THESIS

ODOR ENCODER: COMPUTATIONAL DESIGN OF A NOVEL ALLOSTERIC ENZYME ACTIVATION SYSTEM FOR PROVIDING ENHANCED OLFACTORY ABILITIES TO TRAINED ODOR DETECTING SENTINEL ANIMALS

Submitted by

Michael Scroggins

School of Biomedical Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2022

Master's Committee:

Advisor: Christopher Snow
Co-Advisor: Christie Peebles

Claudia Gentry-Weeks

ABSTRACT

ODOR ENCODER: COMPUTATIONAL DESIGN OF A NOVEL ALLOSTERIC ENZYME ACTIVATION SYSTEM FOR PROVIDING ENHANCED OLFACTORY ABILITIES TO TRAINED ODOR DETECTING SENTINEL ANIMALS

From the perfume of a flower, to the aroma of a favorite food, to what for bioengineers is the all-to-familiar smell of E. coli, olfactory senses play in important role in how animals interact with the world around them. An offensive odor can inform us that an object is unsafe to eat or be around, a familiar scent can recall memories of events from decades in our past, and even our natural body odors can affect our mating selection preferences. Yet there are many chemicals, both natural and synthetic, for which we do not possess the ability for olfactory detection. An everyday example of this is the natural gas that we use in our homes and which is naturally odorless, but which is commonly spiked with the odorant tert-butyl mercaptan (TBM) to provide the characteristic sulfuric smell we associate with natural gas. Because of this added odorant we can rapidly detect a leaking gas via the smell of the TBM and address the situation as needed to ensure the safety of ourselves and our community. Unfortunately, there are some hazardous and odorless chemicals which we cannot simply spike with an odorant molecule, and for these situations it would be ideal to have alternative options for facilitating a rapid olfactory detection. Therein lies the goals of the Odor Encoder project; to create enhanced olfactory abilities via a conditionally activated enzyme which produces a smellable product in the presence of a target odorless molecule. The approach to achieving this goal was creation of a genetically modified bacterial organism which could be engineered for conditional expression of an odorant producing

enzyme in-situ within the nasal microbiome of trained odor detecting animals. The odorant producing enzyme chosen for this purpose was salicylic acid methyltransferase, a.k.a SAMT, which produces the characteristic odorant molecule methyl salicylate via methylation of salicylic acid. The probiotic E. coli strain Nissle 1917 was selected as the bacterial organism for inoculation of the nasal microbiome, and an expression plasmid was created which could produce both salicylic acid and methyl salicylate from endogenously produced metabolites via dual expression of SAMT along with a salicylate synthase enzyme known as irp9. Conditional production of methyl salicylate was achieved via two methods. The first method involved conditional enzyme expression via use of a riboswitch specific to the small molecule theophylline. The second method involved conditional enzyme activity via constitutive expression of a crippled form of SAMT which may potentially have its enzymatic activity restored via theophylline induced allosteric activation. The allosteric rescue method utilized computational design methods to design novel theophylline-specific allosteric cavities in SAMT, and theophylline induced allosteric reactivation of enzyme activity will be investigated via production and screening of the computationally designed enzyme library.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Christopher Snow, for giving me the opportunity to work in his lab and contribute to the greater scientific community. Dr. Snow's energetic enthusiasm for all things bioengineering is impressive, admirable, inspiring. His energy and enthusiasm seem truly limitless, and his passion for bioengineering infallibly passes on to his students and creates not only intelligent and effective scientists, but also individuals who are truly passionate about what they do and why they do it. I appreciate him giving me the chance to join his lab and learn both bioengineering and protein design and learn about and develop myself as a human in the process. I also appreciate his endless patience with, and acceptance of, myself as we traversed the journey of science together. In a project that was anything but straightforward, during a time in which society was experiencing significant global turmoil, Dr. Snow's positive attitude and accepting nature provided stability and motivation at a time it was truly needed. I am grateful for all that I have learned from him academically, professionally, and personally, and I look forward to continuing mentorship from him as I move forward with my scientific career.

I would like to thank my co-advisor Dr. Christie Peebles for guiding me through this project and being an amazing mentor, friend, and "lab-Mom". The first year I was on this project Dr. Snow was out of State, and upon his return we were under Covid quarantine. As such, Dr. Peebles was the one who provided my initial mentoring and introduction to the lab and the overall project and was instrumental in helping me develop my initial laboratory skills and general scientific mentality. Much of this project involved molecular biology rather than computational design, and Dr. Peebles was essential to my successful development and

application of the necessary molecular biology techniques. Dr. Peebles is down-to-earth and realistic in her expectations of her graduate students, but more importantly, she truly desires for her graduate students to grow and develop as both scientists and as human beings. Because of that approach, Dr. Peebles allows her students to truly develop into effective scientists and not just glorified lab techs. On a personal level, Dr. Peebles has shown a wonderful acceptance of myself and my unique and sometimes confusing personality and has shown she cares about myself as an individual and not just an employee. I would like to particularly note Dr. Peebles' kind, conscientious, and understanding response to the passing of my dog which occurred during this project, and which greatly affected me mentally and emotionally. Dr. Peebles showed a particular understanding sympathy of how this affected me and gave me the time and space to focus on myself during this difficult time, as well as showing sincere personal understanding and sympathy for my loss. I am incredibly grateful for Dr. Peebles' mentorship both academically and personally, and for all she has done to help me succeed on this project and in my graduate studies.

I would like to thank Dr. Claudia Gentry-Weeks for her role as the third advisor on my graduate committee. I appreciate her willingness to join my committee last minute due to an unexpected personnel change, and her taking the time and efforts to help with the project with guidance and technical assistance. I would also like to thank Dr. Gentry-Weeks for her lab's assistance in providing purified plasmids from organisms our lab was not permitted to work with. I also would like to thank Dr. Susan DeLong who was originally a member of the graduate committee but who had to unfortunately be replaced due to a change to departmental requirements regarding outside advisor positions. While she was ultimately unable to be a

committee member, her willingness to be one and her technical assistance and guidance while she was on the committee was greatly appreciated.

I would like to thank Dr. Jessica Ayers, Dr. Lon Kendall, Sarah Tan, and other staff of the CSU Laboratory Animal Research center (LAR) for their assistance in performing the animal related portions of the project. I would also like to thank the staff of the CSU sequencing core for their assistance in NGS sequencing of the nasal microbiome.

I would like to thank Dr. Kevin Morey for his assistance and mentorship on this project and in the lab in general. Kevin's experience in molecular biology, and his patience and willingness to share these experiences in order to teach me essential lab techniques and concepts was greatly appreciated. Kevin also helped develop and perform some of the more complex molecular biology tasks required for this project, and by lending his experience and skill to the project helped us be successful and produce results in timely manners.

I would like to thank Dr. Jacob Sebesta for his mentorship during the early stages of the project and my early days in the laboratory. Jake was my primary mentor during the early stages of the project and helped me learn many of the basic techniques which were essential to the project success. More importantly, Jake was incredibly kind, supportive, and generous with his time and efforts when it came to teaching me (and sometimes reteaching me) essential lab techniques, lab etiquette, and how to use most of the major lab equipment. This was particularly relevant during the stages of this project which occurred under Covid quarantine, and which resulted in a significantly limited lab capacity. During this time, Jacob spent hours on the phone talking me through countless techniques and protocols, and without this help there is literally no way I would have been able to be successful on this project. Without Jake's mentorship and friendship, I would not be writing the thesis you are reading today. And this includes a generous

continued assistance and mentorship after his graduation from CSU, for which I am continually grateful.

I would like to thank Jacob DeRoo for his assistance on the computational components of this project. By lending his computational experiences to this project, Jacob allowed me to focus on the theoretical principles of the enzyme design without having to then translate those thoughts into functional Python code. While I do have basic skills with Python, the learning curve necessary to create scripts as functional and effective as those created by Jacob would have required significant investments of time and effort and would have greatly delayed the overall completion of the project. Jacob also contributes to most of the projects in the Snow lab, as well as a handful of other lab groups on campus, and due to the high number of commitments Jacob has highly limited capacity to assist with projects. Despite this, Jacob invested significant amounts of time and effort into this project, including talking me through many concepts which were essential for my understanding of the computational methods used in the project. I am grateful that Jacob committed his time and efforts to this project and did so without complaint and with a smile on his face.

I would like to thank Ethan Shields for his assistance in many of the molecular biology components of the project. Ethan started the project as an undergraduate research assistant and transitioned into a full-time graduate student contributor. His time and efforts were essential to completing the necessary molecular biology tasks required for this project in a timely manner. He also took the lead when it came to development of the expression system which is to eventually be utilized in the animal trainings and took lead on communication with the animal trainers in regard to this component of the project. I would also like to thank Callie Slaughter for her assistance in the genetic sequencing for the nasal microbiome for the project, her work with

developing this system for the rats, and her general contributions to the project. I would also like to thank Michael Burns for his running of the rat training team and his generally impressive ability to independently produce results and lead his team.

I would like to acknowledge the funding this project received from the United States Department of the Navy. I would also like to acknowledge Colorado State University and the other occupants of the Susanne and Walter Scott Jr. Biomedical Sciences research facilities for their cooperation with shared spaces and resources, as well as any cursor contributions which may have arisen in the course of research discussions among peers.

I would like to thank Dr. Mu Feng of the Albert Einstein College of Medicine and Dr. Emmanuel Burgos formerly of the Albert Einstein College of Medicine for providing a sample of TM0936 deaminase free of charge. While this enzyme ultimately was not utilized in the project due to alternative methods being developed, the generosity of Dr. Feng and Dr. Burgos was greatly appreciated.

DEDICATION

I would like to dedicate this thesis to my parents Dave and Michelle in thanks for their support throughout my life and academic career. Without their help and support in so many aspects, there is no way I would be where I am today or able to pursue higher education with the freedom with which I have. I want to particularly thank my father Dave for being a source of guidance and stability over the course of my graduate program and my life overall. The time of my graduate schooling coincided with a fair amount of familial chaos through which my father remained a constant source of strength. This is just one of the many reasons I admire him and hope that my work towards this thesis is both entertaining to him and makes him proud.

I would also like to dedicate this thesis to my Grateful Dead/Furthur Family. In a world were conformity and normality are demanded and reinforced by our societal peers, my Dead family have instead accepted and loved me despite my lack of conformity or normality or general sanity. Growing up I never felt like I belonged, and even today I rarely fit in with normal society. When I met my Dead family, I found a group of humans who not only accepted me for my unique and eccentric personality, but actually appreciated those parts of me which I find so important to my sense of self identity and worth. Without their love and support, I would not be here today to participate in the scientific process, nor would I have the love for life and acceptance of self that I do. The family is large and always growing, but those whom deserve particular note are the Laffies (Andrew, Melissa, Quinn, Stella, Morgan, and Brian), the Deglers (Chris “Rockin’ Dad”, Lindsay, and Zack), Brent “Ch1naCat” DeLaurentis, Chris Poehler, Zach Zidell, Yano “Grandpa Tie-Dye” Harris, Mamma Stacey and Pappa Chris Ulep, Mike Townsley, Bill Carrol, Bobby Chance, Cabot Dimick and Poppy Rockney-Finger, and my dear friend

Jaimie Tosto. A special dedication goes out to the late Dierdre Carrol and Michael Simmons whose lights still shine bright in our hearts and minds despite their departure from these physical forms.

I would also like to dedicate this thesis to the members of the Grateful Dead/Furthur/Dead and Company, and the umbrella of bands which have developed under their guidance. The first dedication goes to musical members of the actual Grateful Dead; Jerry Garcia, Bob Weir, Phil Lesh, Bill Kreutzmann, Ron "Pigpen" McKernan, Mickey Hart, Keith and Donna Godchaux, Brent Mydland, Vince Welnick, and unofficially Tom Constantine and Bruce Hornsby. In one form or another, all of these musicians have contributed to the music which is the soundtrack to my life and my form of religious hymns. Particularly Jerry and Bobby for their elevated contributions to the song composition process. Separately a dedication goes to Mr. Robert Hunter and Mr. John Perry-Barlow for their lyrical contributions to the music. They wrote the words which told the stories which truly made the music of the Grateful Dead stand out from the rest, and their words have played the role of scripture in my life as they have provided guidance, solace, and encouragement throughout my journey down the ever changing river that is life. A separate dedication goes to those who have been particularly important in maintaining the music after the passing of Jerry; John Kadlecik and Joe Russo of Further, John Mayer and Oteil Burbridge of Dead and Company, and Jeff Chimenti and Jay Lane of both Further and Dead and Company, all of whom have played integral roles in the post-Jerry incarnations of the Grateful Dead and which have major formative elements of my life. In addition, a special dedication to the members of Dark Star Orchestra (DSO) for their "deadication" to keeping the music alive. While there are many national circuit dead cover bands

out there, DSO were the first and the best to really recreate the magic, and they continue to do so to this day. Not Fade Away.

The final dedication goes to the victims of the Sars-CoV2 pandemic and the frontline workers and researchers who took on incredible personal risk in their work towards ending the pandemic. The fact that this research thesis was produced over the course of a major global pandemic which caused indescribable changes to the world as we knew it made the entire process of research and graduate education unique and challenging to say the least. However, despite the challenges the pandemic posed for myself and the world in general, these challenges in no way compare to the experiences of those who passed away or lost loved ones to Sars-CoV2. So this thesis is also dedicated to those who lost their lives to Sars-CoV2 as well as those who personally survived but lost loved ones.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND PROJECT MOTIVATIONS

Project Motivation

Olfactory senses are one of the primary ways for animals to detect, perceive, and interact with the world around them. Odors can be used to determine what is likely edible or inedible, can influence mating choices, be used to identify familiar individuals and locations, and can signal the presence of possible harmful compounds in the environment or the presence of possible food items [1,2,16,18]. Olfactory detection occurs when an odorant molecule binds to the olfactory receptors in an animal's nose, signaling to the brain that the specific molecule is present and in what quantity. For a molecule to be detected via olfaction it must have a corresponding olfactory receptor, with each unique mixture of odorants having a unique characteristic scent based on the combination of receptors activated [2]. This ability to distinguish between unique mixtures of odorant compounds is why a molecule such as indole can contribute to diverse odors ranging from the pleasant scent of jasmine flowers to the repulsive scents of feces and rotting flesh [16,17]. However, the olfactory abilities of different species vary based on differences in the number and types of olfactory receptors, with humans in particular having relatively limited olfactory abilities when compared to many other animal species due to the loss of many human olfactory receptor genes due to a combination of mutation and a lack of evolutionary pressure for maintenance of that gene – a phenomenon known as pseudogenization [2, 4]. While the natural range of human olfaction may have been sufficient for navigation in the natural environment, due to modern artificial manipulation of our environment, particularly with chemicals which are olfactorily indetectable, problems can arise due to our natural olfactory limitations.

As olfaction is primarily an investigatory sense, an olfactory limitation can lead to the inability to detect potentially hazardous or harmful chemicals. Sometimes these limitations can arise due to desensitization to an odorant upon repeated exposure, or loss of olfactory abilities due to injury or biological infection, but often olfactory limitations arise from a simple lack of olfactory receptors for that specific chemical [2]. Common examples of these types of odorless hazards include volatile gasses such as the explosive alkane gasses used as domestic and industrial fuel sources, or the lethally toxic carbon monoxide (CO) which is a byproduct of burning these fuels [19,20]. For hazards such as these, we have developed alternative methods of detection such as electronic detection for CO, or the spiking of alkane gasses with volatile compounds with low thresholds of human olfaction such as methyl mercaptan or other sulfuric compounds [21]. However, there are many chemicals of concern for which these methods are, for one reason or another, not feasible methods of alternative detection.

In cases where artificial detection methods are not feasible, humans will often utilize the greater natural olfactory abilities of animal species such as canines and rodents to effectively increase human olfactory detection abilities. These animal sentinel species have both a greater number of possible detectable odorants, as well as increased sensitivity to odorants which may be found in small concentrations. This allows detection of chemicals such as drugs and explosives, many of which are indetectable to the human olfactory system, or for which human exposure would pose too great of a threat. However, these animal sentinels also possess natural limitations to their olfactory abilities, either due to inability to olfactorily detect the compound or due to toxicity issues with the compound of interest [2]. To overcome this limitation, this project aims to enhance the ranges of natural olfactory abilities via allosteric activation of an odor producing enzyme via binding of an analyte with a less intrinsic odor. Deployment of this system

in either humans or animals would theoretically allow tunable olfactory detection systems for chemicals for which normal olfactory detection is not feasible.

Olfactory Detection Mechanism and Limitations

Olfactory detection itself can be characterized as a fairly straightforward mechanism in which the binding of an odorant molecule to an olfactory receptor (OR) induces nerve transduction from the olfactory bulb to the central nervous system, thereby communicating to the brain the presence of that given molecule. So long as a molecule is able to enter the nasal passage, dissolve into the mucous which coats the olfactory epithelium, and bind to an OR, then that molecule should be able to be detected via olfaction. However, this signaling system is made more complex due to the fact that a given odorant molecule can interact with multiple different types of ORs, as well as given ORs having various affinities for various different odorant molecules, which leads to a complex combinatorial signaling process in which a mosaic of neural signals combine to form the final olfactory detection signal. In addition, within a population there are genetic variations in the ORs themselves, with different alleles leading to discrepancies in olfactory sensitivity [2]. As such, while the overall mechanism remains relatively simple and universal, individual variations in olfactory abilities can make it difficult to predict exactly how a given odorant molecule will be detected by a given individual.

Previous works at elucidating the mechanism of olfactory detection have focused on the structural relationship between a given odorant molecule and the ORs it interacts with. These structural investigations have yielded some interesting insights into the complexity of olfactory detection in regard to structure-odor relations. One insight gained was the fact that a given odor type is not necessarily specific to a given molecular structure, with evidence that four relatively different types of molecules can all produce a similar “musky” smell despite these molecules

having drastically different structures from one another. On the other hand, it has also been shown that very minor structural differences can lead to dramatically different olfactory perceptions, with the R/S-enantiomers of many molecules producing highly contrasting olfactory responses. A good example of this would be 2-methyl-butanoic acid, which has either a sweet and fruity odor in its S-enantiomer form, and a “cheesy” or “sweaty” odor when in its R-enantiomeric form [2]. These observations demonstrate the complexity of structure-odor relationships in the olfactory detection system, and showcase how olfactory detection can produce such a wide range of sensations in response to olfactory stimuli.

Another aspect of olfactory detection which leads to variation in olfactory abilities is the genetic variation in ORs that occurs between different individuals of a species, as well as between different species. Because ORs are highly specific to certain odorant molecules, genetic variations in these ORs can lead to variations in both the type and intensity of the resultant odor detected. For example, the odorant molecule androstenone typically has a smell described as similar to animals or urine; however, for some individuals with mutations to the OR7D4 receptor, androstenone has a sweet smell described as reminiscent of honey and vanilla. Another mutation to the same OR7D4 receptor can induce a greatly increased sensitivity to androstenone, demonstrating how variations in OR genes can lead to different olfactory sensations for the same odorant molecule [2]. On the opposite side of this spectrum are cases of genetic linked anosmia such as the X-linked genetic inability to smell the toxic compound hydrogen cyanide, or differences in detection thresholds for the compound cis-3-hexanol that are related to polymorphisms within the OR2J3 gene [2,3]. These examples demonstrate the important relationship between the structure-odor relationship active in olfactory detection, and how even

small changes in the physical structure of an OR can result in drastic changes in olfactory detection abilities.

This relationship between OR genetic variability and olfactory detection abilities is also the primary reason for the differences in olfactory abilities between species. Genomic investigations have shown that human have approximately 600-900 OR genes, while canines, mice, and rats have approximately 1100, 1300, and 1600 OR genes respectively. In addition, humans have a much higher rate of pseudogenization that occurred within their OR genomes. The dramatic differences seen between humans and other mammals is hypothesized to be a result of humans lessening their dependance on olfaction over the course of our evolutionary history. This results in a less variable set of ORs being expressed in humans as compared to other mammalian species, and consequently decreased relative olfactory abilities. While humans may still have the ability to detect certain chemicals, other mammals' increased sensitivity or ability to identify a given compound from a mixture are often due to increased relative OR variety and expression levels [4]. Some common examples of chemicals which canines have been trained to detect are the piperonal in the street drug MDMA, the 2,4-DNT decomposition product of TNT, and the 2-ethyl-1-hexanol in C-4. In addition, canines have been shown to differentiate between the odors of decaying human and swine remains, despite these sources having highly similar volatile chemical profiles [23]. It is for this reason that mammals such as canines have been adapted for the olfactory detection of chemicals which are essentially indetectable to human olfaction. Trained odor scenting canine and other so-called sentinel species are just one of the methods that humans have developed to overcome our own natural olfactory limitations.

The use of chemical smelling mammals is one of the more common and widely applicable methods of enhancing human olfactory abilities due to the fact that if an odorant can

be detected by an animal, and that animal can be trained to demonstrate that detection, that molecule can then be readily detected without having to engineer any novel devices or biological expression systems. However, even sentinel species have their olfactory limitations due to either the lack of ORs to detect a given molecule, low vapor pressure of a compound preventing it from entering the nasal cavity in appreciable concentrations, or due to possible toxicity thresholds that are lower than the olfactory detection thresholds [4,5]. For example, there have been reports of police canines being exposed to synthetic opioids during drug searches that were at a low enough level the canines did not detect the drugs, but were still in significant enough concentrations to induce a life-threatening overdose [5]. Because of these limitations, other methods to enhance human olfactory detection abilities have been developed which do not rely on sentinel species.

The most common examples of these artificial olfactory enhancers would be the spiking of natural gas with highly olfactory detectable sulfuric compounds such as ethyl mercaptan in order to allow humans to detect the otherwise odorless alkane gases, or the use of electronic detector/signal devices such as the carbon monoxide detectors found in our homes. However, even these methods have limitations, as spiking with highly olfactorily detectable volatile molecules is only viable for other highly volatile compounds, and electronic detection methods also require high levels of airborne particles to reach a detection threshold [6]. It is because of these limits that sentinel species are advantageous when attempting to detect low volatility compounds due to their ability to draw these low-to-non volatile compounds into their nasal passage via active inhalation or insufflation. However, as previously mentioned, this is only an effective method of detection if the animal used for detection has ORs which can bind the molecules in question.

In order to overcome the natural limitations in sentinel species olfaction, it would be useful to combine natural animal olfactory abilities with the concept of spiking with a highly detectable volatile compound. In the same way we provide humans with a method for smelling an odorless compound by adding in a volatile odorant, by ensuring an olfactorily detectable molecule is present whenever our target odorless molecule is present we can provide sentinel species with an alternative molecule of detection for which they already possess the necessary ORs. By creating a biological expression system which will produce an odorant molecule when in the presence of a target odorless molecule, we can in essence spike the nasal cavity of our sentinel animals with an olfactorily detectable odorant molecule. This method would overcome the natural limitations of olfaction which arise due to a lack of molecule specific ORs, as theoretically any odorless molecule for which an activation system could be designed could induce downstream production of an olfactorily detectable molecule. Because of this wide ranging possible applications, it is this approach which we decided to use in our attempts at creating an enhanced olfactory system for sentinel species.

Theophylline Background

To test the validity of this concept, we needed to demonstrate that we could design a system which could produce an odorant in the presence of an odorless molecule, and thus it was necessary to select an appropriate odorless molecule for activation of this system. The odorless molecule that was selected for this purpose was a methylxanthine known as theophylline which is structurally and chemically related to caffeine and other dietary methylxanthines. Theophylline has also been used for over 70 years in the treatment of COPD, asthma, and other bronchial inflammatory conditions, although it's use has decreased overtime as more effective substitutes have been discovered. Because of this long history of medicinal use in humans, theophylline has

an established history of relative safety and low toxicity risks when inhaled [11]. This was deemed particularly important since our enhanced olfactory system was designed to be implemented in the nasal cavity of sentinel species, and as such the odorless molecule that is activating the system will necessarily be inhaled by the sentinel animal. By selecting a molecule that was known to be not only safe at low doses, but which was already in use as an inhaled medication which had been tested and confirmed as safe for human use, we were able to have greater confidence in the safety of our design for animal trials.

Another reason that theophylline was considered an attractive choice for our odorless activator molecule was due to the fact that there are preexisting riboswitches which have been designed for activation by theophylline. A theophylline inducible riboswitch has an mRNA structure such that the ribosome will not bind the RBS on the mRNA until theophylline binds to the mRNA theophylline riboswitch (Fig.1). This process ensures that proteins under control of the riboswitch will only have been induced when theophylline is present in sufficient concentrations, and has been shown to be highly effective at controlling gene expression with minimal levels of background expression when compared to the more common but leaky induction systems such as the *lac* or *tet* expression systems. When the theophylline riboswitch was utilized to control RFP production, no RFP was detected when theophylline was not present, and at any given level of IPTG induction, RFP expression occurred on a dose-dependent manner with theophylline induction, with roughly twice the RFP expressed with 1.0mM theophylline induction as was detected with 0.5mM theophylline induction. In addition, the theophylline riboswitch has been shown to drastically decrease overall expression, with expression levels approximately 2.3% of those seen for identical gene expression systems lacking a riboswitch component [12]. Thus, a theophylline induced riboswitch can result in a highly regulated low-

level expression system, something which would be ideal for producing small concentrations of a potent odorant molecule in the nasal cavity of an animal sentinel species.

Working with a theophylline induced riboswitch allows for design of a biological expression system which would express a specific enzyme when exposed to the odorless molecule theophylline. If this enzyme produces an odorant as its product, then activation of this enzyme in the nasal cavity of a sentinel animal would result in the enhanced olfactory abilities the project is aiming to achieve. However, one downside of the riboswitch method of enzyme expression and activation is that there is necessarily a lag time between the activation of the riboswitch by theophylline and ultimate expression and activation of the enzyme. When considering the project applications, this would mean our sentinel species would experience a delay between exposure to the odorless molecule and the delayed detection of an odor. In addition, sustained exposure to an analyte is only consistent with certain applications which could cause the presence of the odorant to be too transitory for easy detection. Alternatively, a sentinel animal may be required to screen multiple samples in relatively short succession in which case delayed detection could confound screening results. As such, while the slower-activated riboswitch method is not without benefit, it does make a more rapid activation system desirable. However; despite the somewhat delayed nature of riboswitch facilitated expression regulation, this does not preclude the use of the riboswitch method as a viable method for achieving the project goals of controlled in-situ enzyme expression.

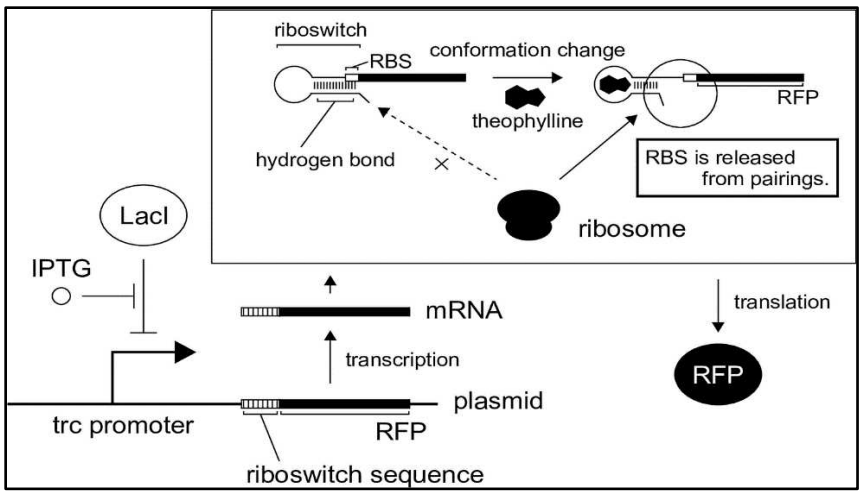


Figure 1. Schematic of the theophylline riboswitch as originally depicted by Kamiura et al. In this schematic, IPTG is used to remove LacI repression of the trc promoter, resulting in transcription of the theophylline-inducible riboswitch mRNA for red fluorescent protein (RFP). In the absence of theophylline, the secondary structure of the mRNA prevents the ribosome from binding the ribosome binding site (RBS) and consequently prevents RFP expression. Upon binding of theophylline to the riboswitch, this secondary structure is modulated such that the RBS becomes available for ribosome binding and the subsequent expression of RFP [12].

A final advantage to using theophylline as our odorless activator molecule is the cyclic nature of theophylline and its relative structural similarity to the cyclic amino acids. Specifically, it is the fact that theophylline has a similar size to some of the larger amino acids, a well as possessing a planar aromatic structure. While the theophylline riboswitch is one of the methods which will be utilized to achieve our stated enhanced olfaction goals via induced expression of our odorant producing enzyme, the other approach used, and the main focus of this thesis, was development of an allosteric activation system which activates the enzyme upon binding of theophylline to a designed allosteric activation site. Because allosteric activation requires binding of the activator into a cavity within the enzyme, having an activator molecule which could readily take the physical space that was previously occupied by a wild-type residue sidechain would theoretically facilitate an easier binding event. For example, it would be much easier to replace a tryptophan side-chain with the theophylline molecule than it would be with a molecule such as carbon monoxide (CO), simply due to the chemical nature of CO in comparison to amino

acid side chains. While theophylline does not have the same level of structural similarity to tryptophan as does a molecule like indole, it still possesses enough of the basic chemical properties which make it a relatively viable replacement for larger amino acids. Figure 2 demonstrates the physical structures of theophylline, tryptophan, indole, and CO for visual comparisons of structural similarities and differences.

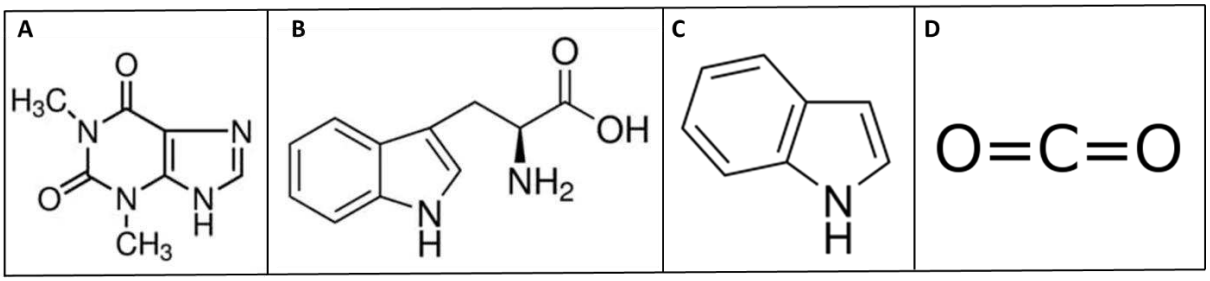


Figure 2. Molecular structures of theophylline (A), tryptophan (B), indole (C), and carbon dioxide (D). Notice how the structure of indole is exactly the same as the tryptophan sidechain, how the structure of theophylline possesses characteristics such as amine groups and a planar cyclic structure which are also found in tryptophan, and how the structure of carbon dioxide is drastically different from the other three molecules depicted.

SAMT Background

When selecting an odorant producing enzyme, considerations to both product odor and preexisting structural knowledge were taken into consideration, and the odorant producing enzyme ultimately selected was salicylic acid methyltransferase (SAMT). This enzyme, originally extracted from the petals of the flowering plant *Clarkia breweri*, is responsible for transferring a methyl group from the coenzyme s-adenosyl methionine (SAM) onto the main substrate salicylic acid (SA) to produce the odorant molecule methyl salicylate (MSA), the molecule responsible for the characteristic odor/flavor known as wintergreen mint (Fig. 3).

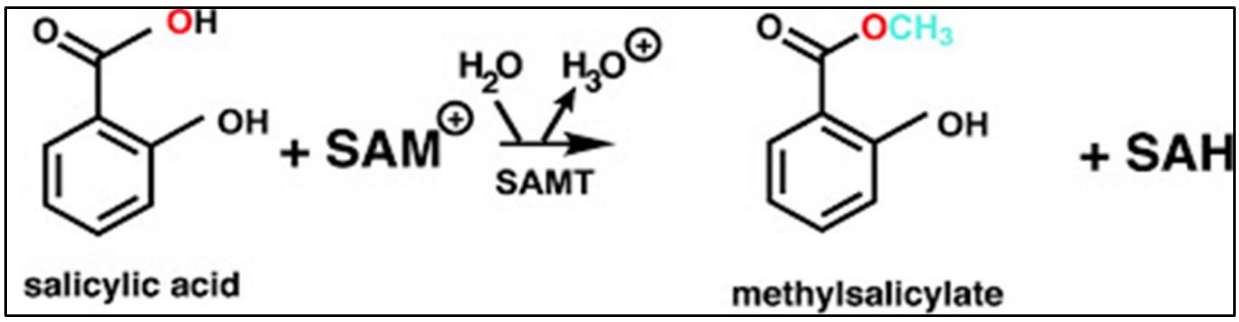


Figure 3: Schematic of the mechanism of action for the SAMT catalyzed methylation of salicylic acid to produce methyl salicylate. S-adenosyl methionine (SAM) is the co-enzyme which acts as the methyl donor, while S-adenosylhomocysteine (SAH) is the byproduct of this reaction. Endogenous pathways recycle SAH back to SAM (pathway not shown). Image originally produced by Zubieta et al [7].

In *C. breweri*, SAMT activity has been associated with increased resistance to pathogens, as well as acting as a chemoattractant for the moth species which are involved in flower pollination [7]. Structurally, SAMT is a homodimeric enzyme which demonstrates high affinity for SA as a substate, but also has moderate methylation activity towards benzoic acid, and limited but measurable activity towards a variety of organic acid substrates. The active site of SAMT is structured such that the methyl group on SAM is positioned approximately 3Å from the carboxylic acid moiety of SA, with this proximity what gives SAMT its high affinity and specificity for SA as a substrate (Fig. 4).

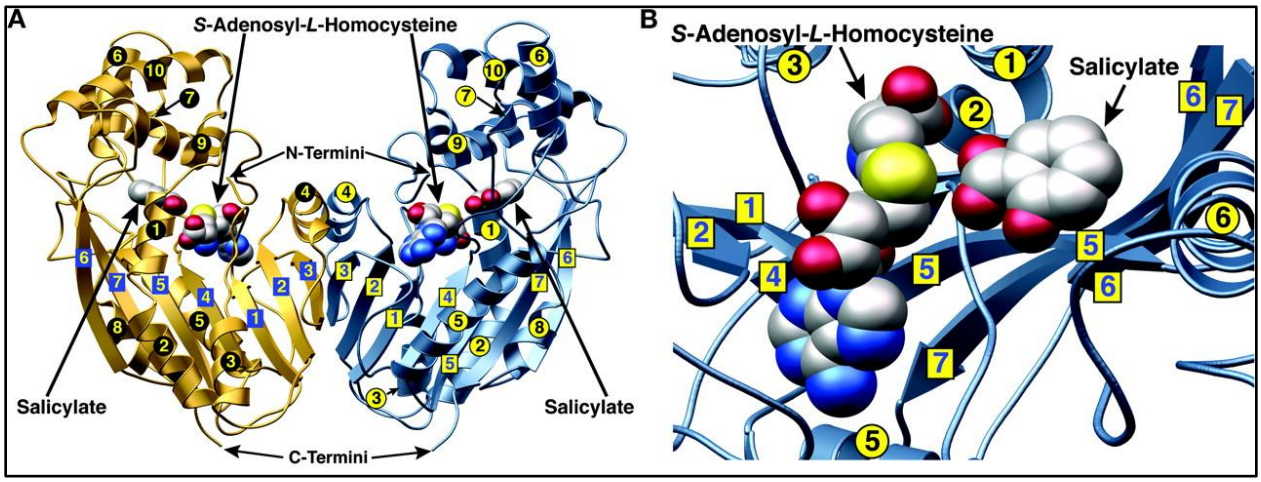


Figure 4: Cartoon depiction of the SAMT crystal structure and active site. **A)** SAMT homodimer with SA and SAH bound in the active site of each monomer. **B)** Close-up of one SAMT active site. Visible is the 3 Å proximity between the methyl group of SAH (yellow) and the carboxylic acid moiety of SA (red) which allows effective catalysis of SA methylation. Both images originally produced by Zubeita et al [7].

SAMT is just one of many different methyl transferase enzymes involved in the production of similar molecules which play similar myriad roles in plant metabolism, communication, and defense such as methyl jasmonate, methyl benzoate, and caffeine. While chemicals like caffeine are not known for their odor, the methyl ester class of compounds are often associated with specific odors due to their highly volatile chemical natures [7]. For example, cis-jasmone – an ester associated with the scent of green tea and jasmine flower – has been found to be one of the major components responsible for the attraction of the New Zealand flower thrip to the scent of Japanese honeysuckle [24,25]. Other examples include certain plant species use of methyl jasmonate and methyl salicylate as airborn communication molecules of pathogenic threats, or the highly diverse impact environmental acetate esters have on the communication and general behavior of the ant species *Solenopsis invicta* [26,27]. In general, esters are known for being highly volatile odiferous molecules which are often utilized for their unique flavor and scent profiles. MSA itself has a relatively low vapor pressure of 0.0343 mmHg at 25°C [28]. It is this highly volatile nature, coupled with a unique characteristic odor, which makes an ester such as MSA a favorable odorant for use in our engineered enhanced olfactory system.

Another benefit to the selecting MSA as our odorant is the fact that MSA can be produced from the central metabolite chorismate using only two enzymes, SAMT and irp9 respectively. Chorismate is an endogenously produced metabolic precursor in the production of aromatic amino acids as well as a variety of other biologically important molecules including coenzymes and co-enzymes [29]. The enzyme irp9 originates in the pathogenic species *Yersinia enterocolitica* and is common to many *Yersinia* species, including the species *Y. pestis* which was responsible for the bubonic plague. However; irp9 is not involved in the direct production of

pathogenic proteins, but instead is involved in a single-step conversion of chorismate to SA. This is preferable to the more common two-step process which occurs in *Pseudomonas* species of bacteria because it allows ready production of SA from a single highly available endogenous precursor (Fig. 5) [30]. Inclusion of the irp9 gene in the SAMT expression circuit will allow both irp9 and SAMT to be expressed simultaneously, and ultimately allow production of MSA without artificial supplementation of non-endogenous substrates.

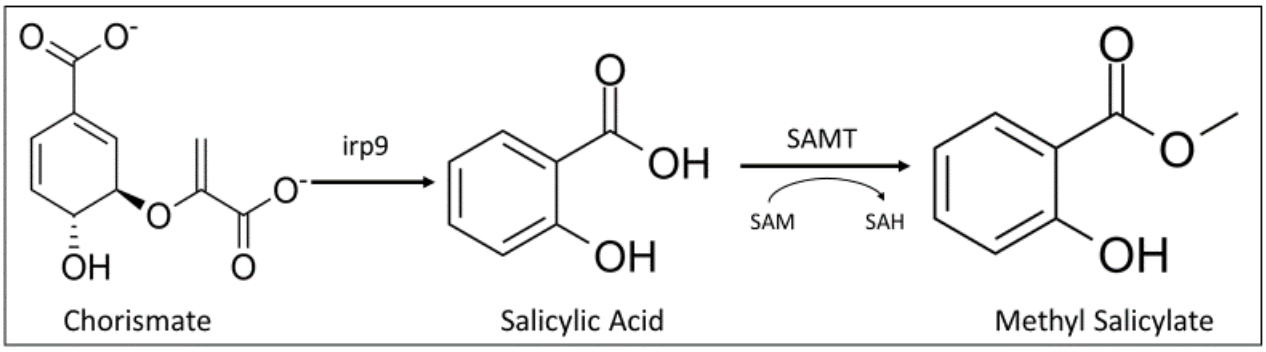


Figure 5. Enzymatic pathway from chorismate to methyl salicylate via the action of irp9 and SAMT respectively. This facilitates the production of MSA from the endogenous chorismate precursor via a two-step enzymatic process.

While SAMT was originally isolated from *C. breweri,* naturally derived MSA is often extracted from the wintergreen plant (*Gaultheria procumbents*), and in low concentrated amounts is commonly used as both a flavoring and perfumery ingredient, often being associated with the smell/flavor of wintergreen mint [8,9]. Concentrations of MSA found in common foods range from 27ppm in iced creams to 8400ppm in wintergreen flavored chewing gum [10]. Due to these uses, we know that at relatively low concentrations, MSA is safe for consumption and can be detected through both olfaction and gustation (taste).

MSA also has medicinal applications as a topically applied anti-inflammatory and analgesic drug due to the conversion of MSA into various salicylate products which have pain relieving effects. These pain-relieving salicylate compounds consist of acetylsalicylate, a.k.a.

aspirin, in addition to other analgesic or anti-inflammatory salicylate compounds. While MSA does has mild skin irritation effects, the resultant muscular pain relief is sufficient to make MSA topical application a viable method for treatment of muscular pain, with 5mL of 98% MSA being roughly equivalent to 21.7 adult aspirin tablets worth of salicylate [8]. While this would be a potentially fatal amount of MSA as a single dose, in appropriate therapeutic volumes MSA should pose no threat upon topical or dermal exposure [10]. As such, methyl salicylate should be a relatively safe compound for low-level expression and exposure in animal nasal cavities, with an LD50 ranging between 887 – 1250 mg/kg in rats, and 2100mg/kg in canines. Despite this relative safety, it should be noted that doses of 600mg/kg have been showed to produce negative effects in canines, and prolonged and/or gastrointestinal exposure to methyl salicylate has been associated with salicylate toxicity [8]. In fact, gastrointestinal exposure is the least safe, with ingestion exceeding 150mg/kg having toxic effects on humans [10]. As such, any expression system for expression of MSA in the nasal cavity must be designed to express in concentrations great enough for olfactory detection but without exceeding the thresholds of toxicity. Fortunately, the low threshold of detection for MSA – determined by Ruth et al. to be 0.62 - 0.87 mg/m3 for human detection – allows for olfactory detection at concentrations below the threshold of toxicity [31]. This is one more reason the highly volatile chemical nature, coupled with the low threshold for olfactory detection, makes MSA such an ideal odorant molecule for our enhanced olfaction system.

Another major benefit of using SAMT as our enzyme design target is the previous elucidation of the crystal structure for the *C. breweri* isoform of SAMT. Determined by Zubieta *et al.* via x-ray crystallography, the crystal structure was determined in the presence of both salicylic acid and the demethylated version of the coenzyme SAM, s-adenosylhomocysteine

(SAH). This structure demonstrates that SAMT is a 41kD protein which exists in solution as a homodimer, and showcases the specific orientation of the SAM/SAH coenzyme in relation to the methylation site on salicylic acid, and the SAMT residues involved in maintaining those binding orientations [7]. This crystal structure is accessible via the online protein databank, accession number 1m6e.

Having a preexisting crystal structure was deemed particularly essential for the computational design component of an allosteric enzyme switch responsive to theophylline. Creation of this allosteric activation system required mutating a cavity within the SAMT enzyme, with the mutation targets chosen based on where in the SAMT structure the theophylline molecule would potentially have space to fit, while simultaneously avoiding residues which were involved in active site binding or coordination. By selecting an enzyme with a preexisting crystal structure, we were able to begin computational design without having to first determine a novel enzyme crystal structure.

Previous SAMT Expression and MSA Production in *E. coli*

Previous work has demonstrated the ability to both express a version of salicylic acid methyltransferase (BSM1) and produce MSA in standard lab strains of *E. coli* utilizing the two-enzyme mechanism which was adapted from *P. fluorescens*. This method utilized this two-step mechanism to produce SA from chorismate, and the BSMT1 benzoic/salicylic acid methyltransferase, to produce significant amounts of MSA in the *E. coli Mach 1* strain. Subsequent codon optimization and precursor supplementation was able to further increase the yields of MSA in comparison to the WT-BSMT1. However, this method did not report any olfactory detection of MSA, with the presence of MSA only confirmed via chromatographic techniques [48]. In addition, this same expression system was replicated during the 2006

International Genetic Engineered Machine (IGEM) competition by the team from MIT, and this team did claim to smell MSA production from their engineered *E. coli*; however, there were no quantitative results reported for this work, and currently the results have not been officially published but simply reported by the IGEM team [49]. As such, there is currently no published work which has explicitly expressed MSA levels at significant enough levels to confirm via human olfactory detection, and our enhanced olfaction project is thus the first known attempt at olfactory detection of *E. coli* produced MSA.

Allostery Background

Because the major linchpin in successful engineering of the enhanced olfactory abilities was the computational design of the theophylline dependent allosteric activation system, it is important to discuss what allosteric regulation is, and why this was the method we opted to use for this project. Allosteric regulation refers to the process in which a enzyme's function is regulated via the binding of a secondary ligand distally to the enzymes substrate-binding active site (Fig.6). Upon binding of this allosteric regulatory ligand, a conformational change will occur in the enzyme such that binding affinity of the substrate ligand becomes either increased or decreased, effects referred to as allosteric activation or inhibition respectively [13]. Allosteric regulation is almost ubiquitous among proteins to some extent, and plays a major role in regulation of many enzymes which have an identical substrate but different biological roles.

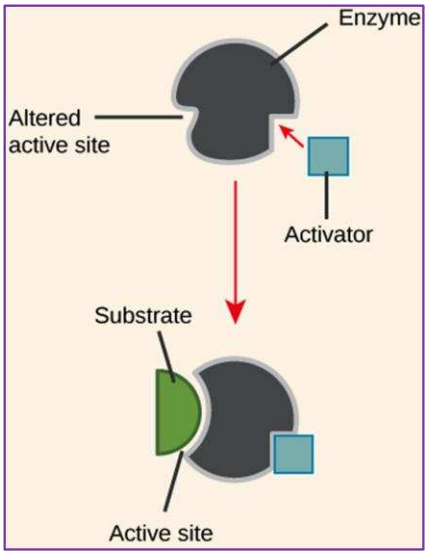


Figure 6. Basic model of allosteric enzyme activation. Prior to the binding of the activator, the active site does not readily bind the substrate. Upon binding of the activator to the allosteric activation site, a conformational change occurs in the active site such that substrate binding and subsequent catalysis can occur. Note how the allosteric binding site is inherently separate from the catalytic active site, a requirement for an activation system to be considered allosteric [15].

Because of the benefits provided by highly tunable enzyme activation, allosteric regulation is commonly utilized when attempting therapeutic drug design. This is done in order to minimize the chances of off-target side effects through unintentional non-specific protein regulation. For example, a cancer drug which is intended to downregulate the activity of a oncogene must not have the effect of unintended downregulation of tumor suppressive genes. By utilizing an allosteric regulator method rather than one which explicitly targets the enzyme's active site, one can design drugs which have highly specific regulatory effects without a high risk of drug toxicity [14]. Having this ability to highly regulate enzyme activity makes allosteric regulation an ideal route of action for attempting design of system which requires conditional activity and would be highly important for a system designed for molecule specific olfactory enhancement.

Some examples of proteins which utilize allosteric regulation to ensure situational activation are the highly ubiquitous membrane bound G-protein coupled receptors (GPCRs) or

ATP dependent kinases, both of which involve binding of fairly ubiquitous co-substrates (GDP/GTP or ATP respectively), and thus are dependent upon allosteric regulation to ensure proper situational activity. If these proteins were allowed to be active whenever their ubiquitous co-substrates were present, then they would almost always be active. By possessing allosteric regulatory mechanisms, these enzymes can ensure they are only active when they are biologically beneficial but can remain dormant when their activity is not needed, even if their substrates and co-substrates are present in significant concentrations [14]. This ability to facilitate highly specific activation of the enzyme allows for highly regulated enzyme control and is very beneficial when it comes to engineering conditionally active biological systems.

In addition to allowing situational enzymatic activation, allosteric activation is also beneficial for the significant increases to enzymatic activity they impart, as well as the speed at which they are able to induce their allosteric activation. For example, the enzyme imidazole glycerol phosphate synthase (HisFH) is an allosterically activated enzyme which has been shown to have a 4500-fold increase in catalytic activity upon allosteric activation. The conformational change for HisFH from inactive to active upon allosteric activation occurs at a rate of 5 s^{-1} , while the transition back from active to inactive occurs at a rate of 9 s^{-1}. This is significantly faster than the rate of turnover for the WT version of HisFH of 0.2 s^{-1}, demonstrating that the rate of allosteric activation is not the rate limiting step in the overall enzymatic mechanism [45]. This rate of allosteric activation can also be modulated via targeted alterations in either the allosteric site or the activator/inhibitor molecules. An investigation into allosteric activation of 3-phosphoinositide-dependent kinase 1 (PDK1) by variations of its regular allosteric activating polypeptides (PIFtides) demonstrated that variations in the sequence of the PIFtides would result in either increased or decreased allosteric activating abilities. All PIFtide variants investigated

demonstrated an ability to allosterically activate PDK1, with one even increasing the fold-activation from 3.9-fold activation to 5.2-fold activation [46]. And with current engineering techniques, allosteric activators can even be modified to fit a desired biological environment. One example of this was the engineering of a version of the clotting protein thrombin which had its allosteric activator molecule changed from Na^+ to K^+. While the resultant mutant still had an allosteric effect of an 18-fold enhancement to enzymatic activity upon binding of Na^+, this was significantly less than the 72-fold activity enhancement which occurred in the K^+ activated mutant [47]. While these examples are all unique to the biochemical mechanisms for each given pair of enzyme and activator(s), they demonstrate the ability of allosteric regulation in general to rapidly and significantly increase the rate of activation and overall enzyme activity for allosterically activated or inhibited enzymes.

Inspiration for this approach came from a previously published engineering of an allosteric switch designed for indole specific allosteric activation of β-glycosidase and β-glucuronidase enzymes. In this work by Dr. Katelyn Deckert and colleagues, a specific tryptophan to glycine mutation in each enzyme, W33G and W492G respectively, resulted in a collapsing of the enzyme active site and a loss of enzymatic activity. Upon subsequent exposure to indole, the enzymes regained their activities in a dose dependent manner, with increased indole concentrations leading to increased enzyme activity in the mutant enzymes. This phenomenon was hypothesized to occur due to restoration of the enzyme active site upon binding of indole to the location where the wild-type tryptophan had once existed (Fig. 7). This would be feasible due to fact the structure of indole is identical to the structure of the tryptophan side chain and thus is able to completely fill the engineered cavity space with a structure which is physically and biochemically identical to the initial tryptophan residue [15]. It was this idea

which was applied to our design approach, in that we attempted to mutate residues in SAMT so that theophylline could bind into the previous location of those residue side-chains in such a way as to fill the cavity space and restore the enzymatic structural integrity. While theophylline does not share an identical structure with any specific amino acid side chain, it does possess enough polar and planar structural aspects that it could theoretically replace naturally occurring amino acids with respect to their role in providing a functional protein structure.

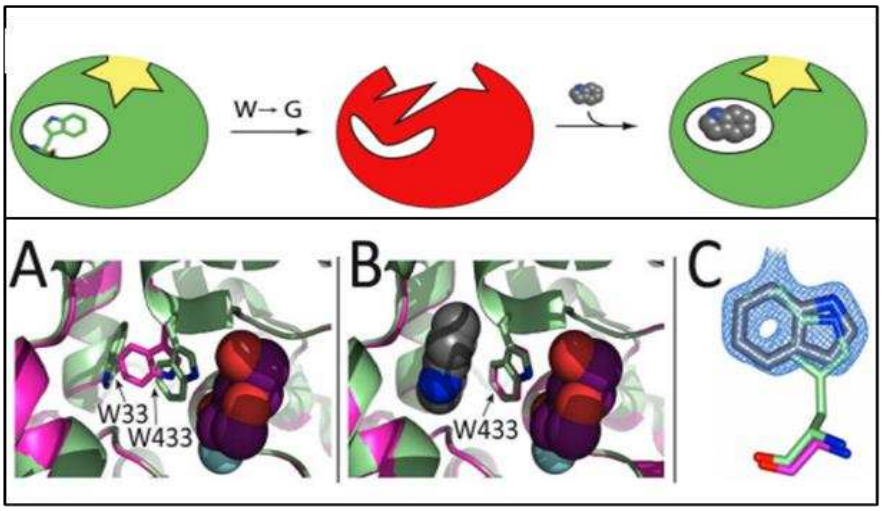


Figure 7. *(Top)* Cartoon depiction of wild-type enzyme with intact Trp residues and functional active site depicted in yellow. After the W33G mutation, a subsequent displacement of W433 causes the active site to collapse and the enzyme to become inactive (red cartoon). Upon exposure to indole, the indole binds to the space previously occupied by the sidechain of W33 within the active site, and in doing so restoring active site structure and function. *(Bottom)* Frame A demonstrates how the W33G mutation causes W433 to shift position and loose contact with the substrate molecule. Frame B shows indole binding in the location of W33, filling the space previously occupied by the sidechain of W33, and consequently restoring the original position of W433 proximal to the substrate to restore enzyme activity. Frame C shows the almost identical nature of the electron densities for indole and the tryptophan side chain, confirming their structural similarities. All images originally produced by K. Deckert et al [15].

Adaptive and Directed Evolution

In the act of engineering an organism which could metabolize SA to MSA, it was deemed important to attempt to create a strain of *E. coli* which had an increased ability to both tolerate and metabolize SA relative to standard lab strains of *E. coli*. Towards that end, basic adaptive and directed evolution techniques were applied in attempts to create these high-SA tolerant *E. coli* strains. Adaptive evolution is simply the process through which a population evolves

through positive selection to better withstand a specific environmental condition, and involves genetic impactors such as genetic drift, immigration/emigration within a population, effective population size, initial frequency of mutation, and an allele's selection coefficient – the coefficient which describes how a given allele type is beneficial towards a given trait while being potentially detrimental against other traits – to determine whether or not a given mutant allele will be preserved or lost within the population [50]. This concept has been adapted in the laboratory to create biological organisms with desired phenotypical and metabolic traits, including increased substrate tolerance, investigations into gene knockouts, and overall increases to desired metabolic abilities [52]. Adaptive evolution is an effective way to induce metabolically advantageous mutations in a protein when there is a lack of knowledge about the protein itself or what regions might be advantageous to target for mutation, as it simply requires the creation of biological conditions which the desired metabolic traits are advantageous and then allows the organisms to "figure it out" on their own. So long as multiple rounds of subsequent culturing and subculturing can be done under the specific biological conditions, eventually it would be expected that alleles which provided advantageous traits towards that given biological condition would be preserved within the population. The result of this is the evolution of organisms which have naturally evolved and preserved the desired metabolic or phenotypical changes. Subsequent genetic sequencing and metabolic investigations can elucidate which genetic changes were associated with which evolved metabolic abilities, and which biomolecules the organisms has either upregulated or downregulated in response to the biological conditions in which it exists [50,52]. However, while the results of adaptive laboratory evolution itself can be quite impactful on the resultant metabolic abilities, this method is limited to natural mutations and evolution with fairly minimal interference by the scientist outside of

creating and maintaining the specific biological conditions and doing the downstream investigations.

In order to increase the effectiveness of evolutionary strain design, directed evolution (DE) techniques can be applied on top of the simpler adaptive evolution techniques. DE is similar to adaptive evolution in that it is a technique for evolving desirable traits in an organism through repeated rounds of mutation and subsequent selection for the desired traits of interest. This is a fairly straightforward iterative process in which the first step is to create a variant library via either targeted or random mutagenesis, followed by a screening and selection step in which library variants with increases in the desired phenotypical or metabolic trait are maintained and utilized as the starting organisms for the next round of mutation and selection. Initially, the results of DE were similar to those of adaptive evolution in that they simply increased the natural metabolic abilities of an organism, usually through the targeting of one specific protein within that metabolic pathway. With recent advancements in biotechnology, DE technique have been applied to entire genomic operons or metabolic pathways, and have even been utilized to create novel biochemical processes which were previously unknown to exist in the natural world. Many of these DE approaches have simply relied on the random mutagenesis process to produce the protein variants which can be maintained through appropriate screening and selection; however, recent applications of computational design to DE has allowed targeted evolution strategies which can use the predictions of computational design programs to help determine where in a protein sequence would be best to target for mutation [51]. This project has primarily utilized the basic adaptive evolutionary techniques in our attempts to create high-SA tolerant strains of *E. coli*, with plans to implement directed evolution as the limits of adaptive laboratory evolution begin to be reached. In addition to attempting to evolve higher SA tolerance

in general, directed evolution techniques will be applied to the larger enzyme library itself, with those library variants that show success at recover of a crippled SAMT being used as the starting templates for future DE of higher activity SAMT variants.

Computational Design Background

To properly design an engineered protein, it is beneficial to supply a known crystal structure to the computational design algorithms due to the large amount of possible variation which occurs upon modelling mutation of even a single residue. Aspects such as residue size, rotational angle, and interactions with neighboring residues are difficult to predict, and each change to one parameter will affect the other parameters in some way, leading to a situation in which there is a combinatorial explosion of possible structures. For example, if you consider one residue and how it energetically interacts with all its proximal atoms, then consider how those interactions could vary if a given residue were to have a different rotational angle or were to be mutated to another residue type, there is a factorial increase in the number of possible result structures for even a single change. Because of this, thorough investigations of theoretically designed protein structures require some form of computational modelling. This is particularly important when attempting to create an allosteric binding cavity that is specific to a specific activator molecule, as this will require mutation of multiple residues, and calculating how the activator binding to the allosteric cavity will affect the overall enzyme structural energy given each possible variation in residue composition. For this project, all design scripts were designed in house by either Dr. Christopher Snow or PhD. candidate Jacob DeRoo.

Overall Project Goal

Due to the natural limitations in human olfactory abilities, coupled with the increasing prevalence of odorless environmental chemical threats, creating a technique for enhancing human olfactory detection abilities is increasingly important for ensuring safe interaction with the potentially hazardous chemicals which we may come into contact with. While some methods of enhanced olfaction, such as spiking natural gas with sulfur compounds or utilizing audible electronic detectors can be utilized directly by humans, detection of other chemical threats necessarily requires the use of animal sentinel species due to their naturally increased sensitivity and perceivability of odorant molecules. This method is in essence enhancing human olfactory abilities through the proxy of the animal sentinel detectors and is one of the most widely applicable methods of enhancing human olfaction, with current limitations being those arising from natural lack of a molecule specific olfactory receptor, or potential toxicity of the compounds being olfactorily detected. By creating a novel biological expression system which can express an odorant producing enzyme and incorporating a molecule specific allosteric activation system to that enzyme, it is possible to increase the natural detection abilities of sentinel animals, and consequently enhance olfactory warning systems. By providing these enhanced olfactory abilities, this project aims to create a system for safer detection of and interaction with potentially hazardous chemicals present in the surrounding environment. This is the first known example of a system which has produced an odorant molecule in situ within the nasal cavity of sentinel animal species, and the first known attempt at creating an allosteric activation system for an enzyme expressed for such purposes.

CHAPTER 2: DETERMINATION OF ENZYME ACTIVITY

Introduction

Due to the fact the project goals rely upon theophylline dependent expression or reactivation of SAMT, it was necessary to determine the basic enzymatic behavior of the wild-type SAMT to use as a baseline for which to compare our engineered systems. For the theophylline dependent riboswitch method of conditional expression, it would be prudent to compare the enzyme activity observed in the theophylline induced strains to the wild-type enzyme activity to determine the ratios of expression levels between the riboswitch-controlled expression system and the constitutively expressed system. Similarly, when investigating the computationally designed mutant SAMT for significant loss of activity, it is prudent to use the activity observed in the constitutively expressed wild type as a baseline of activity for which to compare the activity of the crippled mutants. By that same merit it would be useful to compare the enzymatic activity of the crippled SAMT mutants which do show recovery upon exposure to theophylline to the wild-type activity to numerically compare how effective the allosteric enzymatic restoration effect is. For these reasons, a major component of this project involved the development of *E. coli* SAMT expression strains, and development/validation of methods for which to prove SAMT expression, as well as measure and calculate enzymatic activity in *E. coli* constitutively expressing wild type SAMT, *E. coli* expressing wild type SAMT under control of the theophylline riboswitch, and *E. coli* expressing the computationally designed mutant SAMT both with and without exposure to theophylline.

Materials and Methods

Commercial Gene Procurement

G-blocks containing a theophylline inducible riboswitch-regulated SAMT gene (iSAMT) or a constitutively expressed irp9 gene were purchased from TWIST Bioscience. A G-block containing a constitutively expressed SAMT gene (cSAMT) was purchased from IDT. G-blocks containing gene circuits consisting of irp9, either cSAMT or iSAMT, and GFP were purchased from Genscript.

Chemical Procurement

Methyl Salicylate (MS) was commercially procured from Perfumers Apprentice (Scotts Valley, CA). Miller LB, salicylic acid (SA), and theophylline were commercially procured from Sigma Aldrich. Terrific Broth (TB) was purchased from MP Biomedicals.

Bacterial Strain Procurement

The *dam-/dcm-* strain (dam(-) *E. coli*) was obtained from New England Biolabs, Inc., the E. cloni EXPRESS BL21 (DE3) strain (BL21 *E. coli*) was purchased from VWR, the JW3686 strain was purchased from the Yale University Coli Genetic Stock Center (CGSC), the Nissle strain was generously provided by Dr. Joshua Chan, and the DH5α strain was obtained from preexisting Snow lab *E. coli* stocks [76]. *S. carnosus* TM300 containing pCT20 was obtained from DMSZ [80].

Preparation of Glycerol Stocks

All plasmids which were sequence confirmed were transformed into one or more strains of *E. coli* and glycerol stocked. Glycerol stocks were also made of any relevant plasmid-free strains of *E. coli.* In 14mL round bottomed culture tubes, 5mL of either LB or TB were aliquoted, laced

with antibiotics if appropriate, and inoculated with the organism to be glycerol stocked. Cultures were incubated overnight at 37°C, and stored in 25% (v/v) glycerol at -80°C.

Transformation of Chemical Competent Cells

100uL aliquots of chemical competent cells stored at -80°C were thawed on ice and split into two 50uL aliquots. Purified plasmid DNA was added to one of these 50uL aliquots, with the amount of plasmid DNA varying based on expected transformation efficiency for that plasmid in that strain of *E. coli*. The other 50uL aliquot was maintained free of any foreign DNA sources to be used as a negative control. Tubes were incubated on ice for 30 minutes, subjected to a 30 second heat shock via submergence in a 42°C hot water bath, and allowed to rest once ice for an additional 5 minutes. Post incubation on ice, 950uL of liquid growth media (LB or TB) was added to each tube and tubes were placed into 37°C shaking incubation for 1.5 hours to recover cells. Post recovery incubation, 100uL aliquots of either the experimental transformation reactions or the no-DNA negative controls transformation reactions were plated onto LB/agar without antibiotics (non-selective) and onto LB/agar with the appropriate selective antibiotics for that plasmid and *E. coli* strain. The remaining transformation reaction solutions were centrifuged at 14000rpm at room temp for 1 minute to pellet cells, and supernatants were drawn off to a volume of 100uL. Pelleted cells were then resuspended in this 100uL residual supernatant, and these concentrated transformation reaction solutions were then plated onto LB/agar with antibiotics. All plates were then incubated overnight at 37°C. Subsequent observations of lawn growth on all non-selective plates, a lack of growth on antibiotic selective plates for the no-DNA negative control transformations, and colony growth on the antibiotic selective plates for the

experimental plasmid transformed transformations was considered irrefutable evidence of successful plasmid transformations.

Creation of pSC-b-amp/kan/(AmyP)SAMT plasmid vectors via Blunt End Cloning

Both the cSAMT and iSAMT genes were transferred into pSC-b-amp/kan plasmid backbones using the Agilent Blunt End Cloning kit (Part Number:240207). Gene inserts were PCR amplified using the primers listed in Table 1 below. PCR utilized Phusion polymerase (GoldBio) and reaction cycle was as follows: 5 minutes at 98°C, 35 cycles of 15 seconds at 98°C followed by 30 seconds at 59°C followed by 32 seconds at 72°C, and a 10 minutes final extension at 72°C. PCR products were stored at 4°C prior to use. Primers were commercially produced (IDT). Reaction conditions were as indicated in the kit's provided instructions. Transformants were plated on LB/agar w/ampicillin (100ug/mL) and X-gal (2% v/v), and incubated at 37°C overnight. White transformant colonies were selected and plasmids were purified using the GeneJET Plasmid Miniprep Kit (ThermoFisher). Resultant pSC-b-amp/kan/(AmyP)SAMT plasmids were commercially sequence confirmed (Genewiz).

Table 1. SAMT Blunt End Insert PCR Primers	
cSAMT Forward	*TTGTGAGCGGATAACAATTTGGGCTAACAGGAGGAATTAACC*
cSAMT Reverse	*TTTGGGCTAACAGGAGGAATTAAC*
iSAMT Forward	*TTGTGAGCGGATAACAAGGTACCGGTGATACCAGCATC*
iSAMT Reverse	*GGTACCGGTGATACCAGCATC*

Creation of pSC-b-amp/kan/(PlacZ)SAMT plasmid vectors via SLIM Cloning

Both cSAMT and iSAMT were placed under control of the lac promoter via deletion of the preexisting AmyP promoter using the site-directed ligase independent cloning method (SLIM). Primers for PCR were commercially produced (IDT). Primer sequences are listed in table 2 below. Template DNA was either pSC-b-amp/kan/cSAMT or pSC-b-amp/kan/iSAMT

respectively. PCR reaction cycles consisted of an initial denaturing at 98°C for 30 seconds, followed by 5 cycles of 98°C for 10 seconds, either 68°C (cSAMT) or 69°C (iSAMT) for 30 seconds, and 72°C for 6 minutes, followed by 25 cycles of 98°C for 10 seconds, 71°C for 30 seconds, and 72°C for 6 minutes, followed by a final extension at 72°C for 10 minutes. Resultant PCR products were Dpn1 digested via incubation at 37°C for 4 hours followed by 65°C for 25 minutes. PCR products were then hybridized via incubation at 98°C for 5 minutes, a slow cooling to 38°C over 1 hour, incubation at 65°C for 5 minutes, and final incubation at 30°C for 15 minutes. Resultant hybridization products were transformed into chemically competent JW3686 *E. coli* via a standard heat-shock transformation, with heat shock at 42°C for 30 seconds. Transformants were plated on LB/agar w/100ug/mL ampicillin and incubated overnight at 37°C. Antibiotic selected transformants were harvested, and resultant pSC-b-amp/kan/(PlacZ)SAMT plasmids were purified (GeneJET Plasmid Miniprep Kit) and commercially sequenced confirmed (Genewiz).

Table 2. SLIM Primers	
cSAMT Forward Long	*TTGTGAGCGGATAACAATTTGGGCTAACAGGAGGAATTAACC*
cSAMT Forward Short	*TTTGGGCTAACAGGAGGAATTAAC*
iSAMT Forward Long	*TTGTGAGCGGATAACAAGGTACCGGTGATACCAGCATC*
iSAMT Forward Short	*GGTACCGGTGATACCAGCATC*
Reverse Long	*TTGTTATCCGCTCACAATTCCACAC*
Reverse Short	*TTCCACACAACATACGAGCCGG*

Creation of pSC-b-amp/kan/(PlacZ)cSAMT/mNG Fusion Protein Vectors via HiFi Cloning

HiFi cloning was used to create a gene circuit for constitutive expression of a SAMT-mNG fusion protein containing a Tobacco Etch Virus (TEV) specific cleavable linker between the respective proteins. PCR primers were commercially produced (IDT). Primer sequences are listed in table 3 below. Template DNA for the SAMT gene was pSC-b-amp/kan/(lPacZ)cSAMT,

while template DNA for mNG was pUC19/mNG. PCR consisted of initial denaturing at 98°C for 30 seconds, followed by 30 cycles of 98°C for 10 seconds, either 61.1°C (SAMT) or 67.1°C (mNG) for 30 seconds, and 72°C for 6 minutes, followed by a final extension at 72°C for 10 minutes. Resultant PCR products were Dpn1 digested with incubation at 37°C for 4 hours followed by 65°C for 25 minutes. Hi-Fi reactions were performed at both a 1:1 mass ratio and a 1:2 molar ratio of vector:insert, with the vector being pSC-b-amp/kan/(PlacZ)cSAMT . Reactions were performed using standard commercially produced Hi-Fi cloning mix (NEB), with incubation at 50°C for 1 hour. Resultant Hi-Fi products were transformed into chemically competent DH5α *E. coli* via standard heat shock transformation, with heat shock at 42°C for 30 seconds. Transformants were plated on LB/agar w/100ug/mL ampicillin and incubated overnight at 37°C. Antibiotic selected transformants were harvested, and resultant pSC-b-amp/kan/(PlacZ)SAMT/mNG plasmids were purified (GeneJET Plasmid Miniprep Kit) and commercially sequenced confirmed (Genewiz).

Table 3. HiFi Cloning Primers for Fusion Protein Vector Production	
SAMT Forward	*CATCATCATCACCACCACTAATAACCAGGCATCAAATAAAACGAAAGGCTCAGTCG*
SAMT Reverse	*CTTTCCTTGGAAATACAAGTTTTCATCACTTTTGCGTATTAAAC*
mNG-TEV Forward	*GAAAACTTGTATTTCCAAGGAAAGATGGTGAGCAAGGGCGAGGAGG*
mNG-TEV Reverse	*GATGCCTGGTTATTAGTGGTGGTGATGATGATGCTTGTACAGCTCG*

Creation of SAMT Knockout Plasmid pSC-b-amp/kan/(PlacZ)cSAMT-KO

The Agilent QuikChange Site-Directed Mutagenesis kit (Part Number:200523) was used to introduce mutations M150G & W151G into cSAMT. Template DNA was pSC-b-amp/kan/(PlacZ)cSAMT. Primers were designed using the online Agilent QuikChange primer design software and produced by IDT (Table 4). The QuikChange reaction cycle consisted of 30

seconds at 95°C, followed by 18 cycles of 95°C for 30 seconds, 55°C for 1 minute, and 68°C for 6 minutes. Resultant plasmids were DpnI digested in via incubation at 37°C for 4 hours, and digest products were transformed into kit-provided XL1-Blue supercompetent cells. Transformants were plated on LB/agar w/100ug/mL ampicillin and incubated overnight at 37°C. Antibiotic selected transformants were harvested, and resultant pSC-b-amp/kan/(PlacZ)cSAMT-KO plasmids were purified (GeneJET Plasmid Miniprep Kit) and commercially sequenced confirmed (Plasmidsaurus).

Table 4. QuikChange Mutagenesis Primers	
SAMT_point 2_fwd	*CGTAACACATTACACTTTATTCATTCATCTTATTCTTTAGGGGGGTTATCACAAGTGCCA*
SAMT_point 2_rev	*TGGCACTTGTGATAACCCCCCTAAAGAATAAGATGAATGAATAAAGTGTAATGTGTTAC*

High Performance Liquid Chromatography (HPLC)

Chromatography was performed using the Shimadzu LC-20AB, with a Phenomenex Luna® C18 column, 5um, 250mm x 4.6mm, and with LC solution software used for recording and processing of data. The mobile phase was 85%-orthophosphoric acid, acetonitrile, 18MOm water (2:400:600 V/V/V), and was filter sterilized prior to use (0.2um filter). Separation was achieved under isocratic conditions at a flow rate of 1.0 ml min^{-1}, and the effluent was monitored at 237 nm. Samples were maintained at 10°C, and stationary phase was maintained at 40°C. Run time was 50 min.

HPLC Sample Collection

Cultures had OD600 measurements taken at the point of every sample collection event (Nanodrop 2000C). 1.5mL of culture was centrifuged at 14000rpm at 24°C for 5 minutes (Beckman Coulter Microfuge 22R). Resultant supernatants were immediately syringe filtered

through 0.2um nylon filters (Agilent) and transferred to 1.8mL HPLC sample vials (VWR). Samples were maintained at room temperature until placed into HPLC to prevent salicylic acid from precipitating out of the sample solutions.

Enzyme Expression and Activity Confirmation Assay

50mL solutions of Miller-LB spiked with 100ug/mL carbenicillin and 10ug/mL SA were prepared in 250 flasks and inoculated with either *dam(-) E. coli* w/pSC-b-amp/kan/(PlacZ)mNG or *dam(-) E. coli* w/pSC-b-amp/kan/(PlacZ)cSAMT, or were left non-inoculated (cell-free cultures). Cultures were incubated at 28°C while shaking at 215rpm for 24 hours, with samples taken at the point of inoculation and hours 1, 3, and 24. Each culture variant was cultured and analyzed in triplicate. Sample analysis was performed via described HPLC protocol.

SA Consumption Rate and MSA Accumulation Rate Assays

50mL solutions of terrific broth (TB; MP Biomedicals) spiked with 100ug/mL carbenicillin and 100ug/mL SA were prepared in 250 baffled flasks and inoculated with *dam(-) E. coli* w/pSC-b-amp/kan/(PlacZ)cSAMT. Cultures were incubated at 28°C while shaking at 215rpm for 24 hours, with samples taken at the point of inoculation and hours 12, 14, 16, 18, 20, and 22. Cultures were prepared in biological triplicate, with HPLC sample analysis performed in triplicate via described HPLC protocol. Cultures were also olfactorily screened by humans at every point of sample collection for qualitative confirmation of MSA.

iSAMT Induction Assay

50mL solutions of terrific broth (TB; MP Biomedicals) spiked with 100ug/mL carbenicillin and 100ug/mL SA were prepared in 250 baffled flasks and inoculated with *dam(-) E. coli* w/pSC-b-amp/kan/(PlacZ)iSAMT. Cultures were incubated at 28°C while shaking at 215rpm for 36 hours,

with samples taken at the point of inoculation and hour 36. Culture induction with 5.0mM theophylline occurred after 4 hours incubation. Non-induced cultures were used as comparisons for baseline non-induced expression. Culture variants were prepared in duplicate, with HPLC sample analysis performed in triplicate via described HPLC protocol. Cultures were olfactorily screened at the 36-hour point by 3 humans for qualitative analysis.

cSAMT-KO Activity Assay

35mL solutions of terrific broth (TB; MP Biomedicals) spiked with 100ug/mL carbenicillin and 100ug/mL SA were prepared in 250 baffled flasks and inoculated with *JW3868 E. coli* w/pSC-b-amp/kan/(PlacZ)cSAMT-KO. Cultures were incubated at 28°C while shaking at 215rpm for 36 hours, with samples taken at the point of inoculation and hour 36. Culture variants were prepared in duplicate, with HPLC sample analysis performed in triplicate via described HPLC protocol. Cultures were also olfactorily screened at the 36-hour point by 3 humans for qualitative analysis.

Assay for Glucose Dependence for the pDF-cSAMT/GFP-irp9 Expression Plasmid

50mL solutions of TB (MP Biomedicals) spiked with 100ug/mL carbenicillin were prepared in 250 baffled flasks and inoculated with JW3686 *E. coli* w/pDF-(PblaZ)cSAMT/GFP-(AmyP)irp9. Cultures were incubated at 28°C while shaking at 215rpm for 48 hours, with samples taken at the point of inoculation and hours 24 and 48. After 24 hours incubation, experimental cultures were spiked with glucose to a working concentration of 5%(v/v). Non-spiked cultures were used as comparisons for baseline activity. Sample analysis was performed via described HPLC protocol. Cultures were also olfactorily screened at the 24 and 48-hour points by 3 humans for qualitative analysis.

Salicylic Acid HPLC Standard Curve Determination

0.0815 g of pure SA was dissolved in ethanol (EtOH) and pH adjusted to a final volume of 10.0mL, pH 6.7, to produce a stock solution of 8.150 mg/mL. This 8.150 mg/mL stock was then diluted to make a 0.80mg/mL SA standard, and a series of 2-fold serial dilutions were performed to produce additional standards of 0.40mg/mL, 0.20mg/mL, 0.10mg/mL, 0.050mg/mL, 0.025mg/mL, and 0.0125mg/mL respectively. These solutions were then subjected to triplicate HPLC measurements according to the previously mentioned HPLC protocol. When measuring absorbance at 237nm, HPLC peaks were observed at approximately 9.5 minutes in all SA samples which was not observed in the pure EtOH blank. Average peaks were calculated for each SA concentration triplicate measurement, and linear regression was applied to model concentrations of SA as a function of peak amplitude. The slope for this line equation was calculated to be 3.70572E-08 (unit?) and the intercept was calculated to be 0.008226423 (mg/mL). Figure 8 shows the standard curve produced from this data.

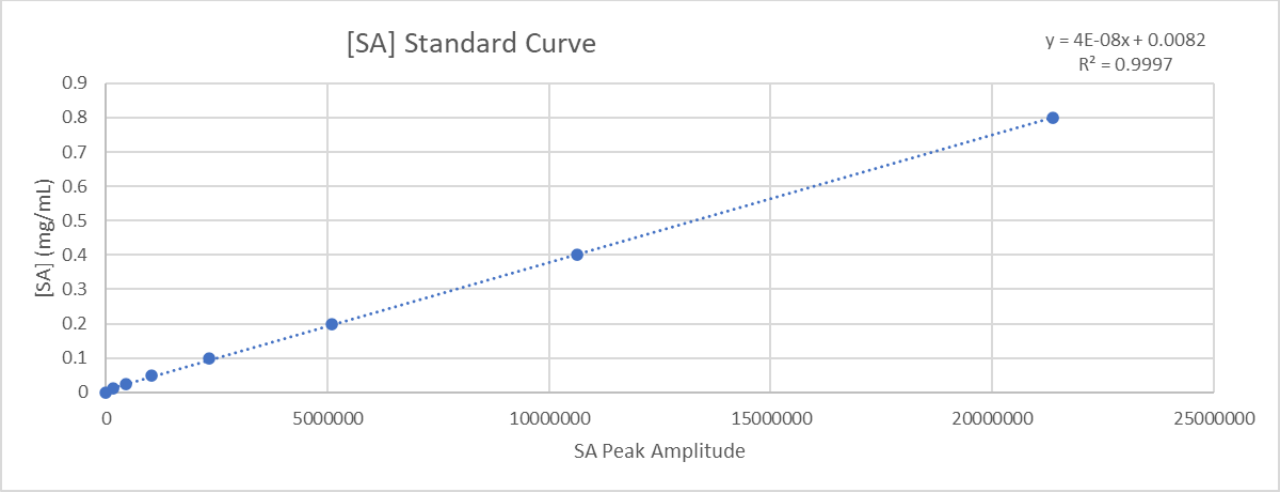


Figure 8. Standard curve for HPLC measurement of SA.

Methyl Salicylate HPLC Standard Curve Determination

10.2 uL of MSA was dissolved in EtOH to a final volume of 12.0mL to produce an initial standard solution of [99.5ug/mL] MSA. This [99.5ug/mL] MSA solution was then subjected to 3x 10-fold serial dilutions to produce MSA solutions of [9.95ug/mL], [0.995ug/mL], and [0.0995ug/mL] respectively. This process was repeated 3 times to produce 3x of each respective standard concentration. These standard solutions were then subjected to HPLC measurement according to previously mentioned HPLC protocol. When measuring absorbance at 237nm, HPLC peaks were observed at approximately 32 minutes in all MSA samples which was not observed in the pure EtOH blank. Average peaks were calculated for each MSA concentration, and linear regression was applied to model concentrations of MSA as a function of peak amplitude. The slope for this line equation was calculated to be 3.07246E-06 (unit?) and the intercept was calculated to be -0.061048303 (ug/mL). Figure 9 shows the standard curve produced from this data.

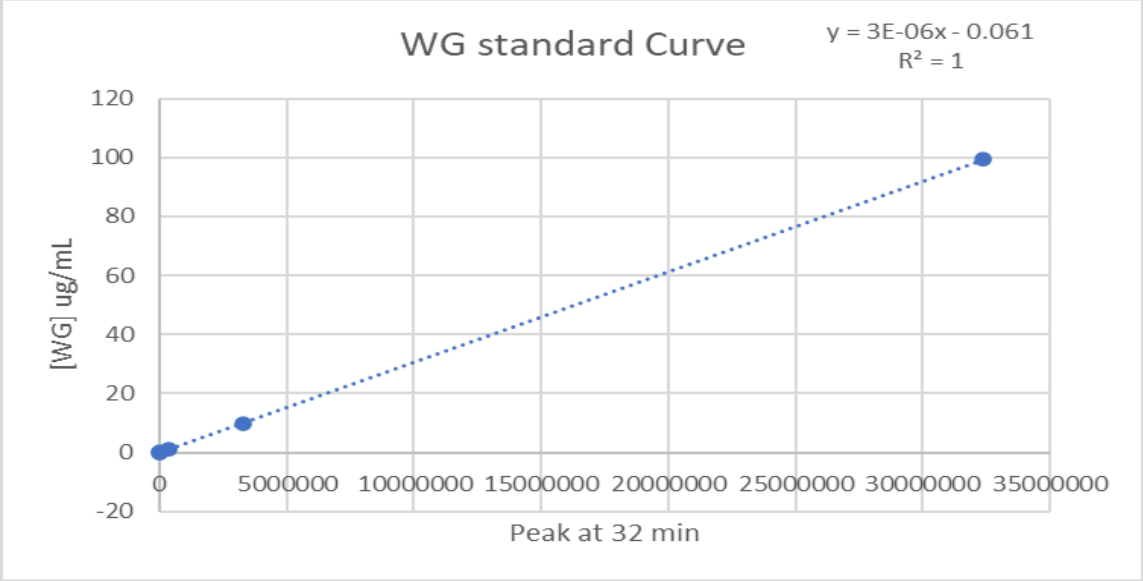


Figure 9. Standard curve for HPLC measurement of MSA.

Olfactory Threshold of Detection Screen

A 1:100 (v/v) solution of MSA in 18MΩ H_2O was produced and serially diluted to produce 1:10,000, 1:1,000,000, 1:100,000,000, 1:10,000,000,000, and 1:1,000,000,000 samples. For each sample concentration, 6x 1mL samples were prepared in 1.5mL microcentrifuge tubes. Samples were incubated at room temperature overnight to allow MSA in solution evaporate into the 0.5mL headspace to a point of thermodynamic equilibrium. Individual concentration series were screened by random lab members for odor, and screeners were not informed of what scent to screen for. Upon detection, the threshold for detection was considered to be somewhere between the first sample in which MSA was detectable and the last sample in which it was indetectable.

Results

Olfactory Threshold of Detection Screen

When screened for olfactory detection of MSA, 75% of those surveyed first detected MSA in the 1:1,000,000 (v:v) solutions, while 25% first detected MSA in the 1:100,000,000 (v:v) solutions. As such, the limit of detection for 75% of respondents is some concentration between 1:1,000,000 and 1:100,000,000 (v:v) concentrations of MSA, and some concentration between 1:100,000,000 and 1:10,000,000,000 (v:v) concentrations of MSA for the other 25% of respondents.

Confirmation of Constitutive SAMT Expression and Activity

Initial confirmation of constitutive SAMT expression and activity was achieved via HPLC confirmation of SA consumption in *dam(-) E. coli* containing the pSC-b-amp/kan/(pLacZ)cSAMT plasmid expression circuit, with a simultaneous lack of SA consumption in cell-free culture media or *dam(-) E. coli* containing the pSC-b-amp/kan/mNG

plasmid expression circuit (Fig. 10). The lack of SA consumption in the cell-free cultures demonstrated that SA did not degrade when kept in the culture media under the specified culture conditions. The lack of SA consumption in the mNG expression culture demonstrated that when an *E. coli* strain that was genetically identical as the cSAMT expression culture, but which constitutively expressed mNG instead of SAMT, was not engaging in any non-SAMT related endogenous SA consumption. In addition, MSA accumulation was observed in the *dam(-)* + pSC-b-amp/kan/(PlacZ)cSAMT cultures, but was never observed in the cell-free or mNG-expressing cultures (Fig. 11). In addition, an identical culture of *dam(-) E. coli* constitutively expressing SAMT but without addition of SA did not show any HPLC peaks associated with either SA or MSA, demonstrating that the *dam(-) E. coli* expressing pSC-b-amp/kan/(PlacZ)cSAMT did not produce endogenous metabolites which could be confused for either SA or MSA (data not shown).

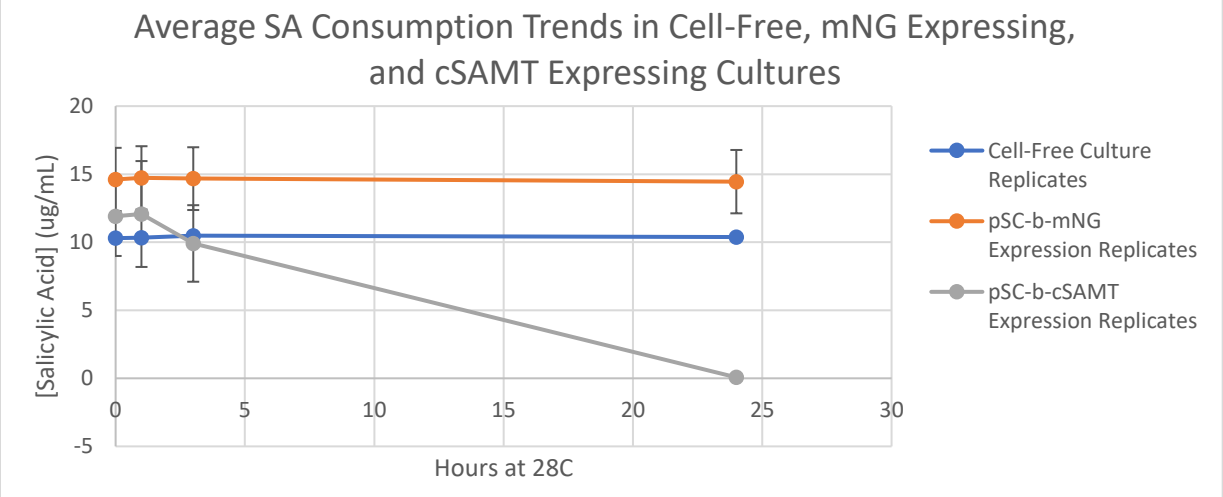


Figure 10. SA consumption trends for cell-free cultures, as well as *dam(-)* cultures constitutively expressing either mNG or SAMT in the pSC-b-amp/kan/(PlacZ) genetic background. SA consumption was only observed in the cultures in which SAMT was expressed.

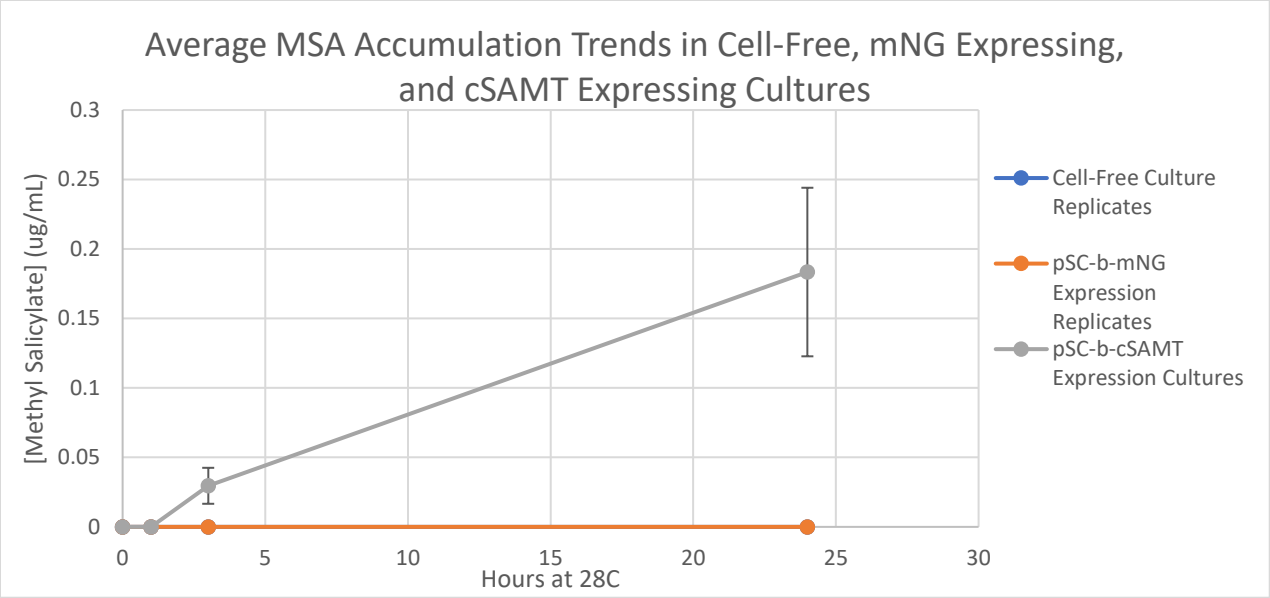


Figure 11. MSA accumulation trends for cell-free cultures, as well as *dam(-)* cultures constitutively expressing either mNG or SAMT in the pSC-b-amp/kan/(PlacZ) genetic background. MSA accumulation was only observed in the cultures in which SAMT was expressed in the presence of SA.

Enzyme Activity for Constitutively Expressed SAMT

The experimentally determined average rate of SA consumption for *dam(-) E. coli* constitutively expressing SAMT was 0.0603 ± 0.000332 ($ug*mL^{-1}*min^{-1}$). The average rate of consumption over the first 12 hours was limited at 0.00341 ± 0.00139 ($ug*mL^{-1}*min^{-1}$). The average rate of consumption increased between hours 12 and 18, with average consumption peaking at 0.102 ± 0.00491 ($ug*mL^{-1}*min^{-1}$) between hours 16 and 18. The average rate of consumption decreased between hours 18 and 22, with the average rate of consumption between hours 20 and 22 being 0.0493 ± 0.00669 ($ug*mL^{-1}*min^{-1}$). The rate of rate-increase was greatest between hours 12 and 16, with the rate of increase decreasing between hours 16 and 18 as overall enzyme rate reached a maximum. Figure 12 shows the trend for the average rates of SA consumption with respect to time.

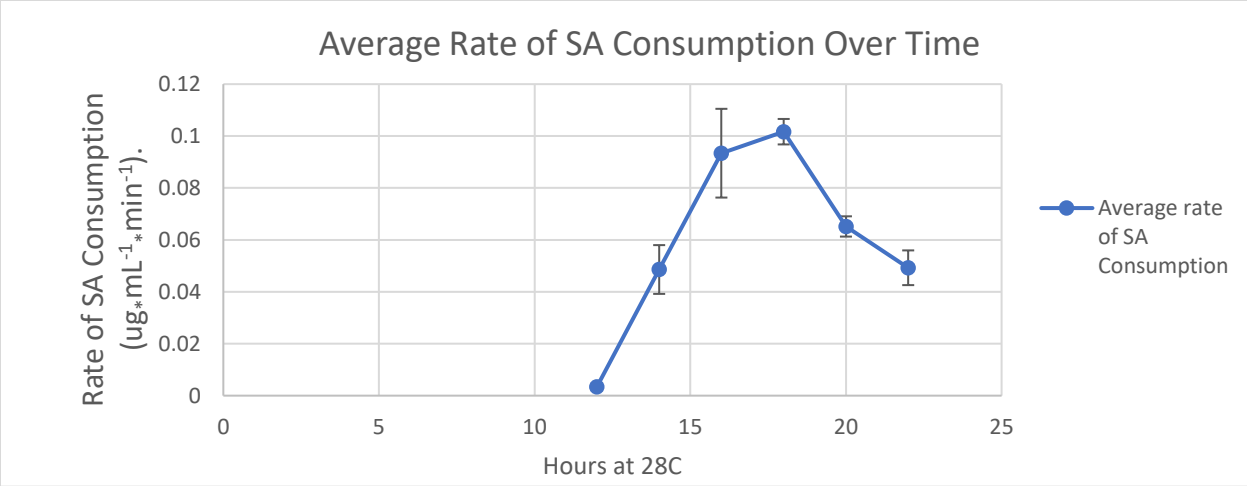


Figure 12. Average rates of SA consumption over 22 hours for *dam(-) E. coli* expressing SAMT at 28°C.

Rate of Methyl Salicylate Product Accumulation

The experimentally determined average rate of MSA accumulation for *dam(-) E. coli* expressing SAMT in the presence of SA was 0.00362 ± 0.00227 ($ug*mL^{-1}*min^{-1}$). The average rate of accumulation over the first 12 hours was limited at 0.000319 ± 0.0000309 ($ug*mL^{-1}*min^{-1}$). The average rate of accumulation increased between hours 12 and 16, with average rate of MSA accumulation peaking at 0.00646 ± 0.000740 ($ug*mL^{-1}*min^{-1}$) between hours 14 and 16. The average rate of accumulation decreased between hours 16 and 22, with the average rate of accumulation between hours 20 and 22 being 0.00221 ± 0.000221 ($ug*mL^{-1}*min^{-1}$). The average final concentration of MSA was 2.74ug/mL. Figure 13 shows the trend for the average rates of MSA accumulation with respect to time.

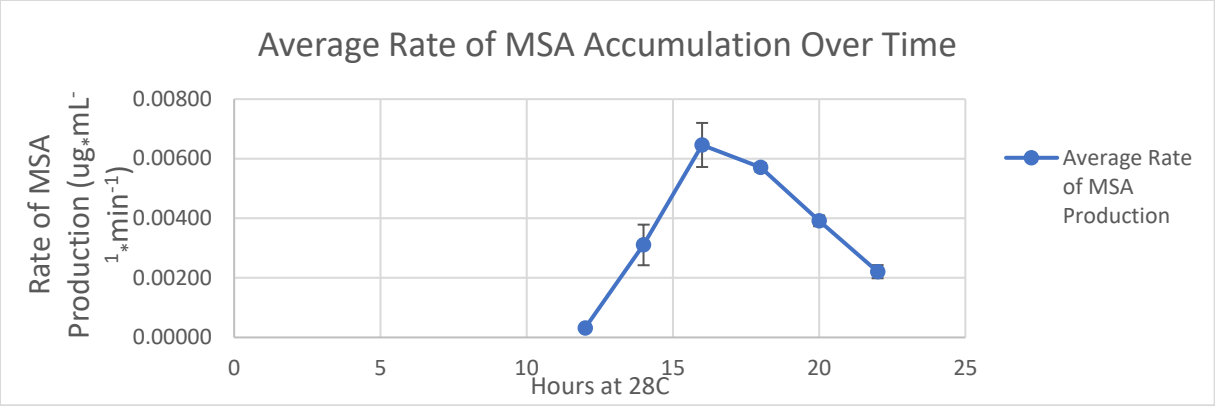


Figure 13. Average rates of MSA accumulation over 22 hours for *dam(-) E. coli* expressing SAMT at 28°C.

pSC-b-amp/kan/(PlacZ)iSAMT Theophylline Induction and Enzyme Activity Assay

Cultures which received theophylline induction at the 4-hour point showed significant consumption of SA and accumulation of MSA after 36 hours incubation. The average rate of SA consumption observed in *dam(-) E. coli* expressing pSC-b-amp/kan/(PlacZ)iSAMT was approximately 0.00158 ($ug*mL^{-1}*min^{-1}$) for the theophylline induced cultures. Of the cultures which did not received theophylline induction, one culture showed no significant SA consumption, while the other consumed SA at an approximate rate of 0.000700 ($ug*mL^{-1}*min^{-1}$). The rate of SA consumption observed in the induced cultures was approximately 1.26-fold higher than the rate of SA consumption in the one non-induced culture. The average MSA accumulated in the theophylline induced cultures was measured to be approximately 6900 ug/mL, while the non-induced cultures did not show any measurable MSA accumulation. Upon human olfactory screening, the theophylline induced cultures presented a mild but noticeable odor of MSA, while the non-induced cultures presented a sour odor associated with highly dilute MSA. The rate of SA consumption observed in the theophylline induced cultures was approximately 2.6% of the SA consumption rate observed in the constitutive SAMT expression cultures, representing an approximately 38-fold decrease in measured SA consumption rates

between theophylline induced expression and constitutive expression of SAMT by *dam(-) E. coli*.

Assay for Glucose Dependence with the pDF-cSAMT/GFP-irp9 Expression Plasmid

Both the glucose spiked experimental cultures and the non-glucose spiked control cultures showed significant production of SA. The SA accumulation was 18.0 ug/mL in the glucose spiked culture and 15.6 ug/mL in the non-spiked culture, representing a 15.3% difference in SA accumulation between glucose spiked and non-spiked conditions. Both the glucose spiked experimental culture and the non-glucose spiked control culture showed significant production and accumulation of MSA. The maximum MSA accumulation was 0.567 ug/mL in the glucose spiked culture and 0.113 ug/mL in the non-spiked culture, representing a 403.5% difference in maximum MSA accumulation between glucose spiked and non-spiked conditions. Maximum MSA levels were observed in the 48-hour time point sample for the glucose spiked culture, whereas maximum MSA levels for the non-spiked cultures were observed in the 24-hour time point sample.

SAMT Knock Out Mutant Confirmation

No MSA accumulation was measured in cultures of JW3686 *E. coli* expressing the SAMT knockout mutant expression plasmid pSC-b-amp/kan/(PlacZ)cSAMT-KO after 36 hours incubation at 28°C. This was supported by a lack of olfactorily detectable MSA when these cultures were olfactorily screened by 3x humans. Both cultures did show significant increases in the HPLC peak which occurred at the 9.5-minute time point and is associated with SA.

Discussion

Expression System Development

Development of the Plasmid Expression Systems

Because the initial goal was expression in *S. carnosus*, but the original cloning and strain development was to be performed in *E. coli* due to its ease of use as a major model organism for genetic engineering, it was deemed important to develop our plasmids as *S. carnosus* to *E. coli* shuttle vectors. Shuttle vectors are genetic expression vectors, usually plasmids, which have been designed specifically for the easy transfer of genetic information between different species. Shuttle vectors are commonly used to transfer genetic expression systems either to or from a more complex expression species such as *Porphyromonas* or *Bacteroides* species to a simpler expression species such as *E. coli* [57]. For that reason, our initial plasmid vectors had our SAMT and irp9 genes under control of the AmyP promoter as this promoter was believed to facilitate expression in both *E. coli* and *S. carnosus.* However, an inability to confirm expression of SAMT in *E. coli* led to the suspicion that the AmyP promoter may have been less than successful at inducing expression in *E. coli*. For this reason, it was decided to replace the AmyP promoter with the promoter for the lac operon (PlacZ).

In addition to the changes to promoter, another significant aspect of the final plasmid design was the transition from the commercially produced plasmid backbones (pIDT and pTWIST) into the pSC-b-amp/kan plasmid backbone (Agilent). This step was performed to allow eventual SAMT expression in the JW3686 strain of *E. coli* due to that strain's indole deficiency, and the belief a lack of indole expression would facilitate an easier olfactory detection of low-level MSA production. However, the mutation utilized to induce indole deficiency in the JW3686 strain also provided resistance to kanamycin, and the commercially provided plasmids utilized kanamycin resistance as their selection markers, and thus JW3686 *E.*

coli transformed with the commercially provided plasmids were unable to be selected for using antibiotic selection on kanamycin [58]. To overcome this issue via providing of ampicillin and carbenicillin resistance, the irp9 and SAMT genes were transferred into the pSC-b-amp/kan backbone using a commercially produced blunt-end cloning kit. This allowed subsequent antibiotic selection for successful transformation of our SAMT expression circuits into JW3686 *E. coli*.

Unfortunately, even with expression being placed under the PlacZ promoter, expression of SAMT in *E. coli* was still unable to be confirmed through SDS-PAGE or biochemical analysis. To address this issue, attempts were made to link the fluorescent reporter protein mNeonGreen (mNG) to the SAMT enzyme via a tobacco etch virus (TEV) cleavable linker. The idea behind this was visual confirmation of plasmid expression via observation of fluorescence in the cells expressing the SAMT-mNG fusion protein. This attempt was successful for the constitutively expressed version of SAMT (cSAMT), but for unknown reasons was unsuccessful for the theophylline-inducible riboswitch version of SAMT (iSAMT). However, the cSAMT-mNG fusion protein expression vector was useful in obtaining an initial visual confirmation of plasmid expression in *E. coli*.

In addition to the pSC-b-amp/kan/(PlacZ) plasmid vectors, the full expression circuit containing both SAMT and irp9 was developed in the both the pSC-b and pDF1 plasmid backbones. Unlike the previously discussed plasmids which were designed to only express SAMT, but which required artificial SA supplementation, the larger SAMT+irp9 expression systems were designed to produce both SA and MSA from the endogenous chorismate precursor, a necessity for the eventual deployment in the nasal microbiome of animal sentinels. The pDF plasmid backbone was introduced due to its increased ability to be expressed in various bacterial

species outside of *E. coli* and the hopes this would allow for a viable *E. coli* to *S. carnosus* shuttle vector. The promoter utilized for controlling expression of SAMT and irp9 in the pDF backbone was AmyP, but with unique promoter cites for SAMT/GFP fusion gene and irp9. While this plasmid was not utilized in the development and assaying of the final enzyme library due to its incompatibility with the assay techniques, it was utilized upon eventual transition into animal studies.

E. coli Strain Choices

Various strains of *E. coli* were employed throughout the plasmid development process, including DH5-alpha, E. cloni EXPRESS BL21-(DE3), JW3686, *dam(-)/dcm(-)*, and *E. coli* Nissle. The initial use of DH5-alpha as our cloning strain was simply due to its prevalence and ready availability in the lab at the start of the project, and its successful use in previous cloning projects. However, SDS-PAGE analysis of the crude lysate and His-tag purified SAMT-mNG fusion suggested protease degradation was occurring in this strain. For this reason, the decision was made to transition into the BL21 strain of *E. coli*, as BL21 is a protease deficient *E. coli* strain which had previous been used with success in the Snow lab for expression of protease sensitive proteins. As previously mentioned, the use of the JW3686 strain was for the purpose of creating a strain of *E. coli* which could produce MSA in the absence of the highly odiferous indole produced by standard *E. coli.* This was deemed ideal for human olfactory screening of cultures producing MSA at lower levels which might be masked by the presence of indole.

Due to issues with successfully transforming our *E. coli* produced shuttle plasmids into *S. carnosus*, the *dam(-)/dcm(-)* strain was obtained due to its deficiency in the dam and dcm methylases which are responsible for in vivo methylation of DNA. Methylation of DNA by these enzymes has been shown to reduce transformation efficiency of *E. coli* produced DNA into other bacterial species, and thus it was believed that use of the *dam(-)/dcm(-)* strain would facilitate

production of more effective *E. coli* to *S. carnosus* shuttle vectors [59]. One potential issue with use of the *dam(-)/dcm(-)* strain of *E. coli* is that when DNA produced in the absence of dam methylation is used at a template for cloning techniques such as PCR, it is unable to be subsequently digested away using the enzyme Dpn1. This enzyme functions by cleaving the in-vivo produced dam methylated DNA, allowing the template DNA to be removed and only the in-vitro produced DNA products to be maintained for further downstream applications [60]. However, an unforeseen benefit of this strain is suspected increased bioavailable SAM due to the fact the dam and dcm methylases utilize SAM as their methyl donor, and thus by having these enzymes inactive there was the potential for increased bioavailable SAM for the SAMT reaction.

Ultimately, due to continued issues with transformation and expression of our shuttle vectors into *S. carnosus,* a decision was made to transition our target organism from *S. carnosus* to the *E. coli* Nissle 1917 strain. The Nissle strain of *E. coli* was selected due to its non-pathogenic nature, its current use as a digestive probiotic, and its widespread presence as a member of the enterobiome of mammals including humans and rats [61]. For this reason, it was believed that *E. coli* Nissle would be a safe organism for inoculation of the sentinel animal's nasal microbiome. This was seen as a good temporary alternative to the use of *S. carnosus* for our animal studies as it provided a biologically safe alternative species for which genetic engineering techniques are widely available, and for which the lab has already established genetic engineering success in. While use of the Nissle strain of *E. coli* facilitated a more rapid ability to engage in animal studies, creating a bypass for the issues with transformation and expression in *S. carnosus*, it does not mean *E. coli* Nissle will completely replace *S. carnosus* in the longer-term project. As mentioned earlier, *S. carnosus* has the ability to be dry-frozen and reconstituted on demand while *E. coli* requires more demanding storage and maintenance conditions. For this reason, it would

be ideal to eventually obtain successful expression of our engineered olfaction system in *S. carnosus* while the *E. coli* Nissle is used in the meantime for the animal studies.

Enzyme Activity Assay Development

Initial Enzymatic Assays

In addition to basic olfactory screening for qualitative confirmation of SAMT expression and activity, the desire for quantitative data led to an attempt at determining SAMT activity via measuring the conversion of the SAM coenzyme into its SAH byproduct. These attempts relied on a combination of specific enzymatic reactions coupled with spectroscopic measurements of product accumulation or depletion. These assays utilized multienzyme processes to convert SAH into spectroscopically measurable products, thereby allowing tracking of methyltransferase activity as well as preventing SAMT reaction stalling due to product inhibition. For example, as assay adapted from Akhtar et al. utilized SAH-nucleosidase (SAH-Nuc) to convert SAH to adenine, xanthine oxidase (XO) to convert the adenine to 8-hydroxyadenine and the 8-hydroxyadeneine into 2,8-dihydroxyadeneine and produce 2 molecules of H_2O_2 in the process, and horseradish peroxidase (HRP) to utilize that H_2O_2 to convert to colorless Amplex Red into the red colored resorufin [62]. Similarly, the commercially produced SAM510 SAM Methyltransferase Assay from G-biosciences utilized multienzyme processes to convert SAH to spectroscopically measurable biomolecules but through a different enzymatic pathway than was adapted from Akhtar. Unfortunately, lack of clear SAMT activity confirmation with these methods led us to adopt an alternative method of confirming and measuring enzyme activity via high performance liquid chromatography (HPLC).

High Performance Liquid Chromatography (HPLC) Assay

Because of the lack of success with use of the spectroscopic investigations of in vitro enzyme activity via metabolism of SAM, the decision was made to instead investigate enzyme activity by measuring depletion of SA from the growth media of *E. coli* expressing SAMT. Like all chromatography techniques, HPLC functions by separate components of a mixture in a time specific manner, with HPLC utilizing differences in a molecule's solubility within the solid or liquid phases to separate compounds in a liquid solution in a very predictable and repeatable manner. In addition to the ability to separate chemical compounds in a very specific manner, HPLC can also be used to determine concentrations of a compound in a solution when measurements are compared to a known standard [64]. This was deemed to be ideal for tracking of SA depletion from the growth media of *E. coli* expressing SAMT due to the ability to determine with high accuracy the concentrations of SA in the growth media and how these concentrations changed when in the presence of SAMT expressing *E. coli.* Subsequent confirmation of the lack of endogenous SA metabolism in our *E. coli* confirmed the viability of the HPLC assay as a manner of confidently tracking changes in SA concentrations when SAMT either was or was not expressed in a given culture.

In addition to being able to track SA levels from the liquid media, HPLC provided the added benefit of being able to confirm MSA production and measure MSA accumulation in the same growth media for which SA depletion was being measured. This was seen as an additional benefit of using the HPLC method of confirming and measuring enzyme activity, as this allowed confirmation of MSA production even when the concentrations of MSA being produced were below the olfactory thresholds for humans. The high accuracy and threshold of detection provided by HPLC was particularly important due to the high volatility of MSA and the consequential tendency for MSA to readily evaporate out of the liquid media. Because of this

issue, MSA quantitative detection methods had to be sensitive enough to measure even low concentrations of MSA with high accuracy and reproducibility. As previous published work had already demonstrated the ability for HPLC to measure both SA and MSA with high accuracy, it was decided that HPLC would be utilized at the primary method for confirming enzyme expression and activity, and subsequent quantitative determination of enzyme activity [65,66]. The determination of HPLC as a valid method for tracking enzyme activity provided the ability to move forward with measurement of enzyme activity in WT-SAMT and the eventual SAMT library variants and confirmed the necessary techniques had been established to allow the production and in-lab investigation into the resultant computationally designed SAMT libraries.

Wild-Type Enzyme Activity in *dam(-) E. coli* Expressing pSC-b-amp/kan/(PlacZ)cSAMT

Results of the HPLC assays demonstrate that SA is being consumed and MSA in being produced and accumulated at significant rates in cultures of *dam(-) E. coli* constitutively expressing SAMT. The conclusion that this SA consumption is due to the expression and activity of SAMT is supported through the measurable presence of MSA in cultures which demonstrate SA consumption in the presence of SAMT expression (*dam(-)* + pSC-b-cSAMT), and the lack of either SA consumption or MSA production in cultures without any SAMT expression system (cell-free or *dam(-)* + pSC-b-amp/kan/mNG).

In cultures with measurable MSA production, SA consumption during the first 12 hours was limited, suggesting that during these first 12 hours the cells had yet to begin expressing SAMT at high enough levels for enzyme activity to be observable. After 12 hours, enzyme activity began to increase at a relatively constant rate for approximately four hours before increases in enzyme activity began to slow and the enzyme reached maximum activity between hours 16 and 18. Between hours 18 and 22, the cultures still displayed significant consumption

of SA; however, the rate of consumption began to decrease, suggesting that peak enzyme activity for SAMT occurs between hours 12 and 18 when expressed in *dam(-) E. coli* grown at 28°C. The observation that the rate of rate-increase begins to decrease between hours 16 and 18 could be indicative of a maximum rate of enzyme expression occurring between hours 12 and 16, and expression levels decreasing after hour 16; however, further analysis would need to be done to confirm this hypothesis.

The detection of MSA in cultures shown to consume SA above a certain rate, and a lack of MSA in cultures not consuming SA, is highly indicative of SAMT activity in these cultures resulting in the conversion of SA into MSA at levels significant enough for detection via HPLC. The lack of any HPLC peak around the 32-minute point in the cultures expressing mNG instead of SAMT confirms that the peaks are truly due to the production and accumulation of MSA as a result of SAMT activity, and not an artificial measurement caused by some other endogenous metabolite. In addition, the presence of significant MSA after approximately 12 hours for was corroborated via qualitative analysis using basic human olfactory detection. In all cultures cultured under the SAMT expression conditions while in the presence of SA – either artificially supplied or produced via irp9 activity) – a definite MSA odor was detected by all other lab members who participated in olfactory screening for MSA (number and identity of screeners varied between screening events).

Due to the one-to-one stochiometric nature of the SA to MSA reaction, it would be expected that for every 1ug of SA which was consumed, roughly 1.1 ug of MSA would be produced; however, this one-to-one ratio was not observed in the HPLC measurements. This was to be expected due to the fact MSA is highly volatile and thus prone to evaporation out of the culture media over the course of prolonged incubation at 28°C. This is why MSA concentrations

and rates were referred to as accumulation rather than production, as the limitations of HPLC measurement prevented measurement of MSA product which had evaporated out of the media prior to sample collection. This evaporative loss is also believed to be why no MSA was detectable in the non-induced iSAMT cultures despite SA consumption being observed (to be discussed), as the very low rate of MSA production which would occur under these leaky expression conditions would likely be lower than the rate of evaporation out of solution, causing MSA concentrations in the culture media to remain beneath the threshold of HPLC detection. However, by applying the one-to-one stoichiometry of the reaction to the measured concentrations of SA and MSA in a given culture at a given time point, it should be possible to roughly determine the amount of MSA which has evaporated out of the culture media and would therefore be accessible for olfactory detection.

The experimentally determined average rate of enzyme activity for constitutively expressed SAMT, as expressed in terms of SA consumption per minute, was 0.0603 ± 0.000332 ($ug*mL^{-1}*min^{-1}$) for *dam(-) E. coli* when cultured at 28°C in TB in the presence of 100ug/mL SA. This was thus considered the wild-type enzyme activity for the 1m6e version of SAMT when cultured under the specified culture conditions in the *dam(-)* strain of *E. coli.* It should be noted that this rate of activity was determined in the *dam(-)* strain of *E. coli*, and could theoretically be different in other strains of *E. coli.* This is due to the fact that *dam/dcm* methylases utilize SAM as the donor for their methylation reactions, and as a result most strains of *E. coli* will have limited amounts of bioavailable SAM for use in the SAMT reaction [23]. Due to a deficiency in these enzymes in the *dam(-)* strain of *E. coli*, it was hypothesized that there would be an increased amount of bioavailable SAM which could be utilized to drive the SAMT reaction [24]. This hypothesis is supported by previous work which demonstrated that methylases which utilize

SAM as a methyl donor have increased activity when bioavailable SAM is increased [25]. Unfortunately, there was not a simple method to measure concentrations of endogenous SAM, but considering the implications of potential increased bioavailable SAM on SAMT activity, it would be plausible to suspect an increased overall enzyme activity in the *dam(-)* strain. Individual cultures of BL21 or JW3686 constitutively expressing SAMT in the presence of SA did demonstrate activities of 0.0610 ($ug*mL^{-1}*min^{-1}$) and 0.0649 ($ug*mL^{-1}*min^{-1}$) respectively; however, due to a lack of biological replicates in this assay, a statistical conclusion to these strain's activities could not be made.

pSC-b-amp/kan/(PlacZ)iSAMT Theophylline Induction and Enzyme Activity Assay

Significant SA consumption was observed in both theophylline induced cultures and one of the two non-induced cultures of *dam(-) E. coli* expressing pSC-b-amp/kan/(PlacZ)iSAMT; however, at no point was MSA accumulation detectable via HPLC in either non-induced culture, while MSA was measurable in both theophylline induced cultures. This is suspected to be due to mild leaky expression of SAMT in the non-induced cultures which resulted in SA consumption and MSA production, but with the rate of MSA production being less than the rate of MSA evaporation from the media, resulting in accumulated MSA concentrations being consistently below the threshold of HPLC detection.

This suspicion is supported by the results of the human olfactory screen which demonstrated a clear odor of MSA in the theophylline induced cultures, with this odor readily described as "minty" by all screeners. In contrast, the non-induced cultures possessed a slightly sour odor reminiscent of Greek yogurt which was notably different than the characteristic odors which emanate from *E. coli* cultures not containing the SAMT gene. This odor was also detected during the early enzyme activity assays in which *E. coli* constitutively expressing SAMT were

cultured in the presence of 10ug/mL SA and which showed full consumption of SA and very minor accumulation of MSA. This supports the idea that this characteristic scent is associated with highly dilute gaseous MSA, and that the presence of this odor in the non-induce iSAMT cultures is representative of low levels of MSA in the headspace despite there being no measurable MSA in the HPLC sample analysis.

The rate of SA consumption in the theophylline induced expression system was approximately 2.6% the rate of SA consumption observed in the constitutively expressed SAMT cultures of the same genetic background. This represents an approximately 13.3% difference between the observed ratio of induced expression to constitutive expression (2.6%) and the ratio of induced expression to constitutive expression reported in the theophylline riboswitch literature (2.3%) [12]. This is a relatively minor difference and demonstrates that the theophylline controlled riboswitch method of inducing expression is effectively able to induce low-level expression such that there is measurable activity and olfactorily detectable MSA, but at levels much lower than would be observed with a constitutive SAMT expression system. This should allow confidence that the amount of MSA produced in the nasal microbiome of the animals will be great enough to be olfactorily detected, yet low enough to not pose a risk of toxicity.

It should be noted that there has not been any explicitly determined toxicity level for mammalian inhalation of MSA as previous toxicity studies involved either topical or gastrointestinal administration, and there also has not been any explicit determination of gaseous MSA concentrations produced by our engineered expression systems, so there is a chance the levels produced are above the range of toxicity, and as such animals being exposed to the expression system should be initially monitored for potential MSA related health effects. This may pose a greater concern with the constitutive expressed computationally designed mutants as

the amount of MSA they will produce is currently unknown, whereas with the inducible cultures the levels will likely always be low enough to be non-toxic simply due to the relatively low-level expression which occurs under the riboswitch.

One thing that is odd in the results for the theophylline induced cultures is the fact that the amount of MSA which accumulated in the media is significantly greater than the amount of SA which was measured to be consumed over the course of the 36 hours. All previous assays into SAMT activity showed a level of SA consumption which was dramatically greater than the measured accumulation of MSA, which makes sense given the fact MSA would be expected to evaporate out of the media solution such that only a fraction of what was produced would remain in the culture media. Furthermore, the initial concentrations of SA in these cultures were measured to be significantly lower than the 100ug/mL which was used as an initial SA concentration. This is odd given the same solution of SA was used to produce all enzymatic assay cultures. This is also not suspected to be due to issues of improper loading of samples into the HPLC due to the fact no samples showed SA concentrations near 100ug/mL which would have occurred if samples were placed into the HPLC in the wrong order, nor is it suspected to be due to inaccurate HPLC measurements since each sample was measured in triplicate and all measurements for a given sample were essentially equal outside of the normal expected variance.

SAMT Knockout Mutant

Results of the SAMT-knockout investigation indicated that the knockout mutations were successful in preventing catalysis of SA to MSA, as evident by both the complete lack of any detectable MSA in the HPLC analysis coupled with the lack of any MSA detection in the qualitative human screenings. This suggests the M150G and W151G mutations were successful in disrupting the active site to such an extent that catalytic ability as completely lost. This is

suspected to be due to the fact that M150 and W151 are both critical active site residues which are involved in securing SA in the appropriate place necessary for methylation to occur, and thus changing these residues to glycine can effectively prevent the binding of SA within the active site. Furthermore, the fact that no MSA was detectable via human olfactory screening indicates that the lack of MSA in the culture media after 36 hours was truly due to a lack of MSA production and not due to MSA production and subsequent evaporation into the headspace. This confirms the viability of the SAMT-KO expression plasmid as a control for MSA production in the ultimate animal inoculation studies.

The increases in HPLC peak amplitude at the 9.5-minute point were unexpected and currently are inexplicable. The assays into enzymatic activity for constitutively expressed WT-SAMT all demonstrated no HPLC peak at the 9.5-minute point in the absence of artificial spiking with SA. Furthermore, SA consumption assays on *E. coli* which did not express SAMT but where cultured in the presence of SA demonstrated no significant change in SA levels, demonstrating that *E. coli* of the same genetic background minus SAMT did not produce or consume SA over the course of incubation. The only time there were observed HPLC peaks at 9.5-minutes where in either cultures which had been artificially provided with SA in the absence of irp9 expression, or which expressed irp9 and thus produced SA endogenously, and the only time the HPLC peaks at 9.5-minutes were observed increasing was in the cultures explicitly expressing irp9 and consequently producing SA endogenously. However, the pSC-b-amp/kan/(Placz)cSAMT-KO plasmids did not possess the irp9 gene and thus would not be expected to produce increased HPLC peak amplitudes at the 9.5-minute point.

Considering the only mutations to this strain occurred explicitly within the SAMT gene, and the only effect they were expected to have was the prevention of reaction catalysis due to an

inability for SA to bind the active site, it would not be expected that these mutations would result in production of novel metabolites which would also possess an HPLC peak around the 9.5-minute point. However, these results were observed in both biological replicates and in all three HPLC measurements of each sample, indicating these results were not due to sampling errors or some abnormality in a singular culture. Why these increases in HPLC peak amplitude at the 9.5-minute point occurred needs to be further investigated to understand what metabolites are producing these peaks and why they are being produced; however, their presence does not negate the fact that the SAMT-KO was unable to produce any detectable MSA and thus can still operate as the negative control for the animal studies.

Assay for Glucose Dependence with the pDF-cSAMT/GFP-irp9 Expression Plasmid

Results for the glucose spiking investigation demonstrate that glucose spiking does not have a dramatic effect on the rate of SA production by irp9, but does seem to produce a significant and dramatic difference in overall MSA production by SAMT. One potential explanation to this could be the fact that chorismate is a product of the Shikimate pathway which itself is dependent upon significant carbon sources in order to produce the large number of essential metabolites which can arise from this important microorganism metabolic pathway. Previous work has demonstrated that glucose spiking can increase the products of the Shikimate pathway in *E. coli*, and it is suspected that glucose spiking could increase the availability of chorismate in order to drive forward the reaction [74]. This hypothesis is somewhat refuted by the observation of similar levels of SA production in both the glucose spiked and non-spiked cultures which would likely only occur if there were similar levels of chorismate available to convert to SA. Furthermore, the somewhat similar final concentrations of SA between the two culture conditions, and the fact that neither culture was artificially provided with SA, indicates

that the activity of irp9 was not dramatically increased due to the presence of glucose. However, it should be noted that the glucose spiked culture showed approximately 15% greater accumulation of SA while also accumulating over 400% greater concentrations of MSA, indicating a significant amount of SA produced in these cultures was converted to MSA which somewhat discredits the similar final concentrations of SA observed between culturing conditions.

An alternative hypothesis to why the spiking with glucose could affect such a dramatic increase in MSA accumulation levels is that the additional carbon provided by the glucose somehow led to additional metabolites being fed into the SAM/SAH recycling pathway known as the active methyl cycle. This pathway is responsible for recycling the demethylated SAH back into SAM, and has been shown to be dependent upon bioavailable levels of methionine and homocysteine. Since both methionine and homocysteine are utilized in metabolic pathways outside the activated methyl cycle, it is possible that additional glucose resulted in increased bioavailability of certain undetermined metabolites which ultimately fed into the activated methyl cycle. Methyltransferases which are dependent upon SAM as a coenzyme have been shown to have their activity increased by increasing bioavailable concentrations of SAM, and if glucose spiking were to ultimately result in increased bioavailable SAM, this could explain the dramatic differences in MSA accumulation between glucose spiked and non-spiked cultures [68]. This would make sense since the results seem to indicate that the real metabolic bottleneck lies in the ability of SAMT to convert the SA produced by irp9 into MSA when glucose either is or is not present, rather than the inability to produce significant SA preventing significant MSA production.

To make concrete conclusions about exactly how the additional glucose was affecting the overall rate of conversion of SA to MSA will require further metabolic investigations; however, it is safe to say that there is a significant and dramatic increase in the rate of MSA production upon spiking with glucose for *E. coli* expressing pDF-(Pblaz)cSAMT/GFP-irp9. This result is important to keep in mind when attempting to utilize this expression plasmid in the nasal microbiome of sentinel animals as it suggests having significant glucose in the cell would be useful, and potentially necessary, to produce olfactorily detectable concentrations of MSA, as as such, further metabolic engineering may be required to facilitate increased glucose production in the final engineering organisms.

Olfactory Threshold of Detection Screen

Results of the olfactory threshold screen indicated that reliable human olfactory detection of MSA in the 0.5mL headspace of a 1.0mL liquid sample occurred when initial MSA liquid concentrations were greater than 1:100,000,000 (v:v) and less then or equal to 1:1,000,000 (v:v), with all individuals surveyed indicating clear MSA detection in the 1:1,000,000 (v:v) sample. Considering the previously determined threshold for human olfaction of MSA is 0.62-0.87 mg/mL, this suggests that the concentration of MSA in the headspace of the 1:1,000,000 (v:v) sample of MSA must be greater than 0.62mg/mL. A concentration of [0.62mg/mL] MSA was calculated to be approximately 530ppm. This is well within the range of MSA concentrations found in common foods and is indicative MSA levels should be non-toxic when at the lower range of olfactory detection. By that same merit, MSA concentrations of 8400ppm – the higher end of those found in food approved for human consumption – would not occur until MSA vapor concentrations reached approximately 98.3mg/mL.

It is worth noting that this screen for olfactory threshold of detection utilized a somewhat low-tech approach with a limited number of samples and individuals surveyed. In addition, the liquid and headspace volumes being investigated were not equal to the volumes and headspaces of the 50mL cultures used in the HPLC analysis and which were olfactorily confirmed to produce significant amounts of MSA. As such, it is somewhat difficult to make comparisons between olfactory detection events in the larger HPLC analyzed cultures versus the smaller olfactory screening samples. For example, it is possible that the limited headspace volume of 0.5mL in the olfactory screening samples prevented sufficient buildup of MSA for olfactory detection, whereas that same volume and concentration of MSA solution may have produced olfactorily detectable concentrations of gaseous MSA with a larger available headspace. Despite this lack of direct comparability, knowledge of olfactory thresholds of detection in any size sample is still important information which confirms the ability for human olfactory detection of MSA at relatively low concentrations.

Another important thing to note about this screen is that it is determining a rough threshold for human olfactory detection of MSA and as such would not be expected to be the same between species. While the threshold for canine and rodent olfactory detection of MSA has not been previously determined, it would be expected that these animals would have a greater sensitivity to low concentration odorants than would humans due the fact non-human mammals tend to have greater olfactory detection sensitivity than do humans [23]. As the final deployment target for this enhanced olfaction system would be mammalian sentinel species, it would be expected that different sentinel species would have different average thresholds of detections for MSA, and thus a unique threshold of olfactory detection should be determined for any species for which the technology is going to be applied. However, the fact that olfactory detection in

humans was confirmed, coupled with the fact that animal sentinel species would be expected to have greater olfactory sensitivity, gives confidence that the amount of MSA produced should be well above the thresholds of detection for any sentinel animal species.

Conclusions

The investigation into enzymatic activity for the various SAMT expression systems demonstrated that there was significant constitutive expression and activity of SAMT for *E. coli* strain expressing the pSC-b-amp/kan/(PlacZ)cSAMT or the pSC-b-amp/kan/(PlacZ)cSAMT-mNG expression plasmids, with a salicylic acid consumption rate of 0.0603 ($ug*mL^{-1}*min^{-1}$) determined for *dam(-) E. coli* expressing pSC-b-amp/kan/(PlacZ)cSAMT. This rate of salicylic acid consumption has been and will continue to be used as the baseline for enzymatic activity when investigating enzymatic activity in the theophylline inducible gene circuit and in the mutant SAMT computational library members.

The measured rate of SA consumption in the theophylline induced expression system was significantly lower than that observed for the constitutive expression system, with the maximum consumption rate observed in the theophylline induced expression systems being approximately 2.6% of that observed in the constitutive expression system. This suggests that the theophylline induced riboswitch is not able to replicate the expression levels observed upon constitutive expression, but were still able to reach a high enough expression level that MSA odor was detectable by all human olfactory screeners. Given the higher general sensitivity to odors possessed by rodents in comparison to human, it is suspected that even at the relatively low expression levels observed with theophylline induced SAMT expression there should still be sufficient enough MSA production to be detected by trained MSA detecting rats, although that is

a speculatory statement at this point due to the lack of fully trained MSA detecting rats for which to test this hypothesis.

For the SAMT knockout mutant, results show a lack of any MSA production in either the HPLC analysis or in the human olfactory screen. This is highly supportive of the fact the SAMT knockout mutant was successful in preventing any catalytic conversion of SA to MSA, and as such it can be utilized as a negative control for the eventual animal studies as it is genetically identically to the constitutively expressed wild-type SAMT outside of two amino acid mutations. While the increase in HPLC peak around the 9.5-minute point is definitely interesting and deserving of further investigation, it should not preclude use of the knockout mutant as a genetically identical negative control in the animal studies.

Future Work

High-throughput SAMT Activity Screening via SA-Repressed mNG Expression System

While use of HPLC to track SAMT activity has been the method of choice for the initial enzyme activity screening, it does have a relatively low throughput which limits the size of the libraries it can effectively screen. To overcome this limitation, a high throughput screening method has been proposed by Dr. Kevin Morey. This method would involve creation of a secondary expression vector for SA induced expression of a repressor protein which in turn would repress expression of a fluorescent reporter protein such as mNG. This idea combines two previously validated expression systems with slight modifications to specifically allow screening for SAMT activity.

The first previous validated component of this screening expression vector is an SA activated expression system in which SA binds with the S *E. coli* transcription factor AraC to induce expression of downstream genes. This system utilizes a mutant SA-specific version of

AraC, AraC-SA, which has been shown to greatly limit expression of downstream of downstream genes in the absence of SA. Furthermore, this system was shown to be highly specific to SA, with little activity shown in the presence of molecules structurally similar to SA such as vanillin or benzoic acid or in the presence of the WT-AraC inducer molecule L-arabinose [69]. While this system utilized SA to allow expression of a desired reporter protein, we will instead be taking this system and modifying it to control expression of a repressor protein which will subsequently repress expression of a fluorescent reporter protein.

The second previously validated component of this proposed expression system involves the chemically alleviated repression of a fluorescent reporter protein, specifically the repressed expression of enhanced yellow fluorescent protein (EYFP) in the presence of the repressor protein TetR. This system constitutively expressed the TetR inhibitor protein to inhibit expression of EYFP, then alleviated that inhibition via treatment with the TetR inactivator molecule anhydrotetracycline (aTc). This system was shown to increase EYFP expression with increased aTc concentrations, providing evidence that the TetR repressor effectively inhibits downstream gene expression while present and active, with these downstream genes subsequently show significant expression when active TetR is absent [70]. While this system utilizes an externally provided chemical to explicitly inactivate the TetR repression of EYFP expression, this system can be modified to instead stop total expression of TetR under specific chemical conditions, thus allowing TetR to naturally degrade and the subsequent expression of the TetR inhibited downstream genes.

The proposed expression system for high-throughput SAMT activity screening combines the two aforementioned expression systems into a single system in which the presence of SA will lead to TetR inhibition of mNG expression (Fig.14). Modifications to the SA induced expression

system will involve swapping out the expressed gene from a fluorescent reporter gene to the TetR repressor. With this modification, expression of the TetR inhibitor will only occur when the AraC-SA transcription factor has been activated by SA, with TetR expression levels decreasing as SA is converted to MSA by SAMT. The secondary component of this expression system will be functionally identical to the previously validated example in which TetR inhibits expression of EYFP, only instead of EYFP the fluorescent reporter gene will be mNG. By combining these two gene elements into a single expression vector, an expression system will be created in which cells with active SAMT will slowly become fluorescent while those with inactive will remain white. This will allow a rapid visual identification of cells with active or inactive SAMT variants via simple observation under UV transillumination.

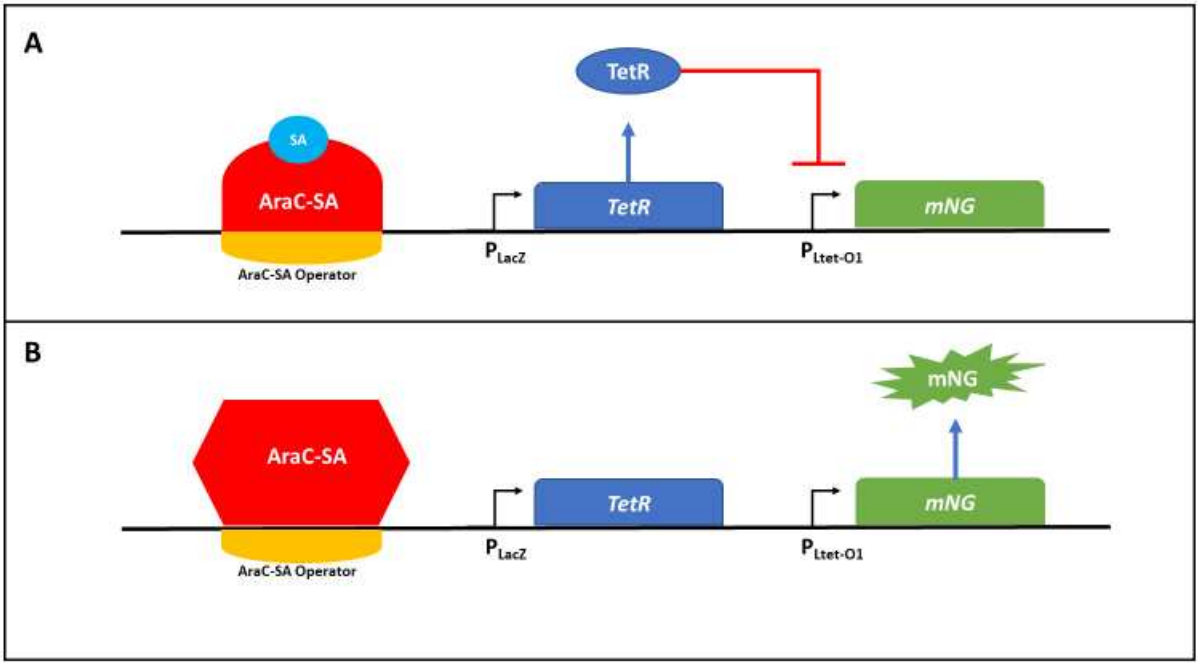


Figure 14: Schematic of the proposed SA inhibited mNG expression system for high-throughput screening for SAMT activity. **A)** When SA is present in the system, it will bind AraC-SA and cause AraC-SA to associate with the operator region upstream of the TetR gene. This induces expression of the TetR inhibitory protein and subsequent inhibition of mNG expression. This would be the expected scenario if SAMT was inactive and would result in cells remaining white in color. **B)** When SA is not present in the system, AraC-SA will unbind from the operator region and expression of TetR will cease. As TetR inhibitory effects cease, expression of mNG will begin and cells will develop green fluorescence. This would be the expected scenario if SAMT was inactive, allowing identification of cells with active SAMT based on the development of fluorescence.

With this expression system, the presence of SA will activate AraC-SA to induce expression of TetR and subsequently inhibit expression of mNG. As SAMT converts SA to MSA, AraC-SA will become inactive and TetR expression and subsequent mNG inhibition will cease, and cells will begin to turn fluorescent. In cells with inactive SAMT, SA levels will not deplete, and as a result of continued inhibition of mNG the cells will remain white. And theoretically, the rate at which a given cell converts SA to MSA would impact the rate at which mNG expression would occur, allowing the time it takes for a cell to display fluorescence to be used as a rough measure of enzyme activity for the different library members. This method should allow for a high-throughput screening method for SAMT activity by simply culturing and plating the library variant transformants on media with artificial supplementation of SA. This will ensure a constant presence of SA during the initial culturing process and the early stages of plate growth and prevent any premature expression of mNG. As the growth plates are allowed to incubate at growth temperature over an extended period of time, cells which contain active forms of SAMT should begin to display fluorescence as the TetR inhibition of mNG ceases. This would allow the plating and subsequent screening of potentially billions of library variants on a single growth plate, with screening facilitated via simple visual selection under UV transillumination. If this method is successfully implemented, it will facilitate high-throughput library screening, and allow far more ambitious library sizes to be investigated with minimal time or resource investment.

Creation of SAM Upregulated *E. coli* strains

As SAM is the methyl donor for the SAMT methylation reaction, a lack of bioavailable SAM would be expected to decrease overall SAMT activity. Support for this hypothesis comes from previous research into SAM-dependent methyltransferases and the effect of low SAM

bioavailability on ultimate enzyme product accumulation. This research investigated the activity of an O-methyltransferase that is responsible for the transformation of protocatechuate to the flavoring compound vanillate in an attempt to increase overall vanillate titers in *E. coli*. Ultimately, this research determined that the rate limitations in the O-methyltransferase were due to a lack of bioavailable SAM, with increases in bioavailable SAM leading to *E. coli*-based production systems in which the final titers of vanillate were significantly increased, while final titers of protocatechuate were significantly decreased [68]. The increases to bioavailable SAM were achieved in two different manners, both of which resulted in improved vanillate:protocatechuate titers, but each of which utilized different biochemical pathways to achieve the increases in bioavailable SAM.

The first method to increase bioavailable SAM involved artificially provision of amino acids which are involved in the endogenous pathway of SAM anabolism, namely methionine (Met) and homocysteine (Hcys). Because *E. coli* have shown an inability to take up SAM from the environment, artificial provision of SAM itself was not feasible; however, the ability of *E. coli* to take up amino acids from the environment allowed for artificial provision of these SAM precursors and resulted in increased overall vanillate titers. Further investigations into the artificial provision of Hcys showed that these overall increases in vanillate titers were not due to simple increases in overall biomass production, but actually due to increased concentrations of vanillate within the biomass itself. To further investigate these results, mutations were made to the normal methionine biosynthesis pathway which led to increased methionine biosynthesis. Specifically, these mutations were the deletion of the methionine biosynthesis regulation gene *metJ*, and the utilization of feedback inhibition repressed genes *MetE** and *CysE**. The *metJ* gene is involved in various regulatory mechanisms of methionine biosynthesis, and deletion of

this gene allowed increased overall methionine production and increased yields of both protocatechuate and vanillate but did not result in vanillate dominated product ratios. The genes *MetE* and *CysE* are responsible for initiating the methionine or cysteine biosynthesis pathways respectively and are both inhibited by high levels of either methionine or cysteine respectively. The variant genes *MetE** and *CysE** are versions which have had this feedback inhibition repressed, allowing for the buildup of higher levels of methionine and its precursor cysteine. Expression of these genes in conjunction with the deletion of *metJ* resulted in product titers that were both increased in total yield, and which showed the desired higher ratio of vanillate to protocatechuate. Considering that the only known relation between the O-methyltransferase reaction and methionine is the utilization of methionine-derived SAM as a methyl donor, these results were highly indicative of SAM bioavailability being the limiting factor in O-methyltransferase activity [68]. However, because this method focused on affecting methionine biosynthesis to theoretically effect SAM bioavailability, it did not explicitly show SAM bioavailability was increased in response to increased bioavailable methionine or that increased SAM bioavailability was explicitly linked to greater methyltransferase activity, but simply supported this hypothesis.

To investigate explicitly the role of SAM bioavailability on methyltransferase activity, a different set of mutations was used to directly increase SAM bioavailability via increasing the rate of recycling of SAH back into SAM. Endogenous SAH recycling occurs via a pathway known as the activated methyl cycle, and involves a number of different chemical intermediates including methionine and Hcys, and begins with the conversion of SAH to Hcys via either the action of SAH-hydrolase (*sahH*) or the duel actions of methylthioadenosine nucleosidase (*mtn*) and s-ribosylhomocysteinease (*luxS*). Mutations were made to upregulate either the *sahH* gene or

both the *mtn* and *luxS* genes, and investigations into final product titers were conducted both with and without the artificial provision of methionine. While the *sahH* upregulated version itself showed reduced vanillate titers compared to a wild-type control, the combined *mtn* and *luxS* upregulations resulted in increased production of both vanillate and protocatechuate, although with similar final concentrations of vanillate to protocatechuate. When artificial methionine supplementation occurred, both the *sahH* and the *mtn/luxS* upregulated strains showed increased overall vanillate titers compared to wild-type strains, and also showed significantly higher titers of vanillate relative to protocatechuate. These results demonstrate clearly that upregulation of the genes involved in recycling SAH to SAM has a significant effect on activity of SAM-dependent methyltransferases, and that product accumulation can be maximized when this upregulation is combined with an increased supply of bioavailable methionine. Furthermore, this research establishes that the limiting aspect of SAM-dependent methyltransferase activity is not due to limitations in the conversion of methionine to SAM, but due to limitations in the upstream methionine biosynthesis and SAH catabolic processes which provide the methionine precursors necessary for SAM biosynthesis [68]. Due to the fact SAMT is also a SAM-dependent methyltransferase, these same biochemical processes should have similar impacts of overall SAMT activity and MSA production.

Due to the general similarities in the SAM-dependent enzymatic mechanisms of O-methyltranferase catalyzed conversion of protocatechuate to vanillate and the SAMT catalyzed conversion of SA to MSA, it would be expected that SAMT and MSA production levels could be improved by applying the same genetic induced metabolic enhancers used to improve vanillate production by O-methyltransferase. As such, creation or procurement of an *E. coli* strain which incorporates the Δ*metJ*, *MetE*, CysE*, Mtn,* and *LuxS* genotypes should result in increased

bioavailable SAM due to deregulation of the upstream methionine biosynthesis inhibitory pathways in conjunction with the upregulation of the SAH recycling enzymes involved in the active methyl cycle. While these mutations could theoretically be induced in any common lab strain of *E. coli*, it would be ideal to create this strain in the *dam(-) E. coli* background due to this strain not using SAM as a methyl donor for *dam/dcm* methylase induced DNA methylation. However, due to the fact all enzymes involved in the aforementioned pathways are already coded for by endogenous *E. coli* genes, incorporation of these mutations into new strains of *E. coli* should be relatively simply without the need for introduction of any novel non-endogenous metabolic pathways.

CHAPTER 3: COMPUTATIONAL ENZYME DESIGN

Introduction

As the crux of the odor encoder project is the computational design of a theophylline binding cavity within SAMT which can transmit allosteric changes that result in the resurrection of a crippled SAMT mutant, the computational design component was integral to overall project success. While the theophylline inducible riboswitch demonstrated successful theophylline dependent MSA production from our engineered *E. coli*, there are minor issues with the riboswitch regulated expression system such as the limited MSA production compared to constitutively expressing SAMT cultures or the delay in MSA production that is inherent to riboswitches in general. As the riboswitch method depends upon theophylline binding to the riboswitch prior to the translation of SAMT mRNA into a functional enzyme, there is an inherent delay in MSA production due to the time required to translate the SAMT mRNA into functional enzyme and express a sufficient concentration of SAMT to produce appreciable levels of MSA. This principle was supported by observations from the enzyme activity assays which demonstrated significantly olfactorily detectable levels of MSA production after approximately 14 hours for cultures with constitutive expression of SAMT, while theophylline induced expression cultures required approximately 24-36 hours to produce mildly olfactorily detectable levels of MSA.

One of the benefits of producing an allosterically activated version of SAMT is that this time delay should be mitigated as there would be constitutive expression of the crippled SAMT mutant which would allow for a pool of crippled enzyme to be produced and held ready for future theophylline induced allosteric activation. This should allow for a more rapid response

time due to the lack of any need for additional protein expression between the theophylline exposure and the enzymatic production of MSAand produce a stronger olfactory signal than the theophylline riboswitch regulated expression. In addition, this system should produce MSA at a similar level as that seen in the constitutive expression cultures if theophylline binding is able to restore activity to wild-type levels due to the relatively high number of SAMT enzymes which this system would produce relative to the riboswitch activated system.

In order to create such a system, it is necessary to design an allosteric activation system which is highly specific to an analyte of interest, and the best approach to facilitating such a design is via use of common combinatorial design programs such as Rosetta or PyMol. These programs work via underlying Python codes and are able to perform a variety of modelling algorithms to design proteins with a variety of desirable traits. For the odor encoder project, these computational design methods were utilized to design the potential theophylline allosteric cavities and to model potential allosteric behaviors in these mutant versions of SAMT both with and without theophylline present. By using combinatorial optimization design methods to determine which mutations would be optimal for facilitating the binding of theophylline into an artificially produced cavity space within SAMT, a library of potential SAMT mutants was determined which demonstrated both the potential for theophylline to bind the designed allosteric cavity within SAMT, and potential perturbation of the SAMT backbone structure to an extent that would potentially facilitate theophylline induced allosteric recovery.

Materials and Methods

Mutation Palette Selection

An allowable mutation palette for each wild-type residue was determined using a combined volumetric and biochemical approach. The volumetric cutoff allowed no residues with volume greater than 5Å^3 larger than the respective wild-type residue. The biochemical limitation prevented non-charged residues from mutating into charged residues or visa-versa. Table 4 lists the mutation palettes for each wild-type amino acid.

Table 4. Mutation Palettes for each Wild Type (WT) Amino Acid

WT ID	Allowable Mutation Palette (Single letter code)
Gly	G
Ala	GAS
Val	GASPNTV
Cys	GAS
Pro	GASPT
Leu	GASPNTVQHMLI
Ile	GASPNTVQHMLI
Met	GASPNTVQHMLI
Trp	GASPNTVQHMLIFYW
Phe	GASPNTVQHMLIFY
Ser	GAS
Thr	GASPNT
Tyr	GASPNTVQHMLIFY
Asn	GASPNT
Gln	GASPNTVQ
His (+)	GASPNTVQHMLIK
Lys (+)	GASPNTVQHMLIK
Arg (+)	GASPNTVQHMLIKR
Asp (-)	GASDPT
Glu (-)	GASDPNTEV

Theophylline Pocket Modelling

Custom Python scripts were created to model potential theophylline binding pockets within SAMT. A 3D-grid representation of theophylline was created and rotated 360° in the x, y, and z dimensions to create a 3D-grid model of all possible rotational positions theophylline could take

on. Another custom Python script was used to create a 3D-grid representation of the cavity space within a fixed position SAMT backbone structure via mutating all residues to glycine and turning the grid off in locations where backbone atoms exist, as well as anywhere within 4Å of the substrate or coenzyme. Another custom Python script created a convex hull around the enzyme surface, with the points of the hull extending from the alpha carbons on surface residues. Any points lying outside this hull, or closer to the hull exterior than any protein alpha carbon, were also excluded from the final 3D-cavity grid. The 3D rotational model of theophylline was then docked within this grid and a complementation score was assigned to each individual theophylline pose. Poses which showed stability in a given location within the 3D-cavity grid were maintained while those which were unstable were eliminated, with the clusters of remaining poses constituting the individual theophylline pockets. Residues proximal to the theophylline poses for a specific pocket were selected based on relative distance to the theophylline molecule, with a 5Å distance used as a cutoff for proximity. In addition, residues constituting a given theophylline pocket were confirmed via a visual inspection in PyMol. Residues targeted for mutation in each theophylline pocket can be found in table 5 below.

Table 5: SAMT Residue Ids for Each Theopyhlline Pocket

Pocket	Pocket Residue IDs
A	2, 4, 8, 151, 155, 226, 251, 254, 315, 319
B	166, 226, 229, 233, 248, 251, 253, 308, 312, 315, 316, 320, 328, 329, 332
C	166, 226, 229, 230, 233, 236, 243, 248, 251, 253, 312, 316, 319, 320, 324, 328
D	55, 67, 70, 71, 74, 92, 94, 96, 107, 112, 114, 117, 121, 123
E	67, 70, 71, 74, 92, 94, 96, 107, 111, 117, 121, 123
F	56, 58, 134, 135, 140, 143, 191, 195, 198, 199, 205
G	130, 145, 149, 188, 191, 192, 205, 207, 209, 262, 266, 272, 352, 354
H	149, 184, 188, 191, 192, 207, 209, 258, 262, 265, 266, 350, 352, 354
I	40, 43, 44, 49, 53, 55, 77, 90, 92
J	29, 32, 36, 40, 49, 69, 73, 76, 142, 144, 206, 208

Computational Library Determination

A mutation pallet for each wild-type residue was determined based on imposed volumetric and biochemical limitations. All theophylline poses from each pocket were modelled 10 times in Rosetta. Combinatorial optimization compared all possible side chain rotamers and all permitted mutations to determine which residue mutations would best stabilize that theophylline pose. Mutations suggested more than 5% of the time were included in the final library design. At locations targeted for mutation, wild-type residues were excluded from the library design unless explicitly stated. A custom Pareto optimization analysis program was run on the designed libraries to produce a set of Pareto optimized libraries for each theophylline pocket, and Pareto optimized libraries was selected based on expected screening throughput limitations. An initial combinatorial optimization which utilized the same comparison of mutations and rotamers as described for the Rosetta designs but with the backbone atoms fixed in place was performed using the systematic hierarchical algorithms for rotamers and proteins on an extended network (SHARPEN) computational design program [22]. Modelling in SHARPEN was performed once for each theophylline pose, and a selection of the resultant SHARPEN designed SAMT mutants were utilized as base structures for subsequent modelling of allostery in Rosetta.

Modelling of Active Site Perturbation

From the SHARPEN designs, a singular representative theophylline pose and was selected from the center of each individual theophylline pocket cluster, and the mutant designed for this theophylline pose was used to investigate potential pocket allostery. The theophylline molecule was allowed to dock anywhere within the space for that specific pocket cavity, and combinatorial optimization of theophylline docking position and SAMT side chain rotamers was performed 100x times. The 100 modelled SAMT structures were relaxed and aligned with the WT-SAMT

structure, and for each structure an RMSD was determined for the alpha and beta carbons of the specified active site residues. Next, the same SHARPEN designed mutant SAMT sans theophylline were modelled 100x, and these 100 structures were relaxed and aligned to the WT-SAMT structure and the RMSDs were determined. The total set of active site residues which were targets for investigation were residues 10, 25, 57, 65, 98, 99, 129, 130, 147, 150, 151, 210, 225, 226, 255, 308, 311, 347. The set of determined RMSDs for a given pocket were compiled into comparative histograms, and a two-sided Welch's t-test was conducted on the RMSD comparative histograms to investigate for statistically significant differences in mean RMSD between SAMT with theophylline absent versus with theophylline present, with P-values < 0.05 considered statistically significant. An ANOVA F-test was conducted to investigate for statistically significant differences in RMSD variance between SAMT with theophylline absent versus with theophylline present, with P-values < 0.05 considered statistically significant. Statistically significant differences were considered to be evidence of significant active site perturbation dependent on the presence of theophylline. Statistical tests were performed using the SciPy Python stats package, specifically scipy.stats.ttest_ind for the Welsh's t-test and scipy.stats.f.cdf for the ANOVA F-statistic . Comparative histograms for each pocket can be found in the Appendix.

Results

Comparative RMSD Modelling of Allostery

Results of the two-sided Welsh's t-test showed significant differences between the means RMSD when theophylline was present and when it was absent for pockets A, E, F, and I. The results of the F-test showed significant differences between the RMSD variance when theophylline was present and when it was absent for pockets A, B, E, F, G, and I. Combining

these results indicates that pockets A, E, F, and I have the greatest potential for theophylline dependent allosteric perturbation of the key active site residues in SAMT. The p-values for the F-test and Welch's t-test, as well as the absolute RMSD variance and mean RMSD both with and sans theophylline are provided in table 6 below, in which pockets with both significantly different mean RMSD and RMSD variances are indicated in green.

Table 6: p-Values for the F-test and Two-sided t-tests analyses of the RMSD Comparative Histograms

Pocket	F statistic p-value	Welsh's t test p-value	variance no Theo	variance w/ Theo	mean no Theo	mean w/ Theo
A	0.00167427	0.00014020	1.21231668	0.66832921	2.66264981	1.33601968
B	0.00321162	0.91741828	0.46022626	0.26482456	0.86811359	0.87293142
C	0.65234996	0.07419014	0.28569315	0.30915933	1.00310400	0.92987676
D	0.98180348	0.18046747	0.35852464	0.54742784	1.00544613	1.10034734
E	0.00007077	0.01410311	0.84424979	0.38845788	1.14626256	0.88912952
F	0.00079571	0.00470319	0.41898229	0.22056264	1.03810689	0.90004212
G	0.00000098	0.07867567	0.70979933	0.26711385	1.03294630	0.89547734
H	0.09153851	0.12679434	0.42608403	0.32569097	1.05818553	0.97321131
I	0.00000000	0.00023248	0.54087467	0.15039226	1.05296187	0.82982271
J	0.00357	0.000908	0.32080319	0.161139175	0.96884751	0.85298143

Enzyme Library Computational Design

The number of designs for pockets A-J were 90, 540, 250, 60, 110, 160, 830, 340, 50, and 20 respectively. After removal of mutations suggested less than 5% of the time, the total number of possible libraries for pockets A-J ranged from as low as 27 for pocket-J to as high as 452,168,325 for pocket-G. Subsequent Pareto optimization reduced the number of possible libraries by including only those which were Pareto optimized. The total number of Rosetta designed variants, number of potential libraries, and total number of Pareto optimized libraries for each pocket are detailed in table 7 below. Note that the total number of potential libraries listed in the aforementioned table are the total number *after* the mutants which occurred at a

frequency of less than 5% were removed due to those mutations which occurred at a frequency of less than 5% not being considered as potential library constituents.

Table 7: Total Number of Potential Libraries for Each Pocket			
Pocket	**Number of Modelled Variants**	**Total Number of Potential Libraries**	**Number of Pareto Optimized Libraries**
A	90	99225	20
B	540	12252303	31
C	250	13395375	32
D	60	972405	20
E	110	205065	16
F	160	4306365	33
G	830	452168325	53
H	340	415256625	45
I	50	3444525	24
J	20	27	3

Pareto optimized libraries will be selected for each pocket based on expected screening capabilities at the time of purchase, with larger libraries being possible once higher throughput screening method is achieved via construction of the SA-repressed mNG expression plasmid. Final library sizes for any pocket containing W226 will be twice as large as indicated by in the Pareto optimized library due to the inclusion of the wild type Trp residue in addition to the Rosetta determined mutations, and as such smaller Pareto libraries will be chosen for these pockets. Figure 15 contains a sequence logo for the complete set of 540 modelled variants for pocket-B and a sequence logo for the same set of 378 modelled variants which remained after discarding mutations suggested at a frequency of less than 5%. Figure 16 shows sequence logos for 5 different potential Pareto optimized libraries for pocket-B, with these 5 Pareto optimized libraries including both the smallest and largest possible Pareto optimized libraries, as well as 3

intermediate sized Pareto optimized libraries which would be within the size range plausible for investigation with current screening methods.

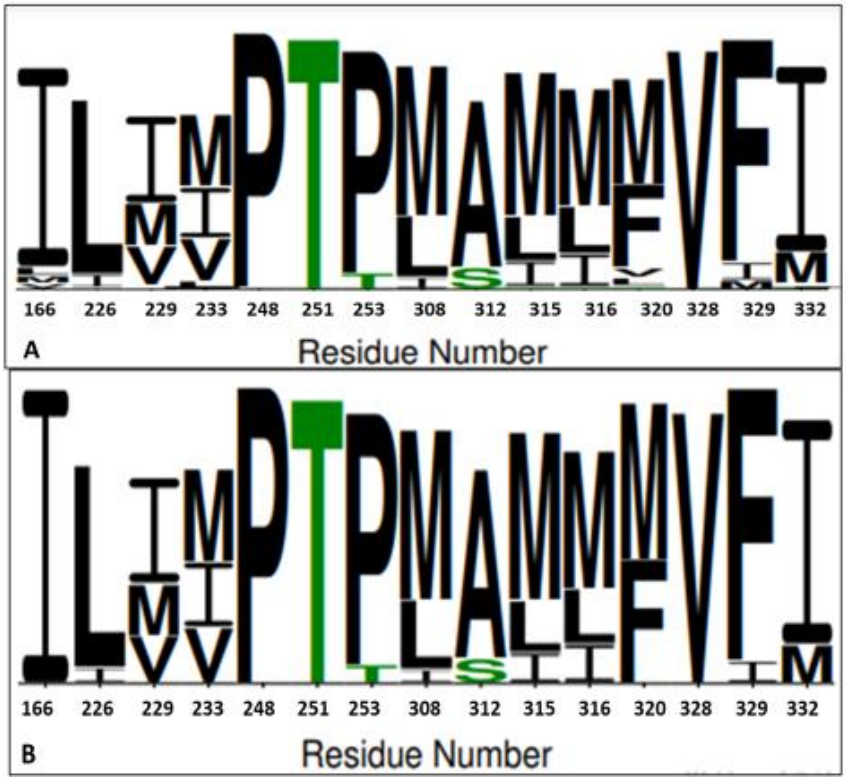


Figure 15: Sequence logos for the entire computatinally determined set of mutations for pocket-B. Image (**A**) shows a sequence logo for the enitre original set of computatinally determined mutations without any reductions, while image (**B**) shows the same set of computationally determined mutations but after those mutations suggested at a frequency les than 5% were eliminated. The relative heights of each letter are representative of the relative proportion at which that mutation was suggested for at that residue site. For example, in (B) the ratio of M/I/V suggested for residue 233 were relatively equal, while the ratio of P/T suggested for residue 253 shows a greater protportion of P than T. The letters which are so small as to be effectively illegible in (A) reprent mutations suggested at very low frequency, and which are absent from logo (B) as a result of the elimimation of low frequency mutations. Logos made via Web Logo 3 [77].

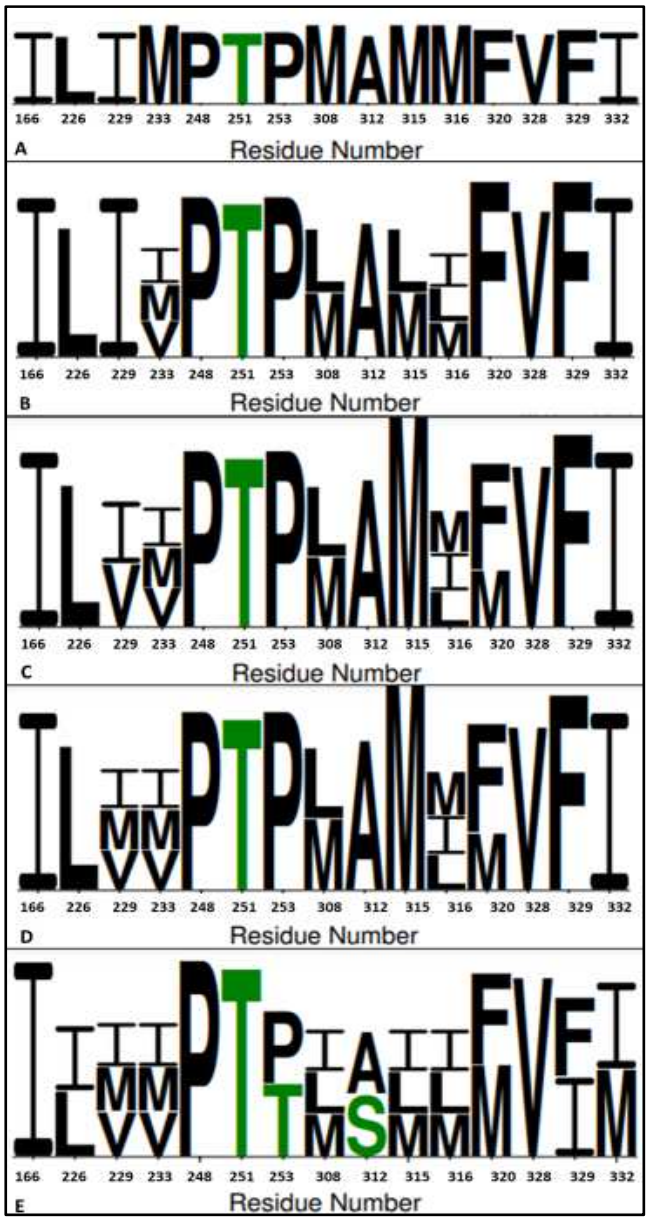


Figure 16: Sequence logos for 5 different Pareto optimized libraries for pocket-B. Image (**A**) shows the sequence logo for the smallest possible Pareto optimized library which would result in a single varient library which would cover 24 of the 378 variants from the reduced original library. Image (**B**) shows the sequence logo for the smaller Pareto optimized library that is reasonable with current screening limitations. This would result in a library of 32 varients which would cover 102 of the 378 variants from the reduced original library. Image (**C**) shows the sequence logo for the medium sized Pareto optimized library that is reasonable with current screening limitations.This would result in a library of 72 varients which would cover 140 of the 378 variants from the reduced original library. Image (**D**) shows the sequence logo for the larger sized Pareto optimized library that is reasonable with current screening limitations.This would result in a library of 108 varients which would cover 164 of the 378 variants from the reduced original library. Image (**E**) shows the sequence logo for the largest possible Pareto optimized library which would not be reasonable with current screening limitations.This would result in a library of 15,552 varients which would cover all 378 variants from the reduced original library. It should be noted that the discrepancies between column height between residue numbers are not indicative of sequences having discrepancies between sequences sizes as all designed sequences spanned the entirety of the SAMT sequence. It should be noted that for residues which show multuple suggested mutation types the heights of the letters in the logo are the same. Logos made via Web Logo 3 [77].

It should be noted that the sequence logos for the reduced original library and the largest possible Pareto optimized library have the same variations indicated at each residue location; however, the sequence logos for the Pareto optimized libraries do not indicate the ratios at which the mutations were suggested while the sequence logos for the complete and reduced original libraries do indicate the ratios at which each mutation type was suggested. This is because any mutation suggested at a certain residue site at a frequency of greater than 5% was included in the final library design which resulted in degeneracy within the reduced original library. However, the subsequent Pareto optimization scripts removed these degenerate variants from the final designs and ensured every variant in the Pareto optimized library will be unique. As such, in a Pareto optimized library for residue sites containing variation the ratios of different mutations will inherently be equal. This degeneracy in the reduced original library is why the number of variants included in a Pareto optimized library can cover a larger number of the total variants from the reduced original library. Table 8 contains one Pareto optimized library example for each pocket, with that example being the one most in line with the current screening methods, with the Pareto optimal library for pocket B being the same one as shown in Figure 16-c. Mutations listed in this table correspond to the same residue IDs in the same numerical order as described in table 5. A complete list of all Pareto optimized libraries for each pocket can be found in the Appendix (Appendix Table 3).

Table 8: Example Pareto Libraries for Each Pocket

Pocket	Pareto Optimized Library Example	Total Number of Library Members	Number of Designed Variants Covered in Library
A	(('G', 'T'), ('I', 'L'), ('I', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	48	70
B	(('I',), ('L',), ('V', 'I'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('M',), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	72	140

C	(('I',), ('L',), ('I', 'L'), ('A',), ('V', 'I'), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V', 'I'), ('V', 'I', 'M'), ('I',), ('V',))	24	38
D	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V', 'L'))	18	16
E	(('L',), ('V',), ('T',), ('V',), ('F', 'Y'), ('M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I', 'L'))	32	72
F	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('I', 'M'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('I',))	48	60
G	(('I',), ('A',), ('L',), ('I', 'M'), ('L', 'M'), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	72	142
H	(('I', 'M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	72	143
I	(('I',), ('I', 'L'), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'L'), ('M',), ('V',), ('T',), ('M', 'L'), ('M',))	32	25
J	(('M',), ('T',), ('A',), ('I',), ('V',), ('A',), ('I', 'L'), ('T',), ('M', 'L'), ('I', 'L'), ('V',), ('T',))	8	20

Discussion

Library Design Rationale

With the previously discussed success at allosteric engineering by Deckert et al. as our inspiration, our approach focused on attempting to cripple the active site of SAMT via targeted mutations in such a way as to allow restoration of enzymatic structure and function upon binding of theophylline. Deckert noted that they could only achieve enzymatic restoration when their mutations occurred in what they referred to as “second-shell” residues, or residues which were directly involved in active site catalysis, but which were proximal enough to active site residues to cause a disturbance to the active site structure upon mutation. In contrast, mutations which targeted active site residues were essentially fatal to the enzyme, with no recovery observed even with saturation levels of indole [15]. For this reason, residues targeted for mutation were primarily focused on these second-shell residues in SAMT, with attempted conservation of residues believed to be essential to active site catalysis. Due to this precaution, second-shell

residues targeted for mutation did not have the wild-type residue included in the final designed library, as it was believed that these locations could withstand any incorporated mutations without losing the potential for overall enzymatic recovery. For residues which were believed to be viable mutation targets, but which also were suspected to be important to active site catalysis, both the wild-type residue and the Rosetta-suggested mutant residue were included in the designed library. For example, the Rosetta output for the Pocket-A designed library consistently suggested a W226L mutation; however, due to the suspected high importance of Trp226 in active site function, both Trp226 and Leu226 were included in the final library design.

Deckert solely utilized Trp to Gly mutations due to the fact it was necessary to make enough space in their designed cavity for indole to successfully enter and bind. Due to the fact indole is essentially a structural analog for the Trp sidechain, this highly aggressive mutation was needed to ensure enough space was created to allow binding of indole in the created cavity [15]. Following this approach, we opted to limit our mutation pallet in a volumetric manner to ensure theophylline had sufficient space for binding into the ultimate designed cavity. Due to the fact theophylline is not an exact structural analog for a specific amino acid, there was no single amino acid deletion which could have singularly accommodated the theophylline molecule, and thus any cavity that could fit theophylline would necessarily require the mutation of multiple amino acids. Due to this necessity, it was possible to be less aggressive in our singular mutations than Deckert, and instead incorporate amino acids which were larger than Gly but still within the volumetric constraints necessary to accommodate the theophylline molecule.

Computational Library Design

Computational design algorithms were designed in house by Dr. Christopher Snow and PhD. candidate Jacob DeRoo. These design scripts were created in Python and utilized Rosetta

for rapid investigation of the SAMT structure and its wild-type residues with respect to potential changes in residue type and rotational isomer (a.k.a rotamers) [67]. Initial custom Python scripts allowed for an investigation of the locations within the SAMT backbone which could spatially accommodate a theophylline molecule, with these locations considered as possible locations for the introduction of a theophylline binding cavity. Subsequent Rosetta Python scripts were used to design a combinatorial optimized mutant library, as well as model potential allostery for the different theophylline pockets. This information was then utilized to determine the final mutant libraries which would be investigated for functional dependency on theophylline in the designed SAMT variants.

First, a series of custom Python scripts were created to investigate where theophylline could potentially fit within the *C. breweri* isoform of SAMT (PDB ID: 1m6e). The first script took the spatial coordinate information for a theophylline molecule and created a 3D grid around the molecule. This 3D grid was designed such that any point within 1.8 Å of a theophylline atom was included in the grid, while those points greater than 1.8 Å from any theophylline atom were excluded from the final grid. The next script took the 3D grid representation of theophylline and rotated that grid 360° in the x, y, and z directions to make a 3D gridded representation of all rotational orientations a single theophylline molecule could exist in. The third script created a 3D grid representation of potential space within the SAMT backbone to examine where theophylline could possible fit within SAMT if side chains were absent but the backbone was fixed in place. This was achieved by first mutating all residue in SAMT to glycine, then creating a 3D grid representation of the resultant cavity space, with the grid active in locations without any backbone atoms but turned off wherever a backbone atom existed.

In addition, a 4 Å exclusion zone was created around SA and SAE to ensure our modelling of potential theophylline binding locations did not place theophylline too close to the substrate or the coenzyme. To ensure the grid was solely representative of potential cavities within SAMT and exclusive of any surface clefts, a 3D convex hull was created around the alpha carbons of SAMT, and any locations which either fell outside of this convex hull or were closer to the exterior of the hull than to any alpha carbon in SAMT were excluded from the final determined 3D cavity grid. The final resultant 3D grid described locations within SAMT where theophylline could potentially fit with the proper sidechain mutations. The final part of this process was modelling docking of the theophylline molecule within the 1m6e cavity grid space to investigate how complementary a theophylline molecule is to a theoretical position within the cavity. This was achieved by scoring how complementary the theophylline molecule grid was to a given location within the SAMT cavity grid. The complementary score is indicated by the coloration of the theophylline molecules in figure 17 below, where more blue coloration is indicative of greater grid complementation.

Investigation of these modelling results demonstrated that there were approximately 6 locations within the backbone of 1m6e in which theophylline molecules could theoretically fit given the previously discussed imposed limitations. These pockets were further broken up into 10 final pockets, denoted as pockets A-J, with some pockets being created by segregating the larger pockets into smaller sub-pockets based on the proximity of the theophylline poses to groups of 1m6e sidechains and other theophylline poses. While this modelling was able to demonstrate where theophylline could potentially bind within the backbone of SAMT, it did not provide any information about whether these locations had potential allosteric interference with

the active site; however, it was the first step in determining where potential induced allosteric cavities could be designed within the SAMT structure.

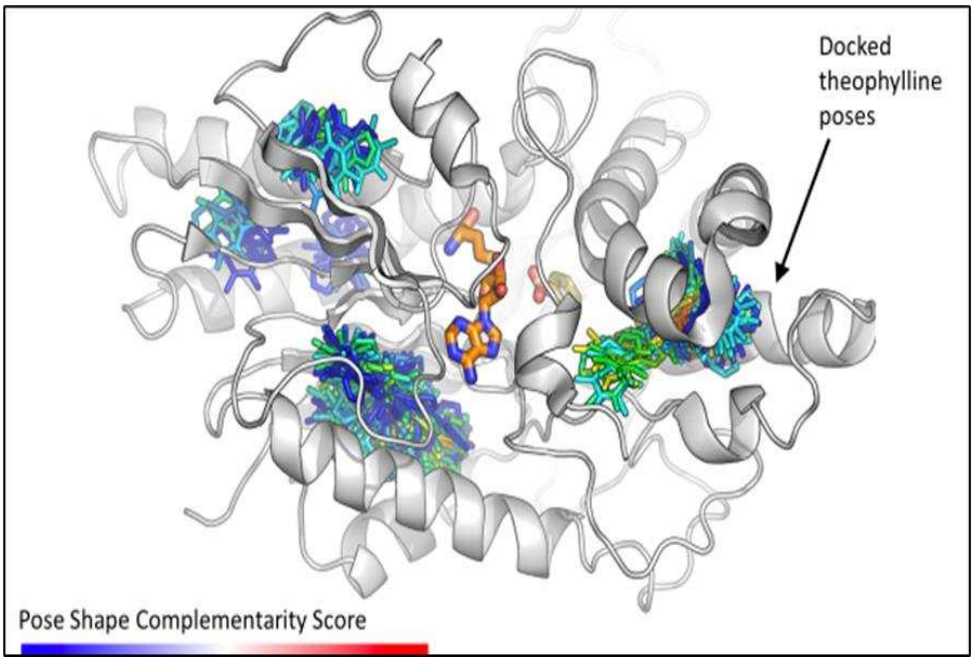


Figure 17. Computationally determined theoretical theophylline docking poses within the backbone of an 1m6e structure in which all residues had been minimized to glycine. The 5 clusters of theophylline poses shown were subsequently broken down into the 10 individual pockets A-J via a manual investigation of individual theophylline pose and 1m6e wild-type residue proximities.

Once theoretical docking positions had been determined, a custom Rosetta script was used to determine which SAMT residue mutations and sidechain rotational orientations would best accommodate a given theophylline pose. This script applied combinatorial optimization to compare all possible side chain rotamers, as well as all possible residue mutations for residues targeted for mutation as allowed by the mutation pallet for that wild-type residue type. This comparison was based on which mutations and sidechain rotamers would best decrease steric clashing between SAMT and a given theophylline pose. The modelling was then performed 10 times for all theophylline poses for a given pocket, and the set of suggested mutations for each targeted residue within that pocket were compiled for further analysis.

For target residues which were designed with a single consistent mutation, that mutation was necessarily included in the final library design; however, for target residues which were designed with multiple suggested mutations, a decision needed to be made as to which mutations would be included in the final library and which would not. It would be possible to include every suggested mutation variant; however, this would increase the number of degenerate library members and cause a combinatorial explosion in which the number of ultimate library members would simply be too great to realistically screen. While this approach would ensure total coverage of all designed library members, and thus theoretically increase the chances of a observing a successful library member, it would also create a library size so large that screening would not be feasible given the financial and temporal limitations of the project. In contrast, the library could have been limited to a minimal number of allowable mutations to decrease the total number of library members to levels which could be easily screened; however, this this would come at a cost of overall library coverage as it would only produce a single library variant, and as a result would greatly decrease the chances of having successful theophylline induced SAMT recovery. As limitations in total library size due to screening throughput limitations made it infeasible to create and investigate all possible designed mutations, only mutation which were suggested in at least 5% of the designs were maintained as potential mutations to decrease the overall number of library variants.

To optimize the effectiveness of the screening process, a Python script was designed which could determine the Pareto front with respect to library member coverage and overall library size. When looking at a scenario in which two parameters are considered highly desirable, the Pareto front will be the set of parameter pairings which optimize both parameters to such an extent that increase of one parameter would result in a decrease in the other parameter [32]. By

only considering mutation combinations which would result in Pareto optimized libraries, it was ensured that the final library size would be small enough to reasonably screen while maximizing the coverage of the library members for that given library size. Figure 18 shows the Pareto front determined for pocket-B, with three examples of Pareto optimized libraries indicated in red. As can be seen, the left-most indicated point is representative of a library which would cover 112 out of 378 Rosetta-designed library members (including some degenerately designed members) whilst only requiring a total library size of 36. The center most indicated point is representative of a library in which 140 out of 378 possible library members would be covered in a library with a total size of 72, while the right most indicated point is representative of a library in which 164 out of 378 possible library members would be covered in a library with a total size of 108. Because standard library screening procedures involve screening each member in triplicate, these three potential libraries would require 108, 216, or 324 individual sample screens for the respective potential libraries. In such a scenario, considerations must be made regarding the accuracy, precision, time, and financial costs of the screening method, and whether the costs of screening is worth the increased chances of successful screening hits.

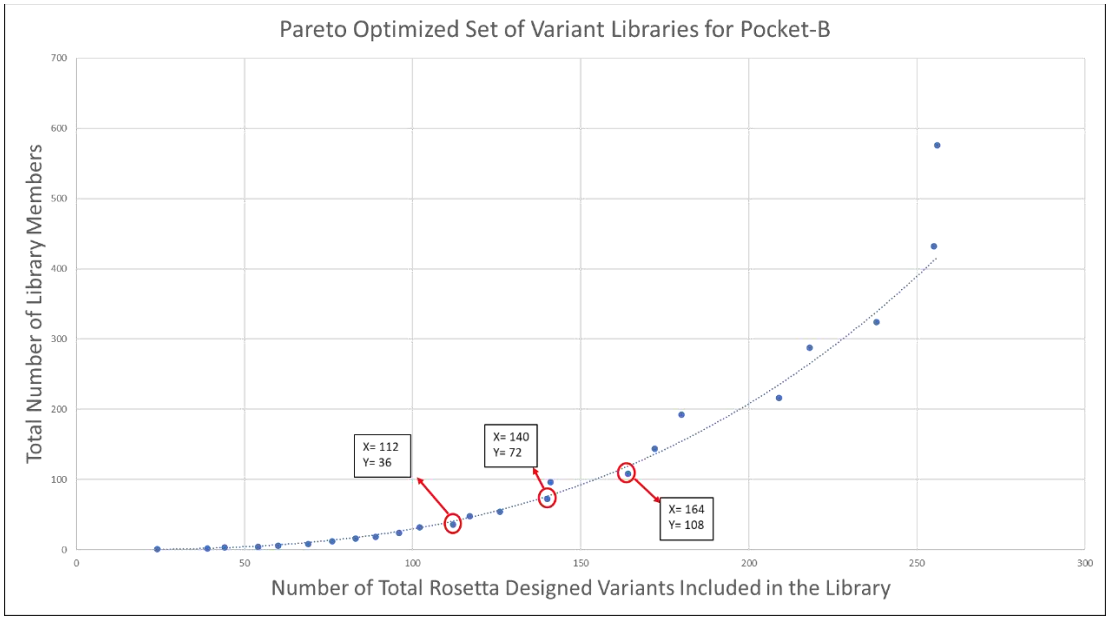


Figure 18. The Pareto optimized set of potential libraries for Pocket-B. The left-most indicated point represents a library of size 36 which would cover 112 out of 378 proposed library members, while the right-most point represents a library of size 108 which would cover 164 out of 378 proposed library members. The library which would correspond to the left-most point would be easier to screen but would have a lower chance of observing a successful screen, whilst the library corresponding to the right-most point would have a greater chance of a successful screening event but would require more time and resources to fully screen.

Innovative Design Aspects

One of the innovative aspects of the computational approach to this project was the imperfect structural match between our analyte molecule and the residues targeted for mutation, and how that impacted the approach at creating the theophylline accommodating pockets. Previous allosteric engineering approaches, such as the previously mentioned work by Deckert et al. and by Rana et al, tended to utilize highly specific residue targeting to explicitly accommodate their novel allosteric analytes. Deckert was able to mutate a single tryptophan residue to glycine with confidence this mutation would create enough space for indole to bind in the space formally occupied by tryptophan, while Rana was able to substitute explicit loops in the transmembrane proteins from those known to be specific to Na^+ to those which were known to be specific towards K^+ [15,47]. In both these cases, the structural changes which were necessary to achieve their modified allosteric abilities were already known, and thus the design had very limited wiggle-room in which to vary their designs. In contrast, theophylline's lack of an amino acid structural analog meant that wiggle room was seen as advantageous as it would theoretically allow more freedom for theophylline to move into the designed cavity. This approach, deemed a "fuzzy design", seems to be unique within the currently published computational designed allosteric regulation approaches.

Modelling of Potential Allostery

Due to the fact the allosteric switch is dependent upon movement within the active site when theophylline binds to the allosteric site, having confidence that there was at least some significant likelihood of this occurring was important prior to investing time and resources into production and screening of the library for a given pocket. This was why the decision was made to model each pocket in Rosetta and investigate for movement within the active site when theophylline was or was not present. This process was performed using two custom Rosetta scripts which investigated potential active site disturbance and recovery upon mutation of wild-type residues and subsequent binding of theophylline. This was made possible due to the ability of Rosetta to model mutations, the predicted movement of the SAMT backbone, and the presence of three small molecules simultaneously. Since our allosteric switch design is dependent upon allosteric binding of theophylline inducing a conformational change in the active site of SAMT that would restore enzyme activity, knowing whether a specific theophylline binding event could plausibly induce this change was important in predicting a given theophylline pocket's possible allosteric abilities.

The allostery investigating Rosetta script examined structural differences between critical active site residues in a Rosetta designed SAMT mutant when theophylline either was or was not bound to the designed allosteric cavity. This script investigated the RMSD between the backbone alpha and beta carbons in specific active site residues of SAMT either with or without theophylline present. The reason RMSDs were determined by looking at perturbations in the alpha and beta carbons of the backbone and not the entire sidechain is due to the fact it is backbone motion which is truly necessary for transmission of allostery. First, the spatial information for the SA substrate and the coenzyme s-adenosylhomocysteine (SAH) – the byproduct of SAM demethylation which was utilized by Zubieta to elucidate the original SAMT

crystal structure – were extracted from the published 1m6e structure. These small molecules were then spliced into the active sites of the Rosetta designed mutant SAMT structures. It should be noted that there were slight perturbations of the exact position of these molecules in the SAMT mutants relative to their positions in the published SAMT structure, but these perturbations were not dramatic enough to be considered issues of concern.

A singular representative theophylline pose was selected from the center of each individual theophylline pocket cluster to be used as the representative pose for that specific pocket. This theophylline molecule was then allowed to dock anywhere within the specific pocket cavity, and combinatorial optimization of theophylline docking and side chain rotamers was performed 100x times. Due to the fact theophylline was allowed to move around in the docking modelling events, and because there were 100 docking events modelled per pocket, the single representative theophylline pose was used in place of analyzing all poses from a pocket due to computational limitations and to prevent combinatorial explosion. The 100 designed mutant SAMT for a given theophylline pocket pose were relaxed and aligned with the WT-SAMT structure, and an RMSD was determined for a select group of critical active site residues. Next, the same mutant SAMT, but this time with theophylline removed, was modelled 100x, those resultant structures were relaxed and aligned to the same WT-SAMT structure, and the RMSD was determined. This process was repeated for each theophylline pocket, and the RMSD values for a give pocket were compiled into a histogram. RMSD histograms which indicated a significantly higher RMSD when theophylline was absent compared to when theophylline was present were taken to be indicative of potential for allostery in that theophylline pocket.

Active site residues that were investigated for potential allosteric translocation were 10, 25, 57, 65, 98, 99, 129, 130, 147, 150, 151, 210, 225, 226, 255, 308, 311, 347. Because these

residues are active site residues that were deemed highly critical to substrate or co-enzyme binding, significant perturbation of these residues upon the binding/unbinding of theophylline was deemed to be evidence of a high potential for translating allosteric effects to the active site. Residues closer to SA were considered better options for potentially inducing allostery due to the fact SA makes fewer natural contacts with SAMT than does SAM. For this reason, it was suspected that effective disruption of SA binding would be easier to achieve than would disruption of SAH binding. As such theophylline pockets with evidence of potential allostery in residues in contact with SA were considered to have greater potential than pockets which affected SAH contacting residues.

The results of this comparative histogram of RMSD values indicated that pockets A, E, F, and I demonstrated the greatest perturbation of the active site residues upon theophylline binding and were thus determined to have the greatest potential for allostery according to the modelling performed. However, this is only a modelling event, and as such it does not supersede basic visual investigations of the theophylline pockets within SAMT which can demonstrate aspects of the protein structure which cannot be properly modelled computationally. As such, despite pockets A, E, F and I showing the greatest potential for allostery in the RMSD analysis, pockets such as pocket B and H which demonstrated relative proximity to the active site and seemed to have plausible routes for transmission of allostery when viewed in PyMol were considered more viable for transmission of allostery and were thus still dominant in the initial investigation into pocket allostery. These potential allostery routes which were observed in PyMol will be discussed in greater detail in the next section.

Initial Investigation into Pocket Allostery

Due to the high financial and resource costs associated with the procurement and screening of a full enzyme library, it would be prudent to perform a preliminary investigation into each pocket's potential for transmitting significant active site allostery. For the overall design to work, the initial mutations must disturb the active site to such an extent that enzyme activity is crippled. If the induced mutations do not lead to a crippling of enzyme activity, the ability to accommodate theophylline binding is irrelevant as the enzymatic production of MSA would occur both with and without theophylline being present. As such, the ability for a given theophylline pocket to transmit allostery to the active site can be initially investigated via screening a select number of sample library variants from each pocket and observing for disruption of enzymatic activity. If screening the representative fraction for a given pocket demonstrates loss of enzyme activity, this would be taken as a sign that this pocket location was indeed able to affect active site catalytic abilities and was thus a candidate for allosteric recovery. If, in contrast, screening the representative fraction does not demonstrate clear loss of catalytic abilities, this would be taken as a sign that this pocket was less likely to transmit allostery to the active site and was not as viable of a target for induced allostery.

From all 10 theophylline pockets a total of 30 unique SAMT variants were selected, with at least one variant included from each pocket. To determine which specific pocket variants would be selected for these initial 30 variants, the results of the Rosetta allostery modelling was used in conjunction with a visual inspection for potential allostery routes in theophylline-bound SAMT in PyMol. This information was used to indicate which pockets showed the greatest allosteric potential and thus deserved greater representation in the initial 30 investigatory variants. From the visual inspection pocket-B demonstrated the most apparent route for allostery due to its proximity to the alpha-helix containing residues 312-320, the sheet-helix containing

residues 248-253, and the beta-barrel containing residues 229-233, which are structurally connected to SA contacting active site residues V311, Y255, and W226 respectively. Pockets C and A shared these potential allostery routes with pocket-B; however, due to these pockets being increasingly more distal to the active site than pocket-B, they were considered a less viable route for allostery than was pocket-B. Pockets G and H both demonstrated potential routes for allostery due to their proximity to residues 146-153 which coordinate directly with both SA and SAM, as well as their ability to potentially disrupt coordination of SAM with residues 129-131 and 59-61. Pocket-F shared these potential allosteric routes; however, due it being more distal from the active site it was also considered a less viable option for allostery than were the more active site proximal pockets. Pockets D and E both demonstrated potential allostery in the Rosetta analyses; however, in the PyMol inspection their highly distal positioning relative to the active site showed no obvious routes through which this allostery might be occurring. For this reason, these pockets were considered to show relatively low potential for allostery. Lastly, pockets I and J both demonstrated no signs of allostery in the Rosetta analyses, and the only possible visually observed route was via possible movement of the alpha-helix containing residues 25-29 which interact with SAM, or possible interact with residues 143-145 which could disrupt the position of active site residue Y147. However, due to the highly distal nature of pockets I and J relative to the active site, these potential allostery routes would need to be translated across a relatively large distance and were thus considered highly unlikely to be successful at inducing allostery. For easy reference, these pocket-specific routes of potential allostery can be found in Appendix Table 1. For the aforementioned reasons, there were 8 variants selected from pocket-B, 4 each from pockets C, G and H, 2 each from A, D, and E, and

1 each from pockets J and I. The variants selected from each pocket where those which would cover the greatest number of Rosetta designs for that pocket.

The logic behind emphasizing the proximity of a theophylline pocket to the active site as a major predictor of potential allostery was based on the idea that with greater distance between the pocket and the active site allowed more opportunities for the predicted allosteric route to fail. A metaphor for this would be the concept of falling dominoes where each domino in the path must correctly hit the domino in front of it to facilitate the next domino falling, and where each additional domino and each additional complexity in the pathway represents another possibility for the falling "chain-reaction" to fail. In contrast, a pocket which was highly proximal to the active site such as pocket B would be expected to have a more reliable transmission of allostery from the pocket location to the active site, increasing the likelihood of successful theophylline induced allosteric regulation. For this reason, when ranking theophylline pockets in order of potential for success, the visual inspection in PyMol was considered slightly more relevant than demonstration of allostery in the Rosetta modelling. Figure 19 shows the relative proximity of each theophylline pocket to the active site of SAMT.

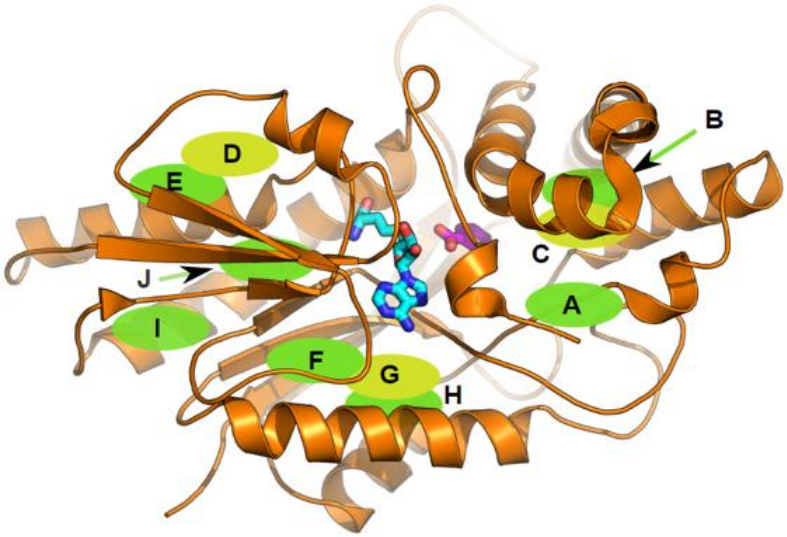


Figure 19: A 2-D representation of SAMT with the locations of pockets A-J indicated. Of note are the relative proximities between the pocket locations and the active site, particularly for pockets which share potential physical routes for allosteric translation from the theophylline binding site and the SAMT active site. Pockets more proximal to the active site are suspected to have a greater allosteric potential while those more distal from the active site are suspected to have lower allosteric potential. This image was produced using PyMol and Inkscape software [78,79]

These 30 selected variants will be commercially synthesized and individually sequence verified to ensure a known sequence for every variant being investigated in this initial investigation. Pocket samples which demonstrate a loss of enzymatic activity will be considered to have the potential for allosteric recovery, and thus will have justification for procurement of a larger library for that specific pocket. Pocket samples which do not demonstrate loss of enzymatic activity will be considered less viable for transmission of induced allostery – although not necessarily incapable of inducing allostery – and library procurement for that pocket will be delayed until necessary.

A metaphor for this initial investigation is the act of prospecting an area for natural minerals before deciding to invest in a larger mining operation. If a prospector's investigation shows there to be minerals in the area, that could be justification to assume there are more minerals of interest in the area, and that investment in further mining is reasonable. However, if a prospector's investigation did not show clear enough evidence of minerals to justify mining, that would not necessarily prove there are no minerals in the area, but further prospecting would need to be done to justify investing in any large-scale mining. In our "prospecting" for potential allostery, a lack of enzyme inactivation in variants sampled from a specific pocket does not necessarily mean that pocket does not have the potential for transmitting induced allostery, but simply that the variants selected from that pocket for the initial investigation did not explicitly show allosteric potential, and that further prospecting is needed to justify working with that specific pocket.

These initial 30 variants will be transformed into *E. coli* and cultured in 10mL of culture media at 28°C for 24-36 hours prior to screening via human olfaction. Any variants which demonstrate olfactorily detectable levels of MSA will be deemed a negative hit due to their lack of effective catalytic crippling in the absence of theophylline, while those which do not demonstrate olfactorily detectable MSA will have their culture media sampled and analyzed via HPLC. Minor amounts of activity in the absence of theophylline might be acceptable so long as the levels are below even those observed under the theophylline riboswitch expression system. Pocket survey variants which do show a lack of activity in the absence of theophylline will be considered positive hits and indicators of allosteric potential for that pocket, and a larger site directed combinatorial library will be ordered for that pocket. These libraries will be commercially produced by TWIST Biosciences using their proprietary SOLD library construction methods.

Conclusions

Currently no explicit conclusions can be drawn about the success of the computational design process as any explicit conclusions are dependent upon the results of laboratory analyses into the library variants. Once the initial investigation into pocket allostery via screening of the initial 30 investigatory variants, and evidence for potential allostery in a pocket is demonstrated by a clear crippling of the enzyme upon mutation, conclusions can be made about the potential each pocket has for transmitting allostery to the active site. This initial investigation may not provide clear evidence that the computational design process created allosteric pockets which can facilitate theophylline dependent allosteric recovery, but it will at least allow conclusions to be made in regard to a given pocket having the ability to effect active site function via hypothetical disturbance of the active site. Once a larger library has been obtained and screened

for theophylline dependent enzymatic recovery further conclusions can be made regarding the success of the overall computational design process.

Conclusions can be made in regard to the effectiveness of the various custom design scripts despite the lack of physical evidence for induced allosteric regulation. Both the Rosetta and SHARPEN designs were able to computationally design large numbers of potential libraries for each theophylline pocket, and the suggested mutations seemed logical when the mutant structures were visually investigated in PyMol. The Pareto optimization script was successfully able to reduce the number of potential libraries and provide a set of Pareto optimized libraries which could be selected from with ease. The scripts for allosteric modelling were successfully able to produce a set of RMSDs and the related comparative histograms for those RMSDs, and while conclusions cannot be drawn regarding how successfully the comparisons were able to actually model allostery, the statistical indications that some libraries had potential for allostery while others did not suggests that the allosteric modelling scripts were able to draw out statistically real information regarding differences between modelled SAMT active site structure when theophylline was present or absent. The success observed in these scripts' functionalities may not provide definitive evidence that the computational design process will be successful when tested in-vivo, but it does give confidence that the design process was based on valid computational design scripts.

Future Work

Initial Investigation into Pocket Allostery

The initial 30 variants will be transformed into *dam(-) E. coli* and cultured in 10mL of culture media containing SA at 28°C for 24-36 hours prior to screening via human olfaction. Any variants which demonstrate olfactorily detectable levels of MSA will be deemed a

negative hit due to their lack of effective catalytic crippling in the absence of theophylline, while those which do not demonstrate olfactorily detectable MSA will have their culture media sampled and analyzed via HPLC. Minor amounts of activity in the absence of theophylline might be acceptable so long as the levels are below even those observed under the theophylline riboswitch expression system. Pocket survey variants which do show a lack of activity in the absence of theophylline will be considered positive hits and indicators of allosteric potential for that pocket, and a larger site directed combinatorial library will be ordered for that pocket. These libraries will be commercially produced by TWIST Biosciences using their proprietary Spread-Out Low Diversity (S.O.L.D.) library construction methods.

When it comes to commercially produced enzyme variant libraries, most companies currently on market offer either a standard random mutagenesis library or more complex libraries which have the ability to create site specific mutations with specific residue mutations and specific residue frequencies. Because the variant library has been computationally designed with site-specific suggested mutations, the best method for producing the desired libraries would be these more complex library options, as this would prevent undesirable random mutations which would not be expected to facilitate theophylline binding and drastically decrease the amount of screening required as compared to a random mutagenesis library. We plan to utilize the S.O.L.D. library preparation method offered by TWIST Biosciences due to the fact the S.O.L.D library allows for the highly specific site-specific mutations at a specific ratio that are desired for the library and was also determined to be the most price competitive company offering this service. The S.O.L.D method will be used to produce a Pareto optimized library of a size which can be readily screened with the current screening methods at the point of purchase, with larger libraries possible once the high throughput screening method has been fully developed.

Library Expression and Screening

Once the initial investigations into pocket allostery are completed and have informed as to which pockets demonstrate the greatest potential for successful transmission of allostery, complete pocket libraries can be procured and screened for loss of SAMT activity and subsequent recovery of activity upon exposure to theophylline. The library will be produced by TWIST and delivered in the form of linear dsDNA which will be cloned into the pSC-b-amp/kan backbone and then transformed into *dam(-) E. coli* and *E. coli* Nissle (and potentially other *E. coli* strains if they are deemed beneficial), then screened for MSA production in the absence of theophylline. As mentioned earlier, for the allosteric activation approach to have an effect, the mutant SAMT must not produce MSA in the absence of theophylline. As such, only library variants which do not demonstrate MSA production in the absence of theophylline will be considered successful in the first round of library screening. The library variants which do demonstrate a lack of enzymatic activity in the absence of theophylline will be graduated to the next round of screening for theophylline conditional enzyme recovery. These variants will have theophylline provided at various concentrations and production of MSA or lack thereof will be investigated via both human olfactory screening and HPLC analysis of the culture media.

The HPLC method of tracking SAMT activity is the current validated method for screening for enzyme activity, and it can be reliably used to investigate either the activity of individual mutant variants or the larger enzyme variant library members. However, as previously discussed, one of the detriments of HPLC analysis is the relatively long time required to produce and analyze the necessary samples for even a single library variant, let alone analysis of an entire library. One way to reduce the amount of screening required for a given library would be to do initial culture screens for olfactorily detectable MSA due to the fact that if a mutant library

variant was still able to produce MSA in the absence of theophylline subsequent allosteric activation with theophylline would not be feasible. This will be a highly useful initial qualitative screen for enzyme crippling upon mutation as it allows library members which have not been crippled to be discarded from the subsequent HPLC screening process. Current plans for both the initial investigation into pocket allostery as well as the investigation into the subsequent complete pocket libraries plan to utilize this dual screening method to facilitate a more efficient screening for SAMT inactivation due to mutation. This will involve culturing *E. coli* containing the library variant to be screened in a 10mL liquid culture within a 15mL falcon tube, providing these cultures with sufficient SA, and investigating these cultures for olfactorily detectable MSA after approximately 24 hours incubation.

Once this initial screen for enzyme deactivation due to mutation is complete, those library variants which displayed a loss of enzymatic activity will then be exposed to theophylline and analyzed for recovery of enzyme activity. A dual-step screening process will occur in which olfactory qualitative screens are used to initially investigate for recovery upon exposure to theophylline, and subsequent HPLC analysis used on library variants which olfactorily demonstrate significant MSA production. There is a chance a library member will have its catalytic ability restored by theophylline but so poorly that MSA production levels are too low to detect olfactorily. Because the overall purpose of the project is to demonstrate catalytic recovery significant enough to produce olfactorily detectable concentrations of the signal odorant, library variants which were technically recovered by theophylline but unable to produce qualitative olfactory detection would still be considered “non-recovered” and would not be worth further HPLC analysis.

Directed Evolution

Once a full library has been purchased and screened for successful theophylline induced allosteric recovery of SAMT activity, directed evolution techniques can be applied to those successful library variants to produce SAMT expression strains with even greater enzymatic activity or other desirable traits. This will be achieved identification and isolation of the successful library variants, subsequent chemical induced random mutagenesis, and a final screening of the mutagens for either increased SAMT activity or whatever other desirable train we would like to select for. While an increase in overall SAMT theophylline dependent catalytic restoration is expected to be the primary trait for which directed evolution would be focused on, alternative traits which could be focused on include increased SA-tolerance, increased irp9 activity, or possible increased heat tolerance for potential use in sentinel animals with natural body temperatures outside the temperature range of standard lab *E. coli.* For example, canines in general have a higher body temperature than do humans, with healthy canine temperatures ranging from 38.3 to 39.2°C [75]. As such, creation of SAMT expression strains of *E. coli* which were evolved to function best within this temperature range could be a desirable trait to evolve using directed evolution.

CHAPTER 4: ANIMAL STUDIES

Introduction

Because the ultimate goal of the odor encoder project is application within the nasal microbiome of trained odor detecting sentinel animals, testing the system in laboratory animals is a necessity for final proof of concept. To achieve this, a training protocol was developed for teaching handling-acclimatized Sprague Dawley rats to signal olfactory detection of MSA. In addition, an investigation into the nasal microbiome of Sprague Dawley rats was performed to better understand the microbiome environment in which we intent to inoculate our engineered organism when eventual in-vivo animal testing occurs. The nasal microbiome information will be used to help establish some basic information about the relative abundance of different bacterial phylum within the natural nasal microbiome of Sprague Dawley rats, as well as establish a proven protocol for the long-term tracking of nasal microbiome colonization and clearance by upon inoculation with the engineered organisms.

Investigation into the Nasal Microbiome

Materials and Methods

To obtain intranasal samples, 50 µL sterile 0.9% NaCl was pipetted into each nostril of 4 isoflurane-anesthetized handling-acclimatized Sprague Dawley rats and samples were stored at -20°C. Collected samples were aliquoted into halfs, and the DNA was extracted using one of two separate kits manufactured by Macherey-Nagel, the bead-based NucleoMag DNA Microbiome kit (Item # 744330.4) or the filter-based NucleoSpin Soil Microbiome kit (Item # 740780.50). The DNA extracted from these two kits were both subjected to 16-s sequencing to compare relative DNA extraction yields. Subsequent sample DNA extractions were performed using the filter based NucleoSpin Soil Microbiome kit, and these new DNA samples were subjected to

Illumina Next Generation Sequencing (NGS). Illumina NGS was performed by the staff at the Colorado State University Sequencing Core, with quality filtering, read alignment, and taxonomy assignments performed using the online QIIME2 software and the Silva RNA sequence database. Organisms were classified at either the phylum or genus level depending on the analysis.

After baseline microbiome analysis was performed, a single rat which was scheduled for humane euthanasia due to development of a tumor was nasally inoculated with 100uL of an overnight culture of *S. carnosus* cultured in Miller-LB. 24 hours post-inoculation, the rat was sacrificed and the mucosal layer was sampled with a swab and lavage in 50uL of sterile 0.9% NaCl, followed by shallow and deep tissue collection. DNA extraction, Illumina NGS, and taxonomic abundance analysis were performed as previously described. This rat was not one of the 4 rats from which the baseline investigation samples were obtained but was from the same cohort of handling-acclimatized Sprague Dawley rats.

Results and Discussion

It should be noted that this investigation into the rat nasal microbiome was performed with a focus upon identification and relative quantification of the phylum *Firmicutes* of which the *Staphylococcus* genus exists because the original plan was to utilize *S. carnosus* as the expression organism which would eventually be inoculated into the rat nasal microbiome. Since the project has transitioned to *E. coli* Nissle as the expression organism, it would be ideal to redo this investigation with a specific focus upon quantifying relative abundance of the phylum *Proteobacteria* of which *E. coli* is a member. None the less, the information gained regarding the background nasal microbiome is still relevant regardless of which expression organism is ultimately in use, as are any insights gained about the increases in relative abundance of an inoculum species post-inoculation.

Protocol optimization for DNA extraction was necessary due to the relatively low biomass environment of the rat nasal microbiome. This was why the initial samples had DNA extracted via two separate kits, and the higher DNA yields obtained from the filter-based Soil Microbiome kit were the justification for exclusively using that extraction kit going forward.

The baseline relative abundance investigation demonstrated that the *Firmicutes* phylum was the most abundant phylum within the nasal microbiome in 75% of the rats sampled, and represented greater than 35% of the nasal microbiome members for all rats sampled. *Proteobacteria* relative abundance was less than *Firmicutes* in all rats sampled, but was still observed in significant abundance in all sampled rats. Figure 20 shows the relative abundance of 9 common bacterial phylum between the 4 sampled rats.

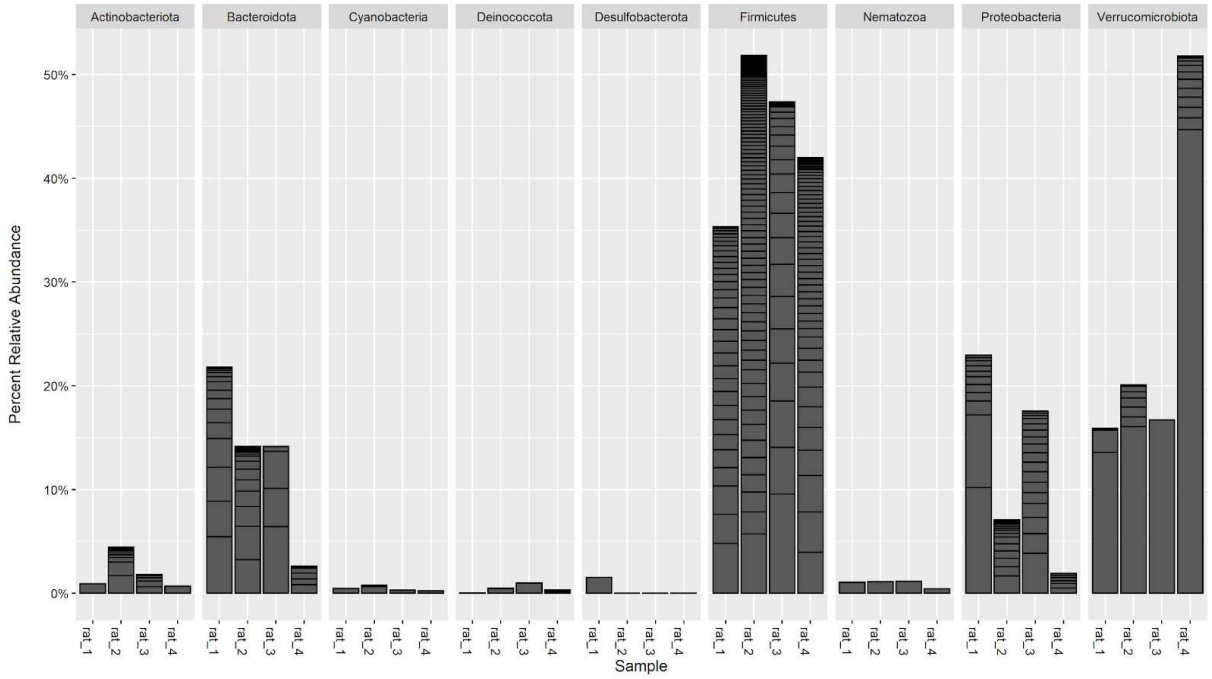


Figure 20: Baseline relative abundance of different nasal microbiome phylum observed in 4 different Sprague Dawley rats. The phylum *Firmicutes* of which *S. carnosus* is a member was the most abundant phylum in 75% of rats, and was at a relative abundance greater than 35% in all rats sampled. The phylum *Proteobacteria* of which *E. coli* is a member was observed in significant relative abundance in all rats, but represented no more than 25% relative abundance in any rats sampled.

Quantification of the relative abundance for *S. carnosus* was analyzed at the genus level, with figure 21 showing the relative abundance of the *Staphylococcus* genus within the nasal microbiome samples. *Staphylococcus* species represented no more than 4% relative abundance within the nasal microbiomes of the 4 rats sampled. This demonstrates that *Staphylococcus* species represent only a small percentage of the nasal microbiome despite the *Firmicutes* phylum being a major representative fraction of the total microbiome.

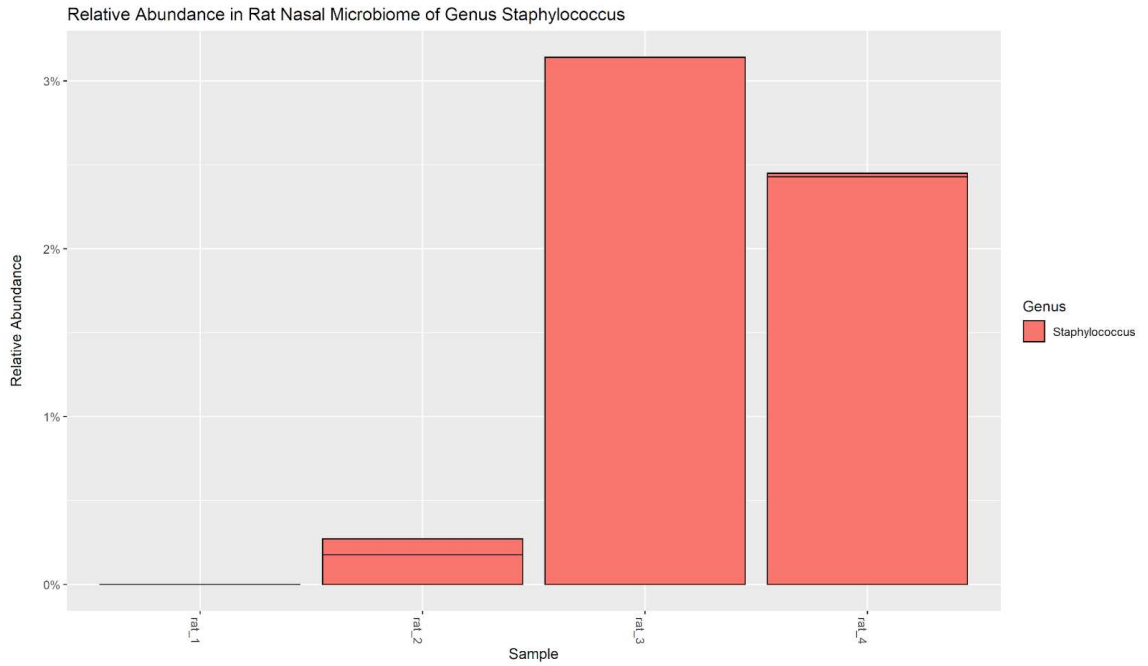


Figure 21: Baseline relative abundance of the genus *Staphylococcus* observed in 4 different Sprague Dawley rats. Relative abundance of *Staphylococcus* species represented less than 4% of the total microbiome in all 4 rats sampled, with one rat having essentially no observable *Staphylococcus* species within their nasal microbiome.

The investigation into the relative abundance of nasal microbiome organisms of a single Sprague Dawley rat 24 hours post nasal inoculation with *S. carnosus* demonstrated *Firmicutes* being the most abundance phylum in the swabbed mucosal sample as well as in both the deep and shallow tissue samples, representing more than 30% relative abundance in all three sample varieties. *Firmicutes* organisms were most abundant in the swabbed mucosal sample and lowest in the shallow tissue sample. The *Proteobacteria* phylum was also showed a relative abundance of greater than 15% in all tissue sample types but was lowest abundance in the deep tissue

sample. There was no obvious trend showing a clear relationship between levels of relative abundance for a phylum and the type of tissue sample, with the location of dominant abundance for various phylum varying between the mucosal swap, the shallow tissue, and the deep tissue. Figure 22 shows the relative abundance of different phylum within the different tissue sample varieties as observed in the single Sprague Dawley rat 24 hours post-inoculation with *S. carnosus*.

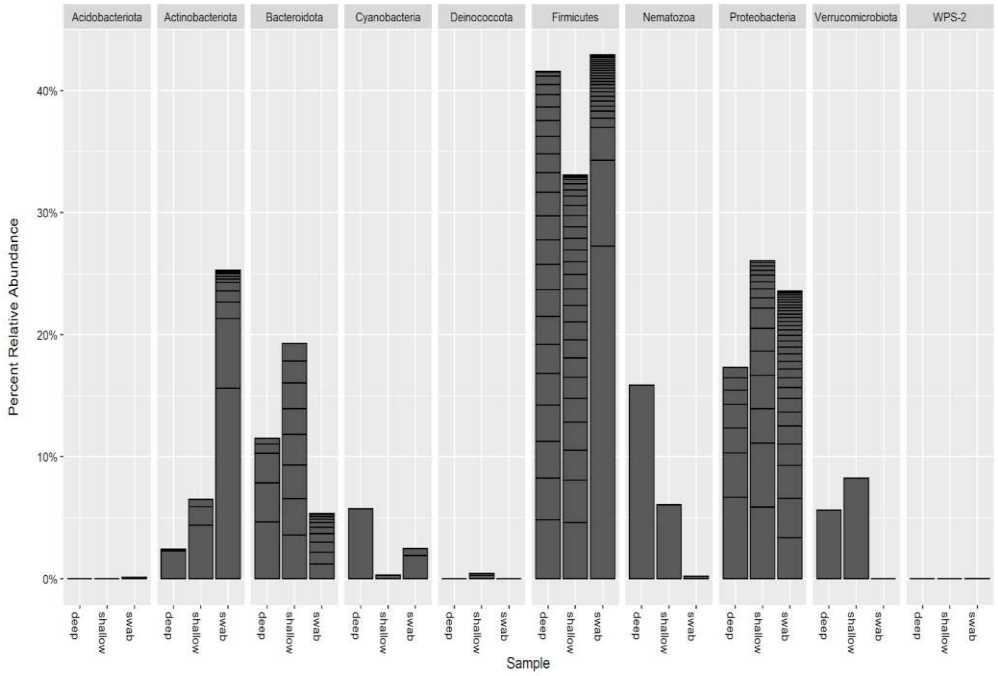


Figure 22: Relative abundance of different nasal microbiome phylum observed in a single Sprague Dawley rat 24 hours post nasal inoculation with *S. carnosus.* The phylum *Firmicutes* was the most abundant phylum observed in the mucosal tissue swabs, the shallow tissue, and the deep tissue, representing greater than 30% relative abundance in all three tissue sample varieties.

In the 24-hour post inoculation tissue samples, the relative abundance of *Staphylococcus* species in the mucosal swabs, shallow tissue, and deep tissue samples were approximately 37%, 5%, and 3% respectively (Fig. 23). The relative abundance of the *Staphylococcus* genus in these samples were significantly higher than those observed in the baseline investigations, although the access to both swabbed and tissue samples for the post-inoculation samples compared to just the nasal lavage samples for the baseline investigation likely impacted the overall sample collection

efficiency between the two assays. Furthermore, the rat which was inoculated with *S. carnosus* was not one of the 4 rats used in the baseline investigation, and there is clear evidence that the relative abundance of nasal microbiome phylum varies between individual rats. Therefore, we cannot draw explicit conclusions about increases in relative abundance between the baseline samples and the post-inoculation samples, but we can take the dramatic increase in the relative abundance of *Staphylococcus* species observed between the single *S. carnosus* inoculated rat and the 4 rats with non-inoculated microbiomes as a positive indication of potentially successful nasal microbiome colonization by the inoculated *S. carnosus*.

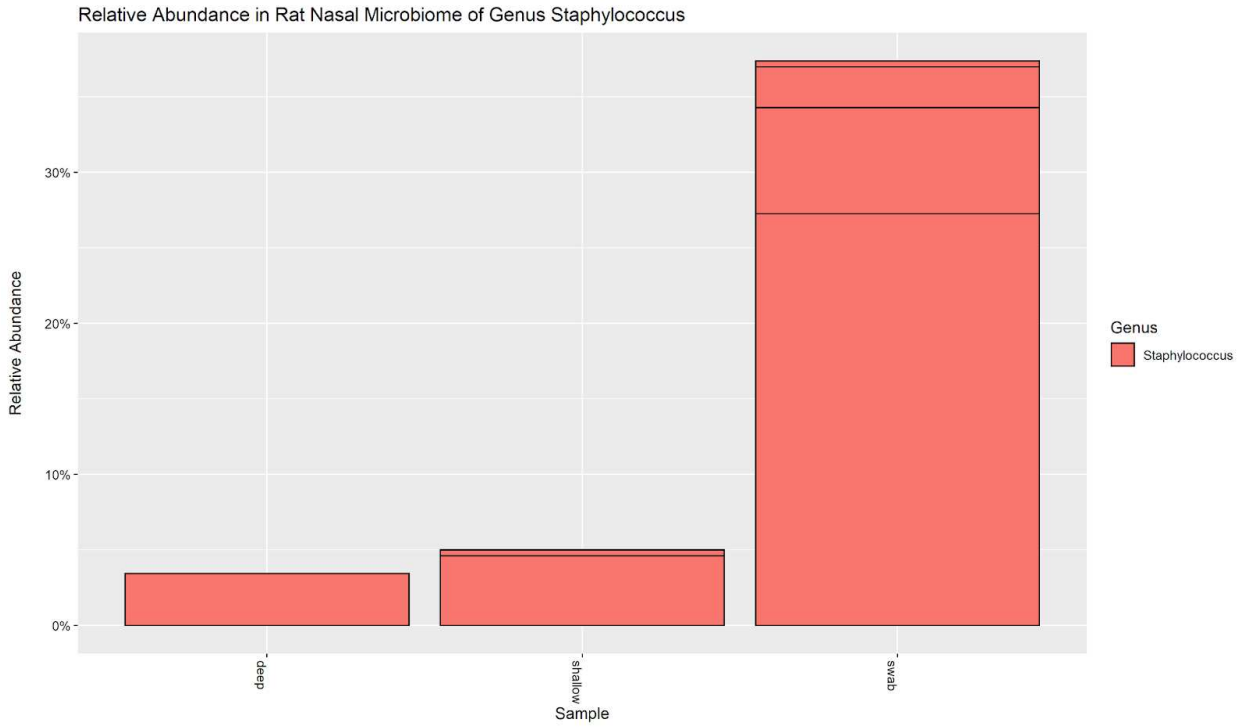


Figure 23: Relative abundance of the genus *Staphylococcus* observed in a Sprague Dawley rat 24 hours after nasal inoculation with *S. carnosus*. Relative abundance of *Staphylococcus* species was 37%, 5% and 3% in the mucosal swab, shallow tissue, and deep tissue samples respectively.

Conclusions

The results of the nasal microbiome relative abundance assay demonstrate that despite the phylum *Firmicutes* commonly beings a dominant member of the natural nasal microbiome of

Sprague Dawley rats, the genus *Staphylococcus* only represents a small percentage of the *Firmicutes* species present in the natural nasal microbiome. The relative abundance of the *Staphylococcus* genus was much higher in the rat which had been subjected to nasal inoculation with *S. carnosus* which suggests inoculated *S. carnosus* are able to effectively colonize the nasal microbiome of Sprague Dawley rats, although this will not be confirmed until comparisons between baseline and post-inoculation relative abundance of *Staphylococcus* species are determined for a single rat. This assay also demonstrated the superior ability of the filter-based Soil Microbiome DNA kit to extract DNA from the relatively low biomass samples obtained from the nasal microbiome lavage samples. This assay also demonstrated that there are differences in the relative abundance of nasal microbiota members observed between the deep tissue, shallow tissue, and mucosal layers of the nasal microbiome, although no clear trends in these differences between phylum was observed. The genus level analysis should be repeated with a focus on the *Escherichia* genus due to the transition to *E. coli* Nissle as the intended expression organism.

Animal Training for Signaling of Olfactory MSA Detection

As the ultimate goal of the Odor Encoder project is to provide enhanced olfactory abilities to trained odor detecting sentinel animals, it is imperative that the engineered biological system is tested and proven effective in animals as a final confirmation of project success and concept validation. A set of Sprague Dawley rats was select for this purpose and have been undergoing training to specifically signal detection of MSA over the course of the project. This effort has been led by undergraduate student Michael Burns utilizing a training protocol developed in-house to train the rats to both detect MSA and successfully signal that detection. This process has involved regular trainings in which rats are placed into an enclosure which

contains two pressure activated levers, exposed to a puff of air either with or without MSA, and allowed to press one of the levers, with one lever being specific to air samples with MSA and another lever being specific to air samples without MSA.

Initially, there is an opaque barrier between the section of the training cage in which the rats are initially placed and the section of the training cage which contains the pressure activated levers. Rats are allowed to sit in this initial preparation chamber until they have adjusted to being inside of the testing cage, with adjustment determined by the rats no longer being actively investigatory towards the testing cage. During this time within the preparation chamber of the larger training cage, there exists an diffuser within the preparation chamber which is constantly emitting an diffused solution of theophylline such that the preparation chamber is believed to be thoroughly saturated with theophylline before the removal of the opaque barrier. This ensures all rats are exposed to theophylline during every training session regardless of the type of air sample they are eventually exposed to. This is done to ensure the rats are being trained to specifically signal detection of MSA and not detection of the theophylline which they will eventually be exposed to during downstream testing for theophylline conditional enzyme expression or activity.

Once rats have been allowed to adjust to the preparation chamber, a sample of air containing either MSA or non-odorant-laced air is administered into the preparation chamber via puffing the air sample through a small hole in the side of the chamber. This is done so that the rats are not aware of air sample administration but are simply detecting what to the rats would be a spontaneous presence of MSA. This should be similar experience to that which would occur upon in-situ production of MSA in the nasal cavity as in this scenario the rats would not detect a specific source of MSA but have in-situ MSA production such that their entire nasal cavity

would be exposed to MSA and anything else they smelled afterwards would necessarily smell like MSA. For this reason, the rats should not be trained to indicate “this object/space smells like MSA” but simply to indicate “yes I do or no I do not smell MSA”. Whether MSA or a no-odorant air sample is provided on a given day is determined via a coin flip to ensure randomness in the training regimen and prevent any pattern in sample administration to influence the rats behaviors.

Once the rats have been exposed to the air sample either with or without MSA, with sufficient time allowed for the air sample to disperse within the preparation chamber and enter the rat’s nasal cavity, the opaque barrier separating the preparation chamber from the lever-containing response chamber is removed and rats are allowed to interact with the levers. As such, it is act of removing this opaque barrier which poses the question to the rats “do you smell MSA or not?”, and it is after the removal of the barrier that the rats are allowed to interact with the levers and indicate detection of MSA or not. Figure 24 shows a rough depiction of the overall training cage. The general training protocol can be found in more detail in the appendix.

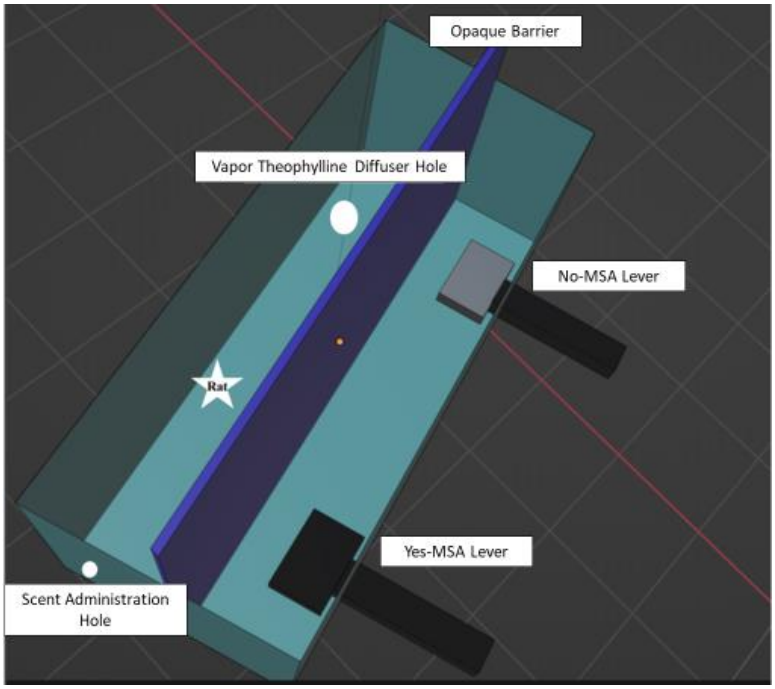


Figure 24: Training cage for the MSA-detection training process. The rear chamber is the preparation chamber in which the rats are initially placed and exposed to both theophylline and the air sample variant for that test. As such, the preparation chamber contains the theophylline diffuser and a hole for covert administration of the air sample. The front chamber is the response chamber and contains the levers which are to be pressed upon removal of the opaque barrier. The grey lever platform is associated with a no-MSA air sample, while the black lever platform is associated with an MSA spiked air sample. The testing cage also has a lid (not depicted) to ensure the theophylline and air samples do not simply diffuse out of the cage and into the surrounding environment. Graphic creation credit goes to Michael Burns.

The rats were initially directed towards the correct lever for a given air sample via presentation with a small food treat on the correct lever after exposure to the air sample in order to develop an association between air samples with or without MSA and the respective correct lever for that air sample type. Eventually this directing was ceased, and rats were allowed to freely interact with either lever upon exposure to an air sample. Rats which interacted only with the correct lever for that air sample type, and did so within a specific set time limit, were rewarded with a food treat and that trial was counted as a "success". In contrast, those rats which interacted with the incorrect lever or did not interact with any lever within the time limit were not given a treat and were counted as a failed trial. The success/fail rate for each rat was tracked over time, with the time limit for lever interaction also decreased as training progressed, and rats which showed clear improvements in their success/fail rate were considered to be successfully demonstrating detection of MSA and were maintained for further training and downstream animal testing. Those rats which did not show clear improvements in their success/fail rate were considered to have not shown potential as MSA detecting animals and were thus removed from future animal studies.

Future Work: Concept Validation via Testing in Trained MSA Detecting Rats

In the future, rats which have passed through this training program and have successfully demonstrated their ability to detect MSA and signal that detection will be inoculated with *E. coli* Nissle with expression plasmids containing either an allosterically activated cSAMT/GFP-irp9 or

iSAMT/GFP-irp9 gene circuits and tested for successful in-situ production and subsequent detection of MSA upon exposure to theophylline. These animal investigations should show definitively that our genetically engineered strains of *E. coli* Nissle are able to successfully produce MSA within the nasal microbiome of a trained odor detecting sentinel animal through conditional expression via the theophylline dependent riboswitch or through conditional activation via theophylline induced allosteric enzymatic restoration. For these future tests, *E. coli* Nissle with non-theophylline dependent constitutively expression of SAMT would be utilized as a genetically equivalent positive control of for in-situ MSA production, while *E. coli* Nissle expressing the SAMT knockout mutant would act as a genetically equivalent negative control. This final animal testing will not occur for a while due to the fact the rats are still undergoing training now, in addition to the fact that the *E. coli* strains and related expression plasmids which will ultimately be utilized for animal inoculation are still being developed. However, the training is already underway, with significant progress already made in regard to development of and optimization of training techniques and production of trained MSA detecting rats.

CHAPTER 5: EXPANDED APPLICATIONS

Alternative SAMT Activator Analytes

While the hopes of the project are to design a crippled version of SAMT which has its activity restored through theophylline induced allosteric activation, it is not expected that every designed library member will actually produce these desired results. However, due to the fact all of the allosteric cavities were designed to accommodate theophylline, there is a chance library members which were not successfully recovered by theophylline could being recovered upon binding of molecules with structural similarities to theophylline. For example, caffeine only differs from theophylline by one methyl group (Fig. 25), is also odorless, and has a low vapor pressure of 9.0×10^{-7} mm Hg at 25 °C [33]. For these reasons, caffeine would be an ideal molecule with which to screen for enzymatic recovery of the library members which were unable to be recovered by theophylline. This could be relevant considering caffeine is currently the most consumed psychoactive compound in the world, and there is evidence in the U.S of caffeine and caffeine metabolites – including theophylline and other structurally similar methylxanthines such as paraxanthine and 1,3,7-trimethyluric acid – being urinarily excreted in significant concentrations [34]. While caffeine usage by humans within recommended dosages is considered safe, there are a number of metabolic functions which are affected by caffeine, and which can be seriously impacted by heavy caffeine intake. These effects include psychological impacts of increased anxiety, noticeable physical conditions such as tachycardia, and metabolic impacts such as increasing the activity of CYP1A2 enzymes in some humans. In fact, toxic effects from caffeine in humans often occur at levels above 15mg/L, although discrepancies do exist between individuals [35]. As such, it could be important to investigate if caffeine contamination in the

environment could be unintentionally impacting nearby human populations, or even for simple tracking of environmental caffeine contamination due to exposure to human sewage systems.

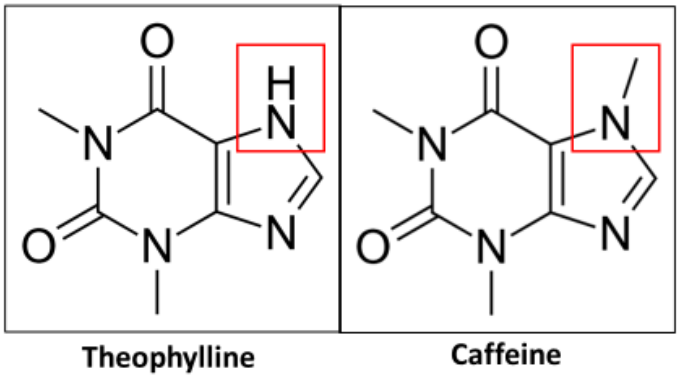


Figure 25: Structures of theophylline and caffeine. Note that caffeine only differs from theophylline by one methyl group, with the location of difference indicated in red.

Furthermore, caffeine can be highly toxic to animals such as canines, with cases of canines being lethally poisoned after consumption of ground meat containing 1% caffeine [36]. Canines have a oral median lethal dose (LD_{50}) for caffeine of only 140mg/kg, while albino rats have been shown to have a caffeine LD_{50} of 367mg/kg [36,37]. This relatively high toxicity in canines has led to instances of canines being intentionally poisoned by being tricked into eating a single 200mg caffeine pill, the strength commonly found in over the counter human-intended caffeine tablets [36]. Due to the fact the high use of caffeine within a population can lead to unintended environmental exposure to caffeine, and due to the fact it can be so toxic to animals including canines, it would be useful to have a ready method of detection for environmental caffeine contaminations. This could be particularly useful in preventing sentinel canines from being intentionally or unintentionally poisoned due to caffeine exposure while performing their duties.

While molecules such as caffeine which share structural similarities to theophylline would be an easy alternative molecule to potentially activate our theophylline-specific designed

allosteric cavity, the idea of designing an analyte specific allosteric regulation system could be expanded to any molecule of interest which has at least some of the aforementioned amino acid-like physical characteristics possessed by theophylline. Some examples of molecules of interest which could potentially match these criteria are fentanyl and rohypnol, both of which are dangerous illicit drugs which are commonly found in drug searches. In fact, across the United States overdoses due to fentanyl have been increasing, often due to fentanyl being mixed with stimulants such as cocaine or methamphetamine without the user's knowledge, with the number of fentanyl laced cocaine samples increasing 300% between 2015 and 2016 [38]. This is particularly dangerous for the police drug detecting canines, with overdoses due to fentanyl exposure occurring often enough that police canine handlers have been trained to administer the opioid-overdose reversing drug naloxone to canines upon potential exposure to fentanyl [39]. There is also clinical evidence that police canine handlers who are administering the naloxone are prone to accidental exposure to powered drugs which may be on the canines themselves, with internasal administration of naloxone shown to be more likely to contaminate the handlers than intramuscular administration [40]. For all these reasons, having a system which could increase the ability of police canines to detect fentanyl or other dangerous drugs in the environment could be useful in preventing accidental exposure and/or overdose of fentanyl or other highly toxic compounds for police drug detecting canines.

In addition to providing drug detection canines with the enhanced olfactory ability to detect dangerous compound such as fentanyl through the detection of an allosterically expressed proxy odorant compound, our enhanced olfaction system also comes with the benefit of allowing increased sensitivity to low concentrations of these compounds due to the natural signal amplification which occurs upon enzyme activation. Theoretically, it would only require a single

analyte molecule to be present and bind the allosteric site to induce continued enzyme activity. As a result, a single activator analyte molecule could cause the activated enzyme to produce large concentrations of the detectable odorant molecule, and a sample with a relatively low level of analyte could still be olfactorily detected. This is particularly relevant for highly toxic compounds such as fentanyl, as it would allow drug detection canines to be exposed to non-toxic levels of these compounds while still producing enough odorant signal molecules for the canines to detect.

Expansion into Non-animal Sentinel Species

While the current project applications are geared towards the enhancement of animal olfactory detection systems, the concept of utilizing sentinel species is not limited to animals, but can also include species such as plants, lichen, and theoretically even bacteria. Plant and lichen have both been utilized as sentinels for environmental contamination, monitoring for atmospheric pollutants, water contamination, or biological pathogens [41,42]. These sentinel species tend to be naturally occurring species which display characteristic responses to specific environmental changes, and as such they are limited to those species which naturally undergo phenotypical changes under these specific conditions [41]. However, through the application of the same engineered allostery concepts being utilized to create the enhanced olfactory system, a similar analyte-specific enzyme activation system could be applied to plant or lichen sentinel species to produce an observable signal in the presence of the chosen analyte. For example, if there was concern for a hazardous industrial chemical leaching into the environment, a plant or lichen species could be engineered so that the chemical of concern allosterically activates a fluorescent reporter protein, a pigment producing enzyme, or even an odorant producing enzyme such as SAMT. As a result, when the chemical of concern was present in the environment, the

sentinel species would either become fluorescent, experience a color change, or begin to produce a noticeable odor respectively. This is just one possible example of how the concepts developed in the enhanced olfaction project can be expanded to plant and lichen sentinel species.

Expansion of this technology to create bacterial sentinel species would be even more straightforward due to the fact this project has already developed these expression systems in *E. coli.* The current system has been designed to produce MSA in the presence of theophylline, which means a simple culture of these engineered *E. coli* could be exposed to a sample suspected to have theophylline in it and produce a detectable amount of MSA if theophylline was indeed present in the sample. By simply altering the activation analyte from theophylline to another chemical of concern, this simple bacterial culture method could be an easy way to screen environmental samples for the chemical of concern and have an easily olfactorily screening method. What's more, this method would not require further engineering of the genetic components necessary for successful genetic expression in plant or lichen species but would simply be able to incorporate alternative allosterically regulated engineered enzymes into the current *E. coli* verified plasmid expression vectors.

Non-Sentinel Based Applications

While the current discussion has revolved around using engineered allostery to provide enhanced abilities to sentinel species, this same technology can be broadly applied to produce a variety of novel allosteric enzyme regulatory systems. Any enzyme could theoretically have this concept applied to it with any activator analyte which fits the necessary biochemical stipulations. One theoretical example of this would be creating an allosterically regulated enzyme for use in enzyme replacement/supplementation therapy. This enzyme could be engineered for activation by an artificially supplied drug, or by a relevant endogenous metabolite, resulting in a system

which could produce highly regulated and rapid activation of the therapeutic enzyme either on demand or under specific biochemical conditions. Another theoretical example could be engineering of a biofilm which could have specific metabolic enzymes activated in the presence of specific environmental pollutants. Biofilms are commonly used in water and sewage treatment systems for their ability to metabolize pollutants out of aqueous solutions, with the variety of species which makeup a biofilm providing that biofilm's filtration abilities [43,44]. By engineering pollutant specific allosteric activation systems into the organisms which constitute a biofilm, a customizable biofilm could theoretically be created which would be able to have any desired metabolic activity activated by a pollutant molecule of interest. This could even be combined with the sentinel species technology to create novel biofilms which are both uniquely detoxifying and able to signal the presence of toxins in the fluids they filter. Ultimately, any enzyme for which highly specific activation conditions would be desirable could theoretically be processed through the same design pipeline as was used for production of the enhanced olfaction system, and the result would be an enzyme with novel, unique, and situation-specific allosteric regulatory abilities.

In addition to the ability to regulate enzymatic catalysis, allosteric regulation can regulate conformational behaviors of non-enzymatic proteins as well. As such, the technology of engineering allostery could be applied to control of proteins for which conformational changes are important to ligand binding. One theoretical example of this could be a conditional scaffold protein which has its aggregation and polymerization events regulated via binding of a chosen allosteric activator. Considering protein polymerization is dependent upon conformational complementation by the various monomers, having monomers which were unable to polymerize in the absence of a specific allosteric activator molecule could be a useful way of regulating the

formation and breakdown of protein scaffolds or other protein polymers. Given the wide-ranging applications of proteins in industry, medicine, and general scientific research, having the ability to create allosteric regulatory systems which are specific to chosen analyte molecules could have wide ranging impacts in the overall field of protein engineering and metabolic regulation.

Another interesting application of this technology is the simple utilization of MSA producing *E. coli* for the benefit of biological researchers. As anyone who has worked with *E. coli* can attest, the metabolism of most lab strains of *E. coli* results in the production of highly indolic metabolites which are usually considered unpleasant to smell. Having a strain of E. coli which can naturally produce and secrete a desirable odorant molecule such as MSA could be an attractive way to reduce the odors associated with *E. coli* metabolism, thereby creating a more pleasant work environment for the numerous biological researchers who utilize *E. coli* as their model organism of choice. While this may not be the most impactful to the wider community, and pleasant-smelling *E. coli* may not be of necessity in the same way as an enhanced olfactory detection system is, production and marketing of MSA producing lab strains of *E. coli* would be relatively simple and could be an additional financially lucrative avenue for this technology.

Nasal Microbiome Expression System

Outside of the enzyme specific innovative components of this project, the work of engineering the nasal microbiome in such a manner is so-far unprecedented. The microbial makeup of the nasal microbiome has been shown to play a role in general human health, with healthy microbiota preventing colonization by potentially pathogenic bacterial species. In addition, competition between non-pathogenic nasal microbiota and the pathogenic species *Staphylococcus aureas* has been utilized to treat *S. aureus* infections by allowing microbiota species which can secrete compounds which kill *S. aureus* and prevent further infection. For

example, some *Corynebacteria* species such as *C. pseudodiphteriticum* were shown to inhibit growth of *S. aureus*, and downregulate genes involved in *S. aureus* colonization and virulence. In contrast, the species *C. accolens* was shown to increase *S. aureus* viability. In addition, many of the competitive regimes seen between *S. aureus* and other nasal microbiome species were shown to be dependent on specific metabolic conditions such as iron or methionine availability, with some *Staph* species showing competitive inhibition when iron levels were limited which could be reversed via supplementation with iron [71]. By engineering nasal microbiome members to engage in specific metabolic processes, an engineered microbiome organism could directly affect the chemical nature of the nasal microbiome to prevent colonization by pathogenic organisms. This would in essence be a nasal probiotic which could be used to create and maintain a healthy nasal microbiome.

Another example of how engineering the nasal microbiome to control overall microbiota variation could be beneficial can be found in the links between nasal microbiota and asthma. There is clinical evidence of statistically significant differences in the ratios of common nasal microbiota between non-asthmatic, non-exacerbated asthmatic, and exacerbated asthmatic individuals. Not only were the overall levels of nasal microbiota members higher in asthmatics compared to non-asthmatic individuals, but there were significant differences in the ratios of the microbiota members investigated between individuals with exacerbated and non-exacerbated asthma [72]. An expansion of the previously mentioned method of controlling the nasal microbiome makeup via engineering of competitively advantageous non-pathogenic nasal microbiota members would be to explicitly control the nasal microbiota of asthmatics to reduce their risk of experiencing exacerbated asthma conditions. While this may not be able to

completely reduce the risk posed by asthma, it could be a useful way to apply the nasal probiotic concept to provide increased protection from severe asthma episodes.

A final interesting expansion of nasal microbiome engineering could be engineering nasal microbiota members which secrete medically useful compounds. For example, there is clinical evidence that nasally administered insulin can improve the cognitive functions of individuals with Alzheimer's. This is believed to be partially due to the nature of insulin in the brain and its ability to readily pass the blood brain barrier when administered intranasally, although the underlying mechanisms behind this phenomenon are not well understood [73]. Still, there is interest in improving the intranasal administration of insulin, and the engineering of an insulin secreting nasal microbiota member could be an effective way of achieving this goal. This could be a system in which insulin was constitutively expressed and secreted at low levels appropriate for a continuous secretion, or one could adapt the allosteric activation method used for the enhanced olfaction project and create an organism that only produced insulin when exposed to a specific activator molecule, that molecule being delivered via simple inhalation or nasal spray. And while this concept is based on the clinical evidence that nasally administered insulin can improve Alzheimer's symptoms, this could possibly even be used as an alternative to the current injection method for providing insulin to type-1 diabetics, although much work would need to be done to validate this concept and fully develop this system.

Covert Communication

A final potential application of this technology which could be particularly attractive to the Department of the Navy or other U.S. military institutes is the potential for a olfactory based covert communication system. Successful covert military operations depend on maintenance of secrecy, and often utilize undercover agents who are not at liberty to openly communicate with

one another. For instances such as this, it would be useful to have method by which these undercover agents could communicate with one another, and which would necessarily only be detectable by allied agents. One possible example would be the application of a “signal cologne” which would be indetectable to normal human olfaction, but which could be detected by allied agents who had been inoculated with the appropriate engineered organism. Another possible example would be a covert “scratch-and-sniff” signal in which an invisible activator molecule could be applied to a specific location on a document or object, and which an agent would know to sniff to covertly receive an odor-specific informative signal. This method of covert communication would be invisible, inaudible, highly specific, and only detectable by allied agents, allowing safe and effective communication between covert agents who must maintain their covert status while still communicating effectively in real time.

CHAPTER 6: OVERALL PROJECT CONLCUSIONS

General Conclusions

This project has demonstrated an effective and reliable workflow for expression of SAMT in *E. coli* and subsequent enzyme activity analyses via HPLC. The validation of HPLC methods for tracking in-vivo enzymatic activity is particularly noteworthy due to both the integral role it plays in the odor encoder project, and as a validation for the HPLC technique as a method for tracking changes in small organic molecules that occur in a liquid sample over time. This allows HPLC to be applied to measurement of many enzymes in a in-vitro method which only relies on tracking depletion of substrate and/or production of product, as well as applying HPLC to numerous possible non-biological processes such as determining degradation of a molecule in solution over time or under various conditions. HPLC is fairly straightforward, and while the initial machinery and software is not cheap to purchase, once it has been obtained the major costs of HPLC use are in the sample preparation methods, and these costs do not seem particularly prohibitive. As such, it is recommended that HPLC be considered for any process which may benefit from tracking the changes in chemical concentration over time.

The project has demonstrated successful development of design scripts which can model the physical structures and interactions between an enzyme and multiple small binding molecules in order to perform combinatorial optimization to produce combinatorically designed mutant enzyme variants which stand up to visual inspection in PyMol. However, the overall validation of the script's successful designing of theophylline binding pockets is unable to be determined until the laboratory screening of designed libraries for theophylline conditional catalytic restoration has been performed. In addition, the script for determining the set of Pareto

Optimized libraries from a larger computationally designed set of libraries was successful in producing easily digestible sets of Pareto optimized libraries when provided with the raw original design data. The script designed to investigate differences between SAMT structure when theophylline either was or was not present was able to produce comparative histograms that captured statistically real trends in RMDS variations between theophylline conditions, although the overall validity of this investigation will require laboratory testing of the libraries from the various pockets. In general, any positive indications of potential theophylline binding induced allostery which are produced from computational algorithms are inherently just modelling events and may not match the physical results observed upon laboratory testing, and as such all of the computational design scripts require physical laboratory data for complete validation.

Constitutive expression of SAMT in *E. coli* was able to result in enough SAMT production that MSA production could be detected both via HPLC quantitative analysis and human olfactory qualitative analysis. Induced expression of SAMT via use of the theophylline dependent riboswitch was able to successfully result in expression of SAMT and production of MSA in significant enough levels to be detected by both HPLC and human olfaction, but with MSA production being significantly less than that which is observed with constitutive expression. However, this is to be expected and is in line with previous reported relative expression levels for theophylline riboswitch expression methods. Non-induction of theophylline riboswitch dependent expression systems did still result is some SA consumption, but at no point was MSA production great enough to be measured via HPLC or definitively detected via human olfaction. This shows that non-induction seems to be successful at preventing significant enzyme expression or activity, although MSA production at levels low beneath the threshold for HPLC

or human olfactory detection might still be within the range of rat olfaction and thus must be monitored for unexpected signs of trace MSA production. If it is desired to increase overall MSA production in the theophylline riboswitch expression system, various previously discussed metabolic engineering approaches could be made to achieve this goal such as directed evolution or upregulation of metabolic pathways involved in SAM biosynthesis.

The full circuit system that includes both SAMT and irp9 genes was able to show successful production of SA and MSA; however, the significantly higher levels of MSA production which occurred in conjunction with glucose spiking indicates this expression system has not yet been perfected and could be greatly improved via further metabolic investigations and engineering. Investigations would need to be performed to determine what role supplemental glucose plays in increasing MSA production and how to engineer strains which are optimized with respect to the elucidated glucose-impacted metabolic pathways. This may be particularly necessary for the eventual in-vivo use in the trained MSA detecting rats as these rats will likely not have an artificial glucose spiking of the nasal microbiome occur post-inoculation and thus will be solely reliant upon the engineered organism's endogenous glucose production to provide enough glucose for high-level MSA production. However, there is also the possibility the lower level of MSA produced in the absence of an artificial glucose spiking would be enough for the trained MSA detecting rat to detect, thus eliminating the need for producing an MSA upregulated strain. Future work in this regard will thus depend on the results from future animal studies and cannot be explicitly determined at the moment of thesis submission.

Progress in the animal training so far seems positive and indicative of certain rats demonstrating increasing abilities to signal the presence or absence of MSA; however, as the initial training process is still in progress and the data so far collected is fairly limited, no real

conclusions about the animal training process can be made until the training has reached a point where the data supports making statistically significant statements in regard to each rat's MSA detection success/fail rate. Despite this, the successful design of the training cage and development of the training protocol were significant steps in allowing the animal training to proceed as smoothly as it has thus far and should be noted as points of success in the overall project

The investigations into the nasal microbiome of Sprague Dawley rats have provided baseline information that has been informative when forming a highly generalized picture of the natural rat nasal microbiome. The first notable success was demonstrating an ability to extract DNA from small volume intranasal lavage samples taken from the low biomass environment of the rat nasal microbiome, as well as determining which DNA extraction kit would work best for such low biomass samples. In addition, the NGS sequencing results provided solid insights into the nature of Sprague Dawley rat nasal microbiomes. While the relative abundance of the different phylum and the *Staph* genus varied between rats, the overall trends of relative abundance were essentially the same between all rats sampled and gives a good general idea of the relative abundance of each phylum and the *Staph* genus likely to occur in all Sprague Dawley lab rats held in this particular laboratory facility. However, the due to the fact longitudinal studies into individual rat nasal microbiomes have not been performed yet, no clear conclusions can be made in regard to how the relative abundance of each phylum or the *Staph* genus might change upon artificial inoculation with *S. carnosus.* It also prevents conclusions from being made into how the relative abundance of each phylum or the *Staph* genus might vary between the mucosal swab, shallow tissue, and deep tissue samples between different rats or for a single rat pre and post inoculation with *S. carnosus*. Lastly, these investigations into the nasal microbiome

were performed with a focus upon *S. carnosus* whereas they should be reperformed focusing on *E. coli* due to the transition into *E. coli* Nissle as the new expression organism.

Innovative Aspects of the Project

One of the innovative aspects of the computationally designed enzyme allostery is that this is the first known example in which this approach has been designed to produce an odiferous enzyme product. As previously mentioned, the work which inspired this by Deckert *et al.* was used to produce an indole induced allosteric activation of β-glycosidase and β-glucuronidase, neither of which produce highly volatile and odiferous compounds [15]. The previous work by Rana *et al.* was able to change the allosteric activator of their enzyme, but once again this enzyme did not produce an odiferous compound, nor was it's activity fully crippled and restored in the manner used by Deckert and in this project [47]. As such, this will be the first time an enzyme with a volatile odiferous product will have been crippled and then allosterically revived to such an extent that the restored enzyme activity can produce clearly detectable levels of odorant.

Another innovative aspect of this project is that it is the first known example of an in-situ expansion of natural olfactory abilities. As previously mentioned, current methods of enhancing olfactory detection include electronic detectors, spiking of the odorless compound with a distinct and easily detectable odorant, or harnessing of the expanded natural olfactory ranges of trained odor detecting animals. While these methods have all proven effective at detecting their odorless compounds of interest, none involve an explicit in-situ production of an odorant molecule within the nasal cavity of an animal. This in essence replicates the principles of enhancing olfaction which are used in the spiking of odorless compounds with odorants, but does not require a

pretreatment of the odorless compounds with a odorant and allows the technology to be applied on demand with respect to both time and location. This on-demand application and ability to detect the odorless compounds wherever they may be present would be a necessity for trained odor detecting animals to perform screening of unknown samples which potentially contain odorless targets of interest.

REFERENCES

[1] R. Chaix, C. Cao, and P. Donnelly, "Is Mate Choice in Humans MHC-Dependent?," PLoS Genet, vol. 4, no. 9, p. e1000184, Sep. 2008, doi: 10.1371/journal.pgen.1000184.
[2] C. A. de March, S. Ryu, G. Sicard, C. Moon, and J. Golebiowski, "Structure–odour relationships reviewed in the postgenomic era," *Flavour and Fragrance Journal*, vol. 30, no. 5, pp. 342–361, 2015, doi: 10.1002/ffj.3249.
[3] D. Gidlow, "Hydrogen cyanide—an update," *Occupational Medicine*, vol. 67, no. 9, pp. 662-663, Dec. 2017, doi: 10.1093/occmed/kqx121.
[4] P. Quignon, M. Rimbault, S. Robin, and F. Galibert, "Genetics of canine olfaction and receptor diversity," *Mamm Genome*, vol. 23, no. 1–2, pp. 132–143, Feb. 2012, doi: 10.1007/s00335-011-9371-1.
[5] "Meet Primus, the K-9 dog who overdosed on drug that killed Prince," *NBC News*. https://www.nbcnews.com/news/us-news/k-9-dogs-overdose-fentanyl-drug-killed-prince-n687611 (accessed Jul. 11, 2022).
[6] J. K. Fink, "Chapter 15 - Odorization," in *Petroleum Engineer's Guide to Oil Field Chemicals and Fluids*, J. K. Fink, Ed. Boston: Gulf Professional Publishing, 2012, pp. 437–458. doi: 10.1016/B978-0-12-383844-5.00015-5.
[7] C. Zubieta, J. R. Ross, P. Koscheski, Y. Yang, E. Pichersky, and J. P. Noel, "Structural Basis for Substrate Recognition in the Salicylic Acid Carboxyl Methyltransferase Family," *The Plant Cell*, vol. 15, no. 8, pp. 1704–1716, Aug. 2003, doi: 10.1105/tpc.014548.
[8] S. Miles, "Methyl Salicylate," in *xPharm: The Comprehensive Pharmacology Reference*, S. J. Enna and D. B. Bylund, Eds. New York: Elsevier, 2007, pp. 1–6. doi: 10.1016/B978-008055232-3.62158-0.
[9] "Perfumers Apprentice - Methyl Salicylate." https://shop.perfumersapprentice.com/p-7325-methyl-salicylate.aspx (accessed Jul. 19, 2022).
[10] PubChem, "Hazardous Substances Data Bank (HSDB) : 1935." https://pubchem.ncbi.nlm.nih.gov/source/hsdb/1935#section=Effluent-Concentrations-(Complete) (accessed Jul. 20, 2022).
[11] P. J. Barnes, "Chapter 50 - Theophylline," in *Asthma and COPD (Second Edition)*, P. J. Barnes, J. M. Drazen, S. I. Rennard, and N. C. Thomson, Eds. Oxford: Academic Press, 2009, pp. 627–638. doi: 10.1016/B978-0-12-374001-4.00050-X.
[12] R. Kamiura, Y. Toya, F. Matsuda, and H. Shimizu, "Theophylline-inducible riboswitch accurately regulates protein expression at low level in Escherichia coli," *Biotechnol Lett*, vol. 41, no. 6, pp. 743–751, Jul. 2019, doi: 10.1007/s10529-019-02672-8.
[13] R. Kivi, K. Solovjova, T. Haljasorg, P. Arukuusk, and J. Järv, "Allosteric Effect of Adenosine Triphosphate on Peptide Recognition by 3′5′-Cyclic Adenosine Monophosphate Dependent Protein Kinase Catalytic Subunits," *Protein J*, vol. 35, no. 6, pp. 459–466, Dec. 2016, doi: 10.1007/s10930-016-9691-9.
[14] E. Guarnera and I. N. Berezovsky, "Allosteric sites: remote control in regulation of protein activity," *Current Opinion in Structural Biology*, vol. 37, pp. 1–8, Apr. 2016, doi: 10.1016/j.sbi.2015.10.004.a

[15] K. Deckert, S. J. Budiardjo, L. C. Brunner, S. Lovell, and J. Karanicolas, "Designing allosteric control into enzymes by chemical rescue of structure," *J Am Chem Soc*, vol. 134, no. 24, pp. 10055–10060, Jun. 2012, doi: 10.1021/ja301409g.
[16] "Roses are Red, Violets are Blue, White Florals Smell Like Feces," MIZU. https://www.mizubrand.com/blogs/news/the-story-of-indole-in-natural-perfumery-white-florals (accessed Sep. 02, 2021).
[17] DEATHSCENT, "The Chemistry of Death and Desire," Death/Scent, Jan. 19, 2016. https://deathscent.com/2016/01/19/the-chemistry-of-death-and-desire/ (accessed Sep. 02, 2021).
[18] D. Purves et al., "Olfactory Perception in Humans," Neuroscience. 2nd edition, 2001, Accessed: Sep. 02, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK11032/
[19] "Alkanes and Cycloalkanes | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/B9780128024447000033?token=FEFCD7A3F7F693FD75E71BACDA8758024AEFCA9C23DAE7EFA49FAF39C76FF69E28663D57A777C72B686A1F88E48EDA61&originRegion=us-east-1&originCreation=20210902104614 (accessed Sep. 02, 2021).
[20]"Industrial Gases," Keen Compressed Gas Co. https://www.keengas.com/gases/industrial-gases/ (accessed Sep. 02, 2021).
[21] "Chapter 15 - Odorization | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/B9780123838445000155?token=B969B098F683EF73180BF3AF3FC0FB0D03D98BBA08066A8F9547465DB40548253D3FF37D6A4C9C3BEF8BB1A0B51F585B&originRegion=us-east-1&originCreation=20210902111512 (accessed Sep. 02, 2021).
[22] I. V. Loksha, J. R. Maiolo, C. W. Hong, A. Ng, and C. D. Snow, "SHARPEN-systematic hierarchical algorithms for rotamers and proteins on an extended network," *J Comput Chem*, vol. 30, no. 6, pp. 999–1005, Apr. 2009, doi: 10.1002/jcc.21204.
[23] N. Lorenzo et al., "Laboratory and field experiments used to identify Canis lupus var. familiaris active odor signature chemicals from drugs, explosives, and humans," Anal Bioanal Chem, vol. 376, no. 8, pp. 1212–1224, Aug. 2003, doi: 10.1007/s00216-003-2018-7.
[24] A. M. El-Sayed, V. J. Mitchell, G. F. McLaren, L. M. Manning, B. Bunn, and D. M. Suckling, "Attraction of New Zealand Flower Thrips, Thrips obscuratus, to cis-Jasmone, a Volatile Identified from Japanese Honeysuckle Flowers," J Chem Ecol, vol. 35, no. 6, pp. 656–663, Jun. 2009, doi: 10.1007/s10886-009-9619-3.
[25] "cis-Jasmone stabilized 488-10-8." http://www.sigmaaldrich.com/ (accessed Aug. 14, 2022).
[26] F. Negre, N. Kolosova, J. Knoll, C. M. Kish, and N. Dudareva, "Novel S-adenosyl-l-methionine:salicylic acid carboxyl methyltransferase, an enzyme responsible for biosynthesis of methyl salicylate and methyl benzoate, is not involved in floral scent production in snapdragon flowers," Archives of Biochemistry and Biophysics, vol. 406, no. 2, pp. 261–270, Oct. 2002, doi: 10.1016/S0003-9861(02)00458-7.
[27] Y. Du, A. Zhou, and J. Chen, "Olfactory and behavioral responses to acetate esters in red imported fire ant, Solenopsis invicta," Pest Management Science, vol. 77, no. 3, pp. 1371–1382, 2021, doi: 10.1002/ps.6152.
[28] "methyl salicylate, 119-36-8." http://www.thegoodscentscompany.com/data/rw1008471.html (accessed Aug. 15, 2022).

[29] F. Dosselaere and J. Vanderleyden, “A metabolic node in action: chorismate-utilizing enzymes in microorganisms,” Crit Rev Microbiol, vol. 27, no. 2, pp. 75–131, 2001, doi: 10.1080/20014091096710.
[30] C. Pelludat, D. Brem, and J. Heesemann, “Irp9, Encoded by the High-Pathogenicity Island of Yersinia enterocolitica, Is Able To Convert Chorismate into Salicylate, the Precursor of the Siderophore Yersiniabactin,” J Bacteriol, vol. 185, no. 18, pp. 5648–5653, Sep. 2003, doi: 10.1128/JB.185.18.5648-5653.2003.
[31] J. H. RUTH, “Odor Thresholds and Irritation Levels of Several Chemical Substances: A Review,” American Industrial Hygiene Association Journal, vol. 47, no. 3, p. A-142-A-151, Mar. 1986, doi: 10.1080/15298668691389595.
[32] A. Jahan, K. L. Edwards, and M. Bahraminasab, Multi-criteria Decision Analysis for Supporting the Selection of Engineering Materials in Product Design. Butterworth-Heinemann, 2016.
[33] PubChem, “Hazardous Substances Data Bank (HSDB) : 36.” https://pubchem.ncbi.nlm.nih.gov/source/hsdb/36#section=Solubility-(Complete) (accessed Aug. 22, 2022).
[34] M. E. Rybak, M. R. Sternberg, C.-I. Pao, N. Ahluwalia, and C. M. Pfeiffer, “Urine Excretion of Caffeine and Select Caffeine Metabolites Is Common in the US Population and Associated with Caffeine Intake,” The Journal of Nutrition, vol. 145, no. 4, pp. 766–774, Apr. 2015, doi: 10.3945/jn.114.205476.
[35] C. Willson, “The clinical toxicology of caffeine: A review and case study,” Toxicol Rep, vol. 5, pp. 1140–1152, Nov. 2018, doi: 10.1016/j.toxrep.2018.11.002.
[36] S. N. Tawde, B. Puschner, T. Albin, S. Stump, and R. H. Poppenga, “Death by Caffeine: Presumptive Malicious Poisoning of a Dog by Incorporation in Ground Meat,” J Med Toxicol, vol. 8, no. 4, pp. 436–440, Dec. 2012, doi: 10.1007/s13181-012-0254-y.
[37] R. H. Adamson, “The acute lethal dose 50 (LD50) of caffeine in albino rats,” Regulatory Toxicology and Pharmacology, vol. 80, pp. 274–276, Oct. 2016, doi: 10.1016/j.yrtph.2016.07.011.
[38] J. N. Park, E. Rashidi, K. Foti, M. Zoorob, S. Sherman, and G. C. Alexander, “Fentanyl and fentanyl analogs in the illicit stimulant supply: Results from U.S. drug seizure data, 2011–2016,” Drug and Alcohol Dependence, vol. 218, p. 108416, Jan. 2021, doi: 10.1016/j.drugalcdep.2020.108416.
[39] J. L. Essler et al., “A Randomized Cross-Over Trial Comparing the Effect of Intramuscular Versus Intranasal Naloxone Reversal of Intravenous Fentanyl on Odor Detection in Working Dogs,” Animals, vol. 9, no. 6, p. 385, 2019, doi: 10.3390/ani9060385.
[40] J. L. Essler, P. G. Smith, C. E. Ruge, T. A. Darling, C. A. Barr, and C. M. Otto, “The first responder exposure to contaminating powder on dog fur during intranasal and intramuscular naloxone administration,” Journal of Veterinary Emergency and Critical Care, vol. 32, no. 1, pp. 18–25, 2022, doi: 10.1111/vec.13113.
[41] C. Damien, “Plants and lichens sentinels of air quality,” Encyclopedia of the Environment, Mar. 20, 2018. https://www.encyclopedie-environnement.org/en/air-en/plants-lichens-sentinels-air-quality/ (accessed Aug. 23, 2022).
[42] R. J. Orth et al., “Submersed Aquatic Vegetation in Chesapeake Bay: Sentinel Species in a Changing World,” BioScience, vol. 67, no. 8, pp. 698–712, Aug. 2017, doi: 10.1093/biosci/bix058.

[43] H. Zhuang, H. Han, S. Jia, Q. Zhao, and B. Hou, “Advanced treatment of biologically pretreated coal gasification wastewater using a novel anoxic moving bed biofilm reactor (ANMBBR)–biological aerated filter (BAF) system,” Bioresource Technology, vol. 157, pp. 223–230, Apr. 2014, doi: 10.1016/j.biortech.2014.01.105.
[44] X. Zhang, J. Li, Y. Yu, R. Xu, and Z. Wu, “Biofilm characteristics in natural ventilation trickling filters (NVTFs) for municipal wastewater treatment: Comparison of three kinds of biofilm carriers,” Biochemical Engineering Journal, vol. 106, pp. 87–96, Feb. 2016, doi: 10.1016/j.bej.2015.11.009.
[45] J. P. Wurm et al., “Molecular basis for the allosteric activation mechanism of the heterodimeric imidazole glycerol phosphate synthase complex,” Nat Commun, vol. 12, no. 1, Art. no. 1, May 2021, doi: 10.1038/s41467-021-22968-6.
[46] J. D. Sadowsky, M. A. Burlingame, D. W. Wolan, C. L. McClendon, M. P. Jacobson, and J. A. Wells, “Turning a protein kinase on or off from a single allosteric site via disulfide trapping,” Proc. Natl. Acad. Sci. U.S.A., vol. 108, no. 15, pp. 6056–6061, Apr. 2011, doi: 10.1073/pnas.1102376108.
[47] S. Rana, N. Pozzi, L. A. Pelc, and E. Di Cera, “Redesigning allosteric activation in an enzyme,” Proceedings of the National Academy of Sciences, vol. 108, no. 13, pp. 5221–5225, Mar. 2011, doi: 10.1073/pnas.1018860108.
[48] T. Chien, D. R. Jones, and T. Danino, “Engineered Bacterial Production of Volatile Methyl Salicylate,” ACS Synth. Biol., vol. 10, no. 1, pp. 204–208, Jan. 2021, doi: 10.1021/acssynbio.0c00497.
[49] “IGEM:MIT/2006 - OpenWetWare.” https://openwetware.org/wiki/IGEM:MIT/2006 (accessed Aug. 23, 2022).
[50] C. F. Olson-Manning, M. R. Wagner, and T. Mitchell-Olds, “Adaptive evolution: evaluating empirical support for theoretical predictions,” Nat Rev Genet, vol. 13, no. 12, pp. 867–877, Dec. 2012, doi: 10.1038/nrg3322.
[51] R. E. Cobb, R. Chao, and H. Zhao, “Directed Evolution: Past, Present and Future,” AIChE J, vol. 59, no. 5, pp. 1432–1440, May 2013, doi: 10.1002/aic.13995.
[52] C. P. Long and M. R. Antoniewicz, “How adaptive evolution reshapes metabolism to improve fitness: recent advances and future outlook,” Curr Opin Chem Eng, vol. 22, pp. 209–215, Dec. 2018, doi: 10.1016/j.coche.2018.11.001.
[53] A. Luqman, M. Nega, M.-T. Nguyen, P. Ebner, and F. Götz, “SadA-Expressing Staphylococci in the Human Gut Show Increased Cell Adherence and Internalization,” Cell Reports, vol. 22, no. 2, pp. 535–545, Jan. 2018, doi: 10.1016/j.celrep.2017.12.058.
[54] J. Löfblom, R. Rosenstein, M.-T. Nguyen, S. Ståhl, and F. Götz, “Staphylococcus carnosus: from starter culture to protein engineering platform,” Appl Microbiol Biotechnol, vol. 101, no. 23, pp. 8293–8307, 2017, doi: 10.1007/s00253-017-8528-6.
[55] M. T. Nguyen et al., “The νSaα Specific Lipoprotein Like Cluster (lpl) of S. aureus USA300 Contributes to Immune Stimulation and Invasion in Human Cells,” PLoS Pathog, vol. 11, no. 6, p. e1004984, Jun. 2015, doi: 10.1371/journal.ppat.1004984.
[56] R. Rosenstein et al., “Genome Analysis of the Meat Starter Culture Bacterium Staphylococcus carnosus TM300,” Appl Environ Microbiol, vol. 75, no. 3, pp. 811–822, Feb. 2009, doi: 10.1128/AEM.01982-08.
[57] K. R. Jones, B. R. Belvin, F. L. Macrina, and J. P. Lewis, “Sequence and characterization of shuttle vectors for molecular cloning in Porphyromonas, Bacteroides and related bacteria,” Mol Oral Microbiol, vol. 35, no. 4, pp. 181–191, Aug. 2020, doi: 10.1111/omi.12304.

[58] “Mutation - tnaA739(del)::kan.” https://cgsc.biology.yale.edu/Mutation.php?ID=99778 (accessed Aug. 27, 2022).
[59] “Dam and Dcm Methylases of E. coli | NEB.” https://www.neb.com/tools-and-resources/usage-guidelines/dam-and-dcm-methylases-of-e-coli (accessed Aug. 03, 2022).
[60] P. E. Carrigan, P. Ballar, and S. Tuzmen, “Site-Directed Mutagenesis,” in Disease Gene Identification: Methods and Protocols, J. K. DiStefano, Ed. Totowa, NJ: Humana Press, 2011, pp. 107–124. doi: 10.1007/978-1-61737-954-3_8.
[61] B. Ou, Y. Yang, W. L. Tham, L. Chen, J. Guo, and G. Zhu, “Genetic engineering of probiotic Escherichia coli Nissle 1917 for clinical application,” Appl Microbiol Biotechnol, vol. 100, no. 20, pp. 8693–8699, Oct. 2016, doi: 10.1007/s00253-016-7829-5.
[62] M. K. Akhtar et al., “Hydrogen Peroxide-Based Fluorometric Assay for Real-Time Monitoring of SAM-Dependent Methyltransferases,” Front Bioeng Biotechnol, vol. 6, p. 146, 2018, doi: 10.3389/fbioe.2018.00146.
[63] “SAM510: SAM Methyltransferase Assay, A Non Radioactive Colorimetric Continuous Enzyme Assay,” p. 5.
[64] I. M. Bird, “High performance liquid chromatography: principles and clinical applications.,” BMJ, vol. 299, no. 6702, pp. 783–787, Sep. 1989.
[65] M. Kowalska, M. Woźniak, M. Kijek, P. Mitrosz, J. Szakiel, and P. Turek, “Management of validation of HPLC method for determination of acetylsalicylic acid impurities in a new pharmaceutical product,” Sci Rep, vol. 12, no. 1, p. 1, Dec. 2022, doi: 10.1038/s41598-021-99269-x.
[66] D. Parker, C. Martinez, C. Stanley, J. Simmons, and I. M. McIntyre, “The Analysis of Methyl Salicylate and Salicylic Acid from Chinese Herbal Medicine Ingestion,” Journal of Analytical Toxicology, vol. 28, no. 3, pp. 214–216, Apr. 2004, doi: 10.1093/jat/28.3.214.
[67] S. J. Fleishman et al., “RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite,” PLOS ONE, vol. 6, no. 6, p. e20161, Jun. 2011, doi: 10.1371/journal.pone.0020161.
[68] A. M. Kunjapur, J. C. Hyun, and K. L. J. Prather, “Deregulation of S-adenosylmethionine biosynthesis and regeneration improves methylation in the E. coli de novo vanillin biosynthesis pathway,” Microbial Cell Factories, vol. 15, no. 1, p. 61, Apr. 2016, doi: 10.1186/s12934-016-0459-x.
[69] S. Qian, Y. Li, and P. C. Cirino, “Biosensor-guided improvements in salicylate production by recombinant Escherichia coli,” Microbial Cell Factories, vol. 18, no. 1, p. 18, Jan. 2019, doi: 10.1186/s12934-019-1069-1.
[70] S. Hooshangi, S. Thiberge, and R. Weiss, “Ultrasensitivity and noise propagation in a synthetic transcriptional cascade,” Proceedings of the National Academy of Sciences, vol. 102, no. 10, pp. 3581–3586, Mar. 2005, doi: 10.1073/pnas.0408507102.
[71] C. Kumpitsch, K. Koskinen, V. Schöpf, and C. Moissl-Eichinger, “The microbiome of the upper respiratory tract in health and disease,” BMC Biol, vol. 17, no. 1, p. 87, Nov. 2019, doi: 10.1186/s12915-019-0703-z.
[72] M. Fazlollahi et al., “The nasal microbiome in asthma,” Journal of Allergy and Clinical Immunology, vol. 142, no. 3, pp. 834-843.e2, Sep. 2018, doi: 10.1016/j.jaci.2018.02.020.
[73] M. Hallschmid, “Intranasal Insulin for Alzheimer’s Disease,” CNS Drugs, vol. 35, no. 1, pp. 21–37, Jan. 2021, doi: 10.1007/s40263-020-00781-x.

[74] Z. Li et al., “Metabolic engineering of Escherichia coli for production of chemicals derived from the shikimate pathway,” Journal of Industrial Microbiology and Biotechnology, vol. 47, no. 6–7, pp. 525–535, Jul. 2020, doi: 10.1007/s10295-020-02288-2.
[75] M. Weir and L. Buzhardt, “Taking Your Pets Temperature | VCA Animal Hospital,” Vca. https://vcahospitals.com/know-your-pet/taking-your-pets-temperature (accessed Sep. 27, 2022).
[76] Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:1-11, 2006
[77] Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, Genome Research, 14:1188-1190, (2004)
[78] Schrödinger, LLC, “The PyMOL Molecular Graphics System, Version 1.8,” Nov. 2015.
[79] Inkscape Project. Inkscape [Internet]. Available from: https://inkscape.org, 2020.
[80] “German Collection of Microorganisms and Cell Cultures GmbH: Details.” https://www.dsmz.de/collection/catalogue/details/culture/DSM-4601 (accessed Oct. 22, 2022).

APPENDIX A: SUPPLEMENTAL RESULTS AND DISCUSSION

Adaptive Evolution of High-SA Tolerant *E. coli* Strains

Material and Methods

E. coli Nissle and *dam(-) E. coli* without plasmids, and *dam (-) E. coli* with either pSC-b-amp/kan/mNG or pSC-b-amp/kan/(PlacZ)_SAMT were inoculated into in 5mL TB or 5mL TB + 100ug/mL carbenicillin for the no-plasmid or plasmid strains respectively. SA was added to a concentration of 1mg/mL and cultures were incubated at 28°C while shaking at 210 rpm. Cultures were allowed to incubate until cell density resulted in the growth media becoming translucent to the eye, the density considered significant, at which point the culture was used to inoculate (1:100 (v/v)) a new culture of the same SA concentration. This process was repeated until significant density was reached in 24 hours for 3 subsequent rounds of subculturing, at which point the concentration of SA was increased by 0.5mg/mL. This process was repeated under increasing concentrations and is still actively in progress.

Results

At the point of thesis submission the adaptive evolution cultures were able to survive in SA concentrations of 6.0 mg/mL and 5.5 mg/mL for the no-plasmid strains and the *dam(-)* w/plasmid culture lineages respectively, each of which have been glycerol stocked for posterity. Sequencing of plasmid and chromosomal DNA for all adaptive evolution strains still needs to be done.

Discussion

The adaptive evolution process demonstrates the basic principles behind adaptive evolution in that the *E. coli* subjected to the high [SA] cultures demonstrated increasing tolerance to SA overtime. When the adaptive evolution process began, the limit at which the *E. coli* were

able to survive was approximately 1.0mg/mL SA, and growth at that concentration took over 48 hours of incubation at 28°C to reach significant outgrowth. As cultures were continually incubated and sub-cultured in the high [SA] cultures, they eventually displayed increased rate of growth such that significant outgrowth in 1.0mg/mL occurred within 24 hours of incubation at 28°C, demonstrating the culture lineages had developed an increased tolerance to SA through some undetermined method. As the concentration of SA was increased over time, this same trend of significant outgrowth reached in approximately 48 hours upon initial increases in SA, and significant outgrowth within 24 hours after numerous rounds of sub-culturing. The number of rounds of sub-culturing which were necessary to reach significant outgrowth in 24 hours varied between the various concentrations of SA, but there was no obvious trend to these numbers.

It was observed that sub-culturing into higher SA concentrations was more successful if the cultures which displayed significant outgrowth in 24 hours were allowed to incubate for an additional 24-48 hours after reaching significant outgrowth. When this occurred, the time required to reach significant outgrowth for the first sub-culture into increased [SA] was approximately 24 hours less than when the sub-culturing event happened without allowing the additional time to culture after reaching significant outgrowth. The reason for this trend is unknown, but one hypothesis would be that allowing additional time to incubate in media that is slowly losing its nutrient content could be leading to a greater rate of genetic mutation or metabolic regulatory changes which results in more rapid adaptive evolution. However, further investigation needs to be done to draw any clear conclusions as to how the adaptive evolution process is occurring on a molecular basis.

The observed increase in tolerance to SA in the adaptive evolution strains is promising in that is allows both a higher level of artificial spiking with SA for in-vitro metabolic assays and

allows the upregulation of irp9 in the SA producing strains of *E. coli*, both of which should allow for a higher level of MSA production so long as other metabolic considerations are met. It should be noted that in the glucose spiked cultures expressing pDF-(PblaZ)cSAMT/GFP-irp9, the levels of SA produced in both cultures was relatively similar while the glucose spiked culture produced significantly more MSA. This indicates that just increasing the concentration of SA is not sufficient to increased MSA production, but this does not change the fact that having additional tolerance to the substrate allows for further downstream development of metabolic routes for increased MSA production.

S. carnosus Plasmid Curing

Materials and Methods

5x 5mL cultures of LB inoculated with *S. carnosus* containing pCT20 were incubated at 40°C while shaking at 215rpm. Cultures with significant outgrowth such that cultures were translucent upon visual inspection were used to inoculate fresh 5mL cultures of LB at a 1:100 inoculum level. Upon every subculturing event, 100uL of culture was plated on both LB/agar with 15ug/mL chloramphenicol and non-selective LB/agar and incubated at 37°C until colony growth was observed on the non-selective LB/agar plates. This process was repeated until colony growth was observed on non-selective plates but not on chloramphenicol laced plates. Colonies were selected from the final non-selective plates and streak-plated onto both non-selective and chloramphenicol laced plates. Streak-plated colonies which grew on non-selective media but did not grow on chloramphenicol media were considered to be fully cured and glycerol stocked.

Results and Discussion

The curing process was ultimately believed to be successful in production of a plasmid-cured strain of *S. carnosus* due to the cured *S. carnosus* meeting the standards for proof indicated

in the methods for this assay, and due to a lack of banding in a subsequent gel electrophoretic analysis of a plasmid purification miniprep product for the cured *S. carnosus.* It is worth noting that there were some previous issues in successful miniprep of the *S. carnosus* which did have plasmids in them, with the resultant purified plasmids being very low in concentration and showing unexplained degradation after even brief storage at -20°C. As such, there is the possibility that there were still plasmids in the "cured" *S. carnosus* but that these plasmids were either too low in concentration or too degraded to show up in the gel electrophoretic analysis. However, this is also in no way indicative of the curing not being truly successful but is simply something to keep in mind in case future work shows unexpected plasmid recovery.

Attempted pSC-b-amp/kan/(PblaZ)iSAMT/GFP-(AmyP)irp9 Plasmid Construction

An attempt was made by Ethan Shield to insert the iSAMT/GFP-irp9 gene circuit into the pSC-b-amp/kan backbone; however, subsequent sequencing indicated that the gene circuit was inserted backwards and as such there was no expression of the gene circuit with this plasmid. In the future, attempts will be made to correct this issue and create a functional pSC-b-amp/kan/(PblaZ)iSAMT/GFP-(AmyP)irp9 plasmid.

Technical Issues of Note

Inability to Successfully Observe Plasmid Expression in *S. carnosus*

Initially the project plan was to utilized *S. carnosus* as the nasal microbiome expression organism due to the benefits of *S. carnosus* discussed earlier; however, an inability to demonstrate expression of the *E. coli* produced expression plasmids in *S. carnosus* led to the transition into *E. coli* Nissle as the expression organism. It is not known if the lack of expression observed in *S. carnosus* was due to a lack of successful expression of the *E. coli* expressed plasmids despite these supposedly being *E. coli-Staph* shuttle vectors, or if there was simply a

failure to transform the plasmids into *S. carnosus* in the first place. Due to *S. carnosus* being a gram-positive bacterium with a thick peptidoglycan cell wall, heat shock transformation of chemical competent cells was not an option and instead electroporative transformation was required. While electroporative transformation of *Staphylococcus* species has been reported in various literature, the in-house achievement of electroporative *S. carnosus* transformation was unable to be confirmed despite numerous attempts. In contrast, the techniques for transformation of *E. coli* are well established and validated and project members were already experienced working with *E. coli.* For that reason, the transition into *E. coli* Nissle was chosen to prevent any workflow bottlenecks associated with *S. carnosus.* Despite this change, it would be ideal to eventually return to *S. carnosus* as the model organism for its benefits in application when compared to *E. coli* such as desiccation tolerance and lack of significant indole production.

Difficulties in Measuring Gaseous MSA

One issue that was unable to be overcame by the point of thesis submission was the inability to directly measure the gaseous concentrations of MSA in the headspace of the MSA producing cultures. While the presence of gaseous MSA was able to be confirmed and a qualitative relative concentration was able to be determined via human olfactory screening, and MSA accumulation in the culture media was able to be confirmed via HPLC, measurement of the gaseous MSA which was escaping into the headspace or outside atmosphere was unable to be achieved. This was simply due to a lack of a clearly established method for which to measure gaseous MSA concentrations. While the amount of MSA which evaporated out of solution should theoretically have been calculatable via application of conservation of mass wrt the 1:1 stochiometric conversion of SA to MSA by subtracting the remaining concentration of SA and the accumulated concentration of MSA from the initial concentration of SA. However, due to the

lack of perfect accuracy that is inherent to HPLC analysis, there would be some error in these calculations. As ultimately only consumption of SA was required to demonstrate enzyme activity, and proof of MSA production could be proven via the detection of MSA in the culture media, the inability to directly measure gaseous MSA was not deemed a critical issue to overcome in order to achieve project success. However, it would be ideal to eventually develop a method for direct measurement of gaseous MSA concentrations in order to obtain quantitative data on the concentrations of MSA which occur under various conditions.

Inability to Confirm SAMT Expression with SDS-PAGE

In addition to the inability to confirm SAMT expression via the in-vitro enzymatic assays, attempts at confirming SAMT expression via SDS-PAGE analysis were also unsuccessful. Crude cell lysates containing from *E. coli* expressing either SAMT, SAMT/mNG, or just mNG were His/Ni^{2+}-NTA purified and the crude lysate as well as representative fractions from the His/Ni^{2+}-NTA purification flow through, lysis buffer rise fractions, wash buffer rinse fractions, and final elution fractions were analyzed via SDS-PAGE. While the elution fractions from both samples showed bands at the expected sizes for their respective expressed proteins, the mNG sample also contained bands around the 72kD and 41kD mark which were representative of SAMT/mNG and SAMT respectively (results not shown). These proteins were not in these samples since they were not in the genetic background of the pure mNG culture strains, but this indicated that either the wash step was simply not efficient enough to fully wash all non-his tagged proteins from the chromatography resin, or that there were other endogenous proteins of these sizes which were expressed in the background genetics of that strain. Issues with overcoming these confounding results led to a transition to the colorimetric enzyme activity assay approach.

Failure of in-vitro Colorimetric Analyses

As earlier discussed, the initial methods for determining SAMT enzymatic activity was via various enzyme based colorimetric assays that functioned by tracking the conversion of SAM to SAH via subsequent downstream reactions which produced colored products while simultaneously preventing recycling of SAH back to SAM or product inhibition due to SAH buildup. Some of these methods were adapted from previously published literature while others utilized commercially available kits specifically designed to measure methyltransferase activity, but unfortunately none were able to definitively confirm SAMT activity, let alone determine any enzyme kinetics. One of the reasons for this failure was the fact there was significant signal produced in the crude extracts from *E. coli* expressing the pSC-b-amp/kan/(PlacZ)mNG expression plasmid which indicated that there were methyltransferases besides SAMT which were endogenously expressed in that genetic background. This produced a spectroscopic signal of similar intensity as *E. coli* expressing either pSC-b-amp/kan/(PlacZ)cSAMT or pSC-b-amp/kan/(PlacZ)cSAMT/mNG which prevented any inferences about differences in signal intensity being due to the presence or absence of SAMT. In addition, assay attempts made with His/Ni^{2+}-NTA chromatographically purified samples of SAMT did produce a statistically significant difference in signal intensity compared to the background control reactions which prevented measurement of enzyme activity from these purified samples. It was not known if this was due to a lack of sufficiently concentrated SAMT to produce a signal, or if perhaps something occurred during the His/Ni^{2+}-NTA purification which caused SAMT to lose its catalytic abilities, however reaction controls indicated the assay reactions were working as intended. Validation of HPLC as a method for confirmation of SAMT expression and activity resulted in a transition away from the in-vitro colorimetric assays; however, it would be good to figure out why these

assays did not work so that they may be utilized in the future for more precise determination of methyltransferase kinetics.

Unexpected Protease Degradation in DH5-alpha *E. coli*

Initial attempts at expression of SAMT in the DH5-alpha strain of *E. coli* were analyzed via SDS-PAGE of His/Ni^{2+}-NTA purified SAMT samples. Unfortunately the elution fractions for these samples showed a high concentration of small protein fragments of around 5kD which were not observed in either the lysate buffer rinse or the wash buffer rinse fractions. This indicated that there was successful capture of His-tagged protein, but that these proteins had somehow been degraded into small fragments prior to capture in the His/Ni^{2+}-NTA resin. Fellow Snow lab member Alec Jones deserves credit for suggesting a transition into alternative strains due to his experiences with unexpected protease activity with the DH5-alpha strain. Due to this, a transition was made into the BL21-Gold *E. coli* as they are known to be protease deficient; however, both the JW3686 and the *dam(-)* strains demonstrated a lack of protease degradation for SAMT despite these strains not being marketed as protease deficient, allowing their use in addition to the BL21-Gold strain.

Failures in Construction of iSAMT-mNG and iSAMT-GFP-irp9 expression plasmids

For unknown reasons the attempt at creating an expression plasmid for and iSAMT/mNG fusion protein via HiFi cloning was unsuccessful despite the same reaction being successful for construction of the cSAMT/mNG fusion protein vector. Sequencing of the iSAMT/mNG plasmids demonstrated odd insertions and deletions at various locations in the gene which did not occur in the cSAMT reactions, and it is suspected this was due to some unexpected interaction between the HiFi reaction primers or reaction mixture components and the riboswitch element of the iSAMT plasmid as this was the only difference between the cSAMT and iSAMT

templates. Production of an iSAMT/mNG expression vector could be reattempted, although currently there is no explicit need for such a vector.

Initial Failures with Cloning Attempts

Early attempts at cloning the SAMT g-blocks into the pUC19 vector backbone were made using basic restriction digestion techniques using enzymes such as EcoRI and XbaI. For unknown reasons, these cloning attempts did not produce the results expected, and gel electrophoretic analysis of the reaction products seems to indicate issues with successful digestion by both restriction enzymes used. Various other restriction enzymes were used in place of EcoRI and XbaI in an attempt to overcome this issue but without success. In addition new restriction enzymes, pipette tips, nuclease fee water, and other reagents/equipment used in the cloning reactions under the hypothesis that there was something wrong with one of the commercially purchased reagents or equipment which were causing the reactions to fail, but none of the mitigating attempts were successful. Due to the continued issues with the restriction digestion cloning methods, the Agilent Strataclone Blunt End cloning kit was used and was successful on the first attempt, eliminating the need for further troubleshooting of the restriction digestion reactions. While it would be interesting to know why these reactions were continually unsuccessful, that information is not necessary for project success at this point in time.

APPENDIX B: SUPPLEMENTAL INFORAMTION AND DATA

S. carnosus background

Initially, the organism for which the final expression system was being designed was the bacterial species *Staphylococcus carnosus*. This organism was chosen to be the species which would be used to inoculate the sentinel animal's nasal microbiome and express the final genetic constructs due to its safety, ease of on-site deployment, and previous development of species-specific engineering techniques. *S. carnosus* is naturally occurring human enterobacteria which has been shown to play a role in diverse actions ranging from basic digestion to production of neuromodulating amine compounds [53]. Regarded as a food safe organism, *S. carnosus* has been used to control and optimized the fermentation processes involved in sausage production since the 1950's, acting to both provide unique flavors as well as preventing the growth of potentially pathogenic bacterial species [54]. In addition, *S. carnosus* has been shown to lack the pathogenic genetics found in other *Staph* species and is commonly used to investigate the pathogenic nature of other *Staph* species [55]. In addition to the general safety of the species, *S. carnosus* was also considered an ideal choice of expression species due to the fact the entire genome of the TM300 strain of *S. carnosus* has been sequenced, allowing an informed understanding of potential genetic modifications and how to best engineer these modifications into the *S. carnosus* TM300 strain [56]. Another benefit is that *S. carnosus* is commonly sold in a freeze-dried form, and this ability to be safely freeze-dried and then reconstituted on demand makes *S. carnosus* ideal for use in scenarios in which culturing conditions for *E. coli* would not be feasible such as on-site deployment. For these reasons, *S. carnosus* was considered a safe organism for inoculation into the nasal microbiome of animal sentinels, and an ideal organism

for potential genetic engineering. However, due to the aforementioned difficulties with successful transgenic expression in *S. carnosus* the choice was ultimately made to transition to *E. coli Nissle* as the expression organisms of choice.

Animal Study Training Protocol Developed by Michael Burns

Preparation

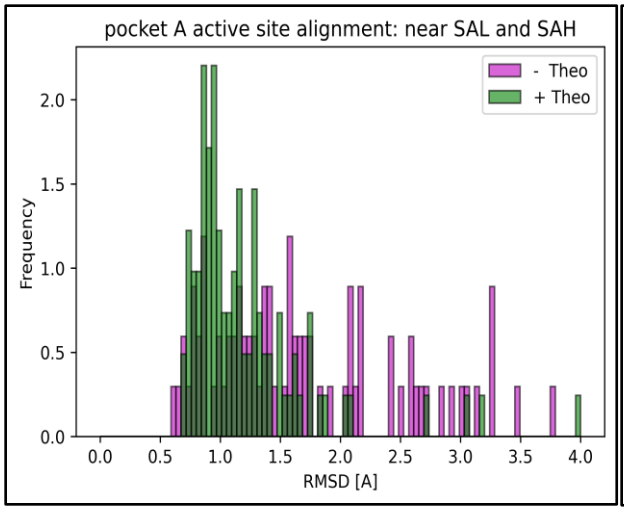
- When first entering the training room, get out the treats and put them into the black tray
- Turn on the diffuser with the theophylline mixture, we will fill the diffuser to the recommended amount. Leave it running while you do the rest of training.
- Turn on the ventilation system as well
- Now put on gloves, we don't want the gloves or the ambient smell of the room to wintergreen. This means that if you are to spray the wintergreen scent with one hand, make sure not to touch anything else with this hand. The scent is quite strong and will get everywhere. The best idea so far is to take the hand that you sprayed wintergreen with, and with glove on, put it in the lab-coat pocket, do not touch anything besides the wintergreen spray bottle with this hand. The other worker with you will take care of the activities you cannot perform with one hand.
- You can go and get one set of rats now, removing the water bottle ahead of time so that they don't get wet.
- One person will administer the scent and remove the divider. The other person will press record on the filming device, give the treat when the correct lever is pressed and remove the levers after the treat is given.

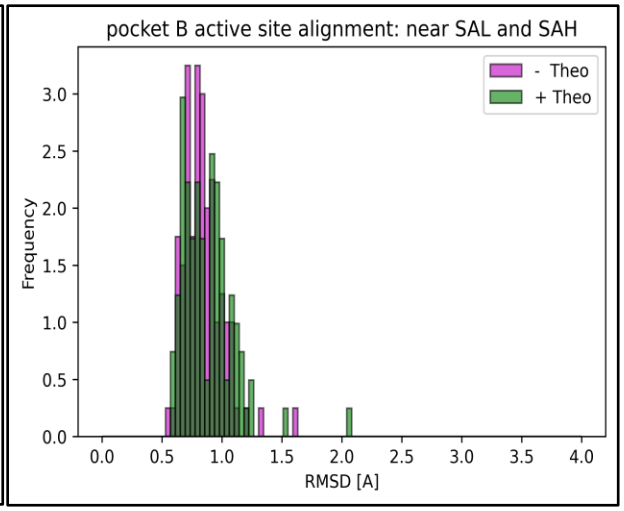
Testing protocol

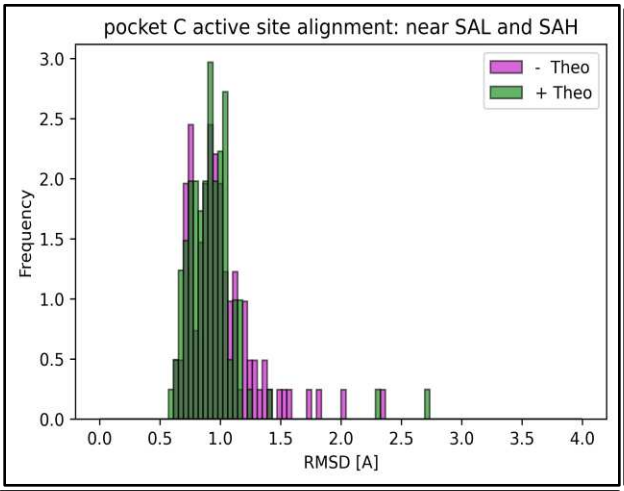
- This is the most important of all the sections. We must keep it consistent.
- We all want to treat these rats like the cute creatures they are, but please do not give verbal praise after they press a lever or any sort of cues for them to pick up on.
- The treat is their only reward and the only stimulus they should have is the scent administer to the cage. AKA no calling to it and no tapping.
- Both trainers should have both hands free, no need to take notes.

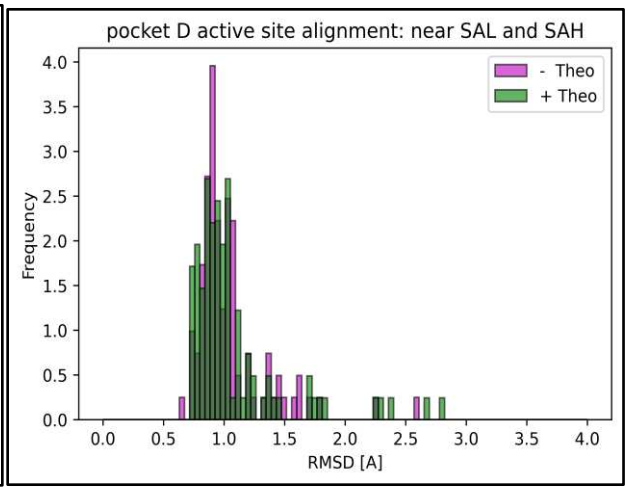
- The scent will either be wintergreen or no scent. Based on a random number generator or coin flip. If the number is 1 on the generator use wintergreen. If the coin flip is heads, use wintergreen.
- Select the marked rat first, and place into the training cage. Put them on the opposite side of the plastic divider. The levers should be pushed inside of the cage, but the plastic blocking the rat from touching them. Give time to acclimate to the training cage.
- Start to record using the filming device so we can verify consistency, and data collection.
- Verbally say whether there is wintergreen or no scent for the camera. Also say last two digits of the cage number and the number on the rats tail. Just loud enough for the microphone to pick it up.
- While the rat is in the cage spray the scent, make sure they smell it. Once you have made sure that the rat has smelled a sufficient amount of the scent, lift the divider and observe.
- You must reward the rat as soon as it presses the correct lever. Nothing happens if the rats are to press the wrong lever.
- Once the rats presses a lever remove the two levers from the cage.
- If the rat got the lever correct, let it finish eating its treat, then place it back in the home cage. If the rat got the lever in incorrect, place it back in the home cage without the reward of a treat. Each rat gets one training attempt for a treat per training session. This means you will have 2 attempts total if there are 2 rats per home cage. (Each rat will be tested once a day as long as we keep to the schedule).
- Make sure to write down in this moment, when there is no rat in the training cage whether the wintergreen scent was used, or the no scent, also write down whether the tail of the rat was marked or not. The rat with the mark is P (for positive), and unmarked rats are N (for negative.)

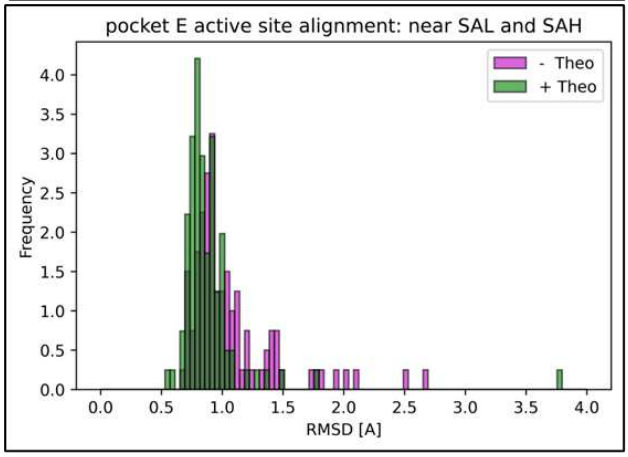
Comparative RMSD Histograms

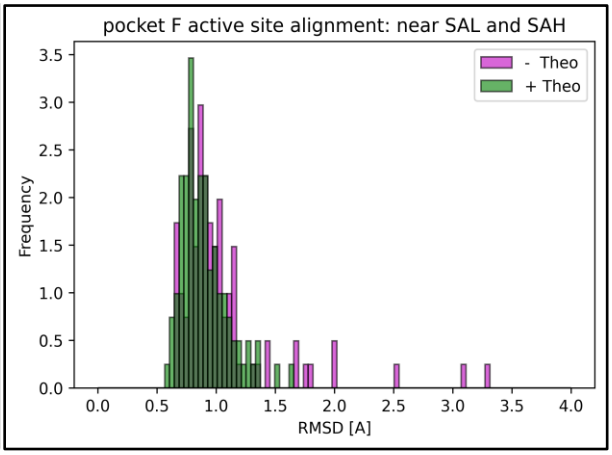












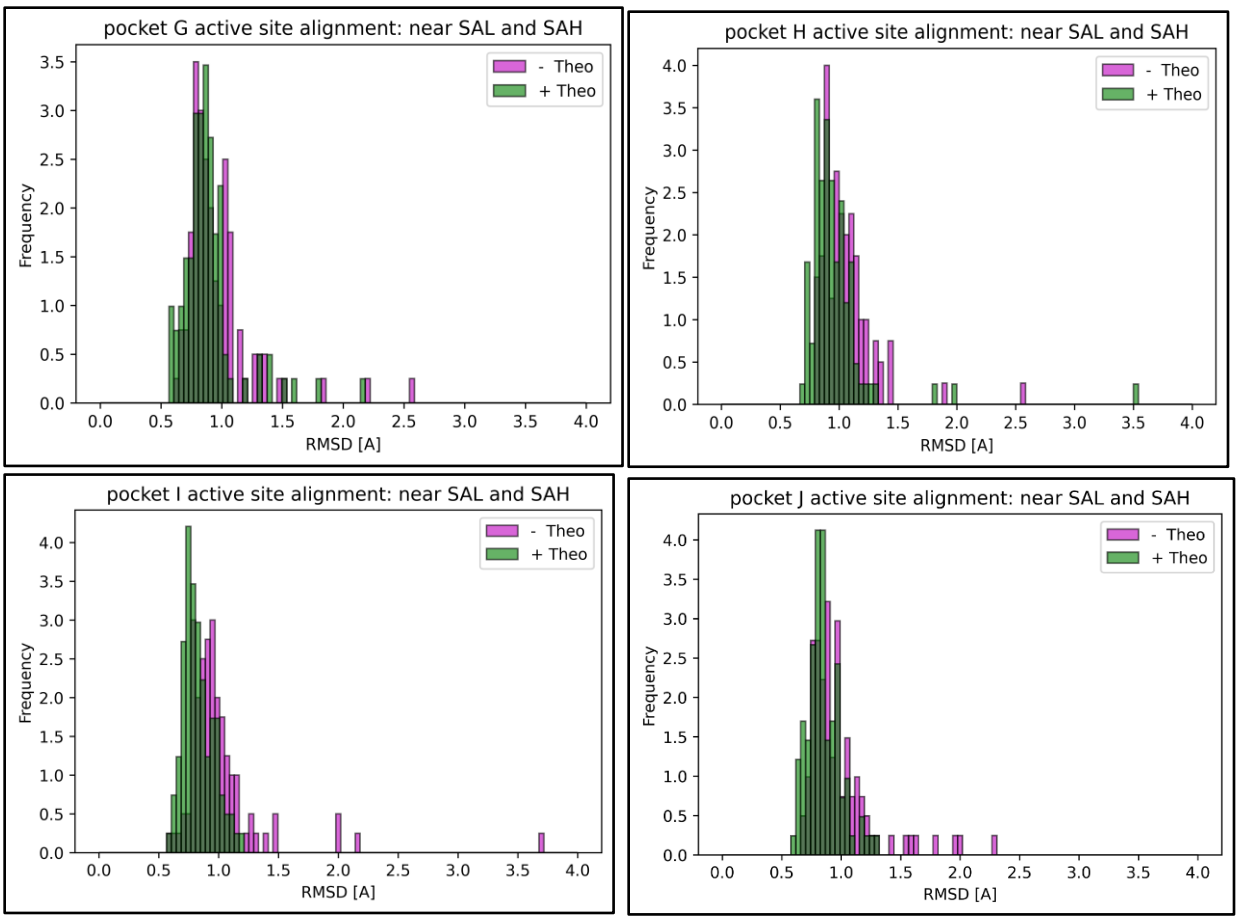
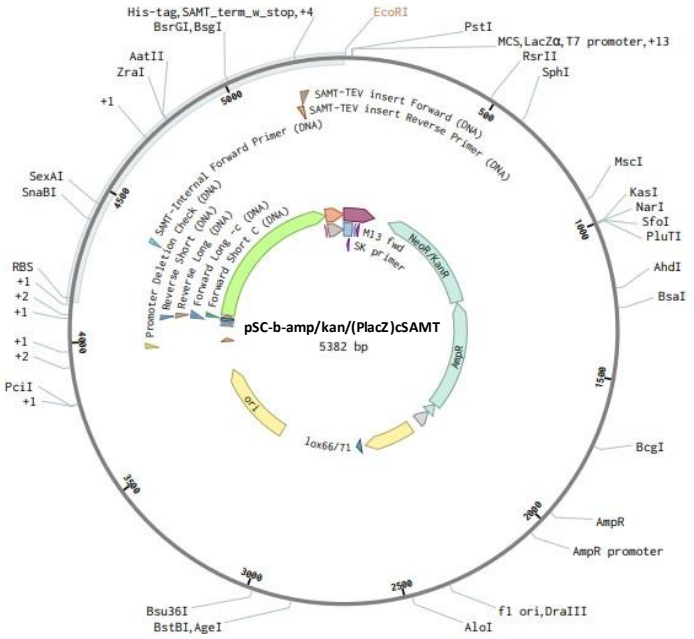
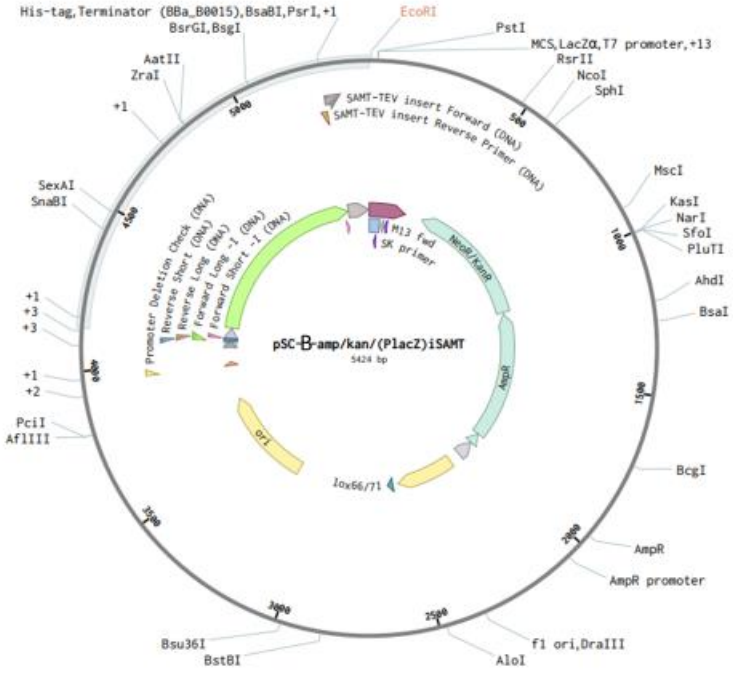


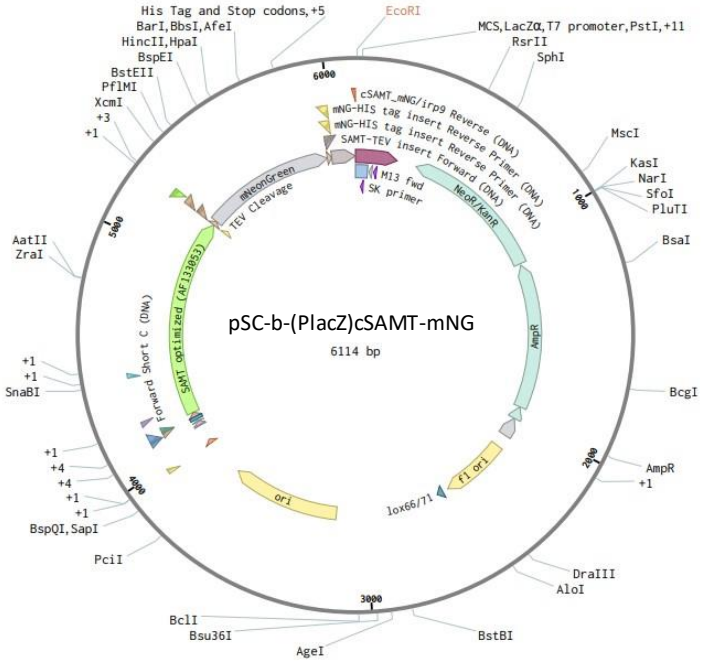
Table of Pocket-Specific Visually Identified Potential Allosteric Routes

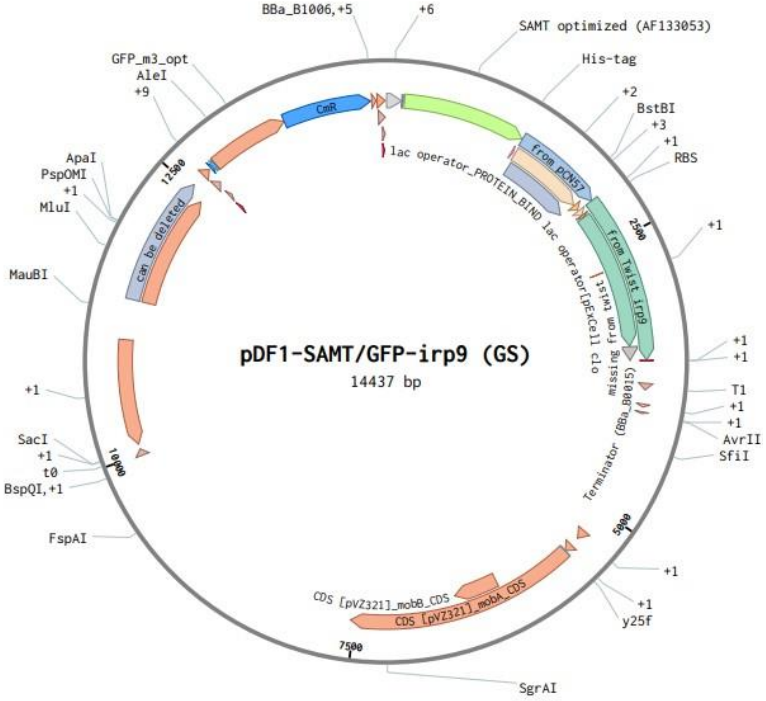
Appendix Table 1: Visually Identified Routes for Potential Allosteric Translation		
Pocket	**Visually Identified Potential Route Residues**	**Active Site Connection**
A B C	alpha-helix (312-320) sheet-helix (248-253) beta-barrel (229-233)	(V311, Y255, and W226) SA structural connection
F G H	Residues (146-153)	(146-153) SA/SAM coordination (129-131) & (59-61) SAM coordination
I J	alpha-helix (25-29); residues (143-145)	(25-29) SAM interaction; Y147 disturbance
D E	N/A	Allosteric Modelling Only

Plasmid Maps






pSC-b-(PlacZ)cSAMT-mNG
6114 bp
His Tag and Stop codons,+5
EcoRI
MCS,LacZα,T7 promoter,PstI,+11
BarI,BbsI,AfeI
HincII,HpaI
BspEI
BstEII
PflMI
XcmI
+3
+1
RsrII
SphI
MscI
KasI
NarI
SfoI
PluTI
BsaI
BcgI
AmpR
+1
DraIII
AloI
BstBI
AgeI
Bsu36I
BclI
BspQI,SapI
+1
+4
+4
+1
+1
PciI
SnaBI
+1
+1
AatII
ZraI
mNeonGreen
TEV Cleavage
SAMT optimized (AF133053)
Short C (DNA)
primer
cSAMT_mNG/irp9 Reverse (DNA)
mNG-HIS tag insert Reverse Primer (DNA)
mNG-HIS tag insert Reverse Primer (DNA)
SAMT-TEV insert forward (DNA)
M13 fwd
SK primer
NeoR/KanR
AmpR
f1 ori
lox66/71
ori
1000
2000
3000
4000
5000
6000


pDF1-SAMT/GFP-irp9 (GS)
14437 bp
BBa_B1006,+5
+6
SAMT optimized (AF133053)
His-tag
+2
BstBI
+3
+1
RBS
+1
GFP_m3_opt
AleI
+9
CmR
lac operator_PROTEIN_BIND
from pCN57
from twist irp9
lac operator
pEvCell clo
missing from twist
Terminator (BBa_B0015)
T1
+1
+1
AvrII
SfiI
+1
+1
y25f
+1
SgrAI
CDS [pVZ321]_mobA_CDS
CDS [pVZ321]_mobB_CDS
FspAI
+1
BspQI,+1
t0
SacI
+1
MauBI
+1
can be deleted
MluI
+1
PspOMI
ApaI
2500
5000
7500
10000
12500

Gene Sequences

Appendix Table 2: Gene Sequences	
SAMT	ATGGATGTTAGACAAGTATTACACATGAAAGGTGGTGCGGGCGAAAATAGTTATGCAATGAAcTCT TTCATTCAACGTCAAGTTATATCAATTACTAAGCCAATCACAGAAGCTGCAATCACGGCATTATATAG CGGAGATACAGTTACAACAAGATTGGCTATTGCAGATCTTGGATGTAGTAGTGGCCCTAACGCATT ATTCGCTGTAACTGAACTTATTAAAACGGTTGAAGAATTACGTAAAAAAATGGGTCGTGAAAACTCA CCGGAATATCAGATCTTTTTAAATGACTTACCTGGTAACGATTTTAACGCAATATTCAGAAGtTTGCC GATCGAAAATGATGTTGATGGAGTGTGTTTTATCAATGGTGTTCCAGGTAGTTTTTATGGACGTTTA TTCCCACGTAACACATTACACTTTATTCATTCATCTTATTCTTTAATGTGGTTATCACAAGTGCCAATA GGTATTGAATCTAACAAAGGCAACATTTATATGGCTAATACATGCCCTCAATCAGTATTAAATGCAT ACTATAAGCAATTTCAAGAAGATCATGCATTATTCTTACGTTGTCGTGCTCAAGAAGTGGTGCCGGG TGGTAGAATGGTTTTAACGATTTTAGGACGTCGTAGTGAAGATCGTGCATCTACAGAGTGTTGCTTA ATTTGGCAATTATTAGCTATGGCTTTGAACCAAATGGTCAGTGAGGGATTAATCGAAGAAGAAAAA ATGGATAAATTCAATATTCCACAATATACTCCTTCACCTACAGAAGTTGAAGCAGAAATTTTGAAAG AAGGTTCTTTTTTAATTGACCATATTGAAGCATCAGAAATTTACTGGAGTAGCTGTACAAAGGATGG TGATGGTGGAGGTTCTGTTGAAGAAGAGGGTTATAATGTAGCAAGATGTATGCGTGCAGTTGCAG AACCTTTATTGTTGGACCATTTCGGTGAAGCGATCATCGAAGATGTTTTTCATCGTTATAAACTTTTA ATCATCGAACGTATGTCTAAAGAAAAAACTAAATTTATCAATGTAATTGTAAGTTTAATACGCAAAA GTGAT
irp9	ATGAAAATTTCAGAGTTTTTACATTTGGCGTTACCGGAAGAACAATGGTTACCTACTATTTCAGGTGT GTTGCGTCAGTTCGCAGAAGAGGAATGTTACGTTTACGAACGTCAACCATGTTGGTATTTAGGTAAA GGATGCCAAGCTagaCTTCATATCAATGCAGACGGTACACAAGCAACATTTATTGATGATGCtGGTG AACAAAAATGGGCTGTTGATTCAATAGCGGATTGTGCACGTCGATTTATGGCACATCCACAAGTTAA AGGTCGTAGAGTTTATGGCCAGGTCGGTTTTAATTTTGCTGCACATGCGCGTGGCATTGCATTTAAC GCtGGTGAATGGCCATTATTAACATTGACAGTACCTCGTGAAGAACTTATTTTTGAAAAAGGAAATG TAACGGTGTATGCTGATAGTGCTGATGGCTGTCGTCGTTTATGCGAATGGGTTAAAGAAGCAAGCA CAACAACTCAAAATGCTCCTTTAGCAGTAGATACGGCATTGAATGGCGAAGCATACAAACAACAAG TAGCACGAGCAGTTGCTGAAATTCGCCGTGGTGAATATGTCAAAGTTATAGTGTCACGTGCAATACC ATTACCAAGTCGTATAGATATGCCAGCTACATTATTATACGGTCGTCAAGCGAACACACCTGTTCGC TCATTTATGTTTCGACAAGAAGGCCGTGAGGCATTAGGTTTCAGCCCAGAATTAGTAATGTCTGTAA CAGGCAATAAAGTCGTGACAGAACCTCTTGCAGGTACTCGAGATCGAATGGGAAACCCAGAACATA ACAAAGCAAAGGAAGCGGAATTATTGCATGATTCTAAGGAAGTTTTGGAACATATCTTATCAGTAA AAGAAGCAATCGCAGAACTTGAAGCAGTATGCCTTCCTGGTTCAGTTGTAGTAGAAGAtCTTATGAG CGTACGTCAACGAGGCTCTGTACAACATTTAGGCAGCGGTGTCTCAGGTCAATTGGCTGAAAATAA AGATGCTTGGGATGCATTTACAGTGTTATTTCCGAGTATCACAGCATCAGGTATCCCTAAAAATGCA GCATTGAATGCGATTATGCAGATTGAGAAAACACCTCGTGAGTTATATAGTGGCGCGATTTTATTAT TGGATGATACGCGATTCGACGCAGCATTAGTGTTACGTAGTGTGTTCCAAGATTCACAAAGATGTTG GATTCAAGCTGGTGCAGGAATCATTGCTCAAAGCACACCTGAACGAGAGTTAACAGAAACACGCGA AAAATTAGCGAGCATTGCACCTTACTTAATGGTTTGA

mNeon Green	ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCTCTCTCCCAGCGACACATGAGTTACACATCTTT GGCTCCATCAACGGTGTGGACTTTGACATGGTGGGTCAGGGCACCGGCAATCCAAATGATGGTTAT GAGGAGTTAAACCTGAAGTCCACCAAGGGTGACCTCCAGTTCTCCCCCTGGATTCTGGTCCCTCATA TCGGGTATGGCTTCCATCAGTACCTGCCCTACCCTGACGGGATGTCGCCTTTCCAGGCCGCCATGGT AGATGGCTCCGGATACCAAGTCCATCGCACAATGCAGTTTGAAGATGGTGCCTCCCTTACTGTTAAC TACCGCTACACCTACGAGGGAAGCCACATCAAAGGAGAGGCCCAGGTGAAGGGGACTGGTTTCCC TGCTGACGGTCCTGTGATGACCAACTCGCTGACCGCTGCGGACTGGTGCAGGTCGAAGAAGACTTA CCCCAACGACAAAACCATCATCAGTACCTTTAAGTGGAGTTACACCACTGGAAATGGCAAGCGCTAC CGGAGCACTGCGCGGACCACCTACACCTTTGCCAAGCCAATGGCGGCTAACTATCTGAAGAACCAG CCGATGTACGTGTTCCGTAAGACGGAGCTCAAGCACTCCAAGACCGAGCTCAACTTCAAGGAGTGG CAAAAGGCCTTTACCGATGTGATGGGCATGGACGAGCTGTACAAG
GFP-mut	ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTA ATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTA AATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGCGTATGGT CTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCG AAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAG TCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGG AAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAA CAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTA GCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCT GTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTT GTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAA

Table of Pareto Optimized Libraries for Each Theophylline Pocket

Appendix Table 3: List of all Pareto Optimized Libraries for All Theophylline Pockets			
Pocket	**Pareto Optimized Libraries**	**# of Variants Covered**	**Total Library Size**
A	(('T',), ('L',), ('I',), ('L',), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	1	14
	(('T',), ('L',), ('I',), ('I', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	2	22
	(('T',), ('L',), ('I',), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	3	23
	(('T',), ('I', 'L'), ('I',), ('I', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	4	29
	(('T',), ('I', 'L'), ('I',), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	6	34
	(('G', 'T'), ('L',), ('I',), ('I', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	8	39
	(('T',), ('I', 'L'), ('I', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	12	44
	(('G', 'T'), ('I', 'L'), ('I',), ('I', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	16	47
	(('T',), ('I', 'L'), ('I', 'M', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('L',))	18	48
	(('T',), ('I', 'L'), ('I', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	24	58
	(('T',), ('I', 'L'), ('I', 'M', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	36	62
	(('G', 'T'), ('I', 'L'), ('I', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	48	70
	(('G', 'T'), ('I', 'L'), ('I', 'V', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	72	74
	(('G', 'T'), ('I', 'L'), ('I', 'V', 'M', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	96	78

	(('G', 'T'), ('I', 'M', 'L'), ('I', 'V', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	108	80
	(('G', 'T'), ('I', 'M', 'L'), ('I', 'V', 'M', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	144	84
	(('G', 'T'), ('I', 'M', 'L'), ('I', 'V', 'M', 'L'), ('I', 'W', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('V', 'L'))	192	85
	(('G', 'T'), ('I', 'M', 'L'), ('I', 'V', 'M', 'L'), ('I', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('I', 'V', 'L'))	216	88
	(('G', 'T'), ('I', 'M', 'L'), ('I', 'V', 'M', 'L'), ('I', 'W', 'M', 'L'), ('V',), ('L',), ('T',), ('V',), ('A',), ('I', 'V', 'L'))	288	89
	(('G', 'T'), ('I', 'M', 'L'), ('I', 'V', 'M', 'L'), ('I', 'W', 'M', 'L'), ('V',), ('L',), ('N', 'T'), ('V',), ('A',), ('I', 'V', 'L'))	576	90
B	(('I',), ('L',), ('I',), ('M',), ('P',), ('T',), ('P',), ('M',), ('A',), ('M',), ('M',), ('F',), ('V',), ('F',), ('I',))	1	24
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('M',), ('A',), ('M',), ('M',), ('F',), ('V',), ('F',), ('I',))	2	39
	(('I',), ('L',), ('I',), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('M',), ('A',), ('M',), ('M',), ('F',), ('V',), ('F',), ('I',))	3	44
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('M',), ('A',), ('M',), ('L', 'M'), ('F',), ('V',), ('F',), ('I',))	4	54
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('M',), ('A',), ('M',), ('L', 'I', 'M'), ('F',), ('V',), ('F',), ('I',))	6	60
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('M',), ('L', 'M'), ('F',), ('V',), ('F',), ('I',))	8	69
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('M',), ('L', 'I', 'M'), ('F',), ('V',), ('F',), ('I',))	12	76
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('I', 'M'), ('L', 'M'), ('F',), ('V',), ('F',), ('I',))	16	83
	(('I',), ('L',), ('I',), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('M',), ('L', 'I', 'M'), ('F',), ('V',), ('F',), ('I',))	18	89
	(('I',), ('L',), ('I',), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'I', 'M'), ('L', 'M'), ('F',), ('V',), ('F',), ('I',))	24	96
	(('I',), ('L',), ('I', 'M'), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('I', 'M'), ('L', 'M'), ('F',), ('V',), ('F',), ('I',))	32	102
	(('I',), ('L',), ('I',), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('L', 'I', 'M'), ('F',), ('V',), ('F',), ('I',))	36	112
	(('I',), ('L',), ('I', 'M'), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'I', 'M'), ('L', 'M'), ('F',), ('V',), ('F',), ('I',))	48	117
	(('I',), ('L',), ('I',), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F',), ('V',), ('F',), ('I',))	54	126
	(('I',), ('L',), ('V', 'I'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('M',), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	72	140
	(('I',), ('L',), ('I', 'M'), ('V', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	96	141
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('M',), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	108	164
	(('I',), ('L',), ('I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	144	172

	(('I',), ('L',), ('I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	192	180
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	216	209
	(('I',), ('L',), ('I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	288	218
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I',))	324	238
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	432	255
	(('I',), ('L',), ('I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A', 'S'), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	576	256
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A',), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	648	288
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A', 'S'), ('L', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	864	293
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A', 'S'), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F',), ('I', 'M'))	1296	332
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'M'), ('A', 'S'), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F', 'I'), ('I', 'M'))	2592	346
	(('I',), ('L',), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'I', 'M'), ('A', 'S'), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F', 'I'), ('I', 'M'))	3888	360
	(('I',), ('L', 'I'), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('P',), ('L', 'I', 'M'), ('A', 'S'), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F', 'I'), ('I', 'M'))	7776	372
	(('I',), ('L', 'I'), ('V', 'I', 'M'), ('V', 'I', 'M'), ('P',), ('T',), ('T', 'P'), ('L', 'I', 'M'), ('A', 'S'), ('L', 'I', 'M'), ('L', 'I', 'M'), ('F', 'M'), ('V',), ('F', 'I'), ('I', 'M'))	15552	378
C	(('I',), ('L',), ('L',), ('A',), ('I',), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V',), ('V',), ('I',), ('V',))	1	15
	(('I',), ('L',), ('L',), ('A',), ('I',), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V',), ('V', 'I'), ('I',), ('V',))	2	24
	(('I',), ('L',), ('L',), ('A',), ('I',), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V',), ('V', 'I', 'M'), ('I',), ('V',))	3	28
	(('I',), ('L',), ('L',), ('A',), ('V', 'I'), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V',), ('V', 'I'), ('I',), ('V',))	4	30
	(('I',), ('L',), ('L',), ('A',), ('V', 'I'), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V',), ('V', 'I', 'M'), ('I',), ('V',))	6	34
	(('I',), ('L',), ('L',), ('A',), ('V', 'I'), ('I', 'M'), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V',), ('V', 'I', 'M'), ('I',), ('V',))	12	35
	(('I',), ('L',), ('I',), ('A',), ('L',), ('M',), ('V',), ('A',), ('T',), ('P',), ('A',), ('M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	16	37
	(('I',), ('L',), ('I', 'L'), ('A',), ('V', 'I'), ('M',), ('V',), ('T',), ('T',), ('P',), ('A',), ('I',), ('V', 'I'), ('V', 'I', 'M'), ('I',), ('V',))	24	38
	(('I',), ('L',), ('I',), ('A',), ('L',), ('M',), ('V',), ('P', 'A'), ('T',), ('P',), ('A',), ('M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	32	43
	(('I',), ('L',), ('I',), ('A',), ('L',), ('M',), ('V',), ('A',), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	48	46
	(('I',), ('L',), ('I',), ('A',), ('L',), ('M',), ('V',), ('P', 'A'), ('T',), ('P',), ('S', 'A'), ('M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	64	51

	(('I',), ('L',), ('I',), ('A',), ('L',), ('M',), ('V',), ('P', 'A'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	96	57
	(('I',), ('L',), ('I',), ('A',), ('L',), ('M',), ('V',), ('P', 'A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	144	58
	(('I',), ('L',), ('I',), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('M', 'L'), ('I', 'L'), ('M', 'F'), ('L',), ('V',))	192	62
	(('I',), ('L',), ('I',), ('A',), ('L', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	256	63
	(('I',), ('L',), ('I',), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('I', 'L'), ('M', 'F'), ('L',), ('V',))	288	71
	(('I',), ('L',), ('I',), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	384	77
	(('I',), ('L',), ('I',), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	432	79
	(('I',), ('L',), ('I', 'L'), ('A',), ('V', 'I'), ('I', 'M'), ('V',), ('T',), ('T',), ('P',), ('S', 'A'), ('I', 'M'), ('V', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	512	80
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('I', 'L'), ('M', 'F'), ('L',), ('V',))	576	102
	(('I',), ('L',), ('I', 'L'), ('A',), ('V', 'I'), ('I', 'M'), ('V',), ('T',), ('T',), ('P',), ('S', 'A'), ('I', 'M'), ('V', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	768	103
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'L'), ('M', 'F'), ('L',), ('V',))	864	112
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	1152	120
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('P', 'A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'L'), ('M', 'F'), ('L',), ('V',))	1296	123
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('P', 'A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('L',), ('V',))	1728	131
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('M', 'F'), ('I', 'L'), ('V',))	2304	138
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'L'), ('V', 'M', 'F'), ('I', 'L'), ('V',))	2592	156
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	3456	176
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	4608	184
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('P', 'A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	5184	187
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'I'), ('I', 'M'), ('V',), ('P', 'A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	6912	195
	(('I',), ('L',), ('I', 'L'), ('A',), ('L', 'V', 'A', 'I'), ('I', 'M'), ('V',), ('P', 'A', 'T'), ('T',), ('P',), ('S', 'A'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('V', 'I', 'M', 'F'), ('I', 'L'), ('V',))	9216	203
D	(('L',), ('I',), ('V',), ('T',), ('I',), ('M',), ('V',), ('L',), ('I',), ('T',), ('V',), ('V',), ('C',), ('I',))	1	7
	(('L',), ('M',), ('V',), ('T',), ('M',), ('L',), ('I', 'L'), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V',))	2	8
	(('L',), ('I', 'M'), ('V',), ('T',), ('M',), ('L',), ('I', 'L'), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V',))	4	12
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V',))	8	14
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'F', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V',))	12	17
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V', 'L'))	16	18

	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'F', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V', 'L'))	24	21
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'L'), ('I', 'L'), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V', 'L'))	32	22
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('V', 'N'), ('C',), ('V', 'I', 'L'))	48	26
	(('I', 'L'), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'F', 'L'), ('V', 'I', 'L'), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V',))	72	30
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'F', 'L'), ('I',), ('L',), ('I',), ('T',), ('V',), ('V', 'A', 'N'), ('C',), ('V', 'I', 'L'))	108	33
	(('I', 'L'), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'F', 'L'), ('V', 'I', 'L'), ('L',), ('I',), ('T',), ('V',), ('N',), ('C',), ('V', 'L'))	144	34
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I'), ('L',), ('I',), ('T',), ('V',), ('V', 'N'), ('C',), ('V', 'I', 'L'))	192	36
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'L'), ('V', 'I', 'L'), ('L',), ('I',), ('T',), ('V',), ('V', 'N'), ('C',), ('V', 'I', 'L'))	216	37
	(('L',), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I'), ('L',), ('I',), ('T',), ('V',), ('V', 'A', 'N'), ('C',), ('V', 'I', 'L'))	288	40
	(('I', 'L'), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I'), ('L',), ('I',), ('T',), ('V',), ('V', 'N'), ('C',), ('V', 'I', 'L'))	384	45
	(('I', 'L'), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I', 'L'), ('L',), ('I',), ('T',), ('V',), ('V', 'N'), ('C',), ('V', 'I', 'L'))	576	50
	(('I', 'L'), ('I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I', 'L'), ('L',), ('I',), ('T',), ('V',), ('V', 'A', 'N'), ('C',), ('V', 'I', 'L'))	864	54
	(('I', 'L'), ('V', 'I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I'), ('L',), ('I', 'L'), ('T',), ('V',), ('V', 'A', 'N'), ('C',), ('V', 'I', 'L'))	1728	55
	(('I', 'L'), ('V', 'I', 'M'), ('V',), ('T',), ('I', 'M'), ('I', 'M', 'F', 'L'), ('V', 'I', 'L'), ('L',), ('I', 'L'), ('T',), ('V',), ('V', 'A', 'N'), ('C',), ('V', 'I', 'L'))	2592	60
E	(('L',), ('V',), ('T',), ('V',), ('F',), ('L',), ('L',), ('I',), ('L',), ('V',), ('A',), ('I',))	1	20
	(('L',), ('V',), ('T',), ('V',), ('F',), ('L',), ('L',), ('I',), ('L',), ('A',), ('C',), ('I', 'L'))	2	26
	(('L',), ('V',), ('T',), ('V',), ('F',), ('L',), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I',))	4	38
	(('L',), ('V',), ('T',), ('V',), ('F',), ('L',), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I', 'L'))	8	52
	(('L',), ('V',), ('T',), ('V',), ('F',), ('M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I', 'L'))	16	63
	(('L',), ('V',), ('T',), ('V',), ('F', 'Y', 'I'), ('L',), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I', 'L'))	24	65
	(('L',), ('V',), ('T',), ('V',), ('F', 'Y'), ('M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I', 'L'))	32	72
	(('L',), ('V',), ('T',), ('V',), ('F', 'Y', 'I'), ('M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'C'), ('I', 'L'))	48	76
	(('L',), ('V',), ('T',), ('V',), ('F', 'Y'), ('I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	72	82
	(('L',), ('V',), ('T',), ('V',), ('F', 'Y', 'I'), ('I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	108	87
	(('L',), ('V',), ('T',), ('V', 'I'), ('F', 'Y'), ('V', 'I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	192	92

	(('L',), ('V',), ('T',), ('V', 'I'), ('F', 'Y', 'I'), ('V', 'I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	288	97
	(('L',), ('V',), ('T',), ('V', 'M'), ('F', 'L', 'Y', 'M', 'I'), ('I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	360	100
	(('L',), ('V',), ('T',), ('V', 'I', 'M'), ('F', 'L', 'Y'), ('V', 'I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	432	101
	(('L',), ('V',), ('T',), ('V', 'I', 'M'), ('F', 'L', 'Y', 'I'), ('V', 'I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	576	106
	(('L',), ('V',), ('T',), ('V', 'I', 'M'), ('F', 'L', 'Y', 'M', 'I'), ('V', 'I', 'M', 'L'), ('L',), ('I',), ('L',), ('V', 'A'), ('A', 'G', 'C'), ('I', 'L'))	720	110
F	(('A',), ('I',), ('L',), ('Y',), ('L',), ('I',), ('L',), ('A',), ('V',), ('V',), ('V',))	1	10
	(('A',), ('I', 'M'), ('L',), ('Y',), ('I',), ('I',), ('L',), ('M',), ('V',), ('V',), ('I',))	2	15
	(('A',), ('I', 'L'), ('L',), ('Y',), ('I', 'L'), ('I',), ('L',), ('I',), ('V',), ('V',), ('I',))	4	20
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I',), ('I',), ('L',), ('M', 'I'), ('V',), ('V',), ('I',))	6	26
	(('A',), ('M', 'L'), ('L',), ('Y',), ('I',), ('I', 'L'), ('L',), ('M', 'I'), ('V',), ('V',), ('I',))	8	27
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('I',), ('L',), ('M', 'I'), ('V',), ('V',), ('I',))	12	35
	(('A',), ('I', 'L'), ('L',), ('Y',), ('I',), ('V', 'I', 'L'), ('M', 'Y', 'L'), ('I',), ('V',), ('V',), ('I',))	18	36
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('I', 'M'), ('L',), ('M', 'I'), ('V',), ('V',), ('I',))	24	50
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('I', 'M', 'L'), ('L',), ('M', 'I'), ('V',), ('V',), ('I',))	36	57
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('I', 'M'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('I',))	48	60
	(('A',), ('I', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('V', 'I'))	64	65
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('I',))	72	73
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('I',))	96	81
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'Y', 'L'), ('M', 'I'), ('V',), ('V',), ('I',))	108	82
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('V', 'I'))	144	92
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'L'), ('M', 'I'), ('V',), ('V',), ('V', 'I'))	192	100
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'L'), ('M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	216	103
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'L'), ('M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	288	111
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'Y', 'L'), ('M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	324	112
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	360	116
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'L'), ('V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	384	118
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	432	120

	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	480	124
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M'), ('M', 'Y', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	540	125
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	576	127
	(('A',), ('I', 'M', 'L'), ('L',), ('Y',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	720	133
	(('A',), ('I', 'M', 'L'), ('L',), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	864	134
	(('A',), ('I', 'M', 'L'), ('L',), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	960	137
	(('A',), ('I', 'M', 'L'), ('L',), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	1152	141
	(('A',), ('I', 'M', 'L'), ('L',), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	1440	147
	(('A',), ('I', 'M', 'L'), ('L',), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I', 'L'))	2160	150
	(('A',), ('I', 'M', 'L'), ('M', 'L'), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I'))	2880	157
	(('A',), ('I', 'M', 'L'), ('M', 'L'), ('Y', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('M', 'Y', 'L'), ('L', 'V', 'M', 'A', 'I'), ('V',), ('V',), ('V', 'I', 'L'))	4320	160
G	(('I',), ('A',), ('L',), ('M',), ('L',), ('L',), ('I',), ('I',), ('I',), ('V',), ('L',), ('F',), ('V',), ('L',))	1	14
	(('I',), ('A',), ('L',), ('M',), ('L',), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('L',), ('F',), ('V',), ('L',))	2	23
	(('I',), ('A',), ('M',), ('I', 'M', 'L'), ('M',), ('L',), ('M',), ('I',), ('I',), ('V',), ('I',), ('F',), ('V',), ('L',))	3	25
	(('I',), ('A',), ('L',), ('I', 'M'), ('L',), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('L',), ('F',), ('V',), ('L',))	4	34
	(('I',), ('A',), ('M', 'L'), ('I',), ('M',), ('I',), ('I',), ('I',), ('I',), ('V',), ('I',), ('F',), ('V',), ('V', 'M', 'L'))	6	37
	(('I',), ('A',), ('M', 'L'), ('I',), ('L', 'M'), ('I', 'M'), ('I',), ('I',), ('I',), ('V',), ('I',), ('F',), ('V',), ('L',))	8	50
	(('I',), ('A',), ('M', 'L'), ('I',), ('L', 'M'), ('I',), ('I',), ('I',), ('I',), ('V',), ('I',), ('F',), ('V',), ('V', 'M', 'L'))	12	56
	(('I',), ('A',), ('M',), ('I', 'M'), ('M', 'I'), ('I', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I',), ('F',), ('V',), ('L',))	16	67
	(('I',), ('A',), ('M', 'L'), ('I',), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I',), ('F',), ('V',), ('L',))	18	68
	(('I',), ('A',), ('L',), ('I', 'M'), ('L',), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	24	82
	(('I',), ('A',), ('L',), ('I', 'M'), ('L', 'M'), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('M', 'L'))	32	92
	(('I',), ('A',), ('L',), ('I', 'M'), ('L',), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	36	100
	(('I',), ('A',), ('L',), ('I', 'M'), ('L', 'M'), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	48	114

	(('I',), ('A',), ('L',), ('I', 'M'), ('L', 'M'), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	72	142
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'M', 'I'), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('M', 'L'))	96	158
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('M', 'L'))	128	172
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'M', 'I'), ('I', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	144	191
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'M', 'I'), ('I', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('M', 'L'))	192	207
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	216	231
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	288	246
	(('I',), ('A',), ('V', 'M', 'L'), ('I', 'M'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I',), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	324	252
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	384	257
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	432	281
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	576	299
	(('I',), ('A',), ('M', 'L'), ('I', 'M', 'L'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	648	314
	(('I',), ('A',), ('M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	864	332
	(('I',), ('A',), ('V', 'M', 'L'), ('I', 'M', 'L'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	972	343
	(('I',), ('A',), ('V', 'M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	1296	363
	(('I',), ('A',), ('V', 'M', 'L'), ('I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	1620	364
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I',), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	1728	377
	(('I',), ('A',), ('V', 'M', 'L'), ('I', 'M', 'L'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('I', 'M'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	1944	381
	(('I',), ('A',), ('M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	2304	389
	(('I',), ('A',), ('M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	2592	405
	(('I',), ('A',), ('V', 'M', 'L'), ('I', 'M', 'L'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	2916	408
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	3072	410

	(('I',), ('A',), ('M', 'L'), ('I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	3240	415
	(('I',), ('A',), ('I', 'M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	3456	432
	(('I',), ('A',), ('I', 'M', 'L'), ('I', 'M', 'L'), ('L', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	3888	439
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('M', 'L'))	4096	444
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	4608	468
	(('I',), ('A',), ('I', 'M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	5184	474
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	5760	480
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	6144	498
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	6912	511
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	7680	513
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I',), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	8640	524
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	9216	557
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	11520	574
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	13824	605
	(('I',), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	17280	623
	(('I', 'Y'), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	27648	655
	(('I', 'Y'), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	34560	673
	(('I', 'Y', 'L'), ('A',), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('I', 'M'), ('V',), ('I', 'L'), ('F',), ('V',), ('V', 'M', 'L'))	51840	717
H	(('M',), ('I',), ('M',), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('L',))	1	33
	(('M',), ('I',), ('I', 'M'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('L',))	2	47
	(('M',), ('I',), ('V', 'I', 'M'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('V',), ('L',))	3	48
	(('M',), ('I',), ('I', 'M'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	4	58

	(('M',), ('I',), ('V', 'I', 'M'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	6	59
	(('M', 'L'), ('I',), ('I', 'M'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	8	71
	(('M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	12	80
	(('M', 'L'), ('I',), ('I', 'M'), ('M',), ('L',), ('L',), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	16	86
	(('I', 'M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('L',), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	18	90
	(('M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('L',), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	24	95
	(('M', 'L'), ('I',), ('I', 'M'), ('M',), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	32	105
	(('I', 'M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('I',), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	36	111
	(('M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	48	122
	(('M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	64	130
	(('I', 'M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	72	143
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	96	151
	(('I', 'M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	108	159
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	144	167
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	192	176
	(('I', 'M', 'L'), ('I',), ('I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	288	189
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('M',), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	384	200
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	480	201
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('L', 'V', 'M'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	576	210
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	720	217
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	768	225
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'V', 'M'), ('L',), ('I', 'L'), ('V', 'I', 'M'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	864	228
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I',), ('V',), ('V',), ('M', 'L'))	960	232

	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M'), ('L',), ('I', 'L'), ('V', 'I', 'M'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	1152	247
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'M'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	1440	257
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	1536	263
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	1920	273
	(('V', 'I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	2560	274
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	3072	276
	(('I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	3840	288
	(('V', 'I', 'M', 'L'), ('I',), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	5120	290
	(('I', 'M', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	5760	296
	(('I', 'M', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	7680	304
	(('V', 'I', 'M', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('G',), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	10240	306
	(('I', 'M', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('L',), ('I', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	11520	313
	(('I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	15360	318
	(('V', 'I', 'M', 'L'), ('I', 'M'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	20480	320
	(('I', 'M', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	23040	334
	(('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	30720	336
	(('I', 'M', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	34560	338
	(('V', 'I', 'M', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('F', 'L', 'V', 'M', 'I'), ('I', 'L'), ('I', 'M', 'L'), ('V', 'I', 'M', 'L'), ('A', 'G'), ('V',), ('V',), ('I', 'L'), ('V',), ('V',), ('M', 'L'))	46080	340
I	(('I',), ('L',), ('I',), ('V',), ('M',), ('I',), ('L',), ('M',), ('V',), ('T',), ('L',), ('M',))	1	7
	(('I',), ('L',), ('I',), ('V',), ('V', 'M'), ('I',), ('L',), ('M',), ('V',), ('T',), ('L',), ('M',))	2	10
	(('I',), ('L',), ('I',), ('V',), ('V', 'M'), ('I',), ('I', 'L'), ('M',), ('V',), ('T',), ('L',), ('M',))	4	12
	(('I',), ('L',), ('I',), ('V',), ('V', 'M'), ('I',), ('T', 'I', 'L'), ('M',), ('V',), ('T',), ('L',), ('M',))	6	13
	(('I',), ('L',), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'L'), ('M',), ('V',), ('T',), ('L',), ('M',))	8	18

	(('I',), ('L',), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'L'), ('M',), ('V',), ('T',), ('L',), ('M',))	12	19
	(('I',), ('L',), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'L'), ('M',), ('V',), ('T', 'A'), ('L',), ('M',))	24	20
	(('I',), ('I', 'L'), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'L'), ('M',), ('V',), ('T',), ('M', 'L'), ('M',))	32	25
	(('I',), ('I', 'L'), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('M',), ('V',), ('T',), ('M', 'L'), ('M',))	48	27
	(('I',), ('I', 'L'), ('I',), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('M',), ('V',), ('T',), ('M', 'L'), ('M',))	64	28
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'L'), ('M',), ('V',), ('T',), ('I', 'M', 'L'), ('M',))	96	34
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'L'), ('M',), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M',))	128	36
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('M',), ('V',), ('T',), ('I', 'M', 'L'), ('M',))	144	37
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('M',), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M',))	192	39
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('M',), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M',))	256	40
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('I', 'M'), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M',))	384	42
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('I', 'M'), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M',))	512	43
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('I', 'M'), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M', 'L'))	768	44
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('I', 'M'), ('V',), ('T',), ('V', 'I', 'M', 'L'), ('M', 'L'))	1024	45
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('I', 'M'), ('V',), ('T', 'A'), ('V', 'I', 'M', 'L'), ('M', 'L'))	1536	46
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('I', 'M'), ('V',), ('T', 'A'), ('V', 'I', 'M', 'L'), ('M', 'L'))	2048	47
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('I', 'V', 'L'), ('I', 'M'), ('V', 'A'), ('T', 'A'), ('V', 'I', 'M', 'L'), ('M', 'L'))	3072	48
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('I', 'M'), ('V', 'A'), ('T', 'A'), ('V', 'I', 'M', 'L'), ('M', 'L'))	4096	49
	(('I',), ('I', 'L'), ('I', 'L'), ('V',), ('V', 'M'), ('I', 'L'), ('T', 'I', 'V', 'L'), ('I', 'M'), ('V', 'A'), ('T', 'A'), ('V', 'I', 'M', 'L'), ('I', 'M', 'L'))	6144	50
J	(('M',), ('T',), ('A',), ('I',), ('V',), ('A',), ('I',), ('T',), ('M',), ('L',), ('V',), ('T',))	1	12
	(('M',), ('T',), ('A',), ('I',), ('V',), ('A',), ('I',), ('T',), ('M',), ('I', 'L'), ('V',), ('T',))	2	17
	(('M',), ('T',), ('A',), ('I',), ('V',), ('A',), ('I', 'L'), ('T',), ('M', 'L'), ('I', 'L'), ('V',), ('T',))	8	20