On the Application of Protein-Observed NMR to Drug Discovery

Erik JB Landin

Submitted to the University of Bristol in fulfilment for the degree of Doctor of Philosophy in the School of Chemistry, Faculty of Science, September 2020.

Word Count:52972

Abstract

Since the advent of structure-activity relationship (SAR) by nuclear magnetic resonance (NMR) in the 90s, NMR has become an essential tool in drug discovery. Though ligand observed NMR is used routinely in drug discovery pipelines, protein-observed NMR is more challenging to apply, but provides more functionally relevant information on ligand binding. This thesis describes the application of protein-observed NMR to two targets with relevance to drug discovery – adenosine 2A receptor ($A_{2A}R$) and alpha-1-acid glycoprotein 2 (AGP2).

The $A_{2A}R$ is a G-protein coupled receptor that has been investigated as a target for novel Parkinson's disease treatments. These efforts have been directed towards developing novel antagonists to improve dopamine signalling, which the $A_{2A}R$ supresses. Here site directed mutagenesis and chemical conjugation are used to incorporate ^{19}F tags for NMR analysis. NMR demonstrated that a novel $A_{2A}R$ antagonist induced a distinct inactive state of the $A_{2A}R$. This study extends previous ^{19}F NMR work on GPCR activation using a combined in silico and in vitro approach to investigate alternative inactive states to integrate the technique more directly into drug discovery pipelines.

The AGP2 is a lipocalin that has a marked detrimental effect on the pharmacokinetics of a broad range of small molecules. Among them, the kinase inhibitor UCN-01, which has been investigated as an anti-cancer treatment, binds AGP2 with affinity of 3.5 nM. Here protein-observed NMR is used in conjunction with X-ray crystallography to elucidate this interaction and facilitate future efforts to re-design UCN-01, abrogating its affinity for AGP2, but maintaining its affinity for its intended target. The NMR data also revealed that a drastic structural change occurs in AGP2 during ligand binding, that may explain AGP2's promiscuity.

These studies serve to illustrate the versatility of NMR in a drug discovery context.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: 14/09/20

Acknowledgements

I wish to thank Lisa Joedicke for her exceptionally committed supervision of the work described in chapter 1 of this thesis, including the initial design of the A_{2A} construct, and intellectual guidance in the design of the experiments. I also wish to thank Lisa for being a strong driving force in publishing and reviewing material that went into this thesis. Further, I wish to thank Lisa for the time and energy she dedicated to developing me as a researcher.

I wish to thank Carl Doyle, Marco Kriek, Marianne Bergin, Neesha Dedi, Prashant Mori, Harry Mackenzie, Christine Prosser, Jeff Kennedy, Rebecca Burnley, Chris Williams, Ash Winter, Tom Burnham, Zhongshu Song, Alice Bochel, Paul Walker, Catherine Spencer and Marija Zacharova for their help and direction in wet lab techniques that I was unfamiliar with.

I wish to thank Deborah Shoemark, Katy Sutcliffe, Sofia Oliveira, Amaurys Avila Ibarra, for their help and direction in computation techniques that I was unfamiliar with.

I wish to thank Ilona Fitos and Agnes Simon for sharing the docked coordinate files of imatinib to AGP1 referred to in their published work (Fitos *et al*, 2012)

I wish to thank Rebecca Burnley for carrying out mass spectrometry fingerprinting of an A_{2A} sample as described in this thesis.

I wish to thank Sara Ryan for carrying out the protein purification of AGP2 that was used for crystallisation.

I wish to thank Nahida Akter for setting up crystallisation screens of AGP2.

I wish to thank Alice Bochel for collecting X-ray diffraction data of the AGP2.

I wish to thank Chris Williams for solving the crystal structure of the AGP2 and for operating the magnet for the 600/700 MHz NMR experiments described in Chapter 2 of this thesis and for his theoretical input.

I wish to thank Christina Redfield for running the 950 MHz NMR experiments described in Chapter 2 of this thesis and for allowing us use of the instrument at a discounted rate.

I wish to thank Zara Sands and Graham Craggs for their theoretical input.

I wish to thank David McMillan for allowing me to work in his research group at UCB Pharma.

I wish to thank Richard Sessions for supervising me and for reviewing material for publication and this thesis.

I wish to thank Richard Taylor for supervising me and for reviewing material for publication and this thesis.

I wish to thank Matthew Crump for supervising me and for reviewing material for publication and this thesis.

I wish to thank UCB Pharma for their generous funding support for my research and funding for my attendance of the 2018 GPCRs in Medicinal Chemistry conference. I also wish to thank UCB for giving me access to their world class research facilities on Bath Road, Slough.

I wish to thank the BBSRC's SWBio DTP for their funding support and training.

I wish to thank the taxpaying Great British Public for funding the BBSRC.

Lastly, I wish to thank my mother, Caroline Landin, and my fiancée, Madeleine Steer, for supporting me during this PhD.

Table of Contents

Introduction **1**
NMR as a Drug Discovery Tool 2
Ligand-Observed NMR in Drug Discovery 3
In-solution NMR of Isotopically Labelled Proteins with Bcl-2 as a Case study 8
Protein-Observed NMR of Membrane Proteins with M2 Influenza A Channel as a Case Study 15
^{19}F Tagging for Conformational Landscape Mapping with β_2-Adrenergic Receptor as a case Study 19
Aims 23
Chapter 1 – Observation of a Novel Inactive state of the A2A Receptor by 19F NMR **24**
Introduction 25
Expression and Purification of $A_{2A}R$ 34
MD Simulations of $A_{2A}R$ 36
Radioligand Binding Assays 41
^{19}F NMR of Tagged $A_{2A}R$ 46
Materials and Methods 53
Chapter 2 – The Binding of AGP2 to the Anti-Cancer Compound UCN-01 ..... **57**
Introduction 58
Trial Expression of AGP2 in Origami Cells 65
Cloning and Expression of New AGP2 Constructs 67
Optimisation of pOPINF-AGP2 Expression 73
Expression and Optimisation of Isotopically Labelled AGP2 for NMR Studies using the Refolding Strategy 77
Protein-Observed Ligand Binding NMR Experiments at pH 6.0 80
NMR Optimisation 84
Purification of AGP2-trunc for NMR Studies 87
NMR Optimisation of AGP2-trunc Construct 89
Full Backbone Assignments of AGP2-trunc 91
UCN-01 Titration 94
Relaxation Data 97
Crystal Structure of AGP2 Bound to UCN091 101
Potential UCN-01 Modifications to Abolish AGP2 Affinity 104
Conclusions 108
Materials and Methods 109
Discussion **115**
^{19}F Tagging of the $A_{2A}R$ in Context 116

Backbone Assignments of AGP2 in Context 118
Conclusion 119
Appendix **120**
References 141

Abbreviations

1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC)

1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)

2,2,2-trifluoroethanthiol (TET)

2-Bromo-N-[4-(trifluoromethyl)phenyl] acetamide (BTFMA)

5'-N-Ethylcarboxamidoadenosine (NECA)

Adenosine A_{2A} Receptor ($A_{2A}R$)

Alpha-1-Acid Glycoprotein 2 (AGP2)

Amitriptyline (AMT)

Bcl-2 homology 3 (BH3)

Beta-mercaptoethanol (BME)

Bromotrifluoroacetone (BTFA)

Carr-Prucell-Meiboom-Gill (CPMG)

Checkpoint1 (Chk1)

Chloropramazine (CPZ)

Cholesteryl Hemisuccinate (CHS)

Disopyramide (DSP)

Electrospray Ionisation Mass spectrometry (ESI-MS)

Ethylenediaminetetraacetic Acid (EDTA)

Fast protein liquid chromatography (FPLC)

Flow Through (FT)

Flow through (FT)

G-Protein Coupled Receptor (GPCR)

Green Fluorescent Protein (GFP)

Heteronuclear Single Quantum Coherence (HSQC)

Histamine binding protein (HBP2_RHIAP)

Human Embryonic Kidney (HEK)

Imatinib (IMT)

Immobilised Metal Affinity Chromatography (IMAC)

Intracellular Loop 3 (ICL 3)

Lauryl Maltose Neopentyl Glycol (LMNG)

Liquid chromatography/Mass spectrometry (LC/MS)

Molecular Dynamics (MD)

Molecular Dynamics (MD)

N utilisation substance A (NusA)

N-ethylcarboxamidoadenosine (NECA)

Nuclear Overhauser effect (NOE)

Orosomucoid 1 (AGP1)

Orosomucoid 2 (AGP2)

Palmitoyloleoylglycerophosphocholine (POPC)

Palmitoyloleoylglycerophosphoethanolamine (POPE)

Paramagnetic relaxation enhancement (PRE)

Protein Kinase A (PKA)

Retinol binding protein (RBP)

Saturation transfer difference (STD)

Seven-transmembrane (7-TM)

Solid state NMR (ssNMR)

Structure activity relationship (SAR)

Target immobilised NMR screening (TINS)

Tobacco Etch Virus (TEV)

Transverse relaxation-optimised spectroscopy (TROSY)

Twin-Streptactin binding (tStrep)

Wild-type (wt)

Introduction

NMR as a Drug Discovery Tool

Nuclear magnetic resonance (NMR) spectroscopy measures the spin excitation in samples containing certain nuclei. The utility of NMR has been exploited for decades across several disciplines including geology, organic chemistry, and structural biology (Timur, 1969; Arnold, Dharmatti & Packard, 1951; Wüthrich, 2001), and continues to be a cutting-edge tool in structural biology today (Alderson & Kay, 2020; Barret *et al*, 2013). NMR is particularly useful as a biophysical tool because it yields parameters describing biochemical properties whilst observing a sample non-destructively. The simplest outputs of an NMR experiment are signal intensity and chemical shift. Electrons experiencing an external magnetic field form a local magnetic field around a nucleus and this local magnetic field modifies the external field experienced by that nucleus altering the observable Larmor frequency (ω) of its spin (Keeler, 2010). The size of this change is expressed as the relative shift from a reference compound in units of ppm and it may be used to infer changes in chemical environment of an atom. Other experimental parameters include rates of re-equilibration of magnetisation (R_1 and R_2), the size of coupling between nuclei through space or bonds, and simple changes in intensity over time (Keeler, 2010).

NMR has become a routine tool for in the pharmaceutical industry to generate and characterise leads in drug discovery. NMR experiments in drug discovery pipelines may be divided into two categories: Ligand-observed NMR and protein-observed NMR. Ligand-observed NMR experiments are comparatively fast, easy to automate and analytically simple due to the lower complexity of small molecules relative to macromolecules (Villar *et al*, 2004). Experiments in this category are workhorses within drug discovery programs and some of the most common types are summarised briefly below. Protein-observed NMR is generally more resource-intensive than ligand-observed NMR. The advantage of observing proteins directly is that they are the usually the targets of drug discovery and so the information gained is more directly relevant to the therapeutic axis and less prone to misinterpretation (Li & Kang, 2017).

This thesis describes the application of protein-observed NMR to two targets relevant within drug discovery: The adenosine 2A receptor ($A_{2A}R$) and human alpha-1-acid glycoprotein (AGP). Protein-observed NMR has been applied across a wide range of targets and includes a number of experiments. To highlight the utility of different protein-observed approaches, their application to the Bcl-2, M2 channel and β_2 adrenergic receptors are described in this section. Through these examples three advantages of protein-observed NMR applied to drug discovery are described:

1. The broad applicability of protein-observed NMR to provide detailed information all the way from hit identification to lead optimisation
2. The flexibility of protein-observed NMR to observe the same target in different environments
3. The extension of the above two advantages to less tractable targets by novel methods

Ligand-observed NMR is first summarised briefly to contrast it to protein observed techniques and draw out the key advantages of each approach.

Ligand-observed NMR in Drug Discovery

Ligand-observed NMR has a great tractability advantage over protein-observed NMR, as isotopically labelled protein, which is most easily generated in bacterial expression systems, may be difficult to produce and the protein may not be soluble at high enough concentration to allow its observation (Cala, Guillière & Krimm, 2013). Typical small molecules found in the compound libraries of pharmaceutical companies contain sufficient ^{1}H spins to be observed easily by simple 1D NMR experiments (Liu *et al*, 2012). Furthermore, ^{19}F-containing compounds are increasingly being used in fragment-based drug discovery processes, thus expanding the scope of NMR applications (Norton *et al*, 2016).

Some of the most common ligand-observed NMR methods used in drug discovery are:

- Saturation transfer difference (STD)
- Water-ligand observed via gradient spectroscopy (WaterLOGSY)
- T_2 relaxation filtering
- Fluorine chemical shift anisotropy and exchange for screening (FAXS)

Saturation Transfer Difference

STD experiments, shown schematically in Figure i.1, use saturation transfer of magnetisation from the protein to a small molecule to distinguish ligands from non-binders (Mayer and Meyer, 1999).

A STD spectrum is the calculated difference between two spectra, the first in which protein spins are not saturated and a second where they are saturated. In the saturated spectrum the protein spins are excited by a series of on-resonance narrow bandwidth NMR pulses (typically ^{1}H methyl spins between 2 and -1 ppm, purple bolt in Figure i.1). Over a period of seconds, these saturated spins transfer magnetisation to neighbouring spins and will eventually transfer to proton spins of a ligand interacting with the protein. The 2 to -1 ppm region is chosen to exclude shifts that commonly occur on ligands found in fragment decks whilst allowing protein methyls to be excited (Pellechia *et al*, 2008; Harner *et al*, 2017). The fast exchange of ligands on and off the protein leads to an accumulation of saturated ligand spins in solution, which then display a lower amplitude free induction decay after a 90° pulse. The off-resonance spectrum uses selective irradiation of some irrelevant part of the spectrum away from both protein and ligand shifts. A subtraction of the on-resonance spectrum from the off-resonance spectrum then reveals which ligands associate with the protein as saturation transfer from the protein to the ligands supresses ligand spin longitudinal magnetisation leaving behind small positive peaks after a 90° pulse. Conversely, there is no effect on non-binding compounds whose signals are therefore removed in the subtraction. This technique typically uses only 0.5-3 μM target protein samples with an excess of the ligand, typically up to 100-fold, to yield a sensitive experiment (Harner *et al*, 2017). Upon its introduction, STD was rapidly applied to a variety of targets

including less tractable integral membrane proteins such as Integrin αIIbβ3 (Meinecke & Meyer, 2001), demonstrating its broad utility.

At its most basic level, an STD experiment can confirm binding or non-binding of a ligand. When applied to small molecules with sufficient observable spins some information on ligand binding pose may even be deduced (Mayer and Meyer, 2001). Spins from ligand motifs more closely associated to the protein receive more magnetisation transfer and so are more intense in the difference spectrum than other spins within the same ligand. This methodology is not, however, reliable, or fast enough to be routinely used in screening pipelines. One source of unwanted variation is that different spins within a small molecule may have significantly different longitudinal relaxation rates that may impact the magnitude of the STD effect and so yield a misleading epitope map (Yan *et al*, 2003).

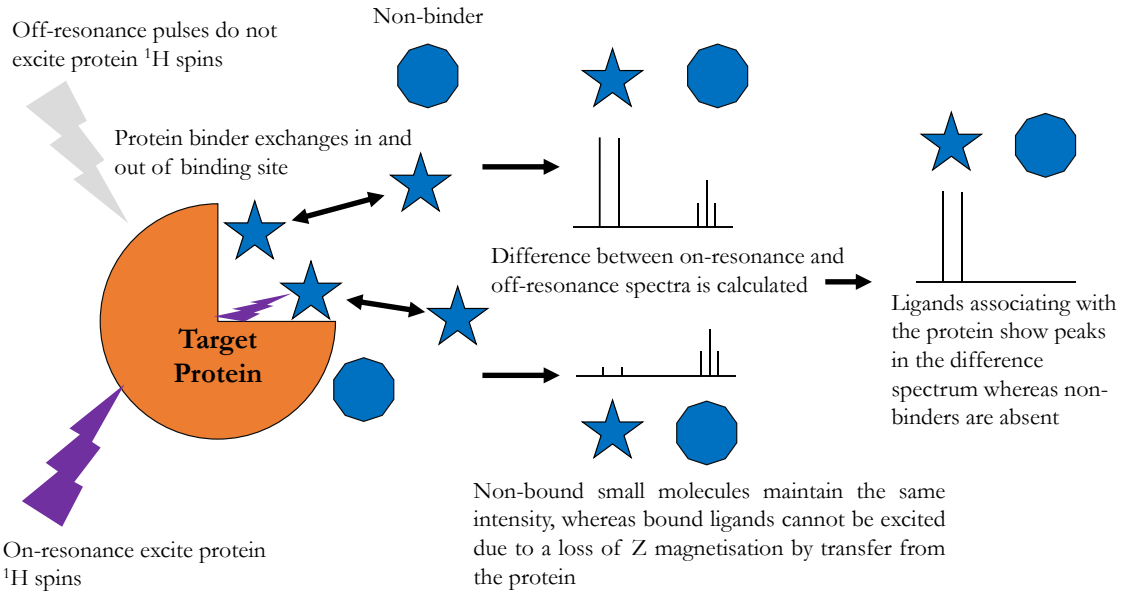


Figure i.1 Schematic of STD experiments to detect weak binders to target proteins. This method is suitable for fragment screening detecting either ^{1}H or ^{19}F spins and is commonly used.

A limitation of STD is that it cannot be applied to ligands that are not soluble enough to be observed in an excess ratio to the protein. Ligands must also have affinities to the protein of between 10^{-3} M and 10^{-8} M, as weaker binders will not be on the protein for enough time to allow magnetisation transfer and tighter binders will not exchange fast enough into solution to allow magnetisation transfer to spread across the entire ligand population (Mayer and Meyer, 1999). The technique is nonetheless widely employed to screen fragment libraries for initial binders as fragments are typically more soluble and have lower affinity than drug-sized molecules.

WaterLOGSY

WaterLOGSY (Shown schematically in Figure i.2.) is an alternative difference spectrum method which uses magnetisation transfer from ^{1}H spins in water to small molecules (Dalvit *et al*, 2000). As with STD experiments the final output is a difference spectrum in which the off-resonance spectrum is subtracted from the on-resonance spectrum.

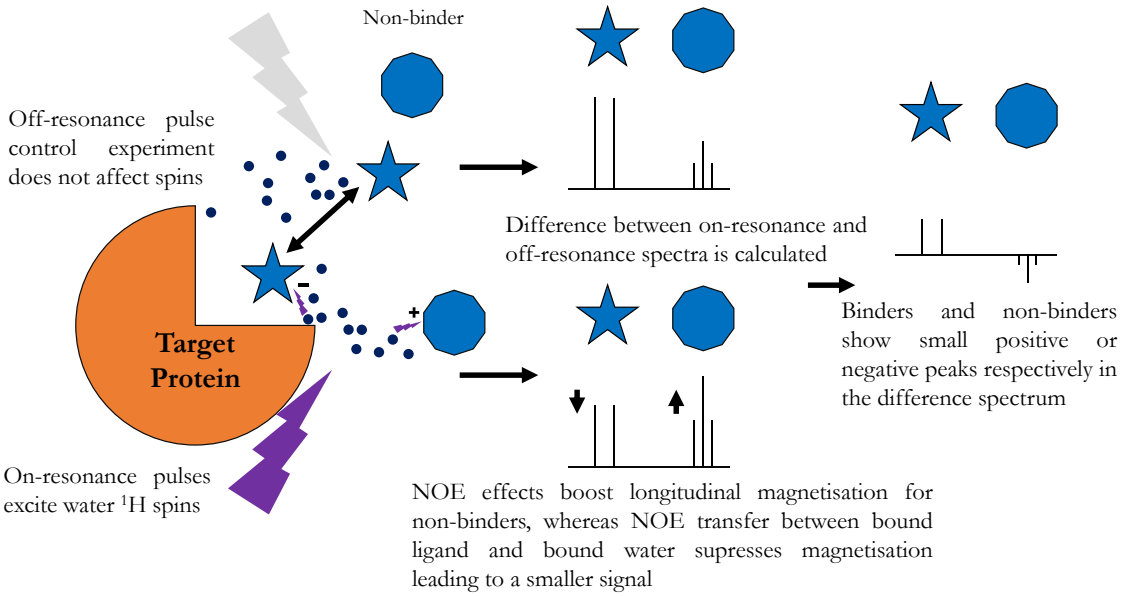


Figure i.2 Schematic representation of fragment screening by WaterLOGSY. Excitation of ^{1}H spins in water causes NOE transfer to protein-bound ligands from protein-bound water which supresses longitudinal magnetisation and yields smaller signals, whereas transfer from free water to unbound ligands increases longitudinal magnetisation due to the faster tumbling rate and so the resultant ligand peaks are larger. Therefore, protein binders yield small positive peaks in the difference spectrum whereas non-binders yield negative peaks.

Prior to excitation of the small molecule spins bulk water is saturated in the on-resonance spectrum or not in the off-resonance spectrum.

Negative NOE enhancement between protein-bound water molecules and protein-bound ligands reduces the population of excitable ligand spins at equilibrium, due to the slow tumbling of the complex. Conversely NOE effects from free water to solvated small molecules increases the proportion of excited spins that are then observed in the resultant 1D spectrum. The difference spectrum from the off-resonance control experiment and the on-resonance spectrum gives negative peaks for non-binders and positive peaks for binders.

This method has many of the same advantages and drawbacks as STD and may be run with similar concentrations of small molecules and protein (Harner *et al*, 2017), but with the added convenience of the non-hits being observable in the final spectrum. This can help rule out a supposed hit being due to broadening of a small molecule signal by some other mechanism.

WaterLOGSY has also been reported to enjoy an advantage over STD in sensitivity (Antanasijevic, Ramirez & Caffrey, 2014), although STD may still be advantageous as it reports directly on magnetisation transfer from target to ligand rather than from associated water.

T_2 Relaxation Filtering

The T_2 relaxation rate of a spin changes with the tumbling rate of that spin in solution. This may be exploited to screen ligand binding events to proteins by a method shown schematically in Figure i.3 (Dalvit, 2009).

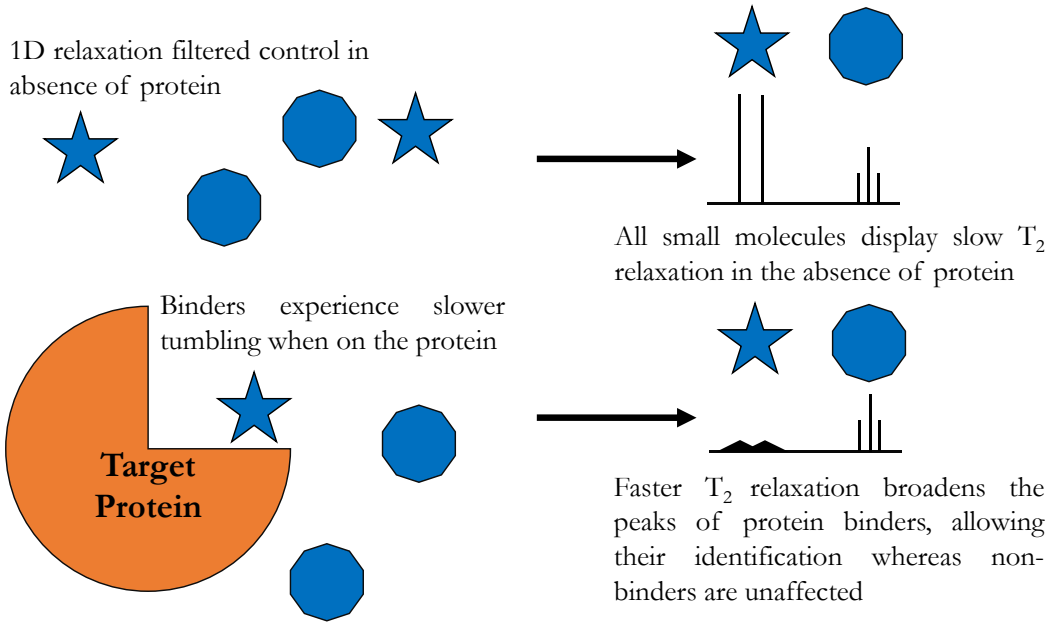


Figure i.3 Schematic representation of fragment screening by T_2 relaxation filtering. The experiment requires two samples with a loss of signal intensity and increased broadening being indicative of a binder

A delay containing a Carr-Prucell-Meiboom-Gill (CPMG) pulse train (Carr & Purcell, 1954; Meiboom & Gill, 1958) between initial excitation and signal observation allows spins with fast relaxation rates to relax to equilibrium and so are reduced or not observed in the test spectrum. This method requires a control in the absence of protein for each test to ensure that ligand signals are broadened/removed significantly relative to signals from solvated ligand.

Variations on T_2 relaxation filtering include target immobilised NMR screening (TINS). This method is employed to reduce the target material consumption of a screening campaign by tethering the target to solid material within the NMR tube, allowing the target protein to be washed and re-used (Vanwetswinkel *et al*, 2005). As in the standard un-tethered methodology the broadening effect of a slowly tumbling complex leads to a reduction in intensity of protein-bound small molecules due to an increased transverse relaxation rate. This effect is intensified when the tumbling of target proteins is restricted further by tethering to a solid substrate within the NMR sample. This method was applied to the $A_{2A}R$ which is the subject of Chapter 1 of this thesis (Chen *et al*, 2012).

The advantages of T_2 relaxation filtering are that it may be carried out with concentrations of ligand and protein as low as 10 μM (Dalvit, 2009) in each experiment, which allows the technique to be applied to less soluble ligands, and it may be applied to stronger binders than saturation transfer or NOE experiments.

FAXS

FAXS (described schematically in Figure i.4) circumvents several limitations of the other ligand-observe techniques by observing ^{19}F in a known binder that acts as a "spy molecule". This molecular probe can then be monitored for changes in behaviour if it is competed out of the protein binding site by the addition of a second ligand (Dalvit *et al*, 2003).

A FAXS experiment is a ^{19}F-detection experiment run by the same principle as T_2 relaxation filtering; experiments consist of 90° excitation pulses followed by a time delay containing a refocusing CPMG pulse train. The time delay between excitation and observation of ^{19}F spins on the spy molecule removes fast-relaxing signals. Chemical shift anisotropy makes a major contribution to ^{19}F transverse relaxation resulting

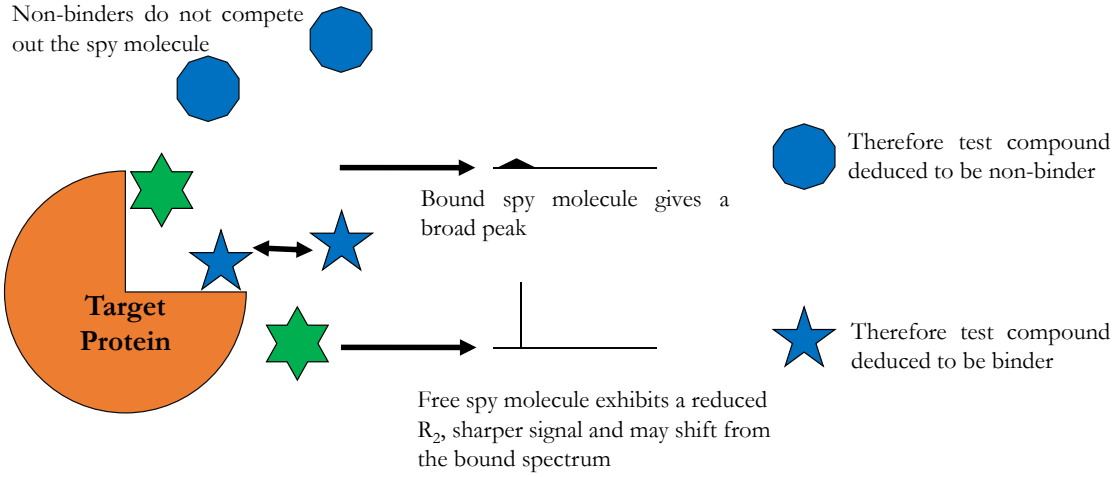


Figure i.4 Schematic representation of fragment screening by FAXS. In the absence of fragment hits the spy ligand is bound to the target and its signal is broadened. A hit is observed by the reappearance of the spy ligand signal.

in significant line broadening when the tumbling rate of the spin slows down, such as when the ligand is in complex with a protein. This effect may be strong enough to entirely remove the peak of the bound spy molecule in a 1D spectrum.

To determine if a test compound binds the protein T_2-filtered signal of the spy molecule in the presence and absence of test compound are collected. The re-appearance of the spy-molecule peak as it is competed out of the ligand binding sight is followed to report on binding events by the ligands being screened.

Observing a spy molecule removes some limitation from ligand properties that may vex other experiments. Ligands prone to non-specific protein binding are not observed as they do not compete the spy molecule

out of the specific binding site. It also shares the advantages of T_2 relaxation filtering in allowing tight binders to be observed and, as it observes ^{19}F spins, eliminates problems with signal overlap.

On the other hand the observation of a specific binder removes the possibility of finding or probing allosteric hits in sites outside of the known orthosteric binding site and limits the possibility of screening fragment cocktails as spy molecule displacement by a single binder may not be distinguished from displacement by multiple binders. It is of course also necessary to possess a known fluorine-containing ligand for the target molecule.

Shared Problems with Ligand Observed Approaches

Although ligand-observed NMR experiments provide efficient methods for screening small molecules for binding to a target, they all have certain limitations. Other than in specific and very favourable circumstances they do not give precise information on ligand binding mode. As noted, STD experiments with ligands that have multiple observable groups can give information on orientation and FAXS by its nature reports on a binding to a particular site. This information, however, falls short of atomic resolution 3D mapping of a binding site that is necessary for rational lead optimisation. The compound decks must also be carefully curated to exclude compounds that aggregate or associate non-specifically with the protein as all the experiments described above other than FAXS are prone to report such compounds as false positives.

To obtain maximum value from NMR as a drug discovery technique protein observation must be employed to transcend the limitations of ligand observed NMR.

In-Solution NMR of Isotopically Labelled Proteins with Bcl-2 as a Case Study

In solution observation of isotopically labelled proteins is a well-established methodology in structural biology with the first NMR structure of the bull proteinase inhibitor IIA published over 30 years ago (Williamson, Havel & Wüthrich, 1985). Observation of the protein target itself is a significantly more powerful approach than ligand-observed NMR in drug discovery, applicable from hit identification through to lead optimisation. This point is illustrated well by work on the Bcl-2 protein family.

The Bcl-2 family members are a set of proteins containing a variable number of Bcl-2 homology 3 domains (BH3) (Kale, Osterlund & Andrews, 2018). These proteins form a signalling network based on BH3 domain dimerization that controls apoptosis. As evasion of apoptosis is a hallmark of cancer anti-apoptotic members of the Bcl-2 family are often upregulated in tumour cells and are therefore a target for small molecule inhibitors (Kale, Osterlund & Andrews, 2018).

HSQC in Drug Discovery

Bcl-2 itself was the target of an extensive fragment screening effort using heteronuclear single quantum correlation (HSQC) monitoring chemical shift perturbation (CSP) using a minimal shift analysis (Petros *et al*, 2010). An HSQC spectrum encodes the chemical shifts of two covalently bonded nuclei. The chemical

shift of the indirectly observed, less sensitive nucleus (typically ^{13}C or ^{15}N) and the signal of the directly observed nucleus (usually ^{1}H) generate a two-dimensional plot described schematically in Figure i.5. ^{1}H-^{15}N HSQC experiments are a common variant of HSQC in structural biology as each protein may ideally yield a predictable and manageable number of well-resolved peaks based on its primary structure, one signal for each backbone amide with the exception of proline. ^{1}H-^{13}C HSQC spectra are also employed, but are generally less well-dispersed unless ^{13}C labelling of specific groups is used (Williamson, 2013). Dispersion of the signals in the 2D plane is determined by the secondary and tertiary structure of the protein in addition to other factors such as pH, temperature and buffer used. Chemical shift is dependent on the chemical environment of a spin, and so the binding of a small molecule ligand close to a protein ^{1}H-Heteroatom pair causes a change in the chemical shifts of those spins known as chemical shift perturbation (CSP). CSPs can arise directly at the ligand binding interface and also from conformational changes induced elsewhere in the protein. CSPs therefore readout on two important mechanisms of action: Orthosteric mechanisms and allosteric mechanisms. Orthosteric binding is the direct blocking of the natural protein binding epitope, the BH3 cleft in the case of the Bcl-2 family (Petros *et al*, 2010). Allosteric binding is the distal binding of a ligand causing a conformational change in the protein's functional site (Manley & Loria, 2012). In fragment screening backbone chemical shift assignments of the ^{1}H-^{15}N HSQC spectrum for the apo protein are typically sufficient. CSPs can then be mapped onto the protein sequence or structural model by assuming that the peaks in a ligand-bound spectrum with the minimum shift from a peak in the assigned apo spectrum correspond to the same residue. There are several methodologies by which CSPs are categorised as significant or insignificant; the application of these methodologies may vary between different research groups. The decision on setting a threshold value for significance may be considered a trade-off between

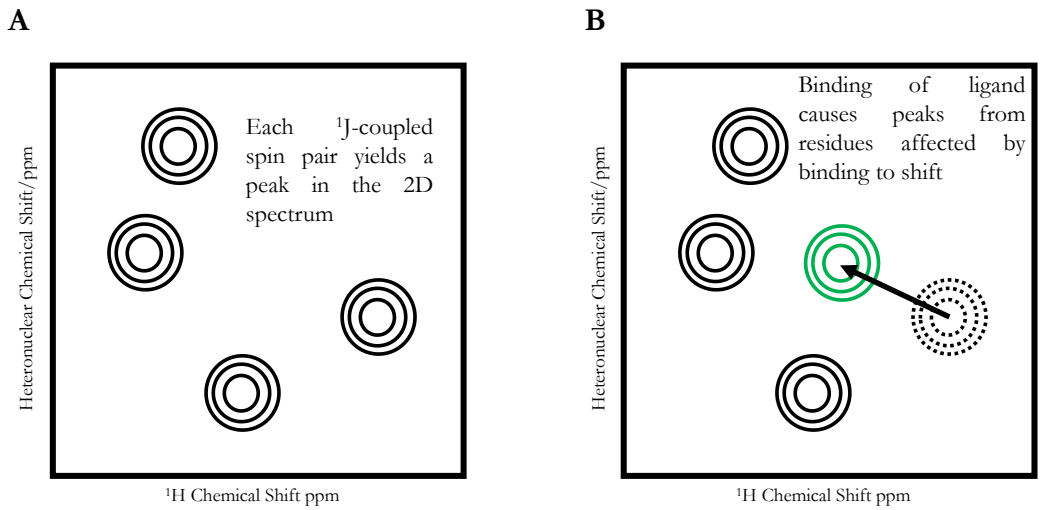


Figure i.5 (A) Apo protein yields a peak for each ^{1}H-Heteroatom pair environment in the sample. The number of peaks may be predicted from the protein primary structure and their identities may be determined from triple resonance experiments. **(B)** An HSQC may be used to screen and characterise the binding of small molecules to the protein targets. By comparing the chemical shifts of the protein before and after addition of small molecules it may be determined whether or not they are binders. Further, the binding site of a ligand may be identified as residues closer to it are likely to undergo more significant shifts.

sensitivity and specificity. One method is to select those residues whose peaks move by more than 1 or 2 standard deviations of the total chemical shift changes (Williamson, 2013). Another method is to iteratively calculate a standard deviation of the maximum number of peaks in a spectrum that all fall within their own standard deviation and to identify those peaks which fall outside of that value as shifting significantly (thus not raising the threshold due to some highly significantly shifted peaks)(Williamson, 2013).

CSP has advantages over previously discussed ligand-observed techniques. The fact that the protein is observed directly serves to rule out ligand aggregation artefacts that may plague STD and WaterLOGSY experiments and the binding site of a small molecule is identified concomitantly with screening. CSP also rules out non-specific association with the protein that may affect STD, WaterLOGSY and T_2 screening. Furthermore, as a screening technique CSP may immediately categorise hits as orthostatic or allosteric which ligand-observed techniques cannot do, indeed FAXS may discard allosteric binders altogether by registering allosteric binders as false negatives.

Bcl-2 binders were identified in a ^{1}H-^{13}C HSQC screen (Petros *et al*, 2010). As protein aliphatic sidechains (which are visible in a ^{1}H-^{13}C HSQC) often form the direct contacts between ligands and proteins this screening method is likely to yield more direct evidence of binding than a ^{1}H-^{15}N HSQC which monitors only backbone perturbations, but the spectra are likely to be more crowded in a uniformly labelled spectra (Hajduk *et al*, 2000). Chemical shift assignments, a prerequisite for mapping binding epitopes, were based on the authors' previous structural study that resulted in a full solution structure (Petros *et al*, 2000). Hajduk and co-workers initially screened 17000 fragment-sized small molecules and identified a potent diphenylmethane hit.

As discussed above a drawback of protein-observed NMR is the large quantity of protein consumed. This Bcl-2 screen, for example, required 50 μM protein samples, and all 17000 fragments most likely had to be screened individually. For a 5 mm NMR probe, as used in the study (Petros *et al*, 2010), 200 μl samples are likely, assuming that 3 mm NMR tubes were used to conserve material. With a 19.23 kDa Bcl-2 construct the full screen would have required more than 3 g of protein from ^{13}C-glucose labelled minimal media, even without including the material necessary for generating the structure and assignments in the first place. Such an expense may be out of reach for many academic groups, but is feasible for a project that may pay large commercial dividends. The success of this method also relies on the protein being stable in the concentration of DMSO necessary to solubilise the fragment in a large excess.

Following the identification of binders to Bcl-2, CSP was further used to determine the affinities of the fragment hits. Under conditions of fast exchange between two states (in this case bound and unbound) an NMR spectrum will represent a population-weighted average chemical shift of those states rather than resolving them as separate signals. By collecting titration data with increasing ligand concentrations the chemical shift variation may be fitted to standard ligand binding equations to determine a K_D (Fielding, 2007). This approach to measuring K_Ds is limited by the kinetics of the binding event as shown schematically in Figure i.6. If the bound and unbound states are in exchange on an intermediate timescale the broadening

effect on the peaks makes fitting an accurate titration curve difficult. Unfavourable chemical exchange timescales may also hinder structural NMR studies of ligand-protein complexes if regions of the protein are in exchange between multiple conformational states. In the case of the Bcl-2 study under consideration the affinity of the primary fragment hit was determined to be 20 μM and so this was not a limitation. However, a drug discovery effort on the related Mcl-1, a member of the Bcl-2 family, illustrates the problem as the experimental temperature had to be altered to improve the chemical exchange conditions (Caenepeel *et al*, 2018). In a robust system modifying temperature to achieve favourable chemical exchange may not be an obstacle, however when less stable targets, such as membrane proteins, are studied these modifications may not be possible and so may limit the application of NMR methods.

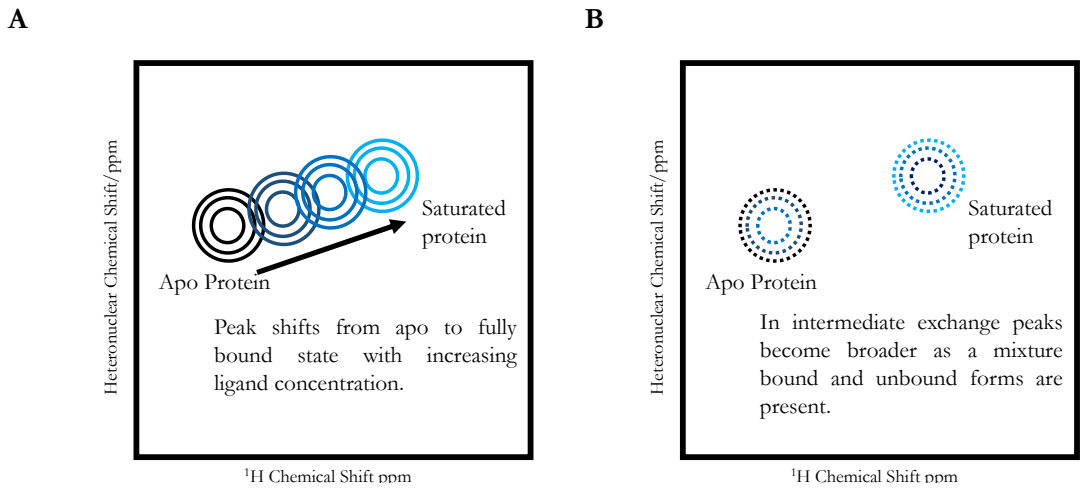


Figure i.6 (A) For binding events in fast exchange, a peak is observed at the weighted average shift between the bound and unbound form. By tracking this average peak, a titration curve may be used to determine the K_D of the ligand to the protein. **(B)** In slower exchange regimes peaks become broadened when a mixture of states is present rather than allowing the observation of an average chemical shift. This may hinder the calculation of an accurate K_D, but for very slow exchange a K_D may be determined by fitting the relative intensities of the 2 peaks to a sigmoidal curve provided the signal/noise ratio is good enough to identify the endpoints accurately.

NMR Structures in Drug Discovery

Having identified a hit to Bcl-2, NMR was employed to carry out a full structural characterisation of the binding of the diphenylmethane fragment to the protein. A solution NMR structure of Bcl-2 (modified with some Bcl-x_L residues to improve solubility) was available before the study and is shown in Figure i.7. (Petros *et al*, 2000).

Although structural determination by NMR has declined in popularity since its peak in 2007 when 965 NMR structures were deposited, it is still an important technique in structural biology with 380 structures deposited to the protein data bank in 2019 (Berman *et al*, 2000). The increasingly routine nature of X-ray crystallography may be part of the reason. NMR is perhaps best applied to those targets, or parts of targets, with particular characteristics that make them intractable to crystallography. Work on the Bcl-2 family

illustrates this as the first structures of the Bcl-2 homologue Bcl-x_L were solved by NMR and X-ray crystallography and described in a single publication (Muchmore *et al*, 1996) with the helix I/II loop observable by NMR, but not by crystallography. Unstructured loops in fast conformational exchange are often visible in NMR spectra giving rise to sharper resonances than the bulk of the protein and so may be observed in solution; on the other hand the conformational heterogeneity of loop regions hinders repeatable crystal packing necessary for their observation by crystallography. NMR is therefore best applied to solving structures of proteins whose functions require them to be dynamic in solution and undergo conformational rearrangement.

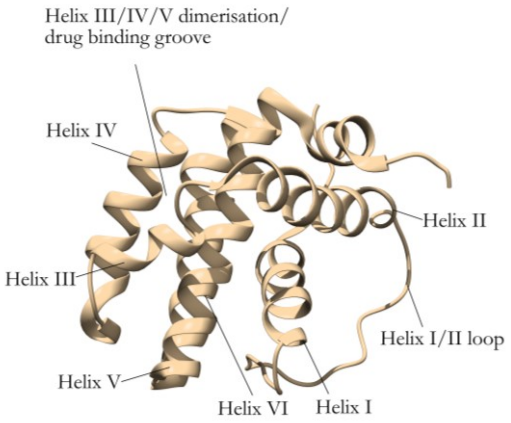


Figure i.7 NMR structure Bcl-2-Bcl-xl hybrid [PDB:1G5M] (Petros *et al*, 2000) The authors went on to use NMR to determine the binding conformations of fragment hits to Bcl-2.

Beyond the ^{1}H, ^{15}N and ^{13}C chemical shift assignment of the protein backbone and side chain resonances the main type of data needed to solve NMR structures is a set of nuclear Overhauser effect (NOE) constraints. NOE constraints for protein structural determination arise from dipolar couplings between nuclei, usually protons, that are proximal in space and not J coupled, resulting in NMR peaks in the off diagonal spectral region with intensities dependent on the proximity of individual nuclei to the power of 6 (Keeler, 2010). The form of a protein NOE experiment is shown schematically in Figure i.8. The peak

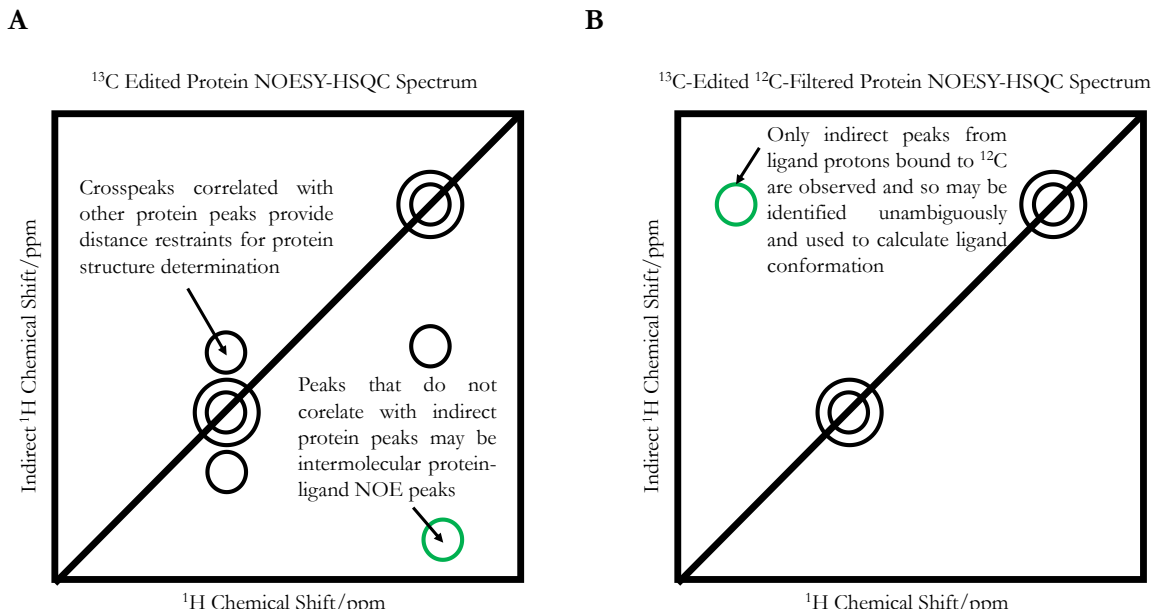


Figure i.8 (A) The structure of a protein may be calculated from distance restraints derived from ^{1}H-^{1}H intramolecular protein cross peaks, which allow the distance between two spins to be calculated. Protein-ligand NOEs appear as cross peaks that are not correlated to any observed diagonal peaks in the indirect dimension. **(B)** When ^{12}C filtering is applied to the indirect dimension crosspeaks solely between ^{1}H spins attached to ^{13}C are not observed and so protein-ligand NOE peaks between ^{1}H-^{13}C and ^{1}H-^{12}C may be identified unambiguously and used to calculate distance restraints for the protein:ligand complex.

intensities can be used to calculate a series of distance constraints between assigned spins. An ensemble of protein structures that violate the minimum number of NOE distance constraints is generated as a structural model.

Further, if the spins in a protein are assigned, NOE data sets may be used to determine the structure of a ligand-protein complex as NOE signals between the protein and the ligand appear as new peaks in the NOE spectrum and can be used to introduced distance constraints of the ligand:protein complex. Protein-ligand NOE's are often, however, more easily identified by using filtered experiments to observe the ligand ^{1}H resonances in the indirect dimension whilst removing intramolecular protein NOE peaks (Otting & Wüthrich, 1990).

Hajduk and co-workers used this methodology to identify the binding site of their diphenylmethane fragment and found it to be in a different part of the Bcl-2 binding pocket compared to known biaryl binders to the homologous Bcl-x_L (Petros *et al*, 2010). NMR was further used to generate distance restraints for a complex of both a biaryl ligand and the diphenylmethane demonstrating that both ligands could bind to Bcl-2 concomitantly.

To produce a ligand with a higher affinity for Bcl-2 the previously known biaryl and newly identified diphenylmethane fragments were joined by an alkyl linker as shown in Figure i.9. The rational assembly of fragment sized ligands, based on NMR data, to form larger molecules is a common approach in fragment-based screening as initial hits that do not have the necessary affinity for their target to act as a drug can be rapidly expanded (Harner, Frank and Fesik, 2013).

The complex of the newly designed ligand bound to Bcl-2 was solved by NMR and confirmed to occupy the two adjacent binding pockets of the individual fragments as shown in Figure i.10 overlaid with the position of the novel diphenylmethane ligand. The affinity of the Bcl-2 ligand after optimisation was 40 nM (Petros *et al*, 2010). The limitation of NMR in determining ligand affinities is highlighted by need for a fluorescence polarisation assay to determine the affinity of this last compound as the binding exchange rate would likely be slow enough to require a potentially less reliable peak intensity plot to determine the affinity.

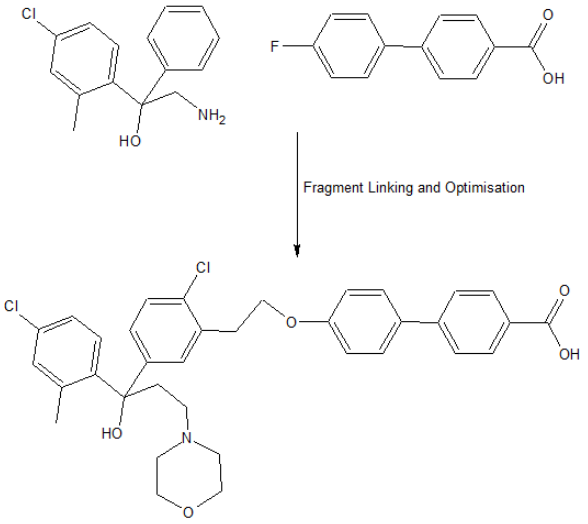


Figure i.9 The diphenylmethane compound (top left) was linked with a previously known biaryl binder (top right) to yield a more potent Bcl-2 ligand (bottom).

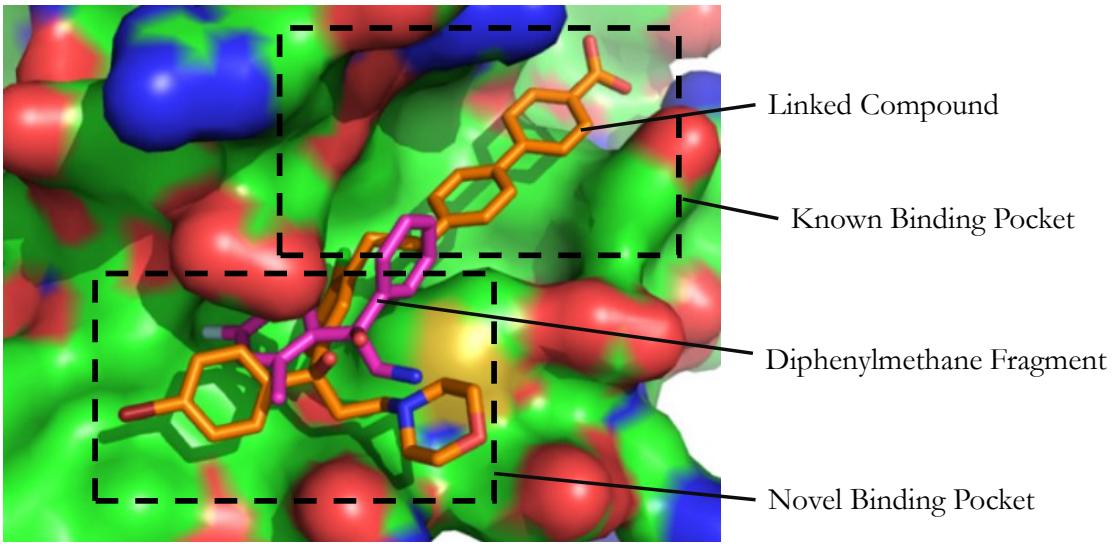


Figure i.10. The optimised compound derived from fragment linking (orange) is shown binding across both the known biaryl pocket and the novel binding pocket. The biphenylmethane moiety of the optimised compound binds in a similar position to the original biphenylmethane ligand (purple) (Adapted from Petros *et al*, 2010).

Before and since this work Abbott has developed other ligands that target the Bcl-2 family of anti-apoptotic proteins. The experimental drug Navitoclax (shown in Figure i.11.), which inhibits Bcl-2 and Bcl-x_L, has been the subject of 16 completed clinical trials and 5 ongoing ones at the time of writing according to the U.S. National Institute of Health (clinicaltrials.gov). Hajduk and co-workers expressed their hope that the compound derived from the described structure activity relationship (SAR) by NMR approach would inform future efforts to develop anti Bcl-2 compounds (Petros *et al*, 2010).

Although the described NMR techniques were applied very effectively to Bcl-2, it should be noted that the Bcl-2 constructs used in these studies were small and soluble, making them well-suited to these NMR techniques. Many targets of drug discovery do not fall within these parameters and so alternative NMR methods must be applied. Membrane proteins, for instance, are not soluble without associated amphipols and so NMR techniques must be adapted for drug discovery studies on these targets. An example of such a protein investigated using alternative NMR techniques is described in the next section.

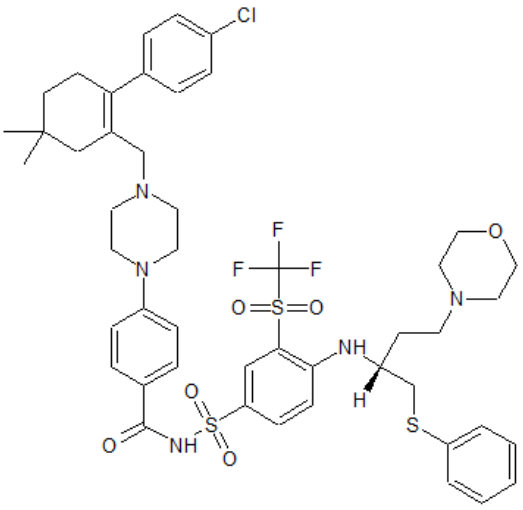


Figure i.11. Structural representation of the candidate oncology drug Navitoclax (ABT 263).

To summarise: In solution observation of proteins is a versatile technique across the entire span of a drug discovery project from screening, characterisation, and final optimisation. The

increased time and resource cost relative to ligand-observed NMR means that protein-observed techniques are often reserved for validation and characterisation of initial hits. However, as this Bcl-2 study demonstrates, protein-observed NMR can be applied to both fragment-sized and drug-sized compounds to generate a valuable output, if the target is tractable to the relevant NMR techniques.

Protein-Observed NMR of Membrane Protein Targets with M2 Influenza A Channel as a Case Study

Drug discovery efforts must often target proteins that are significantly less tractable than Bcl-x_L or Bcl-2. In these cases, the greater flexibility of NMR relative to crystallography in terms of experimental conditions becomes important. Structural biology studies of the influenza A M2 proton channel transmembrane domain illustrate this point.

Membrane proteins include some of the most attractive drug targets in medicine. These proteins are inherently less tractable to biochemical and biophysical techniques than their soluble counterparts due to their extensive hydrophobic transmembrane domains that must be associated with lipids or lipid mimics to maintain the conformational stability, and therefore functional status, of the protein. Membrane proteins are targeted by an extensive range of drugs many of which, such as bisoprolol and morphine, are included in the World Health Organisation's list of essential medicines (World Health Organisation, 2019). It is therefore inevitable that further study of these difficult protein targets must be undertaken to develop new drugs and better understand the mechanisms of existing therapeutics.

The pseudo tetrameric M2 channel from influenza A virus is the target of antiviral drugs to treat influenza, reducing the duration of symptoms and minimising complications (Shen, Lous and Wang, 2015). During cell infection, the influenza virus triggers invasion of the cytoplasm by an internal drop in pH mediated by the M2 proton channel. The adamantane-based drugs for treatment of influenza inhibit the M2 proton channel thus preventing the virus from infecting the cell. The admantane drugs amantadine and rimantadine (shown in Figure i.12) were used as treatments for influenza. The evolution and almost universal spread of mutant influenza viruses resistant to these drugs necessitates the design of new compounds to treat resistant strains (Shen, Lou & Wang, 2015). Rational modification of amantadine and rimantadine necessitates an atomic-level understanding of their mechanism of action, which has been provided by structural studies.

The fact that M2 is a membrane protein complicated crystallisation efforts. Detergent micelles used to solubilise membrane proteins do not form consistent crystal contacts easily and may occlude hydrophilic parts of the protein that are able to mediate such interactions. Therefore, careful detergent selection or extensive screening is necessary. Perhaps as a result of this, the first structural model [PDB: 1NYJ] of M2 was determined by solid state NMR (ssNMR) and is shown in Figure i.13. (Nishimura *et al*, 2002).

ssNMR samples constitute a form of the target protein that is fixed in solid rather than being solubilised. An advantage of this technique for membrane proteins is that the proteins may be prepared in lipid bilayers which more closely replicate physiological membranes than detergent micelles do, although the sample preparation, often in anhydrous conditions, may stress the sample more than solution state sample preparation. For the initial M2 structural study 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), a lipid that spontaneously forms bilayers, was used as a membrane mimetic. The lack of in-solution tumbling in such a sample results in extreme broadening of NMR signals from the effects of CSA. Fortunately, ssNMR samples yield well-dispersed sharp spectra when spun at several thousands of kilohertz at the "Magic Angle" of 54.74° to the B_0 field which averages out the direction-dependant chemical shift anisotropy, allowing data for the determination of protein structures to be collected (McDermott, 2009). This spinning also supresses the structurally useful dipolar couplings between nuclei, but they may be reintroduced by the application of specific radiofrequency pulses to the sample to generate data for structural restraint determination.

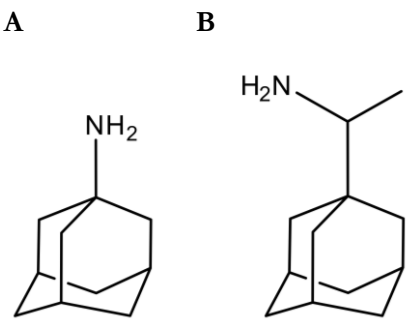


Figure i.12 Adamantane drugs act by inhibiting the influenza A virus M2 proton channel **(A)** Structure of amantadine. **(B)** Structure of rimantadine.

The derivation of structural restraints from dipolar couplings allowed a model of the M2 channel to be proposed (Nishimura *et al*, 2002). This structural model indicated that the pore of the proton channel was sufficiently large to allow the binding of amantadine, supporting this as a mechanism of drug action against the virus.

Subsequently, atomic resolution structures of the M2 tetramer in complex with adamantane drugs were determined by crystallography and solution NMR(Schnell & Chou, 2008; Stouffer *et al*, 2008). The crystal structure of the M2 tetramer contained a single amantadine molecule bound in the channel pore close to the extracellular face, as hypothesised from the ssNMR model. Conversely the solution NMR structure contained 4 rimantadine molecules bound around the internal face of the M2 channel. These alternative structures are shown in Figure i.14.

Figure i.13 Structural model of M2 protein channel [PDB: 1NYJ], based on limited set of dipolar couplings, viewed from extra-viral face. From this model, amantadine binding in the centre of the pore was hypothesised

These alternative binding modes of adamantane drugs were not necessarily in conflict with one another and were both published in the same issue of *Nature*. Subsequent

reviewers have suggested that both structures are possible and consistent with other functional data (Kozakov *et al*, 2010). The X-ray and NMR structures were solved in open and closed conformations respectively, likely due to the different conditions under which they were obtained. The crystals were generated in octyl-β-D-glucopyranoside as a membrane mimetic system with polyethylene glycol (PEG) as a precipitant at pH 5.3 (Stouffer *et al*, 2008). By contrast the solution NMR structure was solved in DHPC micelles, without PEG at a pH of 7.5 (Schnell & Chou, 2008). Differences in membrane mimetic systems between studies are a concern in the study of all membrane proteins, including the $A_{2A}R$ as will be discussed in Chapter 1.

The M2 NMR sample also contained 40 mM of rimantadine whereas the M2 was crystallised with only 0.6 mM amantadine whilst the protein concentrations were comparable at 0.75 mM and 0.5 mM, respectively. The excess ligand in the NMR structure may be necessary to drive occupancy of the putatively lower affinity binding sites around the internal base of the channel. Another possible contributing factor to the difference is the construct used in the NMR structure, which was not truncated to the same extent as the crystallography constructs (Schnell & Chou, 2008). There are, however, no subsequent NMR studies using constructs as large as that used by Chou & co-worker and so it is difficult to draw conclusions relating to this factor. The effects of differences in protein constructs have been investigated in the GPCR field as will be discussed in Chapter 1 (Eddy *et al,* 2016). The discrepancy in the M2 may also be due to the different structures of amantadine and rimantadine.

Although the binding site identified by crystallography was the higher affinity site there is value in identifying alternative sites as they may be explored in further drug discovery efforts. In this respect the flexibility of NMR is a significant advantage in being able to observe transient or less populated conformations which may be present under different conditions from those required for crystallography.

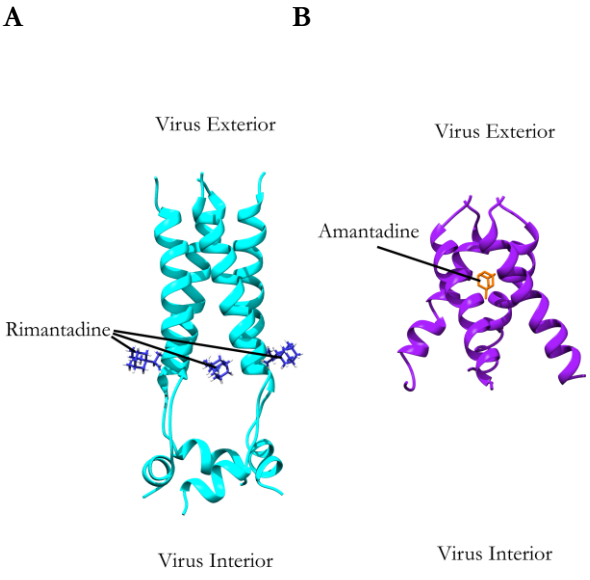


Figure i.14 (A) Solution NMR structure of M2 transmembrane domain in complex with rimantadine bound outside the helical bundle [PDB: 2RLF] **(B)** Crystal structure of M2 transmembrane domain in complex with amantadine bound in the transmembrane pore [PDB: 3C9J]

Following the initial ligand binding studies, the flexibility of NMR as a technique is highlighted by the relative frequency (compared to crystallography) with which new binding studies were published following the initial effort. In 2010 a ssNMR structure of amantadine bound in the channel pore of the M2 protein, was published, as in the crystal structure (Cady *et al*, 2010). The ssNMR study

did also observe a low affinity binding event in the presence of a large excess of amantadine towards the C-terminus of the M2 protein, as observed in the solution state structure. The 2010 structure [PDB: 2KQT] was solved in DMPC bilayers at pH 7.5 and revealed a more tightly packed helical bundle (similar to the "closed" conformation of helices shown in the 2008 solution state structure). By allowing the observation of the M2 channel at a higher pH the NMR studies provide a more complete picture of the M2's structure and interaction with its inhibitors.

Subsequent studies of novel M2 inhibitors exploited the flexibility of NMR to investigate the binding of ligands designed to overcome resistance. Resistant influenza channels contain the mutation S31N (situated close to the in-channel drug binding site) and this mutant was investigated with solution NMR (Wang *et al*, 2013; Wu *et al*, 2014). These studies determined the structures of the potential anti influenza drugs M2JW332 and a compound referred to as "compound 11" (a modified amantadine shown in Figure i.15) in complex with the mutant S31N M2 proton channel transmembrane domain.

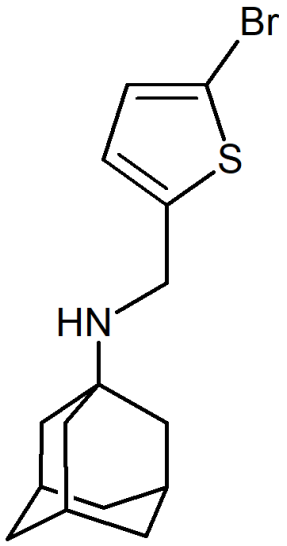


Figure i.15. Structural representation of "compound 11" – a modified amantadine solved in complex with the M2 by NMR

In illustrating the flexibility of NMR these studies are significant as no further crystal structures of the M2 channel in complex with its inhibitors were published in the 10 years between 2008 and 2018, when structures of the channel with 3 adamantanes were solved in lipidic cubic phase crystals, thus highlighting the superior flexibility of NMR in conducting follow-up studies of compounds related to the initially investigated drug (Thomaston *et al*, 2018). Although all the M2 studies discussed in this section were carried out in academia, the analogy to a commercial context highlights that in some cases rational drug design is enabled by taking advantage of the adaptability of NMR to varied experimental conditions to expedite progress where the target is less tractable to crystallography. Although, as discussed in this example, crystallographic studies may often deliver a structure of a target before or concomitantly with NMR, it may be difficult to explore the effects of alternative experimental parameters on the system. Crystal conditions must necessarily be chosen primarily by whether they generate crystals. In the case of the M2 channel, amantadine co-crystallisation conditions were screened by pH and a hit was found at a pH of 5.3 (Stouffer *et al*, 2008). On the other hand, although NMR often benefits from specific pH conditions, these may be varied in a continuous way to explore a greater range of protein environments.

In co-crystallisation studies (such as those of the M2 channel), the presence of a particular ligand itself may be essential for crystal formation, in that certain ligands promote crystallisation whereas others do not. Many apo crystal structures may in fact contain additives which may compete with ligands for the active site. Some protein crystals are sufficiently stable to allow ligands to be soaked in after the fact without

disrupting the lattice indicating that the protein has crystallised in a conformation pre-disposed for ligand binding. If a large conformational change is required to accommodate the ligand crystals may crack and not yield useable diffraction data. Further, the crystallised form may represent only one conformational state of a dynamic protein, and so provide an incomplete understanding.

High resolution NMR is more flexible with respect to studying new ligands as they may simply be added to a protein solution allowing the protein to sample its natural conformational repertoire unimpeded. A drawback of traditional NMR methods is the requirement for isotopically labelled protein and, minimally, assignment of the backbone amide resonances to specific amino acids. Novel selective labelling and tagging approaches have extended the range of protein systems amenable to NMR to new targets such as the one described in the next section.

To summarise: The advantage in adaptability allows NMR to accelerate drug discovery lead optimisation, in many cases faster than other methods, thus saving resources and leading to efficiency in drug development pipelines.

^{19}F Tagging for Conformational Landscape Mapping with β_2-adrenergic Receptor as a Case Study

Studies of the β_2-adrenergic receptor illustrate that alternatives to standard NMR approaches (standard NMR approaches relying on protein backbone assignment) widens the scope of the technique to more challenging drug discovery targets. Interest in the β_2-adrenergic receptor stems from its importance as a target of cardiovascular drugs such as carvedilol (Velmurugan, Baskaran & Huang 2019).

The specialist equipment necessary for solid-state NMR and the challenge of generating isotopically labelled material for difficult-to-express membrane proteins limits the application of NMR to difficult targets. Although chemical shift analysis has been applied to a protein complex in excess of 900 kDa (Fiaux *et al*, 2002) structural determination has generally been considered to be limited to proteins of around 25 kDa (Sugiki, Kobayashi & Fujiwara, 2017). Many drug discovery targets fall outside of this range and so many of the techniques described above are difficult to apply, either because of issues with spectral quality, such as excessive line broadening due to slow molecular tumbling, or the assignment problem associated with severely overlapping signals in uniformly labelled samples.

G-protein coupled receptors (GPCRs), which are described in greater depth in Chapter 1, are 7-transmembrane helical proteins that long eluded structural characterisation due to their intractability in expression, conformational plasticity and instability when extracted from their native membrane environment. However, in recent years advances in NMR and crystallography methods have provided complementary approaches to advance our understanding of the protein family (Shimada *et al*, 2018). The β_2-adrenergic receptor (a structure of which [PDB: 2R4R] is shown in Figure i.12) is an archetypal member of this family and exemplifies some of the tractability problems discussed above. The construct used in the

first crystal structure that did not contain a bacterial fusion protein (as displayed in Figure i.16), despite being truncated, had a mass of around 41 kDa (Rasmussen *et al*, 2007), potentially outside of the mass range of classical NMR techniques.

Advances in understanding the effects of small molecule drugs on the GPCR protein family were made by applying residue specific ^{19}F tagging NMR to study the β_2-adrenergic receptor.

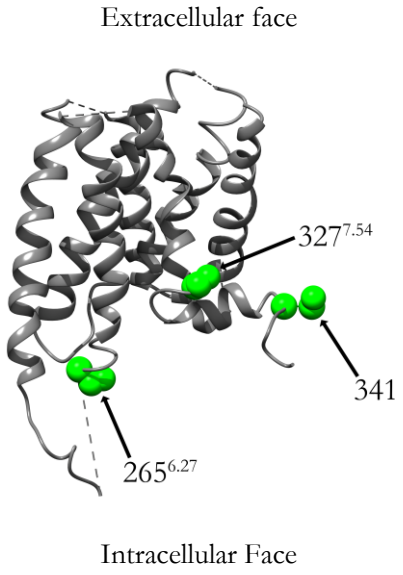


Figure i.16 Ribbon representation of the β_2-adrenergic receptor [PDB: 2R4R] displaying 7-transmembrane helix architecture. The native cysteine residues ($265^{6.27}$, $327^{7.54}$ and 341) may be tagged with ^{19}F-containing compounds.

From a spectroscopic perspective ^{19}F experiments with tagged proteins are extremely simple, even when compared to basic ligand-observed techniques. The studies rely on simple 1D experiments to observe changes in chemical shifts of ^{19}F nuclei between different samples, taking advantage of the exquisite sensitivity of this nucleus to small changes in magnetic environment conferred upon it by its large CSA. ^{19}F atoms do not occur naturally in the translated sequences of proteins. This fact is an advantage as it circumvents the need to assign the ^{15}N, ^{13}C and ^{1}H nuclei, which occur throughout a uniformly labelled protein. The lack of natural ^{19}F in proteins also presents a biochemical challenge, however, as the ^{19}F nuclei must be introduced without disrupting the function of the receptor.

Although ^{19}F-containing non-natural amino acids do exist, the most common method for studying GPCRs using ^{19}F probes is by chemical tagging of protein thiol groups, which are either native or introduced by point mutations. Sulphur atoms possess lone pairs that may react with specific groups of ^{19}F-containing compounds, thus covalently attaching these tags to the protein. The first β_2-adrenergic receptor investigation using ^{19}F NMR tagged native intracellular cysteine residues with the compound 2,2,2-trifluoroethanethiol (TET) (Liu *et al*, 2012) shown in Figure i.17 (A). Wüthrich & co-workers assigned the tagged cysteine residues as $Cys265^{6.27}$, $Cys327^{7.54}$ and Cys341, the locations of which are shown in Figure i.16, by sequential alanine mutations (superscripts refer to Ballesteros/Weinstein numbering of GPCRs (Ballesteros & Weinstein, 1995)). Such an approach would be highly impractical and unnecessary to assign the signals in a 2D spectrum of a uniformly labelled protein, such as Bcl-2, but with only three mutations to make and no multidimensional experiments to assign the ^{19}F signals the mutagenesis approach is appropriate. When making observations of a tag on 1 helix this method requires the other cysteine residues be mutated to alanine so as to observe only 1 tag at a time.

It was found that TET tagged on Cys341 was unresponsive to ligand binding, but chemical shift changes of TET bound to $Cys265^{6.27}$ and $Cys327^{7.54}$ were observed. In this study, and typically across ^{19}F tagging studies of GPCRs in general, the peaks observed in the 1D spectra were consistent with conformational

states in slow exchange with each other, resulting in a change in the proportions of their populations as ligand was added (described schematically in Figure i.17(B)). This method is in principle the same as the CSP method for ligand screening and affinity calculation discussed in the Bcl-2 section of this introduction, but only a one-dimensional spectrum is required due to the low number of or single ^{19}F spin(s).

Wüthrich & co-workers found that the population of the active-associated chemical shift of TET at Cys265$^{6.27}$ was correlated with agonist efficacy. In the same paper the authors had identified this site as exhibiting a large deviation between crystal structures of the receptor bound to agonists and to antagonists, making their results consistent with past studies (Liu *et al*, 2012). Perhaps the more interesting result of their experiments was the observation that TET bound to Cys327$^{7.54}$ could report on changes in conformational state population correlated with the biased agonism of the relevant ligand. A change in the populations of TET bound to Cys327$^{7.54}$ could be observed independently from those associated with the activated state of Cys265$^{6.27}$; this is consistent with a model of receptor activation by certain ligands occurring independently in helix VII without movement of helix VI. It had already been observed pharmacologically that some ligands specifically activate only one or other downstream signalling pathway in a phenomenon referred to as "biased agonism" (Urban *et al*, 2007). Standard agonists primarily activate the G-protein pathway whereas biased agonists activate the β-arrestin pathway more strongly. Carvedilol is one such ligand which was observed to cause the β_2-adrenergic receptor helix VII to populate an active state rather than helix VI. The ^{19}F NMR experimental observations of these structural rearrangements revealed the molecular mechanism that underpins the biased agonism property of some ligands.

The development of drugs exploiting this property has been of particular interest in the GPCR field, given the diverging downstream effects of this protein family. One example is the development of the drug oliceridine by Trevena inc., targeting the μ-opioid receptor (DeWire *et al*, 2013). In opioid signalling the β-arrestin pathway leads to negative side-effects such as respiratory suppression and gastrointestinal dysfunction whereas G-protein signalling causes the analgesic effects. The Trevena group used in vivo cAMP assays to characterise the G-protein signalling of their hits and in vitro complementation assays by chemiluminescence to characterise biased arrestin signalling. Oliceridine is effective in attenuating negative arrestin-induced side-effects (Ok *et al*, 2018) by the mechanism revealed in the described ^{19}F NMR experiments.

Following the observation of structural changes that underpin biased agonism, ^{19}F NMR was further used to identify the thermodynamic parameters of the changes in conformational state of the β_2-receptor by observation at different temperatures (Horst *et al*, 2013). Subsequently the technique was used to hypothesise the existence of an intermediate activation state using the same tagging protocol (Kim *et al*, 2013). Some care must be taken, however, in extrapolating this finding to physiological membranes, since membrane mimetic systems (such as dodecyl maltoside used in the first study and maltose-neopentyl glycol used in the second) are chemically very different from physiological lipid bilayers. It is possible that

imperfect physiological membrane mimics may stabilise conformational states that are not physiologically relevant or result in non-physiological population proportions of conformational states.

Whilst the translatability of in vitro work limits the insights that may be derived from NMR to a degree, the translatability is itself also an area that ^{19}F NMR can explore. The β_2-adrenergic receptor provides an example of how the flexibility of NMR, which has been discussed above as a great advantage, allows it to compare the effects of in vitro experimental conditions themselves. Wüthrich & co-workers used the TET tagging system to investigate the effect of fusing T4 lysozyme into the helix V/VI intracellular loop on the conformational dynamics of the β_2 receptor (Eddy *et al*, 2016). This fusion was present in the first non-rhodopsin GPCR structure. Wüthrich & co-workers compared the ^{19}F NMR spectra of the receptor with and without the fusion protein in intracellular loop 3 and found that with the lysozyme fusion the ^{19}F signal, regardless of the ligand applied, populated exclusively the active state, the chemical shift of which was not affected by the new construct. This result indicates that in vitro experiments in general on the β_2-adrenergic receptor (and perhaps other GPCRs) with the lysozyme fusion would artificially favour the receptor to occupy a more active conformation than would naturally be present if the less rigid wild type intracellular loop sequence were used. The NMR approach was thus able to explore a variety of experimental conditions and provide insight into the most physiologically relevant constructs of the receptor for future experiments potentially stabilising transient conformational states to enable detailed study of them.

To summarise: The example of ^{19}F tagging of the β_2-adrenergic receptor demonstrates how NMR techniques may be extended to more difficult targets than would be possible simply by applying standard uniform isotope labelling. ^{19}F NMR nonetheless, has the same advantage of experimental flexibility as NMR of isotopically labelled proteins, as discussed for the M2 channel, in being able to investigate a wide variety of ligands under different experimental conditions in a way that is difficult to achieve with other structural techniques. ^{19}F NMR applied to membrane proteins is a relatively recent development, however, and so it

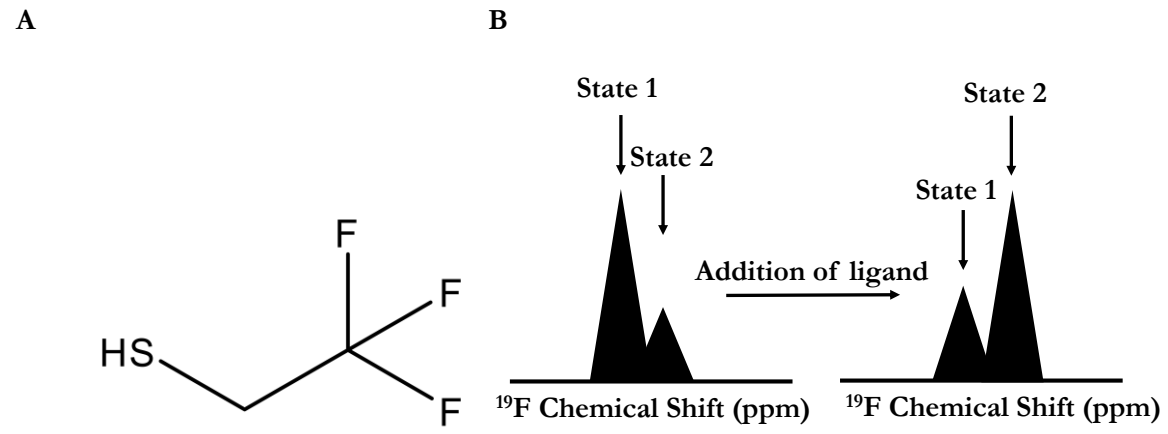


Figure i.17 (A) Structure of TET, a commonly used ^{19}F tag, including in the initial β_2-receptor discussed in this section (Liu *et al*, 2012) **(B)** Schematic of chemical shift change observation of protein tags in 1D ^{19}F spectra. The relative populations of two states, typically in slow exchange, are observed to change when a ligand is added, allowing information about the receptor's mechanism of action to be deduced.

has mainly been applied to rationalise existing pharmacological observations of the action of known compounds. The protein-observed techniques described for Bcl-2 and the M2 receptor were applied in active lead optimisation efforts in a way that has not been done with ^{19}F tagging to date.

Aims

Chapter 1 of this thesis describes the application of ^{19}F NMR to investigate the effect of a compound in development to treat Parkinson's disease on the Adenosine 2A receptor ($A_{2A}R$). This receptor is a GPCR, like the β_2-adrenergic, and a crystallography study found that the receptor occupies a distinct inactive conformation when bound to the novel ligand Cmpd-1(Sun *et al*, 2017). However, crystallisation required the replacement of an intracellular loop with a bacterial cytochrome, limiting the conclusions that could be drawn from the study. ^{19}F NMR has, as discussed, already demonstrated that fusion proteins can drastically affect the conformational landscape of GPCRs and so further evidence was required draw conclusions about Cmpd-1's effects on the $A_{2A}R$ (Eddy *et al*, 2016). Here ^{19}F is used provide evidence for the existence an alternative inactive state of the $A_{2A}R$ in the presence of Cmpd-1 using a tagging site rationally designed by molecular dynamics simulations. It is hoped that this work is a step towards integrating protein-observed ^{19}F NMR into the drug discovery process to address less tractable targets.

Chapter 2 of this thesis describes a structural study of the binding of the anti-tumour compound UCN-01 to alpha-1-acid glycoprotein 2 (AGP2) using more classical NMR techniques more similar to those described for Bcl-2. AGP2 is a blood plasma lipocalin that is a carrier for a broad range of basic drugs (Hervé *et al*, 1998). The strong affinity of AGP2 for small molecule drugs has been implicated in the poor pharmacokinetic properties of those compounds (Fuse *et al*, 1998). In particular, the anti-tumour compound UCN-01, which is a kinase inhibitor that has been investigated for the treatment of cancer, has an affinity of 8 nM for AGP (Welch *et al*, 2007; Kurata *et al*, 2000). Here uniform isotope labelling of AGP2 is used to study UCN-01 binding by NMR in conjunction with the first X-ray crystallography structure of UCN-01 bound to the protein providing a rich view of AGP2's function.

Chapter 1

Observation of a Novel Inactive State of the Adenosine A_{2A} Receptor by ^{19}F NMR

Turnitin Note: This chapter deals with the G-protein coupled receptor project done by this author at UCB Pharma's site. The chapter builds on my publication of that work (Landin *et al*, 2019).

Introduction

G-protein coupled receptors (GPCR(s)) are a protein family of seven-transmembrane (7-TM) helical receptors that is targeted by a larger number of small-molecule drugs than any other single family (Santos *et al,* 2016). GPCRs found in mammals may be divided into five types – rhodopsin family (class A), secretin family (class B), metabotropic glutamate (class C), the adhesin family, and frizzled family (class F) (Alexandere *et al*, 2016). The largest of these is the rhodopsin family (class A) with 719 members. The rhodopsin family has been studied extensively with crystal structures available for many of the class such as rhodopsin, β_2-adrenergic receptor and μ-opioid receptor (Palczewski *et al,* 2000; Cherezov *et al*, 2007; Rasmussen *et al*, 2007; Manglik *et al,* 2012).

GPCR signalling is known to occur mainly via a conserved pathway (Rosenbaum, Rasmussen and Kobilka, 2009). As shown in Figure 1.1, GPCRs receive extracellular stimuli that may be as a result of interaction with drugs such as morphine, endogenous ligands such as adrenalin, activation by light or from proteolytic cleavage. Upon activation the GPCR is stabilised in a conformational state that allows the binding of a heterotrimeric G-protein complex. A single GPCR may be capable of binding to multiple intracellular G-proteins and so their downstream effects may depend upon the G-proteins present at that time in the cell or the receptor's post translational modifications. For example: The β_2-adrenergic receptor has been shown to couple to both the adenylate cyclase activating the G-protein $G_{\alpha s}$ and the inhibitory G-protein $G_{\alpha i}$ depending on the modulation of its phosphorylation status by protein kinase A (PKA) (Daaka, Luttrell and Lefkowitz, 1997). Therefore, GPCRs may transduce the same signal into multiple outputs depending on their cellular context. The G-protein itself is a heterotrimer composed of a G_α, G_β and G_γ subunit. The G_α subunit has intrinsic GTPase catalytic activity and forms most of the intermolecular contacts with a GPCR in its active conformation (Rasmussen *et al*, 2011). Prior to activation the G_α subunit is bound to GDP and exists in complex with the G_β and G_γ. Upon GPCR activation the entire heterotrimeric G-protein binds to the GPCR and exchanges GDP for GTP. The GTP bound heterotrimeric complex then dissociates from the GPCR and itself dissociates into a lone GTP-bound G_α and $G_{\beta\gamma}$ complex. Each of these components has multiple downstream signalling partners. $G_{\alpha s}$ downstream of the β_2 adrenergic receptor may, for example, upregulate adenyl cyclase leading to an increase in cyclic adenosine mono phosphate (cAMP) and therefore activation of PKA (Rosenbaum, Rasmussen and Kobilka, 2009), however the same GPCR can also couple $G_{\alpha i}$ that inhibits adenylate cyclase activity thus having the reverse effect (Daaka, Luttrell and Lefkowitz, 1997). In the case of the β_2 adrenergic receptor these two signalling cascades may even be present in the same cell with the inhibitory $G_{\alpha i}$ protein binding to the GPCR after it has been phosphorylated by PKA upregulated by the earlier $G_{\alpha s}$ signalling. Coupling sequentially to two antagonistic G-proteins therefore attenuates the signal.

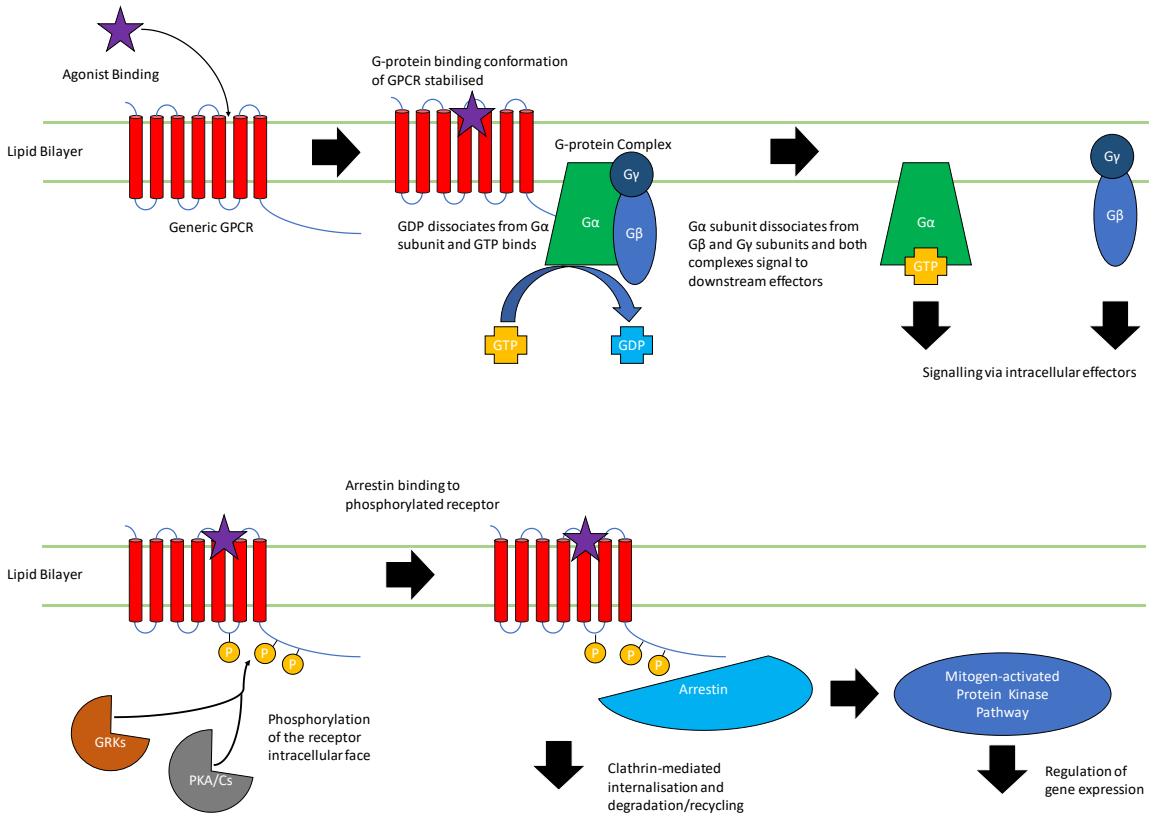


Figure 1.1 Cartoon of generic GPCR signalling pathways. GPCRs are activated by an extracellular stimulus – often the binding of an agonist to the extracellular face of the receptor. Agonist binding stabilises active conformational states of the receptor that allow the heterotrimeric G-protein to bind and exchange GDP for GTP. The GTP-bound G-protein then dissociates into a lone G_{α} subunit and a heterodimer of G_{β} and G_{γ}. Each of the two components of the G-protein heterotrimer can then activate downstream effectors. The GPCR is inactivated both by ligand dissociation and by clathrin mediated endocytosis that is triggered by arrestin binding to the phosphorylated intracellular surface of the GPCR. Arrestin itself also starts a signal cascade leading to changes in gene expression via the mitogen activated protein kinase pathway (Rosenbaum, Rasmussen and Kobilka, 2009; Weis and Kobilka, 2018).

Historically GPCRs have been recalcitrant to investigations of their structure due to their instability and difficulty in forming crystals of sufficient quality for X-ray diffraction. The first direct evidence of their structure came from an electron diffraction projection of bovine rhodopsin in crystalline bilayers (Schertler, Villa and Henderson, 1993). These experiments produced electron density projections consistent with the presence of 7-transmembrane helices as hypothesised prior to that time, which, combined with knowledge of the protein sequence, could be used to assign seven hydrophobic stretches in the sequence to the putative seven α-helices (Baldwin, 1993). An improved projection of frog rhodopsin yielded an improved understanding of GPCR structure (Unger *et al*, 1997), but the first high-resolution structure of a GPCR was bovine rhodopsin solved in 2000 (Palczewski *et al*, 2000).

The structure of inactive bovine rhodopsin (shown in Figure 1.2) confirmed the 7-transmembrane α-helical architecture of GPCRs. The 348-residue receptor is similar in length when compared to other members of the GPCR family and so could be used to provide strong evidence for the secondary structure of other GPCR sequences.

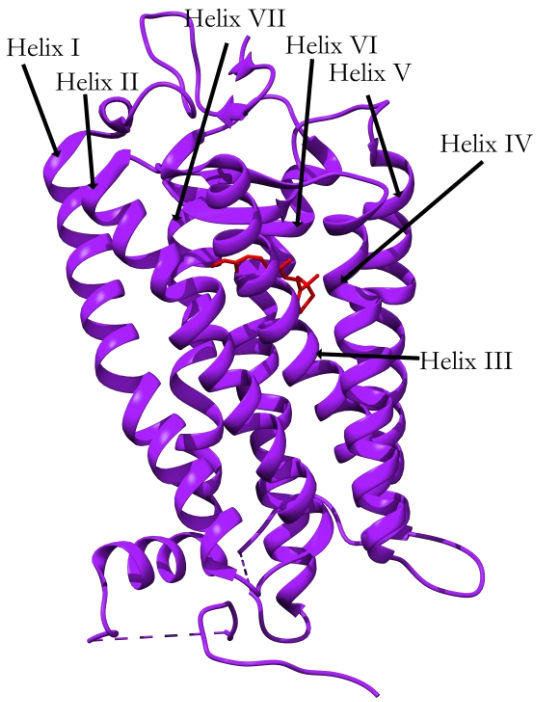


Figure 1.2 Cartoon representation of the X-ray structure of bovine rhodopsin (purple) with its physiological ligand retinal (red) [PDB:1F88] (Palczewski *et al*, 2000) with 11-cis-retinal

One important structural feature of bovine rhodopsin that proved informative for the mechanism of activation of other GPCRs was the function of the highly conserved (D/E)R(Y/W) motif. This motif exists on TM3 of GPCRs. In the published structure Arg135$^{3.50}$ formed a salt bridge to Glu247$^{6.30}$ on TM6 and an additional H-bond to Glu134$^{3.49}$ which was hypothesised to retain the receptor in the inactive state (the superscript refers to the Ballesteros-Weinstein number of the residue; a numbering scheme where the most conserved residue in a helix is assigned to #TM.50 and subsequent residues are counted backwards and forwards from it)(Ballesteros and Weinstein, 1995; Palczewski *et al*, 2000). Further mutational studies on the β_2-adrenergic receptor demonstrated that mutation of the equivalent residues in this receptor resulted in constitutive G-protein signalling, demonstrating the importance of the ionic lock in stabilising the inactive state of GPCRs across the family (Ballesteros *et al*, 2001).

Bovine rhodopsin had several properties that made it more tractable to structural studies; the receptor could be purified in large quantities from bovine retinas and it did not exhibit basal activity. It took the GPCR field seven years to overcome these challenges and crystallise the first non-rhodopsin GPCR (Cherezov *et al*, 2007; Rasmussen *et al*, 2007). The structures of the human β_2-adrenergic receptor were obtained using either the fusion of T4 lysozyme into intracellular loop 3 or an antibody antigen binding fragment (Fab) raised against the native β_2-adrenergic receptor and in complex with the ligand carazolol. The conformational plasticity of the β_2-adrenergic receptor likely necessitated the presence of the fusion protein or Fab to lock the cytoplasmic ends of the transmembrane helices in a particular conformation.

The architecture of the β_2-adrenergic receptor was found to be broadly similar to that of rhodopsin, but with a more open structure. The putative ionic lock of Arg131$^{3.50}$ and Glu247$^{6.30}$ was not closed in either structure of the receptor as the residues were separated by 4.1 Å or 10 Å in the Fab and T4 fusion structures respectively – too distant for a hydrogen bond (Cherezov *et al*, 2007; Rasmussen *et al*, 2007). This makes

the β_2-adrenergic receptor complexed with carazolol structure intermediate between the dark-state structure of rhodopsin and the light-activated structure of rhodopsin (Salom *et al*, 2006).

A structure of a non-rhodopsin GPCR in a fully inactive state was not solved until 2010 with the structure the dopamine D3 receptor in complex with the antagonist eticlopride (Chien *et al*, 2010). This structure [PDB: 1U19] revealed a "locked" conformation of Arg128$^{3.50}$ and Glu324$^{6.30}$ with a hydrogen bonded, as was the case with dark state rhodopsin. It was thought that the T4 lysozyme fusion protein replacing ICL3 used in previous crystal structures impeded the closing of the ionic lock, but new distinct crystal structures of the turkey β_1 adrenergic receptor (Moukhametzianov *et al,* 2011) with the antagonists carazolol, cyanopindolol and iodocyanopindolol revealed that under the right conditions these constructs could also be crystallised in a fully inactive form. The shifting of the receptor conformation to a more active state by the T4 fusion is supported by the ^{19}F NMR study discussed in the introduction.

The GPCR investigated in this study is the adenosine A_{2A} receptor (A_{2A}R) which is shown in Figure 1.3. This receptor is currently the target of two FDA-approved specific small molecule drugs; Regadenoson, developed by Astellas Pharma, is a vasodilator used in cardiac stress tests, (Chen, Eltzschig & Fredholm, 2013) and Istradefylline, developed by Kyowa Hakko Kirin Inc., is a Parkinson's drug approved for use as an add-on treatment with levodopa and carbidopa (Chen & Cunha, 2020). The A_{2A}R has also been the subject of extensive efforts to develop further small molecule treatments to Parkinson's disease (Shook and Jackson, 2011).

The rationale for targeting the A_{2A}R for Parkinson's treatment is supported by studies of the long-term effects of caffeine. Caffeine is one of humanity's most widely consumed psychoactive drugs and has been linked epidemiologically to a reduced risk of Parkinson's disease (Hernan *et al*, 2002). The neuroprotective effect of caffeine as an A_{2A} antagonist has also been supported by studies on the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-treated rodent Parkinson's disease models (Chen *et al*, 2001). Mice treated with the neurotoxin MPTP display comparable symptoms to humans with Parkinson's disease and display improved motor function in response to human treatments such as the dopamine precursor levodopa. These experiments demonstrated that caffeine administration increased residual dopamine levels relative to the non-caffeine-administered control group. Furthermore, this effect is only reproduced with selective A_{2A}R antagonists, thus ruling out caffeine's binding to other adenosine receptors as the mechanism of action (Chen *et al*, 2001a).

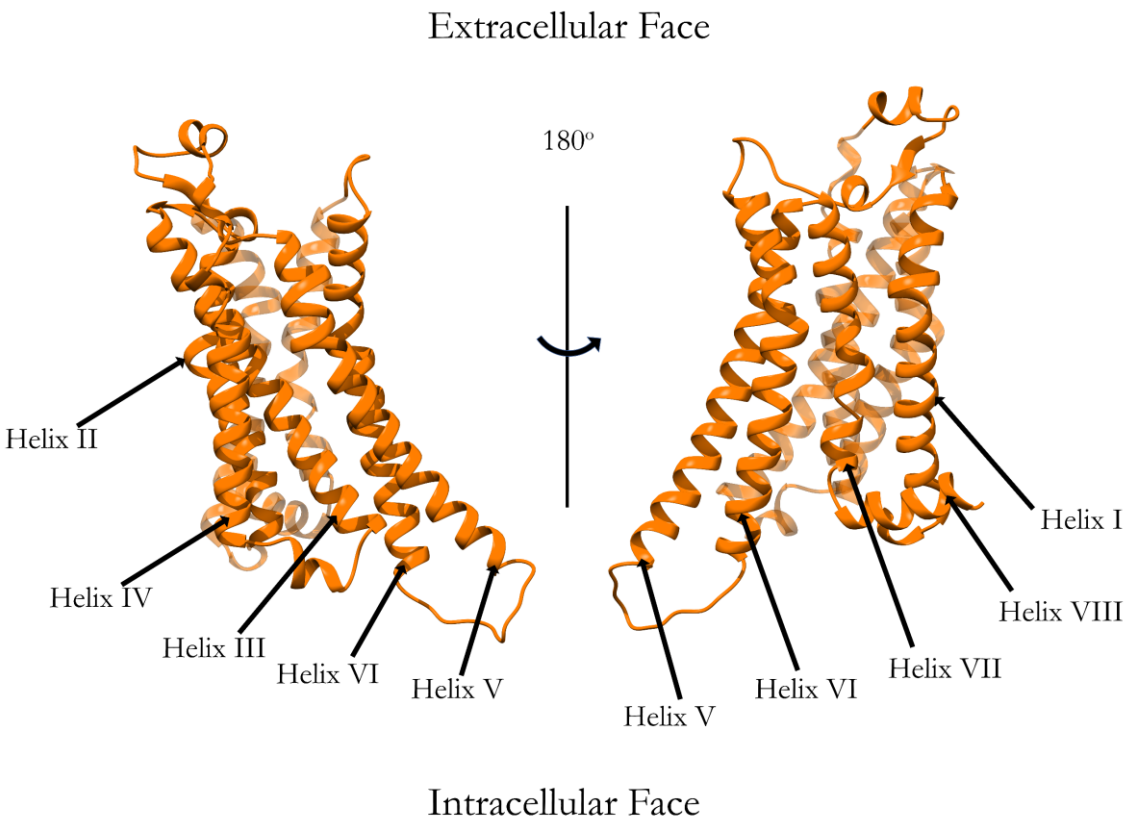


Figure 1.3 Ribbon representation of the $A_{2A}R$ viewed from two positions at 180° relative to each other derived from a computationally completed agonist-bound crystal structure [PDB: 2YDO], with the helix numbers marked.

In addition to their neuroprotective effect $A_{2A}R$ antagonists also improve existing dopamine-depleted symptoms. $A_{2A}R$ agonists have been shown by radioligand binding experiments to decrease the affinity of the dopamine D_2 receptor for its ligands in membranes expressing both receptors (Ferré *et al*, 1993). This leads to suppression of dopamine signalling when the A_{2A} is active due to the weakening of the dopamine D_2 receptor's affinity for its agonists. Furthermore, the $A_{2A}R$ is coupled to stimulatory G_s proteins, whereas dopamine D_2 receptors are coupled to inhibitory G_i proteins and so the receptors also oppose one-another during intracellular signalling (Ferré *et al*, 1993). The $A_{2A}R$ has also been shown to act via dopamine D_2 independent pathways. Dopamine D_2 knockout mice exhibit improved behavioural symptoms when treated with A_{2A} antagonists implying that the $A_{2A}R$ acts in part via an independent pathway.

Taken together these data provide a rationale for treating Parkinson's disease using $A_{2A}R$ antagonists that is described diagrammatically in Figure 1.4.

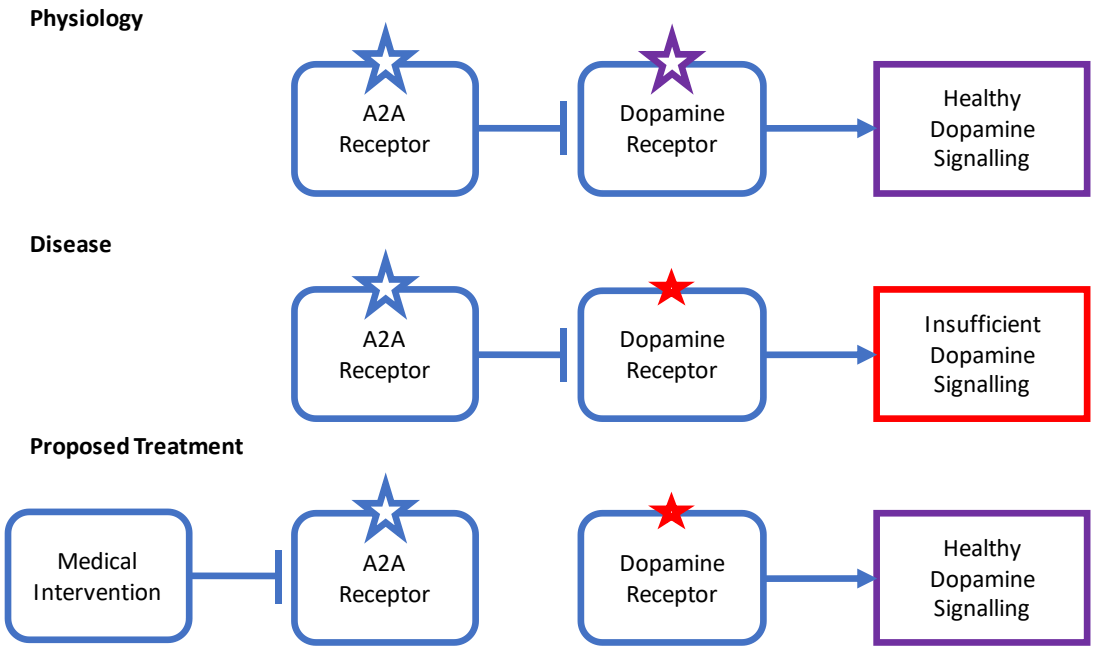


Figure 1.4 Diagrammatic representation of dopamine D_2 signalling in a healthy, Parkinson's disease or treated Parkinson's disease scenario. Dopamine signalling is insufficient in Parkinson's disease, but relieving the inhibition from the $A_{2A}R$ activity can restore more healthy levels of dopamine signalling.

One of the drug discovery campaigns targeting the $A_{2A}R$ led to the development of a novel dual antagonist, which is shown in Figure 1.5(A), to both the A_{2A} receptor and the N-methyl D-aspartate receptor subtype 2B. This antagonist (hereafter referred to as Cmpd-1) was crystallised in complex with the $A_{2A}R$ fused to a cytochrome b562-RIL in intracellular loop 3 (ICL3) (Sun *et al*, 2017). The crystal structure revealed that the methylphenyl and aminotriazole rings of Cmpd-1 occupied a similar position to the heterocycle and furan groups of the well-characterised antagonist ZM 241385. The structure of ZM 241385 is shown in Figure 1.5(A)). In contrast, the *o*-methoxyphenyl group of Cmpd-1 occupies a conformation similar to the phenol group of ZM 241385 when bound to one published thermostabilised $A_{2A}R$ structure [PDB: 3PWH], but different to the conformations of ZM 241385 typical in other structures as seen in Figure 1.5(B)(Doré *et al,* 2011; Segala *et al*, 2016). Cmpd-1 interacts with a unique arrangement of amino acid sidechains from helices I, II and VII. As a result, in the complex with Cmpd-1 the $A_{2A}R$ itself exhibits a distinct conformation when compared to the structures bound to ZM 241385: Notably an outward movement of helices I and V away from the helical bundle as shown in Figure 1.5(C/D). It could not be definitively inferred from this structure that Cmpd-1 causes these rearrangements since the crystallisation of the complex required a different set of crystallisation conditions to those in the complexes with ZM 241385. Previously antagonist-bound crystals were generated in lipidic cubic phase causing the receptor to associate in bilayers. Crystals of the Cmpd-1/$A_{2A}R$ complex, however, required crystallisation using vapour phase diffusion which caused the crystals to pack with individual micelles and contacts through the extracellular domain and the ICL3 fusion protein.

To address the question of whether the conformational changes observed in the crystal structure are due to artefactual crystal contacts or reflect genuine Cmpd-1 induced changes relative to other antagonists we have used a rationally designed ^{19}F NMR tagging approach. Establishing whether Cmpd-1 acts by a novel mechanism to inactivate the $A_{2A}R$ is important commercially as it may raise the value of the patent space of Cmpd-1 and related ligands.

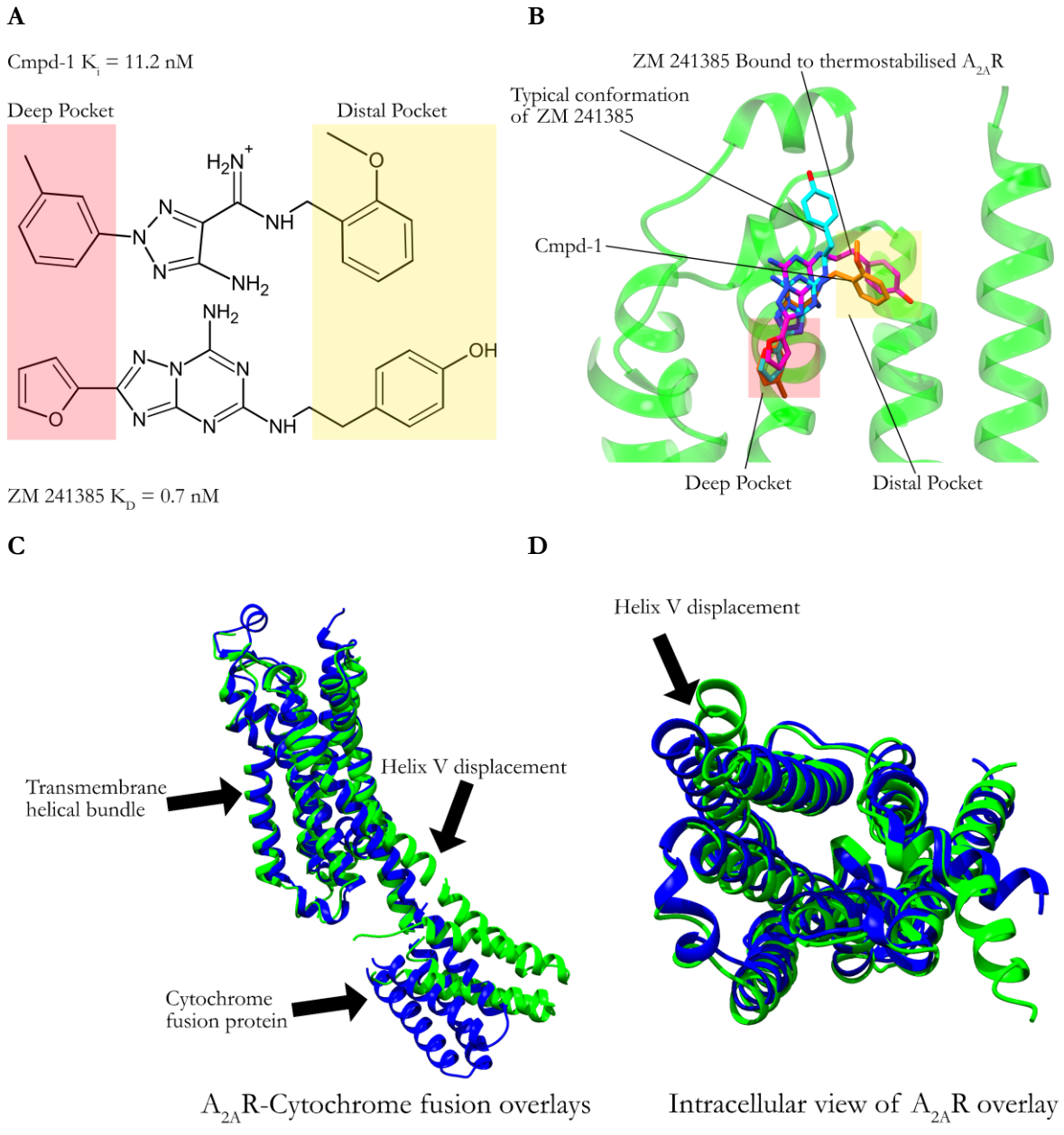


Figure 1.5 (A) The structures of the dual A_{2A} and NRB2 antagonist Cmpd-1 (top) and ZM 241385 (bottom)(Doré *et al*, 2011; Sun *et al*, 2017). **(B)** Structures of ZM 241385 (cyan) in its most common conformation overlaid based on protein backbone alignments, ZM 241385 bound to thermostabilised $A_{2A}R$ (purple), and Cmpd-1 (orange) all shown in a ribbon representation of the $A_{2A}R$ fused to cytochrome bRIL [PDB: 5IU4; 3PWH; 5UIG]. **(C)** Structure of $A_{2A}R$-bRIL in complex with ZM 241385 (blue) and in complex with Cmpd-1 (green) shown with helix V in the foreground **(D)** View of $A_{2A}R$ from the intracellular face with bRIL residues removed. The intracellular terminus of helix V (indicated by the arrows) is displaced in the crystal structure with bound Cmpd-1 relative to that with ZM 241385. ^{19}F NMR is used to observe the above conformational displacement in this study.

As discussed in the introduction, the first GPCR to be investigated by 1D ^{19}F NMR was the β_2 adrenergic receptor (Liu *et al,* 2012). The authors tagged native cysteine residues with the compound trifluoroethanethiol (TET). Since 2012, work has been done comparing the chemical shift sensitivity of various ^{19}F tags to solvent polarity (Ye *et al*, 2015). The most sensitive tag tested in this investigation was 2-Bromo-N-[4-(trifluoromethyl)phenyl]acetamide (BTFMA) as shown in Figure 1.6. The same authors subsequently used this tag to investigate the conformational dynamics of the $A_{2A}R$ (Ye *et al*, 2016). In contrast to previous work on the β_2 adrenergic receptor, a site on helix VI at Val229$^{6.31}$ was selected for tagging based on differences in predicted Cα chemical shift derived from antagonist and agonist structures [PDB: 3EML; 2YDV] respectively calculated by PROSHIFT (Ye *et al*, 2016). This was necessary as, unlike the β_2 adrenergic receptor, the $A_{2A}R$ does not contain native reduced cysteine residues on its intracellular face that could be tagged without mutagenesis. Val229$^{6.31}$ was modified to cysteine and tagged with BTFMA for ^{19}F NMR. These experiments revealed that the $A_{2A}R$ populates multiple active and inactive conformations which are then biased by the addition of appropriate ligands, as schematically shown in Figure 1.7. Prosser & co-workers observed peaks centred on 3 chemical shifts; 61.08 ppm corresponded to the fast exchanging inactive states $S_{1/2}$ (S_1 hypothesised to contain a closed ionic lock and S_2 a broken one) , 61.60 corresponded to the partially active state S_3, and 61.80 corresponded to the fully active state $S_{3'}$ (Ye *et al*, 2016). Curiously, the potent antagonist ZM 241385 yielded spectra with a ~50% population of partially active S_3 state, suggesting that the inactive complex populated an active conformation to some degree. This model of A_{2A} activation is in keeping with the current understanding of GPCR activation in general. A GPCR is not to be understood as a binary switch, but rather signal transducer with multiple outputs that may be affected in different proportions by different ligands (Smith, Lefkowitz and Rajagopal, 2018).

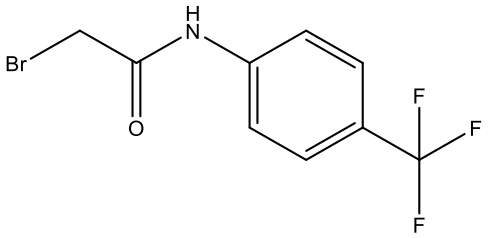


Figure 1.6 The structure of BTFMA, a very chemical shift-sensitive tag that reacts with exposed cysteines via the bromo-alkane group.

The $A_{2A}R$ ligand binding activity has also been investigated by protein observed NMR. One investigation (Clark *et al*, 2017) fed isotope-labelled isoleucine to yeast expressing the receptor and the authors were able to

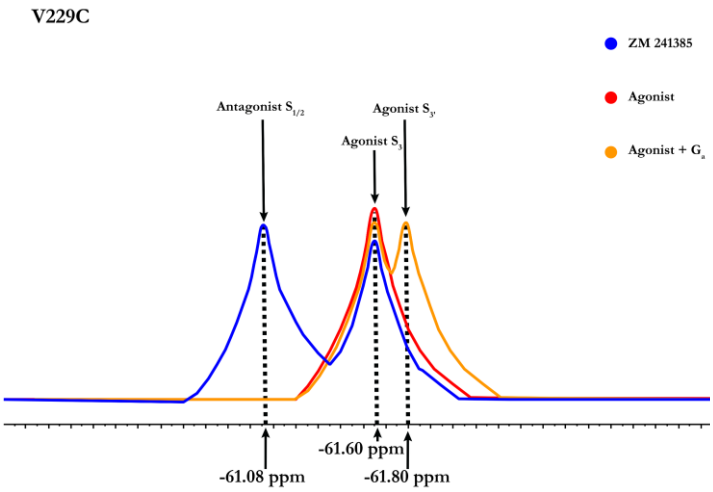


Figure 1.7 Schematic representation BTFMA signals tagged on V229C of helix VI by Prosser & co-workers (Ye *et al*, 2016). Peaks at -61.08, -61.60 and -61.80 were assigned to the inactive $S_{1/2}$ states, partially active S_3 state, and fully active $S_{3'}$ state respectively.

assign four of the protein's native isoleucine residues (Ile92$^{3.40}$, Ile238$^{6.40}$, Ile274$^{7.39}$, Ile292$^{8.92}$) by mutagenesis of individual isoleucine residues. Three of these showed greater conformational flexibility in the agonist bound form relative to the antagonist bound form with the reverse being true for Ile92$^{3.40}$. A further study (Eddy *et al,* 2018) used uniformly ^{15}N, ^{2}H labelled expressed $A_{2A}R$ to observe chemical shift changes in several assigned glycine and tryptophan residues. It was determined that a central tryptophan Trp246$^{6.48}$ (often called the "toggle switch") is functionally coupled to the nearby aspartate Asp52$^{2.50}$ which forms an "allosteric switch" that communicates ligand binding in the extracellular domain to the intracellular domain which adopts different conformations to allow G-protein binding.

In principle it would be possible to conduct our study of the allosteric interactions of Cmpd-1 using isotope labelling of the $A_{2A}R$, rather than a ^{19}F tagging approach, however there are several good reasons to use ^{19}F tagging from both a practical and method development point of view. Isotope labelling requires specialised yeast or insect cell expression systems that can grow in minimal (and deuterated) media. They also require more time for construct preparation by selecting transformed yeast clones or generating new baculovirus stocks respectively. In contrast, the expression of recombinant proteins in mammalian cells can be done easily by lipid-based transfection. This process uses standardised vectors that may be modified by simple site-directed mutagenesis and amplified in bacteria, the vector is then transfected directly into the mammalian cells for high levels of protein expression (Takahashi *et al*, 2010). Improving NMR techniques for systems less tractable to isotope labelling, as discussed in the introduction chapter, is also in itself worthwhile. More recently there have been advances in membrane mimetic systems for GPCRs. These include reconstituting solubilised receptors in nanodiscs (lipid discs bounded by a protein scaffold) and lipodiscs (lipid discs of cell membrane components bounded by amphiphilic polymer scaffold), both of which have been used successfully with $A_{2A}R$ (Jamshad *et al*, 2015; Bocquet *et al*, 2015). Nanodiscs contain additional protein material that may increase transverse relaxation to an extent that makes a protein observe NMR on isotope labelled material intractable, and therefore a ^{19}F tagging approach may be advantageous when applied to these systems. Lipodiscs allow GPCRs to be investigated whilst retained within a small excerpt of membrane excised from the expressing cell which may arguably provide a more native-like environment. This technique could in principle be used for NMR studies of the $A_{2A}R$ in a native membrane environment from a mammalian cell expression system, which would be significantly more physiologically relevant than reconstitution in micelles. Tagging mammalian derived lipodisc-reconstituted GPCRs offers a more tractable approach than isotope labelling to study this family in the most native-like environment currently possible. Certain other members of the GPCR family rely exclusively on mammalian cell expression to produce sufficient amounts of protein for biophysical techniques. Expression of deuterated protein is not possible in these cell lines and therefore ^{19}F tagging extends the application of NMR to these targets beyond mere ligand observed experiments.

In this study a rationally designed novel tagging site on the $A_{2A}R$ is reported at Lys209$^{5.70}$. The site location is based on the putative crystallographic differences induced with Cmpd-1 and ZM 241385 bound, in particular in helix V, and confirmed by molecular dynamics (MD) simulations. The ligand binding activity

of the $A_{2A}R$, carrying the novel mutation is compared with that of a mutant reported in the literature V229C (mutants referred to by single letter code for brevity), both in native membranes with the free cysteine and after tagging with BTFMA. ^{19}F NMR spectra which support the literature resonance assignment for the V229C tagged mutant to inactive and partially active states are also reported alongside spectra for the novel tagged mutant (K209C), which likewise distinguishes active from inactive ligand-induced states based on chemical shift perturbation. In addition, in the presence of Cmpd-1, the K209C mutant reveals a novel inactive conformation of helix V, which remains unresolved by the literature V229C tagged mutant.

The K209C tagged mutant of $A_{2A}R$ represents a unique tool enabling the further development of the Cmpd-1 dual inhibitor series based on its novel mechanism of intervention in a manner not possible with the literature V229C mutation.

We published our key findings from this chapter (Landin *et al*, 2019).

Expression and Purification of $A_{2A}R$

An expression protocol in human embryonic kidney (HEK) cells and purification by immobilised metal affinity chromatography (IMAC) for the A_{2A} receptor was optimised on a green fluorescent protein (GFP) fusion construct (designed by Lisa Joedicke, UCB, Slough)) cleavable by tobacco etch virus (TEV) protease (Fig. 1.8).

The full construct runs low at 48 kDa (actual MW: 67 kDa). After TEV protease cleavage GFP-His_{10} runs at 28 kDa (close to actual MW), whereas $A_{2A}R$ runs at around 30 kDa (actual MW: 38 kDa). After a negative purification, the identity of the $A_{2A}R$ was validated by tryptic cleavage digestion and liquid chromatography/mass spectrometry (LC/MS)(the MS instrument was operated by Rebecca Burnley, UCB, Slough). The results are shown in Figure 1.9. Sufficient cleavage fragments were observed to identify the protein as the $A_{2A}R$.

After this validation experiment, a construct with a cleavable N-terminal twin-StrepII binding (tStrep) tag and cleavable C-terminal His_{10} tag was designed (by Lisa Joedicke, UCB, Slough)) for further expression to allow two-step affinity purification.

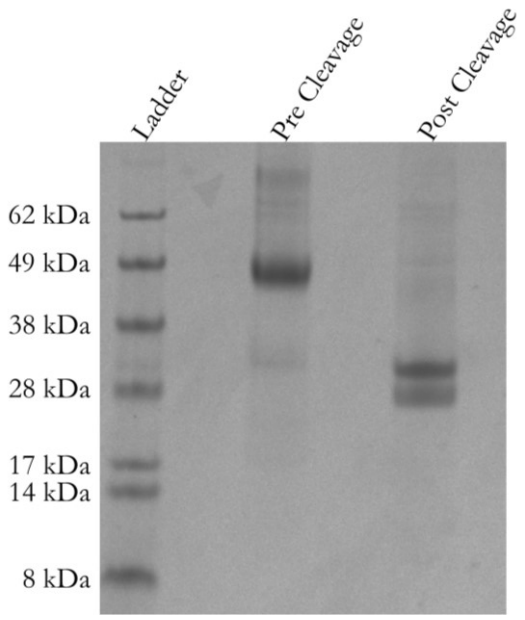


Figure 1.8 SDS PAGE of the $A_{2A}R$-TEV-GFP-His_{10} construct before and after cleavage with TEV protease. GFP-His_{10} and $A_{2A}R$ appear in the second fraction before negative purification of A_{2A}.

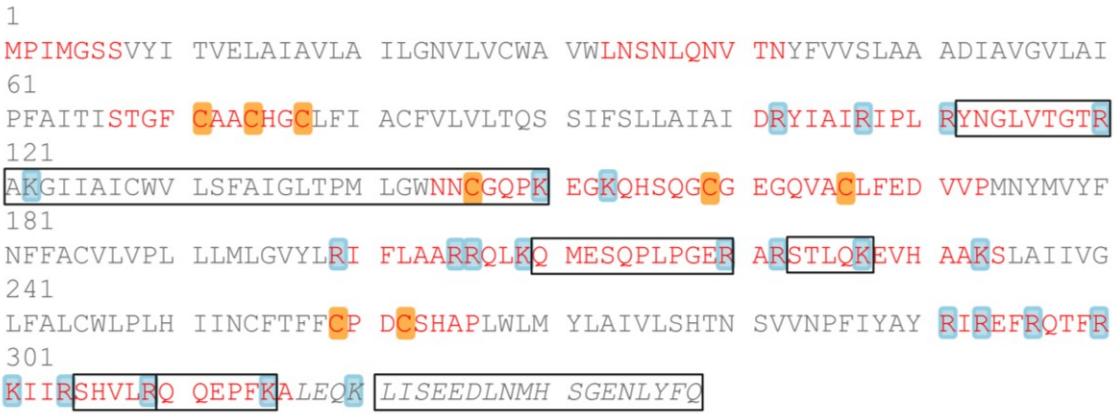


Figure 1.9 Sequence fragments identified by LC-MS of $A_{2A}R$ tryptic digest are enclosed in boxes, with cysteines forming potential disulphide bridges highlighted in yellow and potential cleavage sites highlighted in blue. The sequences shown in red appear as loops in crystal structures. The identified fragments confirm the identity of the protein as the $A_{2A}R$. This mass spectrometry instrument was operated by Rebecca Burnley (UCB, Slough).

The new tStrep $A_{2A}R$ construct (full sequence in methods) was expressed, purified, and analysed by SDS PAGE (Figure. 1.10(A))(full details in methods). The new tStrep-TEV-$A_{2A}R$-TEV-His_{10} construct ran as two main bands with an additional band or double band at ~50 kDa corresponding to a potential SDS stable dimer. The lower of the two bands was probably due to some undesired cleavage of the tStrep tag, since this band can be visualised with an anti-His mouse antibody, but was lost when the sample was purified with streptactin resin (Figure 1.10(B)). The SDS-stable dimer was, however, still present.

The tagging protocol used in this study involved a thiol-tag reaction with the $A_{2A}R$ bound to the IMAC column. To test that this protocol did not result in significant loss of yield it was carried out on a V229C mutant described in the literature (generated by site-directed mutagenesis) (Ye *et al*, 2016). The $A_{2A}R$ was purified by IMAC and incubated with tagging reagents on the column after washing, but before elution.

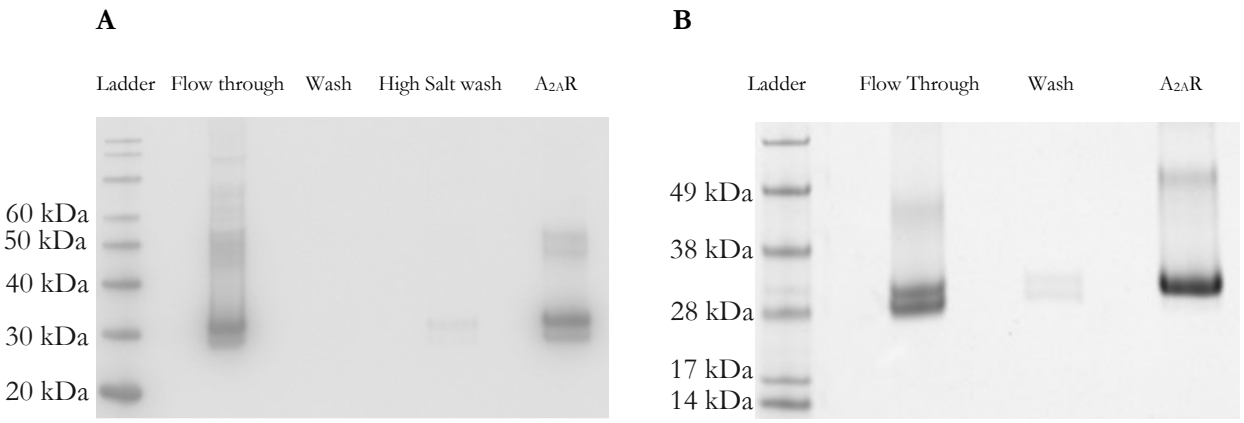


Figure 1.10 (A) Anti-His HRP-conjugated western blot of fractions from an $A_{2A}R$ IMAC purification. The elution shows a double band of A_{2A} at ~30 kDa and a possible SDS-stable dimer at ~50 kDa. **(B)** SDS PAGE image of a streptactin purification of the new construct. The lower band of the double is not present in the final elution suggesting that it is smaller due to missing part of the tStrep tag.

The IMAC column flow through (FT) from each washing and tagging step of the process was analysed by SDS PAGE and the results are shown in Figure 1.11. No bands consistent with $A_{2A}R$ were present in any of the wash or tagging fractions, indicating that the tagging protocol did not affect adhesion of the protein to the IMAC resin. After tagging the $A_{2A}R$ was eluted, desalted, cleaved with in-house TEV protease, and negatively purified yielding a single band, consistent with pure $A_{2A}R$. The remaining protein material on the column from the negative purification was consistent in mass with uncleaved $A_{2A}R$ and TEV protease.

The protocol described in this Figure was initially used with the tag bromo-trifluoroacetone (BTFA), and was subsequently applied to BTFMA.

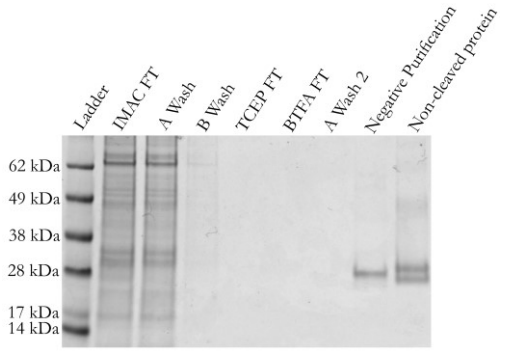


Figure 1.11 Coomassie-stained SDS PAGE of $A_{2A}R$ tagging fractions. A band consistent with pure, cleaved $A_{2A}R$ is present in the final elution with no significant impurities whilst some remains uncleaved as seen (from continued cation affinity) in the imidazole elution from the negative IMAC step.

MD Simulations of the $A_{2A}R$

With expression of the receptor validated the next step was to identify potential tagging sites to distinguish different ligand classes. An in silico approach was used to identify potential mutation sites for ^{19}F tagging of the A_{2A} receptor to follow its conformational changes by NMR.

Preliminary 125 ns MD simulations (full details in methods) were carried out on coordinates of the A_{2A} constructed from antagonist and agonist bound crystal structures [PDB: 5IU4; 2YDO]. The simulated coordinates were constructed to include intracellular loop 3 and helix 1, which were not present in their entirety in the crystal structure due to fusion proteins or poor electron density, but were built into the coordinates sets with a combination of Modeller (Sali and Blundell 1993) and coordinates from the antagonist bound structure 3VG9 (Hino *et al*, 2012). 5IU4 was chosen as the antagonist coordinate set due to its high resolution and 2YDO was chosen as the agonist coordinate set as it contains the native ligand, adenosine. It was assumed that $A_{2A}R$ would not switch from an agonist to antagonist (or vice versa) conformation on this timescale. This assumption was supported by root mean squared deviation (RMSD) comparison between the 2YDO and 5IU4 derived structures across the 125 ns trajectories, which maintained a constant RMSD relative to each other of around 2.5 Å (Fig. 1.12). Residues that maintained distinct conformations between the two simulations were identified by comparing different regions of the receptor against each other more closely (Fig. 1.13).

A **B**

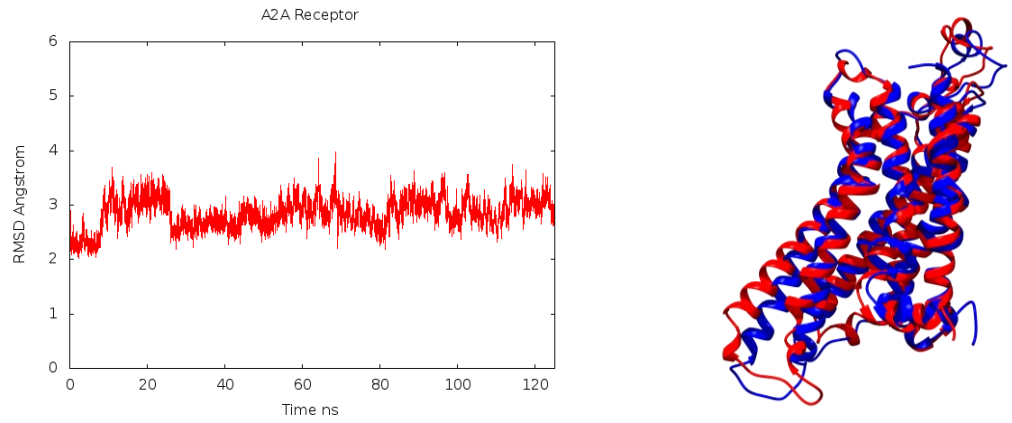


Figure 1.12 (A) RMSD deviation between the agonist and antagonist bound trajectories of the $A_{2A}R$. There is a stable difference of over 2 Å between the two trajectories throughout **(B)** Overlay of ribbon representation of $A_{2A}R$ coordinates from simulations from agonist (red) and antagonist (blue) structures after 125 ns simulation in an artificial lipid bilayer.

These simulations indicated that the terminus of helix 1 and ICL3, between helices 5 and 6, were likely the most appropriate parts of the protein to tag. RMSD deviation of these regions between same the agonist and antagonist trajectories were extracted for a more detailed comparison and the results are shown in Figure 1.13. The RMSD between the trajectories remained above 2 Å for the majority of the simulated time in helix 1 and ICL3. As a comparison the RMSD of helix 4 (a relatively constant parts of all A_{2A} structures used here as a benchmark) remained within 1 Å between the two trajectories throughout the simulation. Comparisons between the structures of the above protein regions are shown in Figure 1.14.

To identify suitable residues in these regions for mutation and tagging, sequence alignments of $A_{2A}R$ and its orthologues and paralogues were prepared using ClustalΩ (Madeira *et al*, 2019) and are shown in Figures 1.15 and 1.16 respectively. It was hypothesised that residues with a greater degree of sequence conservation between homologues would be more likely to affect protein stability or function if mutated. Therefore, residues with poor sequence conservation between homologues were preferred for mutation. Helix I (residues 1-33) shows less sequence conservation at the N-terminus than other regions indicating that these residues might be mutated and tagged without impairing function. ICL3 and the adjacent portions of helices V and VI (residues ~200-240), are well conserved across orthologues. As ICL3 is known to be the G-protein binding site, a greater degree of sequence conservation is expected as the GPCR family signals by a conserved pathway. These alignments indicate that sequence conservation between orthologues (which is much greater than between paralogues) is not necessarily critical to ensure that functional protein can be produced as Val229$^{6.31}$ has been modified previously in the literature, albeit without direct validation of the preservation of ligand binding (Ye *et al*, 2016). The sequence alignments of paralogues in Figure 1.16 demonstrate that across the adenosine receptor family each of the 3 investigated residues vary. Val229$^{6.31}$ is replaced by other hydrophobic residues in other adenosine receptors, whereas Met4$^{1.30}$ and Lys209$^{5.70}$ may

be replaced by residues with very different properties (such lysine -> glutamine). These data indicate that Met4$^{1.30}$ and Lys209$^{5.70}$ may be modified without abolishing receptor function.

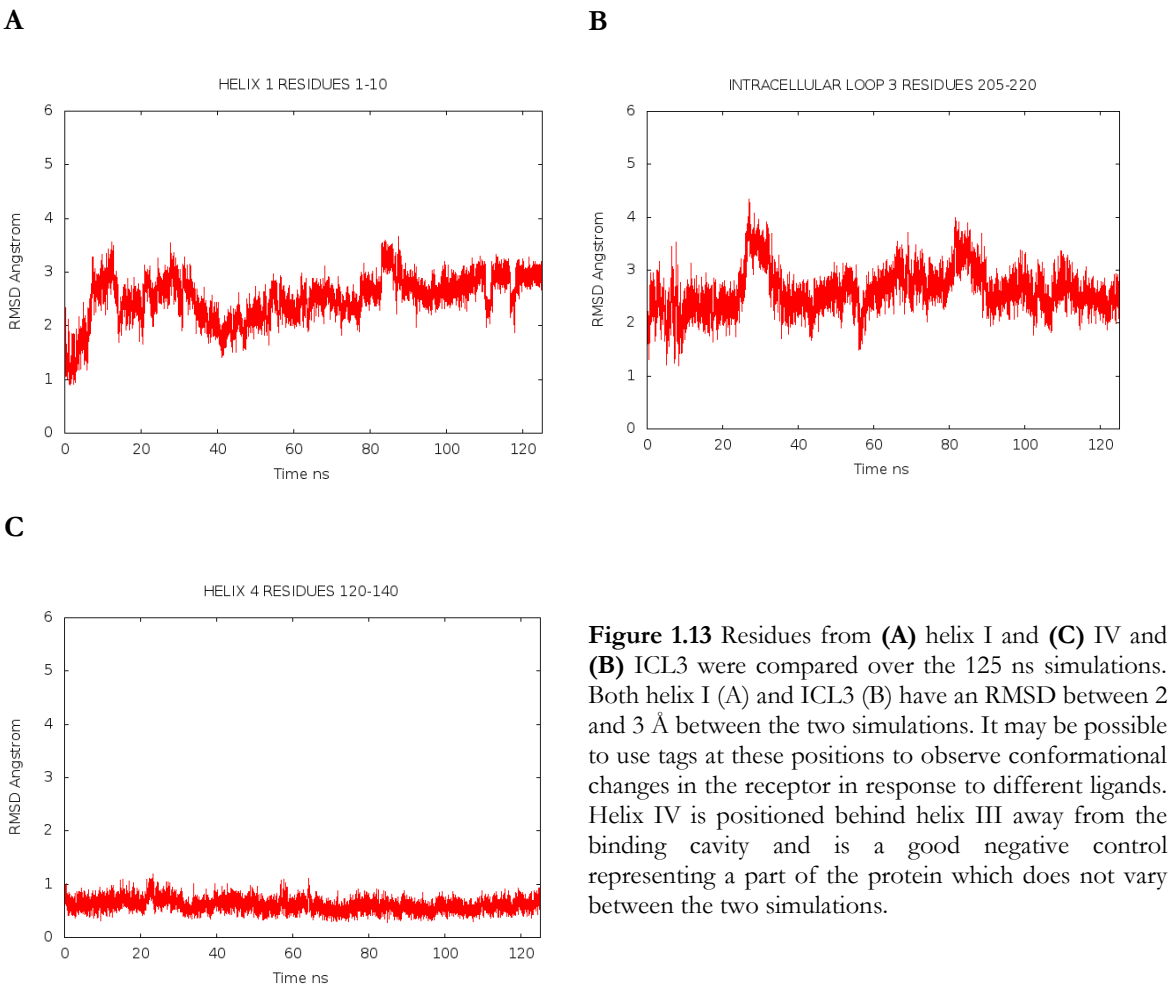


Figure 1.13 Residues from **(A)** helix I and **(C)** IV and **(B)** ICL3 were compared over the 125 ns simulations. Both helix I (A) and ICL3 (B) have an RMSD between 2 and 3 Å between the two simulations. It may be possible to use tags at these positions to observe conformational changes in the receptor in response to different ligands. Helix IV is positioned behind helix III away from the binding cavity and is a good negative control representing a part of the protein which does not vary between the two simulations.

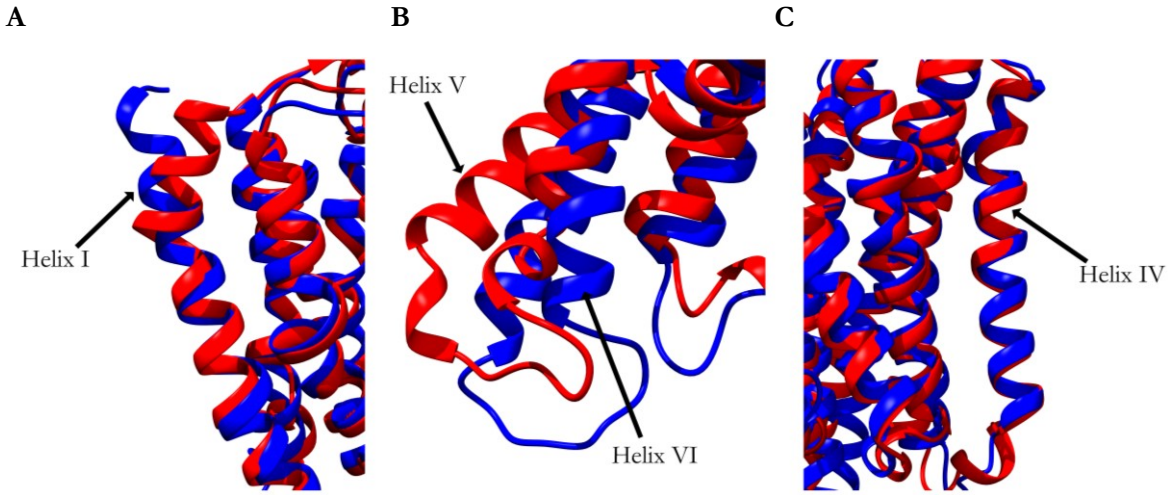


Figure 1.14 Overlays of ribbon representations of $A_{2A}R$ agonist (red) and antagonist (blue) coordinates after 125 ns simulation in an artificial lipid bilayer. **(A)** Helix I exhibits a small deviation between agonist and antagonist trajectories **(B)** ICL3 composed of helices V and VI move outward on agonist binding and retain this deviation over the course of a simulation suggesting that a modification at K209 on helix V may distinguish between different ligand-bound complexes **(C)** Helix IV is included as a negative control that exhibits very little or no deviation over the course of 125 ns.

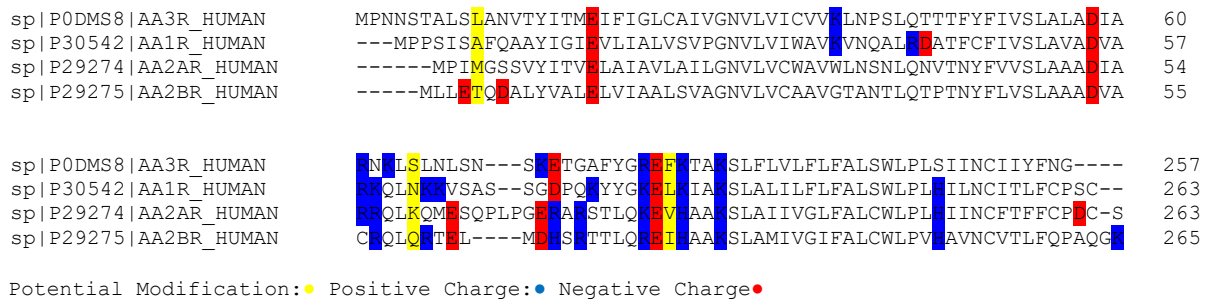


Figure 1.15 Partial sequence alignments of $A_{2A}R$ paralogues with residues investigated for mutation highlighted in yellow (The UniProt Consortium, 2019). None of the highlighted residues are conserved across adenosine receptor paralogues, however the equivalents of $V229^{6.31}$ are invariably hydrophobic residues. Suggesting that its properties are most crucial to the function of the receptor.

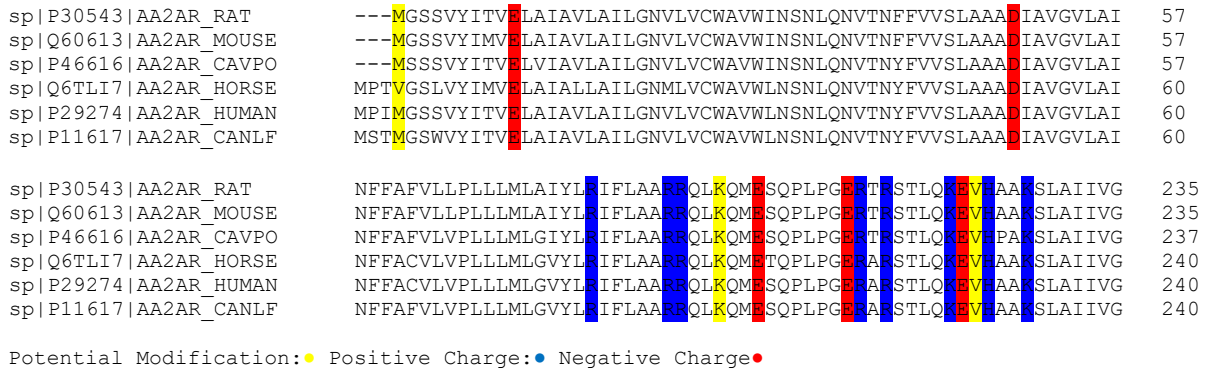


Figure 1.16 Partial sequence alignments of $A_{2A}R$ orthologues with residues investigated for mutation highlighted in yellow (The UniProt Consortium, 2019). $Met4^{1.30}$ (by human numbering) was not conserved in horse $A_{2A}R$ and so may be suitable for mutation. $Lys209^{5.70}$ and $Val229^{6.31}$ were each conserved across related mammals indicating that their identity was important for function, but given the successful study using a mutation at $Val229^{6.31}$ this level of conservation may still allow for point mutations.

Comparisons of the MD trajectories suggested that the residue $Lys209^{5.70}$ could be used to distinguish agonist from antagonist forms of the receptor as seen in Figure 1.14, it is also situated in a region of the receptor sequence that is not well conserved between paralogues (Figure 1.16). The fact that a residue may be different in a closely related protein implies that the presence of a specific amino acid at this location is unlikely to be essential for the protein fold and stability of the 7-TM bundle. After reviewing these data, it was hypothesised that $Lys209^{5.70}$ may, by its position on helix V, distinguish the binding of other antagonists from Cmpd-1 in addition to distinguishing the binding of agonists and antagonists, and this became the main objective of the work.

To further investigate the potential ^{19}F tags to distinguish between antagonist receptor conformations the BTFMA tag was built onto coordinate sets of the $A_{2A}R$ [PDB: 5IU4; 5UIG]. Solvent exposure is a likely mechanism for chemical shift changes of the tag between different conformational states (Ye *et al*, 2016). Therefore simulations of each mutant (V229C and K209C) starting from the coordinates of the $A_{2A}R$ with the ligands ZM 241385 and Cmpd1 [PDB: 5IU4; 5UIG respectively] were run for 625 ns each and the solvent exposure of the ^{19}F atoms in each frame was plotted as shown in Figure 1.17. The solvent exposure of the ^{19}F tags should be considered in a purely qualitative and comparative context as the palmitoyloleoylglycerophosphoethanolamine (POPE)/palmitoyloleoylglycerophosphocholine (POPC) /cholesterol membrane simulated in the MD system is different from both a more complex physiological membrane and from the membrane mimetic, lauryl maltose neopentyl glycol (LMNG) amphipol, used during in vitro experiments. With this caveat, the comparison indicates that the shift in ^{19}F solvent exposure between ligands is greater with the tag at K209C on helix V than at V229C on helix VI. The ^{19}F atoms on the K209C tag are able to sample conformations with solvent exposures of >100 $Å^2$ in the Cmpd-1 bound coordinates, these conformations are not populated to the same degree from the ZM 241385 coordinates. The modal solvent exposure populations of the ^{19}F atoms on the V229C tag change from 61-80 $Å^2$ to 51-70 $Å^2$ based on which ligand is bound, whereas the modal solvent exposure of the tag on K209C changed from 51-70 $Å^2$ to 91-110 $Å^2$. These observations support the hypothesis that K209C can distinguish Cmpd-1 binding from ZM 241385 binding better than a tag at V229C due to the more solvent exposed conformations that can be sampled.

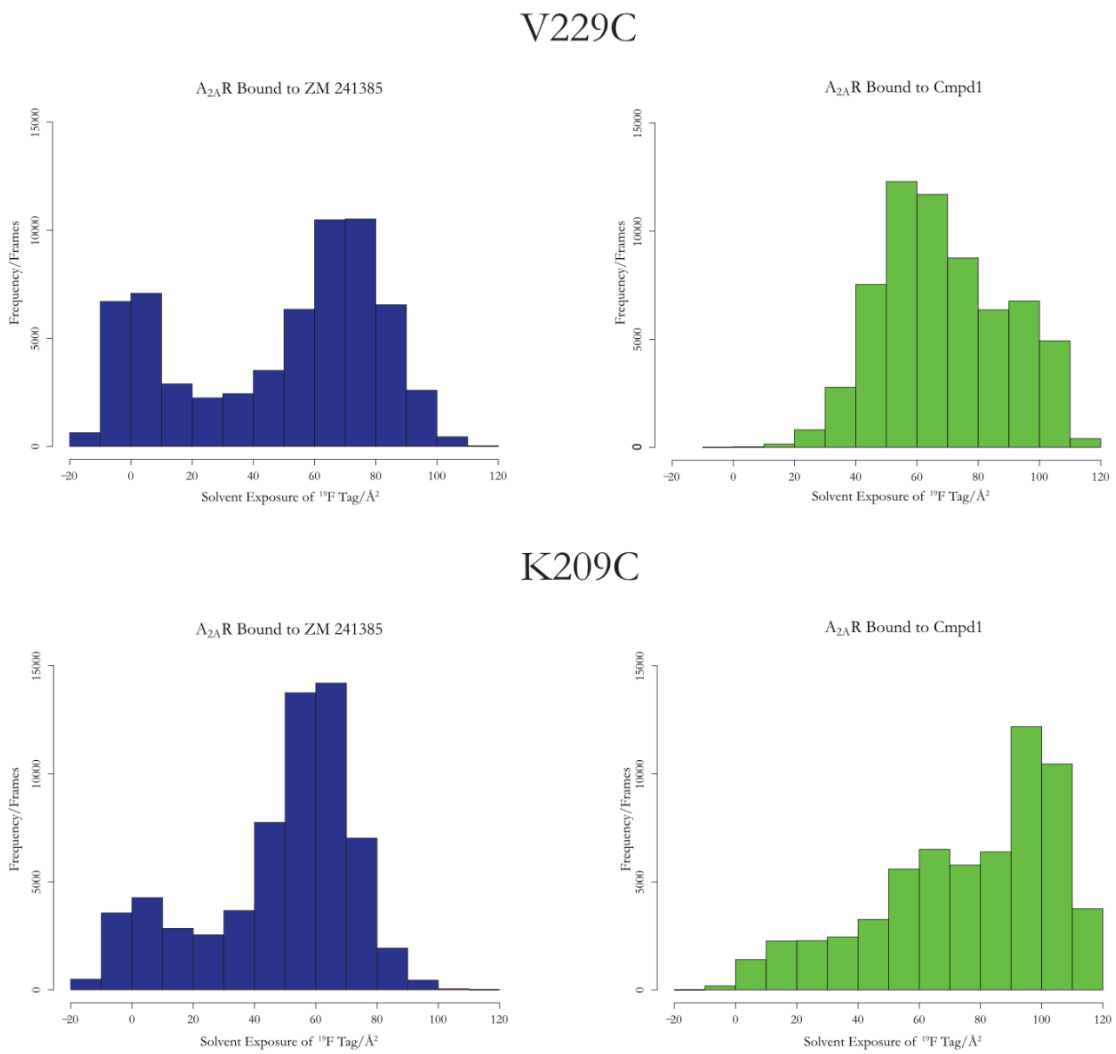


Figure 1.17 Histograms of the solvent exposure of the ^{19}F atoms on the simulated trifluoromethyl groups of the BTFMA tag. When bound to V229C the most populated solvent exposed conformations are between 60 and 80 Å^2 for the standard antagonist ZM 241385 and between 60 and 70 Å^2 for Cmpd-1. When bound to K209C the most populated solvent exposed conformations are between 50 and 70 Å^2 for ZM 241385 and between 90 and 110 Å^2 for Cmpd-1. From the large shift towards being able to sample more solvent-exposed conformations we may hypothesise that Cmpd-1 would be likely to cause a greater chemical shift change to a ^{19}F tag attached to K209C than V229C.

Radioligand Binding Assays

The overarching aim of this study was to observe functionally relevant states of the $A_{2A}R$ by NMR. It is therefore essential that modifications, in this case point mutations, made to the receptor can be confirmed not to interfere with its function. It is also necessary to confirm that the micelle-reconstitution and purification protocol does not disrupt the fold and function of the $A_{2A}R$. Therefore, three types of radioligand binding assay were carried out: Saturation binding experiments, IC50 experiments and displacement experiments on purified protein. These assays used the selective A_{2A} antagonist ZM 241385, originally developed at Zeneca (Poucher *et al*, 1995). ZM 241385 has itself been used to investigate the effect of $A_{2A}R$ antagonists in general on dopamine levels using in vivo models of Parkinson's disease (Gołembiowska *et al*, 2009), but is not itself currently used as a therapeutic for this indication. Saturation binding experiments were carried out on samples of membrane expressing $A_{2A}R$ to fit a saturation binding curve. The binding of radioactive ZM 241385 tends towards a maximum (Bmax) as all the $A_{2A}R$ sites are filled and the curve is fitted to the equation $Y=Bmax \times x / (K_D+x)$. This allows the apparent K_D (the concentration of ligand where half the receptor is in the bound state) to be calculated as an inverse measure of the strength of binding of the ligand. As this property is already well documented, binding of the ligand ZM 241385 serves here as a measure of how similar the mutants are to each other and to wt $A_{2A}R$. If the point mutations do not significantly alter the K_D, then it is reasonable to assume that the receptor's properties are the same in general. IC50 experiments are specific to the set of conditions under which they are measured and determine the concentration of ligand required to remove half of a protein's activity in a given situation. In this case, we calculate IC50s of the ligands competing ZM 241385 out of the $A_{2A}R$ binding pocket which gives a comparative measure of the properties of the ligands and the different $A_{2A}R$ constructs under the conditions we tested them. The amounts of [^{3}H] ZM 241385 that remains bound to the membrane embedded $A_{2A}R$ over a range of increasing concentrations of a competitor ligand, is fitted to a logarithmic curve, and the IC50 is found at its inflection point. The displacement experiment is more conveniently carried out than competition experiments on purified protein samples, but is less precise as the desalting columns used to separate bound from unbound ligand are less reliable than the plate harvester used for in-membrane experiments. This experiment determines if there is significantly more bound radioligand, and therefore functional $A_{2A}R$, in a test sample relative to a control sample, where a large excess of non-radioactive competitor is present. The advantage of this experiment is that the receptor is subjected to identical assay conditions to those used in the NMR experiment.

Saturation binding experiments were carried out in two different buffers (Fig. 1.18). The buffer reported by PerkinElmer (chosen as they are prominent suppliers of A_{2A} and associated reagents) was found to favour specific over non-specific binding and so was chosen for future experiments.

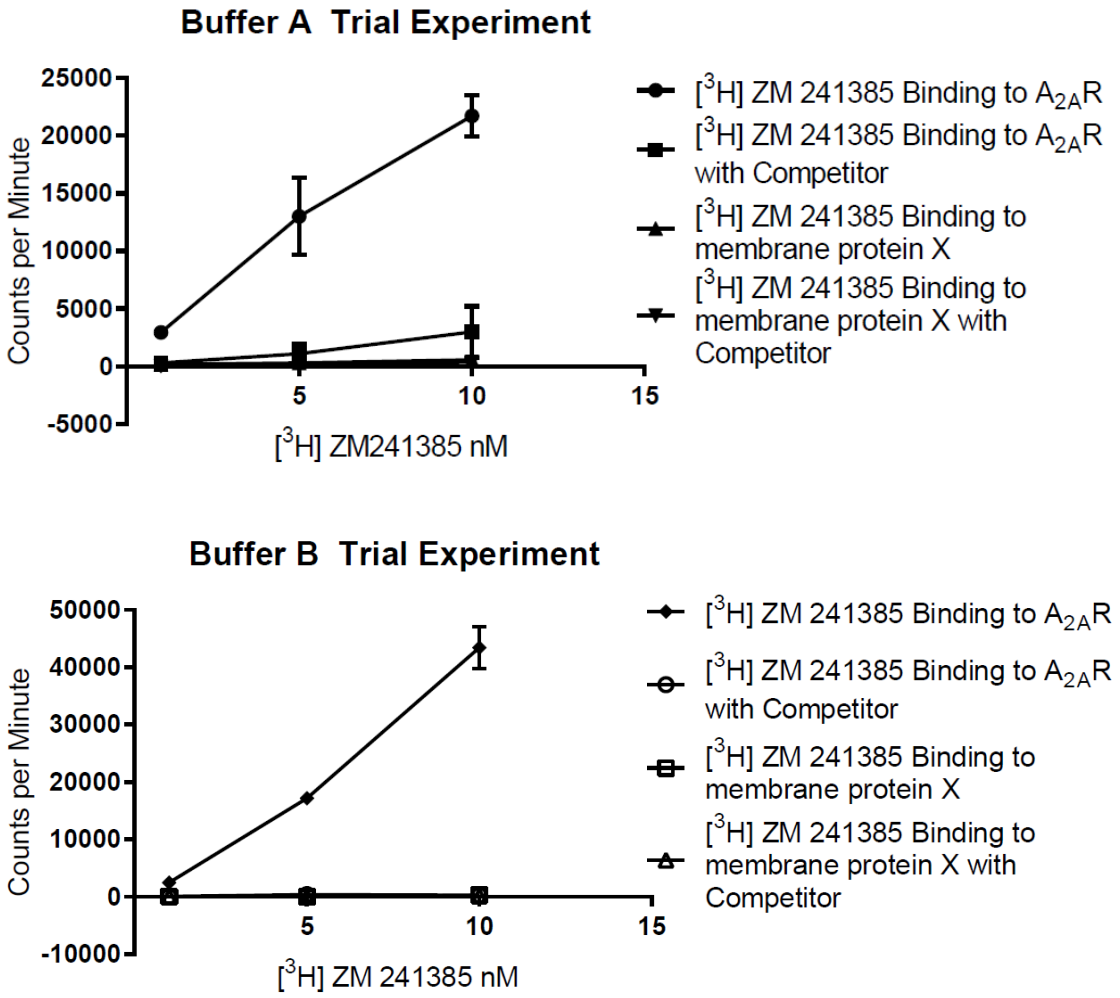


Figure 1.18 Identical saturation binding experiments were carried out in radioligand buffer A (Jaakola *et al*, 2008) and radioligand buffer B (PerkinElmer, 2009). Buffer B was selected for use from this experiment as there was greater differentiation between specific and non specific binding. Membrane protein X is a non-GPCR protein also expressed in HEK 293 membranes as a negative control.

Saturation binding experiments with this buffer were carried out on HEK membranes that had expressed wild type protein, the V229C mutant (Ye *et al*, 2016) and the K209C mutant reported here (Fig. 1.19). Calculated apparent K_Ds from these results were not within a range that could be reliably ascertained from these data. Furthermore, the wild-type (wt) construct appeared to show reduced binding affinity as compared to the mutants. Therefore, a further saturation binding experiment was carried out on the wt construct with an extended radioligand concentration range (Fig. 1.20).

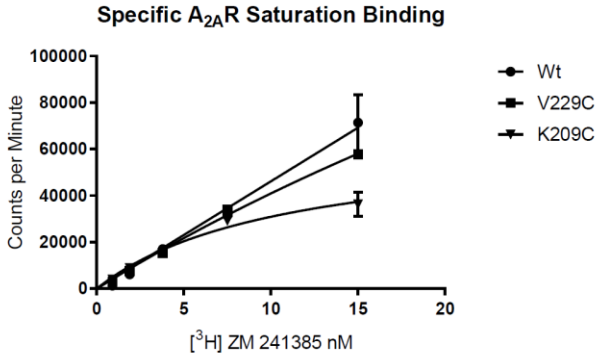


Figure 1.19 Specific [^{3}H] ZM 241385 binding to membranes from HEK cells expressing dual affinity tagged A_{2A} construct with one of two cysteine mutations compared to the truncated wild type construct (n=3). All curves show an apparent K_D at least tenfold higher than the reported values from the literature, with the mutant's possessing apparent K_Ds of V229C: 76 nM, 25 nM – undefined 95% confidence interval; K209C: 11 nM, 5 – 35 nM 95% confidence interval; the wild type construct did not possess a calculable K_D. These graphs where fitted to the single site binding equation (Y= Bmax*x/(Kd + x)) and further assumes a large excess of ligand to target. The error bars show the standard error of the mean.

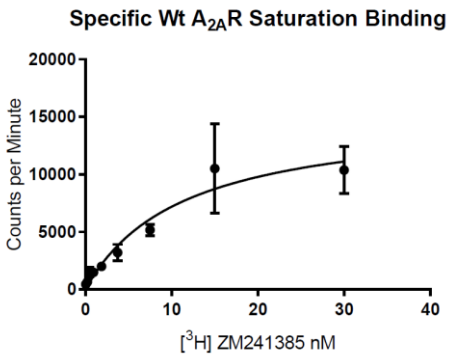


Figure 1.20 Specific [^{3}H] ZM 241385 binding to HEK membranes expressing dual tagged affinity tagged wt construct (n=3). Apparent K_D: 11 nM, 5-25 nM 95% confidence interval. These graphs where fitted to the single site binding equation (Y= Bmax*x/(Kd + x)) and further assumes a large excess of ligand to target. The error bars show the standard error of the mean.

The extended range of measured concentrations allowed an apparent K_D of 11 nM to be calculated. This was around fivefold higher than the 2 nM value reported in the literature (Jaakola *et al*, 2010). The buffer from PerkinElmer used in the experiments of Figures 1.18 did not contain Na^+ ions which are known to be negative allosteric modulators of class A GPCRs (Katrich *et al*, 2014). Since the ZM 241385-bound conformation of the $A_{2A}R$ is stabilised in the presence of bound Na^+ ions then a sub optimal Na^+ ion content may cause a lower proportion of a given concentration of ZM 241385 to be bound in the $A_{2A}R$, thus shifting the point where half of the receptor would be occupied towards an apparently higher ZM 241385 concentration. To test if the lower apparent K_D was due to a lack of Na^+ ions, the buffer was changed to one containing 100 mM NaCl (Uustare *et al*, 2005) and the experiment was repeated (Fig. 1.21). A K_D of 9 nM was calculated from these data, which was still above the 2 nM reported in the literature (though this difference may not be formally significant) (Jaakola *et al*, 2010), and was also not significantly different from the previous experiment. One possible explanation is that literature values were derived from experiments on adenosine deaminase-treated membranes which removed the endogenous agonist, adenosine and allowed higher affinity binding of ZM 241385 to be measured. Here it was not necessary to characterise any ligands de novo with determination of absolute K_D values, it was sufficient to compare the relative affinities in our system and to determine if the cysteine mutants display altered activity. Therefore, competition experiments were performed, from this point on, in the higher NaCl buffer.

The IC50 of ZM 241385 in homologous competition was calculated for each of the wt, V229C and K209C constructs directly in the HEK membranes expressing them (Fig. 1.22). The IC50 of ZM 241385 was 22 nM, 17 nM and 11 nM for Wt, V229C and K209C respectively. These values are very comparable and with

significant overlap of their standard errors of the mean, giving support to transferability of results from mutant to wt protein.

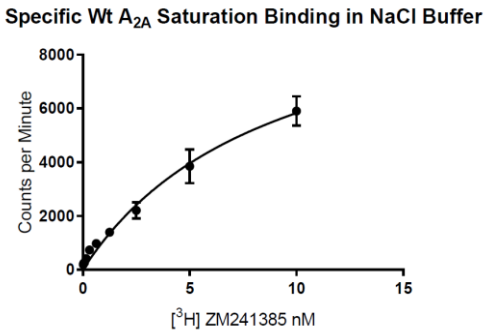


Figure 1.21 Specific [^{3}H] ZM241385 binding to HEK membranes expressing dual affinity tagged $A_{2A}R$ wt construct in an NaCl-containing buffer (n=3). The apparent K_d was 9 nM, 5 – 17 nM 95% confidence interval. Fitting method as above.

These data suggest that one may legitimately draw inferences about the wt receptor from experiments on the cysteine mutant constructs as the variation between the IC50s is smaller than or of a similar size to the standard errors of each experiment. Subsequently it was determined by radioligand binding experiments on purified and tagged receptor that both mutant constructs were active when reconstituted in LMNG/CHS micelles and tagged with BTFMA (Fig. 1.23). The NMR experiments could therefore proceed with confirmation that the receptor was still active in the final samples.

A **B**

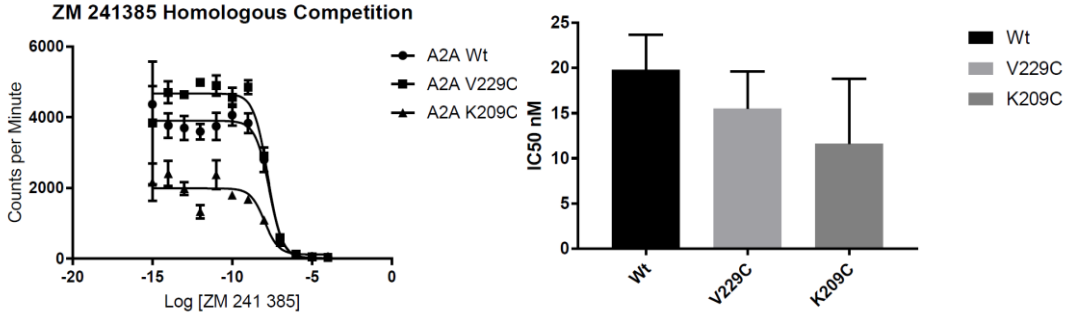


Figure 1.22 (A) Homologous competition curves of [^{3}H] ZM 241385 for the two tagged constructs and wt $A_{2A}R$ (n=3). These fits used the equation for competition binding at one site (Y = Ymin + (Ymax-Ymin)/(1+10^(x-log(IC50))) and the error bars are standard deviations of the mean. **(B)** Chart of calculated IC50's of each mutant. The error bars are the fitting errors for each curve and are consistent with there being no significant difference between the curves. The IC50 of ZM 241385 was 22 nM, 17 nM and 11 nM for Wt, V229C and K209C respectively. The error bars are the standard errors of each fit.

A **B**

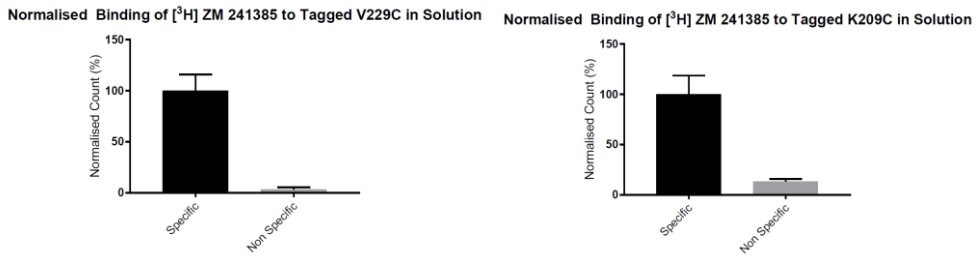


Figure 1.23 Specific and non-specific binding of [^{3}H] ZM 241385 to purified and tagged receptor with the mutations **(A)** V229C and **(B)** K209C normalised in both cases to specific binding. Both mutants still have binding activity in LMNG/CHS micelles after purification and tagging.

^{19}F NMR of Tagged $A_{2A}R$

With the functionality of the K209C and V229C mutants verified, ^{19}F NMR experiments were carried out to characterise the responses of helices V and VI to different ligands. The chemical environment of the tag changes as the protein populates conformational states in different proportions, resulting in the ^{19}F spins becoming more or less shielded in different states and so their chemical shift changes. BTFMA has been chosen as our tag since an investigation of several commercially available tagging compounds (Ye *et al*, 2015) revealed that BTFMA was the most sensitive to changing solvent polarity.

To ensure that meaningful conclusions could be drawn from the ^{19}F experiments, spectra were collected with the antagonist caffeine and agonist 5'-N-Ethylcarboxamidoadenosine (NECA) (the structures of which are shown in Figure 1.24) in addition to ZM 241385 and Cmpd-1. NECA is included to demonstrate that the BTFMA tag can distinguish between inactive and active states of the receptor as this structural transition is the most studied and well characterised, thus validating the ^{19}F tagging data (Xu *et al*, 2011; Lebon *et al*, 2011). Caffeine is chosen as an additional antagonist to ZM 241385 to verify that the ^{19}F data reports on a similar conformation for antagonists in general, but a distinct conformation for Cmpd-1.

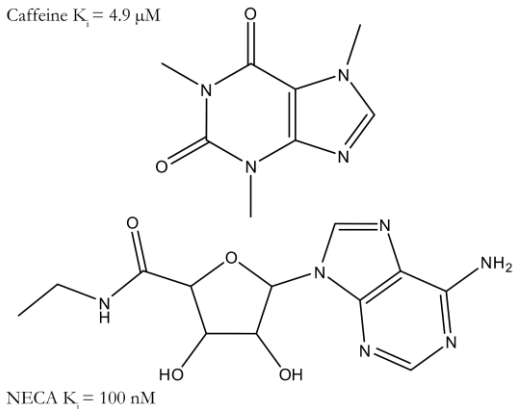


Figure 1.24 Structures of the $A_{2A}R$ antagonist caffeine (top) and the $A_{2A}R$ agonist NECA (bottom) with their affinities (Doré *et al*, 2011)

The NMR experiments were conducted on BTFMA-tagged samples of the V229C (Ye *et al*, 2016) and our K209C mutants as discussed previously in Figure 1.17. The initial ^{19}F NMR experiment was an attempt to reproduce the antagonist to agonist shift reported in the literature (Ye *et al*, 2016). $A_{2A}R$ V229C was tagged on IMAC resin and cleaved, samples were then desalted into buffer A and 5000 scans of 1D ^{19}F data were collected. The high number of scans carried out allowed the $A_{2A}R$ to be studied at very low protein concentrations (20 μM) improving its stability and economising on purification reagents. Figure 1.25 shows spectra of $A_{2A}R$ in solution with 10 mM of the antagonist ZM 241385 (hereafter always in blue) and 4 mM NECA (hereafter always in red). Prosser and co-workers report an inactive ensemble at -61.08 ppm and an active ensemble at -61.60 ppm. These data by contrast show main peaks at -61.72 ppm and -61.80 respectively for ZM 241385 and NECA. There are two similarities to the published spectra: The main agonist peak is shifted upfield relative to the antagonist (but only slightly), and there is a peak around 60.5 ppm that Prosser and co-workers ascribe to non-specific tagging (Ye *et al*, 2016) .

To ascertain which peaks were due specifically to protein bound tags, the antagonist-bound sample was incubated at 25°C for 36 hours to allow aggregation of functional $A_{2A}R$ reducing the corresponding signals and run again for 5000 scans, the resultant spectrum is shown in Figure 1.26.

The peak at -61.72 had decreased in intensity significantly over this time period, suggesting that it was due to the $A_{2A}R$ aggregating over time and becoming invisible to NMR. The other peaks at -60.5, -61.62, and -61.85 on the other hand persisted in intensity over the course of the experiment. This strongly suggests that they were due to material which remained in solution after the $A_{2A}R$ has aggregated and unfolded, and are therefore artefacts. The protocol was amended in such a way that the tagging reactions still proceeded for ~18 hours, but with extensive washing to remove free tag whilst the $A_{2A}R$ was retained on the IMAC resin. Furthermore, in case glycerol could have caused the major peaks to be shifted from the literature values it was omitted in the final buffer (hereafter buffer SG) for the next tagging experiment.

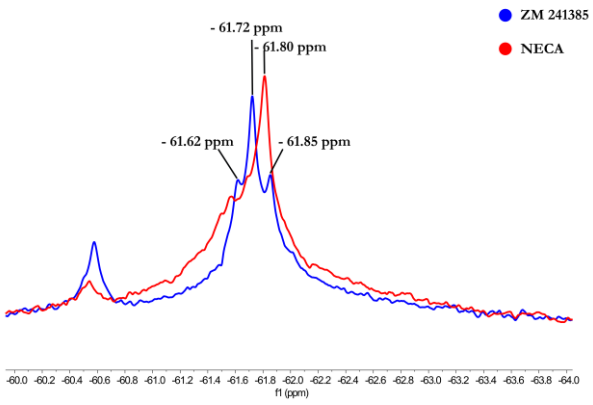


Figure 1.25 1D ^{19}F NMR spectra of $A_{2A}R$ V229C tagged with BTFMA in solution with antagonist ZM 241385 (blue) and agonist NECA (red). The agonist spectrum was broader than that of the antagonist which could be due to the sampling of a broader range of active and partially active conformational states on differing timescales than occurs with the antagonist. Though different these spectra did not confirm the tagging experiments from the literature (Ye *et al*, 2016).

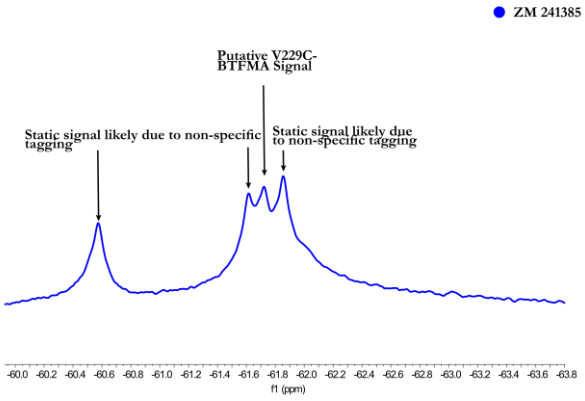


Figure 1.26 1D ^{19}F spectrum of BTFMA-tagged $A_{2A}R$ V229C after 36 hours at 25°C. The peaks 60.5 ppm and 61.85 ppm did not decrease in intensity suggesting that they were not due to tagged $A_{2A}R$, as the $A_{2A}R$ is unstable at this

The $A_{2A}R$ was tagged with the modified protocol and further NMR experiments were acquired with ZM 241385, caffeine, NECA and UCB's proprietary ligand Cmpd-1 (Fig. 1.27). These data are very similar to that reported in the literature. The signal observed with 100-fold or greater concentration of antagonist relative to protein, appears close to the -61.08 ppm reported for the putative S_{1-2} inactive ensemble (Ye *et al*, 2016). The sample with 20-fold agonist present relative to protein shows a peak at -61.55 ppm which is close to the 61.60 reported for the putative S_3 partially active state. The slight variations may be due to our 10-fold lower LMNG

concentration, differences in DMSO concentration, differences in observed vs modelled states or some combination of the above.

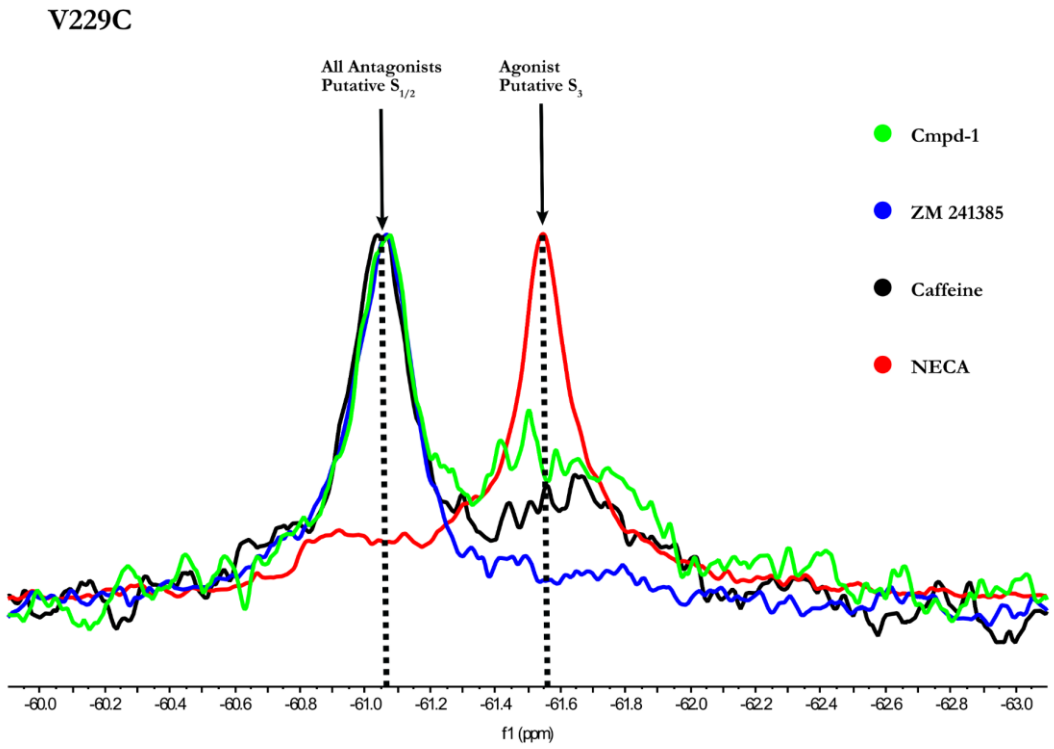


Figure 1.27 1D ^{19}F NMR spectra of $A_{2A}R$ V229C tagged with BTFMA in solution with antagonist ZM 241385 (Blue), antagonist caffeine (Black), antagonist Cmpd-1 (Lime), and agonist NECA (Red) normalised by peak height. The antagonist peaks all appear around -61.06 ppm (putative $S_{1/2}$), whereas the agonist peak appears at -61.55 ppm (putative S_3) (Ye *et al*, 2016).

These data largely confirm the observations reported in the literature around antagonist conformation and demonstrate that Cmpd-1 causes the tag on helix VI to adopt a conformation which is not distinguishable from two other known antagonists as the main peak does not move away from -61.06 just like ZM 241385. On the other hand, the APO spectra reported in the literature could not be reproduced, as protein was too unstable to yield significant signal in our experiment in the absence of ligand, it is hypothesised that this is due to lower detergent concentration used in this work. Therefore, samples of our K209C tagging site on helix V were prepared and tagged in the same way, and an identical set of experiments were run (Fig. 1.28).

The new K209C tagging site shows a reversal in the direction of the shifts between agonist and antagonists compared with the V229C site. Here the agonist shifts the tag signal downfield relative to antagonist and Cmpd-1 shows an upfield shift relative to the other two antagonists. This strongly implies that helix V samples a different conformation when Cmpd-1 binds to the receptor when compared with the other antagonists, and furthermore that this conformation is not detectable by the previously reported tag on helix VI. The main peak in the spectrum of $A_{2A}R$ with Cmpd-1 is broader than those from the $A_{2A}R$ with caffeine or ZM 241385. This may be due to the alternative inactive state existing in an intermediate exchange regime, causing line broadening, or that this peak represents the average of several unresolved states.

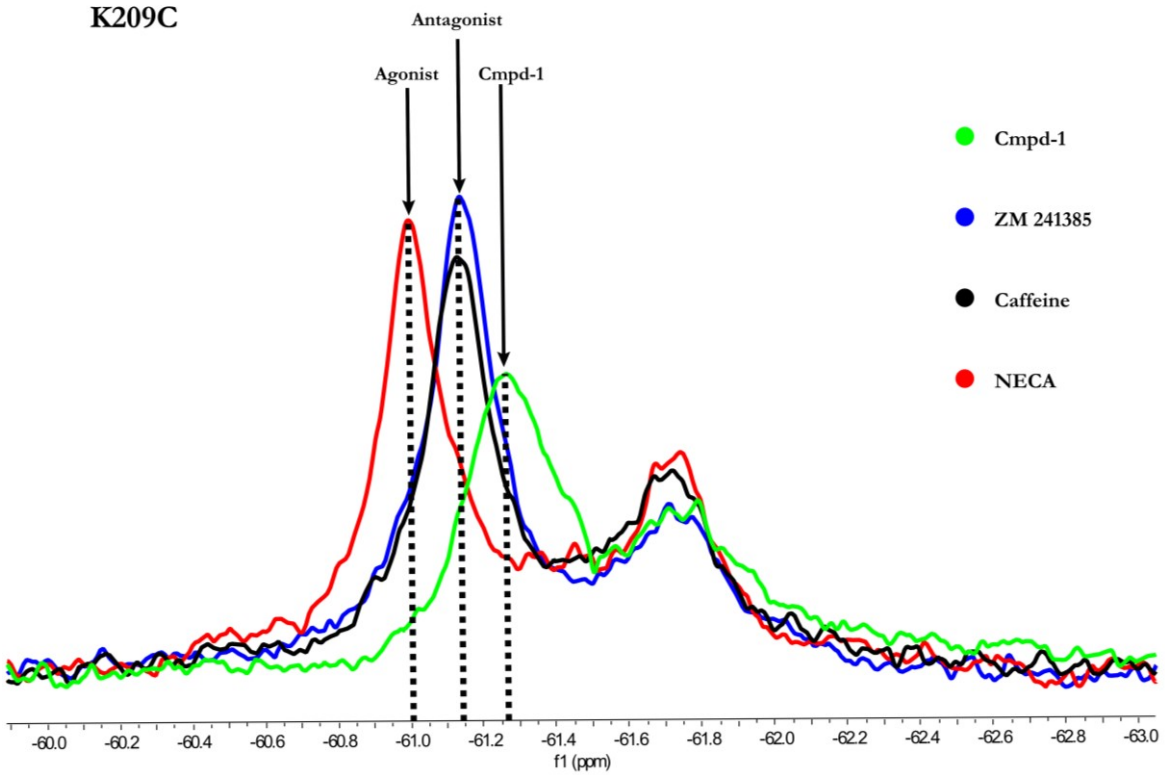


Figure 1.28 1D ^{19}F NMR spectra of $A_{2A}R$ K209C tagged with BTFMA in solution with antagonist ZM 241385 (Blue), antagonist caffeine (Black), antagonist Cmpd-1 (Lime), and agonist NECA (Red). The agonist sample shows a peak at around -60.09 ppm, the antagonists from the literature show a peak at -61.13 and Cmpd-1 shows a peak at around -61.26. This peak is also broader than the others.

Further dynamics experiments could distinguish these interpretations, but either interpretation points to the presence of at least one alternative inactive conformation of the receptor.

In parallel with this work 12 μs of MD simulations and Markov state modelling of the $A_{2A}R$ were carried out by Dr Silvia Lovera as discussed in our publication (Landin *et al*, 2019). These simulations categorised conformations of the $A_{2A}R$ bound to Cmpd-1 into five macrostates. Macrostates four and five (shown in Figure 1.29) are conformations of the ligand that are correlated with the outward movement of helix V. Macrostate four closely resembles the crystal structure, but macrostate five exhibits a more extreme outward movement of helix V concomitantly with the adoption of a metastable ligand binding site outside of the main orthosteric binding pocket. In combination these conformational states cause the outward helix V motion observed in the ^{19}F NMR experiments. The presence of only one resolved peak in the Cmpd-1 bound spectrum is consistent with the macrostates being in fast exchange on the NMR timescale resulting in a single peak, distinguishable from those of other antagonists. The importance of Cmpd-1's structure in forming the allosteric macrostate five interaction is demonstrated by chemical analogues with alternative methoxy substitutions, whose $A_{2A}R$ selectivity is abolished (Sun *et al*, 2017). The importance of allosteric sites in ligand selectivity has been seen more broadly in the field, such as in the case of the M2 muscarinic

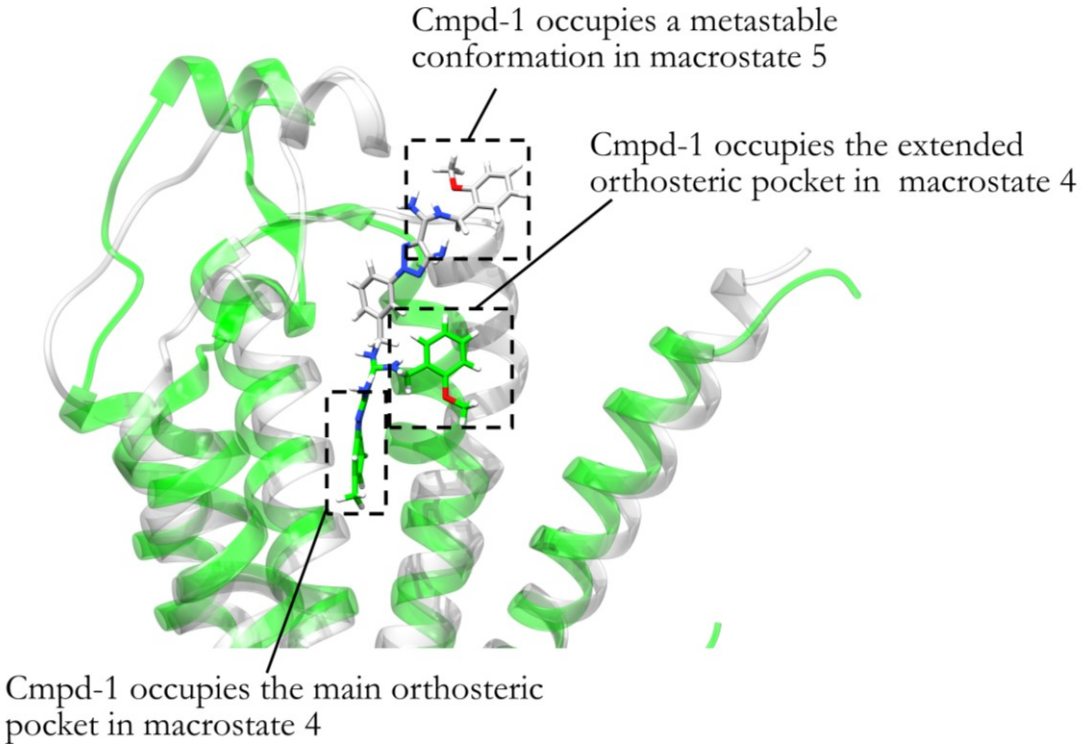


Figure 1.29 Representations of extracellular portion of $A_{2A}R$ showing macrostates five (light grey) and four (lime) causing the $A_{2A}R$ to adopt a conformation with an outward movement of helix V that was evidenced by the shift observed in the ^{19}F tag on K209C (out of frame below image). Cmpd-1 leaves the orthosteric pocket in macrostate five, which further increases the outward movement of helix V. In combination these states result in the alternative inactive conformation of the $A_{2A}R$ observed in the NMR spectra.

acetylcholine receptor (Haga *et al*, 2012). Residues important to allosteric binding to the M2 receptor have been identified and can mediate ligand selectivity for the M2 as opposed to the M4.

As stated above it was hypothesised that the concentration of LMNG/CHS used may influence the proportion of the receptor that populates different states, with our spectra appearing to show more of the inactive ensemble than those of Prosser and co-workers (Ye *et al*, 2016) . It should be noted at this juncture that detergent concentration is a difficult parameter to measure and to control. Although buffers used in membrane protein structural biology include a nominal detergent concentration this very often does not reflect the actual concentration that is present in the final sample. Two of the most recent NMR investigations of the $A_{2A}R$ (Eddy *et al,* 2018; Ye *et al,* 2016) report that sample concentration was the last step in their preparation protocol, being required to deliver a sample suitable for NMR. A potential pitfall of this method is that detergent/CHS molecules may exist as stable, empty micelles which cannot pass readily through a concentrator membrane and therefore become more concentrated concomitantly with the protein when using a spin concentrator. As a result, the detergent concentration may be higher than its nominal value. A further complication is that no convenient methods for detergent quantification on a par with protein quantification exist at present. Therefore, detergent concentrations may vary between experiments from different protein preparations without this being readily ascertainable. More recent work by the previously mentioned Prosser group (Ye *et al*, 2018) uses dialysis as the last step prior to NMR which

would bring the detergent concentration back down to its nominal value. In this study, sets of K209C spectra (including Figure 1.29) have been collected from samples of the same protein prep to ensure a constant detergent concentration even if the absolute value is unknown.

To further investigate the effect the effect of varying detergent on our data, NMR experiments comparing $A_{2A}R$ K209C spectra with bound ZM 241385 and Cmpd-1 were collected with additional detergent added after the protein was concentrated for NMR. These data are shown in Figure 1.30. Increasing concentrations of detergent cause the peak from the Cmpd-1 bound receptor to move towards (but not quite reach) the peak of ZM 241385-bound receptor.

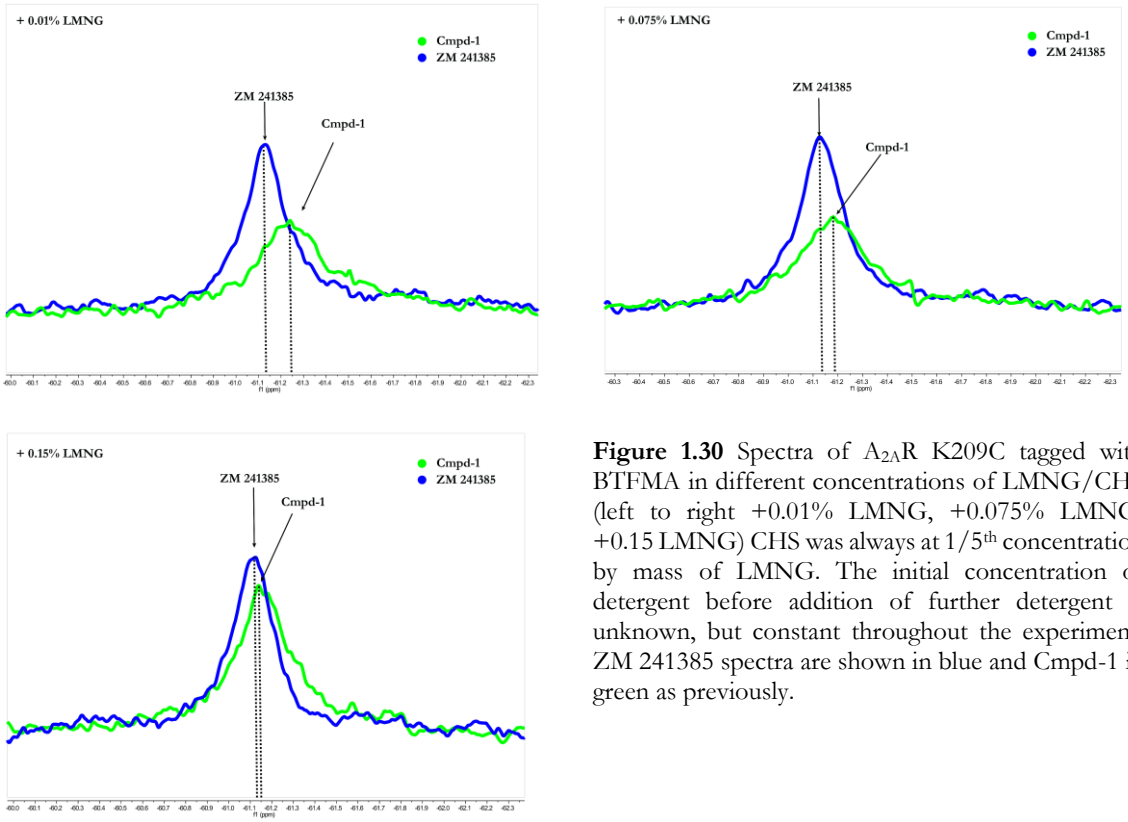


Figure 1.30 Spectra of $A_{2A}R$ K209C tagged with BTFMA in different concentrations of LMNG/CHS (left to right +0.01% LMNG, +0.075% LMNG, +0.15 LMNG) CHS was always at $1/5^{th}$ concentration by mass of LMNG. The initial concentration of detergent before addition of further detergent is unknown, but constant throughout the experiment. ZM 241385 spectra are shown in blue and Cmpd-1 in green as previously.

A variety of explanations for this may be suggested. It may be that the downfield shifted (less negative ppm values) signal represented a more detergent-exposed tag and that, therefore, the addition of more detergent into the solution causes the coalescence of solvent-exposed shifts into the detergent-exposed spectral region as the entire solvent environment becomes more homogeneously detergent-like. One other hypothesis is that it is not the detergent that is influencing the receptor, but the cholesterol. MD simulations have found both in the $A_{2A}R$ and other receptors that several cholesterol binding sites exist on the outer face of the 7-TM bundle (Genheden *et al*, 2017). Since cholesterol is likely to be concentrated as detergent micelle concentration increases it would be consistent with these data to infer that the increased cholesterol concentration in the NMR samples causes the receptor not to adopt the abnormal conformation of helix V to the same degree.

There may be other explanations of these data or the above hypotheses may each be true to some degree. Further experiments would need to be designed and carried out to discriminate between these possibilities. What is clear, however, is that the detergent environment selected for biophysical experiments in drug discovery targeting GPCRs have a profound effect on the outcome. If the detergent environment selected favours an inactive or active state, then the opposing effects of ligands may be supressed to the point that they cannot be observed.

In conclusion these data strongly support the existence of an additional inactive conformation of helix V that is more heavily populated in the presence of the putative allosteric ligand reported by Sun *et al.* These measurements are made in solution at physiological ionic strength and close to physiological temperature, without any fusion or insertion modifications to the protein. Although not explored further in this work, the K209C tagged A_{2A} mutant is a useful tool for further development of novel ligands derived from Cmpd-1; by enabling direct assessment of their allosteric effect on helix V conformational sampling, the K209C tagging site may bridge the gap between ligand binding and antagonistic function. It is also hoped that these data will encourage the further use of rationally designed ^{19}F tagging of novel sites in less tractable target proteins to resolve questions related to the drug discovery and lead development process.

Materials and Methods

Turnitin note: These methods build heavily on our publication of this work (Landin et al, 2019) and my previous progress monitoring reports for this PhD as they refer to the same published experiments that we carried out. Co-authors and collaborators are acknowledged in-text, in the acknowledgements section, and in this section where work was carried out jointly.

Talon Superflow Resin, HisTrap HP columns	GE Healthcare
[^{3}H] ZM241385 10-50 Ci/mmol, 1 mCi/ml, Ethanol:H_20 1:1	Biotrend (ART-0884-50)
Expifectamine 293 Transfection kit, iblot Gel transfer stack, SeeBlue protein ladder, Peroxidase-conjugated anti-His antibody, Nco1, Xho1, DNA digestion and ligase buffer, T4 DNA ligase	ThermoFisher Scientific
Protease Inhibitor Cocktail Tablets	Roche
Lauryl Maltose Neopentyl Glycol >98% purity	Anatrace
Cholesteryl Hemisuccinate >99% purity	Avanti
Theophylline >99% purity, Caffeine >99% purity, N-ethylcarboxamidoadenosine >98% purity, ZM241385 >98% purity, Deuterium Oxide >99% purity, Deuterium-matched Shigemi tube	SigmaAldrich
2-Bromo-N-[4-(trifluoromethyl)phenyl]acetamide >99% purity	Apollo Scientific
50 kDa Protein Spin Concentrator	Amicon
Glass Fibre Filters	PerkinElmer
4-12% Bis-Tris Gradient Gels	NuPAGE

A_{2A}R Constructs (vector was a proprietary UCB HEK293 cell expression vector):

The constructs consist of residues 1-316 of the wild type sequence with an N154Q mutation to remove a putative N154 glycosylation site. The mutants to be used for ^{19}F tagging contained one of the mutations K209C and V229C each. The constructs contain an N-terminal tStrep tag and C-terminal TEV-His_{10}. All sequences for initial constructs were ordered from DNA 2.0 (As it was called then). Cysteine mutations were introduced using a Quickchange Lightening mutatgenesis kit (Agilent). The linkers are commercially recommended sequences. The constructs were designed and ordered by Lisa Joedicke, UCB (Slough). Mutant constructs contained point mutations at the stated points in the sequences.

Wt:

MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAI
TISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSF
AIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLM
LGVYLRIFLAARRQLKQMESQPLPGERARSTLQKEVHAAKSLAIIVGLFALCWLPLHIINCFTFF
CPDCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKALEQKLISE
EDLNMHSGENLYFQGASPVATAVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATY
GKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFK
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI

KVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEF
VTAAGITLGMDELYKGSHHHHHHHHHH

tStrep Wt Construct:

MKTIIALSYIFCLVFAGSSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKSGENLYFQGSMPIMG
SSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAITISTGF
CAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFAIGLT
PMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLGVYL
RIFLAARRQLKQMESQPLPGERARSTLQKEVHAAKSLAIIVGLFALCWLPLHIINCFTFFCPDCS
HAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKALEENLYFQGSG
GSHHHHHHHHHH

Expression:

1 mg DNA per litre of cells was transfected into expiHEK 293 cells at $2.5*10^6$ cells/ml using standard protocols for ThermoFisher Scientific's Gibco ExpiFectamine 293 transfection kit. Enhancers 1 and 2 from the kit were added 16-18 h after transfection as per manufacturer's protocol. Cells were grown at 37 °C in 75% humidity, 8.0% CO_2 and shaking at 120 min^{-1}. The expression was harvested at 48 h after transfection. Cells were harvested by centrifugation (4000 g, 20 min, 4 °C), flash frozen and stored (-80 °C) until use.

Membrane Preparation:

Membranes were prepared by resuspension in ice-cold lysis buffer (25 mM HEPES, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 μM theophylline, pH 7.4 +protease inhibitor cocktail (PIC)) and two passes through a continuous cell disruptor (11.0 kpsi, 4 °C). Debris was removed by centrifugation (7000 g, 20 min, 4 °C). Membranes were spun down (187945 x g, 1 h 30 min, 4 °C), washed by re-suspension in ice-cold high salt buffer (25 mM HEPES, 500 mM NaCl, 1 mM EDTA, 100 μM theophylline, pH 7.4 +PIC), spun down again (187945 x g, 1 h 30 min, 4 °C) and re-suspended in ice cold storage buffer (PBS, 10% glycerol, 100 μM theophylline, pH 7.4, +PIC), before being flash frozen and stored.

A_{2A} Purification and Tagging:

The membranes were solubilised (0.5% LMNG, 0.1% CHS, 3 h, 4 °C) at 5 mg/ml total protein concentration determined by bicinchoninic acid assay (BCA). Insoluble matter was removed by ultracentrifugation (187945 x g, 1 h, 4 °C) and the supernatant was incubated with 1.2 ml buffer-equilibrated (50 mM HEPES, 300 mM NaCl, 10% glycerol + 0.01% LMNG + 0.002% CHS, 4 °C, pH 7.4)TALON Superflow bead slurry per litre cell expression for the total amount of membranes used overnight. The flow-through (FT) was separated on a 2 ml gravity flow column. The resin was washed with 10 CV buffer A (50 mM HEPES, 100 mM NaCl, 10% glycerol + 0.01% LMNG + 0.002% CHS, 4 °C, pH 7.4), 10 CV buffer B (50 mM HEPES, 800 mM NaCl, 20 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol + 0.01% LMNG + 0.002% CHS, 4 °C, pH 7.4), the beads were incubated with 4 CV buffer A + 100 μM TCEP (20 min, 4

°C) and the buffer was removed immediately at the end of the incubation and washed with 20 CV buffer A. The beads were incubated with 100 μM BTFMA (6 h, 4 °C), then with 150 μM fresh BTFMA (overnight, 4 °C) and washed with 10 CV buffer A (50 mM HEPES, 100 mM NaCl, 10% glycerol + 0.01% LMNG + 0.002% CHS, 4 °C, pH 7.4). Protein was eluted with 8 CV buffer E (50 mM HEPES, 100 mM NaCl, 10% glycerol, 200 mM imidazole, 4 °C, pH 7.4), concentrated as required (50 kDa cut-off, Amicon) and desalted in a 3 ml PD10 column (GE Healthcare).

TEV Digest:

The protein was treated with in house TEV protease (1:1 TEV to $A_{2A}R$ mass ratio, 4 h, 4°C) and subsequently incubated on 1 ml equilibrated TALON Superflow bead slurry per mg TEV overnight and the FT containing A_{2A} was collected. A_{2A} was desalted into buffer SG (50 mM HEPES, 100 mM NaCl, 0.01 % LMNG, 0.002% CHS, pH 7.4).

SDS PAGE:

All gels were Bis-Tris 4 - 12 % gradient gels run in MES (ThermoFisher) under reducing conditions. A_{2A} receptor was visualised by InstantBlue Coomassie stain (Expedeon) (> 1 h) with SeeBlue ladder or prepared for western blot by transfer to a polyvinylidene difluoride membrane on a gel transfer stack by standard protocols for iBlot (Thermo) kit (7 min). Western blots were visualised by Anti-His Peroxidase-conjugated monoclonal mouse antibodies (1:4000 dilution).

Coordinate Preparation for MD

MD coordinates were generated from crystal structures with the antagonist ZM 241385 [PDB: 5IU4], the agonist NECA [PDB: 2YDO] or the novel UCB antagonist Cmpd-1 [PDB: 5UIG] (Lebon *et al* 2011; Segala *et al*, 2016; Sun *et al*, 2017). Any cytochrome fusion proteins in ICL3 were removed and loops consisting of the physiological sequences were introduced. For antagonist-bound coordinates the ICL3 coordinates were derived from an antagonist bound structure with ICL3 present [PDB: 3VG9], with the surrounding residues aligned to those in the coordinate structures. For 5UIG and 2YDO there are no analogous ICL3 structures and so the loop was built using Modeller (Sali & Blundell 1993). All thermostabilising mutations in the crystal structures were reverted to the physiological sequence. The simulations were carried out in a POPC:POPE:cholesterol artificial bilayer in a ratio of 5:5:1 built by the CHARMM-GUI tool (Jo *et al*, 2009). The ^{19}F tag was built on using Chimera and hydrogen atoms consistent with pH 7 were added by CHARMM-GUI with the protein coordinates in the lipid bilayer generated by insertion. The protein coordinates were parameterised by the forcefield Amber 14SB and the lipid coordinates were parameterised using lipid14 in LEaP (Maier *et al*, 2015; Dickson *et al*, 2014). The simulation box had a length and width of 85 Å in the plane of the bilayer and a height perpendicular to the bilayer of 120 Å. The system was solvated in water and 0.15 M NaCl. ZM 241385, NECA and Cmpd-1 were parameterised with the generalised Amber forcefield (Wang *et al*, 2004).

MD Simulations

Simulations were run using a 2 fs timestamp with minimisation over 10 000 steps. Heating was carried out in two runs, from 0 to 100 K over 5 ps followed by 100 to 303 K over 100 ps. The box was equilibrated in 10 500 ps runs. Simulations were run for 125 ns each using the Langevin thermostat and Monte Carlo barostat. To generate full 625 ns data for each coordinate set the fresh trajectories were generated by applying the minimisation and heating steps described to the coordinates at 30, 60, 90 and 120 ns from the first 125 ns simulation and then running those for a further 125 ns. The trajectories were analysed using CPPTRAJ (Roe & Cheatham, 2013).

Radioligand Binding Assays on Membranes:

Membranes were prepared as for purification but without the inclusion of theophylline in the storage buffer. Membrane samples containing 10 μg total protein in 200 μl as calculated by (BCA) were incubated with a dilution series of [^{3}H]ZM 241385 with a final concentration of 0.1% DMSO in Jaakola *et al*'s buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 0.5 mM EDTA, pH 7.4), PerkinElmer's buffer (50 mM Tris-HCl, 0.5% BSA, pH 7.4) or Uustare *et al's* buffer (20 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, 1mM EDTA, pH = 7.4) as specified for 1 hour at room temperature (n = 3). To measure non-specific binding duplicate samples (n =2) were incubated in the same way with the addition of 20 μM ZM 241385. The reaction was rapidly stopped by washing with 50 mM Tris (pH 7.4) in a Skatron 1450 Microbeta Trilux onto glass fibre filters which were counted in a Perkinelmer Micro 96 plate reader.

Radioligand Binding Assays on Purified Receptor:

Purified $A_{2A}R$ in buffer SG was incubated at 0.02 mg/ml with 50 nM [^{3}H] ZM 241385 (n = 3) for 1 hour at room temperature and desalted before being counted in a PerkinElmer Tricarb 2910 TR Liquid Scintillation Analyser. Non-specific binding was measured by replicates of this experiment incubated with the addition of 20 μM ZM 241385 (n= 3) which were desalted and counted in the same way.

^{19}F NMR:

Samples of cleaved and tagged A_{2A} were made up at a concentration of 20 μM (measured by BCA) in buffer SG with 2% DMSO and the relevant ligand in a Shigemi tube. Samples were doped with 10% D_2O. All experiments were run on a Bruker AVIII HD spectrometer operating at a magnetic field strength of 14.1 T equipped with a 5 mm QCI-F Helium Cooled Cryoprobe. The experiments were run detecting ^{19}F with a spectral width of 138 ppm, 490 ms acquisition time, a recycle delay of 1 s, 15 000 scans, at a temperature of 298.2 K and processed with 15 Hz of line broadening. Chemical shift referencing was based on the instrument internal standard.

Chapter 2

The Binding of AGP2 to the Anti-Cancer Compound UCN-01

Introduction

The lipocalin protein family is a diverse set of proteins that exhibit an eight-stranded beta-barrel fold, and bind physiological ligands and a variety of exogenous drug-sized small molecules (Du *et al*, 2015; Masi *et al*, 2016). Retinol binding protein 4 (RBP4), also called plasma retinol binding protein, is a well-studied member of this family displaying the archetypal eight-stranded beta barrel structure which is shown in Figure 2.1 (Cowan, Newcomer & Jones, 1990).

RBP4 carries retinol from the liver to peripheral tissues whilst preventing its oxidation (Masi *et al,* 2016). The importance of the lipocalin in this process can be seen in RBP $^{-/-}$ knockout mice that can develop normal vision with a diet sufficiently high in retinol, but unlike RBP$^{+/+}$ cannot maintain normal retinal function after three weeks of a low-retinol diet (Quandro *et al*, 1999). Retinol levels in the liver remain high in RBP $^{-/-}$ knockout mice on a low retinol diet, indicating that RBP is critical for the mobilisation of retinol stores to peripheral tissue via the blood.

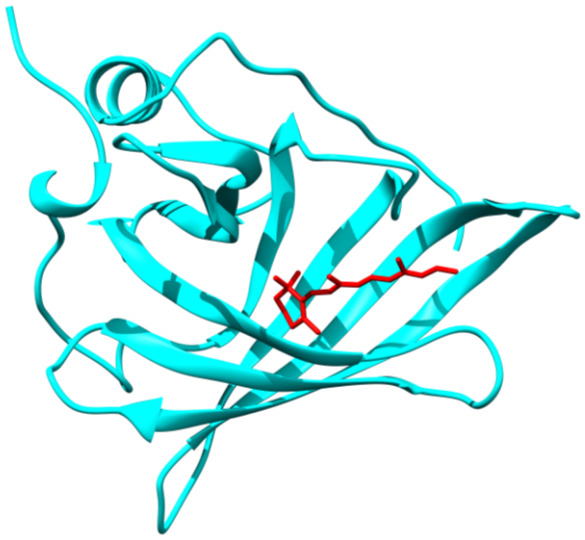


Figure 2.1 Ribbon representation of RBP4 in complex with all-trans retinol [PDB: 1RBP]. The ligand binding pocket is formed by an 8-stranded β-barrel (Cowan, Newcomer & Jones, 1990).

Other members of the lipocalin family bind exogenous ligands. A striking example of a lipocalin exhibiting this property is the histamine binding protein (HBP2_RHIAP) of the tick *Rhipicephalus appendiculatus* (Shown in Figure 2.2). This protein possesses a lipocalin fold and binds two histamine molecules in its cavity with sufficient affinity to outcompete the histamine receptors of the tick's host thus supressing the host's inflammatory response (Paesen *et al*, 1999).

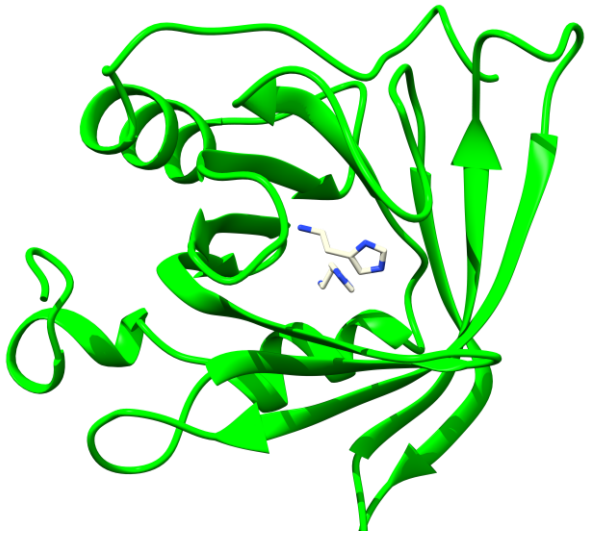


Figure 2.2 Ribbon representation of histamine binding protein [PDB: 1QFV]. Two histamine molecules are present in the ligand binding pocket (Paesen *et al*, 1999).

The human α-1-acid glycoproteins are the lipocalins that are the subjects of this study and possess binding activity against a broad range of exogenous ligands (Masi *et al,* 2016). These proteins are coded by the genes orosomucoid 1 and orosomucoid 2 to produce the F1/F2/S (AGP1) or A (AGP2) variants of α-1-acid glycoprotein respectively (Luo *et al,* 2015). AGP1 and AGP2 are also referred to as Orm1 and Orm2 respectively and their sequences are shown in Figure 2.3. There are around twenty amino acid differences between the two protein variants depending on the alleles compared. AGP1's three allelic variants differ by two amino acids (Q20 in variant F1 and F2, R20 in variant S; V174 in variant F1 and S, M174 in variant F2) (The Uniprot Consortium, 2019).

```
AGP1      MALSWVLTVLSLLPLLEAQIPLCANLVPVPITNATLDQITGKWFYIASAFRNEEYNKSVQ      42
AGP2      MALSWVLTVLSLLPLLEAQIPLCANLVPVPITNATLDRITGKWFYIASAFRNEEYNKSVQ      42

AGP1      EIQATFFYFTPNKTEDTIFLREYQTRQDQCIYNTTYLNVQRENGTISRYVGGQEHFAHLL      102
AGP2      EIQATFFYFTPNKTEDTIFLREYQTRQNQCFYNSSYLNVQRENGTVSRYEGGREHVAHLL      102

AGP1      ILRDTKTYMLAFDVNDEKNWGLSVYADKPETTKEQLGEFYEALDCLRIPKSDVVYTDWKK      162
AGP2      FLRDTKTLMFGSYLDDEKNWGLSFYADKPETTKEQLGEFYEALDCLCIPRSDVMYTDWKK      162

AGP1      DKCEPLEKQHEKERKQEEGES   183
AGP2      DKCEPLEKQHEKERKQEEGES   183
```

Figure 2.3 Sequence alignments of AGP1 variant F1 and AGP2 variant A. The F2 and S variants of AGP1 differ by Q20R and V174M respectively. Residues 1-18 (yellow) are cleaved to make the mature form of the protein and are not included in the count. The differences between AGP1 and AGP2 are highlighted (light blue).

AGP consists of 183 amino acids and adopts a lipocalin type fold based around an archetypal 8-stranded β-barrel structure. The protein is present in blood plasma at a concentration of ~2 μM and binds small molecule drugs such as methadone and endogenous ligands such as serotonin thus acting as a transporter (Masi *et al*, 2016). It has immunosuppressive activities that have been observed in cancer (Elg *et al*, 1997) and neuroinflammation (Jo *et al*, 2017). Although these physiological effects of AGP2 have been observed the link between the binding of natural ligands and AGP2 function is not understood (Bteich, 2019).

In addition to the immunomodulatory effects of AGP there is interest in the function and properties of both variants as they affect the volume of distribution of many therapeutics. Volume of distribution (V_D) is the theoretical volume of blood that would be required to hold an administered dose in its entirety at the blood plasma concentration. V_D is therefore a measure of drug binding to blood plasma. This can have a marked effect on the in vivo efficacy of a particular drug. For instance, the anti-cancer drug Imatinib (IMT) has been shown to enter cells at a reduced concentration in the presence of an elevated concentration of AGP (Gambacorti-passerini *et al,* 2003) and has been found to bind both AGP variants (with K_Ds of 1.25 and 50 μM for the S and A variants respectively) in vitro (Fitos *et al,* 2006).

Past X-ray crystallographic structural studies have focused on establishing the fundamental properties of AGP1 and AGP2 that determine their different ligand binding specificities. In 2008 the crystal structure of AGP1 (Figure 2.4(A)) was solved (Schönfeld *et al*, 2008) and it was followed in 2011 by the structure of AGP2 (Figure 2.4(B)) (Nishi *et al,* 2011). The structure of AGP2 was solved in the 'free' state to 2.10 Å, in complex with PEG, and also in complex with three ligands to a similar resolution: Disopyramide (DSP), amitriptyline (AMT) and chlorpromazine (CPZ). The AGP2 structures revealed that the protein had a smaller ligand binding pocket compared to that of AGP1, thus explaining AGP1's increased binding promiscuity. These structures also showed that certain ligands that were not specific for either variant (such as CPZ) bind via interactions to the conserved Phe-49 and hydrophobic Phe-112/Leu112 residues. Specificity for AGP2 is conferred by the replacement of the sterically occluding Phe114 in AGP1 by the smaller Ser114, which can mediate hydrogen bonds to the ligand, in AGP2.

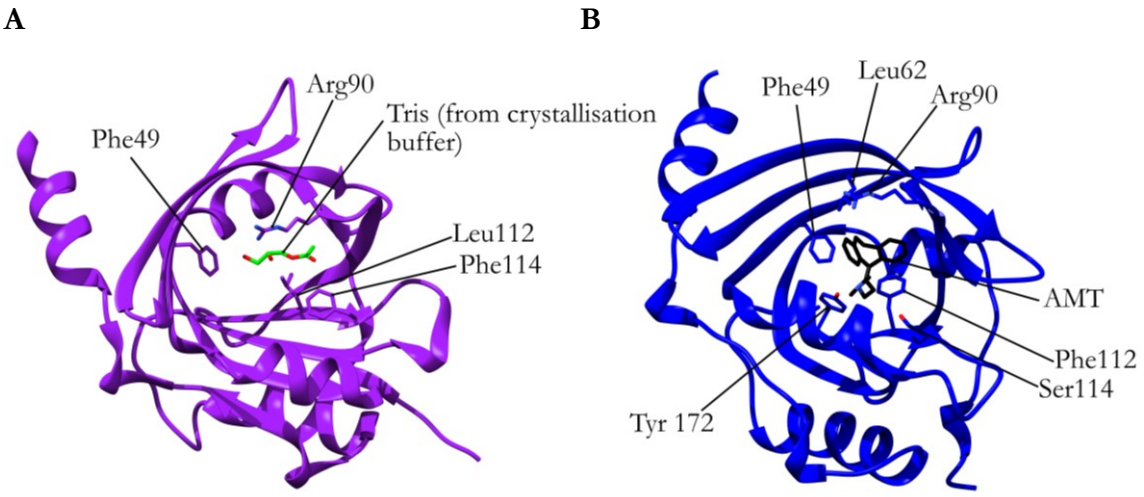


Figure 2.4 (A) AGP-1 [PDB: 3KQ0] in complex with dihydroxypropyl acetate (Tris buffer) (B) AGP-2 [PDB: 3APV] in complex with AMT revealing a highly similar architecture to that of AGP-1 with the AMT ligand binding in the same pocket as the tris buffer in AGP-1.

The structural studies of the AGP variants have not, however, yielded data on their binding to some of the most medicinally important ligands. IMT (mentioned above) is an inhibitor of the break point cluster-Abelson (Bcr-ABL) tyrosine-kinase that is constitutively active and created by a fusion of part of chromosome 9 onto chromosome 22. BCR-ABL produces a constitutive signal that drives cell growth and can be a key oncogenic marker (Savage and Antman, 2002). The existing structure of AGP1 has been used to conduct a molecular docking study on the binding of IMT to AGP1 as shown in Figure 2.5 (Fitos *et al*, 2012), however no experimental studies have confirmed the in silico results. This author wishes to thank Dr Ilona Fitos and Dr Agnes Simon for sharing the ligand coordinates from their docking study by private communication.

In the case of AGP2 the tightest known binders are derivatives of the natural product staurosporine, first isolated in 1977 (Omura *et al,* 1977). The 7(*R*)-hydroxylated derivative of straurosporine, UCN-01 shown

in Figure 2.6, was subsequently found to be a potent specific inhibitor of the protein kinase C (PKC) family, inhibiting all of it members to some degree, but PKCα most strongly (Mizuno *et al*, 1995).

The PKC family of proteins consists of 8 members all of which are promiscuous protein kinases. The PKC family are targets for therapeutic areas such as cardiovascular disease, autoimmune disease, and cancer (Mochly-Rosen, Das and Grimes, 2012). PKCα signalling is associated with cell proliferation making its inhibition a potential route to treat cancer. They each consist of regulatory region and a catalytic region. The regulatory region mediates intramolecular interactions that are relieved by binding to the activatory ligand diacyl glycerol (PKCα, PKCβI, PKCβII and PKCγ also require calcium binding for activation). The kinase region, and in particular its ATP binding site, is the target for a number of clinical compounds including UCN-01.

Figure 2.5 AGP-1 [PDB: 3KQ0] with imantinib docked into the binding pocket. The binding pocket is the same as that found with other ligands with the terminal biaryl in a similar position to the heterocycles of AGP2 ligands in that protein. This author thanks Dr Fitos and Dr Simon (Hungarian Academy of Sciences, Budapest) for sharing the docked coordinates from their publication (Personal communication; Fitos *et al*, 2012).

Following the interest in UCN-01 as a PKC antagonist it was found that the compound can also affect cell cycle control in two separate ways: Abrogating the G2/M checkpoint (Bunch and Eastman, 1996) and by arresting cells in G1 phase (Akiyama *et al*, 1997).

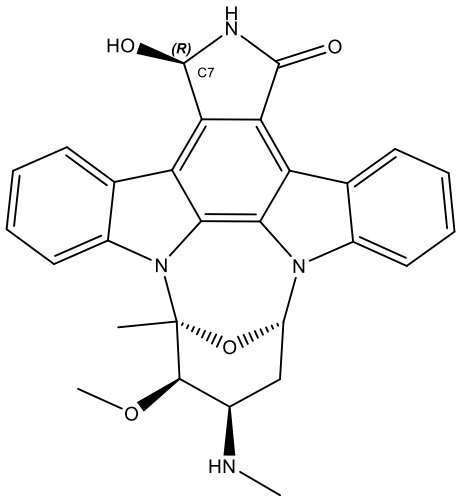


Figure 2.6 The structure of UCN-01. This hydroxylated analogue of staurosporine has been investigated as a cancer treatment due to its inhibitory activity against a broad range of kinases.

The abrogation of the G2/M checkpoint can be employed to increase the toxicity of DNA damage agents by a mechanism described schematically in Figure 2.7. When the cell suffers DNA damage during or immediately following DNA replication the cell cycle may be halted before mitosis by the set of proteins forming the G2/M checkpoint. In the absence of this checkpoint DNA damage that would normally have been repaired

by the cell can prove fatal as the cell attempts to complete mitosis with damaged DNA. Cancer cells are in general more susceptible to DNA damaging agents due to their higher rate of proliferation and genomic instability and so DNA damaging agents have been longstanding cancer treatments. By preventing DNA repair before mitosis UCN-01 increases the potency of these existing treatments (Bunch and Eastman,

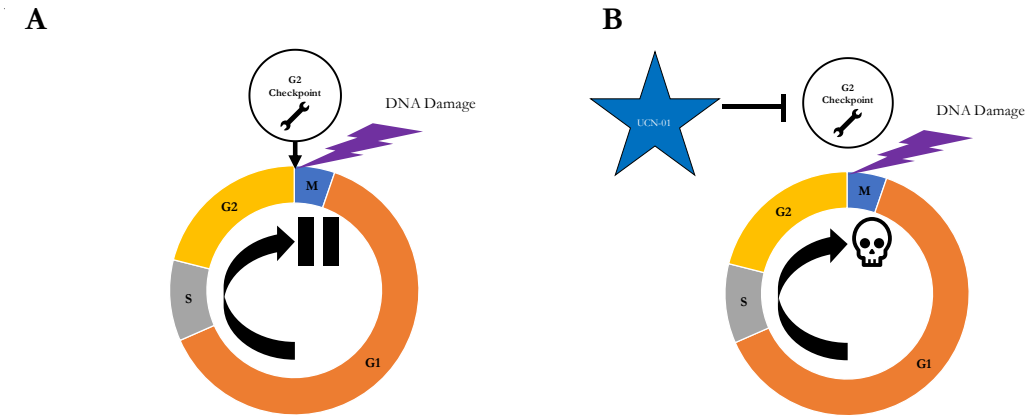


Figure 2.7 (A) If a cell's DNA is damaged (for example by a cytotoxic agent such as cisplatin) during proliferation the G2 checkpoint prevents entry into mitosis to allow the DNA to be repaired. (B) In cultured cells UCN-01 is shown to abrogate the G2 checkpoint and so allows cells with damaged DNA to proceed to mitosis leading to their deaths.

UCN-01's effect on the G2/M checkpoint is mediated by inhibition of checkpoint 1 (Chk1) kinase (Graves *et al,* 2000). A subsequent crystal structure, shown in Figure 2.8, of Chk1 kinase revealed the binding of UCN-01 in the catalytic ATP binding site of the kinase (Zhao *et al*, 2002).

Curiously, the other effect of UCN-01 is to arrest cells in G1 phase, preventing the transition into S-phase (Seynaeve *et al*, 1993). This effect is mediated by the product of the retinoblastoma gene, a tumour suppressor. The retinoblastoma protein inhibits transcription of S-phase promoting genes, but is itself inhibited by phosphorylation. UCN-01 maintains retinoblastoma protein in an active non-phosphorylated state by an unknown mechanism and so prevents cell cycle progression to S-phase (Akiyama *et al*, 1997).

This effect of UCN-01 provides an opportunity to target cancer cells, which have frequently lost retinoblastoma protein function (Dick and Rubin, 2013). A potential method to enhance the efficacy of cytotoxic anti-cancer treatments is shown schematically in Figure 2.9. By arresting predominantly healthy cells in G1, UCN-01 protects them from the effects of inhibitors of DNA synthesis whilst cancer cells are destroyed.

The strategy shown in Figure 2.9 was recently shown to be effective in a human xenograft murine model of breast cancer at improving the tolerated dose of the cytotoxic agent 5-fluorouracil (Mull *et al*, 2020).

Clinical trials on UCN-01 have investigated its potential as a treatment for lymphomas (National Cancer Institute, 2017) and as a dual therapy with topotecan to treat ovarian cancer (Welch *et al*, 2007). The former study was terminated due to funding issues and the latter failed to demonstrate sufficient efficacy.

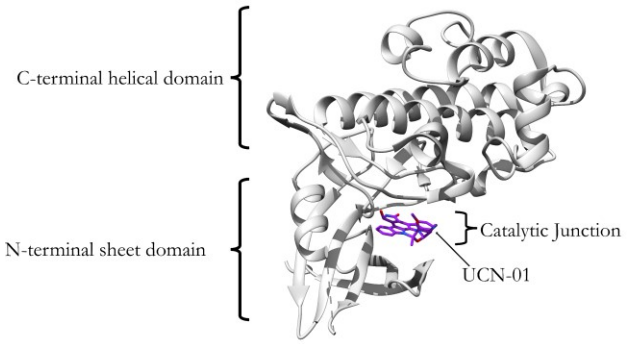


Figure 2.8 The structure of Chk1 kinase in complex with UCN-01 [PDB:1NVQ]. Inhibition of this kinase prevents G2/M checkpoint cell cycle arrest.

Given the drastic effect of AGP2 on UCN-01 pharmacokinetics, it is worthwhile to investigate this interaction as it is likely to be a major hinderance to the efficacy of UCN-01 (Fuse *et al*, 2005).

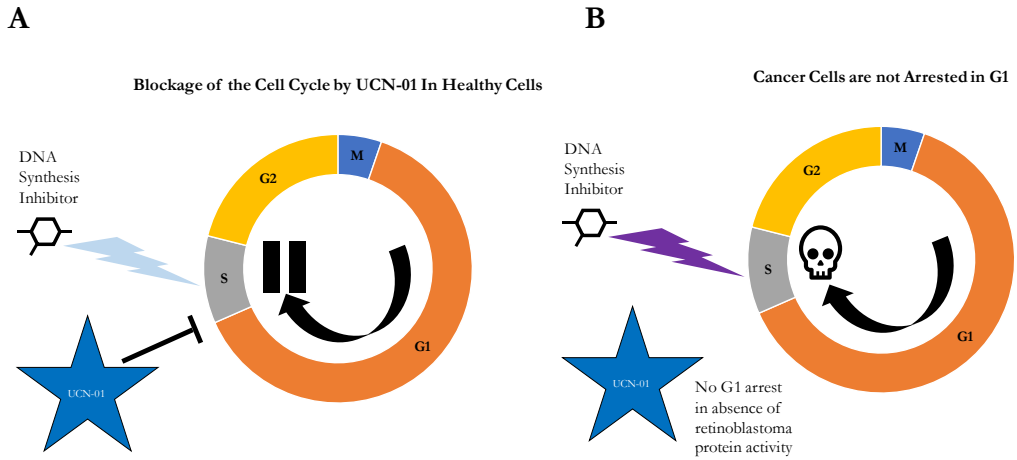


Figure 2.9 (A) Healthy cells are arrested in G1 phase and so are not as sensitive to cytotoxic agents targeting DNA replication. (B) Retinoblastoma protein-deficient cancer cells are not arrested in G1 and so attempt to replicate their DNA in the presence of cytotoxic agents and are killed.

The underlying interest in UCN-01 as a cancer therapy makes understanding its molecular interaction with AGP timely. This understanding could, for example, open opportunities to modify UCN-01 in a way that eliminates AGP2 binding without affecting the interaction with its kinase targets. This might subsequently have an important effect on the compound's therapeutic potential. Furthermore, the structural information gained from this study may be used to inform future drug discovery efforts to avoid creating compounds with high affinity for AGP.

This chapter describes optimisation of bacterial expression and purification of recombinant AGP2 which has led to the production of milligram quantities of pure ^{15}N labelled and $^{15}N/^{13}C$ dual-labelled AGP2. Isotope-labelled AGP2 was subsequently used to generate multi-dimensional NMR spectra which were sufficient for assignment of the protein backbone and significant portion of the side chains. These data were followed by a crystal structure of AGP2 in complex with UCN-01, characterising for the first time the binding of this important ligand to AGP2.

Trial Expression of AGP2 in Origami 2 Cells

The initial aim was to establish a robust in-house protein expression system that could be used for producing isotopically labelled AGP2. AGP2 is a eukaryotic protein containing 2 disulphide bonds and is extensively glycosylated in vivo, with glycans making up as much as 45% of its mass (Masi *et al*, 2016). *E. coli* lack the endomembrane systems of eukaryotes to assist in disulphide bond formation and are not capable of producing eukaryotic glycosylation patterns, however previous crystallographic studies on both AGP variants have successfully produced sufficient folded protein that bound to known ligands in prokaryotic expression hosts and so these problems were not perceived to be insurmountable (Shönfeld *et al*, 2008; Nishi *et al*, 2011).

For previous structural studies, AGP2 was expressed in the cytoplasm of Origami B cells (Nishi *et al*, 2011). Origami cells carry mutations in the thioredoxin reductase (*trxB)* and glutathione reductase (*gor*) genes (Merck, 2019); this results in a more oxidising cytoplasmic environment than that of wt *E. coli* cells or those of other expression strains facilitating disulphide bond formation. This approach allowed Nishi *et al* to produce pure protein at a final working concentration of 18 mg/mL from which it follows that they must have been successful in producing protein quantities on the order of 2 mg to carry out any significant condition screens (one 96-well screen would require 1.7 mg of protein assuming 1 µl drops of protein were used) . A 50 µl sample of 1 mM AGP2 would be sufficient on the instrument available for this study (see methods) and would require 1.1 mg of AGP2. It was therefore concluded that an approach similar to that of Nishi *et al* might be successful.

AGP2 expression was first attempted using Origami 2 cells. This cell line is closely related to the cell strain used in the literature (differing in antibiotic resistance markers), making it a logical starting point (Nishi *et al*, 2011). Full-length (1-183) and short (1-175) versions of AGP2 were synthesized in a pMA-T vector (Thermo Fisher) and transferred by restriction cloning into a pET28 vector. The shortened version was purchased in case a smaller construct was necessary to facilitate NMR assignment later in the project. The crystal structures [PDB: 3APU; 3APV; 3APW; 3APX] indicate that residues after 175 do not form part of the main protein structure and are likely to be dispensable without affecting the protein fold. The AGP2 constructs and pET28 receiver vector were cleaved by Nco1 and Xho1 in a double digestion as shown in the gel image in Figure 2.10. AGP2 was extracted from the gel and ligated into the pET28 receiver vector and amplified in the high copy number strain NEB5α (New England Biolabs). The integrity of the full length AGP2 construct was confirmed by forward sequencing prior to transformation into the Origami 2 cells. AGP2 was expressed and purified from the soluble cell lysate by IMAC. The results were analysed by SDS PAGE and visualised by Coomassie stain and Western blot (Fig. 2.11).

Very little if any AGP2 was detectable on the Coomassie-stained SDS PAGE, but after anti-His Western blotting a significant band consistent with the expected mass was observed. Given that other bands of similar intensity to the putative AGP2 band were observed in the SDS PAGE analysis and the total protein produced was less than 1 mg it was determined that this method of producing Orm2 was sub-optimal.

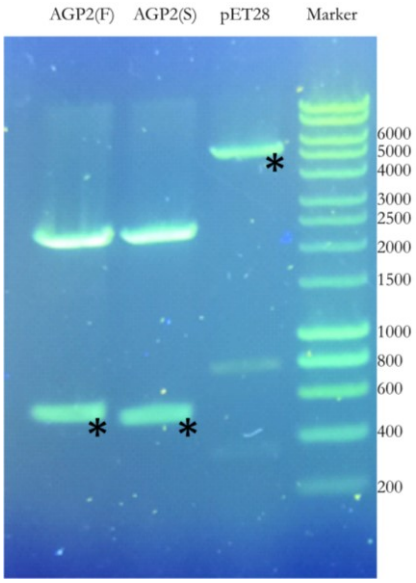


Figure 2.10 Agarose gel of double digestion of AGP2 constructs and the pET28 expression vector. Clear bands are shown for the AGP2 full and short constructs consistent with the expected sizes of 557 bp and 555 bp respectively. pET 28 migration was consistent with the empty vector size of 5231 bp.

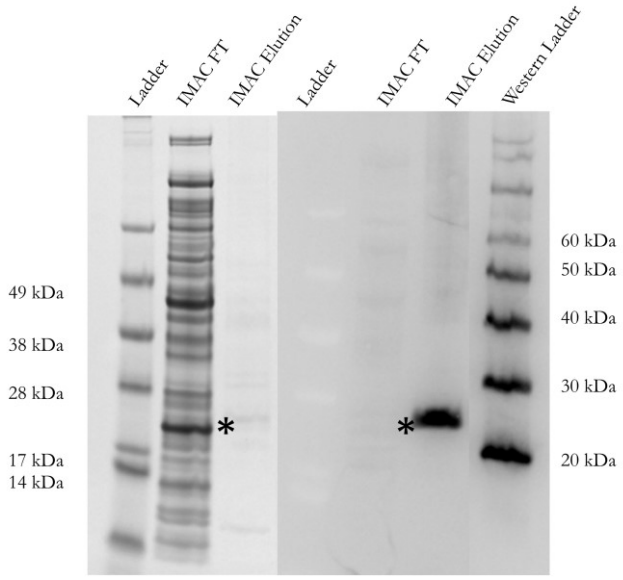


Figure 2.11 SDS PAGE visualised with Coomassie stain (left) and horseradish peroxidase (HRP) linked anti-his antibody after western blotting (right). A very faint band may be detectable by Coomassie staining the identity of which may be confirmed as His_6-AGP2 by the Western blot.

Cloning and Expression of New AGP2 Constructs

The initial expression of AGP2 in the pET28 vector in Origami2 cells did not produce protein in sufficient quantity for efficient purification and subsequent biophysical or NMR studies. Consequently, new expression vectors were explored in efforts to produce sufficient quantities of protein; three strategies were chosen:

- Expression of AGP2 in the bacterial periplasm in a pET26 vector
- Expression of AGP2 in the bacterial cytoplasm with the solubility-enhancing tag N utilisation substance A (NusA) using the pETM60 vector which incorporates a TEV cleavage site for post expression removal of the NusA
- Expression of AGP2 in the bacterial cytoplasm with an N-terminal His_6 tag in a pOPINF vector

These strategies may each respectively be justified as follows:

- Expression in the bacterial periplasm allows the protein to fold in a more oxidising environment assisting in disulphide bond formation, and has been applied successfully to express AGP1 (Schönfeld *et al,* 2008).
- Expression of recombinant protein with a NusA fusion, is an established method to improve both protein solubility and stability. The NusA protein works by recruiting bacterial chaperones and increasing the solubility of the fusion construct (De Marco *et al,* 2004).
- Expression of recombinant protein with an N-terminal His_6 tag rather than a C-terminal addition has been shown to improve expression in some cases (Park *et al*, 2015)

To carry out these strategies the three new constructs first had to be generated.

Generation of 3 New Constructs for Fresh Expression Strategies – pET26, pETM60 and pOPINF

The full AGP2 construct was inserted into the pET26 and pETM60 vectors by restriction enzyme cloning techniques, as before with the pET28 vector, using the Nco1 and Xho1 restriction sites, and ligation with T4 ligase. The double digestion of the AGP2 construct and the pET26 and pETM60 vectors are shown in Figure 2.12. A strong band was present at around 550 kb in each of the AGP2 donor digestions indicating that the construct was cleaved out of the pMA-T vector and strong bands were present at above 5.0 kb in the receiver vector digestions indicating that these vectors had been successfully cut and may receive the AGP2 construct. AGP2 was prepared for insertion into the pOPINF vector by PCR addition of adaptor sequences (5' AAGTTCTGTTTCAGGGCCCG, 3' ATGGTCTAGAAAGCTTTA) across a screen of 5 temperatures as shown in Figure 2.13. A band of around 550 kb, consistent with the AGP2 construct, was present amplification in the range of temperatures tested. These data indicated that all of the temperatures used were suitable. The AGP2 PCR product was inserted into cut pOPINF by homologous recombination by In-Fusion® Enzyme (Takara).

After verification by sequencing AGP2 expressions were carried out for each of these constructs.

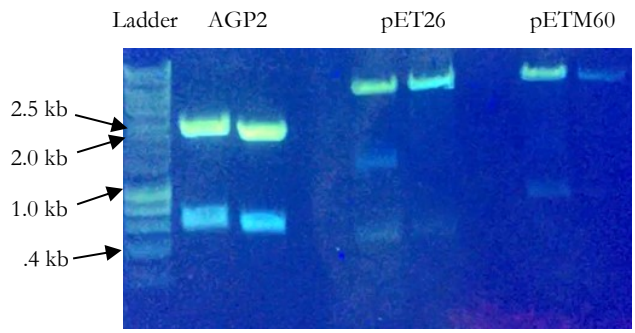


Figure 2.12 Agarose gel of duplicated double digests of AGP2 in pMA-T, pET26 and pETM60. Clear bands are visible for the vectors and AGP2 inserts

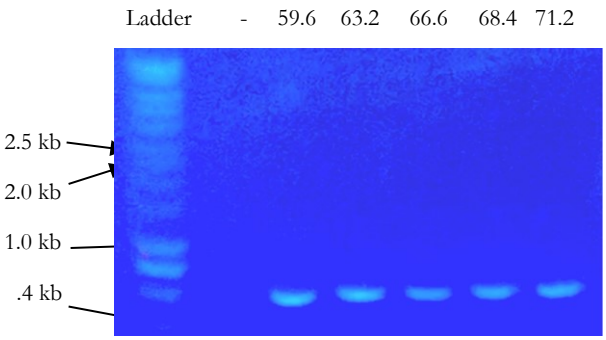


Figure 2.13 Agarose gel of the AGP2 construct amplified and modified to contain pOPINF adapter sequences by PCR across a range of temperatures (in °C)

Periplasmic Expression of AGP2 in pET26

pET26-AGP2 was expressed in BL21 (DE3) cells and a standard periplasmic compartment purification protocol was carried out as shown in the SDS PAGE image of Figure 2.14 and in the UV trace of the IMAC protocol in Figure 2.15. This involved applying hypoosmotic and hyperosmotic buffer to the cells to cause the outer membrane of the cells to rupture and release the recombinant AGP2 into the buffer solution (Schönfeld *et al,* 2008).

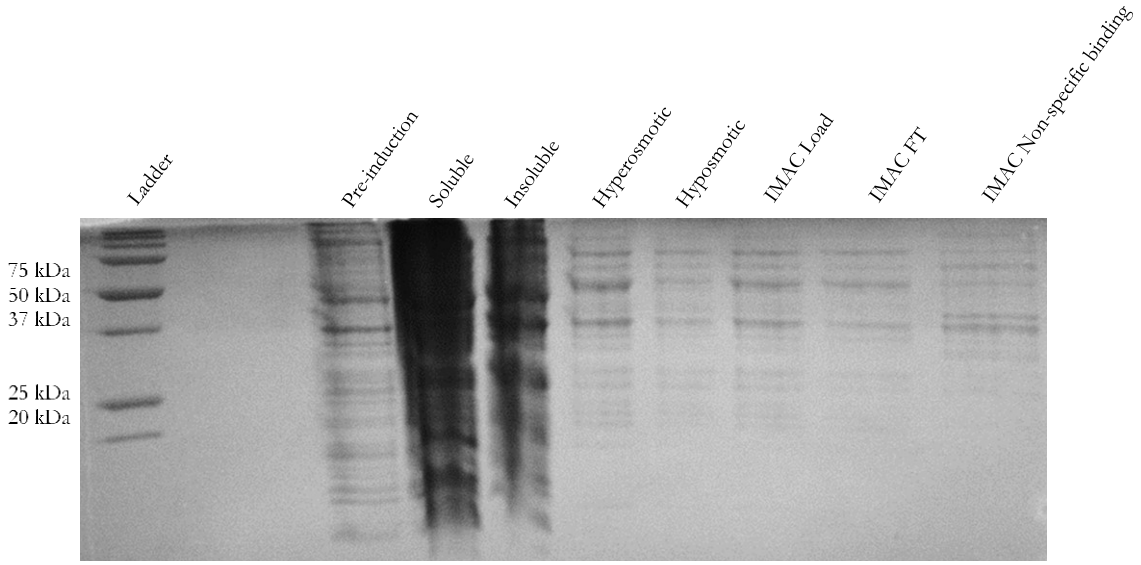


Figure 2.14 Coomassie-stained SDS PAGE of an expression of pET26-AGP2. There is no significant band at the expected mass (~23 kDa) and therefore no evidence of AGP2 expression

There was no evidence of a significant band at the expected mass (~ 23 kDa) in the SDS PAGE nor any significant UV absorption consistent with bound protein in the IMAC elution. As there was no evidence of AGP2 expression the pET26 construct was not used further.

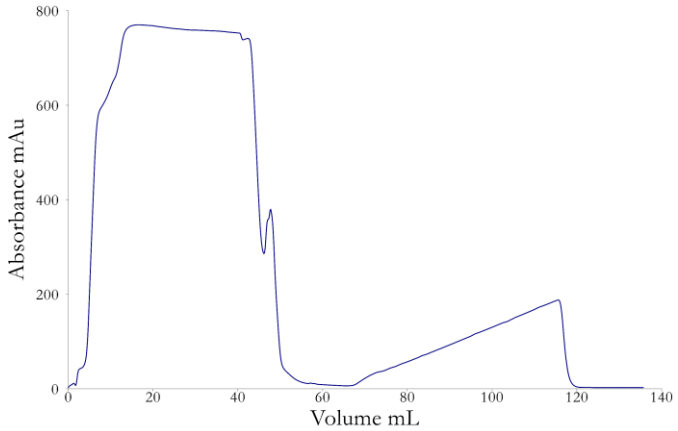


Figure 2.15 UV absorption (280 nm) trace from IMAC fractionation of the pooled hyperosmotic and hypoosmotic fractions of the cell lysis.

Cytoplasmic Expression of NusA-AGP2 Fusion Protein in pETM60

NusA is a highly soluble 50 kDa protein that was expressed upstream of the N-terminus of AGP2 in this construct to assist in the expression of soluble AGP2. Expression and purification of AGP2 with the NusA solubility enhancing tag in the pETM60 vector was carried out and the results were analysed by SDS PAGE as shown in Figure 2.16. A significant band consistent with the expected mass of the construct (~75 kDa) was visible in each of the crude insoluble and soluble cell fractions. The IMAC UV trace (Figure 2.17)

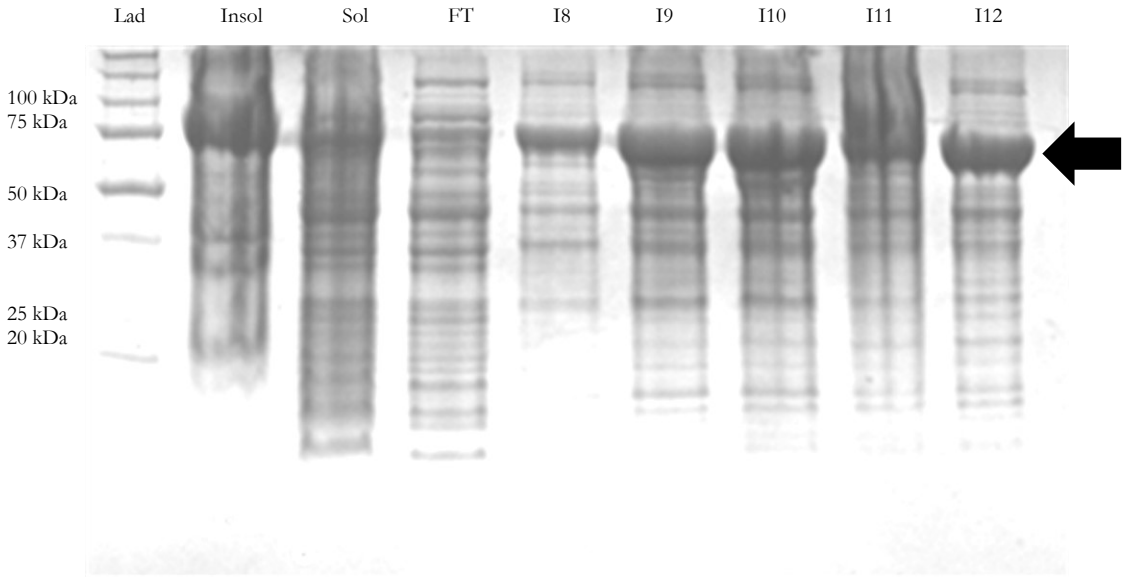


Figure 2.16 Coomassie-stained SDS PAGE of NusA-AGP2(arrow) IMAC purification. There is evidence of expression of the desired protein in the insoluble, soluble and IMAC fractions.

showed a large peak (above the limit of the instrument's UV absorbance detection) in the protein elution. The five fractions I8-I12 of this peak are shown in the SDS PAGE and the same band as in the crude fractions at ~75 kDa consistent with NusA-AGP2 was visible.

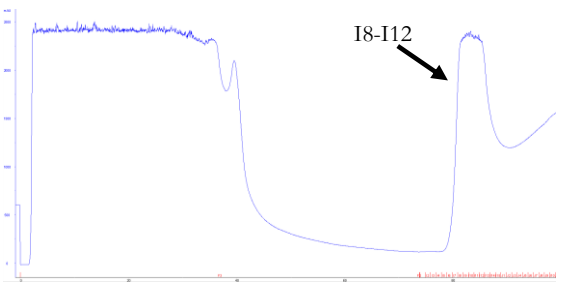


Figure 2.17 UV absorption trace from IMAC of NusA-AGP2 expressing BL21 lysate.

Subsequently the NusA-AGP2 fusion was treated with in-house TEV protease for 16 hours at 4 °C or room temperature and the resulting protein samples were compared to His-AGP2 prepared insolubly (see next section). The results were analysed by SDS PAGE as shown in Figure 2.18. The previously observed band at 75 kDa corresponding to NusA-AGP2 was replaced by a band around 50 kDa, consistent with cleaved NusA. A small band also appeared at around 25 kDa consistent with cleaved AGP2-His_6 strongly suggesting that the NusA-AGP2 protein was successfully cleaved in both reaction conditions.

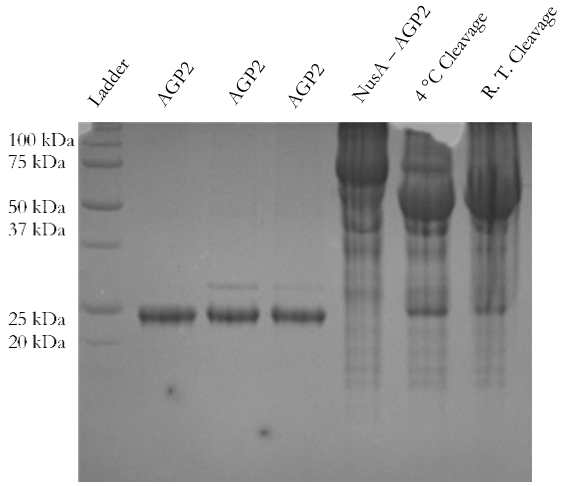


Figure 2.18 Coomassie-stained SDS PAGE image of AGP2 samples compared with NusA-AGP2 before and after TEV cleavage.

To separate AGP2-His_6 from NusA, gel filtrations were carried out on TEV cleavage products of NusA-AGP2 in the presence or absence of a cystamine-cysteamine redox shuffle to allow disulphide bond exchange (Figure 2.19). In both cases significant peaks were observed between an elution volume of 100 and 150 ml-post injection consistent with aggregated protein. Small peaks (A7 and C1 in Figure 2.19) were observed consistent with monomeric protein, however SDS PAGE analysis (shown in Figure 2.20) indicated that the aggregate peaks were composed of NusA and AGP2 cleaved from each other, whereas the fractions consistent with monomeric protein (A7, C1) were composed only of NusA.

From these data it was concluded that whilst NusA successfully brought AGP2 into solution it was nevertheless either improperly folded or partially denatured and therefore aggregated when NusA was removed. Accordingly, the NusA-AGP2 construct was not investigated further.

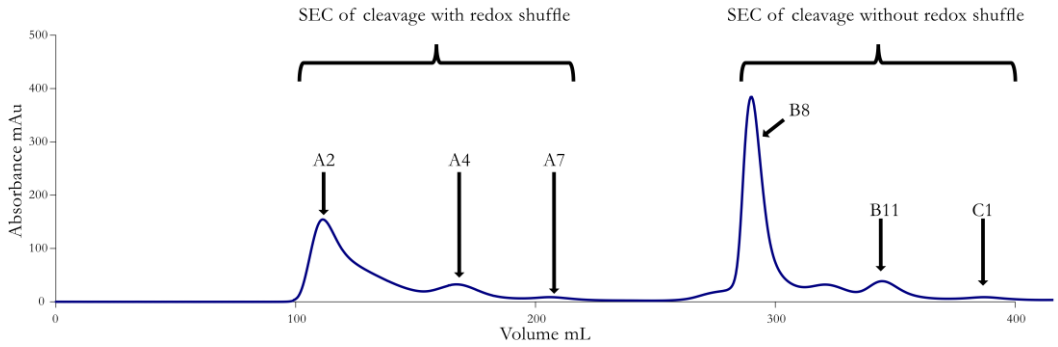


Figure 2.19 UV absorption (280 nm) trace of two in-series gel filtrations of NusA-AGP2 cleavage products with redox shuffle (0-200 ml) and without redox shuffle (200-400 ml).

Although insoluble AGP2 was produced using the pETM60 vector a full protocol to express insoluble AGP2 was developed using the pOPINF vector instead, as the smaller N-terminal His_6 tag expression theoretically should consume fewer cellular resources than expressing a full 55 kDa NusA fusion partner and so can be expected to result in higher yields of AGP2.

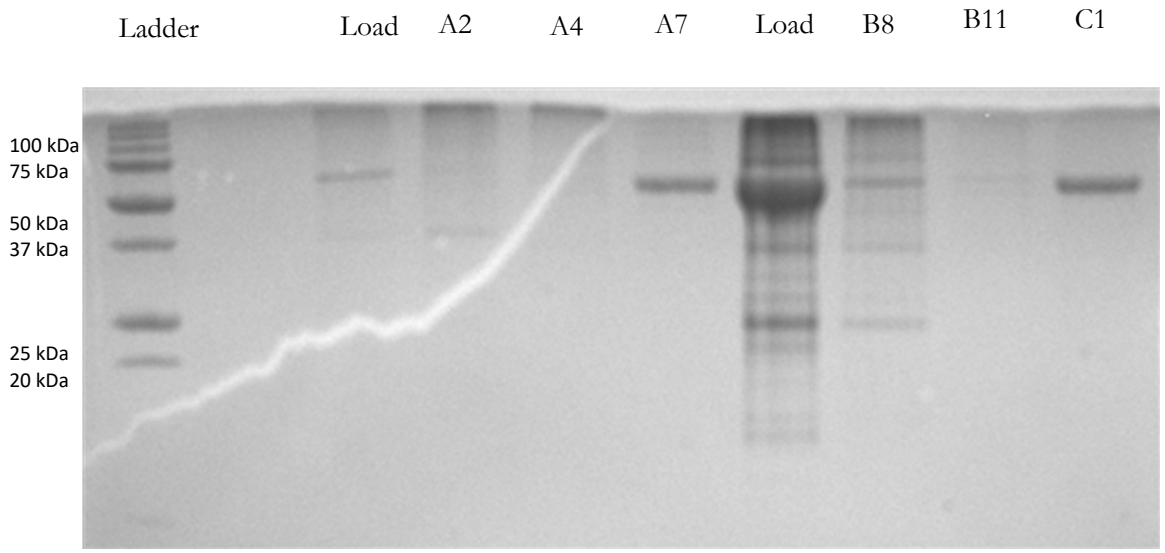


Figure 2.20 Coomassie-stained SDS PAGE of NusA-AGP2 cleavage gel filtration. Bands consistent with NusA are present in both the load and monomeric fractions, but bands for cleaved AGP2 are only present in fractions consistent with aggregates.

Cytoplasmic expression of His_6-AGP2 in pOPINF

AGP2 inserted into the pOPINF vector was expressed in BL21 cells and the lysate fractions were analysed by SDS PAGE as shown in Figure 2.21. A clear band at the expected mass (~23 kDa) was visible in the insoluble fraction of the lysis. This result was consistent with AGP2 being produced as in insoluble inclusion bodies.

The putative AGP2 inclusion bodies were purified by an inclusion body prep, solubilised in 10 M urea buffer, and reduced in preparation for refolding by rapid dilution in 1M arginine containing buffer. The refolded preparation was concentrated, and folded protein was separated from aggregate by gel filtration as shown in Figure 2.22(A). The inclusion bodies, refold mixture, and gel filtration peak were each analysed by SDS PAGE as seen in Figure 2.22(B).

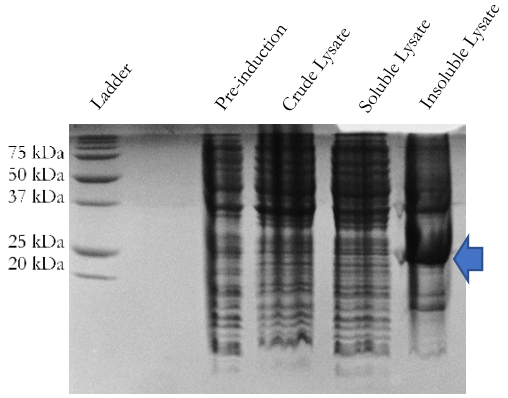
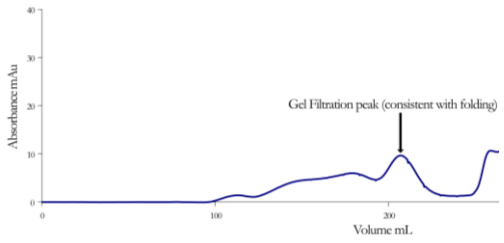


Figure 2.21 Coomassie-stained SDS PAGE of pOPINF-AGP2 expressed in BL21 cells shows a clear band in the insoluble lysate fraction consistent with AGP2 inclusion bodies

A



B

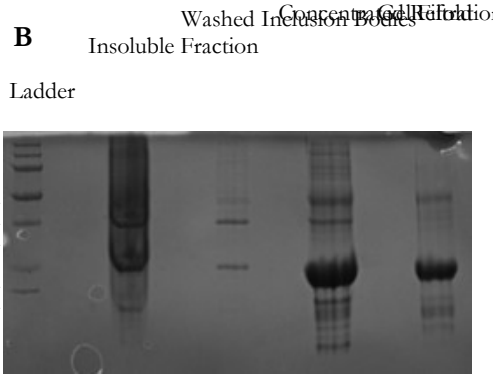


Figure 2.22 (A) Refolded inclusion body prep was separated by gel filtration and a peak UV peak at 280 nm consistent with monomeric protein was observed at an elution volume of 202 ml-post injection, but further validation of identity and correct folding was necessary. **(B)** Coomassie-stained SDS PAGE of the concentrated gel filtration peak compared with the inclusion bodies before gel filtration, the washed inclusion bodies, and the insoluble cell fraction.

To confirm the identity of AGP2, the monomeric protein was analysed by intact mass spectrometry and the resultant spectrum is shown in Figure 2.23. The expected mass of the AGP2 construct, assuming cellular cleavage of the N-terminal methionine, is 23626 Da, however a peak of mass 23529 Da was dominant in the de-convoluted spectrum.

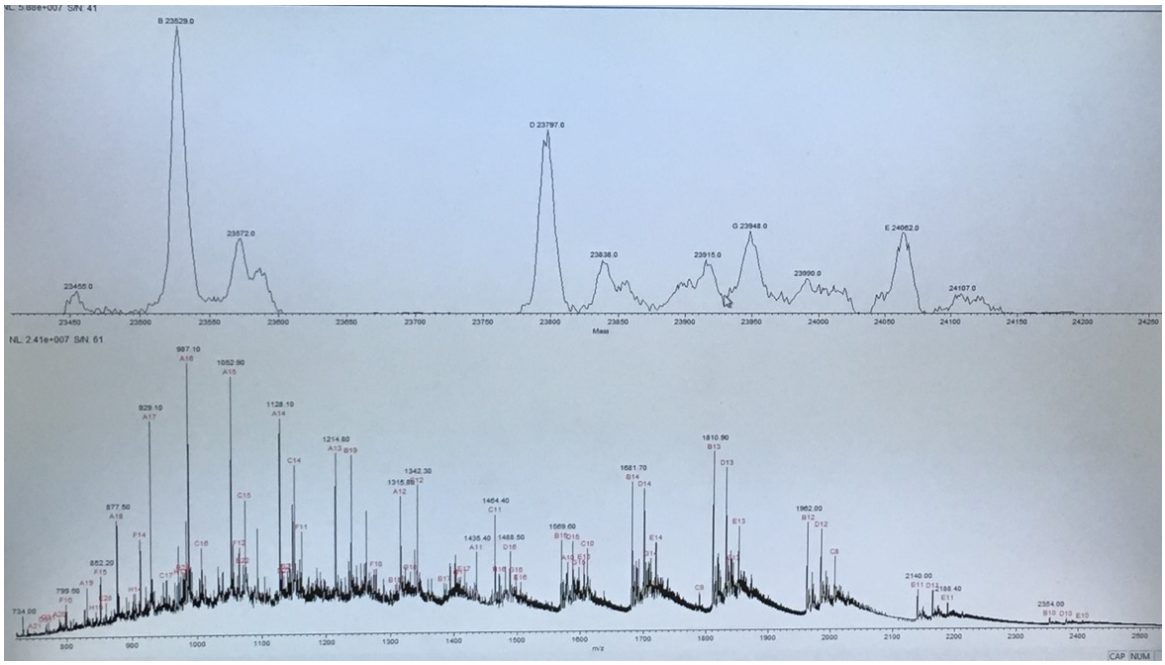


Figure 2.23 Deconvoluted (top) and raw (bottom) spectra of concentrated monomeric protein from gel filtration in electrospray M/S. The main peak at 23529 Da is consistent with a mutated form of the desired construct (expected mass = 23523 Da). Several other peaks that may be due to adducts or impurities are present.

Sequence verification of the pOPINF construct revealed an R->G mutation. This mutation accounted for the mass discrepancy as the mutated construct would yield AGP2 protein with expected mass of 23527 Da which was consistent with the observed mass of 23529 Da. These data confirmed pOPINF-AGP2 as the expression vector of choice for further protein expression.

Optimisation of pOPINF-AGP2 Expression

A replicate of the pOPINF AGP2 homologous recombination insertion, confirmed by N-terminal sequencing to lack the R->G mutation, was chosen to express a fresh sample of AGP2. The construct was expressed in BL21 cells and inclusion bodies were prepared as previously. The protein from the inclusion bodies was refolded by rapid dilution in 1 M arginine-containing buffer, concentrated, and refolded protein was separated from aggregates by gel filtration as shown in Figure 2.24.

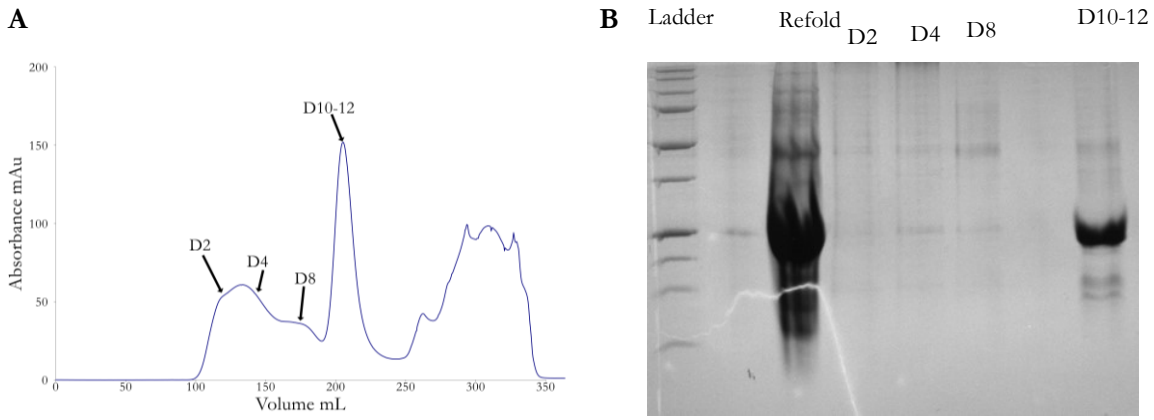


Figure 2.24 (A) UV (280 nm) trace of gel filtration of refolded AGP2 showing a dominant peak consistent with monomeric protein. **(B)** Coomassie-stained SDS PAGE of the AGP2 refold and gel filtration peaks. A significant band is present at around 23 kDa consistent with AGP2

By way of comparison pOPINF-AGP2 was also transformed in SHuffle® (New England Biolabs) cells. This strain is related to the Origami strain, but in addition to the mutations in the thioredoxin reductase (*trxB)* and glutathione reductase (*gor*) genes the SHuffle® strain contains a constitutively expressed chromosomal copy of DsbC – a disulphide bond isomerase (New England Biolabs, accessed 2019). DsbC is usually expressed in the bacterial periplasm, but in this strain DsbC lacks a signal sequence and so is retained in the cytoplasm where it can assist in refolding of incorrectly oxidised disulphides.

SHuffle® cells were grown in rich media and AGP2 was purified by IMAC from the soluble lysis fraction as shown in Figure 2.25. The IMAC elution profile showed 2 distinct peaks and SDS PAGE analysis indicated that the second peak contained a protein of a mass consistent with that His_6-AGP2 (~24 kDa). No significant band of a similar size was detectable in the other fractions suggesting AGP2 is expressed at low levels.

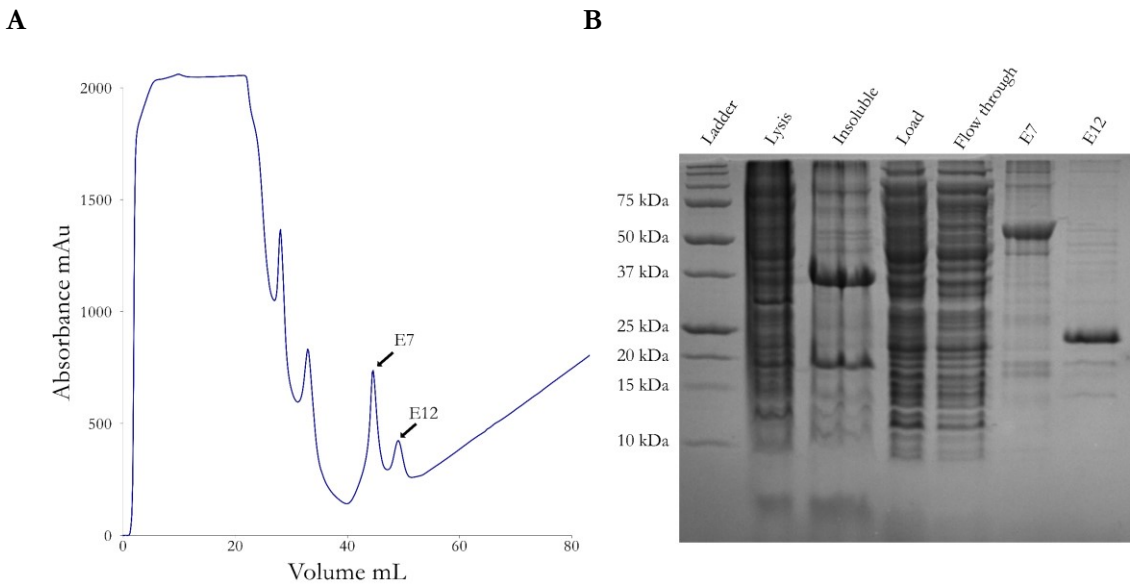


Figure 2.25 (A) UV (280 nm) trace of IMAC purification of AGP2 from the soluble lysate. **(B)** Coomassie-stained SDS PAGE of the SHuffle® cell lysis fractions and IMAC fractions. A band consistent with His_6-AGP2 is observed in fraction E12

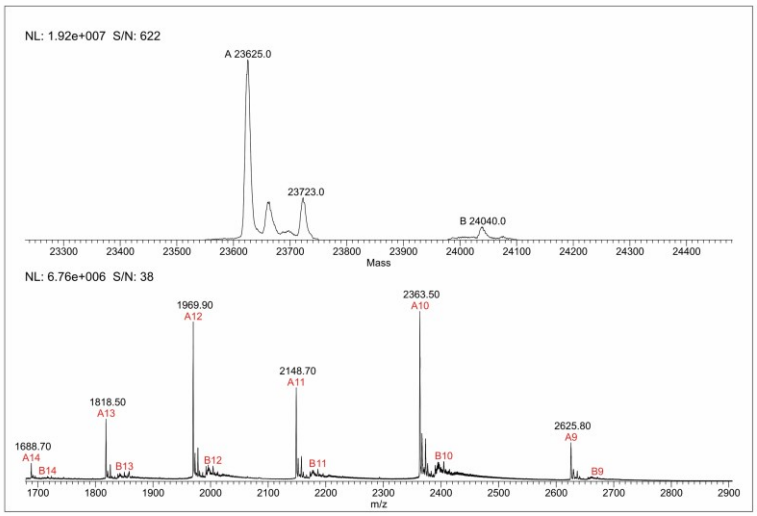
To verify the identity of AGP2 expressed from the pOPINF in both BL21 cells and SHuffle® cells both were analysed by mass spectrometry and the spectra are shown in Figure 2.26. A main peak of mass 23626 Da was observed for each sample. This is consistent with the calculated mass (23622 Da) of the AGP2 construct assuming N-terminal cleavage of methionine. It was therefore concluded that the construct had been successfully expressed in both systems. The 4 Da mass discrepancy may be the result of the sample containing reduced disulphides and so further mass spectrometry data was needed to verify whether this is the case and the refolding conditions may need to be adjusted.

Although the mass spectrometry data demonstrated that the major peak in each sample was consistent with AGP2 there was also evidence from the SDS PAGE analyses of impurities being present.

The analysis of the AGP2 inclusion body prep and refold in Figure 2.26 showed that there were impurities carried through to the final sample and therefore the inclusion body prep was modified to include further wash steps with high salt buffer to facilitate dissociation of soluble protein loosely associated with the insoluble fraction.

To further clean AGP2 produced in the shuffle cell expression, an ionic exchange purification was carried out and the results are shown in Figure 2.27. The purity of the AGP2 fraction was improved by the protocol, but it failed to remove the impurities at ~45 kDa and ~18 kDa which co-eluted with AGP2 and so this purification step was deemed to be sub-optimal, given that some loss of yield would be likely. As the insoluble AGP2 from the inclusion body preparation could be cleaned more easily by additional wash steps of the insoluble material the refolding protocol was taken forward for isotope labelling.

A



B

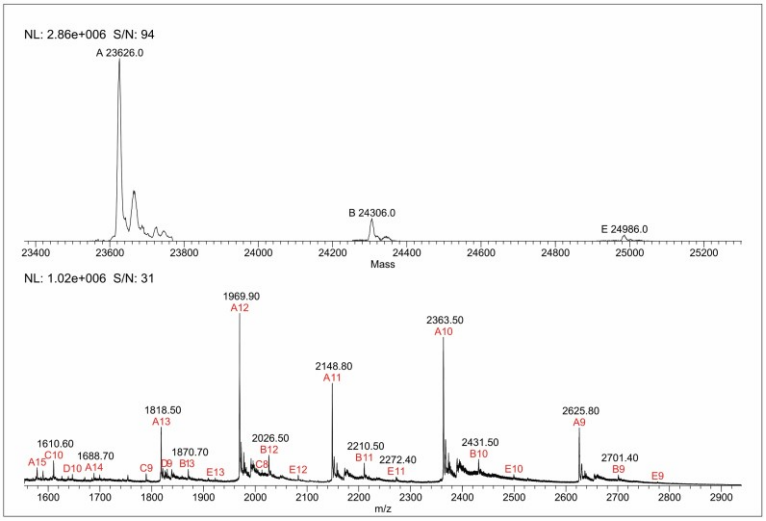
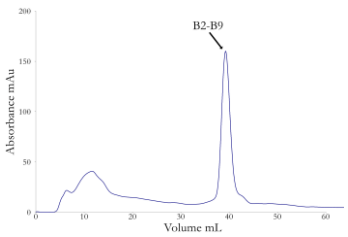


Figure 2.26 (A) Deconvoluted (top) and raw (bottom) electrospray M/S spectra of AGP2 expressed in Shuffle cells with an additional peak at 23723 Da (+98 Da), which is consistent with a sulphate/phosphate adduct **(B)** Deconvoluted (top) and raw (bottom) electrospray M/S spectra of AGP2 expressed in BL21 cells. In both cases the major peak was observed at 23626 Da which is close to the calculated mass of 23622 Da. Further, a peak at ~+40 Da was observed which could correspond to a K^+ adduct. Other peaks were observed at >24000 Da, and were most likely impurities requiring more stringent purification to remove.

A



B

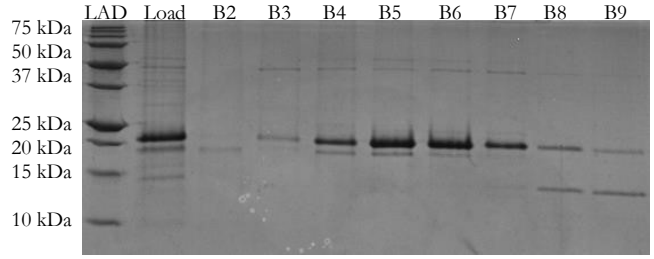


Figure 2.27 (A) UV (280 nm) trace (blue) of NaCl anionic exchange purification with an NaCl elution gradient (green) of AGP2 expressed in shuffle cells. **(B)** Coomassie-stained SDS PAGE analysis of the ionic exchange purification. The protocol was successful in removing some impurities, but significant impurities co-eluted with AGP2.

Expression and Optimisation of Isotopically labelled AGP2 for NMR Studies Using the Refolding Strategy

AGP2 was expressed in BL21 cells grown in rich media and subsequently exchanged into minimal media with $^{15}NH_4Cl$ as its nitrogen source. Inclusion bodies of AGP2 were purified with a more stringent washing protocol compared with the initial rich media test growths in efforts to remove soluble impurities; the results are shown in Figure 2.28.

A dominant band at around 23 kDa was again observable by SDS PAGE. The inclusion bodies were solubilised in 8 M urea solution and rapidly refolded in 1 M buffered arginine as previously. Folded AGP2 was separated from misfolded protein and impurities by gel filtration and analysed by SDS PAGE and the results are shown in Figure 2.29.

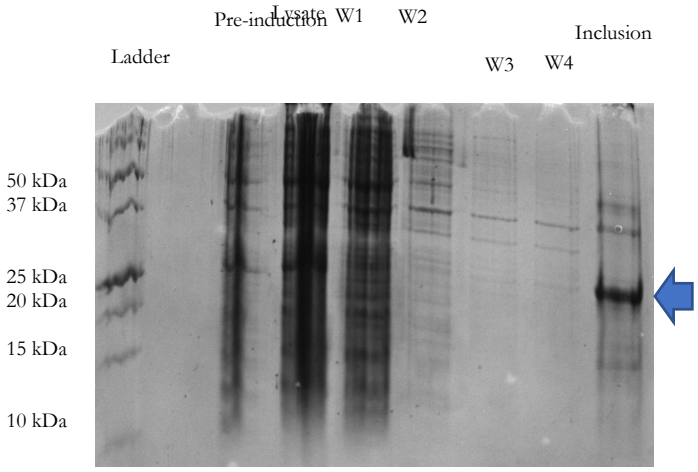
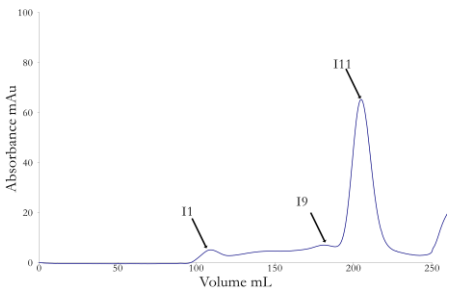


Figure 2.28 Coomassie-stained SDS PAGE of AGP2 inclusion body preparation. The indicated band runs at around 23 kDa as expected and the additional wash steps successfully removed some soluble impurities.

A



B

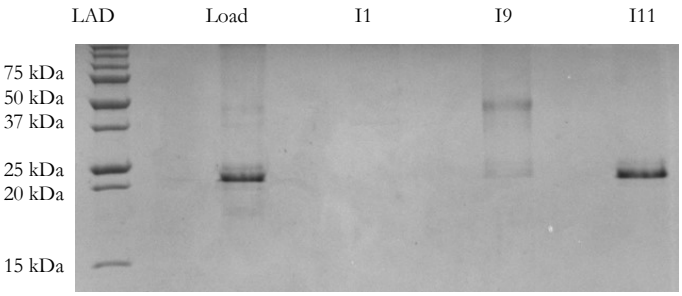


Figure 2.29 (A) UV (280 nm) trace of AGP2 purification from rapid dilution refold. The main peak at around 204 ml post-injection was consistent with folded AGP2 **(B)** Coomassie-stained SDS PAGE analysis of rapid dilution refold gel filtration purification. The main peak from the gel filtration (centred on I11) was significantly purer than the load consistent with the purification step removing both misfolded AGP2 and proteinaceous impurities.

The purified AGP2 from the gel filtration was concentrated and exchanged into 10 mM Na_2HPO_4/NaH_2PO4 buffer at pH 7.4. A ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) transverse relaxation-optimised spectroscopy (TROSY) dataset was collected on a 16.4 T (700 MHz ^{1}H observe frequency) NMR instrument equipped with a 1.7 mm cryogenically cooled TCI probe and the resulting spectrum is shown in Figure 2.30. The dispersion of the H-N peaks is indicative of folding, though further evidence from better-resolved spectra and CD spectroscopy came later. This spectrum and all subsequent NMR spectra were collected by Dr Chris Williams (University of Bristol), whom this author wishes to acknowledge and thank. 195 peaks were expected based on the construct sequence, but only 119 could be confidently identified. This could result from exchange broadening indicative of conformational flexibility in certain regions of the protein or possibly aggregation. Low intensity signals were not included in the peak count, as at this stage it was not possible to rule out that these might correspond to incorrectly folded protein.

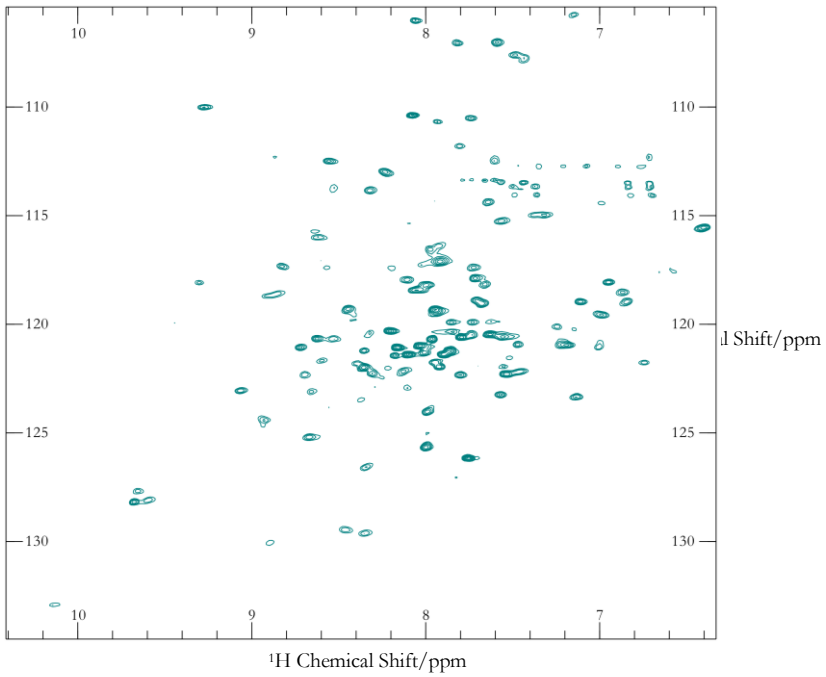


Figure 2.30 ^{1}H-^{15}N TROSY-HSQC solution spectrum of 130 µM apo-AGP2 at pH 7.4 in 10 mM Na_2HPO_4/NaH_2PO4 buffer. Significant peak dispersion was observed consistent with the present of folding protein. 119 significant peaks were counted; this count is substantially lower than the expected 195 peaks expected for this construct and so it was concluded that exchange broadening reducing the intensity of a significant portion of the peaks.

In efforts to further improve the NMR a fresh sample was prepared for which spectra were to be generated with an excess of AMT, which is a known binder with a K_D of 300 nM (Hervé et al, 1998), in an attempt to rigidify the protein, reduce exchange broadening and thereby improve signal resolution. The UV trace of the gel filtration is shown in Figure 2.31.

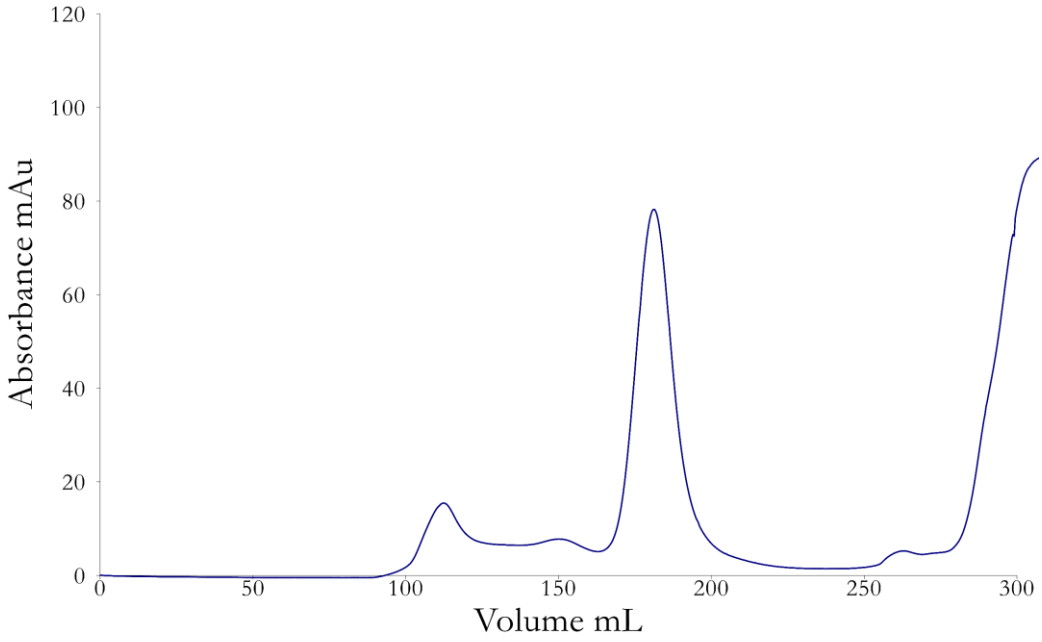


Figure 2.31 UV-trace (280 nm) of AGP2 gel filtration of rapid dilution refold. As previously a significant peak was observed consistent with monomeric AGP2. This run was carried out in a different column to previous preparations and so a peak consistent with monomeric AGP2 was observed at ~195 ml post-injection.

The refold yielded a peak consistent with monomeric AGP2 at around 195 ml post-injection which was concentrated as required and exchanged into 10 mM Na_2HPO_4/NaH_2PO4 buffer for NMR. A TROSY HSQC spectrum was recorded of AGP2 at 300 μM with 3 mM amitriptyline and it is shown in Figure 2.32.

The sample yielded a spectrum with 127 signals that could be counted with confidence indicating that AMT gave some improvement in spectral quality. It was therefore hypothesised that AMT caused AGP2 to become more rigid and conformationally homogeneous resulting in a signal count closer to that which was expected (195). Nonetheless only 65% of the expected peak count could be picked with confidence. It was therefore necessary to carry out further optimisation and gather titration data to validate that AMT yielded CSPs from identical protein samples.

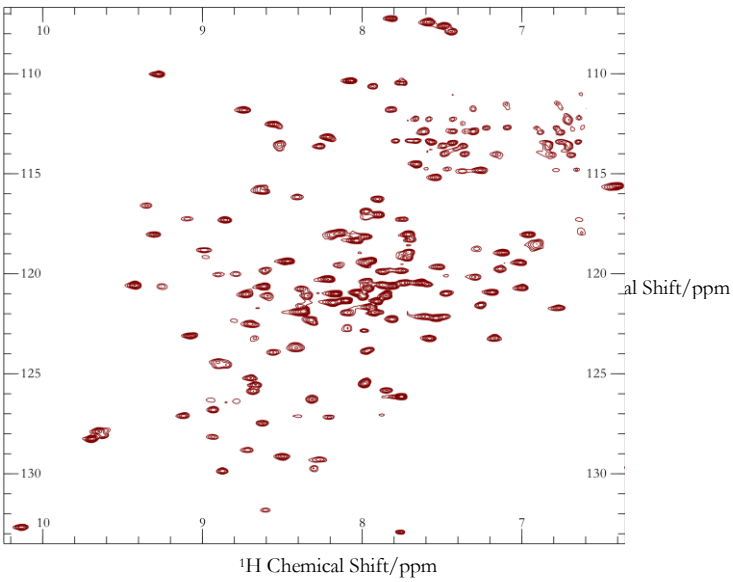


Figure 2.32 ^{1}H-^{15}N TROSY-HSQC solution spectrum of AGP2 at pH 7.4 in 10 mM Na_2HPO_4/NaH_2PO4 buffer with a 3 mM AMT and 300 μM AGP2. 127 signals were picked with confidence indicating that AGP2 was more conformationally homogenous in the presence of AMT than without it.

Protein-observed Ligand Binding NMR Experiments at pH 6.0

Fresh AGP2 was purified from a buffered arginine rapid dilution refolding protocol as previously and the UV trace of the gel filtration is shown in Figure 2.33. In efforts to improve the spectra the NMR buffer pH was changed from 7.4 to 6.0, to reduce signal broadening due to H^+ - NH exchange. AMT was included in the NMR samples at ligand:protein molar ratios of 0:1, 1:1, 2:1 and 10:1 in order to confidently identify and track peak shifts. The TROSY HSQC spectrum of AGP2 with a 10-fold excess of amitriptyline is shown in Figure 2.34. Although less protein was used per sample than in previous spectra, resulting in fewer signals above the noise that could be confidently picked as AGP2 peaks, the relative spectral quality was improved by the addition of AMT with an additional 10% more signals being resolved than the spectrum of the apo protein at a constant concentration. Several signals appeared sharper and exhibited altered chemical shifts. Overlays of the AMT titration from apo to 1:1, 1:2 and 1:10 ratios of AGP2:AMT are shown in Figure 2.35. It was hoped that when the backbone chemical shifts have been fully assigned for the protein it would be possible to map the binding site onto the crystal structure based on minimal shift analysis and potentially identify regions of induced conformational change.

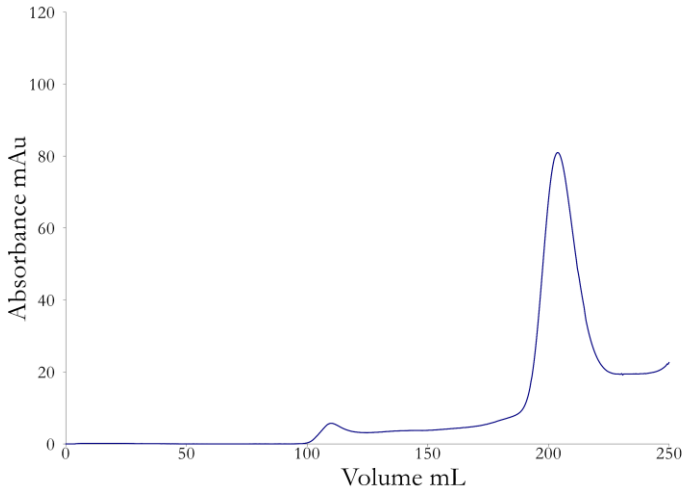


Figure 2.33 UV-trace (280 nm) of AGP2 purification gel filtered directly into 10 mM Na_2HPO_4/NaH_2PO4 buffer at pH 6.0.

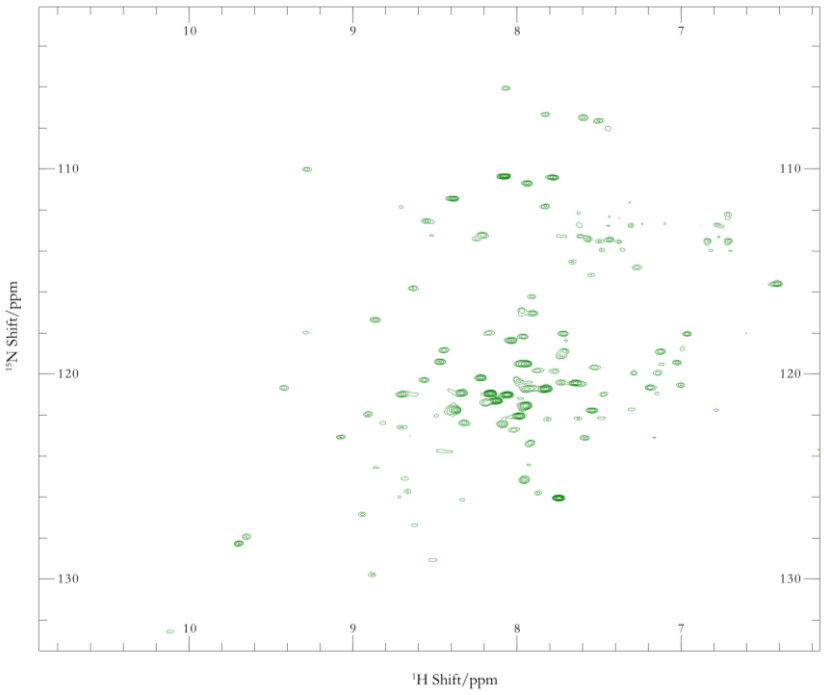


Figure 2.34 ^{1}H-^{15}N TROSY HSQC spectrum of AGP2 (~200 µM) with a 10-fold molar excess of AMT collected at 700 MHz. At a constant protein concentration these conditions resulted in 99 significant peaks compared to 90 peaks for AGP2 in the absence of ligand.

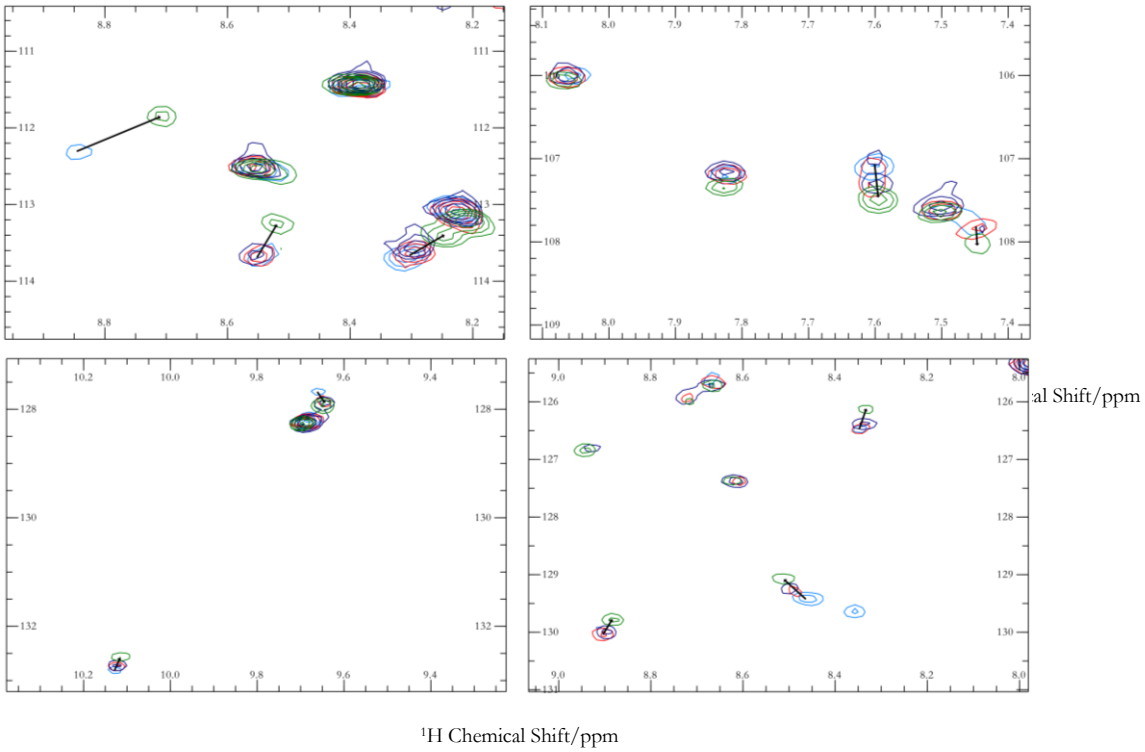
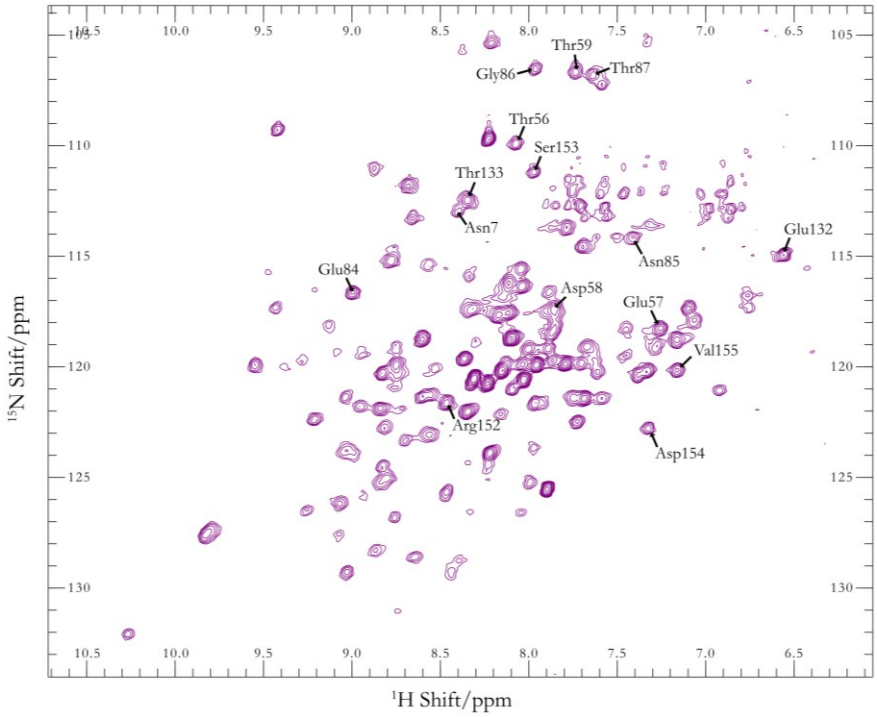


Figure 2.35 ^{1}H-^{15}N TROSY HSQC spectra panels of AGP2 with a ten-fold excess of amitriptyline (green), a two-fold excess of amitriptyline (navy blue), a one-to-one ratio of amitriptyline (red) or without any ligand (cyan) collected at 700 MHz. Several changes in chemical shift between these spectra caused by the addition of the ligand AMT are visible. The reasonable spectral dispersion and combined with CSPs indicated that correctly folded protein with binding activity had been generated.

To enable backbone chemical shift assignment a double labelled sample of AGP2 was expressed in BL21 grown in minimal media with $^{15}NH_4Cl$ and ^{13}C-labelled glucose as sole nitrogen and carbon sources respectively. Double labelled AGP2 was purified by the same protocol as previously. To reduce spectral complexity the N-terminal His_6 tag was cleaved with 3C protease (Generon) and untagged AGP2 was purified by a negative IMAC. The NMR sample was prepared at 0.5 mM AGP2 with a 5-fold excess of amitriptyline in phosphate buffer at pH 6.0. HNCA, HNCACB, HNCOCA and HNCOCACB data sets were collected and examples of the results are shown in Figure 2.36(A). HNCA experiments couple the amide proton with the alpha carbons of the same residue and that of the residue one position earlier in the sequence. HNCOCA experiments couple the amide protons to the alpha carbons of only the preceding residue. A combination of HNCA with HNCOCA experiments enable the sequential chemical shift assignment of the backbone Cα carbons corresponding to these signals. HNCACB and HNCOCACB experiments provide the same data as HNCA and HNCOCA experiments respectively, and in addition generate negative signals corresponding to beta carbons that may also be assigned by the same method. A representative set of assignment strips is shown in Figure 2.36(B).

As the His_6 tag had been cleaved only 179 signals were expected in the spectrum shown in Figure 2.36(A) and 134 were picked with confidence in the HSQC. With 75% of the expected peaks picked, this spectrum was the best one produced up to that time.

A



B

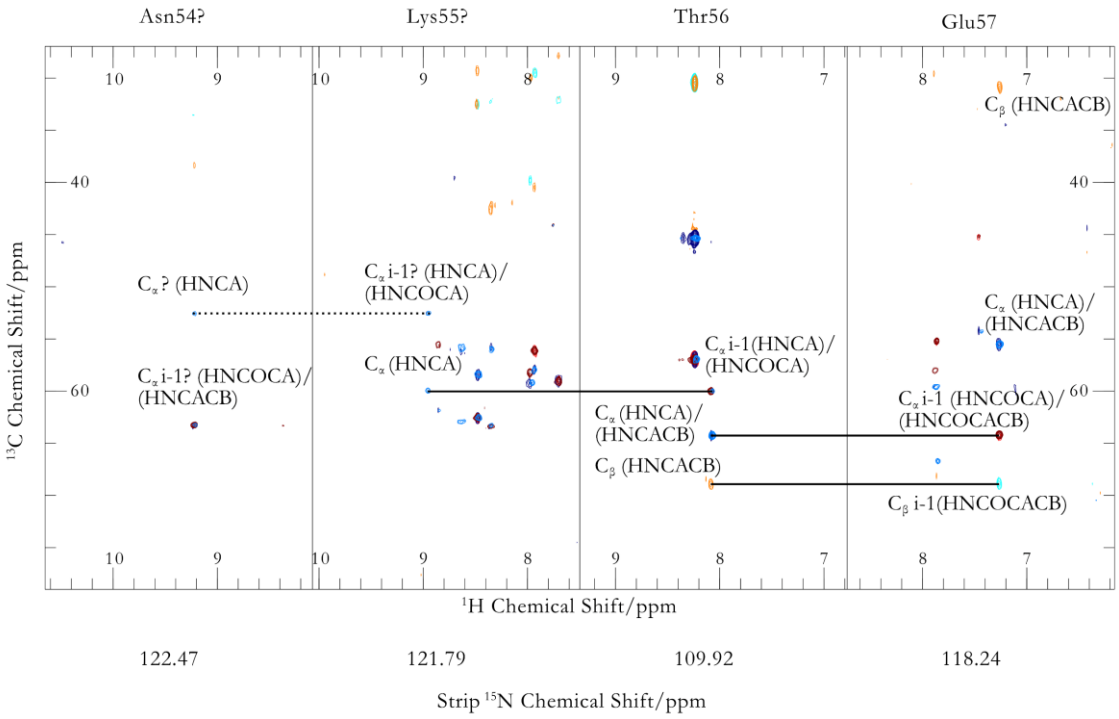


Figure 2.36 (A) ^{15}N-^{1}H TROSY HSQC spectrum of 0.5 mM AGP2 with a five-fold excess of amitriptyline; 134 peaks could be picked confidently. **(B)** Four assignment panels showing HNCA(sky blue), HNCOCA(maroon), HNCACB(navy, orange negative), HNCOCACB (red, cyan negative). Two of the four residues could be assigned with confidence, but declining C_α/C_β signal intensity only allowed Lys55 and Asn54 to be assigned tentatively.

These initial datasets allowed approximately 10% of the residues to be assigned with confidence. Typically, these were polar residues in loop regions of the protein in the crystal structure [PDB: 3APV]. The signal intensity, particularly of β-carbons, was much weaker in spectral regions corresponding to the sequence that forms secondary structure in the crystal structure. Figure 2.36 (B) shows an example of declining signal both of β-carbon signals and some α-signal, in this case in the HNCOCA. The lack of β-carbon signals combined with the presence of many overlapping α-carbon peaks with similar shifts made further confident assignments impossible. Therefore, it became necessary to carry out further optimisation of the NMR conditions to deliver a data set of sufficient quality to allow complete assignment. As the samples at pH 6.0 had not yielded spectra of sufficient quality further optimisation was performed at pH 7.4.

NMR Optimisation

Although AGP2 has only been crystallised with the weaker affinity ligands amitriptyline, disopyramide, and chlorpromazine (Nishi *et al*, 2011), that have K_D's of 0.3 μM, 0.4 μM and 1 μM respectively (Hervé *et al*, 1998), AGP2 is also known to bind the staurosporine class of ligands more tightly. Staurosporine itself is a natural product from *Streptomyces* that was discovered over 4 decades ago (Furusaki, Hashiba and Matsumoto, 1978). Staurosporine and its derivative UCN-01 have K_D's to AGP2 of 88 nM and 3.5 nM respectively (Katsuki *et al*, 2004). It was hypothesised that a ligand that bound AGP2 more strongly than amitriptyline would further reduce the exchange broadening of signals by rigidifying the protein and yield better spectra. ^{15}N-labelled AGP2 was expressed and purified by gel filtration as previously and the resultant UV trace is shown in Figure 2.37. Monomeric AGP2 eluted at around 190 ml post-injection, though in contrast to previous purifications significant UV signal was present between the void and monomeric AGP2 elutions suggesting that some improperly folded protein had been produced in this preparation. Nonetheless, the peak corresponding to monomeric AGP2 was exchanged into 10 mM Na_2HPO_4/NaH_2PO_4 buffer at pH 7.4 and HSQC data sets in the presence and absence of an excess of staurosporine and the spectra are shown in Figure 2.38. 123 peaks, out of the expected 195, were counted. An AGP2 sample from the same prep run without staurosporine resulted in an HSQC spectrum with only 65 significant peaks. Whilst this result demonstrated that staurosporine had a significant effect on the structure of AGP2 in solution each peak count was likely compromised by the poorer quality of the protein preparation indicated by Figure 2.37. Therefore, it was hypothesised that a fresh sample from a superior prep observed with a staurosporine compound would yield good data.

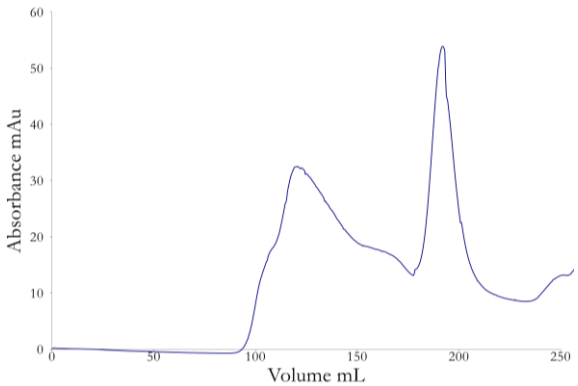


Figure 2.37 UV (280 nm) trace of gel filtration of AGP2 refolded in arginine, significant signal was detected between the void volume at around 120 ml and the AGP2 peak at around 190 ml, this may indicate partially folded aggregates that may contaminate the sample.

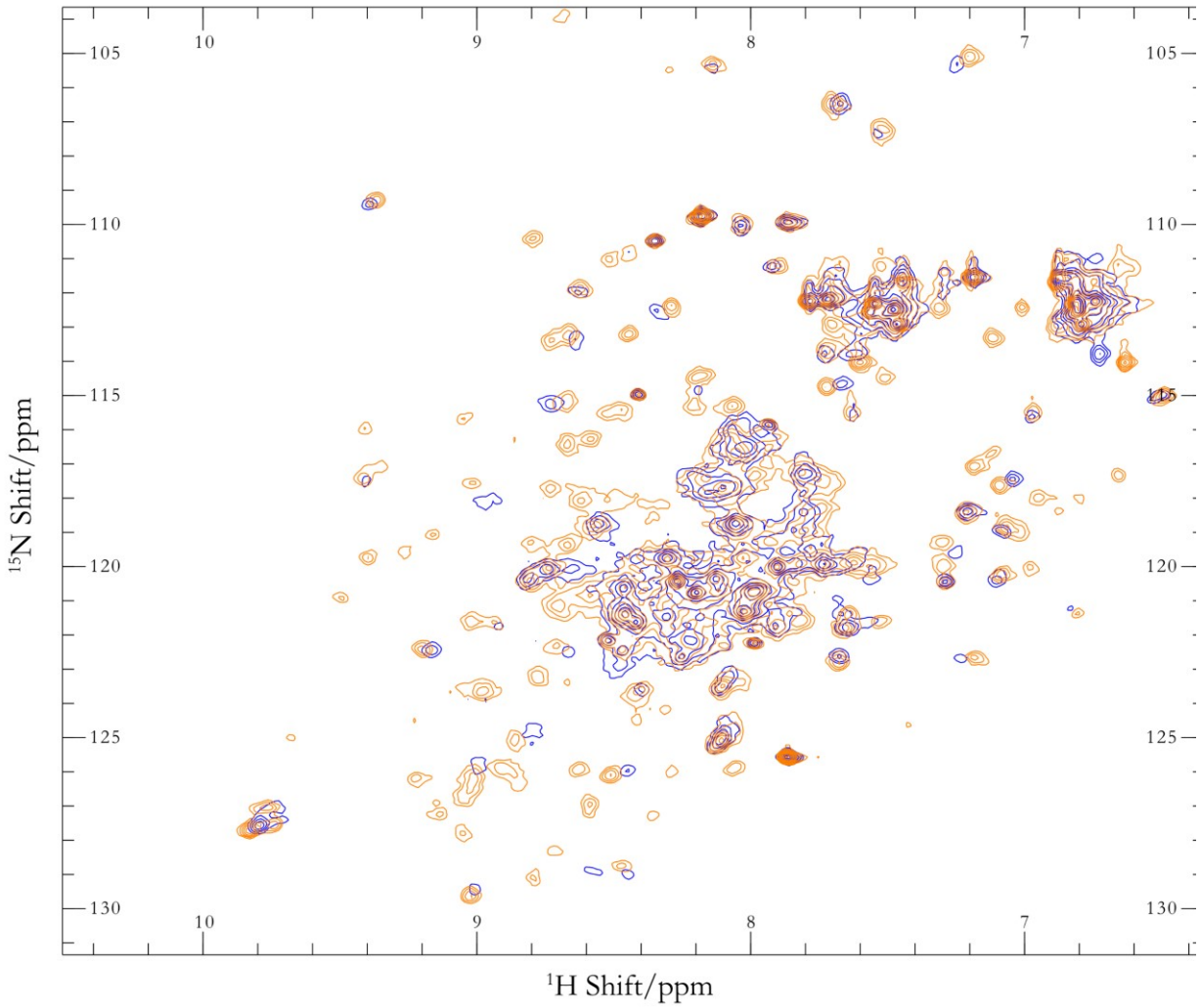


Figure 2.38 Overlay of ^{1}H-^{15}N HSQC spectra of AGP2 (180 μM) with 250 μM Staurosporine (orange) and with an equivalent DMSO concentration without any ligand (blue). 123 significant peaks are present suggesting that staurosporine stabilised a more structurally homogeneous conformation of AGP2 than was present in the spectrum of apo protein that only had 65 peaks. Several peak shifts on the addition of staurosporine are visible, but most peaks in the staurosporine-bound spectrum are not present in the apo spectrum suggesting that the corresponding parts of the protein require a tightly bound ligand to be resolved.

As described above the compound UCN-01 is a hydroxylated analogue of staurosporine and has a higher affinity for AGP2. Given the superior spectral quality of AGP2 in the presence of staurosporine relative to the apo sample of protein from the same preparation, it was hypothesised that UCN-01 as a stronger binder of AGP2 would produce further spectral improvements by forming more rigidifying protein-ligand interactions and so further reduce peak broadening that most likely stems from conformational exchange. Fresh ^{15}N-labelled AGP2 was expressed and purified as previously and the N-terminal His_6 tag was cleaved by in-house 3C protease. AGP2 was purified by negative IMAC and exchanged into 10 mM Na_2HPO_4/NaH_2PO4 buffer at pH 7.4. A TROSY HSQC spectrum was acquired of the AGP2 sample with

a 2-fold excess of UCN-01 and the resulting spectrum is shown in Figure 2.39. The spectrum was well-dispersed and contained 165 peaks out of the expected 179.

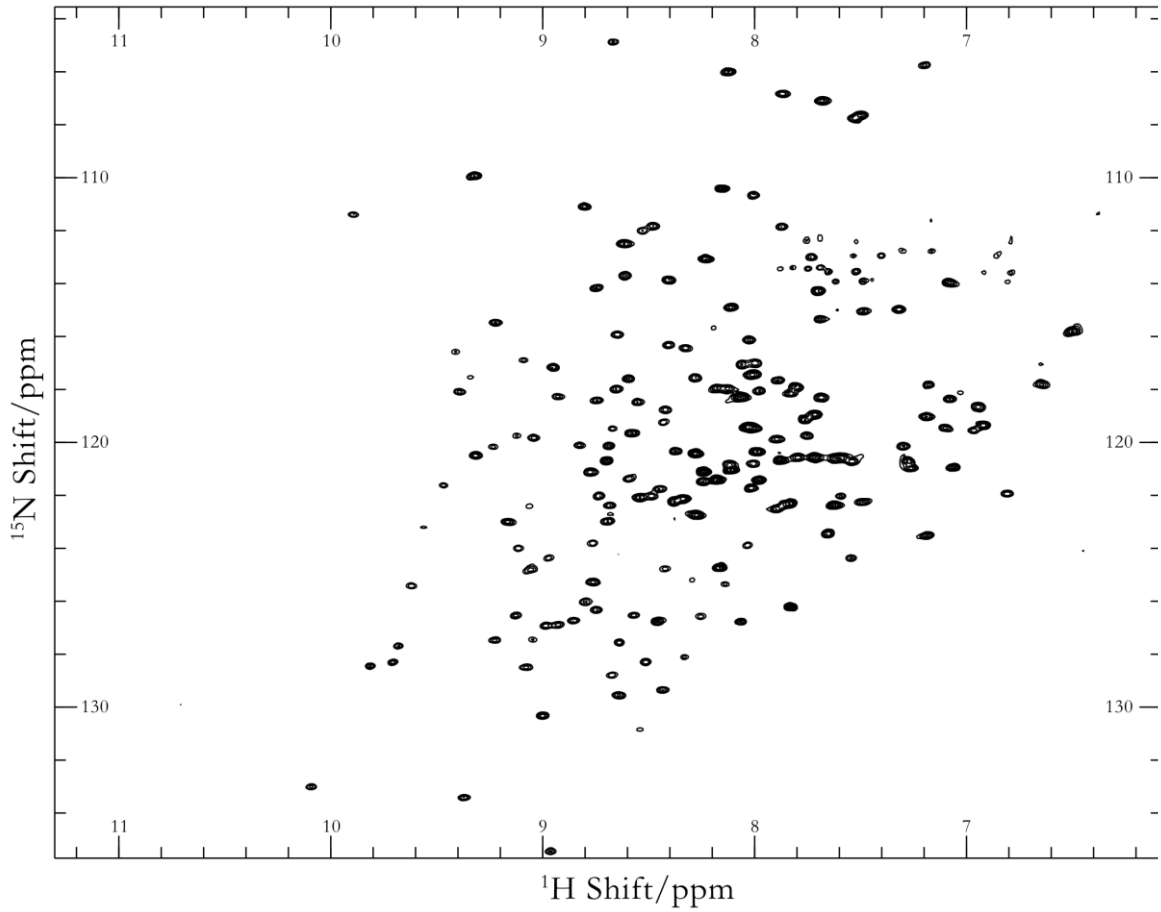


Figure 2.39 ^{1}H-^{15}N TROSY-HSQC of AGP2 (0.5 mM) with 2-fold excess of UCN-01. 165 significant peaks were present out of a total 179 expected peaks collected at 700 MHz. Signals were well-dispersed strongly indicating that the protein was folded and homogeneous.

Purification of AGP2-trunc for NMR Studies

To increase the likelihood of producing a complete set of assignments a modified pOPINF construct was produced by SDM removing ten C-terminal residues (AGP2-trunc). This sequence forms an alpha helix away from the binding pocket (Nishi *et al*, 2011), and so, although we had not confirmed this by a biochemical assay at this stage, it was reasonable to believe that its removal would not disrupt the binding activity of AGP2. The integrity of the construct was verified by sequencing and pure refolded protein was prepared from inclusion bodies by the same method as the full construct (Figure 2.40).

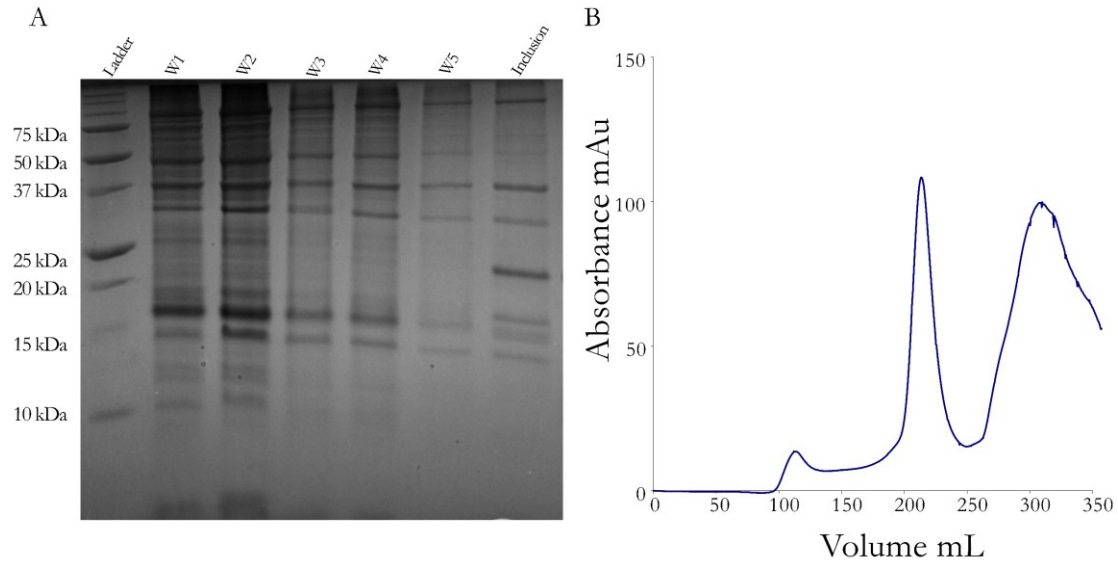


Figure 2.40 (A) SDS PAGE gel of inclusion body preparation of AGP2-trunc. A band between 20 kDa and 25 kDa is consistent with the new construct. **(B)** SEC trace (UV Absorbance 280 nm) of refolded AGP2-trunc inclusion bodies yields a peak at ~215 mL post-injection, consistent with folded AGP-trunc.

The sample corresponding to the main SEC peak was cleaved by 3C protease using the conditions for full-length AGP and analysed by SDS PAGE as shown in Figure 2.41. The hexa-his tag and 3C-protease were removed by negative IMAC to yield the final sample.

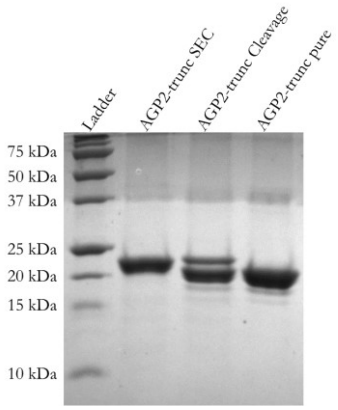


Figure 2.41 SDS PAGE gel of AGP2-trunc cleavage (3C) reaction. The protein was cleaved and yielded a band (lane three) consistent with pure cleaved protein after negative purification.

To verify the identity and oxidation state of the protein from the AGP-trunc refold ESI-MS was carried out on samples prepared for NMR as shown in Figure 2.42. The analysis yielded signals consistent with the cleaved construct. Further, acetylation ESI-MS experiments were carried to demonstrate that the AGP2 thiol groups formed disulphide bonds. These experiments were carried out by comparing MS spectra of AGP2 exposed to the thiol capping reagent iodoacetaminde (IAA) in reducing and non-reducing conditions. In non-reducing conditions IAA would not react with thiol groups involved in disulphide bonding, but if those cysteine residues had not formed

disulphide bonds a reaction would occur. A positive control in reducing conditions demonstrated that when the protein disulphide bonds were broken by DTT before the addition of IAA the observed mass of the

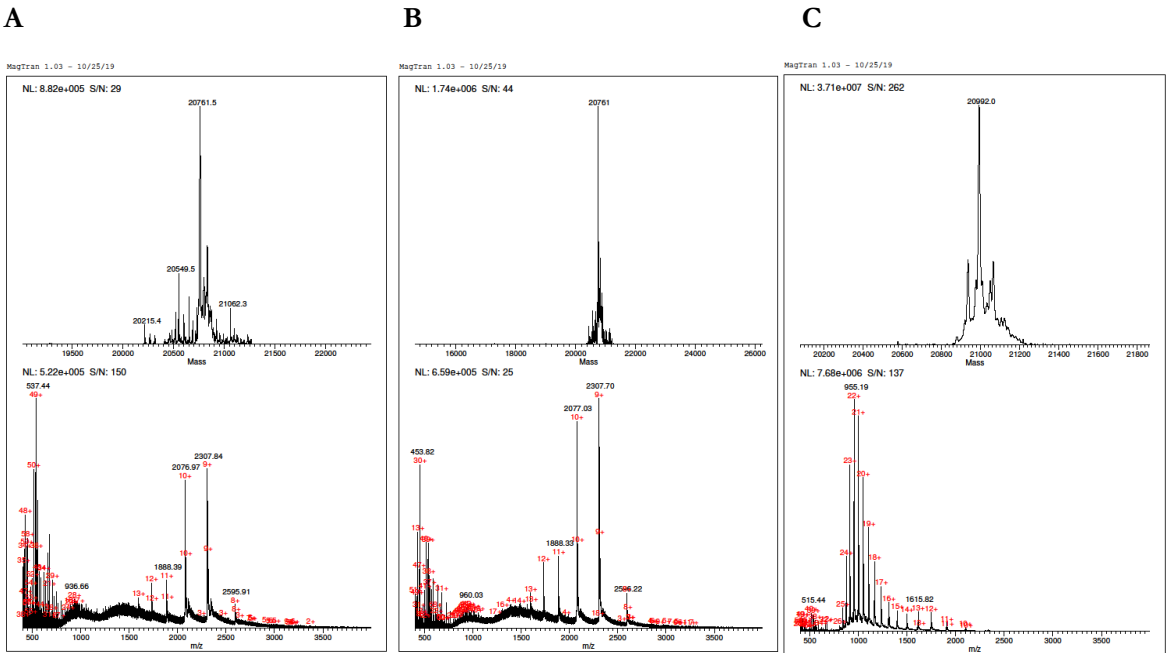


Figure 2.42 Deconvoluted (top) and raw (bottom) ESI-MS spectra of purified ^{15}N-labelled AGP2-trunc **(A)** Untreated AGP2-trunc yielded a peak at 20762 compared to the expected mass of 20769 assuming 100% ^{15}N labelling. Given that isotope labelling is imperfect this discrepancy is acceptable **(B)** ESI-MS of AGP2-trunc treated with IAA without reducing power resulted in an unchanged spectrum with a peak at 20762, indicating that all four thiol groups were oxidised and in disulphide bonds. **(C)** Spectrum of AGP2-trunc reduced with DTT acted as a positive control with a mass increase of 231 to a total of 20922 Da corresponding to the addition of four 58 Da acetamide groups. The raw spectrum also showed a change to a more charged mass envelope due to the presence of additional ionisable amide groups from the assay and likely also from greater exposure of ionisable side chains after disulphide bonds are broken.

protein would increase due to IAA capping (Figure 2.42).

To characterise the stability of the truncated construct, a thermal denaturation was performed and monitored by circular dichroism (CD) spectroscopy. This experiment was performed with both apo AGP-trunc2 and AGP2-trunc in the presence of UCN-01 and the results are shown in Figure 2.43. The addition of UCN-01 resulted in a significant stabilisation of the protein with the melting temperature shifting by 13.7 °C from 61.9 °C to 75.6 °C. Thermal stabilisation is often used as an indicator of binding in small molecule screening as enthalpic interactions between protein and ligand often result in a raised melting temperature of the complex relative to apo protein (Pantoliano *et al*, 2001). The significant conformational stabilisation in these data is consistent with the high affinity binding of UCN-01 to AGP2 reported in the literature (Katsuki *et al*, 2004), and provides strong evidence that the truncated AGP2 construct retains strong ligand binding activity. The thermal stabilisation also supports rigidification as the mechanism of spectral improvement seen upon the addition of UCN-01 to AGP2 NMR samples.

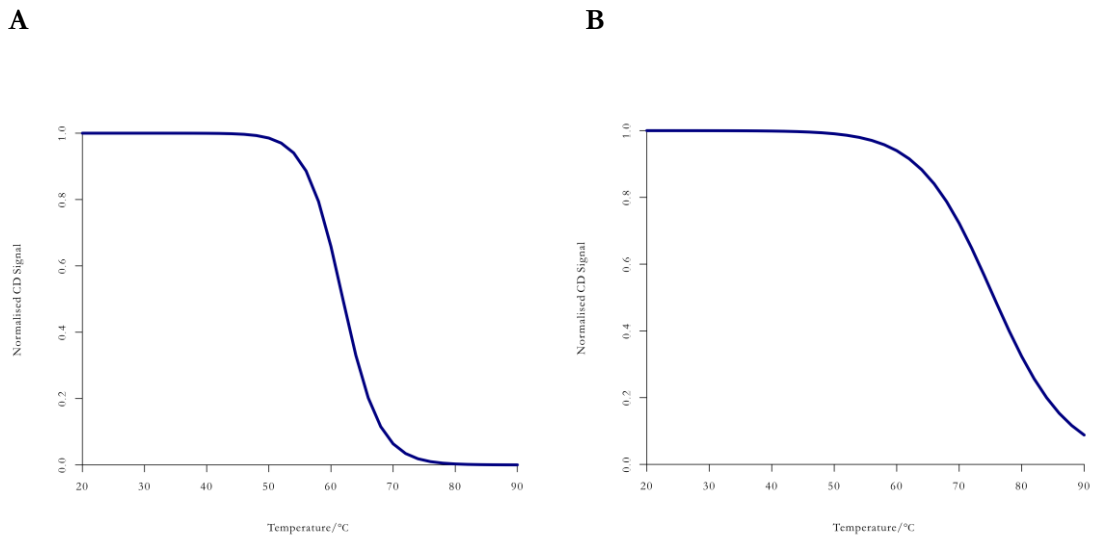


Figure 2.43 (A) CD melt curve of apo AGP2 **(B)** CD melt curve of AGP2 in the presence of the high-affinity binder UCN-01. There was a substantial increase in the melting temperature of the protein by 13.7°C from 61.9°C to 75.6°C indicating that the ligand had a strong stabilising effect on the secondary structure of AGP2. Wavelengths between 185 and 250 nM were observed, and a global fitting algorithm (Global 3 v1.6.0.0 (Applied Photophysics)) was used to determine the overall fit. Each curve is based on one run. This author wishes to thank Dr Richard Taylor and Dr Oliver Durant (UCB, Slough) for operating the CD spectrometer (Chirascan (Applied Photophysics)).

NMR optimisation of the AGP2-trunc Construct

A fresh ^{15}N-labelled sample of AGP2-trunc was prepared by the same method as for unlabelled material and a ^{1}H-^{15}N HSQC spectrum of isotopically labelled cleaved AGP2-trunc was recorded (Figure 2.44(A)). The sample was prepared at a 1 mM protein concentration with a 2-fold excess of UCN-01 and gave excellent peak dispersion and homogeneity. A sample of ^{15}N-labelled full-length AGP2 was prepared as a comparison at the same concentration (Figure 2.44(B)). Amongst the peaks in common between the spectra no significant changes were present in the AGP2-trunc spectrum and so it was decided to proceed with the truncated construct to simplify assignments.

To improve the spectra further, optimisations were carried out by varying pH. The amitriptyline-bound spectra referred to previously were acquired at pH 6.0, whilst the initial UCN-01 spectra were acquired at pH 7.4. Lower pH samples have the advantage of reducing the exchange of backbone amide hydrogen yielding sharper peaks, but a pH closer to 7.4 may be more likely to yield a more natively structured protein as this is the pH of serum. Therefore, pHs of 6.5 and 7.0 were screened by preparing fresh samples of ^{15}N-labelled AGP2-trunc. These pHs were chosen partly to remain above the theoretical pI of 5.46.

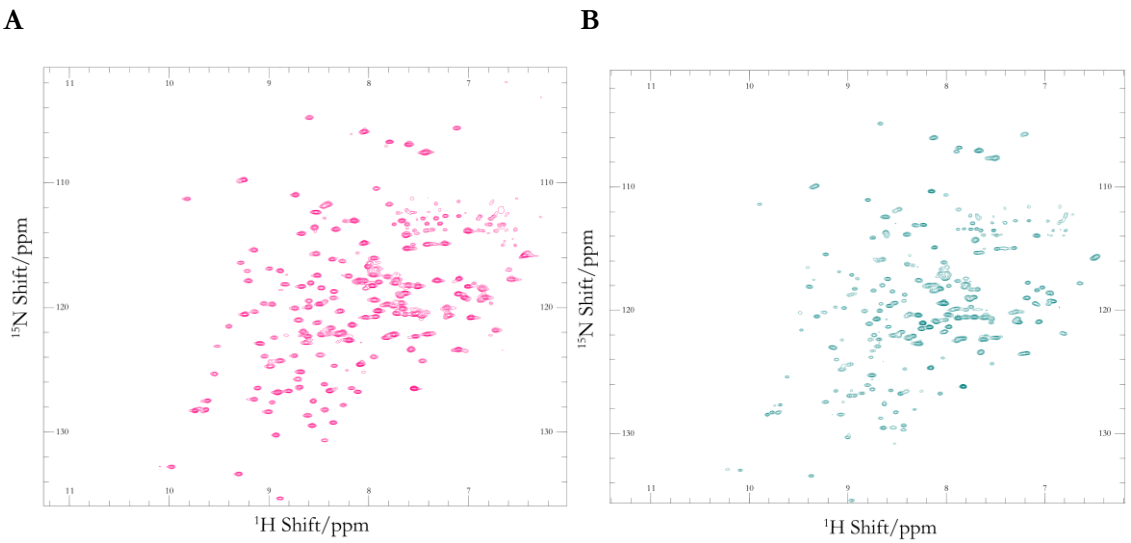


Fig 2.44 (A) ^{1}H-^{15}N TROSY-HSQC spectrum of 1 mM AGP2-trunc with a 2-fold excess of UCN-01 at pH 7.4 collected at 700 MHz. The peaks are well dispersed and 164 out of the expected 169 peaks were counted. **(B)** TROSY-HSQC collected at 700 MHz of comparative sample of full length AGP2 reveals few significant differences.

Figure 2.45 shows test HSQC spectra of AGP2-trunc at pH 7.0 (A) and pH 6.5 (B). This test demonstrated that superior spectra may be generated at pH 6.5 as 169 peaks could be counted out of an expected 169 compared to 164 out of 169 at pH 7.0. Therefore, triple resonance spectra were recorded using the optimised conditions (10 mM NaH_2PO_4/Na_2HPO_4, 100 mM NaCl, pH 6.5)

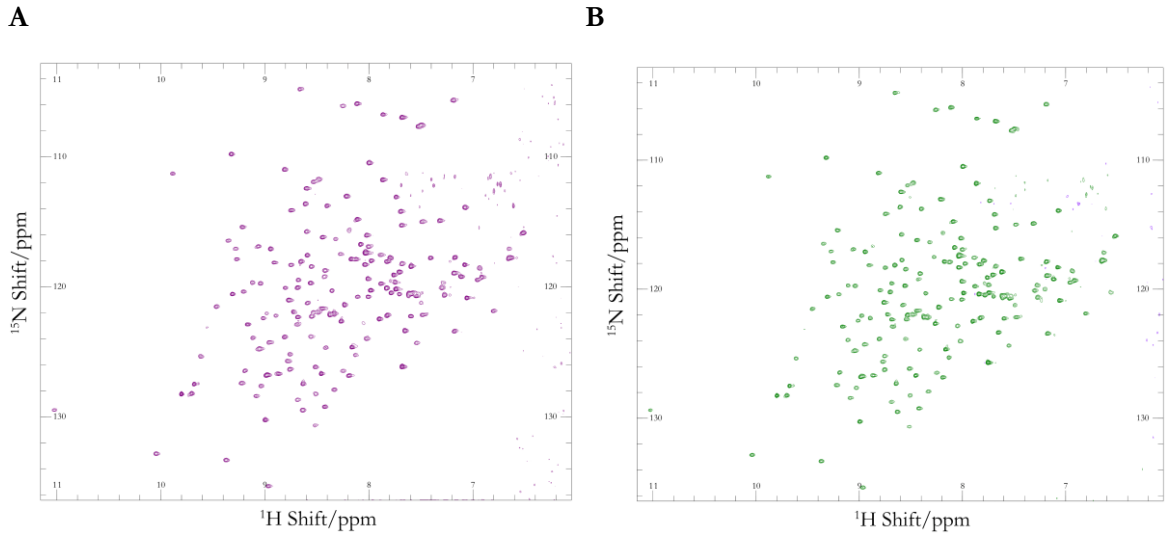


Figure 2.45 (A) ^{1}H-^{15}N TROSY-HSQC spectrum of 300 µM AGP2-trunc with a 3-fold excess of UCN-01 at pH 7.0 collected at 700 MHz. 164 out of an expected 169 peaks could be counted. **(B)** ^{15}N-^{1}H TROSY-HSQC spectrum of 300 µM AGP2-trunc with a 3-fold excess of UCN-01 at pH 6.5 collected at 700 MHz. 169 out of an expected 169 peaks could be counted.

Full Backbone Assignments of AGP2-trunc

To obtain full backbone assignments of the new protein construct additional HNCA, HNCACB, HNCOCA and CBCACONH spectra were recorded at 700 MHz with UCN-01 at pH 6.5. Figure 2.46 shows the same putatively consecutive strips of signals that were partially assigned in previous amitriptyline-bound spectra. The new conditions allowed additional C_α signals to be identified and assigned with confidence whilst the CBCACONH gave greater assurance to the assignments of consecutive residues. As this sample is at the same concentration and similar pH to that used previously with AMT bound, the increased ease of assignment is most likely due to the higher affinity of the ligand and decrease in spectral complexity from the truncated construct. Several regions of AGP2 contain repetitive hydrophobic sequences (most notably: Phe48-Phe49-Tyr50-Phe51 and Leu101-Leu102-Phe103-Leu104) of residues exhibiting very C_α/C_β similar shifts. Consecutive pairs of identical residues with similar shifts were also present at many loci, further complicating assignment (e.g. Glu35-Glu36, Ser76-Ser77, Gly93-Gly94). In other stretches of sequence, the Cα and Cβ signals were challenging to assign due to a lack of signal. In addition to the Asn54-Glu57 region, Ile13-Leu18 and Lys120-Tyr127 could not be assigned immediately due to poor signal-to-noise in those stretches. The former of these regions was further complicated by Pro12. The issues with signal strength generally occurred around the transitions between secondary structure elements and loops. Ile13, Asn54 and Lys120 all occur around these transitions. The reason for the poor signal to noise ratio could be broadening from additional motion of the less rigid loops outside of secondary structure elements. These complexities could not be resolved with existing data and therefore further spectra were collected at 950 MHz (this author wishes to thank Christina Redfield (Oxford University, Biochemistry) who operated the instrument). In particular a CBCANH spectrum at the higher field was able to detect all the remaining Cβ signals and allowed a complete set of unambiguous backbone assignments to be obtained when used in conjunction with previously collected 700 MHz data. The CBCANH experiment yields positive peaks for C_α spins with the exception of glycine, which are negative along with all the Cβ peaks. The negative glycine peaks were particularly helpful to finally assign the Gly93/Gly94 loop, and other elements between two stretches of secondary structure. Figure 2.47 shows how the improved sensitivity and dispersion of the higher field instrument revealed the Cβ peaks, that had previously not been visible above the noise, and so allowed the determination of a complete set of backbone assignments. In the 950 MHz data the newly observed C_β peaks were weak relative to other 950 MHz C_β peaks, but nonetheless strong enough to be used in unambiguous assignments. Assignments were further facilitated by a 3D ^{1}H-^{15}N HSQC-NOESY data set. The geometric disposition of residues in β-sheets in particular results in strong Hα NOE signals detectable from ^{1}H-^{15}N HSQC cross-peaks of succeeding residues. The extensive β-barrel comprising much of AGP2's structure facilitated the identification of many residues based on these NOE peaks. 91 out of 172 Hα atoms were assigned particularly along the β-barrel, in regions such as Asp58-Gln66. The strong NOE signals between amide protons and i-1 Hα protons due to β-sheet geometry made these spins easy to identify and use as validation for backbone assignments where triple resonance peaks may exhibit poor signal-to-noise ratios.

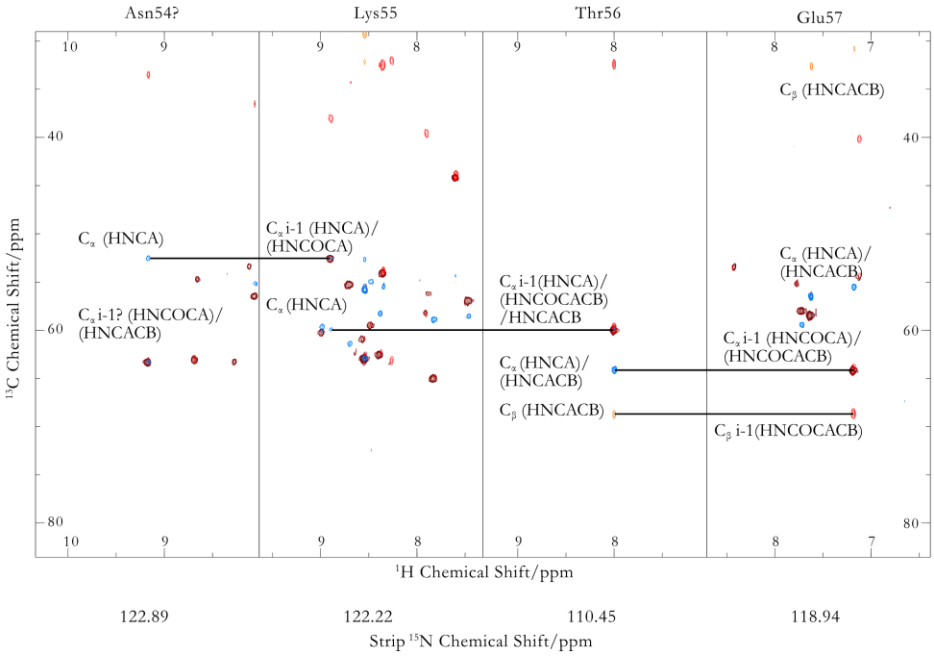


Figure 2.46 HNCA(sky blue), HNCOCA (maroon), HNCACB (navy, orange negative) CBCACOHNH (red) spectra of AGP2-trunc in complex with UCN-01 at 700 MHz. C_α signals were maintained along further stretches of the backbone than in the amitriptyline bound samples. Additionally, more i-1 C_β signals allowing the identities of connecting residues to be established with greater confidence. However, many C_β signals were absent in the HNCACB making assignment difficult in some repetitive hydrophobic regions of the protein.

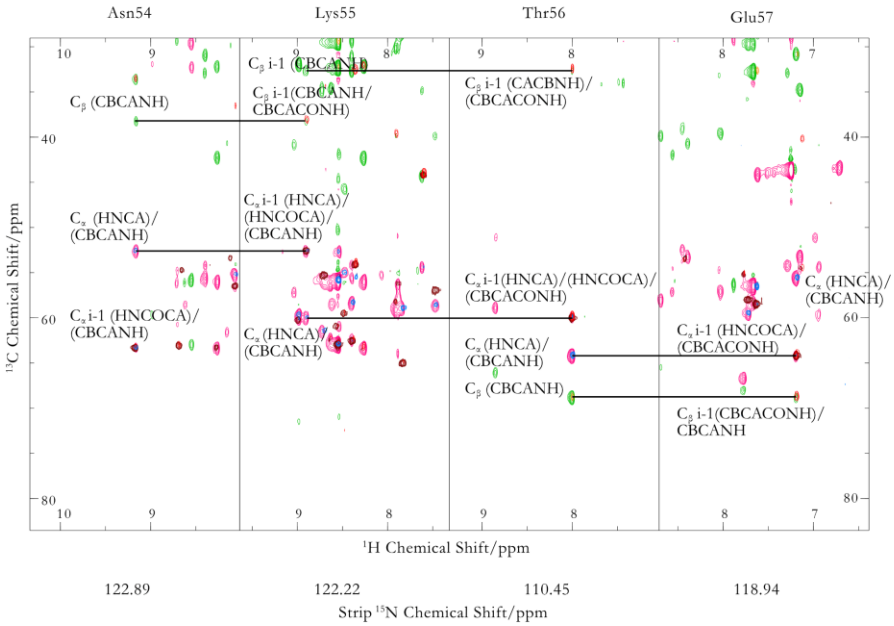


Figure 2.47 HNCA(sky blue), HNCOCA (maroon), HNCACB (navy, orange negative) CBCACONH (red), CBCANH(pink, green negative) spectra of AGP2-trunc in complex with UCN-01. The CBCACNH collected at 950 MHz yielded a drastic improvement in Cβ signal and allowing a complete set of backbone assignments to be determined.

During assignment 38 residues, often in contiguous series, could be assigned to two separate but close chemical shifts. Met111 was one such example and is shown in Figure 2.48. The largest series of doubled peaks were residues Gln1-Leu8, Ser77-Gln82 and Ser89-Arg95. Gln1-Leu8 is in the randomly coiled N-terminal region, Ser77-Gln82 is in β-strand D and Ser89-Arg95 is in the E-F β-hairpin. There is therefore no discernible connection between secondary structure and doubling. This effect was observed at 600 MHz, 700 MHz and 950 MHz and at pH 6.5-7.4 in some, but not all spectra (For example Figure 2.49 shows a spectrum of AGP2 with UCN-01 that does not contain doubling). The cause of this doubling effect could not be identified and so in each case the major peak was assigned.

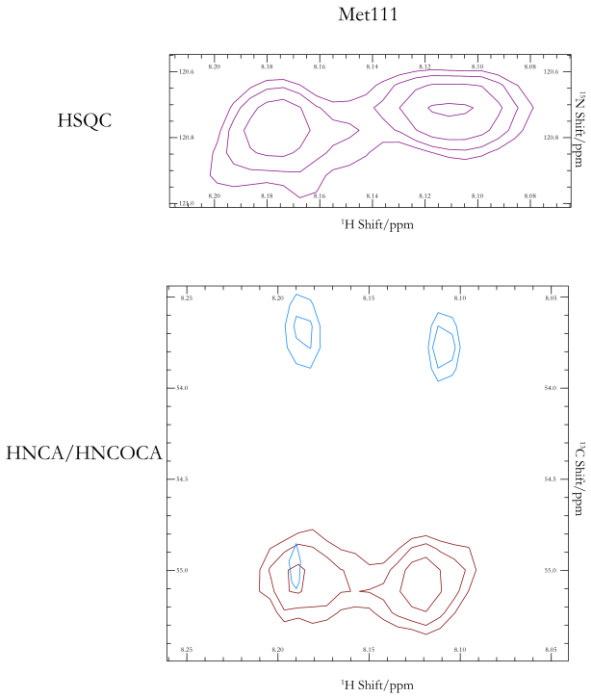


Figure 2.48 2 Sets of peaks in the HSQC (pink), HNCA (sky blue) and HNCOCA (brown) may be assigned to Met111. When this doubling was observed the main peak was assigned.

98% complete backbone assignments of non-proline residues were obtained and are shown in Figure 2.49. A full assignment list is reported in the appendix a.1. The assigned HSQC spectrum is, as discussed in the introduction, a valuable tool in applying NMR to drug discovery. This HSQC spectrum with assignments

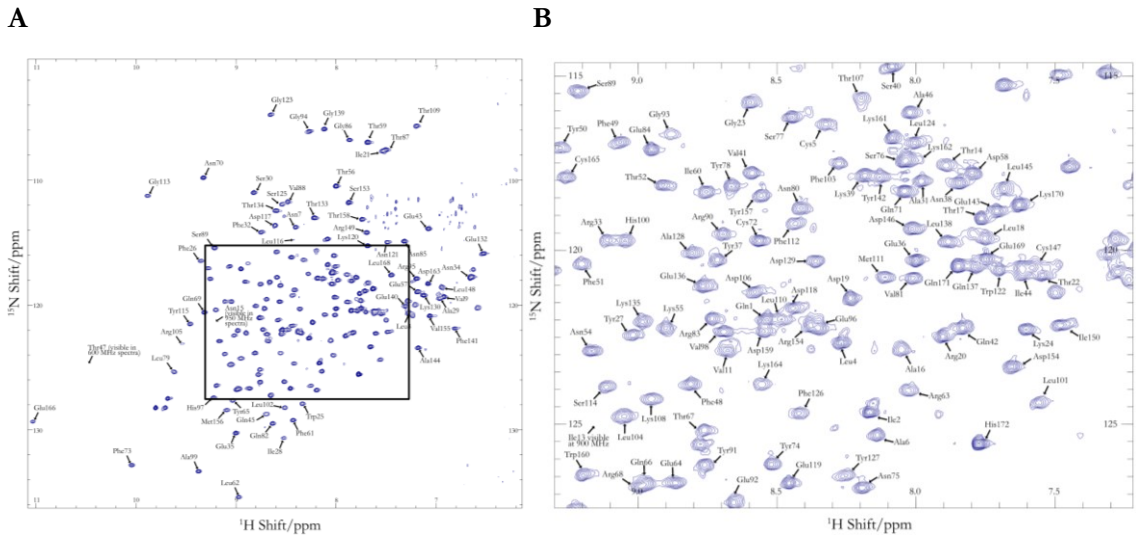


Figure 2.49 ^{1}H-^{15}N TROSY-HSQC spectrum of AGP2 bound to UCN-01 at pH 6.5 with complete backbone assignments. Several peaks are only resolved at lower fields or display multiple peaks at higher field indicating that transitions between slow and medium exchange may occur at these field strengths **(A)** Complete spectrum with peripheral assignments **(B)** expanded central part of the spectrum with assignments.

may now be used to characterise and compare ligand binding to AGP2 without the need for less consistent crystallography approaches that may fail for certain fragments or small molecules.

UCN-01 Titration

To further understand the effect of UCN-01 binding to AGP2, a titration experiment was carried out and the results are shown in Figure 2.50. TROSY-HSQC spectra of a sample of AGP2-trunc were collected at UCN-01 to protein ratios of 0:1, 1:1, 2:1 and 4:1. These data showed that UCN-01 induced a homogeneous bound conformation with very little change in the HSQC data above a stoichiometry of 1:1, as expected for a ligand with a K_D on the order of 10^{-8} M. Therefore, an overlay of the spectra with a 2:1 ratio of UCN-01 to protein and no ligand are shown in Figure 2.50. The data also revealed that the change between the ligand bound and apo form of AGP2 must be extensive as only a portion of the peaks were at a similar or identical chemical shift in each spectrum. The changes were sufficiently extensive to prevent straightforward transfer of backbone chemical shift assignments to the apo form from the bound form.

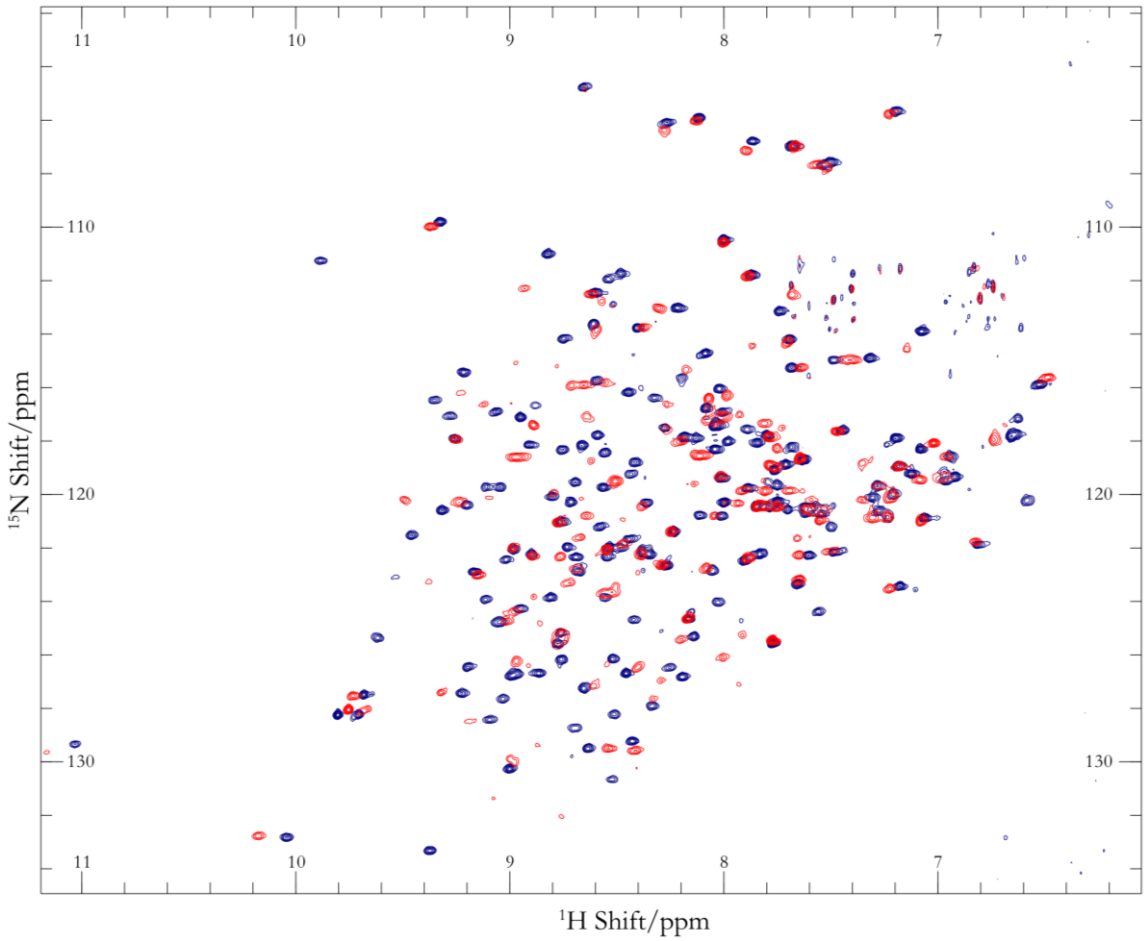


Figure 2.50 ^{1}H-^{15}N TROSY-HSQC AGP2-trunc collected at 700 MHz without ligand (red) and with a two-fold excess of UCN-01. The extensive shifts of backbone peaks and appearance of other from the noise indicate that a radical structural rearrangement occurs upon UCN-01 binding. Spectral changes were minimal upon increasing the UCN-01:AGP2 ratio above 1:1 indicating that the protein saturated with only a minimal excess of ligand as expected for a 10^{-8} M binder. Further, the spectral changes that occurred upon ligand binding were drastic, pointing to a significant conformational rearrangement.

Numerous signals that were resolved in the ligand bound spectrum were missing in the apo-AGP2-trunc spectrum. Resonances in the apo spectrum were also generally broader, indicating that the structure as a whole was less conformationally homogeneous compared to UCN-01-bound protein. These data are consistent with past crystallography results as no truly apo form of AGP has so far been crystallised. AGP either crystallises as protein-ligand complexes with small drug molecules drugs of interest or, if none are present, crystallisation buffer components, such as polyethylene glycol, occupy the binding pocket suggesting that rigidification is critical for crystallisation.

To map the most affected parts of the protein the apo and 2:1 UCN-01:AGP2 HSQC data were compared in a reverse minimal chemical shift approach. This approach was necessitated by the lack of backbone assignments for the lower quality apo spectra where many resonances were either missing, very weak or overlapped due to conformational flexibility. Using the fully assigned ligand bound spectrum the nearest signal in the apo spectrum was assumed to be its pair provided the peaks were within 0.1 ppm ^{1}H and 0.68 ppm ^{15}N (weighting described in materials and methods). The effect of this method is to quantise CSPs into "more significant" or "less significant" categories. It should be noted that this method may yield some errors due to peaks in the apo spectrum by chance being closer to peaks in the UCN-01 bound spectrum that do not correspond to the same residues. Figure 2.51 shows a structure of AGP2 with residues that shift most significantly in the apo spectrum compared to the liganded spectrum highlighted in blue. Residues that are identifiable across the titration are shown in red. Significant shifts between the two forms are not confined to residues that are known to be important for ligand binding (such as Phe112), but also to residues that are distal to the binding site (such as Leu101), which is again indicative of extensive structural rearrangement.

A striking feature of this shift map is the occurrence of strongly shifted residues along adjacent portions of β-sheet around the ligand binding barrel. The H-N pairs that are observed in the HSQC experiment also form the main H-bonding network that holds adjacent antiparallel β-strands together. These data therefore suggest that some deformation of the β-barrel architecture may occur in the apo form of AGP2 such that that the strands with the greatest proportions of shifted residues open to allow ligand entry into the binding site. Where possible, shift distances were plotted against residue number to give a quantitative measure of the difference (Figure 2.51 C).

To validate these hypotheses would require unambiguous backbone assignments and NOE data for the apo form of AGP2. As noted, the shift map may be imperfect since some apo peaks may be mistakenly assigned to specific residues if they happen to move close to a peak from the liganded spectrum by chance. A full set of backbone assignments for the apo protein would also allow a quantitative analysis of peak shifts to be undertaken with confidence. The experiment could then be extended to understand the binding of further ligands to AGP2

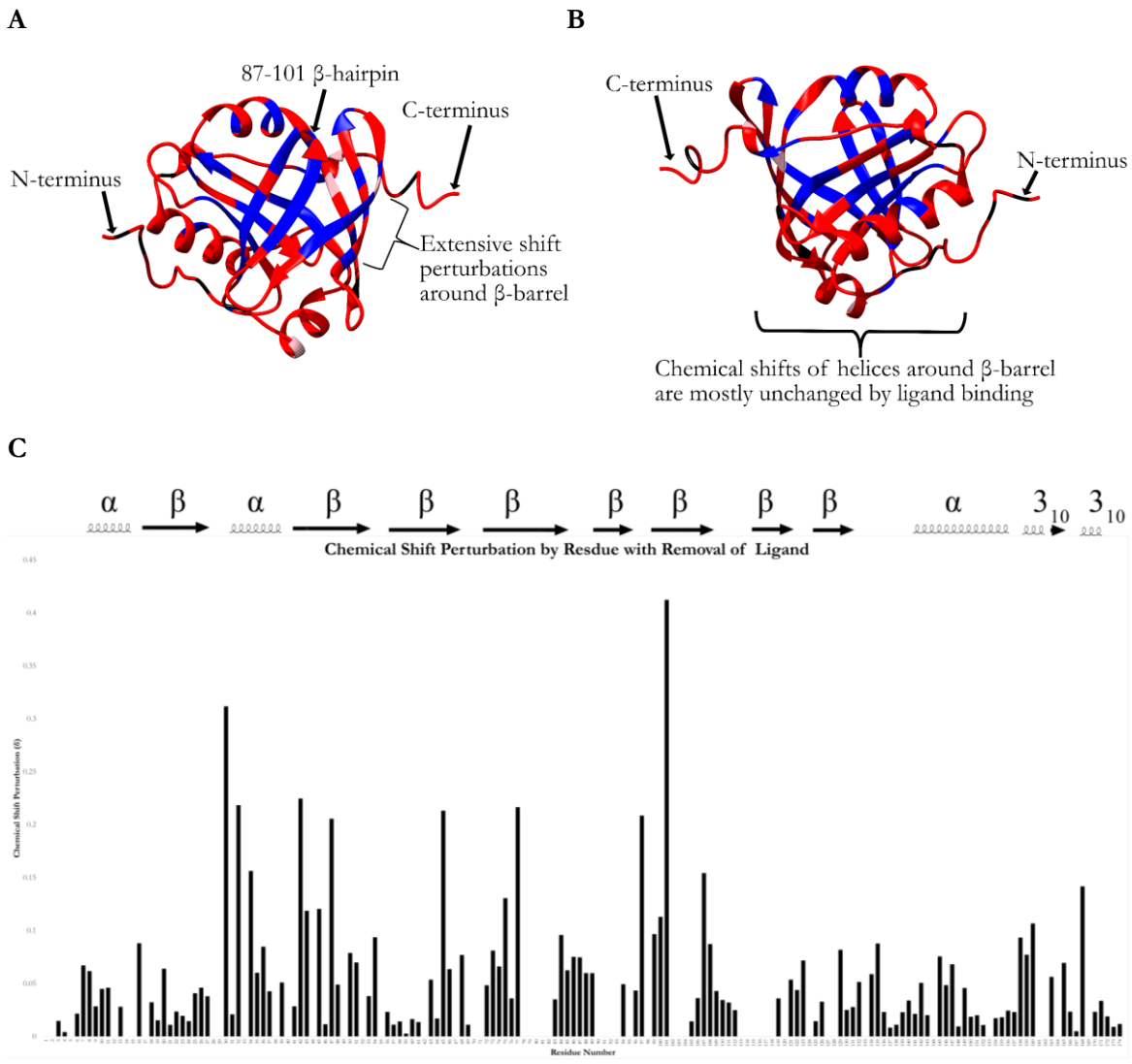


Figure 2.51 Colour map of AGP2 apo->UCN-01 bound titration data. Residues that are identifiable across spectra are coloured red, unidentifiable residues are coloured blue, proline residues are coloured black and peaks that are not visible at 700 MHz or are heavily overlapped are coloured pink. **(A)** The peaks of the β-barrel, particularly the 87-101 β-hairpin, were significantly shifted in the apo spectrum relative to the liganded spectrum indicating that significant conformational changes occur in these parts of the protein. **(B)** The helices arranged around the β-barrel do not exhibit many shifted peaks between the apo and liganded spectra indicating that they do not undergo significant structural changes during ligand binding. **(C)** To produce a quantitative map of chemical shifts, the distances between peaks assigned in the apo spectrum from those in the holo spectrum were plotted. When peaks could not be found in the apo spectrum a value of 0 was used. This chart also shows many of the largest shifts that can be observed in the β-barrel region, although many of those peaks are not identifiable in the apo spectrum.

Relaxation data

To further characterise the UCN-01-AGP2 complex experiments were carried out at 700 MHz to derive the T_1 and T_2 relaxation times and NOE contributions to ligand signals. The AGP2 concentration was 750 μM with a three-fold excess of UCN-01. Each of these measures is linked to the motion of the protein. The best and worst exponential fits for the calculation of T_1 and T_2 time constants are displayed in Figure 2.52, demonstrating that the experiments produced robust decay series for all residues.

The calculated T1, T2 and NOE ratio parameters for each residue are displayed in Figure 2.53. The full relaxation parameter lists are displayed in appendices a.2-a.4. The termini have higher relaxation times and negative NOE ratios indicating independent mobility. The central residues of the protein have relatively uniform dynamics parameters with the exceptions of Asn34 and Leu104 that have elevated T_2 parameters, potentially indicating some independent motion, but given the limited extend of these T_2 deviations, further validation would be necessary.

To calculate valid order parameters for AGP2 further data sets at other field strengths are required, but the relaxation data displayed here indicate that AGP2-trunc-UCN-01 adopts a uniformly rigid conformation along the entire chain with the exception of the termini, which are more mobile. It is likely that greater variation would be observed across the sequence of the apo protein if the difficulty in observing and assigning all of its peaks could be overcome.

Though the relaxation parameters across the main body of the protein are relatively uniform there is a small amount of variation. Asn34 and Leu104, are have elevated T_2 values at 118.6 ms and 83.4 ms respectively. These residues both occur close to secondary structure transitions between β-sheets and coiled residues, potentially implying additional motions due to conformational plasticity of the sheet structure. The termini of protein are typically more flexible that the rest of the chain. These data indicate that the AGP2-trunc-

UCN-01 complex is relatively rigid with only the termini able to be more flexible. Peaks that were extensively overlapped were omitted as they could not be fitted confidently.

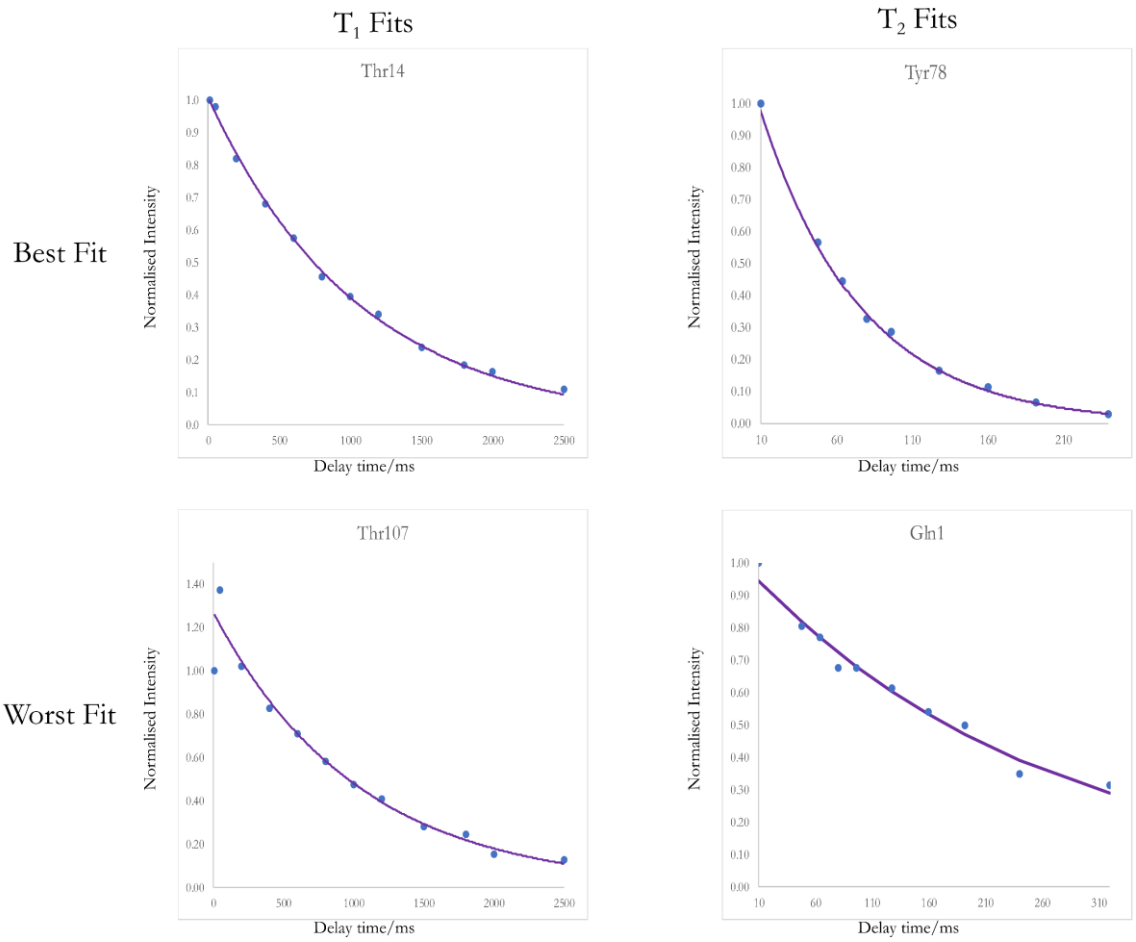


Figure 2.52 Exponential fits with the lowest and highest standard deviation values of all the assigned peaks for both T_1 and T_2 calculations. With all fits in between these extremes in their standard deviations these fits demonstrate that the fits are robust for all of the AGP2 residues.

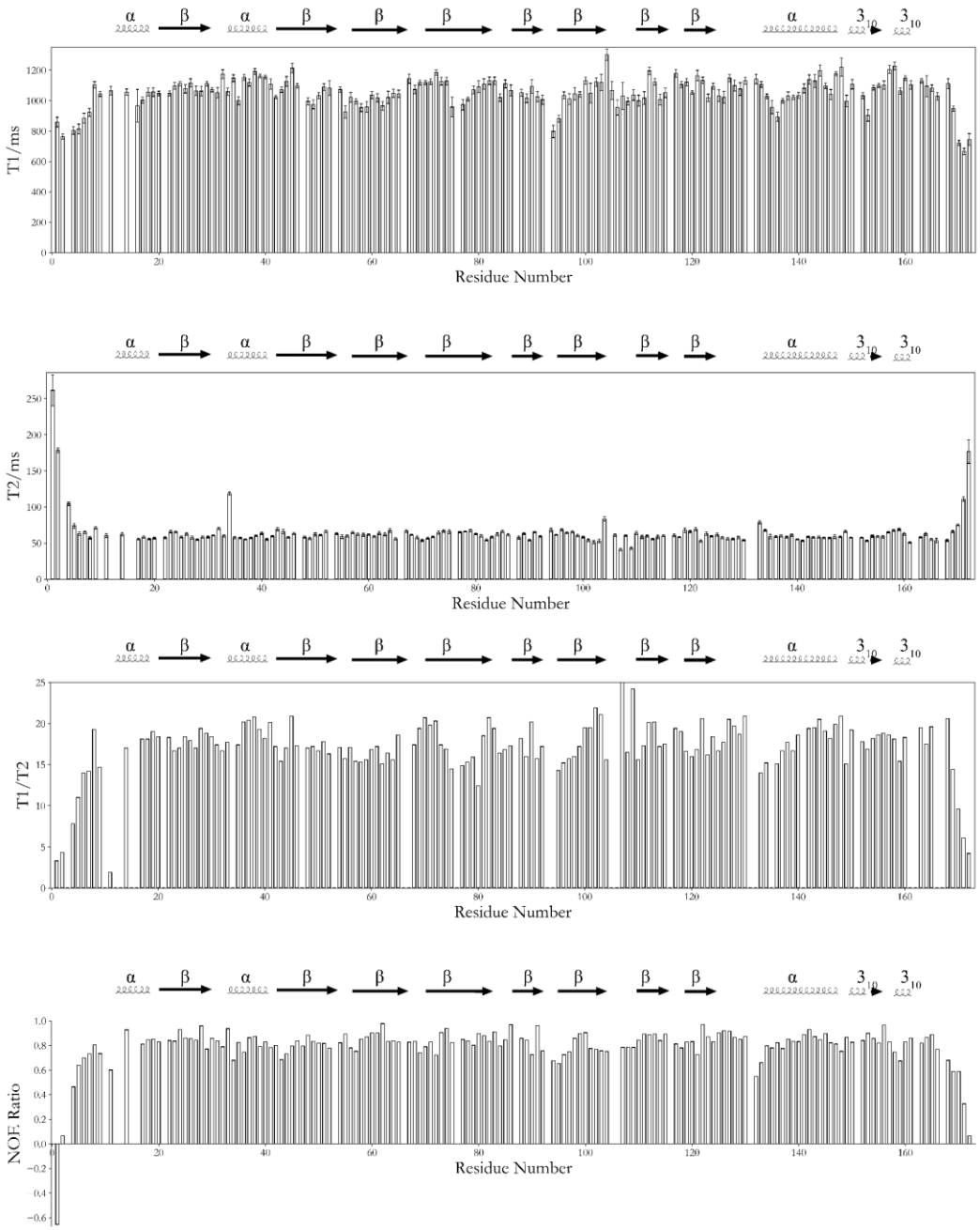


Figure 2.53 Dynamics data at 700 MHz T_1 relaxation times for each backbone ^{15}N spin by residue number, T_2 relaxation times for each backbone ^{15}N spin by residue number, T_1/T_2 ratio, and NOE signal ratios for each backbone ^{15}N spin by residue number. The secondary structure derived from the crystal structure (see below) is displayed above each chart. In each case the parameters are relatively uniform across the primary structure of the protein with parameters at the termini indicative of faster motion relative to the rest of the protein in those regions (residues 1-7 and 171-172). Some smaller deviations are also present in the main body of the protein.

Figure 2.54 shows the structure of AGP2 (hot pink) with residues with low NOE and/or T_1/T_2 ratios highlighted (blue). As noted, the N and C termini of the protein show the most evidence of independent mobility, with NOE ratios turning negative and T_1 and T_2 values becoming more alike. In addition, residues 94-95 and residues 132-133 also display NOE values below 0.7 and relatively low (<15) T_1/T_2 ratios. These residues occur between elements of secondary structure, with 94-95 between β-strands E and F, and 132-133 occurring immediately before helix #3. Combined with the drastic chemical shift changes that occur for signals corresponding to residues from the β-barrel, these data may indicate that, in the UCN-01-bound form, AGP2 is rigid around the ligand with some mobility in the residues surrounding these secondary structure elements.

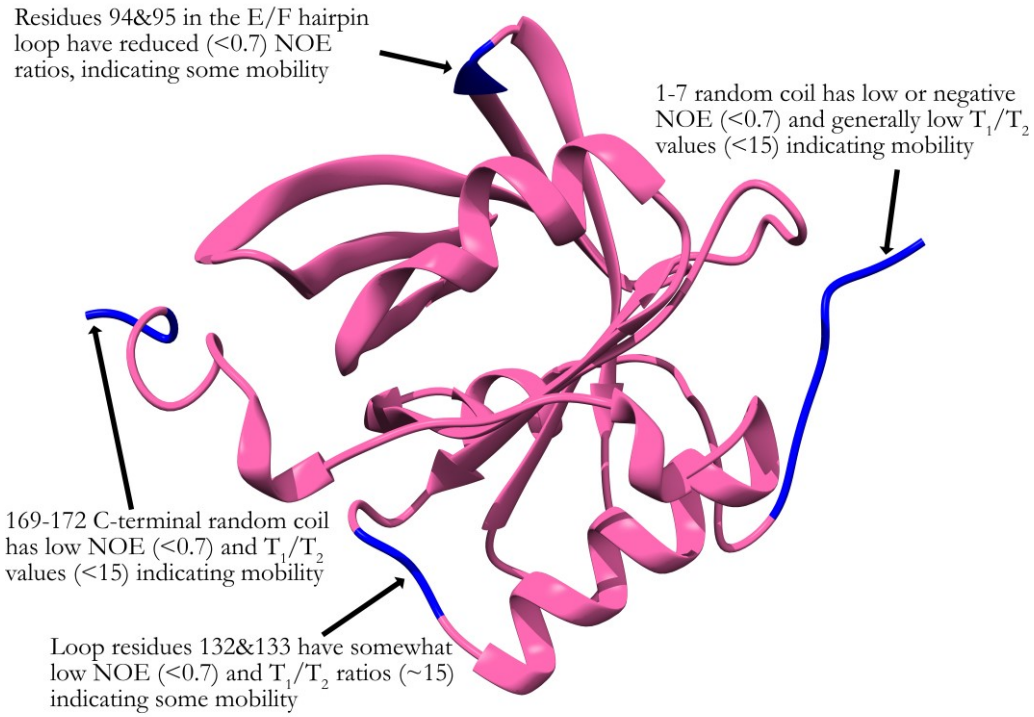


Figure 2.54 Structure of AGP2 (pink) with residues that the data indicate are independently mobile highlighted (blue). Residues that may be more mobile occur around the more rigid secondary structure elements involved in ligand binding or at the termini of the protein.

Crystal structure of AGP2-trunc bound to UCN-01

To further investigate the binding of UCN-01 to AGP2 co-crystallisation screens of AGP2 with UCN-01 were prepared. These yielded conditions suitable for crystal generation from which X-ray data, diffracting at 1.82 Å was collected and the structure, which is shown in Figure 2.55, was solved by molecular replacement compared to AGP2 bound to PEG [PDB: 3APU]. The generation of the crystal structure was a joint effort with Sara Ryan (purified protein), Nahida Akter (set up crystal trays), Alice Bochel (handled and shot crystals), Chris Williams (solved structure), and this author (optimised construct, grew cells, prepared inclusion bodies, coordinated efforts) (all from the University of Bristol).

The structural architecture of the protein is highly analogous to past structures of AGP and lipocalins in general (di Masi *et al*, 2016) with the same archetypal eight-stranded antiparallel β-barrel structure, labelled A-H in Figure 2.55. This β-barrel forms the main ligand binding pocket from its inward-oriented side chains. Three helices, labelled #1-#3(numbering from N-terminus) in Figure 2.55, are present in the structure. Helix #1 occurs before the β-barrel, helix #2 occurs between β-strands A and B and helix #3 occurs after strand H. The previous structure of AGP2 exhibits another helix closer to the C-terminus after

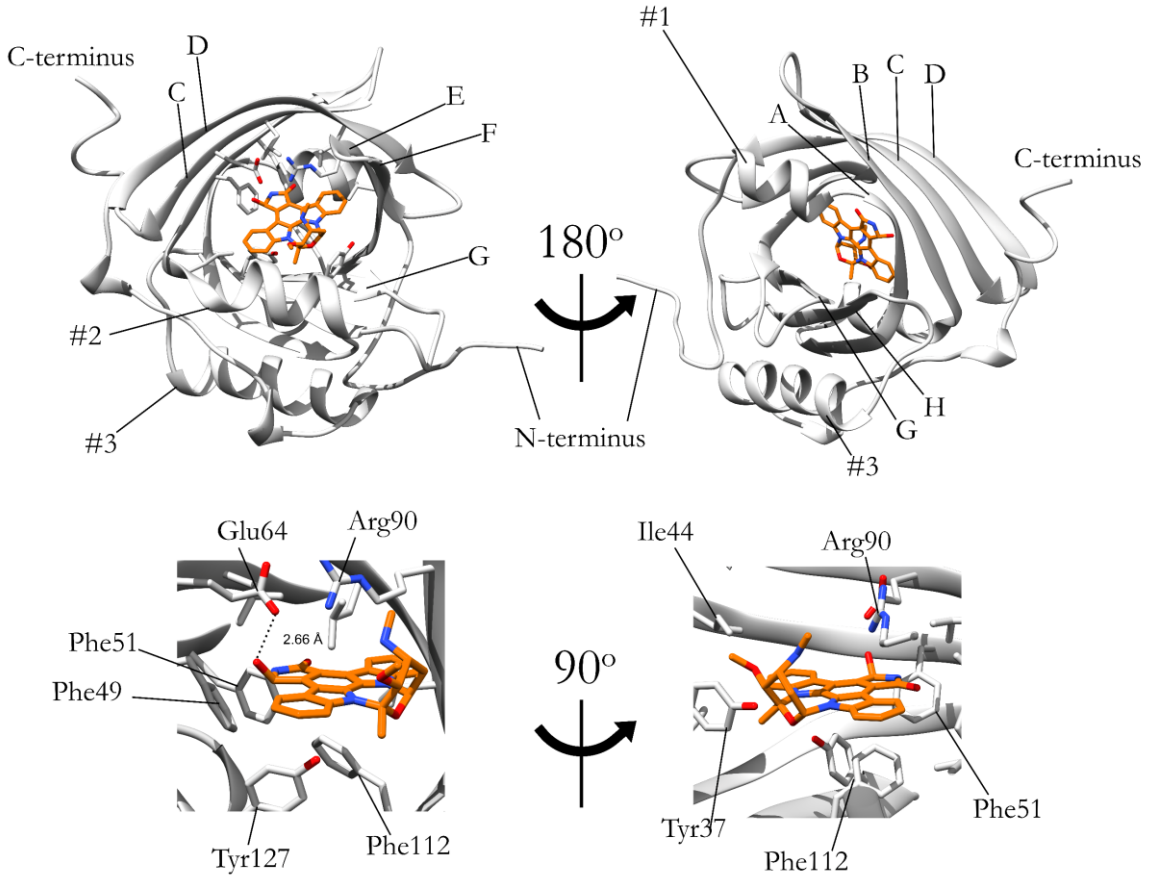


Figure 2.55 Crystal structure of UCN-01 (orange) in complex with AGP2 in its entirety from two faces (top) and viewed from two sides of the binding pocket (bottom). The β-sheets are labelled A-H as in published work (Nishi *et al*, 2011) and form a barrel consisting of antiparallel sheets. The pocket consists of aromatic side chains such as Phe49, a hydrogen bonding interaction from Glu64 and potential cation-π interaction from Arg90. For full table of parameters see appendix a.6.

the β-barrel (Nishi, *et al*, 2011), but the truncated construct presented here does not have this feature. The UCN-01 binding pocket consists of hydrophobic interactions with Phe49, Phe51, and Phe112 as with past structures with other ligands (Nishi *et al*, 2011). In addition, Glu64 forms a hydrogen bond to the 7(*R*) hydroxy group of UCN-01 and Arg90 may form a cation-π bond to main plane of the aromatic ring.

The observed secondary structure elements from the crystal structure were compared to those predicted by DANGLE (Cheung *et al*, 2012) from the chemical shifts of the assigned backbone spins at pH 6.5. The comparison is shown in Figure 2.56 and the DANGLE predictions are listed in appendix a.5.

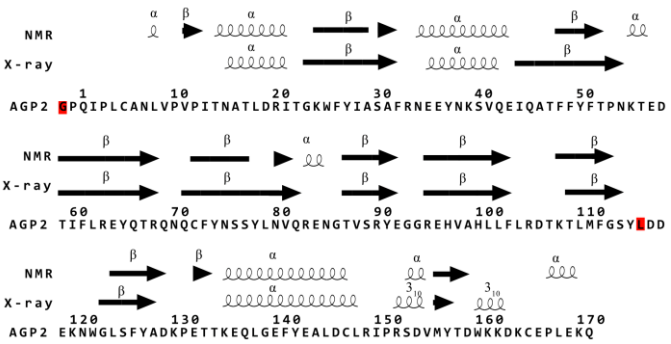


Figure 2.56 Comparison of secondary structure predicted from backbone chemical shifts using DANGLE (Cheung *et al*, 2012) with those observed from the crystal structure as analysed by STRIDE (Frishman & Argos, 1995) reveal broad similarity. The absence of the N-terminal helix in our crystal structure may be due to the effect of pH differences. Residues that were not visible in the considered data sets are highlighted in red.

The secondary structure calculated for AGP2 based on backbone chemical shifts was broadly homologous to that of the crystal structure. One deviation was that the helical secondary structure around residue 9 was predicted by DANGLE, but is not observed in our crystal structure. By contrast past structures do exhibit helicity at these residues (Nishi *et al*, 2011).

Given the limited extent of these predicted secondary structure elements it may not be reasonable to draw any conclusions from them. The secondary structure predictions from the backbone chemical shifts were almost identical to those observed in the crystal structure and so it is reasonable to believe that AGP2 adopts a highly similar conformation in both the UCN-01 bound crystal and under our NMR conditions making the comparisons of inferences from both methods legitimate.

Although solved at pH 7.4 the structure is highly homologous to previous AGP2 structures crystallised at pH 4.6 (Figure 2.57). The conformation of UCN-01 in the binding pocket is most similar to the drug CPZ. Both CPZ and UCN-01 are bound with aromatic ring systems in a similar plane (Figure 2.57 (A)) perpendicular to that of AMT and DSP (Figure 2.57 (C)+(D)). UCN-01 and CPZ are believed to bind to both AGP1 and AGP2 each with comparable affinity for the two AGP variants. In the literature the specificity of binding to one variant of AGP or the other is reported to be due to ligand interactions with Phe112 and Ser114 (residues for AGP2) (Nishi *et al*, 2011). In AGP1 Phe112 is replaced by Leu112 and it is possible that this more planar side chain may prevent AMT and DSP from binding in their preferred orientation by clashing with their ring systems. On the other hand, CPZ and UCN-01 may accommodate this change more easily as their aromatic systems allow more space between themselves and the backbone for the Leu112 side chain. Ser114 is replaced by Phe114 in AGP1 and is believed to clash with AMT and

DSP (Nishi *et al*, 2011). It is unclear from the structure whether this replacement is a specificity determinant. Mutagenesis experiments producing AGP1/2 hybrids in vitro or in silico may help differentiate which of these interactions are more essential.

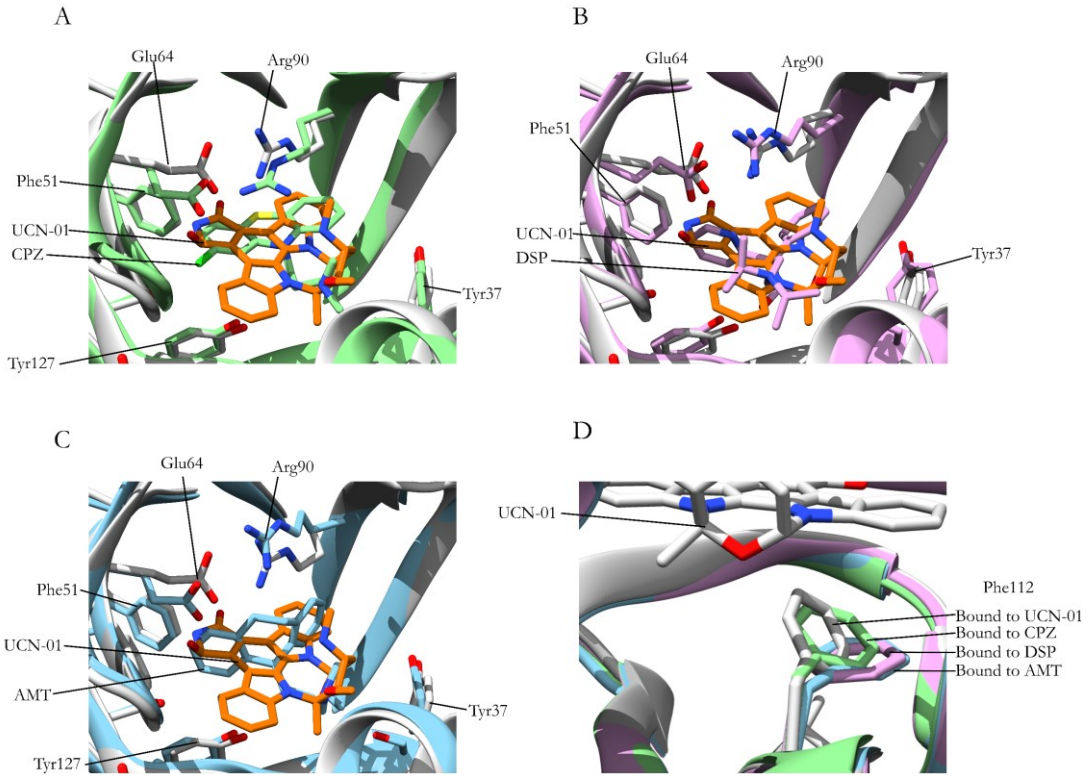


Figure 2.57 (A) Comparison of the binding of UCN-01 and CPZ to AGP2. The aromatic ring systems adopt a similar conformation along the same plane, but UCN's larger size may produce additional stabilising hydrophobic interactions and the hydroxyl group may hydrogen bond with Glu64. These features may explain UCN-01's higher affinity **(B)** Comparison of the binding of UCN-01 and DSP binding do AGP2. The aromatic groups of DSP are tilted perpendicular to UCN-01's aromatic system. This feature may explain why UCN-01 may bind to both AGP1 and AGP2 whilst DSP is specific to AGP2 **(C)** Comparison of the binding of UCN-01 and AMT to AGP2. As with DSP the aromatic groups of AMT are perpendicular those of UCN-01. Arg90 is therefore not able to form a cation-π interaction with the ring system unlike in the UCN-01 structure. **(D)** Comparison of Phe112 conformations between all existing AGP2 structures. The aromatic systems of nonspecific AGP1/2 ligands UCN-01 and CPZ allow the Phe112 sidechain to move further into the binding pocket than it can with AGP2-specific ligands. All alignments were caried out based on protein backbone in Chimera.

The interaction between Arg90 and the UCN-01 aromatic system is of interest as π-stacking interactions are observed more frequently in published structures than T-shaped π-cation interactions, such as the one observed here (Kumar *et al*, 2018). It is likely that a salt bridge with the Glu64 residue influences the orientation of Arg90, making it more likely to adopt the observed conformation. CPZ, which otherwise binds in a similar fashion to UCN-01, interacts with Arg90 by a more typical stacking interaction and concomitantly Glu64 also changes in orientation. The π-systems of the two other ligands, for which crystal structures have been solved, are oriented perpendicular to those of UCN-01 and CPZ and so further comparisons of Arg90-π interactions are unavailable.

In addition to providing information on the specificity of AGP binding to various ligands these data also rationalise the affinities of staurosporine derivatives other than UCN-01 to AGP. The staurosporine ligands

(structures in Figure 2.58) that have been tested for AGP binding differ in the hydroxylation state and stereochemistry on the pyrrole ring.

UCN-01 itself has a hydroxyl group on the same face of its aromatic system as the basic amine group. This hydroxyl is within hydrogen bonding distance (2.66 Å) to Glu64 in our crystal structure (shown in the bottom left panel of Figure 2.55) which may result in energetically more favourable binding. Staurosporine instead does not possess a pyrrole hydroxylation, which may explain its weaker affinity, lacking this enthalpic contribution. UCN-02, which has the reverse stereochemistry at the pyrrole ring has the weakest affinity of the 3 for AGP (Katsuki *et al*, 2004) and this may be due to the unfavourable position of the hydroxyl group in the hydrophobic pocket formed by Phe49 and Tyr127.

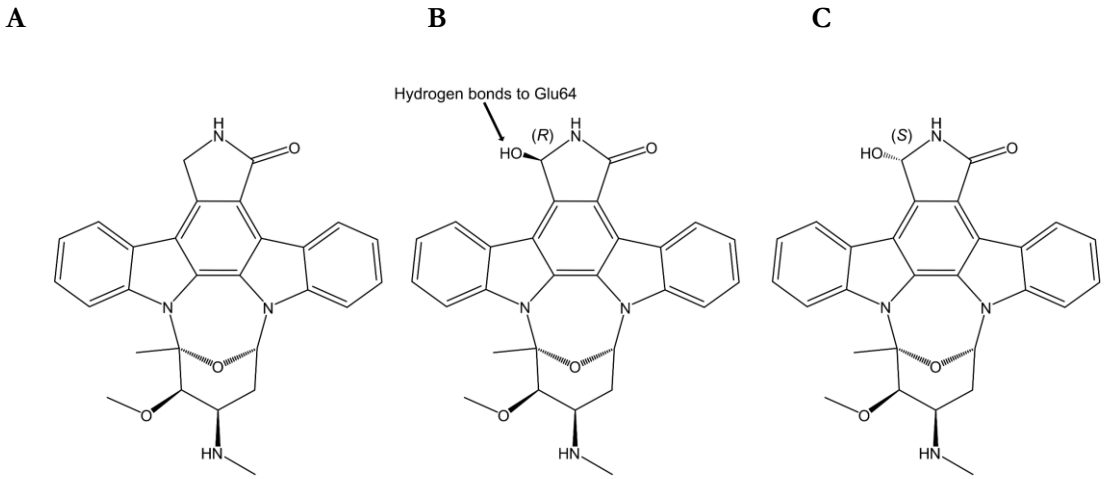


Figure 2.58 (A) Staurosporine **(B)** UCN-01 **(C)** and UCN-02. These compounds differ by the hydroxylation on the pyrrole ring. UCN-01 has a higher affinity (3.5 nM) for AGP2 than its analogues staurosporine (88 nM) and UCN-01 (676 nM). We can rationalise UCN-01's higher potency based on our crystal structure as being due to the hydrogen bonding interaction with Glu64.

Potential UCN-01 Modifications to Abolish AGP2 Affinity

The interest in studying UCN-01's binding to AGP2 stems from the potential to redesign UCN-01 in such a way as to maintain its affinity for its kinase targets whilst reducing its affinity to AGP2.

As discussed in the introduction, UCN-01's targets include the cell cycle proteins CDK1 and Chk1 for which high-resolution crystal structures have been determined. These two crystal structures of UCN-01 with CDK1 and Chk1 together with the AGP2 structure provide a structural framework for rational chemical modification.

Comparison of UCN-01 Binding to Chk1 and AGP2

Figure 2.59 shows the structure of UCN-01 bound to Chk1 compared to the binding pocket of AGP2 aligned to UCN-01's conformation.

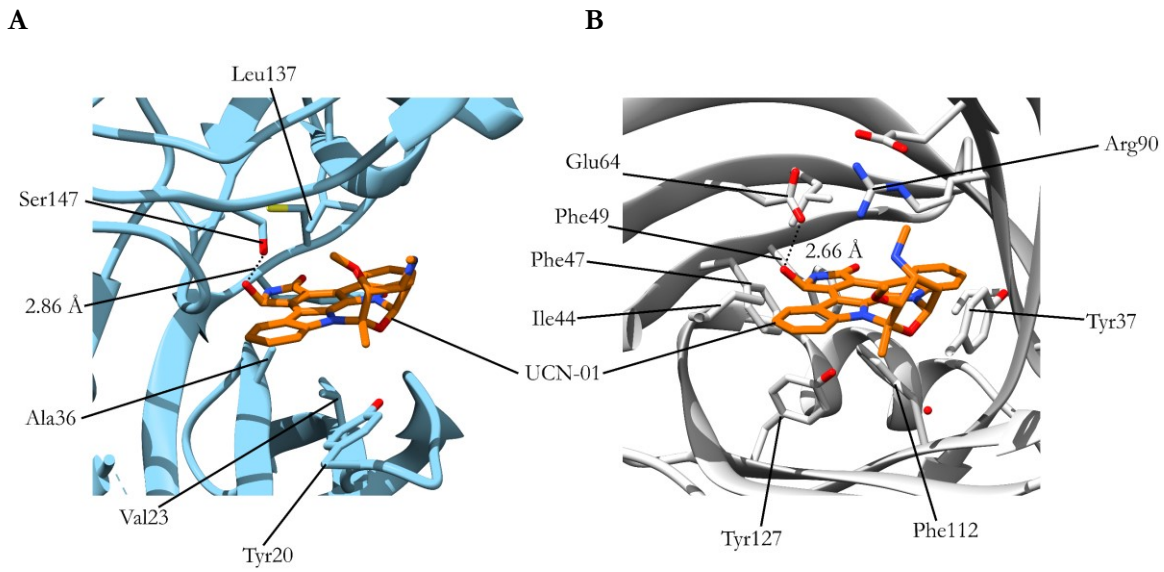


Figure 2.59 (A) Structure of the UCN-01-Chk1 binding pocket [PDB: 1NVQ]. The UCN-01 hydrophobic ring system is "sandwiched" in between hydrophobic residues of Chk1. The hydroxy-pyrrole group of UCN-01 forms a 2.86 Å hydrogen bond to Ser147. **(B)** UCN-01 forms extensive hydrophobic contacts to AGP2 as it does with Chk1 and a hydrogen bond from the hydroxy-pyrrole. The cation-π interaction to the aromatic system differs from Chk1's hydrophobic interactions.

The binding mode of the ligand to these two very different proteins has remarkable similarities, but also some differences that could potentially be exploited. UCN-01 has greater potency as a Chk1 inhibitor than staurosporine (Zhao *et al*, 2003) due to the presence of a hydrogen bonding interaction between the pyrrole hydroxyl group in UCN-01 and Ser147 in much the same way that UCN-01 forms a favourable interaction with Glu64 in AGP2. The effect of this modification on affinity for AGP2 appears more significant than for Chk1 (Katsuki *et al*, 2004), but nonetheless the commonality makes modification less promising than other moieties.

The aromatic system forming the central core of UCN-01 forms hydrophobic interactions with non-polar residues in the binding pocket in both structures. In Chk1 these residues are Ala36, Val23 and Leu137. AGP2 uses a cation-π interaction between Arg90 and the ring on its basic face, but mainly hydrophobic interactions on the reverse face as with the Chk1 complex.

Given the common importance of interactions of the hydroxy-pyrrole and central aromatic system the other parts of UCN-01 must be investigated for modification. Whilst in AGP2, UCN-01 is surrounded by a β-barrel, in Chk1 it is sandwiched between two β-sheets. It may therefore be possible to introduce substitutions on the UCN-01 aromatic system that protrude out of the sides of Chk1's "sandwich" into space but which clash sterically with the sides of AGP2's β-barrel. The positions of such potential substitutions are shown in Figure 2.60.

Modifications at one of these locations could serve a dual purpose of improving the solubility of UCN-01 whilst abrogating AGP2 affinity by appropriate selection of modifications.

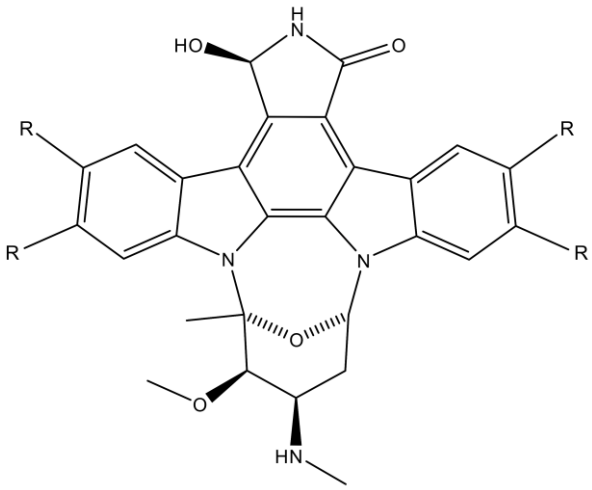


Figure 2.60 The indicated R groups could be substituted from -H to hydrophilic groups to abrogate AGP2 binding sterically and improve UCN-01's solubility.

Figure 2.61 shows the structures of some potential modified UCN-01 analogues. Introducing $-NO_2$ as a ring substituent may increase the steric bulk of the compound whilst improving solubility. An amidine $-C(NH)(NH_2)$ group would serve the same function to a greater degree by introducing a further basic group to the ligand. A carboxylic acid group may also be considered for increased steric bulk, but with the added effect of making the compound a zwitterion at physiological pH which may further assist in abrogating AGP2 binding as AGP2 preferentially binds basic drugs (Hervé et al, 1998).

Abrogating AGP2 affinity sterically is achievable with our knowledge of the crystal structure, but maintenance of Chk1 affinity is less certain. Preliminary in silico docking could be used at low cost to triage modifications likely to be effective, but ultimately synthesis of modified UCN-01 compounds and measurement of their affinities for Chk1 and AGP2 would be necessary.

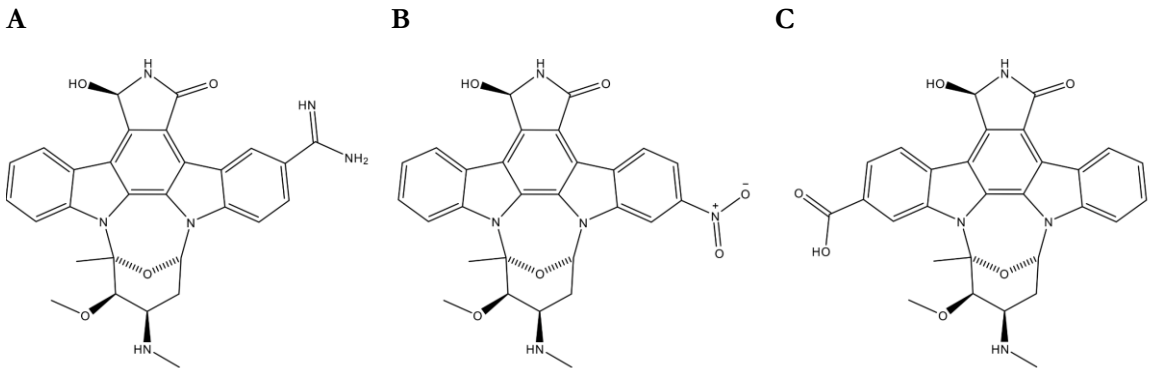


Figure 2.61 Potential substitution of an amidine group **(A)**, nitro group **(B)** or carboxylic acid group **(C)** onto UCN-01. Modifications (A) and (B) would likely clash sterically with the 87-100 β-hairpin of AGP2 whilst protruding into solution when bound to Chk1 thus potentially making UCN-01 more specific for Chk1. Modification (C) would likely clash with the opposite face of the β-barrel to (A) and (B).

Comparison of UCN-01 Binding to CDK2 and AGP2

UCN-01's cell cycle progression inhibition activity is largely due to its binding to CDK2. A comparison between binding of UCN-01 to CDK2 (Johnson e*t al*, 2002) and AGP2 is shown in Figure 2.62.

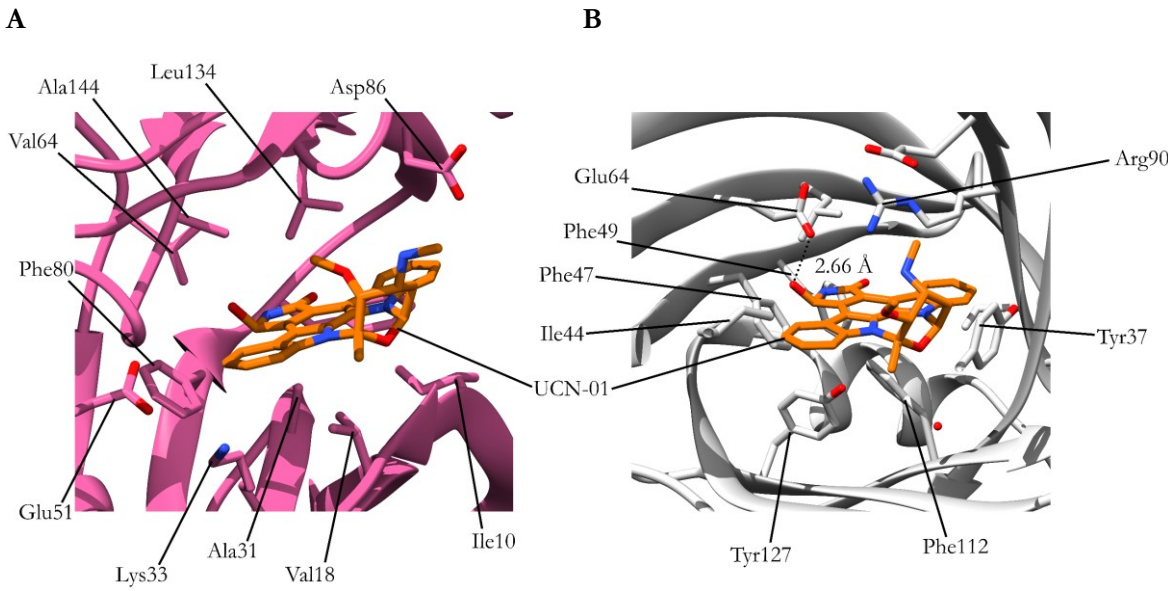


Figure 2.62 (A) Structure of CDK2 in complex with UCN-01 [PDB: 1PKD]. As with Chk1 hydrophobic residues on either side of the aromatic system, in this case Val18 and Leu134, form binding contacts between ligand and protein. Unlike both AGP2 and Chk1 there are no residues near the hydroxy-pyrrole group to form H-bonds. **(B)** Our structure of UCN-01 in complex with AGP2 for comparison. Unlike CDK2 H-bonding from Glu64 to UCN-01 stabilises ligand binding.

Similarly, to Chk1 binding, UCN-01 occupies a site between hydrophobic residues on either face of the aromatic system in the CDK2 binding cleft. In contrast to the binding mode of the compound in both AGP2 and Chk1 there are no hydrogen bond donors or acceptors proximal to the hydroxy-pyrrole group of UCN-01. This fact could be exploited to produce a modified compound whose affinity for AGP2 is abrogated. Some potential analogues are shown in Figure 2.63.

Given the published data on the affinities of staurosporine 7-hydroxy analogues for AGP2 (Katsuki *et al*, 2004) we may believe with a degree confidence that the proposed modifications to UCN-01 would reduce its affinity for AGP2. Less certain is whether affinity for CDK2 would be maintained. As with Chk1, in silico docking followed by in vitro affinity assays should be employed to measure the success of UCN-01 optimisation.

A **B** **C**

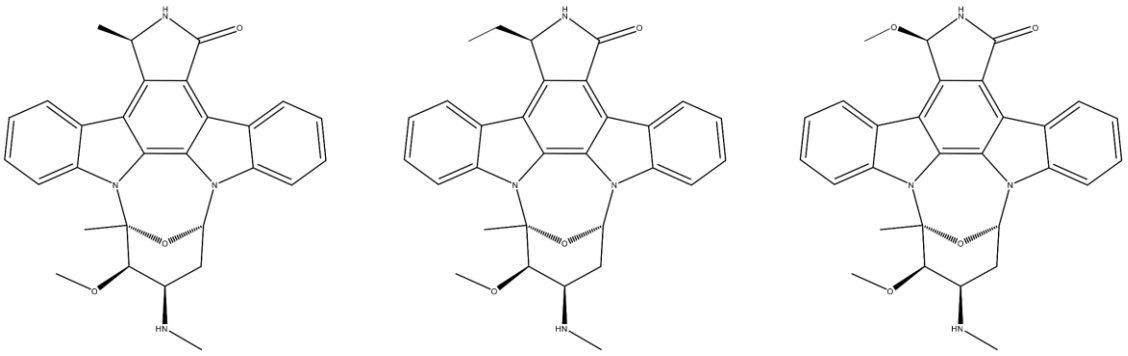


Figure 2.63 Modified UCN-01 to remove its hydrogen bond donor capacity by replacing the hydroxy-pyrrole group with a methyl group **(A)**, ethyl group **(B)**, or methoxy group **(C)**. Each of these modifications removes the capacity of UCN-01 to hydrogen bond to Glu64 in AGP2. (B)&(C) may also clash sterically with Glu64 or the β-barrel backbone of AGP2.

Conclusions

Presented here is a novel method of producing folded AGP2 with ligand binding activity, 98% backbone assignments of the protein with relaxation parameters at 700 MHz, and a crystal structure of the important anti-tumour compound UCN-01 bound to AGP2.

The protein expression and purification method represents an improvement in tractability over the published methods for generating pure recombinant AGP2 as it allows for expression of the protein in BL21(DE3) cells rather than cells specialised for cytoplasmic folding of disulphide bonded proteins.

The NMR data are noteworthy as they indicate that AGP2 may be more disordered in the apo form than would have been assumed based on published crystal structures. The relaxation parameters, on the other hand, show that the main body of the protein displays mostly uniform dynamics when binding UCN-01 in solution. This rigidification is corroborated by CD data showing a significant increase in melting temperature upon ligand binding. NMR spectral quality is also considerably improved in the presence of UCN-01 relative to the apo form, consistent with conformational rigidification upon ligand binding.

The crystal structure reveals the binding mode of the highest known affinity ligand for AGP2 and rationalises how the stereochemistry of the staurosporine pyrrole ring impacts binding affinity. These findings open new avenues for redesigning UCN-01 to maintain its anti-kinase activity whilst avoiding high-affinity plasma binding. It may be that by circumventing AGP2 binding a modified UCN-01 may be a fruitful lead in oncology therapeutics.

Materials and Methods

Precision Protein Marker	Biorad
1 kB DNA Ladder	Bioline
Restriction Enzymes + reagents	ThermoFisher
T4 Ligase	ThermoFisher
KOD polymerase	Novagen
GenElute Gel Extraction Kit	Sigma
BL21 Competent Cells	Novagen
SHuffle® Competent Cells	New England Biolabs
NEB5α competent Cells	New England Biolabs
In-Fusion Cloning Enzyme	Takara
DNA Constructs	ThermoFisher
Amitriptyline >99% purity	Biotechne
Staurosporine >95% purity	VWR
UCN-01 >97% purity	Sigma
3C Protease	Generon
$^{15}NH_4Cl$ >99% purity	Goss Scientific
$^{13}C_6H_{12}O_6$ >99% purity	Goss Scientific
Sequencing	Genewiz
Crystal Screen	Molecular Dimensions

Constructs and Vectors:

Initial constructs ordered from ThermoFisher Geneart were named Full and Short containing residues 1-183 and 1-175 of the mature protein, respectively. As in published crystallography work residue C149 was replaced with arginine to improved sample homogeneity. The restriction sites closest to the AGP2 sequence were Nco1 at the 5' and Xho1 at the 3' resulting in vector artefacts at either end. The final sequences were as follows with vector artefacts highlighted in blue:

Full:

MGQIPLCANLVPVPITNATLDRITGKWFYIASAFRNEEYNKSVQEIQATFFYFTPNKTEDTIFLR
EYQTRQNQCFYNSSYLNVQRENGTVSRYEGGREHVAHLLFLRDTKTLMFGSYLDDEKNWGL
SFYADKPETTKEQLGEFYEALDCLRIPRSDVMYTDWKKDKCEPLEKQHEKERKQEEGESLE

Short:

MGQIPLCANLVPVPITNATLDRITGKWFYIASAFRNEEYNKSVQEIQATFFYFTPNKTEDTIFLR
EYQTRQNQCFYNSSYLNVQRENGTVSRYEGGREHVAHLLFLRDTKTLMFGSYLDDEKNWGL
SFYADKPETTKEQLGEFYEALDCLRIPRSDVMYTDWKKDKCEPLEKQHEKESLE

The constructs were ordered in the pMA-T vector

The pET28 vector introduced MG as a vector artefact at the N-terminus and LE at the C-terminus followed by the His_6 tag

The pET26 vector introduced the PelB signal sequence (MKYLLPTAAAGLLLLAAQPAMA) at the N-terminus for periplasmic secretion, MG as a vector artefact at the N-terminus and LE at the C-terminus followed by the His_6 tag

The pETM60 vector introduced the NusA sequence at the N-terminus followed by the NusA His_6 tag and a TEV cleavage sequence ENLYFQS/GA and an MG vector artefact. At the C-terminus the vector contained the LE vector artefact and His_6 tag

The pOPINF vector introduced a start codon, an N-terminal His_6 tag with and a linker to a 3C protease cleavage site. Together this sequence reads: MAHHHHHHSSGLEVLFQGP. As InFusion cloning does not rely on restriction sites the MG and SLE vector artefacts present in the classically cloned constructs were absent.

The AGP2-trunc construct contained a stop codon replacing Glu173, but was otherwise unchanged from the full length pOPINF construct.

Classical Cloning of AGP2:

For cloning into pET28, pET26 and pETM60 AGP2 donor and receiver vectors was incubated with 5% each of Xho1 and Nco1 enzyme in FastDigest buffer for 2 hours at 37 °C and the fragments were separated on a 0.7% agarose gel. AGP2 construct and receiver vector were purified by GenElute gel extraction kit by manufacturer protocol. Ligations were carried out with T4 ligase by manufacturer protocol. Cloning product was transformed into high copy number 5α cells (NEB) by manufacturer protocols and pure DNA was obtained by miniprep kit (Sigma) by manufacturer protocols.

In-Fusion Cloning of AGP2:

Primers containing the pOPINF vector In-Fusion adaptor sequences (Fwd: aagttctgtttcagggcccg, Rev: atggtctagaaagcttta) and 30-40 bases of the AGP2 construct (ThermoFisher) were used to amplify the AGP2 construct by PCR using KOD polymerase (Novagen) by commercial protocols. AGP2 was inserted into pOPINF by In-Fusion enzyme manufacturer protocols. Cloning product was transformed into high copy number 5α cells (NEB) by manufacturer protocols and pure DNA was obtained by miniprep kit (Sigma) by manufacturer protocols.

Expression of AGP2 in Rich Media:

Cells transformed with the vector containing AGP2 were precultured from single colonies in LB + selection marker for 16h and inoculated into LB/2YT growth media + 2% glucose w/v at a 1:100 volumetric ratio. Protein expression was induced at an OD of 0.7-1.0 with 0.25 mM IPTG for expression of inclusion bodies or 0.1 mM IPTG for soluble expression and the temperature was lowered to 30°C for 16 hours. Cells were harvested by centrifugation (6k rpm, 10 min, 4 °C) and stored (-20°C) until use.

Expression of AGP2 in M9 by Media Exchange:

Cells transformed with the vector containing AGP2 were grown as above in rich media to an OD of 1.0-2.0. Cells were removed from the media by centrifugation (4 krpm, 10 min, 4°C), resuspended in sterile 200 ml M9 (12.9 g $Na_2HPO_4.6H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, pH 7.0), removed from the M9 wash by centrifugation (4 krpm, 10 min, 4°C) and resuspend in 500 ml M9 per initial litre of rich media growth. Cells were maintained in M9 media (30 min, 37°C) to burn ^{14}N. Carbon and nitrogen sources were added (made up to 25 ml H_2O per litre expression, 2.5 % glucose w/v, $MgSO_4$ (for 1 mM final), 1:1000 of final trace metals(20 mM $CaCl_2$, 2 mM $CoCl_2.6H_2O$, 2 mM $CuCl_2.2H_2O$, 60 mM H_3BO_3.HCl, 10 mM $MnCl_2.4H_2O$, 2 mM $Na_2MoO_4.2H_2O$, 2 mM $Na_2SeO_3.5H_2O$, 2 mM $NiCl_2.6H_2O$, 2 mM $ZnSO_4.7H_2O$), 1 g $^{15}NH_4Cl$) and induced with 0.25 mM IPTG for expression of inclusion bodies or 0.1 mM IPTG for soluble expression and the temperature was reduced to 30°C. Cells were incubated for 16 hours and harvested by centrifugation (6k rpm, 10 min, 4 °C). Cell pellets were stored (-20 °C) until use.

Expression of AGP2 in M9:

Cells transformed with the vector containing AGP2 were precultured from single colonies in LB + selection marker for 16h and inoculated directly into M9 + carbon and nitrogen sources (made up to 25 ml H_2O per litre expression, 2.5 % glucose (^{13}C labelled or not as required) w/v, $MgSO_4$ (for 1 mM final), 1:1000 of final trace metals, 1 g $^{15}NH_4Cl$). Cells were grown to OD 0.3-0.6 and induced with .25 mM IPTG for expression of inclusion bodies or 0.1 mM IPTG for soluble expression and the temperature was reduced to 30°C. Cells were incubated for 16 hours and harvested by centrifugation (6k rpm, 10 min, 4 °C). Cell pellets were stored (-20 °C) until use.

Purification of AGP2 by IMAC:

Cells were lysed in 20 mM Tris, 50 mM NaCl, 0.1 % CHAPS, pH 8.0 by Sonication in ice-ethanol bath (3 times 5:10-on:off 2 min 30 sec, amplitude 80%). The lysate was clarified by centrifugation (14.5 krpm, 15 min, 4°C). The soluble lysate was flowed through a 5 ml Ni-NTA column and washed with 20 mM Tris, 50 mM NaCl, 10 mM imidazole, pH 8.0. Soluble AGP2 (for all constructs) was eluted with a 10 mM – 800 mM imidazole gradient. Imidazole was removed by a gel filtration in S75 resin into 20 mM Tris, 150 mM NaCl, pH 8.0.

Preparation of AGP2 Inclusion Bodies:

Cells expressing AGP2 were resuspended in 20 mM Tris, 50 mM NaCl, 1 mM EDTA, 1 mM benzamidine, pH 8.0 and lysed by sonication in ethanol-ice bath (2 times 5:15-on:off 2 min 30 sec, amplitude 75%). Insoluble material was separated by centrifugation (8700 rpm, 30 min, 4°C). The insoluble material was washed by sonication in lysis buffer as previously and removed from the soluble impurities by centrifugation (8700 rpm, 30 min, 4°C). 2 sonication-wash cycles as above were carried out in high salt buffer 20 mM Tris, 500 mM NaCl, 1 mM EDTA, 1 mM benzamidine, pH 8.0. The washed inclusion bodies were separated by centrifugation (8700 rpm, 30 min, 4°C) and dissolved in minimal 8-10 M urea, 50 mM Tris, pH 8.5 + a few drops conc. NaOH. Solubilised inclusion bodies frozen until further use.

Purification of AGP2 by Inclusion Body Rapid Dilution Refold and Size Exclusion Chromatography:

AGP2 inclusion bodies were reduced with 10 mM DTT (1 hour, on ice). Insoluble debris was removed by centrifugation (15 krpm, 10 min). Reduced unfolded AGP2 was added dropwise to 1 M arginine, 50 mM Tris, pH 8.5, 1 mM EDTA, 1 mM benzamidine, 3.5 mM cystamine, 6.5 mM cysteamine. The refold was incubated for 24-72 hours at 4°C. The refolded AGP2 was concentrated in a pressure cell on a 10 kDa membrane as required and monomeric protein was separated by gel filtered in S75 resin into 20 mM Tris, 150 mM NaCl, pH 8.0 and concentrated as required by spin concentrator.

Cleavage and Negative IMAC of AGP2:

AGP2 solution (1-2 mg/ml) in 20 mM Tris, 150 mM NaCl, pH 8.0 was incubated (16 hours, room temperature) with a 10:1 mass ratio to in house 3C protease (stored in 20 mM Tris, 1 mM beta-mercaptoethanol (BME), 150 mM NaCl, 50% glycerol w/v, pH 8.0). Following incubation, the cleavage reaction was passed through a 5 ml Ni-NTA column equilibrated in 10 mM Na_2HPO_4/NaH_2PO_4, 100 mM NaCl, pH 7.4. The negatively purified AGP2 was exchanged into 10 mM Na_2HPO_4/NaH_2PO_4, 100 mM NaCl, pH 7.4 in a spin concentrator and concentrated as required for further experiments.

SDS PAGE Analysis:

Gels were made up as 10% acrylamide (separating) and 4% acrylamide (stacking) in standard tricine buffer. Gels were stained with Coomassie-based stains.

CD Spectroscopy

AGP2 was exchanged into 20 mM NaP, 150 mM NaF, pH 7.2 by Zeba column. CD was run on an Applied Photophysics Chirascan instrument and the analysis was carried out with the instrument's inherent software v. 4.7.0.194.

NMR Experiments:

All 700 MHz and 600 MHZ double resonance experiments (HSQC, and TROSY HSQC) and triple resonance experiments (HNCA, HNCACB, HNCOCA, HNCOCACB) were collected by Chris Williams using standard Topspin settings using a Bruker 16.4 T (700 MHz ^{1}H observe frequency) spectrometer equipped with a 1.7 mm cryogenically cooled TCI probe in 90% H_20, 10% D_2O at 298 K. All spectra were analysed in CcpNmr analysis 2.4.2. AGP2 was observed in 10 mM NaH_2PO_4/Na_2HPO_4, 100 mM NaCl, with the ligand and pH specified in the relevant figure.

950 MHz double and triple resonance experiments were collected by Christina Redfield using standard Topspin settings using an Oxford Instruments 22.3 T (950 MHz ^{1}H observe frequency) spectrometer equipped with a high-sensitivity 5 mm cryogenically cooled TCI probe in 90% H_2O, 10% D_2O at 298 K. AGP2 was observed at a concentration of 1 mM protein, 2 mM UCN-01 in 10 mM NaH_2PO_4/Na_2HPO_4, pH 6.5, 100 mM NaCl.

NMR Titration Chemical Shift Mapping

In a reverse minimal shift approach, peaks in the apo spectrum were assigned based on their proximity to assigned peaks in the UCN-01 bound spectrum. Apo peaks closest to liganded peaks were assigned first and subsequently more distant apo peaks were assigned to the closest remaining unassigned UCN-01 peak provided that the peaks were closer than 0.1 ppm in the ^{1}H dimension and 0.68 ppm in the ^{15}N dimension. Heteronuclear shifts were weighted based on the respective ranges of assigned peak chemical shifts present in the two dimensions (^{1}H:4.504, ^{15}N:30.616) as described in published methods (Williamson, 2013). These ranges yielded a weighting of 0.147 ^{1}H ppm/^{15}N ppm. All visible assigned peaks in the UCN-01 spectrum were greater than 5% of the height of the largest peak (His172) and so peaks in the apo spectrum were only counted if they were more than 5% of the height of the largest apo peak (putative His172).

To produce the minimal chemical shift chart, all peaks in the apo spectrum that could be assigned based on the holo spectrum (including those further away than 0.1 ppm in the ^{1}H dimension and 0.68 ppm in the ^{15}N dimension) were assigned to the same residue as the closest holo peak and the chemical shift distance between the two (with the H:N distance conversion as above) was plotted.

Crystallisation of AGP2 and Structural Resolution

AGP2-trunc (purified by Sara Ryan) was cleaved and crystal conditions were screened from Structure 1+2 (Molecular Dynamics). The successful hit consisted of 0.5 mM AGP2 with a two-fold excess of UCN-01 in a 1:1 sitting drop with 1.4 M sodium citrate, 100 mM HEPES, pH 7.4.

Discussion

The studies of $A_{2A}R$ and AGP2 presented in this thesis are examples of two different approaches drawn from the NMR literature as discussed in the introduction. The techniques applied to the $A_{2A}R$ as a membrane protein are best compared to methodologies previously applied to the M2 proton channel and β_2-adrenergic receptor, whereas the work on AGP2 bears more resemblance to that on the Bcl-2 family. In each case appropriate NMR techniques have been applied to understand the "jigglings and wigglings" (Feynman, 1963) of these proteins and how their motions affect drug discovery campaigns.

^{19}F Tagging of the $A_{2A}R$ in Context

The ^{19}F NMR study reported here extends the application of GPCR ^{19}F tagging to investigate distinct antagonist-induced states and applies it to a novel ligand.

GPCRs in general have been investigated by NMR, but their large size and the necessary presence of a detergent micelle to maintain their stability, make structure determination by solution state NMR extremely challenging. The application of uniform isotope labelling to understand GPCR biology has nonetheless been attempted in the past, in particular a study of the GPCR analogue sensory rhodopsin (pSRII) has been conducted. The structure of this 7-TM helical protein was solved by NMR (Gautier *et al*, 2010) demonstrating that the most in-depth NMR study of 7-TM proteins is in principle possible, but to this author's knowledge it remains the only 7-TM helical protein the structure of which has been solved at atomic resolution by NMR. ssNMR is useful for larger systems such as genuine GPCRs by obviating the increased relaxation rate resulting from slow tumbling in solution. A structural model of the GPCR CXCR1 has been determined using ssNMR (Park *et al*, 2012), just as ssNMR was used to probe the structures of the M2 channel. However, unlike the case of the M2, the determination of the CXCR1 structure relied on bovine rhodopsin as a topology template and cannot be considered a true de-novo NMR structure. The chemical shifts of the protein spins were used to generate structural restraints to produce structure fragments to be mapped onto the template. It is also the case that some reviewers have questioned whether the resolution of the CXCR1 structural model is sufficient for drug discovery applications (Latek *et al*, 2012). The difficulties encountered in generating full NMR models of GPCRs necessitates the application of alternative methods to probe their structures and the ^{19}F study described in this thesis uses one such alternative method along with selective labelling.

Selective labelling or tagging of other GPCRs have been employed as alternatives to full assignment of uniformly labelled proteins to overcome spectral complexity. Chemical shift changes in selectively labelled β_1-adrenergic receptor were used to characterise the effects of multiple ligands on the receptor (Isogai *et al*, 2016; Solt *et al*, 2018). The $A_{2A}R$ itself has also been studied by this method (Clark *et al*, 2017). Just as in a uniformly labelled sample, in selectively labelling GPCRs the isotope labels act as chemical shift probes, but with the identity of the peaks determined by mutagenesis.

By contrast, the ^{19}F tagging approach, whilst being constrained to reporting on one locus per experiment, has a great advantage in tractability as the receptor need not be grown in expensive isotopically labelled

media. Further, the precise location of the label is predetermined by the location of the point mutation and so no further deconvolution is necessary.

For a research question that focusses on a specific part of the protein a single ^{19}F probe may be sufficient to yield all the necessary information. In our case by tagging K209$^{5.70}$ we were able to measure a response for both drug-sized small molecules such as Cmpd-1 and fragment sized caffeine. This single ^{19}F probe was sufficient to address whether helix V adopts a different conformation when in complex with the Cmpd-1 compared to those of other antagonists, negating the need for a full backbone chemical shift assignment. Given the commercially driven time pressures on drug discovery pipeline delivery in pharmaceutical companies, such considerations are of central importance to determine whether or not a method may be used.

The key differences between past studies of β_2-adrenergic receptor or $A_{2A}R$ (Liu *et al*, 2012; Ye *et al*, 2016) and this ^{19}F study are that an alternative method of receptor inactivation was being investigated and that Cmpd-1 is a novel ligand that the study sought to characterise. Past studies have, by contrast, been concerned primarily with characterising different active states and have investigated the receptor rather than the ligand. It is hoped that this ^{19}F study represents an extension of ^{19}F NMR to address a broader range of questions than it has been applied to in the past. The study also advances the integration of non-standard NMR approaches to investigate the properties of lead compounds in drug discovery as some reviewers have noted (Gusach *et al*, 2020).

Although our ^{19}F tagging approach yielded the necessary supporting evidence on the movement of helix V, it did not provide structural detail on the conformation that was adopted. Selective labelling and tagging cannot yield the NOE constraints necessary for a de-novo structure calculation in the same way that full backbone and sidechain assignments of uniformly labelled protein can. However, recently selective labelling has also been used to glean structural information about the β_2-adrenergic receptor by paramagnetic relaxation enhancement (PRE) (Imai *et al*, 2020). PRE causes increased relaxation rates of nuclei by the proximity of a paramagnetic centre, which allows some intramolecular distances to be determined. Paramagnetic tags may be chemically conjugated to a receptor in much the same way as ^{19}F tags, via thiol bonds to introduced cysteine residues. If a more detailed study of the effect of Cmpd-1 on helix V were to be conducted then this approach would be applicable as the introduction of several $A_{2A}R$ cysteine residues for ^{19}F tagging is already validated in terms of feasibility and impact on activity. Such a study could yield distance measurements allowing the movement of helix V to be quantified and so go beyond the qualitative observation of a novel state reported in this work.

Backbone Assignments of AGP2 in Context

The investigation into the binding of UCN-01 to AGP2 was pursued using a very different NMR methodology to that used for the $A_{2A}R$. The smaller size of AGP2 made classical NMR techniques, resembling those discussed for Bcl-2 in the introduction, possible and appropriate. The NMR work is nonetheless unorthodox in some ways.

Whilst NMR data was being gathered the crystal structure of AGP2 bound to UCN-01 was solved revealing its binding orientation. In this regard the work is analogous to that on the M2 proton channel where the binding mode of ligands was determined by crystallography prior to or concomitantly with NMR, but as with these cases the adaptability of NMR allowed a more complete investigation of the backbone dynamics using the NMR assignments than is possible with crystallography alone.

Neither in this study nor in previous AGP work has the crystal structure of AGP2 in the apo state been solved. The applicability of NMR to study apo AGP2 under a range of solution conditions is, therefore, important. In most CSP studies, as described for Bcl-2, the NMR assignments are determined for the apo protein and CSPs are measured by using this as a reference spectrum for comparison to the ligand-bound spectra which are not usually assigned. This work on AGP2 was carried out in the reverse order. This workflow was imposed by the dynamic nature of apo AGP2, evidenced by the loss of NMR signals due to intermediate exchange broadening such that assignment of the complete apo spectrum was not possible. Rigidification of the AGP2 on addition of the UCN-01 ligand yielded data of sufficient quality for full backbone assignments to be made. Therefore, parts of the AGP2 work may be considered to use a "reverse CSP" strategy, in that NMR can observe the apo form of the protein in a way that is not possible by crystallography, and allow inferences about the apo structure to be made based on the assigned ligand-AGP2 complex. If a future drug discovery effort were to target the apo form of AGP2 for investigation, with an antibody for instance, NMR assignments of apo spectra would be highly valued. It may be possible to obtain apo AGP2 assignments by optimising spectral conditions for fresh 3D NMR experiments used in conjunction with the existing UCN-01 bound assignments, but the success of such an approach is not guaranteed.

Future studies may also look at modifying UCN-01 to abolish affinity for AGP2 to yield a compound with more favourable pharmacokinetic properties. During such a process, the full backbone assignments offer the prospect of using HSQC spectra to triage a variety of UCN-01 analogues for AGP2 binding than by crystallography alone. The crystal structures reported to date have been solved in very different buffer conditions at pH 4.6 (Nishi *et al*, 2011) or pH 7.4 (this study). These facts indicate that crystallography may not be sufficiently reproducible to cover the broad range of ligands that NMR can handle with ease. The existence of a complete set of NMR backbone assignments for the UCN-01 bound form makes further NMR studies of new ligands relatively straight-forward, as those assignments may be transferred to recently collected spectra to analyse the binding of new ligands or redesigned UCN-01 derivatives. Future studies may, however, benefit from de-novo assignments of apo AGP2 spectra to confirm the identity of all the

peaks that shift significantly in the ligand bound form. This would allow a more traditional CSP workflow to be adopted, even if some apo residues remain unassigned.

Conclusion

In conclusion protein observed NMR is a powerful tool in drug discovery that has been applied here to advance our understanding of inactive states of the $A_{2A}R$ and the binding of UCN-01 to AGP2. The power of ^{19}F tagging has been extended to study the inhibitory action of a novel antagonist to a relatively intractable GPCR and classical isotope labelling techniques have been used to work backwards from ligand bound AGP2 to understand the apo form of that protein.

It is hoped that this work extends the use of protein observed NMR to accelerate the delivery of novel therapeutic agents currently in drug discovery pipelines creating value for patients in areas of high unmet medical need.

Appendix

Residue	H	N	C	Ca	Cb	Cg	Cd	Ha	Hb	Hd
None	None	None	None	None	None	None	None	None	None	None
0Pro	None	None	None	62.9	32.16	None	None	None	None	None
1Gln	8.54	122.02	None	55.8	29.34	None	None	4.26	1.92,1.99	None
2Ile	8.17	124.55	None	58.62	38.63	17.25,26.82	12.85	4.43	1.81	
3Pro	None	None	None	63.32	32.19	None	None	None	None	None
4Leu	8.26	122.58	None	55.95	42.3	27.31	None	4.18	1.51,1.58	
5Cys	8.34	116.36	None	55.23	39.58	None	None	4.42	2.94,3.21	None
6Ala	8.14	125.25	None	54.79	18.65	None	None	4.43	1.36	None
7Asn	8.4	113.68	None	53.87	36.92	None	None	4.58	2.92	None
8Leu	7.28	120.58	None	53.85	40.94	27.5	24.56,26.13	4.31	1.53,1.58	0.60,0.70
9Val	6.93	119.33	None	59.42	32.92	20.86,20.90	None	4.17	1.99	None
10Pro	None	None	None	63.09	32.52	None	None	None	None	None
11Val	8.7	122.7	None	56.13	29.4	None	None	None	None	None
12Pro	None	None	None	62.7	31.99	None	None	None	None	None
13Ile	9.23	125.07	None	61.77	37.98	None	None	None	None	None
14Thr	7.9	117.36	None	59.31	72.12	None	None	None	None	None
15Asn	9.27	121.08	None	56.54	36.84	None	None	4.69	2.88	None
16Ala	8.07	122.8	None	55.21	17.83	None	None	4.25	1.37	None
17Thr	7.77	119.13	None	66.66	68.01	None	None	3.71	4.43	None
18Leu	7.79	119.51	None	57.55	40.08	None	None	4.26	1.32,1.39	None
19Asp	8.23	121.41	None	58.26	39.64	None	None	4.41	2.73,2.82	None
20Arg	7.9	122.41	None	59.13	29.94	None	None	None	None	None
21Ile	7.53	107.68	None	61.84	37.28	17.25,25.52	14.14	5	2.52	
22Thr	7.54	120.68	None	65.66	69.46	None	None	4.11	4.46	None
23Gly	8.59	115.63	None	44.18	None	None	None	None	None	None
24Lys	7.61	122.3	None	54.32	34.86	None	None	4.45	1.08,1.18	None
25Trp	8.34	127.88	None	54.61	34.29	None	None	4.4	None	None
26Phe	9.35	116.47	None	56.73	41.28	None	None	4.97	None	None
27Tyr	9.03	122.43	None	57.35	40.88	None	None	5.74	None	None
28Ile	8.53	130.58	None	59.61	37.54	None	None	None	None	None
29Ala	6.97	119.45	None	51.02	22.65	None	None	5.07	0.88	None
30Ser	8.82	110.94	None	58.87	65.81	None	None	4.89	3.84,4.00	None
31Ala	7.99	118.05	None	53.19	24.01	None	None	4.49	1.21	None
32Phe	8.75	114.19	None	56.19	44.82	None	None	None	None	None
33Arg	9.11	119.7	None	57.4	30	None	None	None	None	None
34Asn	6.63	117.72	None	54.91	41.19	None	None	None	None	None
35Glu	9	130.25	None	59.95	29.51	None	None	None	None	None
36Glu	8	120.29	None	60.3	28.52	None	None	None	None	None
37Tyr	8.72	120.27	None	58.35	35.4	None	None	None	None	None
38Asn	7.86	118.06	None	57.72	38.99	None	None	4.55	None	None
39Lys	8.18	117.84	None	59.29	32.08	None	None	None	None	None
40Ser	8.09	114.66	None	61.64	63.12	None	None	4.18	3.66,3.77	None
41Val	8.6	117.72	None	65.03	31.68	22.24,22.29	None	4.02	2.43	None
42Gln	7.84	122.17	None	58.93	28.37	None	None	None	None	None
43Glu	7.08	113.86	None	55.91	30.79	None	None	4.37	None	None

44Ile	7.63	120.42	None	62.68	38.4	18.46	14.84	4.09	2.14	
45Gln	8.7	128.72	None	58.15	30.94	None	None	4.44	None	None
46Ala	8.02	116.04	None	51.97	20.5	None	None	5.08	1.68	None
47Thr	10.41	None	None	66.19	77.4	23.54	None	None	None	None
48Phe	8.81	123.78	None	55.71	42.22	None	None	5.74	None	None
49Phe	9.08	116.92	None	55.57	42.18	None	None	5.84	3.07,3.22	None
50Tyr	9.29	117.02	None	56.44	42.22	None	None	5.5	None	None
51Phe	9.2	120.36	None	None	42.61	None	None	5.96	2.88	None
52Thr	8.9	118.12	None	59.92	70.74	None	None	4.98	4.46	None
53Pro	None	None	None	63.3	33.47	None	None	None	None	None
54Asn	9.17	122.91	None	52.55	38.06	None	None	4.94	2.73,3.26	None
55Lys	8.9	122.22	None	59.96	32.41	None	None	None	None	None
56Thr	8	110.44	None	64.12	68.74	22.08	None	None	4.27	None
57Glu	7.18	118.92	None	55.53	30.84	None	None	None	None	None
58Asp	7.79	117.79	None	56.1	37.99	None	None	4.22	2.91,3.21	None
59Thr	7.68	106.93	None	58.94	74.17	None	None	5.57	4.34	None
60Ile	8.76	118.34	None	60.9	42.48	17.86	14.13	4.21	1.35	
61Phe	8.42	129.27	None	57.07	39.51	None	None	4.73	None	None
62Leu	8.98	135.38	None	53.59	44.96	None	24.67,25.31	5.49	1.32,1.73	0.49,1.16
63Arg	8.03	123.99	None	55.57	33.38	None	None	4.44	None	None
64Glu	8.87	126.62	None	56.07	34.73	None	None	4.63	None	None
65Tyr	9.03	127.66	None	57.49	36.67	None	None	4.65	None	None
66Gln	8.97	126.56	None	55.26	34.63	None	None	5.11	None	None
67Thr	8.77	125.53	None	62.81	69.96	None	None	5.52	None	None
68Arg	9	126.78	None	54.91	33.85	None	None	5.07	None	None
69Gln	9.32	120.59	None	57.07	26.33	None	None	None	None	None
70Asn	9.33	109.76	None	54.82	37.41	None	None	4.13	2.88,3.07	None
71Gln	8.04	118.3	None	54.52	32.99	None	None	None	None	None
72Cys	8.56	119.75	None	56.13	43.89	None	None	None	None	None
73Phe	10.04	132.72	None	57.97	40.71	None	None	4.53	None	None
74Tyr	8.52	126.09	None	54.94	40.32	None	None	None	None	None
75Asn	8.19	126.73	None	52.16	42.71	None	None	4.43	2.22,2.30	None
76Ser	8.04	117.39	None	55.81	65.92	None	None	5.16	3.56,3.78	None
77Ser	8.44	116.08	None	57.78	65.33	None	None	4.58	3.65,3.68	None
78Tyr	8.67	118.23	None	58.1	39.91	None	None	5.29	None	None
79Leu	9.62	125.26	None	53.29	43.18	25.98	None	5.01	1.58	
80Asn	8.4	118.88	None	53.04	40.7	None	None	None	None	None
81Val	8.02	120.75	None	61.49	33.54	None	None	None	None	None
82Gln	8.64	129.52	None	55.45	25.65	None	None	4.66	None	None
83Arg	8.73	121.99	None	61.96	34.8	None	None	None	None	None
84Glu	8.96	117.06	None	59.42	28.78	None	None	3.99	None	None
85Asn	7.32	114.9	None	52.76	39.3	None	None	4.67	2.59,2.83	None
86Gly	7.87	106.83	None	47.05	None	None	None	None	None	None
87Thr	7.5	107.68	None	58.45	73.05	None	None	None	None	None
88Val	8.47	111.64	None	58.81	35.37	None	None	5.17	None	None
89Ser	9.22	115.46	None	56.5	67.28	None	None	6.04	3.70,3.90	None

90Arg	8.7	119.59	None	53.81	34.79	25.17	None	None	None	None
91Tyr	8.75	126.2	None	57.1	38.94	None	None	5.02	None	None
92Glu	8.65	127.1	None	55.64	32.19	35.2	None	4.3	1.35	None
93Gly	8.9	116.58	None	46.94	None	None	None	None	None	None
94Gly	8.29	106.15	None	45.11	None	None	None	None	None	None
95Arg	7.25	118.19	None	54.12	32.51	None	None	None	None	None
96Glu	8.36	122.33	None	55.55	31.09	None	None	None	None	None
97His	9.22	127.41	None	55.28	34.4	None	None	4.8	2.62,3.02	None
98Val	8.69	122.25	None	61.42	34.89	20.92,21.16	None	5.09	1.79	None
99Ala	9.38	133.2	None	50.43	20.92	None	None	5.07	None	None
100His	9.05	119.67	None	58.27	31.93	None	None	4.52	None	None
101Leu	7.56	124.34	None	54.91	42.36	None	None	None	None	None
102Leu	8.52	128.14	None	53.41	45.59	None	None	4.59	1.36	None
103Phe	8.29	117.47	None	57.11	41.97	None	None	4.79	None	None
104Leu	9.07	124.74	None	None	None	None	None	None	None	None
105Arg	9.55	123.04	None	59.3	29.47	None	None	None	None	None
106Asp	8.58	121.16	None	54.77	42.85	None	None	None	None	None
107Thr	8.2	115.85	None	63.49	69.01	None	None	4.35	4.55	None
108Lys	8.94	124.28	None	55.75	31.23	None	None	4.58	None	None
109Thr	7.2	105.61	None	59.49	72.71	None	None	5.28	4.25	None
110Leu	8.47	122.04	None	55.05	45.74	None	None	4.26	0.58,0.76	None
111Met	8.12	120.8	None	53.78	38.48	None	None	5.61	1.93	None
112Phe	8.44	119.19	None	52.21	39.16	None	None	5.66	2.28,2.82	None
113Gly	9.88	111.17	None	45.25	None	None	None	None	None	None
114Ser	9.12	123.96	None	59.11	65.59	None	None	4.84	None	None
115Tyr	9.46	121.46	None	57.83	36.22	None	None	4.42	None	None
116Leu	None	None	None	58.14	42.08	None	None	None	None	None
117Asp	8.61	113.65	None	53.44	40.47	None	None	4.67	2.64,2.87	None
118Asp	8.42	121.64	None	52.4	42.71	None	None	4.79	2.42,3.02	None
119Glu	8.46	126.66	None	56.92	28.93	None	None	None	None	None
120Lys	7.68	115.24	None	56.82	32.44	None	None	None	None	None
121Asn	7.48	114.99	None	54.33	45.02	None	None	None	None	None
122Trp	7.7	120.52	None	55.88	30.55	None	None	None	None	None
123Gly	8.65	104.73	None	47.42	None	None	None	None	None	None
124Leu	7.99	116.81	None	53.11	49.62	None	None	5.57	1.43,1.49	None
125Ser	8.54	111.9	None	56.26	65.66	None	None	None	3.11,3.61	None
126Phe	8.43	124.71	None	55.72	42.61	None	None	4.8	None	None
127Tyr	8.29	126.52	None	55.88	42.31	None	None	5.15	None	None
128Ala	8.81	120.04	50.49	50.46	23.45	None	None	5.47	1.37	None
129Asp	8.36	120.31	None	54.46	40.26	None	None	4.17	2.19,2.65	None
130Lys	7.13	119.17	None	53.06	34.81	None	None	None	None	None
131Pro	None	None	None	62.76	30.1	None	None	None	None	None
132Glu	6.54	115.9	None	54.71	31.8	None	None	4.39	None	None
133Thr	8.21	113.02	None	59.64	71.62	None	None	4.57	None	None
134Thr	8.59	112.37	None	60.21	71.48	None	None	4.56	4.34	None
135Lys	8.99	121.99	None	59.54	31.72	None	None	None	None	None

136Glu	8.76	120.96	None	60.35	28.59	None	None	None	None	None
137Gln	7.79	120.41	None	58.84	29.67	None	None	None	None	None
138Leu	7.89	119.71	None	57	41.55	24.74	24.12	3.69	1.29,1.52	
139Gly	8.12	105.87	None	47.41	None	None	None	None	None	None
140Glu	7.32	120.1	None	59.92	30.49	None	None	None	None	None
141Phe	6.82	121.89	None	59.47	36.85	None	None	None	None	None
142Tyr	8.13	117.89	None	57.98	35.84	None	None	None	None	None
143Glu	7.72	118.83	59.5	59.48	29.7	None	None	None	None	None
144Ala	7.18	123.36	None	55.21	18.22	None	None	3.87	1.2	None
145Leu	7.69	118.24	None	57.79	40.66	None	None	None	None	None
146Asp	8.02	119.33	None	57	39.4	None	None	4.25	2.52,2.83	None
147Cys	7.58	120.4	None	60.16	42.45	None	None	None	None	None
148Leu	6.94	118.47	None	54.33	40.9	None	19.91,25.33	3.77	0.75,1.15	-0.67,0.07
149Arg	7.69	114.17	None	56.96	26.59	None	None	None	None	None
150Ile	7.47	122.06	None	58.55	39.87	16.54	14.14	4.52	1.31	
151Pro	None	None	None	62.55	32.47	None	None	None	None	None
152Arg	8.38	122.13	None	58.28	29.31	None	None	None	None	None
153Ser	7.87	111.78	None	59.75	62.59	None	None	None	None	None
154Asp	7.66	123.29	None	54.79	40.08	None	None	4.8	2.59,2.93	None
155Val	7.06	120.85	None	63.18	32.29	23.49,25.81	None	4	1.92	None
156Met	9.09	128.4	None	54.36	34.11	None	None	None	None	None
157Tyr	8.56	118.41	None	57	41.7	None	None	None	None	None
158Thr	7.74	113.15	None	60.94	70.78	None	None	4.32	4.4	None
159Asp	8.55	122.37	None	52.67	41.77	None	None	5.05	None	None
160Trp	9.2	126.4	None	59.1	30.32	None	None	4.62	None	None
161Lys	8.09	116.77	None	58.04	30.93	None	None	None	None	None
162Lys	8.02	117.31	None	55.06	33	None	None	None	None	None
163Asp	7.08	118.27	None	56.93	42.54	None	None	4.03	None	None
164Lys	8.55	123.85	None	53.37	34.39	None	None	None	None	None
165Cys	9.26	117.85	None	54.34	43.73	None	None	None	None	None
166Glu	11.04	129.36	None	62.56	34.49	None	None	None	None	None
167Pro	None	None	None	65.28	31.14	None	None	None	None	None
168Leu	7.45	117.53	None	56.22	41.91	None	None	None	None	None
169Glu	7.75	120.21	None	58.45	29.48	None	None	None	None	None
170Lys	7.63	118.72	None	56.56	32.58	None	None	None	None	None
171Gln	7.85	120.49	None	56	29.12	None	None	None	None	None
172His	7.78	125.53	None	56.82	30.67	None	None	None	None	None
-1Gly	None	None	None	None	None	None	None	None	None	None

a.1 Residue assignments of the AGP2 at pH 6.5

Number	Assign F1	Assign F2	Time Constant	TC Error	Fit Error	Num Peaks	Function
1	1GlnN	1GlnH	859.12198	31.82894	0.39313	16	A exp(-Bx)
2	2IleN	2IleH	763.47149	17.00943	0.59652	16	A exp(-Bx)
3	4LeuN	4LeuH	804.17299	25.66784	0.3263	8	A exp(-Bx)
4	5CysN	5CysH	783.05041	69.08604	0.31496	8	A exp(-Bx)
5	6AlaN	6AlaH	883.89663	33.39728	0.22308	16	A exp(-Bx)
6	7AsnN	7AsnH	923.57311	26.12459	0.33379	16	A exp(-Bx)
7	8LeuN	8LeuH	1104.33927	21.12312	0.20719	16	A exp(-Bx)
8	9ValN	9ValH	1043.75589	15.31983	0.21214	16	A exp(-Bx)
9	14ThrN	14ThrH	1056.33767	19.99133	0.12588	16	A exp(-Bx)
10	16AlaN	16AlaH	966.65201	106.5865	0.17738	8	A exp(-Bx)
11	17ThrN	17ThrH	1001.85123	20.81529	0.32462	16	A exp(-Bx)
12	18LeuN	18LeuH	1055.86121	25.61128	0.20761	16	A exp(-Bx)
13	19AspN	19AspH	1057.47124	29.45026	0.4095	16	A exp(-Bx)
14	20ArgN	20ArgH	1047.05207	14.61138	0.23155	16	A exp(-Bx)
15	22ThrN	22ThrH	1047.3358	15.9004	0.24947	16	A exp(-Bx)
16	23GlyN	23GlyH	1097.17985	23.35668	0.20054	16	A exp(-Bx)
17	24LysN	24LysH	1112.09909	14.91411	0.23427	16	A exp(-Bx)
18	25TrpN	25TrpH	1078.23693	27.53289	0.26535	16	A exp(-Bx)
19	26PheN	26PheH	1116.95195	26.75038	0.19256	16	A exp(-Bx)
20	27TyrN	27TyrH	1064.34591	31.39551	0.25269	15	A exp(-Bx)
21	28IleN	28IleH	1061.89151	32.27857	0.14496	15	A exp(-Bx)
22	29AlaN	29AlaH	1111.36457	15.32199	0.15997	16	A exp(-Bx)
23	30SerN	30SerH	1072.1098	14.26158	0.17842	16	A exp(-Bx)
24	31AlaN	31AlaH	1053.2487	33.92325	0.45326	16	A exp(-Bx)
25	32PheN	32PheH	1174.15922	29.14129	0.28207	16	A exp(-Bx)
26	33ArgN	33ArgH	1058.50824	24.20456	0.18386	16	A exp(-Bx)
27	35GluN	35GluH	999.91044	27.03813	0.39548	16	A exp(-Bx)
28	36GluN	36GluH	1152.31904	18.21528	0.32368	16	A exp(-Bx)
29	37TyrN	37TyrH	1118.92798	23.65899	0.2702	16	A exp(-Bx)
30	38AsnN	38AsnH	1191.48769	19.29484	0.2365	16	A exp(-Bx)
31	39LysN	39LysH	1161.42742	12.76348	0.17221	16	A exp(-Bx)
32	40SerN	40SerH	1154.40609	13.55528	0.17136	16	A exp(-Bx)
33	41ValN	41ValH	1107.80131	32.90179	0.29341	16	A exp(-Bx)
34	42GlnN	42GlnH	1022.20323	10.6005	0.17104	16	A exp(-Bx)
35	43GluN	43GluH	1071.89154	17.74005	0.28745	16	A exp(-Bx)
36	44IleN	44IleH	1126.21677	31.79105	0.49201	16	A exp(-Bx)
37	45GlnN	45GlnH	1213.78226	31.02609	0.14172	16	A exp(-Bx)
38	46AlaN	46AlaH	1096.33811	14.17035	0.19344	16	A exp(-Bx)
39	48PheN	48PheH	994.78891	24.24085	0.30218	16	A exp(-Bx)
40	49PheN	49PheH	974.92523	28.32263	0.32697	16	A exp(-Bx)
41	50TyrN	50TyrH	1032.84423	20.26942	0.1415	16	A exp(-Bx)
42	51PheN	51PheH	1089.11654	21.14562	0.17465	16	A exp(-Bx)
43	52ThrN	52ThrH	1081.83033	47.85542	0.42083	16	A exp(-Bx)
44	54AsnN	54AsnH	1072.54547	17.66621	0.23487	16	A exp(-Bx)

45	55LysN	55LysH	925.94503	41.46333	0.21332	16	A exp(-Bx)
46	56ThrN	56ThrH	1022.08258	31.60084	0.39363	16	A exp(-Bx)
47	57GluN	57GluH	995.33632	17.01003	0.27293	16	A exp(-Bx)
48	58AspN	58AspH	953.68758	25.60208	0.40417	16	A exp(-Bx)
49	59ThrN	59ThrH	958.24781	35.5037	0.59225	16	A exp(-Bx)
50	60IleN	60IleH	1036.44421	23.68901	0.24234	16	A exp(-Bx)
51	61PheN	61PheH	1018.84481	27.26469	0.30862	16	A exp(-Bx)
52	62LeuN	62LeuH	965.61991	28.27951	0.26035	16	A exp(-Bx)
53	63ArgN	63ArgH	1020.78527	40.61561	0.36182	16	A exp(-Bx)
54	64GluN	64GluH	1048.67675	27.351	0.33645	16	A exp(-Bx)
55	65TyrN	65TyrH	1045.01724	23.82829	0.19959	16	A exp(-Bx)
56	67ThrN	67ThrH	1144.7636	30.89965	0.19613	16	A exp(-Bx)
57	68ArgN	68ArgH	1072.95541	26.79116	0.6124	16	A exp(-Bx)
58	69GlnN	69GlnH	1118.40614	15.51881	0.1939	16	A exp(-Bx)
59	70AsnN	70AsnH	1118.98818	12.89404	0.17767	16	A exp(-Bx)
60	71GlnN	71GlnH	1124.4119	17.60843	0.34087	16	A exp(-Bx)
61	72CysN	72CysH	1185.1908	17.1268	0.26019	16	A exp(-Bx)
62	73PheN	73PheH	1127.53563	24.93272	0.25168	16	A exp(-Bx)
63	74TyrN	74TyrH	1128.40071	25.07425	0.25124	16	A exp(-Bx)
64	75AsnN	75AsnH	957.19368	64.28379	0.98591	16	A exp(-Bx)
65	77SerN	77SerH	972.88987	34.20306	0.44175	16	A exp(-Bx)
66	78TyrN	78TyrH	1010.99481	12.11389	0.15444	16	A exp(-Bx)
67	79LeuN	79LeuH	1070.65212	25.74016	0.17941	16	A exp(-Bx)
68	80AsnN	80AsnH	1090.70155	38.60114	0.28337	13	A exp(-Bx)
69	81ValN	81ValH	1109.511	34.20512	0.30113	16	A exp(-Bx)
70	82GlnN	82GlnH	1130.29508	23.99932	0.2246	16	A exp(-Bx)
71	83ArgN	83ArgH	1129.26169	25.28849	0.21639	16	A exp(-Bx)
72	84GluN	84GluH	1020.54878	25.28199	0.31065	16	A exp(-Bx)
73	85AsnN	85AsnH	1110.35487	25.862	0.29959	16	A exp(-Bx)
74	86GlyN	86GlyH	1066.74478	35.79657	0.34402	16	A exp(-Bx)
75	88ValN	88ValH	1051.94764	23.11589	0.18612	16	A exp(-Bx)
76	89SerN	89SerH	1016.39323	28.92422	0.27166	16	A exp(-Bx)
77	90ArgN	90ArgH	1093.04117	44.41905	0.32711	16	A exp(-Bx)
78	91TyrN	91TyrH	1025.279	33.61423	0.40649	16	A exp(-Bx)
79	92GluN	92GluH	1006.4884	30.97041	0.36581	15	A exp(-Bx)
80	95ArgN	95ArgH	881.09439	21.797	0.50583	16	A exp(-Bx)
81	96GluN	96GluH	1035.15257	22.69606	0.37396	16	A exp(-Bx)
82	97HisN	97HisH	1010.59314	35.51851	0.33566	16	A exp(-Bx)
83	98ValN	98ValH	1045.85387	40.48867	0.46938	16	A exp(-Bx)
84	99AlaN	99AlaH	1040.56434	17.40843	0.16748	16	A exp(-Bx)
85	100HisN	100HisH	1130.7395	22.4667	0.16754	16	A exp(-Bx)
86	101LeuN	101LeuH	1049.6963	64.3161	0.3488	15	A exp(-Bx)
87	102LeuN	102LeuH	1122.07749	27.92685	0.15901	16	A exp(-Bx)
88	103PheN	103PheH	1120.33641	52.3439	0.35521	16	A exp(-Bx)
89	104LeuN	104LeuH	1301.47507	36.75987	0.28535	16	A exp(-Bx)
90	105ArgN	105ArgH	1067.638	58.87993	0.0961	11	A exp(-Bx)

91	106AspN	106AspH	954.67309	49.40419	0.2583	8	A exp(-Bx)
92	107ThrN	107ThrH	1030.83481	89.35354	0.31829	16	A exp(-Bx)
93	108LysN	108LysH	994.82462	23.91572	0.21185	16	A exp(-Bx)
94	109ThrN	109ThrH	1036.84623	34.78779	0.33301	16	A exp(-Bx)
95	110LeuN	110LeuH	998.94332	36.86989	0.30939	16	A exp(-Bx)
96	111MetN	111MetH	1016.82516	41.63924	0.33697	16	A exp(-Bx)
97	112PheN	112PheH	1195.09556	24.7324	0.13041	16	A exp(-Bx)
98	113GlyN	113GlyH	1123.23397	23.17653	0.1409	16	A exp(-Bx)
99	114SerN	114SerH	1006.80018	35.15311	0.26671	16	A exp(-Bx)
100	115TyrN	115TyrH	1051.89566	31.03686	0.27502	16	A exp(-Bx)
101	117AspN	117AspH	1179.24966	26.14897	0.27365	16	A exp(-Bx)
102	118AspN	118AspH	1106.20222	19.95265	0.26287	16	A exp(-Bx)
103	119GluN	119GluH	1121.68627	23.272	0.31581	16	A exp(-Bx)
104	120LysN	120LysH	1052.65543	12.13698	0.16368	16	A exp(-Bx)
105	121AsnN	121AsnH	1164.16965	34.7817	0.24961	16	A exp(-Bx)
106	122TrpN	122TrpH	1132.76041	21.59754	0.1962	16	A exp(-Bx)
107	123GlyN	123GlyH	1018.42275	24.79435	0.24535	16	A exp(-Bx)
108	124LeuN	124LeuH	1094.19548	20.0297	0.18695	16	A exp(-Bx)
109	125SerN	125SerH	1031.09403	38.75211	0.28291	16	A exp(-Bx)
110	126PheN	126PheH	1021.54411	39.65845	0.33438	16	A exp(-Bx)
111	127TyrN	127TyrH	1147.45766	24.87868	0.14742	16	A exp(-Bx)
112	128AlaN	128AlaH	1098.62612	36.18784	0.34955	16	A exp(-Bx)
113	129AspN	129AspH	1076.28279	40.95373	0.38415	16	A exp(-Bx)
114	130LysN	130LysH	1130.39317	23.11175	0.42162	16	A exp(-Bx)
115	132GluN	132GluH	1141.60185	31.06778	0.51027	16	A exp(-Bx)
116	133ThrN	133ThrH	1106.01891	18.47247	0.29113	16	A exp(-Bx)
117	134ThrN	134ThrH	1027.89937	15.16915	0.20166	16	A exp(-Bx)
118	135LysN	135LysH	956.56926	43.09572	0.17018	15	A exp(-Bx)
119	136GluN	136GluH	893.43558	30.23463	0.42927	16	A exp(-Bx)
120	137GlnN	137GlnH	1034.00314	29.5011	0.35145	16	A exp(-Bx)
121	138LeuN	138LeuH	1031.26833	26.2903	0.3032	16	A exp(-Bx)
122	139GlyN	139GlyH	1021.76842	14.8512	0.20998	16	A exp(-Bx)
123	140GluN	140GluH	1033.92405	19.62802	0.19296	16	A exp(-Bx)
124	141PheN	141PheH	1080.33327	29.26305	0.28023	16	A exp(-Bx)
125	142TyrN	142TyrH	1139.01394	30.06257	0.29476	16	A exp(-Bx)
126	143GluN	143GluH	1128.64111	43.72245	0.56257	16	A exp(-Bx)
127	144AlaN	144AlaH	1197.44918	36.34605	0.36762	16	A exp(-Bx)
128	145LeuN	145LeuH	1097.70535	17.28938	0.15399	16	A exp(-Bx)
129	146AspN	146AspH	1040.21467	32.80543	0.43593	16	A exp(-Bx)
130	147CysN	147CysH	1176.55292	13.50453	0.19165	16	A exp(-Bx)
131	148LeuN	148LeuH	1219.53078	59.63689	0.42864	16	A exp(-Bx)
132	149ArgN	149ArgH	997.24584	38.57424	0.54075	16	A exp(-Bx)
133	150IleN	150IleH	1107.79789	31.42502	0.36583	16	A exp(-Bx)
134	152ArgN	152ArgH	1032.47434	19.77193	0.27887	16	A exp(-Bx)
135	153SerN	153SerH	902.26539	37.61458	0.5576	16	A exp(-Bx)
136	154AspN	154AspH	1084.51615	15.08616	0.20493	16	A exp(-Bx)

137	155ValN	155ValH	1100.40167	12.58044	0.21836	16	A exp(-Bx)
138	156MetN	156MetH	1102.124	26.94203	0.18325	16	A exp(-Bx)
139	157TyrN	157TyrH	1204.84813	26.62456	0.27299	16	A exp(-Bx)
140	158ThrN	158ThrH	1226.54766	25.81913	0.25353	16	A exp(-Bx)
141	159AspN	159AspH	1062.65882	20.85341	0.20344	16	A exp(-Bx)
142	160TrpN	160TrpH	1146.88491	17.14313	0.15476	16	A exp(-Bx)
143	161LysN	161LysH	1103.35831	27.75991	0.39189	16	A exp(-Bx)
144	163AspN	163AspH	1130.77888	13.36512	0.25731	16	A exp(-Bx)
145	164LysN	164LysH	1095.04573	67.85179	0.36441	16	A exp(-Bx)
146	165CysN	165CysH	1081.39635	21.11555	0.17952	16	A exp(-Bx)
147	166GluN	166GluH	1028.61131	24.24302	0.13111	16	A exp(-Bx)
148	168LeuN	168LeuH	1111.32374	32.83493	0.46251	16	A exp(-Bx)
149	169GluN	169GluH	947.11388	16.0922	0.34129	16	A exp(-Bx)
150	170LysN	170LysH	721.67647	16.46818	0.47189	16	A exp(-Bx)
151	171GlnN	171GlnH	675.39407	31.07253	0.70257	16	A exp(-Bx)
152	172HisN	172HisH	744.50401	38.16403	2.21833	16	A exp(-Bx)

a.2 T1 values of each residue measured at pH 6.5, 750 μM AGP2, three-fold excess of UCN-01 at 700 MHz.

Number	Assign F1	Assign F2	Time Constant	TC Error	Fit Error	Num Peaks	Function
1	1GlnN	1GlnH	261.6321	21.50787	1.00611	14	A exp(-Bx)
2	2IleN	2IleH	178.5108	3.4053	0.56335	14	A exp(-Bx)
3	4LeuN	4LeuH	104.1699	2.26854	0.3639	13	A exp(-Bx)
4	5CysN	5CysH	73.66543	3.55034	0.25994	10	A exp(-Bx)
5	6AlaN	6AlaH	63.14315	2.50493	0.25626	10	A exp(-Bx)
6	7AsnN	7AsnH	65.2203	1.92897	0.33976	11	A exp(-Bx)
7	8LeuN	8LeuH	57.2266	1.66629	0.28939	11	A exp(-Bx)
8	9ValN	9ValH	70.96355	1.74339	0.34034	11	A exp(-Bx)
9	11ValN	11ValH	60.55606	2.71931	0.33785	12	A exp(-Bx)
10	14ThrN	14ThrH	62.29549	2.14818	0.24332	11	A exp(-Bx)
11	17ThrN	17ThrH	55.47479	1.0951	0.28874	11	A exp(-Bx)
12	18LeuN	18LeuH	58.39489	1.47012	0.19802	9	A exp(-Bx)
13	19AspN	19AspH	55.63047	1.01928	0.25639	12	A exp(-Bx)
14	20ArgN	20ArgH	56.80683	1.18503	0.3244	11	A exp(-Bx)
15	22ThrN	22ThrH	57.26078	1.37888	0.35325	11	A exp(-Bx)
16	23GlyN	23GlyH	65.91523	1.7173	0.26781	12	A exp(-Bx)
17	24LysN	24LysH	65.44115	1.33348	0.34258	11	A exp(-Bx)
18	25TrpN	25TrpH	58.50076	1.22762	0.22725	11	A exp(-Bx)
19	26PheN	26PheH	62.53853	1.81054	0.24723	11	A exp(-Bx)
20	27TyrN	27TyrH	57.49354	2.28684	0.30316	9	A exp(-Bx)
21	28IleN	28IleH	54.75278	1.15633	0.09777	8	A exp(-Bx)
22	29AlaN	29AlaH	59.07761	1.93215	0.35695	11	A exp(-Bx)
23	30SerN	30SerH	58.34967	1.23573	0.26765	11	A exp(-Bx)

24	31AlaN	31AlaH	60.60607	0.80777	0.16779	11	A exp(-Bx)
25	32PheN	32PheH	70.12457	1.78245	0.28602	11	A exp(-Bx)
26	33ArgN	33ArgH	59.68319	1.52194	0.20408	10	A exp(-Bx)
27	34AsnN	34AsnH	118.5667	2.17868	0.31935	14	A exp(-Bx)
28	35GluN	35GluH	57.51344	1.34516	0.33372	11	A exp(-Bx)
29	36GluN	36GluH	57.12051	1.36378	0.49561	12	A exp(-Bx)
30	37TyrN	37TyrH	54.93617	0.66759	0.15578	11	A exp(-Bx)
31	38AsnN	38AsnH	57.40002	1.07202	0.25802	11	A exp(-Bx)
32	39LysN	39LysH	60.10143	0.97875	0.25199	11	A exp(-Bx)
33	40SerN	40SerH	63.52428	1.6949	0.36828	11	A exp(-Bx)
34	41ValN	41ValH	55.1858	1.16052	0.21263	10	A exp(-Bx)
35	42GlnN	42GlnH	59.53113	0.87857	0.23907	11	A exp(-Bx)
36	43GluN	43GluH	69.47868	2.20781	0.48966	11	A exp(-Bx)
37	44IleN	44IleH	66.19206	3.05331	0.73651	12	A exp(-Bx)
38	45GlnN	45GlnH	58.08748	1.29298	0.1286	10	A exp(-Bx)
39	46AlaN	46AlaH	63.21534	1.62971	0.35465	12	A exp(-Bx)
40	48PheN	48PheH	58.57069	1.39255	0.26577	12	A exp(-Bx)
41	49PheN	49PheH	56.60452	1.66732	0.31799	10	A exp(-Bx)
42	50TyrN	50TyrH	61.89362	2.2563	0.27137	11	A exp(-Bx)
43	51PheN	51PheH	61.11202	1.14254	0.17404	10	A exp(-Bx)
44	52ThrN	52ThrH	66.31957	1.50353	0.21374	11	A exp(-Bx)
45	54AsnN	54AsnH	62.73361	1.40851	0.30556	11	A exp(-Bx)
46	55LysN	55LysH	58.96963	2.48159	0.21887	10	A exp(-Bx)
47	56ThrN	56ThrH	59.77342	1.43004	0.31271	11	A exp(-Bx)
48	57GluN	57GluH	64.81675	1.44575	0.32867	11	A exp(-Bx)
49	58AspN	58AspH	62.31768	1.55516	0.37038	11	A exp(-Bx)
50	59ThrN	59ThrH	61.50757	1.96034	0.45391	12	A exp(-Bx)
51	60IleN	60IleH	61.79601	1.1806	0.20058	11	A exp(-Bx)
52	61PheN	61PheH	59.0928	1.37577	0.25624	10	A exp(-Bx)
53	62LeuN	62LeuH	63.97755	2.09172	0.27171	12	A exp(-Bx)
54	63ArgN	63ArgH	62.40883	2.12123	0.28472	11	A exp(-Bx)
55	64GluN	64GluH	67.27715	2.80405	0.49282	11	A exp(-Bx)
56	65TyrN	65TyrH	56.09401	1.53951	0.2387	11	A exp(-Bx)
57	67ThrN	67ThrH	66.71972	1.96839	0.21526	11	A exp(-Bx)
58	68ArgN	68ArgH	61.66375	0.83649	0.31468	11	A exp(-Bx)
59	69GlnN	69GlnH	57.72985	1.96162	0.4737	11	A exp(-Bx)
60	70AsnN	70AsnH	54.00573	1.49082	0.40783	10	A exp(-Bx)
61	71GlnN	71GlnH	56.92716	1.26706	0.42836	11	A exp(-Bx)
62	72CysN	72CysH	58.53246	0.98236	0.29251	11	A exp(-Bx)
63	73PheN	73PheH	64.72539	2.22867	0.34983	11	A exp(-Bx)
64	74TyrN	74TyrH	66.59218	1.73451	0.28638	12	A exp(-Bx)
65	75AsnN	75AsnH	65.90558	2.58412	0.51014	12	A exp(-Bx)
66	77SerN	77SerH	65.19692	1.00482	0.19125	11	A exp(-Bx)
67	78TyrN	78TyrH	66.05402	0.4665	0.19009	12	A exp(-Bx)
68	79LeuN	79LeuH	67.49571	1.83949	0.21423	11	A exp(-Bx)
69	80AsnN	80AsnH	62.64291	1.26427	0.1658	10	A exp(-Bx)

70	81ValN	81ValH	60.0555	1.66554	0.25981	11	A exp(-Bx)
71	82GlnN	82GlnH	54.65678	1.47357	0.30779	11	A exp(-Bx)
72	83ArgN	83ArgH	58.79123	1.54841	0.25919	11	A exp(-Bx)
73	84GluN	84GluH	62.10538	2.08255	0.43984	11	A exp(-Bx)
74	85AsnN	85AsnH	66.04556	1.5914	0.31192	11	A exp(-Bx)
75	86GlyN	86GlyH	61.62388	1.69439	0.28377	12	A exp(-Bx)
76	88ValN	88ValH	57.63888	1.97294	0.27954	10	A exp(-Bx)
77	89SerN	89SerH	63.35658	1.45496	0.21961	12	A exp(-Bx)
78	90ArgN	90ArgH	54.01744	1.33247	0.20761	10	A exp(-Bx)
79	91TyrN	91TyrH	65.16303	1.15584	0.20493	12	A exp(-Bx)
80	92GluN	92GluH	59.00442	1.26824	0.24706	11	A exp(-Bx)
81	94GlyN	94GlyH	67.98614	2.62402	0.24887	12	A exp(-Bx)
82	95ArgN	95ArgH	61.56223	0.69537	0.20017	12	A exp(-Bx)
83	96GluN	96GluH	68.29342	1.49434	0.36126	11	A exp(-Bx)
84	97HisN	97HisH	64.32766	1.24226	0.16645	10	A exp(-Bx)
85	98ValN	98ValH	65.57148	1.68004	0.29699	12	A exp(-Bx)
86	99AlaN	99AlaH	60.61592	1.77218	0.26854	12	A exp(-Bx)
87	100HisN	100HisH	57.98147	1.43763	0.21293	10	A exp(-Bx)
88	101LeuN	101LeuH	53.9086	1.69919	0.17024	9	A exp(-Bx)
89	102LeuN	102LeuH	51.17312	2.16173	0.25192	10	A exp(-Bx)
90	103PheN	103PheH	53.0026	2.22747	0.24261	7	A exp(-Bx)
91	104LeuN	104LeuH	83.37568	3.16827	0.32936	11	A exp(-Bx)
92	106AspN	106AspH	61.25367	1.81085	0.2465	11	A exp(-Bx)
93	107ThrN	107ThrH	41.0751	1.67572	0.1304	6	A exp(-Bx)
94	108LysN	108LysH	60.25907	1.24086	0.18274	11	A exp(-Bx)
95	109ThrN	109ThrH	42.88128	1.66847	0.2844	9	A exp(-Bx)
97	111MetN	111MetH	58.6915	2.28615	0.29346	10	A exp(-Bx)
98	112PheN	112PheH	59.60741	1.48153	0.16493	10	A exp(-Bx)
99	113GlyN	113GlyH	55.62553	1.30337	0.17389	11	A exp(-Bx)
100	114SerN	114SerH	58.64507	2.33955	0.30175	10	A exp(-Bx)
101	115TyrN	115TyrH	60.14963	1.18839	0.1821	11	A exp(-Bx)
102	117AspN	117AspH	60.70734	1.85037	0.35075	12	A exp(-Bx)
103	118AspN	118AspH	58.27062	0.88009	0.21027	11	A exp(-Bx)
104	119GluN	119GluH	67.62772	3.03509	0.63535	11	A exp(-Bx)
105	120LysN	120LysH	65.76314	1.65501	0.33027	11	A exp(-Bx)
106	121AsnN	121AsnH	69.42959	2.19221	0.30399	11	A exp(-Bx)
107	122TrpN	122TrpH	56.21889	4.67836	0.8415	11	A exp(-Bx)
108	123GlyN	123GlyH	62.98822	2.24381	0.31859	10	A exp(-Bx)
109	124LeuN	124LeuH	59.53938	0.90992	0.15485	10	A exp(-Bx)
110	125SerN	125SerH	61.74846	2.07671	0.25343	10	A exp(-Bx)
111	126PheN	126PheH	57.59295	1.35035	0.16907	8	A exp(-Bx)
112	127TyrN	127TyrH	55.98289	1.40084	0.16442	11	A exp(-Bx)
113	128AlaN	128AlaH	55.89388	1.1422	0.21675	12	A exp(-Bx)
114	129AspN	129AspH	57.65147	1.75738	0.28659	10	A exp(-Bx)
115	130LysN	130LysH	54.17866	0.88079	0.3078	11	A exp(-Bx)
116	133ThrN	133ThrH	78.93819	2.26149	0.48403	12	A exp(-Bx)

117	134ThrN	134ThrH	67.78292	1.24744	0.2512	12	A exp(-Bx)
118	135LysN	135LysH	59.07886	2.93105	0.22159	9	A exp(-Bx)
119	136GluN	136GluH	59.32664	1.29881	0.2724	11	A exp(-Bx)
120	137GlnN	137GlnH	59.95629	1.43147	0.30078	11	A exp(-Bx)
121	138LeuN	138LeuH	58.3088	1.21082	0.25015	11	A exp(-Bx)
122	139GlyN	139GlyH	61.2646	1.74132	0.35533	10	A exp(-Bx)
123	140GluN	140GluH	55.13054	0.69311	0.11854	10	A exp(-Bx)
124	141PheN	141PheH	53.25469	1.35938	0.26544	10	A exp(-Bx)
125	142TyrN	142TyrH	58.80486	1.02502	0.19838	10	A exp(-Bx)
126	143GluN	143GluH	57.8807	1.04951	0.24295	11	A exp(-Bx)
127	144AlaN	144AlaH	58.32467	1.55333	0.31553	11	A exp(-Bx)
128	145LeuN	145LeuH	57.40867	1.11194	0.18177	10	A exp(-Bx)
129	146AspN	146AspH	57.16081	1.10779	0.26467	12	A exp(-Bx)
130	147CysN	147CysH	59.08766	2.56941	0.67056	12	A exp(-Bx)
131	148LeuN	148LeuH	58.45939	1.10285	0.15701	10	A exp(-Bx)
132	149ArgN	149ArgH	66.20049	1.30374	0.26538	12	A exp(-Bx)
133	150IleN	150IleH	57.6814	0.70244	0.15236	11	A exp(-Bx)
134	152ArgN	152ArgH	58.00401	1.34901	0.34863	11	A exp(-Bx)
135	153SerN	153SerH	53.26337	1.33233	0.3244	11	A exp(-Bx)
136	154AspN	154AspH	59.58501	1.57523	0.35299	11	A exp(-Bx)
137	155ValN	155ValH	59.20227	1.47689	0.46305	11	A exp(-Bx)
138	156MetN	156MetH	58.77794	1.80498	0.22837	11	A exp(-Bx)
139	157TyrN	157TyrH	64.91125	0.95765	0.17194	11	A exp(-Bx)
140	158ThrN	158ThrH	67.92969	1.26092	0.21472	11	A exp(-Bx)
142	160TrpN	160TrpH	62.84437	1.82316	0.30507	12	A exp(-Bx)
143	161LysN	161LysH	50.82732	1.20096	0.32893	10	A exp(-Bx)
144	163AspN	163AspH	58.0513	0.79753	0.2874	11	A exp(-Bx)
145	164LysN	164LysH	62.61586	2.15714	0.19016	11	A exp(-Bx)
146	165CysN	165CysH	55.09495	1.54578	0.25741	11	A exp(-Bx)
147	166GluN	166GluH	53.51348	2.95385	0.29264	8	A exp(-Bx)
148	168LeuN	168LeuH	54.00145	1.51397	0.40327	10	A exp(-Bx)
149	169GluN	169GluH	65.92344	1.52845	0.4569	11	A exp(-Bx)
150	170LysN	170LysH	75.00925	1.1832	0.34117	13	A exp(-Bx)
151	171GlnN	171GlnH	110.9374	3.17931	0.52198	14	A exp(-Bx)
152	172HisN	172HisH	176.9233	16.33071	4.02174	14	A exp(-Bx)

a.3 T2 values of each residue measured at pH 6.5, 750 μM AGP2, three-fold excess of UCN-01 at 700 MHz.

Number	Sat Peak	Ref Peak	1H shift A	1H shift B	15N,H shift A	15N,H shift B	Intensity Ratio
1	1GlnN,H	1GlnN,H	8.52865	8.52865	122.00439	122.00439	-0.65492
2	2IleN,H	2IleN,H	8.15713	8.15896	124.60556	124.66503	0.06583
3	4LeuN,H	4LeuN,H	8.26057	8.26195	122.6946	122.7095	0.46299
4	5CysN,H	5CysN,H	8.33075	8.32975	116.34572	116.36701	0.63905
5	6AlaN,H	6AlaN,H	8.11899	8.11864	125.23878	125.2361	0.69814
6	7AsnN,H	7AsnN,H	8.40318	8.40202	113.80739	113.79387	0.73142
7	8LeuN,H	8LeuN,H	7.28133	7.27923	120.74726	120.77003	0.80384
8	9ValN,H	9ValN,H	6.91578	6.91273	119.34428	119.32208	0.73373
9	11ValN,H	11ValN,H	8.6883	8.68759	122.87769	122.85667	0.60114
10	14ThrN,H	14ThrN,H	7.88395	7.88403	117.61832	117.62439	0.92696
11	17ThrN,H	17ThrN,H	7.75937	7.75899	119.16567	119.17239	0.81153
12	18LeuN,H	18LeuN,H	7.74931	7.74855	119.73473	119.72714	0.8472
13	19AspN,H	19AspN,H	8.23812	8.23845	121.50593	121.51871	0.8516
14	20ArgN,H	20ArgN,H	7.89888	7.89783	122.52163	122.51956	0.82976
15	22ThrN,H	22ThrN,H	7.54237	7.54082	120.7697	120.75874	0.83982
16	23GlyN,H	23GlyN,H	8.59263	8.59236	115.78722	115.78719	0.83508
17	24LysN,H	24LysN,H	7.60715	7.60584	122.42949	122.42506	0.929
18	25TrpN,H	25TrpN,H	8.32662	8.3259	127.93154	127.91949	0.86019
19	26PheN,H	26PheN,H	9.34628	9.34601	116.52196	116.52211	0.85452
20	27TyrN,H	27TyrN,H	9.01695	9.0159	122.52772	122.52867	0.84185
21	28IleN,H	28IleN,H	8.5134	8.51397	130.72025	130.72124	0.95896
22	29AlaN,H	29AlaN,H	6.9532	6.95226	119.53614	119.51033	0.77096
23	30SerN,H	30SerN,H	8.81157	8.81046	111.0602	111.05385	0.85751
24	31AlaN,H	31AlaN,H	7.97778	7.97664	118.1066	118.09939	0.83682
25	32PheN,H	32PheN,H	8.74839	8.74933	114.20418	114.20338	0.78941
26	33ArgN,H	33ArgN,H	9.11534	9.11326	119.77212	119.77593	0.93646
27	34AsnN,H	34AsnN,H	6.63181	6.63345	117.83251	117.84142	0.67927
28	35GluN,H	35GluN,H	9.00263	9.00139	130.34142	130.33654	0.82553
29	36GluN,H	36GluN,H	7.98628	7.98483	120.36285	120.35162	0.74468
30	37TyrN,H	37TyrN,H	8.68953	8.68791	120.18157	120.1855	0.86296
31	38AsnN,H	38AsnN,H	7.81806	7.81809	118.12672	118.12065	0.87477
32	39LysN,H	39LysN,H	8.17409	8.17251	117.97516	117.98561	0.79132
33	40SerN,H	40SerN,H	8.10331	8.10208	114.86783	114.87632	0.82999
34	41ValN,H	41ValN,H	8.58425	8.58478	117.60378	117.61764	0.78361
35	42GlnN,H	42GlnN,H	7.82492	7.82524	122.30046	122.29642	0.79987
36	43GluN,H	43GluN,H	7.07779	7.07733	113.97642	113.99145	0.6841
37	44IleN,H	44IleN,H	7.61631	7.61707	120.65724	120.66072	0.7309
38	45GlnN,H	45GlnN,H	8.68692	8.68311	128.76561	128.76017	0.79534
39	46AlaN,H	46AlaN,H	8.0203	8.01979	116.18532	116.16742	0.83679
40	47ThrN,H	47ThrN,H	10.40517	10.40259	124.40581	124.40783	0.79674
41	48PheN,H	48PheN,H	8.81286	8.81244	123.97204	123.9546	0.88449
42	49PheN,H	49PheN,H	9.06262	9.06284	117.00597	117.00427	0.83367
43	50TyrN,H	50TyrN,H	9.28471	9.28349	117.16297	117.17472	0.81799
44	51PheN,H	51PheN,H	9.20436	9.20455	120.47543	120.4696	0.81673

45	52ThrN,H	52ThrN,H	8.90865	8.9073	118.24146	118.24652	0.77942
46	54AsnN,H	54AsnN,H	9.16185	9.16049	122.95396	122.9542	0.8218
47	55LysN,H	55LysN,H	8.89956	8.90011	122.3406	122.32724	0.89335
48	56ThrN,H	56ThrN,H	7.99668	7.99569	110.53209	110.53377	0.78034
49	57GluN,H	57GluN,H	7.17671	7.17655	118.99511	118.9799	0.75282
50	58AspN,H	58AspN,H	7.79487	7.79389	117.9032	117.90903	0.85192
51	59ThrN,H	59ThrN,H	7.67894	7.67802	107.02745	107.02443	0.8684
52	60IleN,H	60IleN,H	8.74409	8.74553	118.41949	118.42063	0.90256
53	61PheN,H	61PheN,H	8.42859	8.42755	129.33673	129.3338	0.90264
54	62LeuN,H	62LeuN,H	8.95943	8.95404	135.43843	135.42761	0.97786
55	63ArgN,H	63ArgN,H	8.02478	8.02347	124.08449	124.07712	0.83205
56	64GluN,H	64GluN,H	8.86967	8.86856	126.77262	126.7758	0.83868
57	65TyrN,H	65TyrN,H	9.03234	9.03125	127.69617	127.68583	0.82732
58	67ThrN,H	67ThrN,H	8.75271	8.7489	125.25716	125.23692	0.82672
59	68ArgN,H	68ArgN,H	8.98401	8.9831	126.88061	126.8836	0.83416
60	69GlnN,H	69GlnN,H	9.31396	9.31266	120.69066	120.67989	0.74069
61	70AsnN,H	70AsnN,H	9.32339	9.32145	109.88507	109.88031	0.78998
62	71GlnN,H	71GlnN,H	8.03736	8.03638	118.35246	118.35048	0.82888
63	72CysN,H	72CysN,H	8.563	8.56276	119.84108	119.86107	0.72184
64	73PheN,H	73PheN,H	10.03813	10.03724	132.8579	132.84176	0.90464
65	74TyrN,H	74TyrN,H	8.51283	8.51248	126.19928	126.19835	0.93792
66	75AsnN,H	75AsnN,H	8.18123	8.17976	126.85782	126.85782	0.82286
67	77SerN,H	77SerN,H	8.43689	8.43598	116.22745	116.23993	0.84868
68	78TyrN,H	78TyrN,H	8.64995	8.65046	118.24407	118.22644	0.83704
69	79LeuN,H	79LeuN,H	9.61372	9.61167	125.38862	125.3926	0.8022
70	80AsnN,H	80AsnN,H	8.41866	8.41779	118.85241	118.84955	0.89741
71	81ValN,H	81ValN,H	8.00551	8.00432	120.84602	120.8538	0.87824
72	82GlnN,H	82GlnN,H	8.63689	8.63486	129.57834	129.57754	0.8335
73	83ArgN,H	83ArgN,H	8.72906	8.72705	122.04353	122.04442	0.91069
74	84GluN,H	84GluN,H	8.95636	8.95515	117.15224	117.15142	0.79547
75	85AsnN,H	85AsnN,H	7.31125	7.31051	114.96509	114.964	0.84541
76	86GlyN,H	86GlyN,H	7.86147	7.86123	106.82649	106.81115	0.96923
77	88ValN,H	88ValN,H	8.48066	8.4796	111.84189	111.8405	0.85786
78	89SerN,H	89SerN,H	9.21505	9.21384	115.52098	115.52015	0.84118
79	90ArgN,H	90ArgN,H	8.67886	8.67818	119.60378	119.60779	0.72538
80	91TyrN,H	91TyrN,H	8.7559	8.75606	126.41546	126.39608	0.96115
81	92GluN,H	92GluN,H	8.63393	8.63376	127.48134	127.48482	0.75653
82	94GlyN,H	94GlyN,H	8.26075	8.25821	106.17317	106.18229	0.67733
83	95ArgN,H	95ArgN,H	7.18612	7.18592	117.87983	117.87596	0.65331
84	96GluN,H	96GluN,H	8.34263	8.34184	122.22439	122.21626	0.7251
85	97HisN,H	97HisN,H	9.21713	9.21512	127.48812	127.47651	0.74872
86	98ValN,H	98ValN,H	8.68856	8.68822	122.3821	122.38367	0.85785
87	99AlaN,H	99AlaN,H	9.3715	9.37088	133.43502	133.43456	0.89594
88	100HisN,H	100HisN,H	9.04019	9.04075	119.82091	119.82715	0.90449
89	101LeuN,H	101LeuN,H	7.54645	7.54285	124.42456	124.4074	0.77567

90	102LeuN,H	102LeuN,H	8.51167	8.5103	128.29972	128.28288	0.76826
91	103PheN,H	103PheN,H	8.27507	8.27398	117.60509	117.59186	0.75681
92	104LeuN,H	104LeuN,H	9.05389	9.05357	124.84606	124.84504	0.7518
93	107ThrN,H	107ThrN,H	8.1883	8.18969	115.80205	115.79579	0.78524
94	108LysN,H	108LysN,H	8.95181	8.95256	124.36295	124.34979	0.78817
95	109ThrN,H	109ThrN,H	7.18899	7.18887	105.71746	105.72486	0.7848
96	110LeuN,H	110LeuN,H	8.48112	8.47975	122.17092	122.17243	0.84177
97	111MetN,H	111MetN,H	8.10847	8.10686	120.90645	120.93651	0.8919
98	112PheN,H	112PheN,H	8.42815	8.42851	119.3398	119.32843	0.88775
99	113GlyN,H	113GlyN,H	9.88458	9.88419	111.37328	111.3742	0.8935
100	114SerN,H	114SerN,H	9.11271	9.11367	124.015	124.01284	0.84016
101	115TyrN,H	115TyrN,H	9.46909	9.46717	121.5816	121.58668	0.89202
102	117AspN,H	117AspN,H	8.61662	8.61602	113.64726	113.63902	0.81224
103	118AspN,H	118AspN,H	8.44684	8.44648	121.74347	121.76328	0.78001
104	119GluN,H	119GluN,H	8.4647	8.46381	126.77136	126.77602	0.82718
105	120LysN,H	120LysN,H	7.68697	7.68662	115.35816	115.36279	0.83044
106	121AsnN,H	121AsnN,H	7.48118	7.48576	115.01418	115.01799	0.72737
107	122TrpN,H	122TrpN,H	7.70338	7.70344	120.65001	120.66254	0.97019
108	123GlyN,H	123GlyN,H	8.66105	8.65971	104.86141	104.8712	0.86958
109	124LeuN,H	124LeuN,H	8.00293	8.00358	117.00779	117.05032	0.83014
110	125SerN,H	125SerN,H	8.52921	8.52756	111.99609	111.99478	0.9015
111	126PheN,H	126PheN,H	8.41884	8.41731	124.78778	124.74809	0.92121
112	127TyrN,H	127TyrN,H	8.2494	8.24585	126.58543	126.56449	0.91552
113	128AlaN,H	128AlaN,H	8.80886	8.80815	120.16981	120.15715	0.8653
114	129AspN,H	129AspN,H	8.35776	8.35576	120.4035	120.40722	0.8502
115	130LysN,H	130LysN,H	7.12056	7.11854	119.25981	119.26182	0.87492
116	132GluN,H	132GluN,H	6.5163	6.51614	115.95261	115.9481	0.54734
117	133ThrN,H	133ThrN,H	8.21569	8.2135	113.17057	113.15782	0.66

118	134ThrN,H	134ThrN,H	8.59675	8.59535	112.46803	112.46659	0.79768
119	135LysN,H	135LysN,H	8.98685	8.98708	122.14106	122.12458	0.77973
120	136GluN,H	136GluN,H	8.77187	8.77073	121.1154	121.12221	0.82189
121	137GlnN,H	137GlnN,H	7.78684	7.78461	120.57073	120.56816	0.77523
122	138LeuN,H	138LeuN,H	7.888	7.8874	119.83081	119.83893	0.85035
123	139GlyN,H	139GlyN,H	8.11528	8.11466	105.9809	105.97106	0.83217
124	140GluN,H	140GluN,H	7.29682	7.29772	120.13868	120.13623	0.83061
125	141PheN,H	141PheN,H	6.80541	6.80421	121.95456	121.96093	0.88828
126	142TyrN,H	142TyrN,H	8.12187	8.12242	117.97403	117.96501	0.9303
127	143GluN,H	143GluN,H	7.71104	7.71008	118.96701	118.95327	0.87267
128	144AlaN,H	144AlaN,H	7.17647	7.17574	123.47739	123.46735	0.8474
129	145LeuN,H	145LeuN,H	7.68157	7.68017	118.3247	118.31962	0.89692
130	146AspN,H	146AspN,H	8.01797	8.01765	119.49518	119.5115	0.82171
131	147CysN,H	147CysN,H	7.58647	7.5863	120.58054	120.59783	0.81212
132	148LeuN,H	148LeuN,H	6.93721	6.93324	118.567	118.55208	0.75257
133	149ArgN,H	149ArgN,H	7.6942	7.69273	114.3033	114.31076	0.86683
134	150IleN,H	150IleN,H	7.48161	7.48228	122.25295	122.25083	0.82589
135	152ArgN,H	152ArgN,H	8.37825	8.37783	122.2277	122.23785	0.83939
136	153SerN,H	153SerN,H	7.86521	7.86431	111.84319	111.84493	0.89943
137	154AspN,H	154AspN,H	7.65182	7.65087	123.47562	123.47884	0.85745
138	155ValN,H	155ValN,H	7.04975	7.04818	120.90438	120.90873	0.82046
139	156MetN,H	156MetN,H	9.07936	9.0799	128.47098	128.46255	0.96535
140	157TyrN,H	157TyrN,H	8.54689	8.54501	118.53743	118.54583	0.82949
141	158ThrN,H	158ThrN,H	7.74058	7.7406	113.17873	113.17994	0.74674
142	159AspN,H	159AspN,H	8.54538	8.54062	122.37003	122.24058	0.67398
143	160TrpN,H	160TrpN,H	9.18442	9.18218	126.54191	126.52451	0.82848
144	161LysN,H	161LysN,H	8.08047	8.07788	116.85774	116.85366	0.86132
145	163AspN,H	163AspN,H	7.08327	7.08165	118.41549	118.41644	0.81969

146	164LysN,H	164LysN,H	8.55726	8.55588	123.93378	123.94028	0.86648
147	165CysN,H	165CysN,H	9.26291	9.26417	117.96175	117.96099	0.88726
148	166GluN,H	166GluN,H	11.02626	11.02376	129.54848	129.54055	0.76875
149	168LeuN,H	168LeuN,H	7.41845	7.41652	117.80119	117.80581	0.67892
150	169GluN,H	169GluN,H	7.74305	7.7422	120.27623	120.26737	0.58879
151	170LysN,H	170LysN,H	7.60603	7.6051	118.61814	118.61117	0.58801
152	171GlnN,H	171GlnN,H	7.83963	7.84046	120.26049	120.28347	0.32507
153	172HisN,H	172HisN,H	7.69744	7.69787	126.16733	126.09444	0.06446

a.4 NOE ratios of each residue measured at pH 6.5, 750 μM AGP2, three-fold excess of UCN-01 at 700 MHz.

Res Num	Res Type	No. of Islands	SS	Phi	Psi	Phi Upper	Phi Lower	Psi Upper	Psi Lower	Chemical Shifts
1	Gln	1	C	-72.7305	135.0443	-40	-110	179.99	90	CA CB H HA N
2	Ile	1	C	-68.7511	140.7373	-50	-100	170	100	CA CB H HA N
3	Pro	1	C	-59.0506	140.3952	-40	-90	170	110	CA CB
4	Leu	None	None	None	None	None	None	None	None	CA CB H HA N
5	Cys	None	None	None	None	None	None	None	None	CA CB H HA N
6	Ala	1	C	-56.1963	-30.895	-30	-100	10	-60	CA CB H HA N
7	Asn	1	C	-72.7809	-13.7881	-40	-120	20	-50	CA CB H HA N
8	Leu	1	H	-90.7633	-4.96936	-60	-120	30	-40	CA CB H HA N
9	Val	1	C	-71.4673	137.2546	-40	-100	179.99	100	CA CB H HA N
10	Pro	1	C	-64.4339	135.5129	-40	-90	170	110	CA CB
11	Val	1	C	-106.268	118.1109	-60	-160	170	90	CA CB H N
12	Pro	1	E	-67.0671	141.9123	-40	-100	170	110	CA CB
13	Ile	2	C	-104.742	4.7427	-70	-120	20	-40	CA CB H N
14	Thr	1	C	-85.111	165.0008	-60	-120	-170	140	CA CB H N
15	Asn	1	H	-58.1936	-41.2099	-40	-90	-10	-70	CA CB H HA N
16	Ala	1	H	-64.8715	-37.2978	-40	-90	-20	-60	CA CB H HA N
17	Thr	1	H	-64.4555	-44.5175	-40	-90	-20	-70	CA CB H HA N
18	Leu	1	H	-64.7988	-39.8734	-40	-90	-20	-60	CA CB H HA N
19	Asp	1	H	-64.9154	-44.6427	-40	-90	-20	-70	CA CB H HA N
20	Arg	1	H	-64.8817	-39.831	-40	-90	-10	-70	CA CB H N
21	Ile	1	H	-86.3188	-4.14959	-60	-120	30	-40	CA CB H HA N
22	Thr	1	C	-55.3798	135.005	-40	-80	160	110	CA CB H HA N
23	Gly	1	C	84.10316	-4.04445	110	50	40	-40	CA H N
24	Lys	1	E	-114.67	134.9023	-90	-140	160	110	CA CB H HA N
25	Trp	1	E	-134.571	154.4861	-100	-160	179.99	120	CA CB H HA N

26	Phe	1	E	-124.29	134.2781	-80	-160	170	100	CA CB H HA N
27	Tyr	1	E	-102.82	117.8697	-60	-140	150	90	CA CB H HA N
28	Ile	1	E	-96.022	118.7044	-60	-140	160	100	CA CB H N
29	Ala	1	E	-135.516	160.2638	-110	-170	-170	120	CA CB H HA N
30	Ser	1	C	-149.906	160.912	-60	-180	-170	120	CA CB H HA N
31	Ala	1	E	-149.298	160.6636	-90	-180	-170	120	CA CB H HA N
32	Phe	1	E	-136.013	159.7168	-100	-180	179.99	130	CA CB H N
33	Arg	2	C	-70.7123	-24.1753	-40	-100	10	-50	CA CB H N
34	Asn	None	None	None	None	None	None	None	None	CA CB H N
35	Glu	1	H	-57.2629	-41.1997	-40	-90	-10	-70	CA CB H N
36	Glu	1	H	-61.8618	-42.0272	-40	-90	-10	-70	CA CB H N
37	Tyr	1	H	-65.8807	-42.0445	-40	-90	-20	-60	CA CB H N
38	Asn	1	H	-56.8371	-44.9212	-40	-80	-20	-70	CA CB H HA N
39	Lys	1	H	-64.6634	-43.5748	-40	-90	-10	-70	CA CB H N
40	Ser	1	H	-62.5657	-43.0712	-40	-90	-20	-70	CA CB H HA N
41	Val	1	H	-64.6075	-42.9103	-40	-90	-10	-70	CA CB H HA N
42	Gln	1	H	-64.0057	-40.8634	-40	-90	-10	-70	CA CB H N
43	Glu	1	H	-75.2494	-34.9849	-50	-110	10	-60	CA CB H HA N
44	Ile	None	None	None	None	None	None	None	None	CA CB H HA N
45	Gln	1	C	-55.1759	-35.1237	-30	-90	0	-70	CA CB H HA N
46	Ala	1	C	-91.7667	3.52196	-60	-130	40	-30	CA CB H HA N
47	Thr	2	C	-74.7392	125.7425	-40	-130	179.99	100	CA CB H
48	Phe	1	E	-132.013	154.1935	-90	-160	179.99	110	CA CB H HA N
49	Phe	1	E	-133.883	150.9984	-90	-170	179.99	120	CA CB H HA N
50	Tyr	1	E	-125.565	145.7302	-100	-170	179.99	120	CA CB H HA N
51	Phe	1	E	-124.88	130.0311	-90	-150	160	110	CB H HA N
52	Thr	1	E	-116.86	130.9135	-80	-150	160	90	CA CB H HA N
53	Pro	1	C	-65.0603	151.1584	-40	-90	179.99	120	CA CB
54	Asn	None	None	None	None	None	None	None	None	CA CB H HA N
55	Lys	1	H	-56.6078	-33.2814	-30	-80	0	-70	CA CB H N
56	Thr	1	H	-65.3174	-27.071	-40	-90	0	-60	CA CB H N
57	Glu	1	C	-87.0695	-4.06324	-60	-130	30	-40	CA CB H N
58	Asp	1	C	55.01026	44.98974	80	40	60	20	CA CB H HA N
59	Thr	1	E	-132.621	153.4544	-100	-170	179.99	110	CA CB H HA N
60	Ile	1	E	-120.072	127.7119	-80	-150	160	100	CA CB H HA N
61	Phe	1	E	-98.0343	117.34	-70	-140	150	90	CA CB H HA N
62	Leu	1	E	-115.219	138.4524	-80	-150	170	100	CA CB H HA N
63	Arg	1	E	-123.708	127.5574	-80	-160	170	100	CA CB H HA N
64	Glu	1	E	-114.267	134.2422	-80	-150	170	100	CA CB H HA N
65	Tyr	1	E	-118.002	127.4782	-80	-160	170	100	CA CB H HA N
66	Gln	1	E	-120.492	134.1842	-90	-160	179.99	100	CA CB H HA N
67	Thr	1	E	-108.563	124.4852	-80	-140	150	100	CA CB H HA N
68	Arg	1	E	-124.164	137.708	-80	-170	179.99	100	CA CB H HA N
69	Gln	2	C	-64.8949	-15.2949	-30	-100	10	-60	CA CB H N
70	Asn	2	C	-69.9162	-20.4834	-40	-100	10	-50	CA CB H HA N
71	Gln	None	None	None	None	None	None	None	None	CA CB H N

72	Cys	1	E	-97.5529	118.6864	-50	-150	170	90	CA CB H N
73	Phe	1	E	-111.559	127.213	-70	-150	160	100	CA CB H HA N
74	Tyr	1	E	-123.22	125.2346	-90	-150	160	100	CA CB H N
75	Asn	1	E	-117.021	132.9381	-70	-150	170	100	CA CB H HA N
76	Ser	1	E	-118.844	137.1572	-80	-160	170	100	CA CB H HA N
77	Ser	2	E	-139.003	154.2987	-70	-180	-170	110	CA CB H HA N
78	Tyr	1	C	-71.9139	143.1241	-40	-120	170	110	CA CB H HA N
79	Leu	1	C	-114.563	140.2114	-50	-170	179.99	100	CA CB H HA N
80	Asn	1	C	-127.194	135.0366	-60	-160	170	90	CA CB H N
81	Val	1	E	-115.283	126.0413	-70	-150	160	100	CA CB H N
82	Gln	2	C	-65.4388	146.2681	-40	-110	179.99	110	CA CB H HA N
83	Arg	1	H	-55.0221	-34.9513	-30	-80	0	-60	CA CB H N
84	Glu	1	H	-64.9834	-24.9921	-40	-80	0	-50	CA CB H HA N
85	Asn	1	C	-85.0786	-3.17356	-60	-130	20	-20	CA CB H HA N
86	Gly	1	C	85.19527	6.87491	110	40	60	-30	CA H N
87	Thr	1	E	-134.661	156.0381	-110	-160	179.99	130	CA CB H N
88	Val	1	E	-130.966	153.8473	-100	-160	179.99	120	CA CB H HA N
89	Ser	1	E	-127.174	145.9256	-100	-160	179.99	110	CA CB H HA N
90	Arg	1	E	-125.374	134.7661	-100	-160	160	110	CA CB H N
91	Tyr	1	E	-105.175	123.6615	-70	-140	150	100	CA CB H HA N
92	Glu	1	E	-119.728	141.4574	-60	-170	179.99	80	CA CB H HA N
93	Gly	1	C	55.03506	41.53718	80	30	60	10	CA H N
94	Gly	1	C	75.15846	14.55763	100	50	30	-20	CA H N
95	Arg	1	E	-124.034	143.2572	-90	-160	179.99	110	CA CB H N
96	Glu	1	E	-91.1225	134.7643	-60	-150	160	100	CA CB H N
97	His	1	E	-125.818	132.9947	-80	-160	160	100	CA CB H HA N
98	Val	1	E	-108.38	125.3725	-70	-150	150	100	CA CB H HA N
99	Ala	1	E	-121.602	131.9283	-90	-150	170	100	CA CB H HA N
100	His	1	E	-77.1068	136.1821	-50	-140	170	100	CA CB H HA N
101	Leu	1	E	-110.932	125.282	-70	-140	160	100	CA CB H N
102	Leu	1	E	-121.938	137.1655	-80	-160	179.99	100	CA CB H HA N
103	Phe	1	E	-114.993	134.0291	-80	-150	170	100	CA CB H HA N
104	Leu	None	None	None	None	None	None	None	None	H N
105	Arg	1	C	-62.8166	-27.3489	-40	-90	0	-70	CA CB H N
106	Asp	1	C	-71.313	-19.3067	-40	-120	30	-70	CA CB H N
107	Thr	1	C	-65.7655	143.4098	-40	-100	170	110	CA CB H HA N
108	Lys	2	E	-94.4642	135.0134	-40	-130	170	100	CA CB H HA N
109	Thr	1	E	-136.627	155.469	-100	-170	179.99	120	CA CB H HA N
110	Leu	1	E	-127.389	139.6829	-90	-160	170	110	CA CB H HA N
111	Met	1	E	-115.617	135.0257	-90	-160	179.99	100	CA CB H HA N
112	Phe	1	E	-120.005	133.0854	-90	-150	170	100	CA CB H HA N
113	Gly	1	E	-83.4809	133.1042	-50	-160	179.99	80	CA H N
114	Ser	1	C	-70.778	146.0134	-40	-100	179.99	110	CA CB H HA N
115	Tyr	2	C	-65.2954	145.1559	-40	-100	179.99	110	CA CB H HA N
116	Leu	2	C	-77.8672	-16.057	-40	-120	30	-50	CA CB
117	Asp	1	C	-75.7037	-10.9241	-50	-120	30	-40	CA CB H HA N

118	Asp	None	None	None	None	None	None	None	None	CA CB H HA N
119	Glu	1	C	-60.6778	-28.1812	-40	-90	10	-60	CA CB H N
120	Lys	2	C	-66.3325	-14.9175	-40	-110	20	-50	CA CB H N
121	Asn	None	None	None	None	None	None	None	None	CA CB H N
122	Trp	2	C	-91.0674	142.7728	-40	-180	-170	90	CA CB H N
123	Gly	None	None	None	None	None	None	None	None	CA H N
124	Leu	1	E	-142.208	155.4291	-110	-170	-170	130	CA CB H HA N
125	Ser	1	E	-127.322	146.0463	-100	-170	179.99	110	CA CB H N
126	Phe	1	E	-123.235	137.949	-100	-150	179.99	110	CA CB H HA N
127	Tyr	1	E	-125.14	153.9649	-100	-160	179.99	120	CA CB H HA N
128	Ala	1	E	-130.458	144.6134	-80	-180	179.99	100	C CA CB H HA N
129	Asp	2	E	-75.1419	135.9401	-40	-130	179.99	90	CA CB H HA N
130	Lys	1	C	-86.1345	136.4302	-40	-180	179.99	90	CA CB H N
131	Pro	1	C	-62.9018	144.0217	-40	-90	179.99	120	CA CB
132	Glu	None	None	None	None	None	None	None	None	CA CB H HA N
133	Thr	1	E	-116.267	127.4794	-90	-160	179.99	100	CA CB H HA N
134	Thr	1	C	-84.9932	165.0002	-60	-100	-170	150	CA CB H HA N
135	Lys	1	H	-63.6299	-39.7091	-40	-90	-10	-70	CA CB H N
136	Glu	1	H	-64.8832	-44.1474	-40	-90	-20	-70	CA CB H N
137	Gln	1	H	-64.6961	-43.6179	-40	-90	-10	-70	CA CB H N
138	Leu	1	H	-64.4116	-44.1623	-40	-90	-20	-70	CA CB H HA N
139	Gly	1	H	-64.9541	-35.9865	-40	-90	-20	-60	CA H N
140	Glu	1	H	-64.9402	-44.1876	-40	-90	-20	-70	CA CB H N
141	Phe	1	H	-64.8244	-42.479	-40	-90	-20	-70	CA CB H N
142	Tyr	1	H	-64.78	-39.6801	-40	-90	-20	-70	CA CB H N
143	Glu	1	H	-63.4772	-44.8985	-40	-90	-20	-70	C CA CB H N
144	Ala	1	H	-62.559	-41.5139	-40	-90	-10	-70	CA CB H HA N
145	Leu	1	H	-64.9052	-37.4784	-40	-90	-20	-60	CA CB H N
146	Asp	1	H	-64.9261	-44.1091	-40	-90	-20	-70	CA CB H HA N
147	Cys	1	H	-64.5569	-25.3932	-40	-90	0	-60	CA CB H N
148	Leu	1	C	-91.3622	2.17721	-60	-120	30	-30	CA CB H HA N
149	Arg	1	C	57.33877	42.49221	80	40	60	10	CA CB H N
150	Ile	1	C	-84.2765	128.9212	-50	-140	170	90	CA CB H HA N
151	Pro	1	C	-63.7968	143.338	-40	-90	170	120	CA CB
152	Arg	1	C	-55.1946	-34.7466	-40	-90	0	-60	CA CB H N
153	Ser	1	H	-65.606	-14.5181	-40	-100	20	-50	CA CB H N
154	Asp	2	H	-85.3073	-4.91526	-60	-120	20	-40	CA CB H HA N
155	Val	1	C	-79.7523	132.1211	-50	-150	160	90	CA CB H HA N
156	Met	1	E	-94.8891	134.995	-60	-130	160	100	CA CB H N
157	Tyr	1	E	-138.322	153.6266	-90	-180	179.99	120	CA CB H N
158	Thr	1	E	-131.952	139.1649	-90	-160	170	100	CA CB H HA N
159	Asp	1	E	-122.368	130.3725	-70	-160	179.99	90	CA CB H HA N
160	Trp	1	C	-59.8106	-21.2774	-40	-90	10	-50	CA CB H HA N
161	Lys	1	C	-64.9952	-16.5539	-40	-90	10	-60	CA CB H N
162	Lys	1	C	-106.326	6.28499	-70	-130	40	-20	CA CB H N
163	Asp	None	None	None	None	None	None	None	None	CA CB H HA N

164	Lys	2	C	-148.283	153.9789	-40	-180	-170	90	CA CB H N
165	Cys	None	None	None	None	None	None	None	None	CA CB H N
166	Glu	None	None	None	None	None	None	None	None	CA CB H N
167	Pro	1	H	-55.0322	-34.9561	-30	-90	0	-70	CA CB
168	Leu	1	H	-64.9687	-22.4049	-40	-90	0	-50	CA CB H N
169	Glu	1	H	-62.9338	-35.6323	-40	-90	0	-70	CA CB H N
170	Lys	1	H	-84.63	-5.47664	-50	-120	20	-40	CA CB H N

a.5 DANGLE predictions based on chemical shift assignments of AGP2 residues.

	orm2 run42
Wavelength	
Resolution range	44.13 - 1.82 (1.885 - 1.82)
Space group	P 61
Unit cell	88.2555 88.2555 53.7996 90 90 120
Total reflections	371467 (21605)
Unique reflections	21448 (2093)
Multiplicity	17.3 (10.3)
Completeness (%)	99.50 (96.85)
Mean I/sigma(I)	39.04 (6.32)
Wilson B-factor	19.89
R-merge	0.06559 (0.1371)
R-meas	0.06748 (0.1442)
R-pim	0.01571 (0.04361)
CC1/2	0.999 (0.992)
CC*	1 (0.998)
Reflections used in refinement	21447 (2093)
Reflections used for R-free	979 (115)
R-work	0.2122 (0.2706)
R-free	0.2496 (0.2936)
CC(work)	0.952 (0.811)
CC(free)	0.924 (0.809)

a.6. Table of crystallographic parameters. This author wishes to thank Dr Chris Williams for carrying out the structure calculations

References

Akiyama T, Yoshida T, Tsujita T, Shimizu M, Mizukami T, Okabe M & Akinaga S (1997) G1 phase accumulation induced by UCN-01 is associated with dephosphorylation of Rb and CDK2 proteins as well as induction of CDK inhibitor p21/Cip1/WAF1/Sdi1 in p53-mutated human epidermoid carcinoma A431 cells. *Cancer Res.* **57:** 1495–1501

Alderson TR & Kay LE (2020) Unveiling invisible protein states with NMR spectroscopy. *Curr. Opin. Struct. Biol.* **60:** 39–49

Alexander SPH, Davenport AP, Kelly E, Marrion N, Peters JA, Benson HE, Faccenda E, Pawson AJ, Sharman JL, Southan C, Davies JA & Collaborators C (2015) The Concise Guide to Pharmacology 2015/15: G protein-coupled receptors. *British Journal of Pharmacology* **172:** 5744–5869

Antanasijevic A, Ramirez B & Caffrey M (2014) Comparison of the Sensitivities of WaterLOGSY and Saturation. *J. Biomol NMR* **60:** 37–44

Arnold JT, Dharmatti SS & Packard ME (1951) Chemical Effects on Nuclear Induction Signals from Organic Compounds. *J. Chem. Phys.* **28:** 507

Baldwin JM (1993) The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* **12:** 1693–1703

Ballesteros JA & Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors *Methods Neurosci* **25**: 366-428

Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SGF, Shi L, Gether U & Javitch JA (2001) Activation of the β2-Adrenergic Receptor Involves Disruption of an Ionic Lock between the Cytoplasmic Ends of Transmembrane Segments 3 and 6. *J. Biol. Chem.* **276:** 29171–29177

Baltoumas FA, Theodoropoulou MC & Hamodrakas SJ (2013) Interactions of the α-subunits of heterotrimeric G-proteins with GPCRs, effectors and RGS proteins: A critical review and analysis of interacting surfaces, conformational shifts, structural diversity and electrostatic potentials. *J. Struct. Biol.* **182:** 209–218

Barrett PJ, Chen J, Cho MK, Kim JH, Lu Z, Mathew S, Peng D, Song Y, Van Horn WD, Zhuang T, Sönnichsen FD & Sanders CR (2013) The quiet renaissance of protein nuclear magnetic resonance. *Biochemistry* **52:** 1303–1320

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD & Zardecki C (2000) The protein data bank. *Nucleic Acids Res.* **28:** 235–242 www.rcsb.org: accessed 22/05/20

Bocquet N, Kohler J, Hug MN, Kusznir EA, Rufer AC, Dawson RJ, Hennig M, Ruf A, Huber W & Huber S (2015) Real-time monitoring of binding events on a thermostabilized human A_{2A} receptor embedded in a lipid bilayer by surface plasmon resonance. *Biochim. Biophys. Acta - Biomembr.* **1848:** 1224–1233

Bteich M (2019) An overview of albumin and alpha-1-acid glycoprotein main characteristics: highlighting the roles of amino acids in binding kinetics and molecular interactions. *Heliyon* **5:** e02879

Bunch R & Eastman A (1996) Enhancement of Cisplatin-induced Inhibitor1 Cytotoxicity a New by. *Clin. Cancer Res.* **2:** 791–797

Burmann BM, Gerez JA, Matečko-Burmann I, Campioni S, Kumari P, Ghosh D, Mazur A, Aspholm EE, Šulskis D, Wawrzyniuk M, Bock T, Schmidt A, Rüdiger SGD, Riek R, Hiller S (2020) Regulation of α-Synuclein by Chaperones in Mammalian Cells. *Nature* **577:** 127-132

Cady SD, Schmidt-Rohr K, Wang J, Soto CS, Degrado WF & Hong M (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463:** 689–692

Caenepeel S, Brown SP, Belmontes B, Moody G, Keegan KS, Chui D, Whittington DA, Huang X, Poppe L, Cheng AC, Cardozo M, Houze J, Li Y, Lucas B, Paras NA, Wang X, Taygerly JP, Vimolratana M, Zancanella M, Zhu L, et al (2018) AMG 176, a selective MCL1 inhibitor, is effective in hematologic cancer models alone and in combination with established therapies. *Cancer Discov.* **8:** 1582–1597

Cala O, Guillière F & Krimm I (2014) NMR-based analysis of protein-ligand interactions. *Anal. Bioanal. Chem.* **406:** 943–956

Carpenter B, Nehmé R, Warne T, Leslie AGW & Tate CG (2016) Structure of the adenosine A2Areceptor bound to an engineered G protein. *Nature* **536:** 104–107

Carr HY & Purcell EM (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments *Phys. Rev.* **94** 630-638

Chen D, Errey JC, Heitman LH, Marshall FH, Ijzerman AP & Siegal G (2012) Fragment screening of GPCRs using biophysical methods: Identification of ligands of the adenosine A2A receptor with novel biological activity. *ACS Chem. Biol.* **7:** 2064–2073

Chen J, Moratalla R, Impagnatiello F, Grandy DK, Cuellar B, Rubinstein M, Beilstein MA, Hackett E, Fink JS, Low MJ, Ongini E & Schwarzschild MA (2001b) The role of the D2 dopamine receptor (D2R) in A2A adenosine receptor ($A_{2A}R$)-mediated behavioural and cellular responses as revealed by A2A and D2 receptor knockout mice. *PNAS* **98:** 1970–1975

Chen JF & Cunha RA (2020) The belated US FDA approval of the adenosine A2A receptor antagonist istradefylline for treatment of Parkinson's disease. *Purinergic Signal.*

Chen JF, Eltzschig HK & Fredholm BB (2013) Adenosine receptors as drug targets — what are the challenges? *Nat. Rev. Drug Discov.* **12:** 265–286

Chen JF, Xu K, Petzer JP, Staal R, Xu YH, Beilstein M, Sonsalla PK, Castagnoli K, Castagnoli N & Schwarzschild MA (2001a) Neuroprotection by caffeine and A(2A) adenosine receptor inactivation in a model of Parkinson's disease. *J. Neurosci.* **21:** 2–7

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Foon ST, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK & Stevens RC (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* **318:** 1258–1265

Cheung M, Maguire ML, Stevens TJ & Broadhurst RW (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. of Magn. Reason.* **202:** 223-233

Chien EYT, Liu W, Zhao Q, Katritch V, Han GW, Michael A, Shi L, Newman AH, Javitch JA, Cherezov V & Stevens RC (2011) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330:** 1091–1095

Clark LD, Dikiy I, Chapman K, Rödström KEJ, Aramini J, LeVine M V., Khelashvili G, Rasmussen SGF, Gardner KH & Rosenbaum DM (2017) Ligand modulation of sidechain dynamics in a wild-type human GPCR. *Elife* **6:** 1–27

Cowan SW, Newcomer ME & Jones TA (1990) Crystallographic refinement of human serum retinol binding protein at 2Å resolution. *Proteins Struct. Funct. Bioinforma.* **8:** 44–61

Daaka Y & Luttrell, Louis M Lefkowitz RJ (1997) Switching of the coupling of the β_2-adrenergic receptor to different G proteins by protein kinase A. *Nature* **390:** 88–91

Dalvit C (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov. Today* **14:** 1051–1057

Dalvit C, Fagerness PE, Hadden DTA, Sarver RW & Stockman BJ (2003) Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **125:** 7696–7703

Dalvit C, Pevarello P, Tato M, Veronesi M, Vulpetti A & Sundström M (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **18:** 65–68

Dewire SM, Yamashita DS, Rominger DH, Liu G, Cowan CL, Graczyk TM, Chen X, Pitis PM, Gotchev D, Yuan C, Koblish M, Lark MW & Violin JD (2013) A G Protein-Biased Ligand at the m -Opioid Receptor Is Potently Analgesic with Reduced Gastrointestinal and Respiratory Dysfunction Compared with Morphine s. *J. Pharmacol. Exp. Ther.* **344:** 708–717

Dick FA, Rubin SM (2013) Molecular Mechanisms Underlying RB Protein Function. *Nat. Rev. Mol. Cell Biol.* **14:** 297-306

Dickson CJ, Madej BD, Skjevik ÅA, Betz RM, Teigen K, Gould IR & Walker RC (2014) Lipid14: The amber lipid force field. *J. Chem. Theory Comput.* **10:** 865–879

Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M & Marshall FH (2011) Structure of the adenosine A 2A receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **19:** 1283–1293

Du ZP, Wu BL, Wu X, Lin XH, Qiu XY, Zhan XF, Wang SH, Shen JH, Zheng CP, Wu ZY, Xu LY, Wang D & Li EM (2015) A systematic analysis of human lipocalin family and its expression in esophageal carcinoma. *Sci. Rep.* **5:** 1–14

Eddy MT, Didenko T, Stevens RC & Wüthrich K (2016) β2-Adrenergic Receptor Conformational Response to Fusion Protein in the Third Intracellular Loop. *Structure* **24:** 2190–2197

Eddy MT, Lee MY, Gao ZG, White KL, Didenko T, Horst R, Audet M, Stanczak P, McClary KM, Han GW, Jacobson KA, Stevens RC & Wüthrich K (2018) Allosteric Coupling of Drug Binding and Intracellular Signaling in the A_{2A} Adenosine Receptor. *Cell* **172:** 68–80

Elg SA, Mayer AR, Carson LF, Twiggs LB, Hill RB & Ramakrishnan S (1997) α-1 Acid glycoprotein is an immunosuppressive factor found in ascites from ovarian carcinoma. *Cancer* **80:** 1448–1456

Ferré S, Snaprud P & Fuxe K (1993) Opposing actions of an adenosine A2 receptor agonist and a GTP analogue on the regulation of dopamine D2 receptors in rat neostriatal membranes. *Eur. J. Pharmacol. Mol. Pharmacol.* **244:** 311–315

Feynman RP (1963) *Lecutres on Physics* Addison-Wesley

Fiaux J, Bertelsen EB, Horwich AL & Wu K (2002) NMR analysis of a 900K GroEL – GroES complex. *Nature* **418:** 207–211

Fielding L (2007) NMR methods for the determination of protein – ligand dissociation constants. **51:** 219–242

Fitos I, Visy J, Zsila F, Mády G & Simonyi M (2006) Selective binding of imatinib to the genetic variants of human α1-acid glycoprotein. *Biochim. Biophys. Acta - Gen. Subj.* **1760:** 1704–1712

Frishman D & Argos P (1995) Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinforma.* **23:** 566–579

Furusaki A, Hashiba N, Matsumoto T, Hirano A, Iwai Y & Omura S (1978) X-Ray crystal structure of staurosporine: A new alkaloid from a Streptomyces strain. *J. Chem. Soc. Chem. Commun.***:** 800–801

Genheden S, Essex JW & Lee AG (2017) G protein coupled receptor interactions with cholesterol deep in the membrane. *Biochim. Biophys. Acta - Biomembr.* **1859:** 268–281

Gołembiowska K, Dziubina A, Kowalska M & Kamińska K (2009) Effect of adenosine A2A receptor antagonists on L-DOPA-induced hydroxyl radical formation in rat striatum. *Neurotox. Res.* **15:** 155–166

Graves PR, Yu L, Schwarz JK, Gales J, Sausville EA, O'Connor PM & Piwnica-Worms H (2000) The Chk1 protein kinase and the Cdc25C regulatory pathways are targets of the anticancer agent UCN-01. *J. Biol. Chem.* **275:** 5600–5605

Gusach A, Maslov I, Luginina A, Borshchevskiy V, Mishin A & Cherezov V (2020) Beyond structure: emerging approaches to study GPCR dynamics. *Curr. Opin. Struct. Biol.* **63:** 18–25

Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T & Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482:** 547–551

Hajduk PJ, Augeri DJ, Mack J, Mendoza R, Yang J, Betz SF, Fesik SW, Park A, February R V, Re V, Recei M & May V (2000) NMR-Based Screening of Proteins Containing 13 C-Labeled Methyl Groups. **:** 7898–7904

Harner MJ, Frank AO & Fesik SW (2013) Fragment-based drug discovery using NMR spectroscopy. *J. Biomol. NMR* **56:** 65–75

Harner MJ, Mueller L, Robbins KJ, Reily MD & Noe I (2017) NMR in drug design. *Arch. Biochem. Biophys.* **628:** 132–147

Hernán MA, Takkouche B, Caamaño-Isorna F & Gestal-Otero JJ (2002) A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Ann. Neurol.* **52:** 276–284

Hervé F, Caron G, Duché JC, Gaillard P, Abd. Rahman N, Tsantili-Kakoulidou A, Carrupt PA, D'Athis P, Tillement JP & Testa B (1998) Ligand specificity of the genetic variants of human α1-acid glycoprotein: Generation of a three-dimensional quantitative structure- activity relationship model for drug binding to the a variant. *Mol. Pharmacol.* **54:** 129–138

Horst R, Liu JJ, Stevens RC & Wüthrich K (2013) β2-adrenergic receptor activation by agonists studied with ^{19}F NMR spectroscopy. *Angew. Chemie - Int. Ed.* **52:** 10762–10765

I. Fitosa, Á. Simona, F. Zsilaa, G. Mádya, Á. Bencsuraa, Z. Vargab, L. Őrfib, G. Kérib JV (2012) Characterization of binding mode of imatinib to human α1-acid glycoprotein. **50:** 788–795

Imai S, Yokomizo T, Kofuku Y, Shiraishi Y, Ueda T & Shimada I (2020) Structural equilibrium underlying ligand-dependent activation of β2-adrenoreceptor. *Nat. Chem. Biol.* **16:** 430–439

Jaakola V, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman P, Stevens RC & Jolla L (2008) The 2.6 Angstrom Crystal Structure of a Human A. *Science* **322:** 1211-1217

Jaakola VP, Lane JR, Lin JY, Katritch V, IJzerman AP & Stevens RC (2010) Ligand binding and subtype selectivity of the human A2A adenosine receptor: Identification and characterization of essential amino acid residues. *J. Biol. Chem.* **285:** 13032–13044

Jamshad M, Charlton J, Lin Y, Routledge SJ, Bawa Z, Knowles TJ, Overduin M, Dekker N, Dafforn TR, Bill RM, Poyner DR & Wheatley M (2015) G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Biosci. Rep.* **35:** 1–10

Jo M, Kim J-H, Song GJ, Seo M, Hwang EM & Suk K (2017) Astrocytic Orosomucoid-2 Modulates Microglial Activation and Neuroinflammation. *J. Neurosci.* **37:** 2878–2894

Jo S, Lim JB, Klauda JB & Im W (2009) CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97:** 50–58

Johnson LN, De Moliner E, Brown NR, Song H, Barford D, Endicott JA & Noble MEM (2002) Structural studies with inhibitors of the cell cycle regulatory kinase cyclin-dependent protein kinase 2. *Pharmacol. Ther.* **93:** 113–124

Kale J, Osterlund EJ & Andrews DW (2018) BCL-2 family proteins: Changing partners in the dance towards death. *Cell Death Differ.* **25:** 65–80

Katritch V, Fenalti G, Abola EE, Roth BL, Cherezov V & Stevens RC (2014) Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39:** 233–244

Katsuki M, Chuang VTG, Nishi K, Suenaga A & Otagiri M (2004) Tryptophan residues play an important role in the extraordinarily high affinity binding interaction of UCN-01 to human α-1-acid glycoprotein. *Pharm. Res.* **21:** 1648–1655

Keeler J (2010) *Understanding NMR Spectroscopy*, 2nd edn, Wiley

Kim TH, Chung KY, Manglik A, Hansen AL, Dror RO, Mildorf TJ, Shaw DE, Kobilka BK & Prosser RS (2013) The Role of Ligands on the Equilibria Between Functional States of a G Protein-Coupled Receptor. *J. Am. Chem. Soc.* **135:** 9465–9474

Kozakov D, Chuang GY, Beglov D & Vajda S (2010) Where does amantadine bind to the influenza virus M2 proton channel? *Trends Biochem. Sci.* **35:** 471–475

Kumar K, Woo SM, Siu T, Cortopassi WA, Duarte F & Paton RS (2018) Cation-π interactions in protein-ligand binding: Theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **9:** 2655–2665

Landin EJB, Lovera S, de Fabritiis G, Kelm S, Mercier J, McMillan D, Sessions RB, Taylor RJ, Sands ZA, Joedicke L & Crump MP (2019) The Aminotriazole Antagonist Cmpd-1 Stabilises a Distinct Inactive State of the Adenosine 2A Receptor. *Angew. Chemie - Int. Ed.* **241385:** 9499–9503

Latek D, Modzelewska A, Trzaskowski B, Palczewski K & Filipek S (2012) G protein-coupled receptors-recent advances. *Acta Biochim. Pol.* **59:** 515–529

Li Y & Kang C (2017) Solution NMR spectroscopy in target-based drug discovery. *Molecules* **22:** 1399

Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K & Wüthrich K (2012) Biased Signaling Pathways inb2-Adrenergic Receptor Characterized by 19F-NMR. *Science* **1106:** 1106–1111

Liu X, Kolpak MX, Wu J & Leo GC (2012) Automatic analysis of quantitative NMR data of pharmaceutical compound libraries. *Anal. Chem.* **84:** 6914–6918

Luo Z, Lei H, Sun Y & Liu X (2015) Orosomucoid , an acute response protein with multiple modulating activities. **:** 329–340

Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019 Jul;47(W1):W636-W641

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB *J. Chem. Theory Comput.* **11:** 3696-3713

Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM & Sunahara RK (2012) Crystal structure of the m -opioid receptor bound to a morphinan antagonist. *Nature* **485:** 321–326

Manley G & Loria JP (2012) NMR insights into protein allostery. *Arch. Biochem. Biophys.* **519:** 223–231

Marco V, Stier G, Blandin S, Marco A (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochemical and *Biophys. Res. Commun.* **322**: 766-771

Masi A, Trezza V, Leboffe L & Ascenzi P (2016) Human plasma lipocalins and serum albumin: Plasma alternative carriers? *J. Control. Release* **228:** 191–205

Mayer M & Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* **38:** 1784–1788

Mayer M & Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123:** 6108–6117

McDermott A (2009) Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR. *Annu. Rev. Biophys.* **38:** 385–403

Meiboom S & Gill D (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times *Rev. Sci. Instrum.* **29** 688-691

Meinecke R & Meyer (2001) Determination of the Binding Specificity of an Integral Membrane Protein by Saturation Transfer Difference NMR: RGD Peptide Ligands Binding to Integrin αIIbβ3 *J. Med. Chem.* **44** 3059-3065

Merck Group (Accessed 2019) Competent Cells. *Sales Brochure*

Mizuno K, Kumi N, Ueda Y, Hanaki H, Saido TC, Ikuta T, Kuroki T, Tamaoki T, Hirai S, Osada S & Ohno S (1995) UCN-01 , an anti-tumor drug , is a selective inhibitor of the conventional PKC subfamily. **359:** 259–261

Mochly-rosen D, Das K & Grimes K V (2012) Protein kinase C, an elusive therapeutic target ? **11:** 937–957

Moukhametzianov R, Warne T, Edwards PC, Serrano-vega MJ, Leslie AGW, Tate CG & Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a β_1 -adrenergic receptor.**:** 2–6

Muchmore SW, Sattlert M, Liangt H, Meadows RP, Harlant JE, Yoont HS, Nettesheimt D, Chang BS, Thompson CB, Ii SW, Ngll S & Fesikt SW (1996) X-ray and NMR structure of human Bcl-xu an inhibitor of programmed cell death. **381:** 335–341

Mull BB, Livingston JA, Patel N, Bui T, Hunt KK & Keyomarsi K (2020) Specific, reversible G1 arrest by UCN-01 in vivo provides cytostatic protection of normal cells against cytotoxic chemotherapy in breast cancer. *Br. J. Cancer* **122:** 812–822

National Cancer Institute (2017) UCN-01 (7-Hydroxystaurosporine) to Treat Relapsed T-Cell Lymphomas NCT00082017 Available at: https://clinicaltrials.gov/ct2/show/NCT00082017

New England Biolabs (Accessed 2019) C3026 Datasheet. *Competent Cells Webpage*

Nishi K, Ono T, Nakamura T, Fukunaga N, Izumi M, Watanabe H, Suenaga A, Maruyama T, Yamagata Y, Curry S & Otagiri M (2011) Structural insights into differences in drug-binding selectivity between two forms of human α1-acid glycoprotein genetic variants, the A and F1*S forms. *J. Biol. Chem.* **286:** 14427–14434

Norton RS, Leung EWW, Chandrashekaran IR & MacRaild CA (2016) Applications of 19F-NMR in fragment-based drug discovery. *Molecules* **21:** 860

Ok HG, Kim SY, Lee SJ, Kim TK, Huh BK & Kim KH (2018) Can oliceridine (TRV130), an ideal novel μ receptor G protein pathway selective (μ-GPS) modulator, provide analgesia without opioid-related adverse reactions? *Korean J. Pain* **31:** 73–79

Otting G & Wüthrich K (1990) Heteronuclear filters in two-dimensional [1H, 1H]-NMR spectroscopy: Combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q. Rev. Biophys.* **23:** 39–96

Paesen GC, Adams PL, Harlos K, Nuttall PA & Stuart DI (1999) Tick histamine-binding proteins: Isolation, cloning, and three-dimensional structure. *Mol. Cell* **3:** 661–671

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Trong I Le, Teller DC, Okada T, Stenkamp RE, Yamamoto M & Miyano M (2000) Crystal Structure of Rhodopsin : A G Protein – Coupled Receptor. *Science* **289:** 739–746

Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P & Salemme FR (2001) High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screen.* **6:** 429–440

Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM & Opella SJ (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491:** 779–783

Park WJ, You SH, Choi HA, Chu YJ & Kim GJ (2015) Over-expression of recombinant proteins with N-terminal His-tag via subcellular uneven distribution in Escherichia coli. *Acta Biochim. Biophys. Sin. (Shanghai).* **47:** 488–495

Patients L, Gambacorti-passerini C, Zucchetti M, Russo D, Frapolli R, Verga M, Bungaro S, Tornaghi L, Rossi F, Pioltelli P, Pogliani E, Alberti D, Corneo G & Incalci MD (2003) ⌴ 1 Acid Glycoprotein Binds to Imatinib (STI571) and Substantially Alters Its Pharmacokinetics in Chronic Myeloid. **9:** 625–632

PerkinElmer (2009) Human Adenosine A_{2A} Receptor Cell line. *Technical Data Sheet*

Petros AM, Huth JR, Oost T, Park CM, Ding H, Wang X, Zhang H, Nimmer P, Mendoza R, Sun C, MacK J, Walter K, Dorwin S, Gramling E, Ladror U, Rosenberg SH, Elmore SW, Fesik SW & Hajduk PJ (2010) Discovery of a potent and selective Bcl-2 inhibitor using SAR by NMR. *Bioorganic Med. Chem. Lett.* **20:** 6587–6591

Petros AM, Medek A, Nettesheim DG, Kim DH, Yoon HS, Swift K, Matayoshi ED, Oltersdorf T & Fesik SW (2000) Solution structure of the antiapoptotic protein bcl-2. **2000:** 3012-3017

Poucher M, Keddie JR, Singh P, Stoggall SM, Caulkett PWR, Jones G & Collis G (1995) A2a selective adenosine receptor antagonist. *British Journal of Pharmacology***:** 1096–1102

Quadro L, Blaner WS, Salchow DJ, Vogel S, Piantedosi R, Gouras P, Freeman S, Cosma MP, Colantuoni V & Gottesman ME (1999) Impaired retinal function and vitamin A availability in mice lacking retinol-binding protein. *EMBO J.* **18:** 4633–4644

Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI & Kobilka BK (2007) Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450:** 383–388

Rasmussen SGF, Devree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK & Kobilka BK (2011) Crystal structure of the β 2 adrenergic receptor-Gs protein complex. *Nature* **477:** 549–557

Roe DR & Cheatham TE (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9:** 3084–3095

Rosenbaum DM, Rasmussen SGF & Kobilka BK (2009) The structure and function of G-protein-coupled receptors. *Nature* **459:** 356–363

S. Omura, Y. Iwai, A. Hirano, A. Nakagawa, J. Awaya, H. Tsuchiya, Y. Takahashi, R. Masuma (1977) A New Alkaloid Am-2282 Of Streptomyces Origin Taxonomy, Fermentation, Isolation and Preliminary Characterization. **4:** 275–282

Šali A & Blundell TL (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **234:** 779–815

Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, Harris T, Ballesteros JA & Palczewski K (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 16123–16128

Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI & Overington JP (2016) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16:** 19–34

Savage D.G and Antman K.H. (2002) Imatinib Mesylate — A New Oral Targeted Therapy. **346:** 683–693

Schertler GFX, Villa C & Henderson R (1993) Projection structure of rhodopsin. *Nature* **362:** 770–772

Schönfeld DL, Ravelli RBG, Mueller U, Skerra A, Saatzucht A Der & Cedex G (2008) The 1.8-Å Crystal Structure of α 1 -Acid Glycoprotein (Orosomucoid) Solved by UV RIP Reveals the Broad Drug-Binding Activity of This Human Plasma Lipocalin. *J. Mol. Biol.* **384:** 393–405

Seynaeve CM, Stetler-Stevenson M, Sebers S, Kaur G, Sausville EA & Worland PJ (1993) Cell Cycle Arrest and Growth Inhibition by the Protein Kinase Antagonist UCN-01 in Human Breast Carcinoma Cell. *Cancer Res.* **53:** 2081–2086

Shen Z, Lou K & Wang W (2015) New small-molecule drug design strategies for fighting resistant influenza A. *Acta Pharm. Sin. B* **5:** 419–430

Shimada I, Ueda T, Kofuku Y, Eddy MT & Wüthrich K (2018) GPCR drug discovery: Integrating solution NMR data with crystal and cryo-EM structures. *Nat. Rev. Drug Discov.* **18:** 59–82

Shook BC & Jackson PF (2011) Adenosine A $_{2A}$ Receptor Antagonists and Parkinson's Disease. *ACS Chem. Neurosci.* **2:** 555–567

Smith JS, Lefkowitz RJ & Rajagopal S (2018) Biased signalling: From simple switches to allosteric microprocessors. *Nat. Rev. Drug Discov.* **17:** 243–260

Sugiki T, Kobayashi N & Fujiwara T (2017) Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. *Comput. Struct. Biotechnol. J.* **15:** 328–339

Sun B, Bachhawat P, Chu ML-H, Wood M, Ceska T, Sands ZA, Mercier J, Lebon F, Kobilka TS & Kobilka BK (2017) Crystal structure of the adenosine A_{2A} receptor bound to an antagonist reveals a potential allosteric pocket. *Proc. Natl. Acad. Sci.* **114:** 2066-2071

Takahashi H & Shimada I (2010) Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells. *J. Biomol. NMR* **46:** 3–10

The Uniprot Consortium (2019) UniProt: a worldwide hub of protein knowledge Nucleic Acids Res. **47**: D506-515 (2019)

Thomaston JL, Polizzi NF, Konstantinidi A, Wang J, Kolocouris A & Degrado WF (2018) Inhibitors of the M2 Proton Channel Engage and Disrupt Transmembrane Networks of Hydrogen-Bonded Waters. *J. Am. Chem. Soc.* **140:** 15219–15226

Timur A (1969) Pulsed Nuclear Magnetic Resonance Studies of Porosity, Movable Fluid, and Permeability of Sandstones *J. Pet. Technol.* **21** 775-787

Unger VM, Hargrave PA, Baldwin JM & Schertler GFX (1997) Arrangement of rhodopsin transmembrane α-helices. *Nature* **389:** 203–206

Urban JD, Clarke WP, Zastrow M Von, Nichols DE, Kobilka B, Weinstein H, Javitch JA, Roth BL, Christopoulos A, Sexton PM, Miller KJ, Spedding M & Mailman RB (2007) Functional Selectivity and Classical Concepts of Quantitative Pharmacology. **320:** 1–13

Uustare A, Vonk A, Terasmaa A, Fuxe K & Rinken A (2005) Kinetic and functional properties of [^{3}H]ZM241385, a high affinity antagonist for adenosine A_{2A}receptors. *Life Sci.* **76:** 1513–1526

Vanwetswinkel S, Heetebrij RJ, Van Duynhoven J, Hollander JG, Filippov D V., Hajduk PJ & Siegal G (2005) TINS, target immobilized NMR screening: An efficient and sensitive method for ligand discovery. *Chem. Biol.* **12:** 207–216

Velmurugan BK, Baskaran R & Huang C (2019) Biomedicine & Pharmacotherapy Detailed insight on β - adrenoceptors as therapeutic targets. *Biomed. Pharmacother.* **117:** 109039 Available at:

Villar HO, Yan J & Hansen MR (2004) Using NMR for ligand discovery and optimization. *Curr. Opin. Chem. Biol.* **8:** 387–391

Wang J, Wu Y, Ma C, Fiorin G, Wang J, Pinto LH & Lamb RA (2013) Structure and inhibition of the drug-resistant S31N mutant of the M2 ion channel of influenza A virus. *Proc. Natl. Acad. Sci.* **110:** 1315–1320

Wang JM, Wolf RM, Caldwell JW, Kollman P a & Case D a (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25:** 1157–1174

Weis WI & Kobilka BK (2018) The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu. Rev. Biochem.* **87:** 897–919

Widdifield CM, Farrell JD, Cole JC, Howard JAK & Hodgkinson P (2020) Resolving alternative organic crystal structures using density functional theory and NMR chemical shifts. *Chem. Sci.* **11:** 2987–2992

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73:** 1–16

Williamson MP, Havel TF & Wüthrich K (1985) Solution Conformation of Proteinase Inhibitor IIA from Bull Seminal Plasma by ^{1}H Nuclear Magnetic Resonance and Distance Geometry. *J. Mol. Biol.* **182:** 295–315

World Health Organisation (2019) World Health Organization Model List of Essential Medicines www.who.int: accessed 27/05/2020

Wu Y, Canturk B, Jo H, Ma C, Gianti E, Klein ML, Pinto LH, Lamb RA, Fiorin G, Wang J & Degrado WF (2014) Flipping in the pore: Discovery of dual inhibitors that bind in different orientations to the wild-type versus the amantadine-resistant S31N mutant of the influenza a virus M2 proton channel. *J. Am. Chem. Soc.* **136:** 17987–17995

Yan J, He X, Zhang S, Feng C, Wang J, Hu Q, Cai J & Wang M (2020) Sensitive parameters of NMR T2 spectrum and their application to pore structure characterization and evaluation in logging profile: A case study from Chang 7 in the Yanchang Formation, Heshui area, Ordos Basin, NW China. *Mar. Pet. Geol.* **111:** 230–239

Yan J, Kline AD, Mo H, Shapiro MJ & Zartler ER (2003) The effect of relaxation on the epitope mapping by saturation transfer difference NMR. *J. Magn. Reson.* **163:** 270–276

Ye L, Larda ST, Frank Li YF, Manglik A & Prosser RS (2015) A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in 19F NMR studies of proteins. *J. Biomol. NMR* **62:** 97–103

Ye L, Neale C, Sljoka A, Lyda B, Pichugin D, Tsuchimura N, Larda ST, Pomès R, García AE, Ernst OP, Sunahara RK & Prosser RS (2018) Mechanistic insights into allosteric regulation of the A_{2A} adenosine G protein-coupled receptor by physiological cations. *Nat. Commun.* **9:** 1372

Ye L, Van Eps N, Zimmer M, Ernst OP & Scott Prosser R (2016) Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* **533:** 265–268

Zhao B, Bower MJ, McDevitt PJ, Zhao H, Davis ST, Johanson KO, Green SM, Concha NO & Zhou BBS (2002) Structural basis for Chk1 inhibition by UCN-01. *J. Biol. Chem.* **277:** 46609–46615