

**Investigating the Regulation of Indole-3-Acetic
Acid Production By the Plant Associated
Microbe *Pantoea* sp. YR343**

**A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville**

Kasey Noel Estenson

December 2017

Dedication

To my husband, Jason Estenson, and to my children Valen and Soren Estenson.

Acknowledgment

First, I would like to thank my PhD advisors Dr. Jenny Morrell-Falvey and Dr. Valerie Berthelier for giving me the opportunity to carry out my PhD research at UT/ORNL and at UT Medical Center. I would like to express my deep gratitude to Dr. Jenny Morrell-Falvey and Dr. Valerie Berthelier, my research supervisors, for their willingness to give me their time, generous guidance, and valuable/constructive suggestions during the development of my research work. They have provided me with continuous support and resources that have provided me with a solid foundation to be successful upon graduation. It has been a pleasure working with both Jenny and Valerie.

I would like to express my deepest gratitude towards all the group members in the Biological and Nanoscale System Group at Oak Ridge National Laboratory, specifically Dr. Mitchel Doktycz, for allowing me to carry out my research in their laboratories and for their continuous support and advice. I am hugely indebted to Dr. Amber Bible for supplying all the constructs to facilitate my research at ORNL and her kind advice regarding the topic of my research.

Additionally, I would like to thank Dr. Marie Morrison, Dr. Thomas Masi and Mr. James Foster for helping me with my compressive exams. I would like to offer my special thanks to my husband Jason Estenson for his continuous support and love throughout my entire educational career.

Abstract

The auxin indole-3-acetic acid (IAA) plays a central role in plant growth and development and many plant-associated microbes produce IAA. Several IAA biosynthetic pathways have been identified in microbes which use the precursor tryptophan. *Pantoea* sp. YR343, which was isolated from the *Populus deltoides* rhizosphere, is a robust plant root colonizer that produces IAA. Using genomic and metabolomics analyses, we predicted that the indole-3-pyruvate (IPA) pathway is the major pathway in *Pantoea* sp. YR343 for IAA production. To better understand IAA biosynthesis and the effects of IAA exposure on cell physiology, we performed proteomics on *Pantoea* sp. YR343 grown in the presence of tryptophan or IAA. These data indicate that indole-3-pyruvate decarboxylase, a major enzyme in the IPA pathway, is upregulated in the presence of both tryptophan and IAA. Moreover, exposure to IAA induced changes in the proteome, including upregulation of proteins predicted to function in carbohydrate and amino acid transport and EPS biosynthesis. Finally, we constructed a mutant in which the *ipdC* gene was disrupted. The $\Delta ipdC$ mutant showed a significant decrease in IAA production, but was still able to efficiently colonize poplar, suggesting that microbial IAA production is not required for initiation of plant association.

Table of Contents

Chapter 1 Introduction.....	1
Chapter 2 Methods.....	9
Chapter 3 Results.....	13
Chapter 4 Discussion.....	31
Chapter 5 Conclusions and Future Direction.....	34
References.....	36
Appendices.....	46
Appendix A Discovery of a Potent Inhibitor of Z-Alpha1 Antitrypsin Polymerization.....	47
Appendix B Alpha-1 Antitrypsin Mammalian Cell Work	72
Vita.....	85

List of Tables

Table 3.1 Candidate enzymes in IAA biosynthetic pathways, <i>Pantoea</i> sp.YR343 ..	14
Table 3.2 Proteins that are differentially expressed in the presence of tryptophan .	16
Table 3.3 Indole-3-acetic acid statistically significant proteins.....	19
Table A.1 Docking results M α 1AT, Z α 1AT, and M* with S(4nitrobenzyl)	66
Table A.2 Residues interacting with S-(4-nitrobenzyl)-6-thioguanosine.....	67
Table B1. Transfection reagent optimization with various transfection reagents.....	79

List of Figures

Figure 1.1: “Omics Cascade”	2
Figure 1.2: Bacterial IAA influences	5
Figure 1.3: Plant growth promoting characteristics of <i>Pantoea</i> sp. YR343.....	6
Figure 1.4: Tryptophan biosynthetic pathways in <i>Pantoea</i> sp. YR343	8
Figure 3.1: Growth curves of <i>Pantoea</i> sp. YR343 in various conditions.....	17
Figure 3.2: Volcano plots illustrates significantly differentially abundant proteins in <i>Pantoea</i> sp. YR343 in the presence of tryptophan or IAA	23
Figure 3.3 <i>ipdC</i> gene expression in <i>Pantoea</i> sp. YR343.....	26
Figure 3.4 Representative chromatograms of <i>Pantoea</i> sp. YR343, WT and $\Delta ipdC$ using basic extraction to identify tryptamine.....	27
Figure 3.5: Representative metabolomics of <i>Pantoea</i> sp. YR343 and <i>Pantoea</i> sp. YR343 $\Delta ipdC$	28
Figure 3.6: Representative metabolomics of spontaneous degradation of IPA in <i>Pantoea</i> sp. YR343 and <i>Pantoea</i> sp. YR343 $\Delta ipdC$ after 2 days.....	29
Figure 3.7: <i>Populus deltoides</i> root colonization assay by wild type <i>Pantoea</i> sp. YR343 and $\Delta ipdC$	30
Figure A.1: Kinetic diagram of b-PEG-peptide binding to $\alpha 1AT$	54
Figure A.2: Pattern of inhibition resulting from the screening of 80 unknown LOPAC compounds.....	55
Figure A.3: S-(4-nitrobenzyl)-6-thioguanosine inhibits bPEG-peptide binding to Z- $\alpha 1AT$	57
Figure A.4: S-(4-nitrobenzyl)-6-thioguanosine on Z- $\alpha 1AT$ polymerization	58
Figure A.5: The three models of $\alpha 1AT$ protein	59
Figure A.6: The fragments of structures 1QLP (green) and 3T1P (red) used to homology model the M* intermediate state of $\alpha 1AT$	61
Figure A.7: Binding sites for S-(4-nitrobenzyl)-6-thioguanosine	62
Figure A.8: Two crystal structures of $\alpha 1AT$ are superimposed and represented in ribbon diagram	63
Figure B.1: Components of the T-Rex system	74
Figure B.2: pcDNA4/TO expression system.....	75
Figure B.3: Verification of M $\alpha 1AT$ and Z $\alpha 1AT$ intracellular constructs	76

Figure B.4: pcDNA4/TO, M α ₁ AT vector	77
Figure B.5: pcDNA4/TO, Z α ₁ AT vector.....	78
Figure B.6: Transient transfection 48 hours after transfection with lipofectamine 3000 transfection reagent in HepG2 cells	80
Figure B.7: Intracellular construct for both M α ₁ AT and Z α ₁ AT.....	83
Figure B.8: Full-length construct for both M α ₁ AT and Z α ₁ AT	84

Chapter 1

Introduction

In community ecology, it is the assemblage of multiple species that creates the fundamental traits found in complex biosystems. This assemblage, though, is also at the center of the profound consequences these communities have on larger ecological environments [1-3]. In these environments, plants compete for nutrients and space, and this directly affects the microbial and fungal communities which surround the plant's root system. Plants, microbes, and fungi exchange nutrients as a collective in order to maintain a balanced ecological environment that is beneficial for the entire community [4]. Interaction networks within these communities are complex, and include plant-plant, plant-fungal, plant-microbe, fungal-fungal, fungal-microbe, and microbe-microbe interactions [5-10]. How this community is structured and how its members communicate with each other, however, is poorly understood.

Multicellular organisms such as plants and fungi are comprised of complex biological systems, including a variety of microbes, which is referred to as their microbiome. A key area of research is in the understanding of how this microbiome contributes to the phenotypes and responses of the plant to environmental stimuli [11-15]. All microbiome communities, whether they are found on the coasts, in woodlands, or deserts, develop intimate associations with plants coexisting within the same landscape [16-19]. These communities influence each other both directly and indirectly. Furthermore, these communities all form relationships that can be symbiotic, mutual, commensal, or harmful/parasitic [20-22]. Studying the biocomplexity within and between a multiorganismal system allows us to better predict environmental responses to climate challenges over time, benefiting both agriculture and bio-energy production [23, 24].

Biocomplexity is the study of multi-component biological systems and their environmental responses to various chemical and biological signals [25]. As individual organisms, plants, microbes, and fungi each have their own biochemical processes that are influenced by their internal interconnectivity and metabolic processes. These organisms, however, influence each other within their environment by transferring genetic components and biomolecules between one another. Doing so ultimately helps these organisms to construct their collective ecosystem. Although this idea is increasingly recognized, there is still poor understanding of the complexities found between these reactive systems and how they influence each other [2, 26]. Understanding plant microbial communities is extremely difficult task because it involves isolating new organisms and characterizing them quickly. The challenge to understanding these complex systems first happens individually in specific organisms. These attributes are then studied on a community level. To achieve this, plant communities are first isolated from their environment. Then the community is teased apart by isolating the various plant, fungi and bacteria components. Next, the organisms are characterized on an individual level. Finally, the environment is reconstructed and studied at the community level.

Understanding these complex functional processes within a community requires a variety of integrated technologies to reveal an organism's precise internal molecular processes. To understand complex biological systems, it is important to understand what is happening globally within the entire community, as well as on an individual level. The field of omics is utilized to study complex biological systems and to capture all processes within the central dogma. Omics is a field of study in biology used to gather information from DNA, RNA, proteins, and metabolites to better understand the functionality of complex biological systems found under certain conditions [27]. Additionally, omics is also utilized as an analytical tool used to assist in the discovery of unknowns. Omics can be broken down into a variety of fields to capture specific and individual biological processes. This data can then be comparatively analyzed in order to piece relevant biological information back together into a cohesive whole, Figure 1.1.

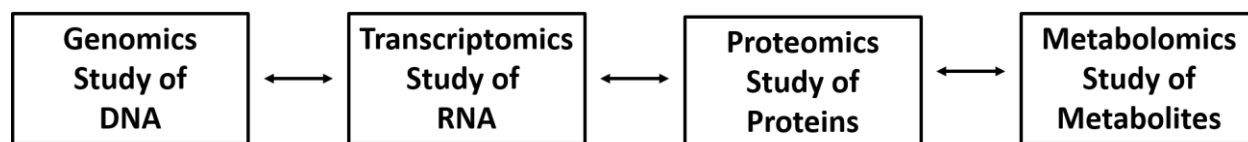


Figure 1.1: “Omics Cascade.” Multiple omics approaches can be used collectively or individually to identify and define a specific phenotype or function of an organism [28].

In a biological system, the comparison of genes via genomics allows for the study of an organism at the genomic level. Genomic resources of these plant-microbe studies are important for researchers. These resources enable researchers to better predict genetic variations, identify gene homologs, predict structural/functional roles, and map genes to specific regions. Based on these genomic predictions, researchers can then go back and perform experiments to validate potential biosynthetic pathways associated with candidate genes [29]. Although genomics is a useful tool used for initial characterization, it is not without its drawbacks. Genomics can only give you predicted functions based on prior genome knowledge. Limitations often include annotation errors, sequence alignment errors, and unknown predictions [30].

Transcriptomics, a field tightly associated with genomics, allows for the investigation of RNA. This field of omics is important because it allows for the detection of specific gene expression networks in complex organisms. Popular approaches to studying RNA transcripts within a biological system include RNA sequencing (RNA-seq), quantitative reverse transcription polymerase chain reaction (qRT-PCR), and reverse transcription polymerase chain reaction (RT-PCR) [31]. The advantages of qRT-PCR vs RT-PCR include: higher resolution, increased signal based on a fluorescent tag, and direct quantification. qRT-PCR and RT-PCR, while effective, are limited to amplifying known genomic regions of the DNA. On the other hand, RNA-seq is also a powerful technology with many advantages such as an easy work flow, detection of unknown transcripts in samples and provides a quick turnaround, it has, as with any technology, drawbacks.

Some of these drawbacks include large data output, low storage capabilities and normalization [31].

Alternatively, Proteomics helps identify proteins. This field can be used to determine protein-protein interactions, examine post-translational modification, identify protein complexes, and identify proteins in, for example, unknown plant or dirt samples [32]. Although proteomics is a powerful approach to identifying functional aspects of organisms, one major disadvantage is that this technique typically only detects 50% of the proteins being expressed at a given time in unicellular organisms and may only detect 10% in higher Eukaryotes, leaving a very narrow window for interpretation [33]. Additionally, proteomic extraction efficiencies may be low and instrument sensitivities may present a problem when detecting reasonable peptides.

Finally, Metabolomics identifies either known or unknown secondary metabolites involved in complex chemical processes. The results of this type of study are often quantified by Gas Chromatography-Mass Spectrometry (GC-MS) [34, 35]. In addition to obtaining functional information, integrative omics can determine information about cellular localization, biological networks, post-translational modifications, and much more [36, 37]. Limitations in metabolomics include the scope of the experiment, measurement accuracy, sample preparation/derivatization, extraction methods, and available metabolite database [38, 39].

Combining these powerful approaches, often referred to as integrative omics, provides detailed information about a biological system in real time. This approach also allows for the collection of functional information about a complex biological organism. That information is then ultimately used to study a community as a cohesive whole. Integrative omics is a powerful approach to studying systems biology. However, it is not without its shortcomings. One of the most apparent drawbacks is that bioinformatic pipelines are limited. Furthermore, data storage is problematic, and only one, single dimensional, analysis of a specific time point is available per experiment [27]. Even with its drawbacks, though, integrative omic approaches provide a powerful and effective means for understanding how communities influence each other and integrate together. It is necessary to use a combination of all of these analyses in order to assert significant biological predictions [40].

The overall scope of this dissertation is to understand what influences the microbiome of *P. deltoides*. *P. deltoides* is a host to an abundant and diverse microbiome, which can influence its growth and productivity. The mechanisms by which plants and microbes influence each other are only starting to be understood. Understanding how beneficial interactions between microbes and plant host are initiated and maintained will inform strategies for improving plant growth and productivity. These efforts will help us determine how these interactions ultimately influence the plant host relationship within *Populus deltoides*, which will help inform strategies to develop more efficient renewable energy resources.

Populus deltoides (poplar) is a model organism used for bioenergy production and agricultural studies. This species was selected for the following reasons: its genome has been sequenced, it's a keystone species, it has broad geographic distribution, fast growth, abundance, and finally, numerous tools have been generated for manipulating its genetic code [41, 42]. Also, *P. deltoides* and its root-associated environment consist of an

integrated infrastructure which contributes to the overall health of the plant [43]. In order to understand how environmental factors influence *P. deltoides*, the rhizosphere and endosphere must be studied. Examination of these microbiomes is important because they are currently thought to contribute to the overall health and viability of *P. deltoides* [5, 44, 45]. The thin layer surrounding the roots of *P. deltoides* is known as the rhizosphere. This area contains rhizobacteria which can influence the plant either indirectly or directly. Some of the bacteria associated with *P. deltoides* are beneficial and can influence the growth and health of the host by acting as bioprotectants, biofertilizers, and/or biostimulants [46, 47].

The beneficial rhizobacteria mentioned previously are known as plant growth promoting rhizobacteria (PGPR) [48-50]. PGPR are bacteria which not only promote plant growth, but also aggressively colonize roots to form symbiotic relationships with its plant host [47, 51]. Root colonization of PGPR is a crucial characteristic of bacteria for the purposes of this study as it allows the bacteria to mediate resource acquisition within the rhizosphere [52, 53]. This mediation allows the bacteria to control various biochemical processes associated within the rhizosphere and helps to maintain a beneficial relationship [53]. Controlled resource acquisition within the rhizosphere directly influences the overall plant viability. Bacteria do this by controlling hormone levels, preventing pathogenicity, and maintaining nitrogen, phosphorus or iron levels as well. Additionally, it helps provide key metabolites, such as indole-3-acetic acid (IAA), which have been shown to improve plant growth and development [54, 55].

The microbiome of *P. deltoides* and *P. trichocarpa* have been the subject of recent studies. In 2009, as a collaborative effort, our research group isolated over 3000 bacterial species from the native *P. deltoides* and *P. trichocarpa* rhizospheres [42]. The specimens were collected across a variety of field sites including the Caney Fork River site and Edgar Evans State park, both located in the state of Tennessee, Yadkin River, North Carolina and sites in Oregon [42, 56]. Following soil extraction, tertiary fine roots were removed. Then, individual bacteria were isolated, sequenced and characterized. One class of the bacteria isolated from the *P. deltoides* rhizosphere is known as *Pantoea*. The *Pantoea* genus is a gram-negative bacteria and consists of approximately 20 different species [57]. *Pantoea* may aid in bioremediation and can act as a biocontrol for plants [58-60]. These factors qualify this species of isolates as a unique niche for experimental development. Various *Pantoea* strains have been isolated from sources including environmental water, soil, insects, animals and humans. Each *Pantoea* strain has the potential to be either pathogenic or non-pathogenic, depending on its host environment [61, 62].

Interestingly, many species within the genus *Pantoea* have been found associated with plants [57]. However, some of these species, such as *P. stewartii* and *P. rodasii*, are known pathogens of corn and eucalyptus respectively [63, 64]. Still, other strains have been shown to have beneficial effects on a plant host. For example, *P. agglomerans* is thought to promote plant growth by enhancing root growth, which can increase water and mineral uptake [65]. Indeed, this promotion of growth is thought to be the result of IAA production. IAA produced by bacteria can influence plant hosts in diverse ways. Such influences include induction of tissue differentiation, cell division, elongation, lateral-root formation, and cambial growth, Figure 1.2 [66, 67]. Genomic analyses of *P. agglomerans* also indicated the presence of a gene encoding indole-3-pyruvate decarboxylase (*Ipdc*).

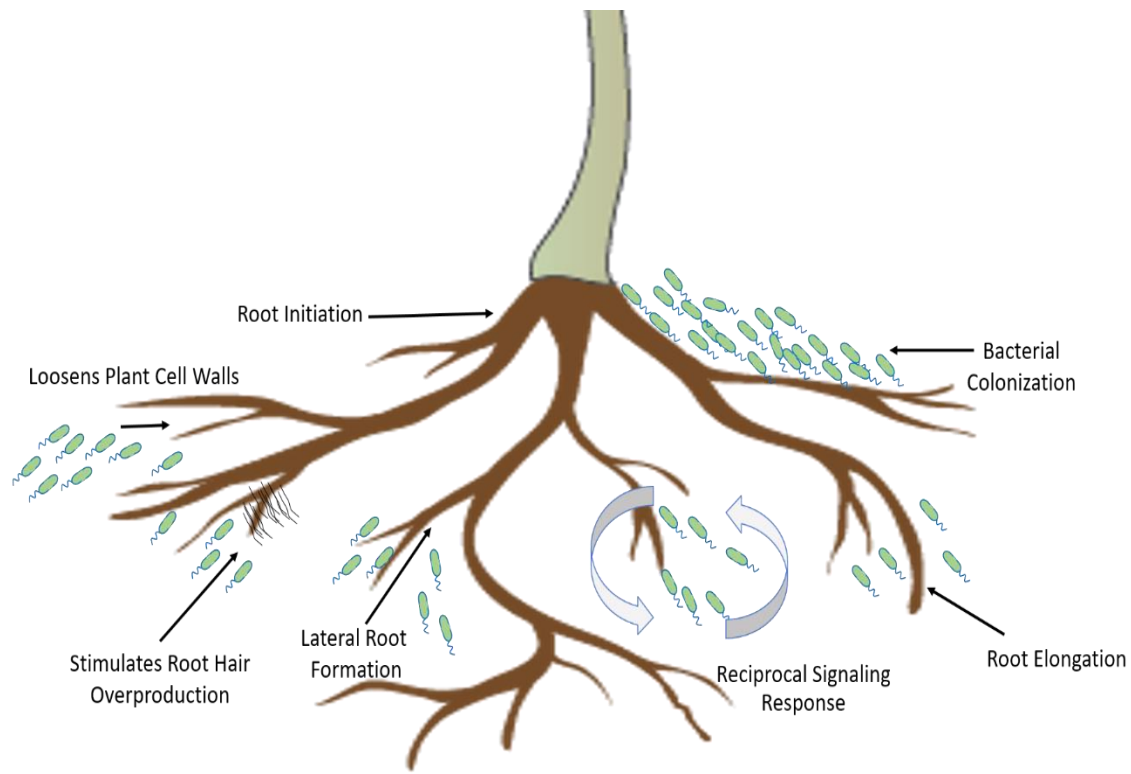


Figure 1.2: Bacteria that produce IAA influence its plant host in a variety of ways including: root initiation, it helps loosen plant cell walls to release exudates, stimulates root hair overproduction, increases lateral root formation, bacterial IAA can act as a signaling molecule to both the plant and bacteria, it aids in root elongation, and bacterial colonization.

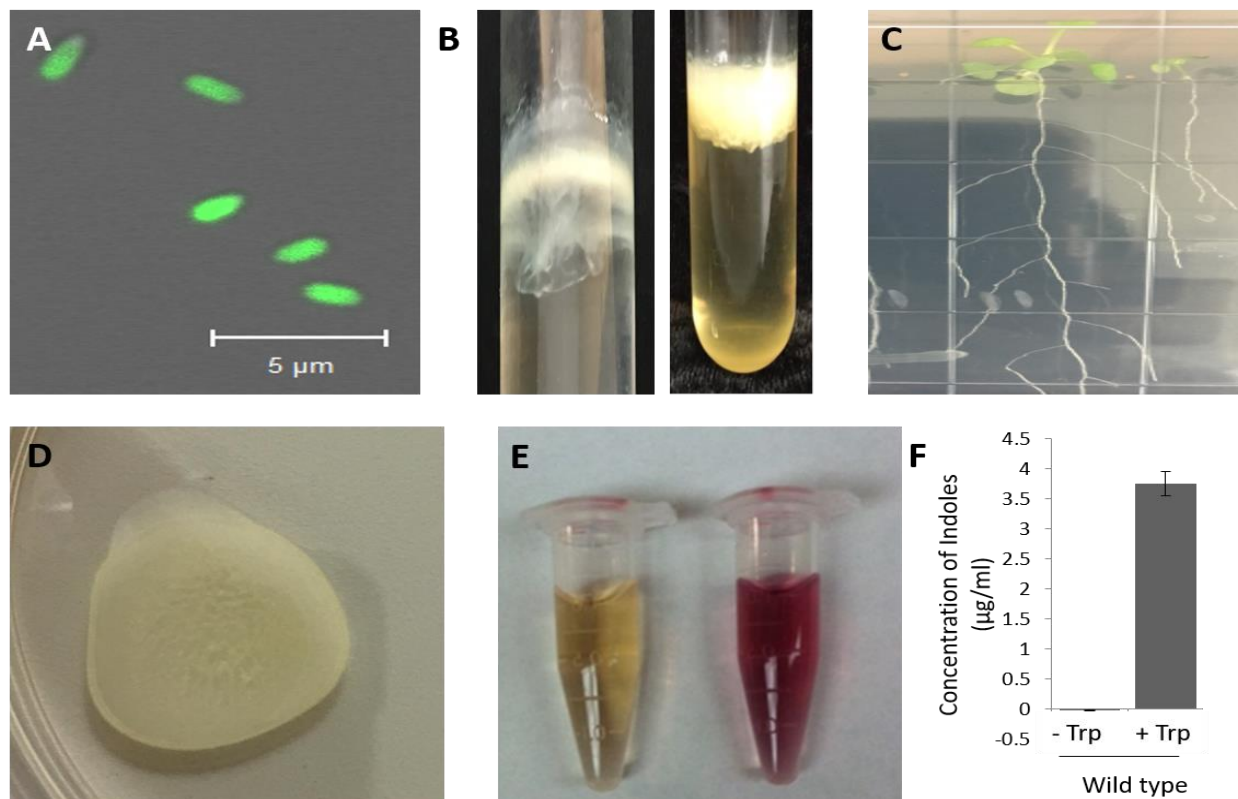


Figure 1.3: Plant growth promoting characteristics of *Pantoea* sp. YR343. A) *Pantoea* sp. YR343 is genetically tractable, confocal image of *Pantoea* sp. YR343 expressing GFP. B) *Pantoea* sp. YR343 forms biofilms in both minimal and rich media. C) *Pantoea* sp. YR343 colonizing the roots of *Arabidopsis thaliana* and initiating lateral root formation. D) *Pantoea* sp. YR343 has swimming and swarming motility. E) Salkowski test for indole production in *Pantoea* sp. YR343. F) *Pantoea* sp. YR343 produces about 4µg/mL of indoles.

This is one of the main genes associated with IAA biosynthesis through the indole-pyruvate (IPA) pathway [65]. Equally important, a metabolite analysis of *P. dispersa* confirmed that it also produces IAA through the IPA pathway [68].

Pantoea sp. YR343, one of our isolates from *P. deltoides*, is a gram-negative γ -proteobacterium [43]. According to our initial analysis, *Pantoea* sp. YR343 is capable of solubilizing phosphate, forming biofilms, and colonizing roots. Furthermore, it is genetically tractable, can initiate lateral root formation in the presence of a plant host, and secretes the phytohormone IAA. As such, *Pantoea* sp. YR343 exhibits key characteristics of a plant growth promoting bacteria, Figure 1.3 [69]. *Pantoea* sp. YR343 can also colonize other plants, including *Triticum aestivum* and *Arabidopsis thaliana*. In fact, the colonization of these plants by *Pantoea* sp. YR343 resulted in increased lateral root production [70]. Similar to *P. agglomerans* and *P. dispersa*, genomic analyses indicated the presence of an *ipdC* homolog (PMI39_00059), suggesting the possibility that *Pantoea* sp. YR343 may synthesize IAA through the IPA pathway.

Multiple IAA biosynthetic pathways have been described in microbes, most of which require the precursor tryptophan [71-73]. The associated tryptophan-dependent pathways are as follows: indole-3-acetonitrile (IAN), indole-3-acetamide (IAM), tryptophan side-chain oxidase (TSO), indole-3-pyruvate (IPA), and finally, the tryptamine pathway, Figure 1.4 [71, 74]. IAA production has also been studied in many different microbes. These microbes include *Azospirillum brasilense*, *Enterobacter cloacae* UW5, *Pantoea dispersa*, *Pantoea agglomerans*, and *Pseudomonas putida* [65, 67, 68, 75, 76]. Evidence of this process at work can be found in the *A. brasilense* microbe, a plant growth promoting bacteria that synthesizes IAA primarily through the indole-3-pyruvate pathway [77]. In *A. brasilense*, a deletion of the *ipdC* gene which encodes indole-3-pyruvate decarboxylase resulted in a drastic reduction in IAA production in the mutant. In fact, IAA production in the mutant was 90 percent below wild type levels [78]. The *ipdC* mutant strain was still capable of IAA production, however. This suggested the presence of a second biosynthetic pathway [79]. Moreover, the presence of IAA itself induced *ipdC* gene expression in *A. brasilense*, further suggesting that IAA biosynthesis is a highly regulated process [76, 80, 81].

To test the prediction that *Pantoea* sp. YR343 produces IAA through the IPA pathway, and to determine how *Pantoea* sp. YR343 responds to the presence of tryptophan and IAA, we performed proteomic analyses of *Pantoea* sp. YR343 growing in minimal media, supplemented with tryptophan or IAA. In the presence of exogenous tryptophan or IAA, we found that the *ipdC* gene is upregulated in *Pantoea* sp. YR343. This finding suggests that the IPA pathway is active under these conditions. Exposure to IAA also induced other changes in the proteome. These changes include the upregulation of proteins which were predicted to function in nutrient and mineral uptake, as well as in EPS biosynthesis. Finally, we constructed a mutant strain in which the *ipdC* gene was disrupted. This mutant strain showed a significant decrease in IAA production, but did not show defects in its ability to colonize poplar. This suggests that microbial IAA production is not required by *Pantoea* sp. YR343 before initiating plant-microbe interactions.

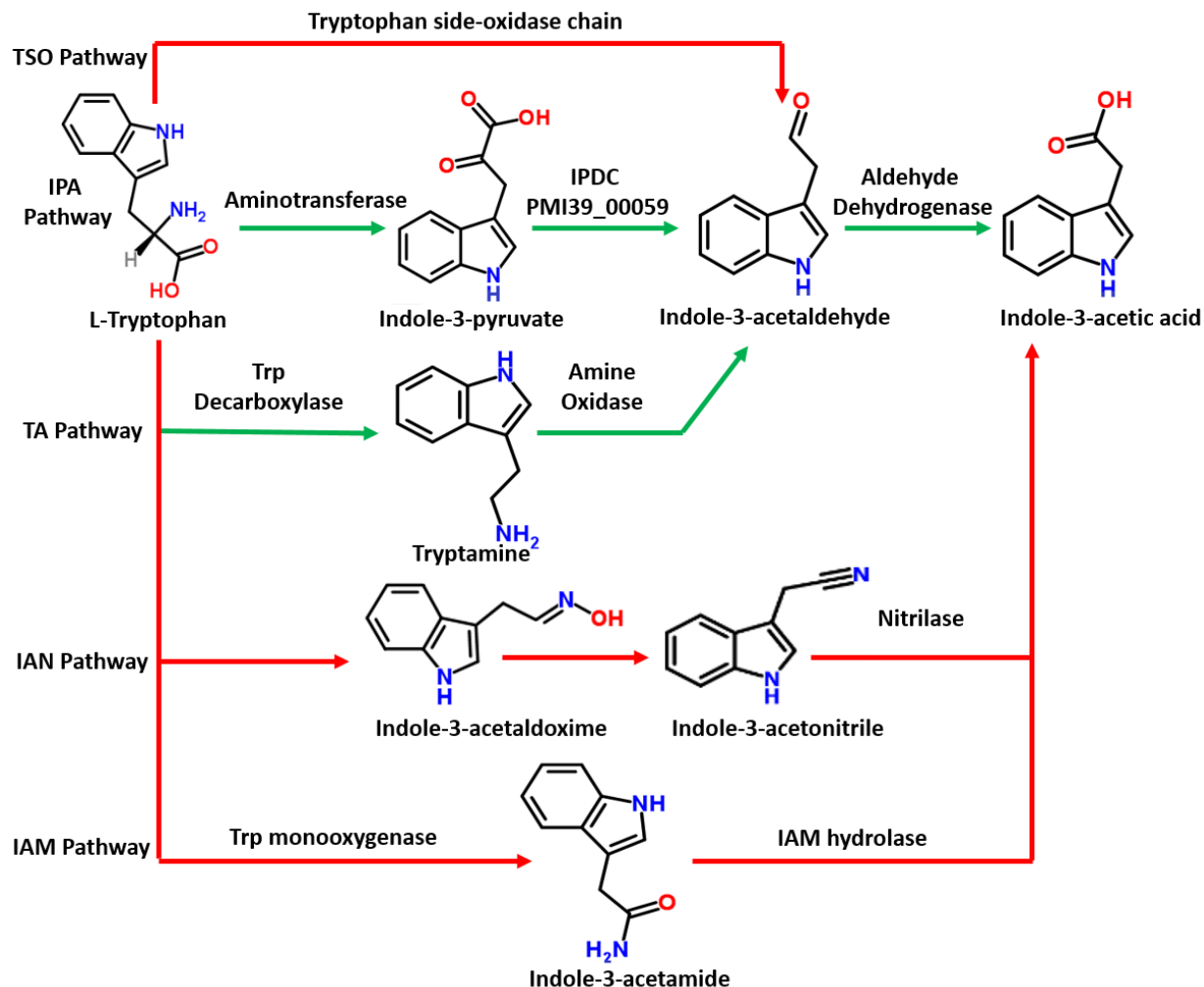


Figure 1.4: Tryptophan biosynthetic pathways in *Pantoea* sp. YR343, red indicates no enzymes found and green indicates enzymes found. Tryptophan side-chain oxidase (TSO) pathway, indole-3-pyruvate (IPA) pathway, tryptamine (TA) pathway, Indole-3-acetonitrile (IAN) pathway, and the indole-3-acetamide (IAM) pathway.

Chapter 2

Methods

2.1 Growth of *Pantoea* sp. YR343

Pantoea sp. YR343 was grown in either R2A (R2A Broth Premix, TEKnova, Inc.) or M9 minimal media (per 1 L, 6 g. NaCl, 1 g NH₄Cl, 10 mL filter sterilized 100 mM MgSO₄, 20% glucose, and 10 mM CaCl₂) with shaking at 28°C. For proteomic analyses, *Pantoea* sp. YR343 was grown in M9 medium supplemented with 1mM L-tryptophan or with 5μM, 50μM or 500μM IAA (Sigma Aldrich).

2.2 Metabolic analyses

Pantoea sp. YR343 wild type and *ΔipdC* cultures were grown in M9 medium with or without L-tryptophan (1 mM) to an OD₆₀₀ of 1. After centrifugation, the supernatant was collected (800 μL) and prepared for metabolite analysis. For acidic and neutral metabolites, the supernatants were acidified to pH 2 with hydrochloric acid and then extracted with ethyl acetate (800 μL). For basic metabolites (specifically, tryptamine), samples were basified to pH 11 with sodium hydroxide and extracted with toluene (800 μL). Samples were vortexed vigorously and briefly centrifuged, after which a 600-μL portion of the organic (top) layer was collected and transferred to a 2-mL glass vial. Next, the solvent was evaporated under a stream of argon gas, and the residue was resuspended in 100 μL of derivatization agent (BSTFA + TMSCI 99:1). These samples were then incubated at 80 °C for 4 h. After derivatization, the sample was diluted with 900 μL of hexane in the same vial.

GC/MS analysis was performed with an Agilent 7890A gas chromatograph equipped with a 7693A automatic liquid sampler, an HP-5ms capillary column (30 m long × 0.25 mm inside diameter with a 0.25-μm capillary film of 5% phenyl methylsilicone) and a 5975C mass-sensitive detector. Splitless injections of 1 μL were made at an inlet temperature of 270 °C with a 15-s dwell time (needle left in the inlet after injection), an initial column temperature of 60 °C and helium carrier gas at a constant flow of 1 mL/min. After 2 min at 60 °C, the temperature was ramped at 20 °C/min to 200 °C, then at 10 °C/min to 270 °C and finally 30 °C/min to 300 °C, with a 2-min hold at 300 °C. The detector was operated with a transfer line temperature of 300 °C, source temperature of 230 °C and quadrupole temperature of 150 °C. After a 5-min solvent delay, electron-impact mass spectra from 50–500 atomic mass units were collected continuously (~3/s) for the duration of the run.

2.3 Promoter-Reporter (GFP) Construct

The *ipdC* (PMI39_00059) promoter region was defined as 350 base pairs upstream from the transcription start site and amplified from genomic DNA using the primers *ipdC*prom_(HindIII)For: 5'-CCCAAGCTTGGCTGTTATCGACGCGCG-3' and *ipdC*prom_(EcoRI) Rev: 5'-CCGGAATTCGCCAACGTTGGGGGTTTT-3'. The fragment was subcloned into pPROBE-NT' [82] and the resulting plasmid was transformed into *Pantoea* sp. YR343 via electroporation and selected on R2A plates containing 50μg

kanamycin mL⁻¹ [70]. GFP expression from cells harboring pPROBE-ipdC was monitored using a Zeiss LSM710 confocal laser scanning microscope. ImageJ was used for image processing [83].

2.4 Sample preparation for proteomics

All chemicals were obtained from Sigma Chemical Co. (St. Louis, MO), unless specified otherwise. High performance liquid chromatography- (HPLC-) grade water and other solvents were obtained from Burdick & Jackson (Muskegon, MI), 99% formic acid was purchased from EM Science (Darmstadt, Germany) and sequencing-grade trypsin was acquired from Promega (Madison, WI).

Frozen cell pellets were ground to a powder under liquid nitrogen and suspended in a detergent-based cell lysis buffer (5% SDS, 50 mM Tris-HCl, 0.15M NaCl, 0.1 mM EDTA, 1 mM MgCl₂, pH 8.5).[84] Cells were lysed via 20 min of heat lysis[84] and transferred to fresh eppendorf tubes. Trichloroacetic acid was added to samples at a final concentration of 25%, and proteins were precipitated by storage at -20 °C overnight. The mixtures were thawed briefly and centrifuged at 21,000 g x 20 min. The resulting cell pellets were washed with chilled acetone and centrifuged at 21,000 g x 10 min. The acetone wash was repeated thrice, discarding supernatants. The cell pellets were air dried, dissolved in guanidine buffer [6M guanidine HCl in Tris-CaCl₂ buffer, pH 8.5 (50mM Tris, 10mM CaCl₂)] and incubated at 60°C for four hours. The total amount of extracted protein was measured using the RC/DC protein estimation kit (Bio-Rad Laboratories, Hercules, CA, USA) per the manufacturer's instructions. Bovine serum albumin supplied with the kit was used as standard for the assay. The samples were diluted six-fold using Tris-CaCl₂ buffer. Proteolysis was carried out by adding modified sequencing grade trypsin (Promega, Madison, WI) at 40 µg/mg protein with overnight incubation at 37 °C and gentle mixing.[85] Digested peptides were stored at -80°C until MS analysis. Approximately 75 µg peptides per sample was loaded on an in-house packed biphasic column [strong cation exchange, SCX (Luna, Phenomenex, Torrance, CA) and reverse-phase C18 (Aqua, Phenomenex, Torrance, CA)] as described earlier [85, 86] and subjected to offline desalting.[87]

2.5 Mass spectrometry

Peptides were analyzed using a two-dimensional liquid chromatography-tandem mass spectrometry approach (2D-LC-MS-MS)[88], implemented as described previously[89]. Peptides were eluted from the SCX trapping column by eleven successive step gradients of increasing ammonium acetate concentration, from 50 mM to 500 mM. Each SCX step gradient eluted a set of peptides onto a 15 cm long reverse-phase (C18) column, where they were separated by a two-hour gradient from 100% solvent A (5% CH₃CN, 0.1% formic acid in water) to 50:50 solvent A:solvent B (70% CH₃CN, 0.1% formic acid in water). Peptides eluting from the RP column were introduced via a nanoelectrospray source (Proxeon, Odense, Denmark) into a linear ion trap mass spectrometer (LTQ-XL, Thermo Scientific, San Jose CA), where data were acquired in data-dependent mode. Following each full-scan mass spectrum, up to 5 tandem mass spectra (MS-MS) were acquired from the most intense ions in the full-scan spectrum. For collision-induced dissociation, precursor isolation width was 3.0 m/z units, and normalized collision energy was 35%. Dynamic exclusion was employed with a repeat count of 1, exclusion list size of 300, repeat duration of 60 sec, and exclusion duration of 180 sec.

2.6 Proteomics data analysis

Peptide identifications were obtained from MS-MS spectra using Myrimatch (version 2.1.138).[90] Myrimatch settings included precursor m/z tolerance of 1.5, fragment m/z tolerance of 0.5, charge states up to +4, TIC cutoff of 98%, cleavage rule “Trypsin/P”, fully tryptic peptides only, with a maximum of two missed cleavage sites per peptide. The protein database for the Myrimatch searches was based on the predicted proteome for *Pantoea* sp. YR343[91], downloaded as 27136.faa from the Department of Energy Joint Genome Institute Integrated Microbial Genomes web site [92] on Aug 25, 2014. Sequences for common contaminant proteins were appended to the *Pantoea* sp. YR343 protein fasta file. The assembled protein fasta file contained sequences for a total of 4945 proteins. Myrimatch employed a decoy database containing sequence-reversed version of these proteins to provide an estimate of peptide false discovery rate.

Protein identifications were assembled from peptide identifications using IDPicker (version 3.1.599). [93] A protein identification required identification of at least two distinct peptides. The maximum false discovery rate for peptide-spectrum matches was set to 2%. Observed peptide and protein false discovery rates were xx and yy, respectively.

Analysis of identified proteins for shared peptides was performed using Microsoft Access and custom scripts in R. Proteins sharing all identified peptides were combined into protein groups. Protein abundance was estimated using spectrum count (number of tandem mass spectra assigned to that protein [94]), adjusted for shared peptides.[95] Normalized spectral abundance factor (NSAF) values [96] were calculated for each protein from adjusted spectrum count, with adjusted spectrum count values of 0 replaced by 0.3. Abundance ratios for proteins for which all adjusted spectrum count values were zero across the 3 biological replicates of one treatment are therefore inaccurate, and were tracked through further analysis steps as “present/absent” proteins (i.e., detected in one treatment and not in the other). Proteins for which an average of < 2.5 tandem mass spectra were identified over the 3 biological replicates of a given treatment were identified as “low abundance” for that treatment;[97] the fraction of NSAF represented by low abundance proteins ranged from 3-7%. Significant changes in protein abundance for a given treatment-control pair were identified using t tests (2-tailed, unpaired, equal variances) on log₁₀ transformed NSAF values, with Benjamini-Hochberg correction for multiple hypothesis testing. [98]

2.7 Reverse transcription polymerase chain reaction (RT-PCR)

Total RNA was extracted from *Pantoea* sp. YR343 cultures grown in M9 minimal media with or without tryptophan using the RNeasy Mini kit (Qiagen) as per manufacturer’s instructions. Following RNA elution, 2.5 units RNase-free DNaseI (1 U/μl) (NEB) was added to the RNA and incubated at 37°C for 30 minutes followed by heat inactivation at 75°C for 5 minutes. cDNA was transcribed from 500 ng RNA using ThermoScript RT-PCR system (Invitrogen) as per manufacturer’s instructions. 5 units of RNaseH was added to each sample and incubated for 20 minutes at 37°C to eliminate residual RNA in the cDNA sample. After RNaseH treatment, RT-PCR was performed using FailSafe PCR PreMix Selection Kit (Epicentre) and specific primers for ipdC (PMI39_00059) (RT_ipdC For: 5’-ATCCCGAAATTGCCTGGGTTG-3’ and RT_ipdCRev: 5’-GCGGTAAAGGTGTCGGTAAAC-3’) and 16S (RT_16s For: 5’-

ACGATCCCTAGCTGGTCT-3' and RT_16s Rev: 5'CTAATCCTGTTTGCTCCC-3') as a control. Gel images were analyzed with ImageJ.

2.8 Construction of ipdC Mutant

To generate a disruption mutant for *ipdC* (PMI39_00059), we first constructed pKnock-ipdC by ligating the pKnock-Tc vector with a 600 base pair ipdC fragment amplified from genomic DNA using ipdC_pknock for (XbaI): 5'-GCTCTAGAACTCCATCAGCAGGTTGCCGCA-3' and ipdC-pknock rev (KpnI): 5'- CGG GGT ACCCCAAAGGCCGCAGTGCCTTGA-3' [99]. The resulting plasmid was verified by restriction digests and transformed into *Pantoea* sp. YR343 by electroporation and selected on R2A plates containing 5 µg tetracycline ml⁻¹. Disruption of the ipdC gene was confirmed by PCR and sequencing.

2.9 Poplar colonization assays

Colonization of *Populus trichocarpa* BESC819 was performed as described previously [70]. Briefly, *Pantoea* YR343-GFP or the $\Delta ipdC$ mutant were grown overnight in R2A media and then the OD₆₀₀ of each culture was adjusted to 0.01 using fresh R2A. Five plants were inoculated for each treatment. For each plant, sterile clay soil with 1X Hoagland's media was inoculated with 10 ml of either sterile R2A media (control), wild type *Pantoea* YR343-GFP, or the $\Delta ipdC$ mutant by first mixing the bacterial culture with the clay soil, then planting a rooted *P. trichocarpa* shoot tip. Plants were incubated in the growth chamber for three weeks. Plant roots were harvested and weighed, then washed with 3 ml PBS and a small amount of glass beads to disrupt bacterial attachment. Colony-forming units (CFU) were counted for wild type *Pantoea* YR343-GFP on R2A agar plates containing 10 µg ml⁻¹ gentamycin. CFUs for the $\Delta ipdC$ mutant were counted on R2A agar plates containing 5 µg ml⁻¹ tetracycline. Control plants showed no background contamination when plated on R2A plates with either gentamycin or tetracycline.

2.10 Biofilm formation assay

Biofilm formation was quantified using a modified protocol from O'Toole and Kolter, 1998 [100]. An overnight culture of *Pantoea* sp. YR343 (WT and $\Delta ipdC$) was diluted 1:100 into M9 minimal media supplemented with 0.4% w/v glucose, supplemented with various treatments (1mM tryptophan, 5µM IAA, 50µM IAA or 500µM IAA), and grown statically in a 96-well plate covered with breatheable tape in place (Brethe-EASIER, Diversified Biotech), at 28°C for 72 hours [70]. After 72 hours, adherent cells were rinsed with water, patted dry and stained with 0.1% w/v crystal violet stain for 10 minutes at room temperature. The stain was then poured off and plates were washed 2 times with water, patted dry and air dried for 30 minutes. After drying, the crystal violet associated with biofilms was dissolved using, 200µl per well, a modified biofilm dissolving solution which contained 10% w/v SDS dissolved in 80% v/v ethanol and incubated at room temperature for 10 minutes [101]. Absorbance was measured at 550 nm using a BioTekSynergy 2 microplate reader.

Chapter 3

Results

3.1 Identification of putative IAA biosynthetic enzymes by genomic analyses

Our previous work showed that *Pantoea* sp. YR343 produces IAA following addition of tryptophan to the growth medium [70]. A number of tryptophan-dependent biosynthetic pathways for IAA have been described in microbes, Figure 1.4; [74]. To determine which pathway(s) may be present in *Pantoea* sp. YR343, we performed genomic analyses to identify putative homologs of known proteins in these pathways. From these analyses, we were able to clearly identify an ipdC homolog (PMI39_00059), which encodes indole-3-pyruvate decarboxylase, an enzyme in the IPA pathway. The first step in the IPA pathway (Trp→indole-3-pyruvate) requires aromatic aminotransferase function and we identified seven genes in *Pantoea* sp. YR343 that are annotated to encode gene products with this function, Table 3.1. Likewise, the final step in this pathway (indole-3-acetaldehyde (IAAld) →IAA) requires an aldehyde dehydrogenase and there are at least 17 gene products annotated with this function, Table 3.1. These results suggest that *Pantoea* sp. YR343 may have a functional IPA pathway.

Similarly, we analyzed the genome for the presence of gene products found in the other IAA biosynthetic pathways. Because we were unable to find a homolog of tryptophan monooxygenase, we ruled out the presence of the IAM pathway in *Pantoea* sp. YR343, Figure 1.3. Likewise, *Pantoea* sp. YR343 lacked enzymes in the IAN pathway, Figure 1.3. Because the TSO pathway has only been described in *Pseudomonas fluorescens* CHA0 [102], we concluded that its presence in *Pantoea* sp. YR343 seemed unlikely. Although the sequence homologies were low, we did find genes encoding candidate enzymes in the tryptamine pathway, suggesting that this pathway may also be present in *Pantoea* sp. YR343.

3.2 Proteomics of *Pantoea* sp. YR343 grown in the presence of tryptophan

To better understand how *Pantoea* sp. YR343 synthesizes IAA, we generated proteomic profiles of cells growing in the presence or absence of tryptophan. We detected 2046 proteins from cultures grown in the presence of tryptophan, with 20 proteins that were significantly differentially abundant under these growth conditions, Table 3.2, Figure 3.1C. Not surprisingly, we found that several enzymes in the tryptophan biosynthesis pathway (encoded by PMI39_02719-PMI39_02724) were downregulated in the presence of excess tryptophan. Conversely, none of the enzymes predicted to be involved in IAA biosynthesis showed statistically significant differences in abundance in the presence of tryptophan, Table 3.1. Despite the results not being statistically significant ($p=0.06$), IpdC did trend toward being more abundant in the presence of excess tryptophan. These results also showed that only four of the seven predicted aromatic aminotransferases in the IPA pathway were detected under these growth conditions, Table 3.1. Likewise, only 9 out of 17 predicted aldehyde dehydrogenases were detected, Table 3.1. In the tryptamine pathway, only one out of the two enzymes was detected under these growth

Table 3.1. Candidate enzymes in the IAA biosynthetic pathways of *Pantoea* sp. YR343. Ratios are based on normalized spectral abundance factors (NSAF) and are calculated as the log2 ratio of experimental condition to control (minimal media). Ratios that are highlighted are statistically significant. ND indicates that the protein was not detected under these conditions.

Candidate Tryptophan Dependent Enzymes				
IPA Pathway				
Locus Tag	Trp	5μM IAA	50μM IAA	500μM IAA
<i>Aromatic Aminotransferases</i>				
PMI39_01811	0.8487	0.4171	0.7565	1.0798
PMI39_01993	0.2275	0.0713	0.3541	-0.2875
PMI39_02094	-1.3658	-2.0033	-0.9197	-3.2724
PMI39_02628	ND	2.5612	1.5838	1.6159
PMI39_02920	ND	ND	ND	ND
PMI39_04274	ND	-0.0348	-0.0686	1.6159
PMI39_04560	0.6005	0.7075	0.1988	1.1081
<i>ipdC</i>				
PMI39_00059	5.1663	-0.0348	1.8268	3.9062
<i>Aldehyde Dehydrogenases</i>				
PMI39_00313	ND	ND	ND	ND
PMI39_00317	ND	ND	ND	ND
PMI39_00354	-0.4284	0.31	0.3592	0.1593
PMI39_00431	ND	ND	ND	ND
PMI39_00617	-0.706	-1.9921	0.3006	1.0616
PMI39_00725	ND	1.6233	2.567	2.0691
PMI39_00794	ND	1.4607	-0.0686	0.1916
PMI39_00977	ND	ND	ND	ND
PMI39_01356	ND	2.2756	-0.0686	0.1916
PMI39_02144	-0.0257	1.1257	1.6944	0.9253
PMI39_02889	0.6333	0.2894	0.3945	1.3988
PMI39_03367	-0.6096	-0.4069	-0.2181	0.5863
PMI39_03939	ND	ND	ND	ND

Table 3.1 Continued

Locus Tag	Trp	5μM IAA	50μM IAA	500μM IAA
PMI39_04111	0.3855	0.0882	0.5504	0.4438
PMI39_04199	-1.2084	-1.4993	1.0188	-1.2728
PMI39_04201	-1.2818	-1.5726	-1.6064	-1.3461
PMI39_04236	-0.0178	0.0281	-0.1852	0.2146
Tryptamine Pathway				
PMI39_03119	-0.2738	1.854	1.854	2.4726
PMI39_03409	ND	-0.0348	-0.0348	3.3216

Note: ND = Not Detected, highlighted grey = Statistically Significant

Table 3.2. Proteins that are differentially expressed in the presence of tryptophan. The log2ratio is based on the normalized spectral abundance factor (NSAF) and is shows as the log2 ratio of trp to control.

Proteins that are Statistically Differentially Abundant in the Presence Of Tryptophan		
Locus Tag	Predicted Function	Trp
PMI39_04926	ATPase-P-type (transporting), HAD superfamily, subfamily IC	3.9928
PMI39_00980	Octopine/nopaline transport system substrate-binding protein	3.9589
PMI39_04398	Protein of unknown function	3.9413
PMI39_01931	Hypothetical protein	3.8306
PMI39_01840	Glycosyltransferase involved in cell wall biosynthesis	3.1865
PMI39_00879	UFP0755 protein	2.9928
PMI39_03779	2-isopropylmalate synthase	-0.7728
PMI39_01718	4-aminobutyrate aminotransferase	-1.0498
PMI39_00111	ABC-type uncharacterized transport system, periplasmic component	-2.481
PMI39_03872	B3/B4 domain-containing protein	-2.481
PMI39_03883	Transcriptional regulator, TetR family	-2.481
PMI39_01285	Hypothetical protein	-2.705
PMI39_00349	Transcriptional regulator, ArgR family	-2.885
PMI39_04501	Glucose-1-phosphate adenylytransferase	-3.2081
PMI39_04638	Membrane dipeptidase	-3.2081
PMI39_02719	Anthranilate synthase, component I	-3.3475
PMI39_00256	DNA repair protein RadA/Sms	-3.4718
PMI39_02040	Polar amino acid tranprot system substrate-binding protein	-3.8029
PMI39_02676	Hypothetical protein	-3.9673
PMI39_02215	Branched-chain amino acid transport system substrate-binding protein	-4.2136

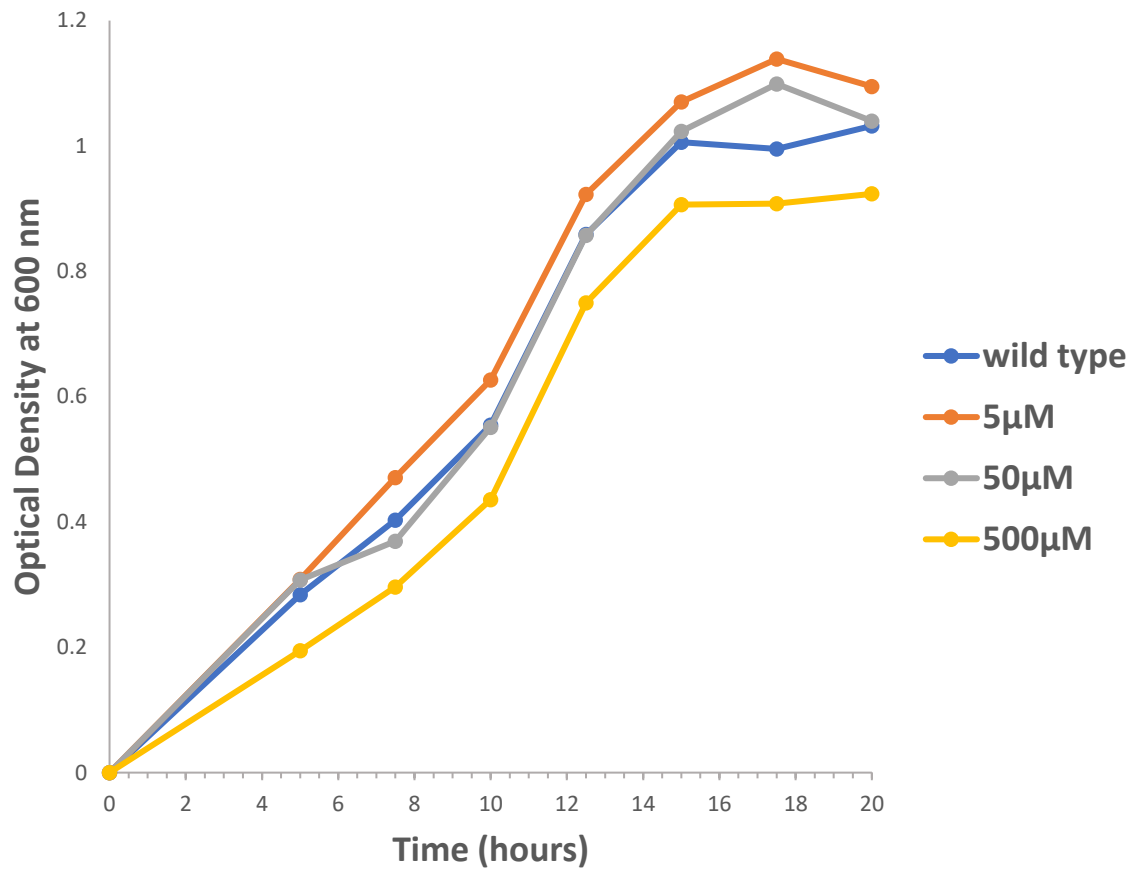


Figure 3.1: Growth curves of *Pantoea* sp. In various conditions. A) Various concentrations of exogenous tryptophan. B) Various concentrations of exogenous IAA.

conditions, Table 3.1

3.3 Proteomics of *Pantoea* sp. YR343 grown in the presence of IAA

Because IAA itself can cause changes in growth of *P. dispersa* [68] and act as a positive regulator of *ipdC* gene expression in *A. brasilense* [76, 81], we wanted to determine the effects of IAA exposure on *Pantoea* sp. YR343. For this experiment, we grew *Pantoea* sp. YR343 in minimal medium supplemented with 5 μ M, 50 μ M, or 500 μ M IAA. Unlike the results with *P. dispersa* [68], we did not observe any differences in the growth of *Pantoea* sp. YR343 cultures in different concentrations of IAA, Figure 3.1. In addition, *Pantoea* sp. YR343 was not able to grow using IAA as the sole carbon source (data not shown). Despite no observable differences in growth, proteomics analyses revealed many differentially abundant proteins under these growth conditions, Table 3.3, and Figure 3.2 A,B,C,D). In these experiments, we detected 2437 proteins across all runs (5 μ M, 50 μ M, and 500 μ M IAA). In the cultures grown with 5 μ M IAA, we detected 34 statistically differentially expressed proteins compared to cells grown in minimal medium, Table 3.3. In cultures grown with 50 μ M and 500 μ M IAA, we detected 26 and 55 statistically differentially expressed proteins, respectively, Table 3.3.

Among the proteins that are differentially regulated, we found several proteins involved in phosphate regulation, including PhoU (PMI39_02797) and PhoH (PMI39_02138) which are both upregulated, and PstB (PMI39_03605) which is down-regulated. In many bacterial systems, PhoU acts as a negative regulator of PhoR, which, together with PhoB, forms a two-component regulatory system that modulates the cellular response to environmental phosphate levels [103]. PhoU also plays a role in regulating the phosphate-specific transporter Pst (PstSCAB), of which PstB is a component [104, 105]. Several proteins involved in inorganic ion transport and metabolism were also upregulated, including two involved in iron transport, (PMI39_01007, PMI39_04647), an ABC-type molybdenum transporter (PMI39_04031), and a copper binding protein (PMI39_04844).

A number of proteins involved in carbohydrate (PMI39_00875, PMI39_02090; PMI39_03310) and amino acid (PMI39_00875, PMI39_02090; PMI39_03310) transport and metabolism were also more abundant in the presence of IAA. In addition, we also observed upregulation of PMI39_00356 which has homology to *E. coli* AaeA, a protein involved in aromatic carboxylic acid efflux [106]. It has been noted that AaeA and other efflux pump proteins are highly active in bacterial biofilms [107]. Growth in the presence of IAA also resulted in upregulation of the gene product encoded by PMI39_01840, which is part of a conserved operon involved in EPS biosynthesis (PMI39_1848-PMI39_01821). Other gene products in that operon were also upregulated, although their abundance was not statistically significant. In *Erwinia amylovora* and *P. stewartii*, the gene products encoded by this operon are required for the synthesis of an acidic extracellular polysaccharide, named *amylovoran* or *stewartin*, respectively [108, 109].

Enzymes predicted to play a role in IAA biosynthesis were differentially abundant in the presence of IAA, including *Ipdc* (PMI39_00059) which was significantly more abundant when cells were grown in 500 μ M IAA, Table 3.1 and 3.3. Of the seven predicted aromatic aminotransferases in the IPA pathway, we detected six when cells were grown in the presence of IAA. Likewise, we detected 12 of the 17 predicted aldehyde dehydrogenases

Table 3.3. Proteins that are differentially expressed in *Pantoea* sp. YR343 with exogenous IAA (5μM, 50μM, and 500μM). The log2ratio is based on the normalized spectral abundance factor (NSAF) and is shows as the log2 ratio of trp to control.

Indole-3-Acetic Acid Statistically Significant Proteins				
LocusTag	Predicted Function	5μM	50μM	500μM
PMI39_01856	Porin, OprB family	1.4607	-0.0686	5.7262
PMI39_01007	Ferritin	3.9464	4.0749	4.914
PMI39_04656	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	-0.0348	-0.0686	4.8976
PMI39_03310	Simple sugar transport system substrate-binding	-0.0348	2.0704	4.6733
PMI39_02999	3-oxoacyl-(acyl-carrier-protein) synthase	3.9491	3.2246	4.5582
PMI39_01826	Hypothetical Protein	3.6444	3.5317	4.4485
PMI39_04053	2-keto-4-pentenoate hydratase	-0.0348	-0.0686	4.2681
PMI39_01840	Glycosyltransferase involved in cell wall biosynthesis	3.7732	2.9993	4.118
PMI39_04447	Hypothetical Protein	1.9249	2.8686	3.9337
PMI39_03722	ABC-type multidrug transport system, ATPase	3.5185	2.6527	3.9062
PMI39_00059	Indolepyruvate decarboxylase	-0.0348	1.8268	3.9061
PMI39_00367	Sulfoxide reductase catalytic subunit YedY	3.9026	3.8041	3.8836
PMI39_01901	S-formylglutathione hydrolase FrmB	3.7233	3.2893	3.8043
PMI39_03606	Phosphoribosylglycinamide formyltransferase-1	2.2421	2.567	3.8043
PMI39_02797	Phosphate transport system protein	2.9169	3.1951	3.7987
PMI39_00352	Protein of unknown function	1.393	3.5549	3.6965
PMI39_01997	Hypothetical Protein	1.393	2.4336	3.6702
PMI39_01799	Putative dehydrogenase	2.6077	-0.0686	3.629
PMI39_04031	Molybdate transport system, ATP-binding protein	2.8853	3.0588	3.5427
PMI39_01989	NADH dehydrogenase subunit B	1.8468	1.4219	3.5203
PMI39_00126	Transcriptional regulator, RpiR family	2.8494	-0.0686	3.5202
PMI39_00132	RNAse III	2.1383	2.2112	3.4907
PMI39_04844	Hypothetical Protein	2.9169	2.5443	3.4907
PMI39_00980	Octopine/nopaline transport system	-0.0348	-0.0686	3.4837
PMI39_00202	CDP-diacylglycerol-serine O-phosphatidyltransferase	2.8997	2.4336	3.4604
PMI39_00847	Phosphatase NudJ	3.702	4.375	3.4604
PMI39_01851	dCTP deaminase	1.4607	2.7951	3.3376
PMI39_00356	p-hydroxybenzoic acid efflux pump subunit AaeA	1.4096	2.1012	3.3298
PMI39_01459	Putative Toxin-antitoxin system antitoxin component	1.8468	2.5207	3.3039
PMI39_04094	D-serine deaminase, pyridoxal phosphate-dependent	-0.0348	1.4219	3.3039

Table 3.3 Continued

LocusTag	Predicted Function	5μM	50μM	500μM
PMI39_00821	Integration host factor subunit alpha	2.2801	1.3711	3.1284
PMI39_02138	Phosphate starvation-inducible protein PhoH	-0.0348	1.4219	3.1284
PMI39_04745	Membrane-bound inhibitor of C-type lysozyme	3.3853	3.2056	3.1284
PMI39_02976	Lysophospholipase, alpha-beta hydrolase superfamily	1.8468	1.4219	3.1284
PMI39_02719	Anthranilate synthase, component I	1.3768	1.9522	2.676
PMI39_04394	Glycerol kinase	-0.1381	0.0453	2.6406
PMI39_03735	Aconitase	0.165	0.6411	1.8773
PMI39_02092	Gluconate 2-dehydrogenase	0.5738	0.3469	1.4872
PMI39_02090	2-dehydro-3-deoxygluconokinase	-0.0857	-0.0329	1.4689
PMI39_00875	PST system D-glucose-specific IIB component, Glc	0.9118	0.5602	1.2818
PMI39_02763	Threonine dehydratase	0.8239	1.205	1.1246
PMI39_00913	Glutaredoxin 2	-0.2758	-0.1577	-1.3042
PMI39_01963	Hypothetical Protein	-0.1487	-0.1923	-2.7535
PMI39_00828	Hypothetical Protein	-0.4603	-1.1449	-2.7802
PMI39_01027	Fimbrial chaperone protein	0.6939	0.3742	-2.9354
PMI39_03605	Phosphate transport system ATP-binding protein	-1.5037	-3.1957	-2.9354
PMI39_02043	Amino acid ABC transporter ATP-binding protein	-3.1896	-3.2234	-2.9631
PMI39_02823	Two-component system, OmpR family	-0.1859	-0.1162	-3.1303
PMI39_04666	Uncharacterized conserved protein, DUF1778 family	-3.3652	-0.4617	-3.1387
PMI39_02094	DNA-binding transcriptional regulator, MocR family	-2.0033	-0.9197	-3.2724
PMI39_03268	S-adenosylhomocysteine hydrolase	0.2842	0.2237	-3.28
PMI39_02182	Flagellar FliL protein	-0.2213	-1.1131	-3.4048
PMI39_02311	5-hydroxyisourate hydrolase	-1.0599	-1.3599	-3.567
PMI39_01065	3,4-dihydroxy-2-butanone 4-phosphate synthase	-0.0869	-1.9795	-3.6733
PMI39_00743	Nicotinamidase-related amidase	0.0547	-0.8319	-3.865
PMI39_02040	Polar amino acid transport system substrate-binding	-1.1558	-4.1275	-3.8672
PMI39_03037	Protein of unknown function	-0.4099	-1.036	-4.0671
PMI39_01330	Nitrogen regulatory protein P-II family	-0.636	-1.0404	-4.2146
PMI39_00439	GTPases-translational elongation factors	0.6619	0.0105	-6.1693
PMI39_01931	Hypothetical Protein	2.2756	4.168	3.3724
PMI39_00919	N-methyl-L-tryptophan oxidase	4.0801	4.012	3.6056
PMI39_04563	Predicted DNA-binding protein, MmcQ/YjbR family	1.4607	3.5317	1.7528
PMI39_04190	SOS-response transcriptional repressor LexA	2.8853	3.4246	1.7917

Table 3.3 Continued

LocusTag	Predicted Function	5μM	50μM	500μM
PMI39_03970	Glutamine amidotransferase	2.2756	3.3734	2.2711
PMI39_01082	Hypothetical Protein	-0.0348	3.0908	3.2954
PMI39_00219	SIR2-like domain-containing protein	2.9026	3.0588	3.3425
PMI39_03880	Predicted O-methyltransferase	1.9249	2.9182	2.6304
PMI39_01725	Tat proofreading chaperone FdhE	3.0634	2.9182	2.0691
PMI39_03657	Protein of unknown function	1.6233	2.9182	2.0691
PMI39_04459	NADPH:quinone reductase	3.4238	2.882	4.2627
PMI39_04877	Gallate dioxygenase	-0.0348	2.882	2.935
PMI39_02474	psiF repeat-containing protein	-0.0348	2.882	0.1916
PMI39_00140	L-aspartate oxidase	3.033	2.8713	3.9684
PMI39_01922	Microcin C transport system ATP-binding protein	3.287	2.8713	2.7549
PMI39_01022	DNA-binding transcriptional regulator, MocR family	-0.0348	2.6683	1.6159
PMI39_00836	Fe-S cluster assembly protein SufA	-0.2913	-2.8056	0.2475
PMI39_03702	Iron-sulfur cluster insertion apoprotein erpA	-2.7718	-2.8056	-0.1065
PMI39_00111	ABC-type uncharacterized transport system	-2.7716	-2.8056	-0.3327
PMI39_01285	Hypothetical Protein	-2.9958	-3.0295	-0.5567
PMI39_04680	Hypothetical Protein	4.1889	2.9921	0.1916
PMI39_04926	ATPase, P-type (transporting), subfamily IC	4.0685	3.6001	4.1146
PMI39_04795	Alanine racemase	3.5155	2.8686	3.5536
PMI39_03844	Methyl-accepting chemotaxis sensory transducer	3.4139	1.8268	0.1916
PMI39_01778	Lipopolysaccharide biosynthesis proteins	3.3874	-0.0686	0.1916
PMI39_03122	Phosphoserine phosphatase	3.2582	2.4336	2.6913
PMI39_04643	Iron complex outermembrane receptor protein	3.1359	2.4336	2.0691
PMI39_03045	Putative transcriptional regulator	3.1359	2.1086	3.5536
PMI39_01209	Lipoyl(octanoyl) transferase	3.1237	3.2893	1.6159
PMI39_01125	16S rRNA C1402 (ribose-2'-O) methylase RsmI	3.1237	1.5838	1.6159
PMI39_02390	DNA recombination protein RmuC	2.9533	1.8854	2.6835
PMI39_02372	ATP-dependent DNA helicase UvrD	2.9533	1.5838	3.1719
PMI39_04744	Pyridoxamine 5'-phosphate oxidase	2.9533	1.4219	2.4135
PMI39_00174	Restriction endonuclease	2.9169	2.9993	2.2267
PMI39_04894	DNA gyrase inhibitor	2.9026	-0.0686	2.4135
PMI39_00549	Formate transporter	2.702	2.8578	0.1916
PMI39_04198	DNA-binding transcriptional regulator, LysR family	2.702	2.7235	1.7528

Table 3.3 Continued

LocusTag	Predicted Function	5μM	50μM	500μM
PMI39_03751	GMP reductase	1.5402	1.1401	1.1029
PMI39_02489	Recombination associated protein RdgC	1.1117	0.7819	0.6966
PMI39_04647	lysine/arginine/ornithine transport system	-3.1619	0.2718	0.6472
PMI39_01157	Hypothetical Protein	-3.3567	-0.7548	-0.9177
PMI39_03855	Enamine deaminase RidA, YjgF/YER057c/UK114 family	-3.7626	-1.0875	-2.1118
PMI39_04118	Peptide/nickel transport system substrate-binding	-4.6936	-0.1795	1.1454

Note: Highlighted Grey = Statistically Significant

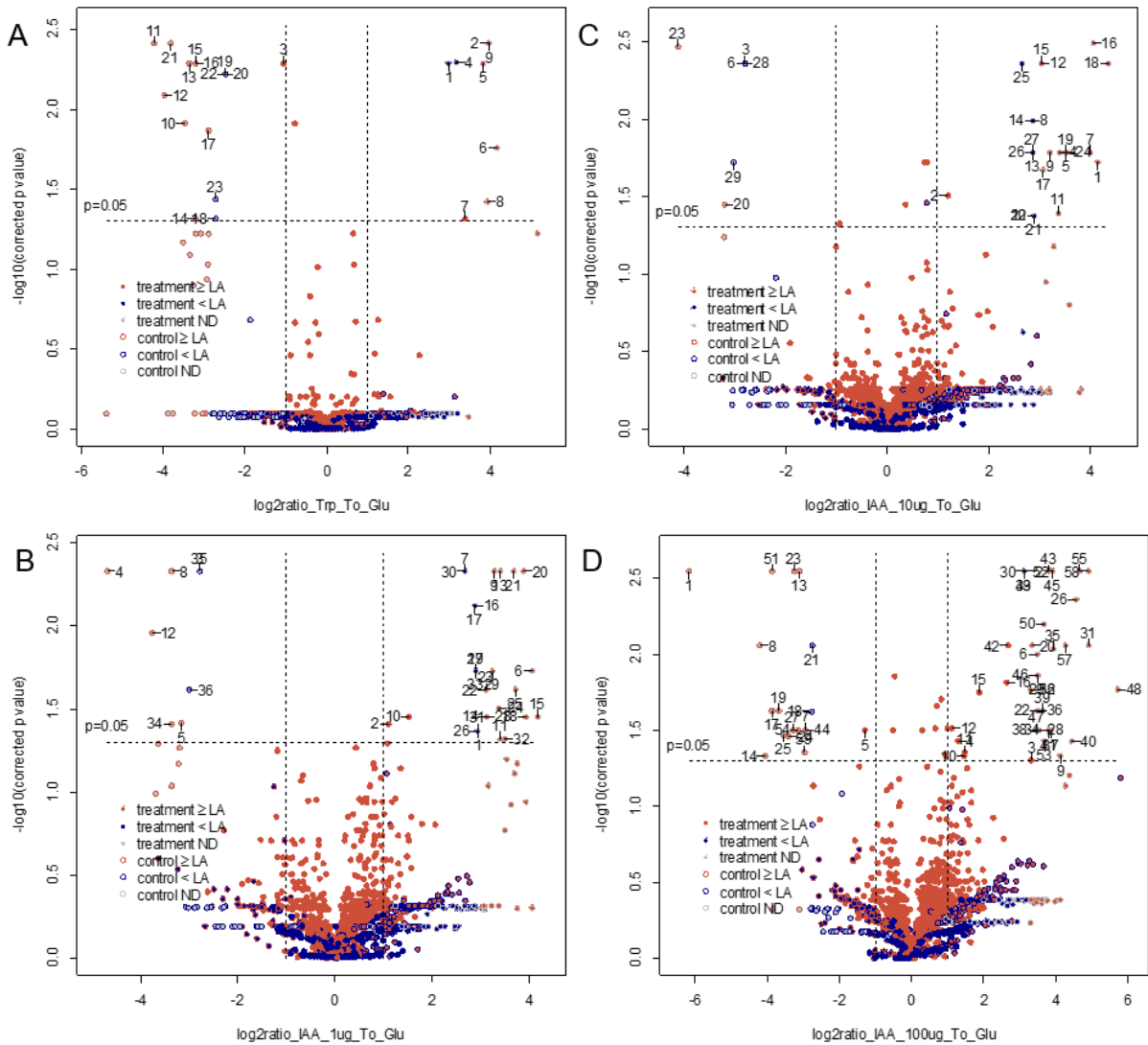


Figure 3.2. Volcano plot illustrates significantly differentially abundant proteins in *Pantoea* sp. YR343 in the present of tryptophan or IAA. The $-\log_{10}$ (Benjamini-Hochberg corrected P value) is plotted against the $\log_2$ (Fold change). A) Tryptophan B) 5 μ M IAA (1 μ g) C) 50 μ M IAA (10 μ g) D) 500 μ M IAA (100 μ g).

under these growth conditions, Table 3.1. Both enzymes predicted to play a role in the tryptamine pathway were also detected under these conditions.

3.4 Induction of *ipdC* gene expression

The proteomics analysis suggests that *Ipdc* is more abundant in the presence of excess tryptophan and high levels of IAA. To determine whether *ipdC* gene expression is upregulated under these conditions, we constructed a reporter plasmid (pPROBE-*ipdC*) in which GFP expression is controlled by the *ipdC* promoter. After transforming this plasmid into wildtype *Pantoea* sp. YR343, we grew cultures in the presence of tryptophan or IAA and measured GFP fluorescence compared to control cultures. These results showed that the *ipdC* promoter was activated in the presence of both tryptophan and IAA, Figure 3.3. An increase in *ipdC* gene expression in the presence of excess tryptophan was confirmed using RT-PCR, which showed a 1.3 fold induction, Figure 3.3.

3.5 Characterization of an *ipdC* mutant strain

Combined, these data suggest that the IPA pathway is the major IAA biosynthesis pathway in *Pantoea* sp. YR343 and that *Ipdc* expression is induced in the presence of tryptophan and IAA. To confirm this, we constructed a mutant strain of *Pantoea* sp. YR343 in which the *ipdC* gene was disrupted. This mutant did not show any defects in growth on plates or in liquid cultures (data not shown). Wildtype and $\Delta ipdC$ mutant cells were grown in minimal media overnight, the cells were removed by centrifugation, and the supernatant was analyzed by GC-MS to determine the metabolite profile. According to this analysis, the dominant metabolites in wildtype cells were indole-3-pyruvate, tryptophol, and IAA, with minor peaks representing indole-3-lactate, and indole-3-carboxyaldehyde (Figure 3.4). Using deuterated-IAA as a standard, these analyses indicate that the wildtype cells produced approximately 2.5 μ g/ml IAA. Tryptophol is produced by the reduction of indole-3-acetaldehyde (IAAld) while oxidation of IAAld produces IAA. IAAld was not detected in the supernatants because the conversion from IAAld to tryptophol is extremely rapid. This metabolite profile is consistent with an active IPA pathway. Moreover, we did not detect any indole-3-acetamide (IAM) or indole-3-acetonitrile (IAN) which would be expected if the IAM or IAN pathways were active in *Pantoea* sp. YR343. Likewise, we prepared samples using a base extraction method, as described in the methods, to look for production of tryptamine but we did not detect this metabolite, suggesting that the tryptamine pathway is not active in *Pantoea* sp. YR343. To ensure tryptamine was not being trapped inside of the cells we also did a whole cell extraction and did not detect this metabolite, again suggesting the tryptamine pathway is not active in *Pantoea* sp. YR343 in the presence of tryptophan, Figure 3.4. Additionally, in the tryptamine extraction we also did not detect any other indolic compounds, the representative peaks are residual acids phenyls, Figure 3.4.

Due to the absence of *ipdC*, the mutant cells showed a different metabolite profile, with the dominant metabolites being indole-3-pyruvate and indole-3-lactate, Figure 3.5. Tryptophol was completely absent in the supernatant from these mutant cells. Since tryptophol is produced from IAAld, we conclude that the $\Delta ipdC$ mutant lacks indole pyruvate decarboxylase activity and is unable to produce IAAld. Interestingly, however, we did detect some IAA in the mutant cells, albeit at much lower levels compared to wild type cells (500 ng/ml). As in the wildtype cells, we also failed to detect tryptamine, IAN, and IAM in the supernatant from mutant cells. Thus, detection of IAA was unexpected

since our genomic, proteomic, and metabolomic data was consistent with the IPA pathway being the only active biosynthetic pathway in *Pantoea* sp. YR343. The presence of IAA in the supernatant from the mutant could certainly suggest that IAA is being synthesized by *Pantoea* sp. YR343, either by an unidentified pathway or one that was erroneously ruled out. It has been noted, however, that IPA is unstable and can spontaneously degrade to IAA [110]. To test whether this may be happening in our samples, we grew cultures of wildtype cells, removed the cells by centrifugation, and collected the supernatant. The supernatant was split into two aliquots, with one aliquot being analyzed by GC-MS immediately and the other aliquot incubated for an additional 24 hours before GC-MS analyses. In the aliquot that was analyzed immediately, IPA was found at about 100µg/ml and IAA at about 860ng/ml. When the supernatant was incubated for an additional 24 hrs in the absence of cells, we saw a decrease in IPA to 3 fold (about 30µg/ml) and an increase in IAA to 2 fold (about 2µg/ml), Figure 3.6. These data are consistent with the spontaneous degradation of IPA into IAA and may explain the accumulation of the IAA in the $\Delta ipdC$ mutant.

3.6 Loss of *Ipdc* activity does not impair root colonization or biofilm formation

Finally, we investigated whether loss of the *ipdc* gene affects other biological behaviors such as plant colonization and biofilm formation. Inactivation of the IPA pathway in *Pantoea* sp. YR343 influenced its ability to associate with plants. The results of these experiments showed that the $\Delta ipdC$ mutant was able to colonize poplar roots as efficiently as the wildtype cells, Figure 3.7 (A). These data suggest that microbial production of IAA is not required to initiate plant association by *Pantoea* sp. YR343. Next, we examined the ability of wild type and $\Delta ipdC$ cells to form biofilms. Using the 96-well plate assay, we found that the $\Delta ipdC$ was not deficient in biofilm formation, Figure 3.7 (B). We also noticed that biofilm formation was not affected by various concentrations of IAA. More importantly we noticed that biofilm formation was increased in the presence of tryptophan for both WT and $\Delta ipdC$. These data suggest that the *ipdc* gene is not required to initiate biofilm formation by *Pantoea* sp. YR343.

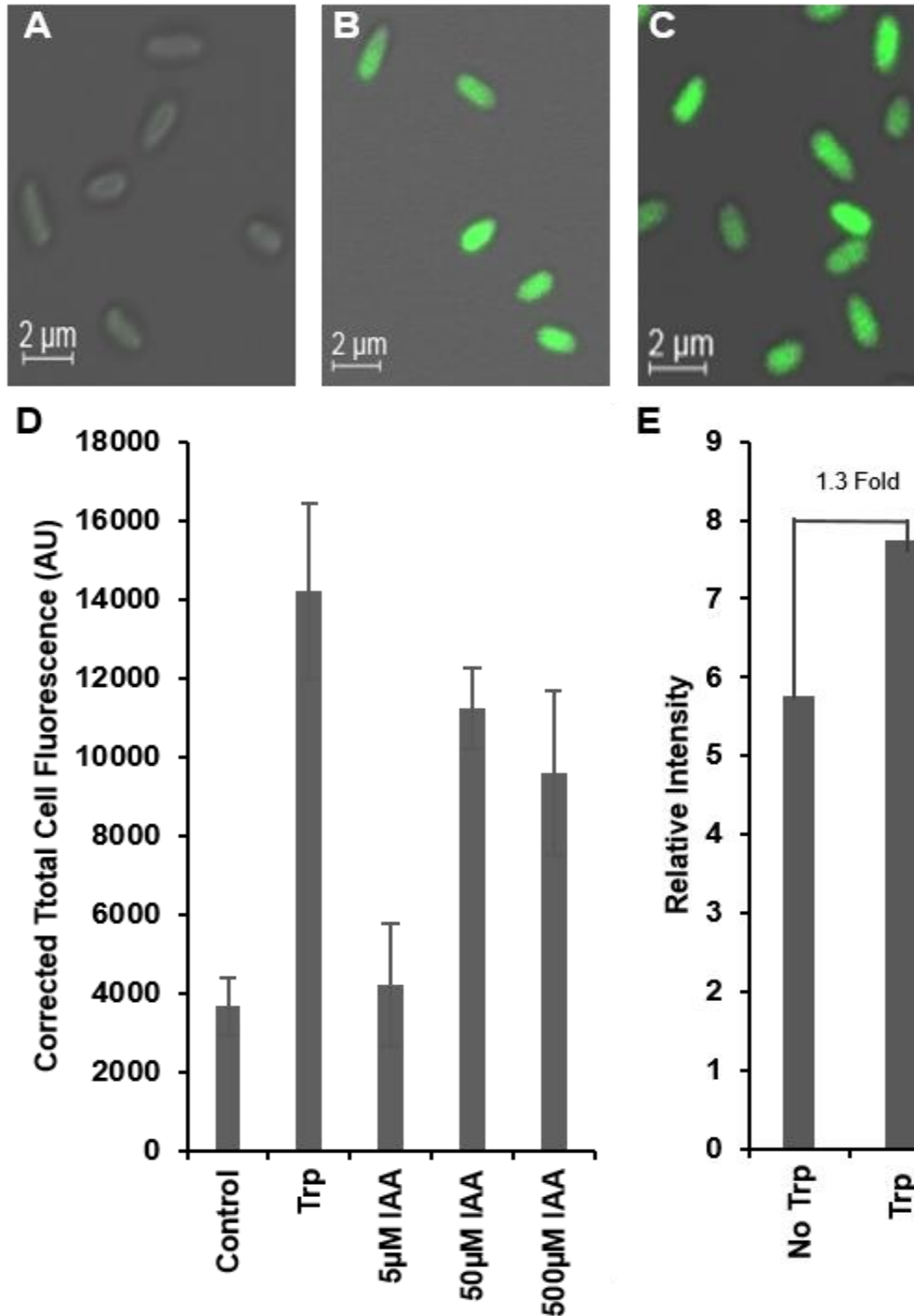


Figure 3.3. *ipdC* gene expression in *Pantoea* sp. YR343. Confocal images of pPROBE-*ipdC* reporter plasmid (A) without tryptophan (B) with tryptophan and (C) with 50 μ M IAA. D) Quantification of pPROBE-*ipdC* fluorescence (Control, Trp, 5 μ M, 50 μ M, and 500 μ M) by confocal microscopy. E) Transcriptomics (RT-PCR) of *Pantoea* sp. YR343 with and without tryptophan, using 16s as a negative control.

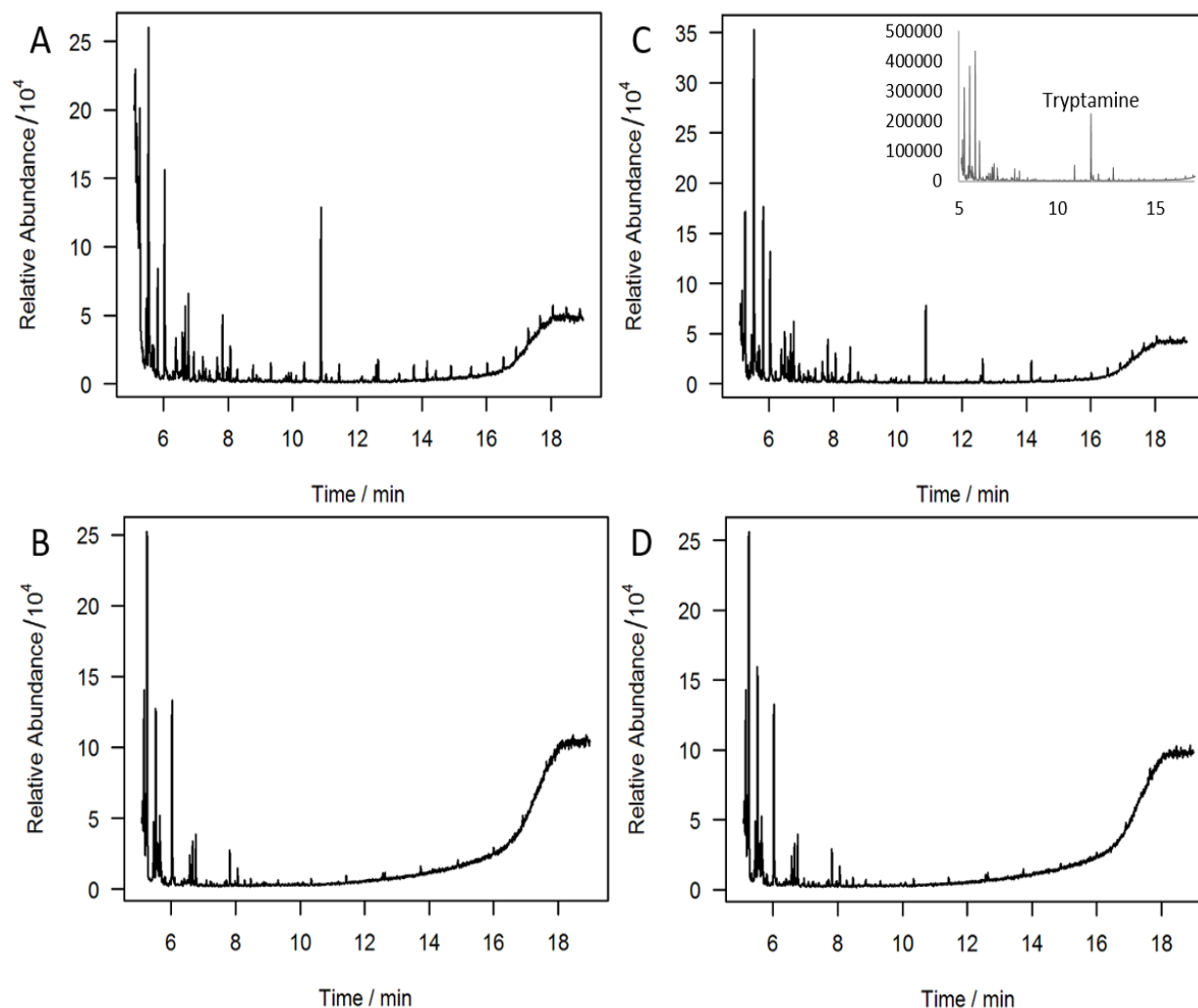


Figure 3.4: Representative chromatograms of *Pantoea* sp. YR343, WT and $\Delta ipdC$ using a basic extraction to identify tryptamine. A) WT supernatant extraction. B) WT whole cell extraction. C) $\Delta ipdC$ supernatant extraction. D) $\Delta ipdC$ whole cell extraction.

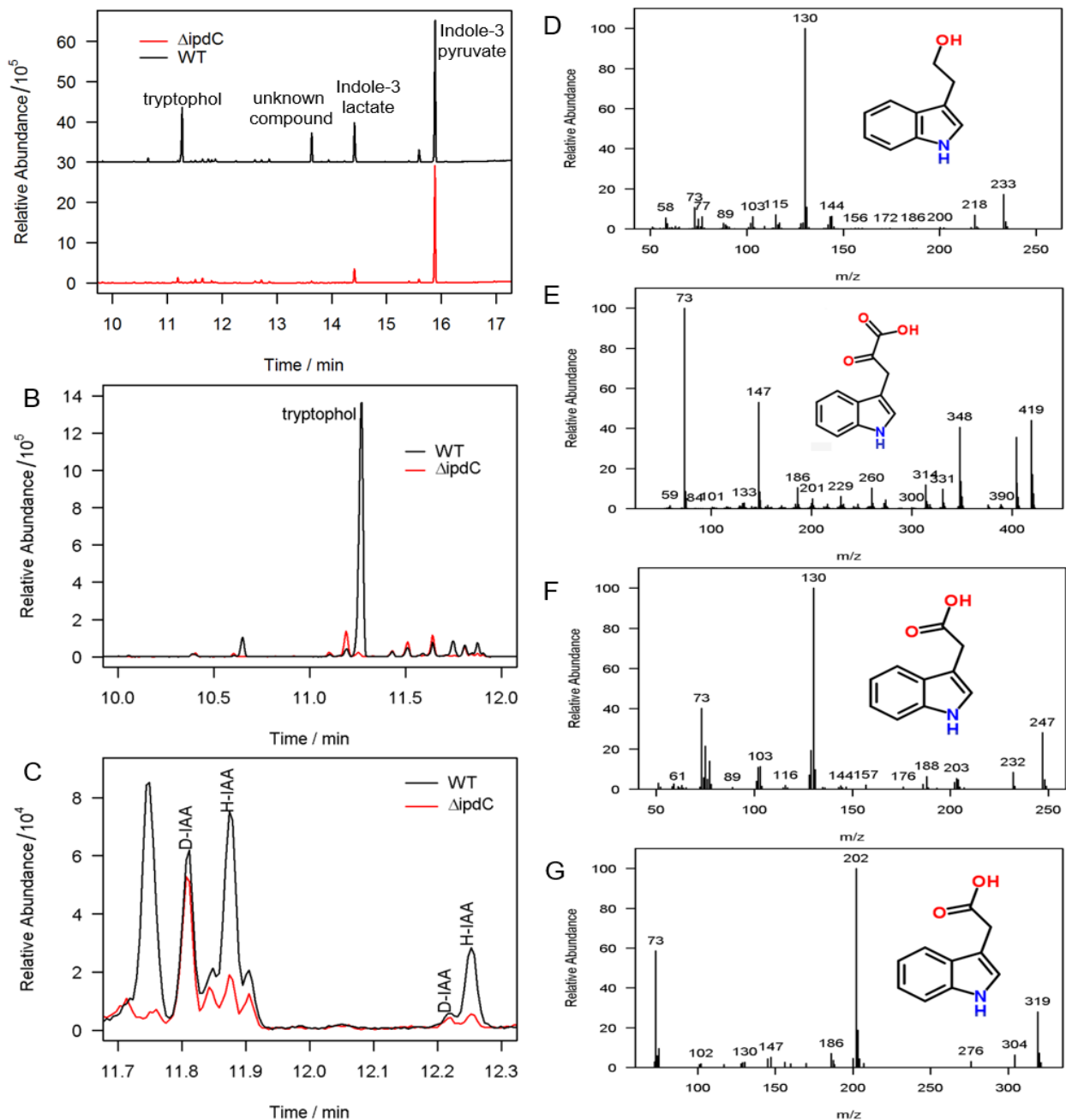


Figure 3.5. Representative metabolomics of *Pantoea* sp. YR343 and *Pantoea* sp. YR343 $\Delta ipdC$. A) Complete chromatogram B) Tryptophol chromatogram C) IAA chromatogram. IAA was singly (m/z 130.1) and doubly (m/z 202.1) derivatized giving two peaks. D) Tryptophol spectra E) IPA spectra F) IAA 130.1 singly derivatized spectra G) IAA 202.1 doubly derivatized spectra. Indolic compound prediction was based on the precursor ion mass and fragmentation patterns through NIST.

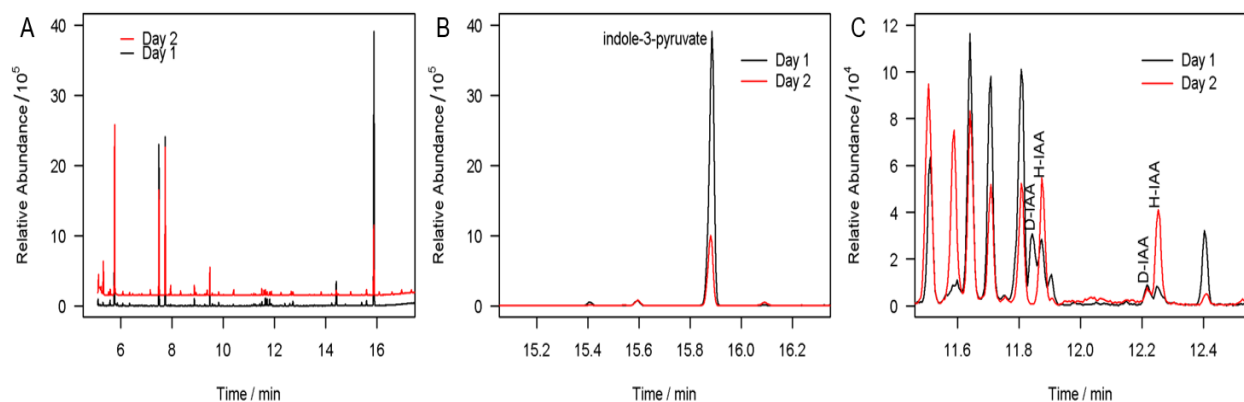


Figure 3.6. Representative metabolomics of spontaneous degradation of IPA in *Pantoea* sp. YR343 and *Pantoea* sp. YR343 $\Delta ipdC$ after 2 days. (A) Complete chromatogram of day 1 and day 2 (B) Chromatogram of IPA alone both day 1 and day 2 (C) Chromatogram of IAA on both day 1 and day 2. IAA was singly (m/z 130.1) and doubly (m/z 202.1) derivatized giving two peaks. Indolic compound prediction was based on the precursor ion mass and fragmentation patterns through NIST.

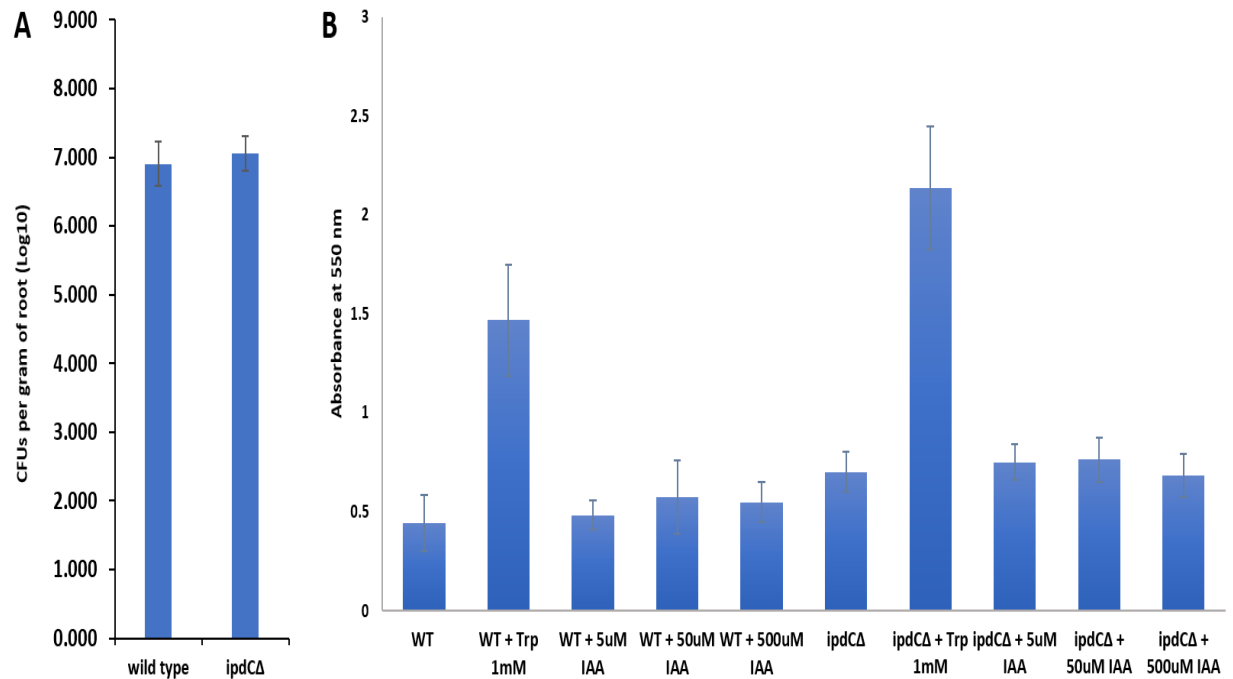


Figure 3.7: *Pantoea* sp. YR343 (WT and $\Delta ipdC$). A) *Populus deltoides* root colonization assay by wild type *Pantoea* sp. YR343 and $\Delta ipdC$ is described as the Log10 value of colony forming units (CFUs) per gram of root material. B) Comparison of biofilm formation between wild type *Pantoea* sp. YR343 and the $\Delta ipdC$ in a plastic 96-well plate measured by the crystal violet assay, as described in the methods

Chapter 4

Discussion

In this dissertation, we use a combination of genomic, transcriptomic, metabolomic, and proteomic analyses to investigate the biosynthesis of IAA by *Pantoea* sp. YR343. The production of IAA by some microbes has been associated with their plant growth promoting behavior [65, 111-114]. In the case of *P. agglomerans*, this growth promotion is thought to be due to enhanced root architecture, which, in turn, can result in enhanced water and nutrient uptake by the plant. The microbe benefits from this interaction by access to nutrients found in plant exudate, as well as increased surface area for colonization. A number of biosynthetic pathways have been described in microbes for IAA production. The ability to assign specific gene products to these pathways using genomic analyses, however, is limited to only a few well-characterized enzymes, such as *Ipdc* and *trp* monooxygenase. Other steps in these IAA synthesis pathways are defined by general functions such as aromatic aminotransferase and aldehyde dehydrogenase, and most bacterial genomes contain many candidate gene products with these annotated functions. Determining which of these candidate gene products is more likely to be involved in IAA biosynthesis is further hindered by the fact that few of the IAA biosynthetic enzymes appear to be co-expressed in operons. Thus, genomic analyses can provide an inventory of candidate gene products, but additional experimentation and/or molecular modeling is required to determine whether these gene products may be involved in IAA biosynthesis.

In *Pantoea* sp. YR343, genomics analyses identified 7 candidate aromatic aminotransferases and 17 candidate aldehyde dehydrogenases that catalyze steps in the IPA pathway, of which only 4 aminotransferases and 9 aldehyde dehydrogenases were detected by proteomics during growth under conditions (excess *trp*) in which IAA is produced. Based on the basic assumption that enzymes involved in IAA biosynthesis should be present under conditions in which cells are producing IAA, the proteins that were identified in our proteomics analyses are more likely to be involved in IAA biosynthesis. It should be noted, however, that none of the enzymes was particularly abundant and the relevant enzymes may be present in cells but not detected in this analysis. Nevertheless, the proteomic results serve to prioritize which enzymes are selected for future analyses. *Ipdc*, on the other hand, was more abundant in the presence of *trp* and IAA, suggesting that its expression is regulated. Indeed, we found that the *ipdc* promoter is activated in the presence of both *trp* and IAA. How *ipdc* gene expression is regulated will be the subject of future studies. Collectively, these data point to a model in which IAA biosynthesis in *Pantoea* sp. YR343 proceeds via the IPA pathway and is regulated by *ipdc* expression.

To better understand the changes to the proteome in the presence of IAA, we grew *Pantoea* sp. YR343 in 5 μ M, 50 μ M or 500 μ M IAA. The localized concentration of IAA typically experienced by *Pantoea* sp. YR343 in the rhizosphere is currently unknown and determining this value presents a measurement challenge that will require advances in chemical imaging methods. Interestingly, growth of *Pantoea* sp. YR343 under each of these conditions produced different sets of differentially abundant proteins. Whether these conditions represent different stages in the initiation and maintenance of plant association is an intriguing possibility, although it is also possible that these results are a reflection of biological variation and instrument sensitivity.

Taken together, it is clear that growth in the presence of IAA results in changes to the proteome in *Pantoea* sp. YR343. These changes are reflected in upregulation of gene products associated with transport of carbohydrates, amino acids, and inorganic ions. In particular, we see increases in proteins annotated to be involved in phosphate, molybdenum, iron, and copper metabolism. Moreover, we see upregulation of gene products involved in EPS biosynthesis. It is tempting to speculate that exposure to IAA induces changes in the physiology of *Pantoea* sp. YR343 that are advantageous for survival in the rhizosphere. For example, sugars and amino acids are primary components of plant exudates; thus, upregulation of these transporters may allow *Pantoea* sp. YR343 to better compete for these resources. Likewise, upregulation of EPS biosynthesis may promote attachment to plant roots and biofilm formation.

Finally, the role of IAA in promoting plant-association was determined by construction of an Δ *ipdC* mutant. That indole pyruvate decarboxylase activity was disrupted was confirmed by metabolite analyses in which the mutant showed no accumulation of tryptophol, which is produced by reduction of IAAld. Conversion of IAAld to tryptophol (reduction) or IAA (oxidation) is likely to be influenced by the redox state of the cell, with growth in minimal media with glucose favoring conversion to tryptophol. In nutrient-starved conditions, such as the rhizosphere, it is likely that conversion of IAAld to IAA is favored. The absence of tryptophol in the mutant also argued against the presence of an active tryptamine pathway since this pathway also produces IAAld. Nevertheless, we did detect some IAA (approximately 20% of wildtype) in the supernatant of the Δ *ipdC* mutant. The simplest explanation for the presence of IAA is that *Pantoea* sp. YR343 harbors more than one pathway for IAA biosynthesis, as has been found in other bacteria [74]. While we cannot say definitively that *Pantoea* sp. YR343 has more than one pathway, our genomic and metabolite data argue against this possibility. Rather, we favor the alternate explanation that IAA is produced in the mutant by non-enzymatic degradation of indole pyruvate under these experimental conditions. Indeed, our experiments show that extended incubation of supernatant collected from *Pantoea* sp. YR343 results in decreasing levels of indole pyruvate and increasing levels of IAA.

Finally, we investigated whether disrupting *ipdC* function had any effect on the ability of *Pantoea* sp. YR343 to colonize poplar roots. Perhaps surprisingly, we found that Δ *ipdC* mutant cells were able to associate with poplar roots and form biofilms as efficiently as wildtype cells, suggesting that IAA is not likely required for initial attachment to the root or biofilm formation. Alternatively, it is possible that another metabolite mediates this interaction or that the mutant is synthesizing IAA by an alternate pathway and this reduced amount of IAA compared to wildtype is sufficient to promote plant association. Another

possibility is that microbial IAA biosynthesis is triggered after plant association and is plays a role in the maintenance of plant-microbe interactions. To better understand plant-microbe association and distinguish between these possibilities, we will need to be able to measure the local concentration and spatial distribution of metabolites produced by microbes during plant association. Indeed, identifying and measuring the temporal and spatial distribution of metabolites in complex systems is an exciting and challenging new area of research that will likely drive technical and scientific advances.

Chapter 5

Conclusions and Future Directions

Understanding biocomplexity within and between a multiorganismal system is a complex matter that requires a multitude of integrated omics approaches. In this dissertation, we presented findings from our omics study on the internal biochemical pathways of indole-3-acetic acid (IAA) in the bacteria *Pantoea* sp. YR343. The essential findings of this dissertation can be summarized as follows: Data extracted from multiple omics approaches, lead to the novel discovery that the main pathway associated with IAA biosynthesis in *Pantoea* sp. YR343 is the indole-pyruvate pathway. The combination of existing experimental data from our omics results indicates that *Pantoea* sp. YR343 produces about 2.5 μg of IAA per mL of culture, primarily through the indole pyruvate pathway. To confirm these results, we disrupted the *ipdC* gene in *Pantoea* sp. YR343 and analyzed its ability to produce IAA in the presence of tryptophan. We found that, empirically, disrupting the *ipdC* gene in *Pantoea* sp. YR343 dramatically reduced its ability to produce IAA in the presence of tryptophan.

Using proteomics, we investigated *Pantoea* sp. YR343 response to exogenous IAA. Previous studies identified that IAA can act a positive regulator of the *ipdC* gene [76, 81], hence our next goal was to determine the effects of IAA exposure on *Pantoea* sp. YR343. Surprisingly, we found that at a very low concentration 50 μM , more than 20 times less than observed in *A. brasilense*, we saw an increase of the *ipdC* gene expression, signifying IAA acts a positive regulator in *Pantoea* sp. YR343 [76]. More importantly, we noticed that the tryptophan biosynthesis genes were also differentially expressed in our IAA samples, indicating that when exposed to IAA *Pantoea* sp. YR343 increases its tryptophan biosynthetic pathways to make IAA.

Another important observation was that apart from the genes involved in IAA biosynthesis (ex *ipdC*), several other proteins were also differentially expressed. In particular, we found pathways associated with phosphate regulation, inorganic ion transport and metabolism, carbohydrate and amino acid transport and metabolism, and aromatic carboxylic acid efflux, all of which could be essential characteristics for bacterial survival in the rhizosphere.

Omics has become an indispensable approach to the studying of biological functions within complex biological ecosystems, specifically in plant-microbe associations. The omics approach in this dissertation allowed us to monitor *Pantoea* sp. YR343 under various external stimuli, revealing specific biosynthetic pathways associated with IAA production in *Pantoea* sp. YR343. We also showed that given a certain stimuli *Pantoea* sp. YR343 proteome changes which may ultimately influence how it responds in the environment given certain ques. The findings presented here only look at a one-dimensional aspect of microbial interactions. Nevertheless, it is with certainty that we are

still just scratching the surface of understanding these complex biochemical interactions in *Pantoea* sp. YR343, let alone the plant-microbe associations on a large scale.

The data presented in this dissertation calls for a new state of mind to consider the relationship between *Pantoea* sp. YR343 and its associated plant host, *P. deltoides*. Predicting IAA was a contributing phytohormone facilitating colonization between *Pantoea* sp. YR343 and *P. deltoides*, we performed colonization tests with our mutant deficient in IAA and found that *Pantoea* sp. YR343 colonization to *P. deltoides* was not affected. This means that there are other biological implications or phytohormones that could be contributing to the colonization of *Pantoea* sp. YR343 to *P. deltoides*. Understanding what types of cues or biochemical responses drive these responses is a fundamental area of research that needs to be addressed. Indeed, identifying and measuring the temporal and spatial distribution of metabolites in complex systems is an exciting and challenging new area of research that will likely drive technical and scientific advances. Currently, it would be difficult to identify the localized concentration of IAA or other phytohormones expressed by *Pantoea* sp. YR343 in the presence of *P. deltoides* because both *Pantoea* sp. YR343 and *P. deltoides* are capable of producing a variety of metabolites.

One approach to measuring the local concentration of the metabolites produced by both *Pantoea* sp. YR343 and *P. deltoides* we be to create a lab on a chip utilizing mass-spectrometry-based monitoring in combination with a microfluidic platform [115]. A microfluidic platform harboring *P. deltoides* roots along with the bacterial inoculum can be developed. The uniqueness of such a platform would be the ability to collect rhizosphere materials through the 20µm ports which would allow us to perform live LC-MS-MS. This platform could help us detect enzymes and other phytohormones, which might give us clues as to how plants and microbes communicate.

Increasing our understanding about bacterial colonization and the cues it utilizes from the environment would greatly improve our knowledge about the complexity between plant-microbe environments. Large plant-microbe community studies have begun to uncover a variety of details about plant-microbe interconnectivity, but have ultimately been limited due to experimental analysis. Utilizing high-throughput technologies and creating new platforms will further the advancements in understanding the complexity within a plant-microbe environment.

References

1. Gross, N., et al., *Plant response traits mediate the effects of subalpine grasslands on soil moisture*. New Phytol, 2008. **180**(3): p. 652-62.
2. Soliveres, S. and F.T. Maestre, *Plant-plant interactions, environmental gradients and plant diversity: a global synthesis of community-level studies*. 2014. **16**(4): p. 154-163.
3. Chapin, F.S., 3rd, et al., *Consequences of changing biodiversity*. Nature, 2000. **405**(6783): p. 234-42.
4. Hahn, M. and K. Mendgen, *Signal and nutrient exchange at biotrophic plant-fungus interfaces*. Curr Opin Plant Biol, 2001. **4**(4): p. 322-7.
5. Turner, T.R., E.K. James, and P.S. Poole, *The plant microbiome*. Genome Biol, 2013. **14**(6): p. 209.
6. Selin, C., et al., *Elucidating the Role of Effectors in Plant-Fungal Interactions: Progress and Challenges*. Front Microbiol, 2016. **7**: p. 600.
7. Kemen, E., *Microbe-microbe interactions determine oomycete and fungal host colonization*. Curr Opin Plant Biol, 2014. **20**: p. 75-81.
8. Kaisermann, A., et al., *Legacy effects of drought on plant-soil feedbacks and plant-plant interactions*. New Phytol, 2017.
9. Bakker, M.G., et al., *Diffuse symbioses: roles of plant-plant, plant-microbe and microbe-microbe interactions in structuring the soil microbiome*. Mol Ecol, 2014. **23**(6): p. 1571-83.
10. Atkins, S.D. and I.M. Clark, *Fungal molecular diagnostics: a mini review*. J Appl Genet, 2004. **45**(1): p. 3-15.
11. Wu, S., et al., *Shifts of microbial community structure in soils of a photovoltaic plant observed using tag-encoded pyrosequencing of 16S rRNA*. Appl Microbiol Biotechnol, 2016. **100**(8): p. 3735-45.
12. Timmers, R.A., et al., *Microbial community structure elucidates performance of *Glyceria maxima* plant microbial fuel cell*. Appl Microbiol Biotechnol, 2012. **94**(2): p. 537-48.
13. Schlatter, D.C., et al., *Plant community richness and microbial interactions structure bacterial communities in soil*. Ecology, 2015. **96**(1): p. 134-42.
14. Chapman, S.K. and G.S. Newman, *Biodiversity at the plant-soil interface: microbial abundance and community structure respond to litter mixing*. Oecologia, 2010. **162**(3): p. 763-9.
15. Hughes, D.T. and V. Sperandio, *Inter-kingdom signalling: communication between bacteria and their hosts*. Nat Rev Microbiol, 2008. **6**(2): p. 111-20.
16. Metrak, M., et al., *Nature's patchwork: How water sources and soil salinity determine the distribution and structure of halophytic plant communities in arid environments of the Eastern Pamir*. PLoS One, 2017. **12**(3): p. e0174496.
17. Khumairoh, U., J.C. Groot, and E.A. Lantinga, *Complex agro-ecosystems for food security in a changing climate*. Ecol Evol, 2012. **2**(7): p. 1696-704.

18. Yang, D., et al., *Diversity and distribution of the prokaryotic community in near-surface permafrost sediments in the Tianshan Mountains, China*. Can J Microbiol, 2008. **54**(4): p. 270-80.
19. Sun, J., et al., *[Classification, species diversity, and species distribution gradient of permafrost wetland plant communities in Great Xing' an Mountains valleys' of northeast China]*. Ying Yong Sheng Tai Xue Bao, 2009. **20**(9): p. 2049-56.
20. Young, H.S., et al., *Introduced Species, Disease Ecology, and Biodiversity-Disease Relationships*. Trends Ecol Evol, 2017. **32**(1): p. 41-54.
21. Schmidt, J., et al., *Effects of plant-symbiotic relationships on the living soil microbial community and microbial necromass in a long-term agro-ecosystem*. Sci Total Environ, 2017. **581-582**: p. 756-765.
22. Adnani, N., S.R. Rajski, and T.S. Bugni, *Symbiosis-inspired approaches to antibiotic discovery*. Nat Prod Rep, 2017. **34**(7): p. 784-814.
23. Shurson, G.C., *The Role of Biofuels Coproducts in Feeding the World Sustainably*. Annu Rev Anim Biosci, 2017. **5**: p. 229-254.
24. Hood, E.E., *Plant-based biofuels*. F1000Res, 2016. **5**.
25. Anderson, O.R., *A model of biocomplexity and its application to the analysis of some terrestrial and marsh eukaryotic microbial communities with an emphasis on amoeboid protists*. J Eukaryot Microbiol, 2003. **50**(2): p. 86-91.
26. Hansen, R.H., et al., *Stochastic Assembly of Bacteria in Microwell Arrays Reveals the Importance of Confinement in Community Development*. PLoS One, 2016. **11**(5): p. e0155080.
27. Rai, A., K. Saito, and M. Yamazaki, *Integrated omics analysis of specialized metabolism in medicinal plants*. Plant J, 2017. **90**(4): p. 764-787.
28. Dettmer, K., P.A. Aronov, and B.D. Hammock, *Mass spectrometry-based metabolomics*. Mass Spectrom Rev, 2007. **26**(1): p. 51-78.
29. Field, B. and A.E. Osbourn, *Metabolic diversification--independent assembly of operon-like gene clusters in different plants*. Science, 2008. **320**(5875): p. 543-7.
30. Chain, P., et al., *An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges*. Brief Bioinform, 2003. **4**(2): p. 105-23.
31. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Rev Genet, 2009. **10**(1): p. 57-63.
32. Yang, W., H. Steen, and M.R. Freeman, *Proteomic approaches to the analysis of multiprotein signaling complexes*. Proteomics, 2008. **8**(4): p. 832-51.
33. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review*. Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.
34. Hoffmann, T., et al., *Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow*. Anal Chem, 2014. **86**(21): p. 10780-8.

35. Baran, R., et al., *Functional genomics of novel secondary metabolites from diverse cyanobacteria using untargeted metabolomics*. Mar Drugs, 2013. **11**(10): p. 3617-31.
36. Varden, F.A., et al., *Taking the stage: effectors in the spotlight*. Curr Opin Plant Biol, 2017. **38**: p. 25-33.
37. Gehlenborg, N., et al., *Visualization of omics data for systems biology*. Nat Methods, 2010. **7**(3 Suppl): p. S56-68.
38. Tebani, A., et al., *Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations*. Int J Mol Sci, 2016. **17**(9).
39. Tebani, A., et al., *Optimization of a liquid chromatography ion mobility-mass spectrometry method for untargeted metabolomics using experimental design and multivariate data analysis*. Anal Chim Acta, 2016. **913**: p. 55-62.
40. Grady, S.L., et al., *A comprehensive multi-omics approach uncovers adaptations for growth and survival of Pseudomonas aeruginosa on n-alkanes*. BMC Genomics, 2017. **18**(1): p. 334.
41. Wullschlegel, S.D., G.A. Tuskan, and S.P. DiFazio, *Genomics and the tree physiologist*. Tree Physiol, 2002. **22**(18): p. 1273-6.
42. Gottel, N.R., et al., *Distinct microbial communities within the endosphere and rhizosphere of Populus deltoides roots across contrasting soil types*. Appl Environ Microbiol, 2011. **77**(17): p. 5934-44.
43. Shakya, M., et al., *A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature Populus deltoides trees*. PLoS One, 2013. **8**(10): p. e76382.
44. Mendes, R., P. Garbeva, and J.M. Raaijmakers, *The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms*. FEMS Microbiol Rev, 2013. **37**(5): p. 634-63.
45. Berendsen, R.L., C.M. Pieterse, and P.A. Bakker, *The rhizosphere microbiome and plant health*. Trends Plant Sci, 2012. **17**(8): p. 478-86.
46. Compant, S., C. Clement, and A. Sessitsch, *Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization*. Soil Biology & Biochemistry, 2010. **42**(5): p. 669-678.
47. Hayat, R., et al., *Soil beneficial bacteria and their role in plant growth promotion: a review*. Annals of Microbiology, 2010. **60**(4): p. 579-598.
48. Haiyambo, D.H., P.M. Chimwamurombe, and B. Reinhold-Hurek, *Isolation and Screening of Rhizosphere Bacteria from Grasses in East Kavango Region of Namibia for Plant Growth Promoting Characteristics*. Curr Microbiol, 2015. **71**(5): p. 566-71.
49. Cecagno, R., T.E. Fritsch, and I.S. Schrank, *The plant growth-promoting bacteria Azospirillum amazonense: genomic versatility and phytohormone pathway*. Biomed Res Int, 2015. **2015**: p. 898592.

50. Amaresan, N., et al., *Plant growth-promoting potential of bacteria isolated from active volcano sites of Barren Island, India*. Lett Appl Microbiol, 2014. **58**(2): p. 130-7.
51. Beneduzi, A., A. Ambrosini, and L.M.P. Passaglia, *Plant growth-promoting rhizobacteria (PGPR): Their potential as antagonists and biocontrol agents*. Genetics and Molecular Biology, 2012. **35**(4): p. 1044-1051.
52. Sabir, A., et al., *Growth and mineral acquisition response of grapevine rootstocks (Vitis spp.) to inoculation with different strains of plant growth-promoting rhizobacteria (PGPR)*. J Sci Food Agric, 2012. **92**(10): p. 2148-53.
53. Glick, B.R., *Plant growth-promoting bacteria: mechanisms and applications*. Scientifica (Cairo), 2012. **2012**: p. 963401.
54. Glick, B.R., et al., *Promotion of plant growth by ACC deaminase-producing soil bacteria*. European Journal of Plant Pathology, 2007. **119**(3): p. 329-339.
55. Bailly, A., et al., *The inter-kingdom volatile signal indole promotes root development by interfering with auxin signalling*. Plant J, 2014. **80**(5): p. 758-71.
56. Shakya, M., et al., *A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature Populus deltoides trees*. PLoS ONE, 2013. **8**(10): p. e76382.
57. Walterson, A.M. and J. Stavriniades, *Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae*. Fems Microbiology Reviews, 2015. **39**(6): p. 968-984.
58. Hebishima, T., et al., *Immune recovery effects of immunopotentiator from Pantoea agglomerans 1 (IP-PA1) on low antibody productions in response to Salmonella enteritidis vaccine and sheep red blood cells in dexamethasone-treated stressed chicken models*. J Vet Med Sci, 2010. **72**(4): p. 435-42.
59. Johnson, K.B., et al., *Assessment of Environmental Factors Influencing Growth and Spread of Pantoea agglomerans on and Among Blossoms of Pear and Apple*. Phytopathology, 2000. **90**(11): p. 1285-94.
60. Yoshida, A., et al., *Improvement of allergic dermatitis via regulation of the Th1/Th2 immune system balance by macrophages activated with lipopolysaccharide derived from Pantoea agglomerans (IP-PA1)*. Anticancer Res, 2009. **29**(11): p. 4867-70.
61. Brady, C.L., et al., *Pantoea allii sp. nov., isolated from onion plants and seed*. Int J Syst Evol Microbiol, 2011. **61**(Pt 4): p. 932-7.
62. Nadarasah, G. and J. Stavriniades, *Quantitative evaluation of the host-colonizing capabilities of the enteric bacterium Pantoea using plant and insect hosts*. Microbiology, 2014. **160**(Pt 3): p. 602-15.
63. Brady, C.L., et al., *Pantoea rodasii sp nov., Pantoea rwandensis sp nov and Pantoea wallisii sp nov., isolated from Eucalyptus*. International Journal of Systematic and Evolutionary Microbiology, 2012. **62**: p. 1457-1464.

64. Stewart, F.C., *A bacterial disease of sweet corn*. N.Y. Agric. Exp. Sta. Bull, 1897. **130**: p. 422-439.
65. Sergeeva, E., D.M. Hirkala, and L. Nelson, *Production of indole-3-acetic acid, aromatic amino acid aminotransferase activities and plant growth promotion by Pantoea agglomerans rhizosphere isolates*. Plant and Soil, 2007. **297**(1-2): p. 1-13.
66. Duca, D., et al., *Indole-3-acetic acid in plant-microbe interactions*. Antonie Van Leeuwenhoek, 2014. **106**(1): p. 85-125.
67. Ryu, R.J. and C.L. Patten, *Aromatic amino acid-dependent expression of indole-3-pyruvate decarboxylase is regulated by TyrR in Enterobacter cloacae UW5*. J Bacteriol, 2008. **190**(21): p. 7200-8.
68. Kulkarni, G.B., et al., *Indole-3-acetic acid biosynthetic pathway and aromatic amino acid aminotransferase activities in Pantoea dispersa strain GPK*. Letters in Applied Microbiology, 2013. **56**(5): p. 340-347.
69. Bible, A.N., et al., *A Carotenoid-Deficient Mutant in Pantoea sp. YR343, a Bacteria Isolated from the Rhizosphere of Populus deltoides, Is Defective in Root Colonization*. Front Microbiol, 2016. **7**: p. 491.
70. Bible, A.N., et al., *A carotenoid-deficient mutant in Pantoea sp. YR343, a bacteria isolated from the rhizosphere of Populus deltoides, is defective in root colonization*. Frontiers in Microbiology, 2016. **7**: p. 491.
71. Spaepen, S., J. Vanderleyden, and R. Remans, *Indole-3-acetic acid in microbial and microorganism-plant signaling*. FEMS Microbiol Rev, 2007. **31**(4): p. 425-48.
72. Spaepen, S. and J. Vanderleyden, *Auxin and plant-microbe interactions*. Cold Spring Harb Perspect Biol, 2011. **3**(4).
73. Patten, C.L., A.J. Blakney, and T.J. Coulson, *Activity, distribution and function of indole-3-acetic acid biosynthetic pathways in bacteria*. Crit Rev Microbiol, 2013. **39**(4): p. 395-415.
74. Spaepen, S. and J. Vanderleyden, *Auxin and Plant-Microbe Interactions*. Cold Spring Harbor Perspectives in Biology, 2011. **3**(4).
75. Patten, C.L. and B.R. Glick, *Role of Pseudomonas putida indoleacetic acid in development of the host plant root system*. Appl Environ Microbiol, 2002. **68**(8): p. 3795-801.
76. Vande Broek, A., et al., *Transcriptional analysis of the Azospirillum brasilense indole-3-pyruvate decarboxylase gene and identification of a cis-acting sequence involved in auxin responsive expression*. Molecular plant-microbe interactions, 2005. **18**(4): p. 311-323.
77. Zimmer, W., M. Wesche, and L. Timmermans, *Identification and isolation of the indole-3-pyruvate decarboxylase gene from Azospirillum brasilense Sp7: sequencing and functional analysis of the gene locus*. Curr Microbiol, 1998. **36**(6): p. 327-31.

78. Van Puyvelde, S., et al., *Transcriptome analysis of the rhizosphere bacterium Azospirillum brasilense reveals an extensive auxin response*. Microbial ecology, 2011. **61**(4): p. 723-728.
79. Carreno-Lopez, R., et al., *Physiological evidence for differently regulated tryptophan-dependent pathways for indole-3-acetic acid synthesis in Azospirillum brasilense*. Molecular and General Genetics, 2000. **264**(4): p. 521-530.
80. Lambrecht, M., et al., *Indole-3-acetic acid: a reciprocal signalling molecule in bacteria-plant interactions*. Trends Microbiol, 2000. **8**(7): p. 298-300.
81. Vande Broek, A., et al., *Auxins Upregulate Expression of the Indole-3-Pyruvate Decarboxylase Gene in Azospirillum brasilense*. Journal of Bacteriology, 1999. **181**(4): p. 1338-1342.
82. Miller, W.G., J.H. Leveau, and S.E. Lindow, *Improved gfp and inaZ broad-host-range promoter-probe vectors*. Mol Plant Microbe Interact, 2000. **13**(11): p. 1243-50.
83. Selinummi, J., et al., *Software for quantification of labeled bacteria from digital microscope images by automated image analysis*. BioTechniques, 2005. **39**(6): p. 859-863.
84. Chourey, K., et al., *Direct Cellular Lysis/Protein Extraction Protocol for Soil Metaproteomics*. Journal of Proteome Research, 2010. **9**(12): p. 6615-6622.
85. Thompson, M.R., et al., *Dosage-dependent proteome response of Shewanella oneidensis MR-1 to acute chromate challenge*. Journal of Proteome Research, 2007. **6**(5): p. 1745-1757.
86. Brown, S.D., et al., *Molecular Dynamics of the Shewanella oneidensis Response to Chromate Stress*. Molecular Cellular Proteomics, 2006. **5**(6): p. 1054-1071.
87. Sharma, R., et al., *Coupling a detergent lysis/cleanup methodology with intact protein fractionation for enhanced proteome characterization*. Journal of Proteome Research, 2012. **11**(12): p. 6008-6018.
88. McDonald, W.H., et al., *Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT*. International Journal of Mass Spectrometry, 2002. **219**(1): p. 245-251.
89. Crosby, H.A., et al., *System-wide Studies of N-Lysine Acetylation in Rhodopseudomonas palustris Reveal Substrate Specificity of Protein Acetyltransferases*. Journal of Biological Chemistry, 2012. **287**(19): p. 15590-15601.
90. Tabb, D.L., C.G. Fernando, and M.C. Chambers, *MyriMatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis*. Journal of Proteome Research, 2007. **6**(2): p. 654-661.
91. Brown, S.D., et al., *Twenty-One Genome Sequences from Pseudomonas Species and 19 Genome Sequences from Diverse Bacteria Isolated from the Rhizosphere and Endosphere of Populus deltoides*. Journal of Bacteriology, 2012. **194**(21): p. 5991-5993.

92. Markowitz, V.M., et al., *IMG 4 version of the integrated microbial genomes comparative analysis system*. Nucleic Acids Research, 2014. **42**(D1): p. D560-D567.
93. Ma, W., et al., *Pharmacological characterization of pannexin-1 currents expressed in mammalian cells*. J Pharmacol Exp Ther, 2009. **328**(2): p. 409-18.
94. Liu, H.B., R.G. Sadygov, and J.R. Yates, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics*. Analytical Chemistry, 2004. **76**(14): p. 4193-4201.
95. Zhang, Y., et al., *Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins*. Analytical Chemistry, 2010. **82**(6): p. 2272-2281.
96. Zybaylov, B., et al., *Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces cerevisiae*. Journal of Proteome Research, 2006. **5**(9): p. 2339-2347.
97. Lochner, A., et al., *Label-free Quantitative Proteomics for the Extremely Thermophilic Bacterium Caldicellulosiruptor obsidiansis Reveal Distinct Abundance Patterns upon Growth on Cellobiose, Crystalline Cellulose, and Switchgrass*. Journal of Proteome Research, 2011. **10**(12): p. 5302-5314.
98. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society Series B-Methodological, 1995. **57**(1): p. 289-300.
99. Alexeyev, M.F., *The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria*. Biotechniques, 1999. **26**(5): p. 824-6, 828.
100. O'Toole, G.A. and R. Kolter, *Initiation of biofilm formation in Pseudomonas fluorescens WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis*. Mol Microbiol, 1998. **28**(3): p. 449-61.
101. Tram, G., V. Korolik, and C.J. Day, *MBDS Solvent: An improved method for assessment of biofilms*. Advances in Microbiology, 2013. **3**(2): p. 200-204.
102. Oberhansli, T., G. Dfago, and D. Haas, *Indole-3-acetic acid (IAA) synthesis in the biocontrol strain CHA0 of Pseudomonas fluorescens: role of tryptophan side chain oxidase*. J Gen Microbiol, 1991. **137**(10): p. 2273-9.
103. Santos-Beneit, F., *The Pho regulon: a huge regulatory network in bacteria*. Frontiers in Microbiology, 2015. **6**(402).
104. Gardner, S.G., et al., *The PhoU Protein from Escherichia coli Interacts with PhoR, PstB, and Metals To Form a Phosphate-Signaling Complex at the Membrane*. Journal of Bacteriology, 2014. **196**(9): p. 1741-1752.
105. Rice, C.D., et al., *Employment of a Promoter-Swapping Technique Shows that PhoU Modulates the Activity of the PstSCAB2 ABC Transporter in Escherichia coli*. Applied and Environmental Microbiology, 2009. **75**(3): p. 573-582.

106. Van Dyk, T.K., et al., *Characterization of the Escherichia coli AaeAB Efflux Pump: a Metabolic Relief Valve?* Journal of Bacteriology, 2004. **186**(21): p. 7196-7204.
107. Kvist, M., V. Hancock, and P. Klemm, *Inactivation of Efflux Pumps Abolishes Bacterial Biofilm Formation.* Applied and Environmental Microbiology, 2008. **74**(23): p. 7376-7382.
108. Bernhard, F., D.L. Coplin, and K. Geider, *Genetic characterization of a DNA region involved in amylovoran synthesis of Erwinia amylovora by complementation of Erwinia stewartii cps mutants.* Plant Pathogenic Bacteria, 1994. **66**: p. 539-544.
109. Langlotz, C., et al., *Biosynthesis of the repeating units of the exopolysaccharides amylovoran from Erwinia amylovora and stewartan from Pantoea stewartii.* Physiological and Molecular Plant Pathology, 2011. **75**(4): p. 163-169.
110. Bentley, J.A., et al., *Some chemical and physiological properties of 3-indolylpyruvic acid.* Biochemical Journal, 1956. **64**(1): p. 44-49.
111. Brandl, M.T. and S.E. Lindow, *Contribution of indole-3-acetic acid production to the epiphytic fitness of Erwinia herbicola.* Applied and Environmental Microbiology, 1998. **64**(9): p. 3256-3263.
112. Dobbelaere, S., et al., *Phytostimulatory effect of Azospirillum brasilense wild type and mutant strains altered in IAA production on wheat.* Plant Soil., 1999. **212**: p. 155-164.
113. Manulis, S., et al., *Differential involvement of indole-3-acetic acid biosynthetic pathways in pathogenicity and epiphytic fitness of Erwinia herbicola pv, gypsophillae.* Molecular Plant-Microbe Interactions, 1998. **11**(7): p. 634-642.
114. Perrig, D., et al., *Plant-growth-promoting compounds produced by two agronomically important strains of Azospirillum brasilense, and implications for inoculant formulation.* Appl Microbiol Biotechnol, 2007. **75**(5): p. 1143-50.
115. Cong, Y., et al., *Mass spectrometry-based monitoring of millisecond protein-ligand binding dynamics using an automated microfluidic platform.* Lab Chip, 2016. **16**(9): p. 1544-8.
116. Gooptu, B., J.A. Dickens, and D.A. Lomas, *The molecular and cellular pathology of alpha(1)-antitrypsin deficiency.* Trends Mol Med, 2014. **20**(2): p. 116-27.
117. Stoller, J.K. and L.S. Aboussouan, *Alpha1-antitrypsin deficiency.* Lancet, 2005. **365**(9478): p. 2225-36.
118. Stoller, J.K. and L.S. Aboussouan, *A review of alpha1-antitrypsin deficiency.* Am J Respir Crit Care Med, 2012. **185**(3): p. 246-59.
119. Abboud, R.T., et al., *Alpha1-antitrypsin deficiency: a clinical-genetic overview.* Appl Clin Genet, 2011. **4**: p. 55-65.
120. Ekeowa, U.I., et al., *Defining the mechanism of polymerization in the serpinopathies.* Proc Natl Acad Sci U S A, 2010. **107**(40): p. 17146-51.
121. Campos, M.A., et al., *Exacerbations in subjects with alpha-1 antitrypsin deficiency receiving augmentation therapy.* Respir Med, 2009. **103**(10): p. 1532-9.

122. Traclet, J., et al., *Augmentation therapy of alpha-1 antitrypsin deficiency associated emphysema*. Rev Mal Respir, 2015. **32**(4): p. 435-46.
123. Lomas, D.A., *Loop-sheet polymerization: the mechanism of alpha1-antitrypsin deficiency*. Respir Med, 2000. **94 Suppl C**: p. S3-6.
124. Lomas, D.A., et al., *Polymerisation underlies alpha1-antitrypsin deficiency, dementia and other serpinopathies*. Front Biosci, 2004. **9**: p. 2873-91.
125. Berthelie, V., et al., *Discovery of an inhibitor of Z-alpha1 antitrypsin polymerization*. PLoS One, 2015. **10**(5): p. e0126256.
126. Saez, E., et al., *Inducible gene expression in mammalian cells and transgenic mice*. Curr Opin Biotechnol, 1997. **8**(5): p. 608-16.
127. Hillen, W. and C. Berens, *Mechanisms underlying expression of Tn10 encoded tetracycline resistance*. Annu Rev Microbiol, 1994. **48**: p. 345-69.
128. Hillen, W., et al., *Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions*. J Mol Biol, 1983. **169**(3): p. 707-21.
129. Alam, S., et al., *Preventing and reversing the cellular consequences of Z alpha-1 antitrypsin accumulation by targeting s4A*. J Hepatol, 2012. **57**(1): p. 116-24.
130. Belorgey, D., et al., *Characterisation of serpin polymers in vitro and in vivo*. Methods, 2011. **53**(3): p. 255-66.

Appendices

Appendix A

Discovery of a Potent Inhibitor of Z-Alpha1 Antitrypsin Polymerization

Valerie Berthelie[†], ^{*§†‡} Jason Harris, ^{§‡††} Kasey Estenson, ^{†‡} and Jerome Baudry^{*‡ ‡††}

[†] Department of Medicine, University of Tennessee Health Science Center – Graduate School of Medicine, Knoxville, TN

[‡] Department of Biochemistry and Cellular and Molecular Biology and [‡] Genome Science and Technology Graduate School, University of Tennessee, Knoxville, TN

^{††} UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, TN

Abstract

Polymerization of the Z variant alpha-1-antitrypsin (Z- α 1AT) results in the most common and severe form of form of α 1AT deficiency (α 1ATD), a debilitating genetic disorder whose clinical manifestations range from asymptomatic to fatal liver or lung disease. As the altered conformation of Z- α 1AT and its attendant aggregation are responsible for pathogenesis, the polymerization process *per se* has become a major target for the development of therapeutics. Based on the ability of Z- α 1AT to aggregate by recruiting on its s4A cavity the reactive center loop (RCL) of another Z- α 1AT, we developed a high-throughput screening assay that uses a modified 6-mer peptide mimicking the RCL to screen for inhibitors of Z- α 1AT polymer growth. A subset of a commercially available library of small compounds with MWs ranging from 300 to 700 Da was used to test the assay's capabilities, and the inhibitor S-(4-nitrobenzyl)-6-thioguanosine was found. To validate S-(4-nitrobenzyl)-6-thioguanosine, an *in silico* strategy was pursued and the intermediate α 1AT M* state modeled to allow molecular docking simulations, which explore various potential binding sites. Docking results predict that S-(4-nitrobenzyl)-6-thioguanosine can bind at the s4A cavity or at the edge of β -sheet A. The former binding site would block RCL insertion whereas the latter site would prevent β -sheet A from expanding between s3A/s5A, and thus indirectly impede RCL binding. Altogether, our investigations have revealed a novel compound that specifically inhibits the formation of Z- α 1AT polymers, as well as *in vitro* and *in silico* strategies for identifying small molecules for treatment of α 1ATD.

A.1 Introduction

Human α 1-antitrypsin (α 1AT) is the most abundant member of the serine protease inhibitor (SERPIN) family. It is a soluble 52-KDa glycoprotein synthesized primarily by hepatocytes and delivered to the lungs to accomplish its critical function: inactivation of the proteinase neutrophil elastase (NE), a mediator of alveolar destruction.(Brantly, Nukiwa, & Crystal, 1988) Defective folding, trafficking and secretion into the plasma of α 1AT are responsible for α 1AT deficiency (α 1ATD).(Bibek Gooptu, Dickens, & Lomas, 2014a; Perlmutter, 2011)

The structural flexibility of α 1AT is important for it to perform its anti-protease function and ensure lung integrity. With a core domain composed of 3 β -sheets A, B and C, and 9 α -helices, α 1AT features an exposed and flexible reactive center loop (RCL) that serves as bait for NE. Upon binding to the proteinase, a dramatic conformational change occurs as RCL is cleaved and translocates into β -sheet A to form the new central and fourth strand, s4A. The translocation event carries along NE from one side to the other of α 1AT, causing its inactivation by forming an irreversible, higher molecular weight suicide complex.(Huntington, Read, & Carrell, 2000)·(Whisstock & Bottomley, 2006) A reduction or lack of this inhibition through loop-sheet insertion and proteolytic cleavage is thought to be the underlying mechanism responsible for α 1ATD.(Dafforn, Mahadeva, Elliott, Sivasothy, & Lomas, 1999)·(Lomas, 2000)

Over 100 genetic variants of α 1AT have been identified with the Z-type being responsible for the most common and severe form of the disease in homozygous patients(de Serres & Blanco, 2012). The punctual mutation E342K in Z- α 1AT renders the anti-protease prone to aggregation and unable to be secreted into the blood stream resulting in a 90% decrease in NE inhibition within the lungs. Accumulation of polymers of Z- α 1AT in the endoplasmic reticulum (ER) of hepatocytes leads to proteotoxic stress and associated liver diseases.(An, Blomenkamp, Lindblad, & Teckman, 2005; Eriksson, Carlson, & Velez, 1986; Hussain, Mieli-Vergani, & Mowat, 1991) In addition to sequestration of polymers in the ER of hepatocytes, the E342K mutation has two additional disease-causing effects. It causes Z- α 1AT to be 5-fold less effective in accomplishing its inhibitory function(Ogushi, Fells, Hubbard, Straus, & Crystal, 1987)·(Lomas, Evans, Stone, Chang, & Carrell, 1993) and it promotes the spontaneous formation of Z- α 1AT polymers within the lungs, thereby further reducing the already depleted levels of α 1AT that are available for alveola protection.(Elliott, Bilton, & Lomas, 1998) Moreover, the conversion of Z- α 1AT from a monomer to a polymer renders it a chemoattractant for human neutrophils.(Mahadeva et al., 2005)·(Parmar et al., 2002) To summarize, emphysema associated with Z- α 1ATD results from a combination of (1) loss of function of the anti-protease, which leads to the absence of circulating α 1AT, decrease of its inhibitory activity, and intra-alveolar

polymerization, and (2) gain of toxic function from the neutrophil chemotactic properties of intra-alveolar polymers.

Preventing formation and accumulation of Z- α 1AT polymers could be crucial to treat α 1ATD.(Lomas, Perlmutter, & Uversky, 2010) For this reason, the mechanisms by which Z- α 1AT form polymers have been under intense investigation. As the substitution of the glutamic acid residue at position 342 by a lysine provokes a perturbation in the native structure by opening the β -sheet A, biochemical evidence reveals the formation of an unstable and polymerogenic intermediate M^* with its own RCL partially inserted.(Mahadeva, Dafforn, Carrell, & Lomas, 2002) The opening of the s4A cavity allows the creation of a sequential β -strand linkage between the RCL of one serpin and β -sheet A of another, leading to the formation of a dimer and then polymers.(Dafforn et al., 1999)(Ekeowa et al., 2010; B Gooptu et al., 2000; Purkayastha et al., 2005) Additional models for Z- α 1AT polymerization have also recently been proposed based on the crystal structures of a dimer of the serpin antithrombin(M Yamasaki, Li, Johnson, & Huntington, 2008) and a trimer of a disulfide mutant of α 1AT,(M Yamasaki, Sendall, Pearce, Whisstock, & Huntington, 2011) suggesting that assembly pathways of Z- α 1AT could be diverse and therefore arising from structurally and/or dynamically distinct polymerogenic intermediates.

Various strategies have been pursued in order to prevent or even attenuate Z- α 1AT polymerization such as increasing the mutant protein secretion with the use of osmolytes(Burrows, Willis, & Perlmutter, 2000; Devlin, Parfrey, Tew, Lomas, & Bottomley, 2001; Teckman, 2004) or by blocking Z- α 1AT polymerization by either filling the s4A cavity with peptides(Mahadeva et al., 2002) or crowding another hydrophobic pocket of Z- α 1AT with small compounds screened virtually.(Mallya et al., 2007) While extensive progress has been made, none of these strategies has been entirely successful so far. To achieve this goal, we developed a set of novel and integrated *in vitro* and *in silico* screenings methods; the *in vitro*, being a high-throughput screening assay using a modified small peptide previously reported as a s4A cavity filler,(Mahadeva et al., 2002) and the *in silico* being a virtual docking model able to predict and help rationalize the binding of compounds to α 1AT, including in the S4A cavity. Here, we present how using these two combined methods we were able to identify, rationalize and validate S-(4-nitrobenzyl)-6-thioguanosine as a specific inhibitor of Z- α 1AT polymerization.

A.2 Experimental procedures

A.2.1. General materials and methods

The peptide acetyl-FLEAIGGG-Q-GKKG containing the 6-mer sequence of the RCL was synthesized by custom solid-phase from the Keck Biotechnology Center at Yale University (<http://info.med.yale.edu/wmkeck/>). A biotinylated version of the peptide

(bPEG-peptide) was obtained by appending a biotinyl-polyethylene glycol spacer on the α -amide group of the glutaminyl residue. The presence of the Lys residues confer a positive net charge to the peptide at neutral pH, enhancing its general solubility. The wild type and Z- α 1AT proteins, prepared according to published protocol,(Parfrey et al., 2003) were graciously provided at a concentration of 1 mg/ml by Professor Lomas, Cambridge Institute for Medical Research, University of Cambridge, UK, and stored at 4 °C. The rabbit anti-human α 1AT antibody (serum fractions IgG) was purchased from Abcam, Cambridge, MA.

The test group RK-001 of the LOPAC library (Library of Pharmacologically Active Compounds, Sigma-RBI, Natick, MA) containing 80 lyophilized chemical compounds was prepared in a 96-well plate format. All compounds were resuspended in 2 ml DMSO at a concentration of approximately 4 mM, based on an estimated MW average of 500 g/moles, and stored at 4 °C.

A.2.2. Preparation of the bPEG-peptide

The synthesized bPEG-peptide was first solubilized in 50% formic acid at a concentration of ~1mg/ml, injected onto a Zorbax C3 Column and purified by RP-HPLC at a rate of 4ml/min. The resulting purified peptide was then lyophilized, resuspended into H₂O and stored at -20 °C. After amino acid analysis of the peptide (Commonwealth Biotechnologies Inc., Richmond, VA), various amounts were injected onto RP-HPLC in order to establish a standard curve, allowing us to determine the exact concentration of each new batch of purified peptide that we prepared.

A.2.3. Preparation of working compound plates

LOPAC compounds were transferred from their original 96-well plates to new 96-well working plates with low evaporation lid (BD Falcon plates non treated, Becton Dickinson Labware, San Jose, CA) in respect to their initial location, and adjusted to a concentration of 1 mM in PBS 1X containing 50% DMSO. First and last columns were filled only with PBS 1X/DMSO (50/50). Working plates were sealed with an adhesive overlay, covered and stored at 4° C until further utilization.

A.2.4. Set up of the microplate screening assay

The assay is based on the principle of a competitive ELISA.(Makarananda & Neal, 1992) Wells were coated by passive adsorption with a 1/1000 solution in PBS 1X of capture α 1AT Ab. The screening microplate was sealed with an adhesive overlay and incubated for 2 h at 37 °C. The wells were then washed three times with extension buffer (PBS 1X and 0.01% Tween 20), blocked for 1 h at 37 °C with 0.3 % gelatin and washed again. Screening results described in this paper were carried out with screening microplates freshly made. However, screening microplates can be filled with PBS 1X, hermetically sealed and stored at 4 °C for one week prior to use.

A.2.5. Z- α 1AT polymerization inhibition assay

In parallel with the preparation of the microplate screening assay, polymerization reactions were carried out in 96-well plates with or without the LOPAC small compounds. A 100 molar excess of bPEG-peptide was used with 4 μ g/well of Z- α 1AT.

Each polymerization reaction plate was organized as follows: the first column contained only bPEG-peptide (background control); the second to eleven columns contained Z- α 1AT, compounds and bPEG-peptide; and the last column contained Z- α 1AT, bPEG-peptide and no compound (reaction control). Assay wells were set up by adding to each well 20 μ l of protein and 20 μ l of compound from a working plate. After 3 min, 160 μ l of bPEG-peptide at 48 μ M was added. The plate was then sealed, shaken on a microplate shaker gently for 5 s to ensure homogeneity of the different reactants, and placed at 37 $^{\circ}$ C for 16 h. All wells contained 5% DMSO.

At the end of the 16 h incubation time, 100 μ l from each well were transferred into the corresponding well of the microplate screening assay. One hour later, the screening plate was washed three times and then incubated in the dark for 1 h at room temperature with 100 μ l/well of 1 ng/ μ l of europium streptavidin (Perkin Elmer, Boston, MA) in 0.5% BSA-extension buffer. Three final washes in extension buffer were carried out and the europium was released from streptavidin by the addition of 100 μ l of enhancement solution (Perkin Elmer). After 5 min, europium fluorescence was measured by time-resolved fluorometry in a Victor 2 counter (Perkin Elmer) and then converted to fmoles of bPEG-peptides recruited into Z- α 1AT. Assays were conducted in triplicate by processing three identical plates in parallel.

A.2.6. Determining IC₅₀ values of inhibitors

As described above, 4 μ g/well of Z- α 1AT were incubated for 3 min with various concentrations of a compound identified as an inhibitor, the highest concentration starting at 400 μ M. The concentrations of the compounds were revised according to the true MW of the molecule. Following the 3 min incubation, 160 μ l of 48 μ M bPEG-peptide were added and the rest of the protocol was applied as described above.

A.2.7. Z- α 1AT polymerization sedimentation assay

A solution of 0.1 mg/ml of Z- α 1AT in PBS 1X was incubated at 37 $^{\circ}$ C with or without 100 μ M of S-(4-nitrobenzyl)-6-thioguanosine. Progress of the polymerization reaction was followed by quantitative RP-HPLC on centrifugation SUPERNATANTS (20 min, 20,000 $\times$ G) OF ALIQUOTES. Quantitative determination of the Z- α 1AT monomer disappearance was calculated according to a pre-established standard curve.

A.2.8. Preparing protein structures for homology modeling and docking simulations

The crystal structures of the mutant Z- α 1AT (PDB code: 3T1P) and the wild type M- α 1AT (PDB codes: 3CWM and 1QLP) were obtained from the RCSB Protein Database.(Elliott, Pei, Dafforn, & Lomas, 2000; Pearce et al., 2008; M Yamasaki et al., 2011) Initial preparation of the receptor structures was carried out with the program MOE(“Chemical Computing Group - Citing MOE,” 2012) (Molecular Operating Environment). Co-crystallized water molecules were deleted from both structures. For the polymerized mutant (3T1P), only the first monomer was retained and the s4A binding cavity was created between the s3A/s5A by deleting residues 345-356, which correspond to the inserted residues of the RCL. The protonation state of atoms was assigned using Protonate 3D(Labute, 2009) utility in MOE at pH 7, 300 K and 0.1 M salt concentration. Solvent effects were implicitly included by using a distance-dependent dielectric. Partial charges were assigned to receptor atoms using MMFF94s(Halgren, T. A. (1999) OBFoercedFieldMMFF94. J. Comput. Chem. 20, n.d.) force field parameters as implemented in MOE.

A.2.9. Homology modeling procedure

Modeling of the M* intermediate state as described below was performed using the Homology Model facility in MOE. The wild type crystal structure of α 1AT (PDB code 1QLP) was used as a template for modeling i) the position of the RCL when not inserted into β -sheet A and ii) the position of the c-terminal loop within β -sheet B when it is not participating in a domain swap. The polymerized Z-mutant structure (PDB code: 3T1P) was used to model the expanded position of β -sheet A but omitting s4A to leave a cavity between s3A/s5A where the RCL would otherwise be found. Fragments from each template structure were joined at transition points selected by superimposing the structures and choosing those residues between fragments with near overlapping atom positions. A total of 25 homology models were generated with unique carbon backbone positions, and for each of those 25 models, 5 additional models (*i.e.* a total of 125 models) were created with alternate side chain positions. These initial models were energy minimized to a gradient of 0.1 kcal/mol·Å. A final M* model (Model 126) was created using the Generalized Born / Volume Integral (GB/VI) energy scoring method(Labute, 2008) to select the best initially packed structure and then further energy minimizing it to a gradient of 0.01 kcal/mol·Å.

A.2.10. Docking simulation procedure

Three-dimensional structures of the 80 *in vitro* tested chemicals, including S-(4-nitrobenzyl)-6-thioguanosine, were obtained in SDF format from the electronic LOPAC library, test group RK-001. Partial charges were added to each ligand atom using MMFF94s forcefield parameters and the structures were energy minimized to a gradient of 0.1 kcal/mol·Å. The energy-minimized ligands were docked into three structural

variations of α 1AT: M* intermediate (Homology Model 126), mutant Z- α 1AT (PDB: 3T1P) and wild type M- α 1AT (PDB: 3CWM) using the docking function built into MOE. Binding sites were identified using the MOE Site Finder facility. Separate docking simulations of the 80 compounds were carried out for each receptor and at each potential binding site. Initial placement of ligand atoms was done with the Triangle Placement method (seeds in 3 atoms at time). The predicted free energy of binding for each initially docked ligand pose was calculated using the London dG("Chemical Computing Group - Citing MOE," 2012) scoring method from within MOE. The top five scoring poses were further energy minimized using the MMFF94s force field, allowing ligand atoms and protein side chains within 6 Å of each docked ligand to be treated as flexible. A tethering weight of 10 kcal/mol/Å² was applied to partially restrain flexible atoms around their original location. A final docking score for each energy-minimized pose was calculated using the Affinity dG("Chemical Computing Group - Citing MOE," 2012) scoring method.

A.3 Experimental results

A.3.1 Principal characteristics of the Z- α 1AT polymerization inhibitor screening assay

Previous studies have shown that a 6-mer peptide whose amino acid sequence contains the RCL sequence FLEAIG can specifically bind the Z-mutant at its opened s4A pocket, but not the wild type. (Mahadeva et al., 2002) The Z- α 1AT high-throughput microplate screening assay is based on this concept. Thus, we designed a similar peptide and added a biotin-polyethylene glycol (bPEG) tag at the C_{term} of the reactive loop sequence as well as some hydrophilic amino acids to increase peptide solubility. The insertion of a PEG-based spacer prevents possible steric hindrance between the peptide and the biotin molecule, resulting in better avidin binding and therefore, a more accurate measurement of the biological activity.

To assess the ability of small molecules to inhibit the recruitment of the bPEG-peptide into Z- α 1AT, microtiter plate wells containing attached Z- α 1AT are subjected for 3 min to library compound before addition of bPEG-peptide. The amount of bPEG-peptides incorporated into the mutant protein is then determined by a europium-streptavidin treatment and time-resolved fluorescence measurements. The inhibition effect of a compound is calculated as a percentage with respect to a reaction control – i.e. Z- α 1AT that has only been exposed to the biotinylated peptide and not to a compound. Any compound showing an inhibitory effect of at least 50% is considered as a hit. Regarding its ability to bind Z- α 1AT, we found the bPEG-peptide association kinetics to be in favor of the mutant proteinase with an initial association rate of 0.22 ± 0.08 fmoles $\cdot$ h⁻¹ vs. 0.042 ± 0.1 fmoles $\cdot$ h⁻¹ for the wild type, Figure A.1A. We also found that an incubation period of 16 hrs for the peptide with Z- α 1AT is an adequate screening end-point for the screening assay as this time period is associated with a high signal-to-noise ratio. In addition, the presence of 5% DMSO in the wells does not affect the bPEG-peptide binding kinetics,

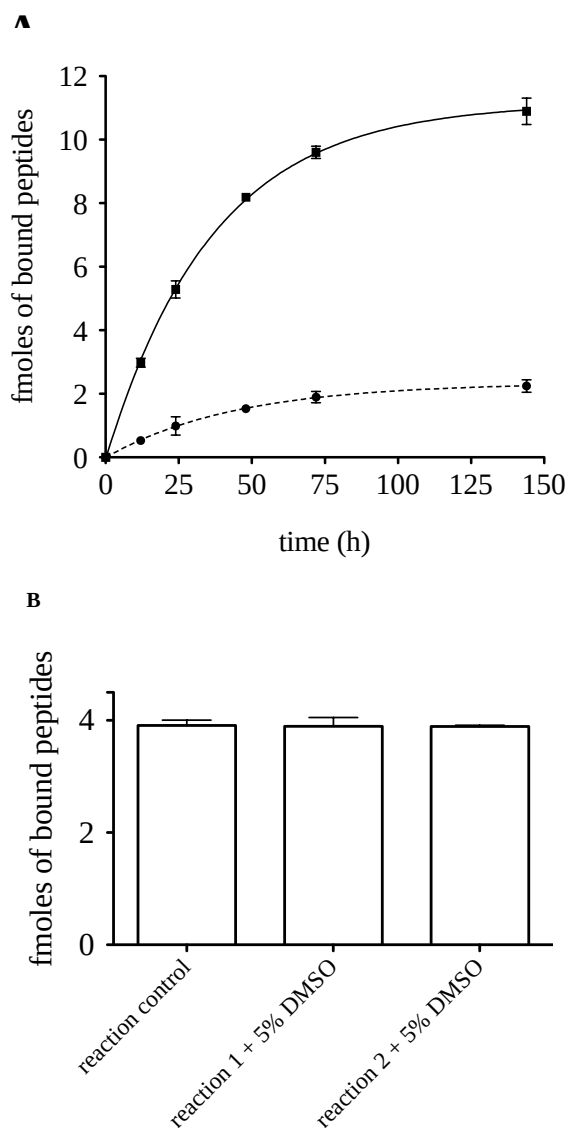


Figure A.1. Kinetic diagram of bPEG-peptide binding to α 1AT. (A) Four micrograms per well of attached (■) Z- α 1AT or (●) M- α 1AT were incubated for various times in presence of 38.4 μ M bPEG-peptide. (B) Z- α 1AT was incubated in presence of 5% DMSO and bPEG-peptide for 16 h. Errors bars reflect the standard deviation of three replicates.

Figure A.1B. Since compound libraries are generally stored in DMSO, this feature makes the assay well suited for a high-throughput screening assay. Finally, this screening assay exhibits very good reproducibility as reflected by the error bars shown in Figure A.2. It requires only small amount of protein and low concentrations of bPEG-peptide, which make it both economical and physiological.

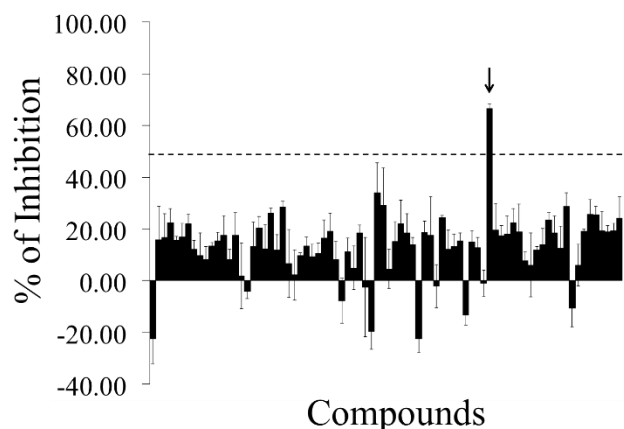


Figure A.2. Pattern of inhibition resulting from the screening of 80 unknown LOPAC compounds. A 96-well plate was coated with 4 $\mu\text{g}/\text{well}$ of Z- α 1AT and incubated for 16 h with 100 μM of various compounds and 38.4 μM of bPEG-peptide. The black arrow indicates the compound that corresponds to S-(4-nitrobenzyl)-6-thioguanosine and gives an inhibition effect of $67 \pm 2\%$ and. The error bars are the standard deviation of three individual experiments.

A.3.2 S-(4-nitrobenzyl)-6-thioguanosine identified as inhibitor of Z- α 1AT polymerization

The test group RK-001 (80 compounds) of the small commercially available LOPAC library containing drug-like molecules was used to test the performance of the screening assay. Figure A.2 shows a typical screening result. As indicated in the figure, only one compound of the tested compound plate appears as a hit, exhibiting a $67 \pm 2\%$ inhibition activity at 100 μM . This compound is S-(4-nitrobenzyl)-6-thioguanosine. To confirm its ability to inactivate Z- α 1AT polymerization, dose-responses curves were carried out and an IC_{50} of $73 \pm 0.12\ \mu\text{M}$ calculated, Figure A.3A. The IC_{50} value obtained is in the micromolar range and matches well with the screening results.

To define a better pharmacophore, and therefore to identify any additional structural element required for inhibiting Z- α 1AT polymerization, we then compared our compound to the entire database that regroups all of the LOPAC molecules. Surprisingly, we found another compound that possesses a very similar structure, differing by a single amino group, but that did not show any inhibitory effect, neither during the original screening nor in the validation assay, Figures A.3A and A.3B.

A.3.3 Validation of the action of S-(4-nitrobenzyl)-6-thioguanosine

A polymerization reaction was set up in presence or absence of 100 μ M of S-(4-nitrobenzyl)-6-thioguanosine and the disappearance of the Z- α 1AT monomer monitored by RP-HPLC – a diminution in monomer concentration indicates that the protein has been recruited into polymers. We found that Z- α 1AT has its polymerization rate decreased 33 times in presence of the compound and that its effect is long lasting, Figure A.4.

An additional approach to validate the action of S-(4-nitrobenzyl)-6-thioguanosine was to carry out an *in silico* molecular modeling and simulation strategy with the intent to gain mechanistic insights into how this small molecule may interact with the protein structure and prevent polymerization. Therefore, virtual docking and homology modeling were used to explore hypothetical binding sites and their putative molecular interactions.

A.3.4 Structural modeling of α 1AT and the M* intermediate

In order to investigate all of the potential binding sites of S-(4-nitrobenzyl)-6-thioguanosine on α 1AT, including the ones located in the s4A cavity at the RCL insertion site, we used two PDB crystal structures, 1QLP and 3T1P, which respectively correspond to the wild type M- α 1AT and polymerized Z- α 1AT states. However, as the s4A cavity does not exist in any crystal structure of α 1AT, a theoretical model comparable to M* had to be created. The M* intermediate state is described to have the following three structural features: i) an expanded β -sheet A with a s4A cavity between s3A/s5A, ii) an RCL at the precipice of inserting between s3A/s5A, and iii) the C_{term} loop inserted within β -sheet B and not participating in a domain swap with another protein. These important features of the M* model are represented in the homology model built from the two available crystal structures of α 1AT, Figure A.5.

To build the M* state homology model, a total of five protein fragments of the two crystal structures, 1QLP and 3T1P, were merged. Figure A.6 shows that fragment 1 consists of residues 1-105 (1QLP) which model the right side of β -sheet B, with respect to beta strands adjacent to the right side of the C_{term} loop. Fragment 3 consists of residues 205-291 which constitute the left side of β -sheet B, with respect to beta strands adjacent to the left side of the C_{term} loop. Together these two fragments model the position of β -sheet B so that the RCL residues from fragment 5 (residues 345-394 from 1QLP) can be placed

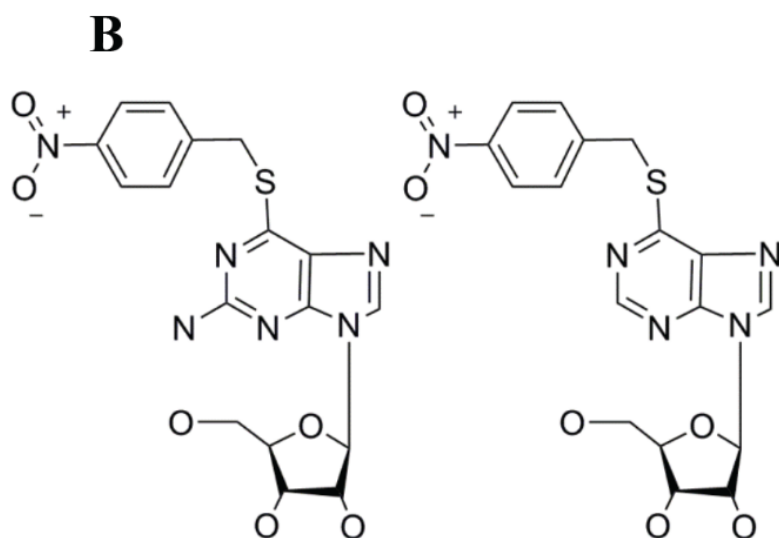
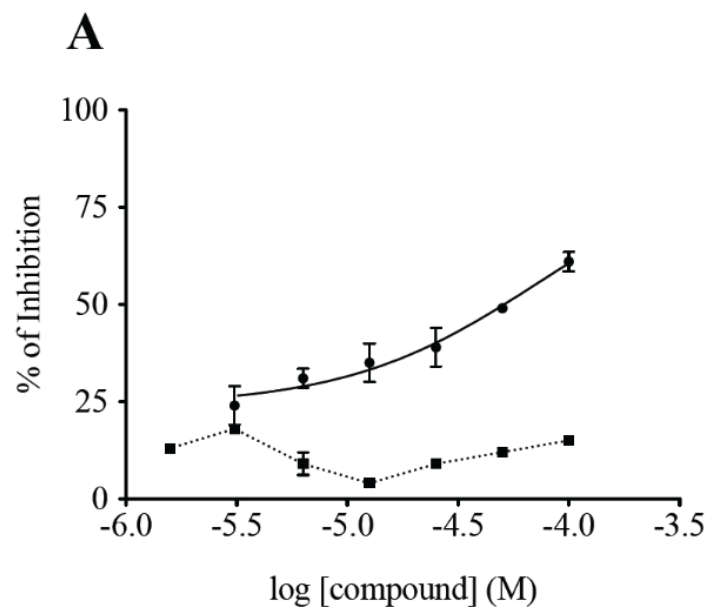


Figure A.3. S-(4-nitrobenzyl)-6-thioguanosine inhibits bPEG-peptide binding to Z- α 1AT. (A) Dose-response curves were assayed for various concentrations of (●) S-(4-nitrobenzyl)-6-thioguanosine and (■) its homologue S-(4-nitrobenzyl)-6-thioinosine. (B) Chemical structures of (left) S-(4-nitrobenzyl)-6-thioguanosine and (right) S-(4-nitrobenzyl)-6-thioinosine. The errors bars are the standard deviation of an experiment conducted in triplicate.

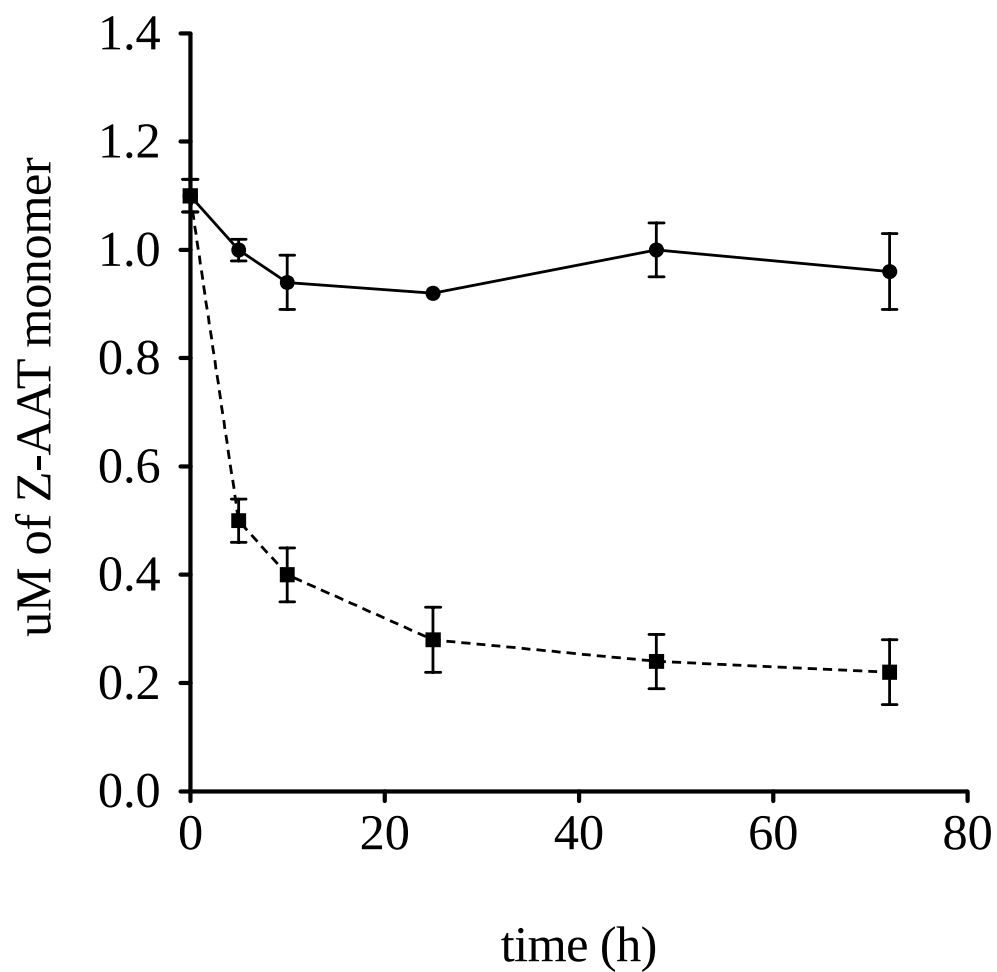


Figure A.4. Effect of S-(4-nitrobenzyl)-6-thioguanosine on Z- α 1AT polymerization. The protein was incubated with (●) or without (■) 100 μM of S-(4-nitrobenzyl)-6-thioguanosine for various time at 37 °C. The error bars are the standard deviation of three separate experiments.

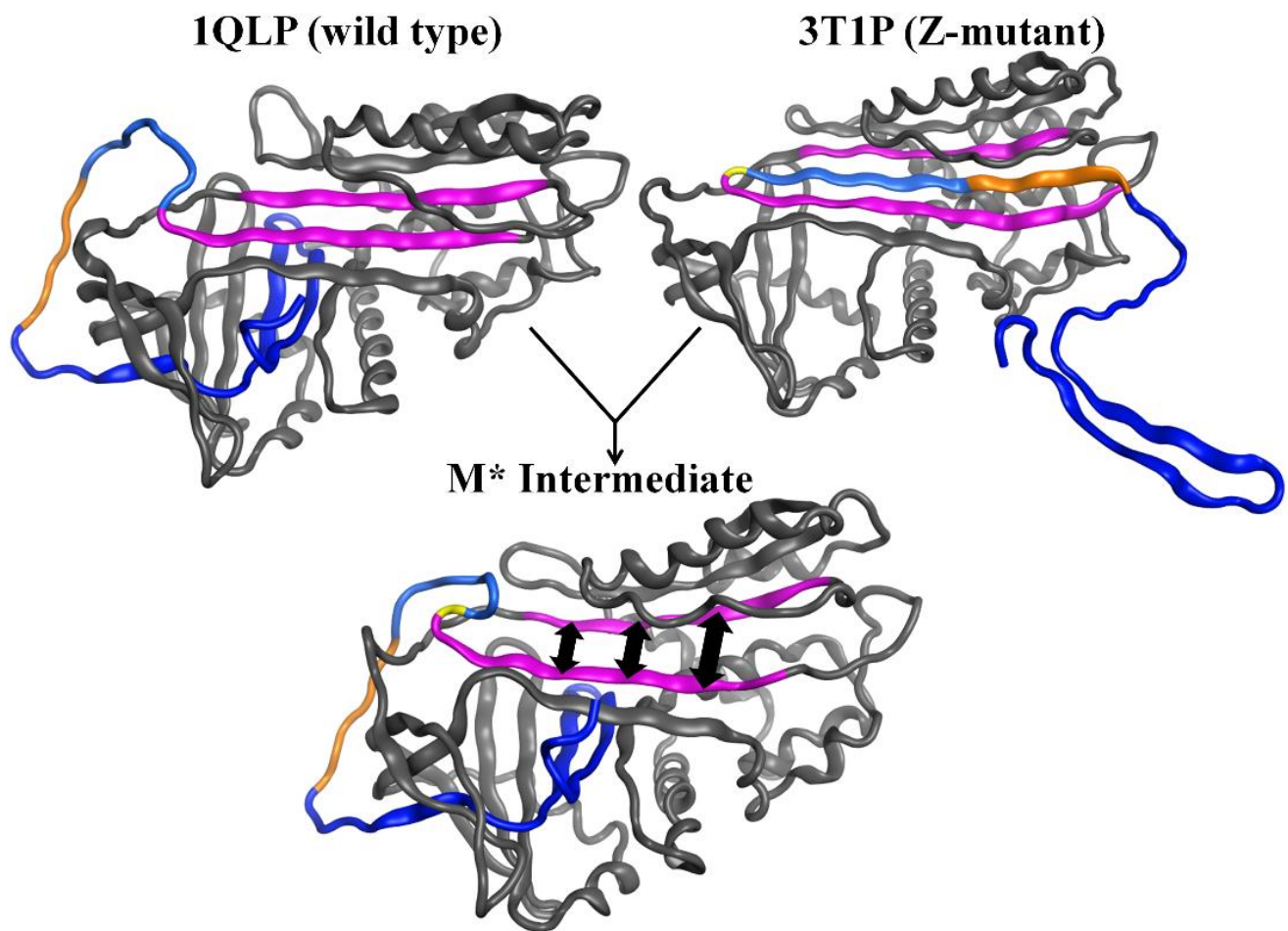


Figure A.5. The three models of α_1 AT protein. (Top left) Structure of wild type (PDB: 1QLP) with the RCL not inserted and β -sheet A not expanded. (Top right) Structure of Z-mutant (PDB: 3T1P) with the RCL inserted and β -sheet A expanded. (Bottom middle) Intermediate M* model with an expanded β -sheet A (retained from structure 3T1P), RCL not inserted into the RCL cavity (retained from structure 1QLP), and C_{term} loop inserted into β -sheet B (retained from structure 1QLP). (Purple) Strands 3 and 5 from β -sheet A. (Dark blue) C_{term} loop within β -sheet B. (Light blue) RCL. (Orange) Residues of the RCL corresponding to the analogous 6-mer peptide. (Black arrows) s4A cavity.

on the outside of the s4A pocket along with the C_{term} loop buried within β -sheet B. The position of strands s1A, s2A, and s3A in β -sheet A are modeled from fragment 2 (residues s4A between s3A/s5A, which would otherwise be the site of RCL insertion. This opened conformation of α 1AT represents one of the possible structures of the unstable M* intermediate state for which experimental methods such as crystallography cannot reproduce.

A.3.5 Analysis of α 1AT structures and their potential binding sites

All of the 80 *in vitro* screened compounds, including S-(4-nitrobenzyl)-6-thioguanosine, were docked into every potential binding site to assess if the computational result is comparable to the *in vitro* screening. A binding site able to dock S-(4-nitrobenzyl)-6-thioguanosine with a lower binding energy than the 79 other compounds would be a promising site for further experimental investigations.

Six putative binding sites were predicted among the three available protein models: M- α 1AT, Z- α 1AT and intermediate M*, Figure A.7. SITE1 and SITE5 were both exclusively available in the M* model and are located in the RCL insertion site. SITE2 was found in all three models. It is also where the compound citrate, previously reported to lower polymerization rates (Pearce et al., 2008) has been observed to bind in the 3CWM wild type structure. Also found in all three models are: SITE3, a large cavity adjacent to SITE2; SITE4 situated near the C_{term} edge of β -sheet A; and SITE6 located near the N_{term} edge of β -sheet A. SITE6 is partially occluded in the M* model due to the expansion of β -sheet A.

A.3.6 S-(4-nitrobenzyl)-6-thioguanosine binds at the RCL insertion site or on the edge of β -sheet A

Docking of all 80 small molecules was performed with each model and at each putative binding site in order to compare how strongly S-(4-nitrobenzyl)-6-thioguanosine binds relative to the 79 other experimentally tested compounds. These results are summarized in Table A.1 and present two possible binding sites where S-(4-nitrobenzyl)-6-thioguanosine can favorably bind to block RCL insertion. Results from docking at SITE5, the RCL insertion site, show S-(4-nitrobenzyl)-6-thioguanosine ranking first among the other 79 ligands which may suggest a mechanism where the RCL is directly blocked at the RCL insertion site. Interestingly, S-(4-nitrobenzyl)-6-thioguanosine is also found to rank first in the wild type model when docked at SITE6. Figure A.8 compares the location of SITE6 in both the M- and Z- α 1AT structures which illustrates how binding of S-(4-nitrobenzyl)-6-thioguanosine at SITE6 may prevent the expansion of β -sheet A and possibly prevent RCL insertion. Lesser sites of interest, which only rank S-(4-nitrobenzyl)-6-thioguanosine in the top 10% of ligands, are SITES 1 and 2. SITE1 is also part of the RCL insertion site. SITE2 has been previously reported as the binding site for citrate

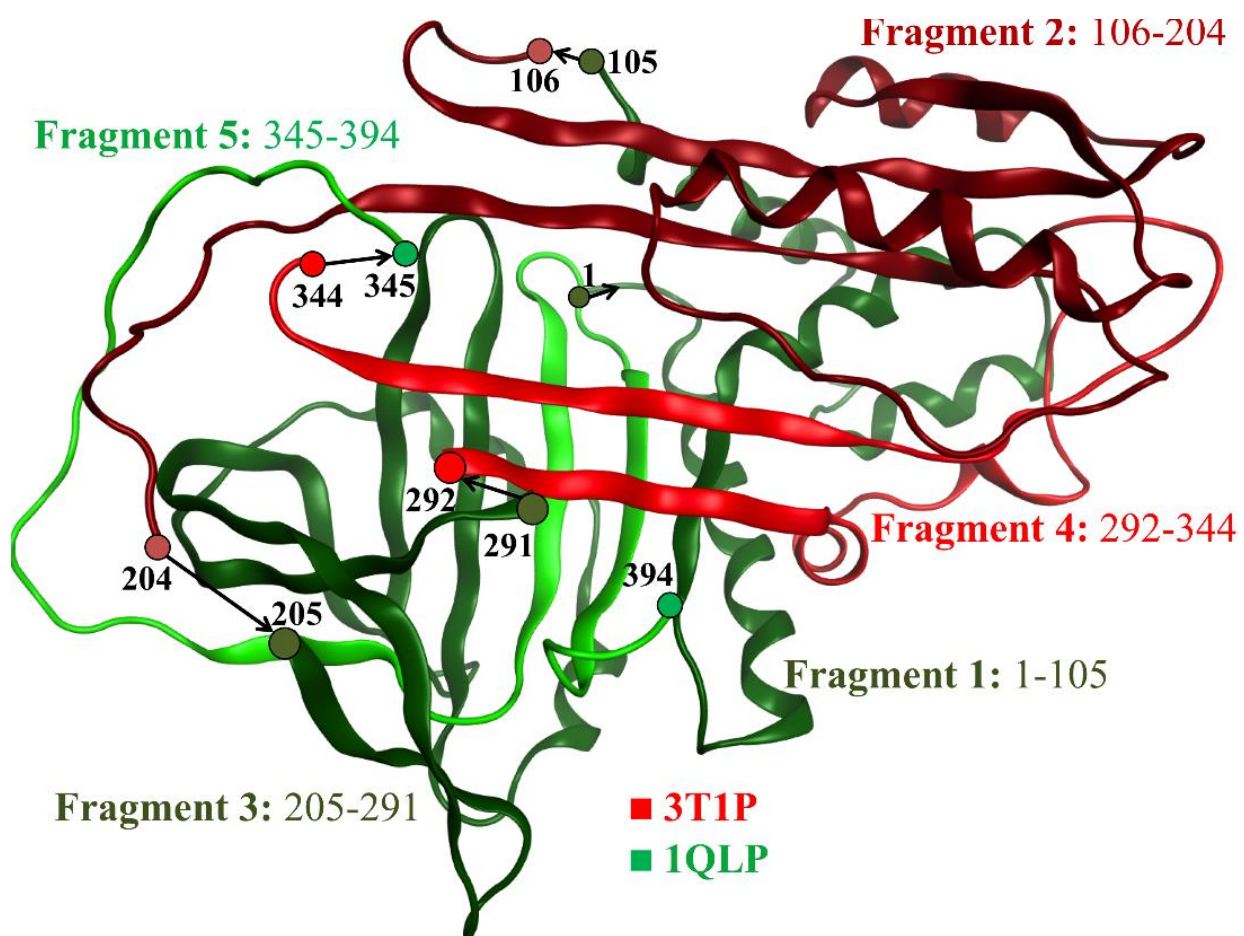


Figure A.6. The fragments of structures 1QLP (green) and 3T1P (red) used to homology model the M* intermediate state of α 1AT. β -sheet A is the red beta sheet across the top half of the model and β -sheet B is the green beta sheet across the bottom of the model. Residue numbers at the start and end of each fragment transition are labeled with an arrow in the N_{term} to C_{term} direction. Shades of green and red distinguish discontinuous fragments from the same initial crystal structure (light/dark green for 1QLP fragments and light/dark red for 3T1P fragments).

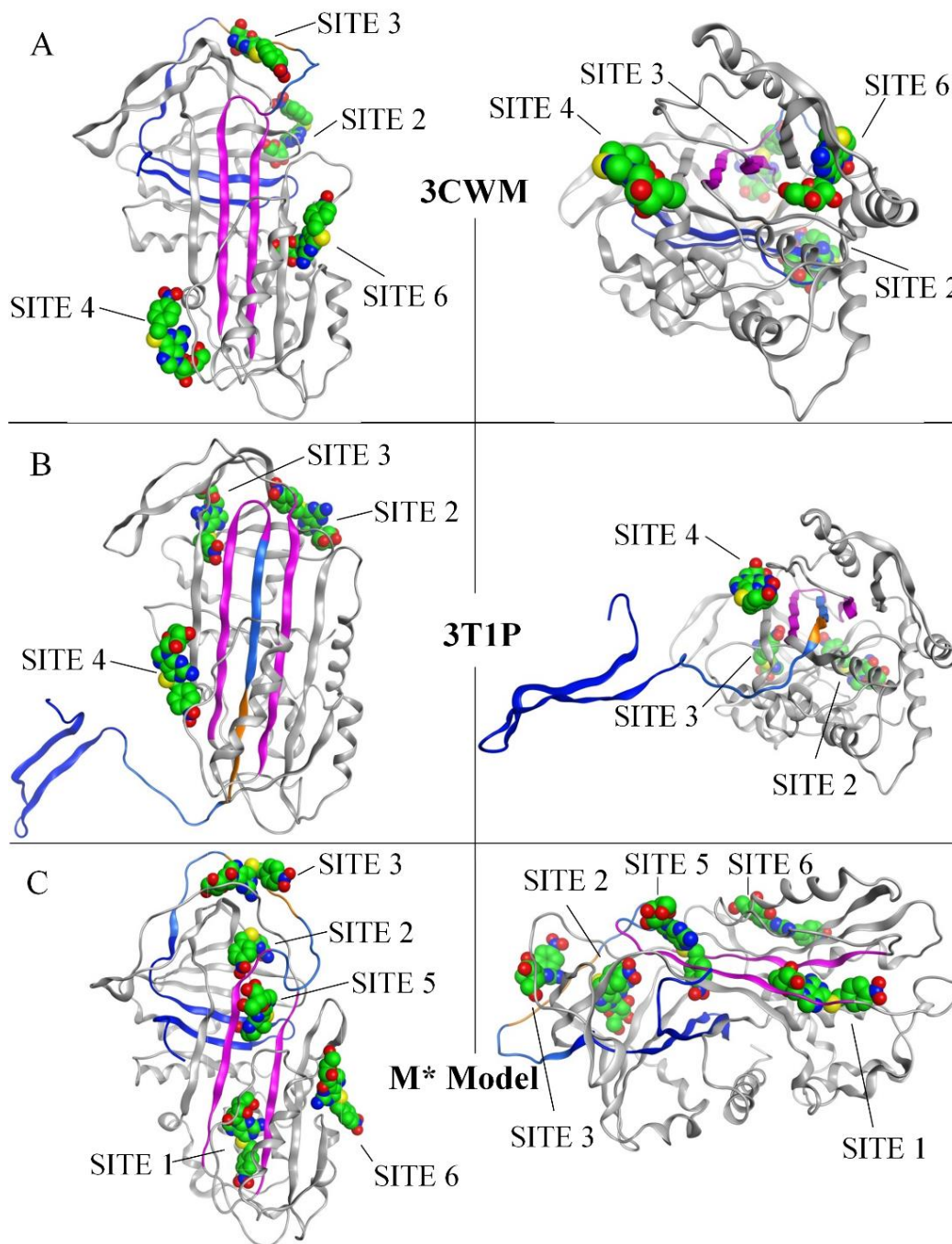


Figure A.7. Binding Sites for S-(4-nitrobenzyl)-6-thioguanosine. Two protein ribbon models are shown for each structure: (A) 3CWM, (B) 3T1P and (C) M* Model. The left model and right representations in each panel are rotated 90° with respect to one another. The best binding poses for S-(4-nitrobenzyl)-6-thioguanosine at each available binding site are shown with space filling atoms with the carbon atoms colored green. (Purple) Strands 3 and 5 from β -sheet A. (Dark blue) C_{term} loop within β -sheet B. (Light blue) RCL. (Orange) Residues of the RCL corresponding to the analogous 6-mer peptide.

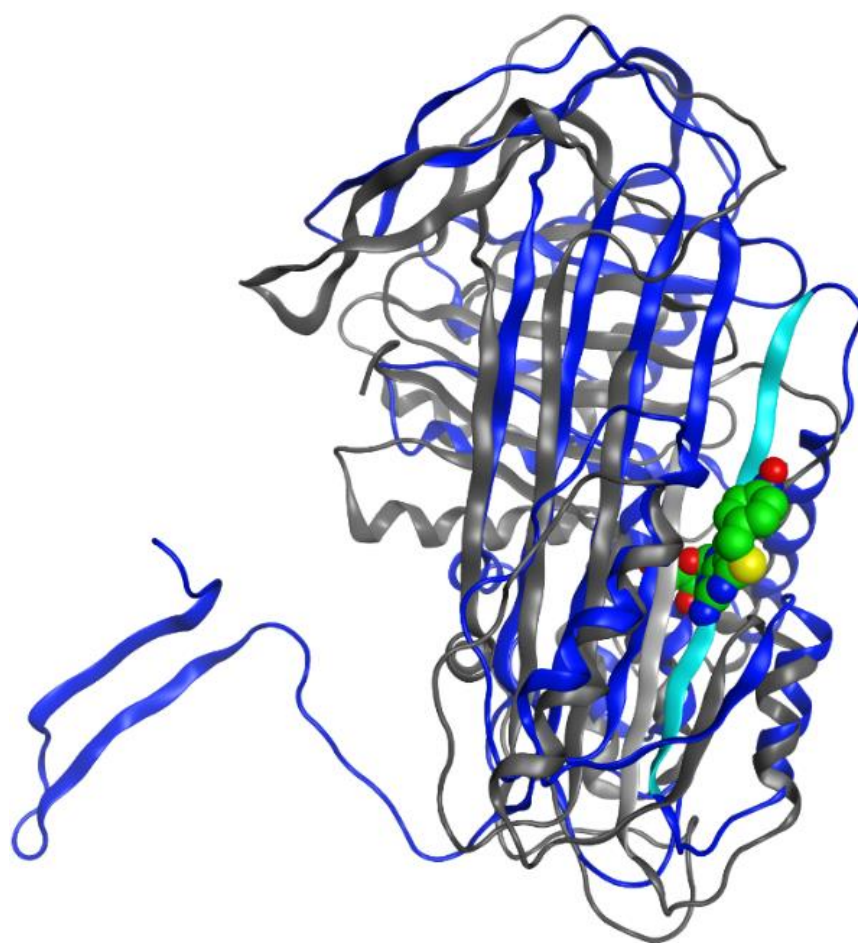


Figure A.8. Two crystal structures for α 1AT are superimposed and represented in ribbon diagram. S-(4-nitrobenzyl)-6-thioguanosine is represented with space filling atoms and positioned at SITE6 for the M- α 1AT structure (1QLP). (Dark blue) Z- α 1AT structure (3T1P) with an expanded β -sheet A. (Dark grey) Wild type structure 3CWM with β -sheet A not expanded into SITE6. (Light blue) expanded β -strand s2A in structure 3T1P, which occupies SITE6. (Light grey) β -strand s2A in structure 1QLP adjacent to SITE6.

which can also prevent polymerization and whose mechanism of action has yet to be determined (Pearce et al., 2008).

A.3.7 Residue interactions with S-(4-nitrobenzyl)-6-thioguanosine

Nearby residues that interact with S-(4-nitrobenzyl)-6-thioguanosine at the top ranking sites (SITE1, SITE2, SITE5) from the M* intermediate model and the single site (SITE6) from the wild type model are described in Table A.2. This information provides the basis for guiding further validations of these binding sites using techniques such as mutagenesis and molecular dynamics. Supplemental Figures S1-S4 contain additional details about the type of interactions formed between individual atoms of S-(4-nitrobenzyl)-6-thioguanosine and the nearby residues listed in Table A.2.

A.4 Discussion

Currently, the only available and effective treatment to correct for the loss of α 1AT function in α 1ATD associated with liver disease is orthotropic liver transplantation. For lung disease, augmentation therapy is the only specific regimen that is thought to slow down disease progression, although this still requires formal proof through well-controlled clinical trials.(Gøtzsche & Johansen, 2010) As these treatments are expensive, labor intensive and associated with side effects, the need for novel treatments are indeed in high-demand. With Z- α 1AT polymerization being responsible for the development of the disease, blocking its aggregation by small molecules (Mallya et al., 2007; Pearce et al., 2008) appears to be a promising strategy to cure Z- α 1ATD.

Here we report an integrated *in vitro* and *in silico* approach which allows discovering and characterizing potent small molecules that disrupt the pathological polymerization of Z- α 1AT. The *in vitro* microplate assay which enables the identification of small molecules able to block the insertion of a modified 6-mer peptide into the s4A cavity, provides quantitative data with reproducibility, sensitivity and rapid throughput. Our results validate the utility of the *in vitro* screening assay and identify S-(4-nitrobenzyl)-6-thioguanosine as inhibitor of Z- α 1AT polymerization. With a molecular weight of 434.43 Da, 4 H-bond donors, 11 H-bond acceptors and a low lipophilicity coefficient (XLogP3 =1.1), this compound presents a strong drug-like profile according to the Lipinski rule of five.(Leeson, 2012) From IC₅₀ determination and structure-activity relationship studies, we also found one of its structural homologues which differs by a single amino group and does not prevent aggregation. This suggests that an interaction with the amino group may be important to counteract the insertion of the modified 6-mer peptide.

The microplate assay has been designed to identify any inhibitor that can impede the insertion of the RCL into the s4A cavity, but it does not exclude the discovery of small molecules that can bind outside of the s4A cavity, causing a conformational rearrangement that still precludes RCL insertion. Molecular docking experiments were

carried out to investigate binding of S-(4-nitrobenzyl)-6-thioguanosine at several potential binding locations, in addition to the s4A cavity. A chimera homology model of the intermediate state, M* model, was built in order to allow investigation of the s4A cavity. Until now, previous studies have only used molecular docking to investigate the binding of small molecules into experimentally resolved structures and only at sites other than the s4A cavity.

The development for the first time of an atomistic M* model reveals a mechanism through which S-(4-nitrobenzyl)-6-thioguanosine may inhibit Z- α 1AT polymerization by either competing with the RCL at the s4A insertion site (SITE5) or by blocking the s4A cavity from forming by binding at a nearby location (SITE6). To definitely discriminate between these two binding sites, additional structural studies will need to be carried out beyond the scope of the present work which will include selected mutagenesis and molecular dynamics (MD). MD simulations, which provide an ensemble of various conformations of M*, will account for the protein flexibility^{38,39} and will aid in refining the docking results.

Table A.1. Docking results from M- α 1AT, Z- α 1AT, and M* model with S-(4-nitrobenzyl)-6-thioguanosine and the 79 other small molecules.

Structure	SITE ^α	Lowest Energy ^β (kcal/mol)	B9 Energy ^γ (kcal/mol)	B9 Rank ^δ
M- α 1AT (3CWM)	SITE1	n/a [⊖]	n/a	n/a
	SITE2	-7.0	-5.6	19 th
	SITE3	-6.2	-5.3	8 th
	SITE4	-6.4	-3.4	25 th
	SITE5	n/a	n/a	n/a
	SITE6	-7.6	-7.6	1 st
Z- α 1AT (3T1P)	SITE1	n/a	n/a	n/a
	SITE2	-6.5	-3.9	34 th
	SITE3	-9.0	-5.8	13 th
	SITE4	-4.6	-3.6	19 th
	SITE5	n/a	n/a	n/a
	SITE6	n/a	n/a	n/a
M* Model	SITE1	-10.7	-7.9	7 th
	SITE2	-10.7	-8.8	5 th
	SITE3	-5.1	-4.6	9 th
	SITE4	n/a	n/a	n/a
	SITE 5	-8.4	-8.4	1 st
	SITE6	-6.2	-4.22	12 th

^α Site number where S-(4-nitrobenzyl)-6-thioguanosine (B9) was docked. ^β Lowest observed binding energy (kcal/mol) for any of the 80 docked compounds. ^γ Predicted binding energy (kcal/mol) for B9. ^δ Rank of B9 relative to the binding energies for all 80 docked compounds. [⊖] Site numbers that are not found in a given model are noted by a not applicable symbol (n/a).

Table A.2. Residues interacting with S-(4-nitrobenzyl)-6-thioguanosine in top scoring binding sites.

M* Model	SITE 1	S34, I35, A37, F38, L41, L149, T157, F159, A160, L161, V162, N163, Y164, L276, F289, L304, K305, L306, K308, A309, V310, H311
	SITE 2	W171, E172, R173, P174, F175, R200, M203, F204, N205, L218, M219, K220, Y221, F229, E256, D257, R258, L263, L265, I317, D318, F329, E331
	SITE 5	F28, K145, I146, I165, F166, F167, K168, V314, L315, C316, I317, D318, E319, K320, G321, T322, E323, A324, M351, F361
M- α 1AT (3CWM)	SITE 6	S56, T59, A60, M63, L100, N104, Q105, L112, T113, T114, G115, N116, G1117, Y138, H139, S140, E141, Y160, G164, N186, Y187, I188

Listed are the interacting residues for binding sites where S-(4-nitrobenzyl)-6-thioguanosine ranks in the top 10% of compounds and less than -7kcal/mol.

Appendix A References

- Belorgey, D., Irving, J. a, Ekeowa, U. I., Freeke, J., Roussel, B. D., Miranda, E., ... Lomas, D. a. (2011). Characterisation of serpin polymers in vitro and in vivo. *Methods (San Diego, Calif.)*, 53(3), 255–66. <https://doi.org/10.1016/j.ymeth.2010.11.008>
- Brantly, M., Nukiwa, T., & Crystal, R. G. (1988). Molecular basis of alpha-1-antitrypsin deficiency. *The American Journal of Medicine*, 84(6A), 13–31.
- Burrows, J. a, Willis, L. K., & Perlmutter, D. H. (2000). Chemical chaperones mediate increased secretion of mutant alpha 1-antitrypsin (alpha 1-AT) Z: A potential pharmacological strategy for prevention of liver injury and emphysema in alpha 1-AT deficiency. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), 1796–801.
- Campos, M. a, & Lascano, J. (2014). α 1 Antitrypsin deficiency: current best practice in testing and augmentation therapy. *Therapeutic Advances in Respiratory Disease*. <https://doi.org/10.1177/1753465814542243>
- Chemical Computing Group - Citing MOE. (2012).
- Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P., & Lomas, D. a. (1999). A kinetic mechanism for the polymerization of alpha1-antitrypsin. *The Journal of Biological Chemistry*, 274(14), 9548–55. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10092640>
- de Serres, F. J., & Blanco, I. (2012). Prevalence of α 1-antitrypsin deficiency alleles PI*S and PI*Z worldwide and effective screening for each of the five phenotypic classes PI*MS, PI*MZ, PI*SS, PI*SZ, and PI*ZZ: a comprehensive review. *Therapeutic Advances in Respiratory Disease*, 6(5), 277–95. <https://doi.org/10.1177/1753465812457113>
- Devlin, G. L., Parfrey, H., Tew, D. J., Lomas, D. a, & Bottomley, S. P. (2001). Prevention of polymerization of M and Z alpha1-Antitrypsin (alpha1-AT) with trimethylamine N-oxide. Implications for the treatment of alpha1-at deficiency. *American Journal of Respiratory Cell and Molecular Biology*, 24(6), 727–32. <https://doi.org/10.1165/ajrcmb.24.6.4407>
- Ekeowa, U. I., Freeke, J., Miranda, E., Gooptu, B., Bush, M. F., Perez, J., ... Lomas, D. A. (2010). Defining the mechanism of polymerization in the serpinopathies. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40), 17146–17151. <https://doi.org/10.1073/pnas.1004785107>
- Elliott, P. R., Bilton, D., & Lomas, D. A. (1998). Lung polymers in Z alpha(1)-antitrypsin deficiency-related emphysema. *American Journal of Respiratory Cell and Molecular Biology*, 18(5), 670–674.
- Elliott, P. R., Pei, X. Y., Dafforn, T. R., & Lomas, D. a. (2000). Topography of a 2.0 Å structure of alpha1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Science : A Publication of the Protein Society*, 9(7),

1274–81. <https://doi.org/10.1110/ps.9.7.1274>

- Eriksson, S., Carlson, J., & Velez, R. (1986). Risk of cirrhosis and primary liver cancer in alpha 1-antitrypsin deficiency. *The New England Journal of Medicine; The New England Journal of Medicine*, 314(12), 736–739.
- Gladkevich, a, Bosker, F., Korf, J., Yenkovyan, K., Vahradyan, H., & Aghajanov, M. (2007). Proline-rich polypeptides in Alzheimer's disease and neurodegenerative disorders -- therapeutic potential or a mirage? *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 31(7), 1347–55. <https://doi.org/10.1016/j.pnpbp.2007.06.005>
- Gooptu, B., Dickens, J. A., & Lomas, D. A. (2014a). The molecular and cellular pathology of α 1-antitrypsin deficiency. *Trends in Molecular Medicine*, 20(2), 116–27. <https://doi.org/10.1016/j.molmed.2013.10.007>
- Gooptu, B., Dickens, J. a, & Lomas, D. a. (2014b). The molecular and cellular pathology of α ₁-antitrypsin deficiency. *Trends in Molecular Medicine*, 20(2), 116–27. <https://doi.org/10.1016/j.molmed.2013.10.007>
- Gooptu, B., Hazes, B., Chang, W. S., Dafforn, T. R., Carrell, R. W., Read, R. J., & Lomas, D. a. (2000). Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. *Proceedings of the National Academy of Sciences of the United States of America*, 97(1), 67–72. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=26617&tool=pmcentrez&rendertype=abstract>
- Gøtzsche, P. C., & Johansen, H. K. (2010). Intravenous alpha-1 antitrypsin augmentation therapy for treating patients with alpha-1 antitrypsin deficiency and lung disease. *The Cochrane Database of Systematic Reviews*, (7), CD007851.
- Huntington, J. a, Read, R. J., & Carrell, R. W. (2000). Structure of a serpin-protease complex shows inhibition by deformation. *Nature*, 407(6806), 923–6. <https://doi.org/10.1038/35038119>
- Hussain, M., Mieli-Vergani, G., & Mowat, a P. (1991). Alpha 1-antitrypsin deficiency and liver disease: clinical presentation, diagnosis and treatment. *Journal of Inherited Metabolic Disease*, 14(4), 497–511. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1749215>
- Issue, I. T. (2016). Welcome Autumn 2016 In This Issue Current status of access to intravenous alpha - 1 antitrypsin augmentation therapy Current status of access to intravenous alpha - 1 antitrypsin augmentation therapy (continued), (15), 1–40.
- Kalsheker, N. (1989). Alpha 1-antitrypsin: structure, function and molecular biology of the gene. *Bioscience Reports*, 9(2), 129–38. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2669992>
- Labute, P. (2008). The generalized Born/volume integral implicit solvent model:

- Estimation of the free energy of hydration using London dispersion instead of atomic surface area. *Journal of Computational Chemistry*, 29(10), 1693–1698. <https://doi.org/10.1002/jcc.20933>
- Labute, P. (2009). Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins: Structure, Function, and Bioinformatics*, 75(1), 187–205. <https://doi.org/10.1002/prot.22234>
- Leeson, P. (2012). Drug discovery: Chemical beauty contest. *Nature*, 481(7382), 455–6. <https://doi.org/10.1038/481455a>
- Lomas, D. A. (2000). Loop-sheet polymerization: the mechanism of alpha1-antitrypsin deficiency. *Respiratory Medicine*, 94 Suppl C, S3-6.
- Lomas, D. A., Evans, D. L., Stone, S. R., Chang, W. S., & Carrell, R. W. (1993). Effect of the Z mutation on the physical and inhibitory properties of alpha 1-antitrypsin. *Biochemistry (John Wiley & Sons); Biochemistry*, 32(2), 500–508.
- Lomas, D. A., Perlmutter, D. H., & Uversky, V. N. (2010). Alpha-1Antitrypsin Deficiency. In M. Ramirez-Alvarado, J. W. Kelly, & C. M. Dobson (Eds.), *Protein Misfolding Diseases - Current and Emerging Principles and Therapies* (Vol. 2010, pp. 403–424). Hoboken, New Jersey: A John Wiley & Sons, Inc.
- Mahadeva, R., Atkinson, C., Li, Z., Stewart, S., Janciauskiene, S., Kelley, D. G., ... Lomas, D. A. (2005). Polymers of Z alpha1-antitrypsin co-localize with neutrophils in emphysematous alveoli and are chemotactic in vivo. *American Journal of Pathology*, 166(2), 377–386.
- Mahadeva, R., Dafforn, T. R., Carrell, R. W., & Lomas, D. A. (2002). 6-mer peptide selectively anneals to a pathogenic serpin conformation and blocks polymerization. Implications for the prevention of Z alpha(1)-antitrypsin-related cirrhosis. *Journal of Biological Chemistry*, 277(9), 6771–6774. <https://doi.org/10.1074/jbc.C100722200>
- Makarananda, K., & Neal, G. E. (1992). Competitive ELISA. *Methods in Molecular Biology (Clifton, N.J.)*, 10, 267–72. <https://doi.org/10.1385/0-89603-204-3:267>
- Mallya, M., Phillips, R. L., Saldanha, S. A., Gooptu, B., Brown, S. C. L., Termine, D. J., ... Lomas, D. a. (2007). Small molecules block the polymerization of Z alpha1-antitrypsin and increase the clearance of intracellular aggregates. *Journal of Medicinal Chemistry*, 50(22), 5357–63. <https://doi.org/10.1021/jm070687z>
- Ogushi, F., Fells, G. a, Hubbard, R. C., Straus, S. D., & Crystal, R. G. (1987). Z-type alpha 1-antitrypsin is less competent than M1-type alpha 1-antitrypsin as an inhibitor of neutrophil elastase. *The Journal of Clinical Investigation*, 80(5), 1366–74. <https://doi.org/10.1172/JCI113214>
- Parfrey, H., Mahadeva, R., Ravenhill, N. A., Zhou, A., Dafforn, T. R., Foreman, R. C., & Lomas, D. A. (2003). Targeting a surface cavity of alpha 1-antitrypsin to prevent conformational disease. *Journal of Biological Chemistry*, 278(35), 33060–33066.

<https://doi.org/10.1074/jbc.M302646200>

- Parmar, J. S., Mahadeva, R., Reed, B. J., Farahi, N., Cadwallader, K. A., Keogan, M. T., ... Lomas, D. A. (2002). Polymers of alpha(1)-antitrypsin are chemotactic for human neutrophils: a new paradigm for the pathogenesis of emphysema. *American Journal of Respiratory Cell and Molecular Biology*, 26(6), 723–730.
- Pearce, M. C., Morton, C. J., Feil, S. C., Hansen, G., Adams, J. J., Parker, M. W., & Bottomley, S. P. (2008). Preventing serpin aggregation: The molecular mechanism of citrate action upon antitrypsin unfolding. *Protein Science*, 17(12), 2127–2133. <https://doi.org/10.1110/ps.037234.108>
- Perlmutter, D. H. (2011). Alpha-1-antitrypsin deficiency: importance of proteasomal and autophagic degradative pathways in disposal of liver disease-associated protein aggregates. *Annual Review of Medicine*, 62, 333–45. <https://doi.org/10.1146/annurev-med-042409-151920>
- Purkayastha, P., Klemke, J. W., Lavender, S., Oyola, R., Cooperman, B. S., & Gai, F. (2005). Alpha 1-antitrypsin polymerization: a fluorescence correlation spectroscopic study. *Biochemistry*, 44(7), 2642–2649. <https://doi.org/10.1021/bi048662e>
- Saez, E., No, D., West, a, & Evans, R. M. (1997). Inducible gene expression in mammalian cells and transgenic mice. *Current Opinion in Biotechnology*, 8(5), 608–16. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9353233>
- Teckman, J. H. (2004). Lack of effect of oral 4-phenylbutyrate on serum alpha-1-antitrypsin in patients with alpha-1-antitrypsin deficiency: a preliminary study. *Journal of Pediatric Gastroenterology and Nutrition*, 39(1), 34–37.
- Whisstock, J. C., & Bottomley, S. P. (2006). Molecular gymnastics: serpin structure, folding and misfolding. *Current Opinion in Structural Biology*, 16(6), 761–768. <https://doi.org/10.1016/j.sbi.2006.10.005>
- Yamasaki, M., Li, W., Johnson, D. J. D., & Huntington, J. a. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature*, 455(7217), 1255–8. <https://doi.org/10.1038/nature07394>
- Yamasaki, M., Li, W., Johnson, D. J., & Huntington, J. A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature*, 455(7217), 1255–1258. <https://doi.org/10.1038/nature07394>
- Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C., & Huntington, J. a. (2011). Molecular basis of α 1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Reports*, 12(10), 1011–7. <https://doi.org/10.1038/embor.2011.171>

Appendix B

Alpha-1 Antitrypsin Mammalian Cell Work

Alpha-1 antitrypsin deficiency (α_1 ATD) is a disorder associated with chronic liver and/or lung disease [116]. In this condition, the lack of circulating alpha-1 antitrypsin in the blood stream leads to emphysema, and the abnormal accumulation of alpha-1 antitrypsin in the liver leads to hepatocellular diseases such as cirrhosis [117, 118]. While there are over 100 variants associated with α_1 AT, the Z variant of ($Z\alpha_1$ AT) is the most common and severe form of Alpha1 Antitrypsin deficiency (α_1 ATD) [119]. $Z\alpha_1$ AT pathogenesis is amplified based on the ability of the disease-protein to form polymers in hepatocytes, inhibiting its secretion into the blood [120]. At present, there is no cure for α_1 ATD. Although many new therapies are in the pipeline to help cure α_1 ATD, the only available therapy to prevent lung disease progression is limited to augmentation therapy and organ transplant [121, 122]. It is hypothesized that breaking these hepatocyte polymers will reduce polymer accumulation in the liver and increase alpha-1 antitrypsin in the blood, thereby reducing the risk of disease [123, 124].

To discover small compounds and polymerization breakers, we utilized an *in vitro* and *in silico* approach to identify potential drug candidates, as identified in appendices A. Based on our results, we identified one excellent candidate B9 (S-(4-nitrobenzyl)-6-thioguanosine), along with 10 other potential candidates from *in silico* experiments that needed to be tested further [125]. To test these potential drug candidates and further test B9, we created a controlled expression system in liver hepatocellular carcinoma (HepG2) cells, to monitor the polymer reduction.

In this regard, Expression Constructs offer an *in vitro* controlled system that allows experimentalists to examine the genes and proteins of interest. For this study, we will be using the T-RExCell Expression system from Life Technologies. This expression system contains two plasmids, namely the inducible expression plasmid pcDNA4/TO and the repressor plasmid pcDNA6/TR that represses the transcription of pcDNA4/TO. Notably, in the absence of tetracycline, the 6/TR repressor binds to 4/TO and transcription is arrested. On the other hand, in the presence of tetracycline, tetracycline binds to the repressor and activates transcription; see Figure B.1 [126]. In this context, creating a controlled system will allow us to control and monitor the production of polymers. Ideally, a stable mammalian construct is the best option for obtaining these desired results, but transient systems are also capable of producing comparable results. To create a mammalian expression system, both $Z\alpha_1$ AT and $M\alpha_1$ AT will need to be ligated to the pcDNA4/TO expression vector designed for the T-REx system [127, 128]. Once

confirmed, mammalian Hep G2 cells will be transfected or co-transfected with pcDNA4/TO, Z α ₁AT/ M α ₁AT and/or pcDNATM6/TR vector to create a stable expression system for testing compounds that prevent polymerization; see Figure B.2.

The mammalian system that we created for this study functions in two parts. The first part of the system has been created to monitor polymer production intracellularly. This construct consists of Z α ₁AT/M α ₁AT protein with a 6xhis-tag and a no signal peptide, which will be used to monitor the breaking of polymers intracellularly. The lack of the signal peptide will allow us to monitor polymer production and reduction in the cytoplasm of this mammalian cell system. The second construct consists of Z α ₁AT/M α ₁AT protein with a 6xhis-tag and a signal peptide to monitor the functionality of the protein after treatment with the potential therapeutics. Upon the addition of the potential drug candidate, the polymerization should reduce significantly and be secreted into the media, where we can then measure the functionality of the protein. For the remainder of this discussion, these constructs will be called intracellular and full-length constructs; M is for wildtype and Z is for the mutant.

B.1 Method and results

B.1.1. Construction of pcDNA4/TO, M α ₁AT (intracellular and full length) and pcDNA4/TO, Z α ₁AT (intracellular and full length)

To begin with, both M α ₁AT and Z α ₁AT DNA intracellular fragments were digested from the pPic3.5K vector with the restriction enzymes BamHI and EcoRI. Then, according to the manufacturer's instructions, the M α ₁AT and Z α ₁AT fragments (1278bp) were subcloned into pcDNA4/TO (5078bp) using T4 DNA ligase. Following ligation, transformation was carried out in MAX Efficiency DH5 α TM *E. coli* cells and selected for 100 μ g/mL Ampicillin. The resulting plasmid was verified by restriction digests (Figure B.3). Moreover, both M α ₁AT and Z α ₁AT DNA full length fragments were synthesized by GenScript and verified, Figure B.4 and B.5. After verification of M α ₁AT and Z α ₁AT in both intracellular and full length, these constructs were ready to be used to transfect mammalian cells.

B 1.2. HepG2 transfection optimization

Importantly, establishing a transfection protocol for each mammalian cell line is critical for transient and stable transfection studies. To establish an efficient protocol, various transfection reagents were used and tested against a GFP expression plasmid; see Table 1. For GFP expression, 1-3 μ g of DNA per well in 6 well plates were used. , As per the manufacturer's directions, the ratios for lipofectamine reagents were 1:3 to 1:5 DNA (μ g). Based on our transfection results, we could identify 4 candidate transfection reagents to be used with our constructs. Next, we next analyzed the top 4 transfection reagents (Lipofectamine LTX reagent with plus reagent, Lipofectamine 3000 transfection reagent, metafectene Pro, and TransIT-LT1) with our constructs pcDNA4/TO M α ₁AT and

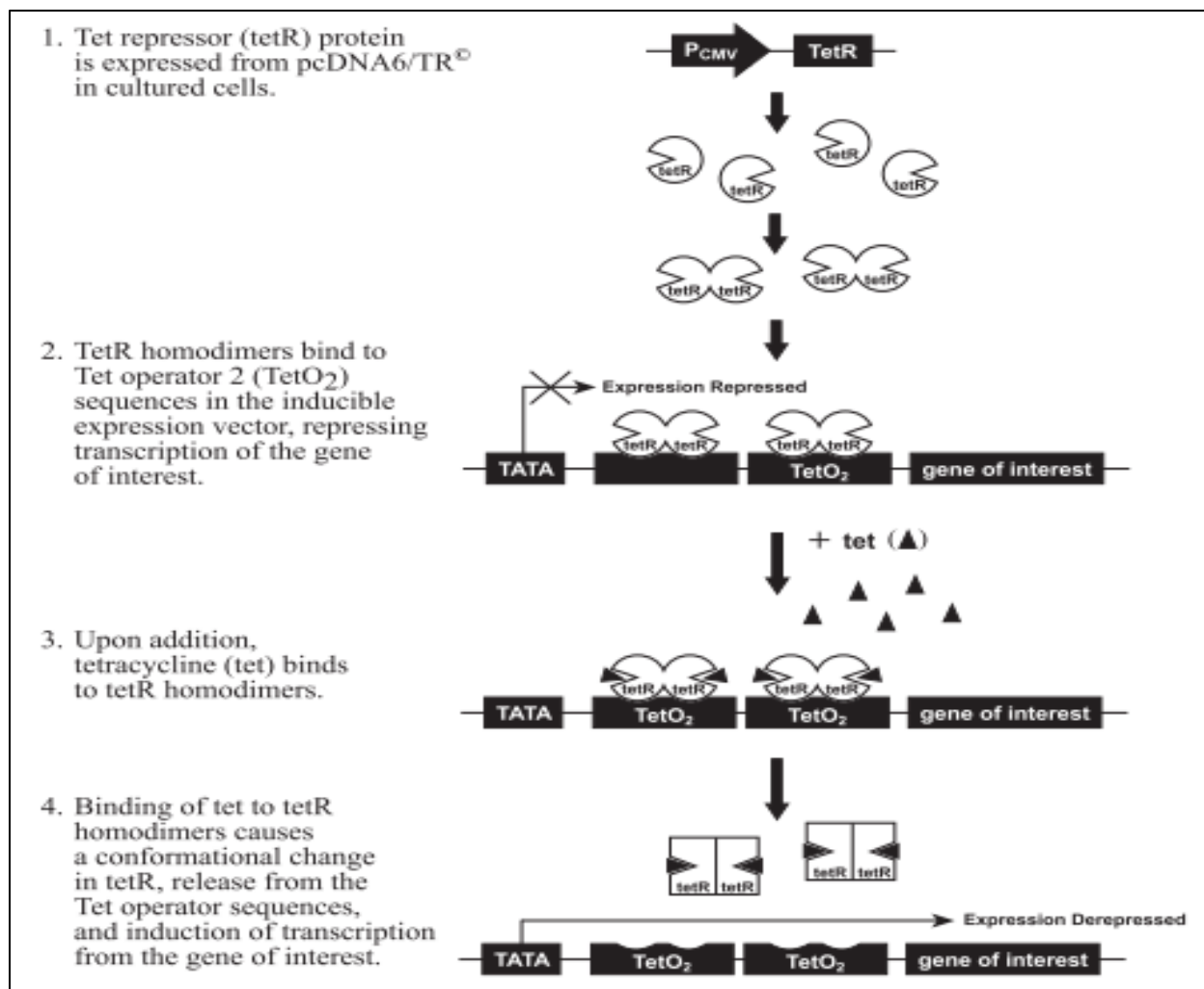


Figure B.1: Figure from Invitrogen, life technologies. Components of the T-REx system, pcDNA6/TR is generating the repressor for pcDNA4/TO, which is the inducible expression plasmid (T-Rex[™] system, 2011).

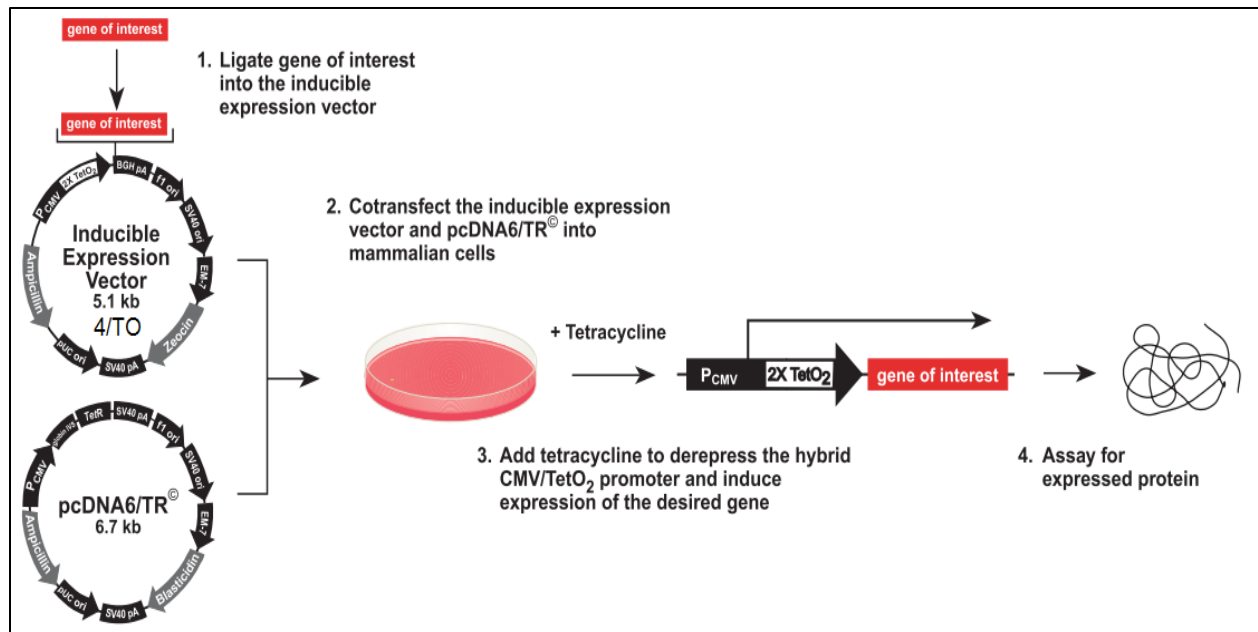


Figure B.2: Figure taken from Invitrogen, life technologies: pcDNA4/TO can be transfected transiently to produce constitutive expression of Z α ₁AT/M α ₁AT or it can be co-transfected with pcDNA6/TR to demonstrate that the system may be inducible (T-Rex™ system, 2011).

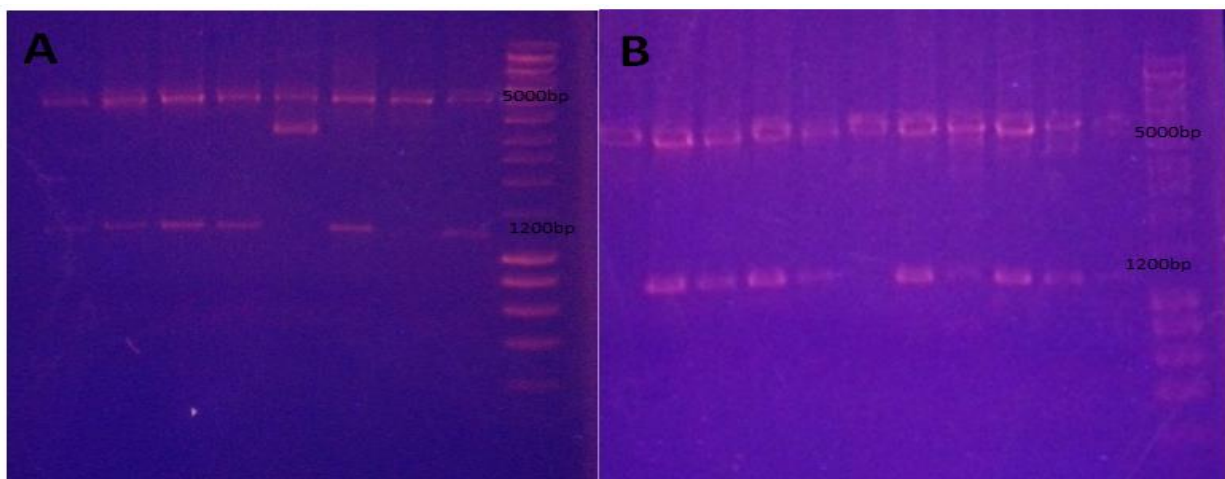


Figure B.3: Verification of (A) pcDNA4/TO $M\alpha_1AT$, (B) pcDNA4/TO $Z\alpha_1AT$ DNA intracellular constructs.

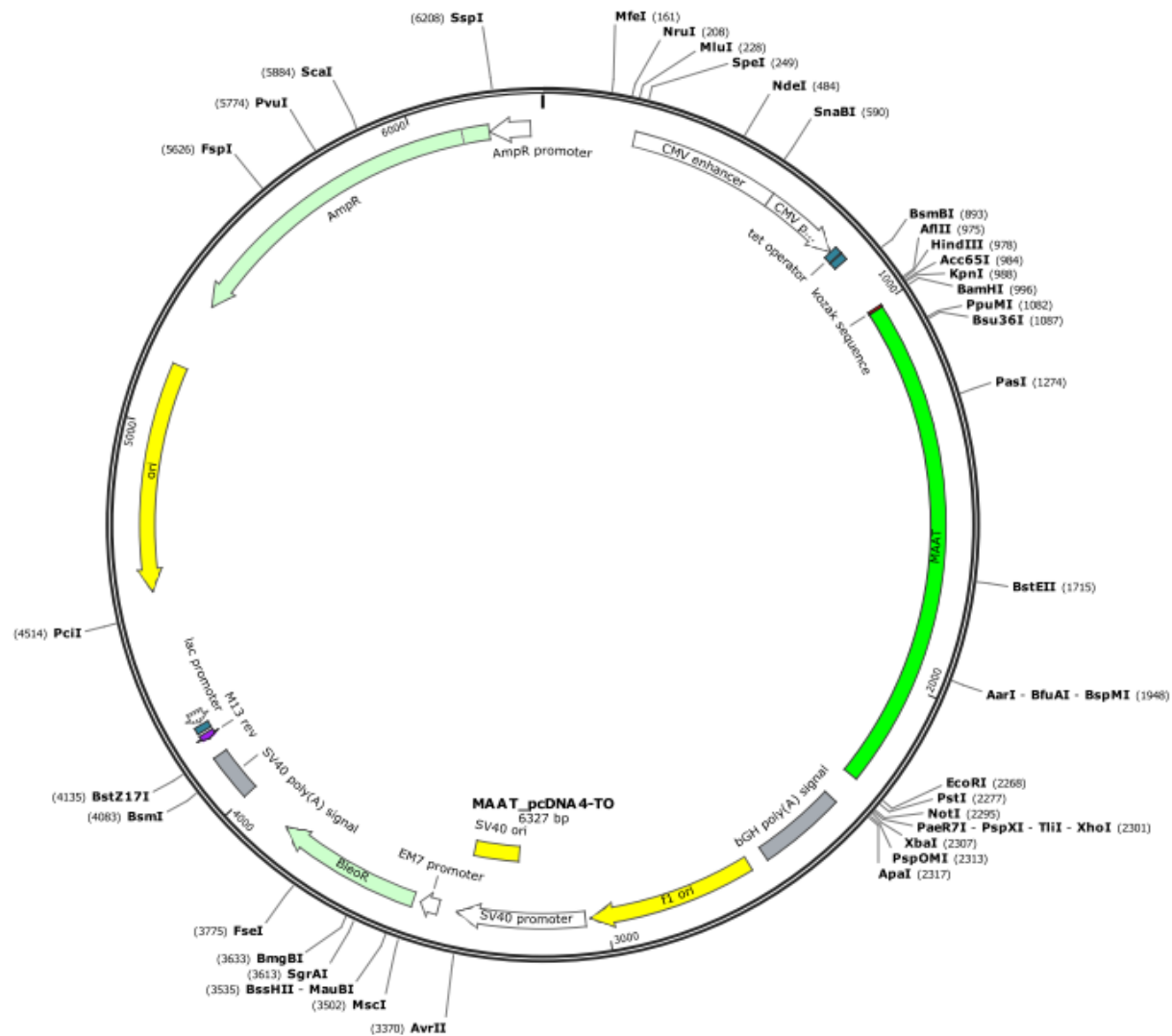


Figure B.4: pcDNA4/TO, Ma₁AT Vector used for expression in HepG2 cells.

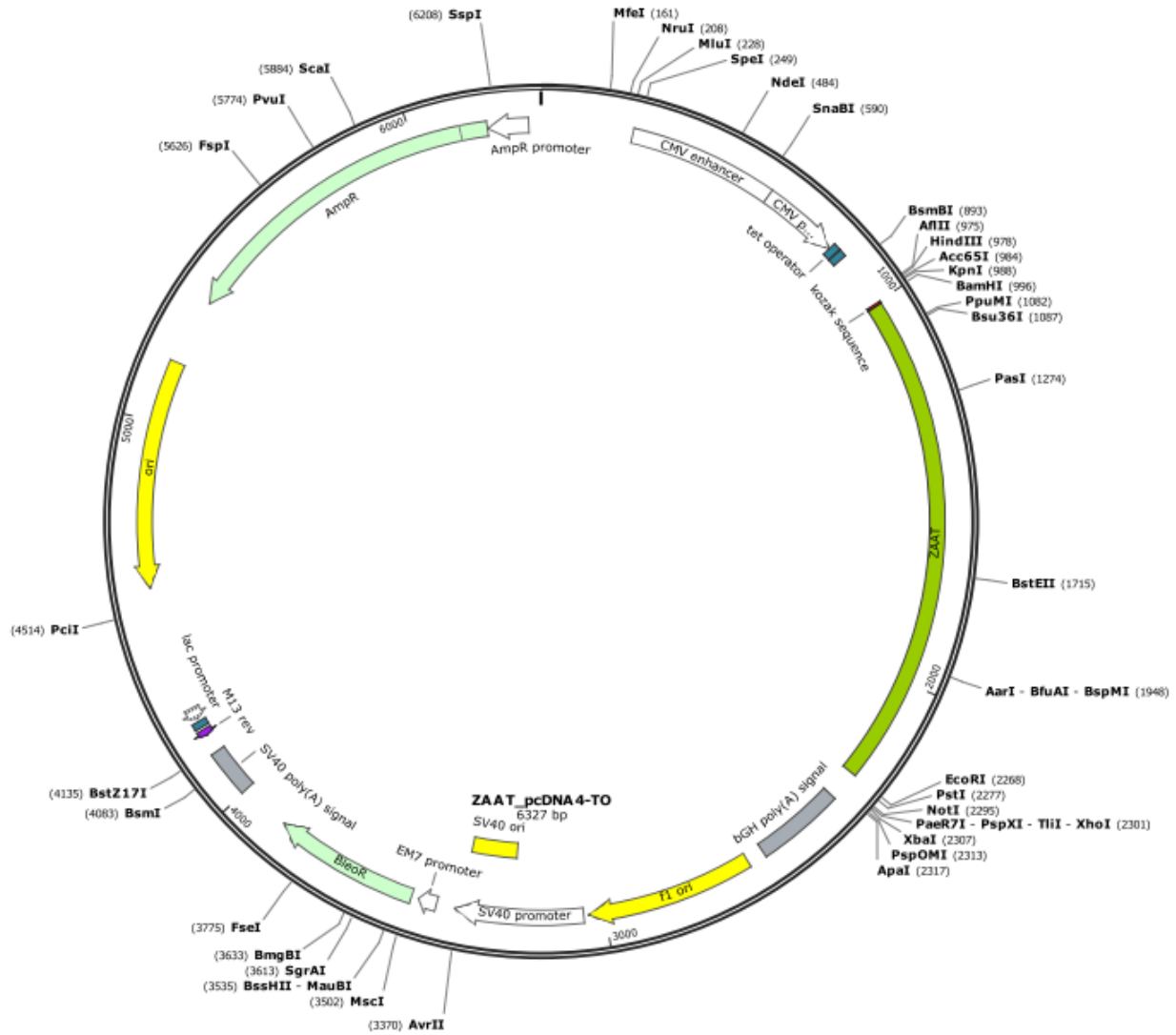


Figure B.5: pcDNA4/TO, Z α 1AT Vector used for expression in HepG2 cells.

Table B.1: Transfection reagent optimization with various transfection reagents. Lipofectamine 3000 transfection reagent responded the best for Ma1AT and Za1AT.

Transfection Reagent	% Transfection Efficiency with GFP	% Transfection Efficiency with MAAT and ZAAT	Cellular Toxicity
Life Technologies: Lipofectamine LTX Reagent with PLUS Reagent	80%	MAAT 10% ZAAT 10%	About 50% death in some cases, but may also be as low as 5%
Life Technologies: Lipofectamine 3000 Transfection Reagent	75%	MAAT 20% ZAAT 10%	About 30% death in some cases but may also be as low as 5%
Biontix: Metafectene Pro	60%	MAAT 10% ZAAT 5-10%	Minimal
Mirus Bio: TransIT-LT1	70%	MAAT 5% ZAAT 5%	Minimal
Mirus Bio: TransIT-X2	35%	N/A	Minimal
Mirus Bio: TransIT-2020	25%	N/A	Minimal

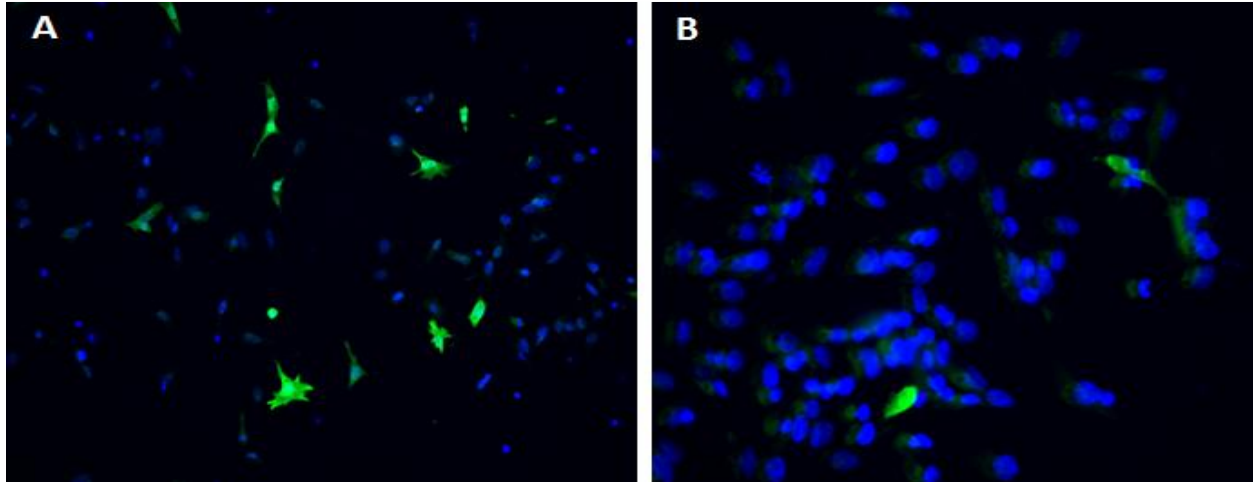


Figure B.6: Transient transfection 48 hours after transfection with lipofectamine 3000 transfection reagent in HepG2 cells. (A) pcDNA4/TO M α ₁AT; (B) pcDNA4/TO Z α ₁AT.

pcDNA4/TO Z α ₁AT. We then transfected HepG2 cells with M α ₁AT and Z α ₁AT, allowed them to express for 48 hours and later imaged them using the Leitz DMRB Fluorescence Microscope (Complete protocol below). It was observed that the Lipofectamine 3000 transfection reagent responded well for both M α ₁AT and Z α ₁AT with a transfection efficiency between ~10 and 20%; see Figure B.6 for details. In addition to testing transfection efficiency, we also wanted to check cellular toxicity with each of the reagents. Trypan Blue (0.4%) was used to count viable cells in each culture. Not only did lipofectamine 3000 have the best transfection with our constructs, but it also showed the minimal cellular toxicity in most cases; see Table 1 for more information.

B.1.3. Transient transfection test with M α ₁AT and Z α ₁AT

Significantly, transient transfections of HepG2 and COS-7 cells have been used to study a variety of serpins and their polymerization mechanisms [129, 130]. Hence, 6 well plates (with cover slips for imaging) were coated with poly-L-lysine (125mg per well) and incubated for an hour; these plates were then allowed to dry overnight in a sterile cabinet. Because HepG2 cells tend to clump, adding poly-L-Lysine aids in the cells forming a monolayer; therefore 200,000 HepG2 cells were plated in a 2 mL culture containing EMEM + 10% FBS. After the cells grew to a confluency of about 60%, the pcDNA4/TO plasmids encoding wild type (M α ₁AT) and mutant (Z α ₁AT) were transfected into HepG2 cells with Lipofectamine 3000 (following transfection ratios from above). Afterward, the transfection serum medium was removed and replaced with EMEM and no FBS, and then incubated at 37°C incubator with 5% CO₂ for 3-5 hours. After incubation, the serum medium was added (EMEM + 10% FBS) back to the wells and assayed for expression between 24-72 hours (protocol below).

B.1.4. Fixing and staining cells for expression

At each time point (24h, 48h, and 72h), cells were fixed using a BD Cytofix Fixation Buffer. First, the cells were washed with 1X PBS. Then, 500ul of BD Cytofix Fixation Buffer was added and the cells were incubated at room temperature for 15-30 minutes; alternatively, a fixing buffer with (0.2%) glutaraldehyde and (4%) formaldehyde can be used. Finally, the cells were washed two more times with 1X PBS and placed back in 1X PBS, until they were ready to be stained with antibodies. Although fixed cells can be stored at 4°C for up to 3 weeks, it is always recommended to do the staining as soon as possible.

B.1.5. Immunofluorescence in HepG2 cells

To detect our constructs in both wild type ($M\alpha_1AT$) and mutant ($Z\alpha_1AT$), we used 2 different antibody conjugates. The first antibody was used to detect alpha-1 antitrypsin in both wild type ($M\alpha_1AT$) and mutant ($Z\alpha_1AT$). Here, we used a Labelled StrepAvidin Biotin (LSAB) method; we also used the goat polyclonal to alpha-1-antitrypsin (HRP) (ab7635) and conjugated that antibody to the secondary antibody (anti goat streptavidin), which in turn was conjugated to a tertiary antibody anti-streptavidin; this combination was conjugated to an Alexa Fluor 594 dye (Red fluorescence) where we could indirectly monitor fluorescence. The second antibody was used to detect the His-tag within the constructs. The his-tag antibody was a mouse monoclonal His.HB to 6X His tag® (ab18184), and it was conjugated to a secondary His-tag antibody, namely the anti-mouse Alexa Fluor 488 (Green fluorescence). All the antibodies used in these experiments were at a working concentration of 1/1000U.

Then, HepG2 cells were permeabilized with 1X PBS and 0.2% Triton X-100 for 5 minutes, and then washed in 1X PBS for 5 minutes 3x. Nonspecific binding sites were blocked using a blocking buffer for 30 minutes; the Blocking Buffer consisted of PBST (1X PBS + 0.1% Tween 20) and 1% BSA. Following incubation, the primary antibody was added and incubated for 1 hour, and then washed with 1X PBS for 5 minutes 3x. The same procedure was repeated for secondary and tertiary antibodies. Then, there was a final wash with 1X PBS for 5 minutes. Finally, a 1X PBS and Hoechst stain (Blue fluorescence, nucleus stain) solution was applied and incubated for 5 minutes; next, the cells were washed again in 1X PBS for 5 minutes 2x. Subsequently, the cells were visualized using the Leitz DMRB Fluorescence Microscope at magnifications between 10-40X.

Afterwards, each of the two constructs, intracellular and full length, were analyzed with both antibody conjugates. The intracellular construct pcDNA4/TO $M\alpha_1AT$ and $Z\alpha_1AT$ was analyzed over a period of 24-72 hours. At each time point, the above protocol was applied with the alpha-1-antitrypsin antibodies. These data showed that after 24 hours of incubation, we see minimal expression of either wild type or mutant. Between 48 and 72, we get the best expression from both constructs, with the mutant expressing the best at 72 hours, see Figure B.7 for additional information. Additionally, we analyzed this

construct using the His-tag antibodies and found the same results; please note that data corresponding to this analysis is not shown. Using the His-tag antibody conjugates, we analyzed the full-length construct for both pcDNA4/TO M α_1 AT and Z α_1 AT at 48 hours; see Figure B.8 for details. These data show that the full-length protein is expressed in the sample. We also measured the alpha 1 antitrypsin antibody and found the same results; the resultant data is not shown.

B.2 Conclusion

Given the seriousness of α_1 ATD and its potentially dire consequences, evaluating the efficacy of treatments to mitigate this disease is vital. Collectively, these aforementioned data from our study have shown that we have an inducible transient system that could be utilized for drug monitoring and drug screening assays. Our intracellular construct, comprising both wildtype and mutant, is sufficient for monitoring the reduction of polymers in the cytoplasm after the treatment with a drug. Our full-length construct, comprising both wild type and mutant, is sufficient for monitoring intracellular expression and extracellular expression after the treatment with a drug. Additionally, this construct could be used to monitor the secretion of Z α_1 AT and may also be used to test the functionality of the Z α_1 AT mutants. The comprehensive data obtained from our system can be used to choose the most efficacious treatment to cure or control α_1 ATD and its attendant risks.

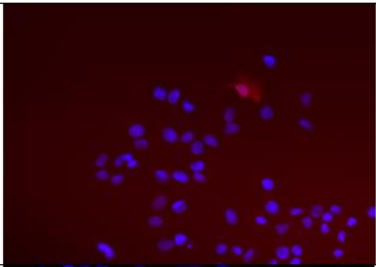
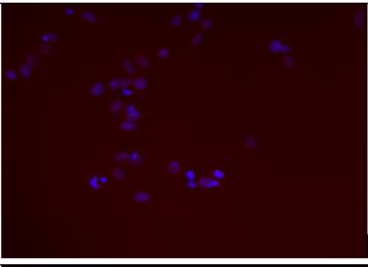
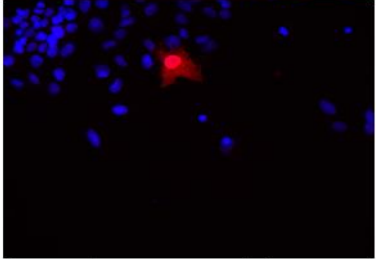
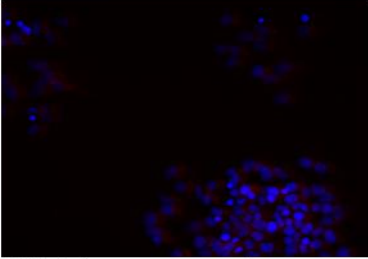
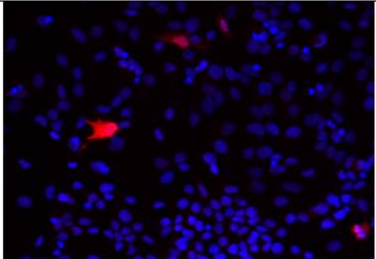
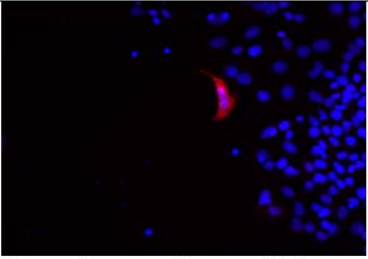
	MAAT	ZAAT
24h (4-22-15)		
48h (4-22-15)	 Expression was more intense	 Little bit more expression
72h (4-22-15)	 Expression was even more intense	 Expression was the most intense at 72

Figure B.7: This is the intracellular construct for both wild type ($M\alpha_1AT$) and mutant ($Z\alpha_1AT$) expressed in HepG2 cells.

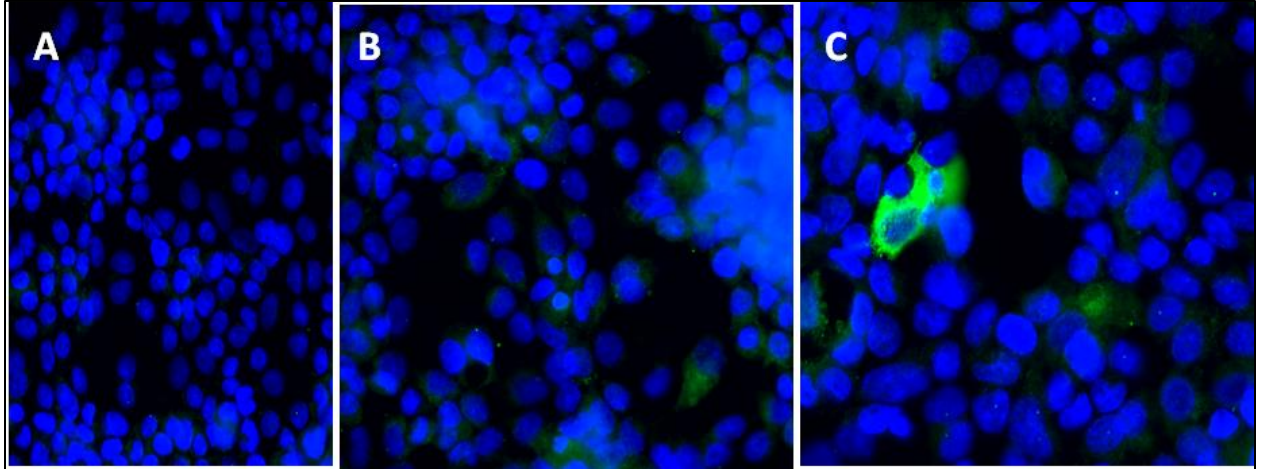


Figure B.8: This is the full-length construct for $M\alpha_1AT$, $Z\alpha_1AT$, and control expressing at 48 hours' post transfection: (A) is control HepG2 along with minimal background; (B) $M\alpha_1AT$; (C) $Z\alpha_1AT$.

Vita

Kasey Estenson studied biochemistry at the University of Tennessee, in Knoxville, Tennessee and graduated with a BS in biochemistry, molecular and cellular biology, May 2012. In June 2012, she joined the Ph.D. program Genome Science and Technology at the University of Tennessee, Knoxville with a major in Life Sciences. During her Ph.D., she performed research at the University of Tennessee Medical Center in Knoxville under the supervision of Dr. Valerie M. Berthelieir. She also performed research the Biosciences Division at the Oak Ridge National Laboratory under the supervision of Dr. Jennifer M. Morrell-Falvey. During this time, she published on research paper as the second author and another research paper is being submitted where she is the first author. Her research interest includes protein misfolding disorders and microbial interface biology.