UNIVERSITY OF CALIFORNIA

Santa Barbara

Hydroxyacid Oxoacid Transhydrogenase and Succinate Semialdehyde Dehydrogenase: Enzymes in the Oxidation of Gamma-Hydroxybutyrate to Succinate

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Chemistry

by

István E. Szabó

Committee in charge:

Professor Brandon Greene, Co-Chair

Professor Stanley M. Parsons, Co-Chair

Professor John Lew

Professor Kevin W. Plaxco

December 2024

The dissertation of István E. Szabó is approved.

John Lew, Member

Kevin W. Plaxco, Member

Stanley M. Parsons, Committee Co-Chair

Brandon Greene, Committee Co-Chair

December 2024

Hydroxyacid Oxoacid Transhydrogenase and Succinate Semialdehyde Dehydrogenase:

Enzymes in the Oxidation of Gamma-Hydroxybutyrate to Succinate

by

István E. Szabó

Acknowledgments

I would like to begin by thanking my Advisor, Professor Stanley M. Parsons, for the extraordinary opportunity to develop a new research project in his laboratory. Prof. Parsons, your faith in my abilities never wavered and I take it as a genuine compliment that you would trust me with such an ambitious project. I very much appreciate the patience shown as we had to rethink approaches, or when we made the difficult choice of abandoning them in favor of new strategies. I am also grateful for your optimism, which was there for me when mine was running low. I especially appreciate the generous gift you personally donated to the project; without this kind gesture I would not have been able to complete the research.

In addition, many thanks to my committee for all their support and kind words over the years. Professor John Lew, though our interactions were limited, your thoughtful comments always left me feeling more confident. During our last meeting your compliments on my kinetics data were very much appreciated, because I struggled to make them perfect under challenging circumstances and was very pleased when you acknowledged this effort. To Professor Kevin W. Plaxco—yes, I know you've always insisted that I call you Kevin—here I must insist on the formality as a sign of the immense respect I have for you, and to thank you for respect you've shown me during all stages of my academic development. Kevin, I always knew I could count on your support, even if I was often too stubborn to recognize when it was needed. Some of my fondest grad school memories will be drinking Margaritas and grading astrobiology exams at Pepe's in Goleta. To Brandon Greene, my co-advisor, thank you from the bottom of my heart for taking time away from the 9 million other things on your plate in order to help me get across the finish line. Although the final months of my research were intense and filled with many last-minute developments, I always valued your opinion even while insisting on my own. Your input brought a needed change to my research style and helped to make me a more efficient scientist when I needed it the most.

Thank you, Professor Yang Hai, for sharing bench space in your lab while I wrapped up my research project. As one of the few unexpected surprises you received after accepting the faculty position, I hope I was among the more pleasant ones. Thank you for your patience, for sharing your knowledge with me, and for our conversations on many other interesting and entertaining subjects.

A very special thank you to Professor Kalju Kahn. You have been a lot of things to me over the years: a lab mate, a teacher, a colleague, but most of all a true friend. I count seven courses for which you were my instructor; no other instructor has ever prepared me as much, and as thoroughly. Much of what I know I've learned from you. Thank you for lending me equipment and materials when I needed them. Most importantly, thank you for giving me feedback on my data, helping me with troubleshooting issues, and make suggestions that greatly improved my results. You are a valuable resource that I came to depend on, and the department is lucky to have you.

To Professor Thomas C. Bruice, whose impact on my life was nothing short of transformative. If I have only one regret, it's that I cannot share this dissertation with you today. You are a legend, and I'm deeply grateful for the opportunity to have worked and published with you. I miss our afternoon walks to the Coral Tree Café for coffee and will treasure the many stories you shared with me—and I hope you don't mind if I've shared a few of them with others since then. Thank you for teaching and advising me, and I will always cherish our friendship for the rest of my life.

Finally, I want to dedicate this work to my family. No one has shaped me into the person I am today more than my parents, Steven and Olivia. I draw my strength and self-reliance from the things you taught me. To my sister Renee, my first advisor, who first inspired my love of science and somehow explained special relativity to me before I could even manage my multiplication tables. Thank you for staying up late and answering my endless questions on the universe. I was probably the only kid on the playground to know that an *erg* is a unit of energy. To my dog Teva, thank you for sticking around for 16 years, seeing me though ups and downs, and not leaving my side until you were sure I'd be okay. To our new kitty, Lhotse, thanks for staying up all night and keeping me company as I completed this manuscript. Lastly, no one has done more to help me though graduate school than my beautiful wife, Mimi. M, you are my best friend, partner in crime and the love of my life. For a *small fry*, you truly possess the strength of Hercules. Whether it's effortlessly belaying me on a 5.10 outdoor climb at the Playground, picking me up from lab at 2 a.m., or making sure I don't leave home without breakfast—you never let me down and kept me alive through challenging times. M, I seriously could not have done this without you. To my wife and family, I dedicate this work to you.

Curriculum Vitae

Education

- Doctor of Philosophy in Chemistry, 2024
  Department of Chemistry & Biochemistry, University of California, Santa Barbara
- Bachelor of Science in Biochemistry, 2008
  Department of Chemistry & Biochemistry, University of California, Santa Barbara

Professional Employment and Research Experience

- **Graduate Research Assistant, 2012 – 2022**
  Stanley M. Parsons Research Group
  Department of Chemistry & Biochemistry, University of California, Santa Barbara
  Explored methods for the expression, purification, and characterization of eukaryotic and prokaryotic enzymes in the metabolic pathway for GHB oxidation. Recombinant DNA sequences were cloned into expression vectors and transfected (or transformed) into mammalian (or bacterial) cell lines for protein expression. Protein purification was accomplished with a variety of chromatographic techniques, depending on the biophysical properties of the enzyme and nature of affinity tags. Characterization of enzymes included techniques such electrophoresis (Western blots), tryptic digest mass spectrophotometry, and determination of kinetic parameters and catalytic mechanisms using enzyme kinetics.
- **Graduate Teaching Assistant, 2012 – 2020**
  Department of Chemistry & Biochemistry, University of California, Santa Barbara
  Teaching assistant for various upper and lower division lecture and laboratory courses, from general chemistry to advanced biochemistry. Instruction on laboratory techniques included advanced spectroscopy (NMR, circular dichroism, mass spec, ultra-violet and fluorescence), interpretation of data using multivariate analysis and nonlinear regression software, computer-based bioinformatics database searches, and computational modelling.

- **Undergraduate Research Assistant and Technician, 1998 – 2008**
  Thomas C. Bruice Research Group
  Department of Chemistry & Biochemistry, University of California, Santa Barbara
  Assisted in the development and synthesis of cationic oligonucleotides (DNA and RNA analogs) for antigene/antisense research. Multistep organic synthesis, purification and characterization of monomers was followed with coupling chemistry to form short oligonucleotides containing positively charged guanidinium-linkages (in place of naturally occurring and negatively charged phosphodiester-linkages). Equilibrium melting of the analogs and their target DNA (or RNA) sequences provided thermodynamic data required to measure the binding affinity.

Publications

- Jain, M.L.; Bruice, P.Y; **Szabo, I.E.**; Bruice, T.C., Incorporation of Positively-Charged Linkages into DNA and RNA Backbones: A Novel Strategy for Antigene and Antisense Agents. Chemical Reviews, **2012**, *112* (3), p. 1284-1309.
- **Szabo, I.E.** and Bruice, T.C., DNG cytidine: synthesis and binding properties of octameric guanidinium-linked deoxycytidine oligomer. Bioorganic & Medicinal Chemistry, **2004**, *12* (15): p. 4233-4244.
- Kojima, N.; **Szabo, I.E.**; Bruice, T.C., Synthesis of ribonucleic guanidine: replacement of the negative phosphodiester linkages of RNA with positive guanidinium linkages. Tetrahedron, **2002**, *58* (5): p. 867-879.
- Linkletter, B.A.; **Szabo, I.E.**; Bruice, T.C., Solid-phase synthesis of oligopurine deoxynucleic guanidine (DNG) and analysis of binding with DNA oligomers. Nucleic Acids Research, **2001**, *29* (11): p. 2370-2376.
- Linkletter, B.A., **Szabo, I.E.**, Bruice, T.C., Solid-phase synthesis of deoxynucleic guanidine (DNG) oligomers and melting point and circular dichroism analysis of binding fidelity of octameric thymidyl oligomers with DNA oligomers. Journal of the American Chemical Society, **1999**, *121* (16): p. 3888-3896.

Manuscripts in Preparation

- **Szabo, I.E.**, Greene, B., Parsons, S.M., Expression of Human Hydroxyacid-Oxoacid Transhydrogenase in *E. coli*, (*communication*)
- **Szabo, I.E.**, Kahn, K., Greene, B., Parsons, S.M., The Steady State Kinetics and pH-Rate Profile of $NADP^+$-dependent Succinate Semialdehyde Dehydrogenase from *E. coli*, (*paper*)

Teaching

- CHEM 1AL-BL-CL, General Chemistry Laboratory
- CHEM 110L, Introductory Biochemistry Laboratory
- CHEM 112L, Biophysical and Bioanalytical Laboratory
- CHEM 125L, Laboratory Techniques in Biochemistry
- CHEM 142A-B, Biochemistry (Lecture)
- CHEM 146, Membrane Biochemistry (Lecture)

Awards

Dean's Honors, Winter 2008

University of California, Santa Barbara

Abstract

Hydroxyacid Oxoacid Transhydrogenase and Succinate Semialdehyde:

Enzymes in the Oxidation of γ-Hydroxybutyrate to Succinate

by

István E. Szabó

Hydroxyacid oxoacid transhydrogenase (HOT) and succinate semialdehyde dehydrogenase (SSADH) are a pair of oxidoreductases that operate along pathways adjacent to the Szent–Györgyi–Krebs cycle and together catalyze the oxidation of γ-hydroxybutyrate (GHB) to succinate. GHB was initially synthesized to study the neurotransmitter GABA but has become infamous as a date rape drug. Succinate is a well-known Krebs Cycle intermediate.

HOT is an iron-dependent group III alcohol dehydrogenase (ADH) and the only member of this little-known, microbial-associated group found in animals. Despite its discovery in 1988, few studies on HOT have been published and the enzyme remains poorly characterized. HOT is proposed to oxidize GHB to succinate semialdehyde with concomitant reduction of a tightly bound NAD^+ *cofactor* that does not exchange with the solvent NAD^+ pool. The NAD^+ cofactor is regenerated in the active site by reduction of an α-ketoglutarate co-substrate to D-2-hydroxyglutarate (D2HG). Since D2HG has been identified as an *oncometabolite* involved in tumor progression, and changes in HOT expression levels have been detected in some cancers, HOT appears to be a potential drug target worthy of continued study.

In published research on HOT, animal organs have been primary source of the enzyme, with cultured animal cells being secondary. However, structural (X-ray) studies demand a relatively good amount of pure enzyme, while functional (enzyme kinetics) studies would benefit from the ability to mutate active site residues. Recombinant HOT containing an affinity

purification tag, transformed into *E. coli* or stably transfected into a mammalian cell line, would ideally produce large amounts of pure and mutable HOT for structural and functional studies. This manuscript describes the heterologous expression of human HOT in *E. coli*, resulting in: 1) soluble and nearly pure recombinant HOT containing an N-terminal maltose binding protein; and 2) soluble and partially purified recombinant HOT with an N-terminal His-tag. Also described is the stable transfection of PC-12 cells (rat pheochromocytoma) with a vector encoding recombinant HOT with a C-terminal HA-tag. Integration of this sequence into the cell's genome resulted in a line of cells that constitutively expressed this construct and, upon lysis and chromatography, yielded moderately pure recombinant HOT.

Lastly, because the HOT-catalyzed reaction produces no measurable change in absorbance or fluorescence under steady state conditions, the isolated reaction cannot be conveniently followed by UV-Vis or fluorescence spectroscopy. However, the HOT reaction may be coupled to the SSADH reaction since the succinate semialdehyde produced by HOT is the substrate for SSADH and the SSADH reaction produces NAD(P)H. Because NAD(P)H has a maximal absorbance at 340 nm, or fluorescence emission at 460 nm, the coupled reaction can be conveniently followed using a UV-Vis spectrophotometer or a spectrofluorometer. *E. coli* SSADH was chosen as the auxiliary enzyme for the coupled assay, though the enzyme kinetics of this SSADH had yet to be thoroughly investigated. Thus, this dissertation also describes the expression, purification and steady-state enzyme kinetics for recombinant *E. coli* SSADH containing an N-terminal His-tag. This effort resulted in the determination of kinetic constants and pH-rate profile for the SSADH-catalyzed oxidation of succinate semialdehyde to succinate.

Table of Contents

List of Equations

List of Figures

List of Tables

Chapter 1

The Metabolism of Gamma-Aminobutyrate and Gamma-Hydroxybutyrate

Abstract

Gamma-aminobutyrate (GABA) is a non-proteinogenic amino acid and the major inhibitory neurotransmitter in the mature mammalian brain, where it binds to $GABA_A$ and $GABA_B$ receptors, which results in the hyperpolarization of post-synaptic membranes and a decrease in neuronal excitability. Gamma-hydroxybutyrate (GHB) is a naturally occurring compound found in the brain, as well as a drug of abuse, which binds to the GHB and $GABA_B$ receptors, also resulting in central nervous system depression.

GABA is produced and maintained by the GABA shunt, a metabolic pathway that branches off the Krebs cycle and then returns to it after a series of four enzymatic reactions. The final reaction that completes the GABA shunt is catalyzed by succinate semialdehyde dehydrogenase (SSADH), a member of a ubiquitous superfamily of enzymes found throughout nature. SSADH oxidizes succinate semialdehyde (SSA) to succinate, which re-enters the Krebs cycle. Although SSA is produced in the GABA shunt, from the transamination of GABA, it can also be produced from the oxidation of GHB. In the mitochondrial matrix, GHB is oxidized to SSA by hydroxyacid-oxoacid transhydrogenase (HOT), a NAD^+-independent alcohol dehydrogenase encoded by the gene ADHFe1.

The catalytic activity of HOT is dependent on α-ketoglutarate, which is reduced to 2-D-hydroxyglutarate (D2HG) during the oxidation of GHB. In recent years, D2HG has come to be recognized as an oncometabolite that promotes carcinogenesis through various

mechanisms. Given that D2HG is a product of GHB oxidation by HOT, or possibly even a substrate for HOT in the reverse reaction, there may be a significant connection between cancers and the enzymatic activity of HOT. Recent studies have found that the ADHFe1 gene is hypermethylated and HOT expression is downregulated in various cancer lines, while a high level of HOT expression has been corelated to patient survival in breast, colon and gastric cancers.

Very little is known about the HOT enzyme mechanism and no experimental structure of HOT exists. Therefore, further studies may prove a valuable activity in light of this enzyme's connection to cancer. SSADH may help in this endeavor, since the HOT reaction does not produce an observable signal for continuous enzyme assays. However, because the two enzymes are connected though the common substrate SSA, SSADH may serve as a reporter enzyme in a coupled-enzyme assay for the study of HOT catalytic activity.

1.1 Introduction

The focus of this dissertation is on the recombinant expression and characterization of two enzymes responsible for the oxidation of gamma-hydroxybutyrate (GHB) to succinate. These two enzymes are: 1) hydroxyacid oxoacid transhydrogenase (HOT); and 2) succinate semialdehyde dehydrogenase (SSADH) (**Figure 1.1**, both in **green**).

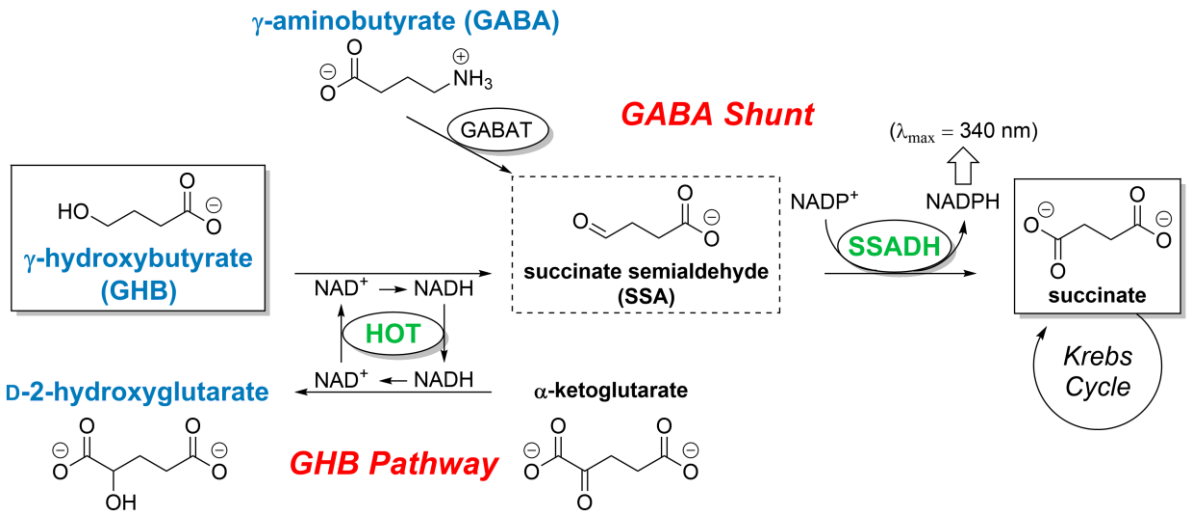

Figure 1.1. The oxidation of GHB to succinate by the enzymes HOT and SSADH. Important metabolic substrates are labeled in **blue**, key metabolic pathways are labeled in **red**, and the enzymes catalyzing reactions are encircled with ovals (HOT and SSADH are labeled in **green**).

HOT oxidizes GHB to succinate semialdehyde (SSA) using a tightly bound NAD^+ *cofactor* that does not exchange with the bulk solvent. Instead, the NADH formed during the reaction must be oxidized back to NAD^+ *in situ*. This step is accomplished through the use of a co-substrate, alpha-ketoglutarate, which is reduced to D-2-hydroxyglutarate in the active site, regenerating the NAD^+ cofactor. Because of the redox cycling of NAD^+/NADH in the active site, there is no net increase in NADH concentration that can be monitored at 340 nm, which is the λ_{max} for NADH and preferred method for following the progress of enzymatic reactions that employ $NAD(P)^+$ coenzymes. In addition, none of the substrates or products involved in the reaction have a UV-Vis absorption, so there is no convenient means of *directly* following

HOT catalysis with a spectrophotometer. However, the SSA produced by HOT is a substrate for SSADH, which oxidizes SSA to succinate with the reduction of $NADP^+$ to NADPH, and the $NADP^+$/NADPH of this enzyme does exchange with the bulk solvent. As a result, the reaction catalyzed by HOT can be followed by observing the increase in absorbance at 340 nm produced by the activity of the second enzyme, SSADH, which serves as a reporter (auxiliary) enzyme for HOT and forms the basis of the coupled enzyme assay depicted in **Figure 1.1**.

In order to fully appreciate the important roles that HOT and SSADH play in biochemistry, *Chapter 1* will: i) introduce three substrates of importance (**Figure 1.1**, in **blue**); ii) describe two vital metabolic pathways these enzymes participate in (**Figure 1.1**, in **red**); and iii) explain how the three substrates and key enzymes in the pathways play critical roles in several serious medical conditions.

Chapter 2 describes the present research on recombinant *E. coli* succinate semialdehyde dehydrogenase (SSADH), an enzyme that oxidizes succinate semialdehyde to succinate in **Figure 1.1** (SSADH labeled in **green**). Although SSADH is the second enzyme in the coupled reaction, it was necessary to express, purify and characterize SSADH before it could be used as an auxiliary enzyme to detect HOT activity in a coupled enzyme assay. Therefore, in keeping with the chronological order of the research, SSADH is presented first. A detailed examination of the steady-state enzyme kinetics is presented, including values for the enzyme kinetics parameters, a pH-rate profile including a rate-dependent active-site pK_a, as well as a possible enzyme mechanism for the reaction.

Chapter 3 and *4* document efforts to express and purify recombinant hydroxyacid oxoacid transhydrogenase (HOT) in *E. coli* and mammalian cells. HOT catalyzes the oxidation of GHB to succinate semialdehyde in the pathway depicted in **Figure 1.1** (HOT colored **green**). A

number of strategies are described, including the use of various vectors, different expression constructs, and a variety of both prokaryotic (*Chapter 3*) and eukaryotic (Chapter 4) lines. Recombinant *mouse* HOT was ultimately expressed in PC-12 (rat pheochromocytoma) cells and purified though a combination of anion-exchange and size exclusion chromatography. Recombinant *human* HOT was expressed in soluble form in a variety of *E. coli* cell lines, using solubility-enhancing tags and co-expressed molecular chaperones, and then purified using affinity and size-exclusion chromatography.

As a final note, the naming of substrates, products and their associated enzymes will indicate their probable ionization state within a physiologically relevant context, defined here as between pH 7.0 and 8.0. The upper value reflects the mildly alkaline pH of the mitochondrial matrix where many of these compounds are found.[1] For example, GABA stands for **g**amma-**a**mino**b**utyric **a**cid, but at physiological pH the carboxylic acid group is deprotonated (pK_a-COOH = 4.03 and pK_a-NH_3 = 10.56). The abbreviation of GABA will still be used for historical reasons but referred to as gamma-aminobutyrate and its structure will be drawn as the expected zwitterion at pH 7–8 (see **Figure 1.2**). When labeling chemical structures in figures, Greek letters may be used to save space (e.g., γ-aminobutyrate). The same rules will apply for the names of most enzymes, such as GABAT, which may be referred to as gamma-aminobutyrate transaminase, or γ-aminobutyrate transaminase. The notable exception will be the use of the name hydroxyacid oxoacid transhydrogenase (HOT), which has a long-standing naming precedent in scientific literature that will be respected here.

1.2 Important Metabolic Substrates

Three substrates (**Figure 1.1** in **blue**; and **Figure 1.2**) warrant a brief introduction because they play a significant role in several widely recognized medical disorders and are either direct substrates for HOT, or in the case of GABA, serve as an intermediate in metabolic pathways influenced by the activities of HOT and SSADH.

1.2.1 Gamma-Aminobutyrate

Gamma-aminobutyrate (GABA), or γ-aminobutyrate (**Figure 1.2**), is a *non-proteinogenic amino acid* and the major inhibitory neurotransmitter in the mature mammalian brain, where it mainly acts to hyperpolarize post-synaptic membranes and decrease neuronal excitability.[2] The name derives from location of the amino functional group (H_3N^+–), which is attached to the C-4 carbon, also called the gamma-carbon (or γ-carbon).

Initially identified in bacteria, yeast, plants, and in some non-neural mammalian tissues,

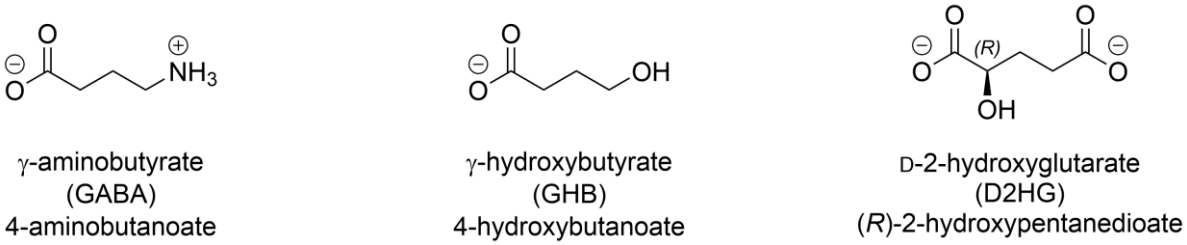

Figure 1.2. Three important metabolic substrates.

it was not until 1950 when GABA was also discovered to be a major amine in *protein-free* brain extracts.[3] In some regions of the human brain the concentrations of GABA are relatively high, in the millimolar range[4], approximately 1,000 times higher than the neurotransmitters dopamine and serotonin.[5] It would take nearly a decade before the role of GABA would be fully understood, when a compound—Factor I—known to have an inhibitory effect on crayfish neurons was confirmed to be GABA.[6]

GABA binds to the active sites of *two general classes of receptors*, $GABA_A$ and $GABA_B$ receptors. $GABA_A$ receptors are *ligand-gated ion channels* that allow the flow of chloride ions across the cell membrane.[7] Some well-known allosteric activators of $GABA_A$ receptors are benzodiazepines, barbiturates and ethanol.[8] $GABA_B$ receptors are *G-protein coupled receptors* linked to potassium channels.[9] In addition to the parent compound GABA, $GABA_B$ receptors are also activated by gamma-hydroxybutyrate (GHB) and the muscle relaxer baclofen.[10] In recent years GABA-$_B$ receptors have been the focus of efforts to treat drug addition, anxiety and pain.[9, 11-13] The metabolic pathway for GABA is discussed in a later section (see § *1.3.1. The GABA Shunt*).

As noted earlier (see § 1.1. Introduction), GABA is a zwitterion (**Figure 1.2**) at physiologically relevant pH values (7.0 – 8.0). Although evidence suggests that the blood-brain barrier (BBB) is impermeable to charged compounds such as GABA[14], and that dietary intake of GABA is unlikely to provide a significant source of this compound to the central nervous system[15] (CNS), this idea remains hotly debated today.[16] Still, the historic belief that GABA as a charged species cannot cross the BBB led to the use of exogenous butyric acid derivatives, such as gamma-hydroxybutyrate, in an attempt alter CNS levels of GABA and study its effects (see next section).

1.2.2 Gamma-Hydroxybutyrate

Gamma-hydroxybutyrate (GHB), or γ-hydroxybutyrate (**Figure 1.2**), is a naturally occurring compound found in the brain, where it acts as a CNS depressant.[17] GHB is also found in extra-neural tissues, such as kidney and muscle, where it is believed to function in metabolism.[18] The name derives from location of the hydroxy functional group (HO–), which is attached to the C-4 carbon, also called the gamma-carbon (or γ-carbon).

Although the synthesis of GHB was first reported by famed chemist Alexander Zaitsev in 1874,[19] it would be nearly a century before GHB would become the focus of intense chemical research. While research on GABA was ongoing during the 1950s, the effects of short-chain fatty acids, including butyrate, were being studied for their ability to induce narcosis when injected into animals.[20] Because a majority of the butyrate underwent β-oxidation after injection, and it was known that GABA does not readily cross the BBB, researchers in the 1960s synthesized GHB from gamma-butyrolactone (GBL) believing that a gamma-hydroxy group would prevent β-oxidation while still allowing the compound to enter the CNS to produce hypnotic effects — a result that was ultimately observed.[21] A 1963 communication in *Nature* reported GHB to be *a normal metabolite* present in the brain tissue of rat, pigeon, and humans, opening the possibility of a biosynthetic pathway for GHB, likely from GABA via a succinate semialdehyde (SSA) intermediate.[22] The discovery of these metabolic pathways for GHB are discussed in a later section (see § *1.3.2. The GHB Pathway*). Subsequent studies confirmed that exogenously administered GHB does not significantly increase cerebral GABA concentrations in sedated subjects as had been expected, suggesting that the observed hypnotic effects were due to the CNS-inhibiting properties of GHB itself.[23]

While it would be eventually determined that GHB binds the $GABA_B$ receptor selectively, albeit with low binding affinity[24] ($K_i = 0.1$ mM), early studies also provided evidence of distinct GHB binding sites in rat cortex cell membranes that were high affinity[25] ($K_d = 95$ nM) and sodium-independent (i.e., not transporters). Two decades later the suspected *GHB receptors* would be confirmed when they were cloned from, and characterized in, rats[26] and humans[27]. By the end of the 1990s, accumulated research on metabolic pathways, transporters, and receptor interactions had established GHB as an endogenous neurotransmitter present at

micromolar concentrations in all regions in the brain and in several peripheral organs.[28] Some arguments against classifying GHB as a neurotransmitter include: (1) it has not been shown that GHB is stored in presynaptic vesicles, ready for exocytosis; (2) GHB is found at higher concentrations in extra-neural tissues; (3) unprecedented for a neurotransmitter, GHB readily crosses the BBB.[17]

Initial therapeutic interest in GHB focused on its use as a hypnotic and as an adjuvant in anesthesia, with the latter falling out of favor because GHB lacks analgesic properties and can produce seizure-type EEG activity in animals.[29] Today, the only FDA-approved use of GHB is in the treatment of narcolepsy, a chronic neurological disorder involving unregulated sleep-wake cycles. In addition to acting as a sedative, GHB can also induce a state of euphoria[30], produce anabolic effects by increasing the secretion of growth hormone[31], and cause physical dependence[32] when abused. For these reasons the use of GHB is tightly regulated as both a Schedule I and Schedule III drug under the United States Controlled Substances Act.[33] To the author's knowledge, GHB is the only endogenously produced neurotransmitter that has the distinction of also being a controlled substance. Despite efforts to restrict non-medicinal applications of GHB, it has become infamous in its use of a date-rape drug, where in high doses it causes sedation and amnesia in victims.[34] In order to combat the surreptitious use of GHB, the Parsons Lab at UCSB has developed and patented a rapid and reliable enzymatic assay for GHB that can be deployed on a "dipstick" and can detect sub-pharmacological doses of GHB in alcoholic beverages.[35]

Despite the negative connotations associated with GHB and its abuse potential, a number of European studies have demonstrated that administration of exogenous GHB is effective in aiding withdrawal from opiates[36-39] and alcohol dependency[40-42]. A recent analysis by the

Cochrane Collaboration, which included data from 13 randomized controlled trials comprising 648 participants, concluded that GHB is more effective than either disulfiram or naltrexone in treating alcohol dependence.[43] One striking observation is that therapeutic doses of GHB do not cause physical dependence.[44] However, the tendency to abuse GHB still is a concern. Alcoholic patients with a history of other drug dependence tend to crave and abuse GHB, while pure alcoholics do not.[45] It remains to be seen if researchers and regulators in the United States are open to similar studies or would authorize the use of GHB in the treatment of substance abuse.

1.2.3 D-2-Hydroxyglutarate

2-hydroxyglutaric acid, also known as alpha-hydroxyglutaric (α-hydroxyglutaric), is a five-carbon dicarboxylic acid whose name derives from location of the hydroxy functional group (HO–), which is attached to the C-2 carbon, also called the alpha-carbon (or α-carbon). Because this α-carbon is a chiral center, the molecule exists as two possible enantiomers. One of these is D-2-hydroxyglutarate (D2HG), or (*R*)-2-hydroxypentandioate (**Figure 1.2**), and the other is L-2-hydroxyglutarate (L2HG), or (*S*)-2-hydroxypentanedioate. The *R*/*S* and D/L nomenclatures refer to the absolute configuration at the chiral carbon based on the Cahn–Ingold–Prelog priority rules and configuration relative to glyceraldehyde, respectively. Both enantiomers are found in a variety of organisms, from bacteria to plants and animals, but their clinical significance at high concentrations appears largely restricted to humans.[46] Because the primary focus of this and other sections is on D2HG, comments on the L-isomer (L2HG) will be kept to a minimum and will be used only when a distinction between the two is important.

The hyperaccumulation of D- and L-hydroxyglutarate is tolerated in some organisms, such as plants, without any serious effects on development.[46] In humans excess D2HG and L2HG

results in the rare and serious metabolic conditions known as D- and L-hydroxyglutaric aciduria, respectively, both of which cause severe neurological and developmental impairment at a very young age.[47] There are two types of D-hydroxyglutaric aciduria. Type I is due to a mutation in the gene for *D-2-hydroxyglutarate dehydrogenase*, which prevents this enzyme from oxidizing D2HG back to AKG.[48] Type II results from *gain-of-function* mutations in mitochondrial *isocitrate dehydrogenase isoform 2* (IDH2) that causes the enzyme to produce D2HG instead of AKG.[47] There is only one form of L-hydroxyglutaric aciduria, resulting from a mutation in the gene for *L-2-hydroxyglutarate dehydrogenase*.[49] A distinct and combined D,L-2-hydroxyglutarate aciduria form has also been identified, with the cause attributed to impaired mitochondrial citrate efflux caused by mutations in the gene encoding a *mitochondrial citrate carrier protein* (SLC25A1).[50]

In addition to D-hydroxyglutaric aciduria, increased levels of D2HG have additional clinical implications. By the late 2000s a correlation between mutations in isocitrate dehydrogenase (IDH) and gliomas (tumors of the CNS) had been well established, though the precise mechanism remained unclear.[51] A 2009 Nature paper by Dang et al. finally revealed these cancer-associated mutations resulted in a *gain-of-function* ability of IDH1 to produce 2DHG (rather than *α*-ketoglutarate), proposing that 2DHG behaved as an *"onco-metabolite"* whose accumulation may drive *oncogenesis*.[52] The following year, similar cancer-associated mutations in IDH2 were also demonstrated to result in the *neomorphic* production of D2HG.[53] Since that time, mutations in IDH1 and IDH2 have been identified in over 20 types of neoplasms, making them the most frequently mutated metabolic genes in human cancers and benign tumors.[54] Interestingly, although Type II D-2-hydroxyglutaric aciduria results from mutations to IDH2, there has been an absence of cancer diagnoses observed in those cases

studied, but the young age of the patients prevents a firm conclusion as to their long-term susceptibility towards cancer.[55] Malignant brain tumors are observed in patients with L-2-hydroxyglutaric aciduria,[56] and L-2-hydroxyglutarate has recently been demonstrated to promote tumor growth in colorectal cancer.[57]

The term *oncometabolite* has now been used to describe the apparent ability of D2HG and other metabolites to promote carcinogenesis through various mechanisms. Because D2HG is an analog of α-ketoglutarate (2-oxoglutarate), it inhibits histone lysine demethylases by acting as an inhibitor in the 2-oxoglutarate-dependent dioxygenase (2OGDD) step.[58, 59] D2HG also inhibits the ten eleven translocation (TET) family of 5-methylcytosine (5mC) hydroxylases by the same mechanism.[60] The net result of inhibiting these enzymes is the epigenetic remodeling of gene expression, resulting in metabolic reprogramming that impacts both adaptive and innate immunity and promotes tumorigenesis. [61-64]

1.3 Metabolic Pathways of GABA, GHB, and D2HG

1.3.1 The GABA Shunt

Given the critical role that the neurotransmitter GABA plays in the central nervous system (CNS), it is not surprising that an important metabolic pathway—the GABA shunt—exists to ensure and control its availability (**Figure 1.4**).

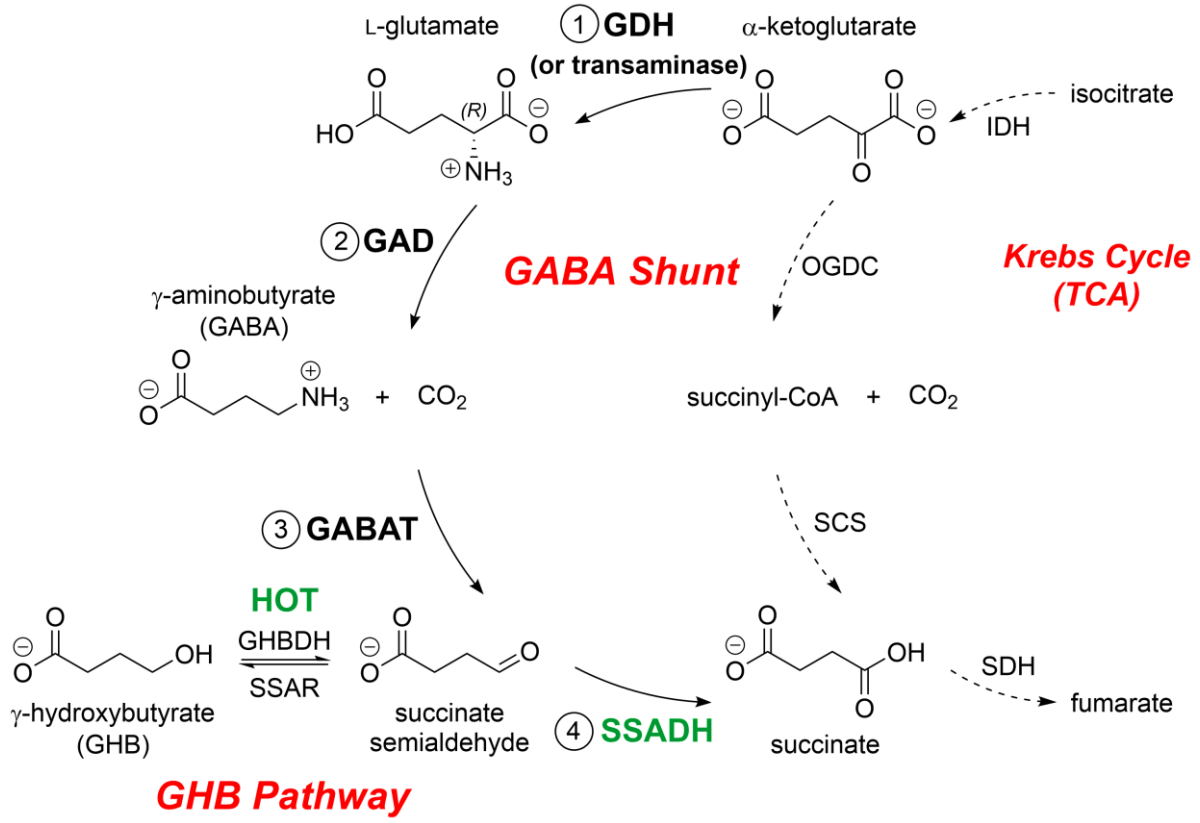

Figure 1.3. The GABA shunt and GHB pathway. (1) GDH (glutamate dehydrogenase); **(2)** GAD (glutamate decarboxylase); **(3)** GABAT (γ-aminobutyrate transaminase); and **(4)** SSADH (succinate semialdehyde dehydrogenase). The adjacent Krebs Cycle is also indicated, along with some of its relevant enzymes: IDH (isocitrate dehydrogenase); OGDC (α-ketoglutarate dehydrogenase); SCS (succinyl-CoA synthetase); and SDH (succinate dehydrogenase). HOT and SSADH, the focus of this dissertation, are highlighted in **green**.

The GABA shunt is a closed-loop process that branches off from the Krebs cycle and is responsible for maintaining the concentration of GABA, though both its *production* and

conservation.[5, 65] The *conservation* becomes apparent when the four steps in the pathway are examined in more detail (**Table 1.1**):

Table 1.1. The four steps of the GABA Shunt.

Step	Enzyme	Reaction
1	glutamate dehydrogenase*	α-ketoglutarate + NAD(P)H + $\overset{\oplus}{N}H_4$ $\overset{GDH}{\rightleftharpoons}$ L-glutamate + $NAD(P)^+$
2	glutamate decarboxylase	L-glutamate + $\overset{\oplus}{H}$ $\overset{GAD\ (PLP)}{\rightleftharpoons}$ γ-aminobutyrate (GABA) + CO_2
3	GABA transaminase	γ-aminobutyrate (GABA) + α-ketoglutarate $\overset{GABAT\ (PLP)}{\rightleftharpoons}$ succinate semialdehyde + L-glutamate
4	SSA dehydrogenase	succinate semialdehyde + $NAD(P)^+$ $\overset{SSADH}{\rightleftharpoons}$ succinate + NAD(P)H + $\overset{\oplus}{H}$
* GABA transaminase (in step 3) can also effectively catalyze step 1.		

The first step of the GABA shunt is the formation of L-glutamate from α-ketoglutarate, a well-known Krebs cycle intermediate. Two enzymes in the pathway can catalyze this reaction: glutamate dehydrogenase (in *Step 1*) and GABA transaminase (in *Step 3*). Later during the GABA shunt, in the third step, GABA transaminase converts GABA to succinate semialdehyde (SSA), preferentially using α-ketoglutarate as the α-keto acid co-substrate and producing L-glutamate in the process. This preference for α-ketoglutarate results in *conservation*: GABA is metabolized only if a precursor is formed in the process.[5]

The importance of the GABA shunt is highlighted by its role in several disorders. The impairment of GABA-mediated inhibition in the brain is one proposed cause for epileptic seizures. Several activators and inhibitors of GABA shunt enzymes, including *SSADH*, have been studied for their ability to increase GABA concentrations, demonstrating an ability overcome this impairment.[66] The GABA shunt is not only critical for maintaining GABA concentrations, but it also functions as an alternative energy pathway that provides carboxylic acids to the Krebs cycle when the glycolytic pathway is compromised. Cerebral amyloid angiopathy caused by Alzheimer's disease can produce ischemia and trigger activation of the

GABA shunt in order to enhance cellular energy metabolism.[67] Thus, the proper functioning of all GABA shunt enzymes is critical for neurological health and a loss of even one of the enzymes can have profound consequences, as in the case of succinate semialdehyde dehydrogenase deficiency (Section 1.4.5).

1.3.2 The GHB Pathways

Aside from the reaction catalyzed by HOT, a number of distinct pathways have been identified that are responsible for the formation and breakdown of GHB, and some of these pathways are illustrated in **Figure 1.4**.[34] SSA can be reduced to GHB by a cytosolic succinate semialdehyde reductase (*SSAR*), an $NADP^+$-dependent enzyme situated on a pathway tangent to the GABA shunt.[68] SSAR is sometimes referenced as *aflatoxin B1 aldehyde reductase member 2*.[69] The oxidation of GHB back to succinate semialdehyde (SSA) can also take place in the cytosol, and this reaction is catalyzed by an $NADP^+$-dependent oxidoreductase known simply as GHB dehydrogenase, an enzyme that couples the reaction to reduction D-glucoronate to L-gluconate.[70] Elevated serum levels of 2,4-dihydroxybutyrate (2,4-DHBA) and

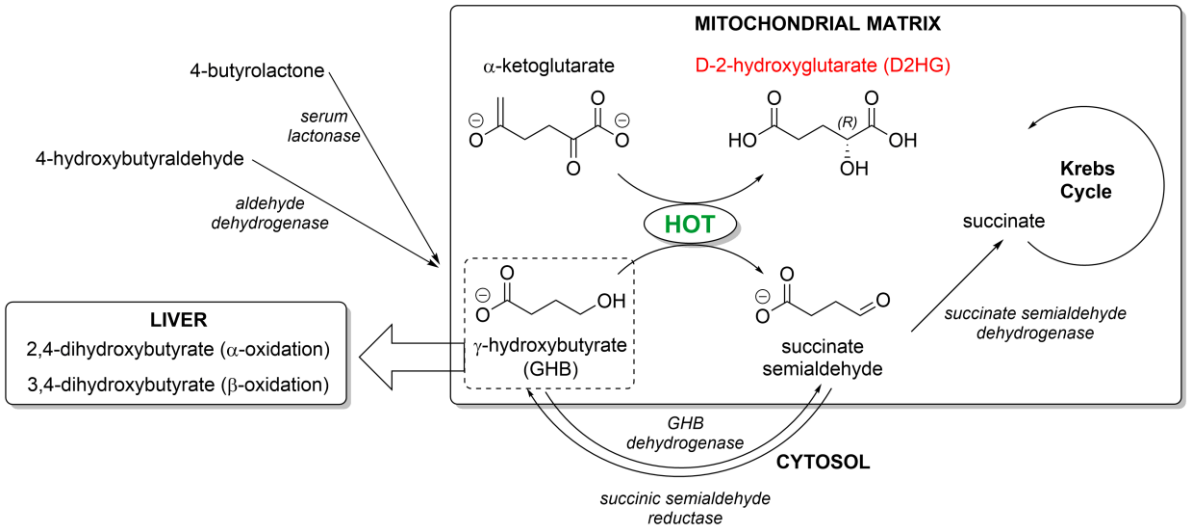

Figure 1.4. The GHB pathways.

3,4-dihydroxybutyrate (3,4-DHBA) have been observed during the progression from mild cognitive impairment to definitive Alzheimer's disease.[71] This increase in serum 2,4-DHBA and 3,4-DHBA appears to result from α-oxidation and β-oxidation of GHB in the liver[72-74] where the source of GHB is thought to come from the GABA shunt intermediate SSA. Indeed, the increased production of 2,4-DHBA and 2,3-DHBA is also observed in succinate semialdehyde dehydrogenase deficiency (see *Section 1.4.5*), where a mutation in SSADH results in excess SSA being converted into GHB by HOT or SSAR.[75] Finally, evidence exists for the formation of GHB from the precursors 4-butyrolacton and 4-hydroxybutyraldehyde by serum lactonase and an aldehyde dehydrogenase, respectively.[34, 76]

1.3.3 Pathways Involving D2HG

There are six possible biosynthetic pathways for D2HG (**Figure 1.5**).[54, 60, 62, 64, 77] Four of the pathways are the result of *normal metabolism*, while the and the fourth and fifth result from the *neomorphic activity* of mutated enzymes. In animals, one *normal* metabolic origin for D2HG is from the degradation of 5-aminolevulinate, a *non-proteinogenic* amino acid involved in heme synthesis.[78] A second normal route, observed in rats, is through the metabolism of 5-hydroxyl-L-lysine.[79] The third *normal* metabolic origin for D2HG is by the action of the mitochondrial, NAD^+-dependent hydroxyacid-oxoacid-transhydrogenase (HOT), which

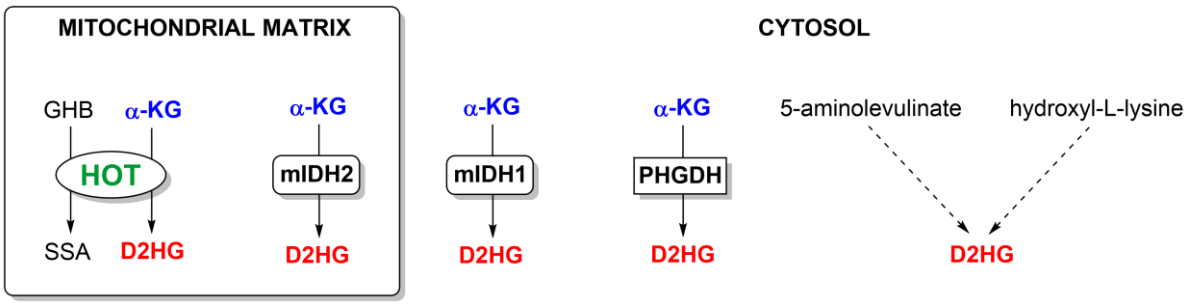

Figure 1.5. Pathways involving D-2-hydroxyglutarate formation.

couples the oxidation of GHB to the reduction of *α*-ketoglutarate, forming succinate semialdehyde (SSA) and D2HG in the process (**Figure 1.1**).[80]

The fourth and fifth routes to D2HG result from *gain-of-function* mutations in cytosolic isocitrate dehydrogenase isoform 1 (IDH1) and mitochondrial isocitrate dehydrogenase isoform 2 (IDH2). The normal function of IDH1 and IDH2 is to decarboxylate and oxidize isocitrate to *α*-ketoglutarate. This *neomorphic* activitiy of IDH1 and IDH2 is highly correlated to several types of cancer and is presumed to result from the ability of D2HG to act as a potent *oncometabolite* (discussed in Section 1.2.3).[53] Once formed, there are only two possible fates for D2HG, either oxidation to *α*-ketoglutarate by the mitochondrial FAD-dependent D-2-hydroxyglutarate dehydrogenase[81], or by mitochondrial HOT catalyzing the reduction of SSA[80], using 2DHG as the co-substrate to produce GHB and *α*-ketoglutarate as products.

L2HG is not an intermediate of any known biosynthetic pathway, appearing to be exclusively formed though the slow and non-specific NADPH-dependent conversion of AKG to L2HG by mitochondrial *L-malate dehydrogenase*, or by *lactate dehydrogenase A* under hypoxic conditions.[82] The only known enzyme to oxidize L2HG back to AKG is mitochondrial, FAD-dependent *L-2-hydroxyglutarate dehydrogenase*, an example of an enzyme whose only function appears to be “metabolite repair”.[83]

1.4 Succinate Semialdehyde Dehydrogenase

1.4.1 SSADH Belongs to the ALDH Superfamily

SSADHs belong to the aldehyde dehydrogenase (ALDH) superfamily of enzymes, which catalyze the oxidation of aldehydes to carboxylic acids. The ALDHs play a vital role in the detoxification of endogenous and exogenous aldehydes, are involved in folate and retinoic acid metabolism, and even act as a crystallin to absorb harmful UV radiation in the cornea.[84] SSADH is ubiquitous in nature, found in every biological kingdom on the planet including Animalia[85-91], Archaea[92, 93], Bacteria[94-99], Fungi[100, 101], and Plantae[102].

Earliest identification and characterization of SSADH (from *Pseudomonas fluorescens*) was made by Jakoby in 1959.[103] Given the 60 intervening years, and the limitations of early biochemical assays, much confusion has been introduced into the literature. Using the Enzyme Commission EC classification system[104-106], SSADHs have been assigned the following numbers (**Table 1.2**), based largely on cofactor preference. Despite this system, misclassifications are still present in older literature. Fortunately, much has been learned about cofactor preference and the role of specific amino acids at a critical site the ALDH cofactor binding pocket, which should help with classification going forward. More than 75% of

Table 1.2. The EC numbers for different SSADH enzymes.

EC 1	Oxidoreductases	
EC 1.2	Acting on the aldehyde or oxo group of donors	
EC 1.2.1	With NAD^+ or $NADP^+$ as acceptor	
EC No.	**Gene**	**Description**
1.2.1.9	?	glyceraldehyde-3-phosphate dehydrogenase ($NADP^+$)
1.2.1.16	*ynel*(*sad*)	succinate-semialdehyde dehydrogenase [$NAD(P)^+$]
1.2.1.24	*aldh5A1*	succinate-semialdehyde dehydrogenase (NAD^+)
1.2.1.76	?	succinate-semialdehyde dehydrogenase (acetylating)
1.2.1.79	*gabD*	succinate-semialdehyde dehydrogenase ($NADP^+$)

NAD^+-dependent ALDHs contain glutamate at a critical position, while 60% of $NADP^+$-dependent ALDHs have a serine and 30% have a threonine.[107] Of course, this makes sense since negatively charged glutamates are more likely to be repelled by the 2′-phosphate of $NADP^+$, while polar serine and threonine can form hydrogen bonds with this moiety.

1.4.2 *E. coli* $NADP^+$-dependent SSADH

Like most organisms, the *E. coli* genome encodes two SSADHs with only 32% sequence identity, one the product of the *yneI* gene (also known as *sad*) and the other the product of the *gabD* gene.[108] The *yneI* product demonstrates no significant preference for coenzyme, while the *gabD* product is $NADP^+$-dependent.[108]

A brief history of E. coli SSADH follows. In 1972, Dover isolated mutant strains of *E.coli* K-12 that could survive on γ-aminobutyrate as the sole source of carbon and nitrogen.[109] In 1980, Cozzani purified two proteins from *E.coli* K-12 3000, both with SSADH activity, but differing in coenzyme preferences as well as other structural and functional characteristics.[110] Likewise, Donnelly in 1981 was studying both NAD^+-dependent and $NADP^+$-dependent SSADH reactions in *E. coli*. The first observation of $NADP^+$-dependent SSADH was found in mutants discovered by Skinner in 1982.[111] In 1988, Marek cloned and expressed an NAD^+-dependent SSADH from E. coli.[112] The first computation and experimental verifications of the *E. coli* SSADH *gabD* and *yneI* genes were made by Fuhrer in 2007.[113] This was followed the next year with NMR saturation transfer difference studies on *E. coli* SSADH substrates and inhibitors by Jaeger.[95] This was finally followed by the first X-ray crystal structure of $NADP^+$-dependent SSADH (*gabD*) by Langendorf in 2010.[96]

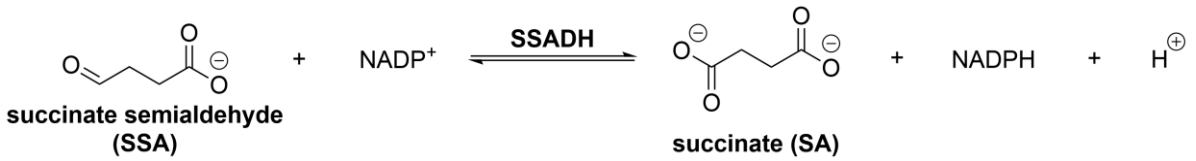

Figure 1.6. The reaction catalyzed by *E. coli* NADP$^+$-dependent SSADH.

The present research described was conducted with Langendorf's NADP$^+$-dependent SSADH (*gabD*), which catalyzes the reaction shown in **Figure 1.6**. The EC and UniProt identifiers for this enzyme are 1.2.1.79 and P25526, respectively.

1.4.3 The Structure of *E. coli* NADP$^+$-dependent SSADH

Native *E. coli* NADP$^+$-dependent SSADH (*gabD*) is a 51,720 kDa protein, comprised of 482 amino acids. The isoelectric point (pI) and grand average of hydropathicity (GRAVY) value for the *E. coli gabD* product are 5.44 and -0.071, respectively.

The first X-ray crystal structure for NADP$^+$-dependent SSADH (*gabD*) was obtained by Langendorf in 2010 and is presented in **Figure 1.7**.[96] The protein crystalizes as a tetramer, with beta sheets that weave their way in-between and connect the monomers. The structure has a

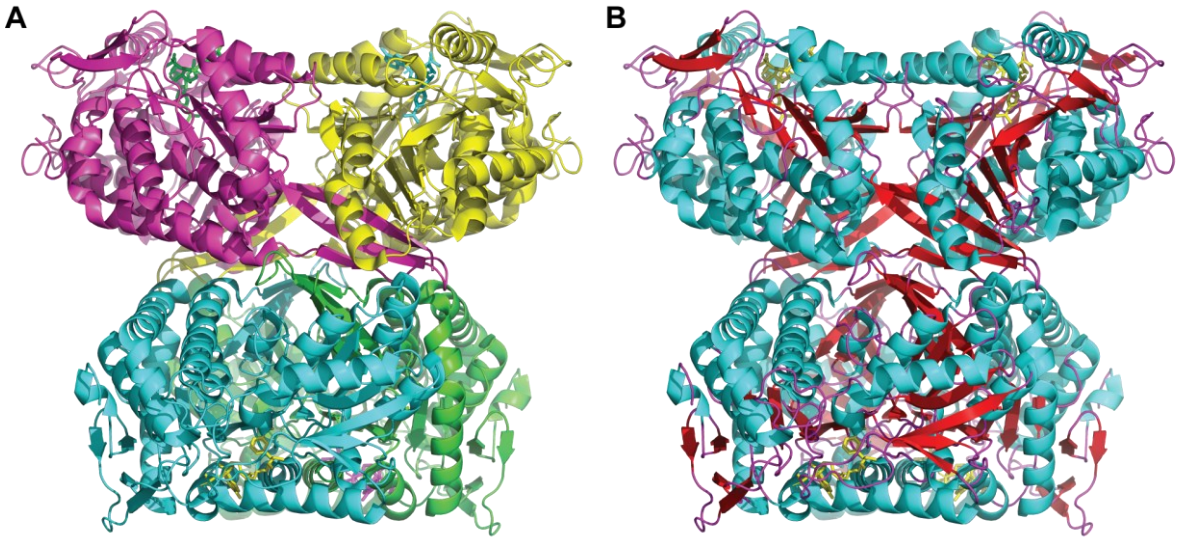

Figure 1.7. The X-ray crystal structure for *E. coli gabD* SSADH. (A) The SSADH tetramer represented by chains A, B, C and D in different colors. **(B)** The tetramer colored by secondary structure, showing that the beta sheets (red) form a pattern that weaves between and connects monomers into a stable tetramer.

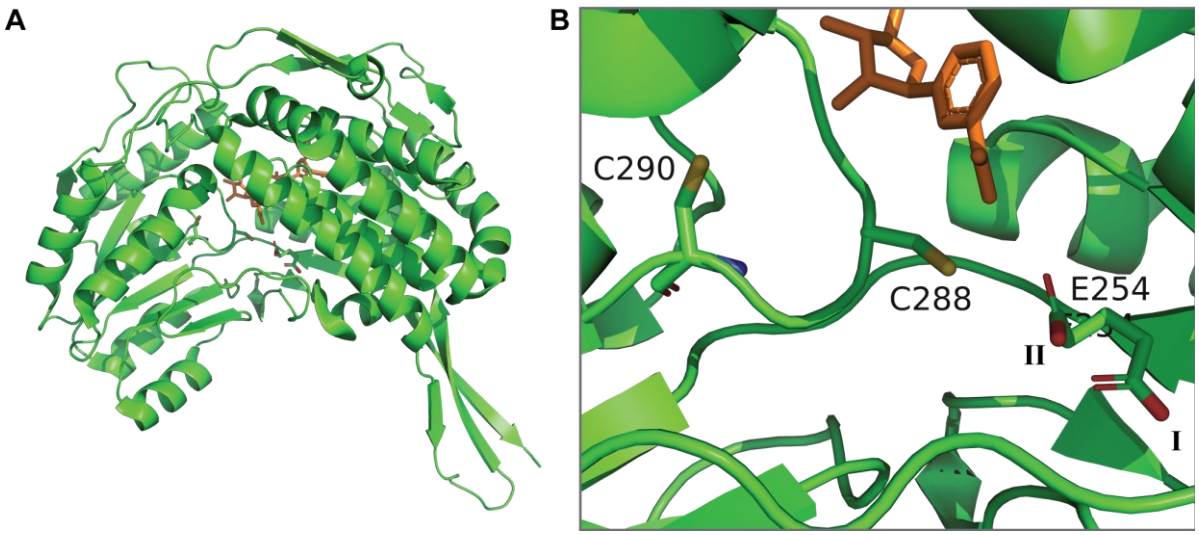

Figure 1.8. The *E. coli gabD* monomer and active site. (A) The *E. coli gabD* monomer. **(B)** The active site of the *E. coli gabD* monomer with $NADP^+$ in orange. The catalytic C288 is shown along with its redox-sensing partner, C290. SSADH crystallized with Glu254 in two conformations. Conformation I represents glutamine conformation during the hydride transfer, while II represents the glutamine position for deacylation.

pair of cysteines (C288 and C290) that may function together to sense the redox environment through reversible disulfide bond formation (**Figure 1.8**).[114] One of the cysteines (C288) is positioned to act as a nucleophile, under reduced conditions, which initiates the attack on the carbonyl carbon of the SSA aldehyde substrate. A glutamate (Glu254; **Figure 1.8**) in the active site is also positioned to facilitate the deacylation step by acting as a general base to deprotonate a water molecule.[115, 116]

1.4.4 The Function of *E. coli* $NADP^+$-dependent SSADH

The catalytic chemical mechanism for *E. coli* $NADP^+$-dependent SSADH is shown in **Figure 1.9**. There are two overall steps in the mechanism, an acylation step followed by deacylation. The reaction is initiated by nucleophilic attack by Cys288 on the carbonyl carbon of SSA to form an oxyanion substrate. As the lone pair of electrons on the oxyanion reform the carbonyl bond, a hydride is transferred from the carbonyl carbon to $NADP^+$, forming NADPH. The resultant thioester is then attacked by hydroxide, generated from water via

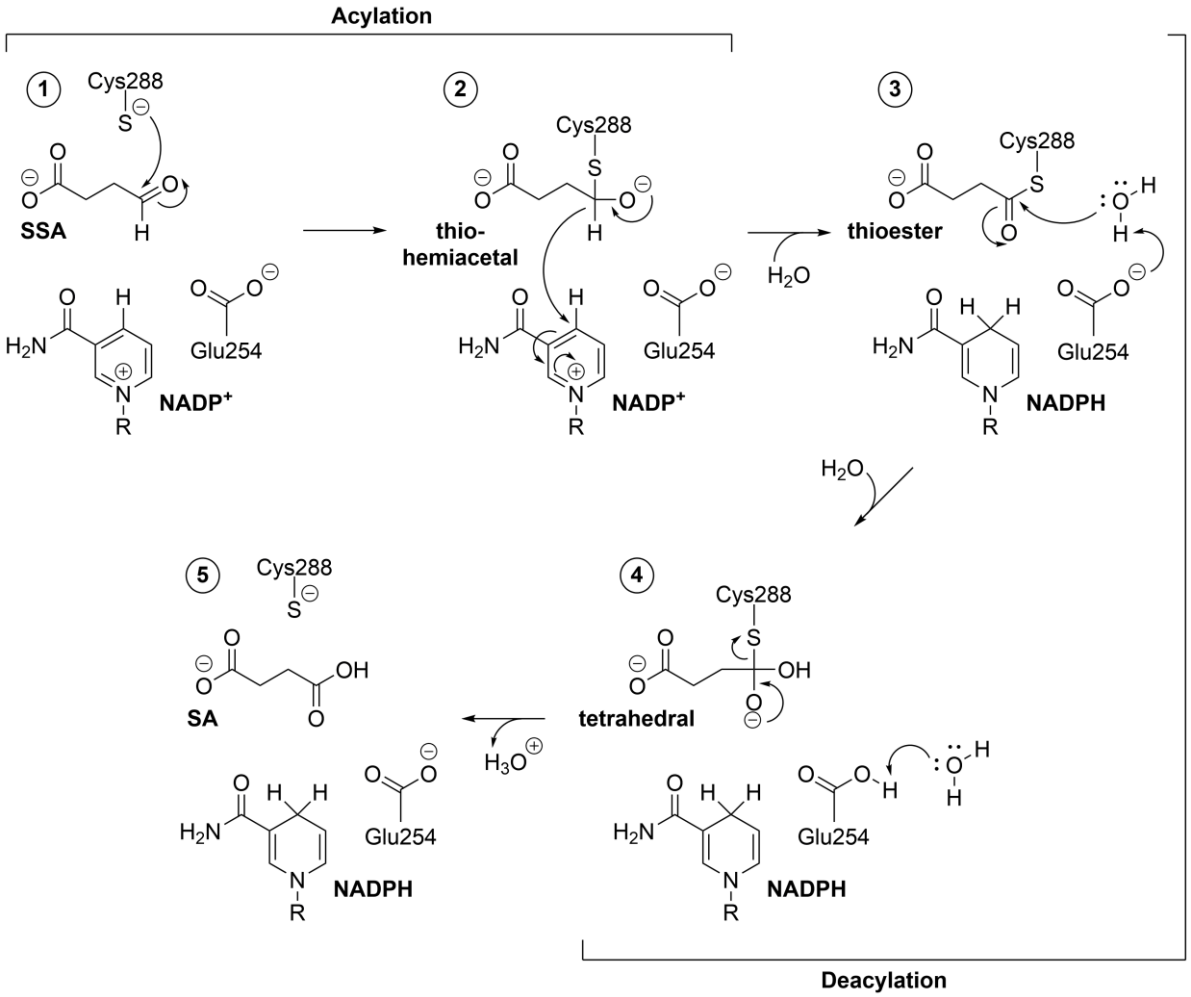

Figure 1.9. The proposed chemical mechanism for NADP$^+$-dependent *E. coli* SSADH.

general base catalysis by Glu254. A second water abstracts the proton from Glu254, returning it to its initial deprotonated state and the tetrahedral intermediate collapses. Due to the reactivity of aldehydes with primary amines, it is advisable to avoid the use of Tris buffer in SSADH assays because Shiff base formation between SSA and the primary amine of Tris can result in noticeably decreased the reaction rates.[117, 118]

1.4.5 Succinate Semialdehyde Dehydrogenase Deficiency

Succinate semialdehyde dehydrogenase deficiency, sometimes referred to a 4-hydroxybutyric aciduria, is an extremely rare autosomal recessive disorder that affects

approximately 350 people worldwide, with the first case being identified in 1981.[119, 120] Symptoms include developmental delay, intellectual disability, ataxia, hypotonia, and epileptic seizures in nearly half of patients with the disorder.[119] SSADH deficiency is caused by two point mutations located at different highly conserved intron/exon boundaries in the human ALD5A1 gene (chromosome locus 6p22).[121] These mutations prevent proper recognition of the normal splice junction, causing exon skipping and resulting in either: 1) a frameshift and premature termination; or 2) an in-frame deletion in the protein.[122] The end result of the mutations is a loss of SSADH activity, which results in a built up of succinate semialdehyde, the net result of which is the formation of GHB by either succinate semialdehyde reductase (SSAR) or possibly the by the HOT enzyme catalyzing the reverse reaction and generating GHB from SSA.[123, 124] GHB is thought to be the cause of the symptoms for succinate semialdehyde dehydrogenase activity, which is why is the disorder is sometimes referred to as 4-hydroxybutyric aciduria.

1.5 Hydroxyacid Oxoacid Transhydrogenase

Hydroxyacid oxoacid transhydrogenase (HOT) is an oxidoreductase, encoded by the ADHFe1 gene. HOT catalyzes the oxidation of GHB to succinate semialdehyde using α-ketoglutarate as a co-substrate, which in turn is reduced to D-2-hydroxyglutarate (D2HG). HOT was first discovered by Elaine E. Kaufman and Thomas Nelson in 1988, as they were researching pathways for the oxidation of GHB in the cytoplasm and mitochondria of rat cells.[80, 125] They noted that the activity of this unknown enzyme did not depend on the concentration of $NAD(P)^+$, but rather it was stoichiometric with the metabolism of α-ketoglutarate.[80] Kaufman and Nelson noted similarities between the reaction catalyzed by their enzyme and a transhydrogenase isolated from bacteria. Malate-lactate transhydrogenase from *Micrococcus lactilyticus* catalyzed a similar reaction, was known to have a tightly bound NAD^+ prosthetic group, and displayed ping-pong (double displacement) enzyme kinetics.[126] Enzymes with a tightly bound $NAD(P)^+$, acting as a *cofactor*, are called nicotinoproteins.[127] Structural information has revealed that tight binding in nicotinoproteins is due to additional hydrogen bonding between the $NAD(P)^+$ cofactor and residues in the binding pocket.[127]

Therefore, Kaufman and Nelson proposed that HOT oxidized GHB to succinate semialdehyde with concomitant reduction of a tightly bound NAD^+ *cofactor* that does not exchange with the solvent NAD^+ pool. The NAD^+ cofactor is regenerated in the active site by reduction of an α-ketoglutarate co-substrate to D-2-hydroxyglutarate (D2HG). They also noted that the enzyme can catalyze the transhydrogenase reaction using a number of different hydroxyacids (e.g., GHB, D/L-lactate, etc.) and oxoacids (e.g., α-ketoglutarate, oxaloacetate, etc.), thus describing the enzyme as a hydroxyacid-oxoacid transhydrogenase. For these reasons, HOT is given the EC number of 1.1.99.24 (**Table 1.3**).

Table 1.3. The EC number for HOT.

EC 1	Oxidoreductases
EC 1.1	Acting on the CH-OH group of donors
EC 1.1.99	With unknown physiological acceptors
EC 1.1.99.24	Hydroxyacid-oxoacid transhydrogenase

In 2002, the gene encoding HOT (ADHFe1) was cloned from human fetal brain mRNA and characterized as a novel human alcohol dehydrogenase (ADH).[128] Multiple sequence alignment resulted in a significant similarity between ADHFe1 product and other Group III ADHs, and no significant similarities with Group I and Group II ADHs.[128] Group III ADHs are a very poorly studied group of ADHs, distributed primarily in the Archaea and Bacteria domains.[129] What is known with certainty about Group III ADHs is that they are iron-containing.[130] A recent BLAST[131, 132] search revealed as of 2024, HOT it is the only iron-containing group III ADH represented in animals.

Some questions regarding the structure, function, and purpose of HOT remain. HOT has yet to be overexpressed as a soluble protein in quantities required for X-ray and NMR structure determination. Thus, the roles of iron and NAD^+ in the mechanism have not been confirmed. Some enzyme kinetics have been done on partially purified HOT from animal cells, but complete multi-substrate enzyme kinetics have not been done and the enzyme mechanism has yet to been determined. Furthermore, the role of HOT in animal biology is still in question, as is how a microbial iron-containing Group III ADH came to be expressed in eukaryotes. It is possible that the HOT reaction plays a role in metabolite repair, converting the oncometabolite 2DHG back to α-ketoglutarate. In fact, the ADHFe1 gene is hypermethylated and HOT expression is downregulated in various cancer lines, while a high level of HOT expression has been corelated to patient survival in breast, colon and gastric cancers.[133, 134]

1.5.1 The Structure and Function of HOT

Native human HOT (ADHFe1) is a 50,308 kDa protein, comprised of 467 amino acids. The isoelectric point (pI) and grand average of hydropathicity (GRAVY) value for HOT are 7.65 and 0.006, respectively. HOT is both slightly hydrophobic and as would have close to a zero net charge in the pH range measured[1] for the mitochonrial matrix. Native mouse HOT differs from human HOT only toward the N-terminus, where the residues Leu36 and Ser37 are deleted in the mouse homolog. These deletions are expected to have no consequence on the reaction mechanism, as the first 48 amino acids in human HOT appear to coorespond to a mitochondrial transit peptide that would be trimmed from the mature protein following import into the mitochondria. UniProt[135] predicts two potential post-translational modifications (PTMs) in human HOT, N^6-acetyllysine at position 445 and phosphoserine at position 452. The same PTMs are predicted for mouse HOT in the corresponding positions.

Experimental evidence indicates that HOT has two primary isoforms, which differ only in the presence of a mitochondrial targeting peptide (MTP) at the N-terminus.[76, 136] *Isoform 1* (M1HOT) is ~50 kDa, while *isoform 2* (M2HOT) is ~45 kDa, with the difference in size being attributed to the 48 amino acid MTP presequence at the N-terminus.

While no experimentally derived structures for HOT exist, some information may be obtained by analyzing the HOT primary amino acid sequence with computer algorithms, such as Phobius[137, 138], SWISS-MODEL[139-143], and AlphaFold[144, 145].

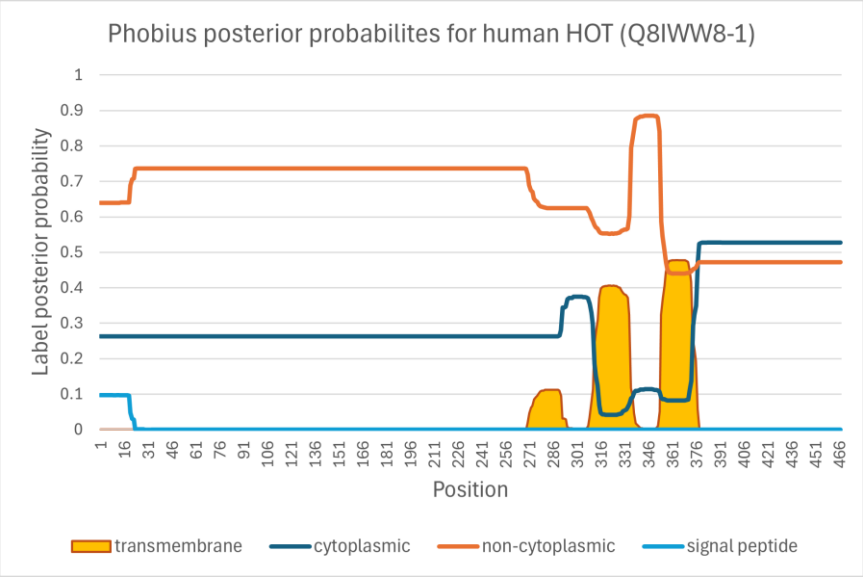

Figure 1.10. A Phobius posterior probability plot for human HOT.

A Phobius posterior probability plot for human HOT is shown in **Figure 1.10** and predicts the presence of a signal sequence at the N-terminus. The prediction also suggests that HOT that the subcellular location of HOT is non-cytoplasmic.

Homology modelling with SWISS-MODEL, using the 2.70 Å X-ray crystal structure of 1,3-propanediol oxidoreductase (PDB 3BFJ)[146] as a template, results in the predicted structure in **Figure 1.11**. Likewise, AlphaFold, which uses artificial intelligence rather than a template,

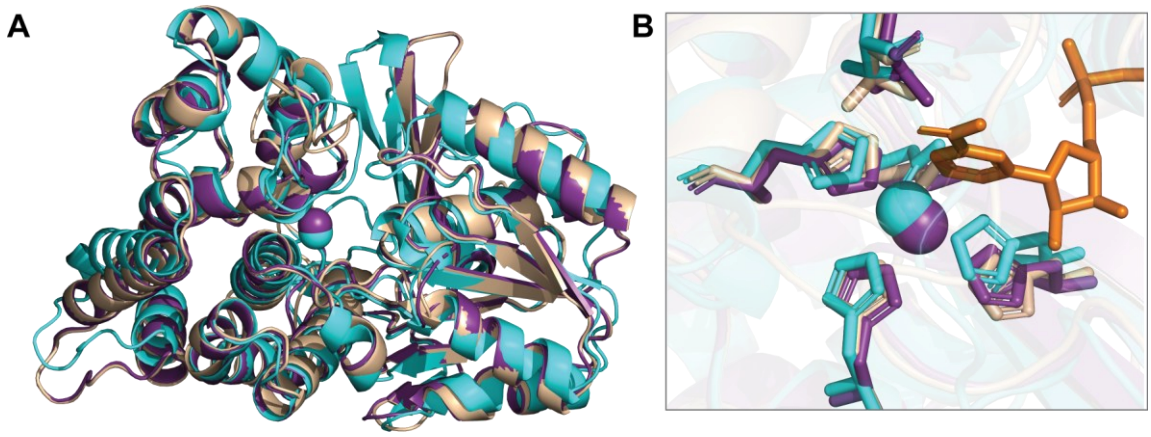

Figure 1.11. The alignments of a homology model and AlphaFold prediction. (A) The 3BFJ template is colored purple, the homology model is colored in wheat, and the AlphaFold prediction is colored in cyan. Note that the homology model cannot predict the iron atom, while AlphaFold can predict both the iron and NAD+ cofactor (orange). **(B)** A closeup of the predicted active site.

also predicts a structure that nicely aligns with both the 3BFJ template and homology model. The RMSD between the homology model and 3BFJ template is 0.383 Å, while the RMSD between the homology model and the AlphaFold prediction is 0.383 Å.

The AlphFold predicted structure for human M1HOT is presented in **Figure 1.12**, with annotations that illustrate the deleted amino acids in mouse M1HOT, the PTMs, disulfide bonds and *cis*-proline residues (Chapter 3), the predicted MTP at the N-terminus, as well as the start of *isoform 1* (M1) and *isoform 2* (M2).

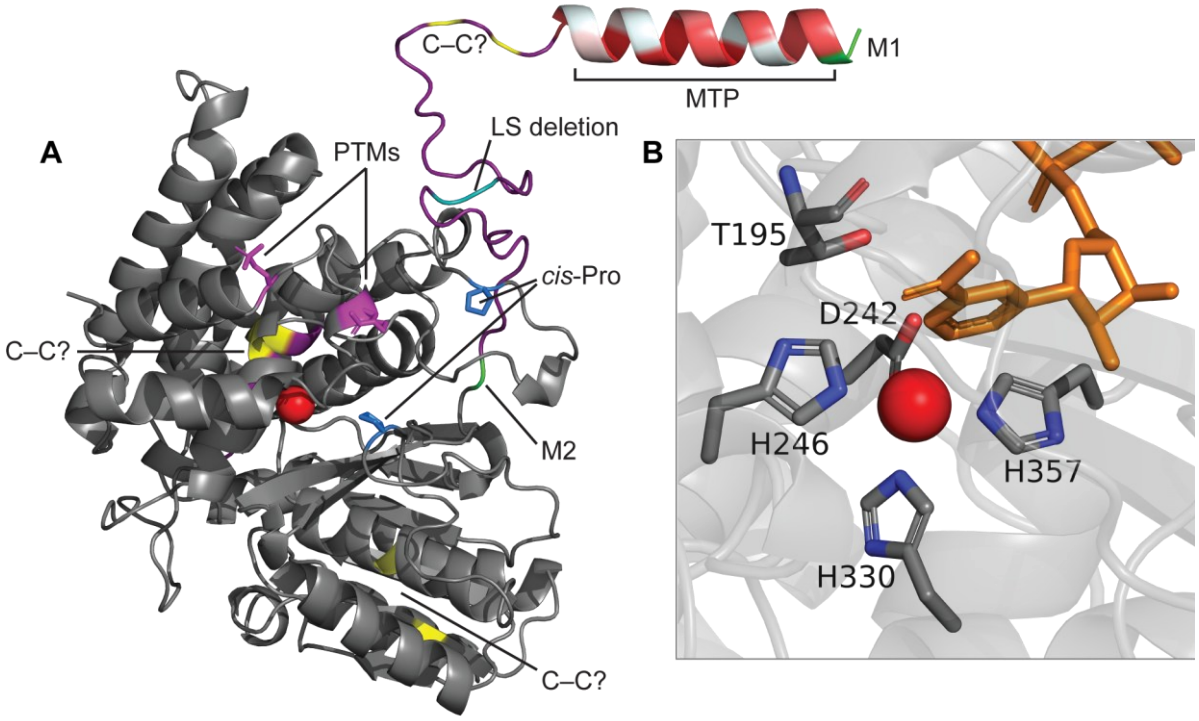

Figure 1.12. The AlphaFold prediction for human M1HOT.

1.6 References

(1) Santo-Domingo, J.; Demaurex, N. Perspectives on: SGP symposium on mitochondrial physiology and medicine: the renaissance of mitochondrial pH. *J Gen Physiol* **2012**, *139* (6), 415-423. DOI: 10.1085/jgp.201110767

(2) Li, K.; Xu, E. The role and the mechanism of gamma-aminobutyric acid during central nervous system development. *Neurosci Bull* **2008**, *24* (3), 195-200. DOI: 10.1007/s12264-008-0109-3

(3) Roberts, E.; Frankel, S. gamma-Aminobutyric acid in brain: its formation from glutamic acid. *J Biol Chem* **1950**, *187* (1), 55-63.

(4) Ke, Y.; Cohen, B. M.; Bang, J. Y.; Yang, M.; Renshaw, P. F. Assessment of GABA concentration in human brain using two-dimensional proton magnetic resonance spectroscopy. *Psychiatry Research: Neuroimaging* **2000**, *100* (3), 169-178. DOI: 10.1016/S0925-4927(00)00075-5

(5) Olsen, R. W.; DeLorey, T. M. GABA Synthesis, Uptake and Release. In *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, 6th ed.; Siegel, G. J. Ed.; Lippincott-Raven, 1999.

(6) Basemore, A. W.; Elliot, K. A.; Florey, E. Isolation of factor I. *J Neurochem* **1957**, *1* (4), 334-339. DOI: 10.1111/j.1471-4159.1957.tb12090.x

(7) Ghit, A.; Assal, D.; Al-Shami, A. S.; Hussein, D. E. E. GABA(A) receptors: structure, function, pharmacology, and related disorders. *J Genet Eng Biotechnol* **2021**, *19* (1), 123. DOI: 10.1186/s43141-021-00224-0

(8) Johnston, G. A. GABAA receptor pharmacology. *Pharmacol Ther* **1996**, *69* (3), 173-198. DOI: 10.1016/0163-7258(95)02043-8

(9) Benarroch, E. E. GABAB receptors: structure, functions, and clinical implications. *Neurology* **2012**, *78* (8), 578-584. DOI: 10.1212/WNL.0b013e318247cd03

(10) Evenseth, L. S. M.; Gabrielsen, M.; Sylte, I. The GABA(B) Receptor-Structure, Ligand Binding and Drug Development. *Molecules* **2020**, *25* (13). DOI: 10.3390/molecules25133093

(11) Tyacke, R. J.; Lingford-Hughes, A.; Reed, L. J.; Nutt, D. J. GABAB receptors in addiction and its treatment. *Adv Pharmacol* **2010**, *58*, 373-396. DOI: 10.1016/s1054-3589(10)58014-1

(12) Vlachou, S.; Markou, A. GABAB receptors in reward processes. *Adv Pharmacol* **2010**, *58*, 315-371. DOI: 10.1016/s1054-3589(10)58013-x

(13) Filip, M.; Frankowska, M.; Sadakierska-Chudy, A.; Suder, A.; Szumiec, L.; Mierzejewski, P.; Bienkowski, P.; Przegaliński, E.; Cryan, J. F. GABAB receptors as a

therapeutic strategy in substance use disorders: focus on positive allosteric modulators. *Neuropharmacology* **2015**, *88*, 36-47. DOI: 10.1016/j.neuropharm.2014.06.016

(14) Kuriyama, K.; Sze, P. Y. Blood-brain barrier to H3-gamma-aminobutyric acid in normal and amino oxyacetic acid-treated animals. *Neuropharmacology* **1971**, *10* (1), 103-108. DOI: 10.1016/0028-3908(71)90013-x

(15) Hepsomali, P.; Groeger, J. A.; Nishihira, J.; Scholey, A. Effects of Oral Gamma-Aminobutyric Acid (GABA) Administration on Stress and Sleep in Humans: A Systematic Review. *Front Neurosci* **2020**, *14*, 923. DOI: 10.3389/fnins.2020.00923

(16) Boonstra, E.; de Kleijn, R.; Colzato, L. S.; Alkemade, A.; Forstmann, B. U.; Nieuwenhuis, S. Neurotransmitters as food supplements: the effects of GABA on brain and behavior. *Front Psychol* **2015**, *6*, 1520. DOI: 10.3389/fpsyg.2015.01520

(17) Bernasconi, R.; Mathivet, P.; Otten, U.; Bettler, B.; Bishoff, S.; Marescaux, C. Part of the pharmacological actions of ??-hydroxybutyrate are mediated by GABAB receptors. *Gamma-hydroxybutyrate: Molecular, Functional and Clinical Aspects* **2002**, 28-63. DOI: 10.4324/9780203300992_chapter_3

(18) Nelson, T.; Kaufman, E.; Kline, J.; Sokoloff, L. The extraneural distribution of gamma-hydroxybutyrate. *J Neurochem* **1981**, *37* (5), 1345-1348. DOI: 10.1111/j.1471-4159.1981.tb04689.x

(19) Saytzeff, A. 4. Ueber die Reduction des Succinylchlorids. *Justus Liebigs Annalen der Chemie* **1874**, *171* (2), 258-290. DOI: 10.1002/jlac.18741710216

(20) Samson, F. E., Jr.; White, R. P. Effects of fatty acid anions on the electroencephalogram of unanesthetized rabbits. *Am J Physiol* **1956**, *186* (2), 271-274. DOI: 10.1152/ajplegacy.1956.186.2.271

(21) Laborit, H. SODIUM 4-HYDROXYBUTYRATE. *Int J Neuropharmacol* **1964**, *3*, 433-451. DOI: 10.1016/0028-3908(64)90074-7

(22) Bessman, S. P.; Fishbein, W. N. GAMMA-HYDROXYBUTYRATE, A NORMAL BRAIN METABOLITE. *Nature* **1963**, *200*, 1207-1208. DOI: 10.1038/2001207a0

(23) Laborit, H. Gamma-hydroxybutyrate, succinic semialdehyde and sleep. *Progress in Neurobiology* **1973**, *1*, 255-274. DOI: 10.1016/0301-0082(73)90014-2

(24) Mathivet, P.; Bernasconi, R.; De Barry, J.; Marescaux, C.; Bittiger, H. Binding characteristics of gamma-hydroxybutyric acid as a weak but selective GABAB receptor agonist. *Eur J Pharmacol* **1997**, *321* (1), 67-75. DOI: 10.1016/s0014-2999(96)00916-8

(25) Benavides, J.; Rumigny, J. F.; Bourguignon, J. J.; Cash, C.; Wermuth, C. G.; Mandel, P.; Vincendon, G.; Maitre, M. High affinity binding sites for gamma-hydroxybutyric acid in rat brain. *Life Sci* **1982**, *30* (11), 953-961. DOI: 10.1016/0024-3205(82)90624-5

(26) Andriamampandry, C.; Taleb, O.; Viry, S.; Muller, C.; Humbert, J. P.; Gobaille, S.; Aunis, D.; Maitre, M. Cloning and characterization of a rat brain receptor that binds the endogenous neuromodulator gamma-hydroxybutyrate (GHB). *Faseb j* **2003**, *17* (12), 1691-1693. DOI: 10.1096/fj.02-0846fje

(27) Andriamampandry, C.; Taleb, O.; Kemmel, V.; Humbert, J. P.; Aunis, D.; Maitre, M. Cloning and functional characterization of a gamma-hydroxybutyrate receptor identified in the human brain. *Faseb j* **2007**, *21* (3), 885-895. DOI: 10.1096/fj.06-6509com

(28) Maitre, M. The gamma-hydroxybutyrate signalling system in brain: organization and functional implications. *Prog Neurobiol* **1997**, *51* (3), 337-361. DOI: 10.1016/s0301-0082(96)00064-0

(29) Agabio, R.; Gessa, G. L. Therapeutic uses of γ-hydroxybutyrate. *Gamma-hydroxybutyrate: molecular, functional and clinical aspects* **2002**, 169-187. DOI: 10.4324/9780203300992_chapter_10

(30) Tunnicliff, G. Sites of action of gamma-hydroxybutyrate (GHB)--a neuroactive drug with abuse potential. *J Toxicol Clin Toxicol* **1997**, *35* (6), 581-590. DOI: 10.3109/15563659709001236

(31) Brailsford, A. D.; Bartlett, C.; Kicman, A. T.; Cowan, D. A. Increases in Serum Growth Hormone Concentrations Associated with GHB Administration. *J Anal Toxicol* **2017**, *41* (1), 54-59. DOI: 10.1093/jat/bkw107

(32) Galloway, G. P.; Frederick, S. L.; Staggers, F. E., Jr.; Gonzales, M.; Stalcup, S. A.; Smith, D. E. Gamma-hydroxybutyrate: an emerging drug of abuse that causes physical dependence. *Addiction* **1997**, *92* (1), 89-96.

(33) United States Controlled Substances Act. 21 U.S.C. Ch. 13 § 801 et seq., 1971.

(34) Parsons, S. M. Date-Rape Drugs with Emphasis on GHB. In *Forensic Chemistry Handbook*, 2011; pp 355-434.

(35) Bravo, D. T.; Harris, D. O.; Parsons, S. M. Reliable, sensitive, rapid and quantitative enzyme-based assay for gamma-hydroxybutyric acid (GHB). *J Forensic Sci* **2004**, *49* (2), 379-387. DOI: 10.1520/JFS2003165

(36) Gallimberti, L.; Cibin, M.; Pagnin, P.; Sabbion, R.; Pani, P. P.; Pirastu, R.; Ferrara, S. D.; Gessa, G. L. Gamma-hydroxybutyric acid for treatment of opiate withdrawal syndrome. *Neuropsychopharmacology* **1993**, *9* (1), 77-81. DOI: 10.1038/npp.1993.45

(37) Gallimberti, L.; Schifano, F.; Forza, G.; Miconi, L.; Ferrara, S. D. Clinical efficacy of gamma-hydroxybutyric acid in treatment of opiate withdrawal. *Eur Arch Psychiatry Clin Neurosci* **1994**, *244* (3), 113-114. DOI: 10.1007/bf02191883

(38) Gallimberti, L.; Spella, M. R.; Soncini, C. A.; Gessa, G. L. Gamma-hydroxybutyric acid in the treatment of alcohol and heroin dependence. *Alcohol* **2000**, *20* (3), 257-262. DOI: 10.1016/s0741-8329(99)00089-0

(39) Maldonado, C.; Rodríguez-Arias, M.; Aguilar, M. A.; Miñarro, J. GHB ameliorates naloxone-induced conditioned place aversion and physical aspects of morphine withdrawal in mice. *Psychopharmacology (Berl)* **2004**, *177* (1-2), 130-140. DOI: 10.1007/s00213-004-1927-2

(40) Addolorato, G.; Cibin, M.; Caputo, F.; Capristo, E.; Gessa, G. L.; Stefanini, G. F.; Gasbarrini, G. Gamma-hydroxybutyric acid in the treatment of alcoholism: dosage fractioning utility in non-responder alcoholic patients. *Drug Alcohol Depend* **1998**, *53* (1), 7-10. DOI: 10.1016/s0376-8716(98)00094-5

(41) Caputo, F.; Vignoli, T.; Maremmani, I.; Bernardi, M.; Zoli, G. Gamma hydroxybutyric acid (GHB) for the treatment of alcohol dependence: a review. *Int J Environ Res Public Health* **2009**, *6* (6), 1917-1929. DOI: 10.3390/ijerph6061917

(42) Stella, L.; Addolorato, G.; Rinaldi, B.; Capuano, A.; Berrino, L.; Rossi, F.; Maione, S. An open randomized study of the treatment of escitalopram alone and combined with gamma-hydroxybutyric acid and naltrexone in alcoholic patients. *Pharmacol Res* **2008**, *57* (4), 312-317. DOI: 10.1016/j.phrs.2008.03.001

(43) Leone, M. A.; Vigna-Taglianti, F.; Avanzi, G.; Brambilla, R.; Faggiano, F. Gamma-hydroxybutyrate (GHB) for treatment of alcohol withdrawal and prevention of relapses. *Cochrane Database Syst Rev* **2010**, (2), Cd006266. DOI: 10.1002/14651858.CD006266.pub2

(44) Addolorato, G.; Leggio, L.; Abenavoli, L.; Gasbarrini, G.; Caputo, F.; Vignoli, T.; Lorenzini, F.; Bernardi, M. Gamma hydroxybutyrric acid (GHB) withdrawal does not occur at therapeutic dosage. *Drug Alcohol Depend* **2005**, *77* (2), 209. DOI: 10.1016/j.drugalcdep.2004.09.003

(45) Caputo, F.; Francini, S.; Stoppo, M.; Lorenzini, F.; Vignoli, T.; Del Re, A.; Comaschi, C.; Leggio, L.; Addolorato, G.; Zoli, G.; et al. Incidence of craving for and abuse of gamma-hydroxybutyric acid (GHB) in different populations of treated alcoholics: an open comparative study. *J Psychopharmacol* **2009**, *23* (8), 883-890. DOI: 10.1177/0269881108094620

(46) Engqvist, M. K.; Eßer, C.; Maier, A.; Lercher, M. J.; Maurino, V. G. Mitochondrial 2-hydroxyglutarate metabolism. *Mitochondrion* **2014**, *19 Pt B*, 275-281. DOI: 10.1016/j.mito.2014.02.009

(47) Kranendijk, M.; Struys, E. A.; Salomons, G. S.; Van der Knaap, M. S.; Jakobs, C. Progress in understanding 2-hydroxyglutaric acidurias. *J Inherit Metab Dis* **2012**, *35* (4), 571-587. DOI: 10.1007/s10545-012-9462-5

(48) Struys, E. A.; Salomons, G. S.; Achouri, Y.; Van Schaftingen, E.; Grosso, S.; Craigen, W. J.; Verhoeven, N. M.; Jakobs, C. Mutations in the D-2-hydroxyglutarate dehydrogenase gene

cause D-2-hydroxyglutaric aciduria. *Am J Hum Genet* **2005**, *76* (2), 358-360. DOI: 10.1086/427890

(49) Rzem, R.; Veiga-da-Cunha, M.; Noël, G.; Goffette, S.; Nassogne, M. C.; Tabarki, B.; Schöller, C.; Marquardt, T.; Vikkula, M.; Van Schaftingen, E. A gene encoding a putative FAD-dependent L-2-hydroxyglutarate dehydrogenase is mutated in L-2-hydroxyglutaric aciduria. *Proc Natl Acad Sci U S A* **2004**, *101* (48), 16849-16854. DOI: 10.1073/pnas.0404840101

(50) Nota, B.; Struys, E. A.; Pop, A.; Jansen, E. E.; Fernandez Ojeda, M. R.; Kanhai, W. A.; Kranendijk, M.; van Dooren, S. J.; Bevova, M. R.; Sistermans, E. A.; et al. Deficiency in SLC25A1, encoding the mitochondrial citrate carrier, causes combined D-2- and L-2-hydroxyglutaric aciduria. *Am J Hum Genet* **2013**, *92* (4), 627-631. DOI: 10.1016/j.ajhg.2013.03.009

(51) Yan, H.; Parsons, D. W.; Jin, G.; McLendon, R.; Rasheed, B. A.; Yuan, W.; Kos, I.; Batinic-Haberle, I.; Jones, S.; Riggins, G. J.; et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **2009**, *360* (8), 765-773. DOI: 10.1056/NEJMoa0808710

(52) Dang, L.; White, D. W.; Gross, S.; Bennett, B. D.; Bittinger, M. A.; Driggers, E. M.; Fantin, V. R.; Jang, H. G.; Jin, S.; Keenan, M. C.; et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **2009**, *462* (7274), 739-744. DOI: 10.1038/nature08617

(53) Ward, P. S.; Patel, J.; Wise, D. R.; Abdel-Wahab, O.; Bennett, B. D.; Coller, H. A.; Cross, J. R.; Fantin, V. R.; Hedvat, C. V.; Perl, A. E.; et al. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **2010**, *17* (3), 225-234. DOI: 10.1016/j.ccr.2010.01.020

(54) Hvinden, I. C.; Cadoux-Hudson, T.; Schofield, C. J.; McCullagh, J. S. O. Metabolic adaptations in cancers expressing isocitrate dehydrogenase mutations. *Cell Rep Med* **2021**, *2* (12), 100469. DOI: 10.1016/j.xcrm.2021.100469

(55) Kranendijk, M.; Struys, E. A.; van Schaftingen, E.; Gibson, K. M.; Kanhai, W. A.; van der Knaap, M. S.; Amiel, J.; Buist, N. R.; Das, A. M.; de Klerk, J. B.; et al. IDH2 mutations in patients with D-2-hydroxyglutaric aciduria. *Science* **2010**, *330* (6002), 336. DOI: 10.1126/science.1192632

(56) Aghili, M.; Zahedi, F.; Rafiee, E. Hydroxyglutaric aciduria and malignant brain tumor: a case report and literature review. *J Neurooncol* **2009**, *91* (2), 233-236. DOI: 10.1007/s11060-008-9706-2

(57) Tabata, S.; Kojima, Y.; Sakamoto, T.; Igarashi, K.; Umetsu, K.; Ishikawa, T.; Hirayama, A.; Kajino-Sakamoto, R.; Sakamoto, N.; Yasumoto, K. I.; et al. L-2hydroxyglutaric acid rewires amino acid metabolism in colorectal cancer via the mTOR-ATF4 axis. *Oncogene* **2023**, *42* (16), 1294-1307. DOI: 10.1038/s41388-023-02632-7

(58) Chowdhury, R.; Yeoh, K. K.; Tian, Y. M.; Hillringhaus, L.; Bagg, E. A.; Rose, N. R.; Leung, I. K.; Li, X. S.; Woon, E. C.; Yang, M.; et al. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* **2011**, *12* (5), 463-469. DOI: 10.1038/embor.2011.43

(59) Losman, J. A.; Koivunen, P.; Kaelin, W. G., Jr. 2-Oxoglutarate-dependent dioxygenases in cancer. *Nat Rev Cancer* **2020**, *20* (12), 710-726. DOI: 10.1038/s41568-020-00303-3

(60) Yang, M.; Soga, T.; Pollard, P. J. Oncometabolites: linking altered metabolism with cancer. *J Clin Invest* **2013**, *123* (9), 3652-3658. DOI: 10.1172/jci67228

(61) Bunse, L.; Pusch, S.; Bunse, T.; Sahm, F.; Sanghvi, K.; Friedrich, M.; Alansary, D.; Sonner, J. K.; Green, E.; Deumelandt, K.; et al. Suppression of antitumor T cell immunity by the oncometabolite (R)-2-hydroxyglutarate. *Nat Med* **2018**, *24* (8), 1192-1203. DOI: 10.1038/s41591-018-0095-6

(62) Ryan, D. G.; Murphy, M. P.; Frezza, C.; Prag, H. A.; Chouchani, E. T.; O'Neill, L. A.; Mills, E. L. Coupling Krebs cycle metabolites to signalling in immunity and cancer. *Nat Metab* **2019**, *1*, 16-33. DOI: 10.1038/s42255-018-0014-7

(63) Faubert, B.; Solmonson, A.; DeBerardinis, R. J. Metabolic reprogramming and cancer progression. *Science* **2020**, *368* (6487). DOI: 10.1126/science.aaw5473

(64) Cai, M.; Zhao, J.; Ding, Q.; Wei, J. Oncometabolite 2-hydroxyglutarate regulates anti-tumor immunity. *Heliyon* **2024**, *10* (2), e24454. DOI: 10.1016/j.heliyon.2024.e24454

(65) Sarasa, S. B.; Mahendran, R.; Muthusamy, G.; Thankappan, B.; Selta, D. R. F.; Angayarkanni, J. A Brief Review on the Non-protein Amino Acid, Gamma-amino Butyric Acid (GABA): Its Production and Role in Microbes. *Curr Microbiol* **2020**, *77* (4), 534-544. DOI: 10.1007/s00284-019-01839-w

(66) Yogeeswari, P.; Sriram, D.; Vaigundaragavendran, J. The GABA shunt: an attractive and potential therapeutic target in the treatment of epileptic disorders. *Curr Drug Metab* **2005**, *6* (2), 127-139. DOI: 10.2174/1389200053586073

(67) Salminen, A.; Jouhten, P.; Sarajärvi, T.; Haapasalo, A.; Hiltunen, M. Hypoxia and GABA shunt activation in the pathogenesis of Alzheimer's disease. *Neurochem Int* **2016**, *92*, 13-24. DOI: 10.1016/j.neuint.2015.11.005

(68) Hearl, W. G.; Churchich, J. E. A mitochondrial NADP+-dependent reductase related to the 4-aminobutyrate shunt. Purification, characterization, and mechanism. *J Biol Chem* **1985**, *260* (30), 16361-16366.

(69) Schaller, M.; Schaffhauser, M.; Sans, N.; Wermuth, B. Cloning and expression of succinic semialdehyde reductase from human brain. Identity with aflatoxin B1 aldehyde reductase. *Eur J Biochem* **1999**, *265* (3), 1056-1060. DOI: 10.1046/j.1432-1327.1999.00826.x

(70) Kaufman, E. E.; Relkin, N.; Nelson, T. Regulation and properties of an NADP+ oxidoreductase which functions as a gamma-hydroxybutyrate dehydrogenase. *J Neurochem* **1983**, *40* (6), 1639-1646. DOI: 10.1111/j.1471-4159.1983.tb08137.x

(71) Orešič, M.; Hyötyläinen, T.; Herukka, S. K.; Sysi-Aho, M.; Mattila, I.; Seppänan-Laakso, T.; Julkunen, V.; Gopalacharyulu, P. V.; Hallikainen, M.; Koikkalainen, J.; et al. Metabolome in progression to Alzheimer's disease. *Transl Psychiatry* **2011**, *1* (12), e57. DOI: 10.1038/tp.2011.55

(72) Zhang, G. F.; Kombu, R. S.; Kasumov, T.; Han, Y.; Sadhukhan, S.; Zhang, J.; Sayre, L. M.; Ray, D.; Gibson, K. M.; Anderson, V. A.; et al. Catabolism of 4-hydroxyacids and 4-hydroxynonenal via 4-hydroxy-4-phosphoacyl-CoAs. *J Biol Chem* **2009**, *284* (48), 33521-33534. DOI: 10.1074/jbc.M109.055665

(73) Zhang, G. F.; Sadhukhan, S.; Ibarra, R. A.; Lauden, S. M.; Chuang, C. Y.; Sushailo, S.; Chatterjee, P.; Anderson, V. E.; Tochtrop, G. P.; Brunengraber, H. Metabolism of γ-hydroxybutyrate in perfused rat livers. *Biochem J* **2012**, *444* (2), 333-341. DOI: 10.1042/bj20112046

(74) Struys, E. A.; Verhoeven, N. M.; Jansen, E. E.; Ten Brink, H. J.; Gupta, M.; Burlingame, T. G.; Quang, L. S.; Maher, T.; Rinaldo, P.; Snead, O. C.; et al. Metabolism of gamma-hydroxybutyrate to d-2-hydroxyglutarate in mammals: further evidence for d-2-hydroxyglutarate transhydrogenase. *Metabolism* **2006**, *55* (3), 353-358. DOI: 10.1016/j.metabol.2005.09.009

(75) Shinka, T.; Inoue, Y.; Ohse, M.; Ito, A.; Ohfu, M.; Hirose, S.; Kuhara, T. Rapid and sensitive detection of urinary 4-hydroxybutyric acid and its related compounds by gas chromatography-mass spectrometry in a patient with succinic semialdehyde dehydrogenase deficiency. *J Chromatogr B Analyt Technol Biomed Life Sci* **2002**, *776* (1), 57-63. DOI: 10.1016/s1570-0232(02)00126-5

(76) Kardon, T.; Noël, G.; Vertommen, D.; Schaftingen, E. V. Identification of the gene encoding hydroxyacid-oxoacid transhydrogenase, an enzyme that metabolizes 4-hydroxybutyrate. *FEBS Lett* **2006**, *580* (9), 2347-2350. DOI: 10.1016/j.febslet.2006.02.082

(77) Fan, J.; Teng, X.; Liu, L.; Mattaini, K. R.; Looper, R. E.; Vander Heiden, M. G.; Rabinowitz, J. D. Human phosphoglycerate dehydrogenase produces the oncometabolite D-2-hydroxyglutarate. *ACS Chem Biol* **2015**, *10* (2), 510-516. DOI: 10.1021/cb500683c

(78) Chalmers, R. A.; Lawson, A. M.; Watts, R. W.; Tavill, A. S.; Kamerling, J. P.; Hey, E.; Ogilvie, D. D-2-hydroxyglutaric aciduria: case report and biochemical studies. *J Inherit Metab Dis* **1980**, *3* (1), 11-15. DOI: 10.1007/bf02312516

(79) Lindahl, G.; Lindstedt, G.; Lindstedt, S. Metabolism of 2-amino-5-hydroxyadipic acid in the rat. *Arch Biochem Biophys* **1967**, *119* (1), 347-352. DOI: 10.1016/0003-9861(67)90463-8

(80) Kaufman, E. E.; Nelson, T.; Fales, H. M.; Levin, D. M. Isolation and characterization of a hydroxyacid-oxoacid transhydrogenase from rat kidney mitochondria. *J Biol Chem* **1988**, *263* (32), 16872-16879.

(81) Achouri, Y.; Noël, G.; Vertommen, D.; Rider, M. H.; Veiga-Da-Cunha, M.; Van Schaftingen, E. Identification of a dehydrogenase acting on D-2-hydroxyglutarate. *Biochem J* **2004**, *381* (Pt 1), 35-42. DOI: 10.1042/bj20031933

(82) Ježek, P. 2-Hydroxyglutarate in Cancer Cells. *Antioxid Redox Signal* **2020**, *33* (13), 903-926. DOI: 10.1089/ars.2019.7902

(83) Rzem, R.; Vincent, M. F.; Van Schaftingen, E.; Veiga-da-Cunha, M. L-2-hydroxyglutaric aciduria, a defect of metabolite repair. *J Inherit Metab Dis* **2007**, *30* (5), 681-689. DOI: 10.1007/s10545-007-0487-0

(84) Jackson, B.; Brocker, C.; Thompson, D. C.; Black, W.; Vasiliou, K.; Nebert, D. W.; Vasiliou, V. Update on the aldehyde dehydrogenase gene (ALDH) superfamily. *Hum Genomics* **2011**, *5* (4), 283-303. DOI: 10.1186/1479-7364-5-4-283

(85) Chambliss, K. L.; Gibson, K. M. Succinic semialdehyde dehydrogenase from mammalian brain: subunit analysis using polyclonal antiserum. *Int J Biochem* **1992**, *24* (9), 1493-1499. DOI: 10.1016/0020-711x(92)90077-e

(86) Kang, J. H.; Park, Y. B.; Huh, T. L.; Lee, W. H.; Choi, M. S.; Kwon, O. S. High-level expression and characterization of the recombinant enzyme, and tissue distribution of human succinic semialdehyde dehydrogenase. *Protein Expr Purif* **2005**, *44* (1), 16-22. DOI: 10.1016/j.pep.2005.03.019

(87) Murphy, T. C.; Amarnath, V.; Gibson, K. M.; Picklo, M. J., Sr. Oxidation of 4-hydroxy-2-nonenal by succinic semialdehyde dehydrogenase (ALDH5A). *J Neurochem* **2003**, *86* (2), 298-305. DOI: 10.1046/j.1471-4159.2003.01839.x

(88) Rivett, A. J.; Tipton, K. F. Kinetic studies with rat-brain succinic-semialdehyde dehydrogenase. *Eur J Biochem* **1981**, *117* (1), 187-193. DOI: 10.1111/j.1432-1033.1981.tb06319.x

(89) Rothacker, B.; Ilg, T. Functional characterization of a Drosophila melanogaster succinic semialdehyde dehydrogenase and a non-specific aldehyde dehydrogenase. *Insect Biochem Mol Biol* **2008**, *38* (3), 354-366. DOI: 10.1016/j.ibmb.2007.12.004

(90) Rothacker, B.; Werr, M.; Ilg, T. Molecular cloning, partial genomic structure and functional characterization of succinic semialdehyde dehydrogenase genes from the parasitic insects Lucilia cuprina and Ctenocephalides felis. *Insect Mol Biol* **2008**, *17* (3), 279-291. DOI: 10.1111/j.1365-2583.2008.00800.x

(91) Rothacker, B.; Werr, M.; Ilg, T. Succinic semialdehyde dehydrogenase from the parasitic cattle tick Rhipicephalus microplus: gene identification, biochemical characterization and

comparison with the mouse ortholog. *Mol Biochem Parasitol* **2008**, *161* (1), 32-43. DOI: 10.1016/j.molbiopara.2008.06.001

(92) Esser, D.; Kouril, T.; Talfournier, F.; Polkowska, J.; Schrader, T.; Bräsen, C.; Siebers, B. Unraveling the function of paralogs of the aldehyde dehydrogenase super family from Sulfolobus solfataricus. *Extremophiles* **2013**, *17* (2), 205-216. DOI: 10.1007/s00792-012-0507-3

(93) Ito, F.; Chishiki, H.; Fushinobu, S.; Wakagi, T. Archaeal aldehyde dehydrogenase ST0064 from Sulfolobus tokodaii, a paralog of non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase, is a succinate semialdehyde dehydrogenase. *Biosci Biotechnol Biochem* **2013**, *77* (6), 1344-1348. DOI: 10.1271/bbb.130119

(94) de Carvalho, L. P.; Ling, Y.; Shen, C.; Warren, J. D.; Rhee, K. Y. On the chemical mechanism of succinic semialdehyde dehydrogenase (GabD1) from Mycobacterium tuberculosis. *Arch Biochem Biophys* **2011**, *509* (1), 90-99. DOI: 10.1016/j.abb.2011.01.023

(95) Jaeger, M.; Rothacker, B.; Ilg, T. Saturation transfer difference NMR studies on substrates and inhibitors of succinic semialdehyde dehydrogenases. *Biochem Biophys Res Commun* **2008**, *372* (3), 400-406. DOI: 10.1016/j.bbrc.2008.04.183

(96) Langendorf, C. G.; Key, T. L.; Fenalti, G.; Kan, W. T.; Buckle, A. M.; Caradoc-Davies, T.; Tuck, K. L.; Law, R. H.; Whisstock, J. C. The X-ray crystal structure of Escherichia coli succinic semialdehyde dehydrogenase; structural insights into NADP+/enzyme interactions. *PLoS One* **2010**, *5* (2), e9280. DOI: 10.1371/journal.pone.0009280

(97) Park, J.; Rhee, S. Structural basis for a cofactor-dependent oxidation protection and catalysis of cyanobacterial succinic semialdehyde dehydrogenase. *J Biol Chem* **2013**, *288* (22), 15760-15770. DOI: 10.1074/jbc.M113.460428

(98) Park, S. A.; Park, Y. S.; Lee, K. S. Kinetic characterization and molecular modeling of NAD(P)(+)-dependent succinic semialdehyde dehydrogenase from Bacillus subtilis as an ortholog YneI. *J Microbiol Biotechnol* **2014**, *24* (7), 954-958. DOI: 10.4014/jmb.1402.02054

(99) Phonbuppha, J.; Maenpuen, S.; Munkajohnpong, P.; Chaiyen, P.; Tinikul, R. Selective determination of the catalytic cysteine pK(a) of two-cysteine succinic semialdehyde dehydrogenase from Acinetobacter baumannii using burst kinetics and enzyme adduct formation. *Febs j* **2018**, *285* (13), 2504-2519. DOI: 10.1111/febs.14497

(100) Cao, J.; Singh, N.; Locy, B. Characterization of the recombinant succinic semialdehyde dehydrogenase from Sacharomyces cerevisiae. *Yeast* **2014**, *31*. DOI: 10.1002/yea.3035

(101) Kumar, S.; Kumar, S.; Punekar, N. S. Characterization of succinic semialdehyde dehydrogenase from Aspergillus niger. *Indian J Exp Biol* **2015**, *53* (2), 67-74.

(102) Busch, K. B.; Fromm, H. Plant Succinic Semialdehyde Dehydrogenase. Cloning, Purification, Localization in Mitochondria, and Regulation by Adenine Nucleotides1. *Plant Physiology* **1999**, *121* (2), 589-598. DOI: 10.1104/pp.121.2.589

(103) Jakoby, W. B.; Scott, E. M. Aldehyde oxidation. III. Succinic semialdehyde dehydrogenase. *J Biol Chem* **1959**, *234* (4), 937-940.

(104) Cornish-Bowden, A. Current IUBMB recommendations on enzyme nomenclature and kinetics. *Perspectives in Science* **2014**, *1* (1), 74-87. DOI: 10.1016/j.pisc.2014.02.006

(105) International Union of Biochemistry and Molecular Biology. Nomenclature Committee.; Webb, E. C. *Enzyme nomenclature 1992 : recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*; Published for the International Union of Biochemistry and Molecular Biology by Academic Press, 1992.

(106) McDonald, A. G.; Boyce, S.; Tipton, K. F. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res* **2009**, *37* (Database issue), D593-597. DOI: 10.1093/nar/gkn582

(107) Yuan, Z.; Yin, B.; Wei, D.; Yuan, Y. R. Structural basis for cofactor and substrate selection by cyanobacterium succinic semialdehyde dehydrogenase. *J Struct Biol* **2013**, *182* (2), 125-135. DOI: 10.1016/j.jsb.2013.03.001

(108) Zheng, H.; Beliavsky, A.; Tchigvintsev, A.; Brunzelle, J. S.; Brown, G.; Flick, R.; Evdokimova, E.; Wawrzak, Z.; Mahadevan, R.; Anderson, W. F.; et al. Structure and activity of the NAD(P)+-dependent succinate semialdehyde dehydrogenase YneI from Salmonella typhimurium. *Proteins* **2013**, *81* (6), 1031-1041. DOI: 10.1002/prot.24227

(109) Dover, S.; Halpern, Y. S. Utilization of γ-aminobutyric acid as the sole carbon and nitrogen source by Escherichia coli K-12 mutants. *J Bacteriol* **1972**, *109* (2), 835-843. DOI: 10.1128/jb.109.2.835-843.1972

(110) Cozzani, I.; Fazio, A. M.; Felici, E.; Barletta, G. Separation and characterization of NAD- and NADP-specific succinate-semialdehyde dehydrogenase from Escherichia coli K-12 3300. *Biochim Biophys Acta* **1980**, *613* (2), 309-317. DOI: 10.1016/0005-2744(80)90085-6

(111) Skinner, M. A.; Cooper, R. A. An Escherichia coli mutant defective in the NAD-dependent succinate semialdehyde dehydrogenase. *Arch Microbiol* **1982**, *132* (3), 270-275. DOI: 10.1007/bf00407964

(112) Marek, L. E.; Henson, J. M. Cloning and expression of the Escherichia coli K-12 sad gene. *J Bacteriol* **1988**, *170* (2), 991-994. DOI: 10.1128/jb.170.2.991-994.1988

(113) Fuhrer, T.; Chen, L.; Sauer, U.; Vitkup, D. Computational prediction and experimental verification of the gene encoding the NAD+/NADP+-dependent succinate semialdehyde dehydrogenase in Escherichia coli. *J Bacteriol* **2007**, *189* (22), 8073-8078. DOI: 10.1128/jb.01027-07

(114) Cremers, C. M.; Jakob, U. Oxidant sensing by reversible disulfide bond formation. *J Biol Chem* **2013**, *288* (37), 26489-26496. DOI: 10.1074/jbc.R113.462929

(115) Marchal, S.; Rahuel-Clermont, S.; Branlant, G. Role of glutamate-268 in the catalytic mechanism of nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase from Streptococcus mutans. *Biochemistry* **2000**, *39* (12), 3327-3335. DOI: 10.1021/bi9914208

(116) Muñoz-Clares, R. A.; González-Segura, L.; Riveros-Rosas, H.; Julián-Sánchez, A. Amino acid residues that affect the basicity of the catalytic glutamate of the hydrolytic aldehyde dehydrogenases. *Chem Biol Interact* **2015**, *234*, 45-58. DOI: 10.1016/j.cbi.2015.01.019

(117) Donnelly, M. I.; Cooper, R. A. Succinic semialdehyde dehydrogenases of Escherichia coli: their role in the degradation of p-hydroxyphenylacetate and gamma-aminobutyrate. *Eur J Biochem* **1981**, *113* (3), 555-561. DOI: 10.1111/j.1432-1033.1981.tb05098.x

(118) Satya Narayan, V.; Nair, P. M. Potato tuber succinate semialdehyde dehydrogenase: purification and characterization. *Arch Biochem Biophys* **1989**, *275* (2), 469-477. DOI: 10.1016/0003-9861(89)90393-7

(119) Pearl, P. L.; Novotny, E. J.; Acosta, M. T.; Jakobs, C.; Gibson, K. M. Succinic semialdehyde dehydrogenase deficiency in children and adults. *Ann Neurol* **2003**, *54 Suppl 6*, S73-80. DOI: 10.1002/ana.10629

(120) Jakobs, C.; Bojasch, M.; Mönch, E.; Rating, D.; Siemes, H.; Hanefeld, F. Urinary excretion of gamma-hydroxybutyric acid in a patient with neurological abnormalities. The probability of a new inborn error of metabolism. *Clin Chim Acta* **1981**, *111* (2-3), 169-178. DOI: 10.1016/0009-8981(81)90184-4

(121) Trettel, F.; Malaspina, P.; Jodice, C.; Novelletto, A.; Slaughter, C. A.; Caudle, D. L.; Hinson, D. D.; Chambliss, K. L.; Gibson, K. M. Human succinic semialdehyde dehydrogenase. Molecular cloning and chromosomal localization. *Adv Exp Med Biol* **1997**, *414*, 253-260.

(122) Chambliss, K. L.; Hinson, D. D.; Trettel, F.; Malaspina, P.; Novelletto, A.; Jakobs, C.; Gibson, K. M. Two exon-skipping mutations as the molecular basis of succinic semialdehyde dehydrogenase deficiency (4-hydroxybutyric aciduria). *Am J Hum Genet* **1998**, *63* (2), 399-408. DOI: 10.1086/301964

(123) Malaspina, P.; Roullet, J. B.; Pearl, P. L.; Ainslie, G. R.; Vogel, K. R.; Gibson, K. M. Succinic semialdehyde dehydrogenase deficiency (SSADHD): Pathophysiological complexity and multifactorial trait associations in a rare monogenic disorder of GABA metabolism. *Neurochem Int* **2016**, *99*, 72-84. DOI: 10.1016/j.neuint.2016.06.009

(124) Didiášová, M.; Banning, A.; Brennenstuhl, H.; Jung-Klawitter, S.; Cinquemani, C.; Opladen, T.; Tikkanen, R. Succinic Semialdehyde Dehydrogenase Deficiency: An Update. *Cells* **2020**, *9* (2). DOI: 10.3390/cells9020477

(125) Kaufman, E. E.; Nelson, T.; Miller, D.; Stadlan, N. Oxidation of gamma-hydroxybutyrate to succinic semialdehyde by a mitochondrial pyridine nucleotide-independent enzyme. *J Neurochem* **1988**, *51* (4), 1079-1084. DOI: 10.1111/j.1471-4159.1988.tb03071.x

(126) Allen, S. H. G.; Patil, J. R. Studies on the Structure and Mechanism of Action of the Malate-Lactate Transhydrogenase. *Journal of Biological Chemistry* **1972**, *247* (3), 909-916. DOI: 10.1016/S0021-9258(19)45693-7

(127) Piersma, S. R.; de Vries, S.; Duine, J. A. Nicotinoprotein alcohol/aldehyde oxidoreductases. Enzymes with bound NAD(P) as cofactor. *Adv Exp Med Biol* **1997**, *414*, 425-434.

(128) Deng, Y.; Wang, Z.; Gu, S.; Ji, C.; Ying, K.; Xie, Y.; Mao, Y. Cloning and characterization of a novel human alcohol dehydrogenase gene (ADHFe1). *DNA Seq* **2002**, *13* (5), 301-306. DOI: 10.1080/1042517021000011636

(129) Elleuche, S.; Antranikian, G. Bacterial group III alcohol dehydrogenases - Function, evolution and biotechnological applications. *OA Alcohol* **2013**, *1*, 3. DOI: 10.13172/2053-0285-1-1-489

(130) Reid, M. F.; Fewson, C. A. Molecular characterization of microbial alcohol dehydrogenases. *Crit Rev Microbiol* **1994**, *20* (1), 13-56. DOI: 10.3109/10408419409113545

(131) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **1990**, *215* (3), 403-410. DOI: 10.1016/s0022-2836(05)80360-2

(132) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **1997**, *25* (17), 3389-3402. DOI: 10.1093/nar/25.17.3389

(133) Chen, Q.; Wu, Q.; Peng, Y. ADHFE1 is a correlative factor of patient survival in cancer. *Open Life Sci* **2021**, *16* (1), 571-582. DOI: 10.1515/biol-2021-0065

(134) Tae, C. H.; Ryu, K. J.; Kim, S. H.; Kim, H. C.; Chun, H. K.; Min, B. H.; Chang, D. K.; Rhee, P. L.; Kim, J. J.; Rhee, J. C.; et al. Alcohol dehydrogenase, iron containing, 1 promoter hypermethylation associated with colorectal cancer differentiation. *BMC Cancer* **2013**, *13*, 142. DOI: 10.1186/1471-2407-13-142

(135) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023**, *51* (D1), D523-d531. DOI: 10.1093/nar/gkac1052

(136) Kim, J. Y.; Tillison, K. S.; Zhou, S.; Lee, J. H.; Smas, C. M. Differentiation-dependent expression of Adhfe1 in adipogenesis. *Arch Biochem Biophys* **2007**, *464* (1), 100-111. DOI: 10.1016/j.abb.2007.04.018

(137) Käll, L.; Krogh, A.; Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **2004**, *338* (5), 1027-1036. DOI: 10.1016/j.jmb.2004.03.016

(138) Käll, L.; Krogh, A.; Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W429-432. DOI: 10.1093/nar/gkm256

(139) Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* **2017**, *7* (1), 10480. DOI: 10.1038/s41598-017-09654-8

(140) Bienert, S.; Waterhouse, A.; de Beer, T. A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* **2017**, *45* (D1), D313-d319. DOI: 10.1093/nar/gkw1132

(141) Guex, N.; Peitsch, M. C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **2009**, *30 Suppl 1*, S162-173. DOI: 10.1002/elps.200900140

(142) Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T. QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics* **2020**, *36* (6), 1765-1771. DOI: 10.1093/bioinformatics/btz828

(143) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **2018**, *46* (W1), W296-w303. DOI: 10.1093/nar/gky427

(144) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2

(145) Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **2024**, *52* (D1), D368-d375. DOI: 10.1093/nar/gkad1011

(146) Marçal, D.; Rêgo, A. T.; Carrondo, M. A.; Enguita, F. J. 1,3-Propanediol dehydrogenase from Klebsiella pneumoniae: decameric quaternary structure and possible subunit cooperativity. *J Bacteriol* **2009**, *191* (4), 1143-1151. DOI: 10.1128/jb.01077-08

Chapter 2

The Enzyme Kinetics of Recombinant *E. coli* SSADH

Abstract

Recombinant *E. coli* $NADP^+$-dependent SSADH (*gabD*) containing an N-terminal His-tag was expressed in BL21(DE3)pLysS cells and then purified using immobilized metal affinity and size exclusion chromatography. Characterization of the pure enzyme by SDS-PAGE, ESI and MALDI-TOF MS revealed a homogeneous sample with an apparent molecular weight of 54 kDa for the monomer. The specific activity of the enzyme obtained from this purification (34.74 U/mg) appears to be approximately 4-fold higher than what has been previously reported. Following optimization of the steady state kinetics assay, SSADH initial velocities were measured while varying the concentrations of the SSA substrate and $NADP^+$ coenzyme and analysis of these rate profiles revealed a clearly-defined sequential mechanism. Global least-squares regression of the data yielded 74.1 ± 1.2 s^{-1}, 10.6 ± 0.7 μM, and 0.13 ± 0.02 mM for the kinetic constants k_{cat}, K_m(SSA), and K_m($NADP^+$), respectively. Analysis of the reaction kinetics as a function of pH revealed that the enzyme requires a group with a pK_a of 7.03 ± 0.01 to be deprotonated for catalytic activity and displays a pH optimum of ≥ 8.25. Additionally, *E. coli* SSADH demonstrated a strong preference for $NADP^+$ over NAD^+, with an 83% decrease in activity when NAD^+ was used as the coenzyme.

2.1 Introduction

2.1.1 Succinate Semialdehyde Dehydrogenase

Succinate semialdehyde dehydrogenase (SSADH) is a ubiquitous enzyme found in nearly every kingdom of life, including Animalia, Plantae, fungi, Archaea and Bacteria. In mammals, SSADH is the last enzyme in the GABA shunt, but in bacteria it also functions in the degradation of p-hydroxyphenylacetate.[1, 2] In humans, succinate semialdehyde dehydrogenase deficiency is a genetic disorder that impacts the degradation of GHB, resulting in a number of debilitating neurological conditions.

2.1.2 Literature on *E. coli* $NADP^+$-dependent SSADH

Despite belonging to the large ALDH superfamily, SSADHs are poorly studied dehydrogenases. To complicate matters, various attempts to classify SSADHs based on their cofactor preference and amino acid sequence have led to establishment of three different Enzyme Commission numbers (1.2.1.16, 1.2.1.24, and 1.2.1.79), attributed to at least two different genes: *gabD* and *yneI* (also known as *sad*). *gabD* encodes an $NADP^+$-dependent SSADH, while *yneI* encodes a $NAD(P)^+$-dependent SSADH. The literature contains only a few enzyme kinetics studies with little consensus on possible kinetic mechanisms and values for kinetic constants.

To date, no rigorous enzyme kinetics have been performed on *E. coli* SSADH from *gabD* or *yneI*, or have mechanisms been proposed for them. There are only three publications on *E. coli* $NADP^+$-dependent SSADH (*gabD*) with disparate substrate and coenzyme K_m values[3-5], two of which also have questionable estimations for the pH optimum[3, 4]. The remainder of the

literature on *E. coli* SSADH is limited to expression[2] and computational prediction[6] of the Yne1 gene product, the role of SSADHs when *E. coli* is grown only on γ-aminobutyrate[1, 7], and a single publication on X-ray structure[5] of the *gabD* gene product.

2.1.3 Present Research on *E. coli* $NADP^+$-dependent SSADH

The aim of the present research was to determine the kinetic constants for *E. coli* $NADP^+$-dependent SSADH, hereon referred to simply as SSADH, before employing it as a reporter enzyme in a coupled assay that would ultimately be used to measure the activity of hydroxyacid-oxoacid transhydrogenase (HOT). The results of this investigation would also be used to contribute towards understanding enzyme mechanism for SSADH. To accomplish these aims, recombinant SSADH would be expressed and purified to a homogeneity, relying on methods previously published.[5] Following purification, a rigorous enzyme kinetics study would be undertaken in order to provide the k_{cat} for SSADH, the K_m for both the substrate and coenzyme, the degree of preference for $NADP^+$ *versus* NAD^+, and a pH-rate profile that would reveal a pH optimum and the pK_a of any ionizable group on which the catalysis depends.

2.1.4 Protein Purification

In order to accurately determine the kinetic constants k_{cat} and K_m for an enzymatic reaction, the enzyme must be isolated from the cellular milieu and purified to a substantial degree, all while maintaining its activity. Steady-state enzyme kinetic measurements produce the parameter V_{max}, and $k_{cat} = V_{max}/[E]_T$, where $[E]_T$ is the total enzyme concentration used in the assay. Determining the exact concentration of enzyme used in the assay is trivial when the sample is pure. Additionally, cell extract contains other molecules that could potentially interfere with kinetics measurements. Other enzymes present might consume common

substrates and cofactors, while other small molecules could serve as inhibitors. In short, carefully controlled rate experiments are only possible when every component is pure. The following purification protocol is adapted from the X-ray paper published by the Whisstock Lab, who generously donated the expression vector.[5]

Immobilized Metal Affinity Chromatography

Because the recombinant SSADH contains an N-terminal 6×His-tag, purification by Immobilized Metal Affinity Chromatography (IMAC) was the logical first step in the overall purification scheme. The principle of IMAC involves the lone pair electrons on the imidazole sidechains of the histidine tag forming coordinate bonds to nickel (II) cations that have been immobilized onto a column matrix though nitriloacetic acid (NTA) linkers. Other metals, including cobalt, copper and zinc may also be used, as well as other means of linking the metal to the column matrix. Ethylenediaminetetraacetic acid (EDTA), a well-known metal chelator, is sometimes included in cell lysis buffers to disrupt membranes and inhibit metalloproteases. However, caution should be exercised to ensure that EDTA is not included in IMAC column buffers because it may remove Ni^{2+} or other metals from the resin, decreasing the capacity of the column for binding 6×His-tagged proteins.

Given that one of the catalytic residues in the active site of SSADH is cysteine, and oxidation of this residue eliminates activity, maintaining a reducing environment during purification is essential for obtaining pure protein with intact enzymatic activity. Some formulations of IMAC resins report a tolerance of the nickel to a limited concentration of reductants during purification, while others discourage their use entirely. The manual for the ThermoFisher HisPur™ Ni-NTA resin used in this purification specifically advises against the use of strong reducing agents, such as β-mercaptoethanol (BME) and dithiothreitol (DTT), as

they will disrupt the function of the nickel resin. Apart from this caution, there is no accompanying chemical explanation for what is occurring or how it might be mitigated.

For the present purification, the relevant interactions are the 6×His···Ni^{2+}-NTA coordination bonds. As long as these coordination bonds exist, the recombinantly 6×His-tagged protein will remain bound to the column while unwanted molecules flow through and exit the column in the eluate. Progress may be followed by monitoring the eluate absorbance at 280 nm after loading the lysate onto the column and applying a wash buffer. Once the absorbance has reached a stable minimum value, it can be assumed that only nickel-binding proteins remain on the column, ideally with the majority being the desired 6×His-tagged recombinant protein. Elution of the desired protein from the Ni^{2+}-NTA column is accomplished by increasing the concentration of free imidazole in the eluent flowing onto the column, either through stepwise increases or by gradient. The presence of free imidazole sets up a competition for Ni^{2+}-NTA coordination sites between the free imidazole in solution and the imidazole rings on the 6×His-tag. At sufficiently high free imidazole concentration the 6×His-tagged protein will be displaced from Ni^{2+}-NTA and flow out of the column to be collected for analysis and further purification.

Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) served as the final polishing step in the purification of 6×His-SSADH, accomplishing the following: i) removal of any remaining contaminants that were sufficiently different in molecular weight from 6×His-SSADH; ii) provided an estimate for the oligomeric state of 6×His-SSADH in solution; and iii) facilitated a complete buffer exchange into the final storage buffer. In SEC, the stationary phase (resin) is composed of small porous gel beads, around and through which the aqueous mobile phase

travels. Beads with a fixed pore size are selected so that only a limited range of molecules in the mobile phase are permitted to enter, with all others being excluded. For two different molecules of sufficiently small size to enter the pores, the smaller of the two will spend more time trapped in the pores and a partitioning between these two different sized molecules occurs. The equilibrium distribution between varying sized particles and their time spent in the pores forms the basis of this separation technique and can be mathematically represented by a partition coefficient. The net result of this separation technique is pure 6×His-SSADH eluting in a completely different elution volume (fraction) from any higher and lower molecular weight contaminants that were initially present. Various protein standards and dyes of known molecular weight can be used to estimate column parameters and efficiency, as well as provide the data required to construct a molecular weight standard curve using the partition coefficient. The standard curve allows the convenient estimate for the molecular weight of the molecules contained in each volume (i.e., 1 mL fraction) that elutes from the column. Since buffer components have the smallest molecular weight in the protein sample, SEC also affords the opportunity to partition 6×His-SSADH into a new buffer optimized for long term storage.

2.1.5 Enzyme Kinetics

Initial velocity measurements are a powerful tool for determining the kinetic mechanism of enzyme catalyzed reactions and their associated kinetic constants k_{cat} and K_m. At the heart of these measurements is the Henri-Michaelis-Menten equation, from which these constants originate. Although the Henri-Michaelis-Menten equation was derived with *single-substrate* reactions in mind, it can be readily applied to *multi-substrate* reactions when all but one of the substrate concentrations are held constant. In this situation, multi-substrate reactions behave as single-substrate reactions. The *rapid-equilibrium* and *steady-state* assumptions are two

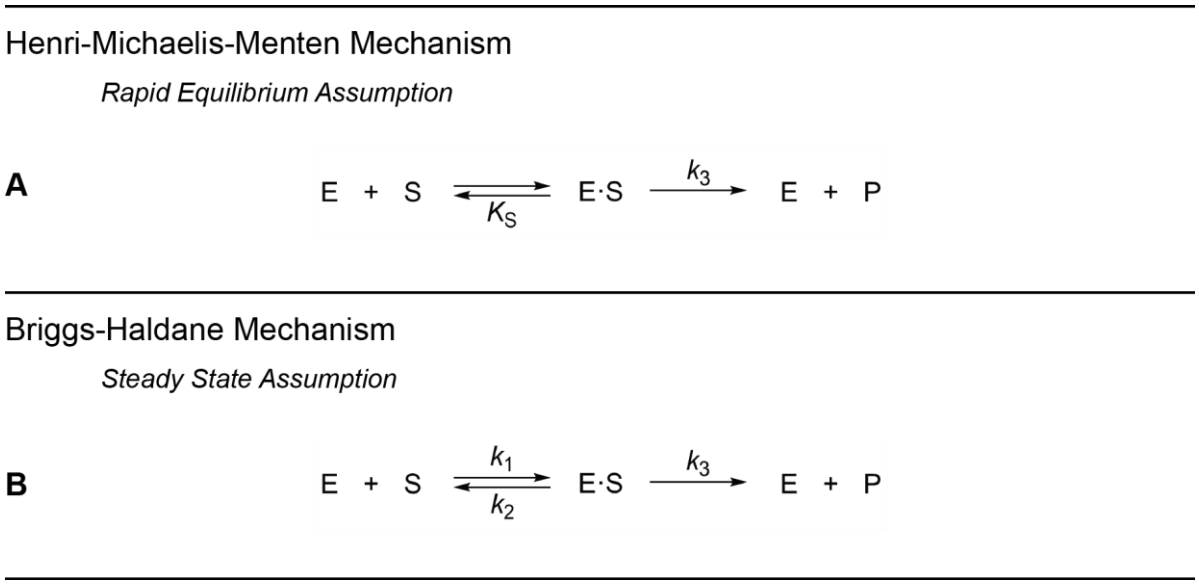

Figure 2.1. Fundamental single-substrate enzyme kinetics mechanisms. (A) The Henri-Michaelis-Menten mechanism, which assumes the enzyme-substrate complex is in thermodynamic equilibrium with the free enzyme and substrate. **(B)** The Briggs-Haldane mechanism, which assumes the concentration of the enzyme-substrate complex is approximately constant over the reaction time.

classical approaches for deriving the Henri-Michaelis-Menten equation. While these two approaches have minimal impact on the final form of the single-substrate equation, they do produce significant differences between some rate equations for multi-substrate reactions, and these differences can prove useful when determining the possible enzyme mechanism.

The Henri-Michaelis-Menten Equation

Two different assumptions set apart what is often seen at the surface as the same equation. Michaelis and Menten initially proposed *rapid-equilibrium assumption* (**Figure 2.1A**), after the work of Henri.[8] This assumption states that an equilibrium is rapidly established between the ES-complex and the free enzyme and substrate, governed by the equilibrium dissociation constant K_S. In this mechanism, $K_S = k_{off}/k_{on} = k_2/k_1$. It is also assumed that $k_2 >> k_3$, where k_3 is the irreversible catalytic step, sometimes referred to as k_{cat}.

The main consequences to the rapid equilibrium derivation of the Henri-Michaelis-Menten equation (**Equation 2.1**) are that K_S here is defined strictly as an equilibrium dissociation

Equation 2.1. The Henri-Michaelis-Menten equation (rapid equilibrium).

Equation	Definitions	Remarks
$v_0 = \frac{V_{max}[S]}{K_S + [S]} = \frac{k_{cat}[E]_T[S]}{K_S + [S]}$	v_0 = initial velocity V_{max} = maximum velocity under saturating [S] [S] = substrate concentration K_S = dissociation constant for [ES] complex $k_{cat} = k_3$ = slow, irreversible catalytic step $[E]_T$ = total enzyme concentration	$K_S = \frac{[ES]}{[E]_T[S]} = \frac{k_2}{k_1}$ $k_2 \gg k_3$

constant equal to k_2/k_1, and the catalytic rate constant k_3 is taken to be much slower than k_2, the rate constant for ES-complex dissociation. The stipulation that $k_3 << k_2$ imposes a critical limitation that fails to account for very efficient enzymes, such as carbonic anhydrase and catalase, which catalyze reactions at rates near the diffusion-controlled limit for the encounter between free enzyme and free substrate. Additionally, it will be shown that deriving a rate equation for a *bireactant sequential mechanism* by applying the rapid equilibrium assumption results in one less term in the denominator than when the rate equation is derived using the steady-state assumption. This difference becomes graphically evident when the data is presented in a Lineweaver-Burk plot (double-reciprocal units).

Briggs and Haldane soon after proposed a derivation that did not require a rapid equilibrium to establish the ES-complex, or that $k_2 >> k_3$.[9] They devised the *steady-state assumption* (**Figure 2.1B**) where, after an initial build-up, the ES-complex concentration remains approximately constant until the substrate concentration becomes significantly depleted. The forward rate of ES formation is perfectly balanced by its loss. Because the initial velocity of the catalyzed reaction was equal to the rate constant multiplied by the concentration of the catalytically species, here being ES-complex, then the initial velocity would also be constant during this steady state ($v_0 = k_3[ES]$).

The main consequences of the state-state approximation used by Briggs and Haldane to formulate their version of the equation (**Equation 2.2**) is that the term K_S was replaced with

Equation 2.2. The Briggs-Haldane equation (steady state)

Equation	Definition	Remarks
$v_0 = \frac{V_{max}[S]}{K_m + [S]} = \frac{k_{cat}[E]_T[S]}{K_m + [S]}$	v_0 = initial velocity V_{max} = maximum velocity under saturating [S] [S] = substrate concentration K_m = Michaelis constant $k_{cat} = k_3$ = irreversible catalytic step $[E]_T$ = total enzyme concentration	$K_m = \frac{k_3 + k_2}{k_1}$ $\frac{d[ES]}{dt} = 0$ $v_0 = k_3[ES]$

K_m (for the Michaelis constant) and was now equal to $(k_3 + k_2)/k_1$. K_m was no longer an equilibrium constant, but a composite of individual rate constants that would be much more challenging to interpret as the mechanisms became more complex. The fact that k_3 was no longer limited to being much less than k_2 also has consequences. If k_3 should be much greater than k_2, then specificity constant k_{cat}/K_m (equivalent to k_3/K_m) would default to k_1, the rate constant for the association of free enzyme and free substrate. This allows the steady-state approximation to account for extremely efficient enzymes, such as carbonic anhydrase and catalase, which operate near the diffusion limit of 10^8 s^{-1} M^{-1}. Enzymes with a k_{cat}/K_m near this near the diffusion limit are probably operating by a steady-state mechanism. As discussed with the rapid equilibrium assumption, rate equations for *bireactant sequential mechanisms* derived with the steady-state approximation have an additional term in the denominator that makes them easy to spot when transformed into Lineweaver-Burk plots (double-reciprocal units).

Key takeaways are that K_s is an equilibrium dissociation constant (rapid-equilibrium assumption), while K_m is a composite of individual microscopic rate constants (steady-state assumption). When deriving bireactant rate equations the assumptions made give rise to differences in the mathematical expressions for the rate equations, which are readily apparent when transformed in double-reciprocal plots. Enzymes with k_{cat}/K_m values that appear diffusion controlled are probably operating by a steady-state mechanism.

Bireactant Enzyme Mechanisms

SSADH is a bireactant enzyme, since both the SSA substrate and the $NADP^+$ coenzyme behave as substrates. A few *relevant* bireactant mechanisms are illustrated in **Figure 2.2**.

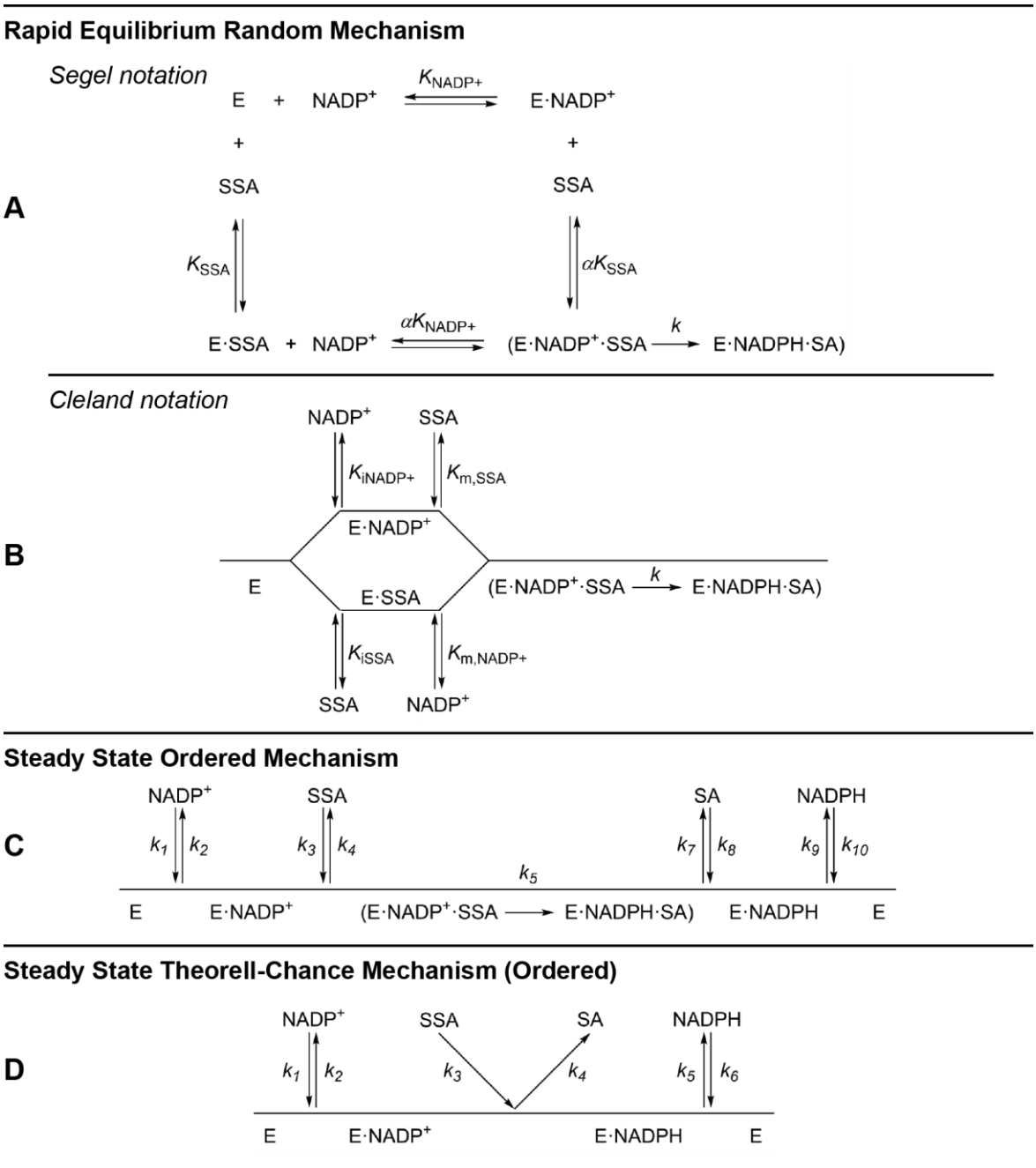

Figure 2.2. Relevant bireactant sequential enzyme mechanisms. Central-ternary complexes, if present, are denoted within parentheses. **(A)** rapid equilibrium random mechanism with Segel notation; **(B)** rapid equilibrium random mechanism with Cleland notation; **(C)** steady state ordered mechanism; and **(D)** steady state Theorell-Chance mechanism.

Equation 2.3. RE Random (Segel).

$$v_0 = \frac{V_{max}[\mathrm{NADP^+}][\mathrm{SSA}]}{\alpha K_{\mathrm{NADP+}} K_{\mathrm{SSA}} + \alpha K_{\mathrm{SSA}}[\mathrm{NADP^+}] + \alpha K_{\mathrm{NADP+}}[\mathrm{SSA}] + [\mathrm{NADP^+}][\mathrm{SSA}]}$$

Equation 2.4. Relationship between *K* constants in Equation 2.3.

$$\frac{\alpha K_{\mathrm{NADP+}}}{\alpha K_{\mathrm{SSA}}} = \frac{K_{\mathrm{NADP+}}}{K_{\mathrm{SSA}}} \quad \text{and} \quad \alpha K_{\mathrm{NADP+}} K_{\mathrm{SSA}} = \alpha K_{\mathrm{SSA}} K_{\mathrm{NADP+}}$$

The rate equation for the rapid equilibrium random (RE Random) mechanism shown in **Figure 2.2A** is given by **Equation 2.3**, with the relationship between the constants given by **Equation 2.4**. In these equations Segel's derivation makes use of an α-term that describes how the binding of one substrate may influence the binding of the other.[10]

A similar rate equation for the rapid equilibrium random mechanism (**Figure 2.2B**) was derived by Cleland using slightly different notation (**Equations 2.5 and 2.6**).[11] What sets the notation in Cleland's derivation apart is that the form of this rapid equilibrium random rate equation is mathematically identical to the rate equations for the steady-state ordered (SS Ordered) and steady-state Theorell-Chance (SS Theorell-Chance) mechanisms in **Figures 2.2C and 2.26D**. The explanation lies in the fact that in the rapid equilibrium random mechanism, the addition of the first substrate to the free enzyme is in rapid equilibrium, where k_{off} for the first substrate is greater than k_{cat}. However, the second substrate to add and form the

Equation 2.5. RE Random (Cleland), SS Ordered, SS Theorell-Chance.

$$v_0 = \frac{V_{max}[\mathrm{NADP^+}][\mathrm{SSA}]}{K_{\mathrm{iNADP+}} K_{\mathrm{m,SSA}} + K_{\mathrm{m,SSA}}[\mathrm{NADP^+}] + K_{\mathrm{m,NADP+}}[\mathrm{SSA}] + [\mathrm{NADP^+}][\mathrm{SSA}]}$$

Equation 2.6. Relationship between *K* constants in Equation 2.5.

$$\frac{K_{\mathrm{iNADP+}}}{K_{\mathrm{iSSA}}} = \frac{K_{\mathrm{m,SSA}}}{K_{\mathrm{m,SSA}}} \quad \text{and} \quad K_{\mathrm{iNADP+}} K_{\mathrm{m,SSA}} = K_{\mathrm{m,NADP+}} K_{\mathrm{iSSA}}$$

final ternary complex may do so either in rapid equilibrium or steady state without changing the form of the rate equation.[11] There is no stipulation that k_{off} for the second substrates departure from the ternary complex be slower than k_{cat}, thus the constant representing the addition of the second substrate is written as a Michaelis constant (K_m). Because mechanisms depicted in **Figures 2.2B**, **2.2C**, and **2.2D** all share the same mathematical form for their rate equations, it is not possible to distinguish between them using Lineweaver-Burk plots (*Section 2.1.4.3*), as their intersecting patterns are similar.

Some bireactant mechanisms have been omitted from discussion (and **Figure 2.2**), as they are not relevant based on the kinetic behavior of SSADH. The *Ping-Pong* mechanism, also known as double-displacement, is a non-sequential mechanism whose rate equation (not shown) gives rise to a set of parallel lines (**Figure 2.3A**) when transformed into a Lineweaver-Burk plot (*Section 2.1.4.3*) which were not observed for SSADH. The *Rapid Equilibrium Ordered* (RE Ordered) mechanism is not relevant, as one set of Lineweaver-Burk plots intersects on the ordinate (**Figure 2.3C**) as a result of the missing ($K_{m,NADP+}$[SSA]) term in the denominator (**Equation 2.7**; compare to **Equation 2.5**), and this behavior was not observed.

Equation 2.7. RE Ordered.

$$v_0 = \frac{V_{max}[NADP^+][SSA]}{K_{iNADP+}K_{SSA} + K_{SSA}[NADP^+] + [NADP^+][SSA]}$$

Finally, the *Steady State Random* mechanism is also not relevant because the rate equation has squared terms of the substrate concentrations (not shown). As a consequence, untransformed data does not have the form of a rectangular hyperbola, and double-reciprocal plots are not linear. No data with those characteristics was encountered and this mechanism is not discussed any further.

Lineweaver-Burk Plots

Long before modern computers made nonlinear regression a trivial thing, researchers relied on mathematical transformations in order to more easily analyze the data. The Lineweaver-Burk plots, which linearize the data by taking the reciprocal of both v_0 and [S], were a convenient tool for determining the values of the kinetic constants using linear regression.[12] Unfortunately, small errors in v_0 measurements are greatly amplified when they are transformed into reciprocal units ($1/v_0$). This is particularly true at low substrate concentrations, where errors in v_0 measurement are more likely to occur. Because the v_0 values collected at low [S] are now at the upper end of the 1/[S] axis when transformed into reciprocal values, and thus they are weighed more heavily during linear regression, this greatly compounds the error in parameter estimation. Thus, it is always preferable to use untransformed data and rate equations such as **Equation 2.2** (single-substrate) or **Equation 2.5** (bireactant) along with nonlinear regression to determine values for V_{max} and K_m.

Despite being largely displaced by nonlinear regression methods, Lineweaver-Burk plots still possess a utility when it comes to diagnosing bireactant mechanisms and different types of enzyme inhibition. Here, one substrate (A) is varied, and the second substrate (B, or inhibitor) concentration is held fixed to produce a dataset. Additional datasets are produced in the same manner, this time at different second substrate (B, or inhibitor) concentrations. When these datasets, resulting from different [B], are transformed into double reciprocal units ($1/v_0$ versus 1/[A]) and plotted together, a pattern of lines emerges (**Figure 2.3**). From these patterns of lines different kinetic mechanisms (or modes of inhibition) may become apparent or ruled out as possible mechanisms.

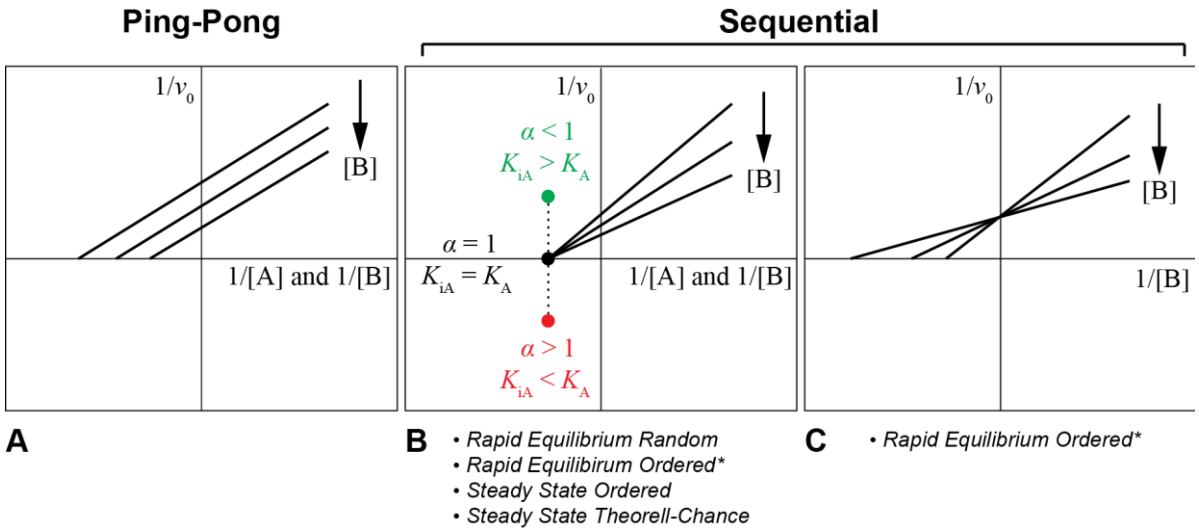

Figure 2.3. Lineweaver-Burk plots for relevant enzyme kinetics mechanisms. (A) Parallel lines are diagnostic of a Ping-Pong mechanism, also known as double-displacement. **(B)** Several sequential mechanisms have plots that intersect to left of the *y*-axis, though position of intersection relative to *x*-axis may vary. The green (•), black (•), and red (•) filled circles indicate the possible locations of the intercept with respect to the *x*-axis, along with the specified conditions, equalities or inequalities, that determine this point. (**For the rapid equilibrium ordered mechanism, intersection to the left of the y-axis occurs only with the 1/v_0 vs 1/[A] plot.*) **C)** The rapid equilibrium ordered mechanism intersects on the *y*-axis only with the 1/v_0 vs 1/[B] plot.

Replots

In much the same way that Lineweaver-Burk plots can be used to find values or diagnose mechanisms, various types of replots have been employed for the same purpose. One example is shown in **Figure 2.4** (adapted from Cleland).[11] From initial velocity plots where the concentration of substrate A is varied at different fixed concentrations of substrate B, the *apparent* V_{max} and *apparent* K_A values are obtained at each concentration of B. The *apparent* K_A values may now be plotted as a function of [B], revealing a number of possible relationships that depend on the value of the α-term (or the magnitudes of K_{iA} and $K_{m,A}$). For example, the substrates may bind synergistically, such that the binding of one substrate increases the affinity

of the enzyme for the second substrate (green curve in **Figure 2.4**). Additionally, with sufficient data, the values for K_{iA} and $K_{m,A}$ may be read directly from the plot.

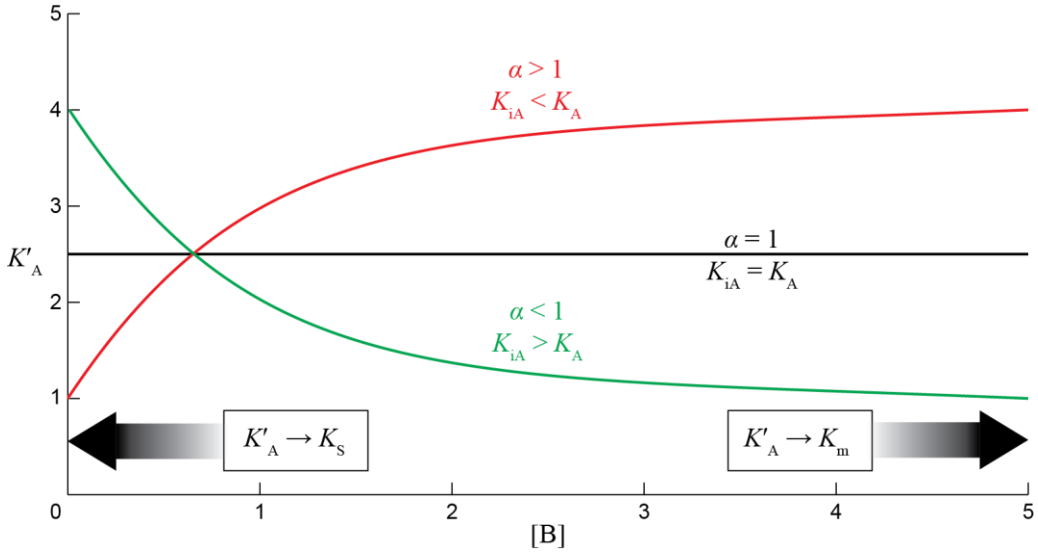

Figure 2.4. Rapid equilibrium random mechanism: K'_A dependence on [B]. A hypothetical look at the dependence of the apparent Michaelis constant for substrate A on the concentration of substrate B. The horizontal **black** line in the middle of the plot indicates the situation when B has no effect on the enzyme's affinity for A (no change in K'_A). The **red** rectangular hyperbola indicates the situation where B decreases the enzyme's affinity for A (K'_A increases). The **green** rectangular hyperbola indicates the situation where B increases the enzyme's affinity for A (K'_A decreases). Note that at low [B], K'_A becomes K_S (equilibrium dissociation constant for A), but at high [B], K'_A becomes K_m (Michaelis constant for A).

2.2 Results and Discussion

2.2.1 Transformation, Protein Expression, and Cell Lysis

After transformation and selection, sequencing of miniprepped plasmid DNA confirmed successful introduction of the pKN12 expression vector into the BL21(DE3)pLysS host cell line. Induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) during a typical protein expression resulted approximately 3.5 g of cell paste per 250 mL of growth media, which provided approximately 150 mg of total protein after cell lysis—a significant fraction of which was recombinant 6×His-SSADH. The results that follow are taken from a single expression and purification, which took place after optimizing the entire process. Lysis of 3.66 g of cell paste resulted in 3.73 mL of clarified lysate with 44.03 mg/mL of total protein (164.23 mg total protein). The specific activity of 6×His-SSADH in the clarified lysate was 10.24 U/mg.

2.2.2 Purification of Recombinant SSADH

Immobilized Metal Affinity Chromatography

Initial attempts to purify 6×His-SSADH omitted the use of BME during IMAC purification, resulting in a 10-fold decrease in specific activity after this step compared to when BME was included. Although introduction of BME immediately after IMAC purification resulted in some recovery of activity, the specific activity of 6×His-SSADH after the final purification step was 5-fold less than when BME was used throughout the entire purification. Generally, the Ni^{2+}-NTA equilibrium and loading buffers always include a small concentration of free imidazole, usually 10 mM, to prevent non-specific interactions between the Ni^{2+} cations and material in the lysate other than the 6×His-tagged protein of interest. It was observed that

5 mM BME can safely be included in the IMAC buffers to prevent oxidation of 6×His-SSADH, without significantly compromising the Ni^{2+} cations, as long as a minimum of 10 mM imidazole was present. Over time there appears to be no more appreciable loss in column binding capacity and performance than is typical for IMAC resins, which are routinely stripped and recharged after frequent use.

The elution profile for the IMAC purification of 6×His-SSADH is presented as a plot of total protein concentration versus fraction number (**Figure 2.5**). The flow-through and first wash fractions concentrations are omitted because their high concentration overwhelms the remainder of the profile when they are included, rendering the gradient fractions nearly featureless on the plot. Another consideration when presenting elution profile data is the use of total protein concentration as units on the ordinate, as opposed to units of total protein mass, as this may be misleading in certain situations. For instance, it is often practical to collect wash fractions in much larger volumes than gradient fractions. From the elution profile, fractions 3 (wash) and 13 (gradient) both have Bradford protein concentrations of approximately 1 mg/mL (**Figure 2.5**; circled in black). However, fraction 3 is approximately 4-times the volume of fraction 13 and therefore contains 4-times more protein by mass, 8.5 mg versus 1.8 mg. This fact is not immediately evident from the profile.

Based on the elution profile, compositional analysis on selected IMAC fractions was performed by SDS-PAGE (**Figure 2.6**). The flow-through fraction appeared to contain a fair amount of recombinant 6×His-SSADH after the initial pass of crude lysate through the resin. A second pass of the flow-through onto the column was not performed, since a large amount 6×His-SSADH had bound on first pass and was subsequently recovered from the column, relatively pure and in sufficient yield for the following step.

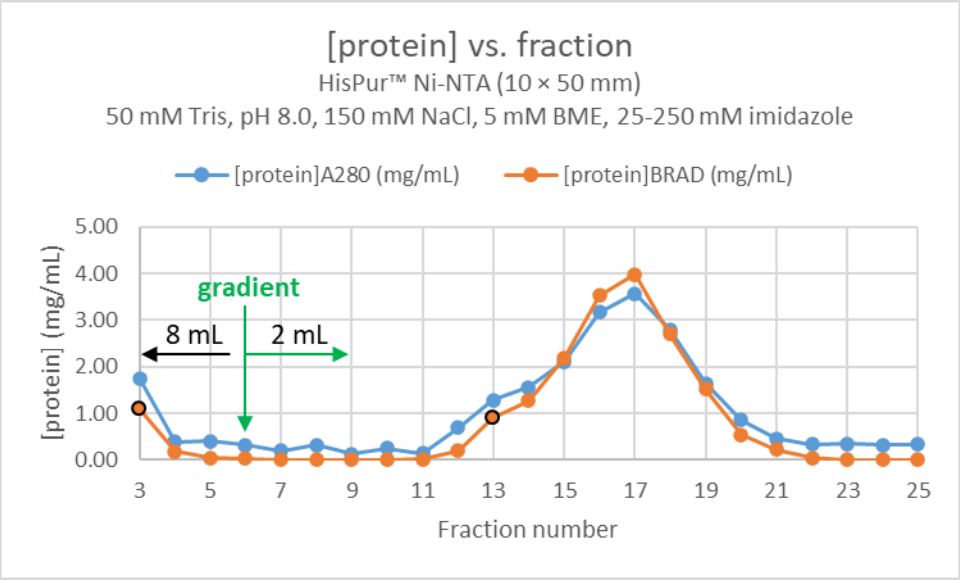

Figure 2.5. The elution profile for the IMAC purification of SSADH. The total protein concentration from fractions eluted from the Ni^{2+}-NTA column is plotted against fraction number to develop an elution profile for the purification. Concentrations were determined using both the A_{280} method (**blue**) and Bradford assay (**orange**). The imidazole gradient began at fraction 6, indicated by the green arrow (↓), with the preceding wash collected at 8 mL/fraction and the gradient at 2 mL/fraction. The Bradford protein concentrations for fractions 3 and 13 are circled in black because they have similar concentrations, but different total protein masses (see text).

E. coli contain a number of endogenous proteins known to bind Ni^{2+}, as well as other similar divalent metals. The presence of these metal-binding proteins in crude lysate leads to competition for the Ni^{2+}-NTA binding sites and effectively decreases the binding capacity of the resin for 6×His-tagged proteins. Examples of these interactions are visible in Lane 6 (G12 fraction) of the SDS-PAGE gel (**Figure 2.6**), where the 6×His-SSADH band is marked with a magenta arrow. The very dark band above 6×His-SSADH, between the 75 and 63 kDa markers, is likely arnA (bifunctional polymyxin resistance protein). The dark band under 6×His-SSADH, just below the 48 kDa marker, is likely DnaJ (Zn^{2+}-binding molecular chaperone). Both arnA and DnaJ protein bands will bind to ThermoFisher's InVision™ His-Tag In-Gel Stain and fluoresce strongly under a UV transilluminator. The InVision stain is comprised of a fluorescent dye conjugated to a Ni^{2+}-NTA complex. Because arnA and DnaJ readily bind the InVision stain, it can be assumed that they also bind the Ni^{2+}-NTA resin through similar

interactions, reducing the available number of Ni^{2+} coordination sites for the 6×His-tagged protein of interest. The remaining contaminant bands at or below 25 kDa are likely a

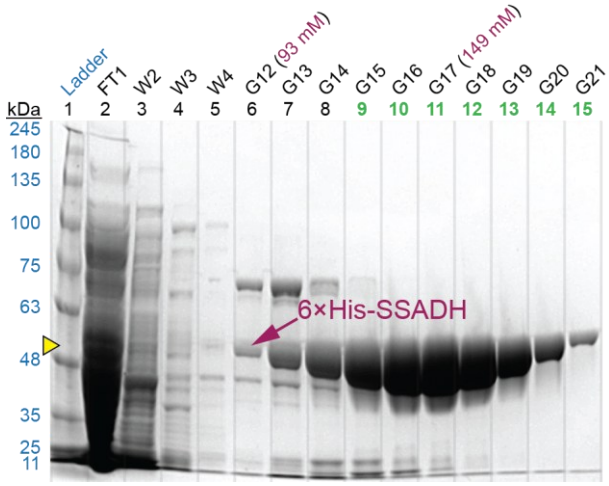

Figure 2.6. SDS-PAGE of fractions from IMAC purification of SSADH. The molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). The bands of interest are indicated with a **magenta** arrow, with the vertical height of the arrow tip (relative to the ladder) indicated by the **yellow** triangle. The calculated molar mass of 6×His-SSADH is 54.2 kDa. Well numbers appear above the lanes, and those colored in **green** indicate the fractions that were collected and pooled for analysis. The labels above the well numbers indicate the column fraction composition and number: FT = flow-through, W = wash, G = gradient. The estimated concentration of imidazole (Imd) in the eluate is given in parentheses for relevant fractions. Fractions without sufficient protein concentration were not analyzed by SDS-PAGE (W5–G11).

combination of additional nickel binding proteins and chaperone. The impact of metal-binding proteins and chaperones during IMAC purification is addressed further in Chapter 3. Due to the presence of these impurities, only *fractions G15–G21* were pooled and advanced to the next purification step.

The advertised binding capacity for HisPur™ Ni-NTA resin is ≥ 60 mg for a 28 kDa 6×His-tagged protein per mL of settled resin, with the protein identity, initial sample purity, and purification conditions unspecified. At 54 kDa, 6×His-SSADH is twice the molar mass of the manufacturer's protein and so a decrease in binding capacity of the resin for 6×His-SSADH was anticipated. A further decrease was expected from the myriad of interactions present in crude lysate, particularly those that may compete with the 6×His-tag···Ni^{2+} association, as well

as from the use of BME. The 4 mL of resin, according to the manufacturer, should result in purification of at least 240 mg of their 28 kDa His-tagged protein. Pooled *fractions G15–G21* contained 28.82 mg total protein (from Bradford assay), nearly all of it 6×His-SSADH.

After centrifugal concentration and buffer exchange, the pooled IMAC fractions had a final volume of 1.31 mL and contained a total protein concentration of 14.30 mg/mL (A_{280}), or 13.61 mg/mL (Bradford). The A_{280} value was chosen, given the overall purity of the pooled and concentrated fractions, providing a total protein mass of 18.76 mg and specific activity of 34.07 U/mg for 6×His-SSADH.

Size Exclusion Chromatography

Inspection of the SDS-PAGE gel from the IMAC purification (**Figure 2.6**) showed evidence of one high and several low molecular weight contaminants in the fractions that were collected and pooled. However, since the gel is run under denaturing and reducing conditions, which disrupt disulfide bonds and associations that hold quaternary structures together, the height of a given band on the gel may not reflect the native molecular weight of that protein in a biological context. For instance, a 100 kDa homotetrameric protein may appear as a 25 kDa monomer on a denaturing gel. The practical consequence of this is that some low molecular weight impurities could run as oligomers on SEC and coelute with 6×His-SSADH. Furthermore, any contaminants with a high binding affinity to 6×His-SSADH, such as chaperones, may also coelute with 6×His-SSADH.

In addition to removing impurities, the SEC also allows for complete exchange of the buffer during the purification. Although buffer exchange can be performed during concentration of the pooled IMAC fractions, the use of a centrifugal concentrator was kept to a minimum. This decision was made because of concerns that highly concentrated proteins

may be prone to aggregation or lost onto the filter membrane, while repeated addition of fresh buffer to highly viscous protein samples can introduce air bubbles that promote protein denaturation. Therefore, after concentration, the sample was diluted into the SEC buffer and loaded directly onto Sephacryl™ S-200 HR resin where the remainder of the buffer exchange would take place during the purification.

The fractionation range of Sephacryl™ S-200 HR is 5kDa to 250kDa, which allows for the purification of both monomeric and oligomeric forms of 6×His-SSADH, assuming that the N-terminal 6×His-tag does not interfere with oligomerization. The column was calibrated with a total of thirteen unique MW standards in order to determine key parameters and generate a standard curve that could reliably estimate molecular weight of proteins based on their retention volume. This estimate assumes that nonspecific ionic interactions between proteins and the resin are negligible, since these interactions can lead to protein adsorption and cause a delay in peak elution. Such interactions can be avoided by the addition of salt to the buffer. Fortunately, these interactions can be avoided by the addition of salt to the mobile phase buffer during purification. Molecules much greater than the resin's fraction range of 250 kDa cannot be resolved. For this column, blue dextran (2,000 kDa) and thyroglobulin (670 kDa) both have a peak elution volume of 45 mL, as do large aggregates of smaller proteins. Therefore, 45 mL is the *void volume* for this column and represents the volume of solutes that are excluded from the resin pores. A well-packed column should have a void volume equivalent to approximately 30% of the total column volume. The *void volume* here was 35% of the total volume (129 mL).

The elution profile for the SEC purification of 6×His-SSADH is shown in **Figure 2.7**. Total protein concentration initially increases between *fractions 30–35*, abruptly plateaus, and then resumes increasing until it reaches a maximum at *fraction 41*. This biphasic elution profile

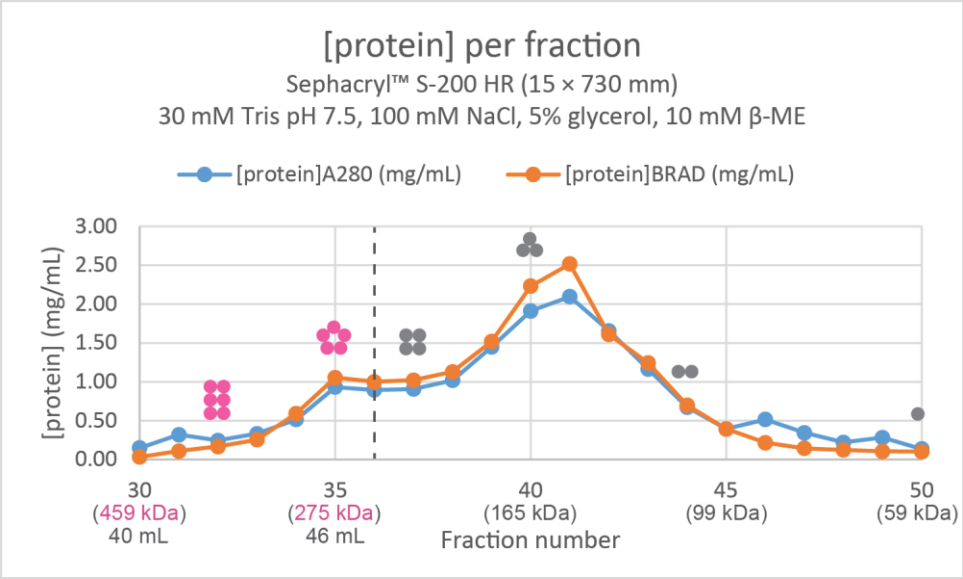

Figure 2.7. The elution profile for the SEC purification of 6×His-SSADH. The total protein concentration from fractions eluted from the Sephacryl™ S-200 HR column is plotted against fraction number to develop an elution profile for the purification. Concentrations were determined using both the A_{280} method (**blue**) and Bradford assay (**orange**). The estimated molecular weight composition of selected fractions is given below in parentheses and is based on a standard curve. Dot patterns (• and •) above the plots indicate the predicted oligomerization of 6×His-SSADH for that fraction. Molecular weights and dots colored in pink are outside the fractionation range of the resin and therefore estimated with very low confidence. The vertical dashed line through fraction 36 denotes the fractionation limit of the resin and corresponds to an elution volume of 48 mL (248 kDa).

results from the resin's 250 kDa upper fractionation limit, which is marked by a vertical dashed line through *fraction 36* (**Figure 2.7**). *Fraction 36* corresponds to an elution volume of 48 mL and has an estimated molecular weight composition of 248 kDa from the standard curve.

Data for the standard curve was restricted to molecular weight standards that fell within the resin's fractionation range, therefore extrapolation of the curve outside of this range cannot produce values with reasonable accuracy. For instance, extrapolation of the standard curve predicts 307 kDa for thyroglobulin (670 kDa), an error of 54%. Since *fraction 35* (46 mL) is on the proximal shoulder of the *void volume* peak (45 mL), its fractional composition can only be speculated to be a combination of large 6×His-SSADH oligomers that occupy the column's excluded volume, with earlier fractions rationalized similarly. Larger 6×His-SSADH

oligomers may be partitioned with use of Sephacryl™ S-300 HR, which has a fractionation range of 10,000 kDa to 1,500,000 kDa.

SDS-PAGE analysis of SEC *fractions 33–46* reveals that 6×His-SSADH is very pure, but traces of low molecular weight impurities are evident across most of the fractions (**Figure 2.8**). These bands appear to be from chaperones that co-purify with 6×His-SSADH, as the intensity of these bands mirrors that of 6×His-SSADH. The possible identities of the chaperones responsible for these band are 40 kDa (DnaJ), 22 kDa (GrpE) and 10 kDa (GroES). Because these bands are significantly less intense than the bands for 6×His-SSADH, amounting to only a few percent of the total protein, their influence on the oligomerization and partitioning of 6×His-SSADH is expected to be minimal, but not zero. The possible consequence of chaperone association would be an increase in the *apparent* molecular weight of 6×His-SSADH, which would result in the peak elution being shifted slightly forward (earlier volume). Also,

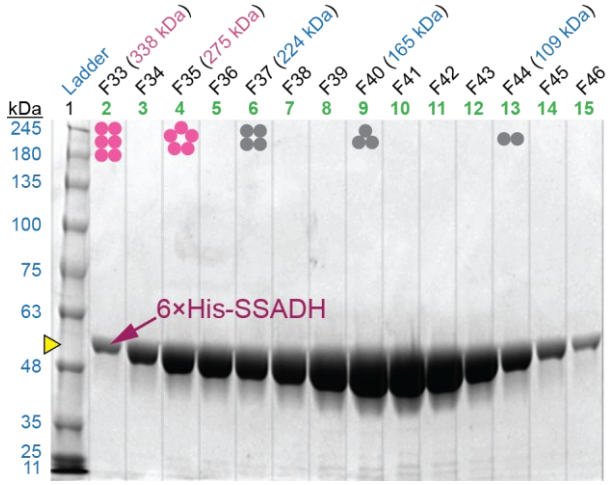

Figure 2.8. SDS-PAGE gel showing the SEC purification of 6×His-SSADH. The molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). The bands of interest are indicated with a magenta arrow (◄), with the vertical height of the arrow tip (relative to the ladder) indicated by the yellow triangle (▶) adjacent the ladder. The calculated molar mass of 6×His-SSADH is 54.2 kDa. Well numbers appear above the lanes, and those colored in **green** indicate the fractions that were collected and pooled for analysis. Labels above the well numbers indicate the column fraction number: F33 = fraction 33. The estimated molecular weight composition for fractions of interest is shown above in parentheses. Dot patterns (● and ●) at the top of the gel indicate the predicted oligomerization of 6×His-SSADH for that fraction. Molecular weights and dots colored in pink are outside the fractionation range of the resin and therefore estimated with very low confidence.

chaperones may preferentially bind a region that is formed or exposed in only one of several possible oligomeric states, impacting the equilibrium between these states. Thus, the net result of significant chaperone interaction could be an elution profile that is slightly shifted and/or broadened from the ideal.

Fractions 33–46 were deemed pure enough to pool for a combined volume of 18.06 mL and a total protein concentration 1.005 mg/mL (A_{280}). The total protein mass was 18.15 mg and the specific activity of 6×His-SSADH was 34.47 U/mg. Fractions were immediately divided into 88×200 µL aliquots, flash-frozen in liquid nitrogen and stored at -80 °C.

A purification summary for 6×His-SSADH is presented in **Table 2.1**. The greatest amount of purification was achieved with the IMAC column, which is to be expected since affinity chromatography is well known to be a powerful separation technique. SEC did not produce a significant increase in 6×His-SSADH purity, but it did afford the ability to exchange the buffer and observe various oligomerization states for 6×His-SSADH. The table also quantifies what is clearly evident by SDS-PAGE analysis, which is that 6×His-SSADH is strongly expressed in *E. coli* BL21(DE3)pLysS cells using the pKN12 vector (29.5% cellular protein abundance).

Table 2.1. 6×His-SSADH purification summary table. "Lysate" is the clarified lysate that was loaded onto the IMAC column. "IMAC" is the eluate from the IMAC column after fractions were pooled and concentrated. "SEC" is the eluate from the SEC column after fractions were pooled. See appendix for individual calculations.

Step	Volume (mL)	[Protein] (mg/mL)	Activity (U)*	Specific activity (U/mg)*	Fold purified	Yield (%)	Cellular Abundance (%)	k_{cat} (s^{-1})*
Lysate	3.73	44.03	1,682.99	10.24	1.0	100.0		
IMAC	1.31	14.30	639.08	34.07	3.3	38.0		
SEC	18.06	1.00	630.41	34.74	3.4	37.5	29.47	31.29
* Values determined at pH 7.5 and room temperature (20–22 °C)								

2.2.3 Enzyme Kinetics of Recombinant SSADH

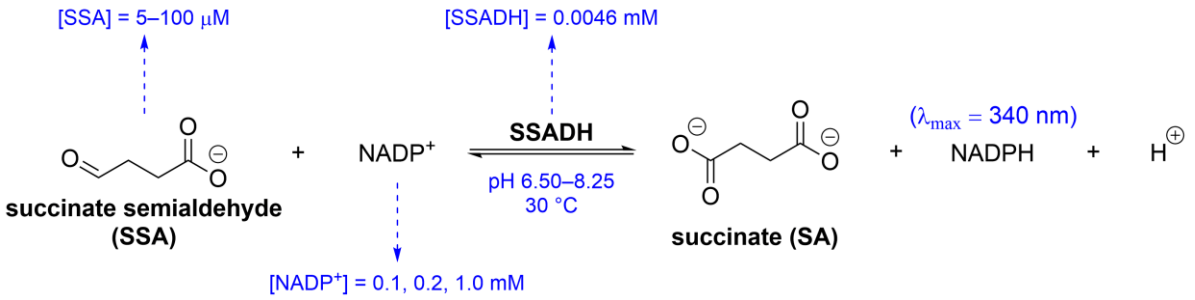

Figure 2.9. The SSADH enzyme kinetics assay.

Introduction to the Rate Experiments

The enzyme kinetics data presented in the following section was obtained by measuring the *initial velocity* (v_0) of SSADH-catalyzed reactions in the *forward direction* (dehydrogenase), where the succinate semialdehyde (SSA) substrate is oxidized to succinate (SA) with the concomitant reduction of the $NADP^+$ coenzyme to NADPH (**Figure 2.9**).

NADPH has a unique absorbance spectrum, with a peak absorbance (λ_{max} = 340 nm) in the ultraviolet A region that is not shared by $NADP^+$ or any of the other substrates and products in the reaction. This distinctive spectral signature provides a convenient way of monitoring the progress of the catalyzed reaction, in either direction. Recording the absorbance of the reaction at 340 nm produces a reaction progress curve that represents the change in NADPH concentration over time. Because the change in product concentration over time is one way to define the reaction velocity, $v = d[P]/dt$, the slope of the progress curve at any point represents the velocity of the enzyme-catalyzed reaction at that time. **Figure 2.10** shows a set of reaction progress curves collected during development of the enzyme assay.

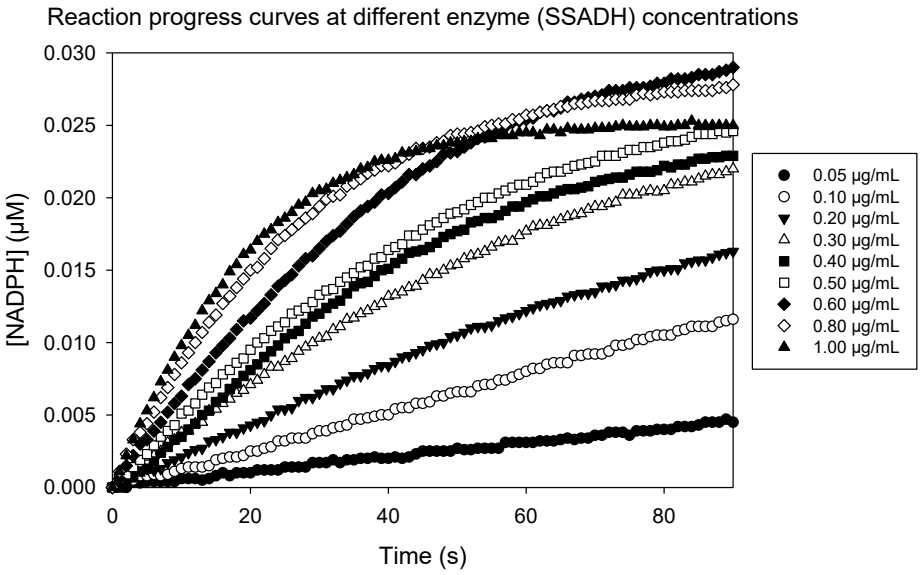

Figure 2.10. A set of reaction progress curves for SSADH. The progress curves represent the increase in NADPH concentration over the 90 seconds following reaction initiation. Each curve reflects a different concentration of enzyme (SSADH) used in the reaction. **Assay conditions:** [SSADH] = 0.05–1.0 μg/mL, [$NADP^+$] = 1.0 mM, [SSA] = 5 μM, in 100 mM sodium phosphate at pH 7.5 and 30 °C.

Progress curves were analyzed using the *initial-rate method*, where initial velocity (v_0) is measured from the progress curve at or near the start of the reaction (t = 0), when substrate depletion is minimal, very little product has formed, and the initial velocity remains essentially constant ($d^2[P]/dt^2 = 0$). This early phase of the reaction, where the enzyme-substrate complex concentration [ES] is also assumed to remain constant ($d[ES]/dt = 0$), has been termed the *pseudo-steady-state* since it takes place after a fast initial transient where the enzyme-substrate complex first builds up (E + S → ES). During the initial transient, the depletion of the free substrate to form the ES-complex is also assumed to be minimal so that the substrate concentration [S] during the steady state may be approximated by the initial substrate concentration $[S]_0$. This is known as the *reactant stationary assumption* ($[S] \approx [S]_0$).

Application of the Henri-Michaelis-Menten equation, or similar equations derived using identical assumptions, necessitates that initial velocities are correctly measured and that the steady-state and reactant stationary assumptions are valid for the experiment. All these criteria

were satisfied by keeping the substrate concentrations orders of magnitude higher than the total enzyme concentration during all the rate experiments. A sufficiently high substrate concentration with respect to the total enzyme concentration ($[S] >> [E]_T$) ensures that less than 5% of the substrate was consumed when the initial velocities were measured, satisfying the steady-state assumption. Additionally, using an enzyme concentration value far below the substrate K_m satisfies the reactant stationary assumption, ensuring that $[S] \approx [S]_0$. Initial velocity conditions were confirmed during development of the enzyme assay (*Section 1.1.3.2*), then these parameters were optimized.

Each experiment consisted of a set of reaction conditions, some of which were allowed to vary while others were tightly controlled. The conditions that were allowed to vary included SSADH concentration (enzyme), SSA concentration (substrate), $NADP^+$ concentrations (coenzyme), as well as buffer pH. The initial velocity data sets were fit to the appropriate *rapid-equilibrium* or *steady-state* derived rate equations to provide: i) the kinetic parameters k_{cat} and K_m; ii) the pH optimum for the catalyzed reaction and the pK_a value of an active site residue involved in catalysis; and iii) a likely mechanism for the catalyzed reaction.

Development of the Enzyme Assay

The first step in the development of an initial-rate assay was to verify that the initial velocity of the catalyzed reaction is linearly dependent on the total enzyme concentration. This need arises because all enzymes displaying Michaelis-like behavior have rate equations of a form similar to $v_0 = k[E]_T/(1 + K_S/[S])$, where the initial velocity is proportional to the total enzyme concentration used in the reaction ($v_0 \propto [E]_T$). Demonstrating this linear dependence is especially crucial at the lowest substrate concentration to be used during the rate experiments, because the substrate will be consumed quickly and the steady-state window will

be very small, making it challenging to obtain initial velocity data without the use of specialized equipment that allows for rapid mixing. Finding an appropriately small enzyme concentration that can provide reliable initial rate measurements at low substrate concentrations is vital in getting good kinetics data. Linearity between v_0 and $[E]_T$ may break down at very low enzyme concentrations for a number of reasons, including an unfavorable change in the oligomeric state of the enzyme upon dilution, the presence of a co-purified inhibitor, and adsorption of the enzyme to the reaction vessel walls.

The linear dependence of initial velocity on total enzyme concentration is demonstrated in the v_0 versus $[E]_T$ plot in **Figure 2.11**, where the initial velocity values used were measured from the reaction progress curves previously shown in **Figure 2.10**.

The K_m for SSADH from various organisms, with SSA as the substrate, ranges from 1–189 μM depending on the gene and sequence. For the recombinant SSADH supplied for this study by Prof. James C. Whisstock, the K_m with SSA as a substrate was reported as 16.94 ±

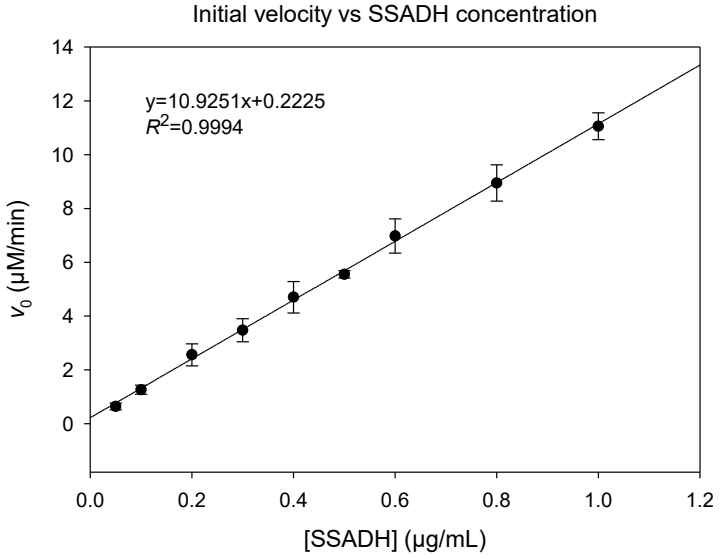

Figure 2.11. Initial velocity versus SSADH concentration. Filled black circles (•) represent the mean initial velocity (v_0) obtained at different concentrations of enzyme in the final reaction mixture. I-bars indicate the standard deviation of each data point. The solid line through the data points is obtained by linear regression, with the resulting equation and R^2 shown on the plot. **Assay conditions:** [SSADH] = 0.05–1.0 μg/mL, [$NADP^+$] = 1.0 mM, [SSA] = 5 μM, in 100 mM sodium phosphate at pH 7.5 and 30 °C. Assays performed in triplicate.

2.2 μM. Based on this information, 5.0 μM SSA would be the lowest substrate concentration assayed. Initial velocity data were obtained using 6×His-SSADH concentrations ranging from 0.05–1.0 μg/mL, with SSA and $NADP^+$ fixed at 5.0 μM and 1.0 mM, respectively. The initial velocity showed a linear dependence with all concentrations of SSADH assayed. A concentration of 0.25 μg/mL (or 4.6 nM) of 6×His-SSADH was selected to be the standard enzyme concentration for the assays. By comparison, the lowest SSA concentration used during the assays was 5 μM, which places the lowest substrate concentration at over 1,000 times excess, satisfying the requirement that $[S] >> [E]_T$.

To verify that enzyme and substrate concentrations satisfied the initial velocity conditions of less than 5% of the substrate consumed during the measurements, a test was performed at pH 8.0 (near pH optimum) with 0.25 μg/mL of enzyme, using the maximum cofactor concentration (1.0 mM $NADP^+$) and the lowest substrate concentration (5 μM SSA). The initial velocity was recorded as 6.11 μM min^{-1} (NADPH produced). For the SSADH reaction the stoichiometry between substrate and product is 1:1, so this is also the rate of substrate consumption. 5% of 5 μM SSA is 0.25 μM, the maximum substrate that can be consumed before the initial velocity conditions are no longer valid. At the recorded initial velocity, it would take approximately 2.5 seconds to consume 5% of the substrate. It took less than 2 seconds to initiate the experiment and begin recording the absorbance using the procedure described in the Methods (*Section 2.4.6.2*).

Maintaining strict control of the reaction temperature is essential when determining enzyme kinetics parameters and, from Arrhenius equation, just a 1 °C increase in temperature can increase the rate of a chemical reaction by as much as 10%. Early attempts at getting consistent data were plagued by dramatic temperature fluctuations in the room where the

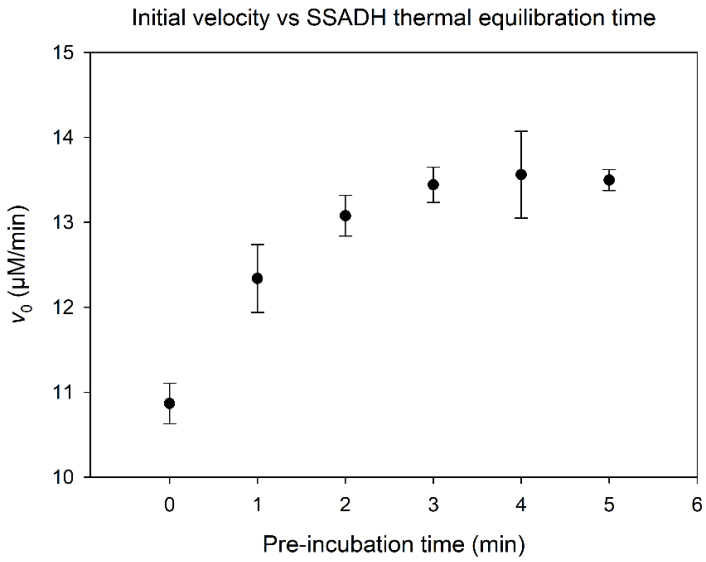

Figure 2.12. Initial velocity vs SSADH thermal equilibration time. Filled black circles (•) represent the mean initial velocity (v_0) obtained after pre-incubating the enzyme for various periods of time, at 30 °C, prior to initiating the reaction. I-bars indicate the standard deviation of each data point. **Assay conditions:** [SSADH] = 0.25 µg/mL, [$NADP^+$] = 1.0 mM, [SSA] = 100 µM, in 100 mM sodium phosphate at pH 7.5. Assays performed in triplicate.

experiments were conducted. It was therefore decided that all reactions should be conducted at 30 °C. Since most of the reaction components were kept on ice during the experiments, it was necessary to determine the length of time required to bring the components into thermal equilibrium with the reaction temperature of 30 °C before initiating the enzyme reaction. The plot of initial velocity versus equilibration time (**Figure 2.12**) illustrates that the components require a pre-incubation of 3–4 minutes at 30 °C in order to obtain reliable initial velocities.

After the final purification step and activity assay, purified 6×His-SSADH was aliquoted, flash frozen in liquid nitrogen and immediately placed into storage at -80 °C. On the day of rate experiments, one aliquot was thawed on ice, diluted to a working concentration with ice cold 100 mM sodium phosphate buffer pH 7.5 with the addition of 1 mM DTT, and then these solutions were placed in a 4 °C refrigerator after a 1 mL sample was withdrawn for experiments. Over the course of 20 days these thawed and diluted aliquots were tested for loss of activity. It is clear from **Figure 2.13** that despite minimal handling, the activity of SSADH

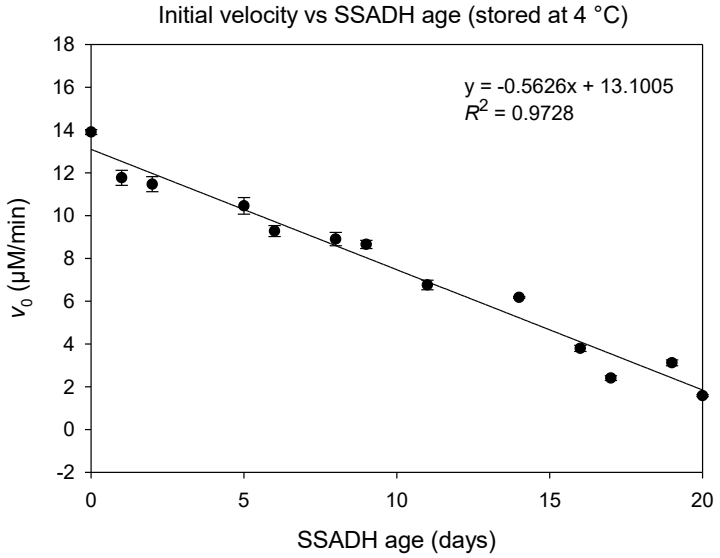

Figure 2.13. Initial velocity vs SSADH age, upon storage at 4 °C. Filled black circles (•) represent the mean initial velocity (v_0) obtained after storing the enzyme, at 4 °C, for up to 20 days. I-bars indicate the standard deviation of each data point. The solid line through the data points was obtained by linear regression, with the resulting equation and R^2 shown on the plot. **Assay conditions:** [SSADH] = 0.25 µg/mL, [$NADP^+$] = 1.0 mM, [SSA] = 100 µM, in 100 mM sodium phosphate at pH 7.5 and 30 °C. Assays performed in triplicate after 3 min pre-incubation.

degrades over time, with an approximately 4% loss in activity over 24 hours. Since the typical set of rate experiments took about 15 hours to complete, a 2% loss in activity over the course of the experiments was deemed acceptable, particularly since the order of the reactions was randomized to account for time-dependent phenomena. Because SSADH contains a pair of adjacent cysteines that function as a redox switch to regulate enzymatic activity based on the surrounding redox state, it was possible that a disulfide bond had formed between these two residues. To test this hypothesis an additional 1 mM DTT was added to the solution, but no increase in activity was observed after incubating for 1 hour. Either the cysteines were oxidized beyond disulfides (i.e., sulfenic, sulfinic, etc.) or there was another phenomenon responsible, possibly loss of native structure or aggregation.

To verify that this enzyme is the product of the *gabD* gene (**EC 1.2.1.79**), and possesses the stated preference for $NADP^+$ as the coenzyme, the initial rates for the SSADH reaction with 100 µM SSA were measured at different concentrations $NADP^+$ and NAD^+ at pH 7.50

and 30 °C. Nonlinear fit of this data to **Equation 2.10** provided a $K_m(NADP^+)$ and $K_m(NAD^+)$ of 0.079 ± 0.013 mM and 1.02 ± 0.50 mM, respectively, a 13-fold increase in K_m when NAD^+ is used in place of $NADP^+$. Additionally, k_{cat} decreased ~83% from 54.58 ± 2.31 s^{-1} to 9.45 ± 1.04 s^{-1} when NAD^+ is used in place of $NADP^+$.

Initial Velocities at pH 8.0 with One Substrate Fixed

Nonlinear regression was used to fit **Equation 2.8** to a dataset from a rate experiment performed at pH 8.0, where [SSA] was varied and [$NADP^+$] was fixed. A plot of the data and the nonlinear fit is presented in **Figure 2.14A**. The data were also transformed, by taking the reciprocal of both the initial velocities and substrate values, then fit to **Equation 2.9** by linear regression, and that data along with the linear fit are presented in **Figure 2.14B**. The values for V'_{max} and K'_{SSA}, that resulted from both fitting procedures, along with the statistics of the fitting, are given in **Table 2.2**. The prime (′) notation on V'_{max} and K'_{SSA} is used to denote these as the

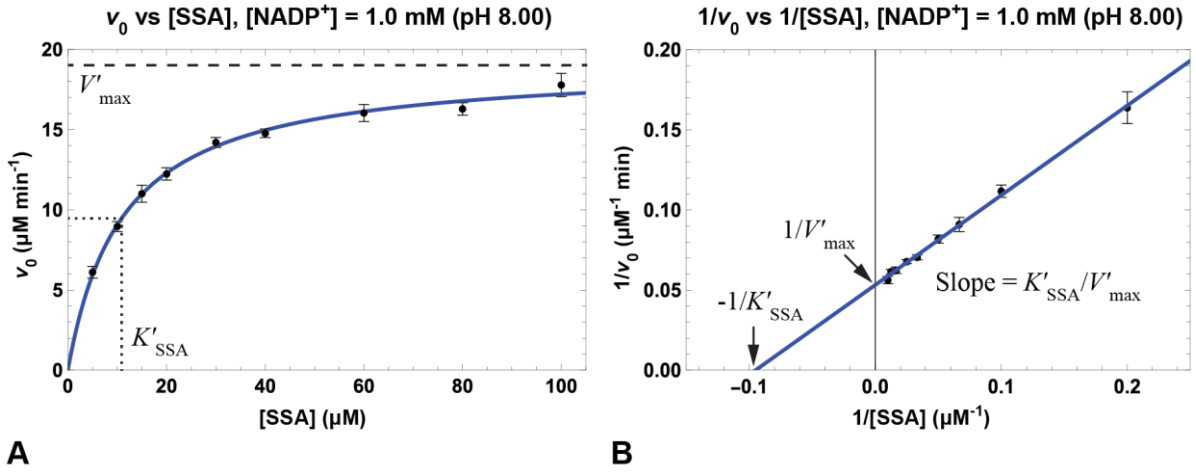

Figure 2.14. SSADH initial velocity vs [SSA], at fixed [$NADP^+$]. (A) Nonlinear regression of SSADH initial velocity (v_0) versus SSA concentration, [SSA]. Filled black circles (•) represent the mean initial velocity (v_0) at each [SSA] value, with I-bars indicating the standard deviation. The **blue rectangular hyperbola** is the nonlinear fit of the Henri-Michaelis-Menten equation to the data. V_{max} and K_{SSA} are indicated by dashed lines. **(B)** Lineweaver-Burk transformation of v_0 versus SSA concentration data. Filled black circles (•) represent inverse mean initial velocity ($1/v_0$) at each value of 1/[SSA], with I-bars indicating the standard deviation. The **blue line** is the linear fit of the Lineweaver-Burk equation to the data. $1/V_{max}$ and $-1/K_{SSA}$ are indicated by arrows; expression for slope is adjacent to the **blue line**. **Assay conditions:** [SSADH] = 0.25 μg/mL, [$NADP^+$] = 1.0 mM, [SSA] = 5–100 μM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

Equation 2.8. Rate equation for SSADH, [SSA] varied at fixed [$NADP^+$].

$$v_0 = \frac{V'_{max}[SSA]}{K'_{SSA} + [SSA]}$$

Equation 2.9. Lineweaver-Burk equation for SSADH, [SSA] varied at fixed [$NADP^+$].

$$\frac{1}{v_0} = \frac{1}{V'_{max}} + \frac{K'_{SSA}}{V'_{max}[SSA]}$$

Table 2.2. Parameters from fitting SSADH initial velocity vs [SSA], at fixed [$NADP^+$].

Henri-Michaelis-Menten (nonlinear)		Lineweaver-Burk (linear transformation)	
V'_{max}	19.11 ± 0.29 μM min^{-1}	$1/V'_{max}$	0.0531 ± 0.0007 μM^{-1} min
V'_{max} % Err	1.52%	$1/V'_{max}$ % Err	1.32%
K'_{SSA}	11.09 ± 0.64 μM	K'_{SSA}/V'_{max}	0.560 ± 0.009 min
K'_{SSA} % Err	5.77%	K'_{SSA}/V'_{max} % Err	1.61%
R^2	0.9995	R^2	0.9997
		V'_{max}*	18.81 ± 0.27 μM min^{-1}
		V'_{max} % Err	1.44%
		K'_{SSA}*	10.53 ± 0.24 μM
		K'_{SSA} % Err	2.28%

*The Lineweaver-Burk values for V'_{max} and K'_{SSA} were calculated from $1/V'_{max}$ and K'_{SSA}/V'_{max} values.

apparent V_{max} and *apparent* K_{SSA}. In order to distinguish them from the absolute V_{max} and K_{SSA}. They are apparent because their value ultimately depends on the concentration of the fixed substrate, which does not explicitly appear in the equations. The mathematical expressions for V'_{max} and K'_{SSA} are presented when their replots are discussed (*Section 2.2.3.5*).

Table 2.2 shows that nonlinear regression produced a V'_{max} of 19.11 ± 0.29 μM min^{-1}, while linear regression on the transformed data estimated V'_{max} as 18.81 ± 0.27 μM min^{-1}, a difference of 1.6%. When considering the standard deviation for both values, there really isn't a difference between the two, statistically. Similarly, the nonlinear regression value and the Lineweaver-Burk estimation of K'_{SSA} are 11.09 ± 0.64 μM and 10.53 ± 0.24 μM, respectively.

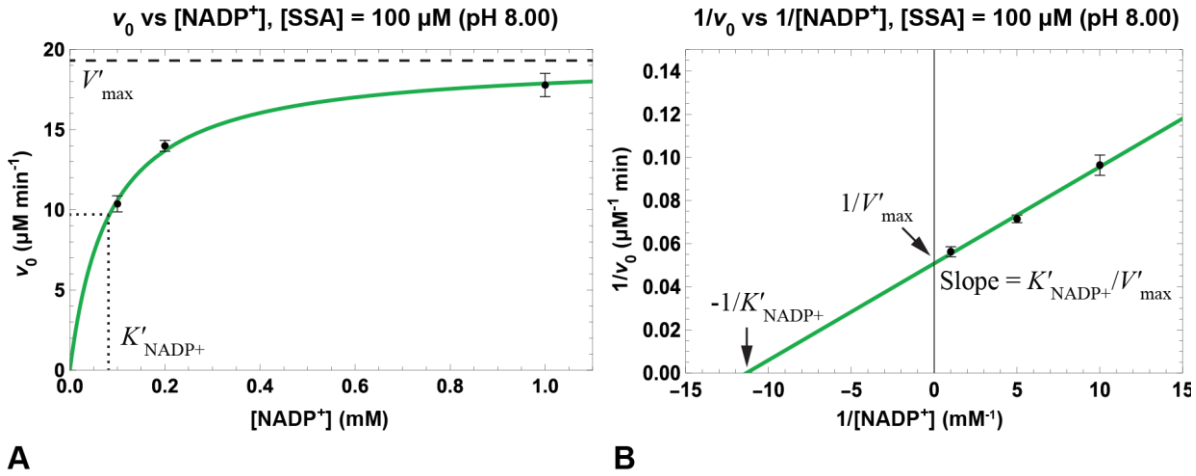

Figure 2.15. SSADH initial velocity vs [NADP$^+$], at fixed [SSA]. (A) Nonlinear regression of SSADH initial velocity (v_0) versus NADP$^+$ concentration, [NADP$^+$]. Filled black circles (•) represent mean initial velocity (v_0) at each [NADP$^+$] value; I-bars indicate standard deviation. The **green rectangular hyperbola** is the nonlinear fit of the Henri-Michaelis-Menten equation to the data. V_{max} and K_{SSA} are indicated by dashed lines. **(B)** Lineweaver-Burk transformation of v_0 versus NADP$^+$ concentration data. Filled black circles (•) represent inverse mean initial velocity ($1/v_0$) at each value of 1/[NADP$^+$]; I-bars indicate standard deviation. The **green line** is the linear fit of the Lineweaver-Burk equation to the data. $1/V_{max}$ and $-1/K_{NADP+}$ are indicated by arrows; expression for slope is adjacent to the **green line**. **Assay conditions:** [SSADH] = 0.25 µg/mL, [NADP$^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 100 µM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

This is a difference of 5.3%, but again, when the standard deviations are considered there is no real difference between the two values. This is largely because the lowest [S] used in the rate assay was not below $0.25K_m$, so the error in estimation of kinetic parameters by the Lineweaver-Burk treatment is minimal.

In order to make use of **Equation 2.10** with the existing experimental data, a new dataset was constructed from three existing rate experiments, where in each experiment [SSA] was varied at a fixed [NADP$^+$]. These three experiments were collected at 0.1, 0.2, and 1.0 mM NADP$^+$. From each of these three [NADP$^+$] datasets, the v_0 value corresponding to the same [SSA] value was extracted to create a new dataset where now [NADP$^+$] varied between 0.1 – 1.0 mM, with [SSA] fixed at a constant value. Such a dataset collected at pH 8.0, with [SSA] fixed at 100 µM and [NADP$^+$] allowed to vary is presented in **Figure 2.15A**, along with the nonlinear fit of **Equation 2.10** to this data. The Lineweaver-Burk transformation of the data,

Equation 2.10. Rate equation for SSADH, [NADP$^+$] varied at fixed [SSA].

$$v_0 = \frac{V'_{max}[NADP^+]}{K'_{NADP+} + [NADP^+]}$$

Equation 2.11. Lineweaver-Burk equation for SSADH, [NADP$^+$] varied at fixed [SSA].

$$\frac{1}{v_0} = \frac{1}{V'_{max}} + \frac{K'_{NADP+}}{V'_{max}[NADP^+]}$$

Table 2.3. Parameters from fitting SSADH initial velocity vs [NADP$^+$], at fixed [SSA].

Henri-Michaelis-Menten (nonlinear)		Lineweaver-Burk (linear transformation)	
V'_{max}	19.36 ± 0.52 μM min^{-1}	$1/V'_{max}$	0.051 ± 0.002 μM^{-1} min
V'_{max} % Err	2.67%	$1/V'_{max}$ % Err	3.92%
K'_{NADP+}	0.083 ± 0.008 mM	K'_{NADP+}/V'_{max}	0.0045 ± 0.0003 min
K'_{NADP+} % Err	9.67%	K'_{NADP+}/V'_{max} % Err	6.66%
R^2	0.9997	R^2	0.9997
		V'_{max}*	19.69 ± 0.84 μM min^{-1}
		V'_{max} % Err	4.26%
		K'_{NADP+}*	0.088 ± 0.007 mM
		K'_{NADP+} % Err	7.95%

*The Lineweaver-Burk values for V'_{max} and K'_{NADP+} were calculated from $1/V'_{max}$ and K'_{NADP+}/V'_{max} values.

and linear regression of it to **Equation 2.11** is presented in **Figure 2.15B**. The V'_{max} and K'_{NADP+} values from both the nonlinear and linear regressions are presented in **Table 2.3** along with the fitting statistics. Despite comprising of only 3 data points, nonlinear regression of the data provides an estimated V'_{max} of 19.36 ± 0.52 μM min^{-1}, which is close to the 19.11 ± 0.29 μM min^{-1} provided by the earlier nonlinear fit for v_0 versus [SSA] at fixed [NADP$^+$].

Combined pH 8.0 Initial Velocity Data as 2D Projections

Data from 3 rate experiments conducted at 0.1, 0.2, and 1.0 mM NADP$^+$, where in each experiment [SSA] varied from 5–100 μM, are plotted together in **Figure 2.16**. It should be noted that the highest [NADP$^+$] (colored in **blue**) used in the experiments was previously

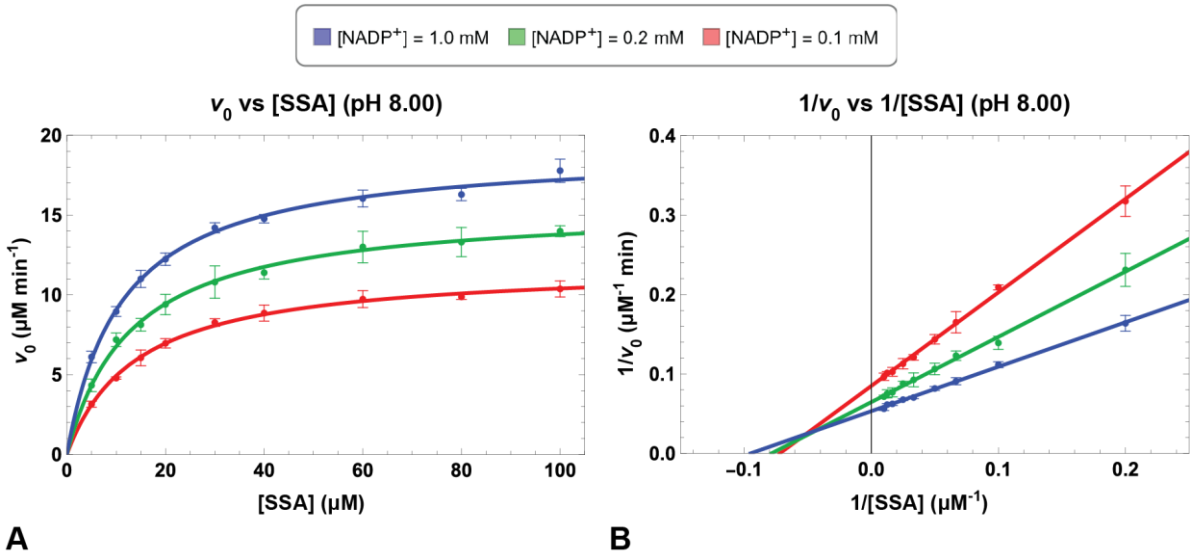

Figure 2.16. Combined fits for SSADH initial velocity vs [SSA], at fixed [NADP$^+$]. The legend appears above the figures. **(A)** Three different data sets, each fit independently with the Henri-Michaelis-Menten equation and then displayed together on the same plot. Filled colored circles represent the mean initial velocity (v_0) at each value for [SSA], with colored I-bars indicating the standard deviation, and colored rectangular hyperbolas are the nonlinear fits for each respective dataset. **(B)** The same data transformed into double-reciprocal format, fit with the Lineweaver-Burk equation and then displayed together on the same plot. Filled colored circles represent the inverse mean initial velocity ($1/v_0$) at each value for 1/[SSA], I-bars indicate the standard deviation, and colored lines are the linear fits for each respective dataset. **Assay conditions:** [SSADH] = 0.25 μg/mL, [NADP$^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 μM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

shown in **Figures 2.14**. Untransformed data fit with **Equation 2.8** (Henri-Michaelis-Menten) is presented in **Figure 2.16A**, where it is clear that an increase in fixed [NADP$^+$] leads to an increase in V_{max}. This point is graphically illustrated with replots (**Figure 2.18A**) in the next section. Data transformed into reciprocal units and fit using **Equation 2.9** is shown in **Figure 2.16B**, where the utility of the Lineweaver-Burk plot is evident. All lines intersect to the left of the ordinate and above the abscissa, in Cartesian quadrant II. This pattern is diagnostic for the following four bireactant mechanisms: i) *rapid equilibrium random*; ii) *rapid equilibrium ordered*; iii) *steady state ordered*; iv) and *steady state Theorell-Chance*.

Data from 9 rate experiments conducted at 5, 10, 15, 20, 30, 40, 60, 80, and 100 μM SSA, where [NADP$^+$] was allowed to vary from 0.1–1.0 mM, is combined in **Figure 2.17**. The untransformed data fit with **Equation 2.10** is shown in **Figure 2.17A**, where the conclusion

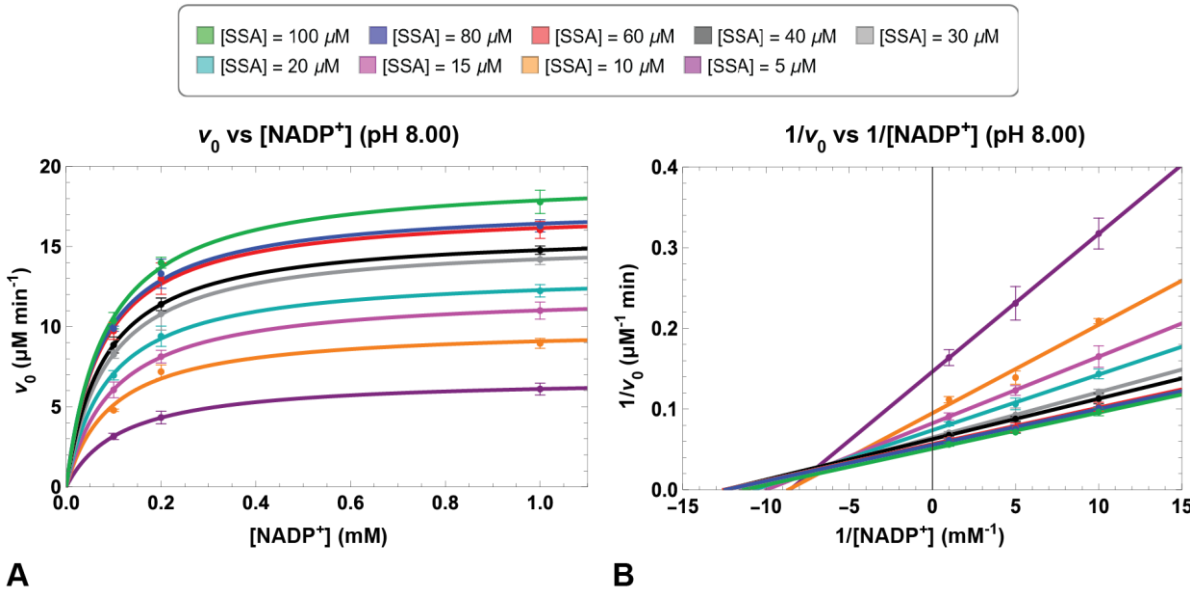

Figure 2.17. Combined fits for SSADH initial velocity vs [NADP$^+$] at fixed [SSA]. The legend appears above the figures. **(A)** Three different data sets, each fit independently with the Henri-Michaelis-Menten equation and then displayed together on the same plot. Filled colored circles represent the mean initial velocity (v_0) at each value for [NADP$^+$], with colored I-bars indicating the standard deviation, and colored rectangular hyperbolas are the nonlinear fits for each respective dataset. **(B)** The same data transformed into double-reciprocal format, fit with the Lineweaver-Burk equation and then displayed together on the same plot. Filled colored circles represent the inverse mean initial velocity ($1/v_0$) at each value for 1/[NADP$^+$], I-bars indicate the standard deviation, and colored lines are the linear fits for each respective dataset. **Assay conditions:** [SSADH] = 0.25 μg/mL, [NADP$^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 μM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

can be made that increasing fixed [SSA] results in an increase in V_{max} (see **Figure 2.18B** in *Section 2.3.3.5*). In **Figure 2.17B**, where the data is transformed into reciprocal units and fit with **Equation 2.11** (Lineweaver-Burk), a pattern of lines intersecting in quadrant II is observed once more. Because neither Lineweaver-Burk plot (**Figure 2.16B** or **2.17B**) has lines that intersect on the ordinate, the *random ordered sequential mechanism* may be ruled out (see **Figure 2.3** in *Section 2.1.5*).

Analysis of Replots from Combined Datasets

Replots of V'_{max} (the apparent V_{max}) versus *fixed substrate* concentration are presented in **Figure 2.18**. **Figure 2.18A** is an alternative representation of the data in **Figure 2.16A**, rotated 90 ° about the v_0-axis, with the [SSA]-axis now compressed (pointing towards the observer)

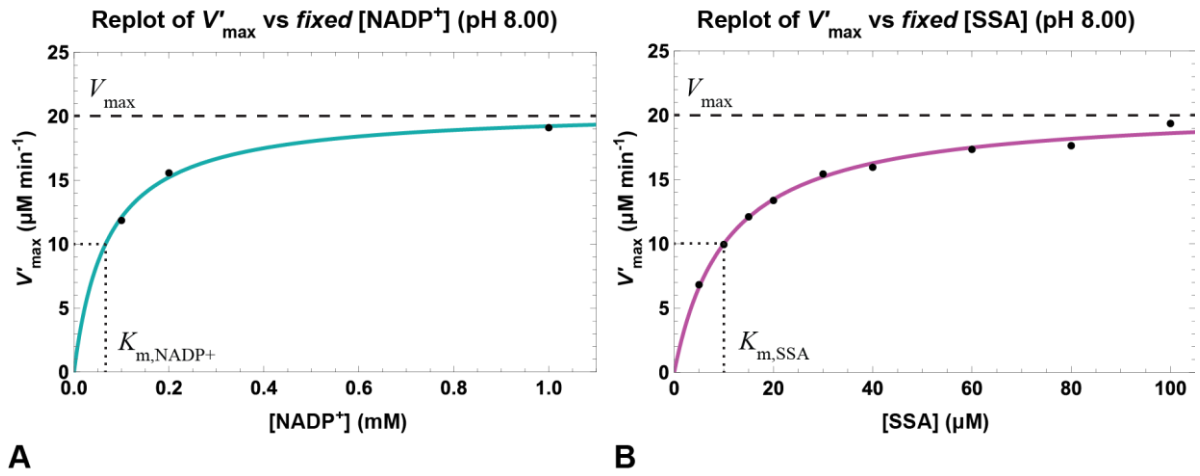

Figure 2.18. Replots of *apparent* maximal velocity (V'_{max}) vs *fixed* substrate concentration. (A) V'_{max} vs *fixed* values for [$NADP^+$]. Filled black circles (•) represent V_{max} values obtained from each set of assays where [SSA] was varied at a fixed [$NADP^+$], and the **teal line** is the nonlinear fit of the data. Dotted lines indicate the positions for V_{max} and $K_{m,NADP+}$. **(B)** V'_{max} vs *fixed* values for [SSA]. Filled black circles (•) represent V_{max} values obtained from each set of assays where [$NADP^+$] was varied at 9 different fixed [SSA], and the **magenta line** is the nonlinear fit of the data. Dotted lines indicate the positions for V_{max} and $K_{m,SSA}$. **Assay conditions:** [SSADH] = 0.25 μg/mL, [$NADP^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 μM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

Equation 2.12. Expression for V'_{max} when [SSA] is varied and [$NADP^+$] is fixed.

$$V'_{max} = \frac{V_{max}[NADP^+]}{K_{NADP+} + [NADP^+]}$$

Equation 2.13. Expression for V'_{max} when [$NADP^+$] is varied and [SSA] is fixed.

$$V'_{max} = \frac{V_{max}[SSA]}{K_{SSA} + [SSA]}$$

and the [$NADP^+$]-axis expanded along the *abscissa*. In other words, **Figure 2.18A** is a view looking *through* **Figure 2.16A** at V'_{max}, where [SSA] → ∞. At this point it starts to become clear that the initial velocity data for bireactant enzymes lies on the surface of a three-dimensional space.

Nonlinear regression of the data in **Figure 2.18A** using **Equation 2.12** provided a V_{max} of 20.58 ± 0.56 μM min^{-1} and K_{NADP+} of 0.070 ± 0.008 mM. Since these values were calculated

using much more data (v_0 for 5–100 μM SSA) than was presented in **Figure 2.15A** (100 μM SSA only), they provide better estimates for the true values of V_{max} and K_{NADP+}. For comparison, the estimates for V_{max} and K_{NADP+} from the nonlinear fit presented in **Figure 2.15A** were 19.36 ± 0.52 μM min^{-1} and 0.083 ± 0.008 mM, respectively.

Likewise, **Figure 2.18B** represents a *transection* of **Figure 2.17A** at V'_{max} (apparent V_{max}), where [$NADP^+$] → ∞. Nonlinear regression of the data in **Figure 2.18B**, using **Equation 2.13,** provided estimates of V_{max} = 20.57 ± 0.35 μM min^{-1} and K_{SSA} = 10.56 ± 0.68 μM. As these estimates resulted from the inclusion of all [$NADP^+$]-datasets (0.1, 0.2, 1.0 mM), they are more accurate than the V_{max} of 19.11 ± 0.29 μM min^{-1} and K_{SSA} of 11.09 ± 0.64 μM calculated from the 1.0 mM $NADP^+$-dataset presented in **Figure 2.14A**.

Replots of K'_m (the apparent K_m) versus *fixed substrate* concentration are presented in **Figure 2.19**. *Synergistic binding* is observed in both **Figures 2.19A** and **2.19B**, where the data was fit using **Equations 2.14** and **2.15**, respectively. This phenomenon was described in *Section 2.1.5* and illustrated by **Figure 2.4**. In **Figure 2.19A**, at low *fixed* [$NADP^+$], K_{SSA} increases towards the value of *equilibrium dissociation constant* for the enzyme-SSA complex (i.e., $K_{SSA} \rightarrow K_{iSSA}$), where K_{iSSA} = [E·SSA]/[E][SSA]. At high *fixed* [$NADP^+$], K_{SSA} decreases towards the value of the *Michaelis constant* for the substrate SSA (i.e., $K_{SSA} \rightarrow K_{m,SSA}$). Here, the interpretation of $K_{m,SSA}$ is not so straightforward, since the Michaelis constant is an aggregate of individual rate constants whose identities are dependent on the proposed mechanism. Similarly, at low *fixed* [SSA], $K_{NADP+} \rightarrow K_{iNADP+}$, and at high *fixed* [SSA], $K_{NADP+} \rightarrow K_{m,NADP+}$ (**Figure 2.19B**). Thus, each substrate increases the other's binding affinity in a synergistic manner.

Nonlinear regression of the data presented in **Figure 2.19A**, using **Equation 2.14**, resulted in the following parameters: K_{iSSA} = 16.36 µM, $K_{m,SSA}$ = 10.22 µM, K_{iNADP+} = 0.265 mM, and $K_{m,NADP+}$ = 0.165 mM. Note, while K_{iSSA} is not explicitly written in **Equation 2.14**, it was derived from the relationship in **Equation 2.6**. Statistics for the fitting could not be obtained

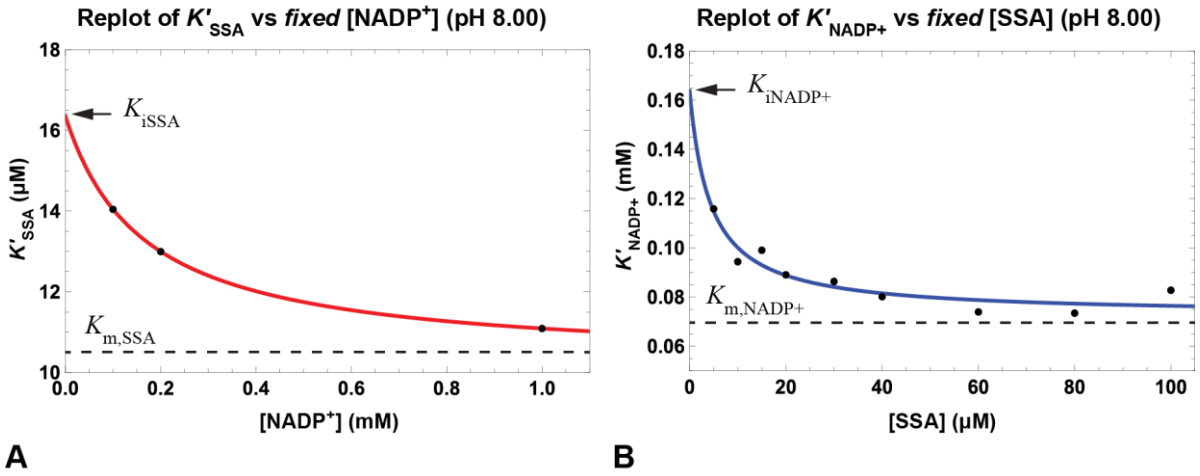

Figure 2.19. Replots of *apparent* Michaelis constant (K'_m) vs *fixed* substrate concentration. (A) K'_{SSA} vs *fixed* values for [$NADP^+$]. Filled black circles (•) represent the K'_{SSA} values obtained from distinct sets of assays where [SSA] was varied (not shown), at a fixed [$NADP^+$] values that were incrementally increased (along *x*-axis), and the red line is the nonlinear fit of the data. The arrow indicates the estimation for K_{iSSA}; dotted line are the estimation for $K_{m,SSA}$. **(B)** K'_{NADP+} vs *fixed* values for [SSA]. Filled black circles (•) represent the K'_{NADP+} values obtained from distinct sets of assays where [$NADP^+$] was varied (not shown), at fixed [SSA] values that were incrementally increased (along *x*-axis), and the blue line is the nonlinear fit of the data. The arrow indicates the estimation for K_{iNADP+}; dotted line are the estimation for $K_{m,NADP+}$. **Assay conditions:** [SSADH] = 0.25 µg/mL, [$NADP^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 µM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

Equation 2.14. Expression for K'_{SSA} when [SSA] is varied and [$NADP^+$] is fixed.

$$K'_{SSA} = \frac{K_{iNADP+}K_{m,SSA} + K_{m,SSA}[NADP^+]}{K_{m,NADP+} + [NADP^+]}$$

Equation 2.15. Expression for K'_{NADP+} when [$NADP^+$] is varied and [SSA] is fixed.

$$K'_{NADP+} = \frac{K_{iNADP+}K_{m,SSA} + K_{m,NADP+}[SSA]}{K_{m,SSA} + [SSA]}$$

because the number of fitting parameters in the model was not less than the number of data points. Apart from the value for K_{SSA} (asymptote parallel to the abscissa), the fit does not provide reliable values for the remaining constants. This is both because there are too few data points and, more importantly, no $NADP^+$ concentrations below the K_m for $NADP^+$ were used during the rate experiment, a fact clearly illustrated in **Figure 2.18A**. Nonlinear regression of the data presented in **Figure 2.19B**, using **Equation 2.15**, resulted in the following parameters: K_{iSSA} 9.71 ± 11.32 μM, $K_{m,SSA}$ = 4.31 ± 4.78 μM, K_{iNADP+} = 0.164 ± 0.057 mM, and $K_{m,NADP+}$ = 0.072 ± 0.004 mM, where K_{iSSA} was derived from the fitting parameters using **Equation 2.6**. Similar to fit in **Figure 2.19A**, the value for K_{NADP+} (asymptote parallel to the abscissa) is estimated well, but the remaining values are unreliable because too few SSA concentrations below the K_m for SSA were used in the rate experiment. The main value of the replots is in the graphical way they illustrate that equilibrium dissociation constants are greater than the Michaelis constants when synergistic binding occurs ($K_S > K_m$). The actual magnitude of the synergism is quantitatively determined in the following section.

Global Least-Squares Regression of Initial Rate Data

A combined dataset, represented by a 27 × 3 matrix, was constructed by combining all three v_0 versus [SSA] datasets (9 × 2 matrices), at fixed [$NADP^+$] of 0.1, 0.2, and 1.0 mM. Global least-squares (multivariable) regression was performed on the combined dataset, using both **Equation 2.3** and **Equation 2.5** (*Section 2.1.5*). The results for the simultaneous fitting of v_0, [SSA] and [$NADP^+$] are presented as a three-dimensional surface plot in **Figures 2.20A** and **2.21A**, along with their companion two-dimensional projections in **Figures 2.20B** and **2.21B**.

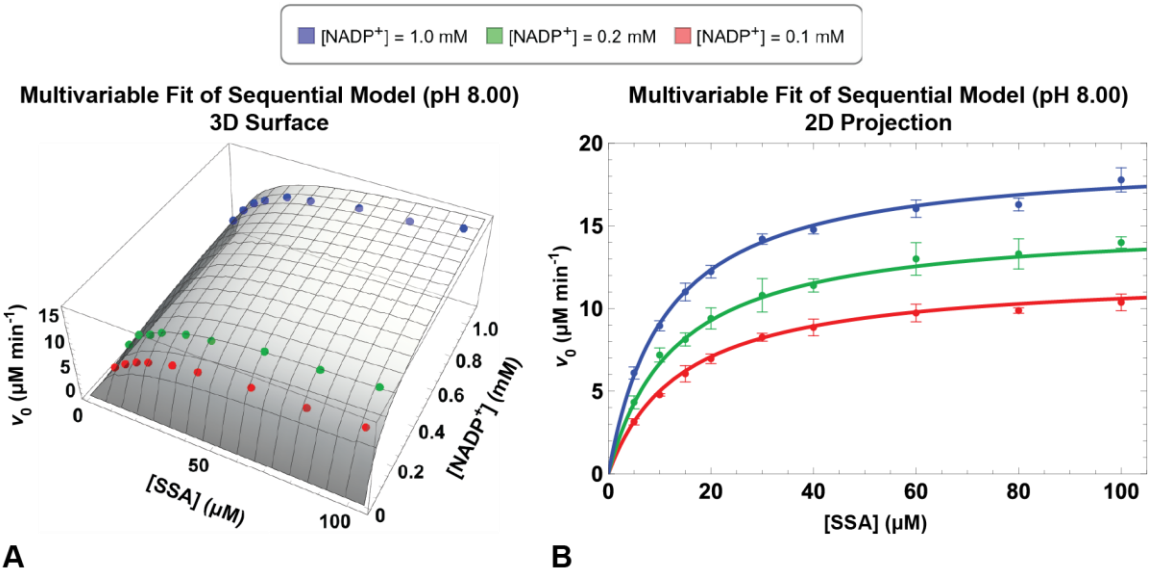

Figure 2.20. Multivariable regression: 3D surface and v_0 vs [SSA] 2D projections. (A) 3D surface from the multivariable nonlinear regression of the [[$NADP^+$]$_x$, [SSA]$_y$, $(v_0)_z$]-matrix. Colored circles represent the mean initial velocities (v_0) used for the fit. **(B)** Compression of the *x*-axis, representing [$NADP^+$], results in a 2D projection showing v_0 vs [SSA] at fixed [$NADP^+$]. Colored circles represent mean initial velocities (v_0), I-bars indicate the standard deviation, and the colored lines are slices through the 3D surface at each fixed [$NADP^+$]. **Assay conditions:** [SSADH] = 0.25 µg/mL, [$NADP^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 µM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

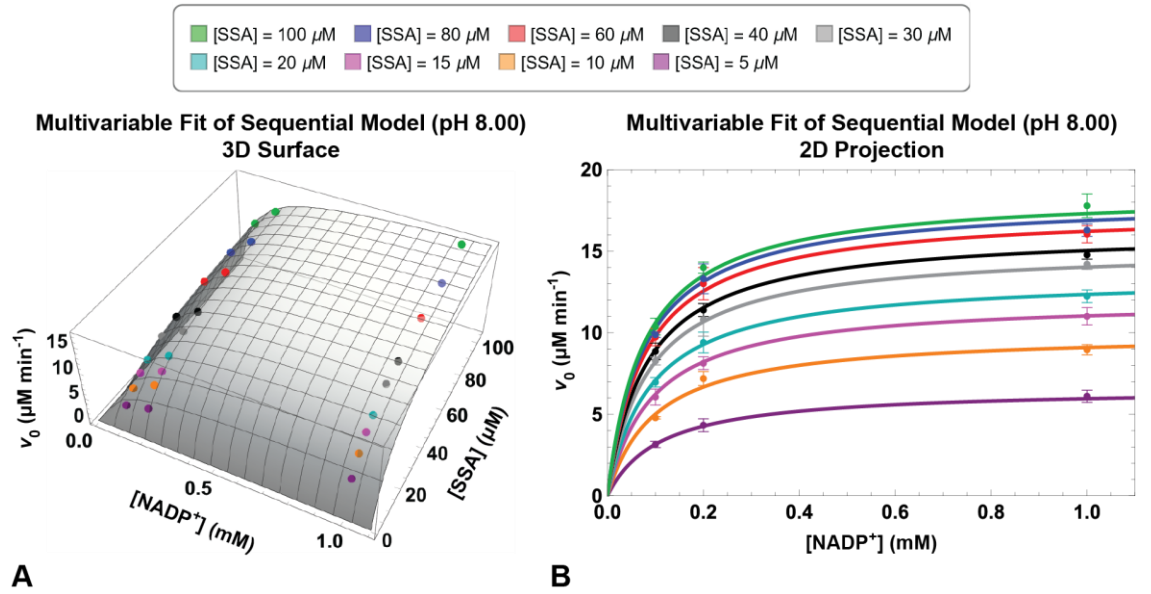

Figure 2.21. Multivariable regression: 3D surface and v_0 vs [$NADP^+$] 2D projections. (A) 3D surface from the multivariable nonlinear regression of the [[SSA]$_x$, [$NADP^+$]$_y$, v_{0z}]-matrix. Colored circles represent the mean initial velocities (v_0) used for the fit. **(B)** Compression of the *x*-axis, representing [SSA], results in a 2D projection showing v_0 vs [$NADP^+$] at fixed [SSA]. Colored circles represent mean initial velocities (v_0), I-bars indicate the standard deviation, and the colored lines are slices through the 3D surface at each fixed [SSA]. **Assay conditions:** [SSADH] = 0.25 µg/mL, [$NADP^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 µM, 100 mM sodium phosphate, pH 8.00, 30 °C. Assays performed in triplicate after 3 min pre-incubation.

The difference between **Figures 2.20** and **2.21** is only the orientation of the [SSA] and [$NADP^+$] axes, with the aim to show the origin of the 2D projections that bear striking similarities to the combined plots presented earlier (**Figures 2.16A** and **2.17A**). The plots are similar, but not identical, as the multivariable regression on the larger dataset produces subtle differences in fitting, with slightly more accurate results. In comparing **Figures 2.20B** and **2.16A**, for [$NADP^+$] = 0.2 mM (**green** filled circles and lines), the positions of the 40 μM SSA data points with respect to the rectangular hyperbola are not the same in these two figures. More dramatically, in **Figure 2.21B** the spacing between the rectangular hyperbolas decreases uniformly from 5 mM to 100 mM SSA, while their spacing in **Figure 2.17A** appears random.

The parameters for the constants resulting from regression using **Equations 2.3** and **2.5**, along with the fitting statistics, are presented in **Table 2.4**. While the derivations of these rate equations and their final mathematical expressions are not the same, both equations produce identical values for the kinetic parameters, albeit with minor differences in presentation. **Equation 2.3** was derived using the *rapid equilibrium assumption*, so it is not surprising that it returns only values for the *equilibrium dissociation constants*, K_{SSA} and K_{NADP+}, along with

Table 2.4. Kinetic parameters from multivariable nonlinear regression of pH 8.00 data.

Rapid Equilibrium Random Model, with the *α*-term (Equation 2.3)						
		Dissociation constants		*Michaelis constants* (*α* × *Dissociation constants*)		
Parameter	V_{max} (μM min^{-1})	K_{SSA} (μM)	K_{NADP+} (mM)	αK_{SSA} (μM)	αK_{NADP+} (mM)	*α*
Estimate	20.57 ± 0.34	19.24 ± 3.15	0.13 ± 0.02	10.58 ± 0.69	0.070 ± 0.005	0.55 ± 0.12
% Error	1.67%	16.37%	15.21%	6.54%	7.45%	21.41%
Sum of Squared Errors = 1.89						
Sequential Model, with four independent constants (Equation 2.5)						
		Dissociation constants		*Michaelis constants*		
Parameter	V_{max} (μM min^{-1})	K_{iSSA} (μM)	K_{iNADP+} (mM)	$K_{m,SSA}$ (μM)	$K_{m,NADP+}$ (mM)	
Estimate	20.57 ± 0.34	19.24 ± 3.15	0.13 ± 0.02	10.58 ± 0.69	0.070 ± 0.005	
% Error	1.67%	16.37%	15.21%	6.54%	7.45%	
Sum of Squared Errors = 1.89						

a value for the α-term. The *Michaelis constants* for SSA and $NADP^+$, $K_{m,SSA}$ and $K_{m,NADP+}$, can be obtained by multiplying each dissociation constant by α. The *steady-state assumption* was used to derive **Equation 2.5**, thus its mathematical expression contains explicit terms for the *Michaelis constants* $K_{m,SSA}$ and $K_{m,NADP+}$, in addition to the *equilibrium dissociation constants*, here termed K_{iSSA} and K_{iNADP+}. One advantage of using the α-term is that no direct comparison between K_s and K_m is required to determine if synergistic binding is occurring, if the binding sites are truly independent, or if there is competition between the substrates; one need only look at the value for the α-term (*Section 2.1.5*). Additionally, the magnitude of α-term may be a useful parameter when comparing homologous enzymes with rapid equilibrium random mechanisms, or when probing such an enzyme's active site using site-directed mutagenesis.

Table 2.5 compares the kinetic parameters from multivariable nonlinear regression with those values produced by methods from the previous sections. With the appropriate rate equation, multivariable nonlinear regression on the complete dataset remains the preferred

Table 2.5. Enzyme kinetics parameters from different methods of data analysis.

Method	Fixed Substrate	V_{max} (μM min^{-1})	K_{iSSA} (μM)	$K_{m,SSA}$ (μM)	K_{iNADP+} (mM)	$K_{m,NADP+}$ (mM)
Single substrate	$NADP^+$	19.11 ± 0.29		11.09 ± 0.64		
	SSA	19.36 ± 0.52				0.083 ± 0.008
V′ replots	$NADP^+$	20.58 ± 0.56				0.070 ± 0.008
	SSA	20.57 ± 0.35		10.56 ± 0.68		
K′ replots*	$NADP^+$		16.36	10.22	0.26	0.17
	SSA		9.71 ± 11.32	4.31 ± 4.78	0.16 ± 0.06	0.073 ± 0.005
Global		20.57 ± 0.34	19.24 ± 3.15	10.58 ± 0.69	0.13 ± 0.02	0.070 ± 0.005

*Statistics were not possible for $NADP^+$ dataset; number of data points was less than number of parameters.

method for obtaining kinetic parameters. Provided the data is complete, *V′* and *K′* replots can also generate comparable values for the kinetics parameters. However, replots should not be trusted when key data is missing. As illustrated in **Table 2.5**, *K′* replots can generate very poor

estimates for constants when an insufficient number of substrate concentrations below the K_m are sampled during the rate experiments.

Using the relationship of $k_{cat} = V_{max}/[E]_T$, the V_{max} value of 20.57 ± 0.34 μM min^{-1} from the multivariable nonlinear regression, and the total enzyme concentration $[E]_T$ of 0.00463 μM (0.25 μg/mL) used in the rate experiments, the k_{cat} for SSADH was 74.12 ± 1.24 s^{-1} at pH 8.0 and 30 °C.

Comparison with Literature Values

The only enzyme kinetics data in the literature for *E. coli* SSADH (*gabD*) belongs to an X-ray crystal structure paper, published in 2010 by the same laboratory that generously donated the expression vector used in the present research. The experimental conditions and results from both the literature and the present research are compared in **Table 2.6**. It should

Table 2.6. Enzyme kinetics from literature and present research.

	Parameter	Literature	Present Research
Conditions	Buffer	100 mM sodium phosphate	100 mM sodium phosphate
	pH	8.0	8.00 (and 7 other pH values)
	Temperature	30 ° C	30 ° C
	[SSADH]	0.037 μM (2 μg/mL)	0.0046 μM (0.25 μg/mL)
	[SSA]	0–400 μM	5–100 μM
	[$NADP^+$]	1.1 mM	0.1, 0.2, 1.0 mM
	Detection	Fluorescence (355 nm/460 nm)	Ultraviolet (340 nm)
Results	V_{max}	40.92 ± 1.3 μM min^{-1} (± 3.2%)	20.57 ± 0.34 μM min^{-1} (± 1.7%)
	k_{cat}	18.43 ± 0.59 s^{-1} (± 3.2%)	74.12 ± 1.24 s^{-1} (± 1.7%)
	K_{iSSA}	—	19.24 ± 3.15 μM (± 16.4%)
	$K_{m,SSA}$	16.94 ± 2.2 μM (± 13.0%)	10.58 ± 0.69 μM (± 6.5%)
	K_{iNADP+}	—	0.13 ± 0.02 mM (± 15.4%)
	$K_{m,NADP+}$	—	0.070 ± 0.005 mM (± 7.1%)

be noted that the V_{max} value in the publication had units of μM, a likely typographical error. The authors were contacted by email for clarification, but they did not reply. Assuming units

of μM min^{-1}, the published V_{max} is double the value reported here, which is not surprising since the $[E]_T$ reportedly used was also eight times higher.

Although the authors did not explicitly report k_{cat}, it was calculated from V_{max} and $[E]_T$ and found to be four times lower than the k_{cat} reported here for same enzyme under nearly the same experimental conditions. One explanation is that the enzyme used in the published research had only one quarter of the activity observed here. It is possible that the enzyme in the present research was purified more thoroughly than the enzyme in the publication, resulting in a higher specific activity. It's also possible that the time-dependent degradation of the enzyme in the present study was much slower, a factor that can be influenced by purity, concentration, handling and storage. Other explanations may lie with differences in how the enzyme assay was developed and performed. For the present research, optimization of the enzyme assay was essential before measuring the initial velocities. This included determining the amount of time required for the reaction components to reach thermal equilibration before initiating the reaction. If the reaction starts before the components reached 30 °C, the rates will appear attenuated. Another explanation for why the published k_{cat} might appear much lower stems from the excessively high substrate concentration (400 μM SSA) in the study, which is four times higher than the one used in the present research (100 μM). Substrate inhibition at high SSA has been reported for SSADH from other organisms.[13] If initial velocity decreased due to inhibition at very high substrate concentrations, and those data points were included in the fitting, the V_{max} would have been underestimated (note the error in V_{max}).

The literature value for $K_{m,SSA}$ is close to the value reported here, although its standard deviation was also twice as large compared to the one in the present study. The publication does not list individual values for the [SSA] used in the rate experiments, only a range from 0

to 400 μM. For the experiments reported here, 5 μM SSA was the lowest substrate concentration that allowed for reliable initial velocity measurements (≤ 5% substrate consumed at t = 0). Even then, this was only possible when using only 0.25 μg/mL SSADH and a procedure that initiated the reaction and started data acquisition in less than 3 seconds. Without a specialized instrument designed for rapid kinetics measurements (i.e., stopped-flow), it would be a significant challenge obtaining reasonably accurate initial velocity measurements at [SSA] values, at or below the K_m, using eight-times more enzyme, as reported in the publication.

It is not this author's intention to diminish the valuable experimental contributions in the 2010 publication, only to propose explanations for the differences that exist between results collected under similar conditions. In defense, the focus of the publication was the X-ray crystal structure of SSADH. The enzyme kinetics were not performed in order to obtain accurate values for k_{cat} and K_m, but to verify that the crystalized enzyme possessed activity. For this purpose, rigorous enzyme kinetics were not required. Lastly, statistical variation and experimental error may cause identical rate experiments conducted on different days to produce V_{max} and K_m values that vary by as much as 10% and 30-50%, respectively.[14]

pH-Rate Dependence, pH Optimum, and pK_a

The previous sections describe the data analysis on SSADH rate experiments conducted at pH 8.0. In order to examine the effect of pH on enzyme catalysis, the same data was collected between pH 6.50 and 8.25, at quarter-pH increments. This range was selected as it is both biologically relevant and compatible with the use of sodium phosphate as the buffer. A summary of kinetic parameters for the pH-rate profile is presented in **Table 2.7**. Only the reaction rate (V_{max} and k_{cat}) displayed a pH dependence that could be easily interpreted in such

a way as to provide insight into the catalytic mechanism. The pH dependencies of K_{iSSA} , K_{iNADP+}, $K_{m,SSA}$ and $K_{m,NADP+}$ are complex, reflecting the interaction between several functional groups that are capable of multiple ionic states. As such, the resulting K versus pH plots do not lend themselves to straightforward interpretation and the same holds true of plots derived from these constants, such as V/K versus pH.

The pH-rate profile for SSADH, using the k_{cat} data in **Table 2.7**, is shown in **Figure 2.22A**. Transformation into $\log_{10}k_{cat}$ versus pH and nonlinear regression with **Equation 2.16** is shown in **Figure 2.22B**. From the pH-rate profile, the pH optimum appears to be ≥ pH 8.25. The exact optimum cannot be stated with certainty because pH 8.25 is the extreme end of phosphate buffering capacity. Accurate determination of the pH optimum would require the use of a buffer such as bicine, which can buffer between pH 7.75 and 9.0, but does not contain a primary amine that would react with the aldehyde moiety of the substrate. Assuming that the ionic strengths of two different buffer systems are similar, the new data could be overlapped to use the existing information. Another option would be to use a combination buffer system that could span a much larger pH range.[15]

Table 2.7. Summary of kinetic parameters from pH-rate profile. Values with magnitudes that do not follow the trend in pH for their respective columns are highlighted in red.

pH	V_{max} (μM min^{-1})	k_{cat} (s^{-1})	K_{iSSA} (μM)	$K_{m,SSA}$ (μM)	K_{iNADP+} (mM)	$K_{m,NADP+}$ (mM)
8.25	20.68 ± 0.41	74.52 ± 1.47	17.86 ± 4.01	11.66 ± 0.85	0.088 ± 0.018	0.057 ± 0.005
8.00	20.57 ± 0.34	74.12 ± 1.24	19.24 ± 3.15	10.58 ± 0.69	0.13 ± 0.02	0.070 ± 0.005
7.75	18.29 ± 0.25	65.89 ± 0.91	20.51 ± 2.75	9.82 ± 0.55	0.14 ± 0.02	0.067 ± 0.004
7.50	16.28 ± 0.27	58.64 ± 0.97	23.65 ± 3.66	9.38 ± 0.65	0.16 ± 0.02	0.065 ± 0.005
7.25	13.61 ± 0.15	49.05 ± 0.53	22.98 ± 2.22	9.14 ± 0.42	0.18 ± 0.02	0.072 ± 0.004
7.00	10.38 ± 0.11	37.40 ± 0.38	18.09 ± 1.93	9.47 ± 0.40	0.12 ± 0.01	0.064 ± 0.003
6.75	7.42 ± 0.08	26.72 ± 0.30	23.12 ± 2.83	6.74 ± 0.38	0.17 ± 0.02	0.049 ± 0.003
6.50	5.13 ± 0.05	18.50 ± 0.20	14.87 ± 2.33	5.33 ± 0.31	0.11 ± 0.02	0.038 ± 0.003

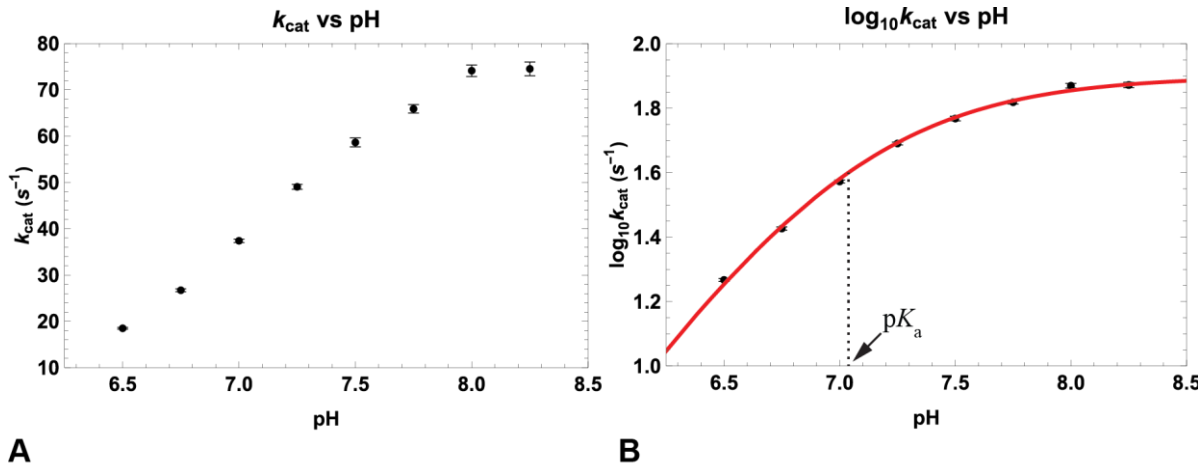

Figure 2.22. pH-Rate profile for SSADH. (A) k_{cat} vs pH data from SSADH enzyme kinetics assays. Filled black circles (•) represent k_{cat} values obtained from multivariable regression of data at different pH values, with I-bars indicating the standard error from each fit. **(B)** $\log_{10} k_{cat}$ vs pH. Filled black circles (•) represent $\log_{10} k_{cat}$ values, with I-bars indicating the propagation of standard error. The nonlinear fit of the data using **Equation 12** is represented by the **red** line. **Assay conditions:** [SSADH] = 0.25 µg/mL, [$NADP^+$] = 0.1, 0.2, and 1.0 mM, [SSA] = 5–100 µM, 100 mM sodium phosphate, 30 °C, at pH 6.50, 6.75, 7.00, 7.25, 7.50, 7.75, 8.00, and 8.25. Assays performed in triplicate after 3 min pre-incubation.

Equation 2.16. Equation for nonlinear regression of $\log_{10}k_{cat}$ vs pH data.

$$\log_{10} k_{cat} = \log_{10}\left(\frac{k_{cat}}{1+\frac{10^{-pH}}{10^{-pKa}}}\right)$$

The parameters from nonlinear regression of $\log_{10}k_{cat}$ versus pH with **Equation 2.16** are listed in **Table 2.8**. The value for k_{cat} is the theoretical turnover number at the pH optimum. Of more importance is the pK_a value of 7.03 ± 0.01. Since k_{cat} represents every step that occurs between formation of the enzyme-substrate complex and reformation of free enzyme, this pK_a value must belong to a titratable group that impacts one of these steps. Neither the SSA substrate nor the $NADP^+$ coenzyme have an ionizable group with a pK_a near 7.0, so this titratable group must belong to an amino acid side chain in the SSADH active site, with the likely candidate being cysteine 288.

Table 2.8. Parameters from nonlinear regression of pH-rate data using Equation 12.

k_{cat} (s^{-1})	k_{cat} % Error	pK_a	pK_a % Error	R^2	Sum of Squared Errors
79.40 ± 1.10	1.38%	7.03 ± 0.01	0.18%	0.9999	0.0006

The pK_a of cysteine has been measured to be 8.55 ± 0.03 in alanine pentapeptides[16], and this is often given as the textbook value. However, a number of environmental factors in the active site of enzymes may perturb side chain pK_a values dramatically, including desolvation effects, Coulombic interactions, and hydrogen bonding.[17] In an analysis of 25 folded proteins in the literature for which the pK_a of cysteine was available, the average pK_a value was determined to be 6.8 ± 2.7.[18, 19] The pK_a a value of 7.03 obtained from the pH-rate profile is less than one quarter-pH unit away from the average pK_a for cysteine in folded proteins. Additionally, the published X-ray crystal structure for *E. coli* SSADH (*gabD*) identifies Cys288 as the nucleophilic side chain that attacks the aldehyde carbonyl carbon of the substrate in the first step of the catalytic mechanism (**Figure 1.7**, *Section 1.4.1*). The cytoplasmic pH of *E. coli* is estimated to be in the range of ~7.5–7.7.[20, 21] In this mildly alkaline environment, lowering the pK_a of cysteine from 8.6 to 7.0 would favor deprotonation of the Cys288 sulfhydryl to the more nucleophilic thiolate. The net result would be a shift in the population of enzymes that are in the correct ionization state for catalytic activity, thus increasing the rate of catalysis.

2.3 Conclusion

Recombinant E. coli $NADP^+$-dependent SSADH (*gabD*) containing an N-terminal His-tag was overexpressed in the E. coli BL21(DE3)pLysS cell line, resulting in an estimated cellular abundance of approximately 30% of all soluble protein. During purification, immobilized metal affinity chromatography resulted in the largest increase in specific activity, with only a marginal improvement in purity coming from the subsequent size exclusion chromatography (SEC) step. SEC did provide the opportunity for a buffer exchange, while a molecular weight standard curve calibrated to the column showed evidence of multiple oligomerization states for SSADH. Characterization by SDS-PAGE, under reducing and denaturing conditions, and mass spectrometry revealed a homogeneous sample that had an apparent molecular weight of 54 kDa. The final purified enzyme had a specific activity of 34.74 U/mg, which was approximately 4-fold higher than what has been previously reported under comparable conditons.

A steady state enzyme assay was developed for SSADH, with careful attention paid to optimal enzyme and substrate concentrations as well as sufficient thermal equilibration prior to initiation of the reaction. During optimization SSADH displayed a strong preference for $NADP^+$ over NAD^+ as the coenzyme, with an 83% decrease in enzyme activity observed at pH 7.5 when NAD^+ was used in place of $NADP^+$. The steady state enzyme kinetics involved measuring initial velocities while varying SSA substrate concentration, at multiple fixed concentrations of the $NADP^+$ coenzyme. These assays were carried out in triplicate at 30 °C, and at seven pH values between 6.5 – 8.25. Global least-squares regression of the data at pH 8.0 yielded 74.1 ± 1.2 s^{-1}, 10.6 ± 0.7 μM, and 0.13 ± 0.02 mM for the kinetic constants k_{cat}, K_m(SSA), and K_m($NADP^+$), respectively. These values differed from the literature values for

the reaction at pH 8.0 under similar conditions. The largest difference was between the k_{cat} value of 74.1 ± 1.2 s^{-1} in the present study and the literature value of 18.4 ± 0.6 s^{-1}, a 4-fold difference. Possible reasons for the difference in measured k_{cat} can be attributed to a number of things, including purity of the enzyme sample, degradation of the enzyme, lack of thermal equilibrium, and enzyme inhibition at high substrate concentrations.

Analysis of the reaction kinetics as a function of pH revealed that the enzyme requires a group with a pK_a of 7.03 ± 0.01 to be deprotonated for catalytic activity and displays a pH optimum of ≥ 8.25. The identity of this group is likely Cys288, which is the nucleophile (thiolate) in a proposed mechanism that attacks the carbonyl carbon of the SSA substrate in the first step. The pK_a of cysteine in alanine pentapeptides is 8.6, while in folded enzymes it has been estimated to be 6.8 ± 2.7. Perturbing the cysteine pK_a to a lower value would increase the fraction of deprotonated cysteines and be an advantage to enzymes that employ cysteine nucleophiles or disulfides bridges for stability.

Analysis of the rate profiles for SSADH reveals a clearly defined sequential mechanism. *Steady state ordered* and *steady state random* have been ruled out since the plotted data do not fit the patterns for those mechanisms. *Steady state Theorell-Chance* is also ruled out because the pH-dependence of k_{cat}, and pK_a assignment to the nucleophilic Cys288 suggests a ternary complex is present. The final possibilities are *rapid equilibrium random* or *steady state ordered*. Synergistic binding suggests no preference in one substrate binding before the other and the X-ray structure shows two distinct binding sites for each substrate, each on the opposite side of the enzyme from the other. While these points do not preclude an order mechanism, they slightly favor a rapid equilibrium random mechanism, but only product inhibition or isotope trapping experiments would be able to say with certainty which mechanism is correct.

2.4 Materials and Methods

2.4.1 Bacterial Transformation and Selection

pKN12 Vector

The expression vector for recombinant SSADH was provided courtesy of Prof. James C. Whisstock of the Department of Biochemistry and Molecular Biology at Monash University, Melbourne, Victoria, Australia.[5] Briefly, the expression vector was based on the Invitrogen™ pRSET-C plasmid, which includes a high copy number ColE1 origin of replication, an ampicillin resistance gene (*bla* gene; β-lactamase; Amp^R) for selection, a bacteriophage T7 RNA polymerase promoter for high expression, an N-terminal 6×His-tag, and an enterokinase (EK) cleavage site for removal of the 6×His fusion tag. Primers were used to modify the pRSET-C plasmid, replacing the EK cleavage site and other intervening sequences ahead of the multiple cloning site (MCS) with the Tobacco Etch Virus (TEV) protease cleavable peptide linker.[22] The *gabD* gene from *E. coli* was ligated into the MCS of pRSET-C/His-TEV to produce the final expression vector, pKN12, which now encoded the 6×His-TEV-SSADH expression product.

XL10-Gold Transformation

The pKN12 vector, which was supplied dried onto filter paper, was taken up in 1X TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and the solubilized plasmid concentration was determined using a NanoDrop. 5 ng of DNA was used to transform a vial containing 50 µL Agilent Technologies XL10-Gold® Ultracompetent *E. coli* cells by following the manufacturer's 30-second heat-pulse transformation protocol. Cells were allowed to recover

in SOC media for 1 hour on a shaker (200 RPM) at 37 °C. 50 µL and 100 µL volumes of cells were streaked onto on LB-agar plates, supplemented with 100 µg/mL of ampicillin (+100 Amp) for selection, and incubated upside down for 24 hours at 37 °C. The transformation efficiency was 1×10^7 cfu/µg. A colony from the plate was picked and grown in 50 mL 2YT broth pH 7.0 (+100 Amp) while shaking (200 RPM) at 37 °C. The bacterial growth was estimated by measuring the optical density at 600 nm (OD_{600}). 0.5 mL of cells was removed at $OD_{600} \approx 0.4$ and added to 0.5 mL of 50% sterile glycerol in a cryotube to produce 25% glycerol stock that was stored at -80 °C. The remainder of the growth continued to shake overnight at 37 °C and a miniprep of the cell pellet resulted in 50 µL of plasmid DNA with a concentration of 317 ng/µL and A_{260}/A_{280} of 1.95. DNA sequencing confirmed the colony contained the pKN12 plasmid encoding 6×His-TEV-SSADH.

BL21(DE3)pLysS Transformation

10 ng of miniprepped pKN12 plasmid was subsequently transformed into Invitrogen™ One Shot® BL21(DE3)pLysS competent *E. coli* cells following the manufacturer's 30-second heat-pulse transformation protocol. Cells were allowed to recover in SOC media as before. BL21(DE3)pLysS cells are used for protein expression and contain the pLysS plasmid that encodes the T7 lysozyme for transcriptional control (see below). pLysS also encodes resistance to chloramphenicol (*cat* gene; chloramphenicol acetyltransferase; Cam^R). Therefore selection of BL21(DE3)pLysS transformants on LB-agar plates required 36 µg/mL of chloramphenicol (+36 Cam) to maintain the pLysS plasmid, in addition to the 100 µg/mL of ampicillin required to select for pKN12. After plating suspended cells and incubating as described earlier, a transformation efficiency of 5×10^6 cfu/µg was observed. A colony from the plate was grown at 37 °C in pH 7.0 2YT broth (+100 Amp/36 Cam) with shaking (200 RPM) until $OD_{600} \approx 0.4$,

at which point 0.5 mL of suspended cells were added to 50% sterile glycerol in a cryotube to make a 25% glycerol stock that was stored at -80 °C.

T7 lysozyme encoded by pLysS catalyzes the hydrolysis of β-(1,4)-glycosidic bonds between *N*-acetylmuramic acid and *N*-acetyl-D-glucosamine residues in the peptidoglycan layer that surrounds the *E. coli* plasma membrane. T7 lysozyme is also a natural inhibitor of T7 RNA polymerase. The small amount of T7 lysozyme supplied by the pLysS plasmid inhibits basal expression of the target gene prior to induction, thereby increasing the tolerance of the *E. coli* host for potentially toxic inserts. Addition of the inducer isopropyl β-D-1-thiogalactopyranoside (IPTG) is sufficient to overcome inhibition of T7 RNA polymerase by T7 lysozyme and begin transcription. Because T7 lysozyme is unable to pass through the plasma membrane, the peptidoglycan layer is unaffected by its production within the cell and no cell lysis is observed under normal conditions. However, the presence of pLysS does facilitate the easy preparation of cell extracts when the *E. coli* cells are subjected to freeze-thaw cycles or high detergent concentrations, both of which will disrupt the plasma membrane enough to allow T7 lysozyme access to the peptidoglycan layer.

2.4.2 The Expression of Recombinant SSADH

Expression of recombinant SSADH was readily accomplished by growing pKN12-transformed BL21(DE3)pLysS cells in Fernbach flasks with 300 mL 2YT pH 7.0 media supplemented with 100 µg/mL of ampicillin and 36 µg/mL chloramphenicol. A 5 mL 2YT pH 7.0 (+100 Amp/36 Cam) overnight culture was started by scraping a pipette tip over the surface of the frozen BL21(DE3)pLysS/pKN12 glycerol stock and ejecting the tip into the culture tube. The following day the overnight was used to seed the 300 mL 2YT, approximately 1:100, until the starting OD_{600} was between 0.05 and 0.10. Cells were initially grown at 37 °C on a shaker

set to 200 RPM to aerate the media. The cells were induced by addition of 0.5 mM IPTG once they entered log phase ($OD_{600} \approx 0.5$), at which point the temperature was lowered to 16 °C and the cells continued shaking for 18 hours before harvesting. At the end of this period the suspension had a measured $OD_{600} \approx 2.3$ and the cells were centrifuged for 10 min at 5000 × *g* and 4 °C. After discarding the supernatant, the pellet washed by resuspending in phosphate buffered saline pH 7.4, repeating the centrifugation and discarding the supernatant. The washed cell pellet, with a mass of 3.66 g, was stored at -80 °C until required for lysis and purification.

2.4.3 Lysis of BL21(DE3)pLysS Cells

Although BL21(DE3)pLysS cells produce a small amount of T7 lysozyme, a catalytic amount of lysozyme (1 mg/mL), along with DNase I (1 mg/mL), was added to a thawed cell pellet (3.66 g) that had been suspended in an equivalent volume of lysis buffer on ice. The composition of the lysis buffer was 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.1% Triton X-100. After gentle agitation for 30 minutes, lysis was achieved by two successive freeze-thaw cycles using liquid nitrogen. DNase I is used to mitigate the viscosity produced by the genomic DNA released during lysis and it is essential that DNase I have time to work before adding BME to the crude lysate. This is because BME will reduce disulfide bonds critical for the conformational stability of DNase I, inhibiting its catalytic activity. This point must be balanced by the need to prevent oxidation of the catalytic cysteine in the SSADH active site. Thus, 5 mM BME was added after the second and final thaw, as soon as the crude lysate behaved as a homogeneous liquid that could be easily pipetted without clogging tips. High-speed centrifugation (30 min at 50,000 × *g* 4 °C) was used to clarify the lysate sufficiently enough for direct application to chromatographic resin without additional preparation, such as ammonium sulfate precipitation. A sample was preserved for analysis.

2.4.4 Protein Purification

Immobilized Metal Affinity Chromatography (IMAC)

HisPur™ Ni-NTA resin from Thermo Scientific was loaded into a 1.0 × 10.0 cm Bio-Rad Econo-Column® to a bed height of 5.0 cm (3.93 mL) and a Bio-Rad Econo-Column® Flow Adaptor was used to pack the column and load the sample. The adapter was connected through tubing to a Bio-Rad Model EP-1 Econo Pump, which was fed by a GibcoBRL/LifeTechnologies™ Model 750 gradient former. A T-valve with Luer-Lok™ fittings was installed part way between the adapter and the pump to interrupt the mobile phase and inject the sample directly onto resin, via the flow adapter. The gradient former (used to make gels) had an inner and outer chamber that held buffers, with a valve that controlled the flow between them. The inner chamber led directly to the pump and was fitted with a stir bar, while the outer chamber was the furthest reservoir from the pump. A dilute solution could be placed in the inner chamber, and an equal volume of concentrated solution in the outer chamber. Uniform mixing was achieved by opening the valve between the two chambers, allowing the more concentrated solution to flow into the inner chamber where the stir bar would mix the two before the combined solution was pumped on to the column. Fractions from the column were collected in VWR disposable borosilicate glass culture tubes (13 × 100 mm) that were placed into a Bio-Rad Model 2110 Fraction Collector. The flow through fraction was collected in a single 50 mL Falcon tube, but the remainder of the IMAC fractions were collected in the glass culture tubes, with 4.34 mL for the sample (twice the sample volume)

In a refrigerated room, the resin on the column was equilibrated with 5 column volumes of 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol (BME), with a flow rate of 0.5 mL/min. The clarified cell lysate was then slowly loaded onto

the column with a syringe that was attached to a Luer-Lok port on the T-valve in the direction that led to the resin. Once the valve was closed, 2 column volumes of equilibration buffer were used to move the sample onto the column at a flow rate of 0.25 mL/min, which would remain the flow rate for the duration of the purification. The column was then washed with 10 column volumes of 50 mM Tris, pH 8.0, 150 mM NaCl, 25 mM imidazole, and 5 mM BME. Lastly, the target protein was eluted with 10 column volumes of 50 mM Tris, pH 8.0, 300 mM NaCl, 25–250 mM imidazole, and 5 mM BME, which was accomplished by filling the inner chamber with the 25 mM imidazole buffer and the outer chamber with the 500 mM imidazole buffer. Every fraction was collected and analyzed, including the flow-through. The column was immediately regenerated with 10 column volumes of 20 mM MES, pH 5.0, 100 mM NaCl and washed with 10 column volumes of nanopure water, both at a flow rate of 0.5 mL/min.

Centrifugal Concentration

Because a large volume (13.8 mL) resulted from pooling IMAC fractions with purified and active SSADH, the volume had to be reduced to about 1% of the size exclusion resin bed volume for optimal resolution in that purification step. The pooled fractions were concentrated using an Amicon® Ultra-15 30kDa MWCO centrifugal filter, spun at 4000 × *g* for 30 minutes at 4 °C in an Eppendorf Centrifuge 5910 R. The filtrate was discarded, the remaining volume diluted four times with cold size exchange chromatography (SEC) elution buffer, and the spin was repeated. The concentrated and partially buffer exchanged sample was carefully pipetted from the concentrator, the filter was rinsed with a 0.5 mL of cold SEC buffer, the washing added to the sample and an aliquot was set aside for analysis.

Size Exclusion Chromatography

GE Healthcare Sephacryl® S-200 High Resolution resin was loaded into a 1.5 × 75.0 cm Bio-Rad Econo-Column® to a bed height of 73.0 cm (129.0 mL) and a Bio-Rad Econo-Column® Flow Adaptor was used to pack the column and load the sample. The adaptor was connected through tubing to a Bio-Rad Model EP-1 Econo Pump. A T-valve with Luer-Lok™ fittings was installed part way between the adapter and the pump to interrupt the mobile phase and inject the sample directly onto resin, via the flow adapter. Fractions from the column were collected in VWR disposable borosilicate glass culture tubes (13 × 100 mm) that were placed into a Bio-Rad Model 2110 Fraction Collector.

In a refrigerated room, the resin was equilibrated with 2 column volumes of the elution buffer: 30 mM Tris, pH 7.5, 100 mM NaCl, 5% glycerol, 10 mM BME. The sample was loaded slowly onto the column with a syringe that was attached to a Luer-Lok port on the T-valve in the direction that led to the resin. The pump was started, and the column was eluted with 1 column volume of elution buffer at a flow rate of 0.44 mL/min (15 cm/hr flow velocity) and the fractions were collected. Fractions that were deemed pure were pooled, aliquoted and flash frozen in liquid nitrogen without concentration, as they were already very concentrated and required dilution in order to assay activity. The resin was regenerated with 1 column volume of 30 mM Tris, pH 7.5, 1 M NaCl. Occasional cleaning in place of the column was performed by flowing 0.3 M NaOH through the column at 15–20 cm/h for 1 to 2 hours.

2.4.5 Protein Analysis

Protein Quantitation

Analysis during the purification consisted of determining the total protein concentration, as well as the purity and enzymatic activity of the expressed recombinant SSADH. Protein quantitation included both the colorimetric dye-based method of Bradford and the UV absorption-based A_{280} method. For crude samples, the Bradford assay with a bovine serum albumin (BSA) standard curve provides a reasonable estimation of the total protein concentration. Its limitations are that the dye reacts with basic and aromatic side chains of amino acids, the content of which differs from one protein to another, and for mixtures of proteins the use of a standard greatly limits the value of the estimation. This limitation extends even to pure samples, where BSA may not be an ideal standard for the unknown protein. The protein A_{280} measurement is a superior method for determining the concentration of a protein when it is sufficiently pure, and its amino acid sequence is known. This method relies on the unique absorption of tryptophan, tyrosine, and cystine (disulfide bond) amino acid side chains at 280 nm and employs a sequence-specific molar absorptivity constant (ε_{280}) that may be determined from the weighted sum of these three amino acids. The molar absorptivity constant for 6×His-SSADH is 44,077 M^{-1} cm^{-1}.

SDS-PAGE

The presence and purity of SSADH in the samples were evaluated by running samples on SDS-PAGE gels, followed by staining the gels with Coomassie Brilliant Blue R-250. Samples were considered for SDS-PAGE after inspecting the elution profiles from the columns, which consisted of both Bradford and A_{280} estimations of total protein concentration for each fraction.

The detection limit for Coomassie Brilliant Blue R-250 is approximately 0.1–0.5 μg of protein. A sample containing 0.5 μg protein that completely fills a 15 μL well would have a total protein concentration of 0.03 mg/mL. Therefore, in general, only samples containing over 0.03 mg/mL of total protein were run on the gels with 15 μL wells. Estimating total protein concentration this low using the Bradford assay is problematic, since it is traditionally held that the BSA standard curve is only valid for samples between 0.2–0.9 mg/mL total protein. More concentrated samples may be easily diluted, but concentrating dilute column fractions just for their consideration on a gel introduces much more time and handling, both of which are enemies in protein purification. It is far more practical to combine Bradford and A_{280} data to develop an elution profile and, combined with some experience, make a judicious selection in borderline cases. SDS-PAGE gels were run at a constant voltage (200 V) for 30 minutes.

2.4.6 Activity Assays and Enzyme Kinetics

Standard Activity Assay During Purification

A standard assay was used to report on the activity of SSADH throughout the purification. For the purification, the standard enzyme unit (U) is defined as the amount that catalyzes the formation of 1 μmol NADPH (product) per minute under the following standard conditions: 100 μM SSA (substrate), 1 mM $NADP^+$ (coenzyme), in a 100 mM sodium phosphate buffer, pH 7.5, at room temperature (20–22 °C). Under these conditions the reaction lasts for several minutes, and so the reactions were initiated by combining the substrate, coenzyme, and enzyme into a quartz cuvette, covering the top with Parafilm M and the mixing by inverting the cuvette 3 times. The cuvette was then inserted into the cell holder on the spectrophotometer, the lid to that compartment was closed and the acquisition started by pressing F9 on the keyboard. he

formation of NADPH was followed by monitoring the sample absorbance at 340 nm using a Shimadzu BioSpec-1601 UV-Vis spectrophotometer running UVProbe v2.61 software. Because the Kinetics Module in UVProbe outputs the data in absorbance units (at 340 nm) per second, this was converted to μmol/min using the Beer-Lambert Law, the extinction coefficient for NADPH, a 1 cm value for the cuvette pathlength and some algebra.

Enzyme Kinetics Assays

Due to the absence of climate control in the room where the assays were being conducted, and the number of electrical instruments and devices that could generate significant amounts of heat, the temperature in the room could fluctuate from 21-26 °C. Because a 5 °C increase in temperature can increase the rate of a chemical reaction by 50%, it was decided that all enzyme kinetics would be conducted at 30 °C. The input/output ports on the cell stage of the Shimadzu BioSpec-1601 UV-Vis spectrophotometer were connected to VWR® Scientific Model 1150 Refrigerated Circulating Bath that could also heat the water in the cell stage to 30 °C. A Torrey Pines Scientific, Inc. Echotherm™ Chilling/Heating Plate Model: IC20 was also used for heating samples to 30 °C in 1.5 mL Eppendorf tubes.

The reaction buffer was 100 mM sodium phosphate, which was prepared as both a pH 7.50 stock for diluting the enzyme, as well as at 7 other pH values from 6.50–8.25. A frozen aliquot of stock enzyme would be thawed on ice and diluted to the working concentration with the pH 7.50 buffer supplemented with 1 mM DTT. A small amount of this working enzyme solution was placed on the bench in ice, and the remainder returned to a refrigerator for long term stability studies. Stock solutions of SSA and $NADP^+$ were likewise thawed, diluted to working concentrations using a 100 mM sodium phosphate buffer that matched the pH for the reactions that day, and also stored on the bench in ice.

The reactions were prepared by dividing the reaction into separate half-volumes. Approximately half the volume of buffer was added to a quartz cuvette, along with the substrate and cofactor, and this was placed into the cell stage to begin warming to 30 °C. The other half-volume was comprised of the remainder of the buffer and the enzyme in a 1.5 mL sealed Eppendorf tube, which was placed into a heating block at 30 °C. A timer was set for 3 minutes, and after that time a P1000 pipette was used to draw up the ~0.5 mL of enzyme in buffer at 30 °C, and add it to the mixture in the cuvette, also now at 30 °C and already seated in the spectrophotometer. The addition was rapid, but controlled, and achieves near rapid mixing with some practice. The moment the plunger on the pipette completed its motion, the lid to the compartment was closed and a single keystroke on the keyboard (F9) began data collection. The entire action took approximately 2 seconds to complete. Reactions were recorded for 90 seconds, though in many cases the reactions were over well before that time and the majority of the initial velocity determinations were made within the first 15 seconds.

The enzyme kinetics assays in their entirety measured, *in triplicate*, initial velocities for SSADH-catalyzed reactions with [SSA] of 5, 10, 15, 20, 30, 40, 60, 80, and 100 μM, at fixed [$NADP^+$] of 0.1, 0.2, and 1.0 mM. These are 96 measurements, several of which would require repeating due to poor mixing or the introduction of an air bubble. For consistency, these reactions all completed in one period, typically over 15 hours. Realtime data analysis allowed the rapid identification of outliers, which could be retested immediately.

Initial Velocity from Progress Curves

Reaction progress curves that follow the change in absorbance with time define the reaction velocity, $v = d[P]/dt$. These progress curves, such as the one shown in **Figure 14**, can easily be fit with a 3rd-order polynomial of the form $f(x) = ax^3 + bx^2 + cx + d$. Taking the

derivative of the resulting fit equation and finding its maximum will provide the value for the initial velocity, assuming the data was correctly measured under initial velocity conditions.

Single-Variable Steady State Enzyme Kinetics

The rate equation for the dependence of initial velocity (v_0) on SSA concentration, when $NADP^+$ is fixed, is given in **Equation 8**. The data can be fit to this through nonlinear regression. The Lineweaver-Burk (double-reciprocal) form of this rate equation, which linearizes the data, is given in **Equation 9**. Linear regression can similarly be used to fit the data with the equation. In both cases, V_{max} is replaced with an apparent maximal velocity, V'_{max}, where the addition of the prime symbol (′) designates the value as apparent. Similarly, the Michaelis constant, K_{SSA}, has been replaced by the apparent Michaelis constant, K'_{SSA}. The *apparent* designations are used to acknowledge that these values are not absolute values, but relative ones that ultimately depend on the concentration of the fixed substrate, $NADP^+$, which is not explicitly represented in the equation. The expressions for V'_{max} and K'_{SSA} when $NADP^+$ is fixed were given in **Equations 12** and **14** (*Section 2.2.3.5*).

The same process may be applied to the situation where the concentration of SSA is fixed at constant value for the set of experiments in which $NADP^+$ is permitted to vary. The rate equation for the dependence of initial velocity (v_0) on $NADP^+$ concentration when SSA concentration is fixed, is given in **Equation 10** along with its double-reciprocal form in **Equation 11**. As before, the V'_{max} and K'_{NADP+} are the apparent values for these kinetic constants because the concentration of SSA is contributing to their values despite not explicitly appearing in the equation. The expressions for V'_{max} and K'_{SSA} with fixed $NADP^+$ were given in **Equations 13** and **15** (*Section 2.2.3.5*).

Global Least-Squares Regression

Multivariable nonlinear regression of the complete steady state enzyme kinetics for the bisubstrate SSADH-catalyzed reaction was performed using global least-squares regression on the initial velocity data with both varying concentrations of substrate and cofactor using **Equations 3** and **5**. ANOVA analysis of the fits, including Sum of Squared Errors were performed to ensure the data fit the proposed model.

pH-Rate Profile

The dependence of initial velocity on pH was explored by repeating the entire set of rate experiments previously described, with varied [SSA] at fixed [$NADP^+$], at different pH values for the 100 mM sodium phosphate buffer. After obtaining the value for k_{cat} from the global least-squares regression, this value was then transformed into the $\log_{10}k_{cat}$ and nonlinear regression was performed on the $\log_{10}k_{cat}$ versus pH data with **Equation 12** (*Section 2.2.3.8*), following the work by Knowels[23] and Kahn.[24] The parameters from the fit provided a pK_a for an ionizable group for which the rate of the SSADH-catalyzed reaction depends on, as well as a theoretical k_{cat}, which would be achieved at the pH optimum.

2.5 References

(1) Donnelly, M. I.; Cooper, R. A. Succinic semialdehyde dehydrogenases of Escherichia coli: their role in the degradation of p-hydroxyphenylacetate and gamma-aminobutyrate. *Eur J Biochem* **1981**, *113* (3), 555-561. DOI: 10.1111/j.1432-1033.1981.tb05098.x

(2) Marek, L. E.; Henson, J. M. Cloning and expression of the Escherichia coli K-12 sad gene. *J Bacteriol* **1988**, *170* (2), 991-994. DOI: 10.1128/jb.170.2.991-994.1988

(3) Cozzani, I.; Fazio, A. M.; Felici, E.; Barletta, G. Separation and characterization of NAD- and NADP-specific succinate-semialdehyde dehydrogenase from Escherichia coli K-12 3300. *Biochim Biophys Acta* **1980**, *613* (2), 309-317. DOI: 10.1016/0005-2744(80)90085-6

(4) Jaeger, M.; Rothacker, B.; Ilg, T. Saturation transfer difference NMR studies on substrates and inhibitors of succinic semialdehyde dehydrogenases. *Biochem Biophys Res Commun* **2008**, *372* (3), 400-406. DOI: 10.1016/j.bbrc.2008.04.183

(5) Langendorf, C. G.; Key, T. L.; Fenalti, G.; Kan, W. T.; Buckle, A. M.; Caradoc-Davies, T.; Tuck, K. L.; Law, R. H.; Whisstock, J. C. The X-ray crystal structure of Escherichia coli succinic semialdehyde dehydrogenase; structural insights into NADP+/enzyme interactions. *PLoS One* **2010**, *5* (2), e9280. DOI: 10.1371/journal.pone.0009280

(6) Fuhrer, T.; Chen, L.; Sauer, U.; Vitkup, D. Computational prediction and experimental verification of the gene encoding the NAD+/NADP+-dependent succinate semialdehyde dehydrogenase in Escherichia coli. *J Bacteriol* **2007**, *189* (22), 8073-8078. DOI: 10.1128/jb.01027-07

(7) Dover, S.; Halpern, Y. S. Utilization of γ-aminobutyric acid as the sole carbon and nitrogen source by Escherichia coli K-12 mutants. *J Bacteriol* **1972**, *109* (2), 835-843. DOI: 10.1128/jb.109.2.835-843.1972

(8) Michaelis, L.; Menten, M. L. Die kinetik der invertinwirkung. *Biochemische Zeitschrift* **1913**, *49*, 333–369.

(9) Briggs, G. E.; Haldane, J. B. A Note on the Kinetics of Enzyme Action. *Biochem J* **1925**, *19* (2), 338-339. DOI: 10.1042/bj0190338

(10) Segel, I. H. Chapter Six - Rapid Equilibrium Bireactant and Terreactant Systems. In *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, Segel, I. H. Ed.; John Wiley & Sons, 1975; pp 273-345.

(11) Cook, P. F.; Cleland, W. W. Chapter 5 - Initial Velocity Studies in the Absence of Added Inhibitors. In *Enzyme Kinetics and Mechanism*, 1st ed.; Garland Science, 2007; pp 59-120.

(12) Lineweaver, H.; Burk, D. The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society* **1934**, *56* (3), 658-666. DOI: 10.1021/ja01318a036

(13) Jang, E. H.; Park, S. A.; Chi, Y. M.; Lee, K. S. Structural insight into the substrate inhibition mechanism of NADP(+)-dependent succinic semialdehyde dehydrogenase from Streptococcus pyogenes. *Biochem Biophys Res Commun* **2015**, *461* (3), 487-493. DOI: 10.1016/j.bbrc.2015.04.047

(14) Purich, D. L. Chapter 4 - Practical Aspects of Measuring Initial Rates and Reaction Parameters. In *Enzyme Kinetics: Catalysis & Control*, Purich, D. L. Ed.; Elsevier, 2010; pp 215-285.

(15) Ellis, K. J.; Morrison, J. F. Buffers of constant ionic strength for studying pH-dependent processes. *Methods Enzymol* **1982**, *87*, 405-426. DOI: 10.1016/s0076-6879(82)87025-0

(16) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK values of the ionizable groups of proteins. *Protein Sci* **2006**, *15* (5), 1214-1218. DOI: 10.1110/ps.051840806

(17) Harris, T. K.; Turner, G. J. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* **2002**, *53* (2), 85-98. DOI: 10.1080/15216540211468

(18) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* **2009**, *18* (1), 247-251. DOI: 10.1002/pro.19

(19) Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* **2009**, *284* (20), 13285-13289. DOI: 10.1074/jbc.R800080200

(20) Krulwich, T. A.; Sachs, G.; Padan, E. Molecular aspects of bacterial pH sensing and homeostasis. *Nat Rev Microbiol* **2011**, *9* (5), 330-343. DOI: 10.1038/nrmicro2549

(21) Saito, H.; Kobayashi, H. Bacterial responses to alkaline stress. *Sci Prog* **2003**, *86* (Pt 4), 271-282. DOI: 10.3184/003685003783238635

(22) Law, R. H.; Irving, J. A.; Buckle, A. M.; Ruzyla, K.; Buzza, M.; Bashtannyk-Puhalovich, T. A.; Beddoe, T. C.; Nguyen, K.; Worrall, D. M.; Bottomley, S. P.; et al. The high resolution crystal structure of the human tumor suppressor maspin reveals a novel conformational switch in the G-helix. *J Biol Chem* **2005**, *280* (23), 22356-22364. DOI: 10.1074/jbc.M412043200

(23) Knowles, J. R. The intrinsic pKa-values of functional groups in enzymes: improper deductions from the pH-dependence of steady-state parameters. *CRC Crit Rev Biochem* **1976**, *4* (2), 165-173. DOI: 10.3109/10409237609105457

(24) Kahn, K.; Tipton, P. A. Kinetic mechanism and cofactor content of soybean root nodule urate oxidase. *Biochemistry* **1997**, *36* (16), 4731-4738. DOI: 10.1021/bi963184w

Chapter 3

The Expression of Recombinant HOT in *E. coli*

Abstract

When expressed in *E. coli*, recombinant human HOT is found exclusively as inclusion bodies amongst the insoluble material that is recovered following the initial purification steps of cell lysis and centrifugation. A single published attempt at expressing human HOT in *E. coli* reported solubilizing the inclusion bodies with urea prior to the chromatographic purification step, but the subsequent effort of refolding the purified protein using dialysis failed to recover the active enzyme. Due to a lack of structural information on HOT, designing a rational strategy to overcome protein insolubility has been challenging. Using AlphaFold, an artificial intelligence program that can accurately predict the structure of a protein from an amino acid sequence, a number of structural features in HOT were identified as being potentially responsible for promoting aggregation or preventing proper folding of the protein. These include a purported amphipathic N-terminal mitochondrial targeting peptide (MTP) as well as several potential disulfide and *cis* X-proline peptide bonds. A strategy was developed that employed two HOT constructs, with and without the N-terminal MTP. These HOT constructs, as well as plasmids co-expressing molecular chaperones, were transformed into several cell lines, including those capable of forming disulfide bonds. Full-length human HOT containing the N-terminal MBP was successfully expressed as a soluble fusion protein in *E. coli* ER2523 cells using an N-terminal maltose-binding protein affinity tag. Additionally, a HOT isoform lacking the N-terminal MTP was successfully expressed as a soluble protein in the BL21(DE3) and NiCo21(DE3) cell lines with the assistance of co-expressed molecular chaperones.

3.1 Introduction

3.1.1 Literature on Recombinant HOT in *E. coli*

An exhaustive literature search produced only one published attempt at expressing recombinant HOT in *E. coli*. The endeavor was documented in a 2007 master's thesis from the University of Strathclyde, Glasgow, Scotland, in which *human* HOT containing an N-terminal His-tag was expressed in the BL21(DE3)pLysS strain of *E. coli*.[1] The expression was performed by growing cells to an OD_{600} of 0.4 in LB media at 30 °C, and then inducing expression with 0.25 mM IPTG for 2 hours. Although a high level of expression was obtained, the majority of recombinant HOT was found only in the insoluble fraction. Modifications to the expression conditions failed to increase in solubility of recombinant HOT, so the inclusion bodies were solubilized with the addition of urea, and recombinant His-tagged HOT was purified on a nickel column in its denatured form. To refold the protein, purified recombinant HOT was dialyzed for 1 hour at 4 °C in pH 8.0 phosphate buffer, but no enzymatic activity was detected.

Critical evaluation of the methods in the thesis suggests flaws in the dialysis methodology and omission of conventional expression strategies used to enhance protein solubility. For example, the purported Fe^{2+} and NAD^+ cofactors were not included in the dialysis buffer, and since phosphate is known to complex iron[2], the choice of phosphate buffer would potentially limit Fe^{2+} availability. Furthermore, dialyzing for one hour won't accomplish anything, since this process is normally carried out stepwise, exchanging the buffer several times over one or more days.[3] Conventional expression strategies for increasing protein solubility were not discussed in the thesis, including significantly decreasing the expression temperature and

adding a solubility-enhancing tag, such as glutathione S-transferase[4] (GST), maltose-binding protein[5] (MBP) or small ubiquitin-like modifier[6] (SUMO).

3.1.2 Present Research on Recombinant HOT in *E. coli*

It is apparent that the heterologous expression of recombinant HOT in *E. coli* results in the formation of insoluble protein aggregates, often referred to as inclusion bodies[7, 8]. As attempted in the Strathclyde thesis[1], one strategy used to recover biologically active proteins from inclusion bodies is refolding[9]. Other strategies, some of which were employed in the present research, aim to prevent or limit the formation of protein aggregates from the start by altering the expression conditions and even the protein of interest itself.[10-15]

The aims of the present research were to eliminate or reduce protein aggregation by:

1. Modifying the HOT primary amino acid sequence to enhance its solubility
2. Modifying the protein expression conditions to limit HOT aggregation

In order to achieve these two aims, the following approaches were taken:

1. For modifications to the HOT primary amino acid sequence, the strategies were to enhance solubility by:
    a. Add a maltose-binding protein tag to the N-terminus
    b. Remove the purported N-terminal mitochondrial targeting peptide
2. For modifications to the protein expression conditions, the strategies to limit protein aggregation were to:
    a. Lower the expression temperature and inducer concentration
    b. Co-express molecular chaperones to aid in protein folding
    c. Express HOT in a cell line that facilitates disulfide bond formation

The rationale for each of these strategies is discussed in the following sections.

3.1.3 Modifications to the HOT Primary Amino Acid Sequence

Addition of an N-terminal Maltose-Binding Protein Tag

A very common strategy for addressing solubility issues in the heterologous expression of eukaryotic proteins in prokaryotic hosts is to add a solubility-enhancing tag to either the N- or C-terminus of the target protein.[16-18] The maltose-binding protein tag (MBP-tag) is one of the most commonly used tags and has been shown to enhance the solubility of proteins[5] in the cytoplasm and periplasmic space[19]. A schematic diagram and predicted structure for the MBP-M1HOT fusion protein is presented in **Figure 3.1**.

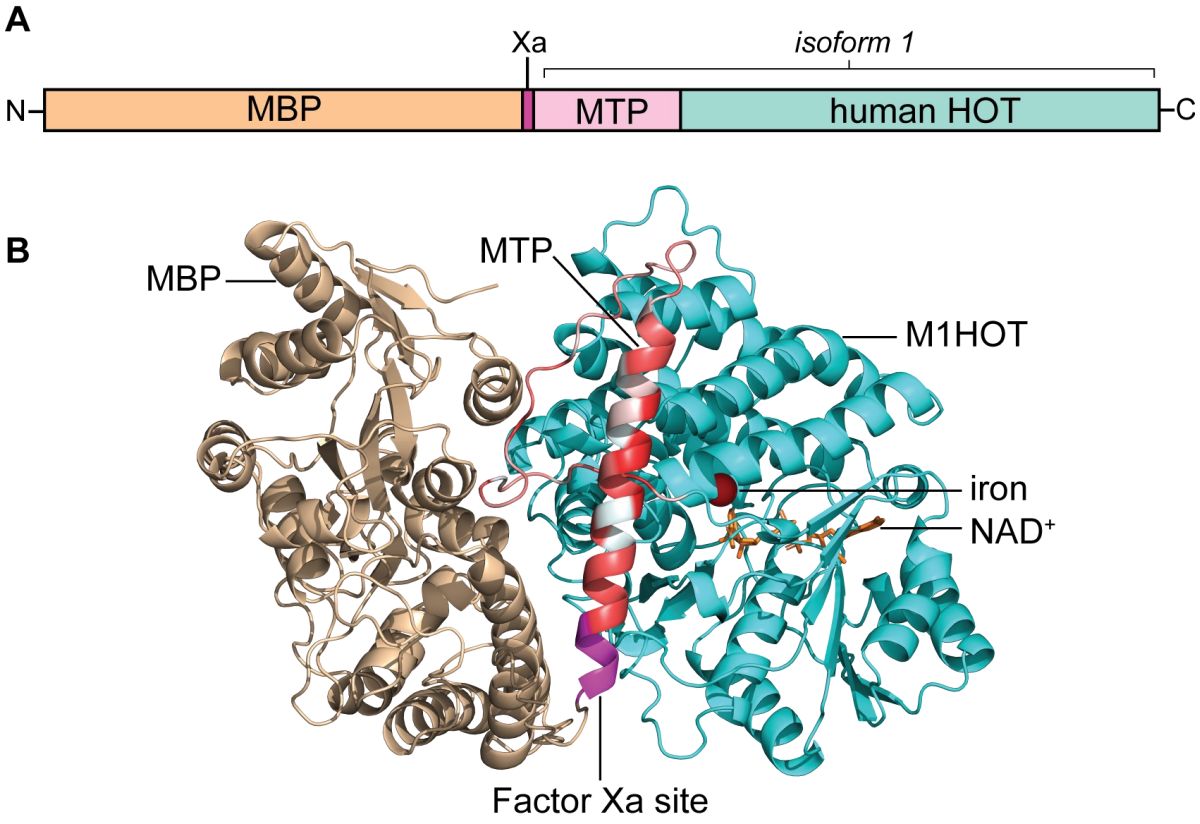

Figure 3.1. MBP-M1HOT *E. coli* expression construct. (A) A schematic diagram of the MBP-M1HOT construct. **(B)** The AlphaFold-predicted structure for MBP-M1HOT. HOT isoform 1 (M1HOT) includes the HOT protein (cyan) and N-terminal MTP (pink in diagram; red hydrophobic and white hydrophilic in structure). The MBP-tag (wheat) is fused to the HOT isoform 1 (M1HOT) N-terminus HOT by a cleavable Factor Xa protease recognition sequence (magenta). The predicted location of iron (red) and NAD^+ (orange) are also shown.

The grand average of hydropathicity (GRAVY) value represents a protein or peptide's tendency towards being either hydrophobic (positive number) or hydrophilic (negative).[20] Like most commonly used solubility-enhancing tags, the MBP-tag has a negative GRAVY value (-0.340).[16] Native human M1HOT has a GRAVY value of 0.006 (hydrophobic), while the recombinant MBP-M1HOT fusion has a GRAVY value of -0.182 (hydrophilic), as calculated by the ProtParam[21] tool on the Expasy Swiss Bioinformatics Resource Portal[22]. Thus, the MBP-tag increases the solubility of the passenger protein by contributing to the overall hydrophilicity of the fusion protein. Aside from adding to the hydrophobic character of the fusion protein, there is evidence that the MBP-tag also functions as a molecular chaperone by binding to hydrophobic folding intermediates of the passenger protein and preventing their self-association.[23] Lastly, because MBP-tag binds maltose it also functions as a purification tag that can be used with in combination with amylose resin during affinity chromatography.

Deletion of the N-terminal Mitochondrial Transit Peptide

There is both experimental[24, 25] and computation[26-30] evidence indicate that HOT contains and N-terminal mitochondrial targeting peptide[31, 32] (MTP). See Chapter 4 (*Sections 4.1.1* and *4.1.2*) for more discussion on the HOT N-terminal MTP. In eukaryotic cells, the MTP presequences are normally removed from the N-terminus by mitochondrial peptidases, at which point the preproteins are folded into their mature and active forms.[33]

Because *E. coli* lack mitochondria, the MTP serves no functional purpose and cannot be cleaved from HOT in order to form the mature protein. Furthermore, the predicted amphipathic helix and random coil structure for the MTP may serve as a nucleation point for aggregation. For these reasons the N-terminal MTP was removed from the expression construct and

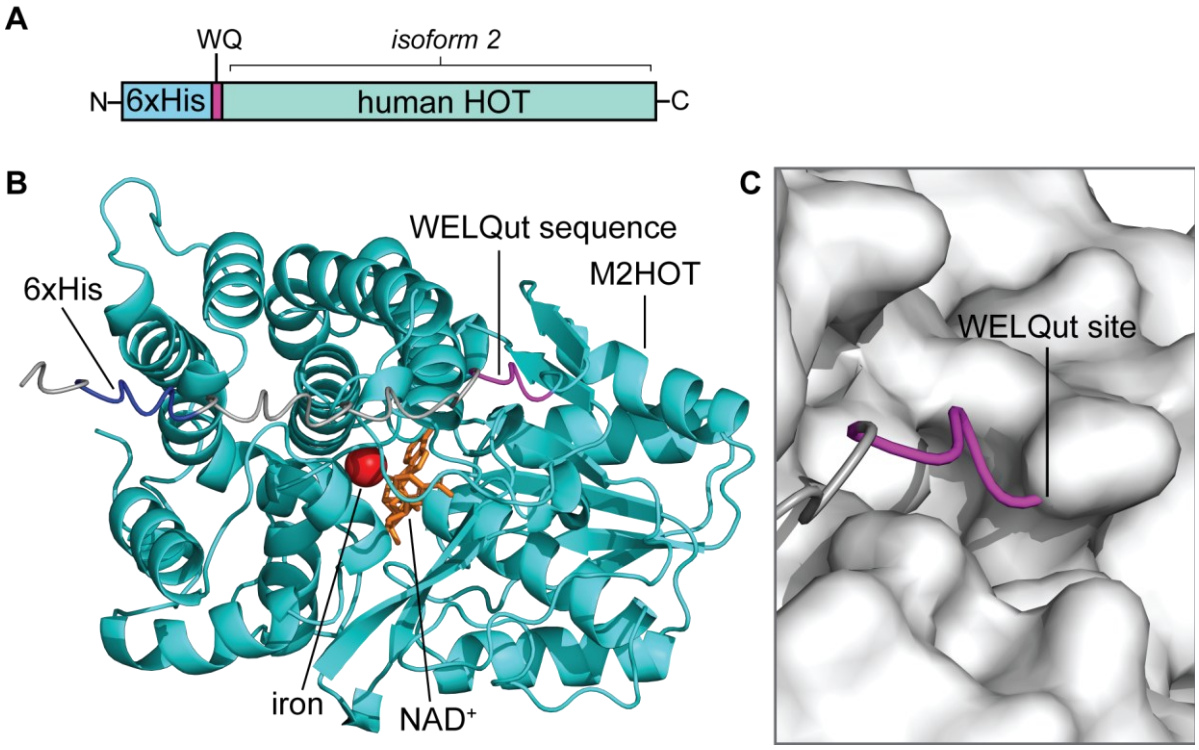

Figure 3.2. 6×His-M2HOT *E. coli* expression construct. (A) A schematic diagram of the 6×His-M2HOT construct. **(B)** The AlphaFold-predicted structure for 6×His-M2HOT. HOT isoform 2 (M2HOT) lacks the N-terminal MTP. The N-terminal fusion (white) containing the hexahistidine sequence (blue) is fused to M2HOT (cyan) by a cleavable WELQut protease recognition sequence (magenta). The predicted location of iron (red) and NAD^+ (orange) are also shown.

replaced with a His-tag. A schematic diagram and predicted structure for the 6×His-M2HOT fusion is shown in **Figure 3.2**.

3.1.4 Modifications to the Protein Expression Conditions

Lowering the Expression Temperature and Inducer Concentration

Generally, the first and easiest strategy used for enhancing protein solubility in *E. coli* is to lower the expression temperature during induction.[34] This due to the kinetic competition between protein folding (first-order process) and aggregation (second-order or higher), both of which are dependent on the rate of protein expression.[35] Lowering the temperature decreases the rate of protein expression, which favors folding over aggregation.

Co-Expression of Molecular Chaperones

Molecular chaperones are a group of proteins that function either to: 1) *hold* nascent polypeptides and prevent their aggregation; 2) actively *fold* (or refold) proteins in ATP-driven reactions; or 3) *disaggregate* protein aggregates, also through ATP-driven reactions.[36] The co-expression of chaperones has increased the solubility for a number of proteins normally expressed as inclusion bodies in *E. coli*.[37, 38]

The two sets of chaperones used in the present research were encoded by Takara's pG-KJE8 and pG-Tf2 plasmids.[39, 40] pG-KJE8 encodes the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, while the pG-Tf2 plasmid encodes *tig* and *gro*ES-*gro*EL chaperones. Because both plasmids contain the *cat* gene (chloramphenicol acetyltransferase) conferring resistance to chloramphenicol (Cam^R), they are incompatible with the BL21(DE3)pLysS *E. coli* cell line that already uses chloramphenicol a selection marker for the pLysS plasmid.

Most peptide bonds adopt the trans isomer configuration to avoid steric clashes between adjacent C_α atoms of the protein backbone. However, the side chain of the amino acid proline is bonded to the C_α atom and so both *cis* and *trans* configurations produce a steric clash, resulting in little energy difference between the two conformations. Therefore, proline adopts both *cis* and *trans* X-Pro peptide bonds. While a protein is being synthesized by the ribosome, all nascent polypeptides have the trans configuration for new peptide bonds. If a folded protein requires the cis conformation, the bond must isomerize. Proline *cis* to *trans* isomerization is extremely slow, with a rate constant of 2.5×10^{-3} s^{-1} at 25 °C.[41] As a result, *cis*/*trans* proline isomerization can be a rate-limiting step in protein folding.[42] However, the isomerization rate can be accelerated 10^3-10^6 fold when catalyzed by prolyl isomerases[43], and the *tig* chaperone (Trigger Factor) on the Takara plasmid pG-Tf2 has peptidyl–prolyl *cis*/*trans* isomerase

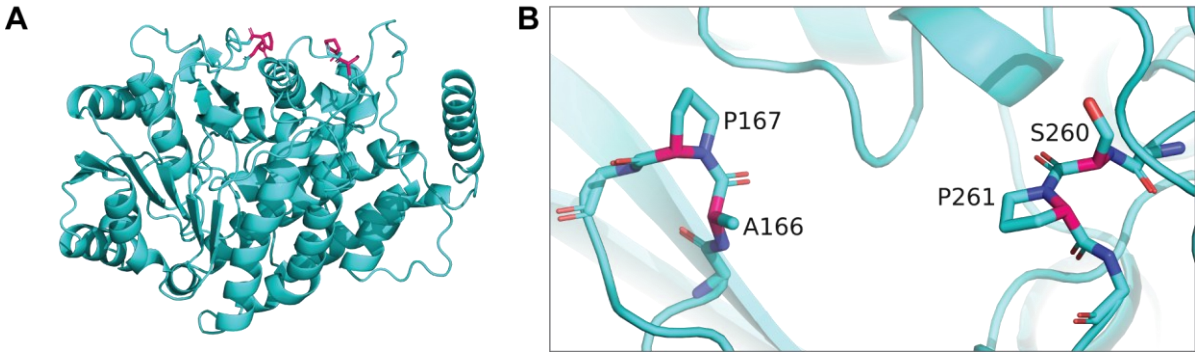

Figure 3.3. Two *cis* X-proline bonds in the predicted human M2HOT structure. (A) The location of the two *cis* X-proline residues in human M2HOT. **(B)** A close-up view of the two *cis* X-proline peptide bonds.

activity.[44, 45] This *tig* chaperone was of particular interest in the present research because examination of the AlphaFold predicted structure for HOT identified two *cis* X-proline peptide bonds that had the potential to limit the folding kinetics (**Figure 3.3**).

Disulfide bonds

Examination of the AlphaFold predicted structure for human M1HOT also uncovered the potential existence of three disulfide bonds (**Figure 3.4**). The same three disulfide bonds were also detected by the DiANNA 1.1 web server, an algorithm that predicts cysteine classification and disulfide connectivity.[46-48] HOT is not a cytoplasmic enzyme and it isn't secreted either, but it is imported into the mitochondria, presumably into the matrix. Studies using redox-sensitive yellow fluorescent protein (rxYFP) indicate that the balance is shifted more towards disulfide formation in mitochondrial matrix than in the cytosol, but the authors note reduction potentials are pH dependent, and that matrix (pH ~7.4) reduction potential is actually greater (-296 mV) than the cytosol (~7.0) reduction potential (-286 mV).[49]

Since literature provided no clear indication as to the probable oxidation states for the three pairs of cysteines, it seemed a worthwhile endeavor to express HOT in *E. coli* cells engineered to form disulfide bonds in the cytoplasm. The SHuffle T7 Express cells from New

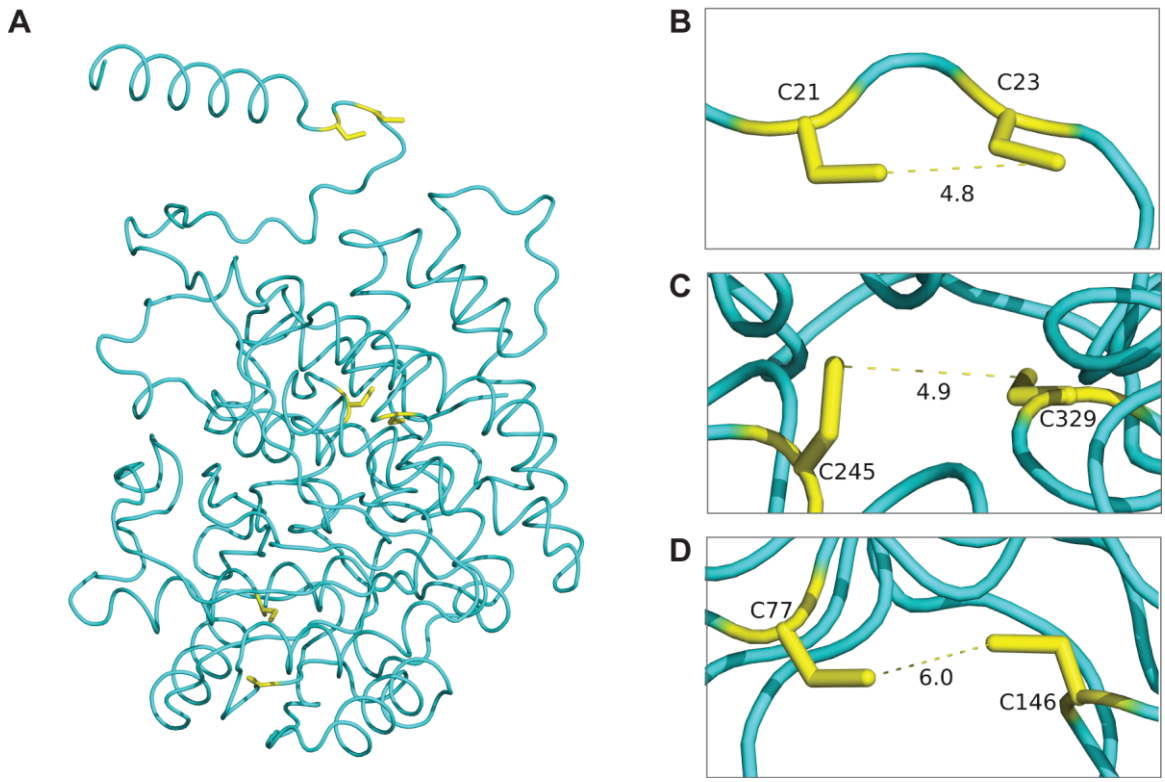

Figure 3.4. The location and distances for potential disulfide bonds in human M1HOT. (A) The location of the three potential disulfide bonds in human M1HOT. **(B)** Heavy atom bond distances for C21 and C23. **(C)** Heavy atom bond distances for C245 and C329. **(D)** Heavy atom bond distances for C77 and C146.

England BioLabs have deletions of the genes for glutaredoxin reductase and thioredoxin reductase (Δgor ΔtrxB), which allows disulfide bonds to form in the cytoplasm, and also expresses the periplasmic disulfide bond isomerase DsbC corrects mis-oxidized bonds and promotes proper folding.[50]

NiCo21(DE3) Cell Line

The His-tag remains one of the most commonly used purification tags for the expression recombinant proteins in *E. coli* and, combined with immobilized metal affinity chromatography (IMAC), the His-tag can often result in highly pure target protein in just one step. However, *E. coli* contains a number of intrinsic metal-binding proteins that can bind to the IMAC resin and then coelute from the column along the protein of interest.[51] To solve this

Table 3.1. Metal-binding proteins in *E. coli*, before and after modification.

Before		After		
Name	**MW (kDa)**	**Name**	**MW (kDA)**	**Modification**
arnA	74	arnA::CBD	84	chitin-binding domain added
glmS	67	glmS6Ala	67	6×His→6×Ala mutation
can	25	can::CBD	32	chitin-binding domain added
slyD	21(28)*	slyD::CBD	28(35)*	chitin-binding domain added
* Numbers in parentheses reflect the apparent molecular weight on SDS-PAGE				

issue, New England BioLabs, has developed a line of engineered *E. coli* strains in which the metal-binding proteins have been modified so they either don't bind the resin, or have been fused with a chitin-binding domain that allows their capture on a chitin column following elution from the IMAC resin.[52] A list of these metal-binding proteins, before and after modification is presented in **Table 3.1**.

3.2 Results and Discussion

3.2.1 Recombinant Human M1HOT with N-terminal MBP-tag

Expression of MBP-M1HOT in ER2523 Cells

The pMAL-c5X plasmid encoding the MBP-M1HOT fusion protein was transformed into the ER2523 *E. coli* cells supplied with the NEB pMAL kit. In addition, the pMAL-c5X plasmid was also transformed into the Thermo Scientific BL21(DE3) and NEB SHuffle T7 Express *E. coli* cell lines, with and without the pG-KJ8 plasmid that co-expresses *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL. Because the ER2523 cell line was engineered by NEB to work with their pMAL expression system, it is not surprising that strong expression and recovery of soluble MBP-M1HOT fusion protein was observed in this cell line.

Although the pMAL-c5X plasmid expresses well in SHuffle T7 Express cells, no enhancement in the solubility of HOT was observed with this cell line. This indicates that a reducing environment favoring disulfide formation may not be a crucial factor in preventing the aggregation of HOT when it is expressed in *E. coli*. Additionally, the co-expression of the pG-KJE8 chaperone set in both BL21(DE3) or SHuffle T7 Express cells did not increase the ratio of soluble MBP-M1HOT fusion protein beyond what the MBP-tag was able to achieve. It may be that this fusion protein, with a molecular weight of 92.8 kDa, may be too large for the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones. The pG-Tf2 chaperone set, which expresses the *tig* and *gro*ES-*gro*EL chaperones, was not tested with the MBP-M1HOT construct.

Supplementation of the lysis buffer with 300 mM KCl significantly decreased MBP-M1HOT solubility, while the inclusion of 5 mM DTT only slightly decreased the

solubility. The addition of 1% Triton X-100 to the lysis buffer had no effect on the solubility of MBP-M1HOT in the absence of KCl and DTT. However, 1% Triton X-100 did manage to compensate for the negative impacts of KCl and DTT on MBP-M1HOT solubility, returning the overall MBP-M1HOT solubility to that observed with the supplemented buffer. Due to the significant biophysical impact that the maltose-binding protein has on the entire fusion protein, it is difficult to rationalize the role KCl, DTT and Triton X-100 have on increasing or decreasing the solubility of M1HOT. Nevertheless, it is clear that the maltose-binding protein alone had the largest impact on M1HOT solubility in the ER2523 cell line.

After the solubilizing properties of the maltose-binding protein, the next most impactful factor that resulted in the recovery of soluble MBP-M1HOT were the expression conditions themselves. The greatest solubility was achieved when the cells were induced at 16 °C for 24 hours using 0.1 mM IPTG as the inducer. Lastly, the lysis method also had a significant impact on MBP-M1HOT solubility. Sonication of ER2523 cells resulted in the greatest recovery of soluble MBP-M1HOT, while freeze-thaw resulted in nearly all of the MBP-M1HOT being found in the insoluble fraction.

Sonication and centrifugation of a 4.08 g ER2523 cell pellet in lysis buffer yielded 25.02 mL of clarified lysate with a total protein concentration of 22.78 mg/mL, as determined by Bradford assay, for a total protein mass of 194.0 mg.

Purification of MBP-M1HOT from ER2523 Cells

Although the N-terminal maltose-binding tag served here primarily as a solubility-enhancing element, it also functions as a very effective affinity-purification tag when combined with a chromatographic separation on an amylose column. **Figure 3.1** shows the elution profile for the purification of MBP-M1HOT on an amylose column, where after binding to the resin

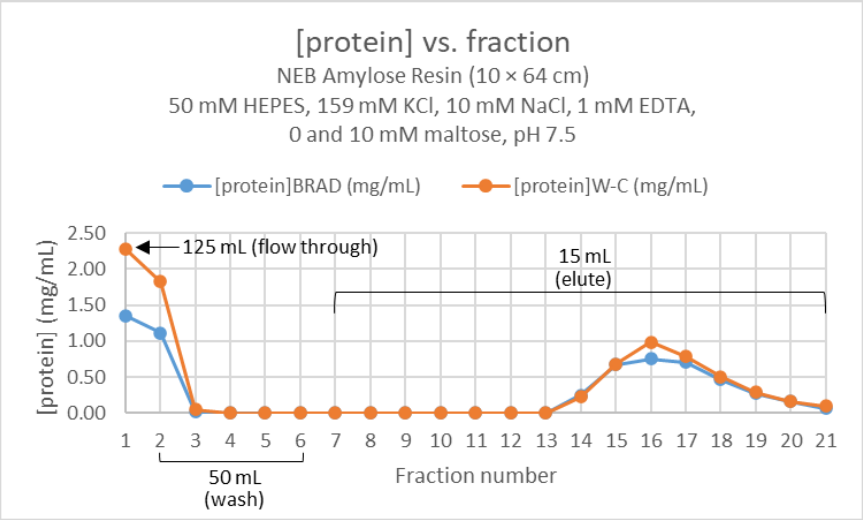

Figure 3.5. Elution profile for the amylose column purification of MBP-M1HOT. The total protein concentration from fractions eluted from the amylose column is plotted against fraction to develop an elution profile for the purification. Concentrations were determined using both the Bradford assay (**blue**) and the Warburg-Christian method (**orange**). The 10 mM maltose elution began at fraction 7.

the MBP-M1HOT was eluted from the column with the addition of 10 mM maltose. SDS-PAGE analysis of the column fractions indicated that MBP-M1HOT eluted nearly pure in fractions E14–E21 (**Figure 3.2**).

The pooled E14–E21 fractions had a total volume of 8 mL, a protein concentration of 0.60 mg/mL based on the A_{280}, and a total protein mass of 4.79 mg (2.5% yield). A sample of the

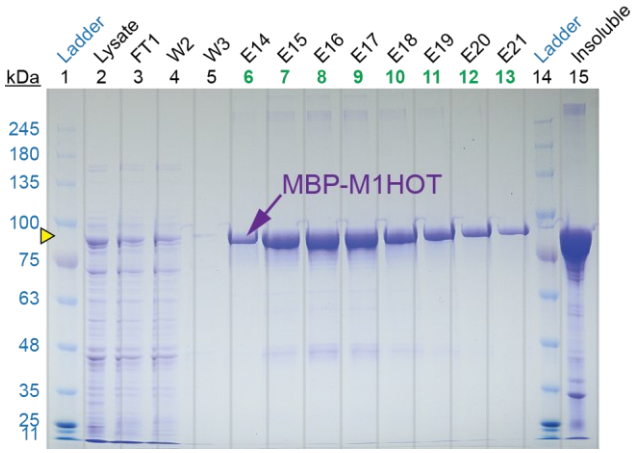

Figure 3.6. SDS-PAGE for the amylose column purification of MBP-M1HOT. The gel was loaded with a maximum of 5 µg total protein/well. The molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). Bands of interest are indicated with a **purple** arrow, with the vertical height of the arrow tip (relative to the ladder) indicated by the **yellow** triangle. The calculated molar mass of MBP-M1HOT is 92.8 kDa. Well numbers appear above the lanes, and those colored in **green** indicate fractions that were collected and pooled for analysis. The labels above the well numbers indicate the column fraction composition and number. **Fraction labels:** FT = flow-though, W = wash, E = elution with maltose.

93 kDa band was cut from the gel, subjected to in-gel digest with reductively methylated trypsin, and analyzed by liquid-chromatography tandem mass spectrometry. Peptide mass fingerprinting confirmed that the 93 kDa band was MBP-M1HOT. A sample of the pooled fractions was also tested for HOT enzyme activity, but none was detected.

3.2.2 Recombinant Human M2HOT with N-terminal His-tag

Expression of 6×His-M2HOT in BL21(DE3) and SHuffle T7 Express Cells

The pLATE52 plasmid encoding the 6×His-M2HOT fusion protein was transformed into the BL21(DE3) and SHuffle T7 Express *E. coli* cells lines, with and without the chaperone plasmids sets pG-KJ8 or pG-Tf2. The pG-KJ8 plasmid co-expresses *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, while pG-Tf2 co-expresses the *tig* and *gro*ES-*gro*EL chaperones.

When 6×His-M2HOT is expressed in SHuffle T7 Express cells alone, all of the protein is recovered in the insoluble fraction of the lysed cells. When co-expressed in SHuffle T7 Express cells with *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, overall expression levels for 6×His-M2HOT are markedly decreased and 50% of 6×His-M2HOT is recovered in the soluble fraction, with the other 50% remaining in the insoluble fraction. When co-expressed in SHuffle T7 Express cells with the *tig* and *gro*ES-*gro*EL chaperones, overall expression levels for 6×His-M2HOT are decreased even further, and 6×His-M2HOT is recovered from the soluble fraction at a comparable level observed with the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, but with almost no detectable 6×His-M2HOT in the insoluble material. It appears that SHuffle T7 Express cells are not compatible with the co-expression of 6×His-M2HOT and the molecular chaperones sets.

When 6×His-M2HOT is expressed in BL21(DE3) cells alone, all of the protein is recovered in the insoluble fraction of the lysed cells, as observed with the SHuffle T7 Express cells. However, when co-expressed in BL21(DE3) cells with the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, a substantial increase in 6×His-M2HOT is observed in the soluble fraction. Moreover, when DTT was used in the lysis buffer the solubility of 6×His-M2HOT further increased, calling into question the necessity of disulfide bonds for the solubility of HOT. There appears to be no decrease in overall levels of 6×His-M2HOT expression with the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, since 6×His-M2HOT is still recovered from the insoluble fraction at comparable levels to when it is expressed without these chaperones. When co-expressed in BL21(DE3) cells with the *gro*ES-*gro*EL and *tig* chaperones, the overall expression level of 6×His-M2HOT follows the pattern observed in SHuffle T7 Express cells: markedly decreased overall expression of 6×His-M2HOT that is barely detectable, with 50% soluble and the remaining 50% recovered in the insoluble fraction.

After examining the expression of 6×His-M2HOT in both SHuffle T7 Express and BL21(DE3) cell lines, along with the presence or absence of the pG-KJ8 and pG-Tf2 chaperone sets, it's clear that the greatest increase in 6×His-M2HOT solubility, without any decrease in overall expression, occurs with the combination of BL21(DE3) cells co-expressing the pG-KJ8 chaperone set (*dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL). Apart from this combination of cell line and chaperone set, the next most impactful factor that resulted in the recovery of soluble 6×His-M2HOT was the expression conditions themselves. The greatest solubility was achieved when the cells were induced at 16 °C for 24 hours using 0.05 mM IPTG as the inducer for 6×His-M2HOT, 2 mg/mL L-arabinose as the inducer for *dna*K-*dna*J-*grp*E, and 10 ng/mL tetracycline as the inducer for *gro*ES-*gro*EL. As for the lysis method, similar to the

MBP-M1HOT in *E. coli* ER2523 cells, the sonication of BL21(DE3) cells resulted in the greatest recovery of soluble MBP-M1HOT, while freeze-thaw resulted in nearly all of the 6×His-M2HOT being found in the insoluble fraction.

Lastly, the lysis buffer was supplemented with various additives in order to assess their ability to aid in the recovery of soluble 6×His-M2HOT expressed in the BL21(DE3) cells without the co-expression of the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones. These included 10 mM DTT, 10% glycerol, and 1% Tween 20. As noted above, the addition of DTT to the lysis buffer produced an observable increase in 6×His-M2HOT solubility. However, the addition glycerol and Tween 20, and various combinations of all three additives did not produce any detectable increase in 6×His-M2HOT solubility. The use of arginine to as a solubilizing agent in the lysis buffer was also explored. Inclusion bodies collected from BL21(DE3) cells expressing 6×His-M2HOT, without co-expression of chaperones, were treated with various buffers that contained combinations of arginine, DTT, glycerol, and Tween 20. Only buffers containing 0.1–2.0 M arginine as the sole additive produced detectable solubilization of 6×His-M2HOT from inclusion bodies. Arginine concentrations of 0.1, 0.25, 0.5 and 1.0 M produced a marginal and equivalent increase in 6×His-M2HOT, as determined by InVision His-tag fluorescent staining, while 2.0 M arginine produced slightly better than marginal increase in solubility. The use of 2.0 M arginine in the lysis buffer was not pursued since the minimal increase in 6×His-M2HOT was not worth the risk in protein denaturation, especially when co-expression of *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones resulted in better 6×His-M2HOT solubility.

Purification of 6×His-M2HOT from BL21(DE3) Cells

6×His-M2HOT was purified from the clarified lysate by immobilized metal affinity chromatography (IMAC). After loading the lysate onto an equilibrated Ni^{2+}-NTA column and washing with a low concentration of imidazole, 6×His-M2HOT was eluted using a gradient of 25 – 250 mM imidazole. Column fractions were analyzed by SDS-PAGE and gels were visualized by staining with both Coomassie Brilliant Blue R-250 and InVision™ His-Tag In-Gel Stain (Thermo Fisher Scientific). The Coomassie-stained SDS-PAGE gel shows that a number of impurities co-purified with 6×His-M2HOT (**Figure 3.3A**), which elutes with approximately 100 mM imidazole. The 6×His-M2HOT band is indicated in **Figure 3.3** by green arrow and text.

The identities of several of the contaminating bands belong to chaperones, and they were deduced both by their molecular weights and by comparing 6×His-M2HOT expressions with

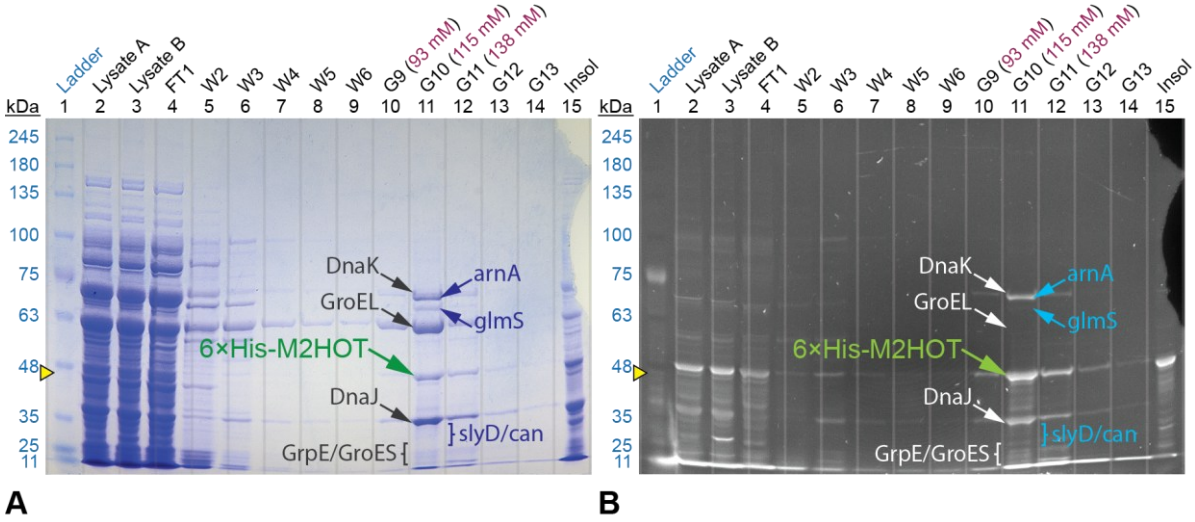

Figure 3.7. SDS-PAGE (InVision) for the IMAC purification of 6×His-M2HOT. The gel was loaded using a maximum of 50 µg/well total protein. Molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). The 6×His-M2HOT bands are indicated by the **green** arrows, with the height of the arrow tips relative to the protein ladder marked by **yellow** triangles. The calculated mass of 6×His-M2HOT is 48.2 kDa. *Chaperones* are labeled with **black** (or white) text/arrows, while *nickel-binding proteins* are labeled with **dark blue** (or **cyan**) text/arrows. The estimated imidazole concentrations (**magenta**) are above the fraction labels. **Fraction labels:** FT = flow-through, W = wash, G = gradient, and Insol = insoluble material. **(A)** Coomassie-stained gel of IMAC fractions showing the co-elution of chaperones and nickel-binding proteins with 6×His-M2HOT. **(B)** InVision (His-tag) staining and fluorescent imaging highlights the location of His-tagged M2HOT, as well as some nickel-binding proteins and DnaJ (chaperone).

and without co-expression of the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperone sets. The contaminating bands from chaperones are indicated in **Figure 3.3A** with black arrows/brackets and text, and in **Figure 3.3B** with white arrows/brackets and text (on the left-hand side of the bands). The remaining contaminating bands belong to nickel-binding proteins, endogenously produced by *E. coli* cells. These nickel-binding proteins are identified in **Figure 3.3A** with dark blue arrows/brackets and text, and in **Figure 3.3B** by cyan arrows/brackets and text (to the right of the bands). The identity of the nickel-binding bands was confirmed by later work with NiCo21(DE3) cells, where these proteins are modified in such a way that they do not bind Ni^{2+}-NTA, or they bind chitin and can be readily eliminated by passing the protein solution though a chitin column (next section).

The same gel is depicted in **Figure 3.3B** after being stained with the InVision dye. The InVision stain is a proprietary fluorescent dye conjugated to a Ni^{2+}-NTA complex. This dye binds to the His-tags in fusion protein bands and causes them to fluoresce when placed on a UV transilluminator. 6×His-M2HOT is readily detected by the dye, as are some of the endogenous nickel-binding proteins such as arnA. Of curious note, the chaperone *Dna*J readily binds the InVision dye and fluoresces quite strongly. This is likely due to the fact that *Dna*J has two non-adjacent Zn^{2+} binding sites, each formed by a motif of four conserved cysteine residues.[53] Similar cysteine motifs are also known to coordinate Ni^{2+}.[54, 55]

While *Dna*J may have co-eluted with 6×His-M2HOT because it genuinely coordinates with the Ni^{2+}-NTA resin, the other chaperones may be directly associated with 6×His-M2HOT or stuck to it due to an incomplete folding cycle. Since chaperones associate with unfolded proteins through hydrophobic interactions, and their protein-folding dynamics are coupled to

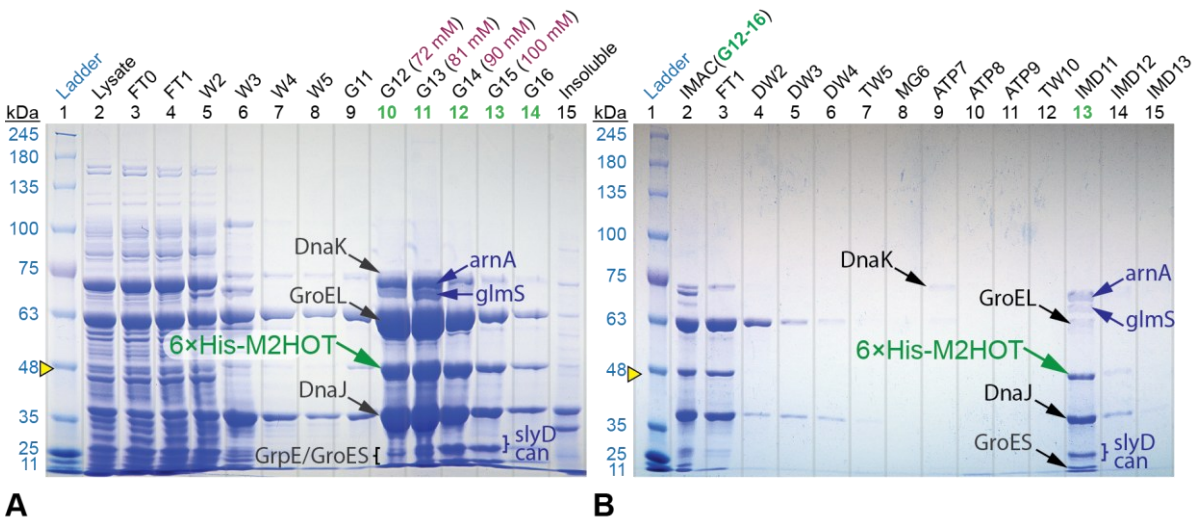

Figure 3.8. SDS-PAGE for the IMAC and HSP purification of 6×His-M2HOT. Molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). The 6×His-M2HOT bands are indicated by the **green** arrows, with the height of the arrow tips relative to the protein ladder marked by **yellow** triangles. The calculated mass of 6×His-M2HOT is 48.2 kDa. Chaperones are labeled with **black** text/arrows, while nickel-binding proteins are labeled with **dark blue** text/arrows. Fractions with **green** well numbers were collected for further analysis or processing. **(A)** Coomassie-stained gel of 1st IMAC purification, loaded with a maximum of 25 µg/well total protein, showing the co-elution of chaperones and nickel-binding proteins with 6×His-M2HOT. The estimated imidazole concentrations are given above the **fraction labels:** FT = flow-through, W = wash, G = gradient. **(B)** Coomassie-stained gel, loaded with a maximum volume per well of 15 µL, showing 2nd IMAC purification employing detergents and ATP/Mg^{2+} washes to remove chaperones (HSPs) prior to the imidazole elution. **Fraction labels:** IMAC(G12-16) = pooled IMAC fractions, FT = flow-through, DW = detergent wash, TW = Tris wash, MG = Mg^{2+} equilibration, ATP = ATP/Mg^{2+} wash, IMD = 200 mM imidazole elution.

the hydrolysis of ATP, it seemed possible to use detergents and ATP/Mg^{2+} washes to help disassociate the chaperones from 6×His-M2HOT while it was bound to the Ni^{2+}-NTA resin.

Sonication and centrifugation of a fresh 1.65 g BL21(DE3) cell pellet in lysis buffer yielded 22.34 mL of clarified lysate with a total protein concentration of 22.43 mg/mL, as determined by Bradford assay, for a total protein mass of 501.09 mg. The IMAC purification was repeated as before, but this time with a shallower (25 – 200 mM) imidazole gradient during elution. This resulted in elution of 6×His-M2HOT at a slightly lower imidazole concentration of 72 mM than in the previous purification. The SDS-PAGE analysis of the column fractions shows a similar story, with Coomassie staining revealing the same contaminating chaperone and nickel-binding protein bands co-eluting with 6×His-M2HOT (**Figure 3.4A**). Column

fractions G12–G16 were pooled for a final volume of 7.89 mL, a total protein concentration of 2.20 mg/mL, and a total protein mass of 17.32 mg (3.46% yield).

Pooled IMAC fractions G12–16 were loaded onto a new Ni^{2+}-NTA column and subjected first to 9 column volumes (CV) of detergent wash (Tween-20 and CHAPS), followed by 9 CV of ATP/Mg^{2+}, and then finally elution with 200 mM imidazole. SDS-PAGE analysis of this heat shock protein (HSP) removal strategy showed promise, as a large amount of *Dna*K and *Gro*EL were removed during this process (**Figure 3.4B**). Although some *Dna*J was also removed, a significant amount of it continued to co-elute with 6×His-M2HOT along with the nickel-binding proteins. Assuming that none of these impurities are physically bound to 6×His-M2HOT, one option is to remove the His-tag from 6×His-M2HOT and pass it back through the Ni^{2+}-NTA column where the impurities should be trapped on the column and M2HOT should elute in the flow-through.

The N-terminal His-tag is connected to M2HOT through a linker that terminates in a novel WELQut protease cleavage site (**Figure 3.5**). Enzymatic cleavage occurs on the outside (C-terminus) of the recognition sequence and does not leave additional amino acids on the N-terminus of M2HOT.

MAGS**HHHHHH**GMASMTGGQQMGRSG**WELQ****AVSNIRYG…**

Figure 3.9. The N-terminus of 6×His M2HOT showing the WELQut cleavage site. The N-term methionine (ATG start) is shown in **green**, the 6×His-tag is in **blue**, the WELQut cleavage site is in **red**, and the start of M2HOT is **bolded**. The WELQut protease cleaves on the C-terminus of glutamine (**Q** in **red**).

Following WELQut protease digestion and release of the 3.2 kDa N-terminal fragment, it was possible to resolve M2HOT (45.0 kDa) from 6×His-M2HOT (48.2 kDa) on SDS-PAGE. The results of the WELQut reaction optimization are presented in **Figure 3.6A** and **3.6B** as Coomassie- and InVision-stained SDS-PAGE gels, respectively. The best results, 25 ug and

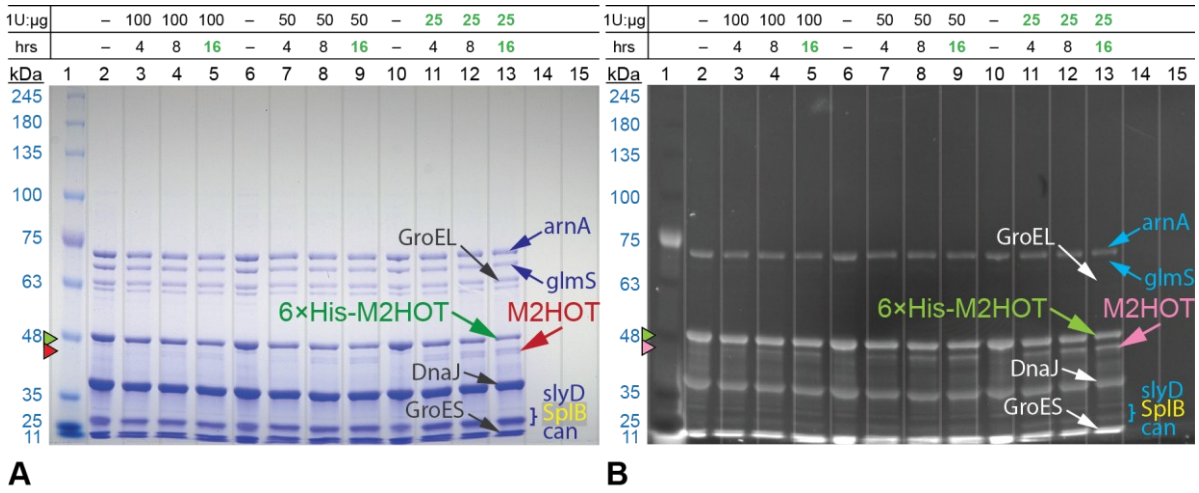

Figure 3.10. SDS-PAGE (InVision) of WELQut protease optimization. Optimization conditions are marked above the well numbers, with "1U:ug" representing the ratio of protease enzyme units (U) to ug of His-tagged protein, while the duration of the reaction at 30 °C is given under "hrs". Molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). The 6×His-M2HOT bands are indicated by the **green** arrows, with the height of the arrow tips relative to the protein ladder marked by **green** triangles. The M2HOT bands are indicated by the **red**/**pink** arrows, with the height of the arrow tips relative to the protein ladder marked by **red**/**pink** triangles. The calculated masses of 6×His-M2HOT and M2HOT are 48.2 kDa and 45.0 kDa, respectively. Chaperones are labeled with **black** (or white) text/arrows, nickel-binding proteins are labeled with **dark blue** (or **cyan**) text/arrows, and WELQut protease (25 kDa) is labeled in **yellow**. **(A)** Coomassie-stained gel of the protease reaction optimization showing that the reaction is largely incomplete even at the highest concentration of protease and longest recommended duration. **(B)** InVision-stained gel for the same reactions, also showing the reaction was largely incomplete. Despite the absence of the His-tag, M2HOT (Fe^{2+} binder) retains the ability to bind the InVision stain in a manner similar to DnaJ (Zn^{2+} binder).

16 hrs, are highlighted in green above the gel images. Even with a high concentration of protease (1 U protease per 25 µg 6×His-M2HOT, or 0.04 U/µg), and for 16 hours at 30 °C, the reaction remained largely incomplete. As a note regarding the gel images, there is significant double banding present with some of the protein bands, especially GroEL and HOT. This was not observed in other optimizations and appears to be an artifact caused by incomplete reduction of disulfide bonds during SDS-PAGE sample preparation.

The lack of complete removal of the His-tag from 6×His-M2HOT, under what should be the most generous and optimal reaction conditions, suggests the possibility that the protease does not have complete access to the cleavage site. Whether this is because the site is occluded due to some protein-protein interaction (i.e., the presence of a chaperone or some other protein), or that the cleavage site is too close to the surface of 6×His-M2HOT remains a

question. It was decided that 6×His-M2HOT needed be further purified by removing the nickel-binding proteins, to rule out the possibility that their presence was impacting the WELQut reaction.

In the end, fraction IMD11 (**Figure 3.4B**) had a volume of 12 mL, and a total protein concentration of 0.34 mg/mL, for a total protein mass of 4.11 mg (0.82% overall yield). From the Coomassie-stained SDS-PAGE gel, the 6×His-M2HOT band represents approximately 15% of the stained proteins in this fraction, placing the estimate of soluble expressed 6×His-M2HOT at 0.5 mg.

Expression of 6×His-M2HOT in NiCo21(DE3) Cells

The pLATE52 plasmid encoding the 6×His-M2HOT fusion protein was transformed into the New England Biolabs NiCo21(DE3) *E. coli* cell line along with the Takara chaperone plasmid set pG-KJ8 that co-expresses *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL. This produces nearly the same expression system as the BL21(DE3)/pG-KJ8 combination described in the previous section, with the exception that several of the contaminating nickel binding proteins have been modified in the NiCo21(DE3) cell line so that they either do not bind the Ni^{2+}-NTA resin or possess a chitin binding domain that allows them to be removed during purification using a chitin column. The expression of soluble 6×His-M2HOT in NiCo21(DE3)/pG-KJ8 cells is therefore comparable to that observed in the BL21(DE3)/pG-KJ8 system. Specifically, the co-expression of the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones in NiCo21(DE3) results in the recovery of soluble 6×His-M2HOT in the lysate supernatant, with some 6×His-M2HOT still observed within the insoluble fraction.

Sonication and centrifugation of an 8.16 g NiCo21(DE3) cell pellet in lysis buffer yielded 41.84 mL of clarified lysate with a total protein concentration of 18.96 mg/mL, as determined by Bradford assay, for a total protein mass of 793.44 mg.

Purification of 6×His-M2HOT from NiCo21(DE3) Cells

The purification of 6×His-M2HOT from NiCo21(DE3) cells followed a similar strategy to the purification from BL21(DE3) cells, with the exception that HSP removal step was integrated into the initial IMAC purification, and this purification was now followed by a chitin column to remove modified nickel-binding proteins.

The results of the initial hybrid IMAC/HSP purification, as analyzed by SDS-PAGE, are presented in **Figure 3.7A**. Chaperones were removed while 6×His-M2HOT was bound to the Ni^{2+}-NTA resin, using detergents and ATP/Mg^{2+} during the first phase of the purification (*lanes 3–6*, **Figure 3.7A**). The subsequent 50 mM imidazole wash step (*lanes 7–10*, **Figure 3.7A**) wasn't nearly long enough and a contaminating band (labeled X in **Figure 3.7**) is visible just below the 6×His-M2HOT band in the elution fractions (*lanes 11–14*). This band is not present in previous purifications and was absent in subsequent purification where the imidazole wash duration was increased. Other notable contaminants that are present in the elution fractions along with 6×His-M2HOT are the nickel-binding protein *arn*A and the chaperone *Dna*J. IMAC/HSP column fractions E11–E13 were pooled and concentrated to a volume of 10.00 mL, with a total protein concentration of 1.69 mg/mL and a total protein mass of 16.94 mg (2.14% yield).

The pooled IMAC fractions E11–E13 were passed through a chitin column to remove all nickel-binding proteins that had been modified to contain a chitin-binding domain and the results of this purification are shown on the SDS-PAGE gel in **Figure 3.7B** (*lanes 5 and 6*),

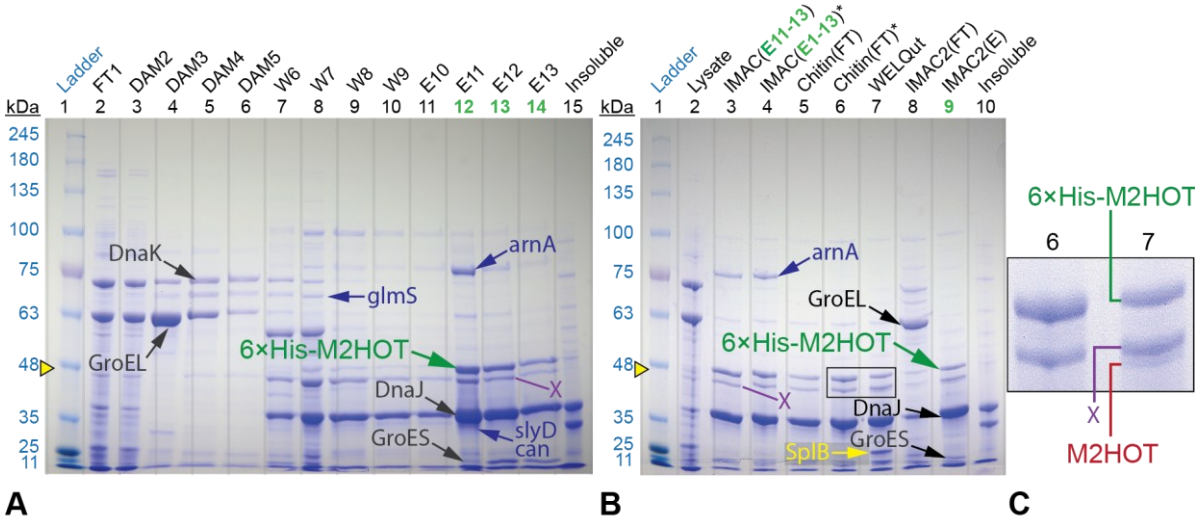

Figure 3.11. IMAC/HSP and chitin purifications of 6×His-M2HOT from NiCo cells. Molecular weights are shown in **blue** to the left of the protein ladder (*Lane 1*). The 6×His-M2HOT bands are indicated by the **green** arrows, with the band height relative to the protein ladder marked by a **yellow** triangle. The calculated mass of 6×His-M2HOT is 48.2 kDa. Chaperones are labeled with **black** text/arrows, nickel-binding proteins are labeled with **dark blue** text/arrows, an unknown impurity is marked with a **purple** X, and the WELQut protease (SplB; 25 kDa) is labeled in **yellow**. Fractions with **green** well numbers were collected for further analysis or processing. **(A)** Coomassie-stained SDS-PAGE gel of IMAC fractions showing removal of chaperones and nickel-binding protein glmS. **Fraction labels:** FT = flow-through, DAM = detergent/ATP/Mg^{2+} wash; W = 50 mM imidazole wash, and E = 150 mM imidazole elution. **(B)** Coomassie-stained SDS-PAGE gel showing results of chitin purification, WELQut digestion, and 2nd IMAC purification. **Fraction labels:** IMAC(E11-13) = pooled IMAC fractions, IMAC(E11-13)* = pooled/concentrated/buffer-exchanged fractions, Chitin(FT) = chitin flow-through, Chitin(FT)* = concentrated flow-through, WELQut = protease digestion, IMAC2(FT) = 2nd IMAC flow-through, IMAC2(E) = 2nd IMAC 150 mM imidazole elution. **(C)** An enlargement of the area bounded by the box in (B). showing two bands prior to WELQut digestion (6×His-M2HOT and impurity X; *Lane 6*), and then the appearance a third band (M2HOT) below impurity X following the WELQut digestion (*Lane 7*).

where the flow-through appeared to be completely free of the *arn*A band. *Dna*J and *band X* were the only remaining significant contaminations after the chitin column. The chitin flow-through was concentrated to a volume of 4.71 mL, with a total protein concentration of 1.87 mg/mL and a total protein mass of 8.79 mg (1.11% yield). The WELQut digestion was repeated with this chitin-purified solution to determine whether the protease reaction would go to completion with the additional contaminants removed. Unfortunately, the digestion remained largely incomplete. *Lane 6* in **Figure 3.7B** is the chitin-purified protein solution prior to digestion, while *lane 7* is the solution after a 16-hour room temperature incubation with a very high concentration of WELQut protease (0.1 U/µg). **Figure 3.7C** shows a closeup of *lanes 6*

and 7 in the region of 50–40 kDa, where a small amount of the M2HOT digestion product is visible, just below contaminant *band X*. It appears that the protease still does not got to completion after the chitin column, suggesting that the recognition site may be too close to the protein surface, or it remains occluded by interaction with one or more chaperones.

The incomplete WELQut digestion was purified by IMAC and this time the 50 mM imidazole wash step, which traditionally follows the collection of flow-through, was skipped in favor of proceeding directly to the 150 mM imidazole elution. At the time *band X* beneath 6×His-M2HOT was believed to be premature proteolytic digest of 6×His-M2HOT to M2HOT, and not an unrelated contaminant as is now suspected. Although this was a missed opportunity to possibly wash away band X with an intermediate imidazole concentration, the second pass through the Ni^{2+}-NTA appears to have liberated additional GroEL, which was observed as a dark band in the flow-through fraction (*lane 8*). It's worth noting here that the SDS-PAGE gel in **Figure 3.7B** was loaded using 5 µg total protein per well. When comparing lanes on gels loaded in this manner, it is common for bands to vary in intensity from lane to lane depending on the overall protein composition of the samples. This is why GroEL suddenly appears in 2nd IMAC flow-through (*lane 8*) despite its *apparent* absence in *lanes 3–7*, where it is just barely visible. 6×His-M2HOT eluted from the Ni^{2+}-NTA in a single fraction (*lane 9*, **Figure 3.7B**). Apart from 6×His-M2HOT (48 kDa), the two other prominently stained protein bands in the elution fraction are *band X* (45 kDa) and *Dna*J (40 kDa). Whether *Dna*J coelutes because it is associated with 6×His-M2HOT or the Ni^{2+}-NTA has yet to be determined. Size-exclusion or hydrophobic-interaction chromatography may be able to answer this question.

The elution fraction had a volume of 5.00 mL, a total protein concentration of 1.27 mg/mL and a total protein mass of 6.35 mg (0.80% yield). Assuming the 6×His-M2HOT band

represents 5% of the stained proteins in this fraction, the estimate of soluble expressed 6×His-M2HOT would be 0.32 mg. A sample of the 45 kDa band was cut from the gel, subjected to in-gel digest with reductively methylated trypsin, and analyzed by liquid-chromatography tandem mass spectrometry. Peptide mass fingerprinting confirmed the 45 kDa band to be 6×His-M2HOT. A sample of the elution fraction was tested for HOT enzyme activity, but none was detected.

3.3 Conclusion

The full-length isoform of human HOT (M1HOT), which includes the purported N-terminal mitochondrial targeting peptide, was successfully expressed in *E. coli* ER2523 cells as a soluble fusion protein with the help of an N-terminal maltose-binding protein (MBP) affinity tag. Besides functioning as an affinity tag for protein purification, the MBP also served to increase the overall solubility of the fusion protein. Expression of the MBP-M1HOT fusion protein in the SHuffle T7 Express *E. coli* cell line, which facilitates the formation of disulfide bonds in the cytoplasm, did not result in an increase in MBP-M1HOT solubility. This suggests that disulfide bond formation may not be critical for HOT solubility. The co-expression of the molecular chaperones *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL also did not enhance the MBP-M1HOT solubility, indicating that the N-terminal maltose-binding protein alone was sufficient for expression of soluble MBP-M1HOT.

A shorter isoform of human HOT (M2HOT), in which the N-terminal mitochondrial targeting peptide was replaced by a 6×His-tag, was successfully expressed as a soluble protein in both BL21(DE3) and NiCo21(DE3) *E. coli* cell with the assistance of co-expressed molecular chaperones. Co-expression of the chaperones *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL resulted in the greatest solubility of 6×His-M2HOT compared to the diminished solubility observed with co-expression of the chaperones *tig* and *gro*ES-*gro*EL. Because the expression of *tig* did not enhance 6×His-M2HOT solubility, X-*cis*-proline bonds and peptidyl-prolyl *cis*/*trans* isomerization do not appear to be factors that influence HOT protein solubility. 6×His-M2HOT was completely insoluble without co-expressed chaperones in both the BL21(DE3) and NiCo21(DE3) lines, indicating that removal of the N-terminal mitochondrial targeting peptide had no effect on HOT solubility. Expression of 6×His-M2HOT in SHuffle

T7 Express cells with the co-expression of chaperones resulted in a decrease in the overall expression and solubility of HOT compared to the BL21(DE3) and NiCo21(DE3) lines. In addition to providing further evidence that disulfide bond formation is not critical for HOT solubility, the overall decrease in 6×His-M2HOT expression indicates that the SHuffle T7 Express cell line may not be totally compatible with the use of these chaperones.

Although analysis of the predicted structure for M1HOT suggested that the N-terminal mitochondrial targeting peptide, three potential disulfide bonds, and two *cis*-proline residues may impact HOT protein folding and solubility, it was found that none of these factors appeared to influence the solubility of HOT to any observable degree. An investigation into the impact of lysis buffer composition on HOT solubility revealed that only supplementation with reductants and high concentrations of arginine produced observable increases in HOT solubility during protein extraction. The most meaningful gains in HOT solubility were made by lowering the concentration of the inducer IPTG to 0.1 mM and expression temperature to 16 °C, and with the co-expression of the chaperones *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL.

MBP-M1HOT was purified nearly to homogeneity with an amylose column, though the yield was poor, with 4.79 mg of pure protein recovered from 194 mg total protein in the cell lysate (2.5% yield). This lysate was obtained from only 4.08 g of cells, so scaling up this purification would be trivial and yield sufficient protein for analysis. In-gel tryptic digestion and liquid chromatography tandem mass spectrometry (LC-MS/MS) were used to confirm the identity of the purified 93 kDa SDS-PAGE band as MBP-M1HOT.

6×His-M2HOT was purified from NiCo21(DE3) cell lysate with Ni^{2+}-NTA and chitin columns, which resulted in a partially purified 6×His-M2HOT with an extremely poor yield. The induced chaperone *Dna*J was a major contaminant, accounting for the majority of the

protein in the recovered column fraction. From the SDS-PAGE bands, it was estimated that 0.32 mg of the protein in the fraction was 6×His-M2HOT, for a yield of 0.04% from 793 mg total protein in the lysate, which in turn was obtained from 8.16 g of cell paste. It may be worth scaling up to determine if 6×His-M2HOT can be separated from *Dna*J by some other technique, such as size-exclusion or hydrophobic-interaction chromatography. In-gel tryptic digestion and LC-MS/MS were used to confirm the identity of the purified 45 kDa SDS-PAGE band as 6×His-M2HOT.

Although both MBP-M1HOT and 6×His-M2HOT were expressed in soluble form, neither displayed enzyme activity when assayed. M1HOT may be incorrectly or partially folded, yet soluble entirely due to the N-terminal MBP. Likewise, M2HOT may also be in some non-native conformation or partially unfolded and bound to the chaperone *Dna*J. Analysis of the folded state could be examined by circular dichroism or fluorescence spectroscopy, in combination with small-angle X-ray scattering. Also, it remains to be seen whether expressed MBP-M1HOT and 6×His-M2HOT contain the NAD^+ and Fe^{2+} cofactors required for enzymatic activity. Electrospray ionization mass spectrometry should be able to detect the presence of NAD^+ in the purified protein sample, while inductively coupled plasma mass spectrometry would be able to detect Fe^{2+}.

3.4 Materials and Methods

3.4.1 MBP-M1HOT in ER2523 Cells

pMAL-c5X Vector

The ADHFe1 gene coding sequence (CCDS 6190.2) from *Homo sapiens* (Taxon ID 9606), which encodes the full-length human HOT (Uniprot Q8IWW8), was codon optimized for expression in *E. coli* and then synthesized by Life Technologies™. The full-length sequence has also been designated on UniProt as *human HOT isoform 1* (Q8IWW8-1). After removing any 3′ extensions from the ADHFe1 insert with the New England Biolabs (NEB) Quick Blunting Kit and cutting it with SBfI, the insert was purified by agarose gel electrophoresis. The purified ADHFe1 insert, with 5′ blunt end and SbfI overhang at the 3′ end, was ligated at room temperature into the pMAL-c5X vector (NEB), which had been previously digested with XmnI and SbfI and also gel purified.

Transformation of DH5α and ER2523 *E. coli* Cells

2 µL of the ligation reaction was mixed with 50 µL DH5α competent *E. coli* cells (ThermoFisher), transformed using the manufacturer's 30-second heat-pulse transformation protocol, and cells were plated onto LB-agar supplemented with 100 µg/mL ampicillin. The transformation efficiency was 2.51×10^5 CFU/µg. Two colonies from the plate were grown overnight at 37 °C in 5 mL of LB broth supplemented with 100 µg/mL ampicillin and miniprepped (303.6 ng/µL; $A_{260/280} = 1.93$) for DNA sequencing, which confirmed that the construct had been successfully inserted into the pMAL-c5X vector. One of these colonies was used to make a 50% glycerol stock of the transformed DH5α cells.

The miniprepped plasmid was also used to transform ER2523 competent *E. coli* cells (NEB) following the manufacturer's 30-second heat-pulse transformation protocol, with selection on LB-agar with 100 µg/mL ampicillin, resulting in a transformation efficiency of 1.00×10^6 CFU/µg. SDS-PAGE analysis of a pilot expression with 0.3 mM IPTG (see next section for expression details) confirmed the presence of the 93 kDa MBP-Xa-M1HOT expression product. A control expression without IPTG was also performed. Upon confirmation that MBP-M1HOT was successfully expressed, the transformed cells were grown to $OD_{600} \approx 0.5$, at which point equal volumes of suspended cells were mixed with sterile 50% glycerol to make 25% glycerol stocks that were stored at -80 °C.

The transformation of pMAL-c5X/M1HOT and pG-KJ8 vectors into BL21(DE3) and SHuffle T7 Express *E. coli* cells, as well as protein expression in these systems, is not covered in the methods because these systems did not result in a significant increase in MBP-M1HOT solubility compared to that observed with pMAL-c5X/M1HOT alone in the NEB ER2523 cell line. The other systems were not discussed beyond reporting these observations in the Results and Discussion section.

MBP-M1HOT Expression in ER2523 Cells

Overnight cultures were grown in sterile round bottom polypropylene culture tubes with two-position vent stoppers, and cultures for protein expression were grown in autoclaved baffled Fernbach flasks. Growth media was autoclaved LB-Lennox, pH 7.0, supplemented with 0.5% glucose and 100 µg/mL ampicillin. Glucose is required to suppress amylase expression, since this enzyme can degrade the amylose affinity resin used in purification. All cultures grew at 37 °C on shakers set to 200 RPM, with the only exception being that the incubation temperature was lowered to 16 °C just prior to induction of protein expression.

Overnight cultures used to seed media in Fernbach flasks at a ratio of 1:100, which generally gave a starting OD_{600} of 0.05 – 0.1. For protein expression, once the OD_{600} reached 0.3–0.4, which took approximately 300 minutes following inoculation, the temperature of the incubation was lowered to 16 °C. 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added when the $OD_{600} \approx 0.5$, which took approximately another 60 minutes after the temperature was lowered. Induction of expression was allowed to proceed at 16 °C for 18 – 20 hours, at which the cells (final $OD_{600} \approx 1.4$) were harvested by centrifugation at 5,000 × *g* for 10 minutes at 4 °C. The supernatant (exhausted media) was discarded and cells washed by resuspending the cell paste in phosphate buffered saline (PBS), pH 7.4, by vortexing. The resuspended cells were aliquoted into *tared* 50 mL Falon tubes and centrifuged at 5,000 × *g* for 10 minutes at 4 °C. After discarding the supernatant, the Falcon tubes were weighed again and the mass of the cell pasted was calculated by subtraction. Aliquots of cell paste were either immediately lysed or stored at -80 °C for future use.

Lysis of ER2523 Cells

The *lysis buffer* was sterile filtered (0.2 μm) 50 mM HEPES, 159 mM KCl, 10 mM NaCl, 1 mM EDTA, pH 7.5 at room temperature. The lysis buffer was supplemented with 10 mM β-mercaptoethanol (BME) and 1 mM phenylmethylsulfonyl fluoride (PMSF) prior to cell lysis. During lysis, an *inclusion body buffer* was used to check pelleted cell debris for insoluble MBP-M1HOT and this buffer hand the following composition: 50 mM Tris, 2% sodium dodecyl sulfate (SDS), pH 6.8 at room temperature. Prior to use, the inclusion body buffer was supplemented with 100 mM dithiothreitol (DTT).

Cell lysis was accomplished by sonication. Ice-cold lysis buffer was added to Falcon tubes of cell paste that had been thawed on wet ice, at a ratio of 5 mL of lysis buffer to 1 g cell paste.

The cell paste was suspended in the lysis buffer with intermittent vortexing. Once the suspended cell paste could be pipetted up and down easily with a serological pipette, the suspension was transferred into a glass or metal beaker of appropriate volume, which was secured by clamps and immersed into an ice bath so that the cell suspension and water levels were at the same height. The 1/2-inch probe of a QSonica Q500 sonicator was lowered into the beaker, the amplitude set to 40%, the pulse time set to 2 seconds ON, 8 second OFF, and the cell suspension was sonicated for a total processing time of 10 minutes. A sample of the sonicated lysate was set aside in a 1.5 mL Eppendorf tube on ice for the purpose of inclusion body detection, while the remainder of the lysate was centrifuged at 15,000 × *g* for 20 minutes at 4 °C in an Avanti J-E centrifuge. The clarified lysate (supernatant) was transferred to a tared 50 mL Falcon tube and the volume of clarified lysate was estimated by a subtraction of masses.

The sample of the crude lysate in the 1.5 mL Eppendorf tube centrifuged in a was centrifuged at 15,000 × *g* for 20 minutes at 4 °C in in a Benchtop Eppendorf Centrifuge 5424R. After discarding the supernatant (soluble fraction), the inclusion body buffer was added to pelleted cell debris at a ratio of 500 µL of buffer to 100 µg of pelleted debris. The sample was solubilized by a combination of placing it into a heating block set to 95 °C and vortexing, with care taken to vent any built-up of pressure in the tube.

MBP-M1HOT Purification by Amylose Affinity Chromatography

The N-terminal MBP affinity-tag was used to purify MBP-M1HOT on amylose resin, supplied with the NEB pMAL™ Protein Fusion & Purification System (E8200S). A Bio-Rad Econo-Column® (1 × 10 cm) was loaded with NEB Amylose Resin High Flow to a bed height of 6 cm, a 5.0 mL column volume (CV). The column was fitted with a Bio-Rad Econo-Column® Flow Adaptor that was used to pack the column as well as load and elute the sample,

and tubing was used to connect the adaptor to a Bio-Rad Model EP-1 Econo Pump. A T-valve with Luer-Lok™ fittings was installed halfway between the adaptor and the pump to interrupt the mobile phase and inject the sample directly onto resin, through the flow adaptor. Fractions from the column were collected in 13 × 100 mm VWR disposable borosilicate glass culture tubes that were placed into a Bio-Rad Model 2110 Fraction Collector.

In a refrigerated room, the amylose resin was equilibrated with 5 CVs of the HEPES column buffer (HCB) at a flow rate of 0.5 mL/min (38.2 cm/hr). HCB was 50 mM HEPES, 159 mM KCl, 10 mM NaCl, 1 mM EDTA, 10 mM BME, and pH 7.5 at room temperature. The clarified lysate was loaded slowly onto the column with a syringe that was attached to a Luer-Lok port on the T-valve. The flow through was collected as a single 25.0 mL fraction in 50 mL Falcon tube. The pump was started, and the column was washed with 10 CVs of HCB at a flow rate of 0.5 mL/min. The column wash was collected as 5 × 10.0 mL fractions in 15 mL Falcon tubes. MBP-M1HOT was eluted with 3 CVs of HEPES elution buffer (HEB) at 0.50 mL/min and 1.0 mL fractions were collected in the borosilicate glass culture tubes using the fraction collector. HEB was just HCB supplemented with 10 mM maltose. The amylose resin was regenerated with 10 CVs of 1% (w/v) SDS in 18 MΩ water, followed by 10 CVs of 18 MΩ water only.

3.4.2 6×His-M2HOT in BL21(DE3) and NiCo(DE3) Cells

pLATE52 Vector

The coding sequence for *human HOT isoform 2* (UniProt Q8IWW8-2), was subcloned out of the pMAL-c5X vector and ligated into the pLATE52 vector using the Thermo Scientific aLICator Ligation Independent Cloning and Expression System Kit 4 (#K1281). The sequence

begins at nucleobase 148 (guanine), after the second in-frame Kozak sequence and ATG start codon. It should be noted that in the *human sequence*, the codon beginning at position 148 is GCT, while in the sequence optimized for expression in *E. coli* it has been modified to GCC. The forward and reverse PCR primers used to subclone *human HOT isoform 2* were:

Forward: 5′–GGTTGGGAATTGCAAGCCGTGAGCAATATTCGTTATGGT

Reverse: 5′–GGAGATGGGAAGTCATTAATACAGTTTCATGCTTGCTTCAAACAG

where the underlined nucleobases belong to coding sequence for *HOT isoform 2*. The primers were HPLC-purified and non-phosphorylated. The insert was PCR amplified using the Phusion™ Plus Green PCR Master Mix (Thermo Scientific) and the manufacturer's 3-step protocol and 25 cycles. The PCR products were purified by electrophoresis on 0.8% agarose gels and, following SYBR™ Safe DNA Gel Stain (Invitrogen), a single band (1,287 bp) was observed and extracted using PureLink® Quick Gel Extraction Kit (Life Technologies). NanoDrop analysis of the purified M2HOT insert gave a concentration of 34.2 ng/μL and $A_{260/280}$ of 2.23.

The purified PCR product (M2HOT) was inserted into the pLATE52 vector by following the ligation independent cloning protocol supplied with the aLICator kit. A control PCR fragment, supplied with the kit, was used to assess the efficiency of the LIC reaction. Briefly, 2.43 μL of the M2HOT insert was combined with the supplied T4 DNA Polymerase in 5X LIC Buffer and nuclease-free water, then briefly vortexed and centrifuged. After a 5-minute room temperature incubation, the reaction was stopped by the addition of EDTA. The annealing reaction was immediately initiated with 1 μL of pLATE52, LIC-ready vector, followed by a brief vortex and centrifugation. The rection was allowed to proceed for 5 minutes, after which the annealed mixture was used to transform DH5α *E. coli* competent cells (ThermoFisher)

following the manufacturer's 30-second heat-pulse transformation protocol. Control transformations were also performed with pUC19 control plasmid and LIC-control (pLATE52 + control PCR fragment). Following overnight incubation at 37 °C on LB-agar + 100 µg/mL, the transformation efficiencies of DH5α with pUC19, LIC-control, and pLATE52/M2HOT were 1.81×10^9, 4.02×10^4, and 1.39×10^6 CFU/µg, respectively.

Colony PCR was performed on 4 colonies from the pLATE52/M2HOT agar plate using the supplied pLATE52 LIC Forward and Reverse sequencing primers, the Thermo Scientific DreamTaq Green PCR Master Mix (2X) and manufacturer's protocol. The sequencing primers will produce a PCR product that is 264 bp, plus insert size, from a correctly annealed pLATE52 vector. Analysis of the PCR mixtures on an agarose gel, following SYBR™ Safe DNA Gel Stain (Invitrogen), showed the presence of a single 1,518 bp band for all four colonies selected, confirming the LIC reaction was successful. Two of the four colonies were grown in LB-broth supplemented with 100 µg/mL ampicillin and miniprepped to produce plasmid DNA for sequencing. The plasmid concentrations (and $A_{260/280}$) for Colony 1 and Colony 2 were 122.0 ng/µL (1.87) and 122.0 ng/µL (1.87), respectively. DNA sequencing of the pLATE52/M2HOT plasmids with the pLATE52 LIC Forward and Reverse sequencing primers confirmed that both Colony 1 and 2 contained the M2HOT sequence correctly inserted into the pLATE52 vector.

Transformation of SHuffle T7 Express, BL21(DE3) and NiCo(DE3) Cells

The miniprep-purified pLATE52/M2HOT vector from Colony 1 was used to transform NEB SHuffle T7 Express and Thermo Scientific BL21(DE3) and competent *E. coli* cells using the manufacturers' 30-second heat-pulse transformation protocols. All selections were performed on LB-agar with 100 µg/mL ampicillin. The transformation efficiencies for SHuffle T7 Express and BL21(DE3) were 1.00×10^6 and 3.01×10^5 CFU/µg, respectively.

For studies with chaperones, both plasmids pG-KJE8 and pG-Tf2 from the Takara Chaperone Plasma Set were used, independently, to co-transform NEB SHuffle T7 Express and Thermo Scientific BL21(DE3) competent *E. coli* cells, along with pLATE52/M2HOT, using the manufacturers' 30-second heat-pulse transformation protocols and selection on LB-agar with 50 µg/mL ampicillin and 20 µg/mL chloramphenicol. The co-transformation efficiencies for BL21(DE3) and SHuffle T7 Express are presented in **Table 3.1** for simplicity.

Table 3.2. Co-transformation efficiencies for BL21(DE3) and SHuffle T7 Express.

		Co-transformation Efficiencies (CFU/µg)	
Plasmid 1	**Plasmid 2**	**SHuffle T7 Express**	**BL21(DE3)**
pLATE52/M2HOT	pG-KJE8	2.51×10^5	1.51×10^5
	pG-Tf2	3.31×10^5	1.60×10^3

The NEB NiCo(DE3) cell line was co-transformed *only* with pLATE52/M2HOT and the Takara pG-KJE8 plasmid, because more soluble 6×His-M2HOT was obtained from BL21(DE3) cells expressing pG-KJE8 than those expressing pG-Tf2. The co-transformation of NiCo(DE3) cells with pLATE52/M2HOT and pG-KJE8 was performed using NEB's 30-second heat-pulse transformation protocol and selection on LB-agar with 50 µg/mL ampicillin and 20 µg/mL chloramphenicol. The co-transformation efficiency for NiCo(DE3) cells with pLATE52/M2HOT and pG-KJE8 was 1.94×10^6 CFU/µg.

For all transformed cells described above, colonies were selected, grown in media supplemented with the appropriate selection agent, and then induced as described in the next section in order to confirm the presence of the expression constructs (i.e., 6×His-M2HOT and chaperones) by SDS-PAGE. Upon confirmation that the proteins of interest were expressed, cells were grown to $OD_{600} \approx 0.5$, at which point equal volumes of suspended cells were mixed with sterile 50% glycerol to make 25% glycerol stocks that were stored at -80 °C.

6×His-M2HOT Expression in SHuffle T7 Express and BL21(DE3) Cells

For the NEB SHuffle T7 Express and Thermo Scientific BL21(DE3) *E. coli* cell lines, 6×His-M2HOT was expressed both alone and with either: a) the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones (pG-KJE8 plasmid); or b) the *tig* and *gro*ES-*gro*EL chaperones (pG-Tf2 plasmid). pG-KJE8 and pG-Tf2 are two of the five plasmids supplied with the Takara Chaperone Plasmid Set (Cat. #3340).

Overnight cultures were grown in sterile round bottom polypropylene culture tubes with two-position vent stoppers, and cultures for protein expression were grown in autoclaved baffled Fernbach flasks. All cultures grew at 37 °C on shakers set to 200 RPM, with the only exception being that the incubation temperature was lowered to 16 °C just prior to induction of protein expression. The base media for all overnight growths and protein expressions was autoclaved 2×YT, adjusted to pH 7.0, then supplemented with the appropriate antibiotic for selection and inducers for protein expression.

When 6×His-M2HOT was expressed alone, the base media was supplemented with 100 µg/mL ampicillin to maintain the pLATE52/M2HOT plasmid. The ampicillin concentration was lowered to 50 µg/mL when either the pG-KJE8 or pG-Tf2 plasmids had been co-transformed into the cell line, since both of these plasmids required supplementation with an additional 20 µg/mL of chloramphenicol.

For induction of 6×His-M2HOT expression, 0.05 mM IPTG was used. For induction of *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperones, both L-arabinose and tetracycline are required, since the two groups of chaperones are under the control of different promotors on the pG-KJE8 plasmid. Initial concentrations for L-arabinose and tetracycline were 0.5 mg/mL and 5 ng/mL, respectively, for pilot studies comparing the different chaperone sets in

BL21(DE3) and SHuffle T7 Express as well as the expression used in the preliminary IMAC purification presented in **Figure 3.3**. L-arabinose and tetracycline were later increased to 2.0 mg/mL and 10 ng/mL, respectively, for all subsequent purifications. The *tig* and *gro*ES-*gro*EL chaperones are under control of a single promotor on the pG-Tf2 plasmid, and expression of these chaperones was induced by the addition of 5 ng/mL tetracycline during the pilot studies.

Overnight cultures of SHuffle T7 Express and BL21(DE3) were used to seed media in Fernbach flasks at a ratio of 1:40, which generally gave a starting OD_{600} of 0.05 – 0.1. When chaperone expression was required, the inducers for the chaperones were present in the media at the time of inoculation. For 6×His-M2HOT expression, once the OD_{600} reached 0.3–0.4, which took 150–180 minutes following inoculation, the temperature of the incubation was lowered to 16 °C. 0.05 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added when the $OD_{600} \approx 0.5$, which generally took an additional 30 minutes after lowering the temperature.

Induction of expression was allowed to proceed at 16 °C for 18 – 20 hours, at which the cells (final $OD_{600} \approx 2.0 – 2.2$) were harvested by centrifugation at 5,000 × *g* for 10 minutes at 4 °C. The supernatant (exhausted media) was discarded and cells washed by resuspending the cell paste in phosphate buffered saline (PBS), pH 7.4, by vortexing. The resuspended cells were aliquoted into *tared* 50 mL Falon tubes and centrifuged at 5,000 × *g* for 10 minutes at 4 °C. After discarding the supernatant, the Falcon tubes were weighed again and the mass of the cell pasted was calculated by subtraction. Aliquots of cell paste were either immediately lysed or stored at -80 °C for future use.

6×His-M2HOT Expression in NiCo21(DE3) Cells

For the NEB NiCo21(DE3) *E. coli* cell line, 6×His-M2HOT was co-expressed with the *dna*K-*dna*J-*grp*E and *gro*ES-*gro*EL chaperone group encoded by the Takara pG-KJE8

plasmid. Therefore, the growth and expression conditions were identical to those used in the latter BL21(DE3) experiments. Briefly, the media was 2×YT supplemented with 50 µg/mL ampicillin, 20 µg/mL chloramphenicol for overnight growths, and the same media supplemented with an additional 2 mg/mL L-arabinose and 10 ng/mL tetracycline when the expression media was inoculated (1:100) with the overnights. All overnights and inoculated media were grown at 37 °C on shakers set to 200 RPM, and incubation temperature was lowered to 16 °C when the OD_{600} reached 0.3–0.4. 6×His-M2HOT expression was induced with 0.05 mM IPTG when the OD_{600} was near 0.5 and the expression continued at 16 °C for 18–20 hours, at which the point the cells were harvested by centrifugation at 5,000 × *g* for 10 minutes at 4 °C. Following a wash with PBS and second spin, the cell paste in 50 mL Falcon tubes was estimated by subtracting the tared tube weight from that of the tube with cell paste. The cells were either immediately lysed or frozen at -80 °C.

Lysis of SHuffle T7 Express, BL21(DE3), and NiCo21(DE3) Cells

Cell lysis was accomplished by sonication in a manner similar to that described for the ER2523 cells (see *Section 3.4.1*), with minor differences mostly involving the compositions of the lysis buffer. Briefly, the similarities are as follows. Cell paste was suspended in ice cold lysis buffer (to be described) by vortexing and transferred to a glass or metal beaker secured by clamps and immersed in an ice water bath to absorb any heat generated. Sonication was performed either with a QSonica Q125 fitted with a 1/8-inch probe, for small scale pilot studies, or a QSonica Q500 fitted with a 1/2-inch probe for scaled up expressions. The amplitude was set to 40%, the pulse time set to 2 seconds ON, 8 second OFF, and the cell suspension was sonicated for a total processing time of approximately 10 minutes, or longer, depending on the cell suspension volume and size of the beaker. A sample of the clarified

lysate was always withdrawn and set aside in a 1.5 mL Eppendorf tube in order to assess the amount of insoluble 6×His-M2HOT.

Both the main volume of crude lysate and the withdrawn sample were centrifuged separately at 15,000 × *g* for 20 minutes at 4 °C. The clarified lysate (supernatant) from the main volume was transferred to a tared 50 mL Falcon tube and the volume estimated by subtraction. The supernatant in the centrifuged 1.5 mL Eppendorf tube was discarded and the insoluble material was dissolved with an *inclusion body buffer*, as described previously.

The lysis buffer used to extract 6×His-M2HOT from SHuffle T7 Express, BL21(DE3) and NiCo(DE3) cells was constantly being optimized in order to increase the yield of soluble protein. Therefore, the curated results presented in the Results and Discussion section do not share a common buffer for lysis and purification. The following paragraphs detail the specific buffers used during cell lysis for the individual experiments that were presented.

The lysis buffers used to gauge the effectiveness of co-expressed chaperones to increase 6×His-M2HOT solubility in pilot studies with SHuffle T7 Express and BL21(DE3) cells were high pH and low pH. High pH was 50 mM Tris, 300 mM NaCl, adjusted to pH 8.0 and supplemented with 1 mM PMSF. Low pH was 50 mM MES, 300 mM NaCl, pH 5.5 and supplemented with 1 mM PMSF.

The lysis buffer for BL21(DE3) cells as a test for the IMAC purification (**Figure 3.3**) was 50 mM HEPES, 300 mM KCl, 10 mM imidazole, 0.1% Tween 20, pH 8.0 and supplemented with 1 mM PMSF.

The lysis buffer for BL21(DE3) cells used for the IMAC purification followed by IMAC/HSP-wash (**Figure 3.4**) was 20 mM Na_3PO_4, 300 mM NaCl, 10 mM imidazole, 2%

3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 0.5% Tween 20, pH 7.4 and supplemented with 10 mM BME and 1 mM PMSF.

Lastly, the lysis buffer for NiCo21(DE3) cells used in the one column IMAC/HSP-wash purification was 50 mM HEPES, 150 mM KCl, 1 mM $ZnCl_2$, 10 mM imidazole, 100 mM arginine, 0.5% Tween 20, pH 8.0 and supplemented with 10 mM BME and 1 mM PMSF.

6×His-M2HOT Purification by IMAC and HSP Wash

Purifications of 6×His-M2HOT were performed almost exclusively by IMAC, with a chitin column used to cleanup nickel-binding proteins for expressions in the NiCo21(DE3) cell line. The IMAC resin was HisPur™ Ni-NTA Resin (ThermoFisher, #88221), which was loaded onto a Bio-Rad Econo-Column® (1 × 10 cm) fitted with a Bio-Rad Econo-Column® Flow Adaptor, which was used to pack the column as well as load, wash and elute the sample. Tubing was used to connect the flow adaptor to a Bio-Rad Model EP-1 Econo Pump. A T-valve with Luer-Lok™ fittings was installed halfway between the adaptor and the pump to interrupt the mobile phase and inject the sample directly onto resin, via the flow adaptor. Fractions from the column were collected in 13 × 100 mm VWR disposable borosilicate glass culture tubes that were loaded into a Bio-Rad Model 2110 Fraction Collector. The Ni^{2+}-NTA resin was packed to a bed height of 6.4 cm, giving a total column volume (CV) of 5.0 mL. The volume flow rate was 0.25 mL/min (19.10 cm/hr) while protein was on the column. In all other cases (e.g., column equilibration and regeneration) the flow rate was 0.50 mL/min (38.20 cm/hr). All purifications were performed in a cold room, at 4 °C.

The gradient former (originally designed to make gradient gels) had an inner and outer chamber that could hold buffer, with a valve that controlled the flow between them. The inner chamber led directly to the pump and was fitted with a stir bar, while the outer chamber was

the furthest reservoir from the pump. A buffer with low imidazole concentration was placed in the inner chamber, while an equal volume of buffer with a high concentration was placed in the outer chamber. Uniform mixing was achieved by opening the valve between the two chambers, allowing the more concentrated solution in the outer chamber to flow into the inner chamber where the stir bar would mix the two, before the combined solution was pumped onto the column. This was the basis for creating the imidazole gradient during the elution step. Following the elution, the column was regenerated with 10 CVs of 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 8.0, and then washed with 10 CVs of 18 MΩ water.

The initial IMAC purification of 6×His-M2HOT from BL21(DE3) cells (**Figure 3.3**) used 5 CVs of an equilibration buffer that was similar to the lysis buffer, specifically 50 mM HEPES, 300 mM KCl and 10 mM imidazole, pH 8.0. Following manual loading of the sample using a syringe and the T-valve installed on the tubing, 2 CVs of flow through were collected as a single fraction and 10 CVs of wash buffer were applied to the column. The wash buffer was 50 mM HEPES, 150 mM KCl, and 25 mM imidazole, pH 8.0, and five 2 CV wash fractions were collected. Finally, the target protein was eluted with a with a 25–250 mM imidazole gradient over 10 CVs with the elution buffer and collected as 10 individual fractions. The elution buffer was created by the gradient by combining low and high imidazole buffers, both having a base of 50 mM HEPES, 150 mM KCl buffer, pH 8.0. Low imidazole was 25 mM, while high was 500 mM.

For the second IMAC purification of 6×His-M2HOT from BL21(DE3) cells (**Figure 3.4**) that was followed by a IMAC purification that included an HSP wash. The initial IMAC purification was the same as that described above, with the exception that the base buffer was phosphate buffer-based, to match the lysis buffer. The equilibration buffer was now 20 mM

Na3PO4, 300 mM NaCl, 10 mM imidazole, 2% CHAPS, 0.5% Tween 20, pH 7.4, supplemented with 10 mM BME. The wash buffer had the same basic composition, but with 25 mM imidazole, and the wash was collected over 12 CVs as 6 fractions. The elution gradient buffer was 20 mM Na_3PO_4, 300 mM NaCl, with no detergents and a low and high imidazole concentration of 25 and 400 mM imidazole, for a gradient of 25 – 212 mM imidazole collected over 10 CVs as 10 × 1 CV fractions. The pooled fractions from the initial IMAC purification were concentrated prior to being loaded back onto the Ni^{2+}-NTA resin.

The HSP purification involved first a detergent wash, followed by an ATP wash, and finally elution of the remaining proteins bound to the resin. After 5 CVs of equilibration with 20 mM Na_3PO_4, 50 mM NaCl, 10 mM imidazole, adjusted to pH 7.4 and supplemented with 10 mM BME, the sample was loaded onto the column and 3 CVs of flow through were collected as one fraction. The column was washed over 9 CVs with equilibrium buffer supplemented with 2% CHAPS and 0.5% Tween 20, and 3 fractions were collected. The column was then equilibrated with 3 CVs of the base for the HSP wash, which was 50 mM TRIS, 50 mM KCl, pH 7.5, collected as one fraction. The buffer was switched from phosphate to TRIS out of concern that phosphate would chelate the Mg^{2+} cations required for ATP activity. This was followed by 3 CVs of HSP buffer with 20 mM $MgCl_2$ and collected as one fraction. The same buffer was supplemented this time with 5 mM ATP and used to wash the column over 9 CVs, collecting three fractions. The HSP wash was repeated over 3 CVs (one fraction) to flush Mg^{2+} from the resin, and then the column was eluted with a 50:50 mixture of the low (25 mM) and high (400 mM) imidazole phosphate-based elution buffers over 9 CVs, collecting 3 fractions.

For the combination IMAC and HSP purification of 6×His-M2HOT from NiCo21(DE3) cells (**Figure 3.5**), the base buffer was again matched to the lysis buffer and designed so that no buffer exchanges were required during this purification. This time the equilibration buffer was 50 mM HEPES, 150 mM KCl, and 10 mM imidazole, 10 mM BME, 0.5% Tween 20, pH 8.0, which ran for 5 CVs. Following the sample injection, and collection of 1 CV fraction, the column was washed with an ATP wash. The ATP wash was 10 CVs, collected over 4 fractions, and was comprised of 50 mM HEPES, 50 mM KCl, 10 mM imidazole, 10 mM BME, 0.5% Tween 20, 1 mM $ZnCl_2$, 10 mM $MgCl_2$, and 5 mM ATP, pH 8.0 collected over 10 CVs and 4 fractions. This was followed by a 10 CV imidazole wash over 4 fractions, comprised of 50 mM HEPES, 150 mM NaCl, 50 mM imidazole, 10 mM BME, pH 8.0, collected over 4 fractions. Finally, the bound proteins were eluted with 6 CVs of 50 mM HEPES, 150 mM NaCl, 150 mM imidazole, 10 mM BM, pH 8.0, and collected over 4 fractions.

WELQut Protease Cleavage

Cleavage of the WELQut linker between the N-terminal 6×His-tag and M2HOT was performed using the Thermo Scientific WELQut protease (#EO0861). The reaction was performed on a solution containing 6×His-M2HOT that had been expressed in BL21(DE3) cells and purified first by traditional IMAC and then a second time by IMAC with a set of HSP washes to remove potentially interfering chaperones. Optimization of protease digestion was performed by following the recommendations in the manufacturer's optimization protocol. The reaction buffer was 20 mM Na_3PO_4, 50 mM NaCl, 10 mM imidazole, adjusted to pH 7.4. The imidazole was included so that the reaction could be immediately loaded onto Ni^{2+}-NTA resin for immediate IMAC purification. The protease to substrate ratios were chosen to be 1 Unit of protease to 100, 50, and 25 μg of 6×His-M2HOT, as estimated using the total protein

concentration of an impure sample. Reactions were performed at room temperature and 30 °C, for durations of 4, 8, and 16 hours and cleavage efficiencies were analyzed by SDS-PAGE.

Centrifugal Concentration

For chromatography fractions that were ≤ 6 mL, desalting and/or concentration was performed with a Sartorius Vivaspin® 6 Centrifugal Concentrator (30 kDa MWCO) at 6,000 × *g* and 4 °C in a Beckman JA-20 rotor mounted in a Beckman J2-HS (and J2-21) centrifuge. For chromatography fractions that were ≥ 6 mL, desalting and/or concentration was performed with an Amicon® Ultra-15 30kDa MWCO centrifugal filter, spun at 5000 × *g* at 4 °C in a Beckman JA-10 rotor mounted in a Beckman J2-HS (and J2-21) centrifuge. When desalting or facilitating a buffer exchange, the spin was repeated after the addition of the new buffer, with this exchange being performed a total of two times.

3.4.3 Protein Analysis

Protein Quantitation

Three different methods of determining protein concentration were used for the experimental results presented in this chapter. The Bradford assay and method of Warburg-Christian were used to determine the total protein concentration of impure protein samples, while the A_{280} method was used for fractions where the sample was relatively pure (e.g. MBP-M1HOT).

The dye-based colorimetric method of Bradford[56] was used to determine the concentration of total protein in the lysate and column fractions. Buffer components are known to interfere[57] with the Bradford assay, detergents[58] in particular. Bio-Rad's Protein Assay Dye Reagent Concentrate (500-0006) states it can be used reliably with most detergents, assuming low

concentrations. The Bio-Rad reagent is also linear between 0.2 and 0.9 mg/mL bovine serum albumin (BSA). The BSA concentrations used for the standard curve were 0 (100 mM NaCl), 0.2, 0.4, 0.6, 0.8 and 0.9 mg/mL BSA. The standards were made by dissolving BSA (Fraction V) Heat Shock Treated (FisherBioReagents, BP1600-100) in filter sterilized 100 mM NaCl. In a plastic cuvette (1 cm diameter), 20 µL unknown sample (or standard) were mixed with 980 µL of 1X dye reagent and incubating at room temperature for 15 minutes, then the absorbance at 595 nm was read. Uknown protein concentrations were determined using the BSA standard curve.

The Warburg-Christian method makes use of both A_{260} and A_{280} absorptions, along with an equation, to estimate the total protein concentration in an impure sample.[59] This equation is quoted in various forms in the literature, with differences being attributed to the coefficients used in the formula. The form of the equation used in this research being given in **Equation 3.1**, with the caveat that the A_{280} must be greater than the A_{260} for the result to be valid.

Equation 3.1. The Warburg-Christian Equation for protein concentration.

$$[\text{protein}]\ (\text{mg/mL}) = [(1.31 \times A_{280}) - (0.57 \times A_{260})] \times \text{dilution factor}$$

The Bradford assay, Warburg-Christian and A_{280} method employed used a UV-Vis spectrophotometer running UVProbe v2.61 software. The A_{280} method also employed the Beer-Lambert Law, a 1 cm value for the cuvette pathlength, and the extinction coefficient of 98,920 M^{-1} cm^{-1} for MBP-M1HOT.

Protein Purity

The presence and purity of MBP-M1HOT and 6×His-M2HOT were evaluated by running samples on SDS-PAGE gels, followed by staining with Coomassie Brilliant Blue R-250. For

6×His-M2HOT, an additional in-gel stain visualized by fluorescence imaging was used to probe the gel for presence of the His-tag. SDS-PAGE gels were 7.5% Mini-PROTEAN® TGX™ Precast Protein Gels, 15-well, 15 µl (Bio-Rad #4561026) run in Bio-Rad Mini-PROTEAN 3 cell using a 1X buffer made from Fisher BioReagents® Tris-Glycine-SDS, 10X (powder; Cat # BP1342-1). The gels were run at a constant 200 volts for 30–35 minutes, with power supplied by a BioRad PowerPac™ 3000 (165-5056).

Following an electrophoretic run, gels were removed from their cassettes, rinsed twice for 30 seconds with 18 MΩ water and placed into Coomassie stain overnight with gentle rocking. The Coomassie stain was 0.1% Coomassie BB R-250, 50% methanol, 40% 18 MΩ water, and 10% glacial acetic acid. The gels were destained the following day with Destain 1 (53% 18 MΩ water, 40% methanol, 7% glacial acetic acid), followed by Destain 2 (88% 18 MΩ water, 5% methanol, 7% glacial acetic acid). Gels were photographed on a lightbox using a Nikon D7000, the images processed using Adobe Photoshop CS6 and annotated using Adobe Illustrator CS6. For detection of His-tags in protein bands, the Invitrogen™ InVision™ His-Tag In-Gel Stain was used prior to the Coomassie stain following the manufacturer's protocol. The InVison-stained gel was imaged with the UV transilluminator (302 nm) and UV tray of a Bio-Rad Gel Doc™ EZ Imager .

Tryptic Digest and Mass Spectrometry

Protein identification for MBP-M1HOT and 6×His-M2HOT was performed by tryptic digest of gel-purified bands, followed by mass spectrometry (MS) and data analysis with a proteome database. The UC Davis Proteomics Core Facility was commissioned to perform the digest, mass spectrometry and analysis. Using a sterile blade, MBP-M1HOT and 6×His-M2HOT protein bands were excised from a freshly run SDS-PAGE gel that had been

Coomassie-stained. Gel slices were placed into 1.5 mL Eppendorf tubes, along with a small amount of liquid (Destain 2), sealed with Parafilm M and shipped to the UC Davis Proteomics Core Facility. At the facility the proteins were reduced, alkylated, digested with in-house reductively methylated trypsin, and dried down in a vacuum concentrator. The material was then solubilized in 2% acetonitrile and 1% trifluoracetic acid and loaded onto a liquid chromatography tandem mass spectrometer (LCMS). Following high performance liquid chromatography (HPLC), peptides were further separated using trapped ion mobility spectrometry (TIMS) and the MS data was acquired using data-dependent parallel accumulation serial fragmentation (DDA-PASEF). Data analysis involved using Fragpipe 18.0 to search the MS raw files using the UniProt proteome database and search results were loaded into the Scaffold 5.0.1 MS/MS proteomics software.

3.5 References

(1) Panopoulos, A. The Role of a Hydroxyacid-Oxoacid Transhydrogenase in GHB Metabolism. University of Strathclyde, Glasgow, Scotland, 2007.

(2) Wilson, J. M.; Carbonaro, R. F. Capillary electrophoresis study of iron(II) and iron(III) polyaminocarboxylate complex speciation. *Environmental Chemistry* **2011**, *8* (3), 295-303. DOI: 10.1071/EN11017

(3) Yamaguchi, H.; Miyazaki, M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **2014**, *4* (1), 235-251. DOI: 10.3390/biom4010235

(4) Smith, D. B.; Johnson, K. S. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **1988**, *67* (1), 31-40. DOI: 10.1016/0378-1119(88)90005-4

(5) Fox, J. D.; Waugh, D. S. Maltose-binding protein as a solubility enhancer. *Methods Mol Biol* **2003**, *205*, 99-117. DOI: 10.1385/1-59259-301-1:99

(6) Butt, T. R.; Edavettal, S. C.; Hall, J. P.; Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* **2005**, *43* (1), 1-9. DOI: 10.1016/j.pep.2005.03.016

(7) Fahnert, B.; Lilie, H.; Neubauer, P. Inclusion bodies: formation and utilisation. *Adv Biochem Eng Biotechnol* **2004**, *89*, 93-142. DOI: 10.1007/b93995

(8) Kane, J. F.; Hartley, D. L. Formation of recombinant protein inclusion bodies in Escherichia coli. *Trends in Biotechnology* **1988**, *6* (5), 95-101. DOI: 10.1016/0167-7799(88)90065-0

(9) Singh, S. M.; Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* **2005**, *99* (4), 303-310. DOI: 10.1263/jbb.99.303

(10) Gupta, S. K.; Shukla, P. Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. *Crit Rev Biotechnol* **2016**, *36* (6), 1089-1098. DOI: 10.3109/07388551.2015.1084264

(11) İncir, İ.; Kaplan, Ö. Escherichia coli as a versatile cell factory: Advances and challenges in recombinant protein production. *Protein Expr Purif* **2024**, *219*, 106463. DOI: 10.1016/j.pep.2024.106463

(12) Khow, O.; Suntrarachun, S. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pac J Trop Biomed* **2012**, *2* (2), 159-162. DOI: 10.1016/s2221-1691(11)60213-x

(13) Rosano, G. L.; Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014**, *5*, 172. DOI: 10.3389/fmicb.2014.00172

(14) Sahdev, S.; Khattar, S. K.; Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **2008**, *307* (1-2), 249-264. DOI: 10.1007/s11010-007-9603-6

(15) Sørensen, H. P.; Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb Cell Fact* **2005**, *4* (1), 1. DOI: 10.1186/1475-2859-4-1

(16) Ki, M. R.; Pack, S. P. Fusion tags to enhance heterologous protein expression. *Appl Microbiol Biotechnol* **2020**, *104* (6), 2411-2425. DOI: 10.1007/s00253-020-10402-8

(17) Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. Comparison of affinity tags for protein purification. *Protein Expr Purif* **2005**, *41* (1), 98-105. DOI: 10.1016/j.pep.2005.01.019

(18) Kosobokova, E. N.; Skrypnik, K. A.; Kosorukov, V. S. Overview of fusion tags for recombinant proteins. *Biochemistry (Moscow)* **2016**, *81* (3), 187-200. DOI: 10.1134/S0006297916030019

(19) Nallamsetty, S.; Austin, B. P.; Penrose, K. J.; Waugh, D. S. Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. *Protein Sci* **2005**, *14* (12), 2964-2971. DOI: 10.1110/ps.051718605

(20) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **1982**, *157* (1), 105-132. DOI: 10.1016/0022-2836(82)90515-0

(21) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker, J. M. Ed.; Humana Press, 2005; pp 571-607.

(22) Duvaud, S.; Gabella, C.; Lisacek, F.; Stockinger, H.; Ioannidis, V.; Durinx, C. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Res* **2021**, *49* (W1), W216-w227. DOI: 10.1093/nar/gkab225

(23) Fox, J. D.; Kapust, R. B.; Waugh, D. S. Single amino acid substitutions on the surface of Escherichia coli maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci* **2001**, *10* (3), 622-630. DOI: 10.1110/ps.45201

(24) Kardon, T.; Noël, G.; Vertommen, D.; Schaftingen, E. V. Identification of the gene encoding hydroxyacid-oxoacid transhydrogenase, an enzyme that metabolizes 4-hydroxybutyrate. *FEBS Lett* **2006**, *580* (9), 2347-2350. DOI: 10.1016/j.febslet.2006.02.082

(25) Kim, J. Y.; Tillison, K. S.; Zhou, S.; Lee, J. H.; Smas, C. M. Differentiation-dependent expression of Adhfe1 in adipogenesis. *Arch Biochem Biophys* **2007**, *464* (1), 100-111. DOI: 10.1016/j.abb.2007.04.018

(26) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023**, *51* (D1), D523-d531. DOI: 10.1093/nar/gkac1052

(27) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2

(28) Käll, L.; Krogh, A.; Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **2004**, *338* (5), 1027-1036. DOI: 10.1016/j.jmb.2004.03.016

(29) Käll, L.; Krogh, A.; Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W429-432. DOI: 10.1093/nar/gkm256

(30) Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **2024**, *52* (D1), D368-d375. DOI: 10.1093/nar/gkad1011

(31) Chacinska, A.; Koehler, C. M.; Milenkovic, D.; Lithgow, T.; Pfanner, N. Importing mitochondrial proteins: machineries and mechanisms. *Cell* **2009**, *138* (4), 628-644. DOI: 10.1016/j.cell.2009.08.005

(32) Wiedemann, N.; Pfanner, N. Mitochondrial Machineries for Protein Import and Assembly. *Annu Rev Biochem* **2017**, *86*, 685-714. DOI: 10.1146/annurev-biochem-060815-014352

(33) Kunová, N.; Havalová, H.; Ondrovičová, G.; Stojkovičová, B.; Bauer, J. A.; Bauerová-Hlinková, V.; Pevala, V.; Kutejová, E. Mitochondrial Processing Peptidases-Structure, Function and the Role in Human Diseases. *Int J Mol Sci* **2022**, *23* (3). DOI: 10.3390/ijms23031297

(34) Schein, C. H. Production of Soluble Recombinant Proteins in Bacteria. *Bio/Technology* **1989**, *7* (11), 1141-1149. DOI: 10.1038/nbt1189-1141

(35) Kiefhaber, T.; Rudolph, R.; Kohler, H. H.; Buchner, J. Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology (N Y)* **1991**, *9* (9), 825-829. DOI: 10.1038/nbt0991-825

(36) Fatima, K.; Naqvi, F.; Younas, H. A Review: Molecular Chaperone-mediated Folding, Unfolding and Disaggregation of Expressed Recombinant Proteins. *Cell Biochem Biophys* **2021**, *79* (2), 153-174. DOI: 10.1007/s12013-021-00970-5

(37) Georgiou, G.; Valax, P. Expression of correctly folded proteins in Escherichia coli. *Curr Opin Biotechnol* **1996**, *7* (2), 190-197. DOI: 10.1016/s0958-1669(96)80012-7

(38) Thomas, J. G.; Ayling, A.; Baneyx, F. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. *Appl Biochem Biotechnol* **1997**, *66* (3), 197-238. DOI: 10.1007/bf02785589

(39) Nishihara, K.; Kanemori, M.; Kitagawa, M.; Yanagi, H.; Yura, T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. *Appl Environ Microbiol* **1998**, *64* (5), 1694-1699. DOI: 10.1128/aem.64.5.1694-1699.1998

(40) Nishihara, K.; Kanemori, M.; Yanagi, H.; Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl Environ Microbiol* **2000**, *66* (3), 884-889. DOI: 10.1128/aem.66.3.884-889.2000

(41) Grathwohl, C.; Wüthrich, K. Nmr studies of the rates of proline cis–trans isomerization in oligopeptides. *Biopolymers* **1981**, *20* (12), 2623-2633. DOI: 10.1002/bip.1981.360201209

(42) Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. Proline cis-trans isomerization and protein folding. *Biochemistry* **2002**, *41* (50), 14637-14644. DOI: 10.1021/bi020574b

(43) Hanes, S. D. Prolyl isomerases in gene transcription. *Biochim Biophys Acta* **2015**, *1850* (10), 2017-2034. DOI: 10.1016/j.bbagen.2014.10.028

(44) Hoffmann, A.; Bukau, B.; Kramer, G. Structure and function of the molecular chaperone Trigger Factor. *Biochim Biophys Acta* **2010**, *1803* (6), 650-661. DOI: 10.1016/j.bbamcr.2010.01.017

(45) Saio, T.; Guan, X.; Rossi, P.; Economou, A.; Kalodimos, C. G. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **2014**, *344* (6184), 1250494. DOI: 10.1126/science.1250494

(46) Ferrè, F.; Clote, P. Disulfide connectivity prediction using secondary structure information and diresidue frequencies. *Bioinformatics* **2005**, *21* (10), 2336-2346. DOI: 10.1093/bioinformatics/bti328

(47) Ferrè, F.; Clote, P. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res* **2005**, *33* (Web Server issue), W230-232. DOI: 10.1093/nar/gki412

(48) Ferrè, F.; Clote, P. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res* **2006**, *34* (Web Server issue), W182-185. DOI: 10.1093/nar/gkl189

(49) Hu, J.; Dong, L.; Outten, C. E. The redox environment in the mitochondrial intermembrane space is maintained separately from the cytosol and matrix. *J Biol Chem* **2008**, *283* (43), 29126-29134. DOI: 10.1074/jbc.M803028200

(50) Ren, G.; Ke, N.; Berkmen, M. Use of the SHuffle Strains in Production of Proteins. *Curr Protoc Protein Sci* **2016**, *85*, 5.26.21-25.26.21. DOI: 10.1002/cpps.11

(51) Bolanos-Garcia, V. M.; Davies, O. R. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta* **2006**, *1760* (9), 1304-1313. DOI: 10.1016/j.bbagen.2006.03.027

(52) Robichon, C.; Luo, J.; Causey, T. B.; Benner, J. S.; Samuelson, J. C. Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol* **2011**, *77* (13), 4634-4646. DOI: 10.1128/aem.00119-11

(53) Linke, K.; Wolfram, T.; Bussemer, J.; Jakob, U. The roles of the two zinc binding sites in DnaJ. *J Biol Chem* **2003**, *278* (45), 44457-44466. DOI: 10.1074/jbc.M307491200

(54) Lukács, M.; Csilla Pálinkás, D.; Szunyog, G.; Várnagy, K. Metal Binding Ability of Small Peptides Containing Cysteine Residues. *ChemistryOpen* **2021**, *10* (4), 451-463. DOI: 10.1002/open.202000304

(55) Sudan, R. J.; Kumari, J. L.; Sudandiradoss, C. Ab initio coordination chemistry for nickel chelation motifs. *PLoS One* **2015**, *10* (5), e0126787. DOI: 10.1371/journal.pone.0126787

(56) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **1976**, *72*, 248-254. DOI: 10.1006/abio.1976.9999

(57) Compton, S. J.; Jones, C. G. Mechanism of dye response and interference in the Bradford protein assay. *Anal Biochem* **1985**, *151* (2), 369-374. DOI: 10.1016/0003-2697(85)90190-3

(58) Friedenauer, S.; Berlet, H. H. Sensitivity and variability of the Bradford protein assay in the presence of detergents. *Anal Biochem* **1989**, *178* (2), 263-268. DOI: 10.1016/0003-2697(89)90636-2

(59) Warburg, O.; Christian, W. Isolierung und Kristallisation des Gärungsferments Enolase. *Naturwissenschaften* **1941**, *29* (39), 589-590. DOI: 10.1007/BF01482279

Chapter 4

The Expression of Recombinant HOT in PC-12 Cells

Abstract

A vector encoding recombinant mouse hydroxyacid-oxoacid transhydrogenase (HOT) and the neomycin resistance gene were transfected into PC-12 (rat adrenal medulla) cells with the use of cationic lipids. Stable transfectants expressing recombinant HOT were isolated under the selective pressure of Geneticin (G-418), since the neomycin resistance gene would be integrated into the host genome along with the mouse HOT cDNA sequence. The recombinant mouse HOT contained a C-terminal HA-tag that allowed for detection of recombinant HOT by Western blotting and could serve as an affinity tag in immunopurification. Despite only a 1.2 kDa difference in molecular weight between *recombinant* HOT and *endogenous* HOT, size exclusion chromatography (SEC) successfully partitioned these proteins into separate column fractions with the apparent molecular weights of 65 kDa and 41 kDa, respectively. When fractions from the SEC purification were subsequently passed through an anion exchange column (AEC), endogenous HOT unexpectedly eluted with the flow though—possibly aggregated—while recombinant HOT eluted at the expected NaCl concentration. When the order of the chromatography steps is reversed and lysate is purified by AEC first, both recombinant and endogenous HOT elute close together at the expected NaCl concentration. In addition, the antibody raised against the full HOT sequence does not bind recombinant HOT with the HA-tag. It appears that the HA-tag at the C-terminus of recombinant HOT may decrease interactions with the SEC resin allowing it to run ahead of endogenous HOT, inhibits aggregation under AEC low-salt loading conditions, and blocks binding of the HOT antibody.

4.1 Introduction

4.1.1 HOT in Mammalian Cells

To date, the only source of enzymatically active hydroxyacid-oxoacid transhydrogenase (HOT) is that produced by animal cells. This includes both endogenously produced HOT, which has been extracted from the cultured tissues and organs of several animal species[1-8], and recombinant HOT, which has been introduced and expressed in mammalian cells that were cultivated in the laboratory[1, 9].

The literature describes only one previous attempt to express human HOT in *E. coli*, with the majority of the HOT expressed as inclusion bodies and, following solubilization with urea, efforts to refold and recover active HOT by dialysis failed.[10] Similar challenges expressing HOT in *E. coli* were encountered during the research presented in this manuscript (Chapter 3). While expression of recombinant proteins is routinely performed in prokaryotic hosts such as *E. coli*, these organisms often lack the proper cellular environment and enzymes required to produce structurally mature and/or functional active eukaryotic proteins. Many eukaryotic proteins depend on covalent *post-translational modifications* (PTMs), including glycosylation and disulfide bond formation, for proper protein folding and to stabilize the right conformation.[11] It is possible that the inability of *E. coli* to express soluble and active HOT is due to the lack of one or more required PTMs. Based on experimental and computational evidence, the UniProt entry Q8R0N6 for mouse HOT infers *N6-acetyllysine* at position 443 (445 in human HOT)[12] and *phosphoserine* at position 450 (452 in human HOT)[13].

In addition, multiple lines of experimental and computational evidence suggest that HOT contains an N-terminal mitochondrial transit peptide (MTP). MTPs are 15 – 50 amino acids

presequences that form amphipathic α-helices and help direct *preproteins* translated in the cytosol to the translocases in the outer and inner mitochondrial membranes.[14, 15] Once across the mitochondrial membranes, these presequences are cleaved from the N-terminus by peptidases and the resulting preproteins are folded into their mature and active forms.[16]

The experimental evidence for the N-terminal MTP is based on the fact that HOT has two in-frame ATG start codons preceded by Kozak consensus sequences.[17, 18] This suggests the existence of *two isoforms* that differ by 46 amino acids in mouse[19], and 48 amino acids in humans[20]. In one study with rat liver, HOT activity coeluted with 50 kDa and 45 bands on SDS-PAGE during size exclusion chromatography.[1] In a separate study the following year, transfection of mammalian cells with cDNA beginning with the first ATG resulted in evidence of 50 kDa and 45 kDa expression products, while transfection with cDNA that began at the second ATG resulted in only the 45 kDa product.[9] The following section details these experimental finding in more detail.

Beside the experimental evidence, amino acid sequence analysis by Phobius[21, 22] and UniProt[23] predict a transit peptide at the N-terminus for both mouse (Q8R0N6) and human (Q8IWW8) HOT (see *Section 1.5.1*). Also, protein structure prediction by AlphaFold[24, 25] generated a structure for HOT that contained an amphipathic α-helix spanning the first 19 amino acids of the sequence, followed by a random coil for another 30 residues. The random coil ends at the second methionine in the amino acid sequence, which encoded by the second in-frame ATG and start of HOT isoform 2 (see *Section 1.5.1*). Given the absence of cellular machinery required to process proteins with MTPs, the heterologous expression of proteins with MTPs in a hosts such as *E. coli* may prove challenging. Indeed, over the past 3 decades a

general strategy when expressing mitochondrial matrix enzymes in *E. coli* has been to remove the MTPs altogether, via truncation of the cDNA.[26, 27]

4.1.2 Literature on Recombinant HOT in Mammalian Cells

There have been only two publications documenting the introduction of *recombinant* HOT into mammalian cells. In one publication, HOT *enzymatic activity* was induced in human embryonic kidney (HEK 293) cells following *transient transfection* with a vector containing the putative cDNA sequence for mouse HOT.[1] HOT enzyme activity was detected only after cell lysate had been partially purified by anion exchange chromatography, with the assays for the forward and reverse reactions providing K_m estimates for all four canonical HOT substrates.[1] Tryptic digest and mass spectrometry of 50 kDa and 45 kDa SDS-PAGE bands that co-eluted with enzyme activity confirmed their identities as HOT, revealing the existence of two isoforms.[1]

In the second publication, COS cells were *transiently transfected* with a recombinant mouse HOT containing a C-terminal HA-tag, in order to determine HOT subcellular localization with *immunocytochemical staining* and to produce expression products for *Western blot analysis*.[9] Two mouse HOT coding sequences were used, one beginning at the first ATG (M1) and including the purported N-terminal MTP, and the other beginning after the MTP at the second ATG (M2). Immunostaining of cells transfected with M1HOT-HA cDNA sequence indicated that HOT was localized to mitochondria.[9] Western blots from transiently transfected COS cells, as well as *in vitro* transcription and translation, demonstrated that M1 sequence can produce both the 50 kDa and 45 kDa isoforms.[9] Although the expression levels of the 45 kDa and 50 kDa isoforms are comparable in mammalian cells, the 45 kDa isoform is a very minor product of *in vitro* transcription and translation.[9] On SDS-PAGE, the

45 kDa expression products from M1 and M2 sequences migrate identically.[9] This suggests that most of the 45 kDa isoform results from post-translational removal of the MTP in mammalian cells, while with *in vitro* translation a small amount of it is from "leaky scanning"[28] of M1 mRNA by ribosomes.

To date, following objectives have not been documented in the literature:

1. Establishment of a mammalian cell line stably expressing high levels of recombinant HOT.
2. Purification of recombinant HOT to homogeneity.
3. Complete multisubstrate enzyme kinetics for recombinant HOT, including determination of k_{cat}, the kinetic mechanism, and the pH-rate profile.

4.1.3 Present Research on Recombinant HOT in Mammalian Cells

Determining the complete enzyme kinetics for HOT requires a pure sample of the enzyme. This is best accomplished with a recombinant HOT containing an affinity tag. Recombinant HOT also has the added benefit that point mutations can be easily introduced in order to determine which active site residues are critical for activity. Because heterologous expression of recombinant HOT in *E. coli* has been largely unsuccessful[10], while transient transfection of recombinant HOT into mammalian cells resulted in the recovery of active enzyme[1], stable transfection in a mammalian cell line is a natural progression of this strategy and far more efficient approach in terms of protein expression.

The aims of the present research were to:

1. Establish a mammalian cell line that stably expresses recombinant HOT.
2. Obtain pure recombinant HOT for enzyme kinetics studies.

To accomplish these aims a *shuttle vector* containing the full (M1) mouse HOT coding sequence, along with a C-terminal HA-tag and the *neomycin resistance* gene, was used to stably transfect PC-12 cells under the selective pressure of the antibiotic Geneticin. Lysate from PC-12 cells stably expressing recombinant HOT was then purified by ion exchange and size exclusion chromatography, with the purification followed by SDS-PAGE and Western blotting with antibodies against both the HA-tag and HOT sequences.

4.1.4 Mammalian Cell Culture

Cell lines and Cell Culture

Three mammalian cell lines were investigated as potential hosts for the expression of recombinant mouse HOT. These were COS-7, SH-SY5Y, and PC-12. COS-7 and SH-SY5Y

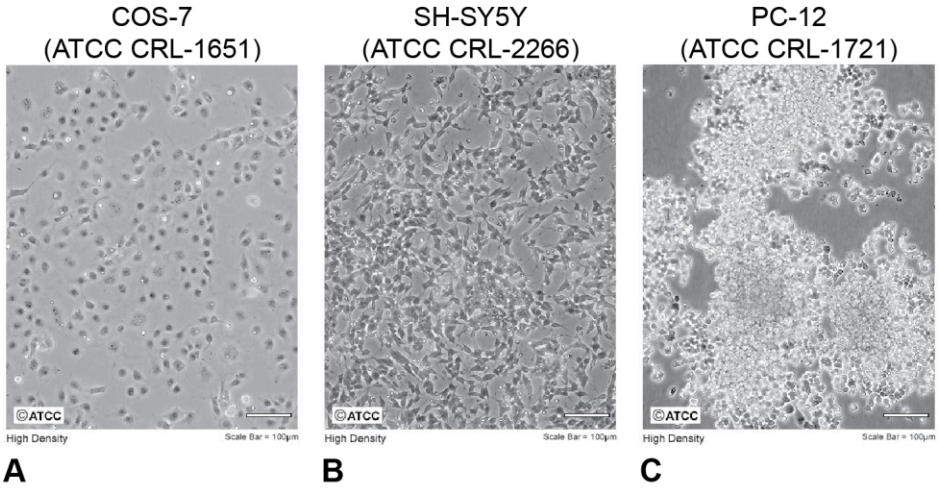

Figure 4.1. Mammalian cell lines used in recombinant protein expression. A) COS-7 from the kidney of an African green monkey; **B)** SH-SY5Y from the bone marrow of a human with neuroblastoma; **C)** PC-12 from a pheochromocytoma of the rat adrenal medulla (adapted from ATCC®).

cells were previously used to study recombinant and endogenous HOT, respectively.[7, 9] COS-7 are kidney cells from the African green monkey (*Cercopithecus aethiops*) that have been immortalized with the SV40 virus.[29] SH-SY5Y is a cancer cell line isolated from the bone marrow of a human patient with neuroblastoma.[30] PC-12 is a cancer cell line that was derived from a pheochromocytoma of the brown rat (*Rattus norvegicus*) adrenal medulla.[31] Although COS-7 and SH-SY5Y cells are adherent, PC-12 cells may grow as either floating aggregates of round cells or adapted to grow as an adherent variant.[32] Images of all three cell lines at high cell confluency (density) are shown in **Figure 4.1**.

Compared to the culturing of prokaryotic cells such as *E. coli*, the culturing of mammalian cells is significantly more expensive, time consuming, and technically demanding. Furthermore, it has been generally regarded that when it comes to target protein expression, mammalian cells give relatively poor yields.[33] Therefore, the use of mammalian cells for recombinant protein expression has been traditionally reserved for situations when the target expression in prokaryotic hosts has proven unsuccessful, or when immunological concerns and the regulations governing the production of pharmaceuticals require it.[34] However, over the past two decades engineering advancements in expression vectors, cell lines, and the use of *suspension cells* in bioreactors has made it possible to obtain g/L yields from target expression in mammalian cells.[35-37]

Shuttle Vector

Gene transfer into mammalian cells can be accomplished in one of two ways: 1) infecting the cell with a virus carrying the target gene (*transduction*); and 2) physically or chemically transferring a vector encoding the target gene into the cell (*transfection*).[38] For the research presented in this dissertation, a non-viral *shuttle vector* (pcDNA3.1) was used to transfect

mammalian cells. Shuttle vectors contain two different origins of replication and two different selection markers, enabling them to be transferred into two distinct organisms.[39] This useful feature permits straightforward manipulation of the DNA when the vector is transformed into bacteria, such as site-directed mutagenesis or amplification.

pcDNA3.1 is a 5.4 kb vector for stable and transient expression of target proteins in mammalian hosts. It contains the human cytomegalovirus (CMV) immediate-early promoter for high-level expression of the recombinant protein, bovine growth hormone polyadenylation signal for transcription termination and polyadenylation of the recombinant transcript, a neomycin resistance gene for selection of stable transfectants, and SV40 early promoter and origin for *episomal replication* in cells that express SV40 large T antigen (i.e., COS-7). Episomes are eukaryotic plasmids that replicate with the cells, associating with metaphase chromosomes, but do not integrate into the genome and are therefore not subject to gene silencing mechanisms in the nucleus.[40] While the CMV immediate-early promoter on the pcDNA3.1 vector results in strong constitutive expression of recombinant proteins in mammalian cells, inducible promoters and engineered mammalian cells have been developed that allow expression of recombinant proteins to be switched on and off, much like the *lac* operon in *E. coli*.[41] The vector also contains the T7 promoter for *in vitro* transcription and sequencing of the insert, and for manipulation in *E. coli* the vector has a pUC origin for high-copy number replication and the *bla* gene for ampicillin resistance.

Transfection

Transfection methods for mammalian cells include: 1) chemical (cationic lipid, calcium phosphate); and 2) physical (electroporation, direct injection).[42] Cationic lipids[43] (Invitrogen™ Lipofectamine 2000™) were used to transfect cultured mammalian cells in the research

presented in this manuscript. The transfection outcome is divided into two categories, *transient* and *stable*, which refer to the length of time the transgene is transcribed and translated.[44] In transient transfection, the transgene is expressed for only a few days after which it is diluted and lost during cell division. In stable transfection the vector encoding the transgene either exists as an episome (see above) or is integrated into the host genome. In most cases, this integration is a rare event that occurs with approximately 1 in 10,000 cells.[45] The integration mechanism is poorly understood but thought to involve a non-homologous end-repair mechanism.[46] Because of the random nature of the integration, the transgene may be subject to silencing if the insertion occurs in condensed heterochromatin or is epigenetically silenced following insertion.[47] Alternatively, transfected DNA can also be integrated at predetermined chromosomal locations with the use of a recombinase, such as FLP from yeast[48], or through CRISPER/Cas9 gene editing[49, 50]. Following transfection, a selectable marker on the vector, with the phenotype generally being antibiotic resistance, is used to select for stable transfectants.[38]

Selection

A number of agents may be used to select for cells that have undergone a stable transfection.[51] The pcDNA3.1 vector used in the present research includes *neo* gene under control of the SV40 promoter.[52, 53] The *neo* gene encodes a bacterial aminoglycoside-3'-phosphotransferase[54] (APH), also known as aminoglycoside kinase, which detoxifies the antibiotic Geneticin® (G418). G418 inhibits the 80S ribosome elongation phase of protein synthesis in mammalian cells.[55] Cells that stably express transgenic sequences in the vector, either as an episome or though integration into the genome, become resistant to effects of G418 and will continue to grow in its presence. With all antibiotics used for selection, a kill curve

should be generated with untransfected cells in order to determine optimal concentration for selection and maintenance of stable cell lines.[35]

Even following selection, the results of a stable transfection can be unpredictable over time due to *genetic drift* and for this reason the *passage number* for a stable line should be kept as low as possible. The passage number increases each time cells are transferred from one flask to another and is one method for tracking the "age" of a cell line. In addition, cellular responses and alterations (i.e., epigenetic, transcriptomic, metabolomic, etc.) may arise due to transgene overexpression, antibiotic selection and other experimental manipulations, impacting the expected phenotype in unforeseen ways.[56]

4.2 Results and Discussion

4.2.1 Mammalian Cell Culture

COS-7 versus PC-12 Cells for Protein Expression

The COS-7 cell line was initially chosen for *stable transfection* because the laboratory that generously donated the pcDNA3.1 vector encoding recombinant HOT published the successful *transient transfection* of this vector in COS cells.[9] Cells were transfected with a mixture of the expression vector and cationic lipids, and the transfectants were subjected to selection with the aminoglycoside antibiotic Geneticin (G418). Stable transfection of COS-7 cells was identified in several surviving colonies by Western blot detection of the C-terminal HA-tag of the recombinant HOT protein using anti-HA antibodies. While scaling up the culturing of COS-7 cells that stably expressed the recombinant HOT-HA, two observations were made about the cell line: 1) the cells grow very slowly, dividing once every two days; and 2) the cell density at 100% confluency is low, with an average of 30 million cells per T-175 flask (175 cm^2 surface for adherent cells). Similar observations were made when culturing the SH-SY5Y cell line, which has an apparent doubling time of 2–3 days and establishes thin neurite-like processes that prevent high cell densities at 100% confluency. COS-7 and SH-SY5Y complete growth media also contains 10% fetal bovine serum (FBS). FBS is extremely expensive, with a $500 - $1000/liter price (pre-2020) that fluctuates with beef prices and can sometimes be difficult to acquire.

Given the time, yield, a cost to grow COS-7 and SH-SY5Y cells, the decision was made to switch to the PC-12 cell line, which at the time were being actively grown in the laboratory for a separate project. PC-12 cells can divide in less than 36 hours and have a polygonal

morphology that allows them to grow to a very high cell density, sometimes growing on top of each other. Up to 90 million cells can be recovered from a single T-175 flask, versus 30 million for COS-7. Lastly, PC-12 complete growth media contains only 5% FBS and 10% donor horse serum (HS). The HS is only around 1/5th the cost of FBS and, because of how it is produced, it is generally always in stock. Therefore, the remainder of the research presented here focuses solely on the expression of recombinant HOT in PC-12 cells.

PC-12 Cell Viability Study with G418

The optimal concentration of G418 required for selection and maintenance of stable transfectants is dependent on the cell line, therefore it was necessary to perform a cell viability study on untransformed PC-12 cells prior to transfection. For *selection*, the optimal concentration of G418 is the minimum dose that results in significant death of *untransfected* cells within 3-7 days. For the viability study, untransfected PC-12 cells were incubated for 12 days in media supplemented with 0–1,500 μg/mL G418 and cell survival was assessed every 3 days.

The results of the viability study are plotted as a kill curve in **Figure 4.2**. By day 3, only 5.2 ± 6.6 % of cells exposed to 500 μg/mL G418 survived, therefore this concentration of G418 was used for the selection of stable transfectants. For *maintenance* of the selected cell line, the general practice is to reduce the G418 concentration to ≤ 50% of the concentration used for selection, which in this case was 250 μg/mL G418. By days 3 and 6, the percent survival for untransfected PC-12 cells in 250 μg/mL G418 was 10.0 ± 3.3% and 6.3 ± 0.7%, respectively. This concentration was deemed sufficient to maintain the population of PC-12 cells that stably express the HOT genes and the resistance-conferring aminoglycoside 3′-phosphotransferase

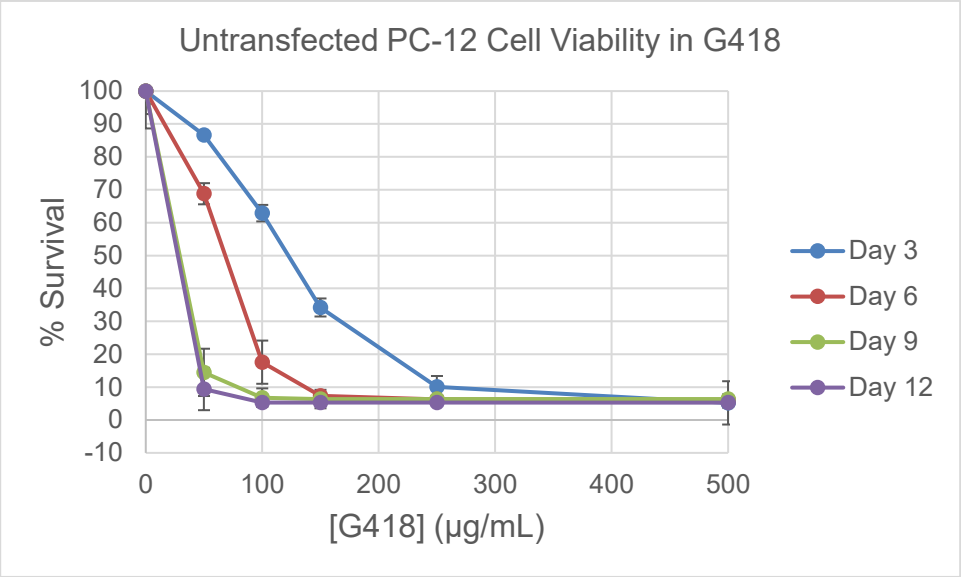

Figure 4.2. G418 kill curve for untransfected PC-12 cells. The results of PC-12 cells incubated with 0, 50, 100, 150, 250, 500, 750, 1000, and 1500 µg/mL G418 for a total of 12 days, six replicates each. The percent survival was based on the control wells, with no G418 (100% survival). The data for 750, 1000, and 1500 µg/mL G418 is not shown because no cells survived in these concentrations for any length of time

(*neo*) genes, assuming that both integrated into the host genome without disruptions or silencing of the HOT gene.

Stable Transfection of PC-12 Cells

PC-12 cells were transfected with the pcDNA3.1 vector encoding M1HOT-HA using Lipofectamine 2000 transfection reagent[57] and the manufacturer's protocol for 6-well plate transfections with 6, 9, 12 and 15 µL of transfection reagent used in 4 wells, plus controls. Following transfection, cells from each well were transferred to 10-cm tissue-culture treated dishes and diluted 1:25 in media. Eighteen distinct colonies were recovered after 6 weeks of selection with 500 µg/mL G418, with each colony originating from a single cell that had successfully acquired a neomycin resistance gene during transfection. Of the 18 recovered colonies, 15 were the result of cells transfected with 9 µL of Lipofectamine 2000, while the other 3 colonies came from cells that were transfected with 12µL of the reagent. No G418-resistant colonies were recovered from transfections with 6 and 15 µL Lipofectamine 2000.

To verify whether PC-12 colonies that acquired neomycin resistance had also integrated the intact HOT-HA coding sequence, cell extract from each colony was analyzed by Western blotting. These blots were visualized by staining with a primary antibody that recognizes the C-terminal HA-tag, followed by a fluorescent dye-conjugated secondary antibody that provided an observable and amplified signal (**Figure 4.3**). Because each sample well on the blot was loaded with 25 µg total protein, the intensity of the fluorescent staining was proportional to the level of HOT-HA expression. The lysate from three colonies (8, 9, and 15) appeared to contain high expression levels of HA-tagged HOT isoform 2, which lacks the N-terminal mitochondrial transit peptide and has a molecular weight of 46 kDa. The Western blots showed no evidence of HA-tagged HOT isoform 1, which contains the N-terminal mitochondrial transit peptide and has a molecular weight of 51 kDa. These observations are consistent with observations that HOT is localized to the mitochondria, where the N-terminal transit peptide is likely removed after being imported into the organelle. Colonies 8, 9, and 15 were cryogenically preserved in liquid nitrogen for further study.

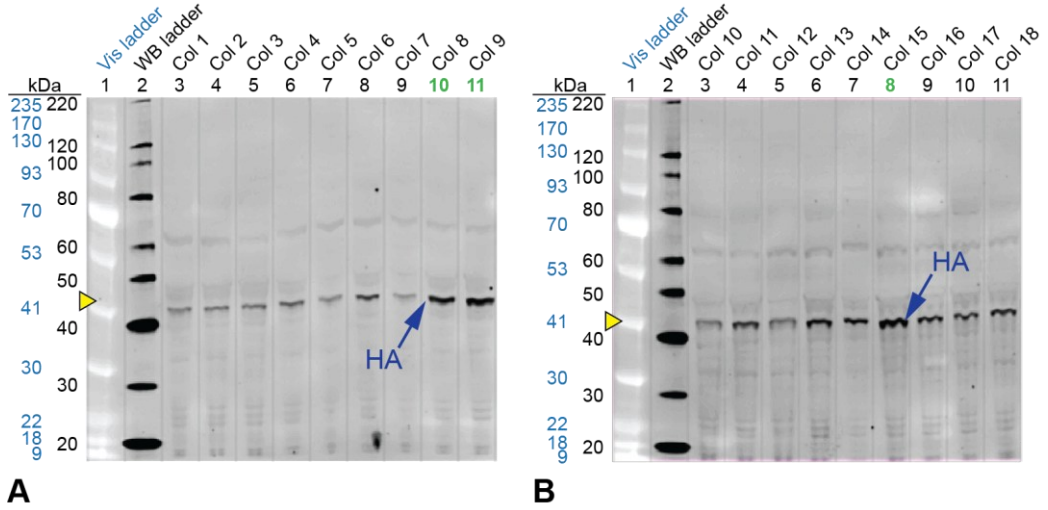

Figure 4.3. Western blots of total protein from stably transfected PC-12 colonies. Fluorescent imaging of blots probed with anti-HA primary antibody and fluorescent-conjugated secondary antibody. Sample lanes were loaded with 25 µg total protein. Lanes 1 and 2 on each blot contain visible (**light blue**) and Western blot (**black**) ladders, respectively. Well numbers in **green** indicate strongly expressing colonies that were cryopreserved. The **yellow** triangle indicates the molecular weight of bands pointed out by the **indigo** arrows. The molecular weight of HOT without the mitochondrial transit peptide is 46 kDa. **(A)** Colonies 1–9; and **(B)** colonies 10–18.

While all of the recovered PC-12 colonies survived G418 selection, implying they all expressed the neomycin resistance gene, Western blots indicated that many of the colonies displayed relatively low expression of recombinant HOT with respect to colonies 8, 9 and 15. Because genomic integration of the pcDNA3.1 vector DNA is a random event, it is possible that some elements required for expression of the target gene might have become disrupted during integration, or the expression construct was integrated into a region of the genome with low transcriptional activity. Furthermore, introduction of the vector and transfection reagent, the expression of the neomycin resistance gene, and exposure of the cells to G418 during selection all have the potential to trigger cellular responses that promote genomic instability and epigenetic remodeling. Such changes to the genomic landscape could impact the potential expression level of the desired protein.

4.2.2 Protein Extraction and Expression Efficiency

HOT has been extracted from animal tissues and cells using both liquid homogenization (i.e. Potter-Elvehjem homogenizer) and repeated freeze-thaw cycles without loss of enzyme activity.(refs) Both methods were evaluated for their ability to recover total protein from PC-12 cells using the same lysis buffer. Liquid homogenization with a Potter-Elvehjem homogenizer yielded significantly more protein (66.38 ± 2.84 mg per billion cells, versus 50.29 ± 3.78 mg per billion cells) than lysis by freeze-thaw.

Using the average of these extraction methods, 100 mg of total protein would require approximately 1,700 million cells (around 2.0 g wet cell pellet). The transfected PC-12 cells used in this study grew to a cell density of approximately 525,000 cells/cm^2 on T-175 cell culture flasks (TC-coated surface of 175 cm^2) for a maximum yield of 90,000,000 cells per flask if grown to 100% confluency. Acquiring 1,700 million cells would require approximately

20 × T-175 flasks, plus a number of T-25 and T-75 flasks for propagating the cells used for seeding. Each T-175 flask should have 50 mL of media supplemented with serum to support the growth of the cells, not counting the media needed to grow the cells required to seed the flasks and the media required to quench the trypsin during harvesting. Expect a requirement of at least 2 liters of PC-12 media, which is D-MEM/F-12 supplemented with 5% donor horse serum, 10% fetal bovine serum, and 250 µg/mL G418. The media costs $100/liter to make, the plastic consumables (flasks, serological pipettes, etc.) are another $50. There are also requirements for specialized equipment, such as a biosafety cabinet with laminar air flow, a temperature and humidity-controlled CO_2 incubator, and liquid nitrogen storage dewars for cryogenic preservation of cell lines. Assuming a week to grow enough cells to seed an incubator with 20 × T-175 flasks, it would take another 3-4 days until the newly seeded cells reach confluency, and then it takes 4 hours just to harvest and wash the cells from 20 × T-175 flasks. This is a lot of work for 100 mg of protein.

Compare this time, effort and cost to the expression of recombinant SSADH in *E. coli* (*Chapter 2*), where 0.3 liters of inexpensive 2YT broth and 24 hours yielded a 3.66 g cell pellet containing 164 mg of total protein, 18 mg of which was pure recombinant SSADH. The contrast could not be clearer. The culturing of mammalian cells for the purposes of recombinant protein expression is a completely inefficient process as performed for this research, particularly when considering that the target protein probably amounts to no more than 0.1 mg of the 100 mg protein expression. Still, if a eukaryotic protein cannot be expressed correctly in a prokaryotic host, this may be the only alternative. While this particular route may not be feasible for structural studies that require milligrams of protein, 0.1 mg of pure enzyme may provide enough material for enzyme kinetics experiments.

4.2.3 Protein Purification Strategy I: AEC Before SEC

Anion Exchange Chromatography as the 1st Step

Anion exchange chromatography was performed on the clarified lysate from transfected PC-12 cells that stably express recombinant M1HOT-HA, along with endogenously produced wild-type HOT. The lysate was obtained by freeze-thaw cycles on 1 gram of cells (800 million) which resulted in 2 mL of lysate with a total protein concentration of 45.33 mg/mL (90.66 mg of total protein). The resin was DEAE-Sepharose Fast Flow and the mobile phase for the separation was 20 mM Tris pH 7.6, with a NaCl gradient of 0–500 mM. The results of the separation were followed by Western blotting using rabbit primary antibodies against both the HA-tag and HOT protein sequence, followed by goat anti-rabbit fluorescent-conjugated secondary antibodies and fluorescence imaging. To determine the quality of the separation, the column fractions were visualized on Coomassie-stained SDS-PAGE gel.

Based on the Western blots stained with the anti-HA antibody (**Figure 4.4**), recombinant HOT begins to elute in fraction G20 (167 mM NaCl), has a peak elution at G23 (243 mM NaCl) and disappears from the column fractions by G25 (301 mM NaCl). The Western blots also reveal *a significant amount of non-specific binding* by either (or both) the primary and secondary antibodies to other protein bands throughout the blot, many of which are significantly larger or smaller than the expected molecular weight for recombinant HOT-HA.

Nonspecific binding of antibodies became a serious issue shortly after this project began when our primary supplier of antibodies, Santa Cruz Biotechnology (SCB), ran afoul of the USDA due to questionable practices in antibody production.[58] SCB anti-HA antibodies were initially used to estimate the expression level of the recombinant HA-tag in PC-12 stable

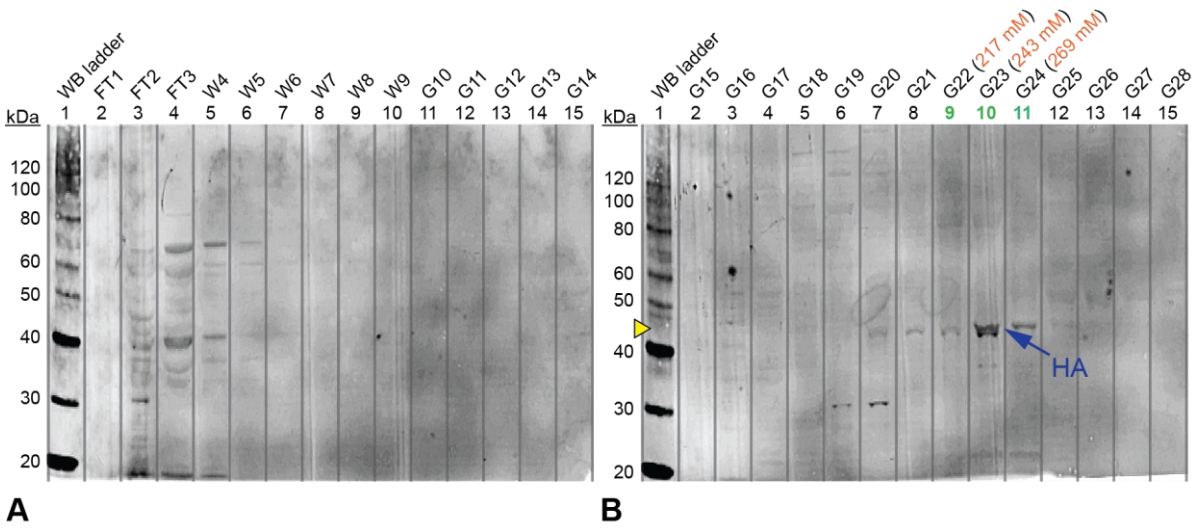

Figure 4.4. Western blots of AEC column fractions with HA 1º antibody. The blots were stained with a primary antibody against the HA-tag sequence. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **blue** arrows. Well numbers in **green** indicate fractions that were collected and advanced to the next purification step. The NaCl concentration for select fractions appears in **orange** after the fraction identifiers (above the well numbers). FT = *flow-through*, W = *wash*, G = *salt gradient*. **(A)** Anion exchange *flow-through* fraction 1 through *salt gradient* fraction 14; **(B)** *salt gradient* fractions 15–28.

transfectants (**Figure 4.3**) with very good results. After the SBC catalog became unavailable, sourcing quality antibodies targeting the HA-tag and the HOT protein sequence became challenging, with most antibodies showing either significant nonspecific binding or worse, they did not bind to any bands reasonably close the expected 40–50 kDa range for recombinant or endogenous HOT.

Based on the protein sequence, the expected molecular weight for recombinant HOT-HA with the N-terminal mitochondrial transit peptide (isoform 1; M1HOT-HA) is 51.1 kDa, while recombinant HOT-HA without the transit peptide (isoform 2; M47HOT-HA) is 46.2 kDa. It would appear that the recombinant HOT-HA observed on the Western blot is isoform 2. The stained band in fraction G23 (lane 10) appears to be two bands running very close together.

The Western blot stained with the anti-HOT antibody (**Figure 4.5**) tells a similar story, but this time with the elution beginning two fractions later, at G22, with peak staining at G23 and disappearance of staining by G25. The expected molecular weights for endogenous HOT

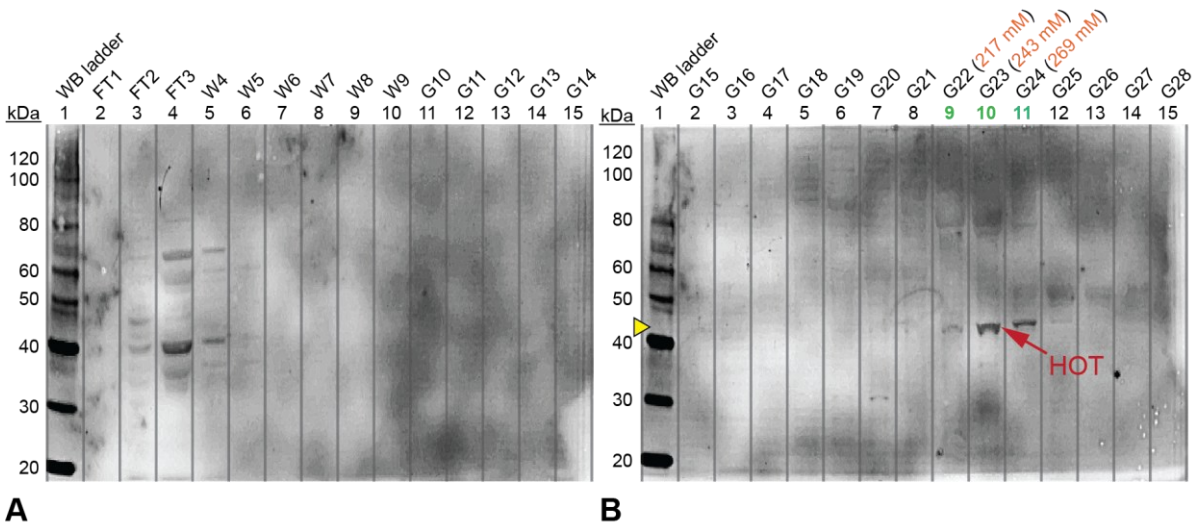

Figure 4.5. Western blots of AEC column fractions with HOT 1º antibody. The blots were stained with a primary antibody against the HOT protein sequence. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **red** arrows. Well numbers in **green** indicate fractions that were collected and advanced to the next purification step. The NaCl concentration for select fractions appears in **orange** after the fraction identifiers (above the well numbers). FT = *flow-through*, W = *wash*, G = *salt gradient*. **(A)** Anion exchange *flow-through* fraction 1 through *salt gradient* fraction 14; **(B)** *salt gradient* fractions 15–28.

are 49.9 kDa with the N- terminal mitochondrial transit peptide (isoform 1; M1HOT) and 45.0 kDa without the transit peptide (isoform 2; M47HOT). As with the anti-HA blot, the apparent molecular weight seems to indicate only the presence of HOT isoform 2.

One critical observation arises when comparing the staining patterns between the anti-HA and anti-HOT Western blots. There is a notable *absence of anti-HOT stained bands* in lanes representing fractions G20 and G21 (**Figure 4.5**). Both of these fractions clearly have recombinant HOT-HA, since the anti-HA antibody in **Figure 4.4** detected the presence of that construct in these lanes. One likely explanation for this is that the presence of the C-terminal HA tag on recombinant HOT-HA obfuscates the epitope recognized by the anti-HOT antibody, resulting in *differential staining* of the two primary antibodies. This phenomenon will be revisited again in *Section 4.2.4.1*, when the effect is more pronounced during Western blot analysis of size exclusion chromatography fractions.

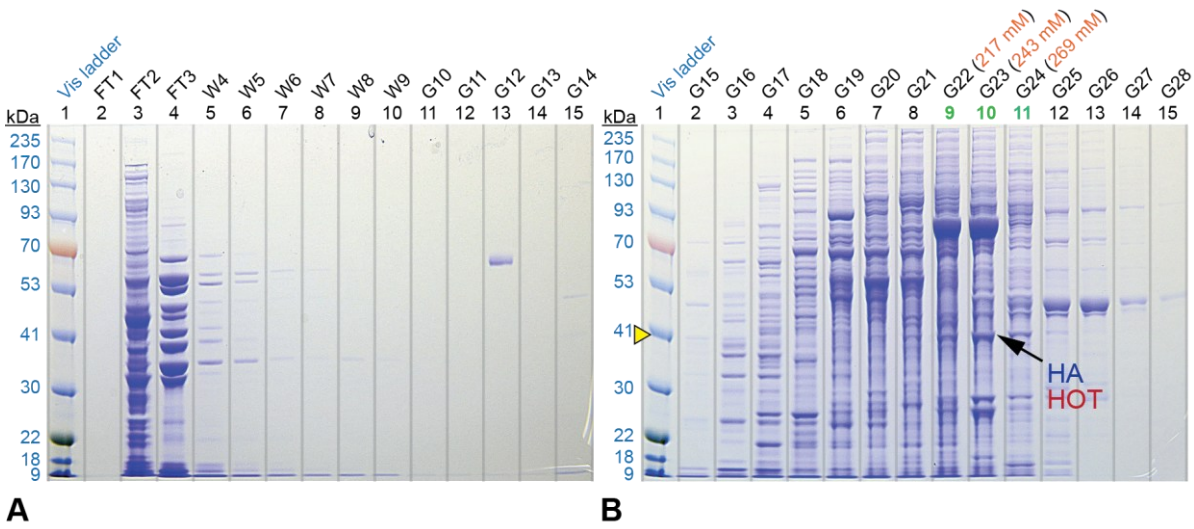

Figure 4.6. SDS-PAGE of AEC column fractions. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **black** arrows. Well numbers in **green** indicate fractions that were advanced to the next purification step. The NaCl concentration for select fractions appears in **orange** after the fraction identifiers (above the well numbers). FT = *flow-through*, W = *wash*, G = *salt gradient*. **(A)** Anion exchange *flow-through* fraction 1 through *salt gradient* fraction 14; **(B)** *salt gradient* fractions 15–28. **Resin:** DEAE-Sepharose FF. **Mobile phase:** 20 mM Tris, pH 7.6. **Gradient:** 0–500 mM NaCl.

Examination of the Coomassie-stained SDS-PAGE gel (**Figure 4.6**) indicates that some separation was achieved during anion exchange, but the fractions indicated by Western blot to contain recombinant HOT-HA and endogenous HOT remained heavily contaminated with other bands. Of particular note is what appears to be a significant number of chaperones from the Hsp90 family. The Hsp70 and Hsp90 family are known to bind proteins bound for the mitochondria and deliver them to mitochondrial translocases for import, as well as help them fold properly once inside.[14-16] To minimize the number of unwanted proteins carried over to the next purification step, only fractions G22–G24 were pooled and concentrated to a 0.5 mL sample with a total protein concentration of 28.18 mg/mL (14.09 mg;14.7% yield). It should be noted that this purification was repeated with the same 20 mM Tris mobile phase, and 0 – 500 mM NaCl salt gradient, but at pH 8.0. The results were nearly identical with the exception that peak elution for recombinant HOT-HA and endogenous HOT occur at 278 mM NaCl, rather than 243 mM NaCl.

Size Exclusion Chromatography as the 2nd Step

The pooled, concentrated and buffer exchanged fractions from anion exchange chromatography, 14.09 mg in a 0.5 mL sample (1.2% of the column volume), were loaded onto a Sephacryl S-200 column and separated using a mobile phase of 20 mM Tris, pH 7.6, 150 mM NaCl. Based on a calibration of the column with several protein standards, fractions F11–F24 were selected for analysis. As with the anion exchange chromatography, the results of the separation were followed by Western blotting using rabbit primary antibodies against both the HA-tag and HOT protein sequence, followed by goat anti-rabbit fluorescent-conjugated secondary antibodies and fluorescence imaging. To determine the quality of the separation, the column fractions were visualized on Coomassie-stained SDS-PAGE gel.

The Western blot stained with antibodies against the HA-tag (**Figure 4.7A**) show bands that run near the expected molecular weight of 46.2 kDa for recombinant HOT isoform 2 (M47HOT-HA), based on the Western ladder in Lane 1. Based on the molecular weight standard curve for the calibrated size exclusion resin, the bands begin to elute from the column with the fraction that would contain proteins with the molecular weight of 46 kDa. The Western blot stained with the anti-HOT antibody (**Figure 4.7B**) also shows bands in the same locations, indicating that both recombinant and endogenous HOT, with a difference in molecular weight of only 1.2 kDa, are distinguishable on this column (1.5 × 24.5 cm; 43.3 mL).

The Coomassie-stained SDS-PAGE gel of the SEC column fractions shows a tremendous gain in purification for the fractions F20–F22 (**Figure 4.8**). The suspected Hsp90 chaperone elutes largely ahead of these fractions, beginning at F14 (192 kDa), peaking at F15 (145 kDa)

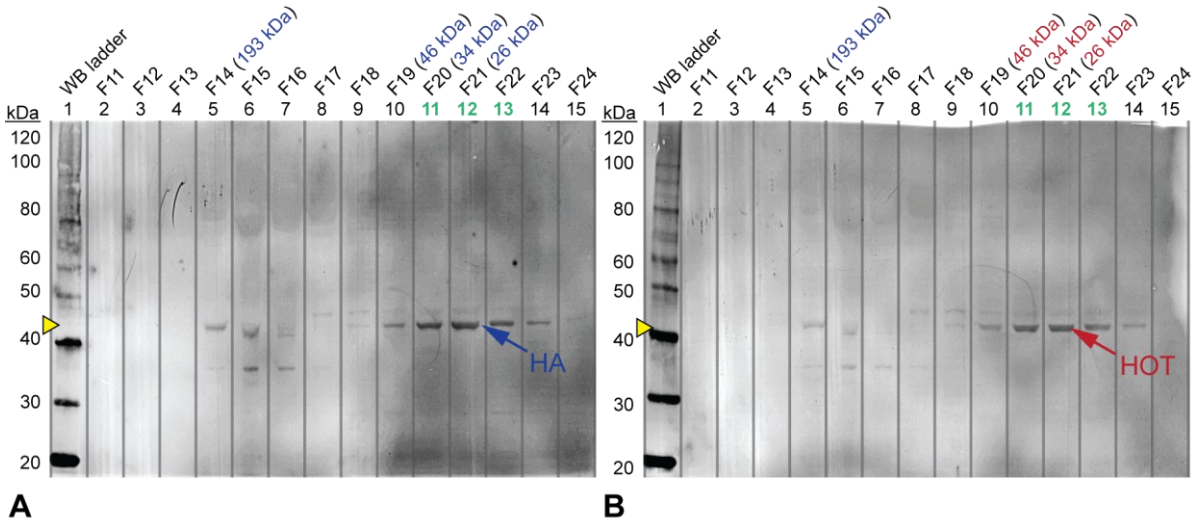

Figure 4.7. Western blots of SEC column fractions with HA and HOT 1° antibodies. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **blue** and **red** arrows. Well numbers in **green** indicate fractions of significant interest. The estimated molecular weight for select fractions appears in **blue** and **red**, after the *fraction* (F) identifiers (above well numbers). **(A)** Size exclusion fractions 11–24, stained with antibody against the HA-tag; **(B)** fractions 11–24, stained with antibody against the HOT protein sequence.

and trailing off by to half the peak intensity by F17 (82 kDa). This suggests that the suspected Hsp90 protein elutes predominantly as a dimer, and to a lesser extent as a monomer. The two bands that appear below the 30 kDa ladder, stained darkly in F18 and F19, also appear to elute as dimers or trimers.

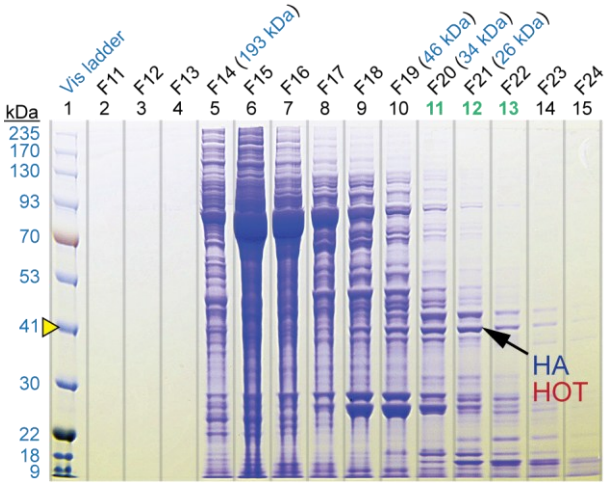

Figure 4.8. SDS-PAGE of SEC column fractions. The gel represents size exclusion column fractions 11–24. The **yellow** triangle indicates the molecular weight of the protein bands pointed out by the **black** arrow. Well numbers in **green** indicate fractions of significant interest. The estimated molecular weight for select fractions appears in **blue**, after the fraction (F) identifiers (above the well numbers).

This information suggests that a much longer SEC column and reduced rate of flow might result in better partitioning, allowing recombinant HOT-HA and endogenous HOT to be collected without the earlier contaminating bands overlapping these fractions. Ultimately, SEC fractions F20–F22 were pooled and concentrated to a volume of 0.5 mL with a total protein concentration of 3.71 mg/mL by Bradford assay (1.8g mg;13.2% yield).

4.2.4 Protein Purification Strategy II: SEC Before AEC

Size Exclusion Chromatography as the 1st Step

In order to observe the native oligomerization state of recombinant and endogenous HOT, clarified lysate from stably transfected PC-12 cells was loaded directly onto a calibrated size exclusion column. This column (1.5 × 62.0 cm; 109.6 mL) was much longer than the column described in *Section 4.2.3.2* (1.5 × 24.5 cm; 43.3 mL), with the additional expectation that better resolution would be obtained with the longer column and a slower flow rate. The mobile phase was once again 20 mM Tris, pH 7.6, 150 mM NaCl. The lysate was obtained by three freeze-thaw cycles on a 2 g PC-12 cell pellet (1.56 billion cells), which were concentrated to give a 1.0 mL sample (0.91% of the column volume) containing 42.16 mg total protein.

Western blots of size exclusion fractions with anti-HA and anti-HOT antibodies are shown in **Figures 4.9A** and **4.9B**. Of particular interest is the apparent *differential staining* of overlapping bands by the two antibodies, which was previously observed during anion exchange chromatography (*Section 4.2.3.1*). Here the effect is much more dramatic. Anti-HA staining begins in fraction F25 and ends in F31 (**Figure 4.9A**). By contrast, anti-HOT staining begins in fraction F29 and continues through the last fraction on the blot, F32 (**Figure 4.9B**). It would make sense that F32 is not stained by the anti-HA antibody, because there is no

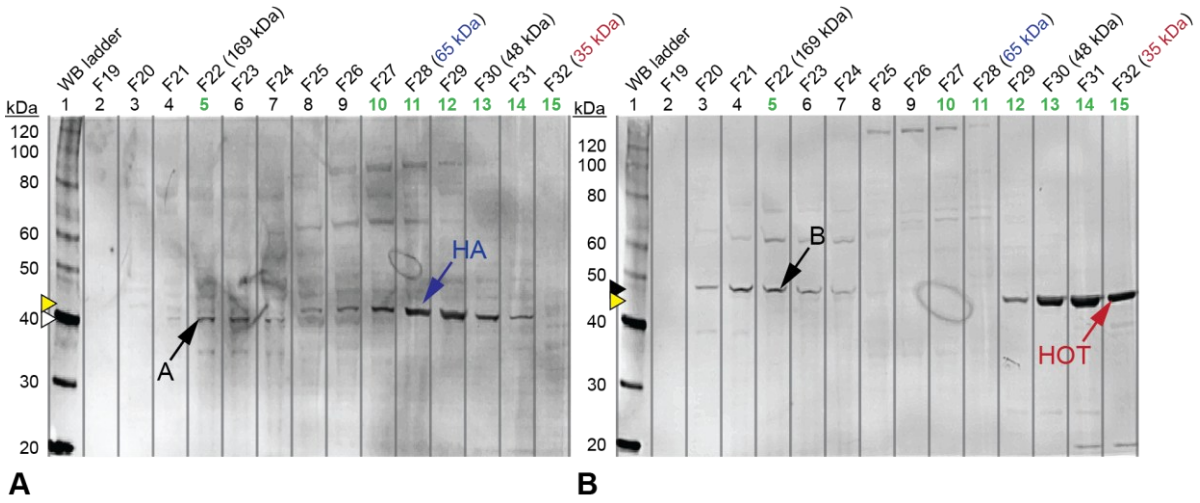

Figure 4.9. Western blots of SEC column fractions with HA and HOT 1º antibodies. The **yellow** triangle indicates the molecular weights of the bands pointed out by the **blue** and **red** arrows, while the **white** and **black** triangles indicate the molecular weights of the bands pointed out by the "A" and "B" arrows, respectively. Well numbers in **green** indicate fractions that were collected. Based on a calibration, the estimated molecular weight for select fractions appears in **blue** and **red**, after the fraction (F) identifiers (above the well numbers). **(A)** Size exclusion fractions 19–32, stained with antibodies against the HA-tag; **(B)** fractions 19–32, stained with antibodies against the HOT protein sequence

detectable recombinant HOT-HA in F32. However, F28 has a very detectable amount of recombinant HOT-HA, and yet the anti-HOT antibody does not appear to bind anything in that fraction. Both antibodies were used at an equivalent and generous 1:500 dilution in the blotting buffer, so the most probable explanation is that the C-terminal HA-tag on the recombinant HOT-HA blocks binding of the anti-HOT antibody to its epitope, which is presumably located near the C-terminus.

Of additional interest are the bands centered around fraction F22, which on first appearance suggests oligomeric forms of recombinant HOT-HA and endogenous HOT. Non-specific interaction of the anti-HA and anti-HOT antibodies was evident on many Western blots, with staining of bands well over 120 kDa under denaturing and reducing conditions. Such staining can clearly be disregarded as artifacts due to poor antibody specificity. However, some bands cannot be dismissed outright when they fall between 40 and 50 kDa, such as the

bands in fraction F22 marked “A” and “B”, in **Figures 4.9A** and **4.9B**, respectively. Fraction F22 was collected in order to analyze these bands further.

The gel used for the Western blot in **Figure 4.9A** seems to have become distorted while assembling the electroblotting stack, a process that involves using a roller to press out air bubbles between the SDS-PAGE gel and the PVDF blotting membrane. Too much pressure can stretch and deform the SDS-PAGE gel, which is what appears to have occurred on the far-right side (lanes 12–15). To correct this, both images in **Figure 4.9** were imported into Adobe Photoshop, colored blue (HA) or red (HOT), made transparent, and overlaid. The images were skewed until the Western blot ladders and other key landmarks in the blots were aligned, appearing black in the process due to the *subtractive color mixing* of blue and red (**Figure 4.10A**). To ensure complete honesty when making comparisons between original and processed figures, the relative positions of the annotations identifying bands in **Figures 4.9** and **4.10** have been fixed; *only the gel or blot images under them have been replaced*. This is readily apparent in **Figure 4.9A**, where the blue arrow adjacent to the HA in does not quite touch the band it points to, unlike in **Figure 4.10A**.

The subtractive coloring in **Figure 4.10A** illustrates the overlapping bands in fractions F29–F31, indicating the presence of both the HA and HOT epitopes. However, fractions F26–F28 show only blue bands resulting from those proteins being stained exclusively by anti-HA antibody, with no evidence of association between those bands and the anti-HOT antibody. Conversely, the single red band in F32 indicates that only the anti-HOT antibody binds to this band, with no evidence that the protein band interacts with the anti-HA antibody. The coloring and overlay of the two Western blot images also reveal the apparent *differential staining* by the two antibodies, providing further evidence that the C-terminal HA-tag may prevent binding

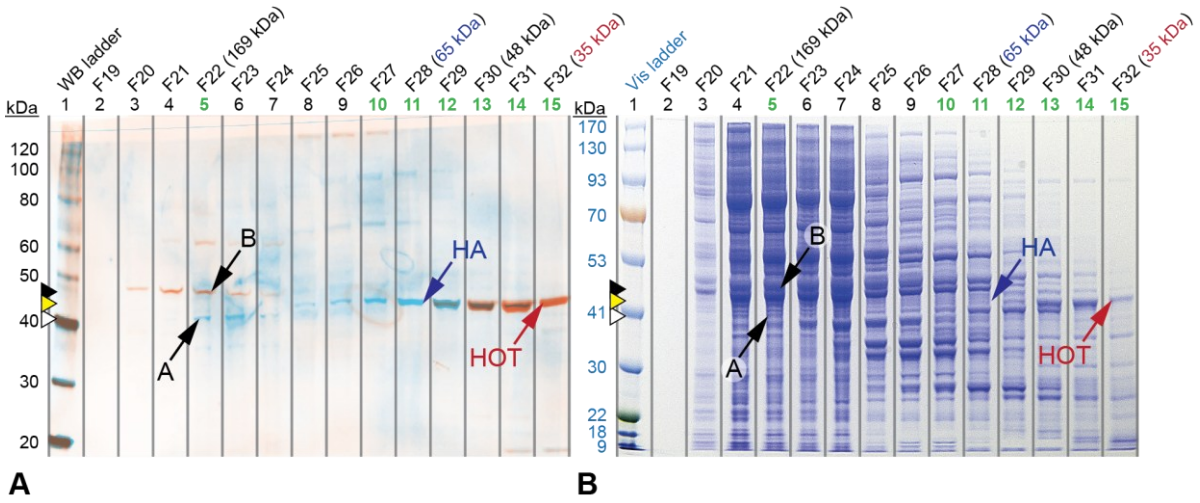

Figure 4.10. Composite Western blot and SDS-PAGE for SEC column fractions. The **yellow** triangle indicates the molecular weights of the bands pointed out by the **blue** and **red** arrows, while the **white** and **black** triangles indicate the molecular weights of the bands pointed out by the "A" and "B" arrows, respectively. Well numbers in **green** indicate fractions that were collected. The estimated molecular weight for select fractions appears in **blue** and **red**, after the fraction (F) identifiers (above the well numbers). **(A)** Size exclusion fractions 19–32 were stained with primary antibodies against the HA-tag and HOT protein sequences, and then digitally overlaid and adjusted to normalize the bands for comparison. **(B)** SDS-PAGE of fractions 19–32, with the annotations in Figure A used without changing their relative positions.

of the anti-HOT antibodies to their epitope. The color and overlay also strongly suggest that the unknown "A" and "B" bands may not be oligomers of recombinant HOT-HA or endogenous HOT, as they seem to run much higher and lower than the bands in fractions F27–F32.

The intention of a larger Sephacryl S-200 column was to partition the much larger molecular weight and oligomeric contaminants away from HOT-HA and HOT, which appear to run only as monomers. It was surprising to discover that HOT-HA and HOT could be resolved from each other, despite only a 1.2 kDa difference, by simply increasing the column size by 2.5 times. Some of this partitioning may result from the intrinsic properties of the HA-tag and its effect on protein interactions with the resin. The HA-tag (YPYDVPDYA) contains two negatively charged aspartates along with a two proline residues. If the prolyl peptide bonds are in the *cis* X-proline conformation, they may cause the tag to protrude from the surface of the recombinant protein with a pair of negative charges and three bulky and electron-rich

tyrosine residues. It would be interesting to see if increasing the NaCl concentration in the buffer from 150 mM to 300 mM would screen some of these interactions and attenuate the observed separation. While the ability to partition recombinant HOT-HA from endogenous HOT was promising, the Coomassie-stained SDS-PAGE gel indicated a significant number of contaminating proteins in all the fractions and further purification would be required (**Figure 4.10B**). Fractions F27–F32 were pooled, concentrated and desalted for the following anion exchange chromatography step. The total protein recovered from the SEC purification was a 6.47 mg total protein in 1 mL (15.3% yield).

Anion Exchange Chromatography as the 2nd Step

Anion exchange chromatography on DEAE-Sepharose described in *Section 4.2.3.1* was performed on the pooled, concentrated and desalted SEC fractions. This mobile phase of 20 mM Tris and 0–500 mM NaCl gradient was buffered at pH 8.0, instead of 7.6. **Figure 4.11** shows the elution of recombinant HOT-HA at 268–284 mM NaCl. This NaCl concentration is

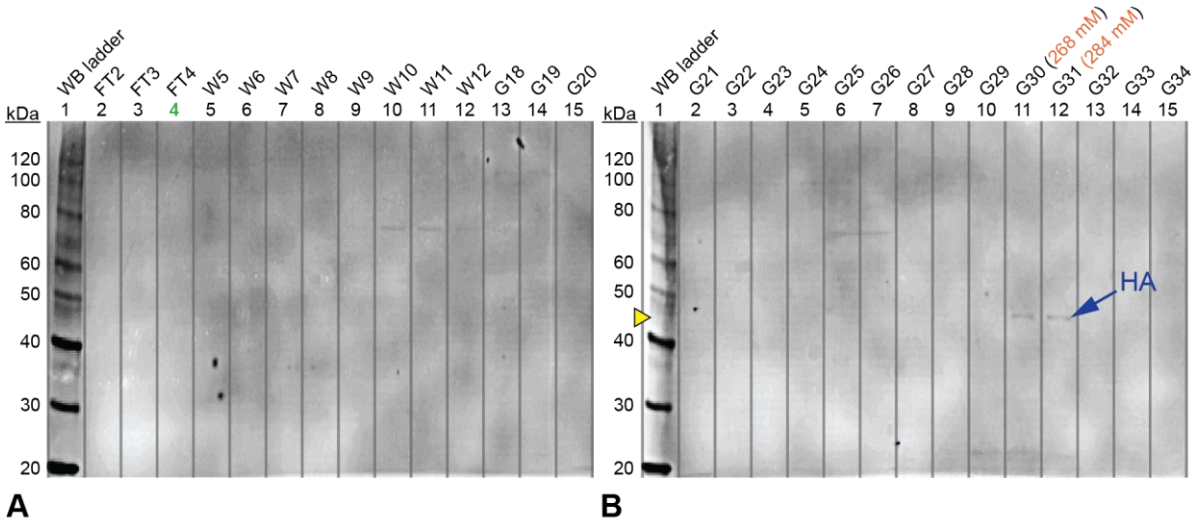

Figure 4.11. Western blots of AEC column fractions with HA 1º antibodies. The blots were stained with a primary antibody against the HA-tag sequence. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **blue** arrows. Well numbers in **green** indicate fractions that were collected for further analysis. The NaCl concentration for select fractions appears in **orange** after the fraction identifiers (above the well numbers). FT = *flow-through*, W = *wash*, G = *salt gradient*. **(A)** Anion exchange *flow-through* fraction 2 through *salt gradient* fraction 20; **(B)** *salt gradient* fractions 21–34.

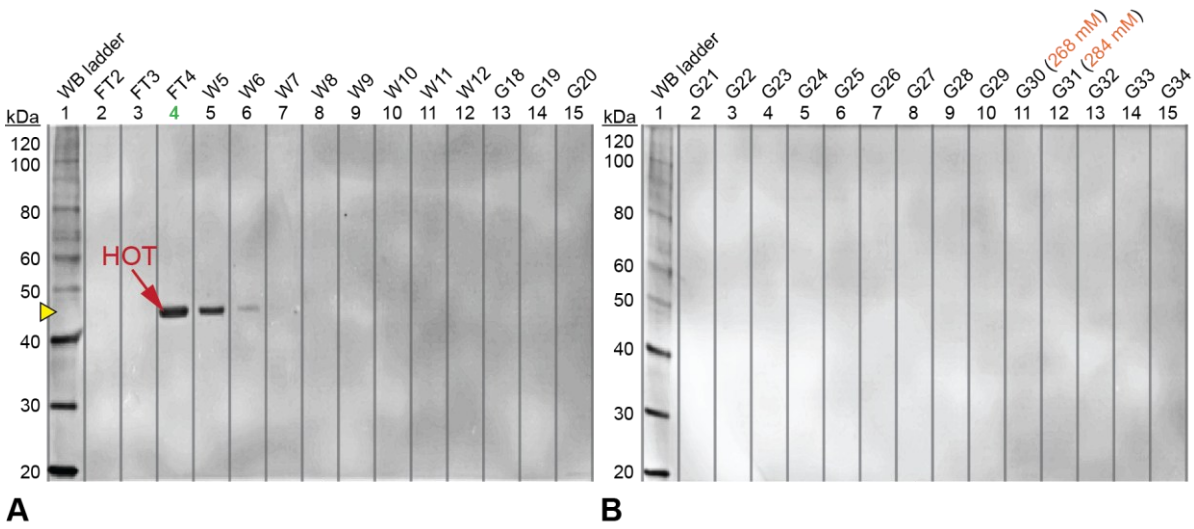

Figure 4.12. Western blots of AEC column fractions with HOT 1º antibodies. The blots were stained with a primary antibody against the HOT protein sequence. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **red** arrows. Well numbers in **green** indicate fractions that were collected for further analysis. The NaCl concentration for select fractions appears in **orange** after the fraction identifiers (above the well numbers). FT = *flow-through*, W = *wash*, G = *salt gradient*. **(A)** Anion exchange *flow-through* fraction 2 through *salt gradient* fraction 20; **(B)** *salt gradient* fractions 21–34.

in close to the 243 mM NaCl that both recombinant HOT-HA and endogenous HOT eluted with when purifying clarified cell lysate at pH 7.6. A surprising observation was that endogenous HOT did not bind the DEAE-Sepharose, even at the increased pH, eluting instead in the tail-end of the column flow-through, in fractions FT4–W6 (**Figure 4.12**). It may be that semi-purified endogenous HOT molecules aggregated after desalting (0 mM NaCl), where the 20 mM Tris buffer provides a limited contribution to ionic strength and charge screening. Aggregated HOT would potentially have fewer accessible negative charges required bind the positively charged diethylaminoethyl groups of the resin. Aggregation may be less of an issue with clarified lysate due, even desalted, since the sample would still contain a substantial variety of charged molecular species and HOT would be effectively less concentrated.

Based on the SDS-PAGE gel, the endogenous HOT that eluted in flow-through fraction FT4 is remarkably pure, with only faint traces of contaminating bands (**Figure 4.13A**). Fraction FT5 also contains a significant amount of endogenous HOT, though more

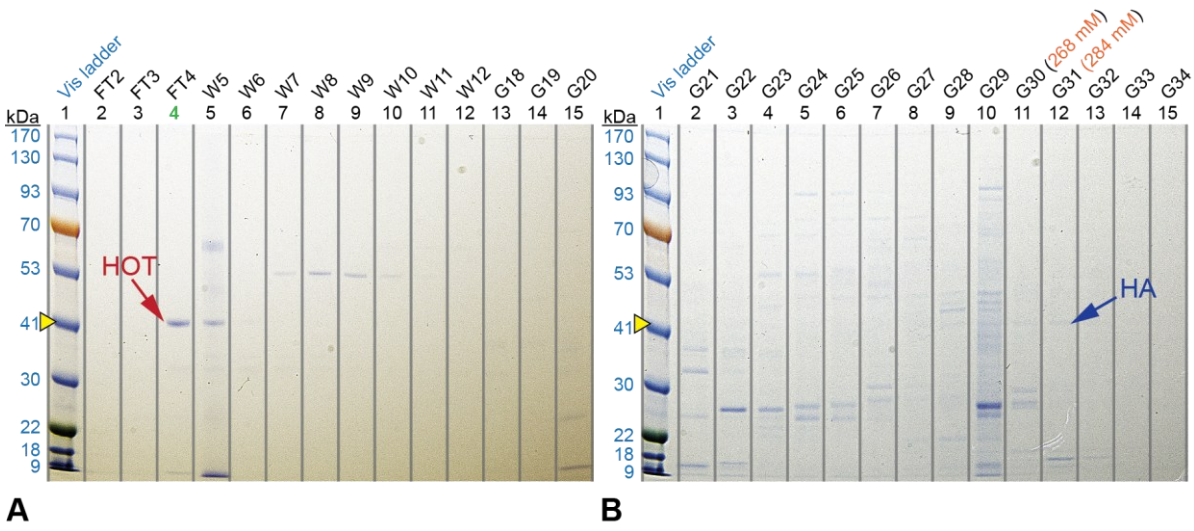

Figure 4.13. SDS-PAGE gel of AEC column fractions. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **red** arrows. Well numbers in **green** indicate fractions that were collected for further analysis. The NaCl concentration for select fractions appears in **orange** after the fraction identifiers (above the well numbers). FT = *flow-through*, W = *wash*, G = *salt gradient*. **(A)** Anion exchange *flow-through* fraction 2 through *salt gradient* fraction 20; **(B)** *salt gradient* fractions 21–34.

contaminating proteins are evident here. With respect to recombinant HOT-HA, gradient fractions G30 and G31 are also devoid of heavy contamination, but the Coomassie-stained HOT-HA bands are also almost too faint to make out (**Figure 4.13B**). Still, the potential increase in purity by simply reversing the purification steps, with SEC preceding AEC, is clearly demonstrated to be a valid strategy.

Fraction FT4 had a volume of 5 mL, with a total protein concentration of 0.13 mg/mL for a 0.65 mg yield of nearly pure endogenous HOT (10.0% yield). The recombinant HOT-HA bands in G30 and G31 were overlooked because they did not readily stand out on the blot and gel until the images were processed some days later. As a result, these fractions were not collected or analyzed for protein concentration.

Since a very pure sample of endogenous HOT was now available, it was used to determine if the unknown bands "A" and "B" in SEC fraction F22 (**Figures 4.9** and **4.10**), which eluted with high molecular weight SEC fractions, represented HOT oligomers or were simply artifacts

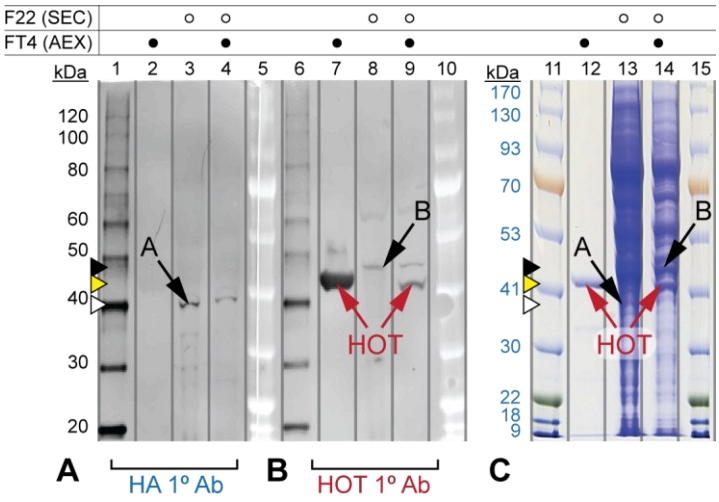

Figure 4.14. Western blot and SDS-PAGE comparison of fractions FT4 and F22. The **yellow** triangle indicates the molecular weights of the protein bands pointed out by the **red** arrows, while the **white** and **black** triangles indicate the molecular weights of the "A" and "B" arrows, respectively. Western blot ladders are in lanes 1 and 6, while visible ladders are in lanes 5, 10, 11, and 15. **(A)** A comparison between fractions FT4 (lane 2) and F22 (lane 3) alone, and together (lane 4), stained with anti-HA antibodies. **(B)** A comparison between fractions FT4 (lane 7) and F22 (lane 8) alone, and together (lane 9), stained with anti-HOT antibodies. **(C)** SDS-PAGE comparison between fractions FT4 (lane 12) and F22 (lane 13) alone, and together (lane 14).

that resulted from poor antibody specificity. SEC fraction F22 was loaded side by side and together with AEC fraction FT4 on Western blots and an SDS-PAGE gel. The results of these comparisons are shown in **Figure 4.14**, and to ensure objectivity, the spacing between the annotations (arrows and labels) is not changed between **Figures 4.14A–C**. For the blot stained with the anti-HA antibody, only band "A" from SEC F22 is detected (**Figure 4.14A**), while endogenous HOT in FT4 (in lane 2) is not stained because it does not have an HA-tag. For the blot stained with the anti-HOT antibody (**Figure 4.14B**), endogenous HOT from AEC FT4 is clearly visible in both lane 7 (run alone) and in lane 9 (run together with SEC F22). The anti-HOT antibody also detects band "B" from SEC F22. The Coomassie-stained gel image for this comparison is shown in **Figure 4.14C**. Based on their migration with respect to endogenous HOT, bands "A" and "B" appear to be entirely different proteins, with their staining by anti-HA and anti-HOT antibodies the result of non-specific binding.

4.3 Conclusion

Both the COS-7 and PC-12 cell lines were transfected to produce lineages of cells stably expressing recombinant HOT-HA. While both cell lines are capable of integrating the recombinant DNA into their genomes, it was determined that protein expression in PC-12 cells is more efficient in terms of time, effort and cost, yielding three times as many cells per cm^2 as the COS-7 cells, and in only half the time.

Cell viability studies of untransfected PC-12 cells grown in the presence of G418 provided the optimal concentrations of the antibiotic for the selection and maintenance of stable transfectants. Cationic lipids were used to deliver the vector encoding antibiotic resistance and recombinant HOT to the cells. Stable transfectants that had integrated these vector elements into their genome were selected with G418. Western blotting was used to assay surviving colonies for the expression of recombinant HOT, which contained a C-terminal HA-tag. Colonies that showed the high expression levels of recombinant HOT-HA were identified, preserved as cryogenic stocks, and cultured for the production of recombinant HOT-HA as well as endogenously produced HOT.

Two methods were used to lyse PC-12 cells for protein extraction, free-thaw and liquid homogenization with a Potter-Elvehjem homogenizer. Liquid homogenization ultimately results in the recovery of more total protein. At the level of protein analysis, which included Western blots and SDS-PAGE, there was no discernible difference in the quality or type of proteins recovered by each method. This is largely due to the addition of 0.1% Triton X-100 to the lysis buffer, which solubilizes the lipid membranes. HOT is reportedly localized to the mitochondria. Liquid homogenization, combined with a detergent-free, isotonic lysis buffer will shear only the outer cell membrane, releasing intact organelles that then can be separated

by differential centrifugation. The isolation of mitochondria from PC-12 cells was in the early stages before this research was interrupted by the 2020 SARS-CoV-2 pandemic.

Preliminary purification of recombinant HOT-HA and endogenous HOT was performed using a combination of anionic exchange chromatography (AEC) and size exclusion chromatography (SEC). Altering the order of these two chromatographic methods resulted in unique observations about how recombinant and endogenous HOT partitions under different conditions. When AEC is performed first both recombinant and endogenous HOT elute at 284 mM NaCl at pH 7.6 (or 243 mM NaCl at pH 8.0). With the following SEC, both recombinant and endogenous HOT elute together, primarily in the 46–26 kDa fractions. However, when SEC is performed first, the pooled and desalted fractions subsequently run on AEC partition very differently, with endogenous HOT nearly pure with the flow through while recombinant HOT elutes at high salt as expected. Additionally, increasing the size of the SEC column resulted in partitioning of recombinant and endogenous HOT into different column fractions, despite only a 1.2 kDa difference in molecular weight. Lastly, Western blotting revealed that some polyclonal antibodies raised against the HOT sequence do not bind recombinant HOT, which differs only in a C-terminal HA-tag. These observations suggest that the HA-tag prevents aggregation of HOT in the absence of salt, deters interactions with Sephacryl (SEC) resin, and can block the major epitope for polyclonal antibodies raised against the HOT protein sequence.

4.4 Materials and Methods

4.4.1 Mammalian Cell Culture

General Practices

Aseptic techniques were strictly practiced. Any cell culture work with exposed cells, such as extracting colonies with cloning rings, harvesting cells, and changing the media in flasks, was performed in a Labconco® Purifier™ Class II biosafety cabinet equipped with Drummond Pipet-Aid® pipette controller. Every surface in the cabinet was wiped with 70% alcohol (ethanol or isopropanol) and the built-in UV light was turned on 30 minutes before and after work. Anything brought in or out of the cabinet was also sprayed with alcohol, including Parafilm M wrapped media tubes and bottles, wrapped serological pipettes, media flasks and gloved hands. Reusable items with internal spaces, such as pipettors and glassware (bottles and Pasteur pipettes), were covered in aluminum foil or canisters and autoclaved for 30 minutes at 121 ºC and 20 psi. All plastic consumables, such as culture flasks and serological pipettes, arrived from the manufacturer in sterile packaging.

All materials used in the culture of immortal cell lines were deemed biohazardous and treated appropriately. While PC-12 cells are classified as *biosafety level* (BSL) 1, COS-7 cells are classified are classified as BSL-2 because they contain *simian vacuolating virus 40* sequences and should be handled with caution. The Tygon® tubing and Erlenmeyer vacuum flask used with the Pasteur pipettes for aspirating spent media were sanitized with a 10% bleach solution, but the Pasteur pipettes themselves were placed into a “sharps” container that was sterilized by autoclave before disposal, along with any plastic consumables that were in contact with cells.

All cell lines were cultured at 37 ºC in Shel Lab® Model 2350T/2350B CO_2 incubators, with 5% CO_2 and equipped with a humidity tray filled with autoclaved water supplemented with VWR Clear Bath® Algae Inhibitor. The incubators were routinely disassembled and sterilized by autoclaving detachable components and/or with 70% alcohol.

Cryogenic Preservation

Cryogenic stocks were maintained in Taylor-Wharton 35VHC and 35VHCB-11M liquid nitrogen (-196 ºC) dewars with Cryo-Sentry low-level alarms. Cryogenically preserved cells were stored in threaded cap cryotubes. Cells were thawed quickly in a 37 ºC water bath for one minute per 1 mL of cryogenically preserved cells. The thawed solution of cells was immediately diluted 1:10 into pre-warmed 37 ºC media, centrifuged at 200 × *g* for 5 minutes, the media aspirated and then replaced with fresh pre-warmed media. This is done to remove DMSO, which is used to prevent damage to the cells from ice crystals but is otherwise toxic to the cells themselves.

COS-7 and PC-12 cryogenic cell stocks were prepared by growing the cells to 50% confluency and then harvesting them with trypsin. Cells were counted and the cell density was adjusted to 2 million cells per mL in freezing media, which is complete media for the respective cell type that was supplemented with 5% DMSO. Cryotubes were placed in Styrofoam holders, then the holders were placed into a Styrofoam box with a lid, and the box was placed into a -80 ºC freezer overnight. The following day the cryotubes were transferred into the liquid nitrogen storage dewars for cryogenic preservation.

Cells and Media

Untransfected COS-7 (CRL-1651) and PC-12 Adh (CRL-1721.1) cells were obtained from ATCC®. Both cell lines were thawed as described and then subcultured twice to provide an ample amount of cryogenic cell stock with a low passage number.

The base media for the COS-7 cell line is Dulbecco's Modified Eagle Medium (DMEM), which contains amino acids, vitamins, glucose, pyruvate and L-glutamine. *Incomplete* COS-7 media was prepared from powdered DMEM (ThermoFisher 12400024) packets, 1 per liter, supplemented with 3.7 g/L sodium bicarbonate. After dissolving with 18 MΩ water, the pH of the incomplete media was adjusted to 7.4 and, while working in the biosafety cabinet, sterile filtered into a 1000 mL disposable vacuum filter/storage bottle with a 0.2 µm polyethersulfone (PES) filter. To be complete, the incomplete base media is supplemented with serum that contains proteins, lipids and growth factors. The *complete* COS-7 media was prepared by adding 10% fetal bovine serum (FBS) and the appropriate antibiotic. Untransfected cells should be supplemented with 100 units/L penicillin and 100 µg/L streptomycin. Transfected cells are supplemented with Geneticin (G418) at a selection or maintenance dose that is determined by a viability study on the cell line.

The base media for the PC-12 cell line is DMEM, which contains amino acids, vitamins, glucose, pyruvate and L-glutamine, plus Ham's F-12 which includes zinc, putrescine, hypoxanthine, and thymidine. The ratio of DMEM to F12 is 1:1. PC-12 *incomplete* media was prepared by dissolving powdered DMEM/F-12 (ThermoFisher 12400024) supplemented with 1.2 g/L sodium bicarbonate in 18 MΩ water, adjusting the pH to 7.4 and filtered in same manner as the COS-7 incomplete media. *Complete* PC-12 media is supplemented with 5%

FBS, 10% donor horse serum, and either penicillin/streptomycin (untransfected cells) or G418 (transfected cells).

All incomplete and complete media were stored at 2 – 4 ºC. The L-glutamine in the growth media is known to rapidly and spontaneously decompose into pyroglutamic acid and ammonia, depriving cells of this important amino acid and exposing them to toxic ammonia.[59, 60] To limit L-glutamine decomposition, all media was used within 3 weeks of preparation. Because media must be pre-warmed to 37 ºC before use with cells, a process that accelerates L-glutamine decomposition, both incomplete and complete media were also aliquoted into 50 mL Falcon tubes to minimize exposure media to repeated heating cycles. Media in 1000 mL bottles was only warmed if daily usage exceeded 250 mL, which also limited potential media contamination. Media was warmed in a 37 ºC water bath that was periodically cleaned and supplemented with quaternary ammonium salts (i.e., Clear Bath).

FBS and HS are shipped frozen, thawed, aliquoted out into 50 mL Falcon tubes, which are sealed with Parafilm M in the biosafety cabinet before being placed in a non-defrosting freezer at -20 ºC. 100X penicillin/streptomycin were aliquoted out in 10 mL Falcon tubes, sealed with Parafilm M and placed at -20 ºC. G418 is shipped on ice in a bottle as a powder and stored at -20 ºC. Each lot of G418 has a unique potency, listed as *purity* in µg/mg. This purity value is used to correct for the actual concentration when making the G418 stock solution. G418 stock solution was prepared combining the appropriate mass of HEPES and G418 powders to make a stock solution that had the final concentrations of 100 mM HEPES and 50 mg/mL (actual) G418. HEPES and G418 powders were dissolved in 18 MΩ water, the pH was adjusted to 7.4, and the solution was filter sterilized (0.22 µm) in the biosafety cabinet. The G418 stock solution was then aliquoted out in 10 mL and 50 mL Falcon tubes, sealed with

Parafilm M and stored at -20 º. Incomplete and complete media were frequently aliquoted out into 50 mL Falcon tubes and sealed with Parafilm M for general use. The use of 1000 mL media bottles, also sealed with Parafilm M, was reserved for daily volumes exceeding 300 mL.

Subculturing and Harvesting of Cells

The COS-7 and PC-12 cell lines are adherent and attached themselves covalently to tissue coated (TC) surfaces. The practice of detaching the cells from the surface of a flask and transferring them to another flask is known as subculturing, passaging, or splitting. Each time cells are subcultured their passage number increases by 1. An alternative and more accurate method for tracking the "age" of a cell culture is to record the doubling number of the cells. This is because splitting a single flask of confluent cells in half will give two new flasks that can double one time before having to be split again, while splitting this flask into four new flasks permits them to double twice. Both sets of flasks have increased their passage number by 1, but the second set of flasks will have seen two rounds of cell division. Cells with high passage or doubling numbers may show unexpected phenotypic changes due to genetic instability or from the selective pressures of the media, and this in turn may affect the outcome of experiments. For this reason, the COS-7 and PC-12 cells were not passaged more than 30 to 40 times before fresh cultures were seeded with cells from the cryogenic stocks. No obvious changes in cell morphology were observed under a confocal microscope and the protein expression patterns appeared stable.

The TC-coated flask sizes used were T-25, T-75, and T-175, where the number indicates the available surface area in cm^2 for cells to grow on. Flasks were equipped with screw caps containing a hydrophobic filter membrane that permitted continuous gas exchange. The volume of media, phosphate buffered saline (PBS) pH 7.4 (without calcium and magnesium),

and trypsin-EDTA (0.05%) use for each flask size is reported in **Table 10**. Complete media is used for quenching the trypsin reaction, as the supplemented serum contains trypsin inhibitors.

Table 4.1. Media, PBS and trypsin volumes for subculturing.

			Complete media	
Flask size	**PBS (mL)**	**Trypsin-EDTA (mL)**	**Trypsin quench (mL)**	**Growth (mL)**
T-25	4	2	4	7
T-75	12	6	12	21
T-175	24	12	24	49

When working with cells, all media, phosphate buffered saline (PBS), and trypsin were pre-warmed to 37 º C in a water bath, then sterilized with alcohol before being transferred to the biosafety cabinet. All culture flasks in the incubator were sprayed with alcohol before being transferred to the biosafety cabinet and again before being returned to the incubator. Cells were typically subcultured when their confluence reached 70–80% in order to keep the cells healthy for propagation and expansion of new cell cultures. If cells were being harvested only for lysis, their confluence was allowed to reach 100%.

To subculture cells, flasks were transferred from the incubator to the biosafety cabinet where the exhausted media was aspirated using an autoclaved glass Pasteur pipette attached to the house vacuum line by Tygon tubing. Two intervening vacuum Erlenmeyer flasks were placed in series along the tubing to collect the hazardous waste and prevent accidental aspiration into the house line. Using a sterile serological pipette, PBS was carefully added to the top surface of the flask, opposite the bottom growing surface, in such a way as to not disturb the attached cells. Flasks were returned to their original position, with the growing surface on the bottom, and the layer of warm PBS was gently moved over the cells by gentle rocking of the flask 3 to 4 times. The purpose of the PBS was to remove traces of the exhausted medium, which had become acidic and contained trypsin inhibitors, both of which would interfere with

the following use of trypsin. DO NOT stand flasks up on their short side (caps pointing up) when there are adherent cells on the flask surface. Vibrations from the biosafety cabinet blower will cause the PBS (or other liquid) meniscus to rapidly oscillate over the cells, detaching them prematurely and resulting in lower cell recovery.

The PBS was carefully aspirated and replaced with the trypsin-EDTA (0.05%) solution. Generally, the warm trypsin solution detached the cells after about 10 minutes with the occasional, gentle rocking of the flasks in the biosafety cabinet. In a few instances where the laboratory was too cold and detachment was incomplete, flasks were returned to the 37 ºC incubators for 5 minutes to assist the trypsin reaction. Despite sheets of cells visibly sloughing off the flask surface, one or two sample flasks were typically inspected under the microscope to ensure all of the cells were detached. Using a sterile serological pipette, fresh complete media was added to the flasks to quench the trypsin. The mixture in the flask was drawn back into the pipette and used to gently wash the cells from the flask surface several times, then transferred into a sterile Falcon tube. The Falcon tube was sealed with Parafilm M, placed into a clinical centrifuge and the cells were spun at 200 × *g* for 5 minutes at room temperature. The Falcon tubes with pelleted cells were returned to the biosafety cabinet where the supernatant was carefully aspirated, and the cells resuspended with enough complete media to fill the tube. 10 µL of suspended cells were removed, placed onto a hemocytometer, and cells within 5 squares were counted under the magnification of a confocal microscope. The cell density of the suspension, in cells per mL, was calculated with **Equation 4.1**.

Equation 4.1. Hemocytometer formula.

$$\text{Cell density (cells/mL)} = \frac{\text{cells counted}}{\text{squares counted}} \times 10{,}000 \times \text{dilution}$$

A record for each cell line was kept that included the number of cells harvested at 100% confluency for any given flask size, providing the maximum supported cell density in cells/cm^2. This value was used to find the optimal cell number required for seeding new flasks, which depended on the purpose. Flasks that were seeded in order to propagate and expand the cell culture were seeded at 15 – 20% confluency, so that after two doublings they would be between 70 – 80% confluent. Flasks that were seeded for the purpose of harvesting cells for lysis were seeded at 25% so they would be 100% confluent after two doublings. Once the volume of suspended cells required to seed a flask at the optimal confluency was determined, this volume was added to flasks containing pre-warmed complete media using a sterile serological pipette and dispersed by gently pipetting up and down three times. The flasks were then capped and returned to the incubator.

When harvesting cells for lysis, the same general procedure for harvesting and counting applied except the Falcon tubes were tared. The counted cells suspended in complete media were centrifuged at 200 × *g* for 5 minutes to pellet the cells so that the complete media could be aspirated. PBS was used to resuspend the cells with gentle pipetting, the cells were centrifuged as before, and the PBS was removed by aspiration. The Falcon tubes with cell pellets were weighted and the pellet mass and number of cells noted. After the base of the lysis buffer was removed the cells were either placed on wet ice in a refrigerator at -2–4 ºC for no more than 24 hours, or flash frozen in liquid nitrogen (first of 3 freeze-thaw cycles) and placed in a -80 ºC freezer until the remainder of the freeze-thaw cycles could be completed.

Transfection of PC-12 Cells and Selection with G418

Approximately 1 day before transfection, enough PC-12 cells were suspended into complete media without antibiotics so they could be added into each well of a 6-well plate at

approximately 45% confluency. The following day, when the cells were 90–95% confluent, they were transfected using Lipofectamine™ 2000 Transfection Reagent (ThermoFisher 11668027) and the 6-well plate transfection protocol provided by the manufacturer for 6, 9, 12 and 15 μL of transfection reagent per well. Two additional wells served as controls (untransfected and mock transfection). The pcDNA3.1 vector was diluted into Opti-MEM™ Reduced Serum Media per the instructions. The 6-well plate was placed into the 37 ºC CO_2 incubator for 24 hours, after which time cells from each well were transferred (trypsin protocol) to 4 × 10-cm tissue culture treated dishes with complete media minus antibiotics. Each well of a 6-well plate is 9.6 cm^2, and each of the four10-cm dishes has an area of 56.7 cm^2, so this was effectively a 1:24 split. After 24 hours the media was replaced with complete media supplemented with 500 μg/mL G418 to begin selection.

After a few weeks of selection with G418, several small colonies were observed on many of the dishes, each resulting from a single cell that had successfully acquired the neomycin resistance gene during transfection. After six weeks under the selective pressure of G418, 18 colonies were large enough to harvest with the use of trypsin and cloning cylinders. Using cloning cylinders, each of the colonies was extracted and transferred to an individual well of a 12-well plate, and once those cells reached 70–80% confluency they were used to seed T-25 flasks. The T-25 flasks were incubated until cells reached approximately 90% confluency, at which point approximately 15% of the cells were used to seed a new T-25 flask for the expansion of potential cryogenic stocks, while the remainder of the cells were harvested for lysis. The cell lysate from each of the 18 colonies was run on SDS-PAGE and analyzed by Western blotting using primary antibodies against the HA-tag. Colonies that showed high expression of the HA-tag were selected for cryogenic preservation.

Cell Lysis

Two methods were used to lyse PC-12 cells, freeze-thaw and liquid homogenization with a Potter-Elvehjem homogenizer. The lysis buffer for both methods was 20 mM Tris, pH 7.6, 150 mM NaCl, 0.1% Triton X-100, and GoldBio ProBlock™ Gold Mammalian (GB-108-2) protease inhibitor cocktail. 10 mL of lysis buffer was used per gram of wet cell pellet, which was approximately 1 billion cells.

For freeze-thaw, cells in a Falcon tube were immersed in liquid nitrogen for 30 seconds, then allowed to thaw at room temperature. This was repeated twice, for a total of 3 freeze-thaw cycles. If the cells had been previously frozen in liquid nitrogen and stored at -80 °C, then this counted as the first freeze and only two additional freeze-thaw cycles were performed. For liquid homogenization, cells suspended in lysis buffer were poured into a glass Potter-Elvehjem homogenizer tube sitting on wet ice. The polytetrafluoroethylene (PTFE) pestle was connected to connected to a motor at a slow speed and the glass tube was slowly lifted from the ice until the pestle reached the bottom of the tube while spinning, and then the pestled was withdrawn by lowering the tube back into the ice. This was performed 10 times and then the cells were checked for lysis with a dye exclusion test which can identify cells with compromised plasma membranes. 0.1 mL of Gibco™ 0.4% Trypan Blue Solution (15250061) was mixed with 0.1 mL of suspended cells, and 10 µL of this mixture was placed into a hemocytometer. Blue cells indicated compromised plasma membranes. Homogenization continued until ≥95% of the cells were blue, which was 20–30 passes of the pestle.

Cells were spun at 20,000 × *g* (~17,300 RPM w/6 cm radius; Beckman Type 70.1 Ti rotor) in Sorvall Discovery 90 ultracentrifuge for 15 min at 4 °C. Supernatant was decanted into a clean tube and spun at 50,000 × *g* (~27,300 RPM) for 30 minutes at 4 °C. A thin and easily

disrupted lipid layer floated at the surface of the supernatant, so a glass pasture pipette was passed though the lip layer and used to carefully withdraw the supernatant without disrupting the lipid layer above and the pelleted debris below. After carefully transferring the clarified lysate to a clean tube it was either concentrated and desalted for anion exchange chromatography, or simply concentrated for size exclusion chromatography centrifuging in a Sartorius Vivaspin® 6 Centrifugal Concentrator (30 kDa MWCO) at 6,000 × *g* and 4 °C in a Beckman JA-20 rotor mounted in a Beckman J2-HS (and J2-21) centrifuge.

4.4.2 Protein Purification

Anion Exchange Chromatography

Anion exchange chromatography was performed using GE Healthcare DEAE Sepharose® Fast Flow (CAS #: 57407-08-6; Sigma-Aldrich: DFF100) resin loaded onto two Bio-Rad Econo-Columns®. One was a 1.5 × 20.0 cm column filled to a height of 7.5 cm (13.25 mL bed volume) and the second was a 1.5 × 20.0 cm column filled to a height of 11.5 cm (20.32 mL bed volume). Both columns were fitted with a Bio-Rad Econo-Column® Flow Adaptor, was used to pack the column as well as load and elute the sample. The adapter was connected through tubing to a Bio-Rad Model EP-1 Econo Pump, which was fed by a GibcoBRL/LifeTechnologies™ Model 750 gradient former. A T-valve with Luer-Lok™ fittings was installed halfway between the adapter and the pump to interrupt the mobile phase and inject the sample directly onto resin, via the flow adapter.

The gradient former (originally designed to make gradient gels) had an inner and outer chamber that held buffer, with a valve that controlled the flow between them. The inner chamber led directly to the pump and was fitted with a stir bar, while the outer chamber was

the furthest reservoir from the pump. A buffer with 0 mM NaCl was placed in the inner chamber, while an equal volume of buffer containing 1000 mM NaCl was placed in the outer chamber. Uniform mixing was achieved by opening the valve between the two chambers, allowing the more concentrated solution in the outer chamber to flow into the inner chamber where the stir bar would mix the two before the combined solution was pumped onto the column. Fractions from the column were collected in VWR disposable borosilicate glass culture tubes (13 × 100 mm) that were placed into a Bio-Rad Model 2110 Fraction Collector. 4.4 mL fractions were collected for the 7.5 cm column, and 5.1 mL fractions were collected for the 11.5 cm column.

In a refrigerated room, the resin on the column was equilibrated with 5 column volumes of 20 mM Tris, pH 7.6 (or pH 8.0), 0 mM NaCl with a flow rate of 1.0 mL/min. A 20 mM Tris buffer at pH 7.6 was used for the 7.5 cm column, while a 20 mM Tris buffer at pH 8.0 was used for the 11.5 cm column. From here on, the method described is for purification performed at pH 7.6. Apart from the bed volume, fraction volumes and pH differences, the methods were the same. The pump was stopped, and the clarified cell lysate was slowly loaded onto the column with a syringe that was attached to a Luer-Lok port on the T-valve in the direction that led to the resin. Once the valve was closed, the pump was restarted and column was washed 2 column volumes of equilibration buffer (0 mM NaCl) at a flow rate of 0.5 mL/min, which would remain the flow rate for the duration of the purification.

The target protein was eluted by opening the valve on the gradient former, allowing the 1000 mM NaCl buffer in the outer chamber to mix with the 0 mM NaCl buffer in the inner chamber, resulting in a 0–500 mM gradient applied to the column over 10 column volumes. Following the gradient, the column was washed with 20 mM Tris, pH 7.6, and 1000 mM NaCl

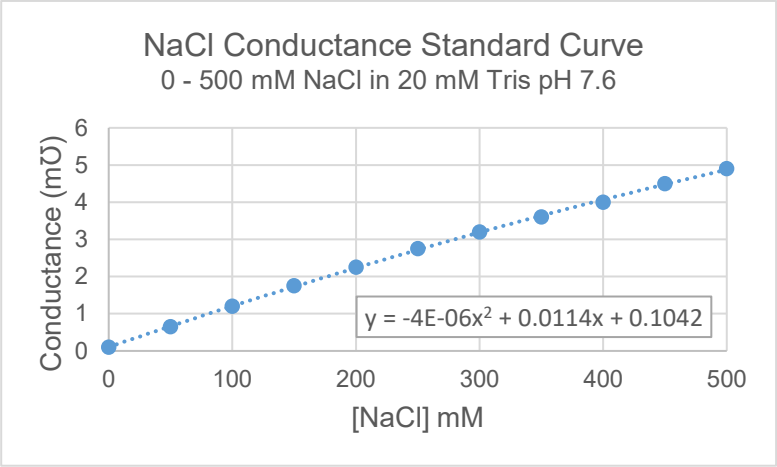

Figure 4.15. Standard curve for NaCl conductance.

for 5 column volumes. Occasional cleaning in place involved washing with 1.0 M NaOH at a flow velocity of 40 cm/h for 1 or 2 hours, and 30% isopropanol for 1 or 2 hours.

A Radiometer Copenhagen Type CDM2e conductivity meter was used to measure the NaCl concentrations in the fractions. To calibrate the meter readings, a standard curve of conductance versus NaCl concentration (**Figure 4.15**) was generated with NaCl standard from 0 to 500 mM NaCl, in 50 mM increments. These standards were made by mixing the appropriate volumes of 20 mM Tris, pH 7.6 buffer and 0 mM NaCl with the same buffer containing 1000 mM NaCl. Using a polynomial fit to the data, the NaCl concentration in any fraction could be estimated and a plot of the gradient could be generated (**Figure 4.16**).

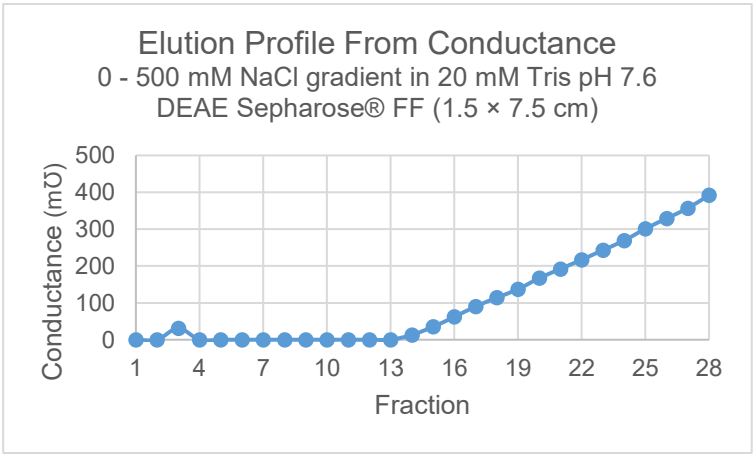

Figure 4.16. Conductance of anion exchange chromatography fractions.

Size Exclusion Chromatography

Size exclusion chromatography was performed with GE Healthcare Sephacryl® S-200 High Resolution resin, which was loaded onto two Bio-Rad Econo-Columns®. One column was a 1.5 × 30.0 cm column filled to a height of 24.5 cm (43.30 mL bed volume) and the other was a 1.5 × 75.0 cm column filled to a height of 62.0 cm (109.56 mL bed volume). Both columns were fitted with a Bio-Rad Econo-Column® Flow Adaptor that was used to pack the column as well as load and elute the sample. The adaptor was connected through tubing to a Bio-Rad Model EP-1 Econo Pump. A T-valve with Luer-Lok™ fittings was installed halfway between the adaptor and the pump to interrupt the mobile phase and inject the sample directly onto resin, via the flow adaptor. Fractions from the column were collected in VWR disposable borosilicate glass culture tubes (13 × 100 mm) that were placed into a Bio-Rad Model 2110 Fraction Collector.

In a refrigerated room, the resin was equilibrated with 2 column volumes of the elution buffer at a flow rate of 1.0 mL/min (35 cm/hr), which consisted of 20 mM Tris, pH 7.6, and 150 mM NaCl. The sample was loaded slowly onto the column with a syringe that was attached to a Luer-Lok port on the T-valve in the direction that led to the resin. The pump was started, and the column was eluted with 1 column volume of elution buffer. The flow rate for the 24.5 cm column was set to 0.50 mL/min (17 cm/hr) and 1.2 mL fractions were collected. The flow rate for the 62.0 cm column was set to 0.45 mL/min (15.2 cm/hr) and 2.0 mL fractions were collected. The resin was regenerated with 1 column volume of 20 mM Tris, pH 7.6, 1 M NaCl. Occasional cleaning in place of the column was performed by flowing 0.3 M NaOH through the column at 15–20 cm/h for 1 to 2 hours.

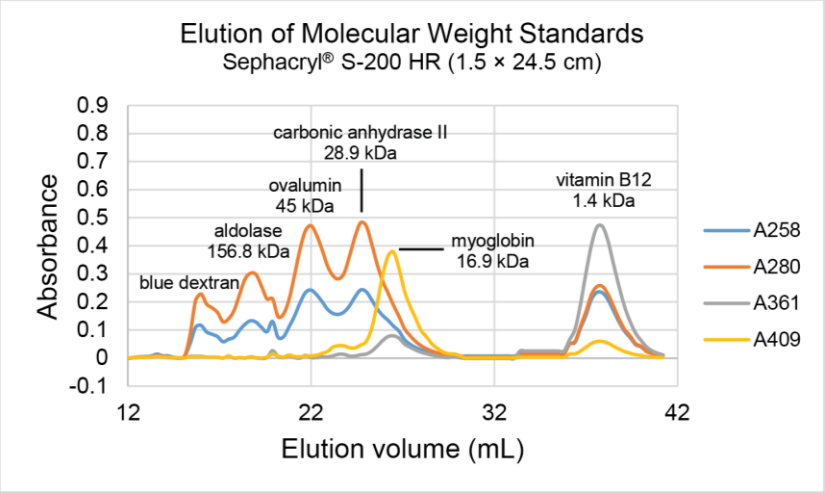

Figure 4.17. Elution of molecular weight standards from 24.5 cm SEC column.

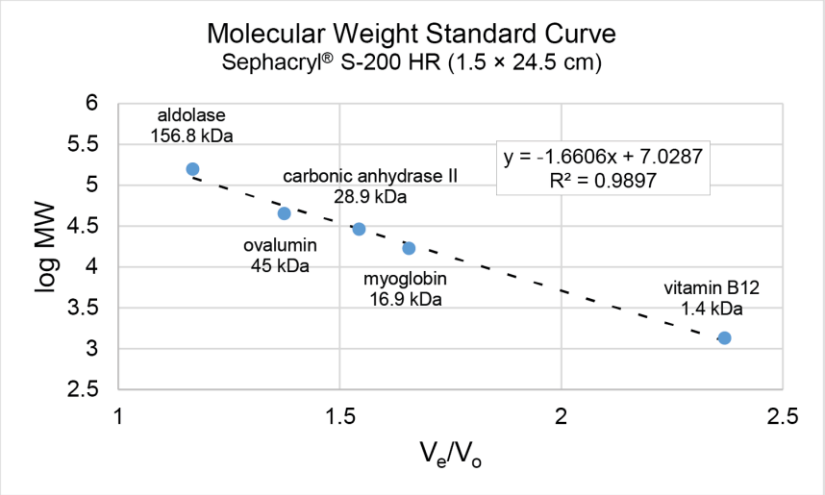

Figure 4.18. Molecular weight standard curve for 24.5 cm SEC column.

Both the 24.5 cm and 62.0 cm columns were calibrated using size exclusion molecular weight standards, which included blue dextran, aldolase, ovalbumin, carbonic anhydrase, myoglobin and vitamin B12. The elution profile for these standards on the 24.5 cm column is shown in **Figure 4.17**. A standard curve was generated using the log of the molecular weight for each standard, the peak elution volume (V_e) for each standard, and the void volume (V_0) for the column as determined by blue dextran. The molecular weight standard curve for the 24.5 cm column is shown in **Figure 4.18**.

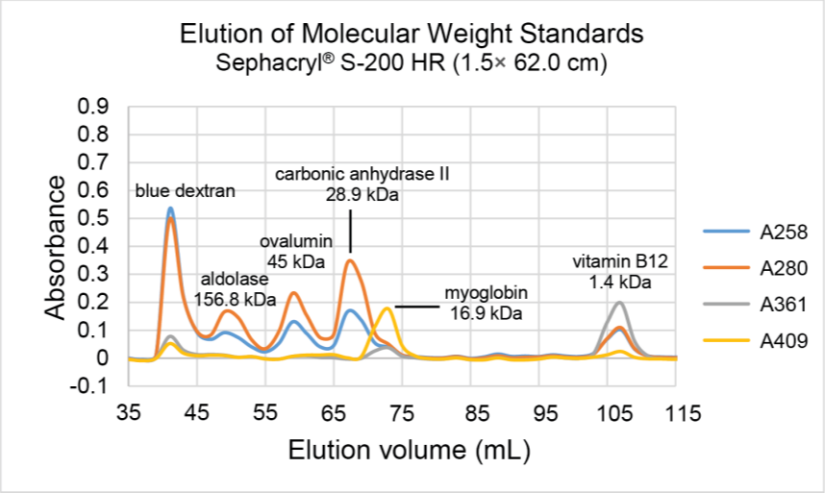

Figure 4.19. Elution of molecular weights from 62.0 cm SEC column.

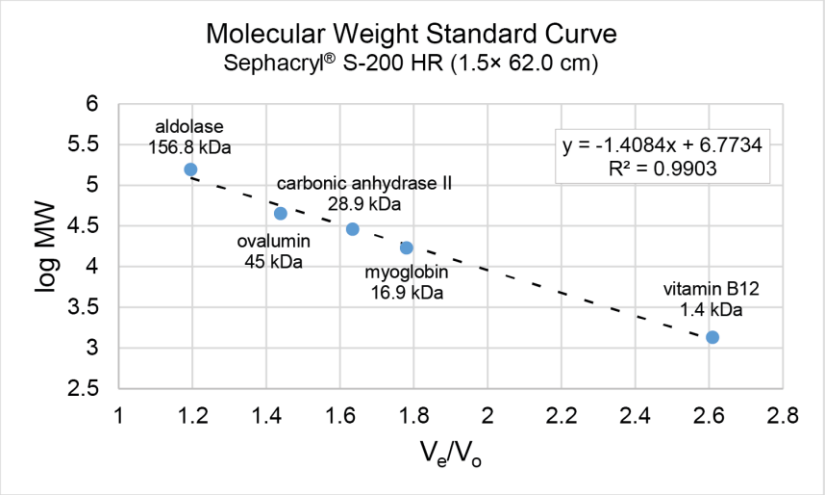

Figure 4.20. Molecular weight standard curve for 62.0 cm SEC column.

The increase in resolution afforded by the 62.0 cm column, using the same set of standards, is evident in the elution profile for this column (**Figure 4.19**). The molecular weight standard curve for the standards on the 62.0 cm column is shown in **Figure 4.20**.

Centrifugal Concentration

For chromatography fractions that were ≤ 6 mL, desalting and/or concentration was performed with a Sartorius Vivaspin® 6 Centrifugal Concentrator (30 kDa MWCO) at 6,000

× *g* and 4 °C in a Beckman JA-20 rotor mounted in a Beckman J2-HS (and J2-21) centrifuge. For chromatography fractions that were ≥ 6 mL, desalting and/or concentration was performed with an Amicon® Ultra-15 30kDa MWCO centrifugal filter, spun at 5000 × *g* at 4 °C in a Beckman JA-10 rotor mounted in a Beckman J2-HS (and J2-21) centrifuge. When desalting or facilitating a buffer exchange, the spin was repeated after the addition of the new buffer, with this exchange being performed a total of two times.

4.4.3 Protein Analysis

Protein Quantitation

The dye-based colorimetric method of Bradford[61] was used to determine the concentration of total protein in the lysate and column fractions. Buffer components are known to interfere[62] with the Bradford assay, detergents[63] in particular. Bio-Rad's Protein Assay Dye Reagent Concentrate (500-0006) states it can be used reliably with up to 0.1% Triton X-100, which was the concentration of Triton used in the lysis buffer. The Bio-Rad reagent is also linear between 0.2 and 0.9 mg/mL bovine serum albumin (BSA). The BSA concentrations used for the standard curve were 0 (100 mM NaCl), 0.2, 0.4, 0.6, 0.8 and 0.9 mg/mL BSA. The standards were made by dissolving BSA (Fraction V) Heat Shock Treated (FisherBioReagents, BP1600-100) in filter sterilized 100 mM NaCl. In a plastic cuvette (1 cm diameter), 20 µL unknown sample (or standard) were mixed with 980 µL of 1X dye reagent and incubating at room temperature for 15 minutes, then the absorbance at 595 nm was read. Uknown protein concentrations were determined using the BSA standard curve.

SDS-PAGE

All SDS-PAGE gels used were Invitrogen™ Bolt™ 4-12% Bis-Tris Plus 1.0 mm × 15-well Mini gels (NW04125BOX). Gels were loaded on a Novex® Bolt® Mini Gel Tank (A25977) using Invitrogen™ Novex™ Bolt™ MOPS SDS Running Buffer (B0001). The visible ladder used was Goldbio® BLUEstain™ Protein Ladder, 11-245 kDa (P007-500), loaded at 3 µL/well, while each protein sample was loaded at 15 µL/well. Protein samples were prepared using 4X Bolt™ LDS Sample Buffer (B0007) and the manufacturer's instructions. The gels were run at a constant 150 volts for 30–35 minutes, with power supplied by a BioRad PowerPac™ 3000 (165-5056). Gels were removed from their cassettes, rinsed twice for 30 seconds with 18 MΩ water, and placed into Coomassie stain overnight. The Coomassie stain was 0.1% Coomassie BB R-250, 50% methanol, 40% 18 MΩ water, and 10% glacial acetic acid. The gels were destained the following day with Destain 1 (53% 18 MΩ water, 40% methanol, 7% glacial acetic acid) followed by Destain 2 (88% 18 MΩ water, 5% methanol, 7% glacial acetic acid). Gels were photographed on a lightbox using a Nikon D7000, the images processed using Adobe Photoshop CS6 and annotated using Adobe Illustrator CS6.

SDS-PAGE gels for Western blots were prepared and run the same way, with the exception that MagicMark™ XP Western Protein Standard (LC5602) was used as the ladder, at 2 µL/well. In addition, after removing gels from their cassettes and rinsing with 18 MΩ water, gels were not stained with Coomassie.

Western Blotting

After washing SDS-PAGE gels with 18 MΩ water, they were placed on top of the PDVF membrane of the electroblotting transfer stack. Both Invitrogen™ Novex™ iBlot™ Transfer Stack, PVDF, regular (Ref: IB401001) and mini size (Ref: IB401002) were used. Regular size

supports the blotting of two mini gels simultaneously. The Invitrogen™ iBlot™ Gel Transfer Device (IB1001) was used to transfer the proteins into the PDVF membrane by electroblotting at 20 volts for 7 minutes. The gel was discarded and Thermo Scientific™ Ponceau S Staining Solution (A40000279) was used to evaluate the membrane for transfer efficiency. The stain was quickly reversed by incubating it for 2 minutes in 0.1% NaOH, followed by a rinse with 18 MΩ water.

To prevent non-specific binding of the antibodies to the PDVF membrane, they were blocked at room temperature by placing them in a 10 mL blocking buffer for 1 hour with gentle rocking. The blocking buffer was 5% BSA in TBST, pH 7.5 (sterile filtered). The blocking buffer was decanted and replaced with 10 mL of the primary antibody solution. The primary antibody solution was 1:500 dilution of antibody in the blocking buffer. For HA-tag detection the primary antibody was polyclonal anti-HA-TAG IgG produced in rabbit (Sigma-Aldrich™; Product number: SAB4300603). For detection of HOT, the primary antibody was polyclonal anti-ADHFe1 IgG produced in rabbit (Sigma-Aldrich™; Product number: SAB1410235). The product documentation of this specific Sigma-Aldrich anti-ADHFe1 IgG lists the immunogen as ADHFE1 (NP_653251.1, ~ 419 a.a.) full-length human HOT protein, therefore the precise nature of the epitope is unknown.. Initial anti-ADHFe1 IgG was obtained from Santa Cruz Biotechnology (item number, clone and immunogen unknown). After adding the primary antibody solution, the blot was incubated overnight at 4 °C with gentle rocking. The following day the primary antibody solution was transferred to a sterile 15 mL Falcon tube and placed in the refrigerator for reuse, since it was possible to use antibody solutions more than once, depending on the circumstances. The blot was washed 4 times for 5 minutes each with TBST, pH 7.5.

After washing TBST, the blot was incubated in 10 mL of secondary antibody solution for 1 hour at room temperature. From this point forward, the container with the membrane was kept wrapped in aluminum foil. The secondary antibody solution was 1:500 dilution of Donkey anti-Rabbit IgG (H+L), DyLight® 488 Conjugate (ImmunoReagents, Inc.; Part number: DkxRb-003-D488NHSX) in blocking buffer. Following the incubation, the secondary antibody solution was transferred to a sterile 15 mL Falcon tube and refrigerated for reuse. The blot was washed 4 times for 5 minutes each with TBST, pH 7.5.

To image the antibody staining, the blot was placed on the Fluor stage of an Amersham™ Typhoon™ 5 Biomolecular Imager with Image Lab v6.1.0 software. The filter was set to Cy2 525BP20 and the blot was excited at 488 nm. The images processed (i.e., crop and contrast/brightness adjustments) with Image Lab, exported as BMP files and saved along with the other instrument specific experiment files. Further image adjustments were made in Adobe Photoshop CS6 and annotations were made with Adobe Illustrator CS6.

4.5 References

(1) Kardon, T.; Noël, G.; Vertommen, D.; Schaftingen, E. V. Identification of the gene encoding hydroxyacid-oxoacid transhydrogenase, an enzyme that metabolizes 4-hydroxybutyrate. *FEBS Lett* **2006**, *580* (9), 2347-2350. DOI: 10.1016/j.febslet.2006.02.082

(2) Kaufman, E. E.; Nelson, T. An overview of gamma-hydroxybutyrate catabolism: the role of the cytosolic NADP(+)-dependent oxidoreductase EC 1.1.1.19 and of a mitochondrial hydroxyacid-oxoacid transhydrogenase in the initial, rate-limiting step in this pathway. *Neurochem Res* **1991**, *16* (9), 965-974. DOI: 10.1007/bf00965839

(3) Kaufman, E. E.; Nelson, T.; Fales, H. M. Isolation and Characterization of a Mammalian Hydroxyacid-Oxoacid Transhydrogenase. In *Biological Oxidation Systems*, Reddy, C. C., Hamilton, G. A., Madyastha, K. M. Eds.; Academic Press, 1990; pp 1059-1069.

(4) Kaufman, E. E.; Nelson, T.; Fales, H. M.; Levin, D. M. Isolation and characterization of a hydroxyacid-oxoacid transhydrogenase from rat kidney mitochondria. *J Biol Chem* **1988**, *263* (32), 16872-16879.

(5) Kaufman, E. E.; Nelson, T.; Miller, D.; Stadlan, N. Oxidation of gamma-hydroxybutyrate to succinic semialdehyde by a mitochondrial pyridine nucleotide-independent enzyme. *J Neurochem* **1988**, *51* (4), 1079-1084. DOI: 10.1111/j.1471-4159.1988.tb03071.x

(6) Liesenfeld, D. B.; Botma, A.; Habermann, N.; Toth, R.; Weigel, C.; Popanda, O.; Klika, K. D.; Potter, J. D.; Lampe, J. W.; Ulrich, C. M. Aspirin Reduces Plasma Concentrations of the Oncometabolite 2-Hydroxyglutarate: Results of a Randomized, Double-Blind, Crossover Trial. *Cancer Epidemiol Biomarkers Prev* **2016**, *25* (1), 180-187. DOI: 10.1158/1055-9965.Epi-15-0697

(7) Lyon, R. C.; Johnston, S. M.; Panopoulos, A.; Alzeer, S.; McGarvie, G.; Ellis, E. M. Enzymes involved in the metabolism of gamma-hydroxybutyrate in SH-SY5Y cells: identification of an iron-dependent alcohol dehydrogenase ADHFe1. *Chem Biol Interact* **2009**, *178* (1-3), 283-287. DOI: 10.1016/j.cbi.2008.10.025

(8) Struys, E. A.; Verhoeven, N. M.; Ten Brink, H. J.; Wickenhagen, W. V.; Gibson, K. M.; Jakobs, C. Kinetic characterization of human hydroxyacid-oxoacid transhydrogenase: relevance to D-2-hydroxyglutaric and gamma-hydroxybutyric acidurias. *J Inherit Metab Dis* **2005**, *28* (6), 921-930. DOI: 10.1007/s10545-005-0114-x

(9) Kim, J. Y.; Tillison, K. S.; Zhou, S.; Lee, J. H.; Smas, C. M. Differentiation-dependent expression of Adhfe1 in adipogenesis. *Arch Biochem Biophys* **2007**, *464* (1), 100-111. DOI: 10.1016/j.abb.2007.04.018

(10) Panopoulos, A. The Role of a Hydroxyacid-Oxoacid Transhydrogenase in GHB Metabolism. University of Strathclyde, Glasgow, Scotland, 2007.

(11) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **2005**, *44* (45), 7342-7372. DOI: 10.1002/anie.200501023

(12) Rardin, M. J.; Newman, J. C.; Held, J. M.; Cusack, M. P.; Sorensen, D. J.; Li, B.; Schilling, B.; Mooney, S. D.; Kahn, C. R.; Verdin, E.; et al. Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proc Natl Acad Sci U S A* **2013**, *110* (16), 6601-6606. DOI: 10.1073/pnas.1302961110

(13) Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villén, J.; Haas, W.; Sowa, M. E.; Gygi, S. P. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **2010**, *143* (7), 1174-1189. DOI: 10.1016/j.cell.2010.12.001

(14) Chacinska, A.; Koehler, C. M.; Milenkovic, D.; Lithgow, T.; Pfanner, N. Importing mitochondrial proteins: machineries and mechanisms. *Cell* **2009**, *138* (4), 628-644. DOI: 10.1016/j.cell.2009.08.005

(15) Wiedemann, N.; Pfanner, N. Mitochondrial Machineries for Protein Import and Assembly. *Annu Rev Biochem* **2017**, *86*, 685-714. DOI: 10.1146/annurev-biochem-060815-014352

(16) Kunová, N.; Havalová, H.; Ondrovičová, G.; Stojkovičová, B.; Bauer, J. A.; Bauerová-Hlinková, V.; Pevala, V.; Kutejová, E. Mitochondrial Processing Peptidases-Structure, Function and the Role in Human Diseases. *Int J Mol Sci* **2022**, *23* (3). DOI: 10.3390/ijms23031297

(17) Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **1987**, *15* (20), 8125-8148. DOI: 10.1093/nar/15.20.8125

(18) Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A* **1990**, *87* (21), 8301-8305. DOI: 10.1073/pnas.87.21.8301

(19) Okazaki, Y.; Furuno, M.; Kasukawa, T.; Adachi, J.; Bono, H.; Kondo, S.; Nikaido, I.; Osato, N.; Saito, R.; Suzuki, H.; et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **2002**, *420* (6915), 563-573. DOI: 10.1038/nature01266

(20) Deng, Y.; Wang, Z.; Gu, S.; Ji, C.; Ying, K.; Xie, Y.; Mao, Y. Cloning and characterization of a novel human alcohol dehydrogenase gene (ADHFe1). *DNA Seq* **2002**, *13* (5), 301-306. DOI: 10.1080/1042517021000011636

(21) Käll, L.; Krogh, A.; Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **2004**, *338* (5), 1027-1036. DOI: 10.1016/j.jmb.2004.03.016

(22) Käll, L.; Krogh, A.; Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W429-432. DOI: 10.1093/nar/gkm256

(23) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023**, *51* (D1), D523-d531. DOI: 10.1093/nar/gkac1052

(24) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2

(25) Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **2024**, *52* (D1), D368-d375. DOI: 10.1093/nar/gkad1011

(26) Ryan, D. G.; Bridger, W. A. Bacterial expression of rat liver succinyl-CoA synthetase alpha-subunit. Factors that contribute to blocked translation of transcripts encoding a mitochondrial signal sequence. *J Mol Biol* **1991**, *219* (2), 165-174. DOI: 10.1016/0022-2836(91)90559-o

(27) Schlachter, C. R.; Klapper, V.; Radford, T.; Chruszcz, M. Comparative studies of Aspergillus fumigatus 2-methylcitrate synthase and human citrate synthase. *Biol Chem* **2019**, *400* (12), 1567-1581. DOI: 10.1515/hsz-2019-0106

(28) Kozak, M. An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* **1991**, *115* (4), 887-903. DOI: 10.1083/jcb.115.4.887

(29) Gluzman, Y. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **1981**, *23* (1), 175-182. DOI: 10.1016/0092-8674(81)90282-8

(30) Biedler, J. L.; Roffler-Tarlov, S.; Schachner, M.; Freedman, L. S. Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. *Cancer Res* **1978**, *38* (11 Pt 1), 3751-3757.

(31) Greene, L. A.; Tischler, A. S. Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc Natl Acad Sci U S A* **1976**, *73* (7), 2424-2428. DOI: 10.1073/pnas.73.7.2424

(32) Wiatrak, B.; Kubis-Kubiak, A.; Piwowar, A.; Barg, E. PC12 Cell Line: Cell Types, Coating of Culture Vessels, Differentiation and Other Culture Conditions. *Cells* **2020**, *9* (4). DOI: 10.3390/cells9040958

(33) Dyson, M. R. Fundamentals of Expression in Mammalian Cells. *Adv Exp Med Biol* **2016**, *896*, 217-224. DOI: 10.1007/978-3-319-27216-0_14

(34) O'Flaherty, R.; Bergin, A.; Flampouri, E.; Mota, L. M.; Obaidi, I.; Quigley, A.; Xie, Y.; Butler, M. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. *Biotechnol Adv* **2020**, *43*, 107552. DOI: 10.1016/j.biotechadv.2020.107552

(35) Hunter, M.; Yuan, P.; Vavilala, D.; Fox, M. Optimization of Protein Expression in Mammalian Cells. *Curr Protoc Protein Sci* **2019**, *95* (1), e77. DOI: 10.1002/cpps.77

(36) Tihanyi, B.; Nyitray, L. Recent advances in CHO cell line development for recombinant protein production. *Drug Discov Today Technol* **2020**, *38*, 25-34. DOI: 10.1016/j.ddtec.2021.02.003

(37) Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **2004**, *22* (11), 1393-1398. DOI: 10.1038/nbt1026

(38) Makrides, S. C. Components of vectors for gene transfer and expression in mammalian cells. *Protein Expr Purif* **1999**, *17* (2), 183-202. DOI: 10.1006/prep.1999.1137

(39) Nora, L. C.; Westmann, C. A.; Martins-Santana, L.; Alves, L. F.; Monteiro, L. M. O.; Guazzaroni, M. E.; Silva-Rocha, R. The art of vector engineering: towards the construction of next-generation genetic tools. *Microb Biotechnol* **2019**, *12* (1), 125-147. DOI: 10.1111/1751-7915.13318

(40) Van Craenenbroeck, K.; Vanhoenacker, P.; Haegeman, G. Episomal vectors for gene expression in mammalian cells. *Eur J Biochem* **2000**, *267* (18), 5665-5678. DOI: 10.1046/j.1432-1327.2000.01645.x

(41) Kallunki, T.; Barisic, M.; Jäättelä, M.; Liu, B. How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* **2019**, *8* (8). DOI: 10.3390/cells8080796

(42) Kim, T. K.; Eberwine, J. H. Mammalian cell transfection: the present and the future. *Anal Bioanal Chem* **2010**, *397* (8), 3173-3178. DOI: 10.1007/s00216-010-3821-6

(43) Hawley-Nelson, P.; Ciccarone, V. Transfection of cultured eukaryotic cells using cationic lipid reagents. *Curr Protoc Cell Biol* **2003**, *Chapter 20*, Unit 20.26. DOI: 10.1002/0471143030.cb2006s19

(44) Mark, J. K. K.; Lim, C. S. Y.; Nordin, F.; Tye, G. J. Expression of mammalian proteins for diagnostics and therapeutics: a review. *Mol Biol Rep* **2022**, *49* (11), 10593-10608. DOI: 10.1007/s11033-022-07651-3

(45) Mortensen, R. M.; Kingston, R. E. Selection of transfected mammalian cells. *Curr Protoc Mol Biol* **2009**, *Chapter 9*, Unit9.5. DOI: 10.1002/0471142727.mb0905s86

(46) Würtele, H.; Little, K. C.; Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene Ther* **2003**, *10* (21), 1791-1799. DOI: 10.1038/sj.gt.3302074

(47) Cabrera, A.; Edelstein, H. I.; Glykofrydis, F.; Love, K. S.; Palacios, S.; Tycko, J.; Zhang, M.; Lensch, S.; Shields, C. E.; Livingston, M.; et al. The sound of silence: Transgene silencing in mammalian cell engineering. *Cell Syst* **2022**, *13* (12), 950-973. DOI: 10.1016/j.cels.2022.11.005

(48) O'Gorman, S.; Fox, D. T.; Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **1991**, *251* (4999), 1351-1355. DOI: 10.1126/science.1900642

(49) Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **2012**, *109* (39), E2579-2586. DOI: 10.1073/pnas.1208507109

(50) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337* (6096), 816-821. DOI: 10.1126/science.1225829

(51) Kumar, P.; Nagarajan, A.; Uchil, P. D. Selective Agents for Stable Transfection. *Cold Spring Harb Protoc* **2018**, *2018* (9). DOI: 10.1101/pdb.top096230

(52) Colbère-Garapin, F.; Horodniceanu, F.; Kourilsky, P.; Garapin, A. C. A new dominant hybrid selective marker for higher eukaryotic cells. *J Mol Biol* **1981**, *150* (1), 1-14. DOI: 10.1016/0022-2836(81)90321-1

(53) Southern, P. J.; Berg, P. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J Mol Appl Genet* **1982**, *1* (4), 327-341.

(54) Wright, G. D.; Thompson, P. R. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Front Biosci* **1999**, *4*, D9-21. DOI: 10.2741/wright

(55) Bar-Nun, S.; Shneyour, Y.; Beckmann, J. S. G-418, an elongation inhibitor of 80 S ribosomes. *Biochim Biophys Acta* **1983**, *741* (1), 123-127. DOI: 10.1016/0167-4781(83)90018-0

(56) Stepanenko, A. A.; Heng, H. H. Transient and stable vector transfection: Pitfalls, off-target effects, artifacts. *Mutat Res Rev Mutat Res* **2017**, *773*, 91-103. DOI: 10.1016/j.mrrev.2017.05.002

(57) Dalby, B.; Cates, S.; Harris, A.; Ohki, E. C.; Tilkins, M. L.; Price, P. J.; Ciccarone, V. C. Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. *Methods* **2004**, *33* (2), 95-103. DOI: 10.1016/j.ymeth.2003.11.023

(58) Reardon, S. US government issues historic $3.5-million fine over animal welfare. *Nature* **2016**. DOI: 10.1038/nature.2016.19958

(59) Heeneman, S.; Deutz, N. E.; Buurman, W. A. The concentrations of glutamine and ammonia in commercially available cell culture media. *J Immunol Methods* **1993**, *166* (1), 85-91. DOI: 10.1016/0022-1759(93)90331-z

(60) Tritsch, G. L.; Moore, G. E. Spontaneous decomposition of glutamine in cell culture media. *Exp Cell Res* **1962**, *28*, 360-364. DOI: 10.1016/0014-4827(62)90290-2

(61) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **1976**, *72*, 248-254. DOI: 10.1006/abio.1976.9999

(62) Compton, S. J.; Jones, C. G. Mechanism of dye response and interference in the Bradford protein assay. *Anal Biochem* **1985**, *151* (2), 369-374. DOI: 10.1016/0003-2697(85)90190-3

(63) Friedenauer, S.; Berlet, H. H. Sensitivity and variability of the Bradford protein assay in the presence of detergents. *Anal Biochem* **1989**, *178* (2), 263-268. DOI: 10.1016/0003-2697(89)90636-2