Overcoming the bottleneck phases in protein crystallography

Thesis submitted in accordance with the requirements
of the University of Liverpool for the degree of Doctor of Philosophy

by

Melissa O'Connor-Farbiarz

October 2009

Declaration

I hereby declare that this thesis has been composed by myself, that all the material presented within it is original unless stated otherwise, and that no part of this work has been submitted for any other degree or professional qualification.

Melissa O'Connor-Farbiarz October 2009

Acknowledgements

The guidance and support provided by my supervisors, Dr. Mark C Wilkinson and Dr. Daniel J Rigden is greatly appreciated and I am very grateful to have been given this opportunity. I have learned so much and I owe that to your lending of expertise and time. You have set me up with a great array of skills for a successful future in science.

The generous hospitality of Prof. Miroslav Papiz while visiting Daresbury laboratories and the assistance with crystallisation set up is very much acknowledged.

I am also extremely grateful to Mr Mark Prescot for the many attempts at obtaining ESI-MS data for metal binding studies and to Prof. Alexander McLennan for assistance with the Ap_4A-H assay.

I really have to thank all members of Lab C for their good humour and support throughout the three years at Liverpool. I will be forever indebted to Dr. Leonora Ciufo, Dr. Debora Ward, Dr. David Fisher and to Dr. Daryl Williams for their encouragement and many scientific discussions.

The assistance of both Holly Birchenough and Lesley-Ann Stevens for their contributions and I hope that they found these studies as rewarding as I did.

To my family, particularly my husband, who I owe many thanks to for the encouragement each of you provided throughout my research and while writing this thesis. Thank you for your patience and support.

Abstract

X-ray crystallography is extensively used for the elucidation of three-dimensional protein structures. Despite the advances in X-ray crystallographic research, single, well-ordered crystals are required and many bottleneck phases are encountered in this process. GPI-PLD was the chosen subject for structural studies given that there is a huge demand for structural information for therapeutic applications as this protein has been implicated in many pathological conditions, including diabetes and malignancy. The common hurdles that are encountered during heterologous protein expression were highlighted and strategies (both common and innovative) were tested to overcome solubility issues and to increase yields. Many parameters were modified in the optimisation of target protein expression including: utilization of both a prokaryotic and a eukaryotic host, variations of induction conditions, plasmid expression vector, length of construct and the employment of large, soluble protein fusions. Unfortunately, it was not possible to obtain sufficient amounts of soluble protein for crystallisation trials. However, studies provided a good account of how studies can be hindered even at the early stages of construct design let alone expression of a soluble protein.

The X-ray crystallography technique also comes with its obstacles. At present, the methods available for solving the 'phasing problem' have drawbacks. The incorporation of metals, particularly those of the lanthanide series, can overcome this flaw by altering diffraction intensities and/or providing a source of anomalous scattering. A metal-binding tag based on the conserved DxDxDG motif of calcium-binding proteins (CaBPs) has been designed to enable lanthanide binding to a target protein. Investigations using gel filtration chromatography and isothermal titration calorimetry have confirmed metal binding abilities of two metal binding tags (LBTA and LBTB). A vector, containing the metal-binding tag (LBTB) region was engineered and a model protein, Asymmetric diadenosine 5', 5'', P1, P4 - tetraphosphate (Ap_4A) Hydrolase was successfully expressed containing the metal binding tag with no consequence to expression levels, solubility or activity when compared to the non-tagged model protein. Numerous crystallisation trials, using a variety of commercial screens have been set up but unfortunately no lead conditions were obtained to date. It is hoped that with further work that this metal-binding tag can be used for heterologous expression of proteins providing a new, general and perhaps automated approach to phasing of protein crystals.

List of Abbreviations

AChE	Acetylcholinesterase
AP	Alkaline-phosphatase
Ap_4A-H	Asymmetric diadenosine 5', 5'', P1, P4 –tetraphosphate hydrolase
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
CaBPs	Calcium binding proteins
CEA	Carcinoembryonic antigen
DAB	Diaminobenzidine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DTNB	Dithiobis-(5,5'-nitrobenzoic acid
DTT	Dithiothreitol
ECL	Enhanced chemical luminescence
EDTA	Ethylenediaminetetraacetic acid
GFC	Gel filtration chromatography
GPI-PLD	Glycosylphosphatidylinositol-specific phospholipase D
GST	Glutathione S-transferase
HDL	High-density lipoproteins
IB	Inclusion bodies
IEX	Ion exchange chromatography
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Luria-Bertani
LBT	Lanthanide binding tag
MALDI-TOF	Matrix assisted laser desorption/ionization - time of flight
MBP	Maltose binding protein
MIR	Multiple isomorphous replacement
MR	Molecular replacement
NI-NTA	Nickel-nitrilotriacetic acid)
NMR	Nuclear magnetic resonance
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PLC	Phospholipase C
PLD	Phospholipase D
PROSO	Protein solubility evaluator
PTMs	Post translational modifications
PVDF	Polyvinylidene fluoride
RNA	Ribonucleic acid
RP-HPLC	Reverse phase high pressure liquid chromatography
SAD/MAD	Single/multi-wavelength anomalous diffraction
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SECRET	Sequence-based crystallisability evaluator
SEG	Segmental analysis
TBS	Tris-buffered saline
TEMED	N, N, N', N'-tetramethyl-ethylenediamine
TFA	Trifluoroacetic acid
TNB	Thionitrobenzoic acid
VSG	Variant surface glycoprotein

List of Tables and Figures

TABLES

FIGURES

CONTENTS

1: Introduction

The elucidation of three dimensional protein structures by X-ray crystallography and NMR spectroscopy has provided a tremendous understanding of fundamental biochemical processes. These insights have rationalised practical applications such as structure-based drug design for proteins of therapeutic interest, providing new opportunities in the medical, biotechnological and pharmaceutical fields. Naturally, with developments in genome sequencing, automation/robotics and bioinformatics, structural biology now has a massive impact on drug discovery (Congreve *et al.,* 2005). However, there are a number of bottleneck phases that must be overcome in the process of protein structure determination. The production of high levels of pure, soluble protein for structural studies can be hindered at any stage. The initial selection of the gene target, the cloning, expression and purification of a given target can all be problematic even before attempting crystallisation trials. Studies herein were aimed at expression of GPI-PLD, which is a protein of therapeutical interest. This in turn, highlighted some of the difficulties faced in structural biology and how they can sometimes be overcome by using both common and innovative strategies.

1.1: Cloning and expression

High-level production of recombinant protein is a prerequisite if studies are to progress to structural analysis and there are many parameters to be considered in this quest. The difficulty is obtaining a concentrated (typically 2 – 20 mg/ml), pure, monodispersed and soluble protein preparation (Baneyx, 1999). Intuitively, expression of a given protein would be best in its native cell type under the desired

physiological conditions. It is unfortunate that few proteins occur naturally in amounts sufficient enough for the rigorous purification procedures required for structural studies. On the other hand, the heterologous expression of a recombinant protein can provide yields that are significantly higher than what would be expected of the native environment.

1.1.1: Target considerations

First to be considered is the target sequence. The amino acid sequence can influence the overall success of crystallisation (Smialowski *et al.*, 2006; Smialowski *et al.*, 2007). The removal or addition of a single amino acid can sometimes be the difference between obtaining high quality protein crystals for X-ray crystallography and failure. Optimal transcription and translation in the chosen host can be ensured by the removal of extended palindromic sequences or repetitive rare codons in the reading frame. If such sequences are present there is a risk of prolonged pausing of the ribosomes that could result in premature termination or mis-incorporation of amino acids (Baneyx, 1999). Additionally, as a result of size limitations in NMR studies and the flexibility of multidomain proteins causing problems in crystallisation it is common to select a single domain. Moreover, it can be very difficult to clone, express and crystallise larger proteins comprising multiple domains.

The Shine-Dalagarno sequence is a prerequisite for translation. Obstruction of this sequence prevents ribosome binding and inhibits translation (Ramesh and Nagaraja, 1994). A Kozak sequence for initiation, (**A/G**)NN*ATG***G**, is generally required by eukaryotic expression systems to ensure translation is initiated from the first ATG

triplet. A purine at positions -3 and a guanine at +4 are particularly important. If this triplet is surrounded by sequence that is different from the optimum content, leaky scanning can occur whereby the first initiation codon is omitted and translation is initiated from a second or third ATG triplet located downstream. A homologous signal peptide sequence for secretion from a yeast expression system would provide the best guarantee for obtaining correctly folded protein (Hitzeman *et al.*, 1990). For example, the replacement of the mammalian signal peptide with homologous α-factor signal sequence has been known to increase soluble protein yields from 1% to 80% (Hitzeman *et al.*, 1990). Many proteins have also been expressed and secreted successfully in yeast using their natural signal sequences. This sequence is not cleaved in *E. coli* and may potentially cause problems during translation and later during crystallisation.

1.1.2: Expression systems

Recombinant protein expression can be hosted by bacteria, yeast, mould, mammalian, plant or insect cells, done in cell free systems or by use of transgenic plants and animals. Each system comes with advantages and also drawbacks (Demain and Vaishnav, 2009). The selection of a system for over-expression of a given target protein is usually influenced by the system's ability to rapidly produce high yields of functionally folded protein. Obviously, this is highly dependent upon the gene sequence and requirements for post-translational modifications (PTMs) such as glycosylation and disulphide bond formations in the folded protein. *E. coli,* the most widely used expression vehicle (Terpe, 2006), is particularly known for its ability to cope with over-expression of recombinant proteins that are usually obtained without post translational modification heterogeneity. However,

the lack of post-translational modification machinery in bacteria poses immense challenges for protein folding that ultimately result in the production of biologically inactive protein expressed as insoluble aggregates, namely inclusion bodies (IBs) (Chou, 2007).

Although recombinant expression by use of eukaryotic expression systems is, in general, time consuming, more costly and often provides lower yields when compared to prokaryotic systems, it is the best route forward for larger proteins, those that may necessitate secretion and proteins requiring glycosylation and disulphide bond formation (Demain and Vaishnav, 2009; Dai *et al.*, 2005). Yeasts secrete and process glycoproteins in much the same way as mammalian cells do. The initial steps of protein N-glycosylation before the glycan is transferred to the Golgi are identical, after which point a divergence of processes occurs between the two organisms. N-glycosylation in *Saccharomyces cerevisiae* involves the addition of numerous mannose sugars throughout the entire Golgi that leads to, in most cases, hypermannosylated glycans with more than 100 mannose residues (Kuruzinska *et al.*, 1987). *Pichia pastoris*, the popular methylotrophic yeast, has been cleverly manipulated to produce a more human-like glycosylation by genetic engineering of its secretory pathway (Choi *et al.*, 2003) and avoid the hyperglycosylation experienced with other yeasts. Another major advantage of employing a yeast system over bacteria is its ability to secrete heterologous proteins, containing the appropriate signal sequence, into the fermentation medium (Dai *et al.*, 2005). This also facilitates isolation and purification of the recombinant protein as less than 1% of native proteins in yeast are secreted into the extracellular broth.

Also capable of secretion of high levels of recombinant proteins are filamentous fungi (moulds). Despite the attractive attributes of this expression host, including PTMs and long-term genetic stability, comparatively low yields can be experienced with some heterologous proteins when secreted. Insect cells boast more sophisticated machinery than the aforementioned organisms (Agathos, 1991) and are capable of carrying out more complex eukaryotic PTMs. This system provides greater flexibility on protein size and is capable of simultaneous expression of multiple gene targets. Thus, they are a good choice for the heterologous production of mammalian proteins. Inefficient secretion, low levels of expression and IB formation are the shortcomings associated with the use of this system. The most logical choice for heterologous expression of mammalian proteins would be a mammalian host. Chinese hamster ovary (CHO) cells are most likely the preferred system for recombinant protein production in mammalian cells, but many others exist (Wurm, 2004). Their ability to express functional, correctly folded proteins with the necessary PTMs can justify the use of mammalian cell cultures in the manufacturing of many biopharmaceuticals at very high costs. However, this system is imperfect in the secretion of foreign proteins and can be devastated by contamination by viruses. The use of this system in labs is generally hindered by the expense and low yields. Transgenic plant and animals are also capable of recombinant protein production but the use of such systems is limited.

Cell-free expression systems offer an attractive alternative to the more traditional approaches. This system can be modified to promote protein folding and disulphide bond formation. Ligands, substrates and protease inhibitors can be incorporated into the system to increase protein stability. Moreover, such systems lend themselves to automation for high-throughput parallel expression of multiple

protein constructs (Sawasaki, *et al.*, 2002; Endo and Sawasaki, 2003). It is no surprise that cell-free systems are considered to be a promising tool for automated production of proteins for structural analysis (Chen *et al.*, 2007; Vinarov *et al.*, 2004). The major drawback to this approach to protein production is its expense, which is cost prohibitive for many laboratories.

Generally, the more complex the expression system, the higher the cost of production and the lower the yields. There are no universal optimal expression requirements available for recombinant protein expression. Each protein poses new challenges and in every case expression has to be optimised in the chosen host.

1.1.3: Host strains

The production of heterologous protein using a prokaryotic expression hosts can more often than not become an exigent undertaking. In addition to inclusion body formation, over-expression of a foreign protein in *E.coli* can also be toxic and often leads to cell death. The protein expressed can be inactive as a result of misfolding. Specialised *E. coli* stains have been developed to tackle these common problems faced during recombinant protein production. Bacterial expression systems, with their well-characterised genetics, offer an advantage in the availability of an ever-growing selection of cloning vectors and mutant host strains. BL21, a *lon* and *ompT* protease-deficient strain, is known to improve the likelihood of intact, full-length protein expression (Studier and Moffatt, 1986) and is routinely used as an ideal starter strain. BL21 derivative strains, mutant C41(DE3) and double mutant C43(DE3) have demonstrated successful expression of proteins previously found to be insoluble in the BL21(DE3) strain (Miroux and Walker, 1996) and have been

particularly useful for expression of membrane proteins (Arechaga *et al.*, 2000). A number of strains have been developed to overcome problems that arise as a result of codon bias in each system. Translation of eukaryotic gene sequences, which may contain large numbers of codons that are rare in bacteria, namely AGG, AGA, AUA, CUA, GGA and CGG, can put a great deal of stress on the translational machinery, which, as a result renders truncation of the recombinant protein, translation frameshifts and mis-incorportion of amino acids (Kurland and Gallant, 1996). Rosetta and BL21-codon plus *E. coli* strains are capable of recognising those rare codons that are not recognised in other strains by providing a supply of tRNAs for each codon (AUA, AGG, AGA, CUA, CCC, GGA and CGG). Enhanced expression of heterologous protein with gene sequences containing a fair number of rarely used codons in *E. coli* has been achieved using these strains (Gustafsson *et al.*, 2004). The Origami strain is another useful tool as it contains a *trxB/gor* mutation and is capable of aiding the construction of disulphide bonds by providing a less-reducing cytoplasmic environment (Prinz *et al.*, 1997). This strain is recommended mainly for the expression of proteins that require disulphide bond formation to achieve their active correctly folded conformation (Xu *et al.*, 2008).

1.1.4: Choice of vector and affinity tags

The choice of vector also plays a pivotal role in expression of high levels of soluble protein. There are a large number of *E.coli* expression plasmids to choose from (reviewed in (Baneyx, 1999; Hannig and Makrides, 1998; Makrides, 1996). The T7 based pET expression plasmids (commercially available from Invitrogen, U.K) have been the most commonly used in recent years. Their attractiveness is owed to the generation of large amounts of mRNA and concomitantly high protein

expression. Usually, the affinity tag selected to aid protein purification influences the choice of vector. The inclusion of an affinity tag is attractive for a number of reasons. In addition to the facilitation of purification the inclusion of affinity tags can also improve recombinant protein yield, increase solubility, aid protein folding and prevent proteolysis (Terpe, 2003). Conversely, tag addition can also reduce yields, inhibit biological activity, change protein conformation and provide undesired flexibility that can be detrimental for structural studies so subsequent removal by restriction proteases is sometimes required (Arnau *et al.*, 2006). His-tags are the most widely used in structural studies with more than 60% proteins produced including a polyhistidine tag (Derewenda, 2004). Interestingly, the simple action of exchanging a C-terminal His-tag to one that is included at the N-terminus can make the difference between expression of an inactive, insoluble protein and a soluble protein with moderate levels of activity (Walker *et al.*, 2001). A popular strategy in addressing this solubility problem is the employment of a large, soluble protein fusion/partner to improve solubility and folding of the protein in question. Such fusion proteins include, maltose binding protein (MBP) (Fox *et al.*, 2003), glutathione S-transferase (GST) (Smith and Johnson, 1988) and thioredoxin (LaVaillie *et al.*, 1993) amongst many others (Arnau *et al.*, 2006). The 3D structure of *Schistosoma japonicum* GST has already been determined (McTigue *et al.*, 1995). Therefore, this structure can be used to provide a search model to solve the crystallographic phasing problem (Discussed in section 1.3.1) by molecular replacement (MR) methods. Furthermore, the GST-fusion can, for smaller peptides of 5-42 residues in length, be crystallised using similar conditions used for the isolated GST (Zhan *et al.*, 2001). The GST fusion is translated first allowing its folding before the nascent C-terminal fusion partner, thereby enabling

interaction with the target protein as it emerges from the ribosome. Although the stabilising role of such large fusion proteins remains unclear (Waugh, 2005), it is agreed that it is most likely the association of the fusion with its target protein counterpart that prevents aggregation of partially folded or incorrectly folded intermediates. The small size and stable fold of the immunoglobulin-binding domain of streptococcal protein G (GB1 domain) (Gronenborn *et al.*, 1991) has also been reported to assist in the correct folding of a target protein. This small, 56 residue protein is known to have a high bacterial expression capability and can limit the requirement for further optimisation of expression (Huth *et al.*, 1997). Naturally, if soluble preparations are desired a number of fusions should be appraised. A variety of commercial combinatorial cloning systems are available to facilitate the transfer of the target protein sequence between plasmids harbouring various tags. Gateway® technology (Invitrogen) is the most commonly employed (Hartley *et al.*, 2000). Not only does this system offer the advantage of facilitating transfer of the cloned gene between plasmid vectors offering alternate affinity tags, but also vectors for additional expression systems.

1.1.5: Modulating expression kinetics

When the rate of synthesis exceeds the host's folding ability there is an accumulation of IBs in the cytoplasm. It is reasonable to assume that a eukaryotic protein will require more complex folding mechanisms (as a result of the PTMs) than allowed by the prokaryotic host, which render them more susceptible to aggregation at higher transcriptional rates. By adjusting the isopropyl β-D-thiogalactopyranoside (IPTG) concentration, expression can be regulated from very

low levels up to robust, fully induced levels. For difficult target proteins a reduced level expression may enhance solubility and maintain activity.

There are many single-case studies in which the lowering of cultivation temperatures shows a marked increase in protein solubility. The hydrophobic interactions that are responsible for IB formation are temperature dependent so it is no surprise that lower temperatures generally lead to increased stability and correct folding. It is also thought that lower temperatures increase the level of chaperones to assist with protein folding. Additionally, the suppression of protease activity at lower temperatures means that the recombinant protein is less susceptible to degradation (Hunke and Betton, 2003; Spiess *et al.*, 1999).

The choice of media available to support bacterial growth can also play an important role in recombinant protein expression levels and solubility. Supplementation of the fermentation media can also be advantageous. Induction of protein expression under osmotic stress by the addition of sorbitol and betaine to cultures has demonstrated some success in increasing the yield of active soluble protein and a reduction in inclusion body formation (Blackwell and Horgan, 1991). Some metalloproteins require the availability of the metal for correct folding of the catalytic domain (Reid *et al.,* 1981) so it is desirable for expression of such proteins to provide a supply of the preferred cofactor in the media.

1.1.6: Cytoplasmic or periplasmic heterologous expression

High yields are usually achieved when protein is expressed and accumulated in the cytoplasm. Thus, it is no surprise that this is often the first choice for over-expression. The downside to this compartment is that the target protein must be

separated from the immense pool of cytoplasmic host proteins. The periplasm, however, offers a purification advantage because there are fewer bacterial proteins compared to the cytoplasm. The oxidising environment of the periplasm provides suitable conditions to aid protein folding, form disulphide bonds and provide sufficient cleavage of the signal sequence during translocation (Hunke and Betton, 2003).

1.1.7: Recovery of protein from IBs by refolding

Although the production of insoluble IBs in the cytoplasm (or the periplasm if recombinant protein has been directed for periplasmic location or secretion) is highly undesirable, they can, on occasion, be recovered by denaturation and re-folding. The high expression levels and compactness of protein aggregates can even facilitate the purification process. A number of refolding methods have been documented that claim to solubilise IBs and obtain a correctly folded, active protein (Swietnicki, 2006). Generally, IBs are solubilised by use of chaotropic agents such as guanidinium hydrochloride and urea amongst others. Milder denaturant conditions are thought to beget superior folding yields with proteins retaining biological activities (Puri *et al.*, 1992). Dilution and dialysis techniques are available and aim to provide an environment to support protein folding (Maeda *et al.*, 1995; West *et al.*, 1998). On-column methods of refolding using chromatography (affinity chromatography, hydrophobic interaction chromatography, ion-exchange chromatography and gel filtration chromatography) can all been considered. Of these methods, gel filtration chromatography (GFC) can provide a more efficient approach to refolding (Lui *et al.*, 2006) by the gradual removal of denaturant in the buffer. Additionally, novel protein refolding strategies have been developed using artificial chaperones, β-cyclodextrin (Daugherty *et al.*,

1998) and β-zeolite (Chicku *et al.*, 2006) in an attempt to obtain high levels of pure protein for structural studies.

1.2: Glycosylphosphatidylinositiol-specific phospholipase D

Glycosylphosphatidylinositiol-specific phospholipase D (GPI-PLD) is an ideal choice for structural studies. The involvement of this protein in a number of disease states and in growth and development (as discussed in section 1.2.5) has generated interest, so there is now a demand for structural information that could provide essential information regarding GPI-PLD catalytic site, which would be highly beneficial for drug discovery. Expression of recombinant GPI-PLD brought to light the difficulties that can hinder protein structural studies in the early stages of molecular cloning and protein expression.

1.2.1: GPI-anchor discovery

Although it was known that certain proteins, namely alkaline phosphatase and erythrocyte acetylcholinesterase (AChE) could be released from the cell surface by bacterial phosphatidylinositol phospholipase Cs (PI-PLC) (Low, 1989) the concept of GPI-anchoring did not emerge until 1985. The first detailed structural analysis of the GPI-anchor soon followed with compositional and structural data from *Trypanosoma brucei* variant surface glycoprotein (VSG) (Ferguson and Williams, 1988; Ferguson, 1999). Since then, studies of GPI-anchor structure, function and biosynthesis have progressed significantly and more than 200 eukaryotic GPI-anchored proteins have been identified (Brown and Waneck, 1992; Paulick and Bertozzi, 2008).

The GPI-anchor is a complex structure comprising a phosphoethanolamine linker, glycan core and a phospholipid tail (Figure 1.1). All GPI-anchors that have been characterized to date share a common glycan core. Modifications to the phosphoinositol, glucosamine and mannose residues within the conserved gylcan core with phosphoethanolamine groups and other sugars (Nosjean *et al.*, 1997; Low, 1999) are common (Figure 1.1). Additionally, variations in the phospholipid tail have also been documented (Figure 1.1). Mammalian GPI-anchor precursors have an additional fatty acid substitution for the hydroxyl of the inositol ring (Urakaze *et al.*, 1992). It is thought that this substitution imparts resistance of such anchors to all PI-PLCs but hydrolysis of GPI-PLD can still be achieved (Toutant *et al.*, 1989; Urakaze *et al.*, 1992). Despite the structural complexity of the GPI moiety its only known function is its capacity for membrane insertion and anchoring of proteins to the cellular membrane.

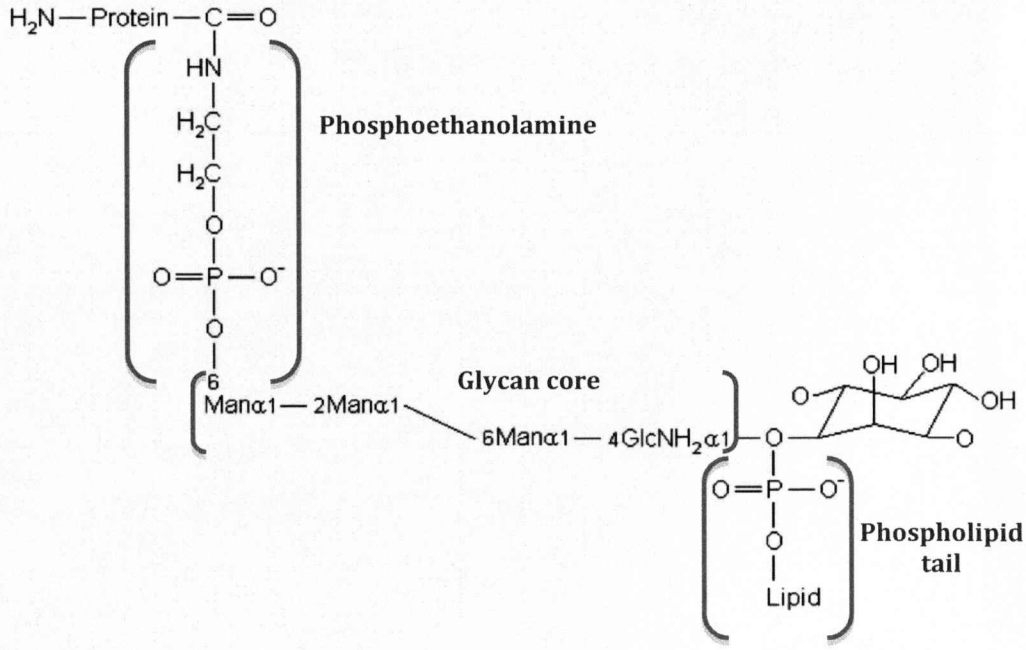


Fig 1.1: The conserved structure of a GPI anchor attached to a target protein.
The lipid may be diacylglycerol, alkylacylglycerol or ceramide. The three mannose (Man) residues and the glucosamine (GlcN) residue make up the glycan, which is attached to the head group of the lipid by a glycosidic linkage of GlcN to position 6 of the inositol ring. Additional structures can be added to the glycan and in some cases added to the inositol ring. The carboxyl terminal of the protein is attached to the glycan via ethanolamine, which is subsequently connected to the first Man residue of the glycan by a phosphodiester bond. Adapted from (Ferguson, 1999; Hooper, 1997).

1.2.2: GPI proteins and their attachment

GPI-anchored proteins, like type I transmembrane and secreted proteins, are synthesized with an N-terminal hydrophobic sequence for direction of the nascent polypeptide into the lumen of the endoplasmic reticulum that is later cleaved by a signal peptidase. They also contain a peptide segment at the C-terminal end of the polypeptide that signals GPI-anchor attachment. This sequence resembles the type I membrane anchor sequence in that it is somewhat hydrophobic and residing at the C-terminal end of the polypeptide sequence. GPI signal sequences have the appearance of a defective or modified transmembrane anchor, containing polar residues in an otherwise hydrophobic region (Waneck *et al.*, 1988). They are often much shorter than the required length and the basic residues required for interaction with the acidic headgroups of phospholipids on the cytoplasmic face are absent (Ferguson *et al.*, 1988; Cross, 1990). This C-terminal sequence is cleaved after translocation across the endoplasmic reticulum (ER) and covalently linked to a pre-formed GPI-anchor by a specific transamidase before being displayed on the cell surface of the plasma membrane (Ferguson *et al.*, 1988; Nosjean *et al.*, 1997).

GPI-linked proteins are a diverse family of proteins with no evident shared functional characteristics to shed light as to why they are GPI-anchored. There are a large number of functionally diverse cell-surface proteins including receptors, cell adhesion molecules, protozoa antigens, catalytic enzymes and lymphocyte antigens that depend on GPI-anchoring for attachment to the surface of the cell (Sevlever *et al.*, 2000). Some GPI-anchored proteins have important roles in immune recognition, complement regulation and intracellular signalling (Brown and

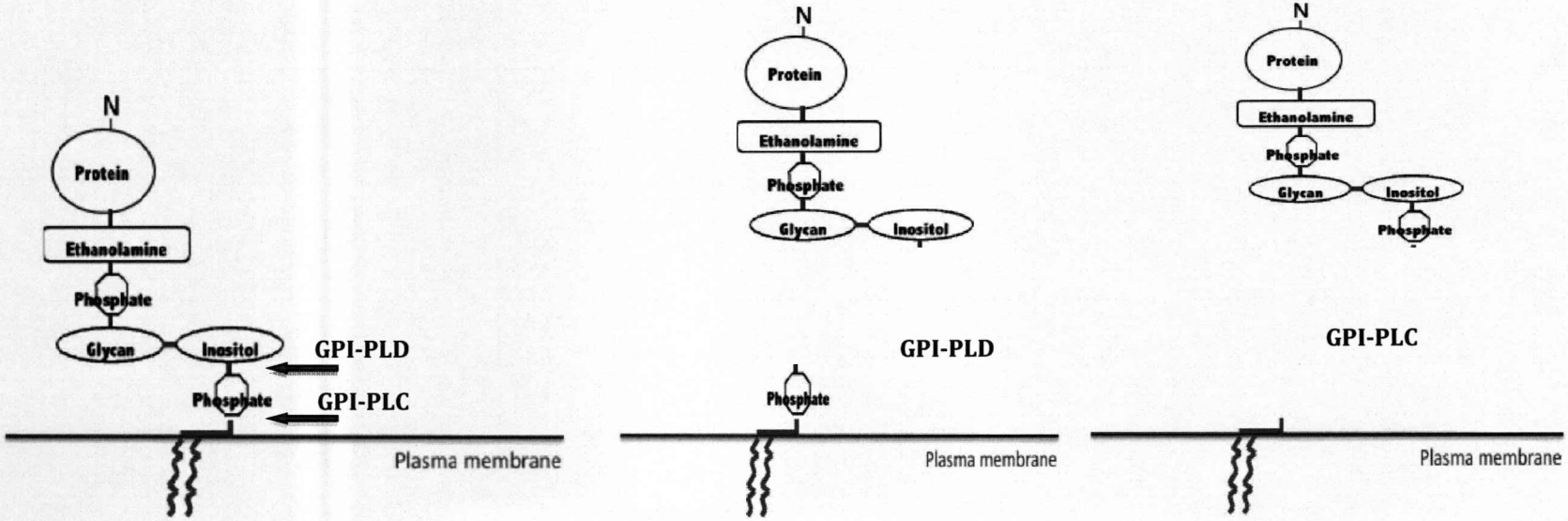
Waneck, 1992; David *et al.*, 1997). Naturally, it is no surprise that GPI-anchored proteins could be implicated in pathogenesis given the diversity of this group.

1.2.3: GPI-anchor hydrolysis

GPI-proteins remain attached and displayed on the cell surface for a period of time. They can subsequently undergo internalisation into the cell or be released into the extracellular space (Nosjean *et al.*, 1997). This liberation of the GPI-protein from the cell surface can be achieved by the hydrolysis of the phosphatidylinositol moiety by two GPI-specific phospholipases, namely GPI-PLD and GPI-PLC. Each has a preferential cleavage site on either side of the phosphate group linking the inositol to glycerol (Figure 1.2).

1.2.4: The role of GPI-PLD

An inactive, latent form of GPI-PLD is relatively abundant in mammalian serum (Davitz *et al.*, 1987; Huang *et al.*, 1990; Low and Prassad, 1988) and varying levels of activity and expression have been detected in a number of different tissues and cell lines (Leboeuf *et al.*, 1998; Schofield and Rademacher, 2000). However, it is thought that the liver is the major source of circulating GPI-PLD (Deeg *et al.*, 2001).



Fig; 1.2: GPI-anchor hydrolysis by GPI-PLC and GPI-PLD
Figure illustrates cleavage sites of GPI-PLC and GPI-PLD.

The vast quantities of this protein, its presence in so many different tissues and cell lines coupled with its involvement in pathological conditions make this protein very interesting but its mechanism of action is not well understood. A number of reports have demonstrated GPI-protein hydrolysis by intracellular GPI-PLD and a co-expressed recombinant enzyme and GPI-anchored protein in the natural environment without perturbing the membrane (Brunner *et al.*, 1994; Metz *et al.*, 1994; Lierheimer *et al.*, 1997; Tsujioka *et al.*, 1998; Wilhelm *et al.*, 1999). Consequently, it has been postulated that an intracellular compartment may provide an apt environment for the activity of the newly synthesised GPI-PLD. Another study demonstrated intracellular cleavage of both GPI-anchored proteins and GPI-anchor intermediates (Mann *et al.*, 2004). This opens up the possibility that GPI-PLD can regulate the steady state levels of GPI-anchored proteins by hydrolysing the anchor before or after its attachment to proteins. It is also a possibility that GPI-PLD remains catalytically inactive until a change occurs in the membrane environment of the substrate (Low and Huang, 1991; Brunner *et al.*, 1994; Metz *et al.*, 1994; Lierheimer *et al.*, 1997; Tsujioka *et al.*, 1998; Wilhelm *et al.*, 1999; Mann *et al.*, 2004; Deng *et al.*, 1996).

GPI-PLD is also known to associate with high-density lipoproteins (HDL) in a small discrete, minor fraction of lipoproteins containing apoA-I and apoA-IV (Deeg *et al.*, 2001). The function of GPI-PLD in serum and the reason why the enzyme associates with HDL is not clear, but apoA-I has been reported to stimulate GPI-PLD activity *in vivo* (Hoener *et al.*, 1993). The majority of GPI-PLD circulating in serum has a molecular weight of approximately 115 kDa. However, a small amount (5-10% of the total GPI-PLD) of an active 47kDa fragment of GPI-PLD has been identified in bovine serum (Hoener *et al.*, 1994). Proteolytic processing of GPI-

PLD by a lysosomal fraction and by cathepsin D produced an enzymically active 47kDa form with an identical N-terminus to the full sized enzyme (Hoener *et al.*, 1994). Furthermore, trypsin cleavage of intact GPI-PLD results in three distinct peptides. GPI-PLD activity was associated with a 38.6 kDa, N-terminal peptide (Hellar *et al.*, 1994). All of the above suggest the presence on the active site at the N-terminus of the protein.

Antibodies directed towards a mid portion region spanning amino acid residues 275-296 of human GPI-PLD appear to increase activity (Deeg and Bowen, 1999). This region contains several trypsin cleavage sites. It has been proposed that activation of GPI-PLD by trypsin or mid-portion antibodies may result from interrupted intramolecular interactions that could be important in its regulation. In addition, there is a significant decrease in the activity of bovine GPI-PLD resulting from phosphorylation of Thr 286. It has been suggested that this is due to steric hindrance of the active site by the additional phosphate group (Civenni *et al.*, 1999).

1.2.5: GPI-PLD in health and disease

It is becoming increasingly important that we learn more about this enzyme as it is thought that it may play a crucial role in a number of pathological conditions, disease states and growth and development. GPI-PLD has been implicated in the process of bone formation during mouse embryogenesis (Gregory *et al.*, 2005). It has been suggested that GPI-PLD also has a role in inflammation and in the pathogenesis of human atherosclerosis (O'Brien *et al.*, 1999). Studies on diabetic

mouse models of type-1 diabetes showed that levels of serum GPI-PLD are altered in diabetic states (Deeg *et al.*, 2001). Furthermore, up-regulation of GPI-PLD expression in immortalized and malignant tumour cell lines appears to be associated with malignancy and tumour progression (He *et al.*, 2002). GPI-PLD mRNA expression has been shown to correlate with tumour progression in human ovarian cancer (Xiaotong *et al.*, 2002). In addition to this, Yamamoto *et al* (2005) suggested that membrane bound carcinoembryonic antigen (CEA), a widely used tumour marker, is GPI-anchored to the cell membrane. It was also suggested in this report that GPI-PLD is responsible for the cleavage and release of CEA from the GPI-anchor, consequently enhancing metastatic potential in colorectal carcinoma cells. Conversely, GPI-PLD expression decreases in hepatocellular carcinoma (HCC) patients. The over-expression of GPI-PLD in HepG2 cells results in increased sensitivity to complement-dependent cytotoxicity (CDC)-killing, impairs proliferative capacity and promotes apoptosis (Jian-Hua *et al.*, 2009). These findings clearly implicate the involvement of GPI-PLD-catalyzed GPI-anchor cleavage, which is for the most part up-regulated in cells under pathological conditions, including cancers and inflammation. This makes cellular GPI-PLD a promising target for monitoring disease and in predicting clinical outcomes.

1.2.6: GPI-PLD sequence and domains

Human GPI-PLD cDNA (Scallon *et al.*, 1991) predicts an 814-817 amino acid protein of approximately 90-115 kDa. Analysis of its primary amino acid sequence reveals a number of potential post-translational modifications, including phosphorylation, N-glycosylation and myristoylation sites (Leboeuf *et al.*, 1998; Scallon *et al.*, 1991; Deeg and Davitz, 1994). These post-translational

modifications may explain the variation in the sizes of GPI-PLD. Such modifications will also have to be considered when choosing an expression host if they are involved in protein folding.

Relatively little can be determined from simple sequence comparisons but, clues relating to the function of bound Ca^{2+} and the location of substrate binding sites come from the high degree of sequence similarity to metal binding domains in the N-terminal domain of integrin α-subunit (Li and Low, 1999). GPI-PLD is made up of 2 functional domains: an N-terminal catalytic domain consisting of amino acids 1-275 (Hellar *et al.*, 1994) and a C-terminal β-propeller (amino acids 376-816) (Figure 1.3). There is increasing evidence that the β-propeller domain plays a major role in maintaining the catalytic activity of GPI-PLD. C-terminal deletions of only two to five amino acids reduce GPI-PLD activity by 70% and deletions of more than five amino acids result in complete loss of enzymic activity. The decrease/suppression of GPI-PLD catalytic activity is likely the result of misfolding or the unravelling of the β-propeller (Stradelmann *et al.*, 1997).

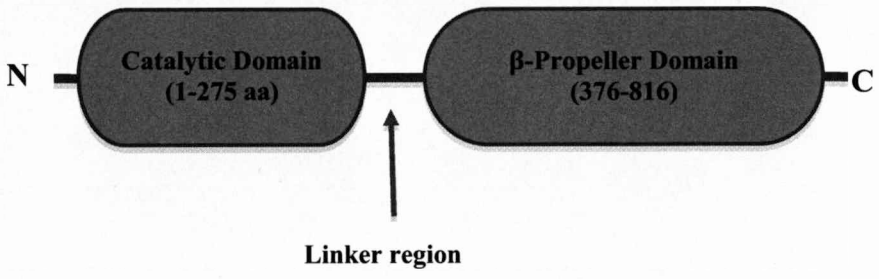


Fig 1.3: Functional Domains of GPI-PLD. The catalytic site resides within the N-terminal catalytic domain consisting of the first 275 amino acids. The C-terminal domain is commonly referred to as the β-propeller domain (amino acids 376-816). A linker region (276-375) separates the two domains.

Despite the high concentration of GPI-PLD in plasma and the extensively studied molecular biology and biochemistry there is limited information regarding its structure and function. The only structural information to date is a computer model derived from a distant relationship between GPI-PLD and bacterial phosphatidylcholine phospholipase C (Rigden, 2004). The model predicts that GPI-PLD contains a catalytic site, which utilizes five His residues for coordinating Zn^{2+} binding (Rigden, 2004). In agreement with the model, mutation studies have identified four histidine residues (29, 125, 133, and 158) that are essential for GPI-PLD catalytic activity (Raikwar *et al.*, 2005).

1.2.7: Aim one

The considerable lack of information regarding the structure and function of GPI-PLD makes it difficult to understand the role of this protein in the aforementioned medical conditions. A greater understanding of the structure and function of GPI-PLD will provide substantial foundations for eventual structure-based design of therapeutics for the conditions in which it has been implicated. Studies herein will be directed at cloning and expressing GPI-PLD: either human or a closely related homologue. If high levels of soluble recombinant protein are obtainable and the protein preparation is at least 90% pure, crystallisation trials will be set up. If crystallisation trials are successful, the structure of this metalloprotein will be solved by exploitation of its metal-binding ability. From here, the catalytic mechanisms can be derived, and possibly confirmed by site directed mutagenesis.

1.3: Bottlenecks in X-ray crystallography

Unfortunately, even when a pure soluble protein preparation is attainable, producing X-ray diffracting crystals is another common bottleneck. Although there has been a recent surge in efforts to develop crystallographic methods, crystallisation is a complex process that is not well understood. To date, no correlations between protein structure or family and the conditions required for crystallisation have been discovered.

1.3.1: Determination of macromolecular three-dimensional structures

The elucidation of three-dimensional structures of proteins is achieved by two main methods, protein crystallography and nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography can provide information on not only the protein structure, but can also give an insight into its activity, mechanism for recognising and binding substrates and any conformational changes it may undergo. Structure determination by protein crystallography is highly dependent on the production of good quality protein crystals. Protein crystals are able to diffract electrons or neutrons creating a diffraction pattern, which is used to deduce the electron density distribution of the crystalline specimen. As there is no convenient material we can use to produce a lens that can focus X-rays a detour of back-transforming X-ray diffraction data mathematically is required to obtain the desired atomic resolution- level information about the molecular structure. To resolve the crystal structure it is necessary to obtain information on phases of scattered waves in addition to the amplitude (Rupp and Kantardjieff, 2008). The diffraction data, the measured diffraction intensities or reflections that form the data set, return only

the magnitude of the diffracted X-rays. Unfortunately, the phase relations between the reflections that carry the majority of the information regarding the macromolecular structure are lost. A strategy must be employed, as this information is paramount in the construction of an electron density map of the protein. If protein-phasing data cannot be obtained or is poor then the structure cannot be solved. This is known as the phasing problem (Taylor, 2003). To resolve the crystal structure it is necessary to obtain information on phases of scattered waves in addition to the amplitude. Phases cannot be measured directly and there are limitations to the common methods available.

Programs such as ACORN (Yao *et al.*, 2006) can be used, but require very high resolution diffraction data and are limited to structure determination of smaller proteins. Molecular replacement relies on the existence and recognition of a comparable structure that has already been solved (Turkenburg and Dodson, 1996). Phase information from the known structure can be used to resolve the structure of interest. However, there tends to be a bias towards the structure of the reference protein. Multiple isomorphous replacement (MIR) and single/multi-wavelength anomalous diffraction (SAD/MAD) are alternative methods when there are no structures of adequate sequence similarity available (Smyth and Martin, 2000). MIR involves the introduction of heavy atoms or groups to the native protein crystal in order to alter diffraction intensities. Diffraction data from both the native protein crystal and the modified crystal is obtained and used to deduce phase angles and determine the crystal structure (Pidcock and Moore, 2001). The modified crystal must be isomorphous to the native crystal, which involves extensive screening of suitable heavy metals that can bind without destruction or distortion of the protein crystal.

SAD/MAD and MIR methods rely on the binding heavy atoms to the crystal to provide a source of anomalous diffraction (Dauter, 2002). Intrinsically bound metals or sulphur atoms in cysteine or methionine residues can be used to obtain this data. For non-metalloproteins selenium can be incorporated as selenomethionine to achieve anomalous scattering (Pidcock and Moore, 2001). Trivalent lanthanide ions have become an attractive alternative to selenium for use in structural biology. They have superior phasing power (almost 4 X greater) than that of selenium (Silvaggi *et al.*, 2007). In addition to this, the requirement for one methionine for every 80 residues in a given target sequence and the fact that selenomethionine may not replace every methonine means that this method is not ideal. Metal is the more favourable supplement, providing a superior magnitude of anomalous signal and does not require the addition of expensive selenomethionine to media. Nevertheless, proteins do not always contain innate metal binding abilities of high affinity.

1.3.2: Metal-binding DxDxDGyyD tag

Recently, there have been a number of developments for use of trivalent lanthanide ions in structural biology. Not only desirable for their luminescent properties, all lanthanides excluding La^{3+} and Lu^{3+} are paramagnetic. This paramagnetism arises from unpaired electrons residing in chemically unreactive f-orbitals and interactions of lanthanides with a ligand are usually ionic in nature. Natural lanthanide-binding sites are relatively rare. Developments in lanthanide tagging of recombinant proteins have enabled the use of powerful techniques in mainstream structural

biology that were previously only available for paramagnetic metalloproteins. Lanthanide use in NMR spectroscopy has enabled advances in our understanding in protein folding (Felitsky *et al.*, 2008) and the means by which proteins are able to recognise and interact with other proteins (Tang *et al.*, 2006; Volkov *et al.*, 2006), DNA 9Iwahara and Clore, 2006) and small ligands (Tang *et al.*, 2007).

To date, a number of lanthanide-binding tags have been developed based on the characteristic calcium-binding motif of calcium binding proteins (CaBPs). It is well established that calcium-binding sites are capable of binding other metals, particularly lanthanides. The lanthanide group, consisting of elements lanthanum (La^{3+}) through to lutetium (Lu^{3+}), possess similar ionic radii to that of Ca^{2+} (Table 1.1). Furthermore, lanthanides exist as trivalent cations in their most stable oxidation state, which often results in binding of greater affinities to that of Ca^{2+} due to an increased electrostatic attraction between the ligand and ion. Additionally, the repulsive properties of the anionic ligands comprising the calcium-binding site are more effectively counteracted by the increased cationic charge (Snyder *et al.*, 1990). Ye *et al* (2005) demonstrated in experiments using transplanted EF-hand-derived motifs that lanthanides have greater metal-binding affinities than Ca^{2+} with Tb^{3+} and La^{3+} binding more tightly by 10 - 20 fold and 5 - 10 fold, respectively. Lead is also accommodated in DxDxDG-style calcium binding sites in both calmodulin (Kern *et al.*, 2000) and in galactose-binding protein (Vyas *et al.*, 1989) despite its larger ionic radius of 133 pm (Table 1.1).

Table; 1.1: Ionic radii of lanthanides

Table adapted from {Xue, 1988 #214}. * not lanthanides

Lanthanide cation	Ionic radius (nm)
La^{3+}	0.114
Ce^{3+}	0.107
Pr^{3+}	0.106
Nd^{3+}	0.104
Sm^{3+}	0.098
Eu^{3+}	0.098
Gd^{3+}	0.097
Tb^{3+}	0.093
Dy^{3+}	0.092
Ho^{3+}	0.091
Er^{3+}	0.089
Tm^{3+}	0.087
Yb^{3+}	0.086
Lu^{3+}	0.085
Ca^{2+} *	0.100
Mg^{2+} *	0.072

1.3.3: Lanthanide binding tags to date

To date paramagnetic tagging of diamagnetic proteins can be achieved through a covalently linked tag (Dvoretsky *et al.*, 2002; Ikegami *et al.*, 2004; Leonov *et al.*, 2005; Haberz *et al.*, 2006; Rodriguez-Castaneda *et al.*, 2006; Su *et al.*, 2006) or by expression as a fusion protein (Ma and Opella, 2000; Wohnert *et al.*, 2003). More freedom of lanthanide positioning can be achieved through covalent attachment by

use of a thiol group of the target protein. However, it is often more desirable to have the lanthanide ion rigidly attached to the target molecule to avoid problems during crystallisation associated with flexible regions and also make the tag useless for phasing. This can be accomplished using a bulkier tag (Su *et al.*, 2008; Su *et al.*, 2006; Martin *et al.*, 2007) or by affixing the tag via two different attachment sites (Prudencio *et al.*, 2004; Keizers *et al.*, 2007; Vlasie *et al.*, 2007). Most of the LBT tags to date are short polypeptides comprising of between 15-25 amino acids and have not demonstrated high affinity binding of lanthanides (Xie *et al.*, 2007). To date, only Silvaggi *et al* (2007) have provided proof of concept of the applicability of double-LBT for macromolecular crystallographic structure determination.

Lanthanides can also be used in the execution of complex NMR experiments as a result of their paramagnetic variations. Such paramagnetic disparity (Dy, Tb, Tm are classified as highly paramagnetic, Er and Yb moderately paramagnetic and Eu, Ce and Sm have weak paramagnetic properties) means that the extent of knockout of NMR spectra can be varied. This allows spectra to be allocated to different regions of the protein, opening up the possibilities of solving the structures of increasingly larger proteins.

1.3.4: Design of the proposed tag

CaBPs often have significant roles within biological systems (Kretsinger, 1976; McPhalen *et al.*, 1991). Many share a common DxDxDG sequence motif despite having evolved independently and they are functionally diverse (Rigden, 2004).

The most common, and by far the most well studied, calcium-binding motif is a helix-loop-helix structure known as the EF-hand. The ligand-accommodating loop consists of three calcium binding aspartic acid residues making up the DxDxDG sequence motif that is flanked by two α-helices. These residues provide the binding loop with a negative formal charge that is ideal for calcium binding (Kirberger *et al.*, 2008). A conserved Gly at residue 6 enables a 90° turn which brings later ligating amino acids into configuration for ion binding (Godzick and Sander, 1989). This motif is separated from a later aspartate or glutamate residue by 5 residues (in the EF-hand) but can be between 2-65 residues in other proteins. The additional ligand (Asp/Glu) reinforces the interaction with the metal.

Initial investigations were aimed at the design of a new lanthanide binding tag, DxDxDGyyD. Only 2 residues were used in the separation of the later Asp. Two metal-binding tags were developed. LBTA with the sequence (N acetyl)DNDKDGHSD and a second, more hydrophobic tag, LBTB, comprising of the sequence (N acetyl)DADKDGWAD. Both LBTs are smaller in size, consisting of only 10 amino acid residues and could be added to the pool of LBTs already available. It is well understood that the addition of a fusion to a target protein can have dire effects on protein stability, solubility, expression and the success of crystallisation. Unfortunately, such consequences are unpredictable. The majority of proteins fold correctly with the common His-tag (Derewenda, 2004). The proposed tag was so small that it is unlikely to have serious repercussions for structure, function and crystallisation of the target protein. It is hoped that this small, and hopefully compact, LBT will provide an approach to phasing where other bulkier tags are not suitable.

Via the modification of the vector pET28a of the pET vector series a derivative containing the metal binding tag, DADKDGWAD was produced. The region of the vector encoding a C-terminal His-tag was substituted for a region encoding the above tag, followed by a stop codon. Any target sequence can be cloned into the modified vector to be expressed with the C-terminal metal-binding tag. Asymmetric diadenosine-5',5'''-P1,P4-tetraphosphate (Ap_4A) Hydrolase was employed as the test/model protein for investigations using the metal-binding LBTB. This mutant is ideal as it was used previously in NMR studies (Swarbrick *et al.*, 2005). This protein was cloned into the modified vector for expression with this C-terminal metal-binding tag as it is has a solvent-exposed C-terminus. This protein was used in the development of this tag, DxDxDGyyD, to establish metal-binding affinities and to assess consequences of tag association on the test/model protein. The availability of the Ap_4A-H structure provides an additional advantage, in that comparisons could be made between the tagged and non-tagged structures providing additional information to determine if fusion with the LBT causes conformational changes on the protein structure.

As mentioned earlier, lanthanide-binding tags can be used, and are currently available, for use in NMR. However, studies herein are focused on how this tag can be used in solving the phasing problem in protein crystallography. The metal-binding tag could be used to complement current approaches used in obtaining information about phases. For the MAD approach the target protein would be crystallised in the presence of saturating concentrations of a lanthanide that can elicit significant anomalous scattering and bind with high affinity to the LBT. Datasets will be obtained at the appropriate wavelengths and the anomalous

scattering of the bound metal/lanthanide can be used to calculate the required phasing information. Using the MIR approach diffraction data will be collected both before, and after occupation of the metal-binding site of the fused tag. The mass difference required by IR can be acquired by growing the crystal in the absence of metal and then soaking in a metal ion solution to achieve binding site occupation or by growing crystals in the presence of heavy metal solution and then removing metal from the metal binding site by soaking in a large volume of metal-free solution or by use of a metal chelator.

1.3.5: Aim two

A lanthanide-binding tag will be designed and an expression vector engineered for generic heterologous expression of proteins in structural biology. The metal binding ability of the tag will be ascertained. The conjugating of the proposed tag should not have any serious consequences on expression, solubility and protein stability. However, this will need to be confirmed. Additionally, studies will need to be conducted to determine the effects of tag addition on protein crystallisation. If good diffracting crystals are obtainable, attempts will be made to obtain phasing data for both the non-tagged and the tagged protein and comparisons will be made. Ideally, if studies involving GPI-PLD are successful, expression with this lanthanide binding tag will be tried. Comparisons could then be made between phasing information obtained using the proteins innate metal-binding ability and that obtained using the lanthanide binding tag.

2.0: Materials and methods

2.1: General Methods

All reagents were of analytical grade and from VWR [Lutterworth, UK] or Sigma [Poole, UK] unless stated otherwise in the text.

2.1.1: SDS-PAGE

SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) was performed essentially according to Laemmli (1970) using a 0.75 mm, 12% polyacrylamide gel. The resolving gel solution was composed of 3.5 ml distilled water, 4.9 ml 30% acrylamide/bis-acrylamide stock (Severn Biotech, UK), 2.5 ml 3 M Tris-Cl (pH 8.85), 0.10 ml 10% (w/v) SDS and polymerisation initiators 0.10 ml 50 mg/ml ammonium persulphate, 10 µl TEMED (N, N, N', N'-tetramethyl-ethylenediamine. The mixture was gently swirled and quickly poured between glass plates of the Mini-Protean II gel electrophoresis apparatus (BioRad) to form two 0.75 mm thick mini-gels. The poured gel solution was overlaid with water-saturated isobutanol and left at room temperature for approximately 30 min to allow for complete polymerisation. Once gels had set, gel surfaces were rinsed with distilled water to remove the isobutanol overlay. Any water remaining was removed using filter paper before a stacking gel was applied. Stacking gel solution was composed of 3.7 ml water, 1.3 ml 30% acrylamide/bis-acrylamide stock (Severn Biotech, UK), 1 ml 1.25 M Tris-Cl (pH 6.8), 0.10 ml 10% (w/v) SDS and polymerisation initiators 0.10 ml 50 mg/ml ammonium persulphate, 20 µl TEMED. Again the mixture was gently swirled and then quickly pipetted between the glass plates onto the resolving gel surface. A plastic comb was inserted into the stacking

gel solution and removed after polymerisation to provide loading wells for samples and protein standards. Wells were rinsed with water before gels were mounted in the electrophoresis apparatus. The inner chamber and the bottom reservoir of the tank were filled with SDS gel running buffer [50 mM Tris, 192 mM glycine, 0.1% (w/v) SDS].

Protein samples to be analysed were prepared by the addition of the appropriate amount of 5X SDS loading buffer [50% (v/v) glycerol, 10 % (w/v) SDS, 25% (v/v) 2-mercaptoethanol, 0.5% (w/v) bromophenol blue in 315 mM Tris-Cl, pH 6.8] and denatured by heating at 100°C for 3 min just prior to loading. A centrifugation step (13 000 x g for 1 min) was used to remove debris from the sample. Wells were loaded with up to 15 µl of each protein sample alongside 7.5 µl of pre-stained protein standards of known molecular weight (Sigma, UK). Electrophoresis was performed at 200 V, 30 mA (per gel) until the dye front reached the end of the resolving gel at which point gels were removed from glass plates. The stacking gels were removed and discarded before proteins were stained or transferred as described in subsequent sections.

2.1.1.1: Detection of proteins in SDS-polyacrylamide gels with Coomassie Brilliant Blue R-250

Gels containing proteins to be stained with Coomassie Brilliant Blue R-250 were transferred to a solution of 0.1% (w/v) Coomassie Brilliant Blue R-250 in a methanol:water:acetic acid mixture (4:5:1, by volume). Gels were left to stain on a slowly rotating platform for 1-2 h at room temperature. The removal of background colour was achieved using a (3:6:1, by volume) methanol-acid mixture

in the absence of Coomassie Brilliant Blue R-250. The destain solution was changed several times until an acceptable level of protein detection was achieved.

2.1.1.2: Detection of proteins in SDS-polyacrylamide gels with silver nitrate

Silver staining was performed essentially according to the method of Heukeshoven and Dernick (Heukeshoven and Dernick, 1985). 0.75 mm gels containing proteins to be stained with silver were first incubated in fixing solution [40% ethanol, 10% acetic acid] for 60 min on a slowly rotating platform for 1-2 h at room temperature. Once proteins were fixed in the gel, a series of washing steps were performed (2 X 5 min in 10% ethanol, followed by 3 X 5 min in water). Gels were then left to stain in 0.2% $AgNO_3$ on a slowly rotating platform for 30 min at room temperature. After this incubation the silver solution was carefully poured off and developer solution [2.5% sodium carbonate, 80 µl/100 ml formaldehyde (37%)] was added until a brown colour developed in the solution. The brown solution was poured off and fresh developer was added. This step was repeated until the desired intensity was obtained. The gel was then rinsed briefly with water before adding 1% acetic acid and incubating with gentle shaking for 5 min. The gel was then thoroughly washed with copious amounts of water (6 X 5 min). The removal of background colour was achieved using reducing solution [20 mM sodium thiosulphate, 5 mM potassium ferricyanide, 5 mM sodium carbonate in 200 ml water]. The gel was briefly rinsed in this solution until background colour cleared and then quickly washed with copious amounts of water. This allowed for detection of protein bands between 20 – 50 ng. If greater sensitivity was required the staining process was repeated from the addition of silver solution. This allowed for detection of proteins of only 1-2 ng amounts.

2.1.2: Western blotting

Proteins to be analysed by western blotting were first separated by SDS-PAGE as described previously then electrophoretically transferred from the gel to a polyvinylidene fluoride [PVDF] membrane. After the initial electrophoresis, the resolving gel (- stacking gel), a gel-size piece of PVDF membrane (previously rinsed with methanol) and two pieces of Whatman 3MM filter paper were "sandwiched" between two transfer pads using the cassette of the Mini-Transblot transfer apparatus (BioRad). The cassette and its contents were assembled while immersed in transfer buffer [1 litre: 0.2 M glycine, 25 mM Tris, 200 ml methanol] and compressed ensuring no bubbles were present. This assembly was then immersed in transfer buffer between the parallel electrodes of the transfer apparatus. A current (30 mA/gel) was passed through the transfer assembly for 1 h. Following electrophoresis, the membrane was submerged in blocking buffer [Tris-buffered saline (TBS: 20 mM Tris-Cl, pH 7.5; 150 mM NaCl), 1% (w/v) dried milk, 0.075% (v/v) Tween 20] for 1 h at room temperature. Before antibody staining, excess blocking buffer was removed by 2 X 30 s washes with TTBS [TBS plus 0.075% (v/v) Tween 20]. Immunoblotting was then performed by incubation with the addition of the required primary antibody diluted in TTBS plus 0.5% (w/v) dried milk at (4°C, 16 h). The membrane was then washed in TTBS (2 X 30 s, then 4 x 5 min) to remove unbound antibody. Incubation with a 1:2000 dilution of horseradish peroxidase-conjugated secondary antibody directed against the IgG of the species that provided the primary antibody proceeded for a minimum of 2 h at room temperature. After incubation the membrane was prepared for detection by washing in TBS 3 X 5 min. Detection using diaminobenzidine (DAB) as a substrate for horseradish peroxidase entailed immersing the blot in a solution of 20

ml TBS containing 0.5 mg/ml DAB and 0.5% (v/v) H_2O_2 until sufficient colour development had occurred. The final blot was rinsed with copious amounts of water then air-dried. When greater sensitivity was required, the method of enhanced chemical luminescence (ECL) was used. The membrane was exposed to equal volumes of both ECL reagents (A and B) (Amersham) for 1 min. Excess reagent was removed by blotting the edge of the membrane before being fixed in Saranwrap and positioned into the developing cassette protein side up. In a dark room, a piece of membrane/blot-size hyperfilm ECL (Amersham) was cut and placed deliberately onto the membrane. A 30 s exposure was typical before film development, which involved a 2 min wash in developer, followed by 3 min in fixer before a final wash in water.

2.1.3: Protein quantification

2.1.3.1: Bradford assay

Where possible, protein concentration was determined by the method of Bradford (Bradford, 1976) using Protein Assay Reagent (Bio-Rad). Protein standards were prepared containing 0.2 – 2.0 mg/ml of bovine gamma globulin (BGG). The Bradford reagent was diluted 1:4 (v/v) with water and filtered before being added to each 10 µl protein sample to make up to 1 ml. Samples were incubated at room temperature for 10 min to allow binding of the dye to the protein. Colour development was assayed by measuring absorbance at 595 nm. Blanks of 10 µl buffer plus 990 µl Bradford reagent were used to zero the spectrophotometer. A standard curve of absorbance versus protein concentration was generated. This was

used to provide an estimation of protein concentration of samples based on absorbance values.

2.1.3.2: Ultraviolet absorption

This method was used when quantification of purified protein concentration was required. A theoretical extinction coefficient for each protein was determined using ProtParam (Gasteiger *et al.*, 2003). Protein concentrations were calculated using the theoretical extinction coefficient (22710 M^{-1} cm^{-1} for Ap_4A-H and 28210 M^{-1} cm^{-1} for Ap_4A-H-LBTB) and the absorbance values at 280 nm for a given sample.

2.1.3.3: BCA method

The colorimetric detection and quantitation of total protein using the bicinchoninic acid (BCA) was performed using a BCA Protein Assay Kit (Pierce) according to manufacturers instructions. A series of dilutions of bovine serum albumin (BSA) were used as protein standards. Dilutions of known concentration (0.25 mg/ml – 2 mg/ml) were assayed alongside the protein sample (unknown concentration) so that a standard curve could be created to determine the concentration of the protein sample. 2.0 ml of the working reagent (WR) was prepared for each sample (50:1, Reagent A:B). For this assay 2.0 ml of the WR was added to 0.1 ml of each standard and unknown sample replicate. Each tube was mixed, covered and incubated at 37°C for 30 min. After this incubation, all tubes were left to cool to room temperature. The absorbance values were taken using a spectrophotometer at 562 nm set to zero using cuvette filled only with water. The average absorbance (at 562 nm) of the blank standard replicates (100 µl 10 mM Tris-HCl, pH 8.0) was subtracted from absorbances taken for all other standards and the purified protein sample replicates. A standard curve was constructed by plotting the average blank-

corrected 562 nm measurement for each BSA standard versus its concentration in µg/ml. This standard curve was used to determine the protein concentration of each unknown sample.

2.1.4: In gel trypsin digestion and peptide extraction from Coomassie stained SDS-polyacrylamide gels

In-gel digestion was carried out using an adaptation of the method of Rosenfeld *et al* (Rosenfeld *et al.*, 1992). Protein bands chosen for analysis by mass spectrometery were excised from a SDS-polyacrylamide gel using a clean scalpel blade. Each gel slice was washed (2 x 30 min) with 50% acetonitrile, 0.2 M ammonium bicarbonate, 2 mM urea, pH 7.8, then partially dried in a rotary evaporator. The slices were re-hydrated using 50 mM ammonium bicarbonate pH 7.8 containing trypsin [0.1 µg per gel piece] and incubated at 37°C for 16 h. Excess buffer was then removed to a second microcentrifuge tube and peptides were extracted from the gel slices in three steps; by washing with 0.1% trifluoroacetic acid (TFA), 30% acetonitrile in 0.1% TFA and then 60% acetonitrile in 0.1% TFA. The latter and the excess buffer were pooled and concentrated by centrifugal evaporation, ready for Mass Spectrometric analysis.

2.1.5: Concentration of proteins

2.1.5.1: Methanol precipitation

A 5-fold excess of methanol (100%) was added to the protein solution (e.g. fermentation media), mixed and incubated at -20°C for 16 h. Protein precipitate was pelleted by centrifugation at 5 000 x g for 15 min. The supernatant was

carefully decanted and the pellet left with tube inverted to air for 5 min. 1.5 ml ice-cold methanol (90%) was used to wash the pellet via re-suspension and re-centrifugation. The methanol was removed and discarded before protein samples were dried under a vacuum. Pellets were re-suspended in an equal volume of 2X SDS-PAGE sample buffer for SDS-PAGE analysis.

2.1.5.2: Vivaspin protein concentrator columns

The protein sample/fermentation medium was added directly to the Vivaspin MWCO 5000 spin concentrator column (GE Healthcare) and concentrated using centrifugation at 4000 x g. 5X SDS-PAGE sample buffer was added to concentrated protein (1 part sample buffer to 4 parts protein) for SDS-PAGE analysis.

2.1.6: RP-HPLC peptide mapping

Internal peptides for sequencing were generated using an adaptation of the 'one-tube' reduction, carboxymethylation and digestion method of Stone *et al* (1992). 20µg protein was dried down by rotary evaporation and 25 µL of 8 M urea in 400 mM ammonium bicarbonate, pH 8.0, was added followed by 2.5 µL 45 mM dithiothreitol (DTT). After incubating at 50°C for 15 min, the sample was cooled to room temperature and 2.5 µL 0.1 M iodoacetic acid was added and left for 15 min in the dark. The mixture was diluted 4-fold, 1µg of sequencing grade trypsin was then added and the sample left at 37°C overnight to digest. The resulting peptide mix was made to 0.5 % in TFA and applied to a PE-Biosystems PepMap C18 RP-HPLC column (100 x 2.1 mm) equilibrated in 0.08% TFA. Peptides were separated

with a 95 min gradient of 0 - 64% acetonitrile in 0.08% TFA; elution was monitored at 214 nm.

2.1.7: Mass spectrometry and protein sequencing

All mass spectrometric analyses were performed using a Matrix assisted laser desorption/ionization - time of flight (MALDI-Tof) instrument (Waters-Micromass). Following trypsin digestion in solution or in-gel, peptides were cleaned up using reverse phase chromatography and then analysed using a saturated solution of alpha-cyano-4 hydroxycinnaminic acid in 50% acetonitrile/0.1% trifluoroacetic acid. Peptides were selected in the mass range of 1000 – 3000 Da. Some proteins immobilised onto PVDF following western blotting were sequenced directly by Edman degradation on a model 471A Protein Sequencer [Applied Biosystems].

2.2: Enzyme assays

2.2.1: Assay for GPI-anchor hydrolysing activity with membrane form acetylcholinesterase (mf-AChE) as a substrate.

The GPI-PLD activity assay, adapted from previously published work (Bordier, 1981; Ellman, 1956), consists of the following steps: cleavage of mf-AChE with the production of soluble AChE (s-AChE); phase separation by Triton X-114 (separating mf-AChE from s-AChE); and finally the s-AChE assay.

2.2.1.1: Preparation of red blood cell ghosts

GPI-PLD anchor-degrading activity was assayed using membrane form, GPI-anchored-AChE from bovine erythrocytes as substrate. These red blood cell ghosts

were obtained from First Link (Birmingham, UK). A 200 µl aliquot of red blood cell 'ghosts' was centrifuged at 10000 x g for 3 min. The supernatant was discarded and the pellet was washed with 100 µl cold water, then centrifuged at 10000 x g for an additional 3 min. The pellet was re-suspended in 1 ml Tris-Cl (10 mM) supplemented with Triton X-100 and incubated on ice for 15 min with shaking. The mixture was centrifuged at 10000 x g for 3 min to pellet remaining cell debris and the supernatant was diluted 1/20 prior to use in the assay.

2.2.1.2: Cleavage of membrane form, GPI-anchored-AChE.

Samples containing GPI-PLD (chromatography fractions or mammalian serum) were incubated at 37°C for 2 h with 200 µl of 1/20 diluted GPI-anchored-AChE in a 1.5 ml microfuge tube.

2.2.1.3: Phase separation.

The product of the reaction, s-AChE, was separated from the substrate by phase separation using Triton X-114 according to Bordier (Bordier, 1981). The incubation mixture was mixed on ice with an equal volume (200 µl) of pre-condensed Triton X-114. Pre-condensation of Triton X-114 was carried out according to the method of Bordier (1981), however, the condensation reaction was performed only two times, not three. The microfuge tube was incubated on ice for 5 min followed by incubation at 37°C for 10 min, mixing thoroughly after 5 min. Centrifugation at 10000 x g for 2 min using a bench-top centrifuge was used to separate the two phases so that 100 µl of the upper aqueous layer could be transferred to a 1.5 ml microfuge tube for the final stage of the assay.

2.2.1.4: s-AChE assay.

To the 100 µl upper aqueous phase, 0.8 ml sodium phosphate buffer (100 mM, pH 7.5) containing 0.2 mM dithiobis-(5,5'-nitrobenzoic acid) (DTNB) and 0.25 mM acetylthiocholine bromide were added. The solution was mixed thoroughly and incubated for 30 min at 37°C. The absorbance was then measured at 412 nm using a spectrophotometer. Anchor degrading activity by GPI-PLD was measured indirectly via thionitrobenzoic acid (TNB) produced in the s-AChE assay using the TNB extinction coefficient of 13 700 M^{-1} cm^{-1} at 412 nm according to Ellman *et al* (1956).

2.2.2: Asymmetric diadenosine-5',5'''-P1,P4-tetraphosphate (Ap_4A) Hydrolase activity assay

The assay was performed using a luciferase-based bioluminescence assay for detection of the ATP product (Prescot *et al.*, 1989). Purified Ap_4A-H and Ap_4A-H -LBT samples were diluted 1/1000 and incubated on ice until required. A series of assays were set up by the addition of 70 µl of the reaction buffer [70 mM Na-HEPES, pH 7.8, 5 mM $MgCl_2$] and 4 µl of substrate [250 mM Ap_4A] in 0.5 ml glass tube and allowed to equilibrate to 25°C by incubation for 5 min in a water bath. Next, 25 µl of ATP-monitoring reagent (Lonza) was added followed by the immediate addition of 1 µl protein sample (Ap_4A-H or Ap_4A-H-LBT). The increase in luminescence was measured (mV/min) over a 5 min period at 25°C with an LKB 1250 luminometer in order to compare Ap_4A-H and Ap_4A-H-LBT samples. For each protein sample the assay was performed 5 times and mean of

luminescence was taken for each.

2.2.3: Isothermal titration calorimetry (ITC)

All ITC experiments were performed using a VP-ITC MicroCalorimeter according to manufacturers manual. Initial titrations were performed to optimise the experiment and to determine the amounts of reactions required for the ITC. Prior to use, both the sample cell and syringe were thoroughly cleaned using degassed water and then pre-rinsed with buffer. Reactants were degassed using a Thermo Vac (Microcal) which was set at 22°C. All ITC runs were set up and controlled using VPViewer. This software was used to program the temperature of the cell, which was set at 25°C and to set injections of 10 µl to be released from the syringe every 300 s. The calorimeter sample cell was filled with a 2 ml degassed sample of peptide or protein (50 µM) in 100 mM Na-HEPES, 150 mM KCl, (pH 7.0) ensuring that no bubbles were incorporated into the cell. Next, the syringe was loaded with 250 µl degassed metal chloride (1.2 mM) dissolved in buffer identical to that used for dissolving the peptide or dialysed protein. A number of purge and re-fill cycles were used to be confident that no air had entered the syringe during loading. The closed syringe was then carefully placed into the cell for the titrations to begin. The entire experiment was computer controlled. As ligand was injected into the sample cell the power was adjusted accordingly to maintain equal temperature between the sample cell and the reference cell. This ultimately provided a record in a series of peaks representing the heat change (enthalpy) associated with injection of ligand into the peptide/protein. The area under each peak represented the total heat associated with the binding of all the material in that injection. A ligand (metal chloride) into buffer ITC experiment was executed and the data subtracted

from final data to eliminate the background enthalpies from buffer interference. ITC data analysis, including setting the baseline, integration and model curve fitting was achieved using Origin® according to the manufacturers guide.

2.3: Molecular biology methods

All reagents were of molecular biology grade and from VWR [Lutterworth, UK] or Sigma [Poole, UK] unless stated otherwise in the text.

2.3.1: Hybridization of oligonucleotides

The oligonucleotides for lanthanide-binding tag construction were combined in water, heated to 95°C for 2 min, and then allowed to cool to ambient temperature for 1 h. Annealed oligonucleotides were added to the expression plasmid at 10-, 100-, and 1000-fold molar excesses. Ligation was achieved using 0.1 U T4 DNA Ligase (Invitrogen) in a total volume of 50 µl using ligase buffer (Invitrogen) at 24°C for 1 h. A second ligation, containing only plasmid DNA and no annealed oligonucleotide insert, was included as a control.

2.3.2: Polymerase chain reaction (PCR)

PCR conditions were first optimised by use of BioRed PCR mix (Bioline) using a gradient PCR within a temperature range, typically 45-65°C (unless stated otherwise), facilitated temperature optimisation for the annealing temperature of the PCR experiment. Target DNA amplification was performed using the Mastercycler® gradient according to the following protocol: 1 cycle of

denaturation at 94°C for 3 min, an amplification of 29 cycles, each consisting of a 45 s denaturation at 94°C, a 45 s annealing at 45-65°C and a 2 min extension at 72°C. This was followed by a final 10 min extension at 72°C.

PCR for molecular cloning was performed with 0.5 μg/50μl of target DNA, 10 μl Platinum Pfx DNA polymerase 10X buffer (Invitrogen), 1 μl $MgSO_4$ (final concentration 1 mM) (Invitrogen), 1.5 μl dNTP mix (Promega) (10 mM of each dNTP), 10 μM (final concentration) each of 5'- and 3'-primers and 1 U Platinum Pfx DNA polymerase (Invitrogen) in a 0.2 ml domed micro-tube containing a total 50 μl PCR volume. DMSO (1% of the final volume) was required to prevent the formation of secondary structures of primers. Presence of the appropriate PCR product was confirmed by determination of molecular weight by agarose gel electrophoresis using a 1% (w/v) agarose gel.

2.3.3: DNA purification

PCR products were purified using the QIAquick PCR purification kit (Qiagen). In some cases, where it was not possible to purify DNA digests using QIAquick gel extraction and purification kit (Qiagen) the PCR purification kit was used instead. 5 volumes of Buffer PB were added to 1 volume of the PCR sample or digested DNA and mixed by inversion. The pH of the mixture was confirmed using pH strips to ensure the mixture was at an optimal pH (pH 6.0-7.0) for DNA binding to the spin column silica membrane. The mixture was then transferred to a QIAquick spin column in a provided 2 ml collection tube and centrifuged (10 000 x g, 45 s). The flow-through was removed from the collection column and discarded before the QIAquick column was placed back into the same tube. The membrane (now

with bound DNA) was washed by the addition of 0.75 ml Buffer PE to the QIAquick column and centrifugation (10000 x g, 45 s). Again the flow-through was discarded and the QIAquick column was placed back in the same collection tube. An additional centrifugation (10000 x g, 45 s) was used to remove residual ethanol from Buffer PE and the flow-through was discarded. The QIAquick column was then placed in a clean 1.5 ml microcentrifuge tube so that the DNA could be eluted by adding 30 µl Buffer EB (10 mM Tris·Cl, pH 8.5) directly to the center of the QIAquick membrane and left to stand for 2 min. A final centrifugation (1000 x g, 1 min) was used to elute the DNA from the column into the 1.5 ml microcentrifuge tube. Purified samples were used immediately if possible or stored at -20°C until required or -80°C for long-term storage.

2.3.4: Agarose gel electrophoresis

Generally, a small 1% (w/v) agarose gel (base 7 x 7 cm, 1.0 cm thick) was used for separation of DNA. 0.25 g of agarose was added to 25 ml of 1X TBE [0.089 M Tris Base; 0.089 M, Boric Acid, 2 mM EDTA, pH 8.3] buffer in a conical flask. The agarose was dissolved by heating in a microwave oven on full power for approximately 2 x 30 s until all the agarose particles were completely dissolved. The agarose gel solution was allowed to cool to 45°C before it was carefully poured into a gel tray and the ends sealed with masking tape, ensuring any bubbles were removed. The gel comb(s) was placed into the appropriate slot(s) of the tray so that the sample wells were near the cathode. After the gel had completely set (approximately 30 min at room temperature) the tank was filled with 1X TBE buffer (0.5 M) so that the gel was covered to a depth of at least 1 mm and the gel combs gently removed. 5.0 µl of 6 X loading buffer (Fermentas) was mixed with

25.0 µl of appropriate DNA sample and as much sample as needed was loaded into the well (typically between 5 µl – 10 µl for PCR DNA). 6 µl of the 1Kb generuler (SM0311, Fermentas) was loaded into the first and last wells leaving the outermost wells empty where it was possible. Once all samples were loaded onto the agarose gel, electrophoresis was performed at 50 mA (constant current), 100-150V until the loading buffer migrated two thirds of the way through the gel. After electrophoresis the gel was removed from the gel tray and incubated in an ethidium bromide solution (0.5 µg/ml) until the DNA had taken up the dye before the DNA was visualized by use of a UV transilluminator. If DNA was to be recovered from the gel for further use, a long wave UV source was used for visualization for the shortest possible time.

2.3.5: DNA recovery from agarose gels

For DNA recovery, deeper wells were created in the 1% (w/v) agarose gel using the appropriate well comb. This allowed for a larger volume of DNA sample (up to 50 µl) to be loaded and separated on the agarose gel. The fragment of interest was cut out of the gel with a sterile scalpel blade using short wavelength UV light for visualization to prevent the introduction of nicks into the DNA. Once the band was excised the DNA was recovered and purified using QIAquick gel extraction and purification kit (Qiagen) following the manufacturer's procedure. This involved the dissolving of the agarose slice containing the DNA in 3 volumes of QG solution at 50°C. Once dissolved, 1 gel volume of isopropanol was added to the solution, which was then immediately mixed and transferred a QIAquickspin column in a 2 ml collection tube. This allowed for the binding of the DNA from the melted

agarose to a silica-gel membrane using centrifugation. A series of washes using QIAquick PE buffer and centrifugation (13000 x g, 1 min) was used to wash the DNA. The flow-through was discarded after each centrifugation except in the case of the final elution of the DNA using QIAquick EB buffer, as the DNA eluate was collected in a 1.5 ml microfuge tube and labeled accordingly. Estimations of the approximate concentration of the DNA obtained was achieved by running 1 µl of the eluate plus 5 µl 6X loading buffer (Fermentas) on a 1% (w/v) agarose gel against the 1Kb generuler (SM0311, Fermentas).

2.3.6: Restriction digests and ligation

Both purified PCR product and purified expression plasmid were digested with complementary restriction enzymes (New England Biolabs). Generally, 5 µl of enzyme was added to 1 µg of purified DNA in a final volume of 100 µl using an appropriate 1X NEBuffer that ensured maximum performance of the restriction enzymes used in the digest. Incubations were for 1 h at the recommended temperature, typically 37°C. Bovine serum albumin (BSA) was added to the reaction at a final concentration of 100 mg/ml (1X) when it was required. The digested DNA was purified by agarose gel electrophoresis and recovered using QIAquick Gel Extraction and purification kit (Qiagen) as detailed earlier. Ligation was achieved using 0.1 U T4 DNA ligase (Invitrogen) in a total volume of 20 µl using ligase buffer (Invitrogen) at 24°C for 1 h. A typical ligation reaction (1:1) consisted of approximately a 2-fold excess of insert DNA to plasmid DNA. A second ligation (1:0), containing only plasmid DNA and no DNA insert, was carried out as a control.

2.3.7: DNA dephosphorylation

As an alternative to gel extraction and purification of DNA, the digested plasmid was dephosphorylated by use of Antarctic phosphatase (NEB). The removal of 5' phosphate groups required by the DNA ligase prevented self-ligation of the plasmid, thereby reducing the background when cloning. This protocol was carried out according to NEB guidelines. 80 µl of Antarctic phosphatase buffer was added to the 100 µl digest along with 1 µl of Antarctic phosphatase. The mixture was gently inverted, and then incubated at 37°C for 15 min. The reaction was purified using a PCR purification kit (Qiagen) as detailed in section 2.3.3 to remove contaminating enzymes before ligation.

2.3.8: Preparation of transformation-competent *E.coli* cells using calcium chloride.

Using an inoculating loop, a glycerol stock culture of the desired *E. coli* (expression or non-expression strain, namely XL1-BLUE (Stratagene) was streaked out onto an LB-agar plate containing selective antibiotic (where it was necessary). Plates were incubated at 37°C for 16-20 h before a single colony was used to inoculate 100 ml LB-broth (containing selective antibiotic where required). The 100 ml culture was incubated at 37°C in a gyratory incubator (190 rpm) until a cell density of OD_{600} = 0.5 was achieved. At this point, the culture was split between 2 sterile, pre-chilled 50 ml centrifuge tubes and incubated on ice for 10 min. Cells were harvested by centrifugation at 4500 x g for 10 min at 4°C. The medium was decanted from cell pellets before re-suspending each pellet in 10 ml of ice-cold $CaCl_2$ (0.1 M) by gentle mixing. A second centrifugation (4 500 x g for 10 min at 4°C) followed and

when the supernatant was removed cell pellets were re-suspended in a smaller volume (2 ml) of ice-cold 0.1 M $CaCl_2$. Re-suspensions were combined in a sterile 25 ml universal tube with the addition of glycerol at a final concentration of 15% (v/v). This suspension was gently mixed. Aliquots of 200 µl were transferred to 1.5 ml microfuge tubes. Tubes were flash-frozen using liquid N_2 and stored at -80°C until required for bacterial transformation.

2.3.9: Bacterial transformation

A recombinant plasmid (expression plasmid containing DNA insert) was used to transform the desired strain of competent *E. coli*. 10 µl of the ligation mix or 50 ng plasmid in a volume of 10 µl was transferred to 200 µl competent *E. coli* cells previously thawed on ice for 10 min. Transformation cultures were left on ice for 20 - 30 min followed by a heat shock step at 42°C for 2 min and then returned to ice for another 5 min. 700 µl LB media was added to each transformation before being incubated with agitation at 37°C for 1 h. A centrifugation step (14000 x g, 3 min) was used to concentrate the transformation. 700 µl LB media was removed and the pellet was re-suspended in the remaining medium and plated onto selective LB agar. Plates were incubated at 37°C for 16 h. Plates were examined to determine if the transformation was successful. The absence of colonies on the control plate (*E. coli* transformed with a ligation of digested plasmid containing no insert DNA) confirmed that only colonies containing the antibiotic resistant plasmid, thus the target insert, were able to grow. Positive colonies were isolated and grown-up overnight in 10 ml LB-broth (plus the appropriate antibiotic, generally at 50 µg/ml) at 37°C in order to carry out a colony PCR. Glycerol stocks

of cells harbouring the recombinant plasmid (or expression plasmid) were made up for storage of the plasmid containing the target insert (in the case of XL1-Blue *E. coli*) or for future use for expression strains of *E. coli* and stored at -80°C.

2.3.10: Assessing validity of transformants by colony-PCR and sequencing

A colony PCR was used to prove the target insert was present in transformants. BioRed Biomix (BioRad) was used to amplify the DNA insert using 5 µl overnight colony cultures and 50 pmol each of 5'- and 3'-primers in a 0.2 ml domed micro-tube containing a total 50 µl PCR plus DMSO (1%). PCR amplification was performed using the Mastercycler® gradient according to the same protocol as detailed in section 2.3.2. The presence of insert was determined by agarose gel electrophoresis. Colonies containing the desired insert of the correct size had to be sequenced before being transformed into a competent *E. coli* strain capable of expressing the ingested gene upon stimulation with IPTG. A 1 ml sample of each colony culture was sent in a 1.5 ml screw-cap micro-centrifuge tube to Cogenics Ltd (Essex, U.K) for sequencing.

2.3.11: Plasmid DNA isolation

Isolation of recombinant plasmid DNA from XL1-Blue cells was achieved using a Plasmid Midi preparation kit (Qiagen) according to the Qiagen handbook. The Qiagen protocol is based on a modified alkaline lysis procedure. A single colony from a freshly streaked selective LB-agar plate was picked and used to inoculate a starter culture of 2 ml LB medium containing the appropriate selective antibiotic. This culture was incubated for 8 h at 37°C with vigorous shaking (approx. 300 rpm)

before an aliquot of 200 µl was used to inoculate 100 ml LB-antibiotic selective medium. An incubation at 37°C for 16 h was required so that the culture could reach a high cell density. At this point, cells were harvested by centrifugation at 6 000 x g for 15 min, then re-suspended in buffer P1 of the Qiagen kit. Incubation of the cell suspension with buffer P2 at room temperature causes cell lysis. The solution was then neutralized by the addition of buffer P3 and incubating on ice for 15 min. The solution was subjected to two centrifugation steps at 20000 x g for 30 min to pellet particulate material and thereby prevent clogging of the QIAGEN-tip which can reduce or eliminate gravity flow. The plasmid DNA suspension was applied to the QIAGEN-tip and bound to the QIAGEN anion-exchange resin under appropriate low-salt and pH conditions provided by the buffers. RNA, proteins, dyes, and low molecular weight impurities were removed by a medium-salt wash using Buffer QC (2 x 10 ml). Plasmid DNA was finally eluted in 5 ml of a high-salt buffer QF and then concentrated and desalted by isopropanol precipitation. This involved the addition of 3.5 ml to the 5 ml eluate and mixing by inversion followed by centrifugation at 15000 x g at 4 °C for 30 min. A final wash with 70% ethanol was required before the plasmid pellet was re-suspended in 80 µl TE buffer [10 mM Tris-Cl, 1 mM EDTA (pH7.5)] by overnight incubation at 4°C.

2.3.12: Induction of target protein expression using an *E. coli* expression host

A single transformed *E.coli* colony was selected from an LB-agar plate (plus selective antibiotic(s)) (stored for no more than 48 h) and used to inoculate 5 ml LB broth containing the required selective antibiotic(s) and grown at 37°C overnight with shaking (200 rpm). 200 µl of the overnight culture was used to inoculate 10

ml LB broth (supplemented with the selective antibiotic(s)), which was incubated at 37°C with shaking until the required cell density was reached, typically OD_{600} = 0.7. The inducer, isopropyl β-D-thiogalactopyranoside (IPTG), was added at a final concentration of 1 mM unless stated otherwise. Cultures were returned for incubation at the desired temperature (15 - 37°C) with shaking for the required length of time to achieve maximum expression of the induced target. IPTG-induced expression of the target protein was monitored by the removal of 1 ml samples from incubations at various intervals. SDS-PAGE was used to analyse expression and solubility states of the target protein.

2.3.13: Cell lysis for preparation of samples for SDS-PAGE

2.3.13.1: Sonication

1 ml samples were centrifuged at 12000 x g for 2 min and the pellet was re-suspended in 100 µl TBS [20 mM Tris-Cl, 0.1 M NaCl]. Cells were sonicated for 16 s, then centrifuged at 13000 x g for 3 min to separate the soluble, supernatant fraction from the insoluble, pellet fraction. For larger cultures of 1 l, cells were harvested by centrifugation at 12000 x g for 5 min then resuspended in 10 ml TBS. Cells were sonicated for 5 x 30 s, incubating on ice in between each round of sonication. SDS-PAGE was used to determine the presence of a protein product of the correct size or to identify inclusion body formation. To prepare for SDS-PAGE, the pellet was re-suspended in 100 µl 1X sample buffer and 5 µl 5X sample buffer was added to 20 µl supernatant fraction.

2.3.13.2: B-PER protein extraction reagent

Cells were harvested from a 1 ml sample of induced culture by centrifugation at 13000 x g. The cell pellet was re-suspended in 50 µl of B-PER (Pierce) and incubated at room temperature with gentle shaking for 10 min. A further centrifugation (13000 x g, 3 min) was required to separate the insoluble, pelletable fraction from soluble material. The supernatant was carefully transferred to a fresh 1.5 ml microtube. 5 µl of 5X sample buffer was added to 20 µl of this fraction for sample preparation for SDS-PAGE as detailed in section 2.1.1. 100 µl of 1X sample buffer was used to re-suspend the cell pellet ready for loading.

2.4: Recombinant protein expression using a eukaryotic host

2.4.1: *S. cerevisiae* BY4741(α) transformation

Transformation of *S. cerevisiae* BY4741(α) cells was essentially performed using the lithium acetate/single stranded carrier DNA/polyethylene glycol (LiAc/ss-DNA/PEG) protocol (Agatep *et al.*, 1998). 50 ml of warm YPD medium was inoculated to a cell density of OD_{600} = 0.1 using approximately 1 ml from an overnight culture. This culture was incubated at 30°C with shaking (200 rpm) until the OD_{600} reached 0.4 – 0.5. Cells were harvested by centrifugation (3000 x g, 5 min) and the pellet re-suspended in 25 ml sterile water. The supernatant was removed after a further centrifugation (3000 x g, 5 min). Cells were re-suspended in 1 ml lithium acetate (100 mM) and transferred to a sterile 1.5 ml microfuge tubes. Cells were, again, pelleted by centrifugation (15000 x g, 15 s) and lithium acetate was removed using a micropipette, then re-suspended in a final volume of 500 µl. Aliquots of 50 µl were transferred to sterile microfuge tubes. For

transformation, cells from one 50 µl aliquot were pelleted by brief centrifugation (6000 x g, 15 s) and the lithium acetate was removed. 240 µl PEG (50% (w/v)) was added first to shield the cells from the detrimental effects due to the high concentration of 36 µl lithium acetate (1 M) of which was added next. An excess (50 µl) of single-stranded DNA (2 mg/ml) (Sigma, U.K) was used to assist in the transformation of BY4741 (α) cells with 10 µg of plasmid DNA (pYES2-GPI-PLD-Sec). Sterile water was added to a total volume of 360 µl. The transformation mix was vortexed vigorously until the cell pellet was completely re-suspended. Incubation at 30°C for 30 min, then a heat shock at 42°C for 1 min followed. The transformation mix was then briefly centrifuged (6000 x g, 15 s) so that the cell pellet could be gently re-suspended in sterile water. This suspension was plated onto two URA-minus plates (Sigma,U.K) One plate contained 200 µl of the 1 ml final yeast inoculum, the other containing the remaining 800 µl concentrated to 200 µl for even distribution onto the plate. Plates were incubated at 30°C for 2 days before a single colony was picked from the first of the transformation plates (the plate containing the lowest cell density). This colony was used to streak a second URA-minus plate so that a clean colony, free from contaminating organisms, could be isolated for expression of the target protein. This second plate was incubated at 30°C for a further 2 days before single colonies were identified for expression and solubility studies.

2.4.2: Expression of GPI-PLD-Sec from BY4741(α) cells

Once a transformant containing the pYES2 construct was isolated, expression of the recombinant protein was induced by the addition of galactose to promote transcription from the *GAL1* promoter. A pYES2 plasmid containing no insert

served as a control in all expression studies using this plasmid and the BY4741(α) strain. An overnight culture was grown up at 30°C in minimal medium [1X Yeast Nitrogen Base (Sigma, U.K), 1X drop out (-URA) (Sigma, U.K)] containing 2% raffinose as a carbon source. Cells were harvested, before being resuspended to an OD_{600} of 0.4 in 50 ml induction medium [1X Yeast Nitrogen Base (Sigma, U.K), 1X drop out (-URA) (Sigma, U.K)] containing 20% galactose. The induction culture was incubated at 30°C with shaking. Samples of 5 ml were transferred from the culture at 0 h, 2 h, 4h, 8 h and at 24 h post induction to a 1.5 ml microfuge tube. Cells were harvested by centrifugation and re-suspended in the required volume of breaking buffer [sodium phosphate pH 7.5, protease inhibitor cocktail (1X)] to achieve an OD_{600} between 50-100. Cells were lysed using excess glass beads and vigorous vortexing for 4 X 30 s interrupted with incubations on ice for 30 s. Holes were pierced at the top and bottom of each microfuge tube, which was then rested at the top of a 15 ml Falcon tube. Centrifugation (2000 x g, 30 s) allowed the cell lysate to filter through glass beads into the 15 ml Falcon tube. A further centrifugation step (13000 x g, 3 min) was performed in order to separate soluble and insoluble fractions. Total cell lysates were assayed by SDS-PAGE to identify induction of recombinant protein expression. Protein concentration was determined to ensure that there was equal loading of samples. SDS-PAGE sample buffer was added to samples at a final concentration of 1X.

2.5: Purification of GPI-PLD

2.5.1: Gel filtration chromatography

For purification of the native target, GPI-PLD, 20 ml sample of bovine serum (previously prepared using 5% (w/v) polyethylene glycol 6000 (PEG 6000) to precipitate protein from solution) was loaded onto a Superdex 200 column (25 cm x 100 cm) or onto a Sephacryl 200 column (1.6 cm x 95 cm) in TBS [50 mM Tris-Cl, 0.15 M NaCl (pH 7.5)] with a flow rate of 2 ml/min. Fractions of 5 ml were collected and assayed for GPI-anchor degrading activity to isolate the fraction(s) displaying the greatest activity. Alternatively, 5.0 ml of active fractions pooled from ion exchange chromatography (IEX) were loaded at a flow rate of 2.0 ml/min onto a Superdex 200 column (2.5 cm x 100 cm) equilibrated in 50 mM Tris-Cl, pH 7.5. Fractions of 6.0 ml were collected. Fractions were assayed for GPI-PLD activity. The anchor degrading activity of GPI-PLD was observed by TNB produced (in pmol ml^{-1} min^{-1}). SDS-PAGE was used to establish the presence of the target protein in the eluted fractions to assess solubility.

2.5.2: Anion exchange chromatography

2.5.2.1: QAE Sepharose

Bovine serum (prepared as detailed above) was subjected to anion exchange chromatography by use of QAE Sepharose. 2 ml of bovine serum (previously dialysed) was loaded at a flow rate of 1.5 ml/min onto a 50 ml QAE Sepharose column equilibrated with 50 mM Tris-Cl, pH 8.0. Bound proteins were eluted with an increasing NaCl gradient [0 - 0.5 M]. Fractions were assayed for activity. The anchor degrading activity of GPI-PLD was observed by TNB produced (in nmol

ml^{-1} min^{-1}). SDS-PAGE was also used to establish the presence of the target protein in the eluted fractions to assess solubility.

2.5.2.2: Mono-Q column

The eluate displaying the highest GPI-PLD activity after purification via the Superdex 200 column was dialysed overnight with 50 mM Tris-Cl (pH 7.5). 4 ml of the dialysed fraction was loaded onto a 1 ml Mono-Q column with a flow rate of 0.5 ml/min. Proteins were eluted with a 0-0.5 M NaCl gradient in 50 mM Tris-Cl (pH 7.5). Fractions of 0.5 ml were collected. Fractions were assayed for GPI-PLD activity. SDS-PAGE was also used to establish the presence of the target protein in the eluted fractions to assess solubility.

2.5.3: Hydrophobic interaction chromatography

Active fractions from ion exchange chromatography (IEX) were pooled and the NaCl concentration adjusted to a final concentration of 2 M. 7.0 ml of this preparation was loaded at a flow rate of 0.5 ml/min onto a 10 ml Phenyl Superose column equilibrated in 50 mM Tris-Cl, pH 8.0, 2 M NaCl. Bound proteins were eluted with a decreasing NaCl gradient and 1.0 ml fractions were collected. The GPI-PLD assay was performed on every fourth fraction and on the material that passed through the column without binding. The anchor degrading activity of GPI-PLD was observed as pmoles TNB produced ml^{-1} min^{-1}. SDS-PAGE was also used to establish the presence of the target protein in the eluted fractions to assess solubility.

2.5.4: Wheat germ lectin affinity chromatography

A sample of 7.5 ml was collected after gel filtration chromatography and IEX and loaded at a flow rate of 0.5 ml/min onto a 1 ml wheat germ lectin agarose (Sigma) column equilibrated in 25 mM Tris-Cl, pH 7.4, 0.15 M NaCl, 2.5 mM $CaCl_2$ and 2.5 mM zinc acetate. Bound proteins were eluted with 25 mM Tris-Cl, pH 7.4, 0.15 M NaCl, 2.5 mM $CaCl_2$ and 2.5 mM zinc acetate supplemented with 0.3 M N-acetylglucosamine and 0.5 ml fractions were collected. 260 µl of both the unbound fraction and remaining load fraction were each concentrated to 50 µl. 50 µl of each eluted fraction, concentrated unbound fraction and concentrated load fraction were assayed for GPI-PLD activity.

2.6: Purification of expressed recombinants

2.6.1.1: Purification of His-tagged proteins using Immobilized Metal Affinity Column (IMAC)

Immobilized metal (Ni^{2+}) affinity chromatography (IMAC) was used to isolate soluble, His-tagged recombinant protein from the supernatant of expression cultures. Such IPTG-induced cultures were scaled-up to 1 l (unless stated otherwise in the text) and the supernatant was passed through a 2 ml or 5 ml HiTrap IMAC HP column with bound Ni^{2+} (GE Healthcare) at a flow rate of 1 ml/min (unless stated otherwise) in 20 mM Tris-Cl, (pH 7.5), 0.1 M NaCl. A wash with buffer [20 mM Tris-Cl, (pH 7.5), 0.1 M NaCl] supplemented with 50 mM imidazole followed to elute weakly and non-specifically bound proteins from the column. Elution was achieved by increasing the concentration of imidazole to 100 mM or 500 mM to remove His-tagged proteins from the column. As a precaution,

the column was stripped by passing EDTA (50 mM) through the column and was reloaded with 0.5 M $NiCl_2$ so that traces of bound protein were removed ensuring that there was no carry over to the purification of the alternative recombinant form. The absorbance at 280 nm was recorded to follow this purification. SDS-PAGE was used to establish the presence of the target protein in the eluted fractions to assess solubility.

2.6.1.2: Small scale Ni-NTA resin binding

This small-scale purification was used as a more rapid approach to identifying his-tagged protein in the supernatant fraction of IPTG-induced cultures. A 1.5 ml microfuge tube was filled with 300 µl Ni^{2+}- NTA Sepharose (Sigma), topped up to 1 ml with Ni^{2+} solution and mixed for 5 min at room temperature. Centrifugation (6000 x g, 2 min) was required for the removal of excess Ni^{2+}, which was carefully decanted. Two washes with 1 ml of buffer [20 mM Tris-Cl, (pH 7.5), 0.1 M NaCl] were carried out before use. A 10 ml IPTG-induced culture was first centrifuged (13000 x g, 3 min) to harvest cells. Cells were lysed using B-PER protein extraction reagent (Pierce) as detailed earlier in section 2.3.13. A further centrifugation (13 000 x g, 3 min) was used to isolate the supernatant fraction. 500 µl of induced culture supernatant was added to 50 µl of the prepared resin and left mixing at room temperature for binding to occur. After 1 h the mixture was centrifuged (13000 x g, 5 min) to separate protein that was bound to the resin from unbound material. The supernatant was transferred to a fresh 1.5 ml microfuge tube for SDS-PAGE analysis. The resin with bound his-tagged protein was washed 4 X with the addition of 1 ml of buffer [20 mM Tris-Cl, (pH 7.5), 0.1 M NaCl] followed by centrifugation (13000 x g, 5 min). The supernatant was kept

after each wash for analysis. The bound protein was released from the resin by a wash with 500 mM imidazole and centrifugation (13000 x g, 5 min) and the supernatant containing his-tagged protein could be assessed using SDS-PAGE.

2.6.2: Gel filtration chromatography

The 5 ml fraction containing highest quantity of pure target protein after IMAC was loaded onto a Superdex 75 (16/60) equilibrated in Tris-Cl (10 mM, pH 8.0) at a flow rate of 1 ml/min. This column was calibrated using a series of protein standards to ensure that the purified protein was monodispersed and not present as a multimer or aggregated. Fractions of 5 ml were collected and immediately stored on ice. The protein concentration in each fraction was calculated using the theoretical extinction coefficient (22710 M^{-1} cm^{-1} for Ap_4A-H and 28210 M^{-1} cm^{-1} for Ap_4A-H-LBTB) at 280 nm. The fraction containing the highest quantity of pure target protein (as indicated by SDS-PAGE, absorbance peak at 280 nm and estimation of protein concentration) was incubated on ice at 4°C for no longer than 16 h before crystallisation screen set up. All other fractions were stored at -80°C for metal binding studies and to assess protein stability.

2.6.3: Purification by IgG affinity chromatography for GB1-tagged proteins

To prepare a column for purification of GB1-tagged protein the ligand, IgG, was covalently attached to a cyanogen bromide (CNBr) – activated Sepharose 4B (Sigma) matrix in a coupling reaction. 2 g freeze-dried CNBr- activated Sepharose 4B was washed with 400 ml of HCl (1 mM). Excess HCl was decanted until cracks began to appear in the gel slurry. The gel was then mixed with the 100 mg IgG

ligand, which was dissolved in coupling buffer [0.2 M $NaHCO_3$, 0.5 M NaCl, pH 8.8] at a final concentration of 10 mg/ml. The coupling solution was left rotating for 2 h at room temperature for coupling to occur. Excess ligand was removed by washing with 5 gel volumes of coupling buffer. A blocking step was required to remove residual active groups that remain on a gel after coupling. This was achieved by hydrolyzing the remaining active groups by re-suspending the gel in Tris-Cl (pH 8.0) and incubating for 20 h. The gel was packed into a 10 ml column. Alternate wash steps with high and low pH solutions were required to remove the excess, uncoupled IgG using 0.1 M sodium acetate buffer (pH 4.0) and coupling buffer respectively. This wash cycle was necessary to ensure that no free ligand remained ionically bound to the immobilized ligand. The column was stored at 4°C until required.

The column was washed with 5 column volumes of buffer [0.1 M Tris-Cl, 0.5 NaCl, 0.05% Tween 20 (pH 7.6)] prior to loading of the supernatant fraction of B-PER lysed cells. The supernatant was diluted 5 X with buffer [0.1 M Tris-Cl, 0.5 M NaCl, 0.05% Tween 20 (pH 7.6)] then, loaded at a rate of 1 ml/min. A wash step with buffer B followed to remove unbound or loosely bound material. Elution of bound GB1-tagged protein was achieved by running elution buffer (0.5 M sodium acetate, pH 3.5) through the column and collecting 1 ml aliquots of the eluate for SDS-PAGE analysis. Each eluate collected was rapidly neutralised with the addition of 1 M Tris-Cl.

2.6.4. Protein refolding in solution

Inclusion bodies collected in the pellet of a 1 l bacterial expression culture (capable of expressing GPI-PLD) were denatured in 10 ml denaturing buffer (50 mM Tris-Cl, pH 8.5, 1 mM $ZnSO_4$, 8 M urea, 5 mM dithiothreitol (DTT)). The denatured protein solution was centrifuged and gradually diluted with 10 ml TBS [20 mM Tris-Cl, pH 7.5, 0.1 M NaCl] over 1 h before being dialysed against 900 ml TBS for 16 h. After a final dialysis for 4 h, a sample was taken, centrifuged and SDS-PAGE analysis was required to determine protein solubility.

2.7: Crystallisation Methods

2.7.1: Sample preparation

Cells from the 1 l expression cultures (BL21(DE3) cells harbouring the expression plasmid for expression of the target protein) were harvested by centrifugation at 6000 x g for 15 min and re-suspended in 20 mM Tris-Cl, 150 mM NaCl (pH 7.5). Cell lysis was achieved by sonication (4 X 30 s) with interruptions of incubation on ice to prevent denaturation caused by heating. The soluble fraction was clarified by centrifugation at 20000 x g and loaded onto a 5 ml HiTrap IMAC HP column (GE Healthcare) and a Superdex 75 column as detailed in section 2.6.1.1 and 2.6.2 respectively. Protein concentration was achieved using a Vivaspin MWCO 5000 spin concentrator column (GE Healthcare) as detailed in section 2.1.5.2. Purified protein preparations were analysed by SDS-PAGE and visualized by Coomassie Brilliant Blue staining.

2.7.2: Crystallisation trials

2.7.2.1: Manual approach to setting up crystallisation screen

A deep-well block was filled with 96 x 1.5 ml screening solutions of the Hampton Index (HR2-114) kit and left to equilibrate to room temperature. Buffers were indexed for identification. A Pipetman® Ultra Multichannel (Gilson) was used to accurately dispense 100 µl of the screening solutions from the deep-well block and into the buffer wells of a 96 well crystallisation tray for sitting-drop vapour diffusion (Greiner). The block was covered where possible to prevent buffer evaporation. All sub-wells of the crystallisation tray were loaded with 1 µl screening solution from corresponding buffer wells using a multichannel pipettor. The sitting drop tray allowed for the screening of three drops per well. The first sub-well was loaded with protein at a concentration of 5 mg/ml, the last at 10 mg/ml with the middle sub-well filled with screening solution from the buffer well. Sample loading into sub-wells made use of a digital automatic repeat pipettor to accurately dispense 1 µl of the protein. Once all wells were loaded the crystallisation tray was covered using sealing tape and stored at room temperature. Drops were viewed using 500X magnification after 24 h, 48 h, 72 h, then every week thereafter. Drops were scored based on extent of precipitation. All 96 wells of each crystallisation screen were scrutinized after 1 week and scores were given for each based on the extent of precipitation, with a score of 5 representing dense precipitation with 1 representing a clear drop. A difference in the precipitation score of 2 or more between LBTB- tagged and non-tagged Ap_4A-H represented an effect on solubility. The number of protein drops affected by the addition of La^{3+} was counted as either causing an increase or a decrease in solubility.

2.7.2.2: Automated approach to setting up crystallisation screen

The Screenweaver 96+8 robot (Innovadyne Techologies Inc.) was utilized for automated crystallisation trial set up. The pre-filled deep-well blocks of Qiagen screens, namely Classics, PEG suite, PEG suite II, pHClear, Mb Class and JCSG+ were removed from storage at 4°C and allowed to equilibrate to room temperature. Protein samples were removed from ice and centrifuged (13000 x g, 2 min) before being set into position on the robot. Next, screening solutions of the desired Qiagen suite and a 96 well crystallisation plate for sitting-drop vapour diffusion (Hampton) were also set in place. The robot transferred 80 µl of the screening suite solutions to the appropriate well. This was followed by the transfer of 0.5 - 1µl from each well to each of the sub-wells. The equivalent volume of protein sample was added to sub-wells. Two different samples were always loaded in sub-wells for comparison. Once all wells were loaded the crystallisation tray was conserved using sealing tape and stored at room temperature. Drops were stored in and viewed using CrystalPro™ HT - Multi-Plate Hotel and Imager (500X magnification) after 24 h, 48 h, 72 h, then every week thereafter. Drops were scored based on extent of precipitation as detailed for manual set up.

2.8: Gel filtration chromatography to determine metal binding of isolated peptides

The Superdex Peptide column (GE Healthcare) is capable of high-resolution separation of smaller biomolecules offering a 100-7000 (Mr) fractional range. Binding of a lanthanide chloride to the peptide results in an increase molecular weight consequently causing earlier elution from the column, shifting the trace to the left. This provided a means of detecting the metal binding ability of the isolated

LBT. Peptides (LBT(A/B)) (2 mg/ml) were incubated at room temperature with an excess of each lanthanide chloride hexahydrate (1 M) (50 µl peptide: 5 µl $LnCl_3 \cdot 6H_2O$) in a 1.5 ml microtube for 1 h or 16 h at room temperature. All metal chlorides were stored in a desiccator over silica. After the required period of incubation (1 h or 16 h) peptide/lanthanide mixtures were subjected to a centrifugation step at 13000 x g for 3 min. A 12 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of the peptide-metal chloride incubation and run at a flow rate of 0.5 ml/min with TBS (20 mM Tris-Cl, 0.1 M NaCl, pH 7.5) or HEPES buffer (100 mM Na-HEPES, pH 7.0, 150 mM KCl). Elution of protein was monitored by absorbance at either 214 nm or 280 nm as indicated in the text.

3.0: RESULTS CHAPTER; Optimisation Strategies For Heterologous Expression Of Glycosylinositol-Specific Phospholipase D (GPI-PLD)

3.0: Introduction

The requirement for high yields of purified protein can often become a bottleneck phase in crystallographic studies. In many instances the rapid production of recombinant proteins leads to the formation of insoluble aggregates designated as inclusion bodies. To date there is no optimal expression system for working with all recombinant proteins. Each protein poses new problems and in every case expression has to be optimised by pragmatic variations of the different parameters.

The aim of this chapter was to address the optimisation of recombinant GPI-PLD expression so that adequate protein concentrations (1-100 mg/ml) could be obtained for structural studies. GPI-PLD was the chosen subject given that there is a huge demand for structural information for therapeutic applications (as discussed in section 1.2.5). Studies involving this GPI-PLD brought to light how studies can be hindered even at the early stages of construct design let alone expression of a soluble protein.

First and foremost, an *in vitro* assay for the detection of the target protein needed to be established. Activity can provide a good indication of protein stability and/or correct folding. GPI-PLD activity can be analysed by its ability to cleave and release GPI-anchored proteins from the cell surface and a number of methodologies using a variety of substrates have been documented. Radiolabeled GPI-linked proteins and membrane form variant surface glycoprotein (mf-VSG) of trypanosomes can be used as substrate (Deeg and Bowen, 1999). However, GPI-anchored enzymes such as membrane form-acetylcholinesterase (mf-AChE) (Heller *et al.*, 1992) and GPI-

alkaline-phosphatase (AP) (Rhode *et al.*, 1995) provide an added advantage in that they can be easily detected. Their reaction products can be detected with minimum intricacy so they were considered for the detection of recombinant GPI-PLD.

There are many parameters to be modified in the optimisation of target protein expression. Induction conditions (temperature, IPTG concentration, time period of induction) are often modified (as detailed herein) so recombinant expression can be optimised in an attempt to maximize protein yields and overcome solubility issues (Baneyx, 1999; Hannig and Makrides, 1998). Ultimately, the choice of vector or even the length of the target amino acid sequence can limit the success of soluble recombinant protein production (Makrides, 1996). So, careful considerations were made from the onset. In addition to this, a number of *E. coli* host strains were employed, including Origami2(DE3) (a strain capable of aiding the construction of disulphide bonds). Although the amplified *Mus musculus* GPI-PLD4a sequence only contains 4 cysteine residues and their involvement in disulphide bridge formation is not yet known this strain was used as an expression host in the event that these bonds were required for the specific folding of the N-terminal domain of this protein. Additionally, this eukaryotic gene sequence contains a number of rare codons (AGG, AGA, AUA, CUA, GGA and CGG). Rosetta2(DE3) *E.coli* strain is capable of recognising rare codons that are not recognised in other strains as they contain the additional tRNAs and were employed in expression studies herein.

Another popular strategy in maximizing expression levels of recombinant targets is the employment of a large, soluble protein fusion/partner to improve solubility and guide the folding of the protein in question. Such fusion proteins include GST (Smith and Johnson, 1988), GB1 (Gronenborn *et al.*, 1991) and their effects on recombinant expression of GPI-PLD were explored.

The production of heterologous protein using prokaryotic expression hosts can more often than not become a difficult undertaking. The use of eukaryotic expression systems is, in general, time consuming, more costly and often provides lower yields when compared to prokaryotic systems, but it can provide a route forward (Demain and Vaishnav, 2009). This expression system offers the added advantage of the glycosylation processes provided by the host and is often preferred for mammalian proteins.

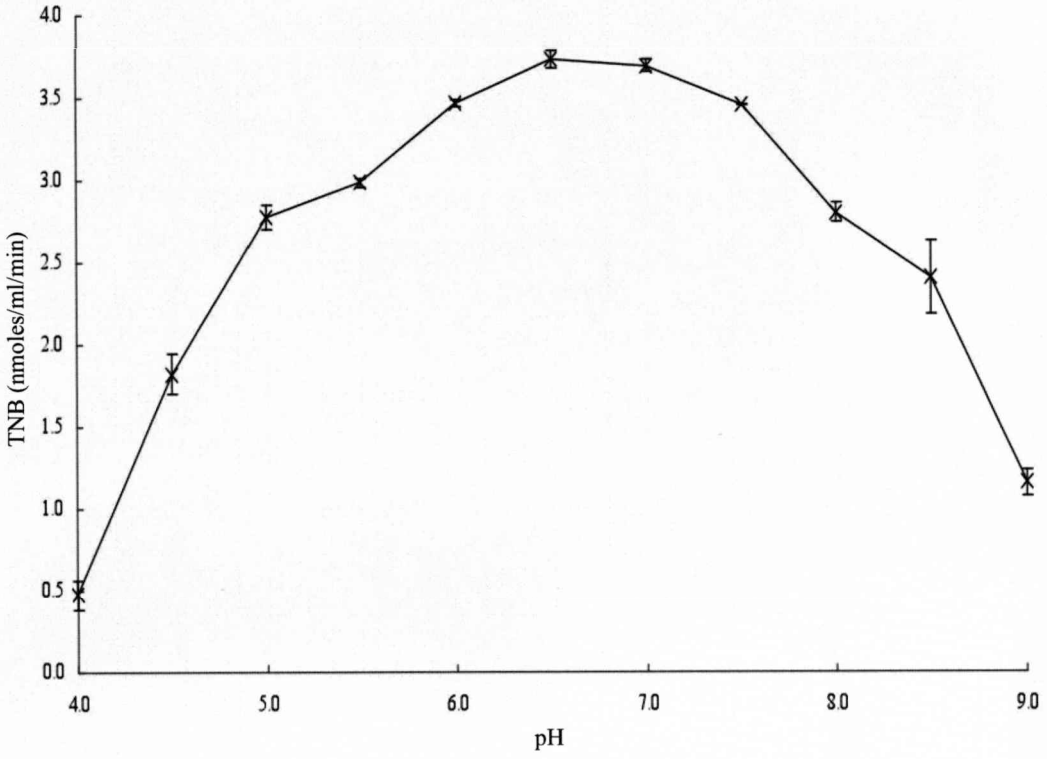
3.1: GPI-PLD Activity Assay Development

Initial investigations used GPI-anchored proteins, FGF and the Fc receptor, CD16. Bovine serum served as a source of GPI-PLD to release FGF and CD16 from the neutrophil cell surface (engineered by and provided by Dr Andy Cross). Evidence of their cleavage was assayed via western blotting (see section 2.1.2 for details). Both diaminobenzidine (DAB) and enhanced chemical luminescence (ECL) methods were used to detect the released protein but neither proved to be successful (results not shown).

Subsequently, a second GPI-PLD activity assay utilising mf-AChE, adapted from previously published work (Bordier, 1981; Ellman, 1956) was investigated. This assay is comprised of the following steps: cleavage of GPI-anchored, membrane form acetylcholinesterase (mf-AChE), producing soluble AChE (s-AChE); phase separation by use of Triton X-114, separating the mf-AChE from that released into the supernatant fraction (s-AChE) and finally an AChE activity assay. This assay (detailed in full in section 2.2.1) had to be optimised for use in this lab, particularly for use on a smaller scale. Thus, preliminary investigations were aimed at optimising the sensitivity of the assay by determining an ideal concentration of red

blood cell ghosts (GPI-anchored AChE), so that the product (TNB) could be detected in the final stage of the assay over a realistic period of time (a maximum of 30 min). For these experiments it was decided that a dilution of 1/20 of red blood cell ghosts would provide a sufficient concentration to allow for detection of s-AChE in the final stage of the GPI-PLD assay.

Studies were carried out in order to ascertain the optimum pH for the first stage of this activity assay (i.e. mf-AChE plus bovine serum at 37°C for 30 min) using a series of buffers in the range pH 4.0 – pH 9.0 in 0.5 increments). Optimum pH of the assay was taken to be between pH 6.5 and pH 7.0 (Figure 3.1). Therefore, all future GPI-PLD activity assays were performed using mf-AChE prepared using the sodium phosphate buffer at pH 7.0.



Fig; 3.1: The effect of pH on the GPI-PLD activity.

The optimum pH for GPI-PLD activity was determined using red blood cell ghosts (mf-AChE) suspended in buffers in the range pH 4.0 – pH 9.0 (according to section 2.2.1). The anchor degrading activity of GPI-PLD from each sample was observed by nmoles TNB produced ml^{-1} min^{-1}. Data points show the mean activity (n =3) with error bars representing standard deviation.

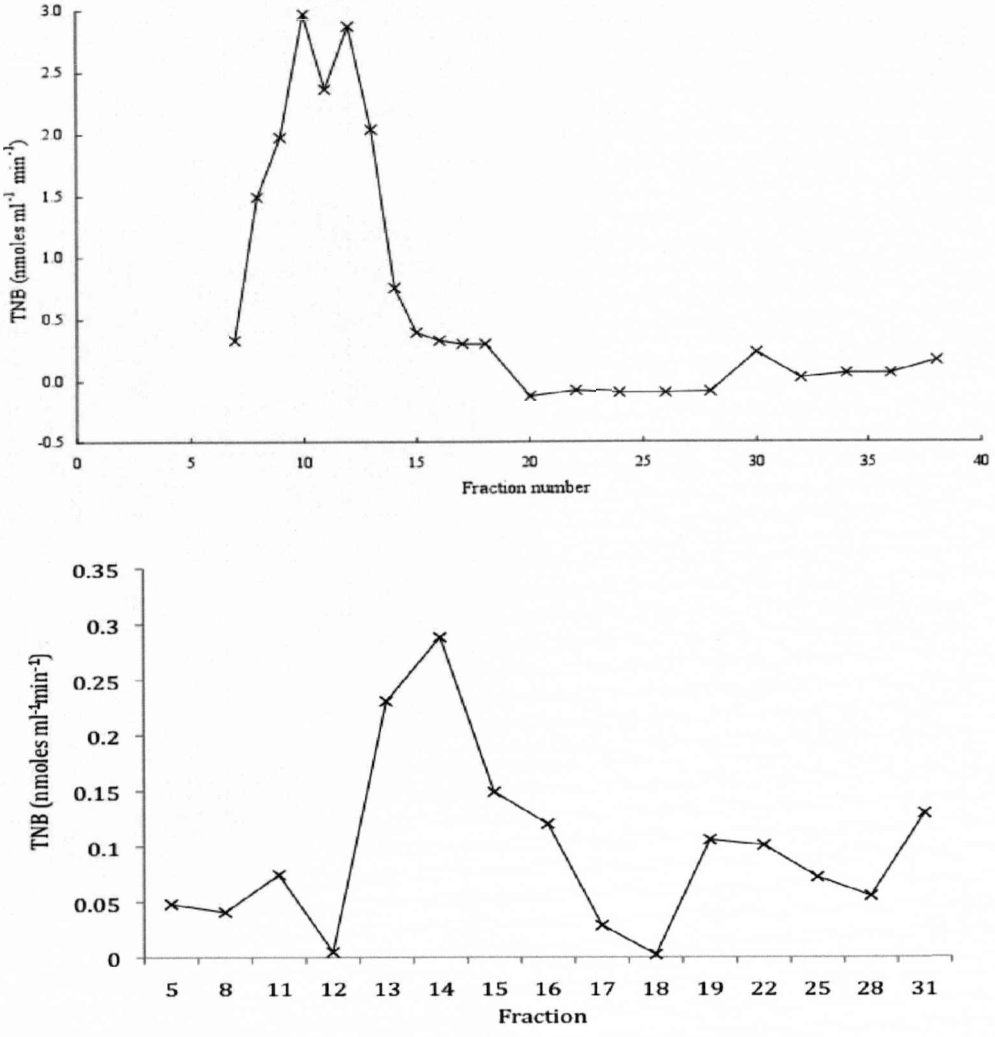
3.1.2: Isolation of native GPI-PLD from bovine serum.

GPI-PLD was purified from bovine serum by taking into consideration the success of protocols from previously published work (Huang *et al.*, 1990; Deeg and Bowen, 1999; Rhode *et al.*, 1995). Various combinations of gel filtration chromatography (using Sephacryl 200 and Superdex 200 columns) and anion exchange chromatography (using Mono Q and QAE Sepharose columns) were tested as possible initial steps. The most successful combination of large-scale gel filtration chromatography followed by high performance anion exchange chromatography is shown in figure 3.2. However, even after a further purification step using a hydrophobic interaction column (Phenyl Superose) SDS-PAGE analysis showed that active fractions contained relatively large amounts of contaminating proteins. Serum albumin contamination was a particular problem. Significant problems were experienced with activity loss during isolation. Consequently, an alternative purification protocol was developed which used affinity chromatography.

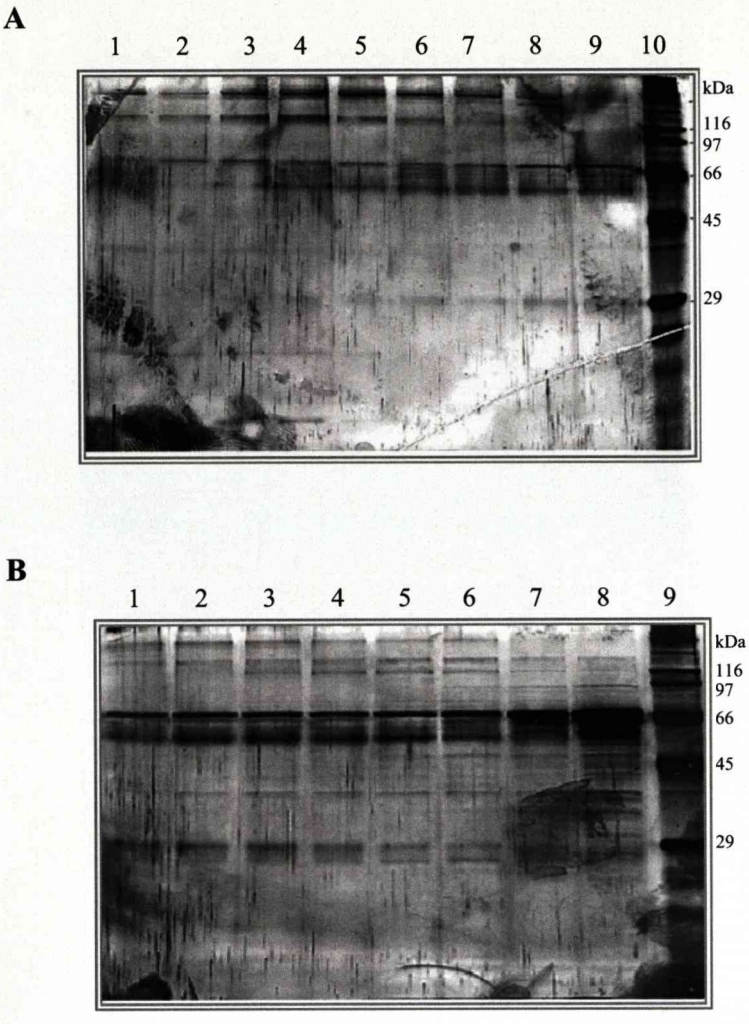
Ultimately, GPI-PLD was partially purified from bovine serum using a series of three chromatography steps. The first of these made use of anion exchange chromatography, but carried out on a larger scale than that used for figure 3.2, using a 100 mL QAE Sepharose column. GPI-PLD activity was eluted predominantly in the 0.15-0.2 M NaCl region of the gradient (results not shown). A small amount of GPI-PLD activity was also present in the unbound fraction. Active fractions were pooled and applied to a Superdex 200 column for gel filtration chromatography. Fractions of 6 ml were collected and 20 µl of each was assayed for GPI-PLD activity. The GPI-PLD assay showed that there were peaks of activity in fractions 6, 11 and 17 with the greatest activity in fraction 17. SDS-PAGE analysis of fractions 4

to 20 collected from the Superdex 200 column showed that fractions 15-20 contained bands around 120 kDa region where GPI-PLD would be expected (Figure 3.3).

The most active fractions [17-19] collected from the Superdex 200 column were pooled and loaded at a flow rate of 0.5 ml/min onto a 1 ml wheat germ lectin agarose (Sigma) column equilibrated in 25 mM Tris-Cl (pH 7.4) 0.15 M NaCl, 2.5 mM $CaCl_2$ and 2.5 mM zinc acetate as described in section 2.5.4. Bound proteins were eluted with buffer [25 mM Tris-Cl (pH 7.4) 0.15 M NaCl, 2.5 mM $CaCl_2$ and 2.5 mM zinc acetate] containing 0.3 M N-acetylglucosamine and 0.5 ml fractions were collected. The eluted fractions were assayed for GPI-PLD activity alongside the load and unbound fractions, both of which had been concentrated 5-fold (Table 3.1). SDS-PAGE analysis was performed on all fractions collected from the wheat germ lectin agarose column, and on the load and unbound material (Figure 3.4). A strong band at approximately 120 kDa was present in all eluted fractions but was more prominent in the most active fractions, 2 and 3.



Fig; 3.2: GPI-PLD activity detected in GF and IEX chromatography fractions
Top) GPI-PLD activity of assayed fractions collected from a Sephacryl 200 column (GFC). 3 ml of bovine serum supernatant was loaded at a flow rate of 1.0 ml min^{-1} onto a Sephacryl 200 column (1.6 cm x 95 cm) equilibrated in TBS [50 mM Tris-Cl, pH 8.0, 0.15 M NaCl]. Fractions of 5.0 ml were collected. *Bottom*) GPI-PLD activities of fractions eluted from anion exchange chromatography. 5 ml fraction 5 from gel filtration chromatography loaded onto Mono-Q column with a flow rate of 0.5 ml/minute. Proteins were eluted with a 0-0.5 M NaCl gradient in buffer A (50 mM Tris-HCl, pH 7.5). Fractions were assayed according to the method described in section 2.5.2.2. The anchor degrading activity of GPI-PLD was measured as nmoles TNB produced ml^{-1} min^{-1}.

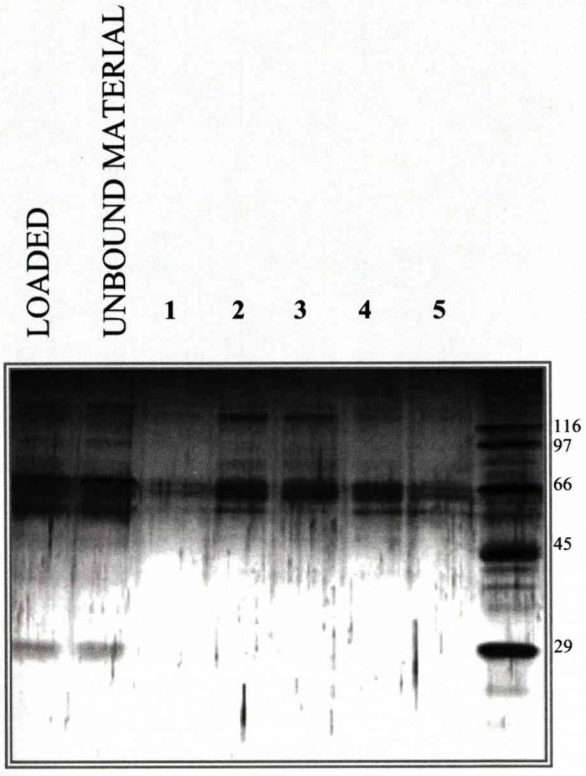


Fig; 3.3: SDS-PAGE of fractions collected from the Superdex 200 column.

Active fractions from GFC were run on 10.5% polyacrylamide gels and visualized by silver staining. **A,** *lanes 1-9,* fractions 4-12; **B**, lanes 1-8, fractions 13-20. Molecular weight markers (kDa) are as indicated (**A**, lane 10; **B**, lane 9).

Table 3.1: GPI-PLD activity of assayed fractions from a 1 ml wheat germ lectin agarose column.

Fraction	Anchor degrading activity (pmoles ml^{-1} min^{-1})
Load	43.8
Unbound	31.6
1	4.9
2	111.9
3	63.3
4	58.4
5	46.2



Fig; 3.4: SDS-PAGE analysis of the fractions collected from wheat germ lectin affinity chromatography.

Samples were run on a 10.5% polyacrylamide gel and visualized by silver staining. Lane 1, loaded fraction; lane 2, unbound fraction; lanes 3-7, eluted fractions 1-5. Molecular weight markers (kDa) are as indicated in lane 8.

A 0.5 ml aliquot of fraction 2 was concentrated 25-fold using a Vivaspin protein concentrator columns. The concentrated fraction was loaded onto two lanes of a 10.5% polyacrylamide gel and SDS-PAGE was carried out in preparation for an in-gel trypsin digest to obtain peptides for identification by mass spectrometry. Unfortunately the protein was lost during the concentrating step and no bands were visible, except for protein standards, when stained with Coomassie blue (data not shown). Thus, it was not possible to sequence the partially purified 120 kDa band in order to confirm that this was in fact GPI-PLD. As bovine GPI-PLD was purified from serum by the successive use of three chromatography steps; anion exchange on QAE Sepharose, GFC on Superdex 200 and wheat germ lectin affinity chromatography it was notable that the level of anchor degrading activity decreased as the purification progressed from column to column. This may have been due to a loss of protein activity and/or a loss of the protein into other fractions collected from the column.

Investigations were carried out to observe the GPI-anchor degrading activity of bovine serum in the presence of protease inhibitors. This was conducted to rule out the possibility that proteases present in serum were solely responsible for, or contributed to, the GPI-anchor degrading activity in this assay. Two methods were used for the protease inhibition, one using EDTA (final concentration, 50 mM), the other using a protease inhibitor cocktail (diluted according to manufacturers instructions) (Sigma). The protease inhibitor cocktail contained aprotinin, E-64, pepstatin A and leupeptin hemisulfate salt giving a broad specificity to serine, cysteine and acid proteases. The presence of EDTA and the inhibitor cocktail resulted in 41% and 43% less anchor degradation (respectively) when compared to that with no added protease inhibitors. Thus it was evident that the GPI-PLD

activities determined by assays carried out in the presence of each protease inhibitor were greatly reduced. This is to be expected, as the addition of EDTA will remove the Zn^{2+} and Ca^{2+} that has been reported to be required for GPI-PLD activity.

3.2: Heterologous expression of GPI-PLD using a prokaryotic expression system.

3.2.0: Identification of a suitable homologue of human GPI-PLD

A number of homologues including *Saccharomyces cerevisiae* (accession NP-012403), *Dictyostelium discoideum* (accession AAO50947) and a collection of mammalian homologues were identified by iterative database searches of human GPI-PLD (Rigden, 2004). Of the two lower eukaryotic homologues, *S. cerevisiae* shares the highest pairwise sequence identity at 36% compared to only 25% for *D. discoideum* (Figure 3.5). The putative protein from *S. cerevisiae* is the product of the non-essential gene, *YJL132W,* that to date has no known function. This GPI-PLD relative was initially employed for heterologous expression and structural studies herein.

The GPI-PLD-specific primers for PCR were designed to span the first 275 amino acid residues (Figures 3.5-3.6). This region is predicted to contain the catalytic domain of the protein (Hellar *et al.*, 1994). The forward primer (Table 3.2) was designed with the intention of amplifying the predicted N-terminal catalytic domain. Reverse primers R275 and R170 (Table 3.2) were devised to amplify the first 275 amino acids and the first 170 amino acids respectively. Both reverse primers were designed to take advantage of the C-terminal his-tag and termination codon offered by the pET22b plasmid. Restriction sites for Nde I and Hind III were embedded in forward and reverse primers, respectively, to facilitate ligation of the insert into the expression plasmid.

Fig; 3.5: Sequence alignment of human GPI-PLD1 homologues

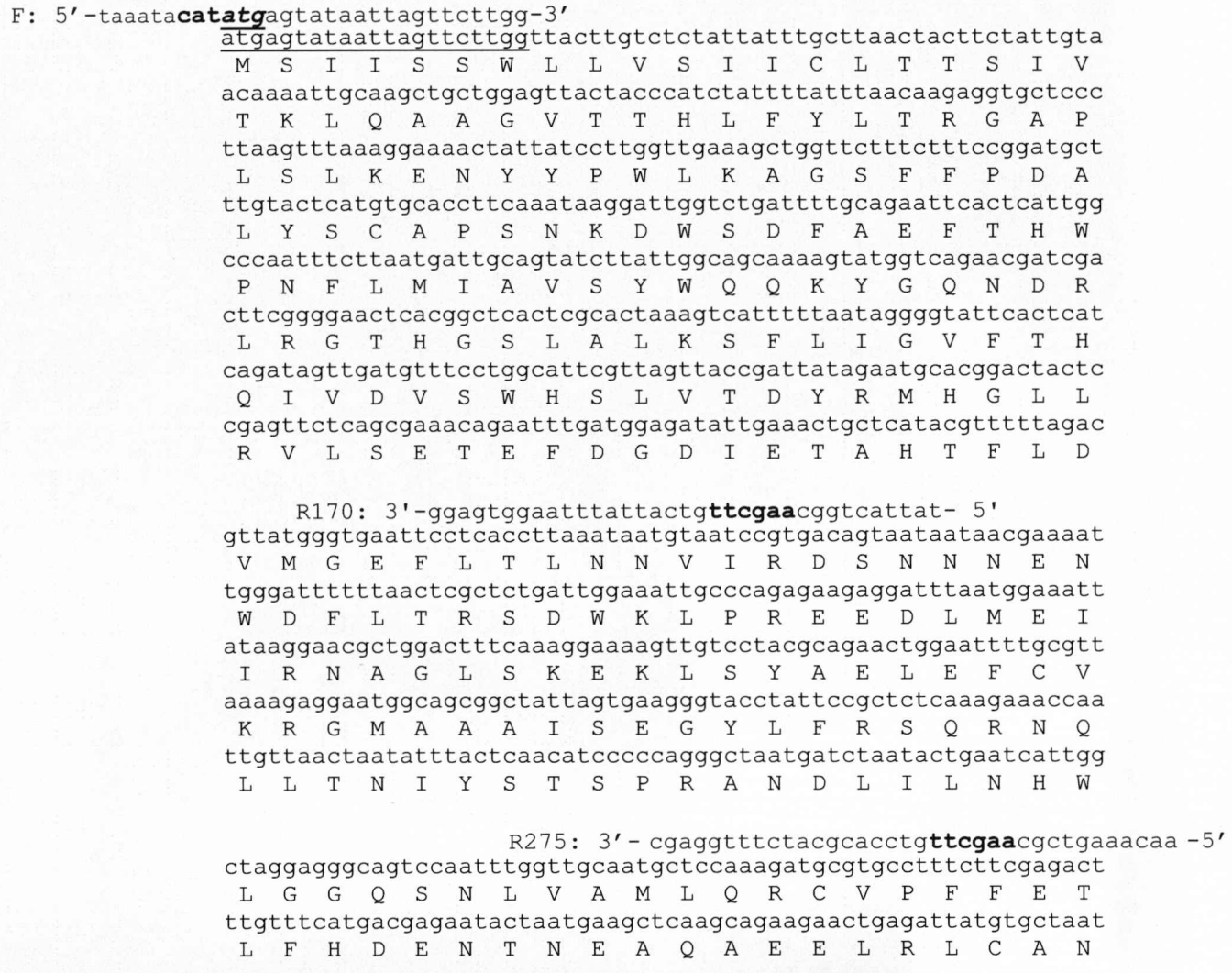
Sequence alignments of human GPI-PLD1 (with accession number, P80108) with *S. cerevisiae* (NP_012403), *D. discoideum* (AAO050947) and *M. musculus* (AAH19146) homologues (each with Genpept accession number) using muscle. Only the initial 275 amino acids of the N-terminal domain were aligned.

TABLE 3.2: Oligonucleotide primers
Primers used for amplification of the catalytic domain of GPI-PLD (Figures 3.2.1 and 3.2.2). Sequence highlighted in bold represent restriction sites. An Nde I site was incorporated into the coding primer, F. The start codon (underlined, italics) was integrated into this restriction site. Both non-coding primers contained Hind III restriction site (bold).

Primer	Oligonucleotide primer sequence
F	5'-TAAATACATATGAGTATAATTAGTTCTTGG-3'
R170	5'-TATTACTGGCAAGCTTGTCATTATTTAAGGTGAGG-3'
R275	5'-AACAAAGTCGCAAGCTTGTCCACGCATCTTTGGAGC-3

Initially, a colony PCR was attempted whereby a single *S. cerevisiae* (NVSCI) colony was smeared inside the 0.2 ml domed microtube to serve as a DNA template. Execution of the colony PCR was as described in section 2.3.2 using annealing temperatures between 40-45°C. Amplification of the catalytic domain of the GPI-PLD gene by the use of oligonucleotide probes (Table 3.2) proved to be unsuccessful.

The alternative approach was to use DNA extracted from *S. cerevisiae* (NVSCI) as a DNA template. Isolation of *S. cerevisiae* genomic DNA was achieved by following the rapid method of extraction of Cheng and Jiang (2006). An annealing temperature of 45°C was optimum for obtaining the PCR product of primers F and R275 (Figure 3.7). The shorter of the two truncations by use of reverse primer, R170, could not be amplified.



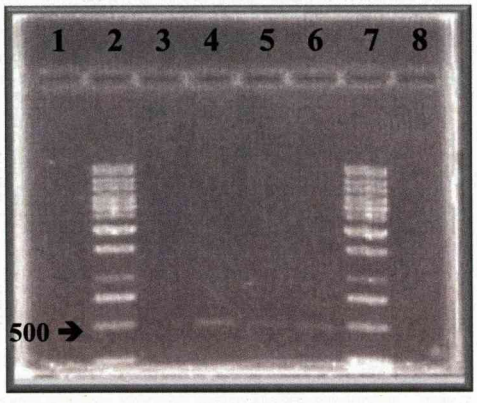
Fig; 3.6: GPI-PLD specific primers and sites for annealing in the corresponding target sequence.

The predicted GPI-PLD catalytic domain from *YJL132W* gene of *S. cerevisiae* was translated using PSIPRED. Primers are shown in blue at the annealing site in the target sequence.

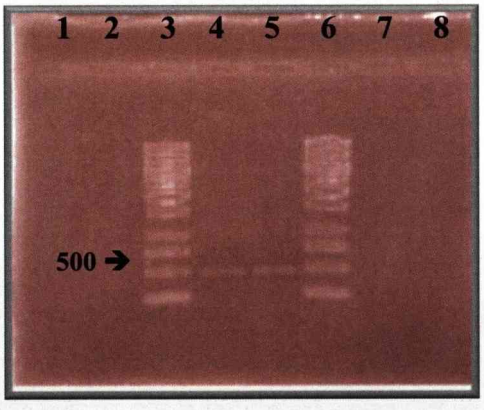
All primers for amplification of *GPI-PLD* gene truncations had a high AT content and had low stability as a consequence. This made it very difficult to amplify a sufficient yield of DNA of the desired length. Success of consecutive PCR reactions was erratic with interludes of non-amplification. Attempts were made to increase the yields of DNA amplification by means of increasing the number of PCR cycles and scaling up the PCR reaction to 400 µl (Figure 3.8). In spite of this, molecular cloning could not be progressed further on account of poor yields and insufficient recovery of DNA after purification, particularly when recovering DNA from agarose gels after restriction digestion.

Persistent problems in the amplification of the N-terminal region of *GPI-PLD* gene from *S. cerevisiae* (NVSCI) by PCR made it necessary to consider an alternative homologue for further studies. It was at this juncture that the *GPI-PLD* cDNA from the *Mus musculus* homologue was considered.

The sequence of the *M. musculus GPI-PLD* homologue (GenBank accession number BC019146) shares 84% pairwise sequence identity with the human form within the first 275 amino acids of the N-terminus. This in itself was an added advantage with respect to the development of specific inhibitors for therapeutic purposes should the catalytic structure of GPI-PLD be deduced. The availability of *GPI-PLD* cDNA (Geneservice, I.M.A.G.E ID: 5052822) meant that it could be readily expressed using an *E. coli* host.



Fig; 3.7: Identification of PCR products from PCR using annealing temperatures of 40-45°C. PCR was performed at annealing temperatures 40-45°C. PCR products were run on a 1% agarose gel. Lane 3 is loaded with the PCR product using an annealing temperature of 40°C, lane 4 at 41°C, lane 5 at 43°C and lane 6 at 45°C. Lanes 1 and 8 remain empty. Lanes 2 and 7 contain 6 µl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250.



Fig; 3.8: Scale-up PCR
A PCR was performed to scale-up the quantity of PCR product obtained using an annealing temperature of 45°C. Four identical 0.2 ml PCR tubes were set up. PCR products were identified by loading 5 µl of each PCR product on a 1% agarose gel (lanes 4 and 5). Lanes 3 and 6 contain 6 µl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250.

3.2.1: PCR optimisation for amplification of GPI-PLD truncations from cDNA

A number of GPI-PLD-specific primers (Table 3.3) were designed in the process of generating multiple GPI-PLD constructs in an attempt to obtain a soluble recombinant protein. Constructs were of varying length encompassing various tags of both the N- and C-termini. Forward primers (F and F-GST) were designed so that the resulting amplicon excluded the signal peptide leader sequence (Figure 3.9). For ease of purification, primers were designed so that the recombinant protein included a fusion-tag, acquired by ligation into the appropriate vector (Novagen). This entailed the insertion of restriction sites for restriction enzymes into the amplified sequence by the use of forward and reverse primers (Table 3.3) to create complementary ends. No termination codons were required in reverse primers for pET22b vector constructs as one follows the C-terminal, hexa-histidine stretch in the plasmid sequence. Termination codons were embedded in reverse primers for all constructs comprising an N-terminal tag/fusion, as they were not offered in the plasmid. Reverse primers were designed to anneal within the linker region to maximize chances of amplifying sequences containing the full catalytic domain.

CONSTRUCT	PRIMERS	EXPRESSION PLASMID	TAG/FUSION PROTEIN	RESTRICTION SITE USAGE
GPI-PLD4a	F 5'- ATAGAATTCATATGTGTGGTCTCTCAACACATGTAGAAATAGG-3' Rev 4a 5'-TCAGCGGCCGCAAGCTTGCAGTCACTGGTCCCGTTCTCCAGC-3	pET22b	C-terminal His6	Nde I + Hind III
GPI-PLD4b	F 5'- ATAGAATTCATATGTGTGGTCTCTCAACACATGTAGAAATAGG-3' Rev 4b 5-TCAGCGGCCGCAAGCTTCTAGTCACTGGTCCCGTTCTCCAGC-3'	pET28a	N-terminal His6	Nde I + Hind III
GPI-PLD3a	F 5'- ATAGAATTCATATGTGTGGTCTCTCAACACATGTAGAAATAGG-3' Rev 3a 5'-AGGCGGCCGCAAGCTTCTCAGGCAGGTTGCAGTCACTGGTCCCG-3'	pET22b	C-terminal His6	Nde I + Hind III
GPI-PLD3b	F 5'- ATAGAATTCATATGTGTGGTCTCTCAACACATGTAGAAATAGG-3' Rev 3b 5'-AGGCGGCCGCAAGCTTCTAAGGCAGGTTGCAGTCACTGGTCCCG-3'	pET28a	N-terminal His6	Nde I + Hind III
GPI-PLD + GST	F-GST 5'-AAGTCATATGCTCGAGATGTGTGGTCTCTCAACACATGTAGAAATAGG-3' R-GST 5'- ATCTGAAGCTTGCGGCCGCTAGTCACTGGTCCCGTTCTCCAGC-3	pGEX-6P1	N-terminal GST	XhoI + Not I
GPI-PLD + GBIHis6	As used for GPI-PLD4a construct	pET24bGB1-His6	N-terminal GBI-His6	Nde I + Hind III

Table 3.3: GPI-PLD-specific primers for the generation of multiple constructs.

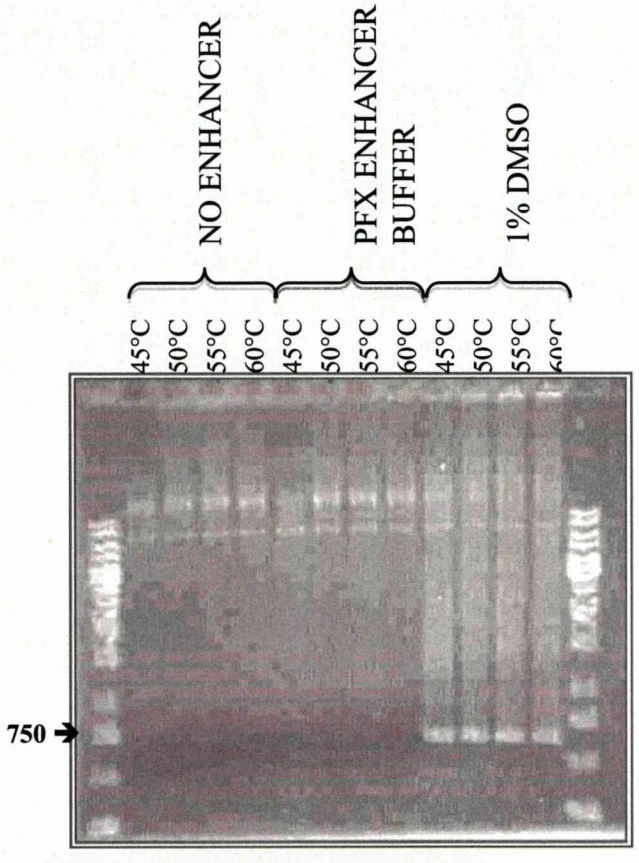
For each construct designed, the primers, expression plasmids and restriction enzymes used are detailed above. The tag or fusion protein of each is also noted.

ccacgcgtccgctgaggaggaatgacaac***atg***tctgcaggcaggctgtggtctagcctgctg
T R P L R R N D N ***M*** *S A G R L W S S L L*
cttctgctgcctcttttctgctctaaaagctca ***tct*** tgtggtctctcaacacatgtagaa
L L L P L F C S K ***S S S C G*** L S T H V E
ataggacacagggctctggagtttcttcggcttcaagatggacgcattaactacaaagag
I G H R A L E F L R L Q D G R I N Y K E
ctgatcttagagcaccaggacgcatatcaggctgggaccgtgtttcctgatgccttttat
L I L E H Q D A Y Q A G T V F P D A F Y
cctagcatctgcaaaagaggaaaatatcatgacgtttctgagaggactcactggactcca
P S I C K R G K Y H D V S E R T H W T P
tttcttaacgccagcatccattatattcgagagaactaccctctgccctgggagaaggac
F L N A S I H Y I R E N Y P L P W E K D
acagagaagttggtggctttcttgtttggaatcacctcccacatggtcgctgacgtgagc
T E K L V A F L F G I T S H M V A D V S
tggcatagcctgggtattgaacaagggttcctcaggacaatgggagctatcgatttttac
W H S L G I E Q G F L R T M G A I D F Y
aactcttactctgacgctcactcggctggtgattttggaggagatgtgttgagccagttt
N S Y S D A H S A G D F G G D V L S Q F
gaatttaattttaattacctctcacggcgctggtacgtgcccgtcagggatcttctgaga
E F N F N Y L S R R W Y V P V R D L L R
atttatgataatctctatggtcggaaagtcatcaccaaagacgtccttgttgattgcacc
I Y D N L Y G R K V I T K D V L V D C T
taccttcagttcctggaaatgcacggggagatgtttgctgtttccaagctctattccacg
Y L Q F L E M H G E M F A V S K L Y S T
tactctacaaagtccccatttctggtggagcagttccaagactatttcctcggaggtctg
Y S T K S P F L V E Q F Q D Y F L G G L
gatgacatggcattctggtccacgaacatttaccgtttgaccagctttatgctggagaac
D D M A F W S T N I Y R L T S F M L E **N**
gggaccagtgactgcaacctgcctgagaaccccctgttcatctcctgtgatggcaggaac
G T S D C N L P E N P L F I S C D G R N
cacaccctcagtggctcaaaagtgcagaaaaatgattttcacaggaatttgaccatgttc
H T L S G S K V Q K N D F H R N L T M F
ataagtagagacatcaggaaaaacctcaattacacagaaagaggcgtgttctacagcaca
I S R D I R K N L N Y T E R G V F Y S T
ggctcctgggccccggaatctgtcacctttatgtaccagactctggagaggaacctgagg
G S W A P E S V T F M Y Q T L E R N L R

Fig; 3.9: Translation of GPI-PLD clone sequence – N terminal domain only

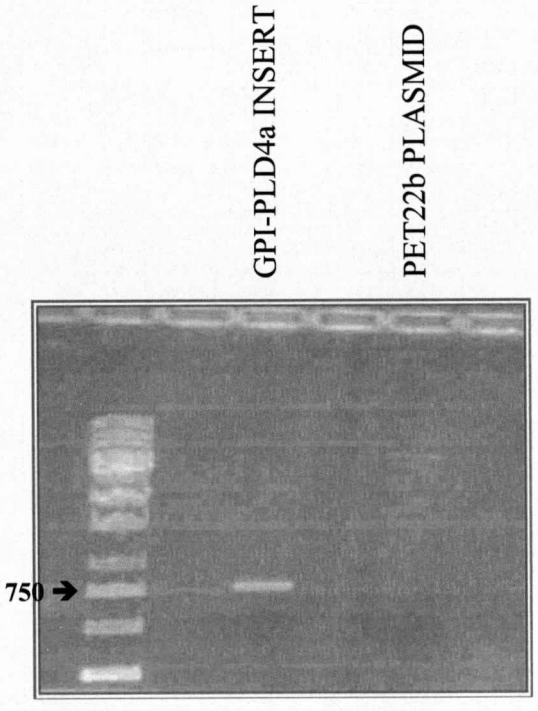
Figure shows the translation of the N-terminal domain GPI-PLD (GenBank Accession number, BC019146) nucleotide sequence highlighting the linker region between N-terminal catalytic domain and C-terminal β-propeller domain (highlighted in green). A cysteine at residue 275 (highlighted in red) was followed with a stop/termination codon where required. The signal peptide sequence is shown in italics. SignalP (Bendtsen *et al.*, 2004) predicts that the signal peptide is cleaved at SSS-CG (blue). The new initiation codon replaces the third S residue (**BOLD**) in the signal peptide.

The first construct, GPI-PLD4a, was designed to produce a 275 amino acid truncation of GPI-PLD including a C-terminal His tag. The GPI-PLD clone served as the DNA template for PCR. An optimum annealing temperature was first determined using BioMix Red reaction mix (BioLine) containing a BIOTAQ DNA polymerase for amplification at annealing temperatures of 45°C, 50°C, 55°C and 60°C (Figure 3.10). All PCR reactions were successful, so the highest annealing temperature of 60°C was selected for all further amplifications to reduce the probability of non-specific binding of primers. A further round of optimisation followed whereby enhancers, DMSO (1% (v/v)) and Platimun Pfx DNA polymerase enhancer buffer (Invitrogen) were added to each PCR reaction to observe their effect on yield of amplification. Only PCR reactions containing DMSO were able to yield a product of the predicted size of 790 bp (Figure 3.10). A PCR product of a sufficient yield was obtained from a 100 µl PCR reaction in the presence of DMSO (1% (v/v)). Digests (using Nde I and Hind III) were performed on the expression plasmid, namely pET22b, and the amplified insert, GPI-PLD4a that had been previously purified using the specified restriction enzymes (Table 3.3). Digestion with restriction enzymes was successful as determined by running each of the 100 µl digests on an agarose gel. Both plasmid and insert DNA were successfully recovered from the gel and purified using a gel extraction and purification kit (Qiagen) (Figure 3.11) followed by ligation and then transformation into competent XL1-Blue *E. coli* (a non-expression host) as described in section 2.3.6 and 2.3.9, respectively). The results of performing a colony PCR followed by agarose gel electrophoresis suggested that all colonies contained an insert of the correct size (Figure 3.12). Consequently, three colonies were selected for sequencing to assess the accuracy of amplification.

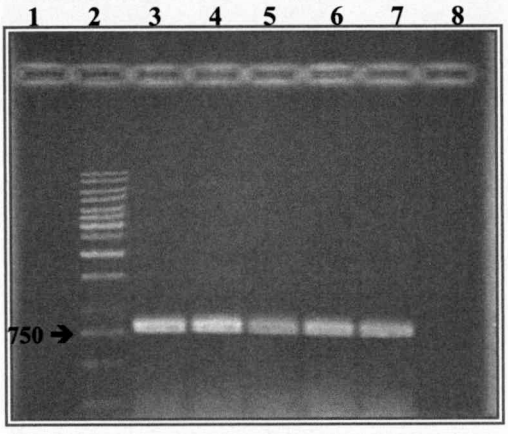


Fig; 3.10: PCR optimization using enhancers

PCR reactions were carried out at 45°C, 50°C, 55°C and 60°C. DMSO (1% (v/v)) and Platimun Pfx DNA polymerase enhancer buffer (Invitrogen) were added to each PCR reaction to observe their effect on yield of amplification. 5 µl of each PCR reaction was loaded onto a 1% agarose gel with 1µl gel loading buffer (Fermentas) to separate DNA products. 6 µl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250 was loaded into end lanes to monitor migration.



Fig; 3.11: Recovery of target insert and plasmid after restriction digestion.
5 µl of the restriction digest reaction of both the target insert and the expression plasmid was loaded onto a 1% agarose gel and separated from any contaminants by electrophoresis. Once separated, bands containing the digested insert and digested plasmid were recovered using a gel extraction and purification kit (Qiagen) as detailed in section 2.3.11. 6 µl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250 was loaded in the first lane to monitor migration of DNA.



Fig; 3.12: Identification of colonies containing GPI-PLD4a using colony PCR
Single colonies from transformed XL1-Blue cells were selected and grown up overnight in LB (50 μg/ml ampicillin). A PCR was performed on 5 μl of each culture. 5 μl of each colony PCR was run on a 1% agarose gel. The gel shows colony PCR of colonies 1-5. All of which contained an insert of the correct size. 6 μl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250 was loaded in the first lane to monitor migration of DNA.

A Blast sequence alignment (NCBI) of GPI-PLD clone sequence against each of the three colony sequences revealed a silent mutation, C-T substitution, in colony 2 (results not shown) that did not alter the coding amino acid. Colony 1 failed to sequence. Colony 3 was a perfect alignment. This colony was propagated overnight at 37°C in 100 ml LB/ampicillin broth and was used for plasmid isolation using a plasmid midi kit (Qiagen) according to manufacturers instructions. The resulting expression vector construct, pET22b-GPI-PLD4a, was then used for transformation of a desired *E. coli* strain capable of expressing the foreign DNA insert.

3.2.3: Variation of host strain of *E. coli* for expression of recombinant GPI-PLD4a

BL21-Gold(DE3)pLysS (Stratagene) cells were initially the chosen strain to host GPI-PLD4a expression. Full details of expression can be found in section 2.3.12. In brief, a 10 ml BL21-Gold(DE3)pLysS -pET226-GPI-PLD4a culture was grown up in LB broth containing ampicillin (50 μg/ml) and chloramphenicol (35 μg/ml)) and induced using isopropyl β-D-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Induced cultures were incubated at 37°C with shaking for 3 h. Samples were removed aseptically at 0 h (immediately prior to IPTG addition), 1 h, 2 h and at 3 h after the point of induction to determine the time at which maximum expression of the recombinant protein was achieved. Cells were lysed by sonication (as detailed in section 2.3.13.1), then centrifuged to separate soluble, supernatant fraction from insoluble, pellet fraction. SDS-PAGE was used to analyse expression and solubility states of the target protein.

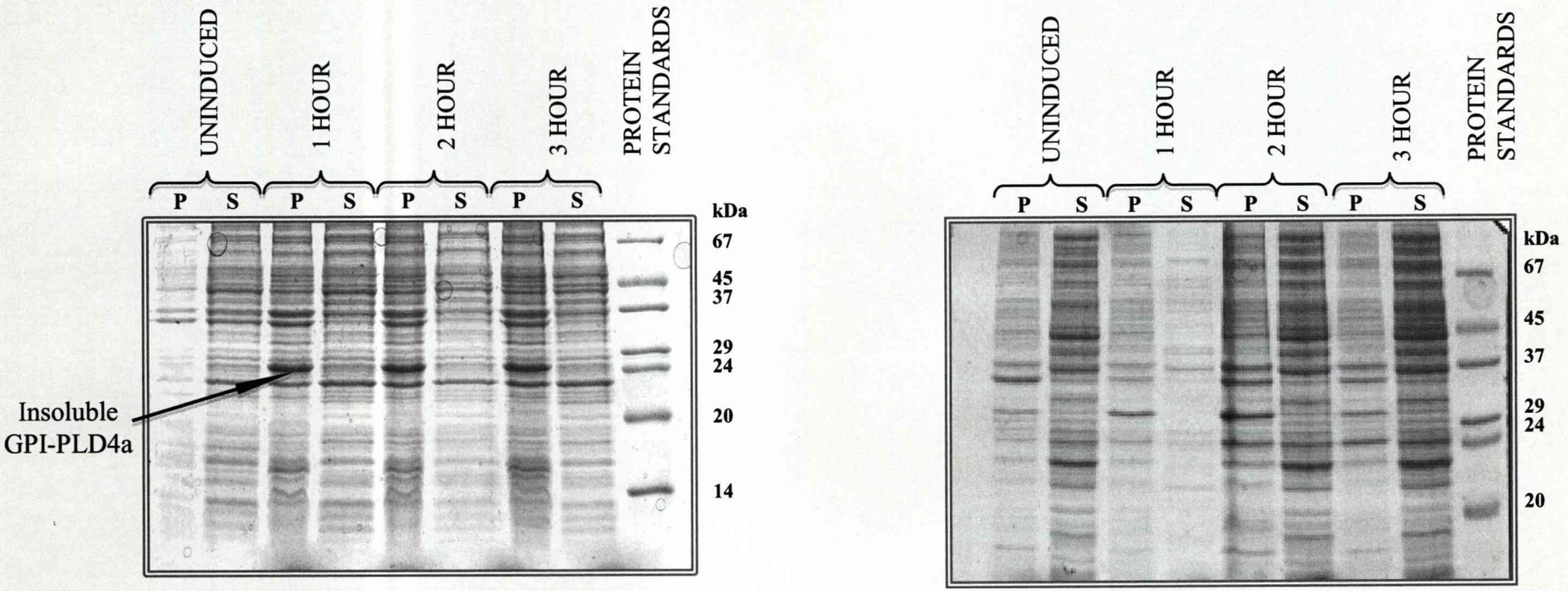
Expression of GPI-PLD4a directed from host strain BL21-Gold(DE3)pLysS resulted in only a small yield of recombinant protein, all of which was found in the pellet

fraction after an incubation of only 1 h post induction (Figure 3.13). This protein migrated to approximately 27 kDa. The occurrence of GPI-PLD4a in the pellet fraction indicates inclusion body formation and a strong probability of an incorrectly folded protein. The temperature at which expression of GPI-PLD4a was induced from the host strain was decreased to 22°C so as to reduce the rate of protein synthesis in the hope that, in doing so, cells would be provided with additional time for the correct folding of the protein product. However, induction at 22°C did not aid proper folding of the recombinant protein and seemed to impede insoluble GPI-PLD4a expression (Figure 3.13). Consequently, at this juncture a number of alternative *E. coli* host strains were employed to achieve expression of soluble, active GPI-PLD4a.

Expression of GPI-PLD4a was induced at 18°C, 25°C and 37°C in both C41(DE3) and C43(DE3) strains (as described in section 2.3.12). These BL21 derivative strains can be used to express proteins where other strains have expressed only inclusion bodies (Miroux and Walker, 1996). SDS-PAGE analysis revealed bands at 25 kDa in lanes loaded with insoluble extracts for cultures induced at 25°C and 37°C using 0.5 mM IPTG for both strains (Figure 3.14). These bands were not apparent in lanes loaded with samples for each un-induced culture and were assumed to be recombinant GPI-PLD4a. Interestingly, bands appeared to be present, although much lower in intensity, in the corresponding supernatant extracts. IMAC purification of the His-tagged-GPI-PLD4a was used to isolate recombinant GPI-PLD4a in the soluble supernatant extract.

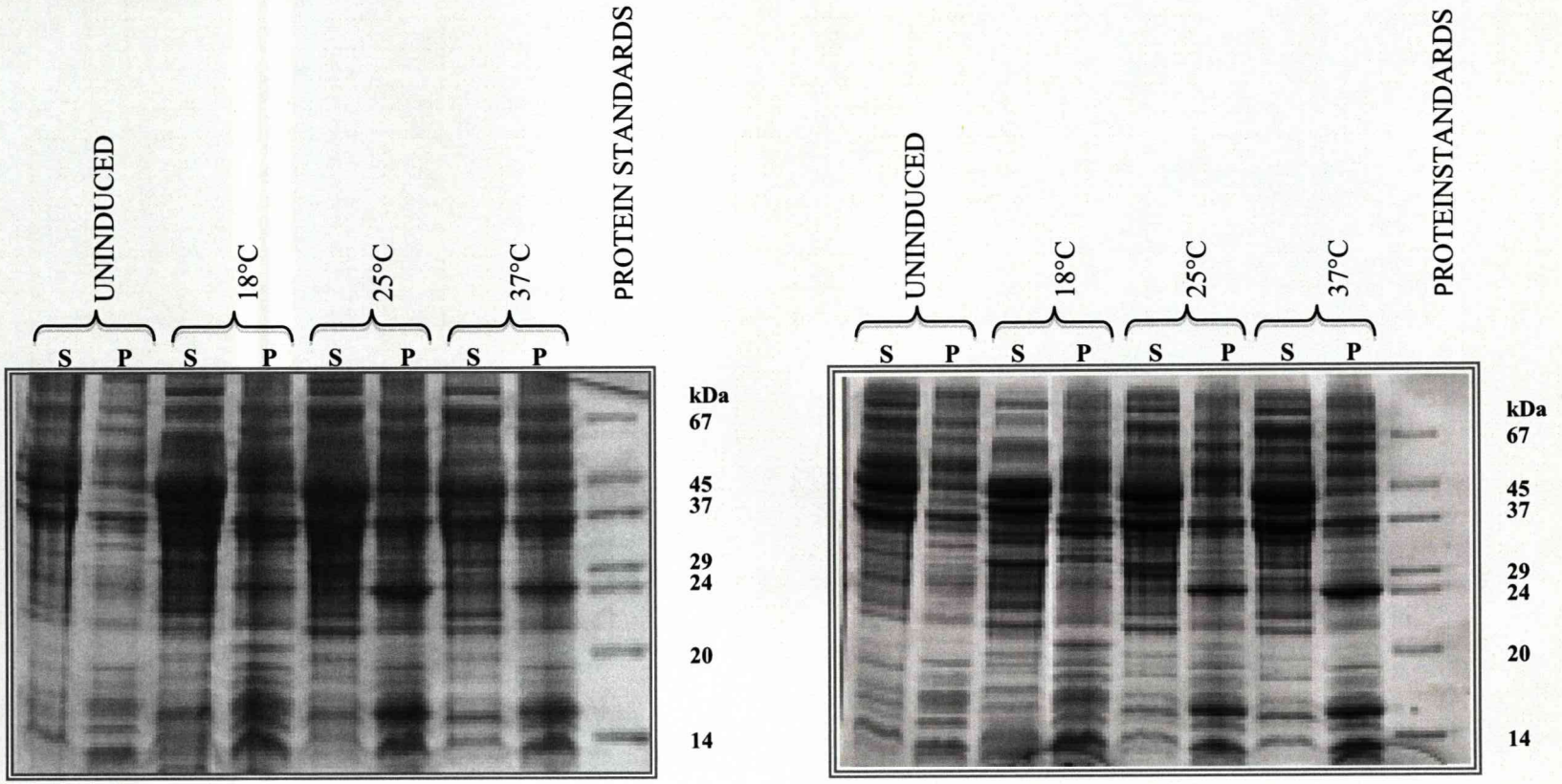
The supernatant cell extract of a 1 l culture of pET22b-GPI-PLD4a-transformed C41(DE3) cells (incubated for a period of 4 h post IPTG induction) was loaded onto

a HiTrap IMAC HP column (2 ml) as described previously in section 2.6.1.1. Bound proteins were eluted with 500 mM imidazole. SDS-PAGE was used to establish the presence of GPI-PLD4a in the final, 500 mM imidazole eluate. Results indicate the presence of a protein of the correct size in the 500 mM imidazole-eluted fraction (Figure 3.15).



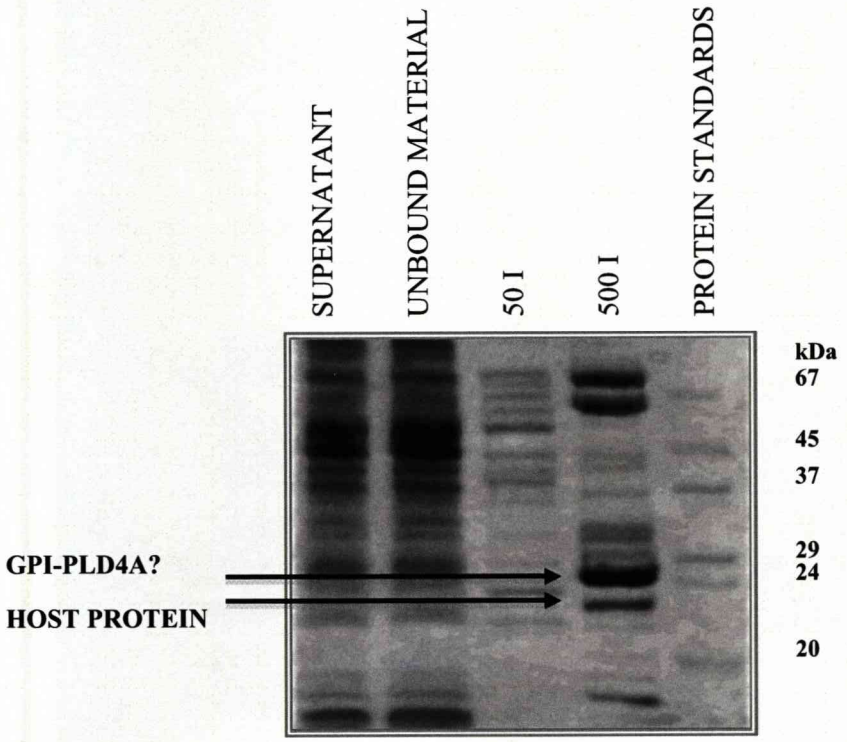
Fig; 3.13: SDS-PAGE analysis of soluble and insoluble fractions of GPI-PLD4a from BL21(DE3)pLysS *E.coli* at 37°C (left) and 22°C (right) over 3 hours.

BL21(DE3)pLysS-GPI-PLD4a cells were grown at 37°C to OD_{600} = 0.7 and then induced using IPTG (1 mM). IPTG-induced cultures were incubated at 37°C (left) and 22°C (right) for 3 h. Samples of 1 ml were taken at 0 h and at each hour for 3 h post induction. Cells were lysed by sonication and resulting extracts, labelled supernatant (s) and pellet (p), were assayed to determine protein solubility by SDS-PAGE.



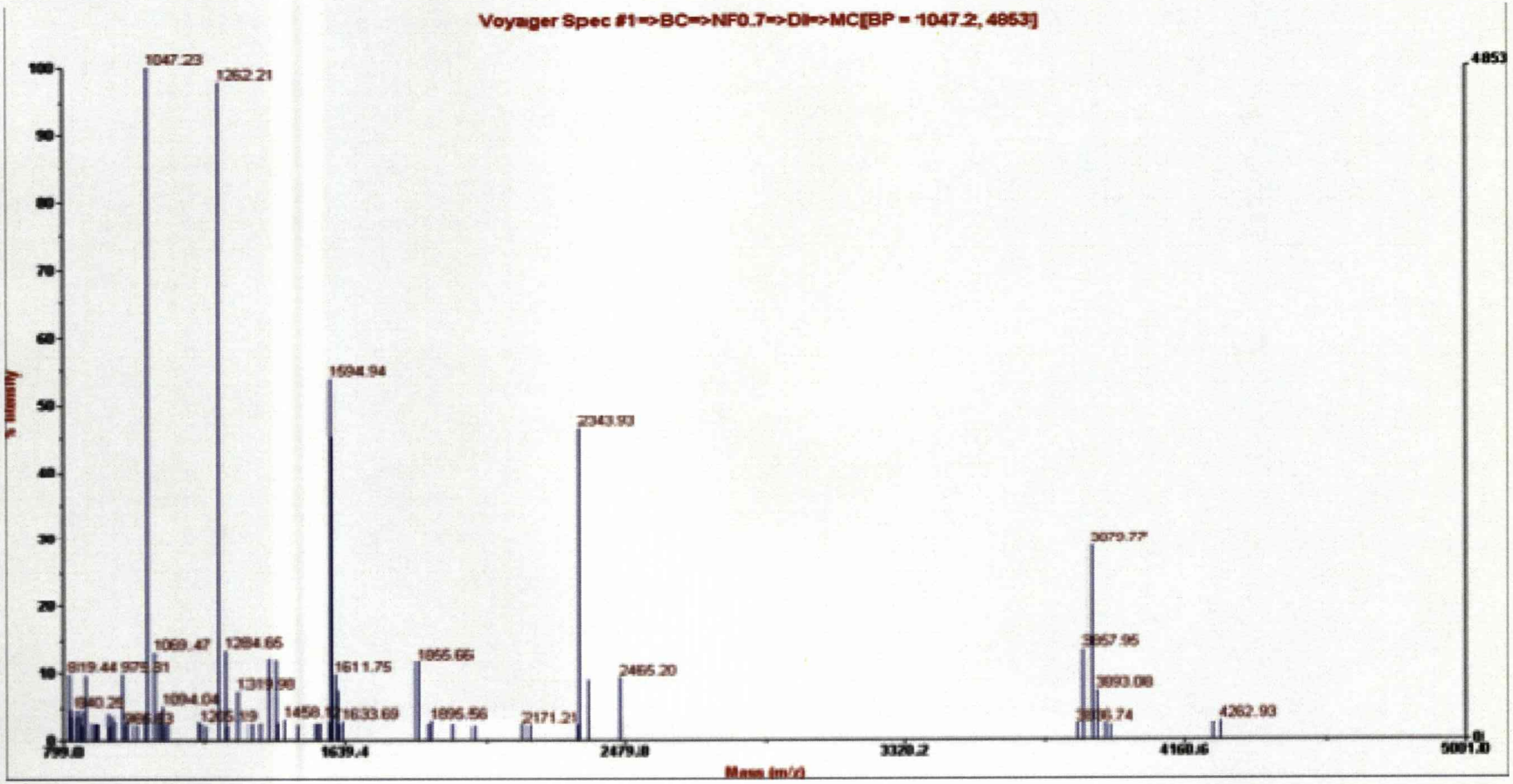
Fig; 3.14: SDS-PAGE analysis of soluble and insoluble fractions of GPI-PLD4a from C41(DE3)GPI-PLD4a (left) and C43(DE3)GPI-PLD4a (right) over 3 hours.

BL21(DE3)pLysS-GPI-PLD4a cells were grown at 37°C to OD_{600} = 0.7 and then induced using IPTG (0.5 mM). IPTG-induced cultures were incubated at 18°C, 25°C and 37°C for 3 h. Samples of 1 ml were taken at 0 h and after 3 h post induction. Cells were lysed by sonication and resulting extracts, (after centrifugation) labelled supernatant (s) and pellet (p), were assayed to determine protein solubility by SDS-PAGE.



Fig; 3.15: SDS-PAGE analysis of IMAC purification fractions from the supernatant of C41(DE3)-GPI-PLD culture.
Cultures were induced at 25°C using IPTG (0.5 mM) over 4 h. Cells were lysed by sonication. The supernatant extract of a 1 l culture was passed through a Ni-NTA column. Samples were removed when the supernatant was loaded, when unbound material passed through, when washed with 50 mM imidazole (50 I) and when finally eluted with 500 mM imidazole (500 I). Samples were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

In order to clarify that this band was representative of soluble GPI-PLD4a an in-gel digest with trypsin (detailed in section 2.1.4) was performed and run on a matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. A theoretical digest was carried out using Peptide Mass, ExPASY, in order to compare to data obtained from the MALDI-MS. MALDI-MS identified more than one protein present in this band (Figure 3.16). Furthermore, data obtained suggested that GPI-PLD4a was not likely to be the major protein in the isolated band. Only three, low intensity peptide masses (819.44, 1409.88, 1855.66) were matched with those predicted for GPI-PLD4a. Mascot identified 4/5 major peaks corresponding to a 21 kDa, native *E. coli* protein, namely FKBP-type peptidyl-prolyl cis-trans isomerase. This metal-binding protein has a histidine rich sequence and binding of this protein to the HiTrap IMAC HP column is plausible.

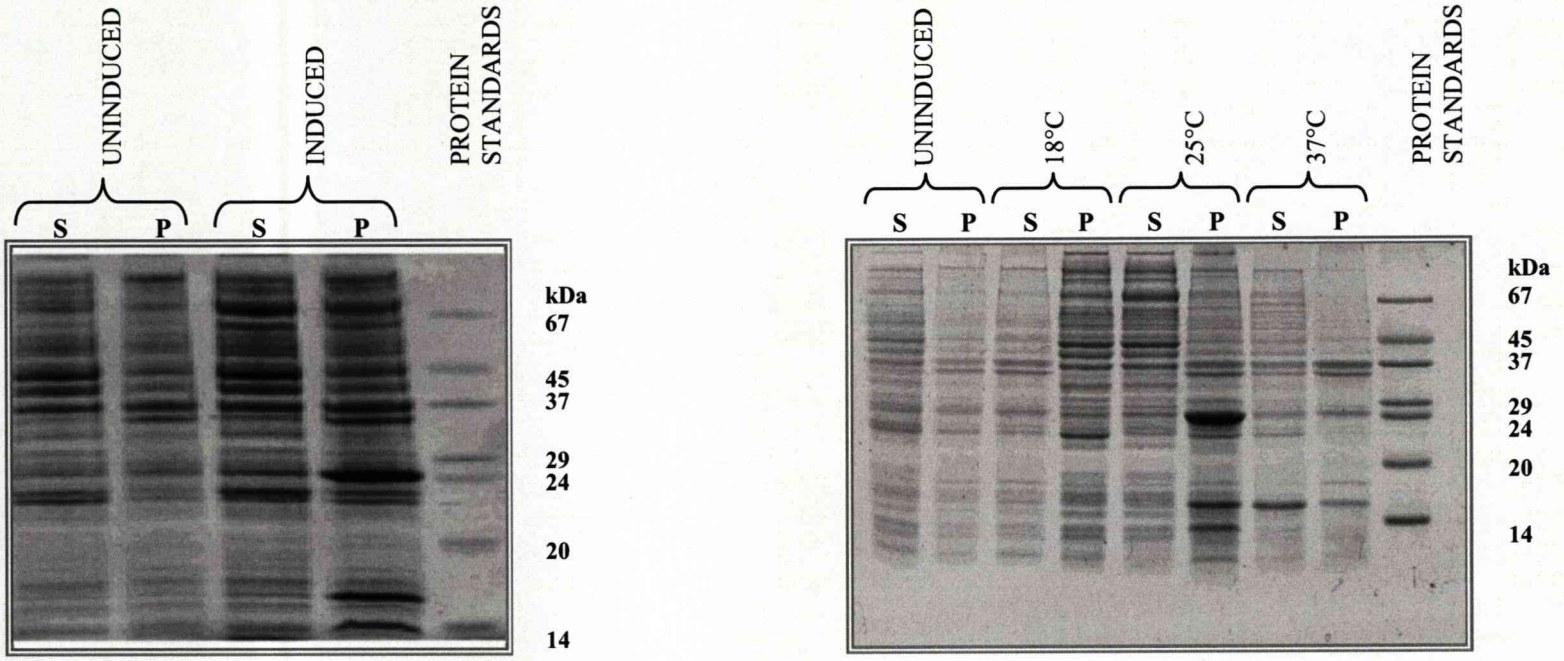


Fig; 3.16: MALDI-MS Data - Peptides Mass Intensities

A 1 l culture of pET22b-GPI-PLD4a-transformed C41(DE3) cells were incubated for a period of 4 h post IPTG induction at 25°C. After cell harvesting and cell lysis by sonication the resulting supernatant cell extract was loaded onto a HiTrap IMAC HP column. Eluates were analysed by SDS-PAGE. An in-gel digest with trypsin was performed on the 26 kDa band suspected to be the recombinant protein that was excised from the polyacrylamide gel and run on a matrix-assisted laser desorption/ionization (MALDI) mass spectrometer.

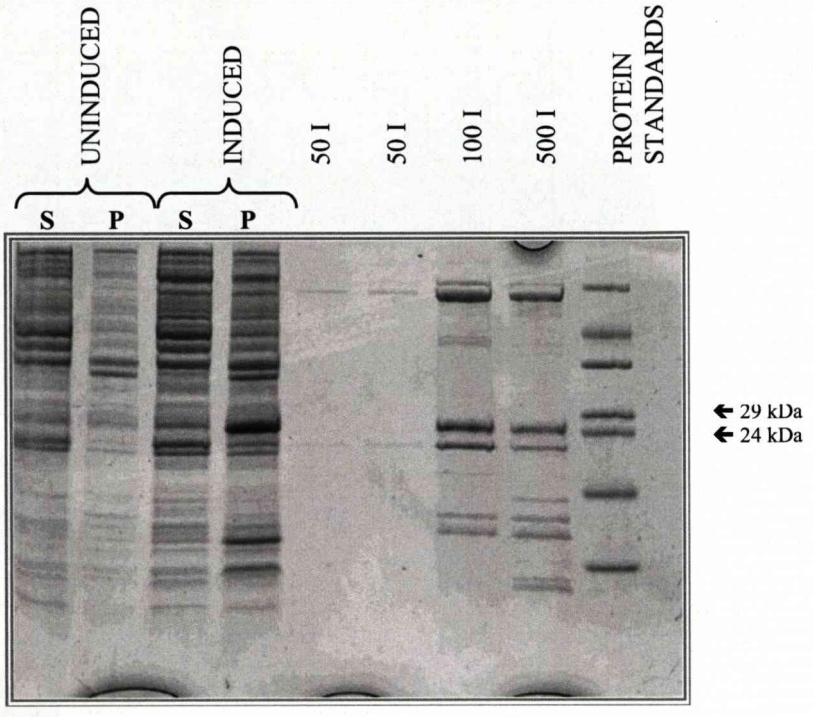
GPI-PLD4a expression was induced at 37°C for 2 h using Origami2(DE3) *E. coli* (a strain capable of disulphide bond formation) . Additional IPTG-induced cultures were incubated at 18°C, 25°C and 37°C for 16 h. As with C41(DE3) and C43(DE3) strains, SDS-PAGE analysis demonstrated expression of a protein at 25 kDa in pellet samples from the 37°C culture that was incubated for only 2 h but, in this case, not for an extended incubation (Figure 3.17). The highest level of target expression is evident in the culture induced at 25°C in the overnight incubation (Figure 3.17). Interestingly, a protein band is present 16 kDa in the pellet and supernatant fractions when expressed at 37°C. This band is not present in the uninduced cultures. This protein is most likely a native chaperone protein. Again, it was difficult to identify the presence of this protein band in the supernatant extract because of the sheer amount of contaminating host protein in this fraction.

A 1 l culture of pET22b-GPI-PLD4a-transformed Origami2(DE3) was cultivated so that the his-tagged recombinant protein could be separated from the majority of host proteins by IMAC (as detailed in section 2.6.1.1). SDS-PAGE analysis of eluted fractions clearly illustrates the presence of two protein bands at between 25.5 kDa and 26 kDa (Figure 3.18). This provides further confirmation that the band isolated from the IMAC-purified supernatant fraction of the C41(DE3)-GPI-PLD 4a (Figure 3.15) was in fact a duplicate band containing both the *E. coli* metal-binding protein (identified by MALDI-MS) and recombinant GPI-PLD4a. The lower band is of a lesser intensity, which is more likely to be the 21 kDa, FKBP-type peptidyl-prolyl cis-trans isomerase. This protein may be less protected from proteases than GPI-PLD thus more readily digested by trypsin.



Fig; 3.17: SDS-PAGE analysis of expression and solubility of GPI-PLD4a directed from Origami2(DE3).

Origami2(DE3)-GPI-PLD4a cells were grown at 37°C to OD_{600} = 0.7 and then induced using IPTG (1 mM). IPTG-induced cultures of 10 ml were incubated at 37°C for 2 h (*left*) and at 18°C, 25°C and 37°C for 16 h (*right)*. Samples of 1 ml were taken at 0 h to serve as an uninduced control and at 2 h or 16 h hour post induction for respective incubations. Cells were lysed by sonication and resulting extracts, labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.



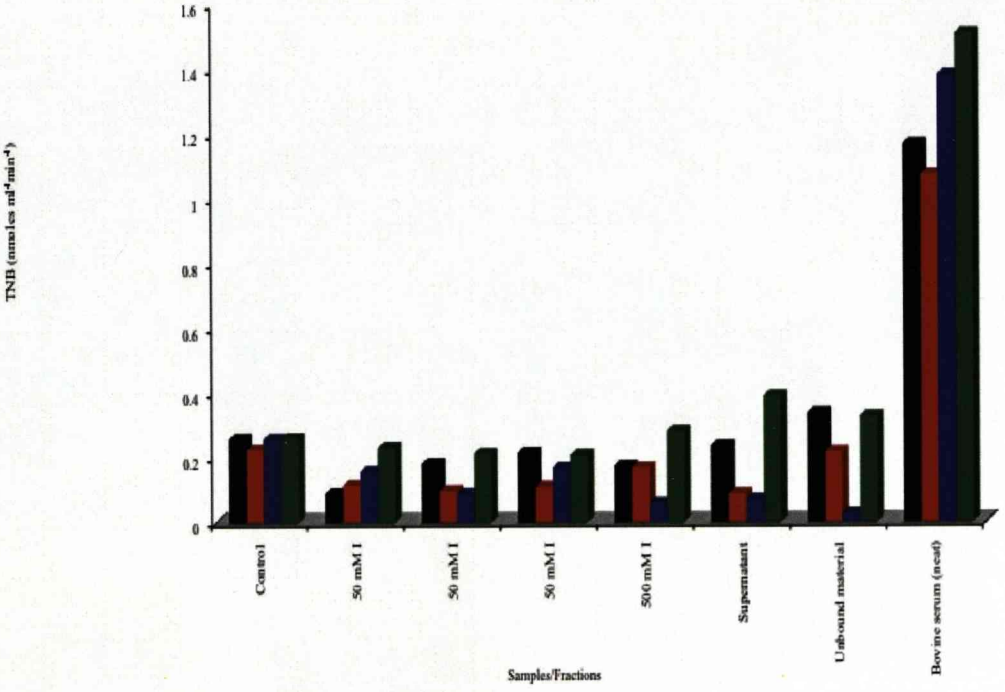
Fig; 3.18: Isolation of GPI-PLD from the supernatant of an IPTG-induced Origami2(DE3)-GPI-PLD4a culture using Ni^{2+} IMAC

A 1 l Origami2)DE3)-GPI-PLD4 culture was incubated at 25°C for 4 h post IPTG (1 mM)-induction. 1 ml samples were taken from cultures at 0 h (to serve as a non-induced control) and 4 h after point of induction. The remaining cells were lysed by sonication and fractions were separated by centrifugation. Samples of resulting extracts labelled supernatant (s) and pellet (p) were taken. The supernatant of a 1 l culture was passed through a HiTrap IMAC HP column. Samples were removed when the supernatant was loaded, washed with 50 mM imidazole and eluted with 100 and 500 mM imidazole in TBS. Samples were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

The fractions collected from IMAC were assayed for GPI-anchor degrading activity using mf-AChE from bovine erythrocytes as substrate to provide evidence for the presence of GPI-PLD. Results from this assay clearly show that no GPI-anchor degrading activity was evident in any of the fractions collected (Figure 3.19). Although Zn^{2+} has been identified as an important component for GPI-PLD catalytic activity (Raikwar *et al.*, 2005), the addition of Zn^{2+} at concentrations of 0.1 mM, 1 mM and 10 mM to the assay had no effect on GPI-PLD activity (Figure 3.19).

GPI-PLD activity could not be found in any of the fractions collected from IMAC purification of the supernatant fraction from a 1 l Origami2(DE3)-GPI-PLD4a culture. Investigations using C41(DE3) and Origami2(DE3) cells have shown a duplicate band comprising of an *E. coli* host protein and possibly GPI-PLD4a. Nevertheless, even if GPI-PLD4a was the major protein in this duplicate band, the problem with poor expression still remains and concentrations were insufficient for further purification and crystallisation trials.

The Rosetta2(DE3) strain was employed in the hope that the benefits of the additional tRNAs would aid the production of a soluble, full length recombinant protein. Expression studies herein resulted in the production of a protein of approximately 26 kDa (as determined by SDS-PAGE), which was smaller than expected for this target. Growth and induction conditions for expression of GPI-PLD4a from Rosetta2(DE3) cells was performed as detailed in section 2.3.12 with a couple of exceptions. Incubation following IPTG induction was carried out for a period of 16 h in an attempt to increase protein yields.



Fig; 3.19: GPI-PLD activity assay on eluted fractions from HiTrap IMAC column

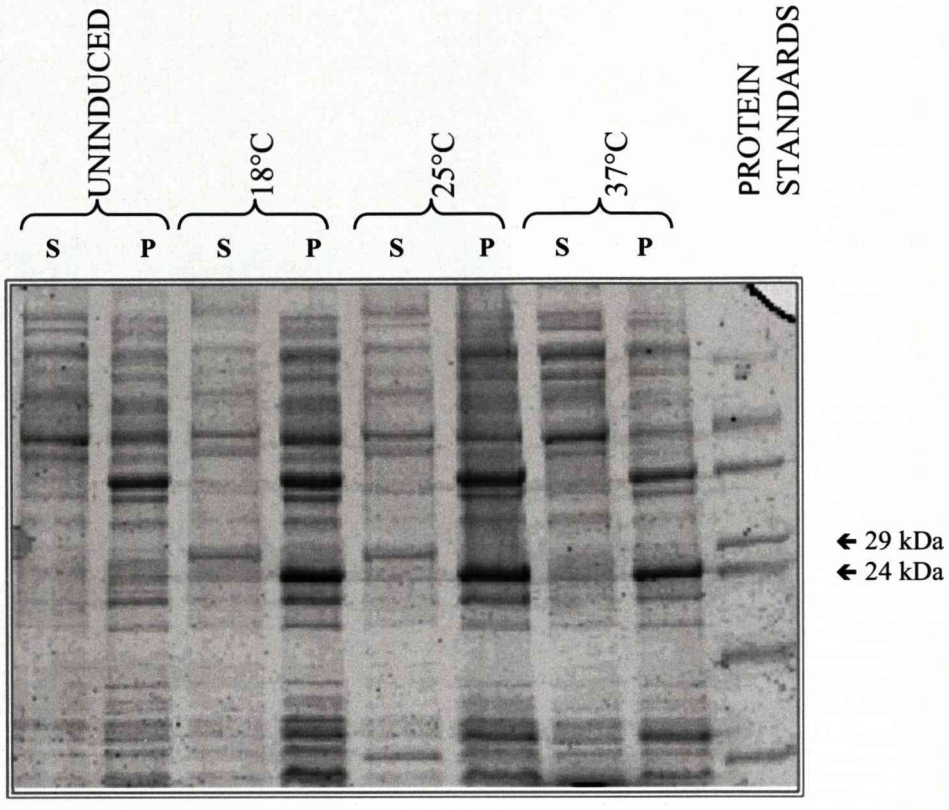
A HiTrap IMAC column was loaded with supernatant from an Origami2(DE3) pLysS culture induced with 1 mM IPTG at 25°C for 3 hours. 10 µl of each chromatography fraction was added to the Ellman assay to determine GPI-PLD activity. Zn^{2+} was added at the desired concentration (0.1 (BLACK), 1.0 (RED), 10 mM (BLUE) and a control with no added Zn^{2+} (GREEN)) for the 2 h incubation with 200 µl mfAChE (red bllod cell ghosts) and 10 µl of each chromatography fraction.

Induced Rosetta2(DE3)-GPI-PLD4a cultures were incubated at 18°C, 25°C and 37°C to determine the effect of temperature of expression (to be discussed later in section 3.2.4). Samples of 1 ml were removed from each culture (18°C, 25°C and 37°C) after 16 h post IPTG induction. A 1 ml sample was also taken immediately prior to induction with IPTG to serve as a non-induced control. SDS-PAGE was used to analyse expression and solubility states of GPI-PLD4a protein. GPI-PLD4a was expressed in this strain, albeit, in an insoluble aggregated form (Figure 3.20). Owing to the potential pitfalls that may be encountered, it was decided that time would be better employed optimising alternative parameters that influence expression and solubility

3.2.4: Optimisation of cultivation conditions

As mentioned earlier, the effect of temperature on expression and solubility of the recombinant protein was explored. Interestingly, a reduction in the temperature at the point of induction had varying effects on protein expression in each of the host strains studied (Figures 3.13, 3.14, 3.17, 3.20).

The length of time for incubation of induced cultures had a significant effect of expression yields. Generally, longer incubations of 16 h produced higher yields, but only of insoluble protein. The exception to this was incubation of IPTG-induced Origami2(DE3) cultures at 37°C due to host degradation in response to over-expression of the recombinant protein. With the aim of obtaining high levels of the target protein, it was decided, at this point, that all future expression cultures were to be incubated for 16 h to compensate for lower yields brought about by reduced expression temperatures.



Fig; 3.20: SDS-PAGE analysis of expression of GPI-PLD4a from Rosetta2(DE3) *E. coli*

Rosetta2(DE3)-GPI-PLD4a cells were grown at 37°C to OD_{600} = 0.7 and then induced using IPTG (1 mM). IPTG-induced cultures were incubated at 18°C, 25°C and 37°C for 16 h. Samples of 1 ml were taken at 0 h to serve as an uninduced control and at 16 h hour post induction for each of the incubations. Cells were lysed by sonication and resulting extracts (separated by centrifugation), labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

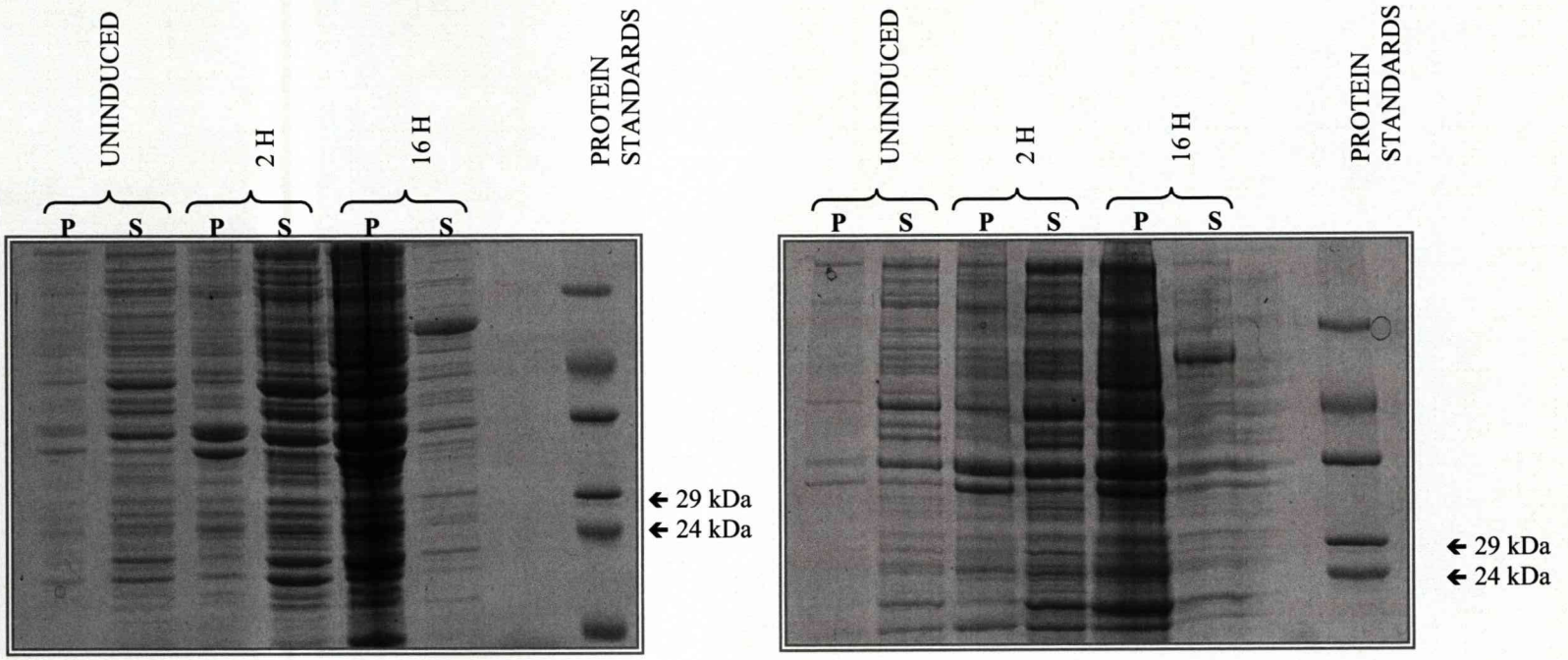
Lower concentrations of IPTG (0.05 mM – 0.2 mM) were also used to reduce translation rates. This presents the GPI-PLD protein with an increased chance of folding into a native state prior to aggregating with folding intermediates. Despite regulating IPTG-induction from host strains, Origami2(DE3) and Rosetta2(DE3), no enhancement in solubility was experienced following protein expression using lower concentrations of the IPTG (Figures 3.21-3.22(A)). Even concentrations as low as 0.05 mM failed to have any effect on solubility of the target when expressed in Rosetta2(DE3) (Figure 3.22(A)).

The point of IPTG-induction of the bacterial culture (harvesting the pET22b-GPI-PLD4 expression plasmid) was varied. IPTG was added at a final concentration of 0.5 mM to each of the expression hosts (C41(DE3), C43(DE3) and Rosetta2(DE3)) at cultures grown to OD_{600} of 0.7 or OD_{600} of 1.3. The addition of IPTG at each cell density did not improve protein solubility nor did it increase target protein levels (Figures; 3.14, 3.22(B)-3.23). Even varying the incubation temperature (18°C, 30°C and 37°C) of each OD_{600} culture for C41(DE3) and C43(DE3) hosts had no effect (Figure 3.14 and 3.23).

The use of alternative medium was also tested as an attempt to reduce expression levels in order to enhanced protein solubility. First, the use of minimal medium was tested. An IPTG-induced Rosetta2(DE3)-GPI-PLD4a culture was incubated in minimal medium for 16 h. Unfortunately, yields were very low and, again, IB formation was evident (Figure 3.24). The next strategy was to increase the cellular concentration of osmolytes. Cells were grown up in LB medium containing 1 M sorbitol and 2.5 mM betaine. The high concentration of such additives is thought to

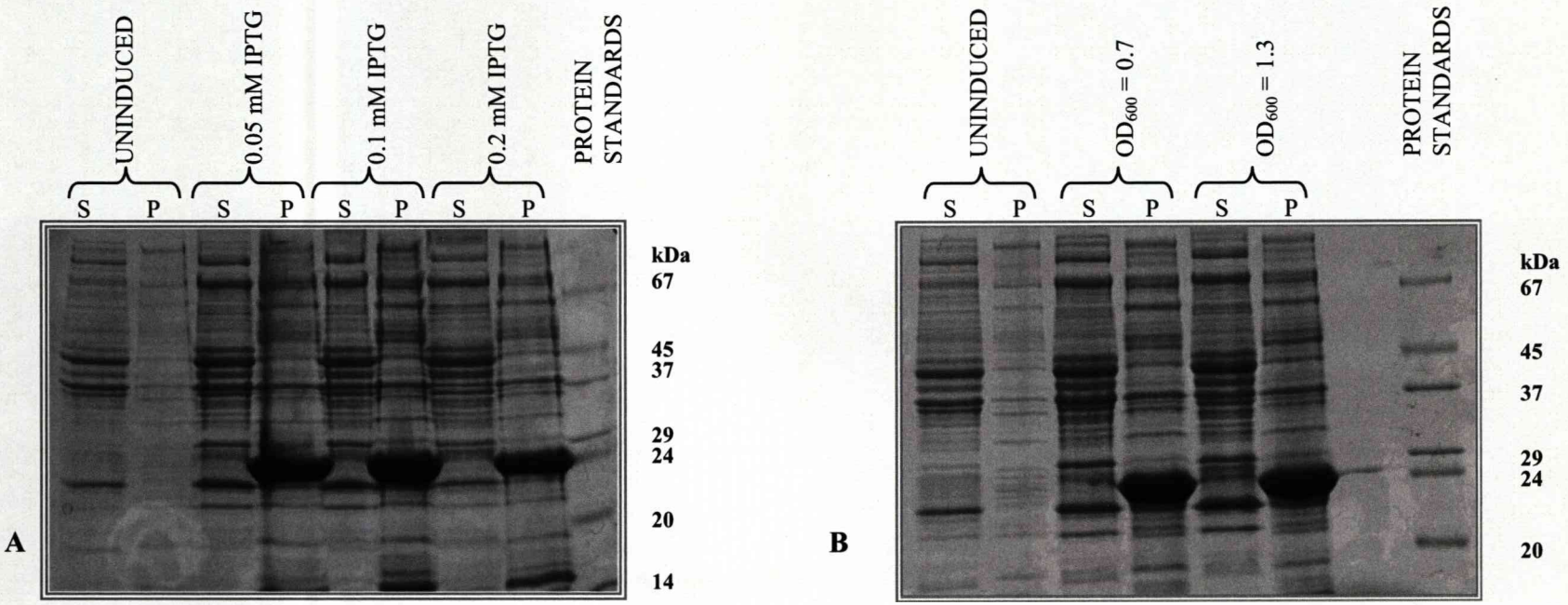
result in an increase in osmotic stress, which facilitates the uptake of the 'compatible solute', glycyl betaine. However, SDS-PAGE revealed that no benefit to protein solubility was experienced when expressed in Rosetta2(DE3) (Figure 3.24) and Origami2(DE3) (Figure 3.25) host strains.

Next, to ascertain if metal-directed folding could promote the correct folding of GPI-PLD4a, Zn^{2+} was added to the culture medium in the form of zinc sulphate ($ZnSO_4$) at a final concentration of 0.1 mM and 10 mM (for Rosetta2(DE3)-GPI-PLD4a). Metal addition to Origami2(DE3) and Rosetta2(DE3) cultures harbouring the expression plasmid for GPI-PLD4a expression did not ameliorate inclusion body formation (Figure 3.26).



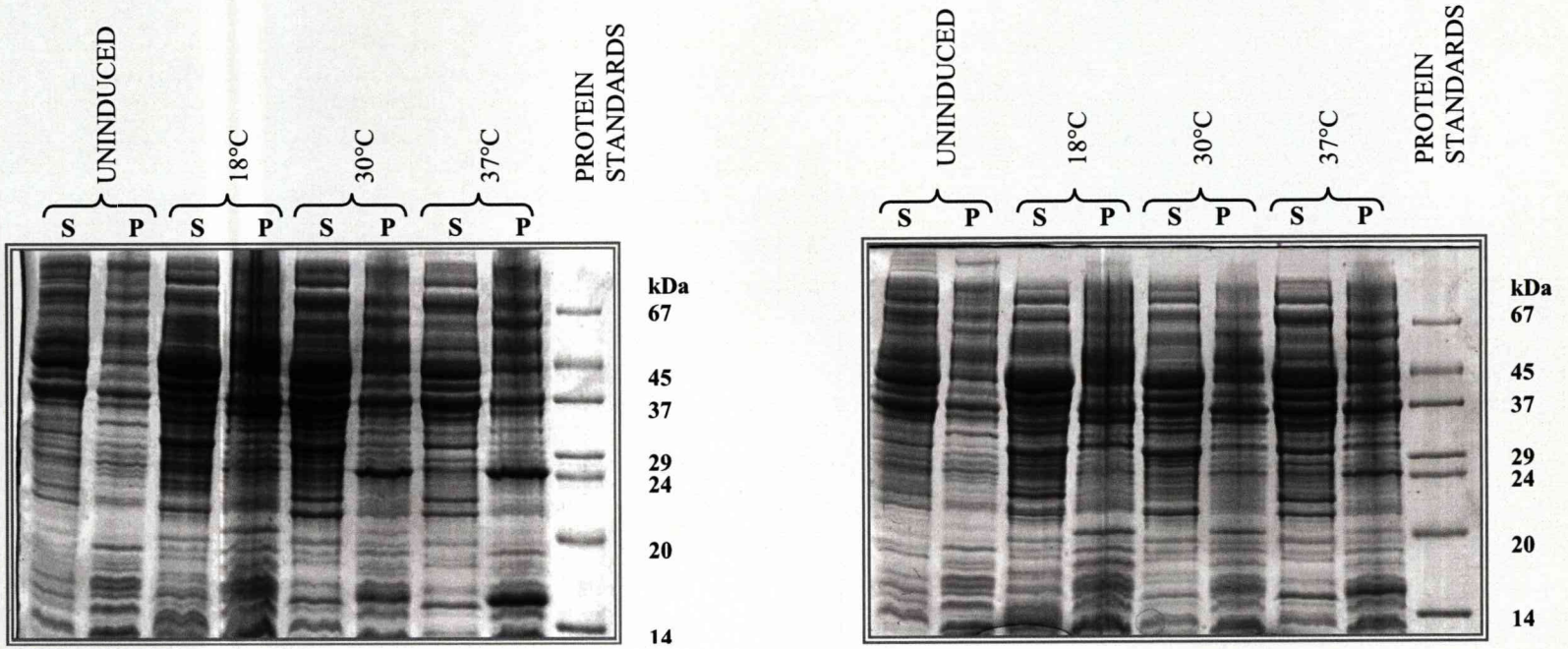
Fig; 3.21: Origami(DE3)pLysS-GPI-PLD4a expression with low IPTG concentrations.

Cultures were grown up at 37°C until OD_{600} = 0.7. The desired concentration of IPTG, 0.1 mM (left) or 0.5 mM (right), was added to induce GPI-PLD4a expression. Cultures were then incubated at 25°C, for 16 h with shaking (220 rpm). Samples of 1 ml were removed at 0 h (to serve as an uninduced control), at 2 h and at 16 h after induction. Cells were lysed by sonication and resulting extracts (separated by centrifugation) labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.



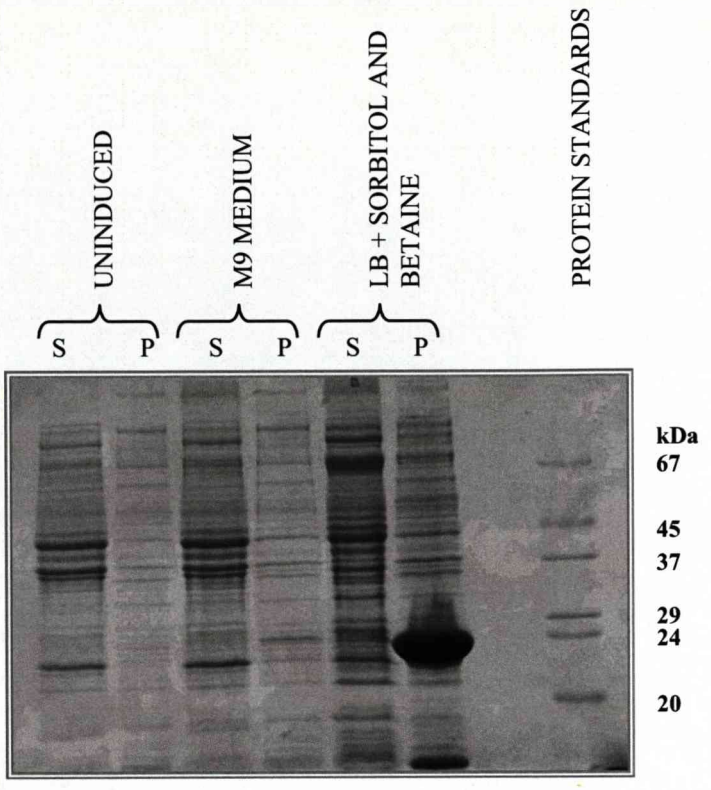
Fig; 3.22: Effect of IPTG inducer regulation on expression and solubility of GPI-PLD4a from Rosetta(DE3) cells

A) A Rosetta2(DE3)-GPI-PLD4 culture was grown at 37°C until OD_{600} = 0.7. Expression was then induced using respective IPTG concentrations (0.05 mM, 0.1 mM and 0.2 mM) and incubated for 16 h at 25°C, with shaking. B) A Rosetta2(DE3)-GPI-PLD4 culture was grown at 37°C until OD600 = 0.7 or 1.3. Expression was then induced in each culture using IPTG at a final concentration of 0.5 mM and incubated for 16 h 25°C, with shaking Samples of 1 ml were removed at 0 h (to serve as a control for uninduced cultures) and after 16 h. Cells were lysed by sonication and resulting extracts, labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.



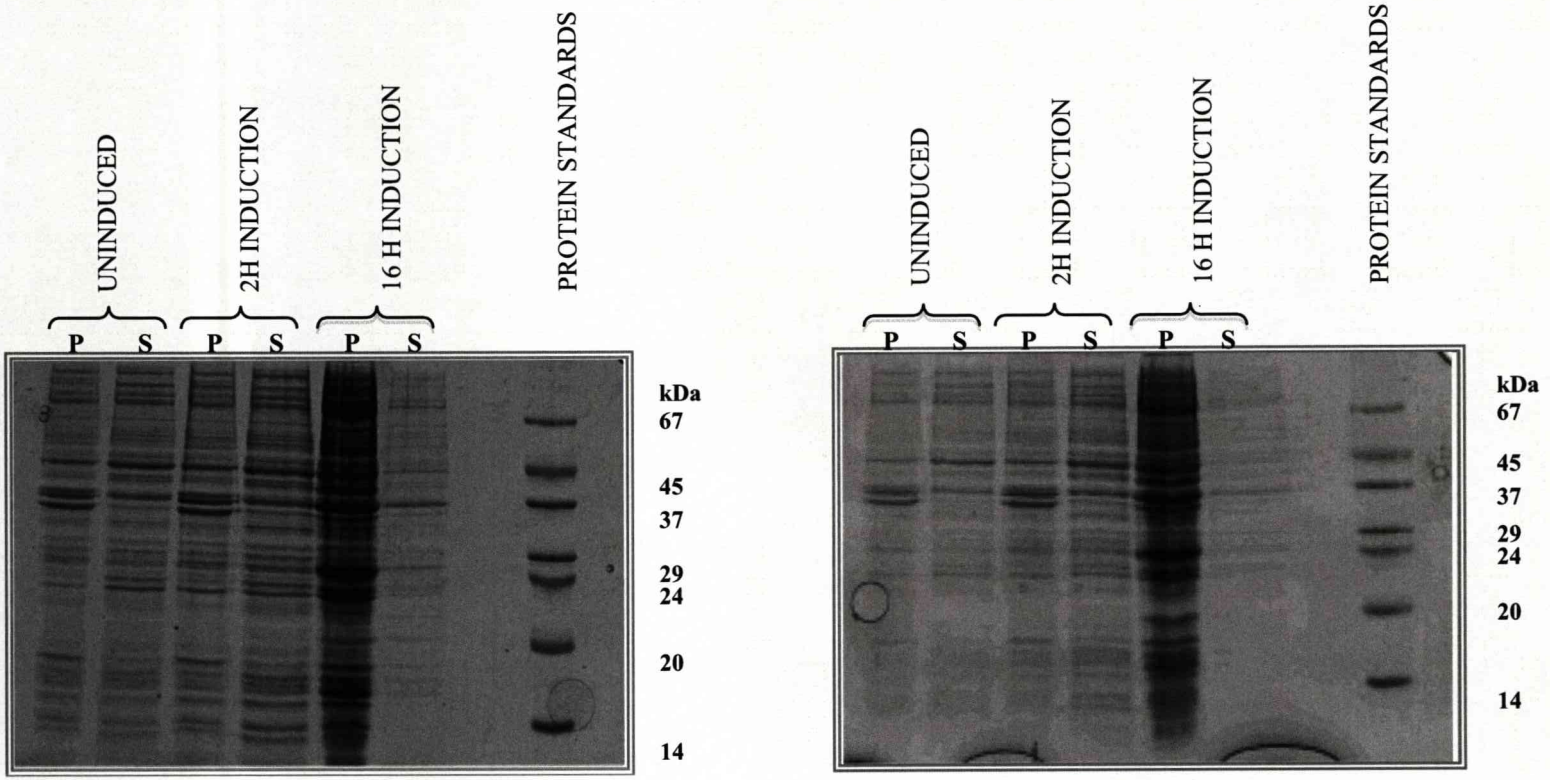
Fig; 3.23: SDS-PAGE analysis of GPI-PLD4a expression and solubility states from C41(DE3)GPI-PLD4a (left) and C43(DE3)GPI-PLD4a (right) induced at a higher cell density

Cultures were grown up at 37°C until OD_{600} = 1.3 and then induced using IPTG (0.5 mM). IPTG-induced cultures were incubated at 18°C, 25°C and 37°C for 16 h. Samples of 1 ml were taken at 0 h (uninduced control) and at 16 h post induction. Cells were lysed by sonication and resulting extracts (after centrifugation), labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.



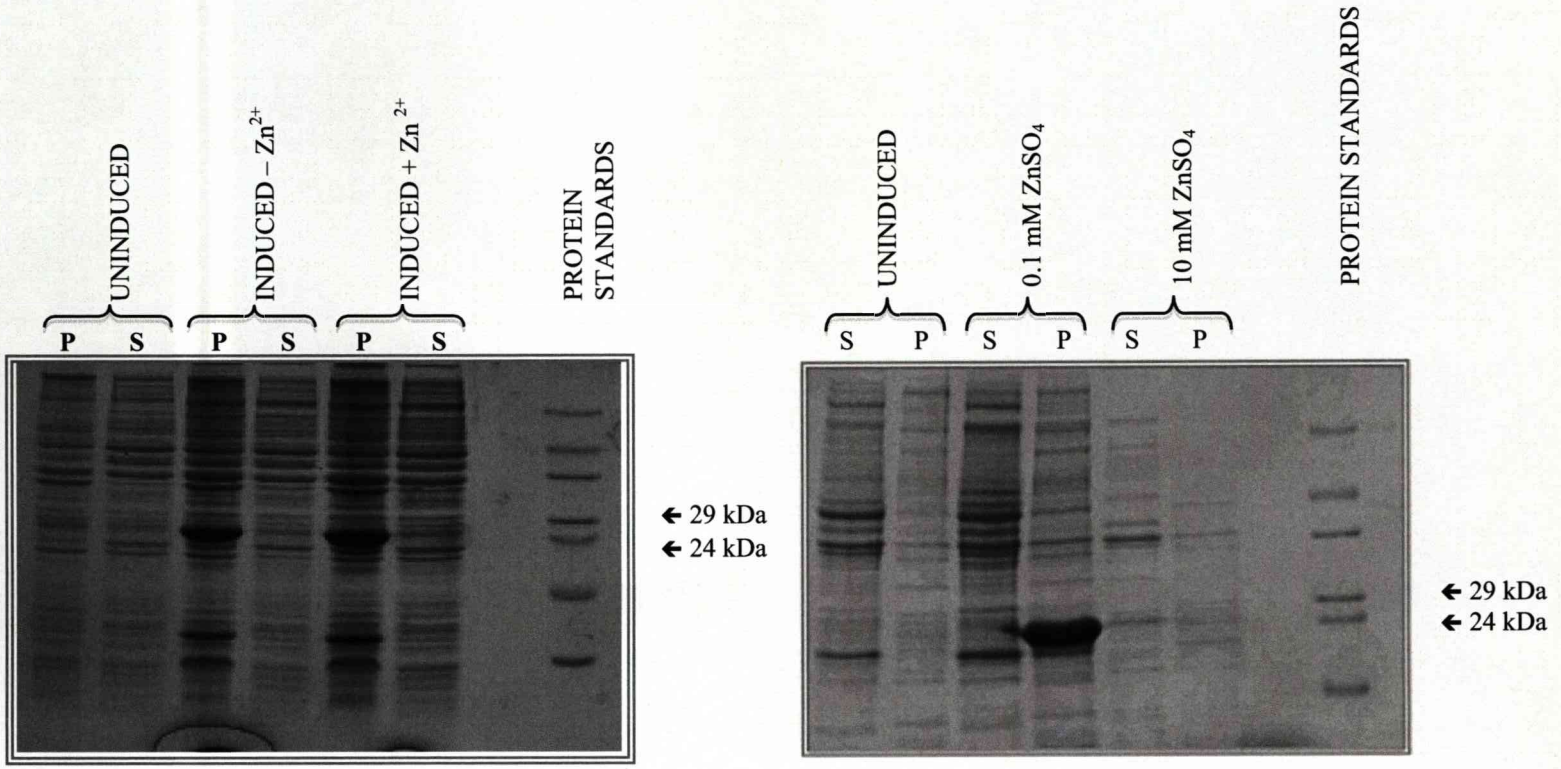
Fig; 3.24: SDS-PAGE analysis of Rosetta2(DE3)-GPI-PLD4a expression using alternative media and additives

Rosetta2(DE3)-GPI-PLD4a cultures were grown at 37°C until OD_{600} = 0.7. Cells were harvested, then resupended in M9 media or LB medium supplemented with sorbitol and betaine. Expression was induced at using IPTG (0.5 mM). IPTG-induced cultures were incubated at 25°C for 16 h. Samples of 1 ml were taken from each culture at 0 h (uninduced control) and at 16 h post induction. Cells were lysed by sonication and resulting extracts (after centrifugation), labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.



Fig; 3.25: Expression of GPI-PLD under osmotic stress

Origami(DE3)pLysS cells were grown at 37°C up to an OD_{600} of 0.7. Cells were harvested and re-suspended in LB media containing sorbitol and betaine. Cultures were induced using 0.5 mM (left) and 1 mM (right) IPTG incubated at 25°C. Samples of 1 ml were taken at 0 h (to serve as an un-induced control) and at 16 h after induction. Cells were lysed by sonication and resulting extracts (after centrifugation) labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.



Fig; 3.26: SDS-PAGE analysis of GPI-PLD$_{4a}$ expression at 25°C with the addition of Zn $^{2+}$.
Origami(DE3)pLysS-GPI-PLD4a cells (*left*) and Rosetta2(DE3) cells (*right*) were induced when cultures reached an OD_{600} of 0.7 using 1 mM IPTG. $ZnSO_4$ was added at a final concentration of 0.1 mM to Origami(DE3) cultures (*left*) and 0.1 mM and 10 mM to Rosetta2(DE3) cultures (*right*). Induced cultures were incubated at 25°C for 16 h with shaking (170 rpm). Samples of 1 ml were taken at 0 h and 16 h after induction. Cells were lysed by sonication and resulting extracts after centrifugation, labelled supernatant (s) and pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

3.2.5: Alternative GPI-PLD constructs with varying fusion proteins/tags

Eukaryotic proteins rarely lend themselves to over-expression in prokaryotic expression systems and so result, as witnessed herein, in inclusion body formation. A popular strategy in addressing this solubility problem is the employment of a large, soluble protein fusion/partner to improve solubility and folding of the protein in question.

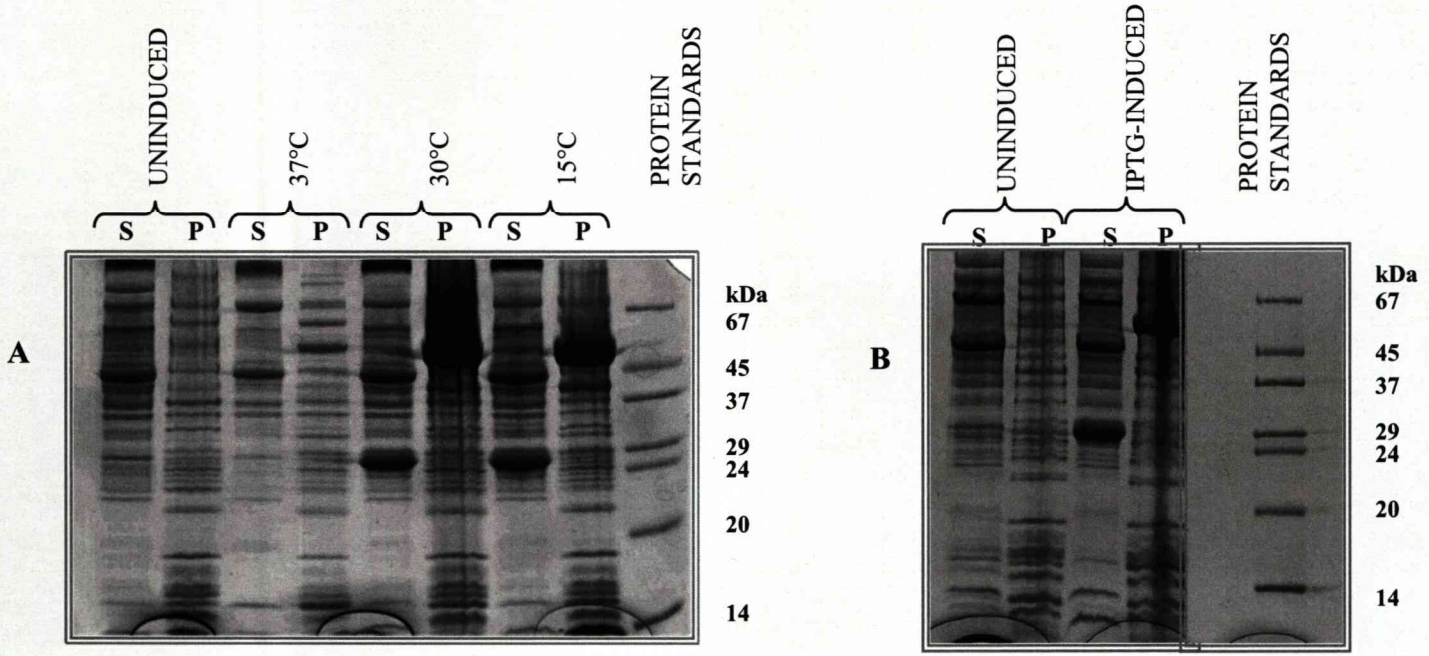
First, an N-terminal His(6X)-GPI-PLD4b fusion was created as described above for GPI-PLD4a with the exception of the reverse primer, which was replaced with Rev-4b, and the use of the plasmid expression vector, pET28a (Table 3.3). Sequencing of the GPI-PLD4b construct obtained from insoluble protein produced in Rosetta2(DE3) and excised from a polyacrylamide gel following electrophoresis, identified the first 6 N-term amino acids (G S S H H H) confirming the identity of N-His(X6) tagged GPI-PLD4b. This confirmed that the target protein was induced.

A GST-GPI-PLD fusion was also generated using the pGEX-6P1 expression plasmid (Amersham Biosciences). Forward and reverse primers were designed so that the GPI-PLD4 insert could be ligated into the plasmid using Xho I and Not I restriction sites, respectively (Table 3.3). Rosetta2(DE3) cells were the chosen host for GPI-PLD4-GST expression. Cells were cultivated at 37°C until the desired culture density (OD_{600} = 0.7) at which point expression was induced using 0.5 mM IPTG at 37°C, 25°C, and 15°C. An additional culture was induced using only 0.1 mM IPTG at 25°C to determine if the use of lower IPTG concentration had any effect on protein solubility. Results show that neither regulation of temperature nor of IPTG concentration had any effect on solubility (Figure 3.27). Results show a significant

band at 26 kDa in the supernatant after cell lysis using B-PER Reagent (Pierce) (detailed in section 2.3.13.2) of cultures induced at temperatures below 37°C. This is more than likely expression of the GST moiety alone rather than GPI-PLD. There are no bands in the corresponding pellet factions. A band at approximately 60 kDa is visible in all insoluble, pellet fractions. It is probable that this protein band is the GST-GPI-PLD4 fusion protein, which suggests, regrettably, that the GST fusion had no benefit to the expression of soluble GPI-PLD.

The next fusion strategy to be tested made use of the small size and stable fold of the immunoglobulin-binding domain of streptococcal protein G (GB1 domain) in preventing aggregation and to assist the correct folding of GPI-PLD4. This plasmid was engineered so that resulting fusion consisted of the GB1 domain at the amino terminus followed by a His(6X) tag, with the target protein counterpart (in this case GPI-PLD) at the carboxyl terminus.

The GPI-PLD4b construct (Table 3.3) served as a template for the GPI-PLD insert. The plasmid vector, pET28a, containing GPI-PLD sequence coding for the predicted active site, was cleaved with Nde I and Hind III restriction endonucleases. This allowed the release of the insert containing Nde I and Hind III overhangs. The expression plasmid, GB1-His6-pET24b was also digested with Nde I and Hind III to create complementary overhangs. To prevent re-ligation of pET28a and the GPI-PLD4b insert of the template, antarctic phosphatase (NEB) treatment was used to dephosphorylate the 5' phosphates of the DNA (as described in section 2.3.7). Nco I (NEB) digestion was also performed as a further precaution to cleave the previous host vector.

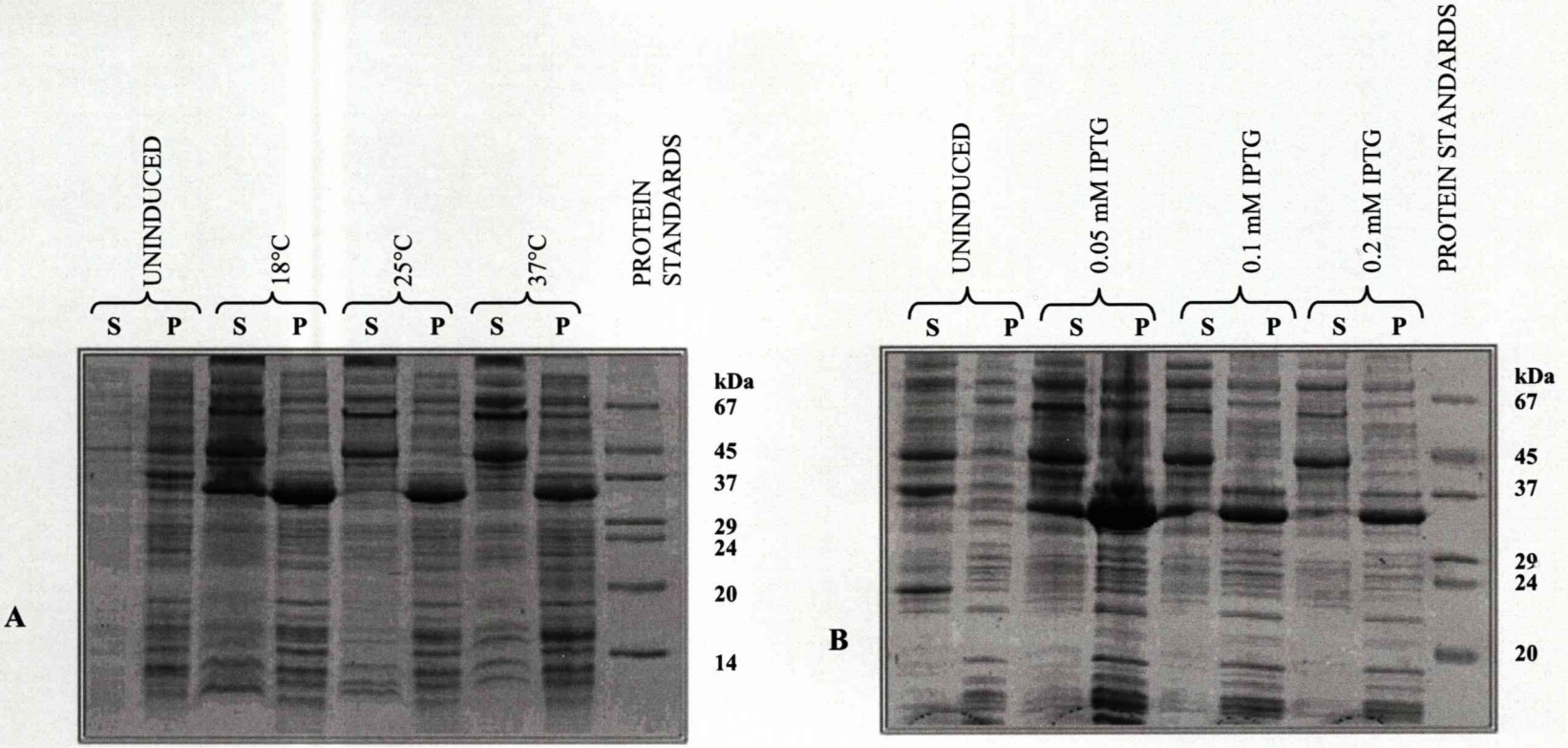


Fig; 3.27: Effect of temperature (A) and IPTG concentration (B) on expression and solubility of GST-GPI-PLD4

Rosetta2(DE3)-GST-GPI-PLD4 cultures was grown at 37° until OD_{600} = 0.7. *A)* Expression was then induced using 0.5 mM IPTG and cultures were incubated for 16 h at 37°C, 30°C and 15°C, with shaking. *B)* Expression was then induced using 0.1 mM IPTG and incubated for 16 h at 25°C with shaking. Samples of 1 ml were removed and cells were lysed using B-PER Reagent (Pierce). A centrifugation step separated soluble, supernatant (s) and insoluble pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

Rosetta2(DE3) *E. coli* was the chosen expression host. Induction of the GB1-His6-GPI-PLD fusion was carried out using low IPTG concentrations of 0.05 mM, 0.1 mM and 0.2 mM at 25°C and using 0.5 mM IPTG cultivations at 18°C, 25°C and 37°C. All inductions were performed over 16 h. It was noticed, at this point, that cell lysis by sonication continually resulted in inclusion body formation. However, when cell lysis was achieved by using B-PER lysis reagent a band of approximately the correct size was visible (Figure 3.28), which may be representative of the soluble GB1-His6-GPI-PLD fusion. The B-PER reagent contains a mild detergent that may help prevent protein aggregation.

These expression and solubility studies also demonstrated that GB1-His6-GPI-PLD4 was expressed as a soluble recombinant protein when protein expression was induced in Rosetta2(DE3) cells at a low temperature (18°C) and/or a low IPTG concentration (0.05 mM) (Figure 3.28). A cold shock step (4°C incubation for 10 min) prior to IPTG-induction was tried in an attempt to increase solubility, but had no effect on the yield of neither soluble nor insoluble recombinant protein (results not shown). Suspected soluble protein was electrophoretically transferred to PVDF membrane for protein sequencing by Edman degradation. The resulting sequence, M Q Y K L I, corresponded to the amino-terminus of the GB1 domain of the fusion, confirming that this was the recombinant protein.



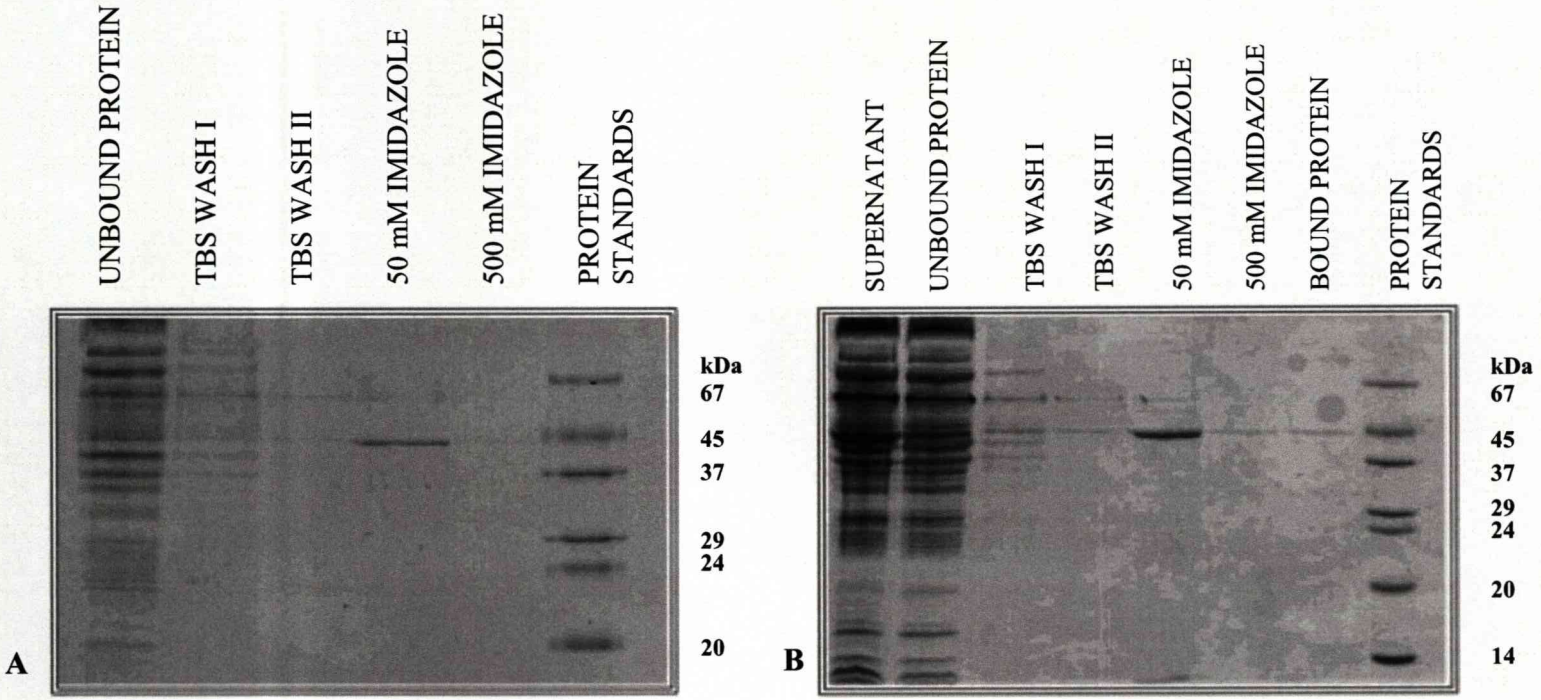
Fig; 3.28: Effect of temperature (A) and IPTG concentration (B) on expression and solubility of GB1-His6-GPI-PLD

Rosetta2(DE3)-pET24bGB1-His6-GPI-PLD cultures were grown at 37°C until OD_{600} = 0.7. *A)* Expression was then induced using 0.5 mM IPTG and cultures were incubated for 16 h at 18°C, 25°C and 37°C, with shaking. *B)* Expression was induced using 0.05 mM, 0.1 mM and 0.2 mM IPTG and incubated for 16 h at 25°C with shaking. Samples of 1 ml were removed and cells were lysed using B-PER(Pierce). A centrifugation step separated soluble, supernatant (s) and insoluble pellet (p) were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

SDS-PAGE analysis of eluates from small-scale IMAC purification of GB1-His6-GPI-PLD concluded that the protein did not bind to the Ni-NTA resin (Figure 3.29). An additional purification was executed with the supplementation of 0.1% Triton X-100 to wash and elution buffers. Figure 3.29 shows that this detergent had no effect on retaining the target protein after purification.

Due to the location of the His6-tag, i.e. inserted between GB1 protein and GPI-PLD protein, it is possible that the hexa-His unit is obscured thereby precluding interaction with Ni-NTA resin. An alternative methodology was to utilize the GB1 domain and its high affinity binding to IgG. Purification can be achieved using IgG (bovine) immobilized to CNBr-activated Sepharose 4B via affinity chromatography. A column was assembled (as described in section 2.6.3) for the purification of the supernatant fraction of a 0.5 l Rosetta2(DE3)-GPI-PLD4-GB1-His6 culture using IgG-affinity chromatography (as described in section 2.6.3). IgG affinity chromatography, in this case, proved to be unsuccessful (Figure 3.30).

As mentioned earlier, the GB1 domain is known to be immensely soluble. With this in mind an attempt at re-folding GB1-His6-GPI-PLD4 using 8M Urea and 25 mM DTT for denaturation, followed by a slow dialysis in buffer [20 mM Tris-Cl, pH 7.5, 0.1 M NaCl] to promote controlled refolding as described in detail in section 2.6.4. A significant amount of protein precipitation was noted after dialysis over 16 h.



Fig; 3.29: Isolation of GB1-His6-GPI-PLD from the supernatant of an IPTG-induced Rosetta2(DE3)pET24bGB1-His6-GPI-PLD culture using IMAC

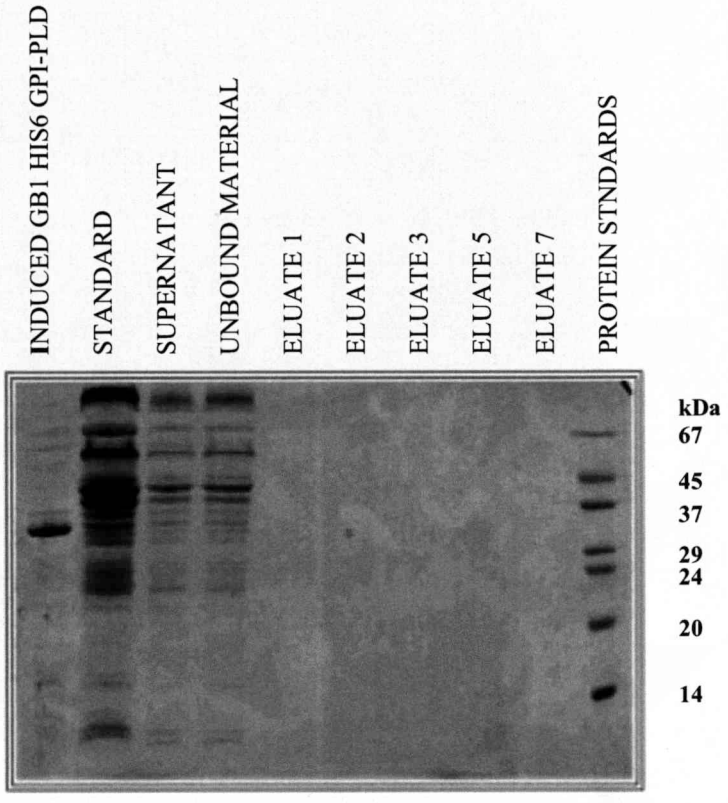
A 1 l Rosetta2(DE3)pET24bGBI-His6-GPID4 culture was incubated at 18°C (A) or 22°C (B) for 16 h post IPTG-induction (0.5 mM). 1 ml samples were taken from cultures at 0 h (to serve as a non-induced control) and 4 h after point of induction. The remaining cells were lysed using B-PER reagent and the soluble fraction separated from insoluble material by centrifugation. The supernatant of this culture was passed through a HiTrap IMAC column at a flow rate of 1 ml/min. Samples were removed when the supernatant was loaded, washed with TBS and 50 mM imidazole and eluted when with 500 mM imidazole in TBS. A sample of unbound material was also taken. 0.1% Triton X-100 was added to TBS to maintain protein solubility in *B*. All samples were loaded onto a 12% polyacrylamide gel for SDS-PAGE analysis.

This was likely to be the GB1-his fusion as this protein was only present in the unfolded, pellet fraction after the re-folding process (Figure 3.31). The decision not to assess the re-folding capabilities with alternative detergents was made based on the fact that refolding did not yield even a small quantity of soluble protein.

3.2.6: Effect of extending the length of GPI-PLD construct

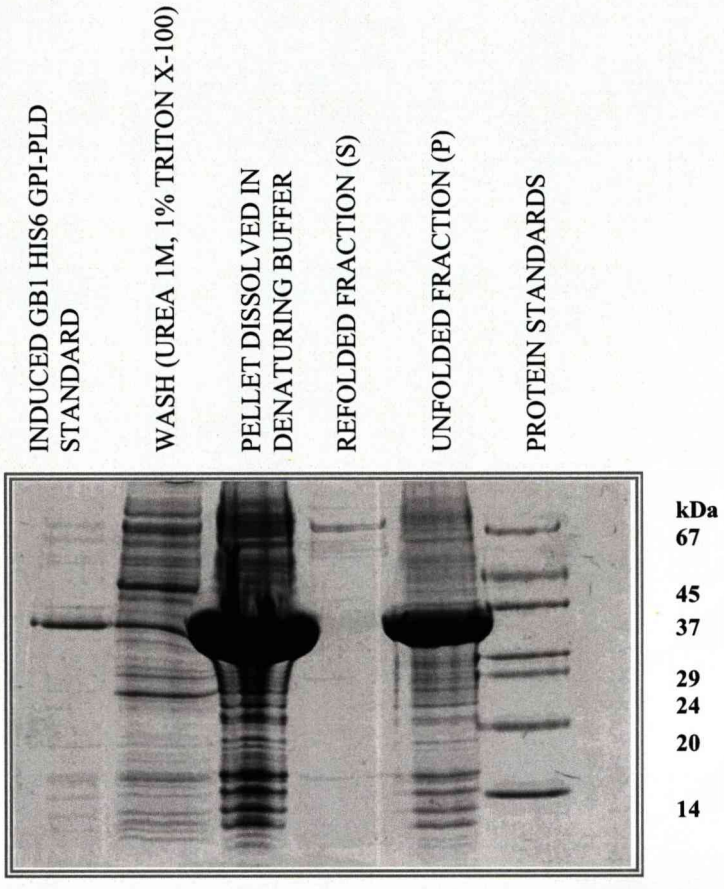
It is common knowledge that protein sequence is a contributing factor to the outcome of recombinant yield, solubility and success of crystallisation (Smialowski *et al.*, 2007; Smialowski *et al.*, 2006). All previous constructs were designed so that the GPI-PLD insert sequence terminated after the Cys residue within the predicted linker region (Figure 3.32). This theoretical structure, as determined by PSIPRED protein structure prediction server, suggested that the linker region consisted of residues 271 through to 305 of the GPI-PLD sequence (Figure 3.32). However, this was only a prediction and the possibility remains that GPI-PLD4 constructs were truncated to a position within the N-terminal, catalytic domain. The GPI-PLD3 construct (Table 3.3) was screened at this point to ensure that the full catalytic domain was being expressed.

The GPI-PLD3 truncation was amplified by PCR using the primers shown in Table 3.3. Primers were designed so that this construct could be successfully inserted into pET28a and pET22b for the expression of N- and C- terminal His fusions, respectively. Restriction enzymes, Nde I and Hind III, were used to create complementary overhangs required for ligation into each expression plasmid. Poor recovery of DNA inserts from agarose gels following digestion meant that molecular biology could not be progressed further.



Fig; 3.30: SDS-PAGE analysis of GPI-PLD-GB1-His6 purified using immobilized IgG affinity chromatography

Immobilized IgG affinity chromatography was used to purify the supernatant fraction collected after lysis (with B-PER reagent) of a 10 ml culture of induced Rosetta2(DE3)-pET24bGBI-GPID4 cells. Protein expression was induced using 0.05 mM IPTG at 22°C for 16 h. The supernatant was diluted 5X with buffer A [Tris-Cl 0.1 M, NaCl 0.5 M, 0.05% Tween 20 (pH 7.6)] then, loaded at a rate of 1 ml/min. A wash step with buffer A followed. Elution of bound GB1-tagged protein was achieved by running elution buffer (sodium acetate 0.5 M, pH 3.5) through the column and collecting 1 ml aliquots of the eluate for SDS-PAGE analysis.



Fig; 3.31: Refolding of GB1-His6-GPI-PLD4 using 8M Urea and overnight dialysis with TBS.

A pellet from a 500 mL lysed Rosetta2(DE3)-GB1-His6-GPI-PLD4 culture was washed in Tris-Cl 50 mM (pH 8.5), 1M Urea, 1% Triton X-100. A centrifugation step at 14 000g separated the washed pellet and the supernatant was discarded. The washed pellet was then re-suspended in denaturing buffer (Tris-Cl 50 mM (pH 8.5), Urea 8 M, DTT 25 mM and left mixing at room temperature for 1 h. A centrifugation step was used to remove cell debris then 5 ml of TBS was added sporadically to the supernatant over 30 min before being transferred to dialysis tubing and left to dialyse against 990 ml TBS (pH 8.5) for 16 h. A further dialysis against 990 ml for 4 h was required before SDS-PAGE analysis to determine if refolding was successful.

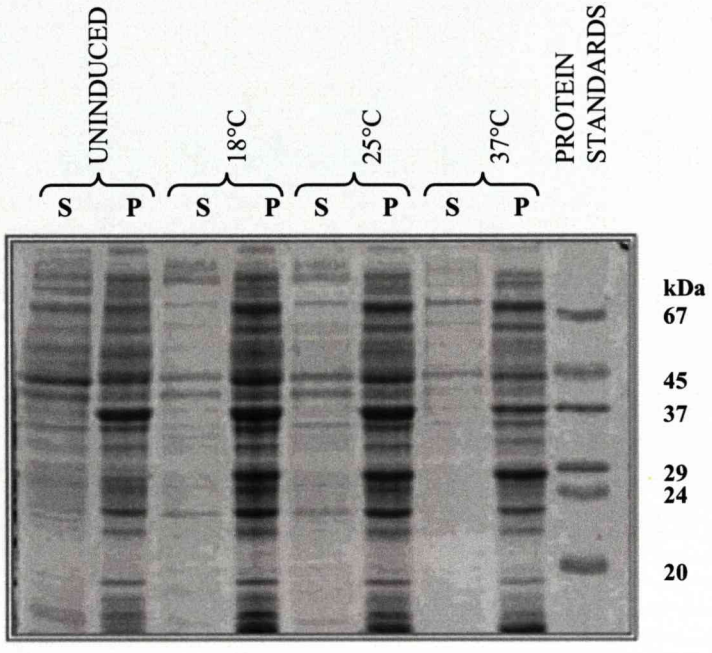
ccacgcgtccgctgaggaggaatgacaac***atg***tctgcaggcaggctgtggtctagcctgctg
T R P L R R N D N ***M*** S A G R L W S S L L
cttctgctgcctcttttctgctctaaaagctcatcttgtggtctctcaacacatgtagaa
L L L P L F C S K ***S S S C G*** L S T H V E
ataggacacagggctctggagtttcttcggcttcaagatggacgcattaactacaaagag
I G H R A L E F L R L Q D G R I N Y K E
ctgatcttagagcaccaggacgcatatcaggctgggaccgtgtttcctgatgccttttat
L I L E H Q D A Y Q A G T V F P D A F Y
cctagcatctgcaaaagaggaaaatatcatgacgtttctgagaggactcactggactcca
P S I C K R G K Y H D V S E R T H W T P
tttcttaacgccagcatccattatattcgagagaactaccctctgccctgggagaaggac
F L N A S I H Y I R E N Y P L P W E K D
acagagaagttggtggctttcttgtttggaatcacctcccacatggtcgctgacgtgagc
T E K L V A F L F G I T S H M V A D V S
tggcatagcctgggtattgaacaagggttcctcaggacaatgggagctatcgatttttac
W H S L G I E Q G F L R T M G A I D F Y
aactcttactctgacgctcactcggctggtgattttggaggagatgtgttgagccagttt
N S Y S D A H S A G D F G G D V L S Q F
gaatttaattttaattacctctcacggcgctggtacgtgcccgtcagggatcttctgaga
E F N F N Y L S R R W Y V P V R D L L R
tactctacaaagtccccatttctggtggagcagttccaagactatttcctcggaggtctg
Y S T K S P F L V E Q F Q D Y F L G G L
atttatgataatctctatggtcggaaagtcatcaccaaagacgtccttgttgattgcacc
I Y D N L Y G R K V I T K D V L V D C T
taccttcagttcctggaaatgcacggggagatgtttgctgtttccaagctctattccacg
Y L Q F L E M H G E M F A V S K L Y S T
gatgacatggcattctggtccacgaacatttaccgtttgaccagctttatgctggagaac
D D M A F W S T N I Y R L T S F M L E **N**
gggaccagtgac**tgc**aacctgcct**gag**aaccccctgttcatctcctgtgatggcaggaac
G T S D C N L P E N P L F I S C D G R N
cacaccctcagtggc**tca**aaagtgcagaaaaatgattttt**cac**aggaatttgaccatgttc
H T L S G S K V Q K N D F H R N L T M F
ataagtagagacatcaggaaaaacctcaattacacagaaagaggcgtgttctacagcaca
I S R D I R K N L N Y T E R G V F Y S T
ggctcctgggccccggaatctgtcacctttatgtaccagactctggagaggaacctgagg
G S W A P E S V T F M Y Q T L E R N L R

Fig; 3.32: Translation of the GPI-PLD clones nucleotide sequence – Highlighting linker region and point for construct extension.

Figure shows the translation of the N-terminal domain GPI-PLD nucleotide sequence highlighting the linker region (highlighted in green) is between N-terminal catalytic domain and C-terminal β-propeller domain. Reverse primers were designed so that constructs were truncated at residues highlighted in red as follows; GPI-PLD4 at the first C highlighted in red, GPI-PLD3 the next at highlighted residue E, GPI-PLD2 at S and GPI-PLD1 at H. The signal peptide sequence is shown in italics. The signal peptide sequence is shown in italics. SignalP (Bendtsen *et al.*, 2004) predicts that the signal peptide is cleaved at SSS-CG (blue). The new initiation codon replaces the third S residue in the signal peptide.

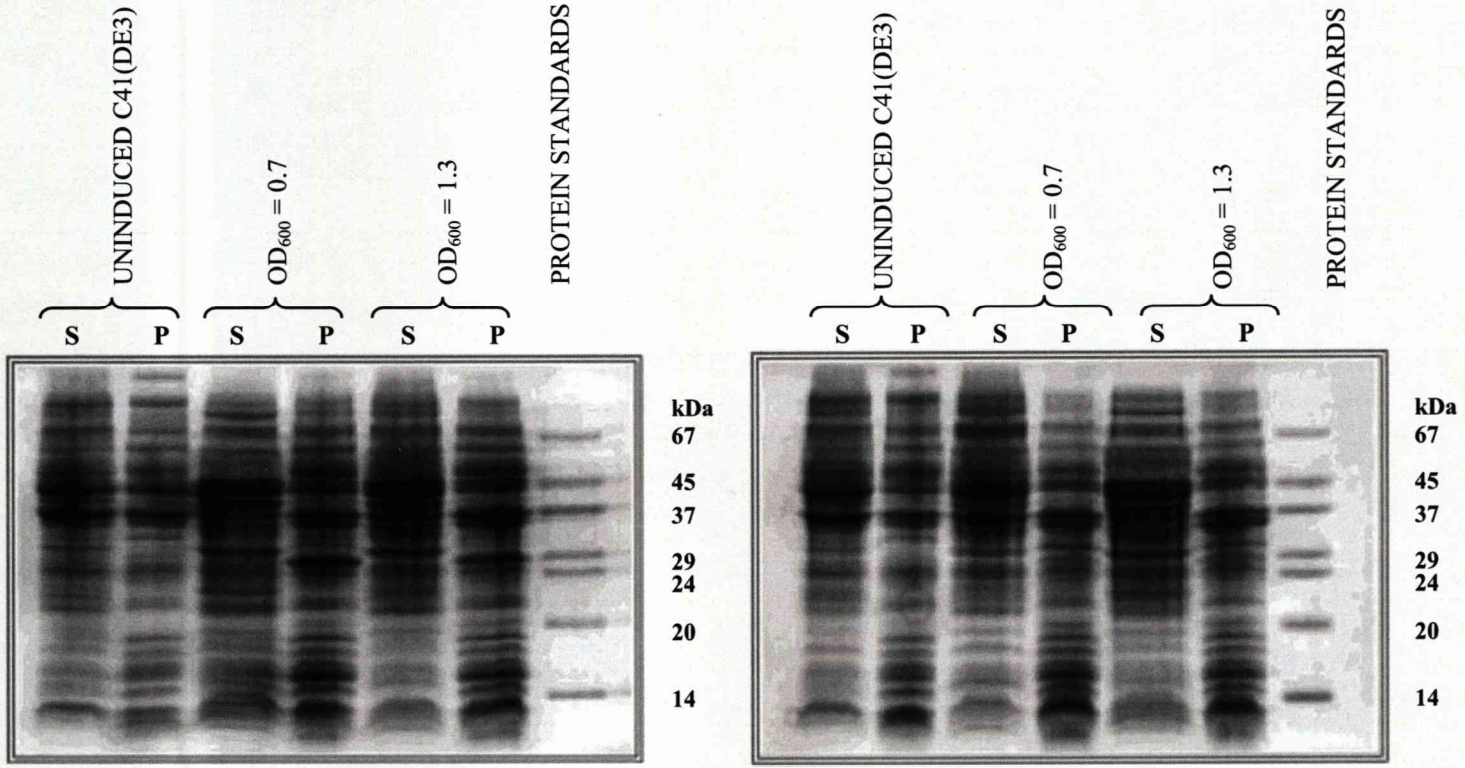
No explanation could be provided for this poor recovery because good yields were present prior to gel extraction and purification using a gel extraction and purification kit (Qiagen) and the pH of the DNA solution was within the required range. So, the method of purification was modified accordingly. Instead of extracting the digested DNA from the agarose gel and then purifying, each digest was purified directly using a PCR purification kit (Qiagen) and the gel extraction step was avoided. Ligations were then successfully carried out. Attempts were made for the construction of constructs GPI-PLD2 and GPI-PLD1, which were of increasing lengths, but both failed at the transformation stage. Colony PCR identified successful colonies containing only the GPI-PLD3b construct, one of which was selected for plasmid preparation and transformation into the Rosetta2(DE3), C41(DE3) and C43(DE3) expression hosts.

Expression and solubility testing of the increased length construct consisted of small-scale expression in Rosetta2(DE3) using 18°C, 25°C and 37°C temperatures for IPTG induction of GPI-PLD3b expression. C41(DE3) and C43(DE3) strains were used to assess the effect of point of IPTG-induction. These cultures were induced at OD_{600} of 0.7 and 1.3, using 0.5 mM IPTG at 25°C. All of the above were cultivated for 16 h using 0.5 mM IPTG for induction of the GPI-PLD3b construct. None of the above experiments yielded soluble protein (Figures 3.33-3.34).



Fig; 3.33: SDS-PAGE analysis of GPI-PLD3b expression from Rosetta2(DE3) strain.

10 ml Rosetta2(DE3)-GPI-PLD3b cultures were grown up at 37°C until OD_{600} = 0.7. Expression was induced using 0.5 mM IPTG and incubated at 18°C, 25°C and 37°C temperatures for 16 h with shaking (200 rpm). Samples of 1 ml were removed from each culture and cells were lysed using B-PER (Pierce). A centrifugation step separated soluble, supernatant (s) and insoluble pellet (p) fractions to assess protein expression and solubility by SDS-PAGE analysis.



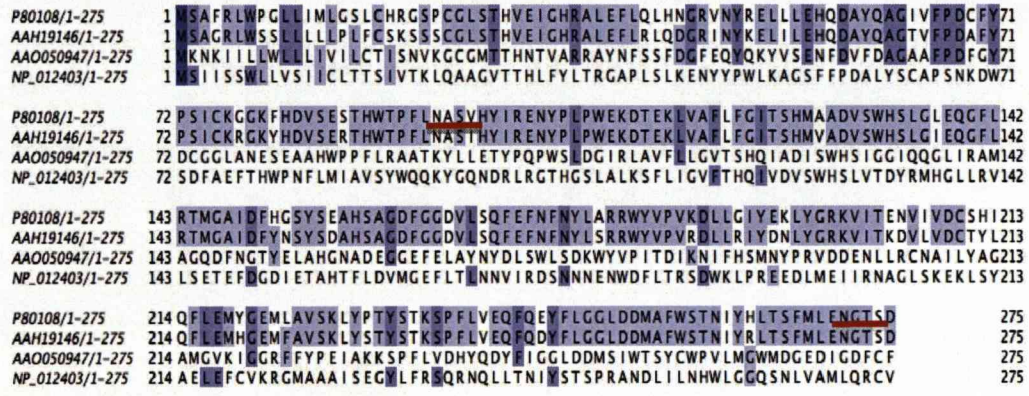
Fig; 3.34: SDS-PAGE analysis of GPI-PLD3b expression from C41(DE3) (left) C43(DE3) (right) *E. coli* strains.

10 ml C41(DE3)-GPI-PLD3b (left) C43(DE3)-GPI-PLD3b (right) cultures were grown up at 37°C until OD_{600} = 0.7 or 1.3. Expression was induced using 0.5 mM IPTG and incubated at 25°C for 16 h with shaking (200 rpm). Samples of 1 ml were removed from each culture and cells were lysed using B-PER (Pierce). A centrifugation step separated soluble, supernatant (s) and insoluble pellet (p) fractions to assess protein expression and solubility by SDS-PAGE analysis.

3.3: Heterologous expression of GPI-PLD4 using a eukaryotic expression system

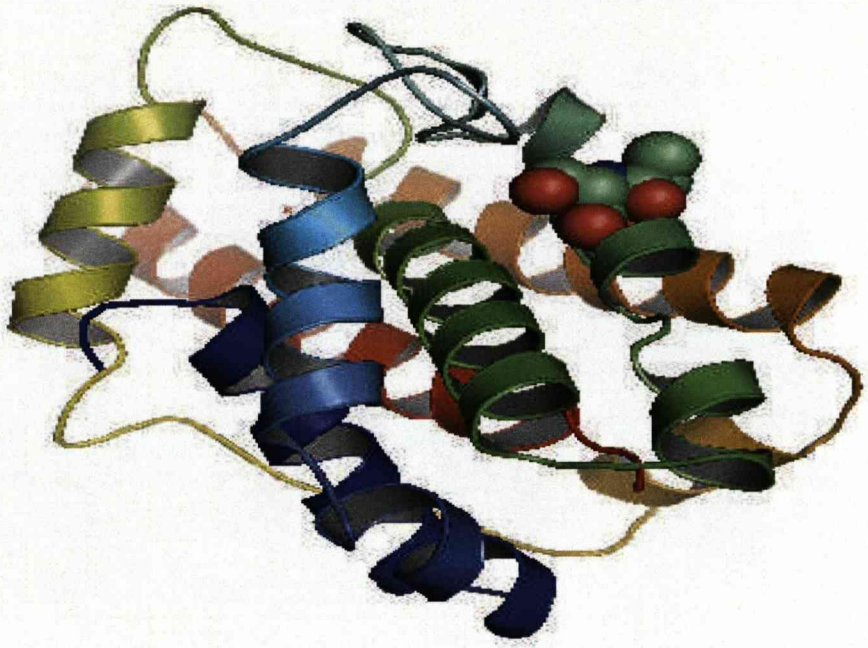
3.3.1: Predicting glycosylation sites in GPI-PLD homologues

N-linked glycosylation involves the addition of glycans to specific asparagine residues of the protein, which occur in an Asn-Xxx-Ser/Thr stretch (where Xxx is any amino acid excluding Pro). The presence of Pro in this motif renders Asn inaccessible thus inhibiting glycosylation, presumably due to conformational constraints. The occurrence of a Cys residue as opposed to a Ser/Thr in the sequence has even been shown to promote glycosylation. N-glycosylation sites were predicted using NetNGlyc (Gupta *et al.*, 2004), for the first 275 residues of the amino terminus of the GPI-PLD *M. musculus* clone. The likelihood of glycosylation site occurrence was based on the above explanation. Two sites were identified as potential glycosylation sites in this sequence (Figure 3.35). Sequence alignments (using muscle) of GPI-PLD homologues revealed that suggested glycosylation sites were conserved between human and mouse sequences (Figure 3.35). The computer model of the human GPI-PLD catalytic domain revealed that one of the sites predicted is in fact at the surface of the structure (Figure 3.36). Glycosylation at this location may prevent aggregation of hydrophobic regions on the external surface. Knowing the location of the predicted glycosylation site and the consequences faced in the absence of this modification, the next strategy was to express GPI-PLD4 using a *S. cerevisiae* expression host.



Fig; 3.35: Sequence alignment of human GPI-PLD1 homologues and conservation of glycosylation sites

Sequence alignment of human GPI-PLD1 (with accession number) with *S.cerevisiae, D. discoideum* and *M. musculus* homologues (each with Genpept accession number). Only the initial 275 amino acids of the N-terminal domain were aligned. Conserved glycosylation sites highlighted in red.



Fig; 3.36: PYMOL model of highlighting predicted catalytic site
Modelling of the N-terminal portion of human GPI-PLD (Swissprot accession number P80108) (Rigden, 2004). The predicted glycosylation site is shown as spheres.

3.3.2: Cloning of GPI-PLD-Sec for secretion in *Saccharomyces cerevisiae*

The construct, designated GPI-PLD-Sec, was designed to include only the first 275 amino acids. This is the suggested catalytic domain determined by limited proteolysis. The sequence was terminated at Cys, residue 275 (highlighted in red in Figure 3.37) by the addition of a stop (TAG) codon in lieu of the 276 Asp residue. A Kozak sequence for initiation, **(A/G)**NN*ATG***G**, is generally required by eukaryotic expression systems to ensure translation is initiated from the first ATG triplet. Therefore the serine triplet, tct, following the *GPI-PLD* start codon was replaced with one that coded for Gly, gga, to create the required Kozak sequence. This amino acid is not likely to affect the function of the signal peptide as it is not charged and is not likely to form a part of the hydrophobic trans-membrane region.

In this case the natural signal peptide sequence (*italics* in Figure 3.37) was included to promote secretion into the fermentation media. A homologous signal peptide sequence for secretion from yeast would provide the best guarantee for obtaining correctly folded GPI-PLD. However, many proteins have been expressed and secreted successfully in yeast using their natural signal sequences (Hitzeman *et al.*, 1990) so the natural signal sequence was used purely because this was the path of less resistance with respect to time.

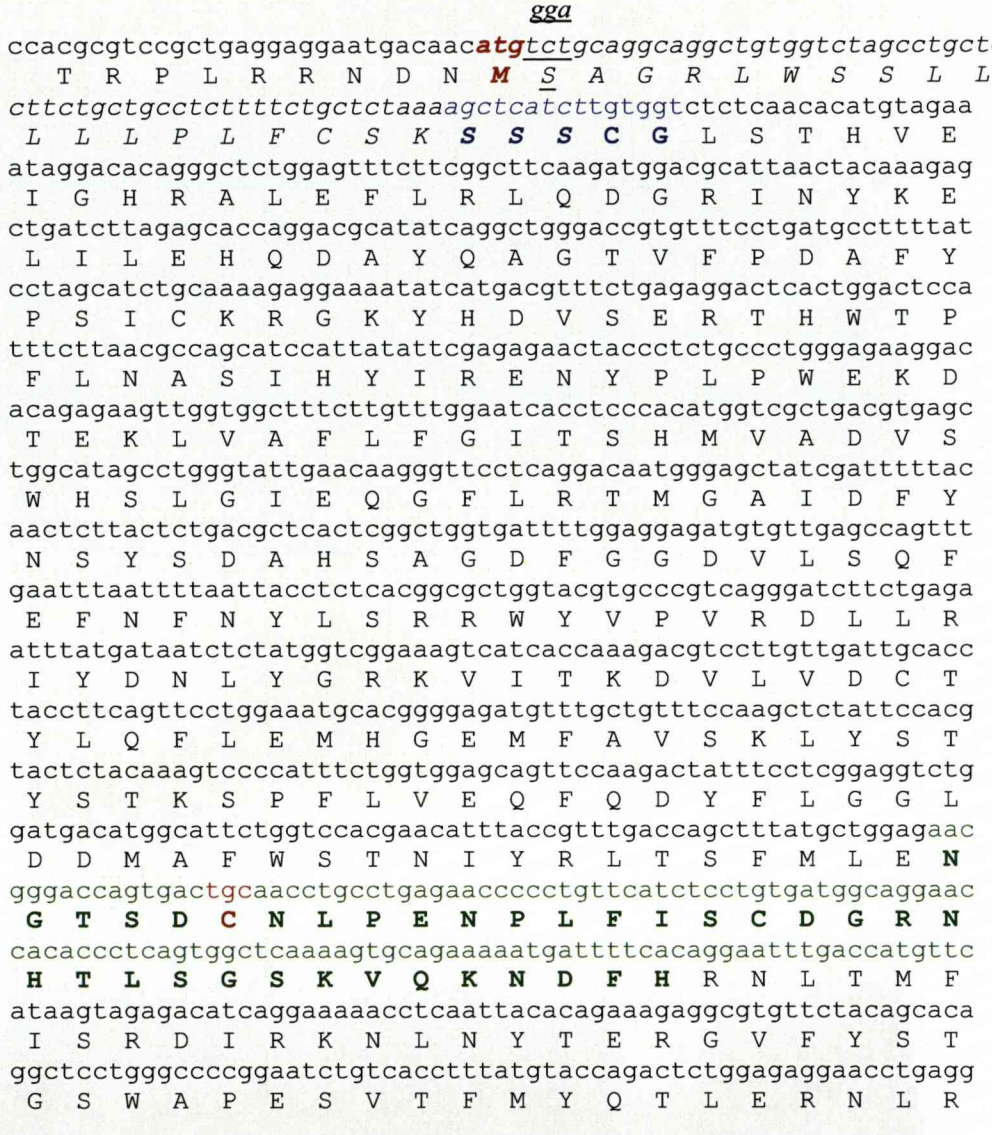
This construct was intended for expression in the *S. cerevisiae* strain, BY4741 (α), using a pYES2 plasmid expression vector (Invitrogen). This construct did not contain a His-tag for purification and identification. Instead, the GPI-PLD activity assay was used to detect the presence of the recombinant protein. Primers (Table 3.4) were designed so that the forward (coding) primer incorporated the initiation codon with the required Kozak sequence. This primer included a Bam HI restriction

site to aid ligation into the pYES2 vector. The sequence was transcribed from the initiation codon within the Kozak sequence in the correct frame, as there was no other start codon present in this vector. The reverse primer contained the end of the N-terminal domain up to the Cys residue at 275. A termination codon was incorporated, followed by an Xho I restriction site for ligation into the expression vector.

PRIMER NAME	OLIGONUCLEOTIDE PRIMER SEQUENCE
secGPI-PLD F	5'- aagcttGGATCCacaacATGgagcaggcaggctgtggtctagc -3'
secGPI-PLD R	5'- ttacagttCTCGAGCAGCTAgcagtcactggtcccgttctcc -3'

Table 3.4: Oligonucleotide primers for secreted expression of GPI-PLD from *S. cerevisiae*

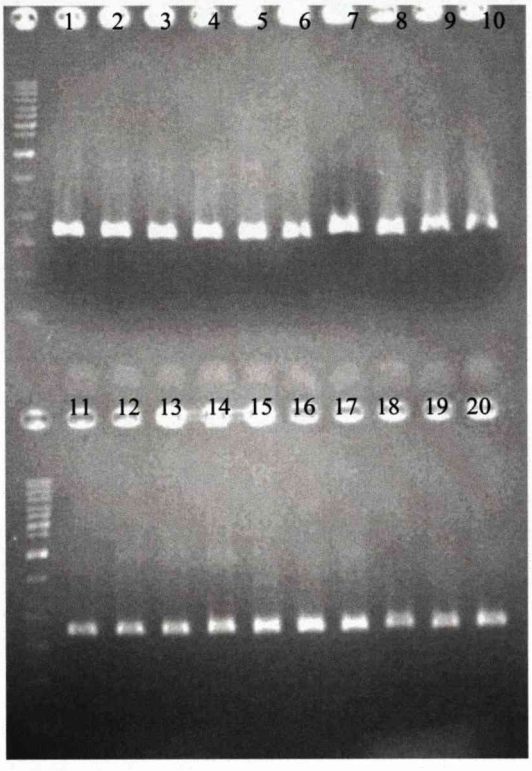
Primers used for amplification of the catalytic domain of GPI-PLD (Figure 3.37). Sequence highlighted in bold represent restriction sites (Bam HI in the forward secGPI-PLD F and Xho I in the reverse secGPI-PLD R primer). The start codon (ATG) is highlighted in bold, underlined, italics in the forward coding primer. This was followed by the codon, gga (bold) for serine to Gly substitution. The stop codon (bold, underlined, italics) was integrated into the reverse primer.



Fig; 3.37: Translation of GPI-PLD *M. musculus* clone – N terminal domain only

Figure shows the translation of the N-terminal domain GPI-PLD nucleotide sequence highlighting the linker region between N-terminal catalytic domain and C-terminal β-propeller domain (highlighted in green). A Cys at residue 275 (highlighted in red) will be followed with a stop/termination codon. The signal peptide sequence is shown in italics. The signal peptide is predicted (by signalP (Bendtsen *et al.*, 2004)) to be cleaved at SSS-CG (blue). The gga, Gly codon for the serine to Gly substitution is shown above the sequence in italics.

PCR amplification was achieved (as detailed in section 2.3.2) using Platinum Pfx DNA polymerase with its 3' proofreading ability. Once a band of the correct size (865 base pairs) was visualized using an ethidium bromide-stained agarose gel, the PCR reaction was scaled-up from 25 µl to 250 µl and the DNA purified using a PCR purification kit (Qiagen) (section 2.3.3 – 2.3.5). Digestion of both the amplified DNA and pYES2 expression plasmid with Bam HI and Xho I restriction enzymes followed (section 2.3.6). Digests were purified and successfully recovered for the ligation reaction using T4 DNA ligase (Invitrogen). The ligation reaction was used to transform XL1-Blue cells and a colony PCR identified that all 20 colonies that were chosen for screening of the insert, revealed PCR products of the correct size (Figure 3.38). Three of the 20 colonies that tested positive for insert incorporation were sequenced (Cogenics Ltd). Sequence alignments of the *M. musculus* GPI-PLD clone and colony, C1 were perfectly aligned with the exception of the intentional incorporated mutations for the insertion of the Kozak sequence and stop codon. This colony was cultivated for a plasmid preparation (section 2.3.11) prior to *S. cerevisiae,* BY4741(α) cell transformation by using the lithium acetate/single stranded carrier DNA/polyethylene glycol (LiAc/ss-DNA/PEG) protocol (Agatep *et al.*, 1998) as detailed in section 2.4.1.



Fig; 3.38: Identification of colonies containing GPI-PLD-Sec using colony PCR
Single colonies from pYES2-GPI-Sec- transformed XL1-Blue cells were selected and grown up overnight in LB (50 μg/ml ampicillin). A colony PCR was performed on 5 μl of each culture. 5 μl of each colony PCR was run on a 1% agarose gel. The gel shows colony PCR of colonies 1-20 all of which displayed a product of the correct size. 6 μl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250 was loaded in the first lane to monitor migration of DNA.

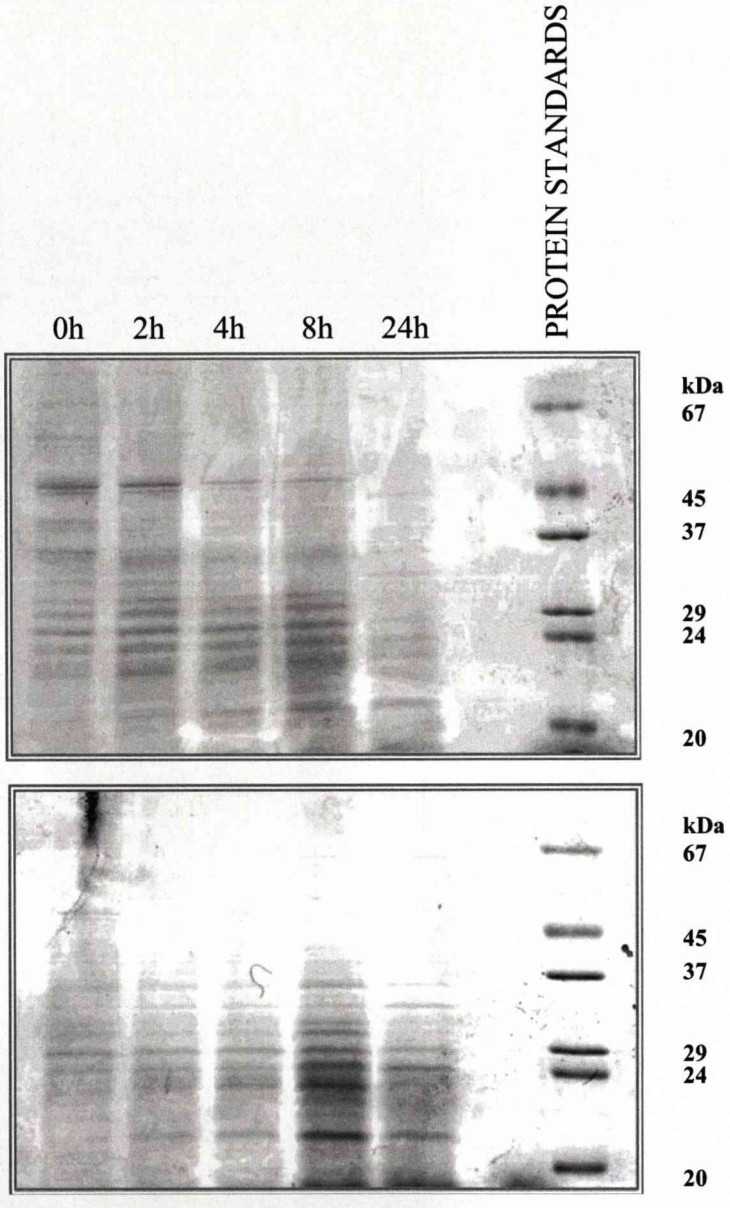
3.3.3: Expression of GPI-PLD-Sec from BY4741(α) cells

Once a transformant containing the pYES2 construct was isolated, expression of GPI-PLD-Sec was induced by the addition of galactose to promote transcription from the *GAL1* promoter. A pYES2 plasmid containing no insert served as a control in all expression studies using this plasmid and the BY4741(α) strain. Full details on expression protocol can be found in section 2.4.2. Expression of GPI-PLD-Sec was induced in 50 ml induction medium containing galactose at 30°C. Samples of 5 ml were removed from the culture at 0 h, 2 h, 4h, 8 h and at 24 h post induction. Cells were harvested by centrifugation and resuspended in the required volume of breaking buffer [sodium phosphate pH 7.5, protease inhibitor cocktail (1X)] to achieve an OD_{600} between 50-100. Cells were lysed using glass beads and vigorous vortexing. Initially, total cell lysates were assayed by SDS-PAGE to identify induction of GPI-PLD-Sec expression. A Bradford assay was used to ensure that there was equal loading of samples. Unfortunately, expression of the target protein could not be detected in the total cell lysate in any of the samples collected over 24 h (Figure 3.39).

The medium of each culture was also assayed for secreted GPI-PLD-Sec. The fermentation medium needed to be concentrated so that the protein could be detected. Two methods of concentration were performed. A methanol precipitation protocol (described in section 2.1.5.1) was carried out on 5 ml sample collected after 24 h. Pellets were re-suspended in a minimal volume (10-20 µl) of 2 X SDS-PAGE sample buffer. This protocol proved to be inadequate for concentration of proteins secreted into the fermentation medium as no protein could be detected by SDS-PAGE even after silver staining (data not shown). The second methodology was to

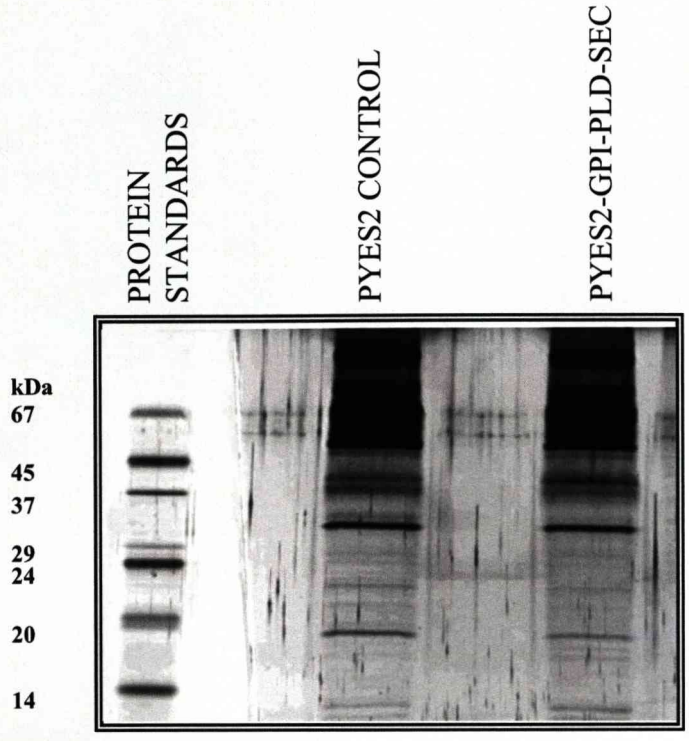
utilize Vivaspin MWCO 5 000 spin concentrators (GEHealthcare) to concentrate the media up to 40 X (described in section 2.1.5.2). Again SDS-PAGE analysis and detection by silver staining did not provide an indication of the presence of GPI-PLD in the culture medium (Figure 3.40).

The GPI-PLD activity assay was considered at this point to provide a more sensitive method of identifying the target protein in the expression medium. Samples of 5 ml were removed from the fermentation medium 24 h post induction and were concentrated to 40X using the Vivaspin concentrator columns to maximize the success of the activity assay. A bovine serum sample provided an additional control to the BY4741(α) cells harbouring the pYES2 plasmid containing no insert. GPI-PLD activity was detected in all samples (Table 3.5). As mentioned earlier, a GPI-PLD homologue has been identified in *S. cerevisiae*. It is more than likely that this homologue is responsible for the activity detected in this assay. With this knowledge coupled with the low levels of expression (if expressed at all) experienced using the *S. cerevisiae* host, the decision was made that time would be better spent providing strategies for tackling another bottleneck in crystallographic protein studies.



Fig; 3.39: SDS-PAGE analysis of GPI-PLD-Sec expression from BY4741(α) cell lysates over 24 h

BY4741(α)-pYES2 (top) and BY4741(α)-GPI-PLD-Sec (bottom) cells from an overnight starter culture were harvested and resuspended to an OD_{600} of 0.4 in 50 ml induction medium containing galactose. The induction culture was incubated at 30°C with shaking (200 rpm). Samples of 5 ml were removed from the culture at 0 h, 2 h, 4h, 8 h and at 24 h post induction. Cells were harvested by centrifugation and resuspended in the required volume of breaking buffer [sodium phosphate pH 7.5, protease inhibitor cocktail] to achieve an OD_{600} between 50-100. Cells were lysed using glass beads and vigorous vortexing. Total cell lysates were assayed by SDS-PAGE to identify induction of GPI-PLD-Sec expression.



Fig; 3.40: SDS-PAGE analysis of proteins secreted into the yeast culture medium.

BY4741(α)-pYES2 (control) and BY4741(α)-GPI-PLD-Sec cells from an overnight starter culture were harvested and resuspended to an OD_{600} of 0.4 in 50 ml induction medium containing galactose. The induction culture was incubated at 30°C with shaking (200 rpm). Samples of 5 ml were removed from the culture at 24 h post induction and cells were harvested by centrifugation. The supernatant/media was then concentrated 40-fold using concentrator Vivaspin columns before SDS-PAGE analysis. Proteins were detected using silver staining.

Table; 3.5: GPI-PLD activity assay for detection of GPI-PLD-sec in yeast fermentation media	
Fraction	TNB (nmoles $ml^{-1}min^{-1}$)
BY4741(α)-pYES2	1.46
BY4741(α)-GPI-PLD-Sec	1.49
Bovine	2.72

3.4: Discussion

Studies within this chapter were initially aimed at the structural and functional analysis of GPI-PLD. The elucidation of this structure would have benefitted the biopharmaceutical industry as it has been implicated in a number of disease states. The heterologous expression of GPI-PLD encountered a number of hurdles and highlighted a common bottleneck faced when expressing a target protein for 3D structural elucidation. Studies were aimed at tackling solubility issues and focusing on how recombinant protein expression can be optimised with the intention of achieving adequate concentrations for structural studies.

3.4.1: Assay and isolation of GPI-PLD

The protocols available for quantifying GPI-PLD activity use GPI-anchored proteins that exhibit enzymic activity, which can in turn be used as an indirect measure of GPI-anchor hydrolysis (Rhode *et al.*, 1995; Hoener and Brodbeck, 1990). For studies herein, the Bordier method (Bordier, 1981) was employed to achieve separation of the mf-AChE and s-AChE with the use of Triton X-114. Preliminary investigations, using bovine, porcine and horse sera, as a source of GPI-PLD, made it possible to determine optimum assay conditions (pH 6.5-7.0) (Figure 3.1).

The purification protocol for the isolation of GPI-PLD from bovine serum made use of three purification steps. Active fractions, purified by anion exchange chromatography using a QAE Sepharose column, were pooled and applied to a Superdex 200 column for separation by gel filtration. Then a final purification by affinity chromatography by use of wheat germ lectin agarose isolated the GPI-PLD activity to a single fraction (Table 3.1). SDS-PAGE analysis and silver staining

displayed a band of approximately 120 kDa in the most active fraction (fraction 2). This band can also be seen in fraction 3 through to fraction 5 with fraction 5 having a band of the lowst intensity. The band in the lane loaded with fraction 3 is almost as intense as fraction 2 but the GPI-PLD activity for this fraction was approximately 50%. This is likely to be the result of human error when loading. The putative GPI-PLD in this fraction could not be sequenced to confirm that the purified protein was in fact GPI-PLD as attempts at concentrating failed.

Levels of anchor degrading activity decreased with successive purification processes due to a loss of protein activity and/or a loss of the protein into other fractions collected from the column. Moreover, the inability to concentrate the purified protein meant that a pure, isolated form of native GPI-PLD would not be practical for use as a control in the GPI-PLD activity assay. Moreover, quantities were not sufficient for further analysis of the native protein. In particular, confirmation of predicted glycosylation sites by detection of glycans would have been extremely useful. Other methods of protein concentration could be investigated, but it was decided that un-purified, undiluted bovine serum would serve as the best control for our assay. This provided the most adequate control given that high levels of recombinant protein would be over-expressed. Purification of the native form reduced its concentration; so non-diluted bovine serum was likely to provide the most uniform control for the assay. Bovine serum served as a positive control to ensure that all reagents of the assay were active and functioning. False positives should not have had any effect on this assay, as recombinant proteins are generally expressed in much higher concentrations than the host native proteins.

The main concern in using the purified putative GPI-PLD was that its identity could not be confirmed. It may be possible that the purified protein was not the protein of interest at all and one of the other purified entities may in fact be responsible for GPI-anchor cleavage. It is possible that glycosylphosphatidylinositol-specific phospholipase C (GPI-PLC)) was responsible for GPI-anchor cleavage of membrane form, GPI-anchored AChE. Findings by Fox *et al* (1987) and Stieger *et al* (1991) reported the presence of GPI-anchor converting activity in rat liver and isolated rat hepatocytes respectively and concluded that anchor-converting activity was that of phospholipase C (PLC) type. Another report confirmed anchor-converting activity of GPI-PLD by the addition of phosphatase inhibitors (NaF at 50 mM and sodium orthovandate 50 mM) and analysis of products (phosphatidic acid and diacylglycerol) (Heller *et al.*, 1992). They identified two forms of anchor degrading activity by separation using anion exchange chromatography. Results herein also show evidence of additional forms of activity. The three active fractions obtained from the Superdex 200 column could also be due to the formation of multimers or complexes with other proteins causing elution of GPI-PLD in the earlier fractions. Low and Prasad {Low, 1988 #5} identified GPI-PLD from rabbit, rat and dog plasma to have an approximate molecular weight of 500 kDa when eluted from a Sephacryl S-300 gel filtration column. Also Huang *et al.* (1990) proposed that purified bovine GPI-PLD had a molecular weight of 100 kDa but had a molecular weight of 200 kDa after being applied to GFC on a Superose 12 HPLC column, suggesting that GPI-PLD may form complexes.

The main flaw in this assay appears to be that it does not distinguish the exact point of GPI-anchor cleavage, thus cannot determine if it is GPI-PLD or GPI-PLC that is

responsible for GPI-anchor cleavage. The conversion of erythrocyte membrane form, GPI-anchored-AChE (non-diluted bovine serum as a source) to soluble form (s-AChE) provided a measure for the anchor degrading activity of GPI-PLD and results have indicated that the assay is working efficiently. Activity is low in fractions where one would expect it to be (i.e. early/later fractions) and activity peaks within a narrow region. So, even though it was not possible to isolate the native GPI-PLD from bovine serum, the assay was tested and its limitations were revealed.

3.4.2: Heterologous expression of GPI-PLD from a prokaryotic host

GPI-PLD could not be isolated from its native cell type under the desired physiological conditions in amounts sufficient enough for the rigorous purification procedures required for structural studies, so investigations were directed at heterologous expression from an *E. coli* host. Results presented in this chapter highlight the many parameters that can be varied to influence expression levels and recombinant protein solubility.

Many specialized *E. coli* strains have been developed to tackle the common problems faced during recombinant protein production. BL21-Gold(DE3)pLysS, a *lon* and *ompT* protease-deficient strain, is known to improve the likelihood of intact, full-length protein expression (Studier and Moffatt, 1986) and is routinely used as an ideal starter strain. This strain and BL21 derivatives, C41(DE3) and C43(DE3), were used in initial expression and solubility studies for expression of GPI-PLD4a but all resulted in inclusion body formation and were of fairly low yields (Figure 3.13 -3.14). IMAC purification and MALDI-MS gave the indication that a small quantity of soluble GPI-PLD4a may be produced when expressed in C41(DE3)

(Figures 3.15-3.16). However, results suggested that the target protein was co-purified with a host protein of a similar size and that this was the major protein in the band. The target protein could be isolated by 2D gel electrophoresis and each protein sequenced to confirm their identity, but the decision was made to try out alternative strains.

GPI-PLD4a expression was then directed from an Origami2(DE3) host containing the trxB/gor mutation recommended only for the expression of proteins that require disulfide bond formation for proper folding. This strain was used in the event that disulfide bond formation was required between the Cys in the protein sequence for correct folding. It was thought that a small quantity may have been expressed as soluble protein at 25°C for 2-4 h (Figure 3.17). IMAC revealed that this protein also appeared to be co-purified with a host protein as experienced in the C41(DE3) strain (Figure 3.18). Unfortunately, GPI-anchor degrading activity was not found in any eluate from this purification (Figure 3.19). This may suggest that GPI-PLD4a was not expressed at all as a soluble protein and all protein was present as inclusion bodies, or, that it was expressed as a soluble protein but is non-functional. In addition to this, there is a possiblitiy of inhibition of the activity assay caused by the presence of imidazole in the eluate.

Perhaps conditions in this particular assay are inadequate for maintaining the activity of the catalytic domain alone. The C-terminal domain has been reported to have a regulatory role and it has also been suggested that this domain also provides stabilization of GPI-PLD (Stradelmann *et al.*, 1997), which may explain the absence of activity in the assay or at worst, the protein may not be correctly folded. Moreover the expressed protein in all of the strains tried up to this point appeared to

be of a smaller size than expected (determined by SDS-PAGE). Truncations of the target may be attributable to stalled expression, particularly where there are two consecutive rare codons in two situations in the target sequence. It is likely that this would result in an unfolded C-terminus, which in turn would cause aggregation of the recombinant protein. The Rosetta2(DE3) strain was used to enhance the expression of this mammalian protein as it contains a pRARE2 plasmid to provide a supply of tRNAs for the codons AUA, AGG, AGA, CUA, CCC, GGA and CGG. The IPTG-induced protein migrated at 25 kDa as experienced with expression of GPI-PLD4a in all other *E. coli* strains tried to date (Figure 3.20). This suggests that expression from each *E.coli* host strain was of a full-length construct, but perhaps the protein's properties hamper its migration during SDS-PAGE.

It is reasonable to assume that a larger protein will require more complex folding mechanisms, which render them more susceptible to aggregation at higher transcriptional rates arising from increased temperatures or higher IPTG concentrations. There are many single-case studies in which the lowering of cultivation temperatures shows a marked increase in protein solubility. This effect was not observed for expression of GPI-PLD at temperatures of 15°C to 37°C. Interestingly though, a reduced temperature at the point of induction had varying effects on protein expression in each of the host strains studied. Variations in temperature appeared to have no effect on either expression nor solubility of GPI-PLD4a when using Rosetta2(DE3) as the expression strain. Protein bands representing GPI-PLD4a expression were of equal intensity in samples of cultures induced at 18°C, 25°C and 37°C (Figure 3.20). When using BL21(DE3) derivatives, C41(DE3) and C43(DE3) as the strain to host expression of GPI-PLD4a there was a clear correlation between decreased temperature and reduced quantities of the

induced recombinant protein (Figure 3.14). Expression of GPI-PLD4a from the Origami2(DE3) host was the most affected by temperature. When incubated at 37°C for 2 h GPI-PLD4a is expressed as aggregated inclusion bodies (Figure 3.17). Yet, it appears that there is no expression of the target protein at 37°C, 16 h post induction (Figure 3.17). This may suggest that at this high temperature such high levels of the target protein beget an increase in degradation by the host. No expression is apparent when induced for 16 h at 18°C (Figure 3.17). Expression at 25°C for 16 h yields the highest levels of the induced protein (Figure 3.17). In the case of expression directed from BL21(DE3)pLysS cells, a temperature reduction to 22°C was enough to impede expression of GPI-PLD4a and bands could not be detected by SDS-PAGE after a period of 3 h (Figure 3.13). It is possible, however, that the amount of induced GPI-PLD4a, at this point, is trivial since the rate of protein synthesis had been reduced and much longer incubations were necessary.

Results herein also suggest that expression of GPI-PLD in all of the strains tested could not be regulated by IPTG control. Using lower concentrations of IPTG (0.05 mM – 0.2 mM), translation rates were reduced presenting the GPI-PLD protein with an increased chance of folding into a native state prior to aggregating with folding intermediates. Despite regulation of IPTG-induction from host strains, Origami2(DE3), Rosetta2(DE3), no enhancement in solubility was experienced following protein expression using lower concentrations of the inducer (Figures 3.21-3.22(A)). Concentrations of the inducer as low as 0.05 mM failed to have any effect on the solubility of the target when expressed in Rosetta2(DE3) (Figure 3.22(A)). Even when regulating the point at which expression was induced by addition of IPTG at different cell densities (OD_{600} of 0.7 or 1.3) there was no

improvement in protein solubility nor did it increase target protein levels (Figures; 3.22(A), 3.3.23 and 3.34). In some instances, optimal expression requires the addition of supplementary salts or alternative medium. It is thought that GPI-PLD requires zinc for catalytic activity but the addition of Zn^{2+} to culture media did not ameliorate inclusion body formation (Figure 3.26). M9 minimal medium was used for incubation of IPTG-induced cultures in order to reduce expression levels thereby enhancing protein solubility but yields were very low and, again, IB formation was evident (Figure 3.24). Induction of protein expression under osmotic stress by the addition of sorbitol and betaine to Rosetta2(DE3)-GPI-PLD4a (Figure 3.24) and Origami2(DE3)-GPI-PLD4a (Figure 3.25) cultures revealed no benefit in increasing the yield of active soluble protein or a reduction in inclusion body formation.

To address solubility issues of the target protein, alternative affinity tags were also considered. The simple action of exchanging a C-terminal 6X his tag to one that is included at the N-terminus can make the difference between expression of an inactive, insoluble protein and a soluble protein with moderate levels of activity (Walker *et al.*, 2001). This was not the case for expression of GPI-PLD4b, which remained insoluble when expressed from the prokaryotic host.

The employment of the large, soluble protein fusion/partner, glutathione S-transferase (GST) (Smith and Johnson, 1988) was used to improve solubility and folding of the protein in question. This fusion has the added advantage as the 3D structure of *Schistosoma japonicum* GST has already been determined (McTigue *et al.*, 1995). This fusion was employed so that its structure could be used to provide a search model to solve the crystallographic phasing problem by molecular replacement (MR) methods. Furthermore, the GST-fusion can, more often than not,

be crystallised using similar conditions used for the isolated GST. Results revealed that the GST fusion had no benefit to expression of soluble GPI-PLD-GST from Rosetta2(DE3) and that neither regulation of temperature nor of IPTG concentration had any effect on solubility (Figures 3.27(A) and (B)).

Expression and solubility studies using the highly expressed and very soluble GB1 fusion may have overcome the solubility issues for expression of GPI-PLD4. SDS-PAGE analysis revealed a protein band that suggests that a small quantity of the target was soluble when protein expression was induced in Rosetta2(DE3) cells at a low temperature (18°C) and/or a low IPTG concentration (0.01 mM) (Figure 3.28). Protein sequencing confirmed that this was the GB1-His6-GPI-PLD fusion. However, the soluble protein could not be isolated using either immobilised metal (Ni^{2+}) or IgG affinity chromatography methods for purification (Figures 3.29-3.30). The detergent, Triton X-100 (0.1%) was added to wash and elution buffers to substitute for the loss of the mild detergent (in the B-PER lysis reagent) during the purification process. Figure 3.31 showed that this had no effect on maintaining the solubility of the target protein during purification. IgG affinity chromatography was use in the event that the hexa-histidine region was concealed by the target protein. Harsh conditions are required for elution of the target protein (pH 3.5), which was very likely to result in denaturation of the GPI-PLD-GB1-His6 fusion. The binding capacity of the IgG column was also to be determined but after closer scrutinization of SDS-PAGE gels (Figures 3.29 – 3.31) it was decided that this might have been the result of human error. The insoluble pellet fraction is much softer after B-PER lysis so it is plausible that contamination from this fraction was present in the supernatant. No further studies were carried out using this construct.

The addition or subtraction of as few as one amino acid residue at the C-terminal domain can be the critical factor in protein structural studies. The theoretical structure for *M. musculus* GPI-PLD, as determined by PSIPRED, protein structure prediction server, suggested that the linker region consisted of residues 271 through to 305 of the GPI-PLD sequence. The length of the target protein was increased to ensure that the full catalytic domain was being expressed. An N-terminal his-tagged-GPI-PLD of increased length, designated GPI-PLD3b (Table 3.3.1) was expressed, but unfortunately could not be optimised to yield soluble protein (Figures 3.33-3.34).

On this level, optimisations can, and did, become very time consuming and ineffective. Even after all of the aforementioned optimisations, there are still many more conditions to be investigated. For example, pH of media, co-expression with chaperones, alternative fusion proteins and combinations of each, plus all of the variables discussed in this chapter. In retrospect, a high throughput system for expression and solubility testing in *E. coli* would have saved time and covered more or all of the parameters for optimisation of expression. Had this system been available in our lab we would have had a clearer indication as to whether expression of a GPI-PLD truncation was even possible in a prokaryotic system. As a high-throughput system was not available for use it was decided that time would be better spend attempting a final expression study using a lower eukaryotic system.

3.4.3: *Saccharomyces cerevisiae* as an expression host

The computer model of the human GPI-PLD catalytic domain revealed a site for glycosylation at the surface of the structure (Figure 3.36). This site is conserved between the human and mouse sequences (Figure 3.35). Glycosylation at this

location may maintain protein solubility and prevent aggregation of hydrophobic regions on the external surface. In addition, the absence of internal glycosylation can result in the incorrect folding during translation that in turn leads to protein aggregation. Moreover, enzyme inactivity can occur, as cellular interactions that rely on the contacts made by means of the carbohydrate moieties of the glycoprotein are not available. This may well explain why GPI-anchor degrading activity could not be detected in semi-purified fractions of the suspected soluble recombinant GPI-PLD. The *S. cerevisiae* host provided an expression system capable of glycosylation. Moreover, it can allow for secreted expression of the recombinant protein and provide a folding environment more like the native conditions than what the prokaryotic host provided. Regrettably, GPI-PLD-Sec could not be detected in the culture medium (Figure 3.40) or in total cell lysates (Figure 3.39) when using the pYES2 plasmid and BY4741(α) strain combination. The very low levels of expression (if expressed at all) experienced maybe enhanced with the replacement of the mammalian signal peptide (*italics* in figure 3.37) with homologous α-factor signal sequence that has been known to increase expression (Hitzeman *et al*, 1990). GPI-PLD activity was detected in the bovine control sample, the control induction culture (BY4741(α) strain harbouring the PYES2 containing no insert) and the target protein induction culture (Table 3.5). As mentioned earlier, a GPI-PLD homologue has been identified in *S. cerevisiae*. It is more than likely that this homologue is responsible for the activity detected in this assay.

To progress further, the recombinant protein must be expressed in a eukaryotic host with a purification tag for detection and purification, if soluble and over-expressed. A hexa-his tag could be incorporated using primers containing a his-rich region. In addition, the homologous α-factor signal peptide sequence for secretion from yeast

would promote secretion into the fermentation media providing the best guarantee for obtaining correctly folded GPI-PLD (Hitzeman *et a.l*, 1990).

The solubility of the GPI-PLD N-terminal domain was estimated using solubility predictor (Smialowski *et al.*, 2007), which indicated that it is unlikely that this target would be expressed as a soluble protein in *E.coli*. Interestingly, a crystallisation predictor (Smialowski *et al.*, 2006) suggested that the same GPI-PLD domain would have a fair chance at crystallisation. Consequently, studies were directed at overcoming an alternative bottleneck in protein crystallographic studies.

Chapter 4: RESULTS CHAPTER; Using Lanthanide Binding Tags For Solving The Phasing Problem

4.0: Development of a lanthanide binding tag

4.1: Introduction

When solving three dimensional protein structures by X-ray crystallography many difficulties can be encountered that can hinder or even prevent a successful outcome. Not only is there the requirement for highly purified protein and good quality crystals but the crystallography experiment itself gives rise to additional hurdles that must be overcome (Rupp and Kantardjieff, 2008). X-ray diffraction data and diffraction intensities are measured directly. However, the information on phases is lost and, unfortunately, this information is paramount to the elucidation of the protein structure. At present, the methods available for solving the 'phasing problem' have their drawbacks including size limitations (Turkenburg and Dodson, 1996), bias in the case of molecular replacement and destruction/distortion of the crystal structure in multiple isomorphous replacement (MIR) (Carvin *et al.*, 2001; Smyth and Martin, 2000; Dauter, 2002). For MIR, crystals are soaked in millimolar solutions of a heavy atom salt, which can often result in loss of isomorphism, so this approach can also be expensive of crystals (Rupp and Kantardjieff, 2008). The single/multi-wavelength anomalous diffraction (SAD/MAD) method can often become problematic as a result of the unpredictable behaviour of bound metals (Dauter, 2002). Although incorporation of selenomethionine does not create such problems, the protein sequence is required to consist of a sufficient number of adequately spaced methionine residues in its sequence in order for this approach to be successful (Hendrickson *et al.*, 1990).

Investigations in this chapter were aimed at the development of a metal-binding tag so that a metal can be bound to a protein in a known position (regarding the chemistry) to facilitate the solving of the phasing problem by SAD/MAD or MIR. Two metal-binding tags were designed based on the conserved DxDxDG motif common to many calcium-binding proteins (CaBPs). These proteins share this common core but differ in the number of residues between the motif and a later Asp or Glu residue; with two being the most common separation within the different classes of CaBPs (Rigden *et al.*, 2003a; Rigden *et al.*, 2003b; Rigden, 2004). The binding of calcium by this motif is thought to be a particularly energetically favourable interaction, thus it is likely that a similar isolated region of polypeptide should be able to maintain the metal interaction. As mentioned earlier (section 1.3.2) it is well known that calcium binding sites are capable of accommodating other metal ions too, particularly those of the lanthanide series. A hypothesis was made that each of the designed metal-binding tags would be capable of binding the lanthanide metals. The binding abilities of each metal-binding tag were assessed using gel filtration chromatography (GFC) and isothermal titration calorimetry (ITC). If this hypothesis is correct, such metal-binding tags can be used for heterologous expression of proteins providing a new, general approach to phasing of protein crystals using either SAD/MAD or MIR approaches.

Although any target sequence could be cloned into the modified expression plasmid (pET28a) for expression with the C-terminal lanthanide binding tag LBT, a stable, highly soluble protein is obviously preferred. Asymmetric diadenosine-5',5'''-P1,P4-tetraphosphate (Ap_4A) Hydrolase was employed as the model for investigations using the metal-binding LBT. Ap_4A hydrolase is a member of the nudix (nucleoside diphosphate linked to x) superfamily (McLennan, 2006). This enzyme maintains

physiological homeostasis by cleaving the metabolite diadenosine tetraphosphate (Ap4A back into ATP and AMP) (McLennan, 2006). The structure of human Ap_4A hydrolase and the *C. elegans* homologue have previously been solved by NMR (Swarbrick *et al.*, 2005) and by X-ray crystallography (Bailey *et al.*, 2002a; Bailey *et al.*, 2002b; Abdelghany *et al.*, 2001) respectively. The C-terminus of the candidate protein is exposed so it is likely that the tag will be available for metal binding.

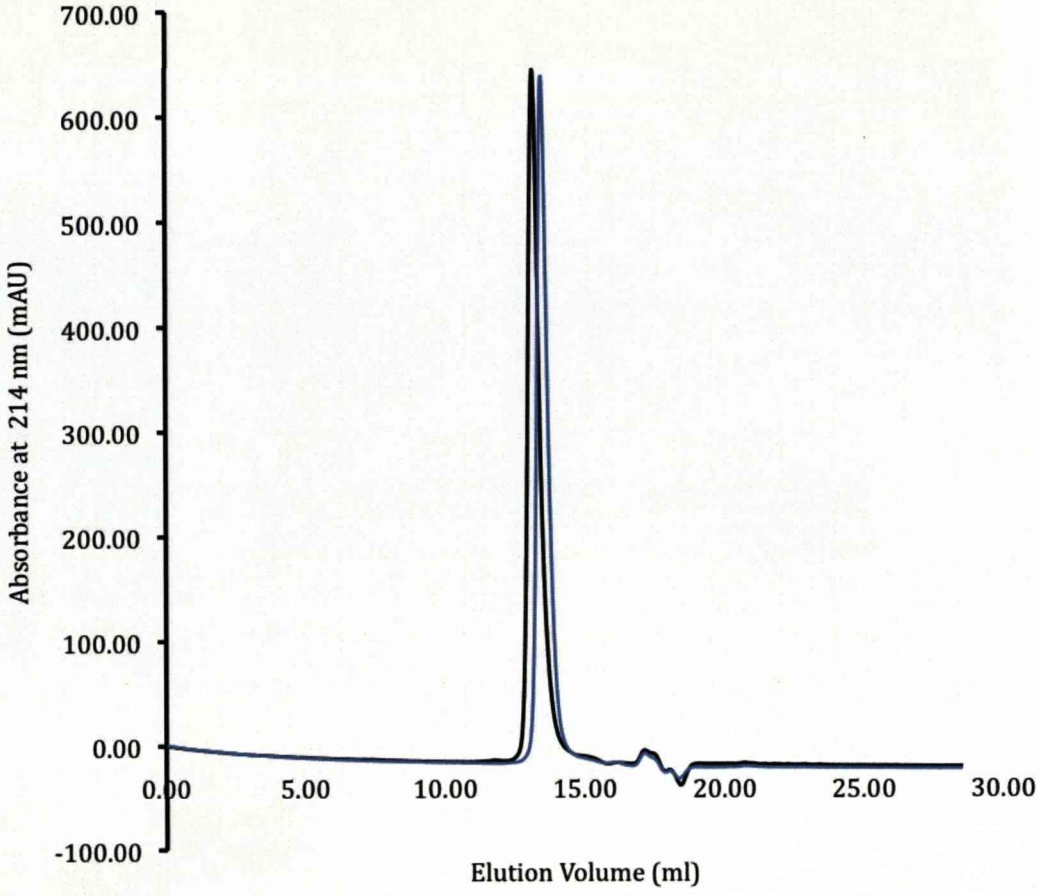
4.2: Metal-binding studies on isolated polypeptide lanthanide binding tag(s)

Two metal-binding tags were developed and synthesised (Sigma-Genosys Ltd). LBTA with the sequence (N acetyl)DNDKDGHSD was the first of the two tags to be synthesised. This tag proved to be extremely soluble during crystallisation attempts so a second, more hydrophobic tag was designed, namely LBTB, comprising of the sequence (N–acetyl)DADKDGWAD.

First, lanthanides, lanthanum (La^{3+}), terbium (Tb^{3+}), ytterbium (Yb^{3+}) and gadolinium (Gd^{3+}) were used in metal binding studies to assess the metal binding ability of isolated LBTA and LBTB using gel filtration chromatography (GFC). Binding of a lanthanide chloride to the peptide would be expected to cause an increase molecular weight consequently causing earlier elution from the column, shifting the peak to the left. This provided a means of detecting the metal binding ability of the isolated LBT.

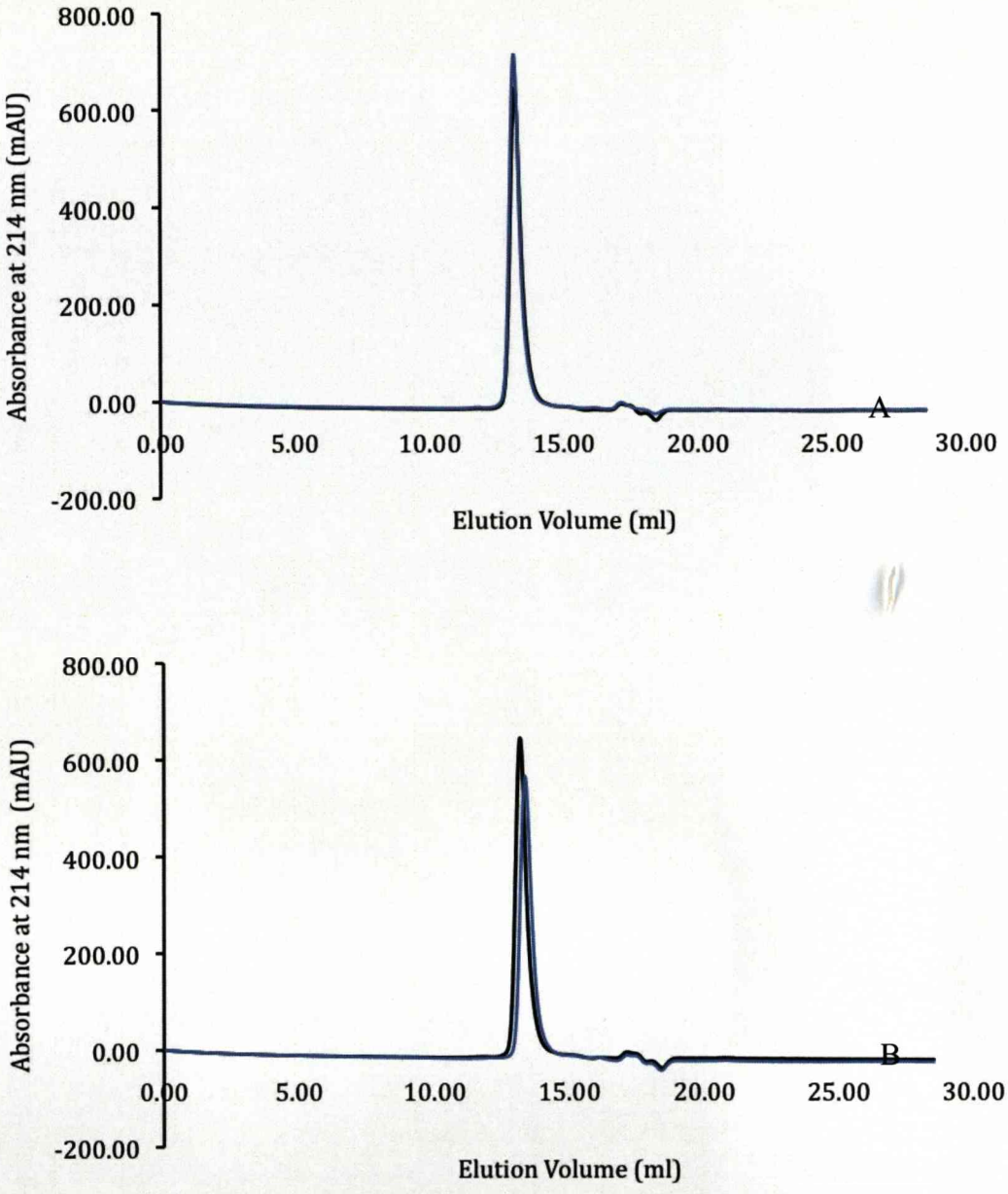
The LBTA peptide (2 mg/ml) was incubated for 1 h at room temperature with an excess of each lanthanide chloride hexahydrate (1 M) (50 µl peptide: 5 µl $LnCl_3 \cdot 6H_2O$) as detailed in section 2.8. Incubations over 16 h were also carried out for detection of site occupancy by lanthanum and ytterbium. An initial run was

conducted with the loading of 20 µl of the LBTA peptide in the absence of any metal in the incubation to serve as a control (Figure 4.1). This provided a trace that would represent the unoccupied LBTA. It appears that the isolated peptide was eluted at approximately 13-14 ml in the absence of hard metal. Following this run, a series of incubations were set up using the aforementioned lanthanides (and $CaCl_2$) and 20 µl of each peptide-metal chloride incubation was loaded onto a 24 ml Superdex Peptide column. Data obtained from this set of metal-binding studies of the isolate LBTA peptide indicate that calcium did not bind to the tag (Figure 4.1). This was also the case for lanthanum despite an incubation of 16 h (Figure 4.2). Interestingly, the binding of the lanthanides terbium, gadolinium and ytterbium appeared to result in a conformational change of the LBT peptide indicated by the later elution that was observed suggesting an overall decrease in size (Figures 4.3-4.4). The binding of these lanthanides may result in a contraction of the overall structure, which is likely to be linear in its unoccupied form due to the charges carried by its constituent amino acids. This assumption is based on the presence of a second peak to the right of the peak representing the unoccupied tag, which suggests a smaller, more compact peptide despite the addition of the bound Ln^{3+}. Ytterbium bound with the greatest affinity as it, of all the studies using LBTA, had the greatest area under the curve representing the occupied peptide (i.e. the second peak) (Figure 4.4). The height of the first peak (i.e. earliest eluate) in each of the metal binding elution profiles was not reduced by an increase in the incubation time (16 h) of the LBTA peptide with ytterbium meaning that no further binding took place.

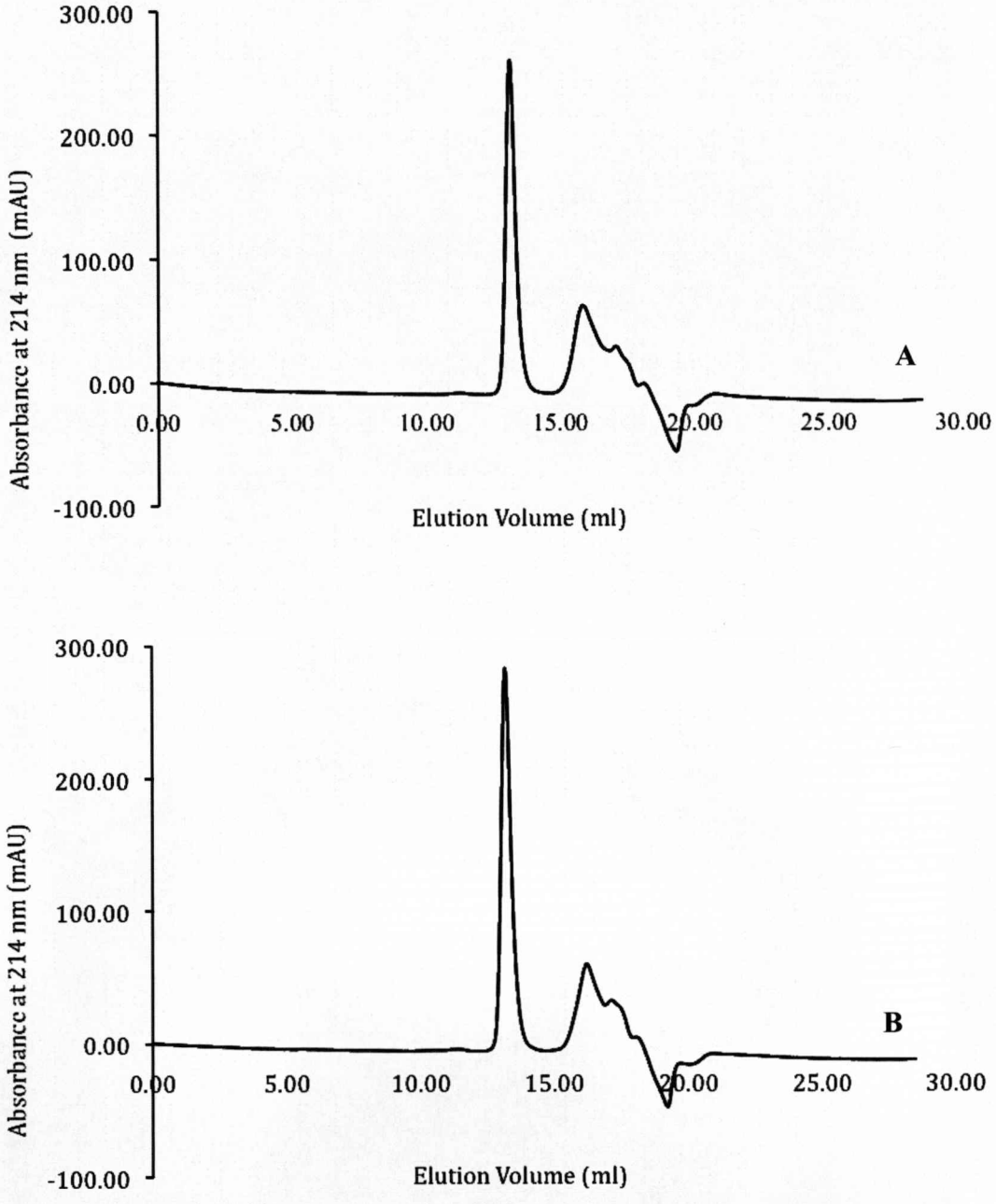


Fig; 4.1: Gel filtration chromatography of LBTA peptide – DNDKDGHSD to determine calcium binding

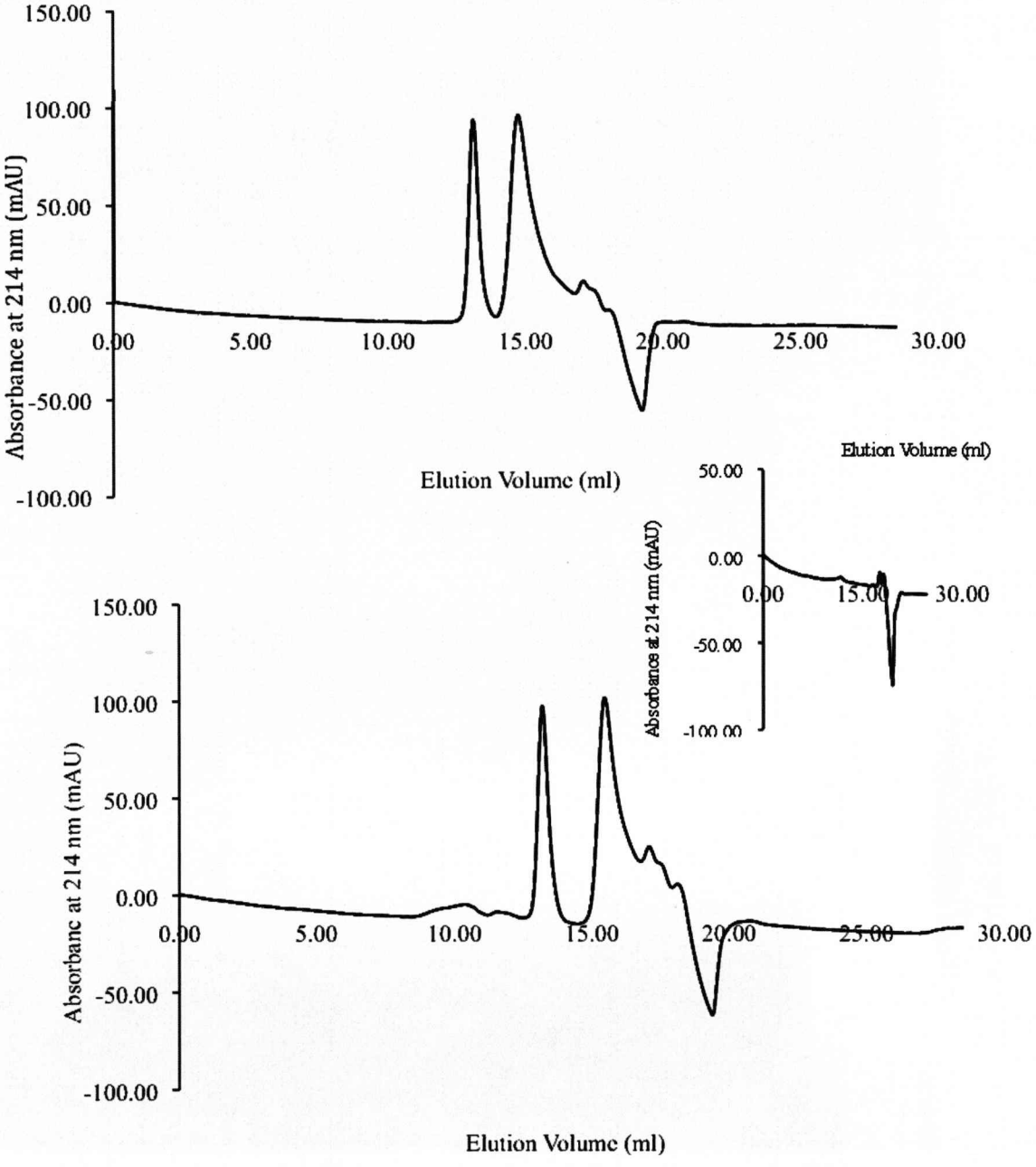
Incubations were set up; one with 45 µl LBT peptide plus 5 µl TBS (Black) and the other with 45 µl LBT peptide plus 5 µl Ca^{2+} (Blue) at room temperature for 1 h. After incubation the peptide/lanthanide mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of the peptide-buffer or peptide-metal chloride incubation and run at a flow rate of 0.5 ml/min with TBS (20 mM Tris-Cl, 0.1 M NaCl, pH 7.5). Trace is representative of typical results for incubation of this peptide with Ca^{2+}.



Fig; 4.2: Determination of binding of lanthanum to LTBA using GFC.
Incubations were set up with 45 µl LBT peptide plus 5 µl La^{3+} at room temperature for 1 h (blue) *(A)* and 16 h (blue) *(B)*. The control incubation curve (45 µl LBTA peptide plus 5 µl TBS) is shown (black) for comparison. After incubation each peptide/metal mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each peptide-metal chloride incubation and run at a flow rate of 0.5 ml/min with TBS. Trace is representative of typical results for incubation of this peptide with La^{3+}.



Fig; 4.3: Binding of terbium and gadolinium to LBTA
Incubations were set up with 45 µl LBT peptide plus 5 µl Tb^{3+} *(A)* and Gd^{3+} *(B)* at room temperature for 1 h (B). After incubation each peptide/metal mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each peptide-metal chloride incubation and run at a flow rate of 0.5 ml/min with TBS (20 mM Tris-Cl, 0.1 M NaCl, pH 7.5). Trace is representative of typical results for incubation of this peptide with Tb^{3+} and Gd^{3+}.



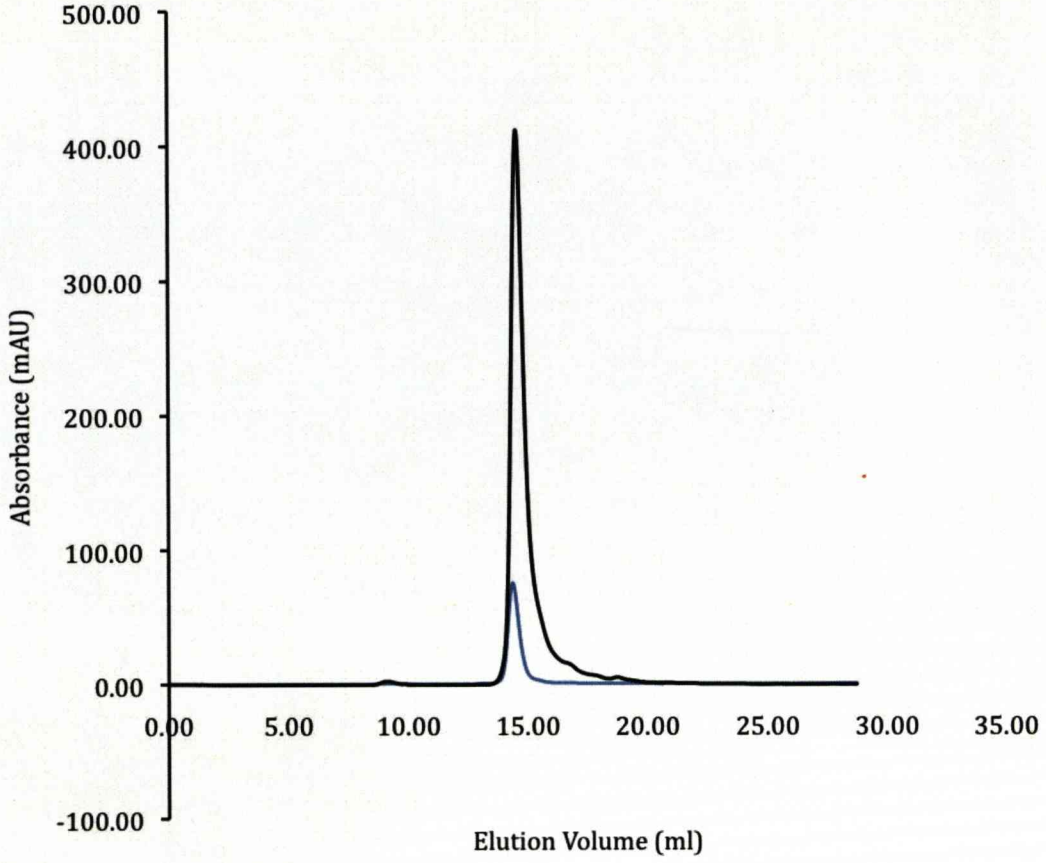
Fig; 4.4: Binding of ytterbium to LTB peptide A

Incubations were set up with 45 µl LBT peptide plus 5 µl Yb^{3+}(1 M) for 1 h *(Top)* and for 16 h *(Bottom)* A Yb^{3+} control incubation was set up with 45 µl TBS plus 5 µl Yb^{3+} to serve as a control for Yb^{3+} alone (*Inset*). After incubation each peptide/metal mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each peptide-metal chloride incubation and run at a flow rate of 0.5 ml/min with TBS (20 mM Tris-Cl, 0.1 M NaCl, pH 7.5). Trace is representative of typical results for incubation of this peptide with Yb^{3+}.

For isolated LBTB, GFC studies were initially executed as described for LBTA using TBS buffer. The preliminary run of unoccupied peptide suggested that there was a high degree of precipitation illustrated by the decrease in peak absorbances (Figure 4.5). Moreover, the addition of metals resulted in total precipitation of the peptide and no 214 nm-absorbing material could be observed during GFC. An alternative buffer had to be considered. 100 mM Na-HEPES, 150 mM KCl, (pH 7.0) was employed for all GFC studies using LBTB based on its stability in this buffer during isothermal titration calorimetry (ITC), which will be discussed later in this section. Peptide elution could not be recorded using an absorbance of 214 nm (as used for detection of LBTA in SEC studies) because the Na-HEPES buffer caused interference at this wavelength. Fortunately, the constituent tryptophan in the LBTB peptide allowed for detection using a wavelength of 280 nm.

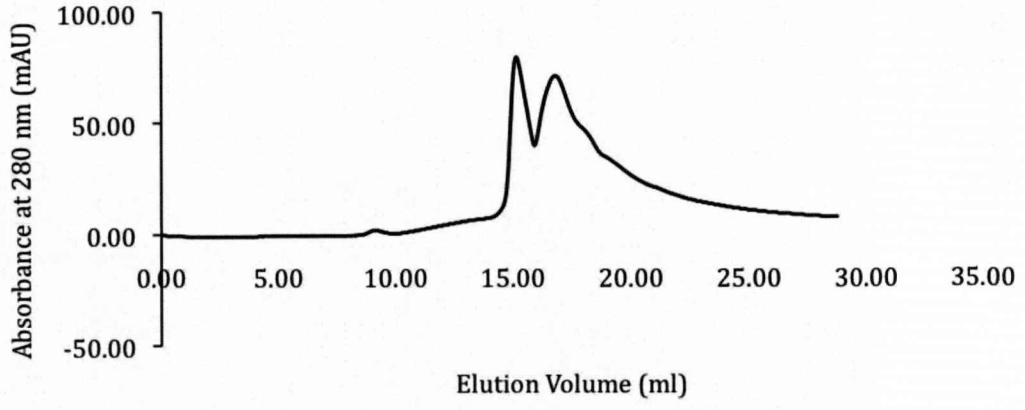
An additional preliminary incubation was set up with the LBTB peptide and Na-HEPES buffer and loaded onto the Superdex Peptide column (Figures 4.5). All further LBTB peptide plus metal incubations followed by GFC were carried out as described for LBTA, with the exception of the buffer, unless stated otherwise. The data obtained from metal binding studies for the isolated LBTB peptide suggest, in contrast to LBTA, that this tag was capable of binding both calcium and lanthanum (Figures 4.6 - 4.7). Incubation with both metals gives rise to a trace shift to a later elution. Moreover, the peak representative of the unoccupied tag is no longer evident in the case of the La^{3+} (1M) incubation (Figure 4.8). All peak absorbances were lower than that obtained from the unoccupied tag, which suggests a high degree of peptide precipitation. Surprisingly, the readout continues to suggest that the occupied tag is gradually being eluted from the column and the absorbance does not return to the baseline even after 30 ml or 75 ml for Ca^{2+} (50 μM) and La^{3+} (1 M),

respectively (Figures 4.6 - 4.7). Binding of Tb^{3+}, Gd^{3+} and Yb^{3+} could not be determined by GFC as all resulted in precipitation, so no curves could be displayed for each. This does not necessarily suggest that these lanthanides could not be bound, although there is this possibility. Perhaps the tag is capable of binding the metal, but has further interactions with other metals that may or may not be associated with the tag. This may also create aggregation causing the peptide to become fixed in the column, which may explain figures 4.6 - 4.7. It is also possible that the hydrophobic tryptophan is exposed resulting in peptide aggregation, which would provide an explanation for the extent of precipitation experienced in these studies.

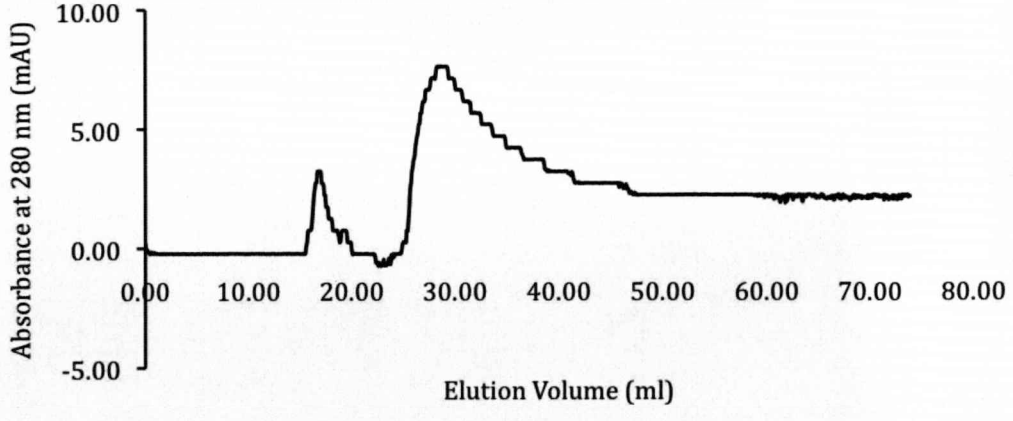


Fig; 4.5: Gel filtration chromatography of unoccupied LBTB using TBS buffer or Na-HEPES buffer.

Incubations were set up with 45 µl LBT peptide (dissolved in either TBS or HEPES) plus 5 µl of TBS buffer *(Blue)* or HEPES buffer *(Black)* for 1 h. After incubation each peptide/metal mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each incubation and run at a flow rate of 0.5 ml/min with TBS (20 mM Tris-Cl, 0.1 M NaCl, pH 7.5) or Na-HEPES 100 mM, KCl 150 mM, (pH 7.0) and absorbances at 214 nm (when using TBS) or 280 nm (when using Na-HEPES). Traces are representative of typical results for incubation of this peptide with each buffer.



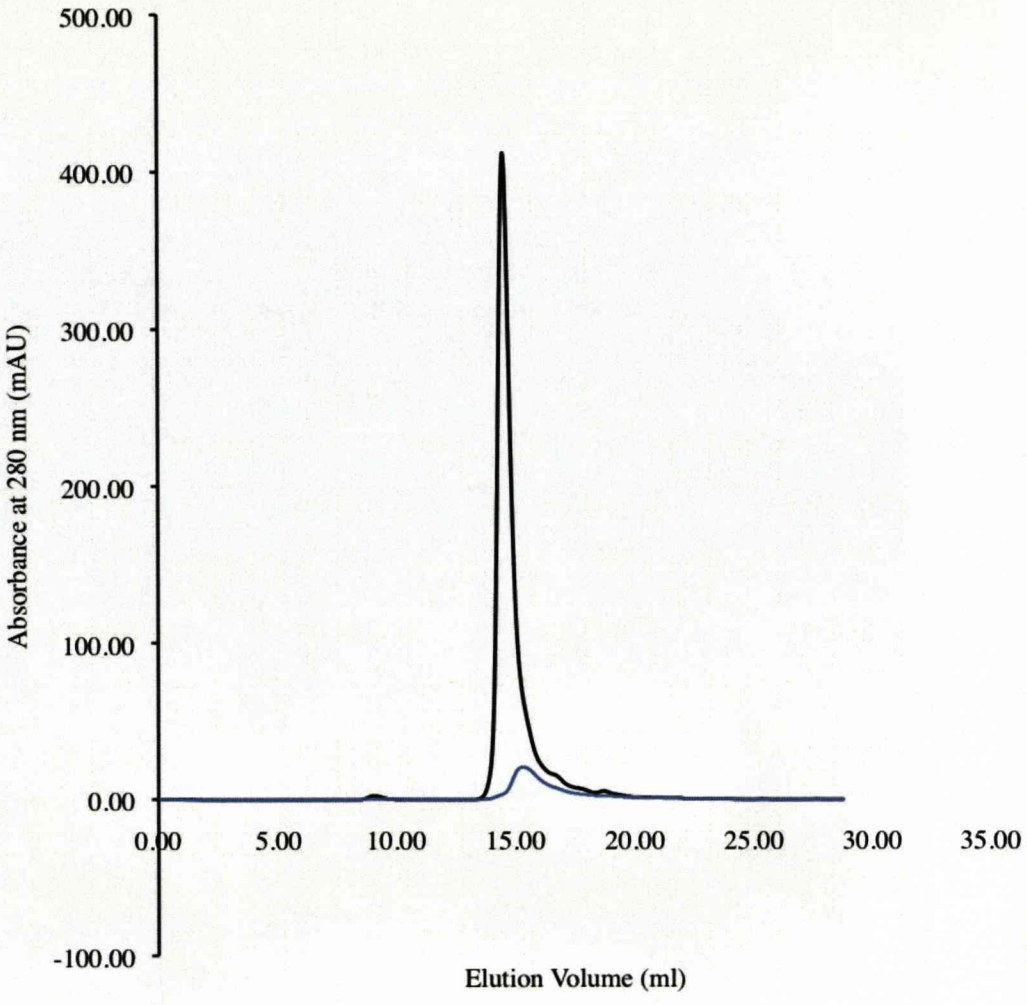
Fig; 4.6: Gel filtration chromatography to determine LBTB binding of calcium. Incubations were set up with 45 µl LBT peptide plus 5 µl of Ca^{2+} (50 µM) for 1 h. After incubation each peptide/metal mixture was subjected to a centrifugation step at 13000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each incubation and run at a flow rate of 0.5 ml/min with Na-HEPES 100 mM, KCl 150 mM, (pH 7.0). Trace is representative of typical results for incubation of this peptide with Ca^{3+}.



Fig; 4.7: Gel filtration chromatography to determine LBTB binding of La^{3+} (1 M). Incubations were set up with 45 µl LBT peptide plus 5 µl of La^{3+} (1 M) for 1 h. After incubation each peptide/metal mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each incubation and run at a flow rate of 0.5 ml/min with Na-HEPES 100 mM, KCl 150 mM, (pH 7.0). Trace is representative of typical results for incubation of this peptide with La^{3+} (1 M).

An additional incubation was set up using La^{3+} at the lower concentration of 50 μM to assess whether the peptide behaved differently when exposed to lower levels of the lanthanide. Figure 4.8 showed that indeed this is the case. A single peak resulted from this GFC. Both La^{3+} occupied and unoccupied LBTB may have been present but eluted at almost the same time as there is very little difference in size. Again, precipitation is evident from the decreased height of the peak.

ITC was used to directly measure the enthalpy changes associated with lanthanide binding to each isolated peptide and also yield binding isotherms allowing quantification of equilibrium binding constants (Kd). In order to obtain meaningful data a sigmoidal binding curve is required so it can be fitted to a binding model, typically a one-site model. Initially, estimations were made for the concentration of ligand and the concentration of the isolated peptide in the calorimeter cell. A number of ITC runs were performed using various concentrations of peptide (LBTB) and ligand (La^{3+}) until suitable enthalpy values were obtained using peptide at a concentration of 1.25 mM and a ligand concentration of 50 μM. The buffer used to dissolve the peptide had to be an exact match (pH, buffer concentration, salt concentration etc.) to the buffer in which the metal chloride was dissolved to avoid spurious heat effects from buffer mixing that would inevitably mask enthalpies arising from binding (Fisher and Singh, 1995).

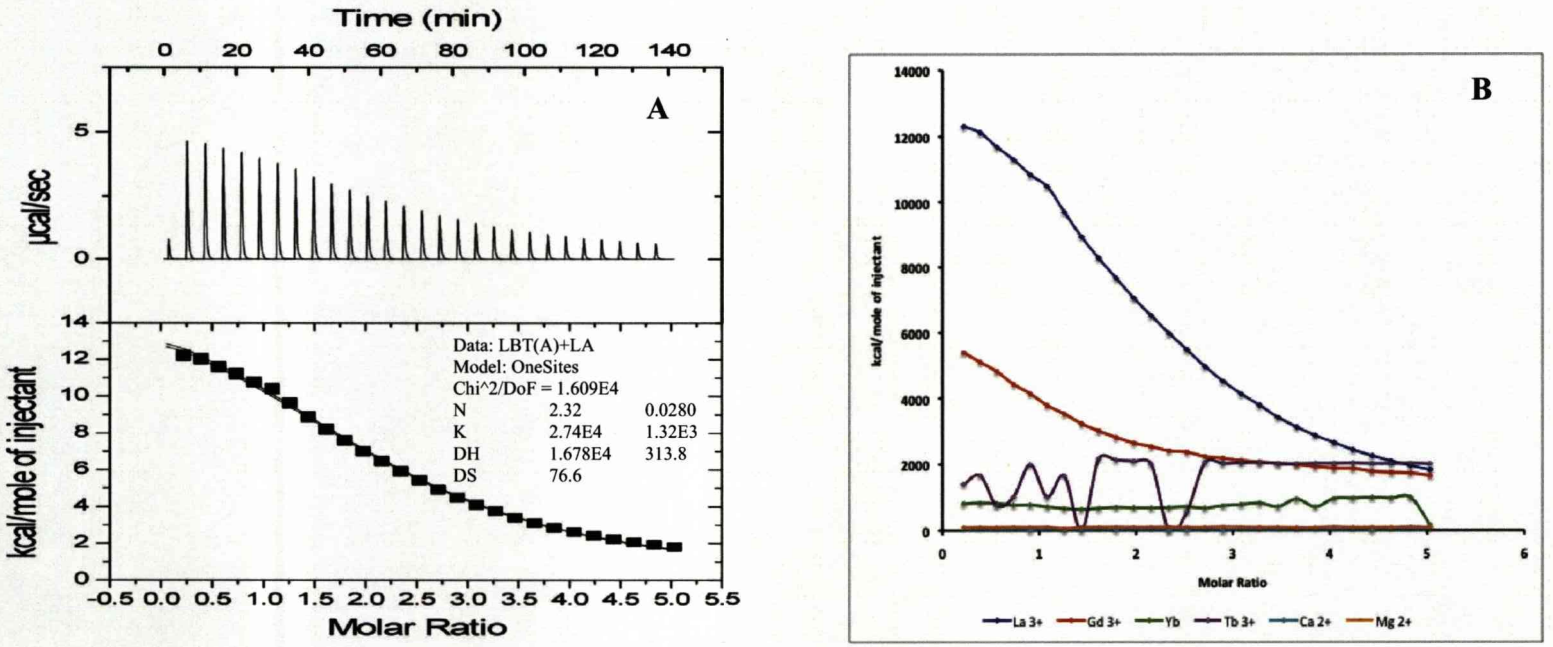


Fig; 4.8: Gel filtration chromatography to determine LBTB binding of La^{3+} (50 µM).

Incubations were set up with 45 µl LBT peptide plus 5 µl of La^{3+} (50 µM) for 1 h. The control incubation curve (45 µl LBTB peptide plus 5 µl TBS) is shown (black) for comparison. After incubation each peptide/metal mixture was subjected to a centrifugation step at 13 000 x g. A 24 ml Superdex Peptide column (GE Healthcare) was loaded with 20 µl of each incubation and run at a flow rate of 0.5 ml/min with Na-HEPES 100 mM, KCl 150 mM, (pH 7.0). Trace is representative of typical results for incubation of this peptide with La^{3+} (50µM).

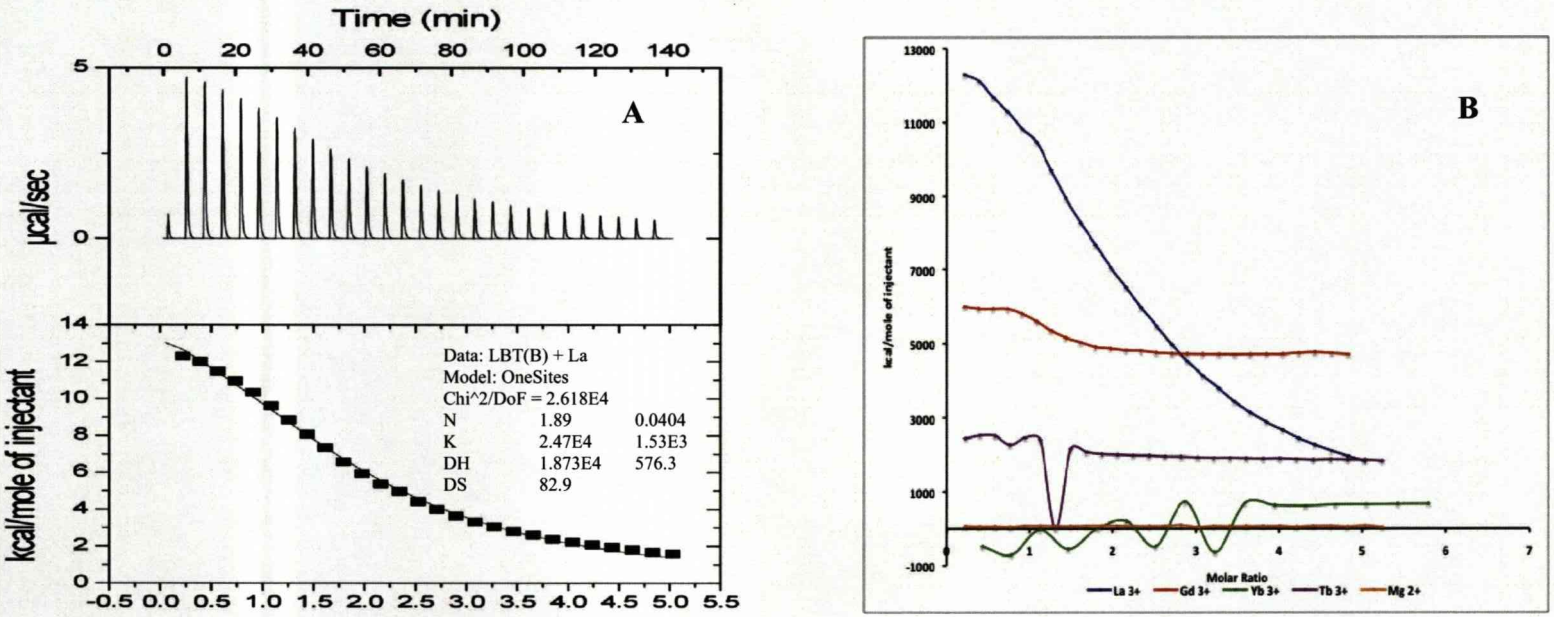
A number of buffers were tried, including; TBS buffer (20 mM Tris-Cl, 0.1 M NaCl, pH 7.5), Tris-Cl (10 mM, pH 8.0) and MES (100 mM, pH 6.0) before a HEPES buffer (Na-HEPES, 100 mM, KCl, 150 mM, pH 7.0) was finally chosen. Less lanthanide precipitation was observed when using the HEPES buffer.

ITC experiments were performed for each isolated peptide using Ca^{2+}, Mg^{2+}, Gd^{3+}, Yb^{3+}, Tb^{3+} and La^{3+}. All data obtained indicated that metal-chloride binding to each isolated peptide was an endothermic process (Figures 4.9 - 4.10). Figures show that only titrations using lanthanum produced the desired sigmoid curve. Kd values for LBTA and LBTB were calculated from the ITC data as 36.5 μM and 40.5 μM respectively (Figures 4.9A - 4.10A). In each case, data suggested that 2 binding sites were available for La^{3+} although the curve could be, and was, fitted to a one-site model. All other ITC titrations using the aforementioned metal chlorides indicated very weak binding, so Kd values could not be calculated. However, the binding abilities could be seen by comparison in figures 4.9B - 4.10B. These figures suggest that both peptides were able to weakly bind Gd^{3+}. All other metal chlorides produced erratic, low enthalpy peaks during ITC, which probably suggest that there was no binding.



Fig; 4.9: ITC analysis of LBT(A) binding using metal chlorides

ITC experiments performed using VP-ITC MicroCalorimeter according to manufacturers manual. The calorimeter sample cell was filled with 2 ml degassed peptide (50 µM) in HEPES buffer [Na-HEPES, 100 mM, KCl, 150 mM, pH 7.0]. The syringe was loaded with 250 µl degassed metal chloride (1.2 mM) dissolved in HEPES buffer identical to that used for dissolving the peptide. The temperature was set at 25°C and injections of 10 µl were released from the syringe every 300 s as programmed by VPViewer software. (A) Binding curve represents enthalpies of binding of La^{3+} addition to the peptide. Binding curve was fit to a one binding site model to determine the binding affinity. (B) Curves obtained for the other metal chlorides (Mg^{2+}, Gd^{3+}, Yb^{3+}, Tb^{3+}) were of low affinity and could not be fit to a binding site model. Binding affinity can be estimated by comparison to that obtained for La^{3+} and to one another.



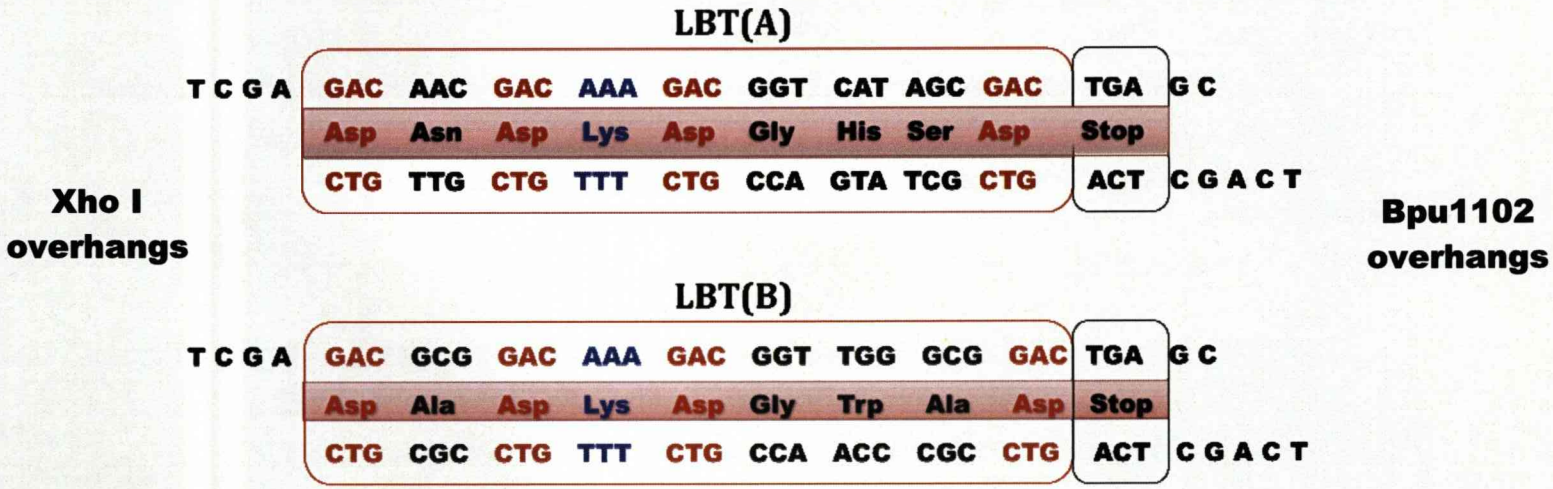
Fig; 4.10: ITC analysis of LBT(B) binding using metal chlorides

ITC experiments performed using VP-ITC MicroCalorimeter according to manufacturers manual. The calorimeter sample cell was filled with 2 ml degassed peptide (50 µM) in HEPES buffer [Na-HEPES, 100 mM, KCl, 150 mM, pH 7.0]. The syringe was loaded with 250 µl degassed metal chloride (1.2 mM) dissolved in HEPES buffer identical to that used for dissolving the peptide. The temperature was set at 25°C and injections of 10 µl were released from the syringe every 300 s as programmed by VPViewer software. (A) Binding curve represents enthalpies of binding of La^{3+} addition to the peptide. Binding curve was fit to a one binding site model to determine the binding affinity. (B) Curves obtained for the other metal chlorides (Mg^{2+}, Gd^{3+}, Yb^{3+}, Tb^{3+}) were of low affinity and could not be fit to a binding site model. Binding affinity can be estimated by comparison to that obtained for La^{3+} and to one another.

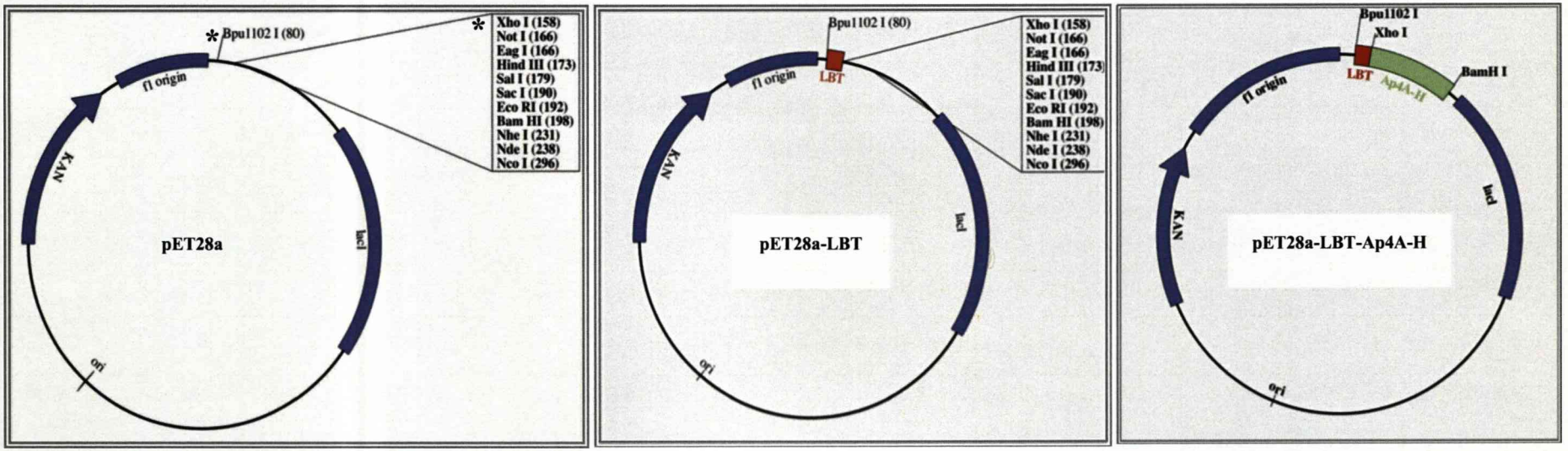
4.3: Cloning and expression of Ap_4A-H-lanthanide binding tag fusion

Once it was discovered that each tag was capable of accommodating metals, they were both considered for use in the heterologous expression of proteins. Two complementary oligonucleotides were designed for each tag so that they could be hybridized to create the designed tag (LBTA or LBTB) (Figure 4.11). The C-terminal His-tag was to be removed from pET28a expression plasmid and replaced with sequence encoding for either LBTA or LBTB followed by a stop codon (Figure 4.12). Xho I and BpuII02 restriction sites of the plasmid were used to achieve this. The desired overhangs were incorporated into each tag sequence to facilitate ligation and to allow for each tag to be expressed in-frame with the model protein. The multiple-cloning region remained intact so options were available for the cloning of the target, model protein (Figure 4.12). The designed oligonucleotides (supplied by MWG) were hybridized by heating to 95°C for 5 min and slowly cooling to room temperature as detailed in section 2.4.1.

For the engineering of the expression plasmid incorporation the LBTB region the expression plasmid, pET28a, was digested with Xho I and BpuII02 restriction enzymes. A ligation reaction followed (as detailed in section 2.4.1) for insertion of hybridized oligonucleotides into the plasmid. 10 µl of each ligation (LBTA and LBTB) was used for transformation of XL1-Blue *E.coli,* which were ultimately used for sequencing (Cogenics Ltd, U.K) of the modified plasmid.



Fig; 4.11: Design of oligonucleotides and hybridization for lanthanide binding tags
For construction of each LBT region two oligonuclotides were hybridized leaving an N-terminal Xho I overhang and a C-terminal Bpu1102 overhang for insertion by ligation into the desired expression plasmid. The coding amino acid sequence is shown through the centre of each oligonucleotide hybrid.



Fig; 4.12: Expression plasmid construction

(*Left*) pET28a plasmid is shown highlighting restriction sites for insertion of the LBT region (*). (*Middle*) The pET28a expression plasmid was engineered to incorporate the LBT region (red) with the full use of all of the restriction sites of the multiple cloning region for future cloning. (*Right*) The pET28a-LBT plasmid is able to incorporate sequence for the model protein, Ap_4A-H (green) between BamH I and Xho I restriction sites for expression of LBT-tagged Ap_4A-H.

Transformants containing the tag insert were determined by colony PCR using a T7 promoter primer and a reverse primer designed for each LBT region. Sequencing revealed that LBTB was successfully incorporated into the vector in the correct frame. This was not possible for LBTA as none of the positive colonies sent for sequencing contained the tag insert. Therefore studies were continued using LBTB only.

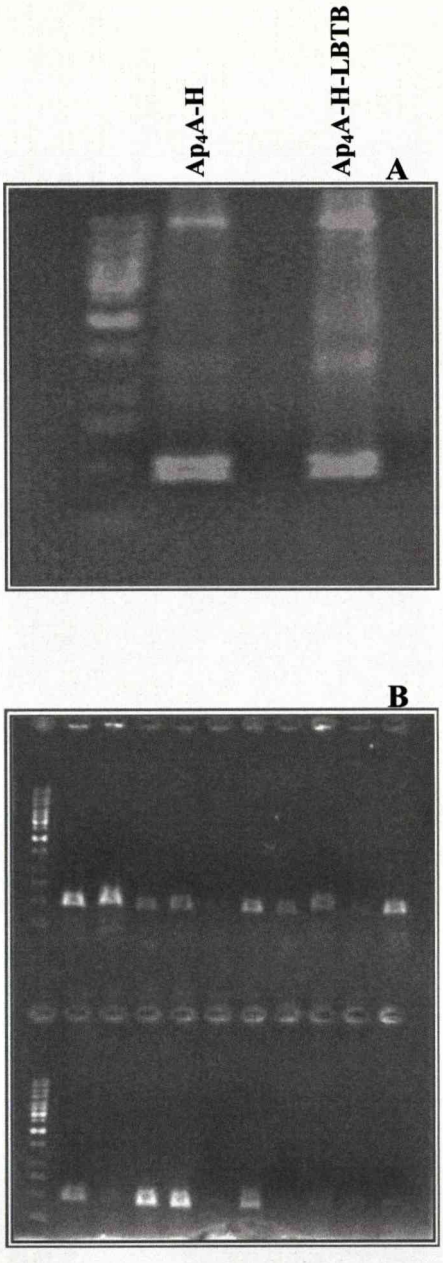
Asymmetric diadenosine-5',5'''-P1,P4-tetraphosphate (Ap_4A) Hydrolase was employed as the model protein for investigations using the metal-binding LBTB. The target sequence (kindly provided by Prof. A. G. McLennan) was encompassed in the pNEBRX1-hygro plasmid (NEB). This insert was previously directionally cloned into this plasmid using Bam HI and Xho I. Although these sites are available in the pET28a plasmid, cloning required the design of primers and PCR amplification to maintain the correct reading frame for expression of Ap_4A-H-LBTB. The forward primer, F-Ap_4A-H (Table 4.1) was designed to amplify the target sequence in the correct frame in the pET28a-LBTB plasmid. Two reverse primers (Table 4.1) were designed so that Ap_4A-H insert could be used in the modified plasmid, pET28a-LBTB (for expression of the LBTB-tagged protein) and also in the original pET28a plasmid (for expression of a non-LBTB-tagged protein) to serve as a control. For expression of non-tagged-Ap_4A-H the reverse primer, R-Ap_4A-H-stop incorporated a TGA stop codon followed by the Xho I restriction site. The second reverse primer was designed to replace the TGA stop codon with GGA coding for Gly followed by the Xho I restriction site in frame with the LBT sequence to enable LBT-tagging. Both tagged and the non-tagged proteins made use of the N-terminal His tag (provided by the plasmid) for purification and detection.

Table; 4.1: Oligonucleotide primers for amplification of Ap_4A-H for expression both with and without the LBTB tag.

Primer	Oligonucleotide sequence
F-Ap4A-H	5 – ATCTGAATTCGGATCCatggccttgagagcatgtggc – 3'
R-Ap4A-H	5 – TACTGAAGCTTCTCGAGtccggcctctatggagcaaag– 3'
R-Ap4A-H-stop	5 – TACTGAAGCTTCTCGAGTtccggcctctatggagcaaag– 3'

The pNEBRX1-hygro plasmid harbouring the sequence of interest served as the DNA template for PCR. First an optimum annealing temperature was deduced using BioMix Red reaction mix (BioLine) containing a BIOTAQ DNA polymerase for amplification at annealing temperatures of 45°C, 50°C, 56°C and 60°C in the presence and absence of DMSO (1% (v/v). All PCR reactions were successful, particularly those containing DMSO, so the highest annealing temperature of 60°C was selected for scale-up amplifications to reduce the probability of non-specific binding of primers. PCR products of approximately the predicted size (542 bp) and at a sufficient yield were obtained from a 250 µl PCR reaction (Figure 4.13A) using Platinum Pfx DNA polymerase in the presence of DMSO (1% (v/v)) (as detailed in section 2.4.2). PCR purification and restriction digests were performed as detailed in section 2.3.3 and 2.3.6. Digests were cleaned up using PCR purification kit (Qiagen) as problems were encountered with DNA recovery using the gel extraction purification in earlier investigations (as detailed in section 2.3.7). A small sample of each purified digest was subjected to agarose gel electrophoresis to ensure recovery from the purification column before ligation of each insert (Ap_4A-H and Ap_4A-H-LBTB) into the respective plasmid expression vectors (pET28a and pET28a-LBTB) by use of T_4 DNA ligase (Invitrogen). A 10 µl aliquot of the ligation mixture was used for a successful transformation into competent XL1-Blue *E. coli* (a non-

expression host, prepared as described in section 2.3.9). A colony PCR followed by agarose gel electrophoresis identified a number of colonies containing inserts of the correct size. Consequently, three of these positive colonies were selected for sequencing (Cogenics Ltd, U.K) to assess accuracy of amplification (Figure 4.13B). The returned sequences were scrutinized to ensure that no errors were incorporated during PCR amplification. Blast sequence alignments (NCBI) were performed to compare amplified DNA and the desired target sequences. Translation of each of the sequences from amplified DNA provided confirmation that the sequence was in frame and the LBT region for construct for LBT-tag addition was also intact and in-frame (Figure 4.14). A colony for each construct was chosen with a perfect alignment. Each colony was propagated overnight at 37°C in 100 ml LB/ampicillin broth and was used for plasmid isolation using a plasmid midi kit (Qiagen). The resulting expression vector constructs, pET28a-Ap_4A-H and pET28a-Ap_4A-H-LBTB, were then used for transformation BL21(DE3) *E. coli* strain as detailed in section 2.3.9 that is capable of expressing the foreign DNA insert.



Fig; 4.13: Agarose gel electrophoresis for identification of PCR products

(A) A scale-up PCR (250 µl) using optimum conditions (60°C, 1% DMSO) was performed for each DNA insert. *(B)* Gel shows colony PCR reactions for identification of colonies containing an insert of the correct size. Ten single colonies from Ap_4A-H (upper gel section) and Ap_4A-H-LBTB (lower gel section) transformed XL1-Blue cells were selected and grown up overnight in LB-broth (50 µg/ml kanamycin). A PCR was performed on 5 µl of each culture. 5 µl of each PCR reaction was loaded onto a 1% agarose gel with 1µl gel loading buffer (6X) (Fermentas) to separate DNA products. 6 µl of 1 kb generuler (SM0311, Fermentas) 14 fragments (in bp): 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500, 250 was loaded into end lanes to monitor migration.

```
catgnttcngncacnctccctctataataattttgtttaactttaaggaaggagatatacc
  M  X  X  X  X  S  L  Y  N  N  F  V  -  L  -  G  R  R  Y  T
atgggcagcagccatcatcatcatcatcacagcagcggcctggtgccgcgcggcagccat
 M  G  S  S  H  H  H  H  H  H  S  S  G  L  V  P  R  G  S  H
atggctagcatgactggtggacagcaaatgggtcgcggatccatggccttgagagcatgt
 M  A  S  M  T  G  G  Q  Q  M  G  R  G  S  M  A  L  R  A  C
ggcttgatcatcttccgaagatgcctcattcccaaagtggacaacaatgcaattgagttt
 G  L  I  I  F  R  R  C  L  I  P  K  V  D  N  N  A  I  E  F
ttactgctgcaggcatcagatggcattcatcactggactcctcccaaaggccatgtggaa
 L  L  L  Q  A  S  D  G  I  H  H  W  T  P  P  K  G  H  V  E
ccaggagaggatgacttggaaacagccctgagggagacccaagaggaagcaggcatagaa
 P  G  E  D  D  L  E  T  A  L  R  E  T  Q  E  E  A  G  I  E
gcaggccagctgaccattattgaggggttcaaaagggaactcaattatgtggccaggaac
 A  G  Q  L  T  I  I  E  G  F  K  R  E  L  N  Y  V  A  R  N
aagcctaaaacagtcatttactggctggcggaggtgaaggactatgacgtggagatccgc
 K  P  K  T  V  I  Y  W  L  A  E  V  K  D  Y  D  V  E  I  R
ctctcccatgagcaccaagcctaccgctggctggggctggaggaggcctgccagttggct
 L  S  H  E  H  Q  A  Y  R  W  L  G  L  E  E  A  C  Q  L  A
cagttcaaggagatgaaggcagcgctccaagaaggacaccagtttctttgctccatagag
 Q  F  K  E  M  K  A  A  L  Q  E  G  H  Q  F  L  C  S  I  E
gccggaactcgagacgcggacaaagacggttgggcggactgagctgagcaataactagca
 A  G  T  R  D  A  D  K  D  G  W  A  D  -  A  E  Q  -  L  A
```

Fig; 4.14: Translation of PCR amplified Ap_4A-H-LBTB

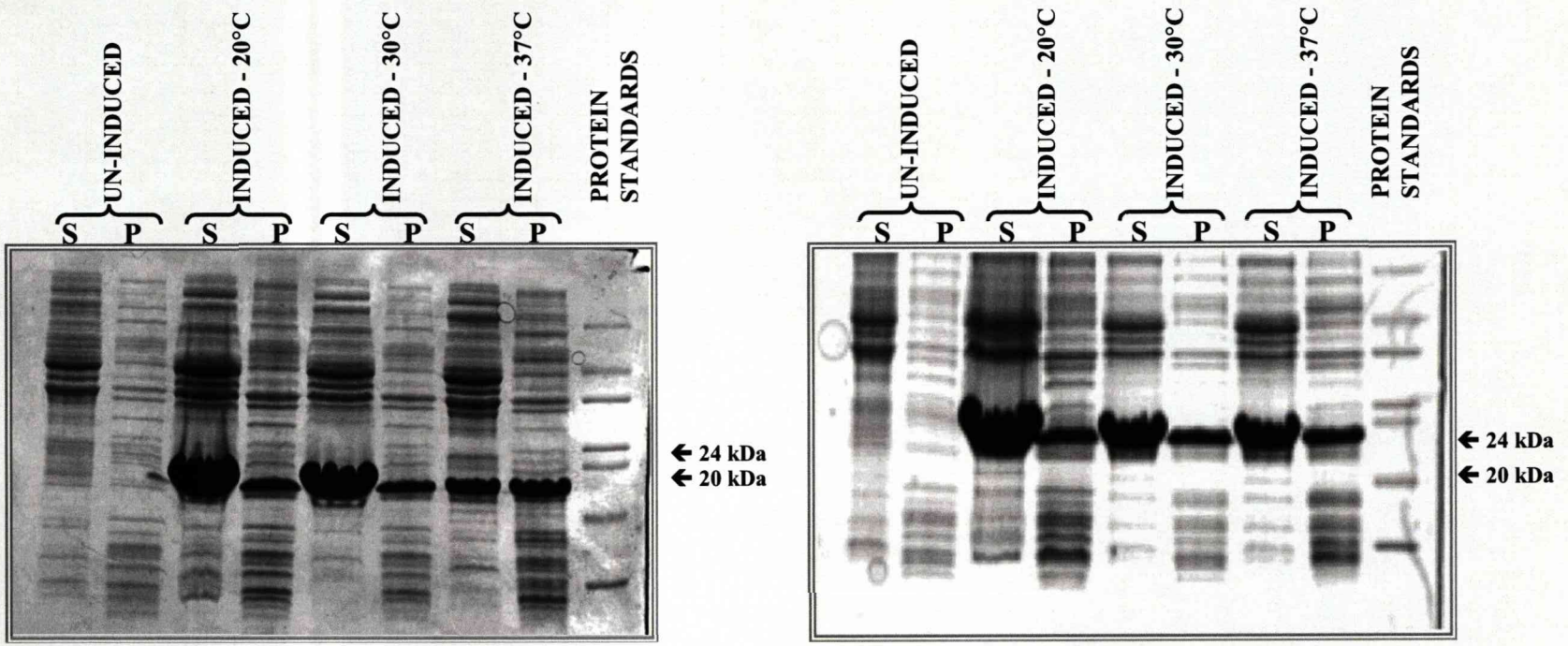
Figure shows the translation of the PCR amplified Ap_4A-H-LBTB nucleotide sequence highlighting the N-terminal His tag (highlighted in red) and the LBTB region at the C-terminal domain (highlighted in Blue) followed with a stop/termination codon.

4.4: Effect of LBTB tag on target protein expression

BL21(DE3) (Stratagene) cells were employed to host the expression of target proteins. These cells are known for directing high levels of protein expression and ease of induction and this was the chosen strain for structural studies of the model protein by Swarbrick *et al* (2005).

Full details for recombinant protein expression can be found in section 2.3.12. B-PER reagent (Pierce) was used to achieve cell lysis (as described in section 2.3.13.2). Results clearly illustrate bands in IPTG-induced cultures for both target proteins (Figure 4.15). Ap_4A-H and Ap_4A-H-LBTB proteins migrating to 22 kDa and 23 kDa respectively. Expression carried out at 20°C produced the highest levels of soluble protein in each case.

When observing soluble protein expression levels over time as a function of bacterial growth it is apparent that levels are sufficient after only 8 h (i.e. 6 h post IPTG-induction) (Figures 4.16 and 4.17). Again it was evident that the addition of the LBTB tag had no inhibitory effect on expression of the recombinant protein. Each protein was highly expressed as the major protein in induced BL21(DE3) cells as indicated by the presence of polypeptide bands migrated to 22 kDa and 23 kDa for Ap_4A-H and Ap_4A-H-LBTB respectively (Figures 4.16 and 4.17).

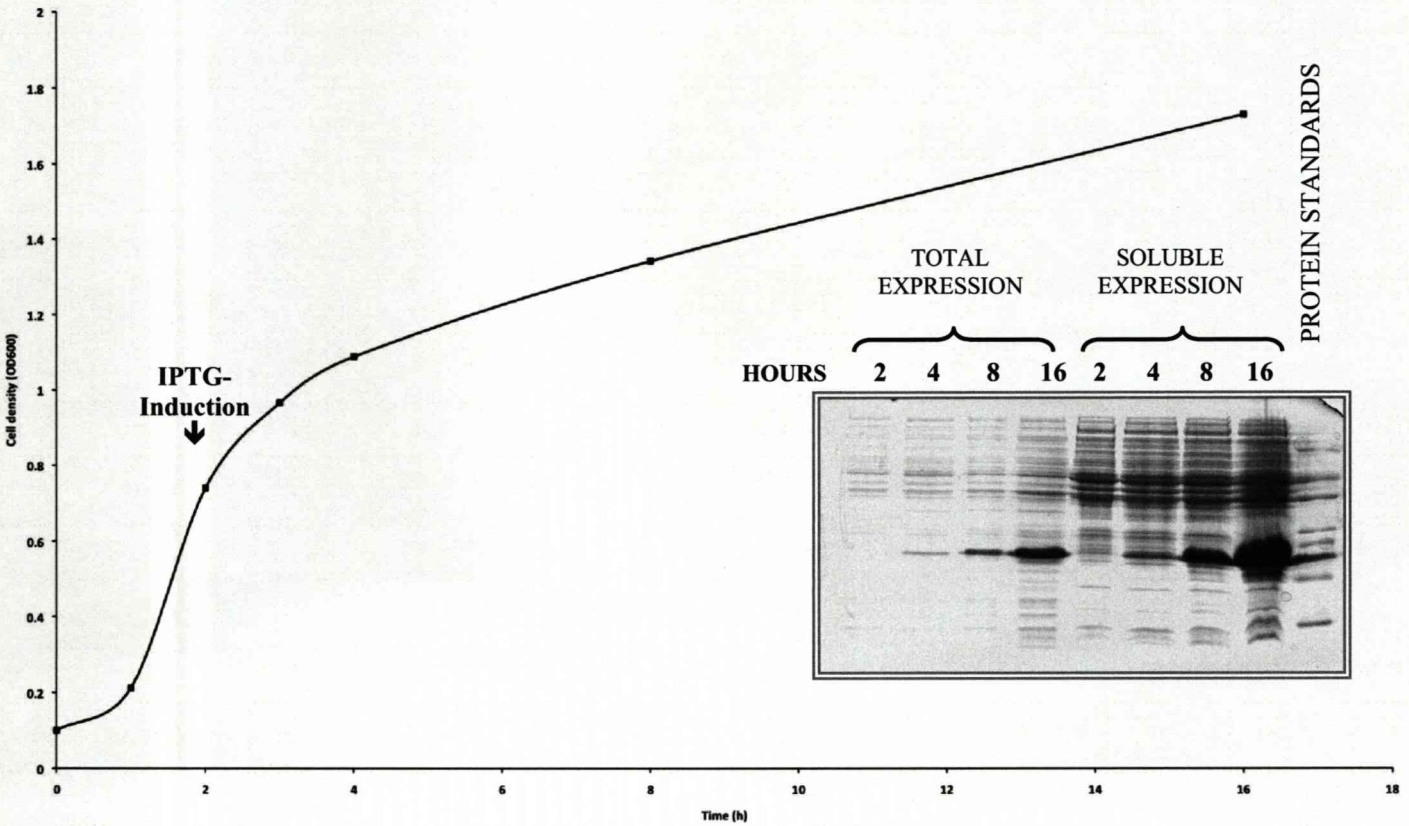


Fig; 4.15: Expression of Ap_4A-H and Ap_4A-H-LBT in Bl21(DE3) at 20°C 30°C and 37°C

Ap_4A-H (left) and Ap_4A-H-LBT (right) expression was induced in Bl21(DE3) *E.coli* at 20°C 30°C and 37°C. Samples were removed aseptically at 0 h (immediately before 0.5 mM IPTG addition) and at 16 h post induction. Cells were harvested by centrifugation at 13 000 x g and lysed by B-PER bacterial protein extraction reagent (Pierce). Soluble supernatant (s) and insoluble pellet (p) fractions were separated by a centrifugation at 13 000 x g. Fractions were analysed by SDS-PAGE and visualized by Coomassie Brilliant Blue. The positions and relative molecular masses of protein standards are indicated on the right. Polypeptide bands of 22 kDa (left) and a 23 kDa (right) were detected in each IPTG-induced *E.coli* strain but not in the un-induced cells.

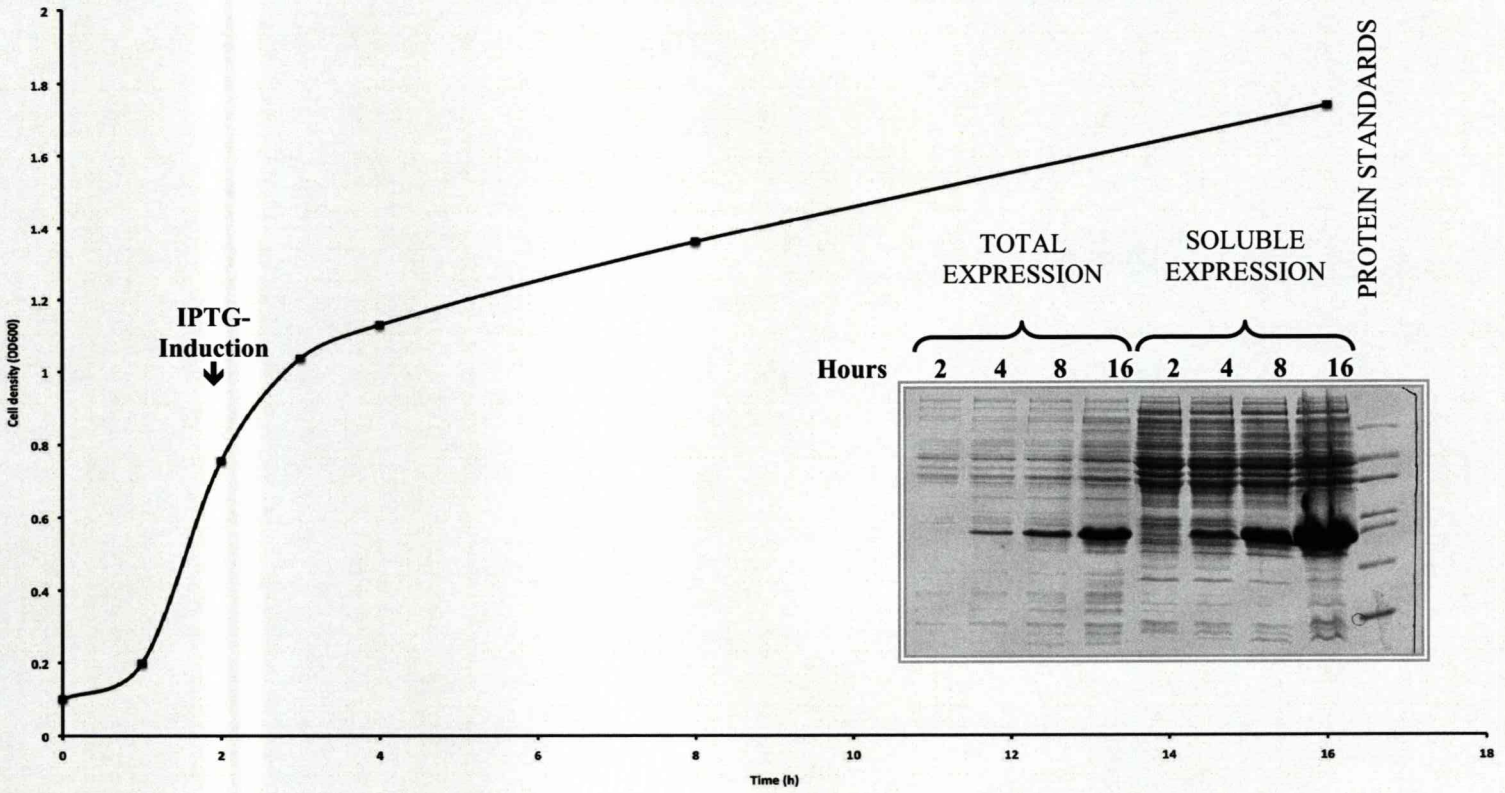
4.5: The effect of LBTB addition on IMAC purification on target proteins

Larger 1 l cultures were induced so that each His-tagged recombinant protein could be separated from the majority of host proteins by IMAC. Indeed, this was assuming that no complications would arise resulting from the addition of the LBTB region. Details for large-scale expression are in full in section 2.3.12. SDS-PAGE analysis of eluted fractions clearly illustrates the presence of two polypeptide bands of 22 kDa and 23 kDa (Figure 4.18) are visible for 100 mM and 500 mM eluates which provided a firm indication that both the tagged and non-tagged Ap_4A-H were successfully isolated using affinity chromatography.



Fig; 4.16: Bacterial growth as a function of time and expression of Ap_4A-H over 16 h

A 1 l culture was cultivated at a starting OD_{600} of 0.1 and incubated at 37°C until cultures reached a cell density of OD_{600} = 0.7 at which point cultures were cooled to 20°C by incubation at 4°C before expression of each protein was induced using 0.5 mM IPTG. Once induced, cultures were returned for incubation at 20°C with shaking (200 rpm). Samples of 1 ml were removed Duplicate samples of 1 ml were removed after 2 h, 4 h, 8h, and 16 h and the absorbance at 600 nm was recorded to measure cell densities. Total expression samples were centrifuged (10 000 x g, 3 min) and the cell pellet re-suspended in 250 µl TBS. 10 µl of this resuspension plus an equal volume of 5X SDS-sample buffer was heated at 95°C for 5 min before loading of 10 µl into the sample well of the polyacrylamide gel. Supernatant samples were prepared using B-PER reagent for cell lysis as detailed in section 2.3.13.2.

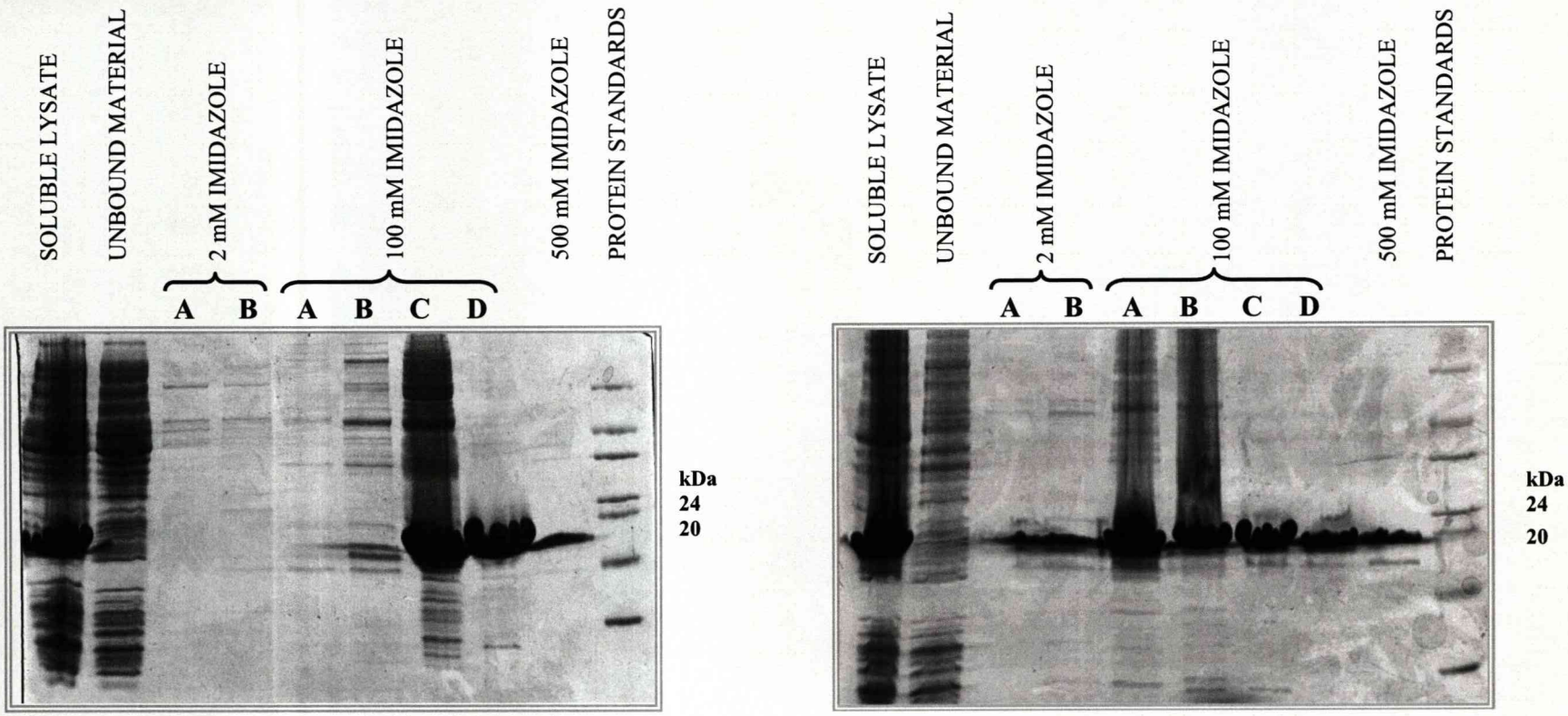


Fig; 4.17: Bacterial growth as a function of time and expression of Ap_4A-H –LBTB over 16 h

A 1 l culture was cultivated at a starting OD_{600} of 0.1 and incubated at 37°C until cultures reached a cell density of OD_{600} = 0.7 at which point cultures were cooled to 20°C by incubation at 4°C before expression of each protein was induced using 0.5 mM IPTG. Once induced, cultures were returned for incubation at 20°C with shaking (200 rpm). Samples of 1 ml were removed Duplicate samples of 1 ml were removed after 2 h, 4 h, 8h, and 16 h and the absorbance at 600 nm was recorded to measure cell densities. Total expression samples were centrifuged (10 000 x g, 3 min) and the cell pellet re-suspended in 250 µl TBS. 10 µl of this resuspension plus an equal volume of 5X SDS-sample buffer was heated at 95°C for 5 min before loading of 10 µl into the sample well of the polyacrylamide gel. Supernatant samples were prepared using B-PER reagent for cell lysis as detailed in section 2.3.13.2.

4.6: Peptide mapping of Ap_4A-H and Ap_4A-H-LBTB using RP-HPLC

The only difference between target proteins was the addition of the LBTB tag of Ap_4A-H-LBTB. In order to confirm the existence of this region in the Ap_4A-H-LBTB construct, internal peptides were generated for sequencing using an adaptation of the 'one-tube' reduction, carboxymethylation and digestion method of Stone *et al* (1992) (detailed in section 2.1.6). A peptide map of each digest (produced by reverse phase high-pressure liquid chromatography (RP-HPLC)) highlighted the presence of an additional peptide in the digested sample of the LBTB-tagged protein (Figure 4.19This peptide was collected when eluted from the column and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was used to confirm its identity. Theoretical digests of Ap_4A-H and Ap_4A-H-LBT were implemented using Peptide Mass, ExPASY, so that comparisons could be made with the data obtained from the MALDI-MS. One major difference evident between theoretical digests was a peptide of 992.4 resulting from the LBTB region (Figure 4.19). MALDI-MS also identified a peptide of approximately the same size at 992.2 (Figure 4.20), which provides evidence that the tag is present and accessible on the surface of Ap_4A-H-LBTB.



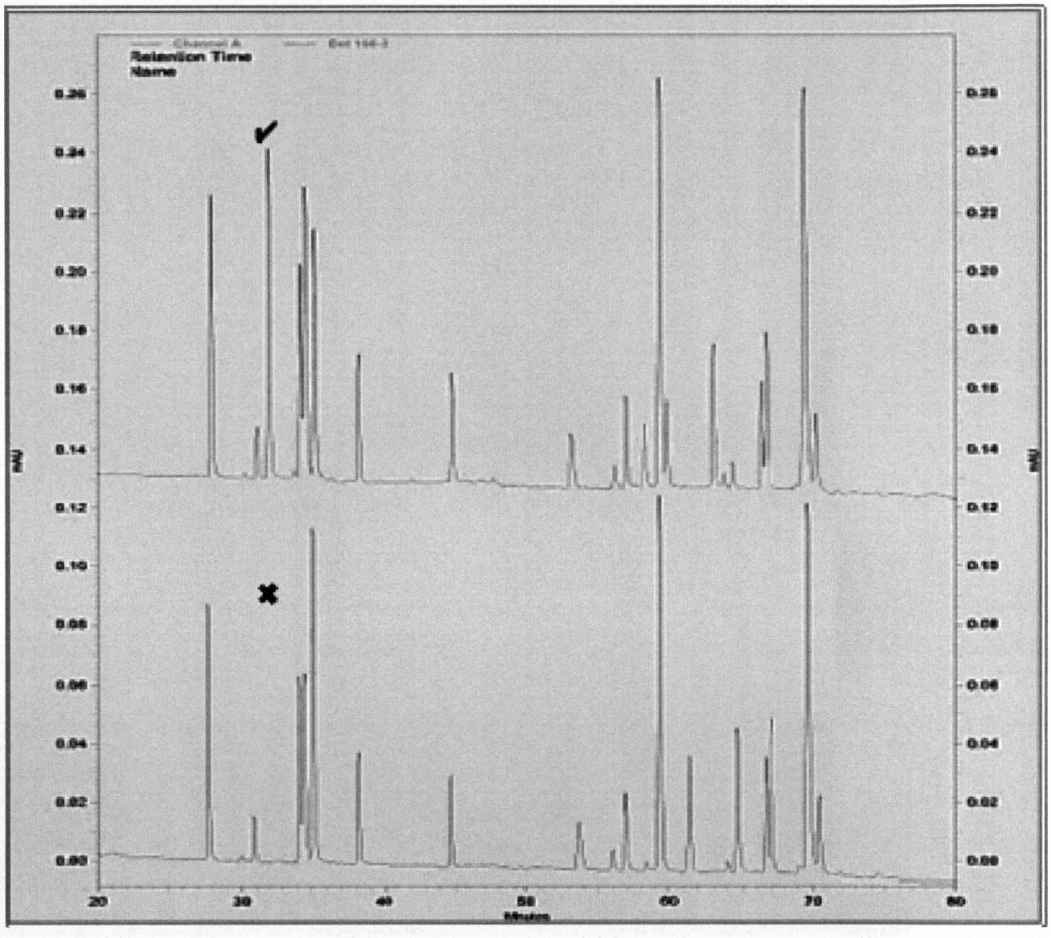
Fig; 4.18: SDS-PAGE analysis of IMAC purification of Ap_4A-H (Left) and Ap_4A-H-LBT (Right)

Ap_4A-H (Left) and Ap_4A-H-LBT (Right) expression was induced by IPTG (0.5 mM) in BL21(DE3) cells at 20°C. For 1 l expression cultures cells were harvested at 6 000 x g and re-suspended in TBS (20 mM Tris-Cl, 150 mM NaCl pH 7.5). Cell lysis was achieved by sonication (4 X 30 s). The soluble fraction was clarified by centrifugation at 20 000 x g and loaded onto a 5 ml Ni-NTA column. Samples of the soluble lysate, the unbound material, 2 mM imidazole washes and of eluates at 100 mM and 500 mM imidazole. IMAC purification samples were analysed by SDS-PAGE and visualized by Coomassie Brilliant Blue. The positions and relative molecular masses of protein standards are indicated on the right. Polypeptide bands of 22 kDa (left) and a 23 kDa (right) are visible for 100 mM-500 mM eluates.

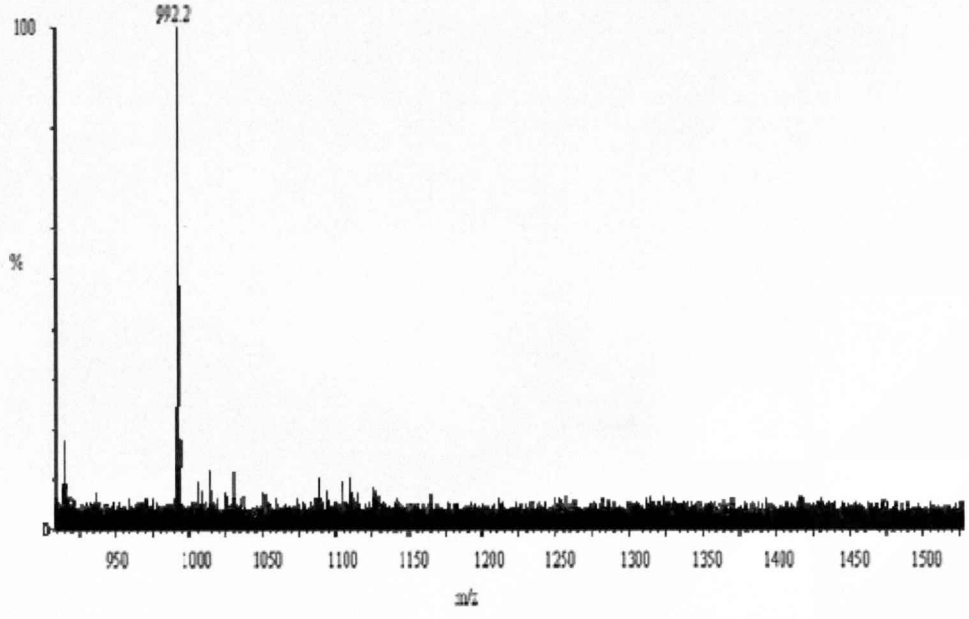
Ap_4A-H-LBT(B)
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMALRACGLIIFRRCLIP
KVDNNAIEFLLLQASDGIHHWTPPKGHVEPGEDDLETALRETQEEAGIEAG
QLTIIEGFKRELNYVARNKPKTVIYWLAEVKDYDVEIRLSHEHQAYRWLGL
EEACQLAQFKEMKAALQEGHQFLCSIEA**GTRDADKDGWAD Stop**

DADKDGWAD
Actual: 992.4 obs: 992.1 peptide 31.2

Ap_4A-H control
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMALRACGLIIFRRCLIP
KVDNNAIEFLLLQASDGIHHWTPPKGHVEPGEDDLETALRETQEEAGIEAG
QLTIIEGFKRELNYVARNKPKTVIYWLAEVKDYDVEIRLSHEHQAYRWLGL
EEACQLAQFKEMKAALQEGHQFLCSIEA Stop



Fig; 4.19: Peptide maps of Ap_4A-H-LBT (upper trace) and Ap_4A-H (lower trace)
Internal peptides for sequencing were generated using an adaptation of the 'one-tube' reduction, carboxymethylation and digestion method of Stone *et al* (1992) as detailed in section 2.1.6. Peptides were separated with a 95 min gradient of 0-64% acetonitrile in 0.08% TFA and elution was monitored at 214 nm. Amino acid sequences for each protein are shown above. The LBTB region is highlighted in **BOLD**.



Fig; 4.20: Mass Spectrum of RPLC purified peptide B of trypsin digested Ap_4A-H-LBT. Analysis was performed using alpha-cyano-4 hydroxycinnaminic acid in 50% acetonitrile/0.1% trifluoroacetic acid. Samples were selected in the mass range of 900 - 3000 Da.

4.7: Ap_4A-H activity assay on both target proteins

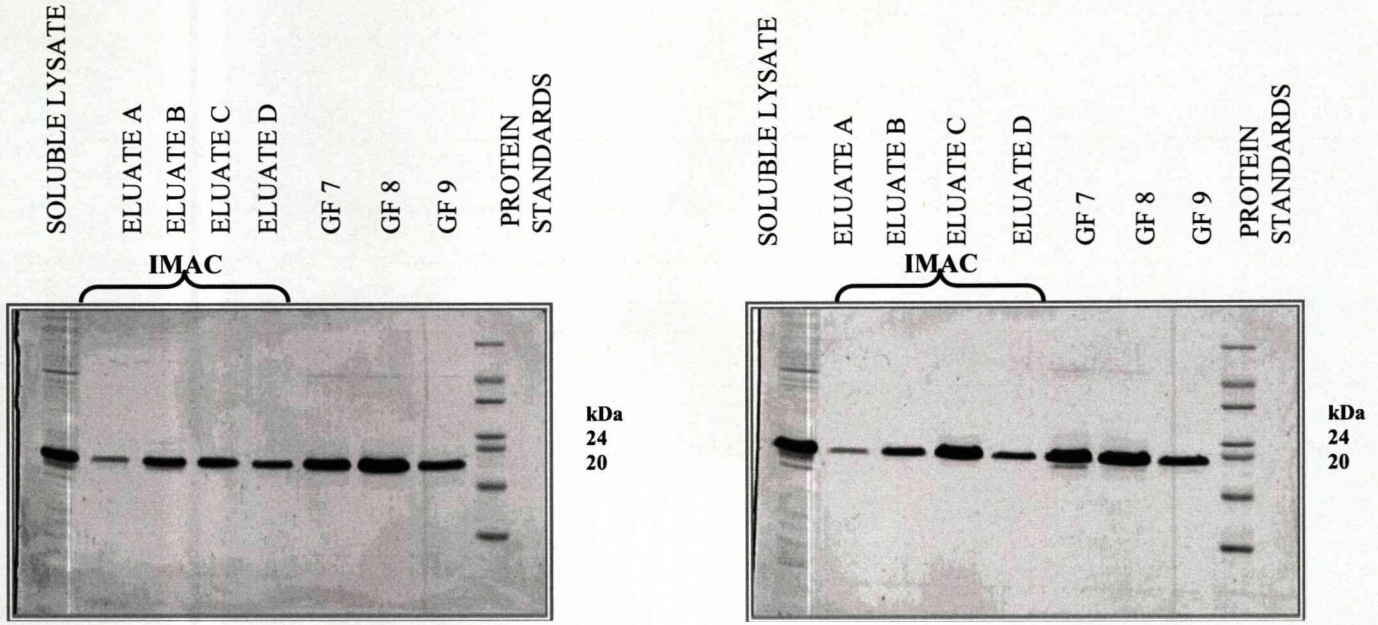
The specific activity of both Ap_4A-H and Ap_4A-H-LBTB was assayed using a luciferase-based bioluminescence assay for detection of the ATP product (Prescot *et al.*, 1989) as detailed in section 2.2.2. Samples of each protein were taken from stocks stored at -80°C of IMAC and GFC purified protein preparations for each assay. Samples were thawed on ice when required. Results showed that Ap_4A-H-LBTB and Ap_4A-H had specific activities of 523 mV/mg/min and 222 mV/mg/min, respectively. Analytical gel filtration chromatography of protein preparations of Ap_4A-H and Ap_4A-H-LBTB (stored at 4°C for 48 h) illustrated that both samples had experienced protein aggregation, with the former experiencing the most (results not shown). It is likely that this aggregation was responsible for the differences noticed in specific activities in this assay.

4.8: Crystallisation screening for each target protein

Cultures of BL21(DE3) harbouring the expression plasmid for the desired crystallisation target were cultivated (as detailed in section 2.3.12) in 1 l LB media containing 50 μg/ml kanamycin until the requisite cell density (OD_{600} = 0.7) was reached. Expression was induced with the addition of IPTG at a final concentration of 0.5 mM and incubation at 20°C for 16 h. Cells were harvested and lysed, and then each preparation purified as detailed in section 2.7.1. Using IMAC it was possible to purify the protein preparation to ≥ 90% (Figure 4.21). An additional purification step using GFC was also required so that the TBS buffer after IMAC (containing high concentrations of imidazole) could be changed for the desired buffer for crystallisation trials (10 mM Tris-Cl, pH 8.0). This provided confirmation that both Ap_4A-H and Ap_4A-H-LBTB IMAC preparations were monomeric and that

no multimers were present. Crystallisation screens were set up immediately where possible, otherwise protein preparations were stored at 4°C for no longer than 12 h.

The screening process began with the manual set up of a crystallisation trial (as detailed in section 2.7.2.1) for non-tagged Ap_4A-H so that a rough estimate of the protein concentration for future crystallisation trials could be made. This preliminary trial was set up using the Hampton Index (HR2-114) kit. The sitting drop plate allowed for the screening of three drops per well. The first sub-well was loaded with protein at a concentration of 5 mg/ml, the last sub-well at 10 mg/ml with the middle filled with buffer only. This screening process did not yield any lead crystallisation conditions (or 'hits'). Drops at 5 mg/ml were mostly clear after 2 weeks at 25°C whereas 70% of protein drops at a concentration of 10 mg/ml had resulted in precipitation after the same period of incubation.

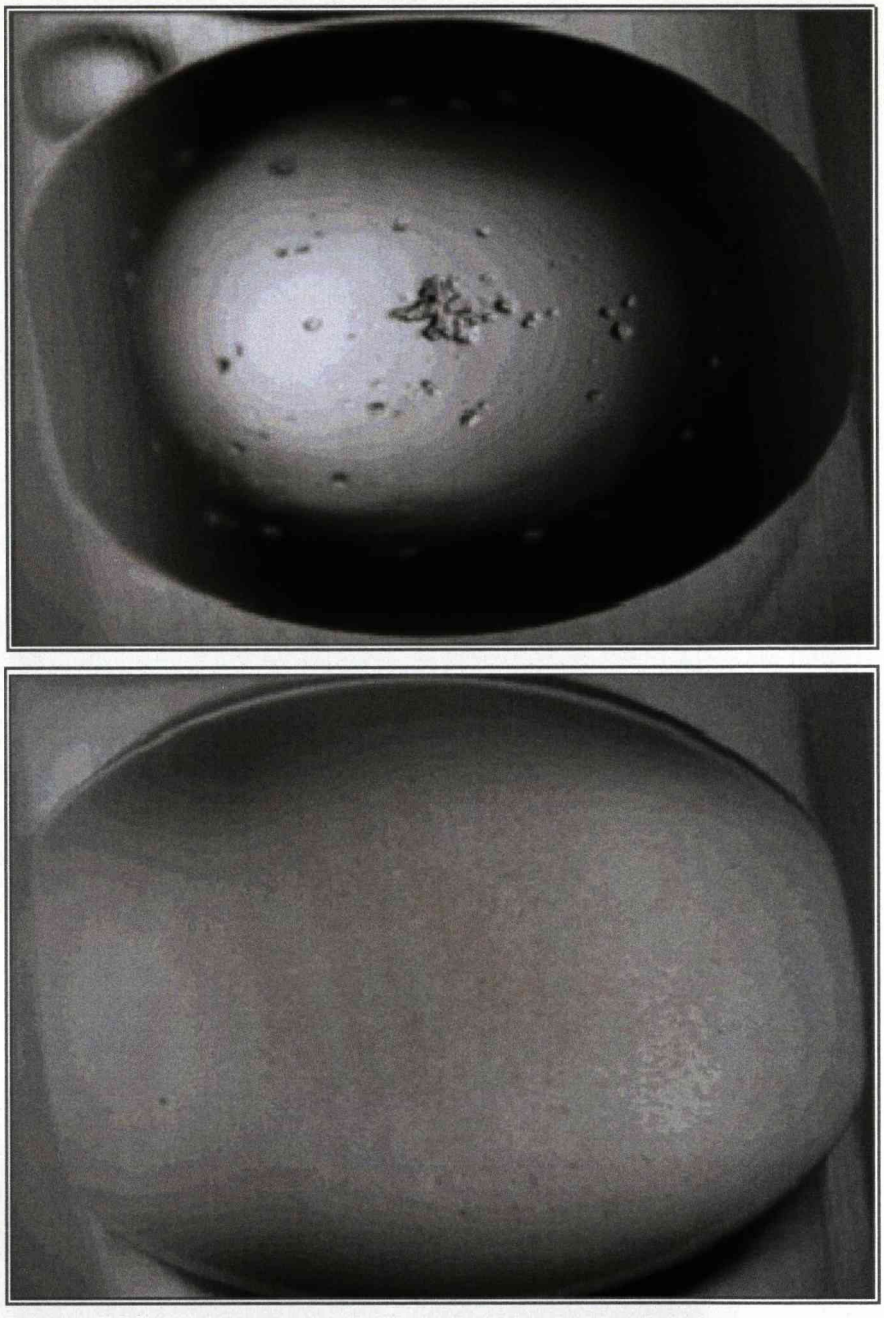


Fig; 4.21: SDS-PAGE analysis of IMAC purification of Ap_4A-H (LEFT) and Ap_4A-H-LBT (RIGHT)

Ap_4A-H (Left) and Ap_4A-H-LBT (Right) expression was induced by IPTG (0.5 mM) in BL21(DE3) cells at 20°C. For 1 l expression cultures cells were harvested at 6 000 x g and re-suspended in TBS (20 mM Tris-Cl, 150 mM NaCl pH 7.5). Cell lysis was achieved by sonication (4 X 30 s). The soluble fraction was clarified by centrifugation at 20 000 x g and loaded onto a 5 ml IMAC column. Samples of the soluble lysate, the unbound material and of eluates at 100 mM imidazole were collected. The 5 ml fraction containing highest quantity of pure target protein after IMAC was loaded onto a Superdex 75 (16/60) equilibrated in Tris-Cl (10 mM, pH 8.0) at a flow rate of 1 ml/min. Peak fractions (as determined by 280 nM absorbance) of 5 ml were collected. IMAC samples (diluted 1/50) and GF samples (diluted 1/10) were analysed by SDS-PAGE and visualized by Coomassie Brilliant Blue. The positions and relative molecular masses of protein standards are indicated on the right.

Automation of crystallisation set up, by utilization of a Screenweaver 96+8 robot (Innovadyne Techologies Inc.) allowed for greater exploration of possible crystallisation conditions. Six Qiagen screens, namely Classics II, PEG suite, PEG suite II, pHClear, Mb Class and JCSG+, were set up for Ap_4A-H and Ap_4A-H-LBTB both at a concentration of 10 mg/ml as detailed in section 2.7.2.2. A suspect growth (Figure 4.22) was noticed in a drop containing LBTB-tagged protein using a screening solution [100 mM sodium acetate, pH 4.6, 12 % PEG 4000] of the PEG suite. The drop containing Ap_4A-H control protein consisted of granular precipitate (Figure 4.22). Images were captured after 6 weeks so it is possible that after such a long incubation this drop had dried out.

An additional screen was set up based upon the crystallisation conditions (pH 5.6, PEG 4000, ammonium acetate) used in the development of X-ray diffracting crystals of the *C. elegans* homologue (Abdelghany *et al.*, 2001). Almost all drops precipitated within 24 h of the screen set up. It was decided at this point to set up a crystallisation screen of the LBTB-tagged protein in the presence and absence of La^{3+}. The binding of the lanthanide to the metal-binding region should, according to our predictions, result in the tag becoming more rigid and may aid crystallisation. This crystallisation screen was carried out, but this time at the lower protein concentration of 5 mg/ml as more than 80 % of protein drops had precipitated in the previous trials. This time, the Ap_4A-H-LBTB protein served as a control and was set up in the adjacent sub-well for comparison. Unfortunately, this screen did not yield any lead crystallisation conditions to date. Moreover, the addition of La^{3+} to the protein preparation was observed to cause a decrease in solubility rather than an increase (Table 4.2).



Fig; 4.22: Comparing crystallisation drops images for Ap_4A-H-LBTB *(top)* and Ap_4A-H *(bottom).*

Crystallisation screening was performed using the sitting drop vapour diffusion. 1µl of each protein sample (10 mg/ml) plus the equivalent volume of the buffer solution [100 mM sodium acetate, pH 4.6, 12 % PEG 4000] of the PEG Suite (Qiagen) was loaded into a sub-well. The buffer well was filled with 80 µl of the same buffer. The crystallisation tray was stored at 25°C. Drops were viewed using 500X magnification after 6 weeks.

Interestingly, the presence of La^{3+} resulted in an increase in protein precipitation in all preparations in Na-HEPES solutions of the PEG suite. This was also the case for ammonium sulphate solutions (0.5 M – 3.5 M) containing protein in the presence of La^{3+}, which also caused an increase in precipitation.

Crystallisation Suite	Increased solubility after 1 week (%)	Decreased solubility after 1 week (%)
pHClear	1	16
Mb Class	7	26
PEG	7	44
PEG II	6	18
JCSG	18	22
Classics II	8	15

Table; 4.2: The effect of La^{3+} addition on the solubility of Ap_4A-H-LBTB
All 96 wells of each crystallisation screen were scrutinized after 1 week and scores were given for each based on the extent of precipitation, with a score of 5 representing dense precipitation with 1 representing a clear drop. A difference in the precipitation score of 2 or more between Ap_4A-H-LBTB preparation with La^{3+} and the one without La^{3+} represented an effect on solubility. The number of protein drops affected by the addition of La^{3+} was counted as either causing an increase or a decrease in solubility.

All drops were scrutinized after 1 week and scores were given for each based on the extent of precipitation, with a score of 5 representing dense precipitation with 1 representing a clear drop. Comparisons could be made based on scores to determine the effect that the LBTB tag had on Ap_4A-H in a vast number of conditions. Where there was a difference in the precipitation score of 2 or more between LBTB- tagged and non-tagged Ap_4A-H the effect on solubility was recorded.

Crystallisation Suite	Increased solubility after 1 week (%)	Decreased solubility after 1 week (%)
pHClear	16	10
Mb Class	22	7
PEG	11	25
PEG II	7	11
JCSG	11	16
Classics II	17	16

Table; 4.3: The effect of LBTB tag addition on solubility.

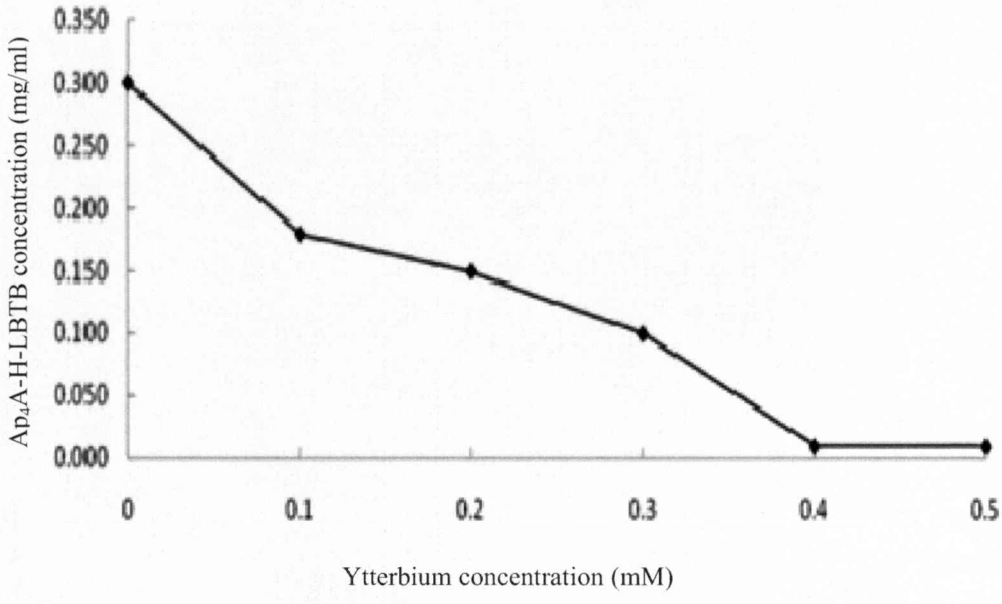
All 96 wells of each crystallisation screen were scrutinized after 1 week and scores were given for each based on the extent of precipitation, with a score of 5 representing dense precipitation with 1 representing a clear drop. A difference in the precipitation score of 2 or more between LBTB- tagged and non-tagged Ap_4A-H represented an effect on solubility. The number of protein drops affected by the addition of the LBTB tag was counted as either causing an increase or a decrease in solubility.

Table 4.3 shows the percentage of drops in each screen suite where an increase or decrease in solubility of the protein arose from the attachment of the LBTB tag to the protein. For the most part, solubility was unchanged: more than 70% of drops in each screen were unaffected by LBTB tag addition. Increased solubility was experienced mostly in drops at neutral pH values (pH 6.5 - pH 8.0). This was particularly apparent in PEG suite and PEG II suite solutions. For protein samples in buffers containing PEG at a lower pH (pH 4.6 - 5.5) the LBTB tag caused a decrease

in solubility. The LBTB-tag had little or no effect on solubility in solutions containing PEG 4000 at 20-32% at any of the pH values tested (pH 4.6 - pH 8.0). After viewing the Classics II screen it was evident that the presence of the LBTB-tag also increased solubility in solutions containing 2 M ammonium sulphate at pH 6.0 - pH 8.0. The tag did not have the same effect on solubility in drops containing the same ammonium sulphate solutions at lower pH values. Interestingly, increased solubility was experienced for Ap_4A-H-LBTB in solutions containing magnesium sulphate (50 mM) and magnesium acetate (50 mM and 200 mM).

4.9: Consequences of lanthanide binding on protein solubility.

To investigate the effects of the lanthanide concentration on protein solubility, a protein assay following $YbCl_3$ addition at various molar concentrations was performed (Figure 4.23). As the concentration of ytterbium is increased the amount of soluble protein present is reduced from 0.300 mg/ml with no metal addition to approximately 0.01 mg/ml at concentrations of ytterbium at 0.4 mM and higher. Even at 0.1 mM ytterbium, the protein concentration is reduced by approximately 40%.



Fig; 4.23: Effects of trivalent metal addition at various concentrations on the solubility of Ap_4A-H-LBT.

$YbCl_3$ was added to Ap_4A-H-LBT at protein concentration of 0.300 mg/ml at various molar concentrations. Following addition of metals samples were centrifuged (13000 x g, 5 min) and the concentration of soluble protein determined.

4.10: Discussion

The 'phasing problem' is another bottleneck faced in X-ray crystallography. Investigations in this chapter were aimed at the development of a metal-binding tag to facilitate protein structure elucidation by providing a means of obtaining phasing information. Studies were carried out to test the hypothesis that each of the designed metal-binding tags, namely LBTA (DNDKDGHSD) and LBTB (DADKDGWAD), would be capable of binding lanthanide metals. Additionally, the consequence on expression, solubility and stability as a result of LBTB-tag addition to Ap_4A-H was explored.

4.10.1: Exploring metal-binding to each tag (LBTA and LBTB)

Studies herein have confirmed the metal binding capabilities of both isolated lanthanide-binding tags. GFC studies demonstrated that the LBTA peptide could accommodate lanthanides, Tb^{3+}, Gd^{3+} and Yb^{3+} with the latter having the greatest affinity (Figures 4.3 - 4.4). There was no evidence of Ca^{2+} or La^{3+} binding (Figures 4.1 - 4.1). La^{3+}, possessing the largest ionic radius, did not bind. Middle block lanthanides Gd^{3+} and Tb^{3+} bound, while, Yb^{3+}, with the smallest atomic radius occupied a greater proportion of the LBTA peptide. This suggests a trend of metal binding based upon atomic radii. This is possibly due to the higher charge-to-size ratio of lanthanides across the series, an effect known as 'lanthanide contraction', and an increase in Lewis acidity, producing tighter ligand-ion interactions and an increasingly well folded peptide (Nitz *et al.*, 2004; Seitz *et al.*, 2007).

Binding studies using GFC for LBTB were not as straightforward. The addition of lanthanides, Tb^{3+}, Gd^{3+} and Yb^{3+} caused the precipitation of the peptide prior to

loading, so GFC could not be used to assess binding. The precipitation of the peptide may indicate that binding has occurred, in turn resulting in aggregation and precipitation. Complete precipitation does not necessarily indicate an inability to bind metal. On the contrary, in fact, the binding of the metal may beget aggregation by exposing hydrophobic regions of the tag (i.e. tryptophan). GFC studies indicated that LBTB is able to bind both Ca^{2+} and La^{3+} (Figures 4.6 - 4.8) and this was not observed for LBTA.

These studies imply that occupation of the LBT-tag may lead to a contraction of the overall structure making the once flexible region a more structured moiety. This is an attractive quality of the LBT tag for crystallisation. This postulate is based on the elution of a second conformation with a smaller Stokes radius than the unoccupied peptide tag, which suggests a smaller peptide despite the addition of the bound Ln^{3+} (Figures 4.3, 4.4, 4.6, 4.7, 4.8). Intriguingly, a single peak resulted from GFC using a lower concentration of La^{3+} (50 μM) for incubation with LBTB (Figure 4.8). The peak was, however, wider than that of the unoccupied tag and there is a possibility that La^{3+} occupied and non-occupied LBTB were eluted at almost the same time. It was evident, however, that the peptide behaved differently when exposed to lower levels of the lanthanide (Figure 4.8) than when flooded with an excess (Figure 4.7). The instability of the peptide may provide an explanation for the absence of metal binding. Small peptides containing DG in the sequence are unstable due to the formation of cyclic imide intermediate, which then hydrolyses (ring opening) into a linear peptide. This ring opening either forms the original correct peptide or proceeds to form the iso-Asp analog. It is probable that this ring opening prevents binding of the lanthanide, which explains the presence of two peaks in GFC studies herein.

Once this tag is conjugated to a protein this would no longer be an issue due to the increase in size of the overall structure.

All data obtained from ITC experiments indicated that metal chloride binding to each isolated peptide was an endothermic process (Figures 4.9 - 4.10) and not a favourable process as we had expected. Interestingly, only titrations using lanthanum produced the desired sigmoid curve. Binding affinity (Kd) of La^{3+} for LBTA and LBTB were calculated from the ITC data as 36.5 μM and 40.5 μM respectively (Figures 4.9A - 4.10A). Gd^{3+} appeared to bind to each of the isolated peptides but binding affinities were so weak that a model could not be fitted and affinities could not be calculated as a result. All other ITC titrations (Tb^{3+}, Gd^{3+} Yb^{3+} Mg^{2+}, Ca^{2+}) produced erratic, low enthalpy peaks during ITC (Figures 4.9B and 4.10B), which probably suggests that there was no binding.

Although a number of buffers, including; 20 mM Tris-Cl, 0.1 M NaCl (pH 7.5), Tris-Cl (10 mM, pH 8.0) and MES (100 mM, pH 6.0) were assessed before 100 mM Na-HEPES, 150 mM KCl (pH 7.0) was finally settled on for ITC titrations, the weak or absent binding of metal ions may suggest that conditions were not adequate to support binding. The use of the wrong buffer or titrations run at a temperature other than what is optimum can affect binding. Unfortunately, the system could not be set at temperatures lower than 25°C and ITC could not be performed on Ap_4A-H-LBTB as the protein became unstable over this length of time at this temperature and precipitated. No precipitation was visible for titrations using each peptide alone.

For each LBT peptide, data suggested that 2 binding sites were available for La^{3+}. For ITC, one is rarely able to obtain data to fit a stoichiometry of 1, exactly (Fisher

and Singh, 1995). However, if true, this would confirm the theory suggested in section 4.2 for GFC studies. The premise suggests that each tag has the ability to bind more than one lanthanide ion or if only one ion occupies the site a region is exposed that can enable interaction with additional peptides (that may or may not have an occupied site). It is possible that the stoichiometry of lanthanide binding to the tag is different for each metal ion. ESI-MS was employed as a tool for evaluating stoichiometry of metal ion (Ca^{2+}, Mg^{2+}, Gd^{3+}, Yb^{3+}, Tb^{3+} and La^{3+}) binding to each isolate peptide (ESI-MS performed by M. Prescot). Unfortunately, results were inconclusive as the background resulting from the buffer masked suspect 2+-charged and 3+-charged species.

Results herein clearly provide an indication of lanthanide binding. Binding studies using gel filtration chromatography showed that each tag displayed very different behaviour when exposed to a variety of metals. It was impossible at this point to identify the most suitable tag as both had their advantages. LBTA is the more hydrophilic of the two tags and more stable than LBTB in metal-binding studies using GFC. However, LBTB may be capable of binding all of the metals used in this study and when fused with the model protein may have benefits to solubility. These investigations rarely take us beyond pointing out the behavioural differences of the two tags and further investigations are required to determine the metal binding affinity of each tag.

4.10.2: Impact of LBTB tag on cloning, expression and purification of Ap_4A-H-LBTB

Now that a plasmid (pET28a) has been engineered, any target sequence can be cloned for expression as an LBTB fusion. Cloning via the use of this expression plasmid would be no different to any other plasmid of the pET vector series. Studies herein confirm that LBTB tag addition does not have a noticeable effect on cell growth (Figures 4.16 - 4.17) or on recombinant protein expression at 20°C, 30°C or 37°C when comparing yields of Ap_4A-H-LBTB to the non-tagged control, Ap_4A-H (Figures 4.15). In fact, both targets were expressed at very high concentrations of soluble protein. Initially, a Bradford assay was the chosen method in an attempt to quantify Ap_4A-H concentration. Values obtained were a great deal higher than expected. After the first step of purification concentrations as high as 100 mg/ml were suggested when using this assay. Protein quantification using the BCA method also resulted in very high estimations. Protein concentrations calculated using a theoretical extinction coefficient (22710 for Ap_4A-H and 28210 for Ap_4A-H-LBTB) (determined by ProtParam) were more than 50% lower than those determined using Bradford reagent.

Despite fears, the tag had no effect on purification using IMAC. Both proteins were successfully purified to more than 90%. No problems were encountered during GFC purification either. Furthermore, investigations suggest that the addition of the LBTB tag actually increased solubility and stability of Ap_4A-H. Results obtained from the luciferase-based bioluminescence assay suggest that the tag appeared to increase the specific activity 2.4 X compared to the control protein. This data also correlated with analytical GFC traces of each protein (after 48 h at 4°C), which

suggested that the control protein without the tag experienced more protein aggregation. More obvious was the extent of precipitation of protein samples when stored for long periods at 4°C or when dialysing samples for ITC. It was very clear that the non-tagged protein was less soluble/stable in the buffers used. The attachment of the LBTB region to Ap_4A-H appears to make the protein more soluble, perhaps experienced as the LBTB tag provides a more hydrophilic C-terminal region. The impact of LBTB addition on solubility using buffers of crystallisation suites was scrutinized (Section 4.3). The pHClear screen revealed that both LBTB-tagged and non-tagged forms were most soluble at pH range pH 7.0 – pH 8.0, even up to pH 9.0. It appears that the tag had differing effects with regards to solubility. No general consensus could be drawn to suggest how the LBTB tag affected solubility. It appears that the buffer solution determines whether the tag will result in an increase or a decrease in the solubility of the protein.

4.10.3: Impact of tag on success of crystallisation

For the most part, the addition of the tag had little influence upon the fate of the protein drop during the crystallisation studies herein. There were, however, several buffers for which the attachment of the LBTB region to Ap_4A-H led to an increase or decrease in protein solubility (Table 4.2). Moreover, a suspect growth was noticed in a drop containing LBTB-tagged protein while the drop containing Ap_4A-H control protein was comprised of granular precipitate (Figure 4.22). Interestingly, the buffer in which the 'growth' was noticed [100 mM sodium acetate, pH 4.6, 12 % PEG 4000, 25°C, sitting drop vapour diffusion] was similar to the conditions used in the crystallisation of the *C. elegans* homologue (0.2 M ammonium acetate and 100 mM sodium citrate, pH 5.6, 32% (w/v) PEG 4000, 25°C, sitting drop vapour diffusion).

This 'growth' had the appearance of an oily crystal form, and if so, could be used for streak seeding to promote crystallisation in future trials. Although this was exciting, the length of time it took for the growth to appear (6 weeks) is somewhat concerning and there was some evidence of drying out in the drop (Figure 4.22). With hindsight, Hampton's IZIT dye or methylene green could have been used in investigations. Dyes such as these are able to stain protein but not salt and would have identified the composition of the growth. Confirmation would have established if the LBTB region was advantageous for crystallisation in this case. A characteristic of this sort is highly favourable for the universal usage of this tag in crystallographic protein studies. Future work would require the expression of a number of model proteins with and without the LBTB fusion to provide conclusive evidence to identify the advantages and disadvantages of the proposed tag. Model proteins with known crystallisation conditions would be the most ideal candidates.

The high degree of precipitation experienced in the screen based upon the crystallisation conditions (pH 5.6, PEG 4000, ammonium acetate) used in the development of X-ray diffracting crystals of the *C. elegans* homologue were most likely the result of protein aggregation. This screen was set up using a protein preparation that had been stored at 4°C for almost 3 days. Analytical gel filtration chromatography of protein preparations of Ap_4A-H and Ap_4A-H-LBTB stored at 4°C for the same duration confirmed that both samples had experienced protein aggregation, with the former experiencing the most.

It was thought that the addition of La^{3+} to the Ap_4A-H-LBTB preparation would maintain the flexible LBTB region in a more rigid conformation, thereby facilitating crystallisation. Unfortunately this was not the case and the addition of this

lanthanide resulted in an increase in the extent of precipitation in some of the crystallisation solutions. Addition of the trivalent metals at high concentrations resulted in the loss of soluble protein (Figure 4.23), consistent with results in previous studies (Lim and Franklin, 2004). This is possibly due to the formation of metal hydroxides resulting in a pH change, causing precipitation of protein.

Crystallisation will require saturation of binding sites with lanthanide ions. Moreover, the loss of protein as a result of precipitation upon addition of lanthanide could prove problematic as high protein concentrations are essential in crystallographic studies (Blundell *et al.*, 2001).

4.10.4: Future work

Despite attempts, crystallisation was not possible for Ap_4A-H or Ap_4A-H-LBTB so the phasing power of such a tag has not yet been explored. Ideally, alternative target proteins with known structures solved by X-ray crystallography (e.g. the *C. elegans* homologue) could be used for LBTB-tagging so that the universal effects on crystallisation and the phasing power of the tag can be determined. Another approach to move investigations forward would be the removal of N-terminal His-tag prior to crystallisation. Additionally, crystallisation trials of the isolated LBTB tag should be performed. Attempts have already been made to crystallise the isolated LBTA peptide but saturation could not be reached even when using very high concentrations, as the peptide is very hydrophilic. Crystallisation of LBTB may prove to be more successful because the LBTB tag is more hydrophobic due to the presence of the tryptophan in the amino acid sequence. LBTB crystallisation trials could be set up with a number of metals, but first extensive studies must be performed to establish the metal binding abilities of this tag in regards to a greater

variety of metals including all of the lanthanide series. Once a metal is found that binds to the tag with high affinity that does not affect the stability of the peptide this will be much more straightforward.

5: General Discussion.

In this post-genomics era greater demands are being placed upon recombinant protein production, expecting the expression of large numbers of proteins in faster timeframes. Currently, strategies are being developed to standardise and automate each process in structural genomics (Dauter, 2002). Studies herein clearly demonstrate the inadequacies of the pragmatic approach of traditional methods. The most concerning of all is the time squandered when embarking on such a path. Many labs are now opting for high throughput expression of large numbers of constructs in the hope that one will be expressed to high levels so that they are able to follow through to crystallisation (Endo and Sawaski, 2003; Stevens, 2000. Had this laboratory been able to employ a high-throughput approach for the expression of GPI-PLD, such as those reviewed in (Stevens, 2000) and (Hunt *et al.*, 2005), it is highly likely that studies would have progressed further. A robot system like the Piccolo robot system (Wollerton *et al.*, 2006) has the capacity for the exploration of hundreds of different combinations of the variables in a prokaryotic expression system. Perhaps this may have identified an ideal condition for the expression of active GPI-PLD, where the low throughput approach herein proved to be unsuccessful for this difficult protein target.

Expression of eukaryotic proteins in *E. coli* is typically associated with lack of solubility. Microbial targets often have higher success rates. It is no surprise that larger proteins are often expressed at low levels and, more concerning, as inclusion bodies. Proteins with higher molecular weights have more opportunity to misfold and aggregate than smaller ones. The pI has also emerged as an important factor in protein solubility for heterologous expression (Mehlin *et al.*, 2006). A more basic

isoelectric point (pI), lack of homology to native *E. coli* proteins and greater protein disorder (segmental analysis, SEG) have all been associated with difficulties in recombinant expression. In light of these difficulties, cell-free systems are becoming increasingly popular. These *in vitro* systems offer several advantages including, the possibility to express toxic gene products and constructs that would otherwise be subjected to proteolytic degradation in an *in vivo* system like *E. coli*. Generally, such systems have higher success rates in regards to solubility and have also been used in the expression of difficult membrane proteins. This system offers a very rapid approach to protein synthesis as expression can be directed from the PCR product. Moreover, these cell free systems lend themselves to automation for large-scale protein production and are currently being used in high-throughput proteomics (Langlais *et al.*, 2007; Yokoyama, 2007).

A sequence-based protein solubility evaluator (PROSO) has recently been developed to provide a guide for choosing the clone with the best chance of soluble expression in *E.coli* (Smialowski *et al.*, 2007). This solubility evaluator is based upon sequence and takes into consideration the subtle differences between soluble proteins from TargetDB and PDB and notoriously insoluble proteins from TargetDB and literature mining. Predictions are also based upon a protein's predicted physiochemical properties and structural properties. It has an overall prediction accuracy of 72%. Furthermore predicted solubility was shown to correlate very well with experimental results on protein solubility upon expression in *E.coli* and, with hindsight, proved to be successful in the prediction for GPI-PLD.

A sequence-based crystallisability evaluator (SECRET) is also available (Smialowski *et al.*, 2006). This sequence-based prediction of protein crystallisability can be used to support the target selection process in structural genomics. The prediction is based on data obtained from three groups of sequences derived from the Protein Data Bank (PDB) corresponding to: a) proteins with X-ray structures, which are hence crystallisable; b) proteins with NMR (Nuclear Magnetic Resonance) structures that also have an X-ray structure, or bear a high sequence similarity to proteins with X-ray structures, and; c) proteins with NMR structures only. Both predictors provide promising and valuable tools for more efficient target selection in structural genomics.

Despite the advances in expression plasmids and strains for expression of difficult proteins, some proteins still express as insoluble aggregates (Heinemann *et al.*, 2003). Recombinant GPI-PLD has proven to be no exception despite attempts at optimisation of expression and the addition of soluble fusion partners to promote solubility (Chapter 3). The only desirable route forward would be the substitution of the surface residues that are predicted to be glycosylated in the native protein (Figure 3.36). Surface residues were substituted by mutagenesis to increase solubility and to aid crystallisation of MAPKAP kinase 2 (Argiriadi *et al.*, 2009). Ap4A-H was successfully expressed at high levels and purified using IMAC for crystallisation trials (Chapter 4). The addition of the LBT region to Ap4A-H had no effect on expression and purification, which made it an ideal candidate for LBT studies.

The success of the proposed LBT tag depends greatly upon the impact it has on the expression, purification and crystallisation of a given protein. In theory, the addition

of a tag of this size should not have any serious repercussions (Arnau *et al.*, 2006). However, in some instances the addition of an affinity tag can alter the conformation of a recombinant protein (Chant *et al.*, 2005), can affect binding properties (Goel *et al.*, 2000), and inhibit (Cadel *et al.*, 2004) or alter activity and/or post-translational modifications (Fonda *et al.*, 2002; Kim *et al.*, 2001). As mentioned earlier, affinity tags can exhibit undesirable flexibility so their removal prior to crystallisation is often required (Smyth *et al.*, 2003). This makes this tag very attractive compared to those already available due to its small size and compact conformation upon metal binding (Figures 4.3, 4.4, 4.6 - 4.8). Furthermore, observations from crystallisation trials using the LBTB tag as a fusion have aroused the idea that this tag may have the benefit of aiding the crystallisation of proteins, perhaps by promoting crystal packing. The structure of *Geobacter sulfurreducens* OmcF was solved as a Strep II-fusion (Lukat *et al.*, 2008). In the crystal, packing interactions in one dimension were exclusively mediated through the Strep-tag II sequence.

There are many considerations to be taken into account for expression with the LBTB tag. It is highly likely that the protein of interest will direct the choice of fusion tag for phasing. Moreover, the positioning of the LBT tag at the C-terminus may not be suitable for the tag to adopt its metal-binding conformation. It may be masked within the protein. In this situation the tag could be used as an N-terminal fusion or a linker could be added.

Results have demonstrated the metal-binding ability of both LBTA and LBTB. The two tags are highly similar in sequence but have very different behaviour with regards to metal binding (detailed in chapter 4) and could provide options depending

upon the requirements of the protein of interest. It remains unclear as to whether the proposed tags could be used as a universal approach to solving the phasing problem. However, other tags have demonstrated the ability to assist in protein phasing. Silvaggi *et al* (2007) provided the first proof of concept of the applicability of double-LBT in X-ray crystallography. Although it may appear that the ability to bind more than one Ln^{3+} is advantageous, it has the potential for creating problems for expression, solubility, purification and/or crystallisation of the target protein. Most of the LBT tags to date are larger than LBTA and LBTB comprising of between 15-25 amino acids. In addition to this, the use of such a tag in providing phasing information for the elucidation of a protein structure has only been demonstrated for one protein (Silvaggi *et al.*, 2007) and may not be universally applicable. This being said, a collection of complementary fusion tags, including LBTA and LBTB, could provide options for a target protein.

Lanthanides can also be in used in the execution of complex NMR experiments as a result of their paramagnetic variations. Such paramagnetic disparity (Dy, Tb, Tm are classified as highly paramagnetic, Er and Yb moderately paramagnetic and Eu, Ce and Sm have weak paramagnetic properties) means that the extent of knockout of NMR spectra can be varied. This allows spectra to be allocated to different regions of the protein, opening up the possibilities of solving the structures of increasingly larger proteins. This means that the LBT fusion could also be used in NMR when crystals cannot be obtained.

There is also the prospect of the tag being used as a purification tag in IMAC by using its metal binding ability. With improvements this tag could enable

standardisation, and perhaps even automation, of purification and phasing leading to a new form of structure determination without the constraints of the other methods. This tag with other could pave the way for a new high-throughput approach to protein crystallography and even provide options for the method of structure determination.

References

Abdelghany HM, Gasmi L, Cartwright JL, Bailey S, Raffery JB, McLennan AG (2001): Cloning, charcterisation and crystallisation of a diadenosine 5',5''',P1,P4-tetraphosphate pyrophosphohydrolase from *Caenorhabditis elegens*. Biochim. Biophys. Acta 1550.

Agatep R, Kirkpatrick RD, Parchaliuk DL, Woods RA, Gietz RD (1998): Transformation of saccharomyces cerevisiae by the lithium acetate/single stranded DNA/polyethylene glycol prptocol. Technical Tips Online 3:133-173.

Agathos SN (1991): Production scale insect cell culture. Biotechnol. Adv. 9:51-68.

Arechaga I, Miroux B, Karrasch S, Huijbregts R, Kruijiff B, Runswick MJ, Walker JE (2000): Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the β subunit of f1f0 ATP synthase. FEBS Lett 482:215-219.

Argiriadi MA, Sousa S, Banach D, Marcotte D, Xiang T, Tomlinson MJ, Demers M, Harris C, Kwak S, Hardman J, Pietras M, Quinn L, DiMauro J, Ni B, Mankovich J, Borhani DW, Talanian RV, Sadhukhan R (2009): Rational mutagenesis to support structure-based drug design: MAPKAP kinase 2 as a case study. BMC Structural Biology 9:16-28.

Arnau J, Lauritzen C, Petersen GE, Pederson J (2006): Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expression and Purification 48:1-13.

Bailey S, Sedelnikova SE, Baker PJ, McLennan AG, Rafferty JB (2002a): The crystal structure of diadenosine tetraphosphate hydrolase from *Caenorhabditis elegans* in free and binary forms. Structure 10:589-600.

Bailey S, Sedelnikova SE, Blackburn GM, Abdelghany HM, McLennan AG, Rafferty JB (2002b): Crystallization of a complex of *Caenorhabditis elegans* diadenosine tetraphosphate hydrolase and a non-hydrolysable substrate analogue, APPCH2PPA. Acta. Cryst. D58:526-528.

Baneyx F (1999): Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol 10:411-421.

Bendtsen JD, Nielsen H, Heijne GV, Brunak S (2004): Improved prediction of signal peptides: SignalIP3.0. J. Mol. biol. 340:783-795.

Blackwell JR, Horgan R (1991): A novel strategy for production of a highly expressed recombinant protein in an active form FEBS Lett 295.

Blundell TL, Harren J, Abell C (2001): High-throughput crystallography for lead discovery in drug design. Nat. Rev. Drug Discov. 1:45-54.

Bordier C (1981): Phase separation of integral membrane proteins in Triton X-114 solution. J. Biol. chem. 256:1604-1607.

Bradford MM (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72:248-254.

Brown D, Waneck G (1992): Glycosyl-phosphatidylinositol-anchored membrane proteins. J. Am. Soc. Nephrol 3:895-906.

Brunner G, Metz C, Nguyen H, Gabrilove J, Patel S, Davitz M, Rifkin D, Wilson E (1994): An endogenous glycosylphosphaitdylinositol-specific phospholipase D releases fibroblastic growth factor-heparan sulphate proteoglycan complexes from human bone marrow cultures. Blood 83:2115-2125.

Cadel S, Gouzy-Darmon C, Peters S, Piesse C, Pham VL, Bienfeld MC, Cohen P, Foulon T (2004): Expression and purification of rat recombinant aminopeptidase B secreted from baculovirus-infected insect cells. Protein Expr. Purif. 36:19-30.

Carvin D, Islam SA, Sternberg MJE, Blundell TL (2001): The preparation of heavy atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering. Int Tables Crystallogr F:247-255.

Chant A, Kraemer-Pecore CM, Watkin R, Kneale GG (2005): Attachment of a histidine-tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein area causes a conformational change at the DNA-binding site. Protein Expr. Purif. 39:152-159.

Chen YJ, Pornillos O, Lieu S, Ma C (2007): X-ray structure of emre supports dual topology model. Proc Natl Acad Sci USA 104:18999-19004.

Cheng H, Jiang N (2006): Extremely rapid extraction of DNA from bacteria and yeasts. Biotech. Letters 28:55-59.

Chicku H, Kawai A, Ishibashi T, Takehara M, Yanai T, Mizukami F, Sakaguchi K (2006): A novel protein refolding method using a zeolite. Anal. Biochem. **348**:307-314.

Choi BK, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH (2003): Use of combinatorial genetic libraries to huminize N-linked glycosylation in yeast *Pichia pastoris*. Proc Natl Acad Sci USA 100:5022-5027.

Chou CP (2007): Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Appl Microbiol Biotech 76:521-532.

Civenni G, Butikofer P, Stradelmann B, Brodbeck U (1999): In vitro phosphorylation of purified gylcosylphosphatidylinositol-specific phospholipase D. Biol. Chem. 380:585-588.

Congreve M, Murray CW, Blundell TL (2005): Structural biology and drug discovery. Drug Discovery Today 10:895-907.

Cross GA (1990): Glycolipid anchoring of plasma membrane proteins. Annu Rev Biochem 6:1-39.

Dai X, Chen Q, Lian M, Zhou Y, Zhou M, Lu S, Chen Y, Luo J, Gu X, Jiang Y, Luo M, Zheng X (2005): Systematic high-yield production of human secreted proteins in *Escherichia coli*. Biochem. Biophys. Res. Comm. 332:593-601.

Daugherty DL, Rozema D, Hanson PE, Gellman SH (1998): Artificial chaperone-assisted refolding of citrate synthase. J. Biol. chem. 273:33961-33971.

Dauter Z (2002): New approaches to high-throughput phasing. Curr Opin Structural Biol 12:674-678.

David RJ, Matias AA, Carmen S, Isabel V (1997): Glycosyl-phosphatidylinositol-phospholipase type D: A possible candidate for the generation of second messengers. Biochem. Biophys. Res. Comm. 223:432-437.

Davitz MA, Hereld D, Shak S, Krakow J, Englund PT, Nussenweig VA (1987): Glycan-phosphatidylinositol-specific phospholipase D in humans. Science 238:81-84.

Deeg M, Biereman E, Cheung M (2001a): GPI-specific phospholipase D associates with an apoA-I and apoA-IV-containing complex. J. Lipid Res. 42:442-451.

Deeg M, Bowen R (1999): Midportion antibodies stimulate glycosylphosphatidylinositol-specific phospholipase D activity. Arch. Biochem. Biophys. 370:278-284.

Deeg MA, Bowen RF, Williams MD, Olson LK, Kirk EA, Leboeuf RC (2001b): Increased expression of GPI-specific phospholipase D in mouse models of type-1 diabetes. Am. J. Physiol. Endocrinol. Metab. 288:E147-E154.

Deeg MA, Davitz MA (1994): Structure and function of the glycosylphosphatidylinositol-specific phospholipase D. In Liiscovitch M (ed): "Signal activated phospholipases." Austin TX: R G Landes, pp 125-138.

Demain AL, Vaishnav P (2009): Production of recombinant proteins by microbes and higher organisms. Biotechnological Advances 27:297-306.

Deng J, Hoylaerts M, Broe MD, Hoof V (1996): Hydrolysis of membrane bound liver alkaline phosphatase by GPI-PLD requires bile salts. Am. J. Physiol. 271:G655-G663.

Derewenda ZS (2004): The use of recombinant methods and molecular engineering in protein crystallization. Methods 34:354-363.

Dvoretsky A, Gaponenko V, Rosevear PR (2002): Derivation of structural restraints using a thiol-reactive chelator. FEBS Lett 528:189-192.

Ellman GL (1956): A colorimetric method for determining low concentrations of mercaptans. Arch. Biochem. Biophys. 74:443-450.

Endo Y, Sawaski T (2003): High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. Biotechnol. Adv. 21:695-713.

Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008): Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. Proc Natl Acad Sci USA 105:6278-6283.

Ferguson MA, Homans SW, Dwek RA, Rademacher TW (1988a): Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. Science 239:753-759.

Ferguson MA, Williams AF (1988b): Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu Rev Biochem 57:285-320.

Ferguson MAJ (1999): The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J. Cell. Sci 112:2799-2809.

Fisher HF, Singh N (1995): Calorimetric methods for interpreting protein-ligand interactions. Methods Enzymol 259:194-221.

Fonda I, Kenig M, Garberc-Porekar AV, Pristovaek P, Menart V (2002): Attachment of histidine tag to tumor necrosis factor alpha drastically changes its properties. Sci. World J. 15:1312-1325.

Fox JA, Soliz NM, Saitiel AR (1987): Purification of a phosphatidylinositol-glycan specific phospholipase c from liver plasma membranes – a possible target of insulin action. Proc Natl Acad Sci USA 84:2663-2667.

Fox JD, Waugh DS (2003): Maltose binding protein as a solubility enhancer. Methods Mol. Biol. 205:99-117.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003): Expasy: The proteomics server for in-depth protein knowledge and analysis Nucleic Acids Res 31:3784-3788.

Godzick A, Sander C (1989): Conservation of residue interactions in a family of Ca-binding proteins. Protein Eng 2:589-596.

Goel A, Colcher D, Koo JS, Booth BJ, Pavlinkova G, Batra SK (2000): Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain fv construct. Biochim. Biophys. Acta 1523:13-20.

Gregory P, Kraemer E, Zurcher G, Genitenetta R, Rohrbach V, Brodbeck U, Andres A, Ziemiecki A, Butikofer P (2005): GPI-specific phospholipase D (GPI-PLD) is expressed during mouse development and is localized to the extracellular matrix of the developin mouse skeleton. Bone. 37:139-147.

Gronenborn AM, Filpula DD, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM (1991): A novel highly stable fold of the immunoglobulin binding domain of streptococcal protein G. Science 523:657-661.

Gupta R, Jung E, Brunak S (2004): Prediction of N-glycosylation sites in human proteins. In preparation.

Gustafsson C, Govindarajan S, Minshull J (2004): Codon bias and heterologous protein expression. Trends Biotechnol 22:346-356.

Haberz P, Rodriguez-Castaneda F, Junker J, Baker S, Leonov A, Griesinger C (2006): Two new chiral EDTA-based metal chelates for weak alignment of proteins in solution. Org Lett 8:1275-1278.

Hannig G, Makrides SC (1998): Strategies for optimizing heterologous protein expression in *Escherichia coli*. Trends Biotechnol 16:54-60.

Hartley JL, Temple GF, Brasch MA (2000): DNA cloning using in vitro site-specific recombination. Genome Res. 10:1788-1795.

He X, Hannocks M, Hampson I, Brunner G (2002): Gpi-specific phospholipase D mRNA expression in tumour cells of different malignancy. Clin. Exp. Metastasis 19:291-299.

Heinemann U, Bussow K, Mueller U, Umbach P (2003): Facillities and methods for high-throughput crystal structural analysis of human proteins. Acc. Chem. Res. 36:157-163.

Hellar M, Butikofer P, Brodbeck U (1994): Generation by limited proteolysis of a catalytically active 39-kDa protein from the 115-kDa form of phosphatidylinositol-glycan-specific phospholipase D from bovine serum. Eur. J. Biochem. 224:823-833.

Heller M, Bieri S, Brodbeck U (1992): A novel form of glycosylphosphatidylinositol-anchor converting activity with a specificity of a phospholipase D in mammalian liver membranes. Biochim. Biophys. Acta 1109:106-116.

Hendrickson WA, Horton JR, LeMaster DM (1990): Selenomethionyl proteins produced fro analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimntional structure. EMBO J 9:1665-1672.

Heukeshoven J, Dernick R (1985): Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. Electrophoresis 6:103-112.

Hitzeman RA, Chen CY, Dowbenko DJ, Renz ME, Lui C, Pai R, Simpson NJ, Kohr WJ, Singh A, Chisholm V, Hamilton R, Chang CN (1990): Use of heterologous and homologous signal sequences for secretion of heterologous proteins from yeast. Methods in Enzymology 421-440.

Hoener M, Golli R, Brodbeck U (1993): Gylcosyl-phosphatidylinositol-specific phospholipase D: Interaction with and stimulation by apolipoprotein A-I. FEBS Lett 327:203-206.

Hoener M, Liao J, Brodbeck U (1994): Conversion of the amphiphillic 115-kDa form of glycosylphosphatidylinositol-specific phospholipase D to an active, hydrophillic 47-kDa, in biological. In F OKJA (ed): "In biological membranes: Structure, biogenesis and dynamics." Heidelberg, Berlin: Springer-Verlag, pp 71-78.

Hoener MC, S S, Brodbeck U (1990): Isolation and characterization of a phosphatidylinositol-glycan-anchor-specific phospholipase D from bovine brain. Eur. J. Biochem. 190:593-601.

Hooper NM (1997): Glycosyl-phosphatidylinositol anchored membrane enzymes. Clinica Chimica Acta 266:3-12.

Huang KS, Li S, Fung WC, Hulmes JD, Reik L, Pan Y (1990): Purification and characterization of glycosyl-phosphatidylinositol-specific phospholipase D. J. Biol. chem. 265:17738-17745.

Hunke S, Betton JM (2003): Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia*. Mol. Microbiol. 50:1579-1589.

Hunt I (2005): From gene to protein: A review of new and enabling technologies for multi-parallel protein expression. Protein Expression and Purification 40:1-22.

Huth JR, Bewley CA, Jackson BM, Hinnenbusch AG, Clore GM, Gronenborn AM (1997): Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. Protein Science 6:2359-2365.

Ikegami T, Verdier L, Sakhaii P, Grimme S, Pescatore B, Fiebig KM, Griesinger C (2004): Novel technique for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. J Biomol NMR 29:339-349.

Iwahara J, Clore GM (2006): Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature 440:1227-1230.

Jian-Hua T, Wang-Jiao H, He H, Chao-Chao T, Qiong D, Kai-Jiia W, Xian0Yu Y, Xu-Jin Z (2009): Important roles of glycosylphosphatidylinositol (GPI) - specific phospholipase D and some GPI-anchored proteins in the pathogenesis of hepatocellular carcinoma. Clinical. Biochemistry 42:400-407.

Keizers PH, Desreux JF, Overhand M, Ubbink M (2007): Increased paramagnetic effect of a lanthanide protein probe by two-point attachment. J Am Chem Soc 129:9292-9293.

Kern M, Wisiewski M, Cabell L, Audesirk G (2000): Inorganic lead and calcium interact positively in activation of calmodulin. Neurotoxicity 21:353-363.

Kim KM, Yi EC, Baker D, Zhang KY (2001): Post-translational modification of the N-terminal his-tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase. Acta. Cryst. 57:759-762.

Kirberger M, Wang X, Deng H, Yang W, Chen G, Yang JJ (2008): Statistical ananlysis of protein Ca^{2+}-binding sites. J. Biol. Inorg. Chem. 13:1169-1181.

Kretsinger RH (1976): Calcium-binding proteins. Ann. Rev. Biochem 45:239-266.

Kurland C, Gallant J (1996): Errors of heterologous protein expression. Curr Opin Biotechnol 7:489-493.

Kuruzinska MA, Bergh MLE, Jackson BJ (1987): Protein glycosylation in yeast. Ann. Rev. Biochem 56:915-944.

Laemmli UK, (1970): Cleavage of structural proteins during the assembly of the head of bacteriaophage T4. Nature 227: 680-685.

Langlais C, Guilleaume B, Wermke N, Scheuermann T, Ebert L, Labaer J, Korn B (2007): A systematic approach for testing expression of human full length proteins in cell-free expression systems. BMC Biotech 7:64-74.

LaVaillie ER, DiBlasio EA, Kovacic S, Grant K, Schendel PF, McCoy JM (1993): A thioredoxin gene fusion expression systam that circumvents inclusion body formation in the *E coli* cytoplasm. Biotechnology 11:187-193.

Leboeuf RC, Caldwelll M, Gue Y, Metz C, Davitz MA, Olson LK, Deeg MA (1998): Mouse glycosylphosphatidylinositol-specific phospholipase D (GPLD1) characterization. Mamm. Genome. 9:714-719.

Leonov A, Voigt B, Rodriguez-Cataneda F, Sakhaii P, Griesinger C (2005): Convenient synthesis of multifunctional EDTA-based chiral metal chelates substituted with an s-mesylcysteine. Chemistry 11:3342-3348.

Li J, Low MG (1999): Studies of the role of the integrin ef-hand, Ca^{2+} binding sites in glycosylphospatidylinositol-specific phospholipase D: Reduced expression following mutagenesis of residues predicted to bind Ca^{2+}. Archives of Biochemistry and Biophysics 361:142-148.

Lierheimer R, Kunz B, Vogt L, Savoca R, Brodbeck U, Sonderegger P (1997): The neuronal cell-adhesion molecule axionin-1 is specifically released by an endogenous glycosylphosphatidyl-specific phospholipase. Eur. J. Biochem. 243:502-510.

Lim S, Franklin SJ (2004): Lanthanide-binding peptides and the enzymes that might have been. Cell. Mol. Life Sci. 61:2184-2188.

Low M (1999): GPI-anchored biomolecules- an overview, in GPI-anchored membrane proteins and carbohydrates. In Hoessli DC, Ilangumaran S (eds) Austin, TX: Landes Company, pp 1-14.

Low MG (1989): The glycosyl-phosphatidylinositol anchor of membrane proteins. Biochim. Biophys. Acta 988:427-454.

Low MG, Huang KS (1991): Factors affecting the ability of glycosylphosphatidylinositol-specific phospholipase D to degrade membrane anchors of cell surface proteins. Biochem. J. 279:483-493.

Low MG, Prassad A (1988): A phospholipase D specific for the phosphatidylinositol anchor of cell surface proteins is abundant in plasma. Proc. Natl. Acad. Sci. USA 85:980-984.

Lui H, Zhou X, Zhang Y (2006): A comparative investigation on different refolding strategies of recombinant human tissue-type plasminogen activator derivative. Biotech. Letters 28:475-463.

Lukat P, Hoffmann M, Einsle O (2008): Crystal packing of the c6-type cytochrome ompCF from *Geobacter sulfurreducens* is mediated by an n-terminal strep-tag II. Acta. Cryst. D64:919-926.

Ma C, Opella SJ (2000): Lanthanide ions bind specifically to an added "EF-hand" And orient a membrane protein in micelles for solution NMR spectroscopy. J Magn Reson 146:381-384.

Maeda Y, Koga H, Yamada H (1995): Effective renaturation of reduced lysozyme by gentle removal of urea. Protein Eng 8:201-205.

Makrides SC (1996): Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev 60:512-538.

Mann K, Hepworth M, Raikwar N, Deeg M, Sevelever D (2004): Effect of gylcosylphosphatidylinositol (GPI)-phospholipase D overexpression on GPI metabolism. Biochem. J. 378:641-648.

Martin LJ, Hahnke MJ, Nitz M, Wohnert J, Sivaggi NR, Allen KN, Schwalbe H, Imperiali B (2007): Double-lanthanide-binding tags: Design and photophysical properties, and NMR applications. J Am Chem Soc 129:7106-7113.

McLennan AG (2006): The Nudix hydrolase superfamily. Cell. Mol. Life Sci 63: 123-143.

McPhalen CA, Strynadka NC, James MN (1991): Calcium-binding sites in proteins: A structural perspective Advan. Protein. Chem. 42:77-144.

McTigue MA, Williams DR, Rainer JA (1995): Crystal structures of a schistosomal drug and vaccine target: Glutatione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziqantel. J Mol. Biol. 246:21-27.

Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb MH, Haji L, Kim D, Liu C, Mueller N, Myler PJ, Reddy JT, Sampson JN, Subramanian E, Voorhis WCV, Worthey E, Zucker F, Hol WGJ (2006): Heterologous expression of proteins from *Plasmodium falciparum*: Results from 1000 genes. Mol Biochem Parasitol 148:144-160.

Metz C, Brunner G, Choi-Muira N, Nguyen H, Gabrilove J, Caras I, Altzuler N, Rifkin D, Davitz M (1994): Release from GPI-anchored proteins by a cell assosiated GPI-specific phospholipase D. EMBO J 3:1741-1751.

Miroux B, Walker JE (1996): Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane and globular proteins at high levels. J. Mol. biol. 260:289-298.

Nitz M, Sherawat M, Franz KJ, Peisach E, Allen KN (2004): Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. Angew Chem Int Ed 43:36-3685.

Nosjean O, Briolay A, Roux B (1997): Mammalian GPI proteins: Sorting, membrane residence and functions. Biochim. Biophys. Acta 1331:153-186.

O'Brien K, Pineda C, Chiu W, Bowen R, Deeg M (1999): Glycosylphosphatidylinositol-specific phospholipase D is expressed by macrophages in human atherosclerosis and colocalizes with oxidation epitopes. Circulation. 99: 2876-2882

Paulick MG, Bertozzi CR (2008): The glycosylphosphatidylinositol anchor: A complex membrane-anchor structure for proteins. Biochemistry 47:6991-7000.

Pidcock E, Moore GR (2001): Structural characteristics of protein binding sites for calcium and lanthanide ions. J. Biol. Inorg. Chem. 6:479-489.

Prescot M, Milne AD, McLennan AG (1989): Characterization of the bis(5'-nucleosidyl) tetraphosphate pyrophosphohydrolase from encysted embryos of the brine shrimp artemia. Biochem. J. 259:831-838.

Prinz WA, Aslund F, Holmgren A, Beckwith J (1997): The role of thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. J. Biol. Chem. 272:15661-15667.

Prudencio M, Rohovec J, Peters JA, Tocheva E, Boulanger MJ, Murphy MEP, Hupkes HK, Kosters W, Impagliazzo A, Ubbink M (2004): A caged lanthanide complex as a paramagnetic shift agent for protein NMR. Chem. Eur. J. 10:3252-3260.

Puri NK, Crivelli E, Cardamone M (1992): Solubilization of growth hormone and other recombinant proteins from *Escherichia coli* by using a cationic surfactant. Biochem. J. 285:871-879.

Raikwar NS, Bowen RF, Deeg MA (2005): Mutating his29, his125, his138 or his158 abolishes glycosylphosphatidylinositol-specific phospholipase D catalytic activity. Biochem. J. 391:285-289.

Ramesh VDA, Nagaraja V (1994): Engineering hyperexpression of bacteriophage mu c protein by removal of secondary structure at the translation initiation region. Protein Eng 7:1053-1057.

Reid RE, Gariepy J, Saund AK, Hodges RS (1981): Calcium-induced protein folding. J Biol Chem 256:2742-2751.

Rhode H, Hoffmann-Blume E, Schilling K, Gehrhardt S, Gohlert A, Buttner A, Bublitz R, Cumme G, Horn A (1995): Glycosylphosphatidylinositol-alkaline phosphatase from calf intestine as substrate for glycosylphosphatidylinositol-specific-phospholipases - microassay using hydrophobic chromatography in pipet tips. Anal. Biochem. 231:99-108.

Rigden D (2004a): A distinct evolutionary relationship between GPI-specific phospholipase D and bacterial phosphatidylcholine-prefering phospholipase C. FEBS Lett 569:229-234.

Rigden DJ, Galperin MY (2004b): The DxDxDG motif for calcium binding: Multiple structural contexts and implications for evolution. J Mol. Biol. 343:971-984.

Rigden DJ, Jedrezejas MJ, Moroz OV, Galperin MY (2003b): Structural diversity of calcium-binding proteins in bacteria: Single-handed EF-hands? Trends Microbiol 11:295-297.

Rigden DJ, Jedrezejas MJ, My MYG (2003a): An extracellular calcium-binding domain in bacteria with a distant relationship to EF-hands. FEMS Microbiol Lett 221:103-110.

Rodriguez-Castaneda F, Haberz P, Leonov A, Griesinger C (2006): Paramagnetting tagging of diamagnetic proteins for solution NMR. Magn Reson Chem 44:S10-S16.

Rosenfeld J, Capdevielle J, Guillemot J-C, Ferrara P (1992): In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. Anal. Biochem. 203:173-179.

Rupp B, Kantardjieff KA (2008): Macromolecular crystallography: "Molecular biomethods handbook." Humana Press, pp 821-849.

Sawasaki T, Ogasawara T, Morishita R, Endo Y (2002): A cell-free protein sysnthesis system for high-throughput proteiomics. Proc Natl Acad Sci USA 99:14652-14657.

Scallon B, Fung WJC, Tsang TC, Li S, Kado-Fong H, Huang KS, Kochan JP (1991): Primary structure and functional activity of a phosphatidylinositol-glycan-specific phospholipase D. Science 252:446-448.

Schofield JN, Rademacher TW (2000): Structure and expression of the human glycosylphosphatidylinositol phospholipase D1 (GPI1). Gene. Biochim. Biophys. Acta 1494:189-194.

Seitz M, Oliver AG, Raymond KN (2007): The lanthanide contraction revisited. J Am Chem Soc 129:11153-11160.

Sevlever D, Chen R, Medof M (2000): Synthesis of GPI-anchor in PHN and GPI-linked proteins San Diago, CA, pp 199-220.

Silvaggi NR, Langdon J, Martin J, Schwalbe H, Imperiali B, Allen KN (2007): Double-lanthanide-binding tags for macromolecular crystallographic structure determination. J Am Chem Soc 129:7114-7120.

Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D (2007): Protein solubility: Sequence based prediction and experimental verification. Bioinformatics 19:2536-2542.

Smialowski P, Schmidt T, Cox J, Kirschner A, Frishman D (2006): Will my protein crystallize? A sequence-based predictor. Proteins 62:343-355.

Smith DB, Johnson KS (1988): Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67:31-40.

Smyth DR, Mrozkiewics MK, McGrath WJ, Listwan P, Kobe B (2003): Crystal structures of fusion proteins with large-affinity tags. Protein Sci. 12:1313-1322.

Smyth MS, Martin JHJ (2000): X ray crystallography. J Clin Pathol: Mol Pathol 53:8-14.

Snyder EE, Buoscio BW, Falke JJ (1990): Calcium(II) site specificity - effect of size and charge on metal-ion binding to an EF-hand like site. Biochemistry 29:3937-3943.

Spiess C, Beil A, Ehrmann M (1999): A temperature dependent switch from chaperone to protease in a widely conserved heat shock protein. Cell:339-347.

Stevens RC (2000): Design of high-throughput methods of protein production for structural biology. Structure 8:R177-R185.

Stieger S, Diem S, Jakob A, Brodbeck U (1991): Enzymatic properties of the phosphatidylinositol-glycan-specific phospholipase C form rat liver and phospatidyliositol-glycan-specific phospholipase D from rat serum. Eur. J. Biochem. 197:67-73.

Stone KL, McNulty DE, LoPresti ML, Crawford JM, Angelis DE, Williams KR (1992): "Elution and internal amino acid sequencing of PVDF-blotted proteins." Academic Press.

Stradelmann B, Butikofer P, Konig A, Brodbeck U (1997): The C-terminus of gylcosylphosphatidylinositol-specific phospholipase D is essential for biological activity. Biochim. Biophys. Acta 1355:107-113.

Studier FW, Moffatt BA (1986): Use of bacteriophage T7 RNA polymerase to direct high-level expression of cloned genes. J Mol. Biol. 189:113-130.

Su XC, Huber T, Dixon NE, Otting G (2006): Site-specific labelling of proteins with a rigid lanthanide-binding tag. ChemBioChem 7:1469-1474.

Su XC, McAndrew K, Huber T, Otting G (2008): Lanthanide binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. J Am Chem Soc 130:1681-1687.

Swarbrick JD, Buyya S, Gunawardana D, Gayler KR, McLennan AG, Gooley PR (2005): Structure and substrate-binding mechanism of human Ap_4A hydrolase. J Biol Chem 280:8471-8481.

Swietnicki W (2006): Folding aggregated proteins into functionally active forms. Curr Opin Biotechnol 17:367-372.

Tang C, Iwahara J, Clore GM (2006): Visualization of transient encounter complexes in protein-protein association. Nature. 444:383-386.

Tang C, Schwieters CD, Clore GM (2007): Open-to-closed transition in apo maltose binding protein observed by paramagnetic NMR. Nature. 449:1078-1082.

Taylor G (2003): The phase problem. Acta. Cryst. 59:1881-1890.

Terpe K (2003): Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotech 60:523-533.

Terpe K (2006): Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotech 72:211-223.

Toutant JP, Roberts WL, Murray NR, Rosenberry TL (1989): Conversion of human erythrocyte acetylcholinesterase from an amphiphilic to a hydrophilic form by phosphatidylinositol-specific phospholipase C and serum phospholipase D Eur. J. Biochem. 180:503-508.

Tsujioka H, Misumi Y, Takami N, Ikehara Y (1998): Post-translational modification of glycosylphosphatidylinositol (GPI)-specific phospholipase D and its activity in leavage of gpi-anchors. Biochem. Biophys. Res. Comm. 251:737-743.

Turkenburg JP, Dodson EJ (1996): Modern developments in molecular replacement. Current opinion in structural Biology 6:604-610.

Urakaze M, Kamitani T, Degasperi R (1992): Identification of a missing link in glycosylphosphatidylinositol anchor biosynthesis in mammalian cells. J Biol Chem 267:6459-6462.

Vinarov DA, Lytle BL, Peterson FC, Tyler EM (2004): Cell-free protein production and labelling for NMR-based structural proteiomics. Nat. Methods 1:149-153.

Vlasie MD, Comuzzi C, Nieuwendijk AMvd, Prudencio M, Overhand M, Ubbink M (2007): Long-range-distance NMR effect in a protein labeled with a lanthanide with a lanthanide-dota chelate. Chemistry 13:1715-1723.

Volkov AN, Worrall JAR, Holtzmann E, Ubbink M (2006): Solution structure and dynamics of the complex between cytochrome c peroxidase determined by paramagnetic NMR. Proc Natl Acad Sci USA 103:18945-18950.

Vyas MN, Jacobson BL, Quiocho FA (1989): The calcium binding site in the galactose chemoreceptor protein. Crystallography and metal-binding studies. J Biol Chem 264:20817-20821.

Walker E, Clarks A, MHewison, Rides J, Stewart P (2001): Functional expression, characterization, and purification of the catalytic domain of human 11--hydroxysteroid dehydrogenase type 1. . J. Biol. chem. 276:21343–21350.

Waneck GL, Stein ME, Flavell RA (1988): Conversion of a pi-anchored protein to an integeral membrane protein by a single amino acid substitution. Science 241:697-699.

Waugh DS (2005): Making the most of affinity tags. Trends Biotechnol 23:316-320.

West SM, Chaudhuri JB, Howell JA (1998): Improved protein refolding using hallow-fibre membrane dialysis. Biotechnol. Bioeng. 57:590-599.

Wilhelm O, Wilhelm S, Escott G, Lutz V, Magdolen V, Schmitt M, Rifkin D, Wilson E, Graeff H, Brunner G (1999): Cellular glycosylphosphatidyl-inositol-specific phospholipase D regulates urokinase receptor shedding and cell surface expression. J. Cell. Physiol. 180:225-235.

Wohnert J, Franz JJ, Nitz M, Imperiali B, Schwalbe H (2003): Protein allignment by a coexpressed lanthanide-binding tag for the measurement of dipolar couplings. J Am Chem Soc 125:13338-13339.

Wollerton MC, Wales R, Bullock JA, Hudson IR, Beggs M (2006): Automation and optimization of protein expression and purification on a novel robotic platform. Journal of the Association for Laboratory Automation 11:291-303.

Wurm FM (2004): Production of recombinant protein therapeutics in cultivated mammalian cells. Nature Biotechnology 22:1393-1398.

Xiaotong H, Hannocks MJ, Hampson I, Brunner G (2002): GPI-specific phospholipase D mRNA expression in tumor cells of different malignancy. Clin. Exp. Metastasis 19:291-299.

Xie J, Liu W, Schultz PG (2007): A genetically encoded bidentate, metal-binding amino acid. Angew Chem Int Ed 46:9239-9242.

Xu Y, Yasin A, Tang R, Scharer JM, Moo-Young M, Chou CP (2008): Heterologous expression of lipase in *Escherichia coli* is limited by folding and disulfide bond formation. Appl Microbiol Biotech 81:79-87.

Xue LA, Chen Y, Brook RJ (1988): The effect of lanthanide contraction on growth grain in lanthanide-doped batio3. Journal of Materials Science Letters 7:1163-1165.

Yamamoto Y, Hirakawa E, Mori S, Hamada Y, Kawaguchi N, Matsuura N (2005): Cleavage of carcinoembryonic antigen induces metastatic potential in colorectal carcinoma. Biochem. Biophys. Res. Comm. 333:223-229.

Yao JX, Dodson EJ, Wilson KS, Woolfson MM (2006): Acorn: A review. Acta. Crystallogr. D62:901-908.

Ye Y, Lee HW, Vang W, Shealy S, Yang JJ (2005): Probing site-specific calmodulin calcium and lanthanide affinity by grafting. J. Am Soc 127:3743-3750.

Yokoyama S (2007): Protein expression systems for structural genomics and proteomics. Curr Opin Biotechnol 7:39-43.

Zhan Y, Song X, Zhou GW (2001): Structural analysis of regulatory protein domains using GST-fusion proteins. Gene 281:1-9.