Designing Protein Expression and Purification Systems for Recombinant Z Alpha1-Antitrypsin using the Methylotrophic Yeast, *Pichia pastoris*

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Monique Jenaie LeMieux
August 2012

DEDICATION

This thesis is dedicated to my mother, Janet LeMieux, who first introduced me to the world of science and to my father, Raymond Joseph LeMieux, who taught me to always go after my dreams.

ACKNOWLEDGEMENTS

First I would like to thank my advisor, Dr. Valerie Berthelier, for inviting me into her lab and allowing me to work on this project. Without her guidance and scientific knowledge, this manuscript would have not been possible.

I would also like to express gratitude to all of the other people who have helped train me and encourage me to become a researcher. I am especially grateful for Dr. Erica Rowe, who taught me all about the joys of molecular biology and who guided me during the initial steps of this project. The guidance, mentorship and training that she provided helped me to overcome many of the initial complications that I faced at that time.

I am grateful for Dr. Thomas Masi who helped me with all of the *Pichia pastoris* experiments. Without his training and knowledge about *Pichia pastoris*, many of the experiments described in this thesis would not have been possible. Also, I would like to express thanks to Dr. Jonathan Phipps who assisted me with the Site-Directed Mutagenesis experiments.

I would like to thank the Genome Science and Technology (GST) Program, the Program for Excellence & Equity in Research (PEER) and the Physicians' Medical Education and Research Foundation (PMERF) for providing financial support for my research and for myself during this time.

Finally I am thankful the most for my family and my friends who have supported me throughout this process. Without their love and faith in me, I would never have been able to finish this thesis.

ABSTRACT

It is well established that improper protein folding often leads to the formation of aggregates whose consequences are cellular impairment and cell death. One example of this is the aggregation of the mutant Z Alpha1-Antitrypsin protein, which results in blocking of its secretion due to inclusion body formation. This can contribute not only to the development of chronic obstructive pulmonary disease but also to hepatitis, cirrhosis and hepatocellular carcinoma. Current treatments are principally limited to intravenous Alpha1-Antitrypsin therapy and organ transplantation. In the scientific community though, it is widely thought that more effective forms of treatments lie within the polymerization process itself. However, in order to conduct the necessary experiments for this, copious amounts of Alpha1-Antitrypsin protein are needed. The amounts far out-weigh those that are obtainable through traditional methods of plasma purification. Therefore there is a high demand for methods that can successfully produce recombinant Alpha1-Antitrypsin protein. Over the years, several systems have been implemented to produce wild-type Alpha1-Antitrypsin protein but it was not until 2009 that recombinant Z Alpha1-Antitrypsin protein was successfully produced. The goal of this thesis was to expand upon this original study in order to produce both untagged and tagged versions of recombinant Z Alpha1-Antitrypsin protein using the methylotrophic yeast, *Pichia pastoris*. Colonies from the protease-deficient strand SMD1163 were selected for multi-copy integrations and then purified using affinity chromatography and anion exchange chromatography. While the results of the purification process proved inconclusive, they do help lay the groundwork for future endeavors.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1:

LITERATURE REVIEW

1.1 Serpins and Disease

a) Overview

With members identified in all major branches of life, Serpins are the most widely distributed family of protease inhibitors (1,2). The inhibitory activity of serpins relies on the structural flexibility of their Reactive Center Loop, which, in its native form, is not in the most stable conformation. In fact, the inhibitory native folded state is in a metastable conformation, making it vulnerable to mutations (1,2). These mutations, alone or in association with other factors, can upset the fragile balance within the serpin structure and promote self-association and tissue deposition. The result of these self-association and tissue depositions is a group of diseases referred to as the serpinopathies. Clinical manifestations of serpinopathies vary greatly, ranging from emphysema to thrombosis to dementia (1,2).

In this section we will go into more detail about what exactly serpins are, their structure, and common mutations that arise within them that lead to disease. This section will also briefly look at the cellular response that occurs as a result of serpin polymers.

b) What are Serpins?

Serine protease inhibitors (Serpins) are the largest family of protease inhibitors and are best known for their conformational flexibility and management of crucial regulatory functions in multicellular eukaryotes. In addition to eukaryotes, serpins have also been identified in bacteria, archaea and even some viruses. To date, over 1,500 members have been identified, 36 of which are found in humans (2). Proteins in the serpin superfamily have an average size of 350-400 amino acids and molecular weight of 40-50kDa (3). While considered to be mid-sized proteins, they are actually relatively large molecules when compared to other protease inhibitors such as basic pancreatic trypsin inhibitor (BPIT, which is about 60 amino acids) (2).

In 2000, a phylogenetic study of the superfamily was conducted, which divided eukaryotic serpins into 16 'clades' (termed A-P) (4). This resulted in a classification system where members are named SERPINXy where X is the clade and y is the number within the clade. Most serpins, however, also have alternative names from before the creation of the classification system. An example of this is SERPINA8 -- member 8 in Clade A --which also known as Angiotensinogen (2).

The name "Serpin" is actually quite misleading. In fact, not all members of this family are serine protease inhibitors and not all serine protease inhibitors are members (1,5). Although most serpins do inhibit serine proteases, some of them have been found to inhibit caspases and papain-like cysteine proteases, while others perform non-inhibitory functions. In fact, serpins with non-inhibitory functions perform vital roles in a broad range of biological capacities, including chromatin packing (the myeloid and erythroid nuclear termination stage-specific protein or MENT), blood pressure regulation

(angiotensinogen), protein folding (heat shock protein 47 or hsp47), tumor progression (mammary serine protease inhibitor or maspin) and hormone transport (cortisol- and thyroxine-binding globulin) (1,5). In addition, there is a subfamily of serpins whose inhibitory activity is increased after binding to heparin or other negatively charged polyanions such as heparin sulfate and dermatan sulphate. This subfamily includes heparin cofactor II, antithrombin, protease nexin 1 and plasminogen activator inhibitior-1 (PAI-1)(2,3). Table 1.1 gives additional examples of serpins and their associated functions.

Despite this diversity in function, the serpin superfamily is highly conserved. In fact, members share more than 30% amino acid sequence homology and a conserved tertiary structure with the archetype serpin, α_1-Antitrypsin (SERPINA1, AAT) (6). This structure typically consists of 9 α-helices, 3 ß-sheets (A, B, C) and an exposed mobile Reactive Center Loop (RCL) that stretches between ß-sheets A and C over a span of about 20 residues (Figure 1.1A). Crystal structures from different serpins have revealed five conformational states –native, cleaved, latent, δ, and polymeric—with differences arising primarily in the structure of the RCL (4). In the native conformation of inhibitory serpins, the RCL acts as a psuedosubstrate for the targeted protease that, upon binding at the scissile bond (P1-P1'), triggers a "mousetrap" mechanism. During this mechanism, the RCL is cleaved and integrated as an additional strand into the ß-sheet A (strand 4 or s4A). The protease, which has been covalently bonded with the RCL, is flipped from the upper pole to the lower pole of the serpin and rendered useless (Figure 1.1 B and C) (1-8).

Table 1.1 Functions and Dysfunctions of Human Serpins: Some of the 36 serpins found in humans along with their function and involvement, if any, in disease. Information was adapted from (2)

Serpin	Alternative name	Protease Target or Function	Involvement in Disease
SERPINA1	Alpha-1 Antitrypsin	Inflammatory response molecule; Neutrophil elastatse inhibition	Deficiency results in emphysema; Polymerization and retention in the ER results in cirrhosis
SERPINA3	Alpha -1 Antichymotrypsin	Inflammatory response molecule; Cathespin C inhibition	Deficiency results in emphysema
SERPINA5	Protein C Inhibitor (PAI-3)	Active Protein C inhibition	Angioedema
SERPINA6	Corticosteroid-Binding Globulin	Hormone transport; cortisol binding	Deficiency linked to chronic fatigue
SERPINA7	Thyroxine-Binding Globulin	Hormone transport; thyroxine binding	Deficiency results in hypothyroidism
SERPINA8	Angiotensinogen	Blood pressure regulation; amino-terminal cleavage by the protease renin results in release of the decapeptide of angiotensin I	Certain variants linked to essential hypertension
SERPINB3	Squamous Cell Carcinoma Antigen-1	Inhibit papain-like enzymes; cross-class inhibition of cathespins L & V	
SERPINB5	Mammary Serine Protease Inhibitor (Maspin)	Prevents metastasis in breast cancer and other cancers through uncharacterized mechanism	Down-regulation and/or intracellular location linked to tumor progression and overall prognosis
SERPINB6	Proteinase-Inhibitor-6 (PI6)	Prevent inappropriate activity of cytotixic apoptotic proteases; Inhibition of cathepsin G	
SERPINB9	Cytoplasmic antiproteinase 9 (PI9)	Prevent inappropriate activity of cytotoxic apoptotic proteases; Inhibition of granzyme B	
SERPINC1	Antithrombin	Thrombin and Factor Xa inhibition	Deficiency results in thrombosis
SERPIND1	Heparin Cofactor II	Thrombin Inhibition	May contribute to thrombotic risk when combined with other deficiencies
SERPINE1	Plasminogen Activator Inhibitor-1 (PAI-1)	Thrombin, uPA, tPA and plasmin inhibition	Abnormal bleeding
SERPINE2	Proteinase nexin I (P17)	uPA and tPA inhibition	
SERPING1	C1 inhibitor	C1 esterase inhibition	Angioedema
SERPINH1	Heat-Shock Protein 47kDa (hsp47)	Protein folding; Chaperone for collagens	
SERPINI1	Neuroserpin (PI 12)	tPA, uPA and plasmin inhibition	Polymerization results in dementia

The change from a native to cleaved conformation, also referred as the Stressed (S) to Relaxed (R) transition, brings the serpin into its lowest energy state and is irreversible ("suicidal"). In addition, the transition from S→ R acts as a signal that the serpin-protease is ready for destruction and is usually taken away through binding to the lipoprotein-related protein (LRP)(2,8).

The latent conformation is an alternative R state where the RCL is uncleaved but is still able to insert into the ß-sheet A. This change in conformation transitions the once native, active serpin into an inactive or non-inhibitory state that can drastically reduce the active lifetime of the serpin. The latent state has been studied the most in PAI-1, but it is also thought to occur in other serpins such as antithrombin, AAT and α1-antichymotrypsin (4,9-11)

The last two conformational states—δ, and polymeric—have only recently been characterized. The δ structure was first observed in α1-antichymotrypsin, where it was caused by a L55P mutation and resembled a 'half-way house' intermediate state. This conformation involved the partial insertion of the RCL into the ß –sheet A and the partial unraveling of Helix F, which inserted itself into the base of ß-sheet A in order to complete the ß-sheet hydrogen bonding (2,3,8,11). The polymeric conformation on the other hand has been the center of much controversy and will be discussed in detail later in this chapter (Section 1.3c).

Regardless of the conformation, there are 4 other structural regions—hinge, breach, shutter and gate-- that are characteristic of inhibitory serpins (Figure 1.1A) (3). The hinge region, a section of the RCL (P15-P9, towards the N-terminal) (12,13), provides mobility during the S→ R transition and therefore is crucial for RCL

conformation change. In fact, almost of all the residues in this region are conserved and the insertion of P14 into ß-sheet A is required for inhibitory activity (14).

The breach and the shutter regions are the two other major areas that aid in sheet opening, allowing insertion of the hinge region (15). The breach region is located at the top of ß-sheet A and is the point of the RCL's initial insertion (16), whereas the shutter domain is near the center of ß-sheet A and contains the highly conserved residues Ser-53 and Ser-56 that play an important role in serpin conformational transitions (17,18).

Lastly, the gate region consists of strands 3 and 4 of ß-sheet C (s3C and s4C respectively) and also contains several highly conserved residues. It has mostly been studied during experiments that focus on PAI-1's transition from an active to latent conformation (9). As discussed before, in the latent state the RCL is uncleaved but fully integrated into ß-Sheet A. In order for this to be accomplished, the RCL has to pass around the ß-turn linking s3C and s4C in the gate region (19). Aside from PAI-1 and antithrombin however, most serpins do not normally transition into the latent state *in vivo*. Therefore it is believed that the residue conservation within the gate region may be linked to native form maintenance and rather than latent state promotion (4).

c) How are they linked to disease?

While the conformational flexibility of the serpin structure is important to its function, it also leaves the structure vulnerable to mutations. Such mutations can result in deficiency or degenerative disorders that are collectively known as serpinopathies.

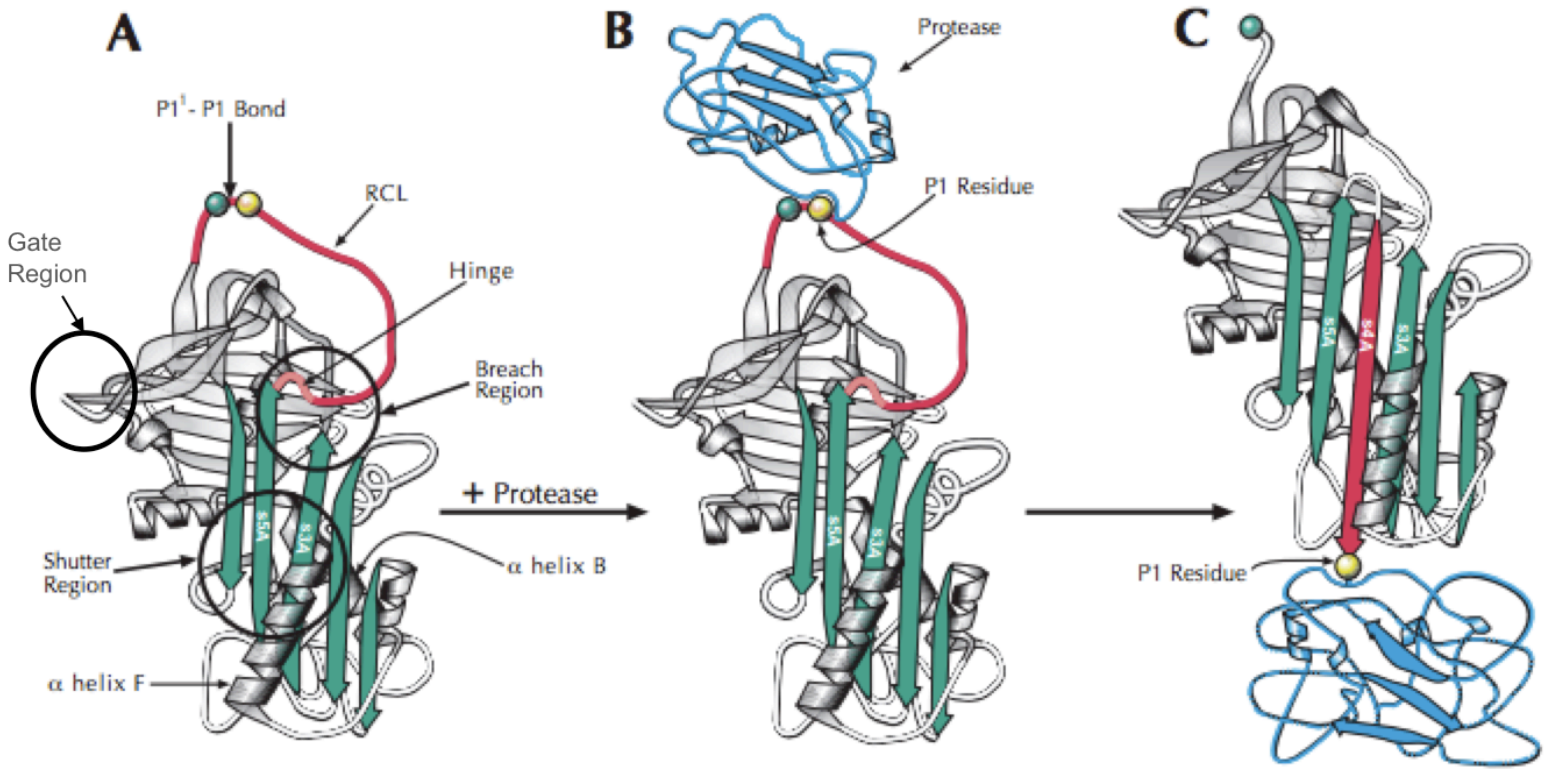


Figure 1.1.1. Serpin Structure and Inhibition Mechanism: A) Conserved Tertiary Structure of a Serpin with the central ß-sheet A highlighted in green and important features as indicated. B) Serpin-Protease Docking complex and in which the protease binds and cleaves at the P1-P1' bond. C) Final Serpin-Protease Complex in which the protease has been flipped over 60Å and its active site is distorted. (Adapted from (1))

Clinically, serpinopathies have been shown to result in uncontrolled activation of proteolytic cascades, thrombosis, angioedema or even chronic obstructive pulmonary diseases (COPD). (1-3,5,6,8,20).

Naturally occurring mutations have been identified throughout the serpin structure. However the most common and typically the most detrimental mutations are located within the shutter region (Table 1.2). These mutations cause an uncharacteristic opening of central ß-sheet A and allow the RCL to partially insert within the opening. Shutter region mutations not only impair the inhibitory mechanism by affecting RCL insertion, but they also predispose the mutant to polymerization *via* a head-to-tail association mechanism. These mutations have been implicated in the development of various clinical symptoms of serpinopathies: angioedema (C1-inhibitor), thrombosis (antithrombin) as well as lung and liver disease (α1-antichymotrypsin and α1-Antitrypsin) (1,3,8).

Although rare, some mutations cause uncharacteristic serpin-protease interactions. One example is the Pittsburg variant (Met358Arg) in AAT, which converts the serpin into an effective thrombin inhibitor. While only one case has been reported, the resulting increase in antithrombin activity led to a severe bleeding and ultimately death during childhood (1,2,21). Another example would be the Enschede variant (ΔAla366) of α2-antiplasmin, where an extra Alanine is added to the RCL. This addition interferes with the RCL's ability to insert into ß-sheet A, causing the protein to behave more like a substrate, rather than an inhibitor, of plasmin (1,2,22).

There are several serpin mutations that have been identified as promoters of the disease-linked latent state. One such mutation is the Wibble variant (Thr85Met) in

antithrombin, where about 10% of the circulating serpin adopts the latent conformation (2,3). Another example is δ-α1-antichymotrypsin which, as previously described, adopts an “intermediate” conformation as a result of the L55P mutation. It is thought that this conformation may exist between native and polymeric conformations and might also explain the loss of α1-antichymotrypsin activity associated with COPD patients with the L55P mutation (2,3,8,11).

Some mutations may lead to diseases if they appear in one serpin but not in another serpin. The mutation that arises at P17 gultamate in serpins is a notable example of this phenomenon. Indeed, a mutation at this location was recently discovered in herparin co-factor II (Glu428Lys) that, while linked to the serpin’s plasma deficiency, was not associated with any clinical disease (8,23). However when this same mutation is expressed in the Z variant of AAT (Glue342Lys), it not only leads to conformational instability and polymerization of the serpin (Figure1.2A and C) but also can manifest clinically as emphysema or cirrhosis (1-3,8). The Z variant and its associated serpinopathy, α1-Antirypsin Deficiency, will be discussed in more detail later on in the chapter (Section 1.3).

Familial Encephalopathy with Neuroserpin Inclusion Bodies (FENIB) is an autosomal dominant form of dementia that, like α1-Antirypsin Deficiency, is the product of serpin polymerization. In FENIB, neuroserpin polymers accumulate and become retained within sub-cortical neurons where they form periodic acid Schiff (PAS) positive, diastase resistant inclusions, called Collin’s bodies (20). These Collin’s bodies are similar to those seen in the Z variant of AAT (Figure 1.2).

Table 1.2. Serpin Mutations Related to Disease

Serpin	Alternative name	Mutation	Region Affected	Disease Associated	Ref
SERPINA3	Alpha -1 Antichymotrypsin	L55P (δ)	Shutter	Creates an 'half-way' house conformation	(2,3,8)
		P228L		Emphysema and cirrhosis	(8)
SERPINC1	Antithrombin	T85M (Wibble)	Shutter	10% of circulating serpin is latent	(2,3)
		T85L	Shutter	Thrombosis	(3)
		F229L	Breach	Spontaneous polymerization at elevated body temperature	(1,8)
		P54S/T	Shutter	Thrombosis	(3,8)
		N158D		Thrombosis	(8)
		C85R	Shutter	Thrombosis	(3)
		L99F	Shutter	Thrombosis	(3)
SERPIND1	Heparin Cofactor II	E428L	Breach	Plasma deficiency; not yet shown to cause disease	(8)
		P433L		Polymerization	(3)
SERPINF2	Alpha-2-antiplasmin	A366 added (Enschede)	RCL	substrate-like activity upon protease binding	(1,2)
SERPING1	C1 inhibitor	F52S	Shutter	Angioedema	(3,8)
		P54L	Shutter	Angioedema	(8)
		A349T	Shutter	Angioedema	(8)
		V366M	Shutter	Angioedema	(8)
		F370S		Angioedema	(8)
		P391S		Angioedema	(8)
		S54L	Shutter	Angioedema	(3)
SERPINI1	Neuroserpin (PI 12)	S49P (Syracuse)	Shutter	FENIB; Dementia, tremor, seizures terminally	(1,3,6,8,20)
		S52R (Portland)	Shutter	FENIB; Myoclonus, status epilepticus, dementia	(1,3,6,8,20)
		H338R	Shutter	FENIB; Myoclonic seizures, dementia, tremor, dysarthia	(1,3,6,8,20)
		G392E	Shutter	FENIB; Myoclonus, status epilepticus, dementia, chorea	(1,3,6,8,20)
		G392R	Shutter	FENIB; Dementia, epilepticus of slow-wave sleep	(20)

There are currently five known mutations that result in FENIB: Ser49Pro, Ser42Arg. His338Arg, Gly392Glu, and Gly392Arg (Table 1.2). The Ser49Pro (Syracuse) mutation was first discovered in a large Irish-American family and had small diffuse intra-neuronal neurosperin inclusions (Figure 1.2E) with an onset of symptoms between 45-60 years old. The second mutation, Ser42Arg (Portland), was shown to be a more severe mutation with larger inclusions (Figure 1.2F) and an onset of about 24 years old. The His338Arg was found in a third family with even more inclusions and an onset of symptoms around 15 years old. The fourth mutation (Gly392Glu) is particularly interesting because it caused the replacement of a conserved residue within the shutter region which produced large multiple inclusions in every neuron (Figure 1.2G), an onset of symptoms around 13 and death before 20 (1,3,6,8,20). Lastly, the Gly392Arg mutation was identified recently in an 8-year-old girl who displayed a drastic intellectual decline, seizures and epilepticus of slow-wave sleep (ESES) (20).

Cellular Response

Serpin polymers, it seems, do not activate the unfolded protein response in cells. This is most likely due to the highly ordered nature of the retained protein. Soluble misfolded monomers, on the other hand, are typically retained in the endoplasmic reticulum (ER), where they are targeted into the ER-associated degradation (ERAD) pathway.

Mutant serpins that are able to escape the ERAD and form homogenous ordered dimers and polymers are eventually trapped inside inclusion bodies. These inclusion bodies are insoluble and are no longer observed by the ER quality control (5).

The accumulation of serpin polymers however, can lead to a calcium-dependent activation of the transcription factor NF-κB via the regulator of G signaling 16 (RGS16) as seen in Figure 1.3 (5,24). This activation appears to eventually trigger the premature apoptosis that is observed in some serpinopathies (1).

1.2 Alpha1-Antitrypsin

a) Overview

Alpha1-Antitrypsin, also known as α_1-antiprotease inhibitor, is viewed as the model for the serpin family. Its main function is to protect the lungs from neutrophil elastase, a serine protease secreted by white blood cells (or neutrophils) that digests damaged and aging cells in addition to bacteria (5,25-28).

Following the standard serpin structure, the α1-Antitrypsin structure consists of nine α-helices, 3 ß-sheets and a mobile reactive center loop. In its native active form, the overall conformation is in a metastable state and it isn't until the serpin is rendered inactive (either through cleavage or latency) that the conformation is in the most energy efficient state (5,25-29).

This section will go into more detail about the α1-Antitrypsin protein, with a focus on its role in the body as well as its gene and protein structures.

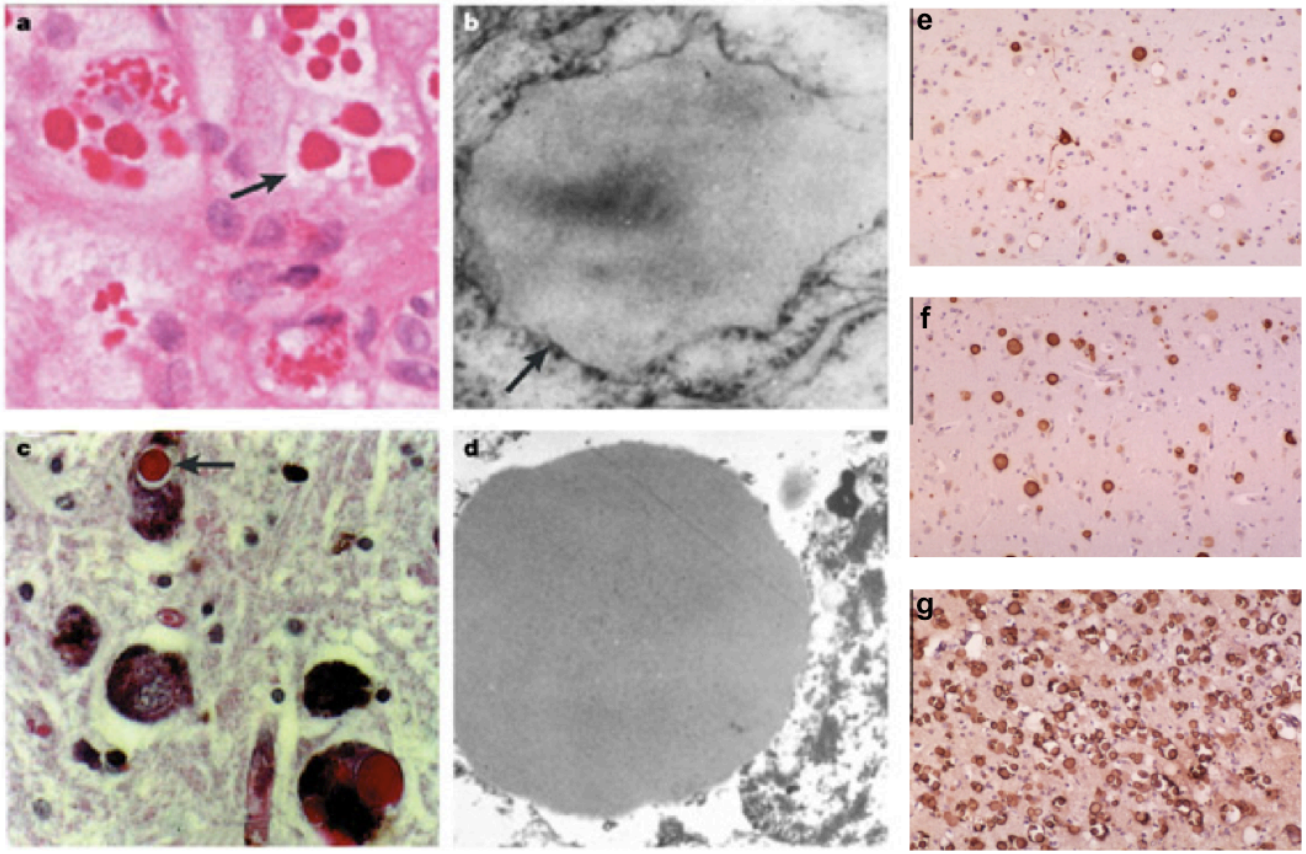


Figure 1.2 α1-Antitrypsin and Neuroserpin Polymers. α1-Antitrypsin and neuroserpin Intracellular inclusions identified in hepatocytes and neurons with PAS staining (a and c, respectively) and endoplasmic aggregates of the abnormal proteins on electron microscopy (b and d, respectively). Histology of neurosperin inclusions post-mortem for Ser49Pro (e), Ser52Arg (f), and Gly329Glu (g) mutations. Adapted from (5,6,20)

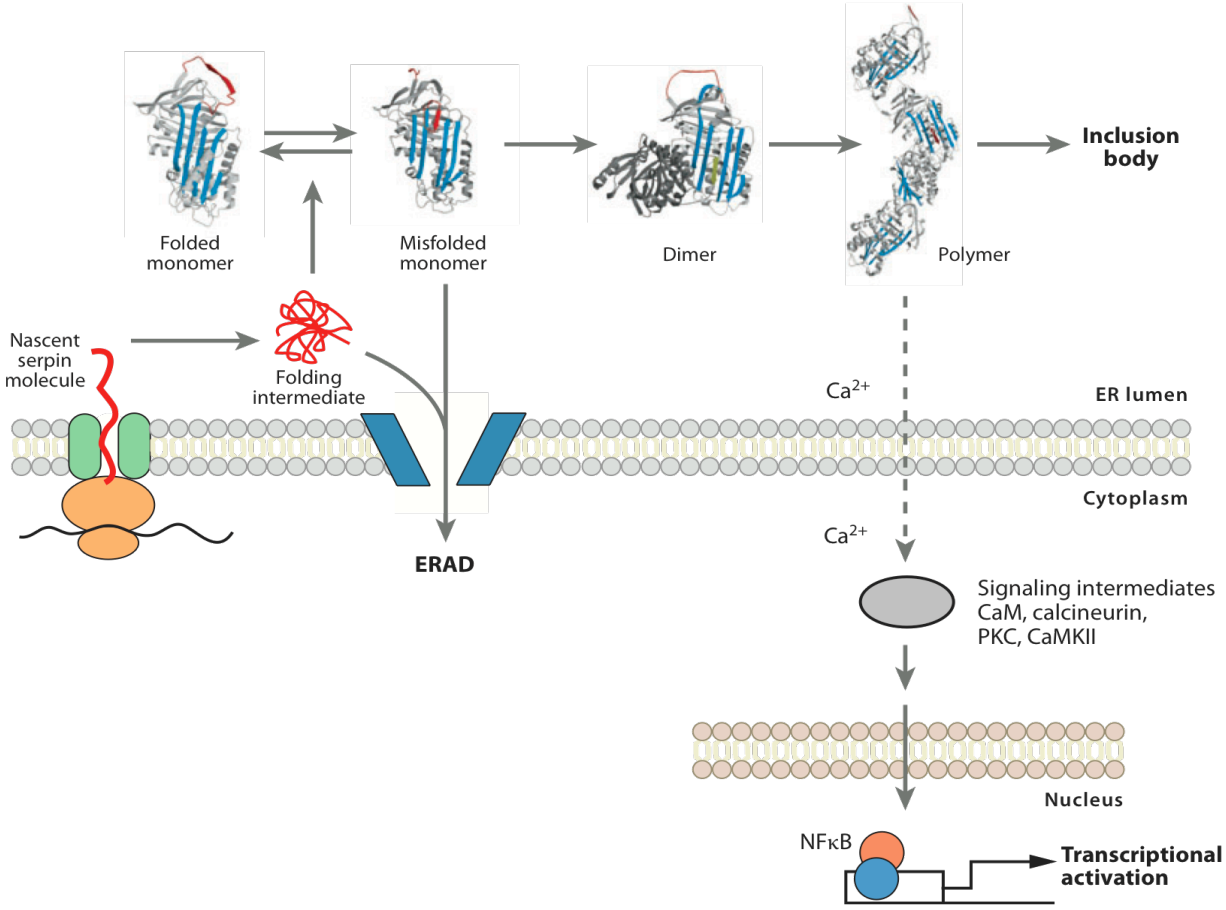


Figure 1.3 Cellular Responses to Serpin Polymers. Correctly folded (native) proteins are packaged into ER-to-Golgi transport vehicles for secretion (not shown), whereas misfolded monomer and folding intermediates are targeted by the ERAD. Some misfolded monomers are able to form dimer or polymers and form inclusion bodies with the ER. Accumulation of these inclusion bodies can cause an increase in cytosolic Ca^{2+} and eventually the activation of the NF-κB pathway. Taken from (5).

b) What is Alpha1 -Antitrypsin?

First identified as part of a clinical disease in 1963, by Laurell and Eriksson (30), Alpha1-Antitrypsin (AAT) is one of the most studied serpins today. AAT is an acute phase glycoprotein that is mainly synthesized in the liver and then secreted into the bloodstream, where its normal plasma concentration is around 1.5g/L. This can almost double during inflammation. In addition to hepatocytes, AAT can also be synthesized by macrophages and intestinal and bronchial epithelial cells. The main function of AAT is to protect tissues from the serine protease, neutrophil elastase (NE), but it also been shown to inhibit trypsin, cathepsin G and proteinase 3 (5,25-28).

Role in the Body

Once in the bloodstream, AAT travels to the lower tracheal region where it protects the alveolar tissue from proteolytic activity and damage produced by neutrophil elastase (NE). This protease is one of three packed within azurophil granules used by neutrophils to break down elastin and other extracelluar matrix components. The ability to break down most matrix proteins is especially helpful when neutrophils need to migrate through such matrixes in order to get to the main site of infection (31,32). However this also has the potential to be very harmful. For example, when AAT plasma levels get too low and there is not enough AAT available to protect the lungs, NE will start to degrade the healthy lung tissue in addition to its normal targets. This can lead to the onset of emphysema or other forms of COPD (5,25-28). The necessary balance between protease, NE, and anti-protease, AAT, lead to the proteinase-antiproteinase hypothesis over 40 years ago (27).

Gene Structure

The AAT gene is located on the long arm of Chromosome 14 (position q31-31.2) and contains 7 exons (5,27,28). One thing that is interesting about its location is that within 200kb of the AAT gene are two homologous DNA sequences, one coding for α1-antichymotrypsin and the other for a psuedogene. Within the AAT gene itself are two tissue-specific promoter regions, one for monocytes and one for hepatocytes. Messenger RNA (mRNA) produced by monocytes differ form hepatocyte mRNA in that the former contains 1-2 more exons than the later. The additional exons more than likely account for differences in mRNA expression levels between the two sources (monocytes express about 1% of the amount that hepatocytes do) as well as influence the rate at which the AAT is translated (28).

Protein Structure

With 394 amino acids and a molecular weight of approximately 54KD, AAT is a highly ordered and large globular polypeptide that is N-glycosylated at positions Asn46, Asn83 and Asn247. This glycosylation is heterogenous, with Asn83 sometimes having two or three N-glycans attached (28,33-35).

Keeping in line with the conserved serpin structure, AAT's structure consists of nine α-helices (A-I), three β-pleated sheets (A-C), and a Reactive Center Loop (RCL) (5,25-29). During AAT's active form, the RCL is thought to be intrinsically disordered; however as soon as the protease forms a Michaelis complex with AAT, the RCL is cut at Met358 and Ser359 (referred to as P1 and P1' respectively) (28). Once this happens, the P1' residue is released and an ester bond is formed between AAT and the protease. The protease is then flipped from the upper pole of AAT to the lower pole, causing the

RCL to be inserted as an extra strand (s4A) into the central β-sheet A (5,25-29). This conformational change is very stable and acts as a signal to the cell that the protease is ready for destruction and is taken away, as illustrated earlier in this chapter.

1.3 Alpha1-Antitrypsin Deficiency

a) Overview

Alpha1-Antitrypsin Deficiency is an often-misdiagnosed yet increasingly common genetic disease that is associated with both liver and lung disease and is thought to have a frequency comparable to cystic fibrosis (28,36). In the fifty years since Laurrel and Eriksson "discovered" it as a clinical disease (30), studies have shown that certain mutations in the serpin α1-Antitrypsin can cause it to misfold and form polymers. While it is believed that this polymerization process is the underlying cause for Alpha1-Antitrypsin Deficiency, the exact mechanism in which this occurs is still unknown. In this section we will discuss what exactly Alpha1-Antitrypsin Deficiency is, including what causes it and how it is currently being treated today.

b) What is Alpha1-Antitrypsin Deficiency?

While Alpha1-Antitrypsin Deficiency was reported in an Alaskan girl who died 800 years ago (37) and it may have been responsible for the premature death of Fredric Chopin in 1849 (38,39), it wasn't until 1963 that it was first describe as a clinical disease. It was at this time that Carl-Bertil Laurell and Sten Eriksson first observed the absence of the α1 protein band in plasma taken from two patients in a respiratory-disease hospital. This absence was later linked to a deficiency in the α1-Antitrypsin

protein. It was this connection that not only helped to reveal an underlying mechanism for emphysema but also introduced scientists to a whole new family of proteins, the serpins (27,40).

As discussed previously, serpins are prone to mutations that lead to polymerization (Section 1.1c) and AAT is no exception to this phenomenon. In fact, the polymerization of AAT has been linked with the conformational disease, Alpha1-antritrypsin deficiency (AATD), and can result not only in the development of chronic obstructive pulmonary disease (COPD) but also hepatitis, cirrhosis and even hepatocellular carcinoma (Figure 1.4) (36,41). AATD has been shown to affect people of virtually all ages, populations and ethnic groups. In fact it is estimated that there are of approximately 116 million carriers and 1.1 million patients with severe AATD worldwide, indicating that the disease is probably one of the most common severe hereditary disorders in the world (42).

Clinical Manifestations

AATD can manifest itself clinically though both pulmonary and hepatic manifestations, although these may vary widely from patient to patient, with both rarely seen in the same patient (43). The dominant manifestation is in lung disease, which begins in early adulthood, and can develop sooner if additional risks are present (e.g. tobacco smoking or occupational air pollutants). Severe lung impairment, such as COPD and emphysema, are predicted to manifest when AAT serum concentration falls below the protective threshold of 35% the normal mean value (≤ 0.8g/L) (43). In addition to low AAT serum levels, it is also thought that AAT polymers themselves may promote

inflammatory processes in the lung. This increases the amount of NE present in the lungs and further contributes to the accelerated destruction of lung tissue (36,43,44).

While AATD is one of the most frequent causes of liver cirrhosis (after viral hepatitis, alcohol abuse, and chronic cholangitis), adults with AATD-related genotypes develop liver disease less frequently than lung disease. The progression of liver disease in AATD patients appears to be slow and typically manifests first as hepatitis (swelling and inflammation of the liver) before developing into fibrosis or cirrhosis. It has also been reported that there is a relatively higher prevalence of hepatocellular carcinoma and cholangiocellular carcinoma associated with AATD than with other hepatic diseases (36,43). Neonatal hepatitis syndrome with cholestais (bile blockage) occurs in a small percentage of newborns that are homozygous for deficiency AATD genotypes. Prolong jaundice after birth with conjugated hyperbilirubinemia (bilirubin concentrations greater than 2mg/dL (45)) are classic clinical signs of this genotype, although the jaundice may resolve itself spontaneously after a few weeks (36,43).

In rare occasions, AATD has been associated with the skin disorder panniculitis, Wegener's granulomatosis (antiproteinase-3-associated vasculitis), and aneurysms of the abdominal aorta and brain arteries (36,43).

c) What causes Alpha1-Antitrypsin Deficiency?

As discussed in Section1.2b, the major function of AAT in the body is to regulate NE activity. However when AAT levels are low, NE activity is left unchecked, allowing it to digest healthy lung tissue in addition to its normal target. This can result in emphysema and COPD, two disease associated with AATD.

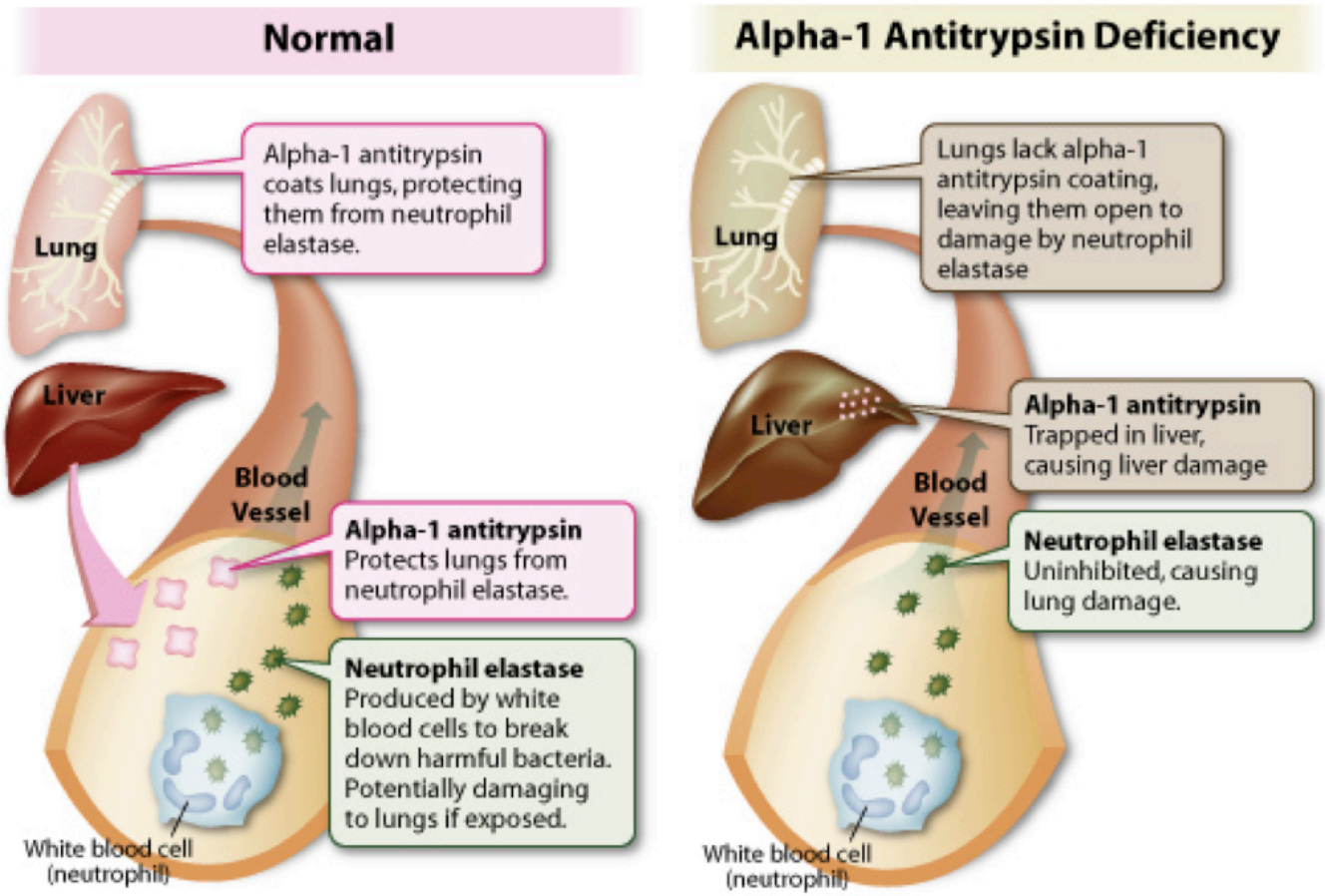


Figure 1.4 The Effects of Alpha1-Antitrypsin Deficiency (46).

Typically a decrease in AAT levels is due to a Z point mutation (E342K) that disrupts a salt-bridge within the structure. When this occurs, the protein misfolds, and forms polymers that accumulate and create inclusion bodies inside the endoplasmic reticulum of hepatocytes where it is produced. With the AAT protein trapped in the liver, it cannot travel to the lungs and therefore cannot protect the lung tissue from damage produced by NE. In addition, the polymers located in the liver can also lead to the destruction of the hepatocytes themselves, therefore leading to diseases such as cirrhosis and hepatitis (5,25-29).

Common mutations and their effect

Among the over 75 different variants of AAT (5), there is a small percentage of variants that negatively affect AAT's structure (Table 1.3). Of these, the Z variant is the most common severe mutation and accounts for 95% of all cases of AATD (47,48). The change from a glutamic acid to a lysine at residue 342 not only disrupts a salt bridge that is normally present between K290 and E342, leading to the misfolding and polymerization of the protein (28,47,48), but it may also be responsible for the failure of the protein to be secreted (28). The plasma levels in the MZ heterozygote are 60% of normal levels (normal levels are 1.4-3.5 g/L) whereas the ZZ homozygote has only 10% of the normal (M) plasma levels (27). The low levels of plasma AAT are because over 80% of the ZZ homozygote protein are not secreted but polymerize and accumulate as inclusions in the rough ER of the liver (29). As such, ZZ homozygotes also have a reduced life span in which they have a 52% chance of being alive at age 50 and a 16% chance at age 60 (normal is 93% and 85% respectively). Smoking further reduces this by 10 years (28). The Z mutation is most prevalent in Northern and Western Europe

with the highest concentration in southern Scandinavia, Denmark, the Netherlands, the United Kingdom and northern France (36,49).

Another common mutation is the S variant (E264V), which is located in the shutter region and reduces plasma AAT concentrations by 40% in homozygotes. While the S variant alone is not affiliated with clinical disease, when it is co-inherited with the Z variant, the two can interact to form heteropolymers within hepatocytes which then develope into inclusions bodies and eventually cirrhosis (27,50). One explanation for the lack of associated disease in homozygous S carriers is that this genotype produces unstable AAT protein that can easily be degraded outside the hepatocyte, which in turn affects its half-life (36). Unlike the Z variant, the S variant is most prevalent in Southern Europe with the highest frequency in the Iberian Peninsula (36,49).

Some other mutations that can lead to AATD include the Siiyama variant (S53F), the Mmalton variant (∆52F), and the I variant (R39C). The Siiyama variant is the most common cause of AATD in Japan whereas the Mmalton (also known as Mnichinan and Mcagliari) variant is the main cause of AATD in the isolated island of Sardinia. Both of these mutations destabilize the ß-sheet A in the shutter region, causing it to form unfolded intermediates and polymers *in vivo* (6,8,27,41). The I variant also affects the shutter region, but does so at a lesser extent. This causes polymer formation to be slowed (when compared to Z polymers) which in turn result in reduced protein retention in hepatocytes, milder plasma deficiency and lack of clinical disease. However, like the S variant, a co-inheritance with the Z variant can lead to liver disease such as cirrhosis (6,8,27,41).

Polymerization Mechanisms

While it is well recognized that the polymerization of AAT is the underlying cause of AATD, the exact mechanism of this polymerization phenomenon is unknown. Three different models have been present over the years that strive to answer this question: 1) s4A linkage 2) s4A/s5A domain swapping and 3) C-terminal swapping (Figure 1.5). However in order to fully understand what these models are proposing, it is important to look at what is currently known about the folding pathway of AAT.

Studies focusing on the unfolding and folding patterns of AAT go back over 20 years. In these initial studies, three species --native (N), unfolded (U) and intermediate (I)—were identified during the unfolding process (47,51). This was shown to be true for both the normal (M) and polymerization-prone (Z) forms of AAT. While both MAAT and ZAAT transitioned from the U to I state at similar rates, they differed in their ability to perform the final step of folding, I→N. It was revealed that ZAAT folding from the intermediate state was significantly slowed, and it was this accumulation of I that resulted in polymerization (47). The structure of the intermediate state, more commonly referred as the polymerogenic monomer M*, has been debated greatly and is the basis for differences seen in the three polymerization models.

Table 1.3 Common α1-Antitrypsin Variations (Compiled from (36,41))

Variant, Mutation	Polymerization Tendency	Circulating Deficiency in Homozygotes	Association with Clinically Significant Liver Disease	Epidemiology
Z, E342K	+++	Severe (10-15% of normal levels)	Yes in homozygotes, with less hepatocyte inclusion bodies present in heterozygotes (no clinical disease associated)	1 in out of 27 North European heterozygotes. Most common severe deficiency variant.
Siiyama, S53F	+++	Severe	Yes	Most common severe variant in Japanese populations
Mmalton, Δ52F	+++	Severe	Yes	Most common severe variant in Sardinian populations
S, E264V	+	Moderate (60% of normal levels in homozygotes)	Reported in SZ AAT compound heterozygotes	1 out of 5 Europeans are heterozygotes. Most common deficiency variant.
I, R38C	+	Mild (extrapolation from levels in heterozygotes)	Case report in IZ AAT heterozygote	Only reported in compound heterozygotes
M	-	Normal levels (1.4-3.5 g/L)	No, although heterozygotes can, on rare occasion, develop clinical symptoms	95% of individuals are homozygotes. Wild-type AAT phenotype

For the longest time, the s4A linkage hypothesis was the accepted mechanism for serpin polymerization, and was not challenge until just recently. This hypothesis is based on the fact that the Z mutation is located on the head of strand 5 of β-sheet A at the base of the RCL (also referred to as the s4A), causing β-sheet A to open and transition to an unstable polymerogenic M* intermediate (26,52,53). The expansion of upper β-sheet A allows the s4A position to be opened, allowing the RCL residues to partially insert, releasing the first strand of β-sheet C (s1C) and partial uncoiling of helix F. Next, the lower part of β-sheet A is expanded, allowing helix F to regain its shape and opening the molecule to accept the RCL from another molecule, thereby forming a chain of loop-sheet polymers (Figure 1.5) (5,53). This hypothesis has been supported by extensive biochemical and biophysical analyses.

Nevertheless, in 2008, this model was challenged when Yamasaki and colleagues presented a model in which the fourth and fifth strand of β-sheet A (s4A and s5A respectively) of one molecule donates itself to another molecule in order to complete the formation of the second molecule (referred to here as s4A/s5A domain swapping, Figure 1.5). This model was based on the crystal structure of a self-terminating antithrombin dimer formed in the presence of the denaturant guanidine hydrochloride (54). The physiological relevance of the s4A/s5A domain swap model was brought into question however, when extensive studies by the Lomas group indicated that only heat-induced polymers are able to correctly mimic those found in hepatocytes (55). This was based largely on the identification and characterization of the 2C1 antibody, which was shown to identify AAT polymers both *in vitro* and *in vivo* (56).

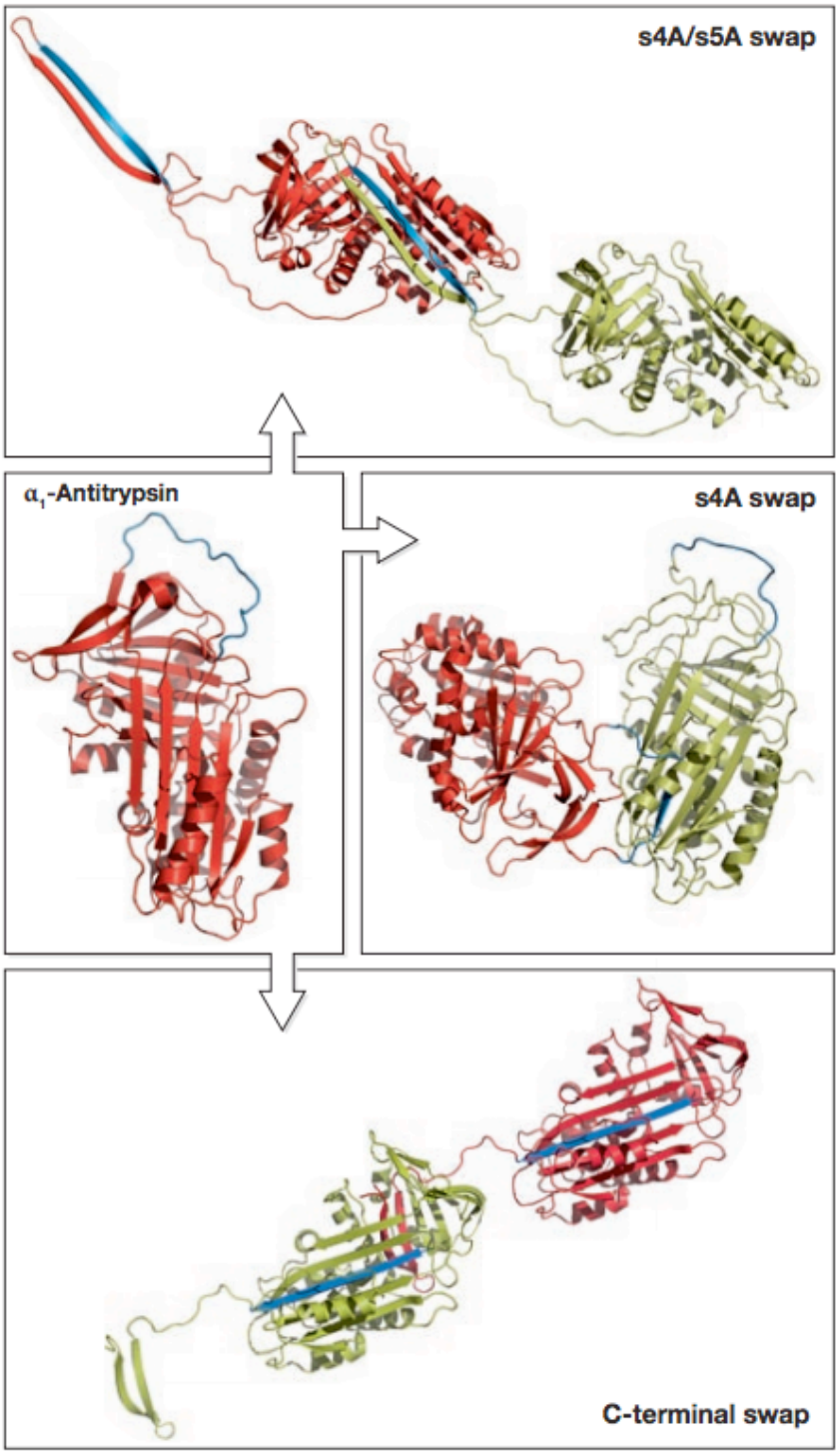


Figure 1.5 Alpha1-Antitrypsin Structure and Its Three Polymerization Models (57).

Based on this, in 2011, Yamasaki and colleagues set out to determine the structure of the heat-induced polymers, the result of which produced a third model of polymerization. In this newest model, the C-terminal swap (Figure 1.5), monomers are linked together via domain swapping that involves the carboxyl (C-) terminus of protein, specifically s1C and strand 4 and 5 from ß-sheet B (s4B and s5B respectively). Polymers produced in this study were shown to bind to the 2C1 antibody and were obtained by heating recombinant protein at 60ºC (58).

It should be noted that the C-term swap polymers were not the only kind present among those analyzed in this study. In fact, the heat-induced polymers contained a mixture of both the s4A/s5A polymers and the C-terminal polymers. This polymer heterogeneity could have important implications in regards to understanding AATD and the development of therapeutics. In addition, it suggests the possibility of more than one polymer structure being present simultaneously and raises the question of if toxic and non-toxic polymers can coexist (57,58). Therefore more research is needed in order to fully understand the types and effects of polymerization in AATD.

d) How is Alpha1-Antitrypsin Deficiency currently treated?

People who suffer from AATD (commonly referred to as Alphas) are very limited in their options for treatment (59). For example, the most common treatment for end-stage lung and liver disease is organ transplantation. Another popular therapy for Alphas is intravenous augmentation of pooled plasma purified AAT, however this process is very costly and often involves weekly treatments (36,43,46,59,60). In the scientific community it is widely thought that more effective forms of treatments lie in the attenuation of the polymerization process itself. There are currently three main

strategies being developed that address this issue: 1) use of chemical chaperones, 2) use of RCL peptides analogues and 3) targeting a surface cavity within the structure (5,41,61).

The ability of chemical chaperones to help stabilize intermediates along the folding pathway have been supported by studies where osmolytes such as trimethylamine oxide, betaine and sarcosine were able to stabilize AAT against polymerization (62). Chemical chaperones were also shown to be effective in cell and animal models for disease (63-65), but failed to show any significant improvement in clinical trials (66). In addition, since chemical chaperones promote protein folding, there lies a risk that it may promote ZAAT to their lowest energy state, i.e polymers (41).

The second approach involves using RCL peptide analogues to compete for binding to ß-sheet A and therefore block polymerization. This is founded on a study that showed an antrithrombin RCL peptide was able to readily insert into the ß-sheet A of AAT, and vice versa (14). While helpful in understanding the polymerization mechanism, the peptides proved to be too large and promiscuous for drug design (5). To compensate, several groups have designed smaller peptides that end polymerization by specifically annealing to the lower part of ß-sheet A (41,67). More research is needed on these peptides before they are ready for therapeutic application.

Lastly, the third approach also uses a structural drug design approach but instead of using peptides to target the RCL, it uses small molecules to block polymerization by filling a surface hydrophobic pocket in AAT. The hydrophobic pocket is surrounded by strand 2A and helices D and E and can be used for allosteric blockade of the conformational transition that underlies polymerization (5,41,61). The pocket is

unobstructed in the native protein but is filled as ß-sheet A accepts the RCL from another molecule. When this pocket was partially filled by mutagenesis (Thr114Phe), scientists saw a 10-fold reduction of polymerization and an increased ZAAT secretion in a *Xenopus* oocyte expression system (68). This has led to use of *in silico* experiments to identify small drug-like molecules that are able to target this pocket and inhibit polymerization (5,41,61).

1.4 Expression of Recombinant Protein

a) Overview

The practice of using different model systems as a vessel to produce heterologous proteins is a very common methodology in both industry and academia. Recombinant proteins have not only been used as tools for cellular and molecular biology but they can also be used as pharmaceuticals and vaccines (69,70).

Not only are recombinant proteins diverse in their applications but they also vary greatly in the way they are produced. Several expression systems are currently available today, which range from the microbial to the mammalian. These systems not only differ in the time, ease and cost of expression but also in the size of protein they are able to produce and the posttranslational processing that can be performed. This section will look at a select few of the systems available today and will discuss the pros and cons of each.

b) Bacteria

Although many different types of prokaryotes have been used for recombinant protein expression, the most common one by far is the Gram-negative bacterium, *Escherichia coli.* Some other popular bacterial systems include *Ralstonia eutropha, Pseudomonas fluorescens, Staphylococcus carnosus,* and certain Bacill strains (70,71). Out of the expression systems discussed in this section, bacterial systems have the lowest costs and the highest yields of protein (72).

Escherichia coli

E. coli was one of the first organisms used for heterologous human protein production and is one of the most common hosts used today (71). For these reasons, *E. coli* is on of the most well understood microorganisms with a vast amount of knowledge known regarding its molecular biology, biochemistry and physiology (69). There are several advantages of using *E. coli* as an expression system, including ease of culture, high product yields, and rapid growth and expression. In fact *E. coli,* can be grown to very high densities (over 100g/L) and requires a very simple nutritional regiment. However when *E. coli* is grown at too high of densities, it can result in toxicity due to acetate formation. Luckily this can be avoided by controlling the oxygen levels of the *E. coli* culture (70). Another advantage of using *E. coli* is that its genome can easily be manipulated in a short amount of time and with high precision. This can be very useful in the transformation process and allows the plasmid copy number to be quickly altered. *E. coli* is also able to thrive under several different environmental conditions and can produce recombinant proteins up to 80% of its dry cell weight (70).

In spite of all the advantages in using *E. coli*, there are also several disadvantages. One such disadvantage is *E. coli's* inability to perform posttranslational modifications (PTM), particularly glycosylation, which can affect protein stability, solubility, folding, biological activity, localization, and circulation half-life (69). In addition proteolytic cleavage and disulfide bond formation rarely occur (69) and refolding of these proteins is extraordinarily difficult (70). The formation of inclusion bodies filled with inactive, aggregated and insoluble protein is also a common issue and often requires the refolding of the protein of interest. While yields are typically higher when produced in the cytoplasm, proteins can also be secreted into the periplasmic space, which can help with disulfide formation and reduction in inclusion bodies (70,71). Additional characteristics of this system can be found in Table 1.4.

Other Bacteria

Secondary to the *E. coli* system, Gram-positive Bacilli strains are one of the most popular bacterial expression systems for recombinant proteins (71). Like *E. coli*, Bacilli are easily manipulated and are genetically well-characterized systems. The *bacillus* system also displays strong secretion with no involvement of intracellular inclusion bodies and has highly developed transformation and gene replacement technologies already in place (70). Some commonly used species include *Bacillus magaterium, Bacillus subtilis, Bacillus brevis* and *Bacillus licheniformis* (70,71). One particular advantage of the *Bacillus* system is that members do not have lipopolysaccharide-containing outer membranes like Gram-negative bacteria do, thereby eliminating the extra step of endotoxin removal normally necessary in the later (70,71).

Some other bacterial expression systems include *Ralstonia eutropha, Pseudomonas fluorescens,* and *Staphylococcus carnosus. R. eutropha* is a Gram-negative bacterium that, with respect to inclusion body formation, appears to be better than *E. coli* (73,74). In fact, organophosphohydrolase, a protein that typically is prone to inclusion body formation, was produced at 10g/L in *R. eutropha*, compared to *E. coli* where it is normally produced at less than 100mg/L (74). A *P. fluorescens*-based platform, referred to as the P*f*enex Expression Technology™ , has recently emerged with high quality profiles and yields that sometimes can be more than 20g/L (75). *S. carnosus,* was shown to produce and secrete a mammalian protein at 2g/L compared to *Streptomyces lvidans* which was only able to express 0.2g/L (70).

c) Yeasts

Humans have used yeasts, the single-cell eukaryotic organisms, since the Neolithic age. While originally used to produce food and beverages (76), yeasts nowadays are used for multiple applications including heterologous protein production. In fact, they are the most common choice for the expression of recombinant proteins that are not successfully produced in *E. coli* due to improper folding or lack of glycosylation (70). While slightly more expensive then bacterial systems, yeast systems are still a relatively cost-efficient and simple expression system for recombinant proteins (Table 1.4) (72). The two most used yeast species are *Saccharomyces cerevisiae* and *Pichia pastoris*.

Saccharomyces cerevisiae

S. cerevisiae was first manipulated for recombinant protein expression use in 1987, when it was used to express human insulin (77). Since then, countless more

proteins have been expressed in this yeast species. Due to its history and its many applications, the molecular biology and physiology of *S. cerevisiae* has been extensively researched and accumulated (69).

S. cerevisiae has many advantages over *E. coli* as an expression system, some of which include its long history of fermentation, ability to secrete proteins into the media and the glycosylation of proteins. Since *S. cerevisiae* naturally secretes very few proteins, the secretion of heterologous proteins acts as a first step in purification, thereby reducing the time needed to purify the protein of interest. Glycosylation by *S. cerevisiae,* however is often intolerable for mammalian cells due to differences in their O-linked oligosaccharides and the fact that *S. cerevisiae* has a tendency to over glycosylate N-linked sites (70). Since glycosylation is species-tissue- and cell-type-specific and necessary for protein stability, solubility, folding, biological activity, localization, and circulation half-life, this issue could drastically affect the usability of certain proteins downstream (69).

Pichia pastoris

With cell densities as high as 130g/L and protein concentrations up to 10g/L, the methylotrophic yeast, *Pichia pastoris* has quickly emerged as a powerful yeast alternative to *S. cerevisiae* (69,70). Some additional advantages of *P. pastoris* over *S. cerevisiae* are avoidance of hyper-glycosylation, growth in high methanol solutions, and integration into chromosomal DNA (70). By far, the greatest advantage of using *P. pastoris* is the fact that it is capable of producing disulfide bonds and glycosylation of proteins. Unlike *S. cerevisiae*, N-glycosylation in *P. pastoris* is less complex and occurs less often (69).

As a methylotrophic yeast, *P. pastoris* is able to metabolize methanol as its sole carbon source, allowing it to grow in high methanol solutions that would normally kill other microorganisms (70). In addition, the *AOX1* promoter, which regulates the *AOX1* gene that expresses the alcohol oxidase (AOX) enzyme needed to utilize methanol, is often used as a promoter in the over-expression of recombinant protein genes introduced via a *Pichia* expression vector (78,79). When grown on glucose, glycerol or ethanol, AOX expression is undetectable in *Pichia* cells, however when they are switched to a methanol solution, expression is dramatically increased. This allows scientists to easily control the expression of heterologous proteins by simply switching the media (80,81). However, care should be taken when using methanol, since it is a flammable reagent and is toxic to humans (69). Lastly, many of the techniques needed for molecular genetic manipulation in *S. cerevisiae*, can also be used in *P. pastoris* thereby making *P. pastoris* an easily manipulated, user-friendly, heterologous protein producing machine (80,81).

d) Insect Cells

A common alternative to using a yeast system for recombinant protein production is to use insect cells (72). Insect cells are able to carry out more complex PTMs than those produced by yeasts and also have one of the best folding machineries for mammalian proteins (70,82). Instead of a plasmid vector, insect cells are commonly infected with baculoviruses that contain relatively large, double-stranded circular DNA genomes. The most common baculovirus used is the *Autographa californica* multicapsid nucleopolyhedrovirus (Ac*M*NPV) that is easy to culture and is able to infect approximately 25 different lepidopteran insects (i.e. moth and butterflies), giving it a

broad host range compared to other baculoviruses (70,83). The most common host is the fall armyworm (*Spodoptera frugiperda*), which in larvae form can be a lot cheaper than mammalian cell cultures (70).

There are several advantages to using the baculovirus- infected insect cells for recombinant protein production. As stated before, this system is able to carry out more complex PTMs which include phosphorylation, correct signal peptide cleavage, acylation, N- and O-glycosylation, and proper proteolytic processing, among others (70,84,85). In addition, insect cells are able to properly fold proteins and allow for disulfide formation. Other advantages include no limit on protein size, ability to express multiple genes at once (86), and high expression levels and possible yields of 11g/L (70). Since the baculoviruses used to infect the insect cells are only able to attack invertebrates and not vertebrates, an added layer of safety is added overall to the system (70).

As with any system, there are some disadvantages in using the insect cell system for heterologous protein production. Expression is often very time-consuming, especially when compared to bacterial systems. In addition, protein yields can vary as a result of several different variables including cell culture medium, virus titer, and expression time after infection (87). While insect cells are able to glycosylate proteins, there have been reports in which this has occurred incorrectly. Compared to the different glycoforms present in mammalian cells, potential N-glycosylation sites for proteins produced in insects are typically either fully glycosylated or not glycosylated at all (70). Additional characteristics of the insect cell system are shown in Table 1.4.

e) Mammalian Cells

Over the past two decades, mammalian cells (Table 1.4) have become one of the dominant systems for recombinant protein production, especially for proteins that require mammalian glycosylation or are not folded correctly in other systems (88). The most frequently used mammalian cell lines are Chinese hamster ovary (CHO) and COS-7 cells [derived from CV-1 (simian) cells in origin and carry the SV40 genetic material] (89). Other popular cell lines include mouse myelomas (NS0 and SP2/0), human embryonic kidney (HEK-293) cells and baby hamster kidney (BHK-21) cells (70,88,90).

One of the major benefits of using mammalian cells is that the proteins they produce are properly folded and glycosylated, thereby removing the need for protein refolding. Aside from glycosylation, mammalian cells can also add fatty acid chains, and phosphorylate Tyr, Thr and Ser hydroxyl groups on heterologous proteins (70). Another benefit of the mammalian system is their high productivity of 20-60pg/cell/day with some protein titers even producing as much as 10g/L in a industry setting (70). In addition, foreign DNA can be inserted into the nucleus of mammalian cells in a multitude of ways including calcium phosphate, electroporation, coprecipitation, lentiviral transduction, and transfection with lipofectamine and other chemical reagents (89).

There are a few disadvantages to using the mammalian cell system. To begin with, mammalian processes are very expensive (especially the media which requires the addition of hormones and growth factor) and are very time-consuming. Mammalian cells are also prone to contamination from viruses or prions, which ultimately adds to the cost of production, and are very poor at being able to secrete proteins (70).

Table 1.4 Characteristics of the Different Recombinant Protein Expression Systems Discussed in this Chapter. (Taken from (72))

Characteristics	*E. coli*	Yeast	Insect Cells	Mammalian Cells
Cell Growth	Rapid (30 min)	Rapid (90 min)	Slow (18-24 hr)	Slow (24 hr)
Complexity of Growth Medium	Minimum	Minimum	Complex	Complex
Cost of Growth Medium	Low	Low	High	High
Expression Level	High	Low-High	Low-High	Low-Moderate
Extracellular Expression	Secretion to Periplasm	Secretion to Medium	Secretion to Medium	Secretion to Medium
Protein Folding	Refolding Usually Required	Refolding May Be Required	Proper Folding	Proper Folding
N-linked Glycosylation	None	High Mannose	Simple, No Sialic Acid	Complex
O-linked Glycosylation	No	Yes	Yes	Yes
Phosphorylation	No	Yes	Yes	Yes
Acetylation	No	Yes	Yes	Yes
Acylation	No	Yes	Yes	Yes
gamma-Carboxylation	No	No	No	Yes
Yield (mg) (per liter culture)	50-500	10-200	10-200	0.1-100
Success Rate (%) (Soluble of Functional)	40-60	50-70	50-70	80-95
Project Cost	Low	Low	Middle	High
Recommended Use	Antigen protein, Protein standards, functional proteins	Proteins with glycosylation, Vaccine, Secreted form, Alternative to insect cell system	Proteins with glycosylation, Assay standards, Secreted form, Alternative to yeast system	Functional study, PTM study, Assay standards, characterization
Advantage	Simple, robust, lowest cost, highest yield	Simple, low cost, good for certain proteins	Relatively higher yield, better PTM	Natural protein configuration, best PTM
Disadvantage	Least PTM	Longer time, less PTM	Longer time, higher cost	Highest cost, lower yield

CHAPTER 2:

PROJECT RATIONALE AND SUMMARY

Alpha1-Antitrypsin Deficiency is an autosomal recessive disorder that can manifest clinically as liver cirrhosis, emphysema and, on rare occasions, as the skin disease panniculitis (27,36,61). The onset of this disorder has been linked to mutations in the archetype serpin, α1-Antitrypsin (AAT), that cause it to misfold and polymerize.

AAT is an acute phase glycoprotein that is primarily synthesized in the liver and then secreted into the bloodstream. Once in the bloodstream, AAT travels to the lower tracheal region where it protects the alveolar tissue from neutrophil elastase, a serine protease that aids in the digestion of bacteria as well as damaged and aging cells (27,31,32,36,41,43,50,61,91).

However when levels of AAT are low, neutrophil elastase will digest healthy lung tissue as well. Typically this decrease in AAT levels is due to a point mutation, referred as the Z variant (E342K), that disrupts a salt-bridge within the structure and causes the protein to misfold. These misfolded proteins can then form polymers that accumulate and create inclusion bodies inside the endoplasmic reticulum of hepatocytes (27,36,41,43,50,61,91). This reduction in available AAT causes a two-fold effect: 1) uncontrolled activity of neutrophil elastase increases the likelihood of emphysema or other chronic obstructive pulmonary diseases developing, and 2) accumulation of AAT

polymers can destroy hepatocytes leading to the on-set of liver diseases, such as cirrhosis and hepatitis (5,25-29).

Current treatments for Alpha1-Antyrpsin Deficiency (AATD) are very limited, the most common of which is substitution therapy with human AAT protein. This method is often very time consuming, requiring weekly intravenous treatments of protein derived from pooled human plasma, and can be very costly (43). Instead of treatments that add AAT proteins to reverse the effects of AATD, most researchers today believe that more efficient and cost-effective treatments will be found by concentrating on the disruption of AAT polymerization. However, even after over twenty years of research on AAT, the exact molecular basis of this polymerization is still unclear. Therefore, in order to find a treatment for AATD, additional research is needed to further understand the polymerization mechanism of AAT.

Most polymerization studies today use AAT protein that has been purified from the plasma of affected patients. The amount of protein acquired using this method is very limited however, making it difficult to perform more in-depth studies on the polymerization mechanism. The ability to produce recombinant AAT protein, on the other hand, could provide a safer alternative to obtain needed protein concentrations.

Since the first recombinant AAT (wild-type, M) was purified from yeast in 1985 (92), numerous expression systems have been used to produce MAAT protein (35,89,93-99). However the Z variant of AAT, with its tendency to polymerize, has long thwarted any attempts at heterologous production. In 2009, a published study indicated the first successful attempt at producing recombinant ZAAT protein (100). This study,

which used *Pichia pastoris* to produce tagged ZAAT protein, paved the way to an unlimited source of AAT protein.

The goal of this thesis research was to build upon the 2009 paper in order to produce copious amounts of protein that could be used for future experiments focusing on Z α_1-Antitrypsin polymerization. In particular, this thesis discusses the attempts made at producing both tagged and untagged versions of recombinant ZAAT protein using *Pichia pastoris*. While the successfulness of the methods implemented during this thesis work remains to be determined, they did provide foundational information that can be used in future endeavors. As is, several modifications and optimizations would need to occur before this methodology could be deemed profitable. Suggestions of what these might include are discussed throughout the thesis.

CHAPTER 3:

EXPRESSION AND PURIFICATION OF UNTAGGED α1-ANTITRYPSIN PROTEINS

3.1 Introduction

As stated in Chapter 2, a paper was published in 2009 by Levina *et al* that showed the Z variant of α1-antitrypsin (AAT) being successfully produced recombinantly using the methylotrophic yeast, *Pichia pastoris* (100). Following the publication of this research, it was decided that this procedure would be adopted in hopes of producing active M (wild-type) and Z (mutant) forms of AAT. Unlike the paper however, which added a Polyhistidine-tag to the N-terminal of the proteins, our proteins would lack any sort of protein tag since it was feared they might interfere with our downstream applications.

Additional alterations were made in terms of vector type, clone selection, as well as induction and harvesting methods. This chapter will cover the different techniques used in our attempt to produce untagged, properly folded, active AAT recombinant protein using two strains of *Pichia pastoris*. The implementation of these techniques helped to lay the groundwork for future attempts within our lab.

3.2 DNA preparation

Before work could begin in *Pichia pastoris*, cDNA was first synthesized for both the wild-type and mutant forms of α1-Antitrypsin and then ligated into a vector that was suitable for expression in *P. pastoris*. This was conducted through a series of experiments as described below.

a) AAT DNA Constructs and Vectors

Gene constructs for both the wild-type (MAAT) and mutant (ZAAT) forms of α1-Antitrypsin were synthesized from the company BlueHeron Biotech (Table 3.1). The DNA used was based on the FASTA sequence published on UniProt (Accession #P01009, http://www.uniprot.org/) and was codon-optimized for expression in *Pichia pastoris*. The sequences were flanked by BamH I and EcoR I restriction sites to aid with integration into the pPic3.5K vector (Invitrogen) used for *P. pastoris* expression. It is likely that ZAAT misfolds in the endoplasmic reticulum (ER) where it is targeted by the ER-associated degradation (ERAD) pathway (5). For this reason, the signal peptide sequence (amino acids 1-24) was removed from our constructs and our proteins were produced in the cyotsol instead of being directed to the ER. Both the cytosol and the ER have chaperones present to assist with folding, therefore no complications were expected as result of this change in location (100). Lastly, the pPic3.5K vector is a transcription fusion vector, which means it does not contain an AUG initiation codon for translation of the open reading frame within the expressed insert (78). Therefore the Kozak consensus sequence was added to the N-terminal of both sequences to ensure that proper translation initiation of our genes occurred in *P. pastoris*.

Table 3.1. Protein Sequences for MAAT and ZAAT. Sequences are based on cDNA generated from Blue Heron Biotech. Restriction enzymes sites are depicted in blue, Kozak sequence for pPic 3.5K in green and Z mutation (E342K) shown in red. Both proteins are predicted to have a molecular weight of approximately 45kDa and an isolectric point of 5.31 (www.expasy.org).

	Protein Sequence
MAAT	**GSTM**EDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFSPVSIATAFAMLSLG TKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYH SEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFEVKDTEEEDFH VDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIITKFLENE DRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNGADLSGVTEEAPLKLSKAVHKAVLTIDEKGTEAAGA MFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK**-EF**
ZAAT	**GSTM**EDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFSPVSIATAFAMLSLG TKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYH SEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFEVKDTEEEDFH VDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIITKFLENE DRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNGADLSGVTEEAPLKLSKAVHKAVLTID**K**KGTEAAGA MFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK**-EF**

Both constructs were synthesized into the Blue Heron pUCminusMCS vector, a 3.2kb plasmid that lacks a multiple cloning site and carries an ampicillin-resistance gene (Amp^R, Figure 3.1). The Amp^R gene can be used to help with selection, replication and maintenance in *E. coli*. However, since pUCminusMCS is not compatible for expression in *P. pastoris*, it was necessary to first insert the DNA into *E. coli* so that it could be easily manipulated and then transferred to the pPic3.5K vector. The pPic3.5K vector is a 9kb vector that expresses genes intracellularly with multi-copy integrations (Figure 3.2). Using this vector, target genes can be integrated into the *P. pastoris AOX1* locus or the *His4* locus (Figure 3.3). pPic3.5K also contains a bacterial Kanamycin-resistance gene (Kan^R from Tn*903*) that confers resistance to the antibiotic G418 in *P. pastoris* and kanamycin-resistance in *E. coli* (101,102). The Kan^R gene can be used to screen for multicopy inserts by testing for increased G418 resistance following transformation. This is different from the pHILD2 vector used in the Levina *et al* paper that could only screen for transformants using the *HIS4* marker. In addition, the two vectors differ in the number of unique sites available for gene ligation. The pHILD2 contains only one unique EcoR I site for blunt-end cloning (103), whereas our pPic3.5K vector has multiple unique cloning sites available for directional cloning (102).

b) Subcloning of AAT cDNA into pPic3.5K

In order to transfer MAAT and ZAAT DNA into the appropriate vector, several steps had to first be taken. These steps included bacterial transformation, plasmid isolation, restriction enzyme digestion and finally ligation. Techniques used for each of these steps are discussed below.

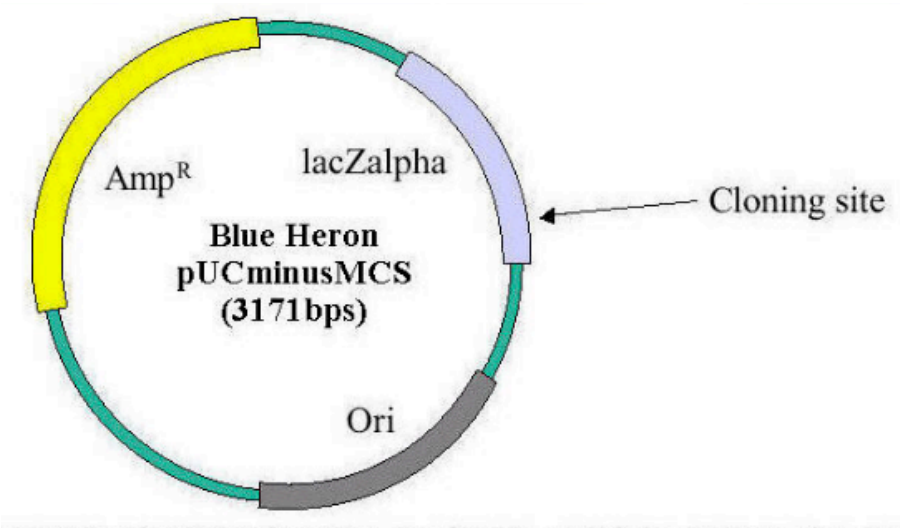


Figure 3.1 Blue Heron pUCminusMCS Vector used for DNA synthesis. (Taken from Blue Heron (104)).

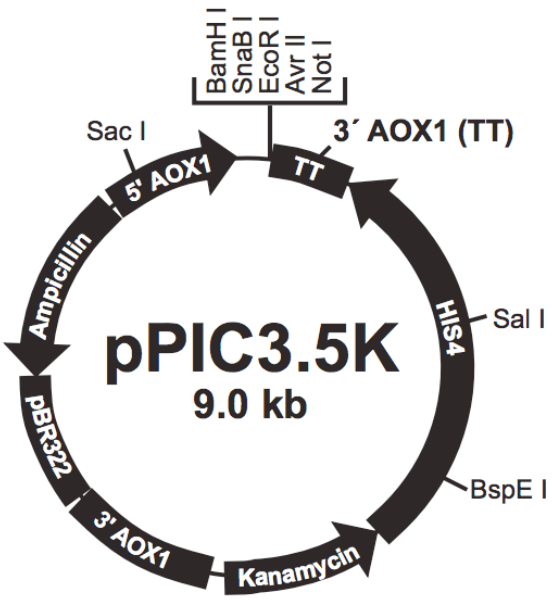


Figure 3.2 pPic3.5K Vector Map used for expression in *Pichia pastoris*. (Taken from Invitrogen (102)).

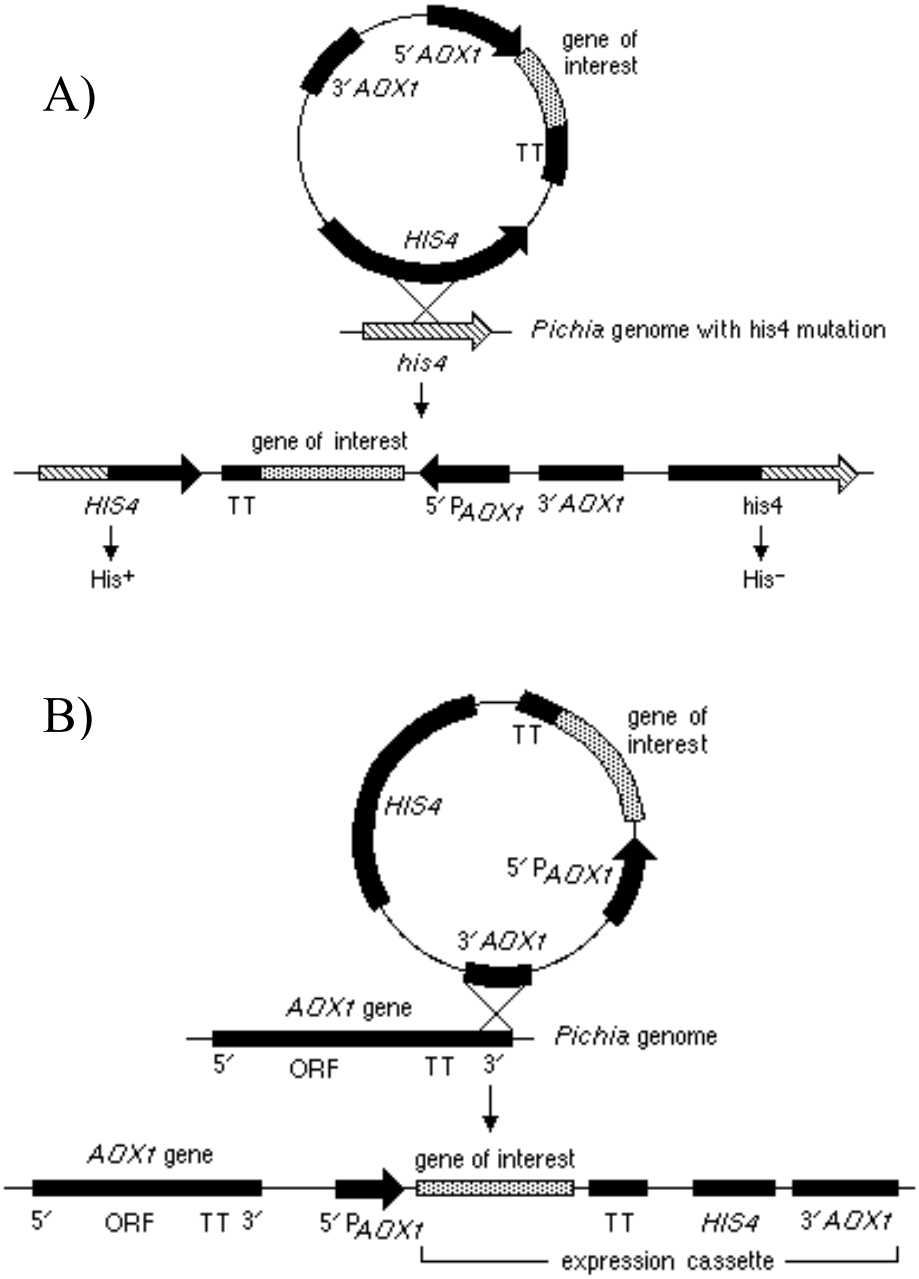


Figure 3.3 Gene Insertion in *Pichia pastoris*. Genes of interest can be inserted either at the (A) *HIS4* or (B) the *AOX1* locus through homologous recombination. Abbreviations: ORF, open reading frame; TT, transcriptional termination sequence (78).

Bacterial transformation was used as a way to amplify the amount of available cDNA: every time the transformed bacteria divided, the plasmid was also replicated (105). Once enough bacteria cells were produced, the cDNA was harvested and then used in the other subcloning steps. Both the MAAT and ZAAT constructs were transformed into *E. coli* using Max Efficiency® DH5α™ Competent cells (Invitrogen), with a stock pUC19 solution (0.01μg/mL) used as a control. To summarize, competent cells that had been thawed on wet ice were aliquoted into pre-chilled 17 X 100mm polypropylene tubes (Falcon® 2059) at 100μL per tube. Approximately 100ng of DNA (MAAT or ZAAT) or 50pg of pUC19 were added to the cells and mixed by gently tapping the tube. Cells were incubated on ice for 30 minutes and then heat-shocked in a water bath at 42°C for 45 seconds. After a 2-minute incubation on ice, 0.9mL of room temperature Super Optimal broth with Catabolite repression (SOC, Invitrogen) was added to each tube and then shaken at 225rpm (37°C) for 1 hour (106).

Although DH5α is naturally susceptible to Ampicillin, both pUC19 and pUCminusMCS contain the Amp^R gene, which confers resistance to this antibiotic. Therefore once transformation was complete and the cells had recovered in the SOC media, they were streaked onto Luria Bertani (LB) agar plates with 100μg/mL Ampicillin. This ensured that only properly transformed cells would be able to survive on the LB plates. Prior to streaking, cells were diluted with SOC media at 1:100 (control) or 1:4, 1:10, and 1:20 (MAAT and ZAAT). Plates were then inverted and incubated overnight at 37°C.

Viable colonies were picked the next day and used to inoculate larger cultures in LB broth overnight. DNA was then isolated from these cultures using either the S.N.A.P. Miniprep Kit (<5mL of culture) or the S.N.A.P. Midiprep (10-100mL of culture) Kit, as per manufacture's instructions (both of which were from Invitrogen).

Briefly, the bacterial cells were first pelleted and then resuspended in resuspension buffer containing RNase A (50mM Tris-HCl, pH 8.0, 10mM EDTA). Next the cells were lysed with Lysis Buffer (0.2M NaOH, 1% SDS) and then protein and genomic DNA from the cells was precipitated using a precipitation salt (3M Potassium acetate, pH 5.2). The lysate/precipitate was then filtered through a filtering column ("Column A"), to which Binding Buffer (7.5M Guanidine-HCl) was added to the flowthrough. This flowthrough was then applied to the binding column ("Column B"), allowing the DNA to bind to the column and everything else to elute into the flowthrough. The bound plasmid was then washed, first with 5M Guanidine-HCl, 50mM MOPS, pH 7.0 and then with 400mM NaCl in 95% ethanol. After the washing was complete, the resin was then dried by centrifugation and the DNA was eluted with sterile Molecular Biology Grade water (107).

Once the DNA was isolated, restriction enzyme digestion was used to verify that our DNA had been properly replicated and isolated from *E. coli.* In addition, restriction enzyme digestion also allowed us to isolate MAAT and ZAAT cDNA from the pUCminusMCS vector so that they could be transferred into the pPic3.5K vector.

Restriction enzyme digestion was performed for 1 hour at 37°C, and involved treating the samples with either the BamH I restriction enzyme by itself (Figure 3.4A, Lanes 2 and 4), BamH I with the EcoR I restriction enzyme (Figure 3.4A, Lanes 3 and

5), or no enzymes at all (not shown). BamH I and EcoR I enzymes were both obtained from Invitrogen. The products from the digest were separated by agarose gel electrophoresis (1% agarose gel), visualized by ethidium bromide staining and then photographed under UV light.

As shown in Figure 3.4A Lanes 2 and 4, a single band (~5kb) was observed after digest with BamH I alone. This indicated that the DNA had been linearized but was still fully integrated into the pUCminusMCS vector. Digestion with both EcoR I and Bam H I resulted in two bands being separated via electrophoresis (Figure 3.4A, Lanes 3 and 5): the top band (~3kb) being the pUCminusMCS vector and the bottom band (~1.2kb) representing our gene of interest.

After the DNA was isolated through restriction enzyme digestion and electrophoresed on the agarose gel, the gel piece containing MAAT or ZAAT DNA was excised and purified using the S.N.A.P. Gel Purification Kit (Invitrogen), as per manufacture's instructions.

The newly isolated DNA was then ligated into the pPic3.5K vector following a modified version of the T4 DNA Ligase (Invitrogen) protocol. Briefly the ligation reaction consisted of five different treatments (Table 3.2) with a 3-to-1 insert-to-vector ratio and was carried out overnight at 4°C. Products were then transformed into *E. coli*, and plated onto LB agar plates with 100µg/mL Ampicillin as described previously.

Figure 3.4 α_1-Antitrypsin DNA manipulations in *E. coli*. Restriction Enzyme Digestion of MAAT and ZAAT cDNA in the (A) pUCminusMCS vector and in the (B) pPic35k vector, where pMAAT and pZAAT represents ligation into the pPic3.5K vector. DNA linearization with (C) Sal I and (D) Sac I. Gels were made with 1% Agarose, stained with ethidium bromide and photographed under UV light. Hyperladder I (Bioline) was used for all DNA gels.

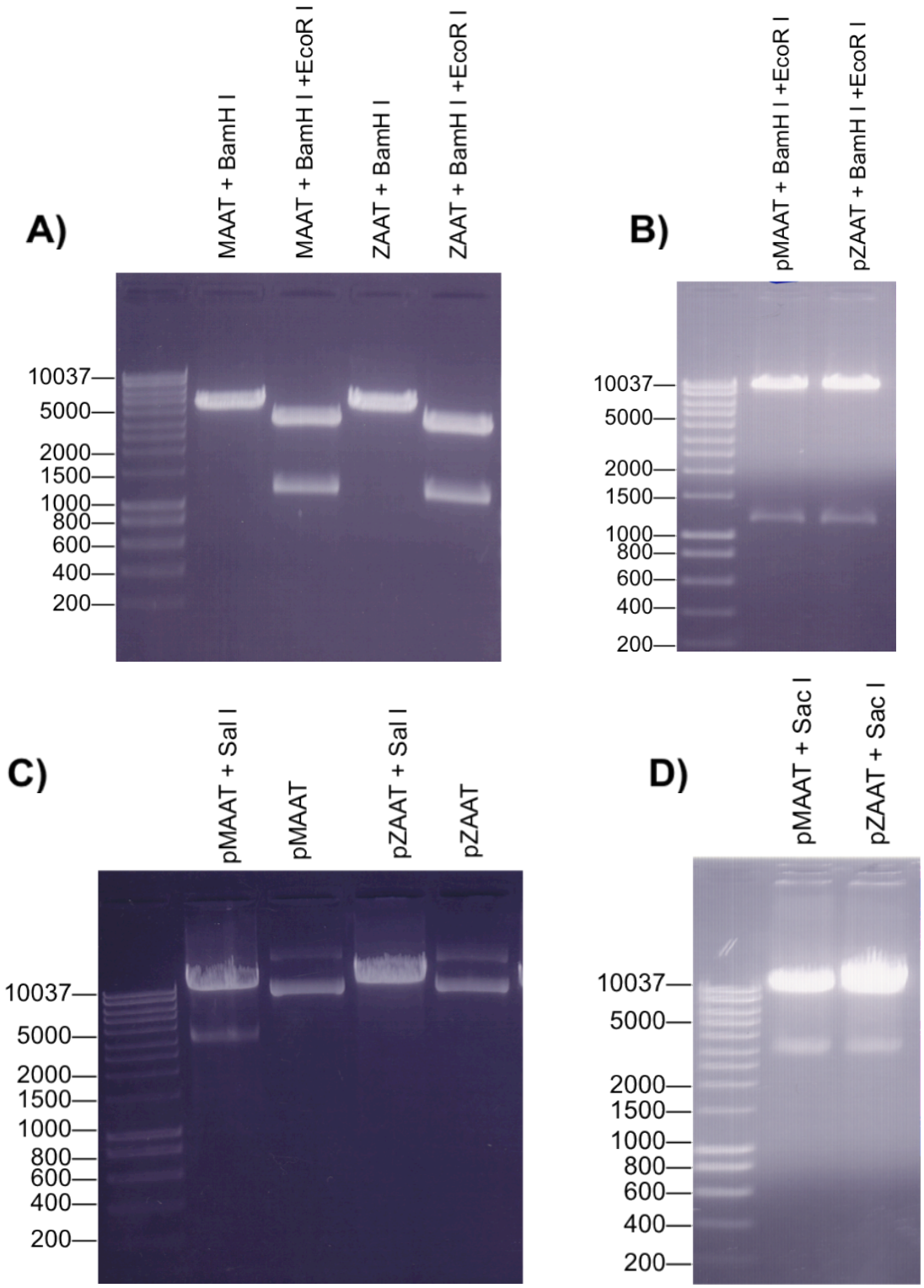


Figure 3.4 continued.

Table 3.2 Ligation Reaction Treatments. Treatments A and B were used as controls to verify the necessity of both DNA's for the reaction. Treatements D and E acted as controls for DNA ligase activity. Set-up was based on protocol in *Molecular cloning: a laboratory manual* (108).

Treatment	DNA Present	DNA Ligase?
A	pPic3.5K	Yes
B	MAAT/ZAAT	Yes
C	pPic3.5K + MAAT/ZAAT	Yes
D	pPic3.5K	No
E	MAAT/ZAAT	No

As expected, only plates that had treatment C (pPic3.5K + MAAT/ZAAT DNA in the presence of the DNA ligase) streaked upon them showed any substantial *E. coli* growth. Colonies from these plates were isolated and grown in an overnight culture of LB Broth plus 100µg/mL Ampicillin. DNA from the overnight culture was then extracted using the S.N.A.P.™ MidiPrep Kit and digested with restriction enzymes as described previously. The double digestion with EcoR I and BamH I (Figure 3.4B) resulted once more in two bands being identified via electrophoresis: the top band (~9kb) indicating the pPic3.5K vector and the bottom band signifying our gene of interests (~1.2kb).

Once the success of the ligation was confirmed, the DNA was then linearized overnight. This was done initially with the restriction enzyme Sac I (Promega), which would allow the DNA to insert at the *AOX1* locus (Figure 3.4C, Lanes 2 and 4), and then later with Sal I (Promega), which inserts the DNA at *HIS4 (*Figure 3.4D). Both forms of digestions resulted in bands at ~10kb after separation by electrophoresis. DNA clean-up was performed afterwards using the Wizard® SV Gel and PCR clean-up System (Promega), as per manufacture's instructions

3.3 Expression in Pichia pastoris strains X33 and SMD1163

Initially two different strains of *P. pastoris* were used to express our recombinant vectors: X33 and SMD1163. Both of these strains contain the *AOX1* promoter, which allows for rapid growth when methanol is utilized as the sole carbon source (methanol utilization plus or Mut^+ phenotype) (78-80).

X33 is a wild-type strain that is a derivative of GS115, one of the most commonly used strains of *P. pastoris*. Unlike GS115, however, the X33 strain has no defects in histidinol dehydogenase activity encoded by the *HIS4* gene (His^+ phenotype) (79,103,109). Due to the His^+ phenotype of X33, selection in this strain is only possible through antibiotic resistance conferred by the vectors that carry the gene of interest (103). In our case, this means that selection for X33 clones was carried out based on its resistance the antibiotic G418.

SMD1163 (*his4, pep4, prb1*), on the other hand, is a His^- and protease-deficient strain that lacks both the *PEP4* and *PRB1* genes. The *PEP4* gene codes for proteinase A, a vascular aspartyl protease that is essential for the activation of carboxypeptidase Y, proteinase B, as well as other vacuolar proteases. Proteinase B can also be produced as a result of *PRB1* gene translation. Therefore, by eliminating these two genes, the activity levels of the three main proteases in *P. pastoris* (proteinase A, carboxypeptidase Y, and proteinase B) are drastically reduced or even eliminated. This can create a more suitable environment for expression of certain heterologous proteins that are prone to protease degradation (78,80). Selection of SMD1163 clones can be carried out either through selecting for His^+ clones, selecting by its resistance to antibiotics, or a combination of both, depending on the vector introduced into SMD1163.

a) Transformation into *Pichia pastoris strains*

While there are several different methods that can be used to transform foreign DNA into *P. pastoris,* spheroplasting and electroporation tend to provide the highest transformation efficiencies for most researchers (10^3 to 10^4 transformants/μg DNA)

(110,111). However, spheroplasting, which involves the partial digestion of the yeast cell wall by Zmolyase (110), can be a very tedious and time-consuming process. Therefore, electroporation was implemented for the transformation of our *P. pastoris* strains. Electroporation and initial selection of *P. pastoris* clones were performed as described below with the help of Dr. Thomas Masi from the Biochemistry, Cellular and Molecular Biology Department at the University of Tennessee in Knoxville, Tennessee.

Prior to electroporation, yeast cells were chemically pretreated with Lithium acetate (LiAC) and dithiothreitol (DTT) according to the procedure outlined by Wu and Letchworth in 2004 (111). The combination of LiAC with a reducing agent, such as DTT, as a chemical pretreatment has been shown to increase the permeability of yeast cells though an unknown mechanism. This results in an increase in transformation efficiency by approximately 150-fold (111). Using this method, the cells ($8x10^8$ cells needed per transformation) were made competent by first washing them with sterile water and then suspending them in 100mM LiAC, 10mM DTT, 0.6 M Sorbitol, and 10mM Tris-HCl. Making sure to keep the cells cold, they were then washed with 1M Sorbitol plus 10mM Tris (pH 7.5) several times at increasing smaller volumes until there was approximately 10^{10} cells/mL in 80µL. Finally the cells were washed once more with sterile water in order to remove all salts present (111). Sorbitol was used throughout the process in order to stabilize osmotic pressure and to support the integrity of the yeast membrane (112).

Once the yeast cells had been prepared, they were then mixed with 3ng of DNA suspended in 1µL of water, and transferred into an ice-cold sterile electroporation cuvette with a 0.2 cm electrode gap width. Next, they were incubated on ice for 5

minutes and then placed into the MicroPulser Electroporator (BioRad). The lower temperature prior to electroporation helps to decrease heat damage to the cells and can increase transformation efficiency. The electroporation was conducted using the "Pic" setting on the machine, with an average time around 5.8ms and voltage of 2.0KV.

Following electroporation, Sorbitol was used once more as osmotic support during cell recovery. Cells were resuspended in Sorbitol and incubated at 30°C for one hour. YEPD (Yeast extract, Peptone, and Dextrose) was then added to the cells so that the Sorbitol and YEPD were in a 1:1 ratio and the cells were incubated at 30°C for another hour. After the second one-hour incubation was complete, 1mL of the transformed cells was spread onto YEPD agar plates with 0.25mg/mL of G418, as recommended by Invitrogen (102). The plates were then placed in an incubator at 30°C and allowed to grow for 3-5 days. This temperature was used because it has been found by researchers to be optimal for protein production in *Pichia pastoris*. Conversely a temperature of 32°C has been shown to inhibit all protein expression (113). Therefore for all of our growing conditions with *P. pastoris,* it was important to maintain the temperature at 30°C.

b) In vivo Screening of Multiple Inserts

It has been previously shown that multiple-copy integrations of recombinant genes in *P. pastoris* can result in increased expression of the desired gene (101,114,115). Such integrations occur spontaneously at a low but detectable frequency—between 1% and 10% in all His^+ transformants. The pPic3.5K vector can be used to generate and isolate multi-copy insertions through *in vivo* methods (78,102).

Once the transformed clones had an adequate amount of time to recover and grow, they were screened for multiple-copy integrations in one of three ways. The first involved screening for G418 resistance, the second focused on selecting His^+ clones and the third identified His^+ and G418 resistant clones.

Screening for G418 resistance

As described before, the pPic3.5K vector contains the bacterial kanamycin-resistance gene that confers resistance to high levels of G418 to clones that contain multiple copies of the vector (102). However, as was quickly discovered during the initial attempts of growing the two strains, only X33 clones were able to survive on the YEPD plus 0.25mg/mL G418 plates following electroporation. This indicated that X33 and SMD1163 might differ in their sensitivity to the antibiotic. The literature supported this finding and revealed that the protease-deficient strain, SMD1163, had a higher sensitivity to G418 when compared to other strains like X33 or GS115 (116). It also indicated that SMD1163 clones that contained just one gene copy were only able to survive on plates supplemented with 0.025mg/mL G418 but clones with multi-copy insertions were able to tolerate 0.05mg/mL or higher (116). Therefore for future transformations we used 0.05mg/mL G418 as the initial cut-off for the SMD1163 clones following electroporation.

While the issue between G418 and the SMD1163 clones was being resolved, experiments with X33 clones began to move forward into the screening process. Large-scale antibiotic-resistance screening of X33 clones was conducted by first filling a 96-well plate with 100µL of sterile water per well. Using sterilized toothpicks, single colonies were then isolated and mixed with the water in a single well. Next, a multi-

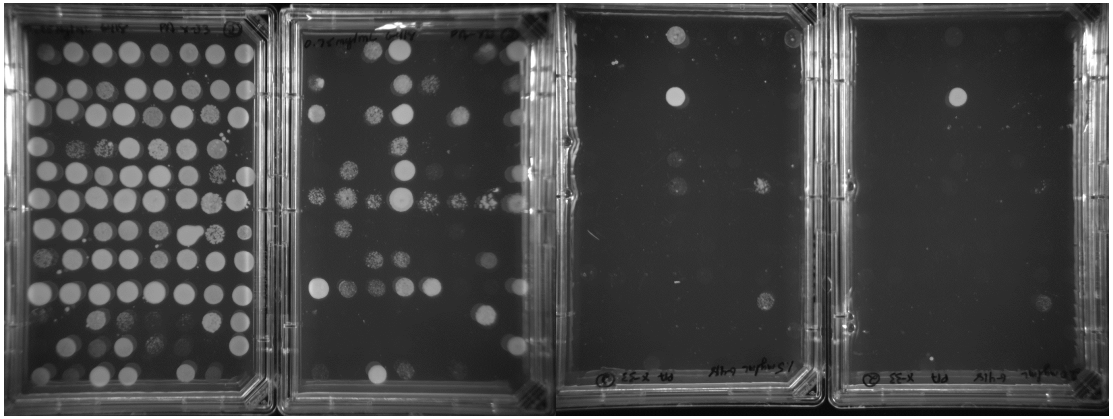
channel pipette was used to dispense 5µL of the yeast/water mixture onto YEPD agar plates that contained concentrations of G418 at 0.25, 0.75, 1.5 or 2.0mg/mL. The plates were then placed in an incubator at 30°C and allowed to grow for 3-5 days. Out of the almost four hundred X33 colonies screened, 9 were able to survive on the 2.0mg/ml G418 plates (6 ZAAT and 3 MAAT, Figure 3.5). These colonies were then isolated and grown in larger volumes for protein extraction (Section 3.5a).

Screening for His^+ clones

As described previously, the SMD1163 lacks the *HIS4* gene and consequently is unable to grow on media that lacks histidine prior to transformation with an appropriate vector. Therefore, the identification of His^+ colonies is commonly one of the first steps in the multi-copy selection process for *his4* mutant strains, like SMD1163 (78). Following electroporation, SMD1163 transformants were streaked directly onto plates that lacked histidine [His (-) plates] and grown for 3-5 days. This resulted in only yeast cells that had been correctly transformed with the pPic3.5K vector being able to survive.

Once the transformants were able to reach sufficient growth, 48 SMD1163 clones (24 per protein) were isolated and re-streaked onto new His (-) plates and grown for an additional 3-5 days. This method, however, resulted in abundance of His^+ clones. While the SMD1163 strain is supposed to have a His^- phenotype, there is a possibility that some endogenous expression may still exist. Therefore 3-amino 1,2,4 triazole (3-AT) was added to the second round of His (-) plates in an attempt to lower the amount of His^+ colonies present on the plates.

A)



B)

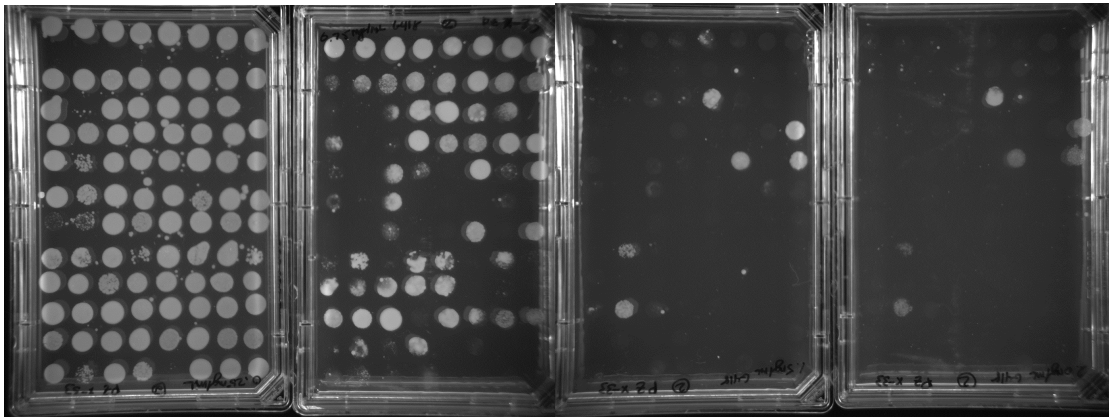


Figure 3.5 Large-Scale Screening with G418. MAAT (A) and ZAAT (B) X33 clones grown on 0.25, 0.75, 1.5 and 2.0 mg/mL G418 (Left to Right respectively).

This was done by spreading 0, 2, 3, 5, 10, 20 or 25mM 3-AT onto the YEPD agar plates using glass beads, and allowing the plates to dry before streaking the colonies onto them (8 colonies per plate). While there was some reduced growth on plates with 10mM, 20mM and 25mM 3-AT, the difference in growth was not significant. Therefore this step was eliminated in future screenings. Overall, colonies that survived both rounds of streaking on His (-) plates were considered His^+ and were used either for protein extraction (Section 3.5b) or for subsequent antibiotic testing.

Selecting for G418 resistant, His^+ clones

SMD1163 transformants were also screened for multi-copy integrations by selecting for His^+, G418 resistant clones as described in the literature (78,81). This was accomplished by first plating the transformants onto His(-) plates directly after electroporation and allowing them to grow for 3-5 days. After the colonies displayed sufficient growth, they were then isolated and mixed with 100µL of sterile water in a 96-well plate. The yeast/water mixture was then spotted onto YEPD agar plates containing G418 at 0.05, 0.10, 0.15 or 0.20mg/mL. The plates were then incubated at 30ºC for 3-5 days.

Using this method, approximately 400 His^+ SMD1163 transformants were isolated and screened for G418 resistance. Thirteen of these, 6 MAAT and 7 ZAAT, were able to survive on plates containing 0.15mg/mL G418 and therefore were chosen for protein extraction (Section 3.5).

3.4 Protein Gel Electrophoresis and Western Blot Analysis

Following the identification of clones that contained multi-copy integrations of our genes, the clones were then induced to promote protein expression. Our proteins of interests were then extracted (Section 3.5) and purified (Section 3.6) through a series of different methods and techniques. During this process, it was important to verify that our protein was still present in the samples being tested. Therefore throughout these experiments, a small portion of our samples were removed and analyzed by protein gel electrophoresis and Western blot analysis. This section will discuss the methods used in order to perform such analyses, the results of which are discussed in their appropriate sections.

a) Protein Gel Electrophoresis

Electrophoresis, in general, is a common method of using an electric field to separate macromolecules. In SDS-polyacrylamide gel electrophoresis (SDS-PAGE) specifically, a polyacrylamide gel is used as a support medium and sodium dodecyl sulfate (SDS) is used to linearize the proteins by denaturing them. Protein separation by SDS-PAGE can be used to estimate the relative molecular mass of proteins as well as the relative abundance, distribution and purity of proteins within a sample (117). SDS-PAGE experiments were conducted using NuPage® Novex ® 4-12% Bis-Tris Mini Gels (Invitrogen) and performed under denaturing conditions (according to manufacture's instructions). Gels were then analyzed by 1) Coomassie stain 2) Silver stain and/or 3) by Western blot analysis. Human α1-Antitrypsin (MP Biomedicals Inc.), which has a molecular weight of 52kDa, was used as a positive control for these analyses.

Coomassie Staining

Coomassie Brilliant Blue staining is a fast, inexpensive and fairly reliable quantitative staining method that has a sensitivity of approximately 100ng (118). Coomassie staining was conducted by covering the gel with the dye (50% methanol, 10% acetic acid, 1% Brillant Blue R-250) and microwaving the gel for 10 seconds. The stain was then decanted and de-stain (50% methanol, 10% acetic acid) was added to the gel and allowed to incubate on a rotating platform for 30 minutes. The de-stain was then replaced with Mili-Q water and incubated on the rotating platform once more until bands reached their desired clarity.

Silver Staining

Silver staining is another popular method of detecting proteins after protein gel electrophoresis. Some advantages of silver staining include, but are not limited to, sensitivity in the low nanogram range and compatibility with downstream processing such as mass spectrometry analysis (118,119).

Silver staining was performed on a rotating platform using the SilverSNAP® Stain Kit II (Pierce) according to the manufacture's instructions. Briefly this involved washing the gel for two 5 minute time periods in ultrapure water and then fixing the gel in a 30% ethanol: 10% acetic acid solution for two 15 minutes incubations. The gel was then washed twice in 10% ethanol for 5 minutes each, followed by two washings with ultrapure water, once again, 5 minutes long each. A sensitizer working solution (50µL sensitizer with 25mL water) was prepared and added to the gel for 1 minute. Following, two 1 minute washes with ultrapure water, a stain working solution (0.5mL Enhancer with 25mL Stain) was then added to the gel for 30 minutes. The gel was then washed

twice with ultrapure water for 20 seconds and then developed with the developer working solution (0.5mL Enhancer with 25mL Developer) until bands appeared. The reaction was then stopped with 5% acetic acid for 10 minutes (120).

b) Western Blot Analysis

Western blotting, also known as immunoblotting, identifies target proteins among a sample by exploiting antibody specificity to the target protein. Proteins are typically separated first by electrophoresis and then transferred onto membranes (usually Polyvinylidene fluoride [PVDF] or nitrocellulose). The membrane, referred to as a blot, is then covered with a primary antibody specific for the target protein and then with a secondary antibody (anti-immunoglobulin) that is labeled with enzymes, radioisoptoes or other marker compounds used for detection (121).

Following separation by SDS-PAGE, proteins were transferred onto 0.45μm nitrocellulose membranes (Invitrogen) as outlined in the NuPAGE® Technical Guide (122). Proteins were transferred at 30V for 1 hour using 1X NuPAGE® Transfer Buffer with 10% methanol in an XCell II™ Blot Module (Invitrogen). Once the proteins were successfully transferred onto the nitrocellulose membrane, they were ready for Western blot analysis, which was performed using the SNAP i.d. Protein Detection System by Millipore.

The procedure was performed as indicated in manufacture's manual (123) using Tris-buffered Saline (TBS) with 0.1% Tween 20 as a wash buffer (TBS-T). Tween 20 was used in the wash buffer to help reduce solution surface tensions and to ensure even antibody distribution during incubation (123). The blot blocking solution, which was

added to help reduce non-specific antibody binding, consisted of either 1% Bovine Serum Albumin (BSA) or 0.5% Casein dissolved in TBS-T. The primary antibody was a mouse Monoclonal Anti-α1-Antitrypsin antibody (Abcam) at 1:4000 dilution and the secondary antibody was a Peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG (H+L) antibody (Jackson ImmunoResearch) at 1:10000 dilution. Detection of the membrane was done using Santa Cruz Western blotting Luminol Reagent and exposing the membrane to the reagent for 1 minute in a dark drawer. The membrane was then exposed to autoradiography film and developed using Kodak supplies.

3.5 Extraction of MAAT and ZAAT proteins from P. pastoris strains

While the aforementioned issues regarding SMD1163's sensitivity to G418 were being resolved, it was decided that experiments with the X33 clones should continue in order to help identify and optimize conditions prior to large-scale production. One such condition was to determine the optimal time post-induction to harvest.

a) Optimizing Induction Time in X33 clones

According to Levina *et al (100)*, the induction period should be between 17 and 72 hours. Given such a broad range of time, an experiment was conducted in order to narrow down the optimal induction time needed for our proteins of interest.

This was accomplished by selecting 1 MAAT and 3 ZAAT clones identified by the large-scale antibiotic resistance screening to inoculate 50mL of YEPD overnight at 30°C shaking (225 rpm). The cells were then transferred to a 50mL conical centrifuge tube, spun down at 4000xg for 5 minutes, washed with sterile water, and spun again at

4000xg for 5 minutes. Once pelleted, they were then re-suspended in 50mL of Induction media (1% yeast extract, 2% peptone, 100mM potassium phosphate, pH 6.0, 1.34% Yeast Nitrogen Base with Ammonium Sulfate without Amino Acids, $4x10^{-5}$% biotin and 0.5% methanol), poured into a 250mL flask and grown in a shaker (225rpm, 30°C) for a total of 96-hours.

After every 24-hours of growth, 10mL of culture was removed from the flask and transferred to a 15mL conical centrifuge tube. Meanwhile, 5% methanol was added to the remaining culture at 1/200 of its volume and then placed back into the shaker (225rpm, 30°C). The 10mL that had been removed was pelleted at 4000xg for 5 minutes, washed with sterile water, and pelleted again at 4000xg for 5 minutes. The pellet was then weighed and stored at -20°C until all samples were ready for lysis.

X33 yeast cells were lysed by mechanically breaking the cells directly into lithium dodecyl sulfate (LDS) sample buffer. This allowed us to load the samples directly onto a SDS-PAGE gel for analysis. Briefly, for every gram of wet cell paste, 1 gram of 0.5mm glass beads, 1mL of NuPAGE® LDS Sample Buffer (Invitrogen), and 50µL of Protease inhibitor Cocktail Set III, EDTA-Free (EMD4Biosciences) were added to the frozen cell paste. DTT was added to the LDS buffer so that the final concentration of DTT was 100mM. The cells were then vortexed twice for 2.5 minutes each, and spun down in a micro-centrifuge at max speed for 1 minute. The supernatant was then removed and saved for further analysis via protein gel electrophoresis and Western blot analysis (Section 3.4b). This allowed us to identify expression levels for each time point.

An initial look at the MAAT samples displayed a faint band at ~55kDa for the 24-hour induction time point but not for the other time points (Figure 3.6A). However when these same samples were re-analyzed alongside those from the three ZAAT samples, no expression was detected at all (Figure 3.6B). Given the band seen in the original Western blot (Figure 3.6A, lane 3), it was decided at the time of the study that harvesting should occur after a 24-hour post-induction time period.

Closer inspection of this Western blot later on, suggested that the band could be the result of a spillover from the positive control since it is at the same exact molecular weight. This would indicate that the monoclonal antibody used at this time was ineffective at detecting our protein. One explanation could be that the antibody, which was raised against Human AAT, might have not been able to detect our proteins, which were synthesized in *P. pastoris,* due differences in glycosylation or protein production within the organisms (discussed further in Section 3.5b).

Another explanation could be that perhaps our genes had not been properly integrated into the *P. pastoris* genome. Although not performed at the time, this could have been confirmed via PCR analysis as described in the Multi-copy *Pichia* Expression Manual (110).

Lastly, the lack of detection could be the result of our proteins being degrading by proteases present in the X33 strain. Therefore, in hopes of resolving some of the inconsistency seen in the Western blots, focus was switched at this point from the X33 clones to the protease-deficient SMD1163 clones.

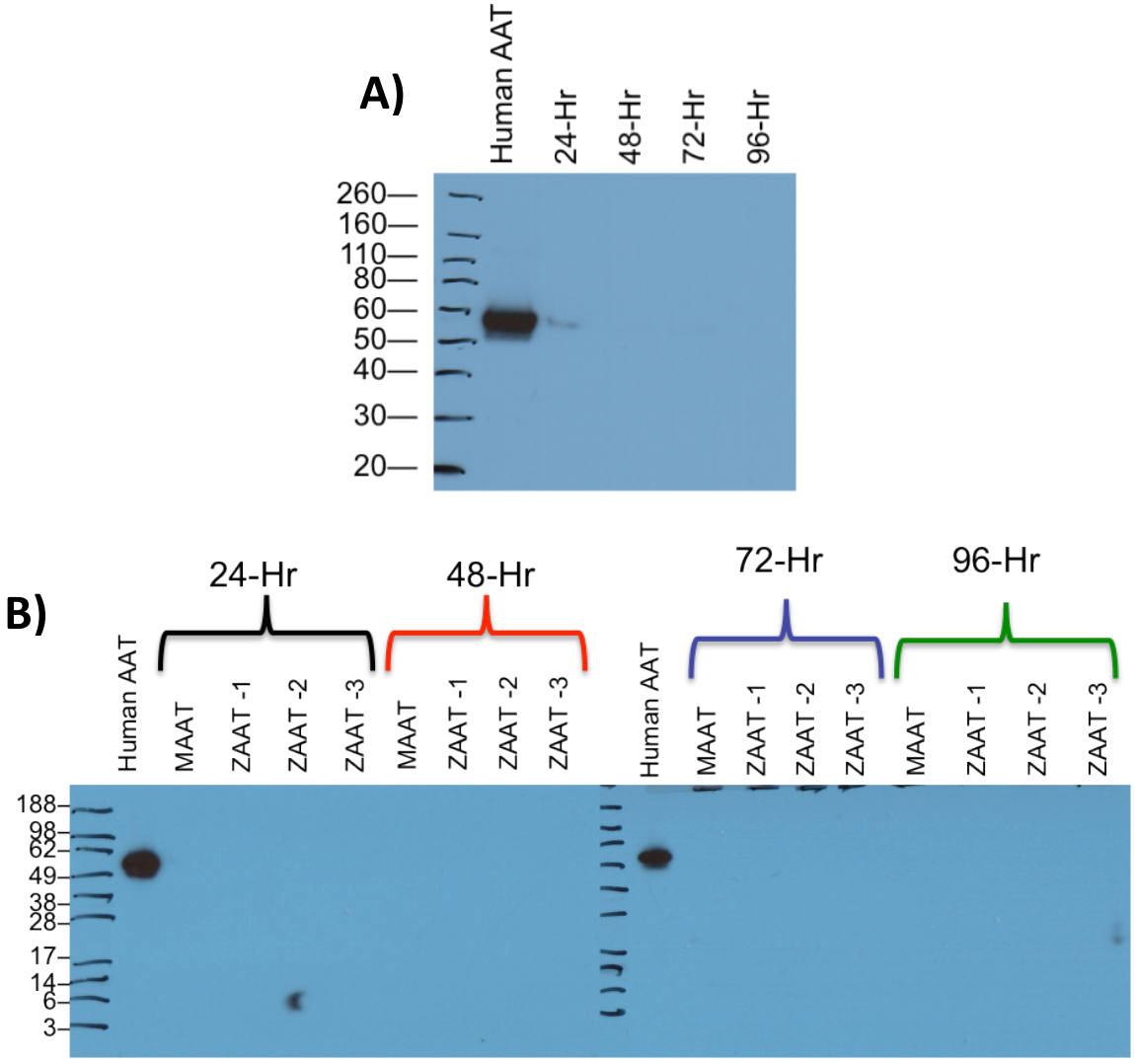


Figure 3.6 Optimizing Induction Time in X33 clones. (A) An initial look at different induction times for MAAT protein production indicated that cells should be harvested 24-hours post induction. (B) However when samples where re-analyzed alongside ZAAT samples, no expression was detected. This later gave rise to the possibility that the band seen 24-hours post induction could have been a result of spillover from the positive control, Human AAT. Nevertheless at the time of the study, it was decided that harvesting would occur after a 24-hour post-induction time period. This was based on the original interpretation of the Western blot seen in (A). A mouse monoclonal α1-Antitrypsin antibody (1:4000 dilution) was used for the detection of AAT proteins.

b) Identification of High-expression SMD1163 clones

Once we were finally able to perform *in vivo* screening for multiple insertions in SMD1163, identified colonies were then isolated and analyzed in order to detect which ones had the highest expression for our proteins of interest. This was performed by first growing the clones in 10mL YEPD overnight in a test tube rotator at 30°C. The following morning, the cells were spun down in a 15mL conical centrifuge tube at 4000xg for 5 minutes and then washed with 10mL of sterile water. The cells were then spun again at 4000xg for 5 minutes and re-suspended in 10mL of Induction media. Once more the cells were placed in a test tube rotator and induced for 24-hours at 30°C. An incubation of 24-hours was used based on our previous induction results with the X33 clones (Figure 3.6A).

After the 24-hour induction period was over, the cells were centrifuged at 4000xg for 5 minutes, and washed with sterile water. The water was then aspirated down to 1mL in which the cells were re-suspended. The mixture was transferred to a 2mL microcentrifuge tube, placed in a microcentrifuge, and spun at max speed for 1 minute. The supernatant was then removed and the pellet was weighed.

In order to determine which clone had the highest expression levels for MAAT or ZAAT, the cells were lysed and then analyzed via protein gel electrophoresis and Western blot analysis. Once more when we used a monoclonal antibody for our Western blots, we observed no expression even after 5 minutes of exposure (Figure 3.7A).

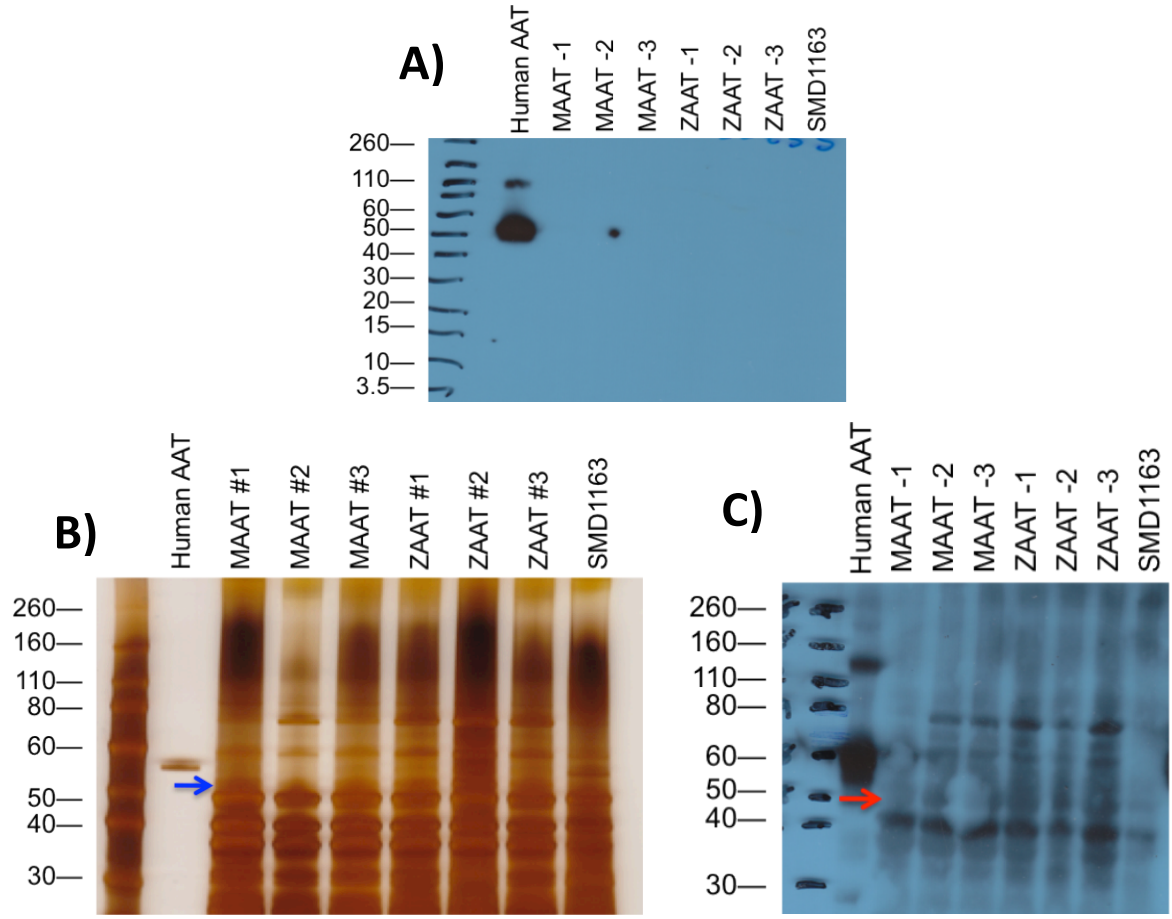


Figure 3.7 Identification of High-Expression SMD1163 Clones. (A) Once more attempts at using a monoclonal antibody were made. However after a 5-minute exposure, protein was still undetectable. (B) Samples were then diluted 1:10 and analyzed via Siler staining. This revealed a ~50kDa band (blue arrow) in each of our samples that was predicted to be our proteins of interests. (C) When the Western blot was repeated, this time using a polyclonal antibody (30 second exposure time), a ~50kDa band (red arrow) was once again observed in each sample. This band was also thought to represent our proteins of interest. Human AAT protein (MP Biomedicals, Inc) served as a positive control for the described experiments.

We next tried concentrating and diluting the samples in order to help increase the resolution of our detection methods. However, when the samples were concentrated 5x by a spin column with a 30kb molecular cut-off, there was still no detection by Western blot and our Silver stain was inconclusive (not shown). On the other hand, when samples were diluted 1:10, we were able to detect a band just slightly below our positive control (around 50kDa) that was believed to be our proteins of interest (Figure 3.7B). Samples were then sent off for mass spectrometry analysis in order to verify this, the results of which indicated that the band present could be our AAT proteins but was not 100% conclusive.

Since the monoclonal antibody proved ineffective in detecting our proteins of interest, we decided to switch to a rabbit polyclonal anti-human α1-Antitrypsin antibody (Sigma) for the primary antibody and a Peroxidase-conjugated AffiniPure Goat Anti-Rabbit IgG (H+L) antibody (Jackson ImmunoResearch) as the secondary. Once the antibody was switched to a polyclonal, we were finally able to see some expression on our Western blots (~50kDa). However this also resulted in increased background noise that could have been the result of non-specific binding (Figure 3.7C). Increasing the TBS-T amount used during washings, as well as increasing its number of applications, helped to resolve this issue. It wasn't until we decided to change lysis methods that were we finally able to determine which clones had the highest expression levels for our proteins of interest.

c) Optimizing Cell Lysis

Although our initial method of lysing the yeast cells directly into LDS buffer was thought to be effective (Figure 3.7B), LDS is an anionic detergent and therefore might interfere with our downstream ion exchange chromatography methods. Ultimately it was only used for quick identification of clones that had the highest concentration of protein expressed.

In hopes of finding a more chromatography-friendly lysis method, a chemical method of lysis was introduced. This method involved adding Y-PER® Yeast Protein Extraction Reagent from Thermo Scientific (2.5µL/mg wet cell paste), DTT (final concentration 100mM), and Protease inhibitor Cocktail Set III, EDTA-Free (50µL/g wet cell paste) to the frozen yeast cells. Next, the cell mixture was agitated at room temperature for 20 minutes and transferred to a 2mL microcentrifuge tube. The tubes were then placed into a microcentrifuge and spun at max speed for 10 minutes. The supernatant was then removed and saved for further analysis (Figure 3.8).

In changing lysis methods it was important to verify that the chemical method was as effective, if not more so, than the physical method. Therefore, a 10mL culture was split into two 5mL aliquots; one that was lysed with LDS and the other was lysed with Y-PER®. The results, as visualized by Western blot (Figure 3.8B) and Comassie stain (Figure 3.8A), indicated that chemical lysis with Y-PER® was in fact a better lysis method then the mechanical one. This can be seen in the fact that both detection methods display bigger bands at ~50kDa for the Y-PER® samples when compared to the LDS samples.

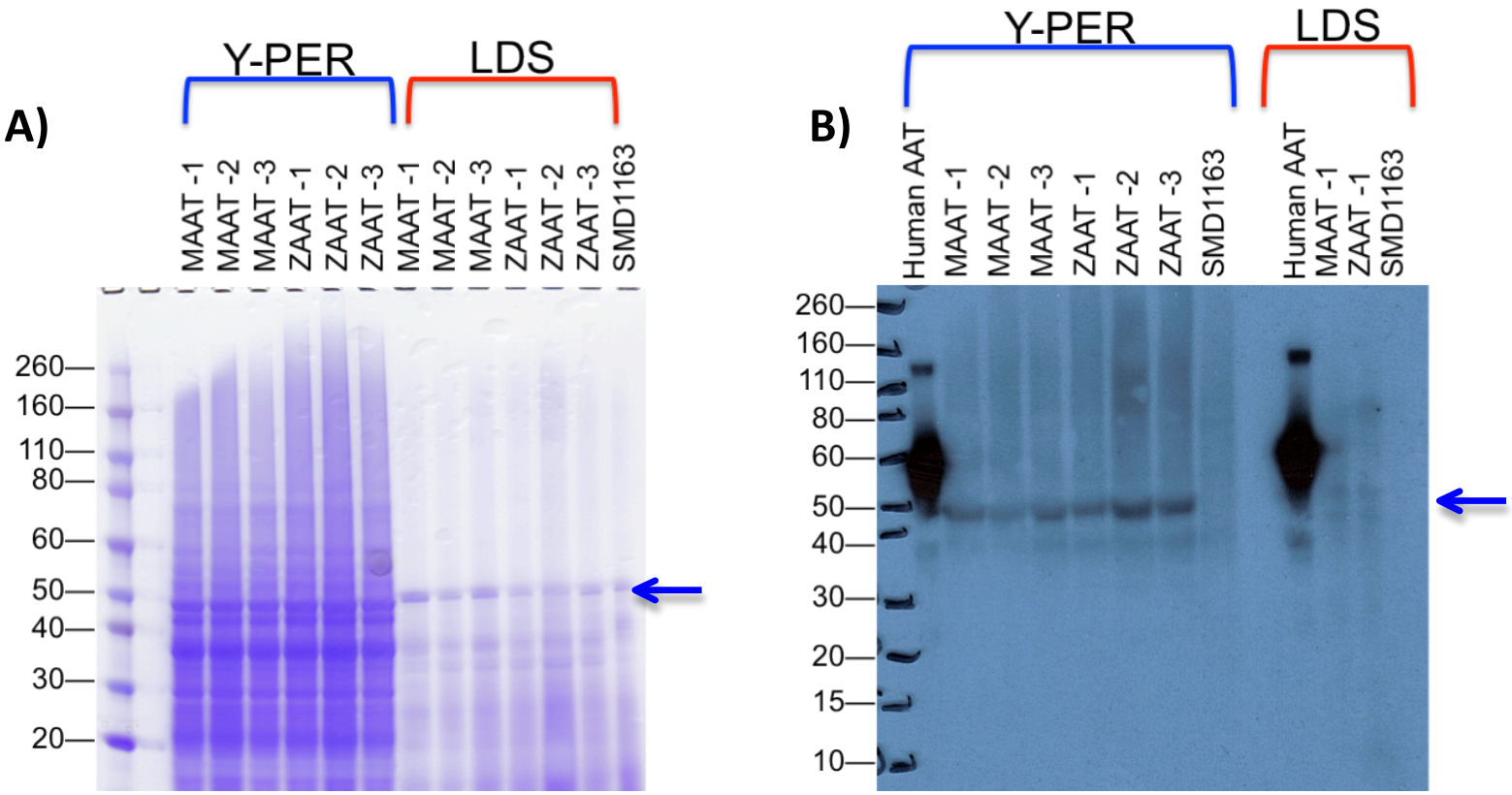


Figure 3.8 Lysis Method Comparisons: (A) Coomassie stain gel and (B)Western blot indicating differences between physical lysis with glass beads and LDS (LDS) and chemical lysis using Yeast Protein Extraction Reagent (Y-PER®). Arrows indicate where AAT proteins were predicted to be located. Polyclonal anti-AAT antibody used in 1:3000 dilutions for Western blot. Human α_1-Antitryspin served as a positive control for experiments.

This is especially true for the Western blot which only showed expression in the Y-PER® samples at ~50kDa but not in the LDS samples (Figure 3.8B). From these results, we were also able to determine that MAAT-1 and ZAAT-2 had the highest protein expression levels of the samples tested. Therefore these two colonies were re-streaked onto YEPD + 0.05mg/mL plates and used for large-scale extraction methods.

d) Large-Scale Induction and Harvesting of SMD1163 clones

Once we were able to determine which clones had the highest expression levels and our conditions were optimized on a small-scale, protein expression was produced on a larger-scale (500-1000mL cultures). First, previously selected MAAT and ZAAT SMD1163 colonies (Figure 3.8) were streaked onto YEPD agar plates with 50µg/mL G418 to verify the viability of clones. Once this was accomplished, they were then grown in 50mL YEPD in a shaker (225rpm) overnight at 30°C. The next day, the overnight culture was transferred to a baffled flask that contained 500-1000mL of YEPD and placed in the shaker (225rpm) to once more grow overnight at 30°C.

The following day, the cultures were split up into 250mL centrifuge containers and spun at 4000xg for 5 minutes. Cells were then washed with sterile water and pelleted down again at 4000xg for 5 minutes. The pelleted cells were then re-suspended in 250mL of Inducing media, transferred to a new sterile baffled flask, and grown for 24-hours in a shaker (225rpm) at 30°C. Once 24-hours had elapsed, the cells were split into 50mL aliquots and centrifuged at 4000xg for 5 minutes. They were then washed with sterile water, spun again at 4000xg for 5 minutes with the resulting pellets weighed and then stored at -20°C until needed for lysis.

3.6 Protein Purification

After large-scale MAAT and ZAAT protein expression was completed, the target proteins were then isolated and purified from the other proteins within the samples as described in this section. In general, proteins can be purified through techniques that separate them based on differences in specific properties such as charge, size, hydrophobicity and ligand specificity (124). Typically protein purification is accomplished through the combination of more than one of these techniques. For our experiments we exploited differences in charge (ion exchange) and size (gel filtration) in order to purify MAAT and ZAAT.

a) Anion Exchange Chromatography

Ion exchange chromatography (IEX) is a widespread protein purification technique that is based on a reversible interaction between charged proteins and an oppositely charged chromatography medium. There are two types of IEX: 1) cation exchange, in which positively charged molecules are attracted to a negatively charged solid support, and 2) anion exchange, where negatively charged molecules are attracted to a positively charged solid support (125,126). Since the isoelectric point of AAT varies from 4.9 to 5.1 (127), anion exchange was used to assist with purification of the target proteins. Specifically Quaternary amine (Q)-Sepharose, a strong anion exchange chromatography matrix, was used for this purification step.

Lab-Made Q-Sepharose Chromatography Columns

Before proceeding to an automated purification process of our proteins, a series of purification experiments using lab-made columns were first performed to ensure that

conditions were optimized. Buffers chosen for these experiments were based on a protocol for plasma AAT purification (128). Approximately 5mL of Q-Sepharose Fast Flow (GE Healthcare Life Sciences) was poured into a 10mL pipette, which had steel wool previously packed into the tip, and allowed to settle. Next 5mL of wash buffer (WB, 2.5M NaCl, 0.1M Tris-HCl, pH 8.0, 5mM EDTA) was added to the column and allowed to flowthrough, followed by 5mL of equilibrate buffer (EB, 0.1M Tris-HCl pH 8.0, 5mM EDTA). Around 500µL of sample was then added to the column and flow through from this point on was saved for further analysis.

The column was then washed twice with 5mL EB and then treated with a salt gradient that would separate the target proteins from other proteins in the sample. This salt gradient consisted of 0, 0.2, 0.4 and 1M NaCl dissolved in our EB. The addition of the gradient to the column consisted of two 5mL washings for each salt concentration, with the exception of 1M NaCl, which had four washings. The column was washed four times with 1M NaCl to ensure that all protein had been removed from the column.

It was predicted that our proteins of interest would elute in the 0.4M NaCl range, therefore the flowthrough for these washings were collected in 1mL aliquots while the other concentrations were collected in 15mL conical centrifuge tubes. Samples of all the collected flowthroughs were run on a SDS-PAGE gel and analyzed with Coomassie, Silver staining and/or Western blotting, as previously described (Section 3.4), in order to verify the presence of our protein.

As shown in Figure 3.9A and C, our Q-Sepharose column was able to reduce the number of proteins present in the sample (ZAAT unpurified versus Q1-4). Among the bands indicted in this figure, the band at ~49kDa, in particular, was thought to be our ZAAT protein. When a Western blot (Figure 3.9B) was done to confirm this however, the band shifted down and was identified at ~40kDa. This could mean a number of different things. The first, is that the sample had been stored at 4°C for about a week before the Western blot analysis was performed. During this time the protein could have been degraded, therefore resulting in the smaller band observed. Doing the Western blot directly after purification and storing the samples with protease inhibitors are two things that could have been implemented to prevent this from occurring.

Another possibility is that our polyclonal antibody was binding to *Pichia pastorsis* proteins instead of our AAT proteins. In fact, a quick BLAST® search (http://blast.ncbi.nlm.nih.gov) revealed several *P. pastoris* proteins that were shown to have sequence homology to our AAT proteins. Among these, two proteins were of particular interest: Mu3-like subunit of the clathrin associated protein complex (AP-3) and a hypothetical protein. The Mu3-like subunit was shown to be 52kDa (same as our positive control) and had a max identity score of 34% (E value of 0.005). The hypothetical protein had a molecular weight of 36.4kDa (similar to the band seen in Figure 3.9B) and a max identity score of 25% (E value of 0.020). Given these parameters, its is possible that our polyclonal antibody (epitope unknown) could be identifying these proteins instead of our AAT proteins.

Figure 3.9 Protein Purification Analysis and Verification. (A) Purification of ZAAT using a lab-made Q-Sepharose chromatography column. Samples were visualized with Silver stain, in which Q1-4 represents fractions collected at 0.4M NaCl from the experiment. (B) Western blot and (C) Silver stain of another Q-Sepharose chromatography expeirment (Q1-4, lab-made column) and a Gel Filtration Chromatography (GFC 1-2) expeirment. Protein in Western blot detected by rabbit polycolonal Human AAT antibody 1:2000 dilution. (D) ZAAT sample purified using a HiTrap Q-FF column followed by Superdex 75 10/300 column and visualized by Coomassie stain. Blue box depicts area in which proteins were predicted to be located. Human AAT protein served as a positive control for the described experiments

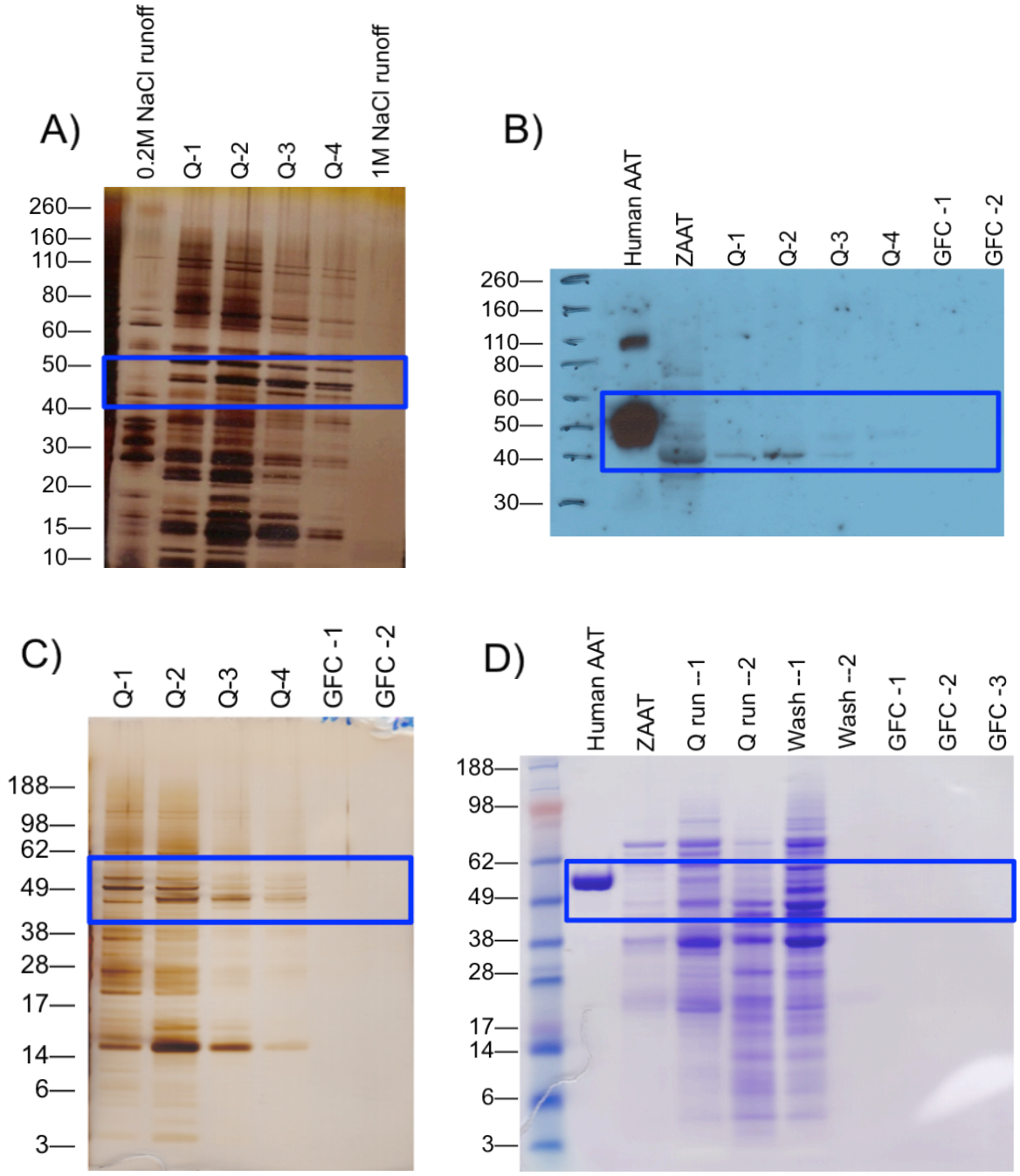


Figure 3.9 Continued.

To verify this, we could have transferred our SDS-PAGE gels to a PVDF membrane instead of a nitrocellulose one. One of the advantages of using a PVDF membrane is that it can be directly used for N-terminal sequencing (129), which would allow us to identify what exactly our Western blot was detecting. N-terminal sequencing could have also been conducted using gel pieces containing the bands of interest that had been extracted from a SDS-PAGE gel stained with Coomassie.

Automated Q-Sepharose Chromatography

Since our Silver stained gels led us to believe that we could elute our proteins of interest with the lab-made column protocol, we decided to move on to a more automated system that would eventually allow us to do large-scale purification. Samples were loaded onto a BioLogic DuoFlow chromatography system (Bio-Rad) that was fitted with a 1-mL HiTrap Q-FF column (GE Healthcare Life Sciences). The column was equilibrated with 1mL of Buffer A (0.1M Tris-HCl pH 8.0, 5mM EDTA) at a flow rate of 1mL/min. Proteins were then eluted with a linear gradient of 0–40% Buffer B (Buffer A + 1 M NaCl) at a flow rate of 1mL/min and a total volume of 20mL. Next, 5mL of Buffer B ran through the column at 1mL/min, followed by 10mL of Buffer A at 1mL/min.

Throughout the program fractions were monitored by a BioLogic QuadTec UV/Vis detector (280nm wavelength) and then collected in 0.5mL aliquots. At the end of the program, a report was generated (Figure 3.10A) that displayed the detector readings, conductivity and the change in Buffer B percentage seen throughout. As shown in Figure 3.10A, during one of our experiments a peak was detected in fractions 44-48 that had a max absorbance of 1.8 AU.

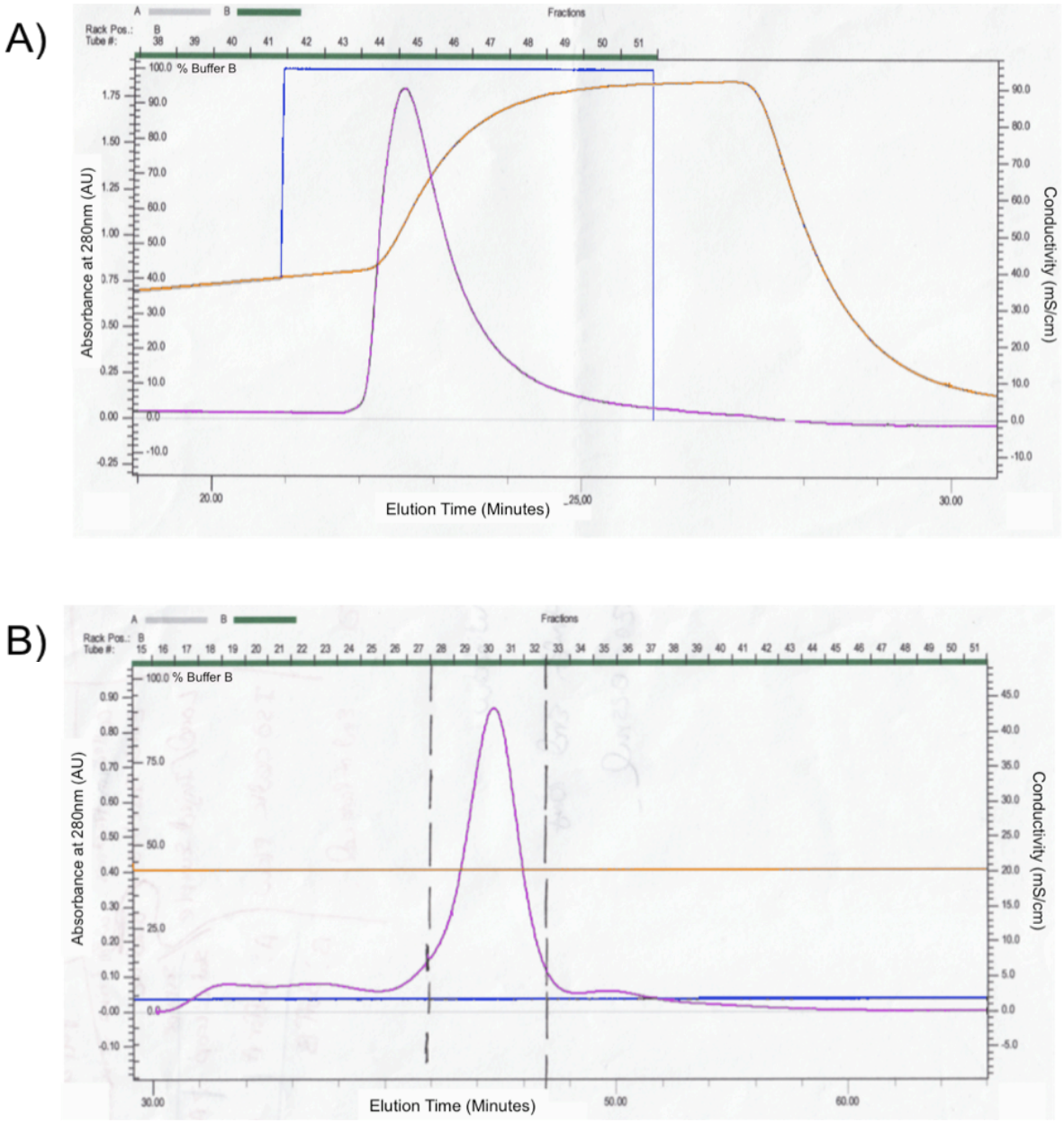


Figure 3.10 BioLogic DuoFlow Run Reports. UV signal (Purple line), % Buffer B (blue line) and conductivity (orange line) for samples run on BioLogic DuoFlow chromatography system (Bio-Rad) that was fitted with (A) HiTrap Q-FF column or (B) a Superdex 75 10/300 column (GE Health Life Sciences). Dotted lines in (B) represent where fractions were collected for analysis.

These fractions were then pooled together and analyzed by protein gel electrophoresis. Run-off from the chromatography experiments were also saved and analyzed.

Two runs were preformed using the HiTrap Q-FF column (Figure 3.9D). However issues arose during the first run (Lane 4) that resulted in most of the protein being lost to the run-off (Lane 6). We were able to resolve the problem during our second run (Lane 5), which resulted in a greater amount of our protein being retained in the fractions and not eluted in the run-off (Lane 7). For all four samples, the protein was predicted to be located at the ~50kDa band.

Given these results, we concluded that we were able to isolate our target proteins using the HiTrap Q-FF column. However, as also indicated by the amount of bands present in Figure 3.9D, the protein was not very pure and would need follow-up purification steps.

b) Size-Exclusion Chromatography

Size exclusion chromatography (SEC), also known as gel filtration chromatography (GFC), separates proteins based solely on differences in their molecular size. Separation is achieved due to the porous matrix packed in a chromatographic column. Molecules applied to the column display different degrees of access within the matrix based on their size. Smaller molecules can easily enter/exit the pores freely, thereby prolonging their navigation through the matrix. Larger molecules, on the other hand, are excluded from entering the pores and are instead swept through the column by the aqueous buffer (mobile phase) that is ran though the column. This

allows the proteins within a specific sample to elute in decreasing order of size, which can be detected by an in-line UV monitor. Due to its low resolution though, SEC is frequently used as a final polishing technique to remove any remaining contaminants from the target protein (130).

SEC was performed using a BioLogic DuoFlow chromatography system fitted with a Superdex 75 10/300 column (GE Health Life Sciences). This column has a separation range for molecules between 3,000 and 70,000kDa and a bed dimension of 10 x 300mm (131). A program was initiated in which 500µL of the sample was injected onto the column at a flow rate of 0.25mL/min. The mobile phase (Water/0.05% TFA) ran through the column at a flow rate of 0.25mL/min for a total of 50mL. Once more fractions were monitored by a BioLogic QuadTec UV/Vis detector (280nm wavelength) and then collected in 0.25mL aliquots throughout the experiment.

At the end of the program, a report was generated (Figure 3.10B) that displayed the detector readings, and conductivity seen throughout. As seen in Figure 3.10B, a peak was detected in fractions 28-32 that had a max absorbance of 0.84 AU. These fractions were then pooled together and analyzed by protein gel electrophoresis and Western blotting (Figure 3.9B-D).

Surprisingly the SDS-Page gels displayed no protein bands for the SEC fractions after treatment with either Coomassie or Silver stain. This was true regardless if the sample had been previously purified with a lab-made Q Sepharose column (Figure 3.9C) or with the HiTrap Q-FF column (Figure 3.9D). Even after Western blot analysis, we were still unable to detect protein following SEC (Figure 3.9B). This could be the result of a number of different things. For example, the proteins may have degraded

during the time they were purified until the time they were analyzed. This could be especially true for the samples purified by a lab-made Q Sepharose column followed by SEC (Figure 3.9B and C), in which almost a week went by between purification and analysis. Although the samples were stored at 4°C during this time, the proteins may have still degraded during this time as discussed previously. For the samples purified by HiTrap Q-FF and SEC (Figure 3.9D) however, analysis was performed with 24-hours of purification, therefore reducing the likelihood of protein degradation for these samples. Another reason for this phenomenon could be as simple as not enough protein was added to the Superdex column and therefore, upon fractionation, the protein became too dilute. This dilution could reduce the protein concentration to levels that are not detectable by our analysis methods. Determining the protein concentration by BCA or Bradford Assay prior to SEC and before protein gel electrophoresis would help to determine if this is indeed the case and if samples need to be concentrated or not.

3.7 Discussion

Upon the completion of the described procedures in this chapter, several different factors regarding the production of M and ZAAT in *P. pastoris* were concluded. For starters, according to Levina et al (100), the induction time for optimal expression is between 17-72 hours. However in our hands, expression was only detectable after 24-hours. While this was later brought into question, we were able to show that this induction time, while maybe not optimal, was sufficient to express our proteins.

In addition, the vector used in the original paper (pHil-D2) only contained the *HIS4* marker and was not suitable for multi-copy integrations. Our vector (pPic3.5K), on

the other hand, contained both the *HIS4* marker and the bacterial Kan^R gene, the combination of which drastically limited the amount of prospective clones by identifying yeast colonies with multi-copy integrations.

Lastly, by comparing a physical method, similar to that seen in the original paper, to a chemical method, such as Y-PER® , we were able to show that the chemical method is actually a better form of lysis. These conditions helped to lay the foundation for future experiments conducted in our laboratory.

Unfortunately the various analyses performed on the pooled fractions from both the Q-Sepharose and SEC experiments proved to be inconclusive overall. After several more attempts at the described procedure, it was determined that the current methods were unable to produce the quantity and purity of protein needed for future experiments. Therefore we decided to take a different approach, in which the cDNA was re-designed to include a Histidine tag that could be exploited during the purification process and hopefully improve our results (Chapter 4 and 5)

CHAPTER 4:

USING SITE-DIRECTED MUTAGENESIS TO CREATE TAGGED α1-ANTITRYPSIN PROTEINS

4.1 Introduction

Not that long ago if scientists wanted to study a specific mutation, they would have to conduct phenotypic screenings after the isolated gene was treated with mutagenic agents. The screenings were not only time-consuming but would often miss lethal or non-phenotypically observable mutations. In addition, treatment with mutagenic agents would commonly result in mutations being randomly distributed throughout the gene, instead of specific locations. (132).

Today however, mutagenesis is typically achieved through either polymerase-chain reaction (PCR) or non-PCR methods. In these methods, mutations are created though codon changes at specified residues in order to produce amino acid substitutions. This process is commonly referred to as Site-Directed Mutagenesis (SDM) (132,133). Over the past three decades, several different SDM methods have been proposed (133-141), some of which are now commercially available as SDM kits.

One of the most common methods of SDM is the QuikChange method, which utilizes complementary primer pairs in a single PCR to rapidly introduce point mutations into sequences of interest (141,142). After thermocycling, the product is then treated with the restriction enzyme DpnI, which targets methylated and hemimethylated DNA.

Since DNA isolated in almost all *E. coli* strains is dam methylated, DpnI only targets the parental DNA, and leaves the newly synthesized DNA intact (136,142).

In this chapter, we will discuss attempts that were made to add a hexa histidine-tag (6xHis-tag) to AAT cDNA through two different versions of the QuikChange method. The first method is a two-step process in which the forward primer (binds to the 'antisense' strand to create a new 'sense' strand) and reverse primer (binds to the 'sense' strand to create a new 'antisense' strand) are placed into separate tubes for one round of PCR and then combined for a second PCR (Figure 4.1). The second method follows the traditional QuikChange approach where both primers are used in a single PCR (Figure 4.2).

4.2 Plasmids and Primers

a) DNA Plasmids

Site-directed mutagenesis experiments were conducted using glycerol stocks of DH5α cells that had been transformed with MAAT or ZAAT cDNA ligated into the pPic3.5K vector (Section 3.2). To increase the amount of available cDNA, cells were grown overnight in 100mL LB media and then isolated using the S.N.A.P. MidiPrep Kit. The viability of the newly isolated cDNA was confirmed through restriction enzyme digestion, prior to the SDM experiments. Digestion was performed for 1 hour at 37°C in the presence of either BamH I by itself, EcoR I and BamH I together, or no restriction enzyme.

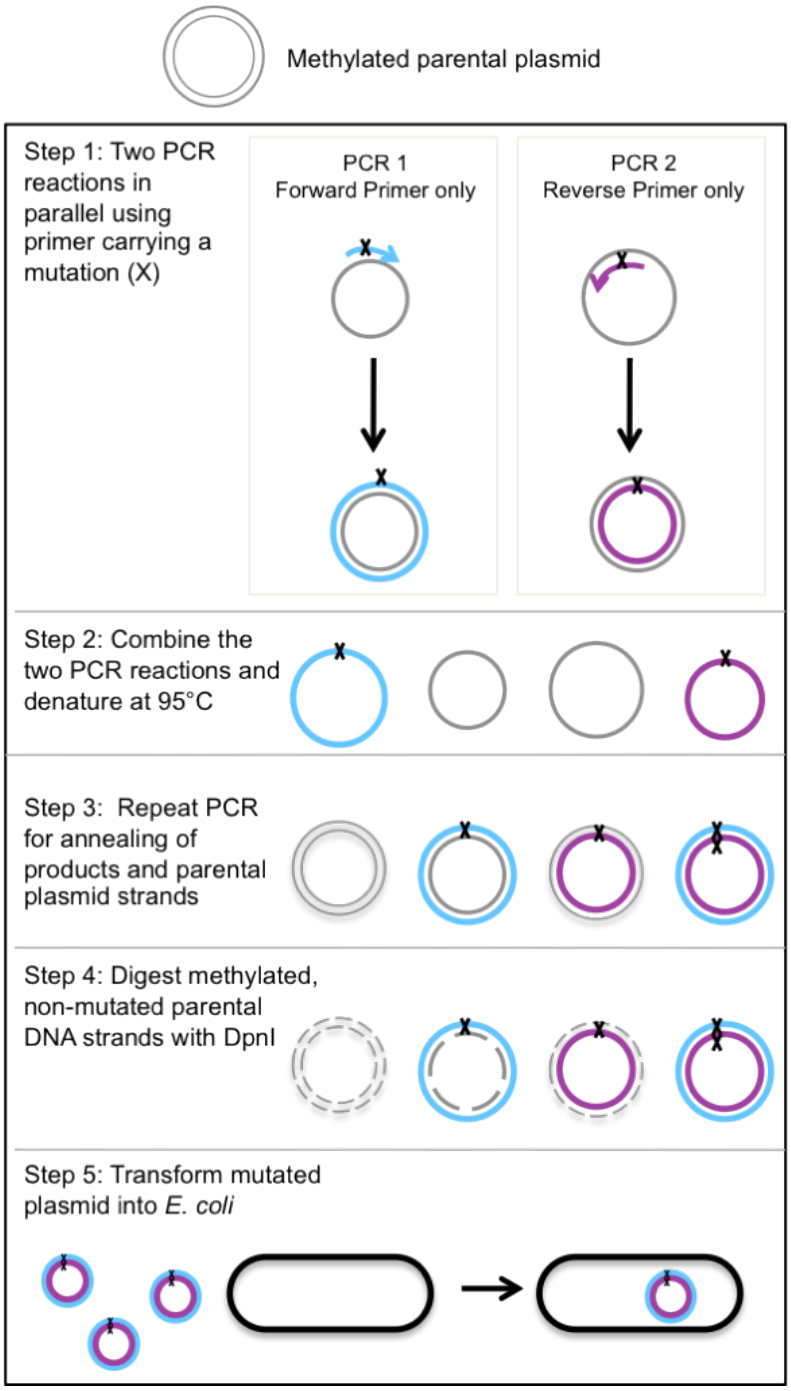


Figure 4.1. Flow chart of the Two-Step Site-Directed Mutagenesis Method. The parental DNA is shown in gray and the two PCR synthesized strands are in blue and purple. The letter X marks the position of the mutation

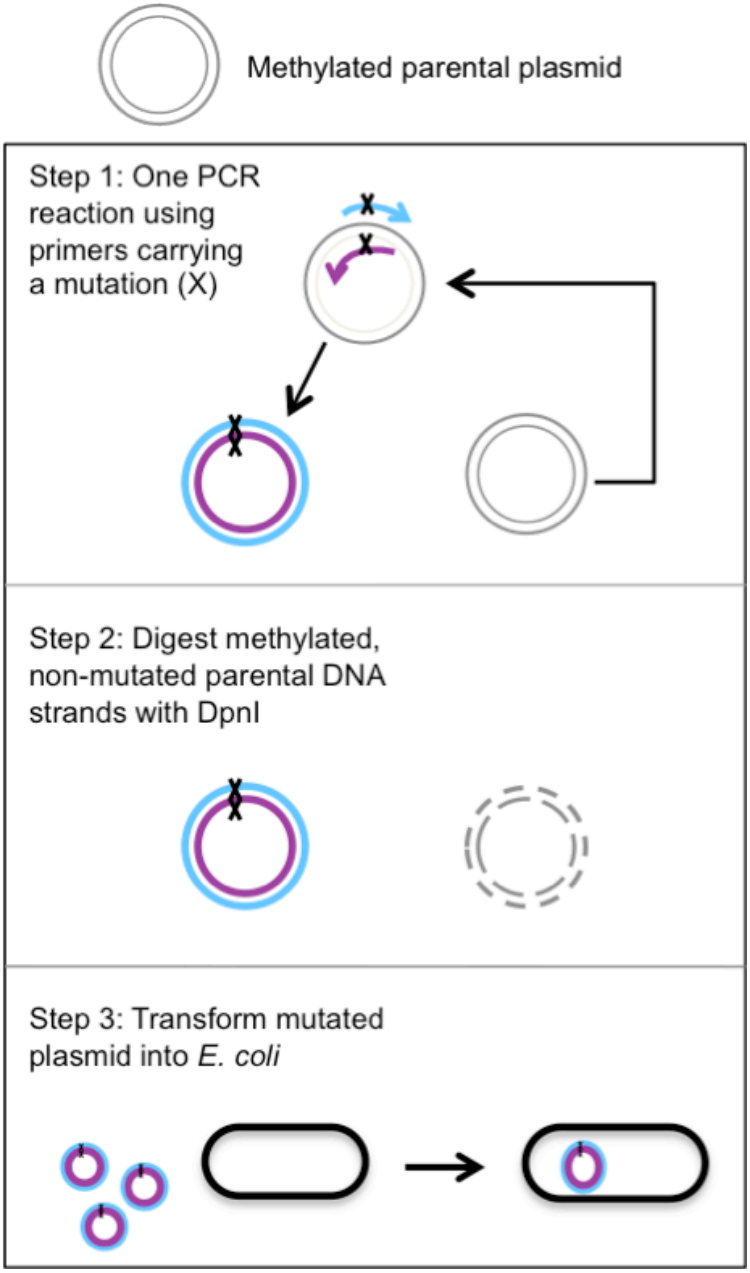


Figure 4.2. Flow Chart of the Single-Step Site Directed Mutagenesis Method. The parental DNA is shown in gray and the PCR synthesized strands in blue and purple. The letter X marks the position of the mutation.

Once complete, a sample of the products was run on 1% agarose gel, which was then stained with ethidium bromide and photographed under UV light (Figure 4.3). As expected, the combined digestion of EcoR I and BamH I (lanes 4 and 7) resulted in two distinct bands, one of which was the pPic3.5K (~9kb) and the other, AAT cDNA (~1.2kb). On the other hand, when the cDNA was treated with BamH I alone (lanes 3 and 6), only one band was visible on the gel (~10kb), indicating that it had been properly linearized. Lastly cDNA that not been treated with any restriction enzymes (lanes 2 and 5) displayed supercoiling and a thick band around 10.2kb. Overall, these results indicated that the cDNA was still viable and could be used for our SDM experiments.

b) Primer Design

The QuikChange Primer Design website (www.agilent.com/genomics/qcpd) was used to create complementary primers. These primers contained an 18bp segment that would add a hexa histidine-tag (6xHis-tag) to the N-terminal of our previously synthesized cDNA constructs (Section 3.2a). This website follows commonly employed guidelines for double-primer PCR reactions, including: 1) primer sequences must be complementary to each other, 2) mutation located in the middle of the primers 3) Primer length should include 10-15 nucleotides (nt) on either side of mutation, 4) a GC content of 40-70%, 5) sequence starts and ends with one or more G or C base, and 6) the melting temperature (Tm) should be ≥78ºC (141,142). Primers (Table 4.1) were ordered from Sigma and used without further purification.

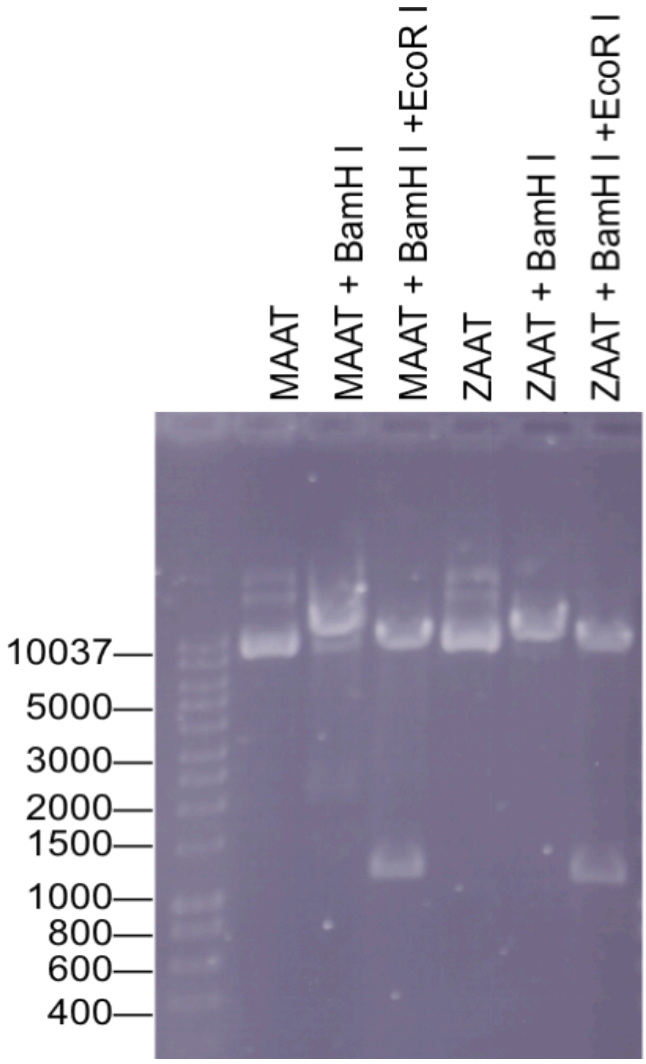


Figure 4.3. Restriction Enzyme Digest of MAAT and ZAAT with BamH I and EcoR I to confirm DNA viability. As expected, uncut MAAT and ZAAT (lanes 2 and 5) resulted in DNA supercoiling whereas a double digest with BamH I and EcoR I (lanes 4 and 7), caused the AAT sequence to be isolated from the pPic3.5K vector (1.2kb and 9kb respectively). HyperLadder I (Bioline) was used as the molecular weight marker in all gels.

Table 4.1 Forward (-F) and Reverse (-R) Primer Sequences for the insertion of an 18nt segment coding for a 6XHis tag. Insertion is indicated in red.

Primer Name	Primer Sequence 5' to 3'
HisAAT-F	GAAGGGATCCACCATGCATCATCATCATCATCATGAAGACCCCCAAGGAG
HisAAT-R	CTCCTTGGGGGTCTTCATGATGATGATGATGATGCATGGTGGATCCCTTC

Table 4.2. PCR Reaction Components and Concentrations. Reaction 1 and 2 are for the Two-Step Site Directed Mutagenesis method and Reaction A is for the Single-Step Site-directed Mutagenesis method,

	Reaction 1	Reaction 2	Reaction A
Template plasmid DNA	~50ng	~50ng	~50ng
Forward primer	40pmol	-	8.4pmol
Reverse primer	-	40pmol	8.4pmol
25mM dNTP mix	1µL	1µL	1µL
10X *PfuUltra*® reaction buffer	5 µL	5 µL	5 µL
QuiKSolution reagent	3 µL	3 µL	3 µL
PfuUltra® HF DNA polymerase	2.5 U	2.5 U	2.5 U
Final Volume	50µL	50µL	50µL

The QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies), which contains the *PfuUltra*® High-Fidelity DNA Polymerase, was used to carry out all SDM experiments. The kit also contains a pWhitescript 4.5-kb plasmid that served as a control for our experiments.

4.3 Two-Step Site-Directed Mutagenesis

The protocol used for the two-step site-directed mutagenesis (2SDM) experiments was based on a procedure from the Soderling lab (Duke University, (143)) that combined elements from the "QuikChange II XL" manual (142) with those from an article published by Wang and Malcolm in 1999 (136). The method starts with the simultaneous PCR of two separate reactions, each one containing separate primers. The products of these parallel reactions are then combined into a single tube and PCR is performed once more (see Table 4.2 for reaction components and concentrations).

a) PCR Conditions

PCR tubes containing the reaction mixes were placed into the thermocylcer, and after an initial denaturation step at 95ºC for 1 minute, PCR was conducted for 10 cycles. Each PCR cycle consisted of denaturation at 95ºC for 1 minute, primer annealing at 55ºC for 90 seconds and DNA synthesis at 68ºC for 10 minutes (one minute per kb of plasmid length). Afterwards, 25µL from each reaction was then mixed together into a fresh 0.2mL PCR tube, along with 0.75µL of fresh *PfuUltra*® High-Fidelity DNA Polymerase.

The tube, now containing both forward and reverse primers, was then placed back into the thermocycler and another round of PCR began. After an initial denaturation step that was carried out at 95ºC for 1 minute, PCR was conducted for 18 cycles with denaturation at 95ºC for 1 minute, primer annealing at 55ºC for 90 seconds and DNA synthesis at 68ºC for 10 minutes (one minute per kb of plasmid length). The reaction was then held at 4ºC until it ready for removal.

b) Amplification Analysis

After the PCR tube was removed from the thermocycler, 1µL of DpnI (200U/µL) was added and the tube was incubated at 37ºC for at least 1 hour. This allowed the DpnI endonuclease to digest the methylated parental plasmid strands, leaving only the newly synthesized strands behind. After DpnI digestion was completed, amplification was confirmed by running a sample of the products on a 1% agarose gel, which was then stained with ethidium bromide and photographed under UV light (Figure 4.4).

Upon the successful implementation of the 2SDM experiment and DpnI digestion, we would expect to see a single band around 10.2kb on the agarose gel. However, as seen in Figure 4.4A, amplification of either MAAT or ZAAT cDNA was undetectable by agarose gel electrophoresis during the first few attempts at 2SDM. On the other hand, we were able to observe a band at approximately 5kb for our control, indicating that our experimental set-up was successful. Therefore the problem was thought to lie either with the cDNA or the primers used for our samples.

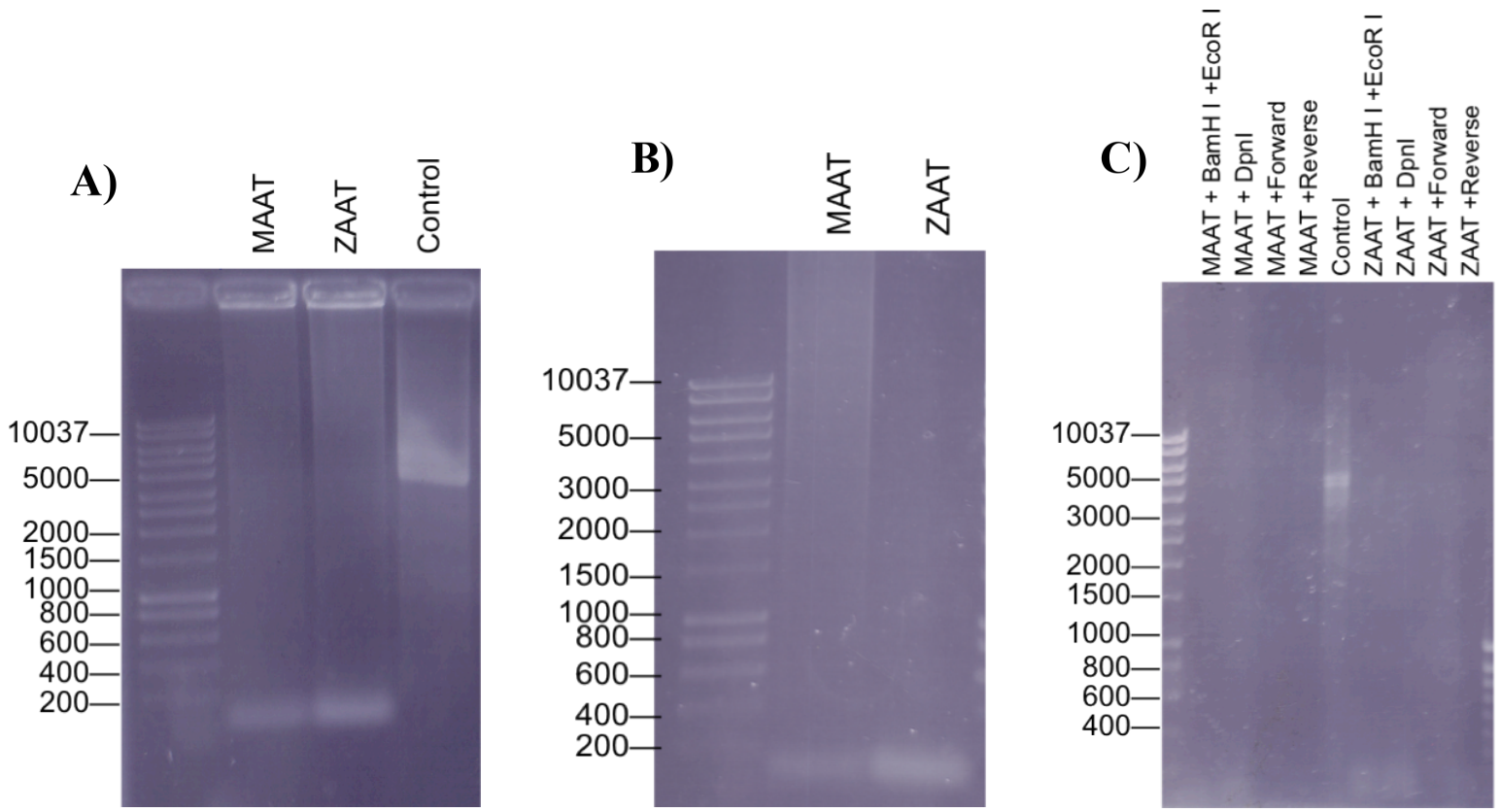


Figure 4.4 Results from Two-Step Site-Directed Mutagenesis. A) Original attempt at 2SDM with wrong primer concentrations. B) Second attempt at 2SDM with corrected primer concentration. C) Further analysis of 2SDM results that included looking at Reaction 1 (forward) and 2 (reverse) alone and digesting the combined reaction with Bam H1 and EcoR I instead of DpnI. HyperLadder I was used as the molecular weight marker in all gels.

A closer look at the primers revealed that there had been a miscalculation during the initial resuspension of the reverse primer that caused the primer concentration to be almost half of what was originally expected. This revelation provided a possible explanation for the lack of amplification observed, and once remedied, 2SDM was attempted again. However, even with the corrected primer concentrations, amplification of our samples was still not detected (Figure 4.4B).

In an attempt to rule out DpnI digestion as a cause for this phenomenon, agarose gel electrophoresis was used to compare the DpnI digested products with those digested by BamH I and EcoR I as well as Reaction 1 and 2 undigested. If successful, we predicted that digestion of the 2SDM products with both BamH I and EcoR I would result in two bands: one at ~9kb and the other at ~1.2kb. Once more we anticipated that DpnI would produce a single band at ~10.2kb and undigested samples would display supercoiling. Unfortunately there was no DNA detected in any of the samples analyzed aside from our control (Figure 4.4C).

Lastly, the integrity of the DNA was once more verified by restriction enzyme digestion with BamH I and EcoR I, which result in the successful separation of the AAT cDNA from the pPic3.5K vector (not shown). Since it seemed that after several attempts the 2SDM method was still unable to provide any detectable amplification of our samples, we transitioned to the traditional single-step site directed mutagenesis method next.

4.4 Single-Step Site Directed Mutagenesis

Unlike the 2SDM method, the single-step site directed mutagenesis (1SDM) method consists of only one PCR reaction that contains both the forward and reverse primers in the same tube. 1SDM experiments were conducted by following the procedure outlined in the QuikChange II XL Site Directed Mutagenesis Kit manual (142). A list of the different reaction components and concentrations used are recorded in Table 4.2. After an initial denaturation step at 95ºC for 1 minute, PCR was conducted for 18 cycles with denaturation at 95ºC for 50 seconds, primer annealing at 60ºC for 50 seconds and DNA synthesis at 68ºC for 10 minutes (one minute per kb of plasmid length). The primers were further allowed to extend at 68ºC for 7 minutes and then held at 4ºC till the reaction was ready for removal.

Once the tube was removed from the thermocycler, 1µL of DpnI was added to the reaction and DpnI digestion was carried out at 37ºC for at least 1 hour. This was then confirmed by agarose gel electrophoresis, as described before. Again, amplification of our samples was not detectable by agarose gel electrophoresis (Figure 4.5A). This proved true even if the PCR product was analyzed prior to digestion (Lanes 2 and 4). In fact, our results mirrored those seen when no DNA was added to our control reactions (Figure 4.5A, Lane 7). Thinking that something may be wrong with our cDNA, we decided to once more validate its integrity through restriction enzyme digestion using EcoR I and BamH I. As expected, this produced two bands on the agarose gel—the first at ~9.0kb (pPic3.5K) and the second at ~1.2kb (isolated AAT)—indicting that our cDNA was still intact and could be isolated from the pPic3.5K vector as before (Figure 4.5B).

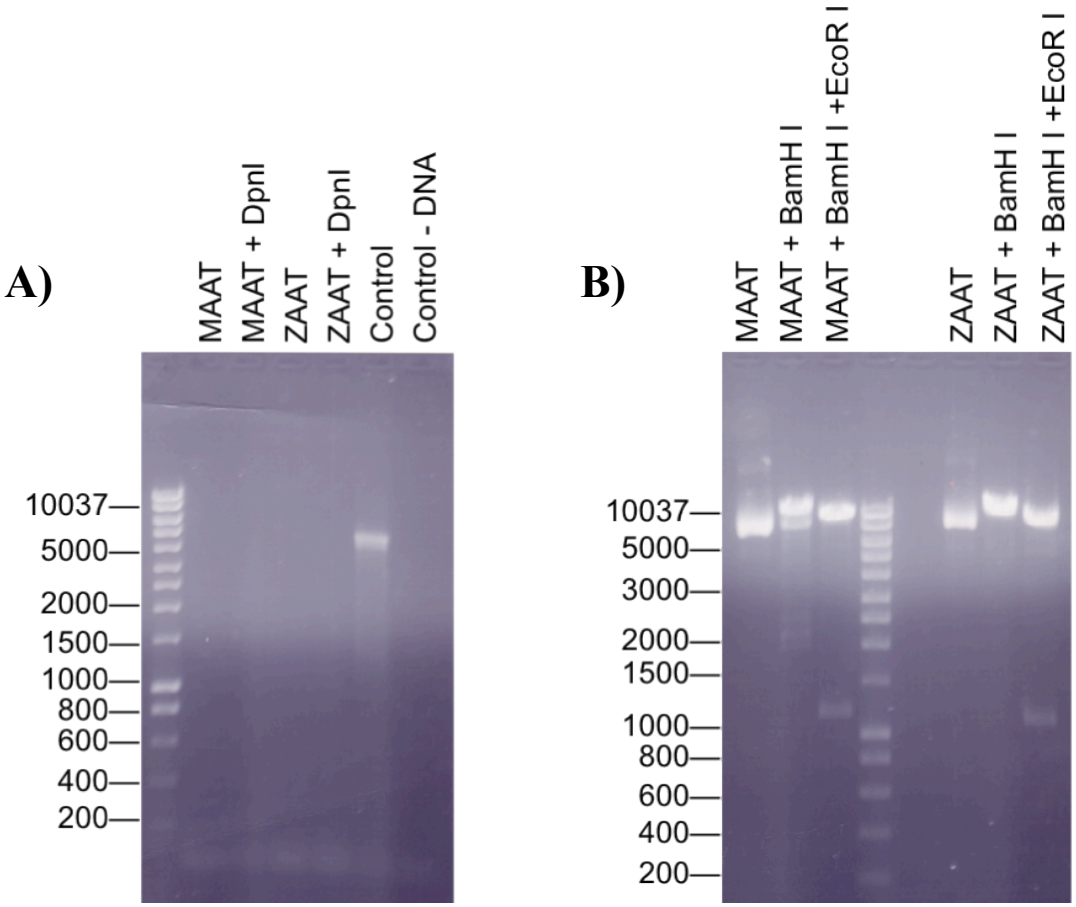


Figure 4.5 Results from Single-Step Site-Directed Mutagenesis. A) Attempt at employing the 1SDM method. Products were either digested with DpnI or ran on the gel without any type of digestion. The control reaction was also performed with and without the provided cDNA. B) Confirmation that our cDNA was still valid and could be digested by BamH I and EcoR I. HyperLadder I was used as a molecular weight marker for both gels.

4.5 Discussion

Overall, our attempts at site-directed mutagenesis were unsuccessful, possibly for a multitude of different reasons. A common issue with using complementary primers, such as those needed for SDM, is the formation of “primer-dimers” occurring within the reaction. These primer-dimers can result from partial annealing of a primer with a second primer in the reaction and can drastically reduce the amount of transformants produced (141). Such a phenomenon could possibly explain why the SDM experiments in this chapter did not perform as well as had been expected. The presence of primer-dimers would also explain the band at ~100kb seen in our agarose gels (Figure 4.4 and 4.5A). One way that we could reduce the formation of primer-dimers is to increase the annealing temperature closer to the melting temperature of our primers. To find the optimal temperature for annealing we could perform a round of touchdown-PCR (a method that gradually decreases the annealing temperature over the course of successive cycles) as described in the literature (144-146).

Another explanation could be that since our primers consisted of 50nt, a relatively long length for a primer, some secondary structure formation could have occurred, which would prohibit it from binding to the cDNA (142). However several studies have been published (136,141,147,148) in which complementary primers that are the same size or longer than ours have been successful in SDM experiments. In order to help resolve this issue, we could once more alter the PCR conditions for the SDM experiments. This could involve optimizing the annealing temperature as described previously or increasing the denaturation temperature to 98°C. Another option

would be to linearize our cDNA prior to PCR, since linear DNA is able to denature more efficiently than supercoiled DNA (149).

Lastly, a mistake during the ordering and synthesis of the primers could account for the lack of amplification seen in our experiments. Closer inspection of our primers revealed that this might have indeed been the case. The sequence that was generated, and thereby used in our experiments, contained an extra guanine (G) in both the forward and reverse primers (Table 4.3) that shifted the whole sequence and disrupted the primer-template duplex (Table 4.4 compared to Table 4.5). Such a disruption would promote primer-dimer formations (100% complementary) while simultaneously reducing primer-template annealing (increased multiple mismatches). If these experiments were to be conducted again, new primers should be ordered and tested prior to the implementation of the other methods described in this section.

Despite several attempts at implementing the 2SDM and the 1SDM methods, ultimately we were unsuccessful at amplifying our target DNA so that it contained the 18nt insertion. Therefore it was decided that the cDNA should be redesigned and reordered so that it contained a 6xHis-tag on the N-terminal of our proteins. The methods and results of which are discussed in the next chapter (Chapter 5).

Table 4.3. Corrected Design for Forward (-F) and Reverse (-R) Primers for the insertion of an 18nt segment coding for a 6XHis tag. Letters in blue indicted missing residues, orange indicates added residues and insertion is indicated in red.

Primer Name	Primer Sequence 5' to 3'
HisAAT-F	CGAAGGGATCCACCATGCATCATCATCATCATCATGAAGACCCCCAAGGAG
HisAAT-R	CTCCTTGGGGGTCTTCATGATGATGATGATGATGCATGGTGGATCCCTTCG

Table 4.4 Primer-Template Duplex for Corrected Primers

Primer Name	Primer-Template Duplex
HisAAT-F	5'-cgaaggatccaccatgcatcatcatcatcatcatgaagacccccaaggag-3' |||||||||||||||| |||||||||||||||| taagcttcctaggtggtac----------cttctgggggttcctctac
HisAAT-R	attcgaaggatccaccatg----------gaagacccccaaggagatg |||||||||||||||| |||||||||||||||| 3'-gcttcctaggtggtacgtagtagtagtagtacttctgggggttcctc-5'

Table 4.5. Primer-Template Duplex for Ordered Primers

Primer Name	Primer-Template Duplex
HisAAT-F	5'- gaagggatccaccatgcatcatcatcatcatcatgaagacccccaaggag-3' ||||| |||||||||||||||| taagcttcctaggtggtac----------cttctgggggttcctctac
HisAAT-R	attcgaaggatccaccatg----------gaagacccccaaggagatg |||||||||||||||| |||||||||||| 3'- cttcctaggtggtacgtagtagtagtagtacttctgggggttcctc-5'

CHAPTER 5:

EXPRESSION AND PURIFICATION OF TAGGED α1-ANTITRYPSIN PROTEINS

5.1 Introduction

Protein affinity tags are highly popular tools for the purification of recombinant proteins. One reason for this is that they can provide a hundred- or even thousand-fold purification increase when compared to untagged purifications techniques. In addition, the use of affinity tags allows a generalized protocol to be implemented to purify a diversity of proteins, thereby saving both time and money (150).

Over the past 20 years, several different types of affinity tags have become commercially available, with the most common being the polyhistidine affinity tag. This tag typically consists of six consecutive histidine residues (hexa-histidine) but can range in length from two to ten histidine residues (150,151). While polyhistidine fusion proteins are conventionally purified using immobilized Ni^{2+}, other metals such as Ca^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+}, or Fe^{3+} can also be used (151,152). This purification process can result in fairly pure protein (>80%) after only one chromatographic step (151).

Initial endeavors at producing untagged, properly folded and active α1-Antitrypsin (AAT) recombinant protein were ultimately unable to produce the desired protein concentrations (Chapter 3). Therefore to increase selectivity and ease of purification, it

was decided that a hexa-histidine tag (6xHis-tag) should be added to the N-terminal of the proteins.

However, as discussed in Chapter 4, several complications arose during preliminary attempts at adding this tag to wild-type (M) and mutant (Z) AAT cDNA. Therefore, new cDNA constructs containing the 6xHis-tag were synthesized for both MAAT and ZAAT. This chapter will discuss the manipulation of these constructs into the *Pichia pastoris* vector, pPic3.5K, as well as the expression, and purification of MAAT and ZAAT protein from *P. pastoris*.

5.2 DNA Manipulation

Prior to expression in *P. pastoris,* cDNA constructs for both the wild-type and mutant forms of AAT were once more synthesized and then subcloned into the pPic3.5k vector using the gram-negative bacterium, *Escherichia coli.* The techniques performed for this process are described below and are similar to those discussed in Section 3.2.

a) Tagged AAT DNA Constructs:

The new MAAT and ZAAT cDNA constructs (Table 5.1), ordered from Blue Heron Biotech, were again flanked by BamH I and EcoR I restriction sites to assist with subcloning as previously described (Section 3.2a). The Kozak sequence was also added to ensure proper translation in *Pichia pastoris* and cytosolic protein production was assured though the removal of the signal peptide sequence. Unlike the previously synthesized cDNA, the new constructs also included a 6xHis-tag and a tobacco etch virus (TEV) cleavage site at the N-terminus of the cDNA.

Table 5.1. Protein Sequences for Hexa-Histidine Tagged MAAT and ZAAT Proteins. Sequences are based on cDNA generated from Blue Heron Biotech. Restriction enzyme sites are depicted in blue, Kozak sequence for pPic3.5K in green, 6xHis-tag in orange, TEV cleavage site in purple and Z mutation (E342K) shown in red. Both proteins are predicted to have a molecular weight of approximately 47kDa and an isolectric point of 5.56 (www.expasy.org).

	Protein Sequences
MAAT	GSTMHHHHHHENLYFQGMEDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFS PVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVD KFLEDVKKLYHSEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFEV KDTEEEDFHVDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDII TKFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNGADLSGVTEEAPLKLSKAVHKAVLTIDEKGTE AAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK-EF
ZAAT	GSTMHHHHHHENLYFQGMEDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFS PVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNGLFLSEGLKLVD KFLEDVKKLYHSEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFEV KDTEEEDFHVDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDII TKFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNGADLSGVTEEAPLKLSKAVHKAVLTIDKKGTE AAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK-EF

To limit any effects that the 6xHis-tag might have had on the already fragile ZAAT structure, a TEV cleavage site was added to the sequence. This site (ENLYFQG) is recognized by a TEV protease that binds and cleaves the cDNA at a specific location (between Q and G), thereby removing the 6xHis-tag as well as anything upstream of it (153). Constructs were first synthesized into the Blue Heron pUCminusMCS vector (Figure 3.1) and then subcloned into the pPic3.5K (Figure 3.2) prior to *P. pastoris* expression.

b) Subcloning of AAT cDNA

Subcloning of MAAT and ZAAT cDNA into the pPic3.5K vector was once more performed though a series of steps that included plasmid amplification and isolation, restriction enzyme digestion, and DNA ligation. As described before (Section 3.2b), plasmid amplification was conducted through the transformation of Max Efficiency® DH5α™ Competent Cells with MAAT and ZAAT cDNA. The amplified AAT DNA was then isolated and cleaned using the S.N.A.P. Midiprep kit, followed by digestion with EcoR I and BamH I restriction enzymes. The restriction enzymes excised MAAT and ZAAT cDNA from its original vector, and the resulting fragment was separated via agarose gel electrophoresis. This can be seen in Figure 5.1A and B where plasmids treated with both EcoR I and Bam H I resulted in two bands, one for the pUCminusMCS vector (~3kb) and the other the AAT DNA fragment (~1.2kb). DNA that was treated with only BamH I was linearized and resulted in a single band at ~4.2kb whereas undigested DNA underwent supercoiling. The DNA fragment was then purified from the gel using the S.N.A.P. Gel Purification kit as per manufacture's instructions.

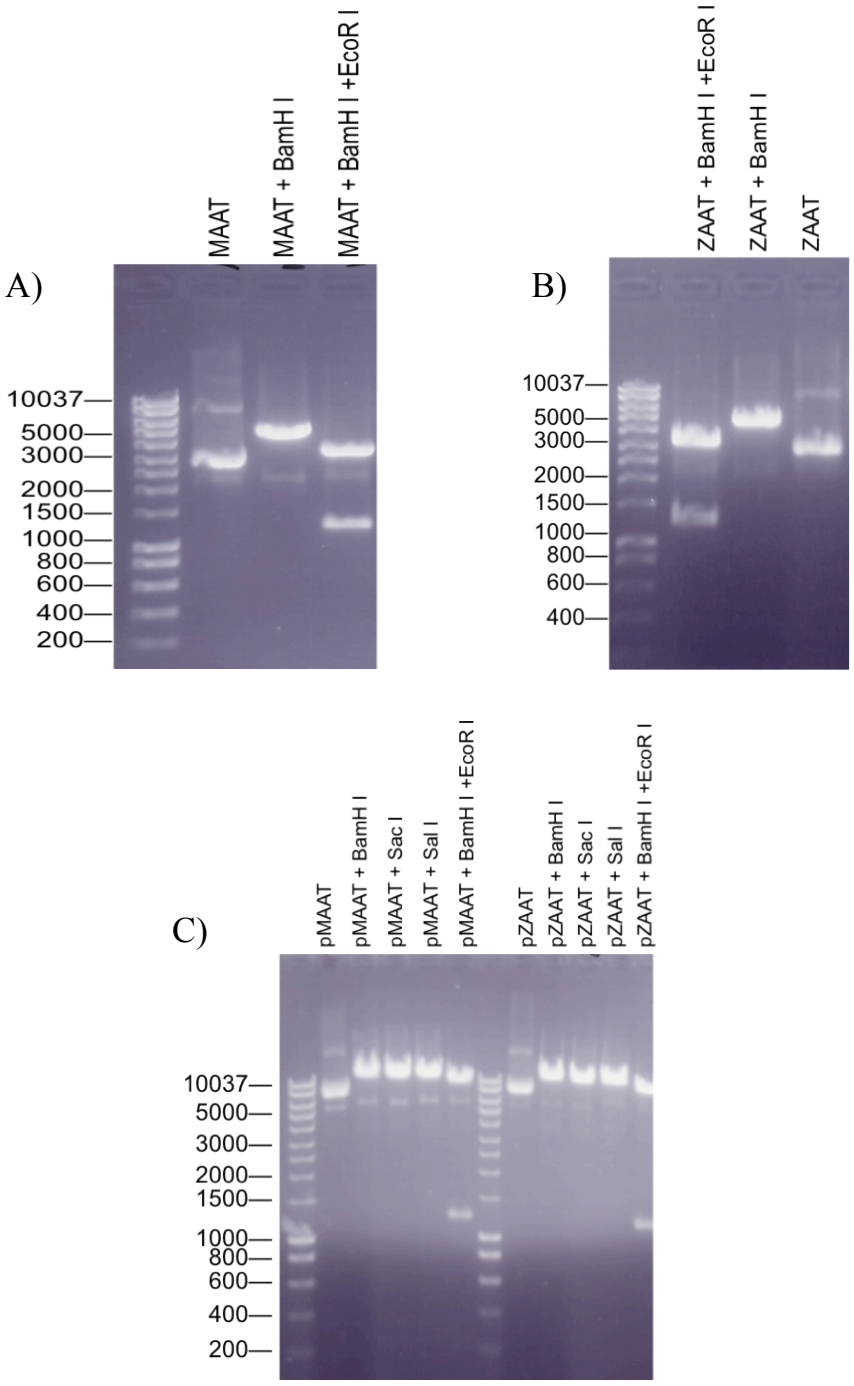


Figure 5.1. Manipulations of Tagged α1-Antitrypsin DNA in *E. coli*. Restriction enzyme digestion of MAAT (A) and ZAAT (B) with EcoR I and BamH I prior to ligation into pPic3.5K. (C) Following ligation, DNA was linearized with Sal I and Sac I restriction enzymes. Gels were made in 1% agarose, stained with ethidium bromide and photographed under UV light. Hyperladder I was used for all DNA gels.

After the DNA fragments were isolated from the agarose gel, they were then ligated into the pPic3.5K vector, as described previously (Section 3.2b). The plasmids were then linearized overnight with either Sac I or Sal I so that the genes can be inserted either at the *AOX1* locus (Sac I) or the *His4* locus (Sal I) within the *P. pastoris* genome. Following linearization, the DNA was separated via agraose gel electrophoresis in which linearized DNA (either by BamH I, Sac I, or Sal I) produced a single band at ~10.2kb. Undigested and double digested (BamH I and EcoR I) were also analyzed as controls (Figure 5.1C). DNA clean-up was once more performed using the Wizard® SV Gel and PCR Clean-Up System, as per the manufacture's instructions.

5.3 P. pastoris *Expression, Induction and Harvesting*

The expression, induction and harvesting of MAAT and ZAAT protein from *P. pastoris* was performed similar to what was described in Sections 3.4 and 3.5. To summarize, SMD1163 cells were first made competent by following the protocol outlined by Wu and Letchworth (111) and then electroporated with linearized MAAT or ZAAT DNA using a MicroPulser Electroporator. Cells were thought to be osmotically sensitive during and following electroporation (103), therefore before plating, cells are incubated in Sorbitol for one hour at 30ºC, followed by an additional hour in a 1:1 mix of Sorbitol and YEPD (Yeast extract, Peptone, and Dextrose). The transformed cells were next spread onto YEPD agar plates with 0.05mg/mL G418 and placed into an incubator at 30ºC for 3-5 days.

a) Screening for multi-copy inserted clones

The pPic3.5K contains the bacterial kanamycin (Kan^R) gene that confers resistance to G418 in *Pichia pastoris*. This resistance is dose dependent in that the more copies of the gene that integrated into the genome, the higher the resistance that is conferred upon the cells (78,102). Therefore, hyper-resistance to G418 was utilized to screen for colonies carrying multiple copies of the Kan^R gene and thereby multiple copies of the AAT genes. This was performed as previously described (Section 3.3b), in which single SMD1163 colonies were mixed with 100µL of sterile water in a 96-well plate and then spotted onto YEPD agar plates containing G418 at either 0.05, 0.10, 0.15 or 0.20mg/mL. The plates were then incubated at 30ºC for 3-5 days. Out of the almost 200 colonies screened, only 10 (4 MAAT and 6 ZAAT) were able to survive on the YEPD agar plates that contained higher amounts of G418. These 10 colonies were isolated and then induced in 10mL of Induction media for 24-hours (based on our previous findings from Section 3.5a). After the 24-hour induction time period, the cells were then pelleted down and stored at -20°C until ready for lysis.

Protein expression levels of each colony were analyzed via Coomassie stain and/or Western blot after lysis was completed (Section 3.5c). As seen in Figure 5.2A, our Western blot analysis was able to detect protein bands just below 50kDa for each sample. This weight is close to our predicted molecular weight of ~47kDa and is comparable to what was observed by Levina *et al* (~45kDa). Of the samples tested, MAAT-4 and ZAAT-6 indicated the highest protein expression levels. Therefore they were re-streaked onto YEPD + 0.05mg/mL G418 plates and used in our extraction methods.

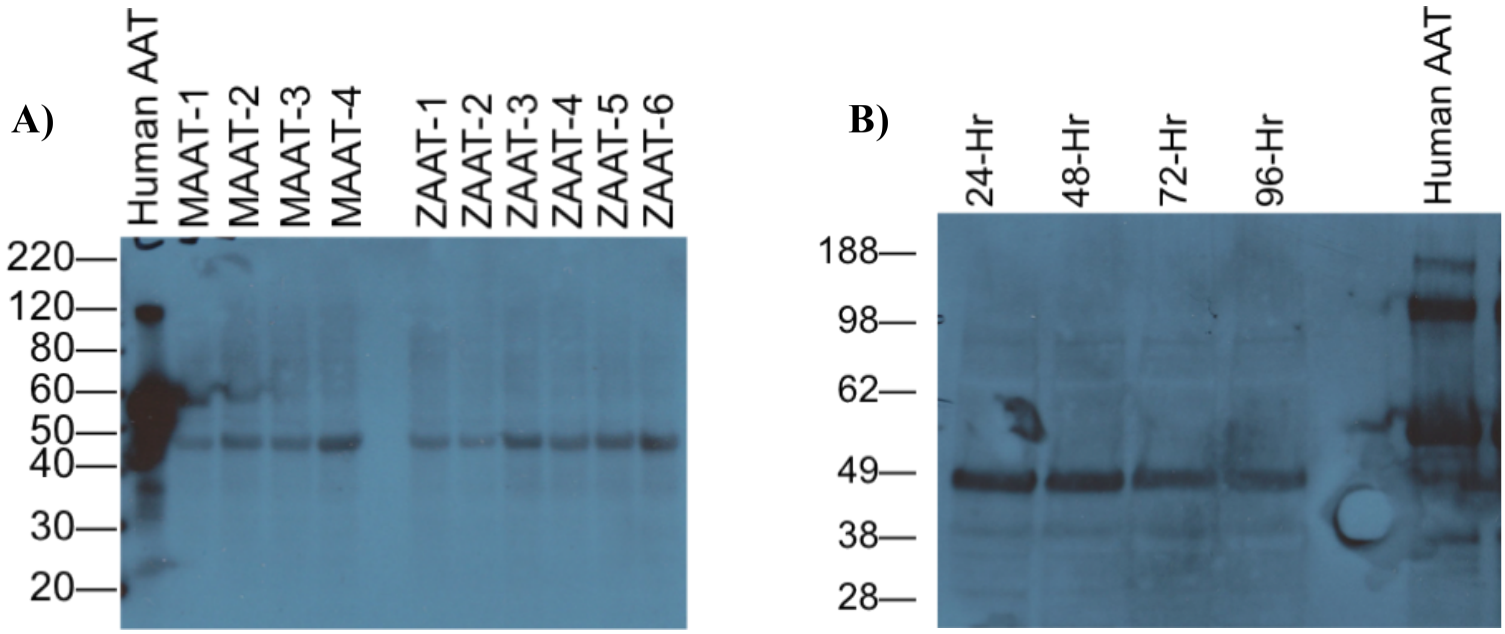


Figure 5.2. Expression, Induction and Harvesting of AAT protein from P. pastoris as analyzed by Western blot. (A) Identification of high expression MAAT and ZAAT clones lysed with Y-PER®. As a result of this analysis, clones MAAT-4 and ZAAT-6 were selected for large-scale expression. (B) Verification of Induction Time for ZAAT-6 protein expression. Cell lysates were prepared using YeastBuster™. Based on this blot, a 48-hour induction time was selected for optimal expression. A polyclonal AAT antibody was used for both Western blots. Human AAT served as a positive control.

b) Verifying length of Induction Time

Previously it was shown that an induction time of 24-hours was needed for optimal expression of MAAT and ZAAT. However this was based on a Western blot that used mouse monoclonal anti-α1-Antitrypsin antibody as the primary antibody (Figure 3.6A). As discussed in Section 3.5b, this antibody ended up not being able to identify our protein of interests. Therefore the induction time needed to be verified using the rabbit polyclonal anti-α1-Antitrypsin antibody.

This was performed as previously described (Section 3.5a), in which high-expression MAAT and ZAAT clones (Figure 5.2A) were first grown overnight (225rpm, 30°C) in 50mL of YEPD and then washed with sterile water. The cells were then resuspended in 50mL of induction media and grown in a shaker (225rpm, 30°C) for a total of 96-hours. After every 24-hours, 10mL of culture was removed, washed with sterile water, and then weighed. Cell pellets were stored at -20°C until all samples were ready for lysis.

The remaining culture received 5% methanol at 1/200 of its volume. This not only helped to replenish the carbon source but also promote expression of our genes though the *AOX1* promoter. Once all the time points were collected, they were lysed and analyzed using Coomassie staining and/or Western blot analysis.

Our previous results indicated that 24-hours was the induction time needed for optimal expression. However, as seen in Figure 5.2B Lane 2, an induction time of 48-hours showed the most expression (band once more at ~49kDa). This was shown to be

true for both ZAAT and MAAT. Therefore the induction time for *P. pastoris* was changed to 48-hours for all future experiments.

c) Preparing Cell Lysates

As described in Section 3.5c, we were able to show that in our hands, the chemical method of cell lysis was more efficient than the mechanical method. In those experiments the Y-PER® reagent was used and seemed to be compatible with the downstream chromatography experiments at the time. However when samples that been extracted using Y-PER® were applied to Ni-NTA columns (Section 5.4a) it was quickly observed that this particular reagent was not compatible with these columns.

Upon the addition of the samples, a discoloration of the resin would occur (would turn brown in color) that would result in most of the protein being eluted directly into the wash (Figure5.3, ~50kDa bands). This discoloration is known to be a sign of Nickel ions being removed or reduced within the resin usually due to reducing agents (i.e. DTT) present in the buffers (154).

Therefore an alternative lysis method was quickly sought out. YeastBuster™ Protein Extraction Reagent (Novagen®), another commercially available lysis reagent, was soon selected based on its known compatibility with the Ni-NTA columns and easy to use protocol (154,155).

Briefly, frozen cell pellets were resuspended in room temperature YeastBuster™ (5mL/g wet cell paste), 100X tris(hydroxypropyl)phosphine solution (THP, 50µL/g wet cell paste) and Protease Inhibitor Cocktail Set III, EDTA-Free (50µL/g wet cell paste).

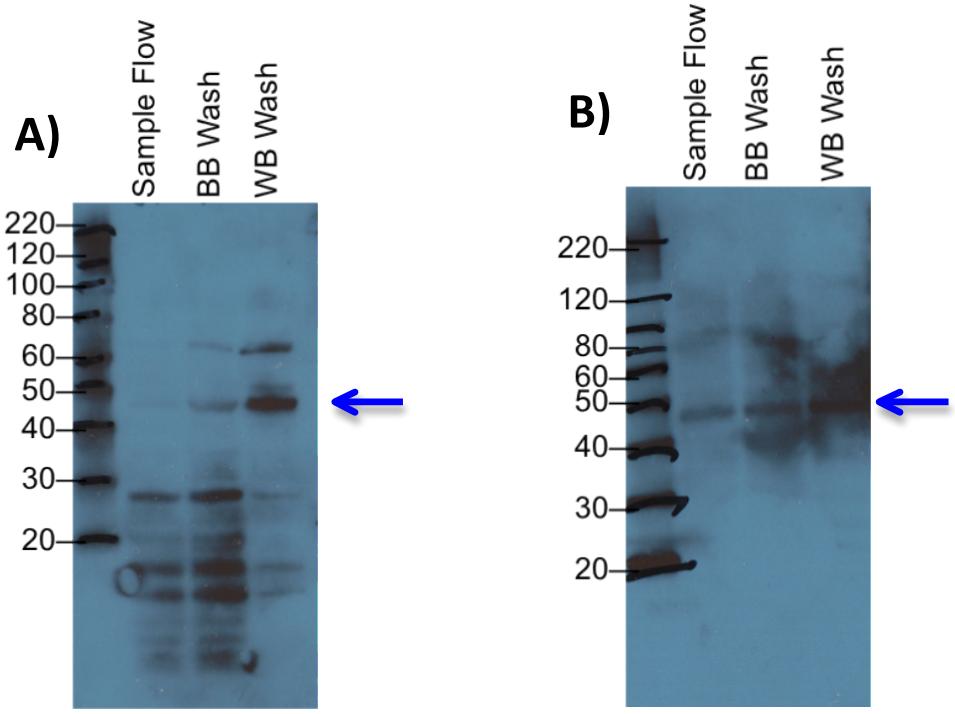


Figure 5.3: Y-PER® Lysed Samples after IMAC Purification. Samples were lysed with the Y-PER® reagent and then purified using Imobilized-metal Affinty Chromatography. The lysis reagent was shown to be incomatible with our columns which resulted in most of the protein being eluted during the washing steps. This was verified by Western blot with an (A) anti-6xHis antibody (1:1000 dilution, Abcam) and a (B) polyclonal AAT antibody (1:2000 dilution, Sigma). Arrows indicate where our proteins were predicated to be located.

The cell suspension was then incubated for 20 minutes at room temperature on either a shaking platform or rotating mixer at a slow setting. Cell debris was removed by centrifugation at 4000xg for 30 minutes at 4°C and the resulting supernatant was transferred to a fresh tube (156). Samples extracted using YeastBuster™ were shown to be comparable to those using Y-PER® (Figure 5.2A versus B).

While we chose to stick to a chemical form of lysis, there are several other alternatives that we could have pursued. The first, and perhaps simplest, would be to lyse the yeast cells with Y-PER® but without the presence of the reducing agent, DTT. While this method was attempted, the results were inconclusive and therefore should be reverified. Another alternative would be to perform a buffer exchange in which the Y-PER® would be replaced with the binding buffer used in the affinity chromatography experiments (Section 5.4).

Lastly, new Ni-NTA columns that are known to be compatible with Y-PER®, such as those provided in the Y-PER® 6xHis Fusion Protein Purification Kit (Pierce), could have been ordered and used for purification. At the time of this study however, it was decided that we should maintain our current columns and instead alter the reagent used for lysis as described above.

d) Large-scale Induction and Harvesting

After the above changes in our methodology were established, MAAT and ZAAT colonies were expressed and extracted from large cultures of 500-1000mL YEPD. This was performed as previously described (Section 3.5d), in which a two-day overnight

culture was produced first. A starter culture of 50mL YEPD was grown overnight and then used to inoculate a second culture in 500-1000mL of YEPD.

Once the final overnight culture was prepared, the cells were washed with sterile water, resupended in Induction media, and induced in a shaker (225rpm, 30°C) for 48-hours. After the first 24-hours, the culture was removed from the shaker so that 5% methanol could be added to it at 1/200 of its volume. The culture was then returned to the shaker for an additional 24-hours of induction. Once the cells had been induced for a total of 48-hours, they were removed from the shaker, washed with sterile and then stored at -20°C until ready for lysis with the YeastBuster™ reagent.

5.4 Protein Purification

Following protein extraction using YeastBuster™, samples were then purified through a combination of Immobilized-Metal Affinity Chromatography and Anion Exchange Chromatography. The techniques implemented during these purification processes will be explained in this section.

a) Immobilized-Metal Affinity Chromatography

As discussed in the introduction of this chapter (Section 5.1), polyhistidine fusion proteins are readily purified using immobilized metals, a technique called Immobilized-Metal Affinity Chromatography (IMAC). This technique exploits the affinity that occurs in an aqueous solution between certain transition metal ions, like Ni^{2+}, and specific amino acids residues, such as histidine (152). Compared to other affinity chromatography techniques, IMAC is not highly specific. Nevertheless, it is often the method of choice

when a combination of high protein purity and quick purification are needed. In general, IMAC has several benefits to its use, including ligand stability, mild elution conditions, high protein loading, and low cost (157).

For our experiments the agarose-based IMAC resin Nickel-Nitriloacetic acid (Ni-NTA) was used to bind and elute the His-tagged MAAT and ZAAT proteins from two different types of columns. The first type was a His-Bind® Column (Novagen®) that was pre-packaged with 1.25mL bed volume of Ni^{2+}-charged His-Bind Resin. The second type involved 1.25mL of Ni-NTA His-Bind Resin (Novagen®) being added to a 5mL pipette that had steel wool previously packed into the tip. The protocol for both of these columns followed the procedure outlined in the His-Bind Kit Manual (155).

First, 20mL of 1x Binding Buffer (H-BB, 50mM NaH_2PO_4, pH 8.0; 300mM NaCl; 10mM imidazole), 10mL of 1x Wash Buffer (H-WB, 50mM NaH_2PO_4, pH 8.0; 300mM NaCl; 20mM imidazole) and 5mL of 1x Elute Buffer (H-EB, 50mM NaH_2PO_4, pH 8.0; 300mM NaCl; 250mM imidazole) were prepared per column. Next the columns were equilibrated with 10mL of H-BB, making sure to allow the entire buffer volume to flow through the column. Once all the H-BB had flowed through the column, the sample was loaded onto the column (up to 10mg of target protein) with the flow through collected from that point forward.

The column was then washed with 10mL of H-BB followed by 10mL of H-WB. AAT protein was then eluted from the column with 5mL of H-EB and captured in 1mL fractions. The fractions and washes that were collected throughout were then ran on a SDS-PAGE gel and analyzed by Coomassie staining and/or Western blot to verify the presence of protein (methods described in Section 3.4).

Initially our protein detection via Coomassie staining was very faint (Figure 5.4A) but it was soon revealed that one of the reasons for this was that we were underwhelming the column and therefore very little protein would be eluted. An example of this is seen in Figure 5.4A fraction B, which resulted in a faint band being identified at ~50kDa (predicted size of our proteins of interest).

Once we figured this out and started to increase our sample load, we were able to better detect our protein on the SDS-PAGE gels. However as seen in Figure 5.4B, when the sample amount was increased, we also increased the amount of contaminants present (bands at ~38kDa). Therefore additional purification steps were necessary.

b) Anion Exchange Chromatography

Previously when we performed anion exchange chromatography using Q-Sepharose (Section 3.6a) we used buffers similar to those involved in purification of AAT from blood plasma (128). However upon closer inspection, and several unsuccessful attempts at purifying tagged protein using this protocol, it was decided that the buffers should be switched to closely mimic those described by Levina *et al* (100,158). Therefore, purification of tagged MAAT and ZAAT using Q-Sepharose chromatography was conducted as follows. First 1.25mL of Q-Sepharose Fast Flow was added to a 5mL pipette that had steel wool previously packed into the tip. Next, 1mL of wash buffer (Q-WB, 50mM Tris-HCL, pH8.0 and 300mM NaCl) was allowed to run through the column followed by two 1mL applications of the equilibrate buffer (Q-EB, 50mM Tri-HCL, pH 8.0, and 75mM NaCl).

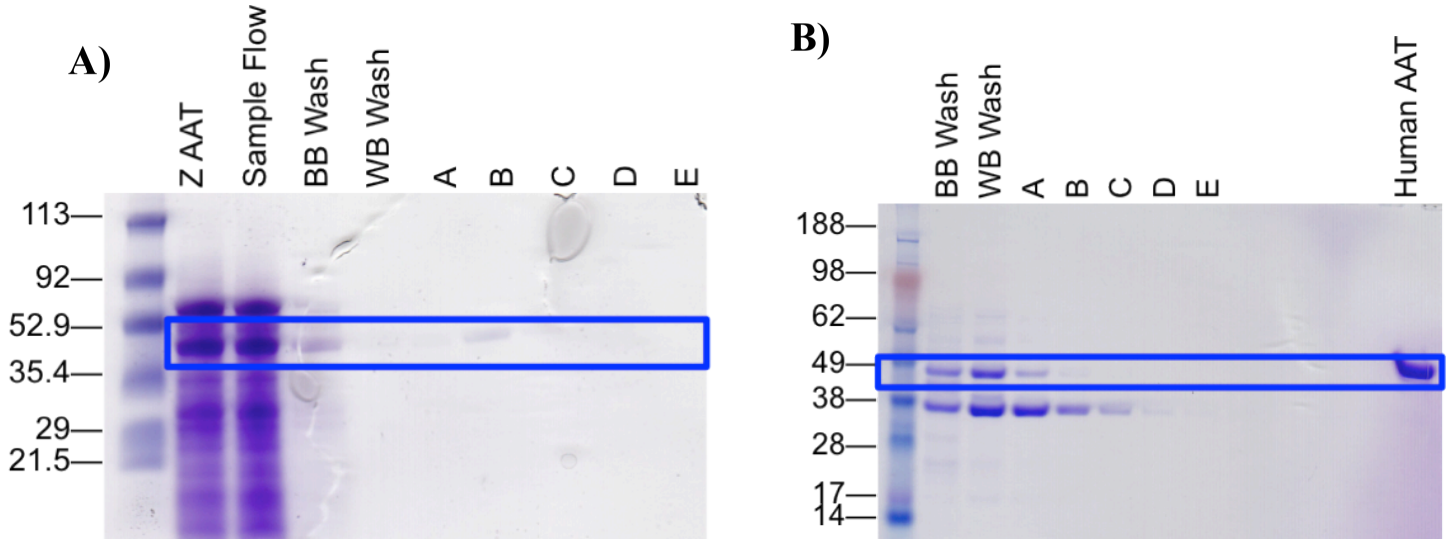


Figure 5.4 IMAC Purification of ZAAT sample. (A) During initial attempts at IMAC purification, small amounts of sample were applied to the column and were therefore underwhelming it. This prevented us from being able to detect our protein very clearly by Coomassie staining. (B) However once we increased the sample amount, we were able to better detect the protein bands. Unfortunately this also increased the amount of contaminants detected as well (bands ~38kDa). Blue box indicates with protein of interest was predicted to be located.

Once the column was equilibrated, the sample (diluted 1:3 with Q-EB) was added to the column and the flow through was collected from that point forward. The column was then washed twice with 1mL Q-EB and then treated with a salt gradient of 75, 100, 125, 150 and 175mM NaCl dissolved in the Q-EB. The gradient was added in two 1mL applications for each NaCl concentration, followed by four 1mL applications of the Q-WB.

Fractions and washes collected throughout the chromatography experiment were separated via SDS-PAGE, analyzed by Coomassie staining (Figure 5.5A) and then confirmed by Western blotting (Figure 5.5B). Although both the Coomassie stained gel and the Western blot showed bands at ~50kDa (Lanes 2-4 and 6), the positive control (human AAT) in the same gels was calculated to be ~75kDa (Figure 5.5A, Lane 10), well above the predicted mass (52kDa).

Fearing that that something could have occurred during electrophoresis that would result in the weight discrepancy, the samples were run again on a new SDS-PAGE gel. However the same phenomenon occurred in the new gels as well. This could mean that there was something wrong with our positive control (such as aggregation) and/or there was an issue with the molecular weight ladder used.

Using new human AAT as our positive control or changing the molecular weight ladder used in our SDS-PAGE set-up could help to rule out these two possibilities. More than likely the proteins at this molecular weight (~50kDa) are *Pichia pastoris* that similar sequence to our AAT proteins and are therefore recognized by our polyclonal antibody as discussed previously in Section 3.6b.

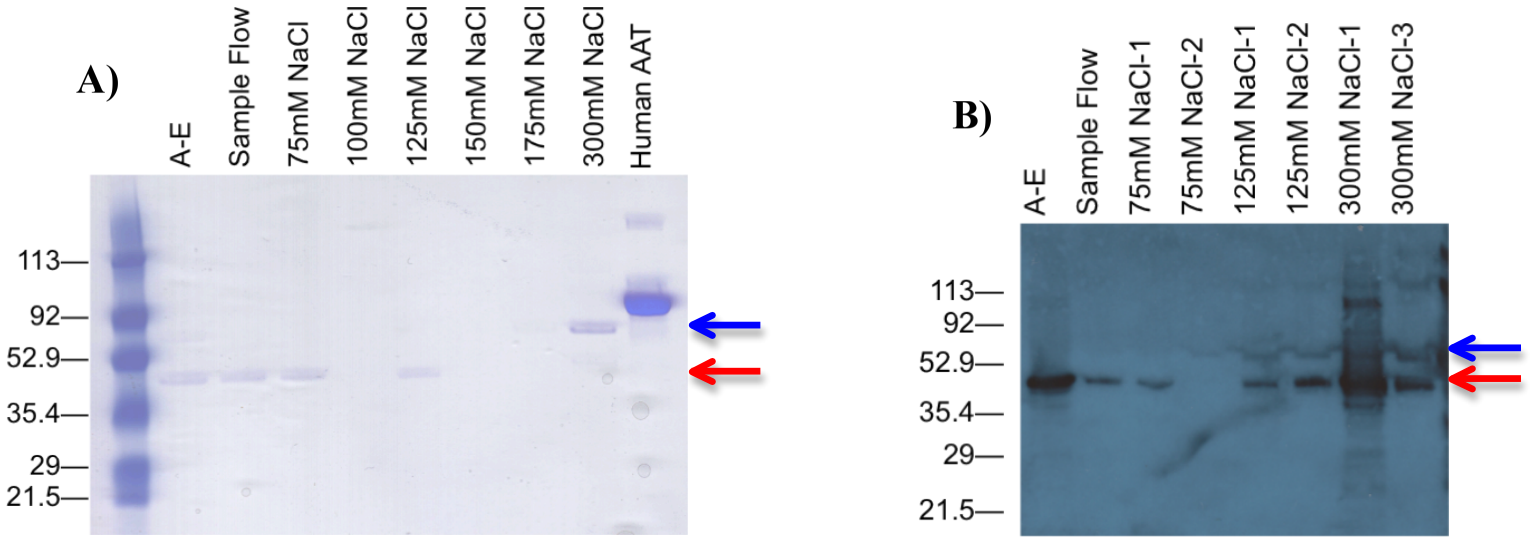


Figure 5.5 Anion Exchange Purification. (A) Coomassie staining and (B) Western blot analysis of samples collected after IMAC and Q-Sepharose chromatography. Letters A-E represent fractions collected during IMAC whereas #mM NaCl indicates Q-Sepharose samples. Blue arrows indicate area where protein is predicted to be located and red arrows reference contaminants.

It should be noted that additional bands (~70kDa) were also observed by Coomassie staining (Figure 5.5A) for our IMAC combined fractions (Lane 2) and our 300mM NaCl elution (Lane 9). Since these bands are closer in molecular weight to our positive control, it is likely that this band is our protein of interest. However since our salt gradient jumped from 175mM NaCl to 300mM NaCl, it is unclear at what concentration the protein was actually eluted from. In addition the presence of several bands in our 300mM NaCl elutions (Figure 3.5B, Lanes 7 and 8) could indicate that we have purified aggregated and monomeric AAT protein together at this concentration. This theory is supported by Levina *et al*, in which their study eluted aggregated protein from 270-290mM NaCl (100). Therefore the experiment would need to be repeated in order to fully cover the missing range and pinpoint the necessary elution concentration.

5.5 Discussion

To summarize the procedures outlined in this chapter, cDNA for both MAAT and ZAAT were re-synthesized to include a 6xHis-tag and TEV cleavage site on the N-terminus. Like the previous constructs, the new cDNA also contained several factors needed throughout the expression and purification processes (Section 5.2a). The cDNA was then subcloned into the pPic3.5K vector (Section 5.2b) and transformed into the SMD1163 strain of *P. pastoris* through electroporation. Following electroporation, cells with multiple copy integration events were selected based on their resistance to the antibiotic, G418. SMD1163 clones that expressed the highest expression levels were then induced for 48-hours followed by protein extraction using YeastBuster™ (Section

5.3). Finally, MAAT and ZAAT proteins were purified by Immobilized-Metal Affinity Chromatography followed by Anion Exchange Chromatography (Section 5.4).

While it is plausible that we were successful at purifying MAAT and ZAAT protein, there are still several conditions that are needed to optimized/ altered before confirming the presence of our protein. The first would be to increase the amount of yeast cells grown prior to lysis. With the methods discussed in this thesis, we were only obtaining 20-30g of cells per liter, which is a low cell density for *Pichia pastoris*. Fermentation, which can increase cell densities in *P. pastoris* to as high as 130g/L (69,70), could be implemented In order to address the issue. By increasing the cell density we would, in theory, also raise the amount of protein produced, thereby helping to enhance our overall yield.

Another thing that could potentially help to boost our overall protein yield would be to increase the amount of Induction media used so that it is equal to the volume of the overnight cultures. Currently the induction media used is only 250mL, which is 25-50% of the overnight culture volume. Therefore increasing the induction media volume would help to ensure that all the yeast cells are induced equally.

Various aspects of the IMAC purification process would also need to be optimized, the first of which would be to alter the buffer compositions. The buffers used for IMAC purification were from the Ni-NTA buffer kit (Novagen®) and differed from those used by Levina *et al*. In comparison, our wash buffer contained less NaCl (300mM vs. 500mM) and slightly less imidazole (20mM vs. 25mM) then was reported in the paper. In addition their elution buffer contained twice the amount of imidazole (250mM vs. 500mM) and half the amount of NaCl (300mM vs.150mM). Since NaCl is

used to prevent ionic interactions and imidazole is used to bind to Ni-NTA and compete with histidine residues in the 6xHis-tag (154), the differences in the buffers could account for some of the obstacles observed during IMAC purification. This would be especially true for the elution buffer in which our conditions might have been to mild for proper elution due the lower imidazole concentration. Such conditions could result in our His-tagged protein to remain bound to the column and therefore reduce the amount seen in our fractions.

Conversely an alternative explanation as to why we observed small amounts of protein being eluted from our Ni-NTA columns is that our 6xHis-tag was hidden, either fully or partially, within the protein. This would explain why we were able to still detect our proteins of interests (~50kb) in the washes even after changing the lysis method (Figure 5.4). To verify that our protein is indeed binding to the column, we could first run the sample through the column as illustrated in Section 5.4a and then cleave the 6xHis-tag using a TEV protease. As previously described, the protease would remove the tag and leave behind a single amino acid residue, causing our proteins to be eluted into the wash during a second application of IMAC. This process would also help to remove contaminants eluted with our protein during the first application. If this proves unsuccessful, the proteins could alternatively be purified under denaturing conditions or the 6xHis-tag could be moved to the C-terminal of the proteins (154).

Lastly, our anion exchange chromatography methods would need to be optimized before we could be confident of protein purification. As discussed in Section 5.4b, it appears as if our protein was eluted at 300mM NaCl, but only after jumping from

175mM NaCl buffer. Therefore, it would be necessary to attempt the described experiments again but making sure to account for missing range of NaCl concentration.

It should be noted however, that even if the proteins elute somewhere between 175-300mM NaCl, it would still be outside of the range described by Levina *et al.* According to their publication, monomeric AAT was eluted between 100-150mM NaCl, cleaved protein was eluted at 150mM (ZAAT) or 190mM (MAAT) and polymers were eluted at 270-290mM NaCl (100).

One explanation for why our proteins seem to be eluting at higher salt concentrations than what has been reported could simply be that we are not adding adequate amount of buffers though the column. Currently we were only applying two bed volumes of buffer to the column at each salt concentration. Increasing this to at least 5 bed volumes would be to ensure proper elution of proteins at each concentration (124). Another reason could be because the samples are transferred directly from a phosphate-based buffer during IMAC purification to a Tris-based buffer in Q-sepharose chromatography. A consistent buffer base (either all phosphate-based or all Tris-based) for our buffers or a buffer exchange between the chromatography steps could help eliminate this issue.

Once all of these conditions have been optimized/altered, then we will be able to confirm if the process we have outlined in this thesis would be successful at producing and purifying active, and monomeric MAAT and ZAAT proteins.

CHAPTER 6:

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

Ever since the connection between Alpha1-Antritrypsin Deficiency and α1-Antritrypsin (AAT) polymerization was established, there has existed a high demand for recombinant AAT protein. This is especially true for the Z variant of AAT, which is very prone to polymerization and thereby difficult to produce recombinantly. Over the years several systems have been implemented to produce wild-type MAAT protein but it was not until 2009 that recombinant ZAAT protein was successfully produced. The goal of this thesis was to expand upon this original study in order to produce both untagged and tagged versions of recombinant ZAAT protein using the methylotrophic yeast, *Pichia pastoris*.

The expression and purification of untagged recombinant MAAT and ZAAT protein was attempted first (Chapter 3). While the techniques used throughout this process ended up being insufficient at producing protein at the quantity or purity needed, they did lay the foundation for protein purification procedures implemented throughout this thesis. These methods included *in vivo* screening of multiple gene inserts, cell lysate preparation as well as induction and harvesting techniques.

Following the unsuccessful attempts at untagged protein purification, focus shifted to the production of His-tagged recombinant ZAAT and MAAT protein. Initially the addition of the 6xHis-tag was attempted via Site Directed Mutagenesis (Chapter 4).

However, after a series of complications and the usage of two SDM methods, it was decided that cDNA for MAAT and ZAAT would be resynthesized to include a 6xHis-tag. Using these new constructs, the SMD1163 strain of *P. pastoris* was transformed by electroporation and selected for multiple-copy insertions based on resistance to the antibiotic, G418. Protein expression was induced for 48-hours prior to harvesting with a chemical reagent. Our proteins were then purified though a combination of affinity chromatography and anion exchange chromatography methods (Chapter 5).

Nonetheless several complications arose throughout this process, particularly during the purification steps. These issues thereby prevent us from being able to state with confidence that we obtained purified AAT protein. Some methods that can be implemented to overcome these obstacles have been described throughout this thesis and will also be reviewed in the next section. Overall, the techniques used and suggested throughout this thesis provide a foundation for future endeavors of AAT expression and purification from *Pichia pastoris*.

6.2 Future Directions

While we have made progress towards expressing and purifying AAT proteins from *Pichia pastoris,* there are several areas within our methodology that would benefit from optimization or modification. Such changes would us to determine if our techniques are actually successful or not. For example, switching the yeast cells to a fermentor instead of growing in a beveled flask in a shaker would increase the amount of cells available and therefore amplify the amount of protein that can be purified. In addition, increasing the amount of Induction media used may also help to enhance protein yields.

As stated before, many of our obstacles occurred during the purification process. Some things that could be attempted to overcome concerns occurring during affinity chromatography include altering the buffers to more closely mimic those used by Levina *et al,* purifying under denaturing conditions, and moving the 6xHis-tag to the C-terminus. Another option would be to run the samples through the affinity chromatography column twice. Prior to the second round however, samples would be cleaved with TEV protease in order to remove the 6xHis-tag, thereby causing the protein of interest to elute into the wash and separating it from any contaminants.

As for complications that arose during anion-exchange chromatography, altering the salt gradient concentrations so that they fully cover the 75-300mM NaCl range and/or increasing the buffer bed volume applied to the column might help to address these obstacles. Lastly changing the buffers used throughout the purification process so that they are all either phosphate-based or Tris-based may help to elevate some of the problems encountered in this thesis.

Once purification of heterologous MAAT and ZAAT protein from *P. pastoris* is shown to be successful, the structural, inhibitory, and polymerization characteristics of the proteins will need to be verified. These characterizations will help to determine if the proteins produced display the same attributes seen in AAT protein purified from Human patient plasma.

LIST OF REFERENCES

1. Serpins, Serpinopathies, and Conformational Diseases. in *PeproTech Focus* (www.peprotech.com ed., PeproTech Inc., Rocky Hill, NJ
2. Law, R. H. P., Zhang, Q. W., McGowan, S., Buckle, A. M., Silverman, G. A., Wong, W., Rosado, C. J., Langendorf, C. G., Pike, R. N., Bird, P. I., and Whisstock, J. C. (2006) *Genome Biol.* **7**
3. Khan, M. S., Singh, P., Azhar, A., Naseem, A., Rashid, Q., Kabir, M. A., and Jairajpuri, M. A. (2011) *Journal of Amino Acids* **2011**
4. Irving, J. A., Pike, R. N., Lesk, A. M., and Whisstock, J. C. (2000) *Genome Research* **10**, 1845-1864
5. Gooptu, B., and Lomas, D. A. (2009) *Annu. Rev. Biochem.* **78**, 147-176
6. Lomas, D. A., and Carrell, R. W. (2002) *Nat Rev Genet* **3**, 759-768
7. Lomas, D. A. (2005) *Clin. Med.* **5**, 249-257
8. Belorgey, D., Hagglof, P., Karlsson-Li, S., and Lomas, D. A. (2007) *Prion* **1**, 15-20
9. Mottonen, J., Strand, A., Symersky, J., Sweet, R. M., Danley, D. E., Geoghegan, K. F., Gerard, R. D., and Goldsmith, E. J. (1992) *Nature* **355**, 270-273
10. Carrell, R. W., Stein, P. E., Fermi, G., and Wardell, M. R. (1994) *Structure* **2**, 257-270
11. Gooptu, B. (2000) *Proc. Natl Acad. Sci. USA* **97**, 67-72
12. Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) *Nature* **407**, 923-926
13. Lawrence, D. A., Olson, S. T., Palaniappan, S., and Ginsburg, D. (1994) *The Journal of biological chemistry* **269**, 27657-27662
14. Chang, W. S. W., Wardell, M. R., Lomas, D. A., and Carrell, R. W. (1996) *Biochem. J.* **314**, 647-653
15. Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J., and Whisstock, J. C. (2001) *The Journal of biological chemistry* **276**, 33293-33296
16. Jeppsson, J. O. (1976) *FEBS letters* **65**, 195-197
17. Krem, M. M., and Di Cera, E. (2003) *The Journal of biological chemistry* **278**, 37810-37814
18. Stein, P. E., and Carrell, R. W. (1995) *Nat Struct Biol* **2**, 96-113
19. Dunstone, M. A. (2000) *Protein Sci.* **9**, 417-420
20. Belorgey, D., Irving, J. A., Ekeowa, U. I., Freeke, J., Roussel, B. D., Miranda, E., P√©rez, J., Robinson, C. V., Marciniak, S. J., Crowther, D. C., Michel, C. H., and Lomas, D. A. (2011) *Methods* **53**, 255-266
21. Owen, M. C., Brennan, S. O., Lewis, J. H., and Carrell, R. W. (1983) *The New England journal of medicine* **309**, 694-698
22. Rijken, D. C., Groeneveld, E., Kluft, C., and Nieuwenhuis, H. K. (1988) *The Biochemical journal* **255**, 609-615
23. Corral, J., Aznar, J., Gonzalez-Conejero, R., Villa, P., Minano, A., Vaya, A., Carrell, R. W., Huntington, J. A., and Vicente, V. (2004) *Circulation* **110**, 1303-1307
24. Hidvegi, T., Mirnics, K., Hale, P., Ewing, M., Beckett, C., and Perlmutter, D. H. (2007) *Journal of Biological Chemistry* **282**, 27769-27780

25. Verbanac, K. M., and Heath, E. C. (1986) *Journal of Biological Chemistry* **261**, 9979-9989
26. Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P., and Lomas, D. A. (1999) *J. Biol. Chem.* **274**, 9548-9555
27. Lomas, D. A., and Parfrey, H. (2004) *Thorax* **59**, 529-535
28. Kalsheker, N., and Morgan, K. (1990) *Thorax* **45**, 759-764
29. Yu, M. H., Lee, K. N., and Kim, J. (1995) *Nature Structural Biology* **2**, 363-367
30. Laurell, C. B., and Eriksson, S. (1963) *Scand. J. Clin. Lab. Invest.* **15**, 132-140
31. Witko-Sarsat, V., Rieu, P., Descamps-Latscha, B., Lesavre, P., and Halbwachs-Mecarelli, L. (2000) *Laboratory investigation; a journal of technical methods and pathology* **80**, 617-653
32. Chua, F., and Laurent, G. J. (2006) *Proceedings of the American Thoracic Society* **3**, 424-427
33. Vaughan, L., Lorier, M. A., and Carrell, R. W. (1982) *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **701**, 339-345
34. Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H.-P., and Altmann, F. (2006) *PROTEOMICS* **6**, 3369-3380
35. Blanchard, V., Liu, X., Eigel, S., Kaup, M., Rieck, S., Janciauskiene, S., Sandig, V., Marx, U., Walden, P., Tauber, R., and Berger, M. (2011) *Biotechnology and Bioengineering* **108**, 2118-2128
36. Fregonese, L., and Stolk, J. (2008) *Orphanet J. Rare Dis.* **3**, 9
37. Kiernan, V. (1995) Warm hearts in a cold land. in *New Scientist*
38. Kuzemko, J. A. (1994) *Journal of the Royal Society of Medicine* **87**, 769-772
39. Kubba, A. K., and Young, M. (1998) *Chest* **113**, 210-216
40. Carrell, R. W. (2004) *Copd* **1**, 71-84
41. Gooptu, B., Ekeowa, U. I., and Lomas, D. A. (2009) *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **34**, 475-488
42. de Serres, F. J. (2002) *Chest* **122**, 1818-1829
43. Kohnlein, T., and Welte, T. (2008) *The American journal of medicine* **121**, 3-9
44. Parmar, J. S., Mahadeva, R., Reed, B. J., Farahi, N., Cadwallader, K. A., Keogan, M. T., Bilton, D., Chilvers, E. R., and Lomas, D. A. (2002) *American journal of respiratory cell and molecular biology* **26**, 723-730
45. Harb, R., and Thomas, D. W. (2007) *Pediatrics in Review* **28**, 83-91
46. Genetic Science Learning Center. (2012) "Alpha-1 Antitrypsin Deficiency". The University of Utah, Learn Genetics
47. Bottomley, S. P. (2010) *Proceedings of the American Thoracic Society* **7**, 404-407
48. Knaupp, A. S., Levina, V., Robertson, A. L., Pearce, M. C., and Bottomley, S. P. (2010) *J Mol Biol* **396**, 375-383
49. Luisetti, M., and Seersholm, N. (2004) *Thorax* **59**, 164-169
50. Carrell, R. W., and Lomas, D. A. (2002) *New England Journal of Medicine* **346**, 45-53
51. Bruch, M., Weiss, V., and Engel, J. (1988) *The Journal of biological chemistry* **263**, 16626-16630
52. James, E. L., and Bottomley, S. P. (1998) *Archives of biochemistry and biophysics* **356**, 296-300

53. Roussel, B. D., Irving, J. A., Ekeowa, U. I., Belorgey, D., Haq, I., Ordonez, A., Kruppa, A. J., Duvoix, A., Rashid, S. T., Crowther, D. C., Marciniak, S. J., and Lomas, D. A. (2011) *The FEBS journal* **278**, 3859-3867
54. Yamasaki, M., Li, W., Johnson, D. J. D., and Huntington, J. A. (2008) *Nature* **455**, 1255-1258
55. Ekeowa, U. I., Freeke, J., Miranda, E., Gooptu, B., Bush, M. F., Perez, J., Teckman, J., Robinson, C. V., and Lomas, D. A. (2010) *Proc Natl Acad Sci U S A* **107**, 17146-17151
56. Miranda, E., Perez, J., Ekeowa, U. I., Hadzic, N., Kalsheker, N., Gooptu, B., Portmann, B., Belorgey, D., Hill, M., Chambers, S., Teckman, J., Alexander, G. J., Marciniak, S. J., and Lomas, D. A. (2010) *Hepatology* **52**, 1078-1088
57. Bottomley, S. P. (2011) *EMBO Rep* **12**, 983-984
58. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C., and Huntington, J. A. (2011) *EMBO Rep* **12**, 1011-1017
59. Alpha-1 Foundation. (2012) Alphas, Friends & Family.
60. Alpha-1 Association. (2012) Newly-Diagnosed.
61. Ekeowa, U. I., Gooptu, B., Belorgey, D., Hagglof, P., Karlsson-Li, S., Miranda, E., Perez, J., MacLeod, I., Kroger, H., Marciniak, S. J., Crowther, D. C., and Lomas, D. A. (2009) *Clin. Sci.* **116**, 837-850
62. Chow, M. K., Devlin, G. L., and Bottomley, S. P. (2001) *Biological chemistry* **382**, 1593-1599
63. Burrows, J. A., Willis, L. K., and Perlmutter, D. H. (2000) *Proc Natl Acad Sci U S A* **97**, 1796-1801
64. Devlin, G. L., Parfrey, H., Tew, D. J., Lomas, D. A., and Bottomley, S. P. (2001) *American journal of respiratory cell and molecular biology* **24**, 727-732
65. Sharp, L. K., Mallya, M., Kinghorn, K. J., Wang, Z., Crowther, D. C., Huntington, J. A., Belorgey, D., and Lomas, D. A. (2006) *The FEBS journal* **273**, 2540-2552
66. Teckman, J. H. (2004) *Journal of pediatric gastroenterology and nutrition* **39**, 34-37
67. Mahadeva, R., Dafforn, T. R., Carrell, R. W., and Lomas, D. A. (2002) *J. Biol. Chem.* **277**, 6771-6774
68. Parfrey, H., Mahadeva, R., Ravenhill, N. A., Zhou, A., Dafforn, T. R., Foreman, R. C., and Lomas, D. A. (2003) *The Journal of biological chemistry* **278**, 33060-33066
69. Palomares, L., Estrada-Mondaca, S., and Ramírez, O. (2004) Production of Recombinant Proteins Recombinant Gene Expression. in *Recombinant Gene Expression* (Balbás, P., and Lorence, A. eds.), Humana Press. pp 15-51
70. Demain, A. L., and Vaishnav, P. (2009) *Biotechnology advances* **27**, 297-306
71. Terpe, K. (2006) *Applied microbiology and biotechnology* **72**, 211-222
72. GenWay Biotech Inc. (2012) Protein Expression Considerations. in *Protein Expression* San Diego, CA
73. Srinivasan, S., Barnard, G. C., and Gerngross, T. U. (2002) *Appl Environ Microbiol* **68**, 5925-5932
74. Barnard, G. C., Henderson, G. E., Srinivasan, S., and Gerngross, T. U. (2004) *Protein Expression and Purification* **38**, 264-271
75. (2012) Platform Technology. Pfēnex Inc., San Diego, CA
76. Dequin, S. (2001) *Applied microbiology and biotechnology* **56**, 577-588
77. Kjeldsen, T. (2000) *Applied microbiology and biotechnology* **54**, 277-286

78. Higgins, D. R. (2001) Overview of Protein Expression in Pichia pastoris. in *Current Protocols in Protein Science*, John Wiley & Sons, Inc. pp
79. Burrowes, O. J., Diamond, G., and Lee, T. C. (2005) *Journal of Biomedicine and Biotechnology*, 374-384
80. Cregg, J. The Pichia System. Research Corp. Technologies, Tucson, AZ
81. Lin Cereghino, G. P., Sunga, A. J., Lin Cereghino, J., and Cregg, J. M. (2001) *Genetic engineering* **23**, 157-169
82. Agathos, S. N. (1991) *Biotechnology advances* **9**, 51-68
83. Jarvis, D. L. (2009) BACULOVIRUS-INSECT CELL EXPRESSION SYSTEMS. in *Guide to Protein Purification, Second Edition* (Burgess, R. R. D. M. P. ed.). pp 191-222
84. Luckow, V. A., and Summers, M. D. (1988) *Nat Biotech* **6**, 47-55
85. Miller, L. K. (1988) *Annual Review of Microbiology* **42**, 177-199
86. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010) *Journal of Structural Biology* **172**, 45-54
87. (2012) Insect Cells/Baculovirus: Protein expression in the Baculovirus system. in *Protein Expression and Purification Core Facility: Protein Expression*, European Molecular Biology Laboratory
88. Wurm, F. M. (2004) *Nat Biotech* **22**, 1393-1398
89. Ghaedi, M., Lotfi, A. S., Soleimani, M., Shamsara, M., Arjmand, S., and Adibi, B. (2009) *Iranian Journal of Biotechnology* **7**, 148-156
90. Hacker, D. L., Nallet, S., and Wurm, F. M. (2008) Recombinant Protein Production Yields from Mammalian Cells: Past, Present, and Future. in *BioPharm International* Advanstar Communications Inc
91. Parfrey, H., Mahadeva, R., and Lomas, D. A. (2003) *International Journal of Biochemistry & Cell Biology* **35**, 1009-1014
92. Travis, J., Owen, M., George, P., Carrell, R., Rosenberg, S., Hallewell, R. A., and Barr, P. J. (1985) *The Journal of biological chemistry* **260**, 4384-4389
93. Kwon, K. S., Song, M., and Yu, M. H. (1995) *Journal of biotechnology* **42**, 191-195
94. Kwon, K. S., Lee, S., and Yu, M. H. (1995) *Biochimica et biophysica acta* **1247**, 179-184
95. Kang, H. A., Sohn, J. H., Choi, E. S., Chung, B. H., Yu, M. H., and Rhee, S. K. (1998) *Yeast* **14**, 371-381
96. Terashima, M., Murai, Y., Kawamura, M., Nakanishi, S., Stoltz, T., Chen, L., Drohan, W., Rodriguez, R. L., and Katoh, S. (1999) *Applied microbiology and biotechnology* **52**, 516-523
97. McDonald, K. A., Hong, L. M., Trombly, D. M., Xie, Q., and Jackman, A. P. (2005) *Biotechnology progress* **21**, 728-734
98. Trexler, M. M., McDonald, K. A., and Jackman, A. P. (2005) *Biotechnology progress* **21**, 321-328
99. Hasannia, S., Lotfi, A. S., Mahboudi, F., Rezaii, A., Rahbarizadeh, F., and Mohsenifar, A. (2006) *Biotechnology letters* **28**, 1545-1550
100. Levina, V., Dai, W., Knaupp, A. S., Kaiserman, D., Pearce, M. C., Cabrita, L. D., Bird, P. I., and Bottomley, S. P. (2009) *Protein Expression and Purification* **68**, 226-232
101. Scorer, C. A., Clare, J. J., McCombie, W. R., Romanos, M. A., and Sreekrishna, K. (1994) *Nat Biotech* **12**, 181-184

102. (2010) pPic3.5K/pAO815: *Pichia* vectors for multicopy integration and intracellular expression (User Manual). Invitrogen by Life Technologies, Carlsbad, CA
103. (2006) Pichia pastoris Expression System (User Manual). Invitrogen by Life Sciences, Carlsbad, CA
104. Blue Heron Biotech. (2012) Blue Heron pUC MinusMCS
105. Slish, D. F. (2009) Bacterial Transformation. State University of New York at Plattsburgh, Plattsburgh, New York
106. (2006) MAX Efficiency® DH5α™ Competent Cells (User Manual). Invitrogen by Life Technologies, Carlsbad, CA
107. (2012) S.N.A.P.™ MidiPrep Kit (Instruction Manual). Invitrogen by Life Technologies, Carisbad, CA
108. Sambrook, J., and Russell, D. W. (2001) Protocol 17: Directional Cloning into Plasmid Vectors. in *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press. pp
109. Blanchard, V., Gadkari, R. A., Gerwig, G. J., Dighe, R. R., and Kamerling, J. P. (2006) N-Glycosylation studies on glycoprotein hormones expressed in two methylotrophic yeast Pichia pastoris strains, GS115 and X-33 in *Structural Features of Free and Covalently Bound Glycans*, Igitur, Utrecht Publishing & Archiving Services. pp
110. (2010) Multi-Copy Pichia Expression Kit: For the Isolation and Expression of Recombinant Proteins from Pichia pastoris Strains Containing Multiple Copies of a Particular Gene (User Manual). Invitrogen by Life Technologies, Carlsbad, CA
111. Wu, S. X., and Letchworth, G. J. (2004) *Biotechniques* **36**, 152-154
112. Hsieh, C.-M., Belk, J. P., Perez, J. M., and Benatuil, L. (2009) Methods for transforming yeast. C40B 50/06 20060101 C40B050/06; C12N 1/19 20060101 C12N001/19 Ed., U.S.A.
113. Cina, J. (2007) High Yield Protein Production From Pichia Pastoris Yeast: A Protocol For Benchtop Fermentation. in *Application Notes* (New Brunswick Scientific ed., Biocompare:
114. Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K., and Romanos, M. A. (1991) *Biotechnology (N Y)* **9**, 455-460
115. Cregg, J. M., Vedvick, T. S., and Raschke, W. C. (1993) *Biotechnology (N Y)* **11**, 905-910
116. Weiss, H. M., Haase, W., Michel, H., and Reilander, H. (1998) *Biochemical Journal* **330**, 1137-1147
117. Manns, J. M. (2005) SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins. in *Current Protocols in Microbiology*, John Wiley & Sons, Inc. pp
118. (2006) Acrylamide gel protein detection methods. (Center, U. o. M.-C. P. ed.
119. Chevallet, M., Luche, S., and Rabilloud, T. (2006) *Nat. Protocols* **1**, 1852-1858
120. (2005) SilverSNAP® Stain Kit II (User Manual). Pierce Biotechnology, Inc, Rockford, IL
121. Magi, B., and Liberatori, S. (2005) Immunoblotting Techniques Immunochemical Protocols. in *Immunochemical Protocols* (Burns, R. ed.), Humana Press. pp 227-253
122. (2010) NuPAGE® Technical Guide: General information and protocols for using the NuPAGE® electrophoresis system. Invitrogen by Life Technologies, Carlsbad, CA
123. (2008) SNAP i.d. Protein Detection System User Manual. Millipore Corporation

124. (2009) *Recombinant Protein Purification Handbook: Principles and Methods*, GE Healthcare Bio-Sciences AB, Sweden
125. (2012) Principles of Ion Exchange Chromatography. in *Technical Support*, Tosoh Bioscience LLC, King of Prussia, PA
126. (2012) Ion Exchange Chromatography (IEX). in *Technology Insights*, GE Healthcare Life Sciences
127. Reh, G., Nerli, B., and Picó, G. (2002) *Journal of Chromatography B* **780**, 389-396
128. Lomas, D. A., Evans, D. L., Stone, S. R., Chang, W. S. W., and Carrell, R. W. (1993) *Biochemistry* **32**, 500-508
129. (2012) Membrane Selection. in *Immunodetection*, EMD Millipore Corporation, Billerica, MA
130. Hagel, L. (2001) Gel-Filtration Chromatography. in *Current Protocols in Protein Science*, John Wiley & Sons, Inc. pp
131. (2012) Superdex 75 GL. in *Gel Filtration Chromatography Products*, GE Healthcare Life Sciences
132. Ling, M. M., and Robinson, B. H. (1997) *Anal Biochem* **254**, 157-178
133. Zhang, B.-z., Zhang, X., An, X.-p., Ran, D.-l., Zhou, Y.-s., Lu, J., and Tong, Y.-g. (2009) *Journal of Zhejiang University - Science B* **10**, 479-482
134. Braman, J., Papworth, C., and Greener, A. (1996) *Methods Mol Biol* **57**, 31 - 44
135. Ke, S. H., and Madison, E. L. (1997) *Nucleic Acids Research* **25**, 3371-3372
136. Wang, W. Y., and Malcolm, B. A. (1999) *Biotechniques* **26**, 680-682
137. (2004) *Nat. Methods* **1**, 181-182
138. Zheng, L., Baumann, U., and Reymond, J. (2004) *Nucleic Acids Res* **32**, e115
139. Chapnik, N., Sherman, H., and Froy, O. (2008) *Anal Biochem* **372**, 255 - 257
140. Liu, H., and Naismith, J. (2008) *BMC Biotechnol* **8**, 91
141. Edelheit, O., Hanukoglu, A., and Hanukoglu, I. (2009) *BMC Biotechnology* **9**, 61
142. (2010) QuikChange II XL Site-Directed Mutagenesis Kit: Instruction Manual. Agilent Technologies, La Jolla, CA
143. Soderling, S. H. Site-Directed Mutagensis. Duke University, Durham, North Carolina
144. (2012) Non-Specific Bands or Primer-Dimers. in *Conventional PCR Doctor*, Bio-Rad Laboratories, Inc.
145. (2011) PCR Problems and Solutions, PCR Help. in *PCR Troublshooting*, PCR Station
146. Korbie, D. J., and Mattick, J. S. (2008) *Nat. Protocols* **3**, 1452-1456
147. Infante, R. E., Abi-Mosleh, L., Radhakrishnan, A., Dale, J. D., Brown, M. S., and Goldstein, J. L. (2008) *The Journal of biological chemistry* **283**, 1052-1063
148. Infante, R. E., Radhakrishnan, A., Abi-Mosleh, L., Kinch, L. N., Wang, M. L., Grishin, N. V., Goldstein, J. L., and Brown, M. S. (2008) *The Journal of biological chemistry* **283**, 1064-1075
149. (2012) Troublshooting Suggestions. in *DNA Sequencing*, Duke University Health System DNA Analysis Facility
150. Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., and Tan, S. (2005) *Protein Expression and Purification* **41**, 98-105
151. Kimple, M. E., and Sondek, J. (2001) Overview of Affinity Tags for Protein Purification. in *Current Protocols in Protein Science*, John Wiley & Sons, Inc. pp

152. Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., and Schafer, F. (2009) IMMOBILIZED-METAL AFFINITY CHROMATOGRAPHY (IMAC): A REVIEW. in *Guide to Protein Purification, Second Edition* (Burgess, R. R., and Deutscher, M. P. eds.), Elsevier Academic Press Inc, San Diego. pp 439-473
153. Waugh, D. S. (2010) TEV Protease FAQ. in *Technical Resources* (Macomolecular Crystallography Laboratory ed., National Cancer Institute, Frederick, MD
154. (2012) Ni-NTA His-Bind® Resins User Protocol. in *Novagen*®, EMD Milipore Coporation, Darmstadt, Germany
155. (2006) His-Bind® Kit User Protocol. in *Novagen*®, EMD Biosciences Inc. , Darmstadt, Germany
156. (2004) YeastBuster™ Protein Extraction Reagent User Protocol. in *Novagen*®, EMD Biosciences Inc., Darmstadt, Germany
157. Gaberc-Porekar, V., and Menart, V. (2001) *Journal of Biochemical and Biophysical Methods* **49**, 335-360
158. Kaiserman, D., Hitchen, C., Levina, V., Bottomley, S. P., and Bird, P. I. (2011) INTRACELLULAR PRODUCTION OF RECOMBINANT SERPINS IN YEAST. in *Methods in Enzymology, Vol 501: Serpin Structure and Evolution* (Whisstock, J. C. B. P. I. ed.). pp 1-12

Vita

Monique Jenaie LeMieux was born in Portsmouth, Virginia on August 8, 1986. She earned a Bachelor of Science in Biology with a minor in Chemistry from Christopher Newport University, Newport News, Virginia in 2008. Just after graduation, she moved to Knoxville, Tennessee and began her graduate studies in the Genome Science & Technology Program at the University of Tennessee, where she obtained her Master of Science in Life Science.