Towards the Production of the *Haemophilus influenzae* IgA1 Protease in *Escherichia coli*

by

William Russell Lotosky

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

H. influenzae is a gram-negative bacterium colonizing the human respiratory tract and is responsible for hundreds of thousands of deaths annually. To aid in persistence and infection in its human host *H. influenzae* produces many secreted virulence factors including an IgA1 protease (IgAP). This enzyme selectively cleaves human and great ape IgA1 in its proline, serine and threonine rich flexible hinge region. This enzyme is a type Va autransporter protein and shares significant structural homology to the serine protease autotransporters of the *Enterobacteriaciae*.

The objective of this thesis is to determine a method for the large scale expression of IgAP to allow for biochemical characterization and aid in understanding its role in infection and potential as a therapeutic for IgA nephropathy. This thesis presents three strategies for the attempted production of this enzyme in *Escherichia coli*, which is complicated by the inability of the protein to tolerate a methionine residue at the N-terminus. Native synthesis of the protein results in an alanine at the N-terminus caused by cleavage of the signal peptide during export to the periplasm. First the gene was cloned into the pBAD vector in frame with a viral signal peptide to target expression to the periplasm and produce a native N-terminus. Expression of IgAP in a simple pET vector was attempted, relying on the native methionine amino peptidase activity of *E. coli*. to cleave the N-terminal methionine. Finally, a mutant form of the enzyme lacking any methionine residues was expressed into insoluble inclusion bodies, purified and after denaturation with guanidine-HCl, the N-terminal methionine was cleaved with CNBr.

Through these studies we present several methods for the production of proteins in *E. coli* particularly applicable to proteins that will not tolerate a methionine at the N-terminus. The methods presented were insufficient to produce an amount of soluble, active IgAP as detectable by SDS-PAGE and cleavage of the IgA1 heavy chain.

Acknowledgements

I must acknowledge the tremendous support both personal and professional from all the graduate students I have worked with in the Holyoak and Rose labs: Marcie, Kyra, Malwina, Marc, Brianna and Iain. My time spent in the lab would have been potentially soul-destroying without you all as comrades. I would like to thank my supervisor Todd for putting up with my inability to be a functioning member of society most days before noon, for caring more about this project than I did at times and for being honest with his praise and criticism. I would also like to acknowledge the advice and counsel of my committee members David and Trevor.

I must acknowledge my mother Ann and sister Maya for their unfailing love and support and for feigning interest in my work so convincingly and finally all of my roommates these past years: Nikko, Cameron, Jen, James and Laura.

Dedication

I would like to dedicate this thesis to my late father Michael for instilling in me a love of science from an early age.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 *Haemophilus influenzae*

The genus *Haemophilus* is defined as having seven member species, all residing in humans or apes, based on sequence analysis of 16S rRNA using the maximum-likelihood method (Rosa & Labedan, 1998). One member species, Haemophilus *influenzae,* a pathogen of the human respiratory tract was erroneously identified in the 1890s as the causative agent of influenza after which it's named. It was later found to be responsible for many infections including: meningitis (Jones, 1937), bacteremic pneumonia, otitis media (Bjuggren & Tunevall, 1952) and septicaemia (De Navasquez, 1942) among others (Turk, 1984). *H. influenzea* is classified into strains based on the serotype of a capsular polysaccharide, with six serologically distinct strains (a-f) (Pittman, 1931) identified to date.

Non-typeable strains, lacking a polysaccharide capsule, are present in the respiratory tract of 40-80% of the population (Turk, 1984) and while these strains do not typically cause disease they act as opportunistic pathogens. The genome of the non-capsular strain Rd KW20 was the first bacterial genome to be sequenced (Fleischmann et al., 1995).

Strains presenting the *H. influenzae* serotype b (Hib) capsule cause most infections, particularly children under the age of 5 of which an estimated 247 000-545 000 a year die annually from Hib infections (Watt et al., 2009). During the 1990s Hib conjugate vaccines were introduced and have been very effective in reducing infections and carriage of *H.influenzae* serotype b strains (Morris, Moss, & Halsey, 2008). Worldwide 26% of children are estimated to be vaccinated, however the rate of vaccination is much higher in developed countries as compared to less developed countries (Morris et al., 2008). As a result other groups including serotype a (Hia) (Ulanova & Tsang, 2014) and non-typeable (NTHi) (Murphy et al., 2009) (Gkentzi, Slack, & Ladhani, 2012) strains are responsible for a greater proportion of infections.

As mentioned previously *H. influenzae* colonizes the human respiratory tract and is an obligate pathogen, well adapted to its niche. This can be demonstrated by comparing *H. influenzae* to

Escherichia coli, another human pathogen and γ-proteobacteria, that unlike *H. influenzae* is able to live in a diverse array of environments (Neidhardt & Curtiss, 1996). *H. influenzae* and *E. coli* share a common ancestor whose genome is believed to be close in size to that of *E. coli,* at 4.7 Mbp while the genome of *H. influenzea* is significantly smaller at 1.8Mb (Rosa & Labedan, 1998).

In order to persist and cause infections *H. influenzae* has developed several strategies against attacks from the human immune system including: capsular polysaccharide, lipoolygosaccharide (LOS), fimbriae, pili and many outer membrane and secreted proteins (Hallström & Riesbeck, 2010). The capsular polysaccharide of serotype b strains, composed primarily of a polymer of polyribosylribitilphosphate (PRP) (Crisel, Baker, & Dorman, 1975), is a major virulence factor (Sukupolvi-Petty, Grass, & St Geme, 2006). The outer membrane of *H. influenzae* contains lipooligosaccharides whose carbohydrate antigens mimic those of human cells in order to avoid an immune response (Mandrell et al., 1992). Adherence by *H. influenzae* to human nasopharyngeal epithelial cells is enhanced in strains possessing fimbriae, although adherence does not absolutely require fimbriae (Loeb, Connor, & Penney, 1988). Pili assist in biofilm formation and allow *H. influenzae* cells to adhere to each other (Murphy & Kirkham, 2002).

1.2 Immunoglobulin A

In human mucosa, including the respiratory tract, immunoglobulin A (IgA) is the predominant immunoglobulin and the second most abundant in the sera after IgG (Chodirker & Tomasi, 1963). IgA has the same basic structure as other immunoglobulins consisting of the two heavy and two light chains (Svehag & Bloth, 1970) the variable regions of both the heavy and light chain form the Fab region. The peptide portion of the Fc region is formed solely by two heavy chains.

Serum IgA is produced by the bone marrow and is monomeric while secretory IgA is synthesized by specialized cells derived from B-lymphocytes in the mucosa aided by other specialized immune cells (Underdown & Schiff, 1986). Compared to serum IgA, secretory IgA has a more variable and complex structure, involving two IgA molecules linked by disulphide bonds to the 15 kDa J-chain

(Koshland, 1985). The J-chain also allows for the pentamerization of IgM and higher order oligomers (tri, tetra and pentamers) of IgA are also produced to a lesser extent. As the mucosal epithelia form a barrier, IgA transport into the lumen is mediated by the polymeric immunoglobulin receptor (pIgR). IgA binds the pIgR at the basolateral membrane of the mucosal epithelium and travels to the apical membrane by transcytosis (Apodaca et al., 1991). During this process a portion of pIgR is cleaved and forms covalent attachments to IgA via disulphide bonds, this cleaved segment of pIgR is termed the secretory component (SC) (Woof & Kerr, 2006).

There are two subclasses of human IgA: IgA1 and IgA2 differing primarily by a 20 amino acid insertion present in the heavy chain of IgA1 (Kerr, 1990). This region comprises a duplicated octapeptide rich in proline, serine and threonine and forms a flexible linker between the conserved (Fc) and antigen binding (Fab) regions. The linker region contains twelve serine and threonine residues for O-glycosylation in addition to two N-linked glycosylation sites at Asn263 and Asn459 in the Fc region (Mattu et al., 1998). Analysis of secretory IgA1 by mass spectrometry has found the O-glycosylation of the hinge region to be quite variable and present studies, while able to identify the glycans present are unable to unambiguously assign the precise site and occupancy of these glycans (Deshpande, Jensen, Packer, & Kolarich, 2010)(Zauner et al., 2013).

IgA1 molecules with reduced galactose content in their hinge region have been implicated in IgA nephropathy (Novak, Julian, Tomana, & Mestecky, 2008). IgA nephropathy is characterized by the accumulation of IgA1 and to a lesser extent IgG and IgM in the glomerulus of the kidney, which causes an autoimmune response leading to reduced kidney function (Wyatt & Julian, 2013).

SIgA is present in human milk and is thought to play a role in the immunity of newborn children (Hanson & Korotkova, 2002). IgA is a poor activator of the complement pathway as compared to IgG and IgM and is only able to activate the complement via the lectin pathway and not the classical or alternative pathways (Roos et al., 2001). This is not to say that IgA is an ineffective member of the immune system, as the human mucosa contains far fewer complement components than the sera

(Kilian, Mestecky, & Russell, 1988). The predominant role of SIgA in mucosal immunity is to inhibit adherence of bacterial cells to the epithelia of the mucosa and allow for removal from the host (Woof & Kerr, 2006).

1.3 Immunoglobulin A1 Proteases

The importance of these molecules in the virulence and persistence of *H. influenzae* was discovered primarily because of their importance in other pathogens. Recently a more systematic approach to determine genes required for *H. influenzae* persistence and virulence using a murine model and a transposon insertion based screen was undertaken (Gawronski, Wong, Giannoukos, Ward, & Akerley, 2009). This screen identified genes primarily involved in the acquisition of nutrients from the host particularly amino acid and nucleotide precursors, proteins protecting against oxidative stress and formation of the LPS as well as outer membrane protein (Wong & Akerley, 2012). Unsurprisingly one gene not identified by this screen is the immunoglobulin A protease (IgAP), which cleaves human IgA1 in its hinge region, which is not present in murine IgA.

IgA proteases are bacterial enzymes that specifically cleave the hinge region of human and ape IgA1 between a proline/serine or proline/threonine peptide bond (Plaut, 1983). The first of these enzymes was initially identified as belonging to a bacterium present in human fecal matter (Mehta, Plaut, Calvanico, & Tomasi, 1973) and eventually found in members of the genus *Neisseria* (Plaut, Gilbert, Artenstein, & Capra, 1975). Both gram-negative and gram-positive pathogens have been found to possess these enzymes, although not all enzymes are evolutionary related as some species contain serine and metallo IgA1 proteases. Serine IgA1 proteases are found in members of the genera *Neisseria* and *Haemophilus* (Plaut, Flentke, Lynch, & Kettner, 1990), while *z*inc-bearing metalloproteases are found in members of the genus *Streptococci* (Plaut, Gilbert, & Heller, 1978), *Clostridium* and *Prevotella* (Frandsen, Kjeldsen, & Kilian, 1997) although the streptococcal enzymes are antigenically distinct form those of the latter two genera.

The precise site of cleavage in the hinge region varies depending on the particular enzyme. For example an enzyme from *H. influenzae* cleaves the peptide bond between proline 231 and serine 232, while Streptococcal enzymes generally cleave between proline 227 and serine 228 (Batten, Senior, Kilian, & Woof, 2003). Proteases found in strains of *S. pneumoniae*, *Streptococcus oralis*, *Streptococcus mitis* and *Streptococcus sanguis* show varying levels of tolerance for substitution of hinge residues. However, it's the *S. pneumoniae* enzyme that demonstrates the highest tolerance for mutated substrates but it will preferentially cleave proline-threonine peptide bonds (Batten et al., 2003). Enzymes found in *Neisseria* and *Haemophilus* have preferred cleavage sites closer to the Fc region than streptococcal enzymes. There are also two distinct types of cleavage, type 1 in which a proline-serine bond is cleaved and type 2 where cleavage occurs at a proline-threonine bond (Batten et al., 2003).

The role that these extracellular enzymes play in pathogenesis expands beyond the cleavage of IgA1, the *H. influenzae* and *N. gonorroheae* enzymes can also cleave other substrates. Lysosomal associated membrane protein 1 (LAMP1), as its name suggests is an integral membrane protein of the lysosome, is cleaved by the gonoccocal IgAP allowing the pathogens to reproduce intracellularly (Lin et al., 1997) (Hauck & Meyer, 1997). The *Neisserial* enzymes also elicit the production of pro-inflammatory cytokines, including tumor necrosis factor-α (TNF-α), in a manner independent of serine protease activity (Jose, Wölk, Lorenzen, Wenschuh, & Meyer, 2000).

1.4 Type V Secretion System

H. influenzae IgAP is exported to the extracellular media using the type V secretion or autotransporter(AT) pathway, which was first discovered in the IgAP of *Neisseria gonorroheae* (Pohlner, J., Halter, R., Beyreuther, K. & Meyer, 1987). More specifically the IgAPs of *N. gonorroheae* and *H. influenzea* are type Va ATs as opposed to type Vb or Vc ATs. Proteins belonging to the type Va family share a general structure characterized by three domains: an N-terminal signal peptide, followed by a passenger and finally the C-terminal β-domain (Dautin & Bernstein, 2007). The passenger domain carries out the proteins intended function, while the signal peptide and β-domain serve as a means of crossing the inner membrane and outer membrane

respectively (Desvaux, Parham, & Henderson, 2004). Type Vb autotransporters consist of two seperate peptides, one containing the signal sequence and passenger domain and another containing the β-domain (Dautin & Bernstein, 2007). All AT passenger domains for which a crystal structure is available contain large β-helices, consisting of 3 parallel sheets occasionally with loops present in between the turns (Dautin & Bernstein, 2007).

The autotransporter pathway is the simplest known method for export of gram-negative bacterial proteins as the proprotein is sufficient for crossing the OM. In the type V secretion system transport across the IM occurs co-translationally via the secretory pathway (Brandon et al., 2003). Transport through the IM involves the passage of the peptide through the SecYEG complex or Sec translocon, however there are two pathways that bring a peptide to the translocon. One method is the signal recognition particle (SRP) pathway, in which the SRP binds the signal peptide and allows for co-translational transport (Luirink & Sinning, 2004), the other is the SecAB pathway in which the protein is maintained in an unfolded state and transported post-translationally (Valent et al., 1998).

Both methods require the presence of an N-terminal signal sequence with three characteristic features: the N-terminal n-region, which is often positively charged followed by the hydrophobic h-region and finally the C-terminal c-region where cleavage by signal peptidase occurs (Martoglio & Dobberstein, 1998). The *H. Influenzae* IgAP signal peptide (MLNKKFKLNFIALTVAYALTPYTEA) follows this general structure. Many ATs contain atypically long signal peptides, which slows their transit through the IM and is believed to reduce the incidence of folding and misfolding in the periplasm (Peterson, Szabady, & Bernstein, 2006). In the periplasm ATs interact with several chaperones: Skp, DegP and SurA to prevent their aggregation and misfolding before export across the OM (Knowles, Scott-Tucker, Overduin, & Henderson, 2009).

Once the AT has crossed the IM into the periplasm the β-domain is responsible for translocation across the OM demonstrated by the fusion of periplasmic proteins with the *N. Gonorroheae* IgAP β-domain, which results in there surface exposure (Klauser, T., Joachim, K., Otzelberger, K., Pohlner and Meyer, 1993). β-domains are roughly 30kDa in size and possess a β-barrel structure consisting of

12-15 strands (Henderson & Navarro-Garcia, 2004). Insertion of the translocator domain into the OM was previously thought to be a spontaneous and thermodynamically favourable, with some assistance from chaperones, peptidylproline isomerases and disulphide isomerases (Tamm, Arora, & Kleinschmidt, 2001). However more recent evidence shows that β-barrel insertion into the OM requires the β-barrel assembly machinery (BAM) (Voulhoux, Bos, Geurtsen, Mols, & Tommassen, 2003) and that AT secretion is also BAM dependent (Jain & Goldberg, 2007).

The mechanism by which the passenger domain traverses the OM resulting in either localization to the OM or cleavage and release to the extracellular media is the least understood part of AT biogenesis. All crystal structures of AT β-domains contain, in addition to the 12 stranded β-barrel, an α-helix present in the cavity of the barrel (van den Berg, 2010) (Zhai et al., 2011) (Barnard, Dautin, Lukacik, Bernstein, & Buchanan, 2007) (Oomen et al., 2004) (Tajima, Kawai, Park, & Tame, 2010) (Barnard et al., 2012) (Meng, Surana, St Geme, & Waksman, 2006). A model in which, the passenger domain is threaded through the central cavity of the β-domain, was first proposed but due to the discovery that disulphide bond containing (Veiga, de Lorenzo, & Fernández, 1999) and protease resistant (Brandon & Goldberg, 2001) substrates can be translocated has led to other models being proposed. Another model of AT transit through the OM involves passage through a large pore formed by an oligomeric ring of β-domains (Veiga, de Lorenzo, & Fernández, 2004).

The hairpin mechanism involves hairpin formation by the region N-terminal to the β-barrel being inserted into the cavity, followed by the threading of the remainder of the molecule in a C- to N-terminal direction into the extracellular media (Junker, Besingi, & Clark, 2009). This is supported by refolding experiments involving the ATs pertactin and Pet, which are mainly β-helical, although Pet contains an N-terminal serine protease domain, that revealed a stable, protease resistant C-terminal core and an unusually slow refolding rate with little aggregation in vitro (Junker et al., 2006) (Renn & Clark, 2008). The folding of this stable core is believed to provide the energy for translocation as there is little ATP in the periplasm and no proton gradient across the OM to couple to as a source of energy (Peterson, Tian, Ieva, Dautin, & Bernstein, 2010). Within this model there is still uncertainty as to whether the α-helix found in the pore of β-domain crystal structures is present in the pore during

translocation or if it is present in the periplasm. The final model of OM transport of AT proteins is the BamA mechanism in which β-domain is still inserted into the OM but the passenger travels through BamA to leave the periplasm (Leyton, Rossiter, & Henderson, 2012).

Some ATs are cleaved and released from the cell, such as IgAP, while others remain associated with the OM via attachment to their β-domain, however amongst cleaved passenger domains there are different methods of cleavage (Henderson & Navarro-Garcia, 2004). An *E. coli* O157:H7 autotransporter, extracellular serine protease plasmid encoded (EspP), has its passenger domain cleaved inside the β-domain cavity and not its serine protease domain (Barnard et al., 2007). The cleavage is mediated by the cyclization of the Asn1023 side chain and carbonyl carbon atom, which is facilitated by residues inside the β-barrel and several water molecules (Barnard et al., 2012), the SPATE protein hemoglobin protease (Hbp) autocatalyitcally cleaves its passenger domain in this manner (Tajima et al., 2010). While EspP is cleaved roughly in the middle of the β-barrel, NalP another AT, is cleaved at the far extracellular end of the β-barrel (Oomen et al., 2004). After cleavage the remainder of the α-helix becomes oriented perpendicular to the membrane and a loop on the periplasmic side folds down to cover the channel (Tajima et al., 2010). The release of the IgAP passenger domain is catalyzed by its serine protease domain rather than residues in the β-domain, as mutants of the active site S288 remain associated with the OM (Qiu et al., 1998) (Plaut, Qiu, & St Geme, 2000).

1.5 Serine Protease Autotransporters of the *Enterobacteraciae* (SPATE)

The sequence similarity amongst all known passenger domains is low, however for all AT with structural data available there is a prominent right-handed β-helix. One group of AT with more significant homology between themselves are the SPATE proteins. While *H. influenzae* IgAP is not technically a member of this group, as the genus *Haemophilus* belongs to the order *Pasteurellales* and not *Enterobacteriales*, its general structure is identical. To date the structure of three SPATE passenger domains: Hbp (Otto et al., 2005), EspP (Khan, Mian, Sandercock, Chirgadze, & Pai, 2011) and Pet (Domingo Meza-Aguilar et al., 2014), as well as *H. influenzae* IgAP (Johnson, Qiu, Plaut, &

Holyoak, 2009) have been solved. SPATE proteins share the same general structure as other AT, they contain an N-terminal signal peptide for crossing the IM, a C-terminal β-domain and finally a passenger domain which in addition to a β-helix possesses an N-terminal chymotrypsin-like serine protease domain (Dautin, 2010).

Serine proteases cleave peptide substrates using a catalytic triad or charge transfer system consisting of an aspartate, histidine and a nucleophilic serine (Blow, Birktoft, & Hartley, 1969), this motif is believed to have evolved four separate times (Dodson & Wlodawer, 1998). The generally accepted mechanism of cleavage by the chymotrypsin family of proteases involves substrate binding followed by formation of a tetrahedral intermediate formed by nucleophilic attack on the P1 carbonyl carbon, by Oγ of the catalytic serine (Blow, 1997). The formation of the oxyanion is stabilized by the positively charged backbone amides, referred to as the oxyanion hole (Henderson, 1970), collapse of the intermediate results in the release of the leaving group, residues P1'-Pn', and the formation of an acyl enzyme intermediate (Dixon, Brennan, & Matthews, 1991)(Press & Sussman, 1991). The histidine of the catalytic triad assists in activating a water molecule to attack the acyl-enzyme intermediate resulting another tetrahedral intermediate followed by collapse and release of P1-Pn (Blow, 1997). While there is direct evidence for the formation of an acyl-enzyme intermediate the evidence for a tetrahedral intermediate consists of inhibitors mimicking the transition state (Radisky, Lee, Lu, & Koshland, 2006), however non-enzymatic hydrolysis reactions are known to involve a tetrahedral intermediate (Bender, 1960).

SPATEs are members of the chymotrypsin family of serine proteases, which contains the most members and are found in all domains of life as well as viruses (Rawlings, Barrett, & Bateman, 2000). A nomenclature was developed for the general description of peptide substrates and the corresponding binding subsites on the protease, the two residues containing the scissile peptide bond are termed P1 and P1' (Schechter & Berger, 1968). Residues on the side of the acyl intermediate are labelled NH3-Pn...P1, while residues of the leaving group are referred to as P2'...Pn'-COO^-. A similar system is used for naming protease residues responsible for substrate binding, the Sn...S2,S1,S1',S2'...Sn' (Schechter & Berger, 1968).

Some SPATEs possess an extremely limited number of substrates while others can cleave a wider array of peptides, the serine IgAPs from *Haemophilus* and *Neisseria*, for example, are known to cleave only two substrates (Lin et al., 1997), while others, such as Pic form Enteroaggregative *E. coli* (EAEC), have as many as eight known substrates (Dautin, 2010). The specificity of chymotrypsin-like serine proteases is determined by the binding subsites as well as surface loops (Perona & Craik, 1997), for example the S1 subsite of trypsin, which prefers substrates with a lysine or arginine in the P1 site, consists of a glycine an aspartate as well other residues. The glycine allows for acomadation of larger side chains while the negatively charged aspartate favours positively charged substrates (Graf et al., 1987). There are eight loop regions in the chymotrypsin fold that play an important role in substrate specificity, loops A, B, C, D and E, which form the basis for certain subsite specificities and loops 1, 2 and 3 generally believed to influence the S1 site (Perona & Craik, 1995). Another important determinant of substrate specificity is the identity of residues 216 and 226 (based on the chymotrypsin numbering), which further splits the chymotypsin family into four sub-groups: chymotrypsin, trypsin, elastase and collagenase (Perona & Craik, 1995).

The elastase sub-group has a preference for small non-polar residues, such as proline, at position P1 of the substrate due to a valine at residue 216 in place of glycine found chymotrypsin (Bode, Meyer, & Powers, 1989). The *H. influenzae* IgAP serine protease domain shows a substrate preference similar to elastase as opposed to chymotrypsin, trypsin or collagenase, however there are several differences in the composition and length of the aformentioned loop regions, particularly loop D (Johnson et al., 2009).

1.6 *H. influenzae* and the *Neisseria* IgA Proteases

In addition to the β-helix and serine protease domain IgAP as well as Hbp contain other globular domains inserted in between the turns of the β-helix, referred to as domains 2, 3 and 4, although Hbp contains only domain 2. These globular insertions into the helical scaffold are hypothesized to play a role in substrate recognition and may account for the extreme selectivity of these enzymes (Johnson et

al., 2009). The only other SPATE passenger domains with known structures, EspP and Pet do not contain a domain 2 but contain small globular insertions into the β-helix similar to domain 3 in Pet and domains 3 and 4 in EspP (Domingo Meza-Aguilar et al., 2014; Khan et al., 2011). SPATEs lacking domain 2 are classified as class-1 while SPATEs containing a chitinase-b like domain 2 are labelled class-2 (Ruiz-Perez & Nataro, 2014). The exact role of domain 2 in Hbp has been examined by generating a mutant form of the enzyme lacking domain 2, which revealed that it was not required for haem binding or cleavage of certain peptide substrates (Nishimura et al., 2010).

The first attempt at producing inhibitors specific for the serine IgAPs were based on the octapeptide region of the IgA1 hinge, the most effective inhibitor with an IC50 of 50μM, generated from this effort contained a cysteine substituted at the P3' position and an acylated N-terminus (Burton, Wood, Lynch, & Plaut, 1988). The next effort towards specific inhibitors were once again similar in composition to the IgA1 hinge but contained an α-aminoboronic acid mimic of proline, the most effective inhibitor from this study (Ac-Ala-Pro-boroPro-OH) had a Ki of 13nm for the *H. influenzae* enzyme (Plaut et al., 1990)

The *H. influenzae* and *Neisseria* IgAPs are able to cleave an altered form of IgA2 which contains a 7 amino acid insertion of the IgA1 hinge region, which suggests that these enzymes only require a portion of the hinge region for cleavage (Senior, Dunlop, Batten, Kilian, & Woof, 2000). More recently the *H. influenzae, Neisseria* and *Streptococcus* IgAPs are able to cleave fluorogenic substrates based on the IgA1 hinge region (Choudary, Qiu, Plaut, & Kritzer, 2013). The smallest substrate the *H. influenzae* enzyme is able to cleave is a hexapeptide (PPAPVY) based on the *N. gonorrheae* autoproteolytic site. Research into the function of these enzymes in pathogenesis, the role of domains 2, 3 and 4 and the design of potent inhibitors are active areas of research.

1.7 Recombinant Protein Expression in *E. coli*

Many biochemical and structural biology techniques require large amounts of highly pure protein, on the order of milligrams, however obtaining that amount of protein from a natural source can be prohibitive because of cost, safety or simply unfeasible due to other factors. Since the molecular

biology revolution recombinant protein expression has become *de rigueur* for almost every lab and can be seen in almost every new publication. There are wide range of hosts, plasmids, selection techniques and strategies used in recombinant protein expression but perhaps the simplest and most widely used system is *E. coli* (Rosano & Ceccarelli, 2014).

Lineages derived from strains K-12 and B are non-pathogenic and are routinely used in a laboratory setting for protein expression (Daegelen, Studier, Lenski, Cure, & Kim, 2009). *E. coli* will grow relatively quickly in a wide range of media, lysogeny broth (LB) is a complex media composed of protease digested casein, yeast extract and NaCl (Bertani, 1951) and is the most common growth media. When grown at 37°C, with agitation in LB, *E. coli* grows with a doubling time of 20 minutes during the exponential phase of growth (Sezonov, Joseleau-Petit, & D'Ari, 2007). However cell growth can be enhanced and protein expression optimized by the use of media supplemented with any number of carbon or nitrogen sources, metals, vitamins, inorganic salts and saccharides (Studier, 2005)

While genes can be introduced directly into the *E. coli* genome, proteins destined for expression are generally transformed into *E. coli* via a plasmid vector, containing at a minimum: an origin of replication, a promoter, an antibiotic resistance gene, a transcription initiation and stop site, a ribosome binding site and translation stop sequence (Sørensen & Mortensen, 2005). In addition proteins may be expressed with a fusion partner to aid in expression, some commonly used fusion partners include: maltose binding protein (MBP) (Kapust & Waugh, 1999), SUMO (Butt, Edavettal, Hall, & Mattern, 2005) and glutathione S-transferase (GST) (Smith, Johnson, 1988). The most widely used expression system is the pET system, which makes use of the T7 bacteriophage promoter and the *lac* operon to regulate expression using the inducer iso-propylthiogalactylpyranoside (IPTG) (Dubendorff & Studier, 1991). A less common system is the pBAD system which makes use of the *araBAD* promoter and uses the L-arabinose operon to control expression (Guzman et al., 1995).

While the expression of soluble protein is the objective of most expression strategies the expression of certain gene products results in insoluble inclusion bodies (Hartley & Kane, 1986), in these cases there are strategies for the solubilization and refolding of these proteins. First the inclusion bodies must be purified and isolated from the remainder of the cellular contents, then solubilized with either high concentrations of urea, guanidine-HCl or detergents, in the presence of reducing agents to prevent intramolecular disulfide formation (Basu, Li, & Leong, 2011). Finally the denaturants and reducing agents must be removed to allow for refolding, there are a number of methods that may be employed including direct dilution (Leong & Middelberg, 2007), dialysis and column refolding (Schmoeger, Wellhoefer, Dürauer, Jungbauer, & Hahn, 2010).

The challenge in expressing *H. influenzae* IgAP recombinantly is that the native N-terminus of the secreted passenger domain is an alanine residue buried in the hydrophobic core of the chymotrypsin like serine protease domain (Johnson et al., 2009), the addition of a methionine at N-terminus results in expression into inclusion bodies. I will present three strategies for cleavage of the undesired N-terminal methionine: cleavage by endogenous *E.coli* methionine aminopeptidase (MAP), co-expression of IgAP with MAP and finally expression of IgAP in inclusion bodies, chemical cleavage of methionine by CNBr and refolding.

1.8 Objectives

The objective of the work presented in this thesis is to determine an expression system able to generate soluble, active *H. influenzae* IgAP in *E. coli* without relying on the enzyme's own catalytic activity in sufficient quantities for biophysical and biochemical characterization. The protein used to determine the structure of the enzyme was obtained from the culture media of *H. influenzae* (Johnson et. al., 2009), however this system is dependent on the enzyme's ability of the passenger domain to cleave itself from the β-domain. Future studies will require mutants potentially lacking in catalytic activity therefore an alternate expression system is required, in addition the *H. influenzae* culture media was provided by a collaborator who is no longer able to do so.

Previous attempts by former members of the Holyoak lab to determine a suitable expression system involved the use of *Bacillus megaterium* as an expression host, co-expressing IgAP with MAP and treatment of IgAP inclusion bodies with MAP in the presence of denaturant. These methods were unable to generate soluble active IgAP in any meaningful amount.

Chapter 2 Materials and Methods

2.1 pBAD-IgAP Cloning and Expression

2.1.1 pBAD-IgAP Cloning

H. influenzae IgA1P was cloned into the pBAD/gIII vector using the 5' NcoI and 3' SalI sites, the gene was amplified from the pET2(b)-IgAP construct. The insert was originally obtained from *H. influenzae* genomic DNA. Sequencing confirmed that the insert had the correct orientation and that no unintended mutations were introduced after cloning. The first two residues immediately following the signal sequence cleavage site were mutated to alanine and leucine respectively, to generate the native N-terminus, using the quick change mutagenesis kit (Stratagene) and the following primer pairs:

Hi IgA pBAD1 for

5' ccgttctatagccatagcgccatggtgagagacgatgtgg 3'

Hi IgA pBAD1 rev

5' ccacatcgtctctcaccatggcgctatggctatagaacgg 3'

Hi IgA pBAD2 for

5' ccgttctatagccatagcgccttggtgagagacgatgtgg 3'

Hi IgA pBAD2 rev

5' ccacatcgtctctcaccaaggcgctatggctatagaacgg 3'

Reaction conditions were determined using the QuickChange Mutagenesis kit (Stratagene).

Polymerase chain reactions (PCR) were incubated with 1µL of *DpnI* restriction enzyme (New England BioLabs) for one hour at 37^{o}C. To verify that the PCR reaction was successful 25 µL of the digested reaction was mixed with 5 µL loading dye and then run on a 0.8% agarose gel containing RedSafe dye (Intron Biotechnology) for visualization under UV light. 1-5 µL of successful reactions was used to transform chemically competent cells of XL-1 Blue *E. coli*., single colonies were used to

inoculate 5mL cultures of LB media containing 100 µg/mL of sodium ampicillin and grown overnight in shaking incubator at 37°C. Plasmid DNA was isolated from the overnight cultures using a GENEJet Plasmid Miniprep Kit (Thermo Fisher) and eluted in 50 µL of elution buffer, a nanodrop spectrophotometer (Thermo Fisher) was used to measure the absorbance of light at 260 nm in order to determine the concentration of DNA present in the sample. To ensure that no unintended mutations were introduced, several clones were sent for DNA sequencing of the entire IgAP gene and successful clones were used to transform BL-21 (DE3) *E. coli* cells for expression.

2.1.2 pBAD-IgAP Expression

100 uL of a 5mL culture of LB media containing 100 µg/mL ampicillin, grown overnight, carrying the pBAD/gIII-IgA1P construct, were used to inoculate a 10 mL LB culture containing 100 µg/mL ampicillin. The cultures were grown in a shaking incubator to an OD_{600} of 1.0 then induced using 100 µL of 0.002-20% w/v L-(+)-arabinose to give a final concentration of 0.00002-0.2% w/v. Expression was carried out at 37°C for hours with 1mL samples of culture media being collected immediately prior to and 4 hours after induction, the OD_{600} was determined using a UV-Vis spectrophotometer (Cary). Another trial was conducted by inducing the cells at 25°C and collecting samples 0, 24, 48 and 72 hours after induction.

The cells collected from the expression media were analyzed centrifuge at 14000 rpm for 1 minute then resuspended in a lysis buffer containing 10mM Tris-HCl pH 8, 1mM EDTA and 20% sucrose to give an OD_{600} of 5. To lyse the outer membrane and release the contents of the periplasm through osmotic shock the cells were incubated on ice for 10 minutes then pelleted via centrifugation at 14000 rpm for 1 minute at 4°C. The supernatant was discarded and the cells were resuspended in a buffer containing 10 mM Tris-HCl pH 8.0 and 1 mM EDTA, incubated on ice for 10 minutes then centrifuged at 14000 rpm for 10 minutes at 4°C. The supernatant, corresponding to the periplasmic fraction of the cells, was precipitated using trichloroacetic acid (TCA). 1 mL of periplasm was mixed with 250 µL of 100% TCA and incubated on ice for 30 minutes then centrifuged for 10 minutes at 14000 rpm. The supernatant was removed with care and the pellet washed with 200 uL ice-cold acetone followed by centrifugation at 14000 rpm for 10 minutes twice. The final pellet was resuspended in 20 µL SDS-PAGE sample buffer and analyzed on a 12% gel.

Expression was also carried out by inducing cells with 20%-0.002% w/v L-(+)-arabinose at 25^{o}C, 1 mL samples were taken immediately before, and 24, 48 and 72 hours after induction.

2.2 pET24(b)-IgAP and pDUET-IgAP-MAP Expression

Expression studies to examine the role of temperature, amount of inducer, *E. coli* strain and expression media were undertaken. *E. coli* BL-21(DE3) cells transformed with the pET24(b)-IgAP construct were expressed at 37°C, 28°C and 21°C and induced with 0.1 mM, 1 mM and 5 mM IPTG, expression was monitored over the course of 4 hours, samples of the cell culture were collected. To analyze the soluble fraction cells were pelleted by centrifugation at 5000 g for 5 minutes and then sonicated for a total of 2 minutes using a cycle of 10 seconds on 50 seconds off.

C41 (DE3) cells were transformed with the pET24(b)-IgAP construct were grown in super broth, LB and M9 minimal media, cells were induced with 0.4 M IPTG and grown at 20°C after induction. To enrich for IgAP samples were incubated with Ni-NTA resin (BioRad) after lysis, the resin was boiled in SDS-PAGE sample buffer to elute the protein and analyzed by SDS-PAGE.

2.3 pET24(b)-IgAP Insoluble Expression and Refolding

2.3.1 Generating a Construct with Fewer Methionine Residues

IgAP contains 6 methionine residues: M74, M117, M167, M635, M647 and M655, using the existing x-ray crystal structure of IgAP it was determined whether the residues were either buried in the hydrophobic core or exposed to solvent. Residues M74, M167, M635, M647 and M655 were found buried in a hydrophobic environment and thus mutated to leucine while M117 was found to be

exposed to solvent and was converted to arginine.

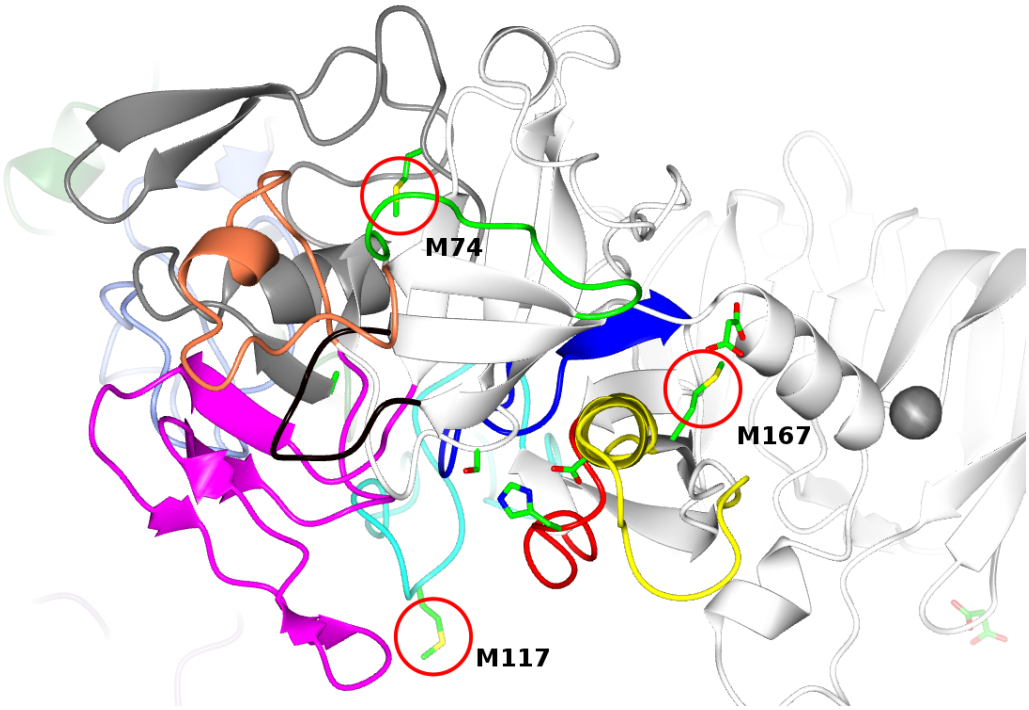

Figure 2-1 Location of the three methioinine residues (M74, M117 and M167) present in the N-terminal serine protease domain of IgAP. Loop regions are coloured identically to the figures in Johnson, 2009 (Johnson, et. al., 2009). The side chains of the methionine residues and the catalytic triad are depicted as cylinders. (PDB: 3H09).

M74 is located in an N-terminal extension not found in elastase, M117 is located on loop E and M167 is located on a β-strand adjacent to the aspartate of the catalytic triad. M635, M647 and M655 are located in domain 2 a globular insertion in the β-helical domain.

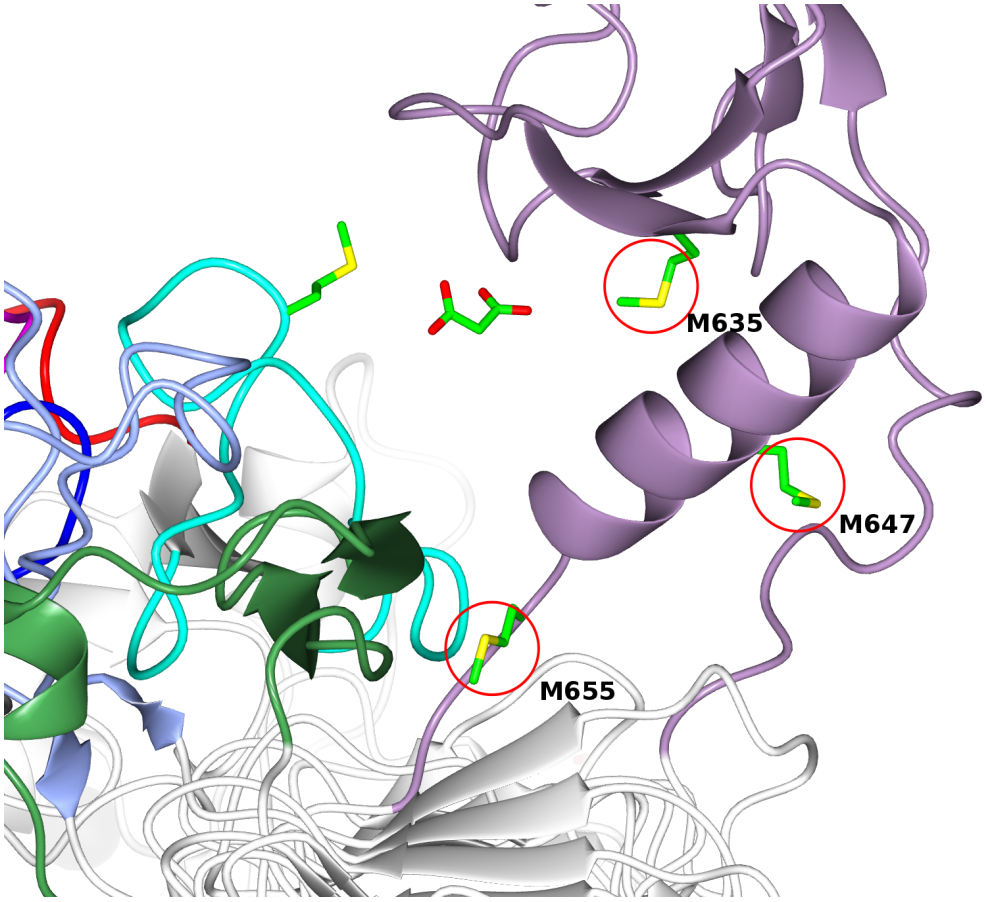

Figure 2-2 Location of methionine residues in Domain 2 of IgAP, loop regions and domains are coloured identically to the figures presented in Johnson, 2009 (Johnson, et. al., 2009). The side chains of the methionine residues are depicted as cylinders. (PDB:3H09).

The mutations were generated using the following primer pairs, the mutated codon is bolded and underlined for emphasis:

M74L

F 5' cctaatggcattccg**ttg**attgattttagtgttgtgg 3'

R 5' ccacaacactaaaatcaat**caa**cggaatgccattagg 3'

M117R

F 5' gggaacttaaatggcaat**agg**aataatggcaatgc 3'

R 5' gcattgccattatt**cct**attgccatttaagttccc 3'

M167L

F 5' ccgtgaagactactat**ctg**ccacgtcttgataaattg 3'

R 5' caatttatcaagacgtgg**cag**atagtagtcttcacgg 3'

M635L

F 5' gaaaattggctatat**ttg**ggtaaaacttccgatgaagcc 3'

R 5' ggcttcatcggaagttttacc**caa**atatagccaattttc 3'

M647L

F 5' gccaaaagaaatgta**ttg**aaccatatcaacaacgtatg 3'

R 5' catacgttgttgatatggtt**caa**tacatttcttttggc 3'

M655L

F 5' ccatatcaacaacgagcgt**ttg**aatggctttaacgg 3'

R 5' ccgttaaagccatt**caa**acgctcgttgttgatatgg 3'

PCR reaction conditions were conditions were made following the QuickChange Mutagenesis kit (Stratagene). The remainder of the cloning process was carried out identically to that of section 2.1 except that 50 μg/mL kanamycin was used to select for transformants instead of 100 μg/mL ampicillin.

2.3.2 pET24(b) Expression of Inclusion Bodies

To generate inclusion bodies 1.5L of LB media containing 50 μg/mL kanamycin was inoculated with a 50 mL culture grown overnight at 37°C. The cultures were grown to an OD_{600} of 1.0 then induced with IPTG to a final concentration of 1 mM and grown for 4 hours after induction. Finally the cells were harvested from the media by centrifugation at 5000 g for 15 minutes, cell pellets were stored at -80°C until needed.

2.3.3 Isolation and Purification of Inclusion Bodies

The cell pellet was lysed by passage through an emulsifier twice at ~17000 psi, 10% Triton X-100 detergent was added to the lysate to a final concentration of 1%v/v. The insoluble fraction was obtained by centrifugation at 24000 g for 15 minutes, the resultant pellet was washed twice in a buffer containing 0.1% Triton X-100 detergent to remove lipids and lipid soluble proteins. Resuspension of the inclusion bodies was achieved by dissolving the pellet, in a buffer containing 6 M Guanidine-HCl and 25 mM HEPES pH 7.5, overnight at 4°C . The sample was centrifuged at 24000 g for 15 minutes once more to remove any material that is still insoluble and then the supernatant is incubated with Ni-NTA resin for 30 minutes while stirring. The resin was washed with roughly 300 mL of 6 M Guanidine-HCl pH7.5 until the A_{280} is less than 0.1, to elute the sample the column is washed in roughly 50 mL of 6 M Guanidine-HCl pH 0.

2.3.4 Cleavage of N-terminal Methionine using CNBr

CNBr was added to the sample for a final concentration of 50 mg/mL and left stirring at room temperature in a sealed light-proof beaker. The reaction was neutralized by the addition of 3 M Tris until the solution yielded a pH of 7.5 by monitoring the reaction with a pH meter (Fisher). To visualize the samples by SDS-PAGE 50 μL of the cleavage reaction was diluted to a final volume of 1 mL in 25 mM HEPES pH 7.5, precipitated by the addition of 200 μL TCA, incubated on ice for 30 minutes, centrifuged at 14000 g and the pellets were dissolved in 20 μL of SDS sample buffer. A construct omitting the M167L mutant was used to monitor the cleavage reaction via SDS-PAGE.

2.3.5 Refolding of CNBr Cleaved Protein

The different methods for refolding of IgAP treated with CNBr are listed in **Table 2.1** below with a more detailed procedure for each of the methods in the following sub-sections.

Table 2-1 Attempted Refolding Conditions: All refolding buffers also contained 25 mM HEPES, the pH was adjusted to 7.5 using either 6 M HCl or 10 M NaOH.

Condition	Method	[Initial]$	[Final]$	GSH*	L-Arg#
A	Column	8 M	0.5 M	N	N
B	Dilution	8 M	0 M	N	N
C	Dilution	8 M	0.5 M	N	N
D	Dilution	8 M	0.5 M	Y	N
E	Dilution	8 M	0.5 M	Y	Y
F	Dialysis	8 M	0.5 M	N	N
G	Dialysis	8 M	0.5 M	Y	N
H	Dialysis	8 M	0.5 M	Y	Y
I	Dialysis	8 M	0.5 M*	Y	Y

*GSH indicates that 5 mM reduced glutathione (GSH) and 0.5 mM oxidized glutathione (GSSG) was included during refolding, #L-Arg indicates that 400 mM L-arginine was included in the refolding buffer. $[Initial] and [Final] refer to the initial and final concentration of urea. In condition **I** dialysis was carried out stepwise with the sample being dialyzed overnight against 4 M, 2 M and finally 0.5 M urea containing buffers.

2.3.5.1 On Column Refolding

After quenching of the CNBr reaction with 3 M Tris the protein was incubated at 4^{o}C with Ni-NTA resin equilibrated in 8 M urea, 25 mM HEPES pH 7.5 for 30 minutes while stirring. The column was washed with a linear gradient of urea from 8 M to 0.5 M over 200 mL, the protein was then eluted with a buffer containing 300 mM imidazole and 25 mM HEPES pH 7.5. To monitor the elution of protein from the column the absorbance of the solution at 280 nm was measured.

2.3.5.2 Refolding by Direct Dilution

CNBr digested IgAP was added directly to the refolding media over the course of 15 minutes, the sample was stirred so that rapid mixing occurred. Digested protein was diluted in a volume 99 times larger than the sample volume equilibrated at 4^{o}C and left overnight to ensure folding was completed. Samples were concentrated in an Amicon nitrogen concentrator equipped with a 30 kDa filter until the sample reached a volume of 1-5mL.

2.3.5.3 Refolding by Dialysis

Samples where refolding was attempted by dialysis were placed in 10 kDa molecular weight cut-off SnakeSkin dialysis tubing (ThermoFisher) overnight. The volume of the refolding solution was always at least 500 times larger than the contents of the dialysis tubing. To remove precipitate formed during dialysis the samples was centrifuged at 5000 g for 15 minutes at 4°C.

2.3.6 Analysis of Refolded Protein

2.3.6.1 IgA1 Cleavage Assay

The primary method used to determine whether the sample was refolded was to assess its enzymatic activity, this was achieved by incubating potentially refolded protein with human IgA1 isolated from human sera (provided by Dr. A. G. Plaut, Tufts University Medical Centre). To determine whether the protein was active 5 µL of IgA1 was incubated with 10 µL of IgAP overnight at 30^{o}C and room temperature then half of the reaction was analyzed by SDS-PAGE to determine if the heavy chain of IgA had been cleaved.

2.3.6.2 Secondary Structure Analysis Using Circular Dichroism Spectrophotometry

CD spectrophotometry was used to assess the secondary structure of any potentially refolded species. Samples were dialyzed overnight into 10 mM PO_4 pH 7.5 and concentrated in a centrifugal concentrator, with a 10 kDa size cutoff, to at least 10 µM for analysis and finally degassed before a spectra was collected. The ellipticity was measured using a Jasco J-715 spectrophotometer scanning from 260-190 nm in 1 nm steps, with a response time of 0.125 s and a path length of 0.1 cm. A

spectra was collected first on a sample of 10 mM PO_4 pH 7.5 in order to generate a blank, each spectra was generated by taking the average of 8 scans.

To calculate the amount of basic secondary structural elements present in the folded wild-type IgAP the PDB 3H09 was submitted to the 2Struc server (http://2struc.cryst.bbk.ac.uk/twostruc).

CD spectra were analyzed using the DichroWeb server: (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml), which provides a single interface for several secondary structure determination programs. Only programs able to interpret data between 190-260 nm were used.

Chapter 3 Results and Discussion

Herein I present three strategies for the expression of the *H. influenzae* IgAP in *E. coli*, all strategies employ different means for cleaving the problematic N-terminal methionine. The first system makes use of the pBAD vector to target IgAP to the periplasm via an N-terminal signal sequence, which generates a native N-terminus. The second strategy makes use of the native methionine aminopeptidase activity of *E. coli* to cleave the protein produced from a pET vector. The final strategy relies upon the ability of CNBr to specifically cleave peptide bonds C-terminal to methionine residues. The construct for this expression strategy required that all methionine residues be mutated to other residues so that CNBr didn't cleave the peptide anywhere other than at the N-terminus. To assess whether the protein was folded and active circular dichroism spectroscopy and a gel based cleavage assay of human IgA1 were employed. Included as a positive control for the IgA1 cleavage assay are two lanes of a 12% SDS-PAGE gel. The first lane contains IgA1 incubated overnight at 30°C and in the second lane IgA1 incubated overnight at 30°C in the presence of IgAP produced by another member of the lab.

3.1 pBAD-IgAP Cloning and Expression

3.1.1 pBAD-IgAP Cloning

The *H. influenzae* IgAP passenger domain (residues 26-989) was successfully ligated into the pBAD vector. Sequencing of the pBAD-IgAP construct revealed that the insert was ligated in the correct orientation and that no unintended mutations were introduced.

3.1.2 pBAD-IgAP Expression

Expression of the pBAD-IgAP construct at 37^{o}C results in a large band, corresponding to the expected size of the IgAP passenger domain, the amount of protein produced appears to demonstrate a positive correlation with the amount of L-(+)-arabinose used to induce the culture.

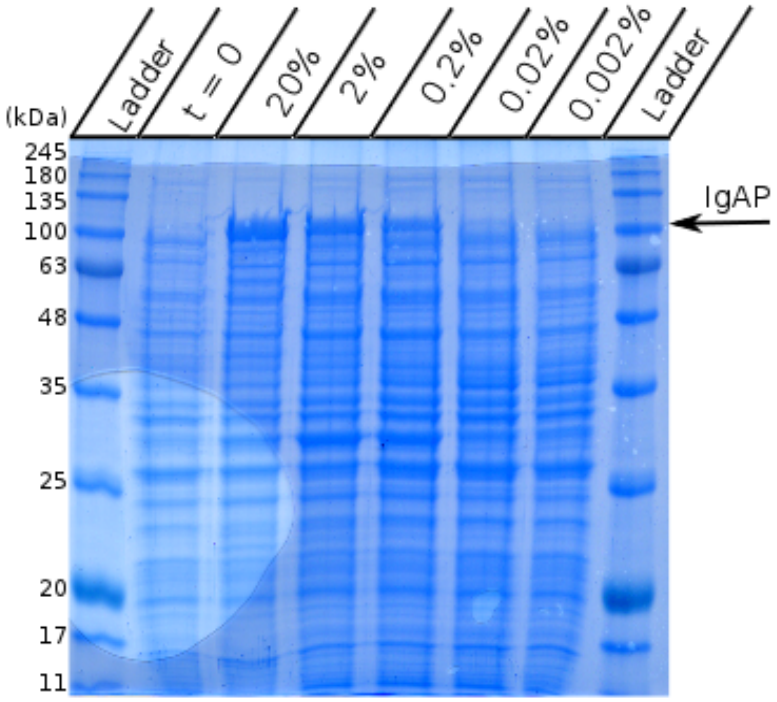

Figure 3-1 12% SDS-PAGE showing total protein at 0 and 4 hours after induction with 0.002-20% L-(+)-arabinose.

Analysis of the periplasmic fraction using osmotic shock to lyse the outer membrane followed by TCA precipitation to concentrate the samples followed by SDS-PAGE analysis did not yield any detectable protein corresponding to the expected size of IgAP. Incubation of the total and periplasmic fractions with IgA did not result in any detectable cleavage of the IgA heavy chain corresponding to IgAP activity.

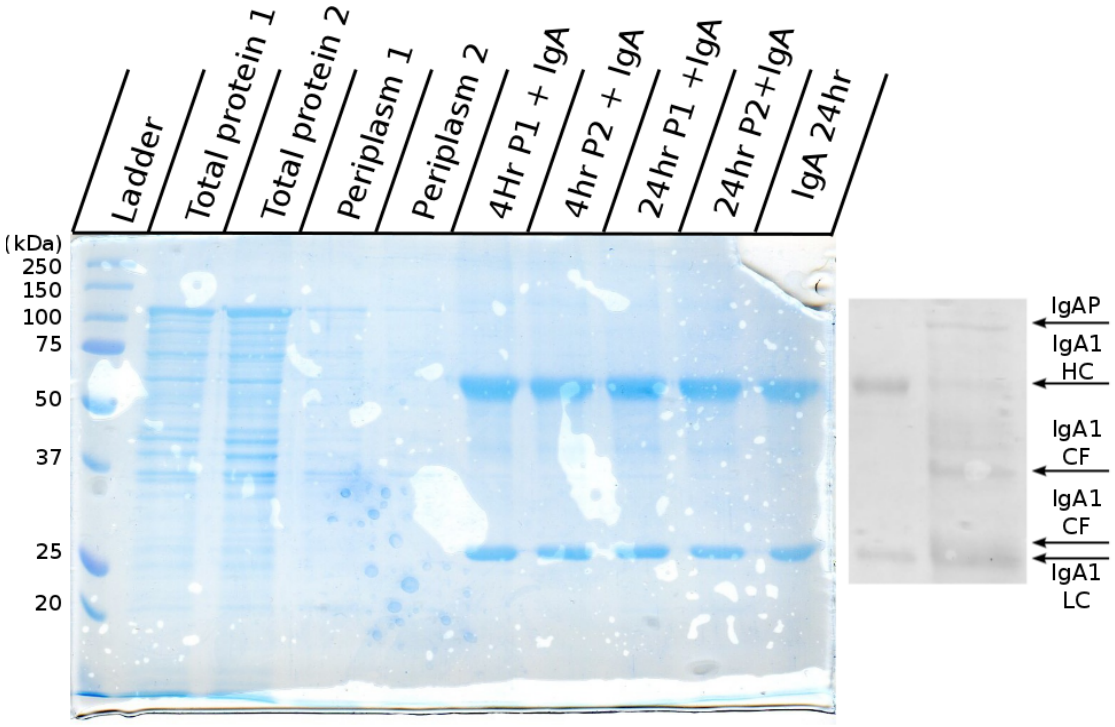

Figure 3-2 12% SDS-PAGE containing total and periplasmic protein as well as IgA incubated with the periplasmic fractions.

The roughly 100 kDa band slightly visible in the periplasmic fractions is attributed to spillover from the lanes containing the total protein fractions and not to any soluble IgA found in the periplasm.

The inability to produce any detectable amount of protein by SDS-PAGE or the IgA cleavage assay and the presence of large amounts of insoluble protein indicates that the protein is being expressed but that misfolding is occurring. It is likely that cleavage of the signal sequence could be the rate limiting factor if the rate of production by the ribosomes overwhelms the machinery of the secretory pathway leading to a surplus of IgAP containing a signal sequence in the cytoplasm. The obvious remedy would be to reduce the rate of protein production by either lowering the temperature of the expression or by lowering the amount of L-arabinose or some combination of both. However as seen

in **Figure 3-1** this decreases the rate of production to such an extent that no protein can be detected soluble or otherwise. Potentially other signal peptides including the native *H. influenzae* IgAP in addition to other signal peptides could have been incorporated into the construct to increase the efficiency of protein over-expression.

3.2 pET24(b)-IgAP and pDUET Soluble Expression

The SDS-PAGE analysis of the expression conditions performed by Matthew Macleod as part of his Biology 499 research project are located in **Appendix B**, the results of those gels are summarized here.

The expression conditions carried out in attempt to generate soluble IgAP using the pET24(b)-IgAP and pDUET-IgAP-MAP constructs were unable to produce any soluble protein as detectable by SDS-PAGE. Attempts to purify the soluble fraction via Ni^{2+} chromatography did not result in any soluble protein visible by SDS-PAGE, in addition the products of these expressions were unable to cleave the heavy chain of IgA1. Attempts at replicating the optimal expression conditions outlined in a previous study, which reported yields of 20-40mg/L, were unsuccessful in generating any soluble protein.

3.3 pET24(b)-IgAP Insoluble Expression and Refolding

3.3.1 Generating a Construct with Fewer Methionine Residues

DNA sequencing of a construct containing mutations: M74L, M117R, M167L, M635L, M647L and M655L and another containing mutations M74L, M117R, M635L, M647L and M655L revealed that all intended mutations were successfully introduced and that no unintended mutations were present. Sequencing data aligned against the IgAP coding sequence is included in **Appendix A**.

3.3.2 pET-24(b)-IgAPΔM

The pET24(b)-IgAPΔM construct was successfully transformed into BL-21(DE3) *E. coli* cells, SDS-PAGE analysis of a culture induced with 1mM IPTG revealed that 4 hours after induction, a large band approximately 100kDa in size appeared.

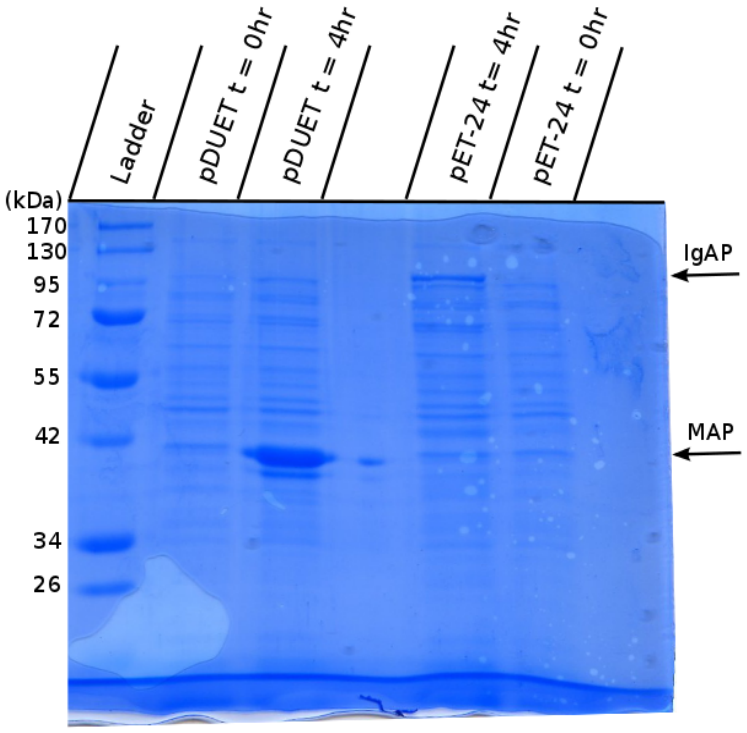

Figure 3-3 Expression of pDUET and pET-24(b) IgA1P constructs induced with 1mM IPTG at 37°C for 4 hours.

The pDUET construct succeeded in producing a large amount of MAP but no detectable amount of IgAP.

3.3.3 Isolation and Purification of Inclusion Bodies

No soluble IgAP was detected by SDS-PAGE or any activity by incubation with IgA1, however there was a large amount of IgAP found in the insoluble fraction likely present as inclusion bodies.

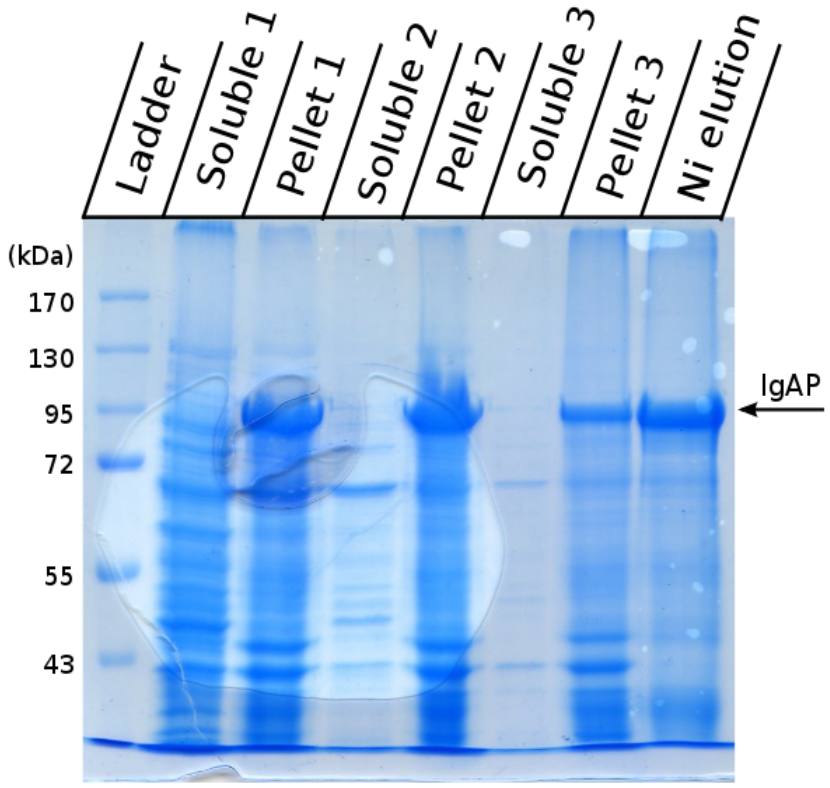

Figure 3-4 10% SDS-PAGE illustrating the isolation and purification of inclusion bodies from *E. coli* BL-21(DE3) cells transformed with the pET-24(b)-IgAPΔM construct.

Washing of the insoluble pellet with detergent to remove lipids and lipid soluble proteins resulted in inclusion bodies with a high degree of purity, Ni^{2+} affinity chromatography resulted in further enrichment of IgAP.

3.3.4 Cleavage of the N-terminal Methionine Using CNBr

To measure the rate and specificity of CNBr cleavage of IgAP a construct omitting the M167L mutation was generated. Incubation with 50mg/mL of CNBr at room temperature showed that the cleavage reaction goes to completion after 4 hours and that no non-specific cleavage occurs after at least 24 hours. This demonstrates that cleavage quickly and with high specificity.

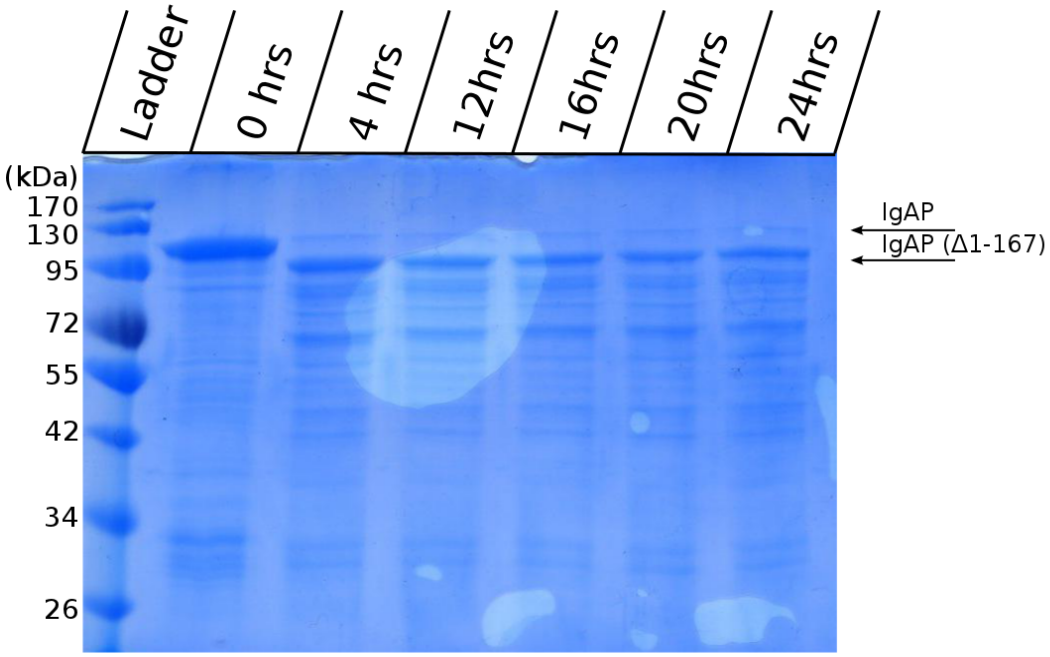

Figure 3-5 10% SDS-PAGE showing IgAP solubilized in 6M Gu-HCl, pH0 incubated for 24 hours with 50mg/mL CNBr to measure the specificity and efficiency of the cleavage reaction.

Incubation of IgAP purified in the same manner, but lacking any methionine other than the N-terminal methionine, does not result in a decrease in size visible by SDS-PAGE as demonstrated by **Figure 3-4** in which a construct containing no methionine residues (IgAPΔM) was subjected to the same chemical treatment.

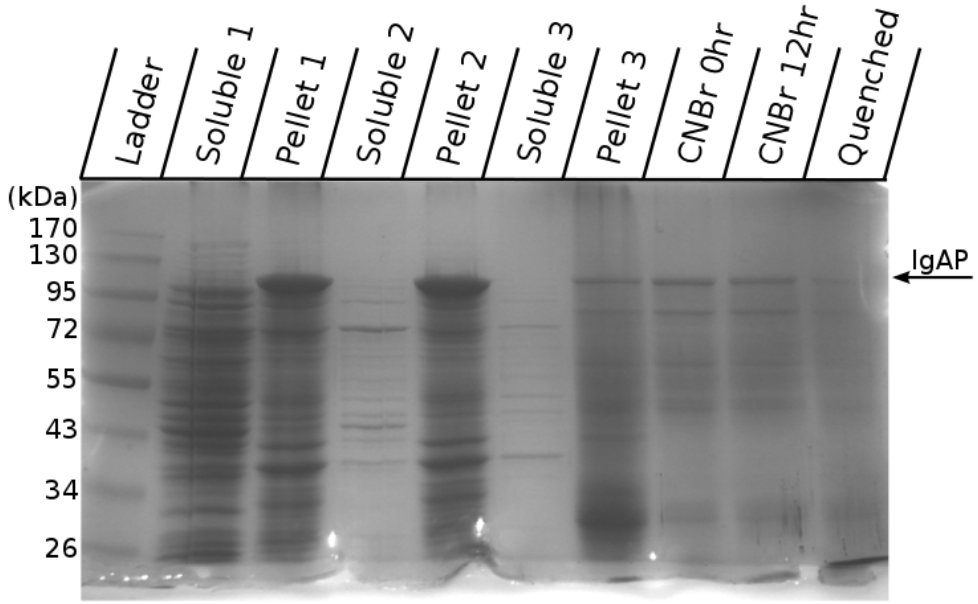

Figure 3-6 10% SDS-PAGE demonstrating the purification, cleavage and quenching of IgAPΔM.

Quenching of the CNBr reaction with 3M TRIS appears to result in a loss of sample but this is attributed to the high concentration of salts interfering with the TCA precipitation, the amount of protein in the sample does not appear to decrease as monitored by the solution's absorbance at 280nm (A_{280})..

3.3.5 Refolding of CNBr Cleaved Protein and IgA Cleavage Assay

Condition A was the only on-column refolding condition employed and was the first method attempted for refolding as we believed that immobilization of the C-terminus would mimic the natural folding mechanism of ATs. After running a gradient from 8M – 0.5M Urea and eluting the column with 300mM imidazole there was no protein detected by A_{280} as well as SDS-PAGE .

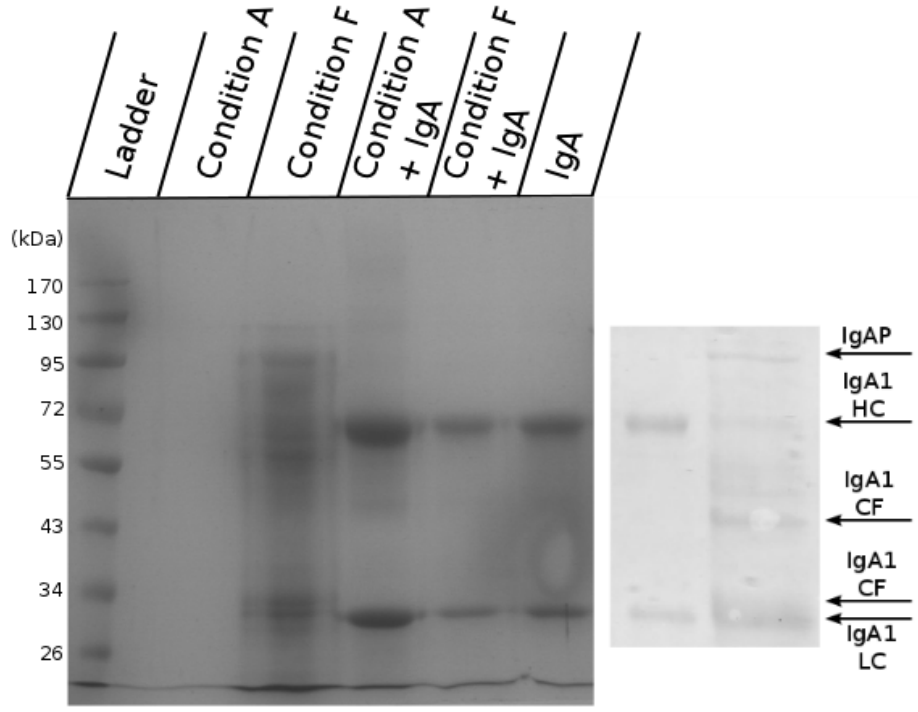

Figure 3-7 10% SDS-PAGE analysis of refolding conditions A **and** F **and IgA cleavage assay**

Unsurprisingly in the absence of any detectable protein **Condition A** was unable to cleave the heavy chain of IgA1.

Condition F was the simplest dialysis experiment attempted for refolding and contained no redox buffer system (GSH-GSSG) or other osmolytes (L-Arg). While there was some detectable soluble protein a large amount of precipitate formed overnight in the dialysis tubing resulting in a decrease of the band corresponding to IgAP as measured by SDS-PAGE. This first attempt at refolding by dialysis resulted in a sample which could not cleave the IgA1 heavy chain.

Condition G was a dialysis method that contained a redox buffer system consisting of GSH and GSSG, once again a large amount of precipitate formed in the dialysis tubing the sample resulting from the clarification of the dialysis by centrifugation was unable to cleave the IgA1 heavy chain as seen in **Figure 3-8.**

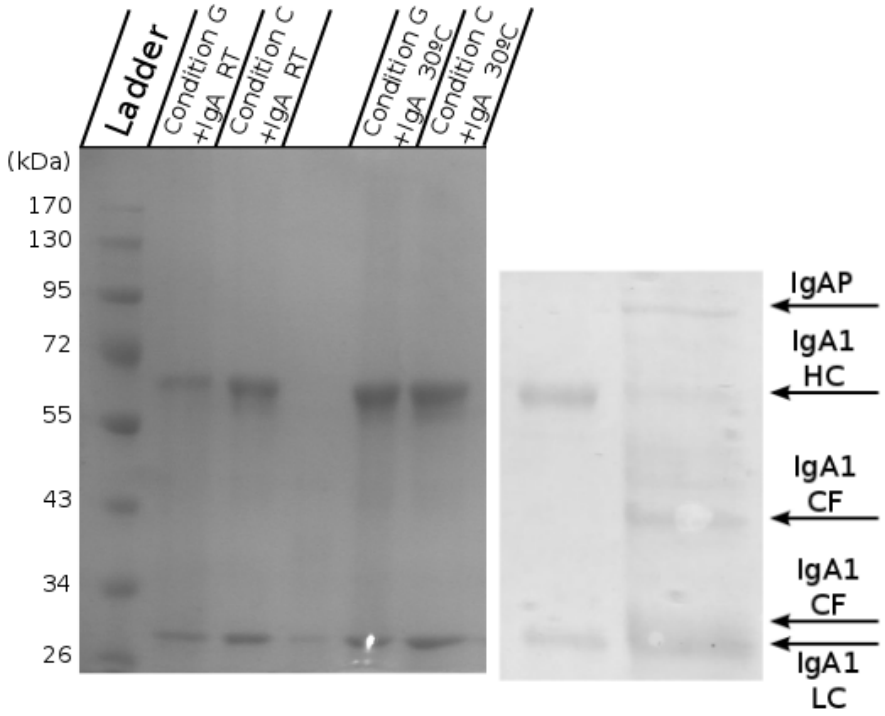

Figure 3-8 10% SDS-PAGE containing the IgA cleavage assay results for Condition G **and** C**.**

Condition C was a direct dilution method, addition of CNBr cleaved protein to the refolding resulted in precipitation, after centrifugation to pellet the insoluble precipitate and concentration no protein was detectable and the sample was unable to cleave IgA.

Condition B was another direct dilution method however in this attempt an endpoint of 0M Urea was chosen as opposed to 0.5M as in **Condition C**, precipitate formed once again after addition of the CNBr cleaved protein as demonstrated in **Figure 3-9** this sample was unable to cleave the IgA1 heavy chain.

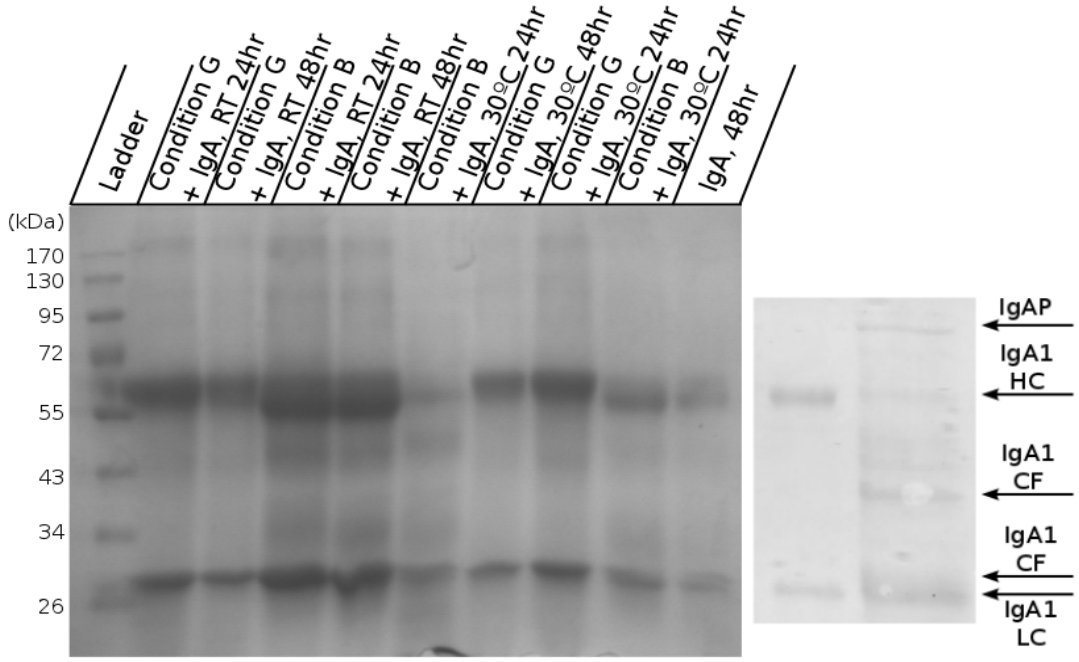

Figure 3-9 10% SDS-PAGE containing the IgA cleavage assays for Condition G **and** B**.**

Refolding **Condition G** was a dialysis method, which used a GSH-GSSG redox buffering system and an endpoint of 0.5M Urea. This sample produced a significant amount of precipitate after incubation overnight and was unable to cleave the IgA1 heavy chain, even after 48 hours of incubation.

Conditions D and **E** were the final direct dilution methods attempted for refolding of CNBr cleaved IgAP, both **Condition D** and **E** contained a GSH-GSSG redox buffer while **Condition E** also contained 0.4M L-Arg as an osmolyte. Once again both conditions produced a large amount of precipitate, analysis by SDS-PAGE did not indicate any soluble protein present in the sample, incubation of both samples with IgA was insufficient to affect cleavage of the heavy chain.

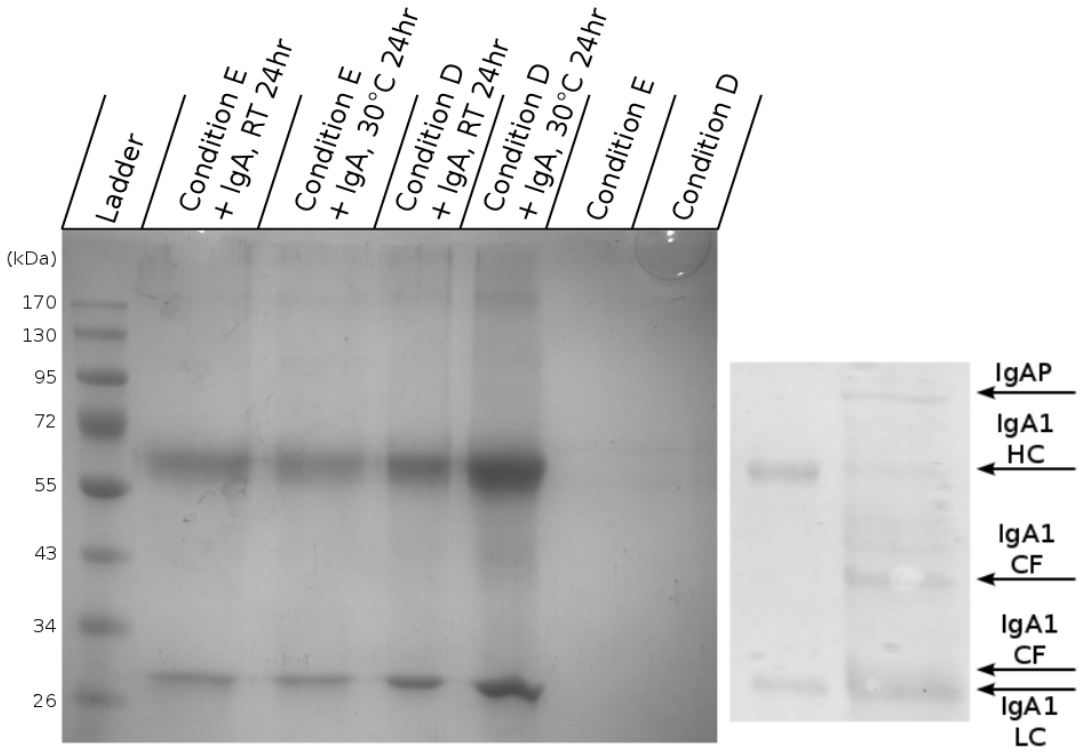

Figure 3-10 10% SDS-PAGE analysis of refolding Conditions D **and** E **via the IgA heavy chain cleavage assay.**

Condition C was a direct dilution method containing GSH-GSSG as a redox buffer, addition of the cleaved IgAP solution to the refolding solution resulted in a white precipitate. As shown in **Figure 3-9 Condition C** is unable to cleave the IgA1 heavy chain.

The direct dilution methods (**Conditions B**, **C**, **D** and **E**) for attempted refolding of IgA protease appear to result in complete loss of the sample due to precipitation caused either by the inadequacy of the refolding buffer to promote folding or from the inability of the protein to adopt it's native structure due to the six amino substitutions required for CNBr cleavage. As there are an almost limitless number of possible additives and permutations possible for refolding buffers given that the conditions attempted covered the general classes of additives (redox buffers and osmolytes) no further conditions were attempted.

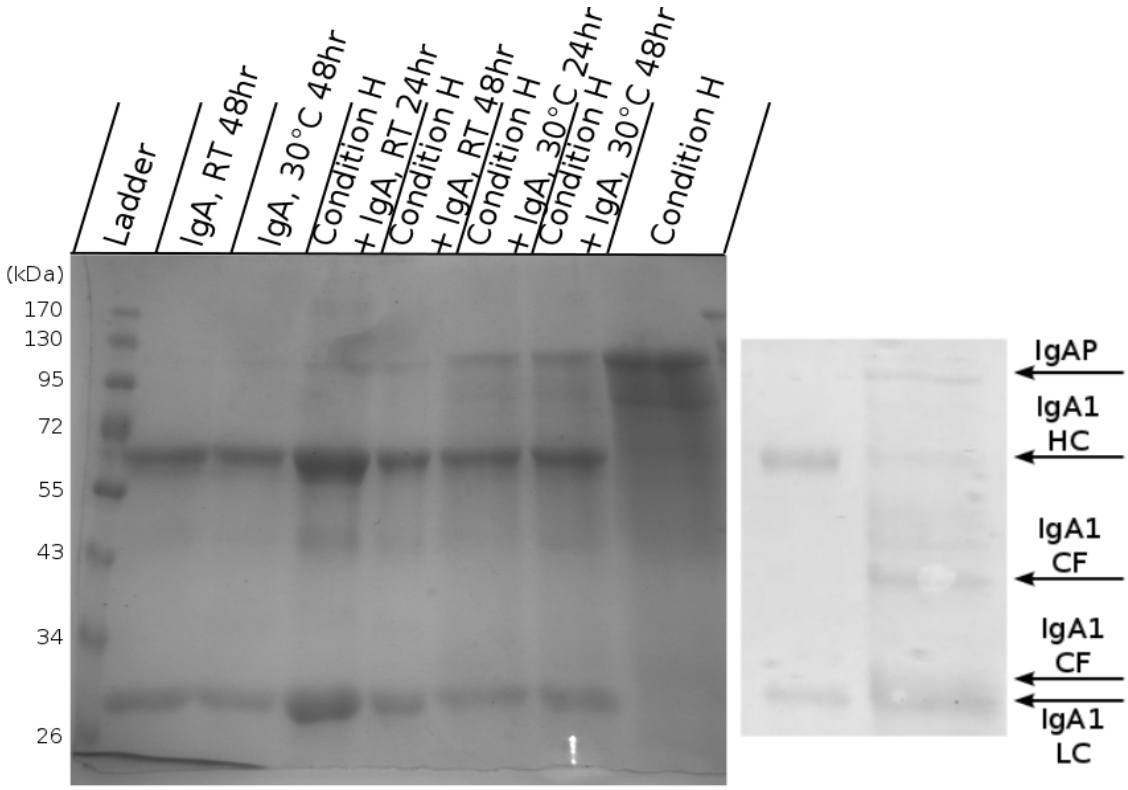

Figure 3-11 10% SDS-PAGE analysis of Condition C **and** H **IgA cleavage assay.**

Condition H employed dialysis with an endpoint of 0.5M urea in addition GSH-GSSG was included as well as L-Arg. This condition produced precipitate in the dialysis tubing but there was still enough protein to be detected by A_{280} readings as well as SDS-PAGE as indicated by **Figure 3-11**. Incubation of this sample with IgA was none the less insufficient for cleavage of the IgA1 heavy chain.

Condition I was a dialysis method similar to **Condition H** except that dialysis was carried out in a stepwise manner with endpoints at 4M, 2M and finally 0.5M urea. Again some precipitate formed in the dialysis bag after incubation overnight, after pelleting the insoluble fraction there was a significant amount of soluble protein visible by SDS-PAGE and A_{280}.

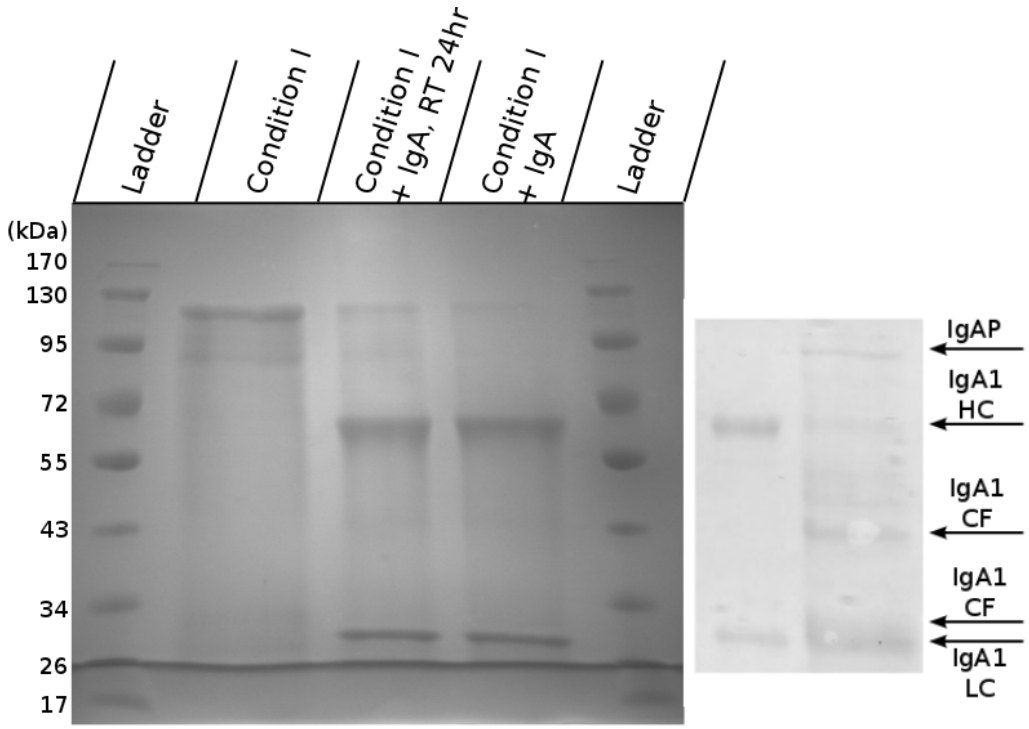

Figure 3-12 10% SDS-PAGE of Condition I **and IgA cleavage assay.**

Condition I produced the least precipitate of all attempted refolding conditions but was still unable to cleave the IgA1 heavy chain. This sample could be concentrated to a concentration of roughly 1.5mg/mL, higher than the samples produced from any other condition, before precipitate started to form on the walls of the centrifugal concentrator. To determine if the aggregation of this species was due to the protein adopting a non-native, but still soluble fold, rather than its native fold or if the six mutations introduced for CNBr cleavage had rendered the native fold unstable CD analysis of **Condition I** was undertaken.

3.3.6 Analysis of Refolded Protein by Circular Dichroism Spectrophotometry

As we're unable to produce significant quantities of the wild-type IgAP to use as a positive control a measurement of the relative abundance of secondary structure elements was obtained using the PDB of the wild-type enzyme and the 2Struc server (2Struc.cryst.bbk.ac.uk).

Table 3-1 Secondary structure analysis of *H. influenzae* IgAP (PDB: 3H09) analyzed by the 2Struc server using the DSSP method.

Structural Element	%
Helix	8.1
Sheet	38.1
Other	52.9

Unsurprisingly there is a larger amount of beta strands and sheets and a smaller amount of α-helices due to the large β-helical stalk and the relatively smaller serine protease domain.

The protein solution produced from **Condition I** was successfully dialyzed into 10 mM PO_4, the blank solution (10 mM PO_4, pH 7.5) produced a small but noticeable signal and was used to normalize the spectra of **Condition I**.

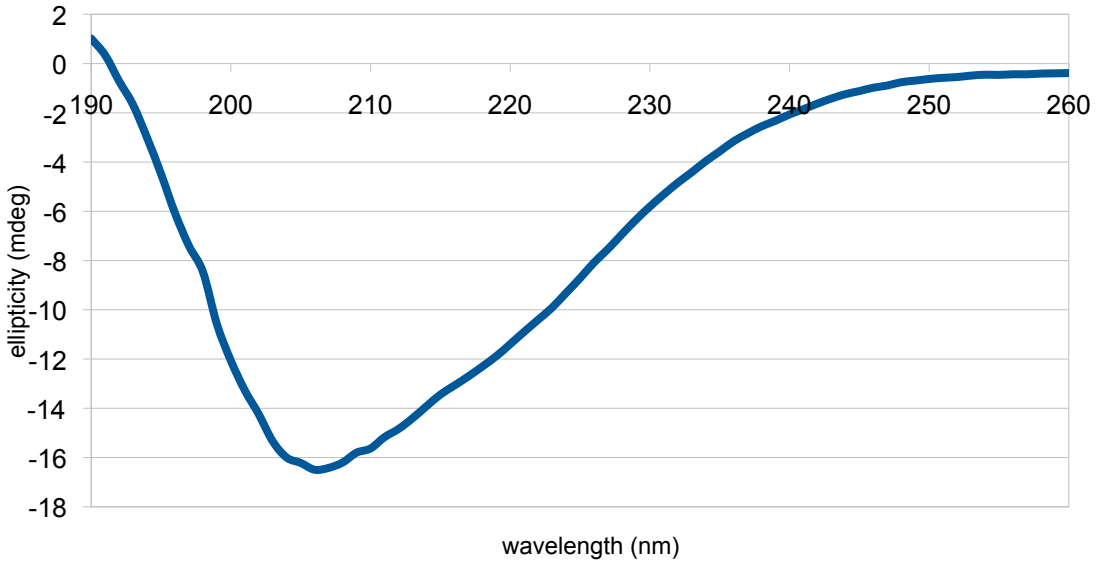

Figure 3-13 Plot of ellipticity vs. wavelength generated from a sample of IgAP Condition I **in 10 mM phosphate, collected on a Jasco J-715 CD spectrophotometer at room temperature.**

The spectra was smoothed using the Jasco spectra analyzer software provided with the instrument and this data was used as the input for analysis using the DichroWeb server to generate an estimate of secondary structure elements.

The programs, which were used for data analysis were SELCON 3, CONTIN and CDSSTR, reference sets 4, 7 and 10 were used for those programs requiring a reference.

Table 3-2 Results of DichroWeb analysis of Condition I CD spectra.

Program	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total	Reference
SELCON3	0.071	0.137	0.181	0.079	0.205	0.383	1.056	4
CONTIN	0.013	0.074	0.227	0.124	0.231	0.331	1	4
CONTIN	0.018	0.058	0.201	0.107	0.187	0.428	0.999	7
CONTIN	0	0.079	0.24	0.136	0.129	0.415	0.999	10
CDSSTR	0	0.09	0.22	0.13	0.24	0.31	0.99	4
CDSSTR	0	0.02	0.27	0.12	0.17	0.38	0.96	7
AVERAGE	0.0107	0.076	0.223	0.116	0.194	0.375	1.001	
STDEV	0.028	0.038	0.031	0.021	0.041	0.046	0.031	

The results of all the successful runs were averaged and the results were plotted against the calculated values generated from the 2Struc server using PDB 3H09 as input.

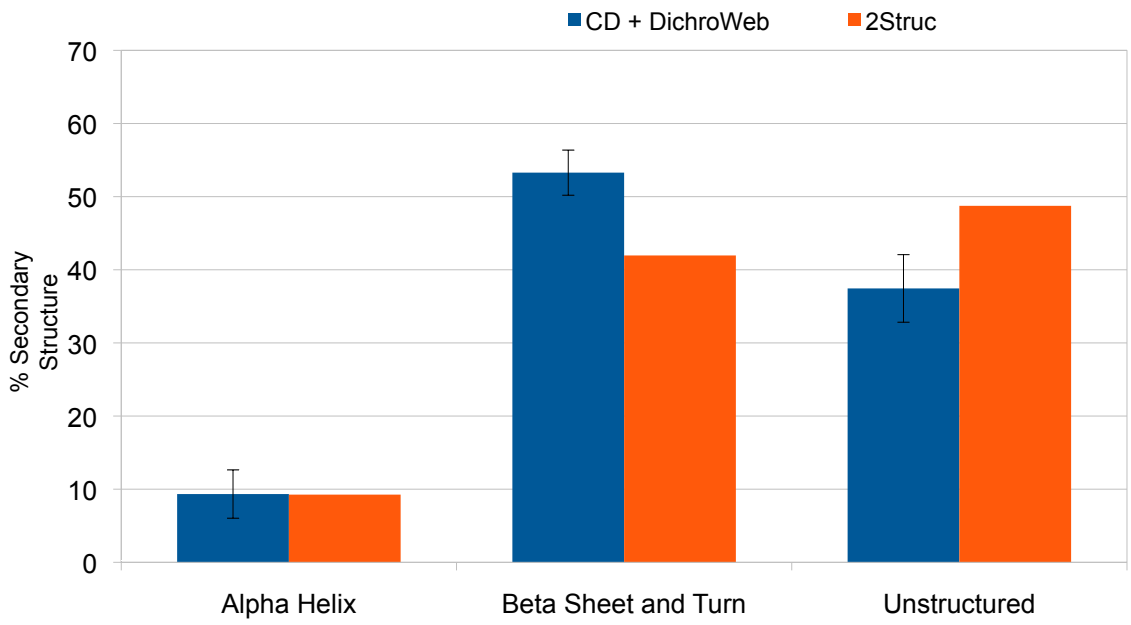

Figure 3-14 Plot of the relative proportion of secondary structure elements as calculated from the PDB:3H09 using the 2Struc server and calculated from the CD spectra of IgAP refolded by Condition I **using the DichroWeb server.**

The results of both methods appear close however the large degree of error present in the measurement of β-sheet/turn and unstructured regions by 2Struc analysis makes the interpretation of the results somewhat ambiguous. However based on the similarity of the secondary structure elements between the wild-type IgAP and methionine deficient IgAP refolded by **Condition I** it is likely that they adopt a similar structure. This indicates that the inability of the refolded enzyme to cleave the IgA1 hinge region is not due to the misfolding of the enzyme but because the six substitutions introduced for CNBr cleavage have abrogated its catalyitc activity. Also of note is the fact that IgAPΔM precipitates in solution above a concentration of 1mg/mL whereas the wild-type enzyme was stable in solution up to roughly 10mg/ml.

Chapter 4 Conclusions and Future Research

4.1 Conclusions

The work presented in this thesis provides several possible methods for expression of proteins in *E. coli*, particularly those that are unable to tolerate an N-terminal methionine residue as well as type V autotransporters. These methods are best suited to bacterial proteins that do not require post-translational modifications.

The pBAD-IgAP construct, under the conditions tested, was unable to generate soluble protein in the periplasm to an extent detectable by SDS-PAGE and the IgA cleavage assay but instead resulted in large amounts of insoluble protein in the cytosol. The pET24(b)-IgAP construct failed to produce an amount of soluble protein detectable by SDS-PAGE or the IgA cleavage assay under the expression parameters carried out. The pET24(b)-IgAP construct was able to produce a large amount of protein as cytoplasmic inclusion bodies, which can be isolated and purified readily. Refolding **Condition H** and **I** were able to produce a somewhat soluble species, stable to a concentration of roughly 1mg/mL, but was unable to cleave the IgA1 heavy chain. In this instance refolding by dialysis appears to be more effective than either on-column refolding or direct dilution. Based on the analysis of the CD spectra collected on this potentially refolded species appears to have a similar secondary structure composition to the wild-type folded protein. This indicates that the relative instability and catalytic deficiency is caused not by the inability of refolding solution to affect refolding but due to the six mutations introduced to allow for CNBr cleavage.

4.2 Future Research

Determining a method for the expression of *H. influenzae* IgAP will require further optimization of the strategies we have adopted here or completely new strategies. The pBAD-IgAP expression system could be optimized by substituting the present signal sequence with any number of other gram-negative signal peptides. The pET-IgAP expression system could be optimized by the supplementing the expression media with any number of additives not limited to cofactors, vitamins, metals and carbon or nitrogen sources. There are also several cell strains which could be used to

potentially generate soluble protein, auto-induction media in which both glucose and lactose is present could be used. This system does not require the addition of an inducer at a specific time, rather expression is induced once glucose is depleted as glucose is antagonistic to lactose in its induction of the *lac* operon. The CNBr cleavage strategy using the pET-IgAP construct is perhaps the least versatile of the attempted strategies, the major aspect which could be optimized is the composition of refolding buffer. There a large number of pH buffers, osmolytes and other additives which could be employed, the denaturant urea could be replaced by a detergent. However based on the CD analysis refolding may not be possible due to the methioinine substitution mutations.

Once acceptable conditions for expression are determined more work to develop a method for the expression of the *H. Influenzae* IgAP on an industrial scale would enable its use as a treatment for IgA nephropathy. Presently there is work being conducted in the pharmaceutical into large scale expression of *H. Influenzae* IgAP, based on publicly available data there has not been any clinical trials studying IgAP as a therapeutic against IgA nephropathy. However there has certainly been interest and IgAP is certainly considered a prospective therapeutic for the treatment of IgA accumulation in the kidneys.

The SPATEs and the related IgA proteases of *H. Influenzae* and *Neisseria* demonstrate extreme substrate selectivity, work is needed to determine which structural elements confer them with this high degree of selectivity. In the case of *H. Influenzae* IgAP, biochemical studies into the particularly role of Loop C and D as well as Domain 2, 3 and 4 are needed to understand their role in determining substrate preference. Deletions of theses loops and globular domains as well as their substitution for analogous regions in related enzymes and the generation of selected amino acid substitutions in these regions could shed light on their function. Determining the crystal structure of the IgAP-IgA1 binary complex would greatly aid the understanding of substrate binding and the catalytic process of the enzyme. This is not be an easy task given the size of both proteins and the potential homogeneity of the IgA1 glycans but would provide a wealth of information about substrate recognition and binding as well as the structure of IgA1 itself. There is presently no crystal structure of the entire protein likely due to the flexible hinge regions flexibility preventing crystallization, binding and immobilization of the IgA1 hinge by IgAP may aid in crystallization.

The role of IgAP in *H. influenzae* infection is still being actively explored and new substrates such as LAMP1 may yet be discovered. There are no animal models for studying IgAP in infection, owing to the fact that only the great apes possess an IgA containing the all important hinge region and developing a disease model in apes would face numerous ethical and practical challenges. There are murine models of *H. influenzae* infection, which are generally useful in studying the process of infection, however they are particularly ill-suited to study the role of IgAP. Presently the most advanced model able to study the role of IgAP is based on a human epithelial carcinoma cell-line, the development of an infection model use non-immortalized, primary cells from the human respiratory tract would present a more relevant model of infection.

Appendix A

DNA Sequencing Data

pET24(b)-IgAPΔM Sequencing

```
IgAP        gcgttagtgagagacgatgtggattatcaaatatttcgtgattttgcagaaaataaaggg
pETiga      gcgttagtgagagacgatgtggattatcaaatatttcgtgattttgcagaaaataaaggg
            ************************************************************

IgAP        agattttctgttggtgcaacaaatgtggaagtgagagataaaaataaccactctttaggc
pETiga      agattttctgttggtgcaacaaatgtggaagtgagagataaaaataaccactctttaggc
            ************************************************************

IgAP        aatgttttacctaatggcattccgatgattgattttagtgttgtggatgtagataaacgc
pETiga      aatgttttacctaatggcattccgttgattgattttagtgttgtggatgtagataaacgc
            ************************ ************************************

IgAP        atcgccacattgataaatccacaatatgtagtaggtgtaaaacacgttagtaacggcgtg
pETiga      atcgccacattgataaatccacaatatgtagtaggtgtaaaacacgttagtaacggcgtg
            ************************************************************

IgAP        agtgaactacattttgggaacttaaatggcaatatgaataatggcaatgctaaatcgcac
pETiga      agtgaactacattttgggaacttaaatggcaataggaataatggcaatgctaaatcgcac
            ********************************** ***************************

IgAP        cgagatgtatcttcagaagaaaatagatatttttccgttgagaaaaatgagtatccaact
pETiga      cgagatgtatcttcagaagaaaatagatatttttccgttgagaaaaatgagtatccaact
            ************************************************************

IgAP        aaattgaatggaaaagcagtaactactgaagatcaaactcaaaaacgccgtgaagactac
pETiga      aaattgaatggaaaagcagtaactactgaagatcaaactcaaaaacgccgtgaagactac
            ************************************************************

IgAP        tatatgccacgtcttgataaatttgttaccgaagttgcaccaatagaggcttcaactgca
pETiga      tatctgccacgtcttgataaatttgttaccgaagttgcaccaatagaggcttcaactgca
            *** ********************************************************

IgAP        agtagtgatgctggcacatataatgatcagaataaatatcctgcttttgtaagactagga
pETiga      agtagtgatgctggcacatataatgatcagaataaatatcctgcttttgtaagactagga
            ************************************************************

IgAP        agtggtagtcaatttatttataaaaaaggagataattacagcttaattttaaataatcat
pETiga      agtggtagtcaatttatttataaaaaaggagataattacagcttaattttaaataatcat
            ************************************************************

IgAP        gaggttggaggcaataatcttaaattggtgggcgatgcctatacctatggtattgcaggc
pETiga      gaggttggaggcaataatcttaaattggtgggcgatgcctatacctatggtattgcaggc
            ************************************************************

IgAP        acaccttataaagtaaaccacgaaaataatggactaattggttttggcaattcaaaagag
```

```
pETiga      acaccttataaagtaaaccacgaaaataatggactaattggttttggcaattcaaaagag
            ************************************************************

IgAP        gaacacagcgatccaaaaggaatattatctcaagatccgcttaccaattatgctgtttta
pETiga      gaacacagcgatccaaaaggaatattatctcaagatccgcttaccaattatgctgtttta
            ************************************************************

IgAP        ggcgacagtggctccccattatttgtatatgatagagaaaaaggaaaatggctttttctt
pETiga      ggcgacagtggctccccattatttgtatatgatagagaaaaagg-aaatggctttttctt
            ******************************************** ***************

IgAP        gggtcttatgatttttgggcaggttataacaaaaaatcttggcaagaatggaatatttat
pETiga      gggtcttatgatttttgggcaggttataacaaaaaatcttggcaagaatggaatatttat
            ************************************************************

IgAP        aaacctgaatttgcaaaaactgttctagataaagatactgcaggttctttaactggttct
pETiga      aaacctgaatttgcaaaaactgttctagataaagatactgcaggttctttaactggttct
            ************************************************************

IgAP        aacacccaatacaattggaatcctactggcaaaacaagcgttatttctaatggttctgaa
pETiga      aacacccaatacaattggaatcctactggcaaaacaagcgttatttctaatggttctgaa
            ************************************************************

IgAP        tctctaaatgttgatttattcgatagtagtcaggatactgactctaagaagaacaatcac
pETiga      tctctaaatgttgatttattcgatagtagtcaggatactgactctaagaagaacaatcac
            ************************************************************

IgAP        ggaaaaagtgtgactcttagaggaagtggaacgcttaccttaaataataatatcgatcaa
pETiga      ggaaaaagtgtgactcttagaggaagtggaacgcttaccttaaataataatatcgatcaa
            ************************************************************

IgAP        ggcgcaggcggcttgttctttgaaggagattatgaagttaaaggcacttctgatagtacc
pETiga      ggcgcaggcggcttgttctttgaaggagattatgaagttaaaggcacttctgatagtacc
            ************************************************************

IgAP        acttggaaaggagctggcgtttctgttgctgatggaaaaacagtaacgtggaaagtacat
pETiga      acttggaaaggagctggcgtttctgttgctgatggaaaaacagtaacgtggaaagtacat
            ************************************************************

IgAP        aacccgaaatctgatcgtttagctaaaatcggcaaaggaacattaattgtagaaggaaag
pETiga      aacccgaaatctgatcgtttagctaaaatcggcaaaggaacattaattgtagaaggaaag
            ************************************************************

IgAP        ggagaaaataaaggttcgctaaaagtgggcgatggtactgttatcttaaaacaacaagct
pETiga      ggagaaaataaaggttcgctaaaagtgggcgatggtactgttatcttaaaacaacaagct
            ************************************************************

IgAP        gatgccaataataaagttaaagccttttcacaagtaggtatagtaagtggtcgctcaact
pETiga      gatgccaataataaagttaaagccttttcacaagtaggtatagtaagtggtcgctcaact
            ************************************************************

IgAP        gttgtacttaatgatgataagcaagtagatccaaattccatttactttggctttagaggt
pETiga      gttgtacttaatgatgataagcaagtagatccaaattccatttactttggctttagaggt
```

```
            ************************************************************

IgAP        ggtcgattagatgccaatggcaataatctcacttttgaacatatccgtaatattgatgat
pETiga      ggtcgattagatgccaatggcaataatctcacttttgaacatatccgtaatattgatgat
            ************************************************************

IgAP        ggcgcaagactagtaaatcacaataccagcaaaacctctactgtaacaattactggggaa
pETiga      ggcgcaagactagtaaatcacaataccagcaaaacctctactgtaacaattactggggaa
            ************************************************************

IgAP        agtctaattacagatccaaatacaattactccatataatatagacgcaccagatgaagat
pETiga      agtctaattacagatccaaatacaattactccatataatatagacgcaccagatgaagat
            ************************************************************

IgAP        aatccttatgcctttcgacggattaaagatggaggacagctctatttaaatttggaaaat
pETiga      aatccttatgcctttcgacggattaaagatggaggacagctctatttaaatttggaaaat
            ************************************************************

IgAP        tacacttattatgcgttaagaaaaggtgcgagcactcgttcagaattacctaaaaatagt
pETiga      tacacttattatgcgttaagaaaaggtgcgagcactcgttcagaattacctaaaaatagt
            ************************************************************

IgAP        ggcgaaagcaatgaaaattggctatatatgggtaaaacttccgatgaagccaaaagaaat
pETiga      ggcgaaagcaatgaaaattggctatatttgggtaaaacttccgatgaagccaaaagaaat
            *************************** ********************************

IgAP        gtaatgaaccatatcaacaacgagcgtatgaatggctttaacggttattttggcgaggaa
pETiga      gtattgaaccatatcaacaacgagcgtttgaatggctttaacggttattttggcgaggaa
            *** ********************** *********************************

IgAP        gagggtaaaaataacggtaatctaaatgtgacttttaaaggcaaaagtgagcaaaatcgc
pETiga      gagggtaaaaataacggtaatctaaatgtgacttttaaaggcaaaagtgagcaaaatcgc
            ************************************************************

IgAP        tttttattaacaggcggaacaaaccttaatggcgatttaaaggttgaaaaaggcacatta
pETiga      tttttattaacaggcggaacaaaccttaatggcgatttaaaggttgaaaaaggcacatta
            ************************************************************

IgAP        ttcctttctggcagaccaacaccgcacgcaagagatattgcaggtatttcttcgacaaaa
pETiga      ttcctttctggcagaccaacaccgcacgcaagagatattgcaggtatttcttcgacaaaa
            ************************************************************

IgAP        aaagatcaacactttgctgaaaataatgaagtggtagtagaagatgactggattaaccgc
pETiga      aaagatcaacactttgctgaaaataatgaagtggtagtagaagatgactggattaaccgc
            ************************************************************

IgAP        aattttaaagcaacaaatattaatgtaaccaataacgcaacccctttattcaggtcgcaat
pETiga      aattttaaagcaacaaatattaatgtaaccaataacgcaacccctttattcaggtcgcaat
            ************************************************************

IgAP        gttgcaaacattacttcaaatatcacagcttctgataatgcaaaagtacatattggctat
pETiga      gttgcaaacattacttcaaatatcacagcttctgataacgcaaaagtacatattggctat
            ************************************ ***********************
```

```
IgAP        aaagcaggcgataccgtttgtgtacgttctgactatacgggctatgtgacttgcactact
pETiga      aaagcaggcgataccgtttgtgtacgttctgactatacgggctatgtgacttgcactact
            ************************************************************

IgAP        gacaagttatccgataaagcccttaatagctttaacgccaccaatgtatctggcaatgta
pETiga      gacaagttatccgataaagcccttaatagctttaacgccaccaatgtatctggcaatgta
            ************************************************************

IgAP        aatttatcaggtaatgcaaactttgtcttaggcaaagctaacttattcggcacaattagc
pETiga      aatttatcaggtaatgcaaactttgtcttaggcaaagctaacttattcggcacaattagc
            ************************************************************

IgAP        ggcacgggaaatagccaagtacgtttaaccgaaaatagccattggcatttaacaggcgat
pETiga      ggcacgggaaatagccaagtacgtttaaccgaaaatagccattggcatttaacaggcgat
            ************************************************************

IgAP        agcaatgttaatcagttaaatttagacaaggggcatattcatttaaatgcacaaaacgat
pETiga      agcaatgttaatcagttaaatttagacaaggggcatattcatttaaatgcacaaaacgat
            ************************************************************

IgAP        gcaaataaagtaactacatataacacgctgactgtgaatagcttatcaggtaacggttct
pETiga      gcaaataaagtaactacatataacacgctgactgtgaatagcttatcaggtaacggttct
            ************************************************************

IgAP        ttctattatttaactgatctttccaataaacaaggcgacaaagttgttgtaactaaatcc
pETiga      ttctattatttaactgatctttccaataaacaaggcgacaaagttgttgtaactaaatcc
            ************************************************************

IgAP        gccacaggtaactttacattacaagtggcagataaaacaggcgagcctacaaaaaatgaa
pETiga      gccacaggtaactttacattacaagtggcagataaaacaggcgagcctacaaaaaatgaa
            ************************************************************

IgAP        ctcacgcttttgatgcgtcaaatgctacaagaaataatttgaatgtgtcattagttggg
pETiga      ctcacgcttttgatgcgtcaaatgctacaagaaataatttgaatgtgtcattagttggg
            ************************************************************

IgAP        aataccgttgatttaggtgcttggaaatataaattacgtaatgttaatggacgttacgat
pETiga      aataccgttgatttaggtgcttggaaatataaattacgtaatgttaatggacgttacgat
            ************************************************************

IgAP        ttgtataaccca
pETiga      ttgtataaccca
            ************
```

Appendix B

SDS-PAGE Results of pET24(b)-IgAP Expression

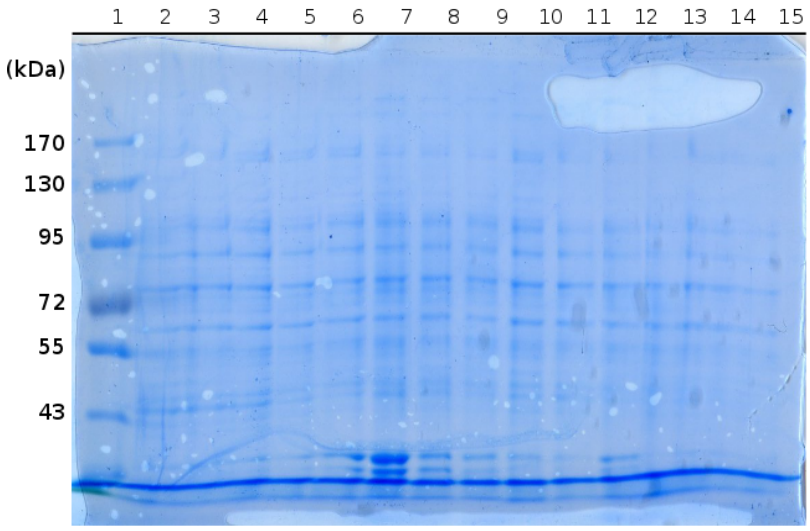

Figure B-4-1 12%SDS-PAGE containing the results of pET-24(b)-IgAP expression in BL-21(DE3) cells conducted at 37°C.

Lane	Condition
1	Ladder
2	1.0 mM IPTG t=0
3	5.0 mM IPTG t=0
4	0.1 mM IPTG t=1
5	1.0 mM IPTG t=1
6	5.0 mM IPTG t=1
7	0.1 mM IPTG t=2
8	1.0 mM IPTG t=2
9	5.0 mM IPTG t=2
10	0.1 mM IPTG t=3
11	1.0 mM IPTG t=3
12	5.0 mM IPTG t=3

13	0.1 mM IPTG t=4
14	1.0 mM IPTG t=4
15	5.0 mM IPTG t=4

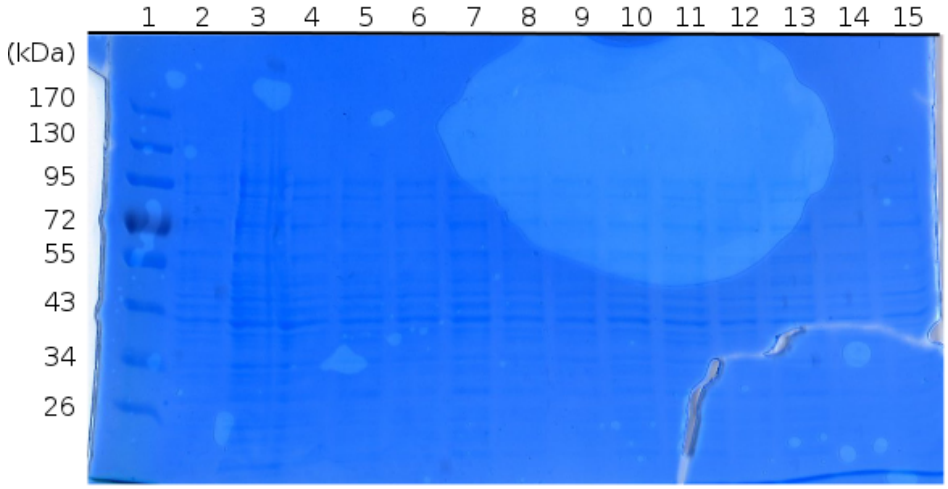

Figure B-4-2 12% SDS-PAGE analysis of pET2(b)-IgAP expression in BL-21(DE3) cells conducted at 28°C

Lane	Condition
1	Ladder
2	1.0 mM IPTG t=0
3	5.0 mM IPTG t=0
4	0.1 mM IPTG t=1
5	1.0 mM IPTG t=1
6	5.0 mM IPTG t=1

7	0.1 mM IPTG t=2
8	1.0 mM IPTG t=2
9	5.0 mM IPTG t=2
10	0.1 mM IPTG t=3
11	1.0 mM IPTG t=3
12	5.0 mM IPTG t=3
13	0.1 mM IPTG t=4
14	1.0 mM IPTG t=4
15	5.0 mM IPTG t=4

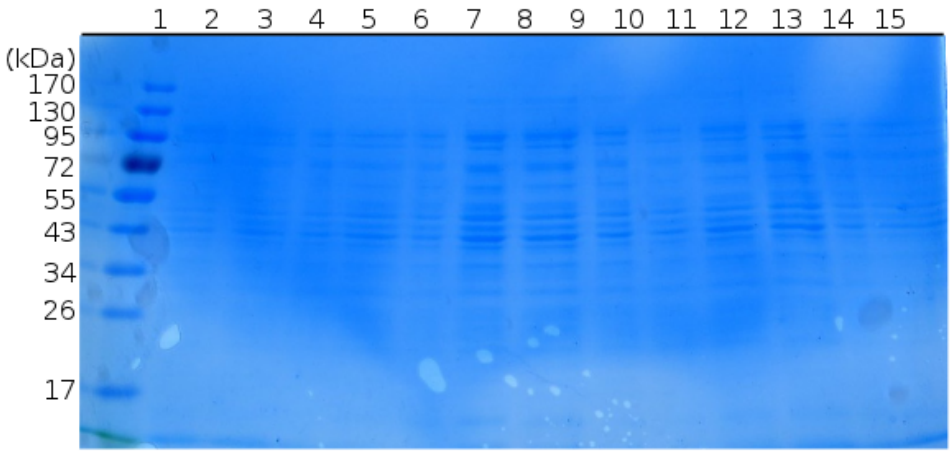

Figure B-4-3 12% SDS-PAGE analysis of pET24(b)-IgAP expression in BL-21(DE3) cells conducted at 21°C

Lane	Condition
1	Empty
2	Ladder
3	1.0 mM IPTG t=0
4	5.1 mM IPTG t=0
5	0.1 mM IPTG t=1

6	1.0 mM IPTG t=1
7	5.0 mM IPTG t=1
8	0.1 mM IPTG t=2
9	1.0 mM IPTG t=2
10	5.0 mM IPTG t=2
11	0.1 mM IPTG t=3
12	1.0 mM IPTG t=3
13	5.0 mM IPTG t=3
14	0.1 mM IPTG t=4
15	1.0 mM IPTG t=4

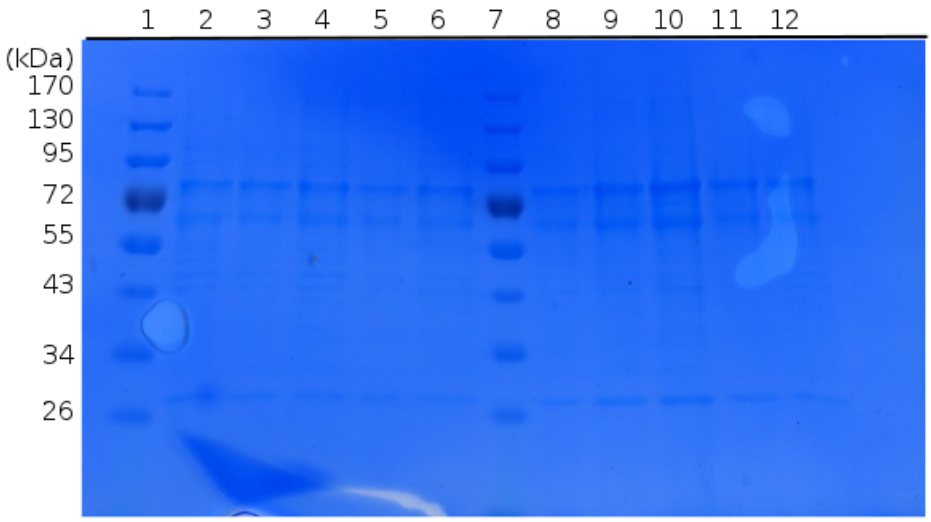

Figure B-4-4 10% SDS-PAGE analysis of IgA cleavage assay conducted with samples in Figure4.2-1.

Lane	Condition
1	Ladder
2	Negative Control t=0 + IgA1
3	Negative Control t=1 + IgA1
4	Negative Control t=2 + IgA1

5	Negative Control t=3 + IgA1
6	Negative Control t=4 + IgA1
7	Dye Marker
8	1.0mM IPTG t=0 + IgA1
9	1.0mM IPTG t=1 + IgA1
10	1.0mM IPTG t=2 + IgA1
11	1.0mM IPTG t=3 + IgA1
12	1.0mM IPTG t=4 + IgA1

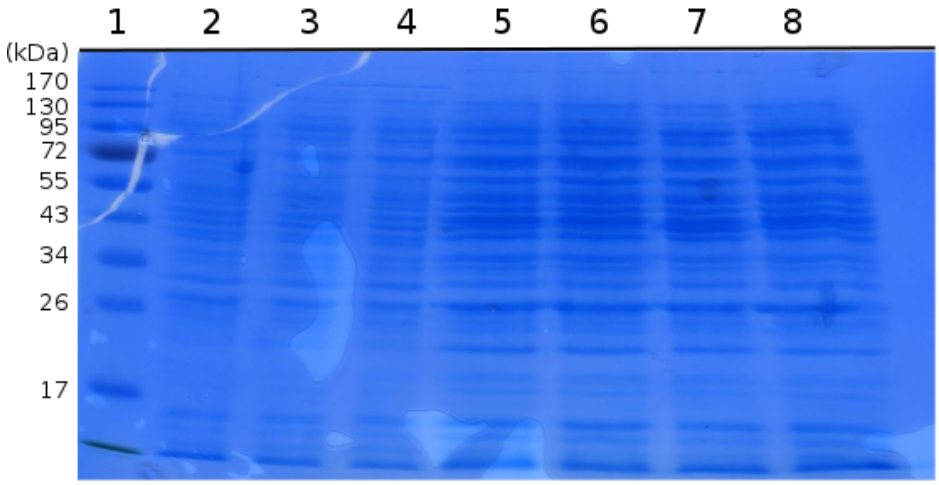

Figure B-4-5 12% SDS-PAGE analysis of pET-24(b)-IgAP expression in C41(DE3) cells carried out at 21°C in superbroth.

Lane	Condition
1	Ladder
2	LB Negative Control t=24
3	LB trial 1 t=24
4	LB trial 2 t=24
5	SuperBroth trial 1 t=0
6	SuperBroth trial 1 t=24
7	SuperBroth trial 2 t=0
8	SuperBroth 2 t=24

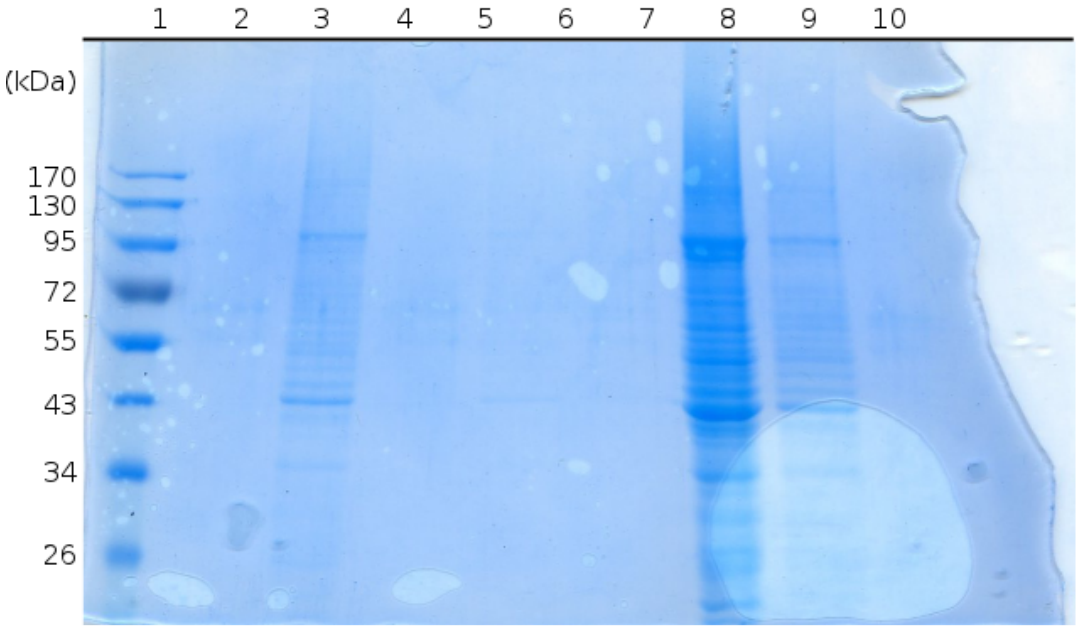

Figure B-4-6 10% SDS-PAGE analysis of the results of Ni^{2+} affinity chromatography of the cultures shown in Figure 4.2-5.

Lane	Condition
1	Ladder
2	Soluble Fraction + Nickel Wash
3	Insoluble Lysate + Nickel Wash
4	Soluble Fraction + Nickel Wash + Imidazole
5	Insoluble Lysate + Nickel Wash + Imidazole
6	Soluble Fraction Nickel Resin
7	Insoluble Fraction Nickel Resin
8	Soluble Fraction Nickel Resin post Imidazole
9	Insoluble Fraction Nickel Resin post Imidazole
10	Empty

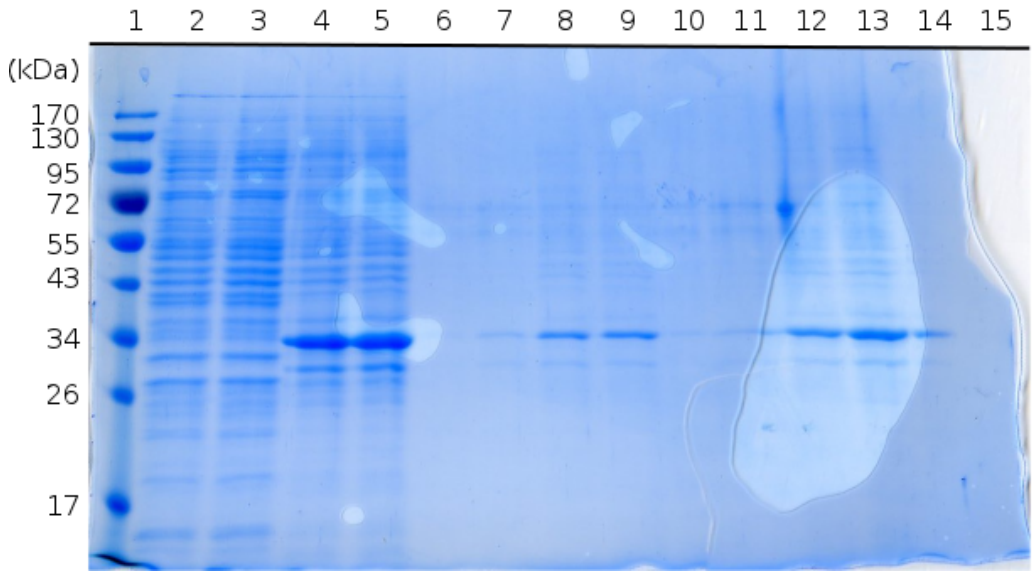

Figure B-4-7 10% SDS-PAGE analysis of the pET24(b)-IgAP construct expressed in C41(DE3) cells carried out at 21°C in LB.

Lane	Condition
1	Ladder
2	Trial 1 - Soluble Fraction
3	Trial 2 - Soluble Fraction
4	Trial 1 - Insoluble Lysate + 1mL 25mM Hepes
5	Trial 2 - Insoluble Lysate + 1mL 25mM Hepes
6	Trial 1 - Soluble Fraction + Nickel Wash
7	Trial 2 - Soluble Fraction + Nickel Wash
8	Trial 1 - Insoluble Lysate + Nickel Wash
9	Trial 2 - Insoluble Lysate + Nickel Wash
10	Empty
11	Empty
12	Trial 1 - Soluble Fraction Nickel Resin
13	Trial 2 - Soluble Fraction Nickel Resin
14	Trial 1 - Insoluble Lysate Nickel Resin
15	Trial 2 - Insoluble Lysate Nickel Resin

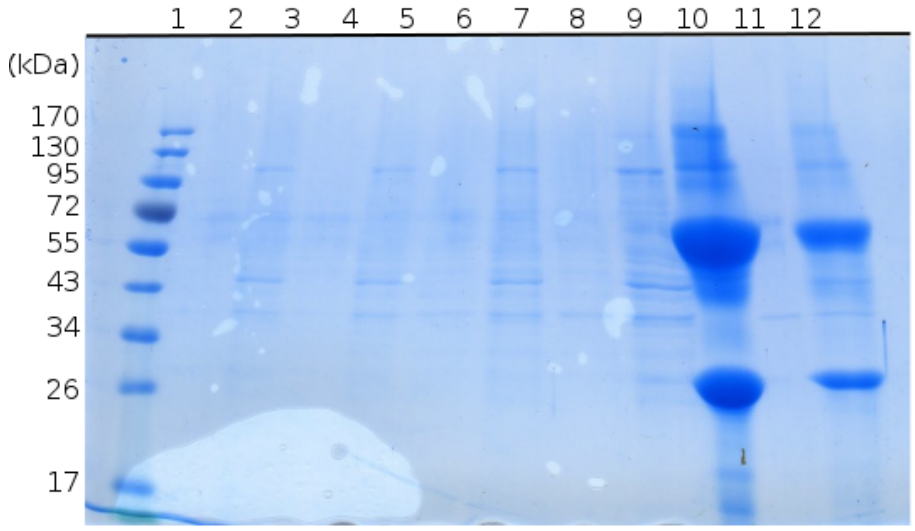

Figure B-4-8 10% SDS-PAGE analysis of the IgA cleavage assay using samples from Figure 4.2-6 and Figure 4.2-7.

Lane	Condition
1	Ladder
2	Soluble Fraction + Nickel Wash
3	Insoluble lysate + Nickel Wash
4	Soluble Fraction + Nickel Wash + Imidazole
5	Insoluble lysate + Nickel Wash + Imidazole
6	Soluble Fraction Nickel Resin
7	Insoluble lysate Nickel Resin
8	Soluble Fraction + Nickel Resin post Imidazole
9	Insoluble lysate + Nickel Resin post Imidazole
10	IgA1 Control
11	Insoluble Fraction + Nickel Wash
12	Insoluble lysate + Nickel Wash + IgA1

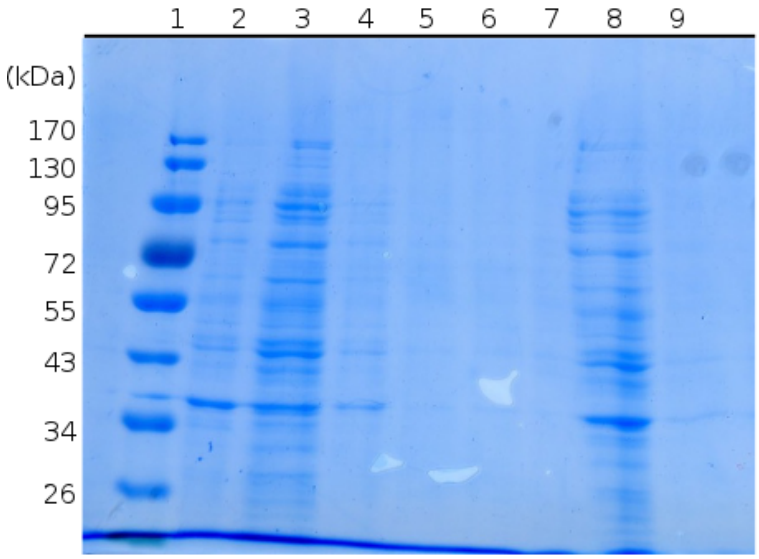

Figure B-4-9 10% SDS-PAGE analysis of pET24(b)-IgAP expression in BL-21(DE3) cells in M9 minimal media carried out at 20°C followed by $Ni^{2=}$ affinity chromatography.

Lane	Condition
1	Ladder
2	Insoluble Lysate + 1mL 25mM Hepes
3	Soluble Fraction
4	Insoluble Lysate + Nickel Wash
5	Soluble Fraction + Nickel Wash
6	Insoluble Lysate + Nickel Wash + Imizadole
7	Soluble Fraction + Nickel Wash + Imizadole
8	Insoluble Lysate Resin
9	Soluble Fraction Resin

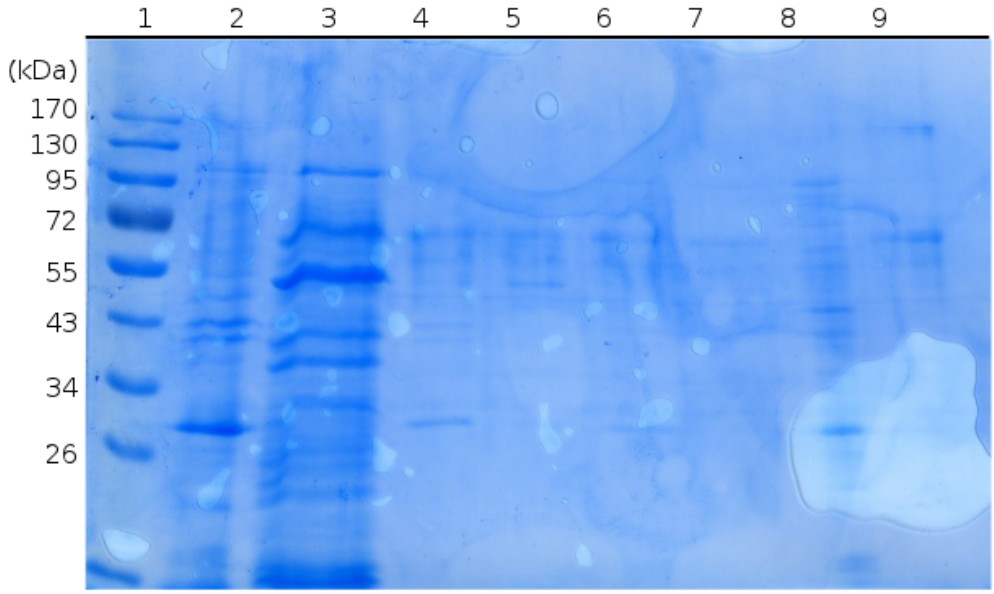

Figure B-4-10 10% SDS-PAGE analysis of pET24(b)-IgAP expression in Bl-21(DE3)pLysS cells carried out in superbroth at 20°C followed by $Ni^{2=}$ affinity chromatography.

Lane	Condition
1	Ladder
2	Insoluble Lysate + 1mL 25mM Hepes
3	Soluble Fraction
4	Insoluble Lysate + Nickel Wash
5	Soluble Fraction + Nickel Wash
6	Insoluble Lysate + Nickel Wash + Imizadole
7	Soluble Fraction + Nickel Wash + Imizadole
8	Insoluble Lysate Resin
9	Soluble Fraction Resin

Bibliography

Apodaca, G., Bomsel, M., Arden, J., Breitfeld, P. P., Tang, K., & Mostov, K. E. (1991). The Polymeric Immunoglobulin Receptor. *The Journal of Clinical Investigation*, *87*(March), 1877–1882.

Barnard, T. J., Dautin, N., Lukacik, P., Bernstein, H. D., & Buchanan, S. K. (2007). Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nature Structural & Molecular Biology*, *14*(12), 1214–20.

Barnard, T. J., Gumbart, J., Peterson, J. H., Noinaj, N., Easley, N. C., Dautin, N., … Buchanan, S. K. (2012). Molecular basis for the activation of a catalytic asparagine residue in a self-cleaving bacterial autotransporter. *Journal of Molecular Biology*, *415*(1), 128–42.

Basu, A., Li, X., & Leong, S. S. J. (2011). Refolding of proteins from inclusion bodies: rational design and recipes. *Applied Microbiology and Biotechnology*, *92*(2), 241–51.

Batten, M. R.,Senior, B. W., Kilian, M., & Woof, J. M. (2003). Amino Acid Sequence Requirements in the Hinge of Human Immunoglobulin A1 (IgA1) for Cleavage by Streptococcal IgA1 Proteases, *71*(3), 1462–1469.

Bender, M. L. (1960). MECHANISMS OF CATALYSIS OF NUCLEOPHILIC REACTIONS. *Chemical Reviews*, *60*, 53–113.

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *Journal of Bacteriology*, *62*(3), 293–300. Retrieved from

Bjuggren, G., & Tunevall, G. (1952). Otitis in Childhood. *Acta Oto-Laryngologica*, *42*(4-5), 311–315.

Blow, D. M. (1997). The tortuous story of Asp His Ser structural analysis of α-chymotrypsin. *Trends in Biochemical Sciene*, *102*(22), 405–408.

Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969). Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature*, *221*, 337–340.

Bode, W., Meyer, E., & Powers, J. C. (1989). Perspectives in Biochemistry Human Leukocyte and Porcine Pancreatic Elastase : X-ray Crystal Structures ,. *Biochemistry*, *28*(5), 74–88.

Brandon, L. D., Goehring, N., Janakiraman, A., Yan, A. W., Wu, T., Beckwith, J., & Goldberg, M. B. (2003). IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed. *Molecular Microbiology*, *50*(1), 45–60.

Brandon, L. D., & Goldberg, M. B. (2001). Periplasmic Transit and Disulfide Bond Formation of the Autotransported Shigella Protein IcsA Periplasmic Transit and Disulfide Bond Formation of the Autotransported Shigella Protein IcsA, *183*(3).

Burton, J., Wood, S. G., Lynch, M., & Plaut, G. (1988). Substrate Analogue Inhibitors of the IgAl Proteinases from, (1), 1647–1651.

Butt, T. R., Edavettal, S. C., Hall, J. P., & Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*, *43*(1), 1–9.

Chodirker, A. W. B., & Tomasi, T. B. (1963). Gamma-Globulins Quantitative Relatlonships in Human Serum and Nonvascular Fluids. *Science*, *12*(3595), 1080–1081.

Choudary, S. K., Qiu, J., Plaut, A. G., & Kritzer, J. a. (2013). Versatile substrates and probes for IgA1 protease activity. *Chembiochem : A European Journal of Chemical Biology*, *14*(15), 2007–12.

Crisel, R. M., Baker, R. S., & Dorman, D. E. (1975). Capsular Polymer of Haemophilus infelunzae, Type b: I. Strucutural Characterization of the Capsular Strain of Serotype Eagan. *The Journal of Biological Chemistry*, *250*(13), 4926–4930.

Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., & Kim, J. F. (2009). Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). *Journal of Molecular Biology*, *394*(4), 634–43.

Dautin, N. (2010). Serine protease autotransporters of enterobacteriaceae (SPATEs): biogenesis and function. *Toxins*, *2*(6), 1179–206.

Dautin, N., & Bernstein, H. D. (2007). Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annual Review of Microbiology*, *61*, 89–112.

De Navasquez, S. (1942). ACUTE LARYNGITIS AND SEPTICAEMIA. *British Medical Journal*, (1928), 187–188.

Deshpande, N., Jensen, P. H., Packer, N. H., & Kolarich, D. (2010). GlycoSpectrumScan : Fishing Glycopeptides from MS Spectra of Protease Digests of Human Colostrum sIgA research articles, (1), 1063–1075.

Desvaux, M., Parham, N. J., & Henderson, I. R. (2004). The autotransporter secretion system. *Research in Microbiology*, *155*(2), 53–60.

Dixon, M. M., Brennan, R. G., & Matthews, B. W. (1991). Structure of gamma-chymotrypsin in the range pH 2.0 to pH 10.5 suggests that gamma-chymotrypsin is a covalent acyl-enzyme adduct at low pH. *International Journal of Biological Macromolecules*, *13*(2), 89–96.

Dodson, G., & Wlodawer, a. (1998). Catalytic triads and their relatives. *Trends in Biochemical Sciences*, *23*(9), 347–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9787641

Domingo Meza-Aguilar, J., Fromme, P., Torres-Larios, A., Mendoza-Hernández, G., Hernandez-Chiñas, U., Arreguin-Espinosa de Los Monteros, R. a, … Fromme, R. (2014). X-ray crystal structure of the passenger domain of plasmid encoded toxin(Pet), an autotransporter enterotoxin from enteroaggregative Escherichia coli (EAEC). *Biochemical and Biophysical Research Communications*, *445*(2), 439–44.

Dubendorff, J. W., & Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of Molecular Biology*, *219*(1), 45–59.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. a, Kirkness, E. F., Kerlavage, a R., … Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (New York, N.Y.)*, *269*(5223), 496–512.

Frandsen, E. V, Kjeldsen, M., & Kilian, M. (1997). Inhibition of Prevotella and Capnocytophaga immunoglobulin A1 proteases by human serum. *Clinical and Diagnostic Laboratory Immunology*, *4*(4), 458–64.

Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V, & Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16422–7.

Gkentzi, D., Slack, M. P. E., & Ladhani, S. N. (2012). The burden of nonencapsulated Haemophilus influenzae in children and potential for prevention. *Current Opinion in Infectious Diseases*.

Graf, L., Craik, C. S., Patthy, a, Roczniak, S., Fletterick, R. J., & Rutter, W. J. (1987). Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry*, *26*(9), 2616–23.

Guzman, L. M., Belin, D., Carson, M. J., Beckwith, J., Guzman, L., Belin, D., & Carson, M. J. (1995). Tight regulation , modulation , and high-level expression by vectors containing the arabinose PBAD promoter . These include : Tight Regulation , Modulation , and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter, *177*(14).

Hallström, T., & Riesbeck, K. (2010). Haemophilus influenzae and the complement system. *Trends in Microbiology*, *18*(6), 258–65.

Hanson, L. Å., & Korotkova, M. (2002). The role of breastfeeding in prevention of neonatal infection. *Seminars in Neonatology*, *7*(4), 275–281.

Hartley, D. L., & Kane, J. F. (1986). Recovery and Reactivation of Recombinant Proteins, *16*, 101–102.

Hauck, C. R., & Meyer, T. F. (1997). The lysosomal/phagosomal membrane protein h-lamp-1 is a target of the IgA1 protease of Neisseria gonorrhoeae. *FEBS Letters*, *405*(1), 86–90.

Henderson, I. R., & Navarro-Garcia, F. (2004). Type V Protein Secretion Pathway : the Autotransporter Story, *68*(4), 692–744.

Henderson, R. (1970). Structure of crystalline alpha-chymotrypsin. IV. The structure of indoleacryloyl-alpha-chyotrypsin and its relevance to the hydrolytic mechanism of the enzyme. *Journal of Molecular Biology*, *54*(2), 341–54.

Jain, S., & Goldberg, M. B. (2007). Requirement for YaeT in the outer membrane assembly of autotransporter proteins. *Journal of Bacteriology*, *189*(14), 5393–8.

Johnson, T. a, Qiu, J., Plaut, A. G., & Holyoak, T. (2009). Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of haemophilus influenzae immunoglobulin A1 protease. *Journal of Molecular Biology*, *389*(3), 559–74.

Jones, H. W. E. (1937). H. influenzae Meningitis. *The British Medical Journal*, 4–6.

Jose, J., Wölk, U., Lorenzen, D., Wenschuh, H., & Meyer, T. F. (2000). Human T-cell response to meningococcal immunoglobulin A1 protease associated alpha-proteins. *Scandinavian Journal of Immunology*, *51*(2), 176–85.

Junker, M., Besingi, R. N., & Clark, P. L. (2009). Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Molecular Microbiology*, *71*(5), 1323–32.

Junker, M., Schuster, C. C., McDonnell, A. V, Sorg, K. a, Finn, M. C., Berger, B., & Clark, P. L. (2006). Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(13), 4918–23.

Kapust, R. B., & Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science : A Publication of the Protein Society*, *8*(8), 1668–74.

Kerr, M. a. (1990). The structure and function of human IgA. *The Biochemical Journal*, *271*(2), 285–96.

Khan, S., Mian, H. S., Sandercock, L. E., Chirgadze, N. Y., & Pai, E. F. (2011). Crystal structure of the passenger domain of the Escherichia coli autotransporter EspP. *Journal of Molecular Biology*, *413*(5), 985–1000.

Kilian, M., Mestecky, J., & Russell, M. W. (1988). Defense mechanisms involving Fc-dependent functions of immunoglobulin A and their subversion by bacterial immunoglobulin A proteases. *Microbiological Reviews*, *52*(2), 296–303.

Klauser, T., Joachim, K., Otzelberger, K., Pohlner and Meyer, T. F. (1993). Characterization of the Neisseria Iga βcore The Essential Unit for Outer Membrane Targeting and Extracellular Secretion. *Journal of Molecular Biology*, *234*(3), 579–593.

Knowles, T. J., Scott-Tucker, A., Overduin, M., & Henderson, I. R. (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nature Reviews. Microbiology*, *7*(3), 206–14.

Koshland, M. E. (1985). The coming of age of the immunoglobulin J chain. *Annual Review of Immunology*, *3*, 425–53.

Leong, S. S. J., & Middelberg, A. P. J. (2007). A Simplified Bioprocess for Human Alpha-Fetoprotein Production From Inclusion Bodies, *97*(1), 99–117.

Leyton, D. L., Rossiter, A. E., & Henderson, I. R. (2012). From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nature Reviews. Microbiology*, *10*(3), 213–25.

Lin, L., Ayala, P., Larson, J., Mulks, M., Fukuda, M., Carlsson, S. R., … So, M. (1997). The Neisseria type 2 IgA1 protease cleaves LAMP1 and promotes survival of bacteria within epithelial cells. *Molecular Microbiology*, *24*(5), 1083–94.

Loeb, M. R., Connor, E., & Penney, D. (1988). A comparison of the adherence of fimbriated and nonfimbriated Haemophilus influenzae type b to human adenoids in organ culture. *Infection and Immunity*, *56*(2), 484–9.

Luirink, J., & Sinning, I. (2004). SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta*, *1694*(1-3), 17–35.

Mandrell, R. E., McLaughlin, R., Aba Kwaik, Y., Lesse, a, Yamasaki, R., Gibson, B., … Apicella, M. a. (1992). Lipooligosaccharides (LOS) of some Haemophilus species mimic human glycosphingolipids, and some LOS are sialylated. *Infection and Immunity*, *60*(4), 1322–8.

Martoglio, B., & Dobberstein, B. (1998). Signal sequences: more than just greasy peptides. *Trends in Cell Biology*, *8*(10), 410–5.

Mattu, T. S., Pleass, R. J., Willis, a C., Kilian, M., Wormald, M. R., Lellouch, a C., … Dwek, R. a. (1998). The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fcα receptor interactions. *The Journal of Biological Chemistry*, *273*(4), 2260–72.

Mehta, S. K., Plaut, a G., Calvanico, N. J., & Tomasi, T. B. (1973). Human immunoglobulin A: production of an Fc fragment by an enteric microbial proteolytic enzyme. *Journal of Immunology (Baltimore, Md. : 1950)*, *111*(4), 1274–6.

Meng, G., Surana, N. K., St Geme, J. W., & Waksman, G. (2006). Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter. *The EMBO Journal*, *25*(11), 2297–304.

Morris, S. K., Moss, W. J., & Halsey, N. (2008). Haemophilus infl uenzae type b conjugate vaccine use and eff ectiveness, 435–443.

Murphy, T. F., Faden, H., Bakaletz, L. O., Kyd, J. M., Forsgren, A., Campos, J., … Pelton, S. I. (2009). Nontypeable Haemophilus influenzae as a pathogen in children. *The Pediatric Infectious Disease Journal*, *28*, 43–48.

Murphy, T. F., & Kirkham, C. (2002). Biofilm formation by nontypeable Haemophilus influenzae: strain variability, outer membrane antigen expression and role of pili. *BMC Microbiology*, *2*, 7.

Neidhardt, F. C., & Curtiss, R. (1996). *Escherichia coli and Salmonella : cellular and molecular biology*. *Journal Of Bacteriology* (Vol. 2, p. 2 v. (xx, 2822 p.)).

Nishimura, K., Yoon, Y. H., Kurihara, A., Unzai, S., Luirink, J., Park, S. Y., & Tame, J. R. H. (2010). Role of domains within the autotransporter Hbp/Tsh. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 12), 1295–300.

Novak, J., Julian, B. a, Tomana, M., & Mestecky, J. (2008). IgA glycosylation and IgA immune complexes in the pathogenesis of IgA nephropathy. *Seminars in Nephrology*, *28*(1), 78–87.

Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., & Gros, P. (2004). Structure of the translocator domain of a bacterial autotransporter. *The EMBO Journal*, *23*(6), 1257–66.

Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., … Tame, J. R. H. (2005). Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic Escherichia coli. *The Journal of Biological Chemistry*, *280*(17), 17339–45.

Perona, J. J., & Craik, C. S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Science : A Publication of the Protein Society*, *4*(3), 337–60.

Perona, J. J., & Craik, C. S. (1997). Evolutionary Divergence of Substrate Specificity within the Chymotrypsin-like Serine Protease Fold. *Journal of Biological Chemistry*, *272*(48), 29987–29990.

Peterson, J. H., Szabady, R. L., & Bernstein, H. D. (2006). An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. *The Journal of Biological Chemistry*, *281*(14), 9038–48.

Peterson, J. H., Tian, P., Ieva, R., Dautin, N., & Bernstein, H. D. (2010). Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17739–44.

Pittman, M. (1931). Variation and Type Specificity in the Bacterial Species Haemophilus Influenzae. *The Journal of Experimental Medicine*, *53*(3), 471–493.

Plaut, a G., Qiu, J., & St Geme, J. W. (2000). Human lactoferrin proteolytic activity: analysis of the cleaved region in the IgA protease of Haemophilus influenzae. *Vaccine*, *19 Suppl 1*, S148–52.

Plaut, A. G. (1983). The IgA1 proteases of pathogenic bacteria. *Annual Review of Microbiology*, *37*, 603–22.

Plaut, A. G., Flentke, G. R., Lynch, M., & Kettner, C. A. (1990). Lynch and C A Kettner Inhibition of IgAl Proteinases from Neisseria gonorrhoeae and Hemophilus influenxae by Peptide Prolyl Boronic Acids *. *Journal of Biological Chemistry*, *265*(7), 3738–3743.

Plaut, A. G., Gilbert, J. V, Artenstein, M. S., & Capra, J. D. (1975). Neisseria gonorrhoeae and Neisseria meningitidis : Extracellular Enzyme Cleaves Human Immunoglobulin A Neisseria gonorrhoeae and Neisseria meningitidis : Extracellular Enzyme Cleaves Human Immunoglobulin A. *Science*, *190*(4219), 1103–1105.

Plaut, A. G., Gilbert, J. V, & Heller, I. (1978). Assay and properties of IgA protease of Streptococcus sanguis. *Advances in Experimental Medicine and Biology*, *107*, 489–495.

Pohlner, J., Halter, R., Beyreuther, K. & Meyer, T. F. (1987). Gene Structure and Extracellular Secretion of Neisseria gonorroheae IgA Protease.

Press, A., & Sussman, L. (1991). y-Chymotrypsin Is a Complex of a-Chymotrypsin with Its Own Autolysis. *Biochemistry*, *30*, 5217–5225.

Qiu, J., Hendrixson, D. R., Baker, E. N., Murphy, T. F., St Geme, J. W., & Plaut, a G. (1998). Human milk lactoferrin inactivates two putative colonization factors expressed by Haemophilus influenzae. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(21), 12641–6.

Radisky, E. S., Lee, J. M., Lu, C.-J. K., & Koshland, D. E. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(18), 6835–40.

Rawlings, N. D., Barrett, A. J., & Bateman, A. (2000). MEROPS: the peptidase database. *Nucleic Acids Research*, *28*(1), 323–325.

Renn, J. P., & Clark, P. L. (2008). A conserved stable core structure in the passenger domain beta-helix of autotransporter virulence proteins. *Biopolymers*, *89*(5), 420–7.

Roos, a., Bouwman, L. H., van Gijlswijk-Janssen, D. J., Faber-Krol, M. C., Stahl, G. L., & Daha, M. R. (2001). Human IgA Activates the Complement System Via the Mannan-Binding Lectin Pathway. *The Journal of Immunology*, *167*(5), 2861–2868.

Rosa, R. De, & Labedan, B. (1998). The Evolutionary Relationships Between the Two Bacteria Escherichia coli and Haemophilus influenzae and their Putative Last Common Ancestor. *Molecular Biology and Evolution*, *17*(1), 17–27.

Rosano, G. L., & Ceccarelli, E. a. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*(April), 172.

Ruiz-Perez, F., & Nataro, J. P. (2014). Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence. *Cellular and Molecular Life Sciences : CMLS*, *71*(5), 745–70.

Schechter, I., & Berger, A. (1968). On the Active Site of Proteases. III. Mapping the Active Site of Papain; Specific Peptide Inhibitors of Papain. *Biochemical and Biophysical Research Communications*, *32*(5), 898–902.

Schmoeger, E., Wellhoefer, M., Dürauer, A., Jungbauer, A., & Hahn, R. (2010). Matrix-assisted refolding of autoprotease fusion proteins on an ion exchange column: a kinetic investigation. *Journal of Chromatography. A*, *1217*(38), 5950–6.

Senior, B. W., Dunlop, J. I., Batten, M. R., Kilian, M., & Woof, J. M. (2000). Cleavage of a Recombinant Human Antibody by Certain Bacterial IgA1 Proteases Cleavage of a Recombinant Human Immunoglobulin A2 (IgA2) -IgA1 Hybrid Antibody by Certain Bacterial IgA1 Proteases, *2*.

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *Journal of Bacteriology*, *189*(23), 8746–9.

Smith, D. B., Johnson, K. S. (1988). Single-step purification of peptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*, *67*, 31–40.

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, *115*(2), 113–28.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234.

Sukupolvi-Petty, S., Grass, S., & St Geme, J. W. (2006). The Haemophilus influenzae Type b hcsA and hcsB gene products facilitate transport of capsular polysaccharide across the outer membrane and are essential for virulence. *Journal of Bacteriology*, *188*(11), 3870–7.

Svehag, S. E., & Bloth, B. (1970). Ultrastructure of Secretory and High-Polymer Serum Immunoglobulin A of Human and Rabbit Origin. *Science*, *168*(3933), 847–849.

Tajima, N., Kawai, F., Park, S.-Y., & Tame, J. R. H. (2010). A novel intein-like autoproteolytic mechanism in autotransporter proteins. *Journal of Molecular Biology*, *402*(4), 645–56.

Tamm, L. K., Arora, a, & Kleinschmidt, J. H. (2001). Structure and assembly of beta-barrel membrane proteins. *The Journal of Biological Chemistry*, *276*(35), 32399–402.

Turk, D. C. (1984). Occasional review the pathogenicity o f, *18*, 1–16.

Ulanova, M., & Tsang, R. S. W. (2014). Haemophilus influenzae serotype a as a cause of serious invasive infections. *The Lancet Infectious Diseases*, *14*(1), 70–82.

Underdown, B. J., & Schiff, J. M. (1986). Immunoglobulin A: strategic defense initiative at the mucosal surface. *Annual Review of Immunology*, *4*, 389–417.

Valent, Q. a, Scotti, P. a, High, S., de Gier, J. W., von Heijne, G., Lentzen, G., … Luirink, J. (1998). The Escherichia coli SRP and SecB targeting pathways converge at the translocon. *The EMBO Journal*, *17*(9), 2504–12.

Van den Berg, B. (2010). Crystal structure of a full-length autotransporter. *Journal of Molecular Biology*, *396*(3), 627–33.

Veiga, E., de Lorenzo, V., & Fernández, L. a. (1999). Probing secretion and translocation of a beta-autotransporter using a reporter single-chain Fv as a cognate passenger domain. *Molecular Microbiology*, *33*(6), 1232–43.

Veiga, E., de Lorenzo, V., & Fernández, L. A. (2004). Structural tolerance of bacterial autotransporters for folded passenger protein domains. *Molecular Microbiology*, *52*(4), 1069–80.

Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M., & Tommassen, J. (2003). Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science (New York, N.Y.)*, *299*(5604), 262–5.

Watt, J. P., Wolfson, L. J., O'Brien, K. L., Henkle, E., Deloria-Knoll, M., McCall, N., … Cherian, T. (2009). Burden of disease caused by Haemophilus influenzae type b in children younger than 5 years: global estimates. *Lancet*, *374*(9693), 903–11.

Wong, S. M. S., & Akerley, B. J. (2012). Genome-scale approaches to identify genes essential for Haemophilus influenzae pathogenesis. *Frontiers in Cellular and Infection Microbiology*, *2*(March), 23.

Woof, J. M., & Kerr, M. a. (2006). The function of immunoglobulin A in immunity. *The Journal of Pathology*, *208*(2), 270–82.

Wyatt, R. J., & Julian, B. a. (2013). IgA nephropathy. *The New England Journal of Medicine*, *368*(25), 2402–14.

Zauner, G., Selman, M. H. J., Bondt, A., Rombouts, Y., Blank, D., Deelder, A. M., & Wuhrer, M. (2013). Glycoproteomic analysis of antibodies. *Molecular & Cellular Proteomics : MCP*, *12*(4), 856–65.

Zhai, Y., Zhang, K., Huo, Y., Zhu, Y., Zhou, Q., Lu, J., … Sun, F. (2011). Autotransporter passenger domain secretion requires a hydrophobic cavity at the extracellular entrance of the β-domain pore. *The Biochemical Journal*, *435*(3), 577–87.