

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

By

Ashley Elliott Cole

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Molecular Bioscience

Middle Tennessee State University

August 2017

Thesis Committee:

Dr. Elliot Altman, Chair

Dr. Anthony L. Farone

Dr. Mary Farone

Dr. Paul Kline

Dr. Preston MacDougall

I dedicate this research to my loving and supportive husband, Stan Cole,
my son, Michael Rhett Cole, and to my parents, John and Susan Elliott.

Without your continuous encouragement and reassurance this dream
would never have been possible.

ACKNOWLEDGEMENTS

I would like to thank the Molecular Biosciences (MOBI) doctoral program, the Biology Department at Middle Tennessee State University (MTSU), and TRIAD for providing the funding for this study. I would also like to thank my advisor, Elliot Altman, for putting up with my crazy antics for the last 6 years. You have been so much more than just a mentor, but also a friend and a confidant. And to my committee, each one of you has been such an influence in my life. Without your help and support, I would not have been able to make this dream a reality.

ABSTRACT

Missense mutants can have devastating effects on proteins and usually act by causing perturbations in the folding of the protein that affects their final tertiary or quaternary forms. In order to better understand how missense mutants affect protein structure, a set of randomly generated nonfunctional missense mutants were isolated in two well-characterized enzymes, β -galactosidase and catechol 2,3-dioxygenase. This collection of missense mutants yielded important information regarding the key structural elements within correctly folded proteins and the ability of protein predictive tools to predict the effects of missense mutants. Many of the missense mutants were found to be salt-correctible and the rescuability of the missense mutants by various salts correlated with the Hofmeister series of ions that affect protein stability. In an additional study the promiscuous or broad-acting sumA 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 was characterized.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1. INTRODUCTION	1
1.1 Protein folding	1
1.2 Amino Acid Hydropathy	5
1.3 Protein Prediction Models	6
1.4 Hofmeister Series and Protein Stabilization	7
1.5 Effect of Missense Mutations	8
1.6 β -galactosidase and catechol 2,3-dioxygenase	9
1.7 Dissertation Synopsis	13
REFERENCES	15
CHAPTER 2: THE PROMISCUOUS SUMA MISSENSE SUPPRESSOR FROM <i>SALMONELLA ENTERICA</i> HAS AN INTRIGUING MECHANISM OF ACTION	20
2.1 Abstract	21
2.2 Introduction	21
2.3 Materials and Methods	24
2.3.1 Media, bacterial strains and plasmids	24
2.3.2 Determining the mutation responsible for sumA	27
2.3.3 Construction of pACYC184- <i>xylE</i>	29
2.3.4 Construction of the proB:: <i>xylE</i> (cat) insertion	30
2.3.5 Generating <i>lacZ</i> and <i>xylE</i> missense mutants	31
2.3.6 Suppression tests	32
2.3.7 β -galactosidase and catechol 2,3-dioxygenase enzyme assays	33
2.3.8 Inheritance test of rifampicin resistant mutants isolated in a sumA missense suppressor	34
2.3.9 Growth rate studies	35
2.3.10 Data availability	35
2.4 Results	35
2.4.1 Determining the mutation responsible for the sumA missense suppressor	35
2.4.2 Characterizing the efficiency and specificity of the sumA missense suppressor	38

2.4.3 Analyzing the missense mutants with codon changes other than glycine to aspartic acid that are suppressible by sumA	45
2.4.4 Inheritance test of mutations generated by sumA	48
2.4.5 The sumA missense suppressor causes a significant reduction in the growth rate	49
2.5 Discussion	50
REFERENCES	58
 CHAPTER 3: LESSONS LEARNED ABOUT PROTEIN STRUCTURE AND FUNCTION FROM A COLLECTION OF INACTIVE MISSENSE MUTANTS IN TWO WELL CHARACTERIZED AQUEOUS CYTOSOLIC PROTEINS.....	
3.1 Abstract	65
3.2 Introduction	65
3.3 Materials and Methods	68
3.3.1 Media and bacterial strains	68
3.3.2 Construction of pET-11a- <i>xylE</i>	68
3.3.3 Purification of catechol 2,3-dioxygenase protein from pET-11a- <i>xylE</i>	69
3.3.4 Generating and sequencing <i>lacZ</i> and <i>xylE</i> missense mutants	70
3.3.5 β -galactosidase and catechol 2,3-dioxygenase enzyme assays	70
3.3.6 Preparation of rabbit anti β -galactosidase and catechol 2, 3-dioxygenase antibody	70
3.3.7 Western Blot Analysis	71
3.3.8 Three dimensional analysis of the <i>lacZ</i> and <i>xylE</i> missense mutants	72
3.3.9 Analysis of the potential effects of the <i>lacZ</i> and <i>xylE</i> 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	72
3.3.10 Analysis of the changes in hydropathy caused by the <i>lacZ</i> and <i>xylE</i> missense mutants	74
3.3.11 Analysis of the potential effects of the missense mutants on the secondary structure of β -galactosidase or catechol 2, 3-dioxygenase.....	75
3.4 Results	75
3.4.1 Generating a collection of inactive missense mutants in <i>lacZ</i> and <i>xylE</i>	75
3.4.2 Analysis of the missense mutants with respect to the location of the mutation within the protein versus the change in hydropathy	83
3.4.3 Analysis of the distribution of the missense mutants in α -helical, β -sheet or coil secondary structures	84

3.4.4 Analysis of the predicted effect of the missense mutants on secondary structures using α -helical, β -sheet or coil propensity scales.....	84
3.4.5 Analysis of the predicted effect of the missense mutants on secondary structures using algorithms that predict the presence of secondary structure.....	85
3.5 Discussion	100
REFERENCES.....	104
CHAPTER 4: A STUDY OF THE ABILITY OF SALTS.....	110
TO STABILIZE PROTEINS <i>IN VIVO</i>	110
4.1 Abstract	111
4.2 Introduction.....	112
4.3 Materials and Methods.....	114
4.3.1 Media and bacterial strains.....	114
4.3.2 Salt suppression plate tests of the β -galactosidase and catechol 2,3-dioxygenase missense mutants.....	114
4.3.3 β -galactosidase and catechol 2,3-dioxygenase enzyme assays	115
4.3.4 Three dimensional analysis of the β -galactosidase and catechol 2,3-dioxygenase proteins	115
4.4 Results.....	116
4.4.1 Determining the ability of different salts to stabilize the β -galactosidase and catechol 2,3-dioxygenase enzymes intracellularly.	116
4.4.2 Determining the ability of salt to suppress β -galactosidase and catechol 2,3-dioxygenase missense mutants	118
4.4.3 Determining the ability of different salts to stabilize the β -galactosidase enzyme produced by the <i>lacZ39</i> missense mutant.....	129
4.5 Discussion	131
REFERENCES.....	135

LIST OF TABLES

Table 2.1. <i>S. enterica</i> strains	25
Table 2.2. Plasmids	26
Table 2.3. Three factor cross results	36
Table 2.4: Characterization of <i>lacZ</i> missense mutants that are suppressible by sumA	40
Table 2.5: Characterization of <i>xylE</i> missense mutants that are suppressible by sumA	41
Table 2.6. Missense mutant neighboring aspartic acid	47
Table 2.7. Data on allele specific versus broad acting missense suppressors.....	52
Table 3.1. Average scale for the propensities of amino acids to be found in the three secondary structures, α -helices (P_α), β -sheets (P_β) and coils (P_c)	73
Table 3.2. <i>lacZ</i> missense mutants	77
Table 3.3. <i>xylE</i> missense mutants	79
Table 3.4. Chou-Fasman, GOR, and Qian-Sejowski propensity scores for <i>lacZ</i> mutant secondary structures	86
Table 3.5. JPred4, PSIPRED, Porter 4.0, and Spider ² consensus scores for <i>lacZ</i> mutant secondary structures	87
Table 3.6. Chou-Fasman, GOR, and Qian-Sejowski propensity scores for <i>xylE</i> mutant secondary structures	88
Table 3.7. JPred4, PSIPRED, Porter 4.0, and Spider ² consensus scores for <i>xylE</i> mutant secondary structures	90
Table 3.8. Chou-Fasman, GOR, and Qian-Sejowski accuracy for <i>lacZ</i> secondary structures	94
Table 3.9. JPred4, PSIPRED, Porter 4.0, and Spider ² accuracy for <i>lacZ</i> secondary structures ...	95
Table 3.10. Chou-Fasman, GOR, and Qian-Sejowski accuracy for <i>xylE</i> secondary structures ...	96
Table 3.11. JPred4, PSIPRED, Porter 4.0, and Spider ² accuracy for <i>xylE</i> secondary structures .	97
Table 4.1. Hofmeister Salt Tables for wild-type β -galactosidase and catechol 2,3-dioxygenase	117
Table 4.2. <i>lacZ</i> mutant activity in the presence of 0.2M or 0.3M NaCl	122
Table 4.3. <i>xylE</i> activity in the presence of 0.2M or 0.3M NaCl	123
Table 4.4. β -galactosidase missense mutants.....	125
Table 4.5. Catechol 2,3-dioxygenase missense mutants.....	127
Table 4.5. Hofmeister Salt Tables for mutant β -galactosidase (<i>lacZ39</i>)	130

LIST OF FIGURES

Figure 1.1. Myoglobin structure.	3
Figure 1.2. Hen egg-white lysozyme and α -chymotrypsin structures.	4
Figure 1.3. β -galactosidase hydrolysis of lactose into glucose and galactose.	10
Figure 1.4. β -galactosidase hydrolysis of X-gal into galactose and the indole resulting in a blue color.	10
Figure 1.6. β -galactosidase structure.	11
Figure 1.7. Linear representation of β -galactosidase	11
Figure 1.8. Catechol 2,3-dioxygenase extradiol cleavage of catechol to form 2-hydroxymuconate semialdehyde.....	12
Figure 1.9 Catechol 2,3-dioxygenase structure.....	13
Figure 1.10. Linear representation of catechol 2,3-dioxygenase	13
Figure 2.1. Linkage map of the <i>sumA</i> , <i>amiB</i> and <i>purA</i> genes in <i>S. enterica</i>	37
Figure 2.2. Suppression of <i>lacZ</i> missense mutants by the <i>sumA</i> missense suppressor	43
Figure 2.3. Suppression of <i>xylE</i> missense mutants by the <i>sumA</i> missense suppressor	44
Figure 3.1. Western blot of wild-type <i>lacZ</i> and mutants of varying amounts of β -galactosidase.....	82
Figure 4.1. Suppression of <i>lacZ</i> missense mutants by 0.3M NaCl.....	119
Figure 4.2. Suppression of <i>xylE</i> missense mutants by 0.3M NaCl.....	120

CHAPTER 1. INTRODUCTION

1.1 Protein folding

Protein folding is a quintessential aspect of molecular biology and ultimately cellular life. To function properly, most proteins must assemble into a tightly folded and accurate 3-D structure (Dobson, 2004). In a crowded cellular environment and guided by the order of their amino acids, proteins must navigate a complex network of mechanisms meant to assist folding or degrade misfolding proteins. In addition, many physical factors contribute to the protein folding which include: hydrogen bonds, van der Waals interactions, backbone angle preferences, electrostatic interactions, hydrophobic/hydrophilic interactions, and chain entropy.

While, according to the laws of thermodynamics, systems tend to move towards the states of lowest free energies. Proteins fold by taking random steps incrementally downhill in energy, and each step needs only be favorable by one to two times the thermal energy to reach the native structure rapidly. Ultimately, protein folding can funnel to a single stable state by multiple routes in conformational space. (Dill, 1997 and Zwanzig *et al.*, 1992).

In the early 1930's the terminology for protein secondary structures were originally coined by William Astbury, α -form and β -form. Astbury discovered a drastic change in the x-ray fiber diffraction of stretched moist wool or hair fibers in which the un-stretched protein molecules formed a helix, α -form, while stretching revealed an uncoiled and more extended state, β -form (Astbury, 1932). Although his models were correct in principle and correspond to the modern element of protein secondary structure, in detail they were inaccurate. However, Astbury's

nomenclature was retained 20 years later when in 1951 Linus Pauling, Robert Corey, and H.R. Branson established two hydrogen-bonded structures, α -helix and β -sheet (Pauling, *et al.*, 1951).

Instead of trying to tackle the entire structure of a protein, Linus Pauling and Robert Corey centered their efforts on bits of proteins and elucidating the structure of each configuration one by one. Secondary structures were first realized by the use of early protein crystallography studies by Pauling *et al.* (1951) who were able to determine the ability of an α -helix to promote stability due to the correct accommodation of hydrogen bonds. Pauling and Corey's α -helix and β -sheet have been found in almost all proteins. Although Pauling and Corey's work was not perfect and since the 1951 paper, it has been found that beta sheets are actually twisted and not flat as originally suggested. In addition, the pi helix, with 5.1 residues per turn, has yet to be found in nature. This lack of discovery has led many to question its existence at all.

In 1962, John Kendrew and Max Perutz were awarded the Nobel Prize in Chemistry for their pioneering work in using x-ray crystallography to determine the globular protein structure of myoglobin and hemoglobin, respectively. In 1958, Kendrew *et al.* determined the globular protein structure of myoglobin at 6Å resolution. And shortly after in 1959, Perutz determined the structure of hemoglobin. Kendrew *et al.* (1958) said, "Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure. Though the detailed principles of construction do not yet emerge, we may hope that they will do so at a later stage of the analysis."

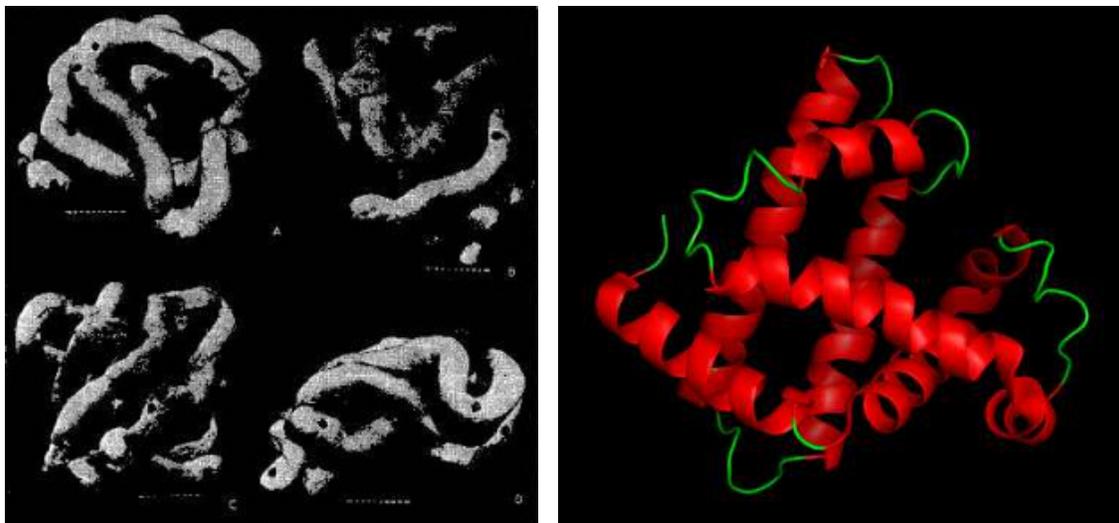


Figure 1.1. Myoglobin structure. The image on the left is the original photo of the myoglobin structure by Kendrew *et al.* (1958). The polypeptide chains are indicated in white. The right image is myoglobin from PyMOL (PDB code 1MBN). The Protein Data Bank was not established until 1971 and in 1973, Watson and Kendrew deposited 1MBN which is a 2.0Å myoglobin structure.

In 1965, David Phillips *et al.* published the high resolution structure of hen egg-white lysozyme. This was the first time that an enzyme structure was solved by X-ray crystallography. The detail in this molecular structure stimulated the suggestion of chemical mechanism for the enzymatic reaction that this protein catalyzed. The next high resolution structures included α -chymotrypsin (Matthews *et al.*, 1967), ribonuclease (Wyckoff *et al.*, 1967) carboxypeptidase (Lipscomb *et al.*, 1969) and *Staphylococcus aureus* nuclease (Arnone *et al.*, 1971).

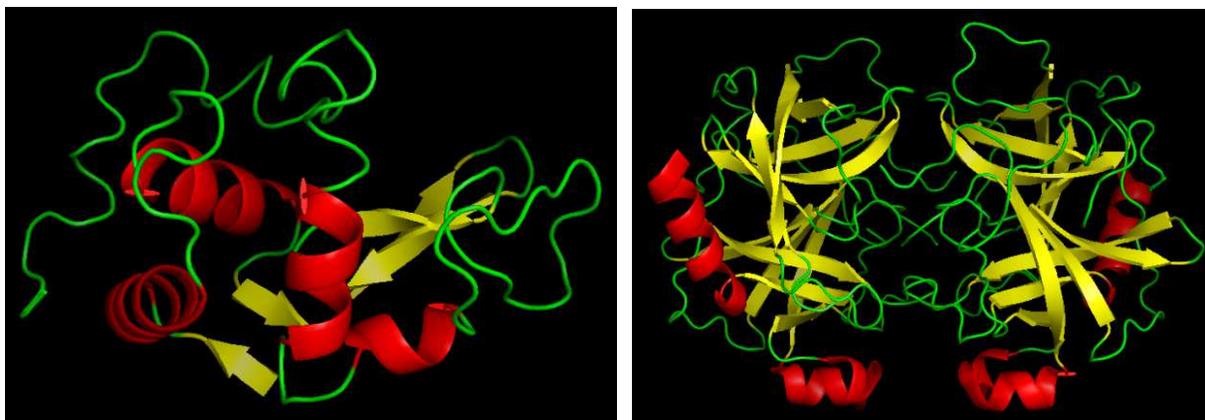


Figure 1.2. Hen egg-white lysozyme and α -chymotrypsin structures. Image on the left is hen egg-white lysozyme from PyMOL (PDB code: 2LYZ), deposited by Diamond, R. and Phillips, D.C. in 1975. Image on the right is α -chymotrypsin at 2Å from PyMOL (PDB code: 2CHA), deposited by Birktoft, J. J and Blow, D. M. in 1972.

Eight years passed before in 1973 when Anfinsen made the discovery of a denatured protein's ability to spontaneously self-assemble back into their native conformation that it was understood that the three-dimensional structure is ultimately determined by the amino acid sequence.

Research has since focused on trying to elicit the molecular mechanism by which proteins fold. Ultimately, it is imperative to recognize the role of each single amino acid which composes the secondary structures and ultimately the final three dimensional structures. A protein's secondary structure is vital to the final configuration of the structural integrity of the protein (Kwok, 2002).

As the three dimensional structures of more and more proteins were determined, two general facts regarding the structures of proteins became clear; first, the propensities of hydrophilic amino acids to be located on or near the surface of the proteins and hydrophobic amino acids to be buried in the protein and second, the importance of α -helices and β -sheets, which have been found in every protein.

1.2 Amino Acid Hydropathy

Several properties play a role in an amino acid's propensity to create a secondary structure. One of more important dominant forces in protein folding is an amino acids' hydropathy. It is generally accepted that two opposing tendencies are reflected in the final folded protein structure. Hydrophilic side-chains access the aqueous solvent while minimizing the hydrophobic side-chains contact with the aqueous environment.

Many studies have been performed in order to create a scale in which amino acids could be classified as hydrophobic or hydrophilic. From the first hydrophobicity scale that was formulated in 1962 by Tanford and Lovrien, to today, there are about 100 hydrophobicity scales. As expected, there is a high variance between them due to the underlying experimental approaches used to determine these scales. This makes it extremely problematic for users to select a suitable hydropathy scale. Several of the well-known scales, Kyte and Doolittle, Janin, Rose *et al.*, Hopp-Woods, and Eisenberg scale all differ in their methods and ultimately disagree in their individual amino acid arrangements. In 1982, Kyte and Doolittle published one of the more well-known hydropathy scales that systematically evaluated the tendencies of a polypeptide chains' amino acids to be hydrophilic or hydrophobic. The program continually determines the average hydropathy of a moving segment as it progresses through to the carboxy terminus. It is also been found useful for predicting transmembrane stretches in proteins. However, several scales such as Janin, 1979 and Rose *et al.* 1985, examined proteins with known structures and defined the hydrophobicity as the tendency for a residue to be found buried within a protein rather than exposed on the surface. These scales have cysteine as the most hydrophobic amino acid as it forms disulfide bonds that occur inside a globular structure. The Hopp-Woods

scale is a hydrophilic index used to identify potential antigenic sites in a protein where nonpolar residues have been assigned a negative value (Hopp and Woods, 1983). The Eisenberg scale is a normalized consensus hydrophobicity scale which shares numerous features with other scales. This is achieved by using the mean hydrophobicity against the mean hydrophobic moment to create a so-called normalized consensus scale (Eisenberg *et al.*, 1984).

1.3 Protein Prediction Models

Since protein folding is dictated by the amino acid sequence and as there are only 20 natural amino acids, it would appear to be a simple task to perform accurate protein structure predictions. However, this is not the case. Currently, the most successful structure-prediction model algorithms are based upon the assumption that similar sequences lead to similar structures.

Two of the earliest and more traditional models, Chou-Fasman developed in 1974 and GOR (Garnier-Osguthorpe-Robson) developed in 1978, are two algorithm based techniques that take the propensities of certain amino acids to produce secondary structures. However the GOR model also takes into account the neighboring amino acid conditional probabilities. In 1988, Qian and Sejnowski developed a neural network prediction model which could use pattern matching abilities that neural networks are designed for using already known structures.

In order to advance the prediction models, in 1994, the Critical Assessment of protein Structure Prediction (CASP) was initiated by Moult and colleagues. In 1994 there were only 224 unique protein folds known and at the last CASP there were over 87,000 structures in the Protein databank which span about 1393 folds. Currently, there have been 12 CASP meetings since its inception (one every 2 years). And several newer models have been developed/tested at CASP

in order to try and more accurately predict secondary structures. Some of these models include JPred4, Porter 4.0, PSIPRED (PSI-blast based secondary structure PREDiction), and SPIDER². The Jpred4 uses the JPred Protein Secondary Structure Server that delivers predictions by the JNet algorithm (Drozdetskiy *et al.*, 2015). Porter 4.0 is a prediction model based on an ensemble of 25 BRNNs (bidirectional recurrent neural networks) (Mirabello, Pollastri, 2013). PSIPRED's algorithm is split into 3 stages: 1) generate a sequence profile generalized by PSIBLAST, 2) predict initial secondary structures using neural networking, and 3) filter predicted structure by generating the highest score for each secondary structure element (McGuffin *et al.*, 2000). SPIDER² is a prediction model that uses iterative Deep Neural Network (DNN) for protein structural properties (Heffernan *et al.*, 2016).

1.4 Hofmeister Series and Protein Stabilization

It is well-known that various environmental factors can impact macromolecules, including proteins. A few of these factors include pH, temperature, and the presence of ions. In living organisms, biomolecular interactions, protein solubility, and protein stability after unfolding can all be altered by the presence of salts according to this unwavering tendency (Baldwin, 1996).

Before the advent of protein chromatography, protein salting out was the main method used in order to purify proteins. In 1888, Hofmeister developed a qualitative ranking of anions towards their ability to precipitate a mixture of hen egg white proteins. This study would become a foundation for many following studies into protein stability. Since Hofmeister's study, numerous studies have focused on elucidating the mechanistic action of these interactions and these have shown that many proteins will have different effects due to certain anion exposure.

Hydrogen bonds play a dominant role in the formation and maintenance of protein structure. However it has also been shown that nonpolar interactions as well as hydrogen bonds play an essential role in protein structure. Devoid of these bonding potentials, neutral salts can also perturb secondary-tertiary protein structures. It has been found that ions can be ranked in order of their increasing effectiveness in lowering the protein melting temperature and ultimately their effectiveness to salt-in or salt-out proteins (Von Hippel and Wong 1964). In 1958, Flory and Garrett showed that melting phenomena were simply the reversible helix to coil transitions in the rigid collagen macromolecule. And in 1962, Hamaguchi and Geiduschek applied the effect of various ions at high concentrations on the temperature of the helix-coil transition of DNA. In 1979, Kohno and Roth made the observations that several neutral salts were effective with NaCl and KCl at rescuing temperature-sensitive mutants (the optimal concentration was 0.2M). Although the neutral salts can easily be ranked in their order of effectiveness on protein stability, there are variances that occur between different proteins.

1.5 Effect of Missense Mutations

Not all proteins will fold into their final, active structure. It is well known that missense mutations cause cellular complications, especially regarding protein folding. These mutations can lead to protein destabilization, decrease the folding rate or even increase the misfolding in such a way that the energetics are no longer practical (Hutt, 2009). According to the Human Genome Mutation Database, missense mutations account for over half (~56%) of all DNA mutations that are known to cause genetic disease (Krawczak *et al.*, 2000 and Stenson *et al.*, 2009). In numerous cases, disease-causing missense mutations affect the ability of the protein to either fold correctly and/or the stability of the folded structure. Three ways that mutations can

affect protein folding include: 1) impairing the folding process which increases degradation of folding intermediates, 2) causing alternative folding to yield stable nonfunctional structures that aggregate and may disturb cellular homeostasis, 3) misfolding proteins into conformations that jeopardize the further assembly of the structural unit. It has been well established that missense mutations can have devastating effects on proteins and generally act by causing perturbations in the folding of the protein that affects their final tertiary or quaternary forms.

1.6 β -galactosidase and catechol 2,3-dioxygenase

The two enzymes we used for our study, β -galactosidase and catechol 2,3-dioxygenase, from the *lacZ* and *xylE* gene respectively, are two well-studied and characterized proteins. Both of these enzymes have fairly simple, robust enzymatic activity assays as well as their 3-dimensional structures are known. These basic factors make these two enzymes ideal for our study.

Exploration of the lactose (*lac*) system in *Escherichia coli* was started in the late 1940s by Jacques Monod at the Pasteur Institute in Paris. The main product of the *lac* operon is β -galactosidase. β -galactosidase hydrolyzes lactose, and other β -galactosides, into monosaccharides (see Figure 1.3). It also has a self-regulating mechanism where it can convert lactose to allolactose which will bind to the *lacZ* repressor. This protein has become a very useful and helpful reporter gene in many studies of gene expression. β -galactosidase is well known for its reaction with X-gal (5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside), a soluble colorless compound consisting of galactose linked to a substituted indole. When X-gal is hydrolyzed, an intense blue color results due to the release of the substituted indole that spontaneously dimerizes (see Figure 1.4).

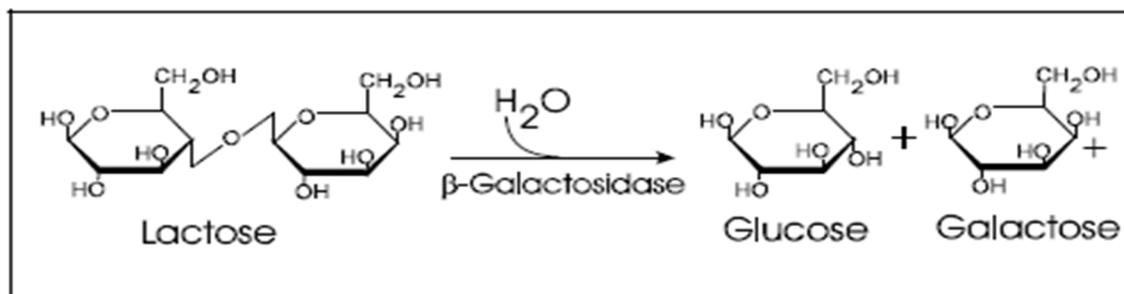


Figure 1.3. β -galactosidase hydrolysis of lactose into glucose and galactose.

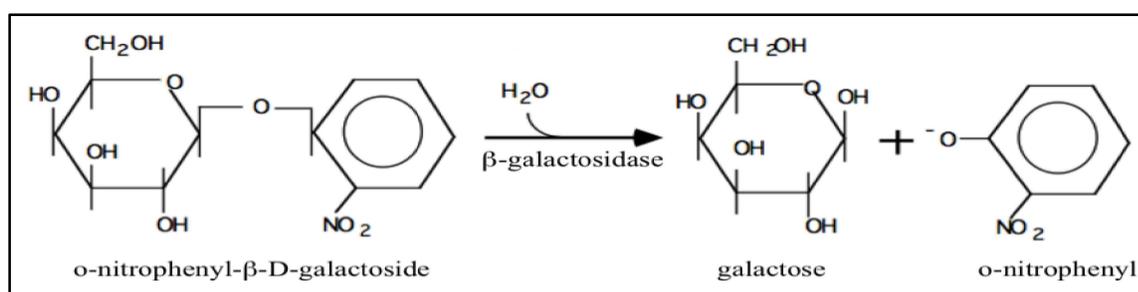


Figure 1.4. β -galactosidase hydrolysis of X-gal into galactose and the indole resulting in a blue color.

The functional form of β -galactosidase is a tetramer consisting of four identical subunits each with 1,023 amino acid residues. β -galactosidase has 5 compact domains (Appel *et al.*, 1965 and Fowler *et al.*, 1978). Each subunit has an active site and the tetramer is the only active form of the enzyme. It is within the third, or central, domain where the TIM (triose phosphate isomerase) or $\alpha_6\beta_8$ barrel with the active site forms a deep pit at the C-terminal end of this barrel. Using PyMOL, the tetramer's dimensions are roughly 174 x 136 x 87Å.

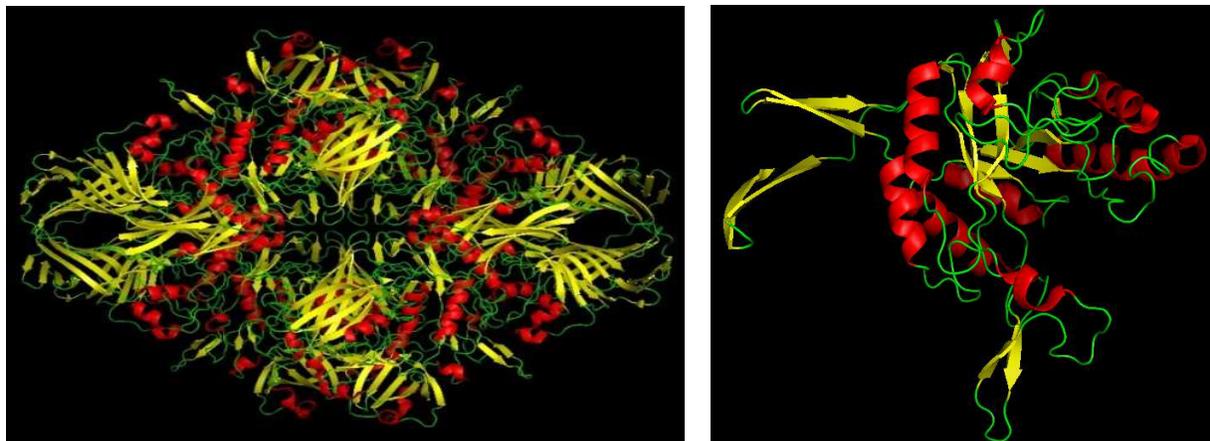


Figure 1.6. β -galactosidase structure. Image on the left is the β -galactosidase structure at 1.7Å (PDB code: 1DP0). Image on the right is the third domain of β -galactosidase. Images generated by PyMOL and α -helices are red, β -sheets are yellow, and coils are green.

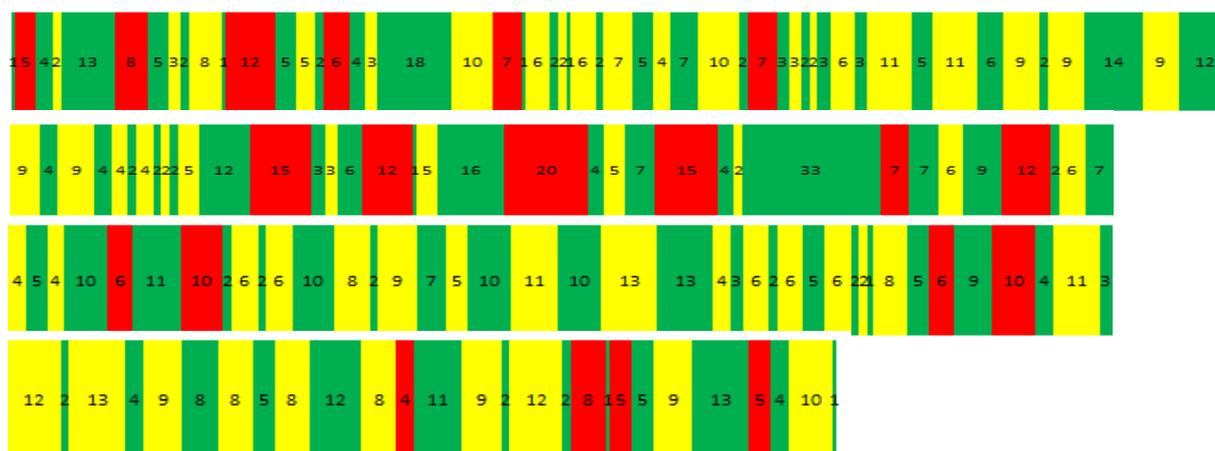


Figure 1.7. Linear representation of β -galactosidase. α -helices (total of 20) shown in red, β -sheets (total of 58) shown in yellow, and coils (total of 77) shown in green.

Extradiol type-dioxygenases are found in a variety of bacteria and involve in aromatic ring fission at *meta* position of dihydroxylated aromatics, intermediates of catabolic pathways of monocyclic and polycyclic compounds. *Pseudomonas putida* is a soil bacterium which has the ability to grow with biphenyl being the sole carbon and energy source (Furukawa *et al.*, 1989).

One extradiol-type dioxygenase, catechol 2,3-dioxygenase, catalyzes the meta cleavage of aromatic rings on catechol formed by metabolic pathways of monocyclic and polycyclic compounds including naphthalene and biphenyl. Catechol 2,3-dioxygenase catalyzes an extradiol cleavage of catechol to form 2-hydroxymuconate semialdehyde with the insertion of two atoms of oxygen (see Figure 1.8).

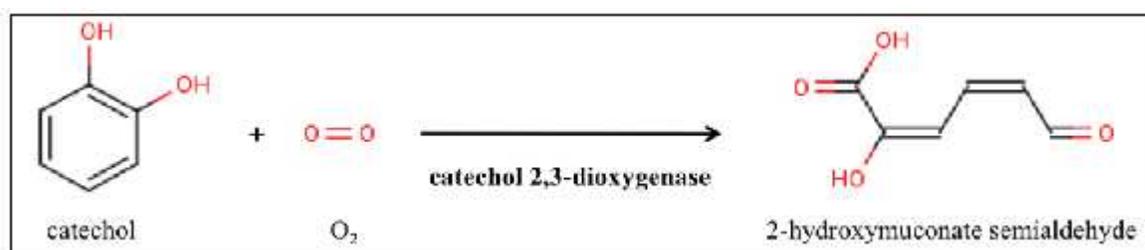


Figure 1.8. Catechol 2,3-dioxygenase extradiol cleavage of catechol to form 2-hydroxymuconate semialdehyde.

Catechol 2,3-dioxygenase is also a tetramer consisting of four identical subunits each with 307 amino acid residues (Kita *et al.*, 1999). The homotetramer prefers small monocyclic substrates and the active site is located in a funnel-shaped space of the C-terminal domain. This site also contains an FE atom directly ligated by three protein residues, His153, His214, and Gly265. The conserved active site residues are His 199 and Tyr255. (Ishida *et al.*, 2002). Using PyMOL, the tetramers dimensions are roughly 95 X 65 X 51Å.

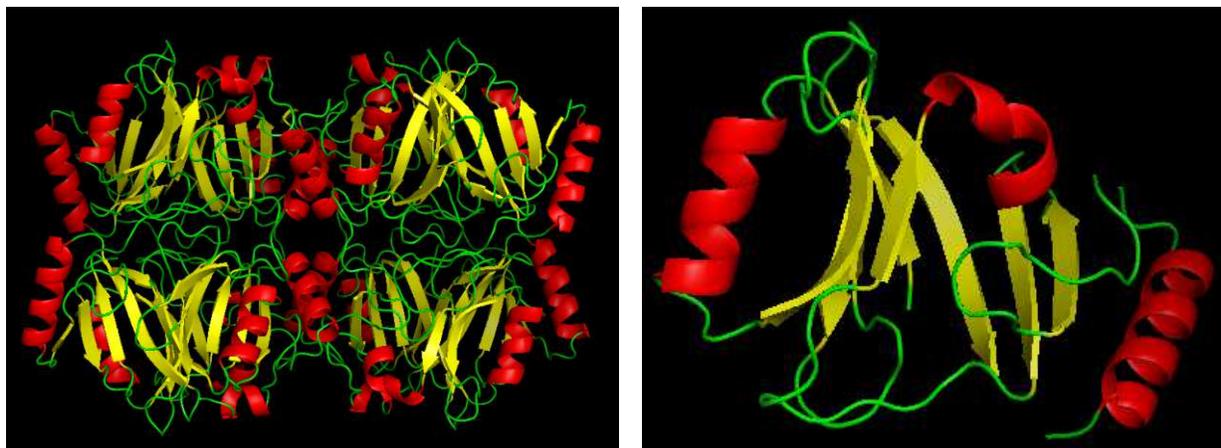


Figure 1.9 Catechol 2,3-dioxygenase structure. Image on the left is Catechol 2,3-dioxygenase at 2.0Å (PDB code:1MPY). Image on the right shows the location of the active site of catechol 2,3-dioxygenase. Images generated by PyMOL and α -helices are red, β -sheets are yellow and coils are green.

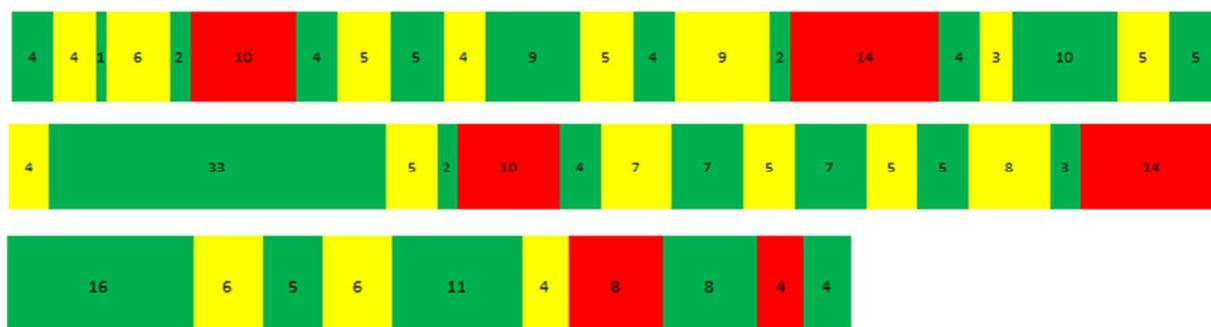


Figure 1.10. Linear representation of catechol 2,3-dioxygenase. α -helices (total of 6) shown in red, β -sheets (total of 17) shown in yellow, and coils (total of 23) shown in green.

1.7 Dissertation Synopsis

For this study, we generated a collection of 70 independent nonfunctional missense mutants, 30 in the *lacZ* gene and 40 in the *xylE* gene, which were sequenced to yield 15 unique mutants in *lacZ* and 27 unique mutants in *xylE*. This mutant collection was used to research how a broad acting missense suppressor functioned, how missense mutants affected protein structure and the

ability of protein predictive tools to anticipate the effects of the missense mutants on protein structure, and the ability of salts based on the Hofmeister series of ions to stabilize proteins.

REFERENCES

- Affinsen CB, Haber E, Sela M, White FH Jr. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA*. 47:1309-1314.
- Appel SH, Alpers DH, Tomkins GM. 1965. Multiple molecular forms of β -galactosidase. *J Mol Biol*. 11:12-22.
- Arnone A, Bier A, Cotton A, Day A, Hazen A, Richardson A, Yonath A, Richardson A. 1971. A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. I. Experimental procedures and chain tracing. *J. Biol. Chem*. 246:2302–2316.
- Astbury WT, Street A. 1932. X-ray studies of the structures of hair, wool, and related fibres. I. General. *Philosophical Transactions, Series A*. 230:75-101.
- Baldwin RL. 1996. How Hofmeister ion interactions affect protein stability. *Biophys. Jour*. 71:2056-2063.
- Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*. 206(4986):757-61.
- Chou PY, Fasman GD. 1974. Prediction of protein conformation. *Biochemistry*. 13(2):222–245.
- Dill KA. 1997. From Levinthal to pathway to funnels. *Nat. Struct. Biol*. 4(1):10-19.
- Dobson CM. 2004. Principles of protein folding, misfolding and aggregation. *Seminars in Cell and Developmental Biology*. 15:3-16.

- Drozdetskiy A, Cole C, Procter J, Barton GJ. (2015). JPred4: a protein secondary structure prediction server. *Nucl. Acids Res.* 43(W1):W389-W394.
- Eisenberg D, Schwarz E, Komaromy M, Wall R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol.* 179(1):125-142.
- Flory P, Garrett RR. 1958. Phase transitions in collagen and gelatin systems. *J. Am. Chem. Soc.* 80(18):4836-4845.
- Fowler AV, Zabin I. 1978. Amino acid sequence of β -galactosidase. *J Biol Chem.* 253:5521-5535.
- Furukawa K, Hayase N, Taira K, Tomizuka N. 1989. *Pseudomonas putida* KF715 *bphABCD* operon encoding biphenyl and polychlorinated biphenyl degradation: cloning, analysis, and expression in soil bacteria. *J. Bacteriol.* 171:5467-5472
- Galitski T, Roth JR. 1996. A search for a general phenomenon of adaptive mutability. *Genetics.* 143:645-659.
- Garnier J, Osguthorpe D, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol.* 120: 97-120.
- Hamaguchi K, Geiduschek EP. 1962. The effects of electrolytes on the stability of the deoxyribonucleate helix. *J. Am. Chem. Soc.* 84(8):1329-1338.
- Heffernan R, Dehzangi A, Lyons J, Paliwal K, Sharma A, Wang J, Sattar A, Zhou Y, Yang Y. 2016. Highly accurate sequence-based prediction of half-sphere exposures of amino acid residues in proteins. *Bioinformatics.* 32(6):843-849.
- Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, Sattar A, Yang Y, Zhou Y. 2015. Improving prediction of secondary structure, local backbone angles, and solvent

- accessible surface area of proteins by iterative deep learning. *Scientific Report*. (5):11476.
- Hofmeister F, Zur Lehre von der Wirkung der Salze. 1888. *Arch. Exp. Pathol. Pharmacol.* (Leipzig) 24 (1888) 247-260; translated in W. Kunz, J. Henle and B. W. Ninham, ' Zur Lehre von der Wirkung der Salze' (about the science of the effect of salts: Franz Hofmeister's historical papers, *Curr. Opin. Colloid Interface Sci.* 9(2004):19-37.
- Hopp TP, Woods KR. 1983. A computer program for predicting protein antigenic determinants. *Mol Immunol.* 20(4):483-489.
- Hughes KT, Roth JR, Olivera BM. 1991. A genetic characterization of the *nadC* Gene of *Salmonella typhimurium*. *Genetics.* 127:657-670.
- Hutt DM, Powers ET, Balch WE. 2009. The proteostasis boundary in misfolding diseases of membrane traffic. *Federation of European Biochemical Societies (FEBS Letters)* 583. 2639-2646.
- 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.
- Janin J. 1979. Surface and inside volumes in globular proteins. *Nature.* 277(5696):491-492.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* 181:662-666.
- Kita A, Kita S, Fujisawa I, Inaka K, Ishida T, Horiike K, Nozaki M, Miki K. 1999. An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3 dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure.* 15(1):25-34.

- Kohno T, Roth JR. 1979. Electrolyte effects on the activity of mutant enzymes *in vivo* and *in vitro*. *Biochemistry*. 18:1386
- Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, Thomas N, Cooper DN. 2000. Human gene mutation database – a biomedical information and research resource. *Hum mutat*. 15 (1): 45-51.
- Kwok SC, Mant CT, Robert HS. 2002. Importance of secondary structural specificity determinants in protein folding: Insertion of a native β -sheet sequence into an α -helical coiled-coil. *Protein Sci*. 11(6):1519-1531.
- Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol*. 157:105-132.
- Lipscomb WN, Hartsuck JA, Quioco FA, Reeke GN. Jr. 1969. The structure of carboxypeptidase A. IX. The x-ray diffraction results in the light of the chemical sequence. *Proc. Natl Acad. Sci. USA*. 64:28–35.
- Matthews BW, Sigler PB, Henderson R, Blow DM. 1967. Three-dimensional structure of tosyl- α -chymotrypsin. *Nature*. 214:652–656
- McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics*. 16:404-405.
- Mirabello C, Pollastri G. 2013. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics*. 29(16):2056-2058.
- Pauling L, Corey RB, Branson HR. 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad.Sci. U.S.A.* 37:205-211.

- Qian N, Sejnowski TJ. 1988. Prediction the secondary structure of globular proteins using neural network models. *J. Mol. Biol.* 202:865-884.
- Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science.* 229(4716):834-838.
- Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. 2009. The human gene mutation database – 2008 update. *Genome Med* 1(1)13.
- Tanford C, Lovrien R. 1962. Dissociation of catalase into subunits. *Joural American Chem Soc* 84:1892-1896.
- Von Hippel PH, Wong KY. 1964. Neutral salts: the generality of their effects on the stability of macromolecular conformations. *Science.* 145(3632):577-580.
- Watson HC. 1969. The stereo chemistry of the protein myoglobin. *Prog. Stereochem.* 4:299.
- Wyckoff HW, Hardman, KD, Allewell NM, Inagami T, Johnson LN, Richards FM. 1967. The structure of ribonuclease-S at 3.5 Å resolution. *J. Biol. Chem.* 242(17):3984–3988.
- Zwanzig R, Szabo A, Baghci B. 1992. Levinthal's paradox. *Proc Natl Acad Sci.* 89(1):20-22.

**CHAPTER 2: 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

Published February 2017 in Genetics:

Cole AE, Hani FM, Altman R, Meservy M, Roth JR, 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 promiscuous or 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 (Hughes *et al.*, 1991; Galitski and Roth, 1996). *sumA* or

su537, like most missense suppressors, was originally isolated as an extragenic suppressor of the *hisC537* missense mutant (Whitfield *et al.*, 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 and 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, 1985; Murgola, 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 and Yanofsky, 1963; Roberts and 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 (Brody and Yanofsky, 1963; Berger and Yanofsky, 1967; Hill *et al.*, 1970; Hill *et al.*, 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 and Yanofsky, 1963; Eggertsson and 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 and 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, 1972) was used as rich or minimal defined media, respectively. The antibiotics chloramphenicol and tetracycline were used at a final concentration of 20 $\mu\text{g}/\text{mL}$ in LB, while rifampicin was used at a final concentration of 50 $\mu\text{g}/\text{mL}$ in LB. Ampicillin was used at 100 $\mu\text{g}/\text{mL}$ in LB to provide selective pressure for Amp^R plasmids or at 30 $\mu\text{g}/\text{mL}$ 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.

Table 2.1. *S. enterica* strains

Name	Genotype	Source
ALS234	<i>hisC537 zjf-3693::Tn10dTet</i> (40% linked to <i>sumA</i>)	this study
ALS1442	<i>proB::xylE(cat)</i>	this study
ALS2241	<i>btuB12::Tn10dCam hisC537 sumA10 zjf-3693::Tn10dTet</i>	this study
ALS2242	<i>btuB12::Tn10dCam hisC537 zjf-3693::Tn10dTet</i>	this study
ALS2583	<i>amiB211::mudA hisC537 purA155</i>	this study
GT2086	<i>zjf-U130::Tn10dTet</i> (99% linked to <i>sumA</i>)	Björk laboratory
GT467	<i>amtB211::mudA (amiB211::mudA)</i>	Björk laboratory
LT2	wild-type	Roth laboratory
TR3359	<i>hisC537</i>	Roth laboratory
TR4780	<i>hisC537 purA155</i>	Roth laboratory
TT2337	<i>hisC527(UAG) leuA414(UAG) supF (UAG, tyr) zde-94::Tn10</i> (50% linked to <i>supF</i>)	Roth laboratory
TT2344	<i>hisC527(UAG) leuA414(UAG) supE (UAG, gln) zbf-604::Tn10</i> (51% linked to <i>supE</i>)	Roth laboratory
TT2839	<i>hisC527(UAG) leuA414(UAG) tyrU90 (supM) zii-614::Tn10</i> (45% linked to <i>supM</i>)	Roth laboratory
TT4029	<i>hisO1242 hisB2135 supU1283(UGA) zhb-736::Tn10</i> (10% linked to <i>supU</i>)	
TT7610	<i>supD501 (UAG, ser) zeb-609::Tn10</i> (60% linked to <i>supD</i>)	Roth laboratory
TT13029	<i>hisC527(UAG) leuA414(UAG) supC80(UAA, UAG, tyr) zde-605::Tn10</i> (54% linked to <i>supC</i>)	Roth laboratory
TT16237	<i>hisC537 sumA10 zjf-3693::Tn10dTet</i> (40% linked to <i>sumA</i>)	Roth laboratory
TT18519	<i>ara-9 hisC10081::MudF(lac+)</i>	Roth laboratory
TT20702	<i>btuB12::Tn10dCam</i> (30% linked to <i>rpoB</i>)	Roth laboratory

Table 2.2. Plasmids

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

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::Tn10dTet* 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::Tn10dCam* 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, $\Delta(\textit{araABOIC-leu})7679 \textit{araD139 hsr- hsm+ galU galK} \Delta(\textit{lac})X74 \textit{rpsL}$, Casadaban and Cohen, 1980.

2.3.2 Determining the mutation responsible for *sumA*

Because the *zjf-UI30::Tn10dTet* insertion was 99% linked to *sumA*, it was expected to be less than 500 bases from the mutation that caused *sumA*. To determine the mutation that caused *sumA*, the tetracycline resistance of the Tn10dTet insertion which was 99% linked to *sumA* was replaced with chloramphenicol resistance from pACYC184 (Chang and Cohen, 1978) to facilitate cloning of the region surrounding the Tn10dTet insertion, using the lambda Red recombination system (Yu *et al.*, 2000; Datsenko and Wanner, 2000). Using the forward primer 5'CTGATGAATCCCCTAATGATTTTGGTAAAAATCATTAAGTTAAGGTGGATTGAGA AGCACACGGTCACA 3' and the reverse primer 5' CTGATGAATCCCCTAATGATTTTGGTAAAAATCATTAAGTTAAGGTGGATTACCTGT GACGGAAGATCAC 3', a 1,163 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 Tn10dTet (del16 del17 TetR, Way *et al.*, 1984) and chloramphenicol resistance from pACYC184. The homology for the pACYC184 chloramphenicol resistance from 419 – 3,601 bp is underlined. The 1,163 bp fragment was gel isolated and electroporated into LT2 pKM201 cells that were prepared as described by Murphy and 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 *Sau3AI* to generate 5,000 bp fragments and ligated into pTrc99A (Amann *et al.*, 1988), which had been digested with *Bam*HI and dephosphorylated with calf intestinal alkaline phosphatase. 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 5,000 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 Tn10dTet insertion that was 99% linked to *sumA* was determined to be at bp 4,596,266 of the *S. enterica* chromosome (McClelland *et al.*, 2001, 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 Tn10dTet 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* (Guest and Yanofsky, 1965; Fleck and Carbon, 1975). 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 *glyV* 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' TGAAGTGGCAACGCTCGAAT 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 *et al.*, 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 *PstI* and *HindIII* and then ligated into the pTrc99A vector which had been digested with the same two restriction enzymes. To construct pACYC184-*xylE*, forward primer 5' ATATCCATAAAGCTTCGCCGACATCATAACGGTTC 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 *HindIII* and *BclI* 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' ATGAGTGACAGCCAGACGCTGGTCGTAAACTCGGCACCAGCGTGCTAACGCCGACATCATAACGGTTCT 3' and the reverse primer 5' TTATCGAGTAATCATGTCATCACGATGAACAGCGACCGGGCCATATTCATTACCTGTGACGGAAGATCAC 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 and Campellone, 2003 and plated on LB chloramphenicol plates at 37°C to kick the pKM201 plasmid. Recombinant colonies were

streaked on LB chloramphenicol plates at 37⁰C 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 (*hisC10081::MudF[lac+]*) or ALS1442 (*proB::xylE[cat]*) were mutagenized with ethylmethane sulfonate (EMS) as described by Miller, 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 *et al.*, 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 *hisC10081::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 *hisC10081::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 and Fasman, 1989), Garnier-Osguthorpe-Robson (Garnier *et al.*, 1978) or Qian-Sejnowski (Qian and Sejnowski, 1988) protein secondary structure analysis, is predicted to be a neutral change.

2.3.6 Suppression tests

Suppression tests were conducted using amber, ochre or opal nonsense suppressor or *sumA* missense suppressor strains, which also harbored a *Tn10* or *Tn10dTet* 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 μ g/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 (Miller, 1972). 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 and Evans (1971).

To conduct the catechol 2,3-dioxygenase assays 100 μ L of a bacterial overnight was added to 3.0 mL of Z buffer that contained 0.27% sodium dodecyl sulfate. 50 μ L 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 μ L 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 (OD368). *XylE* units were determined using the formula, $10,000 \times (\text{OD368 of the 2-hydroxymuconate semialdehyde}) / (\text{T} \times \text{V} \times \text{OD550 of the concentrated cells})$, 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::Tn10dCam* transposon insertion were plated on LB plates that contained rifampicin. *btuB12::Tn10dCam* 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::Tn10dCam* 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::Tn10dCam* 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 OD550 readings were taken every 15 minutes until the OD550 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 OD550 reading, while the doubling times were calculated as $\ln 2/\text{slope}$.

2.3.10 Data availability

All of the data required for confirming the conclusions presented in the article are represented fully within the article. All of the bacterial strains and plasmids that were used in this study are available upon request.

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 and 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 2.1.

Table 2.3. Three factor cross results

Recombinant class	Total number out of 300
<i>sumA amiB::mudA</i>	8
<i>sumA amiB+</i> (wt)	112
<i>sumA+</i> (wt) <i>amiB::mudA</i>	63
<i>sumA+</i> (wt) <i>amiB+</i> (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%.

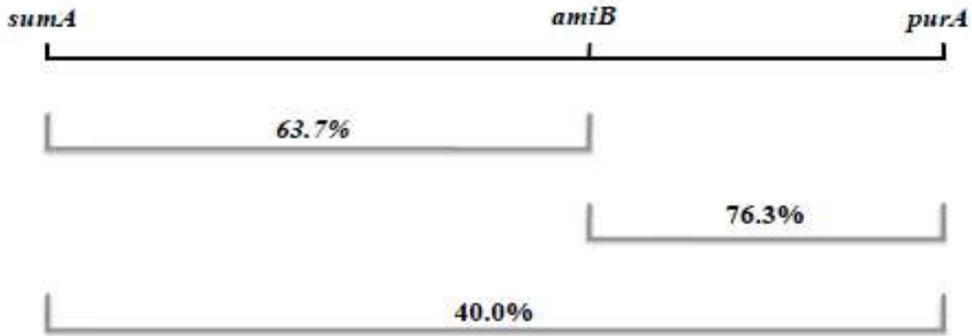


Figure 2.1. Linkage map of the *sumA*, *amiB* and *purA* genes in *S. enterica*. The gene order of the *sumA*, *amiB* and *purA* genes are shown and the linkages based on the data from Table 2.1 are included.

Because the *sumA* mutation was highly linked to the *amiB* locus, we screened a collection of Tn10dTet transposon insertions from the Björk laboratory that mapped to the *amiB* locus and identified a *zjf-U130::Tn10dTet* insertion that was 99% linked to *sumA* (data not shown). The location of the *zjf-U130::Tn10dTet* 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.*, 2001, 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 (Guest and Yanofsky, 1965; Fleck and Carbon, 1975). 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 Greb *et al.*, 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.

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

Mutant ¹	Base change ²	Amino acid change ²	Phenotype ³	Activity Miller units ⁴	with the <i>sumA</i> missense suppressor			
					Phenotype ³	Activity Miller units ⁴	Fold Increase	Percentage of wild-type activity ⁵
<i>lacZ2234</i>	1694 (G to A)	565 (G to D)	W	3.74	LB	99.65	26.64	2.71
<i>lacZ2343</i>	2828 (G to A)	943 (R to H)	VLB	9.78	LB	146.96	15.03	3.87
<i>lacZ2381</i>	2291 (G to A)	764 (G to D)	VLB	6.87	LB	89.17	12.98	2.32
<i>lacZ2382</i>	1697 (G to A)	566 (G to D)	W	3.84	LB	109.67	28.56	2.99
<i>lacZ2396</i>	1643 (G to A)	548 (G to D)	VLB	4.31	LB	65.55	15.21	1.73
<i>lacZ2454</i>	1061 (G to A)	354 (G to D)	W	3.86	LB	122.09	31.63	3.34
<i>lacZ2504</i>	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 (*hisC10081::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*

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

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 4 and Table 5 regarding the ability of the *sumA* missense suppressor to rescue *lacZ* or *xylE* missense mutants are shown in Figures 2.2 and 2.3.

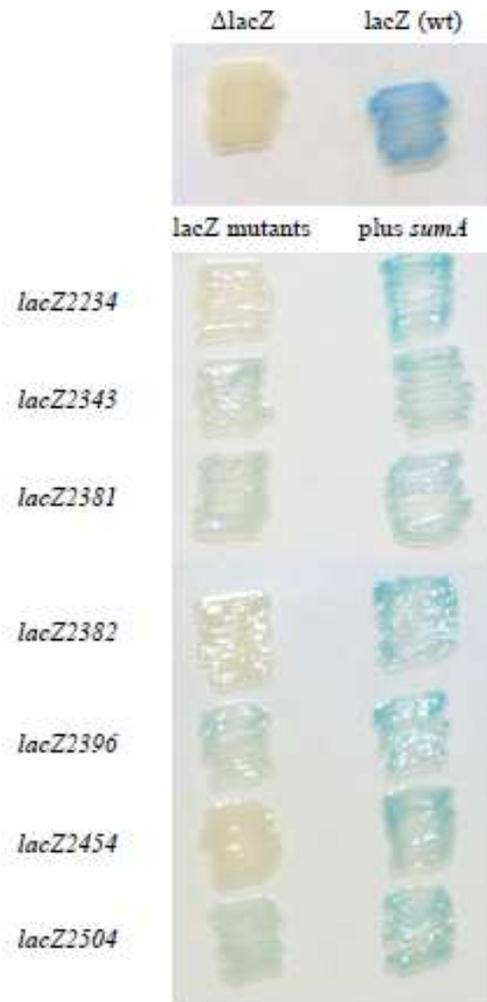


Figure 2.2. Suppression of *lacZ* missense mutants by the *sumA* missense suppressor. *lacZ* missense mutants with or without the *sumA* missense suppressor were patched onto LB, 1 mM IPTG, 40 ug/mL X-Gal plates and incubated at 37°C for 16 hours. Wild-type *S. enterica* (*lacZ*-) and *S. enterica* with the *hisC10081::MudF[lac+]* insertion are included as controls.



Figure 2.3. Suppression of *xylE* missense mutants by the *sumA* missense suppressor. *xylE* missense mutants with or without the *sumA* missense suppressor were patched onto LB, 1 mM IPTG plates and incubated at 37°C for 16 hours after which the plates were sprayed with a light mist of 100 mM catechol. Wild-type *S. enterica* (*xylE*-) and *S. enterica* with the *proB::xylE(cat)* insertion are included as controls.

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 11, 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 *lacZ*₂₄₅₄ 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.

Table 2.6. Missense mutant neighboring aspartic acid

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

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

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 *et al.*, 1990; Slupska *et al.*, 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* (Michaels *et al.*, 1990; Slupska *et al.*, 1996; Murphy and Humayun, 1997; Al Mamun *et al.*, 1999; Al Mamun *et al.*, 2002; Dorazi *et al.*, 2002; Balashov and Humayun, 2004), 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 1×10^8 cells in wild-type *S. enterica* and 13.39 per 1×10^8 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 promiscuous or 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 it's *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 (Brody and Yanofsky, 1963; Berger and Yanofsky, 1967; Hill *et al.*, 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 and 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 (Whitfield *et al.*, 1966; Hughes *et al.*, 1991) 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.

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
<i>glyT</i> tRNA Gly2(GGA/G)	<i>glyT</i> tRNA Gly2(AGA) ³	narrow	GGA (7/1000)	AGA (2/1000)	3.31%
			GGG (9/1000)		
<i>glyU</i> tRNA Gly1(GGG)	<i>glyU</i> tRNA Gly1(GAG) ⁴	broad	GGG (9/1000)	GAG (19/1000)	3.31%
<i>glyV</i> tRNA Gly3(GGU/C)	<i>glyV</i> tRNA Gly3(GAU/C) ⁵	broad	GGU (28/1000)	GAU (33/1000)	6.76%
			GGC (30/1000)	GAC (23/1000)	
<i>serU</i> tRNA Ser2(UCG)	<i>serU</i> 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

The data on the frequency of codon usage is from Maloy *et al.* (1996) and the data on the abundance of the tRNAs is from Dong *et al.*, 1996. If one considers all the factors that determine whether a missense suppressor might be predicted to be broad acting, the data in Table 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 and 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 rescue 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 *et al.*, 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.

A total of 6 out 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. Seven 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 and 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 (Michaels *et al.*, 1990; Slupska *et al.*, 1996; Murphy and Humayun, 1997; Al Mamun *et al.*, 1999; Al Mamun *et al.*, 2002; Dorazi *et al.*, 2002; Balashov and Humayun, 2004). 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 *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.

REFERENCES

- Al Mamun AA, Marians KJ, Humayun MZ. 2002. DNA polymerase III from *Escherichia coli* cells expressing *mutA* mistranslator tRNA is error-prone. *J Biol Chem.* 277:46319–46327.
- Al Mamun AA, Rahman MS, Humayun MZ. 1999. *Escherichia coli* cells bearing *mutA*, a mutant *glyV* tRNA gene, express a *recA*-dependent error-prone DNA replication activity. *Mol Microbiol.* 33:732-740.
- Amann E, Ochs B, Abel KJ. 1988. Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene.* 69:301–315.
- Balashov S, Humayun MZ. 2004. Specificity of spontaneous mutations induced in *mutA* mutator cells. *Mutat Res.* 548:9-18.
- Berger H, Yanofsky C. 1967. Suppressor selection for amino acid replacements expected on the basis of the genetic code. *Science.* 156:394-397.
- Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol.* 62:293-300.
- Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, , Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science.* 277:1453-1462.
- Brody S, Yanofsky C. 1963. Suppressor gene alteration of protein primary structure
Genetics. 50:9-16.

- Casadaban M, Cohen SN. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* 138:179-207
- Chang ACY, Cohen SN. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J Bacteriol.* 134:1141-1156.
- Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA.* 97:6640–6645.
- 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. *Proc Natl Acad Sci USA.* 89:1885-1889.
- Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol.* 260:649–663.
- Dorazi R, Lingutla JL, Humayun MZ. 2002. Expression of mutant alanine tRNAs increases spontaneous mutagenesis in *Escherichia coli*. *Mol Microbiol.* 44:131–141.
- Eggertsson G. 1968. Suppressors causing temperature sensitivity of growth in *Escherichia coli*. *Genetics.* 60:269-280.
- Eggertsson G, Adelberg EA. 1965. Map positions and specificities of suppressor mutations in *Escherichia coli* K-12. *Genetics.* 52:319-40.
- Fleck EW, Carbon J. 1975. Multiple gene loci for a single species of glycine transfer ribonucleic acid. *J Bacteriol.* 122:492-501.
- Galitski T, Roth JR. 1996. A search for a general phenomenon of adaptive mutability. *Genetics.* 143:645-659.

- Garnier JD, Osguthorpe J, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol.* 120:97-120.
- Greeb J, Atkins F, Loper C. 1971. Histidinol dehydrogenase (*hisD*) mutants of *Salmonella typhimurium*. *J Bacteriol.* 106:421-431.
- Guest JR, Yanofsky C. 1965. Amino acid replacements associated with reversion and recombination within a coding unit. *J Mol Bio.* 12:793-804.
- Harayama S, Rekik M, Bairoch A, Neidle EL, Ornston LN. 1991. Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal *benABC* and *Pseudomonas putida* TOL pWVO plasmid *xylXYZ*, genes encoding benzoate dioxygenases. *J Bacteriol.* 173:7540-7548.
- Hill CW. 1975. Informational suppression of missense mutations. *Cell.* 6:419-427.
- Hill CW, Combriato G, Dolph W. 1974. Three different missense suppressor mutations affecting the tRNA^{GGG_{Gly}} species of *Escherichia coli*. *J Bacteriol.* 117:351-359
- Hill CW, Squires C, Carbon J. 1970. Glycine transfer RNA of *Escherichia coli*: I. Structural genes for two glycine tRNA species. *J Mol Biol.* 52:557-569.
- Hughes KT, Roth JR, Olivera BM. 1991. A genetic characterization of the *nadC* Gene of *Salmonella typhimurium*. *Genetics.* 127:657-670.
- Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, Matthews BW. 2000. High resolution refinement of beta-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation. *Protein Sci.* 9:1685-1699.

- Kita A, Kita S, Fujisawa I, Inaka K, Ishida T, Horiike K, Nozaki M, Miki K. 1999. An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure Fold Des.* 7:25-34.
- Maloy SR, Stewart VJ, Taylor RK. 1996. Genetic analysis of pathogenic bacteria: a laboratory manual. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
- McClelland M Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature.* 413:852-856.
- Michaels ML, Cruz C, Miller JH. 1990. *mutA* and *mutC*: Two mutator loci in *Escherichia coli* that stimulate transversions. *Proc Natl Acad Sci USA.* 87:9211-9215.
- Miller JH, 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Minuchehr Z, Goliaei B. 2005. Propensity of amino acids in loop regions connecting beta-strands. *Protein Pept Lett.* 12:379-382
- Murgola EJ. 1985. tRNA, suppression, and the code. *Ann Rev Genet.* 19:57-80.
- Murgola EJ. 1995. Translational suppression: when two wrongs do make a right tRNA: Structure, Biosynthesis, and Function. Edited by Dieter Soli and Uttam RajBhandary. American Society for Microbiology, Washington, DC

- Murphy HS, Humayun MZ. 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. *J Bacteriol.* 179:7507–7514.
- Murphy KC, Campellone KG. 2003. Lambda red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol Biol.* 4:11.
- Prevelige P Jr, Fasman GD. 1989. Chou-Fasman prediction of the secondary structure of proteins: the Chou-Fasman-Prevelige algorithm in Prediction of protein structure and the principles of protein conformation (Fasman GD, editor). Plenum, New York, 391-416.
- Qian N, Sejnowski TJ. 1988. Predicting the secondary structure of globular proteins using neural network models. *J Mol Biol.* 202:865-884.
- Roberts JW, Carbon J. 1975. Nucleotide sequence studies of normal and genetically altered glycine transfer ribonucleic acids from *Escherichia coli*. *J Biol Chem.* 250:5530-5541
- Sala-Tepat JM, Evans WC. 1971. The *meta* cleavage of catechol by *Azotobacter* species: 4-oxalocrotonate pathway. *Eur J Biochem.* 20:400-413.
- Sanderson KE, Hartman PE. 1978. Linkage map of *Salmonella typhimurium*, edition V. *Microbiol Rev.* 42:471-519
- Schmieger H. 1972. Phage P22 mutants with increased or decreased transduction activities. *Mol Gen Genet.* 119:75-88.
- Shestopalov BV. 1988. Amino acid sequence template useful for α -helix-turn- α -helix prediction. *FEBS Lett.* 1:105-108.
- Slupska MM, Baikalov C, Lloyd R, Miller JH. 1996. Mutator tRNAs are encoded by the *Escherichia coli* mutator genes *mutA* and *mutC*: a novel pathway for mutagenesis. *Proc Natl Acad Sci USA.* 93:4380-4385.

- Squires C, Carbon J, Hill CW. 1970. Glycine transfer RNA of *Escherichia coli*. I. Structural genes for two glycine tRNA species. *J Mol Biol.* 52:557-569.
- Su JY, Belmont L, Sclafani RA. 1990. Genetic and molecular analysis of the *SOE1* gene: A tRNA(3Glu) missense suppressor of yeast *cdc8* mutations. *Genetics.* 124:523-531.
- Thorbjarnardóttir S, Uemura H, Dingermann T, Rafnar T, Thorsteinsdóttir S, Söll D, Eggertsson G. 1985. *Escherichia coli supH* suppressor: temperature-sensitive missense suppression caused by an anticodon change in tRNA^{Ser2}. *J Bacteriol.* 161:207-211.
- Way JC, Davis MA, Morisato D, Roberts DE, Kleckner N. 1984. New Tn10 derivatives for transposon mutagenesis and for construction of *lacZ* operon fusions by transposition. *Gene.* 32:369-79.
- Whitfield HJ Jr, Martin RG, Ames BN. 1966. Classification of aminotransferase (C gene) mutants in the histidine operon. *J Mol Biol.* 21:335-355
- Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci USA.* 97:5978-5983.

**CHAPTER 3: LESSONS LEARNED ABOUT PROTEIN
STRUCTURE AND FUNCTION FROM A COLLECTION
OF INACTIVE MISSENSE MUTANTS IN TWO WELL
CHARACTERIZED AQUEOUS CYTOSOLIC PROTEINS**

Ashley E. Cole¹, Fatmah M. Hani¹, Brian W. Allen³, Paul C. Kline², 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

*- corresponding author, ealtman@mtsu.edu

Key words

Protein secondary structure, α -helices, β -sheets, coils

3.1 Abstract

The isolation and characterization of 42 unique inactive missense mutants in the bacterial aqueous cytosolic *β -galactosidase* and *catechol 2,3-dioxygenase* proteins allowed us to examine some of the basic general trends regarding protein structure and function. 20 out of the 42, or 47.62% 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. 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 45.25% of the missense mutants were in the unstructured coiled, turn or loop regions. While α -helical and β -sheets are undeniably important in protein structure, our results clearly indicated that the unstructured regions were just as important. There was generally good consensus among the widely used algorithms to predict the presence of secondary structures and most of them predicted that the great majority of the 42 inactive missense mutants would impact the secondary structures.

3.2 Introduction

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 and Street, 1931; Astbury and Woods, 1934; Astbury and Sisson, 1935). 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 *et al.*, 1951; Pauling and Corey, 1951). Then using more advanced and refined X-ray diffraction techniques

the first α -helical structures were identified in myoglobin by Kendrew *et al.*, 1958 and in haemoglobin by Perutz *et al.*, 1960 and the first β -sheet structure was identified in lysozyme by 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.

The first scale delineating hydrophobic versus hydrophilic amino acids was generated by Tanford and Lovrien, 1962. These scales then became increasingly more widely used in the 1980's and during this period the three most cited scales were developed (Kyte and Doolittle, 1982; Hopp and Woods, 1983; Eisenberg *et al.*, 1984). While these scales can focus either on determining hydrophobicity or hydrophilicity, Kyte and Doolittle, 1982, proposed the general term hydrophathy, to reflect the hydrophobic or hydrophilic nature of the amino acids. To date almost 100 different scales have been developed. For a recent review see Simms *et al.*, 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 developed by Chou and Fasman, 1974 and subsequently modified (Chou and Fasman, 1978). The GOR algorithm based on a different approach was developed by Garnier *et al.*, 1978 and later refined (Garnier *et al.*, 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, 1988 and subsequent widely used algorithms have included JPred (Cuff *et al.*, 1998; Drozdetskiy *et al.*, 2015), PSIPRED (Jones, 1999), Porter (Pollastri and McLysaght, 2005; Mirabello and Pollastri, 2013) and SPIDER² (Heffernan *et al.*, 2015). For general reviews on the various algorithms that have been developed to predict secondary structures in proteins see Pirovano and Heringa, 2010; Pavlopoulou and Michalopoulos, 2011.

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 (Fredericks ZL, Pielak GJ., 1993; He *et al.*, 2004; Conidi *et al.*, 2013). We wanted to see what predictions regarding protein structure could be learned from a study of randomly generated missense mutants in two different well characterized aqueous proteins. The bacterial *β -galactosidase* and *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 (Kita *et al.*, 1999; Juers *et al.*, 2000) and robust simple to use colorimetric enzyme assays have also been developed (Miller, 1972; Cole *et al.*, 2017).

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 *et al.*, 2017). The *Escherichia coli* strain MC1061, $\Delta(\textit{araABOIC-leu})7679 \textit{ araD139 hsr2 hsm+ galU galK} \Delta(\textit{lac})X74 \textit{ rpsL}$ (Casadaban and Cohen, 1980) was used to clone pUC57-*xylE* and pET-11a-*xylE* and the *E. coli* strain BL21(DE3) *fhuA2 lon ompT gal dcm* $\Delta\textit{hsdS}$ □DE3 (Studier and Moffat, 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 pWVO *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 that had been restricted with the same two restriction enzymes.

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

A 30°C overnight LB ampicillin culture of BL21(DE3) pET-11a-*xyIE* 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°C water bath to approximately 0.4 OD₅₅₀ 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% (NH₄)₂SO₄ precipitations were conducted and the pellet from the 60% (NH₄)₂SO₄ 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 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. 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 hydroxyapatite column using a 0.01 – 0.8 M KPO₄ gradient created with 10 mM KPO₄; pH 7.5 and 800 mM KPO₄; pH 7.5. Catechol 2,3-dioxygenase was found to elute at 10 – 18 mM KPO₄. Again each fraction was combined and concentrated. The yield of catechol 2,3-dioxygenase enzyme from 1 L of cells was 3.03 mg.

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 *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 while the 30 *xylE* missense mutants yielded 22 distinct mutations as the same mutation occurred in multiple missense mutants. Four of the *xylE* missense mutants contained double mutants; i.e., changes in two different amino acids, and these mutants were not used. To provide the complete collection of missense mutants shown in Table 3 and Table 4, we also included the 10 *lacZ* and 10 *xylE* missense mutants that were isolated and sequenced by Cole *et al.*, 2017.

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 described by Cole *et al.* 2017.

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

Rabbit anti- β -galactosidase and catechol 2, 3-dioxygenase antibody was prepared by Cocalico Biologicals using their standard protocol. Pure β -galactosidase (G4155-5KU) 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 sodium dodecyl sulfate polyacrylamide gel electrophoresis (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 10% SDS-PAGE gel. The SDS-PAGE 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 2mM ethylenediaminetetraacetic acid (EDTA) and then the blots were washed with distilled water and dried. Images were analyzed with the ChemiDoc MP Imaging System and all intensity calculations were performed using the Image Lab Software Bio-Rad (Bio-Rad, Hercules, CA, USA).

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, and whether the mutated amino acid was located on the surface, located partially on the surface, or buried. 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

We used an averaged propensity scale that predicts the probabilities of amino acids to be found in α -helices (P_α), β -sheets (P_β) or coils (P_c) from the data presented in Costantini *et al.*, 2006, which included propensity scales based on the analysis of 2,216 proteins according to Nakashima *et al.*, 1986 and Chou, 1995 as well as their own criteria (Table 3.1).

Table 3.1. Average scale for the propensities 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

1. Bold numbers indicate the secondary structure in which each amino acid is most likely to be found.

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 and DeGrado, 1990; Pace and Scholtz, 1998) or β -sheets (Minor and Kim, 1994; Smith *et al.*, 1994) using $\Delta\Delta G$ values as well as the survey study conducted by Otaki *et al.*, 2010, which analyzed the frequency of amino acids that were found in α -helices, β -sheets or coils from 1,590 proteins. Each missense mutant was analyzed for how the amino acid change affected the P_α , P_β , 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 hydropathy or hydrophobicity scales (see Simm *et al.*, 2016 for a general review), we utilized the normalized data from 53 scales as analyzed by Cornette *et al.* 1987 and Vihinen and Torkkila, 1993, to generate a consensus hydropathy scale. The order of the amino acids ranging from the most hydrophilic to the most hydrophobic was determined for each of the 53 scales and then averaged to give a consensus ranking. The resulting averaged order from the most hydrophilic to the most hydrophobic amino acids was lys, asp, glu, gln, arg, asn, ser, pro, thr, gly, his, ala, tyr, cys, trp, met, val, leu, ile, phe, with glycine and histidine being the middle amino acids, and was in very good agreement with the order of amino acids predicted by most of the more widely utilized scales (Kyte and Doolittle, 1982; Hopp and Woods, 1983; Eisenberg *et al.*, 1984).

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

The potential effects of each mutant on the secondary structure of β -galactosidase or catechol 2, 3-dioxygenase was analyzed using seven of the most widely utilized algorithms for predicting secondary structure. Analysis using the Chou Fasman, Garnier Osguthorpe Robson (GOR) and the Qian and Sejnowski algorithms were performed using the following website:

<http://cib.cf.ocha.ac.jp/bitool/MIX/>. Analysis using the JPred4, PSIPRED, Porter 4.0 and

SPIDER² 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>. For each mutant to be analyzed by the seven predictive algorithms an amino acid sequence that 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 secondary structure had to contain a complete secondary structure or random coil.

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 3.2 and 3.3 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 hydrophathy caused by the mutation, the enzymatic activity of the missense mutant and the stability of the missense mutant as determined by western analysis.

Table 3.2. *lacZ* missense mutants

<u><i>lacZ</i> mutant</u> ¹	<u>Base pair change</u>	<u>Amino acid change</u> ²	<u>Structural perturbation</u> ³	<u>Mutation favorability</u> ⁴	<u>Protein location</u>	<u>Change in hydrophathy</u> ⁵	<u>Mutant activity</u> ⁶	<u>Protein stability</u> ⁷
39*	3053 T to A	1018 Leu to Gln	Beta Sheet	-	Surface (Partial)	+	9.64	-97.9%
337	2713 G to A	905 Glu to Lys	Coil	+	Surface (Partial)	+	2.55	-8.5%
361*	2698 G to A	900 Gly to Asp	Beta Sheet	-	Buried	+	1.34	+23.0%
364*	624 G to A	208 Gly to Asp	Beta Sheet	-	Buried	+	4.19	+42.0%
2234*	1694 G to A	565 Gly to Asp	Beta Sheet	-	Buried	+	3.74	+21.1%
2343*	2828 G to A	943 Arg to His	Beta Sheet	+	Surface	-	9.78	-88.6%
2381	2291 G to A	764 Gly to Asp	Coil	-	Buried	+	6.87	-65.4%
2382*	1697 G to A	566 Gly to Asp	Beta Sheet	-	Buried	+	3.84	-20.1%
2396	1643 G to A	584 Gly to Asp	Coil	-	Surface	+	4.31	-28.8%
2449*	604 G to A	202 Asp to Asn	Coil	+	Surface (Partial)	-	12.57	-99.9%
2454*	1061 G to A	354 Gly to Asp	Beta Sheet	-	Buried	+	3.86	+2.1%
2456	14 C to T	5 Thr to Met	Alpha Helix	+	Surface	-	15.05	-99.1%
2530	1255 C to T	419 His to Tyr	Coil	-	Surface	-	82.46	-11.2%
2540	11 T to A	4 Ile to Asn	Alpha Helix	-	Surface	+	10.29	-97.5%
2608	1378 G to A	460 Gly to Arg	Coil	-	Buried	+	0.97	-1.0%

1. Only unique mutations are listed in the table. The *lacZ*39 mutation was observed in 2 of the missense mutants, the *lacZ*361 mutation was observed in 3 of the missense mutants, the *lacZ*364 mutation was observed in 2 of the missense mutants, the *lacZ*2234 mutation was observed in 2 of the missense mutants, the *lacZ*2343 mutation was observed in 2 of the missense mutants, the *lacZ*2382 mutation was observed in 6 of the missense mutants, the *lacZ*2449 mutation was observed in 4 of the missense mutants, and the *lacZ*2454 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. 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 *lacZ*2456 and *lacZ*2540, since these mutations were at the extreme amino terminus.

Table 3.2. *lacZ* missense mutants (continued)

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 change in hydrophathy was determined using the averaged hydrophobicity scale. (+) indicates a more hydrophilic change, (–) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydrophathy of the original amino acid significantly and results in a shift of at least 5 amino acids.
6. β -galactosidase activity is given in Miller units. The assays were repeated in triplicate and the standard deviation was less than 10%. For comparison purposes, wild type *lacZ* gave an activity of 3,167.43 Miller units.
7. Protein stability was determined by western blot analysis. The percentage change in 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. *lacZ*2381 produced a proteolytic fragment as seen in Figure 3.1.

Table 3.3. *xyIE* missense mutants

<u><i>xyIE</i> mutant¹</u>	<u>Base pair change</u>	<u>Amino acid change²</u>	<u>Structural perturbation³</u>	<u>Mutation favorability</u>	<u>Protein location</u>	<u>Change in hydrophathy</u>	<u>Mutant activity</u>	<u>Protein stability</u>
1568	316 C to T	106 Arg to Cys	Beta Sheet	+*	Surface (Partial)	-*	8.61	-77.4%
1569	809 G to A	270 Gly to Gln	Coil	-*	Surface (Partial)	+*	19.18	-31.7%
1571*	380 G to A	127 Gly to Gln	Coil	-*	Surface (Partial)	+*	24.78	-49.1%
1574*	698 C to T	233 Ser to Phe	Alpha Helix	+	Surface	-*	10.67	-53.6%
1577	62 C to T	21 Ala to Val	Alpha Helix	-*	Buried	-*	7.74	-54.3%
1581*	88 G to A	30 Gly to Ser	Coil	-	Surface (Partial)	+	10.67	-50.4%
1582	191 G to A	64 Gly to Asp	Beta Sheet	-	Surface	+*	24.06	-36.4%
1583*	473 G to A	158 Gly to Asp	Beta Sheet	-	Buried	+*	24.12	-53.8%
1584	727 C to T	243 Pro to Ser	Coil	-	Surface (Partial)	+	18.05	-68.2%
1585*	25 G to A	9 Gly to Ser	Coil	-	Buried	+	22.01	-91.7%
1587	532 G to A	178 Gln to Lys	Beta Sheet	+	Surface	+	18.71	-66.4%
1588	740 G to A	247 Gly to Asp	Coil	-	Surface (Partial)	+*	22.14	-52.3%
1589*	845 G to C	282 Trp to Ser	Beta Sheet	-*	Surface (Partial)	+*	23.5	-76.4%
1591	22 C to T	8 Pro to Ser	Beta Sheet	+	Buried	+	20.12	-79.7%
1592	578 G to A	193 Ser to Asn	Beta Sheet	-	Buried	+	12.39	-53.8%
1593	733 C to T	245 Arg to Cys	Coil	+	Surface (Partial)	-*	12.83	-6.3%
1665	725 G to A	242 Gly to Asp	Coil	-	Surface (Partial)	+*	5.69	-13.4%
1666	752 G to A	251 Gly to Asp	Coil	-	Surface (Partial)	+*	20.83	-6.1%
1670*	604 G to A	202 Ala to Thr	Beta Sheet	+*	Buried	+	11.56	-64.5%
1677	637 C to T	213 His to Tyr	Beta Sheet	+*	Buried	-	12.85	-79.9%
1680	647 C to T	216 Ser to Phe	Beta Sheet	+*	Buried	-*	13.80	-57.8%
1681	598 G to A	200 Asp to Asn	Coil	+	Buried	-	12.91	-79.5%
1683	343 C to T	115 His to Tyr	Coil	-*	Surface (Partial)	-	13.34	-78.0%
1684	409 G to A	137 Glu to Lys	Coil	+	Surface	+	13.13	-26.2%
1698	454 G to A	152 Asp to Asn	Coil	+	Buried	-	13.33	-34.3%

Table 3.3. *xyIE* missense mutants (continued)

1702	70 C to T	24 His to Tyr	Alpha Helix	+	Surface (Partial)	-	11.42	-91.6%
1710	868 G to A	290 Ala to Thr	Alpha Helix	-*	Surface (Partial)	+	10.13	-51.2%

1. Only unique mutations are listed in this table. The *xyIE1571* mutation was observed in 2 of the missense mutants, the *xyIE1574* mutation was observed in 2 of the missense mutants, the *xyIE1581* mutation was observed in 3 of the missense mutants, the *xyIE1583* mutation was observed in 2 of the missense mutants, the *xyIE1585* mutation was observed in 3 of the missense mutants, the *xyIE1589* mutation was observed in 2 of the missense mutants and the *xyIE1670* mutation was observed in 2 of the missense mutants.
2. The *xyIE* 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 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 change in hydrophathy was determined using the averaged hydrophobicity scale. (+) indicates a more hydrophilic change, (-) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydrophathy of the original amino acid significantly and results in a shift of at least 5 amino acids.
6. Catechol 2,3-dioxygenase activity is given in *XylE* units. The assays were repeated in triplicate and the standard deviation was less than 10%. For comparison purposes, wild type *xyIE* gave an activity of 422.92 *XylE* units.
7. Protein stability was determined by western blot analysis. The percentage change in 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 *xyIE+* parent strain.

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. 35 out of the 42, or 83.33% of the missense mutants produced a stable protein (12 out of 17, or 70.59% for *lacZ* and 25 out of 27, or 92.59% for *xylE*). 7 out of the 42, or 14.29% 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*.

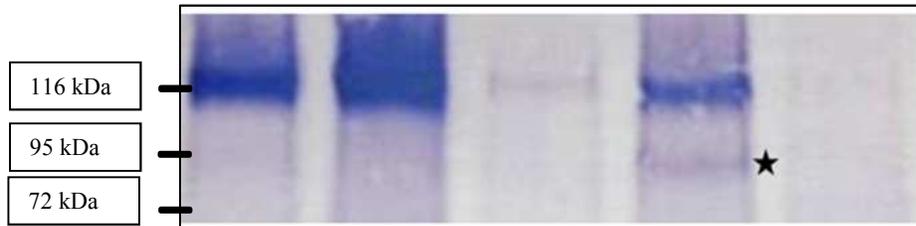


Figure 3.1. Western blot of wild-type *lacZ* and mutants of varying amounts of β -galactosidase.

1. β -galactosidase presence was using Western blot of wild-type *lacZ* (*lacZ232*), mutant *lacZ2234*, mutant *lacZ2343*, mutant *lacZ2381*, and mutant *lacZ2449*. The star indicates a proteolytic fragment present in mutant sample *lacZ2381*.

*lacZ*2381 is the only missense mutant that produces a proteolytic fragment. None of the missense mutants affected amino acids that are known to constitute the active site of either the β -galactosidase or catechol 2,3-dioxygenase proteins and 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 normalized hydropathy scale, 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 the 25, or 36.00% (4 out of 8, or 50.00% for *lacZ* and 5 out of 17, or 29.41% for *xylE*) of the missense mutants located on the surface caused a hydrophobic shift, while the missense mutants that were buried did follow this expectation as 12 out of the 17, or 70.59% (7 out of 7, or 100.00% for *lacZ* and 5 out of 10, or 50.00% for *xylE*) of the missense mutants that were buried caused a hydrophilic shift. In total, 20 out of 42, or 47.62% of the missense mutants caused expected hydrophobic or hydrophilic shifts based on their location (11 out of 15, or 73.33%, for *LacZ* and 9 out of 27, or 33.33% 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 45.25% 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 occur 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 scales (Table 1) 19 out of 42, or 45.24% of the missense mutants affected α -helical, β -sheet or coil secondary structures in which the original amino acid was a preferred amino acid (5 out of 15, or 33.33% for *lacZ* and 14 out of 27, or 51.85% for *xylE*). 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*). 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 4.0 and SPIDER² algorithms. Based on the changes of the propensity scores (Chou-Fasman, GOR and Qian-Sejnowski) or the consensus scores (JPred, PSIPRED, Porter 4.0 and SPIDER²) between the secondary structure in the parent versus the missense mutant, the percentage negative or positive impact was determined. Tables 3.4 and 3.5 lists the results of this analysis for the *lacZ* missense mutants and Tables 3.6 and 3.7 lists the results of this analysis for the *xylE* missense mutants.

Table 3.4. Chou-Fasman, GOR, and Qian-Sejowski propensity scores for *lacZ* mutant secondary structures

<i>lacZ</i> mutant	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR			Qian-Sejowski		
			Parent	Mutant	%Change	Parent	Mutant	%Change	Parent	Mutant	%Change
39	RYHYQLVWC	RYHYQQVWC	103.67	101.67	-1.93%	81.44	66.67	-18.14%	0.24	0.21	-14.43%
337	PQENYPDRLT AA	PQKNYPDRLT AA	103.67	105.25	+1.53%	40.08	50.25	+25.36%	0.72	0.72	+0.06%
361	RVNWLGLG	RVNWLGLR	99.25	100.00	+0.76%	34.00	4.38	-87.13%	0.11	0.09	-14.88%
364	GI	DI	112.50	110.00	-2.22%	-109.00	-161.50	-48.17%	0.19	0.14	-25.00%
2234	LQGGFV	LQDGFV	112.17	109.67	-2.23%	110.17	67.33	-38.88%	0.20	0.15	-23.21%
2343	RCGTRELNY	RCGTHELNY	94.22	93.00	-1.30%	35.44	50.22	+41.69%	0.20	0.18	-10.99%
2381	RQSGFL	RQSGFL	98.83	97.17	-1.69%	-8.33	-10.42	-25.00%	0.50	0.50	-0.66%
2382	LQGGFV	LQGDFV	112.17	108.83	-2.97%	93.67	39.83	-57.47%	0.20	0.15	-26.58%
2396	DENG \underline{N}	DEND \underline{N}	139.80	137.60	-1.57%	125.80	120.30	-4.37%	0.88	0.92	+4.00%
2449	EDQ \underline{D}	EDQ \underline{N}	106.75	109.00	+2.11%	-55.50	-0.25	+99.55%	0.53	0.44	-16.78%
2454	IRGVN	IRDVN	93.00	92.12	-0.95%	0.76	-21.06	-2853.85%	0.14	0.13	-11.95%
2456	MIT	MIM	95.67	109.67	+14.63%	-11.33	79.67	+802.94%	0.23	0.45	+95.68%
2530	ANIETHGMVP MNRLTD	ANIETYGMVP MNRLTD	101.69	102.88	+1.17%	-17.19	-9.38	+45.45%	0.64	0.63	-0.63%
2540	MIT	MNT	95.67	92.33	-3.48%	-11.33	-53.67	-373.53%	0.23	0.21	-10.79%
2608	LG \underline{N} ESGH	LR \underline{N} ESGH	123.57	119.29	-3.47%	21.00	-2.21	-110.54%	0.79	0.70	-11.59%

Table 3.5. JPred4, PSIPRED, Porter 4.0, and Spider² consensus scores for *lacZ* mutant secondary structures

<u><i>lacZ</i></u> <u>mutant</u>	<u>JPred4</u>			<u>PSIPRED</u>			<u>Porter 4.0</u>			<u>SPIDER²</u>		
	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>
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%

Table 3.6. Chou-Fasman, GOR, and Qian-Sejowski propensity scores for *xyIE* mutant secondary structures

<i>xyIE</i> mutant	Parent Secondary Structure	Mutant Secondary Structure	Chou-Fasman			GOR			Qian-Sejowski		
			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 GDYNYPDHKP	G QDYNYPDHKP	123.82	121.09	-2.20%	120.82	110.364	-8.65%	0.91	0.91	0.00%
1571	ADKEYT G KWGL NDVNPEAWPRD LKGMAAVRFDH	ADKEYT Q KWGL NDVNPEAWPRD LKGMAAVRFDH	94.50	89.85	-4.92%	-6.35	-12.39	-95.31%	1.47	0.61	-58.54%
1574	WEDLLRAADL I SMT	WEDLLRAADL I FMT	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	L LGL	L LSL	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 TRHGL THG	DTSIDIG S TRHGL THG	107.38	106.88	-0.47%	6.13	-12.22	-299.49%	0.73	0.60	-18.19%
1585	G	D	99.00	113.00	+14.14%	66.00	41.00	-37.88%	0.75	0.67	-10.07%
1587	YLA E QVL	YLA L VL	95.86	98.43	+2.68%	-23.14	2.57	+111.11%	0.15	0.18	+21.67%
1588	DTSIDIG P TRHGL THG	DTSIDIG P TRH 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	VM R P	VM S P	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	DTSIDIG P TRHGL THG	DTSIDIG P T C HGL THG	107.38	108.88	+1.40%	6.13	25.19	+311.22%	0.73	0.75	+3.34%
1665	DTSIDIG P TRHGL TH	DTSIDI D PTRHGL TH	104.07	103.40	-0.64%	-0.03	1.13	+3500.00%	0.71	0.72	+1.73%
1666	DTSIDIG P TRHGL TH G	DTSIDIG P TRHGL TH D	107.38	106.69	-0.64%	6.13	6.44	+5.10%	0.73	0.73	-0.56%

Table 3.6. Chou-Fasman, GOR, and Qian-Sejowski propensity scores for *xyIE* mutant secondary structures (continued)

1670	V <u>A</u> FIH	V <u>T</u> FIH	103.20	108.80	+5.43%	41.80	92.60	+121.53%	0.24	0.35	+44.44%
1677	RL <u>H</u> HVSFH	RL <u>Y</u> HVSFH	104.67	109.00	+4.14%	35.25	56.63	+60.64%	0.26	0.38	+46.60%
1680	RLHHV <u>S</u> FH	RLHHV <u>G</u> FH	104.67	103.50	-1.11%	35.25	38.63	+9.57%	0.26	0.25	-3.64%
1681	LSTKA <u>H</u> D	LSTKA <u>N</u>	95.56	96.56	+1.05%	-46.67	-37.79	+19.03%	0.46	0.47	+0.97%
1683	APSG <u>H</u>	APSG <u>Y</u>	113.80	117.60	+3.34%	70.90	85.10	+20.03%	0.83	0.82	-0.84%
1684	ADKEYTGKWGL NDVN <u>P</u> EAWPRD LKGMAAVRFDH	ADKEYTGKWGL NDVN <u>K</u> AWPRD LKGMAAVRFDH	101.33	102.06	+0.72%	-12.09	-7.67	+36.57%	0.59	0.58	-1.18%
1698	ADKEYTGKWGL NDVN <u>P</u> EAWPRD LKGMAAVR <u>F</u> DH	ADKEYTGKWGL NDVN <u>P</u> EAWPRD LKGMAAVR <u>F</u> NH	101.33	101.67	+0.33%	-10.89	-11.20	-2.84%	0.60	0.59	-1.39%
1702	MSKALE <u>H</u> YVE	MSKALE <u>Y</u> YVE	117.40	112.50	-4.17%	141.10	99.00	-29.84%	0.77	0.72	-6.94%
1710	TDQLGK <u>A</u> I	TDQLGK <u>T</u> I	106.63	99.13	-7.03%	43.88	-24.88	-156.70%	0.31	0.19	-38.29%

Table 3.7. JPred4, PSIPRED, Porter 4.0, and Spider² consensus scores for *xyIE* mutant secondary structures

<u><i>xyIE</i></u> <u>mutant</u>	<u>JPred4</u>			<u>PSIPRED</u>			<u>Porter 4.0</u>			<u>SPIDER²</u>		
	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%
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%

Table 3.7. JPred4, PSIPRED, Porter 4.0, and Spider² consensus scores for *xylE* mutant secondary structures (continued)

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%
------	------	------	--------	------	------	---------	------	------	---------	-------	-------	----------

Both Chou-Fasman and GOR 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 JPred 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.

Interestingly, when an algorithm predicted a change in the secondary structure between the parent and the mutant, the impact was just 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, GOR analysis predicted that 24 out of the 42, or 57.14% of the missense mutants that affected the secondary structure had a negative impact, 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, JPred analysis predicted that 21 out of the 34, or 61.76% of the missense mutants that affected the secondary structure had a negative impact, 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, Porter 4.0 analysis predicted that 12 out of the 27, or 44.44% of the missense mutants that affected the secondary structure had a negative impact 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.

Given the fact that inactive missense mutants would be predicted to have a significant impact on protein structure and the older algorithms made this prediction more often than the newer algorithms, we also 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 calculated using 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. Tables 3.8 and 3.9 show the results of this analysis for the *lacZ* missense mutants and Tables 3.10 and 3.11 show the results of this analysis for the *xyIE* missense mutants.

Table 3.8. Chou-Fasman, GOR, and Qian-Sejowski accuracy for *lacZ* secondary structures

<u><i>lacZ</i></u> <u>mutant</u>	<u>Chou-Fasman</u>			<u>GOR</u>			<u>Qian-Sejowski</u>		
	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>
39	77.8%	77.8%	0%	66.7%	66.7%	0%	16.7%	0.06%	-99.6%
337	50%	50%	0%	66.7%	66.7%	0%	87.5%	87.5%	0%
361	50%	25.0%	-50%	50%	37.5%	-25.0%	0%	0%	0%
364	100%	100%	0%	0%	0%	0%	0%	0%	0%
2234	66.7%	66.7	0%	50%	33.3%	-25.0%	0%	0%	0%
2343	22.2%	22.2%	0%	22.2%	33.3%	+50%	0%	0%	0%
2381	33.3%	100%	+200.3%	66.7%	50%	-25.0%	75.0%	58.3%	-22.3
2382	66.7%	66.7%	0%	50%	50%	0%	0%	0%	0%
2396	100%	60%	-40%	80%	100%	+25%	50%	50%	0%
2449	0%	0%	0%	0%	0%	0%	25.0%	25.0%	0%
2454	35.3%	11.8%	-66.6	29.4%	11.8%	-59.9%	8.8%	0.06	-99.3%
2456	0%	0%	0%	0%	66.7%	+6670%	66.7%	66.7	0%
2530	68.8%	50%	-27.3	56.3%	62.5%	11.0%	90.6%	96.9%	7.0%
2540	0%	0%	0%	0%	0%	0%	16.7%	0%	-100%
2608	87.5%	87.5%	0%	85.7%	71.4%	-16.7	0%	81.3%	+8130%

Table 3.9. JPred4, PSIPRED, Porter 4.0, and Spider² accuracy for *lacZ* secondary structures

<u><i>lacZ</i></u> <u>mutant</u>	<u>JPred4</u>			<u>PSIPRED</u>			<u>Porter 4.0</u>			<u>SPIDER²</u>		
	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>
39	66.7%	77.8%	+16.6%	88.9%	88.9%	0%	88.9%	88.9%	0%	100%	100%	0%
337	100%	100%	0%	100%	100%	0%	83.3%	75.0%	-10.0%	66.7%	75.0%	+12.4%
361	62.5%	62.5%	0%	50.0%	50.0%	0%	50%	37.5%	-25.0%	62.5%	62.5%	0%
364	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	0%	-100%
2234	33.3%	50.0%	+50.2%	16.7%	16.7%	0%	0%	0%	0%	33.3%	16.6%	-50.2%
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%	-33.4%	83.3%	83.3%	0%	83.3%	75.0%	-10.0%	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%	-20.1	11.8%	11.8%	0%	29.4%	33.3%	+13.3%	17.6%	17.6%	0%
2456	0%	25.0%	+2500%	0%	33.3%	+3300%	0%	0%	0%	0%	33.3%	+3330%
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 3.10. Chou-Fasman, GOR, and Qian-Sejowski accuracy for *xylE* secondary structures

<u><i>xylE</i></u> <u>mutant</u>	<u>Chou-Fasman</u>			<u>GOR</u>			<u>Qian-Sejowski</u>		
	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>
1568	40.0%	80.0%	+100%	80.0%	40.0%	-50.0%	40.0%	50.0%	+25.0%
1569	81.8%	81.8%	0%	63.6%	45.5%	-28.5%	50.0%	50.0%	0%
1571	53.1%	53.1%	0%	15.6%	18.8%	+20.5%	76.6%	73.4%	-4.2%
1574	57.1%	57.1%	0%	78.6%	78.6%	0%	78.6%	85.7%	+9.0%
1577	50.0%	0%	-100%	100%	90%	-10.0%	100%	100%	0%
1581	50.0%	50.0%	0%	0%	0%	0%	0%	0%	0%
1582	33.3%	33.3%	0%	0%	0%	0%	0%	0%	0%
1583	0%	0%	0%	20.0%	0%	-100%	0%	0%	0%
1584	50.0%	25.0%	-50%	25.0%	6.3%	-74.8%	50.0%	71.9%	+43.8%
1585	0%	100%	+1000%	0%	0%	0%	100%	100%	0%
1587	0%	0%	0%	0%	0%	0%	0%	0%	0%
1588	50.0%	50.0%	0%	25.0%	31.3%	+25.2%	50.0%	50.0%	0%
1589	100%	100%	0%	100%	75.0%	-25%	0%	0%	0%
1591	25.0%	50.0%	+100%	50.0%	75.0%	+50.0	0%	0%	0%
1592	80.0%	20.0%	-75.0	0%	0%	0%	20.0%	10.0%	-50.0%
1593	50.0%	50.0%	0%	25.0%	31.3%	+25.2%	50.0%	50.0%	0%
1665	46.7%	46.7%	0%	26.7%	53.3%	+99.6	50.0%	50.0%	0%
1666	50.0%	56.3%	+12.6%	25.0%	43.8%	+75.2%	50.0%	50.0%	0%
1670	0%	60.0%	+6000	0%	80.0%	+8000%	18.2%	60.0%	+229.7
1677	37.5%	62.5%	+66.7	50.0%	62.5%	+50.0%	12.5%	37.5%	+200.0
1680	37.5%	25.0%	-33.3%	50.0%	62.5%	+50.0%	12.5%	12.5%	0%
1681	0%	0%	0%	0.0%	0.0%	0%	27.8%	44.4%	+59.71
1683	80.0%	80.0%	0%	40.0%	40.0%	0%	50.0%	50.0%	0%
1684	42.2%	42.2%	0%	15.2%	21.2%	+38.8%	69.7%	68.2%	-2.2%
1698	51.5%	51.5%	0%	15.2%	15.2%	0%	74.2%	69.7%	-6.1%
1702	50.0%	40.0%	-20.0%	100%	90.0%	-10.0%	100%	90.0%	-10.0%
1710	75.0%	0%	-100%	62.5%	12.5%	-80.0	18.8%	0%	-100%

Table 3.11. JPred4, PSIPRED, Porter 4.0, and Spider² accuracy for *xyIE* secondary structures

<u><i>xyIE</i></u> <u>mutant</u>	<u>JPred4</u>			<u>PSIPRED</u>			<u>Porter 4.0</u>			<u>SPIDER²</u>		
	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>	<u>Parent</u>	<u>Mutant</u>	<u>%Change</u>
1568	100%	100%	0%	100%	100%	0%	100%	100%	0%	100%	100%	0%
1569	84.6%	90.9%	+7.5	81.8%	81.8%	0%	81.8%	100%	+22.2	90.9%	90.9%	0%
1571	90.6%	87.5%	-3.4%	81.8%	84.4%	+3.2%	81.8%	81.8%	0%	78.1%	84.4%	+8.1%
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%	+25.0%	75.0%	75.0%	0%	100%	100%	0%
1582	66.7%	77.8%	+16.6%	77.8%	77.8%	0%	66.7%	66.7%	0%	88.9%	77.8%	-12.5%
1583	80.0%	60.0%	-25.0%	60.0%	60.0%	0%	60.0%	60.0%	0%	80.0%	60%	-25.0%
1584	81.3%	70.6%	-9.3	80.0%	86.7%	+8.4%	87.5%	81.3%	0%	70.6%	62.5%	-11.5%
1585	100.0%	100.0%	0%	50.0%	100.0%	100%	100%	100%	0%	0.0%	100%	+1000%
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%	+1.6%	87.5%	87.5%	0%	70.6%	68.8%	-2.5%
1589	75.0%	75.0%	0%	50.0%	50.0%	0%	75.0%	75.0%	0%	75.0%	25.0%	-66.7%
1591	75.0%	50.0%	-33.3%	0%	0%	0%	50.0%	50.0%	0%	25.0%	50%	100%
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%	+1.6	87.5%	81.3%	-7.1%	70.6%	75.0%	+6.2%
1665	68.8%	73.3%	+6.5%	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%	+16.1%	87.5%	81.3%	-7.1%	75.0%	81.3%	-8.4%
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%	-20.0%	71.4%	71.4%	0%
1683	80.0%	80.0%	0%	100%	100%	0%	100%	100%	0%	80.0%	80.0%	0%
1684	81.8%	81.8%	0%	72.7%	97.0%	+33.4%	72.7%	72.7%	0%	66.7%	69.7%	+4.5%
1698	85.3%	85.3%	0%	97.0%	97.0%	0%	90.9%	87.9%	-3.3%	84.8%	84.8%	0%
1702	100%	100%	0%	100%	100%	0%	100%	100%	0%	100%	100%	0%

Table 3.11. JPred4, PSIPRED, Porter 4.0, and Spider² accuracy for *xyIE* secondary structures (continued)

1710	0.0%	0%	0%	0.0%	0%	0%	37.5%	37.5%	0%	50.0%	37.5%	0%
------	------	----	----	------	----	----	-------	-------	----	-------	-------	----

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. 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 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*), JPred 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*) 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 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*), JPred 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 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*).

3.5 Discussion

We have generated a collection of 42 unique inactive missense mutants in the aqueous cytosolic *β-galactosidase* and *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 (Whitfield *et al.*, 1966; Greb *et al.*, 1971; Truman and Bergquist, 1976). 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.

While the study of aqueous proteins whose three dimensional structures have been determined has reaffirmed the importance of the organized α -helical and β -sheet secondary structures, another trend that was discovered is 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 20 out of the 42, or 47.62% of the missense mutants caused the expected hydrophobic to hydrophilic shift based on their location in the protein.

A total of 19 out of the 42, or 45.45% of the missense mutants occurred in unstructured “coil” or “turn” 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 that was not the case.

The impact of the missense mutants on the secondary structures was just as likely to be positive and result in an improved secondary structure as it was 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 an overwhelming consensus that most of the missense mutants would be predicted to affect the secondary structures. Both Chou-Fasman and GOR 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 JPred 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.

Since the newer JPred, PSIPRED, Porter 4.0 and SPIDER² models are thought to be more accurate than the older Chou-Fasman, GOR and Qian-Sejnowski models (for general reviews see Pirovano and Heringa, 2010; Pavlopoulou and Michalopoulos, 2011) 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 correctly identified 77.67% of all of the amino acids in the secondary structures affected by the missense mutants correctly, followed by JPred at 76.72%, SPIDER² at 75.30%, PSIPRED at 75.06%, Chou-Fasman at 48.69%, Qian-Sejnowski at 47.51% and GOR at 36.34%. Thus the newer models were significantly more accurate at identifying the secondary structures than the older models. Interestingly, the creators of the various algorithms have all calculated the accuracies of their models; Chou-Fasman, 71.0% (Prevelige and Fasman, 1989), GOR, 64.7% (Garnier *et al.*, 1996), Qian and Sejnowski, 64.3% (Qian and Sejnowski, 1988), JPred4, 82.0% (Drozdetskiy *et al.*, 2015), PSIPRED, 76.5% (Jones,

1999), SPIDER2, 82.0% (Heffernan *et al.*, 2015) and Porter 4.0, 82.2% (Mirabello and Pollastri, 2013). Our results suggest that the newer models are indeed more accurate at predicting the presence of secondary structures in proteins than the older models.

REFERENCES

- Anfinsen, CB. 1973. Principles that govern the folding of protein chains. *Science*. 181: 223–230.
- Astbury WT, Sisson WA. 1935. X-ray studies of the structures of hair, wool and related fibres. III. The configuration of the keratin molecule and its orientation in the biological cell. *Proc. R. Soc. Lond.* A150:533-551.
- Astbury WT, Street A. 1931. X-ray studies of the structures of hair, wool and related fibres. I. General, *Trans. R. Soc. Lond.* A230:75-101.
- Astbury WT, Woods HJ. 1934. X-ray studies of the structures of hair, wool and related fibres. II. The molecular structure and elastic properties of hair keratin. *Trans. R. Soc. Lond.* A232:333-394.
- Bertani, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol.* 62:293-300.
- Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Ångstrom resolution. *Nature*. 206(986):757-61.
- Casadaban M, Cohen SN. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* 138:179-207
- Chou KC, Zhang CT. 1995. Prediction of protein structural classes. *Crit Rev Biochem Mol Biol.* 30(4):275-349.
- Chou PY, Fasman GD. 1974. Prediction of protein conformation. *Biochem.* 13(2):222–245.

- Conidi A, van den Berghe V, Leslie K, Stryjewska A, Xue H, Chen YG, Seuntjens E, 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:e76733.
- Constatini S, Colonna G, Focchiano AM. 2006. Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochem Biophys Res Comm*. 342:441-451.
- Cornette JL, Cease KB, Margalit H, Spouge JL, Berzofsky JA, DeLisi C. 1987. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J Mol Biol*. 195(3):659-85.
- Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. 1998. JPred: a consensus secondary structure prediction server. *Bioinformatics*. 14:892-893.
- Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucl. Acids Res*. 43(W1):W389-W394.
- Eisenberg D, Schwarz E, Komaromy M, Wall R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol*. 179(1):125-142.
- Fredericks ZL, Pielak GJ. 1993. Exploring the interface between the N- and C-terminal helices of cytochrome c by random mutagenesis within the C-terminal helix. *Biochemistry*. 32:929-936.
- Garnier J, Gibrat JF, Robson B. 1996. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol*. 266:540-53.

- Garnier J, Osguthorpe D, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol.* 120:97–120.
- Greeb J, Atkins F, Loper C. 1971. Histidinol dehydrogenase (*hisD*) mutants of *Salmonella typhimurium*. *J Bacteriol.* 106:421-431.
- Harayama S, Rekik M, Bairoch A, Neidle EL, Ornston LN. 1991. Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal *benABC* and *Pseudomonas putida* TOL pWWO plasmid *xylXYZ*, genes encoding benzoate dioxygenases. *J Bacteriol.* 173:7540-7548.
- He MM, Wood ZA, Baase WA, Xiao H, Matthews BW. 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.* 13:2716–2724.
- Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, Sattar A, Yang Y, Zhou Y. 2015. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Sci. Rep.* 5:11476.
- Hopp TP, Woods KR. 1983. A computer program for predicting protein antigenic determinants. *Mol Immunol.* 20(4):483-489.
- Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292:195-202.
- Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, Matthews BW. 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.

- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*. 181(4610):662-666.
- Kyte J, Doolittle RF. 1982. A simple method for displaying the hydrophathic character of a protein. *J. Mol. Biol.* 157:105-132.
- Miller JH. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Minor DL Jr, Kim PS. 1994. Measurement of the β -sheet-forming propenisties of amino acids. *Nature* 367:660-663.
- Mirabello C, Pollastri G. 2013. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics*. 29(16):2056-2058.
- Nakashima H, Nishikawa K, Ooi T. 1986. The folding type of a protein is relevant to the amino acid composition. *J Biochem*. 99(1):153-162.
- O'Neil KT, DeGrado W. 1990. A thermodynamic scale for the helix forming tendencies of the commonly occurring amino acids. *Science*. 250(4981):646-651.
- Otaki JM, Tsutsumi M, Gotoh T, Yamamot H. 2010. Secondary structure based on amino acid composition and availability in proteins. *J Chem Inf Model*. 50:690-700.
- Pace CN, Scholtz JM. 1998. A helix propensity scale based on experimental studies of peptides and protiens. *Biophy Journal*. 75:422-427.
- Pauling L, Corey RB. 1951. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* 37(11):729–40.

- Pauling L, Corey RB, Branson HR. 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* 37(4):205-211.
- Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC. 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature.* (4711):416-22.
- Pollastri G, McLysaght A. 2005. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics.* 21:1719-1720.
- Prevelige P Jr, Fasman GD. 1989. Chapter 9: chou-fasman prediction of the secondary structure of proteins: the chou-fasman-prevelige algorithm in prediction of protein structure and the principles of protein conformation (edited by Gerald D. Fasman). Plenum, New York, 391-416.
- Protein Data Bank (PDB). 2017. Yearly Growth of Protein Structures. www.rcsb.org.
- Qian N, Sejnowski TJ. 1988. Predicting the secondary structure of globular proteins using network models. *J. Mol. Biol.* 202:865-884.
- Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for the β -sheet forming tendencies of the amino acids. *Biochem.* 33:5510-5517.
- Tanford C, Lovrien R. 1962. Dissociation of Catalase into Subunits. *J Am Chem Soc.* 84:1892–1896.
- The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
- The UniProt Consortium. 2017. The Universal Protein Resource (UniProt). Current Release Statistics. <http://www.uniprot.org/>.

Truman P, Bergquist PL. 1976. Genetic and Biochemical Characterization of Some Missense Mutations in the lacZ Gene of Escherichia coli K-12. *Journ Bacter* 126(3):1063-1074.

Vihinen M, Torkkila E. 1993. HYDRO: a program for protein hydropathy predictions. *Comp Meth Prog in Biomed.* 41:121-129.

Whitfield HJ Jr, Martin RG, Ames BN. 1966. Classification of aminotransferase (C gene) mutants in the histidine operon. *J Mol Biol.* 21:335-355

**CHAPTER 4: A STUDY OF THE ABILITY OF SALTS
TO STABILIZE PROTEINS *IN VIVO***

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

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

*- corresponding author, ealtman@mtsu.edu

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

Key words

Hofmeister ions, protein stability, salt suppression

4.1 Abstract

Numerous studies have been conducted on the ability of salts to stabilize proteins *in vitro* and 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* as well and that the ability of salts to stabilize these two proteins 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 $(\text{NH}_4)_2\text{HPO}_4$, while $\text{NH}_4\text{CH}_3\text{CO}_2$, $(\text{NH}_4)_2\text{SO}_4$ and NaCH_3CO_2 did not stabilize either of the proteins. We also investigated the ability of salt to rescue a collection of well characterized nonfunctional β -galactosidase and catechol 2,3-dioxygenase missense mutants that our laboratory has created. 92.85% 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. 73.33% of the β -galactosidase missense mutants could be rescued by salt, while only 11.11% 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 compacted as β -galactosidase and thus it is far easier for salts to penetrate and rescue inactive β -galactosidase proteins.

4.2 Introduction

From the seminal studies of Hofmeister (1888) regarding the ability of different salts to precipitate egg white protein, the lyotropic or Hofmeister series of ions was developed in which ions are ordered by their ability to precipitate proteins. For salts with the same cation, $\text{SO}_4^{2-} > \text{HPO}_4^{2-} > \text{F}^- > \text{CH}_3\text{CO}_2^- > \text{Cl}^- > \text{Br}^- > \text{NO}_3^- > \text{I}^- > \text{ClO}_4^- > \text{SCN}^-$ and for salts with the same anion, $(\text{CH}_3)_4\text{N}^+ > \text{Rb}^+ > \text{K}^+ > \text{Na}^+ > \text{Li}^+ > \text{Mg}^{2+} > \text{Ca}^{2+}$. For general reviews see Jungwirth and Cremer, 2014; Salis and 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 by Kunz *et al.*, 2004.

In a series of studies with collagen, gelatin and ribonuclease, Von Hippel and Wong, 1962, 1963, 1964, 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 $(\text{NH}_4)_2\text{SO}_4$ had the greatest stabilizing effects, followed by KCl and NaCl . Subsequent studies by Nandi and Robinson (1972a, 1972b) 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 Pelgram *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, 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 (Good and Pattee, 1970; Russell, 1972; Bilsky and Armstong, 1973; Kohno and Roth, 1979). Good and Pattee, 1970, isolated 15 temperature sensitive *Staphylococcus aureus* mutants that could be corrected by 1 M NaCl. 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, 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, 1979, showed that 56 temperature sensitive and 47 cold sensitive *Salmonella enterica* serovar Typhimurium histidine auxotrophs could all be 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 (Kita *et al.*, 1999; Juers *et al.*, 2000) and their activity can be quantified using robust easy to use colorimetric enzyme assays (Miller, 1972; Cole *et al.*, 2017). 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 *et al.*, 2017).

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 OD₅₅₀ 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. The *S. enterica* strains TT18519, *hisC10081::MudF(lac+)* and ALS1442, *proB::xylE(cat)* were used in this study (Cole *et al.*, 2017).

4.3.2 Salt suppression plate tests of the β -galactosidase and catechol 2,3-dioxygenase missense mutants

The *lacZ* missense mutants were patched onto two YG agar plates that also contained 1 mM IPTG and 40 mg/ml 5-bromo-4-chloro-3-indoyl β -D-galactopyranoside (X-gal) and incubated

overnight at 37°C. One plate contained no additional salt and one contained 0.2 M NaCl. The plates were visually inspected after 16 hours and if a blue color was observed which was more vibrant on the plate containing salt than the plate without salt, the missense mutant was deemed to be salt correctable.

The *xylE* missense mutants were patched onto two YG agar plates that also contained 1mM IPTG and incubated overnight at 37°C. One plate contained no additional salt and one contained 0.2 M NaCl. After 16 hours the plates were lightly sprayed with a 100 mM catechol solution (Cole *et al.*, 2017) and visually inspected. If a yellow color was observed which was more vibrant on the plate containing salt than the plate without salt, the missense mutant was deemed to be salt correctable.

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

β -galactosidase assays were performed as described by Miller (1972) and the catechol 2,3-dioxygenase assays were performed as described by Cole *et al.* (2017) 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.

4.3.4 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, $\text{NH}_4\text{CH}_3\text{CO}_2$ (ammonium acetate), NH_4Cl (ammonium chloride), $(\text{NH}_4)_2\text{HPO}_4$ (ammonium phosphate dibasic), $(\text{NH}_4)_2\text{SO}_4$ (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 (Von Hippel and Wong, 1962, 1963, 1964; Nandi and Robinson, 1972a, 1972b; Pelgram *et al.*, 2010). 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.

Table 4.1. Hofmeister Salt Tables for wild-type β -galactosidase and catechol 2,3-dioxygenase

Salts	0M	0.05M	0.1M	0.15M	0.2M	0.25M	0.3M	0.35M	0.4M
β-galactosidase									
NH ₄ CH ₃ CO ₂	152.11	33.53	25.08	19.59	9.39	NA	NA	NA	NA
NH ₄ Cl	152.11	179.73	174.20	150.20	140.81	141.21	140.26	139.84	136.25
(NH ₄) ₂ HPO ₄	152.11	171.10	105.94	97.84	84.16	81.42	32.19	30.82	20.84
(NH ₄) ₂ SO ₄	152.11	24.41	43.91	65.34	110.49	105.71	103.46	101.93	98.44
NaCH ₃ CO 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 ₂ HPO ₄	152.11	303.21	366.04	384.11	392.20	395.90	409.15	392.54	387.37
Na ₂ SO ₄	152.11	323.43	390.31	473.60	604.90	659.46	770.16	777.96	667.50
Catechol 2,3-dioxygenase									
NH ₄ CH ₃ CO ₂	344.79	32.21	25.39	21.63	22.61	NA	NA	NA	NA
NH ₄ Cl	344.79	368.45	413.99	464.59	464.05	442.13	399.63	428.02	483.88
(NH ₄) ₂ HPO ₄	344.79	658.53	643.23	635.43	613.40	484.03	301.61	247.72	117.15
(NH ₄) ₂ SO ₄	344.79	67.80	333.54	303.95	258.16	152.99	72.86	68.49	70.19
NaCH ₃ CO 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 ₂ HPO ₄	344.79	739.92	847.58	852.98	852.33	884.97	858.91	897.50	952.78
Na ₂ SO ₄	344.79	322.28	415.87	442.38	551.09	881.66	922.25	977.02	142.65

1. β -galactosidase and catechol 2,3 dioxygenase assays were repeated in triplicate and the standard deviation was less than 10%.

Na₂SO₄ was the best salt for stabilizing β-galactosidase, followed by Na₂HPO₄, NaCl, NH₄Cl and (NH₄)₂HPO₄. NH₄CH₃CO₂, (NH₄)₂SO₄ and NaCH₃CO₂ did not stabilize β-galactosidase. Na₂SO₄ was the best salt for stabilizing catechol 2,3-dioxygenase, followed by Na₂HPO₄, (NH₄)₂HPO₄, NH₄Cl and NaCl. NH₄CH₃CO₂, (NH₄)₂SO₄ and NaCH₃CO₂ 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. The 42 missense mutants were initially screened on plates and the results are shown in Figures 4.1 and 4.2.

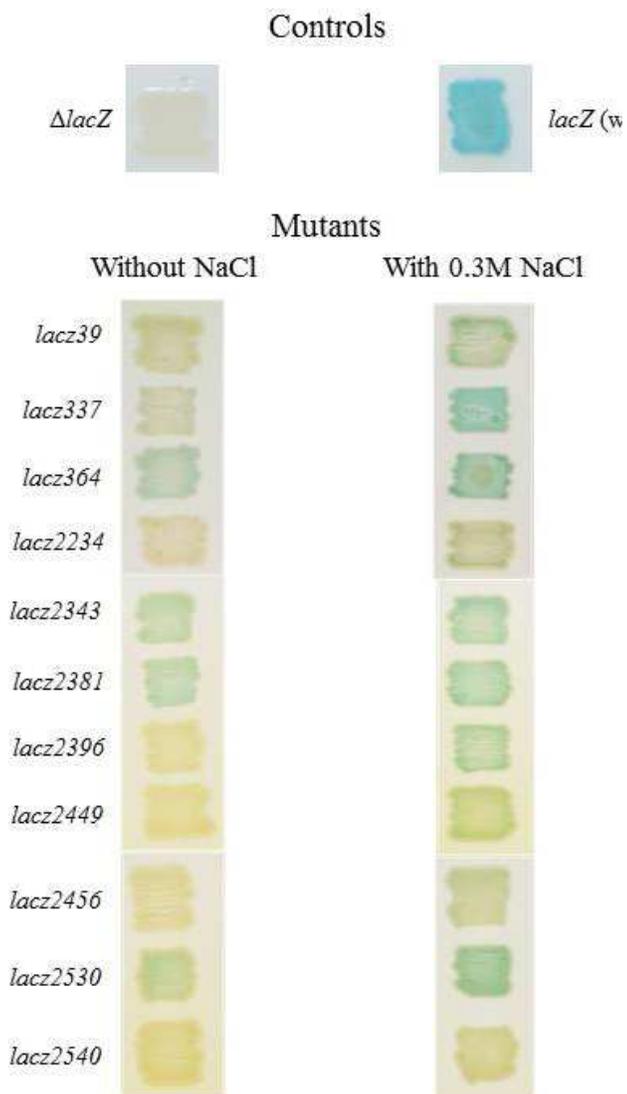


Figure 4.1. Suppression of *lacZ* missense mutants by 0.3M NaCl. *lacZ* missense mutants were patched onto LB, 1 mM IPTG, 40 ug/mL X-Gal plates without and with 0.3M NaCl and incubated at 37°C for 16 hours. Wild-type *S. enterica* (*lacZ*-) and *S. enterica* with the *hisC10081::MudF[lac+]* insertion are included as controls.

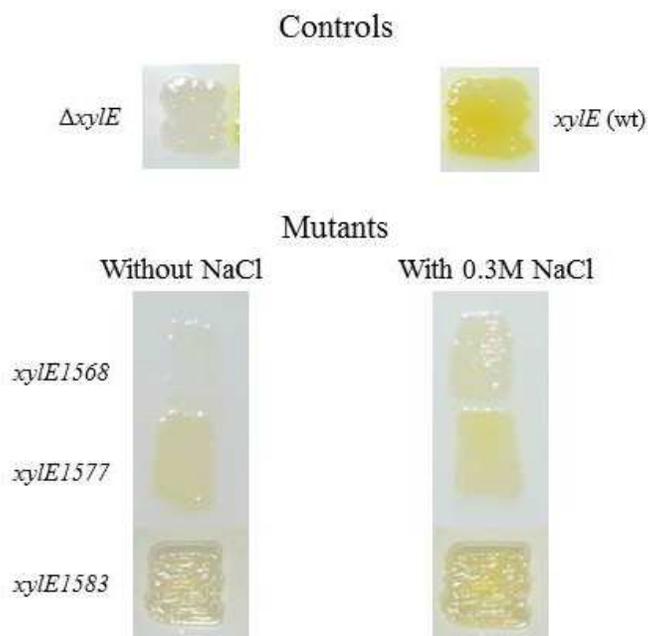


Figure 4.2. Suppression of *xyIE* missense mutants by 0.3M NaCl. *xyIE* missense mutants were patched onto LB, 1 mM IPTG plates without and with 0.3M NaCl and incubated at 37°C for 16 hours after which the plates were sprayed with a light mist of 100 mM catechol. Wild-type *S. enterica* (*xyIE*-) and *S. enterica* with the *proB::xyIE(cat)* insertion are included as controls.

A total of 11 out of the 15, or 73.3% of the nonfunctional β -galactosidase missense mutants could be rescued by the addition of NaCl, but only 3 out of the 27, or 11.11% of the nonfunctional catechol 2,3-dioxygenase missense mutants could be rescued by the addition of NaCl. Enzymatic assays were performed on all of the missense mutants that could be rescued by salt and the results are shown in Tables 4.2 and 4.3.

Table 4.2. *lacZ* mutant activity in the presence of 0.2M or 0.3M NaCl

<i>lacZ</i> mutant	<u>Without NaCl</u> (Average MU)	<u>.2M NaCl</u> (Average MU)	<u>.2M NaCl</u> Fold Increase	<u>.3M NaCl</u> (Average MU)	<u>.3M NaCl</u> Fold Increase
39	1.29	25.39	19.72	34.24	26.60
337	0.65	1.03	1.59	0.90	1.39
364	0.48	1.04	2.18	0.86	1.78
2234	0.70	0.89	1.27	0.91	1.30
2343	14.28	77.96	5.46	87.54	6.13
2381	0.78	2.37	3.06	2.67	3.45
2396	0.24	0.43	1.77	0.31	1.25
2449	1.26	14.27	11.36	12.41	9.88
2456	1.20	10.07	8.40	10.18	8.49
2530	0.79	2.42	3.07	2.61	3.33
2540	0.75	2.26	3.02	2.24	2.99
WT232	151.11	252.71	1.67	255.90	1.70

1. β -galactosidase assays were repeated in triplicate and the standard deviation was less than 10%.

Table 4.3. *xyIE* activity in the presence of 0.2M or 0.3M NaCl

<u><i>xyIE</i> mutant</u>	<u>Without NaCl</u> <u>(Average MU)</u>	<u>.2M NaCl</u> <u>(Average MU)</u>	<u>.2M NaCl</u> <u>Fold Increase</u>	<u>.3M NaCl</u> <u>(Average MU)</u>	<u>.3M NaCl</u> <u>Fold Increase</u>
1568	7.46	10.40	1.39	11.52	1.54
1577	9.24	10.64	1.15	9.83	1.06
1583	8.34	11.24	1.35	11.82	1.42
WT KD1032	344.79	397.93	1.15	389.36	1.13

1. Catechol 2,3 dioxygenase assays were repeated in triplicate and the standard deviation was less than 10%.

For the β -galactosidase missense mutants, the β -galactosidase activity was increased by an average of 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 1.34 fold with the addition of 0.3 M NaCl. *lacZ39* was by far the most suppressible missense mutant as its β -galactosidase activity was increased 26.60 fold with addition of 0.30 M NaCl.

Using data from Chapter 2, Tables 4.4 and 4.5 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 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, and whether the missense mutant can be rescued by salt.

Table 4.4. β -galactosidase missense mutants

<u><i>lacZ</i></u> <u>mutant</u>	<u>Amino acid</u> <u>change</u> ¹	<u>Structural</u> <u>perturbation</u> ²	<u>Mutation</u> <u>favorability</u> ³	<u>Protein</u> <u>Location</u> ⁴	<u>Change in</u> <u>hydropathy</u> ⁵	<u>Salt</u> <u>Suppressible</u>
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.
2. 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.
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 Chapter 2. (+) 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.

Table 4.4. β -galactosidase missense mutants (continued)

5. The change in hydropathy was determined using the averaged hydrophobicity scale from Chapter 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.

Table 4.5. Catechol 2,3-dioxygenase missense mutants

<u><i>xyIE</i></u> <u>mutant</u>	<u>Amino acid</u> <u>change</u> ¹	<u>Structural</u> <u>perturbation</u> ²	<u>Mutation</u> <u>favorability</u> ³	<u>Protein</u> <u>location</u> ⁴	<u>Change in</u> <u>hydropathy</u> ⁵	<u>Salt</u> <u>Suppressible</u>
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	-	-
1683	115 His to Tyr	Coil	-*	Surface (Partial)	-	-
1684	137 Glu to Lys	Coil	+	Surface	+	-

Table 4.5. Catechol 2,3-dioxygenase missense mutants (continued)

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 *xyIE* 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 Chapter 2. (+) 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 hydrophathy was determined using the averaged hydrophobicity scale in Chapter 2. (+) indicates a more hydrophilic change, (-) indicates a more hydrophobic change. An asterisk indicates a mutation that changes the hydrophathy of the original amino acid significantly and results in a shift of at least 5 amino acids.

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 14, or 92.85% 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 4 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 *lacZ39* missense mutant

Because the *lacZ39* missense mutant was easily 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.

Table 4.5. Hofmeister Salt Tables for mutant β -galactosidase (*lacZ39*)

Salts	0M	0.05M	0.1M	0.15M	0.2M	0.25M	0.3M	0.35M	0.4M
NH ₄ CH ₃ CO ₂	1.29	.096	1.05	0.84	0.80	NA	NA	NA	NA
NH ₄ Cl	1.29	3.11	8.97	10.08	13.65	16.88	19.94	21.44	18.47
(NH ₄) ₂ HPO ₄	1.29	2.28	1.91	1.38	1.14	1.08	0.60	0.52	0.43
(NH ₄) ₂ SO ₄	1.29	3.74	12.07	14.42	17.02	2.26	1.91	1.74	1.01
NaCH ₃ CO 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 ₂ HPO ₄	1.29	3.83	4.49	4.55	4.17	3.39	2.90	2.58	2.52
Na ₂ SO ₄	1.29	11.17	22.11	36.39	57.26	48.67	45.18	39.15	34.64

Na₂SO₄ was the best salt for stabilizing β-galactosidase from *lacZ39*, followed by NaCl, NH₄Cl, (NH₄)₂SO₄, NaCH₃CO₂, Na₂HPO₄ and (NH₄)₂HPO₄. Only NH₄CH₃CO₂ 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 Na₂SO₄ 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 NH₄CH₃CO₂, (NH₄)₂SO₄ or NaCH₃CO₂ could stabilize wild-type β-galactosidase, but only NH₄CH₃CO₂ 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 (Von Hippel and Wong, 1962, 1963, 1964; Nandi and Robinson, 1972a, 1972b; Pelgram *et al.*, 2010) could also perform *in vivo* 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, NH₄CH₃CO₂ (ammonium acetate), NH₄Cl (ammonium chloride), (NH₄)₂HPO₄ (ammonium phosphate dibasic), (NH₄)₂SO₄ (ammonium sulfate), NaCH₃CO₂ (sodium acetate), NaCl (sodium chloride), Na₂HPO₄ (sodium phosphate dibasic) and Na₂SO₄ (sodium sulfate). Generally, Na₂SO₄ and Na₂HPO₄ were very effective at stabilizing both proteins, followed by NaCl, NH₄Cl and (NH₄)₂HPO₄, while NH₄CH₃CO₂, (NH₄)₂SO₄ and NaCH₃CO₂ did not stabilize either of the proteins. However, Table 19 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 $(\text{NH}_4)_2\text{HPO}_4$, while $(\text{NH}_4)_2\text{HPO}_4$ 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 (Hawthorne and Friis, 1964; Good and Pattee, 1970; Russell, 1972; Bilsky and Armstrong, 1973; Kohno and Roth, 1979). 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. 14 out of the 42, or 33.33% of the missense mutants were rescuable by salt (11 out of 15, or 73.33% for β -galactosidase and 3 out of 27, or 11.11% for catechol 2,3-dioxygenase). Thus clearly it was much 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. 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. Thus the density of β -galactosidase is 0.377 g/cm³, while the density of catechol 2,3-dioxygenase is 0.725 g/cm³. 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).

Which β -galactosidase and catechol 2,3-dioxygenase missense mutants were rescuable by salt can also be explained by considering the structures of the proteins. 13 out of the 14, or 92.86% of the missense mutants that were suppressible contained amino acid changes that were located on the surface of the proteins and thus would be expected to be more rescuable by salt. For β -galactosidase in which 11 out of 15, or 73.33% of the missense mutants were suppressible by salt, the only missense mutants that were not suppressible affected amino acids that were buried in the protein.

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 , $(\text{NH}_4)_2\text{SO}_4$, NaCH_3CO_2 , Na_2HPO_4 and $(\text{NH}_4)_2\text{HPO}_4$. Only $\text{NH}_4\text{CH}_3\text{CO}_2$ 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 in vivo 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.

REFERENCES

- Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol. 62:293-300.
- Bilski AZ, Armstrong JB. 1973. Osmotic reversal of temperature sensitivity in *Escherichia coli*. J Bacteriol. 113(1):76-81.
- Cole AE, Hani FM, Altman R, Meservy M, Roth JR, Altman E. 2017. The promiscuous sumA missense suppressor from *Salmonella enterica* has an intriguing mechanism of action. Genetics 205(2):577-588.
- Good CM, Pattee PA. 1970. Temperature-sensitive osmotically fragile mutants of *Staphylococcus aureus*. 104(3):1401-1403.
- Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, Matthews BW. 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.
- Hawthorne DC, Friis J. 1964. Osmotic-remedial mutants. A new classification for nutritional mutants in yeast. Genetics 50:829-839.
- Hofmeister F. 1888. Zur Lehre von der Wirkung der Salze. II. Arch. exp. Pathol. Pharmacol. 24:247-260.
- Hofmeister F. 1888. About regularities in the protein precipitating effects of salts and the relation of these effects with the physiological behavior of salts. Arch. Exp. Pathol. Pharmacol. (Leipzig) 24:247-260
- Hofmeister F, Zur Lehre von der Wirkung der Salze. 1888. Arch. Exp. Pathol. Pharmacol. (Leipzig) 24 (1888) 247-260; translated in W. Kunz, J. Henle and B. W. Ninham, ' Zur

- Lehre von der Wirkung der Salze' (about the science of the effect of salts: Franz Hofmeister's historical papers, *Curr. Opin. Colloid Interface Sci.* 9(2004):19-37.
- Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, Matthews BW. 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.
- Kita A, Kita S, Fujisawa I, Inaka K, Ishida T, Horiike K, Nozaki M, Miki K. 1999. An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure.* 15(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.
- Miller JH. 1972. *Experiments in molecular genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Nandi PK, Robinson DR. 1972. The effects of salts on the free energy of the peptide group. *J Am Chem Soc.* 94:1299-1308.
- Nandi PK, Robinson DR. 1972. The effects of salts on the free energies of nonpolar groups in model peptides. *J Am Chem Soc.* 94:1308-1315.
- Pegram LM, Wendorffa T, Erdmanna R, Shkela I, Bellissimoa D, Felitskya DJ, Record MT Jr. Why Hofmeister effects on many salts favor protein folding but not DNA helix formation. *PNAS* 107(17):7716-7721.

- Russell, RRB. 1972. Temperature-sensitive osmotic remedial mutants of *Escherichia coli*. J Bacteriol. 112(2):661-665.
- Von Hippel PH, Schleich T. 1969. Ion effects on the solution structure of biological macromolecules. Acc. Chem. Res. 2:257-265.
- Von Hippel PH, Wong KY. 1962. The effect of ions on the kinetics of formation and the stability of the collagen fold. Biochemistry. 1:664-674.
- Von Hippel PH, Wong KY. 1963. The collagen gelatin phase transition. I. Further studies of the effects of solvent environment and polypeptide chain composition. Biochemistry. 2:1387-1398.
- Von Hippel PH, Wong KY. 1964. The generality of their effects on the stability of macromolecular conformations. 145(3632):577-580.