Using a collection of nonfunctional missense mutants to examine how molecular chaperones and proteases maintain protein homeostasis

By

Fatmah Mirza Hani

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Molecular Bioscience

Middle Tennessee State University

December 2019

Thesis Committee:

Dr. Elliot Altman, Chair

Dr. Anthony L. Farone

Dr. Mary Farone

Dr. Anthony L. Newsome

Dr. Preston MacDougall

I dedicate this research to my father whose soul guided me through my journey, my mother, who always believed in me, to my loving husband, who was beside me all the time and to my lovely daughter, for whom I wish a bright future, and to my family and friends who have always supported me.

ACKNOWLEDGEMNTS

I would like to thank the Molecular Biosciences (MOBI) doctoral program and the Biology Department at Middle Tennessee State University (MTSU) for providing the opportunity for this research study, the Ministry of Higher Education in Saudi Arabia (MOHE) for providing me a full scholarship and my advisor, Elliot Altman, who was my model in dedication, hard work, and ethics. His guidance and continued support could never be thanked enough, I would also like to thank my committee members for their continued help and support.

ABSTRACT

Protein homeostasis is crucial for optimal cellular life and affects thousands of proteins that have to reach their final folded active state to be functional. Protein folding is not trivial in the crowded cellular environment and many proteins cannot fold correctly. These incorrectly folded proteins either have to be rescued by chaperones or unfoldases or have to be destroyed by proteases. This triage by chaperones and proteases is key to the cellular maintenance of proteins. We have studied the interaction between chaperones and proteases using a collection of 42 well-characterized non-functional missense mutants in the bacterial cytosolic enzymes, β-galactosidase and catechol 2,3-dioxygenase. In an initial analysis, we showed that the collection of missense mutants represented proteins that could not fold correctly and that the great majority of them were rescuable by high salt concentrations that are known to stabilize proteins. The two major chaperone complexes, Hsp60 (GroL/GroS) and Hsp70 /Hsp40 (DnaK/DnaJ), were either overproduced in the missense mutants or the two major proteases, Lon and Clp, were deleted. The deletion of the proteases was far more effective at stabilizing the missense mutants than overproducing the chaperones. A total of 16.67% of the missense mutants were rescued by the overproduction of GroL/S, 47.62% of the missense mutants were rescued by the overproduction of DnaK/DnaJ, 78.57% of the missense mutants were rescued by the deletion of Lon and 83.33% of the missense mutants were rescued by the deletion of Clp. Additional analysis of our results supported a number of models that have been proposed concerning protein homeostasis and how chaperone and proteases interact with misfolded proteins.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I. INTRODUCTION

1.1 Protein Homeostasis

Proteins are cornerstone macromolecules in living cells and they are required for most biological functions. Because of their importance to cellular integrity, cells have developed a quality control system to maintain a delicate balance of protein folding, localization, and degradation. This balanced state is often referred to as protein homeostasis or proteostasis (Fernández-Fernández & Valpuesta, 2018; Morimoto, Selkoe, & Kelly, 2012). This sophisticated evolved control system involves two main components: molecular chaperones and proteases (Morimoto et al., 2012). One of the many functions of molecular chaperones is to assist the folding of nascent and unfolded proteins, whereas proteases degrade those that fail to reach the native final folded state. Not all proteins are able to achieve final correctly folded structure; some will misfold in a crowded cellular environment, and as a result, will form protein aggregates. Figure 1.1 summaries the interplay of the quality control system of protein folding. Failure to fold correctly or maintain a native protein state disrupts protein homeostasis and is known to cause several types of disorders including Alzheimer's and Parkinson's diseases (Morimoto et al., 2012). Understanding this complex network of molecular chaperones and proteases in maintaining protein homeostasis could contribute to the overall knowledge of controlling and targeting diseases.

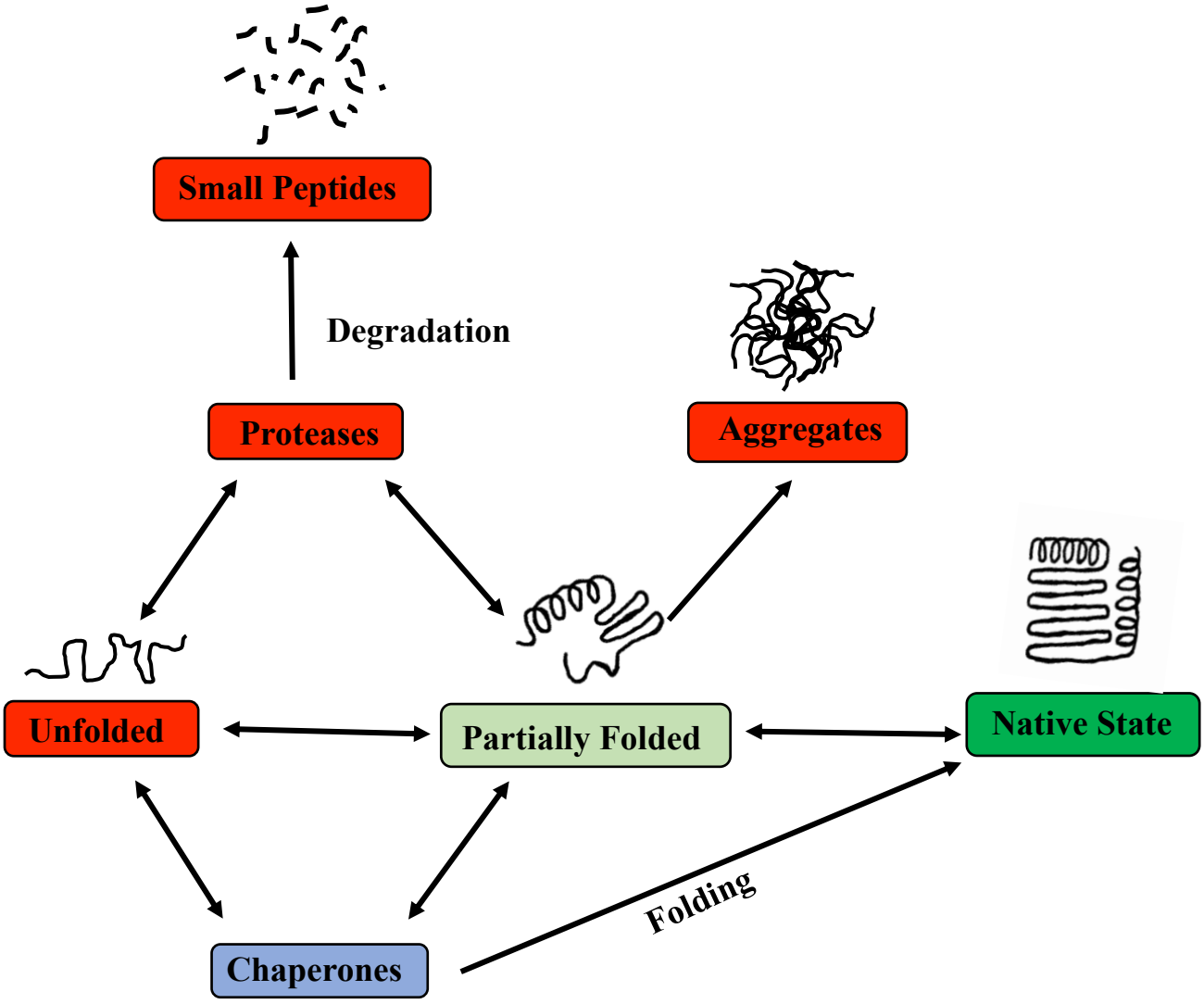

Figure 1.1: Protein homeostasis network. Unfolded and partially folded protein states need the assistance of chaperones to reach their native state. When failing to reach the native state, degradation will take place by proteases. Misfolded proteins will result in protein aggregates which can lead to several cellular and genetic disorders.

1.2 Molecular Chaperones

One of the control systems the cell has evolved is the molecular chaperone system. This system is a large diverse family of proteins with several roles. They assist nascent proteins with folding into the three-dimensional structure, they prevent protein aggregation, and they can also mediate targeted protein unfolding. These chaperones are distributed throughout different locations in the cell and act on a broad range of proteins (Finka, Mattoo, & Goloubinoff, 2016; Saibil, 2013a). They recognize and bind unfolded protein from the exposed hydrophobic amino acids (Fernández-Fernández & Valpuesta, 2018). These chaperones catalyze protein self-assembly. Without the presence of chaperones, proteins would take a long time folding into the native state, or they would be unstable and eventually form aggregates in the crowded cellular environment. A number of chaperones belong to the large group of heat shock proteins (HSP) (Marchenkov & Semisotnov, 2009a). This family is classified according to their molecular weight (HSP60, HSP70, HSP90, HSP100, and small HSPs). Among those, two cytosolic chaperones that are the most well-studied, and for which detailed structure and mechanism have been addressed are the Hsp60 (known as GroLS in bacteria) and Hsp70 (known as DnaK in bacteria) chaperones (Hartl, Bracher, & Hayer-Hartl, 2011).

1.2.1 Hsp60 (GroLS) Chaperone

GroLS is an ATP-dependent chaperone which assists protein folding. GroLS is the only chaperone that is necessary for *E. coli* viability (Morimoto et al., 2012). It binds and assists around 30% of nascent proteins and 40% of denatured proteins

(Marchenkov & Semisotnov, 2009a). It consists of two proteins, GroL and GroS. The GroL component is composed of fourteen identical GroL subunits, each of which is 57kDa. These subunits are arranged as two stacked, seven-member rings with a large solvent filled cavity. The substrate and nucleotide-binding sites are located within this cavity. Unfolded proteins will be encapsulated into the chamber shape complex, which provides a protected environment for assisting protein folding (Saibil, 2013b). Each subunit of GroL has three distinct domains: apical, intermediate, and equatorial. Unfolded proteins will bind to the apical domain, while ATP-binding sites are located on the equatorial domain. The intermediate domain connects the apical with the equatorial (Chaudhuri, Verma, & Maheshwari, 2009). To reach its maximum effectiveness, GroL needs the assistance of a GroS co-chaperone. The GroS component is a ring of seven identical 10 kDa subunits which act as a lid for the GroL chamber. The GroS co-chaperones will capture the hydrophobic exposed polypeptides of the unfolded protein and will encapsulate it into the GroL cylinder. A conformational change is created upon binding of the GroS ring to the GroL chamber, which initiates folding. The mechanism of action is ATP regulated. Upon ATP and GroS binding at the top cis ring, a conformational change occurs, and the binding substrate is encapsulated into the hydrophilic, net negatively-charged cage enabling folding during which ATP is hydrolyzed at the cis ring. ATP then binds in the opposite ring (trans ring), which causes ADP and GroS to dissociate and the folded protein leaves the GroL cage. The same cycle will be repeated when ATP binds to the trans ring, which will be followed by GroS to binding, substrate encapsulation, folding and release (Gupta, Haldar, Miličić, Hartl, & Hayer-Hartl, 2014).

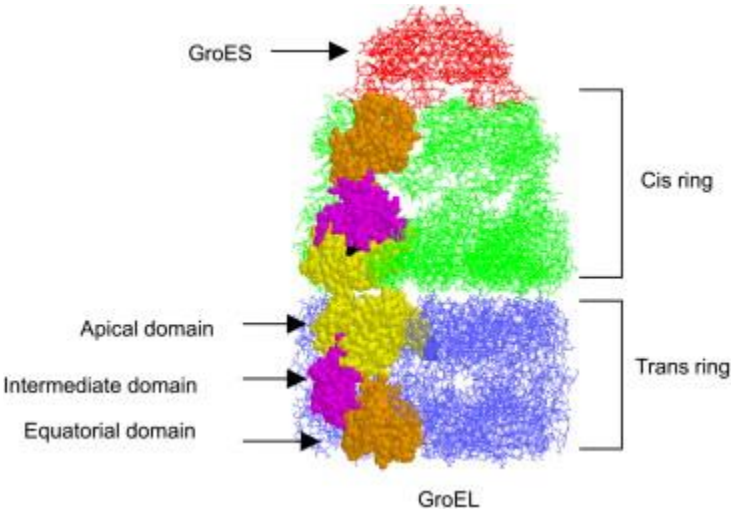

Figure 1.2: Hsp60 (GroLS) chaperone complex. A single subunit in each GroL ring is composed of three domains: apical (brown), equatorial (yellow), and intermediate (purple). ATP binding sites are located on the equatorial domain while the intermediate domain connects the apical with the equatorial. When GroS (red) binds to the GroL-ATP complex, the apical domain moves upward by 60_o and rotates about its axis by 120_o which doubles the central cavity. The cis (green) and trans (blue) rings are distinguished on the basis of size (Chaudhuri et al., 2009).

1.2.2 Hsp70 (DnaK) Chaperone

The HSP70 chaperone is to one of the most conserved proteins amongst all of the biological kingdoms (Finka et al., 2016; Saibil, 2013a). It is the most abundant chaperone in cellular compartments, and it plays a role in stabilizing the unfolded state of proteins in order to reach a natural folding state. Unlike HSP60 which forms a cage inside which unfolded proteins are encapsulated, HSP70 binds to unfolded polypeptides in free solution. If the unfolded protein fails to reach its native state, rebinding to DnaK, or binding to other chaperones can occur. This 70kDa-chaperone

is composed of two domains; a nucleotide-binding domain (NBD) at one end and a substrate-binding domain (SBD) at the other end (Saibil, 2013). The SBD domain will clamp around the target protein, and the NBD domain will use ATP to regulate the process of folding. Both of these domains are connected by a flexible linker (Kityk, Kopp, Sinning, & Mayer, 2012). The SBD is made up of two parts: the base or SBDα and the lid or SBDß (Fernández-Fernández & Valpuesta, 2018). In order to stimulate the activity rate of Hsp70 (DnaK), a 40-kDa co-chaperone, known as (DnaJ), or the J-domain protein, is required. This stimulation of activity is carried out by targeting substrates to DnaK and by enhancing the ATPase activity as well. The DnaKJ chaperone complex exhibits conformational flexibility upon binding/releasing of the substrates, which depends on binding and releasing of an ATP molecule. ATP binding to the NBD domain will induce a conformational change in the SBD to adopt an open state which is achieved by separating the base and the lid subparts (Fernández-Fernández & Valpuesta, 2018). The open state is the low substrate affinity state. When ATP is hydrolyzed, the ADP-state will increase the chaperone affinity for substrates, then the lid and base will come together to close over the bound substrate. Thus, DnaKJ assists protein folding by undergoing an opening and closing reaction of the substrate-binding domain, which is affected by the nucleotide-binding state. In addition, the ATP/ADP exchange is achieved by a 22 kDa nucleotide exchange factor known as GrpE (Bukau & Horwich, 1998). Figure 1.3 shows the HSP70 structure and illustrates the difference between ADP and ATP bound states.

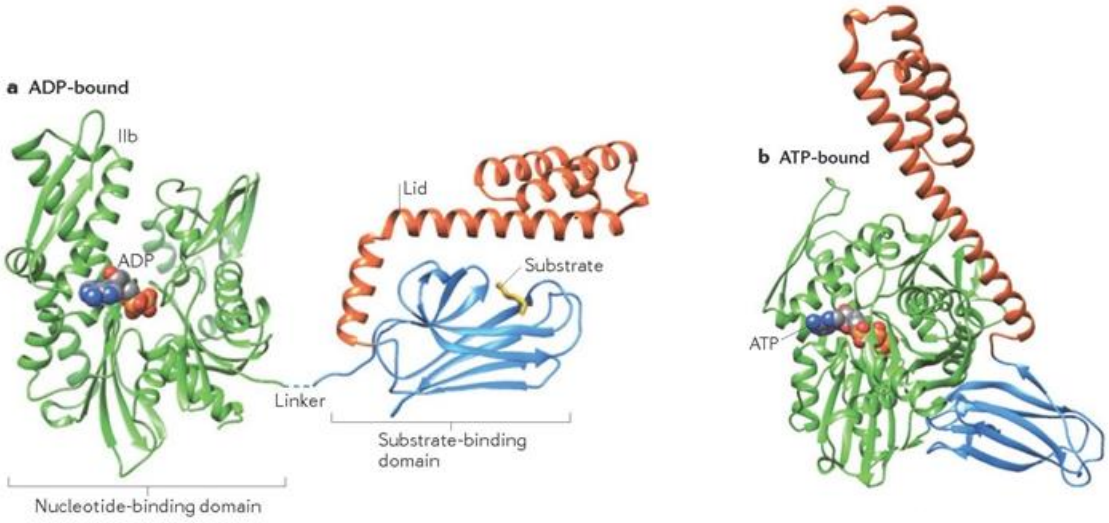

Figure 1.3: Hsp70 (DnaK) chaperone complex. A. ADP-bound state; the closed form of the chaperone. The nucleotide-binding domain (green; Protein Data Bank of HSP70 (PDB) code: 3HSC). A flexible linker connects NBD to the substrate-binding domain (blue; PDB code: 1DKZ). Inside a pocket-like space, a lid-like domain (red) will close and lock a substrate (yellow) into the SBD. **B. The ATP-bound state; the opened form of the chaperone.** A conformational change is triggered upon ATP binding. The NBD cleft closes, and the lid opens, enabling the SBD to widely open and bind to another polypeptide substrate (Saibil, 2013b).

1.3 ATP-dependent Proteases

Proteolysis is a significant quality control mechanism found in living cells. It controls both key regulatory enzymes as well as short-term proteins. This mechanism is carried out by proteases that depend on ATP to function (Porankiewicz, Wang, & Clarke, 1999a). ATP-dependent proteases belong to the AAA+ superfamily (ATPases Associated with a variety of cellular Activities). Proteases degrade proteins by the hydrolysis of peptide

bonds (Bittner, Arends, & Narberhaus, 2016). They degrade not only the damaged or denatured polypeptides, but also transitory regulatory proteins (Maillard et al., 2011a). The AAA+ family has several other roles in addition to proteolysis; they have roles in DNA replication, transcription, and membrane fusion (Lee, Berdis, & Suzuki, 2006). In bacteria, there are five types of ATP-dependent proteases, ClpXP, ClpAP, HslUV, Lon, and FtsH. The proteases of the AAA+ family have a common feature of having the proteolytic domain inside a chamber with a narrow passage. This feature of the multi-subunit protease will ensure the protection of cytosolic proteins from being degraded by AAA+ proteases. An additional controlling mechanism is that proteins must bear a degradation tag in order to be recognized by proteases for degradation to ensure tight regulation of the process (Gur & Sauer, 2008; Gur & Sauer, 2009; Maillard et al., 2011b).

1.3.1 ClpXP Protease

ClpP is a cytosolic serine protease that is conserved amongst bacteria as well as plants and mammalian animals. It is well characterized, and its structure and mechanism of action has been elucidated. The size of the ClpP protease alone is 21.5 kDa and when combined with ClpX is 46 kDa. The ClpP is the proteolytic subunit and is associated with ATPase subunit ClpX that belongs to the molecular chaperone group. ClpP protease contains the catalytic triad residues Ser-His-Asp that are involved in breaking down peptide bonds (Tripathi & Sowdhamini, 2008a). ClpX contains a recognition

domain and has only one ATP binding domain (Porankiewicz et al., 1999a). It is responsible for binding tagged substrates, unfolding the three-dimensional structure and translocating it into the ClpP subunit for degradation (Baker & Sauer, 2012; Flynn, Neher, Kim, Sauer, & Baker, 2003a). When not combined with ClpX, ClpP can degrade polypeptides but with minimal activity. However, when ClpX is by itself, it acts as a chaperone, which depends on ATP to disassemble proteins (Baker & Sauer, 2012). The activity of ClpXP is increased by the induction of stress, including heat shock, salt or oxidative stress. ATP is needed to provide energy for the required conformational changes; however, ATP will not be hydrolyzed; binding ATP alone is required to carry out the process (Baker & Sauer, 2012). Figure 1.4 shows the two Clp domains and substrate recognition followed by degradation.

1.3.2 ClpXP Protease Structure

ClpXP is composed of two seven-membered rings of ClpP. This 14-subunit complex exhibit a barrel-like structure, and one or two hexameric ClpX binds to it to form the full ClpXP protease (Flynn et al., 2003a; Porankiewicz et al., 1999a). The active sites are hydrophobic and are located inside the chamber. This character of hydrophobicity facilitates specific diffusion and degradation of protein substrates (Maillard et al., 2011a). The protease crystal structure has been resolved, as shown in Figure 1.5.

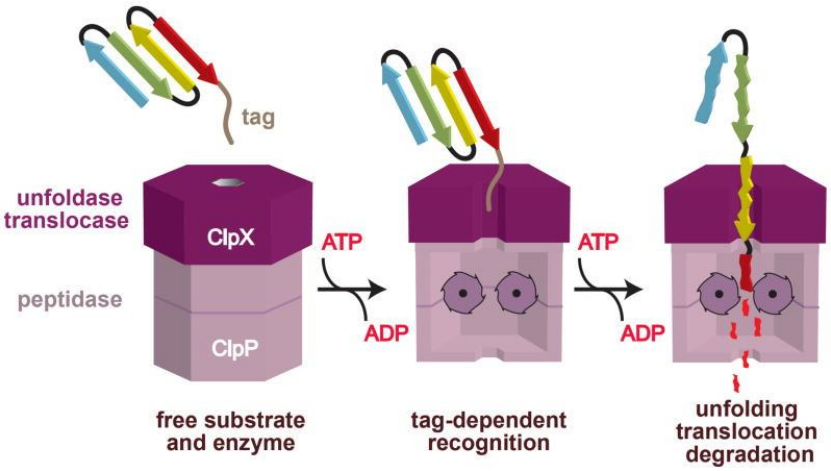

Figure 1.4: Cartoon model of substrate recognition and degradation by the ClpXP protease. From: RT Sauer et al., (2004) Sculpting the proteome with AAA+ proteases and disassembly machines (Sauer et al., 2004).

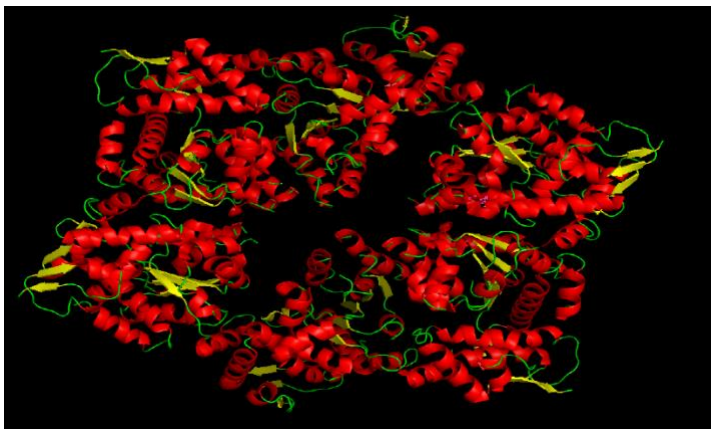

Figure 1.5: ClpX Hexamer. (PDB code: 4I34) generated by PyMOL with α-helical regions indicated in red, β-sheet regions indicated in yellow, and connecting loops indicated in green (Cole, Ashley E. et al., 2017)

1.3.3 Lon Protease

The ATP-dependent Lon protease, which is also known as La, is also a cytosolic serine protease. The name originated from the phenotypic characterization of *lon* mutants in *E. coli* that resemble long undivided filaments when viewed under UV light. The Lon protease maintains protein homeostasis by degrading misfolded and some short-lived regulatory proteins (Chin, Gof, Webster, Smith, & Goldberg, 1988; Lee & Suzuki, 2008a). For each cleavage of a peptide bond in a protein, Lon hydrolyzes two ATP molecules (Chin et al., 1988). The active proteolytic site is located inside the chamber, and a hexameric ring will unfold and translocate the tagged protein substrates into the chamber. Figure 1.6 shows the recognition and degradation sites of Lon protease.

1.3.4 Lon Protease Structure

The *E. coli* Lon protease was the first protease that was characterized. Lon consists of a tetramer of identical subunits that assembles into a ring-shaped complex. Unlike Clp protease that contains the ATPase site and the proteolytic site as separate subunits, each of the Lon subunits carry both an ATP binding site and an active catalytic site (Chin et al., 1988; Lee & Suzuki, 2008a; Tripathi & Sowdhamini, 2008a). Lon proteases involve a Ser-Lys catalytic dyad (Tripathi & Sowdhamini, 2008a). The crystal structure of the

Lon protease has not been identified yet and only the proteolytic fragment has been determined, as shown in Figure 1.7.

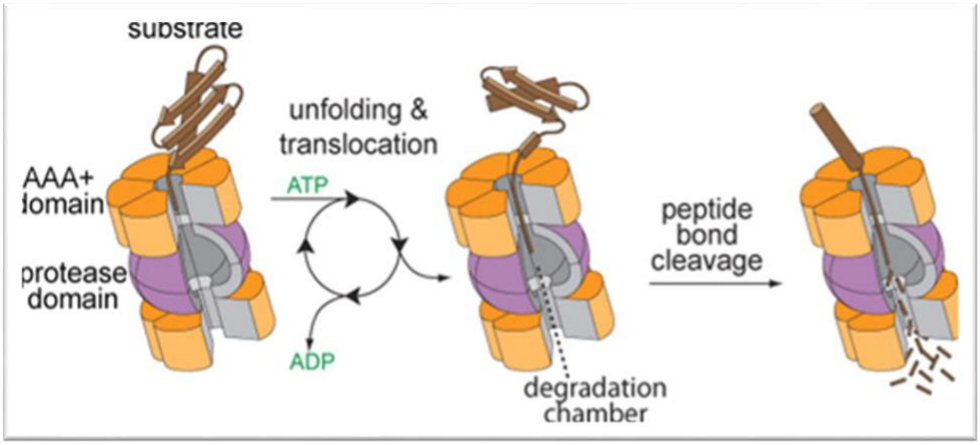

Figure 1.6: Lon protease. ATP-dependent protease degrades substrates by a complex mechanism (Gur, Biran, & Ron, 2011).

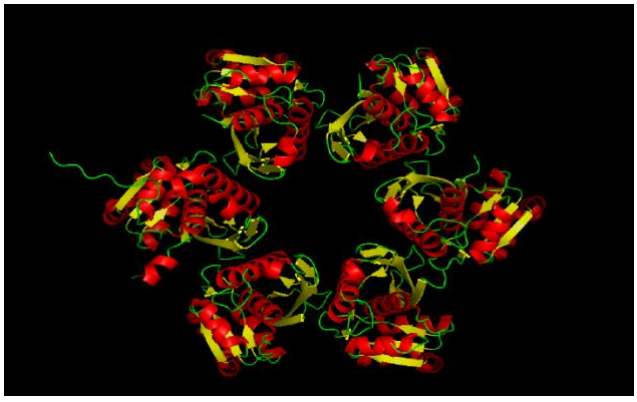

Figure 1.7: Oxidized form of *E. coli* Lon Proteolytic domain. (PDB code: 3WU4) generated by PyMol with α-helical regions indicated in red, β-sheet regions indicated in yellow and connecting loops indicated in green (Cole, Ashley E. et al., 2017)

1.4 Premise for the Study

We generated and characterized a collection of 42 non-functional missense mutants in the cytosolic bacterial β-galactosidase and catechol 2,3-dioxygenase enzymes. In an initial study, we showed that most of these missense mutants were rescuable by high salt, which is known to rescue missense mutants, and thus the collection of missense mutants represented a tool by which the effectiveness of chaperones and proteases could be evaluated.

1.5 The Studied Proteins: β-galactosidase and catechol 2,3-dioxygenase enzymes

The model enzymes for the current study are the two cytosolic bacterial β-galactosidase and catechol 2,3-dioxygenase enzymes. They are a quintessential choice because there is well-established knowledge in the literature regarding their structure, function, and sequence. Moreover, their enzymatic activity can be easily measured by performing quantitative assays.

The *lac* operon in *Escherichia coli* is responsible for the transport and metabolism of the disaccharide lactose. The *lacZ* gene is one component of these clusters of genes, and it codes for the β-galactosidase enzyme. β-galactosidase will hydrolyze the disaccharide lactose into two glucose and galactose monosaccharides. β-galactosidase will also convert lactose to allolactose as a regulation mechanism of the enzyme where then allolactose will bind to the *lacZ* repressor and cease β-galactosidase production. The presence of an active β-galactosidase is easily detected, an advantage that made it the enzyme of choice for many scientific studies. When

available, β-galactosidase hydrolyzes a colorless X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside) to yield galactose and an insoluble blue colored indole compound.

The β-galactosidase structure consists of four polypeptide chains, each subunit with 1,023 amino acid residues (Juers, Matthews, & Huber, 2012). There are five structural domains within each monomer. The active site is located at the C-terminal of a barrel-like central domain, which is also called the a8b8 barrel or triosephosphate isomerase (TIM). The three-dimensional structure generated by PyMOL is shown in Figure 1.8.

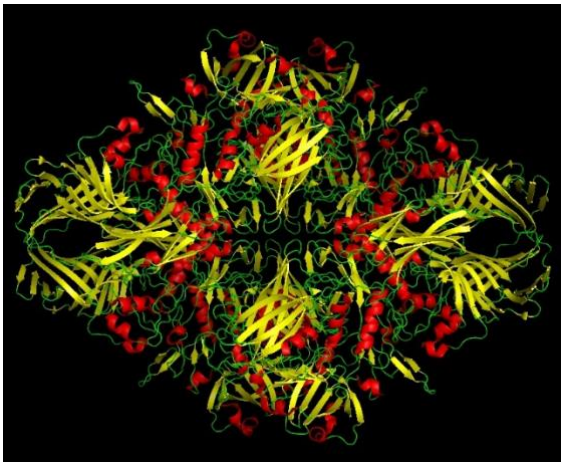

Figure 1.8: β-galactosidase structure. The β-galactosidase structure at 1.7Å (PDB code: 1DP0). Images generated by PyMOL and α-helices are red, β-sheets are yellow, and coils are green (Cole, Ashley Elliott, 2017).

Catechol 2,3-dioxygenase, a product of the *xylE* gene, is produced by the soil bacterium *Pseudomonas puitida*. It plays a role in the cleavage of aromatic compounds found in the environment. Catechol 2,3-dioxygenase specifically catalyzes the extradiol ring cleavage of catechol to form a semialdehyde hydroxymuconic. The degradation is accomplished by incorporating two oxygen molecules into the catechol compound.

Like β-galactosidase, Catechol 2,3-dioxygenase is also a tetramer of identical subunits, but it is a smaller protein. Each subunit is composed of 307 amino acid residues (Kita et al., 1999). The active domain of catechol is found at the C-terminal site in a funnel-shaped space. In this active site, an iron atom is present and is bonded by three protein residues His153, His214, and Glu265 (Ishida, Kita, Miki, Nozaki, & Horiike, 2002). The three-dimensional structure generated by PyMOL is shown in Figure 1.9.

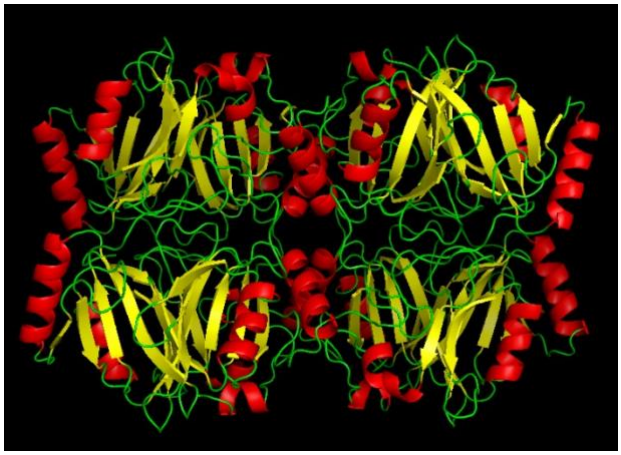

Figure 1.9: Catechol 2,3-dioxygenase structure. Catechol 2,3-dioxygenase at 2.0Å (PDB code:1MPY). Images generated by PyMOL and α-helices are red, β-sheets are yellow, and coils are green (Cole, Ashley Elliott, 2017).

1.6 Dissertation synopsis

The main objective of this research study was to examine the triad between molecular chaperones and proteases in maintaining protein homeostasis using a collection of non-functional missense mutants in β-galactosidase and catechol 2,3-dioxygenase. Initially, we verified the suitability of the missense mutants by determining whether they could be rescued by high levels of salt. Once the suitability of the missense collection was determined, we examined the effect of the two critical components, chaperones and proteases that are involved in protein homeostasis to affect the activity of the missense mutants.

REFERENCES

Baker, T. A., & Sauer, R. T. (2012a). ClpXP, an ATP-powered unfolding and protein-degradation machine. *BBA - Molecular Cell Research, 1823*(1), 15-28. doi:10.1016/j.bbamcr.2011.06.007

Baker, T. A., & Sauer, R. T. (2012b). ClpXP, an ATP-powered unfolding and protein-degradation machine. *BBA - Molecular Cell Research, 1823*(1), 15-28. doi:10.1016/j.bbamcr.2011.06.007

Bittner, L., Arends, J., & Narberhaus, F. (2016). Mini review: ATP-dependent proteases in bacteria. *Biopolymers, 105*(8), 505-517.

Bukau, B., & Horwich, A. L. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell, 92*(3), 351-366. doi:10.1016/s0092-8674(00)80928-9

Chaudhuri, T. K., Verma, V. K., & Maheshwari, A. (2009). GroEL assisted folding of large polypeptide substrates in *Escherichia coli*: Present scenario and assignments for the future. *Progress in Biophysics and Molecular Biology, 99*(1), 42-50. doi:10.1016/j.pbiomolbio.2008.10.007

Chin, D. T., Gof, S. A., Webster, T., Smith, T., & Goldberg, A. L. (1988). Sequence of the lon gene in *Escherichia coli*. A heat-shock gene which encodes the ATP-dependent protease la. *Journal of Biological Chemistry, 263*(24), 11718-11728.

Cole, A. E. (2017). *Using a collection of nonfunctional missense mutants in the β-galactosidase and catechol 2,3-dioxygenase enzymes to better understand the complexity of protein folding* Available from Dissertations & Theses Europe Full Text: Science & Technology.

David, A., & Sternberg, M. J. E. (2015). The contribution of missense mutations in core and rim residues of protein–protein interfaces to human disease. *Journal of Molecular Biology, 427*(17), 2886-2898. doi:10.1016/j.jmb.2015.07.004

Fernández-Fernández, M. R., & Valpuesta, J. M. (2018). Hsp70 chaperone: A master player in protein homeostasis [version 1; referees: 3 approved]. *F1000Research, 7*, 1497. doi:10.12688/f1000research.15528.1

Finka, A., Mattoo, R. U. H., & Goloubinoff, P. (2016). Experimental milestones in the discovery of molecular chaperones as polypeptide unfolding enzymes. *Annual Review of Biochemistry, 85*(1), 715-742. doi:10.1146/annurev-biochem-060815-014124

Flynn, J. M., Neher, S. B., Kim, Y., Sauer, R. T., & Baker, T. A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Molecular Cell, 11*(3), 671-683. doi:10.1016/S1097-2765(03)00060-1

Gupta, A. J., Haldar, S., Miličić, G., Hartl, F. U., & Hayer-Hartl, M. (2014). Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *Journal of Molecular Biology, 426*(15), 2739-2754. doi:10.1016/j.jmb.2014.04.018

Gur, E., Biran, D., & Ron, E. Z. (2011). Regulated proteolysis in gram-negative bacteria — how and when? *Nature Reviews. Microbiology, 9*(12), 839-848. doi:10.1038/nrmicro2669

Gur, E., & Sauer, R. T. (2008). Recognition of misfolded proteins by lon, a AAA+ protease. *Genes & Development, 22*(16), 2267-2277. doi:10.1101/gad.1670908

Gur, E., & Sauer, R. T. (2009). Degrons in protein substrates program the speed and operating efficiency of the AAA+ lon proteolytic machine. *Proceedings of the National Academy of Sciences of the United States of America, 106*(44), 18503-18508. doi:10.1073/pnas.0910392106

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature, 475*(7356), 324-332. doi:10.1038/nature10317

Ishida, T., Kita, A., Miki, K., Nozaki, M., & Horiike, K. (2002). Structure and reaction mechanism of catechol 2,3-dioxygenase (metapyrocatechase). *International Congress Series, 1233*, 213-220. doi:10.1016/S0531-5131(02)00149-8

Kita, S., Kita, A., Fujisawa, I., Inaka, K., Ishida, T., Horiike, K. . . . Miki, K. (1999). An archetypical extradiol-cleaving catecholic dioxygenase: The crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from pseudomonas putida mt-2. *Structure, 7*(1), 25-34. doi:10.1016/S0969-2126(99)80006-9

Lee, I., Berdis, A. J., & Suzuki, C. K. (2006). Recent developments in the mechanistic enzymology of the ATP-dependent lon protease from escherichia coli: Highlights from kinetic studies. *Molecular BioSystems, 2*(1), 477-483. doi:10.1039/b609936j

Lee, I., & Suzuki, C. K. (2008). Functional mechanics of the ATP-dependent lon protease- lessons from endogenous protein and synthetic peptide substrates. *BBA - Proteins and Proteomics, 1784*(5), 727-735. doi:10.1016/j.bbapap.2008.02.010

Lin, Z., Rye, H. S., & Madan, D. (2008). GroEL stimulates protein folding through forced unfolding. *Nature Structural & Molecular Biology, 15*(3), 303-311. doi:10.1038/nsmb.1394

Maillard, R., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C. M. . . . Bustamante, C. (2011). ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell, 145*(3), 459-469. doi:10.1016/j.cell.2011.04.010

Marchenkov, V. V., & Semisotnov, G. V. (2009). GroEL-assisted protein folding: Does it occur within the chaperonin inner cavity? *International Journal of Molecular Sciences, 10*(5), 2066-2083. doi:10.3390/ijms10052066

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P., & Tavtigian, S. V. (2006). Computational approaches for predicting the biological effect of p53 missense mutations: A comparison of three sequence analysis based methods. *Nucleic Acids Research, 34*(5), 1317-1325. doi:10.1093/nar/gkj518

Morimoto, R. I., Selkoe, D. J., & Kelly, J. W. (2012). *Protein homeostasis*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Porankiewicz, J., Wang, J., & Clarke, A. K. (1999). New insights into the ATP-dependent clp protease: Escherichia coli and beyond. *Molecular Microbiology, 32*(3), 449-458. doi:10.1046/j.1365-2958.1999.01357.x

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews. Molecular Cell Biology, 14*(10), 630-642. doi:10.1038/nrm3658

Sauer, R. T., Bolon, D. N., Burton, R. E., Burton, B. M., Flynn, J. M., Grant, R. A., . . . Baker, T. A. (2004). *Sculpting the proteome with AAA+ proteases and disassembly machines*. United States: Elsevier Inc. doi:10.1016/j.cell.2004.09.020

Smith, C. A. (1996). *Cell Biology*. England.

Tripathi, L. P., & Sowdhamini, R. (2008). Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genomics, 9*(1), 549. doi:10.1186/1471-2164-9-549

CHAPTER II: THE PROMISCUOUS SUMA MISSENSE SUPPRESSOR FROM SALMONELLA ENTERICA HAS AN INTRIGUING MECHANISM OF ACTION

Ashley E. Cole*, Fatmah M. Hani*, Ronni Altman*, Megan Meservy, John R. Roth† and Elliot Altman*‡

* Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132

† Department of Microbiology & Molecular Genetics, University of California, Davis, Davis, CA, 95616

Fatmah Hani sequenced approximately half of the missense mutants characterized in this study, conducted most of the β-galactosidase enzyme assays and assisted with the writing of the published manuscript.

Published February 2017 in Genetics:

Cole, A. E., Hani, F. M., Altman, R., Meservy, M., Roth, J. R., & Altman, E. (2017). The promiscuous sumA missense suppressor from salmonella enterica has an intriguing mechanism of action. *Genetics, 205*(2), 577-588.

2.1 Abstract

While most missense suppressors have very narrow specificities and only suppress the allele against which they were isolated, the *sumA* missense suppressor from *Salmonella enterica* serovar Typhimurium is a broad acting missense suppressor that suppresses numerous missense mutants. The *sumA* missense suppressor was identified as a *glyV* tRNA Gly3(GAU/C) missense suppressor that can recognize GAU or GAC aspartic acid codons and insert a glycine amino acid instead of aspartic acid. In addition to rescuing missense mutants caused by glycine to aspartic acid changes as expected, *sumA* could also rescue a number of other missense mutants as well by changing a neighboring (contacting) aspartic acid to glycine which compensated for the other amino acid change. Thus, the ability of *sumA* to rescue numerous missense mutants was due in part to the large number of glycine codons in genes that can be mutated to an aspartic acid codon and in part to the general tolerability and/or preference for glycine amino acids in proteins. Because the *glyV* tRNA Gly3(GAU/C) missense suppressor has also been extensively characterized in *Escherichia coli* as the *mutA* mutator, we demonstrated that all gain of function mutants isolated in a *glyV* tRNA Gly3(GAU/C) missense suppressor are transferable to a wild-type background and thus the increased mutation rates which occur in *glyV* tRNA Gly3(GAU/C) missense suppressors are not due to the suppression of these mutants.

2.2 Introduction

The *sumA* (suppressor of missense) mutation has been used by researchers to identify missense mutants in *Salmonella enterica* serovar Typhimurium based on its ability to restore the enzymatic function of inactive proteins (Galitski & Roth, 1996; Hughes, Roth, & Olivera, 1991). *sumA* or *su537*, like most missense suppressors, was originally isolated as an extragenic suppressor of the *hisC537* missense mutant (Whitfield, Martin, & Ames, 1966). *sumA* suppressed 3 out of 21, or 14.29%, of the *hisC* missense mutants that were characterized in this study but could not suppress *hisC* amber, ochre or frameshift mutants. In a study involving a greater number of missense mutants, *sumA* was shown to suppress 11 out of 72, or 15.28% of *nadC* missense mutants (Hughes et al., 1991). *sumA* has been mapped near the *purA* locus in *S. enterica* (Sanderson & Hartman, 1978).

Missense suppressors are mutated tRNAs that recognize an aberrant codon instead of the usual codon the tRNA is supposed to recognize (for general reviews see (Hill, 1975; Murgola, E., 1985; Murgola, E. J., 1995). For example, the *glyT* tRNA Gly2(AGA) missense suppressor in *Escherichia coli*, the first missense suppressor that was thoroughly characterized, contains a C to U mutation at the 3' end of the wild-type *glyT* tRNA Gly2(GGA/G) anticodon and reads AGA codons instead of GGA codons thus inserting a glycine amino acid instead of arginine which is normally coded for by AGA codons (Brody & Yanofsky, 1963; Roberts & Carbon, 1975). Missense suppressors can be grouped into four classes based on the change that occurs in the tRNA; those that contain a nucleotide substitution in the anticodon, those that contain a nucleotide insertion in the anticodon loop outside of the

anticodon, those that contain a mutation in the amino acid acceptor stem and those that contain a nucleotide substitution in the base-paired region of the D arm. The vast majority of missense suppressors that have been characterized contain a nucleotide substitution in the anticodon of the tRNA.

In general, both the efficiency of missense suppressors, or their ability to restore the functionality of defective missense mutants, and the specificity of missense suppressors, or their ability to rescue different missense mutants, have been found to be quite low, because presumably a highly efficient missense suppressor with broad specificity would be incredibly detrimental to the cell and most likely lethal. In most of the efficiency studies that have been conducted, the efficiency of missense suppressors ranged from 1.10 – 3.60% as measured by the ability of a missense suppressor to restore the enzymatic activity of a mutant inactive gene compared to the enzymatic activity of the wild-type gene (Berger & Yanofsky, 1967; Brody & Yanofsky, 1963; Hill, Combriato, & Dolph, 1974). In the specificity studies that have been conducted, most of the missense suppressors have been found to be allele specific and only suppress one or two alleles (Brody & Yanofsky, 1963; Eggertsson & Adelberg, 1965; Eggertsson, 1968). There are two notable exceptions. The *glyU* tRNA Gly1(GAG) missense suppressor rescued 5 out of the 12, or 35.71%, of the mutants that were tested (Eggertsson & Adelberg, 1965; Hill et al., 1974) and the *sumA* missense suppressor as discussed above, rescued 15.05% of the mutants that were tested.

In this study, we have characterized the *sumA* missense suppressor. The mutation that causes *sumA* was mapped and sequenced and the efficiency and

specificity of the *sumA* missense suppressor was determined using a collection of missense mutants in either the *lacZ* gene from *E. coli*, which codes for the β-galactosidase enzyme, or the *xylE* gene from the *Pseudomonas putida* pWW0 TOL plasmid, which codes for the catechol 2,3-dioxygenase enzyme.

2.3 Materials and Methods

2.3.1 Media, bacterial strains and plasmids

Lysogeny broth (LB) (Bertani, 1951) or M9 (Miller, J. H., 1972) was used as rich or minimal defined media, respectively. The antibiotics chloramphenicol and tetracycline were used at a final concentration of 20 µg/mL in LB, while rifampicin was used at a final concentration of 50 µg/mL in LB. Ampicillin was used at 100 µg/mL in LB to provide selective pressure for Amp^R plasmids or at 30 µg/mL in LB to select for the *amiB*::*mudA* chromosomal insertion. The *S. enterica* strains used in this study are listed in Table 2.1 and the plasmids used in this study are listed in Table 2.2. To construct ALS2583, a P22 HT105/1 *int-201* lysate (Schmieger, 1972) prepared from GT467 was used to transduce *amtB211*::*mudA* into TR4780. LB Amp^R transductants were selected that could not grow on minimal M9 glucose. To construct ALS234, a P22 HT105/1 *int-201* lysate prepared from TT16237 was used to transduce *zjf-3693*::Tn*10*dTet into TR3359. LB Tet^R transductants were selected that could not grow on minimal M9 glucose. To construct ALS2241 and ALS2242, a P22 HT105/1 *int-201* lysate prepared from TT20702 was used to transduce *btuB12*::Tn*10*dCam into TT16237 and ALS234, respectively. The construction of

ALS1442, which contains the *proB*::*xylE*(*cat*) insertion is described later in the Methods section. Two *E. coli* strains were also used in this study, CS520, *glyV50 metB1 relA1 spoT1 trpA58 tyrT58*(AS) from the Carbon laboratory and MC1061, Δ(*araABOIC-leu*)7679 *araD139 hsr- hsm+ galU galK* Δ(*lac*)*X74 rp sL* (Casadaban & Cohen, 1980).

2.3.2 Determining the mutation responsible for sumA

Because the *zjf-U130*::Tn*10*dTet insertion was 99% linked to *sumA*, it was expected to be <500 bases from the mutation that caused *sumA*. To determine the mutation that caused *sumA*, the tetracycline resistance of the Tn*10*dTet insertion, which was 99% linked to *sumA*, was replaced with chloramphenicol resistance from pACYC184 (Chang & Cohen, 1978) to facilitate cloning of the region surrounding the Tn*10*dTet insertion, using the lambda Red recombination system (Datsenko & Wanner, 2000; Yu et al., 2000). Using the forward primer 5′ CTGATGAATCCCCTAATGATTTTGGTAAAAATCATTAAGTTAAGGTGGA TTTGAGAAGCACACGGTCACA 3′ and the reverse primer 5′ CTGATGAATCCCCTAATGATTTTGGTAAAAATCATTAAGTTAAGGTGGATT ACCTGTGACGGAAGATCAC 3′, a 1163-bp fragment was amplified using the polymerase chain reaction (PCR), *Pfu* polymerase, and pACYC184 as a template, which contained the last 50 bases of the inverted repeats from Tn*10*dTet (del16 del17 TetR, (Way, Davis, Morisato, Roberts, & Kleckner, 1984) and chloramphenicol resistance from pACYC184. The homology for the pACYC184 chloramphenicol resistance from 419 to 3601 bp is underlined. The 1163-bp fragment was gel isolated

and electroporated into LT2 pKM201 cells that were prepared as described by (Murphy, K. C. & Campellone, 2003) to express the lambda gam and red recombination genes. Chloramphenicol-resistant colonies were selected and checked for tetracycline sensitivity to ensure that tetracycline resistance had been replaced with chloramphenicol resistance in the resulting strain. Genomic DNA was prepared from the chloramphenicol-resistant strain, partially digested with *Sau*3AI to generate 5000-bp fragments and ligated into pTrc99A (Amann, Ochs, & Abel, 1988), which had been digested with *Bam*HI and dephosphorylated with calf intestinal alkaline phosphatse. MC1061 transformants that were both ampicillin and chloramphenicol resistant were selected. Plasmid DNA was prepared from pTrc99A-*sumA*, one of the clones that harbored a 5000-bp insert and sequenced using the following two primers; 5′ TGTGACCGTGTGCTTCTCAA 3′, a primer that sequences outwards from the beginning of the chloramphenicol resistance region from pACYC184 and 5′ TGATCTTCCGTCACAGGT 3′, a primer that sequences outwards from the end of the chloramphenicol resistance region from pACYC184. The site of the Tn*10*dTet insertion that was 99% linked to *sumA* was determined to be at bp 4,596,266 of the *S. enterica* chromosome (McClelland et al., 2001a), GenBank accession number NC_003197.

Since the *glyV, glyX, glyY* tRNA locus, which could be mutated to generate a missense suppressor, was immediately downstream of the Tn*10*dTet insertion, genomic DNA was prepared from ALS233, which harbors the *sumA* missense suppressor, PCR amplified and sequenced using the 5’ GCGAAAAAATGCGTTCAGGG 3’ and 5’ GCCCTGTGGATAAGTCTGTT 3’

primers designed to amplify the *glyV*, *glyX*, *glyY* region between bp 4596321 – 4597058 of the *S. enterica* chromosome. The mutation in *sumA* was determined to be a C to T change at bp 4,596,687 of the *S. enterica* chromosome.

A *glyV* tRNA Gly3(GAU/C) missense suppressor, which should be identical to the *sumA* missense suppressor from *S. enterica* has also been identified in *E. coli* (Fleck & Carbon, 1975; Guest & Yanofsky, 1965). Since the *glyV* tRNA Gly3(GAU/C) missense suppressor was never sequenced, we did so to confirm this fact. Genomic DNA was prepared from CS520, which harbors the *gly*V tRNA Gly3(GAU/C) missense suppressor and the *glyV*, *glyX*, *glyY* region between bp 4,391,858 and 4,392,795 of the *E. coli* chromosome was PCR amplified using the 5' TGAACTGGCAACGCTCGAAT 3' and 5' CACCGTGCGAAGTTTCTTTG 3' primers and then sequenced using the 5' CGGCGTGATTTTGACGCTAA 3' and 5' CACCGTGCGAAGTTTCTTTG 3' primers. The mutation in *glyV* tRNA Gly3(GAU/C) was determined to be a C to T change at bp 4,392,617 of the *E. coli* chromosome (Blattner et al., 1997), GenBank accession number U00096.3.

2.3.3 Construction of pACYC184-xylE

Initially the *xylE* gene was cloned into the pTrc99A expression vector. To construct pTrc99A-*xylE*, which maximized the expression of *xylE*, forward primer 5' ATC AGA CTG CAG GAG GTA ACA GCT ATG AAC AAA GGT GTA ATG CGA CC 3' and reverse primer 5' TAG CAG TGG CAG CTC TGA AAG CTT TGC ACA ATC TCT GCA ATA AGT CG 3' and *Pfu* polymerase were used to PCR amplify a 1,006 bp fragment from the pXE60 plasmid (Delic, Robbins, &

Westpheling, 1992), which contained the wild-type *Pseudomonas putida xylE* gene isolated from the TOL pWW0 plasmid and a strong Shine-Dalgarno ribosome binding site, (restriction enzyme sites are indicated with a double underline while the regions of homology to *xylE* are indicated by with a single underline). The resulting fragment was gel isolated, digested with *Pst*I and *Hin*dIII and then ligated into the pTrc99A vector which had been digested with the same two restriction enzymes. To construct pACYC184-*xylE*, forward primer 5' ATATCCATAAGCTTCGCCGACATCATAACGGTTC 3' and reverse primer 5' TCATGACCGTGCTGACCTGATGATCATTGGATAT 3' and *Pfu* polymerase were used to PCR amplify a 1,090 bp fragment from pTrc99A-*xylE* that contained the *trc* promoter and the *xylE* gene (restriction enzyme sites are indicated with an underline). The resulting fragment was gel isolated, digested with *Hin*dIII and *Bcl*I and then ligated into the pACYC184 vector which had been digested with the same two restriction enzymes.

2.3.4 Construction of the proB::xylE(cat) insertion

Using pACYC184-*xylE* as template DNA, the forward primer 5' ATGAGTGACAGCCAGACGCTGGTCGTAAAACTCGGCACCAGCGTGCTAAC GCCGACATCATAACGGTTCT 3' and the reverse primer 5' TTATCGAGTAATCATGTCATCACGATGAACAGCGACCGGGCCATATTCAT TACCTGTGACGGAAGATCAC 3' and *Pfu polymerase* was used to PCR amplify a 2,312 bp fragment that contained the start of the *proB* gene along with the following 47 bases, the *trc* promoter and the *xylE* gene through the *cat* promoter region and the

cat gene of pACYC184-*xylE*, and the stop of the *proB* gene along with the preceding 47 bases. The homology to the pACYC184-*xylE* plasmid is underlined. The resulting fragment was gel isolated, electroporated into LT2 pKM201 cells that were prepared as described by (Murphy, K. C. & Campellone, 2003) and plated on LB chloramphenicol plates at 37^0C to kick the pKM201 plasmid. Recombinant colonies were streaked on LB chloramphenicol plates at 37^0C and patched on LB chloramphenicol plates, M9 glucose plates plus or minus proline, and LB ampicillin plates to verify that the desired *proB*::*xylE*(*cat*) insertion had been isolated. The *proB*::*xylE*(*cat*) insertion produced two forms of the catechol 2,3-dioxygenase protein as verified using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The majority of the protein produced was from the wild-type *xylE* gene. A minority protein product was also produced that contained the wild-type *xylE* gene and a 20 amino acid amino terminal extension due to the in frame ATG start from the pTrc99A expression vector.

2.3.5 Generating *lacZ* and *xylE* missense mutants

Strains TT18519 (*hisC*10081::MudF[*lac*+]) or ALS1442 (*proB*::*xylE*[*cat*]) were mutagenized with ethylmethane sulfonate (EMS) as described by (Miller, J. H., 1972). The *lacZ* mutants were isolated using LB plates that contained 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and 40 μg/mL 5-bromo-4-chloro-3-indoyl β-D-galactopyranoside (X-gal). White or very light blue colonies were selected. The *xylE* mutants were isolated on LB plates that contained 1 mM IPTG and then the resulting colonies were sprayed with a light mist of a 100 mM potassium

phosphate buffer, pH 7.5, that contained 100 mM catechol. White or very light yellow colonies were selected. Nonsense mutants due to amber, ochre or opal mutations were identified using the *supD* (TT7610), *supE* (TT2344) and *supF* (TT2337) amber suppressors, the *supC* (TT13029) and *supM* (TT2839) ochre suppressors and the *supU* (TT4029) opal suppressor. Missense mutants suppressible by *sumA* were identified using the *sumA* missense suppressor (TT16237).

The potential *lacZ* and *xylE* missense mutants were sequenced by PCR amplifying the *lacZ* or *xylE* genes using *Pfu* polymerase and then using overlapping primers to sequence both strands of either the *lacZ* or *xylE* genes. The GenBank accession number for the sequence of the *xylE* gene is M64747.1 (Harayama, Rekik, Bairoch, Neidle, & Ornston, 1991). The sequence of the *lacZ* gene was taken from the genomic sequence of MG1655, the first *E. coli* strain to be sequenced (Blattner et al., 1997), GenBank accession number U00096.3. During the sequencing of the *lacZ* missense mutants we noticed that the sequence of the *lacZ* gene in the *hisC*10081::MudF(*lac*+) insertion differed from the sequence of the *lacZ* gene in MG1655 by one codon, The CAA glutamine codon at amino acid 703 in the *lacZ* gene in MG1655 was changed to a UUA leucine codon in the *lacZ* gene in the *hisC*10081::MudF(*lac*+) insertion. The glutamine to leucine codon change is in the last amino acid of a β-sheet and by either Chou-Fasman-Prevelige (Prevelige & Fasman, 1989), Garnier-Osguthorpe-Robson (Garnier, Osguthorpe, & Robson, 1978a) or Qian-Sejnowski (Qian & Sejnowski, 1988b) protein secondary structure analysis, is predicted to be a neutral change.

2.3.6 Suppression tests

Suppression tests were conducted using amber, ocher or opal nonsense suppressor or *sumA* missense suppressor strains, which also harbored a Tn*10* or Tn*10*dTet transposon insertion that was linked to the suppressor. All mutants to be tested were transduced with a P22 lysate prepared from the appropriate suppressor strain. Transduced *hisC* or *hisD* mutants were plated on LB tetracycline plates and suppression was determined by patching 50 transductant colonies onto M9 glucose plates and scoring whether the presence of the suppressor restored the ability of *hisC* or *hisD* mutants to grow on minimal M9 glucose. Transduced *lacZ* mutants were plated on LB tetracycline plates supplemented with both 1 mM IPTG and 40 μg/mL X-gal and suppression was determined by whether the presence of the suppressor restored blue color. Transduced *xylE* mutants were plated on LB tetracycline plates supplemented with 1 mM IPTG and suppression was scored by spraying the transductant colonies with 100 mM catechol and determining whether the presence of the suppressor restored yellow color.

2.3.7 β-galactosidase and catechol 2,3-dioxygenase enzyme assays

β-galactosidase assays were performed as described by Miller (1972). Because the β-galactosidase and catechol 2,3-dioxygenase enzyme assays both utilize a colorless substrate that is converted to a colored product, we optimized the catechol 2,3-dioxygenase assay described by Sala-Trepat and Evans (1971) to generate a more robust easy to use assay where catechol 2,3-dioxygenase activity could be measured in units similar to the units that were developed for the β-galactosidase assay(Cole,

Ashley E. et al., 2017). The optimal buffer was determined to be Z buffer, which is used in β-galactosidase assays, the maximal adsorption of the yellow product, 2-hydroxymuconate semialdehyde, was determined to be 368 nm and the optimal substrate concentration was determined to be 0.067 mM catechol as specified by (Sala-Trepat & Evans, 1971).

To conduct the catechol 2,3-dioxygenase assays 100 µL of a bacterial overnight was added to 3.0 mL of Z buffer that contained 0.27% sodium dodecyl sulfate. 50 µL of chloroform was added and the sample was vortexed thoroughly to lyse the cells and liberate any catechol 2,3-dioxygenase enzyme that was present. After 10 minutes of equilibration at room temperature, 100 µL of 100 mM potassium phosphate buffer, pH 7.5, that contained 2.15 mM catechol was added, the samples were briefly vortexed, and the start time of the assay was recorded. After optimal yellow color development occurred due to the conversion of catechol to 2-hydroxymuconate semialdehyde, the reaction was stopped by the addition of 1.5 mL of methanol to inactivate the catechol 2,3-dioxygenase enzyme and the time of the reaction was recorded. The assay samples were centrifuged to remove cell debris and the 2-hydroxymuconate semialdehyde in the supernatant was measured at 368 nm (OD_{368}). XylE units were determined using the formula, 10,000 x (OD_{368} of the 2-hydroxymuconate semialdehyde)/(T x V x OD_{550} of the concentrated cells), where T was the reaction time in minutes and V was the volume of the concentrated cells that were used in the assay.

2.3.8 Inheritance test of rifampicin resistant mutants isolated in a sumA missense suppressor

0.2 mL of a saturated LB overnight of the isogenic strains ALS2241 (*sumA*) and ALS2242 (wild-type) which both contained the *btuB12*::Tn*10*dCam transposon insertion were plated on LB plates that contained rifampicin. *btuB12*::Tn*10*dCam is 30% linked to the *rpoB* locus, which can be mutated to yield rifampicin resistance. 0.1 mL of 10^{-6} dilutions of the saturated overnights were also plated on LB plates to determine the number of cells present. ALS2242 (wild-type) yielded an average of 4 rifampicin resistant mutants per 1.136×10^{9} cells, while ALS2241 (*sumA*) yielded an average of 102 rifampicin resistant mutants per 7.62×10^{8} cells. P22 HT105/1 *int-201* lysates were prepared from 50 independent rifampicin resistant mutants isolated in ALS2241 (*sumA*) and used to transduce the *btuB12*::Tn*10*dCam into LT2. 200 transductant colonies from each transduction were patched onto LB rifampicin plates to determine whether rifampicin resistance was transferable from ALS2241 (*sumA*) to LT2 (wild-type *S. enterica*). Transferable or inheritable mutations would be transduced 30% of the time, while nontransferable or non-inheritable mutations would not be transduced. All of the rifampicin resistant mutants were transduced at approximately 30%, which is consistent with the linkage of *btuB12*::Tn*10*dCam to *rpoB.*

2.3.9 Growth rate studies

LB overnights of the isogenic strains TT16237 (*sumA*) and ALS234 (wild-type) were diluted 1:200 in fresh LB media and OD_{550} readings were taken every 15

minutes until the OD_{550} reached 0.75. The growth rates (minutes^{-1}) were determined by calculating the slope of a plot of the growth time in minutes versus the natural logarithm of the OD_{550} reading, while the doubling times were calculated as ln2/slope.

2.4 Results

2.4.1 Determining the mutation responsible for the sumA missense suppressor

The *sumA* mutant has been mapped near the *purA* locus in *S. enterica* (Sanderson & Hartman, 1978) and based on this information we conducted a three factor cross with the *amiB, purA* and *sumA* genes to determine the exact location of *sumA*. A P22 HT105/1 *int-201* lysate prepared from a donor *amiB*+, *purA*+, *sumA* missense suppressor strain was used to transduce a recipient *amiB*::mudA, *purA*155, *sumA*+ (wild-type) strain. Table 2.3 shows the results of this cross and a linkage map indicating the location of the *sumA, amiB* and *purA* genes is shown in Figure 1. Because the *sumA* mutation was highly linked to the *amiB* locus, we screened a collection of Tn*10*dTet transposon insertions from the Björk laboratory that mapped to the *amiB* locus and identified a *zjf-U130*::Tn*10*dTet insertion that was 99% linked to *sumA* (data not shown). The location of the *zjf-U130*::Tn*10*dTet and the base pair change that is responsible for the *sumA* mutation was determined as described in the Materials and Methods section.

The mutation in the *sumA* missense suppressor was determined to be a C to T change at bp 4,596,687 of the *S. enterica* chromosome (McClelland et al., 2001b),

GenBank accession number NC_003197, which changes the anticodon of *glyX*, the second of three duplicate copies of the Gly3 tRNA, from GCC to GUC. Thus, the *sumA* missense suppressor should be classified as a *glyV* tRNA Gly3(GAU/C) missense suppressor and is expected to insert a glycine amino acid instead of aspartic acid at GAU or GAC codons. The *glyV* tRNA Gly3(GAU/C) missense suppressor has been previously identified in *E. coli* but not verified by DNA sequencing (Fleck & Carbon, 1975; Guest & Yanofsky, 1965) To show that the *glyV* tRNA Gly3(GAU/C) missense suppressor from *S. enterica* and *E. coli* were identical, we sequenced the *glyV* tRNA Gly3(GAU/C) missense suppressor from *E. coli* as described in the Materials and Methods section. The mutation in the *glyV* tRNA Gly3(GAU/C) missense suppressor from *E. coli* was determined to be a C to T change at bp 4,392,617 of the *E. coli* chromosome (Blattner et al., 1997), GenBank accession number U00096.3, which changes the anticodon of *glyY*, the third of three duplicate copies of the Gly3 tRNA, from GCC to GUC.

2.4.2 Characterizing the efficiency and specificity of the sumA missense suppressor

Initially we tested the well-characterized *hisD* mutant collection that was isolated by (Greeb, Atkins, & Loper, 1971) and found that 14 out of 57, or 24.56%, of the missense mutants analyzed from this collection could be suppressed by *sumA*. To further characterize the efficiency and specificity of the *sumA* missense suppressor, we created a collection of missense mutants in both the *lacZ* and *xylE* genes. These

genes were chosen because they both encode well-characterized enzymes, for which robust colorimetric assays are available.

EMS mutagenesis was employed to isolate 100 *lacZ* and *xylE* missense mutants as described in the Methods section. The *sumA* missense suppressor was able to rescue 15 out of the 100, or 15%, of the *lacZ* missense mutants and 10 out of the 100, or 10%, of the *xylE* missense mutants. To better understand the efficiency and specificity of the *sumA* missense suppressor, 10 of the *lacZ* and *xylE* mutants were sequenced and the enzymatic activities of the rescued β-galactosidase and catechol 2,3-dioxygenase enzymes was determined. Table 2.4 and Table 2.5 lists the *lacZ* and *xylE* missense mutants that were characterized, the results of the mutation on both the gene and protein and gives both the phenotypic and enzymatic activities of the parental missense mutant in the absence or presence of the *sumA* missense suppressor. The fold increase in enzymatic activity due to the *sumA* missense suppressor is also given along with the percentage of wild-type enzymatic activity that is restored by the *sumA* missense suppressor. The phenotypes listed in Table 2.4 and Table 2.5 regarding the ability of the *sumA* missense suppressor to rescue *lacZ* or *xylE* missense mutants are shown in Figure 2.2 and Figure 2.3 As in the missense suppressor studies conducted in *E. coli* with *trpA* missense mutants or in *S. typhimurium* with *hisC* or *nadC* missense mutants, the suppression of the *lacZ* missense mutants could also be analyzed on minimal media, since the *sumA* missense suppressor could restore the growth of *lacZ* missense mutants on M9 lactose plates. However, as seen in Figure 2.2, the use of LB X-gal media to analyze the suppression of the *lacZ* missense mutants was much more sensitive.

Most of the *lacZ* and *xylE* missense mutants contained glycine to aspartic acid codon changes as expected, given that the *sumA* missense suppressor is a glyV tRNA Gly3(GAU/C) missense suppressor that inserts a glycine amino acid instead of aspartic acid. All of the glycine to aspartic acid mutated codons were GGC to GAC changes and no GGU to GAU changes were observed. Interestingly, a large number of the missense mutants, 6 out of 20, or 30%, contained codon changes other than a glycine to aspartic acid. For most of the missense mutants the *sumA* missense suppressor restored an average of 2.53% of the wild-type enzyme activity. There was one notable exception. The *sumA* missense suppressor restored 30.28% of the wild-type enzyme activity of the *lacZ2454* missense mutant. While the *sumA* missense suppressor caused a lower fold increase in the activities of the *xylE* mutants than the *lacZ* mutants, the ability of *sumA* to restore wild-type enzymatic activities for the two proteins was very similar.

2.4.3 Analyzing the missense mutants with codon changes other than glycine to aspartic acid that are suppressible by sumA

The simplest explanation for why the *sumA* missense suppressor could suppress missense mutants with codon changes other than a glycine to aspartic acid was that an adjoining (connecting) or neighboring (contacting) amino acid was an aspartic acid whose conversion to glycine rescued the defective codon change. Using PyMOL we determined whether any of the missense mutants that contained amino acid changes other than a glycine to aspartic acid also contained neighboring amino acids that were aspartic acid. All 6 of the missense mutants in question contained

aspartic acid neighboring amino acids. Table 2.6 lists the codon changes in these missense mutants and the position and distance in Å of the neighboring aspartic acid. Half of the neighboring aspartic acids were coded by the GAC aspartic acid codon and half of the neighboring aspartic acids were coded by the GAU aspartic acid codon, both of which can be suppressed by the *sumA* missense suppressor. The neighboring aspartic acids were an average of 8.85 Å from the affected codon. For comparison purposes, the distance between aspartic acids and adjoining aspartic acids, or asparagine or leucine amino acids, the two amino acids that are structurally very similar to aspartic acid, ranged from 4.65 – 5.88 Å, with an average of 5.30 Å, for the β-galactosidase and catechol 2,3-dioxygenase proteins.

2.4.4 Inheritance test of mutations generated by sumA

The *glyV* tRNA Gly3(GAU/C) missense suppressor has also been identified as the *mutA* mutator in *E. coli* (Michaels, Cruz, & Miller, 1990; Slupska, Baikalov, Lloyd, & Miller, 1996). While it is clear that the mutation rate is significantly higher in the *glyV* tRNA Gly3(GAU/C) missense suppressor than in wild-type *E. coli* (Al Mamun, Abu Amar M, Rahman, & Humayun, 1999; Al Mamun, Abu Amar M, Marians, & Humayun, 2002; Balashov & Zafri Humayun, 2004; Dorazi, Lingutla, & Humayun, 2002; Michaels et al., 1990; Murphy, H. S. & Humayun, 1997; Slupska et al., 1996) no one has determined whether gain of function mutants generated in a *glyV* tRNA Gly3(GAU/C) missense suppressor are inheritable. Because it is quite possible that the increased mutation rate in gain of function mutants was due to the tolerance or potential benefit of aspartic acid to glycine amino acid changes induced

by the *glyV* tRNA Gly3(GAU/C) missense suppressor, we decided to conduct an inheritance test. If some of the gain of function mutations occurred because of the aspartic acid to glycine amino changes generated by the *glyV* tRNA Gly3(GAU/C) missense suppressor then not all of the mutations isolated in a strain containing the *glyV* tRNA Gly3(GAU/C) missense suppressor would be functional in a wild-type strain that lacked the *glyV* tRNA Gly3(GAU/C) missense suppressor.

Initially we confirmed that the mutation rate of gain of function rifampicin resistant mutants was significantly higher in a *S. enterica sumA* strain than in wild-type *S. enterica*. Rifampicin resistant mutants were obtained and quantified as described in the Materials and Methods section and it was determined that rifampicin resistant mutants occurred at a rate of 0.35 per 1 x 10^8 cells in wild-type *S. enterica* and 13.39 per 1 x 10^8 cells in *S. enterica sumA*. Thus, the formation rate of rifampicin resistant mutants increased by 38.26 fold in *S. enterica sumA* cells versus wild-type, a value that is consistent with previous studies. The inheritance test was conducted as described in the Materials and Methods section and we observed that the mutations that caused 50 independent rifampicin resistant mutants in a *S. enterica sumA* strain could all be transduced back into the wild-type *S. enterica* strain and that rifampicin resistance was transferred. Thus, all of the rifampicin resistant mutations isolated in the *S. enterica sumA* strain were inheritable by the *S. enterica* wild-type strain.

2.4.5 The sumA missense suppressor causes a significant reduction in the growth rate

We noticed that strains containing the *sumA* missense suppressor grew significantly slower than strains that lacked the *sumA* missense suppressor under both suppressing and non-suppressing conditions. For this reason, we determined the growth rate of the isogenic strains TT16237 (*sumA*) and ALS234 (wild-type) in LB media. The growth rate and doubling time of the strain containing the *sumA* missense mutation was 1.0486×10^{-2} minutes^{-1} and 66.10 minutes, respectively, while the growth rate and doubling time of the wild-type strain was 1.5460×10^{-2} minutes^{-1} and 44.83 minutes, respectively. Thus, the presence of the *sumA* missense suppressor reduced the growth rate or doubling time of wild-type *S. enterica* by 47.43%.

2.5 Discussion

In this study, we have shown that the *Salmonella enterica* serovar Typhimurium *sumA* missense suppressor is caused by a GCC to GUC change in the anticodon of one of the three copies of the Gly3 tRNA, which can cause a glycine amino acid to be inserted instead of an aspartic acid at GAU or GAC codons. Thus, the *sumA* missense suppressor is more accurately designated as a *glyV* tRNA Gly3(GAU/C) missense suppressor. While the efficiency of the *sumA* missense suppressor was similar to the efficiency of other missense suppressors that have been characterized, it was verified to be broad acting and able to rescue a large number of missense mutants, a trait that has only been observed in one other missense suppressor, the *glyU* tRNA Gly1(GAG) missense suppressor, which can cause a

glycine amino acid to be inserted instead of a glutamic acid. The *sumA* missense suppressor was verified to be identical to the *glyV* tRNA Gly3(GAU/C) missense suppressor that was isolated in *E. coli* and like its *E. coli* counterpart, was shown to be capable of acting as a mutator. Just as the *glyV* tRNA Gly3(GAU/C) *sumA* missense suppressor has proven useful as a research tool in *S. enterica*, it could be used in *E. coli* as well to identify missense mutants.

The suppression efficiencies for the *sumA* missense suppressor ranged from 1.67 – 3.87% of the wild-type enzyme activity for the *lacZ* and *xylE* missense mutants that were tested with one exception. The *sumA* missense suppressor restored 30.28% of the wild-type enzyme activity for the *lacZ2504* missense mutant. In the suppression efficiency studies that have been conducted on other missense suppressors, the suppression efficiencies ranged from 1.10 – 3.60% with two notable exceptions (Berger & Yanofsky, 1967; Brody & Yanofsky, 1963; Hill, Squires, & Carbon, 1970; Hill et al., 1974) The *glyU* tRNA Gly1(AGA) missense suppressor restored 10.50% of the wild-type enzyme activity of the *trpA36* missense mutant and the *glyT* tRNA Gly2(AGA) missense suppressor restored 27.00% of the wild-type enzyme activity of the *trpA36* missense mutant (Brody & Yanofsky, 1963; Hill et al., 1970). Thus, the results of our efficiency studies with the *sumA* missense suppressor are in very close agreement with the efficiency studies for other missense suppressors. With only a couple of exceptions the suppression efficiencies of missense suppressors range from 1.10 – 3.87%.

If one combines the data on the specificity of suppression of the *sumA* missense suppressor from the two previous studies (Hughes et al., 1991; Whitfield et

al., 1966) and our study, *sumA* suppressed 53 out of 320, or 16.56%, of the missense mutants that have been tested. Only one other broad-acting missense suppressor has been characterized, the *glyU* tRNA Gly1(GAG) missense suppressor. Table 2.7 compares the data on the narrow versus broadly acting missense suppressors that have been definitively characterized. The data on the frequency of codon usage is from (Maloy, Stewart, & Taylor, 1996) and the data on the abundance of the tRNAs is from (Dong, Nilsson, & Kurland, 1996). If one considers all the factors that determine whether a missense suppressor might be predicted to be broad acting, the data in Table 2.7 clearly indicates that the *sumA* missense suppressor is expected to be broad acting. Most important amongst these factors is codon usage. The GGC glycine codon, which is the codon most often mutated in the *sumA* suppressible missense mutants that were sequenced, is the sixth most abundant codon found in translated genes and occurs at a frequency of 30 codons per 1,000 total codons. The GGU glycine codon, which can also be mutated to generate *sumA* suppressible missense mutants, is the eighth most abundant codon found in translated genes and occurs at a frequency of 28 codons per 1,000 total codons. Only the CUG leucine codon, the GAA glutamic acid codon, the AAA lysine codon, the GAU aspartic acid codon, and the GCG alanine codon, which occur at frequencies of 52, 44, 38, 33 and 32 codons per 1,000 total codons, respectively, are more widely utilized than the GGC glycine codon. A second factor to consider is how prevalent is the mutated codon change in translated genes. The GAU and GAC aspartic acid codons are both suppressible by the *sumA* missense suppressor and occur at very high frequencies in translated genes. The GAU aspartic acid codon is the fourth most abundant codon

found in translated genes and the GAC aspartic acid codon ranks in the upper third of codons that are found most often in translated genes. A third factor to consider is tRNA abundance, since mutated tRNAs that are more abundant are more likely to be available for suppression. Amongst the tRNAs, only the Arg2, Glu2 and Leu1 tRNAs, which constitute 7.37%, 7.32% and 6.95%, respectively, of the total tRNA population, are more abundant than the Gly3 tRNA, which constitutes 6.76% of the total tRNA population and is modified to form the *sumA* missense suppressor. Additionally, as detailed later in the discussion, aspartic acid to glycine amino acid changes are very tolerable with respect to protein structure.

While not as strong an argument can be made for the *glyU* tRNA Gly1(GAG) missense suppressor being a broad acting missense suppressor compared to the *sumA* missense suppressor based on the data in Table 2.7, a compelling case can be made in comparison to the *glyT* tRNA Gly2(AGA) and *serU* tRNA Ser2(UUG) missense suppressors, which have very narrow specificities. The abundance of the GGG glycine codon in translated genes is below average, but definitely not considered to be a rare codon and the mutated GAG glutamic acid codon ranks in the upper third of codons that occur at very high frequencies in translated genes. Only eight other tRNAs are more abundant than the Gly2 tRNA. Additionally, glutamic acid to glycine changes should also be very tolerable with respect to protein structure. One must exercise caution, however, because the assessment of the broad acting ability of the *glyU* tRNA Gly1(GAG) missense suppressor is based on a single study with a very limited number of missense mutants (Eggertsson & Adelberg, 1965). It would be very interesting to determine whether the *glyU* tRNA Gly1(GAG) missense

suppressor was truly broad acting in a study involving a greater number of missense mutants.

Given that most mutants, which are isolated spontaneously or through the use of ultraviolet radiation or chemical mutagenesis, are either C to T or G to A transitions, if one considers the factors required to yield a broad-acting missense suppressor that were discussed above, one other missense suppressor should be isolatable that would be expected to resciue a large number of missense mutants similar to the *sumA* missense suppressor. A tRNA Glu2(AAA/G) missense suppressor that recognizes AAA or AAG lysine codons could be derived from tRNA Glu2(GAA/G), which normally recognizes GAA or GAG glutamic acid codons. The GAA glutamic acid codon is the second most abundant codon, the AAA lysine codon is the third most abundant codon and the Glu2 tRNA is the second most abundant tRNA. Additionally, missense mutants that resulted from glutamic acid to lysine changes would be very disruptive to protein structure and function due to the acidic to basic change and thus be expected to be quite prevalent. The equivalent of the predicted tRNA Glu2(AAA/G) missense suppressor has been isolated in *Saccharomyces cerevisiae* (Su, Belmont, & Sclafani, 1990). The *S. cerevisiae SOE1* tRNA 3Glu(AAA/G) missense suppressor suppressed 5 out of 8, or 62.5% of the potential missense mutants that were analyzed. The mutants that were analyzed in this study were not known definitively to be missense mutants and it would be interesting to isolate and characterize a tRNA Glu2(AAA/G) missense suppressor in *S. enterica*, where large sets of missense mutants are available for analysis.

Six of the 20, or 30%, of the missense mutants that were suppressible by the *sumA* missense suppressor contained a mutation other than glycine to aspartic acid. Interestingly, all 6 of these missense mutants contained neighboring (contacting) aspartic acids that were in close proximity to the mutated amino acid. These neighboring aspartic acids, which could be changed from an aspartic acid to a glycine by the *sumA* missense suppressor, were within an average of 8.85 Å from the mutated amino acid. To put these distances into perspective, the distances between aspartic acids and adjoining aspartic acids or adjoining asparagine or leucine amino acids, the amino acids most structurally similar to aspartic acid, ranged from 4.65 – 5.88 Å for the β-galactosidase and catechol 2,3-dioxygenase proteins. Thus, in all cases, missense mutants that contained mutations other than a glycine to aspartic acid change also contained a very near neighboring aspartic acid that was in close contact and enabled suppression to occur. This observation reinforces the fact that aspartic acid to glycine changes are very well tolerated in proteins. 7 out of the 9 neighboring aspartic acids were in connecting loops or turns. Glycine is the most prevalent amino acid that is found in connecting loops or turns and the conversion of aspartic acid to glycine by the *sumA* missense suppressor would arguably give the connecting loop or turn greater flexibility. In a study of the amino acids found in turns between two α-helices, glycine was the most prevalent amino acid and occurred 39.39% of the time, followed by valine and serine, both of which occurred at a frequency of 18.18% (Shestopalov, 1988). In a study of the amino acids found in the loops between two β-sheets, glycine was the most prevalent amino acid and occurred 14.42% of the time, followed by aspartic acid, serine and asparagine, which occurred at a frequency of

9.67%, 9.23% and 9.09%, respectively (Minuchehr & Goliaei, 2005). In the β-galactosidase and catechol 2,3-dioxygenase proteins, glycine was the most prevalent amino acid found in connecting loops or turns and occurred 11.26% of the time, followed by proline and aspartic acid, which occurred at a frequency of 9.92% and 8.91%, respectively. While aspartic acid is a prevalent amino acid in connecting loops or turns, glycine is preferred and arguably an aspartic acid to glycine conversion gives the connecting loop or turn greater flexibility.

The *glyV* tRNA Gly3(GAU/C) *sumA* missense suppressor has also been identified and characterized as the *mutA* mutator (Michaels et al., 1990; Slupska et al., 1996) and numerous studies have confirmed that the presence of the *glyV* tRNA Gly3(GAU/C) missense suppressor, or *mutA*, increases the spontaneous generation rate of mutants (Al Mamun, Abu Amar M et al., 1999; Al Mamun, Abu Amar M et al., 2002; Balashov & Zafri Humayun, 2004; Dorazi et al., 2002; Michaels et al., 1990; Murphy, H. S. & Humayun, 1997; Slupska et al., 1996). The glyV tRNA Gly3(GAU/C) missense suppressor appears to cause a change in DNA polymerase III, which acts as a mutator, although the nature of this change has not been determined (Al Mamun, Abu Amar M et al., 2002). In all of the studies that have been conducted on the ability of the *glyV* tRNA Gly3(GAU/C) missense suppressor to act as a mutator, not one has determined whether the increase in the mutational rate is due in part to an increase in mutants whose function would be restored by the presence of the *glyV* tRNA Gly3(GAU/C) missense suppressor, so we conducted a gain of function inheritance test to determine whether this was the case. 50 independent rifampicin resistant mutants were isolated in a *glyV* tRNA Gly3(GAU/C)

missense suppressor and transduction tests were conducted to verify that all 50 of the rifampicin resistant mutants were functional when transduced back into a wild-type background that lacked the *glyV* tRNA Gly3(GAU/C) missense suppressor, thus definitively proving that the *glyV* tRNA Gly3(GAU/C) *sumA* missense suppressor can act as a mutator due to the change in DNA polymerase III, and that the increase in the mutational frequency is not due to an increase in mutants which can be rescued by the *glyV* tRNA Gly3(GAU/C) missense suppressor.

Just as the *glyV* tRNA Gly3(GAU/C) *sumA* missense suppressor can act as a mutator by altering DNA polymerase III, it is possible that the reason *sumA* acts broadly and suppress missense mutants caused by amino acid changes other than a glycine to aspartic acid, is due to its ability to alter the translational process by the ribosome. Alterations in the ribosomal P, A or E sites or ribosomal release or elongation factors could be expected to cause mistranslation. While this might be the case, we feel that it is unlikely, since all of the suppressible missense mutants that were not glycine to aspartic acid amino acid changes as expected on the basis of codon anticodon pairing, contained neighboring aspartic acids that could be converted to glycine by the *sumA* missense suppressor.

Table 2.1. *S. enterica* strains

Name	Genotype	Source
ALS234	*hisC537 zjf-3693*::Tn*10*dTet (40% linked to *sumA*)	this study
ALS1442	*proB*::*xylE*(*cat*)	this study
ALS2241	*btuB12*::Tn*10*dCam *hisC537 sumA10 zjf-3693*::Tn*10*dTet	this study
ALS2242	*btuB12*::Tn*10*dCam *hisC537 zjf-3693*::Tn*10*dTet	this study
ALS2583	*amiB211*::*mudA hisC537 purA155*	this study
GT2086	*zjf-U130*::Tn*10*dTet (99% linked to *sumA*)	Björk laboratory
GT467	*amtB211*::*mudA* (*amiB211*::*mudA*)	Björk laboratory
LT2	wild-type	Roth laboratory
TR3359	*hisC537*	Roth laboratory
TR4780	*hisC537 purA155*	Roth laboratory
TT2337	*hisC527*(UAG) *leuA414*(UAG) *supF* (UAG, *tyr*) *zde-94::*Tn*10* (50% linked to *supF*)	Roth laboratory
TT2344	*hisC527*(UAG) *leuA414*(UAG) *supE* (UAG, *gln*) *zbf-604*::Tn*10* (51% linked to *supE*)	Roth laboratory
TT2839	*hisC527*(UAG) *leuA414*(UAG) *tyrU90* (*supM*) *zii-614*::Tn*10* (45% linked to *supM*)	Roth laboratory
TT4029	*hisO1242 hisB2135 supU1283*(UGA) *zhb-736*::Tn*10* (10% linked to *supU*)	
TT7610	*supD501* (UAG, *ser*) *zeb-609*::Tn*10* (60% linked to *supD*)	Roth laboratory
TT13029	*hisC527*(UAG) *leuA414*(UAG) *supC80*(UAA, UAG, *tyr*) *zde-605*::Tn*10* (54% linked to *supC*)	Roth laboratory
TT16237	*hisC537 sumA10 zjf-3693*::Tn*10*dTet (40% linked to *sumA*)	Roth laboratory
TT18519	ara-9 hisC10081::MudF(*lac*+)	Roth laboratory
TT20702	*btuB12*::Tn*10*dCam (30% linked to *rpoB*)	Roth laboratory

Table 2.2. Plasmids

Name	Relevant Characteristics	Reference
pACYC184	Tet^R, Cam^R, p15A ori	Chang and Cohen, 1978
pACYC184-*xylE*	Tet^R, Cam^R, p15A ori, *trc* promoter / operator, *xylE* gene	this study
pKM201	Amp^R, pAMPts ori, *tac* promoter, lambda *gam* and *red* genes	Murphy and Campellone, 2003
pTrc99A	Amp^R, *trc* promoter / operator, $lacI^Q$, ColE1 ori	Amann et al., 1988
pTrc99A-*sumA*	Amp^R, *trc* promoter / operator, $lacI^Q$, ColE1 ori, *sumA* gene	this study
pTrc99A-*xylE*	Amp^R, *trc* promoter / operator, $lacI^Q$, ColE1 ori, *xylE* gene	this study
pXE60	Amp^R, ColE1 ori, TOL pWWO *xylE* gene	Delic et al, 1992

Table 2.3. Three factor cross results

Recombinant class	Total number out of 300
sumA amiB::*mudA*	8
sumA amiB+ (wt)	112
sumA+ (wt) *amiB*::*mudA*	63
sumA+ (wt) *amiB*+ (wt)	117

Table 2.3. Three factor cross using the *amiB, purA* and *sumA* genes. A P22 HT105/1 *int-201* lysate prepared from a donor *amiB*+, *purA*+, *sumA* missense suppressor strain was used to transduce a recipient *amiB*::*mudA*, *purA155*, *sumA*+ (wild-type) strain, which also contained the *hisC537* mutation, known to be suppressible by *sumA*, to *purA*+ on M9 glucose plates supplemented with histidine. 300 transductants were analyzed and the number of *sumA amiB*::*mudA*, *sumA amiB*+, *sumA*(wild-type) *amiB*::*mudA* and *sumA*(wild-type) *amiB*+ recombinants were scored on M9 glucose plates and LB Amp plates. Based on this information the gene order was determined to be *sumA*, *amiB*, *purA*, due to the rare *sumA amiB*::*mudA* class. The linkage of *purA* to *sumA* was calculated to be 40.0%, the linkage of *purA* to *amiB* was calculated to be 76.3% and based on these linkages, the linkage of *sumA* to *amiB* was predicted to be 63.7%.

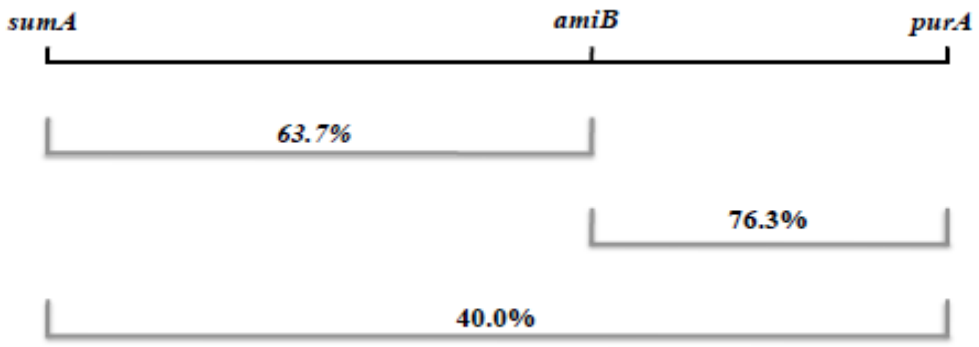

Figure 2.1: Linkage map of the *sumA*, *amiB* and *purA* genes in *S. enterica*. The gene order of the *sumA*, *amiB* and *purA* genes are shown and the linkages based on the data from Table 3 are include

Table 2.4: Characterization of *lacZ* missense mutants that are suppressible by *sumA*

					with the *sumA* missense suppressor			
Mutant [1]	**Base change** [2]	**Amino acid change** [2]	**Phenotype** [3]	**Activity Miller units** [4]	**Phenotype** [3]	**Activity Miller units** [4]	**Fold Increase**	**Percentage of wild-type activity** [5]
lacZ2234	1694 (G to A)	565 (G to D)	W	3.74	LB	99.65	26.64	2.71
lacZ2343	2828 (G to A)	943 (R to H)	VLB	9.78	LB	146.96	15.03	3.87
lacZ2381	2291 (G to A)	764 (G to D)	VLB	6.87	LB	89.17	12.98	2.32
lacZ2382	1697 (G to A)	566 (G to D)	W	3.84	LB	109.67	28.56	2.99
lacZ2396	1643 (G to A)	548 (G to D)	VLB	4.31	LB	65.55	15.21	1.73
lacZ2454	1061 (G to A)	354 (G to D)	W	3.86	LB	122.09	31.63	3.34
lacZ2504	623 (G to A)	208 (G to D)	VLB	4.19	LB	1,076.28	256.87	30.28

1- Of the 10 *lacZ* missense mutants that were sequenced, *lacZ231* was identical to *lacZ2234*, *lacZ2352* was identical to *lacZ2382* and *lacZ2232* was identical to *lacZ2454*, and are not included in the table.
2- The *lacZ* gene is 3,075 bp in length and codes for the 1,024 amino acid β-galactosidase protein. The resulting amino acid changes are given based on the coded protein predicted by the base sequence and not the Protein Data Bank file.
3- The plate phenotypes are depicted as W for white, VLB for very light blue and LB for light blue.
4- β-galactosidase assays were repeated in triplicate and the standard deviation was less than 10%.
5- The percentage of wild-type enzyme activity restored by the *sumA* missense suppressor is calculated with respect to the enzyme activity of wild-type β-galactosidase protein in TT18519 (*hisC*10081::MudF[*lac*+]), which was 3,540.86 Miller units, after subtracting the activity of the non-suppressed mutant.

Table 2.5: Characterization of *xylE* missense mutants that are suppressible by *sumA*

					with the *sumA* missense suppressor			
Mutant [1]	**Base change** [2]	**Amino acid change** [2]	**Phenotype** [3]	**Activity XylE units** [4]	**Phenotype** [3]	**Activity XylE units** [4]	**Fold Increase**	**Percentage of wild-type activity** [5]
xylE1571	380 (G to A)	127 (G to E)	VLY	24.78	LY	34.69	1.40	2.64
xylE1579	26 (G to A)	9 (G to D)	VLY	24.15	LY	30.91	1.28	1.80
xylE1582	191 (G to A)	64 (G to D)	VLY	24.06	LY	30.32	1.26	1.67
xylE1583	473 (G to A)	158 (G to D)	VLY	24.12	LY	31.40	1.30	1.94
xylE1584	727 (C to T)	243 (P to S)	W	18.05	VLY	25.63	1.42	2.02
xylE1585	25 (G to A)	9 (G to S)	W	22.01	LY	31.03	1.41	2.40
xylE1587	532 (G to A)	178 (E to K)	W	18.71	LY	32.74	1.75	3.73
xylE1588	740 (G to A)	247 (G to D)	W	22.14	LY	33.65	1.52	3.06
xylE1589	845 (G to C)	282 (W to S)	W	23.50	LY	29.85	1.27	1.69

1- Of the 10 *xylE* missense mutants that were sequenced, *xylE1570* was identical to *xylE1583* and is not included in the table.
2- The *xylE* gene is 921 bp in length and codes for the 307 amino acid catechol 2,3-dioxygenase protein. The resulting amino acid changes are given based on the coded protein predicted by the base sequence and not the Protein Data Bank file.
3- The plate phenotypes are depicted as W for white, VLY for very light yellow and LY for light yellow.
4- catechol 2,3-dioxygenase assays were repeated in triplicate and the standard deviation was less than 10%.
5- The percentage of wild-type enzyme activity restored by the *sumA* missense suppressor is calculated with respect to the enzyme activity of wild-type catechol 2,3-dioxygenase protein in ALS1442 (*proB*::*xylE*[*cat*]), which was 375.87 XylE units, after subtracting the activity of the non-suppressed mutant.

Table 2.6. Missense mutant neighboring aspartic acid

Mutant	Amino acid change	Neighboring aspartic acid	Distance in Å
lacZ2343	943 (R to H)	660	9.16
		955	5.93
xylE1571	127 (G to E)	51	6.72
		133	8.39
xylE1584	243 (P to S)	295	10.66
xylE1585	9 (G to S)	51	8.31
xylE1587	178 (E to K)	271	8.49
xylE1589	282 (W to S)	182	12.08
		285	9.93

Table 2.6. Position and distance of the neighboring (contacting) aspartic acids in missense mutants that contain codon changes other than a glycine to aspartic acid. PyMOL Molecular Graphics System (Schrödinger, LLC) was used to determine the distance of neighboring aspartic acids. The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) was the source of the three-dimensional crystal structure data. The 1DP0 file was used for the β -galactosidase crystal structure data (Juers et al., 2009) and the 1MPY file was used for the catechol 2,3-dioxygenase crystal structure data (Kita et al., 1999).

Table 2.7. Data on allele specific versus broad acting missense suppressors

Original tRNA	Missense suppressor	Suppression	Original codon abundance [1]	Mutated codon abundance [1]	tRNA abundance [2]
glyT tRNA Gly2(GGA/G)	*glyT* tRNA Gly2(AGA) [3]	narrow	GGA (7/1000)	AGA (2/1000)	3.31%
			GGG (9/1000)		
glyU tRNA Gly1(GGG)	*glyU* tRNA Gly1(GAG) [4]	broad	GGG (9/1000)	GAG (19/1000)	3.31%
glyV tRNA Gly3(GGU/C)	*glyV* tRNA Gly3(GAU/C) [5]	broad	GGU (28/1000)	GAU (33/1000)	6.76%
			GGC (30/1000)	GAC (23/1000)	
serU tRNA Ser2(UCG)	*serU* tRNA Ser2(UUG) [6]	narrow	UCG (8/1000)	UUG (11/1000)	0.53%

1- The frequency that the codon occurs per 1,000 amino acids amongst all *E. coli* proteins. Data is taken from Maloy et al., 1996.
2- The abundance of the tRNA per the total tRNA. Data is taken from Dong et al., 1996.
3- Brody and Yanofsky, 1963; Roberts and Carbon, 1975
4- Eggertsson and Adelberg, 1965; Hill et al., 1974
5- Whitfield et al., 1966; this study
6- Eggertsson and Adelberg, 1965; Thorbjarnardóttir et al, 1985

REFERENCES

Al Mamun, Abu Amar M, Marians, K. J., & Humayun, M. Z. (2002). DNA polymerase III from escherichia coli cells expressing mutA mistranslator tRNA is error-prone. *The Journal of Biological Chemistry, 277*(48), 46319-46327.

Al Mamun, Abu Amar M, Rahman, M. S., & Humayun, M. Z. (1999). Escherichia coli cells bearing mutA, a mutant glyV tRNA gene, express a recA-dependent error-prone DNA replication activity. *Molecular Microbiology, 33*(4), 732-740. doi:10.1046/j.1365-2958.1999.01520.x

Amann, E., Ochs, B., & Abel, K. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in escherichia coli. *Gene, 69*(2), 301-315. doi:10.1016/0378-1119(88)90440-4

Balashov, S., & Zafri Humayun, M. (2004). Specificity of spontaneous mutations induced in mutA mutator cells. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 548*(1), 9-18. doi:10.1016/j.mrfmmm.2003.12.005

Berger, H., & Yanofsky, C. (1967). Suppressor selection for amino acid replacements expected on the basis of the genetic code. *Science, 156*(3773), 394-397. doi:10.1126/science.156.3773.394

Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., . . . Shao, Y. (1997). The complete genome sequence of escherichia coli K-12. *Science, 277*(5331), 1453-1469. doi:10.1126/science.277.5331.1453

Brody, S., & Yanofsky, C. (1963). Suppressor gene alteration of protein primary structure. *Proceedings of the National Academy of Sciences of the United States of America, 50*(1), 9-16. doi:10.1073/pnas.50.1.9

Casadaban, M. J., & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in escherichia coli. *Journal of Molecular Biology, 138*(2), 179-207. doi:10.1016/0022-2836(80)90283-1

Chang, A. C. Y., & Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *Journal of Bacteriology, 134*(3), 1141-1156. Retrieved from http://jb.asm.org/content/134/3/1141.abstract

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America, 97*(12), 6640-6645. doi:10.1073/pnas.120163297

Delic, I., Robbins, P., & Westpheling, J. (1992). Direct repeat sequences are implicated in the regulation of two streptomyces chitinase promoters that are subject to carbon catabolite control. *Proceedings of the National Academy of Sciences of the United States of America, 89*(5), 1885-1889. doi:10.1073/pnas.89.5.1885

Dong, H., Nilsson, L., & Kurland, C. G. (1996). Co-variation of tRNA abundance and codon usage in escherichia coli at different growth rates. *Journal of Molecular Biology, 260*(5), 649-663.

Dorazi, R., Lingutla, J. J., & Humayun, M. Z. (2002). Expression of mutant alanine tRNAs increases spontaneous mutagenesis in escherichia coli. *Molecular Microbiology, 44*(1), 131-141. doi:10.1046/j.1365-2958.2002.02847.x

Eggertsson, G. (1968). Suppressors causing temperature sensitivity of growth in escherichia coli. *Genetics, 60*(2), 269-280.

Eggertsson, G., & Adelberg, E. A. (1965). Map positions and specificities of suppressor mutations in escherichia coli K-12. *Genetics, 52*(2), 319-340.

Fasman, G. D. (1989). *Prediction of protein structure and the principles of protein conformation*. Boston: Springer. doi:10.1007/978-1-4613-1571-1

Fleck, E. W., & Carbon, J. (1975). Multiple gene loci for a single species of glycine transfer ribonucleic acid. *Journal of Bacteriology, 122*(2), 492-501.

Galitski, T., & Roth, J. R. (1996). A search for a general phenomenon of adaptive mutability. *Genetics, 143*(2), 645-659.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology, 120*(1), 97-120. doi:10.1016/0022-2836(78)90297-8

Greeb, J., Atkins, J. F., & Loper, J. C. (1971). Histidinol dehydrogenase (hisD) mutants of salmonella typhimurium. *Journal of Bacteriology, 106*(2), 421-431.

Guest, J. R., & Yanofsky, C. (1965). Amino acid replacements associated with reversion and recombination within a coding unit. *Journal of Molecular Biology, 12*(3), 793-804. doi:10.1016/S0022-2836(65)80328-X

Harayama, Rekik, Bairoch, Neidle, & Ornston. (1991). Potential DNA slippage structures acquired during evolutionary divergence of acinetobacter calcoaceticus chromosomal benABC and pseudomonas putida TOL pWW0 plasmid xylXYZ, genes encoding benzoate dioxygenases. *Journal of Bacteriology, 173*(23), 7540-7548. doi:10.1128/jb.173.23.7540-7548.1991

Hill, C. W. (1975). Informational suppression of missense mutations. *Cell, 6*(4), 419-427. doi:10.1016/0092-8674(75)90031-8

Hill, C. W., Combriato, G., & Dolph, W. (1974). Three different missense suppressor mutations affecting the tRNAGGGGly species of escherichia coli. *Journal of Bacteriology, 117*(2), 351-359.

Hill, C. W., Squires, C., & Carbon, J. (1970). Glycine transfer RNA of escherichia coli: I. structural genes for two glycine tRNA species. *Journal of Molecular Biology, 52*(3), 557-569. doi:10.1016/0022-2836(70)90419-5

Hughes, K. T., Roth, J. R., & Olivera, B. M. (1991). A genetic characterization of the nadC gene of salmonella typhimurium. *Genetics, 127*(4), 657-670.

J W Roberts, & J Carbon. (1975). Nucleotide sequence studies of normal and genetically altered glycine transfer ribonucleic acids from escherichia coli. *Journal of Biological Chemistry, 250*(14), 5530-5541.

Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X., Huber, R. E., Tronrud, D. E., & Matthews, B. W. (2000). High resolution refinement of β-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for α-complementation. *Protein Science, 9*(9), 1685-1699. doi:10.1110/ps.9.9.1685

Kita, A., Kita, S., Fujisawa, I., Inaka, K., Ishida, T., Horiike, K., . . . Miki, K. (1999). An archetypical extradiol-cleaving catecholic dioxygenase: The crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from ppseudomonas putida mt-2. *Structure (London, England : 1993), 7*(1), 25-34.

Maloy, S. R., Stewart, V. J., & Taylor, R. K. (1996). *Genetic analysis of pathogenic bacteria*. New York: Cold Spring Harbor Laboratory Press.

Michaels, M. L., Cruz, C., & Miller, J. H. (1990). mutA and mutC: Two mutator loci in escherichia coli that stimulate transversions. *Proceedings of the National Academy of Sciences of the United States of America, 87*(23), 9211-9215. doi:10.1073/pnas.87.23.9211

Miller, J. H. (1992). *Experiments in molecular genetics* (11. print. ed.). New York: Cold Spring Harbor Laboratory.

Minuchehr, Z., & Goliaei, B. (2005). Propensity of amino acids in loop regions connecting beta-strands. *Protein & Peptide Letters, 12*(4), 379-382. doi:10.2174/0929866053765617

Murgola, E. (1985). tRNA, suppression, and the code. *Annual Review of Genetics, 19*(1), 57-80. doi:10.1146/annurev.genet.19.1.57

Murphy, H. S., & Humayun, M. Z. (1997). Escherichia coli cells expressing a mutant glyV (glycine tRNA) gene have a UVM-constitutive phenotype: Implications for mechanisms underlying the mutA or mutC mutator effect. *Journal of Bacteriology, 179*(23), 7507-7514. doi:10.1128/jb.179.23.7507-7514.1997

Murphy, K. C., & Campellone, K. G. (2003). Lambda red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. *BMC Molecular Biology, 4*(1), 11. doi:10.1186/1471-2199-4-11

Nhan, M., Courtney, L., Stoneking, T., Du, F., Miller, W., Holmes, A., . . . Ryan, E. (2001). Complete genome sequence of salmonella enterica serovar typhimurium LT2. *Nature, 413*(6858), 852-856. doi:10.1038/35101614

Qian, N., & Sejnowski, T. J. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology, 202*(4), 865-884. doi:10.1016/0022-2836(88)90564-5

S Thorbjarnardóttir, H Uemura, T Dingermann, T Rafnar, S Thorsteinsdóttir, D Söll, & G Eggertsson. (1985). Escherichia coli supH suppressor: Temperature-sensitive missense suppression caused by an anticodon change in tRNASer2. *Journal of Bacteriology, 161*(1), 207-211.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews. Molecular Cell Biology, 14*(10), 630-642. doi:10.1038/nrm3658

Sala-Trepat, J. M., & Evans, W. C. (1971). The meta cleavage of catechol by azotobacter species. 4-oxalocrotonate pathway. *European Journal of Biochemistry, 20*(3), 400-413. doi:10.1111/j.1432-1033.1971.tb01406.x

Schmieger, H. (1972). Phage P22-mutants with increased or decreased transduction abilities. *Molecular & General Genetics: MGG, 119*(1), 75-88. doi:10.1007/BF00270447

Shestopalov, B. V. (1988). Amino acid sequence template useful for α-helix-turn-α-helix prediction. *FEBS Letters, 233*(1), 105-108. doi:10.1016/0014-5793(88)81364-4

Slupska, M. M., Baikalov, C., Lloyd, R., & Miller, J. H. (1996). Mutator tRNAs are encoded by the escherichia coli mutator genes mutA and mutC: A novel pathway

for mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America, 93*(9), 4380-4385. doi:10.1073/pnas.93.9.4380

Su, J., Belmont, L., & Sclafani, R. A. (1990). Genetic and molecular analysis of the SOE1 gene: A tRNA(3)(glu) missense suppressor of yeast cdc8 mutations. *Genetics, 124*(3), 523-531.

Way, J. C., Davis, M. A., Morisato, D., Roberts, D. E., & Kleckner, N. (1984). New Tn10 derivatives for transposon mutagenesis and for construction of lacZ operon fusions by transposition. *Gene, 32*(3), 369-379. doi:10.1016/0378-1119(84)90012-X

Whitfield, H. J., Martin, R. G., & Ames, B. N. (1966). Classification of aminotransferase (C gene) mutants in the histidine operon. *Journal of Molecular Biology, 21*(2), 335,IN14,34-346,IN14,355. doi:10.1016/0022-2836(66)90103-3

Yu, D., Ellis, H. M., Lee, E., Jenkins, N. A., Copeland, N. G., & Court, D. L. (2000). An efficient recombination system for chromosome engineering in escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America, 97*(11), 5978-5983. doi:10.1073/pnas.100127597

CHAPTER III: NONFUNCTIONAL MISSENSE MUTANTS IN TWO WELL CHARACTERIZED CYTOSOLIC ENZYMES REVEAL IMPORTANT INFORMATION ABOUT PROTEIN STRUCTURE AND FUNCTION

Key words: protein secondary structure, -helices, β-sheets, unstructured regions, coils, hydropathy

Ashley E. Cole[1], Fatmah M. Hani[1], Brian W. Allen[3], Paul C. Kline[2], Elliot Altman[1]

1- Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132
2- Department of Chemistry, Middle Tennessee State University, Murfreesboro, TN 37132
3- Department of Biomedical Engineering, Duke University, Durham, NC 27708

Fatmah Hani sequenced approximately half of the missense mutants characterized in this study, conducted most of the β-galactosidase enzyme assays and assisted with the writing of the published manuscript.

Published October 2018:

Cole, A., Hani, F., Allen, B., Kline, P., & Altman, E. (2018). Nonfunctional missense mutants in two well characterized cytosolic enzymes reveal important information about protein structure and function. *The Protein Journal, 37*(5), 407-427. doi:10.1007/s10930-018-9786-6

3.1 Abstract

The isolation and characterization of 42 unique nonfunctional missense mutants in the bacterial cytosolic β-galactosidase and catechol 2,3-dioxygenase enzymes allowed us to examine some of the basic general trends regarding protein structure and function. A total of 6 out of the 42, or 14.29% of the missense mutants were in α-helices, 17 out of the 42, or 40.48%, of the missense mutants were in β-sheets and 19 out of the 42, or 45.24% of the missense mutants were in unstructured coil, turn or loop regions. While α-helices and β-sheets are undeniably important in protein structure, our results clearly indicate that the unstructured regions are just as important. A total of 21 out of the 42, or 50.00% of the missense mutants caused either amino acids located on the surface of the protein to shift from hydrophilic to hydrophobic or buried amino acids to shift from hydrophobic to hydrophilic and resulted in drastic changes in hydropathy that would not be preferable. There was generally good consensus amongst the widely used algorithms, Chou-Fasman, GOR, Qian-Sejnowski, JPred, PSIPRED, Porter and SPIDER, in their ability to predict the presence of the secondary structures that were affected by the missense mutants and most of the algorithms predicted that the majority of the 42 inactive missense mutants would impact the α-helical and β-sheet secondary structures or the unstructured coil, turn or loop regions that they altered.

3.2 Introduction

Most proteins are intricate ordered structures dictated by the primary amino acid sequence of the protein that form α-helical or β-sheet secondary structures which then coalesce to yield the final stable tertiary structure. The early X-ray diffraction studies by Astbury's group in the 1930's suggested that proteins contain ordered structures (Astbury, William T. & Street, 1931; Astbury, William Thomas & Woods, 1933; Astbury, William Thomas & Sisson, 1935) and the two primary structures proposed by these studies were better defined to be α-helices and β-sheets by Pauling's group in 1951 (Pauling & Corey, 1951; Pauling, Corey, & Branson, 1951). Using more advanced and refined x-ray diffraction techniques the first α-helical structures were identified in myoglobin by Kendrew's group in 1958 (Kendrew et al., 1958) and in hemoglobin by Perutz's group in 1960 (Perutz et al., 1960) and the first β-sheet structure was identified in lysozyme by Phillips group in 1965 (Blake et al., 1965). The studies by (Anfinsen, 1973), then made it clear that the primary amino acid sequence must be sufficient to enable the final active tertiary structure of a protein to be formed, since denatured unfolded inactive proteins could be refolded into their active form when the denaturant was removed. As the structures of more and more proteins were solved, two general observations were made; 1- the undeniable importance of α-helices and β-sheets and 2- the tendency of amino acids on the surface of the protein to be hydrophilic and the tendency of buried amino acids in the protein to be hydrophobic.

In 1962 Tanford first proposed the ranking of amino acids based on their hydrophobicity (Tanford & Lovrien, 1962) and the first complete scale delineating hydrophobic versus hydrophilic amino acids was generated by Zimmerman and colleagues in 1968 (Zimmerman, Eliezer, & Simha, 1968). These scales then became increasingly more widely used in the 1980's and during this period the three most cited scales were developed, the Kyte-Doolittle scale (Kyte & Doolittle, 1982), the Hopp-Woods scale (Hopp & Woods, 1983) and the Eisenberg scale (Eisenberg, Schwarz, Komaromy, & Wall, 1984).While these scales can focus either on determining hydrophobicity or hydrophilicity, Kyte-Doolittle proposed the general term hydropathy, to reflect the hydrophobic or hydrophilic nature of the amino acids (Kyte & Doolittle, 1982). To date almost 100 different scales have been developed (Simm, Einloft, Mirus, & Schleiff, 2016).

With the accumulation of structures for multiple proteins by x-ray diffraction studies, it became increasingly clear that specific amino acids tended to be found in α-helices, β-sheets or the unstructured coil, turn and loop regions between the α-helices and β-sheets and researchers started to develop models based on proteins whose tertiary structure had been determined by x-ray diffraction studies that could be used to predict the presence of secondary structures in proteins whose structure had not yet been resolved. The first widely utilized secondary structure algorithm based on solved protein structures was the Chou-Fasman method (Chou & Fasman, 1974; Prevelige & Fasman, 1989), which was followed by the GOR method (Garnier,

Gibrat, & Robson, 1996). Although both the Chou-Fasman and GOR algorithms are still widely utilized, most modern approaches to predict secondary structures employ neural networks. The first neural network algorithm was developed by Qian and Sejnowski (Qian & Sejnowski, 1988a) and subsequent widely used algorithms have included JPred (Cuff, Clamp, Siddiqui, Finlay, & Barton, 1998; Drozdetskiy, Cole, Procter, & Barton, 2015) PSIPRED (Jones, 1999) Porter (Mirabello & Pollastri, 2013; Pollastri & McLysaght, 2005) and SPIDER2 (Heffernan et al., 2015a). The various algorithms developed to predict secondary structures in proteins have been compared in two general review articles (Pirovano & Heringa, 2010).

While missense mutants have been widely utilized to determine the importance of secondary structures in proteins, most of the studies have focused on mutations that impact a specific secondary structure deemed to be important (Conidi et al., 2013; Fredericks & Pielak, 1993; He, Wood, Baase, Xiao, & Matthews, 2004). We wanted to see what trends regarding protein structure could be garnered from a study of randomly generated missense mutants in two different well characterized cytosolic proteins. The bacterial tetrameric 465,912 Dalton β-galactosidase and 140,616 Dalton catechol 2,3-dioxygenase enzymes, coded by the *lacZ* and *xylE* genes, respectively, were chosen for this study because the tertiary structure of both proteins has been determined by x-ray diffraction analysis (Juers et al., 2000; Kita et al., 1999) and robust simple to use colorimetric enzyme assays have been developed to quantify the activities of the enzymes (Cole, Ashley E. et al., 2017; Miller, J. H., 1972).

3.3 Materials and Methods

3.3.1 Media and bacterial strains

Lysogeny broth (LB) (Bertani, 1951) was used as the rich media in this study. When required ampicillin was used at a final concentration of 100 µg/mL. To induce the *xylE* and *lacZ* genes, isopropyl β-D-thiogalactopyranoside (IPTG) was added at a concentration of 1 mM. The *lacZ* and *xylE* constructs used in this study were from the *Salmonella enterica* strains TT18519, *hisC10081*::MudF(*lac*+) and ALS1442, *proB*::*xylE*(*cat*), respectively (Cole, Ashley E. et al., 2017). The *Escherichia coli* strain MC1061, Δ*(araABOIC-leu)7679 araD139 hsr- hsm+ galU galK* Δ*(lac)X74 rpsL* (Casadaban & Cohen, 1980) was used to clone pUC57-*xylE* and pET-11a-*xylE* and the *E. coli* strain BL21(DE3) *fhuA2 lon ompT gal dcm* Δ*hsdS* λDE3 (Studier & Moffatt, 1986) was used to express the *xylE* gene from pET-11a-*xylE*.

3.3.2 Construction of pET-11a-*xylE*

Initially the coding sequence of the pWWO *xylE* gene (Harayama et al., 1991) was optimized for expression in *E. coli*, synthesized by GenScript, Piscataway, NJ, and cloned into the *EcoRV* restriction site of pUC57 to generate pUC57-*xylE*. Using pUC57-*xylE* as template DNA and *Pfu* polymerase, the *xylE* gene was amplified using the polymerase chain reaction (PCR) with the forward primer 5’ ATTGGTGTGGTTCATATGAATAAAGGCGTGATGCGTC 3’, which contained the *NdeI* restriction site and the reverse primer 5’ TTGTTGTTGGATCCTTAGGTCAGGACGGTCATAAAAC 3’, which contained

the *BamHI* restriction site. The homologous regions to the optimized *xylE* gene are underlined in the two primers. The resulting 953 bp fragment was restricted with *NdeI* and *BamHI* and cloned into the pET-11a expression vector (Studier & Moffatt, 1986) that had been restricted with the same two restriction enzymes.

3.3.3 Purification of catechol 2,3-dioxygenase protein from pET-11a-*xylE*

A 30$_o$C overnight LB ampicillin culture of BL21(DE3) pET-11a-*xylE* was used to inoculate 1 L of LB ampicillin broth at a 1:100 dilution. The culture was allowed to grow in a shaking 30$_o$C water bath to approximately 0.4 OD_{550} and then induced using 1 mM IPTG. The culture was grown for an additional 3 hours and the cells were harvested by centrifugation. The cell pellet was washed with 10 mM Tris: pH 7, pelleted by centrifugation, resuspended in 10 mL of B-Per Bacterial Protein Extraction Reagent (ThermoFischer Scientific) that contained Protease Inhibitor Cocktail (P8465) (Sigma Aldrich), gently mixed for 10 – 15 minutes and then centrifuged again. The soluble fraction which contained the catechol 2,3-dioxygenase enzyme was brought up to 100 mL using 10 mM Tris; pH 7, sequential 40% and 60% $(NH_4)_2SO_4$ precipitations were conducted and the pellet from the 60% $(NH_4)_2SO_4$ precipitation was resuspended in 5 mL of 10 mM Tris; pH 7. Prior to fast protein liquid chromatography (FPLC), a Zeba desalting column (ThermoFischer Scientific) was used to remove all of the salt. FPLC was conducted with a Sepharose MonoQ 16/10 sepharose column using a 0 – 1 M NaCl gradient created by 10 mM Tris; pH 6.5 and 10 mM Tris; pH 6.5, 1 M NaCl. The flow rate was 3.0 mL/min. Catechol 2,3 dioxygenase was found to elute at 370 – 470 mM NaCl. The fractions containing

catechol 2,3 dioxygenase were combined and concentrated using a Millipore Amicon Pro Affinity Concentrator column and then FPLC was conducted with a Bio-Rad CHT5-1 hydroxyapatite column using a 0.01 – 0.8 M KH_2PO_4 gradient created with 10 mM KH_2PO_4; pH 7.5 and 800 mM KH_2PO_4; pH 7.5. The flow rate was 3.0 mL/min. Catechol 2,3 dioxygenase was found to elute at 10 – 18 mM KH_2PO_4. Again, each fraction was combined and concentrated. The yield of catechol 2,3-dioxygenase enzyme from 1 L of cells was 3.03 mg and the purity was determined to be 98% by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and quantification with the Bio-Rad ChemiDoc MP Imaging System and Image Lab Software (Hercules, CA).

3.3.4 Generating and sequencing *lacZ* and *xylE* missense mutants

Inactive mutants in *lacZ* and *xylE* were generated using ethylmethane sulfonate (EMS) mutagenesis, screened using amber, ochre and opal nonsense suppressors to eliminate the nonsense mutants and then sequenced as described by (Cole, Ashley E. et al., 2017). In total, 20 independently isolated *lacZ* and 30 independently isolated *xylE* missense mutants were sequenced in this study. The 20 *lacZ* missense mutants yielded 11 distinct mutations in which a single nonrepetitive mutation occurred, while the 30 *xylE* missense mutants yielded 22 distinct mutations in which a single nonrepetitive mutation occurred. Four of the *xylE* missense mutants contained double mutants; i.e., changes in two different amino acids, and these were mutants were not used. To provide the complete collection of missense mutants shown in Table 3.3 and Table 3.4, we also included the missense mutants from the 10

lacZ and 10 *xylE* missense mutants isolated and sequenced in a previous study (Cole, Ashley E. et al., 2017) hat were unique from the missense mutants isolated in this study.

3.3.5 β-galactosidase and catechol 2,3-dioxygenase enzyme assays

The *β-galactosidase* and *catechol 2,3-dioxygenase* enzymatic assays were performed as previously described by (Cole, Ashley E. et al., 2017).

3.3.6 Preparation of rabbit anti β-galactosidase and catechol 2, 3-dioxygenase

Rabbit anti β-galactosidase and catechol 2, 3-dioxygenase antibodies were prepared by Cocalico Biologicals (Reamstown, PA) using their standard protocol. Pure β-galactosidase (G4155) was obtained from Sigma Aldrich and pure catechol 2, 3-dioxygenase was isolated as described above.

3.3.7 Western blot analysis

Each missense mutant as well as the wild-type controls were grown in LB plus 1 mM IPTG to an OD_{550} of approximately 0.5 and then 1 mL of cells were pelleted, resuspended in 0.1 mL of SDS-PAGE sample loading buffer and boiled. The total protein levels of each sample were quantified, and normalized amounts of each sample were run on a 4 - 20% gradient gel. The gel was transferred onto an activated nitrocellulose membrane and blocked with Tris buffered saline (TBS) (150 mM NaCl, 500 mM Tris; pH 7.5) that contained 5% nonfat-dried milk and 0.1% Tween for 30 min. The resulting western blots were washed with TBS, incubated

with TBS that contained rabbit anti β-galactosidase or catechol 2, 3-dioxygenase antibody at a 1:1000 dilution for one hour, washed with TBS and incubated in TBS that contained acid phosphatase conjugated goat anti rabbit antibody at a 1:1000 dilution for one hour. The blots were then washed with alkaline phosphatase buffer (100 mM NaCl, 5mM $MgCl_2$, 100 mM Tris; pH 9.5) and incubated with alkaline phosphatase buffer that contained 0.165 mg/mL 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) and 0.330 mg/mL nitro blue tetrazolium (NBT). When the β-galactosidase or catechol 2, 3-dioxygenase protein bands reached their maximum intensity, but before background bands started to appear, the color development was stopped using TBS that contained 2 mM ethylenediaminetetraacetic acid (EDTA) and then the blots were washed with distilled water and dried. Images were analyzed with the Bio-Rad ChemiDoc MP Imaging System and quantification of the proteins was determined with Image Lab Software (Bio-Rad, Hercules, CA).

3.3.8 Three-dimensional analysis of the *lacZ* and *xylE* missense mutants

The PyMOL Molecular Graphics System (Schrödinger, LLC) was used to determine the secondary structure affected by each mutant, whether the mutated amino acid was located on the surface, partially located on the surface, or buried and whether the mutated amino acid was located in the active site, substrate binding pocket or at a point of contact between monomers. The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) was the source of the three-dimensional crystal structure data. The 1DP0 file was used for the β-

galactosidase crystal structure data (Juers et al., 2000) and the 1MPY file was used for the catechol 2,3-dioxygenase crystal structure data (Kita et al., 1999).

3.3.9 Analysis of the potential effects of the *lacZ* and *xylE* missense mutants on secondary structure using amino acid propensity scales that predict the probabilities of amino acids to form α-helices, β-sheets or coils

Propensity scales that predict the probabilities of amino acids to be found in α-helices (P_α), β-sheets (P_β) or coils (P_c) based on the analysis of 2,216 proteins have been developed using three different criteria (Costantini, Colonna, & Facchiano, 2006) and we have generated an averaged propensity scale derived from these values (Table 3.1). The averaged propensity scale only deviated by 1.2% from the three individual scales presented in Costantini et al., 2006 and was in very good agreement with other thermodynamic studies that have predicted the propensity of amino acids to form α-helices (O'Neil & DeGrado, 1990; Pace & Scholtz, 1998) or β -sheets (Minor & Kim, 1994; Smith, Withka, & Regan, 1994) using $\Delta\Delta G$ values, as well as a survey study which analyzed the frequency of amino acids that were found in α-helices, β-sheets or coils from 1,590 proteins (Otaki, Tsutsumi, Gotoh, & Yamamoto, 2010). Each missense mutant was analyzed for how the amino acid change affected the P_α, P_β or P_c value of the wild-type amino acid.

3.3.10 Analysis of the changes in hydropathy caused by the *lacZ* and *xylE* missense mutants

Given the significant disagreement in the ultimate order of amino acids in hydrophobicity, hydrophilicity or hydropathy scales, we compiled the averaged rankings of the amino acids from 98 published scales (Simm et al., 2016) to generate a consensus hydropathy scale (Table 3.2). For each of the 98 scales, the amino acids were assigned a ranking of 1 for the most hydrophilic to 20 for the most hydrophobic and then the rankings from all of the scales were averaged to give a consensus ranking. The resulting averaged order from the most hydrophilic to the most hydrophobic amino acids was lys, asp, arg, gln, glu, asn, ser, pro, his, thr, gly, ala, tyr, met, trp, cys, val, leu, phe, ile with gly and thr being the middle amino acids, and was in very good agreement with the order of amino acids predicted by the more widely utilized Kyte-Doolittle, Hopp-Woods, and Eisenberg scales (Eisenberg et al., 1984; Hopp & Woods, 1983; Kyte & Doolittle, 1982).

3.3.11 Analysis of the potential effects of the missense mutants on the secondary structure of β-galactosidase or catechol 2, 3-dioxygenase using algorithms that predict secondary structures

The potential effects of each mutant on the secondary structure of β-galactosidase or catechol 2, 3-dioxygenase were analyzed using seven of the most widely utilized algorithms for predicting secondary structure. Analyses using the Chou-Fasman (Prevelige & Fasman, 1989), GOR IV (Garnier et al., 1996) and the Qian-Sejnowski (Qian & Sejnowski, 1988b) algorithms were performed using the

following website: http://cib.cf.ocha.ac.jp/bitool/MIX/. Analyses using the JPred4 (Drozdetskiy et al., 2015), PSIPRED (Jones, 1999), Porter 4.0 (Mirabello & Pollastri, 2013), and SPIDER2 (Heffernan et al., 2015b) algorithms were performed using the following websites, respectively: http://www.compbio.dundee.ac.uk/jpred/, http://bioinf.cs.ucl.ac.uk/psipred/, http://distillf.ucd.ie/porterpaleale/ and http://sparks-lab.org/yueyang/server/SPIDER2.

As discussed in the introduction, the more recent Qian-Sejnowski, JPred, PSIPRED, Porter and SPIDER algorithms employ neural networks, while the earliest developed algorithms, Chou-Fasman and GOR, do not. The Qian-Sejnowski, JPred, PSIPRED and Porter algorithms use an initial neural network calculation and then the resulting data is refined in a second neural network calculation, while the SPIDER algorithm uses three neural network calculations. With the exception of the Chou-Fasman algorithm, which uses very short analysis windows, six amino acids for α-helical regions and five amino acids for β-sheet regions (Chou & Fasman, 1974), the rest of the algorithms use much longer amino acid analysis windows, 17 amino acids for GOR (Garnier, Osguthorpe, & Robson, 1978b), 13 amino acids for Qian-Sejnowski (Qian & Sejnowski, 1988b), 17 amino acids for the first neural network calculation and 19 amino acids for the second neural network calculation in JPred (Cuff & Barton, 2000), 15 amino acids for Porter (Pollastri & McLysaght, 2005) and 17 amino acids for SPIDER (Heffernan et al., 2015b). The database of proteins used to predict the propensity of α-helical, β -sheet or unstructured regions in proteins by the different algorithms has increased over time. The most widely used version of the

Chou-Fasman algorithm uses 64 proteins (Prevelige & Fasman, 1989) while SPIDER2 uses 5,789 proteins (Heffernan et al., 2015b).

To accommodate the large analysis windows utilized by most of the algorithms, for each mutant to be analyzed an amino acid sequence that was at least 40 amino acids in length and contained at least 20 amino acids preceding and following the secondary structure in which the mutation occurred was used as the data input, with the caveat that the amino acids preceding and following the impacted secondary structure had to contain a complete secondary structure or random coil and could not end in a truncated secondary structure or random coil. For the *lacZ39* missense mutant where the mutation occurs at the carboxyl terminus of the protein, extra amino acids were added to the 5' end of the input sequence to achieve the minimal input sequence. For the *lacZ2456*, *lacZ2540*, *xylE1585* and *xylE1591* missense mutants where the mutation occurs at the amino terminus of the protein, extra amino acids were added to the 3' end of the input sequence to achieve the minimal input sequence.

3.4 Results

3.4.1 Generating a collection of inactive missense mutants in *lacZ* and *xylE*

As described in Materials and Methods a total of 30 independently isolated inactive *lacZ* missense mutants and 40 independently isolated inactive *xylE* missense mutants were sequenced to generate a collection of 15 unique inactive missense mutants in *lacZ* and 27 unique inactive missense mutants in *xylE*. Tables 3.3 and 3.4

list the *lacZ* and *xylE* mutants, respectively, and gives the base pair change, the corresponding amino acid change, what secondary structure is affected by the missense mutant and what type of impact the mutation is expected to have on the secondary structure according to α-helical, β-sheet and coil propensity scales, whether the mutated amino acid is located on the surface of the protein, partially located on the surface of the protein or buried, the change in hydropathy caused by the mutation, the enzymatic activity of the missense mutant and the stability of the missense mutant as determined by western analysis. A number of the *lacZ* and *xylE* missense mutants contained identical mutations. The mutation in *lacZ2382* occurred in six of the missense mutants and the mutation in *lacZ2449* occurred in four of the missense mutants and potentially are mutations that are easy to isolate because they cause devastating effects to the protein structure. All of the missense mutants had negligible activity compared to the wild-type protein. A total of 35 out of the 42, or 83.33% of the missense mutants produced a stable protein (10 out of 15, or 66.67% for *lacZ* and 25 out of 27, or 92.59% for *xylE*). A total of 7 out of the 42, or 16.67% of the missense mutants produced proteins that were largely degraded with around 90% or higher degradation (5 out of 15, or 33.33% for *lacZ* and 2 out of 27, or 7.41% for *xylE*). Figure 3.1 shows the western blot results of representative *lacZ* missense mutants, *lacZ2234*, *lacZ2343*, *lacZ2381* and *lacZ2449*. *lacZ2381* is the only missense mutant that produced a proteolytic fragment. None of the missense mutants affected amino acids that are known to constitute the active site or substrate binding pocket of either the β-galactosidase or catechol 2,3-dioxygenase enzymes and none of the missense mutants were located at contact points between the monomers of these

proteins that coalesce to form the tetrameric structures. Thus, all of the missense mutants are expected to negatively impact the structure of the proteins.

3.4.2 Analysis of the missense mutants with respect to the location of the mutation within the protein versus the change in hydropathy

According to the consensus hydropathy scale (Table 3.2), 28 out of the 42, or 66.67%, of the missense mutants caused a gain in hydrophilicity (11 out of 15, or 73.33% for *lacZ* and 17 out of 27, or 62.96% for *xylE*). In general, for globular proteins amino acids on the surface tend to be hydrophilic, while buried amino acids tend to be hydrophobic, and one might expect disruptive missense mutants to cause a hydrophilic to hydrophobic shift for amino acids located on the surface and a hydrophobic to hydrophilic shift for buried amino acids. The missense mutants located on the surface did not follow this expectation as only 9 out of 25, or 36.00% of the missense mutants located on the surface caused a hydrophobic shift (4 out of 8, or 50.00 % for *lacZ* and 5 out of 17, or 29.41% for *xylE*), while the missense mutants that were buried did follow this expectation as 12 out of the 17, or 70.59% of the missense mutants that were buried caused a hydrophilic shift (7 out of 7, or 100.00% for *lacZ* and 5 out of 10, or 50.00% for *xylE*). In total, 21 out of 42, or 50.00% of the missense mutants caused hydrophobic or hydrophilic shifts that would be expected to be detrimental based on their location (11 out of 15, or 73.33%, for *lacZ* and 10 out of 27, or 37.04% for *xylE*).

3.4.3 Analysis of the distribution of the missense mutants in α-helical, β-sheet or coil secondary structures

A total of 6 out of the 42, or 14.29% of the missense mutants were in α-helices (2 out of 15, or 13.33% for *lacZ* and 4 out of 27, or 14.81% for *xylE*), 17 out of the 42, or 40.48%, of the missense mutants were in β-sheets (7 out of 15, or 46.67% for *lacZ* and 10 out of 27, or 37.04%, for *xylE*) and 19 out of the 42, or 45.24% of the missense mutants were in random coils (6 out of 15, or 40.00% for *lacZ* and 13 out of 27, or 48.15% for *xylE*). The distribution of amino acids that are in α-helices, β-sheets or random coils for the two proteins is 18.74%, 36.42% and 44.85%, respectively (18.51% α-helices, 38.49% β-sheets and 43.00% random coils for *lacZ* and 19.48% α-helices, 29.55% β-sheets and 50.97% random coils for *xylE*). Thus, the distribution of missense mutations in α-helices, β-sheets and random coils was very much in line with the percentages in which these structures occurred in the two proteins.

3.4.4 Analysis of the predicted effect of the missense mutants on secondary structures using α-helical, β-sheet or coil propensity scales

According to the averaged α-helical, β-sheet or coil propensity values (Table 3.1), 19 out of 42, or 45.24% of the missense mutants affected α-helical, β-sheet or random coil secondary structures in which the original amino acid was a preferred amino acid for that structure (5 out of 15, or 33.33% for *lacZ* and 14 out of 27, or 51.85 % for *xylE*). A total of 26 out of the 42, or 61.90% of the missense mutants were expected to negatively impact the secondary structures (11 out of 15, or 73.33%

for *lacZ* and 15 out of 27, or 55.56% for *xylE*). A total of 3 out of the 6, or 50.00% of the missense mutants in α-helices were expected to cause an unfavorable change (1 out of 2, or 50.00% for *lacZ* and 2 out of 4, or 50.00% for *xylE*), 10 out of the 17, or 58.82% of the missense mutants in β-sheets were expected to cause an unfavorable change (6 out of 7, or 85.71%, for *lacZ* and 4 out of 10, or 40.00% for *xylE*) and 13 out of the 19, or 68.42% of the missense mutants in random coils were expected to cause an unfavorable change (4 out of 6, or 66.67% for *lacZ* and 9 out of 13, or 69.23% for *xylE*).

3.4.5 Analysis of the predicted effect of the missense mutants on secondary structures using algorithms that predict the presence of secondary structure

We analyzed the predicted impact of each missense mutant on the secondary structures using seven of the most widely used algorithms for predicting secondary structure, the older Chou-Fasman, GOR and Qian-Sejnowski algorithms and the newer JPred, PSIPRED, Porter and SPIDER algorithms. Based on the changes of the averaged propensity scores (Chou-Fasman, GOR IV and Qian-Sejnowski) or the consensus scores (JPred4, PSIPRED, Porter 4.0 and SPIDER2) between the affected secondary structure in the parent versus the missense mutant, the percentage negative or positive impact caused by the mutant was determined. As an example, for the *lacZ39* mutant the averaged Chou-Fasman propensity score was 103.67 over the RYHYQ**L**VWC parent sequence and 101.67 over the RYHYQ**Q**VWC mutant sequence and thus the mutant is predicted to cause a -1.93% change according to Chou-Fasman analysis. Tables 5 and 6 lists the results of this analysis for the *lacZ*

missense mutants and Tables 7 and 8 lists the results of this analysis for the *xylE* missense mutants. Both Chou-Fasman and GOR IV analysis predicted that 42 out of 42, or 100.00% of the missense mutants had an impact on the secondary structure (15 out of 15, or 100.00% for *lacZ* and 27 out of 27, or 100.00% for *xylE*), Qian-Sejnowski analysis predicted that 41 out of 42, or 97.62% of the missense mutants had an impact on the secondary structure (15 out of 15, or 100.00% for *lacZ* and 26 out of 27, or 96.30% for *xylE*), both JPred4 and PSIPRED analysis predicted that 34 out of 42, or 80.95% of the missense mutants had an impact on the secondary structure (12 out of 15, or 80.00% for *lacZ* and 22 out of 27, or 81.48% for *xylE*), Porter 4.0 analysis predicted that 27 out of 42, or 64.29%, of the missense mutants had an impact on the secondary structure (9 out of 15, or 60.00% for *lacZ* and 18 out of 27, or 66.67% for *xylE*) and SPIDER2 analysis predicted that 41 out of 42, or 97.62% of the missense mutants had an impact on the secondary structure (14 out of 15, or 93.33% for *lacZ* and 27 out of 27, or 100.00% for *xylE*).

Interestingly, when an algorithm predicted a change in the secondary structure between the parent and the mutant, the impact was almost as likely to be positive as it was to be negative. Chou-Fasman analysis predicted that 25 out of the 42, or 59.52% of the missense mutants that affected the secondary structure had a negative impact (10 out of 15, or 66.67% for *lacZ* and 15 out of 27, or 55.56% for *xylE*), GOR IV analysis predicted that 24 out of the 42, or 57.14% of the missense mutants that affected the secondary structure had a negative impact (10 out of 15, or 66.67% for *lacZ* and 14 out of 27, or 51.85% for *xylE*), Qian-Sejnowski analysis predicted that 29 out of the 41, or 70.73% of the missense mutants that affected the secondary structure

had a negative impact (12 out of 15, or 80.00% for *lacZ* and 17 out of 26, or 65.38% for *xylE*), JPred4 analysis predicted that 21 out of the 34, or 61.76% of the missense mutants that affected the secondary structure had a negative impact (6 out of 12, or 50.00% for *lacZ* and 15 out of 22, or 68.18% for *xylE*), PSIPRED analysis predicted that 17 out of the 34, or 50.00% of the missense mutants that affected the secondary structure had a negative impact (7 out of 12, or 58.33% for *lacZ* and 10 out of 22, or 45.45% for *xylE*), Porter 4.0 analysis predicted that 14 out of the 27, or 51.85% of the missense mutants that affected the secondary structure had a negative impact (4 out of 9, or 44.44% for *lacZ* and 10 out of 18, or 55.56% for *xylE*) and SPIDER2 analysis predicted that 16 out of the 41, or 39.02% of the missense mutants that affected the secondary structure had a negative impact (6 out of 14, or 42.86% for *lacZ* and 10 out of 27, or 37.04% for *xylE*).

Since the missense mutants were nonfunctional one might have expected that the algorithms for predicting secondary structure would have predicted that the missense mutants would have a drastic impact on secondary structure and thus significantly alter the propensity or consensus scores. However, this was not the case for most of the algorithms. Chou-Fasman analysis predicted that 2 out of the 42, or 4.76% of the missense mutants would cause a 10% difference in the propensity scores (1 out of 15, or 6.67% for *lacZ* and 1 out of 27, or 3.70% for *xylE*), GOR IV analysis predicted that 37 out of the 42, or 88.10% of the missense mutants would cause a 10% difference in the propensity scores (14 out of 15, or 93.33% for *lacZ* and 23 out of 27, or 85.19% for *xylE*), Qian-Sejnowski analysis predicted that 25 out of the 42, or 59.52% of the missense mutants would cause a 10% difference in the propensity

scores (11 out of 15, or 73.33% for *lacZ* and 14 out of 27, or 51.85% for *xylE*), JPred4 analysis predicted that 13 out of the 42, or 30.95% of the missense mutants would cause a 10% difference in the consensus scores (6 out of 15, or 40.00% for *lacZ* and 7 out of 27, or 25.93% for *xylE*), PSIPRED analysis predicted that 13 out of the 42, or 30.95% of the missense mutants would cause a 10% difference in the consensus scores (7 out of 15, or 46.67% for *lacZ* and 6 out of 27, or 22.22% for *xylE*), Porter 4.0 analysis predicted that 9 out of the 42, or 21.43% of the missense mutants would cause a 10% difference in the consensus scores (5 out of 15, or 33.33% for *lacZ* and 4 out of 27, or 14.81% for *xylE*) and SPIDER2 analysis predicted that 10 out of the 42, or 23.81% of the missense mutants would cause a 10% difference in the consensus scores (3 out of 15, or 20.00% for *lacZ* and 7 out of 27, or 25.93% for *xylE*).

It was somewhat surprising that there was not more of a general consensus amongst the algorithms in predicting that the missense mutants would significantly impact the secondary structure. For only 12 out of the 42, or 28.57% of the missense mutants did the majority of four out of the seven algorithms predict that the missense mutation would have a significant impact on the secondary structure and cause at least a 10% difference in the propensity or consensus scores (6 out of 15, or 40.00% for *lacZ* and 6 out of 27, or 22.22% for *xylE*). It is worth noting that 4 out of the 7, or 57.14% of the most unstable missense mutants, as judged by the stability of the proteins that they produced, were amongst these mutants (3 out of 5, or 60.00% for *lacZ* and 1 out of 2, or 50.00% for *xylE*).

Given the fact that inactive missense mutants might be predicted to have a significant impact on protein structure and the older algorithms made this prediction more often than the newer algorithms, we analyzed the ability of the algorithms to predict the affected structure in the wild-type protein for the collection of mutants that were characterized in this study to verify that the accuracy of the algorithms in predicting the secondary structures impacted by the missense mutants characterized in this study was consistent with other studies that have calculated the ability of the algorithms to correctly predict protein structure. We first analyzed the ability of the predictive algorithms to accurately predict the secondary structures in which the missense mutants were located. In this analysis, if a predictive algorithm identified at least 50% of the amino acids as being in a specific structure correctly, it was deemed as accurately predicting the secondary structure. For comparison purposes, we conducted this analysis for both the wild-type and mutant structures. As an example, for the *lacZ39* mutant Chou-Fasman analysis predicted that seven out of the nine, or 77.8% of the amino acids in both the RYHYQ**L**VWC wild-type sequence and the RYHYQ**Q**VWC mutant sequence would form the β-sheet structure confirmed by x-ray diffraction studies. Tables 9 and 10 show the results of this analysis for the *lacZ* missense mutants and Tables 11 and 12 show the results of this analysis for the *xylE* missense mutants. Chou-Fasman identified 24 out of the 42, or 57.14% of the secondary structures correctly (9 out of 15, or 60.00% for *lacZ* and 15 out of 27, or 55.56% for *xylE*), GOR IV identified 19 out of the 42, or 45.24% of the secondary structures correctly (9 out of 15, or 60.00% for *lacZ* and 10 out of 27, or 37.04% for *xylE*), Qian-Sejnowski identified 19 out of the 42, or 45.24% of the secondary

structures correctly (5 out of 15, or 33.33% for *lacZ* and 14 out of 27, or 51.85% for *xylE*), JPred4 identified 34 out of the 42, or 80.95% of the secondary structures correctly (9 out of 15, or 60.00% for *lacZ* and 25 out of 27, or 92.59% for *xylE*), PSIPRED identified 33 out of the 42, or 78.57% of the secondary structures correctly (8 out of 15, or 53.33% for *lacZ* and 25 out of 27, or 92.59% for *xylE*), Porter 4.0 identified 35 out of the 42, or 83.33% of the secondary structures correctly (9 out of 15, or 60.00% for *lacZ* and 26 out of 27, or 96.30% for *xylE*) and SPIDER2 identified 35 out of the 42, or 83.33% of the secondary structures correctly (10 out of 15, or 66.67% for *lacZ* and 25 out of 27, or 92.59% for *xylE*).

In a more detailed analysis we analyzed the ability of the predictive algorithms to make all of the calls correctly. In this analysis, the ability of each algorithm to correctly determine that each amino acid was identified as being in a specific secondary structure was analyzed; i.e., in a 9-amino acid β-sheet, were all 9 amino acids identified as being in a β-sheet. In total, the structures affected by the missense mutants contained 421 amino acids. Chou-Fasman identified 205 out of the 421, or 48.69% of the amino acids correctly (59 out of 113, or 52.21% for *lacZ* and 146 out of 308, or 47.40% for *xylE*), GOR IV identified 153 out of the 421, or 36.34% of the amino acids correctly (54 out of 113, or 47.79% for *lacZ* and 99 out of 308, or 32.14% for *xylE*), Qian-Sejnowski identified 200 out of the 421, or 47.51% of the amino acids correctly (38 out of 113, or 33.63% for *lacZ* and 162 out of 308, or 52.60% for *xylE*), JPred4 identified 323 out of the 421, or 76.72% of the amino acids correctly (72 out of 113, or 63.72% for *lacZ* and 251 out of 308, or 81.49% for *xylE*), PSIPRED identified 316 out of the 421, or 75.06% of the amino acids correctly (69

out of 113, or 61.06% for *lacZ* and 247 out of 308, or 80.19% for *xylE*), Porter 4.0 identified 327 out of the 421, or 77.67% of the amino acids correctly (70 out of 113, or 61.95% for *lacZ* and 257 out of 308, or 83.44% for *xylE*) and SPIDER2 identified 317 out of the 421, or 75.30% of the amino acids correctly (72 out of 113, or 63.72% for *lacZ* and 245 out of 308, or 79.55% for *xylE*).

We also analyzed whether the mutants were predicted to cause a change in the secondary structure using the data from Tables 3.9 – 3.12. In this analysis, any mutant that caused a change in the presence of the secondary structure between the wild-type and mutant was noted. As an example, for the *lacZ39* mutant Chou-Fasman analysis predicted that seven out of the nine, or 77.8% of the amino acids in both the RYHYQ**L**VWC wild-type sequence and the RYHYQ**Q**VWC mutant sequence formed a β-sheet and thus the mutant would not be predicted to cause a change in the secondary structure. Chou-Fasman predicted that 17 out of the 42, or 40.48% of the mutants would impact the secondary structure (5 out of 15, or 33.33% for *lacZ* and 12 out of 27, or 44.44% for *xylE*), GOR IV predicted that 27 out of the 42, or 64.29% of the mutants would impact the secondary structure (9 out of 15, or 60.00% for *lacZ* and 18 out of 27, or 66.67% for *xylE*), Qian-Sejnowski predicted that 18 out of the 42, or 42.86% of the mutants would impact the secondary structure (6 out of 15, or 40.00% for *lacZ* and 12 out of 27, or 44.44% for *xylE*), JPred4 predicted that 13 out of the 42, or 30.95% of the mutants would impact the secondary structure (6 out of 15, or 40.00% for *lacZ* and 7 out of 27, or 25.93% for *xylE*), PSIPRED predicted that 9 out of the 42, or 21.43% of the mutants would impact the secondary structure (1 out of 15, or 6.67% for *lacZ* and 8 out of 27, or 29.63% for *xylE*), Porter

4.0 predicted that 9 out of the 42, or 21.43% of the mutants would impact the secondary structure (4 out of 15, or 26.67% for *lacZ* and 5 out of 27, or 18.52% for *xylE*) and SPIDER2 predicted that 16 out of the 42, or 38.10% of the mutants would impact the secondary structure (5 out of 15, or 33.33% for *lacZ* and 11 out of 27, or 40.74% for *xylE*).

3.5 Discussion

We generated a collection of 42 unique inactive missense mutants in the aqueous cytosolic tetrameric 465,912 Dalton β-galactosidase and 140,616 Dalton catechol 2,3-dioxygenase enzymes, which are coded by the *lacZ* and *xylE* genes, respectively. As expected, all of the missense mutants had negligible enzymatic activity and 35 out of the 42, or 83.33% of the missense mutants produced a stable protein as determined by western analysis. Interestingly, 7 out of the 42, or 16.67% of the missense mutants did not produce a stable protein and would most likely have been missed on the basis of the presence of cross reacting material or CRM, which has historically been utilized to identify missense mutants in large mutant collections (Greeb et al., 1971; Jacobson, Zhang, DuBose, & Matthews, 1994; Truman & Bergquist, 1976; Whitfield et al., 1966). The characterization of the 42 unique missense mutants in this study revealed some interesting findings with respect to what types of mutations cause dramatic changes in the structures of aqueous proteins. None of the missense mutants were located at the active site or substrate binding pockets in β-galactosidase or catechol 2,3-dioxygenase and none of the missense mutants were located at contact sites where the monomeric proteins coalesce to form

the final active tetrameric proteins. Thus, all of the missense mutants isolated in this study were expected to cause significant structural perturbations that rendered the missense mutants nonfunctional.

While the study of proteins whose three-dimensional structures had been determined reaffirmed the importance of the organized α-helical and β-sheet secondary structures, another trend that was discovered was the tendency of the internal or buried portion of the protein to favor hydrophobic amino acids and the surface portion of the protein to favor hydrophilic amino acids. Thus, one might have expected some of the missense mutants to cause a hydrophilic to hydrophobic shift if the wild-type amino acid was located on the surface and a hydrophobic to hydrophilic shift if the wild-type amino acid was buried in the protein. A total of 21 out of the 42, or 50.00% of the missense mutants caused the expected hydrophobic to hydrophilic shift based on their location in the protein.

A total of 6 out of the 42, or 14.29% of the missense mutants were in α-helices, 17 out of the 42, or 40.48% of the missense mutants were in β-sheets and 19 out of the 42, or 45.24% of the missense mutants were in unstructured "coil", "turn" or "loop" regions of the protein. While α-helices and β-sheets are undeniably important to the structure of the protein, our results make it clear that mutations in unstructured regions of the protein are just as likely to have devastating consequences. In fact, the overall frequency of the types of missense mutations that we observed in this study based on the structures affected was very close to the overall frequency of distribution of the structures in the β-galactosidase and catechol 2,3-dioxygenase proteins. Had nonfunctional missense mutants been more likely to

occur in α-helices and β-sheets then the frequency of the distribution of the missense mutants should have favored the organized structures and this was not the case.

The significance of unstructured or disordered regions that contain no ordered structure in proteins has become increasingly apparent in recent years and the terms intrinsically disordered proteins (IDPs) and IDP regions or intrinsically disordered regions (IDRs) have been introduced to account for their importance in protein function (53, 54, 55). While the β-galactosidase and catechol 2,3-dioxygenase enzymes that we characterized in this study are clearly proteins that have an overall ordered structure according to x-ray diffraction analysis, both proteins contain a significant amount of unstructured or disordered regions. A total of 43% of the β-galactosidase protein contains unstructured regions while 50.97% of the catechol 2,3-dioxygenase protein contains unstructured regions. Our observation that 45.24% of the non-functional missense mutants that we characterized in this study were in unstructured or disordered regions supports the importance of IDRs.

The predicted impact of the missense mutants on the secondary structures was almost as likely to be positive and result in an improved secondary structure as they were to be negative or disruptive, whether propensity scales that indicate the likelihood of amino acids to be found in α-helical, β-sheet or coil structures or predictive algorithms that predict the presence of secondary structures were utilized. While the seven most popular algorithms for predicting secondary structures gave varied results as to the effect of the missense mutants, there was a consensus that most of the missense mutants would be predicted to affect the secondary structures. Based on changes in the propensity or consensus scores generated by the algorithms,

both Chou-Fasman and GOR IV analysis predicted that 42 out of 42, or 100.00% of the missense mutants had an impact on the secondary structure, both Qian-Sejnowski and SPIDER2 analysis predicted that 41 out of 42, or 97.62% of the missense mutants had an impact on the secondary structure, both JPred4 and PSIPRED analysis predicted that 34 out of 42, or 80.95% of the missense mutants had an impact on the secondary structure and Porter 4.0 analysis predicted that 27 out of 42, or 64.29%, of the missense mutants had an impact on the secondary structure.

Given the fact that all of the missense mutants were nonfunctional and had minimal enzymatic activities and thus might be expected to dramatically alter the secondary structure, it was somewhat surprising that the majority of the algorithms predicted that most of the mutants would only cause subtle changes, whether this was assessed by significant changes in the propensity or consensus scores generated by the algorithms (Tables 3.5 – 3.8) or the ability of the algorithms to predict that the mutant would cause a change in the secondary structure (Tables 3.9 – 3.12). In an analysis to determine whether the algorithms caused a significant change in the propensity or consensus scores Chou-Fasman analysis predicted 2 out of the 42, or 4.76% of the missense mutants would cause a 10% difference in the propensity scores, GOR IV analysis predicted 37 out of the 42, or 88.10% of the missense mutants would cause a 10% difference in the propensity scores, Qian-Sejnowski analysis predicted 25 out of the 42, or 59.52% of the missense mutants would cause a 10% difference in the propensity scores, JPred4 analysis predicted 13 out of the 42, or 30.95% of the missense mutants would cause a 10% difference in the consensus scores, PSIPRED analysis predicted 13 out of the 42, or 30.95% of the missense

mutants would cause a 10% difference in the consensus scores, Porter 4.0 analysis predicted 9 out of the 42, or 21.43% of the missense mutants would cause a 10% difference in the consensus and SPIDER2 analysis predicted 10 out of the 42, or 23.81% of the missense mutants would cause a 10% difference in the consensus scores. In an analysis to determine whether the algorithms predicted that the mutant would cause a change to the secondary structure, Chou-Fasman predicted 17 out of the 42, or 40.48% of the mutants would impact the secondary structure, GOR IV predicted 27 out of the 42, or 64.29% of the mutants would impact the secondary structure, Qian-Sejnowski predicted 18 out of the 42, or 42.86% of the mutants would impact the secondary structure, JPred4 predicted 13 out of the 42, or 30.95% of the mutants would impact the secondary structure, PSIPRED predicted 9 out of the 42, or 21.43% of the mutants would impact the secondary structure, Porter 4.0 predicted 9 out of the 42, or 21.43% of the mutants would impact the secondary structure and SPIDER2 predicted 16 out of the 42, or 38.10% of the mutants would impact the secondary structure. It should be noted that the subtle impacts of the missense mutants, which were predicted by most of the algorithms, might be expected given the intricacy of folded proteins. Most of the secondary structures in proteins are far from optimal according to the predictive algorithms that have been studied and subtle changes that positively or negatively impact a secondary structure could be devastating.

Since the newer JPred4, PSIPRED, Porter 4.0 and SPIDER2 models are thought to be more accurate than the older Chou-Fasman, GOR IV and Qian-Sejnowski models (Pavlopoulou & Michalopoulos, 2011; Pirovano & Heringa, 2010)

it was puzzling that the older models were more likely to predict that the inactive missense mutants would have an impact on the secondary structure than the newer models. For this reason, we analyzed the accuracy of the different models to predict the presence of the secondary structures that were impacted by the missense mutants. Porter 4.0 correctly identified 77.67% of all of the amino acids in the secondary structures affected by the missense mutants correctly, followed by JPred4 at 76.72%, SPIDER2 at 75.30%, PSIPRED at 75.06%, Chou-Fasman at 48.69%, Qian-Sejnowski at 47.51% and GOR IV at 36.34%. Thus, the newer models were significantly more accurate at identifying the secondary structures than the older models. The creators of the various algorithms have all calculated the accuracies of their models; Chou-Fasman, 71.0% (Prevelige & Fasman, 1989), GOR IV, 64.4% (Garnier et al., 1996), Qian and Sejnowski, 64.3% (Qian & Sejnowski, 1988b), JPred4, 82.0% (Drozdetskiy et al., 2015), PSIPRED, 76.5 to 78.3% (Jones, 1999), Porter 4.0, 82.2% (Mirabello & Pollastri, 2013) and SPIDER2, 82.0% (Heffernan et al., 2015a).

Compliance with Ethical Standards

Conflict of interest: The authors declare that they have no conflict of interest.

Table 3.1. Averaged propensity scale values for the likelyhood of amino acids to be found in the three secondary structures, α-helices (P_α), β-sheets (P_β) and coils (P_c)

Amino Acid	P_α	P_β	P_c
Ala	**1.38**	0.75	0.80
Cys	0.77	**1.34**	1.01
Asp	0.89	0.55	**1.35**
Glu	**1.34**	0.72	0.85
Phe	1.01	**1.42**	0.76
Gly	0.48	0.67	**1.65**
His	0.92	0.99	**1.26**
Ile	1.04	**1.70**	0.58
Lys	**1.11**	0.82	1.00
Leu	**1.31**	1.08	0.68
Met	**1.20**	0.98	0.83
Asn	0.77	0.62	**1.41**
Pro	0.50	0.44	**1.76**
Gln	**1.29**	0.76	0.88
Arg	**1.17**	0.90	0.90
Ser	0.82	0.85	**1.25**
Thr	0.77	**1.24**	1.07
Val	0.88	**1.85**	0.63
Trp	1.05	**1.29**	0.79
Tyr	0.95	**1.49**	0.78

Bold numbers indicate the secondary structure in which each amino acid is most likely to be found. The generation of the values is described in experimental procedures.

Table 3.2. Consensus hydropathy scale

Amino Acid	Average Consensus Ranking
Lys	5.46
Asp	5.47
Arg	5.70
Gln	6.08
Glu	6.11
Asn	6.36
Ser	8.18
Pro	8.93
His	9.04
Thr	9.45
Gly	10.16
Ala	11.43
Tyr	11.93
Met	13.93
Trp	14.39
Cys	14.65
Val	15.25
Leu	15.37
Phe	15.78
Ile	16.31

As described in Materials and Methods the rankings of amino acids from 98 hydrophobicity, hydrophilicity or hydropathy scales were averaged to generate the consensus hydropathy scale shown above where lys is the most hydrophilic amino acid and ile is the most hydrophobic amino acid

Table 3.3. *lacZ* missense mutants

***lacZ* mutant number**[1]	**Base pair change**	**Amino acid change**[2]	**Structural perturbation**[3]	**Mutation favorability**[4]	**Protein location**[5]	**Change in hydropathy**[6]	**Mutant activity**[7]	**Protein stability**[8]
39*	3053 T to A	1018 Leu to Gln	Beta Sheet	-	Surface (Partial)	+*	0.30%	2.1%
337	2713 G to A	905 Glu to Lys	Coil	+	Surface (Partial)	+	0.08%	91.5%
361*	2698 G to A	900 Gly to Asp	Beta Sheet	-	Buried	+*	0.04%	123.0%
364*	624 G to A	208 Gly to Asp	Beta Sheet	-	Buried	+*	0.13%	142.0%
2234*	1694 G to A	565 Gly to Asp	Beta Sheet	-	Buried	+*	0.12%	121.1%
2343*	2828 G to A	943 Arg to His	Beta Sheet	+	Surface	-*	0.31%	11.4%
2381	2291 G to A	764 Gly to Asp	Coil	-	Buried	+*	0.22%	34.6%
2382*	1697 G to A	566 Gly to Asp	Beta Sheet	-	Buried	+*	0.12%	79.9%
2396	1643 G to A	584 Gly to Asp	Coil	-	Surface	+*	0.14%	71.2%
2449*	604 G to A	202 Asp to Asn	Coil	+	Surface (Partial)	-	0.40%	0.1%
2454*	1061 G to A	354 Gly to Asp	Beta Sheet	-	Buried	+*	0.12%	102.1%
2456	14 C to T	5 Thr to Met	Alpha Helix	+*	Surface	-	0.48%	0.9%
2530	1255 C to T	419 His to Tyr	Coil	-*	Surface	-	2.60%	88.8%
2540	11 T to A	4 Ile to Asn	Alpha Helix	-	Surface	+*	0.32%	2.5%
2608	1378 G to A	460 Gly to Arg	Coil	-*	Buried	+*	0.03%	99.0%

1. Only unique mutations are listed in the table. The *lacZ39* mutation was observed in 2 of the missense mutants, the *lacZ361* mutation was observed in 3 of the missense mutants, the *lacZ364* mutation was observed in 2 of the missense mutants, the *lacZ2234* mutation was observed in 2 of the missense mutants, the *lacZ2343* mutation was observed in 2 of the missense mutants, the *lacZ2382* mutation was observed in 6 of the missense mutants, the *lacZ2449* mutation was observed in 4 of the missense mutants, and the *lacZ2454* mutation was observed in 3 of the missense mutants.
2. The *lacZ* gene is 3,075 bp in length and codes for the 1,024 amino acid β-galactosidase protein. The resulting amino acid changes are given based on the coded protein predicted by the DNA sequence and not the Protein Data Bank file.

Table 3.3. *lacZ* missense mutants (continued)

3. The secondary structure affected by the mutation was determined using PyMol. Data from Jacobson *et al*. 1994 had to be used to determine the secondary structures affected by *lacZ2456* and *lacZ2540*, since these mutations were at the extreme amino terminus.
4. The likelihood of a mutation to affect the α-helical, β-sheet or random coil structure was determined using P_α, P_β or P_c values from the averaged propensity scale in Table 1. (+) indicates a favorable change and (–) indicates an unfavorable change. An asterisk indicates a mutation that changes the original amino acid from or to one of the preferred amino acids that are found in α-helices, β-sheets or random coils.
5. The location of the mutated amino acids was determined using PyMOL.
6. The change in hydropathy was determined using the averaged hydropathy scale in Table 2. (+) indicates a more hydrophilic change, (–) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydropathy of the original amino acid significantly and results in a shift of at least 5 amino acids.
7. Mutant activities are listed with respect to the wild-type activity (100%). The assays were repeated in triplicate and the standard deviation was less than 10%. Wild type *lacZ* has an activity of 3,167.43 Miller units.
8. Protein stability was determined by western blot analysis. The stability of the mutants are listed with respect to wild-type protein levels (100%). Mutant stability was calculated by comparing the amount of mutant protein that was present to the level of wild-type β-galactosidase protein in the *lacZ+* parent strain. *lacZ2381* produced a proteolytic fragment as seen in Figure 1.

Table 3.4. *xylE* missense mutants

***xylE* mutant number**[1]	**Base pair change**	**Amino acid change**[2]	**Structural perturbation**[3]	**Mutation favorability**[4]	**Protein location**[5]	**Change in hydropathy**[6]	**Mutant activity**[7]	**Protein stability**[8]
1568	316 C to T	106 Arg to Cys	Beta Sheet	+*	Surface (Partial)	-*	2.04%	22.6%
1569	809 G to A	270 Gly to Gln	Coil	-*	Surface (Partial)	+*	4.54%	68.3%
1571*	380 G to A	127 Gly to Gln	Coil	-*	Surface (Partial)	+*	5.86%	50.9%
1574*	698 C to T	233 Ser to Phe	Alpha Helix	+	Surface	-*	2.52%	46.4%
1577	62 C to T	21 Ala to Val	Alpha Helix	-*	Buried	-*	1.83%	45.7%
1581*	88 G to A	30 Gly to Ser	Coil	-	Surface (Partial)	+	2.52%	49.6%
1582	191 G to A	64 Gly to Asp	Beta Sheet	-	Surface	+*	5.69%	63.6%
1583*	473 G to A	158 Gly to Asp	Beta Sheet	-	Buried	+*	5.70%	46.2%
1584	727 C to T	243 Pro to Ser	Coil	-	Surface (Partial)	+	4.27%	31.8%
1585*	25 G to A	9 Gly to Ser	Coil	-	Buried	+	5.20%	8.3%
1587	532 G to A	178 Gln to Lys	Beta Sheet	+	Surface	+	4.42%	33.6%
1588	740 G to A	247 Gly to Asp	Coil	-	Surface (Partial)	+*	5.24%	47.7%
1589*	845 G to C	282 Trp to Ser	Beta Sheet	-*	Surface (Partial)	+*	5.56%	23.6%
1591	22 C to T	8 Pro to Ser	Beta Sheet	+	Buried	+	4.76%	20.3%
1592	578 G to A	193 Ser to Asn	Beta Sheet	-	Buried	+	2.93%	46.2%
1593	733 C to T	245 Arg to Cys	Coil	+	Surface (Partial)	-*	3.03%	93.7%
1665	725 G to A	242 Gly to Asp	Coil	-	Surface (Partial)	+*	1.35%	86.6%
1666	752 G to A	251 Gly to Asp	Coil	-	Surface (Partial)	+*	4.93%	93.9%
1670*	604 G to A	202 Ala to Thr	Beta Sheet	+*	Buried	+	2.73%	35.5%
1677	637 C to T	213 His to Tyr	Beta Sheet	+*	Buried	-	3.04%	20.1%
1680	647 C to T	216 Ser to Phe	Beta Sheet	+*	Buried	-*	3.26%	42.2%
1681	598 G to A	200 Asp to Asn	Coil	+	Buried	-	3.05%	20.5%

Table 3.4. *xylE* missense mutants (continued)

***xylE* mutant number**[1]	**Base pair change**	**Amino acid change**[2]	**Structural perturbation**[3]	**Mutation favorability**[4]	**Protein location**[5]	**Change in hydropathy**[6]	**Mutant activity**[7]	**Protein stability**[8]
1683	343 C to T	115 His to Tyr	Coil	-*	Surface (Partial)	-	3.15%	22.0%
1684	409 G to A	137 Glu to Lys	Coil	+	Surface	+	3.10%	73.8%
1698	454 G to A	152 Asp to Asn	Coil	+	Buried	-	3.15%	65.7%
1702	70 C to T	24 His to Tyr	Alpha Helix	+	Surface (Partial)	-	2.70%	8.4%
1710	868 G to A	290 Ala to Thr	Alpha Helix	-*	Surface (Partial)	+	2.40%	48.8%

1. Only unique mutations are listed in this table. The *xylE1571* mutation was observed in 2 of the missense mutants, the *xylE1574* mutation was observed in 2 of the missense mutants, the *xylE1581* mutation was observed in 3 of the missense mutants, the *xylE1583* mutation was observed in 2 of the missense mutants, the *xylE1585* mutation was observed in 3 of the missense mutants, the *xylE1589* mutation was observed in 2 of the missense mutants and the *xylE1670* mutation was observed in 2 of the missense mutants.
2. The *xylE* gene is 921 bp in length and codes for the 307 amino acid catechol 2,3-dioxygenase protein. The resulting amino acid changes are given based on the coded protein predicted by the DNA sequence and not the Protein Data Bank file.
3. The secondary structure affected by the mutation was determined using PyMol.
4. The likelihood of a mutation to affect the α-helical, β-sheet or random coil structure was determined using P_α, P_β or P_c values from the averaged propensity scale in Table 3.1. (+) indicates a favorable change and (–) indicates an unfavorable change. An asterisk indicates a mutation that changes the original amino acid from or to one of the preferred amino acids that are found in α-helices, β-sheets or random coils.
5. The location of the mutated amino acids was determined using PyMOL.
6. The change in hydropathy was determined using the averaged hydropathy scale in Table 3.2. (+) indicates a more hydrophilic change, (–) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydropathy of the original amino acid significantly and results in a shift of at least 5 amino acids.
7. Mutant activities are listed with respect to the wild-type activity (100%). The assays were repeated in triplicate and the standard deviation was less than 10%. Wild type *xylE* has an activity of 422.92 XylE units.

Table 3.4. *xylE* missense mutants (continued)

8. Protein stability was determined by western blot analysis. The stability of the mutants are listed with respect to wild-type protein levels (100%). Mutant stability was calculated by comparing the amount of mutant protein that was present to the level of wild-type catechol 2,3-dioxygenase protein in the *xylE*+ parent strain.

Table 3.5. Analysis of the impact of the *lacZ* missense mutants on the secondary structures using changes in the propensity scores generated by the Chou-Fasman, GOR IV or Qian-Sejnowski algorithms.

lacZ Mutant Number	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR IV			Qian-Sejnowski		
			Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
39	RYHYQ**L**VWC	RYHYQ**Q**VWC	103.67	101.67	-1.93%	81.44	66.67	-18.14%	0.24	0.21	-14.43%
337	PQ**E**NYPDRLTAA	PQ**K**NYPDRLTAA	103.67	105.25	+1.53%	40.08	50.25	+25.36%	0.72	0.72	+0.06%
361	RVNWLGL**G**	RVNWLGL**R**	99.25	100.00	+0.76%	34.00	4.38	-87.13%	0.11	0.09	-14.88%
364	**G**I	**D**I	112.50	110.00	-2.22%	-109.00	-161.50	-48.17%	0.19	0.14	-25.00%
2234	LQ**G**GFV	LQ**D**GFV	112.17	109.67	-2.23%	110.17	67.33	-38.88%	0.20	0.15	-23.21%
2343	RCGT**R**ELNY	RCGT**H**ELNY	94.22	93.00	-1.30%	35.44	50.22	+41.69%	0.20	0.18	-10.99%
2381	RQS**G**FL	RQS**D**FL	98.83	97.17	-1.69%	-8.33	-10.42	-25.00%	0.50	0.50	-0.66%
2382	LQG**G**FV	LQG**D**FV	112.17	108.83	-2.97%	93.67	39.83	-57.47%	0.20	0.15	-26.58%
2396	DEN**G**N	DEN**D**N	139.80	137.60	-1.57%	125.80	120.30	-4.37%	0.88	0.92	+4.00%
2449	EDQ**D**	EDQ**N**	106.75	109.00	+2.11%	-55.50	-0.25	+99.55%	0.53	0.44	-16.78%
2454	IR**G**VN	IR**D**VN	93.00	92.12	-0.95%	0.76	-21.06	-2853.85%	0.14	0.13	-11.95%
2456	MI**T**	MI**M**	95.67	109.67	+14.63%	-11.33	79.67	+802.94%	0.23	0.45	+95.68%
2530	ANIET**H**GMVP MNRLTD	ANIET**Y**GMVPMNRL TD	101.69	102.88	+1.17%	-17.19	-9.38	+45.45%	0.64	0.63	-0.63%
2540	M**I**T	M**N**T	95.67	92.33	-3.48%	-11.33	-53.67	-373.53%	0.23	0.21	-10.79%
2608	L**G**NESGH	L**R**NESGH	123.57	119.29	-3.47%	21.00	-2.21	-110.54%	0.79	0.70	-11.59%

The propensity scores for both the parent and mutant secondary structures are listed for each predictive model as well as the percentage chang

Table 3.6. Analysis of the impact of the *lacZ* missense mutants on the secondary structure using changes in the consensus scores generated by the JPred4, PSIPRED, Porter 4.0 or SPIDER2 algorithms.

lacZ Mutant Number	JPred4			PSIPRED			Porter 4.0			SPIDER2		
	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
39	6.11	5.89	-3.64%	7.89	7.67	-2.82%	6.67	3.67	-45.00%	0.931	0.909	-2.396%
337	6.33	6.25	-1.32%	5.42	4.333	-20.00%	3.33	3.58	+7.50%	0.565	0.590	+4.35%
361	4.63	12.00	+159.46%	5.88	5.63	-4.26%	4.13	4.13	0.00%	0.653	0.653	0.000%
364	7.00	6.50	-7.14%	1.00	1.00	0.00%	1.50	1.50	0.00%	0.401	0.450	+12.095%
2234	4.00	5.00	+25.00%	4.67	4.83	+3.57%	2.67	2.67	0.00%	0.462	0.372	-19.495%
2343	6.67	6.67	0.00%	3.44	3.44	0.00%	4.78	4.56	-4.65%	0.720	0.774	+7.419%
2381	4.17	1.33	-68.00%	4.17	2.500	-40.00%	4.33	5.00	+15.38%	0.747	0.685	-8.34%
2382	4.00	4.67	+16.67%	4.67	5.17	+10.71%	2.67	2.67	0.00%	0.462	0.453	-1.841%
2396	6.40	6.40	0.00%	0.98	8.200	+733.50%	8.00	8.00	0.00%	0.984	0.985	+0.12%
2449	5.00	5.25	+5.00%	1.60	2.500	+56.25%	2.00	2.25	+12.50%	0.583	0.568	-2.66%
2454	4.89	5.12	+4.68%	5.82	5.76	-1.01%	4.12	4.29	+4.29%	0.288	0.272	-5.298%
2456	5.67	2.00	-64.71%	3.33	2.00	-40.00%	5.00	1.00	-80.00%	0.179	0.265	+47.58%
2530	6.60	6.60	0.00%	6.75	6.75	0.00%	4.44	4.18	-5.88%	0.684	0.700	+2.44%
2540	5.67	6.67	+17.65%	3.33	4.00	+20.00%	5.00	5.67	+13.33%	0.179	0.195	+8.74%
2608	6.29	6.14	-2.27%	7.43	7.29	-1.92%	4.86	4.86	0.00%	0.839	0.853	+1.62%

The consensus scores for both the parent and mutant secondary structures are listed for each predictive model as well as the percentage change.

Table 3.7. Analysis of the impact of the *xylE* missense mutants on the secondary structures using changes in the propensity scores generated by the Chou-Fasman, GOR IV or Qian-Sejnowski algorithms.

xylE Mutant Number	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR IV			Qian-Sejnowski		
			Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1568	**R**VRFQ	**C**VRFQ	102.20	103.00	+0.78%	95.60	77.60	-18.83%	0.40	0.35	-12.34%
1569	G**G**DYNYPDHKP	G**Q**DYNYPDHKP	123.82	121.09	-2.20%	120.82	110.364	-8.65%	0.91	0.91	0.00%
1571	ADKEYT**G**KWGLN DVNPEAWPRDLKG MAAVRFDH	ADKEYT**Q**KWGL NDVNPEAWPRDL KGMAAVRFDH	94.50	89.85	-4.92%	-6.35	-12.39	-95.31%	1.47	0.61	-58.54%
1574	WEDLLRAADL I**S**MT	WEDLLRAADL I**F**MT	111.64	114.43	+2.50%	136.71	164.07	+20.01%	0.57	0.62	+8.16%
1577	MSK**A**LEHYVE	MSK**V**LEHYVE	116.40	113.50	-2.49%	251.10	94.60	-62.33%	0.77	0.73	-5.71%
1581	LL**G**L	LL**S**L	76.75	74.25	-3.26%	-76.38	-95.63	-25.20%	0.24	0.16	-31.77%
1582	**G**MDFMGFKV	**D**MDFMGFKV	106.44	106.11	-0.31%	-114.56	-152.78	-33.37%	0.13	0.11	-13.84%
1583	ALMY**G**	ALMY**D**	95.00	90.80	-4.42%	-64.60	-86.60	-34.06%	0.09	0.08	-19.15%
1584	DTSIDIG**P**TRHGLT HG	DTSIDIG**S**TRHGL THG	107.38	106.88	-0.47%	6.13	-12.22	-299.49%	0.73	0.60	-18.19%
1585	**G**	**S**	99.00	113.00	+14.14%	66.00	41.00	-37.88%	0.75	0.67	-10.07%
1587	YLAE**Q**VL	YLAE**L**VL	95.86	98.43	+2.68%	-23.14	2.57	+111.11%	0.15	0.18	+21.67 %
1588	DTSIDIGPTRH**G**LT HG	DTSIDIGPTRH**D**L THG	107.38	106.81	-0.52%	6.13	7.06	+15.31%	0.73	0.73	+0.60%
1589	VT**W**T	VT**S**T	122.50	117.00	-4.49%	106.75	91.25	-14.52%	0.25	0.18	-28.36%
1591	VMR**P**	VMR**S**	92.75	100.75	+8.63%	-5.00	50.00	+1100.00%	0.10	0.20	+98.75 %
1592	AQFL**S**	AQFL**N**	107.60	102.00	-5.20%	66.00	47.40	-28.18%	0.42	0.36	-16.08%
1593	DTSIDIGPT**R**HGLT HG	DTSIDIGPT**C**HGL THG	107.38	108.88	+1.40%	6.13	25.19	+311.22%	0.73	0.75	+3.34%

Table 3.7. Analysis of the impact of the *xylE* missense mutants on the secondary structures using changes in the propensity scores generated by the Chou-Fasman, GOR IV or Qian-Sejnowski algorithms (continued).

xylE Mutant Number	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR IV			Qian-Sejnowski		
			Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1665	DTSIDI**G**PTRHGLTH	DTSIDI**D**PTRHGLTH	104.07	103.40	-0.64%	-0.03	1.13	+3500.00%	0.71	0.72	+1.73%
1666	DTSIDIGPTRHGLTH**G**	DTSIDIGPTRHGLTH**D**	107.38	106.69	-0.64%	6.13	6.44	+5.10%	0.73	0.73	-0.56%
1670	V**A**FIH	V**T**FIH	103.20	108.80	+5.43%	41.80	92.60	+121.53%	0.24	0.35	+44.44%
1677	RL**H**HVSFH	RL**Y**HVSFH	104.67	109.00	+4.14%	35.25	56.63	+60.64%	0.26	0.38	+46.60%
1680	RLHHV**S**FH	RLHHV**G**FH	104.67	103.50	-1.11%	35.25	38.63	+9.57%	0.26	0.25	-3.64%
1681	LSTKAH**D**	LSTKAH**N**	95.56	96.56	+1.05%	-46.67	-37.79	+19.03%	0.46	0.47	+0.97%
1683	APSG**H**	APSG**Y**	113.80	117.60	+3.34%	70.90	85.10	+20.03%	0.83	0.82	-0.84%
1684	ADKEYTGKWGLNDVNP**E**AWPRDLKGMAAVRFDH	ADKEYTGKWGLNDVNP**K**AWPRDLKGMAAVRFDH	101.33	102.06	+0.72%	-12.09	-7.67	+36.57%	0.59	0.58	-1.18%
1698	ADKEYTGKWGLNDVNPEAWPRDLKGMAAVRF**D**H	ADKEYTGKWGLNDVNPEAWPRDLKGMAAVRF**N**H	101.33	101.67	+0.33%	-10.89	-11.20	-2.84%	0.60	0.59	-1.39%
1702	MSKALE**H**YVE	MSKALE**Y**YVE	117.40	112.50	-4.17%	141.10	99.00	-29.84%	0.77	0.72	-6.94%
1710	TDQLGK**A**I	TDQLGK**T**I	106.63	99.13	-7.03%	43.88	-24.88	-156.70%	0.31	0.19	-38.29%

The propensity scores for both the parent and mutant secondary structures are listed for each predictive model as well as the percentage change.

Table 3.8. Analysis of the impact of the *xylE* missense mutants on the secondary structure using changes in the consensus scores generated by the JPred4, PSIPRED, Porter 4.0 or SPIDER2 algorithms.

xylE Mutant Number	JPred4			PSIPRED			Porter 4.0			SPIDER2		
	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1568	7.80	7.80	0.00%	7.80	7.80	0.00%	7.80	7.80	0.00%	0.962	0.917	-4.673%
1569	4.92	5.73	+16.49%	6.09	6.46	+5.97%	5.73	5.82	+1.59%	0.769	0.813	+5.72%
1571	5.84	5.88	+0.53%	6.46	6.36	-1.41%	4.59	4.75	+3.40%	0.683	0.698	+2.20%
1574	7.43	7.43	0.00%	7.86	7.69	-2.10%	8.14	8.07	-0.94%	0.922	0.925	+0.349%
1577	7.00	6.70	-4.29%	7.50	7.60	+1.33%	8.20	8.20	0.00%	0.979	0.980	+0.123%
1581	4.25	4.25	0.00%	3.50	3.75	+7.14%	3.00	2.75	-8.33%	0.801	0.839	+90.14%
1582	4.67	5.00	+7.14%	6.00	6.11	+1.85%	6.11	6.33	+3.64%	0.816	0.733	-10.263%
1583	5.20	3.60	-30.77%	2.80	2.00	-28.57%	1.80	1.60	-11.11%	0.716	0.645	-9.994%
1584	5.69	5.50	-3.30%	5.36	5.38	-4.44%	4.94	4.75	-3.80%	0.667	0.601	-9.90%
1585	2.00	6.00	+200.00%	2.00	0.00	-100.0%	6.00	6.00	0.00%	0.441	0.679	+53.97%
1587	4.57	3.57	-21.88%	4.43	4.14	-6.45%	4.86	4.86	0.00%	0.889	0.874	-1.656%
1588	5.69	5.38	-5.49%	5.63	6.25	+11.11%	4.94	4.81	-2.53%	0.667	0.677	+1.51%
1589	2.25	2.20	-2.22%	3.50	3.25	-7.14%	1.00	1.40	+40.00%	0.419	0.543	+29.473%
1591	2.50	2.25	-10.00%	2.25	2.25	0.00%	1.50	2.50	+66.67%	0.279	0.322	+15.219%
1592	8.20	8.20	0.00%	5.20	6.00	+15.38%	8.40	8.40	0.00%	0.922	0.893	-3.124%
1593	5.69	5.47	-3.81%	5.63	5.44	-3.33%	4.94	4.75	-3.80%	0.667	0.677	+1.51%
1665	5.47	5.60	+2.44%	5.47	5.87	+7.32%	4.13	3.93	-4.84%	0.661	0.673	+1.86%
1666	5.69	5.94	+4.40%	5.44	5.63	+3.45%	4.94	4.94	0.00%	0.683	0.695	+1.63%
1670	6.80	4.83	-28.92%	4.40	5.20	+18.18%	5.00	5.00	0.00%	0.722	0.686	-4.904%
1677	6.50	5.88	-9.62%	5.38	5.00	-6.98%	5.63	6.00	+6.67%	0.872	0.909	+4.316%
1680	6.50	5.88	-9.62%	5.38	5.50	+2.33%	5.63	6.00	+6.67%	0.872	0.878	+0.746%
1681	4.14	3.14	-24.14%	3.86	3.86	0.00%	4.43	4.43	0.00%	0.665	0.686	+3.18%
1683	5.80	5.80	0.00%	8.60	8.60	0.00%	6.40	6.40	0.00%	0.857	0.821	-4.27%

Table 3.8. Analysis of the impact of the *xylE* missense mutants on the secondary structure using changes in the consensus scores generated by the JPred4, PSIPRED, Porter 4.0 or SPIDER2 algorithms (continued).

xylE Mutant Number	JPred4			PSIPRED			Porter 4.0			SPIDER2		
	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1684	4.76	4.76	-0.15%	5.21	5.21	0.00%	4.58	4.55	-0.66%	0.577	0.586	+1.54%
1698	6.09	6.06	-0.50%	5.91	5.88	-0.51%	5.76	5.82	+1.05%	0.647	0.658	+1.77%
1702	7.00	6.60	-5.71%	7.00	7.10	+1.43%	7.80	7.70	-1.28%	0.979	0.975	-0.419%
1710	4.13	4.14	+0.43%	4.88	5.88	+20.51%	1.71	1.44	-15.74%	0.489	0.415	-15.056%

The consensus scores for both the parent and mutant secondary structures are listed for each predictive model as well as the percentage change.

Table 3.9. Ability of the Chou-Fasman, GOR IV and Qian-Sejnowski algorithms to accurately predict the presence of secondary structures in the *lacZ* missense mutants and the original parent

lacZ Mutant Number	Parent Secondary Structure	Mutant Secondary Structure	Secondary Structure	Chou-Fasman			GOR IV			Qian-Sejnowski		
				Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
39	RYHYQ**L**VWC	RYHYQ**Q**VWC	Beta Sheet	77.8%	77.8%	0%	66.7%	66.7%	0%	16.7%	0.06%	-16.64%
337	PQ**E**NYPDRLTAA	PQ**K**NYPDRLTAA	Coil	50%	50%	0%	66.7%	66.7%	0%	87.5%	87.5%	0%
361	RVNWLGL**G**	RVNWLGL**R**	Beta Sheet	50%	25.0%	-25%	50%	37.5%	-12.5%	0%	0%	0%
364	**G**I	**D**I	Beta Sheet	100%	100%	0%	0%	0%	0%	0%	0%	0%
2234	LQ**G**GFV	LQ**D**GFV	Beta Sheet	66.7%	66.7	0%	50%	33.3%	-16.7%	0%	0%	0%
2343	RCGT**R**ELNY	RCGT**H**ELNY	Beta Sheet	22.2%	22.2%	0%	22.2%	33.3%	+11.1%	0%	0%	0%
2381	RQS**G**FL	RQS**D**FL	Coil	33.3%	100%	+66.7%	66.7%	50%	-16.7%	75.0%	58.3%	-16.7%
2382	LQG**G**FV	LQG**D**FV	Beta Sheet	66.7%	66.7%	0%	50%	50%	0%	0%	0%	0%
2396	DEN**G**N	DEN**D**N	Coil	100%	60%	-40%	80%	100%	+20%	50%	50%	0%
2449	EDQ**D**	EDQ**N**	Coil	0%	0%	0%	0%	0%	0%	25.0%	25.0%	0%
2454	IR**G**VNRHEHHPLHGQVM	IR**D**VNRHEHHPLHGQVM	Beta Sheet	35.3%	11.8%	-23.5%	29.4%	11.8%	-17.6%	8.8%	0.06	-8.74%
2456	MI**T**	MI**M**	Alpha Helix	0%	0%	0%	0%	66.7%	+66.7%	66.7%	66.7	0%
2530	ANIET**H**GMVPMNRLTD	ANIET**Y**GMVPMNRLTD	Coil	68.8%	50%	-18.8%	56.3%	62.5%	+6.2%	90.6%	96.9%	+6.3%
2540	M**I**T	M**N**T	Alpha Helix	0%	0%	0%	0%	0%	0%	16.7%	0%	-16.7%
2608	L**G**NESGH	L**R**NESGH	Coil	87.5%	87.5%	0%	85.7%	71.4%	-14.3 %	0%	81.3%	+81.3%

Table 3.10. Ability of the JPred4, PSIPRED, Porter 4.0 and SPIDER2 algorithms to accurately predict the presence of secondary structures in the *lacZ* missense mutants and the original parent

lacZ Mutant Number	JPred4			PSIPRED			Porter 4.0			SPIDER2		
	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
39	66.7%	77.8%	+11.1%	88.9%	88.9%	0%	88.9%	88.9%	0%	100%	100%	0%
337	100%	100%	0%	100%	100%	0%	83.3%	75.0%	-8.3%	66.7%	75.0%	+8.3%
361	62.5%	62.5%	0%	50.0%	50.0%	0%	50%	37.5%	-12.5%	62.5%	62.5%	0%
364	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	0%	-50%
2234	33.3%	50.0%	+16.7%	16.7%	16.7%	0%	0%	0%	0%	33.3%	16.6%	-16.7%
2343	55.6%	55.6%	0%	44.4%	44.4%	0%	55.6%	55.6%	0%	77.8%	77.8%	0%
2381	50%	33.3%	-16.7%	83.3%	83.3%	0%	83.3%	75.0%	-8.3%	83.3%	83.3%	0%
2382	33.3%	33.3%	0%	16.7%	16.7%	0%	0%	0%	0%	33.3%	33.3%	0%
2396	100%	100%	0%	100%	100%	0%	100%	100%	0%	100%	83.3%	-16.7%
2449	100%	100%	0%	100%	100%	0%	100%	100%	0%	75.0%	75.0%	0%
2454	29.4%	23.5%	-5.9%	11.8%	11.8%	0%	29.4%	33.3%	+3.9%	17.6%	17.6%	0%
2456	0%	25.0%	+25%	0%	33.3%	+33.3%	0%	0%	0%	0%	33.3%	+33.3%
2530	100%	100%	0%	100%	100%	0%	100%	100%	0%	93.8%	93.8%	0%
2540	0%	0%	0%	0%	0%	0%	33.3%	33.3%	0%	0%	0%	0%
2608	100%	85.7%	-14.3%	100%	100%	0%	100%	100%	0%	100%	100%	0%

Table 3.11. Ability of the Chou-Fasman, GOR IV and Qian-Sejnowski algorithms to accurately predict the presence of secondary structures in the *xylE* missense mutants and the original parent.

xylE Mutant Number	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR IV			Qian-Sejnowski		
			Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1568	**R**VRFQ	**C**VRFQ	40.0%	80.0%	+40%	80.0%	40.0%	-40.0%	40.0%	50.0%	+10%
1569	G**G**DYNYPDHKP	G**Q**DYNYPDHKP	81.8%	81.8%	0%	63.6%	45.5%	-18.1%	50.0%	50.0%	0%
1571	ADKEYT**G**KWGLND VNPEAWPRDLKGM AAVRFDH	ADKEYT**Q**KWGLND VNPEAWPRDLKGM AAVRFDH	53.1%	53.1%	0%	15.6%	18.8%	+3.2%	76.6%	73.4%	-3.2%
1574	WEDLLRAADL I**S**MT	WEDLLRAADL I**F**MT	57.1%	57.1%	0%	78.6%	78.6%	0%	78.6%	85.7%	+7.1%
1577	MSK**A**LEHYVE	MSK**V**LEHYVE	50.0%	0%	-50%	100%	90%	-10.0%	100%	100%	0%
1581	LL**G**L	LL**S**L	50.0%	50.0%	0%	0%	0%	0%	0%	0%	0%
1582	**G**MDFMGFKV	**D**MDFMGFKV	33.3%	33.3%	0%	0%	0%	0%	0%	0%	0%
1583	ALMY**G**	ALMY**D**	0%	0%	0%	20.0%	0%	-20%	0%	0%	0%
1584	DTSIDIG**P**TRHGLTH G	DTSIDIG**S**TRHGLTH G	50.0%	25.0%	-25%	25.0%	6.3%	-18.7%	50.0%	71.9%	+21.9%
1585	**G**	**S**	0%	100%	+100%	0%	0%	0%	100%	100%	0%
1587	YLAE**Q**VL	YLAE**L**VL	0%	0%	0%	0%	0%	0%	0%	0%	0%
1588	DTSIDIGPTRH**G**LTH G	DTSIDIGPTRH**D**LTH G	50.0%	50.0%	0%	25.0%	31.3%	+6.3%	50.0%	50.0%	0%
1589	VT**W**T	VT**S**T	100%	100%	0%	100%	75.0%	-25%	0%	0%	0%
1591	VMR**P**	VMR**S**	25.0%	50.0%	+25%	50.0%	75.0%	+25%	0%	0%	0%
1592	AQFL**S**	AQFL**N**	80.0%	20.0%	-60.0%	0%	0%	0%	20.0%	10.0%	-10.0%
1593	DTSIDIGPT**R**HGLTH G	DTSIDIGPT**C**HGLTH G	50.0%	50.0%	0%	25.0%	31.3%	+6.3%	50.0%	50.0%	0%

Table 3.11. Ability of the Chou-Fasman, GOR IV and Qian-Sejnowski algorithms to accurately predict the presence of secondary structures in the *xylE* missense mutants and the original parent (continued).

xylE Mutant Number	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR IV			Qian-Sejnowski		
			Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1665	DTSIDI**G**PTRHGLTH	DTSIDI**D**PTRHGLTH	46.7%	46.7%	0%	26.7%	53.3%	+26.6%	50.0%	50.0%	0%
1666	DTSIDIGPTRHGLTH **G**	DTSIDIGPTRHGLTH**D**	50.0%	56.3%	+6.3%	25.0%	43.8%	+23.8%	50.0%	50.0%	0%
1670	V**A**FIH	V**T**FIH	0%	60.0%	+60%	0%	80.0%	+80%	18.2%	60.0%	+41.8%
1677	RL**H**HVSFH	RL**Y**HVSFH	37.5%	62.5%	+25.%	50.0%	62.5%	+12.5%	12.5%	37.5%	+25%
1680	RLHHV**S**FH	RLHHV**G**FH	37.5%	25.0%	-12.5%	50.0%	62.5%	+12.5%	12.5%	12.5%	0%
1681	LSTKAH**D**	LSTKAH**N**	0%	0%	0%	0.0%	0.0%	0%	27.8%	44.4%	+16.6%
1683	APSG**H**	APSG**Y**	80.0%	80.0%	0%	40.0%	40.0%	0%	50.0%	50.0%	0%
1684	ADKEYTGKWGLND VNP**E**AWPRDLKGM AAVRFDH	ADKEYTGKWGLND VNP**K**AWPRDLKGM AAVRFDH	42.2%	42.2%	0%	15.2%	21.2%	+6%	69.7%	68.2%	-1.5%
1698	ADKEYTGKWGLND VNPEAWPRDLKGM AAVRF**D**H	ADKEYTGKWGLND VNPEAWPRDLKGMA AVRF**N**H	51.5%	51.5%	0%	15.2%	15.2%	0%	74.2%	69.7%	-4.5%
1702	MSKALE**H**YVE	MSKALE**Y**YVE	50.0%	40.0%	-10%	100%	90.0%	-10.0%	100%	90.0%	-10%
1710	TDQLGK**A**I	TDQLGK**T**I	75.0%	0%	-75%	62.5%	12.5%	-50%	18.8%	0%	-18.8%

Table 3.12. Ability of the JPred4, PSIPRED, Porter 4.0 and SPIDER2 algorithms to accurately predict the presence of secondary structures in the *xylE* missense mutants and the original parent

xylE Mutant Number	JPred4			PSIPRED			Porter 4.0			SPIDER2		
	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1568	100%	100%	0%	100%	100%	0%	100%	100%	0%	100%	100%	0%
1569	84.6%	90.9%	+6.3%	81.8%	81.8%	0%	81.8%	100%	+18.2	90.9%	90.9%	0%
1571	90.6%	87.5%	-3.1%	81.8%	84.4%	+2.6%	81.8%	81.8%	0%	78.1%	84.4%	+6.3%
1574	92.9%	92.9%	0%	92.9%	92.9%	0%	92.9%	92.9%	0%	92.9%	92.9%	0%
1577	100%	100%	0%	100%	100%	0%	100%	100%	0%	100%	100%	0%
1581	0.0%	0.0%	0%	50.0%	75.0%	+25%	75.0%	75.0%	0%	100%	100%	0%
1582	66.7%	77.8%	+11.1%	77.8%	77.8%	0%	66.7%	66.7%	0%	88.9%	77.8%	-11.1%
1583	80.0%	60.0%	-20%	60.0%	60.0%	0%	60.0%	60.0%	0%	80.0%	60%	-20%
1584	81.3%	70.6%	-9.7%	80.0%	86.7%	+6.7%	87.5%	81.3%	0%	70.6%	62.5%	-8.1%
1585	100.0%	100.0%	0%	50.0%	100.0%	+50%	100%	100%	0%	0.0%	100%	+100%
1587	85.7%	85.7%	0%	100%	100%	0%	85.7%	85.7%	0%	100%	100%	0%
1588	81.3%	81.3%	0%	80.0%	81.3%	+1.3%	87.5%	87.5%	0%	70.6%	68.8%	-1.8%
1589	75.0%	75.0%	0%	50.0%	50.0%	0%	75.0%	75.0%	0%	75.0%	25.0%	-50%
1591	75.0%	50.0%	-25%	0%	0%	0%	50.0%	50.0%	0%	25.0%	50%	+25%
1592	100%	100.0%	0%	80.0%	80.0%	0%	100%	100%	0%	100%	100%	0%
1593	81.3%	81.3%	0%	80.0%	81.3%	+1.3%	87.5%	81.3%	-6.2%	70.6%	75.0%	+4.4%
1665	68.8%	73.3%	+4.5%	80.0%	80.0%	0%	80.0%	80.0%	0%	80.0%	80.0%	0%
1666	81.3%	81.3%	0%	80.0%	81.3%	+1.3%	87.5%	81.3%	-6.2%	75.0%	81.3%	+6.3%
1670	100%	100%	0%	100%	100%	0%	100%	100%	0%	80.0%	80.0%	0%
1677	87.5%	87.5%	0%	87.5%	87.5%	0%	87.5%	87.5%	0%	87.5%	87.5%	0%
1680	87.5%	87.5%	0%	87.5%	87.5%	0%	87.5%	87.5%	0%	87.5%	87.5%	0%
1681	71.4%	71.4%	0%	57.1%	57.1%	0%	71.4%	57.1%	-14.3%	71.4%	71.4%	0%
1683	80.0%	80.0%	0%	100%	100%	0%	100%	100%	0%	80.0%	80.0%	0%

Table 3.12. Ability of the JPred4, PSIPRED, Porter 4.0 and SPIDER2 algorithms to accurately predict the presence of secondary structures in the *xylE* missense mutants and the original parent (continued).

xylE Mutant Number	JPred4			PSIPRED			Porter 4.0			SPIDER2		
	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
1684	81.8%	81.8%	0%	72.7%	97.0%	+24.3%	72.7%	72.7%	0%	66.7%	69.7%	+3%
1698	85.3%	85.3%	0%	97.0%	97.0%	0%	90.9%	87.9%	-3%	84.8%	84.8%	0%
1702	100%	100%	0%	100%	100%	0%	100%	100%	0%	100%	100%	0%
1710	0.0%	0%	0%	0.0%	0%	0%	37.5%	37.5%	0%	50.0%	37.5%	0%

REFERENCES

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science, 181*(4096), 223-230. doi:10.1126/science.181.4096.223

Astbury, W. T., & Street, A. (1931). X-ray studies of the structure of hair, wool, and related fibres.-i. general. *Philosophical Transactions of the Royal Society of London.Series A, Containing Papers of a Mathematical Or Physical Character, 230*(681-693), 75-101.

Astbury, W. T., & Sisson, W. A. (1935). X-ray studies of the structure of hair, wool, and related fibres-III—The configuration of the keratin molecule and its orientation in the biological cell. *Proceedings of the Royal Society of London.Series A-Mathematical and Physical Sciences, 150*(871), 533-551.

Astbury, W. T., & Woods, H. J. (1933). X-ray studies of the structure of hair, wool, and related fibres. II.-the molecular structure and elastic properties of hair keratin. *Philosophical Transactions of the Royal Society of London.Series A, Containing Papers of a Mathematical Or Physical Character, 232*(707-720), 333-394.

Babu, M. M., Kriwacki, R. W., & Pappu, R. V. (2012). Versatility from protein disorder. *Science, 337*(6101), 1460-1461.

Bertani, G. (1951). Studies on lysogenesis. I. the mode of phage liberation by lysogenic escherichia coli. *Journal of Bacteriology, 62*(3), 293-300. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14888646

Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C., & Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-dimensional fourier synthesis at 2 angstrom resolution. *Nature, 206*(4986), 757-761. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/5891407

Casadaban, M. J., & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in escherichia coli. *Journal of Molecular Biology, 138*(2), 179-207. doi:10.1016/0022-2836(80)90283-1

Chou, P. Y., & Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry, 13*(2), 222-245.

Cole, A. E., Hani, F. M., Altman, R., Meservy, M., Roth, J. R., & Altman, E. (2017). The promiscuous sumA missense suppressor from salmonella enterica has an intriguing mechanism of action. *Genetics, 205*(2), 577-588. doi:10.1534/genetics.116.196550

Conidi, A., Berghe, V., Leslie, K., Stryjewska, A., Xue, H., Chen, Y., . . . Huylebroeck, D. (2013). Four amino acids within a tandem QxVx repeat in a predicted extended α-helix of the smad-binding domain of Sip1 are necessary for binding to activated smad proteins. *PLoS ONE, 8*(10), e76733. doi:10.1371/journal.pone.0076733

Costantini, S., Colonna, G., & Facchiano, A. M. (2006). Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochemical and Biophysical Research Communications, 342*(2), 441-451.

Cuff, J. A., & Barton, G. J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins: Structure, Function, and Bioinformatics, 40*(3), 502-511.

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., & Barton, G. J. (1998). JPred: A consensus secondary structure prediction server. *Bioinformatics (Oxford, England), 14*(10), 892-893. doi:10.1093/bioinformatics/14.10.892

Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: A protein secondary structure prediction server. *Nucleic Acids Research, 43*(W1), W389-W394. doi:10.1093/nar/gkv332

Eisenberg, D., Schwarz, E., Komaromy, M., & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology, 179*(1), 125-142. doi:10.1016/0022-2836(84)90309-7

Fasman, G. D. (1989). *Prediction of protein structure and the principles of protein conformation*. Boston: Springer. doi:10.1007/978-1-4613-1571-1

Fredericks, Z. L., & Pielak, G. J. (1993). Exploring the interface between the N- and C-terminal helices of cytochrome c by random mutagenesis within the C-terminal helix. *Biochemistry, 32*(3), 929-936.

Garnier, J., Gibrat, J. F., & Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods in Enzymology, 266*, 540-553.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology, 120*(1), 97-120. doi:10.1016/0022-2836(78)90297-8

Greeb, J., Atkins, J. F., & Loper, J. C. (1971). Histidinol dehydrogenase (hisD) mutants of salmonella typhimurium. *Journal of Bacteriology, 106*(2), 421-431.

Harayama, Rekik, Bairoch, Neidle, & Ornston. (1991). Potential DNA slippage structures acquired during evolutionary divergence of acinetobacter calcoaceticus chromosomal benABC and pseudomonas putida TOL pWW0 plasmid xylXYZ, genes encoding benzoate dioxygenases. *Journal of Bacteriology, 173*(23), 7540-7548. doi:10.1128/jb.173.23.7540-7548.1991

He, M. M., Wood, Z. A., Baase, W. A., Xiao, H., & Matthews, B. W. (2004). Alanine-scanning mutagenesis of the β-sheet region of phage T4 lysozyme suggests that tertiary context has a dominant effect on β-sheet formation. *Protein Science: A Publication of the Protein Society, 13*(10), 2716-2724. doi:10.1110/ps.04875504.

Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., Wang, J., . . . Zhou, Y. (2015). Improving prediction of secondary structure, local backbone angles and solvent accessible surface area of proteins by iterative deep learning. *Scientific Reports, 5*(1), 11476. doi:10.1038/srep11476

Hopp, T. P., & Woods, K. R. (1983). A computer program for predicting protein antigenic determinants. *Molecular Immunology, 20*(4), 483-489. doi:10.1016/0161-5890(83)90029-9

Jacobson, R. H., Zhang, X., DuBose, R. F., & Matthews, B. W. (1994). Three-dimensional structure of β-galactosidase from E. coli. *Nature, 369*(6483), 761.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology, 292*(2), 195-202. doi:10.1006/jmbi.1999.3091

Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X., Huber, R. E., Tronrud, D. E., & Matthews, B. W. (2000). High resolution refinement of β-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for α-complementation. *Protein Science, 9*(9), 1685-1699. doi:10.1110/ps.9.9.1685

Kendrew, J. C., Phillips, D. C., Dintzis, H. M., Wyckoff, H., Parrish, R. G., & Bodo, H. M. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature, 181*(4610), 662-666. doi:10.1038/181662a0

Kita, A., Kita, S., Fujisawa, I., Inaka, K., Ishida, T., Horiike, K., . . . Miki, K. (1999). An archetypical extradiol-cleaving catecholic dioxygenase: The crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from ppseudomonas putida mt-2. *Structure (London, England : 1993), 7*(1), 25-34.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology, 157*(1), 105-132. doi:10.1016/0022-2836(82)90515-0

Lyons, J., Dehzangi, A., Heffernan, R., Sharma, A., Paliwal, K., Sattar, A., . . . Yang, Y. (2014). Predicting backbone cα angles and dihedrals from protein sequences by stacked sparse auto-encoder deep neural network. *Journal of Computational Chemistry, 35*(28), 2040-2046.

Miller, J. H. (1992). *Experiments in molecular genetics* (11. print. ed.). New York: Cold Spring Harbor Laboratory.

Minor, D. L., & Kim, P. S. (1994). Measurement of the β-sheet-forming propensities of amino acids. *Nature, 367*(6464), 660.

Mirabello, C., & Pollastri, G. (2013). Porter, PaleAle 4.0: High-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics (Oxford, England), 29*(16), 2056-2058. doi:10.1093/bioinformatics/btt344

Oldfield, C. J., & Dunker, A. K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry, 83*, 553-584.

Otaki, J. M., Tsutsumi, M., Gotoh, T., & Yamamoto, H. (2010). Secondary structure characterization based on amino acid composition and availability in proteins. *Journal of Chemical Information and Modeling, 50*(4), 690-700.

Pace, C. N., & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal, 75*(1), 422-427.

Pauling, L., & Corey, R. B. (1951). Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. *Proceedings of the National Academy of Sciences of the United States of America, 37*(11), 729-740. doi:10.1073/pnas.37.11.729

Pauling, L., Corey, R. B., & Branson, H. R. (1951). The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America, 37*(4), 205-211. doi:10.1073/pnas.37.4.205

Pavlopoulou, A., & Michalopoulos, I. (2011). State-of-the-art bioinformatics protein structure prediction tools (review). *International Journal of Molecular Medicine, 28*(3), 295.

Pegram, L. M., Wendorff, T., Erdmann, R., Shkel, I., Bellissimo, D., Felitsky, D. J., & Record, M. T. (2010a). Why hofmeister effects of many salts favor protein folding but not DNA helix formation. *Proceedings of the National Academy of Sciences, 107*(17), 7716-7721.

Pirovano, W., & Heringa, J. (2010). Protein secondary structure prediction. *Data mining techniques for the life sciences* (pp. 327-348) Springer.

Pollastri, G., & McLysaght, A. (2005). Porter: A new, accurate server for protein secondary structure prediction. *Bioinformatics (Oxford, England), 21*(8), 1719-1720. doi:10.1093/bioinformatics/bti203

Prevelige, P., & Fasman, G. D. (1989). Chou-fasman prediction of the secondary structure of proteins. *Prediction of protein structure and the*

principles of protein conformation (pp. 391-416) Springer.

Qian, N., & Sejnowski, T. J. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology, 202*(4), 865-884. doi:10.1016/0022-2836(88)90564-5

Simm, S., Einloft, J., Mirus, O., & Schleiff, E. (2016). 50 years of amino acid hydrophobicity scales: Revisiting the capacity for peptide classification. *Biological Research, 49*(1), 31.

Smith, C. K., Withka, J. M., & Regan, L. (1994). A thermodynamic scale for the. beta.-sheet forming tendencies of the amino acids. *Biochemistry, 33*(18), 5510-5517.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology, 189*(1), 113-130.

Tanford, C., & Lovrien, R. (1962). Dissociation of catalase into subunits. *Journal of the American Chemical Society, 84*(10), 1892-1896. doi:10.1021/ja00869a025

Truman, P., & Bergquist, P. L. (1976). Genetic and biochemical characterization of some missense mutations in the lacZ gene of escherichia coli K-12. *Journal of Bacteriology, 126*(3), 1063-1074.

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., . . . Jones, D. T. (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews, 114*(13), 6589-6631.

Whitfield, H. J., Martin, R. G., & Ames, B. N. (1966). Classification of aminotransferase (C gene) mutants in the histidine operon. *Journal of Molecular Biology, 21*(2), 335,IN14,34-346,IN14,355. doi:10.1016/0022-2836(66)90103-3

Will, G., North, A. C. T., Perutz, M. F., Rossmann, M. G., Muirhead, H., & Cullis, A. F. (1960). Structure of hæmoglobin: A three-dimensional fourier synthesis at 5.5- . resolution, obtained by X-ray analysis. *Nature, 185*(4711), 416-422. doi:10.1038/185416a0

Zimmerman, J. M., Eliezer, N., & Simha, R. (1968). The characterization of amino acid sequences in proteins by statistical methods. *Journal of Theoretical Biology, 21*(2), 170-201.

CHAPTER IV: THE ABILITY OF SALTS TO STABILIZE PROTEINS IN VIVO OR INTRACELLULARLY CORRELATES WITH THE HOFMEISTER SERIES OF IONS

Fatmah M. Hani, Ashley E. Cole, Elliot Altman*

Fatmah M. Hani and Ashley E. Cole contributed equally to this research.

Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132

*- corresponding author, ealtman@mtsu.edu

Fatmah Hani conducted most of the enzyme assays and assisted with the writing of the published manuscript.

Published August 2019 in International Journal of Biochemistry and Molecular Biology:

Hani, F. M., Cole, A. E., & Altman, E. (2019). The ability of salts to stabilize proteins *in vivo* or intracellularly correlates with the Hofmeister series of ions. *International journal of biochemistry and molecular biology*, *10*(3), 23–31.

4.1 Abstract

Numerous studies have been conducted on the ability of salts to stabilize proteins *in vitro* using purified proteins demonstrating the fact that the ability of salts to stabilize proteins correlates with the Hofmeister series of ions. Using the well characterized bacterial aqueous cytosolic β-galactosidase and catechol 2,3-dioxygenase enzymes, we demonstrated that salts can stabilize proteins *in vivo* or intracellularly as well and that the ability of salts to stabilize these two proteins intracellularly also correlates with the Hofmeister series of ions. Na_2SO_4 and Na_2HPO_4 were very effective at stabilizing both proteins, followed by NaCl, NH_4Cl and $(NH_4)_2HPO_4$, while $NH_4CH_3CO_2$, $(NH_4)_2SO_4$ and $NaCH_3CO_2$ did not stabilize either of the proteins. We also investigated the ability of salts to rescue a collection of well characterized nonfunctional β-galactosidase and catechol 2,3-dioxygenase missense mutants that our laboratory has created. 73.33% of the β-galactosidase missense mutants could be rescued by salt, while only 33.33% of the catechol 2,3 dioxygenase missense mutants could be rescued by salt. This observation was explained by the differences in densities for the two proteins. Catechol 2,3 dioxygenase is almost twice as dense or compact as β-galactosidase and thus it is far easier for salts to penetrate and rescue inactive β-galactosidase proteins. 68.42% of the missense mutants that were rescuable by salt contained mutations that affected amino acids on the surface of the protein and is consistent with the likelihood that salt is able to rescue missense mutants that affect amino acids located on the surface of the protein much more readily than salt can rescue missense mutants that affect amino acids buried in the protein.

Key words

Hofmeister ions, protein stability, salt suppression

4.2 Introduction

From the seminal studies of (Hofmeister, 1888), regarding the ability of different salts to precipitate egg white protein, the lyotropic or Hofmesier series of ions was developed in which ions are ordered by their ability to precipitate proteins. For salts with the same cation, SO_4^{2-} > HPO_4^{2-} > F^- > $CH_3CO_2^-$ > Cl^- > Br^- > NO_3^- > I^- > ClO_4^- > SCN^- and for salts with the same anion, $(CH3)_4N^+$ > Rb^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Ca^{2+}. For general reviews see (Jungwirth & Cremer, 2014); and (Salis & Ninham, 2014). Given the great significance of the Hofmeister series of ions, the 1888 Hofmeister article, which was written in German, has been translated in its entirety into English (Kunz, Henle, & Ninham, 2004).

In a series of studies with collagen, gelatin and ribonuclease, Von Hippel and Wong, (Von Hippel & Wong, 1963; Von Hippel & Wong, 1964; Von Hipple & Wong, 1962) showed that the ability of the Hofmeister series of ions to precipitate proteins also correlated with their ability to stabilize proteins. Amongst the salts studied in these experiments, KPO_4 and $(NH_4)_2SO_4$ had the greatest stabilizing effects, followed by KCl and NaCl. Subsequent studies by Nandi and Robinson, (Nandi & Robinson, 1972b; Nandi & Robinson, 1972a), using short peptides confirmed the fact that proteins were stabilized with respect to the Hofmeister series of ions. In these studies, Na_2SO_4 had the greatest stabilizing effect, followed by NaCl and NaBr. More recently Pegram (Pegram et al., 2010) also completed a thorough

study on the ability of Hofmeister salts to stabilize the DNA binding domain (DBD) of *lac* repressor. In these studies, Na_2SO_4 and KF had the greatest stabilizing effects, followed by KCl and NaCl.

The ability of salts to stabilize proteins intracellularly has also been demonstrated in a number of intracellular studies using nonfunctional missense mutants. Hawthorne and Friis (Hawthorne & Friis, 1964), were the first to identify mutants whose functionality could be restored by the addition of salt as they demonstrated that 36 out of 231 *Saccharomyces cerevisiae* auxotrophic mutants could be rescued by the addition of KCl at 0.5 M or 1.0 M and proposed that salt correctable mutants were most likely to be missense mutants. Other researchers showed subsequently that this phenomenon occurred in bacteria as well (Bilsky & Armstrong, 1973; Good & Pattee, 1970; Kohno & Roth, 1979; Russell, 1972). Good and Pattee, 1970, isolated 15 temperature sensitive *Staphylococcus aureus* mutants that could be corrected by 1 M NaCl. Russell, (Russell, 1972), demonstrated that 5 out of 14 temperature sensitive *Escherichia coli* mutants could be rescued by 1% (0.171 M) NaCl. Bilsky and Armstong, (Bilsky & Armstrong, 1973), found that 32 out of 40 temperature sensitive *E. coli* mutants could be rescued by 0.5% (0.086 M) NaCl. Kohno and Roth, (Kohno & Roth, 1979), showed that 56 temperature-sensitive and 47 cold sensitive *Salmonella enterica* serovar Typhimurium histidine auxotrophs could all by rescued by 0.2 M NaCl.

In this study, we have tested the ability of Hofmeister salts to stabilize wild-type proteins intracellularly. The well characterized bacterial aqueous cytosolic β-galactosidase and catechol 2,3-dioxygenase enzymes were chosen for this analysis. The structures for both of these proteins has been determined (Juers et al., 2000; Kita et al., 1999) and their activity can be quantified using robust easy to use colorimetric enzyme assays (Cole, Ashley E. et al., 2017; Miller, J. H., 1972).We have also examined the ability of salts to rescue a collection of inactive missense mutants that our laboratory has generated in these two proteins (Cole, Ashley, Hani, Allen, Kline, & Altman, 2018)

4.3 Materials and Methods

4.3.1 Media and bacterial strains

Lysogeny broth (LB) (Bertani, 1951) was used as the rich media to maintain the strains used in this study. While most studies that have been conducted in *E. coli* and *S. enterica* on the ability of salt to rescue nonfunctional missense mutants have used nutrient broth (NB), the basal salt concentration in nutrient broth is higher than desired. The concentration of sodium and chorine in NB are 0.007 M and 0.003 M, respectively (Difco Manual, 11th Edition. 1998. Difco Laboratories, Sparks, Maryland). For this reason, we utilized yeast glucose broth (YGB), which consisted of 5.0% yeast extract supplemented with 0.2% glucose, as the rich medium in which all the strains were grown to determine the effects of the different salts that were tested. The concentration of sodium and chorine in YGB are 0.003 M and 0.001 M, respectively (Difco Manual, 11th Edition) and thus the basal salt concentration in

YGB is considerably lower than NB. Strains that were grown in YGB achieved similar OD_{550} values as strains that were grown in NB or LB. To induce the expression of either the β-galactosidase enzyme coded by *lacZ* or the catechol 2,3-dioxygenase enzyme coded by *xylE*, isopropyl β-D-thiogalactopyranoside (IPTG) was added at a final concentration of 1 mM using the *S. enterica* strains TT18519, *hisC10081*::MudF(*lac*+) and ALS1442, *proB*::*xylE*(*cat*), respectively (Cole, Ashley E. et al., 2017). Salt solutions to be tested were prepared at 1M concentrations, adjusted to pH 7 if necessary, and added to 2X YGB to achieve the final concentration desired in 1X YGB.

4.3.1 β-galactosidase and catechol 2,3-dioxygenase enzyme assays

β-galactosidase assays and catechol 2,3-dioxygenase assays were performed as described (Cole, Ashley E. et al., 2017; Miller, Jeffrey H., 1992) with two important modifications to eliminate the salt that is present in the buffers used in the assays. Z buffer was replaced with 10 mM Tris; pH 7 and both the ortho-nitrophenyl-β-galactoside (*ONPG*) and the catechol substrate solutions were prepared in 10 mM Tris; pH 7 instead of phosphate buffer.

4.3.2 Three-dimensional analysis of the β-galactosidase and catechol 2,3-dioxygenase proteins

The PyMOL Molecular Graphics System (Schrödinger, LLC) was used to determine the volume of the β-galactosidase and catechol 2,3-dioxygenase proteins.

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) was the source of the three-dimensional crystal structure data. The 1DP0 file was used for the β-galactosidase crystal structure data (Juers et al., 2000) and the 1MPY file was used for the catechol 2,3-dioxygenase crystal structure data (Kita et al., 1999).

4.4 Results

4.4.1 Determining the ability of different salts to stabilize the β-galactosidase and catechol 2,3-dioxygenase enzymes intracellularly

We selected eight salts, $NH_4CH_3CO_2$ (ammonium acetate), NH_4Cl (ammonium chloride), $(NH_4)_2HPO_4$ (ammonium phosphate dibasic), $(NH_4)_2SO_4$ (ammonium sulfate), $NaCH_3CO_2$ (sodium acetate), NaCl (sodium chloride), Na_2HPO_4 (sodium phosphate dibasic) and Na_2SO_4 (sodium sulfate), for these studies based on three criteria. First, they contained both very strong and moderate anions and cations for stabilizing proteins according to the Hofmeister series of ions. Second, these were generally the best salts for stabilizing proteins based on previous studies (Nandi & Robinson, 1972b; Nandi & Robinson, 1972a; Pegram et al., 2010; Von Hippel & Wong, 1963; Von Hippel & Wong, 1964; Von Hipple & Wong, 1962). Third, of the anions and cations that were selected, all of the possible combinations were tested.

The wild-type strains that produced β-galactosidase and catechol 2,3-dioxygenase were grown in YGB plus 1 mM IPTG that contained increasing salt in 0.05 M increments. After 16 hours of growth β-galactosidase or catechol 2,3-dioxygenase enzyme assays were performed and the results are shown in Table 4.1. Na_2SO_4 was

the best salt for stabilizing β-galactosidase, followed by Na_2HPO_4, NaCl, NH_4Cl and $(NH_4)_2HPO_4$. $NH_4CH_3CO_2$, $(NH_4)_2SO_4$ and $NaCH_3CO_2$ did not stabilize β-galactosidase. Na_2SO_4 was the best salt for stabilizing catechol 2,3-dioxygenase, followed by Na_2HPO_4, $(NH_4)_2HPO_4$, NH_4Cl and NaCl. $NH_4CH_3CO_2$, $(NH_4)_2SO_4$ and $NaCH_3CO_2$ did not stabilize catechol 2,3-dioxygenase. Thus, there was very good agreement in the ability of the various salts to stabilize both proteins. The molar concentrations at which the different salts provided their maximal stabilization varied widely and ranged from 0.05 M to 0.035 M for both enzymes.

4.4.2 Determining the ability of salt to suppress β-galactosidase and catechol 2,3-dioxygenase missense mutants

Numerous researchers have shown that nonfunctional missense mutants can be rescued by salt *in vivo*. We wanted to extend these studies by determining whether a collection of 42 unique well-characterized nonfunctional missense mutants, 15 in β-galactosidase and 27 in catechol 2,3-dioxygenase, could be rescued by salt. Because NaCl has been the salt of choice in the previous salt correctable missense mutant studies we used NaCl as well. Both 0.2 M and 0.3 M final concentrations of NaCl were tested to see if salt could rescue the missense mutants. Table 4.2 shows the results of the enzyme assays for missense mutants that had at least a 25% or 1.25 fold increase in enzyme activity upon the addition of salt. Interestingly, NaCl had a negative effect on the missense mutants that were not rescuable by the addition of NaCl. 11 out of the 15, or 73.33% of the nonfunctional β-galactosidase missense mutants could be rescued by the addition of NaCl, but only 9 out of the 27, or 33.33%

of the nonfunctional catechol 2,3-dioxygenase missense mutants could be rescued by the addition of NaCl. For the β-galactosidase missense mutants, the β-galactosidase activity was increased by an average of 5.54 fold with the addition of 0.2 M NaCl and 6.05 fold with the addition of 0.3 M NaCl. For the catechol 2,3-dioxygenase missense mutants, the catechol 2,3-dioxygenase activity was increased by an average of 1.64 fold with the addition of 0.2 M NaCl and 1.83 fold by the addition of 0.3 M NaCl. Thus, the higher concentration of NaCl was slightly more effective at rescuing the nonfunctional missense mutants that could be rescued by salt. *LacZ39* was by far the most suppressible missense mutant as its β-galactosidase activity was increased 19.70 fold with the addition of 0.2 M NaCl and 26.54 fold with the addition of 0.3 M NaCl.

Using data from Chapter 3 (Cole, Ashley et al., 2018), Tables 4.3 and 4.4 list the β-galactosidase and catechol 2,3-dioxygenase missense mutants, respectively, and gives the amino acid change, what secondary structure is affected by the missense mutant and what type of impact the mutation is expected to have on the secondary structure, whether the mutated amino acid is located on the surface of the protein, partially located on the surface of the protein or buried, the change in hydropathy caused by the mutation, and whether the missense mutant can be rescued by salt. Based on our knowledge of protein folding and structure, one would expect that missense mutants which affect amino acids located on the surface of the protein would be far more likely to be rescuable by salt than missense mutants which affect buried amino acids of the protein. 13 out of the 19, or 68.42% of the missense

mutants that were rescuable by salt contained mutations on the surface of the proteins. Amongst the β-galactosidase missense mutants, where the large majority were rescuable by salt, the four missense mutants that could not be rescued by salt all contained mutations which affected amino acids that were buried in the protein.

4.4.3 Determining the ability of different salts to stabilize the β-galactosidase enzyme produced by the *lacZ*39 missense mutant

Because the *lacZ39* missense mutant was the most rescuable by NaCl, we tested the ability of the eight salts we had used previously with wild-type β-galactosidase and catechol 2,3-dioxygenase to see how they affected the inactive β-galactosidase enzyme produced by *lacZ39*. Table 4.5 shows the result of this study. Na_2SO_4 was the best salt for stabilizing β-galactosidase from *lacZ39*, followed by NaCl, NH_4Cl, $(NH_4)_2SO_4$, $NaCH_3CO_2$, Na_2HPO_4 and $(NH_4)_2HPO_4$. Only $NH_4CH_3CO_2$ did not stabilize the β-galactosidase from *lacZ39*. The results clearly demonstrate that the salts affect wild-type β-galactosidase enzyme differently than an inactive mutant β-galactosidase enzyme. While Na_2SO_4 was the best salt at stabilizing both wild-type β-galactosidase and the β-galactosidase produced from *lacZ39*, NaCl was almost as effective at stabilizing the β-galactosidase produced from *lacZ39*. Neither $NH_4CH_3CO_2$, $(NH_4)_2SO_4$ or $NaCH_3CO_2$ could stabilize wild-type β-galactosidase, but only $NH_4CH_3CO_2$ was ineffective at stabilizing the β-galactosidase produced from *lacZ39*.

4.5 Discussion

In this study, we have examined whether the best salts known to stabilize purified proteins *in vitro* based on previous studies (Nandi & Robinson, 1972b; Nandi & Robinson, 1972a; Pegram et al., 2010; Von Hippel & Wong, 1964) could also perform *in vivo* or intracellularly using the two very well characterized β-galactosidase and catechol 2,3 dioxygenase enzymes. To ensure that all the possible combinations of anions and cations were tested based on the salts that have proven the most effective at stabilizing proteins in *in vitro* studies, we tested the following eight salts, $NH_4CH_3CO_2$ (ammonium acetate), NH_4Cl (ammonium chloride), $(NH_4)_2HPO_4$ (ammonium phosphate dibasic), $(NH_4)_2SO_4$ (ammonium sulfate), $NaCH_3CO_2$ (sodium acetate), NaCl (sodium chloride), Na_2HPO_4 (sodium phosphate dibasic) and Na_2SO_4 (sodium sulfate). Generally, Na_2SO_4 and Na_2HPO_4 were very effective at stabilizing both proteins, followed by NaCl, NH_4Cl and $(NH_4)_2HPO_4$, while $NH_4CH_3CO_2$, $(NH_4)_2SO_4$ and $NaCH_3CO_2$ did not stabilize either of the proteins. However, Table 1 clearly show some distinct differences in the ability of the different salts to stabilize the two proteins. Na_2SO_4 and Na_2HPO_4 are equally effective at stabilizing catechol 2,3 dioxygenase, but Na_2SO_4 is much more effective at stabilizing β-galactosidase than Na_2HPO_4. NaCl is much better at stabilizing β-galactosidase than NH_4Cl and $(NH_4)_2HPO_4$, while $(NH_4)_2HPO_4$ is much better at stabilizing catechol 2,3 dioxygenase than NH_4Cl or NaCl.

Numerous researchers have shown that salt can rescue inactive proteins *in vivo* by testing the ability of salt to suppress nonfunctional missense mutants (Bilsky

& Armstrong, 1973; Good & Pattee, 1970; Hawthorne & Friis, 1964; Kohno & Roth, 1979; Russell, 1972). We have expanded these studies by testing the ability of salt to rescue a set of 42 well characterized unique nonfunctional missense mutants, 15 in β-galactosidase and 27 in catechol 2,3 dioxygenase. A total of 19 out of the 42, or 45.24% of the missense mutants were rescuable by salt (11 out of 15, or 73.33% for β-galactosidase and 9 out of 27, or 33.33% for catechol 2,3 dioxygenase). Thus, clearly it was a lot easier to suppress nonfunctional missense mutants in β-galactosidase than it was to suppress nonfunctional missense mutants in catechol 2,3 dioxygenase. An examination of the structure of the two proteins provides an answer.

Both the β-galactosidase and catechol 2,3 dioxygenase proteins are tetrameric. The β-galactosidase monomer is 1,024 amino acids in size and the tetramer's molecular weight is 465,912 g/mol with width, height and depth measurements of 174.10 Å, 136.00 Å and 86.75 Å, respectively. The catechol 2,3 dioxygenase monomer is 307 amino acids in size and the tetramer's molecular weight is 140,616 g/mol with width, height and depth measurements of 94.95 Å, 65.70 and 51.60 Å, respectively. Thus, the density of β-galactosidase is 0.377 g/cm^3, while the density of catechol 2,3 dioxygenase is 0.725 g/cm^3. Since β-galactosidase is a lot less dense or compacted than catechol 2,3 dioxygenase, it should be a lot easier for salts to penetrate and effect β-galactosidase than catechol 2,3 dioxygenase. Additionally, β-galactosidase has a unique structure that consists of a continuous system of channels running along the surface and within the tetramer. These channels appear to be accessible to bulk solvent and vary in width from 5 - 20 Å (Juers et al., 2000).

Several studies have demonstrated the propensity of Hofmeister salts to stabilize proteins via hydrophobic interactions (Geisler et al., 2008; Tadeo, López-Méndez, Castaño, Trigueros, & Millet, 2009; Thomas & Elcock, 2007) and that small cations and large anions facilitate this effect (Schwierz, Horinek, & Netz, 2013). According to these findings NaCl, which has a smaller cation with an ionic radius of .102 nM and a larger anion with an ionic radius of 0.181 nM, would be expected to rescue inactive missense mutants where the mutated amino acid had a hydrophobic shift to a greater degree than missense mutants where the mutated amino acid had a hydrophilic shift. Of the 42 β-galactosidase and catechol 2,3 dioxygenase missense mutants that were characterized in this study, 14 of the mutants had an amino acid change that resulted in a hydrophobic shift and 28 of the missense mutants had an amino acid change that resulted in a hydrophilic shift. A total of 8 out of the 14, or 57.14% of the mutants that resulted in a hydrophobic shift were rescuable by NaCl, while only 12 out of 28, or 42.86% of the mutants that resulted in a hydrophilic shift were rescuable by NaCl. Thus, there was a slight preference for NaCl to rescue mutants that resulted in a hydrophobic shift than to rescue mutants that resulted in a hydrophilic shift.

We further characterized the ability of the different salts to correct the *lacZ39* missense mutant that was highly suppressible by salt. Na_2SO_4 was the most effective salt for stabilizing β-galactosidase from lacZ39, followed by NaCl, NH_4Cl, $(NH_4)_2SO_4$, $NaCH_3CO_2$, Na_2HPO_4 and $(NH_4)_2HPO_4$. Only $NH_4CH_3CO_2$ could not stabilize the β-galactosidase from lacZ39. Interestingly, unlike wild-type β-

galactosidase where NaCl was not nearly as effective as Na_2SO_4, NaCl was almost as effective as Na_2SO_4 at stabilizing the β-galactosidase produced by lacZ39. This finding is consistent with NaCl being the choice of most researchers that have tested the ability of salts to rescue nonfunctional missense mutants.

Historically, most researchers have used NaCl to form the salt gradients in the anionic exchange or hydroxyapatite columns used in fast protein liquid chromatography (FPLC) to purify proteins and the resulting proteins are stored in buffers containing NaCl. The results of our intracellular studies and the numerous *in vitro* studies by other researchers that have investigated the ability of salts to stabilize proteins, suggest that Na_2SO_4 and Na_2HPO_4 would be better choices for stabilizing proteins.

Table 4.1. Impact of different concentrations of Hofmeister salts on the activity of β-galactosidase and catechol 2,3-dioxygenase

Salts	0 M	0.05 M	0.1 M	0.15 M	0.2 M	0.25 M	0.3 M	0.35 M	0.4 M
β-galactosidase									
$NH_4CH_3CO_2$	152.11	33.53	25.08	19.59	9.39	ND	ND	ND	ND
NH_4Cl	152.11	179.73	174.20	150.20	140.81	141.21	140.26	139.84	136.25
$(NH_4)_2HPO_4$	152.11	171.10	105.94	97.84	84.16	81.42	32.19	30.82	20.84
$(NH_4)_2SO_4$	152.11	24.41	43.91	65.34	110.49	105.71	103.46	101.93	98.44
$NaCH_3CO_2$	152.11	88.42	80.72	95.50	110.18	83.64	80.11	76.62	74.15
NaCl	152.11	233.40	225.28	232.12	252.71	248.30	255.90	277.20	311.02
Na_2HPO_4	152.11	303.21	366.04	384.11	392.20	395.90	409.15	392.54	387.37
Na_2SO_4	152.11	323.43	390.31	473.60	604.90	659.46	770.16	777.96	667.50
Catechol 2,3-dioxygenase									
$NH_4CH_3CO_2$	344.79	32.21	25.39	21.63	22.61	ND	ND	ND	ND
NH_4Cl	344.79	368.45	413.99	464.59	464.05	442.13	399.63	428.02	483.88
$(NH_4)_2HPO_4$	344.79	658.53	643.23	635.43	613.40	484.03	301.61	247.72	117.15
$(NH_4)_2SO_4$	344.79	67.80	333.54	303.95	258.16	152.99	72.86	68.49	70.19
$NaCH_3CO_2$	344.79	34.86	21.40	17.99	18.88	18.33	10.38	2.69	3.17
NaCl	344.79	307.45	393.44	385.58	397.93	387.37	389.36	375.36	367.17
Na_2HPO_4	344.79	739.92	847.58	852.98	852.33	884.97	858.91	897.50	952.78
Na_2SO_4	344.79	322.28	415.87	442.38	551.09	881.66	922.25	977.02	142.65

β-galactosidase activities are reported in Miller Units and catechol 2,3-dioxygenase activities are reported in XylE Units. All assays were repeated in triplicate and the standard deviation was less than 10%. ND denotes that these values could not be determined. $NH_4CH_3CO_2$ completely inhibited cell growth at concentrations of 0.25 M or above.

Table 4.2. Enzymatic activity of the salt correctible *lacZ* and *xylE* missense mutants in the presence of 0.2M or 0.3M NaCl

mutant	0.0 M NaCl	0.2 M NaCl	0.2 M NaCl Fold Increase	0.3 M NaCl	.3 M NaCl Fold Increase
β-galactosidase					
lacZ39	1.29	25.39	19.72	34.24	26.60
lacZ337	0.65	1.03	1.59	0.90	1.39
lacZ364	0.48	1.04	2.18	0.86	1.78
lacZ2234	0.70	0.89	1.27	0.91	1.30
lacZ2343	14.28	77.96	5.46	87.54	6.13
lacZ2381	0.78	2.37	3.06	2.67	3.45
lacZ2396	0.24	0.43	1.77	0.31	1.25
lacZ2449	1.26	14.27	11.36	12.41	9.88
lacZ2456	1.20	10.07	8.40	10.18	8.49
lacZ2530	0.79	2.42	3.07	2.61	3.33
lacZ2540	0.75	2.26	3.02	2.24	2.99
Catechol 2,3-dioxygenase					
xylE1568	7.46	10.40	1.39	14.17	1.90
xylE1569	20.59	25.69	1.25	30.58	1.49
xylE1583	8.34	11.24	1.35	11.82	1.42
xylE1585	8.17	17.67	2.16	15.00	1.84
xylE1591	14.16	20.34	1.44	26.75	1.89
xylE1683	6.76	16.82	2.49	11.77	1.74
xylE1698	7.42	10.41	1.4	15.56	2.10
xylE1702	9.60	14.14	1.47	21.09	2.20
xylE1710	6.18	10.92	1.77	11.61	1.88

β-galactosidase activities are reported in Miller Units and catechol 2,3-dioxygenase activities are reported in XylE Units. All assays were repeated in triplicate and the standard deviation was less than 10%.

Table 4.3. β-galactosidase missense mutants

lacZ mutant	Amino acid change[1]	Structural perturbation[2]	Mutation favorability[3]	Protein Location[4]	Change in hydropathy[5]	Salt Suppressible
39	1018 Leu to Gln	Beta Sheet	-	Surface (Partial)	+*	+
337	905 Glu to Lys	Coil	+	Surface (Partial)	+	+
361	900 Gly to Asp	Beta Sheet	-	Buried	+*	-
364	208 Gly to Asp	Beta Sheet	-	Buried	+*	+
2234	565 Gly to Asp	Beta Sheet	-	Buried	+*	+
2343	943 Arg to His	Beta Sheet	+	Surface	-*	+
2381	764 Gly to Asp	Coil	-	Buried	+*	+
2382	566 Gly to Asp	Beta Sheet	-	Buried	+*	-
2396	584 Gly to Asp	Coil	-	Surface	+*	+
2449	202 Asp to Asn	Coil	+	Surface (Partial)	-	+
2454	354 Gly to Asp	Beta Sheet	-	Buried	+*	-
2456	5 Thr to Met	Alpha Helix	+*	Surface	-*	+
2530	419 His to Tyr	Coil	-*	Surface	-	+
2540	4 Ile to Asn	Alpha Helix	-*	Surface	+*	+
2608	460 Gly to Arg	Coil	-*	Buried	+*	-

1. The *lacZ* gene codes for the 1,024 amino acid β-galactosidase protein. The resulting amino acid changes are given based on the coded protein predicted by the DNA sequence and not the Protein Data Bank file.

Table 4.3. β-galactosidase missense mutants (continued).

2. The secondary structure affected by the mutation was determined using PyMOL. Data from Jacobson *et al*. [23] had to be used to determine the secondary structures affected by *lacZ2456* and *lacZ2540*, since these mutations were at the extreme amino terminus.
3. The likelihood of a mutation to affect the α-helical, β-sheet or random coil structure was determined using P_α, P_β or P_c values from the averaged propensity scale in Cole *et al*. (2018). (+) indicates a favorable change and (–) indicates an unfavorable change. An asterisk indicates a mutation that changes the original amino acid from or to one of the preferred amino acids that are found in α-helices, β-sheets or random coils.
4. The location of the mutated amino acids was determined using PyMOL.
5. The change in hydropathy was determined using the averaged hydrophobicity scale from Cole *et al*. (2018). (+) indicates a more hydrophilic change, (–) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydropathy of the original amino acid significantly and results in a shift of at least 5 amino acids.

Table 4.4. Catechol 2,3-dioxygenase missense mutants

xylE mutant	Amino acid change[1]	Structural perturbation[2]	Mutation favorability[3]	Protein Location[4]	Change in hydropathy[5]	Salt Suppressible
1568	106 Arg to Cys	Beta Sheet	+*	Surface (Partial)	-*	+
1569	270 Gly to Gln	Coil	-*	Surface (Partial)	+*	+
1571	127 Gly to Gln	Coil	-*	Surface (Partial)	+*	-
1574	233 Ser to Phe	Alpha Helix	+	Surface	-*	-
1577	21 Ala to Val	Alpha Helix	-*	Buried	-*	-
1581	30 Gly to Ser	Coil	-	Surface (Partial)	+	-
1582	64 Gly to Asp	Beta Sheet	-	Surface	+*	-
1583	158 Gly to Asp	Beta Sheet	-	Buried	+*	+
1584	243 Pro to Ser	Coil	-	Surface (Partial)	+	-
1585	9 Gly to Ser	Coil	-	Buried	+	+
1587	178 Gln to Lys	Beta Sheet	+	Surface	+	-
1588	247 Gly to Asp	Coil	-	Surface (Partial)	+*	-
1589	282 Trp to Ser	Beta Sheet	-*	Surface (Partial)	+*	-
1591	8 Pro to Ser	Beta Sheet	+	Buried	+	+
1592	193 Ser to Asn	Beta Sheet	-	Buried	+	-
1593	245 Arg to Cys	Coil	+	Surface (Partial)	-*	-
1665	242 Gly to Asp	Coil	-	Surface (Partial)	+*	-
1666	251 Gly to Asp	Coil	-	Surface (Partial)	+*	-
1670	202 Ala to Thr	Beta Sheet	+*	Buried	+	-
1677	213 His to Tyr	Beta Sheet	+*	Buried	-	-
1680	216 Ser to Phe	Beta Sheet	+*	Buried	-*	-
1681	200 Asp to Asn	Coil	+	Buried	-	-

Table 4.4. Catechol 2,3-dioxygenase missense mutants (continued).

xylE mutant	Amino acid change[1]	Structural perturbation[2]	Mutation favorability[3]	Protein Location[4]	Change in hydropathy[5]	Salt Suppressible
1683	115 His to Tyr	Coil	-*	Surface (Partial)	-	+
1684	137 Glu to Lys	Coil	+	Surface	+	-
1698	152 Asp to Asn	Coil	+	Buried	-	+
1702	24 His to Tyr	Alpha Helix	+	Surface (Partial)	-	+
1710	290 Ala to Thr	Alpha Helix	-*	Surface (Partial)	+	+

1. The *xylE* gene codes for the 307 amino acid catechol 2,3-dioxygenase protein. The resulting amino acid changes are given based on the coded protein predicted by the DNA sequence and not the Protein Data Bank file.
2. The secondary structure affected by the mutation was determined using PyMOL.
3. The likelihood of a mutation to affect the α-helical, β-sheet or random coil structure was determined using P_α, P_β or P_c values from the averaged propensity scale in Cole *et al*. (+) indicates a favorable change and (–) indicates an unfavorable change. An asterisk indicates a mutation that changes the original amino acid from or to one of the preferred amino acids that are found in α-helices, β-sheets or random coils.
4. The location of the mutated amino acids was determined using PyMOL.
5. The change in hydropathy was determined using the averaged hydrophobicity scale in Cole *et al*. (+) indicates a more hydrophilic change, (–) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydropathy of the original amino acid significantly and results in a shift of at least 5 amino acids.

Table 4.5. Impact of different concentrations of Hofmeister salts on the activity of β-galactosidase for the *lacZ*39 missense mutant

Salts	0 M	0.05 M	0.1 M	0.15 M	0.2 M	0.25 M	0.3 M	0.35 M	0.4 M
$NH_4CH_3CO_2$	1.29	.096	1.05	0.84	0.80	ND	ND	ND	ND
NH_4Cl	1.29	3.11	8.97	10.08	13.65	16.88	19.94	21.44	18.47
$(NH_4)_2HPO_4$	1.29	2.28	1.91	1.38	1.14	1.08	0.60	0.52	0.43
$(NH_4)_2SO_4$	1.29	3.74	12.07	14.42	17.02	2.26	1.91	1.74	1.01
$NaCH_3CO_2$	1.29	3.17	2.84	3.40	4.46	6.36	8.23	8.41	7.66
NaCl	1.29	7.06	10.34	16.06	25.39	31.38	34.24	35.31	40.25
Na_2HPO_4	1.29	3.83	4.49	4.55	4.17	3.39	2.90	2.58	2.52
Na_2SO_4	1.29	11.17	22.11	36.39	57.26	48.67	45.18	39.15	34.64

β-galactosidase activities are reported in Miller Units. All assays were repeated in triplicate and the standard deviation was less than 10%. ND denotes that these values could not be determined. $NH_4CH_3CO_2$ completely inhibited cell growth at concentrations of 0.25 M or above.

REFERENCES

Bertani, G. (1951). Studies on lysogenesis. I. the mode of phage liberation by lysogenic escherichia coli. *Journal of Bacteriology, 62*(3), 293-300.

Bilsky, A. Z., & Armstrong, J. B. (1973). Osmotic reversal of temperature sensitivity in escherichia coli. *Journal of Bacteriology, 113*(1), 76-81. doi:10.20381/ruor-15103

Cole, A. E., Hani, F. M., Altman, R., Meservy, M., Roth, J. R., & Altman, E. (2017). The promiscuous sumA missense suppressor from salmonella enterica has an intriguing mechanism of action. *Genetics, 205*(2), 577-588. doi:10.1534/genetics.116.196550

Cole, A., Hani, F., Allen, B., Kline, P., & Altman, E. (2018). Nonfunctional missense mutants in two well characterized cytosolic enzymes reveal important information about protein structure and function. *The Protein Journal, 37*(5), 407-427. doi:10.1007/s10930-018-9786-6

Geisler, M., Pirzer, T., Ackerschott, C., Lud, S., Garrido, J., Scheibel, T., & Hugel, T. (2008). Hydrophobic and hofmeister effects on the adhesion of spider silk proteins onto solid substrates: An AFM-based single-molecule study. *Langmuir, 24*(4), 1350-1355.

Good, C. M., & Pattee, P. A. (1970). Temperature-sensitive osmotically fragile mutants of staphylococcus aureus. *Journal of Bacteriology, 104*(3), 1401-1403.

Hawthorne, D. C., & Friis, J. (1964). Osmotic-remedial mutants. a new classification for nutritional mutants in yeast. *Genetics, 50*(5), 829-839.

Hofmeister, F. (1888). *Zur lehre von der wirkung der salze: Dritte mittheilung* doi:10.1007/bf01838161

Jacobson, R. H., Zhang, X., DuBose, R. F., & Matthews, B. W. (1994). Three-dimensional structure of β-galactosidase from E. coli. *Nature, 369*(6483), 761.

Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X., Huber, R. E., Tronrud, D. E., & Matthews, B. W. (2000). High resolution refinement of β-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for α-complementation. *Protein Science, 9*(9), 1685-1699. doi:10.1110/ps.9.9.1685

Jungwirth, P., & Cremer, P. S. (2014). Beyond hofmeister. *Nature Chemistry, 6*(4), 261.

Kita, A., Kita, S., Fujisawa, I., Inaka, K., Ishida, T., Horiike, K., . . . Miki, K. (1999). An archetypical extradiol-cleaving catecholic dioxygenase: The crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from ppseudomonas putida mt-2. *Structure (London, England : 1993), 7*(1), 25-34.

Kohno, T., & Roth, J. (1979). Electrolyte effects on the activity of mutant enzymes in vivo and in vitro. *Biochemistry, 18*(7), 1386-1392. doi:10.1021/bi00574a041

Kunz, W., Henle, J., & Ninham, B. W. (2004). ‘Zur lehre von der wirkung der salze’(about the science of the effect of salts): Franz hofmeister's historical papers. *Current Opinion in Colloid & Interface Science, 9*(1-2), 19-37.

Miller, J. H. (1992). *Experiments in molecular genetics* (11. print. ed.). New York: Cold Spring Harbor Laboratory.

Nandi, P. K., & Robinson, D. R. (1972a). Effects of salts on the free energies of nonpolar groups in model peptides. *Journal of the American Chemical Society, 94*(4), 1308-1315. doi:10.1021/ja00759a043

Nandi, P. K., & Robinson, D. R. (1972b). Effects of salts on the free energy of the peptide group. *Journal of the American Chemical Society, 94*(4), 1299-1308.

Pegram, L. M., Wendorff, T., Erdmann, R., Shkel, I., Bellissimo, D., Felitsky, D. J., & Record, M. T. (2010). Why hofmeister effects of many salts favor protein folding but not DNA helix formation. *Proceedings of the National Academy of Sciences, 107*(17), 7716-7721.

Russell, R. R. B. (1972). *Temperature-sensitive osmotic remedial mutants of escherichia coli*

Salis, A., & Ninham, B. W. (2014). Models and mechanisms of hofmeister effects in electrolyte solutions, and colloid and protein systems revisited. *Chemical Society Reviews, 43*(21), 7358-7377.

Schwierz, N., Horinek, D., & Netz, R. R. (2013). Anionic and cationic hofmeister effects on hydrophobic and hydrophilic surfaces. *Langmuir, 29*(8), 2602-2614.

Tadeo, X., López-Méndez, B., Castaño, D., Trigueros, T., & Millet, O. (2009). Protein stabilization and the hofmeister effect: The role of hydrophobic solvation. *Biophysical Journal, 97*(9), 2595-2603.

Thomas, A. S., & Elcock, A. H. (2007). Molecular dynamics simulations of hydrophobic associations in aqueous salt solutions indicate a connection between water hydrogen bonding and the hofmeister effect. *Journal of the American Chemical Society, 129*(48), 14887-14898.

Von Hippel, P. H., & Wong, K. (1963a). The collagen [UNK] gelatin phase transition. I. further studies of the effects of solvent environment and polypeptide chain composition. *Biochemistry, 2*(6), 1387-1398.

Von Hippel, P. H., & Wong, K. (1963b). The collagen [UNK] gelatin phase transition. I. further studies of the effects of solvent environment and polypeptide chain composition. *Biochemistry, 2*(6), 1387-1398. doi:10.1021/bi00906a035

Von Hipple, P. H., & Wong, K. (1962). The effect of ions on the kinetics of formation and the stability of the collagenfold. *Biochemistry, 1*, 664-674.

Zimbro, M. J., Power, D. A., Miller, S. M., Wilson, G. E., & Johnson, J. A. (2009). *Difco & BBL manual: Manual of microbiological culture media, second edition.* BD Diagnostics - Diagnostic Systems, Sparks, MD.

CHAPTER V: EXAMINING THE RELEATIONSHIP BETWEEN MOLECULAR CHAPERONES AND PROTEASES IN MAINTAINING PROTEIN HOMEOSTASIS

Fatmah M. Hani, Ashley E. Cole, Elliot Altman*

Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132

*- corresponding author, ealtman@mtsu.edu

Fatmah Hani conducted all of the enzyme assays and wrote the initial draft of the manuscript that is to be submitted.

Key words

Protein homeostasis, proteases, chaperones

5.1 Abstract

Maintaining protein homeostasis, or proteostasis is crucial for cellular life. Proteostasis is maintained by two cellular systems, chaperones and proteases. Chaperones are responsible for assisting proteins to fold into their active structure. In contrast, proteases are responsible for degrading misfolded and aggregated proteins. Our study focused on the interaction between chaperones and proteases with misfolded proteins using bacteria as a model system. This relationship between the chaperones and proteases was examined by either overproducing the two main chaperones, Hsp70 (DnaKJ) or Hsp60 (GroLS) or removing the two main proteases, ClpX or Lon in the two well-studied cytosolic β-galactosidase and catechol 2,3-dioxygenase enzymes. Overall, out of the total 42 mutants that were examined, 16.67% of the mutants were rescued by GroLS overproduction and 47.62% of mutants were rescued by DnaKJ overproduction. The ClpX knockout rescued 78.57% of the mutants and the Lon knockout rescued 83.33% of the mutants. This study increased our understanding of the triage between chaperones and proteases regarding their interaction with misfolded proteins. Collectively our data demonstrated that proteases are more effective in controlling protein homeostasis than chaperones and that other essential chaperones must be involved in proteostasis in addition to GroLS and DnaKJ.

5.2 Introduction

The necessity of protein folding from the linear primary structure to the functional three-dimensional structure is well-known. In a stressful crowded environment proteins are prone to misfolding, and this failure to fold correctly results in nonfunctional proteins which can lead to several degenerative diseases. Alzheimer's, Parkinson's, and Huntington's diseases are some examples of diseases that can result from cells failing to maintain a delicate proteostatic environment (Dill & MacCallum, 2012). In order to maintain protein quality control or protein homeostasis, cells have developed two protein control systems which sustain this balanced state; the molecular chaperones, and proteases. Chaperones assist refolding of misfolded proteins, whereas proteases facilitate degradation of misfolded and aggregated proteins.

Molecular chaperones were first discovered as heat shock proteins due to the increased production of these proteins when cells undergo stressful situations such as increased temperature. Their role is to assist both native and mature misfolded proteins in folding correctly into their three-dimensional structure. Many achieve this with the help of co-chaperones. Classification of molecular chaperones is based on their molecular weight (Shiber & Ravid, 2014). Approximately 30% of nascent proteins will interact with molecular chaperones to fold correctly (Dunker, Silman, Uversky, & Sussman, 2008). Proteins would remain unstable in the absence of chaperones, leading to incorrectly folded intermediates, and ultimately becoming prone to aggregation (Cooper, 2000).

GroLS (known as Hsp60 in eukaryotes) and DnaKJ (known as Hsp70 in eukaryotes) are two cytosolic chaperone complexes that are conserved among numerous organisms (Finka et al., 2016; Saibil, 2013a). GroLS (Hsp60) is a homo-tetradecameric complex arranged in two seven-membered rings stacked back to back, where each monomer is composed of 57 kDa subunits. The Hsp60 (GroL) co-chaperone is Hsp10 (GroS), which is 10 kDa in size (Lu et al., 2014). DnaK, (Hsp70) is composed of a substrate-binding domain at one end which clamps around the target protein, and a nucleotide-binding domain that uses ATP to regulate the process on the other end. A flexible linker connects the two domains (Kityk et al., 2012). The 40-kDa co-chaperone that is needed for DnaK to function is DnaJ or the J-domain protein. The nucleotide exchange factor, GrpE also mediates DnaKJ chaperone function. Both of these chaperones recognize a hydrophobic sequence of an unfolded protein that is normally not exposed in a correctly folded protein (Fernández-Fernández & Valpuesta, 2018; Gur & Sauer, 2008).

The other destination of misfolded proteins is the degradation pathway or proteolysis. This mechanism is carried out by proteases that depend on ATP to function (Porankiewicz, Wang, & Clarke, 1999b). ATP-dependent proteases belong to the AAA+ family (ATPases Associated with a variety of cellular Activities). In bacteria, there are five types of ATP-dependent proteases. They are ClpXP, ClpAP, HslUV, Lon, and FtsH. All the proteases of the AAA+ family have a common feature of having the proteolytic domain contained inside a chamber with a narrow passage.

This feature of the multi-subunit protease ensures the protection of cytosolic proteins from being degraded by AAA+ proteases. The two most characterized proteases are ClpXP and Lon. ClpXP is a serine protease with a canonical Asp-His-Ser catalytic triad. Lon is also a serine protease; however, it contains a Lys-Ser catalytic dyad instead of a triad (Gur & Sauer, 2008; Gur & Sauer, 2009; Maillard et al., 2011b).

Of the protease families, ClpXP and Lon are responsible for degrading 50% of misfolded proteins. Both proteases are conserved amongst bacteria as well as plants and mammalian animals. They are well characterized, and a lot of research has been conducted to identify their structure and mechanism of action. ClpXP is composed of two seven-membered rings of ClpP. This 14-subunit complex exhibits a barrel-like structure, and one or two hexameric ClpX binds to it to form the full ClpXP protease (Flynn, Neher, Kim, Sauer, & Baker, 2003b; Porankiewicz et al., 1999b). The active sites are hydrophobic and are located inside the chamber. This characteristic of hydrophobicity facilitates specific diffusion and degradation of protein substrates (Maillard et al., 2011a). The Lon protease consists of a tetramer of identical subunits that assembles into a ring-shaped complex. Unlike Clp protease that contains the ATPase site and the proteolytic site in separate subunits, each of the Lon subunits carries both an ATP binding site and an active catalytic site (Chin et al., 1988; Lee & Suzuki, 2008b; Tripathi & Sowdhamini, 2008b).

Given the significance of protein homeostasis, there is an increasing interest in understanding how the regulatory network composed of chaperones and proteases

interact together. We have investigated this by characterizing whether the overproduction of the GroLS or DnaKJ chaperones or the deletion of the Lon or ClpX proteases can restore the activity of 42 nonfunctional missense mutants that our laboratory has generated and characterized in the the β-galactosidase or catechol 2,3-dioxygenase enzymes. Such fundamental understanding is necessary for future clinical applications (Finka et al., 2016).

5.3 Materials and Methods

5.3.1 Media and bacterial strains

Bacterial cultures were all grown in the lysogeny broth (LB) (Bertani, 1951). Isopropyl β -D-thiogalactopyranoside (IPTG) was added at a final concentration of 1 mM to induce the expression of either the β-galactosidase enzyme coded by *lacZ* gene or the catechol 2,3-dioxygenase enzyme coded by *xylE* gene. The *lacZ* gene was isolated from *E. coli*, and the *xylE* gene was isolated from the *Pseudomonas putida* pWW0 TOL plasmid. pOF39 and pAKL1 plasmids were used to overproduce GroLS and DnaKJ, respectively. The *S. enterica* strains TT18519, hisC10081::MudF(lac+) and ALS1442, proB::xylE(cat) were used in this study (Cole, Ashley E. et al., 2017).

5.3.2 Missense mutants used in the study

We used the collection of missense mutants our laboratory has generated in a previous study (Cole, Ashley E. et al., 2017) as described in chapter two. Also, as

shown in chapter 3, seven algorithms were used to predict the destabilization impact of each missense mutant (Cole, Ashley et al., 2018).

5.3.3 Construction of mutants that overproduce the chaperones

The plasmid pOF39 was used to overproduce GroLS, while the plasmid pAKL1 was used to overproduce DnaKJ. The transfer of the plasmid to the bacterial mutant strains was achieved by transducing the strains with P22 lysates of wild-type E. coli that contained pOF39 or pAKL1. Different dilutions of the P22 phage were mixed with the recipient mutant strains, and these multiple dilutions were plated on LB ampicillin. Some strains containing the missense mutants could not be transduced with the P22 lysate. For those strains, elelctroporation was used to transform the pOF39 or PAKL plasmids into the strain.

5.3.4 Overproduction of DnaKJ and GrpE

GrpE was cloned into a pACYC derivatized plasmid our laboratory has constructed which contains the trp-lacZ hybrid promoter from pTrc99A and utilizes the P15 ori. Because the pACYC derivative plasmid our laboratory has constructed is compatible with the pAKL1 plasmid, which utilizes the ColE1 ori, missence mutants containing both overproduced DnaKJ as well as GrpE could be tested. The pAKL1 plasmid with DnaKJ was selected using ampicillin while the pACYC-grpE plasmid was selected using chloramphenicol.

5.3.5 Removal of the two proteases, ClpX, and Lon

Deleting the two proteases was achieved using the Keio knockout technique as described by (Baba et al., 2006).

5.3.6 β-galactosidase and catechol 2,3-dioxygenase enzyme assays

Quantitative enzymatic assays were used to measure the restored enzymatic activity in the missense mutants after overproduction of chaperones or removal of proteases. β-galactosidase assays were performed as described by (Miller, Jeffrey H., 1992) and the catechol 2,3 dioxygenase assays were performed as described by (Cole, Ashley E. et al., 2017).

5.3.7 Western Blot Analysis

We employed the Western blot protocol described by Cole et al. (2018), except that 8-16% gradient gels were used instead of the 4–20% gradient gels. The ChemiDoc MP Imaging System was used to analyze the resulting data, and all intensity determinations were calculated using the Image Lab Software Bio-Rad (Bio-Rad, Hercules, CA, USA).

5.4 Results

5.4.1 Determining the effect of overproduction of either GroLS or DnaKJ and removal of either ClpXP or Lon on *lacZ* mutant strains

Table 5.1 shows the list of the 15 *lacZ* mutant strains, their amino acid change, the mutant enzymatic activity, the activity of the LacZ protein after overproducing GroLS, followed by the percentage change, the activity of the protein after overproducing DnaKJ, followed by the percentage change, the activity of the protein after removal of ClpX, followed by the percentage change and the activity of the protein after removal of Lon, followed by the percentage change. The overproduction of GroLS only restored enzymatic activity in 3 out of the 15, or 20% of the missense mutants (*lacZ*2343, *lacZ*2396, and *lacZ*2530). The overproduction of DnaKJ restored enzymatic activity in 5 out of the 15, or 33.33% of the missense mutants (*lacZ*39, *lacZ*2396, *lacZ*2449, *lacZ*2454, and *lacZ*2456). The deletion of ClpX restored enzymatic activity in 6 out of the 15, or 40% o the missense mutants (*lacZ*39, *lacZ*337, *lacZ*361, *lacZ*2382, *lacZ*2454, and *lacZ*2608), while 8 out of the 15 or 53.33% restored their activity upon Lon protease knockout (*lacZ*39, *lacZ*337, *lacZ*364, *lacZ*2343, *lacZ*2381, *lacZ*2382, *lacZ*2396, and *lacZ*2454).

5.4.2 Determining the effect of overproducing both GroLS and DnaKJ on *lacZ* mutant strains

Table 5.2 shows only the list of 6 *lacZ* missense mutants which were rescuable by the overproduction of either GroLS or DnaKJ. Of those 6 missense

mutants, 2 out of the 6, or 33.33% had a further increase of enzyme activity (*lacZ*2449, and *lacZ*2456) when both chaperones were overproduced.

5.4.3 Determining the effect of overproducing both DnaKJ and GrpE on *lacZ* mutant strains

Overproducing DnaKJ and GrpE together did not induce any further significant increase in the enzymatic activity on *lacZ* mutant strains (data not shown).

5.4.4 Determining the effect of overproduction of either GroLS or DnaKJ and removal of either ClpXP or Lon on *xylE* mutant strains

Table 5.3 shows the list of 27 *xylE* mutant strains, the amino acid changes, the mutant enzymatic activity, the activity of the protein after overproducing GroLS, followed by the percentage change, the activity of the protein after overproducing DnaKJ, followed by the percentage change, the activity of the protein after removal of ClpX, followed by the percentage change and the activity of the protein after removal of Lon, followed by the percentage change. The overproduction of GroLS only restored enzymatic activity in 4 out of 27, or 14.81% of the missense mutants (*xylE*1585, *xylE*1665, *xylE*1666 and *xylE*1677). The overproduction of DnaKJ restored enzymatic activity in 15 out of 27, or 55.56% of the missense mutants (*xylE*1568, *xylE*1574, *xylE*1577, *xylE*1584, *xylE*1587, *xylE*1588, *xylE*1589, *xylE*1665, *xylE*1666, *xylE*1670, *xylE*1677, *xylE*1680, *xylE*1683, *xylE*1684, and *xylE*1710). The deletion of either ClpX or Lon restored enzymatic activity in all 27, or 100% of the missense mutants.

5.4.5 Determining the effect of overproduction of both GroLS and DnaKJ on xylE mutant strains

Table 5.4 shows only the 16 *xylE* missense mutants that were rescued by the overproduction of either GroLS or DnaKJ. Of those 16 missense mutants, 9 out of the 16, or 56.25% had a further increase of enzyme activity (*xylE*1568, *xylE* 1588, *xylE* 1589, *xylE*1670, *xylE*1677, *xylE*1680, *xylE*1684, and *xylE*1710) when both chaperones were overproduced.

5.4.6 Determining the effect of overproduction both DnaKJ and GrpE on xylE mutant strains

As for the *lacZ* strains, overproducing DnaKJ and GrpE together did not induce any further significant increase in the enzymatic activity on *xylE* strains (data not shown).

5.4.7 Determining the restoration of protein stability by chaperone overproduction

Because it could be argued that the restoration of enzymatic activity by the overproduction of chaperones or the removal of proteases is not indicative of the prevention of degradation, we analyzed the restoration of the β-galactosidase and catechol 2,3-dioxygenase proteins by the overproduction of either GroLS or DnaKJ in mutants with protein levels that were less than 20% of wild-type protein levels, as determined by Cole et. al., 2018, whose enzymatic activity could be restored by either GroLS or DnaKJ. Cell cultures of the missense mutants with or without the GroLS or

DnaKJ chaperone were harvested, normalized for total protein levels and subjected to Western analysis. These results are shown in Table 5.5. The protein levels of the 4 β-galactosidase and 7 catechol 2,3-dioxygenase missense mutants whose enzymatic activity could be rescued by the overproduction of GroLS or DnaKJ were all increased by the overproduction of GroLS or DnaKJ. Figure 5.2 shows the Western blot of the 4 *lacZ* mutants with and without the chaperones along with wild-type *lacZ* (*lacZ*232).

5.4.8 Analysis of the rescuability of the missense mutants by the overproduction of chaperones with respect to the hydropathic change of the missense mutant

Because both the GroLS and DnaKJ chaperones act by binding to hydrophobic amino acids, we determined whether missense mutants that were rescuable by the overproduction of a chaperone had a hydrophilic or hydrophobic change as a consequence of the mutation. Change in hydropathy of all mutants was identified using the averaged hydropathy scale as described in chapter 3 (Cole, Ashley et al., 2018). The (+) indicates a more hydrophilic change, and the (–) indicates a more hydrophobic change. For *lacZ* strains, 11 out of the 15 mutants had a more hydrophilic change and out of those 11, 3 or 27.27% were rescued by overproduction of one of the chaperones. The remaining 4 mutants had a hydrophobic change, and all of them (or 100%) were rescued by overproduction of one of the chaperones. For the *xylE* strains, 17 out of the total 27 had a hydrophilic change and out of these 17, 10 or 58.82% were rescued by overproduction of one of the chaperones. The remaining 10 strains had a hydrophobic change, and 6 of them, or

60% were rescued by overproduction of one of the chaperones (Table 5.6). Collectively, 13 out of the 28, or 46.43% of the missense mutants with hydrophilic changes were rescued by overproducing one of the chaperones, while 10 out of the 14, or 71.43% of the missense mutants with hydrophobic changes were rescued by overproducing one of the chaperones. Thus, the overproduction of chaperones were more likely to restore the activity of a missense mutant with a hydrophobic change than a hydrophilic change.

5.5 Discussion

The involvement of chaperones and proteases in the regulation of protein homeostasis is well documented. Chaperones assist the folding of misfolded proteins while proteases destroy misfolded proteins. In this study, we addressed the question of whether the overproduction of the two main chaperones (DnaKJ or GroLS) or the removal of the two main proteases (ClpX or Lon) could rescue a collection of nonfunctional missense mutants in the bacterial cytosolic β-galactosidase and catechol 2,3-dioxygenase enzymes.

A total of 3 out of 15, or 20% of the β-galactosidase missense mutants were rescued by the overproduction of GroLS, while 5 out of the 15, or 33.33% of the β-galactosidase missense mutants were rescued by the overproduction of DnaKJ. A total of 4 out of 27, or 14% of the catechol 2,3-dioxygenase missense mutants were rescued by the overproduction of GroLS, while 15 out of the 27, or 55.56% of the β catechol 2,3-dioxygenase missense mutants were rescued by the overproduction of

DnaKJ. Collectively, of the 42 missense mutants that were examined, 7 out of the 42, or 16.67% could be rescued by the overproduction of GroLS, while 20 out of the 42, or 47.62% could be rescued by the overproduction of DnaKJ. Thus, the DnaKJ chaperone was much more effective at restoring the enzymatic activities of the β-galactosidase and catechol 2,3-dioxygenase enzymes.

This result can be explained by the fact that GroLS is a cage-like chaperone and encapsulation of misfolded proteins is limited to a certain size. The GroLS chaperone can assist unfolded proteins of no more than 100,000 Da in size (Gupta et al., 2014; Marchenkov & Semisotnov, 2009b). The β-galactosidase monomer is 116,478 Da in size, while the tetramer is 465,912 Da. The catechol 2,3-dioxygenase monomer is 35,154 Da in size, while the tetramer is 140,616 Da. Thus neither the β-galactosidase monomer nor tetramer can fit into the GroLS cage, and likely only the GroL complex is involved in rescuing misfolded β-galactosidase proteins, which significantly limits its ability to act on the β-galactosidase protein.

Numerous studies suggest that the GroLS and DnaKJ chaperones can act in tandem to assist the correct folding of misfolded proteins (Saibil, 2013b). To investigate this, we simultaneously overproduced both the GroLS and DnaKJ chaperones in any missense mutant that could be suppressed by the overproduction of either GroLS or DnaKJ alone. Of the 6 β-galactosidase missense mutants which were rescuable by the overproduction of either GroLS or DnaKJ, 2 out of the 6, or 33.33% had a further increase of enzyme activity when both chaperones were overproduced.

Of the 16 catechol 2,3-dioxygenase missense mutants which were rescuable by the overproduction of either GroLS or DnaKJ, 9 out of the 16, or 56.25% had a further increase of enzyme activity when both chaperones were overproduced. Due to the limited ability of the GroLS chaperone to assist the misfolding of β-galactosidase missense mutants, it is not surprising that the overproduction of both the GroLS and DnaKJ chaperones affected to rescue of catechol 2,3-dioxygenase missense mutants more than β-galactosidase missense mutants.

Since the DnaKJ chaperone was much better than the GroLS chaperone in facilitating the rescue of missense non-functional mutants, we also determined whether the overproduction of GrpE, the nucleotide exchange factor for DnaKJ, could improve the function of overproduced DnaKJ. The overproduction of GrpE did not increase the enzymatic activity of missense mutants that were rescued by the overproduction of DnaKJ. While GrpE was essential to the function of DnaKJ, our results confirm that it is a recyclable ancillary factor that is not required in stoichiometric amounts relative to DnaK or DnaJ.

While protein stability and functionality are well-correlated (Bromberg & Rost, 2009), we none the less wanted to confirm that protein stability was enhanced in non-functional missense mutants whose enzymatic activity was increased by the overproduction of the GroLS or DnaKJ chaperone. We analyzed any unstable missense mutant whose protein levels were less than 20% compared to the wild-type protein levels. In every case the overproduction of the appropriate chaperone

significantly increased the amount of protein made by the missense mutants. This result clearly shows that molecular chaperones restore both the native state and the enzymatic function of non-functional proteins by mediating stabilization.

It is well documented that the GroLS and DnaKJ chaperones interact with misfolded proteins by binding to hydrophobic amino acids (Fernández-Fernández & Valpuesta, 2018). We performed a hydropathic analysis of the missense mutants used in this study and found that 28 of the missense mutants caused a hydrophilic shift, while 14 of the missense mutants caused a hydrophobic shift. A total of 13 out of the 28, or 46.43% of the missense mutants with hydrophilic changes were resuced by overproducing one of the chaperones, while 10 out of the 14, or 71.43% of the missense mutants with hydrophobic changes were rescued by overproducing one of the chaperones. These results are consistent with the previous observations that the GroLS and DnaKJ chaperones interact with misfolded proteins by binding to hydrophobic amino acids.

The deletion of ClpX restored enzymatic activity in 6 out of the 15, or 40% of the β-galactosidase missense mutants and 27 out of the 27, or 100% of the catechol 2,3-dioxygenase missense mutants. The deletion of Lon restored enzymatic activity in 8 out of the 15, or 53.33% of the β-galactosidase missense mutants and 27 out of the 27, or 100% of the catechol 2,3-dioxygenase missense mutants. Thus, the removal of either the ClpX or Lon protease was much more effective at restoring the enzymatic activity of non-functional missense mutants and superior to the overproduction of

either the GrLS or DnaKJ chaperone. This result is perhaps not that surprising when one considers that in the case of overproducing a chaperone the proteases are still present and the chaperone must still compete with the proteases, while in the case of deleting a protease, the chaperones have an extended time frame in which to act without competition from the protease.

Collectively our data supports a competitive triad between chaperones and proteases, as chaperones try to refold misfolded proteins while proteases try to degrade misfolded proteins. The fate of misfolded proteins depend on which molecules they interact with first, proteases or chaperones. The collection of missense mutants used in this study can also be utilized to characterize the ability of other chaperones and proteases that are involved in protein homeostasis to stabilize or degrade misfolded proteins.

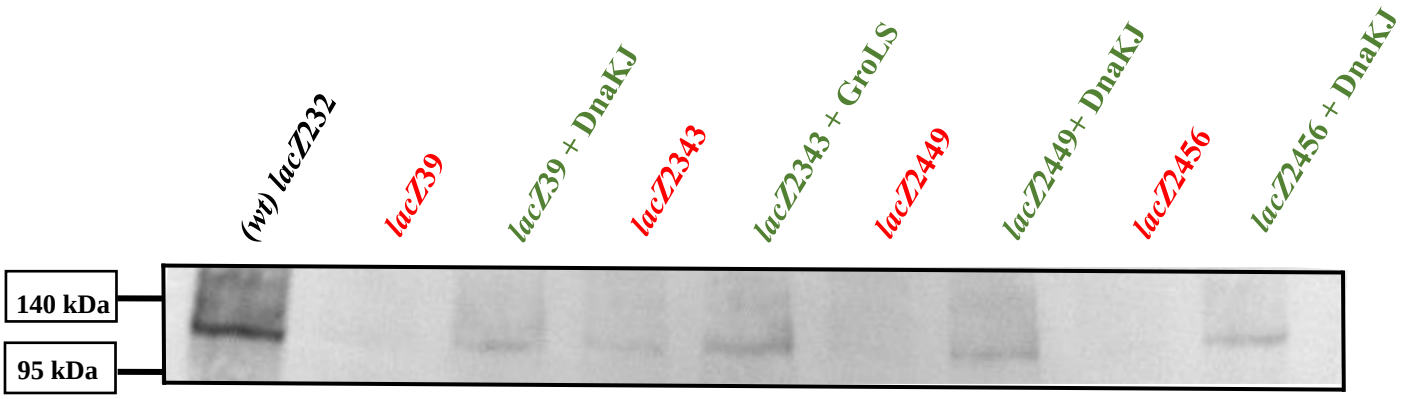

Figure 5.1: Western Blot. Protein β-galactosidase stability was determined by Western blot of wild-type *lacZ* (*lacZ*232), mutant *lacZ*39, *lacZ*39D, mutant *lacZ*2343, *lacZ*2343G, mutant *lacZ*2449, *lacZ*2449D, and mutant *lacZ*2456, and *lacZ*2456.

Table 5.1: Analysis of the effect of overproduction of either GroLS or DnaKJ and removal of either ClpX or Lon *lacZ mutant* strains

***lacZ* mutant number**	**Amino Acid Change**	**Mutant Activity**	**Activity with GroLS**	**% Change**	**Activity with DnaKJ**	**% Change**	**ClpX-KO Activity**	**% Change**	**Lon KO Activity**	**% Change**
39*	1018 Leu to Gln	8.17	3.59	-56.06 % ↓	140.27	+1617.49 % ↑	769.78	+3039.40 % ↑	1416.32	+ 5676.18 % ↑
337	905 Glu to Lys	2.47	2.08	-15.79 % ↓	1.83	-25.91 % ↓	2.71	+ 9.72 % ↑	4.62	+ 87.04 % ↑
361*	900 Gly to Asp	1.34	0.82	-39.03 % ↓	0.84	-37.57 % ↓	1.43	+ 6.72 % ↑	1.25	-6.72 % ↓
364*	208 Gly to Asp	1.70	0.95	-43.77 % ↓	1.23	-27.52 % ↓	1.37	-19.41 % ↓	6.24	+267.06 % ↑
2234*	565 Gly to Asp	1.51	0.52	-65.72 % ↓	1.03	-31.94 % ↓	1.46	-3.31 % ↓	1.22	-19.21 % ↓
2343*	943 Arg to His	120.85	157.28	+30.15 % ↑	10.45	-91.35 % ↓	15.96	-86.79 % ↓	697.56	+477.21 % ↑
2381	764 Gly to Asp	2.11	1.59	-24.48 % ↓	0.90	-57.33 % ↓	1.81	-14.22 % ↓	3.29	+55.92 % ↑
2382*	566 Gly to Asp	1.39	0.56	-59.95 % ↓	0.98	-29.08 % ↓	8.06	+479.86 % ↑	22.05	+1486.33 % ↑
2396	584 Gly to Asp	1.27	1.45	+14.25 % ↑	1.36	+7.27 % ↑	0.67	-47.24 % ↓	1.89	+48.82 % ↑
2449*	202 Asp to Asn	9.17	4.96	-45.89 % ↓	49.76	+442.64 % ↑	7.24	-21.05 % ↓	6.65	-27.48 % ↓
2454*	354 Gly to Asp	0.54	0.41	-23.19 % ↓	0.55	+2.35 % ↑	1.10	+103.70 % ↑	1.02	+88.89 % ↑
2456¥	5 Thr to Met	10.15	3.86	-61.98 % ↓	11.57	+13.99 % ↑	6.09	-40.00 % ↓	5.80	-42.86 % ↓
2530	419 His to Tyr	65.33	81.42	+24.63 % ↑	30.54	-51.43 % ↓	36.37	-44.33 % ↓	47.56	-27.20 % ↓
2540¥	4 Ile to Asn	6.11	4.04	-33.97 % ↓	3.45	-43.62 % ↓	3.77	-38.30 % ↓	4.17	-31.75 % ↓
2608	460 Gly to Arg	1.11	0.81	-26.59 % ↓	0.97	-12.47 % ↓	1.57	+41.44 % ↑	1.12	-0.90 % ↓

Table 5.1: Analysis of the effect of overproduction of either GroLS or DnaKJ and removal of either ClpX or Lon *lacZ mutant* strains (continued)

1. Only unique mutations are listed in the table. The *lacZ39* mutation was observed in 2 of the missense mutants, the *lacZ361* mutation was observed in 3 of the missense mutants, the *lacZ364* mutation was observed in 2 of the missense mutants, the *lacZ2234* mutation was observed in 2 of the missense mutants, the *lacZ2343* mutation was observed in 2 of the missense mutants, the *lacZ2382* mutation was observed in 6 of the missense mutants, the *lacZ2449* mutation was observed in 4 of the missense mutants, and the *lacZ2454* mutation was observed in 3 of the missense mutants.
2. The *lacZ* gene is 3,075 bp in length and codes for the 1,024 amino acid β-galactosidase protein. The resulting amino acid changes are given based on the coded protein predicted by the DNA sequence and not the Protein Data Bank file.
3. Mutant activities are listed with respect to the wild-type activity (100%). The assays were repeated in triplicate and the standard deviation was less than 10%.

Table 5.2: Analysis of the effect of overproduction of both GroLS and DnaKJ on *lacZ mutant* strains

***lacZ* mutant number**[1]	**Mutant Activity**	**Activity with DnaKJ**	**% Change**	**Activity with GroLS**	**% Change**	**Activity with DnaKJ+GroLS**	**% Change**
39*	8.17	140.27	+1617.49 % ↑	3.59	-56.06 % ↓	86.30	+ 956.30 % ↑
2343*	120.85	10.45	-91.35 % ↓	157.28	+30.14 % ↑	5.57	-95.39 % ↓
2396	1.27	1.36	+7.09 % ↑	1.45	+14.17 % ↑	1.05	-17.32 % ↓
2449*	9.17	49.76	+851.27 % ↑	4.96	-45.43 % ↓	117.19	+1177.97 % ↑
2454*	0.54	0.55	+2.35 % ↑	0.41	-23.19 % ↓	0.39	-27.78 % ↓
2456¥	10.15	11.57	+13.99 % ↑	3.86	-61.98 % ↓	80.03	+688.47 % ↑
2530	65.33	30.54	-51.43 % ↓	81.42	+24.63 % ↑	0.71	-98.91 % ↓

1. Only rescuable *lacZ* mutations are listed in the table, on which were tested the effect of overexpressing both chaperones.

Table 5.3: Analysis of the effect of overproduction of either GroLS or DnaKJ and removal of either ClpX or Lon on *xylE mutant* strains

***xylE* mutant number**[1]	**Amino Acid Change**[2]	**Mutant Activity**[3]	**Activity with GroLS**	**% Change**	**Activity with DnaKJ**	**% Change**	**ClpX-KO Activity**	**% Change**	**Lon KO Activity**	**% Change**
1568	106 Arg to Cys	29.11	27.08	-6.96 % ↓	35.22	+20.98 % ↑	55.49	+90.62 % ↑	55.73	+91.44 % ↑
1569	270 Gly to Gln	34.12	26.30	-22.92 % ↓	32.39	-5.10 % ↓	57.97	+69.88 % ↑	81.37	+138.45 % ↑
1571*	127 Gly to Gln	33.63	31.71	-5.71 % ↓	28.29	-15.88 % ↓	65.10	+93.60 % ↑	59.09	+75.72 % ↑
1574*	233 Ser to Phe	33.23	25.11	-24.45 % ↓	39.08	+17.61 % ↑	66.91	+101.34 % ↑	74.75	+124.93 % ↑
1577	21 Ala to Val	28.72	28.27	-1.56 % ↓	32.86	+14.43 % ↑	53.55	+86.48 % ↑	72.27	+151.68 % ↑
1581*	30 Gly to Ser	32.93	26.00	-21.05 % ↓	23.86	-27.54 % ↓	71.56	+117.32 % ↑	49.35	+49.87 % ↑
1582	64 Gly to Asp	33.58	29.46	-12.25 % ↓	27.42	-18.35 % ↓	70.46	+109.84 % ↑	46.91	+39.71 % ↑
1583*	158 Gly to Asp	34.42	24.28	-29.46 % ↓	24.06	-30.09 % ↓	57.18	+66.12 % ↑	56.17	+63.19 % ↑
1584	243 Pro to Ser	31.96	31.03	-2.93 % ↓	34.91	+9.22 % ↑	71.24	+122.88 % ↑	58.05	+81.61 % ↑
1585*	9 Gly to Ser	32.36	41.08	+26.95 % ↑	28.15	-13.00 % ↓	69.47	+114.66 % ↑	53.70	+65.93 % ↑
1587	178 Gln to Lys	26.62	22.01	-17.34 % ↓	30.94	+16.22 % ↑	63.91	+140.07 % ↑	41.39	+55.47 % ↑
1588	247 Gly to Asp	28.78	20.64	-28.30 % ↓	31.17	+8.29 % ↑	123.32	+328.45 % ↑	125.52	+336.09 % ↑
1589*	282 Trp to Ser	31.25	25.11	-19.65 % ↓	34.47	+10.29 % ↑	67.08	+114.62 % ↑	53.82	+72.20 % ↑
1591	8 Pro to Ser	31.40	27.96	-10.98 % ↓	26.54	-15.50 % ↓	66.83	+112.80 % ↑	72.50	+130.86 % ↑
1592	193 Ser to Asn	32.36	26.04	-19.54 % ↓	23.95	-25.99 % ↓	65.60	+102.72 % ↑	55.07	+70.18 % ↑
1593	245 Arg to Cys	31.69	26.77	-15.53 % ↓	30.62	-3.38 % ↓	64.72	+104.21 % ↑	51.47	+62.40 % ↑
1665	242 Gly to Asp	25.35	29.50	+16.38 % ↑	32.62	+28.69 % ↑	61.44	+142.37 % ↑	39.39	+55.38 % ↑
1666	251 Gly to Asp	33.78	36.00	+6.58 % ↑	35.87	+6.20 % ↑	65.89	+95.08 % ↑	69.43	+105.56 % ↑
1670*	202 Ala to Thr	26.17	23.21	-11.33 % ↓	27.91	+6.63 % ↑	73.89	+182.31 % ↑	59.21	+126.22 % ↑
1677	213 His to Tyr	24.04	24.10	+0.26 % ↑	34.75	+44.55 % ↑	72.15	+200.12 % ↑	56.37	+134.48 % ↑
1680	216 Ser to Phe	30.11	23.02	-23.53 % ↓	32.01	+6.32 % ↑	66.91	+122.22 % ↑	51.30	+70.38 % ↑
1681	200 Asp to Asn	32.08	27.66	-13.77 % ↓	28.02	-12.65 % ↓	66.05	+105.92 % ↑	46.48	+44.91 % ↑
1683	115 His to Tyr	32.65	27.20	-16.71 % ↓	42.02	+28.69 % ↑	110.94	+239.74 % ↑	108.19	+231.32 % ↑

Table 5.3: Analysis of the effect of overproduction of either GroLS or DnaKJ and removal of either ClpX or Lon on *xylE mutant* strains (continued).

***xylE* mutant number**[1]	**Amino Acid Change**[2]	**Mutant Activity**[3]	**Activity with GroLS**	**% Change**	**Activity with DnaKJ**	**% Change**	**ClpX-KO Activity**	**% Change**	**Lon KO Activity**	**% Change**
1684	137 Glu to Lys	37.02	25.29	-31.68 % ↓	44.67	+20.66 % ↑	126.73	+242.33 % ↑	99.31	+168.26 % ↑
1698	152 Asp to Asn	31.16	20.89	-32.96 % ↓	23.00	-26.19 % ↓	129.12	+314.42 % ↑	98.50	+216.14 % ↑
1702	24 His to Tyr	33.50	25.51	-23.85 % ↓	22.52	-32.79 % ↓	118.12	+252.62 % ↑	92.84	+177.15 % ↑
1710	290 Ala to Thr	35.14	25.26	-28.12 % ↓	42.34	+20.49 % ↑	116.58	+231.73 % ↑	93.29	+165.46 % ↑

1. Only unique mutations are listed in this table. The *xylE1571* mutation was observed in 2 of the missense mutants, the *xylE1574* mutation was observed in 2 of the missense mutants, the *xylE1581* mutation was observed in 3 of the missense mutants, the *xylE1583* mutation was observed in 2 of the missense mutants, the *xylE1585* mutation was observed in 3 of the missense mutants, the *xylE1589* mutation was observed in 2 of the missense mutants and the *xylE1670* mutation was observed in 2 of the missense mutants.
2. The *xylE* gene is 921 bp in length and codes for the 307-amino acid catechol 2,3-dioxygenase protein. The resulting amino acid changes are given based on the coded protein predicted by the DNA sequence and not the Protein Data Bank file.
3. Mutant activities are listed with respect to the wild-type activity (100%). The assays were repeated in triplicate and the standard deviation was less than 10%.

Table 5.4: Analysis of the effect of overproduction of both GroLS and DnaKJ on *xylE mutant* strains.

***xylE* mutant number**	**Amino Acid Change**	**Mutant Activity**	**Activity with GroLS**	**% Change**	**Activity with DnaKJ**	**% Change**	**Activity with DnaKJ+GroLS**	**% Change**
1568	106 Arg to Cys	29.11	27.08	-6.96 % ↓	35.22	+20.98 % ↑	40.77	+40.05 % ↑
1574*	233 Ser to Phe	33.23	25.11	-24.45 % ↓	39.08	+17.61 % ↑	26.23	-21.07 % ↓
1577	21 Ala to Val	28.72	28.27	-1.56 % ↓	32.86	+14.43 % ↑	23.81	-17.10 % ↓
1584	243 Pro to Ser	31.96	31.03	-2.93 % ↓	34.91	+9.22 % ↑	33.41	+4.54 % ↑
1585*	9 Gly to Ser	32.36	41.08	+26.95 % ↑	28.15	-13.00 % ↓	40.60	+25.46 % ↑
1587	178 Gln to Lys	26.62	22.01	-17.34 % ↓	30.94	+16.22 % ↑	30.04	+12.85 % ↑
1588	247 Gly to Asp	28.78	20.64	-28.30 % ↓	31.17	+8.29 % ↑	60.25	+109.35 % ↑
1589*	282 Trp to Ser	31.25	25.11	-19.65 % ↓	34.47	+10.29 % ↑	39.91	+27.71 % ↑
1665	242 Gly to Asp	25.35	29.50	+16.38 % ↑	32.62	+28.69 % ↑	30.72	+21.18 % ↑
1666	251 Gly to Asp	33.78	36.00	+6.58 % ↑	35.87	+6.20 % ↑	36.03	+6.66 % ↑
1670*	202 Ala to Thr	26.17	23.21	-11.33 % ↓	27.91	+6.63 % ↑	42.65	+62.97 % ↑
1677	213 His to Tyr	24.04	24.10	+0.26 % ↑	34.75	+44.55 % ↑	37.48	+55.91 % ↑
1680	216 Ser to Phe	30.11	23.02	-23.53 % ↓	32.01	+6.32 % ↑	82.69	+174.63 % ↑
1683	115 His to Tyr	32.65	27.20	-16.71 % ↓	42.02	+28.69 % ↑	35.39	+8.39 % ↑
1684	137 Glu to Lys	37.02	25.29	-31.68 % ↓	44.67	+20.66 % ↑	53.19	+43.68 % ↑
1710	290 Ala to Thr	35.14	25.26	-28.12 % ↓	42.34	+20.49 % ↑	60.54	+72.28 % ↑

1. Only rescuable *xylE* mutations are listed in the table, on which we tested the effect of overexpressing both chaperones.

Table 5.5: Protein stability of rescuable mutant strains

***LacZ* Strains**	**Protein stability**
*lacZ*39	2.88%
*lacZ*39D	17.29%
*lacZ*2343	3.87%
*lacZ*2343G	20.24%
*lacZ*2449	0.32%
*lacZ*2449D	19.54%
*lacZ*2456	1.50%
*lacZ*2456D	29.58%
***xylE* Strains**	**Protein stability**
xylE1568	17.31%
xylE1568D	46.24%
xylE1585	26.53%
xylE1585G	59.92%
xylE1589	41.05%
xylE1589D	62.52%
xylE1677	33.34%
xylE1677D	79.10%
xylE1683	32.88%
xylE1683D	50.60%
xylE1684	74.89%
xylE1684D	89.61%
xylE1710	58.09%
xylE1710D	76.96%

Protein stability was determined by Western blot analysis. The stability of the mutants are listed with respect to wild-type protein levels (100%).

Table 5.6: Analysis of mutants' rescuability and change in hydropathy relation

***lacZ* mutant number**	**Change in hydropathy**	**Chaperone correction**
39*	+*	DnaKJ
337	+	None
361*	+*	None
364*	+*	None
2234*	+*	None
2343*	-*	GroLS
2381	+*	None
2382*	+*	None
2396	+*	DnaKJ & GroLS
2449*	-	DnaKJ
2454*	+*	DnaKJ
2456¥	-	DnaKJ
2530	-	GroLS
2540¥	+*	None
2608	+*	None
***xylE* mutant number**	**Change in hydropathy**	**Chaperone correction**
1568	-*	DnaKJ
1569	+*	None
1571*	+*	None
1574*	-*	DnaKJ

Table 5.6: Analysis of mutants' rescuability and change in hydropathy relation (continued)

***xylE* mutant number**	**Change in hydropathy**	**Chaperone correction**
1577	-*	DnaKJ
1581*	+	None
1582	+*	None
1583*	+*	None
1584	+	DnaKJ
1585*	+	GroLS
1587	+	DnaKJ
1588	+*	DnaKJ
1589*	+*	DnaKJ
1591	+	None
1592	+	None
1593	-*	None
1665	+*	DnaKJ & GroLS
1666	+*	DnaKJ & GroLS
1670*	+	DnaKJ
1677	-	DnaKJ & GroLS
1680	-*	DnaKJ
1681	-	None
1683	-	DnaKJ
1684	+	DnaKJ
1698	-	None
1702	-	None
1710	+	DnaKJ

Table 5.6: Analysis of mutants' rescuability and change in hydropathy relation (continued)

The change in hydropathy was determined using the averaged hydrophobicity mentioned in chapter 3, Cole *et al*. (2018). (+) indicates a more hydrophilic change, (–) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydropathy of the original amino acid significantly and results in a shift of at least 5 amino acids

REFERENCES

Baba, M., Takai, Y., Mori, H., Tomita, M., Okumura, Y., Ara, T., . . . Hasegawa, M. (2006). Construction of escherichia coli K-12 in-frame, single-gene knockout mutants: The keio collection. *Molecular Systems Biology, 2*(1), 2006.000-n/a. doi:10.1038/msb4100050

Baker, T. A., & Sauer, R. T. (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. *BBA - Molecular Cell Research, 1823*(1), 15-28. doi:10.1016/j.bbamcr.2011.06.007

Bromberg, Y., & Rost, B. (2009). Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinformatics, 10 Suppl 8*(S8), S8. doi:10.1186/1471-2105-10-S8-S8

Cole, A., Hani, F., Allen, B., Kline, P., & Altman, E. (2018). Nonfunctional missense mutants in two well characterized cytosolic enzymes reveal important information about protein structure and function. *The Protein Journal, 37*(5), 407-427. doi:10.1007/s10930-018-9786-6

Cooper, G. M. (2000). *The cell; a molecular approach* (2. ed. ed.). Washington, DC: ASM Press.

Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science, 338*(6110), 1042-1046. doi:10.1126/science.1219021

Dunker, A. K., Silman, I., Uversky, V. N., & Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology, 18*(6), 756-764. doi:10.1016/j.sbi.2008.10.002

Flynn, J. M., Neher, S. B., Kim, Y., Sauer, R. T., & Baker, T. A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Molecular Cell, 11*(3), 671-683. doi:10.1016/S1097-2765(03)00060-1

Gupta, A. J., Haldar, S., Miličić, G., Hartl, F. U., & Hayer-Hartl, M. (2014). Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *Journal of Molecular Biology, 426*(15), 2739-2754. doi:10.1016/j.jmb.2014.04.018

Lee, I., & Suzuki, C. K. (2008). Functional mechanics of the ATP-dependent lon protease-lessons from endogenous protein and synthetic peptide substrates. *BBA - Proteins and Proteomics, 1784*(5), 727-735. doi:10.1016/j.bbapap.2008.02.010

Lu, R., Tan, M., Wang, H., Xie, A., Yu, J., & Tan, L. (2014). Heat shock protein 70 in alzheimer's disease. *BioMed Research International, 2014*, 1-8. doi:10.1155/2014/435203

Maillard, R., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C. M., . . . Bustamante, C. (2011). ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell, 145*(3), 459-469. doi:10.1016/j.cell.2011.04.010

Marchenkov, V. V., & Semisotnov, G. V. (2009). GroEL-assisted protein folding: Does it occur within the chaperonin inner cavity? *International Journal of Molecular Sciences, 10*(5), 2066-2083. doi:10.3390/ijms10052066

Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., & Gottesman, S. (1990). Sequence and structure of clp P, the proteolytic component of the ATP-dependent clp protease of escherichia coli. *Journal of Biological Chemistry, 265*(21), 12536-12545

Mizobata, T., & Kawata, Y. (2018). The versatile mutational "repertoire" of escherichia coli GroEL, a multidomain chaperonin nanomachine. *Biophysical Reviews, 10*(2), 631-640.

Porankiewicz, J., Wang, J., & Clarke, A. K. (1999). New insights into the ATP-dependent clp protease: Escherichia coli and beyond. *Molecular Microbiology, 32*(3), 449-458. doi:10.1046/j.1365-2958.1999.01357.x

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews. Molecular Cell Biology, 14*(10), 630-642. doi:10.1038/nrm3658

Shiber, A., & Ravid, T. (2014). Chaperoning proteins for destruction: Diverse roles of Hsp70 chaperones and their co-chaperones in targeting misfolded proteins to the proteasome. *Biomolecules, 4*(3), 704-724. doi:10.3390/biom4030704

Tripathi, L. P., & Sowdhamini, R. (2008a). Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genomics, 9*(1), 549. doi:10.1186/1471-2164-9-549

Tripathi, L. P., & Sowdhamini, R. (2008b). Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genomics, 9*(1), 549. doi:10.1186/1471-2164-9-549