibatidine derivatives. All of the compounds showed low-nanomolar affinities to *Ac*AChBP (Table 3.1), which are comparable to those defined in previous literature for epibatidine and are higher affinity than

those for nicotine. These K_d values are several orders of magnitude reduced however when compared to previous [^{3}H]-epibatidine competition assays conducted against the $\alpha_4\beta_2$ nAChR, which showed sub-nanomolar K_i values for each compound (Ondachi *et al.*, 2016). This reduction in affinity is in harmony with previous ligand binding studies using *Ac*AChBP.

3.3.2.0: Epibatidine derivatives crystallographic studies

Dr. Renata Bueno attempted to co-crystallise all six compounds with *Ac*AChBP and was successful in obtaining crystals with (1), (2) and (3). The models with (2) and (3) were built and refined by Dr. Renata Bueno to 2.4 Å and 2.5 Å resolution. The diffraction data for (1) was processed and a model built and refined by the author to 2.2 Å resolution. Crystallographic statistics for (1) are shown in Table 3.2. The models have been deposited in the PDB with codes 6QKK (1), 6QQP (2) and 6QQO (3).



(1) **(2)** **(3)**

Figure 3.2: Polder OMIT maps contoured at 4.0 σ for representative copies of the epibatidine derivatives in the *Ac*AChBP binding site

Table 3.2: Crystallographic statistics for *Ac*AChBP-(1)

PDB Code	6QKK
Unit cell *a*,*b*,*c* (Å) *a*,*b*,*g*, (°)	 211.0 131.6, 131.8 90, 102.8, 90
Space group	*C*2
Source	Diamond Beamline I24 microfocus
Wavelength	0.96858
Subunits per asy. Unit (Å)	10
Resolution range and outer shell	48.83-2.20 (2.24-2.20)
Other ligands	NAG, ethane-1,2-diol, phosphate, oxalate
Total number reflections	611399 (30075)
Unique reflections	177603 (8766)
Redundancy	3.4 (3.4)
R_{merge}	0.172 (0.888)
R_{pim}	0.162 (0.831)
Wilson *B* [Å^2]	15.5
Completeness [%]	99.9 (98.4)
<I/s(I)>	6.6 (2.0)
CC(1/2)	0.95 (0.43)
Refinement	
R_{work}/R_{free}	0.199/0.228
No. reflections for R_{work}/R_{free}	168732/8867
Protein residues	2065
No. ligands	10
No. water molecules	1287
RMSDs	
Bonds [Å]	0.048
Angles [°]	0.97
Ramachandran plot	
Residues in favoured regions	2002
Residues in allowed regions	35
Residues in outlier regions	0
Mean B-factors	
Protein atoms per subunit [Å^2]	18.4/15.7/17.3/ 19.2/17.3/18.4/ 20.5/22.2/22.6/ 19.2
Water molecules [Å^2]	22.2
Ligand [Å^2]	14.5/14.2/16.4/ 14.8/15.3/16.8/ 17.8/20.7/18.1/ 18.1
NAG	68.7
Ethane-1,2-diol	34.4
Phosphate	41.0
Oxalate	41.5

Values in parentheses are for the highest resolution shell.

Two pentamers were present in the asymmetric unit for all three structures and the electron density was good enough to unambiguously place ten copies of each ligand of interest (Figure 3.2). The interaction patterns and orientation are exceedingly similar for each copy within the binding site, so it is sufficient to describe only one. The ligands were applied as an enantiomeric mixture, but the resolution however was insufficient to determine whether the (+), (-) or a mix of forms were present. Therefore, both the (+) and (-) forms were tested being modelled in the electron density.

The (+) enantiomers were eventually positioned as this corresponds to the naturally occurring form of epibatidine. Based on epibatidine (Mu *et al.*, 2006; Dallanoce *et al.*, 2012), it is likely though that there are similar binding and functional properties for the (+) and (-) forms of each analogue.

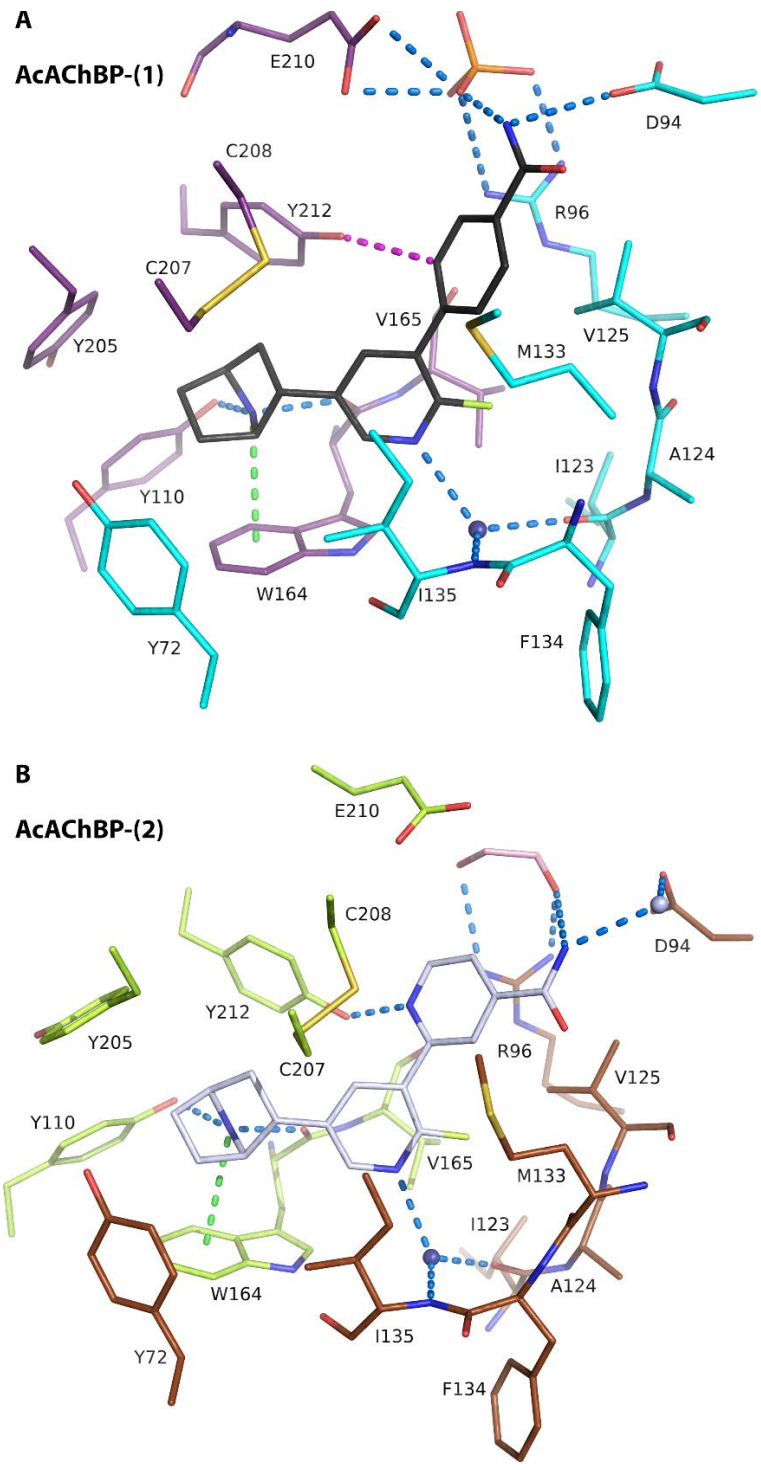


Figure 3.3: Selected *Ac*AChBP residues involved in interactions with analogues (1) and (2)

A: compound (1), **B:** compound (2). A dark blue sphere indicates the position of a conserved water molecule, blue dashed lines represent potential hydrogen bonds and green dashed lines cation-π interactions. In (**A**) the principal subunit residues are purple and those on the complementary side are cyan. Compound (1) is represented as black sticks. A possible C—H···O hydrogen bond is shown as a magenta dashed line and the phosphate atom is orange. In (**B**) the principal subunit residues are light green and those on the complementary side are brown. Compound (2) is represented as silver sticks. The C atoms of EDO (ethane-1,2-diol) are pink and the bridging water between (2) and D94 is represented as a light blue sphere.

In the crystal structures the orientation and pattern of interactions in the common azabicyclo[2.2.1]heptane moiety and the fluoropyridine ring A are similar across the analogues (Figures 3.3, 3.4). The azabicyclo[2.2.1]heptane moiety protonated amine donates hydrogen bonds to the sidechain of Tyr110 and the backbone of Trp164 on loops A and B respectively. A cation-π interaction is suggested between the amine and Trp164, which stabilises the ligand position. This interaction is an important feature for ligand affinity (Shahsavar *et al.*, 2016) which is highly conserved across ligands in PLGIC (Taly *et al.*, 2009; Nys *et al.*, 2013), and particularly nicotinic receptor, binding sites (Zhong *et al.*, 1998). Additional interactions are via van der Waals (VDW) forces with Tyr205, Tyr212, Trp164 and Cys207 and principal side and Tyr72 on the complementary side. The fluoropyridine ring A is orientated toward the complementary side of the binding site, forming van der Waals interactions with Val165 and Ile 135. The fluoropyridine nitrogen furthermore accepts a hydrogen bond from a water molecule which is conserved in all three structures and subsequently bonds respectively with the carboxyl and amino groups of Ile123 and Ile135 on the (-) side and the sidechain N of Trp164 on the (+) side.

Despite the ligands showing mainly antagonist effects, the interactions in the azabicyclo[2.2.1]heptane moiety and the fluoropyridine ring A closely match those of epibatidine, nicotine and other agonists at *Ac*AChBP (Hansen *et al.*, 2005; Rucktooa *et al.*, 2012; Dawson *et al.*, 2019). The ordered water is commonly seen in AChBP structures, but it is unknown how the interactions of ligands (1), (2) and (3) may compare to those of epibatidine in the AChBP binding site as the only available structure is (PDB ID 2BYQ, Hansen *et al.*, 2005) are low resolution and does not show solvent molecules. Neither of two *Ac*AChBP structures (PDB ID 3SQ6,Kouvatsos *et al.*, 2016); PDB ID 5JV engineered to more closely replicate the binding site of nAChRs show any detail either. Some studies (Nys *et al.*, 2013) have previously suggested that agonist activity is consistent with interactions with this ordered water, but the structures with (1), (2) and (3) suggest that this is not always the case.

Epibatidine analogues (1) to (6) all have a similar phenyl or pyridine ring B with a carboxamide at differing positions. The structures of *Ac*AChBP with (1), (2) and (3) suggest VDW interactions with Arg96, Val125 and Met133 on the (+) side and

Cys208 on the (-) side. With (2) and (3), the pyridine ring N is suggested to accept a hydrogen bond from Tyr212 on loop C, but the phenyl ring of (1) is instead suggested to form a possible C—H···O hydrogen bond.

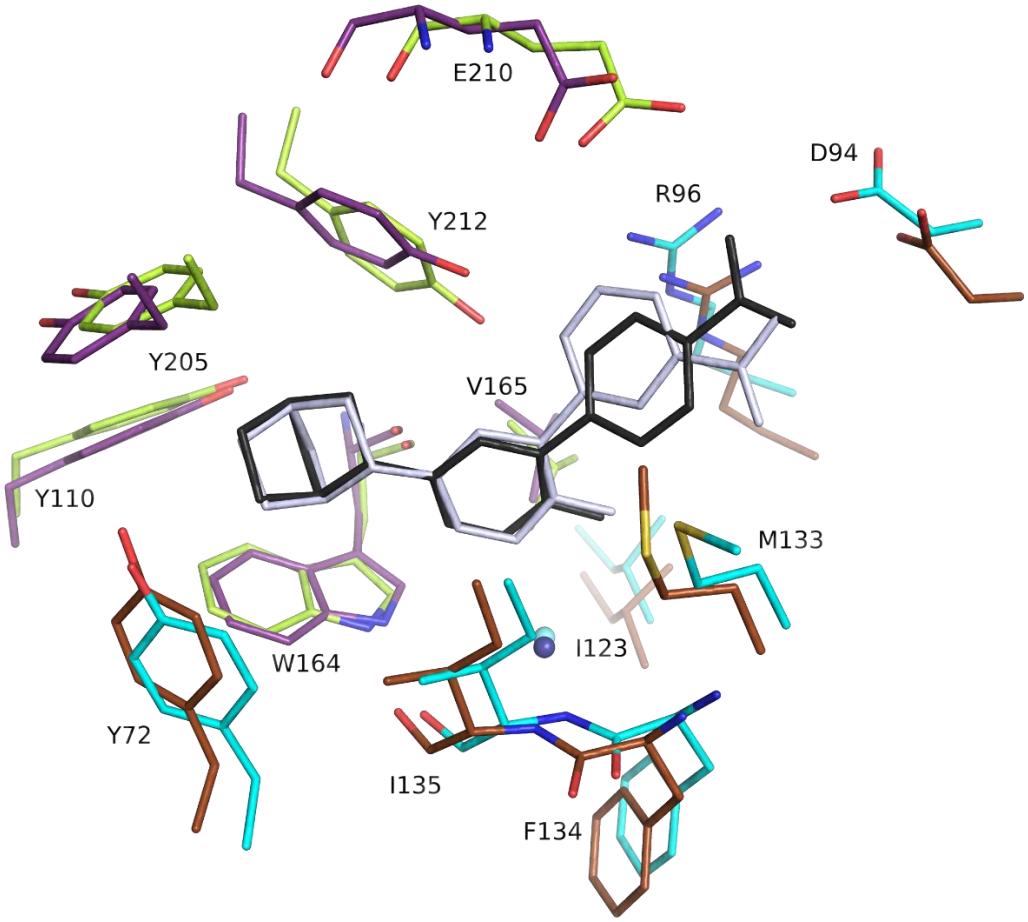


Figure 3.4: Overlay of selected residues for *Ac*AChBP bound with compounds (1) and (2)

Overlaid using the superimpose ligands options in Coot, Principal side residues for the complex with compound (1) are shown in purple and those for the complex with compound (2) are shown in light green. Complementary side residues for the complex with compound (1) are shown in cyan and those for the complex with compound 2 are in brown. The conserved water is shown as a sphere in aquamarine and in dark blue for the complexes with compounds (1) and (2), respectively. Ligand (1) is coloured black and ligand (2) is coloured silver.

In the structures with (1) and (3), the carbonyl groups of the carboxamide is oriented toward and interact with Val125, Thr127 and Ser131 on loop E via bridging water molecules. The amide groups however hydrogen bond with the side chains of Asp94 and Arg96 on the principal side and a phosphate molecule present in the crystallisation mix. Asp94 and Arg96, in addition to Glu170, Glu210, Lys42 and Ser167 from the complementary side, co-ordinate the phosphate. As the protein was crystallised at an acidic pH, this explains why the phosphate is co-ordinated by the Asp and Glu residues. In eight binding sites for (1) and seven

for (3), the amide directly hydrogen bonds with Asp94. In the structure with (2) however, the different position of the carboxamide on ring B orientates the carbonyl group towards loop E and particularly Met133 (Figure 3.4). Ring B however is positioned further away from the Met133 sidechain compared with (1) and (3) and the sidechain rotamer is directed toward rather than away from the ligand. The pyridine ring N is correspondingly closer to Tyr212, which is slightly adjusted in position to maintain the hydrogen bond. The carboxamide is hence in a similar location with all three analogues. Instead of a phosphate, an EDO – used as the cryoprotectant – interacts with the amide of (2) and maintains similar interactions with surrounding residues. In contrast to (1) and (3) though, the carboxamide amide of (2) can only interact with Asp94 via bridging water molecules. The differing position of ring B may prevent the binding of phosphate and it is speculated that the reduction in van der Waals interactions with Met133 by (2) could contribute to it showing the lowest affinity to *Ac*AChBP.

To surmise, two main differences have been identified between the molecular interactions of the epibatidine derivatives at *Ac*AChBP. (2) and (3) can directly hydrogen bond to Tyr212, while (1) can only form a possible C—H···O bond. (1) and (3) can directly hydrogen bond to Asp94, but (2) can only interact via a bridging water because of the differing position of the carboxamide. The slightly altered position of ring B for (2) also orientates the side chain of Met133 toward the ligand rather than away as with (1) and (3). These difference in interactions may explain the differing affinities that the derivatives possess to *Ac*AChBP.

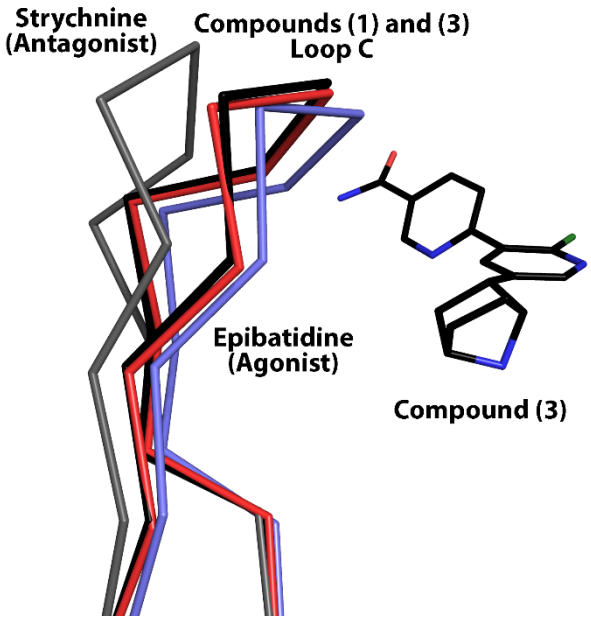


Figure 3.5: Positioning of loop C for compounds (1) and (3) compared with examples for ligands with different pharmacological properties.

Ribbon representation for loop C in *Ac*AChBP bound with the antagonist strychnine (grey, PDB ID 5O8T, Dawson *et al.*, 2019), the agonist epibatidine (blue, PDB ID 2BYQ, Hansen *et al.*, 2005) and compounds (1) (red) and (3) (black). Ligand (3) is shown as black sticks.

Loop C closes over the ligand in all of the structures (Figure 3.5) and contributes to creation of the orthosteric site by directly interacting with residues on the complementary side. The loop has to present with an initially open conformation to enable ligands to ingress and as the complex forms the loop subsequently closes (Jadey and Auerbach, 2012). The averaged solvent-accessible surface areas of unbound compounds (1)–(3) are 504.2, 499.5 and 50.17 Å^2 respectively and 91.4%, 93.9% and 91.9% of these surfaces are buried in the complexes with *Ac*AChBP, with primarily the carboxamide moiety directed towards solvent. In comparison, as epibatidine is smaller 99.4% of the 360.2 Å^2 average surface area is buried after binding.

Comparisons of the binding sites of the *Ac*AChBP complexes to those of the cryo-EM structures of the $\alpha_4\beta_2$ nAChR in complex nicotine (PDB ID 6CNJ, Walsh *et al.*, 2018) and the α_7 nAChR in complex with epibatidine (PDB ID 7KOX, Noviello *et al.*, 2021) and the sequences of human α_4, α_7 and β_2 subtypes were carried out (Figure 3.6). The conservation of sequence and structure of the structure, especially for the five residues comprising the aromatic box, suggests that the orientation of how the analogues bind to *Ac*AChBP would be representative of how they would bind to any nAChR combination. In the α_7 homomer, site directed mutagenesis studies have previously shown that epibatidine binding is abolished

if any of six highly conserved residues are changed (Thompson *et al.*, 2017). In *Ac*AChBP these correspond to Tyr72, Tyr110, Tyr205, Tyr212 and Trp164 in the aromatic box and additionally Ile123.

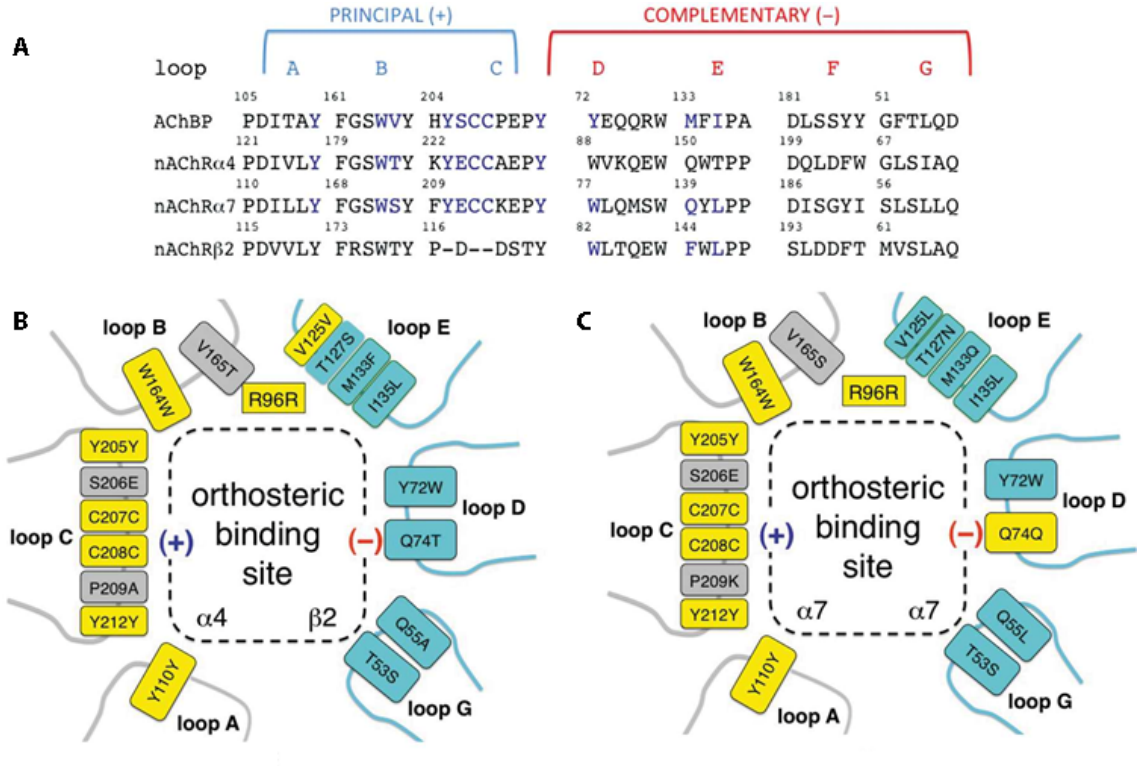


Figure 3.6: Comparison of the key interacting residues in the orthosteric binding site with *Ac*AChBP and the α_4 β_2 and α_7 nAChRs.

A: Alignment of selected sequence segments that form the critical loops on the principal (+) and complementary (-) sides of the orthosteric binding site in *Ac*AChBP and α_4, α_7 and β_2 nAChRs. Residues coloured blue are essential and are discussed in the text. Val125 and Thr127 of *Ac*AChBP are in the N-terminal section of loop E but are left out for clarity. **B:** Schematic of the orthosteric binding site showing the key residues of *Ac*AChBP and the corresponding residues in the α_4(+)β_2(−) nAChR heteromeric site. Substituted residues on the (+) side are in grey and in cyan for the (-). Yellow boxes indicate strict conservation. Loop F has been omitted as it does not form interactions with the compounds. **C**: Schematic for α_7 nAChR homomeric site using the same colour notation as in (**B**).

Overlaying the *Ac*AChBP structures on the nAChR structures (not shown) further suggests that the azabicyclo[2.2.1]heptane moiety and the fluoropyridine ring maintain similar orientations and interactions in both the *Ac*AChBP and nAChR complexes. Of additional note, a similar conserved water molecule is in an analogous place in the α_7 nAChR epibatidine complex to that in the *Ac*AChBP complexes. An analogously positioned water also appears in a structure of the $\alpha_3\beta_4$ nAChR in complex with nicotine (7PV8, Gharpure *et al.*, 2019). While no $\alpha_4\beta_2$ nAChR structures so far have shown solvent molecules in the binding site, a similar ordered water is implied. Approaches using un-natural amino acid

substitution or nicotine analogues with the pyridyl ring replaced with a phenyl to respectively prevent interactions of the complementary face and of the ligand with the putative water have shown analogous several fold increases in the EC50 (Blum *et al.*, 2010).

The conservation of the amino acids which form interactions with the analogues extends beyond the six residues identified by Thompson *et al.* (2017) (Figure 3.6). Val165 of *Ac*AChBP, which participates in van der Waals interactions with the ligands on the principal side, corresponds to Thr183 and Ser172 in the α_4 and α_7 subtypes. These are conservative substitutions in terms of residue size. On loop C, Ser206 is changed to a glutamate in nAChR, but the side chain is directed away from the binding site. Met133 and Ile135 of *Ac*AChBP correspond to Phe144 and Leu146 in nAChR β_2 and Gln139 and Leu141 in α_7 on the complementary side. At the edge of the binding site, Val125 and Thr127 in *Ac*AChBP are Val136 and Ser138 in β_2 and Asn133 and Leu131 in α_7. A prominent change is the Met133Gln substitution, which positions a polar side chain near the ligands in the α_7 nAChR. It is possible that this would alter or maintain the placement of ring B in (1) and (3) by sterically holding the ring in place. Additionally, the glutamine could potentially form a hydrogen bond to the carboxamide substituent when binding (2).

Various suggestions have been made in the literature to explain the structure-activity relationships for agonists and antagonists acting at PLGICs based on AChBP structures (Taly *et al.*, 2009; Lemoine *et al.*, 2012; Bertrand *et al.*, 2015). The residues on the complementary side appear to impart ligand affinity, with cation-π and van der Waals interactions within the aromatic box stabilising the protein-ligand complex. The epibatidine analogues bind within the aromatic box and make interactions as expected for high affinity ligands, so it is likely that their antagonist effects on nicotine-induced antinociception at nAChRs *in-vivo* (Ondachi *et al.*, 2016) is a result of being able to outcompete nicotine for binding.

The conformation of loop C furthermore appears to be of importance for structure-activity relationships at AChBP. Generally, small agonists such as nicotine or epibatidine induce a fully closed and clamped conformation, while bulky antagonists such as strychnine typically induce a fully open conformation. Partial agonists such as varenicline though stabilise an intermediate state. It has

previously been noted that the conformation does not always simply correspond with the ligands pharmacological profile. dihydro-*β*-erythroidine is a notable example of inducing a loop C state similar to that of agonists, despite it being a high potency antagonist (Shahsavar *et al.*, 2012).

It is suggested that loop C for the bound analogues is in the intermediate state, as expected with partial agonists. These ligands are known to demonstrate nearly exclusively antagonist effects at nAChRs, with the exception of (1) and (3) at α_7. It is therefore interesting that loop C demonstrates the same conformation for all of the ligands with *Ac*AChBP. To revisit a previous point, a difference of compounds (1) and (3) with (2) concerns their interactions with Met133, with the latter compound suggested to have less van der Waal interactions compared to the former. This residue is on loop E, which, along with loop D, has been previously linked to contributing to channel opening and desensitisation (Billen *et al.*, 2012). Loop E represents a point of several differences between the $\alpha_4\beta_2$ and α_7 nAChR (Figure 3.6), especially concerning the equivalents of Met133 and Val125 (which also has reduced van der Waals interaction with (2) in AcAChBP due to the position of ring B). In the β2 subunit these are represented by a Val and a Phe respectively, but in the α7 subunit they are a Leu and Glu. It could be speculated therefore that differing interactions between the compounds and loop E in the orthosteric site with the two nAChR forms may be linked to the agonist activity of (1) and (3) at α_7.

On a related note, which also could be linked to the pharmacological differences, the equivalent of Asp94 in *Ac*AChBP is a Lys in the β2 subunit and a Thr in the α7 subunit. (1) and (3) differed from (2) in *Ac*AChBP in being able to directly hydrogen bond with Asp94. The long Lys sidechain of the β2 subunit could offer the potential for forming a hydrogen bond with the ligands, which may not be possible with the shorter Thr sidechain in the α7 subunit.

3.4.0.0: Cytisine derivatives overview

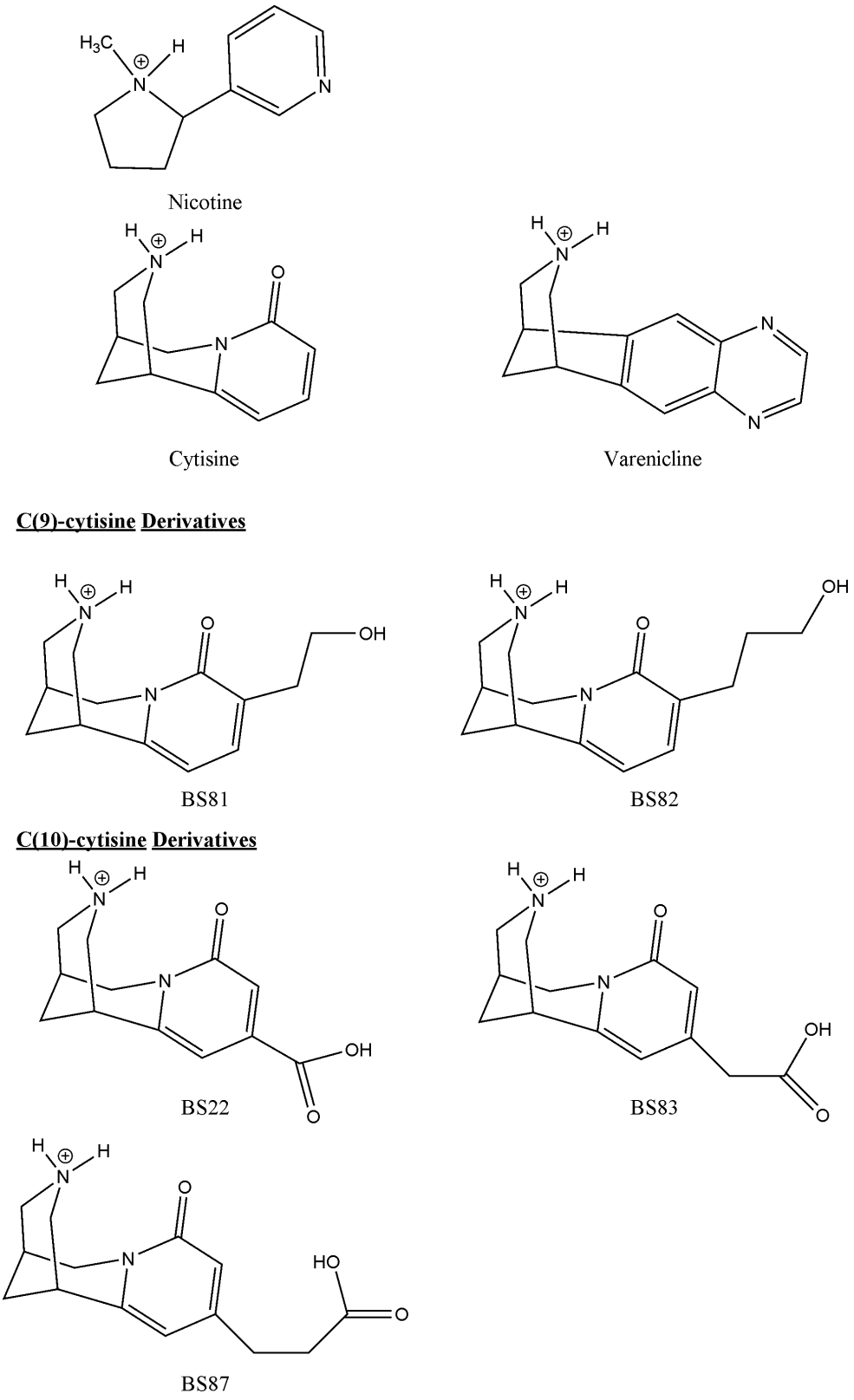


Figure 3.7: Chemical structures of nicotine, cytisine, varenicline and the C(9)- or C(10)-cytisine derivatives.

The protonated form of each ligand is shown as would likely present under physiological conditions.

Nicotine (Figure 3.7) is a natural alkaloid in tobacco leaves which functions as a botanical insecticide (Benowitz *et al.*, 2009). It is the archetypical ligand of the nAChRs, but is most notorious for inducing the addictive effects of tobacco smoking. It is estimated that in 2015 smoking related diseases caused approximately six million deaths worldwide and 150 million disability-adjusted life years, with these estimates increasing year on year (Reitsma *et al.*, 2017).

Nicotine is thought to be addictive as activation of the $\alpha_4\beta_2$ receptor modulates the reward pathway in the brain, leading to the release of dopamine in the nucleus accumbens. Repeated stimulation of dopamine receptors is considered to lead to both a sensitisation of the dopamine response to nicotine and an increase in the number of dopamine receptors in the brain. In the absence of nicotine however, withdrawal symptoms start after approximately 4-6 hours. Aside from a corresponding reduction in the release of dopamine and stimulation of the reward pathway, it is suggested that chronic smokers maintain $\alpha_4\beta_2$ receptors in a desensitised state. Withdrawal symptoms therefore may also be mediated from previously desensitised nAChRs becoming un-occupied and recovering to a responsive state. Smoking cessation agents aim to replicate or replace the effects of nicotine to reduce withdrawal symptoms and allow the brain to eventually revert to a more normal physiology where nicotine is no longer required (Dani and De Biasi, 2001; Foulds, 2006; Benowitz, 2009).

Cytisine (Figure 3.7) is a natural alkaloid expressed in the flowers of *Cytisus laburnum* and related species. It was first noted in 1912 to have similar actions to those of nicotine and is the worlds oldest smoking cessation agent, having been used in former socialist countries in Eastern and Central Europe for over 50 years. It was commercialised as Tabex in these countries, but is relatively unknown elsewhere due to regulatory affairs difficulties, especially regarding a lack of the requisite high quality clinical trials (Prochaska *et al.*, 2013). Small scale trials comparing cytisine to other smoking cessation therapies have suggested that it has an excellent profile in terms of cessation rate, side-effects and economical value. These properties have raised interest in potentially releasing cytisine in newer markets if the regulatory affairs can be satisfied (Hajek *et al.*, 2013; Walker *et al.*, 2014).

Cytisine is structurally related to nicotine, and likewise has agonist activity at neuronal nAChRs, but contrasts in acting as a high affinity partial agonist at the $\alpha_4\beta_2$ nAChR. This has been identified as the hallmark of many clinically successful smoking cessation agents. Partial agonists acting at $\alpha_4\beta_2$ nAChRs are thought to be more successful as they can have a competitive antagonist effect, where nicotine is outcompeted for the orthosteric binding site by the higher affinity partial agonist. Furthermore, when nicotine is absent, partial agonists can reduce withdrawal symptoms by continuing to stimulate the receptor. In either case, only a moderate level of dopamine release is maintained compared to the full agonist actions of nicotine. (Coe *et al.*, 2005; Foulds, 2006; Cahill *et al.*, 2016).

Cytisine is not a perfect molecule however. It also has significant full agonist activity at the $\alpha_3\beta_4$ and α_7 nAChRs, which can induce off target effects (Coe *et al.*, 2005; Rego Campello *et al.*, 2018). Modifications of the cytisine framework have therefore aimed to minimise actions at these additional nAChR subtypes and improve the therapeutic profile. For instance, cytisine provided the lead molecule for the commercially successful smoking cessation agent varenicline (Champix) (Coe *et al.*, 2005).

This section focuses on a set of cytisine derivatives carrying moieties on the C(9) or C(10) of the pyridone ring (Figure 3.2). These had been designed and synthesised by Dr. Hugo Rego Campello in the lab of Professor Tim Gallaher at the University of Bristol. The additional moieties offer a set of potential interactions within the complementary side of the nAChR binding site, where enhanced $\alpha_4\beta_2$ receptor subtype selectivity has been previously noted (Rego Campello *et al.*, 2018). Although un-published binding and functional studies suggested that the presence of these moieties induced changes in selectivity compared to cytisine, it was unclear how the molecular interactions may be altered. *Ac*AChBP was therefore used as a surrogate of the nAChR binding site used for binding and co-crystallography studies with the derivatives.

3.5.0.0: Cytisine derivatives results and discussion

The majority of this section relating to compound BS82 has been published in Davis *et al.* (2020), a copy of which is included in Appendix S2. Contributions by

those other than the author are defined in the text. Selected figures are reproduced or modified from those used in the paper.

3.5.1.0: Cytisine derivatives isothermal titration calorimetry studies

*Ac*AChBP was produced in High Five™ cells with baculovirus using a construct previously designed by Dr. Alice Dawson and purified using procedures as described in section 2.4.1.0. It was attempted to elucidate the thermodynamics of binding and the affinity for the compound set at *Ac*AChBP using ITC. Cytisine appeared to bind with a similar profile to that expected from past literature (Figure 3.8A,C, Table 3.3), with a K_d of 0.6 μM (± 0.3 SE), a ΔH of -15.2 (±1.2 SE) kcal mol^{-1} and a ΔG of -8.6 (±0.3 SE) kcal mol^{-1}. Binding is therefore driven by a favourable enthalpic contribution. This is comparable to results from Rucktooa *et al.* (2012), who fitted a K_d of 1.6 μM, a ΔH of -13.3 kcal mol^{-1} and a ΔG of -8.3 (±0.3 SE) kcal mol^{-1}. The small differences are probably from the use of different buffers - see Celie *et al.* (2004).

Table 3.3: Cytisine and BS82 *Ac*AChBP thermodynamic binding parameters

Ligand	Ligand concentration (μ*M*)	*Ac*AChBP concentration (μ*M*)	No. of sites (N)	K_d (μ*M*)	Δ*G* (kcal mol^{-1})	Δ*H* (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})
Cytisine (n= 2)	100	20	0.6 (0.01)	0.6 (0.3)	−8.6 (0.3)	−15.2 (1.2)	6.6 (1.4)
BS82 (n = 3)	1000	50	0.9 (0.10)	53.3 (19.9)	−6.0 (0.3)	−9.1 (2.5)	3.0 (2.8)

Values in paratheses indicate standard error.

Ligand BS82 unfortunately showed substantially reduced affinity at *Ac*AChBP compared to cytisine (Figure 3.8B, D, Table 3.3). There was a K_d of 53.3 μ*M* (±19.9 SE), a ΔH of -9.1 kcal mol^{-1} (±2.5 SE) and a ΔG of -6.0 (±0.3 SE) kcal mol^{-1}. With cytisine and BS82, there were unfavourable entropic contributions of 6.6 (±1.4 SE) and 3.0 (±2.8 SE) kcal mol^{-1} respectively. Therefore, while the unfavourable entropic contribution is reduced with BS82, the total energy - and hence the affinity - is lower as the enthalpic contribution to binding is also reduced. This is consistent with pyridone-ring modified cytisine derivatives generally lowering the affinity for nAChRs (Rego Campello *et al.*, 2018)

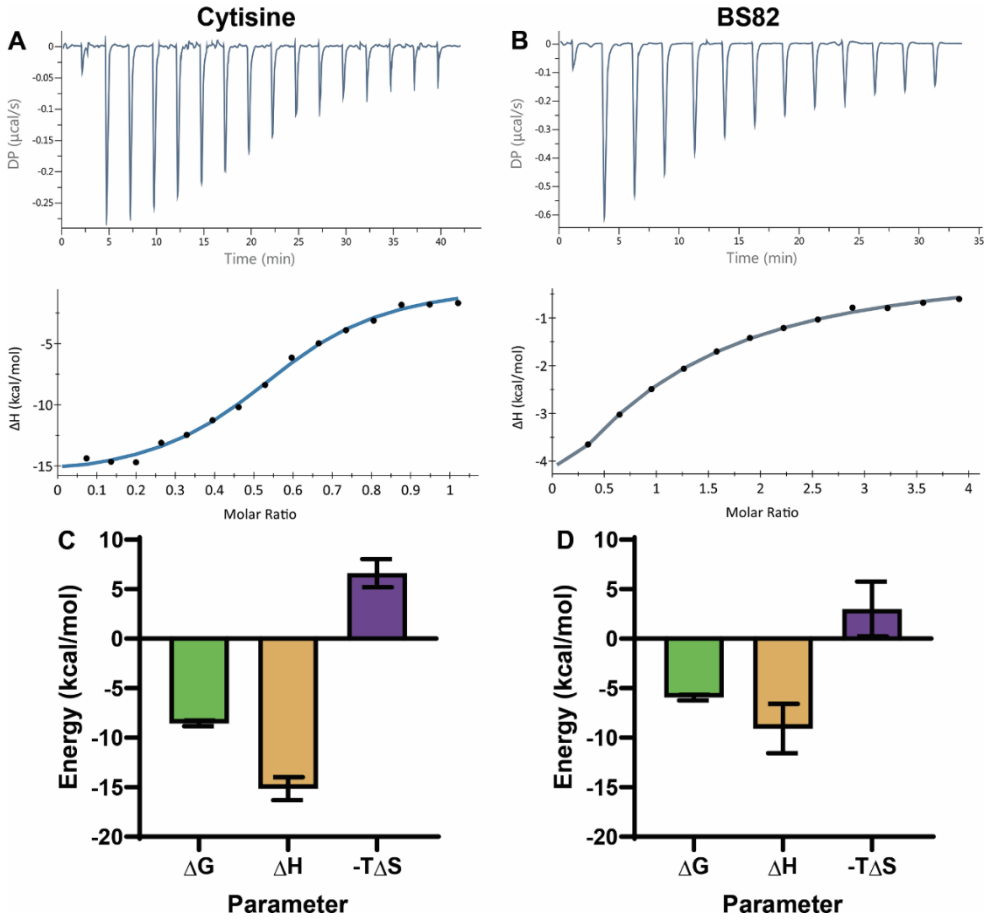


Figure 3.8: Thermodynamics of cytisine and BS82 binding *Ac*AChBP.

A-B: Raw and injection heat-normalised thermodynamic traces for (**A**) cytisine and (**B**) BS82 binding to *Ac*AChBP. The corresponding signature plots are shown in **C-D**, error bars represent standard error. For (**A**) the cell contained 20 µM of *Ac*AChBP and the syringe 100 µM of cytisine, while for (**B**) the cell contained 50 µM of *Ac*AChBP and the syringe 1000 µM of BS82.

It is interesting that cytisine only showed a stoichiometry of ~2.8:1 for interaction, while BS82 displayed a stoichiometry of ~4.5:1. The stoichiometry of BS82 is close to the 5:1 which would be expected with pentameric *Ac*AChBP for a ligand that fully occupies the binding site in a straightforward association. Past ITC studies with *Ac*AChbP have also shown lower than expected molar ratios (Celie *et al.*, 2004) for ligands, although the causes are unclear. Possibilities could include uncertainties in the assumed ligand and/or protein concentrations, the presence of impurities or degradation of the protein/ligand. Modelling by Kantonen *et al.* (2017) suggests that in ITC experiments uncertainty in the N value is particularly affected by errors in the injection heat recorded and the concentration of the titrant in the syringe.

None of the rest of the compound set could be successfully analysed with ITC using standard multiple injection protocols. To establish the effect of each ligand

on the heat change, single injection protocols were therefore tested to determine the ΔH. These showed that none of the additional compounds appeared to induce a heat change (Figure 3.9). This indicates that they do not bind, or either the affinities or heat changes upon binding are too low for ITC to be feasible. It was attempted to analyse some of the derivatives using tryptophan fluorescence quenching assays, but these were unsuccessful as the ligands appeared to show significant autofluorescence at 340 nm when excited at 280 nm. Future efforts therefore concentrated on BS82. It is interesting though that previous competition radioligand binding studies at nAChRs (Table 3.4) have shown all of these analogues to bind to nAChR subtypes and the K_i for cytisine at the α_7 receptor is surprisingly close to that obtained for *Ac*AChBP with ITC.

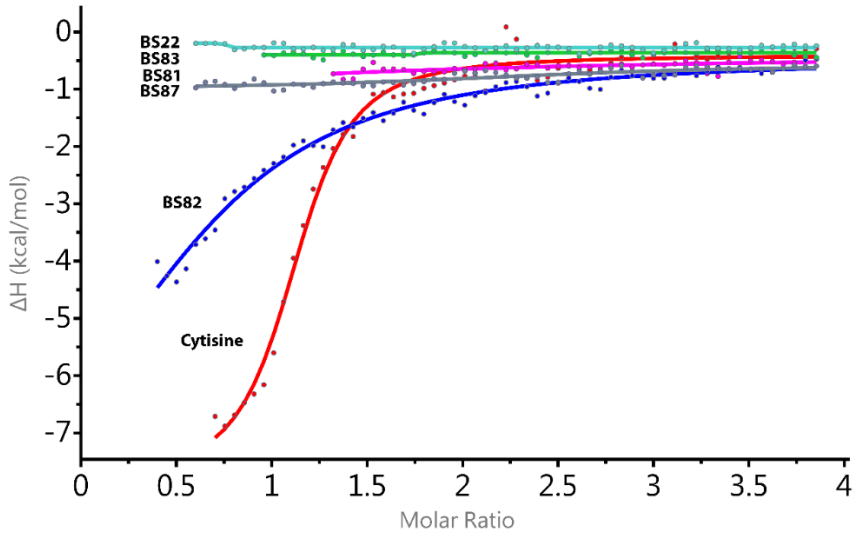


Figure 3.9: Single injection heat-normalised thermodynamic traces for the binding of cytisine and the compound set at AcAChBP.

For each titration, the cell contained 25 μM of *Ac*AChBP and the syringe 500 μM of the respective ligand.

Table 3.4: Cytisine derivatives competition binding assay affinity values at human nicotinic acetylcholine receptors.

Ligand	$\alpha_4\beta_2$ (K_i)	$\alpha_3\beta_4$ (K_i)	$\alpha_4\beta_2/\alpha_3\beta_4$	α_7 (K_i)	$\alpha_4\beta_2/\alpha_7$
Cytisine	3	473	158	1264	421
Varenicline	0.46	171	372	75	163
BS22	39.5	2281	58	45380	1149
BS81	4.3	9205	2141	52380	12181
BS82	7.9	10380	1314	4156	526
BS83	137	30460	222	51170	374
BS87	14.3	ND	ND	7000	490

All K_i values in nM at the human form of the relevant receptor. Unpublished data kindly provided by Professor Tim Gallagher, the assays were performed by Cecilia Gotti (CNR, Neuroscience Institute-Milano, Biometra University of Milan, Milan, Italy). ND, not determined.

3.5.2.0: *Ac*AChBP:BS82 crystallographic studies

It had been demonstrated that BS82 could bind *Ac*AChBP, although with much reduced affinity compared with cytisine. BS82 was therefore co-crystallised with *Ac*AChBP to elucidate how the C(9)-3-hydroxylpropyl moiety may have changed the molecular interactions and caused the reduction in affinity. It proved to not be possible to directly co-crystallise the protein-ligand complex, but a random matrix micro-seeding approach using cytisine bound *Ac*AChBP crystals as a seed stock eventually led to identification of appropriate conditions for obtaining large, high-quality crystals (Figure 3.10, see section 2.9.1.0 for a detailed description of the crystallisation methods). At Diamond I03 an *Ac*AChBP-BS82 crystal (grown in 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 10% glycerol, 0.1 M HEPES pH 7.0 with the drop containing 1.5 µl 12.5 mg/ml protein, 0.3 µl seed stock and 0.2 µl reservoir) diffracted to 1.48 Å resolution and was processed to obtain a final model at 1.72 Å resolution. Crystallographic statistics for the *Ac*AChBP-BS82 complex diffraction and model are in Table 3.5.

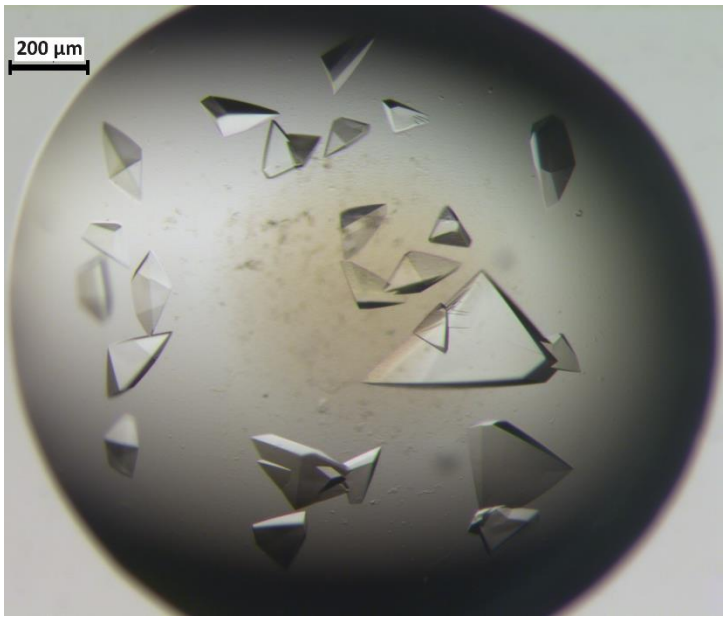


Figure 3.10: *Ac*AChBP-BS82 crystals.

Protein at 12.5 mg/ml starting concentration was vapour diffusion crystallised as hanging drops with a reservoir containing 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 10% glycerol, 0.1 M HEPES pH 7.0. The drop contained 1.5 µl 12.5 mg/ml protein, 0.3 µl seed stock and 0.2 µl reservoir. The largest crystal was diffracted at Diamond I03 and provided the dataset for the model ultimately deposited in the PDB as ID 6T9R (Davis *et al.*, 2020).

Table 3.5: *Ac*AChBP-BS82 crystallographic statistics

PDB Code	6T9R
Unit cell *a*, *b*, *c* (Å) α, β, γ, (°)	209.5, 132.9, 131.1, 90.0, 102.5, 90.0
Space group	*C2*
Wavelength (Å)	0.976
Subunits/asymmetric unit	10
Resolution range (Å)	127.97 – 1.72 (1.75 – 1.72)
Total number reflections	1225936 (25553)
Unique reflections	353432 (12162)
Redundancy	3.5 (2.1)
R_{merge}	0.07 (0.82)
R_{pim}	0.064 (0.794)
Wilson *B*-factor (Å^2)	25.07
Completeness (%)	95.5 (66.6)
$<I/\sigma(I)>$	7.6 (0.6)
CC(1/2)	0.993 (0.435)
Refinement	
Resolution Range (Å)	90.12 – 1.72
R_{work}/R_{free}[f] (%)	16.2/19.2
Number reflections for R_{work} / R_{free}	335977/17388
Number of protein residues	2058
Number of NAGs	10
Number of **BS82** molecules	10
Number of phosphate molecules	10
Number of glycerol molecules	40
Number of water molecules	2899
Number of chloride ions	10
Number of potassium ions	10
RMSDs	
Bonds (Å) / Angles (°)	0.013/1.84
Ramachandran plot	
Residues in favoured regions (%)	98.77
Residues in allowed regions (%)	1.23
Mean B-factors (Å^2)	
Protein atoms by subunit A-J	30.1/29.2/29.7/28.4/30.6/29.8/31.5/31.8/33.4/34.5
NAGs by subunit A-J	93.7/90.2/108.6/87.6/99.5/97.1/104.2/103.6/100.2/92.6
Water molecules	46.0
Ligand **BS82** molecules A-J	28.1/27.9/23.5/22.2/27.9/29.9/23.7/32.4/33.3/31.5
Phosphate ions A-J	58.1/54.2/57.4/63.8/49.2/49.5/56.4/71.9/52.6/58.0
Glycerol molecules	59.9
Chloride ions A-J	33.7/34.6/32.0/32.9/33.5/34.0/37.9/35.2/35.9/37.8
Potassium ions A-J	37.3/37.4/38.6/40.0/39.5/40.5/41.3/40.5/43.2/44.2

Figures in paratheses are for the highest resolution shell

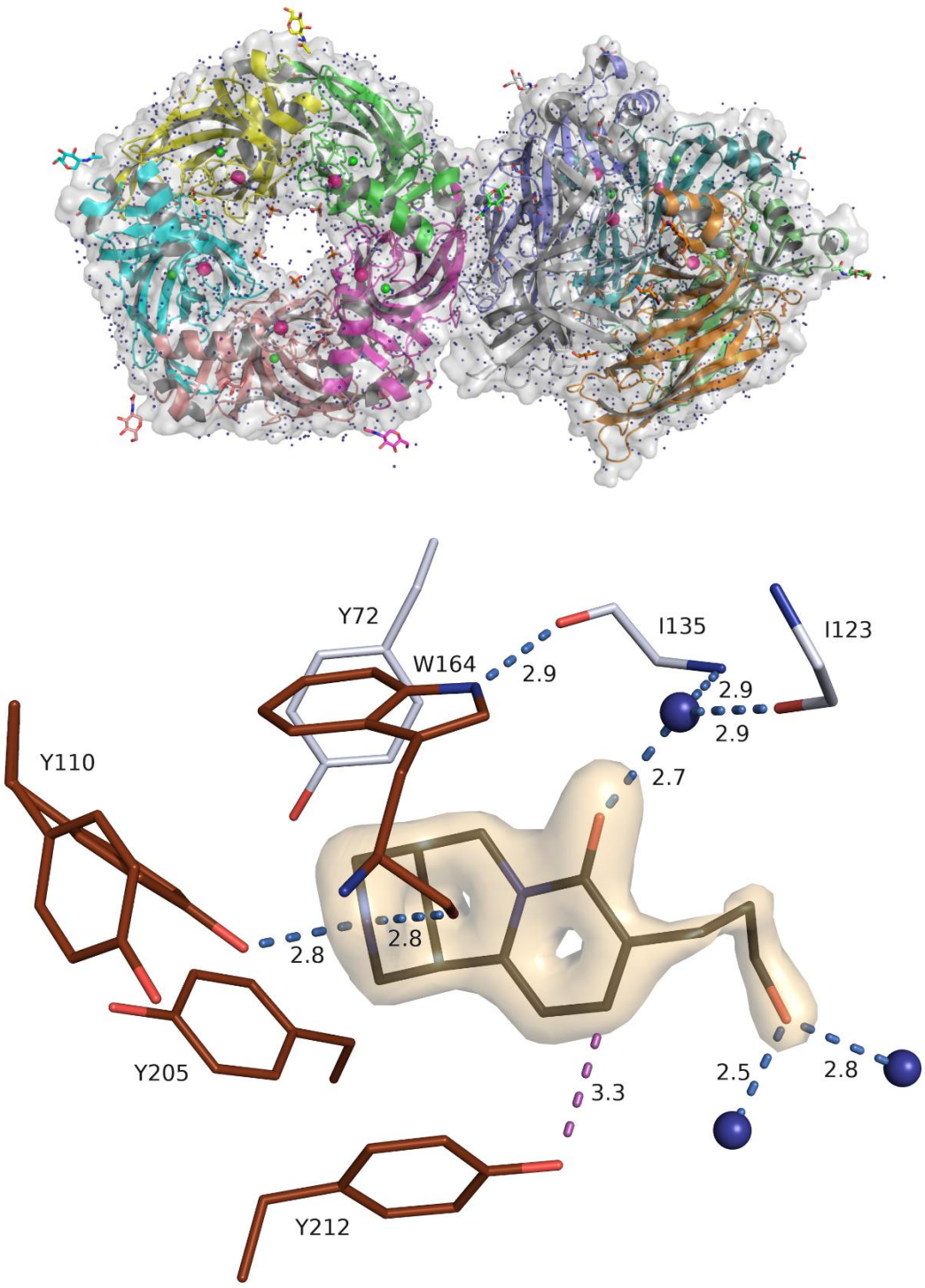


Figure 3.11: Crystal structure of the *Ac*AChBP-BS82 complex and BS82 in a representative binding site with an OMIT map.

A: Cartoon representation with surface overlay of the two pentamers in the asymmetric unit. Each subunit is separately coloured. Water molecules and Cl^- and K^+ ions are shown as dark blue, green and pink spheres respectively. Phosphates are shown as orange sticks and ligand BS82 as black sticks. Glycerols and NAGs are shown as sticks coloured per subunit. **B:** Interactions of ligand BS82 in a representative binding site with an OMIT difference density map for BS82 contoured at 6.5 σ. Residues on the principal side of the binding site are shown in brown, in grey for those on the complementary side and those for ligand BS82 are in black. Three water molecules are depicted as blue spheres. Potential hydrogen bonds are shown as blue dashed lines and a possible C—H···O interaction is designated as a magenta dashed line. Two rotamers for Tyr110 are shown. Distances are in Å.

The monoclinic crystal contained two pentamers in the asymmetric unit (Figure 3.11A), with a high degree of non-crystallographic symmetry even when the NCS restraints were turned off during the later stages of refinement. The ligand could be unambiguously placed in all ten binding sites, aided by the well-defined electron density (Figure 3.11B). Each of these refined with average B-factors close to or less than those for their respective sub-units (Table 3.5). In addition to BS82, phosphate, glycerol and water molecules, the high resolution of the structure also allowed placement of ion molecules. Difference-density maps suggested that density originally assigned as water molecules at the same two positions (outside of the binding site) in each subunit contained additional electrons. This therefore led to the suspicion that they represented ions. Based on the electron density, components of the crystallisation mix, the surrounding chemical environment and the co-ordination chemistry, a Cl^- ion (average B-factor 34.7 $Å^2$) and a K^+ ion (average B-factor 40.3 $Å^2$) were eventually placed in these positions in each sub-unit. Looking at other *Ac*AChBP related structures, a Cl^- ion has been modelled in an analogous position in two (PDB IDs 5OAN, 5OBH, Dawson *et al.*, 2019), although others have more commonly positioned a water molecule instead. No other structures could be found presenting a K^+ ion in a similar position.

The interactions of BS82 in the AChBP binding site are similar to those of cytisine, varenicline and other nAChR ligands (Billen *et al.*, 2012; Rucktooa *et al.*, 2012; Dawson *et al.*, 2019). The Tyr110 hydroxyl group and the carbonyl backbone of Trp164 accept hydrogen bonds from the secondary amine of the BS82 piperidine ring. Cation-π interactions are suggested between the secondary amine and the aromatic groups on Trp164 and Tyr212 based on the orientations and the distances between 4.2 and 4.6 Å. Tyr205 and Tyr212 furthermore interact via van der Waals with the pyridone ring and the Ty212 hydroxyl group may form a potential C—H···O interaction.

The resolution was high enough to place alternate conformations for selected residues throughout the protein and intriguingly in the binding site a dual rotamer for the sidechain of Tyr110 could be placed in seven out of the ten subunits (example in Figures 3.11 and 13.12). The main rotamer with 100% occupancy in three of the subunits, and between 60-80% for the other seven, makes the

contacts described above. While two of these subunits adopt an additional conformer which only imparts a minor difference and therefore makes similar interactions, the other five subunits adopt a rotamer with 20-40% occupancy where the sidechain is orientated at too great a distance for hydrogen bonding with the secondary amine of BS82. This rotamer is also adopted in the apo-form of *Ac*AChBP and can for instance be seen in the empty binding sites of PDB ID 5SYO (Bobango *et al.*, unpublished). These dual rotamers for Tyr110 are consistent with the orthosteric binding site of AChBP and PLGICs having a degree of flexibility, especially in loop C (Dawson *et al.*, 2019), to accommodate ligands of various sizes and properties. This also suggests that may be some degree of conformational freedom in parts of the aromatic cage.

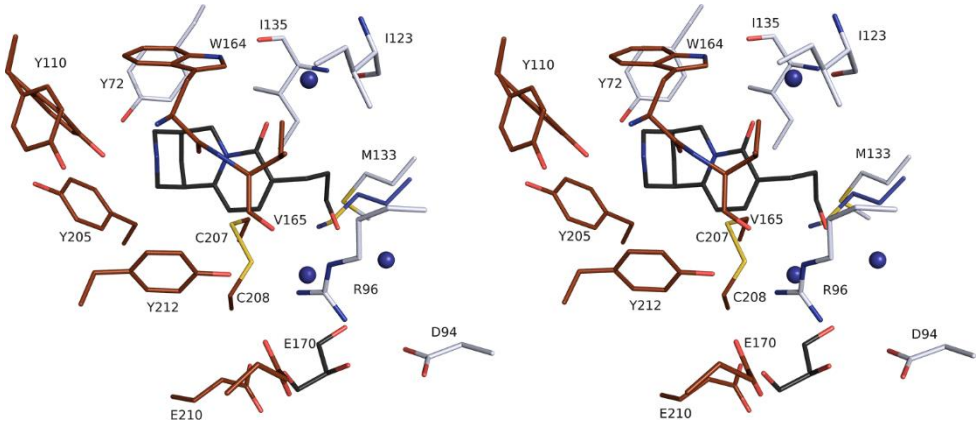


Figure 3.12: Stereo view of the interactions of BS82 in a representative AcAChBP binding site.

Residues on the principal side of the binding site are shown in brown, in grey for those on the complementary side and those for ligand BS82 and glycerol are in black. N and O atoms are coloured blue and red, respectively. M133 from PDB ID 5SYO is overlaid and shown as dark blue sticks. Three water molecules are depicted as blue spheres. Two rotamers for Tyr110 are shown.

The pyridone carbonyl group forms a hydrogen bond with a water molecule which bridges to the carbonyl and amide of residues Ile135 and Ile123 respectively on the complementary side of the binding site. It is also within distance (~3.5 Å) to form a weak hydrogen bond with NE1 of Trp164, but the angle is not appropriate for bonding. As discussed previously with the epibatidine analogues in section 3.3.2.0, a water molecule is highly conserved in AChBP and nAChR structures at this point and appears to partially mediate ligand affinity and possibly mode of action (Billen *et al.*, 2012; Zhang *et al.*, 2012). The only *Ac*AChBP-Cytisine structure in the PDB (PDB code 4BQT, Rucktooa *et al.*, 2012) is of low (2.9 Å) resolution and no water molecules are present in the binding site. However, a

higher resolution (2.0 Å) AChBP:Cytisine structure (PDB code 5SYO, Bobango *et al.*, unpublished) with the protein engineered to possess loop C from the human α_3 nAChR does show a water in an analogous position. This water molecule is hydrogen bonded with the cytisine carbonyl and the complementary side residues as per the BS82 complex structure.

Van der Waals forces between the rings of BS82 and the protein are observed with Trp164, Tyr72, Tyr205, Tyr212, Ile135 and Val165 and additionally the Cys207-Cys208 disulphide bond. The 3-hydroxylpropyl moiety on C(9) interacts with Val125, Ile135, the disulphide between Cys207-Cys208, the Met133 side chain and the backbone of Phe134. The hydroxyl residue however does not form any hydrogen bonds with residues in the binding site and in all ten subunits it is orientated away, and hydrogen bonded to two well-ordered water molecules which interact with other water molecules and a glycerol. This glycerol occupies a highly polar pocket formed by Ap94, Arg96, Glu170 and Glu210 side chains (Figure 3.12).

A factor in the design of BS82, and also BS81, was whether the C(9) moieties could possibly influence or exploit waters forming hydrogen bonding networks associating with the protein. It was additionally considered whether the substituents could bond with the pyridone ring carbonyl to “mimic” a bridging water. In the case of BS82 however, compared to cytisine the additional 3-hydroxylpropyl was associated with an ~90x reduction in affinity to *Ac*AChBP and additionally varying degrees of reduction in affinity to nAChR subtypes. Comparing the conformation of the 3-hydroxylpropyl to structures in the Cambridge Crystallographic Database with the program TorsionAnalyzer (Schärfer *et al.*, 2013) suggests that it has adopted a “staggered” conformation in each binding site. This is the most common conformation in the database and implies that the moiety has formed a preferential low energy conformation.

Despite this, the affinity to *Ac*AChBP was reduced. Comparing the *Ac*AChBP-BS82 structure to those bound with cytisine, it became apparent that the orientation of the Met133 side chain differed. In cytisine structures, the rotamer directs the sidechain into the binding site to form van der Waals interactions with the ligand, but with BS82 the rotamer is forced away by the added 3-

hydroxylpropyl to avoid steric clash (Figure 3.12). This could influence the thermodynamic profile by inducing strain into the protein structure, along with the moiety potentially influencing the bonding pattern of water molecules in the binding site. It has been previously demonstrated with the $\alpha_4\beta_2$ nAChR that disrupting the bonding between a ligand or the complementary side residues and the conserved water is associated with a several fold reduction in the EC50 (Blum *et al.*, 2010). Overall, this raises the point that modification of the cytisine framework needs consider the implications of possible unfavourable interactions in human nAChRs if a large and flexible sidechain is added which interacts within this part of the binding site.

3.4.0.0: General discussion

In this chapter, *Ac*AChBP was used as a surrogate of the nAChR orthosteric binding site for derivatives of epibatidine and cytisine. Binding studies elucidated the affinity of the ligands for AcAChBP and the molecular interactions of the ligands were identified with X-ray crystallography. However, while AChBP offers a powerful tool to obtain this information without the difficulty and expense of working with membrane bound nAChRs or other PLGICs, it needs to be recognised that working with a surrogate imparts certain limitations on any conclusions which can be drawn. While the key residues comprising the aromatic cage are conserved, AChBPs do not directly represent the orthostatic binding site of any nAChR subunit and there are especially differences on the complementary side. Indeed, in the data presented in this chapter there does not appear to be any form of correlation between the ligand affinities at the $\alpha_4\beta_2$ nAChR and *Ac*AChBP. The protein also lacks transmembrane domains, so it is difficult to relate binding to how a ligand may ultimately gate a PLGIC.

It could be argued that AChBP has outlived its usefulness now that there are an increasing number of structures for the nAChR family and other PLGICs, with the general structure of the ECD and orthostatic binding site being well established. This is especially pertinent with the availability of modern *in-silico* docking software which can accurately predict the binding mode of ligands and reduces the need for obtaining structures. Data in the PDB does suggest that the number of AChBP structures being deposited has declined in recent years.

However, aside from the easier production and crystallisation, AChBP does still have certain advantages which are unlikely to be replicated in the near future with full length membrane proteins. ITC for instance is difficult to use with membrane proteins due to uncertainties in matching detergent concentrations in the cell and titrant, plus the high level of protein needed is generally unfeasible (Rajarathnam and Rösgen, 2014). However, ITC is the only way to quantify the thermodynamics of ligand binding in a single experiment and the use of ITC with AChBP is well established (Celie *et al.*, 2004; Rucktooa *et al.*, 2012).

Additionally, there is still only a small number of nAChR and PLGIC structures available in the PDB. Most of these structures are limited to the common subunit combinations and are also bound with typical agonist ligands such as nicotine or epibatidine. Many are also only available in a desensitised state, which limits their usefulness for simulating the binding of antagonists and other ligand types. The resolutions are also typically between 3.0-4.0 Å, whereas AChBP can offer routine high-resolution structures of ~2 Å or greater. There has additionally been a trend in recent years of using engineered AChBP proteins which are intended to more accurately reflect the binding site of human PLGICs and appear to offer an improved correlation (Li *et al.*, 2011; Kesters *et al.*, 2013; Shahsavar *et al.*, 2015; Dawson *et al.*, 2019). It is therefore anticipated that AChBP and modified derivatives will remain a useful tool in the future for binding and co-crystallography studies.

Chapter 4: Identification of Conditions for Small Scale Production of Resistance to Dieldrin

4.1.0.0: Overview

Insecticide resistance is a growing problem worldwide for both agriculture and public health. For example, the World Health Organisation has expressed concern over mosquitos in malarial areas becoming resistant to multiple insecticides of different classes (Sternberg and Thomas, 2018; World Health Organisation, 2018). As was previously described in Chapter 2, RDL represents a target for insecticides which is present in a broad variety of insect species. However, because of their potent toxicity and negative environmental effects, the majority of RDL targeting agents have been withdrawn for many years. The widespread presence of the A2' mutation on TMH2 also causes many insect species to be resistant to the effects of these agents. However, if environmentally safer insecticides that bypass the A2' mutation can be developed, RDL could represent a renewed target of opportunity, as it has not been greatly exploited since the withdrawal of organochlorine agents.

Fipronil is a notable example of a second generation RDL targeting, agent, which is approved for various agricultural and veterinary indications. The A2' mutation does reduce fipronil sensitivity, although not to as great an extent as with organochlorines. Other mutations in sites distinct from the second TMH have been associated with generating further resistance though (Nakao, 2017; Taylor-Wells and Jones, 2017). The use of fipronil is highly restricted in many countries due to concerns about its effects on bees and possible role in "colony collapse" disorder (Zaluski *et al.*, 2015).

Recently, interest in RDL as an insecticide target has been renewed with the development of the novel meta-diamide and isoxazoline classes of insecticides (Figure 4.1). These target insect GABA receptors but are unaffected by the RDL A2' mutation as they act as non-competitive antagonists at a distinctive site. Several agents of these classes are now on the market for selected agricultural and veterinary applications. The meta-diamide broflanilide and the isoxazolines afoxolaner and fluralaner also have high knockdown efficiency against

mosquitos, so they could be useful for disease vector control (Nakao and Banba, 2016; Miglianico *et al.*, 2018; Lees *et al.*, 2020).

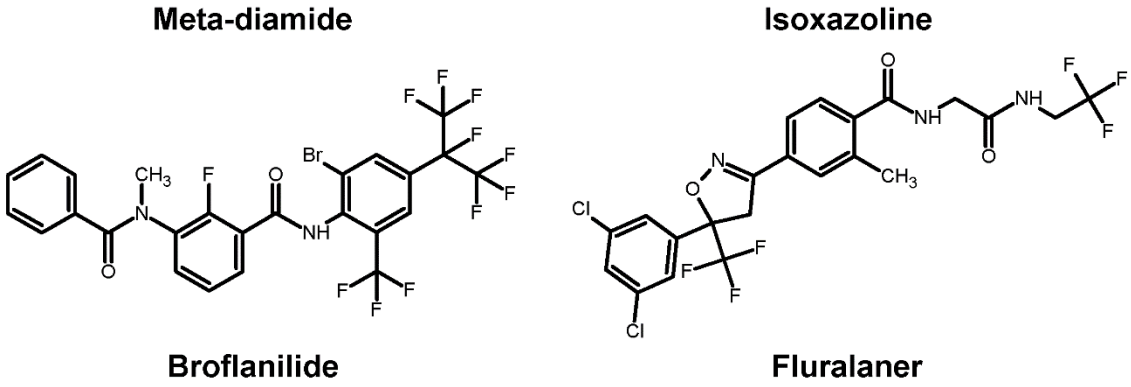


Figure 4.1: Chemical structures of representative meta-diamide and isoxazoline insecticides.

Meta-diamide and isoxazoline agents also appear to have a comparatively low affinity to mammalian $GABA_A$ and glycine receptors compared to previous RDL targeting agents, making them potentially less hazardous to use (Nakao, 2015; Miglianico *et al.*, 2018). While these new classes represent a breakthrough, continuous development of newer agents will be needed to combat resistance inevitably arising eventually. Site directed mutagenesis studies on *Drosophila* cell lines have already identified substitutions which reduce the broflanilide IC_{50} (Nakao and Banba, 2016).

The RDL M3-M4 intracellular loop is of particular interest as a target for insecticides, as it frequently contains species unique differences. Therefore, it has been proposed that exploiting these differences in RDL structure could lead to the development of insecticides which target pest insects, but spare beneficial ones such as bees (Taylor-Wells *et al.*, 2017).

Despite the advances in *in-vitro* screening for ligand development, whole organism 50% lethal dose testing with subsequent identification of the mode of action remains the method of choice for insecticide development. It has been previously suggested that applying biochemical and biophysical screening techniques used in small molecule drug development for ion channels could lead to newer classes of insecticides (Wing, 2020). Obtaining a recombinant source of RDL for use in biophysical assays such as SPR and with X-ray crystallography or cryo-EM structural studies would be of further benefit in the design and study of potential insecticides.

The recent advancements in protein production and cryo-EM have led to unprecedented levels of information on the structural organisation of PLGICs over the past few years and it is therefore seen as feasible to aim to produce recombinant RDL and obtain a structure. The key points regarding recombinant protein expression systems, extraction of membrane proteins and protein structural biology techniques will therefore be reviewed first to give context to the factors which need to be considered.

4.1.1.0: Recombinant protein expression systems

Structural biology has immeasurably contributed to understanding of how proteins are organised and function. In the early days though, it was necessary to obtain proteins of interest from natural sources. For this reason, early structures were of proteins from readily accessible resources and examples include sperm whale myoglobin (Kendrew *et al.*, 1958) and porcine insulin (Adams *et al.*, 1969). While protein obtained in this fashion is in its most native form, the main disadvantage is the requirement of having large amounts of material to extract from. Purification is also complicated and inefficient as the proteins are untagged. The introduction of recombinant protein production was a major factor in the growth of the number of structures being deposited in the PDB year by year (Zhou *et al.*, 2016; Zhu and Gouaux, 2021).

This section aims to briefly review the main advantages and disadvantages of the four main systems used for recombinant protein production, these are namely *Escherichia coli*, yeast, insect cells and mammalian cells. The key points are summarised in Table 4.1.

4.1.1.1: Escherichia coli

As per the PDB, in terms of overall structures deposited the microbe *E. coli* is the most common system for recombinant protein production (Berman *et al.*, 2003). *E. coli* have multiple advantages for producing proteins. They have simple requirements for culturing, rapidly grow to high densities and are straightforward to induce to produce protein. In terms of media requirements and consumables *E. coli* are comparatively inexpensive compared to other methods. Many different strains are available to meet the requirements of different proteins, although

derivatives of the BL21 strain are most commonly used for protein expression (Francis and Page, 2010; Dilworth *et al.*, 2018).

However, *E. coli* have several disadvantages which makes them unsuitable or less appropriate for expression of certain proteins. Insoluble inclusion bodies frequently form due to high levels of mis-folded protein and the probability of obtaining soluble protein decreases with protein size, especially dropping past ~60 kDa (Rosano and Ceccarelli, 2014). Certain genes can also be problematic if they code for products that are toxic to the cells. Disulphide bond formation can be challenging and the majority of *E. coli* strains only have a limited or no ability to make eukaryotic post-translational modifications such as glycosylation and phosphorylation. These are frequently important for the folding and function of eukaryotic proteins (Wagner *et al.*, 2006; Rosano and Ceccarelli, 2014; Dilworth *et al.*, 2018; Birch *et al.*, 2020).

4.1.1.2: Yeast

Yeasts represent protein production systems with similarities to *E. coli,* but, being eukaryotic cells they have enhanced protein production capabilities for eukaryotic samples. Several types of yeast can be used for protein production, but the most common systems are *Pichia pastoris* and *Saccharomyces cerevisiae*. Yeasts possess many of the same advantages as *E. coli*, being relatively simple to culture and having inexpensive costs compared to higher eukaryotic cells. Yeast can efficiently make proteins up to ~150 kDa and be rapidly grown to much higher densities than other Eukaryotic systems. Furthermore, yeast possess more effective protein quality control mechanisms than *E. coli*, so lower proportions of protein are typically mis-folded. They are also able to implement eukaryotic PTMs on proteins, including glycosylating residues (Bill, 2014; Dilworth *et al.*, 2018; Birch *et al.*, 2020).

Yeasts have disadvantages though. With the *Pichia* system extensive screening is required to find a high expressing clone. Yeasts also differ from other eukaryotic systems in using sterols rather than cholesterol, therefore proteins may be poorly produced or inactive if cholesterol is important for their folding. Hyper glycosylation where N-linked glycosylation sites are excessively elongated with mannose residues can also be a problem as this may affect protein function (Tate *et al.*, 2003; Conde *et al.*, 2004; Dilworth *et al.*, 2018; Birch *et al.*, 2020).

4.1.1.3: Mammalian cells

Various mammalian cell lines can be used for recombinant protein production, with HEK293 and Chinese Hamster Ovary (CHO) cells being examples which are frequently used. Chemically mediated transient transfection allows rapid production of protein, and mammalian cells offer the most similar conditions to the native environment of mammalian proteins, with identical or similar PTMs furthermore being added. Mammalian cells can efficiently produce large, complex proteins and efficient quality control mechanisms are present. Therefore, typically higher levels of protein are correctly folded and functional and overall quality is improved compared with other systems (Khan, 2013; Thomas and Tate, 2014; Birch *et al.*, 2018)

Mammalian expression has several drawbacks. Aside from requiring strict aseptic technique to avoid contamination with bacteria or yeast, mammalian cells also generally need a CO_2 atmosphere for growth. The overall costs of media and consumables for mammalian expression are much greater compared to other systems and on larger scales these can quickly become prohibitive. Although protein quality is usually higher with mammalian production, frequently this advantage is offset by lower overall yields of protein; low yields can especially be a problem when using transient transfection. An alternative which can give higher yields of protein is to develop a cell line stably expressing the gene of interest. However, it is difficult and time consuming to develop stable cell lines and constitutive stable transfection may be detrimental to cell health. The virus mediated approaches which are increasingly used for making stable cell lines additionally have safety concerns as they have the potential to infect human cells (Tate *et al.*, 2003; Thomas and Tate, 2014; Schlimgen *et al.*, 2016; Birch *et al.*, 2020).

4.1.1.4: Insect cells

Various insect cell lines can be used for recombinant protein production, with the most used being *Sf*21, *Sf*9 and High Five™ cells. Insect cells have multiple advantages for protein production. Compared to mammalian cells, insect cells can be grown at much higher densities and do not require a CO_2 atmosphere. While insect media costs are higher than for *E. coli* or yeast, they are lower than for mammalian cells (Birch *et al.*, 2018; Furukawa *et al.*, 2021). The most

common method used for protein production in insect cells is baculovirus transduction, where the gene of interest is cloned into a recombinant insect virus. Baculovirus offers a cost-effective method for large scale production and typically results in high levels of protein being produced. Insect cells can furthermore readily fabricate large proteins and multi-subunit complexes with no apparent size limit. Scaling up production is straightforward as higher volumes can be simply infected with baculovirus at the same multiplicity of infection. Insect cells have a lipid profile closer to mammalian cells than yeast and can also make most eukaryotic PTMs, including glycosylation. Quality control mechanisms are furthermore enhanced compared to *E. coli* or yeast. These factors make them potentially suitable for the expression of a wide variety of proteins when lower organisms are unsuitable (Contreras-Gómez *et al.*, 2014; Birch *et al.*, 2018; Furukawa *et al.*, 2021).

Insect cells however have disadvantages for protein production. Like mammalian cells they require strict aseptic technique to avoid contamination. It can also take significant time to raise baculoviruses for protein production. Additionally, while insect cells can glycosylate residues, the pattern differs from those of mammals, and they are typically unable to make higher complexity glycans. This can affect protein function or folding if glycosylation is important. Compared to mammalian cells, insect cells also have less efficient quality control processes. Therefore, although high levels of a protein may be produced, large proportions can be mis-folded or incompletely processed and result in low yields being ultimately obtained. A particular problem can be with residues being incompletely glycosylated (Tate *et al.*, 2003; Thomas and Tate, 2014; Birch *et al.*, 2020).

Table 4.1: Advantages and disadvantages of the main systems used for protein production.

System	Advantages	Disadvantages
Escherichia coli	• Simple to culture • Inexpensive media and consumables • Typically high levels of protein expression	• Typically inefficient with larger proteins (>50-60 kDa) • Inclusion body formation from high levels of un-folded protein • Toxicity from certain products • Limited to no ability to make eukaryotic post translational modifications.
Yeast	• Simple to culture • Inexpensive costs compared to higher Eukaryotic cells. • Can make Eukaryotic post translational modifications • Can efficiently express larger proteins than *Escherichia coli* (up to ~150 kDa)	• Hyper-glycosylation • Use sterols rather than cholesterol. • Prolonged screening to find a high expressing clone with the *Pichia pastoris* system
Insect Cells	• Moderately expensive media and reagents • Can make larger proteins/complexes • Can make Eukaryotic post translational modifications • More similar lipid profile to mammalian cells. • High levels of protein expression with baculovirus infection. • Simple to scale up.	• Require strict aseptic technique • Typically unable to make higher complexity glycans • Raising baculoviruses takes time. • Potentially high levels of mis-folded or incompletely processed protein.
Mammalian Cells	• Most similar system to native mammalian cells • Can make large proteins/complexes • Native or similar post translational modifications • Efficient quality control • Rapid protein production with transient transfection • Proportionally high levels of correctly folded proteins	• Expensive media and reagents • Lower overall yields • Require strict aseptic technique • Require a CO_2 atmosphere

Adapted from references in the text for each expression system.

4.1.2.0: Extraction of membrane proteins

4.1.2.1: The fluid mosaic membrane model

Compared to other types of proteins, membrane proteins have the additional requirement of needing to be extracted from the plasma membrane or other cellular membrane of relevance. As described by the fluid mosaic membrane model (Figure 4.2), biological membranes are formed by a phospholipid bilayer combined with proteins, carbohydrates and other lipids which dynamically move laterally within the bilayer. While the original model assumed a uniform lipid

bilayer with floating proteins randomly inserted, it has since been updated to take into account clusters of ordered lipids within the disordered membrane. These are referred to as lipid rafts, which are membrane microdomains rich in sphingolipids and cholesterol that laterally interact with and order other proteins and lipids (Sonnino and Prinetti, 2010; Lombard, 2014).

While peripherally associated membrane proteins that do not span the bilayer can be usually dissociated using high pH and/or high salt buffers, transmembrane proteins are much harder to release and require detergents or other solubilising agents which remove lipid molecules. After extraction these agents attempt to mimic the native lipidic environment and stabilise the protein (Smith, 2017).

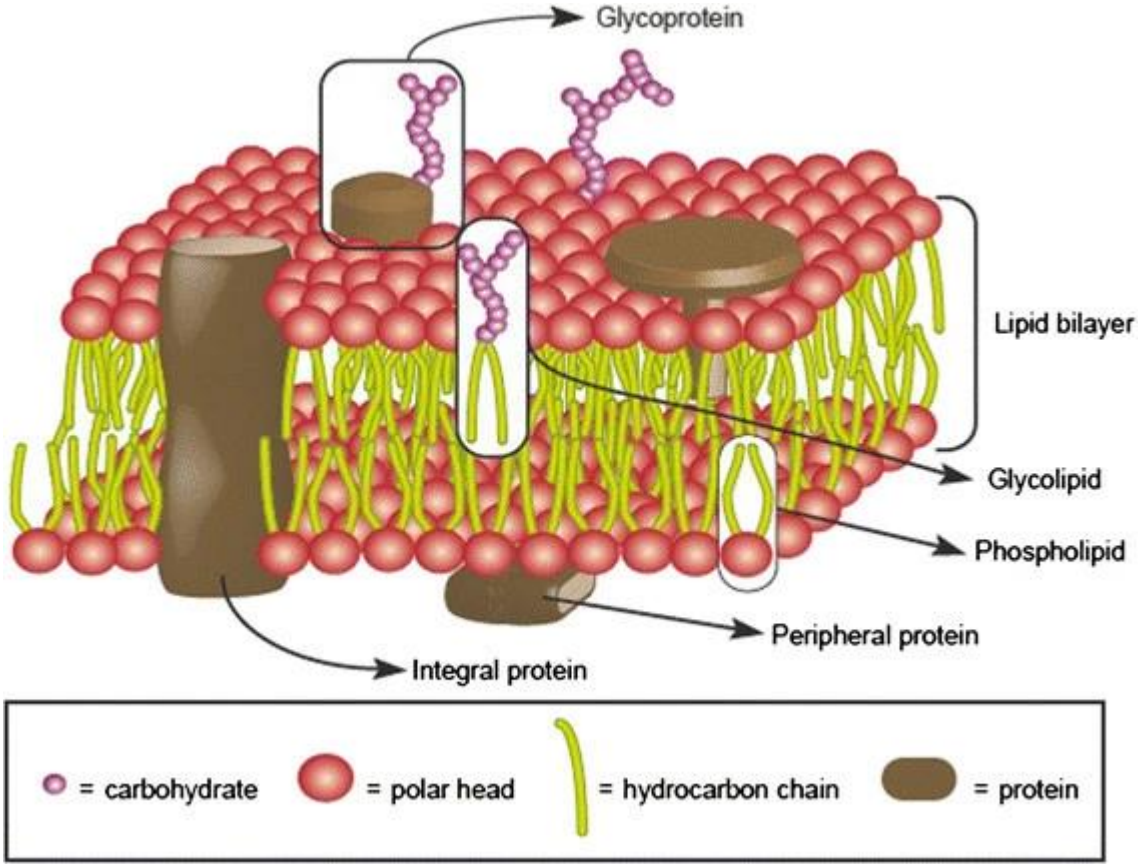


Figure 4.2: Cross section schematic of a cell membrane as described by the fluid mosaic membrane model.

Reproduced from Lombard (2014) under a Creative Commons Attribution License.

4.1.2.2: Detergent extraction of membrane proteins

Detergents represent the most common agents used for extraction and stabilisation of membrane proteins. Detergents, or surfactants as they are more correctly described, are primarily aliphatic molecules which bear a polar hydrophilic headgroup and a hydrophobic tail. The hydrophobic tail can implant into the lipidic plasma membrane, causing disruption of the lipid-protein mosaic and extraction of proteins when at a sufficient concentration. Above the CMC, the polar headgroup will respond by forming micelles and the proteins will become part of a dynamic detergent-lipid-protein complex, with progressive loss of lipid molecules from the protein as the detergent concentration increases. As there is a dynamic equilibrium between free detergent micelles and detergent-protein micelles, it is therefore necessary to always maintain the detergent concentration above the CMC to avoid micelle collapse and loss of the proteins stabilising environment. Aside from the CMC, amongst other properties detergents differ in their aggregation number, which describes the number of detergent monomers per micelle, and correspondingly this informs the size of the micelles which form. (Seddon *et al.*, 2004; Privé, 2007; Stetsenko and Guskov, 2017).

Detergent classes and types

There are many hundreds of different detergents, and it is challenging to find the optimal ones which can not only extract a protein from the lipid bilayer, but stably maintain it in a functional state for downstream studies. While there are certain broad trends for types of detergents towards differing targets and downstream applications, there are no real “universal” detergents which work for all targets. Therefore, screening for the optimal detergent based on the target extraction efficiency, maintained oligomeric state(s), functioning, stability *etc* is essential (Kawate and Gouaux, 2006; Lantez *et al.*, 2015; Kotov *et al.*, 2019)

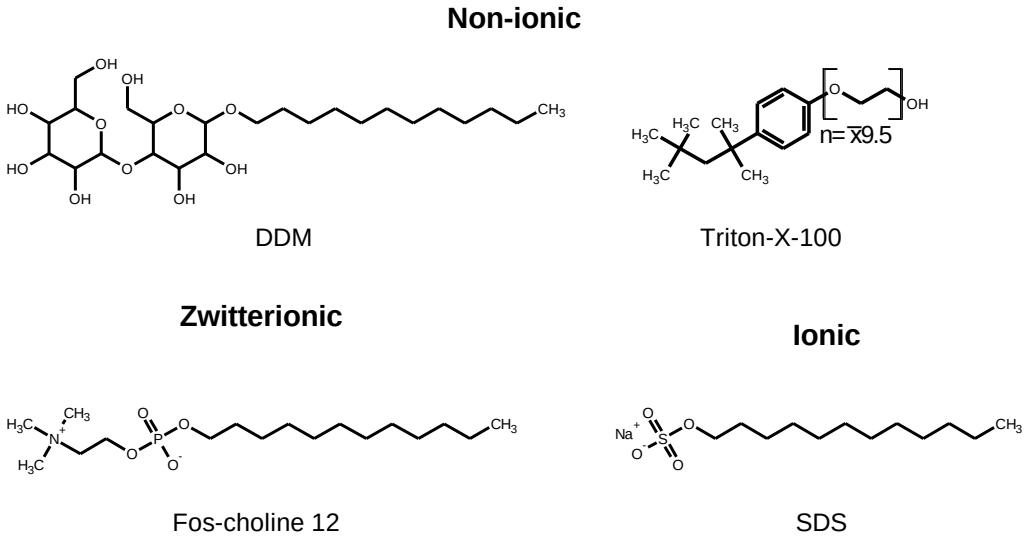


Figure 4.3: Representative common examples of detergents for the three main classes.
DDM, n-Dodecyl-B-D-Maltoside; SDS, sodium dodecyl sulfate.

As described in Seddon *et al.* (2004) and Stetsenko and Guskov (2017), based on their properties, detergents can be split into three main classes as depicted with examples in Figure 4.3:

Non-ionic detergents

Non-ionic detergents possess non-charged hydrophilic headgroups. Examples include Triton-X-100, n-dodecyl-B-D-maltoside (DDM) and n-octyl-β-D-glucopyranoside (OG). Non-ionic detergents are deemed as being “mild” as they can typically maintain the state of extracted proteins by usually only disrupting protein-lipid and lipid-lipid, but not protein-protein interactions. Non-ionic detergents are therefore most frequently used for protein extraction for structural and binding studies where stable, functional protein is essential.

Ionic detergents

Ionic detergents are characterised by their charged head group. Anionic detergents such as sodium dodecyl sulfate (SDS) have negatively charged head groups, while cationic detergents such as cethyltrimethylammonium bromide CTAB possess positively charged ones. Ionic detergents are typically highly efficient at solubilising plasma membranes and releasing transmembrane proteins but compared to non-ionic detergents they are considered as “harsh” because protein-protein interactions are disrupted as well as protein-lipid and lipid-lipid interactions. Proteins solubilised with ionic detergents are therefore

typically in a denatured state and non-functional, so they are rarely used for extracting protein destined for structural or binding studies.

Zwitterionic detergents

Zwitterionic detergents have features of both ionic and non-ionic detergents and carry no net charge. These are seen as being intermediate in terms of their “harshness” and tend to be more deactivating than non-ionic detergents. Prominent examples commonly used for structural biology purposes include 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), Lauryldimethylamine oxide (LDAO) and the fos-choline series.

Detergents for extraction and purification

It is important to note that different detergents are frequently used for the solubilisation and purification steps as these can each require differing detergent properties. While DDM is most commonly used for both, fos-choline 12 for instance is rarely used for solubilisation, but more frequently for purification purposes (Stetsenko and Guskov, 2017). The downstream uses for a protein also dictate the types of detergents used. While DDM for example is the most frequently used detergent for X-ray crystallography and cryo-EM, it is rarely used with NMR studies as the large micelle size induces excessive peak broadening. Small micelle ionic and zwitterionic detergents are therefore preferred for NMR studies as they minimise this effect (Warschawski *et al.*, 2011).

4.1.2.3: Non-detergent agents for extraction and stabilisation

While detergents are the most used agents for extraction and stabilisation of membrane proteins, they do not represent an especially physiologically relevant analogue of the membrane environment as lipid molecules are progressively removed. Lipid molecules can be essential for the correct functioning and stability of membrane proteins and concerns have been raised over the reliability of protein structures which have been obtained in the presence of detergents (Guo, 2020). Free detergent micelles can also complicate downstream assays (Olerinyova *et al.*, 2021) or structural studies and can pose a particular problem for single particle processing with cryo-EM if in excess (Singh and Sigworth, 2015).

Several technologies have been developed as alternatives to detergents for extracting and/or stabilising proteins, including ampiphols, bicelles, lipid nanodiscs and styrene and maleic acid co-polymers (SMA). While these are currently unable to universally replace detergents or require initial detergent extraction, they represent important tools which are likely to become more commonly used as the technologies mature (Seddon *et al.*, 2004; Stetsenko and Guskov, 2017; Guo, 2021). In recent years, lipid nanodiscs and SMA have become particularly prominently used.

Lipid Nanodiscs

Lipid nanodiscs are 8-16 nm diameter disc-shaped lipid bilayers which are stabilised in solution by a surrounding helical protein “belt” which acts as a scaffold. After detergent extraction, before or after purification, the protein is re-constituted into the lipid nanodiscs and the detergent removed. Lipid nanodiscs have been especially successful in recent years for analysis of membrane proteins in a lipid environment which is similar to that of membranes and they pair very well with cryo-EM as free detergent micelles are eliminated. Lipid nanodiscs still require initial detergent extraction however and the native lipids are therefore lost (Schuler *et al.*, 2013; Denisov and Sligar, 2016).

Styrene and maleic acid lipid particles

SMA and related co-polymers work on a similar principle to lipid nanodiscs, with the belt being instead comprised of alternating units of polymers. These go further though in being able to directly extract proteins from the plasma membrane without detergent and therefore capture both the proteins and the native lipid environment. Amongst other applications, SMA technology has been successfully used with crystallography, cryo-EM and circular dichroism, although further improvements are needed for it to widely replace or supplement detergents and lipid nanodiscs (Pollock *et al.*, 2018; Guo, 2021).

4.1.3.0: Protein structural biology techniques

Protein structural biology dates back to the 1930s and in 1957 the first protein structure - of sperm whale myoglobin - was determined by X-ray crystallography (Kendrew *et al.*, 1958). Subsequent protein structures followed, aided by

advancements in both technology and in the understanding of the necessary factors for successful structure determination. The first membrane protein structure – of bacteriorhodopsin- was obtained by 2D electron crystallography in 1975 (Henderson and Unwin, 1975). In 1990, 507 structures were available in the PDB, but this had increased to 172942 at the end of 2020 (when including multiple structures of the same proteins in different states, crystallised with different ligands *etc*) (Berman *et al.*, 2003). This section briefly reviews X-ray crystallography and single particle cryo-EM as these are the dominant techniques currently used for protein structural biology and which have relevance to the data in this thesis.

4.1.3.1: X-Ray Crystallography

Perhaps the most well-known technique used for protein and small molecule structure determination is X-ray crystallography. X-ray crystallography accounts for 88.3% of structures deposited in the PDB up to the end of 2020 (Berman *et al.*, 2003) and has been responsible for many of the key advances in protein structural biology until recently.

The protein of interest is grown as repeating units in a crystal, before recording its X-ray diffraction pattern. For the generation of protein crystals, the purified protein of interest is combined with a crystallisation mix. This includes precipitating agents such as salts and/or polyethylene glycols, as well as buffers and other agents that encourage crystal growth. Ligands or other molecules for co-crystallisation can also be included at this point, or they can be “soaked” into crystals later. The most typical strategy used for finding initial crystallisation conditions in recent years has been sparse matrix screening in multi well vapour diffusion microplates using small (0.1 – 1 µl) drops. Sparse matrix screens are collections of crystallisation mixes which are known to favour crystal growth and offer potential starting points. Typically, larger, higher quality crystals are subsequently generated by bracketing around the conditions identified. The use of additives and/or seeding techniques can also give further improvement (Jancarik and Kim, 1991; Bergfors, 2007a, 2007b; McPherson and Gavira, 2014; Rosa *et al.*, 2020).

Crystal X-ray diffraction can use either low power X-rays (e.g. from a sealed or rotating anode source) or synchrotron radiation (Smyth, 2000). X-rays produced

at a synchrotron are of higher power, which enables more rapid, high quality data collection from smaller crystals. The majority of structures are now determined by data collected using synchrotron radiation (Dauter *et al.*, 2010). By recording and integrating diffraction intensities from the crystal, a 3D atomic model can be determined. However, while the amplitudes resulting from diffraction are recorded, the phases are lost - the so called ”phase problem” - and have to be recovered or estimated first. The structure can then be “solved” using a Fourier transform to calculate an electron density map which can then be interpreted to give a model of the protein. The model is subsequently subjected to rounds of automated refinement with manual input between cycles to improve its quality. This also allows placement of other non-protein molecules such as waters, ions, ligands and components of the crystallisation mix amongst other examples (Ladd and Palmer; Smyth, 2000; Blow, 2004; Egli, 2016).

4.1.3.2 Single Particle Cryo-Electron Microscopy

Although X-Ray crystallography dominates the PDB in terms of structures deposited overall, over the past 10 years the number of structures deposited which used this method has declined. In 2010, 92.4% of the structures deposited in the PDB were determined with X-ray crystallography, but only 80.1% were in 2020. Alternative approaches using electron microscopy (EM) have instead gained favour, accounting for 17.0% of structures deposited in 2020 compared to 0.7% of structures in 2010 (Berman *et al.*, 2003).

Practical issues make it a challenge to image biological samples using electrons. There is a need to analyse under high vacuum conditions and radiation damage is induced by the high energy electrons interacting with the sample and the images are inherently low contrast as elements in biological samples scatter electrons similarly to the water background. Negative stain EM using heavy atom salts such as uranyl acetate was the original solution to these problems, but this has serious drawbacks, the most notable being that the resolution is limited to ~15-20 Å. The most common EM method used for structural studies is single particle cryo-EM, which is the result of a marriage between two separate technologies, namely computational single particle image processing and proteins being rapidly plunge frozen as a thin layer on grids and imaged under cryogenic conditions (Cheng *et al.*, 2015; Nogales, 2016).

Taylor and Glaeser (1974) introduced the freezing of hydrated samples to limit radiation damage, which was revolutionary as it stabilised features and allowed higher resolution data collection compared to negative staining. This methodology was further developed by Dubochet *et al.* (1982) into the process of vitrification, which involves applying protein to EM grids coated with a thin film of holey carbon. Excess liquid is blotted off and surface tension distributes the protein as a thin layer across the holes of the carbon film, after which the grid is rapidly frozen by plunging into liquid ethane. This embeds the protein on the grid in thin ice under random orientations, after which the frozen grid can be imaged on an electron microscope under cryogenic conditions.

In the late 1970s Joachim Frank developed the single particle approach of computationally combining many EM images of a protein at different orientations to determine 2D and 3D structure. In a modern cryo-EM workflow, high quality particles showing the protein of interest are picked either automatically or manually from 2D EM images and initially a 2D classification is carried out to group particles at varying orientations or showing certain features. A subsequent 3D classification aims to integrate these particles to build a 3D map of the protein present, which can then be refined into a model (Cheng *et al.*, 2015; Nogales, 2016).

Molecular size challenges for single particle cryo-EM

Until just under 10 years ago, single particle cryo-EM could only be used with large, or multimeric, proteins of 200 kDa or more and resolutions were typically less than 10 Å. The lower size limit has been steadily decreasing and structures are now routinely reported at 3.0-4.0 Å resolution, with a few reaching <2.0 Å (Berman *et al.*, 2003; Nakane *et al.*, 2020). The key driver in this regard has been the development of direct electron detectors and their rapid readout enabling the recording of “movies”. Software can then deconvolute these into frames to correct for the blurring effect of beam induced sample movement (Wu *et al.*, 2016; Zheng *et al.*, 2017).

Despite impressive developments, proteins smaller than ~100 kDa remain a challenge for single particle cryo-EM. The current record for a molecule alone without any further partners is the 52 kDa biotin-streptavidin complex (Fan *et al.*,

2019). It is important to note that the individual properties of this complex greatly aided reconstruction as streptavidin forms inflexible tetramers and the particles distribute near optimally in thin ice (Wu and Rapoport, 2021). Use of a volta phase plate can further increase the contrast and improve resolution with small proteins, although the usefulness of phase plates has been questioned as a recent comparison study suggested that they increase contrast at the expense of reducing overall data quality (Danev *et al.*, 2021).

Another method to make smaller proteins more tractable to single particle cryo-EM which has been particularly successful is to increase the size of the macromolecule. Fab fragments (Figure 4.4) derived from papain cleaved *in-vivo* generated monoclonal antibodies were one of the first technologies applied to this problem. These are highly specifically targeted and add ~50 kDa to the molecule (Wu *et al.*, 2012). However, the flexibility of Fab fragments between the constant and variable domains can complicate image analysis (Bailey *et al.*, 2018). Production of Fab fragments is also time consuming and expensive as while using display technologies can avoid *in-vivo* generation, they are currently not optimal for creating high affinity fragments as needed with cryo-EM. Fab fragments also cannot be expressed in *E. coli* and need eukaryotic cells for folding (Mukherjee *et al.*, 2020; Kermani, 2021).

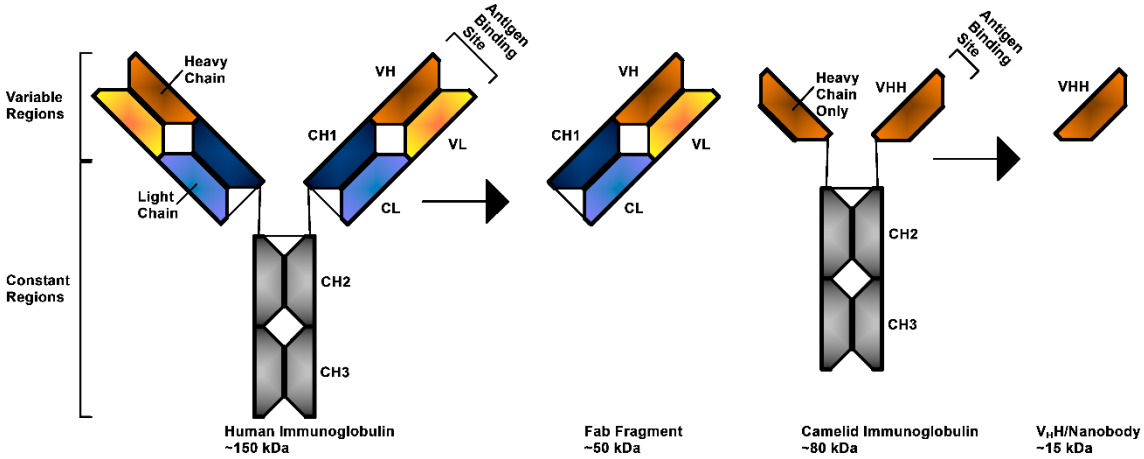


Figure 4.4: Quaternary structures of a human immunoglobulin and a derived Fab fragment and of a camelid immunoglobulin and a derived V_HH/nanobody.

Adapted from Cheloha *et al.* (2020).

Camelids (lamas, alpacas, camels *etc*) naturally generate antibodies which lack the light chains and only have a single variable heavy chain (V_HH, commonly referred to as a nanobody) attached to the constant region of each constant

heavy chain (Figure 4.4). Cleaving the V_HH/nanobody gives an alternative to Fab fragments for labelling proteins. These have the advantage of minimal flexibility and being able to bind into smaller spaces than Fabs. (Cheloha *et al.*, 2020; Bloch *et al.*, 2021). With nanobody yeast libraries, after labelling the protein of interest (e.g. with biotin or a fluorophore), high affinity nanobodies can be raised within weeks using rounds of positive selection with common laboratory techniques such as magnetic and fluorescent cell sorting. After streaking out the yeast on selection plates and sequencing colonies, individual nanobody genes can be sub-cloned and the encoded proteins made in large quantities in *E. coli* (Uchański *et al.*, 2019).

Nanobodies however have a disadvantage in only being ~15 kDa in size and for smaller proteins insufficient mass will likely be added for effective single particle cryo-EM. Several recently developed alternative approaches have been developed to offset this problem. One involves cloning a nanobody into a protein scaffold with a flexible linker, creating a “megabody” of ~50 kDa. These have has been used with several membrane proteins, including a 1.7 Å resolution reconstruction of the $GABA_A$ β_3 homomer. (Nakane *et al.*, 2020). A similar approach dubbed as “legobodies” gives many of the same advantages of megabodies, but instead involves adding two separate scaffolds that universally recognise and rigidly bind to any nanobodies present. In total, this increases the size of the nanobody to ~120 kDa. Legobodies have been recently used for the successful analysis by single particle cryo-EM of two ~20 kDa proteins and it is expected this technique will find much further use (Wu and Rapoport, 2021).

4.1.3.3: Comparison of X-ray crystallography and single particle cryogenic electron microscopy

X-ray crystallography and single particle cryo-EM can both be applied to a wide array of protein targets, but they have strengths and weaknesses in certain areas. These are summarised in Table 4.2.

Table 4.2: Strengths and weaknesses of X-ray crystallography and single particle cryo-EM.

Method	Strengths	Weaknesses
X-ray crystallo-graphy	• Crystallisation typically rigidly fixes protein conformation. • No size limit assuming the protein can be crystallised. • High resolutions of <1.5 Å routinely obtained, near 2.0 Å standard. • Low power X-rays and synchrotron radiation readily accessible.	• High amounts of material required to optimise crystallisation. • Assumes the target can be crystallised, some targets (e.g. membrane proteins, proteins >500 kDa) are difficult in this regard. • Requires highly purified, homogenous protein.
Single particle cryo-EM	• Crystallisation not required. • Large proteins with multiple conformational states. • Less sample required. • Potentially more rapid results as crystallisation not needed. • Resolutions of 3.0-4.0 Å most frequently reported currently but can go to <2 Å . • Increased tolerance for protein heterogeneity.	• Effective analysis of proteins <100 kDa is possible, but difficult without an additional partner protein (e.g a nanobody or Fab) to increase the size. • Limited availability accessing the 300 kV microscopes equipped with direct electron detectors which are needed for high quality structures. • <2.0 Å resolution is uncommon.

(Berman *et al.*, 2003; Skarzynski, 2013; Bai *et al.*, 2015; Vénien-Bryan *et al.*, 2017; Wang and Wang, 2017; Shoemaker and Ando, 2018; Nakane *et al.*, 2020; Yip *et al.*, 2020; Wu and Rapoport, 2021)

While it is generally preferable to obtain an X-ray crystallography structure as the quality is typically higher, the need to grow crystals represents a serious restriction. Until recently, this greatly limited effective structural research on some sample types e.g. membrane proteins. The recent advances in cryo-EM however have fortunately provided a previously unavailable secondary option for when X-ray crystallography is unsuitable. The two techniques complement each other in certain areas, so it is now possible to consider an integrative approach for samples to identify the most appropriate technique (Wang and Wang, 2017; Shoemaker and Ando, 2018).

4.2.0.0: Aims:

- Develop a method for production of purified recombinant RDL.

- Use this recombinant material for assays to characterise the binding of insecticides and other molecules to RDL.

- Obtain a structure of RDL through X-ray crystallography or cryo-EM to characterise the molecular interactions of insecticide binding.

4.3.0.0: Results

4.3.1.0: Production of Resistance to Dieldrin in *Sf*9 insect cells

For production of *Dm* RDL, it was decided to initially use *Sf*9 insect cells with baculovirus transduction. Reasons for this included the likely presence of glycosylation, the ease of scale up with baculovirus and the previous experience of the lab group with the system. There was also the postulated concept that insect cells may represent the most similar environment for an insect related protein.

4.3.1.1: Truncated intracellular domain RDL_{bd} construct

For the initial construct, the RDL_{bd} isoform (Uniprot ID P25123-2) was selected as this is the most common form found in wild-type *Dm*. To simplify the protein and potentially aid future crystallisation efforts, the ICD was removed. Several potential linkers between TMH3 and 4 which have been previously used with PLGICs were assessed by homology modelling. It was eventually decided to use an SQPARAA linker as utilised by the $GABA_A$ β3 homomer structure (PDB ID 4COF) (Miller and Aricescu, 2014) and multiple other $GABA_A$ receptor structures. The gene was synthesised and sub-cloned into the pFastBac 1 vector with the native signal peptide and a C-terminal TEV-His6 tag. This ICD truncated construct is hence termed as $RDL_{bd(ICD\ Trun)}$.

After producing the baculovirus and initial test expressions determining that a His_6 containing protein was visible on western blots, sub-cellular fractionation and indirect immunofluorescence experiments were carried out to determine the localisation of $RDL_{bd(ICD\ Trun)}$. For the sub-cellular fractionation, the cells were mechanically lysed in buffer B (20 mM Tris-HCl, 150 mM NaCl, 2 mM GABA pH 8 [4°C]) plus DNAse I, and protease inhibitors. The extracted membranes/microsomes were homogenised in the same buffer, which was used with all subsequent RDL related proteins. Blotting for the His_6 tag after SDS-PAGE showed a signal at ~35 kDa in the whole cell lysate, which appeared to be amplified in the isolated membranes (Figure 4.5A). No signal was shown in the cytosolic fraction. This therefore suggested that $RDL_{bd(ICD\ Trun)}$ has been successfully produced and trafficked to the cell surface. The signal was at a slightly lower molecular weight than calculated, but membrane proteins frequently

run abnormally on SDS-PAGE due to their high hydrophobicity compared with soluble protein standards (Rath and Deber, 2013).

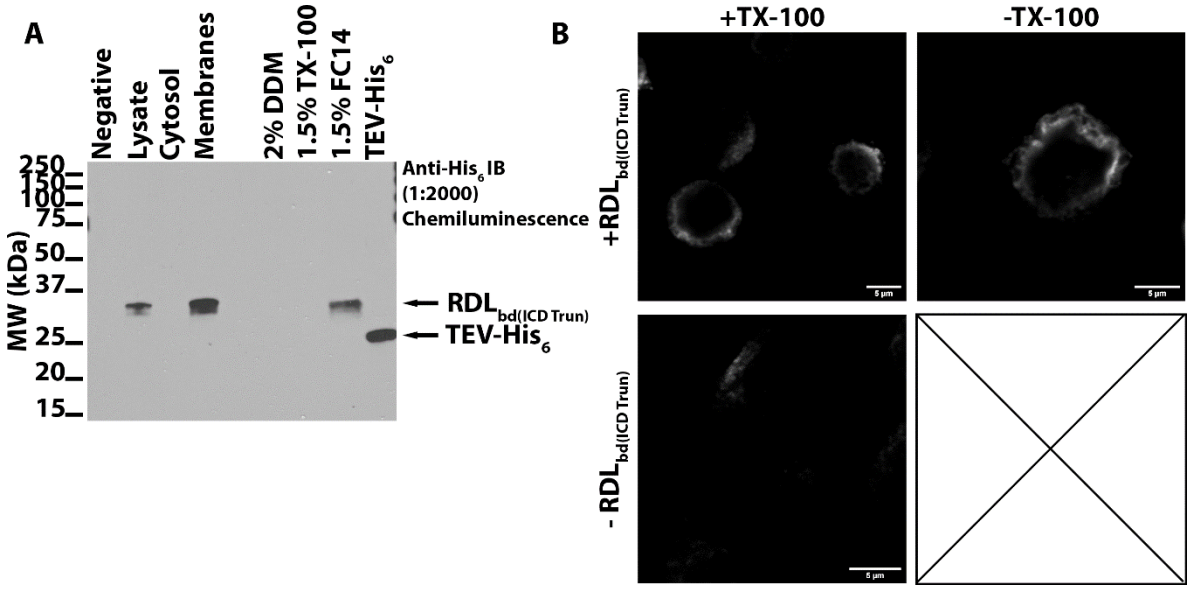


Figure 4.5: Cell localisation of $RDL_{bd(ICD\ Trun)}$.

A: *Sf*9 cells infected with the $RDL_{bd(ICD\ Trun)}$ baculovirus were sub-cellularly fractionated with differential centrifugation. 20 µg of protein was loaded per lane, except for the purified TEV-His_6 positive control where 200 ng was loaded. The negative control consisted of un-infected *Sf*9 cells solubilised in 2% DDM. After transfer the His_6 tag was blotted for. **B:** Anti-His_6 (1:200) indirect immunofluorescence of PFA fixed *Sf*9 cells infected or un-infected with the $RDL_{bd(ICD\ Trun)}$ baculovirus and permeabilised with 0.1% TX-100 or left un-permeabilised as indicated. Scale bars indicate 5 µm. The slides were imaged by Jordana Freemantle.

Anti-His_6 indirect immunofluorescence (Figure 4.5B) with fixed and permeabilised $RDL_{bd(ICD\ Trun)}$ infected *Sf*9 cells likewise suggested that the protein was present on the cell surface and further confirmation of surface expression was given from an identical immunofluorescence pattern being shown with cells which were fixed, but not permeabilised. This also confirmed the extra-cellular localisation of the C-terminus as predicted by homology modelling and following the pattern for other pLGICs.

As a simple initial screen to determine if the protein could be solubilised, samples of membranes were solubilised in buffer B plus 2% DDM, 1.5% TX-100 or 1.5% FC14 for 2 hrs at 4°C before removing insoluble material. On the blot, no signals were present with 2% DDM or 1.5% TX-100, but a strong signal was visible for FC14 (Figure 4.5A). As FC14 is a somewhat harsh detergent which was likely to disrupt protein-protein interactions, further detergent screening was undertaken to find preferentially non-ionic detergents in which $RDL_{bd(ICD\ Trun)}$ was soluble. Initially whole cell samples were screened for solubility with various non-ionic and

zwitterionic detergents, but again only fos-choline series detergents appeared to show any signals on blots (data not shown). Further tests were therefore performed on isolated membranes to try and enhance any signals present. Membranes at 2 mg/ml were solubilised with a panel of seven commonly used non-ionic detergents, plus FC13 as a control detergent known to be able to extract the protein. None of the non-ionic detergents tested however appeared to show any signals on blots (Figure 4.6A-B). The exclusive solubility in fos-choline series detergents raised concerns over the proteins folding state. While these detergents are exceptionally efficient at solubilisation, it has been suggested that membrane proteins in fos-choline micelles are destabilised and represent both folded and un-folded states (Kotov *et al.*, 2019).

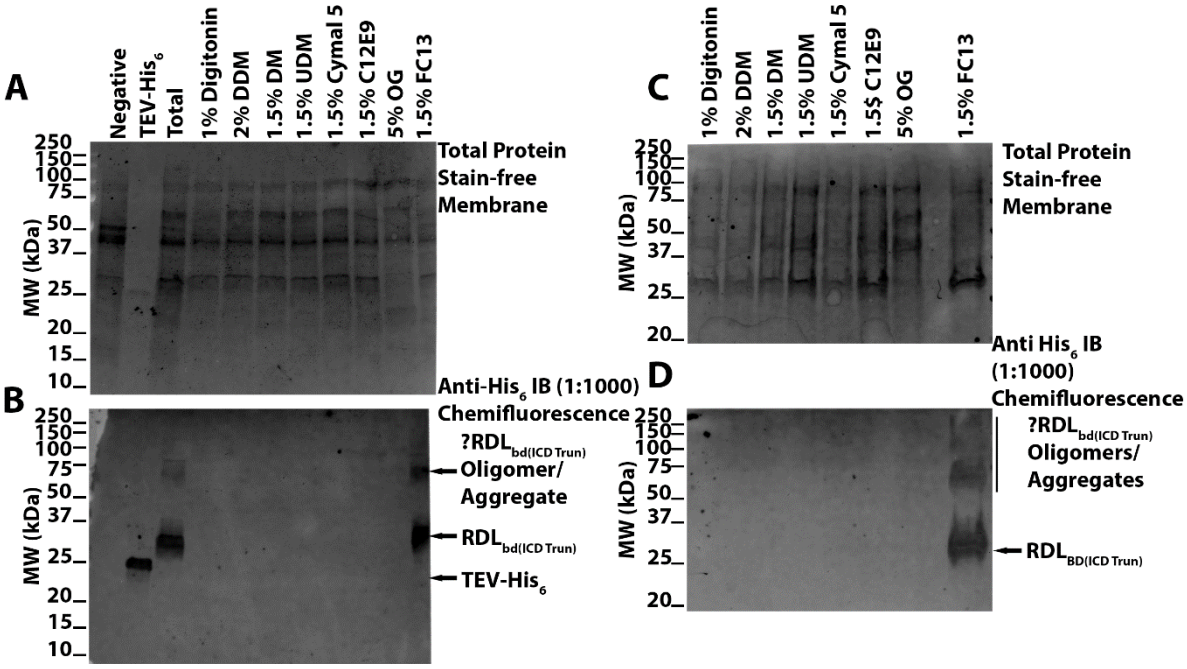


Figure 4.6: $RDL_{bd(ICD\ Trun)}$ detergent solubility screening

A-B: Membranes extracted from *Sf*9 cells infected with the $RDL_{bd(ICD\ Trun)}$ baculovirus were solubilised in the indicated detergents in buffer B at 2 mg/ml for 2 hours at 4 °C. Insoluble material was then removed by ultra-centrifugation. The total sample consisted of membranes which had no detergent added and were not ultra-centrifuged. The negative control was un-infected *Sf*9 cells solubilised in 2% DDM. ~20 µg of protein was loaded per lane except for the purified TEV-His_6 positive control where 500 ng was loaded. Following SDS-PAGE and transfer to a membrane, the stain free total protein on the membrane was recorded (**A**) before proceeding with anti-His6 immunoblotting (**B**).

C-D: Co^{2+} charged resin enrichment of the detergent soluble samples in A. 20 µl of equilibrated 50% Co^{2+} charged resin was added to each sample and gently mixed for 2 hours at 4 °C. Bound protein was then eluted with 4 bed volumes of reducing/denaturing 1x NuPage loading buffer plus 50 mM DTT at 50 °C for 10 minutes with gentle shaking. 15 µl of sample was loaded per lane – due to the dye and DTT in the loading buffer it was not possible to measure the protein concentration for equal protein loading – and after transferring and recording the stain-free total protein (**C**) the His_6 tag was blotted for (**D**).

To try and amplify any signals present, the detergent soluble samples were enriched for $RDL_{bd(ICD\ Trun)}$ by binding to 20 ul of equilibrated 50% Co^{2+} charged resin for 2 hours at 4 °C, followed by elution in 50 °C reducing and denaturing loading buffer for 10 minutes for maximal efficiency. After running the enriched samples on SDS-PAGE it was apparent that the number of proteins visible was reduced compared with the non-enriched samples, although there was still a wide range (Figure 4.6C). After blotting for the His_6 tag however, no signals were visible with any sample except 1.5% FC13 (Figure 4.6D).

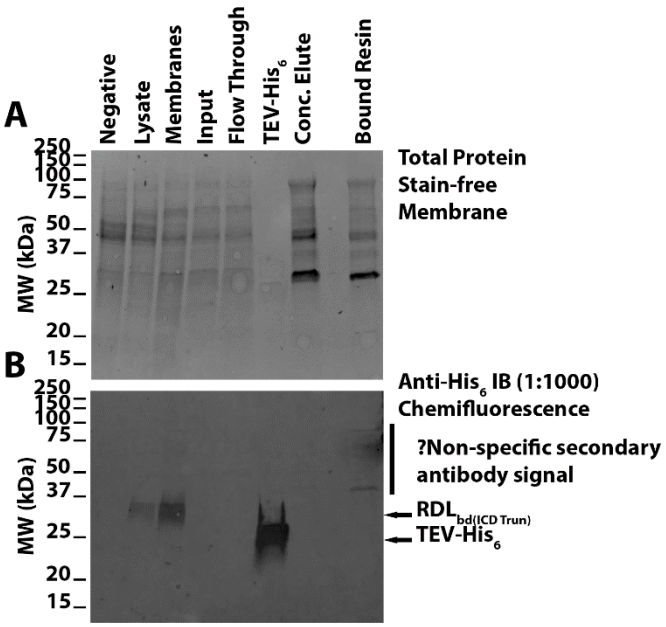


Figure 4.7: $RDL_{bd(ICD\ Trun)}$ purification

Membranes extracted from $RDLbd_{(ICD\ Trun)}$ infected cells were solubilised in 2%/0.4% DDM/CHS for 3 hours at 4 °C. Afterwards insoluble material was removed to obtain the "Input" sample, 1 ml of equilibrated 50% Co2+ charged resin was added and stirred gently for 1.5 hrs at 4 °C. After collecting in a gravity column, the resin was washed with 10 bed volumes of buffer B plus 0.05/0.01% DDM/CHS and the proteins were then eluted with 10 bed volumes of the same buffer also containing 250 mM imidazole. Prior to elution, a 15 ul sample of bound Co2+ charged resin was collected and eluted with 45 ul of denaturing/reducing 1x NuPage loading buffer plus 50 mM DTT at 50 °C for 10 minutes. The samples were then run on SDS-PAGE with ~20µg of protein loaded per lane except for the purified TEV-His_6 positive control where 500 ng was loaded and the bound resin lane where 15 µl was loaded. The negative control consisted of un-infected *Sf*9 cells solubilised in 2% DDM. After transfer the total protein on the stain-free membrane was recorded (**A**) and the His_6 tag then blotted for (**B**).

In a final attempt to see if any soluble $RDL_{bd(ICD\ Trun)}$ could be detected when using non-ionic detergent, membranes from 1 L of culture were extracted and solubilised in buffer B plus 2% DDM and 0.4% CHS. This is a commonly used detergent combination with PLGICs. After insoluble material removal the detergent soluble fraction was purified with Co^{2+} charged resin.

The end result (Figure 4.7A) was similar to those of the enriched samples in the prior detergent screen. Similarities were also shown for a sample of bound resin which was collected prior to elution and had the proteins eluted under denaturing conditions. After transferring and blotting for the His_6 tag (Figure 4.7B), while a signal at the expected MW of $RDL_{bd(ICD\ Trun)}$ could be seen in the lysate and

membrane samples, nothing could be seen in further samples after the detergent solubilisation. There were some faint bands in the purified protein lane, which were also visible more intensely with the sample of resin, but these did not match the expected MW. As a monoclonal primary antibody was used, these are therefore suspected to be from non-specific binding of the secondary antibody.

4.2.1.2: Full Length GFPuv-RDL$_{ac}$

The initial construct of RDL with the ICD replaced with a short linker could not be solubilised aside from when using harsh zwitterionic detergents or denaturing conditions. This raised concerns over the proteins folding state, so it was decided to re-assess the strategy. Although replacement of the ICD is commonly used with PLGICs and does not usually appear to grossly affect the proteins properties, there was the possibility that the replacement affected the folding or assembly. Therefore, it was decided to design a new construct with the ICD retained. It was also decided to switch to using the “ac” isoform of *Dm* RDL (Uniprot ID P25123-1) as this is much more sensitive to GABA, which could be an advantage with functional assays. In terms of sequence details there are only minor changes between the isoforms as previously detailed in Chapter 1.

For this new construct, a strategy modelled on that used by Claxton and Gouaux (2018) for identifying the optimal production parameters of the human $\alpha_1\beta_1$ GABA$_A$ receptor was used. Therefore, a GFPuv followed the signal peptide of *Dm* RDL$_{ac}$ as listed in Uniprot (between residues V44 and Q45) and was coupled to the mature protein with a flexible 13 residue GSGSGS-TEV linker. The tag was to enable in-gel fluorescence and FSEC. The N-terminal position of the GFPuv was used over positioning in the M3-M4 intracellular loop as is commonly used with PLGICs to avoid interrupting the loop and invalidating the hypothesis that truncation of the ICD had been responsible for the poor solubility of RDL$_{bd(ICD\ Trun)}$. GFPuv was used over other XFP derivatives due to its enhanced stability, which is particularly relevant as the RDL C-terminus is extracellular (Ashikawa *et al.*, 2011). On the C-terminus the TEV-His$_6$ purification tag was changed to a twin-strep tag, which has been widely used with membrane proteins due to its exceptional binding affinity to engineered streptactin resin, high purity in a single step and gentle elution with biotin or desthiobiotin (Yeliseev *et al.*, 2017). The construct is hence referred to as GFPuv-RDL$_{ac(INS)}$.

Initial experiments (Figure 4.8) with the GFPuv-RDL$_{ac(INS)}$ construct, after producing the baculovirus, showed that the protein was present in the membranes, but not the cytosol, by in-gel GFP fluorescence and anti-GFP western blot. In-gel GFP fluorescence (Figure 4.8A) appeared to correlate well with results from the blot (Figure 4.8B), with the advantage of rapid results and detection of proteins which were faint or did not appear on the anti-GFP blot (in the cytosol sample for instance). In all the samples on in-gel fluorescence, including the cytosol, there was a band at ~25 kDa and an additional wide band at ~55 kDa. Based on a sample of purified GFPuv used as a positive control, these bands are representative of free GFPuv and likely resulted from proteolytic cleavage of the linker. The 50 kDa band probably represents a dimer, as GFP derivatives that have not had a monomerising mutation (such as A206K) added will naturally dimerise (Zacharias *et al.*, 2002). This experiment used a 10% SDS-PAGE gel, but it was noted subsequently that on AnyKd™ gradient gels (e.g figure 4.8C) this 55 kDa band could be split into two separate bands at ~45 and 55 kDa. This could potentially suggest that shorted forms of GFPuv-RDL$_{ac(INS)}$ were present in addition to GFPuv dimers.

Fluorescence microscopy of both live (Figure 4.8D) and PFA fixed (Figure 4.8E) *Sf*9 cells infected with the GFPuv-RDL$_{ac(INS)}$ baculovirus likewise suggested a presence on the surface. The signals for the N-terminal GFPuv and an anti-StrepII antibody directed against the C-terminal twin-strep tag appeared to colocalise, confirming that full length RDL was being detected rather than just free GFPuv.

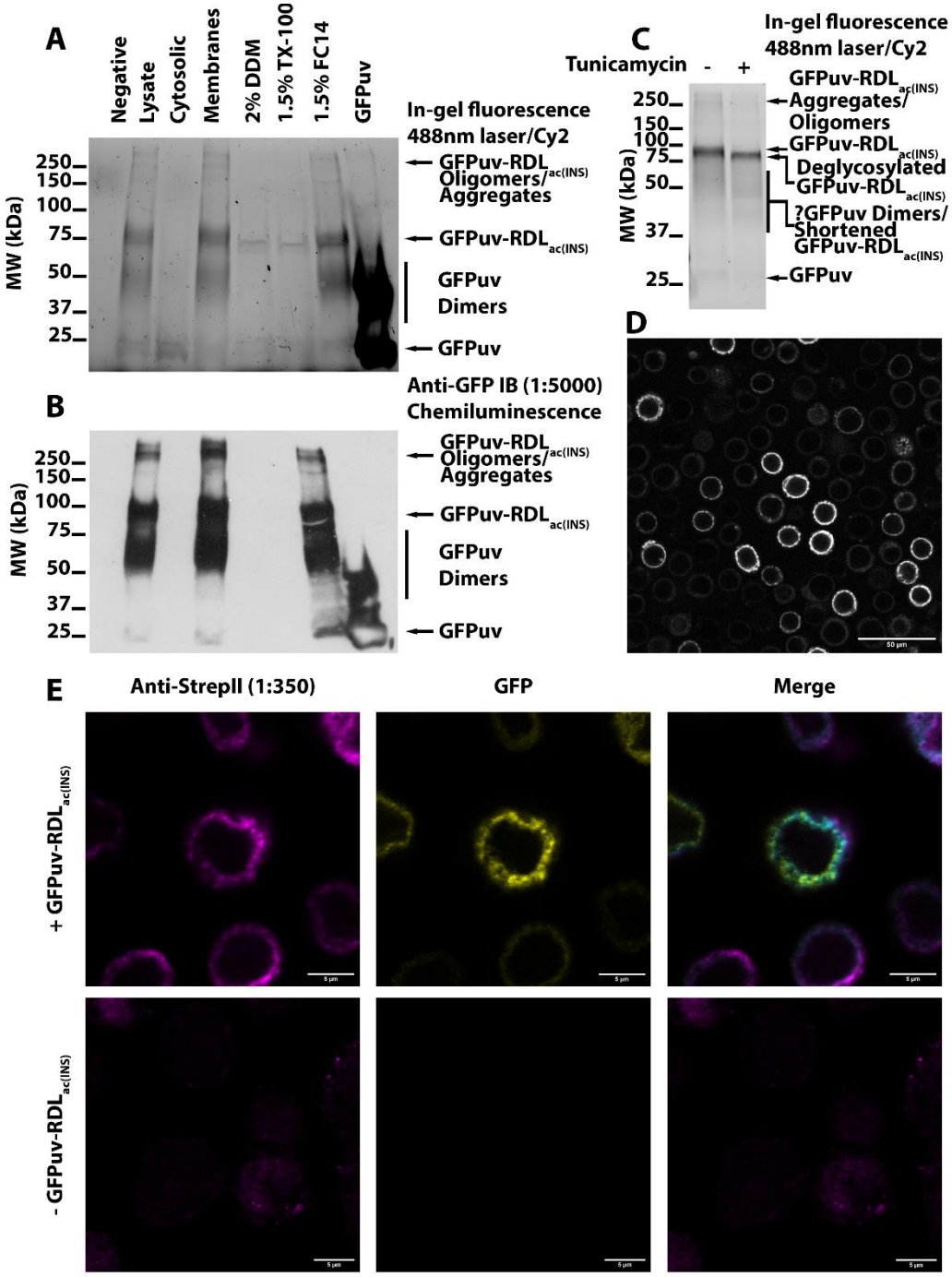


Figure 4.8: Cell localisation of GFPuv-RDL$_{ac(INS)}$

A-B: SDS-PAGE in-gel GFP fluorescence (**A**) and subsequent anti-GFP blot (**B**) of sub-cellularly fractionated *Sf*9 cells infected with the GFPuv-RDL$_{ac(INS)}$ baculovirus and also membrane samples solubilised in the indicated detergents. ~20 µg of protein was loaded per lane, except for the purified GFPuv positive control where 50 ng was loaded. **C:** In-gel GFP fluorescence of mechanically lysed GFPuv-RDL$_{ac(INS)}$ infected *Sf*9 cells treated or un-treated with tunicamycin (5 µg/ml). ~20 µg of protein was loaded per lane. **D:** Live cell fluorescence microscopy of *Sf*9 cells excited at 405 nm at ~48 hours of infection with GFPuv-RDL$_{ac(INS)}$. Scale bar indicates 50 µm. **E:** Indirect immunofluorescence of fixed and permeabilised *Sf*9 cells infected or un-infected with the GFPuv-RDL$_{ac(INS)}$ baculovirus and probed with anti-strepII (1:350). The AF594 secondary antibody was excited at 594 nm and the attached GFPuv at 488 nm. Scale bars indicate 5 µm. The slides were imaged by Jordana Freemantle.

It was noted there appeared to be a double band visible on the in-gel fluorescence and western blots for GFPuv-$RDL_{ac(INS)}$. PLGICs typically contain at least one N-linked glycosylation site and there are two predicted on *Dm* RDL as listed in Uniprot. While a similar double band appearance has been previously noted with the $RDL_{bd(Trun)}$ construct, it was not investigated at the time. Incubation of *Sf*9 cells with 5 µg/ml of tunicamycin 24 hours after infection to induce ER stress and chemically block N-linked glycosylation reduced the mobility of the GFPuv-$RDL_{ac(INS)}$ band detected on in-gel fluorescence compared to a non-treated population of cells (Figure 4.8C). This therefore confirmed the suspicion that the upper band represented a glycosylated form of the protein.

Duplicating the approach previously used with the truncated construct, as an initial detergent solubility screen samples of the extracted membranes were solubilised in 2% DDM, 1.5% TX-100 and 1.5% FC14. The detergent soluble fractions were isolated and then separated by SDS-PAGE. In-gel fluorescence (Figure 4.8A) suggested that a small amount of GFPuv-RDL had been extracted from the membranes by DDM and TX-100, with FC14 pulling out a much greater proportion. However, after transferring the gel and immunoblotting with anti-GFP (Figure 4.8B), while there was still a robust signal for the FC14 sample, nothing was visible with DDM or Triton-X-100. Therefore, this suggests that a low level of GFPuv-$RDL_{ac(INS)}$ can be solubilised, but this is not easily detectable on blots due to loses during transfer to the membrane. In-gel fluorescence was subsequently used as the primary technique for future experiments with detergent screening, as it appeared to be more informative and reliable with samples only containing low levels of solubilised protein.

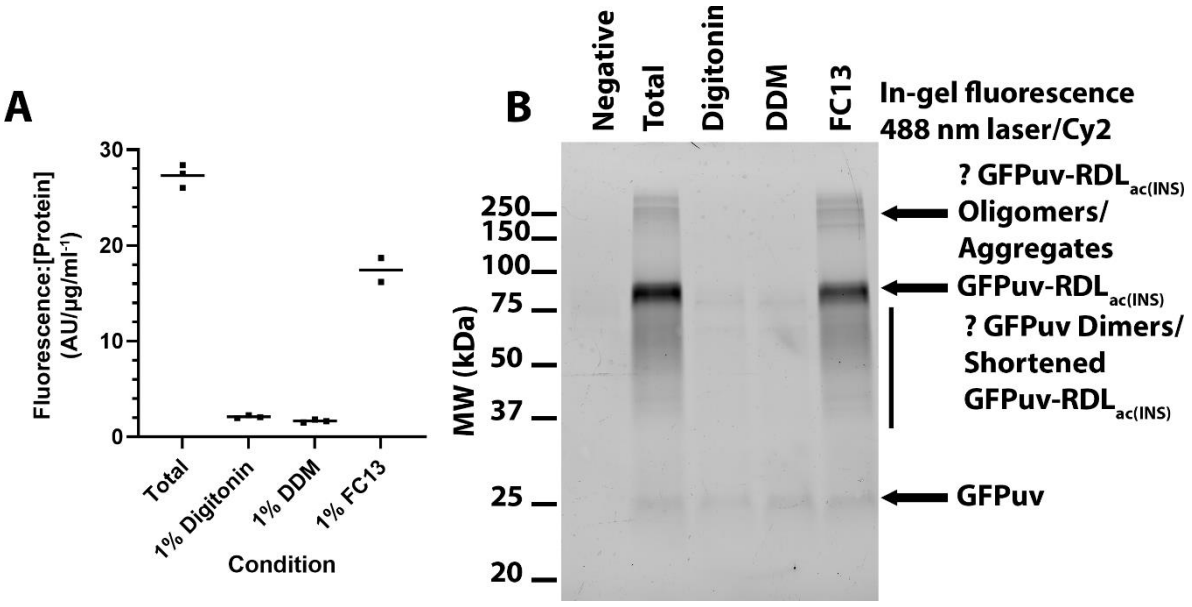


Figure 4.9: GFPuv-RDL$_{ac(INS)}$ folding screen assay.

*Sf*9 cells infected with the GFPuv-RDL$_{ac(INS)}$ baculovirus were solubilised at 1x10^7 cells/ml in buffer B plus 1% digitonin, DDM or FC13 for 3 hours with gentle mixing at 4 °C. They were then ultra-centrifuged to remove insoluble material. The total sample consisted of cells without any detergent added which were lysed by needle sheering and not ultra-centrifuged. **A:** Scatter dot-plot of the samples fluorescence:[protein] measurements, calculated by dividing the fluorescence in AU by the protein concentration in µg/ml as quantified by a BCA assay. Readings were recorded in triplicate (except for FC13 which is a duplicate read) at 405/510 nm with the individual reads shown as dots and the mean as a horizonal bar. **B:** SDS-PAGE in-gel GFP fluorescence of the samples. ~20 µg of protein was loaded per lane. The negative control sample consisted of un-infected *Sf*9 cells lysed using 2% DDM in buffer B. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set.

As a simple initial step to characterise how much of the protein was potentially folded, a folding screen assay adapted from a method described by Thomas and Tate (2014) was carried out. 1% digitonin, DDM and FC13 have ascending abilities to solubilise un-folded protein, ranging from minimal to potent. Comparison of cells or membranes solubilised in these detergents after removing insoluble material to a total un-separated sample (e.g. by blotting or fluorescence measurements) therefore gives a crude semi-quantitative readout of the level of potentially folded protein. Thomas and Tate also found that this assay appeared to correlate well with the number of sites quantified by radioligand saturation binding assays and therefore the level of functional protein. Whole cell samples of GFPuv-RDL$_{ac(INS)}$ infected cells were hence solubilised in 1% digitonin, DDM and FC13 and the fluorescence:[protein] ratio calculated by reading total fluorescence at 405/510 nm and dividing by the sample protein concentration obtained by Rapid Gold BCA assay (Figure 4.9A). This corrected the fluorescence readings for the differing final concentration of protein extracted by

each detergent. These measurements suggested that only minimal amounts of the total present GFPuv-RDL$_{ac(INS)}$ – less than 10% – could be extracted by 1% digitonin or 1% DDM, while ~60% could be by 1% FC13. After separation by SDS-PAGE (Figure 4.9B), in-gel fluorescence showed a similar pattern to the fluorescence:[protein] measurements, with only minimal amounts of GFPuv-RDL$_{ac(INS)}$ in the 1% digitonin and DDM samples compared to 1% FC13. Collectively, this could suggest that the majority of RDL$_{ac(INS)}$ present is in an un-folded state.

Detergent screening on whole cell samples was performed to see if any other detergents could improve extraction efficiency. Five of these (DDM, DM, UDM, Cymal 5 and C12E9) were mild non-ionic detergents which are frequently for extraction of membrane proteins in a functional state (Stetsenko and Guskov, 2017). The harsher zwitterionic detergents LDAO and FC14 were also included, the latter of which as a positive control of a detergent known to be able to efficiently extract GFPuv-RDL$_{ac(INS)}$. CHS was included in each detergent at ~1/10th to 1/5th the detergent concentration as it frequently has a stabilising effect on membrane proteins (Hattori *et al.*, 2012).

The screening however did not appear to show that any of the detergents tested improved extraction efficiency noticeably. The fluorescence:[protein] ratio of the detergent soluble fractions suggested that for all of the non-ionic detergents and LDAO/CHS the signal was less than 10% of the total (Figure 4.10A). FC14/CHS was the exception where the signal was ~40% of the total, although this varied between the two repeats of this screen. Separation of these samples by SDS-PAGE and recording the in-gel fluorescence (Figure 4.10B) corroborated that only a minimal amount of GFPuv-RDL$_{ac(INS)}$ appeared to be present in any sample, except for FC14/CHS. Levels of free GFPuv in each sample furthermore appeared to be similar. Following transfer of this gel to a PVDF membrane and blotting for GFP, as with previous blots no signal was present except with the total, FC14/CHS and purified GFPuv samples (data not shown). Although extraction at 4 °C is typically used with membrane proteins to limit proteolysis, it was also tested if using room temperature (~20 °C) instead could enhance the level of protein extracted. However, there did not appear to be any appreciable change in the fluorescence:[protein] measurements (Figure 4.10C) when using

either 2% DDM/CHS or 1.5% FC14/CHS and in-gel fluorescence (Figure 4.10D) did not indicate any noticeable differences between the two temperatures.

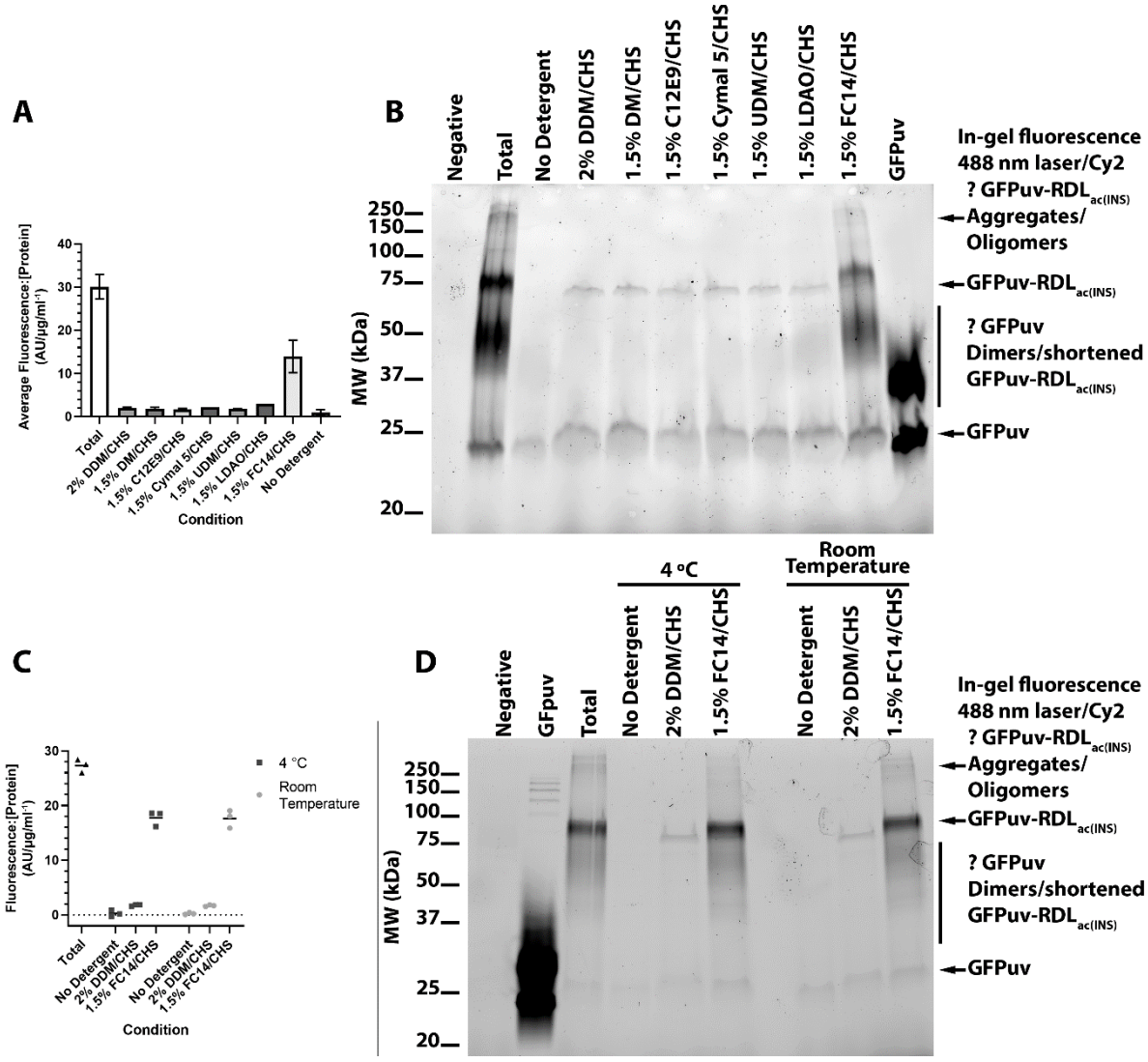


Figure 4.10: GFPuv-RDL$_{ac(INS)}$ detergent solubility screening.

A-B: Detergent screening of whole cell samples expressing GFPuv-RDL$_{ac(INS)}$. *Sf*9 cells infected with the GFPuv-RDL$_{ac(INS)}$ baculovirus were solubilised at 1x10^7 cells/ml in the indicated detergents (each of which included CHS at ~1/10th -1/5th the detergent concentration) made up in buffer B for 2-3 hours with gentle mixing at 4 °C. These were then ultra-centrifuged to remove insoluble material. The no detergent samples were prepared with the detergent replaced with buffer and were lysed by needle sheering. The total sample was prepared likewise but not ultra-centrifuged. **A:** Bar chart of the samples average fluorescence:[protein] measurements, calculated by dividing the fluorescence at 405/510 nm in AU by the protein concentration in µg/ml as quantified by a BCA assay. n=2 except for Cymal 5/CHS and LDAO/CHS where n=1. Error bars represent standard error, where no error bar is visible then it is smaller than the size of the line or n=1. **B:** Representative SDS-PAGE in-gel GFP fluorescence of the samples. ~20 µg of protein was loaded per lane except for the purified GFPuv positive control where 50 ng was loaded. The negative control consisted of un-infected *Sf*9 cells solubilised in 2% DDM. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set. **C-D:** Comparing the effect of temperature on the extraction of GFPuv-RDL$_{ac(INS)}$. The samples were prepared as in A-B except for the room temperature samples being left to solubilise at about 20 °C. **C:** Scatter dot-plot of the samples fluorescence:[protein] measurements calculated as described in A. Readings were recorded in triplicate at 485/510 nm with the individual reads shown as points and the mean as a

horizonal bar. Where no bar is visible then it is within the size of the symbols. **D:** SDS-PAGE in-gel GFP fluorescence of the samples in C as described in A except for the no detergent sample lanes having 10 µg rather 20 µg loaded due to low sample concentrations.

While experiments to this point showed that a small amount of $RDL_{ac(INS)}$ could be extracted with non-ionic detergents and a greater proportion with FC13/14, it was uncertain as to whether this represented protein with the expected pentameric quaternary organisation, monomers, a mix of states or even aggregates. Protein monodispersity is seen as an important pre-requisite for success in crystallisation and other downstream techniques (Newby *et al.*, 2009; Birch *et al.*, 2018). Determination of a proteins oligomeric state in an un-purified form is difficult though as most techniques, for example SEC - alone or combined with multiangle light scattering (SEC-MALS) -, analytical ultra-centrifugation (AUC), dynamic light scattering *etc*, require purified protein (Edwards *et al.*, 2020). One option is to use native PAGE, but separating complex mixtures this way is technically difficult and blotting for the target typically requires a second dimension separation with SDS-PAGE (Kikuchi *et al.*, 2011). Clear native PAGE also has poor resolution as it separates based on the proteins shape, charge and MW. Blue native-PAGE (BN-PAGE) is an improvement as the charge is equalised by the Coomassie G-250 dye in the buffers and samples, but the dye itself can disrupt complexes and the blue background reduces the signal of fluorescently tagged proteins (Wittig *et al.*, 2007).

In the past 15 years, fluorescence size exclusion chromatography has become a widely used technique with membrane proteins for detergent screening and characterisation of a proteins oligomeric state and stability. FSEC relies on using a fluorophore fusion or attached fluorescently labelled antibody/nanobody as a marker for the protein when an un-purified mixture is separated on size exclusion column and detected by fluorescence optics. The main advantages are the small amount of sample required and the native conditions under which the protein is analysed (Kawate and Gouaux, 2006; Hattori *et al.*, 2012; Furukawa *et al.*, 2021).

As an in-line fluorescence detector was unavailable, a simpler alternative FSEC method was developed to analyse GFPuv-$RDL_{ac(INS)}$. This involved fractionating the column elute directly into solid black 96 well plates and reading the fluorescence. A similar approach was later found to have been previously described (Drew *et al.*, 2008). The SEC was carried out at 4 °C to limit proteolysis

and reflect the conditions under which the protein would be purified. Detergent solubilised samples were eluted into buffer B plus 0.05/0.01% DDM/CHS. The use of DDM or other mild non-ionic detergents has previously been shown to not "rescue" harmful destabilising effects of detergents used for solubilisation (Kawate and Gouaux, 2006).

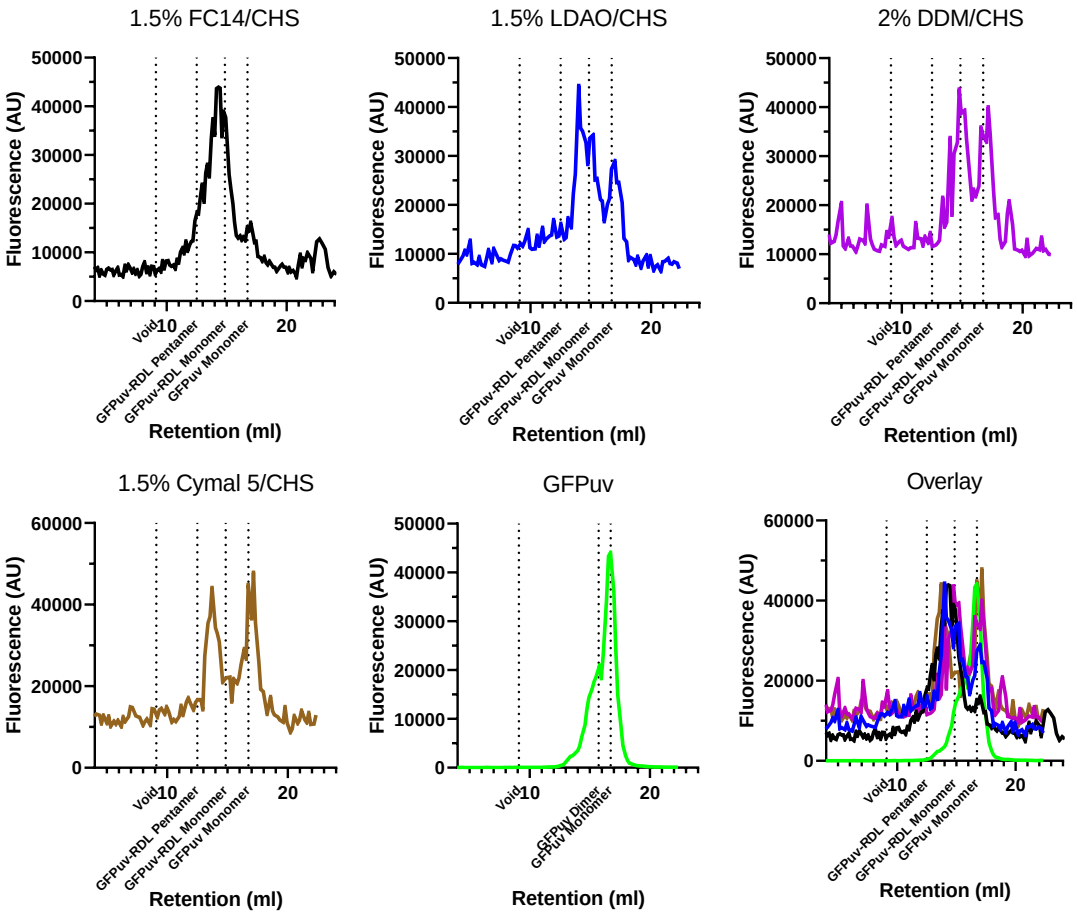


Figure 4.11: Fluorescence size exclusion chromatography of GFPuv-RDL$_{ac(INS)}$ detergent soluble samples

*Sf*9 cells infected with the GFPuv-RDL$_{ac(INS)}$ baculovirus were solubilised at 1x10^7 cells/ml in the detergents indicated in the figure for 2 hours at 4 °C before insoluble material was removed by ultra-centrifugation. At 4 °C each detergent soluble sample was loaded onto a Superose 6 10/300GL size exclusion column equilibrated in buffer B plus 0.05%/0.01% DDM/CHS via a 500 µl loop and eluted into black 96 well microplates as 200 µl fractions (except for the FC14/CHS sample which was eluted as 120 µl fractions over two microplates to improve resolution due to the intense fluorescence in comparison to the other samples). For the purified GFPuv sample, 200 µg was loaded onto the column. After fractionation the fluorescence of each microplate was read at 405/510 nm at a gain calculated for the most intense well to reach 70% of the maximum dynamic range. The traces presented are therefore qualitative rather than quantitative and show the average of every two data-points to reduce noise. The retention volumes for the column void, pentamers (490 kDa) and monomers (98 kDa) of GFPuv-RDL$_{ac(INS)}$ and GFPuv monomers (27 kDa) and dimers (54 kDa) as calculated from standards are indicated. The data in the overlay panel is coloured for each dataset as per the individual entries.

Using FSEC (Figure 4.11), all four of the detergent solubilised samples of GFPuv-RDL$_{ac(INS)}$ appeared to show two main peaks. The second major peak eluting later

on appeared to correspond to a monomer of GFPuv from running a 200 µg sample of purified GFPuv and calculating the expected retention from protein standards. This peak also showed a slight shoulder for what is assumed to represent the dimeric form of GFPuv. The first major peak is postulated to correspond to a form of GFPuv-$RDL_{ac(INS)}$ as each eluted at a retention volume between that calculated for a pentamer (490 kDa) and monomer (98 kDa). Caution should be taken though with assigning oligomeric states to GFPuv-$RDL_{ac(INS)}$ from this FSEC as the protein is non-globular and therefore will likely not elute as per globular protein standards.

Although this peak in 1.5% Cymal 5/CHS especially and 2% DDM/CHS to a degree was fairly monodisperse, 1.5% LDAO/CHS and FC14/CHS showed smaller peaks as part of the main peak postulated to correspond to GFPuv-$RDL_{ac(INS)}$, suggesting the presence of other oligomeric states. It is difficult however to tell if these are real signals or just noise. For the FC14/CHS sample it was necessary to run a sample twice, as on the first run (not shown) collecting 200 µl fractions showed poor resolution and separation because the fluorescence signals were so intense. However, the second run where 120 µl fractions were taken instead showed similar results to the other detergent traces with 200 µl fractions. This illustrates a disadvantage of performing FSEC in this manner: the fractionation causes resolution loss, but there also needs be sufficient signal strength to determine real signals from background noise.

By this point, it was clear that *Sf*9 cells were unlikely to be feasible for production of *Dm* RDL, as only small proportions of the protein could be successfully extracted in what was likely a folded state. This was unfortunate considering that FSEC traces for the four tested detergents were indicative of well-behaved protein that was reasonably monodisperse and did not show evidence of aggregation.

4.3.2.0: Production of Resistance to Dieldrin in mammalian HEK293 cells

All the currently available $GABA_A$ receptor structures have been obtained from protein produced in mammalian HEK293 cells. Therefore, the logical next step was to determine if mammalian expression represented a viable method for production of *Dm* RDL. Due to the high cost and complexity of large volume

mammalian suspension cultures, it was decided to initially carry out a simple pilot study using adherent HEK293 cells with lipofectamine-2000 transfection.

The signal peptide for *Dm* RDL is interesting as it is longer than that for most other PLGICs, being 44 residues compared to ~20-35 for most human PLGIC signal peptides. It contains three possible initiation sites, although the first is accepted as being canonical. The cleavage site is also putative as despite being commonly described in Uniprot and other references as being between residues V44 and Q45, it was not possible to find any literature which had experimentally confirmed this. Running the protein sequence through the SignalP-6.0 server (https://services.healthtech.dtu.dk/service.php?SignalP - Teufel *et al.*, 2022) intriguingly suggested a high probability that RDL may not have a signal peptide. This would differ compared to other PLGICs however and is beyond the scope of this research.

Regardless, as it was not possible to experimentally confirm the cleavage site, it was decided to move the attached GFPuv to a flexible region five residues further along from the assumed cleavage site – between G50 and G51. Sequence alignments with other PLGIC subunits suggested that their cleavage sites were before this point. Aside from moving the GFPuv, the only other change made was to substitute the equivalent of Ala206 in the GFPuv to Lys, which ensures that only monomeric GFPuv can form as assembly of dimers is sterically blocked (Zacharias *et al.*, 2002). It was anticipated that this would ease interpretation of in-gel fluorescence and blotting experiments by eliminating the GFPuv dimers seen previously with the GFPuv-RDL$_{ac(INS)}$ construct. The construct for mammalian expression is henceforth referred to as mGFPuv-RDL$_{ac(MAM)}$.

4.3.2.1: Extraction of mGFPuv-RDL$_{ac(MAM)}$ from adherent HEK293 cells

After sub-cloning into the pcDNA3.4 TOPO vector, mGFPuv-RDL$_{ac(MAM)}$ was transfected into adherent HEK293 cells using lipofectamine 2000 and the cells harvested after 24-48hrs. Fluorescence around the edge of the cells was visible under wide-field fluorescence microscopy after 24 hours and typical transfection efficiencies of 70-80% were suggested using the GFP channel of the Countess II FL cell counter.

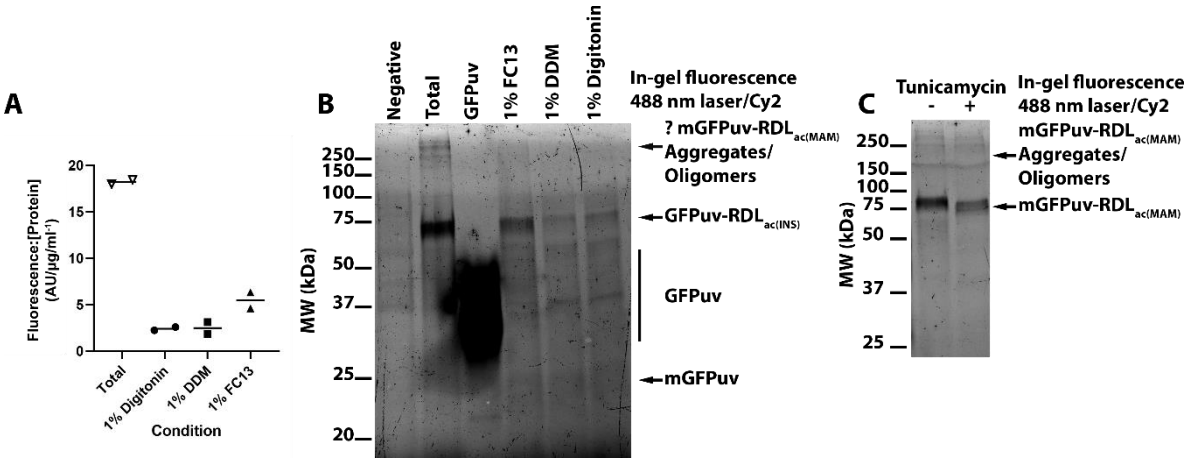


Figure 4.12: mGFPuv-RDL$_{ac(MAM)}$ folding screen assay and tunicamycin treatment

A-B: HEK293 cells transfected with mGFPuv-RDL$_{ac(MAM)}$ were solubilised at 1x10^7 cells/ml in buffer B plus 1% digitonin, DDM or FC13 for 2 hours with gentle mixing at 4 °C. The samples were then ultra-centrifuged to remove insoluble material. The total sample consisted of cells without any detergent added which were lysed by needle sheering and not ultra-centrifuged. **A:** Scatter dot-plot of the samples fluorescence:[protein] measurements, calculated by dividing the fluorescence in AU by the protein concentration in µg/ml as quantified by a BCA assay. Readings were recorded in duplicate at 405/510 nm with the individual reads shown as dots and the mean as a horizonal bar. **B:** SDS-PAGE in-gel GFP fluorescence of the folding screen samples. ~20 µg of protein was loaded per lane except for the purified GFPuv where 20 ng was loaded (note that a sample of mGFPuv was unavailable so GFPuv dimers are shown, although these cannot form in the mGFPuv tagged samples). The negative control consisted of mock transfected HEK293 cells solubilised in 2% DDM. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set. **C:** SDS-PAGE in-gel fluorescence of *Sf*9 cells transfected with mGFPuv-RDL$_{ac(MAM)}$ grown in the presence or absence of 10 µg/ml tunicamycin. The cells were lysed by needle shearing at ~1x10^7 cells/ml. ~20 µg of protein was loaded per lane.

As a first experiment with mGFPuv-RDL$_{ac(MAM)}$, a folding screen assay was carried out as with the previous construct in insect cells to gauge the proportion of potentially folded protein. Fluorescence:[protein] measurements (Figure 4.12A) suggested that approximately 13% of the total mGFPuv-RDL$_{ac(MAM)}$ could be extracted by 1% digitonin or DDM. 1% FC13 extracted approximately 30% of the total. Although not directly comparable as the constructs slightly differ, for 1% digitonin and DDM this was approximately twice that of the previous insect cell construct and potentially suggests higher levels of folded protein. After running these samples on an SDS-PAGE gel (Figure 4.12B), it was possible to see monomeric mGFPuv-RDL$_{ac(MAM)}$ in the total and all the detergent samples by in-gel fluorescence. There appeared to be little free mGFPuv visible. As suggested from the approximately one third lower fluorescence:[protein] ratio of the mGFPuv-RDL$_{ac(MAM)}$ total sample, the signal strength was much weaker than with GFPuv-RDL$_{ac(INS)}$ and the high laser gain required raised the background fluorescence. This illustrates an unfortunate disadvantage of mGFPuv as

although it possesses enhanced stability, compared to later XFP derivatives it is substantially dimmer. Blotting the gel and probing for the twin-strep tag was unsuccessful as the signal strength was too weak to draw any conclusions (data not shown). Growing the cells in the presence of 10 µg/ml tunicamycin (Figure 4.12C) reduced the mobility of the mGFPuv-RDL$_{ac(MAM)}$ monomer on in-gel fluorescence compared to un-treated cells, similar to with GFPuv-RDL$_{ac(INS)}$. This suggested that the protein was N-linked glycosylated in HEK293 cells as expected.

Sub-cellular fractionation experiments were not carried out as for an adherent format it would have been prohibitively expensive in terms of resources to obtain enough cells. Confocal fluorescence microscopy of PFA fixed cells (Figure 9) however suggested the presence of the receptor on the membrane, as an anti-StrepII antibody bound un-permeabilised cells and the signal co-localised with that of the mGFPuv. In permeabilised cells, both of these signals overlapped with that of the plasma membrane marker Na^+K^+ATPase. It was noted though that there did appear to be a significant amount of free mGFPuv in the cytoplasm as well, which did not co-localise with anti-StrepII and is therefore suspected to represent proteolytically cleaved mGFPuv.

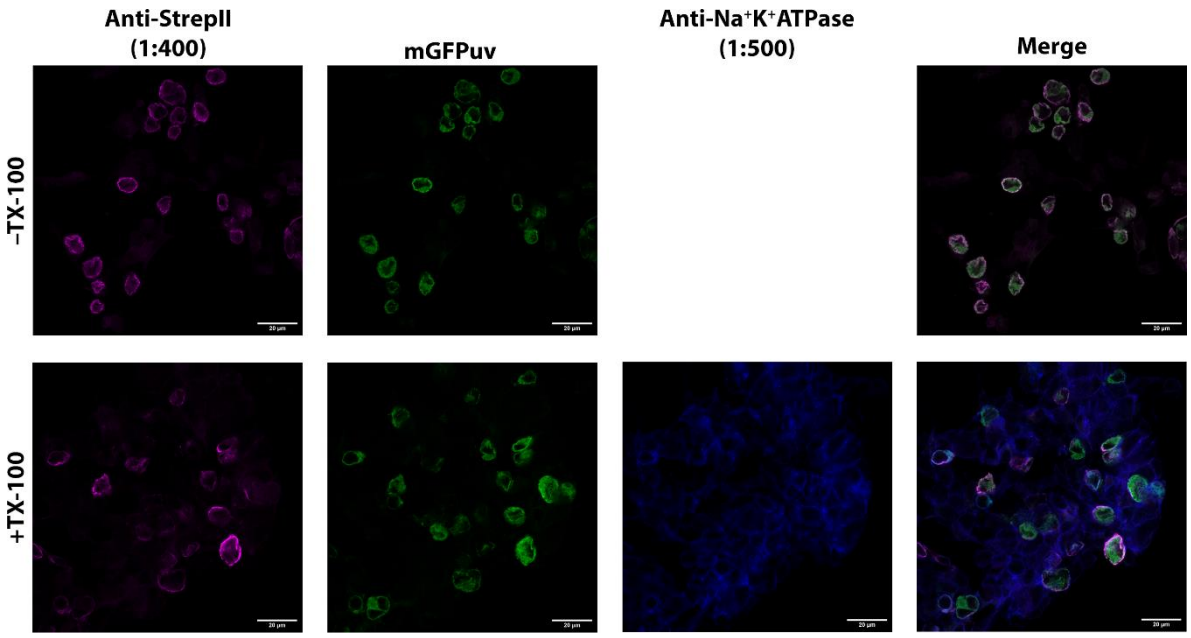


Figure 4.13: HEK293 mGFPuv-RDL$_{ac(MAM)}$ indirect immunofluorescence.

HEK293 cells transfected with mGFPuv-RDL$_{ac(MAM)}$ were fixed with PFA and either permeabilised with 0.1% TX-100 or left untreated. They were then probed with antibodies against the twin-strep tag (AF594 channel in -TX-100, AF680 channel in +TX-100) and the plasma membrane marker Na^+K^+ATPase (AF594 channel). Note that that the Na^+K^+ATPase antibody targets an intracellular

epitope, so it could not be used with the non-permeabilised cells and hence no image is shown. The slides were prepared by Jordana Freemantle and imaged by Dr. Alan Prescott.

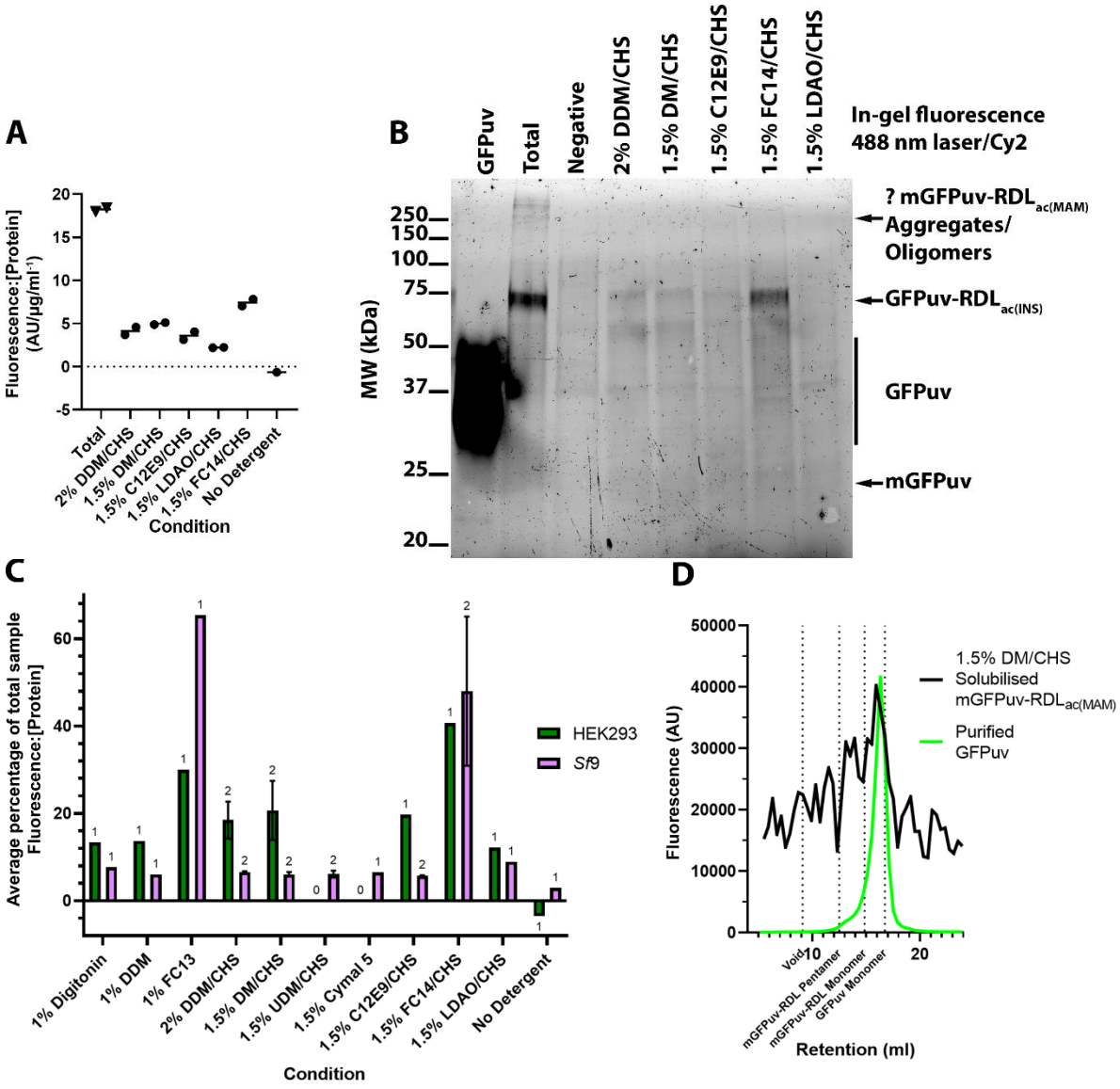


Figure 4.14: mGFPuv-RDL$_{ac(MAM)}$ detergent solubility screening

A-B: Detergent screening of whole cell samples expressing mGFPuv-RDL$_{ac(MAM)}$. HEK293 cells transfected with mGFPuv-RDL$_{ac(MAM)}$ were solubilised at 1x10^7 cells/ml in buffer B plus the indicated detergents (each of which included CHS at ~1/10th-1/5th the detergent concentration) for 2 hours with gentle mixing at 4 °C. The samples were then ultra-centrifuged to remove insoluble material. The total and no detergent samples consisted of transfected cells without any detergent added which were lysed by needle sheering, the no detergent sample only was then ultra-centrifuged. **A:** Scatter dot-plot of the samples fluorescence:[protein] measurements, calculated by dividing the fluorescence in AU by the protein concentration in µg/ml as quantified by a BCA assay. Readings were recorded in duplicate (except for the no detergent sample which was a single read) at 405/510 nm with the individual reads shown as dots and the mean as a horizonal bar. **B:** SDS-PAGE in-gel GFP fluorescence of the detergent screen samples. The negative sample is 2% DDM solubilised mock transfected HEK293 cells. ~20 µg of protein was loaded per lane except for the purified GFPuv where 20 ng was loaded (note that a sample of mGFPuv was unavailable so GFPuv dimers are shown with the positive control, although these cannot form in the mGFPuv tagged samples). The gel was excited with a 488 nm laser and imaged with a Cy2 filter set. **C:** Comparison of the detergent extraction effectiveness for GFPuv-RDL$_{ac(INS)}$ and mGFPuv-RDL$_{ac(MAM)}$ expressed as a percentage of the total samples fluorescence:[protein] ratio. The n for each sample is indicated above each bar, error bars represent standard error. Where no error bar is visible then it is smaller than the size of the line

or n = 1. **D:** FSEC analysis of 1.5% DM/CHS solubilised mGFPuv-RDL$_{ac(MAM)}$ and purified GFPuv. At 4 °C each sample was loaded onto a Superose 6 10/300GL size exclusion column equilibrated in buffer B plus 0.05%/0.01% DDM/CHS via a 500 µl loop and eluted at into black 96 well microplates as 200 µl fractions. For the purified GFPuv sample, 70 µg was loaded onto the column. After fractionation the fluorescence of each microplate was read at 405/510 nm at a gain calculated for the most intense well to reach 70% of the platereaders maximum dynamic range. The traces presented are therefore not quantitative and show the average of every two data-points to reduce noise. The retention volumes for the column void, pentamers (490 kDa) and monomers (98 kDa) of mGFPuv-RDL$_{ac(MAM)}$ and GFPuv monomers (27 kDa) as calculated from standards are indicated.

To identify the best conditions for extraction of mGFPuv-RDL$_{ac(MAM)}$, a detergent screen with a similar panel to that previously used with GFPuv-RDL$_{ac(INS)}$ was used. Fluorescence:[Protein] measurements (Figure 4.14A) suggested that 2% DDM/CHS, 1.5% DM/CHS and 1.5% C12E9/CHS were approximately equipotent in extracting about 20-25% of the signal in the total sample. As expected, 1.5% FC14/CHS was most efficient at ~40%, but surprisingly 1.5% LDAO/CHS was relatively inefficient at ~10%. This pattern was also visible on in-gel fluorescence (Figure 4.14B) and only an extremely faint band was visible for the LDAO/CHS sample. The DDM, DM and C12E9 samples only appeared to show an upper band compared to the total and FC14, so they potentially preferentially extracted the glycosylated form of the protein. When compared to GFPuv-RDL$_{ac(INS)}$ on a normalised scale (Figure 4.14C), the fluorescence:[protein] measurements for mGFPuv-RDL$_{ac(MAM)}$ suggested that non-ionic detergents appeared to extract approximately at least twice as much of the total sample.

Attempts were made to use FSEC analysis with samples of detergent solubilised mGFPuv-RDL$_{ac(MAM)}$ to determine the protein dispersity and oligomeric states, but the low signal strength made traces difficult to interpret from background noise and caution has to be taken with extrapolating from the data. Only a relatively successful trace for protein solubilised in 1.5% DM/CHS was eventually obtained (Figure 4.14D), although the resolution was just sufficient to distinguish mGFPuv and a possible oligomeric species of mGFPuv-RDL$_{ac(MAM)}$ which eluted between that calculated for a pentamer and a monomer. In-gel fluorescence had suggested that free mGFPuv levels were low, contrasting with the microscopy, which FSEC appeared to corroborate better with. Further FSEC attempts were abandoned as it was felt that using further resources was unjustified (for context, each sample used almost a full T-75 flask of HEK293 cells in order to obtain enough cells).

4.3.2.2: Production of mGFPuv-RDL$_{ac(MAM)}$ in suspension Expi293 cells

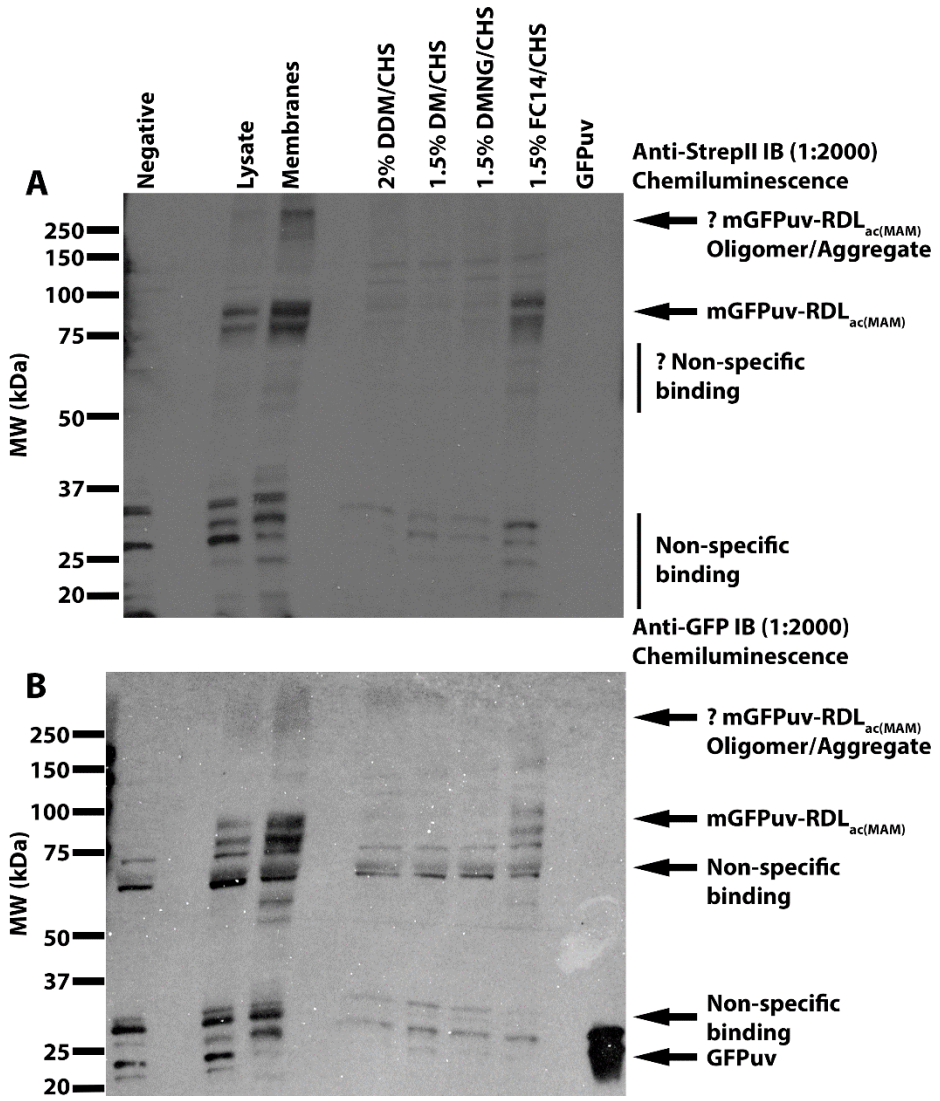


Figure 4.15: Detergent solubilisation of mGFPuv-RDL$_{ac(MAM)}$ in Expi293 membranes

Membranes were extracted from Expi293 cells transfected with mGFPuv-RDL$_{ac(MAM)}$ and solubilised at 1 mg/ml in the buffer B plus indicated detergents for 2 hours at 4 °C. Insoluble material was then removed by ultra-centrifugation. The detergent soluble fractions were subsequently concentrated with 100 kDa centrifugal concentrators to enable loading more protein. The negative control consisted of un-transfected Expi293 cells solubilised in 2% DDM. ~30 µg of protein was loaded per lane except for the purified GFPuv positive control where 200 ng was loaded. Following transferring and blotting for the twin-strep tag with anti-StrepII (**A**), the blot was stripped. After confirming the antibodies removal by re-addition of ECL and exposure to film, the blot was re-probed with anti-GFP (**B**).

Since it was demonstrated that mGFPuv-RDL$_{ac(MAM)}$ could be relatively efficiently extracted, it was decided to switch to using HEK293 cells in a suspension format to gain further material for purification purposes. As it would have been prohibitively expensive to use lipofectamine or a similar lipid agent for transfection purposes with the volumes of cells utilised, these experiments used linear PEI.

Such an approach has been previously used with PLGICs for high level production (Miller and Aricescu, 2014; Kasaragod *et al.*, 2022). Initial experiments (not shown) suggested that using a 1:4 DNA:PEI ratio gave maximal levels of protein when combined with adding sodium butyrate to 10 mM after 12 hours and harvesting after 48 hours.

Extracting the membranes of Expi293 cells transfected with mGFPuv-$RDL_{ac(MAM)}$ showed that the signal was enhanced in this fraction by anti-strepII and anti-GFP Western blot (Figure 4.15). Following detergent solubilisation, it was possible to see a distinct double band at the expected MW for each deterrent sample, although loading at least 30 µg of total protein per lane was needed for a sufficient signal strength with the non-ionic detergent samples. It was therefore necessary to concentrate each detergent soluble sample with 100 kDa centrifugal concentrators to be able to load enough. Although equal amounts of protein were loaded, the samples are as such not directly comparable to each other on the blot as there will have likely been loses of lower MW proteins during the concentration. The expected trend of FC14/CHS being more effective than non-ionic detergents was still distinct though.

Following on, using 200-250 ml cultures of cells transfected with mGFPuv-$RDL_{ac(MAM)}$ it was attempted to purify the protein with a 1 ml Strep-trap column after detergent extraction from membranes. As Expi293 media allows cells to be cultured at higher densities these were equivalent to 400-500 ml cultures of conventional suspension HEK293 cells. 2% DDM/CHS and 1.5% DMNG/CHS were initially tried due to their previous use with PLGICs and being more likely to only extract folded protein. To try and stabilise the protein, the cells were grown in the presence of 10 µM of the channel blocker insecticide lindane (Sigma-Aldrich) and this was also included in all buffers during extraction and purification. Claxton and Gouaux (2018) used a similar strategy with the channel blocker picrotoxin, but lindane was used as a safer alternative. The wash buffer consisted of buffer B plus 0.05/0.005% DDM/CHS and this was also used as the elution buffer with 10 mM desthiobiotin added. For both detergents, only extremely small (~1-2 µg or less) quantities were obtained following concentration of the elute. While faint bands could be seen in the elute sample on gels, the following blots were inconclusive.

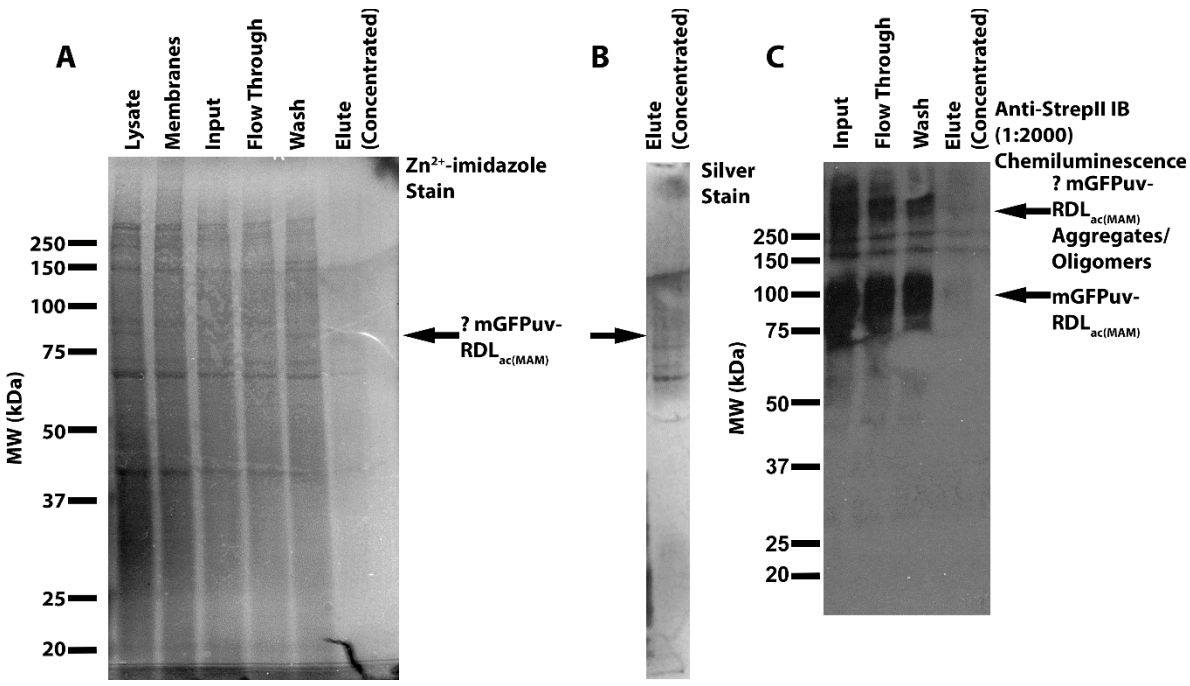


Figure 4.16: mGFPuv-RDL$_{ac(MAM)}$ streptactin purification from FC14/CHS solubilised Expi293 membranes

Membranes were extracted from Expi293 cells transfected with mGFPuv-RDL$_{ac(MAM)}$ and solubilised at 1 mg/ml for 2 hours in buffer B plus 1.5% FC14/CHS at 4 °C. After removal of insoluble material, the detergent soluble fraction (input) was passed through a 1 ml Strep-trap column, the column was then washed with 10 column volumes of buffer B supplemented with 0.05/0.005% DDM/CHS and proteins subsequently eluted with 3 column volumes of the same buffer including 10 mM of desthiobiotin. The eluted protein was then concentrated with 100 kDa cut-off centrifugal concentrators. Samples taken throughout the purification cycle were applied to SDS-PAGE gels and **A:** zinc-imidazole stained **B:** silver stained and **C:** transferred to a membrane and the twin-strep tag blotted for with anti-strepII. ~10 µg of total protein was loaded per lane in A. ~0.75 µg was loaded in the elute (concentrated) lane of A and B. ~30 µg of sample was loaded per lane in C, except for elute (concentrated) where ~0.75 µg was loaded. Note that it was necessary to concentrate the input, flow through and wash samples with 100 kDa cut-off centrifugal concentrators to be able to load enough protein.

At this point it was decided to cease work on trying to isolate purified recombinant *Dm* RDL because it was seen as unlikely to be feasible using Expi293 cells based on the data obtained. As a test exercise it was decided to try purifying membranes solubilised in 1.5% FC14/CHS to determine if the purification procedure could be successful when more material was present, even if a proportion was likely to be un-folded or de-stabilised. Following this, by A280 measurement the eluted protein could not be detected in any of the individual fractions, but when they were combined and concentrated down with a 100 kDa centrifugal concentrator, a concentration of ~0.09 mg/ml in a volume of ~80 µl was obtained. By Zn^{2+}-imidazole (Figure 4.16A) and silver staining (Figure 4.16B) it was possible to see a faint band at the MW expected for the protein. There was heavy smearing visible on the gel with both staining methods, suggesting potential co-purification

of interacting partners. Western blotting for the twin-strep tag (Figure 4.16C) showed the presence of the protein in decreasing levels in the detergent soluble, flow through and wash fractions. This suggested that only a proportion of the soluble material had efficiently bound to the column resin. There was however a faint band visible at the expected MW for the protein in the concentrated elute at the same position as in the silver staining, overall implying successful purification of a small amount of mGFPuv-RDL$_{ac(MAM)}$. Insufficient sample was available for running on a size exclusion column or for BN-PAGE and this experiment pre-dated the availability of mass photometry, so the oligomeric state of the purified FC14/CHS solubilised mGFPuv-RDL$_{ac(MAM)}$ is unfortunately unknown.

To confirm the presence of the protein in the final concentrated sample, it was submitted to the in-house mass spec facility. As the amount of protein available was considered too small for a reliable in-gel digestion approach, an in-solution trypsin digest followed by LC-MS/MS was used for the sample. While detergents are traditionally considered to be unsuitable for mass-spec, it has been determined that DDM/CHS at 0.05/0.005% does not adversely affect the spectra (Zhang and Miller, 2015) so trypsin was added directly to the sample overnight. The mass spectra revealed that 10 statistically significant unique peptide matches corresponding to those expected from trypsin digestion of mGFPuv-RDL$_{ac(MAM)}$ were present. There was 15% sequence coverage, including sections of the mGFPuv, the ECD, the transmembrane helices, the ICD and the last 20 residues of the C-terminus containing portions of twin-strep tag.

4.3.0.0: Discussion

It was attempted to establish a method for recombinant production and purification of the insect GABA$_A$ homolog RDL from *Dm*. Despite extensive efforts, the protein appeared to be only minimally soluble in mild non-ionic detergents when produced in *Sf*9 insect cells using baculovirus and could only be extracted efficiently with harsh zwitterionic detergents. When transfecting the protein into HEK293 mammalian cells however, approximately 2-5x as much soluble protein could be extracted by non-ionic detergents. This was offset though by the lower levels of total protein produced and initial purification trials suggested

that the methods developed to date would be unfeasible for high level production of *Dm* RDL.

These efforts reflect the general difficulty in working to recombinantly produce, detergent extract and purify functional eukaryotic membrane proteins. While there are past studies of *Dm* RDL function in baculovirus infected insect cells (Lee *et al.*, 1993; Zhang *et al.*, 1995), stably transfected HEK293 cells (Gassel *et al.*, 2014) and numerous other systems (Millar *et al.*, 1994; Shotkoski *et al.*, 1996; Grolleau and Sattelle, 2000), literature on solubilisation and purification of *Dm* RDL from insect cells using baculovirus transduction is unfortunately limited to a single study (Lee *et al.*, 1993) and no studies using HEK293 or other mammalian cells could be identified. It is difficult though to make comparisons of the data in this chapter against the results presented by Lee *et al.*, who used a combination of 1% TX-100 plus the denaturing anionic detergents 0.5% deoxycholate and 0.1% SDS for extraction and immunoprecipitated using a polyclonal antibody raised against *Dm* RDL. Their results do show some similarities however, including to whole cell/membrane and fos-choline solubilised samples on western blotting and in-gel GFP fluorescence, particularly the protein running as a distinctive double band. Lee *et al.* speculated that the upper band represented a glycosylated form and in the current work this was confirmed as in both *Sf*9 and HEK293 cells incubation with tunicamcyin reduced the upper band. This matches predictions that at least one N-linked glycosylation site is present on *Dm* RDL (Jones and Sattelle, 2007).

It is curious as to why *Dm* RDL can be relatively efficiently extracted from HEK293 cells using mild non-ionic detergents, but not from *Sf*9 cells. This is despite both producing protein which appeared to be trafficked to the cell surface and was at least partially glycosylated. Although protein function was not tested, studies referred to in the previous paragraph suggest that the protein would likely be active. The reasons for the differences between the two systems are mystifying. While direct comparisons of production in insect and mammalian cells are rare with PLGICs, there is the similar example of Claxton and Gouaux (2018), who found that a construct of the human $\alpha_1\beta_1$ $GABA_A$ receptor could be only minimally solubilised from baculovirus infected *Sf*9 or High Five™ cells. This was despite extensive optimisation of variables including culture temperature/time, multiplicity

of infection and using a wide variety of non-ionic detergents. Production in suspension HEK293 cells using the BacMam baculovirus system in comparison showed successful solubilisation with 2% DDM/CHS and a monodisperse FSEC profile; This eventually led to purified receptor being produced. Interestedly though, successful non-ionic detergent solubilisation of functional mammalian $GABA_A$ receptors produced in insect cells using baculovirus has been previously reported for multiple subtype combinations (Carter *et al.*, 1992; Knight *et al.*, 1998, 2000; Elster *et al.*, 2000); furthermore, in some cases milligram levels of protein are claimed to have been purified (Kang *et al.*, 2008). The reasons for the success in *Sf*9 cells with these studies and not others are unclear.

Data in this chapter from the folding screen assays suggest that a potential reason for this difference in extraction efficiency could be that a higher proportion of *Dm* RDL produced in *Sf*9 cells may be un-folded compared to in HEK293 cells. This was based on an assay using detergents with differing degrees of ability to solubilise un-folded protein. The similar inability of non-ionic detergents to efficiently solubilise mammalian and insect PLGICs produced in insect cells with baculovirus has been a repeated issue with past studies in the Hunter lab (Trumper, 2014; Gottschald Chiodi, 2017; Jones, 2021).

If un-folded protein is the issue though, this is certainly not a problem which is unique to PLGICs. Indeed, using insect cells with baculovirus transduction for membrane protein production has been frequently linked to high levels of inactive, unfolded proteins which are incompletely processed in terms of PTMs, especially glycosylation (Tate *et al.*, 1999, 2003; Higgins *et al.*, 2003; Thomas and Tate, 2014; Lemaitre *et al.*, 2019). In theory, un-folded protein should be retained by the ER and exported to the cytoplasm by the ER-associated degradation (ERAD) system for ubiquitination and targeting to the proteosome for degradation (Brodsky and Skach, 2011). The un-folded protein is likely more a factor of the baculovirus infection though than a property of the insect cells themselves. Baculoviruses act as lytic viruses, which hijack the secretory pathway for their own replication. The secretory pathway through the ER is where membrane proteins are folded, assembled and post-translationally modified before trafficking to the surface. Viral infection and production stresses and slows these processes. (HU, 2005; Brodsky and Skach, 2011). There is also a secondary factor which

further stresses the secretory pathway in that the very late polyhedrin promoter used with most baculovirus vectors, including pFastBac 1 as utilised with RDL, strongly drives gene expression and the levels of mRNA transcribed are likely beyond what the cells can efficiently process. This does not appear to be an issue with some samples though. The late nature of the promoter also means that maximal protein production does not start until the cells are close to death from the virus infection. This also coincides with the secretion of chitinase, which is involved in the cells eventual lysis to release more virus. The combination of the excessive mRNA, secretion of chitinase and the stress on the ER from the viral infection is thought to overload the cellular machinery and quality control processes in the ER, resulting in greater levels of un-folded protein being exported and decreasing levels of protein being correctly processed (HU, 2005; Thomas and Tate, 2014; Lemaitre *et al.*, 2019).

There is also the valid point that the fused GFPuv could affect the protein folding, although this is a near universally used strategy these days for PLGIC research and fluorescent proteins have been successfully fused on the N-terminus (Claxton and Gouaux, 2018; Bernhard and Laube, 2020) previously. As an important addition to this point, unlike in *E. coli* an attached FP cannot be used as a folding indicator for a protein in eukaryotic cells. *E. coli* target un-folded proteins to aggregate in inclusion bodies, in which the GFP is unable to fluoresce (Drew *et al.*, 2001), but eukaryotic cells retain un-folded protein via the ERAD. The FP will still fluoresce however (Illing *et al.*, 2002; Saliba *et al.*, 2002), so while the fluorescence is useful for quantifying protein levels and localisation in eukaryotic cells, it does not act as a folding indicator.

It is accepted though that the folding screen assay used with *Sf*9 and HEK293 cells in this chapter can only show a general semi-quantitative trend in the level of potentially folded protein based on detergent solubility and does not directly measure folded protein. Other techniques therefore will need to be used to confirm if the protein folding state is the cause of the poor solubility of *Dm* RDL produced in *Sf*9 cells.

There are two differences between *Sf*9 and HEK293 cells which could also be factors. The first concerns the differing glycosylation patterns. *Sf*9 cells typically do not make hybrid and complex glycans like mammalian cells, with high

mannose types dominating instead (Shi and Jarvis, 2007). As to whether this would be a factor with *Dm* RDL is questionable though, as *in-vivo Dm* mainly use similar high mannose glycosylation patterns to *Sf*9 cells and other invertebrates (Katoh and Tiemeyer, 2013). The other aspect concerns the cells lipid profiles. Aside from possessing differing distributions of fatty acids and phospholipids, mammalian cells have an approximately 10-20x higher cholesterol/phospholipid molar ratio than *Sf*9 cells (Gimpl *et al.*, 1995; Marheineke *et al.*, 1998). The lipid profile of the plasma membrane can have major effects on protein folding and function and cholesterol in particular is known to be important for the function of $GABA_A$ receptors. In dissociated rat hippocampal neurones for instance, depletion or enrichment of cholesterol caused a reduction in the EC50 for GABA (Sooksawate and Simmonds, 2001). In crystal and cryo-EM structures of $GABA_A$ and nACh receptors, CHS has also been modelled as binding on the TMHs (Laverty, 2016). The lipid profile of *Dm* cells is more similar to that of *Sf*9 cells than mammalian cells (Rietveld *et al.*, 1999; Hammad *et al.*, 2011), with a much reduced level of cholesterol. it is therefore unclear whether the lipid profile of HEK293 cells could be responsible for more *Dm* RDL being extractable and indeed would actually point to *Sf*9 as cells being closer to the proteins native environment.

On a related note though, one other possibility for the detergent insolubility of *Dm* RDL in *Sf*9 cells is that it is partitioned into lipid rafts. These are also less correctly termed as detergent resistant membranes based on the original finding of lipid membranes which could not be solubilised with non-ionic detergents at 4 °C (Brown and Rose, 1992), although this was later found to not be universally applicable and being detergent resistant is now considered as a property of some membranes rather than lipid rafts *per se* (Lichtenberg *et al.*, 2005). *Drosophila* membrane lipid rafts have a lipid profile that curiously seems to match more closely that of mammalian lipid rafts, with an increased cholesterol content compared to membranes of *Drosophila* or other insect cells (Rietveld *et al.*, 1999). No studies could be found investigating if *Dm* RDL is associated with lipid rafts *in-vivo*, but various mammalian ligand gated ion channels have been associated with these or are at least known to be detergent resistant in *ex-vivo* tissue (Li *et al.*, 2007; Nothdurfter *et al.*, 2010). It should be noted though that solubilisation of GFPuv-$RDL_{ac(INS)}$ at room temperature rather than 4 °C did not seem to

improve extraction with either 2% DDM/CHS or 1.5% FC14/CHS. Under room temperature conditions it would be expected that a higher proportion of lipid rafts would be solubilised (Lichtenberg *et al.*, 2005; Ayuyan and Cohen, 2008), although this was not investigated further and would require selective isolation of lipid rafts by sucrose density gradients for instance.

Chapter 5: High Level Recombinant Production of the Purified Human GABA Transporter 1

5.1.0.0: Overview

In the SLC6 neurotransmitter transporter family, while some sub-families such as the monoamine transporters are well characterised in terms of direct structural information and are the targets of multiple clinically relevant drugs, this is not case for other sub-families. The GABA transporter sub-family is a prominent example of this point. Despite the ubiquity of GABA as the main inhibitory neurotransmitter in the CNS, only a single clinically approved agent specifically targets a member of the GABA transporter sub-family. The GABA transporter sub-family also represent six of the 20 SLC6 transporters, yet no structures were available in the PDB for this branch until recently. During the corrections process for this thesis a 3.8 Å resolution cryo-EM structure of GAT1 in complex with tiagabine was published (Motiwala *et al.*, 2022).

As has been previously described in Chapter 1 (1.3.2.2), GAT1 represents a target for the control of epilepsy and neuropathic pain. However, the only drug in clinical use which specifically targets GAT1 is tiagabine and this only has limited uses due to side effects and a lack of efficacy in many epilepsy patients. The mechanism of action for tiagabine at GAT1 is unclear. It has been suggested to have both competitive and non-competitive antagonist properties, as well as potentially acting as an allosteric modulator (Kragholm *et al.*, 2013).

It has proven difficult to develop GABA transporter inhibitors in general, many of those which have been identified show low affinity and poor subtype selectivity (Bröer and Gether, 2012). This difficulty is partly because direct structural information has not been available, so the sub-type specific molecular mechanisms which define the transporters functions are poorly understood. An increasing number of structures of other SLC6 transporters are available in a variety of conformations and bound with differing ligands. Unfortunately, despite the high sequence similarity and analogous structural architecture, it has been challenging to translate the specific details of molecular mechanisms between transporter subtypes (Navratna and Gouaux, 2019; Colas, 2020).

A structure of GAT1 would give an increased understanding of the molecular mechanisms which enable transport of GABA and other substrates. Other novel potential allosteric sites may also be identified. Collectively, these could assist in the design of future GAT1 selective inhibitors with an improved clinical profile compared to tiagabine. A prior necessity however is to establish an efficient system for recombinant production of GAT1 as there have only been sporadic reports in the literature of high level GAT1 production (Hu *et al.*, 2017).

5.2.0.0: Aims

- Recombinantly produce purified GAT1 for structural and binding studies.

- Raise a recombinant nanobody against GAT1 to increase the size of the protein.

- Determine if the GAT1 constructs used represent active protein.

5.3.0.0: Results

5.3.1.0: Production of the human GABA transporter 1 in *Sf*9 cells with baculovirus using a polyhedrin promoter

For production of GAT1, initial efforts used *Sf*9 insect cells with baculovirus transduction using the conventional polyhedrin promoter-based vector pFastBac 1 as with other proteins in this thesis. The gene was ordered with a C-terminal short GGGSGGGS flexible linker, followed by a TEV-mEmerald, a GG linker and a twin-strep tag for purification purposes. This protein is hence referred to as GAT1-mEm. The initiating Methionine was deleted off the fused mEmerald to prevent production of free mEmerald from any erroneous initiation. In case it was necessary to also evaluate production in mammalian cells, the gene was ordered in the pTwist CMV WPRE Neo mammalian expression vector and a mammalian Kozak sequence was placed in front of the initiating methionine. While the C-terminal tags disrupted the PDZ motif, PALS1 is not present in either *Sf*9 or HEK293 cells when performing a BLAST on the genomes, so it was thought to be unlikely that this would cause any major abject effects. Other studies have also successfully used similarly designed constructs (Cai *et al.*, 2005). The gene was sub-cloned into the pFastBac 1 vector, - this is hereafter designated as GAT1-mEm$_{(PFB)}$ -, with the mammalian Kozak sequence retained as literature suggested that this would not cause a problem with insect cells (Hills and Crane-Robinson, 1995). Note that additional data regarding initial extraction and purification of GAT1-mEm$_{(PFB)}$ in SMA co-polymer is presented in the appendix S2.

5.3.1.1: Identifying the optimal conditions for culturing and extracting the human GABA transporter 1 from *Sf*9 cells

After producing baculovirus, initial test Western blots (not shown) with a GFP antibody suggested the presence of a GFP tagged species around 75 kDa, approximately three times the size of an mEmerald monomer (27 kDa). Confocal fluorescence microscopy and sub-cellular fractionation and was used to identify the localisation of the protein. GAT1 appeared to be on the cell surface (Figure 5.1A) and enriched in the isolated membrane fraction on anti-GFP Western blots (Figure 5.1B). These experiments used buffer C (25 mM HEPES, 200 mM NaCl,

1 mM EDTA, 2 mM GABA, 5% glycerol pH 7.5 [4 °C]), except where noted this was used for all GAT1 related experiments. While there appeared to be some protein present in the cytoplasm with microscopy and a small signal at the expected MW for free mEmerald in the cytosolic fraction on Western blot, nothing was visible at the MW for GAT1-mEm and the free mEmerald was only faintly visible in the lysate and membrane samples.

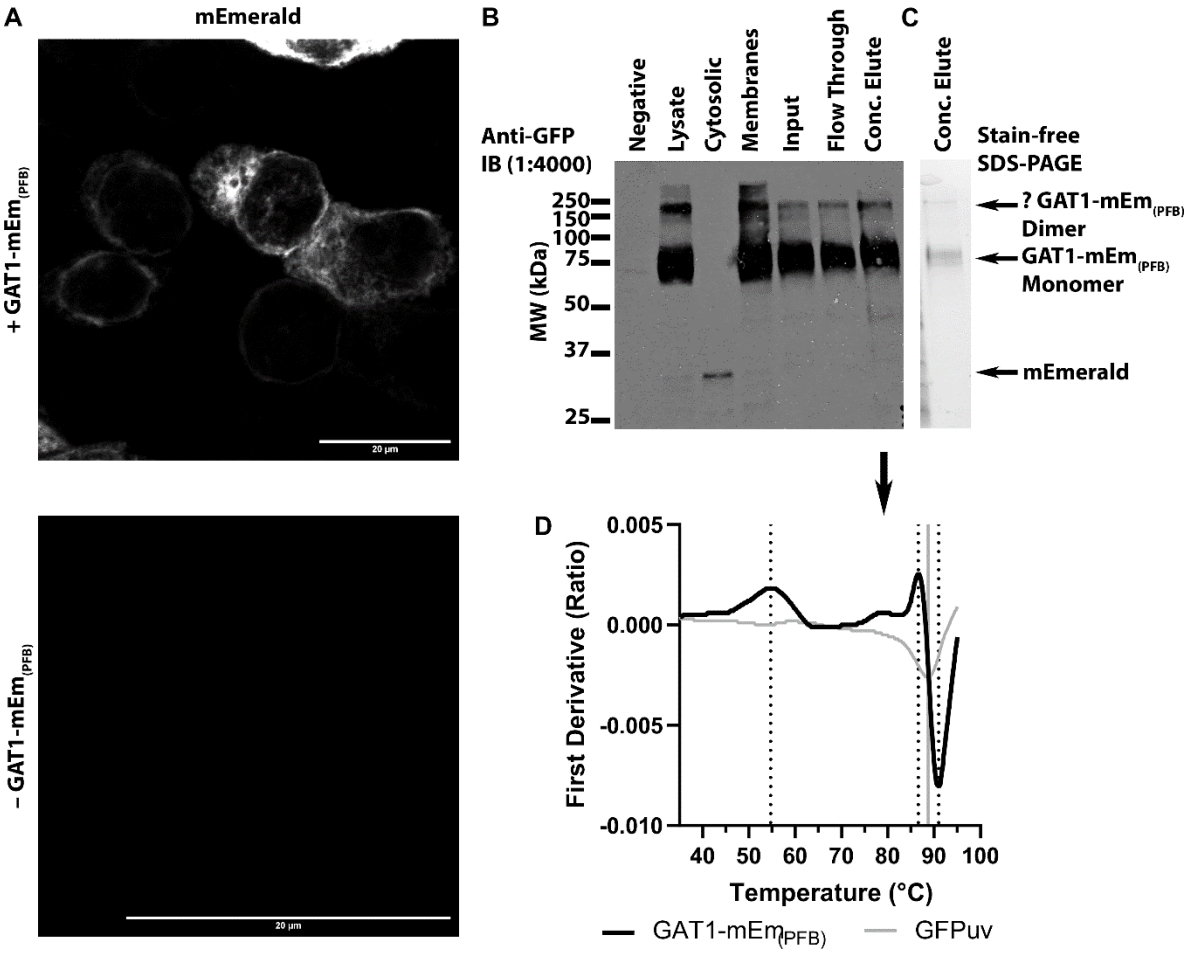


Figure 5.1: Sub-cellular localisation and initial purification of GAT1-mEm$_{(PFB)}$.

A: Confocal fluorescence microscopy of fixed and permeabilised *Sf*9 cells infected or un-infected with the GAT1-mEm$_{(PFB)}$ baculovirus after 36 hours at 27 °C. The attached mEmerald was excited at 488 nm. Scale bars indicate 20 µm. The slides were imaged by Jordana Freemantle. **B-C:** Sub-cellular fractionation, solubilisation and purification of GAT1-mEm$_{(PFB)}$ infected *Sf*9 cells. Extracted membranes were solubilised in 2.0/0.2% DDM/CHS and purified with streptactin-XT resin after removing insoluble material. 2 µg of protein was loaded in the concentrated elute lane and 20 µg in the other lanes. The negative control consisted of un-infected *Sf*9 cells solubilised in 2% DDM. **B:** Anti-GFP Western blot. **C:** stain-free SDS-PAGE of the concentrated elute. **D:** Nanotemper Tycho intrinsic fluorescence 330/350 nm ratio thermal un-folding profiles of the GAT1-mEm$_{(PFB)}$ concentrated elute sample in **B-C** and a sample of purified GFPuv. Fitted inflection temperatures for GAT1-mEm$_{(PFB)}$ are shown as vertical dotted black lines and for GFPuv as a solid grey vertical line.

As an initial test, the membranes were solubilised in 2%/0.4% DDM/CHS. As this appeared to efficiently extract GAT1-mEm PFB, the protein was purified by

binding to 150 µl of buffer equilibrated Streptactin-XT 50% resin suspension for 1 hour at 4 °C. The resin was washed with buffer C plus 0.05/0.005% DDM/CHS and the bound proteins eluted using buffer C with a pH of 8, a raised glycerol content of 15% and including 0.05/0.005% DDM/CHS and 50 mM of *d*-biotin. On an anti-GFP Western blot (Figure 5.1B), GAT1-mEm could be seen in all fractions at ~75 kDa and there was an additional band at ~150 kDa that was postulated to represent a dimer. There appeared to be little to no free mEmerald visible. The concentrated elute sample on stain-free SDS-PAGE (Figure 5.1C) suggested that there was a good level of purity, with only bands for the monomer of GAT1-mEm and the postulated dimer appearing. The monomer appeared as a double band, although the upper form appeared to be more intense. As three N-linked glycosylation sites are known to be present on GAT1 (Bennett and Kanner, 1997), it was suspected that the upper band was the glycosylated form. The section of the gel lane containing the monomer and the postulated dimer was cut out and submitted for mass-spec protein ID. This revealed 234 unique significant peptides which corresponded to GAT1-mEm, covering 46% of the protein sequence and a wide area of the protein. Analysing a sample of the purified protein on the Nanotemper Tycho (Figure 5.1D) fitted three T_i values at 54.7, 86.6 and 91.0 °C. Based on also analysing a sample of purified GFPuv (as a sample of purified mEmerald was unavailable), where a Ti of 88.7 °C was fitted, the latter two T_i values are suggested to be from mEmerald un-folding.

These initial tests had been conducted with the *Sf*9 cells cultured for 72 hours at a constant 27 °C after infection. To identify the best culture conditions for maximal production, a time/temperature study was carried out. In-gel fluorescence after separation by SDS-PAGE (Figure 5.2A) suggested that 48 hours at 27 °C and 24 hours at 27 °C followed by a temperature shift to 18 °C for a subsequent 48 hours gave an approximately equivalent amount of GAT1-mEm PFB. The protein also appeared to run as a double band, as expected from the three glycosylation sites know to be present on the protein. Treatment of cells with 5 µg/ml of tunicamycin 24 hours after infection confirmed this upper band represented the glycosylated form, as it was abolished compared to un-treated cells (Figure 5.2B). As the upper band appeared to be more prevalent and there was less free mEmerald visible, 24 hours at 27 °C followed by a temperature shift to 18 °C for a subsequent 48 hours was selected as the future culture conditions.

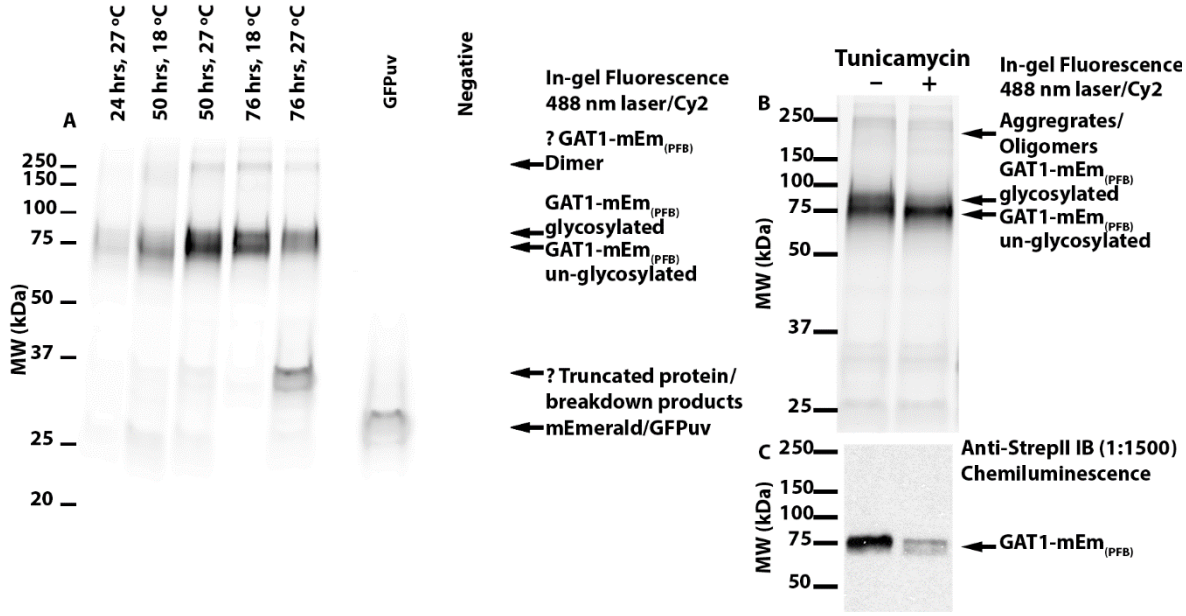


Figure 5.2: GAT1-mEm$_{(PFB)}$ culture time/temperature study and tunicamycin treatment.

A: GAT1-mEm$_{(PFB)}$ time/temperature study in-gel fluorescence. Two cultures of *Sf*9 cells were infected with GAT1-mEm$_{(PFB)}$ baculovirus at 1% and cultured at 27 °C for 24 hours. One culture was then shifted to 18°C. Small samples of cells were taken at the time-points indicated and snap-frozen. After thawing these were solubilised in 2% DDM at 1x10^7 cells/ml for 2 hours, then spun down at 15,000 *g* for 30 minutes. ~15 µg of protein was loaded per lane, except for the purified GFPuv positive control where 50 ng was loaded. The negative control consisted of un-infected *Sf*9 cells solubilised in 2% DDM. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set. **B-C:** Mechanically lysed GAT1-mEm$_{(PFB)}$ infected *Sf*9 cells treated or un-treated with tunicamycin (5 µg/ml). ~20 µg of protein was loaded per lane. **B:** in-gel fluorescence. **C:** anti-strepII Western blot.

As it appeared that GAT1-mEm$_{(PFB)}$ could be efficiently solubilised and was likely folded, further detergent screening was carried out to identify the most suitable agent. A panel of eight detergents was screened against GAT1-mEm$_{(PFB)}$ containing membranes at 1.5 mg/ml, using the fluorescence:[protein] ratio and SDS-PAGE in-gel fluorescence to quantify the amount extracted. To analyse the oligomeric state(s) and dispersity, whole cell samples for seven of the detergents were also run on FSEC.

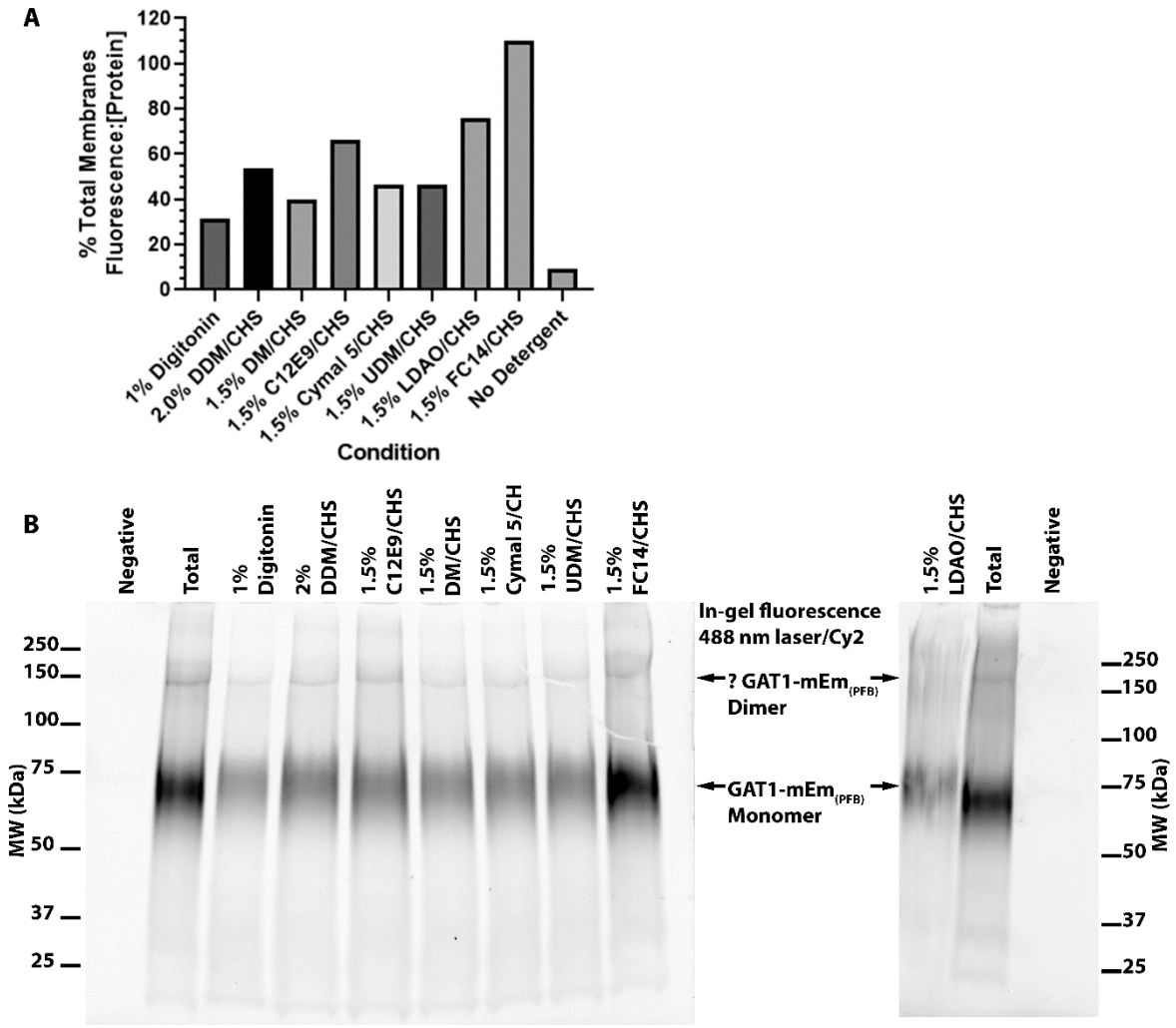


Figure 5.3: GAT1-mEm$_{(PFB)}$ membrane detergent solubility screening.

*Sf*9 membranes containing GAT1-mEm$_{(PFB)}$ were solubilised at 1.5 mg/ml in the indicated detergents made up in buffer C (all of which, except 1% digitonin, contained CHS at ~1/10-1/5th the detergent concentration) for 2 hours with gentle mixing at 4 °C before insoluble material was removed. The total and no detergent samples had buffer added in place of detergent and the total sample was not ultra-centrifuged. **A:** Bar chart of the samples average fluorescence:[protein] ratio expressed as a percentage of that for the total. n= 1 for each bar. **B:** SDS-PAGE in-gel fluorescence of the samples. ~15 µg of protein was loaded per lane. The negative control consisted of un-infected *Sf*9 cells solubilised in 2% DDM. The LDAO sample was separately prepared and loaded on a separate gel with its own total and negative samples. The gels were excited with a 488 nm laser and imaged with a Cy2 filter set.

From the fluorescence:[protein] ratio and in-gel fluorescence (Figure 5.3), the zwitterionic detergents 1.5% FC14/CHS and 1.5% LDAO/CHS appeared be most efficient at extracting GAT1-mEm, to extent that FC14/CHS appeared to pull out >100% of the available protein. With the mild non-ionic detergents, while all appeared to extract >30% of the fluorescence:[protein] ratio for the total sample, 2.0% DDM/CHS and 1.5% C12E9 appeared most efficient at extraction, extracting 53.7% and 66.2 % respectively (Figure 5.3A). There appeared to visually be little difference between the samples however on in-gel fluorescence

(Figure 5.3B), except for the FC14/CHS being more intense as would be expected from the fluorescence:[protein] ratio measurements. All showed a prominent band at the expected position for a GAT1-mEm$_{(PFB)}$ monomer as well as one for the postulated dimer. Free mEmerald was only faintly visible in each sample.

On FSEC (Figure 5.4), all of the tested detergents showed a fairly monodisperse appearance with no signs of aggregation, although with 1.5% DM/CHS and Cymal 5/CHS there appeared to be a small secondary peak. The majority peak, most likely representing GAT1-mEm appeared to elute before that calculated for a monomer or dimer, although little to no signal appeared to elute in the void volume. As the peak did not appear to match the expected retention for a monomer or dimer, likely due to the non-globular shape and formation of protein-detergent micelles, it was difficult to tell the oligomeric state. Curiously, the small secondary peak with DM/CHS and Cymal 5/CHS appeared to elute close to that expected for a GAT1-mEm monomer, although this could be a co-incidence. All samples showed little to no signal at the retention calculated for free mEmerald.

From the detergent screening, it was eventually decided to use 1.5% C12E9/CHS for extraction of GAT1-mEm. This was based on the high level of protein extracted and the monodispersity on FSEC. Being a non-ionic detergent that was unlikely to disrupt protein-protein interactions and more likely to preferentially extract folded protein was another factor in its selection over zwitterionic detergents, despite these extracting more protein. C12E9 was also selected partly for economic reasons as an equivalent amount is approximately 50% of the cost for that of other detergents tested.

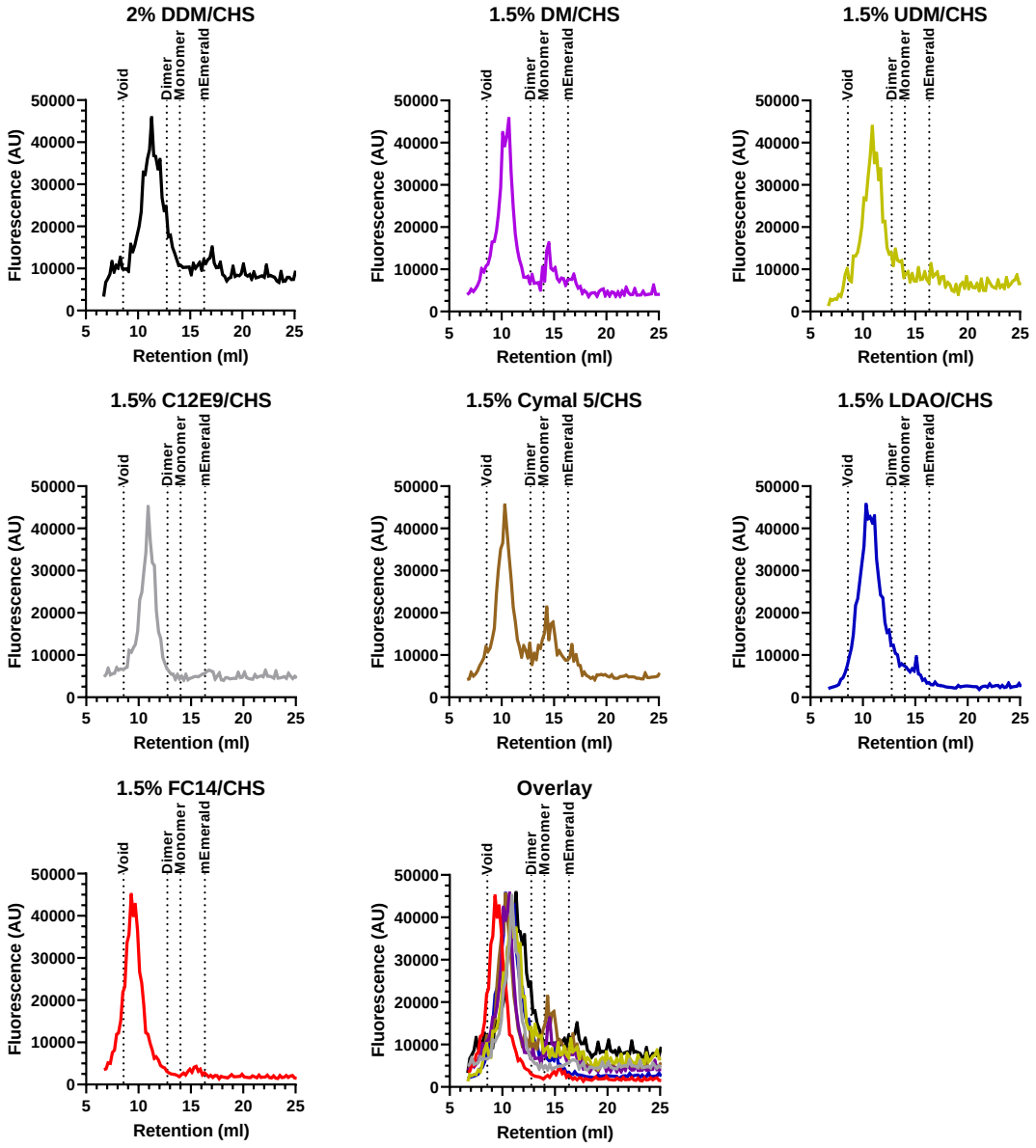


Figure 5.4: Fluorescence size exclusion chromatography of GAT1-mEm$_{(PFB)}$ whole cell detergent soluble samples.

*Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ baculovirus were solubilised at 1x10^7 cells/ml in the detergents indicated in the figure for 2 hours at 4 °C before insoluble material was removed by ultra-centrifugation. At 4 °C 500 µl of each detergent soluble sample was loaded onto a Superdex 200 10/300GL size exclusion column equilibrated in buffer C plus 0.05%/0.01% DDM/CHS via a 1 ml loop and eluted into black 96 well microplates as 200 µl fractions. After fractionation the fluorescence of each microplate was read at 485/510 nm at a gain calculated for the most intense well to reach 70% of the plate readers maximum dynamic range. The traces presented are therefore qualitative rather than quantitative. The retention volumes for the column void, dimers (196 kDa) and monomers (98 kDa) of GAT1-mEm$_{(PFB)}$ and mEmerald monomers (27 kDa) as calculated from standards are indicated. The data in the overlay panel is coloured for each dataset as per the individual entries.

5.3.1.2: Purification of GAT1-mEm$_{(PFB)}$

As the appropriate conditions for culturing and extraction of GAT1-mEm$_{(PFB)}$ had been identified, a 0.8 L culture was grown and processed as described previously. The cells were distinctly light green coloured and after isolation by ultracentrifugation, the membrane pellet was coloured an intense green. This suggested that a high level of GAT1-mEm$_{(PFB)}$ was likely present. The membranes were solubilised in 1.5% C12E9/CHS and after removal of insoluble material, the detergent soluble fraction was purified with 2 ml of buffer washed streptactin-XT 50% resin suspension. The purification procedures were identical to those previously used, aside from the final elution buffer containing 5% glycerol (as in standard buffer C) rather than 15%. After collection of the resin in the gravity flow column, it was noted that the resin was tinted an intense green, indicating that the protein had likely bound. Following elution and concentration with a 100 kDa cut-off centrifugal concentrator, ~44 µg of GAT1-mEm$_{(PFB)}$ were obtained in 250 µl. SDS-PAGE and in-gel fluorescence (Figure 5.5A-B) suggested that the protein was somewhat pure, although additional free mEmerald was also visible as well as some fainter bands. Testing a sample on the Nanotemper Tycho fitted a T_i of 53.7 °C (Figure 5.5C), slightly reduced from the 54.7 °C with the initial DDM solubilised protein, but without any additional T_i values likely corresponding to the mEmerald un-folding. When regenerating the resin, it was noticed that it retained a green tint until several bed volumes of 10 mM NaOH had passed through, so potentially not all the protein had eluted. The NaOH present in the flowthrough precluded analysis of this hypothesis however.

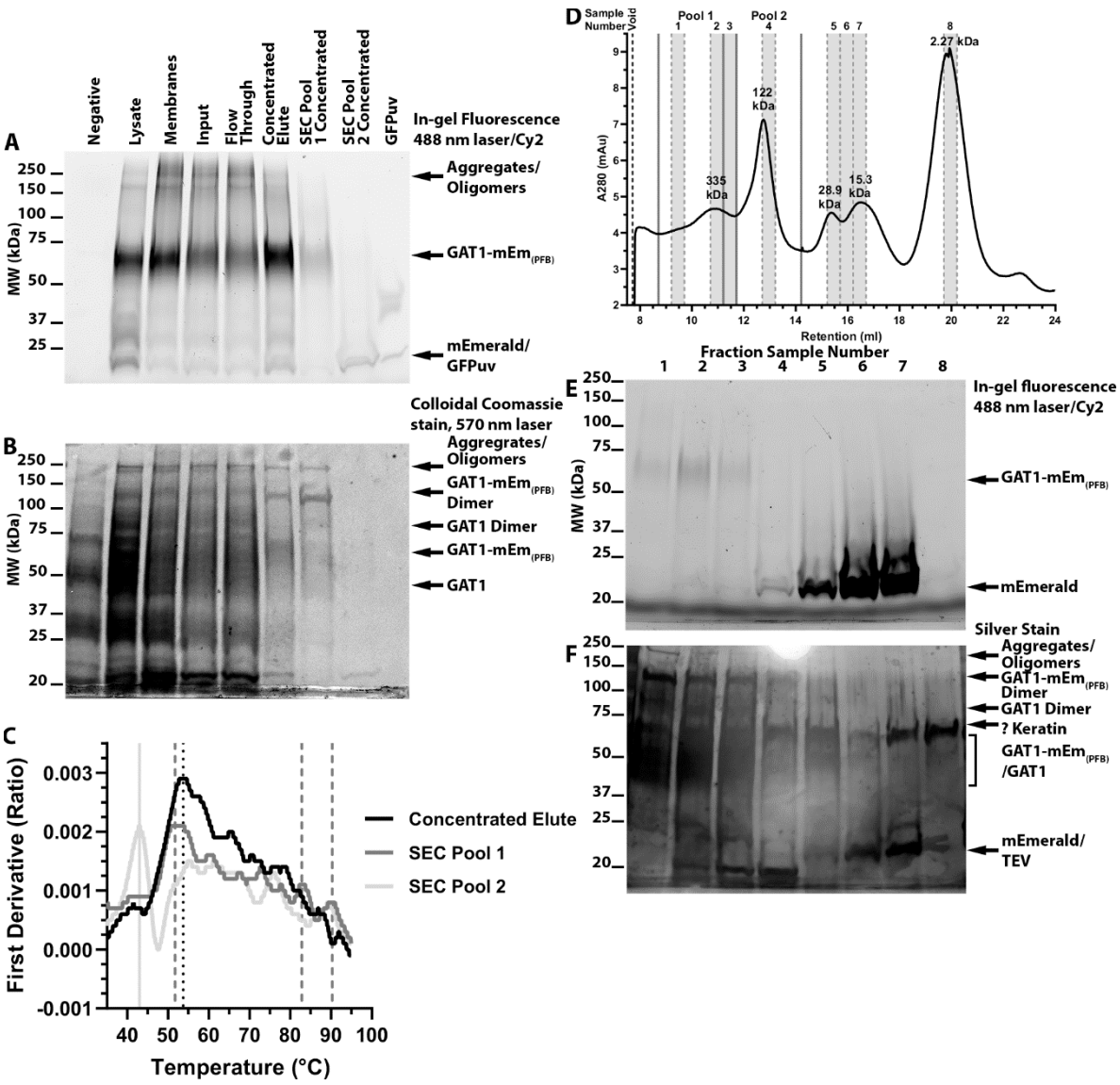


Figure 5.5: Purification of GAT1-mEm(PFB).

Membranes were extracted from *Sf*9 cells infected with the GAT1-mEm(PFB) baculovirus and solubilised in 1.5% C12E9/CHS, then purified with streptactin-XT resin, TEV cleaved and size excluded. **A-B:** SDS-PAGE analysis of samples taken throughout the purification cycle. Details of SEC pool 1 concentrated and SEC pool 2 concentrated are in (**D**). The negative sample consisted of un-infected *Sf*9 cells solubilised in 2% DDM. 20 µg was loaded per each un-purified lane, 3 µg was loaded for concentrated elute and 20 µl of reduced/denatured protein for each concentrated SEC pool sample. **A**: In-gel fluorescence (488 nm excitation, Cy2 filter), **B:** colloidal Coomassie stained gel. **C:** First derivative Nanotemper Tycho intrinsic fluorescence 330/350 nm ratio thermal un-folding profiles of the GAT1-mEm(PFB) concentrated elute sample in **B-C** and SEC pool 1 and pool 2 from (**D**). Fitted inflection temperatures for GAT1-mEm(PFB) concentrated elute are shown as vertical dotted black lines, for SEC pool 1 as vertical dashed dark grey lines and as a solid light grey vertical line for SEC pool 2. **D**: Size exclusion A_{280} trace following TEV cleavage of GAT1-mEm(PFB). The protein was concentrated to ~ 150 µl and loaded on a Superdex 200 10/300GL column at 4 °C and eluted as 0.5 ml fractions. MW values for each peak as calculated from standards are indicated. Fractions shaded in grey and numbered above were analysed by SDS-PAGE in **E-F**. The ranges of fractions pooled and analysed in **A-C** are indicated by flanking solid vertical lines. **E-F:** SDS-PAGE analysis of selected fractions numbered as in (**D**). Following electrophoresis, the gel was analysed by in-gel fluorescence (excitement at 488 nm and imaging with a Cy2 filter) (**F**), followed by silver staining (**G**).

Multiple peaks showed on the SEC A280 trace (Figure 5.5E), but running samples of the fractions in each peak on SDS-PAGE (Figure 5.5F-G) and analysing by both in-gel fluorescence and silver staining (due to the low sample concentration), suggested that cleaved and un-cleaved forms of GAT1-mEm$_{(PFB)}$ were present in the first peak (SEC Pool 1 in Figure 5.5) at ~11 ml, which had eluted at a similar retention to that expected for GAT1-mEm from FSEC. Both monomer and potential dimer forms of GAT1 and GAT1-mEm also appeared to be present. When fractions for this peak were pooled (pool 1) and concentrated, it was clear from in-gel fluorescence that un-cleaved GAT1-mEm$_{(PFB)}$ was in the sample. Cleaved GAT1 was also visible on the stained gel. A T_i of 51.7 °C was recorded (Figure 5.5C), reduced from the 53.7 °C of the un-cleaved protein, and additional T_i values were present around 90 °C which have been previously attributed to mEmerald unfolding. Despite the prominent A280, the second peak at 12.8 ml appeared to contain only a small amount of GAT1 monomers and dimers, plus a small quantity of free mEmerald. Following pooling (pool 2) and concentrating very little was visible on SDS-PAGE. It was un-certain what the fitted T_i of 43 °C corresponded to. The subsequent two peaks appeared to contain free mEmerald and/or TEV, but the identity of the last prominent peak at 19.9 ml was uncertain. From the column calibration this corresponded to a mass of only 2.27 kDa, so it likely ran off the gel. It was noted on the silver stained gel that there was a band at ~55 kDa, but this was also visible in all the other lanes and most likely represents keratin contamination (Ochs, 1983; Burgess, 2009). It was attempted to run the fractions on BN-PAGE as well, but severe smearing (likely because of the salt content being above the recommended maximum of 50 mM) and weak staining even with Zn^{2+}-imidazole counter-staining meant that it provided little usable information (data not shown).

To try to ascertain the oligomeric state(s) of GAT1-mEm$_{(PFB)}$ with greater resolution, samples of the purified protein eluted from the resin and concentrated were analysed by mass photometry (Figure 5.6A). There appeared to be high amounts of noise when analysing samples rapidly diluted 1:400 to beneath the CMC in detergent free buffer to remove free detergent micelles, which complicated analysis. Mass photometry suggested though that gaussians for three forms of the protein could be fit to the data acquired, these being at 98 (σ 30, n=1084, 54%), 136 (σ 40, n=1172, 59%) and 215 (σ 37, n=341, 17%) kDa.

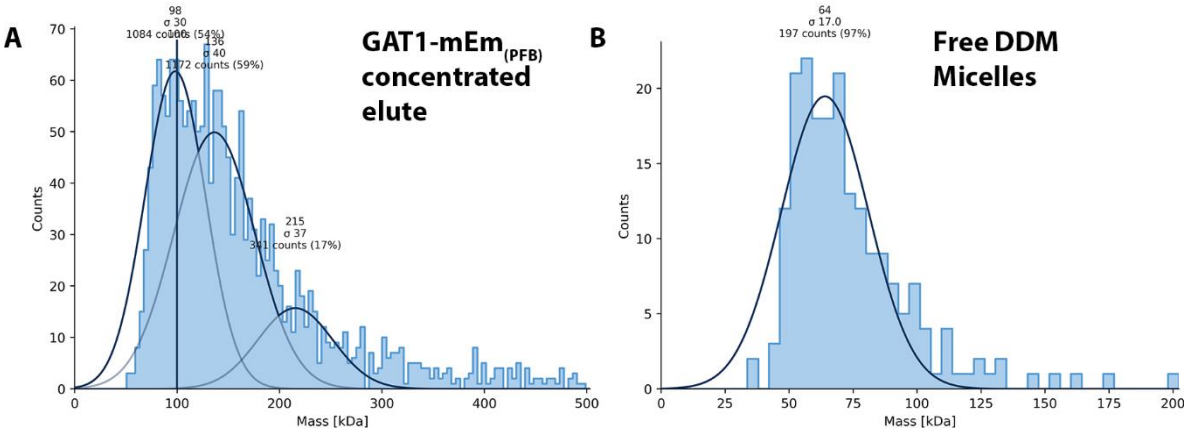


Figure 5.6: Mass photometry analysis of purified GAT1-mEm$_{(PFB)}$.

A-B: Fitted mass photometry of (**A**) purified GAT1-mEm$_{(PFB)}$ and (**B**) free DDM micelles. Rapid 1:400 dilution in detergent free buffer, final concentration ~1.36 nm for (**C**) and 0.0000417% for (**D**). The fitted mass values to each gaussian are indicated.

Based on the mass near exactly fitting that calculated for GAT1-mEm$_{(PFB)}$ (98 kDa), the first was assigned as likely representing a monomer, while the last was assumed to be a dimer. The identity of the 136 kDa species was uncertain. To speculate, it could possibly represent a post-translationally modified form of GAT1-mEm$_{(PFB)}$ or a dimer with the mEmeralds removed as this would be close to the fitted mass. As other faint bands also appeared on the SDS-PAGE, a co-purifying interacting contaminant is a likely possibility When testing a sample of free DDM detergent micelles after an analogous rapid dilution (Figure 5.6B), a peak was fit at 64 kDa (σ 17, n=197, 97%) that was not present with the protein sample, so free DDM micelles did not appear to represent a complicating factor with analysis of the results. This is also close to the literature value of a DDM micelle MW (72 kDa) (Strop and Brunger, 2005). Collectively, this suggests that GAT1-mEm$_{(PFB)}$ purifies as a mainly monomeric species with a small proportion being dimeric.

While this purification was ultimately unsuccessful as large proportions of GAT1 were un-cleaved by TEV, useful information was provided that helped inform subsequent purifications.

Refinement of GAT1-mEm$_{(PFB)}$ purification procedures

It was decided to initially repeat the previous purification of GAT1, but with several modifications to try and improve protein recovery and purity. The culture scale was increased to 2 L, although it was necessary to amplify the GAT1-mEm$_{(PFB)}$ P1 virus stock (to P2, all work to this point had used P1 virus) as it had been

stored at -80 °C for over 12 months due the COVID-19 pandemic disrupting work. A sample of the P0 had unfortunately not been stored frozen after harvesting. Aside from needing more stock due to the increased volumes required, baculovirus titres tend to decline by a small amount after a few months in ultra-cold storage, so it recommended to re-amplify prior to use. 1% P2 virus however appeared to act similarly to P1 virus in terms of cell growth and viability and on in-gel fluorescence gave a similar profile to 1% thawed P1 virus (data not shown), so the infection parameters were not modified.

Extraction proceeded as previously, although when the resin was added to the detergent soluble fraction the newly marketed Streptactin-XT 4Flow resin suspension was used and two hours was allowed for the protein to bind rather than one. The new resin form is advertised as having a tighter affinity for larger proteins and better performance over a wider range of conditions, so it was hoped that this could improve capture of the protein in all oligomeric states present. As with the previous purification, when the resin was collected in a gravity flow column it was intensely green coloured. To try and reduce co-purification of contaminants, the resin was washed with a high salt version of buffer C containing 500 mM NaCl and a raised level of DDM/CHS at 0.1/0.02%. The protein was eluted in buffer C at pH 8 plus 0.05/0.005% DDM/CHS and 50 mM of *d*-biotin (the standard 200 mM NaCl concentration was used). As with the previous purification, the resin remained an intense green, so it was assumed that further protein had not eluted. Elution buffer was therefore added a further four times and allowed to incubate with the resin for 15 minutes on each occasion before collection. After five additions of elution buffer, the reduced resin colour and A280 suggested that the majority of the protein had likely eluted. The elutes were each concentrated down to ~0.5 mg/ml and on SDS-PAGE (Figure 5.7A) and an analytical 15 cm SEC column (representative trace in Figure 5.7C), it was clear that each individual elution contained GAT1-mEm$_{(PFB)}$.

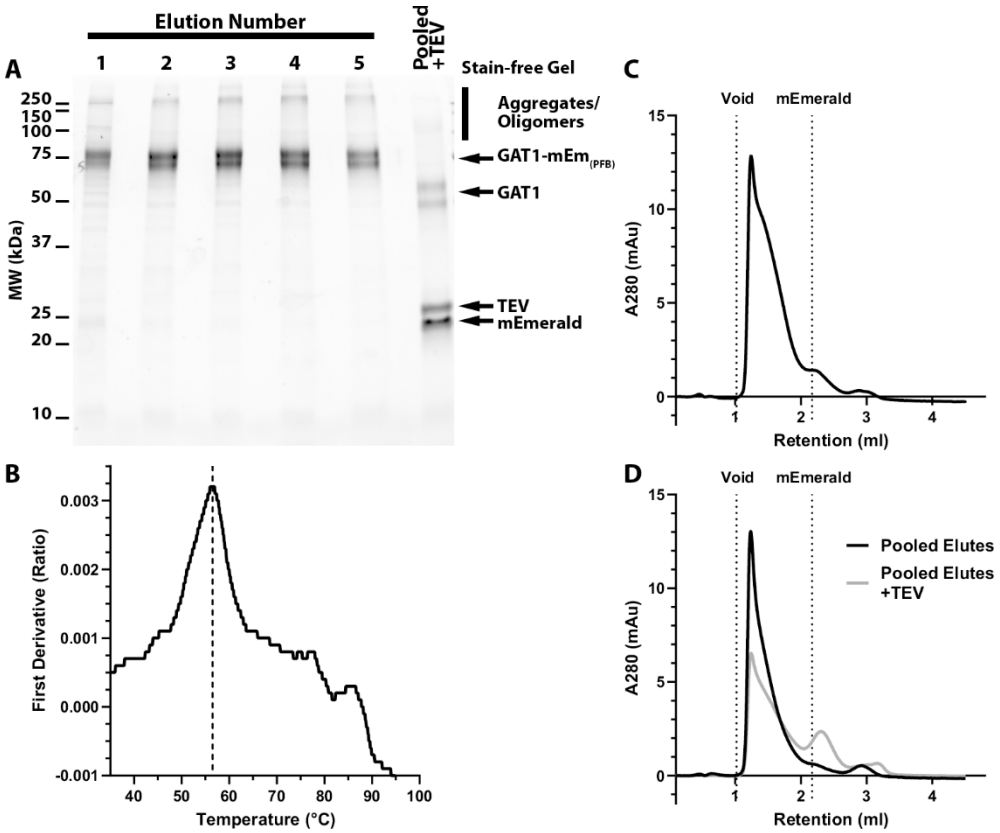


Figure 5.7: Elutions of GAT1-mEm$_{(PFB)}$ from streptactin-XT 4Flow resin and TEV cleavage.

Protein bound to streptactin-XT 4Flow resin was eluted with buffer C plus 0.05/0.005% DDM/CHS and 50 mM *d*-biotin. Elution buffer was added to the gravity flow column five times and the elutes collected, then individually concentrated before addition of TEV protease on a 1:1 ratio at 4 °C. **A:** SDS-PAGE analysis of each elute and the pooled elutes after 18 hours of TEV cleavage. ~3 µg of protein was loaded per lane. **B:** First derivative Nanotemper Tycho intrinsic fluorescence 330/350 nm ratio thermal un-folding profiles of the GAT1-mEm$_{(PFB)}$ elute 1. The fitted inflection temperature is shown as a vertical dotted black line. **C-D:** analytical size exclusion profiles of (**C**) GAT1-mEm$_{(PFB)}$ elute 1 and (**D**) the pooled elutes before and after 36 hours of TEV cleavage. Samples were loaded on a Superdex 200 5/150GL column at 4 °C using a 20 µl loop with 2x overfilling. The column void volume and the calculated retention for mEmerald are indicated.

Analysis on the Nanotemper Tycho fitted a T_i of 56.5 °C for the first elution, raised approximately 3 °C from the elute of the previous purification (Figure 5.7B). This suggests that 200 mM of NaCl has a greater stabilising effect than the 100 mM previously used for elution, as this was the only major difference between the two purifications except for the high salt/detergent wash, additional elutions and the use of P2 virus. After pooling and concentrating, approximately 1.8 mg of GAT1-mEm$_{(PFB)}$ was obtained. TEV protease was applied to the concentrated protein on a 1 mg:1 mg basis, raised from the previous 1:5 ratio used, and left to cleave at 4 °C for 36 hours. Later running of a sample collected at 18 hours on SDS-PAGE suggested though that ~95% of the protein had been cleaved by that point (Figure 5.7A, D).

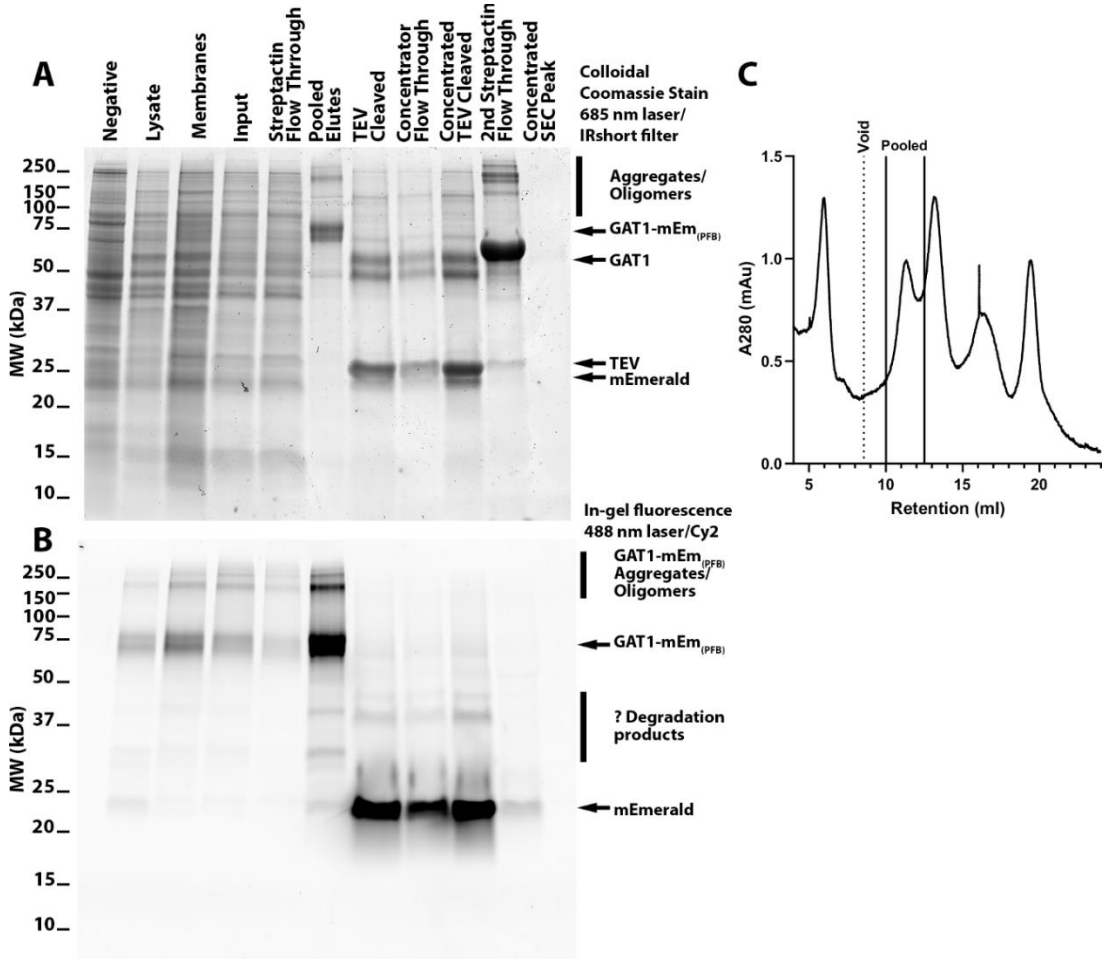


Figure 5.8: Refined purification and TEV cleavage of GAT1-mEm$_{(PFB)}$.

Membranes were extracted from *Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ baculovirus and solubilised in 1.5% C12E9/CHS in buffer C, then purified with streptactin-XT 4Flow resin. The protein was next TEV cleaved, and following removal of the *d*-biotin, passed over Streptactin-XT resin and size excluded. **A-B:** SDS-PAGE analysis of samples taken throughout the purification cycle. Details of concentrated SEC pool are in (**C**). The negative sample consisted of un-infected *Sf*9 cells solubilised in 2% DDM. 20 µg was loaded per each un-purified lane, 3 µg was loaded for the pooled elutes and an equal volume to this was reduced/denatured and loaded for the subsequent lanes up to the concentrated SEC peak lane, where 15 µl of sample was reduced/denatured and loaded. **A**: Colloidal Coomassie stained gel, **B:** In-gel fluorescence (488 nm excitation, Cy2 filter). **C**: Size exclusion A_{280} trace following TEV cleavage of GAT1-mEm$_{(PFB)}$ and a second reverse streptactin purification. The protein was concentrated to ~200 µl and loaded on a Superdex 200 10/300GL column at 4 °C and eluted as 0.5 ml fractions. The range of fractions pooled and analysed in **A-B** is indicated by flanking solid vertical lines.

As it was previously indicated that SEC could not reliably separate cleaved and un-cleaved GAT1-mEm$_{(PFB)}$, the protein was passed over 500 µl of equilibrated 50% streptactin-XT resin to bind away un-cleaved protein and twin-strep tagged free mEmerald. The *d*-biotin present in the buffer was first removed though by concentration followed by dilution in buffer C plus 0.05/0.005% DDM/CHS three times in 100 kDa cut-off centrifugal concentrators. Theoretically, after this reverse purification only GAT1 and any remaining TEV protease should have been present in the protein sample and in-gel fluorescence suggested that mEmerald

tagged proteins had been mostly depleted (Figure 5.8B). Following further concentration with a 100 kDa concentrator, which should have further reduced the amount of TEV present, the protein was size excluded. Unfortunately, the A280 was extremely faint on the chromatogram compared to previously (Figure 5.8C). Pooling the fractions assumed to correspond to the GAT1 from previous purifications and maximally concentrating led to a barely detectable A280 signal and a band on SDS-PAGE that was scarcely visible even under infrared laser excitation of the colloidal Coomassie (Figure 5.8A). It was unclear what happened to the protein, but a significant peak was also present in the column void, indicating aggregation. This was surprising as aggregates had not been seen before with GAT1-mEm$_{(PFB)}$. It was additionally noted on SDS-PAGE that substantial levels of GAT1 appeared in the 100 kDa centrifugal concentrators flow during concentration after TEV cleavage, so the buffer exchange cycles and final concentration may have lost most of the protein.

While an efficient method for obtaining comparatively large amounts of relatively pure GAT1-mEm$_{(PFB)}$ that could be successfully TEV cleaved had been identified, there were substantial losses of protein throughout the purification cycle. With a subsequent purification of another 2 L culture, the methods were therefore altered to try and minimise protein losses. The experiment was repeated identically until the addition of TEV protease, which was only left for 18 hours as cleavage had been found to be maximal by this point previously. Only 1 mg of protein was obtained however after eluting 5 times off the resin compared to 1.8 mg previously. After TEV cleavage, following concentrating down with a 50 kDa cut-off centrifugal concentrator the biotin was removed by SEC and the protein fractionated. Using <100 kDa cut-off filters can be problematic as free detergent micelles are typically too large to pass through, which concentrates the free micelles as well as the protein; the micelle size of DDM is 72 kDa for instance (Strop and Brunger, 2005).

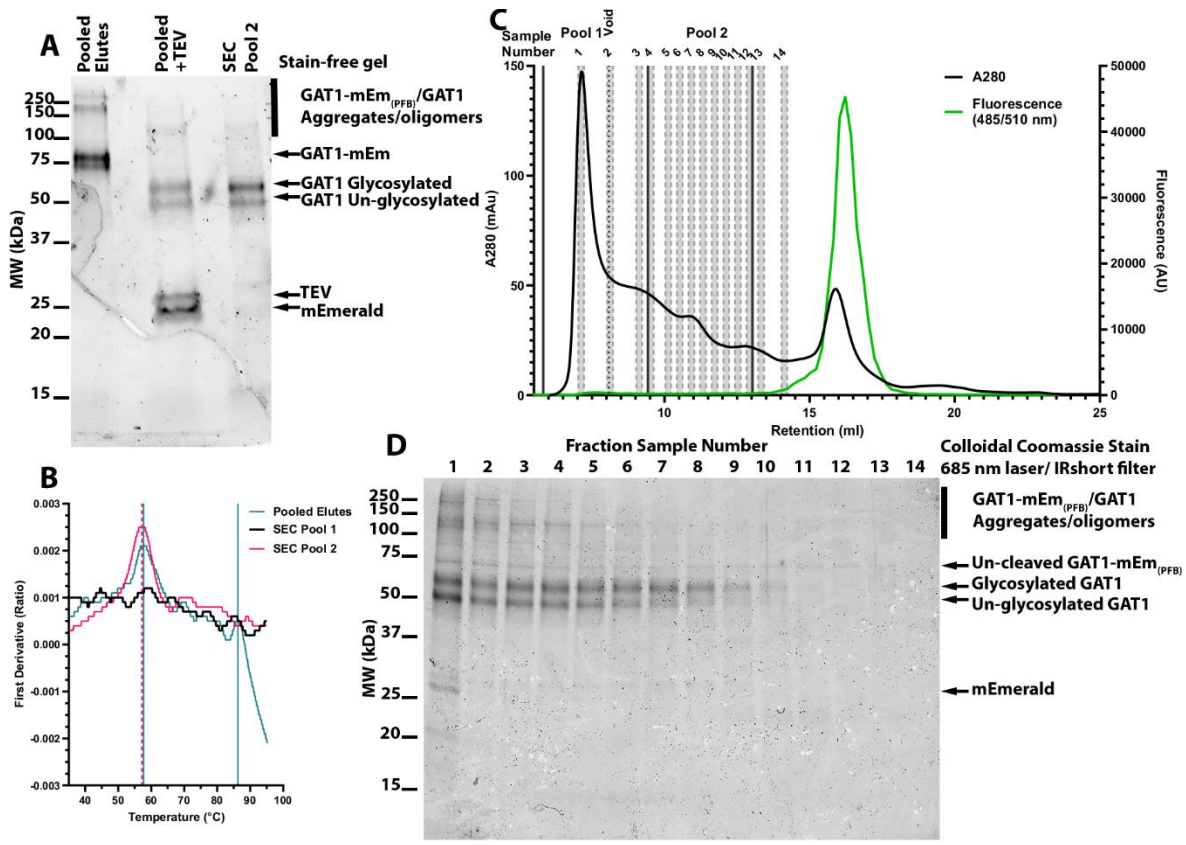


Figure 5.9: Production and analysis of purified GAT1 from GAT1-mEm$_{(PFB)}$.

Membranes were extracted from *Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ baculovirus and solubilised in 1.5% C12E9/CHS, then purified with streptactin-XT 4Flow resin. The protein was then TEV cleaved and size excluded. **A:** Stain-free SDS-PAGE analysis of purified samples. Details of concentrated SEC 2 pool are in (**C**). 3 µg was loaded for the pooled elutes lane and an equal volume was reduced/denatured and loaded for the pooled+TEV lane. 3 µg was loaded for the concentrated SEC pool 2 lane. **B:** First derivative Nanotemper Tycho intrinsic fluorescence 330/350 nm ratio thermal un-folding profiles of samples taken throughout the purification. The fitted inflection temperatures for the pooled elutes are shown as vertical solid teal lines and for SEC pool 1 as a dashed vertical dashed pink line. **C**: Size exclusion A_{280} and fluorescence traces following TEV cleavage of GAT1-mEm$_{(PFB)}$. The protein was concentrated to ~200 µl and loaded on a Superdex 200 10/300GL column at 4 °C and eluted as 0.2 ml fractions in a black walled clear bottomed 96 well microplate. Fluorescence of the fractions was then read at 485/510 nm at a gain sufficient to reach 70% of the plate readers maximum dynamic range. Fractions shaded in grey and numbered above were analysed by SDS-PAGE in (**D)**. The ranges of fractions pooled and analysed in **A-B** are indicated by flanking solid vertical lines. **E:** Colloidal Coomassie stained SDS-PAGE gel of selected fractions numbered as in (**C**).

As 100 kDa concentrators previously caused large protein losses though, it was considered appropriate to use a 50 kDa concentrator, especially as the protein was going to be buffer exchanged regardless during the size exclusion. Unfortunately, as with the previous purification a large proportion of the protein eluted in the void volume during size exclusion (Figure 5.9C), indicating aggregation. During the return to the baseline there was a minor peak around the retention where GAT1 was thought to have eluted during the previous purification. A second major peak followed at ~15-18 ml. FSEC detergent screening had

previously suggested that protein solubilised in C12E9/CHS detergent eluted as a monodisperse peak with no signal in the column void and past analytical SEC of TEV cleaved GAT1 did not show any elution in the void. Therefore, it was strange as to why the protein appeared to have aggregated.

Running samples of the fractions on SDS-PAGE (Figure 5.9D) indicated the presence of GAT1 in all fractions of the major peak, including those in the void volume and in the minor peak during the return to the baseline. The fractions had been collected in a 96 well microplate as used with FSEC and plotting the fluorescence of all the fractions suggested that the second major peak at ~15-18 ml corresponded to free mEmerald (Figure 5.9C). Levels of un-cleaved protein were implied to be low as there was only a barely detectable fluorescence signal in the other fractions and this was also apparent on in-gel fluorescence (not shown). Testing pooled fractions of the major peak (SEC pool 1 in Figure 5.9) in the void volume on the Nanotemper Tycho after concentrating with a 100 kDa cut-off filter did not produce an inflection temperature despite showing a high sample brightness (Figure 5.9B), implying that it represented protein which was already aggregated prior to the temperature ramp. Concentrating the fractions for the minor peak before the mEmerald (SEC pool 2 in Figure 5.9) however gave an analogous un-folding curve to that of GAT1-mEm prior to TEV cleavage and the inflection temperature of 57.2 was only reduced by 0.3 °C. On SDS-PAGE (Figure 5.9A), the protein in this pool was shown as relatively pure and ran as a double band with the glycosylated and un-glycosylated species of approximately equal intensity. The intention had been to pass the pooled fractions over streptactin-XT resin again before concentration to remove any remaining mEmerald or un-cleaved protein, but the low quantity/concentration of sample available (~0.06 µg in 200 µl) precluded this.

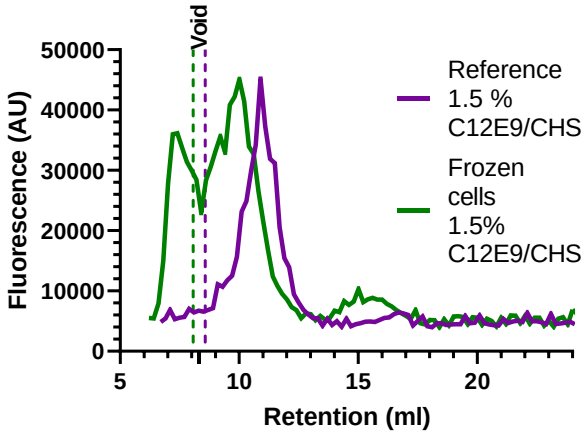


Figure 5.10: Comparison FSEC traces for GAT1-mEm$_{(PFB)}$ infected *Sf9* cells used for purification and for previous detergent screening.

Sf9 cells infected with the GAT1-mEm$_{(PFB)}$ baculovirus were solubilised in 1.5% 12E9/CHS for 2 hours at 4 °C with gentle agitation before insoluble material was removed. The samples were then loaded on a Superdex 200 10/300GL column at 4 °C and eluted into black microplates as 200 µl fractions. The fluorescence was then read at 485/510 nm at a gain sufficient to reach 70% of the plate readers dynamic range. **A:** Previous reference FSEC trace for 1.5% C12E9/CHS solubilised cells during detergent screening. **B:** Frozen sample of cells used for purification solubilised in 1.5% C12E9/CHS.

The void volumes for each trace are shown as vertical dashed lines in the respective colours. Note that the reference sample used a 1 ml loop with partial filling for loading, while the purification cells sample was loaded using a 500 µl loop with overfilling. The FSEC setup had also been optimised to minimise the flow path length since the reference had been recorded. These explain the slightly differing retention volumes.

While it was assumed that the aggregation was a result of the concentration for loading on the column and/or removal of the mEmerald, running a frozen sample of the cells used for this purification on FSEC surprisingly appeared to show an analogous trace with high proportions of protein eluting in the void volume after solubilisation in 1.5% C12E9/CHS (Figure 5.10). This differed from the initial FSEC detergent screening, where none of the detergent soluble samples appeared to show any aggregation. Repeating the FSEC with another culture infected with GAT1-mEm$_{(PFB)}$ appeared to show the aggregation as well, although the aggregates appeared to be abolished when FC14/CHS was used instead (data not shown).

5.3.2.0: Production of the human GABA transporter 1 in *Sf*9 cells with baculovirus using the EarlyBac system

In parallel with attempting to purify GAT1 that had been produced using the conventional baculovirus system with a polyhedrin promoter as used with other proteins in this thesis, another baculovirus approach was also tested later. The EarlyBac (EB) system was recently developed by the Furukawa Lab at Cold Spring Harbour Laboratory and uses a *Drosophila* Hsp70 early promoter in place of the polyhedrin or P10 very late promoters which are most frequently used with baculovirus for protein production (Furukawa *et al.*, 2021). As was previously

discussed in Chapter 4, while these very late promoters strongly drive protein production, they likely contribute to large proportions of produced proteins being frequently un-folded and incompletely processed. This is probably because of the excessive mRNA transcribed. Furthermore, their very late nature means that maximal protein production at overlaps with when baculovirus infected cells are close to death from the viral infection and the secretory path is overwhelmed (Thomas and Tate, 2014). One way around these two problems is to use an earlier, and therefore weaker, promoter as baculovirus infection works as a cascade to transcribe the early followed by the late and very late genes. Examples in order of increasing lateness and strength include the intermediate early 1 (*ie*1), the delayed early *39K* and the late *6.*9 promoters (Lin and Jarvis, 2013). Literature which has used earlier promoters has shown that this can lead to higher proportions of correctly folded and processed proteins being produced, although overall levels may be lower (Higgins *et al.*, 2003; Lin and Jarvis, 2013; Furukawa *et al.*, 2021).

The EarlyBac vectors work by the Bac-to-Bac approach as used with pFastBac 1 and can be substituted in place without any changes in procedures to raise the baculovirus. The Hsp70 promoter acts earlier, with maximal efficiency by ~24-36 hours compared to ~48-72 hours with a polyhedrin promoter (Lee *et al.*, 2000). To maximally boost expression, the EarlyBac vectors also flank the multiple cloning site with a 5' Hr1 enhancer and a P10 3' un-translated region after the stop codon as this combination were found to raise the levels of test proteins by at least six-fold. Furthermore, being based off the multi-bac system, multiple donor and acceptor plasmids can be combined for simultaneous production of different proteins in one virus (Furukawa *et al.*, 2021).

The GAT1-mEm gene was therefore sub-cloned into the EB PFP10 vector (a kind gift from Professor Furukawa, Cold Spring Harbour Laboratory) and baculovirus produced. This construct is hence referred to as GAT1-mEm$_{(EB)}$. Comparison of the fluorescence:[protein] ratio for GAT1-mEm$_{(PFB)}$ and GAT1-mEm$_{(EB)}$ whole cell lysates taken at 24, 48 and 72 hours implied that despite the ratio being raised for EB samples after 24 hours, ultimately the PFB samples were approximately two thirds more fluorescent by 48 hours (Figure 5.11). There appeared to be only a minor variance in the fluorescence:[protein] ratio between EB samples which

had been moved from 27 °C to 18 °C after 24 hours (repeating the ideal conditions identified for PFB produced protein) and those which stayed at 27 °C throughout. This was except for the constant 27 °C samples fluorescence:[protein] ratio saturating earlier at 48 hours and slightly declining by 72 hours (Figure 5.11B). For future cultures using the GAT1-mEm$_{(EB)}$ baculovirus the cells were therefore kept at 27 °C and harvested at 48-60 hours. This also simplified logistics as a separate 18 °C incubator was not required.

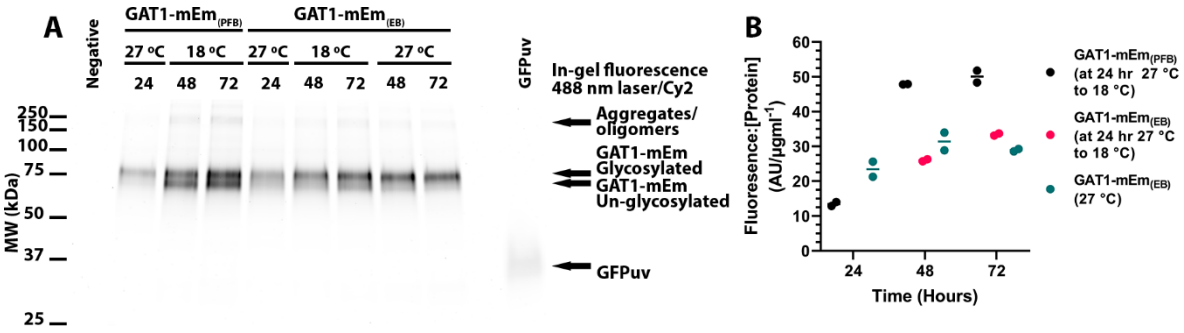


Figure 5.11: Time-course of protein production in GAT1-mEm$_{(PFB)}$ and GAT1-mEm$_{(EB)}$ baculovirus infected *Sf*9 cultures.

Samples of *Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ or GAT1-mEm$_{(EB)}$ baculovirus were collected every 24 hours after infection and snap frozen until analysed. The GAT1-mEm$_{(PFB)}$ baculovirus infected culture was shifted from 27 °C to 18°C after 24 hours, replicating the previously identified optimal procedures. The GAT1-mEm$_{(EB)}$ baculovirus infected culture however was split in half after 24 hours, with one half remaining at 27 °C and the other being changed to 18°C. After thawing, the samples were solubilised for 2 hours at 1x10^7 cells/ml in buffer C plus 2% DDM with gentle mixing at 4 °C. They were then centrifuged at 15,000 *g* for 30 minutes. **A:** SDS-PAGE in-gel fluorescence of the samples. ~15 µg of protein was loaded per lane, except for the purified GFPuv positive control where 50 ng was loaded. The negative control sample consisted of un-infected *Sf*9 cells lysed using 2% DDM in buffer C. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set. **B:** Scatter dot-plot of the samples fluorescence:[protein] measurements, calculated by dividing the fluorescence in AU by the protein concentration in µg/ml as quantified by a BCA assay. Readings were recorded in duplicate at 485/510 nm with the individual reads shown as points and the mean as a horizonal bar. Where no bar is shown then it is within the size of the symbols.

Attractively, GAT1-mEm$_{(EB)}$ differed from GAT1-mEm$_{(PFB)}$ in mainly only showing a single band at the expected point for the glycosylated form on in-gel fluorescence (Figure 5.11A). While a lower band for the non-glycosylated form was also faintly detectable, GAT1-mEm$_{(PFB)}$ contrasted in universally running with the bands for each form being approximately equal intensity. Therefore, it is suggested that a greater proportion of GAT1-mEm produced with the EarlyBac system is correctly glycosylated.

As to whether a greater proportion of GAT1-mEm$_{(EB)}$ was additionally correctly folded compared to GAT1-mEm$_{(PFB)}$ was explored by running folding screen assays. As a positive control of a system known to be able to successfully

produce the protein in a functional form, lipofectamine 2000 transfected HEK293 cells (hence referred to as GAT1-mEm$_{(HEK293)}$) were also compared. In whole cell samples, there appeared to be little difference between the PFB and EB systems, with 1% digitonin, 1% DDM and 1% FC13 respectfully extracting ~30%, ~60% and ~80% of the total sample fluorescence:[protein] ratio (Figure 5.12B), although the EB system may have been slightly more efficient. With GAT1-mEm$_{(HEK293)}$ though, compared to both baculovirus systems extraction was increased ~20% for 1% digitonin and ~10% for 1% DDM respectively. 1% FC13 however showed little difference in extraction. Overall, this suggests that a similar proportion of the protein was likely folded in the PFB and EB systems, but higher levels were in HEK293 cells.

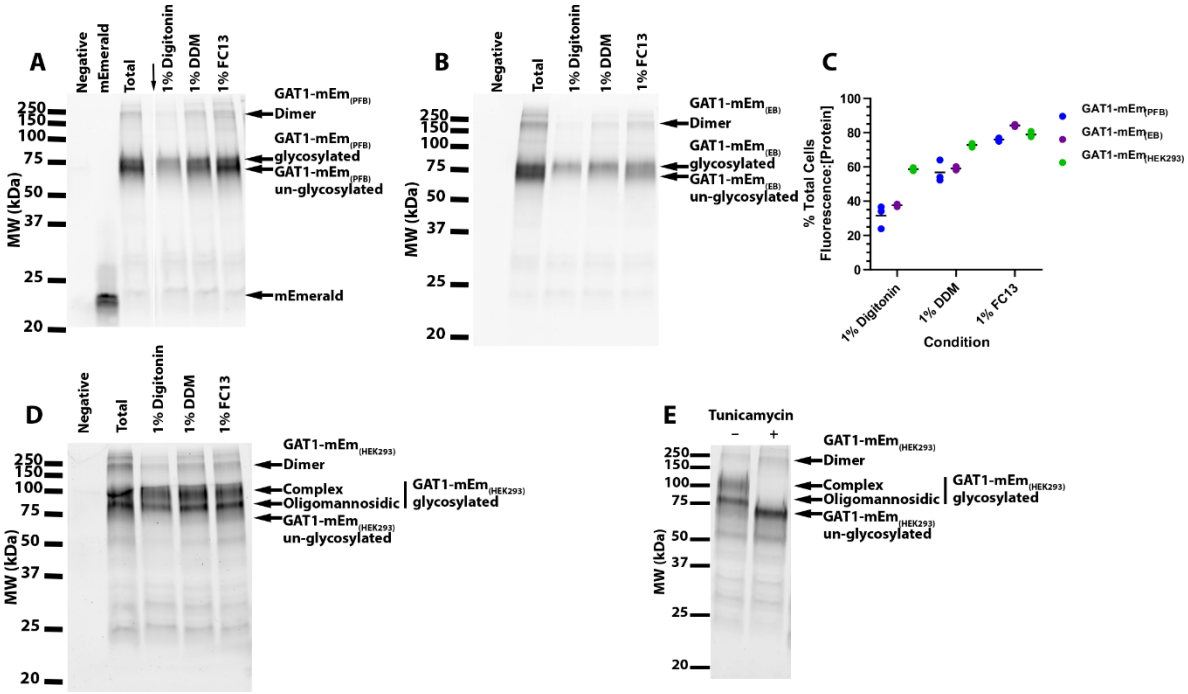


Figure 5.12: Folding screen assays with whole cell samples of GAT1-mEm$_{(PFB)}$, GAT1-mEm$_{(EB)}$ and GAT1-mEm$_{(HEK293)}$ and tunicamycin treatment of GAT1-mEm$_{(HEK293)}$ transfected cells.

A-D: *Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ (**A**) or GAT1-mEm$_{(EB)}$ (**B**) baculovirus and HEK293 cells transfected with GAT1-mEm$_{(HEK293)}$ (**D**) were solubilised at 1x10^7 cells/ml in buffer C plus 1% digitonin, DDM or FC13 for 2 hours with gentle mixing at 4 °C. They were then ultra-centrifuged to remove insoluble material. The total sample consisted of cells without any detergent added which were lysed by needle sheering and not ultra-centrifuged. **A-B, D:** SDS-PAGE in-gel fluorescence of the samples. ~20 µg of protein was loaded per lane. The negative control samples consisted of un-infected *Sf*9 or mock transfected HEK293 cells lysed using 2% DDM in buffer C. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set. Please note that in (**A)** a gel lane has been removed, the position of which is identified with an arrow. **C:** Scatter dot-plot of the folding screen assay samples fluorescence:[protein] measurements, calculated by dividing the fluorescence in AU by the protein concentration in µg/ml as quantified by a BCA assay. Readings were recorded in duplicate at 485/510 nm with the individual reads shown as dots and the mean as a horizonal bar. Where no bar is shown then it is within the size of the symbols. **E:** SDS-PAGE in-gel fluorescence of the tunicamycin treated and un-treated samples. For the

treated sample 10 µg/ml tunicamycin was added at the time of transfection. After 24 hours the cells were harvested and lysed by needle shearing, ~20 µg was loaded per lane. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set.

Although less apparent, these trends in the levels of extraction were also visible on in-gel fluorescence (Figure 5.12A, B, D), but intriguingly GAT1-mEm$_{(PFB)}$ and GAT1-mEm$_{(EB)}$ differed in the levels of glycosylated protein extracted. While the total samples for both showed glycosylated and un-glycosylated forms, with GAT1-mEm$_{(PFB)}$ 1% digitonin and DDM appeared to extract a roughly equal proportion of each form. FC13 extracted more of the un-glycosylated protein. With GAT1-mEm$_{(EB)}$ however, while the overall intensity was similar, 1% digitonin and DDM near exclusively extracted glycosylated protein and FC13 appeared to extract a roughly equal proportion of both forms. GAT1-mEm$_{(HEK293)}$ was similar in this regard, with all samples showing a similar pattern of two main bands at ~80 kDa and 100-110 kDa, as well as a barely visible lower band at ~70 kDa. Treatment with tunicamycin abolished the upper two bands (Figure 5.12E), confirming that they represented glycosylated forms of the protein. These been previously identified as forms containing complex and oligomannosidic *N*-glycans respectively (Cai *et al.*, 2005).

However, in a folding screen assay performed on isolated membranes (Figure 5.13), there appeared to be a dramatic difference between the PFB and EB vectors (HEK293 cell membranes were not tested due to the excessive number of cells which would have been required). 1% digitonin extracted 35.3% of the total membranes fluorescence:[protein] ratio for GAT1-mEm$_{(PFB)}$, but 72.4% for GAT1-mEm$_{(EB)}$. A similar difference was apparent with 1% DDM (62.9% compared to 95.5%), although for 1% FC13 practically 100% was extracted with both systems. 1.5% C12E9/CHS was similar at ~75% for both.

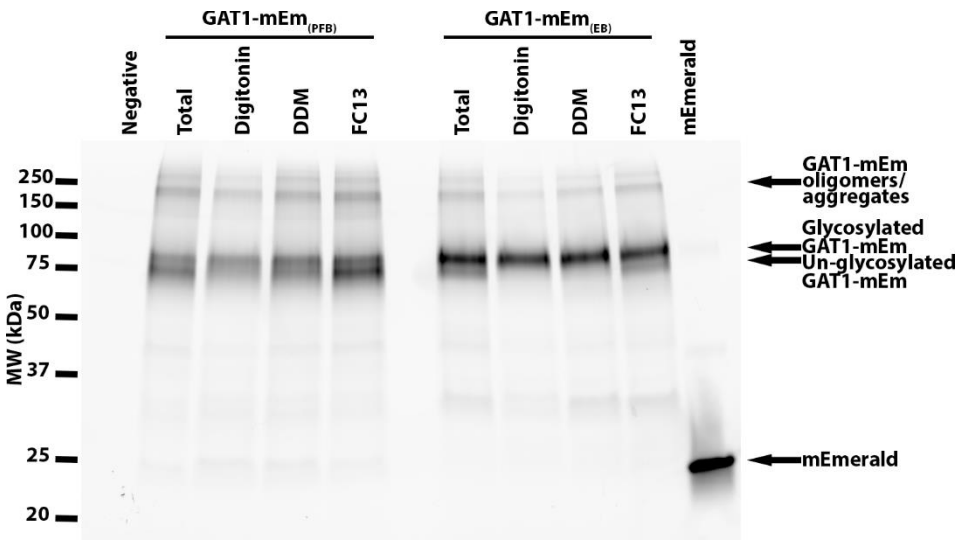


Figure 5.13: Folding screen assays with membranes of GAT1-mEm$_{(PFB)}$ and GAT1-mEm$_{(EB)}$ infected *Sf*9 cells.

Membranes were extracted from *Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ or GAT1-mEm$_{(EB)}$ baculovirus and solubilised at 1.5 mg/ml in buffer C plus 1% digitonin, DDM or FC13 for 2 hours with gentle mixing at 4 °C. The samples were then ultra-centrifuged to remove insoluble material. The total sample consisted of membranes without any detergent added and not ultra-centrifuged. SDS-PAGE in-gel fluorescence of the samples. The negative control sample consisted of un-infected *Sf*9 cells lysed using 2% DDM in buffer C. ~10 µg of protein was loaded per lane, except for the purified mEmerald positive control where 40 ng was loaded. The gel was excited with a 488 nm laser and imaged with a Cy2 filter set.

It was interesting that GAT1-mEm$_{(EB)}$ appeared to preferentially be extracted nearly exclusively as a single glycosylated form by non-ionic detergents in both whole cell and membrane samples. There were similarities in this regard to whole cell samples of GAT1-mEm$_{(HEK293)}$, although the average fluorescence:[protein] ratios for the total samples suggested that approximately 40% greater levels of GAT1-mEm$_{(EB)}$ were present compared to GAT1-mEm$_{(HEK293)}$. Comparing FSEC profiles for protein produced using the EB system to that with HEK293 cells showed that they behaved surprisingly similar (Figure 5.14). Each showed a single peak with 1% DDM and FC13. 1.5% C12E9/CHS also appeared to give a sharp single peak with both detergents, like what had been identified with GAT1-mEm$_{(PFB)}$ during the initial detergent screening. It was therefore decided to continue using this detergent combination for extraction of GAT1-mEm$_{(EB)}$ from membranes for purification.

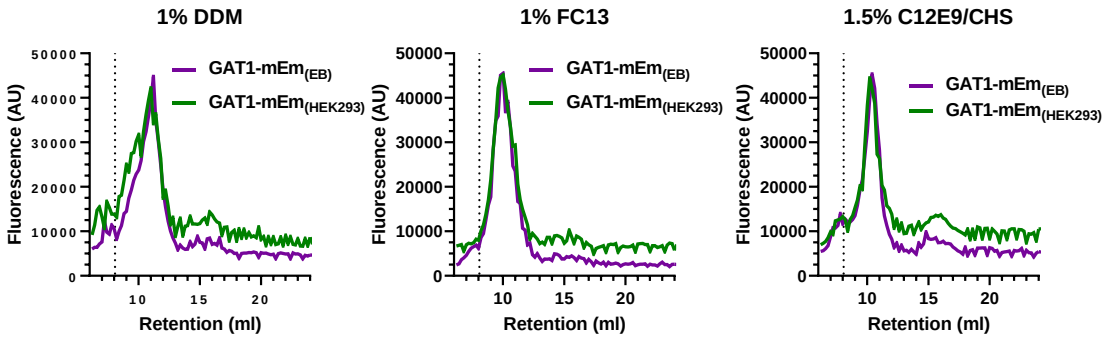


Figure 5.14: FSEC analysis of GAT1-mEm$_{(EB)}$ and GAT1-mEm$_{(HEK293)}$.

*Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ baculovirus were solubilised in 1% DDM, 1% FC13 or 1.5% 12E9/CHS for 2 hours at 4 °C with gentle agitation before insoluble material was removed. The samples were then loaded on a Superdex 200 10/300GL column at 4 °C and eluted into black microplates as 200 µl fractions. The fluorescence was then read at 485/510 nm at a gain sufficient to reach 70% of the plate readers dynamic range - note that the results are therefore not comparable quantitatively. The void volume is shown in each trace as a vertical dotted line.

5.2.2.1: Purification of GAT1 from GAT1-mEm$_{(EB)}$ infected *Sf*9 cells

Following on, a 2 L culture of *Sf*9 cells was infected with the GAT1-mEm$_{(EB)}$ baculovirus and harvested after 60 hours at 27 °C. The membranes were extracted and solubilised in 1.5% C12E9/CHS and then purified as with GAT1-mEm$_{(PFB)}$, the last 2 L culture of which (shown previously in Figure 5.9) was processed and purified in parallel to allow direct comparison. After five elutions off the resin and pooling and concentrating, ~1.7 mg of GAT1-mEm$_{(EB)}$ was obtained as assessed by A280 with the protein extinction co-efficient and MW. Analysis on the Nanotemper Tycho fitted an average inflection temperature of 58.9 °C compared to 57.5 °C for a sample of PFB GAT1-mEm$_{(PFB)}$, therefore suggesting that the EB produced protein was more stable (Figure 5.15B).

TEV was added for ~18 hours on a 1:1 basis, before concentrating with a 50 kDa cut-off centrifugal concentrator to ~200 µl and loading onto a Superdex 200 10/300GL column. GAT1 eluted as a single majority peak on A280 at ~11 ml, with a small amount of protein in the void volume and a secondary peak assumed to represent the cleaved mEmerald and TEV (Figure 5.15C). The SEC A280 elution profile was similar to that of GAT1 from GAT1-mEm$_{(PFB)}$, except for only a comparatively small proportion eluting in the void volume. Fluorescence of the fractions likewise confirmed that the secondary peak was cleaved mEmerald and there was only a barely detectable fluorescence signal (less than approximately 0.5% the height of the cleaved mEmerald) in the A280 peak at ~11 ml. As with

GAT1-mEm$_{(PFB)}$, this inferred that only a small amount of un-cleaved GAT1-mEm$_{(EB)}$ was present.

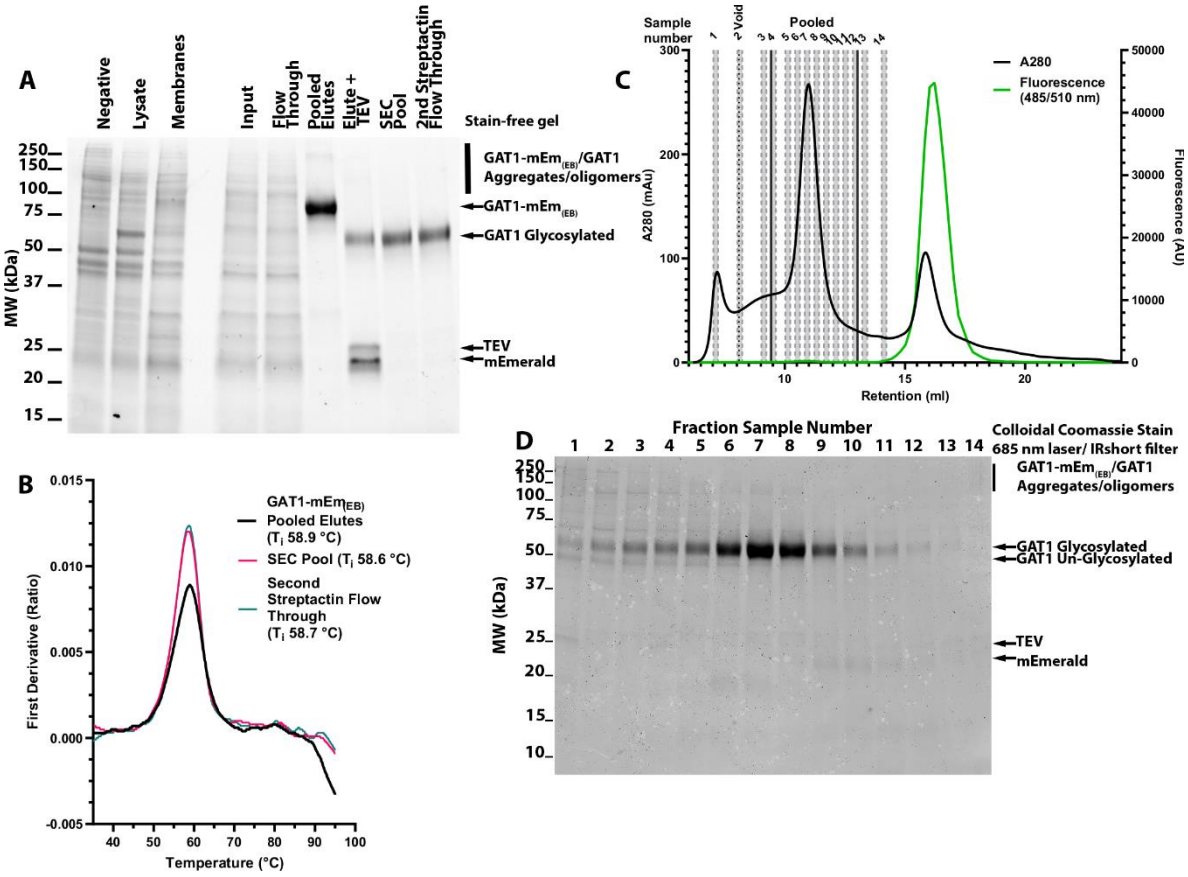


Figure 5.15: Production and analysis of purified GAT1 from TEV cleaved GAT1-mEm$_{(EB)}$.

Membranes were extracted from *Sf*9 cells infected with the GAT1-mEm$_{(EB)}$ baculovirus and solubilised in 1.5% C12E9/CHS in buffer C, then purified with streptactin-XT 4Flow resin. The protein was then TEV cleaved and size excluded. **A:** Stain-free SDS-PAGE analysis of purified samples. Details of the SEC pool sample are in (**C**). 15 µg was loaded per each un-purified sample lane. 3 µg was loaded for the pooled elutes lane and an equal volume was reduced/denatured and loaded for the pooled+TEV lane. 3 µg was loaded for the SEC pool and 2nd streptactin flow through lanes. **B:** First derivative Nanotemper Tycho intrinsic fluorescence 330/350 nm ratio thermal un-folding profiles of samples taken throughout the purification. **C**: Size exclusion A280 and fluorescence traces following TEV cleavage of GAT1-mEm$_{(PFB)}$. The protein was concentrated to ~200 µl and loaded on a Superdex 200 10/300GL column at 4 °C and eluted as 0.2 ml fractions in a black walled clear bottomed 96 well microplate. Fluorescence of the fractions was then read at 485/510 nm at a gain sufficient to reach 70% of the plate readers maximum dynamic range. Fractions shaded in grey and numbered above were analysed by SDS-PAGE in (**D).** The range of fractions pooled and analysed in A-B are indicated with flanking vertical solid grey lines. The void volume is indicated with a vertical black dotted line. **E:** Colloidal Coomassie stained SDS-PAGE gel of selected fractions numbered as in (**C**).

Upon separating selected fractions by SDS-PAGE (Figure 5.15D), in gel fluorescence (not shown) only showed a faint signal at the expected MW for GAT1-mEm$_{(EB)}$ in the fractions. In the final fraction analysed just before the cleaved mEmerald peak, where a strong signal was also present for mEmerald.

After colloidal Coomassie staining however, it was apparent that GAT1 was present in all fractions analysed. The glycosylated form of GAT1 appeared to be the most predominant, although a much smaller proportion of un-glycosylated protein and a form possibly corresponding to un-cleaved GAT1-mEm$_{(EB)}$ could also be seen. Fractions (Figure 5.15C) were pooled and then passed twice across 500 µl of 50% streptactin-XT resin in a gravity flow column to capture any un-cleaved protein or free mEmerald. It was then attempted to concentrate the protein using a 100 kDa cut-off centrifugal concentrator, but unfortunately the protein appeared to pass into the flow-through. A 50 kDa cut-off concentrator was therefore used, although as this likely concentrated free DDM/CHS micelles as well, the concentration of detergent present was unknown. By A280, it was estimated that ~0.6 mg of GAT1 was obtained at 1.5 mg/ml. SDS-PAGE (Figure 5.15A) suggested that the protein was pure, although on in-gel fluorescence (not shown) there was still a faint band present at the position expected for GAT1-mEm$_{(EB)}$. Total protein loading quantification however suggested that the level of un-cleaved protein had been depleted by ~80% compared to before passing over the resin and any mEmerald present had been removed. Analysis on the Nanotemper Tycho (Figure 5.15B) fitted an average inflection temperature of 58.7 °C to the final protein, which was similar to that measured previously (58.5 °C) when the fractions were pooled.

After treatment of purified GAT1-mEm$_{(EB)}$ and the resulting purified GAT1 with PNGase F, it was clear that practically 100% of the protein appeared to be glycosylated (Figure 5.16). The bands for the treated samples showed substantial movement compared to un-treated samples. On a 3-8% tris-acetate gel to separate high MW proteins, it was also apparent that possible oligomeric species of GAT1 up to perhaps pentameric forms were visible and the position of these likewise shifted with PNGase F treatment. Treatment of purified GAT1-mEm$_{(PFB)}$ also showed similar results (data not shown). As to whether these are actual functional oligomeric forms or LDS resistant aggregates is unclear however. After re-running on another gel, the PNGase F treated and un-treated bands for the GAT1 monomer and for the treated dimer were excised and sent for mass spec protein ID to confirm the protein identity. All three bands submitted were confirmed to contain GAT1, with 28-32% sequence identity and good coverage of the N- and C-termini. It was also attempted to confirm the glycosylation of the

three known N-linked sites, both by searching the peptides for the mass shift associated with linked glycans and also from the deamidation of the Asn residues to Asp in the PNGase F treated samples. Unfortunately, sequence coverage within the core of the protein was poorer and none of the peptides identified contained the relevant section with the assumed sites.

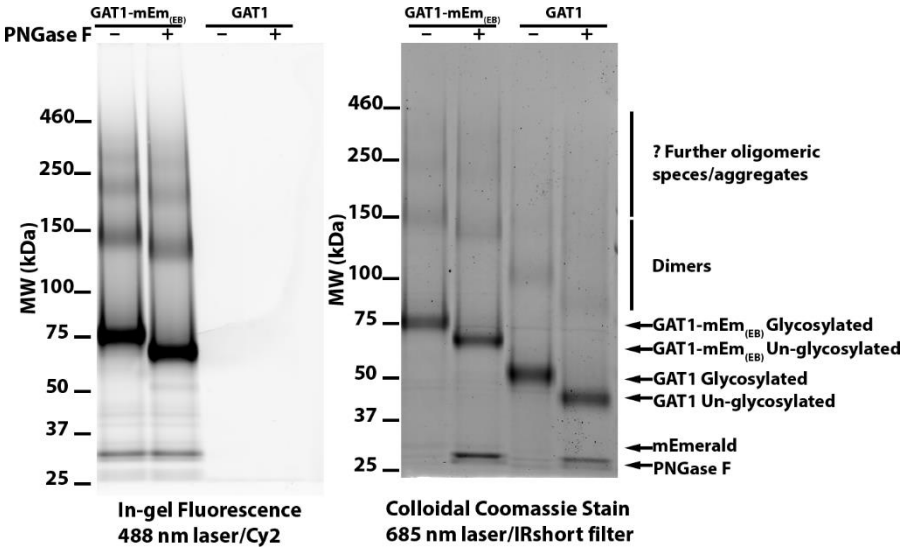


Figure 5.16: PNGase F treatment of purified GAT1-mEm$_{(EB)}$ and GAT1.

Samples of purified GAT1-mEm$_{(EB)}$ and the subsequent purified GAT1 were denatured and treated with 5x10^{8} units of PNGase F enzyme (+) or water (-). The reactions were stopped with SDS-PAGE loading buffer and ~1.5 µg of protein per lane loaded on SDS-PAGE. The gels were first imaged on in-gel fluorescence (488 nm laser, Cy2 filter) and subsequently colloidal Coomassie stained.

Initial negative stain EM grids (Figure 5.17) of GAT1 produced from GAT1-mEm$_{(EB)}$ suggested that the protein was readily visible and the particles appeared well distributed and fairly heterogenous. It was possible to see particles in multiple orientations. While some aggregation was present, this was probably from the low pH of the uranyl acetate staining. Overall, the grids looked promising and give optimism that the protein may be suitable for future cryo-EM studies.

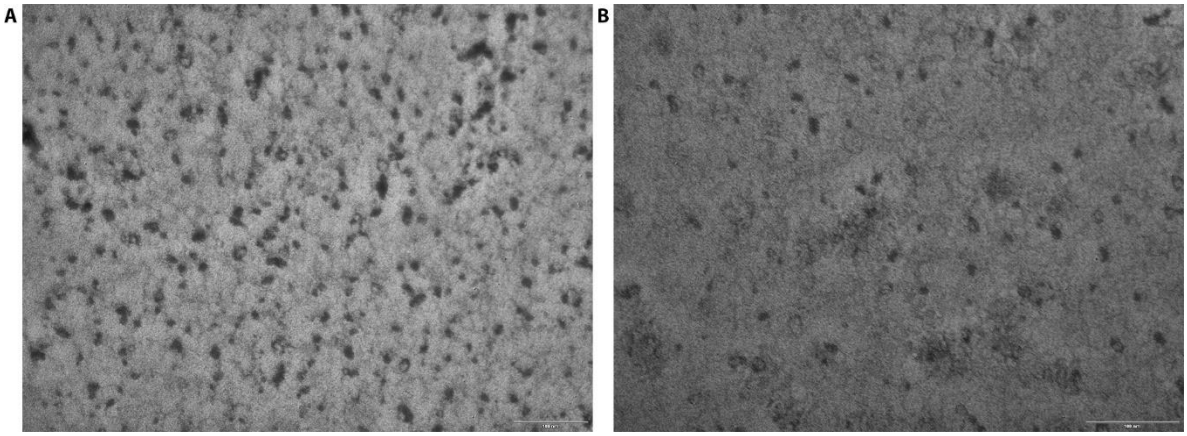


Figure 5.17: Representative negative stain electron microscopy of purified GAT1.

After exchanging into fresh buffer with a de-salting column to remove excess detergent micelles, GAT1 produced from GAT1-mEm$_{(EB)}$ was diluted to 0.5 mg/ml, applied to glow discharged carbon grids and negative stained with uranyl acetate. The grids were imaged on a Jeol JEM 1200EX at 80 kV and **A:** 120,000x, **B:** 150,000x magnification. Scale bar indicates 100 nm. The grids were prepared and imaged by Dr. Alan Prescott.

5.3.3.0: Raising a recombinant nanobody against the human GABA transporter 1

Structural studies of GAT1 were one of the primary objectives. However, it was seen as unlikely that GAT1 could be crystallised for X-ray diffraction without extensive sequence modification. Crystallising membrane proteins is exceptionally challenging and has always represented a bottleneck for structural studies (Kermani, 2021). Single particle cryo-EM however has matured into a complementary technology which has proved especially useful for study of membrane protein structure (Thonghin *et al.*, 2018). While single particle cryo-EM is a powerful technique where the minimum size limit has been steadily decreasing, successful reconstructions of proteins such as GAT1 which are under 100 kDa are still challenging. As more fully described previously in section 4.1.3.2, one common method to reduce this problem is to use a binding partner protein such as a nanobody to increase the molecular size. In this section, the raising of recombinant nanobodies against GAT1 using a yeast nanobody library is described.

GAT1 produced from GAT1-mEm$_{(EB)}$ was labelled with Sulfo-NHS-LC-LC-Biotin and labelling was confirmed from the electrophoretic shift on non-reducing SDS-PAGE after addition of streptavidin. After the first two tetrameric antigen selection rounds using 400 nM of biotinylated GAT1 in the positive selection steps to enrich

for avidity, flow cytometry recorded 43.89 % of the post LS column sort yeast being positive for AF647 fluorescence compared to 5.08% in the stained pre-sort and 4.3% in the LS column flow through.

Round three was the initial monomeric antigen selection round to enrich for affinity rather than avidity. This round used 2 µM of biotinylated GAT1 during the positive selection and for the subsequent round four 500 nM was used. Compared to 10.30% in the stained pre-sort and 1.17% in the LS column flow through, 46% of the post LS column sort yeast were positive for AF647 fluorescence (Figure 5.18) after round four. Following round five using 100 nM of biotinylated GAT1, 9.39% of the stained pre-sort, 1.24% of the LS column flow through and 38.95% of the post sort yeast were AF647 fluorescence positive.

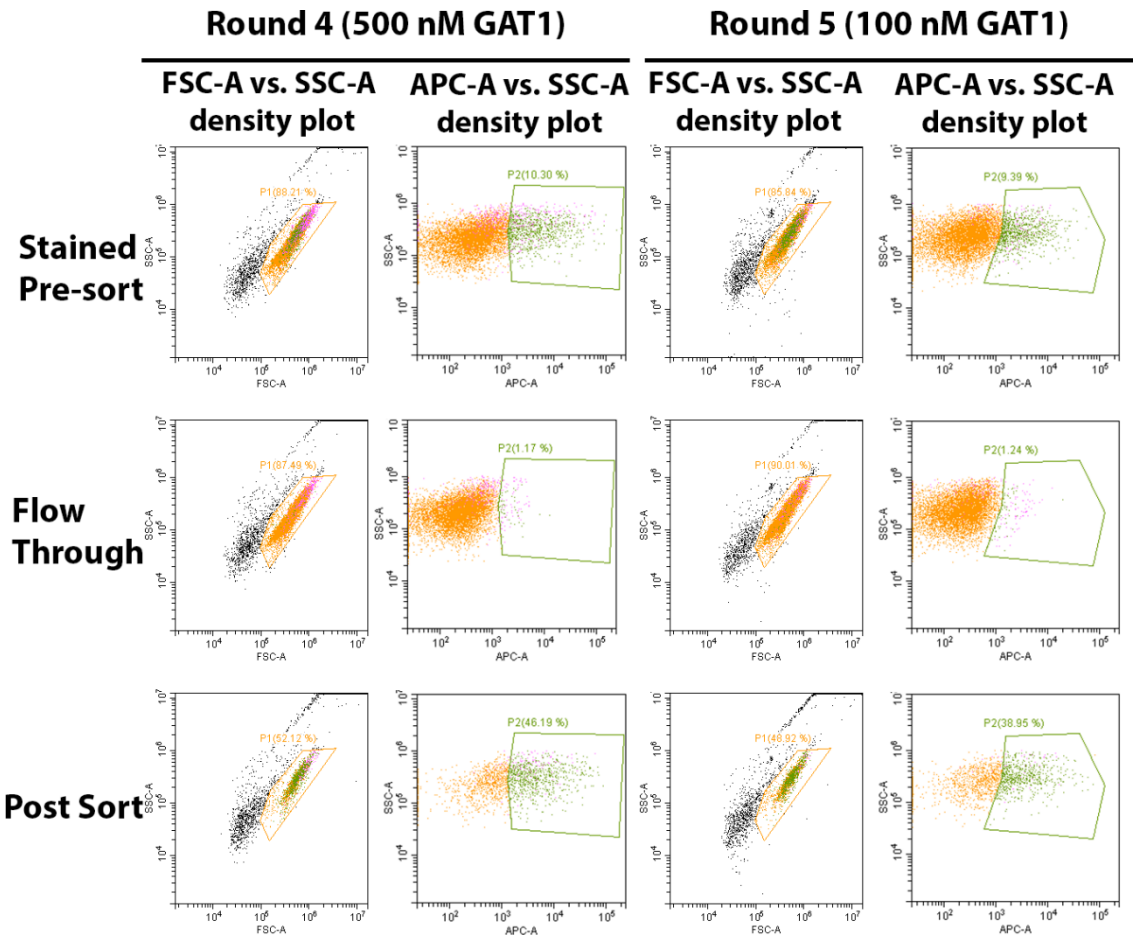


Figure 5.18: Flow cytometry density plots for yeast taken throughout selection rounds 4 and 5.

During each selection round, samples of yeast were taken before sorting on the LS column, from the column flow through and from the post-sort. The yeast were then analysed for AF647 fluorescence in the APC channel on the Cytoflex 5. The gates were set automatically from the SSC-A vs. FSC-A plot.

Following the last selection round, subsequent work was carried out by Dr. Tom Eadsforth and Mr Greg Stewart. The yeast were plated on YPD media and 24 individual colonies were subsequently picked, grown up in liquid culture and plasmid DNA extracted. After transforming into *E.*coli, plating and growing up in liquid cultures, plasmid DNA was extracted and sequenced. Full length sequences were obtained for 18 nanobodies (Figure 5.19), all of which were unique. Most variation occurred in the three complementarity-determining regions, although some nanobodies also showed substitutions or deletions in other parts of the sequence.

Unfortunately, there was insufficient time to be able to sub-clone the nanobody genes for growth in *E.coli* and purification. The specificity and strength of any interaction for these candidate nanobodies against GAT1 is therefore unknown.

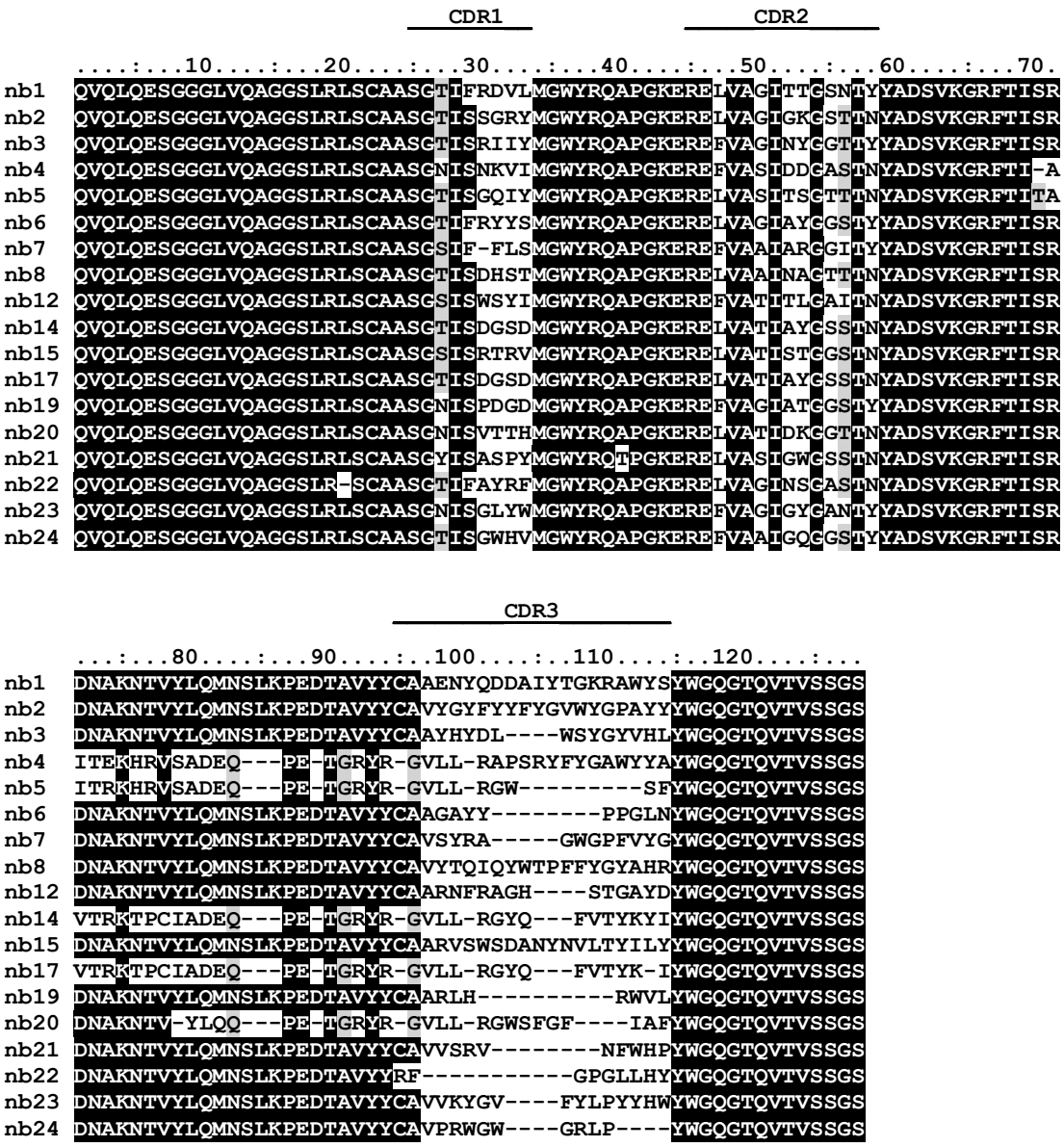


Figure 5.19: Amino acid sequences for the nanobodies raised against GAT1.

The candidate nanobody plasmid DNA sequences were obtained using flanking forward and reverse primers with Sanger sequencing. The corresponding amino acid sequences are shown aligned with Clustal Omega and shaded with pyBoxshade to indicate conservation. Directly conserved residues are shaded black and conservative substitutions in grey. Residues comprising the three complementarity-determining regions (CDR) in the nanobody are indicated above.

5.3.4.0: Assessing the function of the human GABA transporter 1 with a fluorescent membrane potential assay

While research described so far indicated that comparatively high yields of GAT1 could be produced at high purity using the EarlyBac system, it was unclear if this represented functional protein which was correctly folded and able to transport GABA. Measuring the uptake of a radiolabelled substrate in cells, isolated membranes or in a purified transporter after reconstitution in proteoliposomes is the traditional technique used to study transporter function (Weinglass *et al.*, 2008). However, as using radioligands was not possible due to the prohibitive expense, training requirements and safety factors, alternative approaches were examined. Preferentially, this was going to be an assay which could work with cells or isolated membranes as it would avoid needing a steady supply of purified GAT1 for analysis. Eventually it was decided to use a fluorescent membrane potential assay from Molecular Devices. The FLIPR membrane potential blue assay labels live cells with a fluorescent dye so that during membrane depolarisation the fluorescent signal intensity increases as dye trails the positively charged ions into the cell. The fluorescent signal correspondingly decreases in intensity during hyperpolarisation as positively charged ions leave the cell, with the dye following.

5.3.4.1: Assessment of GAT1-mEm$_{(HEK293)}$ with the FLIPR membrane potential blue assay

Initially, assays were performed with GAT1-mEm$_{(HEK293)}$ as HEK293 cells have been used extensively for functional assays with GAT1 in previous literature, including with attached fluorophores. This was intended to act as a positive control both for the protein construct and for the assay in general, as it was unclear if the dye would work with *Sf*9 cells.

As a control to check the function of the assay, mock transfected and GAT1-mEm$_{(HEK293)}$ transfected cells were depolarised by addition of KCl (Figure 5.20A). Previous tests (data not shown) with the dye in un-transfected HEK293 cells had shown that the fluorescence signal increased along a concentration-response relationship with KCl, presumably in response to depolarising and the opening of endogenous ion channels. In mock transfected HEK293 cells, KCl gave a

concentration response curve with a pEC50 of 2.02 (±0.08 SEM) and a Hill slope of 2.20 (±0.75 SEM). In a single curve of GAT1-mEm$_{(HEK293)}$ transfected cells, a similar pEC50 of 2.37 (±0.08 SEM) was fit. The fluorescence signal also appeared to be supressed compared to mock transfected cells. The reason for this was uncertain, but as this assay was not repeated it was not possible to determine if this suppressive effect resulted from transfection of the transporter.

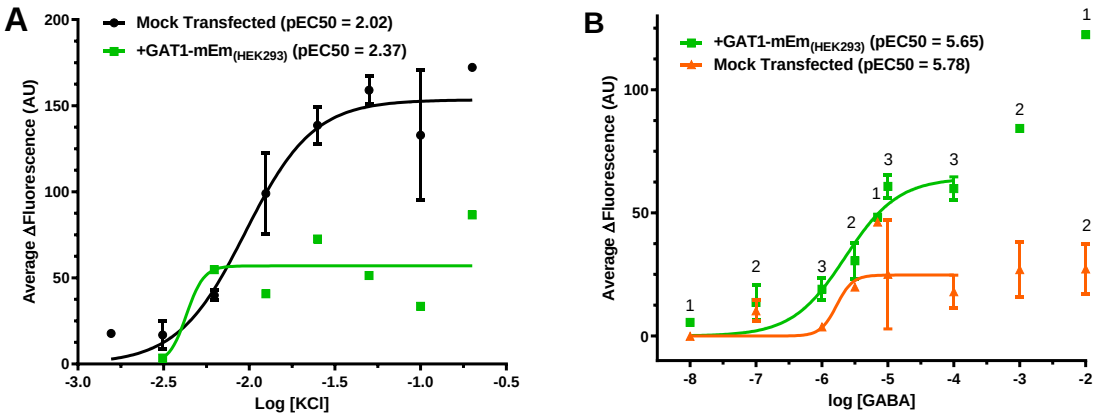


Figure 5.20: FMP blue assay with mock transfected and GAT1-mEm$_{(HEK293)}$ transfected HEK293 cells.

HEK293 cells either mock transfected or transfected with GAT1-mEm$_{(HEK293)}$ were incubated with FMP blue dye and subsequently stimulated with **A:** KCl and **B**: GABA. Each point represents an average and (**A**) n=1 for +GAT1-mEm$_{(HEK293)}$ and n=2 for mock transfected, except at -0.7 and -2.8 log[KCl] where n=1. In (**B**) the n is indicated above a datapoint and is the same for both datasets unless indicated. Error bars represent standard error, where no error bar is visible then it is within the size of the point or n=1. The Hill equation is fitted with the Hill slope left un-constrained. Response to buffer only has been subtracted off each point and the bottom constrained to zero. The fitted pEC50 values are indicated within each figure. For (**B**) the curve is only fit up to -4 log[GABA] and the subsequent points are given for reference only.

Addition of GABA to GAT1-mEm$_{(HEK293)}$ transfected cells caused a concentration dependant increase in fluorescence (Figure 5.20B). From three averaged curves, a pEC50 of 5.65 (±0.16 SEM) and a Hill slope of 1.02 (±0.37 SEM) were fitted. The signals were variable between datasets, but individual analysis fitted similar pEC50 values of around 5.5. There was also surprisingly a concentration-response like relationship with GABA in mock transfected cells. While the maximum signal strength was approximately a third of that for transfected cells and varied between datasets, this background response possibly contributed to

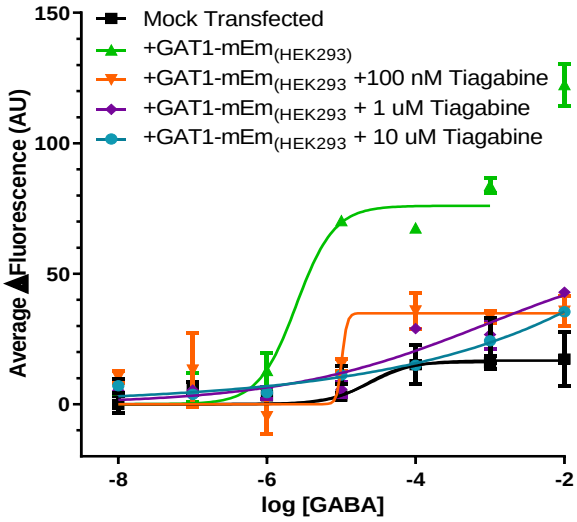


Figure 5.21: The effect of tiagabine on the response of GAT1-mEm$_{(HEK293)}$ transfected HEK293 cells to GABA.

HEK293 cells either mock transfected or transfected with GAT1-mEm$_{(HEK293)}$ were incubated with FMP blue dye and tiagabine at the indicated concentration. **A:** KCl and **B**: GABA. Each point represents the average of two reads. Error bars represent standard error, where no error bar is visible then it is within the size of the point. The Hill equation is fitted with the Hill slope left un-constrained. Response to buffer only has been subtracted off each point and the bottom constrained to zero. The fitted pEC50 values are indicated within each figure. For GAT1-mEm$_{(HEK293)}$ transfected cells not in the presence of tiagabine the curve is only fit up to -5 log[GABA] and the subsequent points are given for reference only.

the unpredictability between datasets with the transfected cells. It was also noted that after saturating at ~10 μM there appeared to be a second phase to the transfected cells response to GABA past 100 μM. The significance of this is unknown. Curves were therefore only fit up to 100 μM and while it was attempted to fit a bi-phasic curve, this was unsuccessful.

As confirmation that the fluorescence signal from the FMP blue dye was from transport of GABA and exchange for ions, tests were also made with GABA in the presence of the GAT1 specific inhibitor tiagabine (Figure 5.21). Cells were incubated with tiagabine at 100 nM, 1 μM and 10 μM, with the intention of using Schild analysis to determine the mode of antagonism and calculate the pA_2 if it was competitive. However, in addition to a rightward shift in the GABA concentration-response curve, tiagabine at all concentrations appeared to supress the E_{max} for GABA. Although this meant that Schild analysis could not be carried out, the suppression suggested a non-competitive mode of action.

It was attempted to repeat this assay with other concentrations of tiagabine which did not suppress the signal as greatly, but it proved difficult to capture reliable data which could be fit with concentration-response curves as there was a high level of variability between datasets. The E_{max} suppression and the rightward shift also necessitated the use of GABA concentrations past 100 μM, at which point both the intrinsic response to GABA from the cells and potentially the putative

second phase from the transfected cells response to GABA likely contributed to the variability. To focus time and resources, it was eventually elected to suspend investigating this potential non-competitive antagonist activity of tiagabine, although it represents an interesting point to look at in future.

5.3.4.2: Assessment of GAT1-mEm$_{(EB)}$ with the FLIPR membrane potential blue assay

As it had been determined that the FMP blue assay worked with GAT1-mEm in HEK293 cells, it was next attempted to repeat using GAT1-mEm$_{(EB)}$ in *Sf*9 cells. As a suitable channel or transporter protein which was known to be active in *Sf*9 cells was unfortunately not available to hand, application of KCl was tested. While it was uncertain if this would change the membrane potential as the endogenous channels and transporters in *Sf*9 cells are poorly defined, excess extracellular KCl will frequently cause depolarisation. As a blank baculovirus was unavailable, a baculovirus for eGFP was used as a negative control.

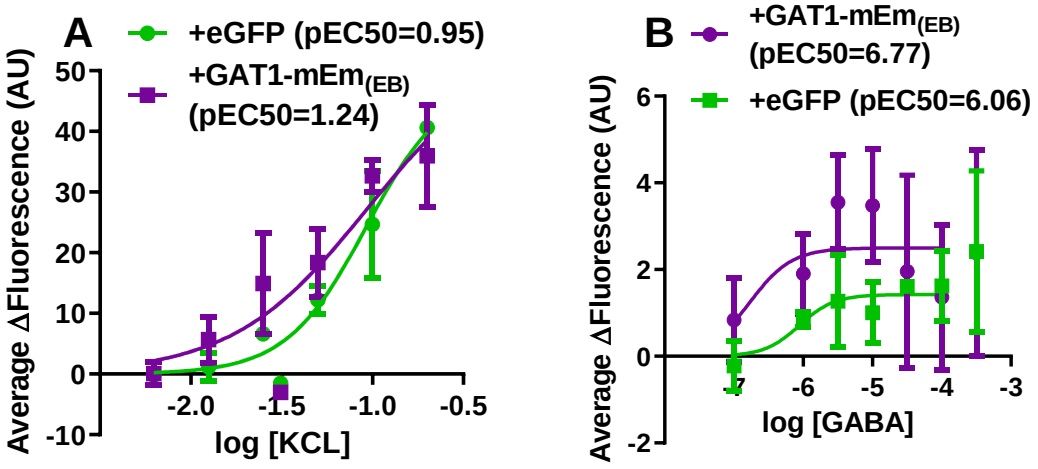


Figure 5.22: FMP blue assay with eGFP and GAT1-mEm$_{(EB)}$ infected *Sf*9 cells.

*Sf*9 cells infected with eGFP or GAT1-mEm$_{(EB)}$ baculovirus for 36 hours at 27 °C were incubated with FMP blue dye and subsequently stimulated with **A:** KCl and **B**: GABA. Each point represents an average and n=2 for each point. Error bars represent standard error, where no error bar is visible then it is within the size of the point. The Hill equation is fitted with the Hill slope left un-constrained. Response to buffer only has been subtracted off each point and the bottom constrained to zero. The fitted pEC50 values are indicated within each figure. For (**B**) the curve is only fit up to -4 log[GABA] and the subsequent points are given for reference only.

KCl appeared to cause a concentration dependent increase in fluorescence with either eGFP or GAT1-mEm$_{(EB)}$, although several fold smaller in intensity than with

HEK293 cells and with pEC50 values at least one order of magnitude smaller (Figure 5.22A). Unfortunately, application of GABA to cells infected with GAT1-mEm$_{(EB)}$ only induced a minor increase in fluorescence and a similar response was given by eGFP infected cells (Figure 5.22B). There was also excessive variability in the data points. While it was possible to fit concentration response curves to these, the extra sum of squares F-test in GraphPad Prism suggested that statistically there were no difference between any of the fitted parameters for either eGFP or GAT1-mEm$_{(EB)}$.

As it appeared to be unfeasible to use the FMP blue assay with *Sf*9 cells due to the low fluorescence response, future attempts at using it to establish the function of GAT1-mEm$_{(EB)}$ were abandoned.

5.4.0.0: Discussion

It was attempted to establish a method for recombinant production and purification of the human GABA transporter 1, a currently poorly exploited drug target for epilepsy and other related disorders. Although there were some challenges after a promising start with baculovirus infection of *Sf*9 cells, a successful method was eventually developed to extract and purify the receptor at a yield and purity potentially suitable for downstream structural studies and other biophysical techniques. Further work involving raising a recombinant nanobody against the purified protein also showed progress, although efforts to assess the transporter function in *Sf*9 cells with a membrane potential assay were unsuccessful.

As discussed in the previous chapter, trials and tribulations are inherent to working with membrane proteins. Using a conventional polyhedrin based baculovirus system however had initially appeared to show a promising start, with GAT1-mEm$_{(PFB)}$ being expressed at a high level in and successfully purified on the first attempt. GAT1-mEm$_{(PFB)}$ could also be efficiently extracted from membranes at a high level in mild non-ionic detergents and showed a monodisperse appearance on FSEC with the majority of the detergents tested. Overall, these were indicative of a well-behaved protein.

When moving on to larger scale purification however, the protein was more problematic. A systematic series of improvements to the purification procedures

eventually led to isolation of purified GAT1 from TEV cleaved GAT1-mEm$_{(PFB)}$, although there were ultimately problems with large proportions being aggregated. There is extensive literature on the function of GAT1 and its interaction with other proteins, trafficking *etc,* but there is very little describing its purification and how it behaves as a purified protein. This also extends to other members of the GABA sub-family of the SLC6 transporters.

At the time of submitting this thesis, the only published study (Hu *et al.*, 2017) which could be found describing the purification of GAT1 used *Sf*9 cells with a conventional polyhedrin based baculovirus. The strategy of a C-terminal GFP fused to GAT1 was remarkably like that used with GAT1-mEm$_{(PFB)}$. This paper shows multiple similarities to results obtained with GAT1-mEm$_{(PFB)}$, including the formation of possible dimers and oligomers on SDS-PAGE and similar SEC elution traces. This paper was an offshoot of a PhD thesis which describes further purification experiments with GAT1-GFP, including from mammalian cells (Hu, 2010). However, an unusual purification strategy of large scale immunoaffinity purification against the GFP followed by elution off the resin at high pH was used with these studies, compared to maintaining closer to physiological buffering conditions throughout for purification of GAT1-mEm in this chapter.

Although it was ultimately possible to purify a small amount of GAT1 from extracted GAT1-mEm$_{(PFB)}$, it is puzzling as to why it started to show large proportions of protein as aggregated in later purifications. The initial FSEC and purification studies did not show any appreciable aggregation at all. P2 baculovirus stock is the most likely candidate for this, as the aggregation only started to appear after this began to be used instead of working from P1 infected cell pellets stored at -80 °C. This was despite the P2 appearing to act similarly to the thawed P1 stock in terms of cell growth, viability at harvest etc and appearing to visually produce a similar level of GAT1-mEm on SDS-PAGE in-gel fluorescence. As the stocks were not titred though, it is uncertain as to what the ultimate MOI may have been and the aggregation could be related to having excessive virus present compared to previous cultures. The point at which the aggregation occurred though is unclear, as analytical SEC analysis of the streptactin resin elute for the initial use of P2 virus for large scale culture did not suggest aggregation either before or after TEV cleavage, but showed a sizable

proportion on the final SEC. There is course the possibility that the protein aggregated under the high concentration needed for SEC. Unfortunately, a sample of the harvested cells was not frozen for this culture, so unlike the subsequent purification it was not possible to use FSEC to see if aggregation was present after whole cell detergent solubilisation. The T_i stayed fairly consistent throughout the purification cycles for GAT1-mEm$_{(PFB)}$, so the protein was not suggested to have lost stability at any point.

A related possibility is that the removal of the mEmerald or the TEV cleavage overnight destabilised the protein. Production as a GFP fusion followed by purification and cleavage of the GFP has been frequently used though for other SLC6 transporters (Penmatsa *et al.*, 2013; Shahsavar *et al.*, 2021). Typically, FSEC studies with un-purified proteins also show similar results to those of a purified un-tagged protein, aside from the increase in mass (Kawate and Gouaux, 2006).

A structure of GAT1 was published during the course of making corrections to this thesis (Motiwala *et al.*, 2022). Unexpectedly, to produce the protein this study was able to successfully use a similar GFP tagged construct of GAT1 in pFastBac 1 with *Sf*9 cells. There were however differences in the extraction and purification strategy. After DDM/CHS was determined to be unsuitable, Motiwala *et al.* used LMNG/CHS was used for extraction. The protein was then purified into a low concentration of LMNG/CHS after binding to GFP affinity resin and performing an on-column proteolytic cleavage off the resin. On SEC, no signals appeared in the void volume, suggesting that aggregates were not present like with GAT1-mEm$_{(PFB)}$. Thermofluor assays also suggested that GAT1 purified into DDM/CHS had a melting temperature of ~48 ºC, compared to ~59 ºC for protein purified into LMNG/CHS. Surprisingly, the reported melting temperatures for LMNG/CHS purified GAT are very similar to the ~56-59 ºC T_i values obtained in this thesis for GAT1 purified into DDM/CHS produced from GAT1-mEm$_{(PFB)}$ and also from GAT1-mEm$_{(EB)}$. Although not directly comparable as differing methods were used, this suggests that the samples were of similar stability.

It is uncertain why the study of Motiwala *et al.* and the research presented in this thesis differ drastically in some fashions, despite using similar techniques. A main difference of course was the use of LMNG/CHS for both extraction and

purification and it is possible that this prevented GAT1 from aggregating during purification. However, Motiwala *et al.* only obtained ~0.1 mg of GAT1 per litre of initial culture (Gati, 2022). When compared to the ~0.3 mg per L for GAT1 ultimately produced from GAT1-mEm$_{(EB)}$ in this thesis, their protein production method appears to have been several fold less efficient.

Using the EarlyBac system appeared to minimise many of the problems associated with purifying GAT1-mEm$_{(PFB)}$ and it was possible to purify GAT1 at a high yield as a single majority peak on SEC, with only a small proportion eluting in the column void. The final protein appeared homogenous and well distributed on negative stain EM, with only a small amount of aggregation which may have been induced by the negative staining. In addition, the results from the EarlyBac system are of interest as practically 100% of the protein extracted with non-ionic detergents was glycosylated, while with pFastBac 1 there was significant heterogeneity in the level of glycosylated and un-glycosylated protein in all extracted samples. This homogeneity with the EB system is consistent with the assumption that higher levels of correctly processed proteins can be obtained by using an earlier, and therefore weaker, promoter that is maximally active before the cellular machinery and ER quality control processes become overwhelmed by virus production (Higgins *et al.*, 2003; Lin and Jarvis, 2013; Furukawa *et al.*, 2021). As glycosylation has significant impact on the function of GAT1 (Bennett and Kanner, 1997; Cai *et al.*, 2005), this could suggest that the protein produced from the EB system will be of higher quality, although the type of glycosylation will still differ in being of a high mannose type rather than the complex glycans of mammalian cells.

In whole cell samples, the level of potentially folded protein appeared to be roughly equal between the two systems in whole cell samples based on the folding screen assay. However, in isolated membranes there was a dramatic difference, with 1% digitonin being able to extract almost twice as much GAT1-mEm$_{(EB)}$ as GAT1-mEm$_{(PFB)}$. The reason for the difference between whole cell and membrane samples is uncertain, but it could infer that a greater amount of correctly folded protein is present on cell membranes with the EB system compared to PFB. It is important to note though that the 200,000 *g* step during the differential centrifugation as used to isolate the membranes also extracts

multiple other cellular organelles, including parts of the rough and smooth ER. (Graham, 2002). This could therefore also isolate ER retained mis-folded protein, which would still be able to fluoresce and influence the assays. To repeat a previous criticism of the folding screen assay, it is based solely on the assumption of a detergents ability to solubilise un-folded protein and is only an indirect semi-quantitative measure at best. Using a functional assay to compare the level of transport or a saturation binding assay to quantify the number of binding sites present on the surface would illuminate more information on this.

When compared to GAT1-mEm$_{(HEK293)}$, production of GAT1-mEm$_{(EB)}$ was implied to less efficient in terms of the level of folding, but this was outweighed by the greater levels of protein produced based on fluorescence:[protein] ratio measurements and in-gel fluorescence. On FSEC, protein produced in the two systems appeared remarkably similar in terms of dispersity. Being able to use *Sf*9 cells instead of mammalian cells yet gaining similarly behaving protein with enhanced yields is advantageous from a cost and scale up perspective. The fact that GAT1-mEm$_{(EB)}$ ran as a single glycosylated band is also a potential advantage over mammalian production, where there were two forms of glycan present. The reduced heterogeneity of GAT1-mEm$_{(EB)}$ could be an advantage during cryo-EM data processing or crystallisation, although it is currently unknown how the high mannose glycans may affect the proteins function. Of course, the glycans could be enzymatically removed with PNGase F or similar before any structural studies.

Efforts to raise a nanobody against GAT1 unfortunately did not reach the stage of being able to define the selectivity and affinity of the nanobody candidates to GAT1. However, the data obtained so far from the flow cytometry is indicative of an increasing enrichment of higher affinity binders as a similar proportion of AF647 labelled yeast were isolated by the LS column, even when dropping the biotinylated protein concentration five-fold between each positive selection round.

As expected, the most variation in the nanobody sequences occurred in the three CDRs and especially in CDR3, which is a critical region for antigen recognition (Mcmahon *et al.*, 2018). Large sequence truncations appeared to be the most common feature of the nanobody candidates in this region, with there being no clear pattern to the differences between each nanobody. It is interesting however

that five nanobodies also showed similar deletions and analogous non-conservative substitutions throughout the framework region three separating CDR2 and CDR3. Typically this region is less sequence variable (Mitchell and Colwell, 2018), so it is intriguing that such a large proportion of the nanobodies showed differences here. It is conjectured if this may represent a defining feature for a class of nanobodies potentially binding at a specific site on GAT1.

Functionally, it was suggested from the FMP blue assay that GAT1-mEm$_{(HEK293)}$ was active in transporting GABA. The calculated pEC50 of 5.62 for GABA from the FMP blue assay is similar to values reported in literature using radioligand uptake studies with GAT1 in mammalian cell lines (Dhar *et al.*, 1994; Giraudo *et al.*, 2019; Motiwala *et al.*, 2022).

The cause of the background fluorescent response to application of GABA in mock transfected cells, especially above 1 μM is unclear. While HEK293 cells do possess mRNA for the $GABA_A$ β_3 subunit, which can form homomers, electrophysiology studies have suggested that the cells have no functional endogenous $GABA_A$ receptors (Davies *et al.*, 2000) . No data could be found suggesting that HEK293 cells endogenously possess GABA transporter proteins. It is therefore suggested that this background is a non-specific or off-target response to the application of GABA. This would be a complicating factor if using the FMP blue assay for ligand screening purposes against GAT1, although it is possible that other mammalian cells (e.g CHO cells) may be suitable.

In the presence of the GAT1 specific inhibitor tiagabine, with GAT1-mEm$_{(HEK293)}$ transfected cells there appeared to be a decrease in the pEC50 and a suppression of the E_{max}. This is consistent with a possible non-competitive mode of action (Kenakin, 2014), although further investigation of this was complicated by the high variability of the FMP blue assay and the suppression of the fluorescence signal.

Despite being a clinically used drug, the mechanism of action for tiagabine is poorly defined. Kragholm *et al.* (2013) found a mixed competitive/non-competitive action of tiagabine against human GAT1 in a [^{3}H]-GABA uptake assay and suggest that the drug could function by targeting an allosteric site on GAT1 in addition to the orthosteric site, although a more complex mechanism is another

possibility. Combining their cryo-EM structure of GAT1 in the inward-open conformation with an AlphaFold model of the transporter in the outward-open conformation and [^{3}H]-GABA saturation uptake assays, Motiwala *et al.*, (2022) have recently proposed that tiagabine works by a two stage mechanism. First, tiagabine, Na^+ and Cl^- ions are bound in the substrate-binding site, with the GAT1 in the outward-open conformation. At this stage, tiagabine acts as a competitive inhibitor. This is then followed by a conformational change to the inward-open conformation, where the ions are released and tiagabine acts as a non-competitive inhibitor by eventually arresting GAT1 in the inward-open conformation. This hence prevents GAT1 from cycling through the subsequent stages of the alternating access model and prevents transport of GABA.

It was unfortunate that the FMP blue assay appeared to be unsuitable for studying the function of GAT1-mEm$_{(EB)}$ in *Sf*9 cells. Although no studies using the assay with *Sf*9 cells could be found, studies were located using it with *Drosophila* S2 cells to study ion channel function (Nakao *et al.*, 2012; Nakao and Banba, 2016). This had given hope that it could also work with other insect cell lines, but ultimately this proved to not be the case. It was unclear though if GAT1-mEm$_{(EB)}$ was non-functional or if the signal strength from the *Sf*9 cells membrane potential change was simply too weak to be a feasible readout. Concerningly, Hu *et al.* (2017) found that for a baculovirus mediated GAT1-GFP fusion in *Sf*9 cells, uptake of [^{3}H]-GABA was barely raised from that of control *Sf*9 cells. Motiwala *et al.*, (2022) used *Sf*9 cells for producing the GAT1 which ultimately resulted in a cryo-EM structure. Unfortunately, they did not report any functional assays with the insect cell produced protein and only tested the construct in mammalian cells. An important point though is that it was not possible to test the FMP blue assay with a positive control protein known to be functional with *Sf*9 cells by electrophysiology *etc*. While KCl appeared to depolarise the cells, this was not really an appropriate control as the endogenous channels in *Sf*9 cells are poorly defined and the mechanism of the depolarisation is therefore uncertain.

Chapter 6: Conclusions and potential future work

Channels and transporters are essential proteins for the control, mediation and termination of neurotransmission. These are implicated in numerous pathological conditions and represent the targets of a multitude of drugs, both clinically useful and otherwise. However, detailed structural information is only available for a small proportion of the channels and transporters and many structure-activity relationships are unclear. This has impeded the development of improved and novel ligands which act at these proteins. This thesis therefore aimed to investigate structure and binding at neurotransmitter channels and transporters, using a variety of approaches.

In Chapter 3, novel epibatidine and cytisine analogues were studied using *Ac*AChBP as a surrogate of the human nAChR binding site. For both sets of ligands, binding affinities were identified and molecular interactions elucidated by co-crystallisation. While interactions with the highly conserved “aromatic box” appeared to be responsible for the affinity of the analogues, interactions with complementary side residues appeared to mediate the compounds properties and also at least partially affect the affinity. In particular, Met133 on loop E was suggested to be important with both compound sets, as differences in the sidechain orientation were associated with changes in ligand affinity. These findings can now be factored into the design of future potential ligands and have potential relevance to nAChR subtype selectivity. The equivalent residue to Met133 is a Phe in the β_2 subunit and a Gln in the α_7, therefore offering the probability of different interactions at each subtype. These differences between *Ac*AChBP and human nAChR subunits however do raise the point that caution needs to be used with data obtained using a surrogate. There did not appear to be any correlation for instance between the affinities measured for the derivatives at *Ac*AChBP and those previously reported at the $\alpha_4\beta_2$ nAChR.

There are several possibilities for future work to further characterise the potential properties of the two sets of derivatives at nAChRs. Firstly, only a subset of the ligands in each set could be successfully analysed with both binding assays and crystallography. While ITC was suggested to be unfeasible with the remaining cytisine derivatives, an alternative technique such as surface plasmon resonance (SPR) could be used. This would not be able to give the detailed thermodynamic

information provided by ITC however. Crystals were not obtained for epibatidine derivatives (4)-(6) and it was not attempted to grow crystals for the remaining cytisine derivatives as it was unclear if they could bind *Ac*AChBP. Implementing random matrix microseeding into crystallisation screening may potentially assist with growing co-crystals of the remaining ligands.

The studies of these derivatives used *Ac*AChBP without any modifications. While this offers the likely general interactions of a ligand in an nAChR binding site, *Ac*AChBP does not directly represent any human nAChR subunit combination. There are especially differences on the complementary side of the binding site, where many of the derivatives were found to make ligand specific interactions. One course of action therefore could be to engineer the *Ac*AChBP binding site to be more similar to that of the interface for the nAChR subunit combination of choice. This could give more representative information about the binding affinity and molecular interactions, while maintaining many of the advantages of using *Ac*AChBP.

It would be most directly relevant though if the molecular interactions of the ligands could be characterised in the binding site of actual nAChRs. In recent years, it has become increasingly feasible to obtain structures of nAChRs and other PLGICs by cryo-EM and ultimately this could become a routine technique for characterising PLGIC-ligand interactions. An increasing number of nAChR structures are available in the PDB though in various conformations and bound with differing ligands. In due course, if high quality structures are available, then it may not be absolutely necessary to obtain structures with the ligands bound. *In-silico* docking and molecular dynamics simulations could offer a viable alternative.

In Chapter 4, it was attempted to establish a method for recombinant production of the insect GABA receptor RDL, with an objective of conducting eventual structural and binding studies. Initial production attempts in insect cells were problematic however. Despite being readily trafficked to the plasma membrane, only a small proportion of the protein could be extracted with mild non-ionic detergents. As harsh zwitterionic detergents could extract several fold more protein, it was implied that a large proportion of the protein was in an un-folded state. Production in mammalian HEK293 cells was therefore investigated.

Solubility screening tests in adherent HEK293 cells suggested that non-ionic detergents could extract a much higher proportion of protein compared to the previous insect cell experiments. This therefore implied that a higher level of protein was correctly folded, although total levels of protein appeared to be much lower. Initial attempts at extraction and purification from suspension HEK293 cells were therefore made. These were complicated however by the low level of protein present and only a microscale amount was ultimately able to be purified when FC14 was used for extraction.

While attempts to produce sufficient levels of purified *Dm* RDL for structural and binding studies were ultimately unsuccessful, initial results as presented suggest that production in mammalian cells could represent a viable approach if improvements can be made to the levels of protein produced.

Several possible avenues for future work with RDL have been identified. Firstly, the folding screen assay utilised is unable to directly measure the level of correctly folded protein and is only a crude indicator. In future, combining this assay with radioligand saturation binding assays to measure the level of functional protein in each detergent soluble sample compared to the total in cells/membranes would provide more direct evidence of if the folding state is a factor in the difference between insect and mammalian cells. This would also provide direct quantitative measurements of which detergents extract the greatest level of functional protein.

While there have been multiple studies into the functioning of RDL using electrophysiology in a wide variety of systems, there have been relatively few investigations into its underlying biochemistry in terms of cell localisation, trafficking, PTMs, interacting partners, chaperone proteins etc. Identification of these in native tissue and *in-vitro* could help to inform on how protein constructs and culturing conditions could be modified to increase yields. Identification and selective removal of glycosylation sites and truncations of the N- and C-termini or ICD could represent starting points.

The linear PEI transfection approach appeared to induce production of only a small amount of total protein in HEK293 cells, although a higher proportion could be extracted compared to insect cells. This was surprising as several studies

have used linear PEI for high level production of PLGICs (Miller and Aricescu, 2014; Kasaragod *et al.*, 2022). Using BacMam (Claxton and Gouaux, 2018; Phulera *et al.*, 2018; Zhu *et al.*, 2018) or developing an inducible stable cell line (Zhang and Miller, 2015; Miller *et al.*, 2018; Laverty *et al.*, 2019) are potential alternatives for producing RDL which have both been widely used for producing PLGICs in HEK293 cells. Other potential factors which could be optimised include the culture temperature and the use of enhancers. Several groups (Claxton and Gouaux, 2018; Zhu *et al.*, 2018) describing high level production of human $GABA_A$ receptors in HEK293 cells lowered the culture temperature from 37 °C to 32 °C or even 30 °C after transfection as this was found to raise yields and improve protein monodispersity. Due to the logistical problems of being unable to dedicate a separate lower temperature CO_2 incubator, this was unfortunately not possible to test with mGFPuv-$RDL_{ac(MAM)}$ in HEK293 or Expi293 cells. The use of enhancers such as histone deacetylase inhibitors (e.g. sodium butyrate or valproic acid) to boost production are other aspects which could be tested. While sodium butyrate was added to 10 mM 12-hours after transfection when producing mGFPuv-$RDL_{ac(MAM)}$ in Expi293 cells, there was not a systemic study on the concentration used or time-point of addition. Optimisation of both of these has been suggested to be important for obtaining maximal yields (Claxton and Gouaux, 2018).

The high total, but low extractable amount of RDL in *Sf*9 insect cells with baculovirus is intriguing. The obtained FSEC traces pointed towards the small amount of detergent soluble protein as being relatively well behaved with little to no aggregation and a fair to good degree of monodispersity for two out of the four detergents tested. If the proportion of extractable protein could be increased, then *Sf*9 cells could represent a much more economical method than mammalian cells for producing *Dm* RDL. There are multiple potential factors which could be investigated. One could be to use a baculovirus with the gene encoding RDL under the control of an earlier and weaker promoter, which has been successful in enhancing yields by several fold of some membrane proteins that worked poorly with a conventional very late polyhedrin or P10 promotor (Higgins *et al.*, 2003; Furukawa *et al.*, 2021). This approach could solve the postulated folding issues and was later particularly successful with GAT1 in Chapter 5. Another could be to co-infect insect cells with a baculovirus encoding mammalian ER

quality control chaperones such as calnexin or calreticulin. These have been demonstrated in some examples to improve yields of correctly processed and functional and therefore folded membrane proteins in insect cells, including the *Dm* shaker voltage gated potassium channel (Tate *et al.*, 1999). Finally, the cholesterol content of *Sf*9 cell membranes could be altered to see if increased or decreased membrane cholesterol can improve folding and extraction of *Dm* RDL (Gimpl *et al.*, 1995).

In Chapter 5, after much optimisation of the culturing, extraction and purification procedures, a method was established for recombinant production of the human GABA transporter 1. Initial experiments used baculovirus infection of *Sf*9 cells, with a construct of GAT1 in pFastBac 1 vector under the control of a very late polyhedrin promoter. With this method, while the protein could be successfully extracted with mild non-ionic detergents, during purification there were problems with aggregation and only a small amount of purified protein could be ultimately obtained. Switching to the EarlyBac system with GAT1 under the control of an *Dm* Hsp70 early promoter however led to the successful high level production of GAT1 from *Sf*9 cells in a pure form. This protein will hopefully be appropriate for future structural studies by cryo-EM and initial negative stain EM micrographs looked promising. A set of candidate recombinant nanobodies against purified GAT1 were also isolated for potential future use as a binding partner for GAT1 with cryo-EM. However, it was not possible to determine if the protein was active and capable of transporting GABA when produced in *Sf*9 cells, as the FMP blue assay appeared to be unsuitable for this purpose.

There are multiple possible threads of work that could be picked up in future to expand on the success of producing purified GAT1. Firstly, the nanobodies raised against GAT1 need to be grown in *E.coli* and purified so that their binding specificity and affinity to GAT1 can be identified. There are multiple methods which could be used for this, ranging from simple, such as seeing an electrophoretic shift on native-PAGE or a difference in retention on SEC, to complex e.g., SPR and radioligand binding assays. If a suitable nanobody can be identified, then it may be of assistance in increasing the size of the protein for future cryo-EM studies.

Although a cryo-EM structure of GAT1 has now been obtained by Motiwala *et al.*, (2022), this only represents the inward-open conformation of the transporter and it is in complex with tiagabine. As such, the ion binding pockets are empty and details are still uncertain regarding how GABA binds in the substrate binding pocket. Further structures in other conformations and bound with differing molecules therefore need to be obtained. It is technically impressive that this structure was able to be obtained using the full-length protein with no sequence modifications and without any other partner molecules to increase the size. However, the overall resolution of 3.8 Å is lower than for other transporter structures. Although the binding of a nanobody has the potential to disrupt transport activity, they can also lock the transporter in a single conformation and the increased size of the protein will likely help enable obtaining higher resolution structures of GAT1. Such resolution will likely be necessary to ascertain details of ions binding when the protein is in the relevant conformations.

Another possibility is that it may be possible to crystallise GAT1 using the lipidic-cubic phase (LCP) technique. Shahsavar *et al.* (2021) used LCP crystallisation to obtain micro-crystals of GlyT1, which were diffracted on a micro focus beamline. They were ultimately able to construct a 3.4 Å model after merging multiple datasets. It should be noted however that GlyT1 in this study was extensively modified to remove all except for one glycosylation site (which was later removed with PNGase F) and to shorten the length of certain loops. A nanobody could also be of potential assistance with LCP by stabilising parts of GAT1. Nanobodies are frequently used to stabilise GPCRs during crystallisation for instance (Steyaert and Kobilka, 2011).

The initial negative stain EM images looked to be promising, showing what visually appeared to be a fairly monodisperse distribution. However, it is unclear as to which oligomeric state(s) GAT1 is present in. Even on SDS-PAGE, the protein appeared to run with a readily visible additional dimer band, as well as fainter bands up to potentially pentameric or greater forms. It is uncertain though if these are true oligomers or LDS resistant aggregates. Mass photometery could give an initial indication of the distribution of oligomeric states which are present, but this technique only has limited suitability with detergent containing samples. A sample therefore ideally needs to be analysed with a technique such as SEC-

MALS or AUC. So far there is only very limited direct structural information on SLC6 transporters in a dimeric form. As cryo-EM is not as sensitive to deviations from monodispersity compared to other techniques, it may be possible to obtain structures of multiple GAT1 oligomeric states from a single grid.

On a related note, the protein was only tested being purified into DDM/CHS. If there is a wide distribution of oligomeric states, using another detergent such as LMNG/CHS (as used by Motiwala *et al.*, (2022) may reduce the heterogeneity or encourage formation of a particular oligomer state. An extension would be to exchange the initially detergent solubilised protein into lipid nanodiscs. This has been particularly successful with other membrane proteins (Masiulis *et al.*, 2019), plus details of lipid interactions would be available. Furthermore, protein could be directly extracted from cells into SMA co-polymer. Initial tests in section S1 suggest that may be a viable method.

When tested in HEK293 cells, the GAT1-mEm construct appeared to be active from the FMP blue assay. The functionality of the protein produced from the EarlyBac system however also needs to be established. It was unclear from the FMP blue assay as to whether the protein was inactive or the assay was unsuitable with *Sf*9 cells. Although expensive and having safety issues, experiments measuring uptake of radiolabelled GABA from cells or membranes likely represent the best way forward. Another option could be to fluorescently label the purified protein for microscale thermophoresis measurements to measure the binding affinity of GABA or other ligands, which is a method that has been used by some studies (Tsai *et al.*, 2020).

References

Adams, M.J., Blundell, T.L., Dodson, E.J., Dodson, G.G., Vijayan, M., Baker, E.N., *et al.* (1969). Structure of rhombohedral 2 zinc insulin crystals. Nature *224*: 491–495.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular biology of the cell (New York, NY: Garland Science).

Alexander, S.P.H., Kelly, E., Mathie, A., Peters, J.A., Veale, E.L., Armstrong, J.F., *et al.* (2019a). The concise guide to pharmacology 2019/20: transporters. Br. J. Pharmacol. *176*: S397–S493.

Alexander, S.P.H., Mathie, A., Peters, J.A., Veale, E.L., Striessnig, J., Kelly, E., *et al.* (2019b). The concise guide to pharmacology 2019/20: ion channels. Br. J. Pharmacol. *176*: S142–S228.

Andersen, N., Corradi, J., Sine, S.M., and Bouzat, C. (2013). Stoichiometry for activation of neuronal α7 nicotinic receptors. Proc. Natl. Acad. Sci. *110*: 20819–20824.

Aronstein, K., Auld, V., and Ffrench-Constant, R. (1996). Distribution of two GABA receptor-like subunits in the *Drosophila* CNS. Invertebr. Neurosci. *2*: 115–120.

Aronstein, K., and Ffrench-Constant, R. (1995). Immunocytochemistry of a novel GABA receptor subunit Rdl in *Drosophila melanogaster*. Invertebr. Neurosci. *1*: 25–31.

Aronstein, K., Ode, P., and Ffrench-Constant, R.H. (1995). PCR based monitoring of specific *Drosophila* (Diptera: Drosophilidae) cyclodiene resistance alleles in the presence and absence of selection. Bull. Entomol. Res. *85*: 5–9.

Ashby, J.A.A., McGonigle, I.V. V., Price, K.L.L., Cohen, N., Comitani, F., Dougherty, D.A.A., *et al.* (2012). GABA binding to an insect GABA receptor: a molecular dynamics and mutagenesis study. Biophys. J. *103*: 2071–2081.

Ashikawa, Y., Ihara, M., Matsuura, N., Fukunaga, Y., Kusakabe, Y., and Yamashita, A. (2011). GFP-based evaluation system of recombinant expression through the secretory pathway in insect cells and its application to the

extracellular domains of class C GPCRs. Protein Sci. *20*: 1720–1734.

Ayuyan, A.G., and Cohen, F.S. (2008). Raft composition at physiological temperature and pH in the absence of detergents. Biophys. J. *94*: 2654.

Bai, X., McMullan, G., and Scheres, S.H.W. (2015). How cryo-EM is revolutionizing structural biology. Trends Biochem. Sci. *40*: 49–57.

Bai, X., Moraes, T.F., and Reithmeier, R.A.F. (2017). Structural biology of solute carrier (SLC) membrane transport proteins. Mol. Membr. Biol. *34*: 1–32.

Bailey, L.J., Sheehy, K.M., Dominik, P.K., Liang, W.G., Rui, H., Clark, M., *et al.* (2018). Locking the Elbow: Improved antibody Fab fragments as chaperones for structure determination. J. Mol. Biol. *430*: 337.

Baptista-Hon, D.T., Deeb, T.Z., Lambert, J.J., Peters, J.A., and Hales, T.G. (2013). The Minimum M3-M4 loop length of neurotransmitter-activated pentameric receptors Is critical for the structural integrity of cytoplasmic portals. J. Biol. Chem. *288*: 21558–21568.

Basak, S., Gicheru, Y., Kapoor, A., Mayer, M.L., Filizola, M., and Chakrapani, S. (2019). Molecular mechanism of setron-mediated inhibition of full-length 5-HT_{3A} receptor. Nat. Commun. 2019 101 *10*: 3225.

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47*: D506–D515.

Bauer, J., and Cooper-Mahkorn, D. (2008). Tiagabine: efficacy and safety in partial seizures - current status. Neuropsychiatr. Dis. Treat. *4*: 731–736.

Bennett, E.R., and Kanner, B.I. (1997). The membrane topology of GAT-1, a (Na^++Cl^-)-coupled γ-aminobutyric acid transporter from rat brain. J. Biol. Chem. *272*: 1203–1210.

Benowitz, N.L. (2009). Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu. Rev. Pharmacol. Toxicol. *49*: 57–71.

Benowitz, N.L., Hukkanen, J., and Jacob, P. (2009). Nicotine chemistry, metabolism, kinetics and biomarkers. In handbook of experimental pharmacology, (NIH Public Access), pp 29–60.

Bergfors, T. (2007a). Screening and optimization methods for nonautomated crystallization laboratories. In methods in molecular biology: macromolecular crystallography protocols: volume 1, S. Doublié, ed. (Humana Press Inc), pp 131–151.

Bergfors, T. (2007b). Succeeding with seeding: some practical advice. In evolving methods for macromolecular crystallography, R.J. Reed, and J.L. Sussman, eds. (Springer), pp 1–10.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nat. Struct. Mol. Biol. *10*: 980–980.

Bernhard, M., and Laube, B. (2020). Thermophoretic analysis of ligand-specific conformational states of the inhibitory glycine receptor embedded in copolymer nanodiscs. Sci. Reports 2020 101 *10*: 1–11.

Bertrand, D., Lee, C.-H.L., Flood, D., Marger, F., and Donnelly-Roberts, D. (2015). Therapeutic potential of α_7 nicotinic acetylcholine receptors. Pharmacol. Rev. *67*: 1025–1073.

Bill, R.M. (2014). Playing catch-up with escherichia coli: Using yeast to increase success rates in recombinant protein production experiments. Front. Microbiol. *5*: 85.

Billen, B., Spurny, R., Brams, M., Elk, R. van, Valera-Kummer, S., Yakel, J.L., *et al.* (2012). Molecular actions of smoking cessation drugs at $\alpha_4\beta_2$ nicotinic receptors defined in crystal structures of a homologous binding protein. Proc. Natl. Acad. Sci. *109*: 9173 LP – 9178.

Birch, J., Axford, D., Foadi, J., Meyer, A., Eckhardt, A., Thielmann, Y., *et al.* (2018). The fine art of integral membrane protein crystallisation. Methods *147*: 150–162.

Birch, J., Cheruvara, H., Gamage, N., Harrison, P.J., Lithgo, R., and Quigley, A. (2020). Changes in membrane protein structural biology. Biology (Basel). *9*: 401.

Bloch, J.S., Mukherjee, S., Kowal, J., Filippova, E. V., Niederer, M., Pardon, E., *et al.* (2021). Development of a universal nanobody-binding Fab module for fiducial-assisted cryo-EM studies of membrane proteins. Proc. Natl. Acad. Sci. U. S. A. *118*:.

Blow, D. (2004). Outline of crystallography for biologists (Oxford University Press).

Blum, A.P., Lester, H.A., and Dougherty, D.A. (2010). Nicotinic pharmacophore: The pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. Proc. Natl. Acad. Sci. U. S. A. *107*: 13206–13211.

Bobango, J., Wu, J., Talley, I.T., and Talley, T. Crystal structure of a chimeric acetylcholine binding protein from Aplysia californica (Ac-AChBP) containing loop C from the human α_3 nicotinic acetylcholine receptor in complex with Cytisine. *To be published. http://doi.org/10.2210/pdb5SYO/pdb*

Bondarenko, V., Tillman, T., Xu, Y., and Tang, P. (2010). NMR structure of the transmembrane domain of the n-acetylcholine receptor β2 subunit. Biochim. Biophys. Acta - Biomembr. *1798*: 1608–1614.

Brams, M., Pandya, A., Kuzmin, D., Elk, R. van, Krijnen, L., Yakel, J.L., *et al.* (2011). A structural and mutagenic blueprint for molecular recognition of strychnine and d-tubocurarine by different Cys-loop receptors. PLoS Biol. *9*: 01034.

Brejc, K., Dijk, W.J. van, Klaassen, R. V., Schuurmans, M., Oost, J. van der, Smit, A.B., *et al.* (2001). Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. Nature *411*: 269–276.

Brodsky, J.L., and Skach, W.R. (2011). Protein folding and quality control in the endoplasmic reticulum: recent lessons from yeast and mammalian cell systems. Curr. Opin. Cell Biol. *23*: 464–475.

Bröer, S., and Gether, U. (2012). The solute carrier 6 family of transporters. Br. J. Pharmacol. *167*: 256–278.

Brown, D.A., and Rose, J.K. (1992). Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. Cell *68*: 533–544.

Brunzell, D.H., Stafford, A.M., and Dixon, C.I. (2015). Nicotinic receptor contributions to smoking: insights from human studies and animal models. Curr. Addict. Reports *2*: 33–46.

Buckingham, S.D., Biggin, P.C., Sattelle, B.M., Brown, L.A., and Sattelle, D.B. (2005). Insect GABA receptors: splicing, editing, and targeting by antiparasitics and insecticides. Mol. Pharmacol. *68*: 942–951.

Buckingham, S.D., Hosie, A.M., Roush, R.L., and Sattelle, D.B. (1994). Actions of agonists and convulsant antagonists on a *Drosophila melanogaster* GABA receptor (Rdl) homo-oligomer expressed in Xenopus oocytes. Neurosci. Lett. *181*: 137–140.

Bueno, R. V., Davis, S., Dawson, A., Ondachi, P.W., Carroll, F.I., and Hunter, W.N. (2022). Interactions between 2′-fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues and acetylcholine binding protein inform on potent antagonist activity against nicotinic receptors. Acta Crystallogr. Sect. D Struct. Biol. *78*: 353–362.

Burgess, R.R. (2009). Chapter 44 important but little known (or forgotten) artifacts in protein biochemistry. In methods in enzymology, (Elsevier Inc.), pp 813–820.

Cahill, K., Lindson-Hawley, N., Thomas, K.H., Fanshawe, T.R., and Lancaster, T. (2016). Nicotine receptor partial agonists for smoking cessation. Cochrane Database Syst. Rev. *2016*:.

Cai, G., Salonikidis, P.S., Fei, J., Schwarz, W., Schülein, R., Reutter, W., *et al.* (2005). The role of N-glycosylation in the stability, trafficking and GABA-uptake of GABA-transporter 1. FEBS J. *272*: 1625–1638.

Carson, R. (1962). Silent Spring (London, England: Penguin Books).

Carter, D.B., Thomsen, D.R., Im, W.B., Lennon, D.J., Ngo, D.M., Gale, W., *et al.* (1992). Functional expression of $GABA_A$ chloride channels and benzodiazepine binding sites in baculovirus infected insect cells. Biotechnology. (N. Y). *10*: 679–81.

Carvelli, L., Blakely, R.D., and DeFelice, L.J. (2008). Dopamine transporter/syntaxin 1A interactions regulate transporter channel activity and dopaminergic synaptic transmission. Proc. Natl. Acad. Sci. *105*: 14192–14197.

Casida, J.E., and Durkin, K.A. (2013). Neuroactive insecticides: targets, selectivity, resistance, and secondary effects. Annu. Rev. Entomol. *58*: 99–117.

Casida, J.E., and Durkin, K.A. (2015). Novel GABA receptor pesticide targets. Pestic. Biochem. Physiol. *121*: 22–30.

Castellanos-Serra, L., and Hardy, E. (2006). Negative detection of biomolecules separated in polyacrylamide electrophoresis gels. Nat. Protoc. *1*: 1544–1551.

Caton, M., Ochoa, E.L.M., and Barrantes, F.J. (2020). The role of nicotinic cholinergic neurotransmission in delusional thinking. Schizophrenia *6*: 16.

Celie, P.H.., Rossum-Fikkert, S.E. van, Dijk, W.J. van, Brejc, K., Smit, A.B., and Sixma, T.K. (2004). Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. Neuron *41*: 907–914.

Cheloha, R.W., Harmand, T.J., Wijne, C., Schwartz, T.U., and Ploegh, H.L. (2020). Exploring cellular biochemistry with nanobodies. J. Biol. Chem. *295*: 15307–15327.

Chen, R., Belellit, D., Lambert, J.J., Peters, J.A., Reyes, A., and Lan, N.C. (1994). Cloning and functional expression of a *Drosophila* y-aminobutyric acid receptor. Proc. Natl. Acad. Sci. *91*: 6069–6073.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., *et al.* (2010). MolProbity : all-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*: 12–21.

Cheng, Y., Grigorieff, N., Penczek, P.A., and Walz, T. (2015). A primer to single-particle cryo-electron microscopy. Cell *161*: 438–449.

Claxton, D.P., and Gouaux, E. (2018). Expression and purification of a functional heteromeric gabaa receptor for structural studies. PLoS One *13*: e0201210.

Coe, J.W., Brooks, P.R., Vetelino, M.G., Wirtz, M.C., Arnold, E.P., Huang, J., *et al.* (2005). Varenicline: An $\alpha_4\beta_2$ nicotinic receptor partial agonist for smoking cessation. J. Med. Chem. *48*: 3474–3477.

Colas, C. (2020). Toward a systematic structural and functional annotation of solute carriers transporters - example of the SLC6 and SLC7 families. Front. Pharmacol. *11*: 1229.

Coleman, J.A., Green, E.M., and Gouaux, E. (2016). X-ray structures and

mechanism of the human serotonin transporter. Nature *532*: 334–339.

Coleman, J.A., Yang, D., Zhao, Z., Wen, P.-C., Yoshioka, C., Tajkhorshid, E., *et al.* (2019). Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport. Nature *569*: 141–145.

Colquhoun, D., and Lape, R. (2012). Allosteric coupling in ligand-gated ion channels. J. Gen. Physiol. *140*: 599–612.

Conde, R., Cueva, R., Pablo, G., Polaina, J., and Larriba, G. (2004). A Search for hyperglycosylation signals in yeast glycoproteins. J. Biol. Chem. *279*: 43789–43798.

Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E., and Chisti, Y. (2014). Protein production using the baculovirus-insect cell expression system. Biotechnol. Prog. *30*: 1–18.

Corbel, V., and N'Guessan, R. (2013). Distribution, mechanisms, impact and management of insecticide resistance in malaria vectors: a pragmatic review. In Anopheles mosquitoes - new insights into malaria vectors, S. Manguin, ed. (London: InTech), pp 579–631.

Corringer, P.J., Poitevin, F., Prevost, M.S., Sauguet, L., Delarue, M., and Changeux, J.P. (2012). Structure and pharmacology of pentameric receptor channels: From bacteria to brain. Structure *20*: 941–956.

daCosta, C.J.B., and Baenziger, J.E. (2013). Gating of Pentameric Ligand-Gated Ion Channels: Structural Insights and Ambiguities. Structure *21*: 1271–1283.

Dallanoce, C., Matera, C., Amici, M. De, Rizzi, L., Pucci, L., Gotti, C., *et al.* (2012). The enantiomers of epiboxidine and of two related analogs: Synthesis and estimation of their binding affinity at α4β2 and α7 neuronal nicotinic acetylcholine receptors. Chirality *24*: 543–551.

Danev, R., Belousoff, M., Liang, Y.-L., Zhang, X., Eisenstein, F., Wootten, D., *et al.* (2021). Routine sub-2.5 Å cryo-EM structure determination of GPCRs. Nat. Commun. *12*: 4333.

Dani, J.A., and Biasi, M. De (2001). Cellular mechanisms of nicotine addiction. Pharmacol. Biochem. Behav. *70*: 439–446.

Dauter, Z., Jaskolski, M., and Wlodawer, A. (2010). Impact of synchrotron radiation on macromolecular crystallography: a personal view. J. Synchrotron Radiat. *17*: 433–444.

Davies, P.A., Hoffmann, E.B., Carlisle, H.J., Tyndale, R.F., and Hales, T.G. (2000). The influence of an endogenous β_3 subunit on recombinant GABAA receptor assembly and pharmacology in WSS-1 cells and transiently transfected HEK293 cells. Neuropharmacology *39*: 611–620.

Davies, P.A., Wang, W., Hales, T.G., and Kirkness, E.F. (2003). A novel class of ligand-gated ion channel is activated by Zn^{2+}. J. Biol. Chem. *278*: 712–717.

Davies, T.G.E., Field, L.M., Usherwood, P.N.R., and Williamson, M.S. (2007). DDT, pyrethrins, pyrethroids and insect sodium channels. IUBMB Life *59*: 151–162.

Davis, S., Rego Campello, H., Gallagher, T., and Hunter, W.N. (2020). The thermodynamic profile and molecular interactions of a C(9)-cytisine derivative-binding acetylcholine-binding protein from Aplysia californica. Acta Crystallogr. Sect. F Struct. Biol. Commun. *76*: 74–80.

Dawson, A., Trumper, P., Souza, J.O. de, Parker, H., Jones, M.J., Hales, T.G., *et al.* (2019). Engineering a surrogate human heteromeric α/β glycine receptor orthosteric site exploiting the structural homology and stability of acetylcholine-binding protein. IUCrJ *6*: 1–10.

Deken, S.L., Beckman, M.L., Boos, L., and Quick, M.W. (2000). Transport rates of GABA transporters: regulation by the N-terminal domain and syntaxin 1A. Nat. Neurosci. 2000 310 *3*: 998–1003.

Denisov, I.G., and Sligar, S.G. (2016). Nanodiscs for structural and functional studies of membrane proteins. Nat. Struct. Mol. Biol. *23*: 481–486.

Dhar, T.G.M., Borden, L.A., Tyagarajan, S., Smith, K.E., Branchek, T.A., Weinshank, R.L., *et al.* (1994). Design, synthesis and evaluation of substituted triarylnipecotic acid derivatives as GABA uptake inhibitors: identification of a ligand with moderate affinity and selectivity for the cloned human GABA transporter GAT-3. J. Med. Chem. *37*: 2334–2342.

Dilworth, M. V., Piel, M.S., Bettaney, K.E., Ma, P., Luo, J., Sharples, D., *et al.*

(2018). Microbial expression systems for membrane proteins. Methods *147*: 3–39.

Domingues, L.N., Guerrero, F.D., Becker, M.E., Alison, M.W., and Foil, L.D. (2013). Discovery of the Rdl mutation in association with a cyclodiene resistant population of horn flies, *Haematobia irritans* (Diptera: Muscidae). Vet. Parasitol. *198*: 172–179.

Drew, D., Newstead, S., Sonoda, Y., Kim, H., Heijne, G. von, and Iwata, S. (2008). GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. Nat. Protoc. *3*: 784.

Drew, D.E., Heijne, G. Von, Nordlund, P., and Gier, J.W.L. De (2001). Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. FEBS Lett. *507*: 220–224.

Dubochet, J., Lepault, J., Freeman, R., Berriman, J.A., and Homo, J.-C. (1982). Electron microscopy of frozen water and aqueous solutions. J. Microsc. *128*: 219–237.

Edwards, G.B., Muthurajan, U.M., Bowerman, S., and Luger, K. (2020). Analytical ultracentrifugation (AUC): an overview of the application of fluorescence and absorbance auc to the study of biological macromolecules. Curr. Protoc. Mol. Biol. *133*: e131.

Egli, M. (2016). Diffraction techniques in structural biology. Curr. Protoc. Nucleic Acid Chem. *65*: 7.13.1-7.13.41.

Elster, L., Schousboe, A., and Olsen, R.W. (2000). Stable $GABA_A$ receptor intermediates in *Sf*-9 cells expressing α_1, β_2 and γ_2 subunits. J. Neurosci. Res. *61*: 193–205.

Emsley, P., and Cowtan, K. (2004). Coot : model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*: 2126–2132.

Es-Salah, Z., Lapied, B., Goff, G. Le, and Hamon, A. (2008). RNA editing regulates insect γ-aminobutyric acid receptor function and insecticide sensitivity. Neuroreport *19*: 939–943.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. *69*: 1204–1214.

Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J.-C., Zhao, L., *et al.* (2019). Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. Nat. Commun. *10*: 2386.

Farhan, H., Reiterer, V., Korkhov, V.M., Schmid, J.A., Freissmuth, M., and Sitte, H.H. (2007). Concentrative export from the endoplasmic reticulum of the γ-aminobutyric acid transporter 1 requires binding to SEC24D. J. Biol. Chem. *282*: 7679–7689.

Ffrench-Constant, R.H., Mortlock, D.P., Shaffer, C.D., MacIntyre, R.J., and Roush, R.T. (1991). Molecular cloning and transformation of cyclodiene resistance in *Drosophila*: an invertebrate γ-aminobutyric acid subtype A receptor locus. Proc. Natl. Acad. Sci. *88*: 7209 LP – 7213.

ffrench-Constant, R.H., and Rocheleau, T.A. (1993). *Drosophila* γ-aminobutyric acid receptor gene Rdl shows extensive alternative splicing. J. Neurochem. *60*: 2323–2326.

Ffrench-Constant, R.H., Rocheleau, T.A., Steichen, J.C., and Chalmers, A.E. (1993a). A point mutation in a *Drosophila* GABA receptor confers insecticide resistance. Nature *363*: 449–451.

Ffrench-Constant, R.H., Steichen, J.C., Rocheleau, T.A., Aronstein, K., and Roush, R.T. (1993b). A single-amino acid substitution in a γ-aminobutyric acid subtype A receptor locus is associated with cyclodiene insecticide resistance in *Drosophila* populations. Proc. Natl. Acad. Sci. *90*: 1957–1961.

Focke, P.J., Wang, X., and Larsson, H.P. (2013). Neurotransmitter transporters: structure meets function. Structure *21*: 694–705.

Foulds, J. (2006). The neurobiological basis for partial agonist treatment of nicotine dependence: varenicline. Int. J. Clin. Pract. *60*: 571–576.

Francis, D.M., and Page, R. (2010). Strategies to optimize protein expression in E. coli. Curr. Protoc. Protein Sci. *61*: 5241.

Furukawa, H., Simorowski, N., and Michalski, K. (2021). Effective production of

oligomeric membrane proteins by EarlyBac-insect cell system. In methods in enzymology, (Academic Press), pp 3–19.

Galanopoulou, A. (2008). $GABA_A$ receptors in normal development and seizures: friends or foes? Curr. Neuropharmacol. *6*: 1–20.

Gao, K., Oerlemans, R., and Groves, M.R. (2020). Theory and applications of differential scanning fluorimetry in early-stage drug discovery. Biophys. Rev. *12*: 85–104.

Gassel, M., Wolf, C., Noack, S., Williams, H., and Ilg, T. (2014). The novel isoxazoline ectoparasiticide fluralaner: Selective inhibition of arthropod γ-aminobutyric acid- and l-glutamate-gated chloride channels and insecticidal/acaricidal activity. Insect Biochem. Mol. Biol. *45*: 111–124.

Gati, C. (2022). https://twitter.com/CorneliusGati/status/153503254.

Gharpure, A., Teng, J., Zhuang, Y., Noviello, C.M., Walsh, R.M., Cabuco, R., *et al.* (2019). Agonist selectivity and ion permeation in the $\alpha_3\beta_4$ ganglionic nicotinic receptor. Neuron *104*: 501-511.e6.

Gimpl, G., Klein, U., Reilaender, H., and Fahrenholz, F. (1995). Expression of the human oxytocin receptor in baculovirus-infected insect cells: high-affinity binding is induced by a cholesterol-cyclodextrin complex. Biochemistry *34*: 13794–13801.

Giraudo, A., Krall, J., Bavo, F., Nielsen, B., Kongstad, K.T., Rolando, B., *et al.* (2019). Five-membered N-heterocyclic scaffolds as novel amino bioisosteres at γ-aminobutyric acid (GABA) type A receptors and GABA transporters. J. Med. Chem. *62*: 5797–5809.

Gisselmann, G., Plonka, J., Pusch, H., and Hatt, H. (2004). *Drosophila melanogaster* GRD and LCCH3 subunits form heteromultimeric GABA-gated cation channels. Br. J. Pharmacol. *142*: 409–413.

Gonzales, A.L., Lee, W., Spencer, S.R., Oropeza, R.A., Chapman, J. V., Ku, J.Y., *et al.* (2007). Turnover rate of the γ-aminobutyric acid transporter GAT1. J. Membr. Biol. *220*: 33–51.

Gotti, C., Clementi, F., Fornari, A., Gaimarri, A., Guiducci, S., Manfredi, I., *et al.*

(2009). Structural and functional diversity of native brain neuronal nicotinic receptors. Biochem. Pharmacol. *78*: 703–711.

Gotti, C., Zoli, M., and Clementi, F. (2006). Brain nicotinic acetylcholine receptors: native subtypes and their relevance. Trends Pharmacol. Sci. *27*: 482–491.

Gottschald Chiodi, C. (2017). A biophysical and computational study of the $GABA_AR$ orthosteric site. University of Dundee.

Gottschald Chiodi, C., Baptista-Hon, D.T., Hunter, W.N., and Hales, T.G. (2018). Amino acid substitutions in the human homomeric β_3 $GABA_A$ receptor that enable activation by GABA. J. Biol. Chem. *294*: 2375–2385.

Graham, J. (2002). Preparation of crude subcellular fractions by differential centrifugation. Sci. World J. *2*: 1638–1642.

Grau-Bové, X., Tomlinson, S., O'Reilly, A.O., Harding, N.J., Miles, A., Kwiatkowski, D., *et al.* (2020). Evolution of the Insecticide target Rdl in African *Anopheles* is driven by interspecific and interkaryotypic introgression. Mol. Biol. Evol. *37*: 2900–2917.

Greenfield, L.J. (2013). Molecular mechanisms of antiseizure drug activity at $GABA_A$ receptors. Seizure *22*: 589–600.

Grolleau, F., and Sattelle, D.B. (2000). Single channel analysis of the blocking actions of BIDN and fipronil on a *Drosophila melanogaster* GABA receptor (RDL) stably expressed in a *Drosophila* cell line. Br. J. Pharmacol. *130*: 1833–1842.

Guastella, J., Nelson, N., Nelson, H., Czyzyk, L., Keynan, S., Miedel, M.C., *et al.* (1990). Cloning and expression of a rat brain GABA transporter. Science *249*: 1303–1306.

Guo, Y. (2020). Be cautious with crystal structures of membrane proteins or complexes prepared in detergents. Crystals *10*: 86.

Guo, Y. (2021). Detergent-free systems for structural studies of membrane proteins. Biochem. Soc. Trans. *49*: 1361–1374.

Guyton, A.C., and Hall, J.E. (2006). Textbook of medical physiology. (Philadelphia, Pennsylvania: Elsevier Saunders), pp 45–56.

Hammad, L.A., Cooper, B.S., Fisher, N.P., Montooth, K.L., and Karty, J.A. (2011). Profiling and quantification of *Drosophila melanogaster* lipids using liquid chromatography/mass spectrometry. Rapid Commun. Mass Spectrom. *25*: 2959–2968.

Hansen, S.B., Sulzenbacher, G., Huxford, T., Marchot, P., Taylor, P., and Bourne, Y. (2005). Structures of *Aplysia* AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. EMBO J. *24*: 3635–3646.

Harrison, J.B., Chen, H.H., Sattelle, E., Barker, P.J., Huskisson, N.S., Rauh, J.J., *et al.* (1996). Immunocytochemical mapping of a C-terminus anti-peptide antibody to the GABA receptor subunit, RDL in the nervous system of *Drosophila melanogaster*. Cell Tissue Res. *284*: 269–278.

Harvey, R.J., Schmitt, B., Hermans-Borgmeyer, I., Gundelfinger, E.D., Betz, H., and Darlison, M.G. (1994). Sequence of a *Drosophila* ligand-gated ion-channel polypeptide with an unusual amino-terminal extracellular domain. J. Neurochem. *62*: 2480–2483.

Hattori, M., Hibbs, R.E., and Gouaux, E. (2012). A fluorescence-detection size-exclusion chromatography-based thermostability assay to identify membrane protein expression and crystallization conditions. Structure *20*: 1293.

Hediger, M.A., Clémençon, B., Burrier, R.E., and Bruford, E.A. (2013). The ABCs of membrane transporters in health and disease (SLC series): Introduction. Mol. Aspects Med. *34*: 95–107.

Henderson, J.E., Soderlund, D.M., and Knipple, D.C. (1993). Characterization of a putative γ-aminobutyric-Acid (GABA) receptor β-subunit gene from *Drosophila melanogaster*. Biochem. Biophys. Res. Commun. *193*: 474–482.

Henderson, R., and Unwin, P.N.T. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. Nature *257*: 28–32.

Henry, C., Cens, T., Charnet, P., Cohen-Solal, C., Collet, C., van-Dijk, J., *et al.* (2020). Heterogeneous expression of GABA receptor-like subunits LCCH3 and GRD reveals functional diversity of GABA receptors in the honeybee *Apis mellifera*. Br. J. Pharmacol. *177*: 3924–3940.

Hibbs, R.E., and Gouaux, E. (2011). Principles of activation and permeation in an anion-selective Cys-loop receptor. Nature *474*: 54–60.

Higgins, M.K., Demir, M., and Tate, C.G. (2003). Calnexin co-expression and the use of weaker promoters increase the expression of correctly assembled Shaker potassium channel in insect cells. Biochim. Biophys. Acta - Biomembr. *1610*: 124–132.

Hilf, R.J.C., and Dutzler, R. (2008). X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature *452*: 375–379.

Hilf, R.J.C., and Dutzler, R. (2009). Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. Nature *457*: 115–118.

Hills, D., and Crane-Robinson, C. (1995). Baculovirus expression of human basic fibroblast growth factor from a synthetic gene: role of the Kozak consensus and comparison with bacterial expression. Biochim. Biophys. Acta - Gene Struct. Expr. *1260*: 14–20.

Holz, R., and Fisher, S. (1999). Synaptic transmission. In basic neurochemistry: molecular, cellular and medical aspects. 6th Edition, G.J. Siegel, B. Agranoff, R. Albers, S. Fisher, and M. Uhler, eds. (Lippincott-Raven 1999).

Hosie, A., Sattelle, D., Aronstein, K., and Ffrench-Constant, R. (1997). Molecular biology of insect neuronal GABA receptors. Trends Neurosci. *20*: 578–583.

Howard, R.J. (2021). Elephants in the dark: insights and incongruities in pentameric ligand-gated ion channel models. J. Mol. Biol. *433*: 167128.

Hu, J. (2010). Role of terminal sialic acid of GABA transporter 1 in GABA uptake & purification and characterization of this transporter for structural analysis. Freie Universität Berlin.

Hu, J., Weise, C., Böttcher, C., Fan, H., and Yin, J. (2017). Expression, purification and structural analysis of functional GABA transporter 1 using the baculovirus expression system. Beilstein J. Org. Chem. *13*: 874–882.

HU, Y. (2005). Baculovirus as a highly efficient expression vector in insect and mammalian cells. Acta Pharmacol. Sin. *26*: 405–416.

Huynh, K., and Partch, C.L. (2015). Analysis of protein stability and ligand

interactions by thermal shift assay. Curr. Protoc. Protein Sci. *79*: 28.9.1-28.9.14.

Hyman, S.E. (2005). Neurotransmitters. Curr. Biol. *15*: R154–R158.

Illing, M.E., Rajan, R.S., Bence, N.F., and Kopito, R.R. (2002). A rhodopsin mutant linked to autosomal dominant retinitis pigmentosa is prone to aggregate and interacts with the ubiquitin proteasome system. J. Biol. Chem. *277*: 34150–34160.

Imoukhuede, P.I., Moss, F.J., Michael, D.J., Chow, R.H., and Lester, H.A. (2009). Ezrin mediates tethering of the γ-aminobutyric acid transporter GAT1 to actin filaments via a C-terminal PDZ-interacting domain. Biophys. J. *96*: 2949–2960.

Jadey, S., and Auerbach, A. (2012). An integrated catch-and-hold mechanism activates nicotinic acetylcholine receptors. J. Gen. Physiol. *140*: 17–28.

Jancarik, J., and Kim, S.H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Crystallogr. *24*: 409–411.

Jayaraj, R., Megha, P., and Sreedev, P. (2016). Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. Interdiscip. Toxicol. *9*: 90–100.

Jayaraman, K., Das, A.K., Luethi, D., Szöllősi, D., Schütz, G.J., Reith, M.E.A., *et al.* (2021). SLC6 transporter oligomerization. J. Neurochem. *157*: 919–929.

Jette, N., Cappell, J., VanPassel, L., and Akman, C.I. (2006). Tiagabine-induced nonconvulsive status epilepticus in an adolescent without epilepsy. Neurology *67*: 1514–1515.

Jin, X.T., Galvan, A., Wichmann, T., and Smith, Y. (2011). Localization and function of GABA transporters GAT-1 and GAT-3 in the basal ganglia. Front. Syst. Neurosci. *5*: 1–10.

Jones, A.K. (2018). Genomics, cys-loop ligand-gated ion channels and new targets for the control of insect pests and vectors. Curr. Opin. Insect Sci.

Jones, A.K., Buckingham, S.D., Papadaki, M., Yokota, M., Sattelle, B.M., Matsuda, K., *et al.* (2009). Splice-variant- and stage-specific RNA editing of the *Drosophila* GABA receptor modulates agonist potency. J. Neurosci. *29*: 4287–4292.

Jones, A.K., and Sattelle, D.B. (2007). The cys-loop ligand-gated ion channel gene superfamily of the red flour beetle, *Tribolium castaneum*. BMC Genomics *8*: 327.

Jones, M. (2021). Characterisation of protein-ligand interactions of glycine binding protein (GBP) as a model of the heteromeric glycine receptor. University of Dundee.

Jorgenson, J.L. (2001). Aldrin and dieldrin: a review of research on their production, environmental deposition and fate, bioaccumulation, toxicology, and epidemiology in the United States. Environ. Health Perspect. *109*: 113–139.

Joseph, D., Pidathala, S., Mallela, A.K., and Penmatsa, A. (2019). Structure and Gating Dynamics of Na^+/Cl^- Coupled Neurotransmitter Transporters. Front. Mol. Biosci. *6*: 1–16.

Kabsch, W. (2010). XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*: 125–132.

Kalamida, D., Poulas, K., Avramopoulou, V., Fostieri, E., Lagoumintzis, G., Lazaridis, K., *et al.* (2007). Muscle and neuronal nicotinic acetylcholine receptors. FEBS J. *274*: 3799–3845.

Kang, S.U., Fuchs, K., Sieghart, W., and Lubec, G. (2008). Gel-Based Mass Spectrometric Analysis of recombinant $GABA_A$ receptor subunits representing strongly hydrophobic transmembrane proteins. J. Proteome Res. *7*: 3498–3506.

Kantonen, S.A., Henriksen, N.M., and Gilson, M.K. (2017). Evaluation and minimization of uncertainty in ITC binding measurements: heat error, concentration error, saturation, and stoichiometry. Biochim. Biophys. Acta - Gen. Subj. *1861*: 485–498.

Kasaragod, V.B., Mortensen, M., Hardwick, S.W., Wahid, A.A., Dorovykh, V., Chirgadze, D.Y., *et al.* (2022). Mechanisms of inhibition and activation of extrasynaptic αβ $GABA_A$ receptors. Nature *602*: 1–5.

Katoh, T., and Tiemeyer, M. (2013). The N's and O's of *Drosophila* glycoprotein glycobiology. Glycoconj. J. *30*: 57.

Kawate, T., and Gouaux, E. (2006). Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins.

Structure *14*: 673–681.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*: 845.

Kenakin, T.P. (2014). Orthosteric drug antagonism. In a pharmacology primer, (Elsevier), pp 119–154.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-Ray analysis. Nature *181*: 662–666.

Kermani, A.A. (2021). A guide to membrane protein X-ray crystallography. FEBS J. *288*: 5788–5804.

Kesters, D., Thompson, A.J., Brams, M., Elk, R. van, Spurny, R., Geitmann, M., *et al.* (2013). Structural basis of ligand recognition in 5-HT_3 receptors. EMBO Rep. *14*: 49–56.

Khan, K.H. (2013). Gene expression in mammalian cells and its applications. Adv. Pharm. Bull. *3*: 257.

Kikuchi, S., Bédard, J., and Nakai, M. (2011). One- and two-dimensional blue native-PAGE and Immunodetection of low-abundance chloroplast membrane protein complexes. In chloroplast research in arabidopsis. Methods in molecular biology (methods and protocols), R.P. Jarvis, ed. (Totowa, NJ: Humana Press), pp 3–17.

Kita, T., Ozoe, F., Azuma, M., and Ozoe, Y. (2013). Differential distribution of glutamate- and GABA-gated chloride channels in the housefly *Musca domestica*. J. Insect Physiol. *59*: 887–893.

Knight, A.R., Hartnett, C., Marks, C., Brown, M., Gallager, D., Tallman, J., *et al.* (1998). Molecular size of recombinant $\alpha_1\beta_1$ and $\alpha_1\beta_1\gamma_2$ $GABA_A$ receptors expressed in *Sf*9 cells. Receptors Channels *6*: 1–18.

Knight, A.R., Stephenson, F.A., Tallman, J.F., and Ramabahdran, T. V (2000). Monospecific antibodies as probes for the stoichiometry of recombinant $GABA_A$ receptors. Receptors Channels *7*: 213–226.

Korkhov, V.M., Farhan, H., Freissmuth, M., and Sitte, H.H. (2004). Oligomerization of the γ-aminobutyric acid transporter-1 is driven by an interplay of polar and hydrophobic interactions in transmembrane helix II. J. Biol. Chem. *279*: 55728–55736.

Kotov, V., Bartels, K., Veith, K., Josts, I., Subhramanyam, U.K.T., Günther, C., *et al.* (2019). High-throughput stability screening for detergent-solubilized membrane proteins. Sci. Rep. *9*: 10379.

Kouvatsos, N., Giastas, P., Chroni-Tzartou, D., Poulopoulou, C., and Tzartos, S.J. (2016). Crystal structure of a human neuronal nAChR extracellular domain in pentameric assembly: ligand-bound α_2 homopentamer. Proc. Natl. Acad. Sci. U. S. A. *113*: 9635–40.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. Mol. Biol. *196*: 947–950.

Kragholm, B., Kvist, T., Madsen, K.K., Jørgensen, L., Vogensen, S.B., Schousboe, A., *et al.* (2013). Discovery of a subtype selective inhibitor of the human betaine/GABA transporter 1 (BGT-1) with a non-competitive pharmacological profile. Biochem. Pharmacol. *86*: 521–528.

Kudryavtsev, D., Shelukhina, I., Vulfius, C., Makarieva, T., Stonik, V., Zhmak, M., *et al.* (2015). Natural compounds interacting with nicotinic acetylcholine receptors: from low-molecular weight ones to peptides and proteins. Toxins (Basel). *7*: 1683.

Kumar, A., Basak, S., Rao, S., Gicheru, Y., Mayer, M.L., Sansom, M.S.P., *et al.* (2020). Mechanisms of activation and desensitization of full-length glycine receptor in lipid nanodiscs. Nat. Commun. *11*: 3752.

Ladd, M., and Palmer, R. Structure determination by X-ray crystallography: analysis by X-rays and neutrons (Boston, MA: Springer US).

Lantez, V., Nikolaidis, I., Rechenmann, M., Vernet, T., and Noirclerc-Savoye, M. (2015). Rapid automated detergent screening for the solubilization and purification of membrane proteins and complexes. Eng. Life Sci. *15*: 39–50.

Laverty, D. (2016). Structural and functional studies of a chimeric $GABA_A$ receptor. University College London.

Laverty, D., Desai, R., Uchański, T., Masiulis, S., Stec, W.J., Malinauskas, T., *et al.* (2019). Cryo-EM structure of the human $\alpha_1\beta_3\gamma_2$ $GABA_A$ receptor in a lipid bilayer. Nature *565*: 516–520.

Laverty, D., Thomas, P., Field, M., Andersen, O.J., Gold, M.G., Biggin, P.C., *et al.* (2017). Crystal structures of a $GABA_A$-receptor chimera reveal new endogenous neurosteroid-binding sites. Nat. Struct. Mol. Biol. *24*: 977–985.

Lee, D.-F., Chen, C.-C., Hsu, T.-A., and Juang, J.-L. (2000). A baculovirus superinfection system: efficient vehicle for gene transfer into *Drosophila* S2 cells. J. Virol. *74*: 11873–11880.

Lee, H.-J., Rocheleau, T., Zhang, H.-G., Jackson, M.B., and Ffrench-Constant, R.H. (1993). Expression of a *Drosophila* GABA receptor in a baculovirus insect cell system. FEBS Lett. *335*: 315–318.

Lee, S.C., Knowles, T.J., Postis, V.L.G., Jamshad, M., Parslow, R.A., Lin, Y., *et al.* (2016). A method for detergent-free isolation of membrane proteins in their local lipid environment. Nat. Protoc. *11*: 1149–1162.

Lees, K., Musgaard, M., Suwanmanee, S., Buckingham, S.D., Biggin, P., and Sattelle, D. (2014). Actions of agonists, fipronil and ivermectin on the predominant in vivo splice and edit variant ($RDL_{bd,}$ I/V) of the *Drosophila* GABA receptor expressed in *Xenopus laevis* oocytes. PLoS One *9*: e97468.

Lees, R.S., Ambrose, P., Williams, J., Morgan, J., Praulins, G., Ingham, V.A., *et al.* (2020). Tenebenal: a meta-diamide with potential for use as a novel mode of action insecticide for public health. Malar. J. *19*: 398.

Legocki, R.P., and Verma, D.P.S. (1981). Multiple immunoreplica technique: Screening for specific proteins with a series of different antibodies using one polyacrylamide gel. Anal. Biochem. *111*: 385–392.

Lemaitre, R.P., Bogdanova, A., Borgonovo, B., Woodruff, J.B., and Drechsel, D.N. (2019). FlexiBAC: a versatile, open-source baculovirus vector system for protein expression, secretion, and proteolytic processing. BMC Biotechnol. *19*: 1–11.

Lemoine, D., Jiang, R., Taly, A., Chataigneau, T., Specht, A., and Grutter, T. (2012). Ligand-gated ion channels: new insights into neurological disorders and

ligand recognition. Chem. Rev. *112*: 6285–6318.

Li, S.-X., Huang, S., Bren, N., Noridomi, K., Dellisanti, C.D., Sine, S.M., *et al.* (2011). Ligand-binding domain of an α_7-nicotinic receptor chimera and its complex with agonist. Nat. Neurosci. *14*: 1253.

Li, X., Serwanski, D.R., Miralles, C.P., Bahr, B.A., and Blas, A.L. De (2007). Two pools of Triton X-100-insoluble $GABA_A$ receptors are present in the brain, one associated to lipid rafts and another one to the postsynaptic GABAergic complex. J. Neurochem. *102*: 1329.

Lichtenberg, D., Goñi, F.M., and Heerklotz, H. (2005). Detergent-resistant membranes should not be identified with membrane rafts. Trends Biochem. Sci. *30*: 430–436.

Liebschner, D., Afonine, P. V., Moriarty, N.W., Poon, B.K., Sobolev, O. V., Terwilliger, T.C., *et al.* (2017). Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallogr. Sect. D Struct. Biol. *73*: 148–157.

Lin, C.-H., and Jarvis, D.L. (2013). Utility of temporally distinct baculovirus promoters for constitutive and baculovirus-inducible transgene expression in transformed insect cells. J. Biotechnol. *165*: 11–17.

Liu, X., Krause, W.C., and Davis, R.L. (2007). $GABA_A$ receptor RDL inhibits *Drosophila* olfactory associative learning. Neuron *56*: 1090.

Lombard, J. (2014). Once upon a time the cell membranes: 175 years of cell boundary research. Biol. Direct *9*: 32.

Lombardo, S., and Maskos, U. (2015). Role of the nicotinic acetylcholine receptor in Alzheimer's disease pathology and treatment. Neuropharmacology *96*: 255–262.

Madsen, K., White, H.S., Clausen, R.P., Frolund, B., Larsson, O.M., Krogsgaard-Larsen, P., *et al.* (2007). Functional and pharmacological aspects of GABA transporters. Handb. Neurochem. Mol. Neurobiol. Neural Membr. Transp. 285–303.

Marheineke, K., Grünewald, S., Christie, W., and Reilander, H. (1998). Lipid composition of *Spodoptera frugiperda* (*Sf*9) and *Trichoplusia ni* (*Tn*) insect cells

used for baculovirus infection. FEBS Lett. *441*: 49–52.

Masiulis, S., Desai, R., Uchański, T., Serna Martin, I., Laverty, D., Karia, D., *et al.* (2019). $GABA_A$ receptor signalling mechanisms revealed by structural pharmacology. Nature *565*: 454–459.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*: 658–674.

McGonigle, I., and Lummis, S.C.R. (2010). Molecular characterization of agonists that bind to an insect GABA receptor. Biochemistry *49*: 2897–902.

McHugh, E.M., Zhu, W., Milgram, S., and Mager, S. (2004). The GABA transporter GAT1 and the MAGUK protein Pals1: Interaction, uptake modulation, and coexpression in the brain. Mol. Cell. Neurosci.

Mcmahon, C., Baier, A.S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J.X., *et al.* (2018). Yeast surface display platform for rapid discovery of conformationally selective nanobodies.

McPherson, A., and Gavira, J.A. (2014). Introduction to protein crystallization. Acta Crystallogr. Sect. F Struct. Biol. Commun. *70*: 2–20.

Ménard, C., Folacci, M., Brunello, L., Charreton, M., Collet, C., Mary, R., *et al.* (2018). Multiple combinations of RDL subunits diversify the repertoire of GABA receptors in the honey bee parasite *Varroa destructor*. J. Biol. Chem. *293*: 19012–19024.

Miglianico, M., Eldering, M., Slater, H., Ferguson, N., Ambrose, P., Lees, R.S., *et al.* (2018). Repurposing isoxazoline veterinary drugs for control of vector-borne human diseases. Proc. Natl. Acad. Sci. *115*: E6920–E6926.

Millar, N.S., Buckingham, S.D., and Sattelle, D.B. (1994). Stable expression of a functional homo-oligomeric *Drosophila* GABA receptor in a *Drosophila* cell line. Proc. R. Soc. London. Ser. B Biol. Sci. *258*: 307–314.

Miller, P.S., and Aricescu, A.R. (2014). Crystal structure of a human $GABA_A$ receptor. Nature *512*: 270–275.

Miller, P.S., Masiulis, S., Malinauskas, T., Kotecha, A., Rao, S., Chavali, S., *et al.*

(2018). Heteromeric $GABA_A$ receptor structures in positively-modulated active states. BioRxiv 338343.

Mitchell, L.S., and Colwell, L.J. (2018). Comparative analysis of nanobody sequence and structure data. Proteins Struct. Funct. Bioinforma. *86*: 697–706.

Morales-Perez, C.L., Noviello, C.M., and Hibbs, R.E. (2016). X-ray structure of the human $\alpha_4\beta_2$ nicotinic receptor. Nature *538*: 411–415.

Moroni, M., Zwart, R., Sher, E., Cassels, B.K., and Bermudez, I. (2006). $\alpha_4\beta_2$ nicotinic receptors with high and low acetylcholine sensitivity: pharmacology, stoichiometry, and sensitivity to long-term exposure to nicotine. Mol. Pharmacol. *70*: 755–768.

Moss, F.J., Imoukhuede, P.I., Scott, K., Hu, J., Jankowsky, J.L., Quick, M.W., *et al.* (2009). GABA transporter function, oligomerization state, and anchoring: correlates with subcellularly resolved FRET. J. Gen. Physiol. *134*: 489–521.

Motiwala, Z., Aduri, N.G., Shaye, H., Han, G.W., Lam, J.H., Katritch, V., *et al.* (2022). Structural basis of GABA reuptake inhibition. Nature *606*: 820–826.

Mu, L., Drandarov, K., Bisson, W.H., Schibig, A., Wirz, C., Schubiger, P.A., *et al.* (2006). Synthesis and binding studies of epibatidine analogues as ligands for the nicotinic acetylcholine receptors. Eur. J. Med. Chem. *41*: 640–650.

Mukherjee, S., Erramilli, S.K., Ammirati, M., Alvarez, F.J.D., Fennell, K.F., Purdy, M.D., *et al.* (2020). Synthetic antibodies against BRIL as universal fiducial marks for single−particle cryoEM structure determination of membrane proteins. Nat. Commun. 2020 111 *11*: 1–14.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., *et al.* (2011). REFMAC 5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*: 355–367.

Naffaa, M.M., Hung, S., Chebib, M., Johnston, G.A.R.R., and Hanrahan, J.R. (2017). GABA-ρ receptors: distinctive functions and molecular pharmacology. Br. J. Pharmacol. *174*: 1881–1894.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., *et al.* (2020). Single-particle cryo-EM at atomic resolution. Nature *587*: 152–156.

Nakao, T. (2015). Study on the structure and insecticide sensitivity of the RDL GABA receptor. J. Pestic. Sci. *40*: 152–159.

Nakao, T. (2017). Mechanisms of resistance to insecticides targeting RDL GABA receptors in planthoppers. Neurotoxicology *60*: 293–298.

Nakao, T., and Banba, S. (2016). Broflanilide: A meta-diamide insecticide with a novel mode of action. Bioorg. Med. Chem. *24*: 372–377.

Nakao, T., Hama, M., Kawahara, N., and Hirase, K. (2012). Fipronil resistance in Sogatella furcifera: Molecular cloning and functional expression of wild-type and mutant RDL GABA receptor subunits. J. Pestic. Sci. *37*: 37–44.

Navratna, V., and Gouaux, E. (2019). Insights into the mechanism and pharmacology of neurotransmitter sodium symporters. Curr. Opin. Struct. Biol. *54*: 161–170.

Newby, Z.E.R., O'Connell, J.D., Gruswitz, F., Hays, F. a, Harries, W.E.C., Harwood, I.M., *et al.* (2009). A general protocol for the crystallization of membrane proteins for X-ray structural investigation. Nat. Protoc. *4*: 619–637.

Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. Nat. Methods *13*: 24–27.

Nothdurfter, C., Tanasic, S., Benedetto, B. Di, Rammes, G., Wagner, E.-M., Kirmeier, T., *et al.* (2010). Impact of lipid raft integrity on 5-HT_3 receptor function and its modulation by antidepressants. Neuropsychopharmacology *35*: 1510–1519.

Novak, V., Kanard, R., Kissel, J.T., and Mendell, J.R. (2001). Treatment of painful sensory neuropathy with tiagabine: a pilot study. Clin. Auton. Res. 2001 116 *11*: 357–361.

Noviello, C.M., Gharpure, A., Mukhtasimova, N., Cabuco, R., Baxter, L., Borek, D., *et al.* (2021). Structure and gating mechanism of the α_7 nicotinic acetylcholine receptor. Cell 1–14.

Nys, M., Kesters, D., and Ulens, C. (2013). Structural insights into Cys-loop receptor function and ligand recognition. Biochem. Pharmacol. *86*: 1042–1053.

O'Toole, K.K., and Jenkins, A. (2011). Discrete M3-M4 intracellular loop

subdomains control specific aspects of γ-aminobutyric acid type A receptor function. J. Biol. Chem. *286*: 37990–37999.

Ochs, D. (1983). Protein contaminants of sodium dodecyl sulfate-polyacrylamide gels. Anal. Biochem. *135*: 470–474.

Olerinyova, A., Sonn-Segev, A., Gault, J., Eichmann, C., Schimpf, J., Kopf, A.H., *et al.* (2021). Mass photometry of membrane proteins. Chem *7*: 224–236.

Olsen, R.W., and Sieghart, W. (2009). GABAA receptors: Subtypes provide diversity of function and pharmacology. Neuropharmacology *56*: 141–148.

Ondachi, P.W., Castro, A.H., Luetje, C.W., Wageman, C.R., Marks, M.J., Damaj, M.I., *et al.* (2016). Synthesis, nicotinic acetylcholine binding, and in vitro and in vivo pharmacological properties of 2′-fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues. ACS Chem. Neurosci. *7*: 1004–1012.

Papworth, C., Greener, A., Braman, J., and Wright, D.. (1996). Site-directed mutagenesis in one day with >80% efficiency. Strategies *9*: 3–4.

Penmatsa, A., Wang, K.H., and Gouaux, E. (2013). X-ray structure of dopamine transporter elucidates antidepressant mechanism. Nature *503*: 85–90.

Penmatsa, A., Wang, K.H., and Gouaux, E. (2015). X-ray structures of *Drosophila* dopamine transporter in complex with nisoxetine and reboxetine. Nat. Struct. Mol. Biol. *22*: 506–508.

Phulera, S., Zhu, H., Yu, J., Claxton, D.P., Yoder, N., Yoshioka, C., *et al.* (2018). Cryo-EM structure of the benzodiazepine-sensitive $\alpha_1\beta_1\gamma_{2S}$ tri-heteromeric $GABA_A$ receptor in complex with GABA. Elife *7*: 163–170.

Platt, N., Kwiatkowska, R.M., Irving, H., Diabaté, A., Dabire, R., and Wondji, C.S. (2015). Target-site resistance mutations (kdr and RDL), but not metabolic resistance, negatively impact male mating competiveness in the malaria vector *Anopheles gambiae*. Heredity (Edinb). *115*: 243–252.

Pless, S.A., and Sivilotti, L.G. (2018). A tale of ligands big and small: an update on how pentameric ligand-gated ion channels interact with agonists and proteins. Curr. Opin. Physiol. *2*: 19–26.

Pollock, N.L., Lee, S.C., Patel, J.H., Gulamhussein, A.A., and Rothnie, A.J. (2018). Structure and function of membrane proteins encapsulated in a polymer-bound lipid bilayer. Biochim. Biophys. Acta - Biomembr. *1860*: 809–817.

Pramod, A.B., Foster, J., Carvelli, L., and Henry, L.K. (2013). SLC6 transporters: Structure, function, regulation, disease association and therapeutics. Mol. Aspects Med. *34*: 197–219.

Privé, G.G. (2007). Detergents for the stabilization and crystallization of membrane proteins. Methods *41*: 388–397.

Prochaska, J.J., Das, S., and Benowitz, N.L. (2013). Cytisine, the world's oldest smoking cessation aid. BMJ *347*:.

Purves, D., and Williams, S.M. (2001). Two families of postsynaptic receptors. In neuroscience, (Sinauer Associates 2001).

Quick, M.W. (2002). Substrates regulate γ-aminobutyric acid transporters in a syntaxin 1A-dependent manner. Proc. Natl. Acad. Sci. *99*: 5686–5691.

Radian, R., Bendahan, A., and Kanner, B.I. (1986). Purification and identification of the functional sodium- and chloride-coupled γ-aminobutyric acid transport glycoprotein from rat brain. J. Biol. Chem. *261*: 15437–15441.

Rajarathnam, K., and Rösgen, J. (2014). Isothermal titration calorimetry of membrane proteins - progress and challenges. Biochim. Biophys. Acta - Biomembr. *1838*: 69–77.

Rang, H.P., Dale, M.M., Ritter, J., Flower, R., and Henderson, G. (2011). Rang and Dale's Pharmacology (Edinburgh; New York: Elsevier/Churchill Livingstone).

Rath, A., and Deber, C.M. (2013). Correction factors for membrane protein molecular weight readouts on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Anal. Biochem. *434*: 67–72.

Rayes, D., Rosa, M.J. De, Sine, S.M., and Bouzat, C. (2009). Number and locations of agonist binding sites required to activate homomeric Cys-loop receptors. J. Neurosci. *29*: 6022–6032.

Rego Campello, H., Villar, S.G. Del, Honraedt, A., Minguez, T., Oliveira, A.S.F., Ranaghan, K.E., *et al.* (2018). Unlocking Nicotinic selectivity via direct C–H

functionalization of (−)-Cytisine. Chem *4*: 1710–1725.

Reitsma, M.B., Fullman, N., Ng, M., Salama, J.S., Abajobir, A., Abate, K.H., *et al.* (2017). Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: A systematic analysis from the global burden of disease study 2015. Lancet *389*: 1885–1906.

Remnant, E.J., Good, R.T., Schmidt, J.M., Lumb, C., Robin, C., Daborn, P.J., *et al.* (2013). Gene duplication in the major insecticide target site, Rdl , in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. *110*: 14705–14710.

Remnant, E.J., Morton, C.J., Daborn, P.J., Lumb, C., Yang, Y.T., Ng, H.L., *et al.* (2014). The role of Rdl in resistance to phenylpyrazoles in *Drosophila melanogaster*. Insect Biochem. Mol. Biol. *54*: 11–21.

Rietveld, A., Neutz, S., Simons, K., and Eaton, S. (1999). Association of sterol- and glycosylphosphatidylinositol-linked proteins with *Drosophila* raft lipid microdomains. J. Biol. Chem. *274*: 12049–12054.

Ritter, L., Solomon, K.R., Forget, J., Stemeroff, M., and O 'leary, C. (1995). A review of selected persistent organic pollutants DDT, aldrin, dieldrin, endrin, chlordane heptachlor, hexachlorobenzene, mirex, toxaphene, polychlorinated biphenyls, dioxins and furans.

Rosa, N., Ristic, M., Thorburn, L., Abrahams, G.J., Marshall, B., Watkins, C.J., *et al.* (2020). Tools to ease the choice and design of protein crystallisation experiments. Cryst. *10*: 1–19.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. Front. Microbiol. *5*: 172.

Rowlett, J.K., Platt, D.M., Lelas, S., Atack, J.R., and Dawson, G.R. (2005). Different $GABA_A$ receptor subtypes mediate the anxiolytic, abuse-related, and motor effects of benzodiazepine-like drugs in primates. Proc. Natl. Acad. Sci. *102*: 915–920.

Rucktooa, P., Haseler, C.A., Elk, R. van, Smit, A.B., Gallagher, T., and Sixma, T.K. (2012). Structural characterization of binding mode of smoking cessation drugs to nicotinic acetylcholine receptors through study of ligand complexes with acetylcholine-binding protein. J. Biol. Chem. *287*: 23283–23293.

Rucktooa, P., Smit, A.B., and Sixma, T.K. (2009). Insight in nAChR subtype selectivity from AChBP crystal structures. Biochem. Pharmacol. *78*: 777–787.

Rudell, J.C., Borges, L.S., Yarov-Yarovoy, V., and Ferns, M. (2020). The MX-helix of muscle nAChR subunits regulates receptor assembly and surface trafficking. Front. Mol. Neurosci. *13*: 48.

Saboury, A.A. (2014). A review on the ligand binding studies by isothermal titration calorimetry. J. Iran. Chem. Soc. *3*: 1–21.

Saliba, R.S., Munro, P.M.G., Luthert, P.J., and Cheetham, M.E. (2002). The cellular fate of mutant rhodopsin: quality control, degradation and aggresome formation. J. Cell Sci. *115*: 2907–2918.

Schärfer, C., Schulz-Gasch, T., Ehrlich, H.-C., Guba, W., Rarey, M., and Stahl, M. (2013). Torsion angle preferences in druglike chemical space: a comprehensive guide. J. Med. Chem. *56*: 2016–2028.

Schlessinger, A., Yee, S.W., Sali, A., and Giacomini, K.M. (2013). SLC classification: an update. Clin. Pharmacol. Ther. *94*: 19–23.

Schlimgen, R., Howard, J., Wooley, D., Thompson, M., Baden, L.R., Yang, O.O., *et al.* (2016). Risks associated with lentiviral vector exposures and prevention strategies. J. Occup. Environ. Med. *58*: 1159–1166.

Schmid, J.A., Scholze, P., Kudlacek, O., Freissmuth, M., Singer, E.A., and Sitte, H.H. (2001). Oligomerization of the human serotonin transporter and of the rat GABA transporter 1 visualized by fluorescence resonance energy transfer microscopy in living cells. J. Biol. Chem. *276*: 3805–3810.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*: 671–675.

Scholze, P., Freissmuth, M., and Sitte, H.H. (2002). Mutations within an intramembrane leucine heptad repeat disrupt oligomer formation of the rat GABA transporter 1. J. Biol. Chem. *277*: 43682–43690.

Schuler, M.A., Denisov, I.G., and Sligar, S.G. (2013). Nanodiscs as a new tool to examine lipid-protein interactions. Methods Mol. Biol. *974*: 415.

Scott, H., and Panin, V.M. (2014). The role of protein N-glycosylation in neural

transmission. Glycobiology *24*: 407–417.

Seddon, A.M., Curnow, P., and Booth, P.J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. Biochim. Biophys. Acta - Biomembr. *1666*: 105–117.

Shahsavar, A., Ahring, P.K., Olsen, J.A., Krintel, C., Kastrup, J.S., Balle, T., *et al.* (2015). Acetylcholine-binding protein engineered to mimic the a_4-a_4 binding pocket in a_4b_2 nicotinic acetylcholine receptors reveals interface specific interactions important for binding and activity. Mol. Pharmacol. Mol Pharmacol *88*: 697–707.

Shahsavar, A., Gajhede, M., Kastrup, J.S., and Balle, T. (2016). Structural studies of nicotinic acetylcholine receptors: using acetylcholine-binding protein as a structural surrogate. Basic Clin. Pharmacol. Toxicol. *118*: 399–407.

Shahsavar, A., Kastrup, J.S., Nielsen, E.Ø., Kristensen, J.L., Gajhede, M., and Balle, T. (2012). Crystal structure of *Lymnaea stagnalis* AChBP complexed with the potent nAChR antagonist DHβE suggests a unique mode of antagonism. PLoS One *7*: e40757.

Shahsavar, A., Stohler, P., Bourenkov, G., Zimmermann, I., Siegrist, M., Guba, W., *et al.* (2021). Structural insights into the inhibition of glycine reuptake. Nature *591*: 677–681.

Shi, X., and Jarvis, D. (2007). Protein N-glycosylation in the baculovirus-insect cell system. Curr. Drug Targets *8*: 1116–1125.

Shoemaker, S.C., and Ando, N. (2018). X-rays in the cryo-electron microscopy era: structural biology's dynamic future. Biochemistry *57*: 277–285.

Shotkoski, F., Zhang, H.-G., Jackson, M.B., and Ffrench-Constant, R.H. (1996). Stable expression of insect GABA receptors in insect cell lines: promoters for efficient expression of *Drosophila* and mosquito Rdl GABA receptors in stably transformed mosquito cell lines. FEBS Lett. *380*: 257–262.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Mol. Syst. Biol. *7*: 539.

Singh, S.K., and Sigworth, F.J. (2015). Cryo-EM: Spinning the Micelles Away. Structure *23*: 1561.

Skarzynski, T. (2013). Collecting data in the home laboratory: evolution of X-ray sources, detectors and working practices. Acta Crystallogr. Sect. D Biol. Crystallogr. *69*: 1283–1288.

Smit, A.B., Brejc, K., Syed, N., and Sixma, T.K. (2003). Structure and function of AChBP, homologue of the ligand-binding domain of the nicotinic acetylcholine receptor. Ann. N. Y. Acad. Sci. *998*: 81–92.

Smith, S.M. (2017). Strategies for the purification of membrane proteins. In methods in molecular biology, pp 389–400.

Smyth, M.S. (2000). X-Ray crystallography. Mol. Pathol. *53*: 8–14.

Song, C., Zhang, S., and Huang, H. (2015). Choosing a suitable method for the identification of replication origins in microbial genomes. Front. Microbiol. *6*: 1-18.

Sonnino, S., and Prinetti, A. (2010). Lipids and membrane lateral organization. Front. Physiol. *1*: 153.

Sooksawate, T., and Simmonds, M.A. (2001). Effects of membrane cholesterol on the sensitivity of the $GABA_A$ receptor to GABA in acutely dissociated rat hippocampal neurones. Neuropharmacology *40*: 178–184.

Soragna, A., Bossi, E., Giovannardi, S., Pisani, R., and Peres, A. (2005). Functionally independent subunits in the oligomeric structure of the GABA cotransporter rGAT1. Cell. Mol. Life Sci. *62*: 2877–2885.

Spande, T.F., Garraffo, H.M., Edwards, M.W., Yeh, H.J.C., Pannell, L., and Daly, J.W. (1992). Epibatidine: a novel (chloropyridyl)azabicycloheptane with potent analgesic activity from an Ecuadoran poison frog. J. Am. Chem. Soc. *114*: 3475–3478.

Spang, J.E., Bertrand, S., Westera, G., Patt, J.T., Schubiger, P.A., and Bertrand, D. (2000). Chemical modification of epibatidine causes a switch from agonist to antagonist and modifies its selectivity for neuronal nicotinic acetylcholine receptors. Chem. Biol. *7*: 545–555.

Sternberg, E.D., and Thomas, M.B. (2018). Insights from agriculture for the

management of insecticide resistance in disease vectors. Evol. Appl. *11*: 404–414.

Stetsenko, A., and Guskov, A. (2017). An overview of the top ten detergents used for membrane protein crystallization. Crystals *7*: 197.

Steyaert, J., and Kobilka, B.K. (2011). Nanobody stabilization of G protein-coupled receptor conformational states. Curr. Opin. Struct. Biol. *21*: 567–572.

Stillwell, W. (2016a). Membrane-associated processes. In an introduction to biological membranes, (Elsevier), pp 381–421.

Stillwell, W. (2016b). Membrane transport. In an introduction to biological membranes, (Elsevier), pp 423–451.

Stilwell, G.E., and Ffrench-Constant, R.H. (1998). Transcriptional analysis of the *Drosophila* GABA receptor gene resistance to dieldrin. J. Neurobiol. *36*: 468–484.

Strop, P., and Brunger, A.T. (2005). Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. Protein Sci. *14*: 2207–2211.

Taly, A., Corringer, P.-J., Guedin, D., Lestage, P., and Changeux, J.-P. (2009). Nicotinic receptors: allosteric transitions and therapeutic targets in the nervous system. Nat. Rev. Drug Discov. *8*: 733–750.

Tate, C.G., Haase, J., Baker, C., Boorsma, M., Magnani, F., Vallis, Y., *et al.* (2003). Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. Biochim. Biophys. Acta - Biomembr. *1610*: 141–153.

Tate, C.G., Whiteley, E., and Betenbaugh, M.J. (1999). Molecular chaperones stimulate the functional expression of the cocaine-sensitive serotonin transporter. J. Biol. Chem. *274*: 17551–17558.

Taylor-Wells, J., Hawkins, J., Colombo, C., Bermudez, I., and Jones, A.K. (2017). Cloning and functional expression of intracellular loop variants of the honey bee (*Apis mellifera*) RDL GABA receptor. Neurotoxicology *60*: 207–213.

Taylor-Wells, J., and Jones, A.K. (2017). Variations in the insect GABA receptor, RDL, and their Impact on receptor pharmacology. In advances in agrochemicals:

ion channels and g protein-coupled receptors (GPCRs) as targets for pest control, A.D. Gross, Y. Ozoe, and J.R. Coats, eds. (Washington, DC: American Chemical Society), pp 1–21.

Taylor, K.A., and Glaeser, R.M. (1974). Electron diffraction of frozen, hydrated protein crystals. Science (80-.). *186*: 1036–1037.

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., *et al.* (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 2022 1–3.

Thomas, J., and Tate, C.G. (2014). Quality control in eukaryotic membrane protein overproduction. J. Mol. Biol. *426*: 4139–4154.

Thompson, A.J., Metzger, S., Lochner, M., and Ruepp, M.-D. (2017). The binding orientation of epibatidine at α_7 nACh receptors. Neuropharmacology *116*: 421–428.

Thompson, M.J., and Baenziger, J.E. (2020). Structural basis for the modulation of pentameric ligand-gated ion channel function by lipids. Biochim. Biophys. Acta - Biomembr. *1862*: 183304.

Thonghin, N., Kargas, V., Clews, J., and Ford, R.C. (2018). Cryo-electron microscopy of membrane proteins. Methods *147*: 176–186.

Trattnig, S.M., Gasiorek, A., Deeb, T.Z., Ortiz, E.J.C., Moss, S.J., Jensen, A.A., *et al.* (2016). Copper and protons directly activate the zinc-activated channel. Biochem. Pharmacol. *103*: 109–117.

Traynor, J.R. (1998). Epibatidine and pain. Br. J. Anaesth. *81*: 69–76.

Trumper, P. (2014). Development of a high throughput ligand screening method and structural studies of pentameric ligand gated ion channels. University of Dundee.

Tsai, J.-Y., Chu, C.-H., Lin, M.-G., Chou, Y.-H., Hong, R.-Y., Yen, C.-Y., *et al.* (2020). Structure of the sodium-dependent phosphate transporter reveals insights into human solute carrier SLC20. Sci. Adv. *6*: eabb4024.

Uchański, T., Zögg, T., Yin, J., Yuan, D., Wohlkönig, A., Fischer, B., *et al.* (2019). An improved yeast surface display platform for the screening of nanobody

immune libraries. Sci. Reports *9*: 1–12.

UN Environment Programme (2002). Stockholm convention on persistent organic pollutants (POPs).

Unwin, N. (2005). Refined structure of the nicotinic acetylcholine receptor at 4Å resolution. J. Mol. Biol. *346*: 967–989.

Uthman, B.M., Rowan, A.J., Ahmann, P.A., Leppik, I.E., Schachter, S.C., Sommerville, K.W., *et al.* (1998). Tiagabine for complex partial seizures: a randomized, add-on, dose-response trial. Arch. Neurol. *55*: 56–62.

Vénien-Bryan, C., Li, Z., Vuillard, L., and Boutin, J.A. (2017). Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. Acta Crystallogr. Sect. F Struct. Biol. Commun. *73*: 174–183.

Villar, S. del, and Jones, A. (2018). Cloning and functional characterisation of the duplicated RDL subunits from the pea aphid, *Acyrthosiphon pisum*. Int. J. Mol. Sci. *19*: 2235.

Wagner, S., Bader, M.L., Drew, D., and Gier, J.-W. de (2006). Rationalizing membrane protein overexpression. Trends Biotechnol. *24*: 364–371.

Walsh, R.M., Roh, S., Gharpure, A., Morales-Perez, C.L., Teng, J., and Hibbs, R.E. (2018). Structural principles of distinct assemblies of the human $\alpha_4\beta_2$ nicotinic receptor. Nature *557*: 261–265.

Wang, D., Deken, S.L., Whitworth, T.L., and Quick, M.W. (2003). Syntaxin 1A inhibits GABA flux, efflux, and exchange mediated by the rat brain GABA transporter GAT1. Mol. Pharmacol. *64*: 905–913.

Wang, H.W., and Wang, J.W. (2017). How cryo-electron microscopy and X-ray crystallography complement each other. Protein Sci. *26*: 32–39.

Wang, J., and Lindstrom, J. (2018). Orthosteric and allosteric potentiation of heteromeric neuronal nicotinic acetylcholine receptors. Br. J. Pharmacol. *175*: 1805–1821.

Wang, Z.-J., Deba, F., Mohamed, T.S., Chiara, D.C., Ramos, K., and Hamouda, A.K. (2017). Unraveling amino acid residues critical for allosteric potentiation of

(α_4)3(β_2)2-type nicotinic acetylcholine receptor responses. J. Biol. Chem. *292*: 9988–10001.

Warschawski, D.E., Arnold, A.A., Beaugrand, M., Gravel, A., Chartrand, É., and Marcotte, I. (2011). Choosing membrane mimetics for NMR structural studies of transmembrane proteins. Biochim. Biophys. Acta - Biomembr. *1808*: 1957–1974.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*: W296–W303.

Weinglass, A., Kaczorowski, G., and Garcia, M. (2008). Technologies for transporter drug discovery. Channels *2*: 312–321.

Wilkerson, J.L., Deba, F., Crowley, M.L., Hamouda, A.K., and McMahon, L.R. (2020). Advances in the *in vitro* and in vivo pharmacology of $\alpha_4\beta_2$ Nicotinic receptor positive allosteric modulators. Neuropharmacology *168*: 108008.

Wing, K.D. (2020). Pharmaceutical technologies with potential application to insecticide discovery. Pest Manag. Sci. ps.6075.

Winter, G., Waterman, D.G., Parkhurst, J.M., Brewster, A.S., Gildea, R.J., Gerstel, M., *et al.* (2018). DIALS: implementation and evaluation of a new integration package. Acta Crystallogr. Sect. D Struct. Biol. *74*: 85–97.

Wittig, I., Karas, M., and Schägger, H. (2007). High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. Mol. Cell. Proteomics *6*:.

Wondji, C.S., Dabire, R.K., Tukur, Z., Irving, H., Djouaka, R., and Morgan, J.C. (2011). Identification and distribution of a GABA receptor mutation conferring dieldrin resistance in the malaria vector *Anopheles funestus* in Africa. Insect Biochem. Mol. Biol. *41*: 484–491.

Wooltorton, J.R.A., Moss, S.J., and Smart, T.G. (1997). Pharmacological and physiological characterization of murine homomeric β_3 gaba a receptors. Eur. J. Neurosci. *9*: 2225–2235.

World Health Organisation (2018). Global report on insecticide resistance in malaria vectors: 2010–2016.

Wu, J., and Lukas, R.J. (2011). Naturally-expressed nicotinic acetylcholine receptor subtypes. Biochem. Pharmacol. *82*: 800–807.

Wu, S., Armache, J.-P., and Cheng, Y. (2016). Single-particle cryo-EM data acquisition by using direct electron detection camera. Microscopy *65*: 35–41.

Wu, S., Avila-Sakar, A., Kim, J., Booth, D.S., Greenberg, C.H., Rossi, A., *et al.* (2012). Fabs enable single particle cryoEM studies of small proteins. Structure *20*: 582–592.

Wu, X., and Rapoport, T.A. (2021). Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (legobodies). Proc. Natl. Acad. Sci. *118*: e2115001118.

Xu, J., Sun, H., Huang, G., Liu, G., Li, Z., Yang, H., *et al.* (2019). A fixation method for the optimisation of western blotting. Sci. Rep. *9*: 6649.

Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., and Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. Nature *437*: 215–223.

Yeliseev, A., Zoubak, L., and Schmidt, T.G.M. (2017). Application of strep-tactin xt for affinity purification of twin-strep-tagged CB2, a G protein-coupled cannabinoid receptor. Protein Expr. Purif. *131*: 109.

Yip, K.M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. Nature *587*: 157–161.

Young, G., Hundt, N., Cole, D., Fineberg, A., Andrecka, J., Tyler, A., *et al.* (2017). Quantitative mass imaging of single molecules in solution. Science (80-.). *427*: 423–427.

Zacharias, D.A., Violin, J.D., Newton, A.C., and Tsien, R.Y. (2002). Partitioning of lipid-modified monomeric gfps into membrane microdomains of live cells. Science. *296*: 913–916.

Zaluski, R., Kadri, S.M., Alonso, D.P., Martins Ribolla, P.E., and Oliveira Orsi, R. de (2015). Fipronil promotes motor and behavioral changes in honey bees (*Apis mellifera*) and affects the development of colonies exposed to sublethal doses. Environ. Toxicol. Chem. *34*: 1062–9.

Zare, M., Soleimani-Ahmadi, M., Davoodi, S.H., and Sanei-Dehkordi, A. (2016). Insecticide susceptibility of *Anopheles stephensi* to DDT and current insecticides in an elimination area in Iran. Parasit. Vectors *9*: 571.

Zhang, H.G., Lee, H.J., Rocheleau, T., Ffrench-Constant, R.H., and Jackson, M.B. (1995). Subunit composition determines picrotoxin and bicuculline sensitivity of *Drosophila* γ-aminobutyric acid receptors. Mol. Pharmacol. *48*: 835–40.

Zhang, X., and Miller, K.W. (2015). Dodecyl maltopyranoside enabled purification of active human GABA type A receptors for deep and direct proteomic sequencing. Mol. Cell. Proteomics *14*: 724–38.

Zhao, Y., Liu, S., Zhou, Y., Zhang, M., Chen, H., Eric Xu, H., *et al.* (2021). Structural basis of human α_7 nicotinic acetylcholine receptor activation. Cell Res. *31*: 713–716.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 2017 144 *14*: 331–332.

Zhong, W., Gallivan, J.P., Zhang, Y., Li, L., Lester, H.A., and Dougherty, D.A. (1998). From ab initio quantum mechanics to molecular neurobiology: a cation-binding site in the nicotinic receptor. Proc. Natl. Acad. Sci. *95*: 12088–12093.

Zhou, R. Bin, Lu, H.M., Liu, J., Shi, J.Y., Zhu, J., Lu, Q.Q., *et al.* (2016). A Systematic analysis of the structures of heterologously expressed proteins and those from their native hosts in the RCSB PDB archive. PLoS One *11*: e0161254.

Zhou, Y., and Danbolt, N.C. (2013). GABA and glutamate transporters in brain. Front. Endocrinol. (Lausanne). *4*: 165.

Zhu, H., and Gouaux, E. (2021). Architecture and assembly mechanism of native glycine receptors. Nat. | *599*: 513.

Zhu, J., and Reith, M.E.A. (2008). Role of the dopamine transporter in the action of psychostimulants, nicotine, and other drugs of abuse. CNS Neurol. Disord. - Drug Targets *7*: 393–409.

Zhu, S., Noviello, C.M., Teng, J., Walsh, R.M., Kim, J.J., and Hibbs, R.E. (2018).

Structure of a human synaptic GABAA receptor. Nature *559*: 67–72.

Appendices

S1: Extraction and Purification of the Human GABA Transporter 1 with Styrene and Malic Acid Co-polymer

S1.1 Overview

In addition to extracting and purifying GAT1 using a conventional detergent-based approach, some initial work was also conducted using SMA co-polymer to extract GAT1-mEm$_{(PFB)}$ along with the native lipids present in *Sf*9 cell membranes. This work ultimately did not go beyond initial purifications due to needing to concentrate time and resources on the detergent extraction method, but some interesting preliminary results were obtained.

S1.2 Results

SMA extraction screens were run on an analogous basis to detergent screening. The cells were cultured normally and infected with the GAT1-mEm$_{(PFB)}$ baculovirus, but after extraction of the membranes one of three SMA co-polymers was used in place of detergent. These were respectively SL30010P SMA co-polymer (referred to as SL30 from here), SL25001P (SL25) and SL40005P (SL40). These were kind gifts from Polyscope Polymer BV and the co-polymers have since been commercialised for protein extraction by the spin off company Orbiscope. The polymers differed in the ratios of styrene and maleic acid groups, respectfully being 3:1 for SL25, 2.3:1 for SL30 and 1.4:1 for SL40.

SMA co-polymer was added to membranes at 1.5 mg/ml to a 2.5% final concentration. For each co-polymer two parallel samples were setup: one was gently shaken at room temperature for 2 hrs, while the other was gently mixed at 4 °C overnight. Insoluble material was then removed by ultra-centrifugation. While detergent extraction typically requires low temperature to minimise proteolysis, room temperature extraction is recommended with SMA due to the increased membrane fluidity enhancing excision of the protein from the lipid environment; lower temperatures simply slow this process (Lee *et al.*, 2016). The samples were analysed by the fluorescence:[protein] ratio and SDS-PAGE in-gel

fluorescence. FSEC was attempted with these samples, but insufficient signal strength was present to obtain any usable data.

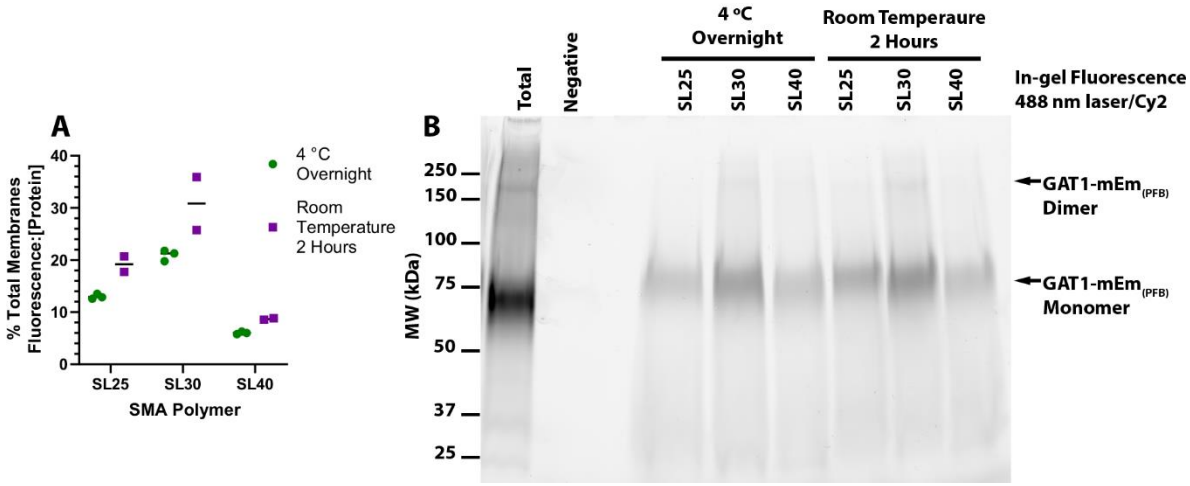


Figure S1.1: SMA extraction screening of GAT1-mEm$_{(PFB)}$.

*Sf*9 membranes containing GAT1-mEm$_{(PFB)}$ were incubated at 1.5 mg/ml with the indicated SMA polymers at 2.5% made up in buffer C for 2 hours at room temperature or over-night at 4 °C with gentle mixing before insoluble material was removed. The total had buffer added in place of SMA and was not ultra-centrifuged. **A:** Bar chart of the samples average fluorescence:[protein] ratios expressed as a percentage of that for the total. Reads were taken in triplicate for the 4 °C samples and in duplicate for the room temperature samples. Each point represents an individual read and the mean is shown as a horizontal bar. Where no bar is visible then it is within the size of the symbols. **B:** SDS-PAGE in-gel fluorescence of the samples. ~15 µg of protein was loaded per lane. The negative control consisted of un-infected *Sf*9 cells lysed by needle shearing. The gel was excited with a 488 nm laser and imaged with a Cy2 filter.

The fluorescence:[protein] ratio measurements showed that less appeared to be extracted at 4 °C overnight compared to room temperature for 2 hours (Figure S1.1A). SL30 was more efficient than SL25 and SL40 SMA co-polymers at both temperatures. Although more variable than the other polymers, at room temperature SL25 extracted ~30% of the total membranes fluorescence:[protein] ratio, similar to that of 1% digitonin detergent previously tested. Running the samples on SDS-PAGE with in-gel fluorescence (Figure S1.1B) appeared to corroborate the total fluorescence measurements and little or no free mEmerald was visible for any of the co-polymers, but postulated dimers were also seen as with detergents. Therefore, it was suggested that SMA co-polymers represent a viable method for extraction of GAT1-mEm in a native membrane environment.

Taking advantage of having excess membranes leftover from the initial larger scale 0.8L GAT1-mEm detergent purification (see Chapter 5 section 5.3.1.2), a small scale (~100 ml culture) extraction and purification was performed using SL30 SMA at room temperature for 2 hrs. After removal of insoluble material the

soluble fraction was bound to 200 µl of Streptactin-XT resin for 1 hr and collected in a gravity flow column. Except for omitting any detergents, the purification procedure was identical to that of the detergent solubilised preparation, with the resin washed with buffer C and the proteins eluted once using buffer C at pH 8 with a lowered 100 mM NaCl content and containing 50 mM *d*-biotin. After elution the protein was concentrated with a 30 kDa centrifugal concentrator (used so that empty ~6.5 kDa SL30 co-polymer would pass through) down to ~100 µl. The concentration was measured using a BCA assay, as SMA absorbs strongly at 280 nm. This suggested that there was ~32 µg of total protein present in the 100 µl.

Loading samples on an SDS-PAGE gel (Figure S1.2A, B) indicated that GAT1-mEm had been extracted from the membranes as it was visible in every fraction on both stain-free total protein imaging and in-gel fluorescence. However, it was also indicated that other proteins had co-eluted off the resin. The signal strength for the elute lane was weak with the stain free imaging, so it was attempted to colloidal Coomassie stain the gel for improved sensitivity. The concentrated elute did not appear to take up the stain though. It was attempted to analyse the concentrated elute on the Nanotemper Tycho (Figure S1.2C), but it was unable to fit any T_i values. The un-folding curve did show some similarities however to previous ones obtained with detergent purified protein, including the rapid drop in the first derivative ratio at ~90 °C.

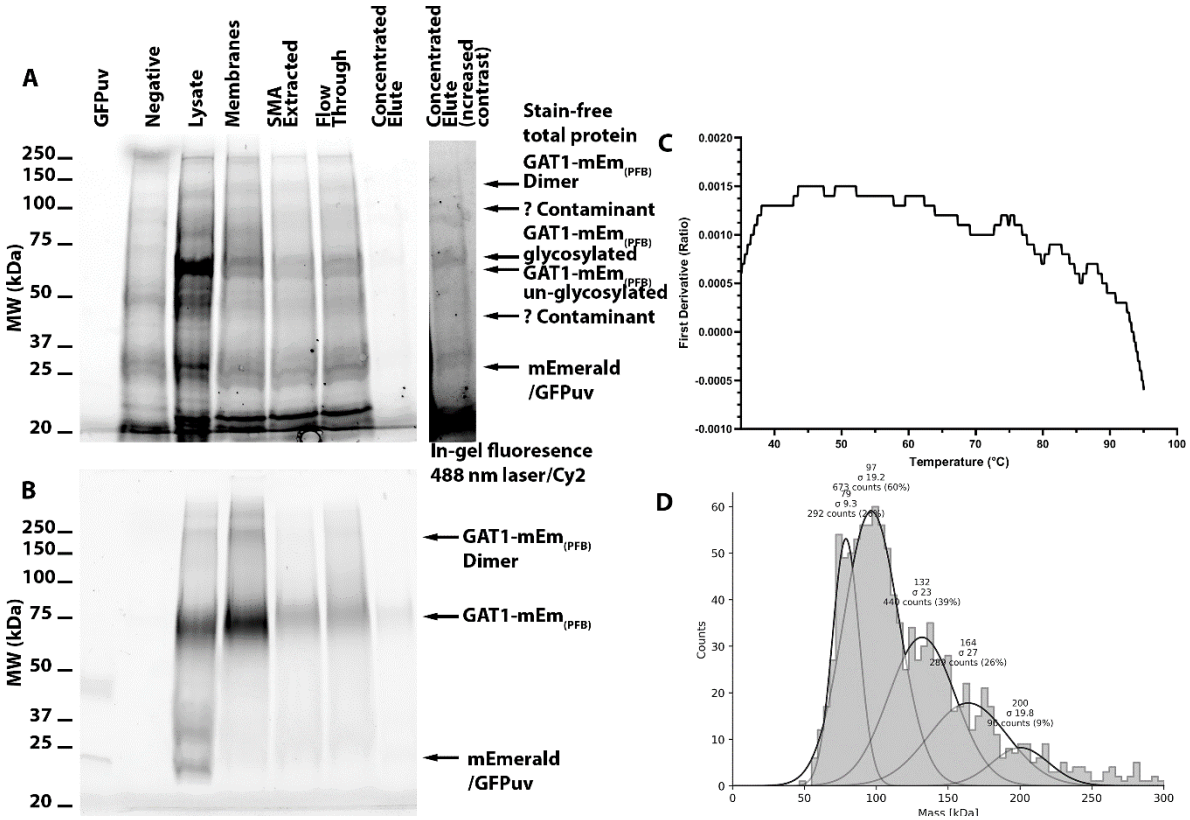


Figure S1.2: Purification of GAT1-mEm$_{(PFB)}$ extracted in SL30 SMA.

Membranes were extracted from *Sf*9 cells infected with the GAT1-mEm$_{(PFB)}$ baculovirus and proteins in their native lipid environment extracted with 2.5% SL30 SMA for 2 hours at room temperature with gentle shaking before insoluble material was removed. The protein was then purified with streptactin-XT resin **A-B:** SDS-PAGE analysis of samples taken throughout the purification cycle. The negative sample consisted of un-infected *Sf*9 cells lysed by needle shearing. 20 µg was loaded per each un-purified lane, 3 µg for concentrated elute and 50 ng for the purified GFPuv positive control. **A**: Stain-free SDS-PAGE gel, **B:** In-gel fluorescence (488 nm excitation, Cy2 filter). **C:** First derivative Nanotemper Tycho intrinsic fluorescence 330/350 nm ratio thermal un-folding profiles of the concentrated elute sample in **A-B**. **D:** Fitted mass photometry of the concentrated elute sample in **A-B.** Rapid 1:100 dilution in detergent free buffer, final concentration ~0.0032 mg/ml. The fitted mass values to each gaussian are indicated.

Mass photometry measurements (Figure S1.2D) of the concentrated elute diluted 1:100 gave similar results to detergent purified protein, with gaussians fit at 97 (σ 19.2, n=673, 60%), 132 (σ 23, n=440, 39%) and 200 (σ 19.2, n=673, 60%) kDa. Previously, in detergent samples similar gaussians at 98, 215 and 136 kD were assigned as possible GAT1-mEm$_{(PFB)}$ monomers, dimers and dimers with both mEmeralds removed respectively. Several additional gaussians could also be fit though with the SMA extracted protein and there appeared to be less noise, possibly as many empty SMA co-polymers would have already been removed by passing into the concentrator flow through. It was attempted to also measure the mass of empty SMA for comparison, but this was unsuccessful, likely as the small size (6.5 kDa) was beneath the detection limit of the instrument. The identity of the 79 (σ 9.3, n=292, 26%) kDa gaussian is unclear, but it could represent GAT1-

$mEm_{(PFB)}$ where the mEmerald has been proteolytically cleaved (theoretical mass ~71 kDa). Likewise, the gaussian fitted at 164 (σ 27, n=289, 26%) kDa could represent a dimer with one mEmerald removed.

S1.3: Conclusions

The initial results using SMA for extraction of GAT1-$mEm_{(PFB)}$ showed promise for future work. It would be exciting to see how using the improved purification procedures that were developed using detergents could fare when applied to SMA extracted protein. It would also be especially interesting to see whether GAT1-$mEm_{(EB)}$ shows a similar improvement in SMA as it did in detergent. Newer SMA polymers with improved properties have also been released since receiving the samples from Polyscope and it would be noteworthy to see if they may be of benefit.

S2: Copies of publications

Bueno, R. V., Davis, S., Dawson, A., Ondachi, P.W., Carroll, F.I., and Hunter, W.N. (2022). Interactions between 2′-fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues and acetylcholine binding protein inform on potent antagonist activity against nicotinic receptors. Acta Crystallogr. Sect. D Struct. Biol. *78*: 353–362.

Davis, S., Rego Campello, H., Gallagher, T., and Hunter, W.N. (2020). The thermodynamic profile and molecular interactions of a C(9)-cytisine derivative-binding acetylcholine-binding protein from Aplysia californica. Acta Crystallogr. Sect. F Struct. Biol. Commun. *76*: 74–80.

STRUCTURAL BIOLOGY
ISSN 2059-7983

Interactions between 2′-fluoro-(carbamoyl-pyridinyl)deschloroepibatidine analogues and acetylcholine-binding protein inform on potent antagonist activity against nicotinic receptors

Renata V. Bueno,[a]‡ Samuel Davis,[a]‡ Alice Dawson,[a] Pauline W. Ondachi,[b] F. Ivy Carroll[b] and William N. Hunter[a]*

[a]Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom, and [b]Research Triangle Institute, PO Box 12194, Research Triangle Park, Durham, NC 27709, USA. *Correspondence e-mail: w.n.hunter@dundee.ac.uk

Received 27 September 2021
Accepted 21 January 2022

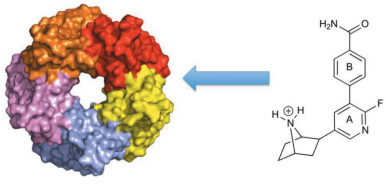
Edited by R. C. Garratt, University of Sâo Paulo, Brazil

‡ Joint first authors.

Keywords: acetylcholine-binding protein; biolayer interferometry; crystal structure; epibatidine derivatives; ligand-gated ion channels; nicotinic acetylcholine receptors.

PDB references: *Ac*AChBP, complex with compound **1**, 6qkk; complex with compound **2**, 6qqp; complex with compound **3**, 6qqo

Supporting information: this article has supporting information at journals.iucr.org/d



Low-nanomolar binding constants were recorded for a series of six 2′-fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues with acetylcholine-binding protein (AChBP). The crystal structures of three complexes with AChBP reveal details of molecular recognition in the orthosteric binding site and imply how the other three ligands bind. Comparisons exploiting AChBP as a surrogate for $\alpha4\beta2$ and $\alpha7$ nicotinic acetylcholine receptors (nAChRs) suggest that the key interactions are conserved. The ligands interact with the same residues as the archetypal nAChR agonist nicotine yet display greater affinity, thereby rationalizing their *in vivo* activity as potent antagonists of nicotine-induced antinociception. An oxyanion-binding site is formed on the periphery of the AChBP orthosteric site by Lys42, Asp94, Glu170 and Glu210. These residues are highly conserved in the human $\alpha4$, $\beta2$ and $\alpha7$ nAChR sequences. However, specific sequence differences are discussed that could contribute to nAChR subtype selectivity and in addition may represent a point of allosteric modulation. The ability to engage with this peripheral site may explain, in part, the function of a subset of ligands to act as agonists of $\alpha7$ nAChR.

1. Introduction

Nicotinic acetylcholine receptors (nAChRs) are cation-selective pentameric ligand-gated ion channels (pLGICs) gated by the neurotransmitters acetylcholine and choline. They are also the targets of non-endogenous molecules, with the best known being the archetypal agonist nicotine. The availability of a variety of subunits enables neuronal nAChRs to assemble in different homopentameric or heteropentameric combinations (Gotti *et al.*, 2009; Taly *et al.*, 2009; Bertrand & Terry, 2018). The most common types in the mammalian central nervous system are an $\alpha4\beta2$ combination, which has a high affinity for nicotine, and the $\alpha7$ homomer (Taly *et al.*, 2009). The pLGIC family presents a number of therapeutic targets for neurological conditions; specifically, nAChR subtypes are key targets for the development of compounds with use in the treatment of nicotine addiction and also of pain (Bertrand *et al.*, 2015; Dineley *et al.*, 2015; Bagdas *et al.*, 2018) The discovery of epibatidine [(1*R*,2*R*,4*S*)-2-(6-chloro-3-pyridinyl)-7-azabicyclo[2.2.1]heptane; Fig. 1], a highly potent but relatively nonselective agonist of nAChR that displays powerful non-opiate-mediated antinociceptive effects, elicited great excitement (Spande *et al.*, 1992; Traynor, 1998). A serious liability due to toxicity rules out therapeutic use; nevertheless, with high ligand efficiency and potency the compound has provided a basis for the development of

analogues that target nAChR (Spang *et al.*, 2000; Carroll, 2004; Mu *et al.*, 2006; Ondachi *et al.*, 2016). One study focused on the position of the chloropyridine ring N atom and resulted in conversion from an agonist to an antagonist profile (Spang *et al.*, 2000).

Our interest centres on a series of 2′-fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues (compounds **1**–**6**; Fig. 1), characterized as having high affinity (K_i < 1 n*M*) for $\alpha4\beta2$ nAChR, that display differing degrees of subtype selectivity and novel pharmacological effects compared with epibatidine (Ondachi *et al.*, 2016). Electrophysiological measurements of ion-channel activity indicated that compounds **1**–**6** have little or no agonist activity on recombinant $\alpha4\beta2$ or $\alpha3\beta4$ nAChRs. In contrast, the parent compound epibatidine is a full agonist. The compound set demonstrated antagonist activity on $\alpha4\beta2$ and $\alpha3\beta4$ nAChRs. However, puzzling effects were observed on the homomeric $\alpha7$ nAChR. Compounds **1** and **3** act as mixed partial agonists of $\alpha4\beta2$, whilst compound **2** is only an antagonist of $\alpha7$ nAChR, with no agonist properties detected. Furthermore, compounds **4** and **5** were selective for $\alpha4\beta2$ over $\alpha7$, and *in vivo* studies indicated that these analogues all antagonize the antinociceptive action of nicotine with, in the case of compound **2**, a potency approaching that of varenicline, a well studied partial agonist of $\alpha4\beta2$ receptors and a full agonist of $\alpha7$ nAChR (Ondachi *et al.*, 2016). The differing actions at nAChR subtypes are perplexing and we sought to investigate further.

The biological target nAChRs follow the standard structural arrangement of pLGICs, with five subunits creating a central ion pore. Each subunit possesses an extracellular domain (ECD) followed by four transmembrane α-helices and intracellular contributions from the inter-helical segments (Taly *et al.*, 2009; Bertrand & Terry, 2018). The orthosteric binding site, where agonists and competitive antagonists bind, is created by contributions from the ECD at the interface between two subunits. One subunit contributes the principal (+) face, which is formed of three loops known as A, B and C. On an adjacent subunit, loops D, E, F and G comprise the complementary (−) side of the binding site (Sixma & Smit, 2003; Corringer *et al.*, 2012; daCosta & Baenziger, 2013; Nys *et al.*, 2013; Sauguet *et al.*, 2015). Acetylcholine-binding protein, which is found in the cholinergic synapse of gastropods including *Aplysia californica* (*Ac*AChBP), shares 20–25% sequence identity with the ECD of nAChR sequences. In *Ac*AChBP, 44 residues are involved in the orthosteric binding site and the identities of these residues in human nAChRs range from 32% ($\beta2$) to 45% ($\alpha4$ and $\alpha7$). In addition, AChBP and nAChR display closely related structures and similar ligand-binding properties (Celie *et al.*, 2004; Hansen *et al.*, 2005; Lemoine *et al.*, 2012; Rucktooa *et al.*, 2012; Shahsavar *et al.*, 2016). Despite the hugely impressive developments in studies of membrane-bound pLGIC forms, for example the cryo-EM structure of the heteromeric human $\alpha4\beta2$ nAChR (Walsh *et al.*, 2018), *Ac*AChBP remains a valued surrogate for the study of ligand–receptor interactions. This is due to the convenience of working with a stable, soluble protein for which efficient purification protocols exist and where ordered single crystals can be obtained.

We sought to exploit *Ac*AChBP in this manner to investigate the interactions and affinities of epibatidine analogues **1**–**6** (Fig. 1). Biolayer interferometry (BLI) provided K_d values, and the crystal structures of three complexes revealed the mode of binding and interactions of compounds **1**–**3**, providing models for complexes with compounds **4**–**6**. Comparisons and modelling using nAChR sequences and structures allowed us to consider the bioactivity of the compound series and to comment on structure–function relationships

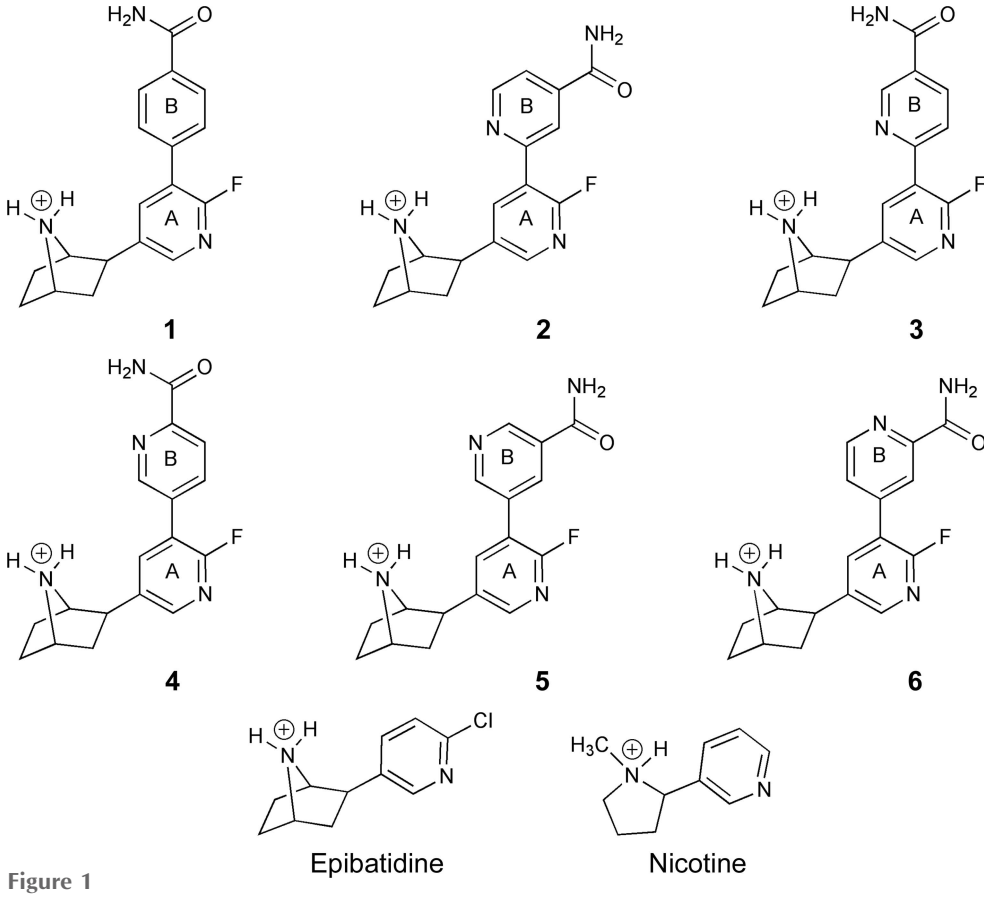


Figure 1
The structures of epibatidine, nicotine and analogues **1**–**6** used in this work. In all cases the protonated form, which is most likely under physiological conditions, is shown. The first aromatic ring after the azabicycloheptane (a fluoropyridine) is labelled A and the second ring labelled B is a benzylcarboxamide in compound **1** and a pyridinecarboxamide in compounds **2**–**6**.

that may be exploited in the design of new nAChR ligands as chemical tools and/or with therapeutic potential.

2. Materials and methods

2.1. Recombinant protein production

A gene encoding *Ac*AChBP with a C-terminal His_6 tag was cloned into the pFastBac1 vector (Thermo Fisher). The amino-acid sequence was derived from the *A. californica* genome (https://www.broadinstitute.org/aplysia/aplysia-genome-project) and is similar to UniProt entry Q8WSF8 except that two alanine residues are replaced by Val60 and Val155. This construct was expressed using the Bac-to-Bac system (Thermo Fisher) in *Spodoptera frugiperda* (*Sf*9) cells maintained in shaker flasks at 27°C using Insect-XPRESS medium (Lonza) supplemented with 2 m*M* L-glutamine and 100 U ml^{-1} penicillin/streptomycin (Thermo Fisher). The *Ac*AChBP baculovirus was generated through transfection of 500 ng bacmid DNA into adherent *Sf*9 cells at a density of 8×10^5 cells per millilitre using Insect Genejuice reagent (Novagen). Following seven days of incubation at 27°C, the virus was amplified twice in *Sf*9 suspension cultures before being harvested from the medium via centrifugation (1000*g*, 10 min, 4°C).

For protein production, *Sf*9 cells (15×10^5 cells per millilitre) were infected with virus and incubated for approximately 48 h before being separated from the medium by centrifugation (1000*g*, 10 min, 12°C). The recombinant *Ac*AChBP is secreted, so the supernatant was clarified by further centrifugation (4000*g*, 10 min, 12°C), concentrated and exchanged into buffer *A* (50 m*M* Tris–HCl, 250 m*M* NaCl pH 7.5) using a Sartojet system (Satorius) with a Sartocon Slice 10 kDa cutoff microfiltration cassette. *Ac*AChBP was purified by immobilized metal-ion chromatography with Ni^{2+} HisTrap columns (GE Healthcare) and eluted using a linear gradient of buffer *B* (50 m*M* Tris–HCl, 250 m*M* NaCl, 800 m*M* imidazole pH 7.5). Sample purity was assessed using stain-free SDS–PAGE gels (Bio-Rad). The fractions containing *Ac*AChBP were pooled, buffer *B* was exchanged for buffer *A* using 10 kDa centrifugal concentrators and the protein was concentrated to 6 mg ml^{-1}.

2.2. 2′-Fluoro-(carbamoylpyridinyl)deschloroepibatidine analogues

Compounds **1**–**6** were synthesized as racemic mixtures as described previously (Ondachi *et al.*, 2016) and then dissolved in DMSO as 100 m*M* stock solutions.

2.3. K_d determination by BLI

The K_d values of compounds **1**–**6** were determined using an Octet RED system (Forté-Bio). *Ac*AChBP (0.5 mg ml^{-1} in buffer *A*) was immobilized on Ni^{2+}–NTA sensors for 600 s, followed by the removal of non-immobilized protein (600 s) and baseline stabilization (120 s). Association and dissociation measurements were obtained at 25°C for six concentrations of each compound (1.7–410 n*M*) diluted in buffer *A* plus 1% DMSO. BLI assays were conducted with a baseline measurement in buffer (60 s), an association measurement in the well containing the compound (120 s) and a dissociation measurement in buffer (120 s). Data processing and analysis were performed with *Octet RED Data Analysis* version 7.1 (Forté-Bio), where the background response of immobilized *Ac*AChBP in buffer was subtracted.

2.4. Co-crystallization of *Ac*AChBP with compounds

*Ac*AChBP (4 mg ml^{-1} in buffer *A*) was incubated with 2 m*M* of compounds **1**–**6** for 1 h at room temperature and this mixture was then used in crystallization trials. Compounds **4**–**6** did not yield suitable crystals. Small well formed crystals of *Ac*AChBP with compounds **1**–**3** were obtained using hanging drops consisting of 2 µl complex solution plus 1 µl reservoir solution equilibrated against 800 µl reservoir solution for 24 h at 18°C. The reservoirs consisted of 0.2 *M* NaCl, 0.1 *M* phosphate–citrate pH 4.2 and PEG 8000 at 8% (compound **2**), 10% (compound **3**) or 12% (compound **1**). These conditions allowed us to prepare microseed stocks by transferring crystals to microcentrifuge tubes containing the appropriate reservoir solution, 40% glycerol and a Seed Bead (Hampton Research), and vortexing the sample for 1 min. A cleaned human eyelash was dipped into the seed stock and then passed through freshly assembled crystallization drops. Well ordered multi-faced prisms, with maximum dimensions of about 20 µm, appeared after four days.

2.5. Crystallographic analyses

Crystals were harvested using a nylon loop, cryoprotected with reservoir solution adjusted to contain 20% ethane-1,2-diol and then flash-cooled in liquid nitrogen. Diffraction was recorded on Diamond Light Source (DLS) microfocus beamline I24 using a PILATUS 6M-F detector (Dectris) and the images were indexed and integrated using *XDS* (Kabsch, 2010). The data were scaled using *AIMLESS* (Evans & Murshudov, 2013) from the *CCP*4 suite (Winn *et al.*, 2011) and the structures were solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007), utilizing wild-type protein in complex with strychnine at 1.91 Å resolution (PDB entry 2xys; Brams *et al.*, 2011) as the model for *Ac*AChBP–**1** and the complex with nicotine at 2.20 Å resolution (PDB entry 5o87; Dawson *et al.*, 2019) as that for *Ac*AChBP–**2** and *Ac*AChBP–**3**. Ligand models and restraints were generated with the *grade* server (Global Phasing; http://grade.globalphasing.org/cgi-bin/grade/server.cgi). Multiple rounds of automated restrained refinement were completed using *REFMAC*5 (Murshudov *et al.*, 2011), with manual refinement and model building in *Coot* (Emsley *et al.*, 2010). The epibatidine analogues were well defined in electron and difference density maps (Supplementary Fig. S1). Asn91 is glycosylated and *N*-acetyl-D-glucosamine (NAG) was modelled onto several subunits. It became clear that additional ligands were present and these were assigned and refined satisfactorily as phosphate, ethane-1,2-diol and oxalate. The latter is likely to be present as a contaminant in PEG 8000 (Fyfe *et al.*, 2010). Strict noncrystallographic symmetry restraints were applied throughout to

Table 1
Crystallographic statistics for the *Ac*AChBP–ligand complexes.

Values in parentheses are for the highest resolution shell.

	*Ac*AChBP–**1**	*Ac*AChBP–**2**	*Ac*AChBP–**3**
PDB code	6qkk	6qqp	6qqo
Data collection			
a, *b*, *c* (Å)	211.0, 131.6, 131.8	209.5, 136.9, 131.5	209.5, 136.9, 131.5
α, β, γ (°)	90, 102.8, 90	90, 102.7, 90	90, 102.6, 90
Space group	*C*2	*C*2	*C*2
Source	DLS microfocus beamline I24	DLS microfocus beamline I24	DLS microfocus beamline I24
Wavelength (Å)	0.96858	0.96858	0.96858
Subunits per asymmetric unit	10	10	10
Resolution range (Å)	48.83–2.20 (2.24–2.20)	46.82–2.40 (2.44–2.40)	46.40–2.50 (2.54–2.50)
Other ligands	NAG, ethane-1,2-diol, phosphate, oxalate	NAG, ethane-1,2-diol, phosphate	NAG, ethane-1,2-diol, phosphate
Total No. of reflections	611399 (30075)	405440 (19537)	254487 (12673)
Unique reflections	177603 (8766)	139199 (6884)	109552 (5537)
Multiplicity	3.4 (3.4)	2.9 (2.8)	2.3 (2.3)
R_{merge}	0.172 (0.888)	0.089 (0.510)	0.118 (0.559)
$R_{p.i.m.}$	0.162 (0.831)	0.088 (0.497)	0.114 (0.539)
Wilson *B* factor (Å^2)	15.5	29.0	13.6
Completeness (%)	99.9 (98.4)	98.7 (98.7)	89.7 (91.8)
$\langle I/\sigma(I)\rangle$	6.6 (2.0)	8.2 (2.5)	6.9 (2.6)
$CC_{1/2}$	0.95 (0.43)	0.95 (0.65)	0.93 (0.46)
Refinement			
R_{work}/R_{free}	0.199/0.228	0.188/0.210	0.193/0.211
No. of reflections for R_{work}/R_{free}	168732/8867	131926/6903	104105/5421
Protein residues	2065	2054	2053
No. of ligands	10	10	10
No. of water molecules	1287	924	659
R.m.s.d.s			
Bond lengths (Å)	0.048	0.011	0.012
Angles (°)	0.97	1.49	1.56
Ramachandran plot			
Residues in favoured regions	2002	1981	1964
Residues in allowed regions	35	30	35
Residues in outlier regions	0	1	0
Mean *B* factors (Å^2)			
Protein atoms per subunit	18.4/15.7/17.3/19.2/17.3/18.4/20.5/22.2/22.6/19.2	34.2/30.7/33.1/29.9/32.2/37.7/40.4/41.2/34.4/34.0	19.7/17.3/19.1/17.2/18.9/22.2/24.0/23.6/19.3/19.6
Water molecules	22.2	36.1	17.1
Ligand	14.5/14.2/16.4/14.8/15.3/16.8/17.8/20.7/18.1/18.1	32.8/31.0/28.7/26.1/35.4/38.9/43.7/40.7/30.2/35.6	17.9/19.8/22.0/14.4/26.8/28.1/35.1/27.0/16.3/29.0
NAG	68.7	112.7	87.6
Ethane-1,2-diol	34.4	46.7	36.6
Phosphate	41.0	—	54.9
Oxalate	41.5	—	—

*Ac*AChBP–**2** and *Ac*AChBP–**3** but were relaxed towards the end of the refinement for *Ac*AChBP–**1**. Ligand 3D solvent-accessible surface areas were calculated using *AREAIMOL*, with a probe solvent radius of 1.4 Å and 100 surface points per Å^2. Omit maps were generated by removal of the ligands and bulk-solvent corrections before recalculating the $F_o - F_c$ maps presented in Supplementary Fig. S1. Graphics were rendered using the *PyMOL* molecular-graphics system (Schrödinger). Crystallographic statistics are presented in Table 1. The amino-acid sequences used include human α4 (UniProt code P43681), α7 (P36544) and α2 (P17787) nAChR subtypes.

3. Results and discussion

3.1. Binding properties

The affinity of *Ac*AChBP for compounds **1**–**6** was assessed by BLI (Table 2; representative sensorgrams are shown in Supplementary Fig. S2). All compounds displayed K_d values in the low-nanomolar range, comparable to that of epibatidine (Ondachi *et al.*, 2016), but show a higher affinity than that seen for nicotine. The ligands can be placed into three groups based on affinity for *Ac*AChBP: compounds **1** and **6** display K_d values close to 10 n*M*, while the values for compounds **3**, **4**, **5** are around 30 n*M* and that for compound **2** is 60 n*M*. For comparison, the K_i values determined with a [^{3}H]-epibatidine displacement assay against $\alpha4\beta2$ nAChR are presented in Table 2. Consistent with previous *Ac*AChBP ligand-binding studies (Hansen *et al.*, 2005), the affinities are reduced significantly compared with the subnanomolar levels observed with actual $\alpha4\beta2$ nAChR (Ondachi *et al.*, 2016).

3.2. Crystallographic analyses

We attempted to co-crystallize all six compounds with *Ac*AChBP and success with three led to structures of the *Ac*AChBP–**1**, *Ac*AChBP–**2** and *Ac*AChBP–**3** complexes at resolutions of 2.2, 2.4 and 2.5 Å, respectively. The monoclinic

Table 2
Ligand-binding properties.

Values are given with standard errors.

Ligand	*Ac*AChBP, K_d (n*M*)	$\alpha_4\beta_2$ nAChR, K_i† (n*M*)
1	9.8 ± 0.03	0.12 ± 0.020
2	60.0 ± 0.14	0.28 ± 0.010
3	29.0 ± 0.05	0.94 ± 0.070
4	33.0 ± 0.08	0.07 ± 0.002
5	30.0 ± 0.06	0.28 ± 0.030
6	9.8 ± 0.01	0.67 ± 0.280
Epibatidine	14.0‡	0.02 ± 0.001
Varenicline	342.0§	0.12 ± 0.002
Nicotine	835.0§	0.95¶

† [^{3}H]-Epibatidine competition assay, K_d = 0.02 n*M* (Ondachi *et al.*, 2016). ‡ Tryptophan fluorescence-quenching assay (Hansen *et al.*, 2005). § Isothermal titration calorimetry (Rucktooa *et al.*, 2012). ¶ [^{3}H]-Nicotine competition assay (Coe *et al.*, 2005).

crystals obtained in each case are isomorphous, with two pentameric assemblies comprising the asymmetric unit. A high degree of noncrystallographic symmetry was evident and was thus maintained in the refinement calculations, although for *Ac*AChBP-**1**, the highest resolution structure, these restraints were released in the final calculations. The ligands of interest are well defined in the electron density observed in each of the ten binding sites of the structures (Supplementary Fig. S1) and were refined with average *B* factors that were lower than or close to the values noted for their associated subunits (Table 1). Within each complex, the orientation of the ligand and the pattern of interactions within each binding site are essentially identical and it is only necessary to describe one. Enantiomeric mixtures of compounds **1**–**6** were used, and in the structures with compounds **1**–**3** we tested both forms in modelling to the electron density. The resolution of the crystal structures was insufficient to distinguish (+) and (−) enantiomers or whether a mixture was present, so we used the former to match that of the naturally occurring form of epibatidine, the parent compound. Note that the (+) and (−) forms of epibatidine have very similar binding and biological properties (Mu *et al.*, 2006; Dallanoce *et al.*, 2012) and we judge it likely that this also applies to compounds **1**–**6**.

There are three crystal structures containing epibatidine in the Protein Data Bank (PDB) that are relevant to our study. These are a low-resolution (3.4 Å) complex with *Ac*AChBP (PDB entry 2byq; Hansen *et al.*, 2005), a 3.2 Å resolution structure with α2 (PDB entry 5fjv; Li *et al.*, 2011) and a 2.8 Å resolution complex with an α7 chimera (PDB entry 3sq6; Kouvatsos *et al.*, 2016). Compounds **1**–**3** and epibatidine share the azabicyclo[2.2.1]heptane moiety and this part of the ligand binds in a similar way in all structures, deep in a hydrophobic part of the binding site formed primarily on the principal side. Here, the orthosteric site is dominated by the presence of aromatic residues (Fig. 2). The protonated amine donates hydrogen bonds to Tyr110 OH and the Trp164 carbonyl group. The position of Tyr110 OH is fixed by hydrogen-bonding interactions with the Ser163 carbonyl and a network of ordered water molecules that form bridges through to Tyr205, and Asp214 (not shown). There are van der Waals interactions with Tyr72, Tyr205, Tyr212, Trp164 and Cys207. The amine is around 4 Å distant from the face of the Trp164 side chain and is positioned to suggest the presence of a cation–π interaction. This interaction is a common and important feature of pLGIC ligand complexes (Taly *et al.*, 2009; Nys *et al.*, 2013) and specifically nAChRs (Zhong *et al.*, 1998). The halogen-substituted pyrimidine ring A, which is common to the series of compounds and epibatidine, is directed towards the complementary side of the orthosteric site, with one face of the aromatic system forming van der Waals interactions with Val165 from the principal side and the other face interacting with Ile135 on the complementary side. A side-on interaction with Cys208 is evident. In the complexes of *Ac*AChBP with compounds **1**, **2** and **3**, the fluorine is directed into a shallow pocket and forms van der Waals interactions with the main chain of Ala124 and Phe134 and the side chains of Val125 and Ile135, all from the complementary side. The pyridine N accepts a hydrogen bond from a water molecule that is at one end of an ordered solvent chain extending to the surface of the protein, forming hydrogen bonds to residues in the orthosteric site, for example to Ile123 O and Trp164 NE1. These parts of the ligands replicate key features found in AChBP–nicotine complex structures; in particular, the presence of an ordered water molecule linking the pyridine N atom to the protein is noted repeatedly (Hansen *et al.*, 2005; Nys *et al.*, 2013; Dawson *et al.*, 2019). The available AChBP–epibatidine complex structures are at low resolution and lack solvent molecules in the binding sites. However, this hydration pattern is strictly conserved across the three structures reported here and is observed in other structures of AChBP complexes (Hansen *et al.*, 2005; Nys *et al.*, 2013; Dawson *et al.*, 2019). It was previously thought that the presence of such an ordered water molecule correlates with agonist activity of the ligand (Nys *et al.*, 2013), but our structures suggest that such a conclusion does not apply in all cases, a point that we will revisit below.

The structures in the series **1**–**6** have a similar substituent: a phenyl or pyridine ring (termed ring B), with a carboxamide substituent, at the 4′ position of ring A (Fig. 1). These represent extensions of the epibatidine scaffold. Compound **1** has a phenyl group, and compounds **2**–**6** have pyridines with the N atom at position 2, 3 or 4. A further variation is a carboxamide substituent attached to the pyridine N atom at the *meta* or *para* position. The structures of the complexes of *Ac*AChBP with compounds **1**–**3** all suggest that ring B participates in van der Waals interactions with Arg96, Val125 and Met133 from the complementary side and Cys208 on the principal side (Fig. 2). In the structure of *Ac*AChBP–**1** the alignment and distances (around 3.3 Å) of the Tyr212 OH group and a carbon on ring B suggest the possibility of a C—H···O hydrogen bond (Fig. 2*a*). In the complexes with compounds **2** and **3** (Fig. 2*b*) the pyridine N atom participates in a hydrogen bond to Tyr212 with distances of around 2.6–3.0 Å.

In the *Ac*AChBP–**1** and *Ac*AChBP–**3** complexes the carboxamide is directed outwards towards bulk solvent. The carbonyl groups participate in water-mediated links to Val125, Thr127 and Ser131 (not shown). The amide groups contribute to a phosphate-binding site, with the oxyanion also coordinated by the side chains of Asp94 and Arg96 on the

complementary side and then by Glu170, Glu210, Lys42 and Ser167 from the principal side. This is further detailed below. Phosphate was a component of the crystallization mixture, and the acidic pH used explains why aspartate and glutamate side chains coordinate the oxyanion. In eight binding sites in *Ac*AChBP–**1** and seven in *Ac*AChBP–**3** the amide forms direct hydrogen bonds to Asp94 on the (−) side, perhaps helped by the ordering influence of the oxyanion. Due to the difference in the ring position of the carboxamide substituent in compound **2**, in *Ac*AChBP–**2** the carbonyl group of the ligand is tilted towards loop E on the complementary side, in particular Met133, but ring B is positioned further away from the methionine side chain. An adjustment of the Tyr212 side chain is observed, serving to maintain the hydrogen bond to the ring B N atom. The result is to position the carboxamide in roughly the same position in all three complexes. The pyridine ring B of compound **2** in comparison is posed closer to Tyr212, the position of the amide group is adjusted and our analysis indicated that ethane-1,2-diol, the cryoprotectant, and not phosphate interacts with the amide and engages with the same residues that form the phosphate-binding pocket. A link from the amide of compound **2** to Asp94 is retained but *via* bridging water molecules (not shown). The position of ring B may provide a disruptive block to phosphate binding, and the reduction in van der Waals interactions with Met133 may contribute to compound **2** displaying the lowest affinity for *Ac*AChBP.

Loop C, on the (+) side, contributes significantly to the creation of the orthosteric site and interacts directly with residues on the (−) side. This loop has to present an open conformation to allow ligands to enter the site and then close as the complex forms (Jadey & Auerbach, 2012). When ligands **1**–**3** occupy the site, loop C closes over them (Fig. 3). The solvent-accessible surface areas of unbound compounds **1**–**3** are approximately 500 Å^2, and between 91% and 94% of this surface is buried in the complexes with *Ac*AChBP, with primarily the carboxamide moiety directed towards solvent. Epibatidine is smaller (Fig. 1), with a solvent-accessible area of 360 Å^2, and is almost entirely buried (>99%) deep in the binding site, with loop C in the agonist-bound state (Fig. 3; Spang *et al.*, 2000).

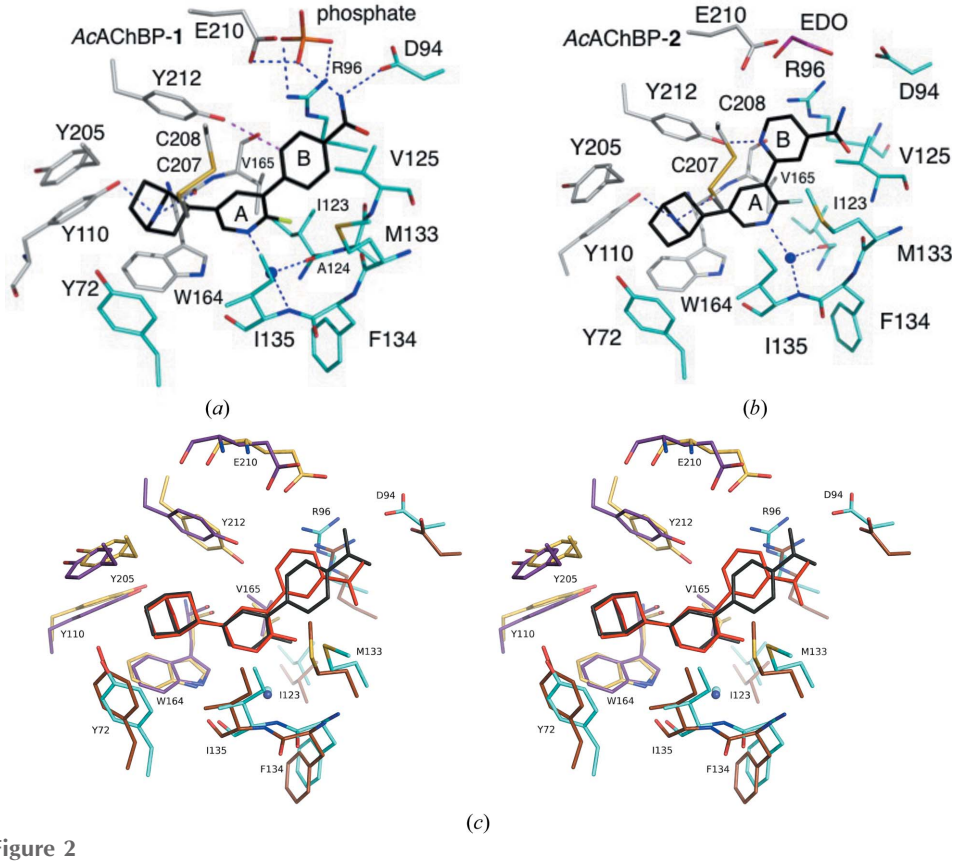


Figure 2
Selected residues involved in interactions with ligands. (*a*) *Ac*AChBP–**1**. (*b*) *Ac*AChBP–**2**. Atoms are coloured using the following scheme: C atoms of the ligands are black, those in the principal subunit are grey and those on the complementary side are cyan, O atoms are red, N atoms are blue, S atoms are yellow and F atoms are green. A marine sphere indicates the position of a conserved water molecule. Blue dashed lines represent potential hydrogen bonds. Rings A and B are labelled. In (*a*) a possible C—H···O hydrogen bond is shown as a red dashed line and the P atom is orange, and in (*b*) the C atoms of EDO (ethane-1,2-diol) are pink. Water molecules, main-chain atoms of some residues, for example Cys207-Cys208, and Ala124 in (*b*) are omitted for clarity. (*c*) Stereoview overlay of the two structures. Principal side residues for the complex with compound **1** are shown in violet/purple and those for the complex with compound **2** are shown in yellow/orange. Complementary side residues for the complex with compound **1** are shown in cyan and those for the complex with compound **2** are in brown. The conserved water is shown as a sphere in aquamarine and in dark blue for the complexes with compounds **1** and **2**, respectively. Ligand **1** is coloured black and ligand **2** is coloured red. The coordinates were overlaid using the superpose ligands option in *Coot*.

Compounds **2**, **5** and **6** differ only in the position of the N atom on pyridine ring B. The *Ac*AChBP–**2** complex indicates a displacement of ring B compared with the other structures and we speculate that this might influence the interactions in and around the observed oxyanion-binding site. The placement of the ring B N atom in compound **6** would position a hydrogen-bond acceptor in an ideal position to participate in formation or organization of the anion-binding site. Using an overlay of complexes with compounds **1** and **2** as a template, the N atom in ring B would be about 2.4 Å from a phosphate in compound **6** and around 2.9 Å in compound **5**.

Comparisons of the *Ac*AChBP complexes with the cryo-EM structure of nAChR in complex with nicotine (PDB entry 6cnj; Walsh *et al.*, 2018) and the sequences of human α4, α7 and β2 subtypes were carried out (Fig. 4). We considered the orthosteric sites of human

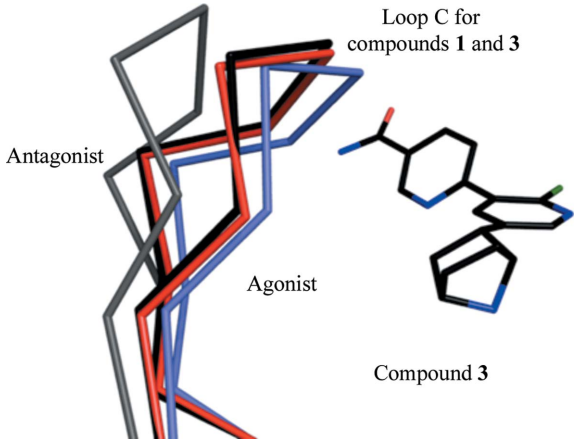


Figure 3
Loop C of *Ac*AChBP adopts different conformations. Ribbon representation of loop C segments with the bound agonist epibatidine (cyan, PDB entry 2byq; Hansen *et al.*, 2005), antagonist strychnine (grey, PDB entry 5o8t; Dawson *et al.*, 2019) and the *Ac*AChBP–**1** and *Ac*AChBP–**3** complexes (red and black, respectively). Compound **3** is shown and coloured with the following scheme: C, black; O, red; N, blue; F, dark green. The overlay was calculated using all C^{α} atoms of a single subunit.

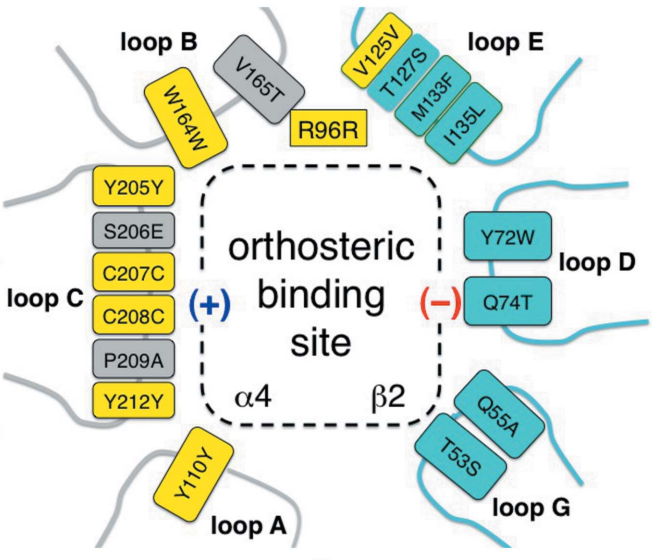
heteromeric $\alpha4(+)/\beta2(-)$ nAChR and homomeric $\alpha7$ nAChR and first asked how similar the mode of binding of compounds **1**–**3**, and by inference compounds **4**–**6**, might be. The conservation of sequence and structure, in particular considering the aromatic cage of the binding site, suggests that the orientation of the epibatidine analogues is representative of how this compound series would bind to any nAChR. This conclusion is supported by site-directed mutagenesis studies with $\alpha7$ nAChR, where changes to six residues abolished epibatidine binding (Thompson *et al.*, 2017). These six residues, which are highly conserved as a group, correspond to Tyr72, Tyr110, Tyr205, Tyr212, Ile123 and Trp164 in *Ac*AChBP (Figs. 2 and 4).

The similarity of the amino acids implicated in interactions with the compound set **1**–**6** extends beyond these six residues. On the principal side, Val165 of *Ac*AChBP, which participates in van der Waals interactions with the ligands, aligns with Thr183 and Ser172 in the $\alpha4$ and $\alpha7$ subtypes, respectively. In terms of size, these represent conservative substitutions. On loop C, Ser206 is changed to a glutamate in nAChR but the side chain is directed away from the binding site. On the complementary side, Met133 and Ile135 of *Ac*AChBP are

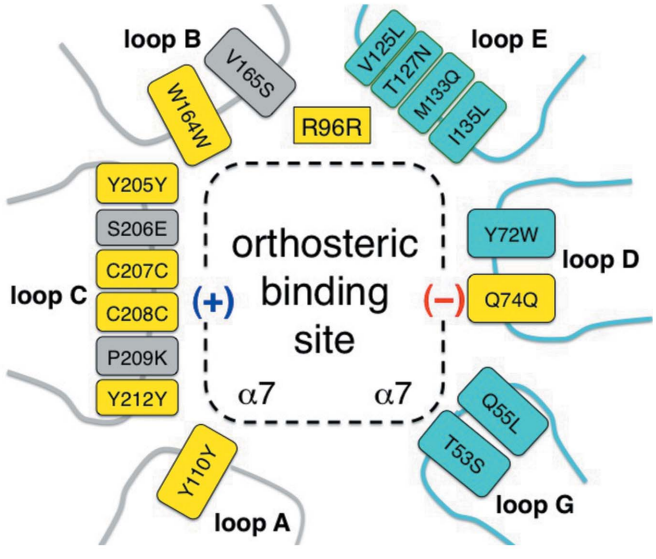
PRINCIPAL (+) COMPLEMENTARY (–)

loop	A	B	C	D	E	F	G
	105	161	204	72	133	181	51
AChBP	PDITAY	FGSWVY	HYSCCPEPY	YEQQRW	MFIPA	DLSSYY	GFTLQD
	121	179	222	88	150	199	67
nAChRα4	PDIVLY	FGSWTY	KYECCAEPY	WVKQEW	QWTPP	DQLDFW	GLSIAQ
	110	168	209	77	139	186	56
nAChRα7	PDILLY	FGSWSY	FYECCKEPY	WLQMSW	QYLPP	DISGYI	SLSLLQ
	115	173	116	82	144	193	61
nAChRβ2	PDVVLY	FRSWTY	P-D--DSTY	WLTQEW	FWLPP	SLDDFT	MVSLAQ

(*a*)



(*b*)



(*c*)

Figure 4
(*a*) Alignment of selected sequence segments that form the orthosteric binding sites in *Ac*AChBP and $\alpha4$, $\alpha7$ and $\beta2$ nAChRs. Loops are labelled and are split into principal (+) and complementary (−) sides. Residues coloured blue are key and are discussed in the text. Val125 and Thr127 of *Ac*AChBP are in the N-terminal section of loop E but are left out for clarity. (*b*) A schematic of the orthosteric binding site listing the key residues of *Ac*AChBP and the corresponding residues in the $\alpha4(+)\beta2(-)$ nAChR heteromeric site. Loop F is out of range of the ligands discussed in this work and has been omitted. Arg96 is included given its role in binding the anion (see text). Boxes are coloured yellow to highlight strict conservation, grey for the (+) side and cyan for the (−) side. (*c*) Schematic with the same format for the $\alpha7$ nAChR homomeric site.

Phe144 and Leu146 in nAChR β2 and Gln139 and Leu141 in α7, respectively. At the periphery of the binding site the combination of Val125 and Thr127 in *Ac*AChBP is Val136 and Ser138 in β2 and Asn133 and Leu131 in α7. The noteworthy difference is the Met133Gln substitution, which in α7 nAChR will place a polar side chain near the ligands. It is possible that the latter would provide a steric restriction on the placement of ring B in **1** and **3**, and also that the glutamine could form a hydrogen bond to the carboxamide substituent when binding **2**.

Different ideas have been proposed to explain the activity of agonists and antagonists of pLGICs based on the crystal structures of AChBP–ligand complexes (Taly *et al.*, 2009; Lemoine *et al.*, 2012; Bertrand *et al.*, 2015). The use of *Ac*AChBP provides binding and structural data that inform on affinity and aspects of molecular recognition, but it is important to recognize the limitations of this surrogate (for example it has no transmembrane domains) and exercise caution when considering aspects of channel opening and closing. At least three factors appear to contribute to the distinctive responses of ligands that bind to the orthosteric site of pLGICs. Firstly, residues on the complementary side are the primary determinants of ligand affinity; in particular, by combining the presence of cation–π and van der Waals interactions with the aromatic cage to stabilize the complex. This area of the orthosteric site is well targeted by compounds **1**–**6**; hence, they are high-affinity ligands and their potent antagonist effects on nicotine-induced antinociception at nAChRs (Ondachi *et al.*, 2016) is likely to be a consequence of being bioavailable and able to outcompete nicotine (Table 2).

Secondly, the conformation of loop C appears to be relevant. In general, this loop displays three states. There is a fully contracted, closed and clamped conformation when binding small agonists (for example nicotine or epibatidine), a fully extended and open form in the presence of larger antagonist ligands (for example strychnine) and an intermediate state that is stabilized in the presence of partial agonists (for example varenicline). However, the conformation of loop C does not always correspond in such a simple fashion to the pharmacological profile of a ligand. The noteworthy exception is the high-potency antagonist dihydro-β-erythroidine, which induces a loop C conformation similar to that of agonists (Shahsavar *et al.*, 2012).

In a study of the drug varenicline, Billen *et al.* (2012) drew attention to a third factor in ligand response. Partial agonists can desensitize pLGICs with the ability to induce opening with higher affinity but with lower efficacy than a full agonist. By targeted site-directed mutagenesis and electrophysiology on α4β2 nAChR they showed that residues on loops D and E, the complementary side of the orthosteric site, contribute to desensitization and channel opening. In particular, interactions of β2 nAChR residues Trp82 and Leu146 are important for channel opening. In *Ac*AChBP these correspond to Tyr72 on loop D and Ile135 on loop E (Fig. 4). These residues are conserved and interact with compounds **1**–**3**, and by implication also with compounds **4**–**6**. Compounds **1** and **3** represent an anomaly, being agonists of α7 nAChR (Ondachi *et al.*, 2016). The loop C conformations we observe are similar

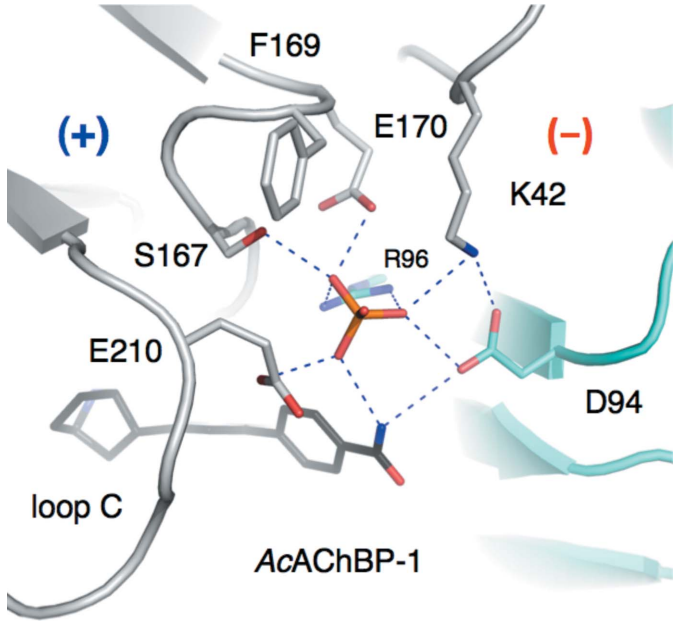


Figure 5
Oxyanion binding at the entrance to the orthosteric site. The polypeptide is shown in ribbon format coloured grey for the principal subunit, which is also marked (+). The complementary side (−) is coloured cyan. Specific residue side chains are shown as sticks with C-atom positions coloured according to the subunit to which they belong, O atoms in red, N atoms in blue, the P atom in orange and C atoms of compound **1** coloured black. Blue dashed lines represent potential hydrogen-bonding interactions.

and in an intermediate state. The *Ac*AChBP complex structures with compounds **1** and **3** suggest van der Waals interactions with Val125 and Met133, which are on loop E. These two residues differ in the α7 and β2 nAChR forms (Fig. 4). In β2 Val125 and Met133 correspond to a valine and a phenylalanine, respectively. In the α7 form they are a leucine and glutamine, respectively, and we speculate that different interactions between the ligands and this part of the orthosteric site may be linked to the agonist property of compounds **1** and **3**.

Allosteric control is an important facet of pLGIC function and has been studied extensively in nAChRs (Taly *et al.*, 2014; Chatzidaki & Millar, 2015). Of note is the role of Ca^{2+}, which increases the affinity for agonists and potentiates their activity on nAChRs, producing an increase in current amplitudes. A number of residues at distinct sites on nAChR structures are implicated in Ca^{2+} binding, and this is indicative of multiple sites of regulation (Galzi *et al.*, 1996). In this context, the identification of oxyanion binding in two of the structures reported here, involving charged and conserved residues at the subunit interface near the orthosteric site and which bring the key loop C into play, is intriguing (Fig. 5). These residues have not been previously discussed or investigated in the context of Ca^{2+} binding or allosteric regulation to the best of our knowledge. It may seem counterintuitive to invoke cation binding to a site where an oxyanion is found. However, our crystals were grown in the presence of phosphate and at acidic pH. The bound phosphate may be an artefact of these conditions, but nevertheless draws attention to a group of charged residues that are able to interact with each other or

with an ion, which is exactly the type of feature that is likely to be involved in allosteric regulation.

Asp94 and Glu170 in *Ac*AChBP, which as mentioned are likely to be protonated, are replaced by residues which are well established as phosphate-interacting amino acids in nAChR sequences. In human $\beta2$, $\alpha4$ and $\alpha7$ nAChR, Asp94 corresponds to Lys104, Ser110 and Thr99, respectively, whilst Glu170 corresponds to Glu182, Lys188 and Ser177, respectively. These are adjacent to the strictly conserved Arg96 (Arg110 and Arg101 in $\beta2$ and $\alpha7$, respectively) on the complementary side. From the principal side *Ac*AChBP Glu210, on loop C, aligns with Ser219 in the $\beta2$ form, Glu228 in $\alpha4$ and Glu215 in $\alpha7$ (Fig. 3). Note, however, that loop C in the $\beta2$ form is potentially very different from those of $\alpha4$ and $\alpha7$ (Fig. 4). Adjacent to Glu210 is Ser167, which corresponds to Asp185 ($\alpha4$), Gly174 ($\alpha7$) and Asp179 ($\beta2$). It would be too speculative to align Lys42 in *Ac*AChBP with a specific residue in $\alpha4$ or $\alpha7$ since it is in a region where, as shown by an overlay with the cryo-EM structure of an nAChR (not shown), conservation is lacking. Nevertheless, the residue types in this location would be compatible with an anion-binding site and such an event may influence the activity of compounds **1**–**6**. This site could represent a point of allosteric control to gating.

4. Conclusions

Using *Ac*AChBP as a surrogate of nAChRs, we characterized the binding affinity of a series of epibatidine analogues and were able to determine three crystal structures to inform on protein–ligand interactions. These data may inform the design of new nAChR targeting ligands with defined pharmacological properties. Differing interactions with Tyr212 and Asp94 of *Ac*AChBP appeared to mediate ligand affinities, with interactions at Asp94 potentially also explaining their subtype-selective nAChR pharmacological profiles due to the differing equivalent residues. It may be possible to modify the epibatidine framework, or indeed some other scaffold, to build interactions at a putative ion-binding site to investigate further. Data derived from compounds **1** and **3** suggest that distinction between $\alpha4\beta2$ and $\alpha7$ nAChR forms might involve two residues on loop E, the complementary side of the orthosteric binding site, and it could be instructive to test this hypothesis. Future work will seek to address these issues.

Acknowledgements

We thank Diamond Synchrotron Light Source, especially local contacts, and Dr Paul Fyfe for support.

Funding information

This work was supported by the São Paulo Research Foundation (FAPESP, Brazil) under grant 2016/16038-5 (RB), the Medical Research Council UK (studentship to SD) and National Institute on Drug Abuse Grant DA12001 to FIC.

References

Bagdas, D., Gurun, M. S., Flood, P., Papke, R. L. & Damaj, M. I. (2018). *Curr. Neuropharmacol.* **16**, 415–425.
Bertrand, D., Lee, C. L., Flood, D., Marger, F. & Donnelly Roberts, D. (2015). *Pharmacol. Rev.* **67**, 1025–1073.
Bertrand, D. & Terry, A. V. (2018). *Biochem. Pharmacol.* **151**, 214–225.
Billen, B., Spurny, R., Brams, M., van Elk, R., Valera-Kummer, S., Yakel, J. L., Voets, T., Bertrand, D., Smit, A. B. & Ulens, C. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 9173–9178.
Brams, M., Pandya, A., Kuzmin, D., van Elk, R., Krijnen, L., Yakel, J. L., Tsetlin, V., Smit, A. B. & Ulens, C. (2011). *PLoS Biol.* **9**, e1001034.
Carroll, F. I. (2004). *Bioorg. Med. Chem. Lett.* **14**, 1889–1896.
Celie, P. H., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B. & Sixma, T. K. (2004). *Neuron*, **41**, 907–914.
Chatzidaki, A. & Millar, N. S. (2015). *Biochem. Pharmacol.* **97**, 408–417.
Coe, J. W., Brooks, P. R., Vetelino, M. G., Wirtz, M. C., Arnold, E. P., Huang, J., Sands, S. B., Davis, T. I., Lebel, L. A., Fox, C. B., Shrikhande, A., Heym, J. H., Schaeffer, E., Rollema, H., Lu, Y., Mansbach, R. S., Chambers, L. K., Rovetti, C. C., Schulz, D. W., Tingley, F. D. & O'Neill, B. T. (2005). *J. Med. Chem.* **48**, 3474–3477.
Corringer, P. J., Poitevin, F., Prevost, M. S., Sauguet, L., Delarue, M. & Changeux, J.-P. (2012). *Structure*, **20**, 941–956.
daCosta, C. J. B. & Baenziger, J. E. (2013). *Structure*, **21**, 1271–1283.
Dallanoce, C., Matera, C., Amici, M. D., Rizzi, L., Pucci, L., Gotti, C., Clementi, F. & Micheli, C. D. (2012). *Chirality*, **24**, 543–551.
Dawson, A., Trumper, P., de Souza, J. O., Parker, H., Jones, M. J., Hales, T. G. & Hunter, W. N. (2019). *IUCrJ*, **6**, 1014–1023.
Dineley, K. T., Pandya, A. A. & Yakel, J. L. (2015). *Trends Pharmacol. Sci.* **36**, 96–108.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Fyfe, P. K., Dawson, A., Hutchison, M.-T., Cameron, S. & Hunter, W. N. (2010). *Acta Cryst.* D**66**, 881–888.
Galzi, J. L., Bertrand, S., Corringer, P. J., Changeux, J.-P. & Bertrand, D. (1996). *EMBO J.* **15**, 5824–5832.
Gotti, C., Clementi, F., Fornari, A., Gaimarri, A., Guiducci, S., Manfredi, I., Moretti, M., Pedrazzi, P., Pucci, L. & Zoli, M. (2009). *Biochem. Pharmacol.* **78**, 703–711.
Hansen, S. B., Sulzenbacher, G., Huxford, T., Marchot, P., Taylor, P. & Bourne, Y. (2005). *EMBO J.* **24**, 3635–3646.
Jadey, S. & Auerbach, A. (2012). *J. Gen. Physiol.* **140**, 17–28.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kouvatsos, N., Giastas, P., Chroni-Tzartou, D., Poulopoulou, C. & Tzartos, S. J. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 9635–9640.
Lemoine, D., Jiang, R., Taly, A., Chataigneau, T., Specht, A. & Grutter, T. (2012). *Chem. Rev.* **112**, 6285–6318.
Li, S.-X., Huang, S., Bren, N., Noridomi, K., Dellisanti, C. D., Sine, S. M. & Chen, L. (2011). *Nat. Neurosci.* **14**, 1253–1259.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mu, L., Drandarov, K., Bisson, W. H., Schibig, A., Wirz, C., Schubiger, P. A. & Westera, G. (2006). *Eur. J. Med. Chem.* **41**, 640–650.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nys, M., Kesters, D. & Ulens, C. (2013). *Biochem. Pharmacol.* **86**, 1042–1053.
Ondachi, P. W., Castro, A. H., Luetje, C. W., Wageman, C. R., Marks, M. J., Damaj, M. I., Mascarella, S. W., Navarro, H. A. & Carroll, F. I. (2016). *ACS Chem. Neurosci.* **7**, 1004–1012.
Rucktooa, P., Haseler, C. A., van Elk, R., Smit, A. B., Gallagher, T. & Sixma, T. K. (2012). *J. Biol. Chem.* **287**, 23283–23293.

Sauguet, L., Shahsavar, A. & Delarue, M. (2015). *Biochim. Biophys. Acta*, **1850**, 511–523.

Shahsavar, A., Gajhede, M., Kastrup, J. S. & Balle, T. (2016). *Basic Clin. Pharmacol. Toxicol.* **118**, 399–407.

Shahsavar, A., Kastrup, J. S., Nielsen, E. Ø., Kristensen, J. L., Gajhede, M. & Balle, T. (2012). *PLoS One*, **7**, e40757.

Sixma, T. K. & Smit, A. B. (2003). *Annu. Rev. Biophys. Biomol. Struct.* **32**, 311–334.

Spande, T. F., Garraffo, H. M., Edwards, M. W., Yeh, H. J. C., Pannell, L. & Daly, J. W. (1992). *J. Am. Chem. Soc.* **114**, 3475–3478.

Spang, J. E., Bertrand, S., Westera, G., Patt, J. T., Schubiger, P. A. & Bertrand, D. (2000). *Chem. Biol.* **7**, 545–555.

Taly, A., Corringer, P. J., Guedin, D., Lestage, P. & Changeux, J.-P. (2009). *Nat. Rev. Drug Discov.* **8**, 733–750.

Taly, A., Hénin, J., Changeux, J.-P. & Cecchini, M. (2014). *Channels*, **8**, 350–360.

Thompson, A. J., Metzger, S., Lochner, M. & Ruepp, M.-D. (2017). *Neuropharmacology*, **116**, 421–428.

Traynor, J. R. (1998). *Br. J. Anaesth.* **81**, 69–76.

Walsh, R. M., Roh, S., Gharpure, A., Morales-Perez, C. L., Teng, J. & Hibbs, R. E. (2018). *Nature*, **557**, 261–265.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.

Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A. & Dougherty, D. A. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12088–12093.

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

The thermodynamic profile and molecular interactions of a C(9)-cytisine derivative-binding acetylcholine-binding protein from *Aplysia californica*

Samuel Davis,[a] Hugo Rego Campello,[b] Timothy Gallagher[b] and William N. Hunter[a]*

[a]Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, and [b]School of Chemistry, University of Bristol, Bristol BS8 1TS, England. *Correspondence e-mail: w.n.hunter@dundee.ac.uk

Received 21 November 2019
Accepted 28 January 2020

Edited by N. Sträter, University of Leipzig, Germany

Keywords: acetylcholine-binding protein; crystal structure; cytisine; ligand-gated ion channel; nicotine.

PDB reference: acetylcholine-binding protein–ligand complex, 6t9r

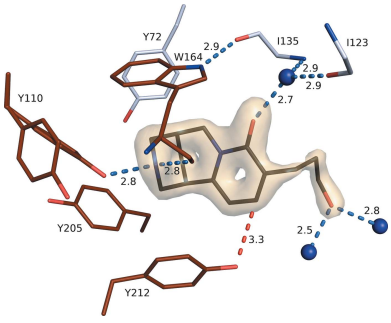
Supporting information: this article has supporting information at journals.iucr.org/f

Cytisine, a natural product with high affinity for clinically relevant nicotinic acetylcholine receptors (nAChRs), is used as a smoking-cessation agent. The compound displays an excellent clinical profile and hence there is an interest in derivatives that may be further improved or find use in the treatment of other conditions. Here, the binding of a cytisine derivative modified by the addition of a 3-(hydroxypropyl) moiety (ligand **4**) to *Aplysia californica* acetylcholine-binding protein (*Ac*AChBP), a surrogate for nAChR orthosteric binding sites, was investigated. Isothermal titration calorimetry revealed that the favorable binding of cytisine and its derivative to *Ac*AChBP is driven by the enthalpic contribution, which dominates an unfavorable entropic component. Although ligand **4** had a less unfavorable entropic contribution compared with cytisine, the affinity for *Ac*AChBP was significantly diminished owing to the magnitude of the reduction in the enthalpic component. The high-resolution crystal structure of the *Ac*AChBP–**4** complex indicated close similarities in the protein–ligand interactions involving the parts of **4** common to cytisine. The point of difference, the 3-(hydroxypropyl) substituent, appears to influence the conformation of the Met133 side chain and helps to form an ordered solvent structure at the edge of the orthosteric binding site.

1. Introduction

Nicotine (**1**; Fig. 1) is the archetypical ligand of nicotinic acetylcholine receptors (nAChRs), a family of excitatory pentameric ligand-gated ion channels (pLGICs) that contribute to the function of the human peripheral and central nervous system. This family of receptors is being studied as they may represent therapeutic targets for Alzheimer's disease, Parkinson's disease and anti-nociception (Quik & Wonnacott, 2011; Umana *et al.*, 2013; Lombardo & Maskos, 2015). Nicotine is notorious as the agent responsible for the addictive effects of tobacco smoking, which worldwide is estimated to have caused six million deaths and 150 million disability-adjusted life years in 2015, and these numbers are increasing (GBD 2015 Tobacco Collaborators, 2017).

Cytisine (**2**; Fig. 1) is a natural product widely distributed in *Cytisus* and *Laburnum* species and is the active component in the smoking-cessation agent called Tabex, which is used widely in Central and Eastern Europe. Small-scale trials comparing cytisine with other smoking-cessation therapies have suggested that it has an excellent profile in terms of cessation rate, limited side effects and economical value, and this has raised interest in the potential use of cytisine in the wider global market (Hajek *et al.*, 2013; Walker *et al.*, 2014;



West *et al.*, 2011). Cytisine, which is structurally related to nicotine, is a potent agonist at neuronal nAChRs but differs in displaying only partial agonist activity at $\alpha_4\beta_2$ nAChRs, the high-affinity nicotine subtype. It is this combination of high affinity and partial agonism which has been identified as a hallmark of clinically successful smoking-cessation agents. Cytisine, however, also has significant full agonist activity at the heteromeric $\alpha_3\beta_4$ and in particular the homomeric α_7 nAChRs, which can induce off-target effects (Coe *et al.*, 2005; Rego Campello *et al.*, 2018). It is of interest therefore to understand how cytisine and its derivatives might interact with nAChR subtypes and to use this information to guide modifications that may improve its therapeutic profile. In one such approach cytisine provided the lead for varenicline (**3**; Fig. 1), a smoking-cessation agent marketed as Champix (Coe *et al.*, 2005). The C(9)-substituted ligand **4** (Fig. 1) is a derivative of cytisine carrying a 3-(hydroxypropyl) moiety on the pyridone ring. This offers an additional set of potential interactions involving the variable complementary β-subunit, and it is this region of the cytisine scaffold that has been associated with enhanced receptor-subtype discrimination (Rego Campello *et al.*, 2018). In this study, we sought to compare the binding parameters and structural interactions of cytisine and ligand **4** using acetylcholine-binding protein (AChBP). AChBP is a soluble, highly conserved homolog of the nAChR extracellular domain (ECD) in mollusks such as the Californian sea hare (*Aplysia californica*) and can therefore be used as a surrogate for crystallography and binding studies (see, for example, Shahsavar *et al.*, 2016; Dawson *et al.*, 2019; Smit *et al.*, 2003).

2. Materials and methods

2.1. Protein production

A recombinant source of *Ac*AChBP (UniProt ID Q8WSF8) with a C-terminal Tobacco etch virus cleavage site and His_6 tag was produced in baculovirus-infected *Sf*9 insect cells using the Bac-to-Bac system (Thermo Fisher). Suspension High Five insect cells, cultured in Express Five medium plus 100 U ml^{-1} penicillin/streptomycin and 2 m*M* L-glutamine (Thermo Fisher), were used for protein production. 15×10^5 cells per millilitre were infected with 5% of the baculovirus and incubated at 27°C in shaking flasks for 48 h before being harvested by centrifugation (1500*g* for 10 min at 12°C followed by 4000*g* for 10 min at 12°C). The *Ac*AChBP was secreted out to the medium and, using a Sartojet system with a 10 kDa cutoff Sartocon Slice filter (Sartorius), the medium was exchanged for buffer *A* (50 m*M* Tris–HCl, 250 m*M* NaCl pH 7.5) and the sample was concentrated. The protein solution was applied onto a 5 ml Ni^{2+} HisTrap column (GE Life Sciences) equilibrated in buffer *A* for immobilized metal-affinity chromatography. The column was washed with 15 column volumes of buffer *A* plus 7.5% buffer *B* (50 m*M* Tris–HCl, 250 m*M* NaCl, 800 m*M* imidazole pH 7.5) and the product was then eluted over 30 column volumes using a combination of a stepped and a linear gradient of buffer *B*. Fractions were analyzed by stain-free SDS–PAGE gels (Bio-Rad) and those containing the desired product were pooled, exchanged into buffer *A* and concentrated using 10 kDa centrifugal concentrators (Pall). Protein destined for use in ITC was dialyzed against buffer *A* using SnakeSkin 10 kDa dialysis tubing (Thermo Fisher) prior to concentration.

2.2. Isothermal titration calorimetry (ITC)

ITC was carried out on a MicroCal PEAQ-ITC instrument (Malvern Panalytical). Ligands were dissolved in DMSO as 100 m*M* stocks and the concentration of DMSO was then matched in the titrant and cell solutions to minimize heat changes from buffer mismatch. Experiments utilized 12 × 3.0 µl or 15 × 2.5 µl injections, with the reference response of ligand titrated into buffer being subtracted. All experiments used a cell temperature of 25°C and the data were fitted with a one-site model using the manufacturer's software.

2.3. Crystallization

Initial crystallization trials were carried out by mixing 0.1 µl protein solution with 0.1 µl reservoir solution using MRC sitting-drop vapor-diffusion plates. A Mosquito crystallization robot (TPP) was used to prepare the drops, utilizing the JCSG-*plus* and Morpheus crystal screens (Molecular Dimensions), with protein solutions of 4 and 10 mg ml^{-1} *Ac*AChBP incubated with cytisine (3 m*M*) or ligand **4** (5 m*M*) for 1 h at room temperature. The synthesis of **4** will be described elsewhere, but is based on the use of enantiomerically pure N-Boc 9-bromocytisine as a substrate for a Pd(0)-mediated cross-coupling to an alkyl boronate ester (Rouden *et al.*, 2014). Seeking to increase the number of hits and improve the quality of the crystals, random matrix microseeding was implemented as described by Shaw Stewart *et al.* (2011). Seeded plates used a protein solution of 12.5 mg ml^{-1} plus 6 m*M* ligand **4** in 0.6 µl drops (0.3 µl protein solution plus 0.2 µl reservoir plus 0.1 µl seed-stock mixture). Subsequent

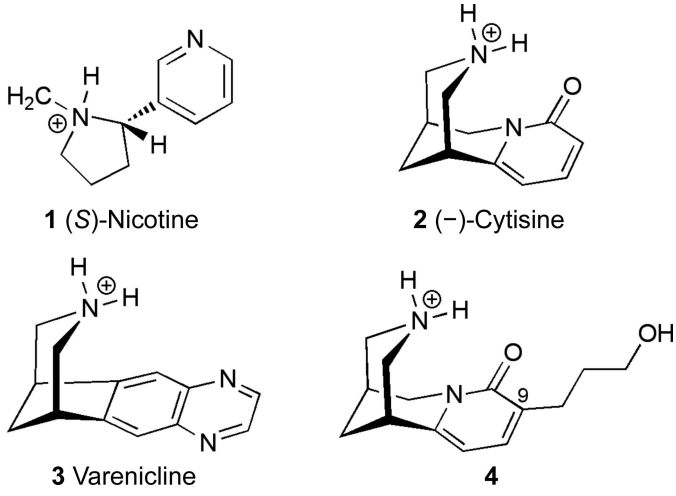


Figure 1
The chemical structures of nicotine {**1**; 3-[(2*S*)-1-methyl-2-pyrrolidinyl]-pyridine}, cytisine [**2**; {(1*R*,5*S*)-1,2,3,4,5,6-hexahydro-1,5-methano-8*H*-pyrido[1,2a][1,5]diazocin-8-one}, varenicline {**3**; (1*R*,12*S*)-5,8,14-triaza-tetracyclo[10.3.1.0^{2,11}0.0^{4,9}]hexadeca-2(11),3,5,7,9-pentaene} and ligand **4** {(1*R*,5*S*)-9-(3-hydroxypropyl)-1,2,3,4,5,6-hexahydro-1,5-methano-8*H*-pyrido[1,2a][1,5]diazocin-8-one}.

optimization of the identified conditions used 24-well hanging-drop plates with 2 µl drops (all including 0.3 µl seed stock), screening the effect of drop ratios, precipitant concentration and additives (Table 1, Fig. 2*a*). The crystal used for analysis grew in a reservoir consisting of 0.8 *M* NaH_2PO_4, 0.8 *M* KH_2PO_4, 10% glycerol, 0.1 *M* HEPES pH 7.0, with the drop consisting of 1.5 µl protein solution and 0.2 µl reservoir solution.

2.4. Crystallographic analyses

Crystals were harvested with a nylon loop, cryoprotected in 20% glycerol made up in the reservoir solution and then snap-frozen in liquid N_2. Diffraction data were recorded on beamline I03 at Diamond Light Source (DLS; Table 2),

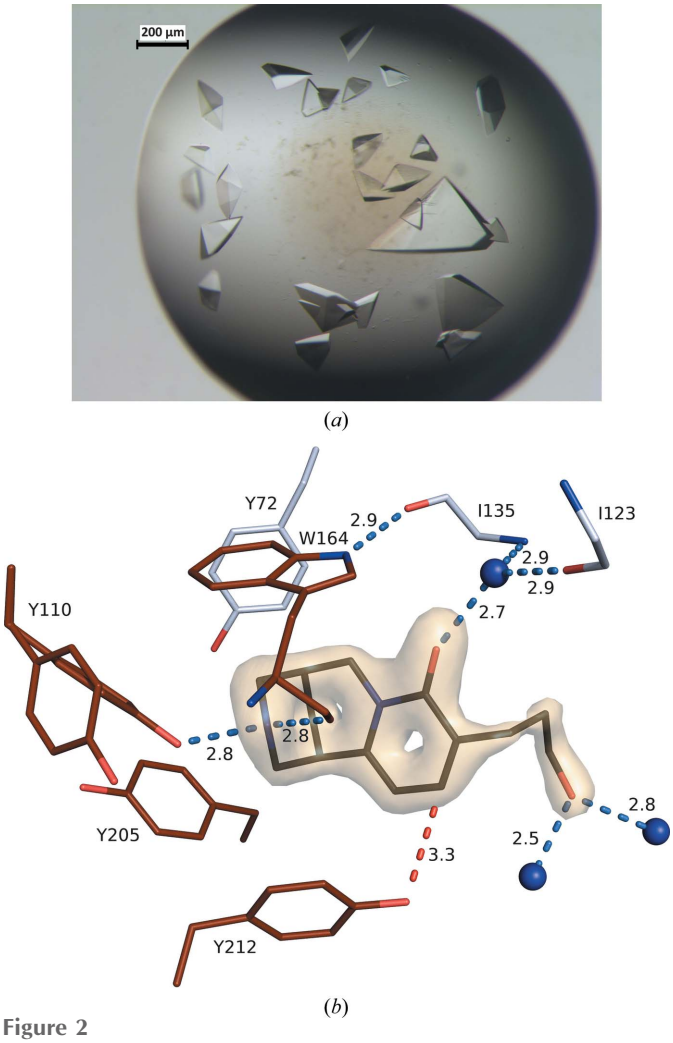


Figure 2
(*a*) Hanging drop containing crystals of the *Ac*AChBP–**4** complex. The largest crystal provided the data used for structure determination. (*b*) An OMIT difference density map for ligand **4** contoured at 6.5 standard deviations (associated with subunit *A*). C-atom positions of residues on the principal side of the binding site (subunit *A*) are shown in brown, those of residues on the complementary side (subunit *D*) are in gray and those for ligand **4** are in black. N and O atoms are colored blue and red, respectively. Three water molecules are depicted as blue spheres. Potential hydrogen bonds are identified with blue dashed lines and a C—H···O interaction is identified with a red dashed line. Two rotamers for Tyr110 are shown. Distances are in Å.

Table 1
Crystallization of the *Ac*AChBP–**4** complex.

Method	Vapor diffusion
Plate type	Hanging drop (Hampton Research)
Temperature (K)	295
Protein concentration (mg ml^{-1})	12.5
Buffer composition of protein solution	50 m*M* Tris–HCl, 250 m*M* NaCl pH 7.5 + 6 m*M* ligand **4**
Composition of reservoir solution	0.8 *M* NaH_2PO_4, 0.8 *M* KH_2PO_4, 10% glycerol, 0.1 *M* HEPES pH 7.0
Volume and ratio of drop	1.5 µl protein solution:0.2 µl reservoir solution:0.3 µl seed stock
Volume of reservoir (µl)	650

Table 2
Diffraction data-collection parameters.

Diffraction source	Beamline I03, DLS
Wavelength (Å)	0.976
Temperature (K)	100
Detector	EIGER2 XE 16M
Crystal-to-detector distance (mm)	288.18
Rotation range per image (°)	0.1
Total rotation range (°)	180

indexed and integrated with *DIALS* (Winter *et al.*, 2018) and scaled and merged in *AIMLESS* (Evans & Murshudov, 2013). The structure was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the 2.20 Å resolution structure of *Ac*AChBP in complex with an epibatidine derivative as the search model (PDB entry 6qkk; S. Davis, R. V. Bueno, A. Dawson & W. N. Hunter, unpublished work). *REFMAC* v.5.8.0257 (Murshudov *et al.*, 2011) was used for multiple rounds of automated restrained refinement, with manual refinement and model building in *Coot* (Emsley *et al.*, 2010). *MolProbity* (Chen *et al.*, 2010) was used for Ramachandran analysis. Restraints and models for ligand **4** were generated using the *Grade* server (Global Phasing; http://grade.globalphasing.org/cgi-bin/grade/server.cgi) and graphics were rendered using the *PyMOL* molecular-graphics system (Schrödinger). Two pentamers were present in the asymmetric unit and the ligand **4** was unambiguously placed in every binding site on the basis of electron and difference density (see, for example, Fig. 2*b*). The protein is glycosylated and *N*-acetyl-D-glucosamine was modeled onto Asn91 in every subunit. Strict local noncrystallographic symmetry restraints were initially applied, but were then relaxed during the course of refinement. It became evident that multiple conformations were visible for certain residues and these were manually assigned, with different occupancies being tested until the difference density maps suggested appropriate modeling. During the placement of water molecules it was apparent that ordered ions and other ligands were also present, and these were assigned as Cl^-, K^+, glycerol or phosphate. Crystallographic statistics are given in Table 3.

3. Results and discussion

3.1. Thermodynamic parameters

The ITC measurements for cytisine (**2**) binding to *Ac*AChBP (Fig. 3, Table 4) are similar to those previously

Table 3
Crystallographic statistics for the *Ac*AChBP–**4** complex.

Values in parentheses are for the highest resolution shell.

Data statistics	
Unit-cell parameters (Å, °)	a = 209.5, b = 132.9, c = 131.1, α = 90.0, β = 102.5, γ = 90.0
Space group	*C*2
Wavelength (Å)	0.976
Subunits per asymmetric unit	10
Resolution range (Å)	127.97–1.72 (1.75–1.72)
Total No. of reflections	1225936 (25553)
Unique reflections	353432 (12162)
Multiplicity	3.5 (2.1)
R_{merge}†	0.07 (0.82)
$R_{p.i.m.}$	0.064 (0.794)
Wilson B factor (Å^2)	25.07
Completeness (%)	95.5 (66.6)‡
$\langle I/\sigma(I)\rangle$	7.6 (0.6)§
$CC_{1/2}$¶	0.993 (0.435)
Refinement	
Resolution range (Å)	90.12–1.72
R_{work}/R_{free}†† (%)	16.2/19.2
No. of reflections for R_{work}/R_{free}	335977/17388
No. of protein residues	2058
No. of NAG molecules	10
No. of molecules of **4**	10
No. of phosphate molecules	10
No. of glycerol molecules	40
No. of water molecules	2899
No. of chloride ions	10
No. of potassium ions	10
R.m.s.d.s	
Bond lengths (Å)	0.013
Angles (°)	1.84
Ramachandran plot	
Residues in favored regions (%)	98.77
Residues in allowed regions (%)	1.23
Mean B factors (Å^2)	
Protein atoms (subunit *A*–*J*)	30.1/29.2/29.7/28.4/30.6/29.8/31.5/31.8/33.4/34.5
NAG molecules (subunit *A*–*J*)	93.7/90.2/108.6/87.6/99.5/97.1/104.2/103.6/100.2/92.6
Water molecules	46.0
Ligand **4** (subunit *A*–*J*)	28.1/27.9/23.5/22.2/27.9/29.9/23.7/32.4/33.3/31.5
Phosphate ions (subunit *A*–*J*)	58.1/54.2/57.4/63.8/49.2/49.5/56.4/71.9/52.6/58.0
Glycerol molecules	59.9
Chloride ions (subunit *A*–*J*)	33.7/34.6/32.0/32.9/33.5/34.0/37.9/35.2/35.9/37.8
Potassium ions (subunit *A*–*J*)	37.3/37.4/38.6/40.0/39.5/40.5/41.3/40.5/43.2/44.2
PDB code	6t9r

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all i measurements. ‡ Completeness was <70% in the highest resolution shell owing to the use of a square detector for data collection. § $\langle I/\sigma(I)\rangle$ = 2.0 at 1.90 Å resolution. ¶ Pearson correlation coefficient. †† $R_{work} = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitude and F_{calc} is the structure-factor amplitude calculated from the model. R_{free} is the same as R_{work} except that it was calculated using a subset (5%) of data that were excluded from refinement calculations.

published (Rucktooa *et al.*, 2012). For example, the values of the thermodynamic dissociation constant K_d are 0.6 (±0.3) and 1.6 µ*M*, respectively, and a similar thermodynamic profile is obtained that indicates that the binding is dominated by a favorable enthalpic contribution. In the present work ΔH was measured as −15.2 (±1.2) kcal mol^{-1}, whilst in the previous study the value was −13.3 kcal mol^{-1}. The small differences observed are likely to be owing to the use of different buffers (see, for example, Celie *et al.*, 2004). The affinity of ligand **4** is significantly reduced compared with that of **2**, with K_d values of 53.3 (±19.9) and 0.6 (±0.3) µ*M*, respectively. For both **2** and **4** there is an unfavorable entropic contribution of 6.6 (±1.4) and 3.0 (±2.8) kcal mol^{-1}, respectively, but binding is driven by gains in enthalpy, as discussed above. For **4**, although the entropic term is less unfavorable, the enthalpic contribution is reduced and therefore the affinity is lowered. This is consistent with the general reduction in binding affinity of nAChRs for cytisine derivatives possessing pyridone-ring modifications (Rego Campello *et al.*, 2018).

Ligand **4** displays a stoichiometry of approximately 4.5:1 for interaction with the pentameric *Ac*AChBP (Table 3), which is close to the expected 5:1 ratio for a ligand that fully occupies the binding site in a straightforward association. We note, however, that for **2** the stoichiometry is about 2.8:1. Our data are strikingly consistent with the previous study of *Ac*AChBP interacting with **2** (see Fig. 3B in Rucktooa *et al.*, 2012). It has previously been noted, again using ITC (Celie *et al.*, 2004), that carbamylcholine binds the pentameric *Ac*AChBP at a molar ratio of 2.5:1. The agonist carbamylcholine displays a similar affinity for *Ac*AChBP as **2**, 7.6 ± 0.4 µ*M*, and the reason for the low molar ratio is unclear. The finger of suspicion would point towards experimental issues such as the presence of impurities, uncertainty in the concentrations of the ligand and/or protein, the degradation of reagents or precipitation/aggregation (a particular problem at high concentrations). Alternatively, cooperativity or allosteric transitions can complicate the analyses. We have not uncovered any proof for such behavior in *Ac*AChBP in this or previous studies (Dawson *et al.*, 2019; Jones *et al.*, 2020) and are unable to shed further light on these observations relating to carbamylcholine or **2**.

3.2. A seeding route for suitable crystals

Attempts to directly co-crystallize ligand **4** with *Ac*AChBP were unsuccessful. Therefore, *Ac*AChBP was co-crystallized with 3 m*M* cytisine for random matrix microseeding. Small crystals grew in a drop consisting of 5 mg ml^{-1} protein and 0.4 *M* NaH_2PO_4, 0.4 *M* KH_2PO_4, 0.05 *M* HEPES pH 7.5. These crystals were crushed and used to seed drops consisting of 6.25 mg ml^{-1} *Ac*AChBP plus 3 m*M* ligand **4** and 3% DMSO. This generated crystals in the same reservoir conditions as the originating seed stock. Subsequent optimization in 24-well hanging-drop plates led to crystals with maximal dimensions of ∼400 × 400 × 300 µm in 0.8 *M* NaH_2PO_4, 0.8 *M* KH_2PO_4, 10% glycerol, 0.1 *M* HEPES pH 7.0 (Fig. 3). The drop consisted of 1.5 µl protein solution (12.5 mg ml^{-1} *Ac*AChBP, 6 m*M* ligand **4**, buffer *A*), 0.2 µl reservoir solution (0.8 *M* NaH_2PO_4, 0.8 *M* KH_2PO_4, 10% glycerol, 0.1 *M* HEPES pH 7.0) and 0.3 µl seed stock (Fig. 2*a*). High-quality diffraction data were measured using synchrotron radiation and the structure was solved and refined to our satisfaction.

3.3. Structure of the *Ac*AChBP–4 complex

The monoclinic crystal has two pentameric assemblies in the asymmetric unit. A high degree of noncrystallographic

symmetry is evident even though NCS restraints were released in the refinement calculations. Electron density for the ligand is well defined in all ten binding sites and they refined with average *B* factors less than or close to the values noted for their associated subunits (Table 3). The position of the ligand and interactions formed with the protein are highly conserved in each binding site and it is only necessary to detail one.

There are two crystal structures with cytisine in the Protein Data Bank (PDB) that are relevant to our study; a low-resolution (2.9 Å) complex with *Ac*AChBP (PDB entry 4bqt;

Table 4
Thermodynamic parameters.

Values in parentheses indicate the standard errors of the mean.

Ligand	Ligand concentration (μ*M*)	*Ac*AChBP concentration (μ*M*)	No. of sites	K_d (μ*M*)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})
Cytisine (**2**) (*n* = 2)	100	20	0.6 (0.01)	0.6 (0.3)	−8.6 (0.3)	−15.2 (1.2)	6.6 (1.4)
4 (*n* = 3)	1000	50	0.9 (0.10)	53.3 (19.9)	−6.0 (0.3)	−9.1 (2.5)	3.0 (2.8)

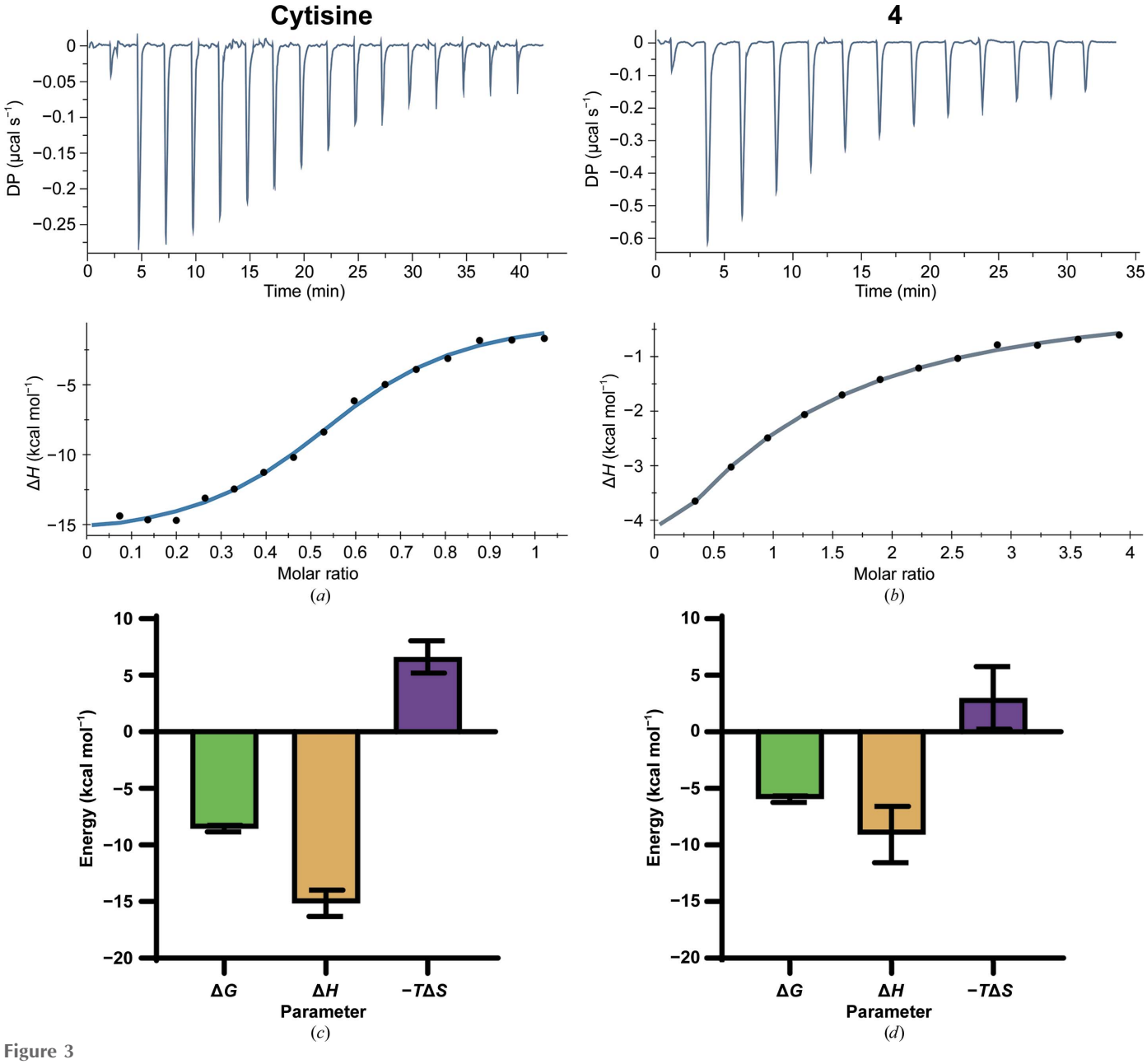


Figure 3
Raw and injection heat-normalized thermodynamic traces for the binding of (*a*) cytisine and (*b*) ligand **4** to *Ac*AChBP and (*c*, *d*) the corresponding signature plots. Error bars represent the standard errors of the mean.

Rucktooa *et al.*, 2012) and a 2.0 Å resolution structure (PDB entry 5syo; J. Bobango, J. Wu, I. T. Talley & T. T. Talley, unpublished work) in which loop C has been engineered to incorporate amino acids from the human α_3 nAChR. Also of note is the 2.0 Å resolution complex of varenicline with *Capitella telata* AChBP (PDB entry 4afg; Billen *et al.*, 2012).

The interactions of ligand **4** in the *Ac*AChBP binding site are similar to those of cytisine (**2**) and other nAChR ligands, such as varenicline (**3**; Fig. 4; Billen *et al.*, 2012; Rucktooa *et al.*, 2012; Dawson *et al.*, 2019). The Tyr110 OH group and the carbonyl backbone of Trp164 accept hydrogen bonds from the secondary amine of the ligand, indicating the presence of a protonated group. The Tyr110 OH group also donates a hydrogen bond to the carbonyl group of Ser163. The alignment and distances of between 4.6 and 4.2 Å of the aromatic groups on Trp164 and Tyr212, respectively, from the nitrogen suggest that cation–π interactions contribute to binding. Interestingly, dual rotamers of Tyr110 were modeled in seven out of ten subunits. The main rotamer, with full occupancy in three subunits and an occupancy of between 60% and 80% in the remainder, is that which interacts as described above and is shown in Fig. 4. In two cases the second rotamer is only slightly different and still participates in similar interactions. The rest, in five subunits, adopt a conformer with the OH group too far away for hydrogen bonding to the ligand. This rotamer, with occupancies of between 20% and 40%, is that observed in the apo-form binding site; for example in the three vacant binding sites of PDB entry 5syo. The orthosteric binding site of the pLGIC family possesses a degree of plasticity to accommodate ligands of differing properties, and this is noted in particular for those residues that constitute part of loop C (see, for example, Dawson *et al.*, 2019). That we also note the presence of different rotamers for the side chain of Tyr110 suggests that here also deep in the binding site there may be a degree of conformational freedom for part of the aromatic cage.

Tyr205 and Tyr212 participate in van der Waals interactions with the pyridone ring and with a potential C—H···O interaction involving the Tyr212 OH group. The pyridone C═O is about 3.5 Å from Trp164 NE1 but the geometry is not optimal for hydrogen bonding. The ligand carbonyl group accepts a hydrogen bond, with a distance of 2.7 Å, donated from a water molecule that then bridges to Ile135 (via the carbonyl) and Ile123 (via the amide) on the complementary side of the binding site. This hydration point is highly conserved in *Ac*AChBP and nAChR structures, and appears to partially mediate ligand affinity and possibly the mode of action (Billen *et al.*, 2012; Zhang *et al.*, 2012). The only available *Ac*AChBP–cytisine structure (PDB entry 4bqt; Rucktooa *et al.*, 2012) is of low resolution (2.9 Å) and does not show any water molecules in the binding site. However, the higher resolution engineered *Ac*AChBP–cytisine structure (PDB entry 5syo) with loop C altered to mimic the human α_3 nAChR structure shows a water molecule in the same position hydrogen-bonded to the cytisine C═O and the complementary side residues as per the ligand **4** complex structure.

Ligand **4** is primarily hydrophobic, and van der Waals interactions, which are a major determinant of binding, between the ring systems and the protein involve Trp164, Tyr72, Tyr205, Tyr212, Ile135 and Val165, and the disulfide between Cys207 and Cys208. The 3-hydroxylpropyl substituent interacts with Val125, Ile135, the Cys207–Cys208 disulfide, the side chain of Met133 and the main chain of Phe134. The hydroxyl residue on **4** is directed out of the binding site, forming hydrogen bonds to two well ordered water molecules, which in turn interact with other waters and glycerol. A glycerol is observed in all ten orthosteric binding sites of the asymmetric unit and occupies a highly polar pocket created by the side chains of Asp94, Arg96, Glu170 and Glu210.

The possibility of influencing affinity by virtue of exploiting waters that are able to form hydrogen-bonding networks to link the ligand to the protein was a consideration in the design of **4**. However, although the presence of such well ordered waters may contribute to the formation of a stable complex, the ITC data indicate that the 3-propylalcohol substituent leads to a reduced ligand affinity. The program *Torsion-Analyzer* (Schärfer *et al.*, 2013) was used to investigate the conformation of the 3-propylalcohol substituent and to compare it with structures in the Cambridge Crystallographic Database. The two torsion angles relevant to the conformation

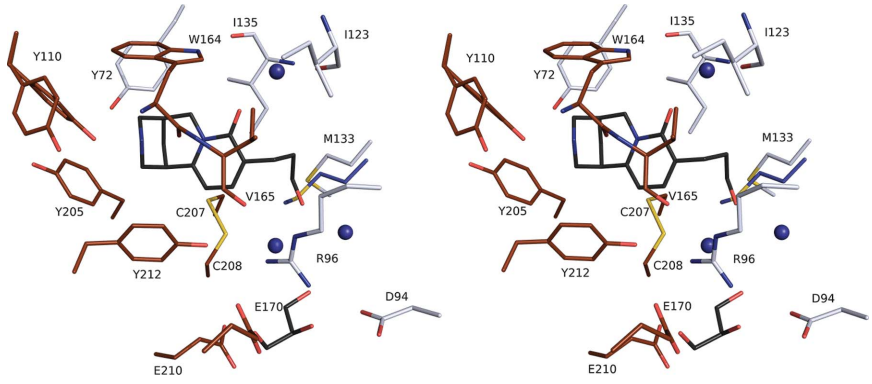


Figure 4
Stereoview of ligand **4** in a representative *Ac*AChBP binding site. In this case, the site is formed between subunit *A* (principal side) and subunit *D* (complementary side). A similar color scheme to Fig. 2(*b*) is used with the following additions: S-atom positions are shown in yellow, glycerol C atoms are shown in black and the side chain of a Met133 rotamer from a complex with cytisine (PDB entry 5syo) directed into the binding site is shown with blue C atoms.

of the aliphatic C atoms display values close to −60° for all ten ligands in the asymmetric unit. This matches the well characterized minima expected for a staggered conformation; moreover, this represents the most prevalent conformation in the database. This suggests that the ligand has adopted a preferred low-energy conformation. Nevertheless, the inclusion of the substituent has reduced the affinity for *Ac*AChBP. An overlay of the complexes with cytisine and **4** identifies structural perturbation involving Met133. In the cytisine complexes the side chain of this residue displays a rotamer that is directed into the binding site and is able to form van der Waals interactions with the ligand. In the *Ac*AChBP–**4** complex we note that the 3-hydroxylpropyl substituent appears to force the Met133 side chain to adopt different rotamers and thus avoid steric clash. This may influence the thermodynamic profile of ligand **4** compared with the cytisine by virtue of introducing a degree of strain into the protein structure coupled with a change to the hydration structure within the binding site. Of note is that in a human $\alpha_4\beta_2$ nAChR orthosteric site the residue that corresponds to Met133 in *Ac*AChBP aligns with Gln150 or Phe144 in the α_4 and β_2 subtypes, respectively. Modification of the cytisine framework to enhance activity might have to consider the likelihood of unfavorable interactions with a sizable and flexible side chain in this part of the binding site.

4. Conclusions

We have characterized the binding of cytisine and ligand **4**, a novel 9-substitued cytisine variant, to *Ac*AChBP using ITC and reported the high-resolution crystal structure of the *Ac*AChBP–**4** complex. Ligand **4** carries a 9-(3-hydroxypropyl) modification and this reduces its affinity for *Ac*AChBP. Structural comparisons indicate that the molecular interactions involving the identical components of cytisine and ligand **4** are conserved. There is only a minor perturbation of the protein structure at the side chain of Met133, where a steric clash is likely to contribute to the reduced affinity of **4** compared with the parent compound. The complex structure suggests that recovery of affinity may be possible with substituents that interact with a polar patch at the periphery of the binding site.

Acknowledgements

Author contributions were as follows. Participated in research design: SD, WNH, HRC and TG. Conducted experiments: SD and HRC. Performed data analysis: SD and WNH. Wrote or contributed to the writing of the manuscript: SD, HRC, TG and WNH. We thank the Diamond synchrotron staff, especially local contacts, Dr Tracey Gloster for recording the diffraction data, Dr Paul Fyfe and Dr Alice Dawson for advice and support, and Achieve Life Sciences for a gift of (−)-cytisine.

Funding information

This work was supported by The Wellcome Trust (grant 094090), the Medical Research Council (UK) Doctoral Training Programme (SD), the Engineering and Physical Sciences Research Council (EP/N024117/1) and a Block Allocation Grant for synchrotron access at Diamond (Reference No. 19844).

References

Billen, B., Spurny, R., Brams, M., van Elk, R., Valera-Kummer, S., Yakel, J. L., Voets, T., Bertrand, D., Smit, A. B. & Ulens, C. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 9173–9178.
Celie, P. H., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B. & Sixma, T. K. (2004). *Neuron*, **41**, 907–914.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Coe, J. W., Brooks, P. R., Vetelino, M. G., Wirtz, M. C., Arnold, E. P., Huang, J., Sands, S. B., Davis, T. I., Lebel, L. A., Fox, C. B., Shrikhande, A., Heym, J. H., Schaeffer, E., Rollema, H., Lu, Y., Mansbach, R. S., Chambers, L. K., Rovetti, C. C., Schulz, D. W., Tingley, F. D. & O'Neill, B. T. (2005). *J. Med. Chem.* **48**, 3474–3477.
Dawson, A., Trumper, P., de Souza, J. O., Parker, H., Jones, M. J., Hales, T. G. & Hunter, W. N. (2019). *IUCrJ*, **6**, 1014–1023.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
GBD 2015 Tobacco Collaborators (2017). *Lancet*, **389**, 1885–1906.
Hajek, P., McRobbie, H. & Myers, K. (2013). *Thorax*, **68**, 1037–1042.
Jones, M. J., Dawson, A., Hales, T. G. & Hunter, W. N. (2020). *ChemBioChem*, https://doi.org/10.1002/cbic.201900680.
Lombardo, S. & Maskos, U. (2015). *Neuropharmacology*, **96**, 255–262.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Quik, M. & Wonnacott, S. (2011). *Pharmacol. Rev.* **63**, 938–966.
Rego Campello, H., Del Villar, S. G., Honraedt, A., Minguez, T., Oliveira, A. S. F., Ranaghan, K. E., Shoemark, D. K., Bermudez, I., Gotti, C., Sessions, R. B., Mulholland, A. J., Wonnacott, S. & Gallagher, T. (2018). *Chem*, **4**, 1710–1725.
Rouden, J., Lasne, M.-C., Blanchet, J. & Baudoux, J. (2014). *Chem. Rev.* **114**, 712–778.
Rucktooa, P., Haseler, C. A., van Elk, R., Smit, A. B., Gallagher, T. & Sixma, T. K. (2012). *J. Biol. Chem.* **287**, 23283–23293.
Schärfer, C., Schulz-Gasch, T., Ehrlich, H.-C., Guba, W., Rarey, M. & Stahl, M. (2013). *J. Med. Chem.* **56**, 2016–2028.
Shahsavar, A., Gajhede, M., Kastrup, J. S. & Balle, T. (2016). *Basic Clin. Pharmacol. Toxicol.* **118**, 399–407.
Shaw Stewart, P. D., Kolek, S. A., Briggs, R. A., Chayen, N. E. & Baldock, P. F. M. (2011). *Cryst. Growth Des.* **11**, 3432–3441.
Smit, A. B., Brejc, K., Syed, N. & Sixma, T. K. (2003). *Ann. N. Y. Acad. Sci.* **998**, 81–92.
Umana, I. C., Daniele, C. A. & McGehee, D. S. (2013). *Biochem. Pharmacol.* **86**, 1208–1214.
Walker, N., Howe, C., Glover, M., McRobbie, H., Barnes, J., Nosa, V., Parag, V., Bassett, B. & Bullen, C. (2014). *N. Engl. J. Med.* **371**, 2353–2362.
West, R., Zatonski, W., Cedzynska, M., Lewandowska, D., Pazik, J., Aveyard, P. & Stapleton, J. (2011). *N. Engl. J. Med.* **365**, 1193–1200.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.
Zhang, H. K., Eaton, J. B., Yu, L. F., Nys, M., Mazzolari, A., van Elk, R., Smit, A. B., Alexandrov, V., Hanania, T., Sabath, E., Fedolak, A., Brunner, D., Lukas, R. J., Vistoli, G., Ulens, C. & Kozikowski, A. P. (2012). *J. Med. Chem.* **55**, 8028–8037.