Establishing Viable Systems for the Recombinant Production of *Haemophilus influenzae* Immunoglobulin A1 Protease

by

Norman Tran

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2020

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

Immunoglobulin A1 proteases (IgAPs) are a family of enzymatic virulence factors secreted from a wide variety of pathogenic bacteria known to infect human mucosal surfaces. Once secreted, IgAPs cleave and inactivate the predominant mucosal antibody, immunoglobulin A1 (IgA1), to hinder the host's adaptive immune response. As such, these enzymes are heavily implicated in these species' pathogenicities. Efforts are therefore underway to understand how IgAPs select for and cleave IgA1 to aid in the development of drugs to treat bacterial lower respiratory tract infections, among other clinical and biochemical applications.

The IgAP secreted from *Haemophilus influenzae* is of particular interest due to its high specificity for IgA1. Since its crystal structure was solved over a decade ago, there has been slow progress in uncovering its mechanism of substrate recognition and the structure-function relationships underlying these mechanisms. This has been due in part to problems that arise from transitioning the protein expression system from that in the native host, *H. influenzae*, to one more amenable to large-scale enzyme production. To this end, my thesis details the successful development and verification of two *H. influenzae* IgAP expression and purification systems, one in *Pichia pastoris* and the other in *Escherichia coli*. With these new enzyme sources, future studies can take a deeper dive into the biochemical intricacies of the *H. influenzae* IgAP, with the hope of making advances in clinical medicine.

Acknowledgements

I would like to thank Dr. Todd Holyoak for taking me on as a graduate student, providing plentiful amounts of research advice and opportunities, and helping me develop personally and academically. Thank you to Dr. David Rose and Dr. Brendan McConkey for being a part of my committee, asking insightful questions, and supporting my research endeavors.

I would also like to thank members of the Holyoak and Rose Lab for their support and friendship: Matthew McLeod, Sarah Barwell, Nicole Fraser, Nardo Nava Rodriguez, and Anna Jewczynko. Special shout-outs to Matthew McLeod and Nicole Fraser for being fantastic mentors open to emotional and scientific discussions.

Thank you to Jameson Ngo, Julia Solonenka, and Amy Xu for their endless support, keeping me sane, and making sure I had some semblance of a work-life balance during my studies.

On the experimental side, thank you to Aranksha Thakor (previously in Dr. Susan Lolle's and now in Dr. Trevor Charles' research group) for helping me with the anti-His tag dot blot.

This project would not have come to fruition without the opportunity to intern at the University of Colorado, Anschutz Medical Campus. Thank you to Dr. Elan Z. Eisenmesser for allowing me to integrate into his lab and giving me the chance to help with several of his other projects. Your scientific curiosity and work ethic continues to inspire me to this day. Finally, I could not have accomplished what I did in Denver without the gracious compassion and support of Mai Chi To and Khinh To.

Finally, thank you to Mitacs and NSERC for funding this project.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

LRTI	-	Lower respiratory tract infection
IgAP	-	Immunoglobulin A1 protease
IgA1	-	Immunoglobulin A1
IgAN	-	Immunoglobulin A nephropathy
HI-IgAP	-	*Haemophilus influenzae* immunoglobulin A1 protease
MAP	-	Methionine aminopeptidase
IPTG	-	Isopropyl β-D-1-thiogalactopyranoside
PMSF	-	Phenylmethanesulfonyl fluoride
α-MF	-	*Saccharomyces cerevisiae* alpha-mating factor
PAD	-	*Pichia pastoris* adenine-lacking media
rpm	-	Rotations per minute
BMGY	-	Buffered medium (containing) glycerol and yeast extract
BMMY	-	Buffered medium (containing) methanol and yeast extract
BMWY	-	Buffered medium (containing) water and yeast extract
TCA	-	Trichloroacetic acid
MeOH	-	Methanol
Ni-IMAC	-	Nickel-chelated immobilized metal-affinity chromatography
SDS-PAGE	-	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
NMR	-	Nuclear magnetic resonance
SUMO	-	Small ubiquitin-like modifier
ULP1	-	Ubiquitin-like-specific protease 1
SUMO-HI-IgAP	-	Small ubiquitin-like modifier-fused HI-IgAP
PCR	-	Polymerase chain reaction
MWCO	-	Molecular weight cutoff
SEC	-	Size-exclusion chromatography
FPLC	-	Fast protein liquid chromatography
GdnHCl	-	Guanidinium hydrochloride salt
HEPES	-	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
EDTA	-	Ethylenediaminetetraacetic acid
OD_x/A_x	-	Optical density/absorbance at x nm of light
CD	-	Circular dichroism
TCEP	-	Tris(2-carboxyethyl)phosphine

Section 1 – Immunoglobulin A1 Proteases

1.1 – Biomedical Significance

Lower respiratory tract infections (LRTIs) are one of the global leading causes of death for all ages,[1] with individuals lacking adequate health care access,[2] infants,[3] the elderly,[3] and the immunocompromised being most susceptible to the disease and its outcomes.[4,5] The vast majority of LRTIs are caused by the bacterial pathogens *Streptococcus pneumoniae* and *Haemophilus influenzae*.[1,2] Because LRTIs resulting from bacterial infection, particularly *S. pneumoniae*, are typically more serious and have a higher risk of mortality, antibiotics are often prescribed preemptively even though symptoms may not be related to LRTIs or may be due to viruses.[6] This consequently escalates the global issue of antibiotic resistance.[7] With three LRTI-associated bacterial pathogens, *S. pneumoniae*, *H. influenzae*, and *Neisseria gonorrhoeae*, already identified by the World Health Organization as top-priority concerns for antibiotic resistance,[8] the development of alternative treatments for LRTIs is of great importance.

1.2 – Immunoglobulin A1 Proteases as Bacterial Virulence Factors

Immunoglobulin A1 proteases (IgAPs) are a family of proteases which cleave immunoglobulin A1 (IgA1), the main antibody protecting mucosal surfaces.[9,10] Because of IgA1's role in defending lower respiratory tract surfaces, LRTI-associated bacteria often use IgAPs as a virulence factor to aid in immunosuppression and tissue colonization.[9]

IgA1 is cleaved at its hinge region (*Fig. 1.2.1a*, red arrow), the linker between its antigen-binding and immune-system-inducing domains.[9,10] By decoupling the processes of pathogen detection and immune reaction, IgAPs are able to effectively subdue and subvert the adaptive immune response.[9,10] All IgAPs cleave IgA1 C-terminal to a hinge-region proline, although IgAPs produced from different species or isozymes found within a single species may cleave at separate hinge sites (*Fig. 1.2.1b*).[9] Once IgA1 has been cleaved, free antigen-binding domains coat the bacterium surface, protecting the pathogen from opsonization by competing with intact IgA1 for epitope binding.[11] A more detailed description of the general features of monomeric and oligomeric IgA1 and their relation to IgAP function can be found in Section 1.6.

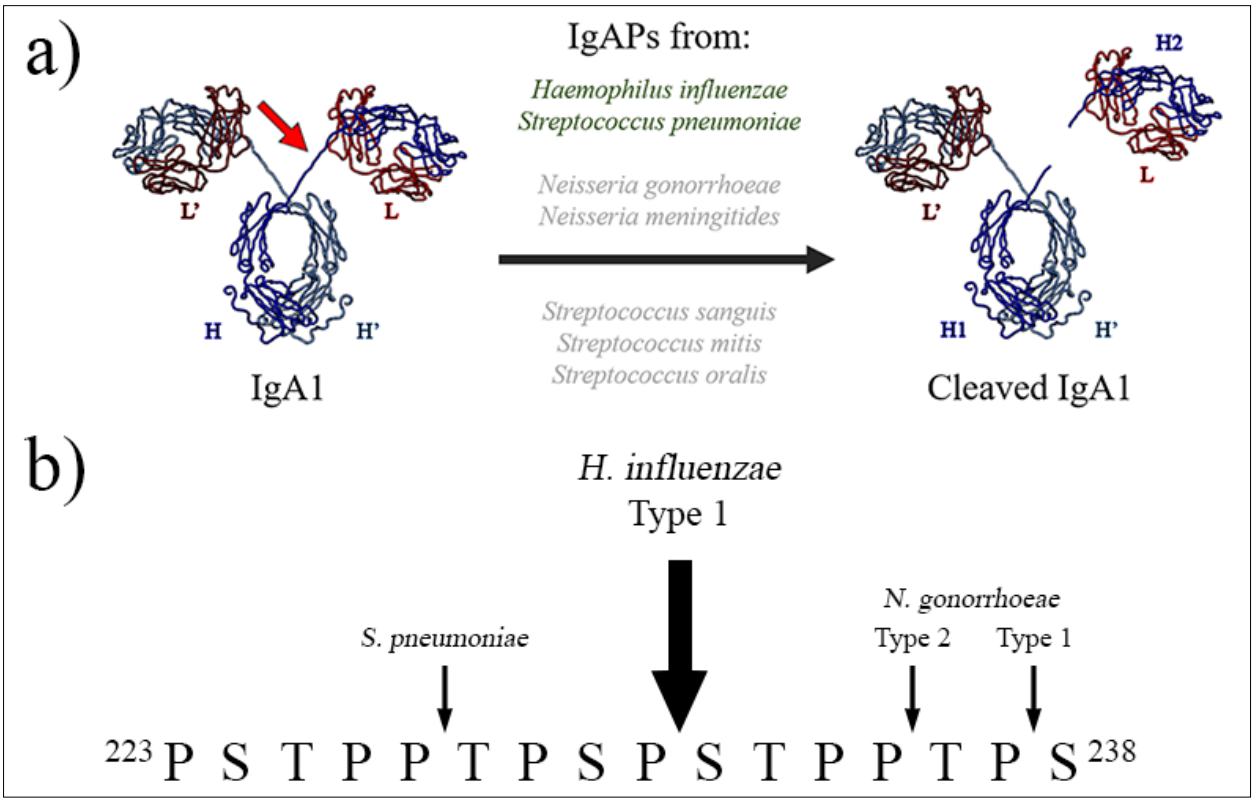

Figure 1.2.1 – General reaction catalyzed by IgAP. a) Reaction overview: IgAPs cleave IgA1 (PDB 3CHN) at the heavy-chain hinge region (red arrow); Although the reaction is depicted with monomeric IgA1 for simplicity, the same site is cleaved in the secreted dimeric IgA1 (Section 1.6); L/L' and H/H' represent the light and heavy chains, respectively; H1/H2 are the two heavy-chain fragments after cleavage; IgAP-producing species of highest clinical interest are in green. b) Cleavage sites of several IgAPs: Primary sequence of the IgA1 hinge is shown with residues numbered according to Toraño and Putnam;[12] the enzyme of interest in this thesis is the Type 1 *H. influenzae* IgAP and its cleavage site is indicated by the large arrow.

These two means of protection offered by IgAPs highlight the enzymes' importance in the infection pathway of LRTI-associated pathogens. The fact that pathogenicities of *H. influenzae* and *N. gonorrhoeae* clinical isolates were determined solely by the production of functional enzyme strongly suggests that IgAPs are critical to establishing successful infections.[13–15] This likewise supports the use of IgAPs as the next set of antibiotic drug targets to treat bacterial LRTIs.[9,14] As such, a deeper understanding of how IgAPs function at a molecular and atomic level is essential to develop IgAP-inhibiting molecules as well as vaccines for the long-term prevention of LRTIs.

1.3 – The Immunoglobulin A1 Protease Family

Although all IgAPs have the same general function, there are stark differences in enzyme structure and mechanism between some members of the IgAP family. Each family member can be classified under one of the following three categories based on their catalytic mechanism: those involving activated serine, cysteine, or metal-activated nucleophiles.[9] Evidence for the existence of each class was mainly based on the effects of general protease inhibitors on these enzymes, with additional support from structural homology to other protease types and primary-sequence functional motifs.[9] In particular, serine IgAPs are inactivated by serine-specific peptides like prolyl boronic acid motifs,[16] cysteine IgAPs with cysteine-specific inhibitors like iodoacetamide,[17] and metal-dependent IgAPs with metal chelators like ethylenediaminetetraacetic acid (EDTA).[18] In addition, cysteine IgAPs require reducing agent to retain activity.[17]

Given the lack of primary sequence homology between each enzyme class, IgAPs are likely products of convergent evolution.[9] This is not surprising as there is a great evolutionary advantage in cleaving IgA1 to facilitate infection.[13–15] Structural divergence within the IgAP family is particularly evident when comparing the globular-domain architecture of the *S. pneumoniae* metal-dependent IgAP[19] to the elongated rod-like *H. influenzae* serine IgAP.[20] Unfortunately, this means that structures of at least three representative IgAPs, one from each mechanism class, will be needed to gain a general understanding of this enzyme family. Furthermore, we can expect inhibitors specific to one enzyme to have limited applicability to members of other mechanism classes. With this in mind, let's take a brief dive into the current progress of IgAP therapeutics and what needs to be done to continue their development.

1.4 – Preliminary Immunoglobulin A1 Protease Inhibitors

Ever since the discovery of IgAPs, and their role in bacterial pathogenesis, there has been rising interest in developing IgAP inhibitors to treat diseases caused by IgAP-secreting pathogens. The collaborative efforts of synthetic organic chemists alongside cell biologists have led to the discovery of many novel small-molecule inhibitors, which act on a variety of IgAPs.

The first set of inhibitors were designed to inhibit the *N. gonorrhoeae* serine IgAP by mimicking the IgA1 hinge-region primary sequence.[21] Despite being a competitive active-site inhibitor, it did not inhibit the *H. influenzae* serine protease,[21] which likely shares a conserved

active site to the *N. gonorrhoeae* IgAP based on its homologous primary sequence and identical structural elements.[9] This highlights the need for high-resolution structural information, even within an IgAP class, as small structural differences clearly lead to different biochemical properties. These hinge-region peptides were further improved by brute-force optimization, leading to the development of inhibitors with low-micromolar half-maximal inhibitory concentrations (IC_{50}).[22] A similar methodology used prolyl boronic acids to develop hinge-peptide mimics for *N. gonorrhoeae* and *H. influenzae* IgAPs.[16] As these molecules closely resembled one of the reaction's transition states, they showed much stronger inhibition compared to the previous molecules, with inhibition constants (K_d) in the low-nanomolar range.[16]

Non-active-site IgAP inhibitors were only discovered fairly recently, a few decades after the initial set of inhibitors were established, when a semi-high-throughput inhibitor screening method using gold nanoparticles was developed.[23] The first inhibitors for the *S. pneumoniae* metal-dependent IgAP were discovered using this method.[23] A second high-throughput inhibitor screening method was developed using fluorescent probes, leading to the discovery of several allosteric inhibitors of the *H. influenzae* IgAP.[24,25]

Unfortunately, none of these competitive or allosteric inhibitors have been further enhanced or characterized since their initial discovery. This has been largely due to the lack of high-resolution structural information for many of the enzymes targeted, which prevented the use of structure-based drug design strategies for drug optimization.

1.5 – Immunoglobulin A Nephropathy

The clinical relevance of IgAPs is not only limited to LRTIs and inhibitor design. Instead of being thought of as a drug target, IgAPs can be viewed as a protein-based drug for the treatment of immunoglobulin A nephropathy (IgAN). IgAN is a chronic renal disease characterized by the accumulation of immunoglobulin aggregates, primarily composed of IgA1, in kidney glomeruli.[26,27] As these aggregates impede glomerular filtration, IgAN is often comorbid with other systemic diseases and complications, with the disease eventually leading to kidney failure.[28] Due to the unknown and multi-faceted factors underlying IgAN, current treatment options have been unable to treat the underlying causes of the disease, only providing temporary relief by slowing down IgA1 buildup rather than removing existing immunoglobulin aggregates.[28]

Owing to their IgA1-degrading capabilities, IgAPs have promise as an IgAN therapeutic as they may be able to reverse symptoms by directly targeting the symptoms' source.[29] This potential was supported in a proof-of-concept experimental mouse model. To simulate the kidneys of chronic IgAN patients, human IgA1 aggregates were deposited into mouse glomeruli via the bloodstream.[30] These aggregates were then effectively cleared following injections of *H. influenzae* IgAP.[30] Similar kidney mouse models were used to show the viability of *N. gonorrhoeae* and *Neisseria meningitidis* IgAPs for IgAN treatment, recapitulating the same promising results.[31]

The use of IgAPs as an IgAN therapeutic provides two benefits in addition to the one previously stated. Some IgAPs have high substrate specificity, evident by their inability to cleave the IgA1 hinge peptide in isolation.[32–34] This will reduce the likelihood of off-target effects, a frequent concern for small-molecule therapeutics.[35] A second layer of substrate specificity may also be obtained by engineering IgAPs to recognize aggregated IgA1 over circulating or mucosal IgA1, as these molecules typically have under-glycosylated O-linked glycans, a structural variation that is thought to be the molecular cause of IgAN.[27]

As promising as this appears, these proof-of-concept mouse models failed to consider side-effects associated with the enzyme's high immunogenicity, as the enzyme has bacterial origin.[29] Engineering IgAPs, or any enzyme for that matter, to recognize alternate forms of its substrate is also not an easy feat. An in-depth understanding of IgAP structure, regions of immunogenicity, as well as properties of the IgAP-IgA1 interaction are all needed to redesign IgAPs into a clinically viable protein-based therapeutic.

1.6 – Secretory and Serum Immunoglobulin A1

To fully appreciate the nuances and potential problems involved in moving IgAP research towards LRTI treatment, where the aim is to weaken or prevent IgAP-IgA1 interactions, or IgAN treatment, where the aim is to create more specific and stronger interactions, we must first take into account the structures of the physiologically relevant forms of IgA1.

Although IgA1 is found in serum and on mucosal surfaces, its major quaternary state differs between these environments.[36] Serum IgA1 is mainly monomeric (*Fig. 1.6.1a*), while mucosal IgA1 primarily takes on a dimeric form (*Fig. 1.6.1c*).[36] As IgA1 has the ability to form oligomeric

states as a part of its secretion pathway, through its association with other polypeptide chains, mucosal IgA1 is secreted as a mixture of oligomeric states, primarily of the dimeric form, but with small populations of larger oligomers up to and including a pentameric state.[37]

A brief summary of the secretion pathway for IgA1 is as follows: mucosal B cells produce and export the immature dimeric IgA1 precursor (*Fig. 1.6.1b*) out into the extracellular space basolateral to the mucosal epithelial lining.[38] This precursor consists of two covalently linked monomeric IgA1 molecules attached by their constant region to a separate joining chain (J-chain).[38] Receptors present on the basolateral epithelial surface recognize and transcytose this precursor to the apical surface.[38] Through this process, this dimeric precursor reaches maturation by combining with the transcytosis receptor's ectodomain, which forms the secretory component (SC) of the mature dimer.[37] Larger secreted oligomers can also form during transcytosis, whereby additional IgA1 monomers add to the existing dimeric scaffold.[37]

When we eventually wish to apply IgAP research towards the two previously stated clinical end goals, it will be important to identify the most physiologically relevant form of IgA1 under each situation. As IgAP inhibitors should be designed to function on mucosal surfaces, the structure-function relationship of the dimeric IgA1-IgAP interaction should be the focus of that research. Under a similar logic, IgAPs designed to treat IgAN should be developed with a focus on the monomeric IgA1-IgAP interaction. That being said, IgAPs are able to recognize and cleave both monomeric and dimeric IgA1.[31] Although a comparison between the kinetics and binding interactions of IgAPs with each form of IgA1, including oligomers larger than a dimer, have yet to be done, monomeric IgA1 still remains the easiest substrate to characterize for initial studies from both an economic and structural standpoint. As such, my thesis will use myeloma-produced monomeric IgA1 (Athens Research & Technology) for all experiments involving IgA1.

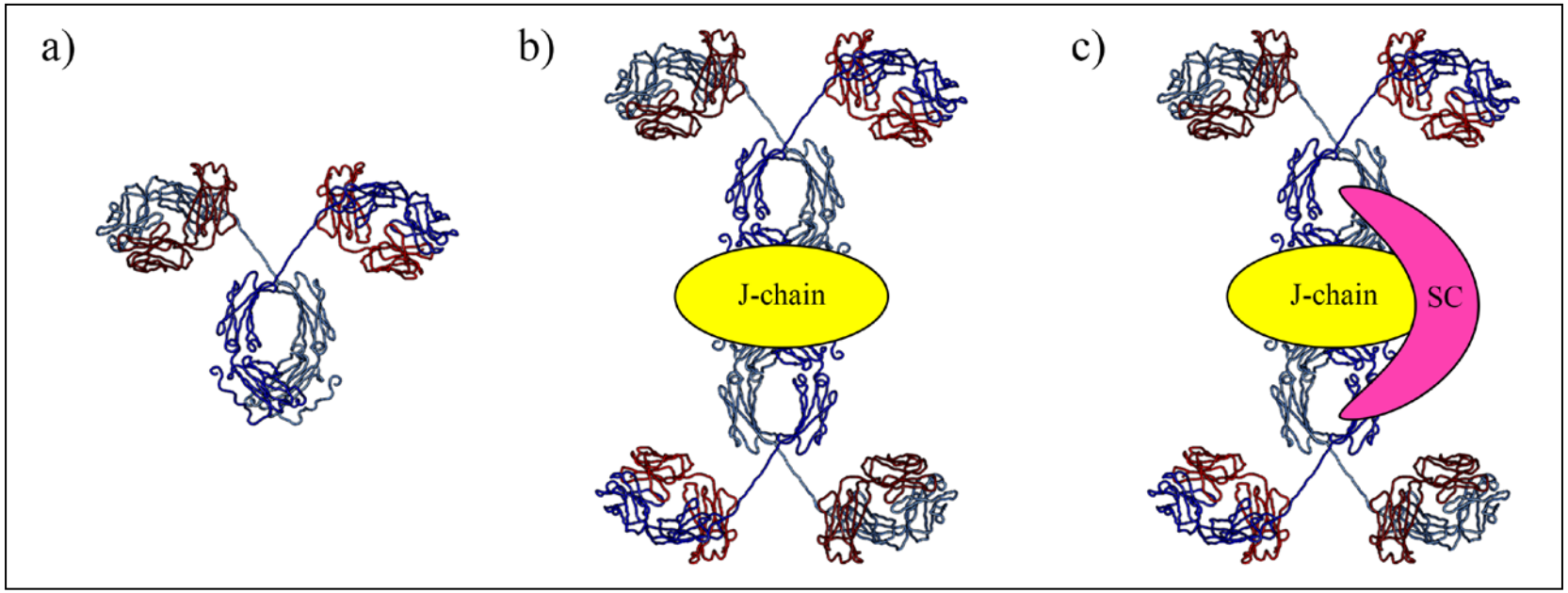

Figure 1.6.1 – Oligomeric forms of IgA1. a) Monomeric IgA1 (PDB 3CHN). The heavy and light chains are colored according to *Fig. 1.2.1a*. b) Secreted by basolaterally located B cells, this immature dimeric IgA1 consists of two IgA1 monomers covalent linked via their constant region to the joining chain (J-chain). c) The mature IgA1 dimer found in mucosal secretions. The secretory component (SC) is added during the maturating transcytosis process.

1.7 – The Haemophilus influenzae Immunoglobulin A1 Protease

Of all of the IgAPs, the *H. influenzae* IgAP (HI-IgAP) is currently the most understood as it is the only IgAP with a published high-resolution structure (*Fig. 1.7.1*).[20] This makes it a good starting point for investigating how IgAPs function. Unfortunately since the discovery of this crystal structure over a decade ago, there has been little to no progress in uncovering the deeper workings of the enzyme. This has been largely due to barriers in transitioning the protein expression system from the *H. influenzae* native system, previously used as the enzyme source to solve the crystal structure,[20] to one more suitable for large-scale enzyme production. Fortunately, the results of my research have overcome this barrier, providing a foundation for future research into the biochemical intricacies of the HI-IgAP.

This thesis will first outline the aforementioned barrier and previous unsuccessful attempts at overcoming it (Section 2). The development and verification of two new protein-production systems, one in *Pichia pastoris* (Section 3) and the other in *Escherichia coli* (Section 4), will be described, followed by a brief recap and discussion of future research aims (Section 5).

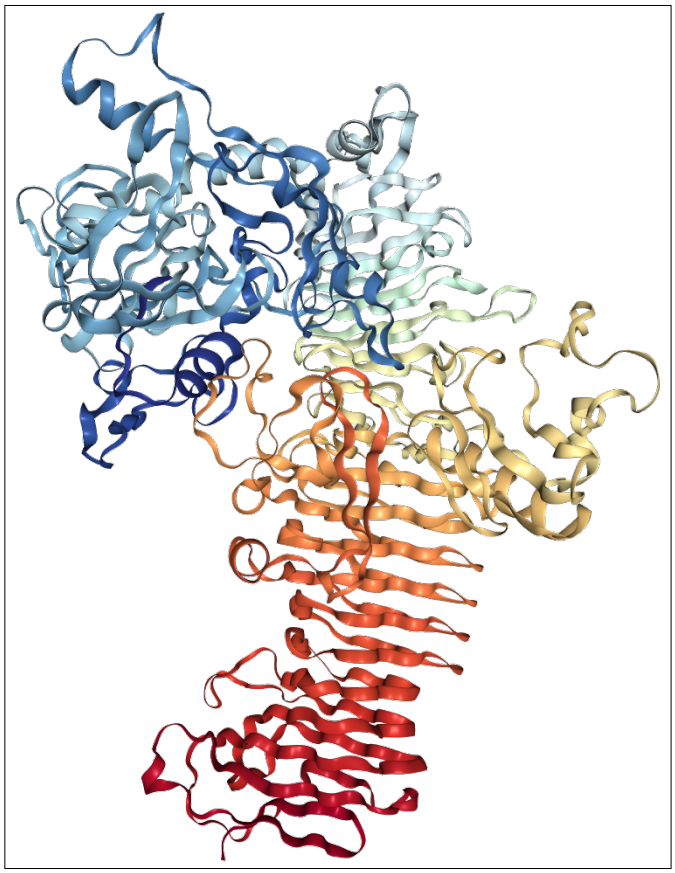

Figure 1.7.1 – HI-IgAP crystal structure (PDB 3H09).[25]

Section 2 – Previous *H. influenzae* Immunoglobulin A1 Protease Expression Systems

To understand the inherent problems in moving the protein production system from *H. influenzae* to one more optimized for protein production, we must first appreciate some nuances of the *H. influenzae* IgAP (HI-IgAP) secretion pathway and crystal structure.

2.1 – Secretion Pathway of the Haemophilus influenzae Immunoglobulin A1 Protease

As *H. influenzae* is Gram-negative, secreted proteins must cross both inner and outer cell membranes before they exit into the extracellular space.[39] Although there are many secretion pathways available for Gram-negative bacteria, the HI-IgAP is secreted using the Type Va autotransporter system (*Fig. 2.1.1*):[39]

The newly translated pre-pro-peptide contains an N-terminal signal peptide that is recognized by a protein complex associated with the inner membrane, the Sec complex, which translocates the protein into the periplasm.[39] During this process, the signal peptide is cleaved to form the pro-peptide.[39] Instead of using existing cellular machinery to cross the outer membrane, the C-terminal region of the pro-peptide freely folds into a beta barrel, which embeds itself into the outer membrane.[39] The N-terminal region of the pro-peptide is then fed through the outer membrane with the help of one or more of these C-terminal β-domains, the specifics of which remain unclear, and the fully folded N-terminal protease domain is then released following membrane-bound folding and autoproteolysis.[39] It is this N-terminal domain which constitutes the active protease, whose structure was determined by crystallography (*Fig. 1.7.1*).[20] Part of the impetus to develop a non-native protein production system was the need to investigate inactive variants, which would have been impossible to produce in *H. influenzae* given the need for proteolytic activity during protein secretion.

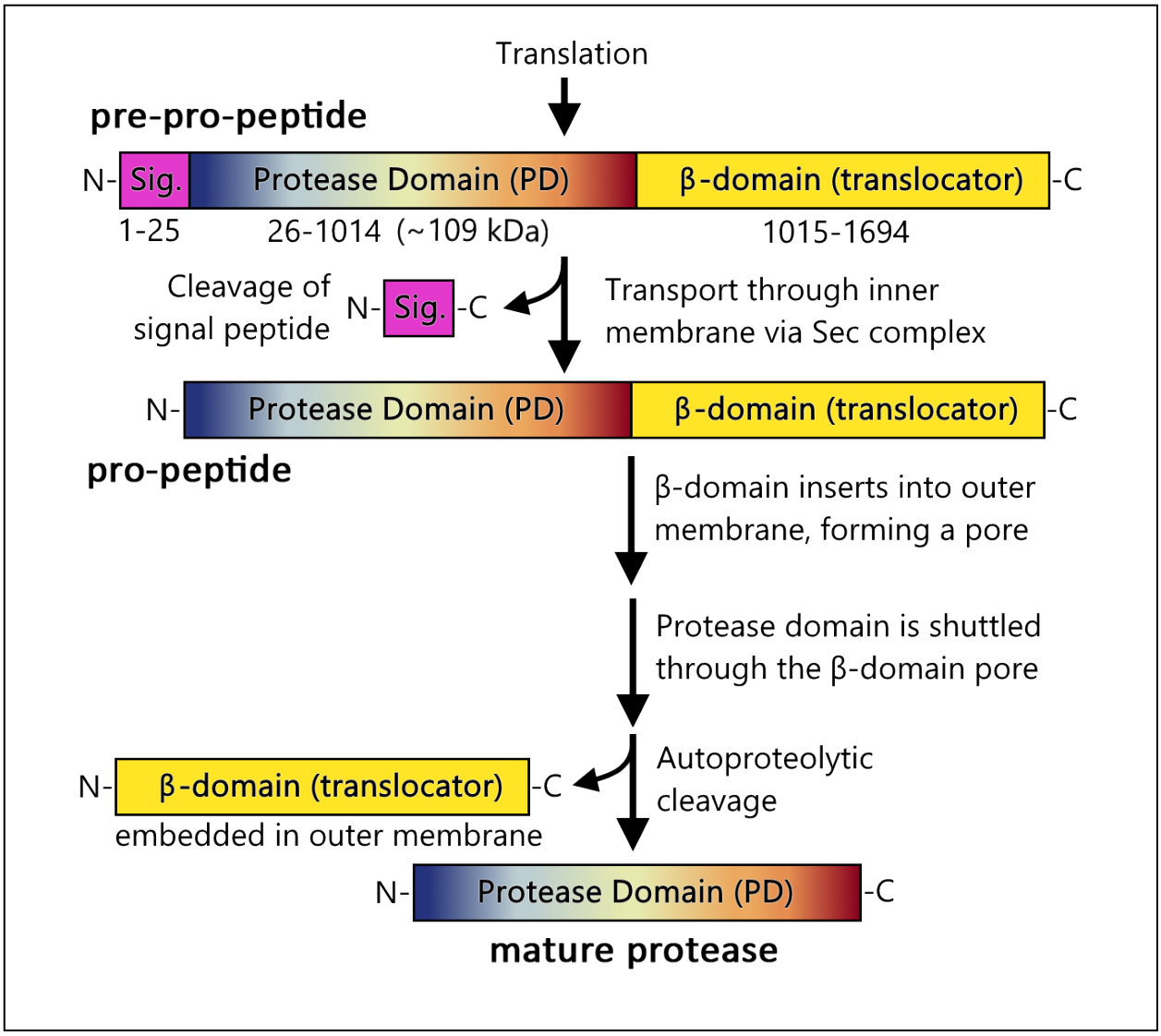

Figure 2.1.1 – Schematic of the HI-IgAP Type Va secretion pathway. This figure traces the autotransporting path of HI-IgAP from the cytosolic pre-pro-peptide, through to the periplasmic pro-peptide, and finally to the fully active protease (*Fig. 1.7.1*). The signal peptide and beta barrel are depicted in purple and yellow, respectively. Residues are numbered based on UniProtKB entry #P45384.

2.2 – The Protein-Production Quandary

The crystal structure of the secreted protease (*Fig. 1.7.1*) follows a typical fold for Type Va autotransporters,[20,39] with an N-terminal functional domain linked to a characteristically elongated C-terminal beta helix.[39] Despite having high structural homology to HI-IgAP, other proteins under the Type Va autotransporter family have vastly different functions, ranging from being a cell-cell adhesion molecule for biofilm formation[40] to proteases that, like HI-IgAP, aid in bacterial pathogenicity, albeit using differing mechanisms.[41,42] Again, this highlights the need for high-resolution structural information to understand how IgAPs work and what elements make them function differently to these structurally homologous proteins.

The HI-IgAP N-terminal domain has a chymotrypsin-like fold and is where the active-site serine is located.[20] The protease's C-terminal fold seems to confer specificity towards IgA1 by using several loops that protrude from the beta helix to selectively interact with the IgA1 constant fragment, as supported by crude molecular-docking experiments.[20] Beta helices are very common among other proteins in the Type Va autotransporter family, likely functioning as a folding nucleus to allow for protein folding following autoproteolysis starting from the C-terminus towards the N-terminal protease domain.[39]

Problems with the recombinant production of HI-IgAP arise from structural features of the N-terminal chymotrypsin-like domain. Unlike most other proteins which have unstructured N-termini, the HI-IgAP N-terminus is hydrophobic and buries itself into its chymotrypsin-like domain core, forming an essential part of the N-terminal fold (*Fig. 2.2.1b*).[20] This hydrophobicity is not an issue during native expression as the proper N-terminus is revealed after signal peptide cleavage during secretion (Section 2.1). However when the enzyme is recombinantly expressed in the cytosol, the initiator methionine cannot be effectively removed, resulting in aggregation and the formation of insoluble inclusion bodies. Despite this, a miniscule amount of this recombinant HI-IgAP remains active and soluble due to basal activity of methionine aminopeptidase (MAP), an enzyme that cleaves N-terminal methionines.[43] Armed with this knowledge, we developed our first-generation *E. coli* cytosolic expression system in hopes of getting around this N-terminal problem without relying on more complicated expression systems.

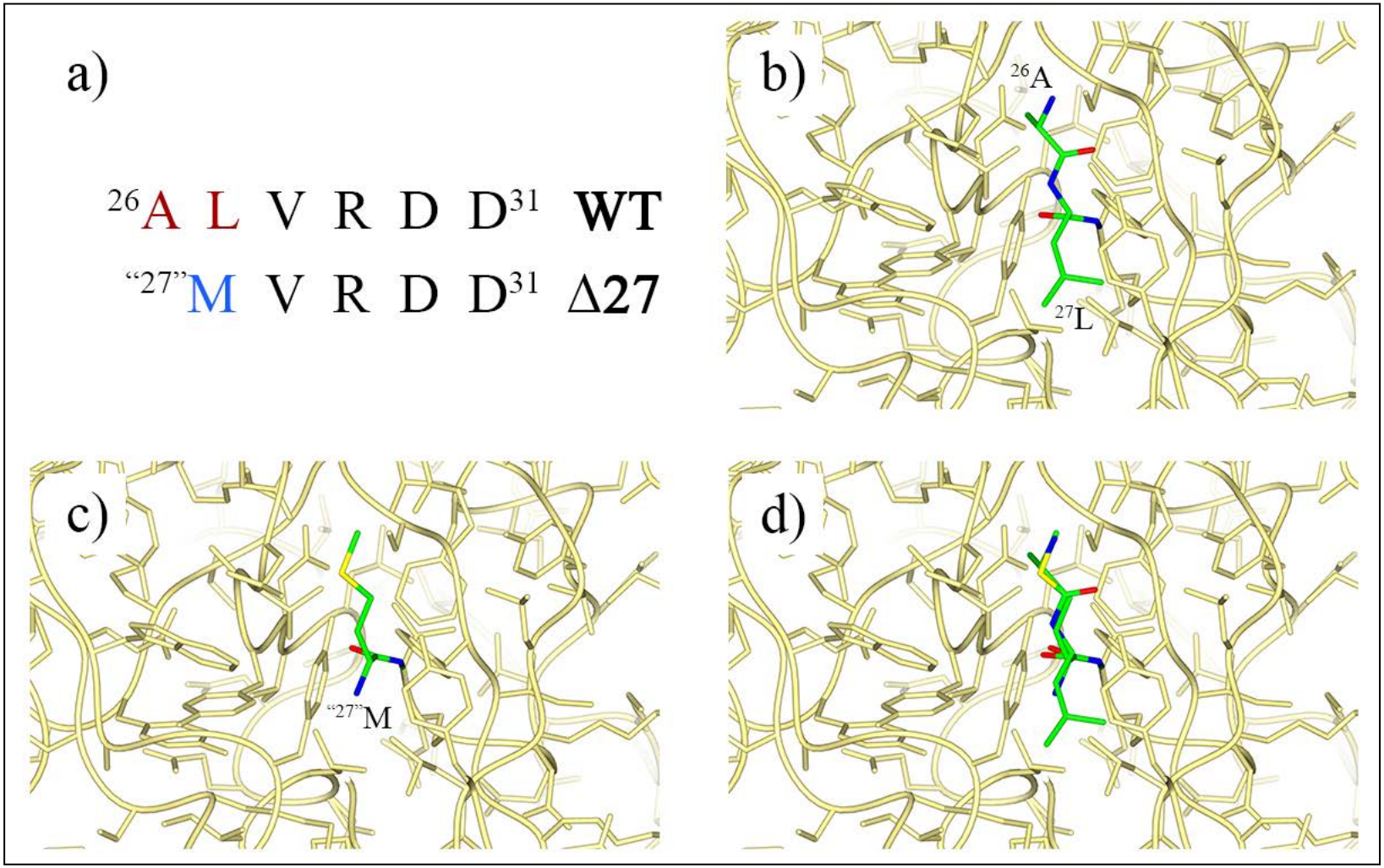

Figure 2.2.1 – The N-terminal protein-production problem. a) Comparing primary sequences of the mature wild-type (WT) and Δ27 HI-IgAP constructs: two N-terminal residues (red) were truncated to accommodate for the N-terminal formyl-methionine (blue) added during bacterial cytosolic expression; sequence was numbered based on UniProtKB entry #P45384. b) N-terminal fold of the WT HI-IgAP: the tightly packed N-terminus (A26 and L27) is highlighted. c) Hypothetical N-terminal fold of the mature Δ27 HI-IgAP: A26 and L27 were replaced by the N-terminal methionine; this methionine is theoretically well accommodated within the pocket created by removing these residues when solely considering the steric bulk of the surrounding crystal-structure residues (d); the effect of this mutation on the protein's other biophysical properties was not taken into consideration for the theoretical structure presented here.

2.3 – The First-Generation Cytosolic Expression System in Escherichia coli

The first-generation *E. coli* cytosolic system for producing recombinant HI-IgAP expressed the native non-codon-optimized HI-IgAP protease domain (ENA #AAA24966.1; residues 26-1006 based on UniProtKB entry #P45384) in BL21(DE3) cells. The gene was cloned into pET24b using the NdeI/XhoI sites. Various strategies to remove the initiator methionine during expression or purification were attempted by a previous Masters' student in the lab: dual overexpression of MAP and HI-IgAP, post-expression treatment of MAP under denaturing conditions, cyanobromide to chemically remove the N-terminal methionine, and inclusion body refolding.[44] Ultimately, no soluble or active protease was produced.[44]

I began considering other approaches to this problem. Instead of attempting to remove the initiator methionine, we designed a construct that would accommodate the methionine as part of its N-terminal fold. This new construct started two amino acids C-terminal to the mature N-terminus (*Fig. 2.2.1a*), theoretically allowing the initiator methionine to take up the space freed up by removing these residues (*Fig. 2.2.1c* and *2.2.1d*). Termed the Δ27 HI-IgAP, this construct was able to produce soluble (*Fig. 2.3.1a*) and active protease (*Fig. 2.3.1b*) with a two-day autoinduction expression at 20°C (Section 2.4).[45] Even though this was an improvement over the previous system, only a small fraction of the total enzyme was soluble (*Fig. 2.3.1a*, red arrow).

Further experimentation also suggested that the Δ27 HI-IgAP was incorrectly folded as attempts to optimize its purification failed due to reoccurring and unavoidable proteolytic degradation (*Fig. 2.3.2a*) and co-purifying contaminants (*Fig. 2.3.2b*), both of which could not be separated using ammonium sulfate precipitation, hydrophobic interaction, anion exchange, or size exclusion chromatography (data not shown).

As this construct lacked the ability to produce enzyme with sufficient purity for biochemical analyses, I began investigating recombinant secretory expression systems as an alternative to the *E. coli* cytosolic system, which had proven to be troublesome.

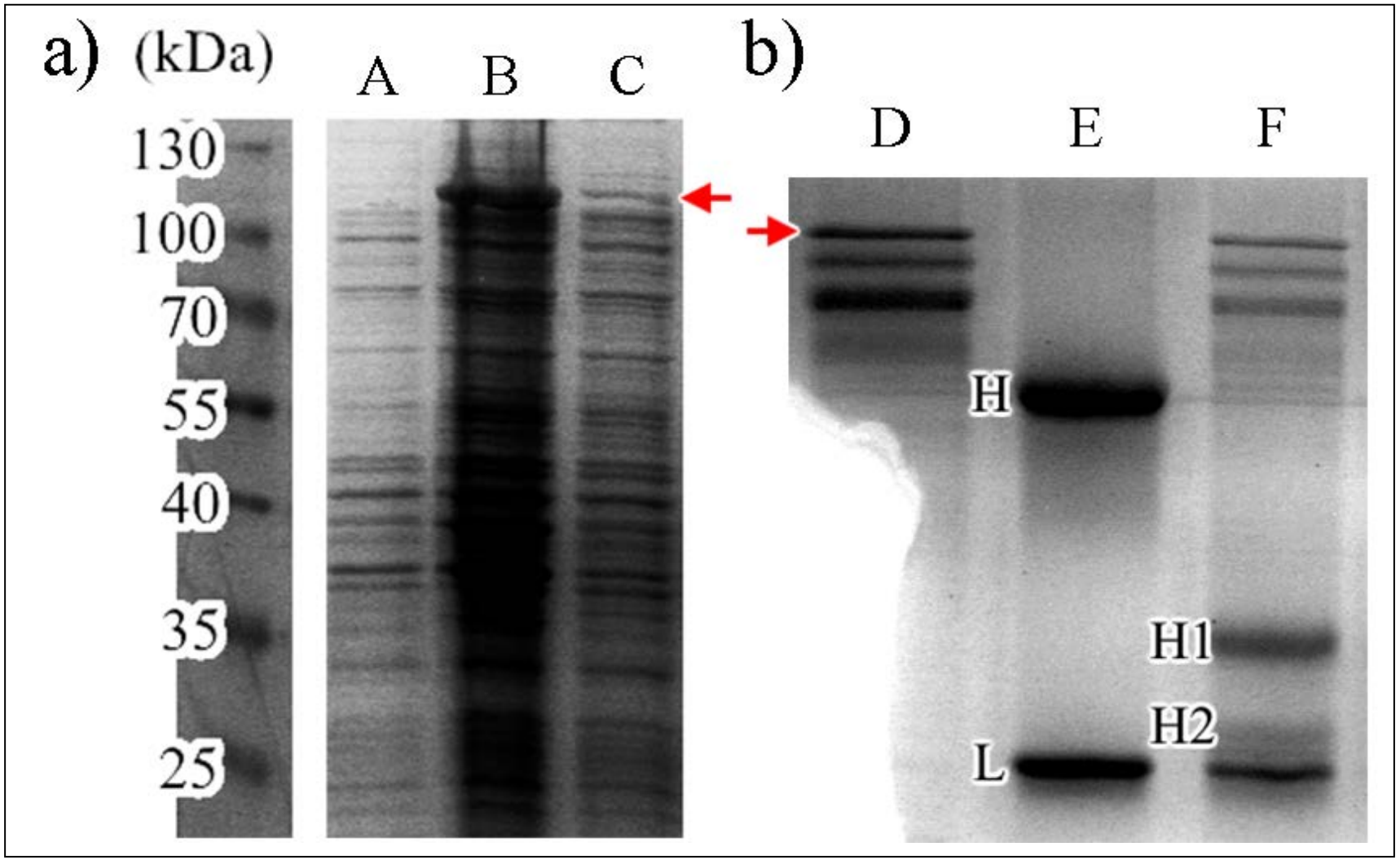

Figure 2.3.1 – Solubility and activity of the Δ27 HI-IgAP construct. a) Solubility of the Δ27 HI-IgAP: the red arrow highlights the Δ27 HI-IgAP in the soluble fraction (C); only a tiny fraction of the total protein (B) is soluble; A is the uninduced lane, which is essentially identical to the soluble fraction of the wild type HI-IgAP. b) IgA1 cleavage assay: crudely purified and mostly degraded Δ27 HI-IgAP (lane D; full-length protein is the top band indicated by a red arrow) was incubated with IgA1 (E), showing activity by cleaving the heavy chain into two fragments, H1 and H2 (F); fragment and chain labels correspond to those described in *Fig. 1.2.1*.

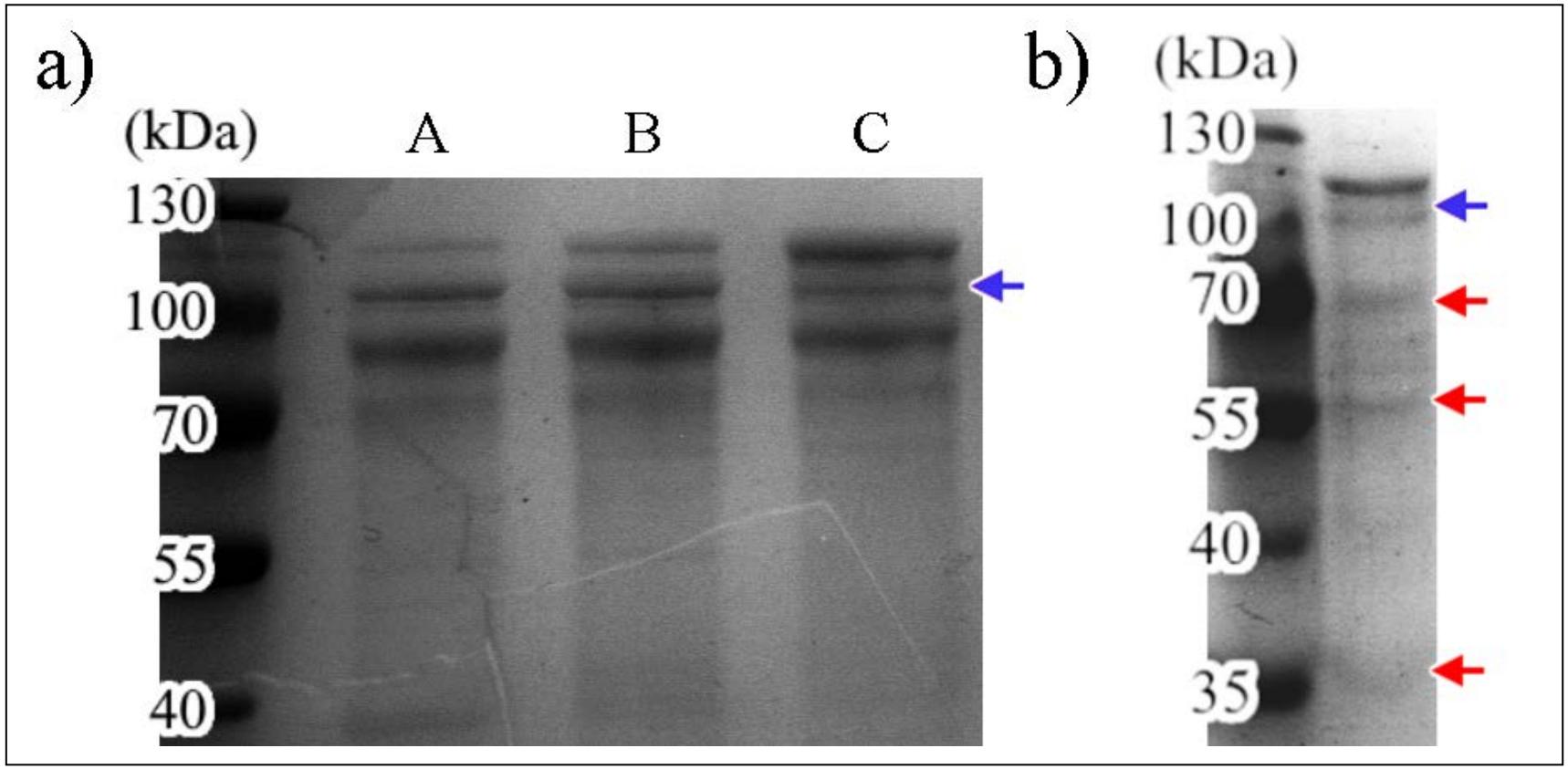

Figure 2.3.2 – Δ27 HI-IgAP degradation products. a) Partially inhibiting degradation during purification: the major degradation product (blue arrow) was limited by adding 50 mM EDTA (C) to the protein sample (A); PMSF, a serine-protease inhibitor, did not prevent degradation (B). b) Final enzyme purity: even with the addition of EDTA throughout the entire purification, metalloprotease-associated degradation was still seen (blue arrow), albeit to a lesser degree than in panel a); other minor, but still significant, degradation products (red arrows) were also seen in the final product and could not be separated by simple chromatography.

2.4 – Expression and Purification Protocol for the Δ27 HI-IgAP

A general overview for the production of Δ27 HI-IgAP is described below. The components of each buffer are listed in *Table 2.4*. All steps were done at 4°C unless otherwise specified.

1. Grow an overnight BL21(DE3) *E. coli* starter containing the Δ27 expression construct at 37°C from a glycerol stock. All media were supplemented with 50 μg/mL kanamycin.
2. Add 50 mL starter into each liter of ZY-5052 media.[45] Cells were incubated at 20°C for 40-48 hours and were induced via autoinduction.[45]
3. Cells were harvested using centrifugation for 10 minutes at 6,000xg.
4. Cells were resolubilized in Resuspension Buffer and lysed via two passes through a FRENCH pressure cell (Thermo) at 1100 psi.
5. The lysate was centrifuged for one hour at 12,000xg and the supernatant was incubated with EDTA-resistant Ni-Penta resin (Marvelgent) pre-equilibrated in Resuspension Buffer for 1 hour at 4°C.
6. Ni-Penta resin was washed with 1 L Ni Wash Buffer followed by 1 L Resuspension Buffer.
7. Ni-Penta resin was eluted with Ni Elution Buffer until A_{280} readings were below 0.05.
8. Eluent was concentrated to less than 5 mL, using a 30 kDa MWCO centrifugal concentrator (Amicon), and loaded onto a manually packed Bio-Gel P-6DG desalting column (Bio-Rad) pre-equilibrated in Storage Buffer.
9. The buffer-exchanged fractions were concentrated to less than 1 mL using a 30 kDa MWCO centrifugal concentrator and flash frozen in aliquots with liquid nitrogen.

Table 2.4 – Purification buffers for the Δ27 HI-IgAP. HEPES, NaCl, EDTA, and imidazole were purchased from BioBasic and IgePal CA-630 from Sigma. All buffers have a pH of 7.5 at room temperature. All buffers were equilibrated to 4°C before use.

Buffer Name	Components
Resuspension	25 mM HEPES + 0.5 M NaCl + 5 mM imidazole + 50 mM EDTA
Wash	25 mM HEPES + 0.1% (v/v) IgePal CA-630 + 5 mM imidazole + 50 mM EDTA
Elution	25 mM HEPES + 0.5 M NaCl + 300 mM imidazole + 50 mM EDTA
Storage	25 mM HEPES + 50 mM EDTA

Section 3 – Developing the *Pichia pastoris* Secretory Expression System

3.1 – Pichia pastoris as a Secretory Overexpression System

Pichia pastoris is a methylotrophic single-cellular yeast that has been optimized as a recombinant protein secretory system.[46] Its relative ease of use, ability to form high-density cell cultures, and proven success in producing difficult proteins made *P. pastoris* a logical next step in the search for a viable method of HI-IgAP production.[46,47]

A typical *P. pastoris* expression vector is shown in *Fig. 3.1.1*. Three main elements are of interest:

1. *Selectable marker – ADE2 gene*

   P. pastoris strains used for protein expression are typically auxotrophic, meaning they have been genetically engineered to lack a particular or set of genes essential to *de novo* synthesis of a certain macromolecular nutrient.[47] These auxotrophic strains are then transformed with expression vectors that complement the missing gene and grown on media lacking the auxotrophic nutrient to select for transformants.[47]

   Some auxotrophic *P. pastoris* strains lack *ADE2*, a gene which codes for an enzyme in the adenine biosynthetic pathway.[48] Because of this knockout, untransformed *ADE2*-lacking cells accumulate a red metabolic adenine precursor when grown on adenine-lacking media.[48] This simplifies the process of identifying successful transformants as colonies containing the expression vector are theoretically colorless after transformation.[48]

2. *Induction method – AOX1 promoter*

   In addition to preferentially using high-energy carbon sources such as glucose and glycerol, *P. pastoris* can use methanol, not only as a nutrient source but also to induce gene expression via the AOX1 promoter.[46,47] The use of this promoter also enables glycerol-mediated transcription repression, allowing for the growth of high-density cell cultures prior to induction using glycerol-amended media.[47]

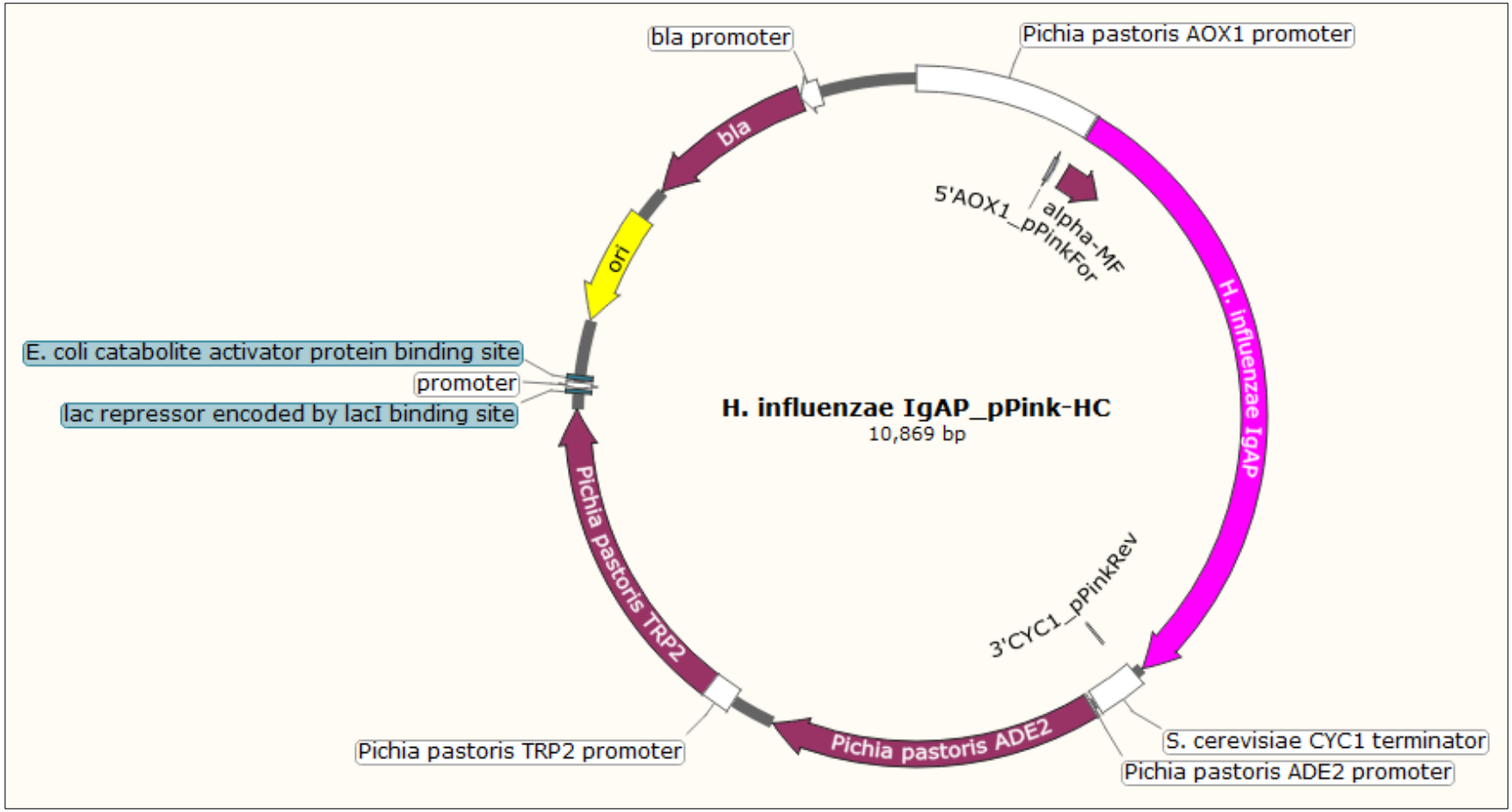

Figure 3.1.1 – Schematic of a typical P. pastoris expression vector.

3. *Method of secretion – α-mating factor*

The *Saccharomyces cerevisciae* α-mating factor (α-MF) is the most common signal peptide used for recombinant protein secretion in *P. pastoris*.[46,47] Expressing a protein of interest as a C-terminal fusion with an N-terminal α-MF allows the protein to shuttle into the endomembrane system for secretion.[49] The α-MF is then cleaved in a two-step process, leaving behind no residual α-MF amino acids on the protein's N-terminus.[50] This property makes this particular signal peptide a great match for the recombinant expression of HI-IgAP (Section 2.2).

3.2 – Vector Design and Transformation

The pPINK-HC vector (Invitrogen) was used as the backbone for the HI-IgAP *P. pastoris* expression vector. The recombinant construct consisted of an 85-residue N-terminal α-MF fused to residues 26-1006 of the HI-IgAP sequence (UniProtKB entry #P45384), followed by a C-terminal hexahistidine tag (His tag). This sequence was codon-optimized for *P. pastoris*, synthesized, and cloned into the EcoRI/Acc65I sites of pPINK-HC using the services of GenScript USA. The codon-optimized gene sequence can be found in *Appendix A* and its corresponding primary sequence in *Appendix B.*

Invitrogen's PichiaPink Strain 4 *P. pastoris* cells (genotype *ade2*$^{-}$ *prb1*$^{-}$ *pep4*$^{-}$) were used for protein expression while *E. coli* DH5α cells (Thermo) were used for DNA maintenance. To prepare the plasmid, the vector was first linearized with EcoRI (Thermo) and dephosphorylated with alkaline phosphatase (FastAP; Thermo) in a one-pot reaction at 37^{o}C for an hour. The enzymes were then heat inactivated at 95^{o}C for 10 minutes. EcoRI was chosen as the linearization enzyme as it flanked the AOX1 promoter used for host plasmid integration. Competent cells were made and used according to Lin-Cereghino *et al.*'s condensed protocol.[51] 40 μL of fresh electrocompetent cells were electroporated with 100 ng linearized plasmid at 1.5 kV in a 0.2 cm MicroPulser electroporation cuvette (BioRad) using an Eppendorf 220V 940000017 electroporator. Measured electroporation time constants were between 4 and 5 msec.

Electroporated cells were then grown on adenine-lacking (PAD) plates (United States Biological) at 30^{o}C for three days to develop colonies. Successful transformants were selected by color. If no negative transformants were identified, the plates were placed at 4^{o}C until pigmentation

was seen. Positive transformants were further verified by colony PCR using universal AOX1/CYC1 primers, according to the protocol in the Invitrogen PichiaPink manual. Colonies verified by colony PCR were then grown to saturation in PAD media, diluted with 50% (v/v) glycerol to a final concentration of 25% (v/v) glycerol, and placed at -80^{o}C for long-term storage.

3.3 – Optimizing Protein Expression

PCR-verified transformants were then used in protein expression experiments as summarized in the Invitrogen PichiaPink manual. The following descriptions outline the main steps, their importance, and the parameters used in the first successful induction of HI-IgAP:

1. *Starter culture*

   A starter culture is often used as the source of cells for initiating protein expression experiments as it allows for control over cell density, phase of cell growth, and consequently the growth rates prior to and during induction.[52] To grow a HI-IgAP *P. pastoris* starter culture, a volume of PAD media was inoculated with a PCR-verified colony or glycerol stock. The volume of media inoculated varied depending on the desired volume of induced culture (see the "Log-phase growth and starvation" section below). The starter culture was then grown at 30^{o}C for three days in 1 L vented-cap baffled flasks (VWR) with shaking at 150 rpm. A 3:1 headspace to media ratio was used to ensure sufficient aeration for optimal growth.

2. *Log-phase growth and starvation*

   P. pastoris cell density was then increased by inoculating glycerol-amended nutrient-rich media (BMGY; G for glycerol) with the starter culture at a 1:10 PAD inoculum to BMGY ratio. For example, 1 L of BMGY was inoculated with 100 mL of PAD starter culture, for a final total volume of 1.1 L. The glycerol was used as a high-energy supplement to facilitate high-density cell growth prior to protein induction.[47] Cells in BMGY were grown at 30^{o}C for 18-24 hours in vented-cap baffled flasks, with a 3:1 headspace to media ratio and shaking at 150 rpm. The media was then centrifuged for 10 minutes at 3,000xg to harvest the cells. The cell pellet was resuspended in BMGY without glycerol (BMWY; W for water) to starve the cells, allowing for the total consumption of residual glycerol, which would otherwise inhibit methanol induction.[46,47] Cells in BMWY were incubated under identical conditions to BMGY.

3. *Methanol induction*

After an overnight growth, the media was centrifuged for 10 minutes at 3,000xg and resuspended in BMWY with 1% (v/v) methanol (BMMY; M for methanol) to begin induction. Cells in BMMY were grown at 30°C in 1 L vented-cap baffled flasks, with a 3:1 headspace to media ratio and vigorous shaking at 200 rpm. After induction, the cells were centrifuged for 10 minutes at 3,000xg and the supernatant was immediately used for purification.

The first set of inductions took place over five days, with an additional daily dose of 100% methanol added to the culture to a final concentration of 1% (v/v) methanol. This was done under the assumption that all of the methanol was metabolized prior to the addition of new methanol, as methanol concentrations were not measured or controlled for during expression.

Samples of media supernatant were taken on each day of induction to test for protein expression using trichloroacetic acid (TCA) precipitation[53] and activity by monitoring IgA1 cleavage using SDS-PAGE. *Fig. 3.3.1* shows the promising results of HI-IgAP expression and activity after the first day of induction. HI-IgAP supernatant concentrations however, failed to rise after the first day (*Fig. 3.3.2*), suggesting that no new protease was produced past the initial burst of protein expression. The enzyme even seemed to degrade over the 5-day induction period, evident by the simultaneous disappearance of HI-IgAP band intensity and appearance of many degradation products (*Fig. 3.3.2*, red arrow). To prevent further degradation, the overnight growth in BMWY was removed as it did not impact protein induction or yield. As the glycerol carried over from BMGY into BMMY was still preferentially used over methanol, residual glycerol was still depleted to allow for methanol induction without switching to BMWY.[54]

Ultimately, these attempts to increase the enzyme yield during induction were unsuccessful. Because I did not quantify cell culture densities or methanol concentrations over the course of the experiment, the reasons for these issues still remains inconclusive. However, the observed phenomena does suggest a possible cause, which is discussed in this section's conclusion (Section 3.6).

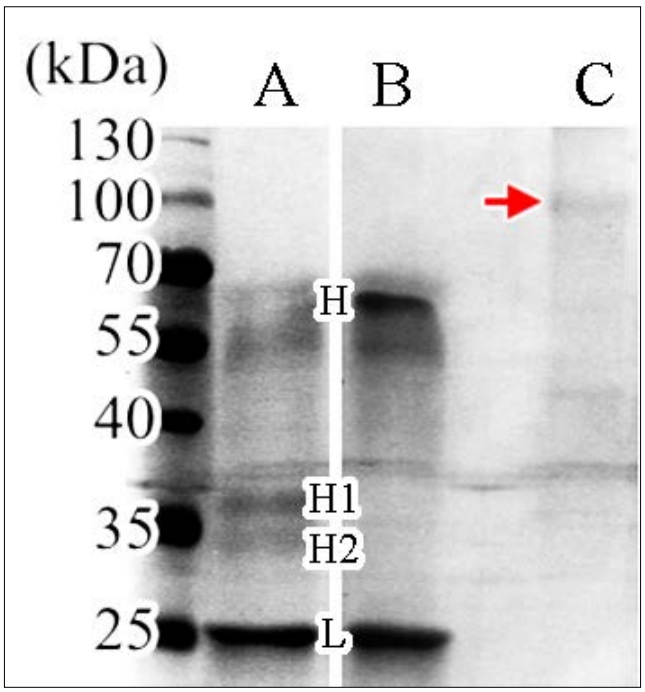

Figure 3.3.1 – Expression and activity of HI-IgAP produced from <u>P. pastoris</u>. Lane A shows the activity of *P. pastoris*-produced HI-IgAP via IgA1 cleavage. 5 µL of induced supernatant was incubated with 15 µL of 0.3 mg/mL IgA1 (Athens Research; uncleaved in lane B) at 37ºC for 30 minutes. The IgA1 heavy chain (H) was fully cleaved into fragments H1 and H2 (labelled according to *Fig. 1.2.1*). 1 mL of induced supernatant was TCA precipitated and loaded into lane C. The red arrow highlights the HI-IgAP band.

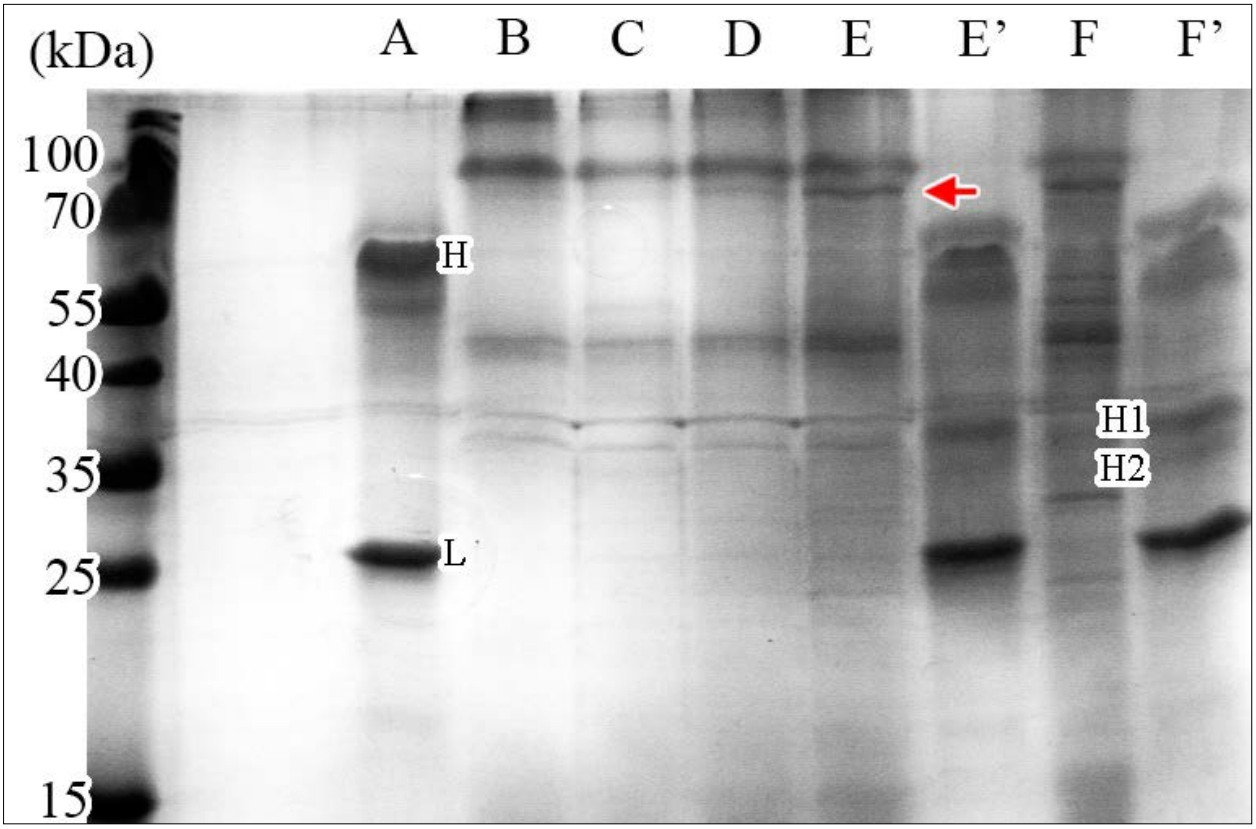

Figure 3.3.2 – Time course of the <u>P. pastoris</u> HI-IgAP induction profile. 1 mL of supernatant was TCA precipitated on each day of the induction period, with lanes B to F corresponding to days 1 to 5 post-induction. HI-IgAP activity on days 4 (E') and 5 (F') were analyzed as previously described. Lane A is the uncleaved IgA1 negative control. The red arrow highlights one of the many degradation products that appeared in significant amounts starting on day 3 (lane D).

Instead of attempting to optimize enzyme yield during expression, my focus shifted towards finding ways to increase yield by changing conditions prior to induction. As HI-IgAP expression seemed to occur all-at-once, I hypothesized that further increasing the cell density past a one-day BMGY growth may improve yields as a result of the larger cell mass available for enzyme production. This did not hold true as enzyme yield significantly decreased when cells were induced after two or more days of BMGY growth (*Fig. 3.3.3*, left). This was possibly due to the cells no longer being in the exponential phase of growth optimal for protein expression.[46,50] Changing the methanol concentration used during induction showed a greater effect on yield than BMGY growth time (*Fig. 3.3.3*, right). However, differences in yield between the first (*Fig. 3.3.2*) and this (*Fig. 3.3.3*, 2.0% lane) expression were negligible. Unfortunately, since the culture's growth density and cell viability were not measured, it was hard to conclude whether this lack of expression was due to changes in growth phase or because of other confounding variables.

Despite my attempts at optimizing expression, little to no increase in yield was seen. A more in-depth study of the factors underlying expression therefore needs to be done to fully understand the nuances of this *P. pastoris* system.

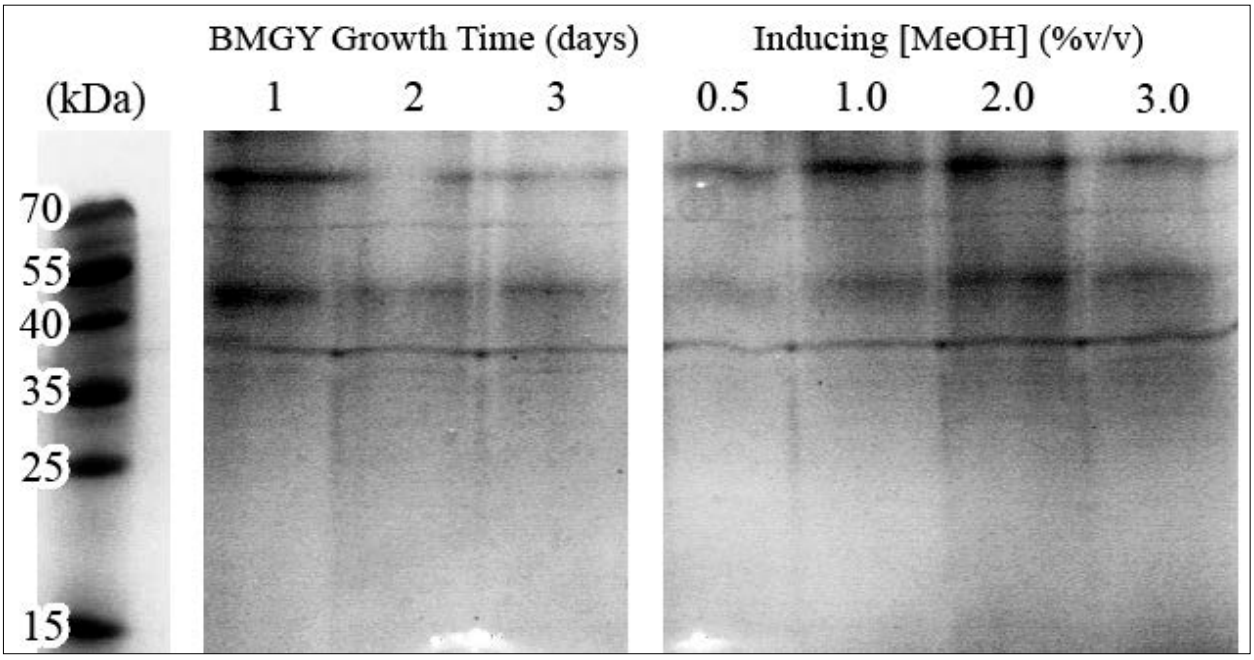

Figure 3.3.3 – Optimizing <u>P. pastoris</u> HI-IgAP production. Cells were grown under different expression conditions to optimize for HI-IgAP yield. This gel shows the TCA-precipitated profile of 1 mL supernatant collected one day after induction. The best cell density for expression, out of the experimental conditions, was obtained after only one day of BMGY growth. The best methanol (MeOH) concentration for a one-day BMGY growth, out of the experimental conditions, was around 2.0% methanol.

3.4 – Optimizing Protein Purification

Before the purification of HI-IgAP from *P. pastoris* cell culture supernatant is described, an important aspect in this section's SDS-PAGE gels must first be addressed: the apparent molecular weight of the HI-IgAP is shown to shift, ranging from just under 100 kDa up to 130 kDa. This is seen between enzyme produced from the same cell stock but in different expression batches. The cause of this inconsistency was never resolved, as mass spectrometry was never performed on the shifting bands. It is therefore assumed that the predominant band of the expression profile represented the HI-IgAP. As fractions containing this band always showed activity, and those lacking this band showed no activity, it was a reasonable assumption to make.

Each purification attempt used a minimum of 2 L culture to compensate for the poor yield of each expression batch (*Fig. 3.3.1*). Affinity chromatography was therefore used as the go-to purification method to take advantage of the construct's C-terminal His tag. Our lab has previous experience using nickel-chelated immobilized metal affinity chromatography (Ni-IMAC) to capture HI-IgAP secreted from *H. influenzae*.[20] As such, we knew that the components in sugar-rich media, including those typically used to grow *H. influenzae* and *P. pastoris*, easily outcompeted HI-IgAP for affinity resin binding. An excessively large amount of Ni-IMAC resin was therefore used to bind the enzyme directly from the supernatant. 100 mL of settled Ni-IMAC resin (Ni-NTA; Molecular Cloning Laboratories) equilibrated in 25 mM HEPES (BioShop), pH 7.5 was incubated with the supernatant for an hour at 4^{o}C. The resin was then washed with 2 L HEPES buffer before being eluted with a buffer containing 25 mM HEPES and 0.3 M imidazole, pH 7.5.

Due to non-specific binding of unidentified contaminants within the media, the Ni-IMAC eluent had a dark yellow to deep brown hue. Adding 0.5 to 1 M NaCl to the wash buffer did not noticeably change this discoloration. Spectrophotometric methods (i.e. A_{280} measurements) thus could not be used to determine the protein concentration as the media had partial absorbance at 280 nm. Instead, concentrations were semi-quantitatively assessed by comparing SDS-PAGE band intensities to other bands with known concentrations.

Despite a C-terminal His tag present in the construct, the Ni-IMAC eluent contained no enzyme as the eluent showed no activity and HI-IgAP was not seen in its TCA precipitation (*Fig. 3.4.1*, lane A). The opposite was true for the flow through (*Fig. 3.4.1*, lane B). I hypothesized that this lack of binding was due to the relatively low pH of the media supernatant. As BMMY is heavily buffered at pH 6.0 for optimal growth, the His tag's imidazole groups were likely only semi-deprotonated as their approximate pKa of 6.6 comes too close to this pH.[55] Unfortunately the enzyme still did not bind even after the supernatant pH was raised to 8.0 using 1 M pH 12.0 K_3PO_4 post-expression. An anti-His-tag dot blot done on the induced supernatant and cell pellet (*Fig. 3.4.1*, right) suggested that the secreted HI-IgAP did not contain a His tag despite it being part of the expression construct. This may have been due to non-specific autoproteolysis of the C-terminus after secretion, which is known to occur in the native *H. influenzae* system.[20]

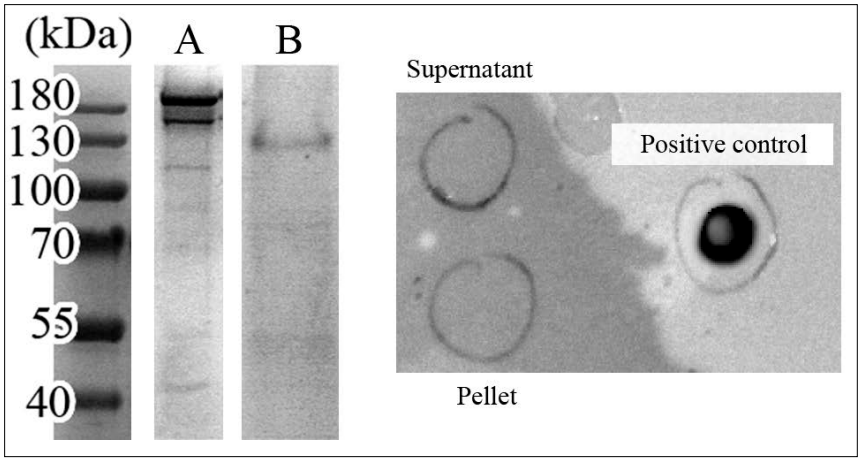

Figure 3.4.1 – Ni-IMAC eluent and flow-through profiles. This gel shows the non-specific binding of high-molecular-weight contaminants (A) to the Ni-IMAC resin. Lane B is the 1 mL TCA-precipitation of the flow through, which contained all of the secreted HI-IgAP. The anti-His tag dot blot (right) shown is an overlay of the samples' position and corresponding fluorescence, highlighting the lack of a His-tag on the secreted HI-IgAP.

A strong anion-exchange resin (Q Sepharose; GE) was used as an alternative to Ni-IMAC. This was used to concentrate the enzyme for downstream purification while removing problematic media components. Q Sepharose's much higher binding capacity, relative to Ni-IMAC, also allowed the enzyme to bind even with non-specific interactions with the media. Since BMMY has high ionic strength, ammonium sulfate precipitation was first used to lower the ion concentration by precipitating and resuspending the enzyme in no-salt buffer. 75 g of ammonium sulfate was added to every 100 mL of BMMY supernatant. The mixture was stirred at 4°C for at least an hour and formed an opaque solution. The precipitated slurry was centrifuged at 10,000xg and 4°C for an hour and the supernatant was discarded. The pellet was dissolved in and diluted to 2L with 25 mM Tris base (BioBasic), pH 7.5 and incubated with 200 mL settled Q Sepharose resin equilibrated in the same buffer. Q Sepharose resin was then washed with 2 L Tris buffer and eluted

with 25 mM Tris, pH 7.5 supplemented with 2 M NaCl. Although the eluent (*Fig. 3.4.2*, lane A) still retained its strong yellow color, most of the color was removed through this process. The Q Sepharose eluent still would not bind to Ni-IMAC resin despite removing most of the contaminating media components, reinforcing the idea that the His tag was post-translationally removed during expression.

At this point, the approximate yield of a 2 L induction was less than a milligram of enzyme. As any additional steps led to a significant loss in yield, I decided to forego further purification until the issue of yield could be resolved. Thus, the anion-exchange eluent was immediately concentrated using a 30 kDa centrifugal concentrator (Amicon) to under 1 mL and flash frozen with liquid nitrogen. The concentrated protein (*Fig. 3.4.2*, lane B) still retained a strong yellow color despite efforts to de-color the solution with prolonged and successive dialyses.

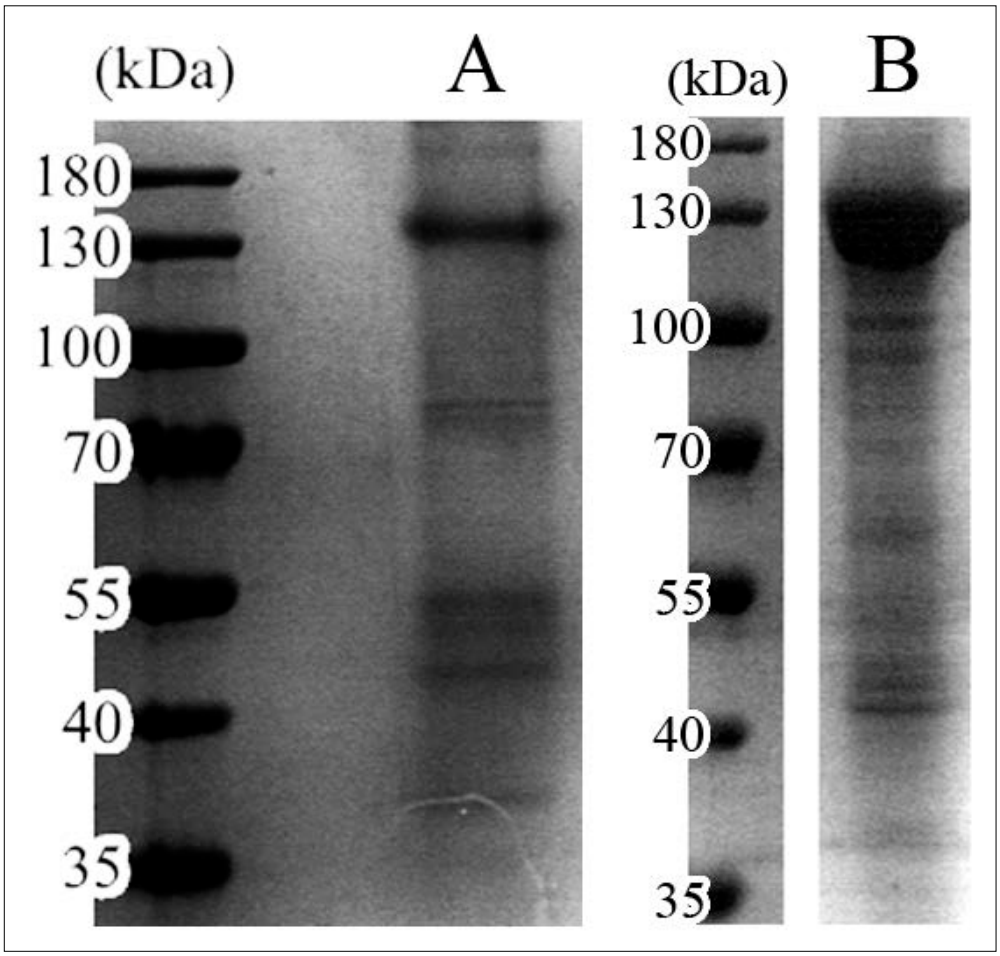

Figure 3.4.2 – Q Sepharose elution profile. The protein profile of 1 mL TCA-precipitated Q Sepharose eluent is shown in lane A and is similar to that shown in *Fig. 3.3.4*. The final purity of the enzyme after concentration is shown in lane B.

3.5 – Finalized P. pastoris Protein Production Protocol

My general *P. pastoris* expression and purification workflow is summarized below. All pH values were measured at room temperature. All cultures were grown in 1 L baffled flasks with a 3:1 headspace to media ratio. Flasks were shaken at 150 rpm unless otherwise specified.

1. Grow a PAD starter culture from a glycerol stock or colony for three days at 30°C.
2. Inoculate BMGY at a 1:10 PAD inoculum to BMGY ratio and grow overnight at 30°C.
3. Cells were centrifuged at 3,000xg for 15 minutes and transferred to BMMY containing 2% (v/v) methanol. Cells were grown overnight at 30°C with shaking at 200 rpm.
4. The supernatant was harvested using centrifugation at 3,000xg for 15 minutes and the cell pellet discarded.
5. 75 g of ammonium sulfate was added to every 100 mL supernatant. The mixture was stirred at 4°C until the solution turned opaque.
6. The precipitated enzyme was harvested by centrifugation at 10,000xg for 1 hour at 4°C.
7. The pellet was dissolved and incubated in 100 mL Q Sepharose (GE) resin pre-equilibrated in 25 mM Tris (BioBasic), pH 7.5.
8. The Q Sepharose resin was washed with 1 L 25 mM Tris, pH 7.5 and eluted with 1 L 25 mM Tris, pH 7.5 supplemented with 1 M NaCl.
9. Eluent was concentrated to less than 5 mL, using 30 kDa MWCO stirred-cell and centrifugal concentrators (Amicon), and loaded onto a manually packed Bio-Gel P-6DG desalting column (Bio-Rad) pre-equilibrated in 25 mM HEPES, pH 7.5.
10. The buffer-exchanged fractions were concentrated to less than 1 mL using a 30 kDa MWCO centrifugal concentrator and flash frozen in aliquots with liquid nitrogen.

The final yield was approximately 0.9 mg enzyme from a 2 L expression, as semi-quantitatively estimated by SDS-PAGE. Based on inspection by eye, the final purified enzyme was more than 90% pure and is shown in lane B of *Fig. 3.4.2*.

3.6 – Shortcomings and Next Steps

There are several shortcomings to this system. Most notably, its low yield coupled with co-purification of media components made the already lengthy purification even more difficult and resource-intensive. The lack of a His tag suggested that the C-terminus can be re-engineered to prevent autoproteolysis, potentially mitigating some loss during purification. Yields could also be improved by using a bioreactor, which optimizes oxygen and methanol concentrations in real time to form high-cell-density cultures that express better than shaking vented-cap flasks.[47]

The fact that expression occurred as a one-time response to methanol suggested that the vector failed to integrate into the *P. pastoris* genome, as it should have, and thus was discarded mid-induction.[54,56] When the cells were transferred from PAD into BMGY media, the auxotrophic pressure of PAD media was removed as BMGY is nutrient-rich. Without any selective pressure, the plasmid was readily discarded to lessen the energetic cost associated with its expression.[54] Linearizing the expression vector within the *TRP2* gene may have been a better alternative to host integration over linearizing the AOX1 promoter. Positive transformants can also be tested for successful genomic integration via PCR. Furthermore, the number of copies integrated should be screened for either by PCR or small-scale expressions before protein expression is scaled up.

Section 4 – Developing the Second-Generation *E. coli* Cytosolic Expression System

I was fortunate to have an opportunity to intern in our collaborator's lab (Dr. Elan Z. Eisenmesser; University of Colorado, Denver, Anschutz Medical Campus) through our existing partnership on the IgAP project. His lab uses nuclear magnetic resonance (NMR) to look at the dynamics and structure-function relationships of immune-system proteins. Because of the notable differences between how they use NMR and how our lab uses X-ray crystallography to answer similar biochemical questions, I was able to learn a complimentary set of techniques that sparked the idea for a second-generation *E. coli* HI-IgAP production system. Particularly, their methods of denatured purification, used as part of their NMR toolkit, was something that our lab never had success with and therefore rarely used. Although we attempted denaturing and refolding techniques during the development of our first-generation *E. coli* system (Section 2.3),[44] my internship experiences showed me that we needed to make major adjustments to our previous approach before we even had a chance at success.

4.1 – Alternative Approach to the E. coli Cytosolic Expression System

My new experiences with protein denaturation made me reconsider our first-generation *E. coli* system, precisely because the HI-IgAP produced almost always formed insoluble inclusion bodies (Section 2.2 and 2.3). Although these aggregates could not be previously resolubilized,[44] their insolubility would provide several advantages over traditional methods of soluble purification, given that a feasible method of resolubilizing and refolding could be developed. Insolubility not only protected the enzyme from proteolysis during expression, as most molecules in inclusion bodies are not accessible to solvent,[57] but also allowed for the separation of a vast majority of soluble contaminants by isolating and purifying just the insoluble cell fraction. The success of this new inclusion-body purification was contingent on two criteria, both of which were lacking in our first-generation approach. One, all exogenous N-terminal residues must be removed before the enzyme is refolded (Section 2.2). And two, the refolding process must be altered to slow the rate of precipitation and aggregation to allow for proper protein folding.

The solution to the first criterion involved modifying the first-generation expression construct to better suit a denatured approach. This new construct (*Fig. 4.1.1*; sequences in *Appendices C* and *D*) comprised of a small ubiquitin-like modifier (SUMO; UniProtKB entry #A6ZZ98) N-terminally fused to the HI-IgAP protease domain. After cleavage with ubiquitin-like-

specific protease 1 (ULP1; UniProtKB entry #Q02724), which recognizes the SUMO tertiary structure and cleaves C-terminal to a double-glycine motif (*Fig. 4.1.1*),[58] no SUMO residues are left behind. Without any exogenous N-terminal residues, the protease is now able to properly fold. This is similar to the N-terminal processing involved in the *P. pastoris* α-MF secretory pathway (Section 3.1). The plasmid expressing ULP1 was a gift from Dr. Christopher Lima (Sloan Kettering Institute) and its purification protocol is described in *Appendix E*.

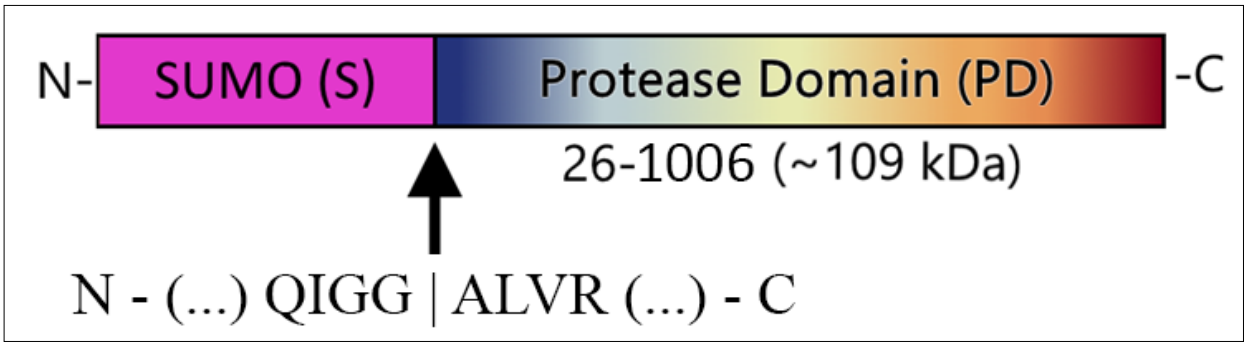

Figure 4.1.1 – Schematic representation of the second-generation <u>E. coli</u> HI-IgAP construct. The primary sequence surrounding the ULP1 cleavage site ("|") is shown. After cleavage, the primary sequence starts at A26, the first amino acid of the true secreted protein (*Fig. 2.2.1*).

4.2 – Vector Design

The second-generation *E. coli* expression vector (*Fig. 4.1.1*; SUMO-HI-IgAP) was created in two steps. First, the sequence coding for untagged SUMO was cloned upstream of the HI-IgAP gene. The second round of cloning simultaneously modified the vector in two ways: it ensured that the HI-IgAP sequence after ULP1 cleavage started at A26, and it moved the His tag from the HI-IgAP C-terminus to the SUMO N-terminus. The location of the His tag was switched to allow for the separation of His-tagged contaminants from untagged and cleaved HI-IgAP (*Fig. 4.3.1*).

The first-generation Δ27 *E. coli* expression vector (Section 2.3) was used as the starting point for modifications. To prevent complications with autoproteolysis,[20] the inactive mutant (Δ27 S288A) was used. The Takara Bio In-Fusion Cloning kit was used in both cloning steps to simplify Gibson assembly methods. Hence, all primers contained a 20 to 30 bp 5' overhang engineered to be complimentary to the sequences flanking the insertion site (*Table 4.2*). For the first stage of cloning, the SUMO gene was PCR amplified from an expression vector containing a SUMO-fused protein. The PCR product was then gel purified (Qiagen) and cloned into the NdeI site of the Δ27 S288A pET24b vector using the In-Fusion kit.

Table 4.2 – Primers used to create the second-generation <u>E. coli</u> expression vector. NdeI was used to linearize all vectors. Its restriction site is highlighted in bold. Primers were ordered dry from Integrated DNA Technologies.

Sequence (5' to 3')	Type	Function
ATT TTG TTT AAC TTT AAG AAG GAG ATA TA**C ATA TG**T CGG ACT CAG AAG TCA ATG AAG	forward	To amplify the SUMO gene for Gibson assembly into the NdeI site upstream of the Δ27 HI-IgAP in pET24b
GAT AAT CCA CAT CGT CTC TCA C**CA TAT G**AC CTC CAA TCT GTT CGC GGT G	reverse	
CTT TAA GAA GGA GAT ATA **CAT ATG** ATG GGT CAT CAC CAT CAT CAT CAC GGG TCG GAC	forward	To generate a large forward primer which includes the A26, L27, and N-terminal His tag
GAT AAT CCA CAT CGT CTC TCA CTA AGG CAC CTC CAA TCT GTT C	reverse	
The ~1,100 bp sequence generated from the PCR amplification of above	forward	To amplify the SUMO-HI-IgAP gene with an N-terminal His tag for Gibson assembly into the NdeI site of an empty pET21 vector
CAC CAG TCA TGC TAG C**CA TAT G**TT AAG GTG TTG TGA TAT TTG TCG TAT CGA CAG TTT G	reverse	

As the Δ27 S288A construct was used as the template, residues A27 and L28 were missing and had to be added to the sequence after the ULP1 cleavage site (*Fig. 4.1.1*). This was done with two consecutive PCR reactions. The first reaction amplified a DNA fragment which coded for an N-terminal His tag, SUMO, and part of the HI-IgAP sequence containing A27 and L28. This large fragment was used as the forward primer for the subsequent PCR reaction, where the rest of the HI-IgAP sequence, excluding the C-terminal His tag, was amplified. This was then inserted into the NdeI site of a pET21 empty vector (Novagen). BL21(DE3) *E. coli* cells were transformed with the expression vector and subsequently stored in 25% (v/v) glycerol for future expressions. The construct sequence was verified after each cloning step (Integrated DNA Technologies).

4.3 – Overview of the Second-Generation Cytosolic Production System in <u>E. coli</u>

A general overview of the preliminary protein production protocol is described below and summarized in *Fig. 4.3.1*. The components of each preliminary buffer are listed in *Table 4.4, column A*. All steps were done at 4°C, unless otherwise specified.

1. 4L culture was grown and induced in ZY-5052 media via autoinduction at 20°C for 40-48 hours.[45] All media were supplemented with 50 μg/mL ampicillin.
2. Cells were harvested using centrifugation for 10 minutes at 6,000xg.
3. Cells were resolubilized in 100 mL Resuspension Buffer, lysed via two passes through a FRENCH pressure cell (Thermo) at 1100 psi, and centrifuged for an hour at 12,000xg.
4. The supernatant was discarded and the pellet was fully dissolved in 50 mL Denaturing Buffer.

5. The protein was incubated with 10 mL settled Ni-IMAC resin (NiNTA; Molecular Cloning Lab), pre-equilibrated in Denaturing Buffer, for an hour with agitation.
6. The Ni-IMAC resin was washed with 1 L Wash Buffer and eluted with Elution Buffer until A_{280} readings were under 0.05.
7. Eluent was dialyzed overnight with 14 kDa MWCO dialysis tubes (Aldon Corporation) into Arginine Buffer with a 10-fold-excess dialysis volume.
8. The dialysis buffer was changed to Refold Buffer and the protein was dialyzed overnight.
9. The dialyzed solution was incubated with 3 mg homemade ULP1 for three hours at room temperature on a rocking shaker. Precipitated protein was removed by centrifugation at 12,000xg for 15 minutes.
10. 50 mL of Ni-IMAC resin, equilibrated in Storage Buffer, was incubated with the ULP1-treated protein for an hour.
11. The resin was rinsed with two column volumes of Storage Buffer and the flow through and rinse solution were concentrated with 30 kDa stirred-cell and centrifugal concentrators (Amicon) to less than 5 mL.
12. The concentrated protein was injected onto a Sephacryl S-200 HR (GE) size-exclusion chromatography (SEC) column, equilibrated in Storage Buffer. The column was operated by a Biologic DuoFlow FPLC (Bio-Rad) and ran at 0.5 mL/min.
13. Fractions containing non-aggregated protein were concentrated using a 30 kDa centrifugal concentrator to less than 1 mL and frozen with liquid nitrogen.

Refolding protocols commonly use a one-step dialysis to transition the protein from its unfolded state, in high concentrations of denaturant (e.g. guanidinium hydrochloride; GdnHCl), to a solution with no denaturant.[59] Through this process, the protein ideally follows a thermodynamically favorable path towards a proper fold.[59] This was how we attempted our first HI-IgAP refold (Section 2.3).[44] However, this one-step refold is often unsuccessful because of a problem arising from the kinetic relationship between protein precipitation and refolding. There are two paths available to the denatured protein when denaturants are removed: one towards the desired folded state, assuming it can occur under the refolding conditions, and the other towards aggregation.[60] Although both paths are available, the rate of aggregation is typically much higher than that of folding due to the greater frequency of unwanted protein-protein interactions that occur when core side chains are exposed.[60] This is exacerbated by the abnormally high concentrations

of unfolded protein during the refold, compared to that of the natural folding environment.[59] Overall, this leads to the vast majority of protein precipitating instead of refolding.

A potential solution is the addition of an intermediate refolding step to bridge the gap between the totally denatured and natively folded states. The HI-IgAP was therefore dialyzed from 5 M GdnHCl first into 1 M arginine before subsequently dialyzed into a soluble buffer. Rather than acting as a molecular chaperone, arginine works as a molecular crowder, simulating the packed cellular environment in which protein folding typically takes place.[60] This reduces the frequency and scale of unwanted protein-protein interactions, slowing the rate of precipitation enough to allow for protein folding to take place.[60] Since HI-IgAP naturally folds after passing through the outer membrane (Section 2.1), it stood to reason that it would do the same *in vitro*. We were lucky that this was the case.

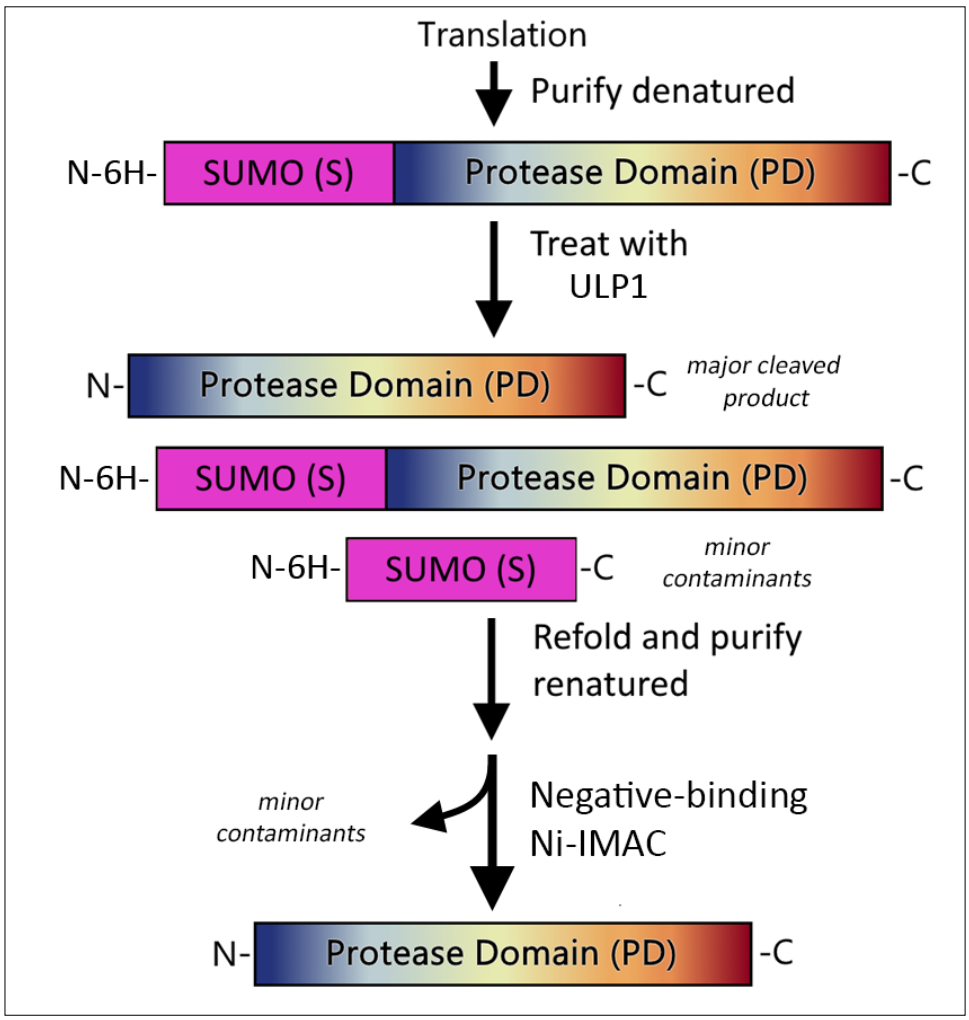

Figure 4.3.1 – Flowchart of the denatured HI-IgAP purification.

A typical denatured HI-IgAP purification is shown in *Fig. 4.3.2*. Samples with GdnHCl or arginine were prepared using ethanol precipitation: 1 mL of -80°C 100% ethanol was added to 50 μL of sample and set to precipitate for an hour at -80°C. The mixture was centrifuged at 14,000xg for a minute, the supernatant discarded, and the pellet was washed in -20°C 70% (v/v) ethanol. The sample was centrifuged at 14,000xg for a minute, the supernatant discarded, and residual ethanol was removed by heating the sample at 80°C for several minutes. SDS loading dye and water was added to a total volume of 200 μL. The sample was heated at 80°C until the pellet was completely dissolved. 10 μL of sample was added to each gel lane.

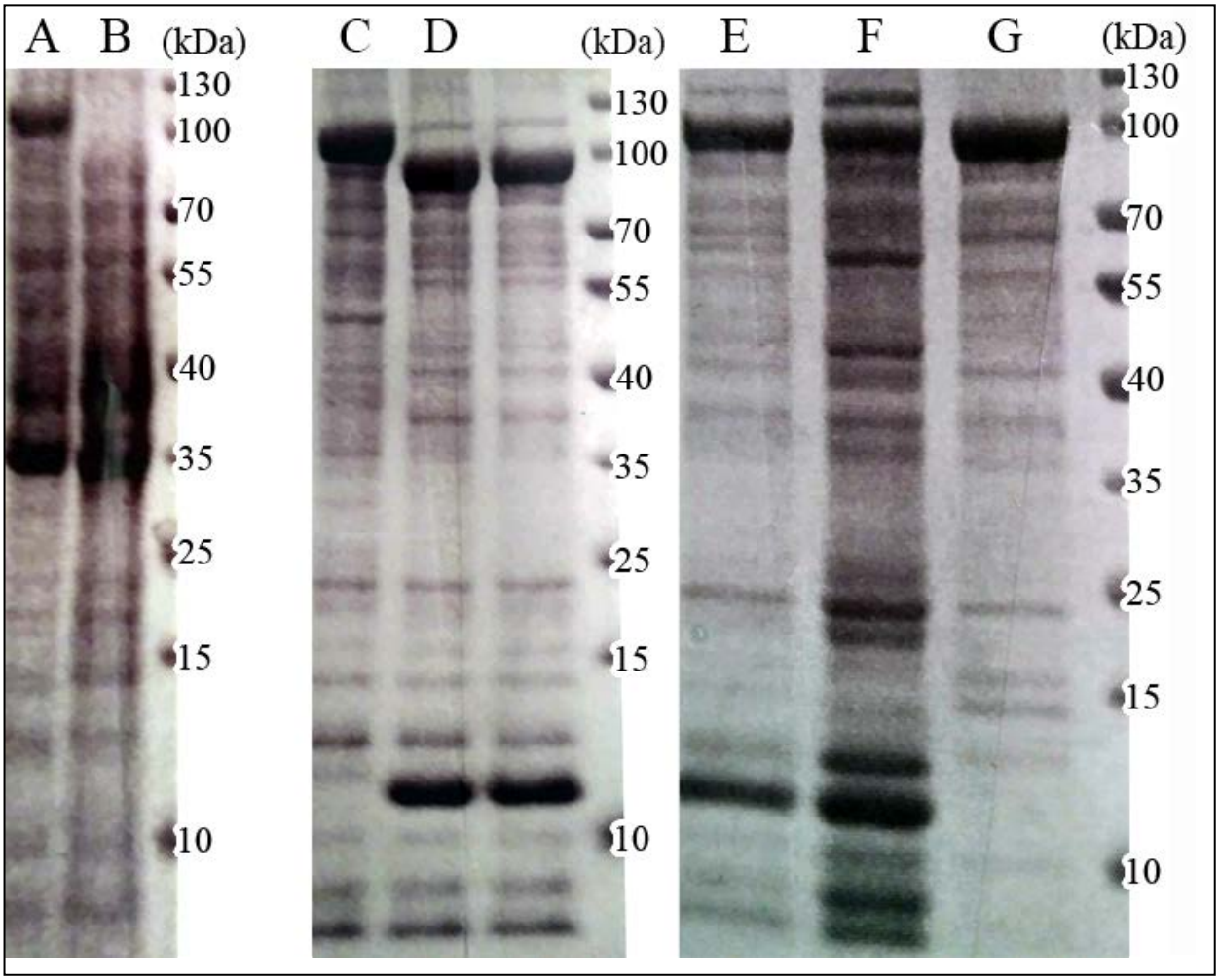

Figure 4.3.2 – Preliminary denatured HI-IgAP purification. With an N-terminal SUMO (~12 kDa), the new HI-IgAP construct was ~120 kDa. Lane identities include: (A) total induced; (B) uninduced; (C) 1st Ni-IMAC eluent; (D) ULP1-cleaved Ni-IMAC eluent; (E) Before the 2nd Ni-IMAC; (F) 2nd Ni-IMAC eluent; (G) 2nd Ni-IMAC flow through.

4.4 – Optimizing Protein Expression and Purification

This section focuses on optimizing several key steps of the purification (Section 4.3). Each key step will be detailed within its own subsection. All of the buffers referenced below have been amended to reflect this optimization. Their components are listed in *Table 4.4*, *column B*.

Protein Expression (Step #1)

Autoinduction was initially used to express this second-generation construct based on our previous success using it over IPTG in our first-generation system (Section 2.3). However, it turned out that the total amount of protein produced per volume of cell culture was almost identical between a two-day autoinduction at 20°C and a three-hour induction with 1 mM IPTG at 37°C. As such, there was no reason to continue using autoinductive methods due to its longer growth time and higher contaminant background. Furthermore, a higher incubation temperature also facilitated inclusion body formation, allowing for a greater level of protection against proteolysis during expression compared to lower temperatures.[57] That said, incubation temperatures beyond 37°C, and up to 42°C, still need to be tested to see if protein yield and purity can be further improved.[57]

Table 4.4 – Denatured HI-IgAP purification buffers. HEPES, Na_2HPO_4, NaCl, imidazole, and disodium EDTA dihydrate were purchased from BioBasic, GdnHCl and arginine from ChemImpex, and IgePal CA-630 from Sigma. All buffers have a pH of 7.5 at room temperature. All buffers were equilibrated to 4°C before use.

Buffer Name	**Components**	
	Preliminary (A)	**Optimized (B)**
Resuspension	25 mM HEPES	
Denaturing	50 mM sodium phosphate + 0.5 M NaCl + 10 mM imidazole + 5 M GdnHCl	
Wash	Same as Denaturing Buffer	Same as Denaturing Buffer but with 1% (v/v) IgePal CA-630 and no NaCl
Elution	Same as Denaturing Buffer but with 300 mM imidazole	
Arginine	100 mM Tris + 0.3 M NaCl + 1 M arginine	100 mM Tris + 0.3 M NaCl + 1 M arginine + 50 mM EDTA
Refold	50 mM sodium phosphate + 0.15 M NaCl	(*Refold A*) 50 mM sodium phosphate + 0.15 M NaCl + 0.3M GdnHCl + 50 mM EDTA
		(*Refold B*) 25 mM HEPES + 0.3M NaCl
Storage	25 mM HEPES	

Lysis (Steps #3 and 4)

Because SUMO-HI-IgAP is highly insoluble, the lysate pellet can be resuspended in a non-denaturing buffer to remove residual soluble contaminants. Thus after centrifuging the initial lysate, the pellet was resuspended in Resuspension Buffer for a second time and passed through the FRENCH press system to maximally solubilize contaminants. This mixture was centrifuged at 12,000xg and the supernatant was discarded. After the second centrifugation, there was a noticeable decrease in pellet size with no noticeable loss in IgAP. This pellet was then dissolved in Denaturing Buffer.

First Ni-IMAC (Step #5 and 6)

A large proportion of protein did not bind to the first denatured Ni-IMAC, even though the solution loaded onto the resin was homogenous and free of visible particles. Binding was significantly improved when the solution was first vigorously sonicated to break up and fully denature non-visible aggregates. This resulted in a 10-fold increase in Ni-IMAC binding.

Because of the Ni-IMAC was done under denaturing conditions, the resin washes could be modified to higher stringency to remove sticky contaminants that were bound through non-specific electrostatic and hydrophobic interactions. Those bound with weak electrostatic interactions were removed using the same high-salt Denaturing Buffer. Hydrophobically interacting contaminants were removed using the new Wash Buffer, which contained a high concentration of non-ionic detergent and no additional salt. Ten column volumes of Denaturing Buffer were used to remove residual detergent, and the protein was eluted using the same Elution Buffer.

Two-Step Refold (Steps #7 and 8)

Proteins undergoing refolding are particularly susceptible to proteolysis, particularly if other proteases fold faster than the protein of interest.[57] Thus, protease inhibitors are often added in the refold buffers to reduce degradation.[57] Initial refolding of the HI-IgAP displayed a consistent set of degradation products that would appear and increase in intensity throughout the refold (*Fig. 4.4.1*, red arrow). Upon addition of a metalloprotease inhibitor, EDTA, to the Arginine and Refold Buffers, degradation no longer occurred.

Even with an intermediate arginine step, a large proportion of the total enzyme precipitated when it was dialyzed from the Arginine Buffer into the preliminary Refold Buffer. This was addressed in two ways. The Ni-IMAC eluent was first diluted three-or-more fold with Denaturing

Buffer to lower the protein concentration and hence the rate of aggregation.[60] For example, a typical volume of 40 mL Ni-IMAC eluent was diluted to 200 mL for dialysis. A third refolding step was also added to facilitate folding and aid in ULP1 cleavage (see below).

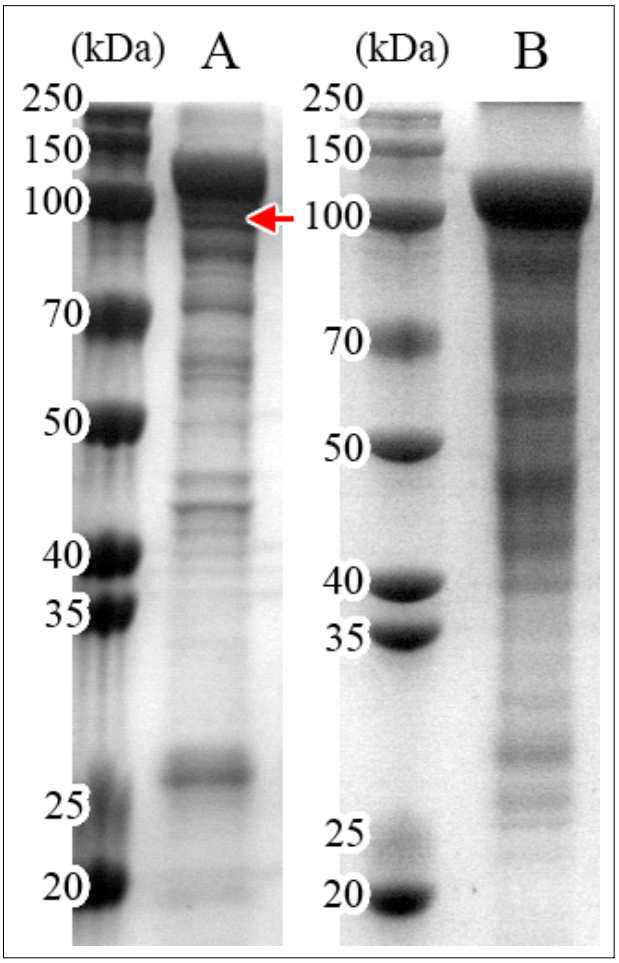

Figure 4.4.1 – Effect of EDTA on HI-IgAP refolding. This figure shows the effect of EDTA on degradation during refolding. Eluent from the denatured Ni-IMAC was dialyzed against Arginine and Refold Buffer with (lane B) and without EDTA (lane A). The red arrow highlights degradation products that are not present when the refold is done under EDTA.

ULP1 Cleavage (Step #9)

In addition to protein loss during refolding, heavy precipitation was seen after ULP1 treatment. As cleavage of SUMO is necessary for proper folding, it is not surprising that aggregation was especially prevalent during this step. To increase the solubility and assist in the folding of cleaved HI-IgAP, ULP1 cleavage was tested under various mildly denaturing conditions. The condition which provided the best solubility with the least amount of degradation was 0.3 M GdnHCl (*Fig. 4.4.2*, lanes C and C'). Unwanted degradation occurred when SUMO was cleaved under urea (*Fig. 4.4.2*, red arrow). Therefore, Ni-IMAC eluent was first dialyzed into Arginine Buffer then into buffer with 0.3 M GdnHCl (Refold A Buffer), cleaved with ULP1 at room temperature, then dialyzed once more time into the Refold B Buffer with no denaturant.

The tradeoff of cleaving SUMO under semi-denaturing conditions was the inevitable slowing of ULP1 activity.[58] Monitoring the reaction over several hours at room temperature showed that full SUMO cleavage could not be achieved using only one three-milligram dose of ULP1 (*Fig. 4.4.3*). A minimum of two doses and overnight incubation were necessary for near-complete cleavage. As HI-IgAP seemed to be stable at room temperature, given the lack of degradation products in *Fig. 4.4.3*, lane E, further optimizing cleavage conditions will only reduce the reaction time and amount of ULP1 used but is not essential to obtaining pure HI-IgAP.

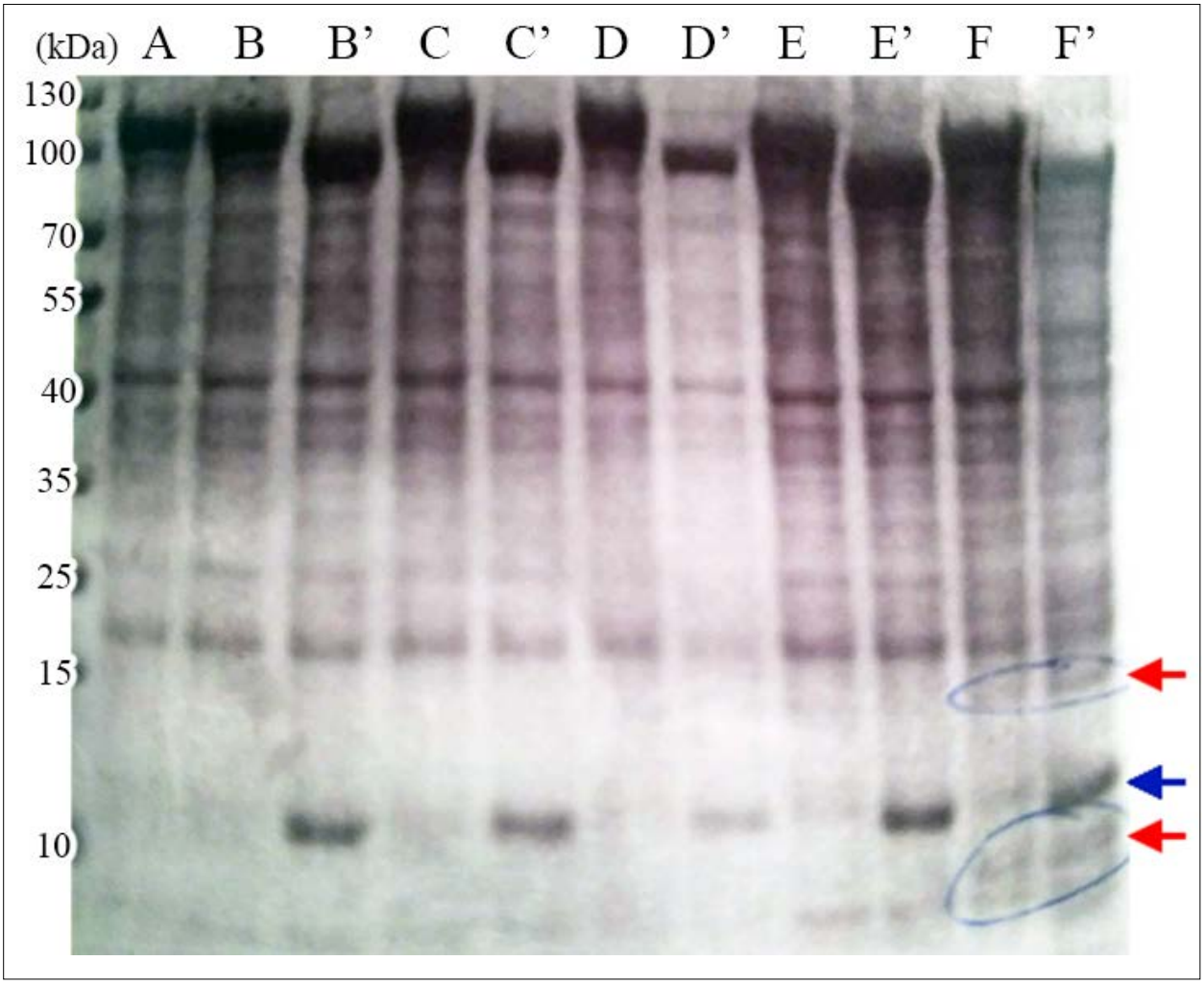

Figure 4.4.2 – Effect of mild denaturant on ULP1 cleavage. Ni-IMAC eluent was first dialyzed into Arginine Buffer then into the following buffers. All buffers contained 50 mM phosphate, pH 7.5, 150 mM NaCl, and a mild denaturant. All lanes with prime labels (') represent post-ULP1-treated protein. Noticeable degradation products (red arrows) were seen in buffers containing urea. The blue arrow indicates the cleaved SUMO. The identities of the mild denaturants for each lane are: (A) Protein in Arginine Buffer; (B) No denaturant; (C) 0.3 M GdnHCl; (D) 0.6 M GdnHCl; (E) 1.5 M urea; (F) 2.5 M urea.

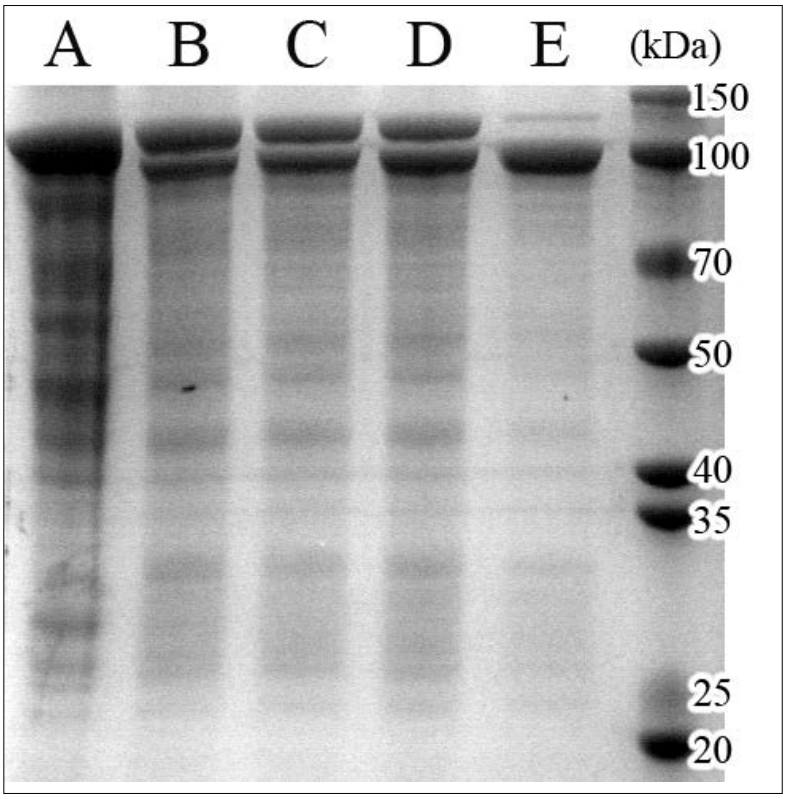

Figure 4.4.3 – Time course of ULP1 cleavage in Refold A Buffer. This figure shows the progress of SUMO cleavage over several hours. 3 mg ULP1 (one dose) was added to the protein (A) at room temperature. Samples were taken at 1 (B), 2 (C), and 3 (D) hours post-incubation. Lane E represents the overnight cleavage after 3 more milligrams of ULP1 were added.

Size-Exclusion Chromatography (Step #12)

Using size-exclusion chromatography (SEC) to separate HI-IgAP aggregates from correctly folded protein was very successful (*Fig. 4.4.4* and *4.4.5*). HI-IgAP that was uncleaved, and therefore improperly folded (*Fig. 4.4.4*, red arrow), formed large aggregates by associating with itself and other properly folded proteins. These complexes eluted at the void volume and were able to be separated from properly folded and non-aggregated HI-IgAP (*Fig. 4.4.5*). However, initial SEC runs showed that a large proportion of the properly folded protein was being pulled into the void volume by uncleaved HI-IgAP (*Fig. 4.4.4* and *4.4.5a*). When the protein was extensively dialyzed into a no-salt SEC buffer before running the column, there was a notable shift towards a non-aggregated state (*Fig. 4.4.5b*). At the moment, it is unclear whether this shift was due to changes in ionic strength or because of the prolonged dialysis. Hence, more experimentation needs to be done to optimize this purification step.

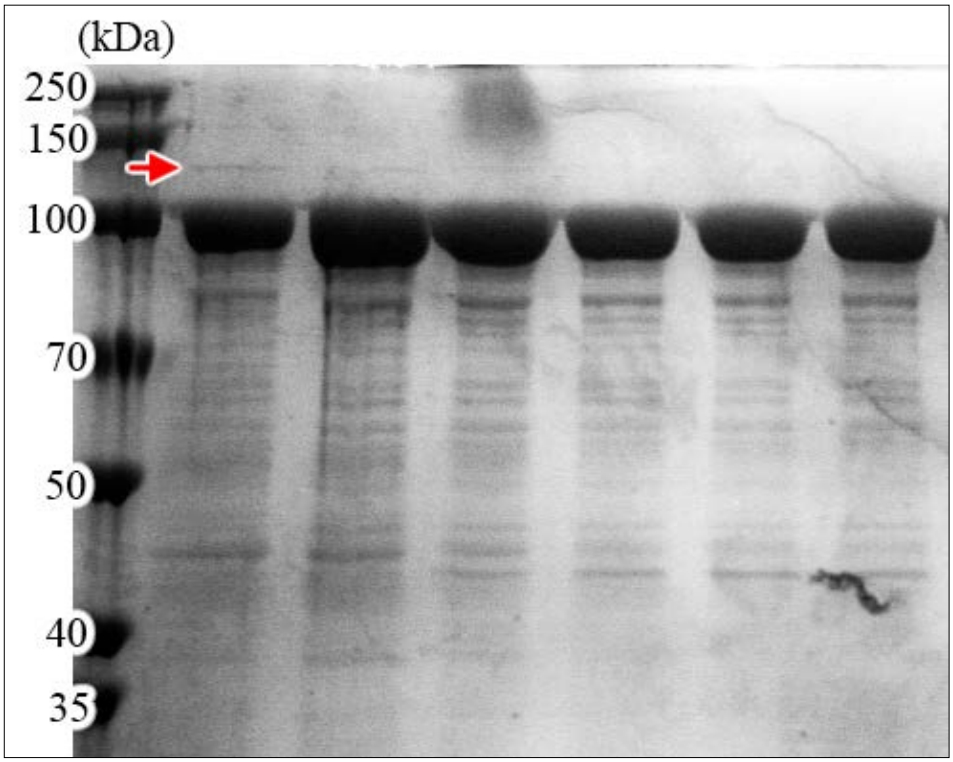

Figure 4.4.4 – S200 SEC fractions. This figure highlights the pull-down effect of aggregate-inducing uncleaved HI-IgAP (red arrow) on a S200 SEC column. The gel lanes span the two peaks of the *Fig. 4.4.5a* trace, where the void peak is represented in the left-most three lanes and the elution peak is represented in the right-most three lanes.

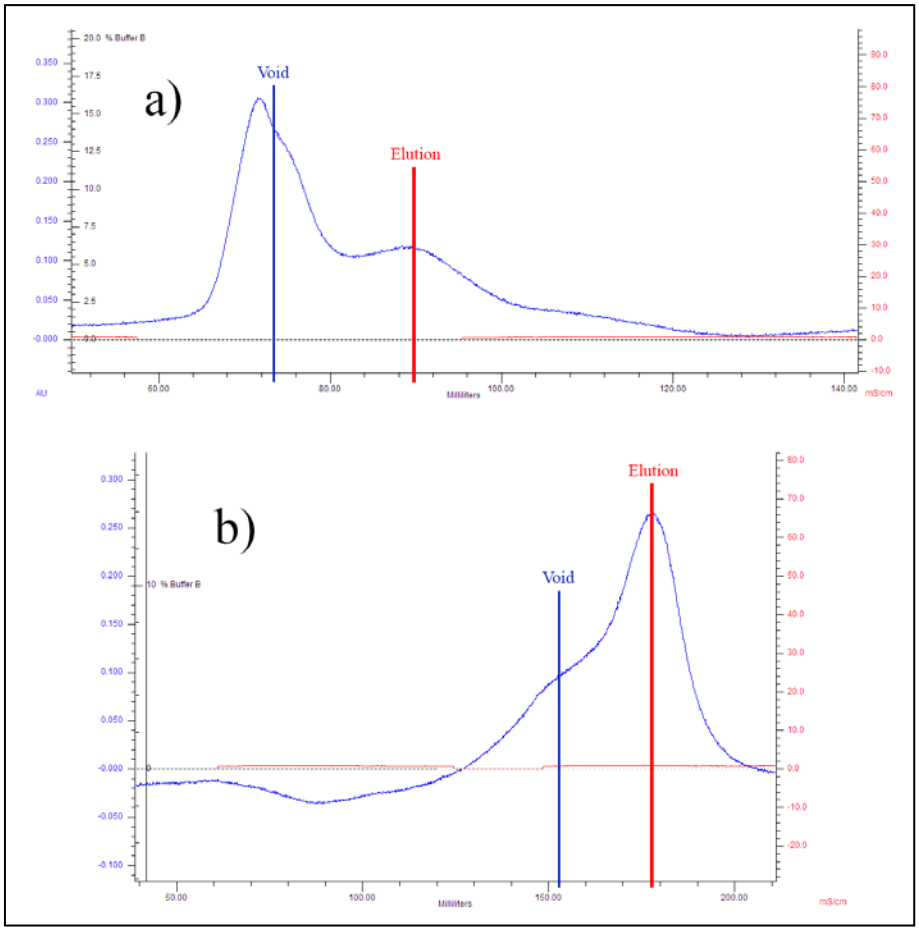

Figure 4.4.5 – S200 aggregation. This figure shows the effect of salt in the injected protein buffer on the relative intensities of the void and elution peaks. Both SEC runs were equilibrated in 25 mM HEPES, pH 7.5 and run at 0.5 mL/min. The different absolute void and elution volumes between the two runs were due to different bed volumes.

a) Protein was injected in buffer containing 0.15 M NaCl

b) Protein was injected with in buffer containing no salt, which was obtained by using prolonged dialysis to remove residual salt

4.5 – Finalized SUMO-HI-IgAP E. coli Protein Production Protocol

My second-generation denatured HI-IgAP workflow is summarized below. A typical expression and purification yields 2 mg HI-IgAP per liter of *E. coli* culture with a purity greater than 99%, as determined by eye (*Fig. 4.5.1*). All of the buffers below refer to the optimized buffers in *Table 4.4, column B*. Steps were performed at 4°C, unless otherwise specified.

1. Grow an overnight SUMO-HI-IgAP BL21(DE3) *E. coli* starter at 37°C from a glycerol stock. All media were supplemented with 50 μg/mL ampicillin.
2. Inoculate 50 mL starter into each liter of LB media.
3. Grow LB at 37°C until an OD_{600} of 0.6-0.9 and induce at 37°C with 1 mM IPTG for three hours. A typical purification involves using cells from 6-8 L *E. coli* culture.
4. Harvest the cells by centrifugation at 6,000xg for 10 minutes.
5. Resuspend the pellet in 100 mL Resuspension Buffer until mix until homogenous.
6. Lyse the cells via two passes through a FRENCH pressure cell (Thermo) at 1100 psi.
7. Centrifuge the lysate for one hour at 12,000xg, discard the supernatant, and resuspend the lysate pellet in 100 mL Resuspension Buffer and mix until homogenous.
8. Repeat Step #6.
9. Centrifuge the second lysate for one hour at 12,000xg and discard the supernatant.
10. The pellet was vigorously sonicated on ice in 50 mL Denaturing Buffer using a Qsonica Q500 sonicator and 1/2”-diameter tip (Qsonica) at 60% maximum amplitude with a 20/40 sec on/off cycle. The total sonication time was 10 minutes.
11. The sonicated mixture was loaded onto a Bio-Rad DynaLoop 90 injection loop pre-equilibrated in Denaturing Buffer. The dynamic loop was then hooked up in parallel to a Biologic DuoFlow FPLC (Bio-Rad) and was injected onto 20 mL Ni-IMAC resin (NiNTA; Molecular Cloning Lab) for denatured purification. The resin was extensively washed with 1 L Wash Buffer followed by 1 L Denaturing Buffer to remove residual detergent. SUMO-HI-IgAP was eluted isocratically with Elution Buffer until A_{280} readings were below 0.05.
12. Eluent was diluted three-fold using Denaturing Buffer and dialyzed overnight in 14 kDa MWCO dialysis tubes (Aldon Corporation) against 10-fold-excess Arginine Buffer. The protein was further dialyzed overnight into 10-fold-excess Refold A Buffer.

13. The dialyzed protein solution was incubated with homemade ULP1 at room temperature overnight for SUMO cleavage on a bench-top rocker. 3 mg ULP1 was added at the start of cleavage and dosed in every three hours, with a minimum of one extra dose.
14. Precipitate was removed by centrifugation at 12,000xg for 15 minutes.
15. Cleaved HI-IgAP was dialyzed against a 10-fold excess of Refold B Buffer overnight.
16. Protein was incubated with 100 mL Ni-IMAC resin equilibrated in Refold B Buffer for an hour. The resin was rinsed with 500 mL Refold B Buffer.
17. The flow through and rinse solutions were concentrated to less than 10 mL using a 30 kDa MWCO stirred-cell concentrator (Amicon).
18. The protein was extensively buffer exchanged into Storage Buffer using the stirred-cell concentrator with repeated dilutions and concentrations, overnight dialysis, or a desalting column (Bio-Gel P-6DG; Bio-Rad).
19. The protein was concentrated to less than 5 mL using a 30 kDa MWCO centrifugal concentrator (Amicon) and injected onto a manually packed S200 HR SEC column (GE) equilibrated in Storage Buffer.
20. Aggregates were removed by discarding fractions within the void-volume peak. The limits of this aggregate peak were liberally estimated by inspecting the FPLC A_{280} trace.
21. Fractions containing properly folded HI-IgAP were concentrated using a 30 kDa centrifugal concentrator to less than 1 mL.
22. Protein was flash frozen in aliquots using liquid nitrogen.

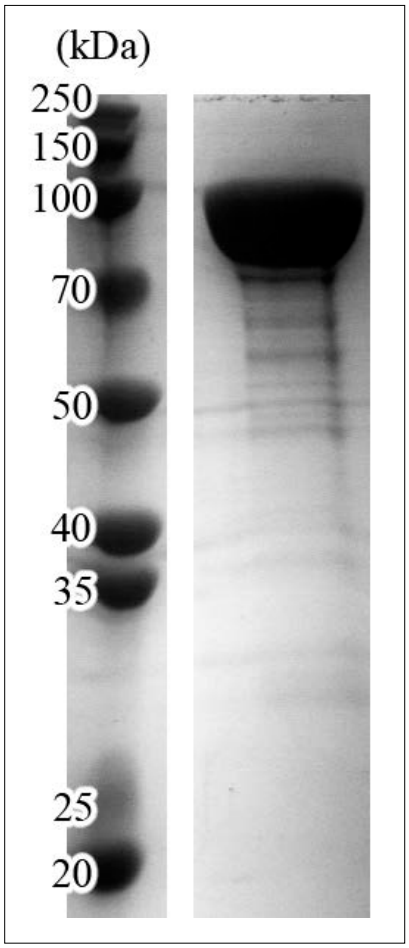

Figure 4.5.1 – Final HI-IgAP purity.

4.6 – Protein Validation

Even though the HI-IgAP purified from the second-generation denatured *E. coli* system was extremely pure and non-aggregated, further verification was needed to convince us that the protein was properly folded and biochemically identical to the enzyme purified from *H. influenzae*. The following four techniques confirmed that this was the first system to produce viable and recombinant HI-IgAP in sufficient biochemical amounts.

a) *Circular Dichroism*

A circular dichroism (CD) spectrum of HI-IgAP was taken using a JASCO 815 CD Spectrometer. A 0.6 mg/mL solution of HI-IgAP buffered in 25 mM HEPES, pH 7.5 was scanned from 200 to 250 nm at 50 nm/min. The resulting spectrum is shown in *Fig. 4.6.1*.

The CD spectrum was submitted to the BeStSel server to calculate theoretical magnitudes of each secondary structure component.[61] The fitted parameters indicate a strong parallel beta-sheet signal (*Table 4.6*), which suggests a properly folded C-terminal beta helix based on the known crystal structure of the enzyme (*Fig. 1.7.1*).

Table 4.6 – Theoretical model parameters for the HI-IgAP CD spectrum. The BeStSel server was used to generate a model based on differently weighted secondary structure elements.[61] A scale factor of 2 was used for this model.

Secondary Structure Element	Percentage of Element (%)
Helix	6.8
Antiparallel	9.8
Parallel	34.2
Turn	0.0
Others	49.1

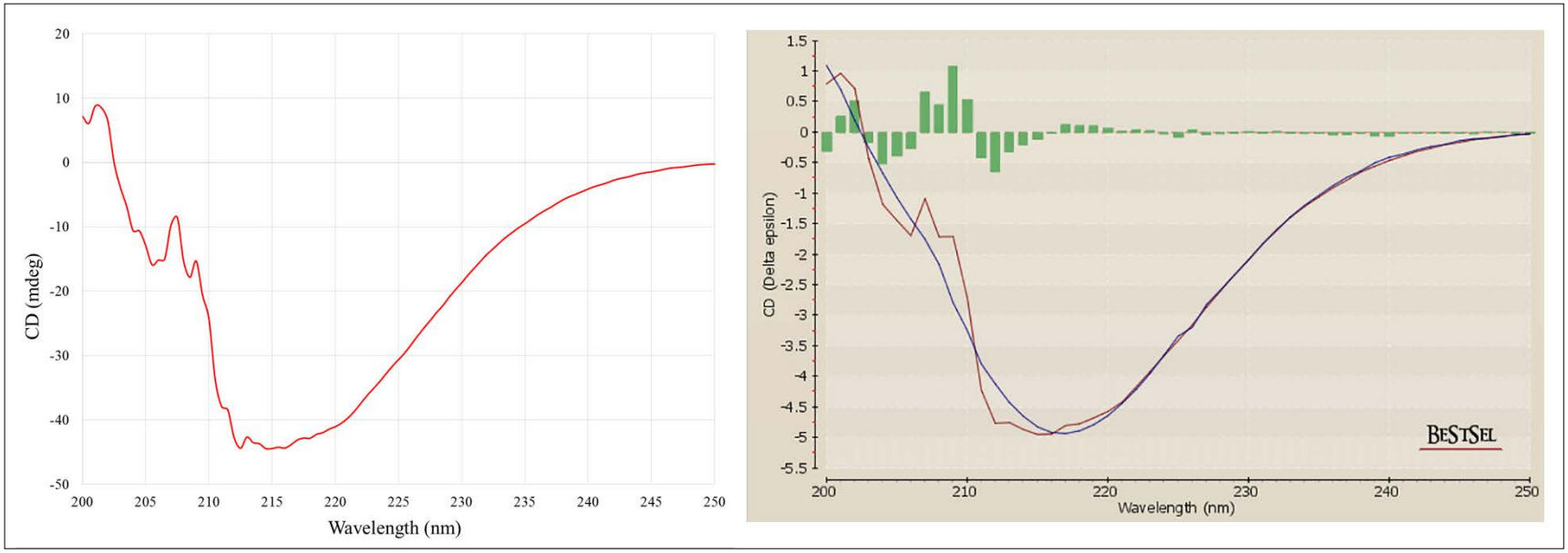

Figure 4.6.1 – CD spectrum of HI-IgAP. The experimental CD spectrum (left) shows a strong beta-sheet signal. The experimental data (red; right) were fit to a theoretical model (blue; right) based on weighted secondary structure elements. The fitting was done using the BeStSel server.[61] Model residuals are shown in green and its parameters are summarized in *Table 4.6*.

b) *Analytical Size-exclusion Chromatography*

Analytical SEC was used to assess the ability of recombinant HI-IgAP to bind to IgA1. Recombinant inactive HI-IgAP and IgA1 (Athens Research) were injected onto an ENrich SEC650 (Bio-Rad) column using a Biologic DuoFlow FPLC (Bio-Rad). The IgA1 peak of the inactive enzyme-substrate mixture eluted earlier than that of the free IgA1 peak (*Fig. 4.6.2*). As the concentrations of HI-IgAP and IgA1 were identical for all injections, the increase in absorbance for the IgA1 peak of the mixture also suggested that recombinant HI-IgAP bound IgA1.

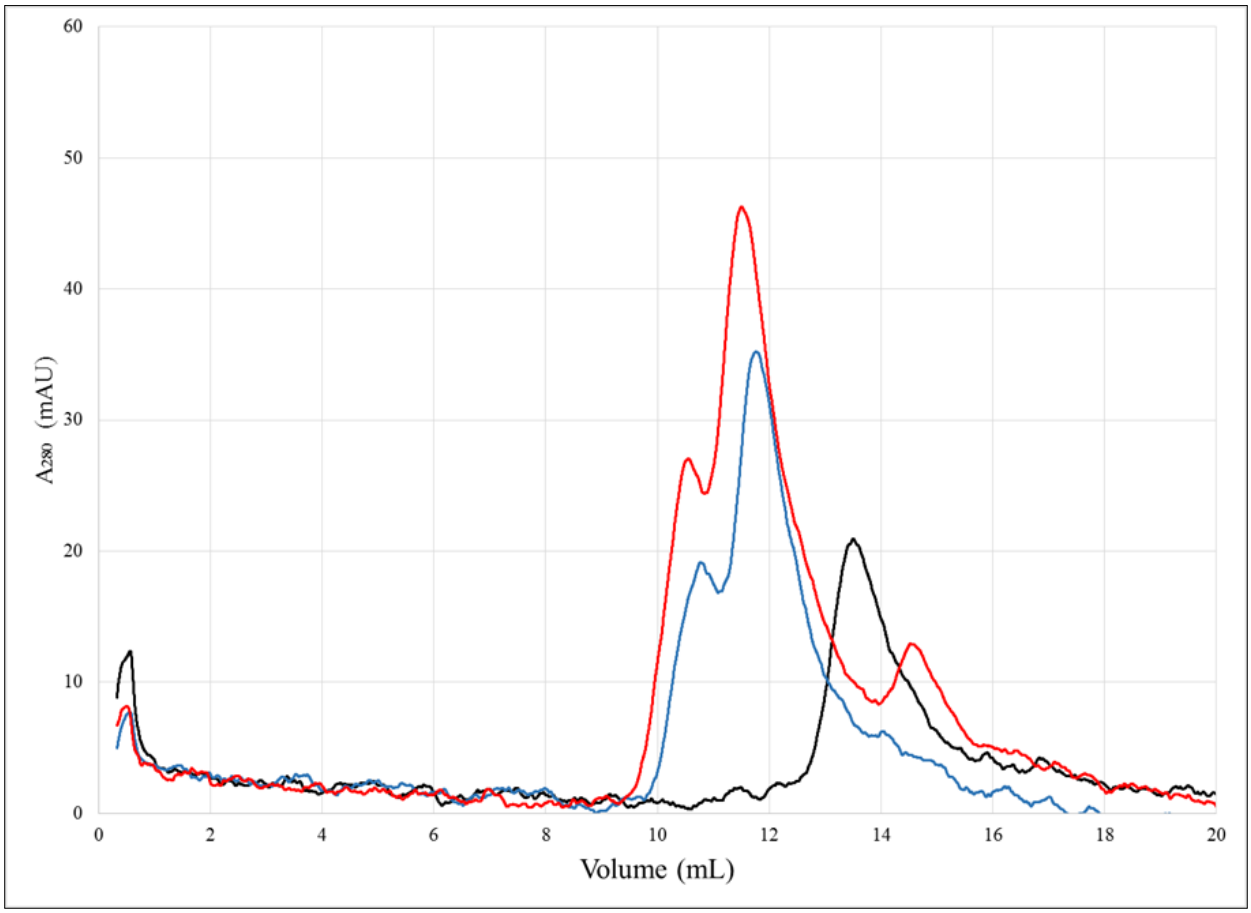

Figure 4.6.2 – Analytical SEC traces of the HI-IgAP-IgA1 interaction. 100 µL of 10 µM HI-IgAP (black) and 10 µM IgA1 (blue) were injected separately and at a 1:1 molar ratio (red). The SEC used 25 mM HEPES, pH 7.5 for its running buffer.

c) *X-ray Crystallography*

The ultimate test for confirming the validity of the protein-production system is to solve the HI-IgAP crystal structure using recombinant protein. Crystal drops were set up as previously described.[20] The largest crystals formed from drops with HI-IgAP concentrations greater than 7 mg/mL and PEG 20,000 concentrations less than 10% (w/v) (*Fig. 4.6.3a*). Small crystals were visible after two days of incubation. They were then left to grow for two weeks prior to diffraction. HI-IgAP crystals were cryo-cooled as previously described.[20] Diffraction images were collected using a Rigaku rotating copper anode X-ray generator and an R-axis IV++ detector with a ten-minute exposure time. Although they did not diffract very well (*Fig. 4.6.3b*), low-intensity and low-resolution diffraction peaks were visible. Synchrotron radiation and further optimization of the crystal conditions will likely allow us to collect a reasonable dataset to solve the crystal structure of the recombinant HI-IgAP. The fact that crystals with weak diffraction were successfully replicated strongly suggests the production of properly folded HI-IgAP.

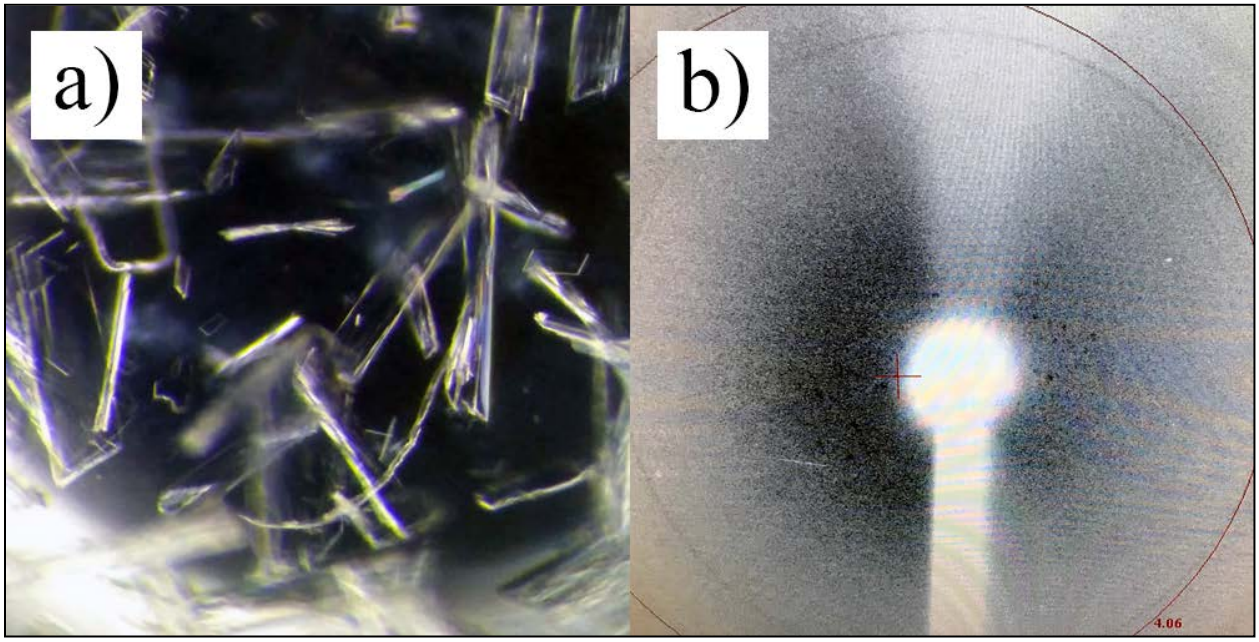

Figure 4.6.3 – HI-IgAP crystals. This figure shows the diffraction pattern (b) of a HI-IgAP crystal grown (a) under previously discovered crystal conditions. Low-resolution diffraction peaks were only visible when the crystal was exposed to prolonged radiation (10 minutes).

d) Enzymatic Activity

The activity of the SUMO-HI-IgAP construct was recently reconstituted by reverting the active-site point mutant back to the wild-type sequence (*Appendices C* and *D*). This was done with the services of GenScript USA. The active enzyme was then expressed and purified identically to the inactive mutant (Section 4.5) and showed activity following incubation with IgA1 (*Fig. 4.6.4*).

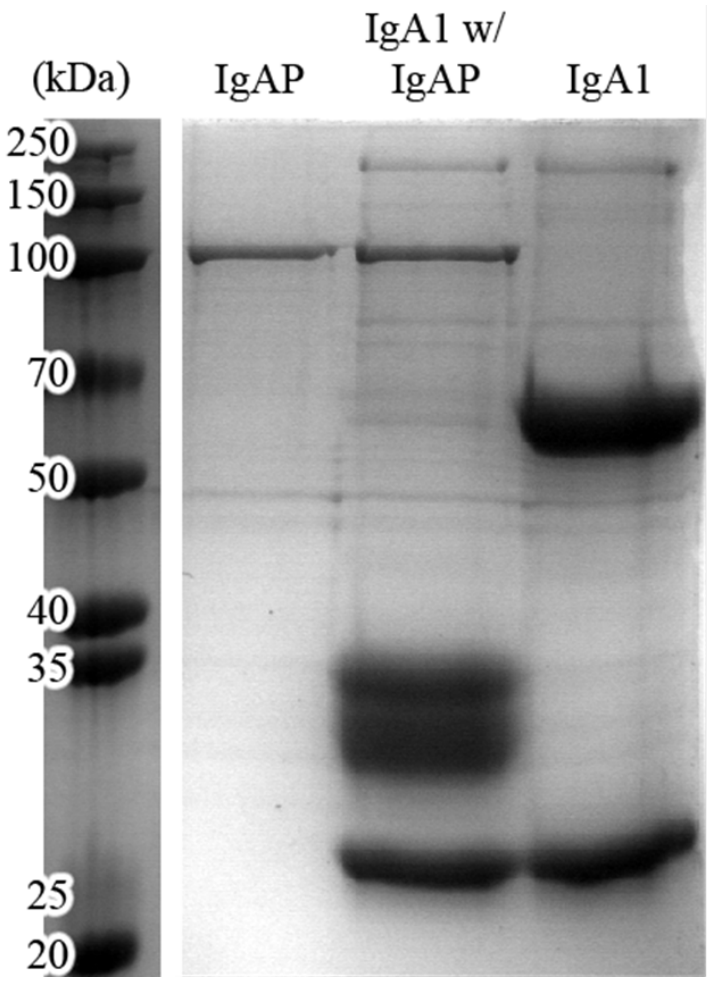

Figure 4.6.4 – SUMO-HI-IgAP activity. This figure shows the activity of recombinantly expressed SUMO-HI-IgAP. Recombinant IgAP was incubated with IgA1 for 30 minutes at 37°C.

4.7 – Conclusions and Next Steps

My second-generation denatured *E. coli* system successfully produced recombinant HI-IgAP, a task that has been unsuccessful ever since the crystal structure was solved a decade ago. The next step is to optimize the current HI-IgAP crystals to examine the binding sites of existing inhibitors by co-crystallization or crystal soaking. Improvements to existing small-molecule inhibitors can then be made using structure-based drug design strategies. Furthermore, the kinetic and thermodynamic parameters of the HI-IgAP can now be quantified and compared to other IgAPs and well-characterized proteases.

Section 5 – Thesis Summary

Both HI-IgAP *P. pastoris* and *E. coli* protein production systems were able to successfully produce recombinant HI-IgAP. It is clear, however, that the second-generation denaturing *E. coli* system is vastly superior to the *P. pastoris* system, at least under laboratory-scale conditions. Given that the *P. pastoris* system still may be improved by reconsidering alternative transformation and expression methods (Section 3.6), particularly the use of a bioreactor, the most-suitable system for large-scale protein production has yet to be determined. This is assuming the enzyme can be redesigned as an IgAN therapeutic and scaling up becomes a necessity.

This strategy of designing an expression construct with an N-terminal SUMO fusion with the intent performing denatured purification lends itself well to other problematic proteins that require native N-termini to properly fold. In particular, this strategy can be used to recombinantly produce the structurally homologous *N. gonorrhoeae* and *N. meningitidis* serine IgAPs, which also have the same N-terminal problem.

With this new enzyme source, we can now begin to study the biochemical intricacies of the enzyme by interrogating the important questions surrounding HI-IgAP-IgA1 interactions and IgA1 selectivity, which will lay the foundation for the design of therapeutics for LRTIs and IgAN.

References

1. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, Fullman N, Thompson RL, Abajobir A, Ahmed M, Alemayohu MA, Alvis-Guzman N, Amare AT, Antonio CA, Asayesh H, Avokpaho E, Awasthi A, Bacha U, Barac A, et al. Estimates of global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis*. 2018;18(11):1191-1210. doi:10.1016/ S1473-3099(18)30310-4

2. Brooks LRK, and Mias GI. Streptococcus pneumoniae's virulence and host immunity: Aging, diagnostics, and prevention. *Front Immunol*. 2018;9(1366). doi:10.3389/fimmu.2018.01366

3. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers C, and Black RE. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet*. 2016;388(10063):3027-3035. doi:10.1016/S0140-6736(16)31593-8

4. Wong JL, and Evans SE. Bacterial pneumonia in cancer patients: novel risk factors and current management. *Clin Chest Med*. 2017;38(2):263-277. doi:10.1016/j.ccm.2016.12.005

5. Carroll KC, and Adams LL. Lower Respiratory Tract Infections. *Microbiol Spectr*. 2016;4(4). doi:doi:10.1128/microbiolspec.DMIH2-0029-2016

6. Örtqvist Å. Treatment of community-acquired lower respiratory tract infections in adults. *Eur Respir J*. 2002;20(Suppl. 36):40s-53s. doi:10.1183/09031936.02.00309002

7. World Health Organization. Global antimicrobial resistance surveillance system (GLASS) report: early implementation 2016-2017. *World Heal Organ*. Published online 2017. https://apps.who.int/iris/handle/10665/259744

8. World Health Organization. Prioritization of Pathogens to Guide Discovery, Research and Development of New Antibiotics for Drug-Resistant Bacterial Infections, Including Tuberculosis. *World Heal Organ*. Published online 2017. https://www.who.int/medicines/areas/rational_use/prioritization-of-pathogens/en/

9. Mistry D, and Stockley RA. IgA1 protease. *Int J Biochem Cell Biol*. 2006;38(8):1244-1248. doi:10.1016/j.biocel.2005.10.005

10. Batten MR, Senior BW, Kilian M, and Woof JM. Amino Acid Sequence Requirements in the Hinge of Human Immunoglobulin A1 (IgA1) for Cleavage by Streptococcal IgA1 Proteases. *Infect Immun*. 2003;71(3):1462-1469. doi:10.1128/IAI.71.3.1462-1469.2003

11. Mansa B, and Kilian M. Retained Antigen-Binding Activity of Fab(α) Fragments of Human Monoclonal Immunoglobulin A1 (IgA1) Cleaved by IgA1 Protease. *Infect Immun*. 1986;52(1):171-174. doi:10.1128/IAI.52.1.171-174.1986

12. Toraño A, and Putnam FW. Complete amino acid sequence of the α2 heavy chain of a human IgA2 immunoglobulin of the A2m(2) allotype. *Proc Natl Acad Sci U S A*. 1978;75(2):966-969. doi:10.1073/pnas.75.2.966

13. Mulks MH, Kornfeld SJ, and Plaut AG. Specific Proteolysis of Human IgA by Streptococcus pneumoniae and Haemophilus influenzae. *J Infect Dis*. 1980;141(4):450-456. doi:10.1093/infdis/141.4.450

14. Kilian M, Mestecky J, and Schrohenloher RE. Pathogenic Species of the Genus Haemophilus and Streptococcus pneumoniae Produce Immunoglobulin A1 Protease. *Infect Immun*. 1979;26(1):143-149. doi:10.1128/IAI.26.1.143-149.1979

15. Mulks MH, Kornfeld SJ, Frangione B, and Plaut AG. Relationship between the Specificity of IgA Proteases and Serotypes in Haemophilus influenzae. 1982;146(2):266-274. doi:10.1093/infdis/146.2.266

16. Bachovchin WW, Plaut AG, Flentke GR, Lynch M, and Kettner CA. Inhibition of IgA1 proteinases from Neisseria gonorrhoeae and Hemophilus influenzae by peptide prolyl boronic acids. *J Biol Chem*. 1990;265(7):3738-3743.

17. Mortensen SB, and Kilian M. Purification and Characterization of an Immunoglobulin A1 Protease from Bacteroides melaninogenicus. *Infect Immun*. 1984;45(3):550-557. doi:10.1128/iai.45.3.550-557.1984

18. Kilian M, Mestecky J, Kulhavy R, Tomana M, and Butler WT. IgA1 proteases from Haemophilus influenzae, Streptococcus pneumoniae, Neisseria meningitidis, and Streptococcus sanguis: comparative immunochemical studies. *J Immunol*. 1980;124(6):2596-2600.

19. Chi Y, Rahkola JT, Kendrick AA, Holliday MJ, Paukovich N, Roberts TS, Janoff EN, and Eisenmesser EZ. Streptococcus pneumoniae IgA1 protease: A metalloprotease that can catalyze in a split manner in vitro. *Protein Sci*. 2017;26:600-610. doi:10.1002/pro.3110

20. Johnson TA, Qiu J, Plaut AG, and Holyoak T. Active site gating regulates substrate selectivity in a chymotrypsin-like serine protease. The structure of Haemophilus influenzae IgA1 protease. *J Mol Biol*. 2009;389(3):559-574. doi:10.1016/j.jmb.2009.04.041

21. Burton J, Wood SG, Lynch M, and Plaut AG. Substrate Analogue Inhibitors of the IgA1 Proteinases from Neisseria Gonorrhoeae. *J Med Chem*. 1988;31(8):1647-1651. doi:10.1021/jm00403a027

22. Wood SG, Lynch M, Plaut AG, and Burton J. Tetrapeptide Inhibitors of the IgAl Proteinases from Type I Neisseria gonorrhoeae. *J Med Chem*. 1989;32(10):2407-2411. doi:10.1021/jm00130a030

23. Garner AL, Fullagar JL, Day JA, Cohen SM, and Janda KD. Development of a High-Throughput Screen and Its Use In the Discovery of Streptococcus pneumoniae Immunoglobulin A1 Protease (IgA1P) Inhibitors. *J Am Chem Soc*. 2013;135(27). doi:10.1038/jid.2014.371

24. Choudary SK, Qiu J, Plaut AG, and Kritzer JA. Versatile Substrates and Probes for IgA1 Protease Activity. *ChemBioChem*. 2013;14:2007-2012. doi:10.1002/cbic.201300281

25. Shehaj L, Choudary SK, Makwana KM, Gallo MC, Murphy TF, and Kritzer JA. Small-Molecule Inhibitors of Haemophilus influenzae IgA1 Protease. *ACS Infect Dis*. 2019;5:1129-1138. doi:10.1021/acsinfecdis.9b00004

26. Julian BA, Waldo FB, Rifai A, and Mestecky J. IgA Nephropathy, the Most Common Glomerulonephritis Worldwide. *Am J Med*. 1988;84:129-132. doi:10.1016/0002-9343(88)90019-8

27. Novak J, Barratt J, Julian BA, and Renfrow MB. Aberrant Glycosylation of the IgA1 Molecule in IgA Nephropathy. *Semin Nephrol*. 2018;38(5):461-476. doi:10.1016/j.semnephrol.2018.05.016

28. Lafayette RA, and Kelepouris E. Immunoglobulin A Nephropathy: Advances in Understanding of Pathogenesis and Treatment. *Am J Nephrol*. 2018;47(suppl 1):43-52. doi:10.1159/000481636

29. Xie LS, Huang J, Qin W, and Fan JM. Immunoglobulin A1 protease: A new therapeutic candidate for immunoglobulin A nephropathy. *Nephrology*. 2010;15:584-586. doi:10.1111/j.1440-1797.2010.01278.x

30. Lamm ME, Emancipator SN, Robinson JK, Yamashita M, Fujioka H, Qiu J, and Plaut AG. Microbial IgA Protease Removes IgA Immune Complexes from Mouse Glomeruli in vivo: Potential Therapy for IgA Nephropathy. *Am J Pathol*. 2008;172(1):31-36. doi:10.2353/ajpath.2008.070131

31. Wang L, Li X, Shen H, Mao N, Wang H, Cui L, Cheng Y, and Fan J. Bacterial IgA protease-mediated degradation of agIgA1 and agIgA1 immune complexes as a potential therapy for IgA Nephropathy. *Sci Rep*. 2016;6(October 2015):1-13. doi:10.1038/srep30964

32. Senior BW, and Woof JM. The Influences of Hinge Length and Composition on the Susceptibility of Human IgA to Cleavage by Diverse Bacterial IgA1 Proteases. *J Immunol*. 2005;174(12):7792-7799. doi:10.4049/jimmunol.174.12.7792

33. Senior BW, and Woof JM. Sites in the CH3 Domain of Human IgA1 That Influence Sensitivity to Bacterial IgA1 Proteases. *J Immunol*. 2006;177(6):3913-3919. doi:10.4049/jimmunol.177.6.3913

34. Senior BW, and Woof JM. Effect of Mutations in the Human Immunoglobulin A1 (IgA1) Hinge on Its Susceptibility to Cleavage by Diverse Bacterial IgA1 Proteases. *Infect Immun*. 2005;73(3):1515-1522. doi:10.1128/IAI.73.3.1515-1522.2005

35. Kintzing JR, Filsinger Interrante M, and Cochran JR. Emerging Strategies for Developing Next-Generation Protein Therapeutics for Cancer Treatment. *Trends Pharmacol Sci*. 2016;37(12):993-1008. doi:10.4172/2327-4360-c1-011

36. Woof JM, and Kerr MA. IgA function - variations on a theme. *Immunology*. 2004;113(2):175-177. doi:10.1111/j.1365-2567.2004.01958.x

37. Kumar N, Arthur CP, Ciferri C, and Matsumoto ML. Structure of the secretory immunoglobulin A core. *Science*. 2020;367(6481):1008-1014.

38. Wines BD, and Hogarth PM. IgA receptors in health and disease. *Tissue Antigens*. 2006;68:103-114. doi:10.1111/j.1399-0039.2006.00613.x

39. Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, and Ala'Aldeen D. Type V Protein Secretion Pathway: the Autotransporter Story. *Microbiol Mol Biol Rev*. 2004;68(4):692-744. doi:10.1128/mmbr.68.4.692-744.2004

40. Meng G, Spahich N, Kenjale R, Waksman G, and St Geme III JW. Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *EMBO J*. 2011;30(18):3864-3874. doi:10.1038/emboj.2011.279

41. Maldonado-Contreras A, Birtley JR, Boll E, Zhao Y, Mumy KL, Toscano J, Ayehunie S, Reinecker HC, Stern LJ, and McCormick BA. Shigella depends on SepA to destabilize the intestinal epithelial integrity via cofilin activation. *Gut Microbes*. 2017;8(6):544-560. doi:10.1080/19490976.2017.1339006

42. Otto BR, Sijbrandi R, Luirink J, Oudega B, Heddle JG, Mizutani K, Park SY, and Tame JRH. Crystal Structure of Hemoglobin Protease, a Heme Binding Autotransporter Protein from Pathogenic Escherichia coli. *J Biol Chem*. 2005;280(17):17339-17345. doi:10.1074/jbc.M412885200

43. Ben-Bassat A, Bauer K, Chang SY, Myambo K, and Boosman A. Processing of the initiation methionine from proteins: Properties of the Escherichia coli methionine aminopeptidase and its gene structure. *J Bacteriol*. 1987;169(2):751-757. doi:10.1128/jb.169.2.751-757.1987

44. Lotosky WR. Towards the Production of the Haemophilus influenzae IgA1 Protease in Escherichia coli by. *UWSpace*. Published online 2014. http://hdl.handle.net/10012/8986

45. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005;41(1):207-234. doi:10.1016/j.pep.2005.01.016

46. Juturu V, and Wu JC. Heterologous Protein Expression in Pichia pastoris: Latest Research Progress and Applications. *ChemBioChem*. 2018;19(1):7-21. doi:10.1002/cbic.201700460

47. Ahmad M, Hirz M, Pichler H, and Schwab H. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol*. 2014;98(12):5301-5317. doi:10.1007/s00253-014-5732-5

48. Du M, Battles MB, and Nett JH. A color-based stable multi-copy integrant selection system for Pichia pastoris using the attenuated ADE1 and ADE2 genes as auxotrophic markers. *Bioeng Bugs*. 2012;3(1):32-37. doi:10.4161/bbug.3.1.17936

49. Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, Urdea MS, Valenzuela P, and Barr PJ. α-Factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 1984;81(15):4642-4646. doi:10.1073/pnas.81.15.4642

50. Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, Jiang Z, Zhou Y, Wang Y, Xu A, Li S, Liu P, and Wu D. Enhanced Production of Recombinant Secretory Proteins in Pichia pastoris by Optimizing Kex2 P1' site. *PLoS One*. 2013;8(9):1-11. doi:10.1371/journal.pone.0075347

51. Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, and Lin-Cereghino GP. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *Biotechniques*. 2005;38(1):44-48. doi:10.2144/05381BM04

52. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, and Li Q. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci*. 2009;18(5):936-948. doi:10.1002/pro.102

53. Koontz L. *TCA Precipitation*. Vol 541.; 2014. doi:10.1016/B978-0-12-420119-4.00001-X

54. Higgins DR, and Cregg JM. *Pichia Protocols*. Vol 103.; 1998. doi:10.1007/978-1-62703-239-1_1

55. Grimsley GR, Scholtz JM, and Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci*. 2009;18(1):247-251. doi:10.1002/pro.19

56. Vogl T, Gebbie L, Palfreyman RW, and Speight R. Effect of Plasmid Design and Type of Integration Event on Recombinant Protein Expression in Pichia pastoris. *Appl Environ Microbiol*. 2018;84(6). doi:10.1128/AEM.02712-17

57. Rinas U, Garcia-Fruitós E, Corchero JL, Vázquez E, Seras-Franzoso J, and Villaverde A. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem Sci*. 2017;42(9):726-737. doi:10.1016/j.tibs.2017.01.005

58. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, and Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics*. 2004;5:75-86. doi:10.1023/B:JSFG.0000029237.70316.52

59. Singh SM, and Panda AK. Solubilzation and Refolding of Bacterial Inclusion Body Proteins. *J Biosci Bioeng*. 2005;99(4):303-310. doi:10.1263/jbb.99.303

60. Arakawa T, and Tsumoto K. The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochem Biophys Res Commun*. 2003;304:148-152. doi:10.1016/S0006-291X(03)00578-3

61. Micsonai A, Wien F, Bulyáki É, Kun J, Moussong É, Lee YH, Goto Y, Réfrégiers M, and Kardos J. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res*. 2018;46(W1):W315-W322. doi:10.1093/nar/gky497

Appendices

Appendix A – Codon-optimized P. pastoris HI-IgAP Gene Sequence

ATGAGATTTCCATCCATTTTTACCGCTGTCCTTTTTGCTGCCTCCTCCGCACTTGCTGCTCCTGTCAACACCACT
ACTGAAGACGAGACTGCTCAAATTCCAGCTGAGGCTGTTATTGGTTACTCTGATTTGGAAGGAGATTTTGATGTT
GCTGTTTTGCCTTTCTCTAATTCTACTAACAACGGTTTGTTGTTTATTAACACTACTATCGCTTCTATCGCTGCT
AAAGAAGAGGGTGTTTCTTTGGAAAAGAGAGCTTTGGTTAGAGATGATGTTGATTACCAAATTTTTAGAGATTTC
GCTGAGAACAAGGGTAGATTTTCTGTTGGTGCTACTAACGTTGAAGTTAGAGATAAGAACAACCATTCTTTGGGT
AACGTTTTGCCAAATGGTATTCCTATGATCGATTTCTCTGTTGTTGATGTTGATAAGAGAATCGCTACTTTGATT
AATCCACAATACGTTGTTGGTGTTAAGCATGTTTCTAACGGTGTTTCTGAATTGCACTTCGGTAATTTGAACGGT
AACATGAACAACGGTAACGCTAAGTCTCACAGAGATGTTTCTTCTGAAGAGAACAGATACTTCTCTGTTGAGAAG
AATGAATATCCTACTAAGTTGAACGGTAAAGCTGTTACTACTGAGGATCAAACTCAAAAGAGAAGAGAAGATTAC
TACATGCCAAGATTGGATAAGTTTGTTACTGAGGTTGCTCCTATTGAAGCTTCTACTGCTTCTTCTGATGCTGGT
ACTTACAACGATCAAAATAAGTACCCAGCTTTCGTTAGATTGGGTTCTGGTTCTCAATTCATCTACAAGAAAGGA
GATAACTACTCTTTGATCTTGAACAACCATGAGGTTGGTGGTAACAATTTGAAATTGGTTGGAGATGCTTACACT
TATGGTATTGCTGGTACTCCATACAAGGTTAACCATGAAAACAACGGTTTGATCGGTTTTGGTAACTCTAAAGAA
GAGCACTCTGATCCAAAGGGTATTTTGTCTCAAGATCCTTTGACTAACTACGCTGTTTTGGGAGATTCTGGTTCT
CCTTTGTTCGTTTATGATAGAGAGAAGGGTAAATGGTTGTTTTTGGGTTCTTACGATTTCTGGGCTGGTTATAAT
AAGAAATCTTGGCAAGAGTGGAACATCTACAAACCAGAATTTGCTAAGACTGTTTTGGATAAGGATACTGCTGGT
TCTTTGACTGGTTCTAACACTCAATATAACTGGAATCCTACTGGTAAAACTTCTGTTATTTCTAACGGTTCTGAA
TCTTTGAACGTTGATTTGTTCGATTCTTCTCAAGATACTGATTCTAAGAAAAACAACCATGGTAAATCTGTTACT
TTGAGAGGTTCTGGTACTTTGACTTTGAACAACAACATCGATCAAGGTGCTGGTGGTTTGTTTTTCGAGGGAGAT
TACGAAGTTAAAGGTACTTCTGATTCTACTACTTGGAAGGGTGCTGGTGTTTCTGTTGCTGATGGTAAAACTGTT
ACTTGGAAAGTTCACAATCCAAAGTCTGATAGATTGGCTAAGATTGGTAAAGGTACTTTGATTGTTGAGGGTAAA
GGTGAAAACAAAGGTTCTTTGAAGGTTGGAGATGGTACTGTTATTTTGAAGCAACAAGCTGATGCTAACAACAAG
GTTAAGGCTTTTTCTCAAGTTGGTATCGTTTCTGGTAGATCTACTGTTGTTTTGAACGATGATAAGCAAGTTGAT
CCTAACTCTATCTACTTTGGTTTCAGAGGTGGTAGATTGGATGCTAACGGTAACAACTTGACTTTCGAACATATC
AGAAACATCGATGATGGTGCTAGATTGGTTAACCACAACACTTCTAAGACTTCTACTGTTACTATCACTGGAGAG
TCTTTGATCACTGATCCAAACACTATCACTCCTTACAACATTGATGCTCCAGATGAAGATAACCCTTACGCTTTT
AGAAGAATTAAGGATGGTGGTCAATTGTACTTGAACTTGGAGAACTACACTTACTACGCTTTGAGAAAGGGTGCT
TCTACTAGATCTGAATTGCCAAAGAACTCTGGAGAGTCTAACGAAAATTGGTTGTACATGGGTAAAACTTCTGAT
GAGGCTAAGAGAAACGTTATGAACCATATCAACAACGAAAGAATGAACGGTTTTAATGGTTATTTCGGTGAAGAG
GAAGGTAAAAATAACGGTAACTTGAACGTTACTTTTAAGGGTAAATCTGAGCAAAACAGATTCTTGTTGACTGGT
GGTACTAACTTGAATGGAGATTTGAAGGTTGAAAAAGGTACTTTGTTCTTGTCTGGTAGACCAACTCCTCATGCT
AGAGATATTGCTGGTATCTCTTCTACTAAGAAAGATCAACACTTTGCTGAGAATAACGAAGTTGTTGTTGAAGAT
GATTGGATCAACAGAAACTTCAAGGCTACTAACATCAACGTTACTAACAACGCTACTTTGTACTCTGGTAGAAAC
GTTGCTAACATCACTTCTAACATCACTGCTTCTGATAACGCTAAAGTTCACATTGGTTATAAGGCTGGAGATACT
GTTTGTGTTAGATCTGATTACACTGGTTACGTTACTTGTACTACTGATAAGTTGTCTGATAAGGCTTTGAACTCT
TTTAATGCTACTAACGTTTCTGGTAACGTTAATTTGTCTGGTAACGCTAACTTCGTTTTGGGTAAAGCTAATTTG
TTCGGTACTATTTCTGGTACTGGTAATTCTCAAGTTAGATTGACTGAAAACTCTCATTGGCACTTGACTGGAGAT
TCTAACGTTAACCAATTGAACTTGGATAAGGGTCATATTCACTTGAATGCTCAAAACGATGCTAATAAGGTTACTA
CTTACAACACTTTGACTGTTAATTCTTTGTCTGGTAACGGTTCTTTCTACTATTTGACTGATTTGTCTAACAAGCAAG
GAGATAAGGTTGTTGTTACTAAATCTGCTACTGGTAACTTTACTTTGCAAGTTGCTGATAAAACTGGAGAGCCAAC
TAAGAATGAATTGACTTTGTTCGATGCTTCTAACGCTACTAGAAACAACTTGAACGTTTCTTTGGTTGGTAACACTG
TTGATTTGGGTGCTTGGAAGTACAAGTTGAGAAACGTTAACGGTAGATACGATTTGTACAATCCTGAGGTTGAAAA
GAGAAACCAGACCGTTGACACTACTAACATCACTACTCCTCTTGAGCATCATCATCACCATCACTAA

Appendix B – Primary Sequence of the P. pastoris HI-IgAP Protein Fusion

The α-MF, HI-IgAP, and His tag sequences are colored in red, blue, and purple, respectively.

MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSL
EKRALVRDDVDYQIFRDFAENKGRFSVGATNVEVRDKNNHSLGNVLPNGIPMIDFSVVDVDKRIATLINPQYVVGVKH
VSNGVSELHFGNLNGNMNNGNAKSHRDVSSEENRYFSVEKNEYPTKLNGKAVTTEDQTQKRREDYYMPRLDKFVTEV
APIEASTASSDAGTYNDQNKYPAFVRLGSGSQFIYKKGDNYSLILNNHEVGGNNLKLVGDAYTYGIAGTPYKVNHENN
GLIGFGNSKEEHSDPKGILSQDPLTNYAVLGDSGSPLFVYDREKGKWLFLGSYDFWAGYNKKSWQEWNIYKPEFAKTV
LDKDTAGSLTGSNTQYNWNPTGKTSVISNGSESLNVDLFDSSQDTDSKKNNHGKSVTLRGSGTLTLNNNIDQGAGGLFF
EGDYEVKGTSDSTTWKGAGVSVADGKTVTWKVHNPKSDRLAKIGKGTLIVEGKGENKGSLKVGDGTVILKQQADAN
NKVKAFSQVGIVSGRSTVVLNDDKQVDPNSIYFGFRGGRLDANGNNLTFEHIRNIDDGARLVNHNTSKTSTVTITGESLI
TDPNTITPYNIDAPDEDNPYAFRRIKDGGQLYLNLENYTYYALRKGASTRSELPKNSGESNENWLYMGKTSDEAKRNV
MNHINNERMNGFNGYFGEEEGKNNGNLNVTFKGKSEQNRFLLTGGTNLNGDLKVEKGTLFLSGRPTPHARDIAGISST
KKDQHFAENNEVVVEDDWINRNFKATNINVTNNATLYSGRNVANITSNITASDNAKVHIGYKAGDTVCVRSDYTGYV
TCTTDKLSDKALNSFNATNVSGNVNLSGNANFVLGKANLFGTISGTGNSQVRLTENSHWHLTGDSNVNQLNLDKGHIH
LNAQNDANKVTTYNTLTVNSLSGNGSFYYLTDLSNKQGDKVVVTKSATGNFTLQVADKTGEPTKNELTLFDASNATR
NNLNVSLVGNTVDLGAWKYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTPLEHHHHHH*

Appendix C – Gene Sequence of the SUMO-HI-IgAP Construct

The coding sequence for the inactive (A288) SUMO-HI-IgAP construct is shown below. This was the construct used for most of the experiments detailed in this thesis, with the exception of the activity assay in Section 4.6. The sequence of the active enzyme (S288) is the same as the one shown below but the codon in brackets is replaced by "AGT".

ATGGGTCATCACCATCATCATCACGGGTCGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAA
GTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCA
CTCCTTTGCGTAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACGA
CGGTATTCGCATTCAAGCTGATCAGGCCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACCGC
GAACAGATTGGAGGTGCCTTAGTGAGAGACGATGTGGATTATCAAATATTTCGTGATTTTGCAGAAAATAAAGGG
AGATTTTCTGTTGGTGCAACAAATGTGGAAGTGAGAGATAAAAATAACCACTCTTTAGGCAATGTTTTACCTAATG
GCATTCCGATGATTGATTTTAGTGTTGTGGATGTAGATAAACGCATCGCCACATTGATAAATCCACAATATGTAGT
AGGTGTAAAACACGTTAGTAACGGCGTGAGTGAACTACATTTTGGGAACTTAAATGGCAATATGAATAATGGCAA
TGCTAAATCGCACCGAGATGTATCTTCAGAAGAAAATAGATATTTTTCCGTTGAGAAAAATGAGTATCCAACTAAA
TTGAATGGAAAAGCAGTAACTACTGAAGATCAAACTCAAAAACGCCGTGAAGACTACTATATGCCACGTCTTGAT
AAATTTGTTACCGAAGTTGCACCAATAGAGGCTTCAACTGCAAGTAGTGATGCTGGCACATATAATGATCAGAATA
AATATCCTGCTTTTGTAAGACTAGGAAGTGGTAGTCAATTTATTTATAAAAAAGGAGATAATTACAGCTTAATTTT
AAATAATCATGAGGTTGGAGGCAATAATCTTAAATTGGTGGGCGATGCCTATACCTATGGTATTGCAGGCACACCT
TATAAAGTAAACCACGAAAATAATGGACTAATTGGTTTTGGCAATTCAAAAGAGGAACACAGCGATCCAAAAGGA
ATATTATCTCAAGATCCGCTTACCAATTATGCTGTTTTAGGCGAC**(GCT)**GGCTCCCCATTATTTGTATATGATAGAG
AAAAAGGAAAATGGCTTTTTCTTGGGTCTTATGATTTTTGGGCAGGTTATAACAAAAAATCTTGGCAAGAATGGAA
TATTTATAAACCTGAATTTGCAAAAACTGTTCTAGATAAAGATACTGCAGGTTCTTTAACTGGTTCTAACACCCAAT
ACAATTGGAATCCTACTGGCAAAACAAGCGTTATTTCTAATGGTTCTGAATCTCTAAATGTTGATTTATTCGATAGT
AGTCAGGATACTGACTCTAAGAAGAACAATCACGGAAAAAGTGTGACTCTTAGAGGAAGTGGAACGCTTACCTTA
AATAATAATATCGATCAAGGCGCAGGCGGCTTGTTCTTTGAAGGAGATTATGAAGTTAAAGGCACTTCTGATAGTA
CCACTTGGAAAGGAGCTGGCGTTTCTGTTGCTGATGGAAAAACAGTAACGTGGAAAGTACATAACCCGAAATCTG
ATCGTTTAGCTAAAATCGGCAAAGGAACATTAATTGTAGAAGGAAAGGGAGAAAATAAAGGTTCGCTAAAAGTGG
GCGATGGTACTGTTATCTTAAAACAACAAGCTGATGCCAATAATAAAGTTAAAGCCTTTTCACAAGTAGGTATAGT
AAGTGGTCGCTCAACTGTTGTACTTAATGATGATAAGCAAGTAGATCCAAATTCCATTTACTTTGGCTTTAGAGGT
GGTCGATTAGATGCCAATGGCAATAATCTCACTTTTGAACATATCCGTAATATTGATGATGGCGCAAGACTAGTAA
ATCACAATACCAGCAAAACCTCTACTGTAACAATTACTGGGGAAAGTCTAATTACAGATCCAAATACAATTACTCC
ATATAATATAGACGCACCAGATGAAGATAATCCTTATGCCTTTCGACGGATTAAAGATGGAGGACAGCTCTATTTA
AATTTGGAAAATTACACTTATTATGCGTTAAGAAAAGGTGCGAGCACTCGTTCAGAATTACCTAAAAATAGTGGCG
AAAGCAATGAAAATTGGCTATATATGGGTAAAACTTCCGATGAAGCCAAAAGAAATGTAATGAACCATATCAACA
ACGAGCGTATGAATGGCTTTAACGGTTATTTTGGCGAGGAAGAGGGTAAAAATAACGGTAATCTAAATGTGACTTT
TAAAGGCAAAAGTGAGCAAAATCGCTTTTTATTAACAGGCGGAACAAACCTTAATGGCGATTTAAAGGTTGAAAA
AGGCACATTATTCCTTTCTGGCAGACCAACACCGCACGCAAGAGATATTGCAGGTATTTCTTCGACAAAAAAAGAT
CAACACTTTGCTGAAAATAATGAAGTGGTAGTAGAAGATGACTGGATTAACCGCAATTTTAAAGCAACAAATATT
AATGTAACCAATAACGCAACCCTTTATTCAGGTCGCAATGTTGCAAACATTACTTCAAATATCACAGCTTCTGATA
ATGCAAAAGTACATATTGGCTATAAAGCAGGCGATACCGTTTGTGTACGTTCTGACTATACGGGCTATGTGACTTG
CACTACTGACAAGTTATCCGATAAAGCCCTTAATAGCTTTAACGCCACCAATGTATCTGGCAATGTAAATTTATCA
GGTAATGCAAACTTTGTCTTAGGCAAAGCTAACTTATTCGGCACAATTAGCGGCACGGGAAATAGCCAAGTACGTT
TAACCGAAAATAGCCATTGGCATTTAACAGGCGATAGCAATGTTAATCAGTTAAATTTAGACAAGGGGCATATTCA
TTTAAATGCACAAAACGATGCAAATAAAGTAACTACATATAACACGCTGACTGTGAATAGCTTATCAGGTAACGG
TTCTTTCTATTATTTAACTGATCTTTCCAATAAACAAGGCGACAAAGTTGTTGTAACTAAATCCGCCACAGGTAACT
TTACATTACAAGTGGCAGATAAAACAGGCGAGCCTACAAAAAATGAACTCACGCTTTTTGATGCGTCAAATGCTAC
AAGAAATAATTTGAATGTGTCATTAGTTGGGAATACCGTTGATTTAGGTGCTTGGAAATATAAATTACGTAATGTT
AATGGACGTTACGATTTGTATAACCCAGAGGTGGAAAAAAGAAATCAAACTGTCGATACGACAAATATCACAACA
CCTTAA

Appendix D – Primary Sequence of the SUMO-HI-IgAP Construct

The primary sequence for the inactive (A288) SUMO-HI-IgAP construct is shown below. This was the construct used for most of the experiments detailed in this thesis, with the exception of the activity assay in Section 4.6. The sequence of the active enzyme (S288) is the same as the one shown below but the residue in brackets is replaced by "S". The His tag, SUMO, and HI-IgAP sequences are shown in purple, red, and blue, respectively.

MGHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRI
QADQAPEDLDMEDNDIIEAHREQIGGALVRDDVDYQIFRDFAENKGRFSVGATNVEVRDKNNHSLGNVLPNGIPMIDFS
VVDVDKRIATLINPQYVVGVKHVSNGVSELHFGNLNGNMNNGNAKSHRDVSSEENRYFSVEKNEYPTKLNGKAVTTE
DQTQKRREDYYMPRLDKFVTEVAPIEASTASSDAGTYNDQNKYPAFVRLGSGSQFIYKKGDNYSLILNNHEVGGNNLK
LVGDAYTYGIAGTPYKVNHENNGLIGFGNSKEEHSDPKGILSQDPLTNYAVLGD**(A)**GSPLFVYDREKGKWLFLGSYDF
WAGYNKKSWQEWNIYKPEFAKTVLDKDTAGSLTGSNTQYNWNPTGKTSVISNGSESLNVDLFDSSQDTDSKKNNHGK
SVTLRGSGTLTLNNNIDQGAGGLFFEGDYEVKGTSDSTTWKGAGVSVADGKTVTWKVHNPKSDRLAKIGKGTLIVEGK
GENKGSLKVGDGTVILKQQADANNKVKAFSQVGIVSGRSTVVLNDDKQVDPNSIYFGFRGGRLDANGNNLTFEHIRNI
DDGARLVNHNTSKTSTVTITGESLITDPNTITPYNIDAPDEDNPYAFRRIKDGGQLYLNLENYTYYALRKGASTRSELPK
NSGESNENWLYMGKTSDEAKRNVMNHINNERMNGFNGYFGEEEGKNNGNLNVTFKGKSEQNRFLLTGGTNLNGDLK
VEKGTLFLSGRPTPHARDIAGISSTKKDQHFAENNEVVVEDDWINRNFKATNINVTNNATLYSGRNVANITSNITASDNA
KVHIGYKAGDTVCVRSDYTGYVTCTTDKLSDKALNSFNATNVSGNVNLSGNANFVLGKANLFGTISGTGNSQVRLTEN
SHWHLTGDSNVNQLNLDKGHIHLNAQNDANKVTTYNTLTVNSLSGNGSFYYLTDLSNKQGDKVVVTKSATGNFTLQV
ADKTGEPTKNELTLFDASNATRNNLNVSLVGNTVDLGAWKYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTP*

Appendix E – Expression and Purification ULP1 Protocol

The following describes the protocol for ULP1 production. All pH values were measured at room temperature. All steps were performed at 4°C unless other specified.

1. Grow an overnight ULP1 BL21(DE3) *E. coli* starter at 37°C from a glycerol stock.
2. Inoculate 50 mL starter into each liter of LB media.
3. Grow LB at 37°C until an OD_{600} of 0.6-0.9 and induce at 37°C with 1 mM IPTG for three hours.
4. Harvest the cells by centrifugation at 6,000xg for 10 minutes.
5. Resuspend the pellet in Resuspension Buffer until homogenous.
6. Lyse the cells via two passes through a FRENCH pressure cell (Thermo) at 1100 psi.
7. Centrifuge the lysate for one hour at 12,000xg and incubate the supernatant with Ni-IMAC resin (NiNTA; Molecular Cloning Lab) pre-equilibrated in Resuspension Buffer for 1 hour at 4°C.
8. Wash the resin with 1 L Wash Buffer then 1 L Resuspension Buffer.
9. Elute the resin with Elution Buffer until an A_{280} reading of less than 0.05.
10. The protein was concentrated to less than 5 mL using a 30 kDa MWCO centrifugal concentrator (Amicon) and loaded onto a Bio-Gel P-6DG desalting resin (Bio-Rad) pre-equilibrated in Storage Buffer.
11. Buffer-exchanged fractions were concentrated using a 30 kDa centrifugal concentrator to less than 1 mL and flash frozen in aliquots using liquid nitrogen.

Table E.1 – ULP1 purification buffers. HEPES, NaCl, and imidazole were purchased from BioBasic, TCEP from ChemImpex, and IgePal CA-630 from Sigma. All buffers have a pH of 7.5 at room temperature. All buffers were equilibrated to 4°C before use. TCEP was added to the buffer right before its use.

Buffer Name	Components
Resuspension	25 mM HEPES + 0.5 M NaCl + 10 mM imidazole + 2 mM TCEP
Wash	25 mM HEPES + 0.1% (v/v) IgePal CA-630 + 10 mM imidazole + 2 mM TCEP
Elution	25 mM HEPES + 0.5 M NaCl + 300 mM imidazole + 2 mM TCEP
Storage	25 mM HEPES + 0.5 M NaCl + 2 mM TCEP