

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMISTRY

Entanglements Present in Protein Sub-Populations Impact the Protein's Structure and Function

TAYLOR ROSE BEREK
SPRING 2023

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Chemistry
with honors in Chemistry

Reviewed and approved* by the following:

Edward P. O'Brien
Professor of Chemistry and of the Institute for Computational and Data Sciences
Thesis Supervisor

David D. Boehr
Associate Professor of Chemistry
Honors Adviser

C. Denise Okafor
Assistant Professor of Biochemistry and Molecular Biology, and Chemistry
Faculty Reader

* Electronic approvals are on file.

ABSTRACT

This thesis investigates how the existence of near-native like misfolded states influence the overall specific activity and structural stability of a protein population. The rate of co-translational folding is influenced by synonymous mutations, which can make certain proteins more prone to co-translational entanglements. In the protein D-alanine D-alanine ligase B and hexokinase, computational analysis has identified synonymous mutations with high probabilities of entanglements. This was proven experimentally, by growing genetically engineered cultures, isolating the protein of interest, and measuring the activity. For each case, a wildtype protein is compared to mutants with both faster and slower rates of co-translational folding. The slow mutant repeatedly displayed lower enzymatic function than the fast mutant, proving that co-translational folding rate can impact protein function. The third protein, pyruvate kinase, was exposed to a denaturing environment where it was forced to unfold and later refold. When compared to the native state, the refolded protein had similar enzyme substrate binding affinity with a maximal reaction rate less than half than that of the native state. The results of the pyruvate kinase refolding assay, along with the D-alanine D-alanine ligase B and hexokinase study, prove that a sub-population of misfolded proteins exists in the cell, and is similar enough to the native state to have some degree of function, but not to the extent or efficiency levels seen the native state.

TABLE OF CONTENTS

LIST OF FIGURES	iii
LIST OF TABLES	iv
ACKNOWLEDGEMENTS	v
Chapter 1 Introduction to Synonymous Mutations and Effects on Protein Function ..	1
Protein Formation	1
Synonymous mutations	4
Motivation.....	6
Chapter 2 Synonymous Mutations in DDLB	8
Introduction.....	8
<i>Computational Analysis of DDLB</i>	9
<i>Cell Growth and Protein Expression</i>	9
<i>Protein Purification Procedure</i>	11
Results and Discussion	13
<i>Growth and Purification Results</i>	13
<i>DDLB Specific Activity Results</i>	17
Chapter 3 Synonymous Mutations in Hexokinase	19
Introduction.....	19
Methods.....	21
<i>Coupled Assay of Hexokinase Activity</i>	21
<i>Chemotaxis of Enzyme Substrate Interaction</i>	22
Results and Discussion	23
<i>Hexokinase Growth and Specific Activity</i>	23
<i>Chemotaxis Results</i>	25
Chapter 4 Refolding of Pyruvate Kinase	26
Introduction.....	26
Methods.....	28
<i>Pyruvate Kinase Activity Measurements</i>	28
<i>Unfolding and Refolding Process</i>	29
Results and Discussion	29
<i>Native Activity Measurements</i>	29
<i>Refolded Activity Measurements</i>	30
Chapter 5 Conclusion.....	34

LIST OF FIGURES

Figure 1 – Visual comparison of the native and misfolded structure of the prion protein, created by <i>The Mayo Clinic</i> ⁴	2
Figure 2 – Example of a polypeptide chain shown with N-terminus(left) and C-terminus(right), the amino acids alternate between glycine(bottom) and alanine(top). Created using ChemDraw software	3
Figure 3 – Structure of D-alanine D-alanine ligase B, from “How synonymous mutations alter enzyme structure and function over long time scales” <i>Nature Chemistry</i> ⁶	8
Figure 4 – Graph comparing the increasing cell mass (optical density) of the three biological triplicates of the slow variant as a function of time.	13
Figure 5 – Protein concentration of fast variant DDLB represented as a function of IPTG concentration during induction	14
Figure 6 – Protein concentration of wildtype DDLB in each of the four imidazole washes represented as a function of inducing agent pulse length (in minutes).....	15
Figure 7 – Protein concentration remaining in samples of lysis buffer after passing through the resin used for column chromatography shown for both fast and slow variant DDLB	16
Figure 8 – Graphic of DDLB variants fast (orange) and slow (blue), with graphs of the translation rates across the sequence and the specific activity of the fast and slow protein, from “How synonymous mutations alter enzyme structure and function over long time scales” <i>Nature Chemistry</i> ⁶	17
Figure 9 – Chemical reaction facilitated by hexokinase, created on ChemDraw using data from the Protein Data Base ¹⁰	19
Figure 10 – Representation of the microfluidics device with the particles entering through the two inlets (right) and exiting through the five outlets (left), created on Notability	23
Figure 11 – Graphs of hexokinase specific activity, coupled assay,	24
Figure 12 – Protein concentrations (in mg) per outlet of hexokinase chemotaxis study.....	25
Figure 13 - Chemical reaction facilitated by Pyruvate Kinase, created on ChemDraw	26
Figure 14 – Lineweaver Burk plot showing linear trends in double reciprocal data (1/V as a function of 1/[ADP]) of pyruvate kinase native structure	30
Figure 15 Lineweaver Burk plot showing linear trends in double reciprocal data (1/V as a function of 1/[ADP]) of pyruvate kinase refolded structure	31
Figure 16 – Reaction rate curves of both native state and refolded pyruvate kinase,	32

LIST OF TABLES

Table 1 – Buffer Solutions Used in Purification Protocol	11
Table 2 – Pyruvate Kinase Data: velocity per substrate concentration, V_{\max} and K_M for both native and refolded enzyme samples	33

ACKNOWLEDGEMENTS

I would like to thank Dr. O'Brien and all the members of the O'Brien Research Group, for taking a chance on an undergraduate and supporting me throughout this process. I am so grateful for the opportunity to work with such incredible people on such groundbreaking research.

I would not be where I am today without my mentor, Ian Sitarik. For more than two years now, he has guided me every step of the way. He has consistently gone out of his way to listen to my ideas, correct my mistakes, and teach me everything he can. The other members of the initial experimental group—Palak Kashyap, Rong Shi, Priya Pradhan, and Paulina Frutos—have also shaped so much of the person I am today and gave me the confidence to excel in this field.

The O'Brien Lab worked very closely with the Squire Booker Lab throughout this project. The initial design of the techniques and protocols used in the cell growth and protein purification were created by Syam Neti of the Booker Lab. The final results of the DDLB experiments are credited to the brilliance and perseverance of Syam Neti and Yang Jiang.

Lastly, I am so incredibly proud of all of the authors and contributors who have been a part of the recent Nature Chemistry and Nature Communications publications and I am so thankful that I got the opportunity to be a part of it. 8

Chapter 1

Introduction to Synonymous Mutations and Effects on Protein Function

Protein Formation

Protein formation is one of the principle building blocks of cellular function and is essential in the biological life cycle. In the canonical process of protein formation, the ribosome translates each triplet nucleotide set, or codon, of a messenger RNA segment into its corresponding amino acid¹. The ribosome moves along the mRNA strand, connecting the sequential amino acids to an elongating nascent polypeptide chain. This linear sequence of amino acids is known as primary structure, however different patterns of folding and hydrogen bonding between the amino acid functional groups creates more complicated conformations known as secondary and tertiary structure¹.

Protein kinetics refers to the study of the rates at which proteins themselves are synthesized or degraded and the rate of the reaction catalyzed by the enzyme. The production, modification and degradation of proteins are all tightly regulated processes that are critical to cellular function. The process by which proteins are maintained in the cell is known as proteostasis².

Occasionally, the protein folds incorrectly, resulting a different structure than the native state. In humans, the severity of protein misfolding can be seen in Alzheimer's disease, caused by a segment of the beta amyloid protein that is only about forty residues long¹. This short, mutated polypeptide chain can be dangerous because after cleaved from the larger protein, it can aggregate—to attach itself to other like molecules—causing a buildup known as amyloid

plaque¹. The plaque builds through a positive feedback loop, triggering more misfolded beta amyloid proteins to aggregate¹. Cells have mechanisms to degrade aggregated proteins, however in the case of Alzheimer's disease, the amyloid plaque builds faster than it can be degraded and interfere with cellular and biological processes which causes the symptoms seen in people with Alzheimer's disease¹. Other similar cases are seen in Bovine Spongiform encephalopathy—more commonly known as mad cow disease—and Creutzfeldt-Jakob disease, both of which are rare but fatal¹. Stanley Pruisner famously attributed these types of diseases to the aggregation of the prion protein and won the 1997 Nobel Prize in Physiology and Medicine³. The difference in structure between the native state and the misfolded state can be seen in Figure 1⁴. One of prion's three alpha helixes does not fully form, leading to the development of a beta helix instead³. Although it is now known that the prion protein is only involved in a handful of protein aggregation diseases, Pruisner's discovery paved the way for understanding insoluble fibrous protein aggregates¹.

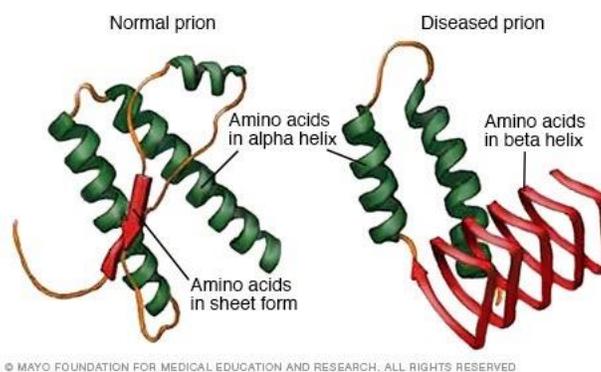


Figure 1 – Visual comparison of the native and misfolded structure of the prion protein, created by *The Mayo Clinic*⁴

Another type of mutation, seen in the collagen protein, can leave the overall structure unchanged, but destroy the molecular stability. In humans, the disease known as Osteogenesis imperfecta weakens the collagen protein, which supports hair, skin, bone, and muscle tissue¹. This single base pair mutation causes an alanine amino acid to take the place of a glycine, and this miniscule change leads to brittle bones and fragile skin¹. When an alanine group takes the place of a glycine group in a nascent polypeptide chain, the hydrogen bonding is disrupted which distorts the structure of the helix⁵. Without a strong tertiary structure, these collagen molecules

cannot provide the underlying structural support in the body, so when force is applied to the body, the tissue is more likely to shatter¹.

The mutation seen in prion changes the proteins' secondary structure—preventing it from even forming the third alpha helix—while the mutation in collagen impacts the tertiary structure—the alpha helix can still form, however changes in orientation and location of hydrogen bonding between the structures within a protein cause decreased stability. The highest level of protein structure, quaternary structure, refers the bonds between protein subunits and structures; these complexes are disturbed by major changes in structural interactions, such as protein entanglements. The term entanglement refers to the crosslinking between protein structures; the entanglements studied in this research are self-entangled, meaning the one area of a polypeptide chain is looped around a different area of the same protein⁵.

While the cell has regulatory chaperons designed to identify misfolded proteins, our lab hypothesized that there are some proteins that have nearly identical structure to the native state (lowest free energy state) however do contain changes in topological entanglement that can cause slight differentiation in structure and function². These near-native-like proteins are predicted to bypass the cellular quality control mechanisms and exist within the cell for extended time periods². This occurs because removing the entanglements forces the protein to unfold and refold some of the major structures, which requires a large amount of energy⁶.

Synonymous mutations

It is known that there are groups of codon sequences that all correlate to the same amino acid. When translated the results should be identical regardless of which codon is used, however that is not what is seen. The speed at which the codon is translated can vary and codons with the highest predicted translation elongation rate—or in other words the most efficient—are used the most⁷, and therefore have the highest tRNA populations. This phenomenon, known as codon usage bias, appears most significantly in the top one hundred most abundant proteins, indicating a correlation between efficiency of translation and efficiency of protein production⁷. This trend is not universally seen across all proteins, for example, the *Neurospora* FRQ protein presents decreased function and decreased stability when optimized for codon bias usage⁷. The FRQ protein is commonly referred to as the “FRQ clock protein” as it maintains an organism’s internal clock by regulating the circadian rhythm⁷. When the codon optimized version of the FRQ protein was translated at 25°C, the resulting protein was unable to rescue the circadian rhythm in darkness, lost molecular rhythmicity, and was unable to abolish the phosphorylation profile which led to prolonged hyperphosphorylation⁷. While the functional output was impaired, the overall concentration of proteins was consistent for both the optimized and non-optimized populations, proving that the synonymous mutation caused variation of the functionality of the protein, not simply on the protein production.

In addition to function, the codon usage also impacted the structural stability of the FRQ protein. In the hyperphosphorylated state, the FRQ protein is more sensitive to being denatured by trypsin digest and cycloheximide⁷. When a protein becomes denatured it is no longer functional, and must be either repaired through refolding, or discarded and reproduced. The fact that the FRQ protein existed in the cell in this altered state means that it was able to bypass the

biological quality controls that typically repair or discard misfolded and damaged proteins. The production of the FRQ protein is chaperoned by a heterodimer transcription factor identified as “WC” and typically has a positive feedback loop with the protein⁷. When the FRQ protein is in the hyperphosphorylated state, however, WC levels were also very low⁷. This means that not only is the protein less stable, but it also is less capable of reproducing which will lead to drastic decreases in protein concentration. The change in structural stability proves that the codon optimization directly caused a structural change in the formation of the FRQ protein.

To investigate how the structural variation occurred, the experiment was repeated at 18°C⁷. By decreasing the temperature, the ribosomal translation was also slowed, which decreases the impact that codon optimization has on the overall translation rate. These FRQ proteins were seen to be more temperature resistant than those translated at 25°C when enduring freeze thaw cycles⁷. This confirms the correlation between the codon’s impact on translation speed and the structural stability of the protein. To investigate these structural variations further, the experiment was repeated by optimizing codons from the center of the nucleic acid chain, as opposed to the N-terminus. This resulted in proteins with improved structural stability, due to an increased ability to abolish hyperphosphorylation, making them less susceptible to trypsin digests and cycloheximide⁷. While this increased the protein stability, the function was still impaired, as they exhibited the same arrhythmic conidiation as the other codon optimized mutants⁷. The differences present between the mid-strand and N-terminus optimized proteins proves that the functional impacts of codon optimization are location specific. The FRQ protein is evidence that not only do synonymous mutations change translational speed, but that the resulting structure is dependent on where within the mRNA sequence synonymous mutations

occur, indicating that the functional and structural changes seen are dependent on the changes in folding kinetics of the amino acid strand.

Motivation

My research aims to provide evidence proteins can adopt near-native like conformations with changes in entanglement, and that they have a measurable impact on the function of the protein. By genetically engineering cell cultures to push the extremes of protein folding rates and synthesize proteins at the fastest and slowest translational speeds possible, I hope to identify a population with near-native-like structure and non-native like entanglements. These entanglements will alter the proteins' structure, meaning a change should be seen in the stability and activity of the protein. Based on computational data, several proteins have been identified as candidates for experimental testing; this paper discusses the methods and results of studies done on three proteins, D-alanine D-alanine ligase B, hexokinase, and pyruvate kinase.

Producing evidence of stable protein subpopulations that differ in function or structure for long time periods will provide evidence supporting simulations that show non-native entanglements caught in long-lived kinetic traps⁶. Correlating the presence of these subpopulations to synonymous mutations confirms the theory that the entanglements are created by changes in protein folding rates during translation. Most synonymous mutations are considered silent because they do not impact the final protein structure, so further proof that the alterations in translational speed from synonymous mutations can lead to misfolding and long-lived entanglements would have a large impact on the field of biochemistry. Not only will it allow us to more accurately predict how protein function will be altered by genetic mutations, it

will also improve efficiency of genetically engineered proteins by optimizing the folding kinetics. As mentioned earlier, protein misfolding is the direct cause of many deadly diseases. While proving that synonymous mutations can decrease functionality is not a cure, it does provide insight as to why certain mutations have drastic consequences for protein function while others have no effect at all, and why certain misfolded proteins aggregate and must be degraded by the cell while others exist for long periods of time and can still catalyze reactions in their misfolded state.

Chapter 2

Synonymous Mutations in DDLB

Introduction

This experiment investigates the long-lived kinetic traps of the protein D-alanine D-alanine ligase B (DDLB) that we predict are caused by changes in non-covalent lasso entanglements. While these entangled states are not the global free energy minimum, the large energy barriers to unfolding and refolding to the native state are often too high for the ambiently available thermal energy⁶.

To test these results experimentally, the RNA plasmid must contain the sequence of the protein of interest and must be engineered using the lac operon to manipulate the protein expression. The lac operon is the mechanism used to turn on and turn off translation of the part of the genome where the lactase enzyme is typically encoded. The presence of lactose, the sugar commonly found in milk, triggers the expression of the lactase enzyme to digest the molecule. In the design for the DDLB plasmids, the lac operon is used, but the sequence pertaining to lactase is replaced with the sequence for DDLB, allowing the production of the desired protein to be controlled by the addition of lactose—or synthetic lactose such as isopropyl beta-D-1thiogalactopyranoside, also known as IPTG.

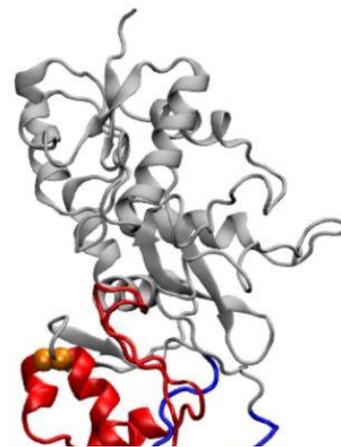


Figure 2 – Structure of D-alanine D-alanine ligase B, from “How synonymous mutations alter enzyme structure and function over long time scales” *Nature Chemistry*⁶

Methods

Computational Analysis of DDLB

Using coarse grain molecular dynamic simulations, the O'Brien lab identified D-alanine D-alanine ligase B as a prime candidate for observing kinetically stable co-translational entanglements. The coarse-grain model used the crystal structure PDB 4v9d as the E. coli ribosome and PDB 5jte as the A- and P-site tRNA molecules⁶. A continuous synthesis protocol was used to model the A-site tRNA binding, peptidyl transfer, tRNA translocation, and ribosome trafficking⁶. The synthesis of different monomeric proteins that can be expressed in E. coli were rated based on the scoring function seen below in equation (1), where τ_F^{post} is the mean post-translational folding time.

$$\text{Score} = \frac{1}{2} \times \left[\left(1 - \left(1 + \langle \tau_F^{\text{post}} \rangle \right)^{-0.2} \right) + \theta_{\text{binding}} \right] \times 100, \quad (1)$$

D-alanine D-alanine ligase B was the highest scoring enzyme, with a score of 99.5, indicating the high possibility of long-lived kinetic traps existing in the enzyme and specifically occurring near the substrate binding location—represented by the term θ_{binding} ⁶. The simulation was repeated on the fastest and slowest translating mutations, with the goal of finding specific activity differences using the Markov state modeling procedure⁶.

Cell Growth and Protein Expression

A primary stock of DDLB was expressed in the K-12 strain of Escherichia coli using the plasmid designed by the O'Brien and Booker Lab. The cell line was grown in 1-liter cultures using Lysogeny Broth. This broth, often referred to as LB Media, contains 10 grams of tryptone, 10 grams of sodium chloride, and 5 grams of yeast per liter. After being sanitized in the autoclave, 50 $\mu\text{g}/\text{mL}$ of the antibiotic, kanamycin is added to the LB media. A small volume, ~ 200 mL is set aside, and agar is added, so when cooled the media hardens to a soft solid state. Agar plates are made in petri dishes, and bacteria strains

are grown on the surface of these agar plates. From the cryo-stock of the original cell line, frozen cells were streaked on LB media agar plates under sterile conditions. The plates were streaked four times in each direction; one plate was made for each cell line: wild-type, fast mutation, and slow mutation. The plates were incubated for 12-24 hours at 37° Celsius, and then a single colony was chosen for inoculation of each of the seed cultures. By beginning a cell culture on an agar plate, the growth of different cell colonies can be observed as small individual dots on the surface of the agar. Each colony is initiated by a single E. coli cell that is replicated repeatedly. To eliminate genetic diversity, one singular colony was selected and moved to a new media to begin growing again, theoretically creating a new culture in which all the cells are genetically identical. Mutations are bound to occur as the cells reproduce, however single colony selection helps to minimize the genetic variety of the culture¹.

The seed cultures consisted of 50 mL of LB media and 50 µg/mL kanamycin in 125 mL flasks and were incubated for 12 hours mixed at 200 rotations per minute (rpm) to maintain circulation throughout the media. The main growth, 1 liter of LB media treated with kanamycin, was inoculated with 10 mL of the seed culture. The cultures were grown in triplicate, so each run contained three 1-liter cultures for each of the three cell lines. The cultures were incubated at 37° Celsius and mixed at 180 rpm. Every 30 minutes a 1 mL sample was collected and the ocular density at 600 nm was recorded. When the ocular density reached an absorbance value of 0.6, the flasks were removed from the incubator and cooled to 25° Celsius. Once cooled, the cultures were induced with 5 mL of 0.25 mM solutions of IPTG. The cultures were kept at 25° Celsius for 5 hours to allow for complete protein expression, and then centrifuged at 15,000 g for 15 minutes in 500 mL increments. The supernatant was discarded, and the accumulated solid cell mass was flash-frozen via submersion in liquid nitrogen for 10-30 seconds. The cell samples were labeled and stored at -80° Celsius until used for purification.

Protein Purification Procedure

The protein samples were purified using a fast protein liquid chromatography machine (FPLC) fit with a nickel resin column. The resin column was cleaned overnight with a solution of 20% ethanol and 80% deionized water, at 4 degrees Celsius. Six buffers were made according to the concentrations listed in Table 1. For each purification, 10 mg of the protein sample were defrosted in 25 mL of lysis buffer at 4° Celsius until the mixture is fully homogenized. The cells were then lysed via sonification. With the sample in the ice bath the Bradson Sonifer 450 ran fifteen cycles of 45 second pulses—with emission lasting 0.4 seconds at an output power of 7—and 120 second intervals between pulses. The lysate is then centrifuged at 45,000g for 45 minutes, 4° Celsius, in 500 mL increments.

Table 1 – Buffer Solutions Used in Purification Procedure

Lysis Buffer		
Reagent	Concentration	Mass per 1 L
Tris HCl	20 mM	3152 mg
NaCl	200 mM	11688 mg
Imidazole	5 mM	340.385 mg
MgCl ₂	5 mM	476.055 mg
BME	2 mM	140 uL

Wash Buffer A		
Reagent	Concentration	Mass per 500 mL
Tris HCl	20 mM	1576 mg
NaCl	200 mM	5844 mg
Imidazole	20 mM	680.77 mg
MgCl ₂	5 mM	238.0275 mg
BME	2 mM	70 uL

Wash Buffer B		
Reagent	Concentration	Mass per 500 mL
Tris HCl	20 mM	1576 mg
NaCl	200 mM	5844 mg
Imidazole	40 mM	1361.54 mg
MgCl ₂	5 mM	238.0275 mg
BME	2 mM	70 uL

Wash Buffer C		
Reagent	Concentration	Mass per 500 mL
Tris HCl	20 mM	1576 mg
NaCl	200 mM	5844 mg
Imidazole	80 mM	2723.08 mg
MgCl ₂	5 mM	238.0275 mg
BME	2 mM	70 uL

Elution Buffer		
Reagent	Concentration	Mass per 500 mL
Tris HCl	20 mM	1576 mg
NaCl	200 mM	5844 mg
Imidazole	300 mM	10211.55 mg
MgCl ₂	5 mM	238.0275 mg
BME	2 mM	70 uL

Storage Buffer		
Reagent	Concentration	Mass per 500 mL
Tris HCl	20 mM	1576 mg
NaCl	200 mM	5844 mg
MgCl ₂	5 mM	95.211 mg
BME	2 mM	70 uL
Glycerol	10 %	50 mL

The resin column was equilibrated with three column volumes of lysis buffer and then two column volumes of supernatant—having been filtered at 0.2 μm —was loaded onto the column at a flow rate of 3-5 mL/min. All the prepared buffers were also filtered at 0.2 μm and three column volumes of each of the three wash buffers were rinsed through the column consecutively. When the final buffer, the elution buffer, was rinsed through column, the runoff was collected in 1 mL fractions. The internal spectrophotometer on the FPLC was used to select the fractions with the highest protein concentration. These samples were then concentrated using 10 kDa filter tubes centrifuged at 3,500 rcf in two-minute increments. After the elution buffer had drained, and the protein was transferred to storage buffer and cleaned using PB 10 columns to decrease the salt concentration. The isolated proteins can be analyzed or frozen from this point, typically concentrations were measured using Bradford analysis and the protein identity and size was affirmed using an electrophoresis SDS-PAGE.

As more experiments were run, the protocol was adjusted to receive the best results. The growth was optimized by controlling the pulse chase—the amount of time the cells are exposed to IPTG—and the concentration of IPTG used. The pulse chase was performed by growing eighteen cultures of the wildtype protein simultaneously and assigning six pulse lengths in triplicate. The pulses ranged from 10 minutes to 350 minutes, and the protein concentration of each later were measured and compared. The concentration optimization procedure was done using six concentrations of IPTG, ranging from 0.01 μM to 0.25 μM . The purification was optimized by testing the protein concentration in the runoff of each buffer addition to the column. The results were compared to determine what stages of the purification have the most protein exiting the column.

Results and Discussion

Growth and Purification Results

To prove that any differences seen in protein structure were caused by the genetic mutation and not environmental factors, it is necessary to maintain similar growth conditions for each culture. Using biological triplicates was essential in monitoring this error; the difference in growth rates between triplicates was indicative of an outside variable or an unequal growth environment. Figure 3 shows the growth curves of biological triplicates of the slow mutation. Each of the cell cultures were grown side-by-side in the same incubation machine and mixed with the same shaking device, however the optical density values clearly show some variation, specifically starting at 125 minutes after inoculation. While this difference is not significant, it was important to monitor the optical density as often as possible so that the variation can be accounted for when moving on to the next step.

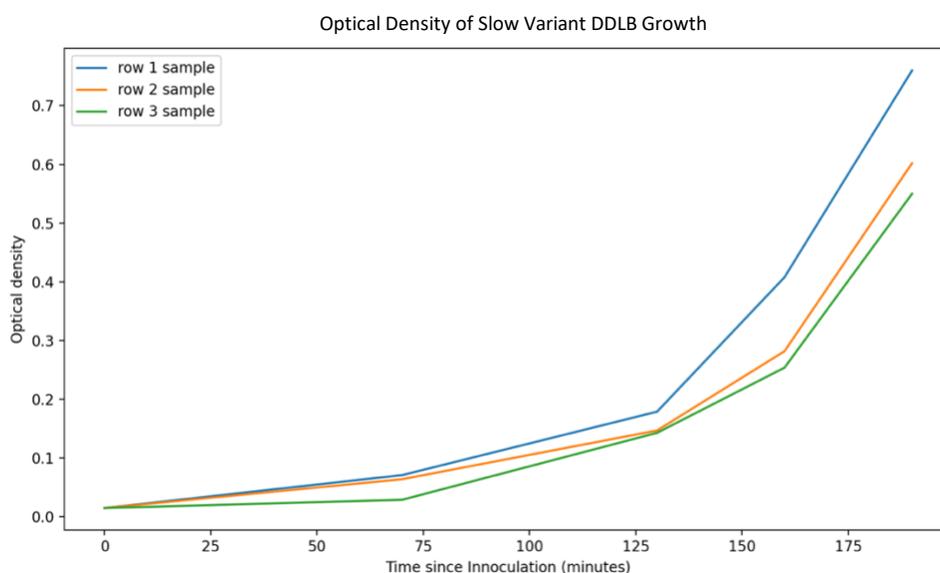


Figure 3 – Graph comparing the increasing cell mass (optical density) of the three biological triplicates of the slow variant as a function of time.

The results of the IPTG optimization experiment are seen in Figure 4. There is a direct correlation between IPTG concentration and protein concentration when looking at low IPTG concentrations, however for concentrations of IPTG greater than 0.05 mM the protein concentration decreases as a function of IPTG concentration. This is due to the stress imposed on the cell from too high concentrations of the synthetic sugar⁸. While IPTG is an easier and more efficient way to induce protein expression in a laboratory setting, it contains toxins that have been proven to directly correlate with a drop in cell vitality⁸.

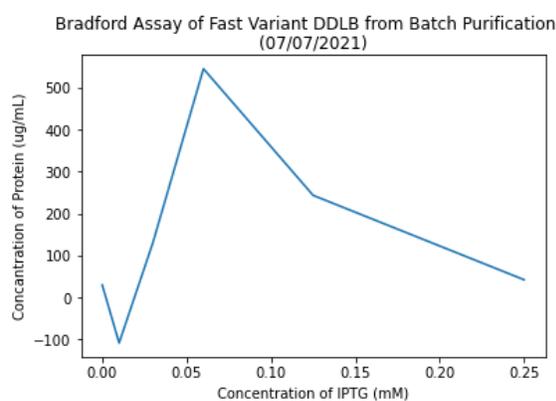


Figure 4 – Protein concentration of fast variant DDLB represented as a function of IPTG concentration during induction

The next variable optimized was pulse length. As seen in Figure 5, there are clearly two peaks in protein expression: one occurring 15 minutes after induction and the other beginning more than 2 hours after induction. Comparing the concentrations per imidazole wash argues that the second peak occurring at 125 minutes contains protein with stronger nickel interactions—such as genetically engineered DDLB—than the protein in the first peak because it withstands washes A and B but is pushed off the column by the stronger washes, C and the elution buffer. Figure 5 not only demonstrates that the proteins must be exposed to IPTG for several hours to see full expression of the desired protein, but also models how using four increasingly

concentrated wash buffers can strip away unwanted proteins while still maintaining high concentrations of the intended protein.

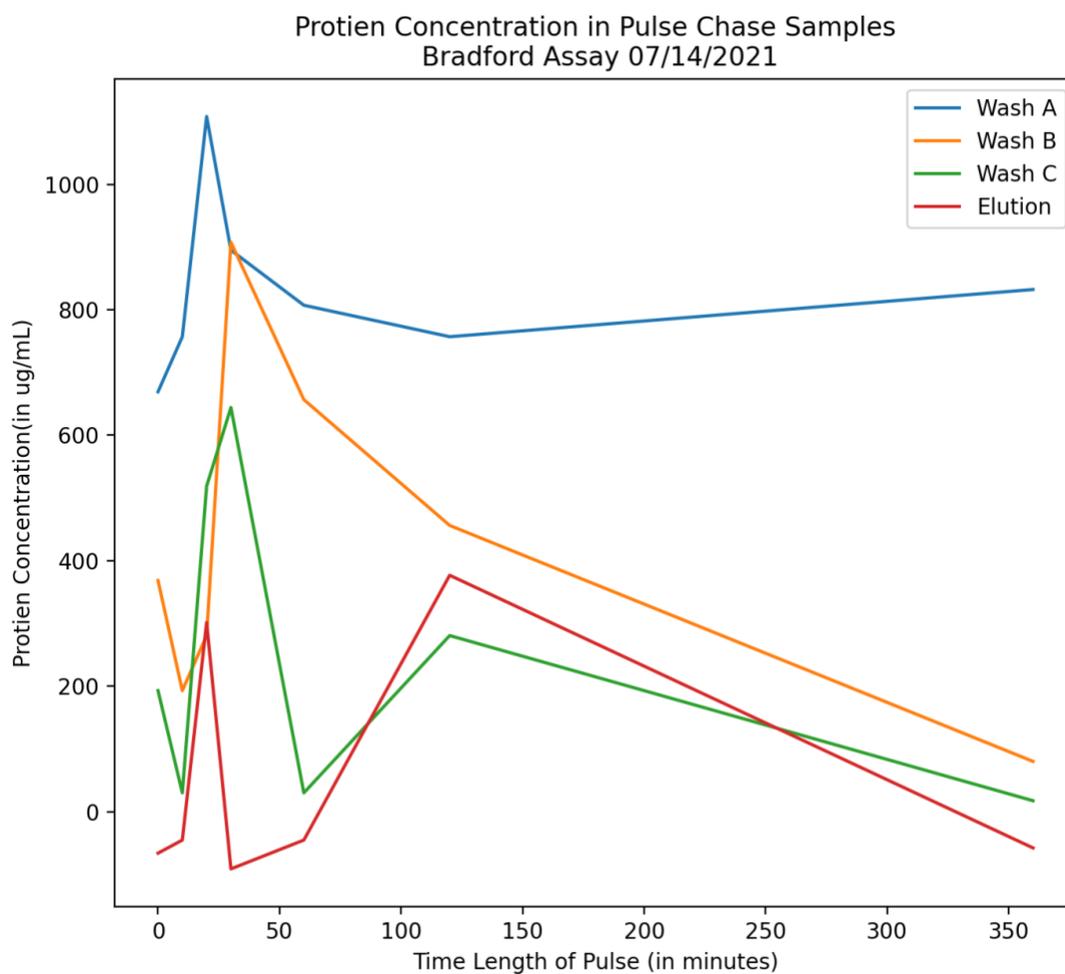


Figure 5 – Protein concentration of wildtype DDLB in each of the four imidazole washes represented as a function of inducing agent pulse length (in minutes)

The final optimization experiment performed was to confirm that the maximal amount of protein was being loaded onto the column during the prewash stages of the purification. The results in Figure 6 show the protein concentration of the lysis buffer before being loaded onto the column (initial reading) and then the protein concentrations of the liquid that has already passed through

the column. After the first addition of lysis buffer to the column (prewash 1), the liquid that passes through has a much lower protein concentration, indicating that a significant amount of protein from that sample was loaded onto the column. This process is repeated with an addition of new lysate (prewash 2) which shows that some protein has been removed again from the lysate and loaded onto the column, however, on average the amount of protein added to the column from the second addition is less than that of the first addition. When the third addition (prewash 3) is performed, the protein concentration after passing through the column is almost as high as it was initially, meaning little to no protein is added to the column in the third addition. It can be concluded that the amount of protein capable of being stored on the column is reached with just two rinses of lysate buffer; more than two additions of lysate buffer add nothing to the purification or protein sample.

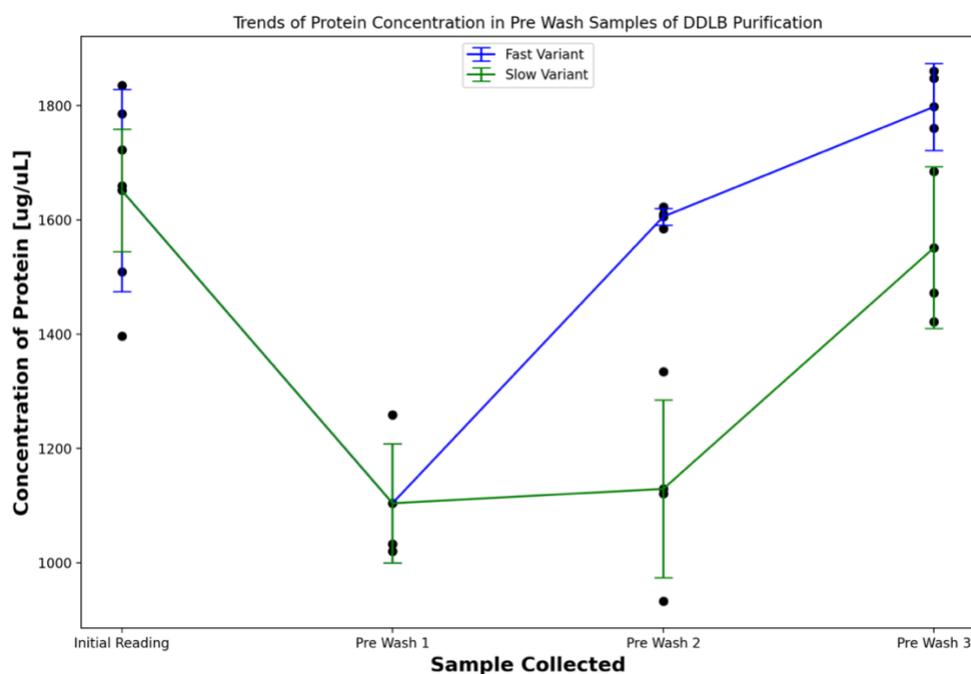


Figure 6 – Protein concentration remaining in samples of lysis buffer after passing through the resin used for column chromatography shown for both fast and slow variant DDLB

DDLB Specific Activity Results

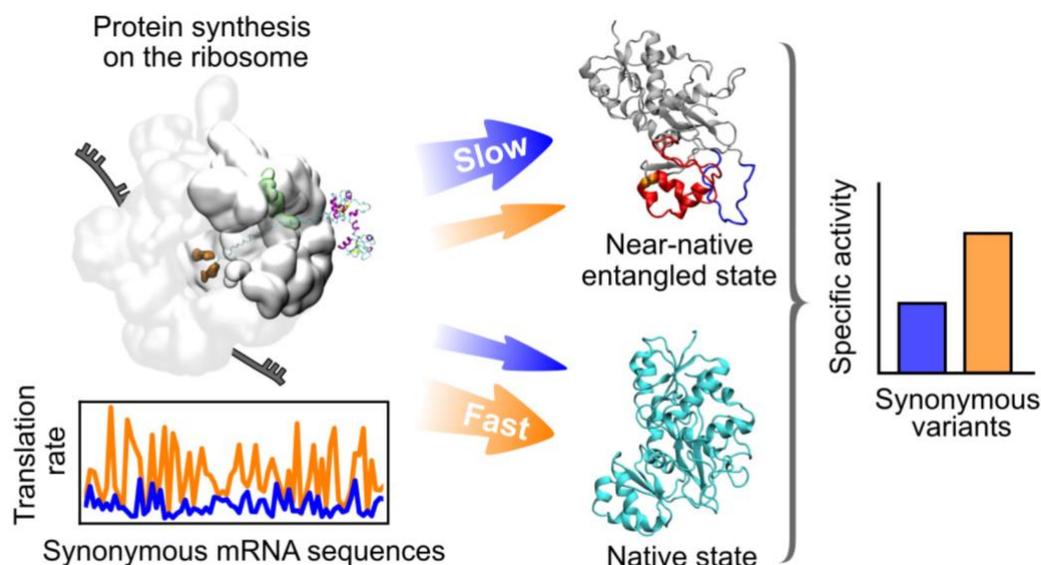


Figure 7 – Graphic of DDLB variants fast (orange) and slow (blue), with graphs of the translation rates across the sequence and the specific activity of the fast and slow protein, from “How synonymous mutations alter enzyme structure and function over long time scales” *Nature Chemistry*⁶

Overall, it was observed that the cell cultures containing the fast co-translational folding mutation had higher levels of protein expression than those containing the slow mutation⁶. This supports the idea that the mutation is increasing the rate of translation. Structural analysis revealed that 37% of the protein population produced by cells containing the slow mutation were entangled, and 30% of the fast mutation’s population were entangled⁶, which means that although the fast mutation produced more protein, the slower mutation had a higher percentage of entangled proteins. The specific activity measurements and additional growths were performed by other members of the O’Brien Lab and Booker Lab in corroboration with the experiments described above. In a study of the reaction turnover rate, they discovered that

across five biological replicates, the slow mutant's activity levels were 88% that of the fast mutants⁶. Using a 95% confidence interval, the range of activity rates (measuring $K_{cat}SLOW/K_{cat}FAST$) is 81.3% to 94.8%⁶. In Figure 7, the slow mutated DDLB, represented in blue, has a lower amplitude in terms of both translation rate and specific activity than the fast mutated DDLB, represented in orange.

Chapter 3

Synonymous Mutations in Hexokinase

Introduction

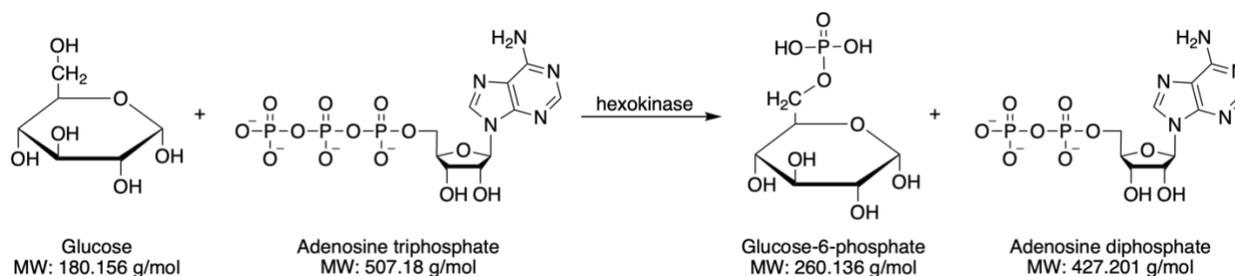


Figure 8 – Chemical reaction facilitated by hexokinase, created on ChemDraw using data from the Protein Data Base¹⁰

Hexokinase (BRENDA: EC 2.7.1.1) is an essential enzyme in the process of glycolysis, the primary energy-producing pathway in most living organisms⁹. As the first enzyme in this pathway, hexokinase catalyzes the phosphorylation of glucose to glucose-6-phosphate, which initiates metabolism¹⁰. The enzyme has been extensively studied for its role in glucose homeostasis, and a direct link has been found between hexokinase deficiency and diseases of the blood—specifically hemolytic anemia⁹. While it is normal for the enzymatic activity of hexokinase to decrease as blood cells age⁹, genetic mutations and protein misfolding can cause long lasting decreases in hexokinase activity. For example, it was recently discovered that the inability to repair unfolded hexokinase proteins may be the cause of insulin resistance in those with diabetes¹¹. High concentrations of methylglyoxal and methylglyoxal-mediated protein glycation are correlated with protein misfolding and the activation of unfolded protein response, which describes a state of stress where the endoplasmic reticulum cannot degrade proteins fast enough, creating an accumulation of unfolded or misfolded proteins¹². The endoplasmic

reticulum is overwhelmed because the glycation of hexokinase by methylglyoxal attacks the chaperonin complexes—which are needed for the protein to fold correctly—and impacts proteasome subunits—which mark and transport proteins for degradation—so the cells ability to repair the damage is limited¹¹. A therapeutic treatment to increase expression of the Glo1 gene may have the potential to prevent protein glycation by methylglyoxal, which would treat diabetes related complications and may have the potential to reverse the disease all together¹¹.

To measure the specific activity of the enzyme Hexokinase in the UV-visual spectrometer, the reaction must be coupled with the reaction mechanism of Glucose-6-phosphate-Dehydrogenase. One of the products of Glucose-6DH is NADPH, which absorbs light at a wavelength of 340 nm¹⁰. The high reaction rate of Glucose-6DH means that the coupled reaction rate will be almost entirely dependent on the Hexokinase reaction rate. By measuring the absorbance overtime of 340 nm light, we can calculate the rate at which the NADPH concentration increases, which is indicative of the Kcat of Hexokinase.

Chemotaxis is an additional method of measuring reaction specific reaction rates. The main principle of chemotaxis is that when there is a large difference in concentration within a confined area, referred to as a chemical gradient, the sensitivity of a molecule or cell can trigger movement across the gradient that differs from the natural diffusion⁵. The goal of this experiment is to monitor the lateral movement of enzyme and substrate to separate the native proteins from the non-native like proteins. The enzymes that travel further across the gradient have a stronger chemical attraction to the substrate. This experiment is motivated by the desire to boost the low populations of self-entangled states. By using chemotaxis to separate the protein

populations according to substrate attraction, the samples that present the weakest substrate interaction will have a higher concentration of non-native proteins.

Methods

In this experiment, two mutations of hexokinase were selected based on computational studies as most likely to alter the rate of translation, and therefore have the highest probability of entanglement, while still retaining an overall native-like structure. Using the software Gene script, a plasmid was designed containing the genetic sequence for hexokinase—identified as 3B8A in the protein database—containing the synonymous mutations identified in the computational models. The protein was designed in the same manner as DDLB, using *E. Coli* K12 and the corresponding lac operon. The cell cultures were grown in triplicate and hexokinase was expressed and purified following the same procedure used for the D-alanine D-alanine ligase B cultures.

Coupled Assay of Hexokinase Activity

The specific activity was measured by a spectrophotometer at 340 nm. Four solutions were used in the reaction. The buffer solution, 20 mM Tris HCl, 200 mM NaCl and 10 mM MgCl₂, was mixed and brought to a pH of 7.4. The substrate solution contained 2.0 mM ATP and 0.5 mM NADP, using the buffer solution as the solvent. The hexokinase solution contained 5.0 nM hexokinase and the GDH solution contained 2 units of glucose-6-dehydrogenase (100

units per mL). The two-enzyme solution can be frozen in 1.0 mL and 0.5 mL samples using liquid nitrogen if glycerol is added to at least 15% of the total. The glycerol will help preserve the protein structure by disrupting the crystal structure of ice, however it cannot preserve the substrate solutions, those must be made on the day of the experiment.

To begin each run, the cuvette (1 cm pathlength) was filled with 855 uL of buffer solution, 25 uL of hexokinase solution, and 20 uL of GDH solution. This solution was thoroughly mixed, it must be homogenous. The sample is spiked with 100 uL of the substrate solution and the change in absorption is monitored.

Chemotaxis of Enzyme Substrate Interaction

After the specific activity trials, an additional experiment was performed with the purpose increasing the concentration of the non-native-like proteins in the samples. This was done through chemotaxis, using a microfluidics device. As seen in Figure 9, a model of the device, there are two inlet holes, one inlet will contain the enzyme and the other will contain the substrate. It is essential to keep a steady flow rate for each inlet, so the solutions were kept in a syringe, with the needle connected to the inlet via delicate tubing and the plunger compressed by a syringe infusion pump. When the solutions from the inlets enter the microfluidics device, the mixture travels through a groove inside the resin of the microfluidics device and exits through one of five different outlets.

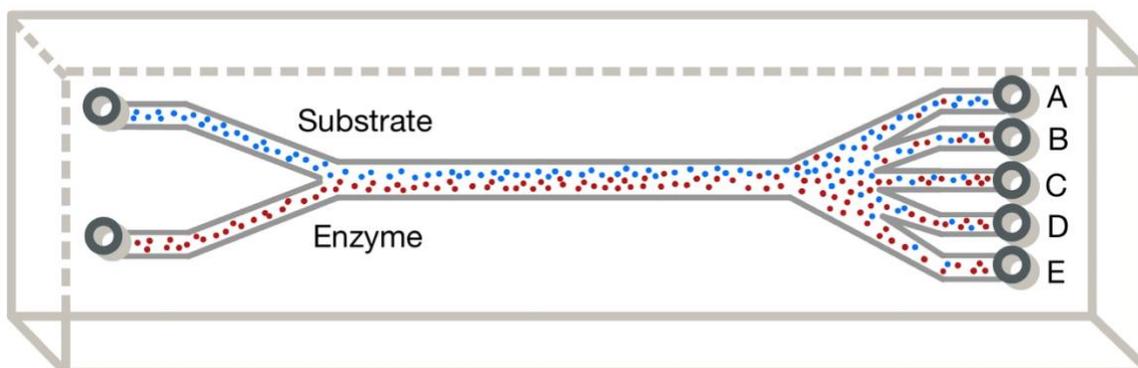


Figure 9 – Representation of the microfluidics device with the particles entering through the two inlets (right) and exiting through the five outlets (left), created on Notability

The experiment was performed using a control group—substrate solution with zero concentration—and a treatment group—substrate solution with substrate concentration. The enzyme concentration and the flow rate are kept consistent across trials.

Results and Discussion

Hexokinase Growth and Specific Activity

While the growth and purifications were successful for hexokinase, the specific activity measurements posed challenges. Using Beer's Law to transform the absorption values into concentrations, the change in NADP can be measured as a function of time. The molar extinction coefficient for both NADH and NADPH at 340 nm light is $6.220 \text{ cm}^{-1}\text{mol}^{-1}$, this is specific to each molecule and each wavelength¹³. In equation (2) shown below, A represents absorption (which is a unitless value), epsilon represents the molar extinction coefficient (mol^{-1}

$l\text{cm}^{-1}$), b represents the pathlength of the photon through the sample (cm), and c represents the concentration (mol)¹.

$$A = \varepsilon \times b \times c \quad (2)$$

Figure 10 shows the results after the absorption was converted to NADPH concentration using Beer's Law. In the first graph on the left of the Figure, the rate of NADPH production (in μmols) is plotted as a function of time. The derivate of those plots are shown in the middle graph of Figure 10, which has the rate of NADPH production per minute plotted as a function of time. Both graphs have each of the five technical replicates performed, and the final graph on the right side of Figure 10 has the maximum rate of NADPH production per minute per mg of hexokinase plotted for each of the five replicates. It is clear in all three of these graphs that replicates 2, 3, and 4 had a higher reaction rate than replicates 1 and 5. This trend, is due to the temperature sensitivity of hexokinase. The specific activity values were reasonably close to the expected value of $120 \mu\text{mol}/\text{min}/\text{mg}$ ¹¹ but in the future the error margins could be decreased by keeping the temperature controlled. Unfortunately, due to supply chain issues the hexokinase assay could not be finished at this time this was the extent of the specific activity results.

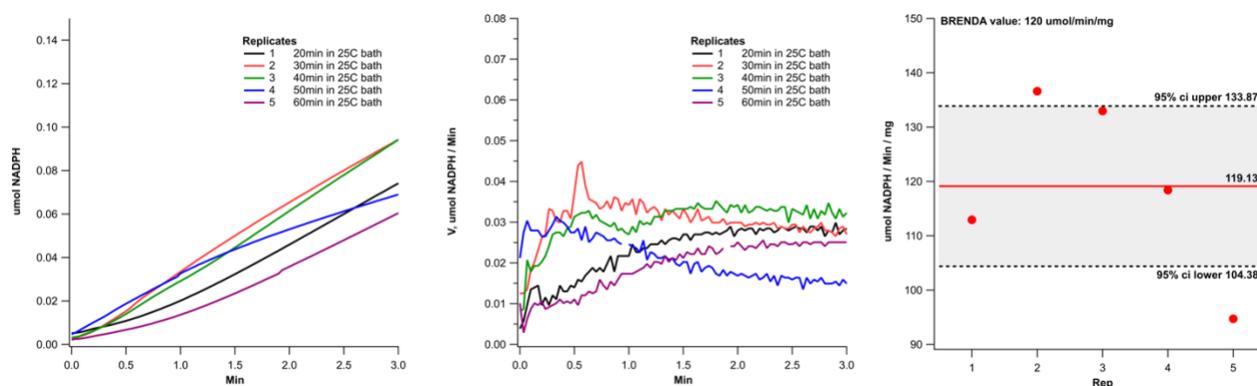


Figure 10 – Graphs of hexokinase specific activity, coupled assay, created by Ian Sitarik of the O'Brien Research Group

Chemotaxis Results

The results of the chemotaxis experiment clearly display a difference in enzyme concentration across the five outlets. Figure 11 displays the protein concentration across the five outlets for both the treatment case—shown in blue—and the control group—shown in orange. Outlet 5 contains the proteins that traveled the farthest across the gradient, shown as outlet A in Figure 9, and outlet 1 contains the proteins that traveled the least, corresponding to outlet E in Figure 9. The treatment case showed a larger concentration of enzymes migrate from their initial position, representing a significant pull from enzyme substrate interactions. The enzyme control—which did contain substrate—would not have any substrate enzyme interaction occurring, which is why this pull was not seen in the control case. This technique can be used in the future to isolate protein samples with lower activities which are more likely to contain entangled proteins.

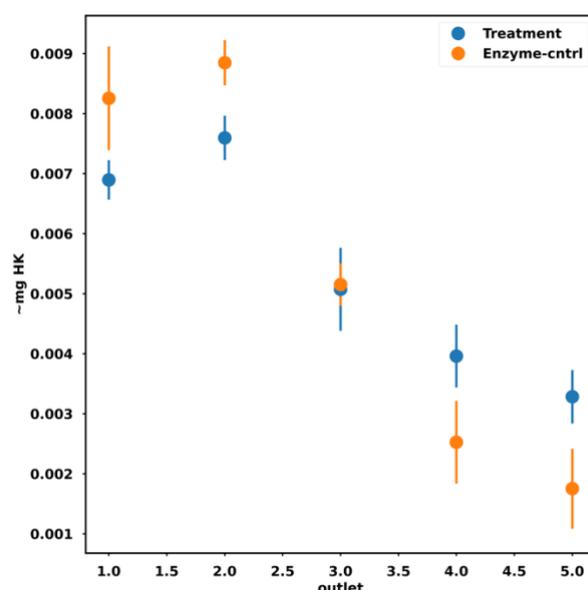


Figure 11 – Protein concentrations (in mg) per outlet of hexokinase chemotaxis study comparing treatment case (in blue) to the control case (in orange), created by Ian Sitarik of the O’Brien Research Group

Chapter 4

Refolding of Pyruvate Kinase

Introduction

Another enzyme that is essential to the regulation of glycolysis is pyruvate kinase (BRENDA: EC 2.7.1.40), whose function is to catalyze the phosphorylation of Adenosine Diphosphate to Adenosine Triphosphate and convert phosphoenolpyruvate to pyruvate. In this reaction pyruvate kinase is both producing energy in the form of ATP and acting as a key determining factor of the specificity and sensitivity of kinase phosphorylation throughout glycolysis¹⁴.

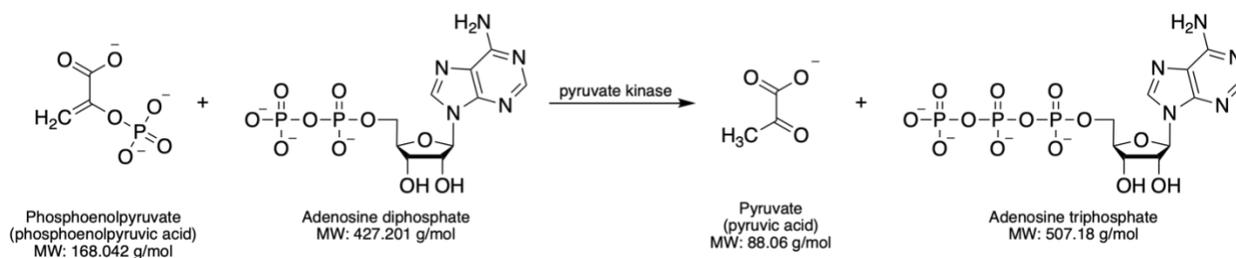


Figure 12 - Chemical reaction facilitated by Pyruvate Kinase, created on ChemDraw using data from the Protein Data Base¹⁵

The reaction catalyzed by pyruvate kinase, shown in Figure 12, is the final step in glycolysis. The enzyme facilitates the transphosphorylation from phosphoenolpyruvate to adenosine diphosphate in two principal steps. In the first step, one of the negatively charged, nucleophilic oxygens on the outer phosphate group of ADP attacks the phosphorus in the phosphoenolpyruvate molecule, causing the existing oxygen-phosphorus bond in

phosphoenolpyruvate to break and creating an enolate¹⁶. In the following step, the enolate is protonated to produce pyruvate. This reaction had previously been thought of as irreversible due to the extensive Gibbs free energy produced, however, recent findings reveal the exothermic reaction to have a magnitude of only 27kJ¹⁶. This reaction is significant because it is considered one of the three rate determining steps in glycolysis. The enzyme can be safely stored at -20°C without losing activity¹⁷. The cations Mg²⁺ and K⁺ must be present to measure the maximum reaction rate of pyruvate kinase¹⁷.

Mutated forms of pyruvate kinase have been documented; an in vivo study of glycolysis enzymes in bacteria cells induced co and post translational changes in the protein structure. Although no results were recorded on the impact on protein activity, the study proves that pyruvate kinase can be phosphorylated inside the cell. Phosphorylation is the most common type of post-translational modification seen in proteins and does not impact co-translation modifications, which are more relevant to this study as the entanglement forms during the protein folding process¹⁸. In this experiment, pyruvate kinase will be forced to unfold, or denature, and then allowed the opportunity to refold again. The activity of the protein post refolding will be compared to the activity seen in identical protein that were not exposed to the denaturing process. If we see a decrease in enzyme function in the refolded protein when compared to the native state, it will prove that misfolding occurred in the refolding process.

Methods

Pyruvate Kinase Activity Measurements

Commercially produced samples of the protein, pyruvate kinase was obtained from gene script and frozen via liquid nitrogen in 500 microliter samples at a concentration of 0.1 mg/mL. The buffer used in all the specific activity experiments consisted of 50 mM imidazole, 120 mM potassium chloride, and 62 mM magnesium sulfate. To calculate the reaction's maximum velocity, the substrates must be at a concentration at least ten times greater than their K_m value, so solutions of 45 mM ADP, 45 mM PEP, and 6.6 mM NADH were assembled. The reactions were performed in a 700 uL quartz cuvette and absorbance of photons at 340 nm was measured. For each experiment the 700 uL sample consisted of: 630 uL of the buffer, 50 uL ADP solution, 25 uL NADH solution, 10 uL of LDH and 10 uL of the Pyruvate kinase solution and was induced with 50 uL of the PEP solution. The absorbance value at 340 nm was measured before the PEP was added, and then for 2-3 minutes after the addition. The PEP was mixed into the solution using a 200 uL pipette for 10 seconds and then inverted by hand—sealed with a lid to prevent sample loss or contamination—for an additional 10 seconds. The total time over which the PEP was added and mixed into the solution was no more than 30 seconds. The decrease in 340 nm absorbance post PEP addition was recorded and used to calculate the reaction rate.

The initial study contained a concentration of ADP and PEP high enough that the reaction rate seen represented the maximal velocity, however in following experiments, the concentration of ADP was varied. To solve for the K_m value, several ADP concentration levels were used: 0.1 mM, 0.3 mM, 0.7 mM, 1.5 mM, 3.0 mM, and 6.0 mM. Stock solutions of 84 mM, 21 mM, and 4.2 mM ADP were prepared and diluted to create the 0.1-6.0 mM range used.

Unfolding and Refolding Process

To denature the protein, a solution of 6M guanidinium chloride and 100mM β -mercapto-ethanol are mixed with 200 uL of the 1mg/mL pyruvate kinase stock. This solution is left at room temperature (25°C) for 12 hours and then diluted one hundred-fold and stored at 4°C. A similar solution containing 200 uL pyruvate kinase—taken from the same stock solution fraction as the original sample—and mixed with corresponding amounts of β -mercapto-ethanol and guanidinium chloride but immediately diluted one hundred-fold and placed at 4°C. Both samples now contain the same concentrations of 0.1 micrograms per milliliter pyruvate kinase, 0.06 Molar guanidinium chloride, and 1 millimolar β -mercapto-ethanol, however the first sample was denatured before being diluted and cooled. 50 uL of the protein solution is used per reaction, and the K_m protocol described above is used for both protein samples.

Results and Discussion

Native Activity Measurements

The results of this experiment showed that in the native state, the maximum velocity matched the expected value, however when in the refolded state, the maximum velocity achieved dropped significantly. The absorbance value at 340nm was recorded throughout each reaction and converted to concentration of NADH using Beer's Law⁵. To calculate the K_m and V_{max} values, a Lineweaver Burk Plot was used by graphing the inverse of the reaction rate as a function of inverse substrate concentration⁵ and is shown in Figure 13.

Lineweaver Burk Plot of Native State Pyruvate Kinase Reaction

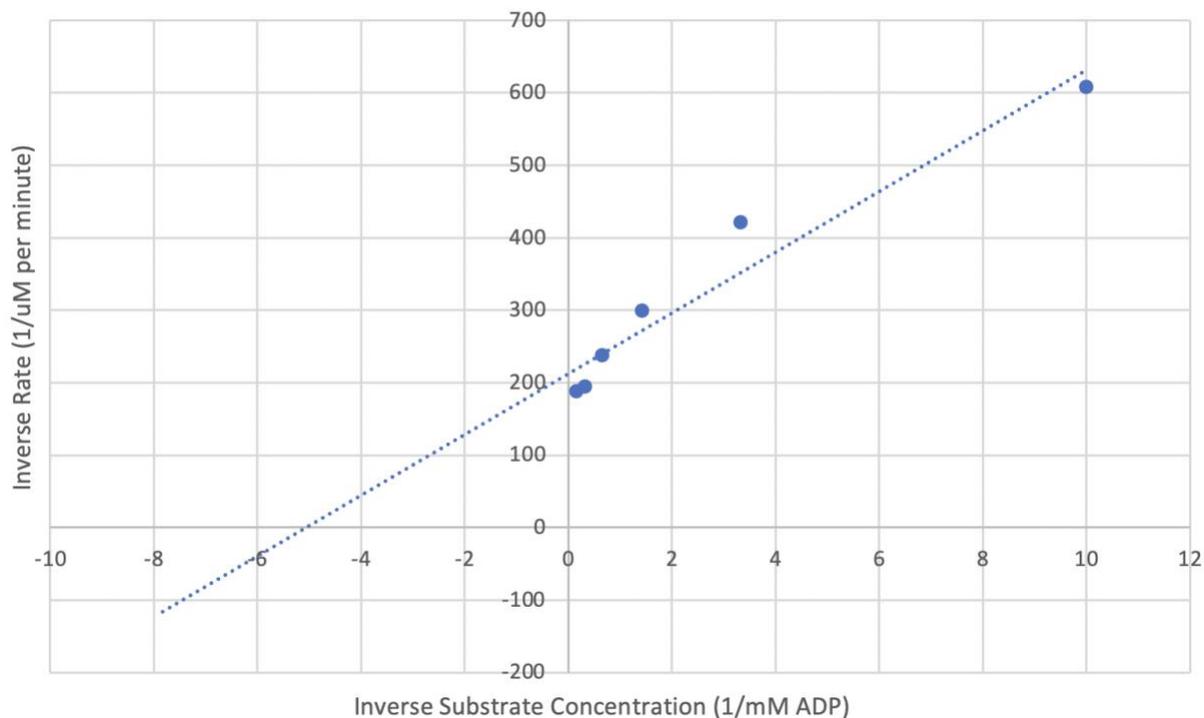


Figure 13 – Lineweaver Burk plot showing linear trends in double reciprocal data ($1/V$ as a function of $1/[ADP]$) of pyruvate kinase native structure

The inverse of the Y-intercept represents the V_{max} which has a value of $0.0046 \mu\text{mol}/\text{min}$ according to this data. The inverse of the X-axis represents the K_M value of 0.23 mM ADP . This value is low in comparison to other studies performed at optimum conditions of 37° Celsius and pH of 7.4 where a K_M value of $0.37 \pm 0.02 \text{ mM}$ was observed¹⁷.

Refolded Activity Measurements

The results of the refolded enzymes were compared to the results of the native enzymes of experiments performed on the same day with pyruvate kinase taken from the same protein sample and identical solution concentrations. The Lineweaver Burk plot, Figure 14, was created in the same manner

as described above. The double reciprocal calculated V_{\max} for the refolded protein is 0.00268 $\mu\text{mol}/\text{minute}$ and the K_M is 0.167 mM ADP.

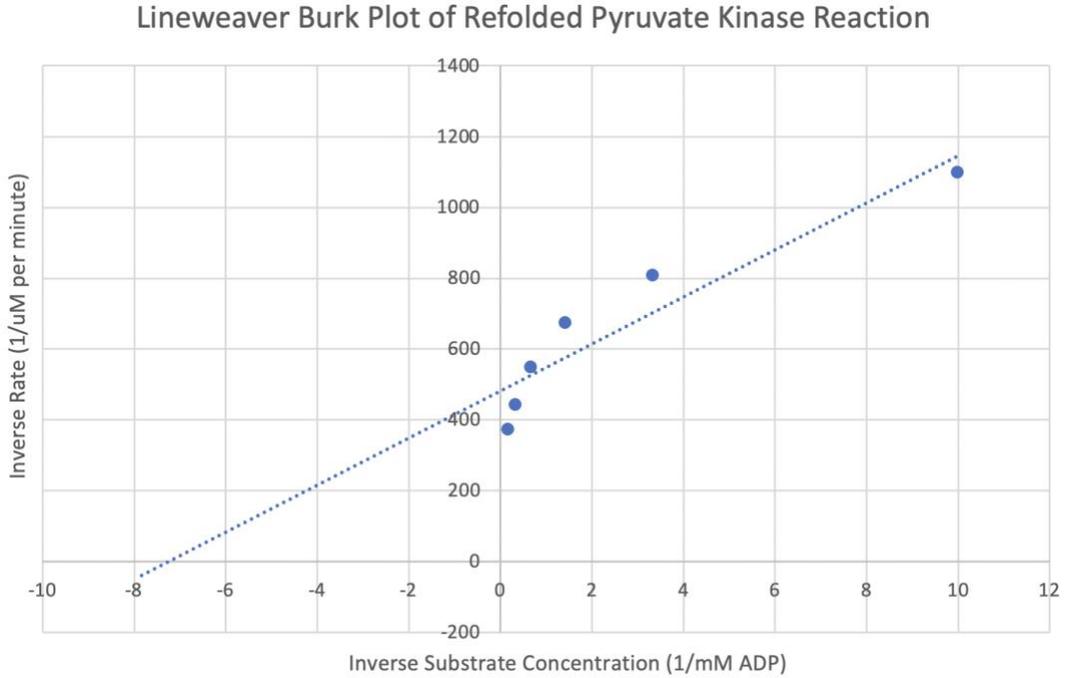
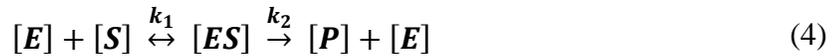


Figure 14 Lineweaver Burk plot showing linear trends in double reciprocal data ($1/V$ as a function of $1/[ADP]$) of pyruvate kinase refolded structure

The initial data from both the refolded pyruvate kinase and native state pyruvate kinase specific activity assays were plotted side by side and fitted with an exponential function calculated by the Michaelis-Menten model of enzyme kinetics¹⁹ derived in equations (3-13).

$$[E_{total}] = [E] + [ES] \quad (3)$$



$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] \quad (5)$$

$$k_1[E][S] = (k_{-1} + k_2)[ES] \quad (6)$$

$$K_m = \frac{k_{-1} + k_2}{k_1} = \frac{[S][E]}{[ES]} = \frac{[S][E_{total} - ES]}{[ES]} \quad (7)$$

$$K_m = \frac{[S][E_{total}]}{[ES]} - [S] \quad (8)$$

$$[ES] = \frac{[S][E_{total}]}{K_m + [S]} \quad (9)$$

$$V = k_2[ES] \quad (10)$$

$$V = k_2 \frac{[S][E_{total}]}{K_m + [S]} \quad (11)$$

$$V_{max} = k_2[E_{total}] \quad (12)$$

$$V = \frac{V_{max}[S]}{K_m + [S]} \quad (13)$$

When the reaction rate defined in equation (13) is plotted with the experimental data using the maximum rate of NADH depletion in $\mu\text{mol}/\text{min}$ versus the substrate concentration of ADP in μM , as seen in Figure 15, the constants V_{max} and K_M can be solved for. The maximum velocities were plotted using 95% confidence intervals

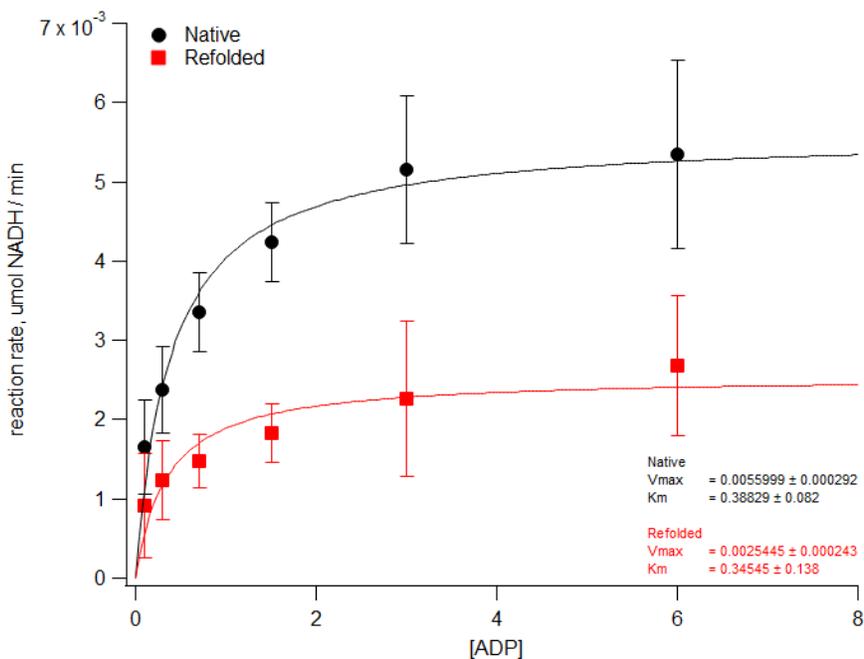


Figure 15 – Reaction curves of both native and refolded PK, using 95% CI for the reaction rates, graphed with the assistance of Ian Sitarik of the O'Brien Research Group

The V_{\max} calculated using the Michaelis-Menten model were 0.0055999 ± 0.000292 $\mu\text{mol}/\text{min}$ for the native state pyruvate kinase and 0.0025445 ± 0.000243 $\mu\text{mol}/\text{min}$ for the refolded pyruvate kinase. The K_M values calculated using the Michaelis-Menten model were 0.38829 ± 0.082 μM ADP for the native state pyruvate kinase and 0.34545 ± 0.138 μM ADP for the refolded state. The Michaelis-Menten model is considered to be a better tool for estimating enzyme kinetics than the Lineweaver Burk plot, so the Michaelis-Menten data was used moving forward. The average reaction rate across substrate concentrations is compiled with the final V_{\max} and K_M data in table 2.

Table 2 – Pyruvate Kinase Data: velocity per substrate concentration, V_{\max} and K_M for both native and refolded enzyme samples

Reaction Rate Velocities of Pyruvate Kinase (in micromoles per minute)								
	ADP Concentrations						V_{\max}	K_M
	0.1 μM	0.3 μM	0.7 μM	1.5 μM	3.0 μM	6.0 μM		
Native PK	0.001646	0.002377	0.003359	0.004231	0.005157	0.005343	0.0055999	0.38829 μM
Refolded PK	0.000910	0.001238	0.001484	0.001823	0.002258	0.002687	0.0025445	0.34545 μM

The decrease in V_{\max} without a significant decrease in K_M could be interpreted two different ways, the first being that the refolding process did not alter the enzymes' ability to bind to the substrate but did alter the enzymes' ability to catalyze the reaction. The second possibility is that the enzymes impacted by the degradation process were not able to properly refold and instead aggregated and are no longer functional. The active protein concentrations were not measured after the degradation process, so it cannot be determined whether the functionality of a subpopulation was decreased or if the total population of all functional proteins was decreased but it was most likely a mix of both factors.

Chapter 5

Conclusion

In each of the three proteins examined in this thesis, D-alanine D-alanine ligase B, hexokinase, and pyruvate kinase, there is evidence of a subpopulation of proteins possessing a different structural distribution and overall enzymatic activity than that of the native population. Each of these enzymes were identified by the quantum mechanics and molecular mechanics analysis of the coarse grain simulation of protein synthesis as having long lived kinetic traps that alter functionality. It was predicted that synonymous mutations led to changes in co-translational folding rates that can cause near-native-like protein entanglements which can cause a decrease in enzyme activity but remains similar enough to the native structure that it bypasses the cell's proteostasis mechanisms.

In the protein D-alanine D-alanine ligase B, the computational studies indicated that a non-covalent lasso entanglement that both altered function and locked the structure in a kinetic trap could be caused by changing the co-translational folding rate via a synonymous mutation. The computational results were tested experimentally by expressing a fast co-translational folding enzyme, a slow co-translational folding enzyme, and the native enzyme. The results show that 30% to 40% of the proteins expressed by the synonymous mutations differed structurally from the native state. Between the fast and slow mutations, it was seen that the slow mutants have only 88% of the enzymatic activity seen in the fast mutants. These results prove that for DDLB, synonymous mutations are correlated with subpopulations of proteins with structural and functional differences.

After the success with DDLB, the same procedure was applied to a more prominent, but more complex enzyme: Hexokinase. The synonymous mutations of hexokinase were selected,

and the protein was produced. The specific activity was measured through a coupled assay, however inconsistent results led to the incorporation of another technique, chemotaxis. The differences in migration of hexokinase enzymes across a substrate-enzyme concentration gradient allowed for samples of high activity enzymes to be separated from the low activity enzymes. The samples that moved less due to less enzyme substrate interaction contain higher concentrations of proteins with decreased functionality which can be attributed to non-native structural distributions.

The final protein analyzed, pyruvate kinase, demonstrated how an enzymes activity changes after being denatured and refolded. The reaction rates of the native state and the refolded proteins, measured under identical conditions, show that refolding the enzyme decreases its maximum reaction rate by 50%. While it cannot be proven from these experiments alone what causes this decrease, it can be inferred that a significant population of the protein sample was unable to return to its native structure due to some form of misfolding.

This thesis combines the results of DDLB, hexokinase, and pyruvate kinase, to allude to the conclusion that protein entanglements lead to decreased function. By manipulating the speed at which a polypeptide chain folds, whether it be through synonymous mutations or a forced denaturizing and reformation process, subpopulations of proteins with altered function and structure may be created and exist in the cell for extended periods of time.

BIBLIOGRAPHY

- (1) Voet, D.; Voet, J.; Pratt, C. *Fundamentals of Biochemistry: Life at the Molecular Level*, 4th ed.; Wiley: Hoboken, NJ, 2013.
- (2) Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P. Universal Protein Misfolding Intermediates Can Bypass the Proteostasis Network and Remain Soluble and Less Functional. *Nat Commun* **2022**, *13* (1), 3081. <https://doi.org/10.1038/s41467-022-30548-5>.
- (3) Stephenson, J. Nobel Prize to Stanley Prusiner for Prion Discovery. *JAMA: The Journal of the American Medical Association* **1997**, *278* (18), 1479. <https://doi.org/10.1001/jama.1997.03550180027011>.
- (4) Mayo Foundation For Medical Education and Research. *How Prions Fold*. The Mayo Clinic.
- (5) Voet, D.; Voet, J.; Pratt, C. *Fundamentals of Biochemistry: Life at the Molecular Level*, 4th ed.; Wiley: Hoboken, NJ, 2013.
- (6) Jiang, Y.; Neti, S.; Sitarik, I.; Pradhan, P.; To, P.; Xia, Y.; Fried, S.; Booker, S.; O'Brien, E. How Synonymous Mutations Alter Enzyme Structure and Function over Long Time Scales. *Nat Chem preprint*.
- (7) Zhou, M.; Guo, J.; Cha, J.; Chae, M.; Chen, S.; Barral, J. M.; Sachs, M. S.; Liu, Y. Non-Optimal Codon Usage Affects Expression, Structure and Function of Clock Protein FRQ. *Nature* **2013**, *495* (7439), 111–115. <https://doi.org/10.1038/nature11833>.
- (8) Dvorak, P.; Chrast, L.; Nikel, P. I.; Fedr, R.; Soucek, K.; Sedlackova, M.; Chaloupkova, R.; de Lorenzo, V.; Prokop, Z.; Damborsky, J. Exacerbation of Substrate Toxicity by IPTG in *Escherichia Coli* BL21(DE3) Carrying a Synthetic Metabolic Pathway. *Microb Cell Fact* **2015**, *14* (1), 201. <https://doi.org/10.1186/s12934-015-0393-3>.

- (9) Beutler, E. Disorders Due to Enzyme Defects in the Red Blood Cell; 1972; pp 131–160.
<https://doi.org/10.1016/B978-0-12-027306-5.50010-3>.
- (10) Kuser, P.; Cupri, F.; Bleicher, L.; Polikarpov, I. *3B8A - Crystal structure of yeast hexokinase PI in complex with glucose*. RCSB Protein Data Base.
- (11) Rabbani, N.; Xue, M.; Thornalley, P. J. Hexokinase-2-Linked Glycolytic Overload and Unscheduled Glycolysis—Driver of Insulin Resistance and Development of Vascular Complications of Diabetes. *Int J Mol Sci* **2022**, *23* (4), 2165.
<https://doi.org/10.3390/ijms23042165>.
- (12) Hetz, C. The Unfolded Protein Response: Controlling Cell Fate Decisions under ER Stress and Beyond. *Nat Rev Mol Cell Biol* **2012**, *13* (2), 89–102. <https://doi.org/10.1038/nrm3270>.
- (13) Bergmeyer, H. U. [New Values for the Molar Extinction Coefficients of NADH and NADPH for the Use in Routine Laboratories (Author's Transl)]. *Z Klin Chem Klin Biochem* **1975**, *13* (11), 507–508.
- (14) Johnson, L. N. Structural Basis for Substrate Recognition and Control in Protein Kinases. In *Data Mining in Structural Biology*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2001; pp 47–69.
https://doi.org/10.1007/978-3-662-04645-6_3.
- (15) Morgan, H. P.; O'Reilly F.; Palmer, R.; McNae, I. W.; Wear, M. A.; Fothergill-Gilmore, L. A.; Walkinshaw, M. D. *Human M1 pyruvate Kinase: 3SRF*. RCSB Protein Data Base.
- (16) Haferkamp, P.; Tjaden, B.; Shen, L.; Bräsen, C.; Kouril, T.; Siebers, B. The Carbon Switch at the Level of Pyruvate and Phosphoenolpyruvate in *Sulfolobus Solfataricus* P2. *Front Microbiol* **2019**, *10*. <https://doi.org/10.3389/fmicb.2019.00757>.
- (17) Boehme, C.; Bieber, F.; Linnemann, J.; Breitling, R.; Lorkowski, S.; Reissmann, S. Chemical and Enzymatic Characterization of Recombinant Rabbit Muscle Pyruvate Kinase. *bchm* **2013**, *394* (5), 695–701. <https://doi.org/10.1515/hsz-2012-0334>.

- (18) Bendt, A. K.; Burkovski, A.; Schaffer, S.; Bott, M.; Farwick, M.; Hermann, T. Towards a Phosphoproteome Map of *Corynebacterium Glutamicum*. *Proteomics* **2003**, 3 (8), 1637–1646. <https://doi.org/10.1002/pmic.200300494>.
- (19) Roskoski, R. Michaelis-Menten Kinetics☆. In *Reference Module in Biomedical Sciences*; Elsevier, 2015. <https://doi.org/10.1016/B978-0-12-801238-3.05143-6>.

ACADEMIC VITA

Taylor Rose Berek

EDUCATION	Bachelor of Science in Chemistry , Schreyer Honors College <i>The Pennsylvania State University</i> , University Park, PA Anticipated Graduation: May 2023 Relevant Coursework Organic Chemistry I & II Physical Chemistry: Thermodynamics & Quantum Chemistry Bioengineering: Signals and Systems Advanced Analytical Chemistry: Chemical Spectroscopy
WORK EXPERIENCE	Research Assistant in the O'Brien Research Group , Penn State Chemistry Department Fall 2020-Current (15 hours per week while in school, full-time during breaks) <ul style="list-style-type: none">Expressed proteins through an <i>E. coli</i> cell culture grown in a 5L BioreactorGenetically engineered mRNA plasmids to produce non-native entangled proteinsPurified the protein using affinity and size exclusion columns through the FPLCAnalyzed the expression with Bradford assays and SDS PAGE gel electrophoresisMeasured the specific activity of Hexokinase, Urease, DDLB, DFHR, and Pyruvate kinase Intern for the Hypoglycemia and Adverse Pregnancy Outcomes Study Northwestern Medical Research Center, Chicago, IL June-Aug 2018 <ul style="list-style-type: none">Organized the genomic sequencing lab results of thousands of DNA samplesFormulated coding in R to be used for statistical analysis of dataCompared statistical trends in the genetic data and patient health, using confidential medical IDs
VOLUNTEER EXPERIENCE	Amigos de las Americas Volunteer La Casa de la Juventud, Perez Zeledón, Costa Rica July-Aug 2017 <ul style="list-style-type: none">Lived with a host family in a rural, secluded communityFacilitated local youth in executing environmental sustainability projects, including clearing fields, building water canals, and paving walkways
SKILLS	Programming Languages Python, R, MATLAB, HTML, CSS Laboratory Training/Experience Bacterial culture growths, Bioreactor (5L), FPLC, Manual and Spin Column purification, Microfluidizer, Sonification, HDX analysis, Cryo-EM imaging, UV-vis, Activity measurements, Chemical extractions, Microfluidic devices, Metal Nanoparticle production and oxidation
LEADERSHIP	Executive Board Member, Vice President, Society of Industrial Biotechnology 2020-2021 Executive Board Member, Secretary, Biomedical Engineering Society 2020-2021 Executive Board Member, Philanthropy and Community Service, Alpha Omicron Pi 2020-2021
OTHER INVOLVEMENT	Berek's Book Store, self-run Amazon storefront reselling used books, 2021-current Learning Assistant, Intro to Organic Chemistry (CHEM 202) for Dr. Kyle Schmid, 2020-2021 Schreyer Honors College Student Counsel 2019-2021 Society of Women Engineers 2019-2021 <i>Rince Na Leon</i> (Irish Dance Club Team) 2019-2021
HONORS	Schreyer Scholar, Honors College of Penn State University, 2019-2023 Penn State Provost Award, 2019-2023 Erickson Discovery Grant, undergraduate research, 2021 Alfred W. Pond and North Chemistry Award, Eberly College of Science, 2021-2022 Center of Excellence in Industrial Biotechnology, Innovation Award, 2021-2022 Academic Excellence Award, 2021-2023