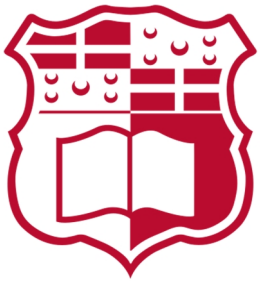
# Protein-Protein Interactions of AIP (Aryl Hydrocarbon Receptor Interacting Protein)

A dissertation presented to the Faculty of Medicine and Surgery in fulfilment of the requirements for the Degree of Doctor of Philosophy at the University of Malta



**Supervisor: Prof. Gary J. Hunter**
**Co-supervisor: Prof. Josanne Vassallo**

**Marita Vella**
**2019**

# Ethics Declaration

This research is in conformity with the University of Malta's research code of practice and research ethics review procedures. It has been approved by FREC, **UNIQUE FORM ID: 342:06.12.18**

_____

Ms Marita Vella

# Declaration of Authenticity

I, the undersigned, **MARITA VELLA**, declare that this dissertation is based on my original work carried out under the supervision of **PROF. G. J. HUNTER** as principal supervisor and **PROF. J. VASSALLO** as co-supervisor, within the Department of Physiology and Biochemistry, to fulfil the aims and objectives of this study.

_____

Ms Marita Vella

*To my late grandmother,*

*Cetta*

***Thanks for all your love, support and prayers***

## ACKNOWLEDGMENTS

The research work disclosed in this publication is partially funded by the Endeavour Scholarship Scheme (Malta). This Scholarship is part-financed by the European Union - European Social Fund (ESF) under Operational Programme II – Cohesion Policy 2014-2020, "Investing in human capital to create more opportunities and promote the well being of society".

European Union- European Structural and Investment Funds
Operational Programme II – Cohesion Policy 2014-2020
*Investing in human capital to create more opportunities and promote the well being of society*

Scholarship part-financed by the European Union European Social Fund (ESF)

Co-financing rate: 80% EU Funds; 20% National Funds

***Investing in your future***

## ACKNOWLEDGMENTS

This journey certainly would not have been possible without the support and understanding of various people to whom I am sincerely grateful. First and foremost, my deepest gratitude goes to my supervisor Prof. Gary J. Hunter for his patient instruction and exceptional guidance throughout this research. Thank you for always finding the time to share your knowledge and discuss any issue regarding the planning and development of this project. My heartfelt appreciation also goes to Prof. Thérèse Hunter who was always readily available, providing constructive suggestions for the way forward and words of comfort whenever needed. I honestly cannot thank you enough. Their expertise in the field of biochemistry and enthusiasm were fundamental for the completion of this research. The invaluable knowledge that they both shared with me will surely be of benefit in my life and career and for which I will always be grateful.

I would also like to show my appreciation towards Mr Brandon Seychell who also started his PhD journey four years ago within the Laboratory of Biochemistry and Protein Science. Thank you for the laughs, support, lengthy discussions and valuable help especially in the analyses and interpretation of X-ray crystallography data.

Moreover, I would like to extend my gratitude towards my co-supervisor, Prof. Josanne Vassallo for her optimistic approach on the potential of this research and her words of encouragement. Thanks also go to Dr Robert Formosa for his willingness to help whenever possible and for providing the AIP cDNA. I am also thankful towards the members of the Allied Research Unit, particularly Prof. Anthony Fenech and the members of the Laboratory of Molecular Oncology for allowing me to utilise equipment within their laboratories.

A number of experiments in this study were carried out at the Astbury Centre for Structural Molecular Biology at the University of Leeds. Therefore, I would like to thank the Director of the Astbury Centre, Prof Sheena Radford, for kindly allowing me to use the facilities. Special thanks go to our collaborator Dr Chi Trinh for facilitating the research visits, his kind assistance throughout and providing constant support and advice. I would also like to express my gratitude towards facility managers Dr Iain Manfield and Dr G. Nasir Khan for their assistance in the use of laboratory equipment

and help in data interpretation. The Mass Spectrometry Facility instrumentation used in this work is supported by BBSRC – BB/M012573/1 and the analysis was carried out by Dr James Ault and Rachel George. Circular dichroism experiments were performed on a Chirascan funded by the Welcome Trust, grant code 094232. Research visits to the University of Leeds were financially supported by COST action CM1306 "Understanding Movement and Mechanism in Molecular Machines" and the Internalisation Partnership Awards Scheme through MCST awarded to Prof. Thérèse Hunter.

I would also like to thank the staff members at B21, Diamond Light Source for their kind assistance during beamtime. Special thanks go to Dr Robert Rambo for his invaluable help in the interpretation of SAXS data and for performing modelling analyses that would have been difficult to accomplish without his expertise.

Last but not least, I would like to express my appreciation towards my family, especially my mother and brother who have always believed in me and supported me every step of the way and my late father who I'm sure gave me strength to persevere during difficult times. Special thanks go to my boyfriend Malcolm, for always being supportive and patient throughout these years despite the bad moods and long hours at the lab. Thanks to all my friends, especially Deborah and Donatella for their constant support and motivation for which I am sincerely grateful. Finally, it is done!




## ABSTRACT

Aryl hydrocarbon receptor interacting protein (AIP) is a cytoplasmic molecular co-chaperone and tumour suppressor that assists in protein stability and complex formation. It is involved in biochemical pathways, such as xenobiotic metabolism, cyclic AMP signalling, mitochondrial import and apoptosis. Germline mutations in the AIP gene predispose to pituitary tumourigenesis with patients exhibiting an aggressive clinical phenotype. This study was focused on the structural and functional characterisation of AIP to investigate whether N-domain mutations, of clinical relevance, affect the ability of the protein to interact with client binding partners. A purification protocol for AIP was successfully devised that maintains the protein in a stable homogenous state. Similarly, variants of full length AIP harbouring N-domain mutations were purified from *E.coli* to the same level of purity. Circular dichroism showed that the mutations did not significantly affect the thermal stability of the protein and caused no overall disruptive effect in the protein structure. However, through ITC and Biacore experiments, these mutations lowered, to different extents, the binding affinity of AIP towards two of its binding partners, Hsp90$\beta^{(513\text{-}724)}$ and PDE4A5, also purified in this study. The possible biological implications of such disruptions are the destabilisation of complexes requiring Hsp90 and changes in cellular cAMP levels respectively. The latter was further demonstrated through a PDE enzymatic assay in which the mutants failed to attenuate the enzymatic activity of PDE to the same degree as the wild type protein. Through the use of small angle X-ray scattering, the full length model of AIP was obtained and provided valuable information on protein shape, flexibility and inter-domain distances.

This study provides clear evidence that AIP N-domain mutations have a significant role in protein-protein interactions and although they may not necessarily contribute directly to pituitary tumourigenesis, the complex interactome of AIP suggests that any observable change in one or more of its binding partners cannot be disregarded as it may have repercussions on other biochemical pathways. To further investigate the AIP protein interactome, a novel protocol for the expression of soluble AhR in *E.coli* has also been established. AhR is another binding partner of AIP that displays both oncogenic and tumour suppressor abilities and is currently at the focus of various studies.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AC | Adenylyl cyclase |
| Ab | Antibody |
| ACTH | Adrenocorticotropic hormone |
| AhR | Aryl hydrocarbon receptor |
| AIP | Aryl hydrocarbon receptor interacting protein |
| APS | Ammonium persulfate |
| ARNT | Aryl hydrocarbon receptor nuclear translocator |
| AU | Absorbance units |
| BA | Benzyl alcohol |
| BCM | Barycentric mean fluorescence |
| bHLH | Basic-helix-loop-helix |
| bp | Base pairs |
| BPER | Bacterial protein extraction reagent |
| cAMP | Cyclic adenosine monophosphate |
| CD | Circular dichroism |
| CM5 | Carboxymethylated |
| CNC | Carney complex |
| Dmax | Maximum diameter of protein |
| dNTP | Deoxyribonucleotide triphosphate |
| DTT | Dithiothreitol |
| EDC | 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide |
| EDTA | Ethylenediaminetetraacetic acid |
| EPR | Electron paramagnetic resonance |
| FIPA | Familial isolated pituitary adenomas |
| FSH | Follicle stimulating hormone |
| GF | Gel filtration |
| GH | Growth hormone |
| GR | Glucocortiocoid receptor |
| $G\alpha_i$ | G-alpha inhibitory protein |
| $H_6$ | Hexa-histidine |
| Hsp90 | Heat shock protein 90 |
| $Hsp90\beta^{(513\text{-}724)}$ | Heat shock protein 90 beta isoform, residues 513-724 |

| | |
|---|---|
| IAP | 3-(2-Iodoacetamido)-PROXYL |
| IBMX | 3-isobutyl-1-methylxanthine |
| IFS | Isolated familial somatotropinoma |
| IMAC | Immobilised metal affinity chromatography |
| IPTG | Isopropyl-β-D-1-thiogalactopyranoside |
| ITC | Isothermal titration calorimetry |
| $K_A$ | Equilibrium association constant |
| $K_D$ | Equilibrium dissociation constant |
| kDa | Kilo daltons |
| KP | Potassium phosphate |
| LH | Luteinizing hormone |
| LOH | Loss of heterozygosity |
| MBP | Maltose binding protein |
| *MEN1* | Menin gene |
| MEN1 | Multiple endocrine neoplasia type 1 |
| MEN4 | Multiple endocrine neoplasia type 4 |
| MLPA | Multiplex ligation dependent probe amplification |
| MWCO | Molecular weight cut-off |
| NFPA | Non-functional pituitary adenomas |
| NHS | N-hydroxysuccinimide |
| $OD_{600}$ | Optical density at 600 nm wavelength |
| P20 | Polysorbate 20 |
| PAP | Pituitary adenoma predisposition |
| PAS | Per ARNT-Sim |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PDE4A5 | Phosphodiesterase isoform 4A5 |
| PDEs | Phosphodiesterases |
| PI | Isoelectric point |
| PKA | Protein kinase A |
| PMSF | Phenylmethylsulfonyl fluoride |
| PPIase | Peptidyl (cis-trans)Prolyl isomerase |
| PRL | Prolactin |

| | |
|---|---|
| rASA | Relative accessible surface |
| RET | Rearranged during transfection |
| Rg | Radius of gyration |
| RLU | Relative luminescence units |
| RU | Response (resonance) units |
| SAXS | Small angle X-ray scattering |
| SDM | Site directed mutagenesis |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide electrophoresis |
| SEC-SAXS | Size exclusion small angle X-ray scattering |
| SHG | Second harmonic generation |
| SLS | Static light scattering |
| SONICC | Second order nonlinear imagining of chiral crystals |
| SPR | Surface plasmon resonance |
| SSA | Somatostatin analogue |
| SSTRs | Somatostatin receptors |
| SUMO | Small Ubiquitin-like-modifier |
| TAE | Tris-acetate-EDTA |
| TBE | Tris-borate-EDTA |
| TCDD | 2,3,7,8-tetrachlorodibenzo-*p*-dioxin |
| TCEP | Tris (2-carboxyethyl)phosphine |
| TEMED | N,N,N',-Tetramethylethylenediamine |
| $T_m$ | Melting temperature |
| TPR | Tetratricopeptide |
| TSH | Thyroid stimulating hormone |
| UCR | Upstream conserved regions |
| Vp | Porod volume |
| Wnt | Wingless-related integration site |
| XAP2 | X-associated protein 2 synonym for AIP |
| XRE | Xenobiotic response elements |

## AMINO ACID SIDE CHAIN ABBREVIATIONS

| Group | Donor | Acceptor |
|---|---|---|
| Main chain | NH | CO |
| Aspartic acid (D) | | OD1, OD2 |
| Methionine (M) | | SD |
| Arginine (R) | NE, NH1, NH2 | |
| Glutamine (Q) | NE2 | OE1 |
| Histidine (H) | ND1, ND2 | ND1 |
| Lysine (K) | NZ | |

# Chapter 1

# Literature Review

## 1.1. Introduction to pituitary adenomas

Pituitary adenomas are benign tumours that arise in the anterior pituitary lobe. Up to 15%-25% of the reported intracranial tumours cases are pituitary adenomas, affecting one individual per 1,000 in the general population (Daly *et al*., 2006; Beckers *et al.*, 2013; Agustsson *et al*., 2015). The enclosed and sensitive location of these tumours causes significant morbidities to the afflicted patients. Two-thirds of pituitary adenomas are defined as functional because they result in an increase in production and secretion of endogenous hormones that disrupts normal endocrine function. Functional pituitary tumours are classified according to the type of hormone they over-produce. Prolactinomas over-secrete prolactin (PRL) and cause sexual and reproductive dysfunction, whereas somatotropinomas result in gigantism in children and acromegaly in adults due to the abnormal production of growth hormone (GH). Cushing's syndrome is caused by the over-secretion of the adrenocorticotropic hormone (ACTH) and though less frequent, pituitary tumours that secrete thyroid stimulating hormone (TSH), follicle stimulating hormone (FSH) and/or luteinizing hormone (LH), have also been reported (Heaney and Melmed, 2004). The clinical morbidities observed in patients suffering from non-functional pituitary adenomas (NFPAs) result from the pressure that the growing tumour exerts on the surrounding tissues, particularly on the cranial nerves. This can cause visual problems and headaches. NFPAs are generally diagnosed late because of normal endocrine functions, by which time the tumour would have developed into a macroadenoma (Arafah and Nasrallah, 2001; Greenman and Stern, 2009).

The size and the location of the tumour may hinder its removal by transsphenoidal surgery. Consequently, pituitary tumour patients may require long-term medical treatment that may involve combinational therapy and frequent hormonal screening. When surgery is an option, tumour re-growth may demand multiple surgical interventions, particularly in patients with a genetic predisposition to the disease (Daly *et al.*, 2010).

The prevalence of pituitary adenomas in Malta (76 per 100,000) is comparable to the rates reported in Switzerland (81 per 100,000) and the United Kingdom (77 per 100,000) (Fernandez, Karavitaki and Wass, 2009; Fontana and Gaillard, 2009). In Malta, a standardised incidence rate (SIR) of 4.27/100, 000 per year has been reported. This suggests that in a population of approximately 420,000, 16-17 cases of pituitary tumours are diagnosed annually (Gruppetta, Mercieca and Vassallo, 2013). The mean duration of disease burden has been estimated to be approximately 18 years (Gruppetta, Mercieca and Vassallo, 2013).

### 1.2. Genetic predisposition to pituitary adenomas

Heaney and Melmed (2004) report that almost 94-95% of pituitary tumours cases occur sporadically with the afflicted patients lacking any apparent family history. The rest of the cases however, are familial in which at least two family members are diagnosed with a pituitary tumour. Familial cases are typically characterised by a genetic predisposition caused by a germline mutation that inactivates a tumour suppressor gene (Karhu and Aaltonen, 2007; Daly *et al.*, 2007). Four such genes have been identified in familial pituitary adenomas; *PRKAR1A, MEN1, CDKN1B,* and *AIP* which predispose to the conditions known as Carney Complex (CNC), multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4) and pituitary adenoma predisposition (PAP) respectively. The latter only affects the pituitary gland as opposed to the conditions of CNC, MEN1 and MEN4 that have an effect on more than one endocrine organ (Karhu and Aaltonen, 2007).

### 1.3. *AIP* gene: its role as a tumour suppressor

The cytogenetic location of the aryl hydrocarbon receptor interacting protein (*AIP*) gene is 11q13, which is the same chromosomal location of another tumour suppressor gene, *MEN1*. *MENI* codes for the production of the menin protein, a tumour suppressor protein, whose function is not entirely understood, but has been found to play a significant role in DNA replication, repair and apoptosis (National Library of Medicine, 2019). The *AIP* gene codes for the production of the AIP protein that is a ubiquitously expressed cytoplasmic co-chaperone. AIP is also known as ARA9 and XAP2 (National Library of Medicine, 2019).

The involvement of AIP in functional pituitary adenomas associated with acromegaly was first reported in Finland in 2006, where members from three families, showed a loss of heterozygosity at the 11q13 locus (Vierimaa *et al*., 2006). As none of the patients exhibited symptoms related to the MEN1 condition, this suggested the involvement of a different tumour suppressor gene. Two germline heterozygous mutations in the *AIP* gene were identified; a nonsense mutation, Q14X and an intronic site mutation, IVS3-IG>A, which were present in 16% of the Finish patients suffering from acromegaly. A nonsense mutation (R304X) was also identified that results in a truncated non-functional protein that lacks the last 26 amino acids from the C-terminus of the protein (Vierimaa *et al.*, 2006).

The tumour suppressor role of AIP was initially identified by Leontiou and coworkers (2008) who observed that transfection and over-expression of AIP in rat pituitary cell lines (GH3), HEK293 human fibroblasts and primary human TIG3 (with normal cell cycle regulation), exhibited an anti-proliferative effect in all cell types. The same decrease in cellular proliferation was also observed on primarily pituitary cells, surgically removed from tumours (Formosa, Xuereb-Anastasi and Vassallo, 2013). In both studies, AIP mutants failed to repress cell growth (Leontiou *et al*., 2008; Formosa, Xuereb-Anastasi and Vassallo, 2013). Moreover, most of the reported cases of AIP-positive pituitary tumours are characterised by loss of heterozygosity of the wild type allele. This is in agreement with the Knudson two-hit hypothesis of tumourigenesis (Leontiou *et al.*, 2008; Salvatori *et al.*, 2014). In fact *AIP* heterozygous mice were shown to have a higher predisposition to pituitary tumours, especially somatotropinomas and displayed full penetrance by the age of 15 months (Raitila *et al*., 2010).

Given the involvement of AIP in PAs and the fact that AIP is ubiquitously expressed, the possibility of *AIP* gene mutations in other cancer types was investigated (Georgisti *et al*., 2007). Mutation analyses were carried out on a large cohort of patients (499) who had either breast, colorectal or prostate cancer. A missense mutation (R16H) was identified into two patients with colorectal cancer. Other sequence variations detected in other patients were classified as natural polymorphisms, thus concluding that AIP mutations are uncommon in these cancer types (Georgisti *et al*., 2007). A recent study has suggested that this apparent specificity to the pituitary and the tumour

suppressor function of AIP may be due to interactions occurring specifically in the pituitary gland (Hernández-Ramirez *et al*., 2018).

### 1.4. Tumour suppressor role of AIP: proposed mechanisms of action.

Owing to the vast repertoire of proteins that AIP interacts with, it is unlikely that a single pathway may explain the mechanism by which AIP mutations promote tumourgenesis. It would be more plausible to postulate that AIP inactivation due to germline mutations affects more than one biochemical pathway and that this cumulative effect ultimately predisposes to pituitary adenomas. In fact, although the involvement of AIP in pituitary tumours is well established, the exact molecular mechanisms involved have remained quite elusive and to this day they are not fully understood (Barry *et al*., 2019). It was only recently that putative mechanisms of action have been proposed which provide a better understanding of the role of AIP in AIP-positive tumours.

#### 1.4.1. Cyclic AMP signaling pathway

One pathway that has been closely associated with the mode of action of AIP is the cAMP pathway. Under normal physiological conditions, the pituitary gland is pivotal in controlling cell differentiation, cell growth and hormone production and secretion. The activation of G-protein coupled-receptors (GPCRs) in pituitary cells results in the activation of adenylyl cyclase (AC), which in turn increases the intracellular levels of cyclic AMP (Lania, Mantovani and Spada, 2012). The activity of AC is negatively regulated by a specific type of G-protein, Gα that ensures that the pituitary levels of cAMP do not exceed a certain threshold (Tuominen *et al*., 2014).

Over-expression of AIP was observed to diminish cAMP signalling and GH secretion in GH3 cells treated with forskolin, therefore maintaining low levels of intracellular cAMP. Conversely, knock-down experiments of the endogenous *AIP* increased cAMP signalling even at basal conditions (Formosa, Xuereb-Anastasi and Vassallo, 2013). The same tumour-suppressor role of AIP was supported in a study by Tuominen *et al*., (2014) who showed that AIP deficiency in GH-secreting tumours increases cellular levels of cAMP through a defective $G\alpha_i$-cAMP signalling pathway, also affecting the

phosphorylation action of protein kinases. The enzyme adenylyl cyclase (AC) is responsible for converting ATP into cAMP thus regulating the intracellular cAMP levels. Under normal physiological conditions, the activity of AC is regulated by $G\alpha_i$ proteins. Deficiency of AIP correlates with lower levels of $G\alpha_{i-2}$ and $G\alpha_{i-3}$ proteins with a consequent increase in AC activity and cAMP levels. Through the use of the ingenuity pathway analysis, AIP deficiency was found to have an effect on cAMP signalling pathways, cellular proliferation and immune-inflammatory responses (Tuominen *et al*., 2014).

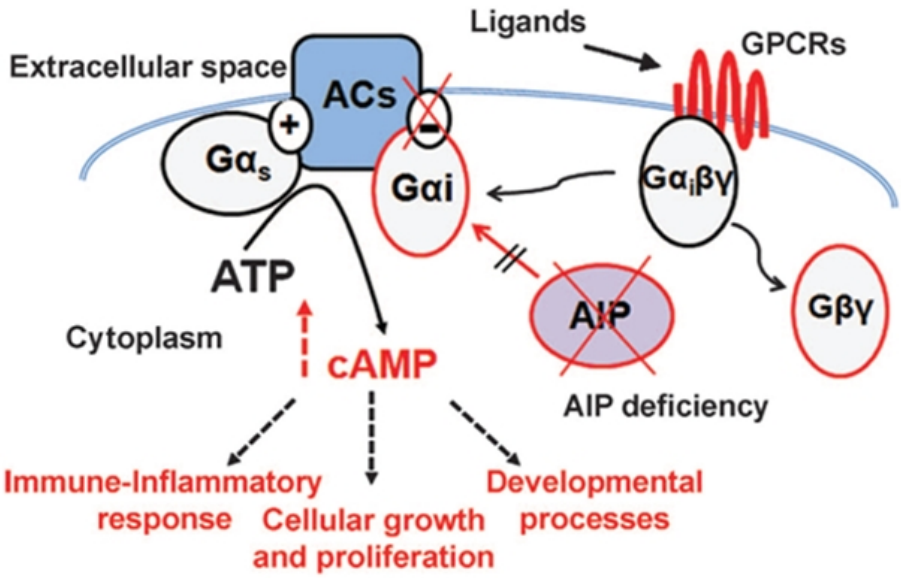



**Figure 1.1: Effect of AIP on $G\alpha_i$-cAMP signalling pathway.** AIP inactivation by mutations and consequent deficiency disrupts the $G\alpha_i$-cAMP signalling pathway, leading to an upregulation of the intracellular cAMP levels (Tuominen *et al*., 2014).

### 1.4.2. Tumour microenvironment

Interestingly, a recent study conducted by Barry *et al*., (2019) demonstrated that the microenvironment surrounding the tumour tissue in AIP-positive tumours is most likely responsible for its invasive potential. Through the use of gene expression profiling, human pituitary tumour samples that had germline *AIP* mutation/s were found to have higher macrophage content when compared to pituitary tumours with functional AIP protein. An upregulation in the production of the cytokine CCL5 was also detected as well as an alteration in the epithelial-to-mesenchymal transition (EMT) pathway.

During this process epithelial cells are transformed into mesenchymal ones and as a result lose their polarity and cell-to-cell adhesion properties but gain an increase in mobility. The same was also observed in tissue obtained from mice that specifically lacked AIP in their pituitary gland. This suggests that the absence of AIP enhances the infiltration of macrophages and invasion in pituitary tumour tissue (Barry *et al*., 2019). Whilst macrophages serve as the first line of defence against disease, their cumulative effect in the cell can promote tumorigenesis and invasion (Ramanthan and Jagannathan, 2014), by secreting enzymes including matrix metalloproteases that degrade the extracellular matrix (ECM) (Ojalvo *et al*., 2010).

### 1.4.3. *ZAC1* co-expression

Somatotropinomas and lactotropinomas which over-produce GH and PRL respectively are characterised by a high expression levels of dopamine and somatostatin receptors (SSTRs). The latter have been the target of a number of developed therapeutic agents to reduce hormonal secretion and tumour growth through somatostatin analogue (SSA) therapy (Heanley and Melmed, 2004). Under normal physiological conditions, SSTRs activate glycogen synthase kinase 3b, which regulates the expression of genes that are involved in cell cycle control. One such gene is *ZAC1* that encodes a tumour suppressor protein highly expressed in anterior part of pituitary gland. Zac1, a zinc-finger transcription factor, is believed to complement AIP in its tumour suppressor role (Chahal *et al*., 2012). The over-expression of AIP in normal pituitary cells was positively correlated with an increase in the mRNA expression of Zac1. This was not the case with mutant AIP, as none of the mutants investigated, upregulated the expression of Zac1 (Chahal *et al*., 2012). This finding suggests that AIP may exhibit its anti-proliferative effect through the co-expression of yet another tumour suppressor protein, Zac1.

A recent study, conducted on three family members from a Chinese family, all of whom harboured a missense *AIP* gene mutation at residue 171 (p.T171I), were found to express less somatostatin receptor type 2 (SSTR2), which is usually targeted in SSA therapy. Lower levels of Zac1 were also detected *(*Cai *et al.,* 2019*).* All this may account, at least partially, for the fact that pituitary adenoma patients harbouring germline *AIP* gene mutations are usually less responsive to SSA treatment.

### 1.4.4. Wnt Pathway

The Wnt pathway (wingless-related integration site) is a conserved signalling pathway that is central for the regulation of cell fate, cellular migration, polarity, the development of organs and stem cell renewal (Reya and Clevers, 2005; Komiya and Habes, 2008; Vlad-Fiegen *et al*., 2012). Considering this pivotal role in cell functioning and normal development, any dysregulation of this pathway can be deleterious. Aberrant Wnt signalling, particularly the canonical (β-catenin dependent) pathway has in fact been implicated in the initiation and/or progression of various types of cancer (Shtutman *et al*., 1999; Polakis, 2000, 2012; Duchartre, Kim and Kahn, 2015). The involvement of the Wnt pathway in pituitary tumours has also been observed (Woloschak and Roberts, 1994; Hibberts *et al*., 1999; Semba *et al*., 2001; Gaston-Massueta *et al*., 2001). A local study showed that activation of the Wnt signalling in pituitary adenomas occurs independently of β-catenin (through a non-canonical pathway) with a significant over-expression of cyclin D1 and MYC-proteins in tumour samples (Formosa *et al*., 2012). The same study also showed that over-expression of AIP reduced the expression of the MYC oncogene, thus suggesting another putative mechanism for the tumour suppressor role of wild type AIP (Formosa *et al*., 2012).

## 1.5. Prevalence of *AIP* mutations

### 1.5.1. Familial isolated pituitary adenoma cohorts

Around 2% of the reported cases of pituitary tumours occur within families, generally described as familial isolated pituitary adenomas (FIPA). The term 'FIPA' has been coined and initially described by Daly *et al*., (2006) as a clinical condition where at least two members of the same family suffer from a pituitary tumour and have no genetic mutations or clinical symptoms of other endocrine syndromes, such as Carney complex or MEN1. Within FIPA, individuals that harbour an *AIP* mutation are grouped into the subset of pituitary adenoma predisposition (PAP). On the other hand, FIPA patients where all the afflicted members have GH-secreting adenomas as the only pituitary tumour type are termed as isolated familial somatotropinoma (IFS). PAP and IFS are two overlapping conditions within FIPA (Figure 1.2).

Various studies demonstrate that *AIP* mutations occur primarily within a familial setting. Leontiou and co-workers (2008) report that 34.6% of the 26 families studied harboured an *AIP* mutation. Similarly Igerja *et al*., (2010) report an *AIP* mutation prevalence rate of 30.6% in FIPA families, of whom 8.3% had a large genomic deletion that was initially missed by direct sequencing but later identified through multiplex ligation-dependent probe amplification (MLPA). A study carried out on Brazilian kindred show a prevalence rate of 33% (Naves *et al*., 2007). A lower rate of 15% was reported by Daly *et al*., (2007) that was carried out on 73 European and America families. However, this study did not make use of MLPA and thus may have missed large genomic deletions. MPLA has in fact been shown to detect large genomic *AIP* deletions in up to 9.5% of pituitary tumour cases, which were not detected by conventional direct sequencing (Igerja *et al*., 2010).

In summary, up to a third of the pituitary adenoma patients within the FIPA kindreds harbour a heterozygous *AIP* mutation. The prevalence rate increases up to 40% when only IFS patients are considered (Daly *et al.*, 2007; Beckers *et al*., 2013). In fact, the occurrence of GH-secreting tumours in two or more family members, particularly if they display an early disease onset, is an almost clear indication of *AIP* gene mutations (Daly *et al.*, 2010; Georgisti *et al.*; 2008; Korbonits, Storr and Kumart, 2012).

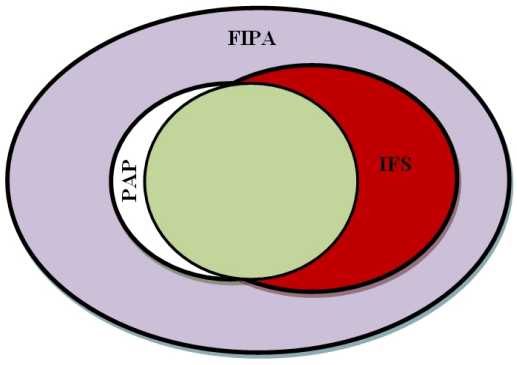



**Figure 1.2: FIPA and its subsets.** PAP (pituitary adenoma predisposition) and IFS (isolated familial somatotropinoma) are two co-existing and overlapping groups. Approximately 30% of patients within the FIPA (familial isolated pituitary adenomas) group harbour an *AIP* gene mutation and fall within the PAP subset. The rate increases to 40% in IFS patients. In addition, approximately 85-90% of PAP patients have a somatotropinoma and are within the IFS subset (Figure reproduced from Llyod and Grossman, 2014)

GH-secreting tumours with/without prolactin co-secretion are the most common tumours types within FIPA cohorts, followed by prolactinomas and NFPAs. While the rank order of frequency remains essentially the same, the presence of GH-secreting tumours increases significantly in FIPA patients who harbour an *AIP* mutation. Although, not shown in Figure 1.3 below, *AIP* gene mutations have also been reported in a few cases of corticotropinomas, characterised by Cushing disease (Cazabet *et al*., 2007; Georgisti *et al*., 2010), however these comprise less than 1% of the total cases. The majority of corticotropic adenomas with positive AIP mutations are sporadic or apparently sporadic patients (Cazabat *et al*., 2012).

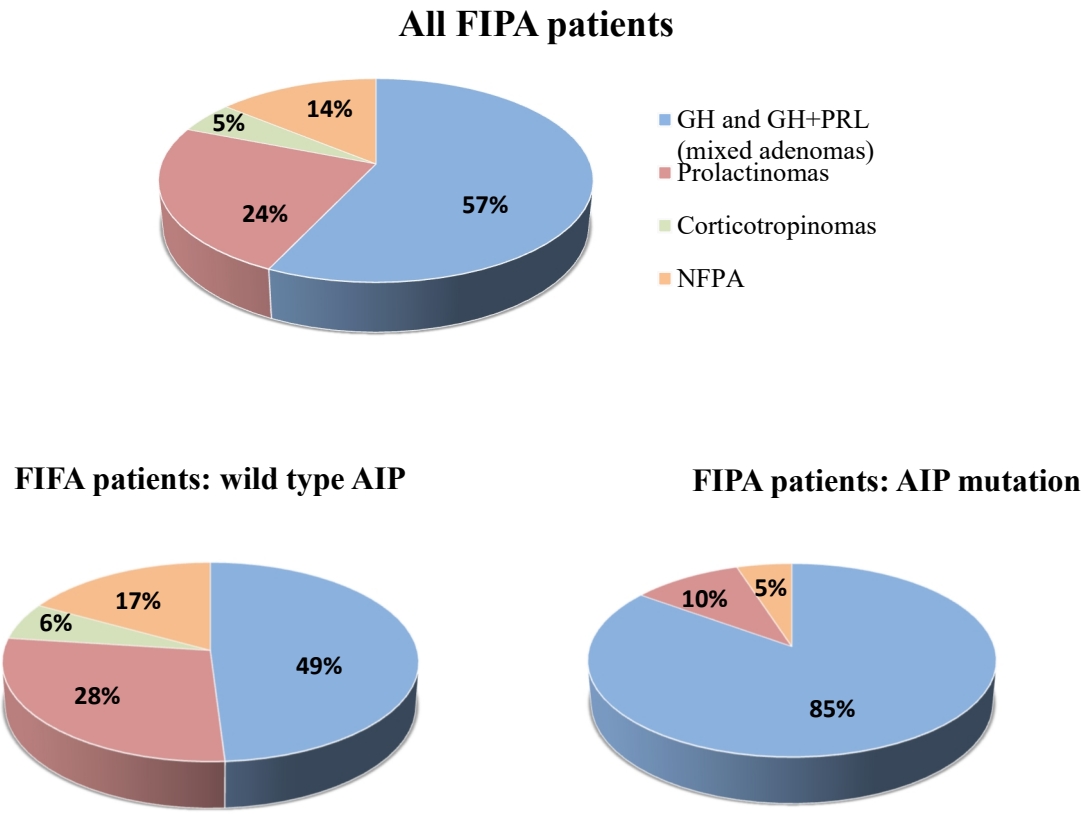



**Figure 1.3:**Pituitary tumours in FIPA cohorts. GH; Growth Hormone, PRL; prolactin, NFPA; non-functional pituitary adenomas. This figure was produced using the data published by Igerja *et al*., 2010; Korbonits, Storr and Kumart, 2012; Llyod and Grossman, 2014)

### 1.5.2. Sporadic pituitary adenoma cohorts

AIP mutations amongst sporadic patients[1] are less common. Though, various independent studies have reported different prevalence values, they are all relatively low. Moreover, the clinical phenotype is usually more varied and the prolactinomas are usually the most common PA type (Arafah and Nasrallah, 2001; Tichomirowa *et al.*, 2011). Leontiou *et al*. (2008) did not identify any *AIP* mutations in sporadic pituitary tumour patients while Barlier *et al*. (2007) recorded a prevalence of only 0.7%. Slightly higher prevalence values of 2% and 3% were reported in geographically dispersed and genetically heterogeneous cohorts (Georgitsi *et al*.,2007; Cazabat *et al*., 2007). A later study conducted by Cazabet and coworkers (2012); who screened a large cohort of 443 apparently sporadic patients, identified a total of 16 *AIP* mutations (3.6%); 6 with acromegaly, 6 with prolactinomas, 3 with corticotripinomas and 1 patient harboured an NFPA. All AIP mutation positive patients were younger than 40 years when diagnosed.

While a general low prevalence of *AIP* mutations has been found when screening unselected sporadic patients, this rate significantly increases when studying selected age-related cohorts. This was shown by Tichomirowa *et al.*, 2011, who carried out a selective screening on a large cohort of 163 sporadic patients. All the studied patients started showing symptoms at a relatively young age (>30 years), had a macroadenoma and no family history of pituitary tumours. Of these, 25% were paediatrics. AIP mutations were reported in 17.2% of the total cohort and in 20.5% of the paediatric sub-group, clearly suggesting that there is a positive correlation between AIP prevalence and an early disease-onset (Tichomirowa *et al.*, 2011).

[1] In some cases, patients are classified as being apparently sporadic due to poor knowledge of family history and as such the possibility of other unknown and/or undiagnosed familial PAs cannot be out-ruled (Llyod and Grossman, 2014).

## 1.6. Clinical characteristics of AIP-mutation positive patients

In general, AIP inactivation is associated with pituitary adenomas that have a more aggressive clinical phenotype. A global study on 96 pituitary tumour patients, both within a familial and sporadic setting, showed that when compared to a control group, patients with a somatropinoma and who also harbour an *AIP* mutation had larger tumours (22.5 mm *vs*. 16 mm) and elevated levels of GH hypersecretion (28.5 ng/mL *vs*. 17.4 ng/mL). In addition, somatostatin analogue (SSA) therapy, which is usually administered to counteract the effect of GH-hypersecretion, was less effective (Daly *et al*., 2010).

Patients exhibit less tumour shrinkage and require additional medication to re-establish endocrine homeostasis, particularly the levels of GH and insulin-like growth factor 1 (Daly *et al.*, 2010; Oriola *et al*., 2013; Salvatori *et al.*, 2014). Reinforcing this is the fact, that gigantism also appears to be more frequent in PA patients with *AIP* mutations (Daly *et al.*, 2010; Leontiou *et al.*, 2008; Korbonits, Storr and Kumart, 2012). In the case of patients suffering from somatotropinoma, this is not surprising, as the patients are usually diagnosed at a young age, at which point the epiphyseal plates are still open (Daly *et al.*, 2010). For PAP, the mean age of disease-onset is reported to be 23.5 years, as opposed to 40 years that is the average age for FIPA patients with wild type AIP (Igerja *et al*., 2010). Another characteristic of AIP-positive tumours is their higher invasive potential (Daly *et al*., 2010; Martucci, Trivellin and Korbonitis, 2012), a clinical aspect which may be explained, at least in part, by the 'unique microenvironment' of the tumour in patients with *AIP* mutations (Barry *et al*., 2019).

## 1.7. Characterisation of *AIP* gene mutations

Till present more than 90 germline *AIP* variants have been identified. These include; nonsense, missense, deletions, intronic splice site, frame-shift insertions, large genomic deletions and promoter region mutations. Almost half of these are either nonsense or frameshift mutations that result in a non-functional protein and are thus termed as pathogenic. *AIP* gene mutations found in conserved regions of the C-domain, in particular within the TPR motifs or the terminal α-7 helical structure, are also clinically relevant as they have an effect on the ability of the protein to interact with its

client partners (Ozfirat and Korbonits, 2010; Beckers *et al*., 2013). Nonsense pathogenic variants in the N-terminal domain have also been reported, with the most frequently encountered being; Q14X (exon 1), E24X (exon 1) R81X (exon 2) and Q142X (exon 3) (FIPA, 2011). The most reported *AIP* gene mutation is the R304X (exon 6), a nonsense pathogenic mutation that has been identified in 35 patients, both within a familial and sporadic setting. The same mutation has also been identified in Charles Byrne, a known giant in medical history who lived in Northern Ireland between 1761 and 1783. His 7 ft 7 inch skeleton is on display at the Hunterian Museum in London. The same mutation was identified in four Irish families, who are believed to share ancestry with Byrne (Chahal *et al.,* 2011).

Most of the studies on the characterisation of *AIP* gene mutations have focused on the ones that are found within the C-domain of the protein. These mutations have to some extent, been prioritised over N-domain mutations, due to the direct role of the TPR motifs in mediating protein-protein interactions. However, the presence of N-domain *AIP* mutations, requires equal consideration. Although, the N-domain might not always be involved in protein interaction/s, it is still essential for the overall stability of the protein (Kazlauskas *et al*., 2002; Linnert *et al*., 2013). This suggests that although an N-terminal variation might not be within a binding site, it may destabilise regions of the protein and thus indirectly affect protein binding. Moreover, while truncating mutations are always pathogenic in nature, the medical implication (if any) of missense mutations requires further characterisation to fully assess their impact on cell function.

A representation of all the *AIP* gene mutations is shown in Figure 1.4. This figure is a modified version of the one published on the FIPA official website (2011). The modifications include the addition of four novel missense mutations that were discovered after the publication of this figure. These are i) R9Q, initially identified in a sporadic Maltese patient in 2009 (Formosa *et al*., 2010, 2012); ii) I13N identified in a 19-year old patient with a large apparently sporadic somatotropinoma (Salvatori *et al*., 2014;); iii) R341, also characterised in a young acromegalic patient (Baciu *et al*., 2013) and iv) T171I, identified in a familial setting (Cai *et al*., 2019). This study will focus on the characterisation of four N-terminal missense mutations; R9Q, R16H, V49M and K103R. A detailed overview of their occurrence in patients is represented in Table 1.1.

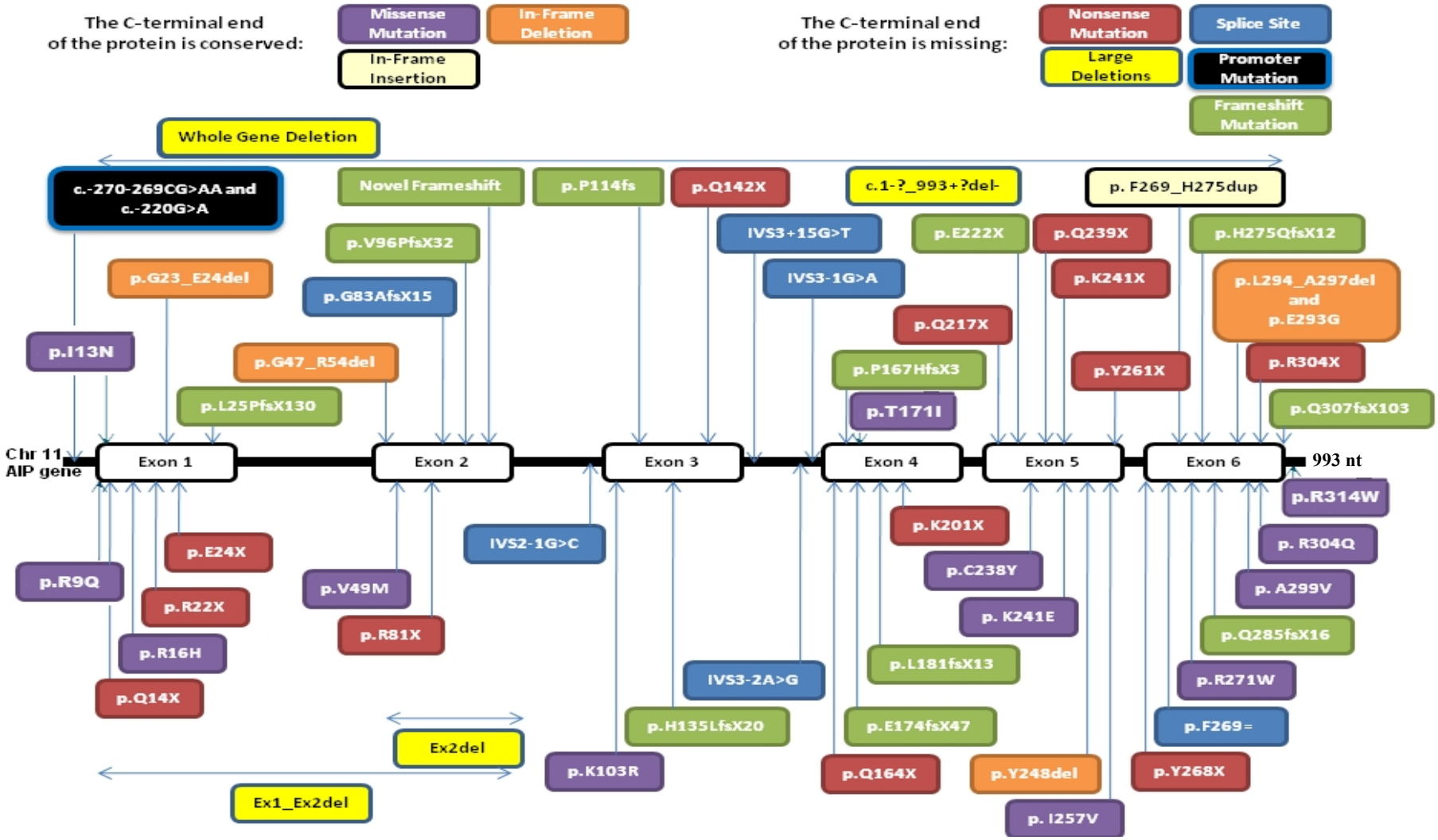



**Figure 1.4: *AIP* gene mutations**. Figure taken from FIPA, c.2011 with modification. Mutation types are colour coded.

**Table 1.1:Occurrence and clinical features of AIP missense mutations; R9Q, R16H,V49M and KI03R.**

| Mutation | Patients identified | Pituitary Adenoma type | Gender | Age at diagnosis | Familial/ Sporadic | References |
|---|---|---|---|---|---|---|
| c.26G>A (p.R9Q) | 4 | GH-secreting | Male | 63 | Sporadic | Formosa *et al*., 2010 |
| | | GH-secreting macroadenoma | Female | 20 | Sporadic | Puig-Domingo *et al*., 2011 |
| | | ACTH-secreting (Cushing's disease) | Female | 39 | Sporadic | Cazabat *et al*., 2012 |
| | | PRL-secreting macroadenoma | Female | 14 | Sporadic | Cazabat *et al*., 2012 |
| c.47G>A (p.R16H) | 7 | GH-secreting microadenoma | N/A | 46 | *Familial | Daly *et al*., 2007 |
| | | GH-secreting microadenoma | N/A | N/A | *Familial | Daly *et al*., 2007 |
| | | Non-functional macroadenoma | Male | 55 | Sporadic | Buchbinder *et al*., 2008 |
| | | Non-functional macroadenoma | Female | 54 | **Familial | Buchbinder *et al*., 2008 |
| | | GH-secreting | Female | 55 | ***Familial | Guaradli and Salvatori, 2011 |
| | | Non-functional | Female | 22 | Familial | Zatelli *et al*., 2013 |
| | | ACTH-secreting macroadenoma | Male | 50 | Sporadic | Dinesen *et al*., 2015 |
| c.145G>A (p.V49M) | 1 | GH-secreting | Male | 28 | Sporadic | Iwata et al., 2007 |
| c.308a>G (p.K103R) | 1 | ACTH-secreting microadenoma (Cushing's disease) | Male | 6 | Sporadic | Beckers *et al*., 2008 |

*The two patients harbouring the R16H mutation who were diagnosed with a GH-secreting microadenoma were first cousins.
**The 28 year old son of the patient also harboured the R16H mutation. However, clinical examination and MRI imagining performed at the time of his mother's diagnosis, showed no sign of pituitary adenomas.
***Patient had a family history of acromegaly.

## 1.8. Aryl Hydrocarbon Receptor Interacting Protein (AIP)

### 1.8.1. Cellular location

AIP is ubiquitously expressed in human tissues with higher levels found in the heart, muscle, brain, cerebellum and kidney (Kuzhandaivelu *et al*., 1996; Leontiou et al., 2008). In the normal pituitary, AIP is localised in lactotrophs and somatotrophs, within the secretory vesicles that release prolactin and growth hormone respectively. However, in sporadic pituitary tumours the expression of AIP was detected in all pituitary cell types and is not only limited to GH- and PRL-secreting cells. In the case of somatotropinomas, AIP is found in the secretory vesicles, as also observed in normal somatotrophs, whilst in prolactinomas AIP is distributed within the cytoplasm (Leontiou *et al*., 2008; Jaffrain *et al.*, 2009). Although, not present in gonadotroph and corticotroph cells under normal physiological conditions, AIP is found to be relatively abundant in corticotropinomas, although not within the secretary vesicles. Similarly, AIP is present in substantial amounts in NFPAs, localised within the cytoplasm (Leontiou *et al*., 2008). Whilst some pituitary tumour cells types exhibit an abnormal AIP expression with changes in its sub-cellular location, the presence of AIP in pituitary tumours harbouring *AIP* mutations varies depending on the type of mutation. For instance, nonsense mutations give rise to a truncated protein that is rapidly degraded (Georgisti *et al*., 2007) thus resulting in AIP deficiency in pituitary cell types.

Previous studies, conducted by Carver and Bradfield (1997), have shown that that the levels of AIP mRNA are elevated in the heart tissue, suggesting that AIP is possibly also involved in the early stages of heart development. Supporting this is the fact that *AIP* knockout mice died during embryonic development as a result of limited blood flow to the limbs and heart, the latter resulting in cardiovascular defects (Lin *et al.*, 2007).

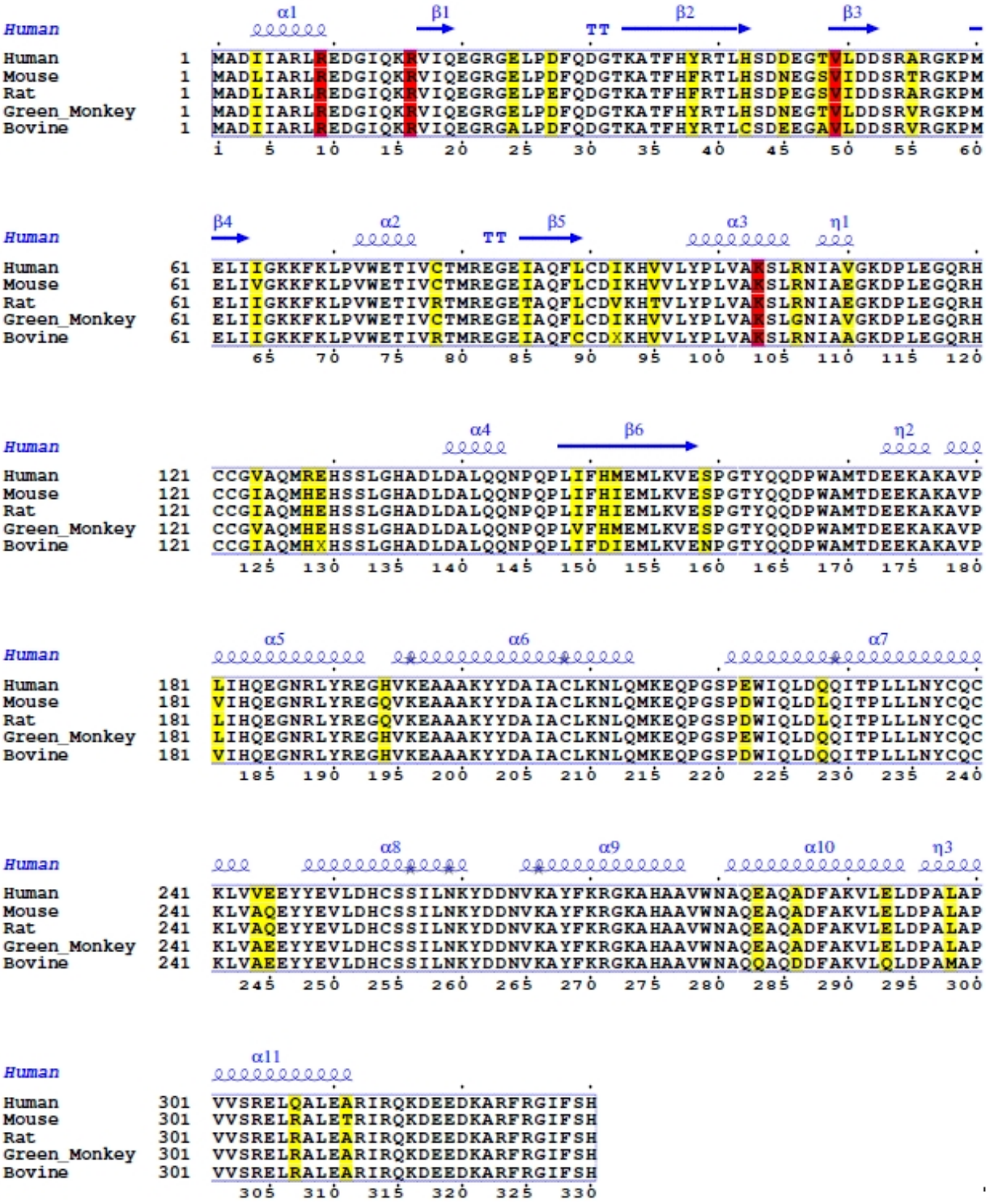



**Figure 1.5: Alignment of selected eukaryotic AIP amino acid sequences.** Amino acid variations are highlighted by a yellow box. The red box indicates the amino acid residues that have been investigated in this study. These residues are conserved amongst species. Sequences were obtained from the UniProt Database, 2019. The alignment was generated using the online server ESPript version 3.0 (Robert and Gouet, 2014).

### 1.8.2. Protein function

The role of AIP in the normal pituitary and tumour cells is not fully understood. However, being a molecular co-chaperone, it mediates a number of interactions, either directly or indirectly by acting as a scaffolding protein, with various cytoplasmic and nuclear proteins. Through these interactions, AIP is likely to regulate cell growth, cell division and ensure cell survival, as well as exert its role as a tumour suppressor (Carver and Bradfield, 1997; Cai *et al*., 2011). The *AIP* gene is highly conserved amongst vertebrates, as illustrated in Figure 1.5. In addition, the vital function of the protein is further reinforced by the fact that none of the PAP patients known harbour a homozygous *AIP* gene mutation.

### 1.8.3. Protein structure

AIP is a monomeric protein consisting of 330 amino acids having a molecular weight of 37.5 kDa. It shares sequence homology with two large protein families; FK506-binding proteins (FKBPs) and tetratricopeptide repeat (TPR)-containing proteins (Linnert *et al.*, 2013). FKBPs are a conserved group of proteins known for their ability to bind to immunosuppressant drugs such as FK506. These immunophilins function as chaperones and are involved in several cellular processes that require protein folding. In addition, most display peptidyl-prolyl *cis-trans* isomerase (PPIase) activity (Kang *et al*., 2008). TPR motifs are characterised by a high helical content. They are commonly found in proteins to mediate protein-protein interactions and the formation of multi-protein complexes. The consensus TPR is composed of 34 amino acids. TPR-containing proteins may have three to sixteen TPRs however the folding pattern and packaging differs amongst TPR-containing proteins (D'Andrea and Regan, 2003). The C-domain of AIP contains three TPRs.

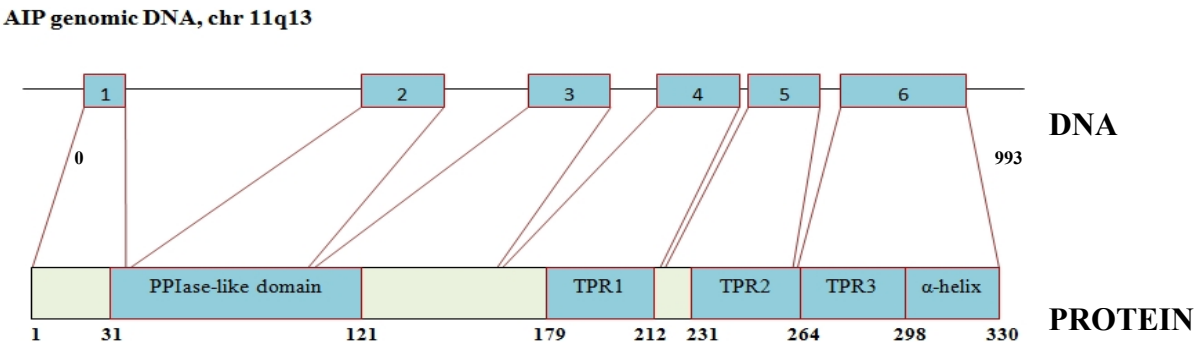



**Figure 1.6: Schematic diagram of AIP**. The *AIP* gene (top) is 993 nucleotides in length. It consists of 6 exons that code for a protein with 330 amino acid residues (bottom). Exons are shown as blue boxes and the numbering of amino acids is shown below the protein schematic. (Amino acid numbering is reproduced from The UniProt Consortium (2019).

Structural information is available on the two separate domains of AIP. The N-domain has been studied through nuclear magnetic resonance (NMR) (Linnert *et al.*, 2013) whilst the crystal structure of the C-domain has been solved to 2 Å resolution (Morgan *et al*., 2012). However, till present there is no structural data available of the complete full-length protein.

### 1.8.3.1. The N-domain

The N-domain of AIP is comprised of 167 amino acid residues. This domain shares 30% sequence homology to FKBP52 and 25% homology to another immunophilin, FKBP12. Similar to peptidylprolyl isomerase (PPIase) domains, this region of AIP consists of four α-helices and five anti-parallel β-sheets forming a half beta-barrel around a central α-helix (Figure 1.7). However, despite structural similarity, AIP does not function as an immunophilin, it lacks the ability to bind to the common immunosuppressant, FKB506 and exhibits no PPIase activity (Carver *et al*., 1998, Laenger *et al*., 2009). Furthermore, NMR spectroscopy revealed two structural features that have not been reported in any other FKBP protein and only pertain to AIP. These are an N-terminal helix referred to as α0 and a long insert designated as '*βD-βE* extension' (Linnert *et al.*, 2012, 2013).

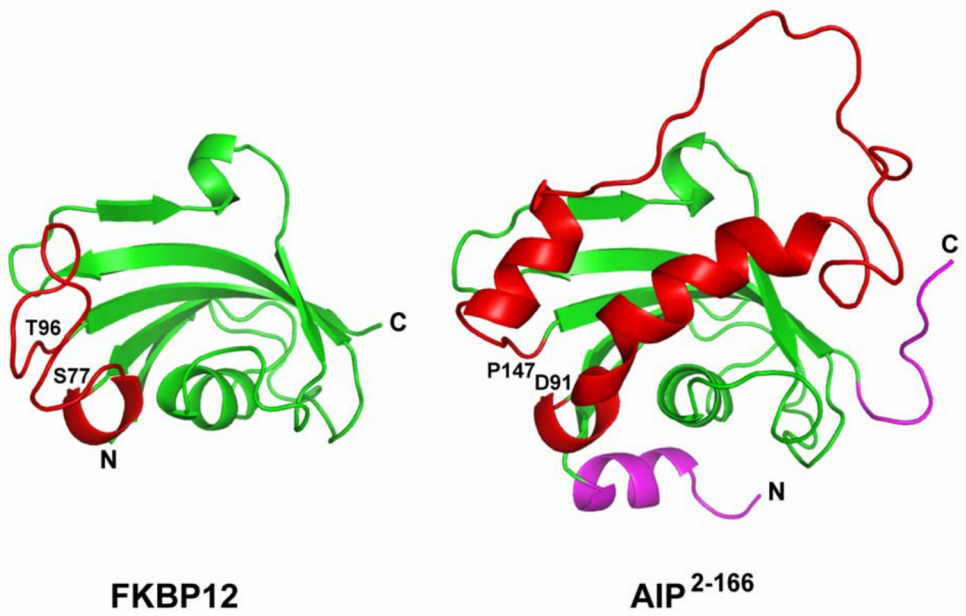



**Figure 1.7: FKBP12 and AIP$^{2-166}$ structural comparison**. Structurally, AIP displays two unusual and unique structural elements, an N-terminal helix (α0) and a long insert ('*βD-βE* extension'), which are displayed in magenta and red respectively (Linnert *et al*., 2013).

The N-terminal helix (α0) is 8 residues in length, from I4 to N11 and lies almost parallel to the first β-strand. As demonstrated by Linnert *et al*. (2012), this structural motif is crucial for protein stability, as its removal resulted in significant aggregation and degradation during purification. The *βD-βE* extension spans a total length of 57 residues and connects the two strands of the barrel, designated as *βD* and *βE,* hence its name. It is composed of three major segments; a long helical region comprised of 19 residues (αIII I92-V110), a random coil with no defined secondary structure (G111-H135) and a terminal atypical α-helical segment, designated as αIV (A136-Q143).

Similar FKBP-proteins have the PPIase active site in the same region where the αIV and αIII helical structures are localised in AIP. The absence of PPIase activity in AIP is thus partially attributed to the fact that these two unusually long helical structures are covering the putative active site (Linnert *et al.*, 2012, 2013).

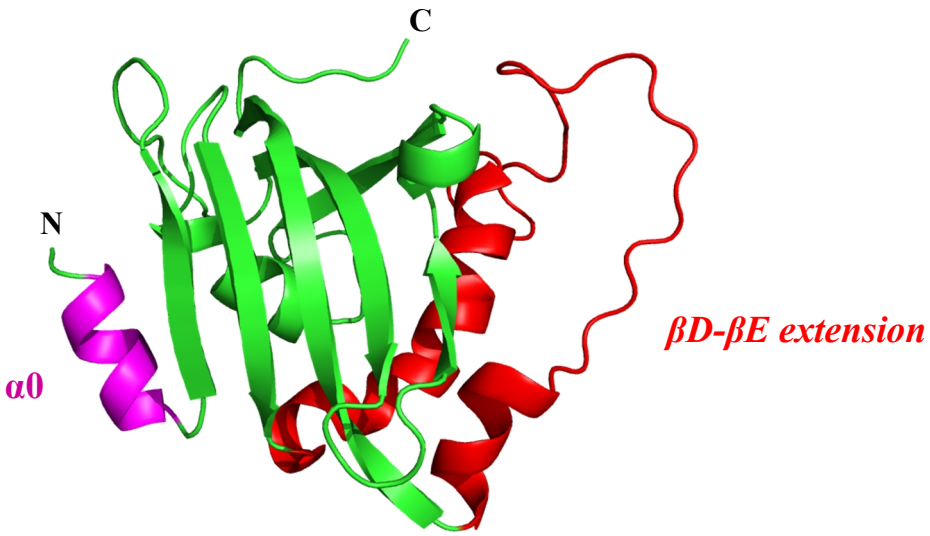



**Figure 1.8: Cartoon representation of the solution structure of human AIP$^{2-166}$.** The green structure represents the typical FKPB-type domain, whereas the structures shown in red and magenta correspond to the N-terminal α0 and *β*D-*β*E extension respectively (PDB ID: 2LKN, Linnert *et al*., 2013). Figure `was reproduced using PyMOL version 2.3.0 (Schrödinger, 2015).

### 1.8.3.2. The C-domain

The C-domain of AIP is comprised of 151 amino acid residues. It consists of three TPR motifs followed by a terminal α-7 helix. Each individual TPR is 34 amino acids in length and consists of two anti-parallel α-helices (Figure 1.9). The terminal α-7 helical segment (residues 301-330) is a highly conserved sequence of hydrophobic and positively charged amino acids. Aside from enhancing protein stability and solubility, this structural feature is also apparently crucial for mediating protein-protein interactions. The last five amino acids, GIFSH (G326-H330), are essential for AIP to interact with its binding partner, aryl hydrocarbon receptor (AhR). *AIP* gene mutations identified in this region resulted in a complete loss of function (Bell and Poland, 2000; Daly *et al*., 2010). These amino acids however, do not appear to directly participate in the interaction with peptides of Hsp90 or Tomm20 that were co-crystallised with the C-domain of AIP (Morgan *et al*., 2012). The AIP C-domain (residues 172-315) has been co-crystallised with short peptide fragments of Hsp90 (SRMEEVD), PDB ID: 4AIF, and Tomm 20 (AEDDVE) (PDB ID: 4APO, to resolutions of 2.0 Å and 1.9 Å respectively) (Morgan *et al*., 2012).

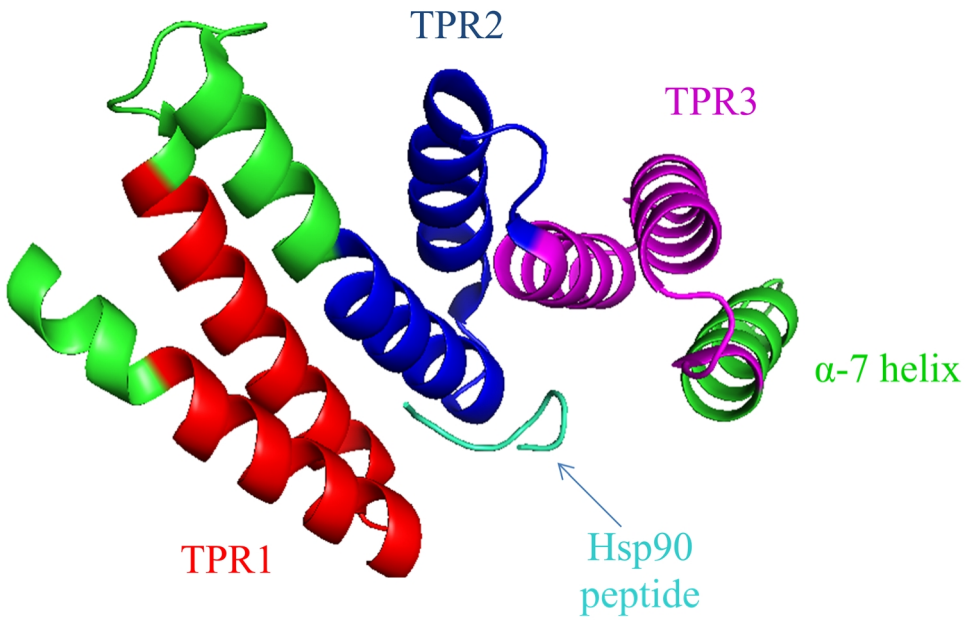



**Figure 1.9: X-ray crystallography structure of the C-domain of human AIP$^{172-315}$.** The three TPR domains are represented in red, blue and magenta respectively. The terminal α-7 helix is shown in green. The domain is shown in association with a synthetic peptide fragment derived from human Hsp90β (SRMEEVD), shown in cyan. PDB ID: 4AIF (Morgan *et al*., 2012).The figure was generated using PyMOL version 2.3.0 (Schrödinger, 2015).

Although the N-and C-domains have been characterised individually, the structural relationship between the two domains is currently unknown. Furthermore, structural data for residues 167-171 and 324-330 is absent. The former represents part of the inter-domain linker, whilst the latter, form an important part of the α-7 helical segment, required for binding to some client proteins.

## 1.9. The AIP protein interactome

Though AIP is inactive as a chaperone (Li *et al*., 2013), it functions as a co-chaperone to other chaperones and cellular proteins, therefore resulting in a vast protein interactome (Figure 1.10). It is best known for its interaction with AhR that responds to exogenous ligands. Aside, from xenobiotic-metabolising pathways (induction of cytochrome P450 CYP1A1 and CYP1B1 by activated AhR), AIP also plays a role in the regulation of cellular cAMP levels, oestrogen receptor signalling and cell survival, through its interaction with phosphodiesterases, Gα-proteins, the oestrogen receptor and RET and survivin respectively (Trivellin and Korbonits, 2011). The signalling and biochemical processes which require AIP as a co-chaperone are indicated in Figure 1.10.

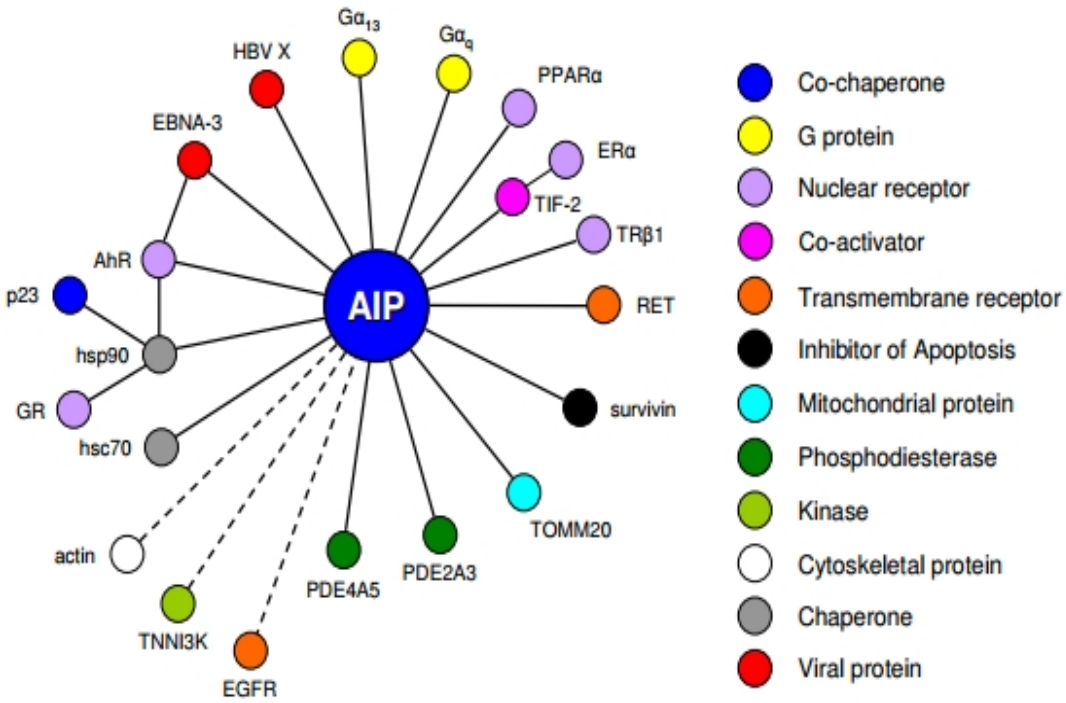



**Figure 1.10: The interacting partners of AIP, c2011.** Solid lines represent a direct interaction, while putative interactions are represented by dashed lines. Interacting partners are grouped and colour coded (Trivellin and Korbonits, 2011).

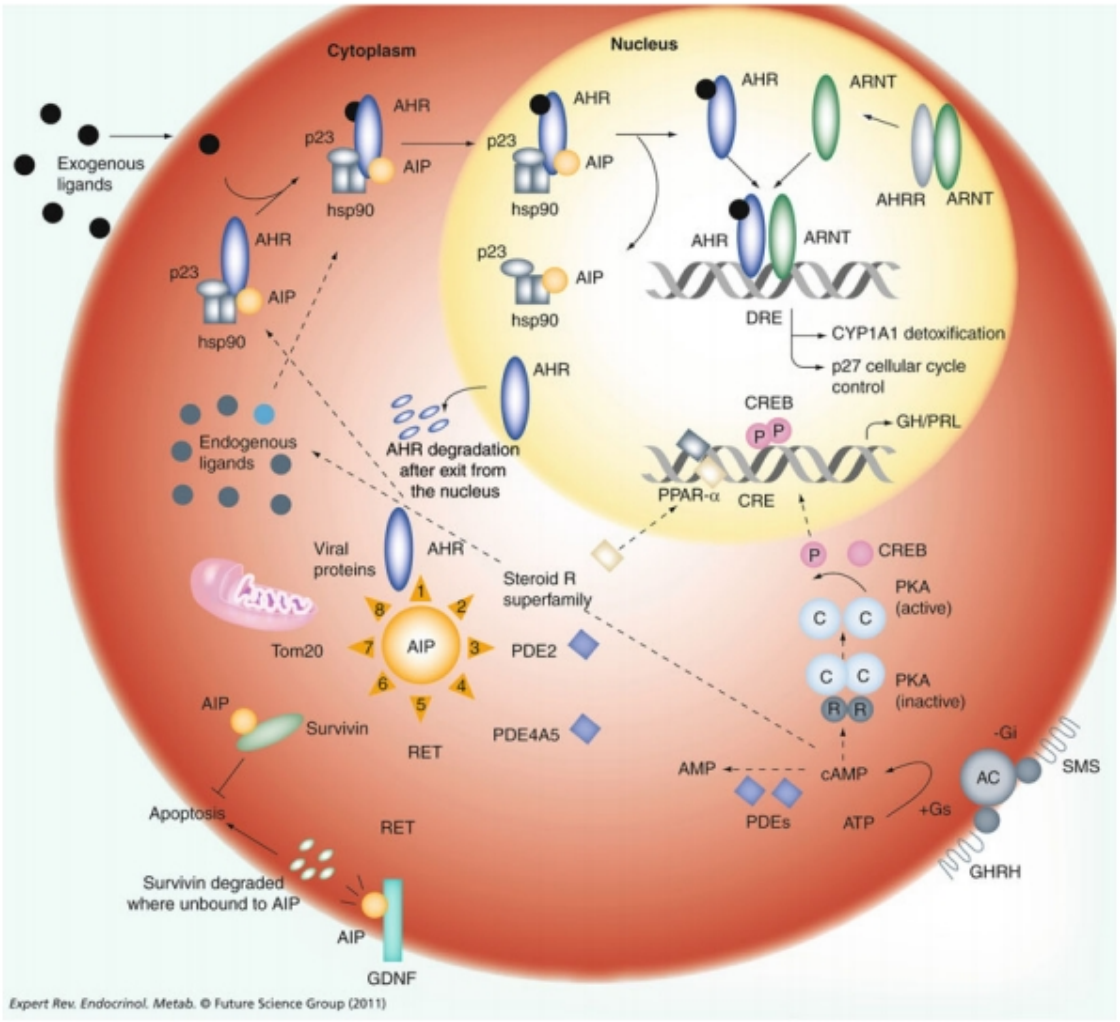



**Figure 1.11: AIP interactome and related pathways.** AIP interacts with (1) the transcription factor AhR (Aryl hydrocarbon receptor) involved in xenobiotic metabolism, (2) members of the steroid receptor superfamily including the transcription factor PPARa, the thyroid receptor and the glucocorticoid receptor, (3,4) phosphodiesterases PDE2 and PDE4A5 that regulate the cellular levels of cAMP (5) the RET proto-oncogene, (6) survivin, the AIP-survivin interaction regulates the stability of survivin, resulting in an increase in the anti-apoptotic threshold of the cell (7) Tom20, has role in mitochondrial import and (8) viral proteins that include the Epstein Barr Virus and hepatitis B virus X antigen (Jaffrain and Beckers, 2013).

### 1.9.1. Hsp90

Hsp90 is an ATP-dependent chaperone that assists in the correct folding and maturation of specific proteins, particularly those involved in cell growth and normal cell functioning, including kinases, transcription factors and steroid receptors. The mutant form of the p53 tumour suppressor protein as well as the breast cancer associated HER2 protein are also maintained in a stable active form by Hsp90, which has thus become an additional target of certain chemotherapeutic drugs (Goodsell, 2008; Pearl, 2016). At a molecular level it exists as a homodimer. Each monomeric component is comprised of three domains, an N-terminal domain with an ATP-binding site, a middle domain and a C-terminal domain that mediates protein dimerisation (Pearl, 2016). Physiologically, Hps90 normally occurs as part of multi-protein complex, as exemplified by the cytoplasmic complexes of AIP/AhR/Hsp90 and AIP/Hsp90/Glucocortiocoid receptor (GR) (Trivellin and Korbonits, 2011).

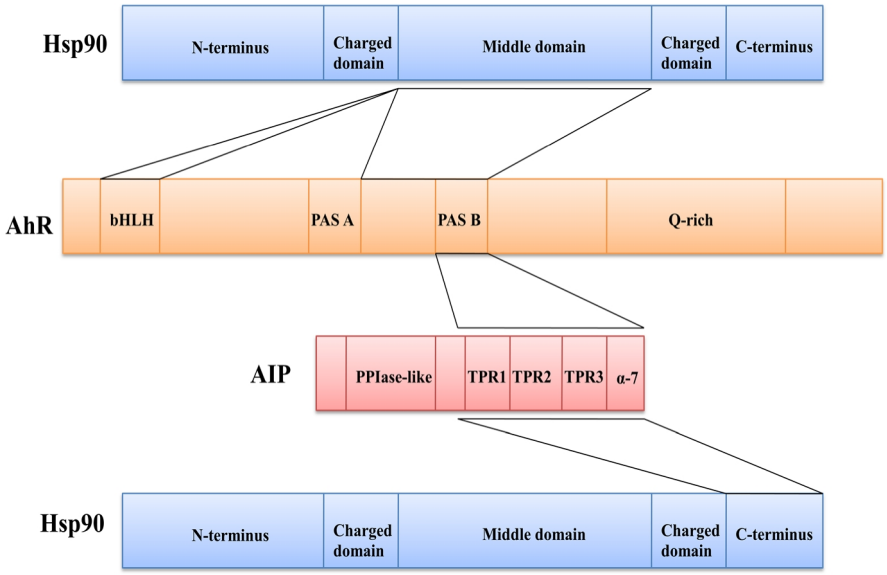



**Figure 1.12: Interacting domains of the AIP-Hsp90-AhR complex**. Within the cell, AIP is found in complex with two molecules of Hsp90, AhR and p23 (the latter not shown in figure). The interacting domains of each protein in the complex are indicated by solid lines. The size of each domain is drawn to proportion with respect to the number of amino acid involved. bHLH: basic-helix-loop helix; PAS: Per-ARNT-Sim homology domain and Q-glutamine rich region. Figure was reproduced from Trivellin and Korbonitis, 2011.

AIP and Hsp90 have been shown to interact, even in the absence of ATP, both *in vitro* and *in vivo.* The interaction occurs through hydrogen bonding and was observed both in the presence and absence of AhR (Carver and Bradfield, 1997; Bell and Poland, 2000 and Li *et al.*, 2013).

Various studies have reported that the N-domain of AIP is not involved in the interaction with Hsp90 and AhR and that this interaction is mediated through the TPR motifs of the protein (Carver *et al*., 1998; Meyer and Perdew, 1999, Kazlauskas *et al*., 2000), as also represented in Figure 1.12. The N-domain was believed to be solely responsible for conferring protein stability and regulating the location of AhR within the cell (Kazlauskas *et al*., 2002). This was however contradicted by Linnert and co-workers (2013), who identified additional Hsp90 binding contacts within the PPIase domain that necessitate the *βD-βE* extension loop. None of the other proteins within the FKBP family showed an interaction with Hsp90, being uniquely observed in AIP. The same group showed that AIP interacts with Hsp90 in pull-down assays both when using the full length AIP protein and the N-domain alone. Though the interaction between Hsp90 and the N-domain of AIP was relatively weak and could not be measured by isothermal titration calorimetry (Linnert *et al*., 2013), this novel discovery, suggests that the N-domain has more functional roles, in addition to enhancing protein stability. This adds to the relevance of investigating the effect of N-domain mutations on the co-chaperone function of AIP.

### 1.9.2. Phosphodiesterases

Phosphodiesterases (PDEs) are conserved enzymes responsible for the degradation of cyclic nucleotides, such as cyclic adenosine 3’5’-cylic monophosphate (cAMP) and guanosine 3’5’-cyclic monophosphate (cGMP), thereby regulating their abundance in the cell. Cyclic AMP is a universal regulator molecule that is pivotal for a number of cellular functions in response to activated GPCRs. Within the pituitary gland, the levels of cAMP are the key to regulating the synthesis and secretion of hormones (Kits and Mansvelder, 2000; Peverelli *et al*., 2013). Dysregulation of the cAMP-pathways is associated with pituitary tumourigenesis and response to treatment (Lania, Mantovani and Spada, 2012; Hernández-Ramirez, Trivellin and Stratakis, 2017).

Human PDEs are classified into 11 families that are encoded for by 24 genes, in total producing more than 100 different isoforms that vary in their cellular distribution, kinetics, substrate specificity and susceptibility to inhibitors (Francis, Blount and Corbin, 2011; Bizzi *et al*., 2019). All PDE isoforms share a conserved catalytic domain that is 300 amino acids in length and most isoforms have regulatory regions in the N-domain that are family-specific (Conti and Beavo, 2007). Through mRNA studies, the PDE1A, PDE2A, PDE4 (A, B, C, D) PDE8B and PDE11A are the most highly expressed isoforms of PDEs in the normal pituitary gland (Michibata *et al*., 2001, Persani *et al*., 2001; Mackenzie *et al*., 2008; Peverelli *et al*., 2009). The expression of these isoforms was found to increase in GH-secreting pituitary tumours. Since PDEs are the only enzymes that are able to hydrolyze cAMP to its inactive 5'-AMP form, this up-regulation in PDE expression observed in adenomas is possibly a compensatory mechanism to tumourigenesis (Bolger *et al*., 2016). However, AIP mutations were found to interfere with the expression of these PDEs (Bizzi *et al*., 2018, 2019).

Compared to other PDE families, the PDE4A and PDE2A families have a close association with AIP (Bolger *et al*., 2003; de Oliveria *et al.*, 2007). In mammals, PDE4 isoforms are encoded by 4 different genes, (PDE4A/B/C/D). In contrast to other PDE families, PDE4s have specific catalytic regions (Conti *et al*., 2003; Houslay and Adams, 2003) and two 'signature' regions that are referred to as upstream conserved regions (UCRs), located in the N-domain (Beard *et al*., 2000). The interaction between AIP and the rat isoform PDE4A5 (that is analogue to the human PDE4A4) has been studied and is known to be inhibitory for the enzyme. Aside from attenuating PDE4A5 catalytic activity, AIP also decreased the ability of the enzyme to be phosphorylated by protein kinase A (PKA). Through pull-down assays, the AIP-PDE4A5 interaction was shown to require the TPR domains of AIP (Bolger *et al*., 2003).

AIP also interacts with PDE2A, an interaction that is mediated by the C- domain of AIP and a central regulatory domain in PDE2A (GAF-B). This interaction does not seem to alter the catalytic activity of the enzyme however it does regulate AhR mobility and its ability to translocate to the nucleus when it is activated by exogenous ligands (de Oliveria, *et al*., 2007).

### 1.9.3. Aryl hydrocarbon receptor

#### 1.9.3.1. AhR function

Aryl hydrocarbon receptor (AhR) is a ligand-activated transcription factor that is activated by exogenous and endogenous ligands. The former are generally dioxins and other environmental pollutants that diffuse into the cell owing to their lipophilic nature.

AhR is ubiquitously expressed with higher levels detected in the lungs, liver, spleen, kidney, placenta and the skin (Döhr *et al*., 1996; Yamamoto *et al*., 2004). In the absence of ligands, AIP regulates the localisation, ligand receptivity and stability of AhR (Kazlauskas *et al*., 2000). AhR resides in the cytoplasm as a complex with AIP, two molecules of Hsp90 (dimer), p23 and the proto-oncogene tyrosine-protein kinase SRC. The best characterised function of AIP is in fact that of stabilising AhR in the cytoplasm. It exerts its stabilising role through two mechanisms; i) by inhibiting AhR from interacting with importin-β which would otherwise translocate it to the nucleus (Petrulis *et al*., 2003; Ramadoss *et al*., 2004) and ii) by protecting AhR from degradation via the ubiquitin-proteasomal pathway (Kazlauskas *et al.* 2000; Petrulis and Perdew, 2002). Low cellular levels of AIP correlate with low levels of AhR as observed in the majority of pituitary tumours (Jaffrain-Rea *et al.*, 2009; Nukaya *et al*., 2010). The AhR complex is further stabilised by the Hsp90 and p23 binding clients. Similar to AIP, the co-chaperone p23 prevents AhR degradation and maintains the protein in a stable conformation that is receptive to ligand binding (Kazlauskas *et al*. 2001; Kudo *et al*., 2018).

When activated by ligands, AIP dissociates from the rest of the complex, exposing parts of the N-terminal domain of AhR that signal for its translocation to the nucleus. This "nucleocytoplasmic shuttling" process is mediated through an importin-β-dependent pathway (Ikuta *et al*., 2000), although a later study highlighted the involvement of a cAMP-dependent pathway (de Oliviera *et al*., 2007). Within the nucleus, AhR binds to the aryl hydrocarbon receptor nuclear translocator (ARNT), forming an active heterodimer. The AhR-ARNT complex binds to xenobiotic response elements (XRE) (also known as dioxin-response elements, DRE) onto the DNA and activates the transcription of metabolising enzymes, such as CYP1A1 and CYP1B (Murray Petterson and Perdew, 2014). The activity of AhR is regulated by the AhR-

repressor (AhRR) that competitively binds to ARNT (forming an AhRR-ARNT complex) and/or XRE in preference to AhR (Hahn, Allan and Sherr, 2009). Activated AhR that fails to bind to ARNT is unstable and targeted for degradation by the ubiquitin-proteasome pathway (Ma and Baldwin, 2000). These mechanisms exert control on the activated receptor and prevent over-simulation by AhR agonists.

It is still unclear whether the other chaperones (Hsp90 and p23) of the AhR complex dissociate from it prior to translocation or within the nucleus. Studies performed on mice showed that activated AhR dissociates from all of its chaperones and co-chaperones within the cytoplasm (Ma and Whitlock, 1997), whereas recent studies in human tissue suggest that Hsp90 translocates to the nucleus together with AhR when the latter is activated by xenobiotic ligands, such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (Beckers *et al*., 2013; Tsuji *et al.*, 2014).

The first documented exogenous ligands that activate AhR were synthetic halogenatic aromatic hydrocarbons such as polychlorobiphenyls, polychlorinateddibenzo-para dixoins and polycyclic aromatic hydrocarbons, with the most commonly studied dioxin being TCDD. These molecules are present in the air and fatty foods and are relatively stable molecules, posing an addition health hazard due to their ability to accumulate within the body (Denison, Fisher and Whitlock, 1988). Natural ligands such as indole derivatives and flavonoids such as quercetin and resveratrol have also been classified as AhR ligands that can act as both agonists and antagonists (Ciolino, Daschner and Yeh, 1999; Guyot *et al*., 2013). Recently, endogenous ligands have also been identified, namely arachidonic acid metabolites, metabolites of the kynurenine pathway and kynurenine itself (Nguyen and Bradfield, 2008; Seok *et al*., 2018).

Aside from its role in xenobiotic metabolism, AhR is essential for other signalling pathways that are critical to the normal cell physiology, including cell proliferation, differentiation and cell motility. It is involved in varies other physiological roles, as exemplified by AhR knock-out (KO) mice that show a significant reduction in survival rate, while those that survived had an impaired immune system and produced less lymphocytes in their lymph nodes and spleen. In addition, their liver was reduced in size by almost half and their bile ducts exhibited fibrosis (Fernandez-Salguero *et al*., 1995). Similarly, in other independent studies, AhR-deficient mice showed hepatic defects

(Schmidt *et al*., 1996); fertility problems (Abbott *et al*., 1999) and vascular abnormalities (Fernandez-Salguero, Ward, Sundberg and Gonzalez, 1999; Lahvis *et al*., 2005). In addition, KO-mice models also confirmed the role of AhR in cell adhesion and mobility (Mulero-Navarro *et al*., 2005).

AhR has also been found in the soft-shell clam, *Mya arenaria* (Butler *et al*., 2001) and *Caenorhabditis elegans* (Qin and Powell-Coffman, 2004) although in both cases, AhR failed to bind to TCDD and beta-napthoflavone, two well-known agonists of AhR. No ligand has yet been identified that activates AhR in invertebrates (Larigot *et al*. 2018) clearly suggesting that the role of AhR within the cell goes beyond toxin detoxification and that this function, though the most characterised in vertebrates, has mostly likely been acquired throughout the course of evolution.

#### 1.9.3.2. AhR structure

AhR is a multiple domain protein that forms part of the bHLH-PAS family (basic helix-loop-helix- Per-ARNT-Sim). At the N-terminal there is a characteristic bHLH region that is responsible for the protein's ability to dimerise with ARNT once inside the nucleus (Seok *et al*, 2017). The binding of AhR to DNA and to Hsp90 is also mediated through this domain that also directs its translocations to the nucleus (Whitelaw *et al*., 1995). The PAS domain is composed of two structural repeats, PAS A and PAS B respectively; the former is involved in dimerisation with ARNT whilst the latter is the site of ligand binding (Fukunaga *et al*.,1995; Chapman-Smith, Lutwyche and Whitelaw, 2004.). The C-domain comprises three sub-domains, an acidic region that is characterised by a high proportion of glutamate and aspartate residues, a Q rich which is enriched with glutamine and a final sub-domain comprised mostly of proline, serine and threonine. This C-domain is important in regulating the function of AhR as it encompasses the binding site of AhR co-activators and co-repressors, such as the AHR-repressor (AHRR) (Jain, Dolwick, Schmidt and Bradfield, 1994; Ko, Okino, Ma and Whitlock, 1997; Baba *et al*., 2001).


**Figure 1.13: Schematic diagram of AhR.** The amino acid numbering of domains is reproduced from The UniProt Consortium (2019).

At present, there is no published method that yields milligram quantities of stable native AhR and recent studies have acknowledged this fact by commenting on the difficulty in purifying the full length protein (Tsuji *et al*., 2014; Schulte *et al*., 2017; Sasaki-Kudoh, 2018). As such, *in vitro* work has been centred on the purification of domains, particularly the bHLH and PAS domains to investigate the ability of the protein to dimerise with ARNT (Seok *et al*., 2016), its ligand binding affinity and association to Hsp90 (Tsuji *et al*., 2014). AhR encoding amino acid residues 23-273, thus including both the bHLH and PAS A domains, has been crystallised in complex with a 12mer dsDNA, to a resolution of 3.3 Å, PDB ID: 5NJ8 (Schulte *et al*., 2017).

### 1.9.3.3. Involvement of AhR in cancer pathology

AhR has an important role in the major stages of tumourigenesis- initiation, promotion, progression, and metastasis (Murray, Petterson and Perdew, 2014). Nonetheless, there are conflicting studies on the role of AhR in cancer pathology. Various studies demonstrate that AhR favours cell proliferation, thus promoting cancer progression (Moennikes *et al*., 2004; Feng, Cao and Wang, 2013) while other studies have shown AhR to have anti-proliferative effects and tumour suppressive roles (Schmidt *et al*., 1996; Kolluri *et al*., 1999; Puga *et al*., 2000; Fan *et al*., 2010). A recent review has clearly illustrated that the role of AhR in cell proliferation varies and is most likely dependent on the cell type being investigated, its cell cycle stage and the developmental period of animals, when these are used as models for the experiment (Larigot *et al*., 2018).

AhR has been found to be a potential drug target for treating oestrogen receptor (ER) negative breast cancer patients (Jin, Lee, Pfent and Safe, 2014). High-levels of

AhR have been found in breast cancers (Vacher *et al*., 2018), while down-regulation of AhR has been reported in cases of the acute lymphocytic leukaemia (Jaffrain-Rea and Beckers, 2013). A recent study conducted by Stanford *et al*. (2016) suggests that AhR has a significant role in the development of cells with "cancer stem-like qualities" that might instigate different types of cancer including breast cancer. This effect is enhanced in the presence of exogenous AhR ligands such as dioxins and other polyaromatic hydrocarbons which are thus likely to exacerbate the progression of breast cancer (Stanford *et al*. 2016). AhR was also found to act as a potential tumour suppressor in liver cancer (Fan *et al*., 2010) melanomas (Troca *et al*., 2013) and pituitary adenomas (Formosa, Borg and Vassallo, 2017). A direct interaction between AhR and the tumour suppressor p53 has also been reported (Su, Lin, Wang and Chang, 2009). Emerging evidence is also showing a cross-talk between AhR signalling pathway and the Wnt pathway with studies reporting that an activated AhR attenuated the signalling of the Wnt canonical pathway (Kawajiri *et al.,* 2009; Procházková *et al.,* 2011; Faust *et al.,* 2013). The innate function of AhR seems to be required to "fine-tune" the signalling of the Wnt pathway and maintaining it within the required levels of activation. Since various cancers are characterised with an up-regulated Wnt canonical pathway, this intersection with AhR, further reinforces its importance in the understanding of disease and development of treatment plans (Schnedier, Branam and Peterson, 2014).

All current findings suggest that AhR is a potential candidate in cancer pathology and thus its cellular levels are crucial for normal cell functioning. Considering the close association that exists between AhR and AIP, studying their interaction and how this is influenced by the presence of AIP mutations is an important physiological aspect that requires investigation. Presently, there is no kinetic data that describes the nature of this interaction.

## 1.10. Aims and Objectives

The involvement of AIP in pituitary tumour predisposition has been demonstrated in a number of independent experiments. However, the precise molecular mechanisms through which AIP exerts its tumour suppressor role have not been fully understood. The overall aim of this study was to determine the effect of clinically-relevant N-domain mutations of AIP to specific client binding partners. This study will focus on investigating the structure and physiochemical properties of the full length AIP and N-mutants, R9Q, R16H, V49M and K103R. Although, these mutations have been previously reported (Igerja *et al.*, 2010, Formosa *et al*., 2017), no reports are currently available that compare protein stability and/or the binding kinetics of AIP with those of its mutant counterparts.

The aims and objectives of this study are:

- To design an experimental protocol to purify and stabilise AIP and AIP mutants.

- To investigate the effect of N-domain mutations on the thermal stability and overall structure of the protein.

- Being a molecular co-chaperone with no enzymatic activity, the function of AIP is best studied through protein-protein interaction analyses with client partners. For this purpose, two AIP binding partners will be purified. These include Hsp90 and PDE4A5, chosen on the basis of their stabilising influence and catalytic activity respectively. Binding studies will be carried out using surface plasmon resonance technology (SPR) and isothermal titration calorimetry (ITC).

- To investigate whether the binding affinities correlate with enzymatic activity.

- To determine the structure of full length AIP. Crystallisation conditions will be investigated as well as other techniques such as small angle X-ray scattering (SAXS) to investigate the potential role of N-domain mutation on protein structure and function.

# Chapter 2 Methodology

## 2.1. Materials and General Methods

### 2.1.1. General buffers and solutions

**Table 2.1: General buffers and solutions**

| Buffer/ Solution | Components (for 1 L solution) |
|---|---|
| 1 M Tris-Cl pH 7.5, 8.0 | 121.14 g of Trizma® base<br>pH adjusted with conc. HCl |
| 0.5 M Ethylenediaminetetraacetic acid (EDTA) pH 8.0 | 186.1 g EDTA<br>pH adjusted to 8.0 using NaOH pellets |
| 50x Tris-acetate-EDTA (TAE) | 242 g Trizma® base<br>57.1 mL glacial acetic acid<br>100 mL of 0.5 M EDTA pH 8.0 |
| 1x Phosphate buffered saline (PBS) pH 7.5 | 14.24 g $Na_2HPO_4$<br>2.76 g $NaH_2PO_4.2H_2O$<br>5.84 g NaCl |
| 0.1 M Potassium phosphate buffer (KP) pH 7.8 | 17.4 g $K_2HPO_4$ in 1 L deionised water (0.1 M)<br>13.6 g $KH_2PO_4$ in 1 L deionised water (0.1 M)<br>$KH_2PO4$ was gently added to $K_2HPO4$ until the required pH was reached. |
| 10x Tris-Glycine-SDS | 30 g of Trizma® base<br>188 g Glycine<br>20 g Sodium dodecyl sulfate (SDS) |
| 30 % Acrylamide | 30 g acrylamide<br>0.8 g N'N'-bis-methylene-acrylamide |
| Coomassie Brilliant Blue stain | 10% (w/v) R250 Coomassie Brilliant Blue<br>40.5% (v/v) methanol<br>11% (v/v) glacial acetic acid |
| Coomassie colloidal staining solution | 0.08% (w/v) Coomassie Brilliant Blue G250<br>10% (w/v) citric acid<br>8% ammonium sulfate<br>20% (v/v) methanol |

| | |
|---|---|
| Destain solution | 20% (v/v) isopropyl alcohol (IPA)<br>10% (v/v) acetic acid |
| Transfer buffer | 39 mM glycine,<br>48 mM Trizma base<br>0.0375% (w/v) sodium Dodecyl sulfate (SDS)<br>20% (v/v) methanol |

Ammonium sulfate, glycine, SDS, EDTA, Trizma®, acrylamide, N'N'-bis-metheylene-acrylamide, dithiothreitol (DTT) and imidazole were purchased from Sigma-Aldrich. Dipotassium phosphate ($K_2HPO_4$), potassium dihydrogen phosphate ($KH_2PO_4$), NaOH, disodium phosphate ($Na_2HPO_4$), sodium dihydrogen phosphate dihydrate ($NaH_2PO_4.2H_2O$), methanol, R250/G250 Coomassie Brilliant Blue, Tween-20, glycerol, and NaCl were from VWR. Citric acid, IPA, HCl and glacial acetic acid were of Analar® grade.

### 2.1.2. Media

All bacterial cultures utilised in this project were grown in tryptone yeast extract medium, which will be referred to as 2TY media in the sections that follow. Sterilisation of media was carried at 121 °C and 15 psi for a 20 min period. Antibiotics were added to the media before use. Sterile antibiotics were used at the following working concentrations; ampicillin (100 $\mu g.mL^{-1}$); kanamyin (50 $\mu g.mL^{-1}$), tetracycline (15 $\mu g.mL^{-1}$) and chloramphenicol (34 $\mu g.mL^{-1}$ for *E.coli* B strains and 170 $\mu g.mL^{-1}$ for the other *E.coli* strains). All antibiotics were purchased from Sigma-Aldrich. Tryptone, yeast extract and bacteriological agar were purchased from Oxoid.

**Table 2.2: Bacterial Media Preparation**

| Media | Component/s (for 1 L solution) |
|---|---|
| 2TY liquid broth | 16 g tryptone<br>10 g yeast extract<br>5 g NaCl |
| 2TY agar plates | 18.8 g of bacteriological agar per litre of 2TY media |

**Table 2.3: Bacterial Strains**

| Bacterial Strain | Genotype | Supplier |
|---|---|---|
| **XL1-Blue** | *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI$^q$ ZΔM15 Tn10 (Tet$^r$)]* | Strategene |
| **XL10-Gold Ultracompetent** | *Tet$^r$Δ (mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacI$^q$ZΔM15 Tn10 (Tet$^r$) Amy Cam$^r$]* | Agilent Technologies |
| **HI-Control 10G** | *mcrA Δ (mrr-hsdRMS-mcrBC) endA1 recA1 Φ80dlacZ ΔM15 ΔlacX74 araD139 Δ (ara,leu)7697 galU galK rpsL (Str$^R$) nupG λ$^-$ tonA /Mini-F lacIq$^1$(Gent$^R$)* | Lucigen |
| **BL21(DE3) CodonPlus** | *E. coli* B *F$^-$dcm ompT Ion hsdS*($_{rB^-mB^-}$) *gal* λ(DE3) *end*A *Hte* CodonPlus cells harbour a plasmid containing extra copies of *argU*, *ileY* and *leuW* tRNA genes that are rare in *E. coli*. | Novagen |

### 2.1.3. Bacterial expression systems and vectors

#### 2.1.3.1. Expresso$^{TM}$ SUMO Cloning and Expression System

This expression system makes use of an engineered form of the SUMO protein (Small Ubiquitin-like MOdifier) to facilitate the purification of proteins and enhance their solubility within a bacterial expression system. SUMO is a 100 amino acid protein that is derived from the yeast SMT3 gene product. The SUMO tag incorporates six N-terminal histidine residues ($H_6$ tag) as part of its sequence that enables the purification of the protein by metal affinity chromatography (IMAC) (Marblestone *et al*., 2006). The SUMO tag can be efficiently cleaved by a specific SUMO Express Protease available from Lucigen.

#### 2.1.3.2. pETite $H_6$-SUMO expression vector

pETite $H_6$-SUMO (2,535 bp) is a bacterial expression vector that is available from Lucigen as part of the Expresso$^{TM}$ SUMO Cloning and Expression System kit (provided as pETite-NHis-SUMO-kan). The T7 promoter controls the expression of the cloned gene. In addition to the SUMO tag, this vector harbours the gene that encodes an

aminoglycoside phosphotransferase, aph(3')-Ia. Bacterial cells transformed with this vector are thus resistant to kanamycin. Vector replication initiates at the *ori* site. This vector was provided in a linearised form, ready for co-transformation with the DNA of interest.

#### 2.1.3.3. pTH-1 expression vector

pTH-1 (4,196 bp) is a bacterial expression vector based on pTrc99A, genetically engineered for the production of authentic proteins from N-terminal hexa-histidine ($H_6$) tagged proteins (Hunter and Hunter, 2013). Downstream of the $H_6$ tag there is a StuI cut site that is ideal for blunt end cloning. A Factor Xa recognition sequence (Ile-Glu-Gly-Arg) between the $H_6$ and the StuI allows the efficient cleavage of this tag. Expression of the gene of interest is under the control of the *trc* promoter. pTH-1 also has *lac*$I^q$ sequence, a CG -> TA mutation at the -35 promoter region of the *lac*I gene. This increases the expression of the repressor *lac*I by 10-fold (Muller-Hill, Crapo and Gilbert, 1968), so preventing uninduced protein expression. pTH-1 also contains the *bla* ampicillin resistance gene and the *ori* site of replication derived from the pBR322 plasmid.

#### 2.1.3.4. pET28a expression vector

pET28a (5,369 bp) is a bacterial expression vector with a high expression level, controlled by the bacteriophage T7 promoter. This vector contains an $H_6$ tag on either side of the multiple cloning site, thus making it possible to add the tag to the gene of interest on either the N- and/or C-terminal, depending on the cloning method of choice. The kanamycin resistance gene provides selectivity for the transformed cells.

pET28a harbouring the cDNA for Hsp90$\beta^{(513\text{-}724)}$ was kindly provided by Professor Thomas Ratajczak (Department of Endocrinology & Diabetes, Western Australia).

#### 2.1.3.5. pET28a $H_6$-SUMO

This bacterial expression vector is similar to pET28a, incorporating a T7 promoter and a kanamycin resistance gene. It also appends an $H_6$-SUMO tag to the N-terminus of the protein similar to pETite. However, the SUMO is not engineered and standard SUMO protease can be used to remove it from tagged proteins. This vector was a gift from Dr Chi Trinh (University of Leeds).

### 2.1.4. Bacterial transformation

Chemically competent cells of XL1 Blue and BL21 (DE3) CodonPlus were prepared using the $CaCl_2$ method, following the procedure described by Sambrook and Russell (2001). All other bacterial cells (Table 2.3) were purchased as competent cells.

For each transformation reaction, 100 µL of competent bacterial cells (XL1 Blue and BL21 cells) were mixed with 10-50 ng of recombinant DNA. The mixture was incubated for 30 min on ice, followed by a 3 min heat shock at 42°C. 2TY media (400 µL) was added and each mixture was incubated for an hour at 37°C, prior to plating on a 2TYagar plate supplemented with the required antibiotic.

HI-Control 10G chemically competent cells were transformed as described in the Expresso™ SUMO T7 Cloning and Expression System Manual (Lucigen). XL10-Gold Ultracompetent cells were also transformed using the heat shock method, following the instructions according to the QuikChange Site-Directed Mutagenesis Manual (Agilent Technologies).

### 2.1.5. Plasmid DNA extraction

Transformed colonies harbouring the recombinant clone of interest were cultured in 10 mL 2TY media and DNA was extracted using the ZymoPure™ Plasmid MiniPrep kit (Zymo Research). RNAseA was added to the re-suspended cell pellets at a working concentration of 10 µg $mL^{-1}$ and incubated for 15 min at 37°C. The rest of the protocol was carried out according to manufacturer's instructions. Size and quality of DNA was determined by agarose gel (0.7% w/v) electrophoresis using TAE (Table 2.1) as running buffer.

### 2.1.6. DNA concentration

DNA concentration was quantified spectrophotometrically according to the equation:

$$\text{DNA conc. } (\mu g.mL^{-1}) = A_{260} \text{ x dilution factor x } X \text{ } \mu g \text{ } mL^{-1} \quad (2.1)$$

Where $X$ = 50 for double stranded DNA (dsDNA)
30 for single stranded DNA (ssDNA) and oligonucleotides

### 2.1.7. Sanger DNA Sequencing

Plasmid DNA was ethanol precipitated and samples were prepared in sterile water at a concentration of 150 ng.$\mu L^{-1}$ and 10 µL final volume. The DNA was sequenced by Bioneer (South Korea) using the appropriate primers (5 pmol.$\mu L^{-1}$). Sequence of primers is given in Table 2.8.

### 2.1.8. Polyacrylamide Gel Electrophoresis (PAGE)

#### 2.1.8.1. Sodium Dodecyl Sulfate-PAGE (SDS-PAGE)

The molecular weight and purity of the proteins of interest were analysed by SDS polyacrylamide gel electrophoresis (Laemmli, 1970) using vertical electrophoresis (mini-Protean, Bio-Rad). The resolving and stacking gel mixtures were prepared as described by Sambrook and Russell (2001) (Table 2.4).

**Table 2.4: Composition of SDS-PAGE for protein analysis**

| Chemical | 8 % Resolver gel Volume (mL) | 15% Resolver gel Volume (mL) | 5% Stacking gel Volume (mL) |
|---|---|---|---|
| AnalaR® water | 4.6 | 2.3 | 2.7 |
| 1.5 M Tris pH 8.8 | 2.5 | 2.5 | 0.5 |
| 30% Acrylamide | 2.7 | 5.0 | 0.67 |
| 0.5 % (v/v)Trichloroethanol | 0.05 | 0.05 | - |
| 10% (w/v) SDS | 0.1 | 0.1 | 0.04 |
| 10% (w/v) Ammonium persulfate (APS) | 0.1 | 0.1 | 0.04 |
| TEMED | 0.006 | 0.004 | 0.004 |

The protein sample was mixed with an equal volume of 2x SDS loading dye (50 mM Tris-Cl pH 6.8, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol, 0.25 M DTT) and boiled for 3 min prior to loading. Electrophoresis was performed at a constant voltage of 200 V using 1x Tris-glycine-SDS as the running buffer (Table 2.1). The resulting gels were evaluated under UV illumination and then stained with Coomassie Brilliant Blue stain solution. Excess dye was removed using a destain solution (Table 2.1). Protein bands were recorded using a Doc-IT LS image acquisition system (UVP) or a ChemiDoc™ Imaging System (BioRad).

#### 2.1.8.2. Native-PAGE

Native-PAGE electrophoresis was carried out to assess the state and nature of the purified proteins. Native-PAGE resolver gels (8%) were prepared as outlined in Table 2.4, excluding the addition of SDS and without boiling the samples during preparation. Samples were mixed with 6x loading dye (50 mM Tris-Cl pH 6.8, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol) prior to loading. Gels were stained with Coomassie Brilliant Blue stain, destained and visualised (Section 2.1.8.1).

### 2.1.9. Western blotting (Immunoblotting)

Immunoblotting was performed using a semi-dry transfer technique. Following SDS-PAGE electrophoresis, the gel was incubated for 20 min in transfer buffer (Table 2.1) with gentle agitation. Nitrocellulose membrane (Hybond-C extra, (GE Healthcare), with the same dimensions as the gel, was equilibrated in deionised water for 10 min followed by 10 min incubation in transfer buffer. The electroblotting unit was assembled according to the manufacturer instructions (Multiphor, LKB Pharmacia). Electrophoresis was performed at a constant current of 40 mA for 90 min. The membrane was blocked overnight at 4°C using Starting Block Reagent (Pierce). Following an extensive wash with PBS-T (1x PBS + 0.1% Tween-20), the membrane was incubated with a monoclonal anti-his tag primary antibody (Aviva Systems Biology) for 60 min at room temperature, with agitation. The membrane was thoroughly washed with PBS-T prior to incubation with a secondary antibody, an IRDye® 800CW Anti-Mouse IgG polyclonal antibody (LI-COR), at the same incubation conditions. Primary and secondary antibodies were both diluted in PBS in ratios of 1:3,000 and 1:10,000 respectively. Detection and imaging was performed using an Odyssey® CLx Infrared Imagining System (LI-COR).

## 2.2. DNA Cloning Techniques

### 2.2.1. AIP cDNA

The pcDNA3 vector harbouring the full length human AIP cDNA (pcDNA3-hAIP) was provided by Dr Robert Formosa (University of Malta), who originally obtained this expression plasmid from Prof. Marta Korbonits (St. Bartholomew's Hospital, London). Since pcDNA3 is a mammalian expression vector, the AIP coding region was sub-cloned into a vector suitable for bacterial systems as hosts. The vectors of choice were pETite and pTH-1 (Section 2.1.3).

### 2.2.2. Amplification of the AIP cDNA for Blunt-end Cloning

Amplification of the cDNA was performed using the GeneAmp High Fidelity PCR System (Perkin Elmer). The PCR mix was prepared as shown in Table 2.5.

**Table 2.5: PCR mixture: GeneAmp High Fidelity kit**

| Component | Volume (µL) |
|---|---|
| Gene High Fidelity 10x PCR buffer | 5 |
| dNTPs (200 µM) | 1 |
| Forward primer (300 nM) | 0.5 |
| Reverse primer (300 nM) | 0.5 |
| GeneAmp High Fidelity Enzyme Mix (2.5 U) | 1 |
| $MgCl_2$ (1mM) | 1 |
| DNA template (0.5 µg) | 2 |
| AnalaR® $H_2O$ | 38 |

The PCR reaction was performed in a GeneAmp PCR system 9600 thermal cycler (Perkin Elmer) at the following cycling conditions:

Denaturation step: 94°C for 2 min

30 cycles: 94°C for 15 sec, 55°C for 15 sec, 68°C for 1 min

Following the amplification reaction, the mixture was incubated for 30 min at 37 °C in the presence of Klenow enzyme (Roche $2U.\mu L^{-1}$). This step was necessarily to ensure

that the ends of the PCR product are blunt-ended and extension is complete (Haqqi, 1992; Hunter and Hunter, 1998).

### 2.2.3. Sub-cloning of AIP cDNA into pTH-1

pTH-1 vector (0.5 µg) was cleaved with 1 µL of StuI (Roche®) for 1 hour at 37°C. The linearised pTH-1 vector (0.5 µg) and AIP PCR product were both gel purified from agarose gel (0.7% w/v) using an Ultrafree DA DNA centrifugal filter device according to manufacturer's instructions (EMD Millipore) and eluted in 10 mM Tris-Cl pH 8.5. The DNA was concentrated by centrifugation at 12,000 $g$ for 5 min using a Microcon 30 concentrator (EMD Millipore) with a nucleotide cut-off of 50 bp (dsDNA).The two DNA samples were mixed in a vector to insert ratio of 1:3 respectively, in the presence of T4 DNA ligase (1 U) and 10x T4 DNA Ligase Buffer, supplemented with 10 mM ATP (Promega), in a final volume of 10 µL. The conversion of molar ratios to mass ratios of DNA was calculated using the equation on the T4 DNA Ligation kit Manual (Promega).

$$ng\ of\ DNA\ insert\ to\ use = \frac{\text{ng of vector x kb size of insert}}{kb\ size\ of\ vector}\ x\ molar\ ratio\ of\ \frac{[insert]}{vector} \qquad (2.2)$$

The ligation mixture was incubated at 14°C for 16 h. An aliquot (1 µL from a 5x diluted ligation mix) was used for the transformation of XL1 Blue competent cells.

### 2.2.4. Mini-prep analysis of DNA

The success of the blunt-end cloning was assessed by mini-prep analysis. The colonies obtained after transformation with the ligation mixture were picked and streaked on an agar plate to prepare a master plate. A clump of cells from the master plate streaks were picked and re-suspended in 100 uL of TE buffer. To this suspension, 50 µL of phenol: chloroform (1:1 v/v) and 10 µL of 6x loading dye (30% (v/v) glycerol, 0.25% (w/v) xylene cyanol, 0.25% (w/v) bromophenol blue) were added. Samples were vortexed for 1 min and centrifuged at room temperature for 5 min at 13,000 $g$ (Sambrook and Russell, 2001). Agarose gel electrophoresis was performed at a constant voltage of 6.5 $V.cm^{-1}$, using 1x TAE (Table 2.1) as electrophoresis buffer. Cells transformed with empty pTH-1 vector were used as standards.

### 2.2.5. Cloning by homologous recombination- Primer design for AIP gene amplification

To clone into the pETite $H_6$-SUMO vector, the cDNA of AIP was first amplified with PCR primers that add sequences adjacent to the cloning site that correspond to the ends of the linearised vector. The presence of these homologous flanking sequences enabled the cloning of AIP in pETite vector without the use of enzymes.

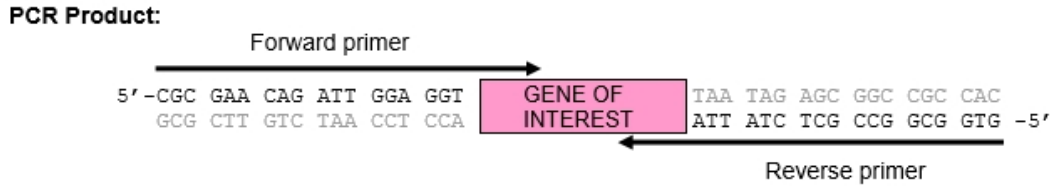



**Figure 2.1: Ligation-free cloning with the pETite vector.** PCR primers add flanking sequences that are correspond to the sequence of the vector, adjoining the site of insertion. This enables the blunt ended PCR product to recombine with the linearised vector in the absence of a ligase. (Figure taken from Expresso T7 SUMO cloning Lucigen Manual).

### 2.2.6. Amplification of the target gene

pcDNA3-hAIP clone was used as a template for the amplification reaction. The AmpliTaq DNA polymerase was supplied with the Expresso$^{TM}$ SUMO T7 Cloning and Expression System kit (Lucigen). The forward and reverse primers used for AIP were pETite-AIP-F and pETite-AIP-R (Table 2.8).

**Table 2.6: PCR mixture: Expresso$^{TM}$ SUMO T7 kit**

| Component | Volume (µL) |
|---|---|
| 10x Reaction buffer | 5 |
| dNTPs (2.5 mM each) | 1 |
| Forward primer (10 µM) | 1 |
| Reverse primer (10 µM) | 1 |
| AmpliTaq DNA polymerase | 0.5 |
| DNA template (80 ng.µl$^{-1}$) | 1 |
| AnalaR® $H_2O$ | 40.5 |

The PCR reaction was performed at the following cycling conditions:

Denaturation step: 94°C for 2 min

25 cycles: 94°C for 15 sec, 60°C for 15 sec, 72°C for 1 min /kb

Extension step: 72°C for 10 min

### 2.2.7. *In vivo* Homologous recombination

The size and quality of the PCR product was assessed by agarose gel electrophoresis. The DNA band was excised, gel purified and concentrated as described in Section 2.2.3 The purified DNA insert (100 ng) was mixed with 25 ng of linear pETite vector and transformed directly into chemically competent HI-Control 10G cells (Section 2.1.4). Homologous recombination occurs *in vivo* with high efficiency. DNA from transformed colonies was tested by mini-prep analysis (Section 2.2.4) and the whole cDNA sequence was confirmed by Sanger DNA sequencing prior to further use.

### 2.2.8. Site-directed mutagenesis (SDM)

Recombinant pTH-$H_6$-AIP plasmid was used as DNA template for the mutagenesis reaction, to generate four clinically relevant AIP mutants. SDM reactions were performed as described on the QuikChange II XL Site-Directed Mutagenesis Kit Manual (Agilent Technologies) using specific mutagenic primers (Table 2.8). Mutagenesis reactions were performed in a GeneAmp (Perkin Elmer) PCR system 9600 thermal cycler at the following conditions:

Denaturation step: 95°C for 1 min

18 cycles: 95°C for 50 sec, 60°C for 50 sec, 68°C for 1 min/kb

Extension step: 68°C for 7 min

At the end of the mutagenesis reaction, each sample was incubated at 37°C for 1 hour in the presence of DpnI (1µL from 10 $U.\mu L^{-1}$stock) to digest parental non-mutated DNA. DpnI-treated DNA was diluted ten-fold and the resultant mix (2 µL) was used for the transformation of XL10-Ultracompetent cells (Section 2.1.4).

The above procedure was also repeated using the recombinant pETite $H_6$-SUMO-AIP as template, to generate $H_6$-SUMO-AIP mutants, in addition to $H_6$-AIP mutants. The introduction of each mutation was confirmed by Sanger DNA sequencing (Bioneer).

### 2.2.9. Heat shock protein 90β (Hsp90β)$^{(513-724)}$

#### 2.2.9.1. Hsp90 cDNA

The DNA construct pET28a, harbouring the human heat shock protein 90β (residues 513-724) with an N-terminal $H_6$-tag, required no sub-cloning procedures and was used as provided. BL21 (DE3) CodonPlus competent cells were transformed with the recombinant plasmid using the calcium chloride method (Section 2.1.4).

### 2.2.10. Phosphodiesterase 4A5 (PDE4A5)

#### 2.2.10.1. PDE4A5 cDNA

The pMALR6 expression vector harbouring the full-length PDE4A5 cDNA (rat isoform) as a fusion to maltose-binding protein (MBP) was provided by Professor Graeme Bolger (University of Alabama USA). The rat isoforms of PDE4A5 used in this study is a homologue of the human PDE4A4. PDE4A5/4 has been shown to have a unique interaction with the co-chaperone AIP (Bolger *et al*., 2003; Bizzi *et al*., 2019)

#### 2.2.10.2. Sub-cloning of the PDE4A5 cDNA into pETite vector

PDE4A5 cDNA was sub-cloned in pETite $H_6$-SUMO vector by homologous recombination. The amplification of the cDNA and cloning procedure were similar to those described previously for AIP (Sections 2.2.6 and 2.2.7). pMALR6 expression vector harbouring the PDE4A5 cDNA was used as a template. pETite-PDE-F and pETite-PDE-R were used as the forward and reverse primers respectively (Table 2.8).

### 2.2.11. Aryl hydrocarbon receptor (AhR)

#### 2.2.11.1. AhR cDNA

The human full length AhR cDNA was purchased from Sino Biological and was encoded within the mammalian expression vector pCMV3. It was therefore sub-cloned into the pET28-SUMO expression vector for use within a bacterial system.

#### 2.2.11.2. Amplification of the AhR gene

Amplification of AhR cDNA was performed using the Invitrogen™ Platinum SuperFi™ Green DNA polymerase kit, using primers that have a 23-bp homology to each end of the linearised vector. The PCR mixture was prepared as described in Table 2.7 and the reaction performed at the following cycling conditions:

Denaturation step: 98°C for 2 min
10 cycles: 98°C for 8 sec, 67°C for 15 sec, 72°C for 1.5 min
25 cycles: 98°C for 8 sec, 72°C for 15 sec, 72°C for 1.5 min
Extension step: 72°C for 5 min

**Table 2.7: PCR mixture: Invitrogen™ Platinum SuperFi™ Green kit**

| Component | Volume (µL) |
|---|---|
| 5x Green buffer | 10 |
| 10 mM DNTPs mix | 1 |
| AhR-F (10 µM) | 2.5 |
| AhR-R (10 µM) | 2.5 |
| DNA template (5 ng) | 0.5 |
| Platinum™ SuperFi™ Green DNA Polymerase (2 $U.\mu L^{-1}$) | 0.5 |
| $H_2O$ | 33 |

#### 2.2.11.3. *In-vitro* Homologous recombination

Homologous recombination was carried out using the one-step sequence and ligation independent cloning (SLIC), following the method reported by Jeong *et al* (2012). This technique makes use of T4 DNA polymerase that exhibits 3'-> 5'exonuclease activity, to generate single-stranded 5'-overhangs that are complementary between vector and insert. Unlike the homologous recombination, described in Section 2.2.5 which is enzyme-free, this recombination occurs *in vitro* with the aid of a T4 DNA polymerase.

Gel purified AhR cDNA and pET28a-SUMO vector linearised by EcoRV restriction digest, were mixed together in a molar ratio of 4:1 respectively, in the presence of 1x buffer supplemented with bovine serum albumin (NEB Ligation kit). The mixture (10 µL final volume) was treated with 0.6 U of T4 DNA polymerase (NEB) and incubated at room temperature for 2.5 min. This was followed by 10 min incubation on ice to inactivate the T4 DNA polymerase activity and bacterial transformation in XL1 Blue competent cells (Section 2.1.4).

**Table 2.8: Oligonucleotides used for DNA sequencing, PCR and mutagenesis**

| Name | Nucleotide Sequence | N | % GC | $T_m$ °C |
|---|---|---|---|---|
| | **DNA sequencing** | | | |
| PKPRO | CTCGTATAATGTGTGGAATTGTGAGCGG | 28 | 46 | 76 |
| PKTERM | CCTGACCCCATGCCGAACTCAGAAG | 25 | 60 | 79 |
| SUMO-F | ATTCAAGCTGATCAGACCCCTGAA | 24 | 46 | 72 |
| SUMO-R | CTCAAGACCCGTTTAGAGGC | 20 | 55 | 70 |
| T7 TERM | GCTAGTTATTGCTCAGCGG | 19 | 53 | 67 |
| | **PCR** | | | |
| AIP-F | GCGGATATCATCGCACGCCT | 20 | 60 | 72 |
| AIP-R | TCAATGGGAGAAGATCCCCC | 20 | 55 | 70 |
| pETite-AIP-F | CGCGAACAGATTGGAGGTGCGGATATCATCGCACGCCTCCGG | 42 | 62 | 90 |
| pETite-AIP-R | GTGGCGGCCGCTCTATTAATGGGAGAAGATCCCCCGGAACCG | 42 | 62 | 90 |
| pETite-PDE-F | CGCGAACAGATTGGAGGTGAGCCTCCGGCCGCCCCCTCGGAA | 42 | 69 | 91 |
| pETite-PDE-R | GTGGCGGCCGCTCTATTAGGCAGGGTCTCCGCCTGACCCCCA | 42 | 69 | 93 |
| AhR-F | GCTCACAGAGAACAGATTGGTGGAAACAGCAGCAGCGCCAACAT | 44 | 52 | 82 |
| AhR-R | GTGGTGGTGGTGCTCGAGTGCGGCTTACAGGAATCCACTGGATGTC | 46 | 59 | 85 |
| | **Site-Directed Mutagenesis** | | | |
| AIP[R9Q]-F | ATCGCACGCCTCC**A**GGAGGACGGGATC | 27 | 67 | 84 |
| AIP[R9Q]-R | GATCCCGTCCTCC**T**GGAGGCGTGCGAT | 27 | 67 | 84 |
| AIP[R16H] – F | GATCCAAAAAC**A**TGTGATACAGGAAGG | 27 | 41 | 73 |
| AIP[R16H]-R | CCTTCCTGTATCACA**T**GTTTTTGGATC | 27 | 41 | 73 |
| AIP[V49M]-F | ACGAGGGCACC**A**TGCTGGACGAC | 23 | 65 | 79 |
| AIP[V49M]-R | GTCGTCCAGCA**T**GGTGCCCTCGT | 23 | 65 | 79 |
| AIP[K103R]-F | CTGGTGGCCA**G**GAGTCTCCGC | 21 | 71 | 79 |
| AIP[K103R]-R | GCGGAGACT**C**CTGGCCACCAG | 21 | 71 | 79 |

N denotes the number of nucleotides in the sequence. For homologous recombination nucleotides equivalent to the vector are underlined, while for Site-Directed Mutagenesis mutagenic (mismatching) nucleotides are shown in bold. F denotes forward and R reverse (complementary) sequences relative to the coding gene sequence.

Each primer was synthesized by Bioneer and was supplied as a lyophilized sample. It was reconstituted with the specified volume of sterile deionised water to a final concentration of 100 pmoles $\mu L^{-1}$. Primer concentration was then checked spectrophotometrically as described in Section 2.1.6. Where required, the conversion of nanograms to picomoles of oligonucleotides was calculated using the following equation;

$$X\ pmoles\ of\ oligos\ = \frac{\text{ng of oligo}}{330\ x\ no.of\ bases\ in\ oligo}\ x\ 1000 \qquad (2.3)$$

The melting temperature of each primer was calculated using the equation from Sambrook and Russell (2001):

$$81.5 + 0.41(\%GC) - \frac{675}{N} \qquad (2.4)$$

For mutagenic primers, the Tm was calculated as described on the QuikChange Site-directed Mutagenesis Protocol, using equation 2.5 below:

$$81.5 + 0.41(\%GC) - \frac{675}{N} - \%\ Mismatch \qquad (2.5)$$

The values for % GC and % mismatch were taken as whole numbers.

For each individual amplification reaction, the optimal annealing temperature was calculated as described by Thermo Fischer (Allawi, Santa Lucia, 1997). In the case of primers used for homologous recombination, only the nucleotide sequence matching to the cDNA of the protein of interest was used for calculation of the annealing temperature.

## 2.3. Protein Purification and Characterisation

### 2.3.1. Optimisation of recombinant protein expression

Small-scale experiments (10 or 50 ml cultures) were performed to determine the optimum conditions for recombinant protein expression. The conditions investigated included; the incubation period prior to induction ($OD_{600}$), the concentration of the inducer (isopropyl β-D-1-thiogalactopyranoside, IPTG, LabScientific Inc.), the final cell density at which cells are harvested (length of growth period after induction) and the temperature of incubation after induction.

### 2.3.2. Large-scale protein expression of target protein

For large scale expression, 500 mL 2TY media in a 1 liter flask supplemented with the appropriate antibiotic was inoculated with 5 mL of an overnight culture of *E.coli* transformed with the vector of choice. Cultures were incubated at 37°C with shaking until growth had reached the mid-log phase ($OD_{600}$ = 0.4-0.8). Expression of protein of interest was performed based on the optimal value of IPTG concentration, incubation temperature and the length of growth period after induction, as determined from the optimisation experiments (Section 2.3.1). Cultures were harvested by centrifugation at 6,000 *g* for 15 min at 4°C.

### 2.3.3. Protein Purification: Metal Affinity Chromatography

All the cDNAs utilised in this study were cloned to include either a N-terminal $H_6$-SUMO-tag or a N-terminal $H_6$-tag in the expression construct. Immobilized metal affinity chromatography (IMAC) was therefore the method of choice for the purification of all recombinant proteins.

IPTG-induced cultures (transformed BL21 (DE3) CodonPlus competent cells) were harvested (Section 2.3.2) and re-suspended in 30 mL of ice-cold START buffer (20 mM Tris-Cl pH 7.5, 100 mM NaCl). This cell suspension was lysed by French pressing (French® Pressure Cell Press, SLM AMINCO) at an applied cell pressure of 16,000 $Nm^2$. One cOmplete protease inhibitor tablet (Roche®) and 10 mM phenylmethylsulfonyl fluoride (PMSF) were added to 30 mL of lysate. DNA was

degraded by DNase A (0.01 mg.mL$^{-1}$) supplemented with 5 mM $MgCl_2$ followed by sonication on ice (8 cycles: 10 seconds on, 50 seconds off) at 15 Amplitude microns using a Soniprep II 150 (MSE). The lysate was clarified by centrifugation at 15,000 *g* (SS-34 rotor, Sorvall RC-5C centrifuge) for 30 min at 4°C and filtered through a 0.45 µm cellulose acetate Whatman™ filter prior to loading on a cOmplete™ His-Tag Purification Column (GE Healthcare). The column was pre-charged with 0.1 M solution of $NiSO_4$ and equilibrated with START buffer. The bound protein was eluted using an imidazole step gradient with increasing imidazole concentration. The optimal imidazole concentration for the elution of the protein of interest was monitored by measuring the absorbance at 280nm of collected fractions. The protein fractions were dialysed against the appropriate storage buffer (Table 2.9). Protein samples were then analysed by SDS-PAGE electrophoresis and purity was assessed through a densitometry technique using ImageJ v1.52a (Rasband, 2019). Unless stated otherwise, the purified protein was stored at -20°C in 1 mL aliquots for future use.

**Table 2.9: Protein storage buffers**

| Protein | Storage buffer |
|---|---|
| AIP and mutans (both $H_6$-SUMO- and $H_6$-tagged) | 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 10% v/v glycerol, 2 mM DTT |
| Hsp90β$^{(513–724)}$ | 10 mM Tris-Cl pH 7.4, 100 mM NaCl, 10% v/v glycerol, 5 mM DTT |
| $H_6$-SUMO-PDE4A5 | 20 mM Tris-Cl pH 8.0, 100 mM NaCl, 5% v/v glycerol, 1 mM TCEP |

Table 2.10: Summary of the proteins of interest and their properties.

| Protein | Organism | Expression vector | Type of protein tag | Amino acids | Molecular weight (Da) | Isoelectic point (pI) | Molar Extinction Coefficient $M^{-1}$ $cm^{-1}$ at pH 7.0 |
|---|---|---|---|---|---|---|---|
| AIP | Human | - pETite<br>- pTH | - $H_6$-SUMO<br>- $H_6$ | 437<br>342 | 49,808<br>39,041 | 5.8<br>6.6 | SUMO_AIP and mutants: 40,880<br>$H_6$-AIP and mutants: 39,390 |
| AIP-R9Q | Human | - pETite<br>- pTH | - $H_6$-SUMO<br>- $H_6$ | 437<br>342 | 49,780<br>39,014 | 5.7<br>6.5 | |
| AIP-R16H | Human | - pETite<br>- pTH | - $H_6$-SUMO<br>- $H_6$ | 437<br>342 | 49,790<br>39,022 | 5.8<br>6.6 | |
| AIP-V49M | Human | - pETite<br>- pTH | - $H_6$-SUMO<br>- $H_6$ | 437<br>342 | 49,841<br>39,073 | 5.8<br>6.6 | |
| AIP-K103R | Human | - pETite<br>- pTH | - $H_6$-SUMO<br>- $H_6$ | 437<br>342 | 49,837<br>39,069 | 5.8<br>6.6 | |
| Hsp90β$^{(513-724)}$ (monomer) | Human | pET28a | $H_6$ | 218 | 24,017 | 5.1 | 10,345 |
| PDE4A5 | Rat | pETite | $H_6$-SUMO | 951 | 105,590 | 5.1 | 81,705 |
| AhR | Human | pET28a N-His SUMO | $H_6$-SUMO | 954 | 108,299 | 6.0 | 71,470 |

**The pETite vector refers to the pETite $H_6$-SUMO expression system described in Section 2.1.3.2. *E.coli* BL21 (DE3) CodonPlus cells were used as the expression system. The wild type AIP protein and mutants were purified both as a SUMO-construct and a $H_6$-construct. Except for the Hsp90β construct (amino acid residues 513-724) all other constructs encoded the full-length protein. Protein molecular weight and other physical parameters were computational determined by Protean (DNASTAR, Lasergene), based on the protein sequence of the sequenced gene.**

### 2.3.4. Removal of the $H_6$-SUMO-tag

The $H_6$-SUMO-tag was removed using the SUMO Express Protease (Lucigen). The purified $H_6$-SUMO-tagged AIP protein was dialysed for 24 h against 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 10% (v/v) glycerol. The protein was diluted in cleavage buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% (v/v) glycerol, 2 mM dithiothreitol (DTT) and incubated overnight in the presence of the protease (1 U/100 µg of fusion protein) at 4°C with gentle agitation. The mixture was dialysed extensively to remove the DTT and loaded on a (1 ml or 5 ml) IMAC column to separate the $H_6$-SUMO-tag, SUMO Express protease and any uncleaved fusion protein from the untagged AIP protein. The cleavage efficiency and the state of the resulting protein was analysed by SDS-PAGE electrophoresis (Section 2.1.8).

### 2.3.5. Purification of recombinant protein from inclusion bodies

Recombinant PDE4A5 has the tendency of forming inclusion bodies when expressed in bacterial hosts. The methodology used to recover soluble protein from inclusion bodies is a modification of the technique reported by Richter, Hermsdorf and Dettmer (2002).This involves a denaturing technique, whereby the protein is initially denatured through the use of a high concentration of denaturing chemicals followed by re-folding.

Cells from IPTG-induced cultures were harvested and re-suspended in 30 mL of START buffer (Section 2.3.3). The re-suspended cells were lysed by French Press and sonicated on ice (Section 2.3.3). DNase A (0.01 $mg.mL^{-1}$) supplemented with 5 mM $MgCl_2$ was added to the resulting suspension and was stirred at room temperature for 30 min. The inclusion bodies were pelleted by a 10 min centrifugation at 10,000 *g* at 4°C. The pellet was re-suspended in washing buffer (0.1 M Tris-HCl pH 7.0, 20 mM EDTA pH 8.0) and re-centrifuged at 10,000 *g* for 10 min at 4°C. This centrifugation step was repeated three times to obtain a clean pellet of inclusion bodies. The pellet was re-solubilised in 6 mL of solubilisation buffer (0.1 M Tris-HCl pH 8.0 containing 6 M guanidine hydrochloride, 100 mM DTT and 1 mM EDTA pH 8.0) and incubated at room temperature for 2 h on a rotating mixer. The solution was centrifuged at 30,000 *g* for 10 min and the clear supernatant was poured directly into the refolding

buffer (0.1 M Tris-HCl pH 7.0 containing 40 µM $ZnSO_4$, 20 mM $MgCl_2$, and 10 mM DTT) and stirred for 16 h at 10°C. After refolding, the sample was dialysed against 20 mM Tris-Cl pH 8.0, 100 mM NaCl and centrifuged for 30 min at 10,000 g at 4°C to remove aggregates. The soluble supernatant was loaded on a pre-equilibrated Ni-charged HiTrap Chelating Column (GE Healthcare) and purified through IMAC (Section 2.3.3).

### 2.3.6. Protein aggregation and stability studies

Some proteins such as AIP have a tendency to form aggregates under certain experimental conditions and upon storage. The formation of aggregates was monitored by 8% Native-PAGE to visualise the presence of high molecular weight aggregates. Stability experiments were performed on an *Optim*®1000 instrument (Avacta Innovative Analysis UK). This helped determine the storage buffer conditions for the protein.

AIP that had been purified as described in Section 2.3.3 and stored as 500 uL aliquots in a buffer containing 20 mM Tris-Cl pH 7.5, 100 mM NaCl was left stirring at 4°C in the presence of a specific additive. Following overnight incubation, soluble protein was separated from any protein aggregates by centrifugation at 16,000 *g* for 15 min. The supernatant was analysed on 8% Native-PAGE. Gels were stained with the Coomassie brilliant blue solution. This methodology is a modification of the procedure outlined by Bondos and Bicknell (2003).

Further stability experiments were performed on an *Optim*®1000 instrument. This is a sensitive, high throughput instrument that assesses the stability of micro quantities of proteins under a vast array of conditions. Twelve samples of $H_6$-SUMO-AIP at a final concentration of 1 $mg.mL^{-1}$ were prepared in 20 mM Tris-Cl pH 7.5, 100 mM NaCl, to which different additive concentrations was added. Aliquots of each sample (10 µL) were loaded onto the *Optim*®1000 compatible micro-cuvette arrays (MCAs). The machine was programmed to monitor the static light scattering (SLS) from a 266 nm and 473 nm laser source. The sample temperature was increased from 15°C to 90°C in 1°C step intervals with a 30 sec hold prior to each temperature change. The SLS signal was recorded at each temperature point.

The change in fluorescence emission of tryptophan was measured through the barycentric mean fluorescence ($\lambda_{BCM}$) that is calculated by the following equation:

$$\lambda\, BCM = \frac{\Sigma\, I(\lambda) x\, \lambda}{\Sigma\, I\, (\lambda)} \quad (2.5)$$

Where $\lambda_{BCM}$ is the mean barycentric fluorescence in nm, λ represents any given wavelength and I(λ) is the fluorescence intensity at a particular wavelength.

When the barycentric mean fluorescence is considered as a function of temperature, the melting temperature ($T_m$) of the protein can be calculated. The mathematical representation of this is shown in Equation 2.6 below (Garstka *et al.*, 2014).

$$Tm = \; max \frac{dBCM}{dT}(T) \quad (2.6)$$

Where Tm is the transition temperature, *max* is the local maximum, dBCM/dT (T) is the first derivative of the barycentric mean fluorescence as a function of temperature (T) in nm / °C.

The generated data was analysed using the *Optim*®1000 Analysis Software version 2.0 (Avacta Analytical).

### 2.3.7. Gel filtration chromatography

Gel filtration was used for an estimate of molecular size and to separate monomeric from multimeric and/or aggregated protein. Concentrated samples were loaded on a Superdex™ 75 (S75) column connected to an AKTA™ Prime chromatography unit, set at a flow rate of 2 mL.min$^{-1}$. Column was equilibrated with 20 mM Tris-Cl pH 7.5, 0.1 M NaCl, 5% (v/v) glycerol, 2 mM DTT used as the running buffer. Column equilibration and protein elution was monitored in real time using the Unicorn Software 3.0 that is supplied with the AKTA™ purification system. Eluted proteins were collected in 2 mL fractions.

### 2.3.8. Mass Spectrometry

For molecular weight determination, samples of AIP and AIP mutants were subject to mass spectrometry using an Electro-Spray Mass Spectrometer. This analysis was carried out at the Astbury Centre for Structural Molecular Biology, at Leeds University. Samples were prepared in 50 mM ammonium acetate pH 7.4 at a concentration of 20 μM and 20 μL final volume. $H_6$-AIP was also analysed by in-gel trypsin digest mass spectrometry. The sample was separated on a 4-20% Mini-PROTEAN® TGX™ precast protein gel (Bio-Rad), under denaturing conditions. The gel was stained overnight with Coomassie colloidal staining solution (Table 2.1) and destained extensively with Analar® water. In-gel trypsin digest mass spectrometry was performed at EMBL (Heidelberg).

### 2.3.9. Circular dichroism

Far-UV and near-UV CD-spectroscopy data was collected to asses secondary and tertiary structure respectively. The melting temperature of each sample was determined by a continuous temperature scan, from 5°C - 90°C with 5°C increase increments and a 2 min hold prior each temperature change. Experimental parameters (Table 2.11) were set using the Pro-Data Chirascan Software. Protein samples were prepared at a concentration ranging between 0.1-0.2 $mg.mL^{-1}$ in 10 mM potassium phosphate (KP) pH 7.8. This buffer was used as the reference buffer. CD data was collected using a Chirascan™ CD-Spectrometer (Applied Photophysics) at the University of Leeds and analysed using Graphpad Prism 6.0.

**Table 2.11: Circular dichroism parameter settings**

| | Far-UV | Near-UV | Temperature Ramp Scan |
|---|---|---|---|
| Cell path length (mm) | 1 | 10 | 1 |
| Bandwidth (nm) | 2.0 | 2.0 | 2.0 |
| Wavelength range (nm) | 180-260 | 250-320 | 180- 260 |
| Set temperature /°C | 5 | 20 | 5 |
| Temperature ramp /°C | - | - | 5 - 90 - 5 |
| Approximate scan time (min) | 4 | 4 | 126 |

### 2.3.10. Protein-protein interaction studies

Protein-protein interaction studies between AIP and its client proteins were conducted by surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC). Both techniques enable the determination of affinity binding constants. SPR and ITC experiments were performed at the Astbury Centre for Structural Molecular Biology, at the University of Leeds.

Surface plasmon resonance studies were performed at 25°C using a Biacore 2000 instrument (GE Healthcare Life Sciences). All buffers were degassed and sterile-filtered. The protein samples were centrifuged for 10 min at 16,000 *g* prior to injection onto a CM5 sensor chip (GE Healthcare Life Sciences). Hsp90β$^{(513\text{-}724)}$ and $H_6$-SUMO-PDE4A5 (ligands) were immobilized on different flow cells of the CM5 chip, while AIP and AIP mutants (analytes) were injected across the surface. The system was primed first in water and then 0.1 M sodium acetate buffer pH 5.6 that was used as the running buffer during pH scouting.

#### 2.3.10.1. pH Scouting

Prior to the activation of the CM5 chip, 'electrostatic pre-concentration' experiments were performed. This step was imperative in order to determine the ideal pH at which to conduct protein coupling whilst avoiding protein precipitation on the surface of the chip during the experiment.

#### 2.3.10.2. Chip Derivatisation

Chip derivatisation is the process of activating the surface of the CM5 chip and immobilisation of protein. Protein immobilisation was performed using amine coupling chemistry. For surface activation, equal volumes of 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and 0.05 M N-hydroxysuccinimide (NHS), were used. Both reagents were components of the Amine Coupling Kit (GE Healthcare Life Sciences).

During surface activation, EDC first reacts with the carbonyl groups of the dextran matrix to form an unstable O-acylisourea ester. The ester bond is subsequently attacked by the electrophilic nitrogen of the NHS moiety, forming a stable amine-

reactive ester on the dextran matrix. Proteins injected across the activated surface react with the NHS-ester through their primary amine groups, displacing the NHS moiety and forming a covalent amide bond (GE Healthcare, 2012). The chemistry of the immobilisation process is illustrated in Figure 2.2.

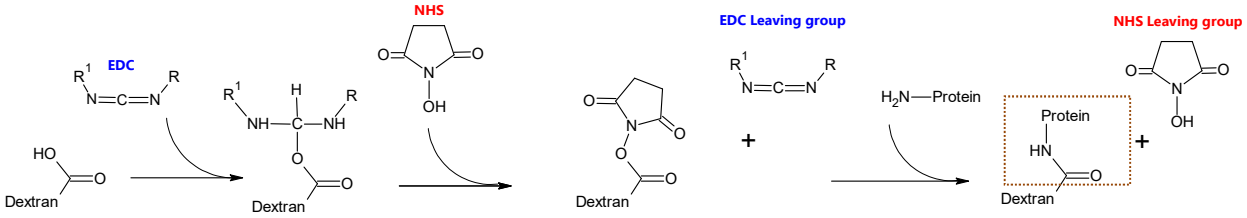



**Figure 2.2: The chemistry of surface activation of a CM5 chip by amine coupling.** EDC and NHS are both regenerated by the end of the reaction and can thus be regarded as catalysts that facilitate the approach and covalent binding of proteins to the carbonyl groups of the dextran matrix. The mechanism was reproduced using ACD/ChemSketch version 2018.2.1.

Hsp90β$^{(513\text{-}724)}$ and $H_6$-SUMO-PDE4A5 (50 µg.mL$^{-1}$) were injected for 5 min at a constant flow rate of 5 µL.min$^{-1}$ over the NHS/ECD-activated surface to generate the required surface densities on flow cells-2 and 3 respectively. Unreacted material was eluted in PBS/ NaCl salt solution. Subsequently, ethanolamine (1 M, pH 8.5) was injected to quench any remaining unreacted ester groups. No protein was immobilised on flow cell-1 as this was used as the reference cell.

### 2.3.10.3. Binding Assays

Interaction analysis was performed at 25°C in 1 x PBS pH 7.4 containing 0.1% surface polysorbate 20 (P20). Analyte was injected for 3 min at a flow rate of 40 µL min$^{-1}$, followed by a 5 min dissociation phase. The concentration of each analyte was gradually increased from 20 nm to 20 µM. The surface of the chip was extensively washed with a solution of 1x PBS, 1M NaCl, 0.05% (w/v) SDS prior to the next injection. Binding was observed in real-time. Curve fitting and kinetic data analyses were evaluated with BIAevaluation software version 3.1. Details of each step are summarised in Table 2.12.

**Table 2.12: Summary of chip derivatisation steps**

| Chip Derivatisation Summary | Flow rate ($\mu L\ min^{-1}$) | Concentration | Contact Time (min) | Volume injected ($\mu L$) |
|---|---|---|---|---|
| **pH scouting** | | | | |
| a) Ligand injection | 5 | 50 $\mu g.mL^{-1}$ | 5 | 25 |
| b) Salt wash (PBS/NaCl) | 5 | 1x PBS, 1 M NaCl | 2 | 10 |
| c) Ethanolamine wash | 5 | 1M | 2 | 10 |
| **Surface activation** | | | | |
| a) EDC/NHS injection | 5 | 0.2 M EDC/0.05 M NHS | 7 | 35 |
| b) Ligand injection | 5 | 50 $\mu g.mL^{-1}$ | 5 | 25 |
| c) Salt wash (PBS/NaCl) | 5 | 1x PBS, 1 M NaCl | 4 | 20 |
| d) Ethanolamine wash | 5 | 1 M | 7 | 35 |
| **Binding Assays** | | | | |
| a) Analyte injection | 40 | Variable | 3 | 120 |
| b) Salt wash (PBS/NaCl) | 5 | 1x PBS, 1 M NaCl, 0.05% SDS | 2 | 10 |

### 2.3.11. Isothermal Titration Calorimetry (ITC)

ITC experiments were performed out at 25°C in an ITC200 micro-calorimeter (MicroCal). Hsp90β$^{(513-724)}$ in 1x PBS, 5% (v/v) glycerol at a concentration of 150 µM was titrated into the reaction cell containing $H_6$-AIP (16 µM), dialysed against the same buffer. The same injection was repeated for every AIP sample. The heat of dilution of titrant into water and titrant into buffer were used as controls. Another set of ITC experiments was performed using 20 mM Tris-Cl pH 7.5, 5mM NaCl, 5% (v/v) glycerol as the protein buffer. Equilibrium dissociation constants were calculated using the MicroCal ORIGIN software

### 2.3.12. PDE-Glo™ Phosphodiesterase Assay

The effect that AIP binding may have on the activity of PDE4A5 was measured using the luminescence based PDE-Glo™ phosphodiesterase assay (Promega). The chemistry of the PDE assay is shown in Figure 2.3.

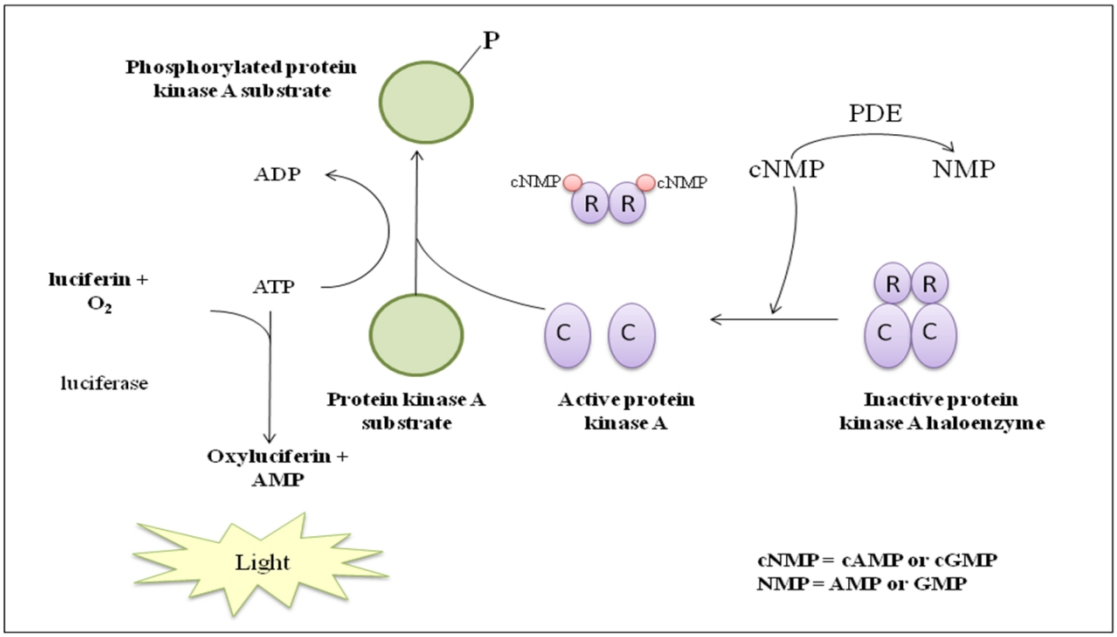



**Figure 2.3: Chemistry of the PDE-Glo™ Phosphodiesterase Assay.** When active PDE is present, cAMP is hydrolysed to AMP resulting in a reduction in the amount of active protein kinase A (PKA). This leaves more ATP to be used in the luciferase reaction, resulting in a higher luminescence signal. The luminescence signal is thus proportional to ATP levels that correlate with PDE activity. The figure has been reproduced using the chemical illustration described on the PDE-Glo™ Kit Manual (Promega).

The enzymatic reaction was initiated by addition of cAMP (2 µM) to PDE4A protein (Bio-Techne) that was previously diluted to the required concentration, in 1x PDE-Glo™ reaction buffer. The solution was incubated at 37°C for 30 min and stopped by the addition of PDE-Glo™ termination buffer. The rest of the procedure was performed according to manufacturer's instructions (Promega). Each assay was performed in a white 96-well plate (Perkin Elmer) and luminescence readings were measured with a plate-reading luminometer, Mithras LB940 (Berthold Technologies).

**Table 2.13: Composition of the PDE-Glo™ Phosphodiesterase assay**

| Reagent | Volume per reaction (µL) |
|---|---|
| Pure PDE4A diluted in 1x PDE-Glo™ Reaction buffer | 12.5 |
| Substrate, cAMP (2 µM) | 12.5 |
| 1x PDE-Glo™ Termination buffer | 12.5 |
| 1x PDE-Glo™ Detection buffer | 12.5 |
| Kinase-Glo® Reagent | 50 |

#### 2.3.12.1. Titration of cyclic nucleotide Phosphodiesterase and Inhibition Assays

A PDE titration was performed to empirically determine the optimal enzymatic concentration to use. The reaction was performed as described in Section 2.3.12, with increasing concentrations of PDE, prepared by serial dilution. The PDE assay was also repeated with increasing amounts of $H_6$-AIP to determine the concentration that exerts maximal enzymatic inhibition. The concentration of PDE was kept at a constant optimal concentration, previously determined experimentally. Inhibition assays were performed using all $H_6$-AIP mutants. Experiments were carried out in triplicates for reliability.

Data was graphically plotted and analysed using GraphPad Prism software Version 6.0. Statistical analyses were carried out using SPSS Statistics Version 20 (IBM). The distribution of the data was checked using the Shapiro-Wilk and Kolmogorov-Smirnov test. Parametric tests were used for normally distributed data. These included the independent t-test and one-way ANOVA, set at a confidence level of 95%.

### 2.3.13. Size-Exclusion Chromatography Small Angle X-ray Scattering (SEC-SAXS)

Small angle X-ray scattering is a powerful technique currently at the forefront of protein research, which is suitable for studying flexible systems in solution. Although, this technique is limited in terms of resolution (10-20 Å) when determining models *ab initio*, it is especially informative when coupled with high-resolution data, obtained through X-ray crystallography or NMR and/or homology models. This allows for the generation of molecular envelopes and more detailed atomistic models (Mertens and Svergun, 2010).

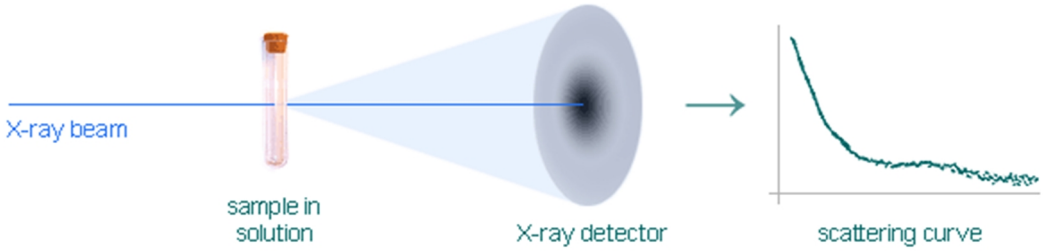



**Figure 2.4: Small Angle X-ray scattering** (BIOSAXS, GmbH, 2016)

Samples of various concentrations of pure AIP (both $H_6$- and $H_6$-SUMO-tagged) in 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM DTT, 5% (v/v) glycerol were initially analysed by a BioSAXS robot (Arinax, EMBL-designed). Samples were centrifuged at 16,000 *g* prior to use. During a SAXS experiment, focused monochromatic X-rays are directed onto the sample in solution and the scattered radiation is detected and used to generate a scattering profile. Aliquots of the AIP samples that showed the best scattering profile were loaded on a pre-calibrated Shodex KW403-4F gel filtration column (Agilent 1200 HPLC system), set at a constant flow rate of 0.08 $mL.min^{-1}$. To eliminate the issue of buffer mismatch, the buffer used to dialyse the proteins was re-used for column equilibration. Experimental parameters are given in Table 2.14.

**Table 2.14: SAXS experimental parameters**

| Technique | Injection Volume (µL) | Flow rate mL $min^{-1}$ | Radiation exposure time) (sec) | Time of run (min) |
|---|---|---|---|---|
| BIOSAXS Robot | 45 | 0.16 | 3 | 30 |
| SEC-SAXS | 55 | 0.08 | 6 | 60 |

SAXS data was analysed using the program ScÅtter Version 3.0, a JAVA-based application (Rambo, 2019a). The quality of the data was assessed through the Kratky plot, Guinier Analysis and Porod plot. The SAXS scattering profile was used to generate a low resolution bead density model using DAMMIN/F, a modelling algorithm that runs from the ScÅtter graphical user interface. The data was further analysed to generate models of $H_6$-AIP and $H_6$-SUMO-AIP. SWISS-MODEL (Waterhouse, 2018) was employed to generate PDB files that match regions of the input sequence that are indexed correctly. PDB files were used to generate the distance constraints for folding into a single polypeptide chain. Crystallography and NMR Systems (CNS) was also used to make an extended chain from the provided sequence and to define the constraints between domains. Secondary structure prediction and constraints for the unknown regions of the protein sequence were determined using SPIDER2 (Khashan, Zheng and Tropsha, 2012; Fleishmann *et al*., 2011). The theoretical scattering profile of the generated models was computationally compared to the experimental one and the best model identified using FoXS, a fast SAXS profile computation program that employs the Debye Formula (Schneidman-Duhovny et al., 2013). The data of $H_6$-SUMO-AIP and $H_6$-AIP was also analysed using MONSA (Svergun, 1999), to visualise the arrangement of the protein with respect to the $H_6$-SUMO tag. A total of eleven runs were performed with MONSA to generate a bead model that compliments the atomistic models generated by FoXS.

The modelling analyses were kindly performed by Dr Robert Rambo, who is the author of the program ScÅtter and science group leader for the soft condensed matter village at Diamond Light Source. All SAXS data was collected at beamline 21 (B21) at Diamond Light Source, UK's national synchrotron science facility situated at the Harwell Centre for Research and Innovation, Oxford.

### 2.3.14. Crystallisation Trials

Crystallisation trials of AIP ($H_6$- and $H_6$-SUMO-tagged in 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 5 mM DTT, with and without 10% (v/v) glycerol at a concentration of 8 $mg.mL^{-1}$) were prepared using both the hanging drop vapour diffusion technique and the sitting drop crystallisation technique (Section 4.6). The latter was performed at the University of Leeds, employing the use an NT8, a fast nanoliter-volume liquid handler with a drop-dispensing unit (Formulatrix). The following commercially available crystallisation screens were tested; Crystal Screen 1 & 2, Index HT Screen, SaltRx HT (Hampton Research), Wizard Classic 1-4 HT-96 MD15-W3-T, Morpheus HT-96, PACT Premier HT-96, Midas MD1-59 and JCSG Cores I-IV (Molecular Dimensions).

# Chapter 3

# Results: Protein Expression and Purification

## 3.1. Aryl hydrocarbon receptor interacting protein (AIP)

### 3.1.1. Generation of pETite $H_6$- SUMO-AIP by homologous recombination

The cDNA of AIP was amplified using specific PCR primers that add flanking sequences adjacent to the cloning site that correspond to either end of the linearised vector. This enabled the cloning of AIP cDNA within the pETite vector to occur by homologous recombination, without the use of enzymes.

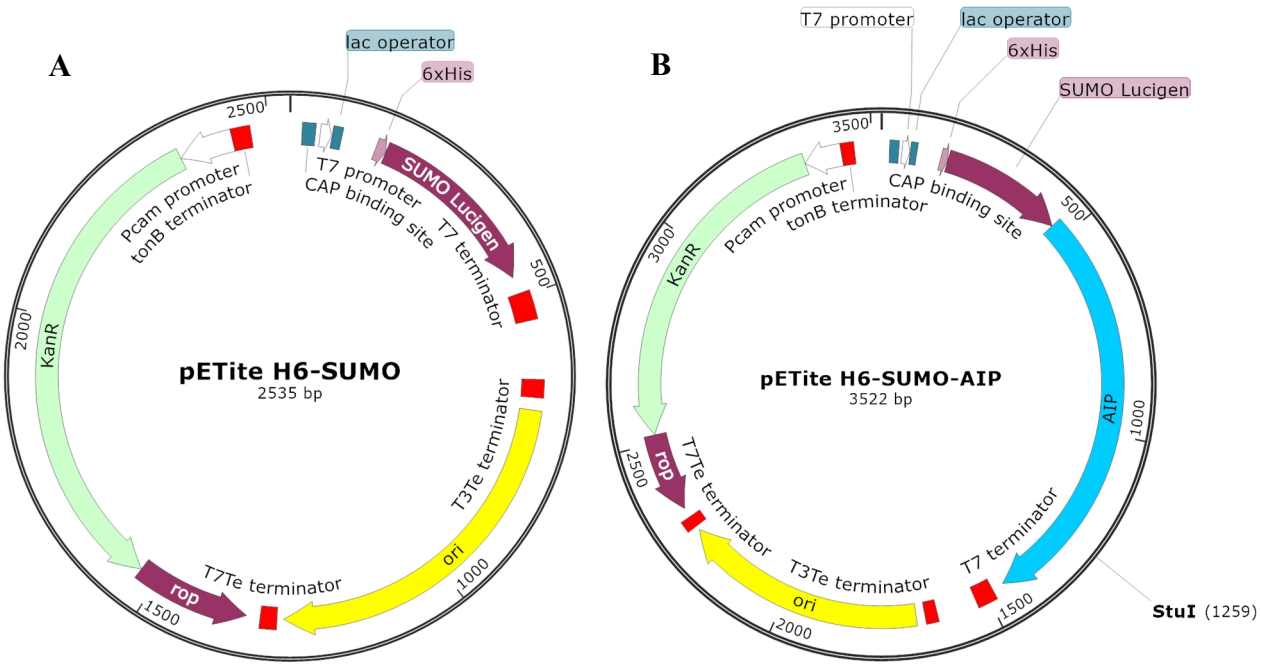



**Figure 3.1: A:pETite $H_6$-SUMO vector. B:pETite $H_6$-SUMO-AIP recombinant vector.** *In-silico* cloning of the human *AIP* coding sequence within pETite $H_6$-SUMO vector to create recombinant pETite $H_6$-SUMO-AIP. *AIP* was cloned downstream of the SUMO-tag. The StuI restriction cut site is a unique site within the *AIP* coding region. *In-silico* cloning was performed using SnapGene® Viewer.

The amplification reaction of the *AIP* cDNA was carried out using pcDNA3-hAIP as a DNA template and pETite-AIP-F and pETite-AIP-R as primers (Table 2.8). The success of the amplification and integrity of the pETite vector (Lucigen) were checked by gel electrophoresis. A PCR product of the correct size (~1,000 bp in length) was obtained as the only product.

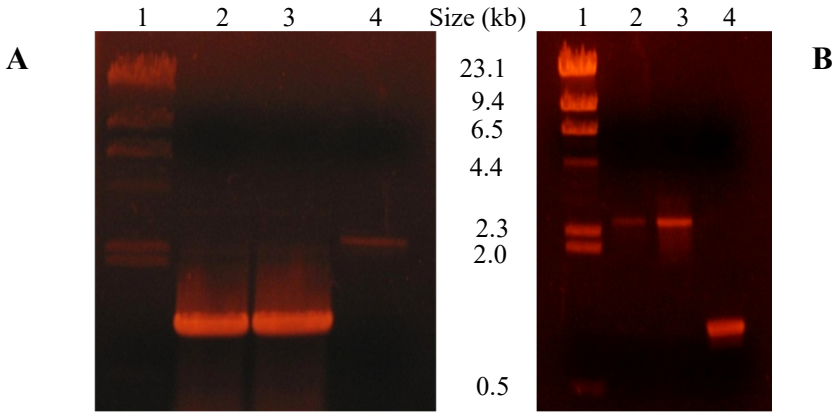



**Figure 3.2: A: Amplification of *AIP* cDNA.** Lane 1: Lambda-phage DNA HindIII digested (1 µg); Lanes 2-3: AIP PCR product; Lane 4: Linearised pETite $H_6$-SUMO vector (25 ng). **B: After gel purification.** Lane 1: Lambda-phage DNA HindIII digested (1 µg); Lanes 2-3: Gel purified linearised pETite $H_6$-SUMO vector, 25 ng, 50 ng respectively; Lane 4: Gel purified AIP PCR product (200 ng).

Randomly picked transformed colonies were grown for plasmid purification. The success of recombination was confirmed by restriction digest with StuI which is a unique cut site in the AIP cDNA and thus only found in the clone.

The use of the pETite cloning system has the advantage of maintaining a low background of empty vector transformants. In fact all twenty DNA samples analysed (only two are shown in Figure 3.3) were positive recombinant clones. Restriction digest by StuI produced a linearised DNA fragment of the correct expected size for a positive clone, 3,522 bp. Since homologous recombination is a directional type of cloning, there was only one possible orientation in which the insert could have been ligated within the vector.

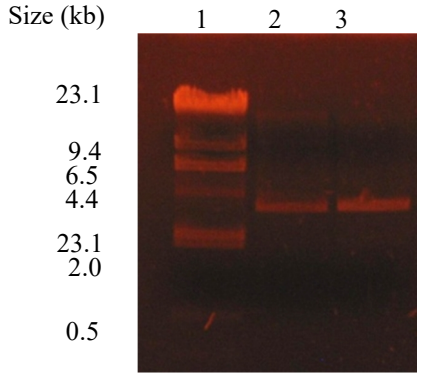



**Figure 3.3: StuI Restriction digest. Lane 1:** Lambda-phage DNA HindIII digested (1 µg); Lanes 2-3: StuI digested plasmid from two randomly chosen colonies after homologous recombination and transformation (0.2 µg).

### 3.2. Optimisation of protein expression: $H_6$-SUMO-AIP

BL21 (DE3) Codon Plus cells harbouring the recombinant pETite $H_6$-SUMO-AIP vector were grown in 10 mL cultures of 2TYmedia supplemented with kanamycin (50 µg.mL$^{-1}$). Growth and protein expression were carried out at 37°C. The optimization conditions for growth and protein expression of $H_6$-SUMO-AIP were experimentally determined (Figure 3.4

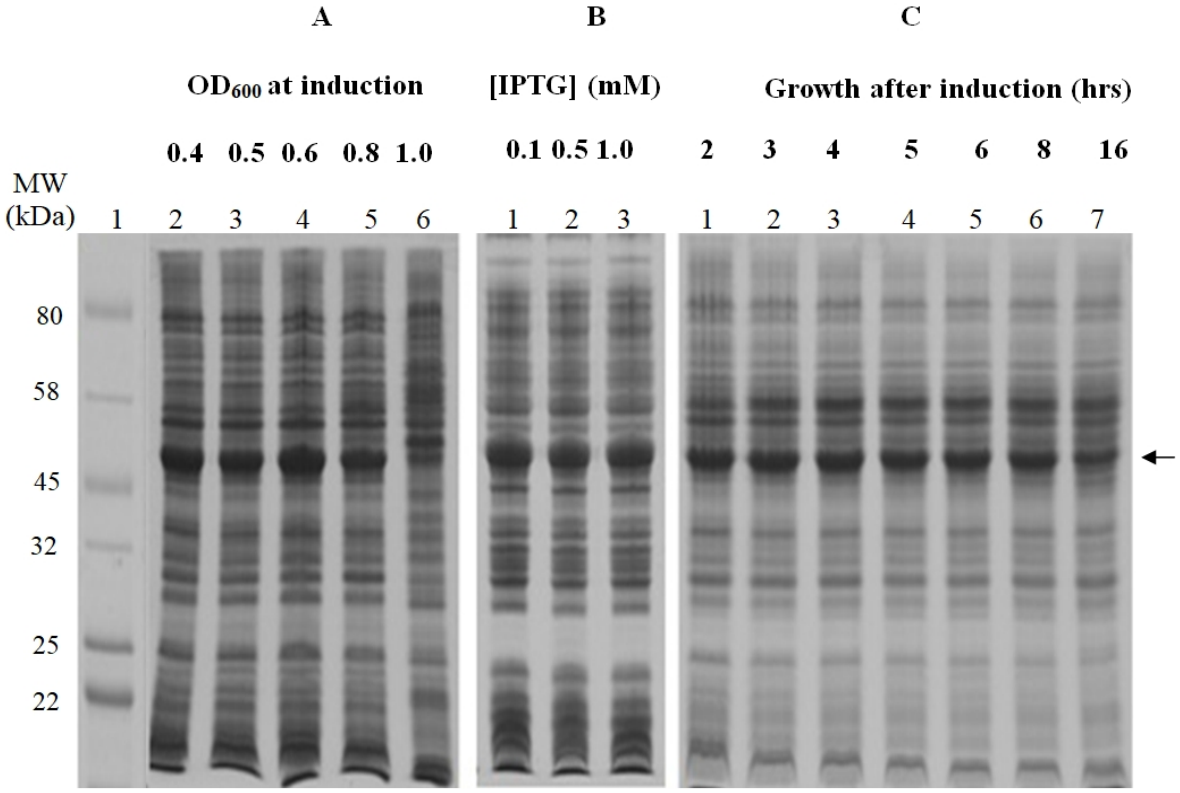



**Figure 3.4: $H_6$-SUMO-AIP Protein expression optimisation: SDS-PAGE (15%): A: $OD_{600}$ at induction:** Lane 1: Colour prestained protein standards, broad range (3 µl) (NEB); Lanes 2-6: IPTG-induced cultures at $OD_{600}$ values of 0.4-1.0 respectively. **B: Inducer (IPTG) concentration.** Lanes 1-3: Cultures induced at $OD_{600}$ of 0.6 with 0.1 mM, 0.5 mM and 1 mM IPTG respectively. In both panel A and B, induced cultures were allowed to grow a further 2 hrs at 37°C following induction **C: Growth after induction:** Lanes 1-7: Cultures induced with 0.1 mM IPTG at an $OD_{600}$ of 0.6 and grown for 2-16 hrs respectively. All samples were standardised with respect to cell growth (diluted with 2TY to the same $OD_{600}$ value), prior to loading. In each lane 10 µL of cell lysate were loaded (~ 20 µg). Over-expressed protein is indicated by an arrow.

$H_6$-SUMO-AIP appears as a distinctive, over-expressed band with an approximate molecular weight of 49kDa, as predicted from *in-silico* analysis using Protean (DNASTAR, Lasergene). When considering panel A, the ideal $OD_{600}$ was found to be 0.6, though similar protein levels were obtained at $OD_{600}$ of 0.4-0.5. No discernible band was observed when inducing at an $OD_{600}$ of 1.0 (Figure 3.4A, Lane 6), implying that expression of AIP is best initiated during the mid-log phase of bacterial growth. Similarly, when the culture was left expressing protein overnight, a less intense AIP band was observed, suggesting that overall less protein was being produced per cell or more likely that the recombinant protein was being degraded (Figure 3.4C, Lane 7). With respect to IPTG, the concentration of the inducer did not seem to have any effect on protein expression between 0.1 and 1.0 mM.

### 3.3. Preparative protein expression and purification of $H_6$-SUMO-AIP

Two liter culture flasks containing 500 mL 2TY supplemented with kanamycin (50 $\mu g.mL^{-1}$) were grown at 37°C till $OD_{600}$ of 0.5-0.6 and then induced with 0.1 mM IPTG. Protein expression was carried out at the same temperature for a further 4 h prior to harvesting. Cells were collected, lysed and clarified by centrifugation. $H_6$-SUMO-AIP was purified by IMAC using a cOmplete™ His-Tag purification column (Section 2.3.3). Unbound proteins were washed off the column using 20 mM Tris-Cl pH 7.5, 300 mM M NaCl, 12 mM imidazole. Bound protein was eluted with 25 mM imidazole.

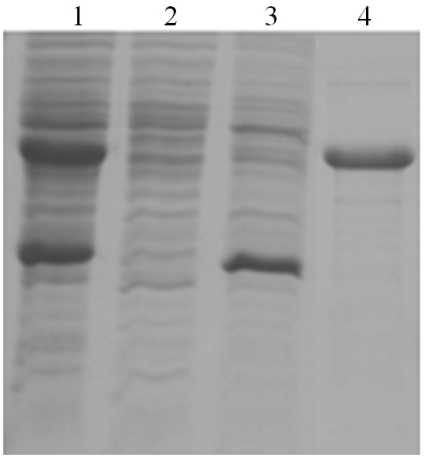



**Figure 3.5: Purification of $H_6$-SUMO-AIP: SDS-PAGE (15%) analysis:** Lane 1: Clarified lysate loaded on cOmplete™ His-Tag Purification Column (20 µg); Lane 2: Through wash (10 µg); Lane 3: Sample eluted during column washing with 12 mM imidazole (10 µg); Lanes 4: $H_6$-SUMO-AIP eluted with 25 mM imidazole (10 µg).

**Table 3.1: Purification Table: $H_6$-SUMO-AIP**

| Purification step (from 1 L culture) | Volume (mL) | [2]Concentration (mg $mL^{-1}$) | Total protein (mg) | % Purity |
|---|---|---|---|---|
| Total cell lysate | 30 | 57.0 ± 2.3 | 1.7 $x10^3$ | 8 |
| Clarified lysate | 26 | 46.1 ± 1.9 | 1.2 x $10^3$ | 13 |
| IMAC | 32 | 0.8 ± 0.1 | 27.0 | 91 |

A total of 27 mg of pure protein was obtained from a 1 L culture flask, with purity levels of 91% after one round of purification on a cOmplete$^{TM}$ HiTrap Chelating column. Purified protein was dialysed extensively against 20 mM Tris-Cl pH 7.5 and 100 mM NaCl and stored at -20°C. Some aliquots of protein were also left at 4°C.

### 3.4. Protein solubility and stability assays: $H_6$-SUMO-AIP

As AIP is a molecular co-chaperone, with a natural tendency to self-associate (Hollingshead *et al*. 2004) and bind to other proteins, the purification of the monomeric state was challenging. Signs of aggregation appeared in purified samples of $H_6$-SUMO-AIP (in 20 mM Tris-Cl pH 7.5 and 100 mM NaCl) upon short-term storage at both 4°C and -20°C (Section 3.3). Protein aggregation also occurred during centrifugal concentration as the concentration reached and exceeded 1 $mg.mL^{-1}$. In order to determine the optimum storage conditions and prevent aggregation upon storage and concentration, a number of chemical additives were assessed, to monitor the solubility and stability of the protein (Section 2.3.6). The additives analysed included: NaCl (500 mM), DTT (10 mM), sucrose (1% w/v), glycerol (10% v/v), a combination of 10% glycerol and 1% sucrose, a combination of 10% glycerol and 10 mM DTT.

[2] Protein concentration of lysate was determined through the bicinchoninic acid assay (BCA) whilst the concentration of pure samples after IMAC was determined spectrophometrically by measuring the $A_{280}$.

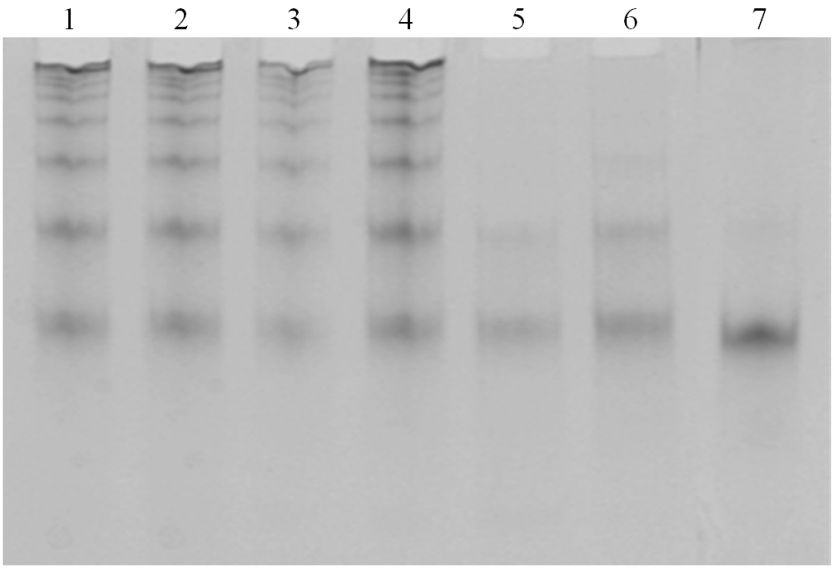



**Figure 3.6: $H_6$-SUMO-AIP aggregation assay: Native-PAGE (8%) analysis:** Lane 1: $H_6$-SUMO-AIP incubated in the absence of any additives (10 µg). Lanes 2-7: $H_6$-SUMO-AIP (10 µg) incubated overnight in the presence of 500 mM NaCl (Lane 2), 10 mM DTT (Lane 3), 1% sucrose (Lane 4), 10% glycerol (Lane 5), 10% glycerol and 1% sucrose (Lane 6) and 10% glycerol and 10 mM DTT (Lane 7).

Whilst the protein appeared pure on SDS-PAGE following purification (Figure 3.5), the use of native-PAGE gels was fundamental to assess the state of the protein. Figure 3.6, clearly shows that the protein formed various multimeric states in the absence of additives. Glycerol and DTT were found to effectively stabilise the native protein in its monomeric form.

To further investigate the stability of the protein in the presence of glycerol and/or DTT, a series of experiments were performed using an *Optim*®1000 (Section 2.3.6). Protein unfolding was measured by monitoring the changes in the intrinsic protein fluorescence spectrum as a function of temperature. Changes in fluorescence emission resulted from the exposure of tryptophan residues and other aromatic amino acids (excited at 266 nm) and differences in the polarity of their environment (Avacta Analytical).

The *Optim*®1000 was programmed to record the barycentric mean fluorescence (BCM) and the static light scattering (SLS) at two wavelengths, 266 nm and 473nm, that are sensitive to the scattering of small and larger sized aggregates respectively. BCM provides the transition temperature of the protein, designated as $T_m$, while SLS enables the determination of the temperature of aggregation ($T_{agg}$). Initially, the $T_m$ of AIP in the absence of additives was determined.

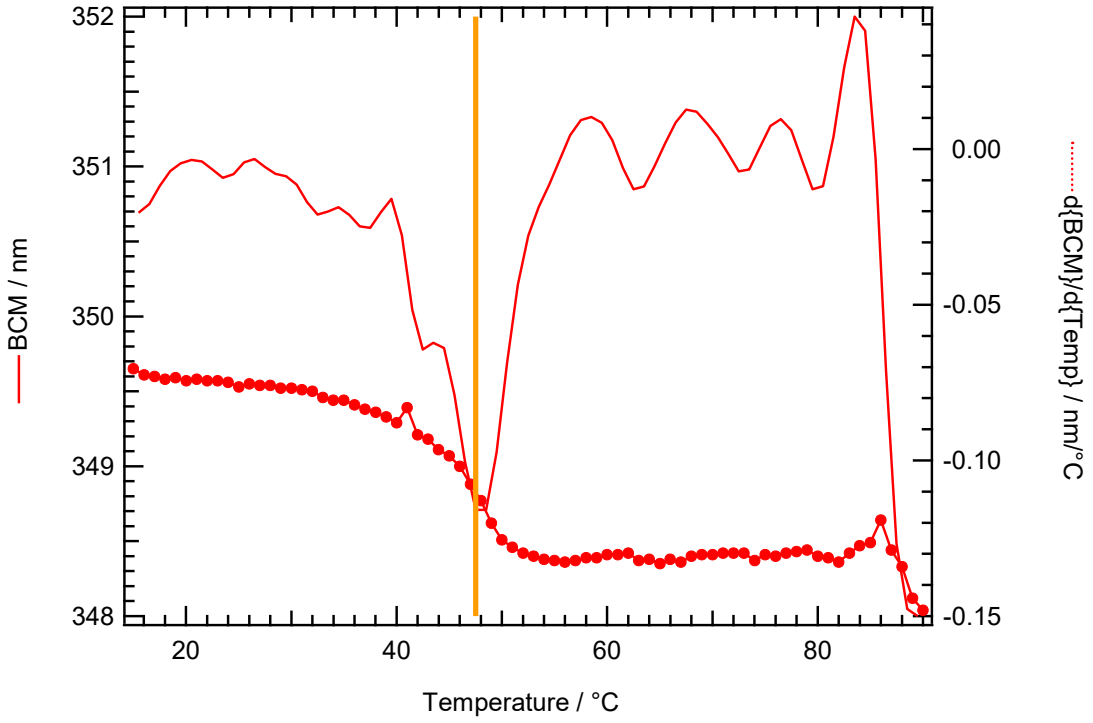



**Figure 3.7: Plot of BCM vs. Temp: Graph of intrinsic fluorescence conformational stability.**

The graph is showing the BCM in nm, (solid red circle) and the differential signal as a function of temperature (dotted line). The point where the peak occurs in the differential signal plot is equivalent to the melting temperature of the protein, represented by an orange line. In this case, only one transition state was observed which corresponds to a temperature of 47.5°C.

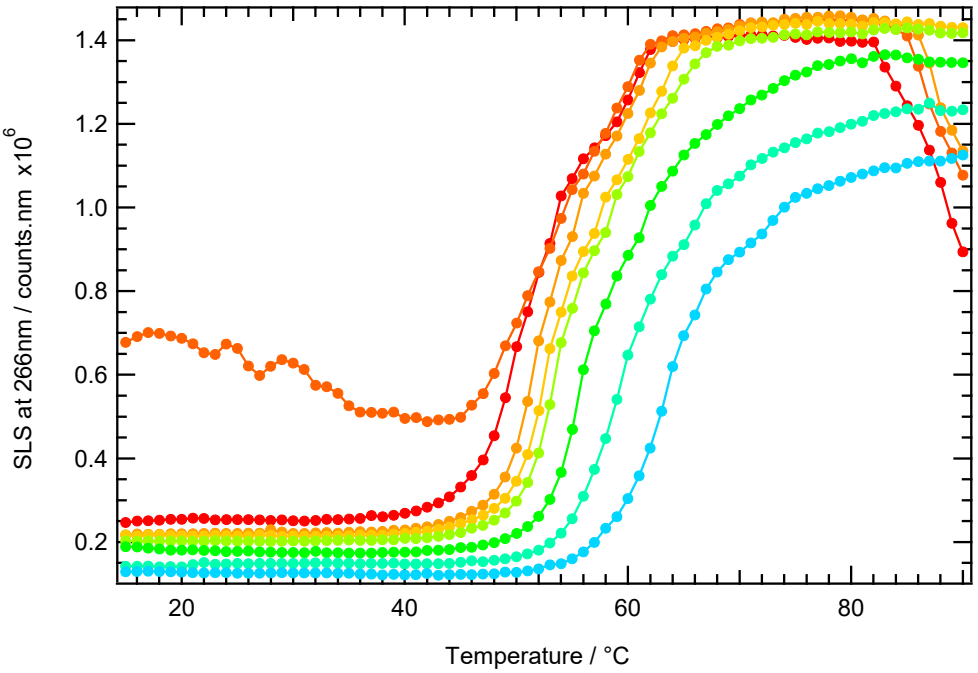



**Figure 3.8: Graph of colloidal stability and aggregation of $H_6$-SUMO-AIP in the presence of glycerol.** Static light scattering (SLS) at 266 nm was monitored with increasing concentration of glycerol ranging from 0% (red line) to 30% (cyan line) over a temperature range, 15°C to 90°C. Each sample was left incubating for 1 hr at room temperature in each of the test buffers, prior to measuring. An explanation of the curves is presented in Table 3.2.

**Table 3.2: Table of colloidal stability and aggregation (A).**

| | **$H_6$-SUMO-AIP ( diluted to 1 mg.mL$^{-1}$) was present in each well** | |
|---|---|---|
| **Well** | **Glycerol concentration (%) (v/v)** | **Temperature at onset of aggregation ($T_{agg}$) / °C** |
| **•A** | 0 | 43.5 |
| **•B** | 2.5 | 44.7 |
| **•C** | 5.0 | 46.5 |
| **•D** | 7.5 | 47.3 |
| **•E** | 10.0 | 48.1 |
| **•F** | 15.0 | 50.0 |
| **•G** | 20.0 | 52.2 |
| **•H** | 30.0 | 54.1 |

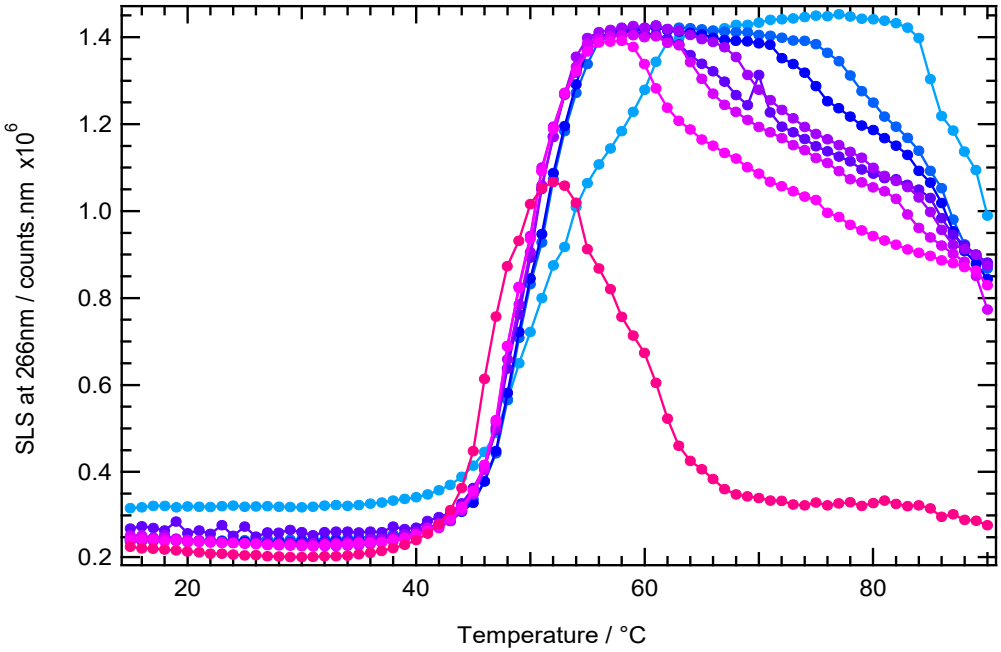



**Figure 3.9: Graph of colloidal stability and aggregation of $H_6$-SUMO-AIP in the presence of DTT.** SLS at 266 nm was monitored with increasing concentration of DTT ranging from 0 mM (cyan line) to 50 mM (pink line) over a temperature range, 15°C to 90°C. Each sample was left incubating for 1 hr at room temperature in each of the test buffers, prior to measuring. An explanation of the curves is presented in Table 3.4.

**Table 3.3: Table of colloidal stability and aggregation (B).**

| | **$H_6$-SUMO-AIP ( diluted 1 mg.mL$^{-1}$) was present in each well** | |
|---|---|---|
| **Well** | **DTT concentration (mM)** | **Temperature at onset of aggregation ($T_{agg}$) / °C** |
| **•A** | 0 | 43.3 |
| **•B** | 0.5 | 44.0 |
| **•C** | 1.0 | 43.8 |
| **•D** | 2.0 | 43.9 |
| **•E** | 5.0 | 44.5 |
| **•F** | 10.0 | 41.5 |
| **•G** | 20.0 | 43.6 |
| **•H** | 50.0 | 42.2 |

In both experiments, the increase in SLS signal was a direct measure of protein aggregation, due to an increase in particle size. In the absence of additives, the onset of protein aggregation ($T_{agg}$) for $H_6$-SUMO-AIP in 20 mM Tris-pH 7.5, 100 mM NaCl, was 43.5°C. This temperature increased with each consecutive addition of glycerol, reaching a value of 54.1°C at a glycerol concentration of 30% (v/v) (Figure 3.8 and Table 3.2). This confirmed the stabilising effect that glycerol has on AIP. A less noticeable effect was observed with DTT. The highest increase in $T_{agg}$ was of 1.2°C with 5 mM DTT. Further increase in DTT concentration did not improve the thermal stability of the protein. A low concentration of DTT (0.5 mM) had essentially the same effect as a high concentration (20 mM) clearly suggesting that DTT on its own does not have any significant effect in preventing aggregation and/or enhancing protein stability. The presence of 50 mM DTT was denaturing to the protein as it started to precipitate out of solution beyond 50°C. This is indicated by the decrease in SLS signal (Figure 3.9). However, when DTT (2 mM) was added with glycerol, the $T_{agg}$ value at each glycerol concentration was slightly higher (Table 3.4), than the values obtained when glycerol was the sole additive in the buffer (Table 3.2).

**Table 3.4: Table of colloidal stability and aggregation (C)**

| | **$H_6$-SUMO-AIP ( diluted 1mg.mL$^{-1}$) was present in each well** | |
|---|---|---|
| **Well** | **Sample conditions** | **Temperature at onset of aggregation ($T_{agg}$) / °C** |
| **A** | 2.5% glycerol, 2 mM DTT | 48.5 |
| **B** | 5.0% glycerol, 2 mM DTT | 49.1 |
| **C** | 10.0% glycerol, 2 mM DTT | 50.5 |
| **D** | 15.0% glycerol, 2 mM DTT | 51.1 |
| **E** | 20.0% glycerol, 2 mM DTT | 53.2 |
| **F** | 25.0% glycerol, 2 mM DTT | 55.1 |
| **G** | 30.0% glycerol, 2 mM DTT | 57.3 |

Subsequent purifications of AIP and AIP mutants were therefore carried out in the presence of these additives, with 5 mM DTT added during cell lysis and 10% (v/v) glycerol added after sonication. The same additives were also included in the storage buffer; however, in this case, the final concentration of DTT was reduced to 2 mM.

### 3.5. Concentration of $H_6$-SUMO-AIP

Protein characterisation requires relatively high amounts of protein. Isothermal titration calorimetry and small angle X-ray scattering both require protein concentrations that range between 4 to 7 mg $mL^{-1}$. The success of both characterisation techniques relies on the homogeneity of the sample. Multimeric protein states, especially aggregates significantly interfere with the quality of the data.

The concentration of AIP was a monitored process and aliquots of concentrated protein were taken at regular intervals to assess the state of the protein throughout. Samples of pure $H_6$-SUMO-AIP were concentrated using Amicon Ultra centrifugal filter units, in the presence of 10% (v/v) glycerol and 2 mM DTT. The importance of these additives was determined experimentally (Section 3.4).

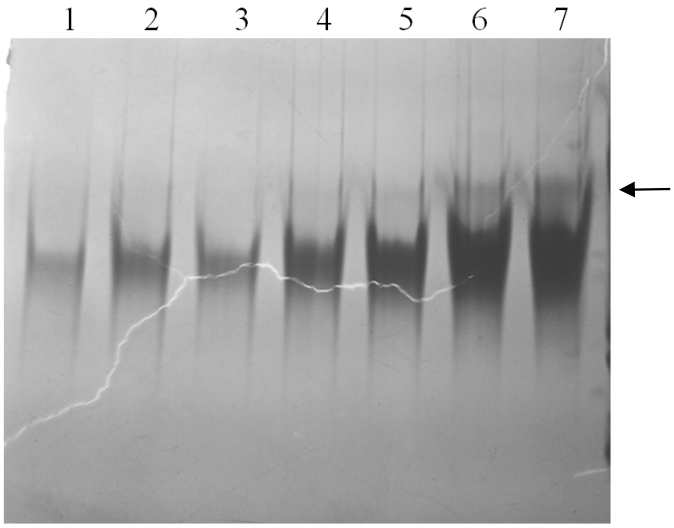



**Figure 3.10: Monitoring the concentration of $H_6$-SUMO-AIP**: **Native-PAGE (8%) analysis:** The gel shows $H_6$-SUMO-AIP at increasing concentrations ranging from 1 $mg.mL^{-1}$ (Lane 1) to 7 $mg.mL^{-1}$ (Lane 7). The presence of a major band indicates monomeric protein and that the additives were successful in preventing protein aggregation. Lanes 6 and 7 indicate the possibility of multimerisation at these concentrations (shown by arrow). Lane number also indicates the concentration of protein in $mg.mL^{-1}$.

### 3.6. Gel filtration chromatography

To ensure that the protein sample is homogeneous and monomeric, concentrated $H_6$-SUMO-AIP was loaded on a gel filtration column (S75) connected to an AKTA Prime chromatography unit (Section 2.3.7). The running buffer was 20 mM Tris-Cl pH 7.5, 0.1 M NaCl, 5% (v/v) glycerol and 2 mM DTT.

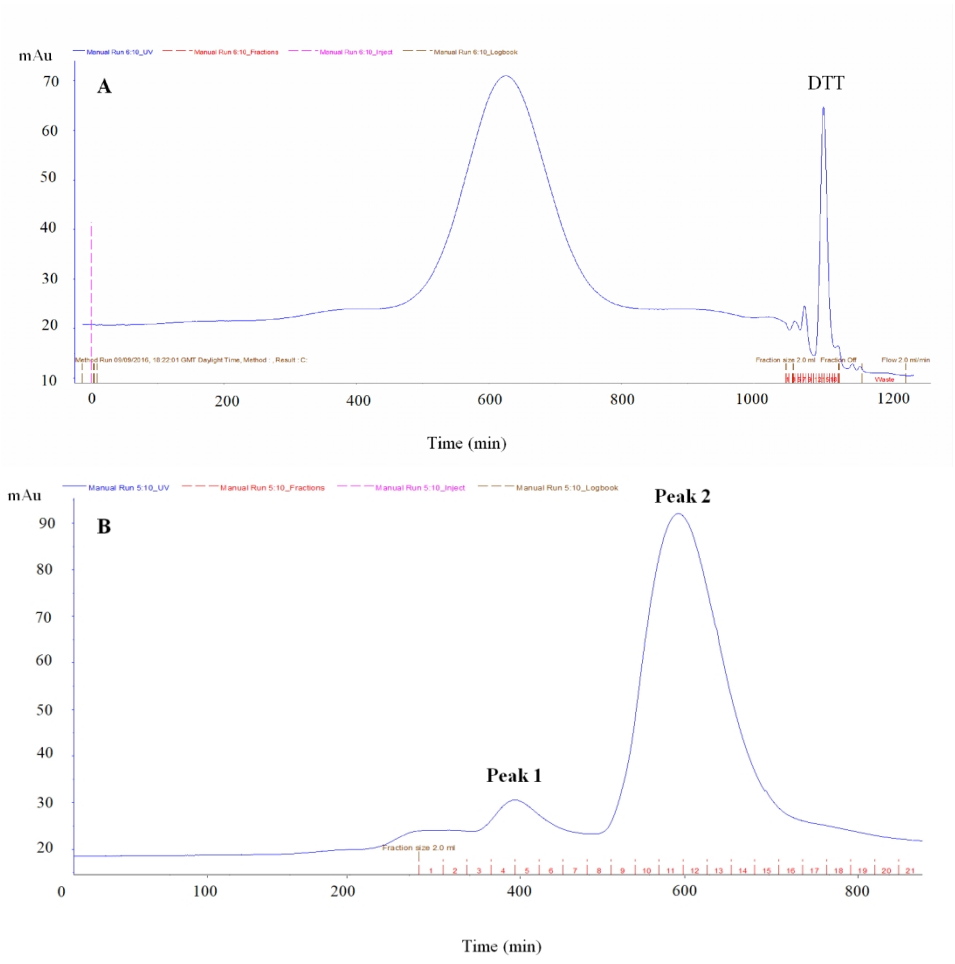



**Figure 3.11: Gel filtration chromatogram of $H_6$-SUMO-AIP.** **A:** Sample at 4 mg.mL$^{-1}$; **B:** Sample at 7 mg.mL$^{-1}$.

In the first chromatogram (A), only one major peak was eluted, indicating a pure homogeneous monomeric sample. The peak at the far right of the spectrum is due to the presence of DTT that also absorbs at 280 nm. The second chromatogram (B) shows two peaks, a low intensity peak 1 with an estimated molecular weight of 140 kDa and peak 2 with an estimated molecular weight of 50 kDa, the latter corresponding to the monomeric form of $H_6$-SUMO-AIP. The DTT peak is not shown in chromatogram B as the image was taken prior to the elution of DTT.

### 3.7. Removal of the SUMO-tag

The $H_6$-SUMO-tag was removed using the SUMO Express Protease (Section 2.3.5. The cleavage efficiency was analysed on 15% SDS-PAGE and the state of the resulting protein was visually assessed on Native-PAGE (Figure 3.12).

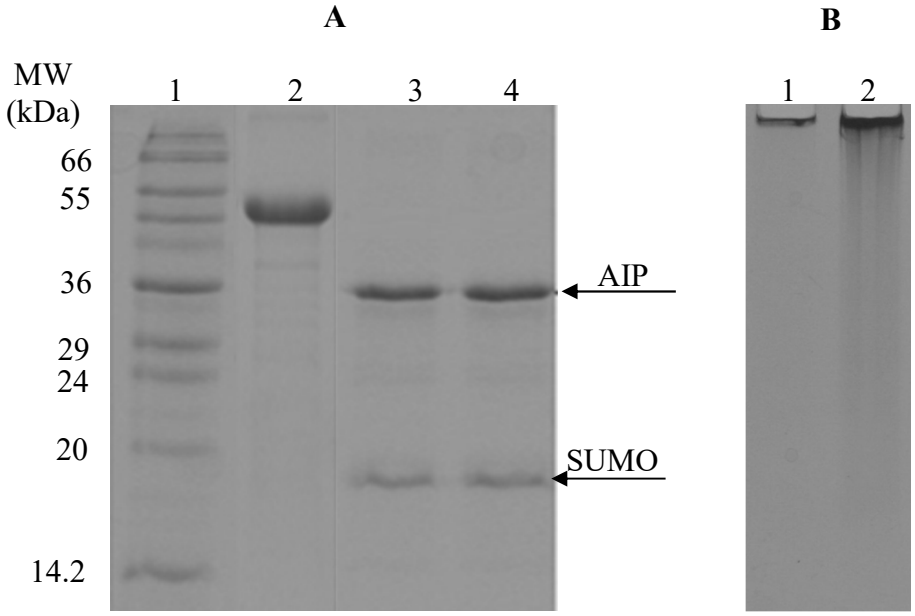



**Figure 3.12: Removal of the SUMO-tag**. **A: SDS-PAGE (15%) analysis:** Lane 1: Low range molecular weight standards, (3 µL) (Sigma-Aldrich); Lane 2: $H_6$-SUMO-AIP (5µg); Lanes 3-4; Sample after cleavage containing AIP and SUMO (5µg). **B: Native-PAGE analysis (8%)**. Lanes 1-2: Untagged AIP protein (5 µg and 10µg respectively).

The tag was successful cleaved, obtaining pure untagged AIP. The free SUMO tag has an expected molecular weight of 12 kDa, however on SDS-PAGE it migrates as a 15-18 kDa band (Expresso T7 SUMO cloning Expresso T7 SUMO cloning Lucigen Manual) as observed in Figure 3.12A. Nonetheless, despite the success in removing the tag, the solubility and stability of the protein was adversely affected. Native-PAGE analysis showed that the untagged protein was significantly aggregated to the point that it remained in the wells and failed to enter the resolving gel, despite the presence of the stabilising additives. Therefore, for functional and structural studies the SUMO-tag was retained and where necessary SUMO-protein[3] was used as a control.

[3] Pure SUMO-protein was a generously provided by Dr Chi Trinh, University of Leeds.

### 3.7.1. Generation of pTH-H$_6$-AIP by blunt-end sub-cloning

Similar to the previous cloning procedure, pcDNA3-hAIP was used as DNA template for the PCR experiment (Section 2.2.2).

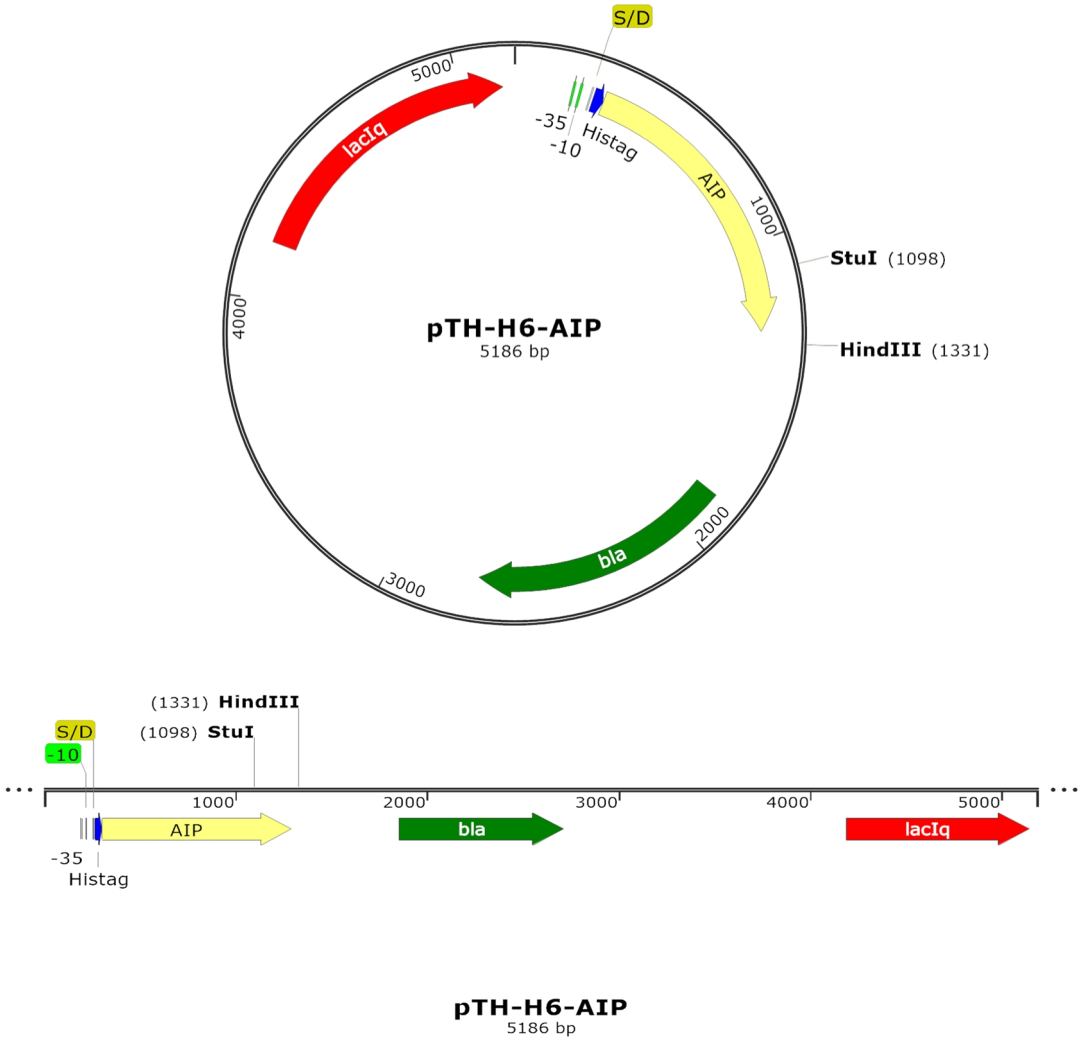



**Figure 3.13: pTH-H$_6$-AIP recombinant clone.** *In-silico* cloning of the human AIP coding sequence within pTH-1 vector. The StuI restriction cut site is a unique site within the AIP coding region, while HindIII recognizes a unique site within the pTH-1 vector (SnapGene® Viewer).

The forward and reverse amplification primers used were AIP-F and AIP-R (Table 2.8). A PCR product of approximately 1,000 bp was obtained, gel purified and ligated with gel purified StuI digested vector. The success of blunt-end cloning was analysed by gel electrophoresis (Figures 3.14 and 3.15)

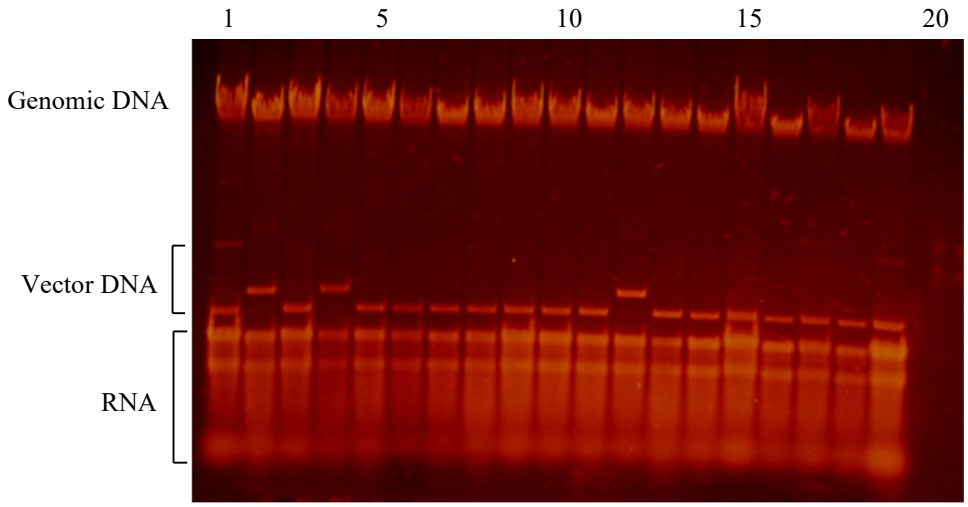



**Figure 3.14: Mini-prep analysis of transformants.** Lane 1 and 19 contain pTH-1 vector as a control. Lanes 2-18 contain nucleic acid extracted from bacterial colonies that were transformed with the ligation mixture.

The gel shows that the vector DNA in lanes 2, 4 and 12 contains insert. Positive recombinant clones can be identified by their larger size (5,186 bp) when compared to control pTH-1 vector (4,196 bp) in lanes 1 and 19. A total of 50 colonies were analysed of which only 3 (shown in Figure 3.14) were positive. The size of the recombinant pTH-$H_6$-AIP vector was further confirmed by restriction digest (Figure 3.15).

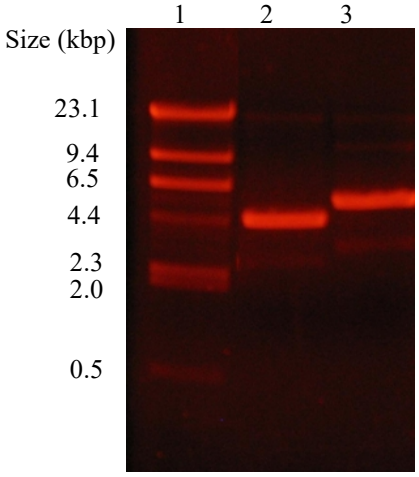



**Figure 3.15: Confirmation of insert by restriction digest analysis.** Lambda-phage DNA HindIII digested (1 µg); Lane 2: pTH-1 vector digested with HindIII (0.5 µg). Lane 3: Recombinant pTH-$H_6$-AIP digested with StuI (0.5 µg).

As expected, the DNA in lane 3 is larger in size (5,186 bp) than the DNA in Lane 2 (4,196 bp).

Given the non-directionality of blunt-end cloning, a positive clone does not necessarily indicate a functional one, as the DNA insert may be incorporated in the wrong orientation. In order to ascertain that the DNA insert was cloned in the correct orientation and was functional, the positive colonies shown in Figure 3.14 were grown in 10 mL of 2TY media and assessed for protein expression following IPTG induction.

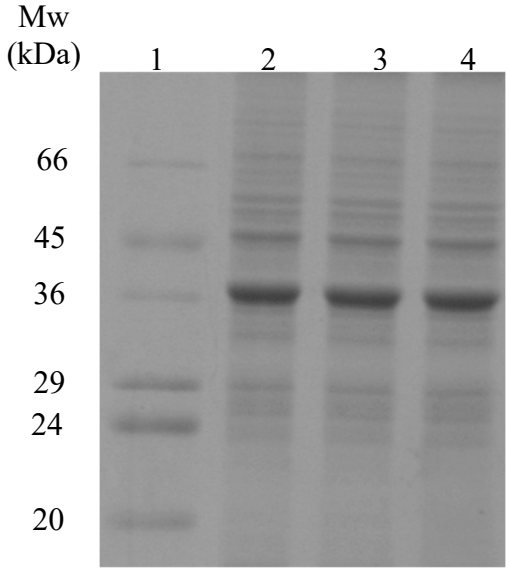



**Figure 3.16: Confirmation of $H_6$-AIP protein expression: SDS-PAGE (15%) analysis:** Lane 1: Low range molecular weight markers (3 µL) (Sigma-Aldrich); Lanes 2-4: Lysate from IPTG-induced cells (20 µg).

Induced cultures that harboured the recombinant positive clone produced an over-expressed band of an approximate molecular weight of 39 kDa. This coincides with the expected molecular weight of $H_6$-AIP, as determined from the protein sequence using Protean (DNASTAR, Lasergene).

The integrity of the DNA inserts of all the positive clones obtained for pETite-$H_6$-SUMO-AIP and pTH-$H_6$-AIP was assessed by Sanger sequencing (Bioneer). The plasmid DNA was extracted as described in Section 2.1.5. For pETite-$H_6$-SUMO-AIP DNA, the sequencing primers used were SUMO-F, SUMO-R and T7TERM (Table 2.8). These primers enabled the sequencing of the *SUMO* and the *AIP* DNA in both directions. The primers designated as PKPRO and PKTERM were used to sequence the

DNA of pTH-$H_6$-AIP in both directions (Table 2.8). The full nucleotide sequence of *AIP* is found in Appendix A.1.

After confirmation of the DNA sequences, the DNA of both pETite-$H_6$-AIP and pTH-$H_6$-AIP served as templates for site-directed mutagenesis (Section 2.2.8). Four missense N-domain mutations designated as R9Q, R16H V49M and K103R were successfully introduced using each pair of mutagenic primers, AIP[R9Q]-F and AIP[R9Q]-R; AIP[R16H]-F and AIP[R16H]-R; AIP[V49M]-F and AIP[V49M]-R; AIP[K103]-F and AIP[K103R]-R respectively. The sequence of each primer is given Table 2.8. After each mutagenic reaction and transformation with DpnI treated DNA, ten transformed colonies were randomly picked and grown for plasmid DNA extraction (Section 2.1.5). The DNA was prepared as described in Section 2.1.7, and sent for Sanger DNA sequencing (Bioneer).

Sequencing analyses revealed that in the case of R9Q, R16H and K103R each of the ten colonies investigated had DNA harbouring the mutation of interest. In the case of V49M, only one colony had wild-type DNA, while the DNA from the other nine colonies had the required mutation at position 49. Based on the number of colonies investigated, the mutagenic efficiency ranged between 90-100%.

Alignment of the DNA sequence of AIP-mutants with the wild type *AIP* sequence, confirmed that mutations only occurred at the site of interest and that no additional mutations were mistakenly introduced during the SDM reaction. Results of the analyses are shown in Figures 3.17-3.19 respectively.

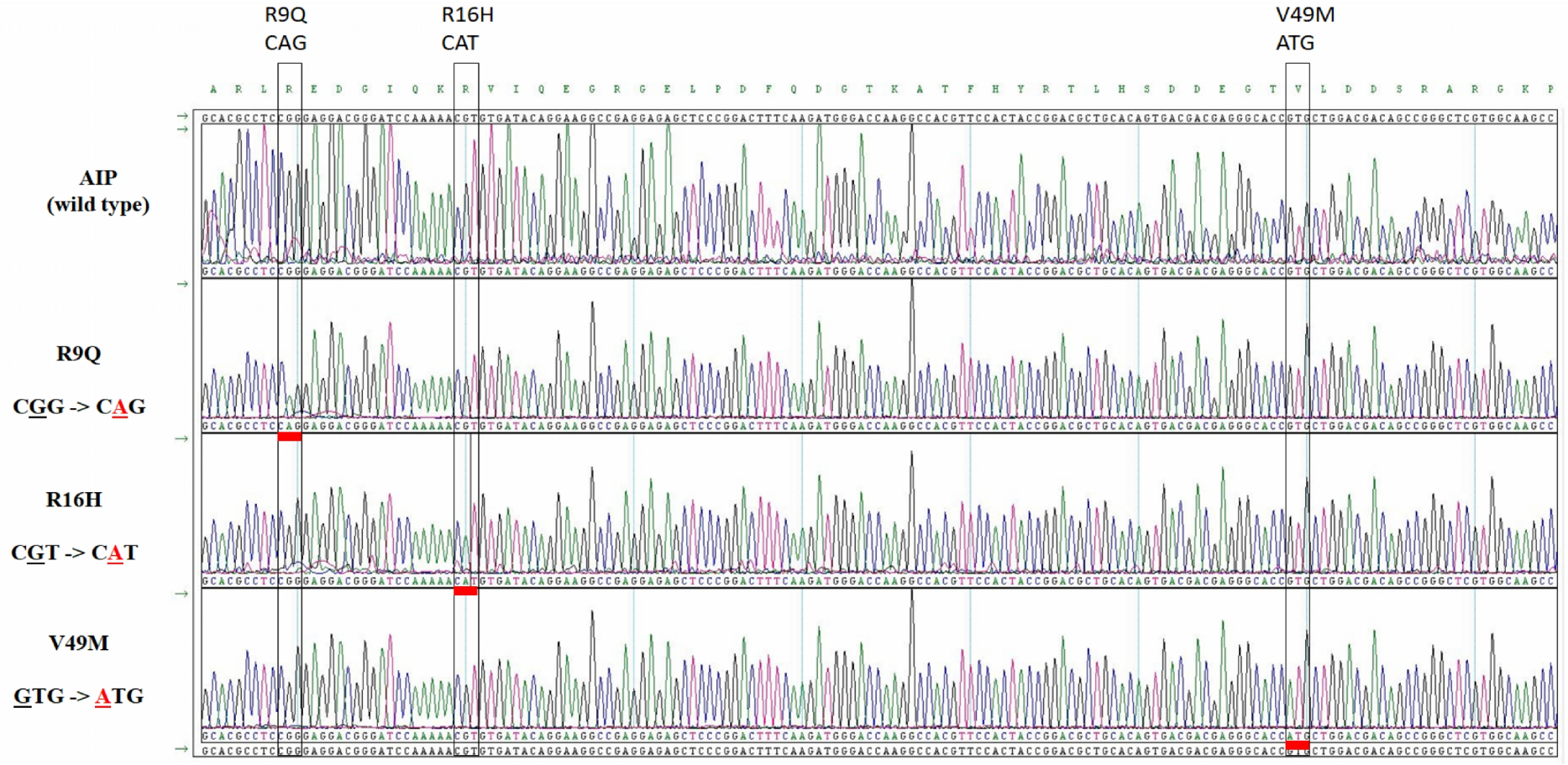



**Figure 3.17: DNA sequencing results of $H_6$-SUMO-AIP wild type and mutants R9Q, R16H and V49M.** The mutated codons are boxed and underlined.

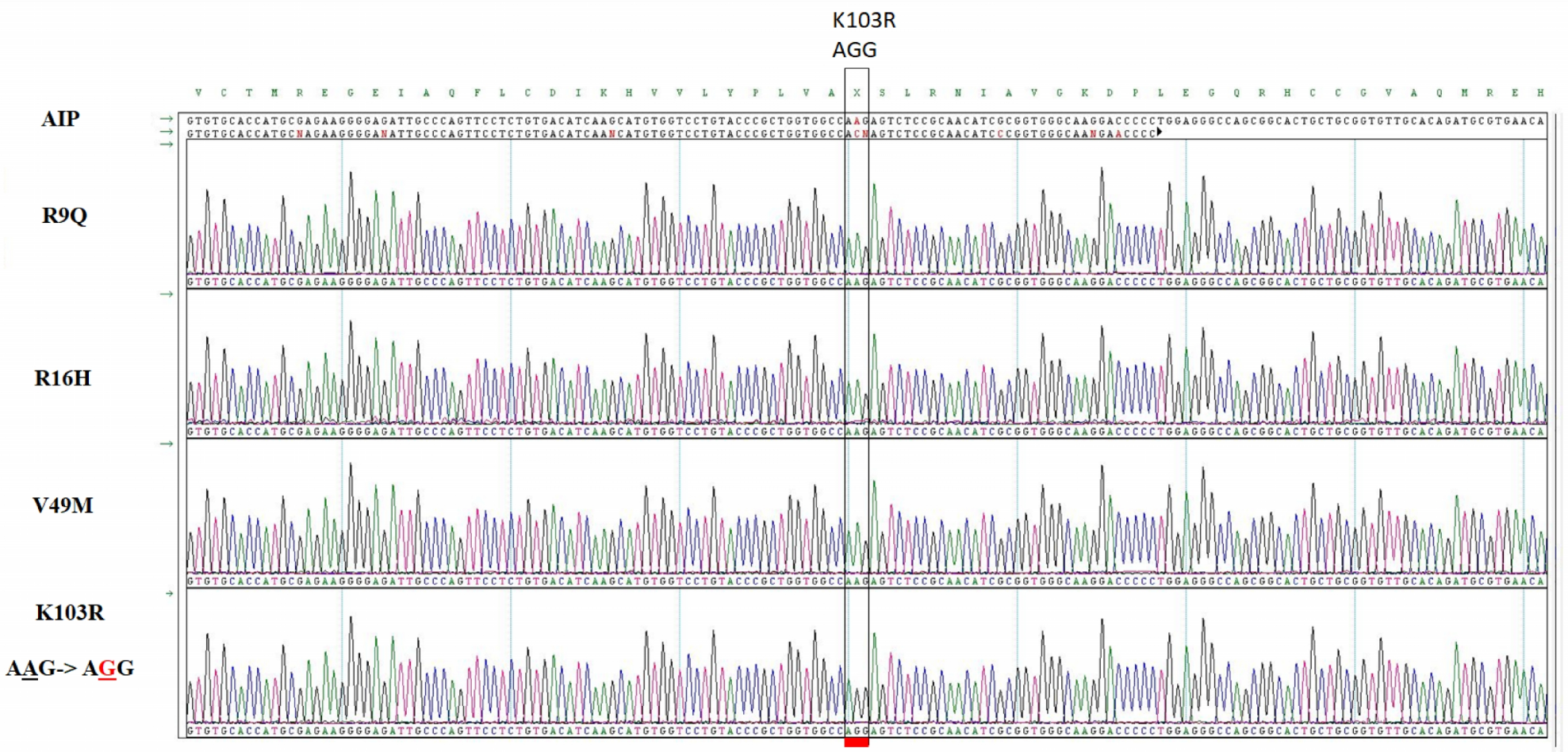



**Figure 3.18: DNA sequencing results of $H_6$-SUMO-AIP wild type and mutant K103R.** The mutated codon is boxed and underlined. A similar alignment for the $H_6$-AIP and mutant counterparts can be found in Appendix A.1.

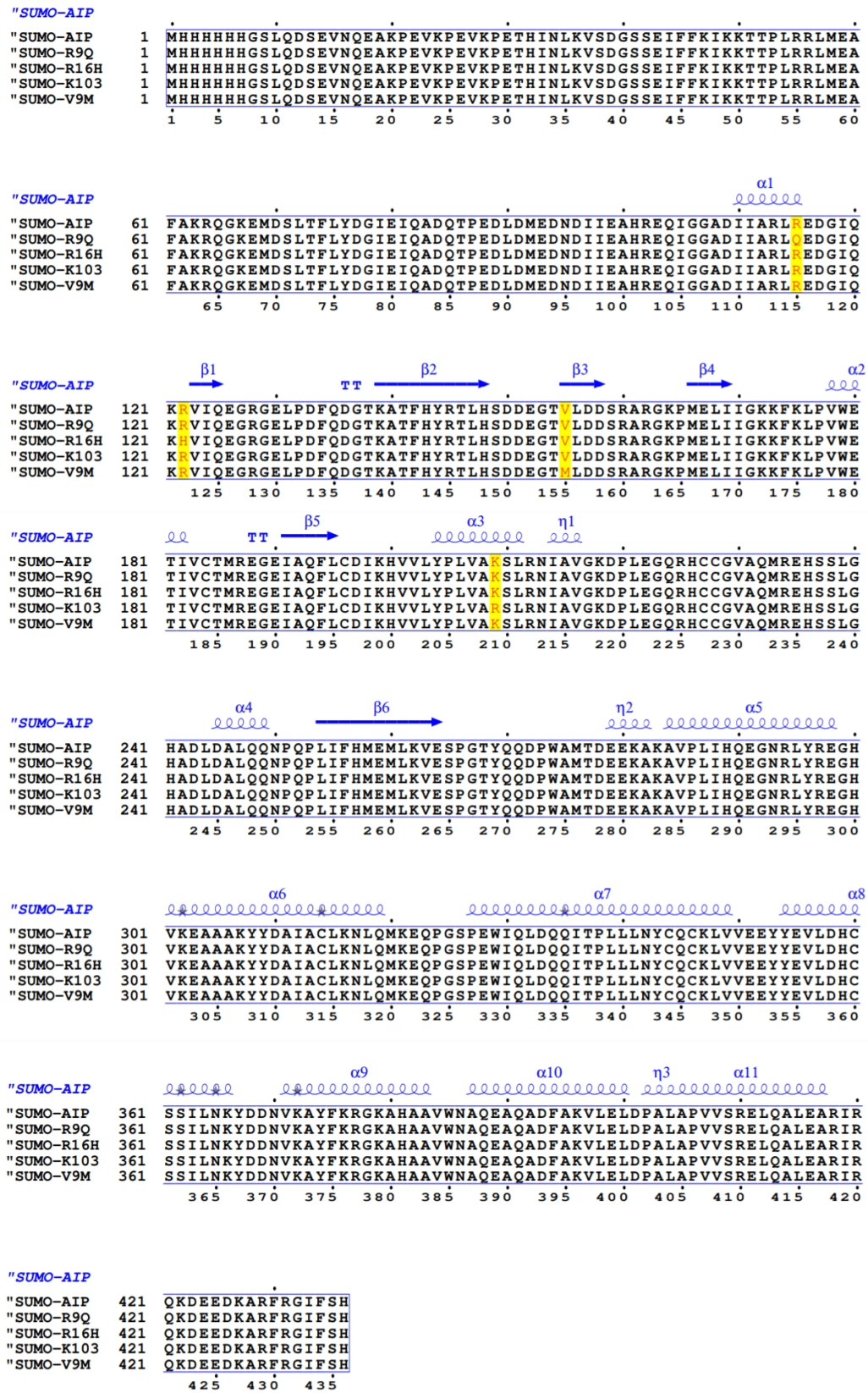



**Figure 3.19: Alignment of $H_6$-SUMO-AIP and mutant $H_6$-SUMO-AIP amino acid sequences.** The alignment was generated using the online server ESPript version 3.0 (Robert and Gouet, 2014).

### 3.8. Protein expression and purification of $H_6$-AIP

The optimisation conditions for growth and protein expression of $H_6$-AIP were identical to the ones used for $H_6$-SUMO-AIP (Section 3.2), with the exception of adding ampicillin (100 µg.mL$^{-1}$), instead of kanamycin to the growth media. Cultures were induced with IPTG to a final concentration of 0.1 mM and protein expression carried out for 4 h. DTT (5 mM) and glycerol (10% v/v) were added to the clarified lysate and loaded on a cOmplete$^{TM}$ His-Tag purification column (Section 2.3.3). $H_6$-AIP protein was eluted with 25 mM imidazole. Native-PAGE demonstrates that under these purification conditions AIP does not aggregate and immunoblotting confirmed the presence of the $H_6$-tagged protein.

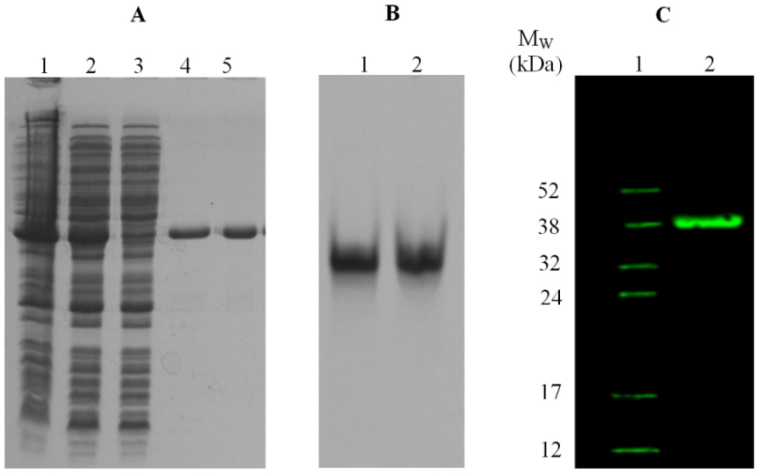



**Figure 3.20: Purification of $H_6$-AIP: A. SDS-PAGE (15%) analysis**: Lane 1: Total cell lysate (20 µg), Lane 2: Clarified lysate (20 µg); Lane 3: Colum through wash (15 µg); Lanes 4-5: Purified $H_6$-AIP fractions (5 µg). **B. Native-PAGE (8%) analysis**: Lanes 1-2: purified $H_6$-AIP (5 µg). **C: Immunoblotting using anti-his tag as 1° antibody:** Lane 1: Rainbow molecular weight markers (5 µL) (GE Healthcare), Lane 2: Pure $H_6$-AIP (10 µg).

**Table 3.5: Purification table of $H_6$-AIP**

| Purification step (from 1 L culture) | Volume (mL) | Concentration (mg.mL$^{-1}$) | Total protein (mg) | Purity (%) |
|---|---|---|---|---|
| Total cell lysate | 30 | 81.2 ± 1.4 | 2.4 x10$^3$ | 8 |
| Clarified lysate | 26 | 75.1 ± 0.5 | 1.9 x10$^3$ | 11 |
| IMAC | 30 | 1.4 ± 0.1 | 42.0 | 93 |

### 3.9. Purification of AIP mutants

BL21 (DE3) Codon Plus cells harbouring mutant DNA (pTH-$H_6$[R9Q], pTH-$H_6$[R16H], pTH-$H_6$[V49M] and pTH-$H_6$[K103R]) were grown in separate culture flasks as described in Section 3.2. Similarly, the purification of AIP mutants, both $H_6$- and $H_6$-SUMO tagged, was carried out using the same methodology described for the wild type AIP protein, in the presence of DTT (5 mM) and glycerol (10% v/v) as additives (Section 2.3.3). Purified protein was extensively dialysed against storage buffer (Table 2.9) and stored at -20°C at a final concentration of 1 $mg.mL^{-1}$.

**Table 3.6: Purification table of-$H_6$-SUMO-AIP mutants**

| Purification step (from 1 L culture) | Volume (mL) | Concentration ($mg.mL^{-1}$) | Total protein (mg) | Purity (%) |
|---|---|---|---|---|
| | | **R9Q** | | |
| Total cell lysate | 30 | 53.3 ± 2.3 | 1.6 $x10^3$ | 7 |
| Clarified lysate | 26 | 50.1 ± 0.9 | 1.3 $x10^3$ | 11 |
| IMAC- Round 1 | 25 | 1.1 ± 0.1 | 27.5 | 78 |
| IMAC-Round 2 | 25 | 0.9 ± 0.1 | 22.5 | 88 |
| | | **R16H** | | |
| Total cell lysate | 30 | 46.9 ± 1.2 | 1.4 $x10^3$ | 5 |
| Clarified lysate | 26 | 40.1 ± 1.1 | 1.0 $x10^3$ | 9 |
| IMAC- Round 1 | 25 | 0.9 ± 0.1 | 22.5 | 82 |
| IMAC-Round 2 | 25 | 0.6 ± 0.3 | 15.0 | 89 |
| | | **V49M** | | |
| Total cell lysate | 30 | 44.5 ± 1.1 | 1.3 $x10^3$ | 7 |
| Clarified lysate | 27 | 34.2 ± 0.8 | 0.9 $x10^3$ | 9 |
| IMAC-Round 1 | 25 | 0.8 ± 0.5 | 20.0 | 79 |
| IMAC-Round 2 | 25 | 0.5 ± 0.1 | 12.5 | 88 |
| | | **K103R** | | |
| Total cell lysate | 30 | 60.2 ± 2.3 | 1.8 $x10^3$ | 7 |
| Clarified lysate | 27 | 57.1 ± 1.3 | 1.5 $x10^3$ | 10 |
| IMAC-Round 1 | 25 | 1.2 ± 1.3 | 30.0 | 83 |
| IMAC-Round 2 | 25 | 1.0 ± 0.4 | 25.0 | 90 |

**Table 3.7: Purification table of $H_6$-AIP mutants**

| Purification step (from 1 L culture) | Volume (mL) | Concentration ($mg.mL^{-1}$) | Total protein (mg) | Purity (%) |
|---|---|---|---|---|
| | | **R9Q** | | |
| Total cell lysate | 30 | 77.2 ± 1.4 | 2.3 $x10^3$ | 9 |
| Clarified lysate | 27 | 74.1 ± 1.6 | 2.0 $x10^3$ | 11 |
| IMAC | 25 | 1.7 ± 0.1 | 42.6 | 91 |
| | | **R16H** | | |
| Total cell lysate | 30 | 61.3 ± 1.1 | 1.8 $x10^3$ | 8 |
| Clarified lysate | 25 | 50.1 ± 2.1 | 1.3 $x10^3$ | 9 |
| IMAC | 25 | 0.9 ± 0.2 | 23.5 | 89 |
| | | **V49M** | | |
| Total cell lysate | 30 | 60.8 ± 1.9 | 1.8 $x10^3$ | 7 |
| Clarified lysate | 25 | 49.1 ± 0.5 | 1.2 $x10^3$ | 10 |
| IMAC | 20 | 1.4 ± 0.2 | 28.0 | 90 |
| | | **K103R** | | |
| Total cell lysate | 30 | 84.2 ± 0.9 | 2.5 $x10^3$ | 10 |
| Clarified lysate | 27 | 82.8 ± 0.8 | 2.2 $x10^3$ | 11 |
| IMAC | 25 | 1.5 ± 0.1 | 37.5 | 90 |

AIP and its mutant counterparts were successfully expressed and purified to homogeneity from bacterial hosts, reaching purity levels ranging from 88-93%. $H_6$-tagged proteins reached their maximal purity levels after one round of purification on the IMAC column, whereas $H_6$-SUMO-tagged counterparts required two consecutive rounds to attain similar levels of purity. In between runs, the protein was dialysed against 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 10% (v/v) glycerol and 1 mM DTT to remove the imidazole present in the elution buffer. For each mutant, higher protein levels were obtained when using the $H_6$-tagged variants rather than the $H_6$-SUMO-tagged proteins. For both tags, the lowest protein yield was obtained for R16H and V49M. The final yield of the other mutants, R9Q and K103R was comparable to that of the wild type, AIP. Each protein was analysed for purity and solubility by SDS-PAGE and Native-PAGE respectively prior to the functional studies. Purified protein was dialysed against storage buffer (Table 2.9) and stored in 1 mL aliquots at -20°C. Despite

the presence of additives in the storage buffer, the final concentration of the stored AIP and AIP mutants did not exceed the 1 $mg.mL^{-1}$. When a higher concentration was required, the protein was concentrated as described in Section 3.5, prior to use.

### 3.10. Purification of Hsp90β $^{(513-724)}$

Hsp90β$^{(513-724)}$ was kindly provided by Professor Thomas Ratajczak (Department of Endocrinology & Diabetes, Western Australia) as part of pET28a expression construct. This construct is compatible with a bacterial system and as such, no cloning / sub-cloning procedures were required. Protein expression and purification were carried as reported by Ward *et al*., 2002, with some modification.

2TY culture flasks supplemented with kanamycin (50 $\mu g.mL^{-1}$) were grown at 37°C till $OD_{600}$ of 0.5 and induced with 0.4 mM IPTG. Growth was allowed for a further 3 h at the same temperature. Cells were harvested and lysed, DTT (5 mM) and glycerol (10% v/v) were added after sonication and the lysate was clarified by centrifugation. Hsp90β$^{(513-724)}$ was purified by IMAC (Section 2.3.3). Unbound proteins were washed through the column by 20 mM Tris-Cl pH 7.5, 0.3 M NaCl, 0.2% (v/v) Triton X-100, 10 mM imidazole. Elution of the Hsp90β$^{(513-724)}$ fraction was achieved with 25 mM imidazole. Pure protein was achieved after two successive rounds of purification on HiTrap Chelating Column. The sample eluted from the first round in 25 mM imidazole was dialysed against 10 mM Tris-Cl pH 7.5, 100 mM NaCl, 5 mM DTT prior to re-loading on the second column. Purified protein was dialysed against storage buffer (Table 2.9) and stored at -20°C in 1 mL aliquots at 2 $mg.mL^{-1}$, final concentration.

Hsp90β$^{(513-724)}$ was successfully expressed in *E.coli* BL21(DE3) cells and purified by two rounds of IMAC purification. A total of 29 mg of purified protein was extracted from a one liter culture at 95% purity. The purity was confirmed by 15% SDS-PAGE, while 8% Native-PAGE demonstrated the homogeneity of the purified protein. The estimated molecular weight of the pure protein from SDS-PAGE was 24 kDa. This is equivalent to the size of the monomer as predicted from the protein sequence using Protean (DNASTAR, Lasergene). However, the molecular weight as determined from gel filtration was that of 48-49 kDa. This confirmed that the protein purified as a dimer. Physiologically Hsp90 does exist as a homodimer and the ability to dimerise is

mediated by the C-domain of the protein, residues 599-724 (Pearl, 2016). Therefore, it was expected that the protein will purify in its dimeric form.

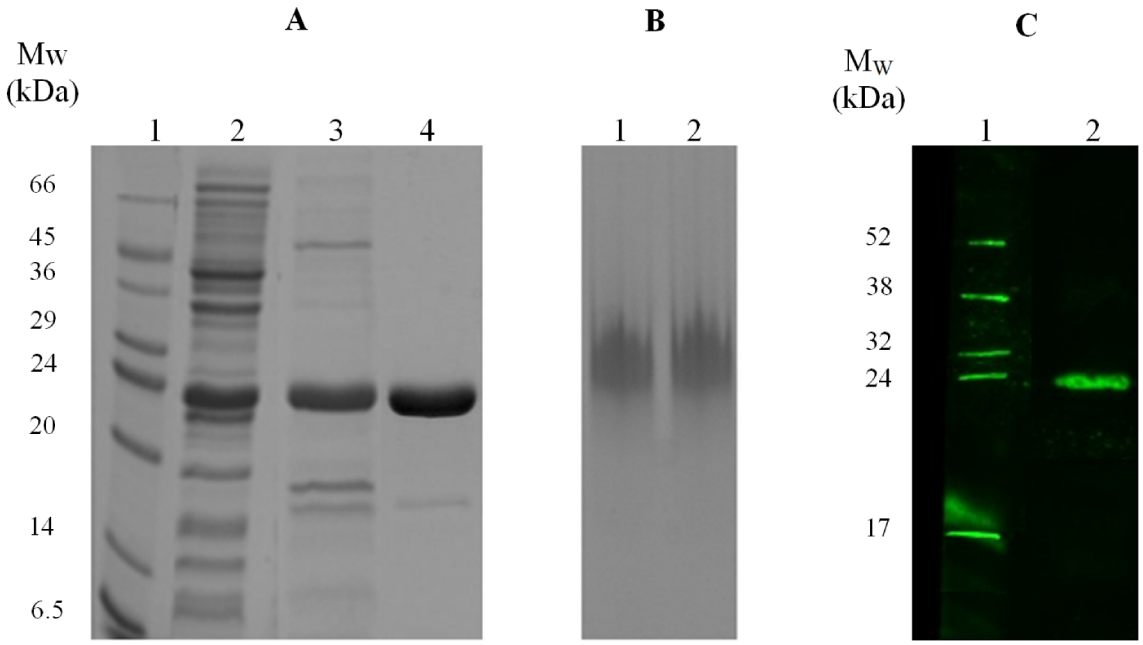



**Figure 3.21: Purification of Hsp90β $^{(513-724)}$. A: SDS-PAGE (15%) analysis:** Lane 1: Low range molecular weight standards (5 µL) (Sigma-Aldrich); Lane 2: Total protein lysate (20 µg); Lanes 3-4: Hsp90β$^{(513-724)}$ eluted with 25 mM imidazole after the first and second round IMAC respectively (15 µg). **B: Native-PAGE (8%) analysis:** Lanes 1-2: Pure Hsp90β$^{(513-724)}$ (15 µg). **C: Immunoblotting using anti-his tag as 1° antibody.** Lane 1: Rainbow molecular weight markers (5 µL) (GE Healthcare), Lane 2: Pure Hsp90β$^{(513-724)}$ (5 µg).

**Table 3.8: Purification Table of Hsp90β $^{(513-724)}$**

| **Purification step (from 1 L culture)** | **Volume (mL)** | **Concentration (mg.mL$^{-1}$)** | **Total protein (mg)** | **Purity (%)** |
|---|---|---|---|---|
| Total cell lysate | 30 | 91.2 ± 1.3 | 2.7 x10$^3$ | 10 |
| Clarified lysate | 27 | 86.1 ± 0.9 | 2.3 x10$^3$ | 13 |
| IMAC Round 1 | 20 | 1.92 ± 0.1 | 38.4 | 88 |
| IMAC Round 2 | 20 | 1.45 ± 0.2 | 29.0 | 95 |

## 3.11. Phosphodiesterase 4A5

### 3.11.1. Generation of pETite H6-SUMO PDE by homologous recombination

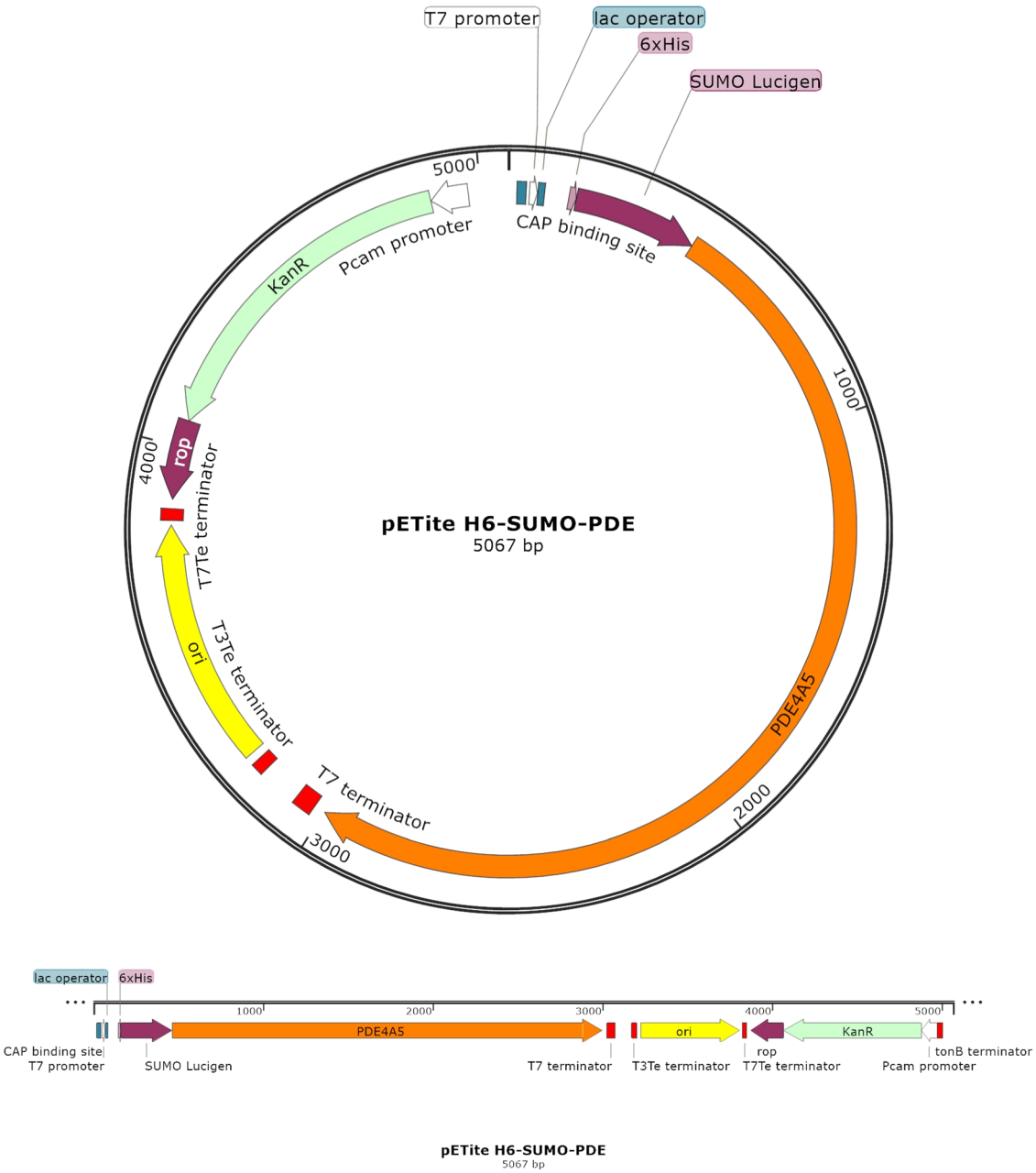



**Figure 3.22: pETite-H6-SUMO-PDE recombinant clone.** *In-silico* sub-cloning of the rat cDNA PDE4A5 sequence within pETite H6-SUMO vector generated using SnapGene® Viewer.

The *PDE4A5* cDNA was amplified using the primers designated as pETite-PDE-F and pETite-PDE-R as the forward and reverse primers respectively (Table 2.8), using the PCR cycling parameters;

Denaturation step: 94°C for 2 min;

25 cycles: 94°C for 15 sec, 60°C for 15 sec, 72°C for 2.5 min;

Extension step: 72°C for 4 min.

A broad band of approximately 2.5 kb in length was obtained, in agreement with the size of *PDE4A5* cDNA (2,535 bp) as predicted from *in-silico* cloning performed in SnapGene® Viewer. The PCR product was gel purified prior to the homologous recombination in HI-Control 10G cells. Twelve of the transformed colonies were randomly picked for mini-prep analyses of DNA. Of these, nine where shown to have recombinant DNA. Five of these were grown in 2TY media supplemented with kanamycin (50 $\mu g.mL^{-1}$) for plasmid extraction. DNA (100 $ng.\mu L^{-1}$) in 10 µL final volume was sent for Sanger DNA sequencing (Bioneer), which confirmed the success of cloning by homologous recombination.

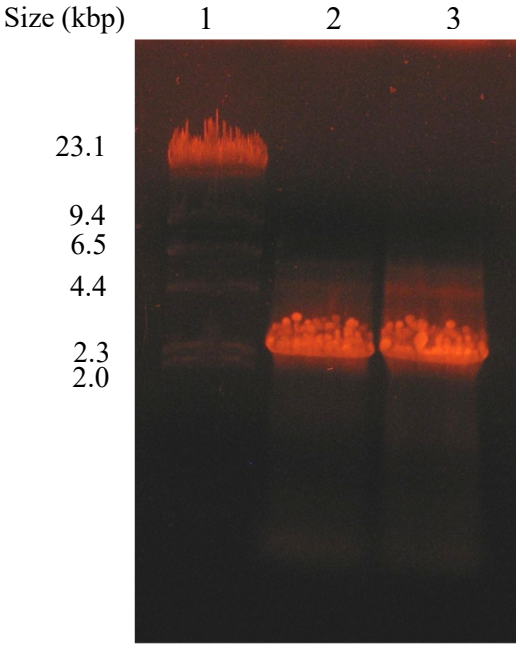



**Figure 3.23: Amplification of *PDE4A5* cDNA by PCR.** Lane 1: Lambda-phage DNA HindIII digested (1 µg); Lanes 2-3: PDE4A5 PCR product (each well contains 25 µL from 50 µL PCR mix, approximately 2 µg per lane).

### 3.11.2. Optimisation of protein expression and purification of $H_6$-SUMO-PDE4A5

The original PDE-construct provided had an MBP-tag upstream of the PDE coding sequencing. Various experiments were initially carried out to purify PDE4A5 as an MBP-tagged protein, employing the use of MBP-Trap affinity columns filled with an amylose resin to which MPB-tagged proteins display affinity. However, although the MBP-tag did enhance protein solubility, the protein was not binding selectively to the resin. This made the purification laborious and the purity of the protein obtained never exceeded 70% (Vella, 2015). The same issue has been reported elsewhere stating that MBP-tagged proteins do not bind effectively to the amylose resin and that amylose affinity chromatography does not yield a large quantity of pure recombinant protein (Pryor and Leiting, 1997; Routzahn and Waugh, 2002). Consequently the *PDE4A5* cDNA was sub-cloned into the pETite vector by homologous recombination (Section 2.2.10), with the scope of improving the purification process and enhancing the final level of purity attained.

Analytical small-scale cultures (10 mL) were set up to test and optimise protein expression. Cultures were grown until an $OD_{600}$ of 0.4-0.6 and induced with IPTG at a final concentration of 0.1 mM. Sorbitol (400 mM) was included into the growth media of one of the cultures under test. The addition of sorbitol has been reported to improve the solubility of recombinant proteins produced in *E.coli* (Prasad, Khadatare and Roy, 2011). The pH of the media containing sorbitol was maintained constant by the addition of KP buffer pH 7.8, at a final concentration of 0.1 M. Induced cultures were harvested after 3 h of growth. The re-suspended pellets were standardised with respect to cell growth, by dilution with 2TY media to the same $OD_{600}$ value.

The result obtained indicates that sorbitol, in this case, had no positive effect. The culture grown in the absence of sorbitol at 30°C had a higher yield of protein than the culture grown with sorbitol at the same temperature. The best protein yield was obtained at 37°C (Figure 3.24).

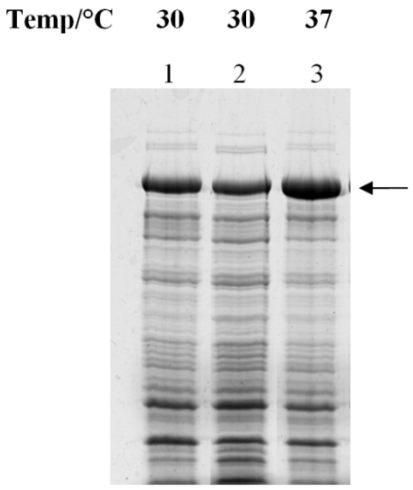



**Figure 3.24: Optimisation of $H_6$-SUMO-PDE4A5 protein expression (A): SDS-PAGE (8%) analysis:** Lane 1: Lysate from cells grown at 30 °C (15 µg); Lane 2: Lysate from cells grown at 30°C in the presence of 400 mM sorbitol (15 µg); Lane 3: Lysate from cells grown at 37°C (15 µg). Over-expressed $H_6$-SUMO-PDE4A5 is indicated by an arrow.

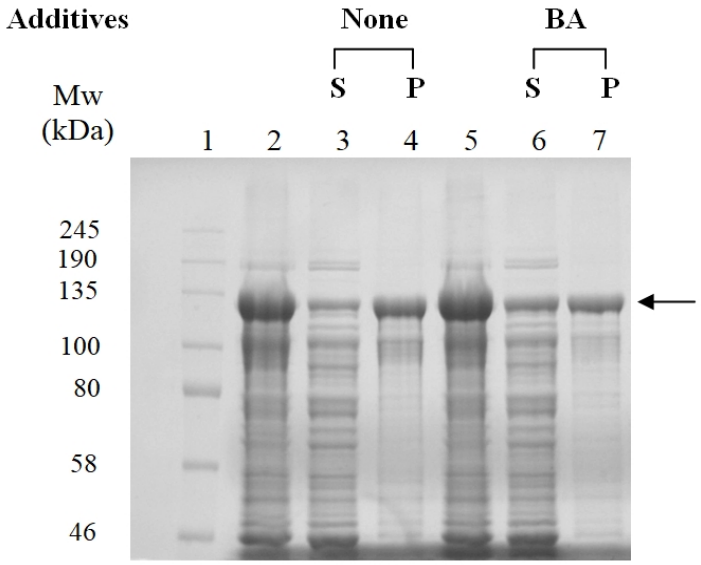



**Figure 3.25: Optimisation of $H_6$-SUMO-PDE4A5 protein expression (B): SDS-PAGE (8%) analysis:** Lane 1: Colour prestained protein standards, broad range (3 µl) (NEB); Lane 2: Lysate from cells induced with 0.1 mM IPTG (20 µg); Lane 3: Soluble fraction (S) of lysate in lane 1 (15 µg); Lane 4: Re-solubilised pellet (P) obtained after the centrifugation of lysate in lane 1 (15 µg); Lane 5: Lysate from cells grown in 10 mM benzyl alcohol (BA) and induced with 0.1 mM IPTG (20 µg); Lane 6: Soluble fraction of lysate in lane 5 (15 µg); Lane 7: Re-solubilised pellet (obtained after the centrifugation of lysate in lane 5 (15 µg). Both cultures were grown at 37°C. The arrow indicates a protein band migrating with the approximate size of $H_6$-SUMO-PDE4A5.

Benzyl alcohol (BA), when present in low doses, has been shown to cause transient fluidisation of the *E.coli* cell membrane and up-regulates the transcription of target heat shock genes (Shigapova *et al*., 2005). BA treatment thus induces endogenous chaperone expression that may improve the yield and solubility of proteins expressed in an *E.coli* system (Shigapova *et al*., 2005; Marco *et al*., 2005). This is particularly useful when expressing relatively large proteins with a higher level of complexity. The results obtained indicate, that the expression of $H_6$-SUMO-PDE4A5 was highly efficient even in the absence of BA. In fact, the addition of BA did not seem to affect and/or improve the overall yield of the recombinant protein. However, it did enhance the proportion of soluble PDE4A5, as a higher amount is observed in Lane 6, when compared to Lane 3 in Figure 3.25.

Large-scale preparative protein expression was carried out in 2TY 2 L culture flasks supplemented with kanamycin (50 $\mu g.mL^{-1}$). BA (10 mM) was added 20 min prior to induction with 0.1 mM IPTG. $ZnSO_4$ (40 μM) was added at the point of induction, as it is required for the folding of active PDE (Pryor and Leiting, 1997; Routzahn and Waugh, 2002). Protein expression was allowed to continue at 37°C for 3 h.

$H_6$-SUMO-PDE4A5 was purified by IMAC (Section 2.3.3), with the exception of using 1 mM Tris(2-carboxyethyl)phosphine (TCEP) instead of DTT. The clarified cell lysate was loaded onto a 5 mL nickel pre-charged HiTrap Chelating Column HP (GE Healthcare), the column washed with equilibration buffer (20 mM Tris-Cl pH 8.0, 0.5 M NaCl, 1 mM TCEP) containing 50 mM imidazole. Elution was achieved with 200 mM imidazole. Pure dialysed PDE4A5 was stored at -80°C in 20 mM Tris-Cl pH 8.0, 100 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, at a concentration of 0.5 $mg.mL^{-1}$ for future experimentation.

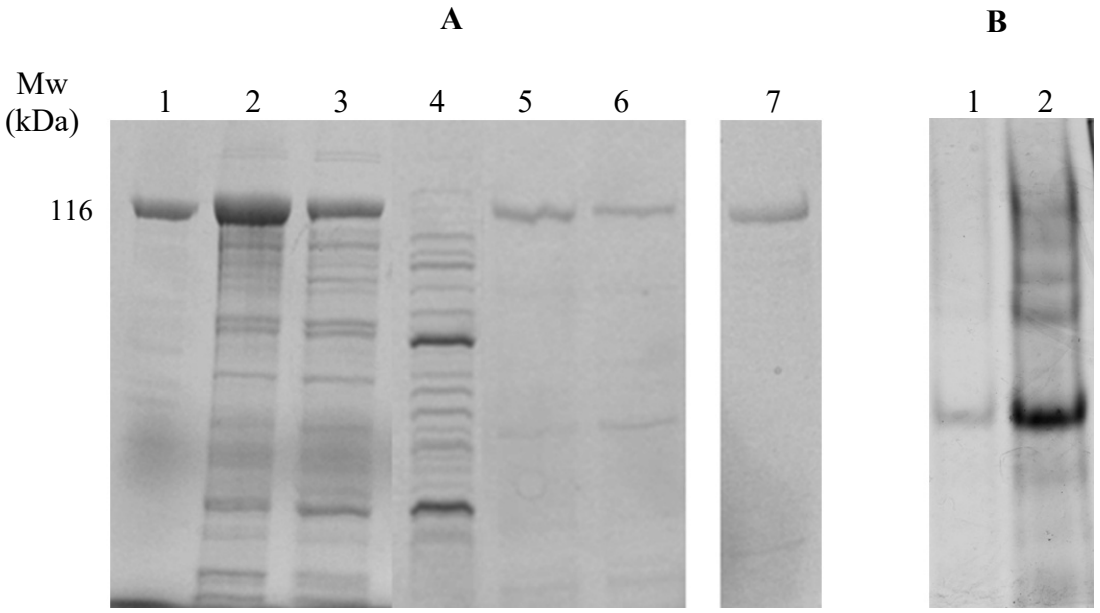



**Figure 3.26: Purification of $H_6$-SUMO-PDE4A5.A: SDS-PAGE (8%) analysis:** Lane 1: β-galactosidase standard (10 µg) (Sigma-Aldrich); Lane 2: Total protein lysate 1 (20 µg); Soluble fraction from lysate loaded on column (15 µg); Lane 4: Column through-wash with 50 mM imidazole (10 µg); Lanes 5-6; $H_6$-SUMO-PDE4A5 eluted with 200mM imidazole (10 µg); Lane 7: $H_6$-SUMO-PDE4A5 eluted 200mM imidazole after the second round of purification by IMAC (10 µg). **B: Native-PAGE (8%) analysis:** Lane 1: Purified $H_6$-SUMO-PDE4A5 at 1 $mg.mL^{-1}$ (10 µL); Lane 2: Purified $H_6$-SUMO-PDE4A5 at 3 $mg.mL^{-1}$ (10 µL).

Concentration of pure $H_6$-SUMO-PDE4A5 was carried out using an Amicon Ultra centrifugal filter unit, with a molecular weight cut off of 30 kDa. Despite the presence of glycerol and TCEP in the buffer, the protein started to form multimeric states even at a concentration of 3 $mg.mL^{-1}$. From a 1 L culture, a total of 5.6 mg of protein, with a purity of 87% was attained.

The precipitation of PDEs as inclusion bodies when grown in *E.coli* is not uncommon (Lugnier, 2005) and the proportion of PDE protein observed in the insoluble fraction (pellet) was higher than the amount present in the soluble portion (Figure 3.25). In an attempt to obtain a higher yield of pure protein, an experiment was designed in an attempt to purify the SUMO-tagged protein from the pellet. The methodology employed is a modification of the technique reported by Richter, Hermsdorf and Dettmer (2002) (Section 2.3.5).

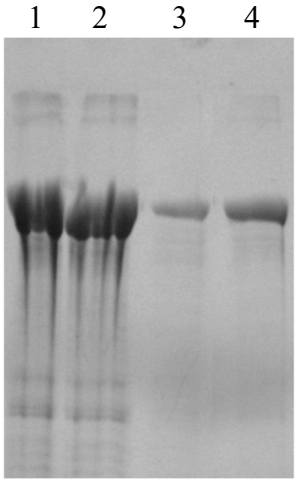



**Figure 3.27: Purification stages of $H_6$-SUMO-PDE4A5 from inclusion bodies: SDS-PAGE (8%) analysis:** Lane 1: $H_6$-SUMO-PDE4A5 in solubilisation buffer (30µg); Lane 2: Soluble unfolded $H_6$-SUMO-PDE4A5 after centrifugation (30 µg); Lanes 3-4: Soluble $H_6$-SUMO-PDE4A5 after refolding (10-15 µg).

The purification of $H_6$-SUMO-PDE4A5 from inclusion bodies produced a slightly higher protein yield of superior purity. However owing to the denaturing nature of this technique and the difficulties associated with the re-folding of proteins, the purified $H_6$-SUMO-PDE4A5 was assessed for secondary structure and protein stability by CD spectroscopy (Section 4.2.1).

**Table 3.9: Purification Tables of $H_6$-SUMO-PDE4A5**

| **Purification step (from 1 L culture)** | **Volume (mL)** | **Concentration ($mg.mL^{-1}$)** | **Total protein (mg)** | **% Purity** |
|---|---|---|---|---|
| Total Cell lysate | 30 | 58.6 ± 4.5 | $1.8 \times 10^3$ | 8 |
| Clarified lysate | 27 | 26.3 ± 2.1 | $7.1 \times 10^2$ | 10 |
| IMAC Round 2 | 15 | 0.9 ± 0.1 | 10.5 | 80 |
| IMAC Round 2 | 8 | 0.7 ± 0.1 | 5.6 | 87 |
| | | | | |
| Total cell lysate from inclusion bodies | 30 | 30.6 ± 4.51 | $9.2 \times 10^2$ | 15 |
| Solubilised protein | 6 | 5.5 ± 1.51 | 33.4 | 78 |
| Protein in re-folding buffer | 50 | 0.2 ± 0.06 | 10.0 | 85 |
| IMAC | 15 | 0.3 ± 0.09 | 8.4 | 93 |

## 3.12. Aryl hydrocarbon receptor (AhR)

### 3.12.1. Sub-cloning of AhR cDNA into pET28a-H$_6$-SUMO by homologous recombination

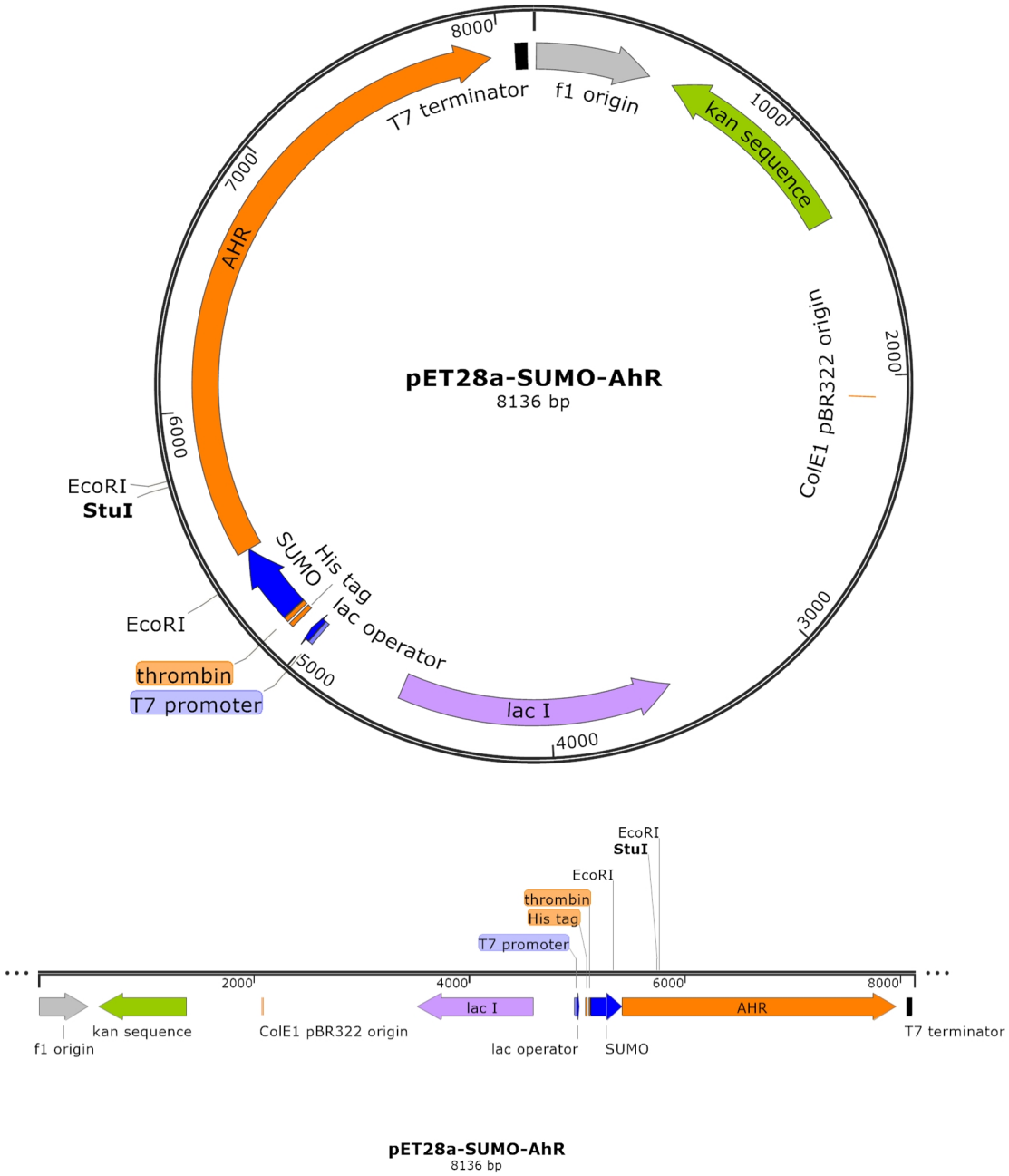



**Figure 3.28: pET28a-SUMO-AhR recombinant clone.** *In-silico* cloning of the human AhR sequence within pET28a-H$_6$-SUMO vector generated using SnapGene® Viewer.

StuI is only present within the *AhR* cDNA and recognises a single site at position 5,746. This results in the linearization of the vector. EcoRI has two restriction cuts sites, at positions 5,766 and 5,341 resulting in two DNA fragments of sizes 425 bp and 7,711 bp.

*AhR* cDNA (2,544 bp) was successfully amplified by PCR using AhR-F and AhR-R as the PCR primers (Table 2.8). Homologous recombination was 71 % efficient, with 20 out of 28 colonies analysed showing a higher molecular weight plasmid, when compared to the controls (Lanes 1 and 30, Figure 3.29). Recombinant positive clones were grown for plasmid extraction, analysed for the presence of the insert by restriction digest (Figure 3.30) and sequenced for confirmatory purposes.

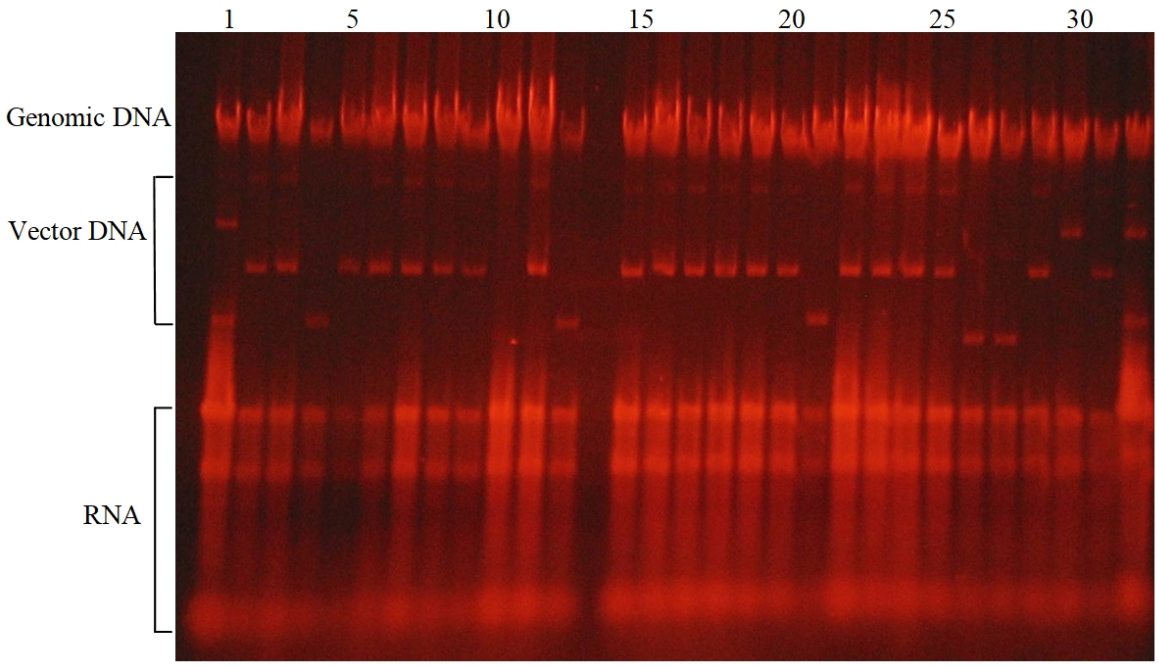



**Figure 3.29: Mini-prep analysis of the DNA of transformed bacterial colonies.** Lane 1 and 30 contain DNA from cells transformed with empty pET28a vector as control. Lanes 2-29 contain nucleic acid from cells that were transformed after homologous recombination.

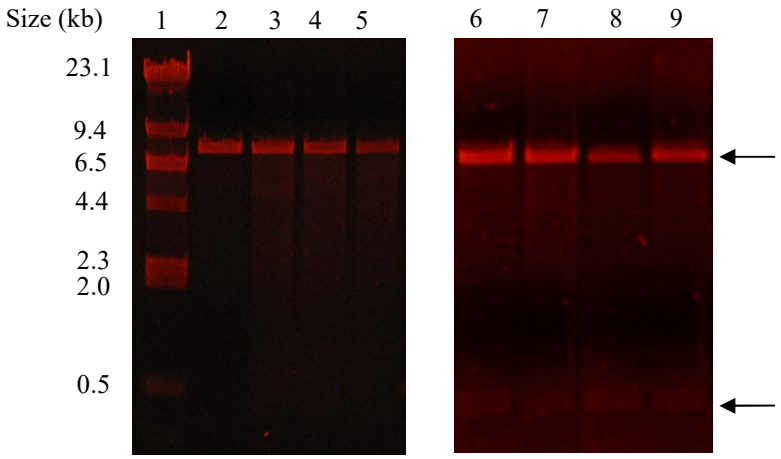



**Figure 3.30: Restriction digest of pET28a-SUMO-AhR to confirm insert.** Lane 1: Lambda-phage DNA HindIII digested (1 µg); Lanes 2-5: DNA (0.3 µg) after StuI restriction digest. Lanes 6-9: DNA (0.3 µg) after EcoRI restriction digest.

All DNA samples in lanes 2-5 produced a fragment corresponding to the expected size of the linearised vector, 8,136 bp. Similarly, DNA in lanes 6-9 resulted in fragments of the expected size for a positive recombinant clone, equivalent to 7,711 bp and 425 bp as predicted through *in-silico* cloning. The location of each fragment is indicated by arrows (Figure 3.30).

### 3.13. Optimisation of protein expression of $H_6$-SUMO-AhR

BL21 (DE3) cells transformed with pET28a-SUMO-AhR were grown in 10 mL 2TY cultures and IPTG induced with some modifications to the method described in section 3.2. Cells were harvested by centrifugation at 3,000 *g* for 10 min and lysed by lysozyme (1 mg $mL^{-1}$). PMSF (1 µL from 0.17µg $mL^{-1}$ stock) was added to inhibit the proteolytic effect of any serine proteases that may be released after cell lysis. DNA was degraded by sonication on ice. Soluble protein was separated from cellular debris and inclusion bodies by centrifugation at 10, 000 *g* and analysed by SDS-PAGE electrophoresis. A monoclonal anti-his antibody (Aviva Systems Biology) was used for detection by immunoblotting (Section 2.1.9).

Cells were grown at 37°C and protein expression was monitored over time. No apparent AhR band (of approximate 110 kDa in size) could be observed on SDS-PAGE and thus immunoblotting was performed. This showed that whilst AhR was being produced in relatively small quantities, it was completely degraded after 4 h of growth (Figure 3.31, lane 4). The absence of Ab-positive bands in the uninduced culture (Figure 3.31 lane 2) clearly suggests that the lower molecular weight bands observed under other conditions (Figure 3.31, lanes 3-5) are not due to non-specific binding of the antibody, but are most likely AhR degradation fragments or prematurely stopped translation products containing the $H_6$-segment.

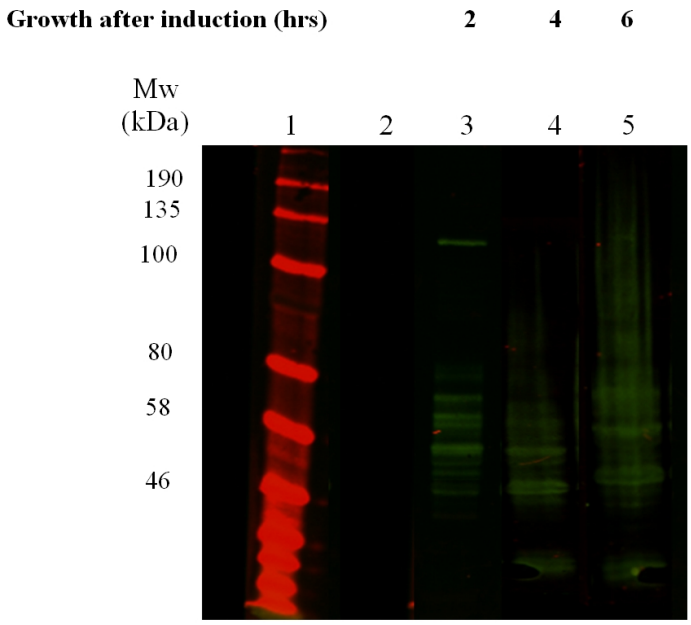



**Figure 3.31: Immunoblotting: $H_6$-SUMO-AhR protein expression.** Lane 1: Colour prestained protein standards, broad range (3 µl) (NEB); Lane 2: Lysate from uninduced culture; Lanes 3-5 represent cells grown for 2, 4 and 16 hours respectively. All cells were transformed with the recombinant vector, pET28a-SUMO-AhR and were grown at 37°C. Each well contains 10 µg of protein. Proteins were separated on 8% SDS-PAGE, transferred to a nitrocellulose membrane and the blot probed with anti-his tag as 1° antibody.

Considering the extent of protein degradation and/or fragmentation a number of experimental variables were tested to improve protein quality. The effects of benzyl alcohol, PMSF, temperature, time course of protein expression and inducer concentration were investigated. The use of PMSF during protein expression was

employed following the findings of Prouty and Goldberg, (1972) that showed that PMSF added during cell growth, can inhibit the rate of protein breakdown intracellularly in *E.coli* cells. The use of BA was once again investigated due to its reported ability to induce the expression of endogenous chaperones in *E.coli* (Shigapova *et al*., 2005; Marco *et al*., 2005).

BL21 (DE3) Codon Plus cell harbouring pET28a-SUMO-AhR vector, were grown in 2TY media at 37°C, until an $OD_{600}$ of 0.5. BA was added to a final concentration of 10 mM and the culture grown for an additional 20 min before inducing with IPTG, 0.5 mM final concentration. The $OD_{600}$ at the point of induction was ~0.8-0.9. Protein expression was then carried out at 18°C. Similarly, the experiments were repeated in the presence of 1 mM PMSF. After addition of PMSF or BA, cell growth was monitored by taking aliquots at regular time intervals and reading the $OD_{600}$. Neither BA nor PMSF affected bacterial cell growth.

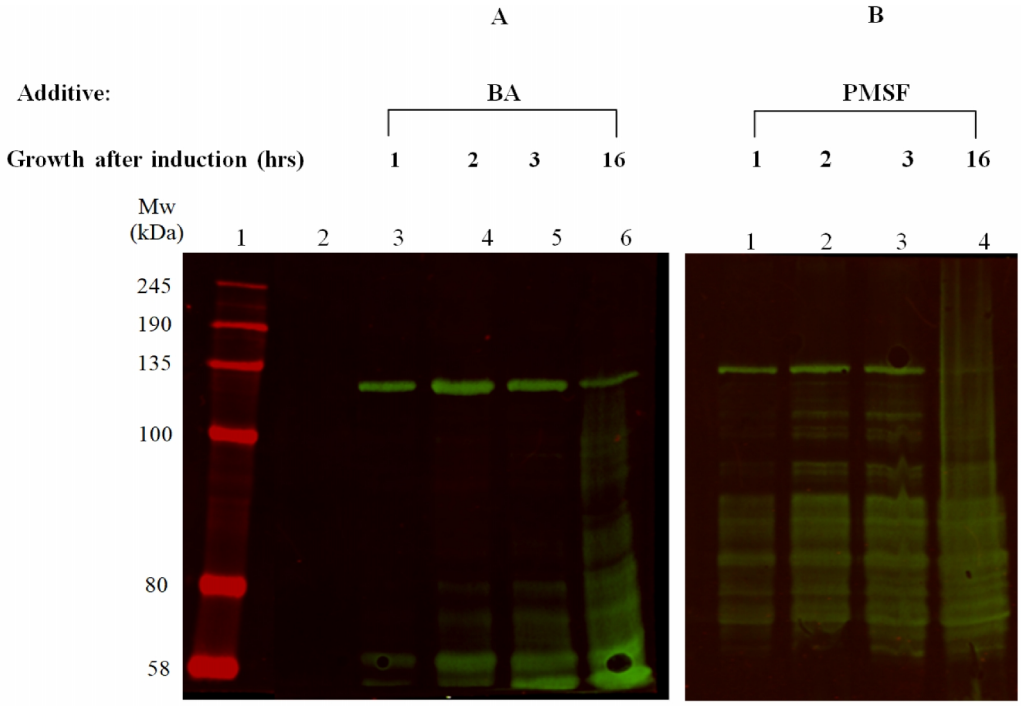



**Figure 3.32: Immunoblotting: Optimisation of $H_6$-SUMO-AhR protein expression. A: Effect of benzyl alcohol**. Lane 1: Colour prestained protein standards (3 µl) (NEB ); Lane 2: Un-induced culture; Lanes 3-6: Cultures grown in 10 mM BA, induced with 0.5 mM IPTG, grown for 1 hr, 2 hrs, 3 hrs and 16 hrs respectively. **B: Effect of PMSF:** Lanes 1-4: Cultures grown in 1 mM PMSF, induced with 0.5 mM IPTG, grown for 1 hr, 2 hrs, 3 hrs and 16 hrs respectively. In each case, 10 µL of lysate was loaded for testing (~ 20 µg of protein). Proteins were separated on 8% SDS-PAGE and the blot probed with anti-his tag 1° antibody.

BA significantly enhanced protein expression and reduced the extent of protein degradation with the best expression observed after 2 h of growth after induction. On the contrary, PMSF had no apparent effect on the protein degradation pattern. The presence of an AhR band in the both BA and PMSF-treated cells, as opposed to the untreated cells (Figure 3.31), shows that protein expression is better carried out at 18°C than at 37°C as was initially carried out. To confirm this observation, small scale cultures (10 mL) were grown in the presence of BA (10 mM). Each culture was induced with 0.5 mM IPTG (final concentration) at an $OD_{600}$ of 0.8 and incubated at 18°C, 30°C and 37°C. Aliquots were taken after 2 and 4 h and analysed by SDS-PAGE electrophoresis (Figure 3.33). Similarly, the effect of IPTG concentration was investigated (Figure 3.34).

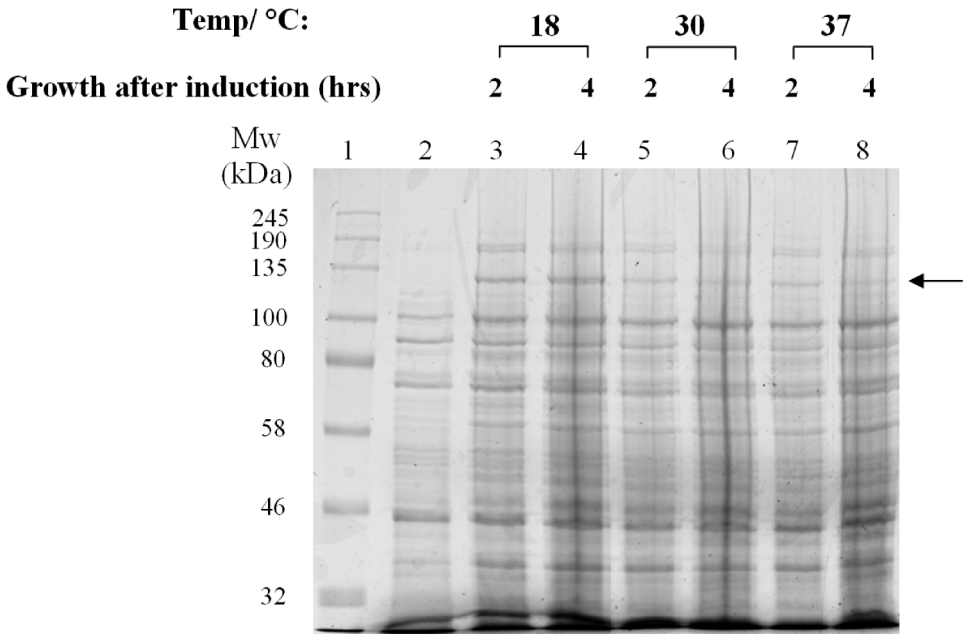



**Figure 3.33: Effect of growth temperature on $H_6$-SUMO-AhR production in the presence of benzyl alcohol**: **SDS-PAGE (8%) analysis**: Lane 1: Colour prestained protein standards, broad range (3 µl) (NEB); Lane 2: Lysate from an uninduced culture; Lane 3-8: Induction for 2 hrs and 4 hrs at 18°C, 30°C and 37°C respectively. Each of lysate contained around 10~15 µg of protein per well. For reliable comparison, samples were standardised with respect to cell growth by dilution to the same $OD_{600}$ value. The arrow indicates a protein band migrating with the approximate molecular weight of $H_6$-SUMO-AhR.

The result obtained further suggests that expression of $H_6$-SUMO-AhR is best carried out at low temperatures. Higher temperatures of 30°C and 37°C resulted in less protein production possibly due to substantial protein degradation (Figure 3.33).

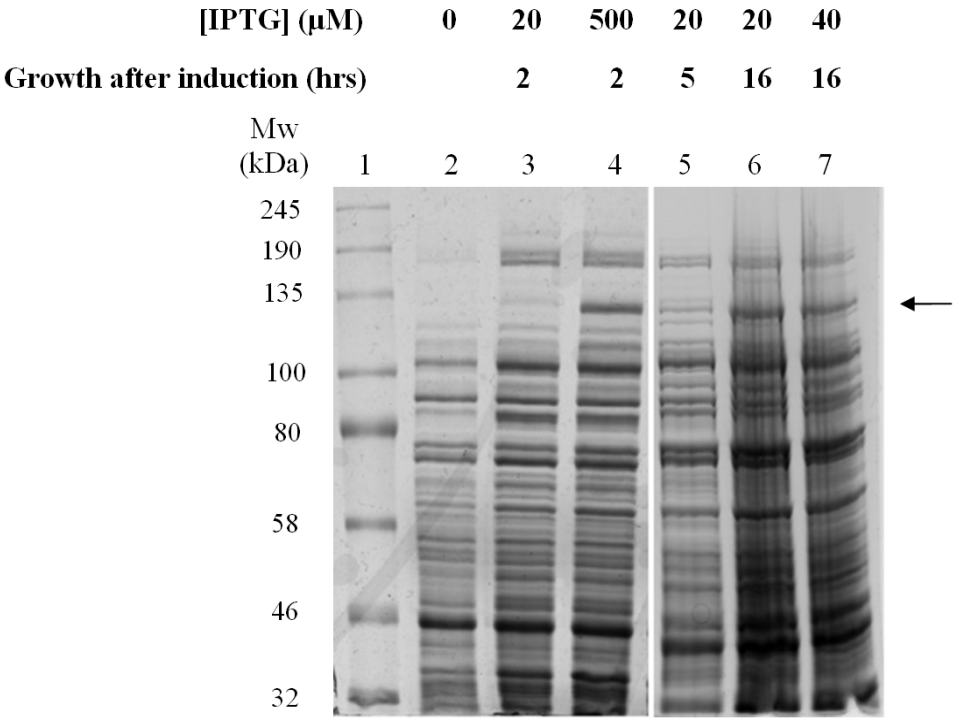



**Figure 3.34: Effect of IPTG concentration on $H_6$-SUMO-AhR production: SDS-PAGE (8%) analysis:** Lane 1: Colour prestained protein standards (3 µl) (NEB); Lane 2: Lysate from an un-induced culture; Lane 3: Induced with 20 µM IPTG for 2 hrs; Lane 4: Induced with 500 µM IPTG for 2 hrs; Lane 5: Induced with 20 µM for 5 hrs; Lane 6: Induced with 20 µM for 16 hrs; Lane 7: Induced with 40 µM for 16 hrs. All cells were transformed with the recombinant vector, pET28a-SUMO-AhR and were grown at 18°C. Each well contains around 15~20 µg of cell lysate. The arrow indicates the band corresponding to $H_6$-SUMO-AhR.

The results indicate that when using low inducer concentration, there is minimal protein production even after 5 h from induction time. However, a discernible band was observed following overnight incubation (Figure 3.34, Lanes 6-7). In the light of these and previous results, all subsequent cultures were grown in the presence of BA (10 mM) at 37°C until the point of induction (at $OD_{600}$ of 0.8), induced with IPTG at a final concentration of 0.5 mM and then transferred to 18°C, for 2 h, for protein expression.

After having established the conditions that enhance protein expression, the effect of chemical additives was also investigated. Following BA addition, cultures were supplemented with chemical additives 5 min prior to IPTG induction. The additives tested were; ethanol (3% v/v); sorbitol (400 mM) and arginine (50, 100 mM). Ethanol was included as one of the additives, after the findings of Chhetri *et al*., (2015) who have proven that ethanol (at a final concentration of 3% v/v) can significantly enhance the production of recombinant proteins in *E.coli.* The inclusion of sorbitol and arginine in the growth medium is a modification of the conditions reported by Prasad, Khadantare and Roy (2011) and Leibly *et al*., (2012) in their attempts to improve the solubility of recombinant proteins in *E.coli*. Stock solutions of arginine and sorbitol were both prepared in 50 mM KP buffer pH 7.8, to prevent changes in the pH of the 2TY media. Drastic changes in pH hinder bacterial growth and encourage protein precipitation and the formation of inclusion bodies, thereby negating the scope of the experiment. The final pH of the medium after 2 hours of induction was 8.0 ± 0.2. Each pellet was re-suspended in 1 mL of Bacterial Protein Extraction Reagent (B-PER), briefly sonicated on ice and centrifuged. Soluble and pelleted fractions were analysed by SDS-PAGE.

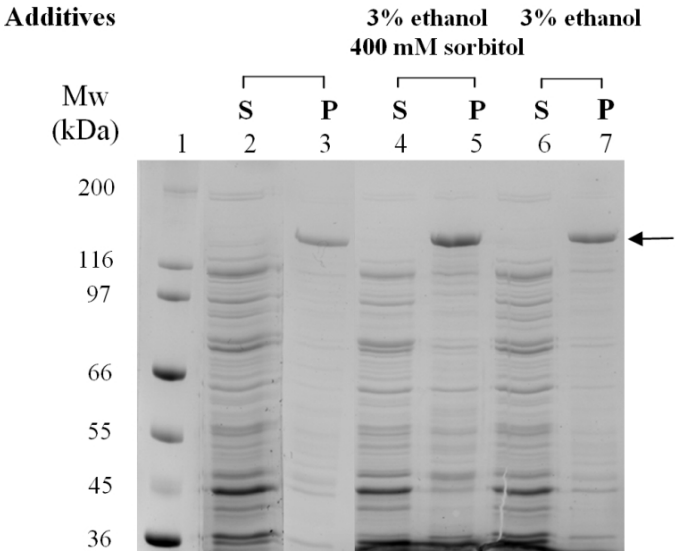



**Figure 3.35: Effect of additives (A): SDS-PAGE (8%) analysis:** Lane 1: Wide range molecular weight markers (3 µL) (Sigma-Aldrich); Lanes 2-3: No additives; Lanes 4-5: Induced cells in the presence of 3% (v/v) ethanol and 400 mM sorbitol; Lanes 6-7: Induced cells in the presence of 3% (v/v) ethanol. In each case; S denotes the soluble fraction; whilst P refers to the re-solubilised pellet. Equal volumes of each fraction were loaded in each well (10 µL). The arrow indicates the band corresponding to $H_6$-SUMO-AhR.

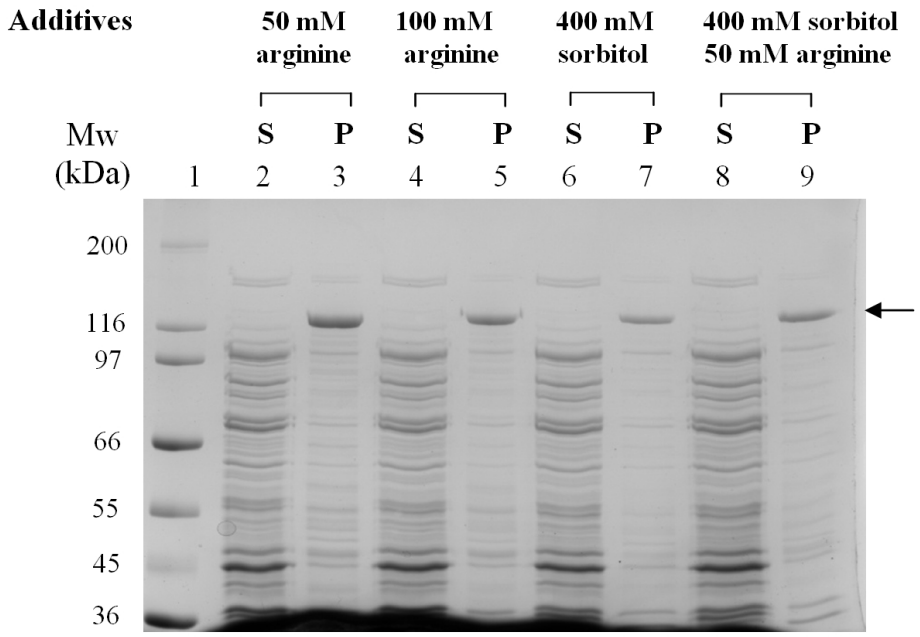



**Figure 3.36: Effect of additives (B): SDS-PAGE (8%) analysis:** Lane 1: Wide range molecular weight markers (3 µL) (Sigma-Aldrich); Lanes 2-3: Induced cells in the presence of 50 mM arginine; Lanes 4-5: Induced cells in the presence of 100 mM arginine; Lanes 6-7: Induced cells in the presence of 400 mM sorbitol; Lanes 8-9: Induced cells in the presence 400 mM sorbitol and 50 mM arginine. In each case; S denotes the soluble supernatant fraction; whilst P refers to the re-solubilised pellet. Equal volumes of each fraction were loaded in each well (10 µL). The arrow indicates the band corresponding to $H_6$-SUMO-AhR.

The highest AhR expression was observed in the presence of 3% ethanol and 400 mM sorbitol (Figure 3.35, Lane 5) and in the presence of 50 mM arginine (Lane 3, Figure 3.36). Ethanol is an amphipathic molecule that can influence membrane fluidity and composition (Ingram and Buttke, 1984), with possible enhancement of DNA synthesis that is believed to be responsible for the higher expression of inducible proteins (Chhetri *et al*., 2015). Sorbitol and arginine are both known for their role as a protein stabiliser and aggregation suppressor respectively, the latter exhibiting its effect by interacting with surface residues of the protein (Liu *et al*., 2007; Prasad *et al*., 2011). However, their combined addition was not effective (Figure 3.36, Lane 9). Similarly, protein expression in the presence of either ethanol (Figure 3.35, Lane 7) or sorbitol only (Figure 3.36, Lane 7) was comparable to that observed in the untreated cells. On the other hand, when both chemicals were in the growing media, the expression showed a considerable increase (Figure 3.35, Lane 5), suggesting a combinational effect.

While the use of certain additives did enhance protein expression, under all condition investigated, AhR precipitated in the pellet with minimal amount, if any, remaining in the soluble fraction. In order to increase the amount of soluble AhR, the next experimental variable to be investigated was the composition of the lysis buffer.

The lysis buffer is known to significantly affect the solubility of a protein. A combination of additives and buffers (EMBL, Heidelberg, 2019) were thus employed to investigate the effect of lysis buffer on AhR solubility. Individual pellets from 10 mL IPTG-induced cultures supplemented with 50 mM arginine, were re-suspended in 1 mL of each solubilisation buffer (Table 3.10). The cell suspension was incubated at room temperature for 10 min to enhance lysozyme activity. Following sonication, the suspension was centrifuged for 5 min at 10,000 *g*. For fair comparison, each pellet was re-solubilised in an equal volume of the original solubilisation buffer and loaded on SDS-PAGE for analyses.

**Table 3.10: Composition of lysis buffers.**

| Lysis buffer | Components |
|---|---|
| **A** | 50 mM Tris-Cl pH 7.5, 50 mM NaCl, 5 mM EDTA, 0.5 M urea, 1 mg $mL^{-1}$ lysozyme |
| **B** | 20 mM Tris-Cl pH 7.5, 50 mM NaCl, 0.2% v/v Nonidet-P40, 1 mg $mL^{-1}$ lysozyme |
| **C** | 50 mM Tris-Cl pH 8.5, 50 mM NaCl, 5 mM EDTA, 1 mg $mL^{-1}$ lysozyme |
| **D** | 50 mM Tris-Cl pH 7.5, 50 mM NaCl, 5 mM EDTA, 1 M urea, 1 mg $mL^{-1}$ lysozyme |
| **E** | 20 mM Tris-Cl pH 7.5, 50 mM NaCl, 0.2% v/v Triton X100, 1 mg $mL^{-1}$ lysozyme |
| **F** | 20 mM Tris-Cl pH 7.5, 50 mM NaCl, 0.2% v/v Tween-20, 1 mg $mL^{-1}$ lysozyme |
| **G** | 50 mM Tris-Cl pH 7.5, 50 mM NaCl, 10% v/v glycerol, 1 mg $mL^{-1}$ lysozyme |
| **H** | 50 mM Tris-Cl pH 8.5, 50 mM NaCl, 5 mM EDTA, 1 M urea, 1 mg $mL^{-1}$ lysozyme |

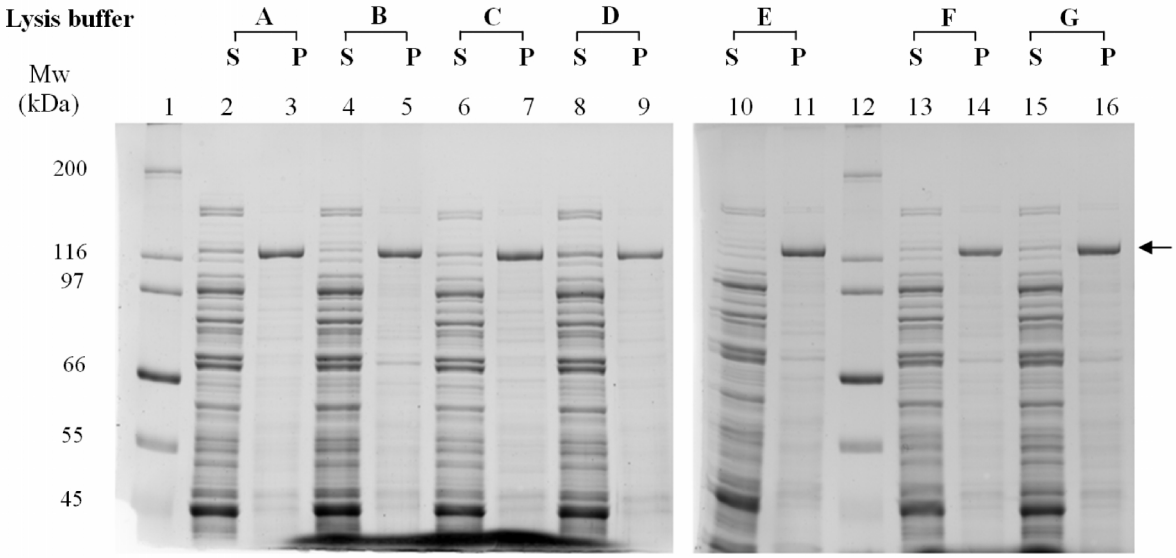



**Figure 3.37: Effect of solubilisation buffer: SDS-PAGE (8%) analysis:** Lanes 1 and 12: Wide range molecular weight markers (3 µL) (Sigma-Aldrich). All the other lanes contain equal volumes (10 µL) of supernatant and solubilised pellet were loaded alongside each other, for each of the 7 solubilisation buffers A-G. S denotes the soluble supernatant fraction; whilst P refers to the re-solubilised pellet**.** The arrow indicates the band corresponding to $H_6$-SUMO-AhR.

In all cases, the majority of $H_6$-SUMO-AhR precipitated out of solution and was seen in the pellet, as an almost pure band. However, as opposed to the previous result (Figures 3.35 and 3.36), a fraction of $H_6$-SUMO-AhR was visualised in the supernatant of Lanes 2, 6 and 8 and was thus soluble. This implies that the presence of urea (found in lysis buffers A and D) and alkaline pH (lysis buffer C), encourages the solubility of the AhR. The efficiency of lysis was confirmed by the fact that all other bacterial proteins were in the soluble fraction, which would not be the case if lysis was incomplete. Another buffer (H) was therefore prepared (Table 3.10). This new buffer contained urea and an alkaline pH so combining the additives of lysis buffer A and D.

Moreover, after observing that a substantial amount of protein was precipitating in the pellet as inclusion bodies, the use of low concentration of IPTG was once again explored. In some cases, low inducer concentration over a prolonged period of time has been reported to promote protein folding, particularly when the expression is carried out

under low temperature (Rosano and Ceccarelli, 2014). Buffer H was used on freshly grown cultures (Figure 3.38).

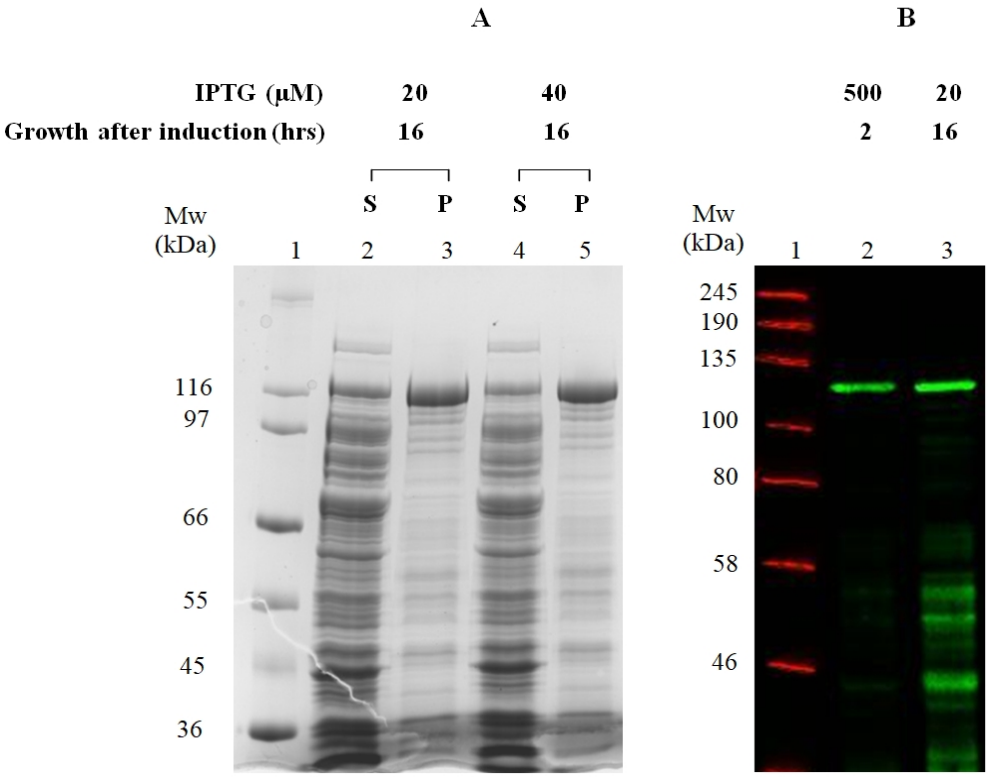



**Figure 3.38: Optimisation $H_6$-SUMO-AhR solubility**. **A: 8% SDS-PAGE analysis**: Lane 1: Wide range molecular weight markers (3 µL) (Sigma-Aldrich); Lanes 2-3; supernatant (S) and pellet (P) from an overnight culture grown in the presence of arginine (50 mM), BA (10 mM) and induced with 20 µM IPTG at 18°C (15 µg). Lanes 4-5: Same culture conditions but induced with 40 µM IPTG at 18°C (15 µg). Harvested cells were both re-suspended in lysis buffer H (Table 3.10).

**B: Immunoblotting:** Lane 1: Colour prestained protein standards (3 µl) (NEB); Lane 2: Soluble fraction containing AhR from a culture grown in arginine (50 mM), BA (10 mM), induced with 500 mM IPTG for 2 hrs (10 µg); Lane 3: Soluble fraction containing AhR from a culture grown in arginine (50 mM), BA (10 mM), induced with 20 µM IPTG, overnight (10 µg). In both cases, protein expression was carried out at 18°C and lysis buffer H was used for re-suspension. Blotting was performed using anti-his tag as the 1°antibody.

Although the overall expression of $H_6$-SUMO-AhR improved when using low IPTG concentration over an overnight induction period, immunoblotting revealed that inducing with high IPTG for a short period still seems to be the best option to prevent protein degradation and/or fragmentation (Figure 3.38B, Lane 2). With regards to the expression of soluble $H_6$-SUMO-AhR, it is evident that the use of lysis buffer H significantly improved protein solubility (Figure 3.38, Lanes 2 and 4), when compared to previously tested buffers in which the presence of AhR in the soluble fraction was merely noticeable on SDS-PAGE (Figure 3.37, Lanes 2, 4, 6, 8, 10, 13, 15).

Based on these results, the ideal conditions for growth and protein expression of $H_6$-SUMO-AhR that were experimentally determined are summarised in Table 3.12.

**Table 3.11: Conditions for $H_6$-SUMO-AhR protein expression**

| | |
|---|---|
| **Point of induction** | High $OD_{600}$ (0.7-.8) |
| **Concentration of inducer** | 0.5 mM IPTG |
| **Temperature** | Bacterial culture is grown at 37°C until the point of induction. Protein expression is carried out at 18°C |
| **Additives** | 10 mM benzyl alcohol<br>50 mM arginine OR 3% (v/v) ethanol + 400 mM sorbitol |
| **Growth after induction** | 2 h |
| **Lysis buffer** | 50 mM Tris-Cl pH 8.5, 50 mM NaCl, 5 mM EDTA, 1 M urea, 1 mg $mL^{-1}$ lysozyme |

Considering the degradation susceptibility of AhR and fact that AIP is known to stabilise AhR in the cytoplasm (Petrulis and Perdew, 2002), the possibility of co-expression was also investigated. Chemically competent BL21 (DE3) cells were co-transformed using the heat-shock method (Section 2.1.4), with the modification of using

20 ng of each plasmid DNA, pTH-$H_6$-AIP and pET28a-$H_6$-SUMO-AhR. The success of co-transformation was assessed by DNA gel electrophoresis.

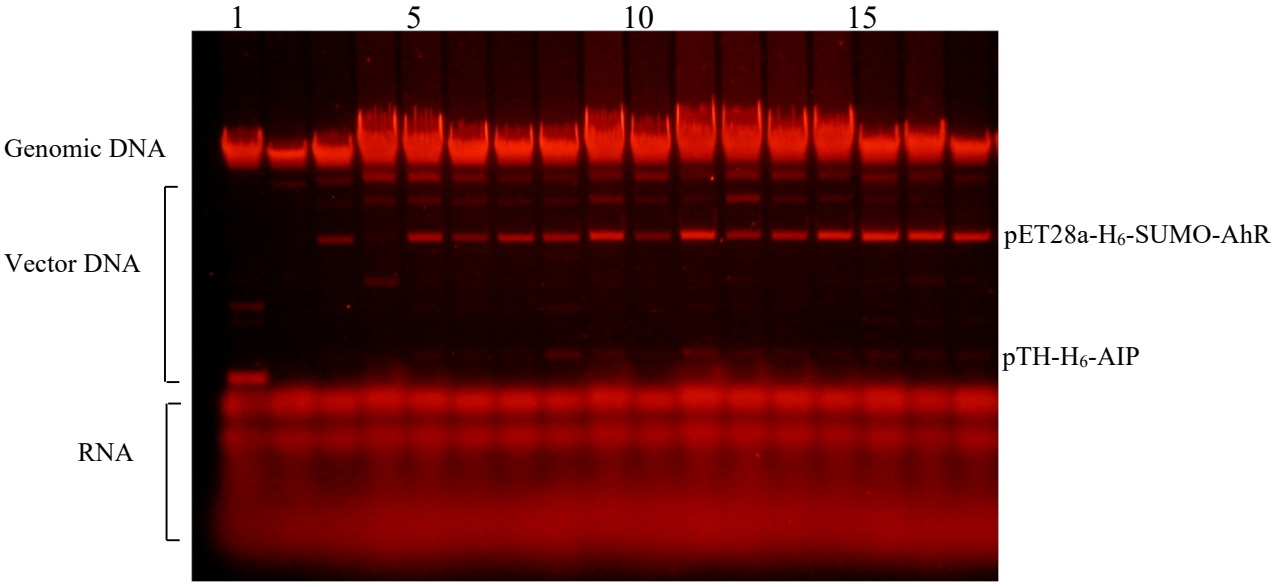



**Figure 3.39: Co-transformation DNA analysis:** Lane 1: DNA from cells transformed with pTH-$H_6$-AIP (control); Lanes 2-3: DNA from cells transformed with pET28a-$H_6$-SUMO-AhR (control). Lanes 4-17: DNA from cells that were co-transformed with pTH-$H_6$-AIP and pET28a-$H_6$-SUMO-AhR.

Though the band corresponding to the recombinant plasmid pTH-$H_6$-AIP is not as pronounced as that of pET28a-$H_6$-SUMO-AhR, samples in lanes 8, 9, 11-17 contain both DNA constructs. Successful co-transformed *E-coli* cells were grown in the presence of both ampicillin and kanamycin and protein expression carried under the established ideal conditions to minimize AhR degradation. Whilst an AhR band was observed at 18°C, no AIP was expressed at this low temperature. This was confirmed by immunoblotting. On the other hand, when expression was carried at 37°C, AIP was produced but no apparent AhR band was detected on SDS-PAGE. Immunoblotting showed extensive AhR degradation as had been previously observed in the temperature-controlled experiments at this temperature (Figure 3.33).

# Chapter 4

# Results: Protein Characterisation

## 4.1. Mass Spectrometry

All samples were subject to mass spectrometry to confirm their molecular weight. This was conducted on an electro-spray ionisation mass spectrometer that uses a soft ionisation technique without fragmentation and is thus ideal for determination of molecular weight of native proteins.

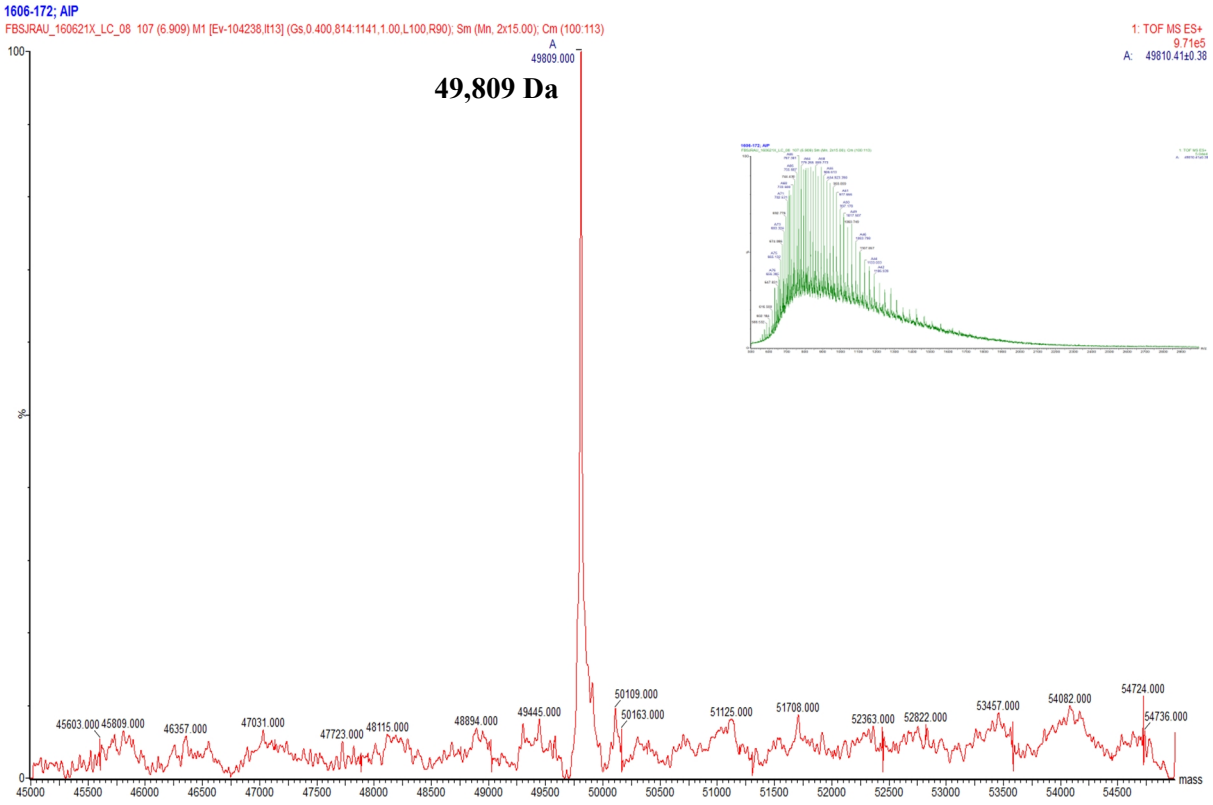



**Figure 4.1: Mass spectrum of $H_6$-SUMO-AIP.** A mass-to charge spectrum (m/z) shown in green transformed to a mass-only scale spectrum, shown in red.

The heaviest ion has a molecular weight of 49,809 Da that is in agreement with the expected molecular weight of $H_6$-SUMO-AIP using Protean (DNASTAR, Lasergene), based on the sequenced gene. The other $H_6$-SUMO tagged mutant counterparts also gave the expected molecular size; R9Q: 49,781 Da, R16H: 49,793 Da, V49M: 49,839 Da and K103R: 49,841 Da. The mass spectrum of each protein can be found in Appendix A.2.

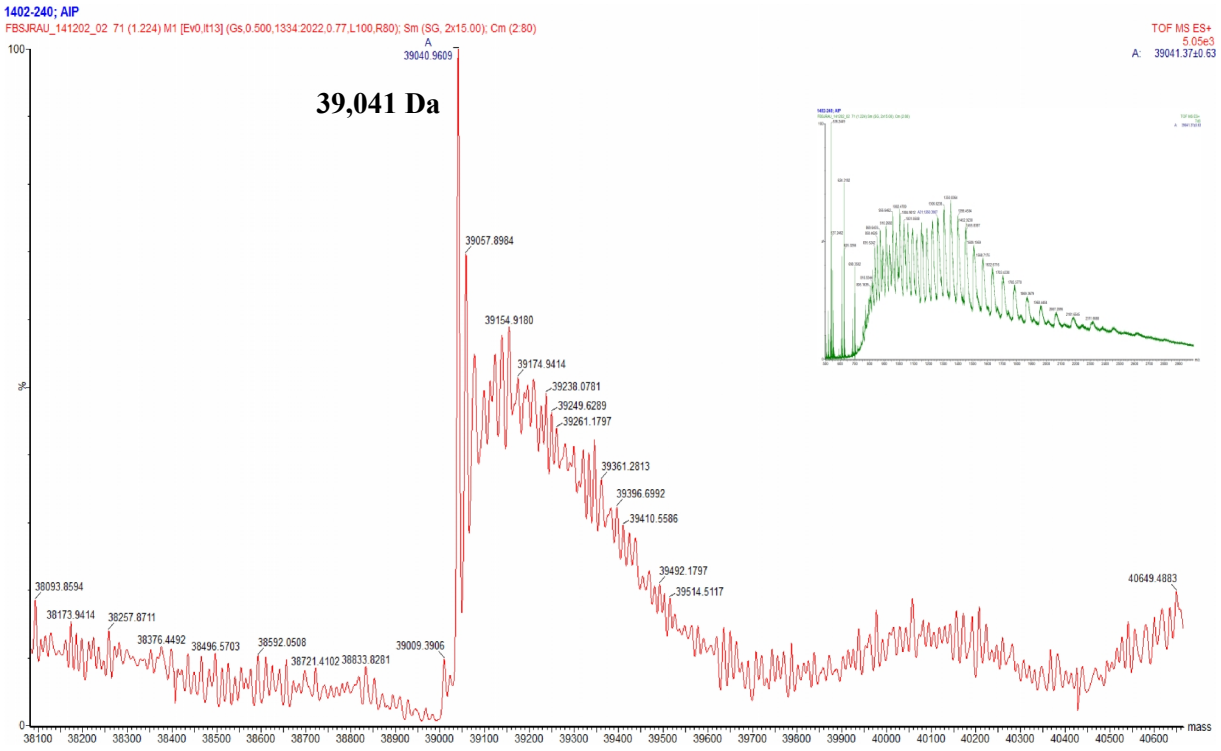



**Figure 4.2: Mass spectrum of $H_6$-AIP.** A mass-to charge spectrum (m/z) shown in green transformed to a mass-only scale spectrum, shown in red.

The heaviest ion has a molecular weight of 39,041 Da that is in perfect agreement with the predicted molecular weight from the sequence gene. Similarly, the $H_6$-mutants were analysed and found to have the expected molecular weight; R9Q: 39,014 Da, R16H: 39,022 Da, V49M: 39,073 Da and K103R: 39,069 Da.

Unlike the spectrum of the $H_6$-SUMO-AIP that showed a distinct sharp peak, the peaks for the $H_6$-tagged sample are quite broad indicating some heterogeneity with respect to the mass. This is most likely due to adducts to the protein, such as sodium ions which were present in all the buffers used throughout purification. In addition, the presence of peaks of lower molecular weight is probably due to degradation fragments that may suggest that the $H_6$-tagged protein is less stable in ammonium acetate, than the $H_6$-SUMO-AIP counterpart. The SUMO-tag may have conferred additional stability to the protein under the conditions used for mass spectrometry. The molecular weight of $H_6$-AIP was also determined by in-gel tryptic digestion followed by liquid chromatography mass spectrometry (LC-MS) (EMBL-Heidelberg), which confirmed sample purity and its molecular weight, as 39,041 Da. Data is shown in Appendix A.2.

## 4.2. Circular dichroism spectroscopy

### 4.2.1. Far-UV spectra

All proteins that were utilised in this study were initially characterised by circular dichroism spectroscopy. As eukaryotic proteins may exhibit misfolding when over-expressed in a bacterial host cell system, it was pivotal to ascertain that the secondary structure of the protein was maintained after purification. $H_6$-AIP and $H_6$-SUMO-PDE4A5 were the proteins used for each of the spectra presented in this section. For simplicity, they will be referred to as AIP and PDE4A5 respectively in Figures 4.3-4.10.

The molar ellipticity [θ] was calculated using the following equation, adapted from the University of Leeds website, circular dichroism section (2019).

$$[\theta] = \frac{\theta}{10\, c\, l} \tag{4.1}$$

Where; θ = ellipticity in degrees, c is the molar concentration of the protein sample (M) and l is the pathlength of the cuvette in cm.

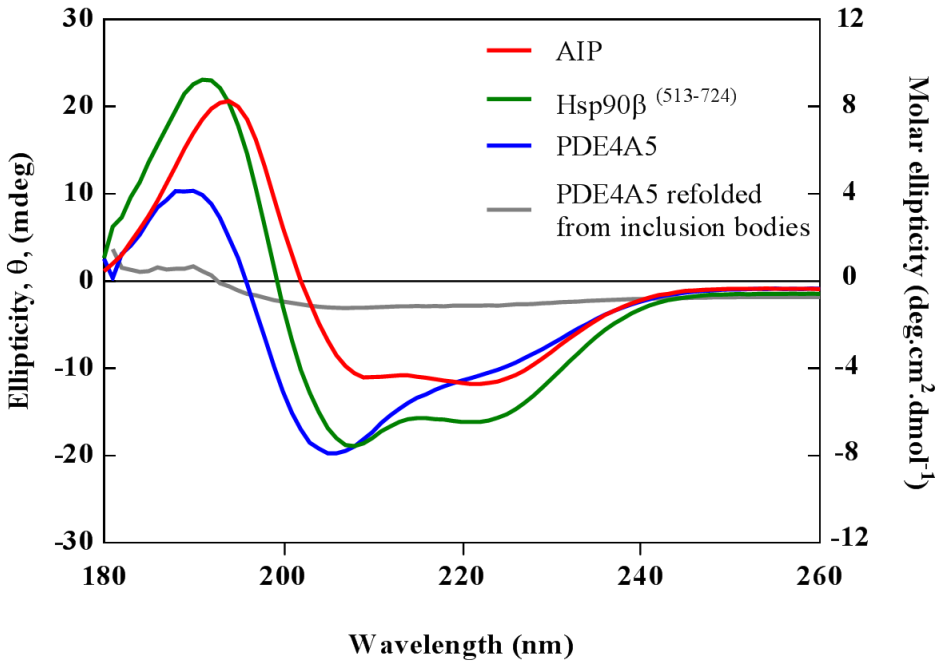



**Figure 4.3: Far-UV CD spectrum of AIP and binding partners.** All proteins were prepared at an equal molar concentration of 2.5 µM.

$H_6$-AIP and [4]Hsp90β$^{(513-724)}$ both have a higher proportion of α-helices as reflected by the characteristic peak maxima at 222 nm, 208 nm and 190 -193 nm. The spectrum of PDE4A5 shows a negative peak of great magnitude at 206-207 nm, a shoulder at around 222nm and a positive peak at 190 nm, reflecting a combination of random coil, α-helices and β-sheets. The latter generally exhibit a negative band around 210-220 nm and a positive peak within the region of 195-200 nm (Kelly *et al*., 2005). The low ellipticity (θ) values obtained in the case of the refolded PDE4A5 indicate that the purification protocol utilised in denaturing and refolding of the protein (Section 2.3.3.1) was not successful in yielding stable functional protein with secondary structure. All other proteins had an ordered secondary structure, reflecting a successful purification technique.

For a more accurate representation of secondary structure, the same data was analysed by CDNN (Bohm, Muhr and Jaenicke, 1992).

**Table 4.1: Secondary structure prediction**

| **180-260 nm** | **Secondary structure content (%)** | | | | | |
|---|---|---|---|---|---|---|
| **Protein** | **α-Helix** | **Antiparrallel β-sheets** | **Parallel β-sheets** | **β-Turn** | **Random Coil** | **Total Sum** |
| **AIP** | 33.2 | 8.2 | 8.8 | 16.8 | 33.2 | 100.1 |
| **Hsp90β $^{(513-724)}$ (dimer)** | 49.8 | 5.0 | 5.7 | 14.2 | 23.8 | 98.5 |
| **PDE4A5** | 23.0 | 13.1 | 11.4 | 19.5 | 33.9 | 100.9 |

[4] AIP purifies as monomer whilst Hsp90β purifies as a dimer. All the concentrations stated for Hsp90β are calculated based on the dimer molar concentration. A ratio of 2:1 thus signifies that two AIP molecules are present in solution per every dimer of Hsp90β.

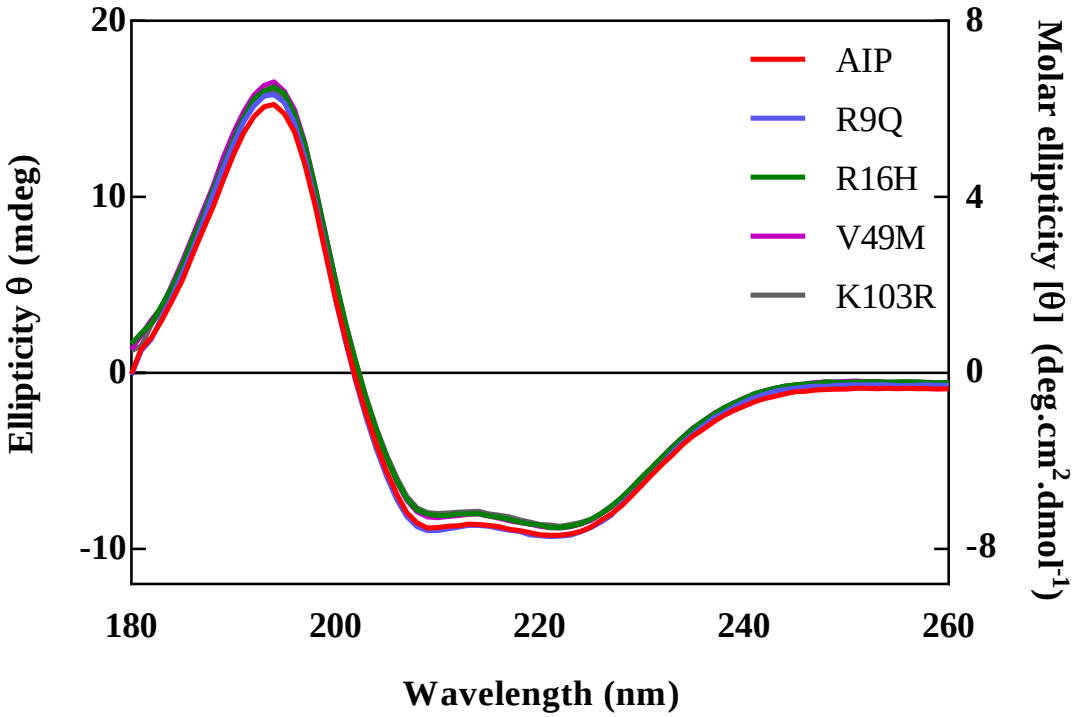



**Figure 4.4: Far-UV CD spectrum of $H_6$-AIP and $H_6$-AIPmutants.** All proteins were prepared at an equal molar concentration of 2.5 µM that is equivalent to 0.1 mg.mL$^{-1}$.

All the mutant AIP proteins exhibit a spectrum identical to the wild-type protein that is typical of a protein with a higher proportion of α-helices, as shown by the negative peaks at 222 nm and 208 nm of similar magnitude and the positive ellipticity at 193 nm.

The melting temperature of AIP and AIP mutants, both $H_6$- and $H_6$-SUMO-tagged was determined using the circular dichroism data at 222 nm over the temperature range of 5-90°C. CD ellipticity data (θ) was normalised as shown in Equation 4.2;

$$\mathrm{CD}norm(\mathrm{T}) = \frac{CD\ (T) - fL\ (T)}{fH\ (T) - fL\ (T)} \qquad (4.2)$$

where CD*norm* is the normalised CD value, CD is the experimental ellipticity value, $f_H$ and $f_L$ represent the ellipticity value at the highest and lowest temperature (T) respectively.

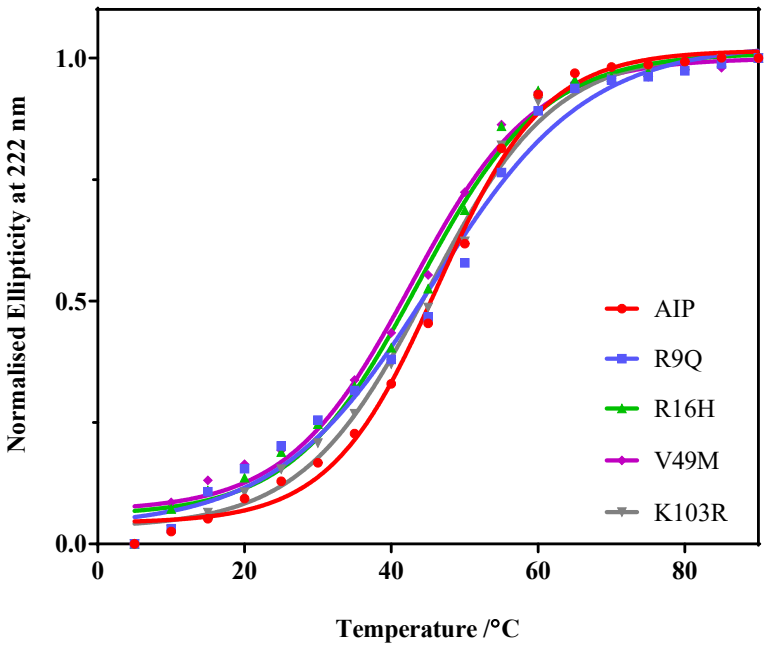



**Figure 4.5:Melting temperature determination.** Normalised CD data at 222 nm for $H_6$-AIP and mutants. Samples were prepared at a concentration of 0.2 mg.mL$^{-1}$.

The melting temperature of AIP and mutant counterparts was determined using CD data at a fixed wavelength of 222 nm. Data was normalised using the equation 4.2 and fitted using a non-linear regression with a least squares fitting method and a Boltzmann Sigmoidal equation. The confidence level was set at 95%. The $R^2$ value was used as a representation of the goodness of fit. All CD analysis was performed using GraphPad Prism Version 6.0.

**Table 4.2: Melting temperature determination of AIP and mutant derivatives.**

| Sample | Melting Temperature / °C | | $R^2$ value | |
|---|---|---|---|---|
| | $H_6$- | $H_6$-SUMO- | $H_6$- | $H_6$-SUMO- |
| **AIP** | 46.7 ± 0.7 | 46.3 ± 0.6 | 0.998 | 0.996 |
| **R9Q** | 45.4 ± 0.7 | 45.5 ±1.3 | 0.997 | 0.990 |
| **R16H** | 44.3 ± 0.6 | 43.6 ± 0.9 | 0.997 | 0.994 |
| **V49M** | 42.1 ± 0.6 | 42.7 ± 0.9 | 0.996 | 0.994 |
| **K103R** | 43.5 ± 0.5 | 45.0 ± 0.7 | 0.998 | 0.996 |

The results indicate that the mutations did not significantly affect the thermal stability of the protein, as each mutant had a melting temperature similar to the wild type AIP.

Aside from secondary structure prediction and melting point determination, circular dichroism was also used to study protein-protein interactions by monitoring changes in the far-UV (180-260 nm) and near-UV (250 -320 nm) resulting from conformational changes (Greenfield, 2004). Mixtures AIP with binding partners Hsp90β$^{(513-724)}$ and PDE4A5 were analysed at different incubation conditions.

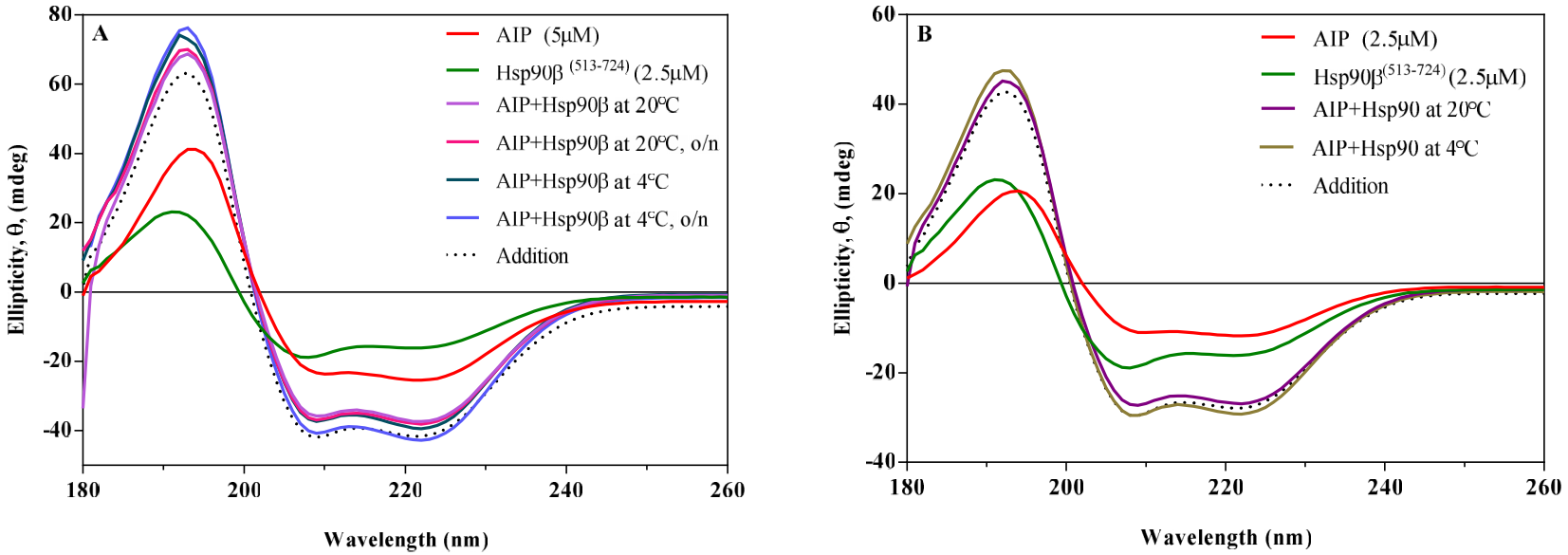



**Figure 4.6: Far-UV CD spectra of AIP, Hsp90β $^{(513-724)}$ and an equimolar mixture of the two purified proteins A**: Solutions of AIP (10 μM) and Hsp90β$^{(513-724)}$ (5 μM) were mixed in equal volumes and mixtures incubated at 4°C and 20°C for 1 hr. The same samples were re-measured after overnight incubation at the same temperature. **B:** Solutions of AIP (5 μM) and Hsp90β$^{(513-724)}$ (5 μM) were mixed in equal volumes and incubated for 1 hr at 4°C and 20°C. In each mixture, the final molar concentration of each protein was diluted by half, by the presence of the second protein. The CD spectrum of each mixture was compared to the spectrum of the sum of the individual proteins (dotted line).

In both figures, the far-UV spectra show that there is relatively minimal change (if any) in the secondary structure upon protein mixing. This could be interpreted either as the absence of protein binding under the prevailing conditions used or that binding did occur however it caused no changes in secondary structure that could be observed in the far-UV spectrum. When compared to the sum of the individual spectra (dotted line), the spectra of the mixtures show only slight changes in ellipticity at wavelengths 192, 208 and 222 nm. The overall shape of the spectra remained the same with no observable shifts in the peak maxima and minima. Nonetheless, for a fairer comparison, the ellipticity values at 192, 208 and 222 nm for each of the experimental mixtures were noted and the $\theta_{222}/\theta_{208}$ ratio was calculated which gives insight on the spatial distribution of the α-helix.

**Table 4.3: Table of Ratio $\theta_{222\ nm}$ / $\theta_{208\ nm}$**

| | **Ratio $\theta_{222\ nm}$ / $\theta_{208\ nm}$** | |
|---|---|---|
| **Incubation conditions** | **Mixture of AIP-Hsp90β$^{(513-724)}$** | **Addition of the individual spectra of AIP and Hsp90β$^{(513-724)}$** |
| | Ratio: AIP: Hsp90β – 2:1 | |
| **1 hr at 20°C** | 1.06 | 1.00 |
| **1 hr at 4°C** | 1.07 | |
| **16 h (o/n) at 20°C** | 1.05 | |
| **16 h (o/n) at 4°C** | 1.07 | |
| | Ratio: AIP: Hsp90β- 1:1 | |
| **1 hr at 20°C** | 0.99 | 0.95 |
| **1 hr at 4°C** | 0.99 | |

An $\theta_{222}/\theta_{208}$ ratio of ≥ 1.0 usually indicates the presence of helices within a coiled coil structure whereas values ≤ 0.9 reflect the presence of helices in solution (Lau, Taneja and Hodges, 1984; Shepherd, Hoang, Abbenante and Fairlie, 2005). The main observation that can be inferred from these values is that in all cases, the $\theta_{222}/\theta_{208}$ ratio of the mixture was higher than the expected value from the sum of the individual spectra, thus favouring the presence of helices as part of a coiled coil structure. However, since these changes are subtle, whether or not these represent complex formation remains inconclusive.

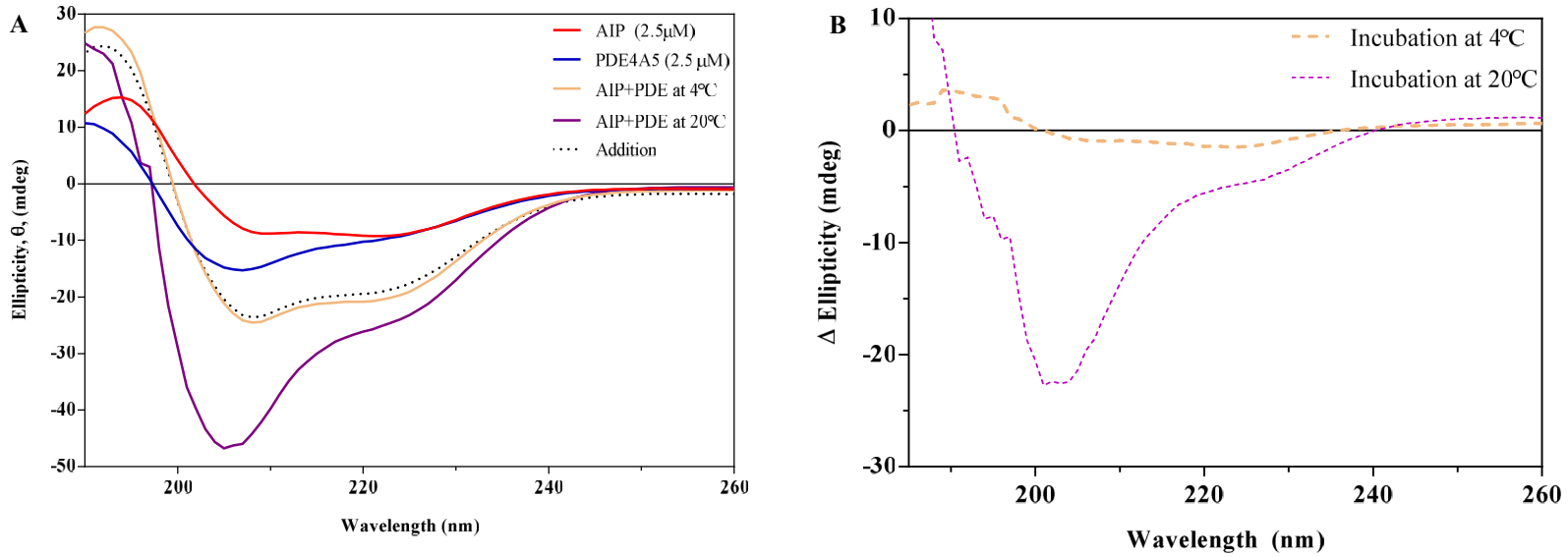



**Figure 4.7: Far-UV CD spectra of AIP, PDE4A5 and equimolar mixtures of the two purified proteins. A:** Solutions of AIP (5 μM) and PDE4A5 (5 μM) were mixed in equal volumes and incubated for 1 hr at 4°C and 20°C. **B:** Spectra showing the difference in ellipticity between the AIP-PDE4A5 mixture and the sum of the individual pure components at the two temperatures investigated, 4°C and 20°C. This was obtained by subtracting the CD spectrum of each mixture from the spectrum of the sum of the unmixed proteins.

In panel A, the mixture incubated at 4°C overlaps almost precisely with the expected ellipticity spectrum from the sum of the two individual proteins. However, when the mixture was incubated for the same amount of time at room temperature, the spectrum exhibited a significantly large minimum around 205-208 nm. The mixture also exhibited an increase in ellipticity at 222 nm (from -20.04 to -25.15 x$10^{-3}$ °). These differences can be interpreted as changes in secondary structure upon complex formation and are better represented in panel B.

Following the changes observed in the far-UV of AIP-PDE4A5 mixture, the same incubation experiment was carried out using the mutant counterparts of AIP. Each individual protein was at a constant concentration of 2.5μM and mixed in a ratio of 1:1. Mixtures were incubated at 20°C for 1 hr.

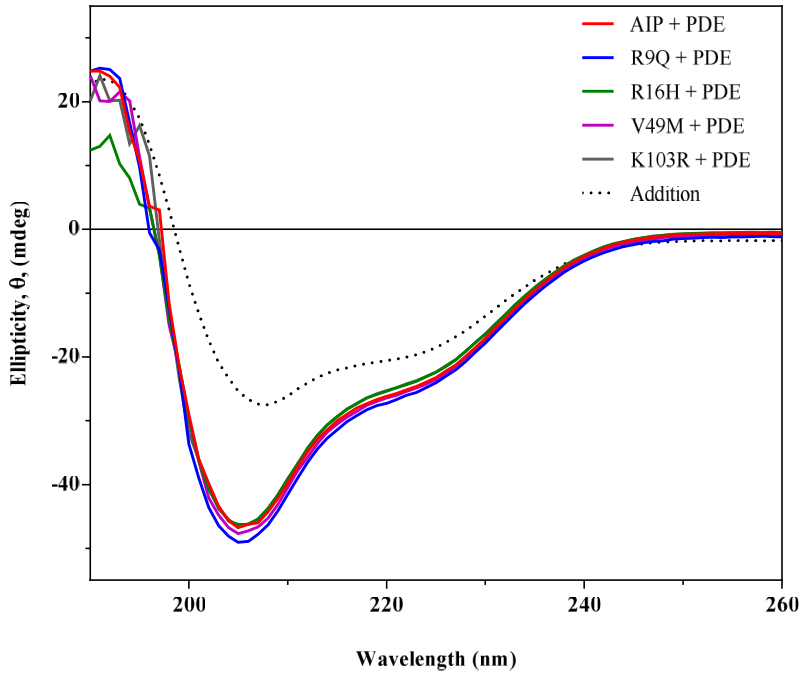



**Figure 4.8: Far-UV spectra of AIP and $H_6$-AIP-mutants in an equimolar mixture with PDE4A5.**

The far-UV spectra of AIP and all mutant counterparts in mixture with PDE4A5 overlap almost precisely. In all cases, there is a negative ellipticity peak at around 205-208nm, which is larger than it would be expected from the sum of the individual proteins together. A peak is observed at 190-192 nm. The intensity of this peak is lower in the case of the R16H mutant, followed by V49M, which may be due to a concentration issue. Mutants K103R and R9Q have an intensity equal to the wild type AIP protein, which coincides with the expected ellipticity spectrum (dotted line). This shows that any changes in secondary structure that occur in the wild type protein upon binding to PDE4A5 are also exhibited by the mutant counterparts. Binding affinities were then investigated through SPR experiments (Section 4.4).

### 4.2.2. Near-UV spectra

AIP, Hsp90β$^{(513-724)}$, PDE4A5 and their respective combinations were also analysed in the near-UV (250-320 nm). Absorption within this region is due to aromatic amino acids in the protein sample and is considered as the tertiary structure 'fingerprint' of proteins (Kelly *et al*., 2005).

**Table 4.4: Table: Aromatic amino acids in AIP, Hsp90β$^{(513-724)}$ and PDE4A5**

| | AIP | Hsp90β$^{(513-724)}$ (per monomer) | PDE4A5 |
|---|---|---|---|
| **Trp** | 4 | 1 | 10 |
| **Tyr** | 11 | 3 | 17 |
| **Phe** | 9 | 4 | 27 |

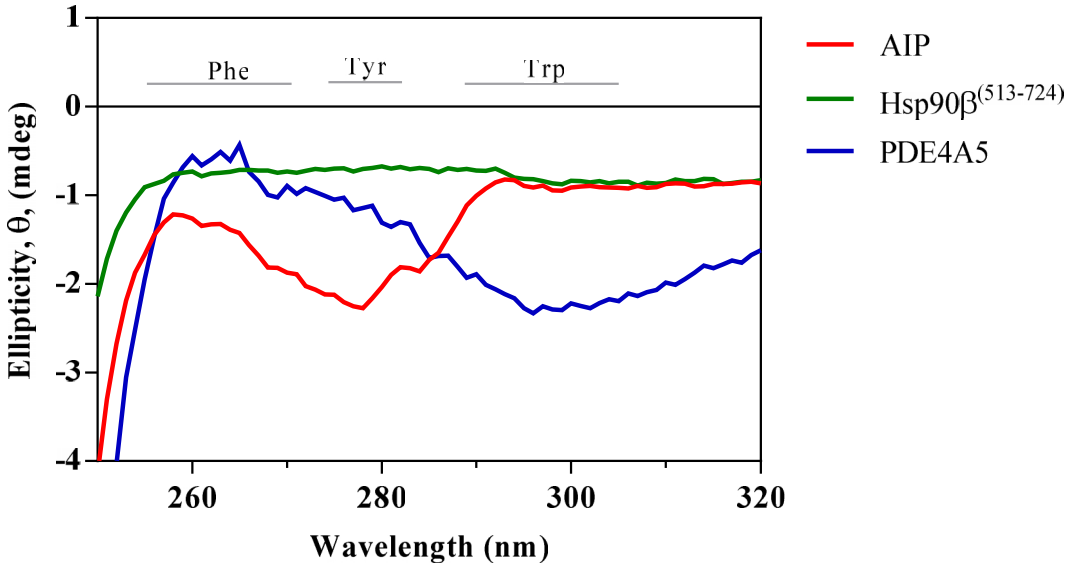



**Figure 4.9: Near-UV CD spectra of AIP and its binding partners, Hsp90β$^{(513-724)}$ and PDE4A5.** All proteins were set at a concentration of 10 µM.

The spectrum of AIP exhibits a peak at 258 nm, followed by a minimum at 278 nm and a shoulder at 290 nm, the latter being the characteristic peak maxima of tryptophan. PDE4A5 shows a peak with fine structures within the phenylalanine absorption spectrum (255-270 nm), which reflects the higher number of Phe present in the protein. No absorbance peaks are observed in the spectrum of Hsp90β$^{(513-724)}$ which is not surprising when considering the low amount of aromatic amino acids present (Table 4.4). The shape and magnitude of the spectrum does not only depend on the

number of aromatic amino acids present, but also on their mobility, electrostatic environment and spatial disruption within the protein (Kelly, Jess, Nicholas and Price, 2005).

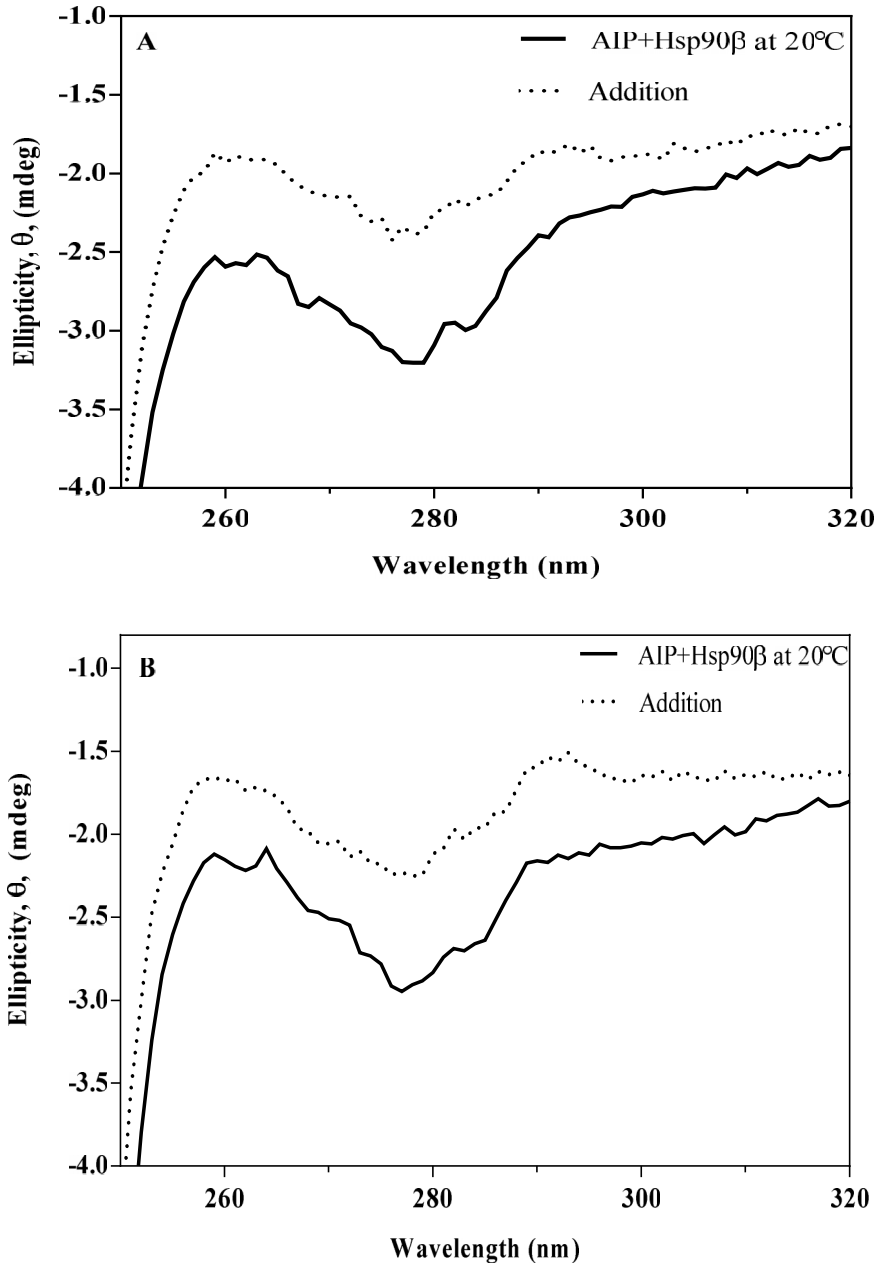



**Figure 4.10: Near UV CD spectra of AIP-Hsp90β$^{(513-724)}$ mixtures.** A: AIP (10 μM) and Hsp90β$^{(513-724)}$ (5 μM) were mixed in equal volumes and incubated for 1 hr at room temperature, B: AIP (10 μM) and Hsp90β$^{(513-724)}$ (10 μM) were mixed in equal volumes and incubated for 1 hr at room temperature. The CD spectrum of the resulting AIP-Hsp90β$^{(513-724)}$ mixture is compared with the sum of the spectra of the individual proteins (dotted line).

In both cases, the mixture showed more negative ellipticity than the expected sum of the individual proteins, which is likely not due to a concentration issue. Furthermore, both addition spectra showed a peak at 290 nm, which is characteristic of the amino acid tryptophan that is missing in the spectra of the mixtures. This suggests protein binding resulting in subtle changes in tertiary structure. The peak at 290 nm is more prominent in panel B, indicating that an AIP:Hsp90β$^{(513\text{-}724)}$ ratio of 1:1 is probably better than 2:1 ratio for complex formation.

The near-UV CD spectra of combinations of AIP-PDE4A5 mixed in equimolar ratios and incubated at room temperature, were also investigated. The spectrum of the mixture showed the same shape as the spectrum of the sum of the individual protein, suggesting that complex formation alters the secondary structure, as observed in Figure 4.7, but has no apparent detectable effect on the tertiary structure of the proteins.

## 4.3. Protein-protein interactions: Isothermal titration calorimetry (ITC)

ITC is a quantitative measure of the thermodynamic properties of protein-protein interactions. When a ligand is titrated into the sample cell containing the binding partner, the amount of heat energy that is released or absorbed by the interaction is detected and quantified. It is the only technique whereby in a single experiment, the equilibrium binding association ($K_A$) and dissociation constants ($K_D$) are obtained, together with the stoichiometry (*n*) and the enthalpic ($\Delta H$) and entropic ($\Delta S$) parameters that define the binding interaction (Pierce, Raman and Nall, 1999).

This technique was only utilised to study the binding kinetics of the interaction between AIP and Hsp90β$^{(513-724)}$. Due to the difficulties in obtaining a high yield of pure monomeric $H_6$-SUMO-PDE4A5 and since ITC requires high concentrations of protein, it was not experimentally feasible to measure the binding interaction of PDE to AIP by this technique. The interaction between AIP and $H_6$-SUMO-PDE4A5 was characterised through surface plasmon resonance (SPR) that requires lower quantities of pure protein. SPR will be discussed in the following section (Section 4.4).

The interaction of Hsp90β $^{(513-724)}$ was initially analysed using both $H_6$-SUMO-AIP and $H_6$-AIP. All proteins were in 1x PBS, 5% (v/v) glycerol (Section 2.3.10). DTT was completely removed for this experiment as it may cause an irregular baseline and may also produce a heat signal that would mask the enthalpy of binding. Similarly, the glycerol concentration was reduced from 10% (v/v) (storage conditions) to 5% (v/v). Glycerol was not completely removed from the protein sample as its presence prevents the formation of multimeric states and or/ aggregates particularly in the case of Hsp90β$^{(513-724)}$ that was used as the titrant in this experiment and was thus prepared at a high concentration of 150 µM. The experimental heat can be significantly affected by the presence of reactant degradation and or aggregates, leading to measurement inaccuracies (Velazquez-Campoy, Leavitt and Freire, 2015).

The heat of interaction of water into water, buffer into buffer and Hsp90β$^{(513-724)}$ into buffer were measured at the start of the experiment as controls. This ensured that the system was clean, all solutions had been properly degassed and that the buffer composition was suitable for the experiment (Figure 4.11).

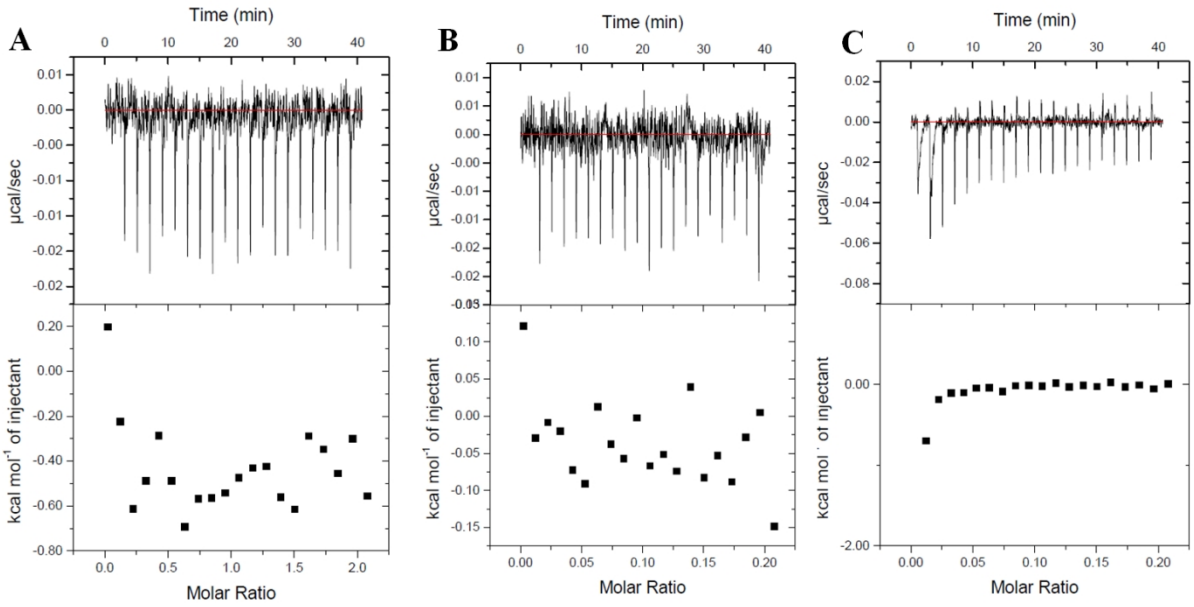



**Figure 4.11: ITC thermograms of control reactions: A:** Injection of water into water, **B:** Injection of buffer into buffer; **C:** Injection of Hsp90β$^{(513-724)}$ into buffer. For each measurement, the upper panel represents the raw heat signal versus time and the lower panel shows the integrated heat peaks of each injection as a function of the molar ratio.

The data generated from Hsp90β$^{(513-724)}$ into buffer titration was used as a reference and subtracted from the experimental titrations of Hsp90 $^{(513-724)}$ into AIP. Data was fitted using a non-linear least squares curve fitting algorithm (Microcal Origin Software).

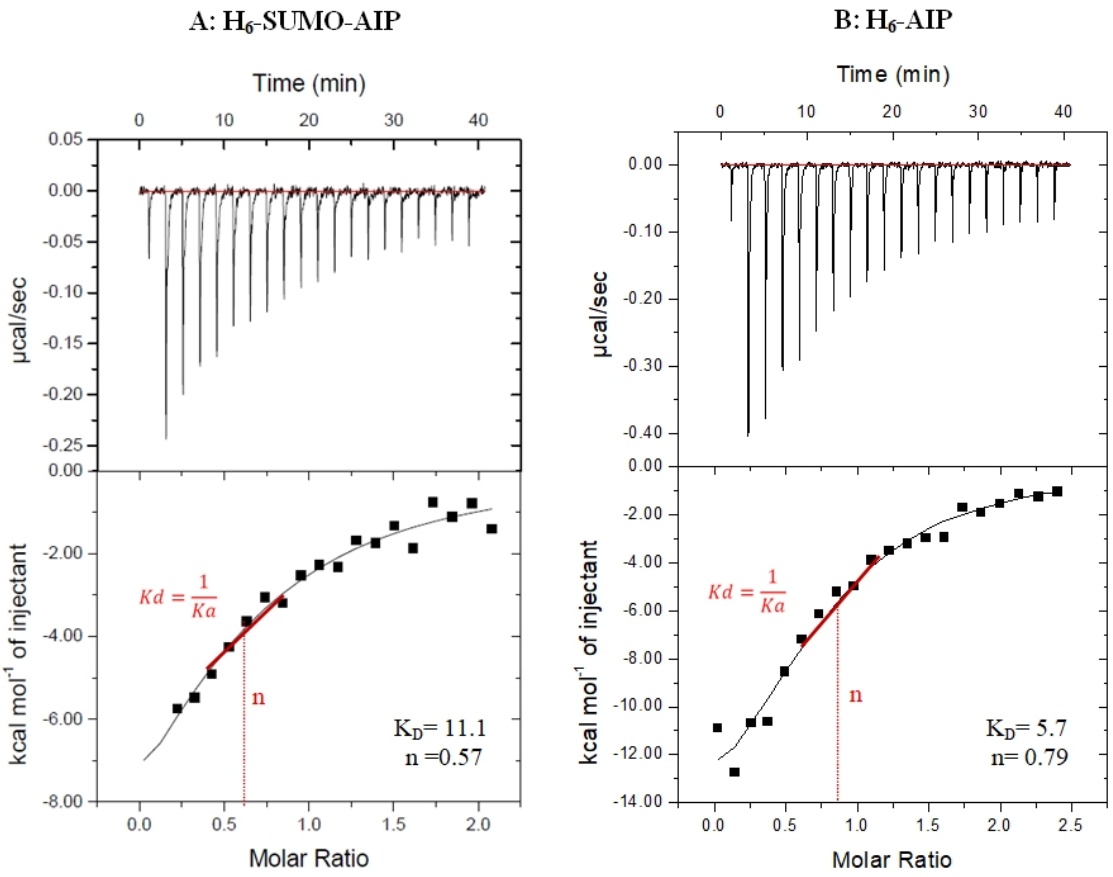



**Figure 4.12: ITC thermogram.** The upper panel shows the thermogram of the interaction between A: $H_6$-SUMO-AIP and Hsp90β$^{(513-724)}$ and B: $H_6$- AIP and Hsp90β$^{(513-724)}$. The titration thermogram is represented as heat per unit time that is released with every injection of the ligand, Hsp90β$^{(513-724)}$ when titrated into sample cell containing AIP. The lower panel represents the integrated heat peaks as a function of the molar ratio. The molar ratio is expressed as a fraction of the syringe to cell concentration (Hsp90β$^{(513-724)}$ and AIP respectively). The squares represent the experimental data whilst the line corresponds to the best fitting model. The binding parameters are also indicated ($K_D$; equilibrium dissociation constant and n, stoichiomtry factor).

Injections were repeatedly performed as indicated by each peak. The size of the peaks eventually became smaller due to protein saturation. The constant size of the peaks at the end of the titration indicates complete saturation of AIP, at which point the peaks are only representing the heats of dilution of Hsp90β$^{(513-724)}$ into buffer.

The thermograms show that the interaction is exothermic as indicated by the presence of negative peaks. In both cases the curve is not evidently sigmoidal, a clear indication of a weak binding affinity (within the μM range). Integration and curve fitting showed a 2-fold difference between $K_D$ values. The experiment was repeated thrice for $H_6$-AIP and twice for $H_6$-SUMO-AIP for reproducibility (Table 4.5). Owing to the fact that the $H_6$-SUMO is a relatively large peptide tag (12 kDa) at the N-domain of the protein, it was considered appropriate to perform further ITC experiments using the $H_6$-proteins. ITC offers the advantage that the proteins are free to move in solution and so the use of the $H_6$-proteins eliminated the possibility that any potential binding sites in the N-domain may be masked or in any way affected by the presence of the $H_6$-SUMO-tag.

**Table 4.5: Thermodynamic ITC parameters (A)-AIP-Hsp90β$^{(513-724)}$ injection**

**Buffer:** **1x PBS pH 7.5, 5% (v/v) glycerol**

| Protein | n | $K_A$ ($M^{-1}$) (x $10^5$) | $K_D$ (μM) | ΔH (kcal $mol^{-1}$) | ΔS (cal $mol^{-1}$ $deg^{-1}$) |
|---|---|---|---|---|---|
| | | | **$H_6$-SUMO-AIP** | | |
| **Exp 1** | 0.57 ± 0.21 | 0.89 | 11.2 | -15.3 ± 6.8 | -28.6 |
| **Exp 2** | 0.41 ± 0.13 | 0.93 | 10.5 | -35.0 ± 19.1 | -94.8 |
| | | | **$H_6$-AIP** | | |
| **Exp 1** | 0.79 ± 0.07 | 1.76 | 5.7 | -17.9 ± 2.1 | -36.2 |
| **Exp 2** | 0.61 ± 0.05 | 2.07 | 4.8 | -16.8 ± 2.2 | -32.1 |
| **Exp 3** | 0.69 ± 0.05 | 2.39 | 4.2 | -15.9 ± 1.6 | -28.6 |
| | | | **$H_6$-AIP mutants** | | |
| **R9Q** | 0.65± 0.10 | 0.91 | 10.9 ± 2.3 | -18.9 ± 4.8 | -40.8 |
| **R16H** | 0.46 ± 0.08 | 0.53 | 18.8 ± 2.5 | -31.8 ± 8.6 | -85.1 |
| **V49M** | 0.36 ± 0.21 | 0.47 | 21.3 ± 4.1 | -38.6 ± 10.3 | -65.1 |
| **K103R** | 0.55 ± 0.10 | 1.27 | 7.9 ± 0.8 | -21.1 ± 6.6 | -47.2 |

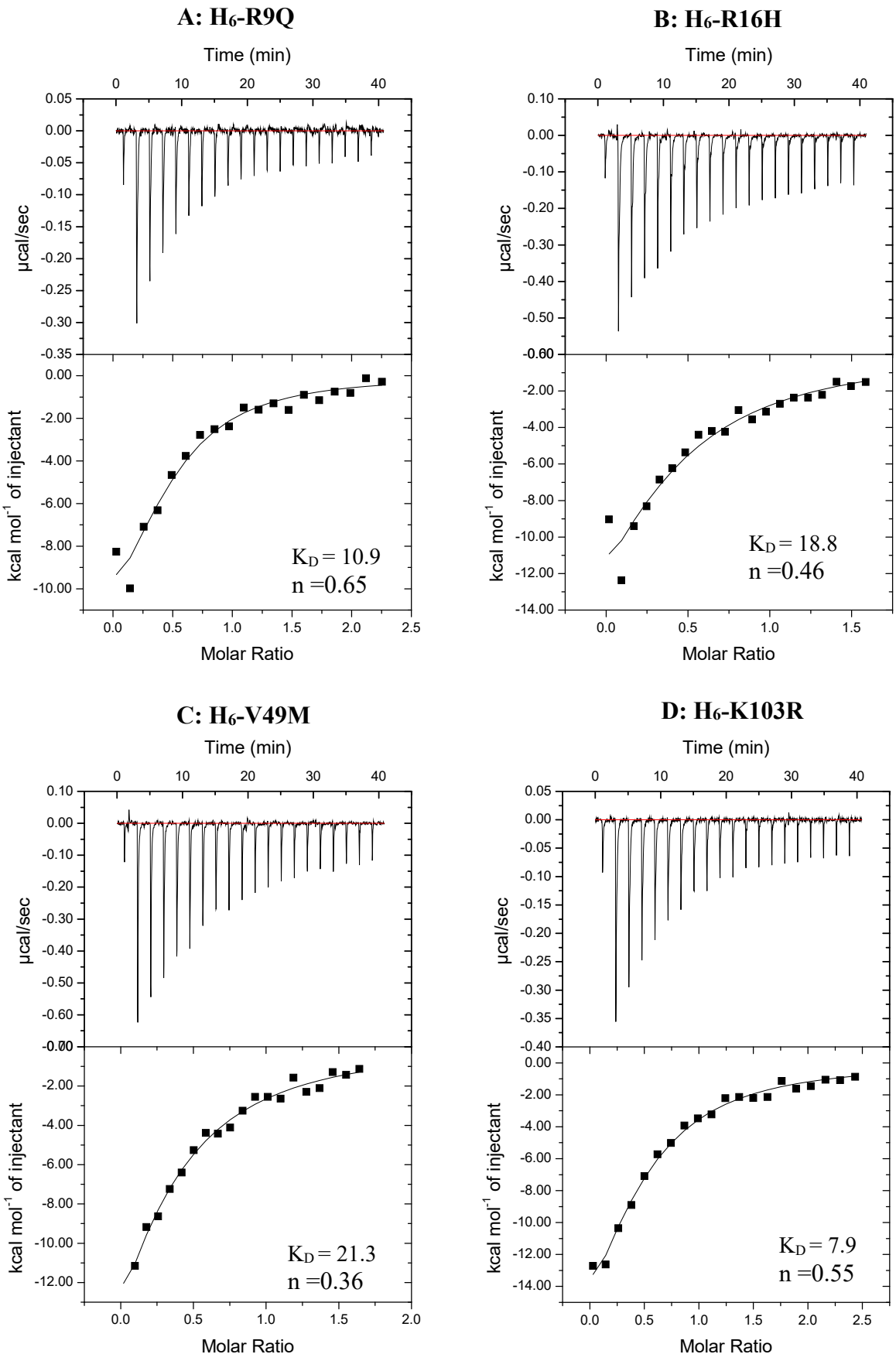



**Figure 4.13: ITC thermograms of H6-AIP mutants.** In each figure (A-D), the upper panel represents the raw data and the bottom panel shows curve fitting of the integrated heat peaks as a function of the Hsp90$^{(513\text{-}724)}$/H$_6$-AIP mutant molar ratio.

The results obtained demonstrate that the wild type AIP has a higher affinity to its binding partner than the mutant counterparts, R9Q, R16H and V49M. The R9Q mutant exhibits a 2-fold decrease in binding affinity, while a more pronounced difference was observed for the R16H and V49M mutants, that both displayed 4-fold lower affinity. The K103R mutation does not seem to have any effect on the binding to Hsp90β$^{(513-724)}$, as the $K_D$ value obtained is very comparable to that of the wild type protein.

Another notable difference was in the value of the stoichiometry factor (n). All mutants displayed an 'n' value close to 0.5, suggesting a stoichiometry ratio of 1:2, that corresponds to 2 molecules of $H_6$-AIP mutants binding to one dimer molecule of Hsp90β$^{(513-724)}$. This is only valid however under the assumption that all the protein molecules used during the experiment (both in the cell and syringe) were fully active (Malvern Panalytical, 2018). Since the wild type protein showed n-values that are closer to 1 (0.79 and 0.69 respectively), it may be possible that under these experimental conditions, the mutants were less stable and so less suitable for binding. To evaluate this possibility, the ITC experiments were repeated using different buffer components (20 mM Tris-Cl pH 7.5, 5mM NaCl, 5% v/v glycerol that are similar to the storage buffer (Table 2.9) and is comparable to the buffer components used by other groups (Morgan *et al*., 2012). These results showed an increase in the n-value for all the mutants, as well as lower $K_D$ values. In comparison to the wild type AIP however, each mutant exhibited the same fold difference in $K_D$ as observed when using 1x PBS as dilution buffer.

**Table 4.6: Thermodynamic ITC parameters (B)- AIP-Hsp90β$^{(513-724)}$ injection**

| **Buffer:** | **20 mM Tris-Cl pH 7.5, 5 mM NaCl, 5% v/v glycerol** | | | | |
|---|---|---|---|---|---|
| **$H_6$ tag** | **n** | **$K_A$ ($M^{-1}$) (x $10^5$)** | **$K_D$ (μM)** | **ΔH (kcal $mol^{-1}$)** | **ΔS (cal $mol^{-1}$ $deg^{-1}$)** |
| **AIP** | 0.74 ± 0.16 | 4.56 | 2.19 | -11.5 ± 3.2 | -12.7 |
| **R9Q** | 0.70 ± 0.15 | 1.85 | 5.40 | -8.7 ± 2.6 | -5.13 |
| **R16H** | 0.53 ± 0.29 | 1.27 | 7.80 | -23.7 ± 13.1 | -16.1 |
| **V49M** | 0.60 ± 0.40 | 0.93 | 10.70 | -10.3 ± 9.34 | -11.7 |
| **K103R** | 0.69 ± 0.07 | 3.47 | 2.88 | -13.0 ± 1.99 | -18.4 |

## 4.4. Protein-protein interactions: Surface plasmon resonance (SPR)

Binding interactions were also analysed using the technique of surface plasmon resonance, whereby Hsp90β$^{(513-724)}$ and $H_6$-SUMO-PDE4A5 were immobilised on the surface of a CM5 chip. This chip is composed of a glass slide that is coated with a thin film of gold, onto which carboxymethylated dextran is covalently bound. Dextran enables the covalent immobilisation of the ligand molecule and provides a hydrophilic environment to facilitate protein interactions between the immobilised ligand and the injected analyte flowing across its surface. The gold film is needed for the generation of the SPR response signal that is observed in real-time on a sensorgram (GE Healthcare, 2013).

### 4.4.1. pH scouting

A number of pre-concentration experiments (pH scouting) were conducted to identify the ideal pH that provides a sufficient degree of immobilisation of the ligand on the surface of the chip. The interaction between the ligand and the negatively charged dextran surface is electrostatic. The carboxylated dextran matrix carries an overall negative charge at pH values higher than 3.5. As such, the pH of the protein (ligand) should ideally be above 3.5 and below its isoelectric point (PI) to maximise the electrostatic adsorption of the protein to the dextran matrix (Murphy, Jason-Moller and Bruno, 2006).

Aliquots of Hsp90β$^{(513-724)}$ and $H_6$-SUMO-PDE4A5 were diluted in 10 mM sodium acetate at varying pH values (4.0, 4.5 and 5.5). Another aliquot of each protein diluted in 1x PBS pH 6.5 was also prepared. The low ionic strength of the buffer was necessarily to ensure that any binding observed is due to the positive charges on the protein (protonated amine groups) and not from any competing buffer ions (Leonard *et al*., 2017). Each aliquot was individually injected over the surface of the underivatised chip (non-chemically activated) and the degree of electrostatic binding was monitored in real-time. In between injections, a high salt wash (1x PBS, 1 M NaCl) was performed to remove the electrostatically bound ligand prior to the next injection. Buffers at pH of 5.0 were avoided as this pH corresponds to the pI of both Hsp90β$^{(513-724)}$ and $H_6$-SUMO-PDE4A5 (pI value of 5.1, based on protein sequence) at which point

their solubility is minimal and are therefore more likely to precipitate on the surface of the chip.

The results obtained indicate that pH 6.5 is the ideal buffer pH for protein immobilisation. Although higher levels of protein binding were observed when using sodium acetate at pH of 5.5, the resonance/response units (RUs) remained relatively high even after the high-ionic strength salt wash, a clear indication that the electrostatically bound protein was not being entirely removed from the surface of the chip. Lower levels of binding were observed particularly at pH 4.0. Similarly, the RU signal did not return to baseline following the salt wash, suggesting that the protein was most likely precipitating at these low pH values. An ethanolamine wash (1 M) had to be performed to re-establish the baseline RU signal (Figure 4.14).

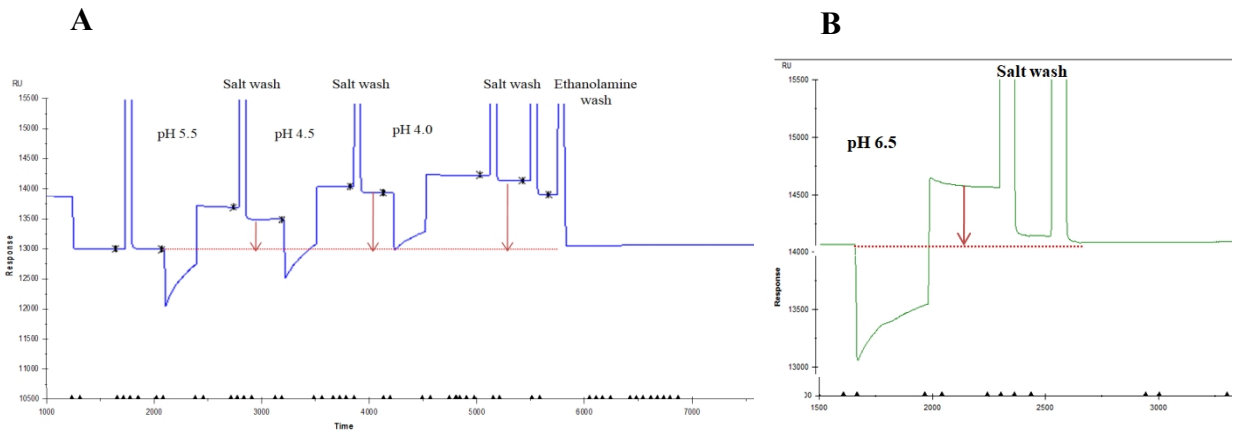



**Figure 4.14: Sensorgram of $H_6$-SUMO-PDEA45 during pH scouting.** The start and end of each injection is marked by report points (*). At each point, the vertical lines represent changes in refractive index at the interface due to protein binding. These alter the angle at which the reduced-intensity polarised light reflects. The change in refractive angle caused by binding or dissociation of protein is proportional to the mass of protein bound (GE Healthcare, 2019). The baseline RU is indicated by a red dotted line.

The pH value of 6.5 was the lowest pH that produced sufficient binding levels and where the electrostatically bound protein was efficiently washed off with a re-establishment of baseline RU signal. This pH was used for the subsequent ligand immobilisation step.

### 4.4.2. Ligand immobilisation

Protein immobilisation on the surface of the CM5 chip was carried out through amine coupling using an equal mixture of EDC and NHS (Section 2.3.9.2). This technique involves three major steps, i) the chemical activation of the surface, ii) the injection of the ligand over the activated surface until the required immobilisation level is reached and finally iii) the unreacted sites are blocked so they will not interfere during binding assays.

All four flow cells on the surface of the CM5 chip were chemically activated. Hsp90β$^{(513-724)}$ was diluted to 50 µg.mL$^{-1}$ in 1x PBS buffer pH 6.5 and was injected over the activated surface of flow cell 2. Protein binding was observed in real-time using the BIAevaluation software Version 3.1. Injection was manually stopped when a signal of over 500 RU was reached. Following a salt wash (Table 2.12), any other active sites on the surface of the chip were chemically blocked using 1 M ethanolamine pH 8.5. Similarly, $H_6$-SUMO-PDE4A5 was diluted to the same concentration in 1x PBS pH 6.5 and immobilised on flow cell 3. Flow cell 4 was immobilised with another sample of $H_6$-SUMO-PDE4A5 that served as a replicate. Since the molecular weight of PDE is almost twice in size that of Hsp90β$^{(513-724)}$, the injection of $H_6$-SUMO-PDE4A5 on both flow cells 3 and 4 was stopped when a signal of over 1000 RU was obtained. This ensured that the amount of immobilised ligand on flow cells 3 and 4 was comparable to the level of Hsp90β$^{(513-724)}$ immobilised on flow cell 2. Flow cell 1 was chemically treated as the other three flow cells, with the exception that a solution of 1x PBS pH 6.5 was injected in place of protein. This reference cell was necessarily to subtract the effect of buffer from each response signal value both during the immobilisation stage and from the binding data. The final immobilisation levels of flow cells 2-4 after subtraction are shown in Table 4.7.

**Table 4.7: CM5 chip immobilisation levels**

| Flow cell | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **Immobilised protein** | Reference cell | Hsp90β$^{(513-724)}$ | $H_6$-SUMO-PDE4A5 | $H_6$-SUMO-PDE4A5 |
| **Response signal (RUs)** | 40 | 580 | 1150 | 1050 |
| **Final density (RUs)** | 0 | 540 | 1110 | 1010 |

When the ligand is immobilised on the surface of the dextran matrix, it causes an increase in the refractive index at the surface of the chip that changes the SPR angle. This change is directly proportional to the amount of protein bound. An SPR angle change of around 0.1º corresponds to 1000 RU (Biacore AB, 2002). Since some of the amine groups of the protein are being used for binding to the negatively charged dextran, it is inevitable that a fraction of the ligand is rendered inactive during the immobilisation process, as their active and/or binding site/s would be inaccessible. Ligand immobilisation by amine coupling normally results in a surface activity of > 75% (Murphy, Jason-Moller and Bruno, 2006).

### 4.4.3. Binding assays

For the binding assays, the system was primed with 1x PBS pH 7.4 containing 0.1% of the surfactant polysorbate 20 (P20) as the running buffer. PBS was the buffer of choice as it resembles physiological conditions while the addition of the non-ionic detergent, P20 was required to avoid non-specific binding of proteins (GE Healthcare LifeSciences, 2012). A titration no-regeneration method was programmed whereby proteins were passed over the derivatised surface of the same CM5 chip in the following order; buffer (1x PBS pH 7.4), SUMO, K103R, V49M, R9Q and AIP. AIP and mutants were all $H_6$-SUMO tagged. It was therefore necessarily to use pure SUMO protein as the control to investigate whether the SUMO-tag itself was contributing to the measured binding signal. The order of injection for mutants was completely arbitrary, but once chosen it was maintained the same throughout the entire experiment. $H_6$-SUMO-AIP was deliberately injected last on the hypothesis that it would have a higher binding affinity than its mutant counterparts.

All samples were diluted in running buffer to the required concentration and injected over the derivatised flow cells (2, 3 and 4) and the underivatised surface (flow cell 1) as a background control. The flow rate was set at 40 μL.min$^{-1}$ with a 3 min injection, followed by a 5 min dissociation phase with running buffer.

In instances when the $K_D$ of the interaction is known from previous studies, it is recommended to use an analyte concentration well above the $K_D$ value to ensure that the interaction does occur and is observable (Murphy, Jason-Moller and Bruno, 2006). However, since in this case, there was no information available regarding the $K_D$ of the AIP-PDE4A5 interaction, a series of injections was performed with increasing analyte concentration, starting with 20 nM as the first injection. Subsequent injections were performed at a concentration of 40 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2.5 μM, 5 μM, 10 μM and 20 μM.

In the case of AIP, a positive SPR indicative of binding (> 20 RU) was obtained from 1 μM onwards. This was not the case for all the other mutants, especially V49M that showed a significant SPR signal only at 20 μM. Consequently, this concentration was chosen for the data fitting of all the AIP mutant proteins.

#### 4.4.3.1. Flow cell 2: Hsp90β$^{(513-724)}$ as ligand

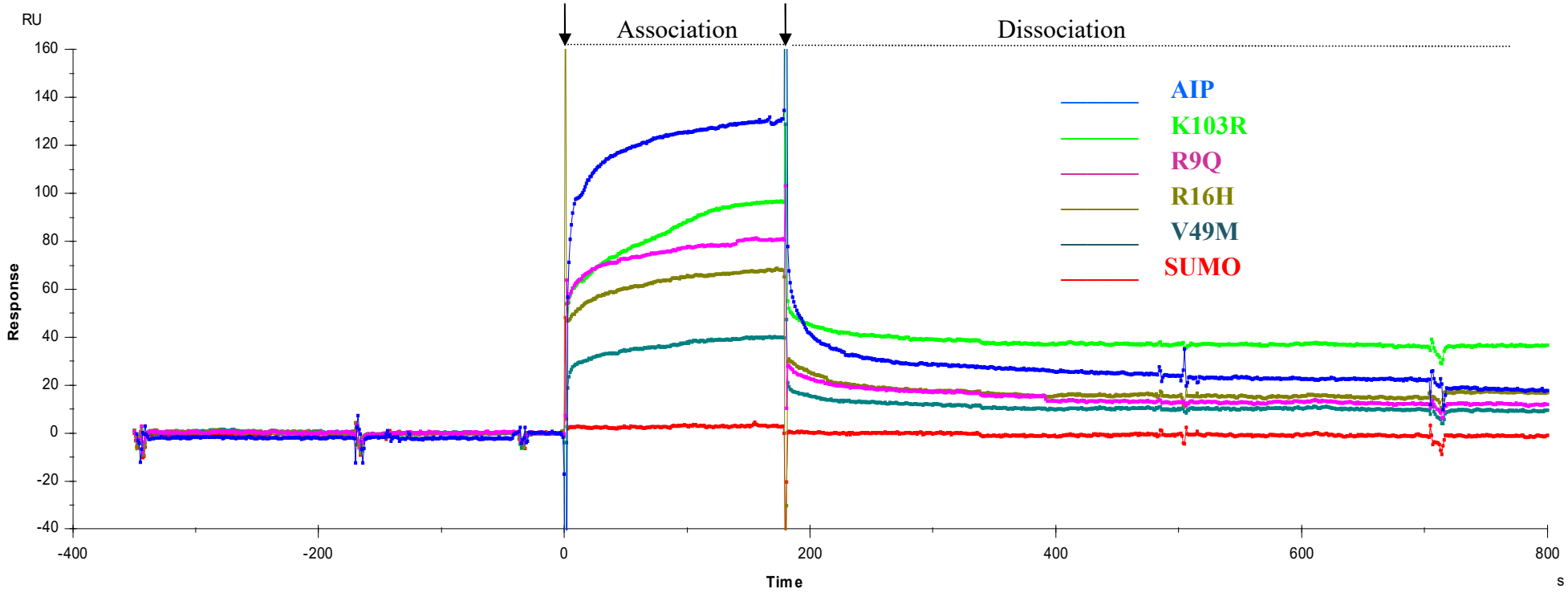



**Figure 4.15: Sensorgram showing the binding of H$_6$-SUMO-AIP and mutants to Hsp90β$^{(513-724)}$ at 20 µM.** The binding signal from flow cell 1 (blank) was subtracted from the signal of the other flow cells. For ease of comparison of the sensorgrams, the overlaid curves were x- and y-transformed. This aligned each curve at the start of injection and also adjusted them to the same baseline. Only the first minute of the dissociation phase (185-245 min) was used for curve fitting analyses. The remaining four minutes were carried out to ensure that the analyte was washed off the surface prior to the next analyte injection (chip-regeneration).

Curve fitting and kinetic data analyses were evaluated using the BIAevaluation software version 3.1. The result of the fitting gives the association and dissociation rate constants, ka and kd respectively as the fitting parameters. The equilibrium association and dissociation constants ($K_A$ and $K_D$) are also provided and are related through the following equations:

$$K_D = \frac{kd}{ka}$$
$$K_A = \frac{ka}{kd} \qquad (4.3)$$

To confirm reliability of the experimental data, the binding kinetics of $H_6$-SUMO-AIP to Hsp90β$^{(513\text{-}724)}$ were fitted using the concentrations at 5µM, 10 µM and 20 µM. The data was fitted to the 1:1 Langmuir model. The goodness of the fit was evaluated using the Chi-squared value ($\chi^2$) and the T-values for the rate constants ka and kd. T-values were calculated by dividing each parameter (ka and kd) by its respective standard error (SE) and can thus be regarded as a type of normalised inverse standard error value. T-values above 10 and low $\chi^2$ values provide confidence in the curve fit (GE Healthcare, 2019).

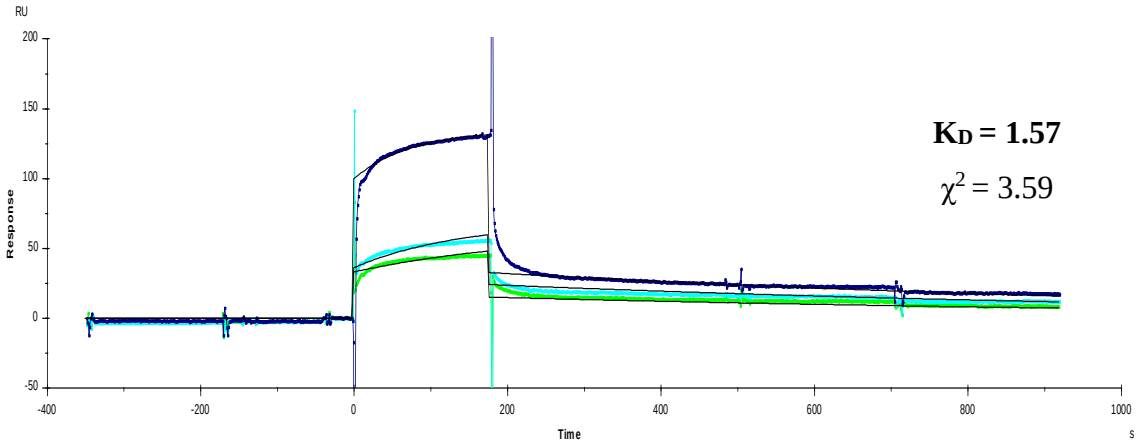



**Figure 4.16: Overlay of the sensorgrams of $H_6$-SUMO-AIP over flow-cell 2.** The curves at different concentrations (5-20µM) were fitted using the 1:1 Langmuir model. $\chi^2$ value increases with multiple curves overlay.

The overlay indicates that the SPR data fit well the 1:1 Langmuir model. However, the same overlaid fitting could not be performed for the AIP mutants as most of them did not show a significant and usable RU signal at 5 and 10 µM. As a result,

curve fitting analyses for mutant data was performed using only the concentration at 20 μM. At this concentration all samples gave an RU signal > 20 RU, thus enabling curve fitting. Moreover, to ensure fair comparison, during curve fitting, the same points for injection (0 min), association (5-170 min), end of injection (180 min) and dissociation (185-245 min) were used. The same data was also analysed using the bivalent analyte model. This model is based on the concept that the analyte has two potential binding sites so it can bind twice to the immobilised ligand. The results obtained are shown in Table 4.8.

**Table 4.8: Curve fitting: Binding constants and statistical parameters**

| Analyte | $K_A$ ($M^{-1}$) | $K_D$ (μM) | $Chi^2$ | T value (ka) | T value (kd) |
|---|---|---|---|---|---|
| | **Langmuir 1:1 Model** | | | | |
| **AIP** | $8.16 \times 10^5$ | 1.23 | 0.93 | 40 | 30 |
| **AIP (5, 10, 20 μM overlay)** | $6.38 \times 10^5$ | 1.57 | 3.59 | 70 | 66 |
| **R9Q** | $3.04 \times 10^5$ | 3.29 | 1.13 | 26 | 23 |
| **R16H** | $2.52 \times 10^5$ | 3.97 | 0.84 | 33 | 12 |
| **V49M** | $1.75 \times 10^5$ | 5.72 | 0.61 | 21 | 13 |
| **K103R** | $6.21 \times 10^5$ | 1.61 | 0.87 | 26 | 13 |
| | **Bivalent Analyte Model** | | | | |
| **AIP** | $4.01 \times 10^5$ | 2.45 | 0.94 | 28 | 21 |
| **R9Q** | $1.72 \times 10^5$ | 5.83 | 1.26 | 13 | 5 |
| **R16H** | $1.15 \times 10^5$ | 8.72 | 1.18 | 29 | 14 |
| **V49M** | $8.61 \times 10^4$ | 11.60 | 0.61 | 20 | 12 |
| **K103R** | $3.12 \times 10^5$ | 3.20 | 0.89 | 26 | 13 |

Since $H_6$-SUMO-AIP showed the largest SPR response, the data at 20 μM concentration was also analysed for mass-transfer limitations. This occurs when the association rate is faster than the diffusion rate, resulting in the analyte dissociating and re-binding during the association phase. Curve fitting however, showed that this was not the case, as the model fitted the data very data poorly with a $\chi^2$ value of 22 (Appendix A.3).

For each fitting, the raw data of ka, kd and SE values for each parameter is found in Appendix A.3. For the bivalent analyte model only $K_D1$ is shown that was obtained by dividing the values of kd1 and ka1 respectively. $K_D2$ could not be determined accurately as it was very high, suggesting that if a second binding does occur, it is very weak. The R9Q did not fit the bivalent model as well as the other proteins, especially for the dissociation phase (T value obtained is <10). The 1:1 Langmuir model fitted the data slightly better than the bivalent analyte model as suggested from the $\chi 2$ values and the higher T–values for both ka and kd.

Although the two models generated slightly different values, the relative difference between the wild type protein and each mutant is similar. In both models, the R9Q mutant exhibited an almost 2-fold decrease in binding affinity. The R16H showed a 2-3 fold decrease, whilst the V49M appears to bind with a 3.5-5 fold lower affinity. K103R showed no change in affinity with respect to wild type. These fold-differences are comparable to the differences observed through ITC experiments. A graphical comparison is represented in Figure 4.17. This indicates that despite the use of two different techniques, the relative differences in affinities are comparable in both methods, therefore adding more confidence to the results obtained.

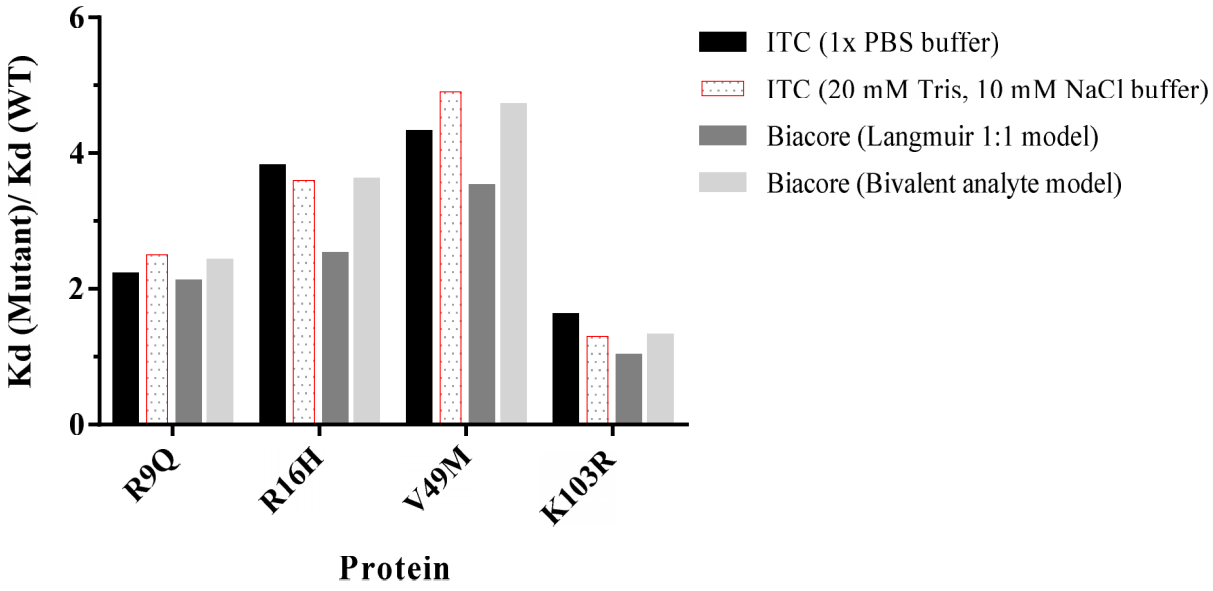



**Figure 4.17: Relative binding affinities of AIP mutants to Hsp90β$^{(513-724)}$ compared to the wild-type.**

#### 4.4.3.2. Flow cell 3: $H_6$-SUMO-PDE4A5 as ligand

For all proteins, including wild-type, no apparent binding was observed in the nM range. Wild type AIP showed SPR values higher than 20 RU at a concentration of 1 μM. For curve fitting the sensorgrams at 5, 10 and 20 μM were overlaid and analysed. This gave a $K_D$ value of 2.64 μM. For all the mutants, the $K_D$ value was obtained by fitting the sensorgram generated at 20 μM. For R9Q and V49M, this concentration was the first to produce a reasonable SPR signal that could be used for kinetic analyses. Overlays of sensorgrams at each concentration are presented in Appendix A.3.

**Table 4.9: Curve fitting: Binding constants and statistical parameters for flow cell 3 using the Langmuir 1:1 model**

| Analyte | $K_A$ ($M^{-1}$) | $K_D$ (μM) | $Chi^2$ | T value (ka) | T value (kd) |
|---|---|---|---|---|---|
| **AIP (Overlay 5, 10, 20)** | $3.79 \times 10^5$ | 2.64 | 0.90 | 31 | 17 |
| **R9Q** | $9.52 \times 10^4$ | 10.5 | 0.59 | 28 | 41 |
| **R16H** | $1.24 \times 10^5$ | 8.04 | 0.88 | 31 | 27 |
| **V49M** | $6.90 \times 10^4$ | 14.5 | 1.46 | 19 | 21 |
| **K103R** | $9.94 \times 10^4$ | 10.1 | 1.26 | 30 | 27 |

When the data at 20 μM was tested using the 1:1 model with mass transfer, $\chi^2$ values > 10 were obtained. This is an indication that the data at this concentration was not being influenced by mass transfer limitations. The low $\chi^2$ values (<2) obtained for the 1:1 Langmuir model suggest a good fit. The high T values for both ka and kd reinforce this, as high T values are proportional to low SE values for that particular parameter.

The highest binding affinity was observed for the wild type protein, as also suggested by the pronounced SPR signal when compared to the mutant counterparts. Mutants exhibited a 3-fold (R16H), 4-fold (R9Q and K103R) and 5-6 fold (V49M) decrease in binding affinity towards the immobilised ligand, $H_6$-SUMO-PDE4A5.

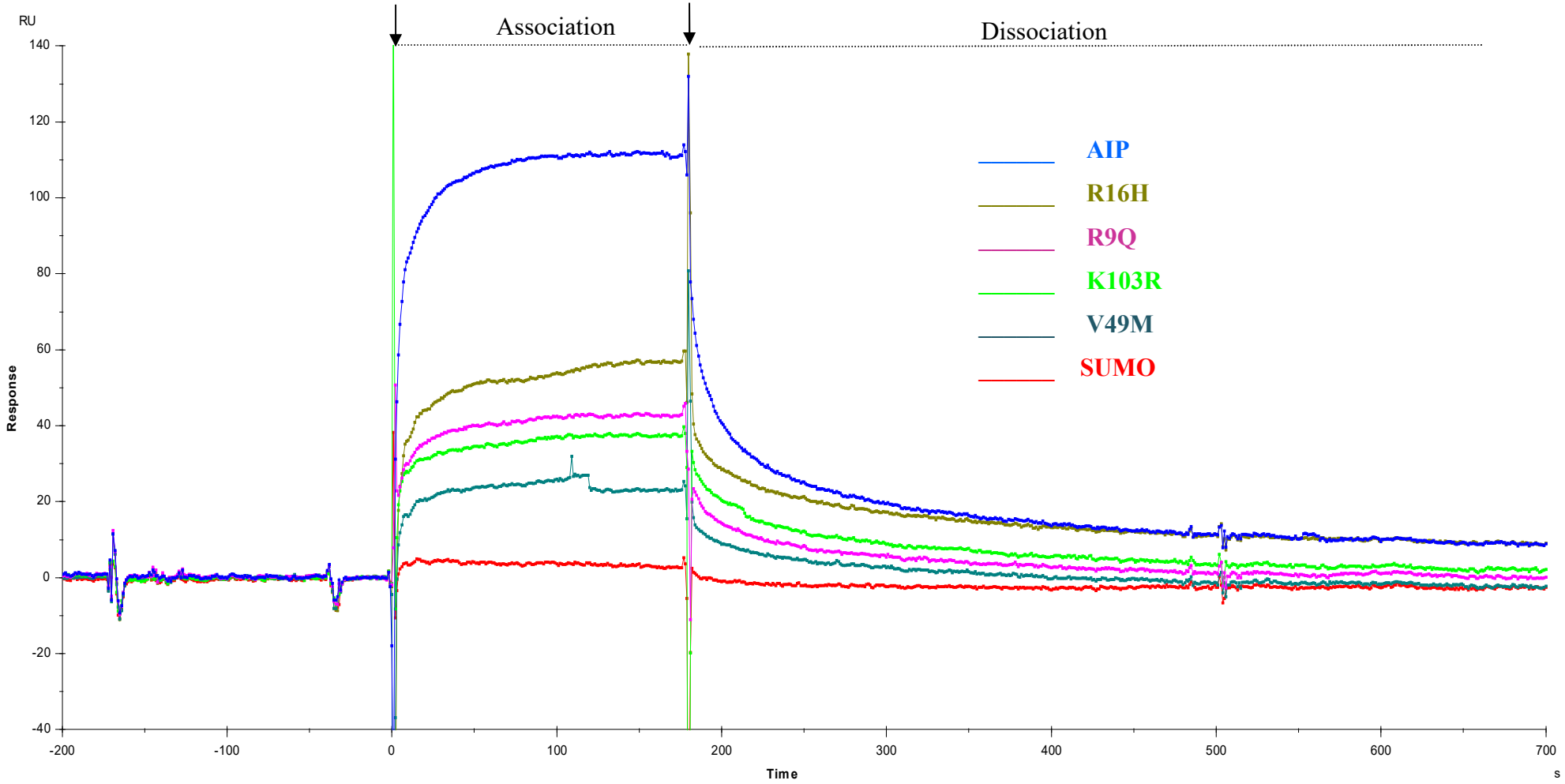



**Figure 4.18: Sensorgram showing the binding of $H_6$-SUMO-AIP and mutants to PDE4A5 at 20 µM**

A similar trend in binding affinities was obtained on flow cell 4 that also had $H_6$-SUMO-PDE4A5 as a ligand (replicate cell), except for R9Q. On flow cell 4, R9Q showed an SPR signal that was similar to that of R16H and both exhibited a 2-fold decrease in binding affinity when compared to wild type. Since AIP and the other mutants showed similar binding affinities, it is unlikely that the differential behaviour observed for R9Q is due to $H_6$-SUMO-PDE4A5 immobilised on flow-cell 4. The sensorgram and kinetic data of flow cell 4 can be found in Appendix A.3.

**Table 4.10: Summary of binding data**

| **Binding Experiment** | | **AIP-Hsp90β$^{(513-724)}$** | | **AIP-$H_6$-SUMO-PDE4A5** | |
|---|---|---|---|---|---|
| **Circular dichroism** | **Far-UV** | No changes observed | | Changes observed indicative of binding | |
| | **Near-UV** | Subtle changes observed indicative of binding | | Not analysed | |
| **SPR** | **Wild type** | Strongest binding affinity (lowest $K_D$ value) | | Strongest binding affinity (lowest $K_D$ value) | |
| | ***Mutants** | R9Q | ↓ 2x | R9Q | ↓ 4x |
| | | R16H | ↓ 2-3x | R16H | ↓ 3x |
| | | V49M | ↓ 3-4x | V49M | ↓ 5-6x |
| | | K103R | 1x | K103R | ↓ 4x |
| **ITC** | **Wild type** | Strongest binding affinity (lowest $K_D$ value) | | Not experimentally feasible due to difficulties in concentrating $H_6$-SUMO-PDE4A5 | |
| | ***Mutants** | R9Q | ↓ 2x | | |
| | | R16H | ↓ 2-4x | | |
| | | V49M | ↓ 3-5x | | |
| | | K103R | 1x | | |

*All mutants exhibited a reduction in the binding affinity towards the binding partners, Hsp90 and PDE4A5. Values given represent the fold differences in binding affinity between wild type and mutants.

## 4.5. PDE-Glo™ Phosphodiesterase Assay Protocol

Each reaction mixture was prepared as described in Section 2.3.12, Table 2.13. To test the efficiency of the PDE assay two control reactions were set up, both without the addition of the PDE enzyme.

| | | |
|---|---|---|
| Positive Control | (-) PDE, (-) cAMP | Luminescence reading: High |
| Negative Control | (-) PDE, (+) cAMP | Luminescence reading: Low |

As illustrated in Figure 4.12, the presence of high cAMP activates PKA that phosphorylates its substrate at the expense of ATP molecules. A reduction in the ATP molecules correlates with a low luminescence reading. On the contrary, the absence of cAMP signifies that a high amount of ATP is available for the luciferase reaction, resulting in a high luminescence reading. These two controls represent two extremes and thus experimental values higher than the positive control or lower than the negative control were considered as erroneous and omitted from statistical analyses. Since the luminescence values obtained from this assay are arbitrary values, it was imperative to have both control reactions in every experiment as an assessment of the assay performance and to ensure that all kit components are active. The experimental luminescence readings were corrected by subtracting from the background luminescence value.

### 4.5.1. Titration of cyclic nucleotide phosphodiesterase 4A (PDE4A)

The optimal concentration of PDE4A to use in the assay was determined experimentally. A PDE activity that is either too high or too low is unsuitable for inhibition studies, as any subtle changes in activity will not be observed. Seven solutions of PDE4A were prepared ranging from 1086 nM to 0.1086 nM respectively.

**Table 4.11:Dilutions of PDE4A**

| Sample | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Dilution factor | 0 | 0.1 | 0.05 | 0.03 | 0.01 | 0.001 | 0.0001 |
| [PDE] (nM) | 1086.96 | 108.70 | 65.22 | 21.74 | 10.87 | 1.09 | 0.11 |
| [PDE] (log) | 3.04 | 2.04 | 1.81 | 1.33 | 1.03 | 0.04 | -0.96 |

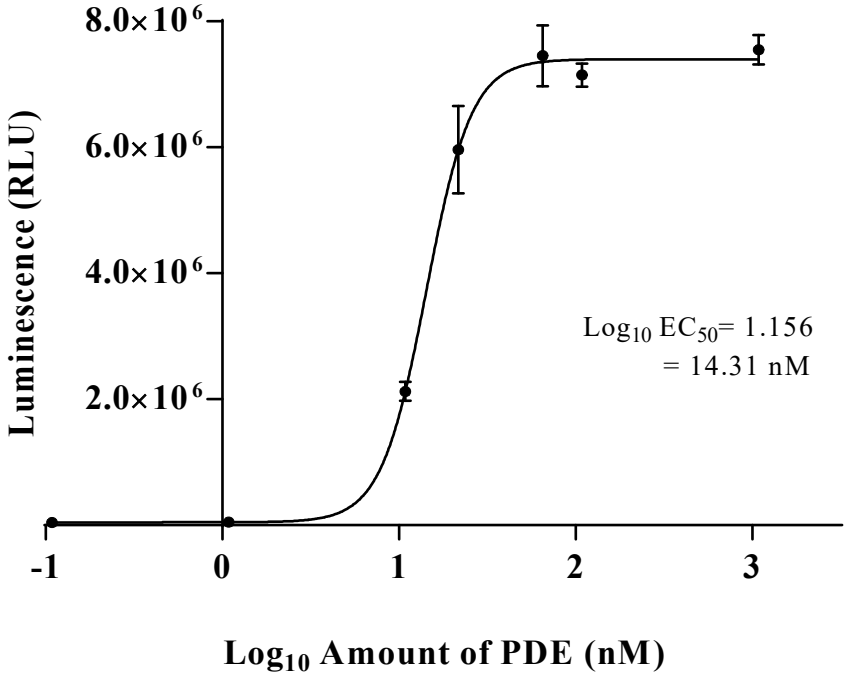



**Figure 4.19: Titration of PDE4A activity.** Data was analysed using a sigmoidal dose-response (variable slope) equation. The data represent the average of three replicates (n=3). Error bars represent the standard deviation (SD).

This result indicates that the optimal PDE concentration is around 14 nM. This value lies within the slope of the curve, making it possible to detect any changes in PDE activity when in the presence of additives.

The interaction of AIP and PDE4A5/4 has been shown to be inhibitory (Bolger *et al.*, 2003, 2016). This finding was used as the basis of the next experiment, whereby the activity of

PDE4A was assayed in the presence of AIP and AIP mutants. The scope of this experiment was to investigate whether the presence of AIP N-domain mutations affects the degree of enzyme inhibition. A constant concentration of PDE4A (14 nM) was used for all the inhibition assays.

### 4.5.2. Inhibition Assays

Prior to starting the inhibition assays, an additional control reaction was set up to investigate if the presence of AIP affects the assay performance. The reaction was prepared as described in Table 2.13 but substituting PDE4A for AIP. Both $H_6$-AIP and $H_6$-SUMO-AIP were tested.

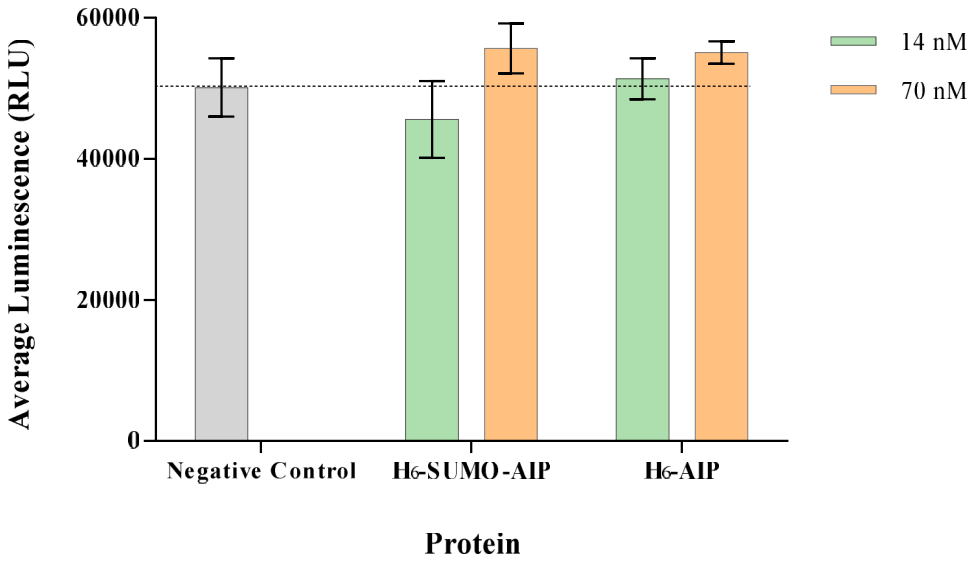



**Figure 4.20: Effect of $H_6$-AIP and $H_6$-SUMO-AIP on assay performance.** The bars represent the average of two separate experiments. In each experiment, samples were prepared in triplicates (n=3). Error bars represent the standard deviation (SD).

The luminescence values obtained in the presence of low (0.0014 μM) and high (0.007 μM) concentration of protein were compared to the negative control reaction that had the same composition as the experimental samples, but without the addition of AIP. The luminescence readings were analysed by an independent t-test that showed that the subtle differences observed are not statistically significant ($p > 0.05$) thus implying that the presence of $H_6$-AIP and $H_6$-SUMO-AIP has no effect on the kit components, at the concentrations tested.

Initially a concentration-dependent inhibition assay was performed. The reaction mixture was prepared as described in Table 2.13, with increasing concentration of $H_6$-AIP. $H_6$-AIP was added to PDE4A in a final volume of 12.5 µL and incubated for 20 min at room temperature, prior to the addition of cAMP (2 µM). The rest of the procedure was carried out as described in Section 2.3.11. Four AIP:PDE stoichiometric ratios were investigated, 5:1, 2:1, 1:1 and 0.5:1 respectively. The activity of PDE4A in the absence of $H_6$-AIP was used as control and the luminescence value obtained was considered as maximal PDE activity (100%).

**Table 4.12: AIP: PDE titration: Luminescence data and calculated activity and inhibition (%)**

| **AIP:PDE Ratio** | **0.5:1** | **1:1** | **2:1** | **5:1** |
|---|---|---|---|---|
| **Molarity AIP (nM)** | **7** | **14** | **28** | **70** |
| **PDE4A** | | **AIP + PDE** | | |
| 9479834 | 5182334 | 2981333 | 1931015 | 2113218 |
| 8482318 | 5568908 | 3676564 | 2234562 | 1872234 |
| 9642169 | 5611124 | 3091125 | 1877785 | 2298722 |
| Average **9201440** | | | | |
| **%Activity** | 56.23 | 32.40 | 20.99 | 22.96 |
| | 60.52 | 39.96 | 24.29 | 20.34 |
| | 60.98 | 33.59 | 20.41 | 24.98 |
| Average | **59.28** | **35.32** | **21.89** | **22.77** |
| Standard Deviation | 2.57 | 4.06 | 2.09 | 2.32 |
| Standard Error | 1.49 | 2.35 | 1.21 | 1.34 |
| **% Inhibition** | 43.68 | 67.59 | 79.01 | 77.03 |
| | 39.48 | 60.04 | 75.72 | 79.65 |
| | 39.02 | 66.41 | 79.59 | 75.02 |
| Average | **40.73** | **64.68** | **78.11** | **77.24** |
| Standard Deviation | 2.57 | 4.06 | 2.09 | 2.32 |
| Standard Error | 1.49 | 2.35 | 1.21 | 1.34 |

Due to the arbitrary nature of the luminescence readings, the activity of PDE4A was tested in every experiment. The average luminescence was considered as the maximal PDE activity for that particular experiment, thus making it possible to compare inhibition values obtained from different inhibition assays.

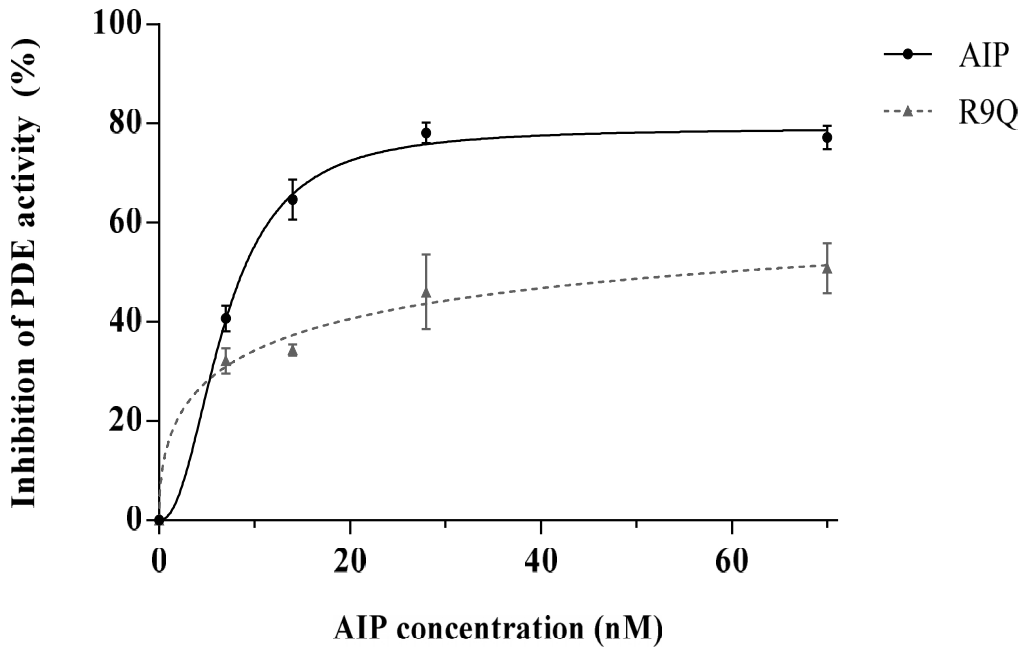



**Figure 4.21: Concentration-dependent PDE4A inhibition assay.**
Data was analysed using a non-linear regression, one-site specific binding with Hill slope equation. Each point represents the average of three replicates each (n =3). Error bars represent the standard deviation (SD).

At all concentrations, the activity of PDE4A was inhibited more in the presence of the wild type AIP than its mutant counterpart. At a protein ratio of 1:1 (AIP to PDE), the activity of PDE4A was inhibited by 64.7 ± 4.1 % by $H_6$-AIP whereas the observed inhibition with $H_6$-R9Q was almost 2-fold less, 34.6 ± 1.1 % (mean ± standard deviation, S.D.). Similarly, at an AIP: PDE ratio of 2:1, an inhibition of 78.1 ± 2.1 % was obtained with $H_6$-AIP, as opposed to 46.1 ± 13.1 % with $H_6$-R9Q. At this concentration, both proteins exerted a saturable inhibitory effect on PDE4A activity.

While an AIP: PDE ratio of 1:1 is physiologically relevant, maximal inhibition was observed at a ratio of 2:1 respectively. Based on these findings, the effect of the other AIP-mutants (R16H, V49M and K103R) on PDE4A activity was analysed separately using both ratios.

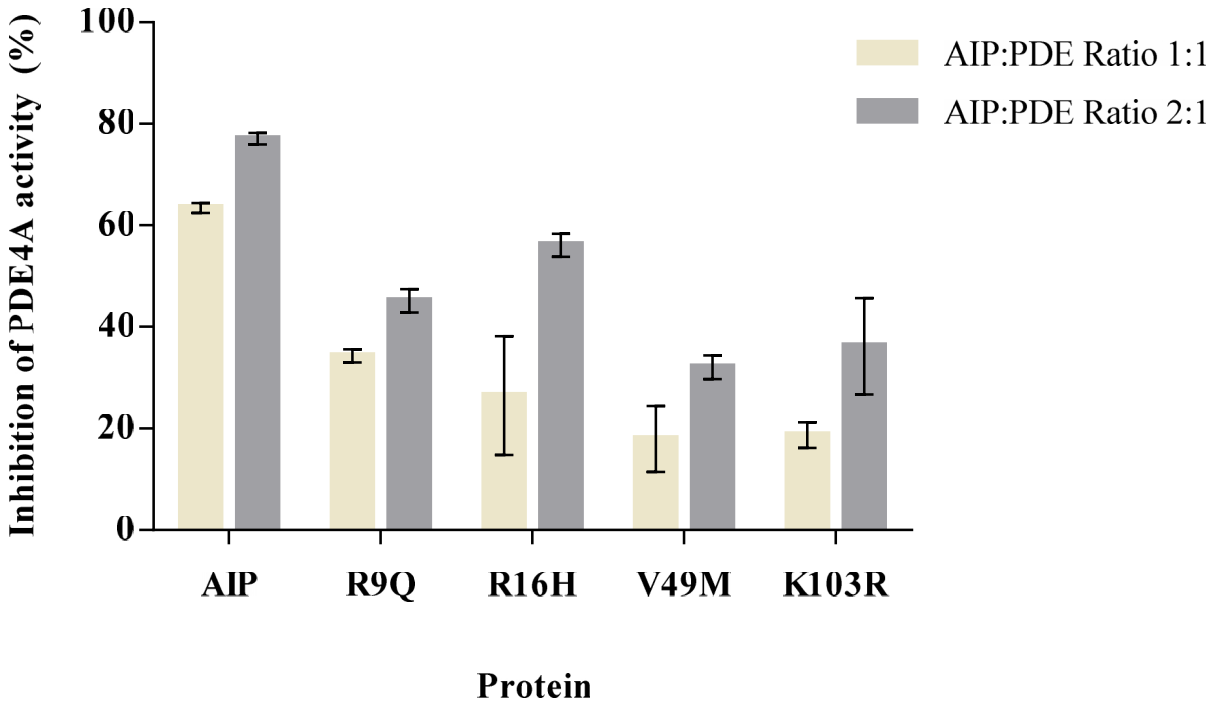



**Figure 4.22: PDE4A inhibition assay-effect of AIP mutations.**
The bars represent the average of three separate experiments carried out on separate days with different batches of the same protein type. In each experiment, samples were prepared in triplicates for reliability. Error bars represent standard deviation. AIP and AIP-mutants used in this experiment were $H_6$-tagged.

AIP exerted the highest inhibitory effect on PDE4A activity followed by R9Q and R16H.However, the inhibitory effect exerted by R16H exceeded that of R9Q by 8% when present at a higher concentration (2:1). V49M and K103R were shown to least affect PDE4A activity, with the enzyme retaining more than 80% of activity when present in a 1:1 ratio with these AIP-mutants. PDE4A % inhibition/activity values for all the AIP-mutants at both ratios are presented in Tables 4.12 and 4.13.

**Table 4.13:Inhibition of PDE4A in the presence of AIP and AIP mutants, ratio 1:1**

| | % PDE4A Inhibition | | | | |
|---|---|---|---|---|---|
| | AIP + PDE | R9Q+ PDE | R16H + PDE | V49M + PDE | K103R+PDE |
| [5]Exp 1 | 62.59 | 33.02 | 36.51 | 24.89 | 18.38 |
| Exp 2 | 64.43 | 35.61 | 29.33 | 12.02 | 21.30 |
| Exp 3 | 63.24 | 34.36 | 13.63 | 16.93 | 16.33 |
| **Average** | **63.42** | **34.33** | **26.49** | **17.95** | **18.67** |
| Std Deviation | 0.93 | 1.30 | 11.70 | 6.50 | 2.50 |
| Std Error | 0.54 | 0.75 | 6.76 | 3.75 | 1.44 |
| | **% PDE4A Activity** | | | | |
| Exp 1 | 37.41 | 66.98 | 62.17 | 75.11 | 81.62 |
| Exp 2 | 35.57 | 64.39 | 70.67 | 87.98 | 78.70 |
| Exp 3 | 36.76 | 65.64 | 86.37 | 83.07 | 83.67 |
| **Average** | **36.58** | **65.67** | **73.07** | **82.05** | **81.33** |
| Std Deviation | 0.93 | 1.30 | 12.28 | 6.50 | 2.50 |
| Std Error | 0.54 | 0.75 | 7.10 | 3.75 | 1.44 |

**Table 4.14: Inhibition of PDE4A in the presence of AIP and AIP mutants, ratio 2:1**

| | % PDE4A Inhibition | | | | |
|---|---|---|---|---|---|
| | AIP + PDE | R9Q+ PDE | R16H + PDE | V49M + PDE | K103R +PDE |
| Exp 1 | 77.23 | 46.08 | 55.08 | 30.00 | 32.32 |
| Exp2 | 78.11 | 46.83 | 58.74 | 31.56 | 29.26 |
| Exp 3 | 75.88 | 42.54 | 54.44 | 34.61 | 47.05 |
| Average | **77.07** | **45.15** | **56.09** | **32.06** | **36.21** |
| Std Deviation | 1.12 | 2.29 | 2.32 | 2.34 | 9.51 |
| Std Error | 0.65 | 1.32 | 1.34 | 1.35 | 5.50 |
| | **% PDE4A Activity** | | | | |
| Exp 1 | 22.77 | 53.92 | 44.92 | 69.99 | 67.68 |
| Exp2 | 21.89 | 53.17 | 41.26 | 68.44 | 70.74 |
| Exp 3 | 24.12 | 57.46 | 45.56 | 65.39 | 52.95 |
| Average | **22.93** | **54.85** | **43.91** | **67.94** | **63.79** |
| Std Deviation | 1.12 | 2.29 | 2.32 | 2.34 | 9.51 |
| Std Error | 0.65 | 1.32 | 1.34 | 1.35 | 5.50 |

[5] The value of each experiment represents the average of three replicates (n=3)

**Table 4.15: Tests of Normality using data from protein: PDE ratio of 1:1**

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| AIP | .190 | 9 | .200* | .911 | 9 | **.322** |
| R9Q | .217 | 9 | .200* | .938 | 9 | **.566** |
| R16H | .187 | 9 | .200* | .917 | 9 | **.365** |
| V49M | .182 | 9 | .200* | .917 | 9 | **.368** |
| K103R | .191 | 9 | .200* | .936 | 9 | **.538** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Table 4.16: Tests of Normality using data from protein: PDE ratio of 2:1**

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| AIP | .231 | 8 | .200* | .861 | 8 | **.122** |
| R9Q | .142 | 8 | .200* | .971 | 8 | **.902** |
| R16H | .238 | 8 | .200* | .882 | 8 | **.196** |
| V49M | .124 | 8 | .200* | .976 | 8 | **.941** |
| K103R | .174 | 8 | .200* | .922 | 8 | **.447** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The results indicate that the data was normally distributed (all p-values > 0.05) and was therefore analysed using a two-tailed independent t-test and one-way ANOVA. A confidence interval of 95% was set for both parametric tests. The difference observed in inhibition percentile, at both ratios, was found to be statistically significant ($p < 0.0001$) for all the mutants. Descriptive statistics and the output of both parametric tests can be found in Appendix A.4.

## 4.6. Structural analysis of AIP

Throughout the course of this research project, various commercially available crystallisation screens, with and without additives (Section 2.3.14) were used for crystallisation trials of purified full-length AIP protein. Crystallisation trays (288 wells) were set up using the sitting drop technique at the Astbury Centre for Structural Molecular Biology, at the University of Leeds and were stored at 20°C in an imaging and detection machine, SONICC (Formulatrix) for a period of 12-16 months. This machine makes use of ultraviolet two photon excited fluorescence (UV-TPEF) and second harmonic generation (SHG) to identify and distinguish chiral protein crystals from salt crystals. Similarly, the plates that were prepared manually using the hanging drop technique were stored at 18°C in temperature-controlled rooms (University of Malta and University of Leeds). All trays were routinely checked for crystal growth. However, despite testing over 5500 different conditions, both with $H_6$-AIP and $H_6$-SUMO AIP, none of these conditions yielded crystals. It was therefore necessary to employ other techniques to obtain structural information on the native full-length protein structure of AIP.

### 4.6.1. Protein structure prediction through I-TASSER

I-TASSER is an algorithm that predicts protein structure and function using various computational analyses (Zhang, 2008; Yang, *et al*., 2015). The protein sequence of $H_6$-AIP was uploaded to the I-TASSER website and compared to a set of reference proteins whose structures have been determined and deposited in the PDB repository.

I-TASSER predicted five models for AIP. The one shown in Figure 4.24, has the highest confidence value and is therefore the most likely conformation adopted by AIP from the ones predicted.

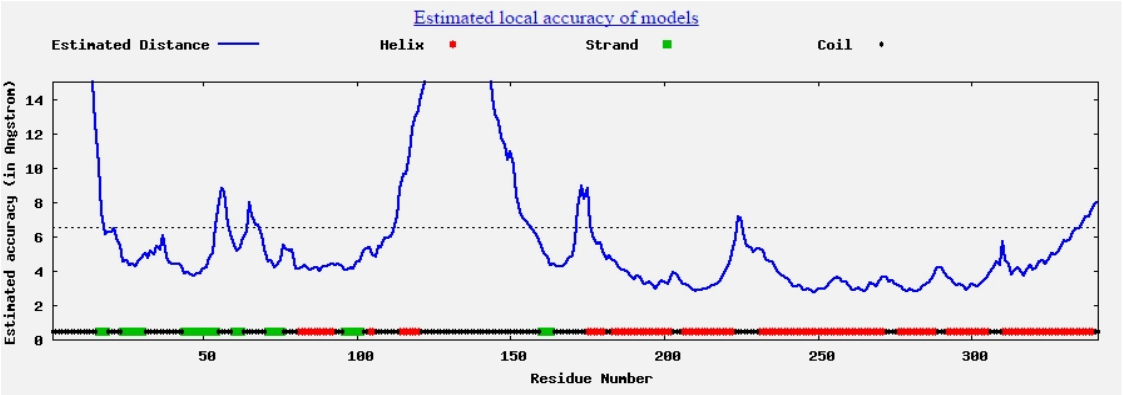



**Figure 4.23: Estimated local accuracy of the model structure of AIP produced using I-TASSER** (Zhang, 2008; Yang, *et al*., 2015). The estimated accuracy of every amino acid residue is graphically depicted in blue. As expected, unstructured coil regions are characterised with higher estimated distances which correlates with lower structural accuracy.

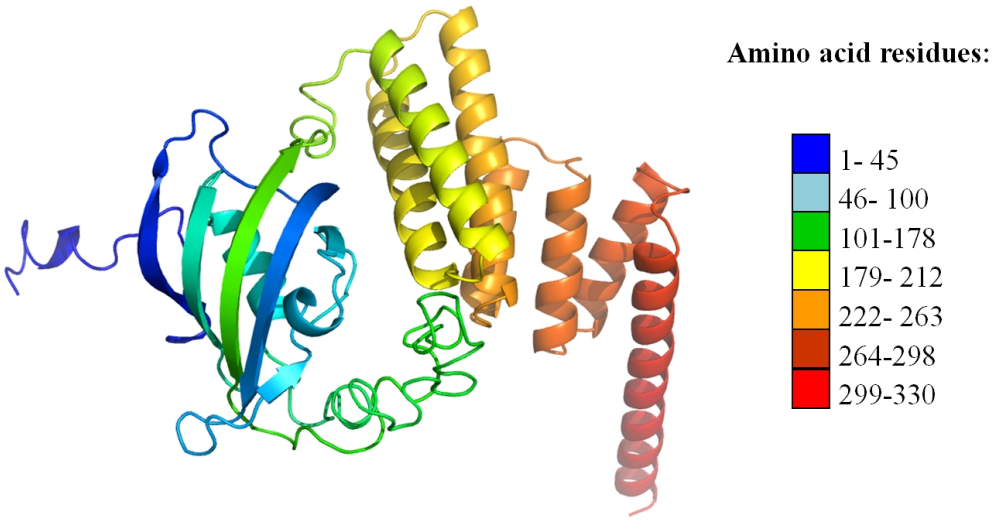



**Figure 4.24: Cartoon representation of the predicted model of $H_6$-AIP by ITASSER.**

Structural prediction of $H_6$-AIP through I-TASSER showed that the N-domain contains regions of disordered structure, particularly from residues Gly111-His135 that represent the unstructured loop region of the *βD-βE* extension. This is a unique structural feature of AIP. The linker joining the N to the C-domain is also a flexible disordered region. The C-domain is mainly composed of alpha helices. This correlates well with the X-ray crystallography structures (PDB ID: 4AIF and 4APO) that demonstrate that the C-domain of AIP is composed of three TPR motifs.

## 4.7. Small Angle X-Ray Scattering (SAXS)

SAXS is a relatively recent technology, useful for characterising the structure of biological macromolecules in solution. It is mainly employed for challenging biomolecules whose structure cannot be obtained by other conventional techniques, such as X-ray crystallography or NMR, usually due to having flexible domains and/or disordered inter-domain linkers. Proteins that are intrinsically disordered are also good candidates for this technique (Feigin and Svergun, 1987, Svergun *et al*., 2013; Tria *et al*., 2015).

Aside from web-based predictions, the presence of naturally disordered regions within AIP was also confirmed through analyses of the far-UV spectrum by CDNN that showed $H_6$-AIP to contain 33% of random coil (Table 4.1). The occurrence of these regions probably accounts for the fact that the crystallisation of the native protein was unsuccessful and that the N-domain has only been structurally characterised through NMR. Moreover, the NMR structure of the N-domain (PDB ID: 2LKN) is represented by twenty states which indicate the presence of a highly mobile loop region that is part of the $\beta$D-$\beta$E extension. SAXS was thus applied to obtain structural information on full-length AIP at Diamond Light Source, B21 beamline.

### 4.7.1. Structural analyses using ScÅtter

Samples of $H_6$-SUMO-AIP were prepared at final concentrations of 4.5 $mg.mL^{-1}$ and 7 $mg.mL^{-1}$ in 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM DTT and 5% (v/v) glycerol. Both samples were initially tested on the BioSAXS Robot using the experimental parameters described in Section 2.3.13. The inclusion of glycerol and DTT in the sample buffer was essential both to minimise protein aggregation at high concentration as well as protect the sample against radiation damage (BIOSAXS, GmbH, 2016). Owing to the natural tendency of the protein to form multimeric states, this preliminary test was conducted to establish the optimum concentration of $H_6$-SUMO-AIP that generates a good scattering profile and is not compromised by the presence of protein aggregates. Data was analysed using the ScÅtter Version 3.0 (Rambo, 2019a).

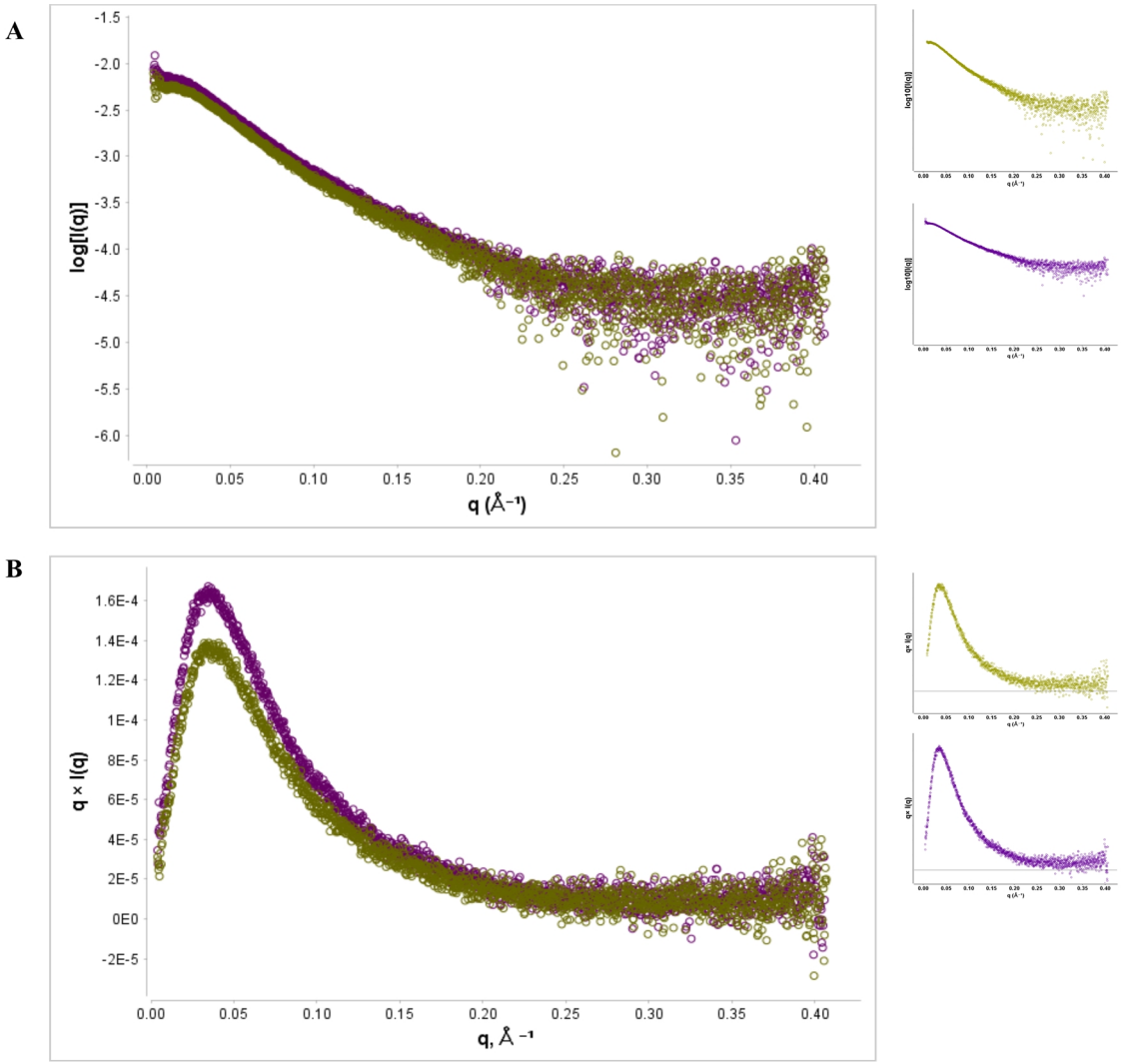



**Figure 4.25:A: Log$_{10}$ SAXS intensity versus scattering angle, *q*. B: Total scattered intensity plot.** In both figures, the scattering profiles of $H_6$-SUMO-AIP at 4.5 mg.mL$^{-1}$ (olive green) and at 7 mg.mL$^{-1}$ (purple) are superimposed. The individual profiles are shown on the right.

Superposition of log$_{10}$ SAXS scattering curves from samples of different concentrations can be used to show concentration-dependent scattering. If the sample contains aggregates, there will be a plateau followed by an obvious decline in intensity at the very small q values (far left of the plot). This would be indicative of inter-particle repulsion (Putnam *et al.*, 2007). In the case of $H_6$-SUMO-AIP, neither of the two scattering curves showed such an apparent decline in intensity. Further data analyses using the Guinier fitting were thus performed to assess the homogeneity of the sample.

#### 4.7.1.1. The Guinier Analysis - the Guinier fitting and Guinier peak analysis

The Guinier analysis investigates the SAXS scattering curve at the lowest scattering angles. It enables the determination of the radius of gyration ($R_g$) that is a measure of overall size of the protein. Therefore it is also useful in the detection of protein multimerisation and/ or aggregation (Rambo, 2019b).

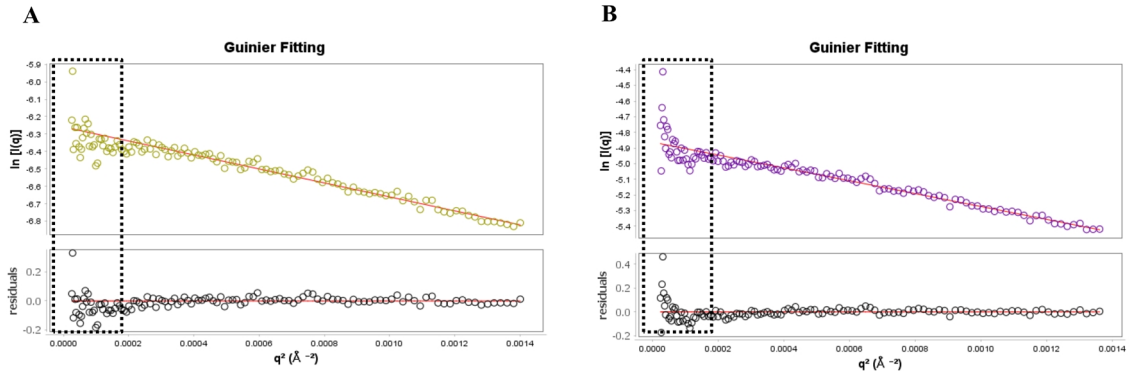



**Figure 4.26: Guinier analysis fitting plot.**

**A**: $H_6$-SUMO-AIP at 4.5 mg $mL^{-1}$, **B**: $H_6$-SUMO-AIP at 7 mg $mL^{-1}$. The upper plot shows the SAXS data and fit plotted against ln [I (q)] as a function of $q^2$. The lower plot represents the corresponding residuals and is an indication of the goodness of fit. I = SAXS intensity; q = scattering angle.

A non-linear relationship between the dependent ($q^2$) and independent variable [ln I (q)] is a direct indication of multimerisation in the sample (Putnam *et al*., 2007). Figure 4.26 shows that the sample at 7 $mg.mL^{-1}$ deviates from linearity within the small q range (indicated by a dotted black box) implying that multimierisation might have started to occur at that concentration. This was further reinforced through the Guinier-Peak analysis plot (Figure 4.27). This plot is the normalised Guinier and is more sensitive to sample heterogenity (Putnam *et al*., 2007).

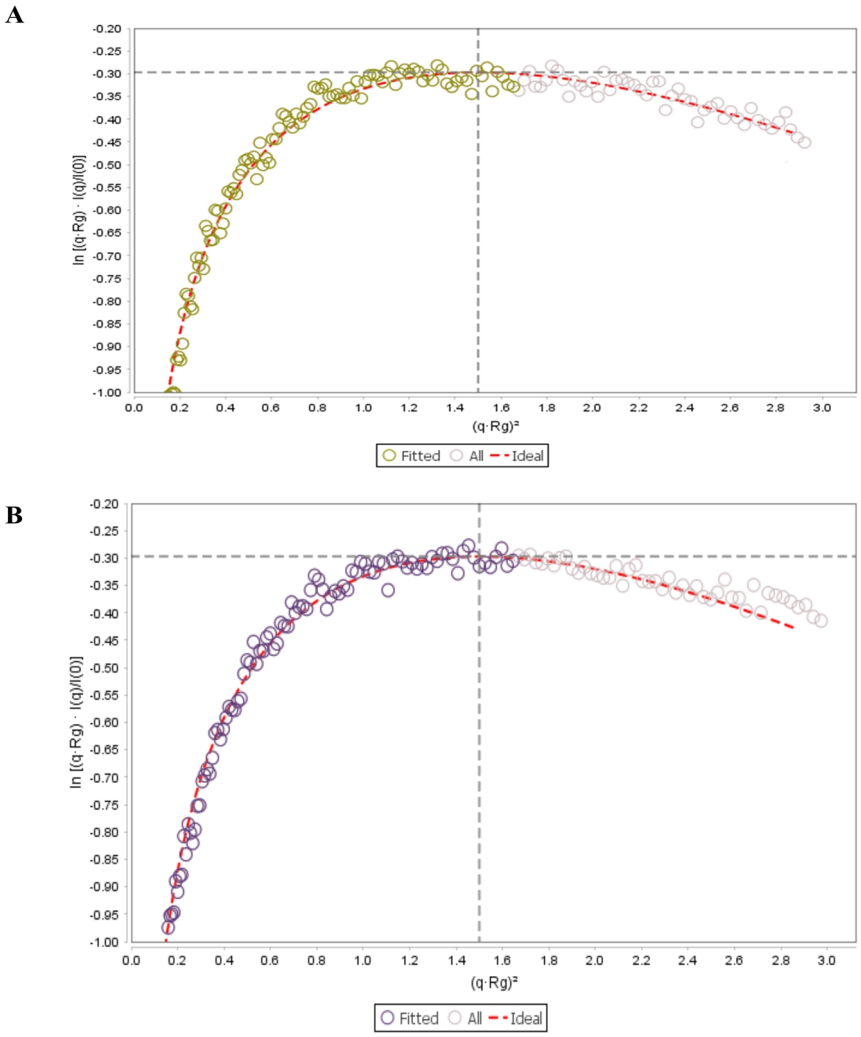



**Figure 4.27: Guinier Peak Analysis Plot: A:** $H_6$-SUMO-AIP at 4.5 mg $mL^{-1}$; **B:** $H_6$-SUMO-AIP at 7 mg $mL^{-1}$.

Ideal globular proteins normally have a peak at $(q \times Rg)^2$ of around 1.5 on a normalized Guinier plot. Experimental data obtained from a well-behaved sample would follow an ideal model that is indicated by a dashed red line (Rambo and Tainer, 2011; Putnam, 2016).

A comparison of the two Guinier peak analysis plots shows that while the sample at 4.5 $mg.mL^{-1}$ fits exactly the ideal model, the sample at the higher concentration deviates from ideal behaviour at the upper q-regions. Within this region the experimental data is skewed away from the dashed red-line. Given the sensitivity of SAXS, further experiments with $H_6$-SUMO-AIP and $H_6$-AIP were carried out at a protein concentration of 4.5–5 $mg.mL^{-1}$ and the use of SEC-SAXS was employed

(Section 2.3.13). The coupling of SEC to SAXS ensures that the sample is pure and mono-disperse an essential prerequisite for determination of protein structure by SAXS.

Following SEC-SAXS, samples of $H_6$-AIP were also analysed by Guinier analyses. The experimental data fitted the ideal model both in the Guinier fitting and in the normalized Guinier plot (Figure 4.28 B). This confirmed the homogeneity of the sample.

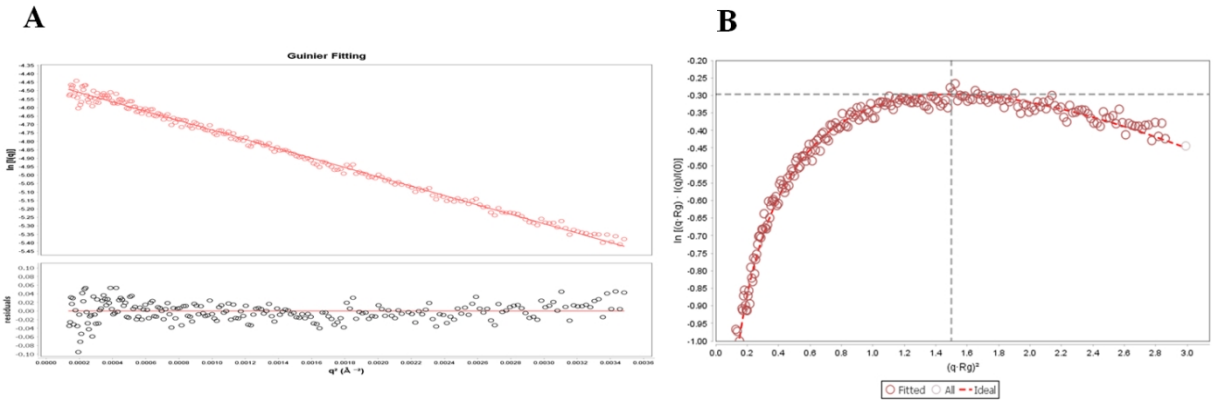



**Figure 4.28: Guinier analysis of $H_6$-AIP at 4.5 mg.mL$^{-1}$.**
**A:** Guinier fitting, the upper plot shows the experimental SAXS data plotted against ln [I (q)] as a function of $q^2$. The lower plot represents the corresponding residuals. **B:** Guinier peak analysis (that is equivalent to the normalised Guinier plot).

Through the same analyses, an Rg value of 30.0 Å was obtained for $H_6$-AIP, as opposed to 35.9 Å for $H_6$-SUMO-AIP at the same concentration, thus reflecting the larger size of the latter. This is expected due to the additional SUMO-tag which is 107 amino acids in length with an approximate molecular weight of 12 kDa. Furthermore, the estimated Dmax (maximum diameter of the protein) value from the experimental data of $H_6$-AIP was 110.5 Å that is 21.5 Å smaller in size than the Dmax value $H_6$-SUMO-AIP (132.0 Å). The experimental SAXS data was further analysed for molecular shape and flexibility through the Kratky plot and Porod volume respectively (Putnam, 2007).

### 4.7.1.2. The Kratky plot

The Kratky plot is a qualitative assessment of structural disorder versus compactness of biological macromolecules. This representation provides a clear idea of the degree of flexibility, or 'un-foldness' within the scattering protein in solution. This differentiation is possible because different biological macromolecules have characteristic representations. Globular ideal proteins such as bovine serum albumin (BSA) exhibit a typical bell-shaped curve. On the other hand, extended, flexible or unfolded peptides lack this distinctive peak and usually have a plateau at high q-values that is indicative of flexibility and/or disorder. An intrinsically disordered protein would show a characteristic "hyperbolic plateau" (Rambo and Tainer, 2011).

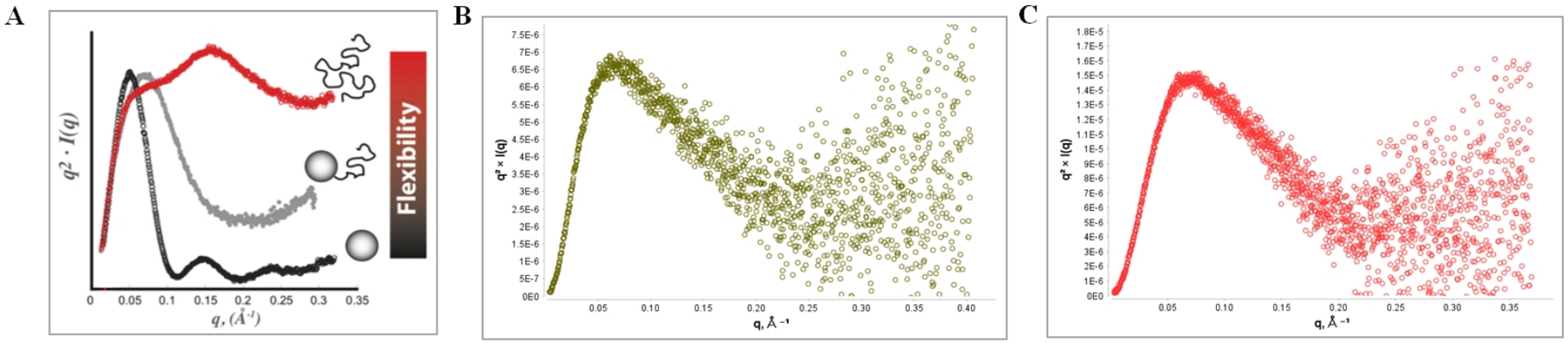



**Figure 4.29: Kratky plot. A:** The scattering data shows the difference in the behavior of the curve for a globular folded protein (black), partially folded protein due to flexible regions (grey) and completely unfolded or intrinsically disordered proteins (red) (Rambo and Tainer, 2011). **B:** Kratky plot based on the experimental scattering data of $H_6$-SUMO-AIP (olive green) and C: $H_6$-AIP (red).

By comparing the experimental Kratky plots to the standard plot it is evident that the scattering curve of both AIP samples lies in between the flexibility range displayed in panel A. The protein is not entirely compact and folded because it is characterised with regions of disorder and flexibility. This is in agreement with structural predictions (Figure 4.24).

#### 4.7.1.3. Flexibility and P(r) distribution analyses

Whilst the Kratky plot, is a qualitative measure of the compactness or 'random-coil' nature of a protein, this can be further supported by quantitative metrics obtained through flexibility analyses. This analysis is based on the Porod-Debye fourth law power law that describes a relationship between the scattering angle, q and the observed intensities, I(q). This law is an approximation that holds within a limited range of q-values, known as the Porod-Debye region that describes the decay of the scattering intensity (Rambo and Tainer, 2011).

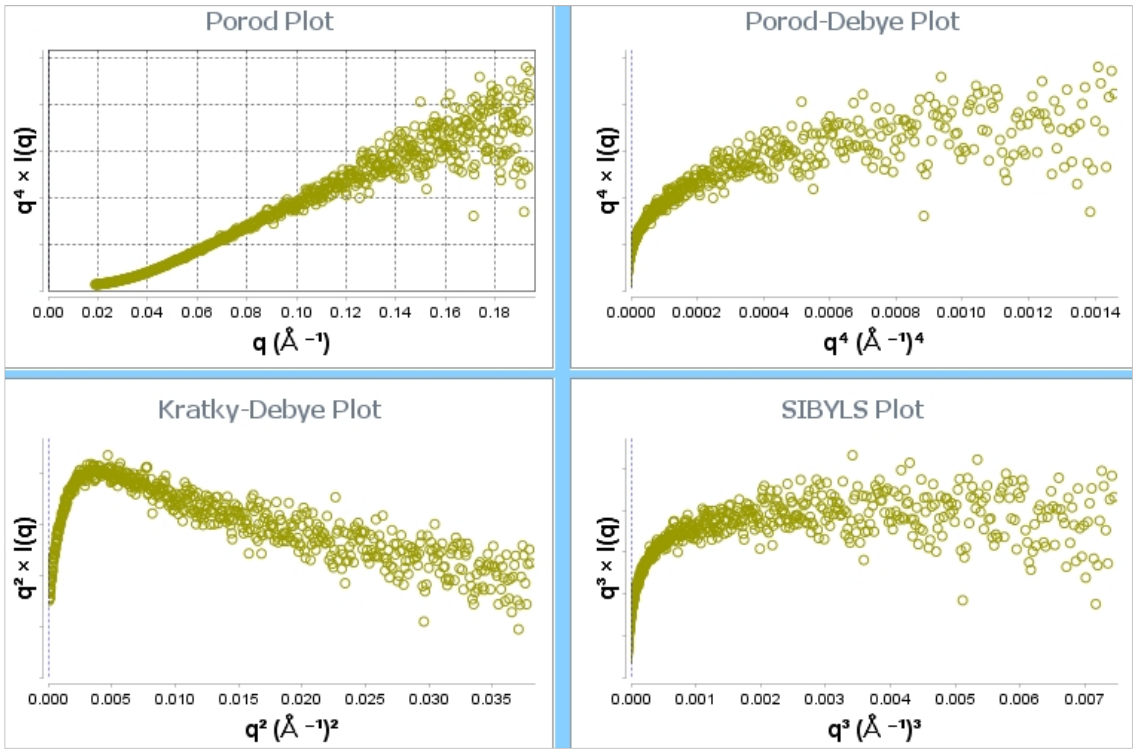



**Figure 4.30: Flexibility analyses of $H_6$-SUMO-AIP.** The flexibility analyses function displays four panels, each with a different plot depending on relationship between the scattering angle, q (raised to the appropriate power) against the scattering intensity I (q) at the same angle.

During this analysis the value of q-max is decreased until the data is observed to plateau after the "hyperbolic rise". As a general rule, if the plateau is first seen in Porod-Debye plot, the protein is globular and compact. On the other hand, a flexible, extended and/or disordered protein would exhibit the first plateau in the Kratky-Debye plot (Rambo, 2019b).

In this case, the Kratky-Debye plot is excluded from further consideration as it shows a negative slope. This implies that the protein is neither unfolded nor entirely flexible. The SIBYLS plot was the first to plateau at q-max of 0.19579 (read from the Porod Plot), therefore suggesting that $H_6$-SUMO-AIP is globular with regions that exhibit flexibility. This is in agreement with the Kratky plot (Figure 4.29). $H_6$-AIP showed the same trend.

Following the flexibility test, the volume of the protein (Porod volume) and its Dmax were calculated, for both AIP samples. Dmax was obtained from the pair-distance, P(r) distribution function tool. It is equivalent to the position where the P(r) function returns back to zero, at large *r* values. The output of the SAXS parameters is summarised in Table 4.16.

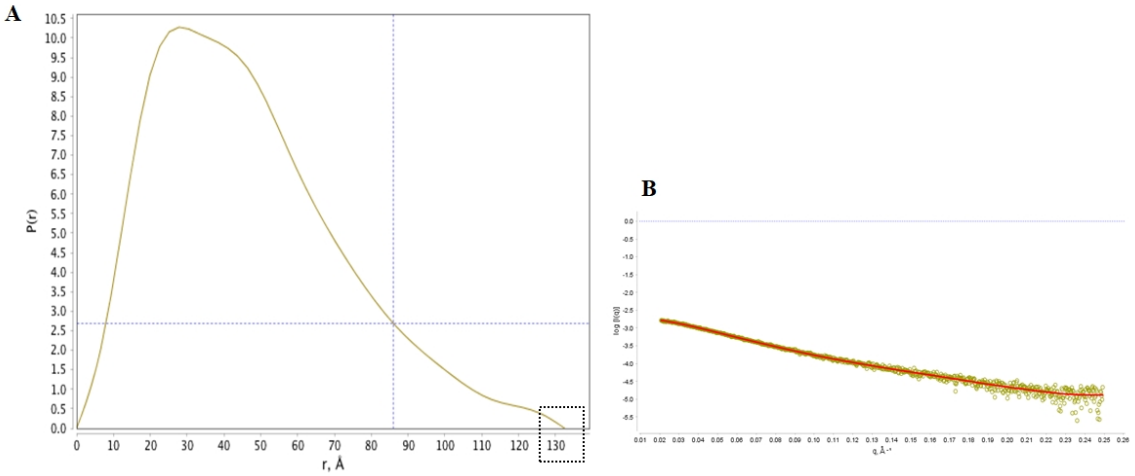



**Figure 4.31: A: Pair-distance, P(r), distribution function of $H_6$-SUMO-AIP.** Maximum diameter, Dmax, is the largest non-negative value that shows a smooth distribution function, indicated by a dotted black box. **B: Representation of the goodness of the fit.** Experimental scattering data is shown in dark yellow circles. Red line indicates the expected behaviour (Rambo, 2019b).

The skewness of the curve also indicates the presence of flexible regions and suggests an extended conformation. Globular compact proteins have a characteristic bell-shaped curve (Putnam *et al*., 2007; Rambo, 2019b).

**Table 4.17: Comparison of the SAXS parameters for $H_6$-AIP and $H_6$-SUMO-AIP.**

| | **Protein** | |
|---|---|---|
| | **$H_6$-AIP** | **$H_6$-SUMO-AIP** |
| q-min ($Å^{-1}$) | 0.0182 | 0.0208 |
| q-max ($Å^{-1}$) | 0.1998 | 0.1957 |
| Radius of gyration, Rg (Å) | 30.0 ± 0.44 | 35.9 ± 0.83 |
| Maximum diameter, Dmax (Å) | 110.5 | 132.0 |
| [6]Porod particle volume, $V_p$ ($A^3$) | 68,582 | 88,793 |
| Calculated molecular weight (Da) | 41,314 | 53,490 |
| Expected molecular weight (Da) based on protein sequence | 39,041 | 49, 787 |

The molecular weight of the protein was calculated using the equation developed by Rambo and Tainer (2011). In this study most of the investigated protein samples had a protein density within 0.9 and 1.0 gm·$cm^{-3}$, most likely due to the presence of flexible regions and/ or his-tagged extensions within them (Rambo and Tainer, 2011).

$$d = 1.66\, n\, \frac{M}{V}$$

Where:

d= particle (protein) density in $g.mL^{-1}$

n = number of subunits in a multimeric complex

$V_p$ = Porod volume in $Å^3$

$M_r$ = molecular weight in Da

In the case of AIP, the protein density was assumed to be 1 $g.mL^{-1}$ and n was taken as 1, since the AIP samples used were monomeric and not part of a complex. The molecular weight was obtained by dividing the Porod volume by 1.66.

[6] 1 $nm^3$ = 1000 $Å^3$

The calculated molecular weights of $H_6$-AIP and $H_6$-SUMO-AIP protein were marginally higher than the expected value. The observed discrepancy is most likely due to the fact that the value of $V_p$ is based on the user's input to determine the linear region in the Porod-Debye Plot that is limited by q-max (previously obtained from the flexibility analyses). This might introduce an element of subjectivity. Nonetheless the difference is not high enough to suggest the presence of multimeric states, therefore confirming that both proteins were present as a monomer in solution.

### 4.7.2. Molecular modelling using SAXS data

There are two main methods of how the SAXS scattering profile can be utilised for protein modelling. In the absence of structural information of the full-length protein or any of its domains, only the *ab initio* method can be used for protein structure prediction. This will generate a low resolution bead model that will give an idea of the overall shape that the protein under investigation can adopt. However, when the structure of subunits or domains is known, rigid body modelling can be employed which results in an atomic model of higher resolution (Petoukhov and Svergun, 2006).

In this study, both techniques were employed. Initially, an *ab initio* model of both $H_6$-SUMO-AIP and $H_6$-AIP was generated using DAMMIN/F. This was followed by rigid body modelling using various docking methods, performed by Dr Robert Rambo (Section 2.3.13). The generated models were analysed further using PyMOL version 2.3.0 (Schrödinger, 2015).

### 4.7.3. DAMMIN/F- ab initio modelling

DAMMIN/F is a modelling algorithm that uses the scattering SAXS data to generate an *ab initio* low resolution model that accommodates the scattering volume of the protein. A number of independent DAMMIF runs were carried out and then averaged to generate a bead density model (Rambo, 2019b). The output file shown here is the damaver.pdb file. This file represents the probability map of seventeen different models that were aligned together for the best overlap and used to generate a molecular envelope. The data of $H_6$-SUMO-AIP was also analysed using MONSA (Svergun, 1999). This additional step was performed to visualise the arrangement of the protein with respect to the $H_6$-SUMO tag. A total of eleven runs were performed with MONSA to generate a bead model.

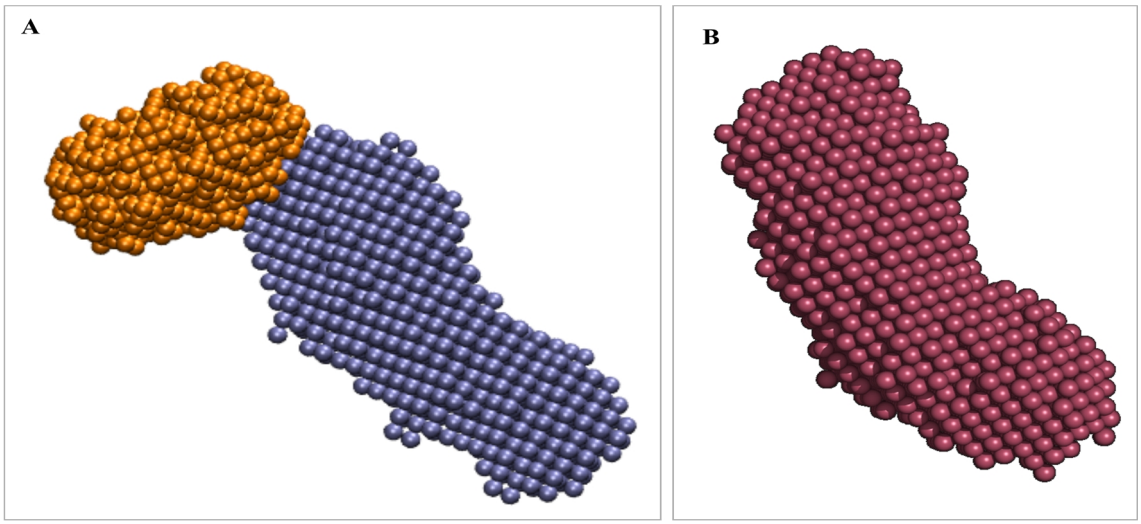



**Figure 4.32: Low resolution bead models.**
**A:** $H_6$-SUMO-AIP-Molecular envelope after refinement by MONSA. The low resolution model initially obtained by DAMMIN/F was further refined using MONSA. The position of the $H_6$-SUMO tag relative to the rest of the protein is highlighted in orange. **B:** $H_6$-AIP-Molecular envelope of the DAMAVER output.

The damaver bead model is the average of seventeen other models aligned together for the generation of a probability map (the molecular envelope). As such, the dimensions of the damaver bead model are always larger than the dimensions of subsequent refined models. This accounts for larger size observed for $H_6$-AIP bead model when compared to the refined bead model of $H_6$-SUMO-AIP.

### 4.7.4. Rigid body modelling of $H_6$-SUMO-AIP

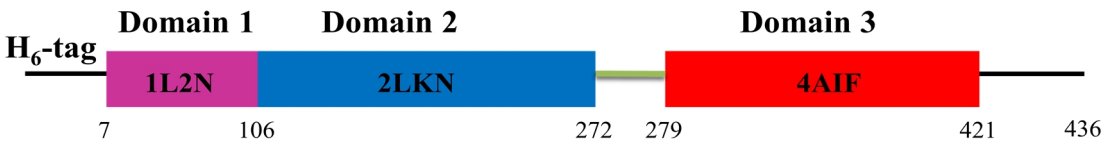



**Figure 4.33: Schematic representation of $H_6$-SUMO-AIP.**
Domains labelled as 1, 2 and 3 represent the SUMO, N-domain and C-domain of AIP respectively. The N- to C-domain linker is shown in green. Amino acid numbering of each domain is shown beneath the schematic.

Modelling of $H_6$-SUMO-AIP required the use of three separate PDB files that contain structural information on the SUMO, N-domain and C-domain of AIP respectively. SWISS-Model (Waterhouse, 2018) was used to generate three individual PDB files, in which the each residue of the PDB was correctly numbered based on the sequence of $H_6$-SUMO-AIP that was inputted. The co-ordinates of each file were copied into a single PDB file to create a single polypeptide chain. The single domain structures were assembled into a multi-domain protein by FoXS (Schneidman-Duhovny *et al*., 2013).

After building the initial model with all the domains linked on a single polypeptide chain, the dimensions obtained exceeded the value of Dmax (132 Å) of the protein. This value was previously determined through analyses of the SAXS scattering profile (Table 4.16). This suggested that two ends of the protein were too far apart and did not fit the scattering data obtained.

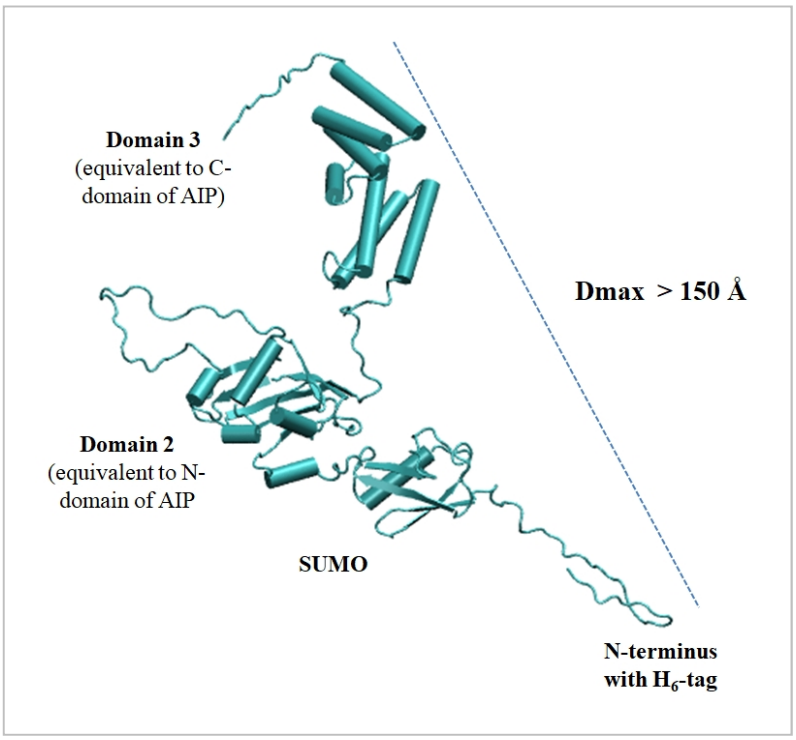



**Figure 4.34: Initial built model of $H_6$-SUMO-AIP.**

This model was considered as inaccurate as the Dmax value was >150 Å, considerably larger than the Dmax value calculated from solution scattering (132 Å). This proved that further modelling was required. Secondary structure constrains were applied between domains 2 and 3 and a limit was placed on how far the two ends can be with respect to each other.

Since the 3 PDB files (12LN, 2LKN and 4AIF) used all had missing structural information of the flexible regions, the sequence was evaluated through SPIDER2 (Khashan, Zheng and Tropsha, 2012; Fleishmann *et al*., 2011). This was necessarily to obtain a prediction of the secondary structure of the whole protein. Moreover, it also provides the surface area accessibility of each amino acids as well as the main–chain torsional angles (Yang et al., 2017). This made it possible to estimate the distance restraints for the flexible regions.

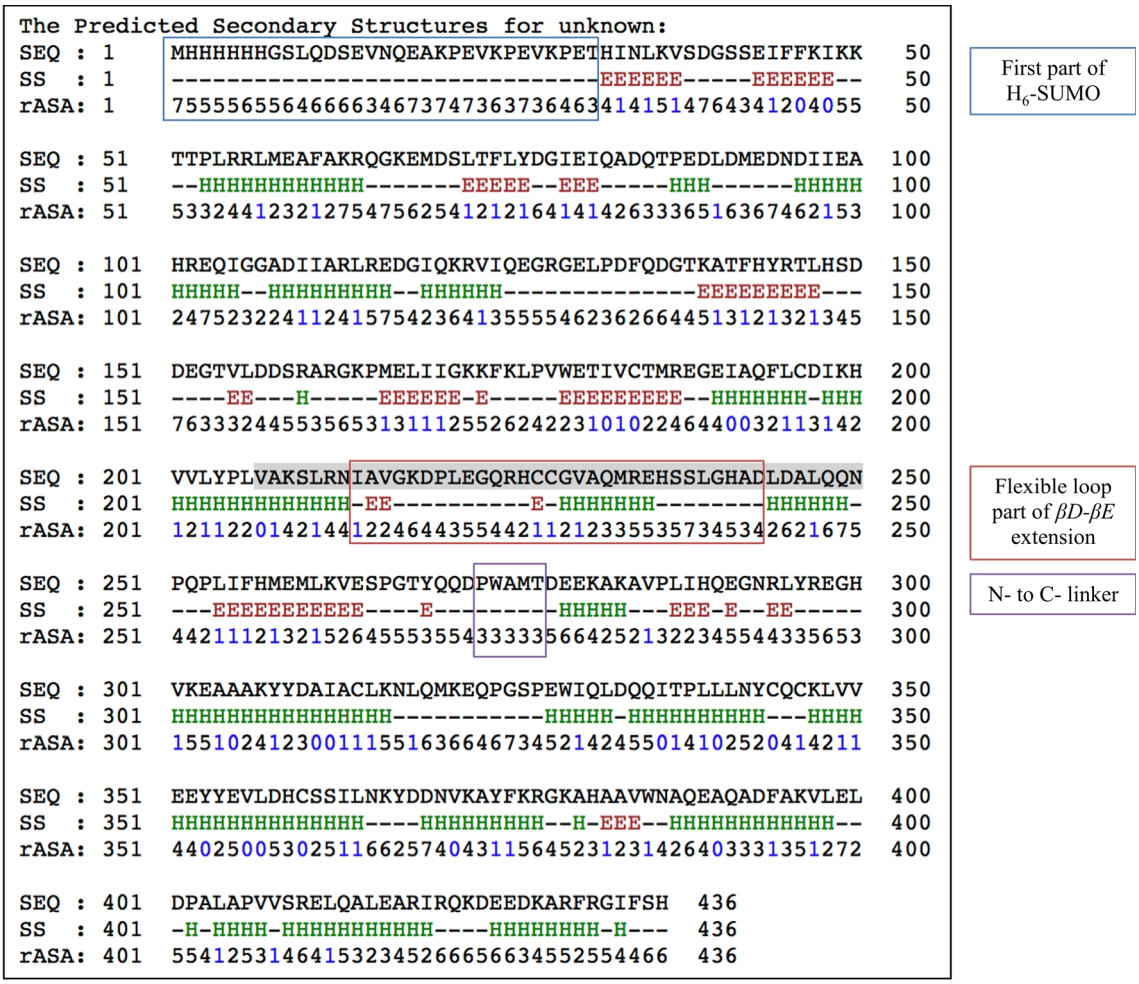



**Figure 4.35: SPIDER2 protein structure prediction.** The sequence is depicted at the top. Secondary structure (SS) is denoted by H (helix), E (beta-sheet) or – (random coil). The relative accessible surface area (rASA) is presented by a number, 0-9 with 9 being the most accessible. Buried residues with rASA <20% are labeled in blue. Selected regions that were predicted to have no secondary structure are shown in coloured boxes with a matching description on the right hand side of the figure.

The results obtained indicate the first part of N-domain does not have a defined secondary structure. This represents the first region of the $H_6$-SUMO tag. Moreover, the N-domain of AIP (domain 2 in this model) is characterised with other flexible regions, the most notable of which is the loop that is part of the *βD-βE* extension. The residues in these mobile regions were constrained based on the Dmax value (132 Å).

FoXS was once again used for the structural characterisation of the flexible regions and the generation of a complete protein model, using simulated annealing runs. For each model FoXS computed a theoretical scattering profile and fitted it to the experimental one. The output resulted in a total of twenty models. The best model was chosen on the basis of fit of the scattering curves and the $Chi^2$ value.

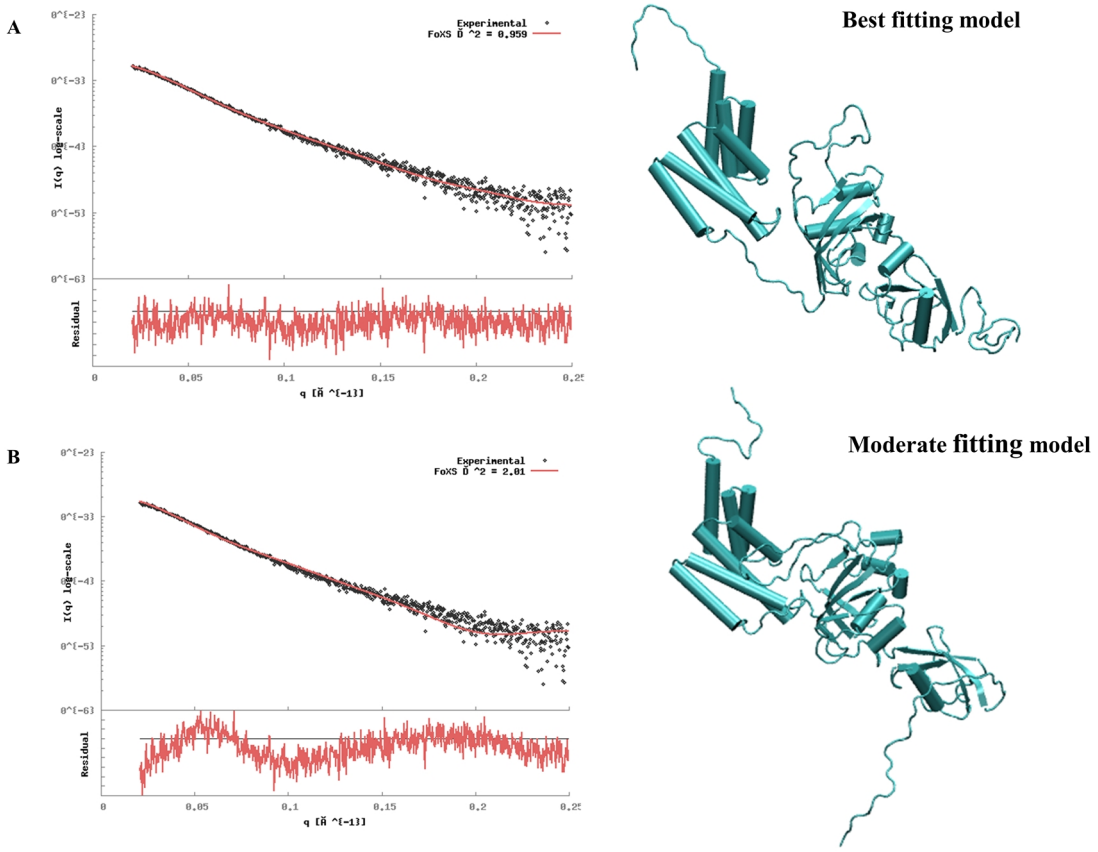



**Figure 4.36: Atomistic models of $H_6$-SUMO-AIP generated by FoXS.**
**A:** Best fitting model with a $Chi^2$ value of 0.959. **B:** Moderate fitting model with a $Chi^2$ value of 2.01. On the left hand side of each figure, there is an overlay of the theoretical (red line) and experimental scattering profile (black dots), with the residuals plot underneath.

As can be observed in Figure 4.38, panel A, the theoretical scattering profile of the model fits well the experimental data as reinforced by the $Chi^2$ value and the residuals plot. The most notable difference between the two models is the position of the flexible loop in domain 2 (which represents the N-domain of the AIP) and the C-terminal tail. A closer analysis of the best fitting model is shown in Figure 4.37.

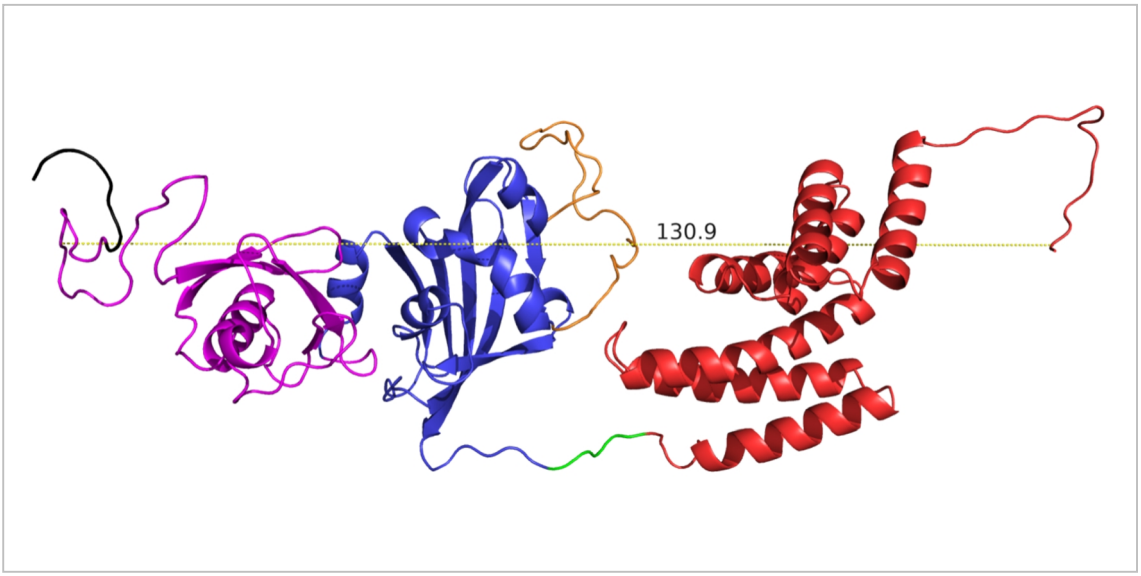



**Figure 4.37: Cartoon representation of the best fitting model of $H_6$-SUMO-AIP.** The $H_6$ sequence is shown in black. The three domains coloured in magenta, blue and red, represent the SUMO, N-domain and C-domain of AIP respectively. The flexible loop in the N-domain of AIP is shown in orange. The N- to C-domain linker between is coloured in green. This figure was generated using PyMOL version 2.3.0 (Schrödinger, 2015).

The result showed that the protein is relatively linear. Other generated models in which the domains were more compact, fitted the SAXS data poorly, thereby adding confidence on the linear orientation of the N-domain and C-domain of AIP (domains 2 and 3 in this model), with respect to each other. Interestingly, the different orientations of the flexible loop in domain 2 (shown in orange, Figure 4.37) have large effects on the fit. This adds further credibility to the relative orientation of this loop as identified in the best fitting model for $H_6$-SUMO-AIP. The measured Dmax of the model is 130.9 Å which is very close to the Dmax value (132 Å) as estimated through the P(r) distribution function on ScÅtter (Table 4.16).

Despite the good fitting observed, the size of the $H_6$-SUMO domain and the presence of a long disordered region at the N-terminus did present a challenge when modelling. It may also have had a direct effect on the precise orientation of the N- and C-domain of AIP with respect to each other. As a result, it was considered appropriate to also perform the same modelling analyses on the $H_6$-AIP protein.

### 4.7.5. Rigid body modelling of $H_6$-AIP

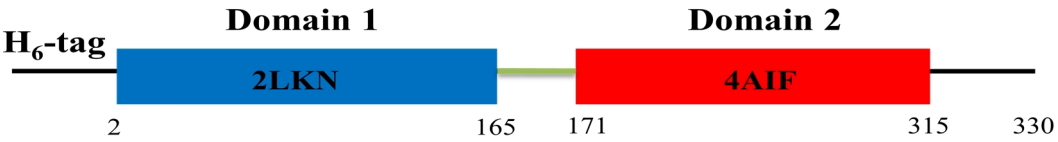



**Figure 4.38: Schematic representation of $H_6$-AIP.** Domains labelled as 1 and 2 represent N-domain and C-domain of AIP respectively. The N- to C-domain linker is shown in green. Amino acid numbering of each domain is shown beneath the schematic. For simplicity purposes, numbering is based on the non $H_6$-tagged AIP sequence.

The model of $H_6$-AIP was carried out as explained for the $H_6$-SUMO tagged protein using FoXS. Secondary structure information from homology domains was used to constrain the regions that lacked stable secondary structure.

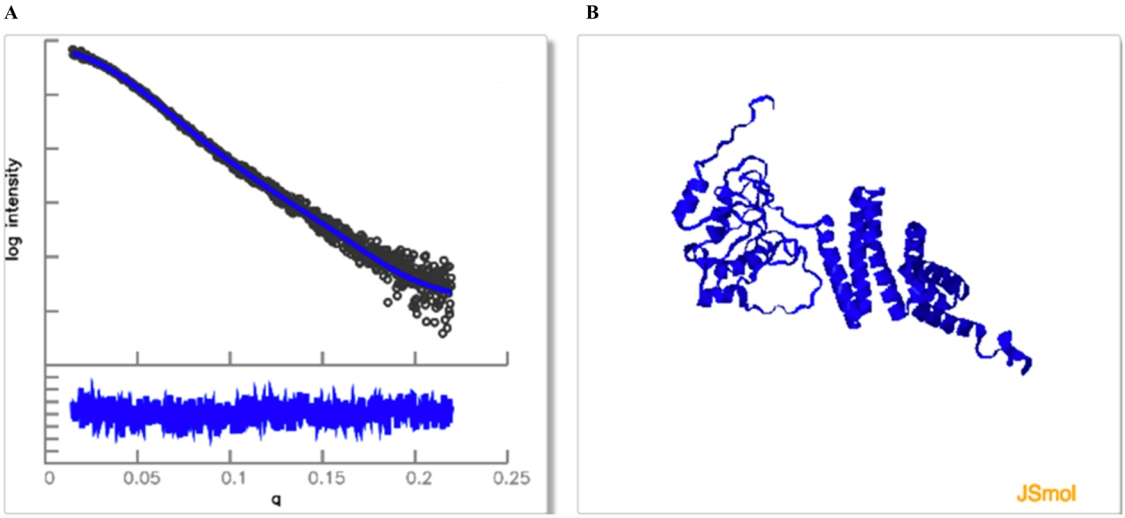



**Figure 4.39:Atomistic model of $H_6$-AIP generated by FoXS. A**: Profile fit together with the residuals (experimental vs. theoretical). The input experimental data is displayed with black dots and the theoretical data is shown as a blue line. **B:** The model displayed in a JSmol window.

The best fitting atomistic model generated by FoXS was further refined using CNS. A total of 50 simulated annealing runs were performed. Two top models were identified that had a $Chi^2$ value <2.

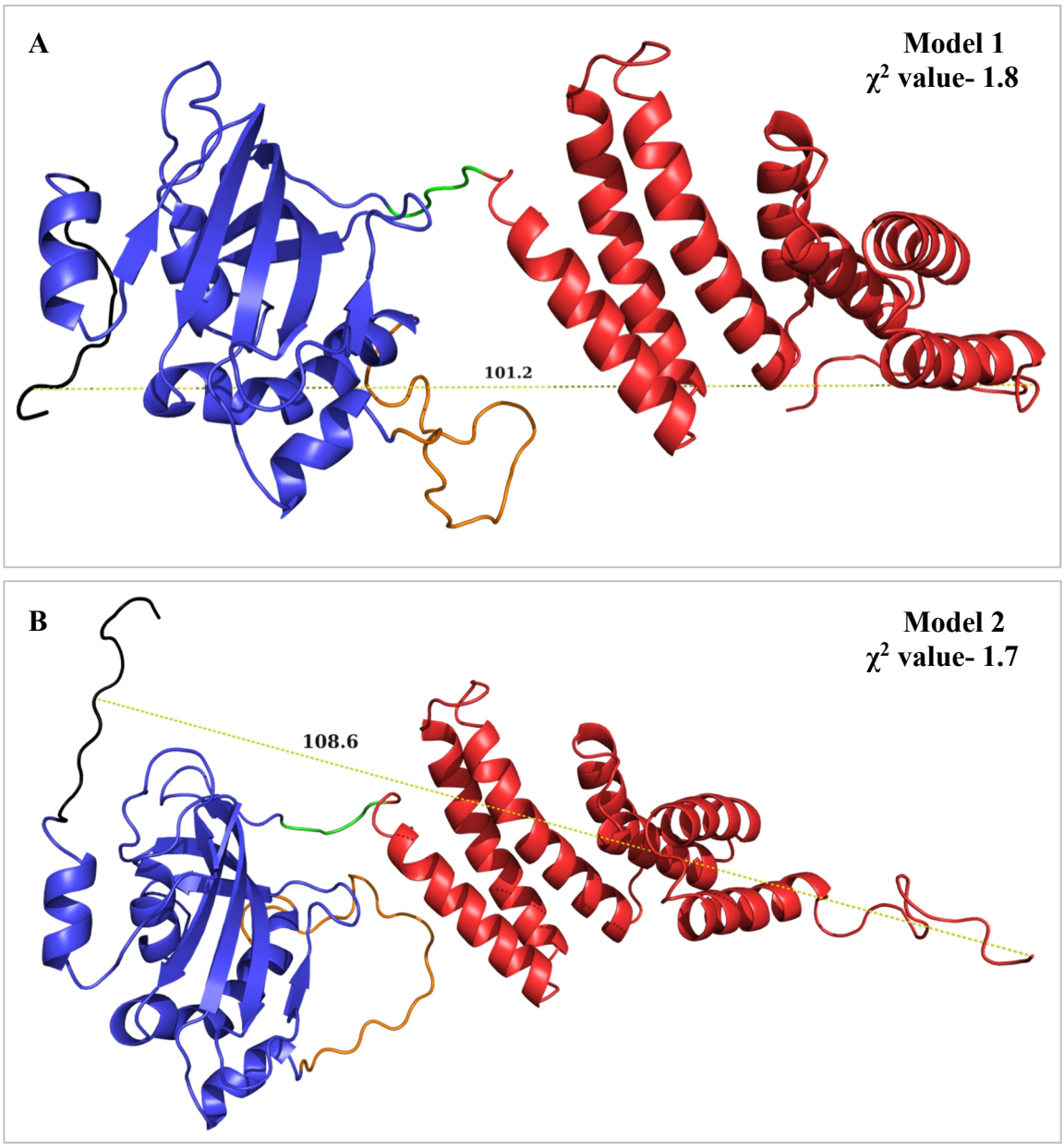



**Figure 4.40: Cartoon representation of the two best fitting models of $H_6$-AIP.** $H_6$ sequence is shown in black. The N-domain and C-domain of AIP are shown in blue and red respectively, connected by a short linker (green). The flexible loop in the N-domain of AIP is shown in orange. This figure was generated using PyMOL version 2.3.0 (Schrödinger, 2015).

The Dmax of both models, particularly of model 2 is comparable to the Dmax value (110.5 Å) determined from ScÅtter (Table 4.16) thus reinforcing the validity of the models. Both models show a linear arrangement of the two domains. This result complements the modelling of the $H_6$-SUMO-AIP and shows that the SUMO-tag did not affect the relative position of the N- and C-domain of AIP with respect to each other.

### 4.7.6. Structural analyses of each domain of $H_6$-AIP

The N-domain consists of five antiparrallel *β*-strands (*β*A-*β*E) that form half a *β*-barrel enveloping a central helix designated as αII. In the SAXS generated model (Figure 4.41) only part of the *β*A is shown as a beta-sheet. The *β*C strand is interrupted by a loop that contains αI. The PPIase-like domain of AIP (residues G31-C121) is shown in red and does not include the structural elements *β*A, *β*E, αIV and α0, the latter being uniquely found in AIP. The insert that connects *β*D and *β*E, known as the *β*D-*β*E extension is another unique feature of AIP that is 37 amino acid residues longer than found in typical FKBPs (Linnert *et al*., 2013). This insert is composed of helices αIII and αIV that are connected by a long loop region (G111-H135) that is mostly unstructured. The SAXS generated models showed conformational variability in this region. This is in agreement with the published NMR data where a large number of residues within this region, were non-assigned, an indication of flexibility (Linnert *et al*., 2013). The C-domain of AIP consists of three pairs of TPR motifs followed by a terminal Cα-7 helix.

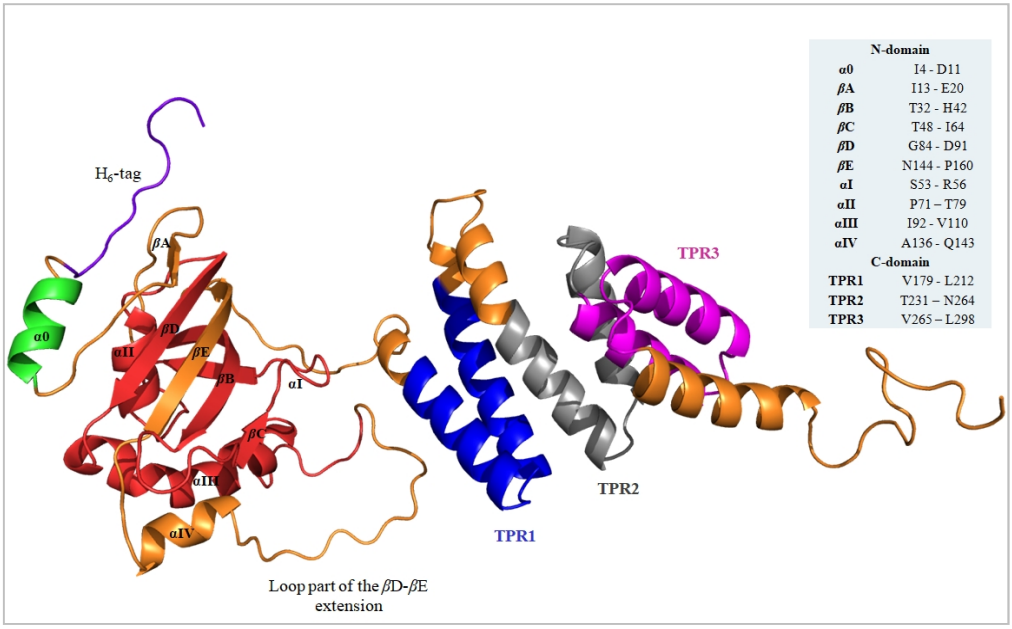



**Figure 4.41: Cartoon representation of $H_6$-AIP Model 2 with domain annotations.** The N-domain PPIase-like region is coloured in red, while each TPR motif is coloured in blue, dark grey and magenta respectively, followed by a C-terminal α-7 helix. For ease of comparison, numbering of each structural element and motif is based on the untagged AIP sequence. This figure was generated using PyMOL (Schrödinger, 2015).

Each TPR motif consists of two antiparrallel helices, with each helix being 17 amino acids in length. A number of residues were classified as being highly conserved due to their direct involvement in mediating protein-protein interactions, including E216, P232, L242 (TPR2), K266, R271 and D287 (TPR3) (Morgan *et al*., 2012).

### 4.7.7. Structural comparison of models

In order to investigate how each model compares with the other, they were aligned with respect to each domain.

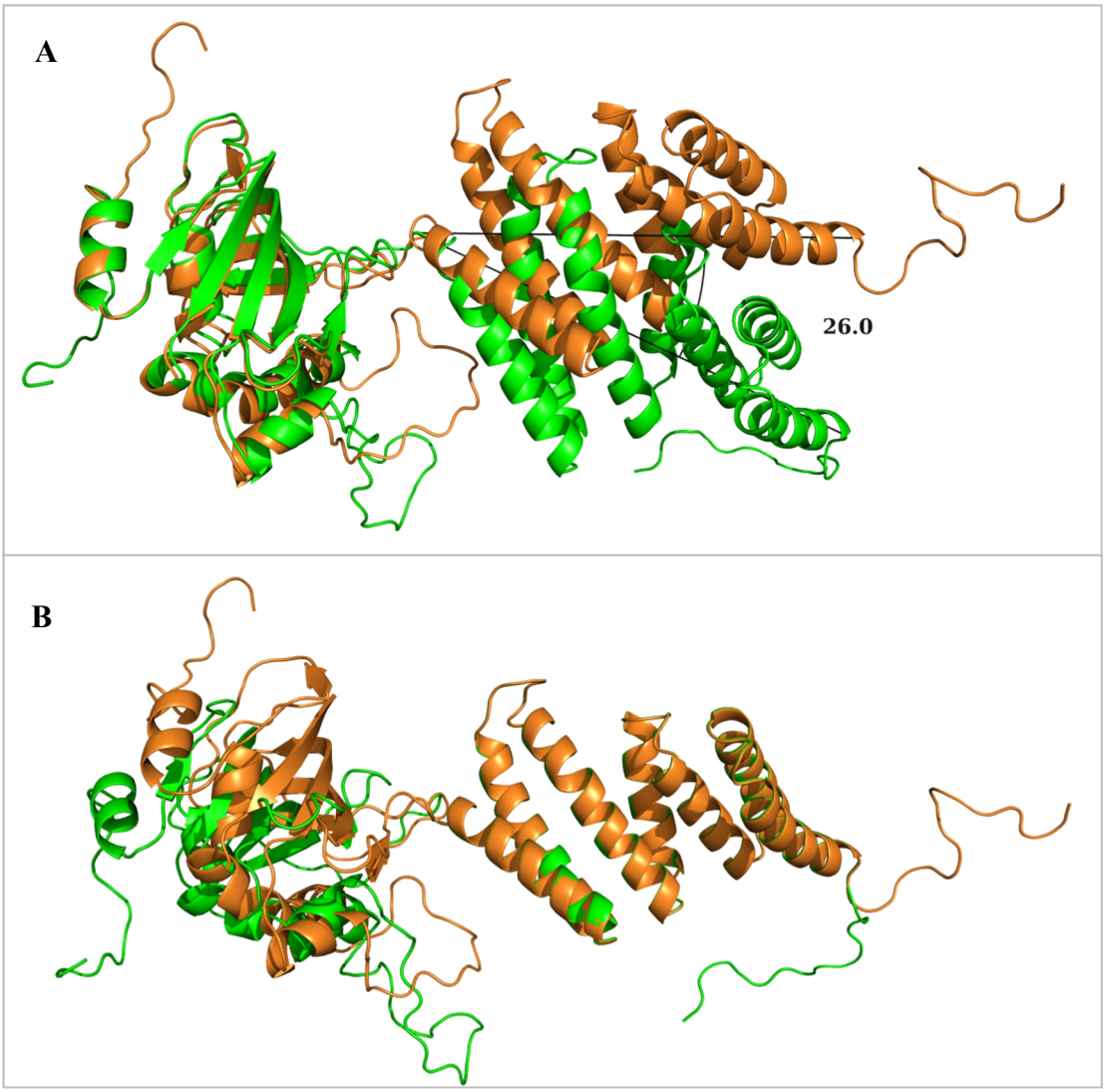



**Figure 4.42: Model alignment. A:** N-domain alignment of the two best models of $H_6$-AIP; **B:** C-domain alignment of the same models. Model 1 is shown in green while model 2 is shown in orange. Model alignment was performed using PyMol (Schrödinger, 2015)

Both of these models fitted well the scattering profile and are in agreement with the calculated Dmax, suggesting that AIP can be in either conformation in solution. The alignment analysis shows that the relative position of the N-domain loop has an important role in the orientation of the two domains with respect to each other. Panel A, demonstrates that when both models are aligned at their N-domain, the C-domain is misaligned by an angle of 26º. In both cases only one domain from each model could be precisely aligned. The C-terminal tail is also highly flexible and can take different orientations in solution. Interestingly, all the other forty-eight models that were generated by FoXS had the flexible loop and C-domain tail region in different positions. However, only the two shown in Figures 4.42, had an acceptable $Chi^2$ value. All other models fitted the scattering data poorly, clearly suggesting that although these regions have more flexibility than the other secondary structure elements; these two conformations of AIP are more likely to exist in solution in the absence of binding partners. This finding indicates that the protein is not entirely flexible with the two domains moving in varying directions, but rather it is an extended molecule, with regions of flexibility. In model 2 (shown in orange), the N-domain loop is closer to the first TPR motif in the C-domain, with the closest measured distance being that of 4.8 Å, that is 6.7 Å closer than the distance measured for model 1.

To further investigate the structural differences between the two best models, the interface face residues of both models were determined. To do so, the N- and C-domain were considered as two separate chains and a cut-off value of 1.0 Å was set. This determines the difference in area over which residues are considered as interface residues. Amino acids were considered interface residues if the difference in their accessible surface area from the complex to a single chain was greater than 1.0 Å. As expected, the number of interface residues between domains is higher in model 2 than model 1 due to the closer proximity of the loop in the former model. In the case of model 1, none of the amino acid residues found in the N-domain loop qualified as an interface residue.

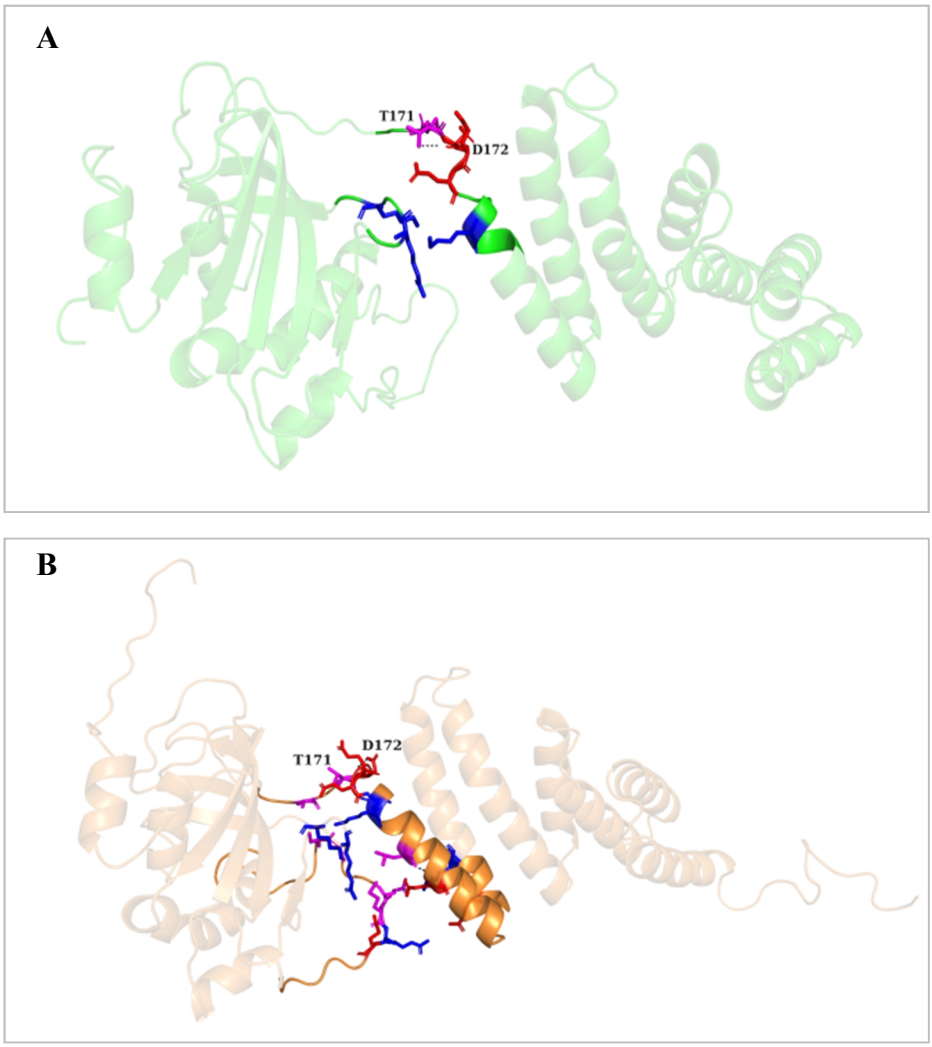



**Figure 4.43: Inter-domain interface residues. A:** $H_6$-AIP Model 1; **B:** $H_6$-AIP Model 2. In each model, negatively and positively charged residues are coloured in red and blue respectively. Other amino acids are shown in purple. In both models, only the residues T171 and D172 formed a weak electrostatic interaction of 4 Å. The interface face residues were generated using the InterfaceResidues.py script in PyMOL version 2.3.0 (Schrödinger, 2015).

Although model 2 (Figure 4.43 B) displays a higher number of interface residues, no polar contacts (that include hydrogen bonds and salt bridges), were identified within this region, implying that any potential interacting residues are not within bonding distance (>4 Å). This further suggests an elongated structure. Similar analysis was also carried out on the $H_6$-SUMO-AIP model and the model predicted from I-TASSER. Results are displayed in Table 4.17.

**Table 4.18: Comparison of model dimensions**

| Parameter | Protein model | | | |
|---|---|---|---|---|
| | **$H_6$-SUMO-AIP** | **$H_6$-AIP Model 1** | **$H_6$-AIP Model 2** | **I-TASSER Model** |
| Dmax (Å) | 130.9 | 101.2 | 103.8 | 82.1 |
| [7]Length of linker (Å) | 13.9 | 13.2 | 12.7 | 8.2 |
| [6]Width of N-domain (Å) (excluding loop) | 33.8 | 36.0 | 35.5 | 30.8 |
| [6]Width of C-domain (Å) (excluding tail region) | 56.8 | 51.8 | 54.9 | 47.7 |
| Closest distance between N-domain flexible loop and C-domain (Å) | 4.8 | 26.4 | 11.5 | 8.1 |
| Number of interface residues | 11 | 8 | 19 | 45 |

The predicted model of AIP using I-TASSER (Zhang, 2008; Yang, *et al*., 2015) was based on the $H_6$-AIP sequence. The most notable difference that can be observed on comparing this model to the two $H_6$-AIP models obtained from SAXS is that the predicted model is more compact with a closer orientation of the N- and C-domain with respect to each other. The width of each domain is also relatively smaller and the linker region is intertwined, rather than fully extended as observed in the SAXS generated models. This is further observed through the higher number of inter-domain interface residues that were identified in this predicted model.

Since the SAXS models are based on solution scattering of the protein, the generated models have more physiological relevance.

[7] For these measurements, the same amino acid pairs were chosen for fair comparison amongst models.

### 4.7.8. SAXS models validation

As a final analysis, the models were validated using MolProbity (Davis *et al*., 2007; Chen *et al*., 2010). This online structure-validation server evaluates the model quality based on hydrogen placement, all-atom contact and analyses of covalent-geometry and torsion angles (Chen *et al*., 2010). Although this service is essentially designed for structures obtained through X-ray crystallography or NMR, it was considered appropriate to upload the models obtained by SAXS to obtain the Ramachandran plot for each model.

The Ramachandran plot gives an overview of the amino acids that are in a stable allowed conformation and highlights the presence of any outliers. Within this plot each amino acid is represented as a circle on a graph of the dihedral torsion angles, psi (ψ) against phi (φ). The dimension of each circle represents the van der Waals radii for that particular amino acid. Allowed regions that signify stable amino acid conformations are shown by coloured contours. Amino acids outside of these regions have ψ and φ angles which are not sterically favoured and are therefore considered as outliers (Cooper, 1995; UCSF, 2017).

The Ramachandran plot as displayed by MolProbity is shown as six panels of probability contours, coloured in blue. The first designated as 'General case' represents all the amino acids except for glycine and proline. Amino acids that precede proline residues are also excluded and are shown in a panel labelled as 'pre-proline'. Similarly, glycine and proline amino acids are displayed separately in an individual panel. Both glycine and proline have a higher likelihood of occurring in a disallowed region than other amino acids. The former because of its high flexibility and the latter due to its cyclic side group that significantly restricts rotation and the φ values (Ho and Brasseur, 2005). The results obtained for $H_6$-AIP Model 2 are shown in Figures 4.46-4.47 respectively. The results of $H_6$-AIP Model 1 and $H_6$-SUMO-AIP can be found in Appendix A.5.

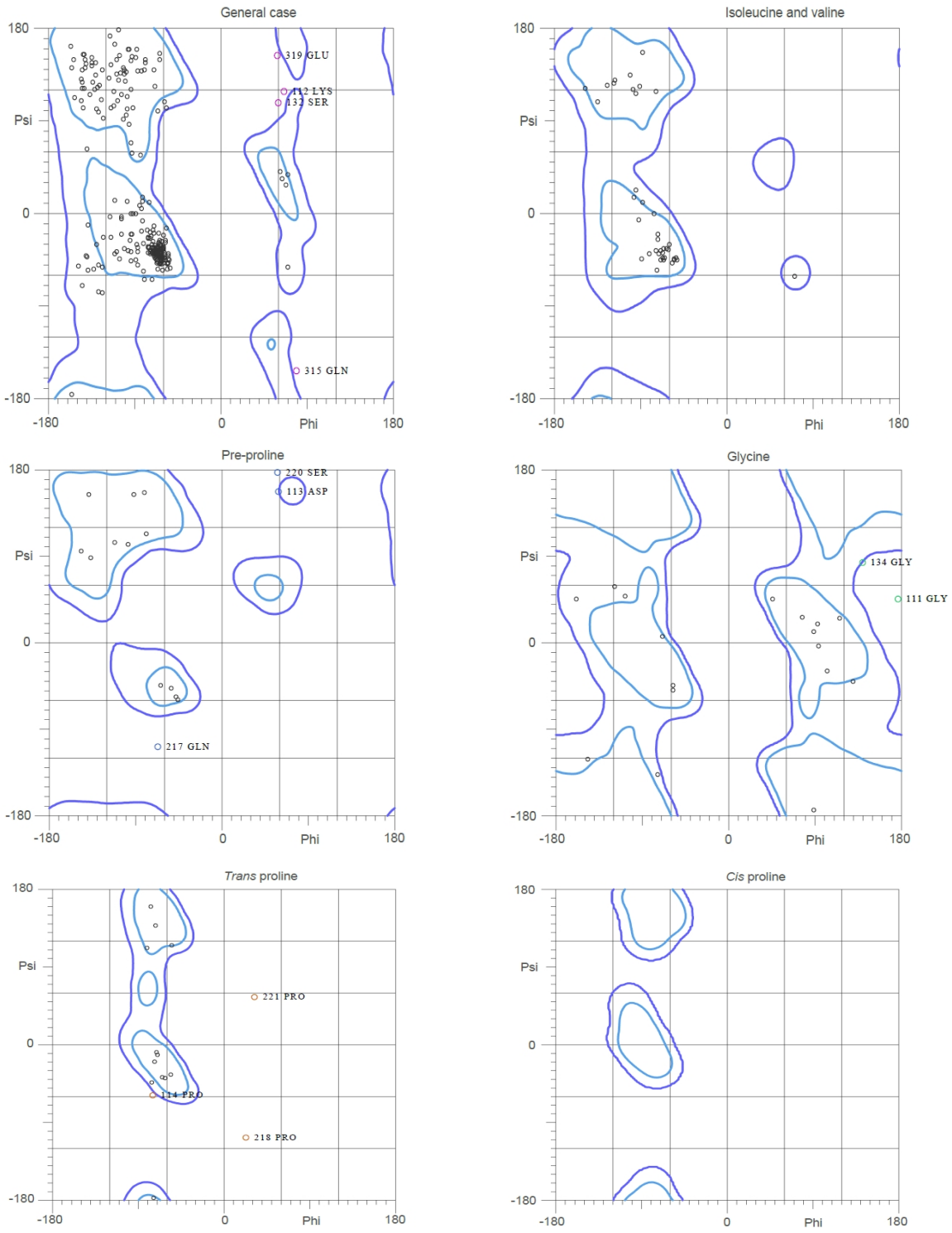



**Figure 4.44: Ramachandran plot of $H_6$-AIP Model 2.**

Analysis of the plot shows that 83.7 % of all the amino acid residues (296/339) were found in favoured (98 %) regions, while. 96.5 % (327/339) were in allowed (> 99.8 %) regions. A total of 12 outliers were identified. This plot was generated using MolProbity (Davis *et al*., 2007; Chen *et al*., 2010).

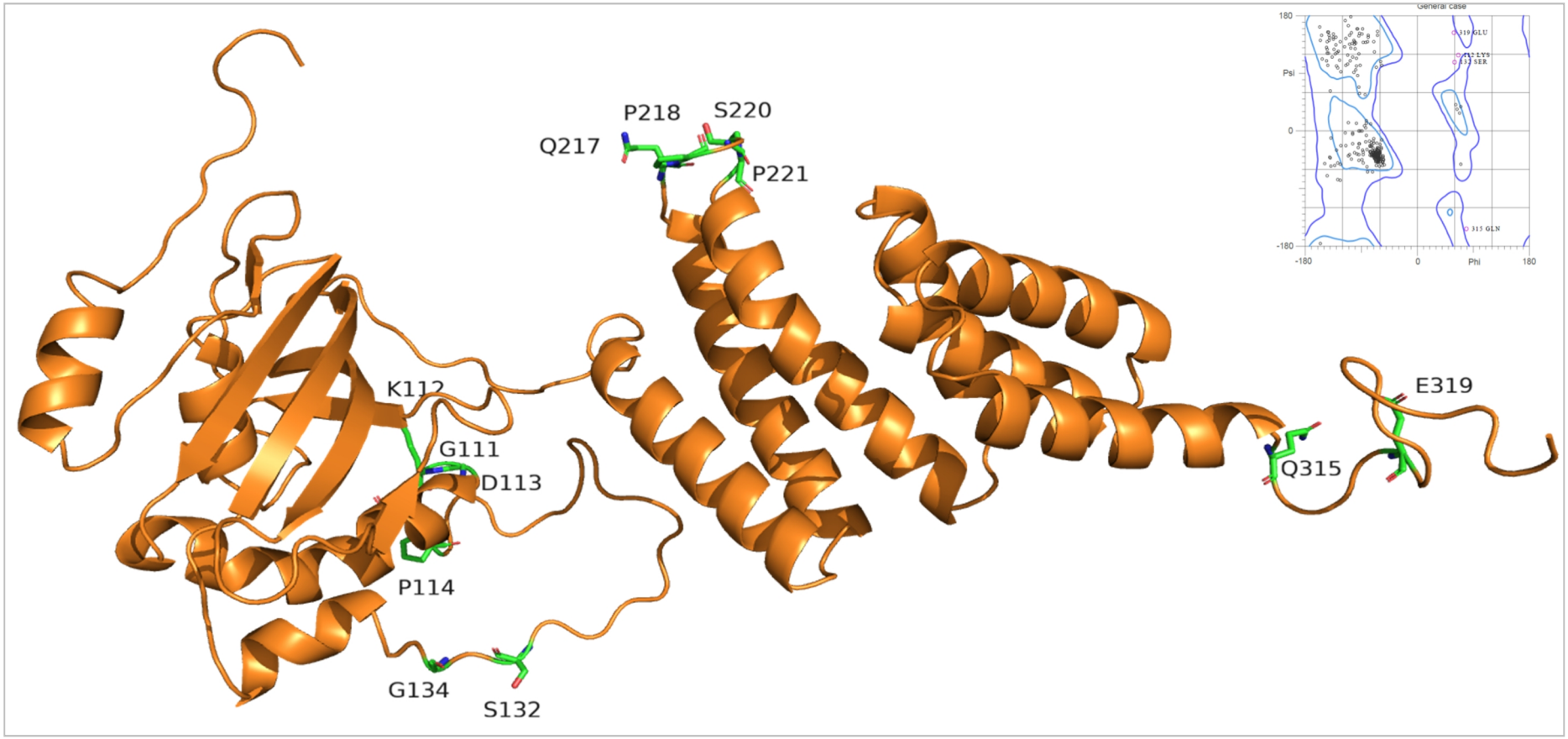



**Figure 4.45: Cartoon representation of $H_6$-AIP Model 2.**

The location of each outlier is displayed as sticks. Each amino acid is coloured according to element C, O, N, S. Hydrogrens are not shown. This figure was generated using PyMOL v2.3 (Schrödinger, 2015). The general case Ramachandran plot is shown on the bottom right hand side.

For each of the three models, the number of outliers would be considered high by crystallographic standards, as these aim for an outlier value of <0.05%. However, upon closer observation of each model, it can be noted that each outlier is within a flexible disordered region of the protein. With respect to the $H_6$-SUMO-AIP model, 45 % of the outliers are in the disordered region of the $H_6$-SUMO tag and 36% are within the disordered loop region of the *βD-βE* extension. Similarly, 47 % of the outliers found in $H_6$-AIP model 1 and 50 % of those in model 2 are also within this flexible region. For both $H_6$-AIP models, 15-20 % of the other outliers were observed at the C-terminal tail of the protein and the rest are proline and pre-proline residues. Unlike α-helices and β-sheets, loop regions are less restrained and occupy a broader range in the Ramachandran plot. As such, it is more likely that during protein modelling, the dihedral angles of these regions occur in conformations that may not be sterically permitted in most proteins and thus listed as outliers.

### 4.7.9. Effect of the mutations on the structure of AIP

After having obtained a model for the full length AIP protein, the next step was to investigate any possible effect the N-domain mutations may have on the structure of AIP.

Each mutation was separately introduced in the SAXS generated model using PyMOL version 2.3.0 (Schrödinger, 2015). In each case, the rotamer of least strain was selected. Each of the four models with the respective N-domain mutation was further refined by energy minimization performed on SWISS-PDB Viewer version 4.1 (Guex and Peitsch, 1997). Only the mutated residue was selected for energy minimization using 20 steps of steepest descent. This additional step was performed to obtain the most energetically favourable conformation at the site of mutation. Wild type AIP was aligned with the model harbouring the mutation and the amino acid residues that are 4 Å away from the site of interest were selected. Intra-molecular distances within this region were identified and measured. Moreover, a qualitative electrostatic representation of each model was generated. This was necessarily to assess and compare the local electrostatic potential at the site of mutation and confirm that it is in agreement with the ionisation state as predicted from the amino acid composition. This representation was generated using the 'vacuum-electrostatics' function on PyMOL that generates an image by averaging the charges using a semi-Coulombic shape

convolution function and illustrates the local charge distributions on the protein (Pan and Aller, 2015). A summary of the results obtained is represented in Figure 4.48.

The first N-domain mutation analysed was the R9Q, in which the positively charged arginine was substituted with the neutral glutamine. The side chain of the arginine at position 9 is projecting outwards and does not seem to be involved in any intramolecular interactions with neighbouring residues (Figure 4.48A). Secondary structure prediction using SPIDER2 (Figure 4.35) assigned a score value of 5 to this residue, thus suggesting that 50% of its surface is exposed. The mutation to glutamine not only results in the loss of surface positive charge, which might be involved in interactions due to its relative accessibility, but also creates a [8]moderate electrostatic hydrogen bond (3.2Å) between OE1 of Q9 and main chain N of K15. This hydrogen bond thus connects the Q9 residue in the α0 helical region with the first *β*-strand (*β*A) of the half *β*-barrel, possibly limiting the mobility of the α0 region. Moreover, when the whole α0 region (I4-D11) was analysed for polar contacts none of the residues in the α0 region are in close enough proximity to form such interactions with any of the residues on the *β*A strand. This further adds to the uniqueness of the interaction introduced by the Q9 mutation.

The R16H mutation also resulted in the partial loss of positive charge, which can be confirmed by the electrostatic potential representation (Figure 4.48B). Similarly, the introduction of histidine resulted in the formation of a moderate, mostly electrostatic interaction (2.9 Å) between ND1 of H16 and O of V17. Both H16 and V17 are part of the *β*A strand. Circular dichroism results (Table 4.2) showed that this interaction did not increase protein thermal stability.

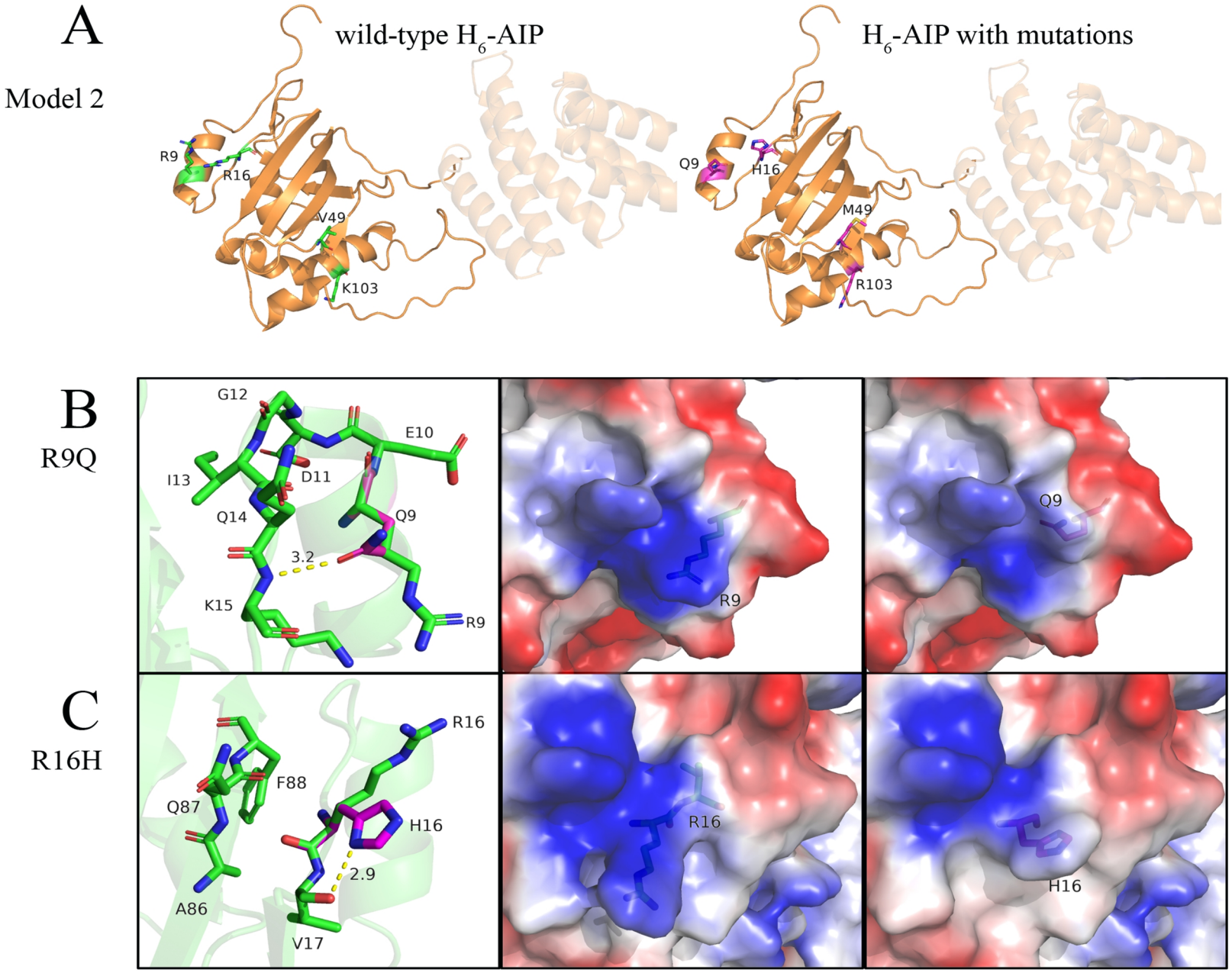
The other N-domain mutation involved the substitution of valine to methionine at position 49. The classification of methionine is in some cases considered as dichotomous. Although it has a strong dipole moment exerted by the divalent sulfur atom and so unlike valine, it is not entirely non-polar, methionine residues are not often involved in hydrogen bonding (Gregoret *et al*., 1991).The C-S-C group of methionine

[8] The categorization of hydrogen bonds between donors and acceptors, as strong (mostly covalent), moderate (mostly electrostatic) or weak electrostatic are based on the classification published by Jeffrey (1997).

side chain residue was often classified as a hydrophobic moiety, until findings showed that the divalent sulfur atom of methionine can form non-bonded interactions with oxygen atoms, in which the sulfur atom acts as the electrophile (Pal and Chakrabarti, 2001; Iwaoka and Isozumi, 2012). This makes methionine relatively unique as it appears to engage in hydrophobic interactions within the protein core and also interact with polar oxygen atoms either of the main chain carbonyl or side chain carboxylate group (Pal and Chakrabarti, 2001).

The substitution of the non-polar hydrophobic valine residue with methionine caused no apparent change in electrostatics (Figure 4.48C). Secondary structure prediction showed valine to be relatively buried, with an accessibility scoring of 3 (Figure 4.35). Since the side chain of valine cannot be involved in polar contacts, the introduction of methionine did not result in the loss of any interactions. On the contrary, the methionine side chain was in close enough proximity to form an interaction (2.6 Å) between SD of M49 and OD1 of D52, both residues form part of the *β*C strand.

The overall surface charge was minimally affected in the case of the K103R mutation, as the positively charged lysine was replaced by another positively charged residue. Polar contact analyses demonstrated that in the wild type protein there is a moderate hydrogen bond of 3.4 Å between NZ of K103 and O of P99. Both K103 and P99 are found on the αIII helix that forms the first structural element of the *β*D-*β*E extension. While, this interaction is lost in the R103 mutant, a different interaction is formed that involves NH2 of R103 and OD1/OD2 of D139. Interesting, D139 is located on the αIV helical region that marks the end of the *β*D-*β*E extension. In the wild type protein, none of the amino acids in the αIII region form hydrogen bonds and/or salt bridges with the residues on the αIV helix, since the distances are longer than 4 Å. The absence of such interactions might partially account for the flexibility observed in the loop region that is found in between these two helical structures and therefore suggests that the interaction that is introduced in the R103 mutant might limit this mobility.


A
Model 2
wild-type $H_6$-AIP
$H_6$-AIP with mutations
R9
R16
V49
K103
Q9
H16
M49
R103
B
R9Q
G12
E10
I13
D11
Q14
Q9
3.2
K15
R9
R9
Q9
C
R16H
R16
F88
Q87
H16
2.9
A86
V17
R16
H16

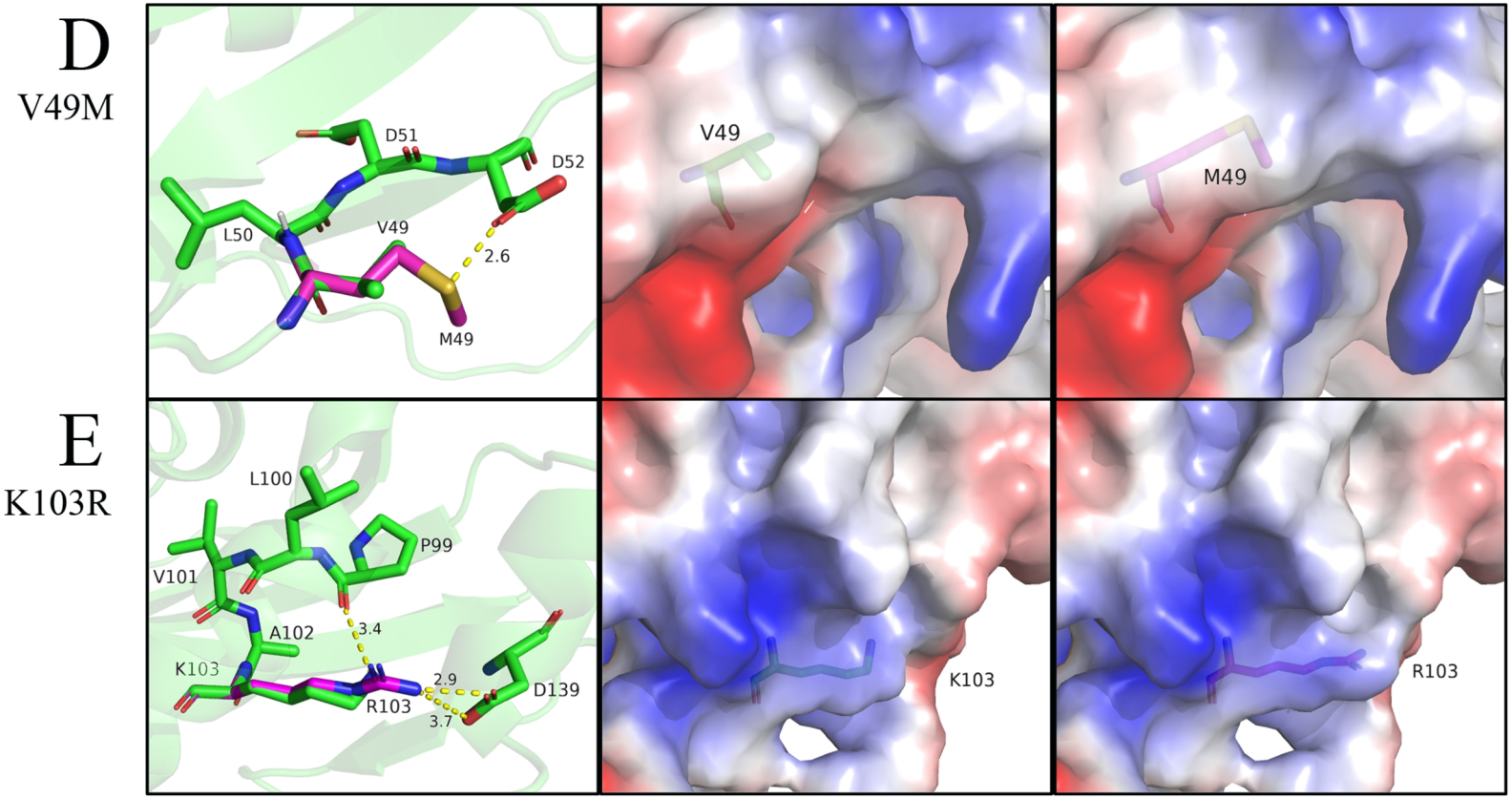



**Figure 4.46: Structural analyses of wild type AIP and AIP mutants. A**: Visual representation of the position of each N-domain mutation. **Panels B-E:** Close up evaluation of each mutation (R9Q, R16H, V49M and K103R). Figures from left to right represent the amino acid interactions and the electrostatic potential around the amino acid of interest for wild type AIP protein and the mutant counterpart respectively. The carbon chain in the wild type protein is depicted in green and the side chain of the mutated reside is overlaid and carbon atoms are coloured in purple. For both wild type and mutated amino acid residue, N, O and S atoms are shown in blue, red and yellow respectively. These analyses were performed using the SAXS generated AIP Model 2.

# Chapter 5 Discussion

## 5.1. Discussion

The scope of this study was to investigate the structure of the full length AIP protein and determine whether N-domain mutations affect the protein's ability to interact with client binding partners. AIP is a molecular co-chaperone with a vast repertoire of interacting proteins and is thus involved in various biochemical processes within the cell that may collectively or partially, contribute to its role as a tumour suppressor protein. Over the past decade, more than 50 pathogenic *AIP* gene mutations have been identified (FIPA, 2011; National Library of Medicine, 2019). Although truncating mutations usually have a deleterious effect, the impact of missense mutations requires further investigation. Mutations in the C-domain of AIP have attracted vast attention due to their direct involvement in protein-protein interactions (Trivellin and Korbonits, 2011). However, recent studies have identified the N-domain to also play an important role in protein interactions, particularly with Hsp90 (Linnert *et al*, 2013) and the scaffolding protein CARMA1, the latter being solely dependent on the N-domain of AIP for interaction (Schimmack *et al*., 2014). This new emerging evidence further suggested the need to investigate the effect of N-domain mutations in protein-protein interactions.

The mutations selected for this project were all in the N-domain and with clinical relevance. The R9Q mutation was first characterised in Malta in a patient suffering from acromegaly (Formosa *et al*., 2010) but has later been reported in other pituitary adenoma patients (Puig-Domingo *et al*., 2011; Cazabat *et al*., 2012). The three other selected mutations were R16H, V49M and K103R. A summary of the clinical heterogeneity associated with each mutation is described in Table 1.1. A study carried out locally on these four missense mutations, demonstrated that the R9Q, R16H and V49M lose their tumour suppressing ability 96 h post-transfection in GH3 pituitary adenoma cell line, when compared to the wild type AIP (Formosa and Vassallo, 2017). These findings further reinforced the need to characterise these mutants *in-vitro* through their effect on binding to other client proteins.

To this end, AIP was successfully purified as a full-length protein. This study is the first to describe the use of glycerol and DTT as additives employed during both purification and protein storage to ensure sample homogeneity. In their absence, the protein was observed to multimerise and aggregate significantly with time, even when

stored at -20°C (Figure 3.6). The protein was initially expressed and purified as a fusion to a $H_6$-SUMO tag. This tag has recently been shown to enhance the solubility of various proteins expressed in an *E.coli* expression system (Kuo, Nie and Courey, 2014). However, its removal detrimentally affected both the stability and solubility of AIP, rendering it impractical to work with the untagged protein and its mutant counterparts (Figure 3.13). This led to the cloning, expression and purification of $H_6$-AIP. Although the $H_6$-tag was also not cleaved from the protein, the smaller size of this tag when compared to $H_6$-SUMO, made it less likely to interfere with the functional studies that were carried out to characterise AIP mutants. In fact, for most functional and structural studies carried out on other proteins, the small size of the $H_6$-tag was found to have minimal or no interference with the folding and structure of the target protein and in some cases, the $H_6$-tag actually assisted in crystal formation by enhancing protein stability. For this reason, many independent groups working with $H_6$-tagged proteins opt to retain the tag even after purification (Carson *et al*., 2007; Malhotra, 2009).

AIP and mutants were purified using the same procedure, attaining purity levels ranging from 88-93%. The homogeneity of each sample was confirmed by Native-PAGE. For all proteins, there was a significant decrease by almost two-fold in the final yield of the $H_6$-SUMO tagged proteins when compared to the $H_6$-tagged counterparts, suggesting that in this case the $H_6$-SUMO tag had a negative effect on protein expression. Circular dichroism analyses showed that was no difference in thermal stability between $H_6$- and $H_6$-SUMO tagged proteins. $H_6$-AIP showed a melting temperature of 46.7°C ± 0.7 (Table 4.2) when analysed by circular dichroism at 222 nm.

Irrespective of the purification tag, AIP, R9Q and K103R were expressed and purified to a similar yield. The yield obtained from the purification of R16H and V49M was comparatively lower (Tables 3.7-3.8), suggesting that these two mutations may have an effect on protein expression by destabilising the protein in the cell. Supporting this is the fact that the V49M mutant has been classified to have a 'short' half-life in HEK293 cells due to enhanced proteasomal degradation (Hernandez-Ramirez *et al*., 2016). Similarly, cycloheximide chase analyses on GH3 cells transfected with wild-type and N-domain mutants show that the mutants R9Q, R16H and V49M have a significantly higher rate of degradation, 28 h post-treatment, with reported half lives of 18.5 h, 18.0 h and 16.0 h respectively (Formosa and Vassallo, 2017). Nonetheless,

despite the differences in yield, all purified AIP mutants showed CD spectra identical to the spectrum of the wild type protein, indicating that all proteins had an ordered structure and that the mutations did not cause any disruptive effect on the overall structure of AIP (Table 4.3). Moreover, the melting temperature of each mutant was also very similar to the value of the wild type, with the differences ranging from 1.3°C (R9Q) to a maximum of 4.6°C (V49M). Degradation was however observed on mass spectra, whereby V49M showed considerable fragmentation (Appendix A.2). This is a common feature of some proteins and can be useful for identification using slightly different techniques in the mass spectrometer.

While, cell biology studies suggest that N-domain mutations destabilise AIP (Hernandez-Ramirez *et al*., 2016; Formosa and Vassallo, 2017), the results herein suggest that protein stability, per se, is not necessarily the cause and that biological stability may therefore depend on other factors. This study shows how protein over-expression causes aggregation of AIP and mutants, which can successfully be prevented through the use of additives. However, this is not possible *in vivo,* so it is quite likely that over-expression of protein in cell studies may cause aggregation of mutant AIP, triggering its subsequent removal by the protein "quality control" mechanisms, such as the ubiquitin-proteasome pathway. Within the cell, native AIP may not be so susceptible to aggregation due to over-expression because its binding capacity for its partners is not affected. The reduced half-life and enhanced proteasomal degradation observed *in vivo* for N-domain mutants, particularly for V49M (Hernandez-Ramirez *et al*., 2016), could thus be attributed to their reducing binding affinity to client partners as demonstrated in this study.

The binding partners in this study were selected from the list of proteins described by Trivellin and Korbonits (2011). An updated version of the protein interactome of AIP is illustrated in Figure 5.1. One enzyme (phosphodiesterase), PDE4A5, was chosen to enable evaluation of the effect of AIP and mutants on the catalytic activity of its binding partner. Hsp90β was chosen based on the fact that to date there are no studies that have characterised the *in-vivo* or *in-vitro* interaction of AIP mutants with this binding partner (Bizzi *et al*., 2019).

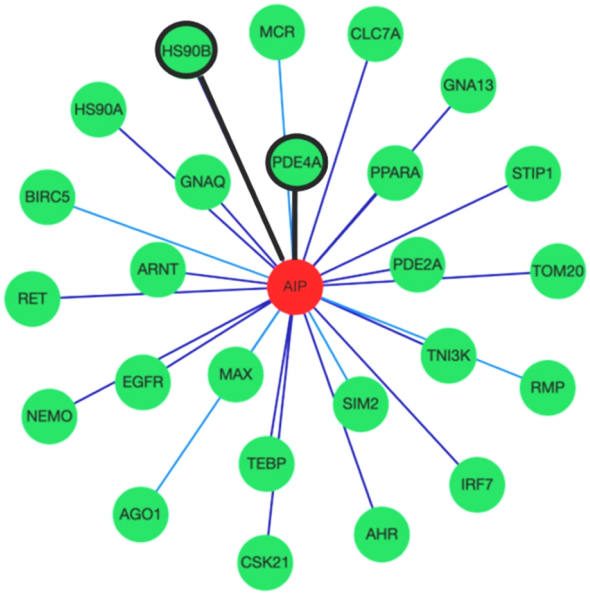



**Figure 5.1: AIP protein interactome c.2019.** The interactions that were investigated in this study are indicated with solid black lines. This interactome was generated using the human protein-protein interaction database, PICKLE version 2.3 (Gioutlakis, Klapa and Moschonas, 2017; Klapa *et al*., 2013).

Hsp90β$^{(513\text{-}724)}$ was purified with protein yields of 29 mg.L$^{-1}$ of media and purity levels of 95%. The purification of $H_6$-SUMO-PDE4A5 was more laborious and although the expression was optimal, a significant amount of protein was produced as inclusion bodies and subsequently lost throughout the purification, a problem that appears to be common for PDE4s (Richter, Hermsdorf and Dettmer 2002; Lugnier, 2005). Since the CD spectrum of the refolded $H_6$-SUMO-PDE4A5 lacked appreciable secondary structure (Figure 4.3), the purification procedure was optimised to maximise the fraction of soluble PDE. The best yield obtained was of 5.6 mg.L$^{-1}$ with purity levels of 87%.

Protein binding was initially analysed using circular dichroism by monitoring changes in the secondary and tertiary structure respectively. The binding of $H_6$-AIP to Hsp90β$^{(513\text{-}724)}$ resulted in a subtle change in tertiary structure, whereas binding to $H_6$-SUMO-PDE4A5 caused an observable change in the far-UV spectrum, indicative of secondary structure change upon binding. Binding kinetics data was then obtained through SPR and ITC experiments. Both techniques were carried out using in 1x PBS pH 7.4 as running and/or dilution buffer. This buffer was chosen on the basis that it

mainly resembles physiological conditions and is thus a better representative of the intracellular environment. Although, ITC experiments were also repeated using Tris-Cl pH 7.5 as dilution buffer, the fold difference in binding affinity between wild type and mutant proteins was essentially the same for both buffers. In addition, while all proteins were purified and stored in the presence of glycerol and DTT as additives, the use of such additives was kept to a minimum when investigating binding studies, especially when working with low concentration of protein as is required for SPR experiments. In the case of ITC experiments, the presence of 5% (v/v) glycerol in the sample had to be retained, as its removal would have resulted in protein multerimerisation and aggregation, resulting in unusable ITC data. Although one might argue that the presence of glycerol might in itself have had an effect on protein binding, all tested samples had the same buffer composition and thus any differences observed in binding affinities between wild type and the mutant counterparts is due to the N-domain mutation under investigation.

The interaction between AIP and Hsp90β$^{(513-724)}$ was characterised both through ITC and SPR experiments. The $K_D$ values obtained from SPR using the Langmuir model (1.23 μM) and bivalent analyte model (2.45 μM) are comparable to the value (2.28 μM), published by other groups (Li *et al*., 2013). The binding affinity of Hsp90 and the TPR domain of AIP (residues 166-330) had been previously characterised through ITC with reported $K_D$ values of 13.3 μM and 18.6 μM for full length Hsp90 and an Hsp90 peptide (amino acid sequence: EDASRMEEVD) respectively (Morgan *et al*., 2012). The lower $K_D$ values obtained in this study (2.19 μM) using full length AIP, reinforce the fact that full length proteins show binding specificity that may not always be reflected when using the individual domains or peptides (Musacchio, 2003). In contrast to previous studies, whereby only the C-domain of AIP was considered essential for Hsp90 binding (Kazlauskas *et al*., 2002; Morgan *et al*., 2012), this study indicates that the N-domain has a significant effect on the protein's binding affinity to its client partner, Hsp90. This is in agreement with the findings of Linnert *et al*. (2013) who also report the involvement of the N-domain in mediating the interaction between AIP and Hsp90.

Binding kinetics of the N-domain mutants showed that the V49M mutation decreased the protein's binding affinity by 4-5 fold. This was followed by the R16H that displayed a binding affinity around 3-4 fold lower than wild type. The R9Q mutation showed a 2-fold difference in $K_D$ while the K103R mutant bound to Hsp90β$^{(513-724)}$ with an affinity comparable to that of the wild type protein (Figure 4.17). From a physiological view point, the lowering in binding affinity, observed in the R9Q, R16H and V49M mutants suggests that these mutation may destablise any complexes that require the interaction of AIP and Hsp90. Within the AIP interactome, Hsp90 is present as part of AIP-Hsp90-AhR-p23 complex where it assists complex stability and maintains the AhR in a structural conformation that is receptive to ligand binding (Kazlauskas *et al*. 2001; Tsuji *et al*., 2014; Kudo *et al*., 2018). Destabilisation of such a complex might therefore correlate to a lowering in AhR stability and a consequent reduction in its biological function. Considering the involvement of AhR in cancer pathology, including its potential role as a tumour suppressor in the pituitary gland (Formosa, Borg and Vassallo, 2017), any mutation that might directly or indirectly destabilise it, will have a detrimental effect on the cell. Similarly, Hsp90 mediates the binding of AIP to the glucocorticoid receptor (GR), to form a multiprotein heterocomplex that also contains the co-chaperone p23, protein phosphate 5 (PP5) and FKBP52 (Grad and Picard, 2007). In the presence of ligands (glucocorticoids), GR translocates to the nucleus where it homodimerises and activates the transcription of target genes (Heitzer *et al*., 2007), a mechanism similar to the one observed for AhR. Destabilisation of the complex due to a lowering of the AIP-Hsp90 binding affinity might therefore affect GR-mediated transcription and signaling. This hypothesis can be tested by studying the GR-mediated transcription and signalling pathway *in vivo*, for instance through the use of a luciferase reporter gene assay.

The other interaction investigated was between AIP and PDE4A5. PDE4A5/4 has a unique interaction with the co-chaperone AIP. This interaction inhibits the enzymatic activity of PDE4A5, resulting in an increase in cellular cAMP and activation of protein kinase A (PKA). Therefore the presence of AIP mutations that result in loss and/or a reduction in protein function are expected to increase PDE4A5/4 activity, lower cellular cAMP levels and down-regulate PKA (Bolger *et al*., 2003, 2016).

This study is the first to kinetically characterise the interaction between AIP and PDE4A5 using full length proteins that show a $K_D$ value of 2.64 μM (Table 4.9). This is in contrast to a previously published study whereby the interaction between the TPR domain of AIP (residues 166-330) and a short peptide of PDE4A5 (amino acid sequence: TLEELDW) was studied through ITC and reported a $K_D$ value of 64.5 μM (Morgan *et al*., 2012). This discrepancy in value clearly highlights the binding specificity that is exhibited by full length proteins as opposed to single domains and/or short peptide sequences. While the use of full length proteins may not also be feasible due to the difficulties associated with expressing and purifying stable forms of the protein, native proteins provide a more accurate representation of the physiological state. Single domains and/or short peptides are usually easier to purify (Tsuji *et al*., 2014) and can help to elucidate potential binding sites in a protein-protein complex. However, one cannot exclude the possibility of other binding contacts that may only be observed when using full length proteins, thus reflecting the importance of working with native proteins whenever possible. Reinforcing this is the fact that the interaction of AIP-PDE4A5 was believed to solely require the TPR domains of AIP (Bolger *et al*., 2003). The results herein however, categorically indicate that the N-domain mutations, including near N-terminal mutations like the R9Q, have a negative effect on the AIP-PDE4A5 interaction.

All mutants exhibited a lower binding affinity than the wild type. Considering the data from flow-cells (3 and 4) in the Biacore experiment (Section 4.4.3), K103R and R16H showed 4-5 fold and 3-fold lowering in binding affinity respectively. This is comparable with the data published by another group who studied the interaction in yeast cells through β-galactosidase assay and report a >5-fold and 2-fold difference in binding for K103R and R16H respectively (Igerja *et al*., 2010). This current study supports this data and provides $K_D$ values that describe the specific interaction between AIP and PDE4A5 as the sole proteins present in the mixture. Similar to the AIP-Hsp90β$^{(513-724)}$ interaction, the V49M mutant showed the least binding affinity, with 5-6 fold difference from wild-type. This notable difference was not observed in the β-galactosidase assay (Igerja *et al*., 2010). The reduction in binding affinity observed through SPR for AIP mutants correlates with their reduced ability to inhibit PDE4A, as measured through a PDE enzymatic assay. In other words, PDE4A was more active in the presence of AIP mutants when compared to wild type AIP. The lowest inhibition of

PDE4A activity was observed for the V49M and K103R mutants that also showed the lowest binding affinity. Similarly, AIP showed the strongest binding affinity and exerted the highest inhibition of enzyme activity, achieving a maximal inhibitory value of 77 ± 2 % (mean ± S.D). An inhibition value of 62 ± 9% has been previously reported for AIP mixed with COS7 cells over-expressing PDE4A5 (Bolger *et al*., 2003). This is comparable to the inhibition value (64 ± 4%) observed in this study, at 1:1 protein ratio. The use of lysate as opposed to pure proteins, may account for the differences observed between studies.

While the mutants displayed the same trend in binding and inhibition studies, the fold differences observed were not always directly proportional. This signifies that a 2-fold decrease in binding affinity did not necessarily correspond to a 2-fold decrease in enzyme inhibition, implying that not all of the bound molecules lead to inhibition (Table 5.1). This suggests that binding affinity alone cannot be used to accurately determine the extent of PDE inhibition effectiveness. For accurate measurements, inhibition studies must therefore be carried out rather than inferred from binding experimental data.

**Table 5.1: Differences in binding and inhibition of PDE4A by AIP mutants compared to wild type AIP.**

| AIP mutant | PDE 1:1 binding (fold difference) | PDE 1:1 inhibition (fold difference) |
|---|---|---|
| **R9Q** | 2-4 | 2.0 |
| **R16H** | 2-3 | 2.3 |
| **V49M** | 5-6 | 3.5 |
| **K103R** | 4-5 | 3.5 |

From a physiological view point, these results would suggest a lowering in cellular cAMP levels. Although, there is evidence that the cAMP signaling pathway is dependent on the interaction of AIP with $G\alpha_i$ proteins and does not necessarily require PDE (Tuominen *et al*., 2014; Formosa *et al*., 2013), this does not lessen the importance and relevance of the AIP-PDE4A5 interaction. While it may not necessarily play a direct role in tumourigenesis, mutations such as the ones investigated that affect PDE

activity and therefore its ability to regulate the cellular levels of cAMP, can affect other disease-associated pathways. Furthermore, when considering the vast protein interactome of AIP, disease predisposition may result due to various pathways disrupted to varying extents by the presence of AIP mutations. Interestingly PDE4A4 (the human homologue of the rat PDE4A5 used in this study) was found to be under-expressed in sporadic somatotroph adenomas with AIP mutations when compared to the protein levels present in the same adenoma sub-types with functional AIP. This decreased protein expression suggests that the lack of functional AIP hinders the expression of PDE4A4 and consequently alters the balance of the cAMP-PDE pathway. This may play a role in pituitary tumourigenesis (Bolger *et al*., 2016; Bizzi *et al*., 2018, 2019). This further reinforces the rationale that any change that may alter the homeostatic levels of cAMP in the cell cannot be disregarded.

In attempts to better understand the structure and function of AIP, the structure of the full length protein was determined through SAXS. Although the N- and C-domains of the protein had been previously characterised through NMR and X-ray crystallography respectively (Linnert *et al*., 2012; Morgan *et al*., 2013), the relative orientation of the two domains with respect to each other was unknown. The results indicate that AIP has an extended structure with a linear arrangement of the two domains. This is in contrast to the predicted structures obtained through I-TASSER (Figure 4.24) that suggested a more compact structure, with a smaller Dmax and a significantly higher number of interface residues (Table 4.17). In agreement with NMR data (Linnert *et al*. 2013), the unstructured loop region (G111-H135) that is part of the *β*D-*β*E extension showed conformational variability. However while NMR data revealed multiple conformations of the unstructured loop, most likely indicating a high degree of flexibility and mobility of this region, the SAXS generated models show that some of these conformations are not physiologically relevant due to classical steric hindrance with the C-domain (Appendix A.5). In addition, SAXS revealed that the relative position of this loop has a significant effect of the structure solution and although it can take different conformations, only two models fit the SAXS scattering profile, revealing the most stable to occur in solution.

Analyses of the intermolecular interactions at the site of mutation showed that in all cases, the mutated residue resulted in the formation of an electrostatic interaction (Figure 4.52), making it less likely to act as a potential binding contact. This is particularly applicable to the R9Q mutant as its residue is the most accessible. On the contrary, V49M is rather buried, with an accessibility score of 3 (Figure 4.35) and thus less likely to interact directly. Nonetheless, this mutant exhibited the least binding with both binding partners. The replacement of valine by methionine resulted in an interesting S-O interaction. Mutagenesis experiments published by other groups showed that when adjacent core residues in T4 lysozyme such as leucine and isoleucine were replaced with methionine, the stability of the single mutants was lowered due to additional side chain flexibility (Gassner *et al*., 1996). This would suggest the V49M to have a destabilising effect on protein structure. Although in this study, circular dichroism did not show a significant reduction in the thermal stability of this mutant, significant fragmentation was observed on mass spectrometry. Similarly, the low SPR signals observed for this mutant, as well as the low 'n' stoichiometry factor seen in ITC might suggest protein instability. Although, circular dichroism does not confirm this, it can be argued that any interaction that may potentially destabilise the mutant may not form under all experimental conditions. This might account for the differential behaviour shown by this mutant as also described in other studies (Igerja *et al*., 2010; Formosa and Vassallo, 2017). Another interesting scenario was observed in the case of the K103R that resulted in the loss of a hydrogen bond with another residue on the same αIII helix and the formation of a new bond with an amino acid residue on the αIV helix. This newly formed interaction is unique, in the sense that none of the residues on the αIII helix interact with those on the αIV. As the flexible loop region is found between these two structural elements (together forming the *β*D-*β*E extension) it can be argued that the lack of interactions between the αIII and αIV favours loop mobility. Conversely, the formation of such an interaction in the K103R mutant might hinder the mobility of this loop, with possible consequence on protein-protein interactions. In view of these results, it would be interesting to study the structure of protein complexes to further understand protein-protein interactions and their significance. The study of protein-protein interactions is becoming more important particularly for drug design whereby small molecule inhibitors are being designed to target binding interfaces that only occur in the protein complex thus ensuring specificity in the drug's mode of action (Arkin, Tang and Wells, 2014). In addition, the K103 site is an ubiquitinylation site,

with a high score confidence. This post-translation modification is lost in the K103R mutant (Yuan Ze University, 2014). Although the consequence of this affect could not have been observed in this study, due to the use of bacterial expression system, it might have a physiological effect *in vivo*.

This study also shows that recombinant human AhR can be successfully expressed within an *E.coli* system. While AhR is currently at the centre of various studies due to its involvement in cancer pathology (Section 1.9.3.2), structural and functional studies have focused on the use of AhR domain/s due to difficulties in expressing and purifying the native protein (Tsuji *et al*., 2014; Schulte *et al*., 2017; Sasaki-Kudoh, 2018). The recombinant expression of human AhR in *Pichia pastoris* has been investigated. However, the expressed protein was only detectable by immunoblotting and the amount recovered from 1 L of *P. Pastoris* culture was roughly translated into 40 µg of AhR protein (Zheng *et al.,* 2016), that is insufficient for the majority of protein characterisation techniques. The results herein provide an alternative improvement and show that through controlled growth and the use of chemical additives that induce the production of endogenous chaperones and suppress aggregation, AhR is over-expressed, stabilised against degradation and can be observed as an intense band both by immunoblotting and Coomassie staining (Figure 3.38).

## 5.2. Conclusions

Clinically-relevant N-domain AIP mutant proteins have been expressed to homogeneity and used to study protein-protein interactions directly using ITC and SPR techniques. The results clearly demonstrate the effect of mutations in the lowering of binding affinities to each of two client proteins, Hsp90β$^{(513\text{-}724)}$ and PDE4A5, also purified in this work. These effects are further supported by the differential inhibition of PDE enzyme activity by the mutants. Given the large repertoire of binding partners for AIP it is likely that the demonstrable effect of mutations as observed in this study, will also affect other pathways that require functional AIP.

To obtain structural information on native AIP, crystallisation trials were initially performed. Despite, numerous attempts with varying concentrations, additives and conditions, crystallisation of AIP was unsuccessful, most likely due to the presence of

flexible loops within the protein. Although the removal of such loops can encourage protein crystallisation, this option was not explored as it would result in a truncated form of the protein, thus contradicting a major principle of this study, that of working with full length proteins, particularly in the case of AIP. The full length structure of AIP was however obtained through SAXS, whilst retaining the protein in solution. Although at present, it does not suggest an obvious explanation for the observed differences in binding, it provides valuable information on the intermolecular interactions present in the mutant forms that might affect protein stability and/or mobility. While the main aims of this study have been reached, the structure of protein complexes will be required to fully elucidate the structural role of these N-domain mutations in protein-protein interactions.

This study also describes a novel protocol for the expression of AhR in a bacterial expression system. The use of additives combined with a short induction period at low temperature (18°C), significantly enhanced the expression and reduced the extent of protein degradation.

## 5.3. On-going work and suggestions for future work

The results obtained in this project are encouraging and have opened the possibility for further investigation both from a structural and functional perspective.

### 5.3.1. Electron paramagnetic resonance (EPR) experiments

An experimental procedure that is currently underway is the use of electron paramagnetic resonance (EPR) to obtain additional structural information on the full length AIP protein. EPR experiments are being performed in collaboration with Dr Fraser Macmillan at the School of Chemistry, University of East Anglia.

This technique relies on the presence of accessible cysteine residues that can be spin labelled and their relative signals within a magnetic field used to characterise protein structure and dynamics (Smirnova and Smirnov, 2015). AIP has a total of 8 cysteine residues, 4 on the N-domain (C78, C90, C121 and C122) and 4 on the C-domain (C208, C238, C240, C254), of which C121 and C122 are the most accessible as they form part of the flexible loop of the $\beta$D-$\beta$E extension. For a more targeted site-directed spin labelling, $H_6$-AIP cysteine mutants were generated through site-directed mutagenesis. In each mutant, two cysteine residues were deliberately substituted by a glycine residue. The rationale behind the use of such mutants is to have one accessible cysteine reside on each domain of the protein that can be efficiently labelled. The introduction of mutations was confirmed by Sanger DNA sequencing and both $H_6$-AIP cysteine mutants were purified as described in Section 2.3.3, employing the use of 10% (v/v) glycerol and 5 mM DTT throughout the purification (Appendix A.6).

Preliminary continuous wave (cw) EPR experiments have already been carried out to assess the labelling efficiency of $H_6$-AIP and the two cysteine mutants, designated as C78G/C121G and C121G/C122G. The results obtained indicate that the wild type protein labels with the 3-(2-Iodoacetamido)-PROXYL (IAP) spin label. IAP is a nitroxide probe that labels cysteine residues. The C78GC121G mutant also generated a positive EPR intensity signal that is indicative of spin-labelled cysteine residue/s. However, when both cysteine residues at position 121 and 122 were mutated to glycine, there was no EPR signal, clearly suggesting that at least one of these two residues has to be retained for efficient labelling. Pulsed electron-electron double resonance

(PELDOR) experiments will be performed next to provide distance measurements between the spin-labelled cysteine residues (spin-spin distances). This would provide inter-domain distance measurements that will complement the structural information obtained through the SAXS generated model.

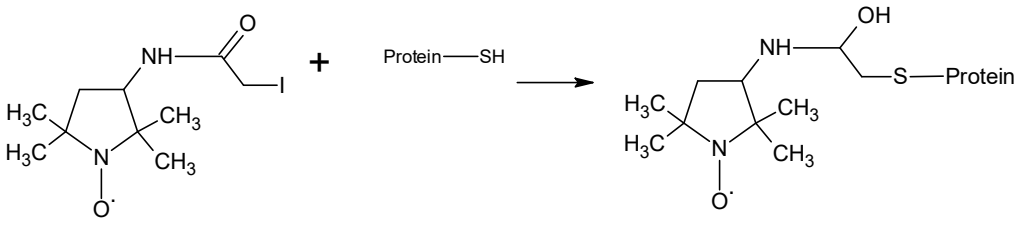



**Figure 5.2: IAP spin labelling of cysteine residues.** Only accessible cysteine residues in the protein can be efficiently labelled. The chemical illustration was drawn using ACD/ChemSketch version 2018.2.1.

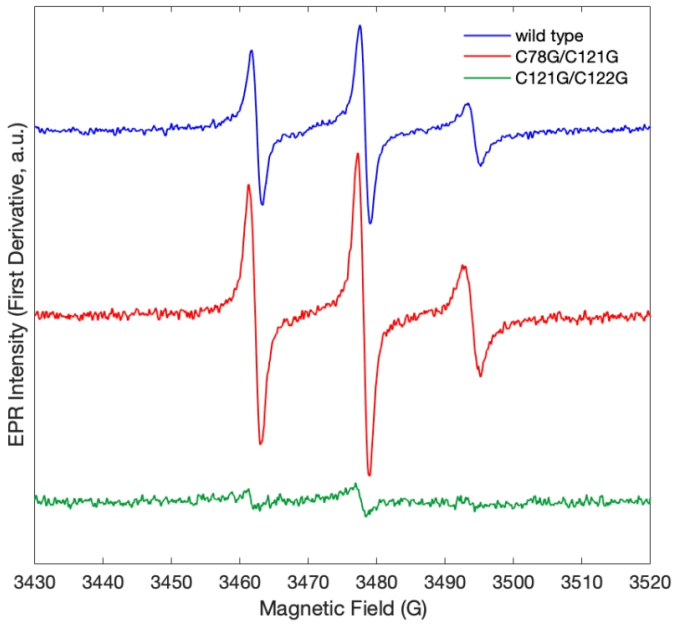



**Figure 5.3: Continuous-wave EPR.** AIP was spin labelled using IAP and cw-EPR spectra were recorded of wild-type protein (blue), and two selected mutants: C78G/C121G (red) and C121G/C122G (green). The EPR spectra were recorded using the following parameters: Centre Field: 3477G, Sweep width: 100G, Modulation amplitude: 0.1 mT, Modulation frequency: 100 kHz, Microwave frequency: 9.751 GHz, Microwave power: 0.05 mW (wt) or 0.025 mW (mutants), Scans: 128 (wild type) and 64 (mutants).

### 5.3.2. Small angle X-ray scattering of protein complexes

The use of SAXS will be extended to studying protein complexes particularly the complex obtained between $H_6$-AIP-Hsp90β$^{(513-724)}$. A preliminary experiment performed at B21, Diamond Light Source using a mixture of $H_6$-AIP-Hsp90β$^{(513-724)}$ confirmed their interaction and the formation of a protein complex.

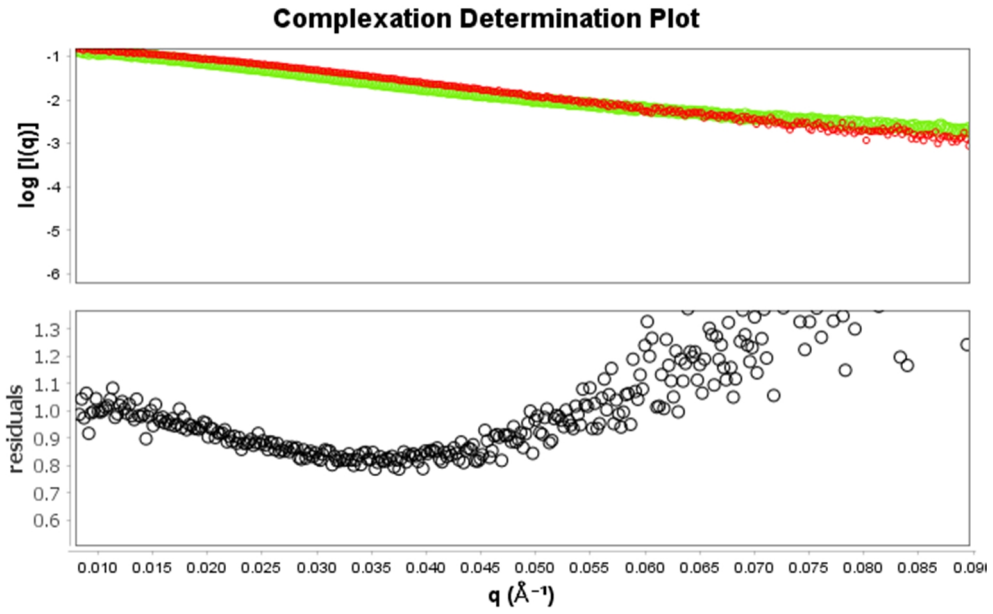



**Figure 5.4: Complexation Determination Plot.**

The upper panel shows the scattering intensity, log [I(q] as a function of the scattering angle, q (Å$^{-1}$). Complex formation is inferred by determining the best linear combination of the two individual monomeric SAXS curves of $H_6$-AIP and Hsp90β $^{(513-724)}$ (green) to explain the complex SAXS curve (red). A perfect fit means no complex formation. The non-linearity observed in the residuals plot indicates the presence of a complex.

While this data proved complex formation, further optimisation is required. In an ideal situation, the two protein components interact in such a way that the protein-protein complex exists as the major species in solution. In this case, however, SEC-SAXS of the mixture revealed that the complex was actually the minor species present and most of the protein still existed as unbound $H_6$-AIP and Hsp90β $^{(513-724)}$. Since the SAXS curve of the mixture is the summation of all the species present, it is necessarily

to optimise the condition/s of complex formation. This would enable the determination of a low resolution *ab initio* model of the $H_6$-AIP-Hsp90β$^{(513-724)}$ complex that would provide valuable information on the conformational change/s associated with binding interaction. This research project proved that $H_6$-AIP has a rather elongated structure with minimal contact between N- and C-domain. Nonetheless, N-domain mutations had a significant effect on binding affinity, which therefore suggests that $H_6$-AIP is likely to have a different conformational structure when bound to client partners. The model of the complex would thus provide valuable biological information and shed more light as to how N-domain mutations might exert their effect. Attempts to crystallise the Hsp90β $^{(513-724)}$ will also begin. If successful, the coupling of X-ray crystallography to SAXS would make it possible to model the complex through rigid-body modelling, in addition to *ab initio* modelling, providing atomistic detail on the nature of the interaction and binding contacts.

### 5.3.3. AhR purification and characterisation

After having established a successful protocol for the expression and stabilisation of AhR within a bacterial expression system, further work will be performed to purify and characterise the protein, whilst ensuring that the integrity and stability of the protein is retained. Once purified, the interaction between AIP and AhR will be investigated through ITC and the conditions for complex formation in solution optimised for structural analyses by SAXS.

AhR has recently attracted attention due to its conflicting role/s in cancer pathology both as a potential tumour promoter and suppressor (Section 1.9.3.2). The dependence of AhR on AIP has been documented in various independent studies that demonstrate how the cytoplasmic levels of AIP in pituitary adenomas are significantly correlated to the levels of AhR (Jaffrein-Rea *et al*., 2009) and ARNT (Heliovaara *et al*., 2009). The expression of AIP in hepatocytes is required to maintain high levels of functional AhR in the cytoplasm (Nukaya *et al*., 2010). Moreover, when AhR is activated by dioxin, the up-regulation of the cytochrome P450 *(Cyp1b1)* and AhR-repressor (*AhRR)* genes also requires the presence of functional AIP in the cytoplasm (Nukaya *et al*., 2010). The cytochrome P450 enzymes are needed to metabolise the harmful dioxin ligands while AhRR regulates the transcription of AhR-dependent gene

expression by negative feedback. Moreover, AhRR is recently being considered as a potential tumour suppressor, due to its ability to inhibit cellular proliferation, regulating inflammatory responses and increasing the apoptotic susceptibility of cancerous cells (Vogel and Haarmann-Stemman, 2017). All this evidence suggests that destabilisation of AIP protein or dysregulation of its protein expression by the presence of germline AIP mutation/s as is mostly observed in PA (Jaffrein-Rea *et al*., 2009) has a cascade effect that may collectively contribute to tumourigenesis.

While the *in-vivo* aspect of AhR biology and its related molecules is being extensively studied, the *in-vitro* protein characterisation and determination of the full length molecular structure of AhR would provide a complete view of the mode of action of this protein and pave the way towards a targeted therapy approach based on protein structure and/ or protein-protein interactions.

# References


Abbott, B.D., Schmid, J.E., Pitt, J.A., Buckalew, C.R., Wood, G.A., Held, J. and Diliberto, J. (1999). Adverse reproductive outcomes in the transgenic Ah receptor-deficient mouse. *Toxicol Appl Pharmacol*., **155** (1), 62-70.

ACD/ChemSketch, version 2018.2.1, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2019.

Agustsson, T.T., Baldvinsdottir, T., Jonasson, J.G., Olafsdottir, E., Steinthorsdottir, V., Sigurdsson, G., Thorsson, A.V., Carroll, P.V., Korbonits, M. and Benediktsson, R. (2015). The epidemiology of pituitary adenomas in Iceland, 1955-2012: a nationwide population-based study. *Eur J Endocrinol*., **173** (5), 655-664.

Allawi, H. T., and SantaLucia, J. (1997). Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry*, **36** (34), 10581-10594.

Arafah, B.M. and Nasrallah, M.P. (2001). Pituitary tumors: pathophysiology, clinical manifestations and management. *Endocr Relat Cancer*., **8** (4), 287-305.

Arkin, M.R., Tang, Y. and Wells, J.A. (2014). Small-molecule inhibitors of protein-protein interactions: progressing towards the reality. *Chem Biol*., **21** (9), 1102-1114.

Baba, T., Mimura, J., Gradin, K., Kuroiwa, A., Watanabe, Y., Matsuda, J., Sogawa, K. and Fujii-Kuriyama, Y. (2001). Structure and expression of the Ah receptor repressor gene. *J Biol Chem*., **276** (35), 33101-33110.

Barhoover, M.A., Hall, J.M., Greenlee, W.F. and Thomas, R.S. (2009). Aryl hydrocarbon receptor regulates cell cycle progression in human breast cancer cells via a functional interaction with cyclin-dependent kinase 4. *Mol Pharmacol*., **77** (2), 195-201.

Barlier, A., Vanbellinghen, J.F., Daly, A.F., Silvy, M., Jaffrain-Rea, M.L., Trouillas, J., Tamagno G... and Beckers, A. (2007). Mutations in the aryl hydrocarbon receptor interacting protein gene are not highly prevalent among subjects with sporadic pituitary adenomas. *J Clin Endocrinol Metab*., **92** (5), 1952-1955.

Barry, S., Carlsen, E., Marques, E., Stiles, C.E., Gadaleta, E., Berney, D.M., ... Korbonits, M. (2019). Tumor microenviroment defines the invasive phenotype of *AIP*-mutation-positive pituitary tumours. *Oncogene.* https://doi.org/10.1038/s41388-019-0779-5

Beard, M.B., Olsen, A.E., Jones, R.E., Erdogan, S., Houslay, M.D. and Bolger, G.B. (2000). UCR1 and UCR2 domains unique to the cAMP-specific phosphodiesterase family form a discrete module via electrostatic interactions. *J Biol Chem*., **275** (14), 10349-10358.

Beckers, A., Aaltonen, L., Daly, A.F. and Karhu, A. (2013). Familial isolated pituitary adenomas (FIPA) and the pituitary adenoma predisposition due to mutations in the aryl hydrocarbon receptor interacting protein (AIP) gene. *Endocrine Reviews*, **34** (2), 239-277.

BioSAXS GmbH. (2012). Medical and industrial applications of small angle X-ray scattering. Retrieved from: http://biosaxs.com/

Bizzi, M.F., Bolger, G.B., Korbonits, M. and Ribeiro-Oliveira Jr., A. (2019). Phosphodiesterases and cAMP pathway in Pituitary Diseases. *Frontiers in Endocrinology*, **10** (141).

Bizzi, M.F., Brant Pinheiro, S.V., Bolger, G.B., Riberio de Oliveria, J., Giannetti, V., Dang, M.N., Ribeiro-Oliveria Jr, A. and Korbonits, M. (2018). Reduced protein expression of the phosphodiesterases PDE4A4 and PDE4A8 in *AIP* mutation positive somatotroph adenomas. *Mol Cell Endocrinol*., **476**, 103-109.

Bohm, G., Muhr, R. and Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.*, **5** (3), 191-195.

Bolger, G.B., Bizzi, M.F., Brant Pinheiro, S.V., Trivellin, G., Smoot, L., Accavitti, M.A., Korbonits, M. and Ribeiro-Oliveira Jr., A. (2016). cAMP-specific PDE4 Phosphodiesterases and AIP in the Pathogenesis of Pituitary Tumors. *Endocr. Relat. Cancer*, **23** (5), 419-431.

Bolger, G.B., Peden, A., H., Steele, M.R., MacKenzie, C., McEwan, D.G., Wallace, D.A., Huston, E., Baillie, G.S. and Houslay, M.D. (2003). Attenuation of the activity of the cAMP-specific Phosphodiesterase PDE4A5 by interaction with the immunophilin XAP2. *J Biol Chem*., **278** (35), 33351-33363.

Bondos, S.E. and Bicknell, A. (2003). Detection and prevention of protein aggregation before, during and after purification. *Anal Biochem.,* **316**, 223-231.

Butler, R.A., Kelly, M.L., Powell, W.H., Hahn, M.E. and Van Beneden, R.J. (2001). An aryl hydrocarbon receptor (AHR) homologue from the soft-shell clam, *Mya aremaria*: evidence that invertebrate AHR homologous lack 2,3,7,8-tetrachlorodibenzo-p-dioxin and beta naphtaflavone binding. *Gene*, **278** (1-2), 223-234.

Cai, F., Hong, Y., Xu, J., Wu, Q., Reis, C., Yan, W., Wang, W. and Zhang, J. (2019). A novel mutation of aryl hydrocarbon receptor interaction protein gene associated with familial isolated pituitary adenoma mediates tumour invasion and growth hormone hypersecretion. *World Neurosurg*., **123**, 45-59.

Cai, W., Kramarova, T.V., Berg, P., Korbonits, M. and Pongratz, I. (2011). The immunophilin-like protein XAP2 is a negative regulator of estrogen signalling through interaction with estrogen receptor α. *PLoS One,* **6** (10), e25201.

Carson, M., Johnson, D.H., McDonald, H., Brouillette, C. and Delucas, J. (2007). His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr*., **63** (Pt 3), 295-301.

Carver, L.A. and Bradfield, C.A. (1997). Ligand-dependent interaction of the aryl hydrocarbon receptor with a novel immunophilin homolog *in vivo*. *J Biol Chem*., **272** (17), 11452-11456.

Carver, L.A. and Bradfield, C.A. (1997). Ligand-dependent interaction of the aryl hydrocarbon receptor with a novel immunophilin homolog *in vivo*. *J Biol Chem*., **272** (17), 11452-11456.

Cazabat, L., Bouligand, J., Salenave, S., Bernier, S.G., Parker, F., Young, J., GuiochonMantel, A. and Chanson, P. (2012). Germline AIP mutations in apparently sporadic pituitary adenomas: prevalence in a prospective single-centre cohort of 443 patients. *J Clin Endocrinol Metab*, **97** (4), 663-670.

Cazabat, L., Libè, R., Perlemoine, K., René-Corail, F., Burnichon, N., Glimenez-Roqueplo, A.P., Dupasquier-Fediaevsky, L., Bertagna, X... and Raffin-Sanson, M.L. (2007). Germline inactivating mutations of the *aryl hydrocarbon receptor-interacting protein* gene in a large cohort of sporadic acromegaly: mutations are found in a subset of young patients with macroadenomas. *Eur J Endocrinol.*, **157** (1), 1-8.

Chahal, H.S. Stals, K., Underlander, M., Balding, D.J., Thomas, M.G., Kumar, A.V... Korbonits, M. (2011). AIP mutation in pituitary adenomas in the 18th century and today. *N Engl J Med*., **364**, 43-50.
Chahal, H.S., Trivellin, G., Leontiou, C.A., Alband, N., Fowkes, R.C., Tahir, A., Igerja, S.C, Chapple, J.P., Jordan, S., Lupp, A., Schultz, S., Ansorge, O., Karavitaki, N., Carlsen, E., Wass, J.A., Grossman, A.B. and Korbonits, M. (2012). Somatostain analogs modulate AIP in somatotroph adenomas: the role of the Zac1 pathway. *J Clin Endorinol Metab*., **97** (8), 1411-1420.

Chapman-Smith, A, Lutwyche, J.K. and Whitelaw, M.L. (2004). Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix loop-helix PAS transcriptional regulators. *J Biol Chem*., **279** (7), 5353-5362.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Karpal, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogra*., **66** (Pt 1), 12-21.

Ciolino, H.P., Daschner, P.J. and Yeh, G.C. (1999). Dietary flavonols quercetin and kaempferol are ligands of the aryl hydrocarbon receptor that affect CYP1A1 transcription differentially. *Biochem J*., **340** (3), 715-722.

Conti, M., Richter, W., Mehats, C., Livera, G., Park, J.Y. and Jin, C. (2003). Cyclic AMP-specific PDE4 phosphodiesterases as critical components of cyclic AMP signalling. *J Biol Chem.*, **278**, 5493-5496.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76** (6), 2879-2886.
D'Andrea, L.D. and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem Sci.*, **28** (12), 655-662.

Daly, A., Vanbellinghen, J.F. and Beckers, A. (2006). Characteristics of familial isolated pituitary adenomas. *Expert Rev Endocrinol Metab*, **2** (6), 725-733.

Daly, A.F., Rixhon, M., Adam, C., Dempegioti, A., Tichomirowa, M.A. and Beckers, A. (2006). High prevalence of pituitary adenomas: a cross-sectional study in the province of Liege, Belgium. *J Clin Endocrinol Metab*., **91** (12), 4769-4775.

Daly, A.F., Tichomirowa, M.A., Petrossians, P., Heliövaara, E., Jaffrain-Rea, M.L., Barlier, A., Naves, L.A... and Beckers, A. (2010). Clinical characteristics and therapeutic responses in patients with germline AIP mutations and pituitary adenomas: An international collaborative study. *J Clin Endocrinol Metab.*, **95** (11), E373-E383.

Daly, A.F., Vanbellinghen, J.F., Khoo, S.K., Jaffrain-Rea, M.L., Naves, L.A., Guitelman, M.A., Emy, P... and Beckers, A. (2007). Aryl hydrocarbon receptor-interacting protein gene mutations in familial isolated pituitary adenomas: Analysis n 73 families. *J Clin Endocrinol Metab.*, **92** (5), 1891-1896.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeynik, J., Richardson, J.S. and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*., **35** (Web Server issue): W375-383.

De Oliveira, S.K., Hoffmesiter, M., Gambaryan, S., Muller-Esterl, W., Guimaraes, J.A. and FSmolenski, A.P. (2007). Phosphodiesterase 2A forms a complex with the co-chaperone XAP2 and regulates nuclear translocation of the aryl hydrocarbon receptor. *J. Biol Chem*., **282**, 13656-13663.

Denison, M.S., Fisher, J.M. and Whitlock, J.P. (1988). The DNA recognition site for the dioxin-Ah receptor complex. Nucleotide sequence and functional analysis. *J Biol Chem*., **263** (33), 17221-17224.

Döhr, O., Li, W., Donat, S., Vogel, C., and Abel, J. (1996). Aryl hydrocarbon receptor mRNA levels in different tissues of 2,3,7,8-tetrachlorodibenzo-p-dioxin-responsive and nonresponsive mice. *Adv. Exp. Med. Biol*. **387**, 447–459.

Duchartre, Y., Kim, Y.M. and Kahn, M. (2016). The Wnt Signaling Pathway in Cancer. *Crit Rev Oncol Hematol*., **99**, 141-149.

EMBL, Heidelberg. (2019). Protein Purification: Extraction and Clarification, Choice of Lysis Buffer and Additives. Retrieved from: https://www.embl.de/pepcore/pepcore_services/protein_purification/extraction_clarification/lysis_buffer_additives/

Familial isolated pituitary adenoma (FIPA) patients. (2011). *AIP gene mutations.* Retrieved from: http://www.fipapatients.org/aip/

Fan, Y., Boivin, G.P., Knudsen, E.S., Nebert, D.W., Xia, Y and Puga, A. (2010). The aryl hydrocarbon receptor functions as a tumour suppressor of liver carcinogenesis. *Cancer Res*., **70** (1), 212-220.

Faust D., Vondráček J., Krčmář P., Šmerdová L., Procházková J., Hrubá E., Hulinková P., Kaina B., Dietrich C., Machala M. AHR-mediated changes in global gene expression in rat liver progenitor cells. *Arch. Toxicol*., **87** (4), 681–698.

Feigin, L.A. and Svergun, D.I. (1987). Structure analysis by small angle X-ray and neuron scattering. New York: Plenum Press.

Feng, S., Cao, Z and Wang, X. (2013). Role of aryl hydrocarbon receptor in cancer. *Biochim Biophys Acta*., **1836** (2), 197-210.

Fernandez, A., Karavitaki, N. and Wass, J.A. (2009). Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clin Endocrinol (Oxf)*, **72**, 377-382.

Fernandez-Salguero, P., Pineau, T., Hilbert, D.M., McPhail, T., Lee, S.S., Kimura, S., Nebert, D.W., Rudikoff, S., Ward, J.M. and Gonzalez, F.J. (1995). Immune system impairment and hepatic fibrosis in mice lacking the dioxin binding Ah receptor. *Science*, **268** (5211), 722-726.

Fernandez-Salguero, P.M., Ward, J.M., Sundberg, J.P. and Gonzalez, F.J. (1997). Lesions of aryl hydrocarbon receptor-deficient mice. *Vet Pathol*., **34** (6), 605-614.

Fontana, E. and Gaillard, R. (2009). Epidemiology of pituitary adenoma: results of the first Swiss study. *Rev Med Suisse*, **5**, 2172-2174.

Formosa, R. (2012). *Genetic predisposition and functional analysis of pituitary adenomas*. (Doctoral dissertation). Faculty of Medicine and Surgery, University of Malta, Malta.

Formosa, R. and Vassallo, J. (2014). cAMP signalling in the normal and tumorigenic pituitary gland. *Mol Cell Endocrinol.,* **392** (1-2), 37–50.

Formosa, R., Borg, J. and Vassallo, J. (2017). Aryl hydrocarbon receptor (AHR) is a potential tumour suppressor in pituitary adenomas. *Endocr Relat Cancer*., **24** (8), 445-447.

Formosa, R., Farrugia, C., Xuereb-Annastasi, A., Korbonits, M., Vassallo, J. (2010). Aryl hydrocarbon receptor-interacting protein: mutational analysis and functional validation in primary pituitary cell cultures. *Endocrine Abstracts*, **22**, P436.
Formosa, R., Gruppetta, M., Falzon, S., Santillo, G., DeGaetano, J., Xuereb-Anastasi, A. and Vassallo, J. (2012). Expression and Clinical Significance of Wnt Players and Survivin in Pituitary Tumours. *Endocr Pathol*., **23**, 123-131.

Formosa, R., Xuereb-Anastasi, A. and Vassallo, J. (2013). Aip regulates cAMP signalling and GH secretion in GH3 cells. *Endocr Relat Cancer*, **20** (4), 495-505.

Fukunaga, B.N., Probst, M.R., Reisz-Porszasz, S. and Hankinson, O. (1995). Identification of functional domains of the aryl hydrocarbon receptor. *J Biol Chem*., **270** (49), 2927-29278.

Garstka, M.A., Fish, A., Celie, P. H. N., Joosten, R.P., Janseen, G. M. C., Berlin, I., ... Neefjes, J. (2015). The first step of peptide selection in antigen presentation by MHC class I molecules. *PNAS*, **112** (5), 1505-1510.

Gaston-Massuet, C., Andoniadou, C.L., Signore, M., Jayakody, S.A., Charolidi, N., Kyeyune, R., Vernay, B., Jacques, T.S., Taketo, M.M., Le Tissier, P., Dattani, M.T. and Martinez-Barbera, J.P. (2011). Increased Wingless (Wnt) signaling in pituitary progenitor/stem cells gives rise to pituitary tumours in mice and humans. *Proc.Natl. Acad. Sci. USA*., **108** (28), 11482-11487.

GE Healthcare. (2012). Biacore™ Assay Handbook. Retrieved from: https://www.gelifesciences.co.jp/contact/pdf/BiacoreAssayHandbook.pdf

Georgitsi, M., Raitila, A., Karhu, A., Tuppurainen, K., Makinen, M., Vierimaa, O., Paschke, R... and Aaltonen, L. A. (2007). Molecular diagnosis of pituitary adenoma predisposition caused by aryl hydrocarbon receptor-interacting protein gene mutations. *PNAS*, **104** (10), 4101-4105.

Gherardi, E., Sandin, S., Petoukhov, M.V, Finch, J., Youles, M.E., Ofverstedt, L.G., Miguel, R.N., Blundell, T.L., Vande Woude, G.F., Skoglund, U. and Svergun, D.I. (2006). *Proc Natl Acad Sci U S A*. **103** (11), 4046-4051.

Gioutlakis, A., Klapa, M.L., Moschonas, N.K. (2017). PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLos ONE*, **12** (10): e0186039.

Goodsell, D. (2008). Hsp90: Molecule of the month. Retrieved from: https://pdb101.rcsb.org/motm/108

Greenfield, N.J. (2004). Circular dichroism analysis for protein-protein interactions. *Methods Mol Biol*., **261**, 55-78.

Greenman, Y. and Stern, N. (2009). Non-functioning pituitary adenomas. *Best. Pract. Res. Clin. Endocrinol. Metab*., **23** (5), 625-638.

Gregoret, L.M., Rader, S.D., Fletterick, R.J. and Cohen, F.E. (1991). Hydrogen bonds involving sulfur atoms in proteins. *Proteins*, **9** (2), 99-107.

Gruppetta, M., Mercieca, C. and Vassallo, J. (2013). Prevalence and incidence of pituitary adenomas: a population based study in Malta. *Pituitary*, **16**, 545-553.

Guex, N. And Peitsch, M.C. (1997). Swiss-Model and the Swiss-Pdb Viewer: An environment for comparative protein modelling. *Electrophoresis*, **18**, 2714-2723.

Guyot, E., Chevallier, A., Barouki, R. and Coumoul, X. (2013). The AhR twist: ligand dependent AhR signalling and pharmaco-toxicological implications. *Drug Discov Today*, **18** (9-10), 479-486.

Hahn, M.E., Allan, L.L. and Sherr, D.H. (2009). Regulation of constitutive and inducible AHR signaling: complex interactions involving the AHR repressor. *Biochem Pharmacol*., **77** (4), 485-497.

Heaney, A.P. and Melmed, S. (2004). Molecular targets in pituitary tumours. *Nat Rev Cancer*, **4** (4), 285-295.

Heliövaara, E., Raitila, A., Launonen, V., Paetau, A., Arola, J., Lehtonen, H., Sane, T., Weil, R.J.,Vierimaa, O., Salmela, P., Tuppurainen, K., Makinen, M., Aaltonen, L.A. and Karhu, A. (2009). The expression of AIP-related molecules in elucidation of cellular pathways in pituitary adenomas. *Am J Pathol.*, **175** (6), 2501-2517.

Hernandez-Ramirez, L.C. Trivellin, G. and Stratakis, C.A. (2017). Cyclic 3', 5-adenosine monophosphate (cAMP) signaling in the anterior pituitary gland in health and disease. *Mol Cell Endocrinol*., **463**, 72-86.

Hibberts, N.A., Simpson, D.J., Bicknell, J.E., Broome, J.C., Hoban, P.R., Clayton, R.N. and Farrell, W.E. (1999). Analysis of cyclin D1 (CCND1) allelic imbalance and overepxression in sporadic human pituitary tumours. *Clin Cancer Res*., **5** (8), 2133-2139.

Ho, B. K. and Brasseur, R. (2005). The Ramachandran plots of glycine and pre-proline. *BMC Structural Biology*, **5** (14).

Houslay, M.D. and Adams, D.R. (2003). PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. *Biochem J*., **370** (Pt 1), 1-18.

Igerja, S., Chahal, H.S., King, P., Bolger G.B., Srirangalingam, U., Guasti, L., Chapple, J.P... and International FIPA consortium. (2010). Characterisation of aryl hydrocarbon receptor interacting protein (AIP) mutations in familial isolated pituitary adenoma families. *Hum Mutat.*, **31** (8), 950-960.

Ikuta, T., Tachibana, T., Watanabe, J., Yoshida, M., Yoneda, Y. and Kawajiri, K. (2000). Nucleoctytoplasmic shuttling of the aryl hydrocarbon receptor. *J Biochem*., **127** (3), 503-509.

Ingram, L.O. and Buttke, T.M. (1985). Effects of Alcohols on Micro-Organisms. *Advances in Microbial Physiology*, **25**, 253-300.

Iwaoka, M. and Isozumi, N. (2012). Hypervalent nonbonded interactions of a divalent sulfur atom. Implications in protein architecture and the functions. *Molecules*, **17** (6), 7266-7283.

Jaffrain-Rea, M.L. and Beckers, A. (2013). The role of aryl hydrocarbon receptor (AhR) and AhR-Interacting Protein (AIP) in the pathogenesis of pituitary adenomas. In Hayat, A.M (Ed.), Tumours of the central nervous system (Vol. 10). Springer Science and Buisness Media: Dordrecht.

Jaffrain-Rea, M.L., Angelini, M., Gargano, D., Tichomirowa, M.A., Daly, A.F., Vanbellinghen, J.F., D'Innocenzo, E… and Beckers, E. (2009). Expression of aryl hydrocarbon receptor (AHR) and AHR-interacting protein in pituitary adenomas: pathological and clinical implications. *Endocrine-Related Cancer*, **16**, 1029-1043.

Jain, S., Dolwick, K.M., Schmidt, J.V. and Bradfield, C.A. (1994). Potent transactivation domains of the Ah receptor and the Ah receptor nuclear translocator map to their carboxyl termini. *J Biol Chem*., **269** (50), 31518-31524.

Jeong, J.Y., Yim, H.S., Ryu, J.Y., Lee, H.S., Lee, J.H., Seen, D.S. and Kang, S.G. (2012). One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl Environ Microbiol.,* **78** (15), 5440-5443.

Jin, U.H., Lee, S.O., Pfent, C. and Safe, S. (2014). The aryl hydrocarbon receptor ligand omeprazole inhibits breast cancer cell invasion and metastasis. *BMC Cancer*, **14**, 498.

Kang, C.B., Hong, Y., Dhe-Paganon, S. And Yoon, H.S. (2008). FKBP family proteins: immunophilins with versatile biological functions. *Neurosignals*, **16** (4), 318-325.

Karhu, A. And Aaltonen, L.A. (2007). Susceptibility to pituitary neoplasia related to *MEN-1*, *CDKN1B* and *AIP* mutations: an update. *Hum Mol Genet.*, **16** (1), 73-79.

Kawajiri, K., Kobayashi, Y., Ohtake, F., Ikuta, T., Matsushima, Y., Mimura, J., Petterson, S., Pollenz, R.S., Sakaki, T. and Hirokawa, T. (2009). Aryl hydrocarbon receptor suppresses intestinal carcinogenesis in $Apc^{min/+}$ mice with natural ligands. *Proc Natl Acad Sci USA*., **106** (32), 13481-13486.

Kazlauskas, A., Poellinger, L. And Pongratz, I. (2002). Two distinct regions of the immunophilin-like protein XAP2 regulate dioxin receptor function and interaction with HSP90. *J Biol Chem.*, **277** (14), 11795-11801.

Kelly, S.M., Jess, T.J. and Prince, N.C. (2005). How to study proteins by circular dichroism. *Biochim Biophys Acta*., **1751** (2), 119-139.

Klapa, M.I., Tsafou, K., Theodoridis, E., Tsakalids, A. and Moschonas, N.K. (2013). Reconstruction of the experimentally supported human protein interactome: what can we learn? *BMC systems biology*, **7** (96).

Ko, H.P., Okino, S.T., Ma, Q. and Whitlock, J.P. (1997). Transactivation domains facilitate promoter occupancy for the dioxin-inducible CYP1A1 gene *in vivo*. *Mol Cell Biol*., **17** (7), 3497-3507.

Kolluri, S.K., Weiss, C., Koff, A. and Gottlicher, M. (1999). p27(Kip1) induction and inhibition of proliferation by the intracellular Ah receptor in developing thymus and hepatoma cells. *Genes Dev*., **13** (13), 1742-1753.

Komiya, Y. and Habes, R. (2008). Wnt signal transduction pathways. *Organogenesis*, **4** (2), 68-75.

Korbonits, M., Storr, H. and Kumart, A.V. (2012). Familial pituitary adenomas- who should be tested for *AIP* mutations. *Clin Endocrinol.*, **77**, (3), 351-356.

Kudo, I., Hosaka, M., Haga, A., Tsuji, N., Nagata, Y., Okada, H., Fukuda, K., Kakizaki, Y., Okamoto, T., Grave, E. and Itoh, H. (2018). The regulation mechanisms of AhR by molecular chaperone complex. *J. Biochem.*, **163** (3), 223–232.

Kuo, D., Nie, M. and Courey, A.J. (2014). SUMO as a solubility tag and in vivo cleavage of SUMO fusion proteins with Ulp1. *Methods Mol Biol*., **1177**, 71-80.

Kuzhandaivelu, N., Cong, Y.S., Inouye, C., Yang, W.M. and Seto, E. (1996). XAP2, a novel hepatitis B virus X-associated protein that inhibits X transactivation. *Nucleic Acids Res*., **24** (23), 4741-4750.

Laenger, A., Lang-Rollin, I., Kozany, C., Zschocke, J., Zimmermann, N., Ruegg, J., Holsboer, F., Hausch, F. and Rein, T. (2009). XAP2 inhibits glucocorticoid receptor activity in mammalian cells. *FEBS Letters*, **583** (9), 1493-1498.

Lahvis, G.P., Pyzakski, R.W., Glover, E., Pitot, H.C., McElwee, M.K. and Bradfield, C.A. (2005). The aryl hydrocarbon receptor is required for developmental closure of the ductus venosus in the neonatal mouse. *Mol Pharmacol*., **67** (3), 714-720.

Lania A, Mantovani G, Spada A. (2012). cAMP pathway and pituitary tumorigenesis. *Ann Endocrinol.*, **73** (2), 73-75.

Larigot, L., Juricek, Dairou, J. and Coumoul, X. (2018). AhR signalling pathways and regulatory functions. *Biochimie Open*, **7**, 1-9.

Lau, S.Y., Taneja, A.K. and Hodges, R.S. (1984). Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilisation and formation of two-stranded alpha-helical coiled-coils. *J Biol Chem*., **259** (21), 13253-13261.

Leibly, D. J., Nguyen, T. N., Kao, L.T., Hewitt, S.N., Barrett, L. and Van Voorhis, W.C. (2012). Stabilising additives added during cell lysis aid in the solubilisation of recombinant proteins. *PLOS ONE*, **7** (12), e52482.

Leonard, P., Hearty, S., Ma, H. and O'Kennedy, R. (2017). Measuring protein-protein interactions using Biacore. *Methods Mol Biol*., **1485**, 339-354.

Leontiou, C.A., Gueorguiev, M., van der Spuy, J., Quinton, R., Lolli, F., Hassan, S., Chahal, H.S... and Korbonits, M. (2008). The role of the aryl hydrocarbon receptor-interacting protein gene in familial and sporadic pituitary adenomas. *J Clin Endocrinol Metab*, **93** (6), 2390-2401.

Li, J., Zoldak, G., Kriehuber, T., Soroka, J., Schmid F.X and Richter, K. (2013). Unique proline-rich domain regulates the chaperone function of AIPL1. *Biochemistry*, **52**, 2089-2096.

Lin, B.C., Sullivan, R., Lee, Y., Moran, S., Glover, E. and Bradfield,C.A. (2007). Deletion of the aryl hydrocarbon receptor-associated protein 9 leads to cardiac malformation and embryonic lethality. *J Biol Chem.*, **282** (49), 35924-35932.

Linnert, M., Haupt, K., Lin, Y.J., Kissing, S., Paschke, A.K., Fischer, G., Weiward, G. and Lücke, C. (2012). NMR assignments of the FKBP-type PPIase domain of the human aryl hydrocarbon receptor-interacting protein (AIP). *Biomol NMR Assign*, **6**, 209-212.

Linnert, M., Lin, Y.J., Manns, A., Haupt, K., Paschke, A.K., Fischer, G., Weiward, M. and Lücke, C. (2013). The FKBP-type domain of the human aryl hydrocarbon receptor-interacting protein reveals an unusual HSP90 interaction. *Biochemistry*, **52** (12), 2097-2107.

Lloyd, C. And Grossman, A. (2014). The *AIP* (aryl hydrocarbon receptor-interacting protein) gene and its relation to the pathogenesis of pituitary adenomas. *Endocrine*, **46** (3), 387-396.

Ma, Q. and Baldwin, K.T. (2000). 2,3,7,8-tetrachlorodibenzo-p-dioxin-indunced degradation of aryl hydrocarbon receptor (AhR) by the ubiquitin-proteasome pathway. Role of the transcription activation and DNA binding of AhR. *J Biol Chem*., **275** (12), 8432-8438.

Mackenzie, K.F., Topping, E.C., Bugaj-Gaweda, B., Deng, C., Cheung, Y.F., Olsen, A.E....Bolger, G. B. (2008). Human PDE4A8, a novel brain-expressed PDE4 cAMP-specific phosphodiesterase that has undergone rapid evolutionary change. *Biochem J.,* **411** (2), 361-369.

Maghathe, T., Miller, W.K., Mugge, L., Mansour, T.R. and Schroeder, J. (2018). Immunotherapy and potential molecular targets for the treatment of pituitary adenomas resistant to standard therapy: a critical review of potential therapeutic targets and current developments. *J.Neurosurg Sci*., **7**.

Marco, A., Vigh, L., Diamant, S. and Golobinoff, P. (2005). Native folding of aggregation prone recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol- overexpressed molecular chaperones. *Cell Stress Chaperones*, **10** (4), 329-339.

Malhotra, A. (2009). Tagging for protein expression. *Methods in Enzymology*, **463**, 239-258.

Mertens, H.D. Svergun, D. (2010). Structural characterisation of proteins and complexes using small angle X ray solution scattering. *J Struct Biol*., **172** (1), 128-141.

Michibata, H., Yanaka, N., Kanoh, Y., Okumura, K. and Omori, K. (2001). Human Ca2+/ calmodulin-dependent phosphodiesterase PDE1A: novel splice variants, their specific expression, genomic organization, and chromosomal localisation. *Biochim Biophys Acta*., **1517** (2), 278-287.

Moennikes, O., Loeppen, S., Buchmann, A., Andersson, P., Ittrich, C., Poellinger, L. and Schwarz, M. (2004). A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice. *Cancer Res*., **64** (14), 4707-4710.

Morales, R., Berna, A., Carpentier, P., Contreras-Martel, C., Renault, F., Nicodeme, M...Chabriere, E. (2006). Serendipitous discovery and X-ray structure of a human phosphate binding apolipoprotein. *Structure*, **14** (3), 601-609.

Morgan, R.M.L., Hernández-Ramírez, L.C., Trivellin, G., Zhou, L., Roe, M., Korbonits, M. and Prodromou, C. (2012). Structure of the TPR domain of AIP: lack of client protein interaction with the C-terminal α-7 helix of the TPR domain of AIP is sufficient for pituitary adenoma predisposition. *PLoS ONE*, **7** (12), e53339.

Mulero-Navarro, S., Pozo-Guisado, E., Pérez-Mancera, P.A., Alvarez-Barrientos, A., Catalina-Fernández, I., Hernández-Nieto, E., Sáenz Santa-Maria, J., Martinez, N., Rojas, J.M., Sánchez-Garcia, I. and Fernandez-Salguero, P.M. (2005). Immortalized mouse mammary fibroblasts lacking dioxin receptor having impaired tumorigenicity in a subcutaneous mouse xenograft model. *J Biol Chem*., **280** (31), 28731-28741.

Murphy, M., Jason-Moller, L., Bruno, J. (2006). Using Biacore to measure the binding kinetics of an antibody-antigen interaction. *Curr Protoc Protein Sci*., Chapter 19: Unit 19.14.

Murray, I.A, Patterson, A. D.and Perdew, G. H. (2014) Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nature Reviews Cancer*, **14**, 801–814.

National Library of Medicine. (2019). Genetics Home Reference. Bethesada (MD): The Library, 2013. Retrieved from: https://ghr.nlm.nih.gov/gene/AIP

Nguyen, L.P. and Bradfield, C.A. (2008). The search for endogenous activators of the aryl hydrocarbon receptor. *Chem Res Toxicol*., **21** (1), 102-116.

Nukaya, M., Lin, B.C., Glover, E., Moran, S.M., Kennedy, G.D. and Bradfield, C.A. (2010). The aryl hydrocarbon receptor interacting protein (AIP) is required for dioxin-induced heptatoxicity but not for the induction of the Cyp1a1 and Cyp1a2 genes. *J Biol Chem*., **285** (46), 35599-35605.

Ojalvo, L.S., Whittaker, C.A., Condeelis, J.S. and Pollard, J.W. (2010). Gene expression analysis of macrophages hat facilitate tumour invasion supports a role for Wnt-signalling in mediating their activity in primary mammary tumours. *J. Immunol*., **184** (2), 702-712.

Oriola, J., Lucas, T., Halperin, I., Mora, M., Perales, M.J., Alvarez-Escolá, C., Novoa Paz, M... and Puig-Domingo, M. (2013). Germline mutations of *AIP* gene in somatotropinomas resistant to somatostatin analogues. *Eur J Endocrinol*., **168** (1), 9-13.

Ozfirat, Z. and Korbonits, M. (2010). AIP gene and familial isolated pituitary adenomas. *Mol Cell Endocrinol*., **326**, (1-2), 71-79.

Pal, D. and Chakrabarti, P. (2001). Non-hydrogen bond interactions involving the methionine sulfur atom. *J Biomol Struct Dyn*., **19** (1), 115-128.

Pan, L. and Aller, S.G. (2015). Tools and procedures for visualisation of proteins and other biomolecules. *Curr Protoc Mol Biol*., **110**, (19-2), 1-47.

Pearl, L.H. (2016). Review: The HSP90 molecular chaperone-an enigmatic ATPase. *Biopolymers*, **105** (8), 594-607.

Persani, L., Borgato, S., Lania, A.,Filopanti, M., Mantovani, G., Conti, M., Spada, A. (2001). Relevant cAMP-Specific Phosphodiesterase Isoforms in Human Pituitary: Effect of Gsα Mutations. *J Clin Endocrinol Metab*., **86** (8), 3795-3800.

Petoukhov, M.V. and D.I. Svergun. (2006). Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. *Eur Biophys. J*., **35** (7), 567-576.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D., Konarev, P.V. and Svergun, D.I. (2012). *J Appl Crystallogr*., **45** (Pt 2), 342-350.

Petoukhov, M.V., Svergun, D.I., Konarev, P.V., Ravasio, S., van den Heuvel, R.H., Curti, B. and Vanoni, M.A. (2003). Quaternary structure of *Azosspirillum brasilense* NADPH-dependent glutamate synthase in solution as revealed by synchrotron radiation X-ray scattering. *J Biol Chem*., **278** (32), 29933-29939.

Petrulis, J. R., Kusnadi, A., Ramadoss, P., Hollingshead, B. and Perdew, G. H. (2003). The hsp90 co-chaperone XAP2 alters importin beta recognition of the bipartite nuclear localisation signal of the Ah receptor and represses transcriptional activity. *J Biol Chem.,* **278** (4), 2677–2685.

Petrulis, J.R. and Perdew, G.H. (2002). The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chem Biol Interact.*, **141** (1-2), 25-40.

Peverelli, E., Ermetici, F., Filopanti, M., Eli, F.M., Ronchi, C.L., Mantovani, G., Ferrero, S., Bosari, S., Beck-Peccoz, P., Lania, A. and Spada, A. (2009). Analysis of genetic variants of phosphodiesterase 11A in acromegalic patients. *Eur J Endocrinol*., **161** (5), 687-694.

Peverelli, E., Mantovani, G., Lania, A.G. and Spada, A. (2013). cAMP in the pituitary: an old messenger for multiple signals. *J. Mol. Endocrinol*., **52** (1), 67-77.

Peverelli, E., Mantovani, G., Lania, A.G. and Spada, A. (2014). cAMP in the pituitary: an old messenger for multiple signals. *J Mol Endocrinol*., **52** (1), 67-77.

Pierce, M.M., Raman, C.S. and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods*, **19** (2), 213-221.

Polakis, P. (2000). Wnt signaling and cancer. *Genes and Dev*., **14**, 1837-.1851.

Polakis, P. (2012). Wnt Signaling in Cancer. *Cold Spring Harb Perspect Biol*., **4** (5), a008052.

Prasad, S., Prashant, B., Khadatare, B. and Roy, I. (2011). Effect of chemical chaperones in improving the solubility of recombinant proteins in *Escherichia coli*. *Appl Environ Microbiol*., **77** (13), 4603-4609.

Procházkova J., Kabátkova M., Bryja V., Umannová L., Bernatík O., Kozubík A., Machala M., Vondráček J. (2011). The interplay of the aryl hydrocarbon receptor and β-catenin alters both AHR-dependent transcription and Wnt/β-catenin signaling in liver progenitors. *Toxicol*., **122** (2), 349–360.

Prouty, W.F. and Goldberg, A.L. (1972). Effects of Protease Inhibitors on Protein Breakdown in *Escherichia coli. J Biol Chem.*, **247** (10), 3341-3352.

Pryor, K.D. and Leiting, B. (1997). High-level expression of soluble protein in Escherichia coli using a $His_6$-tag and maltose-binding-protein double affinity fusion system. *Protein Expr. Purif*., **10**, 309-319.

Puga, A., Barnes, S.J. Dalton, T.P., Chang, C.Y., Knudsen E. and Maier, M.A. (2000). Aromatic hydrocarbon receptor interaction with the retinoblastoma protein potentiates repression of E2F-dependent transcription and cell cycle arrest. *J. Biol Chem*., **275** (4), 2943-2950.

Putnam, C. D. (2016). Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. *J App Crystal.*, **49** (5), 1412-1419.

Putnam, C. D., Hammel, M., Hura, G. L. and Tainer, J. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*., **40** (3), 191-285.

Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007). X-Ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution, *Q Rev Biophys*., **40** (3), 191-285.

Qin, H. and Powell-Coffman, J.A. (2004). The *Caenorhabditis elegans* aryl hydrocarbon receptor, AHR-1, regulates neuronal development. *Dev Biol*., **270** (1), 64-75.

Raitila, A., Lehtonen, H.J., Arola, J., Heliövaara, E., Ahlsten, M., Georgitsi, M., Jalanko, A., Paetau, A., Aaltonen, L.A. and Karhu, A. (2010). Mice with inactivation of aryl hydrocarbon receptor-interacting protein (*Aip*) display complete penetrance of pituitary adenomas with aberrant ARNT expression. *Am J Pathol*., **177** (4), 1969-1976.

Ramadoss, P., Petrulis, J. R., Hollingshead, B. D., Kusnadi, A. and Perdew, G. H. (2004). Divergent roles of hepatitis B virus X-associated protein 2 (XAP2) in human versus mouse Ah receptor complexes. *Biochemistry*, **43** (3), 700–709.

Ramanathan, S. And Jagannathan N. (2014). Tumor associated macrophage: A review on the phenotypes, traits and functions. *Iran J Cancer Prev*., **7** (1), 1-8.

Rambo, P.R. (2019a). ScÅtter Version 3.0. Supported by Diamond Light Source, Didcot, Oxfordhire, UK.

Rambo, P.R.. (2019b). BIOISIS Tutorial. Retrieved from: http://www.bioisis.net/tutorial
Rambo, R.P. and Tainer, J.A. (2011). Characterizing Flexible and Instrinsically Unstructured Biological Macromolecules by SAS using the Porod-Debye Law. *Biopolymers*, **95** (8) 559-571.

Rasband, W. (2019). ImageJ v1.52a. National Institute of Health Sciences, USA.

Reya, T. and Clevers, H. (2005). Wnt signalling in stem cells and cancer. *Nature*, **434** (7035), 843-850.

Robert, X. And Gouet, P. (2012). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*., **42** (1) 320-324.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*., **5** (172).

Routzahn, K.M. and Waugh, D.S. (2002). Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. *J. Struct. Funct. Genomics*, **2**, 83-92.

Salvatori, R., Daly, A.F., Quinones-Hinojosa, A., Thiry A. and Beckers, A. (2014). A clinically novel *AIP* mutation in a patient with a very large, apparently sporadic somatotrope adenoma. *Endocrinol Diabetes Metab Case Rep*. 2014; 2014:140048.

Sambrook, J., & Russell, D.W. (2001). *Molecular Cloning. A Laboratory Manual* (3rd ed.). New York: Cold Springer Harbor Laboratory Press: Cold Spring Harbor.

Sasaki-Kudoh, E., Kodo, I., Kakizaki, Y., Hosaka, M., Ikeda, S., Uemura, S., Grave, E., Togashi, S., Sugawara, T., Shimizu, H. and Itoh, H. (2018). Cisplatin Inhibits AhR Activation. *Am J Mol Biol*., **8** (1), 69-82.

Schmidt, J.V., Su, G.H., Reddy, J.K., Simon, M.C. and Bradfield, C.A. (1996). Characterisation of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proc NatI Acad Sci. USA*, **93** (13), 6731-6736.

Schneider, A.J., Branam, A.M. and Peterson, R.E. (2014). Intersection of AHR and Wnt Signaling in Development, Health, and Disease. *Int J Mol Sci*., **15** (10), 17852-17885.

Schulte, K.W., Green, E., Wilz A., Platten, M. and Daumke, O. (2017). Structural basis for aryl hydrocarbon receptor mediated gene activation. *Structure,* **25** (7), 1025-1033.

Semba, S., Han, S.Y., Ikeda, H. and Horii, A. (2001). Frequent nuclear accumulation of beta-catenin in pituitary adenoma. *Cancer*, **19** (1), 42-48.

Seok, S.H., Lee, W., Jiang, L., Molugu, K., Zheng, A., Li, Y., Park, S., Bradfield C.A and Xing Y. (2016). Structural hierarchy controlling dimerisation and target DNA recognition in the AHR transcriptional complex. *Proc Natl Acad Sci U.S.A*., **114** (21), 5431-5436.

Seok, S.H., Ma, Z.X., Feltenberger, J.B., Chen, H., Scarlett, C., Lin, Z, Satyshur, K.A., Cortopassi, M., Jefcoate, C.R., Ge. Y., Tang, W., Bradfield, C.A. and Xing, Y. (2018). Trace derivatives of kynurenine potently activate the aryl hydrocarbon receptor (AHR). *J Biol Chem*., **293** (6), 1994-2005.

Shepherd, N.E., Hoang, H.N., Abbenante, G. and Fairlie, D. P. (2005). Single turn peptide alpha helices with exceptional stability in water. *J Am Chem Soc*., **127** (9), 2974-2983.

Shigapova, N., Török, Z., Balogh, G., Goloubinoff, P., Vigh, L. and Horváth, I. (2005). Membrane fluidization triggers membrane remodelling which affects the thermotolerance in *Escherichia coli*. *Biochem Biophys Res Commun*., **328** (4), 1216-1223.

Shtutman, M., Zhurinsky, J., Simcha, I., Albanese, C., D'Amico, M., Pestell. R., Ben-Ze'ev, A. (1999). The cyclin D1 gene is a target of the beta-catenin/LEF-1 pathway. *Proc Natl Acad Sci USA*., **96** (10), 5522-5527.

Smirnova, T. I. and Smirnov, A.I. (2015). Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Part B. In Qin, P.,Z and Warncke,K. (Eds), *Methods Enzymol.*, 564, 219-258.

Stanford, E.A., Wang, Z., Novikov, O., Mulas, F., Bollag, E.L., Monti, S., Smith, B. W., Seldin, D.C., Murphy, G.J. & Sherr, D.H. (2016). The role of the aryl hydrocarbon receptor in the development of cells with the molecular and functional characteristics of cancer stem-like cells. *BMC Biol*, **14**, (20).

Su, J.M., Lin, P., Wang, C.K. and Chang, H. (2009). Overexpression of cytochrome P450 1B1 in advanced non-small cell lung cancer: a potential therapeutic target. *Anticancer Res*., **29** (2), 509-515.

Svergun, D.I. Koch, M.H.J., Timmins, P.A. and May, R.P. (2013). Small-Angle X-ray and neutron scattering from solutions of biological macromolecules. USA: Oxford University Press.

The UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge**.** *Nucleic Acids Res*., **47**, 506-515.

Tichomirowa, M.A., Barlier, A., Daly, A.F., Jaffrain-Rea, M.L., Ronchi, C., Yaneva, M., D Urban, J... and Beckers, A. (2011). High prevalence of AIP gene mutations following focused screening in young patients with sporadic pituitary macroadenomas. *Eur J Endocrinol*., **165** (4), 509-515.

Tria, G., Mertens, H.D., Kachala, M. and Svergen, D.I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, **2** (Pt 2), 207-217.

Trivellin, G. and Korbonits, M. (2011). AIP and its interacting partners. *J Endocrinol.,* **210** (2), 137-155.

Troca, C.M., Alvarez-Barrientos, A., Barrasa, E., Rico-Leo, E.M., Catalina-Fernandez, I., Menacho-Marquez, Bustelo, X.R., Garcia-Boron, J.C…and Fernández-Salguero, P.M. (2013). The dioxin receptor has tumor suppressor activity in melanoma growth and metastasis. *Carcinogenesis*, **34** (12), 2683-2693.

Tsuji, N., Fukuda, K., Nagata, Y., Okada, H., Haga, A., Hatakeyama, S., Yoshida, S... and Itoh, H. (2014). The activation mechanism of the aryl hydrocarbon receptor (AhR) by molecular chaperone HSP90. *FEBS Open Bio*., **16** (4), 796-803.

Tuominen, I., Heliövaara, E., Raitila, A., Rautiainen, M.R., Mehine, M., Katainen, R., Donner, I.,. and Karhu, A. (2014). AIP inactivation leads to pituitary tumorigenesis through defective Gαi-cAMP signalling. *Oncogene*, 1-11.

UniProt Consortium. (2002-2019). 000170- AIP_HUMAN. Retrieved from: http://www.uniprot.org/uniprot/O00170#section_seq

Vacher, S., Castagnet, P., Chemlali, W., Lallemand, F., Mesuere, D., Pocard, M., Bieche, I. And Perrot-Applanat, M. (2018). High *AHR* expression in breast tumours correlates with expression of genes from several signalling pathways namely inflammation and endogenous tryptophan metabolism. *PLOS ONE*, **13** (1), e0190619

Velazquez-Campoy, A., Leavitt, S.A. and Freire, E. (2015). Characterisation of protein-protein interactions by isothermal titration calorimetry. *Methods Mol Biol*., **1278**, 183-204.

Vella, M. (2015). *Structural-Functional Analyses of Aryl Hydrocarbon Receptor Interacting Protein (AIP) and mutant derivative associated with Pituitary Adenomas in Malta* (Master Dissertation). Faculty of Medicine and Surgery, University of Malta, Malta.

Vierimaa, O., Georgitsi, M., Lehtonen, R., Vahteristo, P., Kokko, A., Raitila, A., Tuppurainen, K.... and Aaltonen, L.A. (2006). Pituitary adenoma predisposition caused by germline mutations in the *AIP* gene. *Science*, **312**, 1228-1130.

Vlad-Fiegen, A., Langerak, A., Eberth, S. and Muller, O. (2012). The Wnt pathway destabilises adherens junctions and promotes cell migration via β-catenin and its target gene cyclin D1. *Febs Open Bio*, **2**, 26-31.

Vogel, C.F.A. and Haarmann-Stemmann, T. (2017). The aryl hydrocarbon receptor repressor – more than a simple feedback inhibitor of AhR signaling: Clues for its role in inflammation and cancer. *Curr. Opin, Toxicol*., **2**, 109-119.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*., **46** (W1), 296-303.

Whitelaw, M.L., McGuire, J., Picard, D., Gustafsson, J.A. and Poellinger, L. Heat shock protein hsp90 regulates dioxin receptor function in *vivo. Proc Natl Acad Sci U.S.A*., **92** (10), 4437-4441.

Woloschak, M., Roberts, J.L. and Post, K. (1994). c-myc, c-fos, and c-myb gene expression in human pituitary adenomas. *JCEM*, **79** (1), 253-257.

Yamamoto, J., Ihara, K., Nakayama, H., Hikino, S., Satoh, K., Kubo, N., Iida, T., Fujii, Y., and Hara, T. (2004). Characteristic expression of aryl hydrocarbon receptor repressor gene in human tissues: organ-specific distribution and variable induction patterns in mononuclear cells. *Life Sci*. **74** (8), 1039–1049.

Yang, Y., Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., Wang, J., Satter, A. and Zhou, Y. (2017). SPIDER2: A package to predict secondary structure, accessible surface area and main-chain torsional angles by deep neural networks. *Methods Mol Biol*., **1484**, 55-63.

Yuan Ze University. (2014). UbiSite: A web server for identifying Ubiquitination sites: Version 1.0. Accessed through http://csb.cse.yzu.edu.tw/UbiSite/index.php.

Zheng, Y., Xie, J., Huang, X., Dong, J., Park, M.S. and Chan, W.K. (2016). Binding studies using *Pichia pastoris* expressed human aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator proteins. *Protein Expr Purif*., **122**, 72-81.

## Addendum

Another part of this study that merits at least a mention is the unexpected formation of protein crystals of the human phosphate binding apolipoprotein (ALP) while attempting to crystallise AIP. These crystals formed in 0.1 M HEPES pH 7.0, 10% v/v PEG 6000, almost ten months after the tray had been set up. They displayed a positive signal both in UV and SHG and diffracted at a resolution of 1.5-1.7 Å. These protein crystals were initially thought to be of AIP and since the N-domain and C-domain had already been characterised by other groups, solving the structure of the full length protein was not expected to be challenging, especially at such an optimal resolution. However, it definitely turned out to be quite difficult as the crystals were not of AIP. Data was analysed through molecular replacement using various programs available on CCP4 Online. Through, sequence-independent molecular replacement (SIMBAD) it was identified that the crystal (space group: C 2 2 21) was in fact of a human apolipoprotein, PDB ID: 2V3Q. This 38 kDa protein belongs to the family of eukaryotic proteins collectively known as DING that are ubiquitously expressed. Interestingly, the group that first described and crystallised this apolipoprotein, also termed its discovery as being "serendipitous", as it co-purified with their enzyme of interest, paraoxonase (Morales *et al*., 2006). The crystal structure of this protein shares structural homology with the prokaryotic phosphate solute binding proteins (Morales *et al*., 2006).

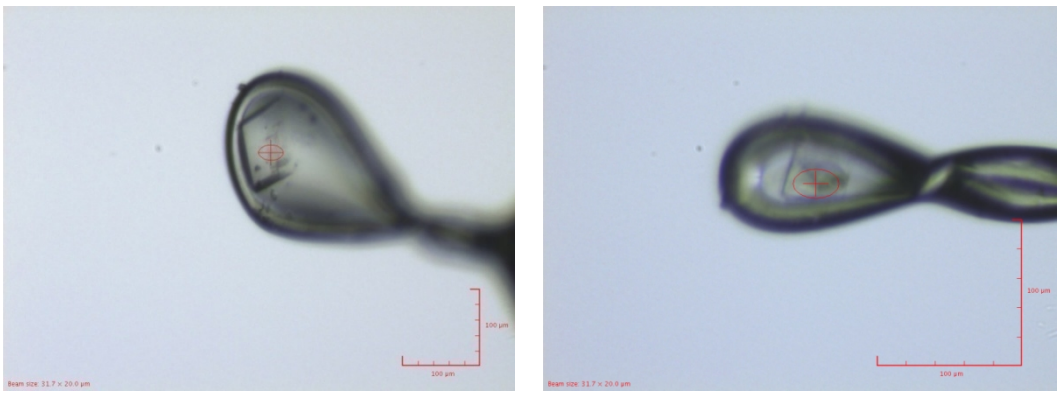


**Figure 5.5: Loaded crystals before being diffracted by X-rays.**

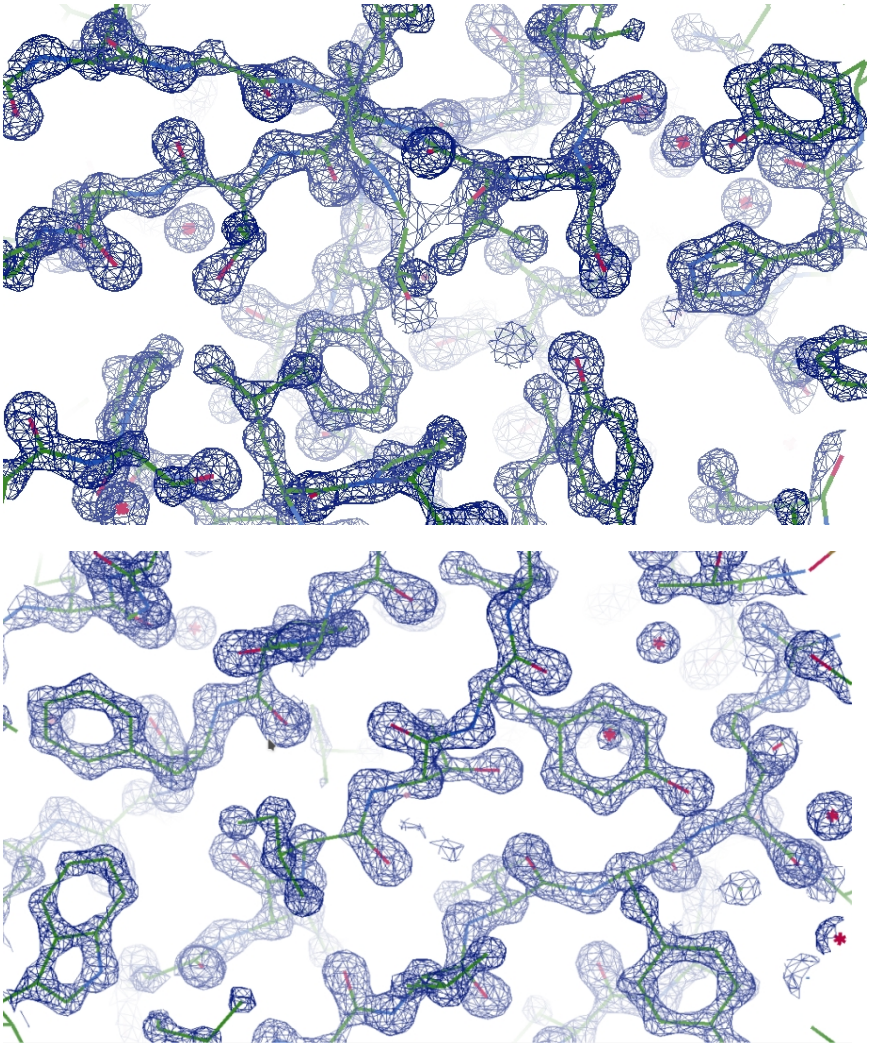


**Figure 5.6: Electron density map.** Parts of the electron density map of the diffracted crystal after refinement, resolution 1.3Å. This electron density map matches precisely the sequence in the pdb model of the 2V3Q protein.

Since no one in the lab locally works or has ever worked with this protein before, the exact origin of this contaminant remains uncertain. This contaminant was not observed on SDS-PAGE, Native-PAGE or mass spectrometry, after purification. While it is curious and almost unbelievable that the presence of a low concentration contaminant can crystallise instead of the protein of interest, the same surprising result has also been obtained by other groups, who traced its origin to a contaminated tubing of an AKTA$^{(TM)}$ purification system. Similar to what was observed in this study, the protein crystallised after several months (J. Baumann, personal communication, 2019).

***Moral of the story: Never trust a protein crystal that has taken months to form...***