A Study of Mitochondrial Malate Dehydrogenase in *Gallus gallus* and Other Avian Species

Kendall Reidenbach

Submitted in Partial Fulfillment of the Requirements for Graduation from the Malone University Honors Program

Advisor: Kathryn Huisinga, Ph.D.

April 28, 2020

Abstract:

Malate dehydrogenase (MDH) functions as a catalyst for the NAD^+/NADH-dependent reversible reaction between malate and oxaloacetate. The mitochondrial form of the enzyme (MDH2) is important in the citric acid cycle, a key part of aerobic metabolism. Previous studies into avian MDH2 have focused on studying enzyme gel mobility between taxonomic families, activity differences between migratory and non-migratory species, and activity differences of a species at high versus low altitudes. Individual enzyme kinetics and structural data on the wild-type MDH2, however, are not documented. A cDNA library was utilized to obtain the gene for *Gallus gallus* (chicken). The Gibson Cloning assembly was used to insert the MDH2 gene into the pET28(a)+ expression vector for expression of the protein in *E. coli* cells. Initial experiments to test expression conditions indicated that codon optimization may be required. Additionally, protein modeling software was employed to predict the 3D structure of the *G. gallus* and other avian species mitochondrial MDH proteins. These comparisons not only give insight on how differences in the amino acid sequence affect the structure, but provide clues to what mutations to the *Gallus gallus* expression vector may mimic the MDH2 enzyme of another avian species providing future research direction.

Acknowledgements:

I'd like to thank my parents for encouraging me from a young age to develop a love for science and learning. In addition, I want to thank my family and friends for being supportive and for listening to me everytime that I went on a tangent about my thesis. Thank you to all the Math and Science faculty members who opened the lab doors for me. I am grateful for both Dr. Courter and Dr. Frazier for being members of my thesis committee and taking time out of their busy schedules to help make this possible. Lastly, I want to thank Dr. Huisinga for her help and guidance throughout the whole thesis process.

Table of Contents:

Chapter 1: Introduction

The class Aves encompasses all the species of birds. There are several obvious variations in both physical and behavioral characteristics between many birds. As a biochemist, the differences that involve biochemical pathways are of most interest. Metabolism is an ideal candidate for in vitro analysis. Given the various sizes and migration patterns of the different orders within Aves, there are differences in their energy requirements (Hudson, 2013). These differences in the metabolic needs of the various species provide several areas of study, whether the focus is on normal physiological differences or specific metabolic events such as migration.

To study these differences in energy requirements, a particular part of metabolism should be chosen for analysis. Glycolysis, the citric acid cycle, and oxidative phosphorylation are three potential candidates. Glycolysis is the anaerobic breakdown of glucose into pyruvate. The citric acid cycle is a key set of eight reactions utilized during aerobic metabolism. Oxidative phosphorylation is the generation of ATP (adenosine triphosphate) through the transfer of electrons (from NADH/$FADH_2$ to O_2). When the difficulty of research, accessibility, and the literature was considered, I selected the final step of the citric acid cycle as the place to study metabolic differences among birds. A single semester of research did not allow for all parts of metabolism to be thoroughly studied. Additionally, accessibility to reagents and the ease of experimental design played a huge role. The last step of the citric acid cycle fits these criteria (Figure 1). Future studies that extend beyond this single reaction would be expected to shed additional light on any potential metabolic differences within the Aves order.

The analysis in this thesis will specifically examine variations between the enzyme malate dehydrogenase. Malate dehydrogenase (MDH) functions as a catalyst for NAD+/NADH-dependent reversible redox reaction between malate and oxaloacetate (Figure 1).

The enzyme is highly conserved in structure and key sequences among species ranging from bacteria to humans (Goward, 1994). In prokaryotic cells, MDH has a single form. For some eukaryotes, there are at least three forms with the most commonly found two forms being cytoplasmic and mitochondrial. These two forms are similar, but not identical. They participate in different pathways in the cell. The cytoplasmic form is present in the aspartate-malate shuttle, among others. The mitochondrial form is active in the citric acid cycle. The third, less common form is found in the glyoxysomes of yeast and plants. The proteins are encoded by different genes and have primary structure variance yet they share many crucial structures. The enzyme is either found as a dimer or a tetramer, depending on the form and the organism. The stability and binding of the dimer and tetramer can affect the activity of MDH2. If the whole enzyme cannot form, it will not function. The active sites, the coenzyme binding sites, and quaternary structures are conserved. Secondary structural differences exist outside of these (Minarik, 2002).

Saccharomyces cerevisiae (baker's yeast), as well as in *Homo sapiens* and *Sus scrofa* (pig) MDH have been well studied. The mitochondrial 3D crystal structure is solved in both *S. cerevisiae* and humans, while both structures are solved in pigs (Taylor, 2001; Ugochukwu, N.D.; Birtoft, 1989; Roderick, 1986). The solved human structure in Figure 2 provides an example of what a solved model looks like. Additionally, the single prokaryotic MDH form has been researched in *Escherichia coli* and *Salmonella enterica*. The glyoxysomal version in *Citrullus lanatus* (watermelon) has also been studied significantly (Cox, 2005). These crystal structures can be difficult to generate, as the enzymes must undergo purification and crystallization. During this, proteins may degrade, be difficult to purify, or be time-consuming to crystallize. Due to the growing ease of sequencing, thousands of species have the sequences of both MDH forms known. This was not due to studies investigating MDH, but rather the result of

genome-wide sequencing projects. As a result, very little has been studied about the MDH forms and their structural information in many organisms, especially in avian species.

The few studies that have investigated the mobility and the activity of MDH in birds have focused on both intraspecies and interspecies differences. MDH has been studied comparing migratory versus non-migratory birds of a single species, the red-headed bunting (*Emberiza bruniceps*). The researchers found that the birds about to migrate have higher concentrations and activity levels of MDH compared to their nonmigratory counterparts (Banerjee, 2016). A higher activity means a higher rate of catalysis and more product. These were maximal enzyme activity values determined in a UV/Vis spectrophotometer, using an assay measuring the rate of oxidation of NADH to NAD+. MDH is more prominent and more active in pre-migratory tissues (Figure 3). Given the high energy requirements for migration, an increase in a metabolic enzyme is not entirely surprising (Banerjee). Another study has shown that altitude also plays a role in the activity level of MDH. Torrent ducks (*Merganetta armata*), a species living in the Andes Mountains, have a higher MDH activity when living at elevated altitudes compared to the individuals of the same species at lower altitudes (Dawson, 2016). It is speculated that the difference is due to an increased reliance on the malate-aspartate shuttle or the increased need for the regeneration of NAD+/NADH for other biochemical pathways. Similar to Banerjee, this study used an assay to measure the catalytic ability of MDH. The MDH catalyzes at a higher rate at higher altitudes in torrent ducks than at lower altitudes (Dawson, 2016). Additionally, the enzyme's size and mobility have been compared between various avian orders using native gel electrophoresis. In this method, proteins will travel further on the gel, depending on their charge and size (Arndt, 2012). Although they did not quantitatively estimate the enzyme size, the electrophoresis data produced a trend that more advanced families such as Psittacidae (parrots)

and Corvidae (crows, jays, and ravens) have faster gel mobility than the less intelligent families (Kakizawa, 1982). These studies are useful for highlighting the presence of differences between the enzymes, such as charge and activity level of the enzymes, but there is a limited explanation for the cause at the molecular level. Additionally, individual information about each MDH form is still missing. There are still informational gaps about the enzyme kinetics and enzymatic activity of wild-type MDHs. Furthermore, there is limited data about the three-dimensional structures of the enzyme. Researching both the kinetics and the structures can provide a new comparison tool for human MDHs. Beyond that, this data is important to study due to its ability to help improve understanding of avian metabolism and the potential to reveal new information about MDH. To conduct a study with this focus would require selecting a species that has a sequenced genome and easily accessible cDNAs to utilize for study.

Several avian species have fully sequenced genomes. Pigeons, ducks, and budgerigars have been successfully sequenced (Shapiro, 2013; Huang, 2013; Ganapathy, 2014). These three species, however, lack a cDNA library. One species that fulfills both criteria is *G. gallus* or the domesticated chicken. *G. gallus* has both a cytoplasmic (MDH1) and a mitochondrial (MDH2) form, but the focus of this study will be on the mitochondrial as no cDNA in the library contains the complete sequence for the cytoplasmic enzyme (Boardman, 2002). This thesis aims to gain a further understanding of the *G. gallus* MDH2 by investigating its enzyme kinetics and protein structure. As a starting point, this lays a foundation for future studies into the *G. gallus* MDH2. The gene can be mutated to test the effects of post-translational modifications on the enzyme's function. The wild-type kinetics can also be compared to other species, like humans, to identify key differences and potential explanations for the differences. The MDH2 protein must be

expressed in an organism like *E. coli* and purified as the cDNA itself does not produce the target protein and purifying the protein from the native source is not always possible.

Studying mitochondrial MDH in chickens can help to provide a better understanding of the enzyme and how well it is conserved between species beyond their amino acid sequences. Additionally, a deeper understanding of different MDH structures could provide insight into species-specific differences and enzyme region optimization for varying energy requirements. Beyond avian species, this type of research could eventually be beneficial to individuals that are ill. Studies in MDH2 in humans have shown that MDH2 mutants can cause encephalopathy (Ait-El-Mkadem, 2017). A better understanding of structure could show how these mutants are capable of causing harm.

Figures:

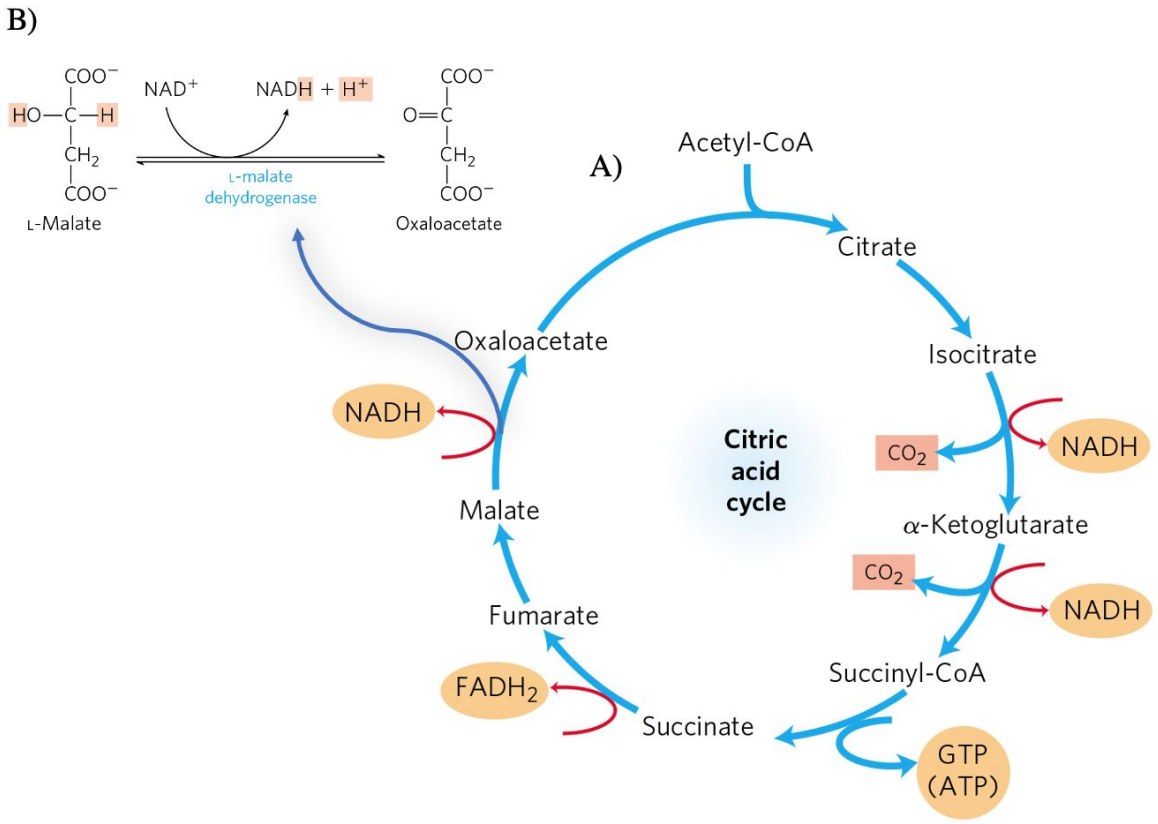


Figure 1: **Simplified Citric Acid Cycle and the Reaction between Malate and OAA**. A) The citric acid cycle and its intermediates generate electron carriers, carbon dioxide, and GTP with each turn. B) The reversible redox conversion of oxaloacetate and malate is catalyzed by malate dehydrogenase using NAD+/NADH as a cofactor.

Source: Lehninger, 2017.

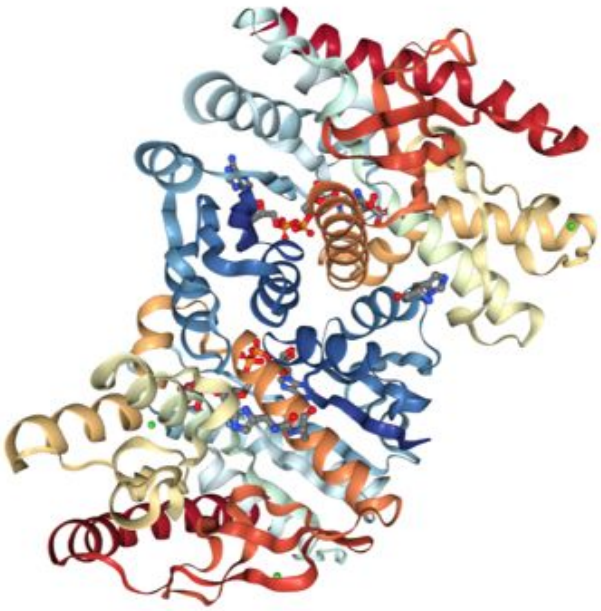


Figure 2: **Mitochondrial MDH crystal structure in *Homo sapiens*.** Using x-ray crystallography, three-dimensional protein structures like this one can be solved. Source: Ugochukwu, N.D.

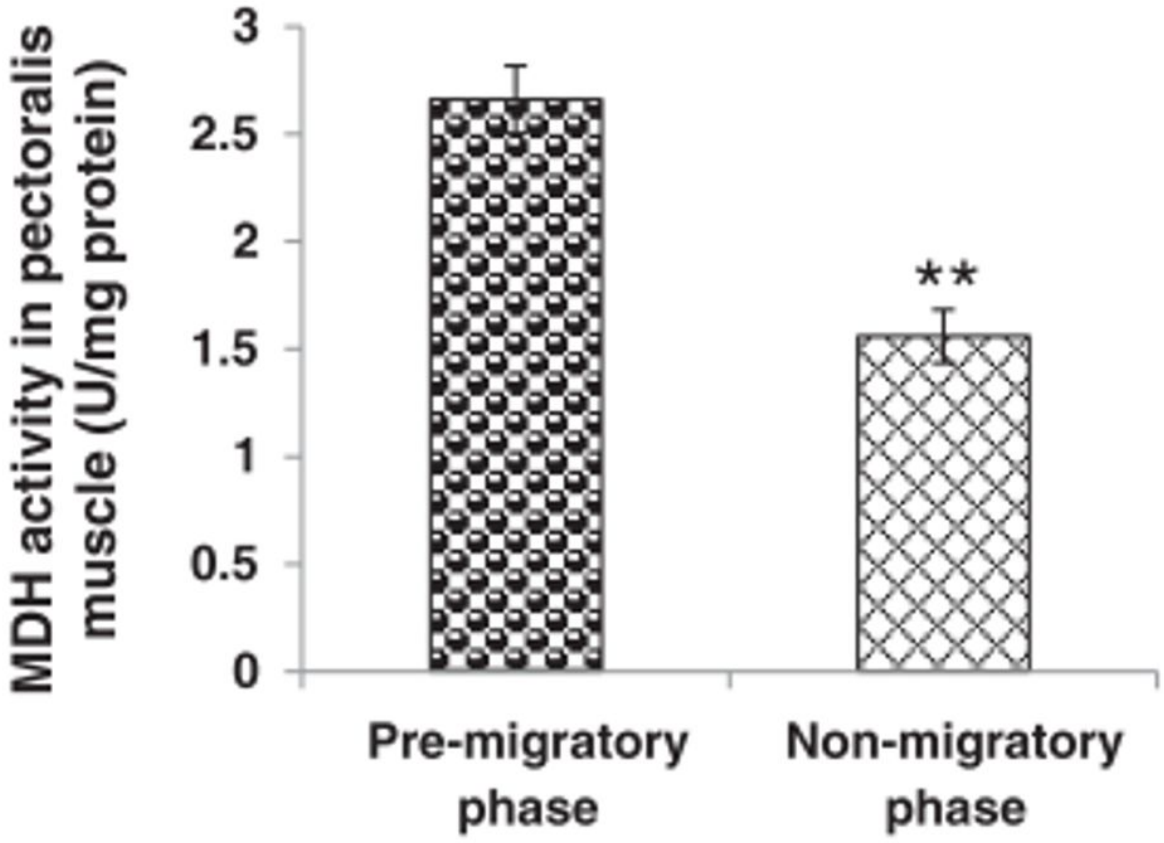


Figure 3: **Comparison of MDH activity in *Emberiza bruniceps*.** MDH activity is higher in the pectoralis muscles before migration than in muscles of non-migratory birds.

Source: Banerjee, 2016.

Chapter 2: Cloning a *Gallus gallus* MDH Expression Construct

Introduction:

Producing an expression vector requires several steps. The gene received from the cDNA library created by Boardman (2002) was not ready for protein expression. The MDH2 gene from the cDNA must be isolated and cloned from the storage vector. Additionally, the pET28(a)+ expression vector backbone requires cloning as well. These cloned pieces can then be combined. Gibson cloning assembly was utilized to combine these two pieces of DNA. The PCR primers used have a portion that anneals to the template and a portion that doesn't but is homologous to the other template (Figure 4). The DNA fragment (MDH2 gene) and the recipient plasmid (pET28(a)+) undergo a polymerase chain reaction (PCR) to create 2 products with overlapping regions.

These overlapping regions are cut by a T5 exonuclease to form complementary sticky ends when running the Gibson cloning assembly. These then anneal and the nucleotides are filled in with ligation.

Materials and Methods:

Before running, four PCR primers were designed using the Gibson Cloning Assembly feature on Benchling. The four primer sequences are

5'-CGGGCGAGGCGGGACAGCATGGTA

TATCTCCTTCTTAAAGTTAAACAAAATTATTTCT-3’ (pET28(a)+ FWD),

5'-ATTTTGCGAA GAACTTCAAGCACCACCACCACCACCACTG-3' (pET28(a)+ REV),

5'-CAGTGGTGGT GGTGGTGGTGCTTGAAGTTCTTCGCAAAATCCTCTCCTTTCTTAATAGA-3' (MDH2 FWD), 5'-CTTTAAGAAGGAGATATACCATGCTGTCCCGCCTCGCCCG-3' (MDH2 REV). The primers were ordered from IDT. These primers lie on the border between the MDH2 insert and the pET28(a)+ backbone and further details about the construct can be found in a cartoon diagram (Figure 5).

Several PCR reactions were performed with all the reactions using the same general concentrations. PCR reaction buffer final concentration was 1X, the forward and reverse primers were 0.2 μM, the dNTPs were at a final concentration of 0.2 mM, and the Sigma-Aldrich REDTaq DNA Pol was at 0.5 U/μL to a final volume of 50 μL. The DNA added varied with each reaction. The PCR reactions for the MDH2 insert were conducted with the following cycling conditions: 5 minutes at 94℃ and 1 minute at 72℃ during the pre-PCR phase; 30 cycles of denaturation at 94℃ for 30 seconds, annealing for 1 minute at 51℃ with the temperature increasing by 0.3℃ to reach 60℃, and extension for 2 minutes at 72℃; and finally hold for 5 minutes at 72℃ and hold at 4℃ until removed (Figure 6a).

After amplification of the MDH2 PCR products, they were purified using Qiagen's Qiaquick PCR purification kit following the manufacturer's provided protocol. The PCR reactions for the pET28(a)+ backbone were conducted with the following cycling conditions: 5 minutes at 94℃ and 1 minute at 72℃ during the pre-PCR phase; 30 cycles of denaturation at 94℃ for 30 seconds, annealing for 1 minute at 51℃ with the temperature increasing by 0.3℃ to reach 60℃, and extension for 5 minutes and 30 seconds at 72℃; and finally hold for 5 minutes at 72℃ and hold at 4℃ until removed (Figure 6b). Two more trials were run for the pET28(a)+, with the only change being to the annealing temperature. The first used a set annealing

temperature of 54 ℃. The pET28(a)+ PCR products did not produce a single band during analysis so the product was excised from the gel and purified using the Qiagen gel purification kit.

The Gibson Cloning Assembly Kit (NEB) was used to join the PCR products to form the expression construct. The reaction was set up following the manufacturer's guidelines with a 2:1 ratio of the DNA insert (MDH2) to vector backbone (pET28(a)+) as recommended by NEB (See Table 1). Additional trials followed using 1:1, 2:1, and 3:1 ratios utilizing a quantified combination of all the PCR tubes. The calculator provided by NEB online (http://nebiocalculator.neb.com/#!/ligation) was used to determine the ideal amount of insert to vector required to fall in the optimal cloning efficiency range of 50-100 ng. The DNA samples were then diluted to match the amount of DNA calculated. The experimental DH5α transformation was plated on LB+Kan (50 μg/mL) plates while the positive control was plated on LB+Amp (100 μg/mL) using 50 μL, 150 μL, and the rest of the transformed cells. The plates were incubated overnight at 37℃.

A subset of colonies were sampled from successful transformations to analyze their DNA. These were prepared using boiling lysis. The DNA was quantified and a restriction digest with EcoRI was performed using about 1 μg of DNA and 20 units of EcoRI to check the identity of the products.

After preliminary analyses of the first set of minipreps depleted the samples, 12 new minipreps were made using 12 new colonies from the second assembly plates as before. These minipreps were prepared using Qiagen miniprep spin columns. A double restriction digest was conducted with HindIII-HF and ApaI to check the identity of the plasmids. About 1 μg of DNA

was used with 50 units of ApaI and 20 units of HindIII-HF for each reaction. These restriction sites were selected because one is located outside the insert region while the other is found in both the parent vector and the DNA insert. The original miniprep from pET28(a)+ was also digested as a control for comparison. Any samples that appeared to have the correct construct were sent out to Genewiz for sequencing using the T7 and T7 terminator primers.

Three samples from the 24 minipreps were selected for transformation into BL21 *E. coli* cells. The NEB BL21 (DE3) Competent Cells and protocol C2527 were used. Using serial dilutions, the transformed cells and negative control were plated on 5 LB+Kan plates each and incubated overnight at 37℃. The samples were sent out for sequencing using the T7 and T7 terminator primers. These primers lie on each side of the MDH2 insert region and sequencing tests for the proper insertion of the MDH2 fragment during the Gibson Cloning Assembly.

Glycerol freezer stocks for both expression strains were generated for the correct *G. gallus* MDH2 clone. One DH5α freezer stock and 3 BL21 freezer stocks were produced. The first freezer stock of the BL21 cells, FS1, was then used to streak out a LB+Kan plate to be used for protein expression.

Results and Conclusions:

To create the *G. gallus* expression vector using Gibson Assembly Cloning, two PCR products needed to be generated. One product is the MDH2 gene insert. The protocol for amplifying the MDH2 insert was effective, producing the desired product when visualized on a 1% agarose gel (Figure 8). The MDH2 insert is expected to be 1011 base pairs in length, which matches with the size of the fragment observed in lanes 3 and 4 on this gel. This PCR product

was purified and ran on a 1% gel which confirmed the fragment had no free nucleotides (Figure 9).

The second PCR fragment needed for the Gibson Cloning method is the pET28(a)+ vector backbone. The second trial for pET28(a)+ yielded the desired product, but a competitive byproduct was present when ran on a 1% agarose gel. A faint band can be seen below 1 kB. The desired product is just over 5 kB in length which matches the top fragment in lanes 3 and 4 (Figure 9). Due to the presence of a competitive byproduct, the annealing temperature of the PCR cycling conditions was altered. Changing to 54℃ annealing temperature was designed to encourage annealing to the desired region and not the competitive site. This 54 ℃ trial yielded the same result as before. The top band was the desired product just above 5 kB, while the competitive product was at 1kB. To purify this product, gel excision and purification were used to isolate the pET28(a)+ PCR product found in the top fragment band (Figure 10).

The 2:1 Gibson cloning assembly conducted first yielded no experimental colonies after overnight incubation. The positive control had 7 colonies on the 50 µL plate, 10 on the 150 µL plate, and 16 on the "remainder" plate under the same incubation conditions. This suggests that the assembly reagents were sufficient and the problem was with the experimental samples.

The experimental PCR products were combined into a single tube and quantified before running a second set of Gibson Assembly trials. These new trials were conducted with three ratios of insert (MDH2) to vector (pET28(a)+)— 1:1, 2:1, and 3:1— to ensure that ratios were not the cause of the failure in the previous trial. Every plate from the second trial of Gibson Cloning Assembly had colonies (Table 2). All aspects of the experimental design including transformation were successful.

A subset of colonies transformed from the Gibson cloning assembly was selected for analysis to check the identity of the plasmid. A digest with EcoRI was chosen. The digest bands from the EcoRI restriction digest appeared to be at the appropriate location if the cloning assembly had worked (Figure 11). The sequencing data showed otherwise. The samples did not align with the assembled expression vector but with the pET28a+ parent sequence. An EcoRI site exists in the parent vector and the predicted digest results in similar sizes, making it possible that the digest bands were from the parent vector.

New minipreps were made due to low volume and digested using HindIII-HF and ApaI. The digest of a construct with the MDH2 insert will have two bands, one just below 5 kB and one just around 1. 6 kB. The digest of the pET28(a)+ parent vector will have a band just above 4 kB and a second band just above 1 kB (Figure 14). The test revealed that one sample, 1.1.8, appeared to have the MDH2 insert. It matches the virtual digest as the second band is higher on the gel than all the other samples, including the control pET-28(a)+ (Figure 15). After sequencing data was analyzed, sample 1.1.8 aligned with the assembled expression vector (Figure 16).

After the successful transformation into DH5α cells, Samples 1.1.4 and 2.1.2 were successfully transformed into BL21 cells with the number of colonies being found in Table 3. These samples were not used further because they did not contain the MDH2 insert gene. After confirmation from the HindIII-HF and ApaI digest, the sample 1.1.8 was successfully transformed into BL21 cells with the plates having over 100 colonies. The sample was stored as freezer stocks for future testing.

The Gibson Cloning Assembly at a 1:1 insert/backbone ratio yielded at least one colony with the correct insert and BL21 transformation was successful. Sample 1.1.8 is ready for protein

expression and induction tests. A chicken mitochondrial MDH expression vector has been successfully constructed.

Figures:

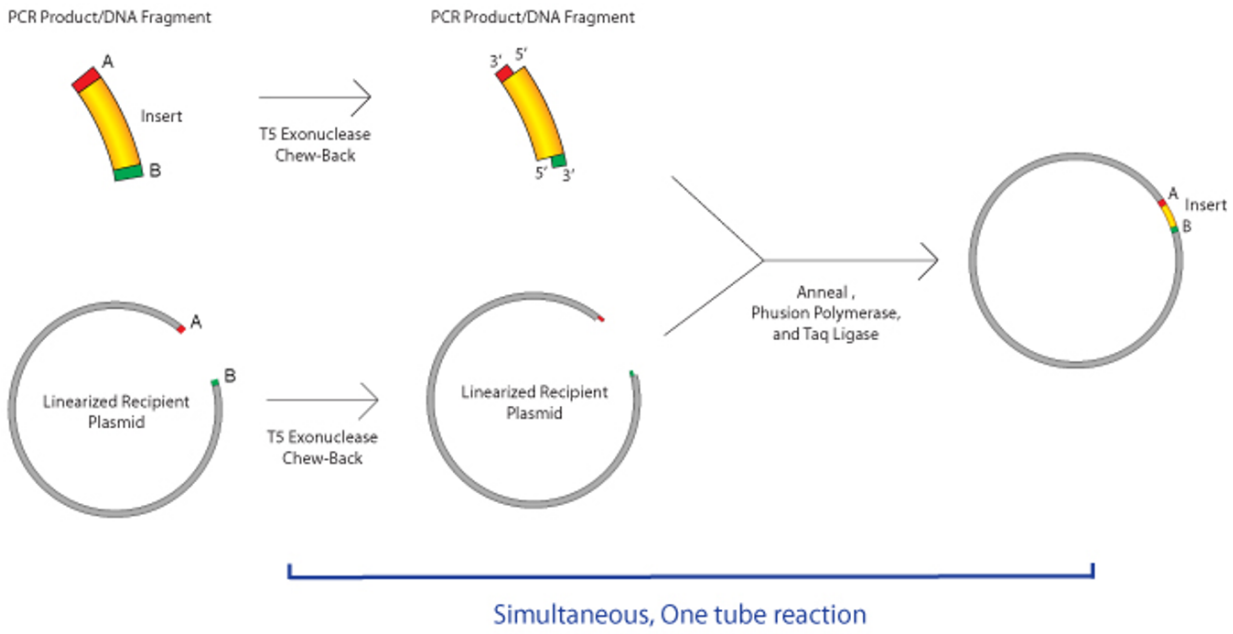


Figure 4: **Gibson Cloning Assembly.** Using PCR, overlap regions are created on both the DNA insert and the vector backbone. A T5 exonuclease creates sticky ends allowing for the combination of the two.

Source: Addgene

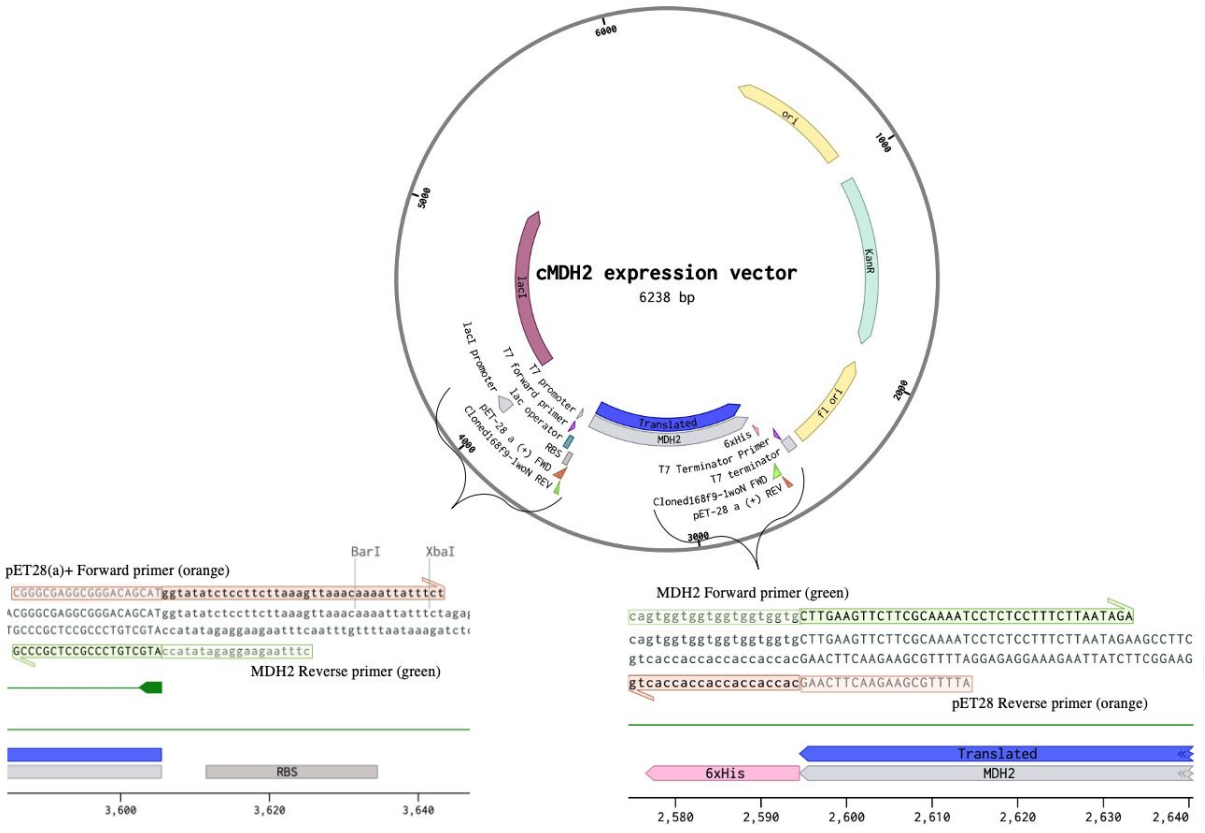


Figure 5: ***G. gallus*** **expression vector and the Gibson Assembly Primers.** A constructed MDH2 expression vector with key annotations. The sequences and the locations on the vector for the PCR primers are provided. Also worth noting are the T7 forward and T7 terminator primers utilized for sequencing.

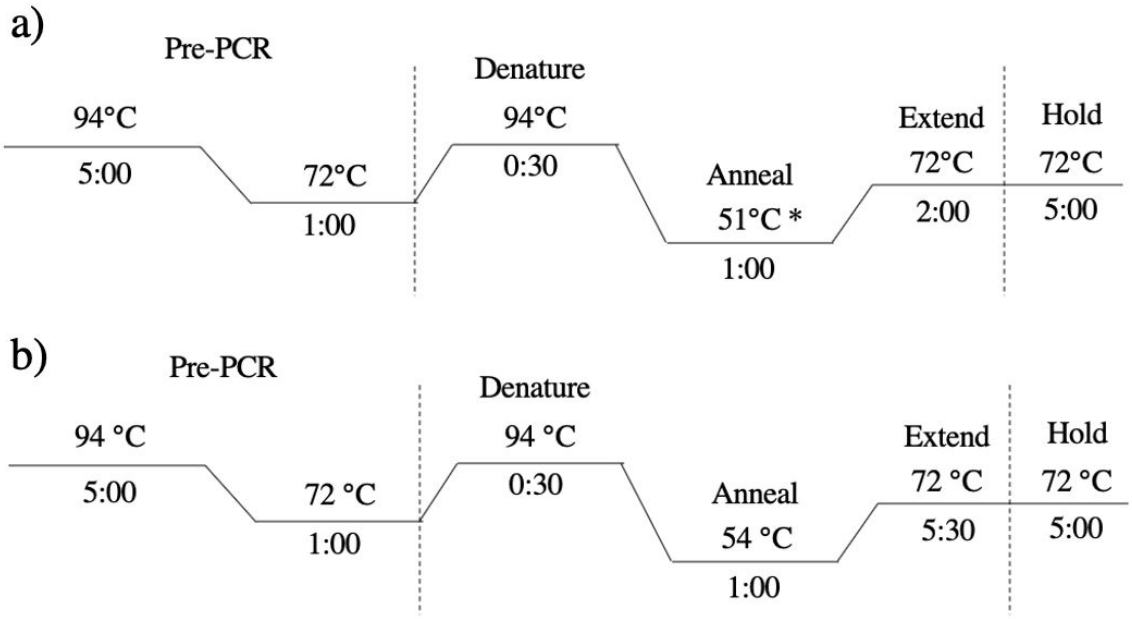


Figure 6: **Final PCR Conditions.** a) The Conditions for the Amplification of the *G. gallus* MDH2 gene insert. The annealing temperature increased by 0.3℃ over 30 cycles to reach 60℃. b) The conditions for the amplification of the pET28(a)+ PCR product utilized for the gel filtration trial.

	Amount of Fragments Used for Gibson Cloning Assembly	
	2-3 Fragment Assembly	Positive Control
Total Amount of Fragments	0.02–0.5 pmols X µl	10 µl
Gibson Assembly Master Mix (2X)	10 µl	10 µl
Deionized H_2O	10-X µl	0
Total Volume	20 µl	20 µl

Table 1: **Amounts of Fragments Utilized for Gibson Trial.** The total amount of fragments used for both the positive and experimental reactions. Molar ratios of the fragments were calculated first and then dilutions were used to match the recommended moles.

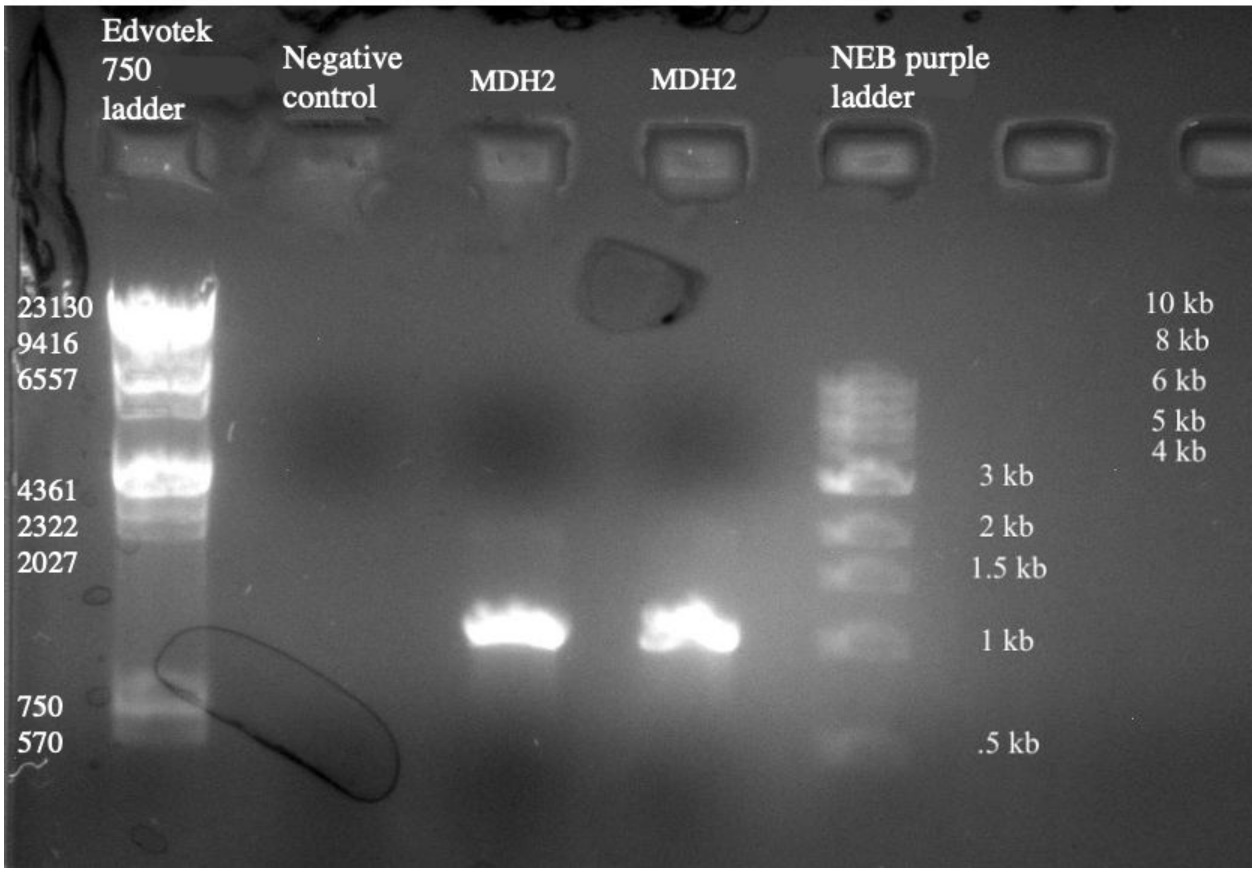


Figure 8: **MDH2 Insert PCR Visualization.** The MDH2 insert PCR products are visualized on 1% agarose gel. The amplification is successful as the insert is 1011 base pairs in length and the band lies just above the 1 kB ladder.

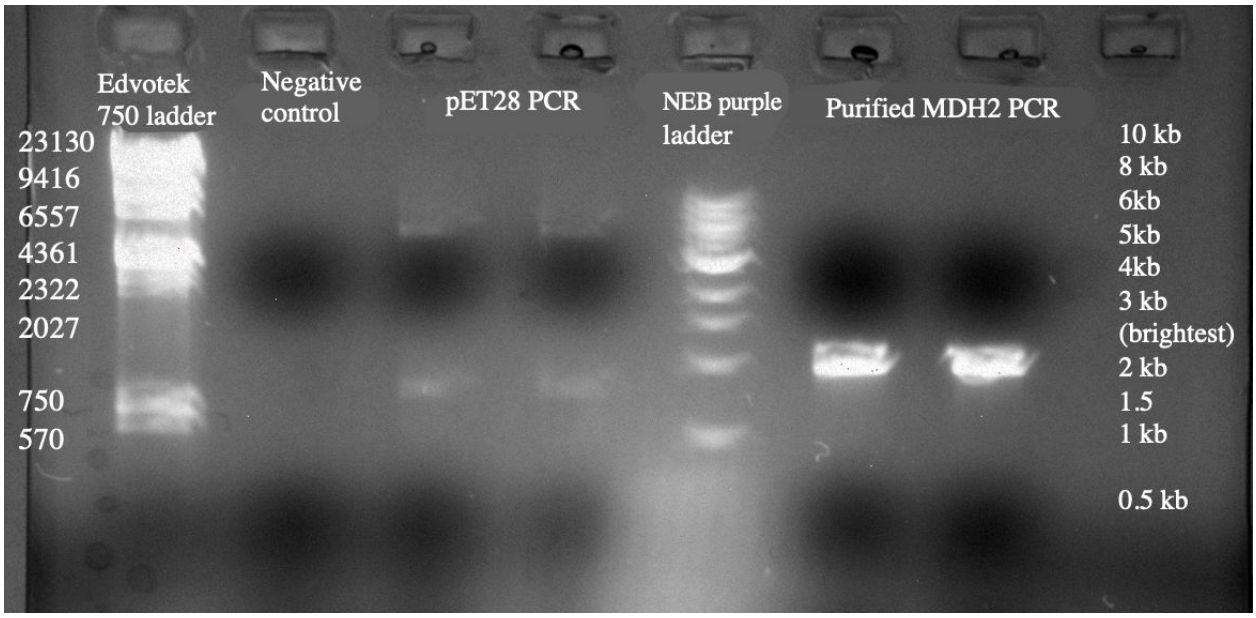


Figure 9: **Modified pET28(a)+ PCR Product and Purified MDH2 PCR Product Visualization.** The PCR amplification of the pET28(a)+ backbone has two products. The desired product is 5227 bp in length. It is the top band between the 6557 and 4361 bands on the Edvotek 750 ladder. There is a competitive, unwanted product just above the 750 band. The purification of the MDH2 insert removed any free nucleotides.

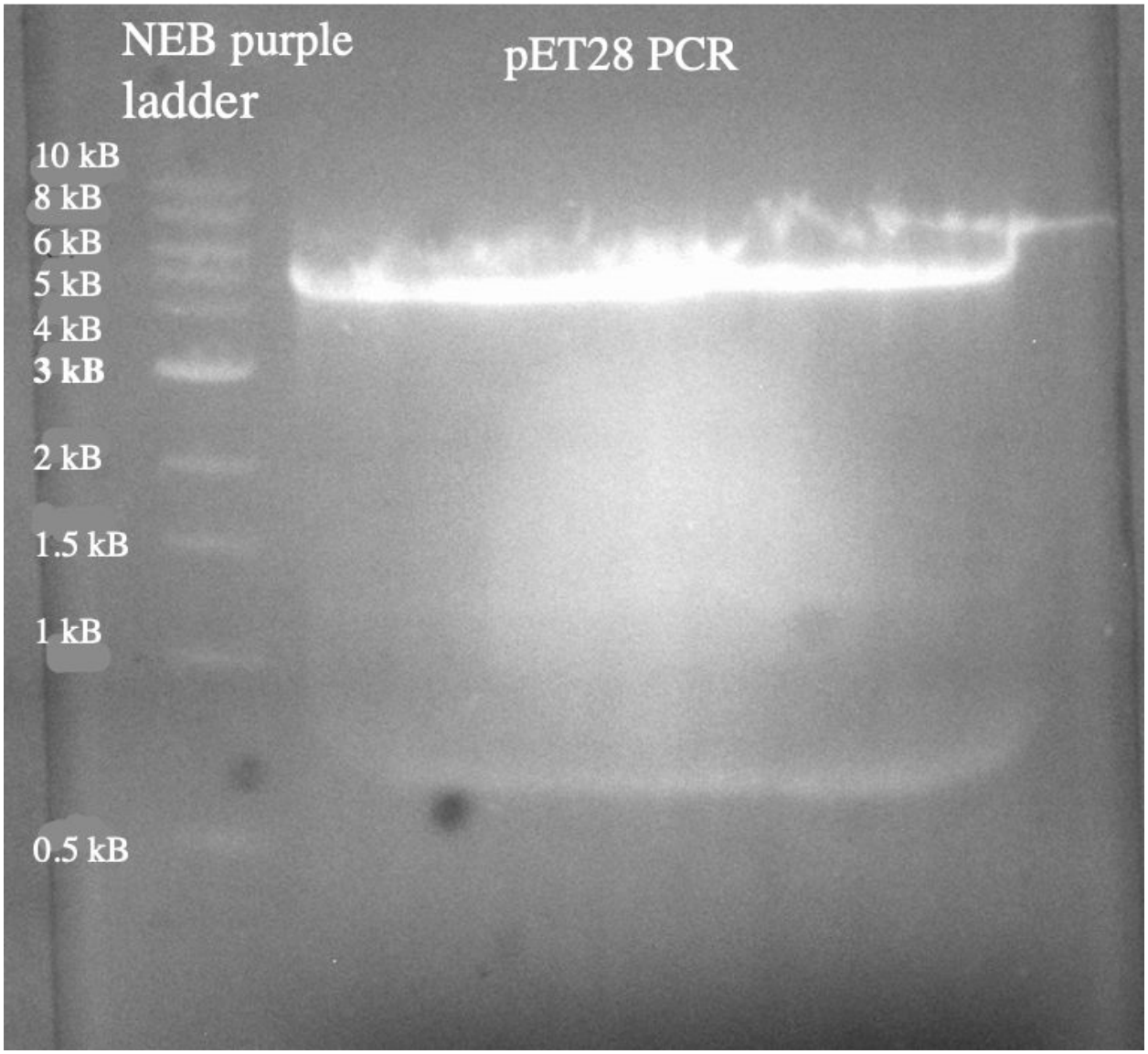
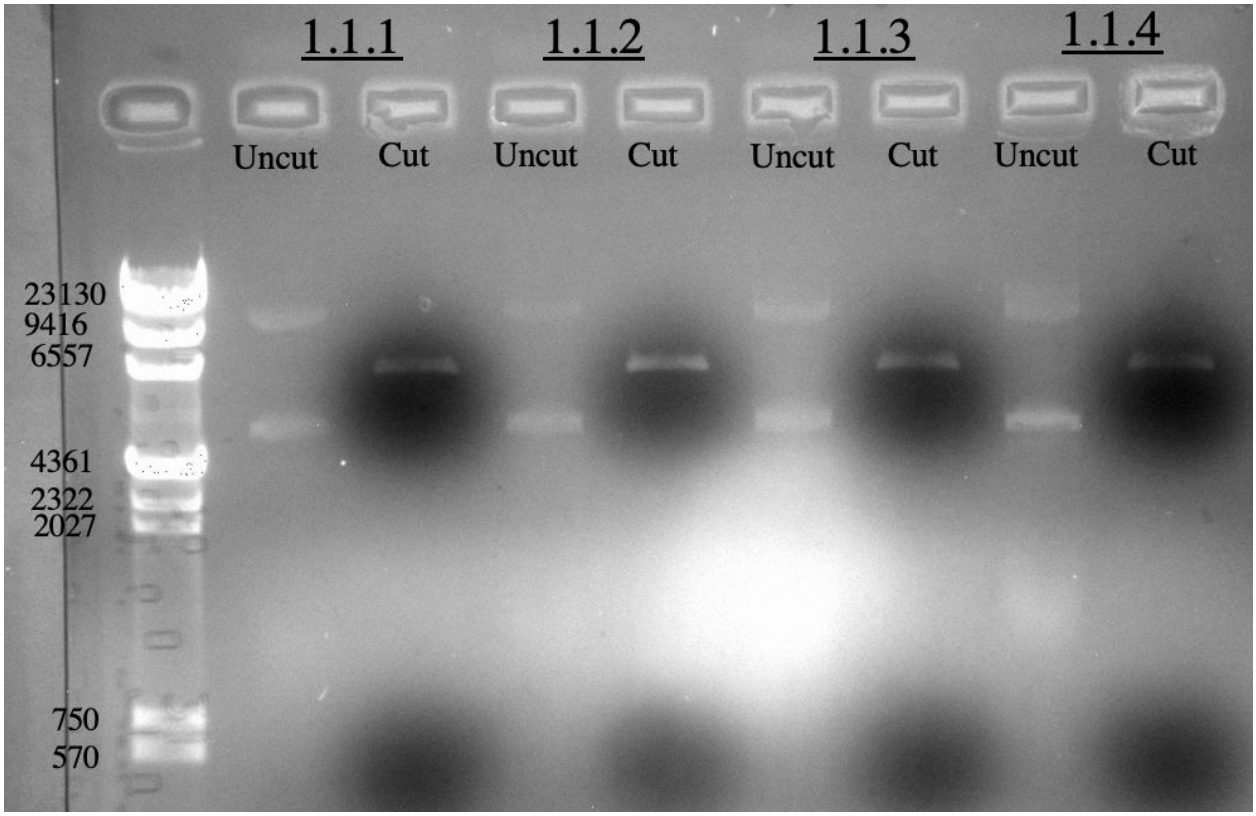
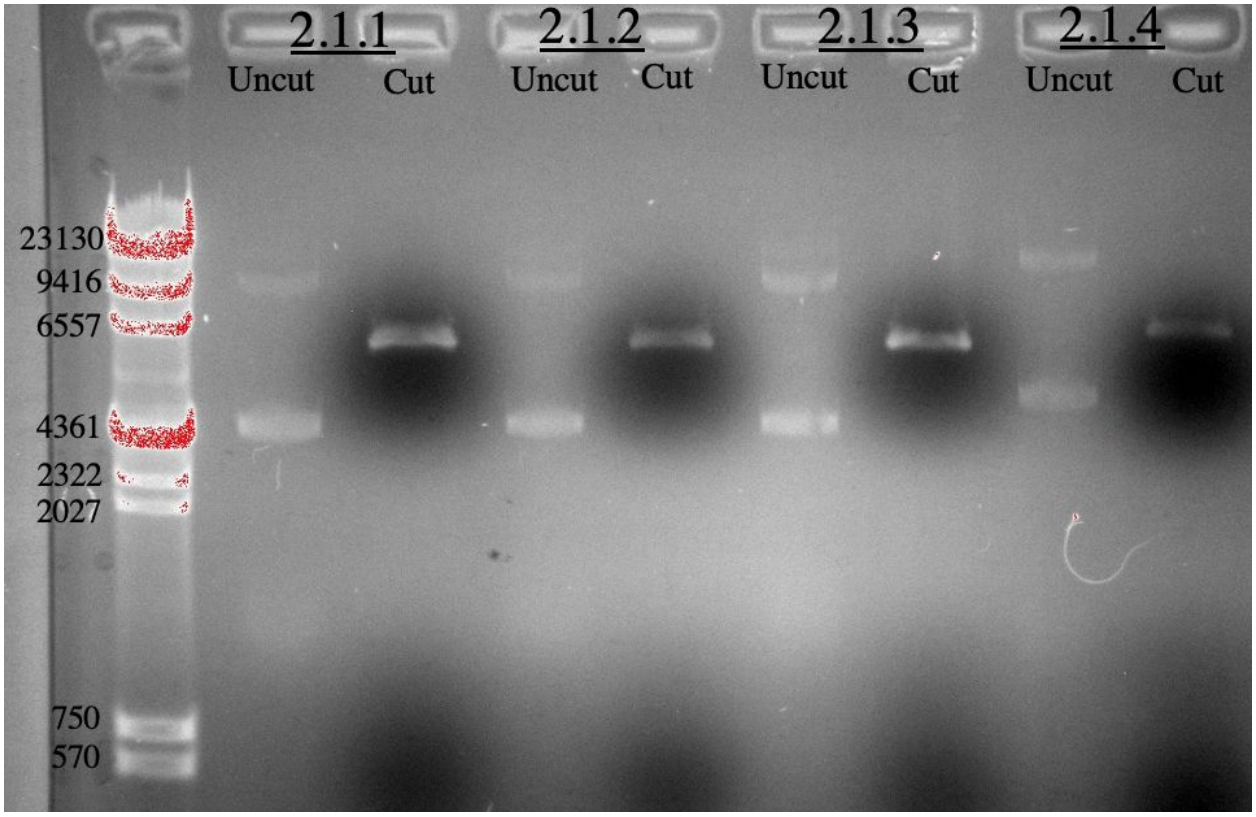


Figure 10: **Gel Purification of pET28(a)+ PCR Product.** Amplification of the pET28(a)+ backbone has two products and the usage of gel filtration allows for the excision of the desired product. The gel confirms that the large band is the backbone.

a)


1.1.1
1.1.2
1.1.3
1.1.4
Uncut
Cut
Uncut
Cut
Uncut
Cut
Uncut
Cut
23130
9416
6557
4361
2322
2027
750
570

b)


2.1.1
Uncut
Cut
2.1.2
Uncut
Cut
2.1.3
Uncut
Cut
2.1.4
Uncut
Cut
23130
9416
6557
4361
2322
2027
750
570

c)

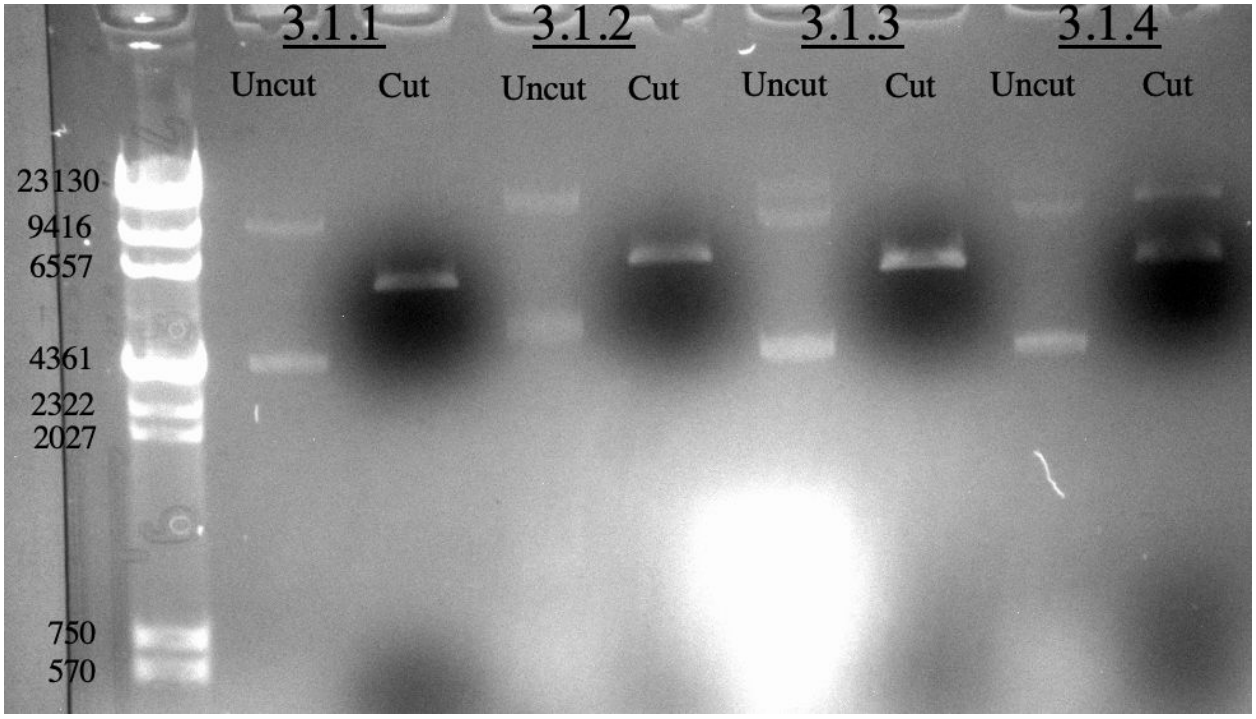

3.1.1
Uncut
Cut
3.1.2
Uncut
Cut
3.1.3
Uncut
Cut
3.1.4
Uncut
Cut
23130
9416
6557
4361
2322
2027
750
570

Figure 11: **Gibson Cloning Assembly EcoRI Restriction Digest Agarose Gel.** a) The EcoRI Restriction digest visualized on the 1% agarose gel for the 1:1 ratio of DNA insert to vector backbone produced the same results for all four preps. b) The EcoRI Restriction digest visualized on the 1% agarose gel for the 2:1 ratio of DNA insert to vector backbone produced the same band for all four preps. c) The EcoRI Restriction digest visualized on the 1% agarose gel for the 3:1 ratio of DNA insert to vector backbone produced similar results to the 2:1 and 1:1 digests except sample 3.1.4 had a distinct second band after being digested.

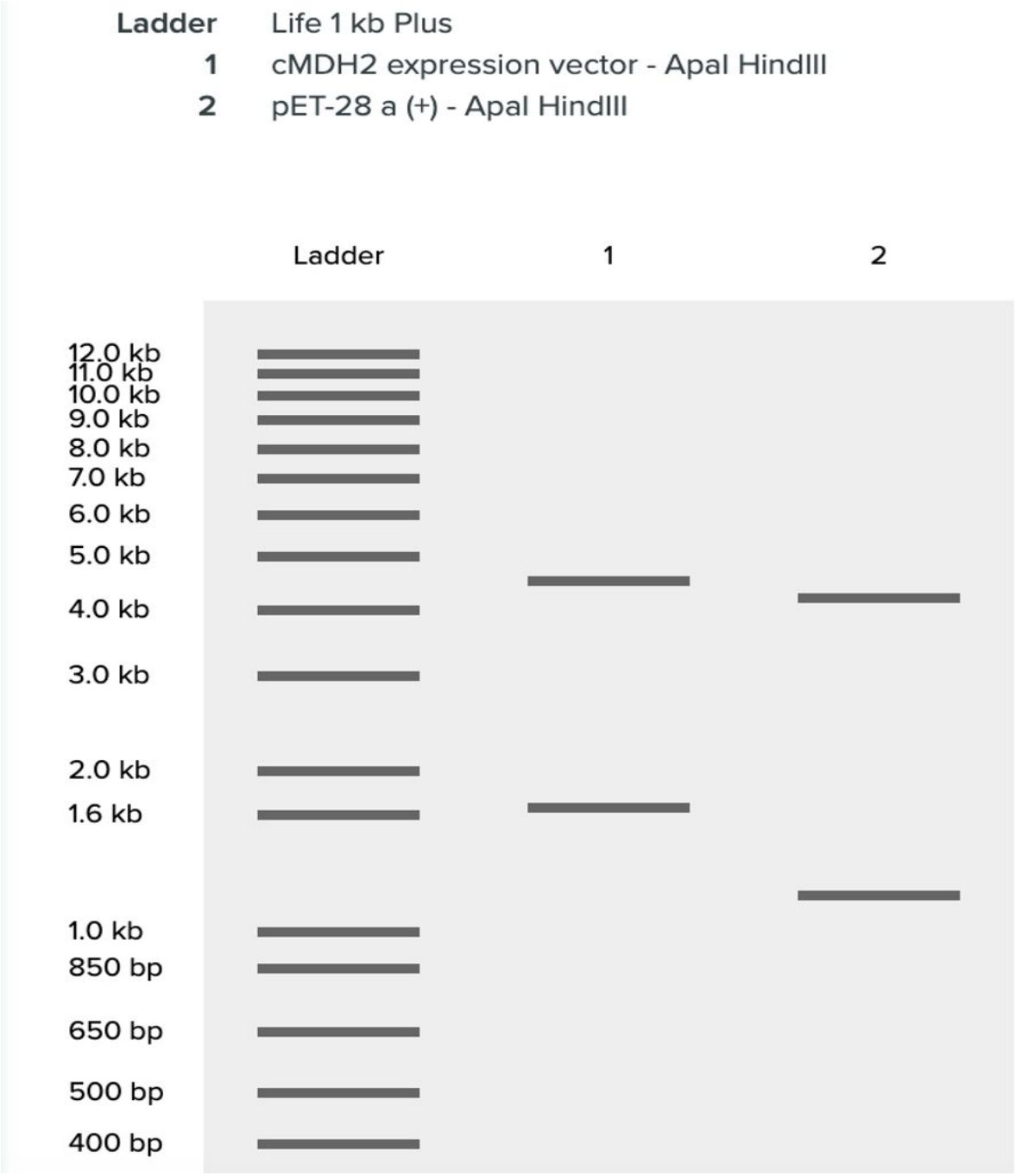


Figure 14: **Virtual ApaI and HindIII Digest Results.** The virtual digest shows the expected results of the double digest with both the constructed MDH2 expression vector and pET28(a)+ parent vector. Any preps containing the MDH2 gene insert will have a second band just below 2 kB while the parent vector has a band just above 1 kB.

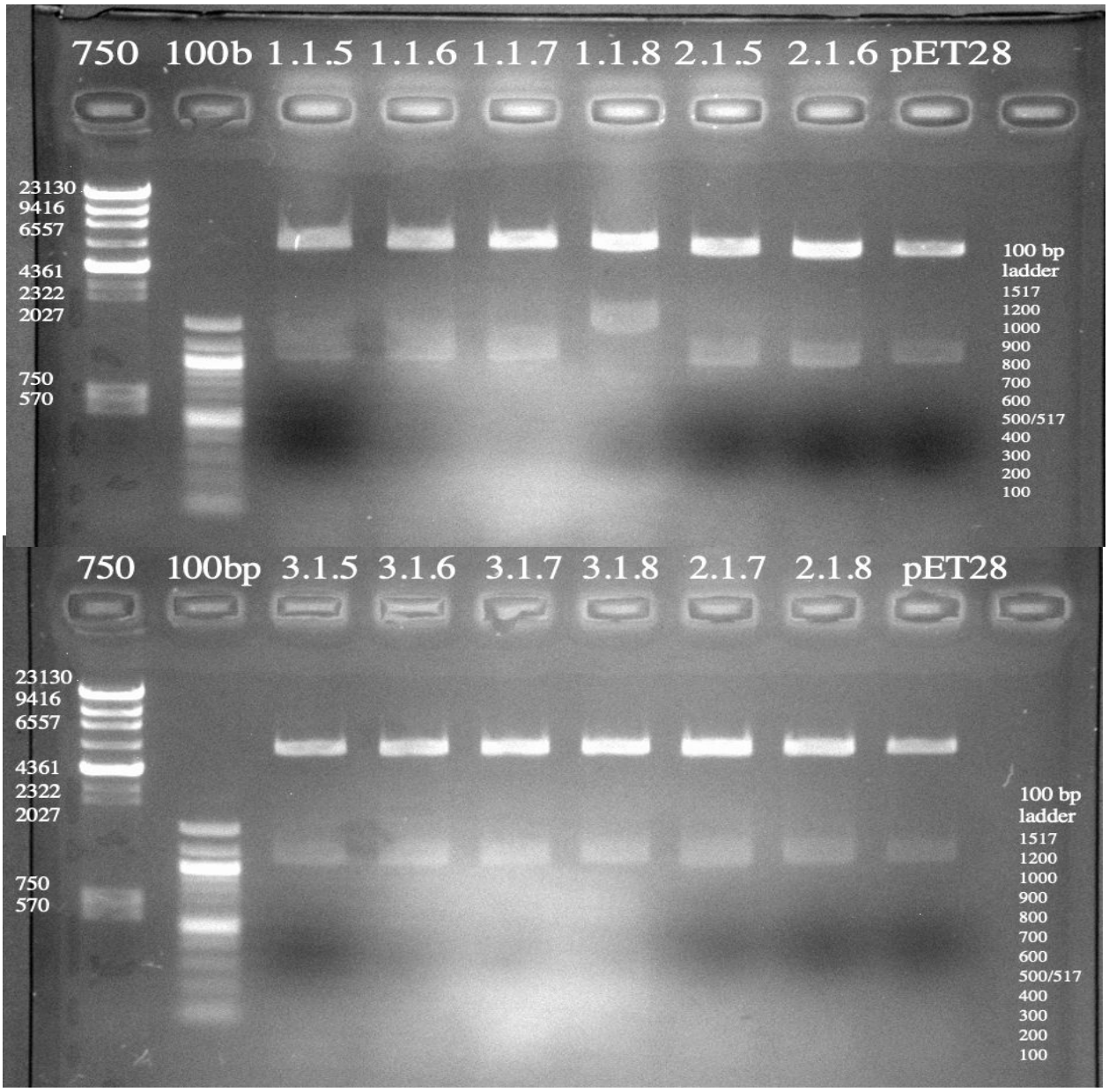


Figure 15**: Visualization of the New Prep Double Restriction Digest Using HindIII-HF and ApaI.** The double digest of the 12 new minipreps from the Gibson Cloning Assembly visualized on a 1% agarose gel revealed that one sample, 1.1.8, appeared to have the correct insert. All other sample digests matched the pET28(a)+ parent digest results.

Number of Colonies on Gibson Cloning Assembly Transformation to DH5α Cells				
Volume of Cells Added	Amp Positive Control	1.1 ratio	2:1 ratio	3:1 ratio
50 µL	35	3	4	4
100 µL	47	6	6	15
Rest	70	45	35	30

Table 2: **Number of Colonies on DH5α Transformation Plates.** The total number of colonies found on each Gibson Cloning transformation plate after overnight incubation at 37℃.

Number of Colonies on BL21 Transformation Plates

Serial Dilution	Negative Control	1.1.4	2.1.2
undiluted	0	24	200
1/10	0	1	34
1/100	0	0	8
1/1000	0	0	1
1/10000	0	0	0

Table 3: **Number of Colonies on BL21 Transformation Plates.** The total number of colonies found on each serial dilution transformation plate after overnight incubation at 37℃.

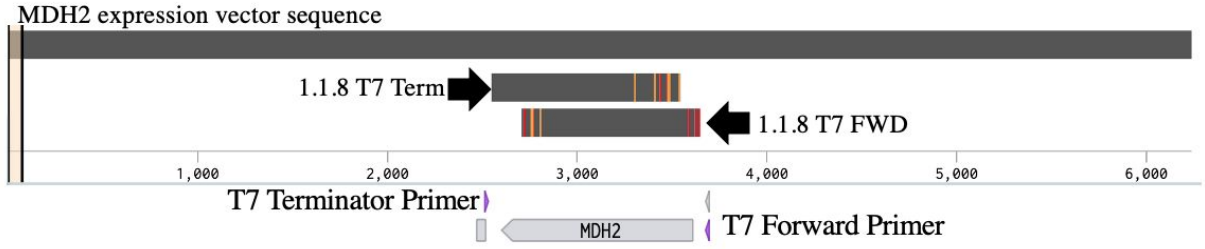


Figure 16: **Sequence Alignment of Sample 1.1.8 With MDH2 Expression Vector.** Using the T7 Forward and T7 Terminator Primers, the DNA from sample 1.1.8 was sequenced. The data

aligned well with the DNA insert section of the constructed expression vector confirming the successful insertion of the MDH2 gene.

Chapter 3: Protein Expression

Introduction:

Protein expression in *E. coli* cells requires the use of an inducing agent that activates transcription. The plasmid vector pET28(a)+ contains the lac operon. In the absence of lactose (or a synthetic mimic), this operon is repressed. In living cells, the presence of lactose and the absence of glucose signal for the repressor to be removed, and transcription begins. In the case of pET28(a)+, a molecular mimic, isopropyl β- d-1-thiogalactopyranoside (IPTG) activates the lac operon. This then transcribes the mRNA for the gene of choice that is located slightly downstream on the DNA sequence. The mRNA is translated into protein (Figure 17). The process of protein induction can be difficult, given that several parameters can be the cause of a low amount of protein production. The induction temperature and induction time can vary for each expression vector and protein gene.

Another key factor in the expression of proteins is codon usage. These nucleotide triplets signal for a tRNA that carries a particular amino acid. There are 64 codons, with 3 coding for sequence termination, and 20 amino acids. This means that some codons will signal for the same amino acid. The amount of tRNAs available for each codon varies from organism to organism, resulting in codon usage variation (Brule, 2017). The expression of a *G. gallus* gene in an *E. coli* cell may be difficult as a result.

Materials and Methods:

To test for protein induction in the *G. gallus* MDH2 expression vector, a constant temperature, 37℃, was selected with induction time being the variable. A 10 mL LB+Kan (50 μg/mL) culture inoculated from a single MDH2 BL21 colony and a 10 mL LB+Kan (50 μg/mL) culture inoculated from a single human MDH1 BL21 colony were incubated at 37℃ and 250 rpm until the OD600 was between 0.5-0.6. To induce expression, IPTG was added to a final concentration of 0.5mM. Cell samples were collected before induction, after 3, 5, and 18 hours of induction. The samples were run on a 1.2% bisacrylamide SDS-PAGE and stained with coomassie blue to examine the protein expression.

Using Benchling, codon optimization was performed on the MDH2 gene insert to determine what codons could be changed to improve protein induction. To create a more realistic codon optimization protocol for future usage, key codons identified by the EMBL that are often the cause of translational problems in *E. coli* were selected for optimization (Protein Expression).

Results and Conclusions:

The SDS-PAGE gel revealed the results of the 37℃ induction test to check the expression of the *G. gallus* MDH2 expression construct. The cytoplasmic *Homo sapiens* MDH was used as a control for induction as it has been optimized for induction at 37℃ for 3 hours. The MDH protein is expected to have a size of around 30-35 kD. In lanes 2 and 3, the uninduced cell samples appear to be comparable to each other. In the induction lanes 4 through 9, there is clearly strong induction from the human version, but there is little induction for the chicken

MDH2. The solid band in lanes 4, 6, and 8 lies just below the 37 kD ladder band confirming the protein produced is MDH1. There is no bold, distinct band present in the MDH2 induction lanes (Figure 18).

Due to the limited expression during the induction test at 37°C, codon optimization becomes an option to improve the ability of *E. coli* cells to produce the target protein. When the expression vector is codon optimized, there are 74 codons that could be changed. Even with a number so large, this only reduces the number of rare codons from 12 to 5. Changing 74 codons was an unrealistic amount so the number was reduced to 9 codons. These 9 codons lie throughout the entire MDH2 gene (Figure 19). The codons picked were identified by EMBL and coded for arginine, proline, and leucine as they often give the most trouble when expressing proteins in *E. coli* (Protein Expression). The arginine codon changes utilize two codons, while the others only use one (Table 4).

Due to the COVID-19 pandemic, the continuation of wet labs was not possible. Testing temperature variables during the induction conditions would have been conducted before codon optimization.

When compared with optimized human cytoplasmic MDH expression, it is clear that these induction conditions or the codon usages are not ideal for chicken MDH2 in *E. coli*. Further tests of induction conditions and codon optimization are required. One potential alternative explanation is alteration outside the region that has been sequenced. The promoter region should be analyzed by sequencing.

Figures:

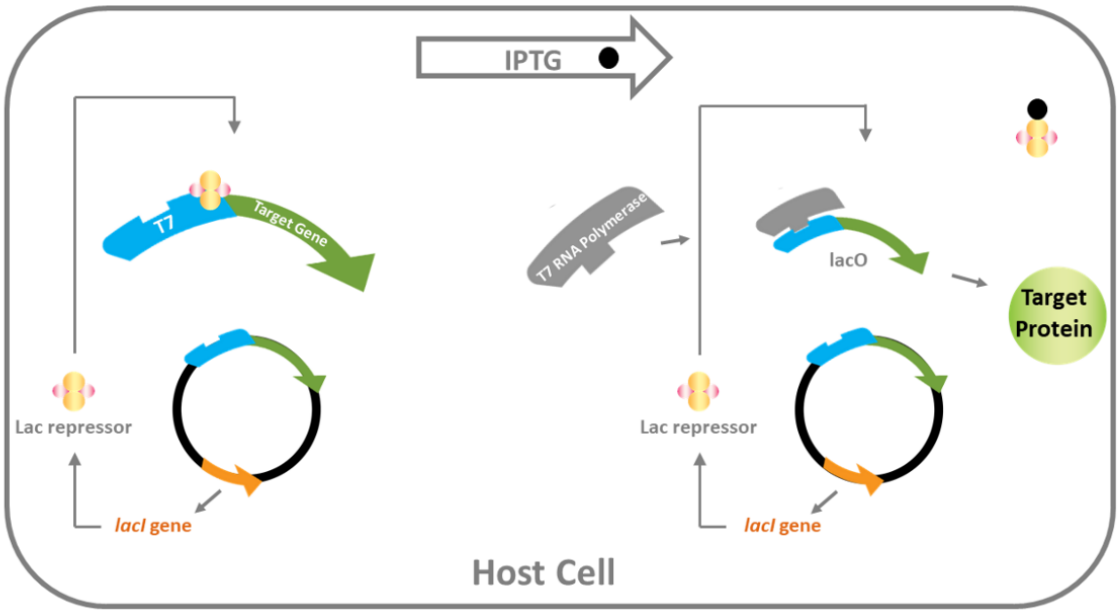


Figure 17: **Simplified diagram of IPTG protein expression.** Transcription begins when IPTG signals for the removal of the Lac repressor. The gene of interest is transcribed to mRNA and translated into the target protein.

Source: Gold Biotechnology, 2018.

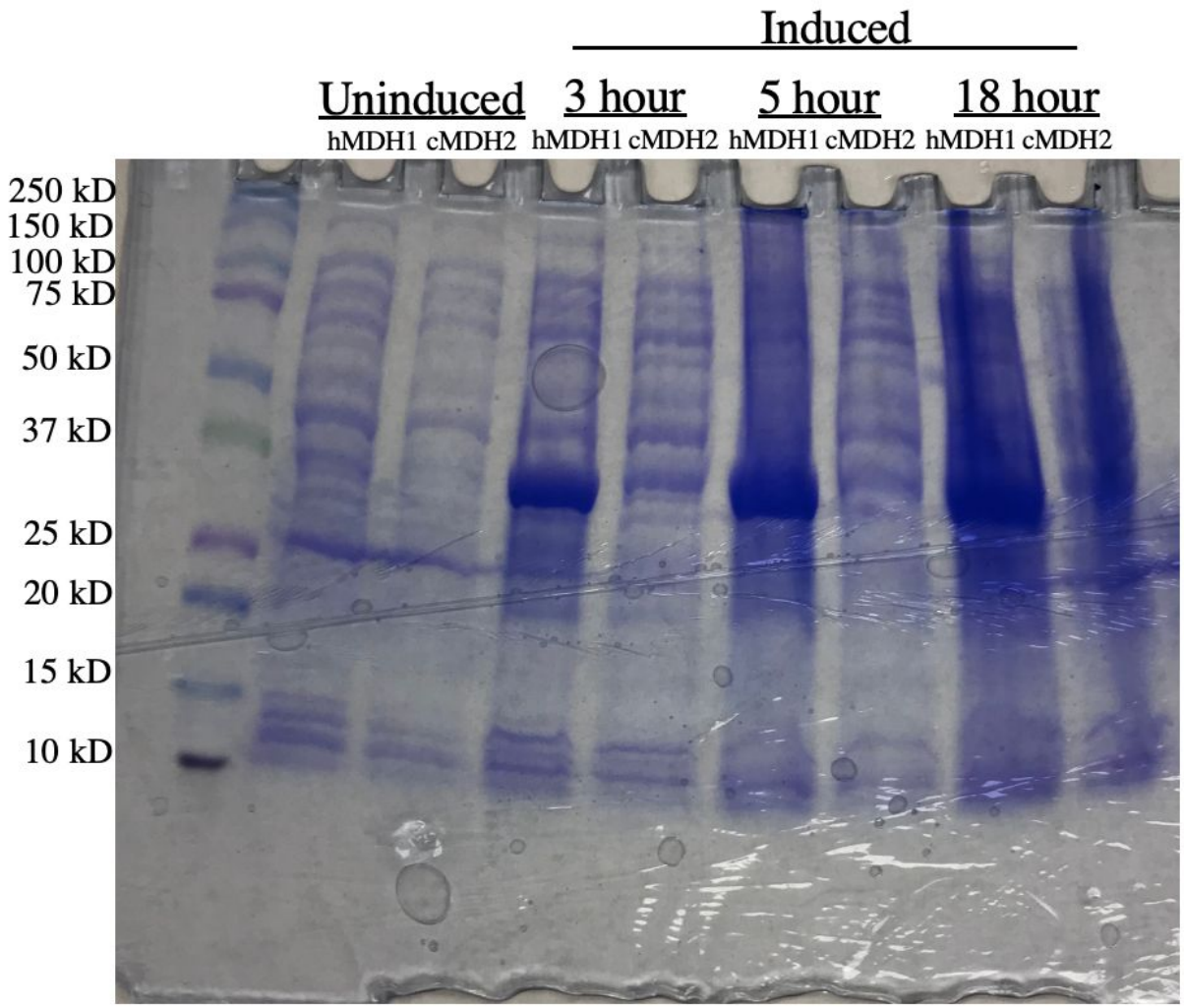


Figure 18: ***G. gallus*** **MDH2 expression test SDS-PAGE.** Samples of protein were collected during the induction of MDH at 37℃. The first samples are uninduced proteins, before IPTG was added. The next 3 sets of samples were collected after 3, 5, and 18 hours of induction. hMDH1 is the cytoplasmic MDH in *Homo sapiens*, while cMDH2 is the mitochondrial MDH in *G. gallus.* The human MDH1 was used as a control to compare for ideal induction.

Range 1: 1 to 1011 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1784 bits(966)	0.0	996/1011(99%)	0/1011(0%)	Plus/Plus

```
Query  1     ATGCTGTCCCGCCTCGCCCGTCCCGCCGCTGTCCTGTGCCGCGGCCTGGCCACCAGCGCC  60
Sbjct  1     ......................G.....................................  60

Query  61    CAGAACAATGCCAAGGTGGCAGTGCTGGGAGCATCAGGAGGGATTGGGCAGCCTCTCTCC  120
Sbjct  61    ............................................................  120

Query  121   CTTCTCCTGAAGAACAGCCCGCTGGTGAGCAGGCTCACCCTTTACGATATTGCGCACACT  180
Sbjct  121   ............................C.T.........................  180

Query  181   CCTGGTGTTGCAGCTGACCTCAGTCACATCGAGACAAGAGCGAATGTTAAAGGTTTCCTG  240
Sbjct  181   ................................C.C.....................  240

Query  241   GGACCCGAGCAGCTGCCAGAATGTTTGAAGGGCTGTGATGTTGTAGTGATTCCTGCAGGA  300
Sbjct  241   ............................................................  300

Query  301   GTCCCTAGGAAACCAGGTATGACCCGTGATGACCTGTTCAACACCAATGCTAGCATTGTT  360
Sbjct  301   ......C.C...................................................  360

Query  361   GCTACTTTAACAACTGCCTGTGCAAAGCACTGCCCAGAAGCCATGATCTGTATTATTTCC  420
Sbjct  361   ............................................................  420

Query  421   AACCCGGTAAATTCAACCATACCCATAACTTCAGAAGTTTTCAAGAAGCATGGTGTGTAT  480
Sbjct  421   ............................................................  480

Query  481   AATCCCAATAGAATCTTTGGTGTAACAACACTGGACATCGTCAGAGCGAATACTTTTGTG  540
Sbjct  481   .....G.......................................................  540

Query  541   GCTGAGCTAAAGGGTTTAGATCCAGCTCGAGTAAGTGTTCCAGTTATTGGTGGCCATGCA  600
Sbjct  541   ......T.G.....................T..............................  600

Query  601   GGGAAAACCATCATCCCTCTGATCTCCCAGTGCACACCGAAAGTGGACTTCCCTCAGGAT  660
Sbjct  601   ............................................................  660

Query  661   CAACTGGAAAAGCTTACAGGGAGAATTCAAGAAGCTGGCACTGAAGTTGTCAAAGCCAAA  720
Sbjct  661   ......................C.T.....................................  720

Query  721   GCAGGAGCAGGATCTGCCACCTTGTCTATGGCTTATGCTGGCGCTCGGTTTGTGTTCTCT  780
Sbjct  721   ............................................................  780

Query  781   CTGGTGGATGCCATGAATGGAAAGGAAGGAGTTATTGAATGTTCCTTTGTTCGCTCGGAA  840
Sbjct  781   ............................................................  840

Query  841   GAGACGGAGAGCCCATACTTCTCTACACCTCTGCTACTGGGAAAAAATGGAATTGAGAAG  900
Sbjct  841   ............................................................  900

Query  901   AACCTAGGCATTGGCAAGATCACCCCCTTTGAAGAGAAGATGGTTGCTGAGGCCATGGCT  960
Sbjct  901   ...T.G.......................................................  960

Query  961   GAGCTGAAGGCTTCTATTAAGAAAGGAGAGGATTTTGCGAAGAACTTCAAG  1011
Sbjct  961   ................................................  1011
```

Figure 19: **Alignment of codon- optimized MDH2 gene with the unoptimized gene.** The alignment shows which nucleotides change in order to optimize the 9 codons. Table 4 provides further detail into each codon. Query: Unoptimized MDH2 Subject: Optimized MDH2

Codon Optimization of *G. gallus* MDH2 Expression Vector			
Nucleotide Number	Initial Codon	Optimized Codon	Amino Acid Residue
991	CCC	CCG	P
862	AGG	CGC	R
796	AGG	CGC	R
706	AGG	CGC	R
529	CCC	CCG	P
466	CTA	TTG	L
445	CGA	CGT	R
331	AGA	CGT	R
109	CTA	TTG	L

Table 4: **Codon Optimization Chart.** The table provides the initial and optimized codon, nucleotide number, and the residue.

Chapter 4: Protein Modeling and Bioinformatics

Introduction:

Bioinformatics, protein modeling, and other *in silico* experiments serve a key role in enzymatic studies. When studying avian MDH2, these options are helpful for avoiding tissue collection, protein isolation, and construction of expression vectors, all procedures that can be time-consuming and expensive. Amino acid sequence alignments are the first step in comparing different enzymes. An alignment comparing MDH2 sequences from various avian species could reveal potential relationships between the enzyme and different physiological characteristics. Beyond the primary structure, proteins can be studied in regard to their 3D structure. Using programs like Pymol and Phyre2, a predicted 3D model can be produced. These models are not perfect, but provide an excellent tool for understanding the protein's secondary structure and other important enzyme characteristics. Differences in secondary structure or characteristics like binding properties may account for some of the results found in previous studies of avian MDH2.

Materials and Methods:

The protein sequences were aligned for *Zonotrichia albicollis* (white-throated sparrow, XP_014128222) and *Ficedula albicollis* (collared flycatcher, XP_005056570) and *G. gallus* (XP_415765) were selected from the avian MDH2 sequences available on NCBI's protein database (https://www.ncbi.nlm.nih.gov/protein/). Using ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/), these three sequences were aligned.

An additional sequence alignment was created using ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with the three avian species (XP_014128222,

XP_005056570, XP_415765) and the *Homo sapiens* MDH2 sequence (P40926) obtained from Uniprot to predict active and catalytic sites, NAD+/NADH binding sites, and dimerization interfaces.

Using the FASTA sequences for the avian species, three predicted models were generated by Phyre2's intense mode (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) (Kelley, 2015).

For a quantified comparison of the structures, root-mean-square-deviation was calculated in Pymol (https://pymol.org/2/). This calculation is the average distance between selected aligned atoms. The Python command "align 'name1', 'name2', cycles=0, transform=0" was used to generate RMSD without rejecting any outliers. Using the Python command "align 'name1', 'name2', cycles=5, transform=0" runs through 5 cycles of removing outliers to provide a more accurate RMSD value. The same calculation was done using the Phyre2 structure of *Homo sapiens* cytoplasmic MDH.

PyMol (https://pymol.org/2/) was used to generate a dimer of the avian species based on the *Homo Sapiens* MDH2 structure (P40926).

Molprobity (http://molprobity.biochem.duke.edu/index.php) was used to generate Ramachandran plots of the 3 predicted enzyme models (Williams, 2018).

H++ (http://biophysics.cs.vt.edu/index.php) allows for pKa estimation of protein amino acid side chains (Gordon, 2005). The three sequences were entered into the system using default settings with the exception of pH being set at 8.0 and the AMBER set for cubic.

Results and Conclusions:

When selecting species for a sequence alignment, *G. gallus* would serve as the non-migratory species, and two additional species were chosen to represent migratory birds. *Z. albicollis* (white-throated sparrow) and *F. albicollis* (collared flycatcher) were selected from the avian MDH2 sequences available on NCBI's protein database. A sequence alignment between the three avian species— *Z. albicollis*, *F. albicollis*, and *G. gallus*— provides insight into differences in the primary sequence between the enzymes (Figure 20). There were 14 amino acid differences that either had a change in polarity, charge, or phosphorylation ability. There are 4 additional serine and threonine residues in both *Z. albicollis* and *F. albicollis* not found in *G. gallus* while two serine residues are present in *G. gallus* that the other two species do not have.

The active sites, catalytic sites, substrate binding sites, and dimer interface regions have been added to the sequences by aligning to the human mitochondrial MDH (MDH2) which already has a solved crystal structure (Figure 21). The dimer interface regions are critical to the enzyme's ability to function, as hindered dimerization inhibits the enzyme. All active and catalytic site residues are conserved. There is an alanine residue present in the final NAD+/NADH binding region that is serine in the avian species at residue number 140. The dimer interface regions have two residue differences at residues 73 and 238.

Phyre2 creates predicted 3-D models with an amino acid sequence. These files were exported into Pymol for visualization and into other online programs for further analysis. The model generated for each species is displayed individually with *Z. albicollis* in blue, *F. albicollis* in green, *G. gallus* in pink, and an overlay of all three (Figure 22). The predicted models had variations in the secondary structure. Alpha helices, beta-sheets, and other regions were not always conserved. However, the majority of these variations were located at the beginning and

end of the sequences which are not modeled well. The solved structures used to generate the models do not have these regions in them, which results in the varying structure.

When comparing models, root-mean-square-deviation provides a way to quantify the differences between the generated Phyre2 models. The smaller the RMSD value, the more similar the structures. Identical structures have an RMSD of 0. The RMSD values with outliers are 5.819 between *G. gallus* and *F. albicollis*, 4.509 between *G. gallus* and *Z. albicollis*, and 4.401 between *Z. albicollis* and *F. albicollis* (Table 5). These indicated that the models have noticeable differences due to RMSD ≥ 2Å (Bordogna, 2011). RMSD has been shown to be easily manipulated and skewed by outliers (Kufareva, 2012). When adjusted for the outliers, the RMSD values for all three comparisons drop significantly to 0.126, 0.163, and 0.163, suggesting that the models are highly similar (Table 5). To account for variation as a result of the modeling output from Phyre 2, a triplicate RMSD test of the same amino acid sequence for *Homo sapiens* cytoplasmic MDH was conducted. All three alignments generated an RMSD of zero which shows that the variation seen is the result of the sequence differences and not errors in Phyre2.

MDH2 is found as homodimeric across most species. Each monomer has two distinct functional regions, but without dimerization, the enzyme loses catalytic function (Minarik, 2002). The dimer interface is made of α-helices that fit together (Minarik, 2002). The models generated by Phyre2 were monomers only, therefore Pymol was used to create them. The dimer model for *Z. albicollis* is blue and rose gold, *F. albicollis* is pink and yellow, and *G. gallus* is green and cyan (Figure 23). The alpha helices along the dimer interface are conserved in all three models. The substitution of the glutamine residue for lysine at residue number 238 in *Z. albicollis* and *F. albicollis* does not appear to be directly in the interface so it may not have a

huge impact on dimerization ability. The circled areas on the overlay show the location of these residues (Figure 23).

Ramachandran plots provide the angles and locations of the amino acid residues of each model and provide a plot to compare the three. The Ramachandran plot generated by MolProbity, a server designed to further analyze the structure of individual proteins, shows that the *Z. albicollis* model has 4 outliers and 98.8% of all residues lie in allowed regions (Figure 24). The *G. gallus* has 1 outlier residue (Figure 25). The *F. albicollis* model has 5 outlier residues (Figure 26). 7 of the outliers lie within the first 25 residues (Ala11, Ala13, Ala16, Gly19, Ala24) or the last 20 residues (Phe336). These regions are not well modeled which may account for these outliers.

After analysis of all the protonation sites and their effect on pKa, the data was similar across all three species. However, a few differences arose because the collared flycatcher had 6 fewer protonation sites and there is a His-22 in the white-throated sparrow that is not present in the other two species. These differences, which lie outside the conserved regions, could affect the pKas at the conserved regions. If the pKas shift, it can affect the protonation state of a residue. The change in protonation could affect the charge on a residue and, in turn, change the pKas and binding abilities of other residues. The pKas of the residues in the active sites and binding sites are compared between the three species (Figure 27). They are nearly identical but slight variation exists at R-109 in the chicken (113 and 108 for the Flycatcher and Sparrow, respectively) and K-202 (206 and 201). R109 is predicted to be in the substrate binding region while K-202 is estimated to be in the active/catalytic site.

When the pKa decomposition data for R109 was considered, there are 3 residue differences affecting its pKa. Ile residues at site 200 in *F. albicollis* and 195 in *Z. albicollis* are

not present in *G. gallus*. Gly-227 in *G. gallus* is an alanine residue in both *Z. albicollis* and *F. albicollis*. Lastly, there is a lysine residue present at residue number 238 in *G. gallus* that is not in the other two species.

The same idea was applied to compare the pKas of the dimer interface residues (Figure 28). Like the active site residues, the pKas are similar across all three species. Arg-73, Glu-235, and Lys-240 all have differences between the three species. It is important to note the lack of lysine at residue 238 in *F. albicollis* and *Z. albicollis*.

The sequences from *G. gallus*, *Z. albicollis*, and *F. albicollis* are similar, with predicted catalytic sites being conserved across all three species. The data suggest that the differences at the dimer interface do not play a significant role in changing the enzyme structure. The enzymes are quite similar overall in structure, as supported by the RMSD outlier-adjusted calculations. The structures themselves likely do not account for differences required for varying physiological needs, but rather other regulation. The differences that do exist in the sequences suggest that regulation of these enzymes occurs outside the active sites, with the potential for phosphorylation to play a role. The addition of serine or threonine residues in the migratory species outside the predicted interface regions may affect the enzyme's ability to dimerize and function when phosphorylated.

The Ramachandran plots suggest that the predicted structures are relatively realistic as nearly all the residues were in favorable or allowed positions. The differences in the 3D structures are captured by the different plot points between the three and show that the residues lie at different angles between the species.

The differences in pKas in the active sites and dimer interface do not seem to be significant but could play a role depending on the microenvironment of the enzymes. The

Arg-109 pKa differences should be investigated in more detail due to its key role in substrate binding. The differences in pKas do suggest that there is variation between the three enzymes and these could be explored further. These variations may provide a molecular explanation to how the different metabolic needs of various bird species are regulated.

Figures:

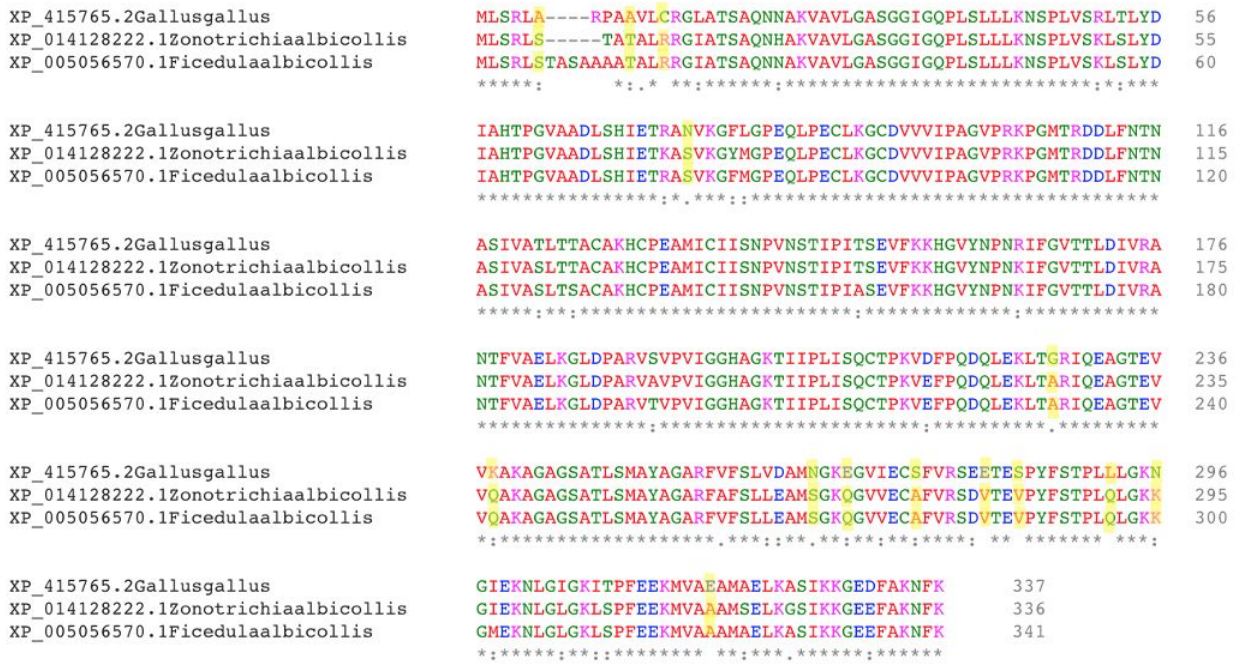


Figure 20: **Sequence Alignment of the Three Avian Species.** The various colors represent the types of amino acids. Highlighted are several differences of interest.

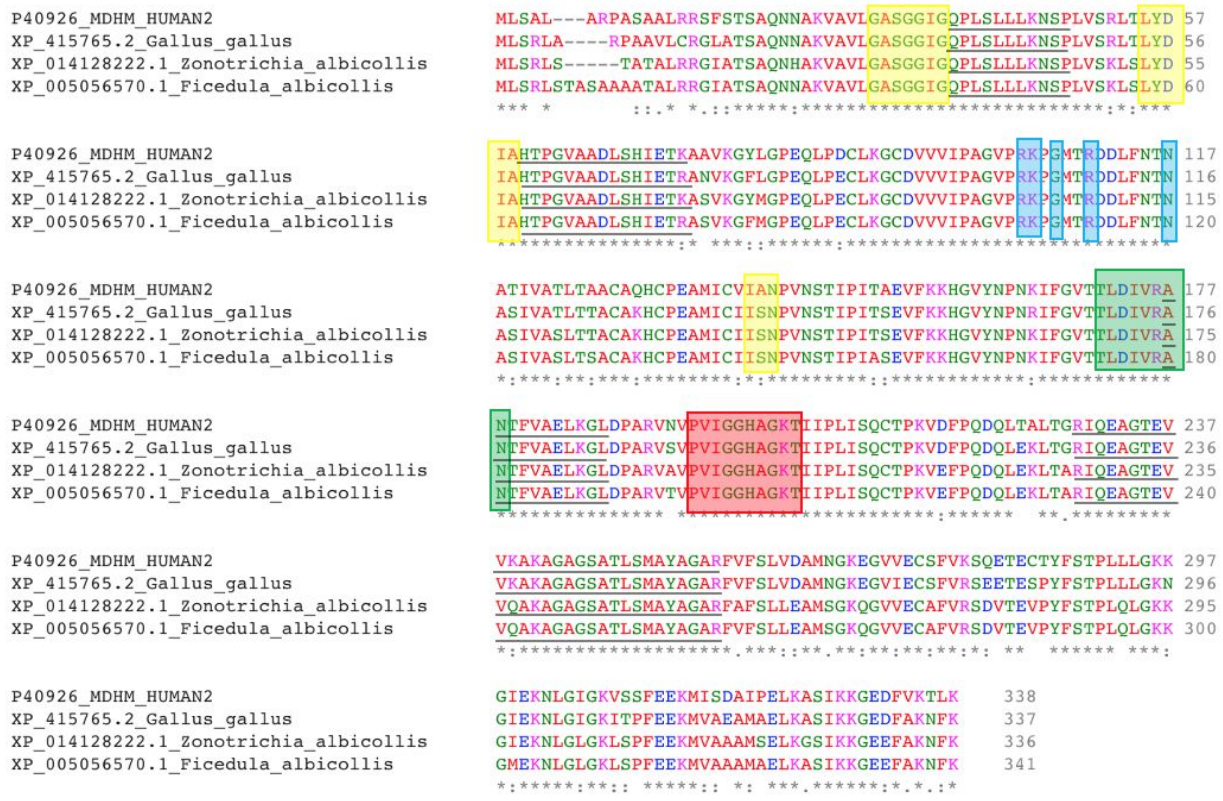


Figure 21: **Sequence Alignment of the Three Avian Species With *Homo sapiens* MDH2**. This alignment allows for prediction of the locations of key enzyme sites. KEY: Nucleotide (NAD+/NADH) binding, Substrate binding, Active site/Catalytic site(D/R), Active site/Catalytic site(Base), XX: Interface regions

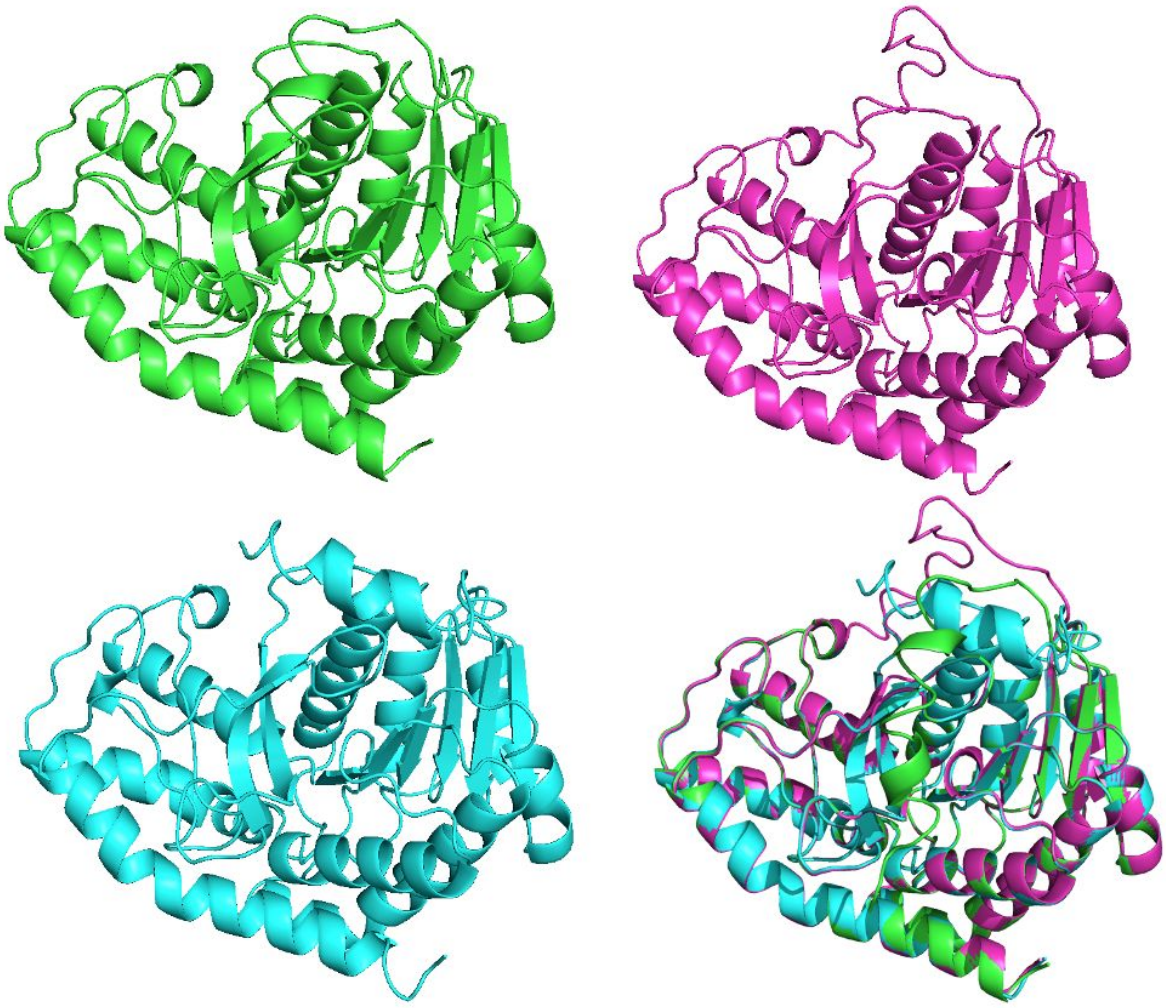


Figure 22: **Predicted Three-Dimensional Models of *G. gallus*, *Z. albicollis*, *F. albicollis*.** These models were created using Phyre2 and visualized in Pymol. An overlay of the structures is in the bottom right. Key: Green= *F. albicollis* Blue= *Z. albicollis* Pink= *G. gallus*

Model Alignment	RMSD with Outliers	RMSD without Outliers
Chicken versus Flycatcher	5.819	0.126
Chicken Versus Sparrow	4.509	0.163
Flycatcher Versus Sparrow	4.401	0.163

Table 5: **RMSD Values for Phyre2 Structures.** Root-mean-square-deviation measures the average distance between selected atoms. The 3-D structures were aligned and the RMSD values were measured using Pymol. Atomic outliers increase the deviation between the structures and when removed, the models are highly similar.

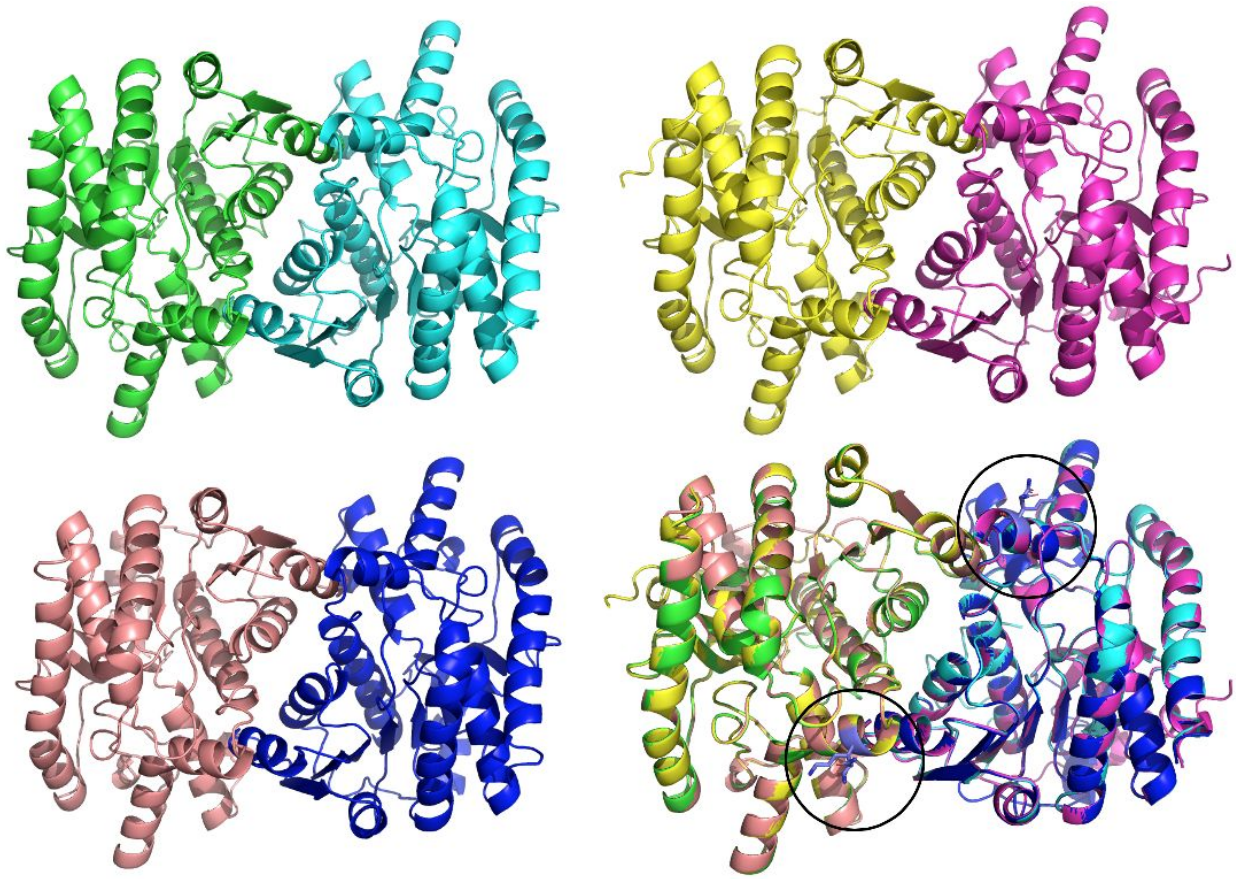


Figure 23: **Predicted Three-Dimensional Dimer Models of *G. gallus*, *Z. albicollis*, *F. albicollis*.** These models were created using Phyre2 and visualized in Pymol. An overlay of the structures is in the bottom right. All three dimer are similar in the interface, with the alpha helices being conserved. Key: Green/Cyan= *G. gallus* Blue/Rose Gold= *Z. albicollis* Pink/Yellow=*F. albicollis*

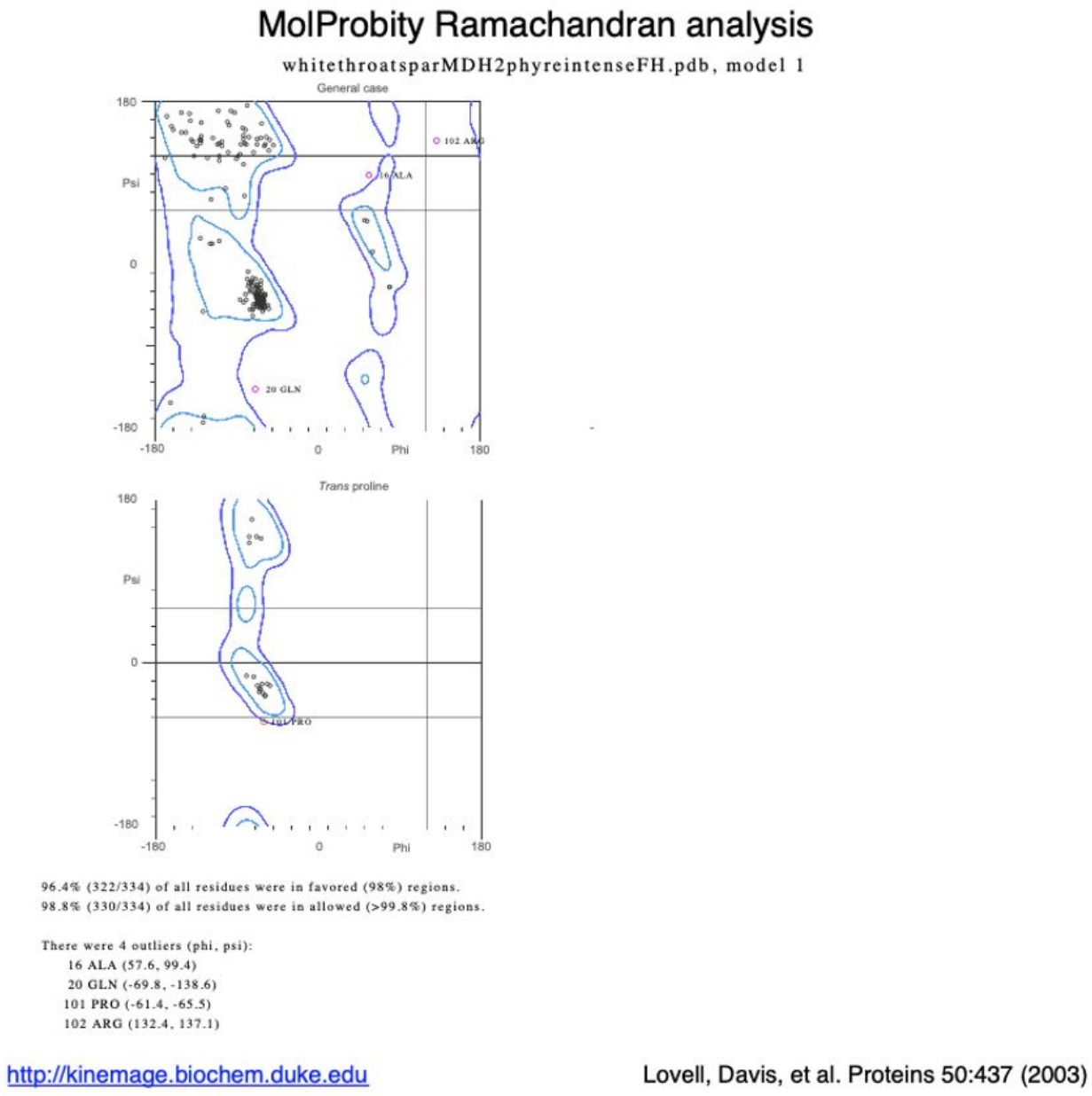


Figure 24: **Phyre2 *Z. albicollis* MDH2 model Ramachandran Plot**. The Ramachandran plot generated by MolProbity for *Z. albicollis* shows that nearly all the residues (98.8%) were in allowed regions. The Phyre2 model only has 4 outlier residues that are not at allowed angles.

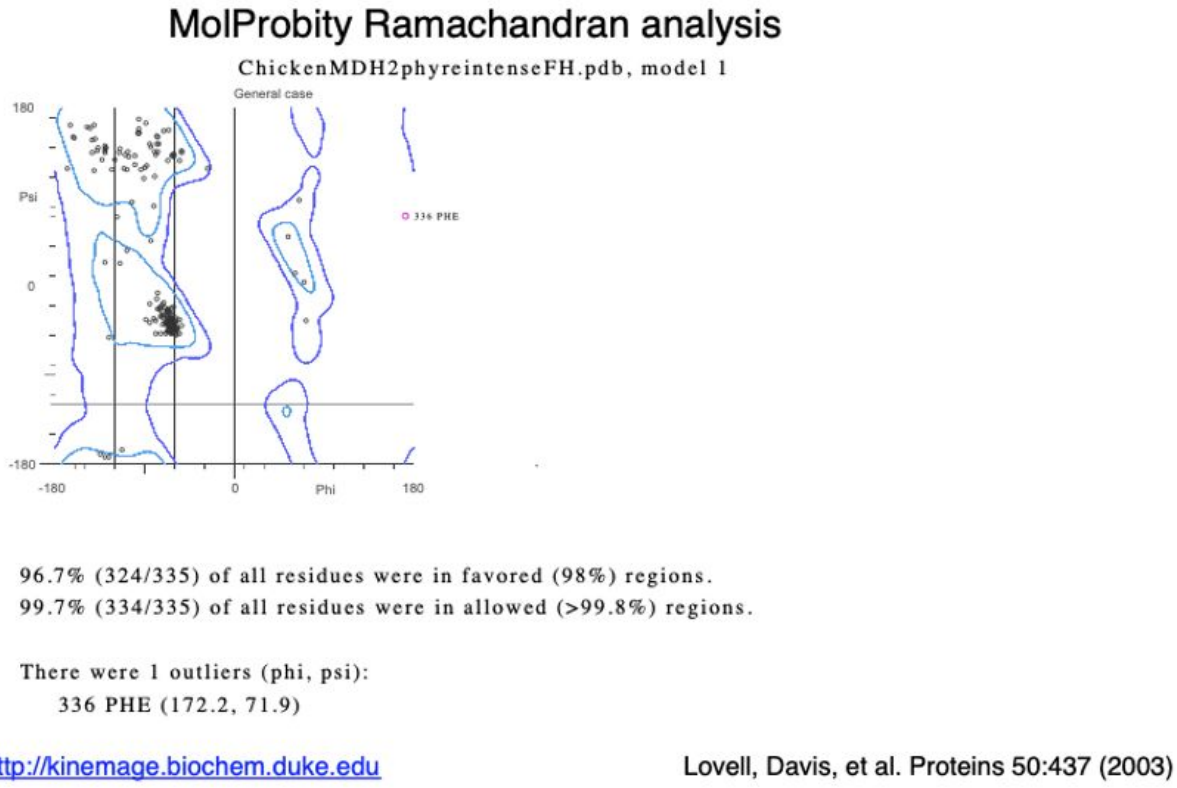


Figure 25: **Phyre2 *G. gallus* MDH2 model Ramachandran Plot**. The Ramachandran plot generated by MolProbity for *G. gallus* shows that nearly all the residues (99.7%) were located in allowed regions. The Phyre2 model only has 1 outlier residue that is not at allowed angles.

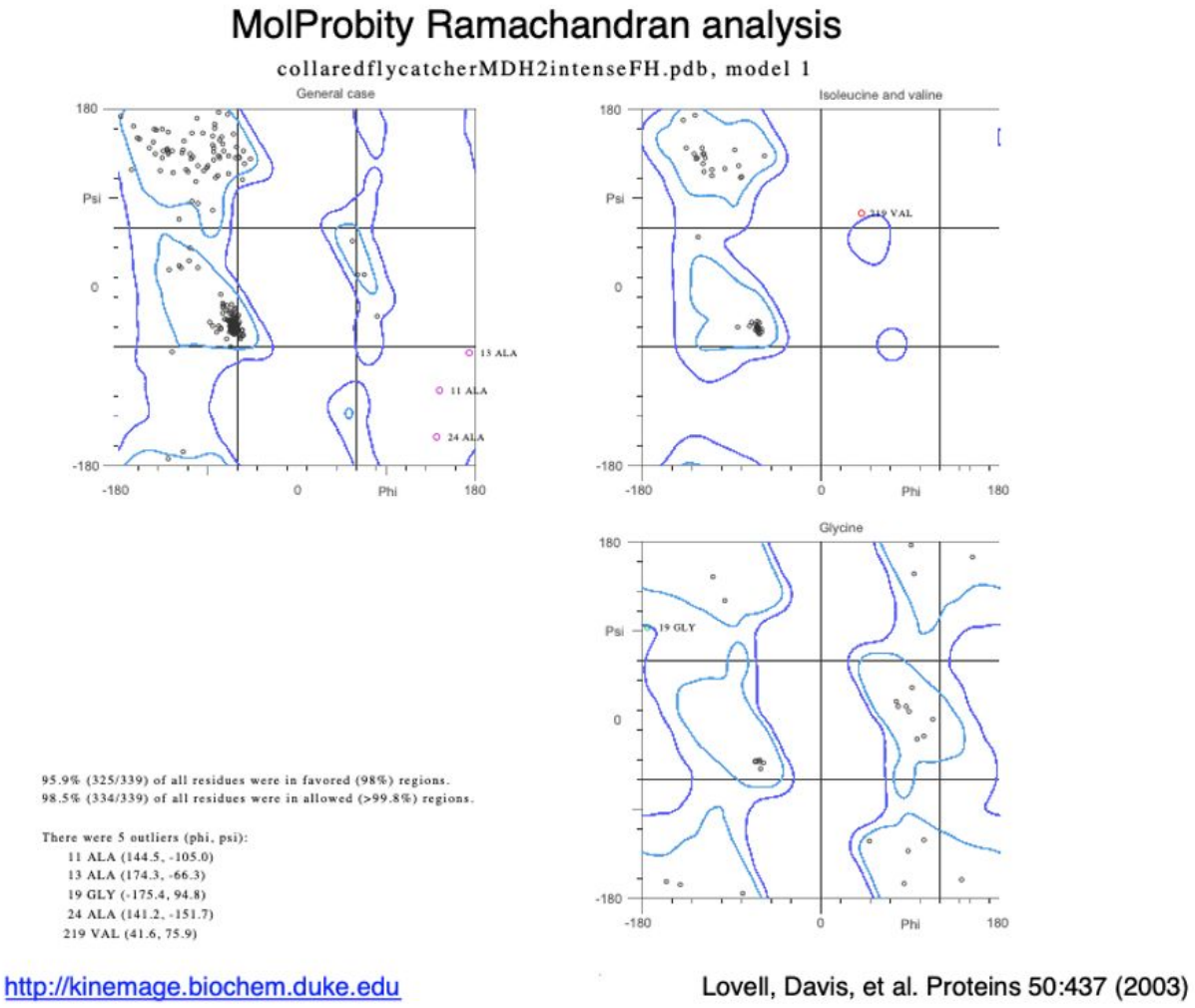


Figure 26: **Phyre2 *F. albicollis* MDH2 model Ramachandran Plot**. The Ramachandran plot generated by MolProbity for *F. albicollis* shows that nearly all the residues (98.5%) were located in allowed regions. The Phyre2 model only has 5 outlier residues that are not at allowed angles.

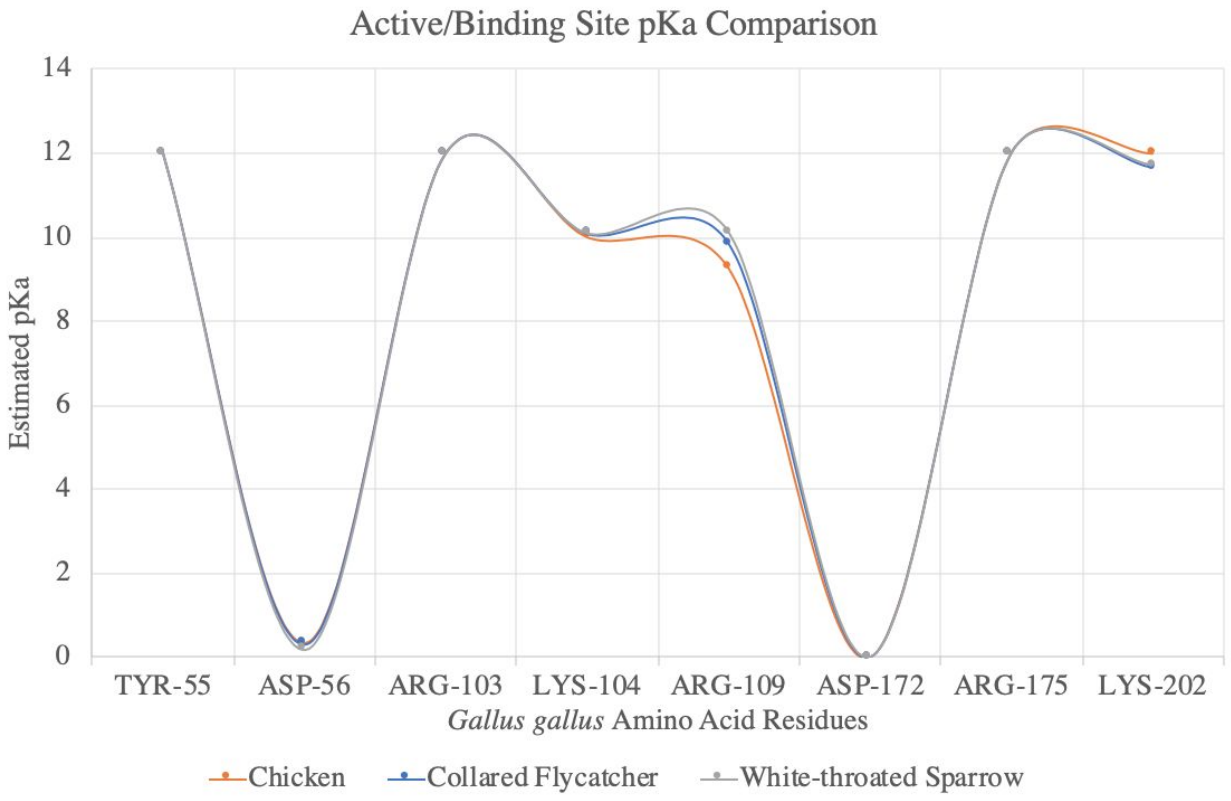


Figure 27: **Comparison of pKa Values at Active and Binding Sites.** Using the predicted active and binding sites from the alignment to human MDH2, the pKas of relevant residues of the three species were plotted. This revealed that there is some variation in pKas at Arg-109 and Lys-202.

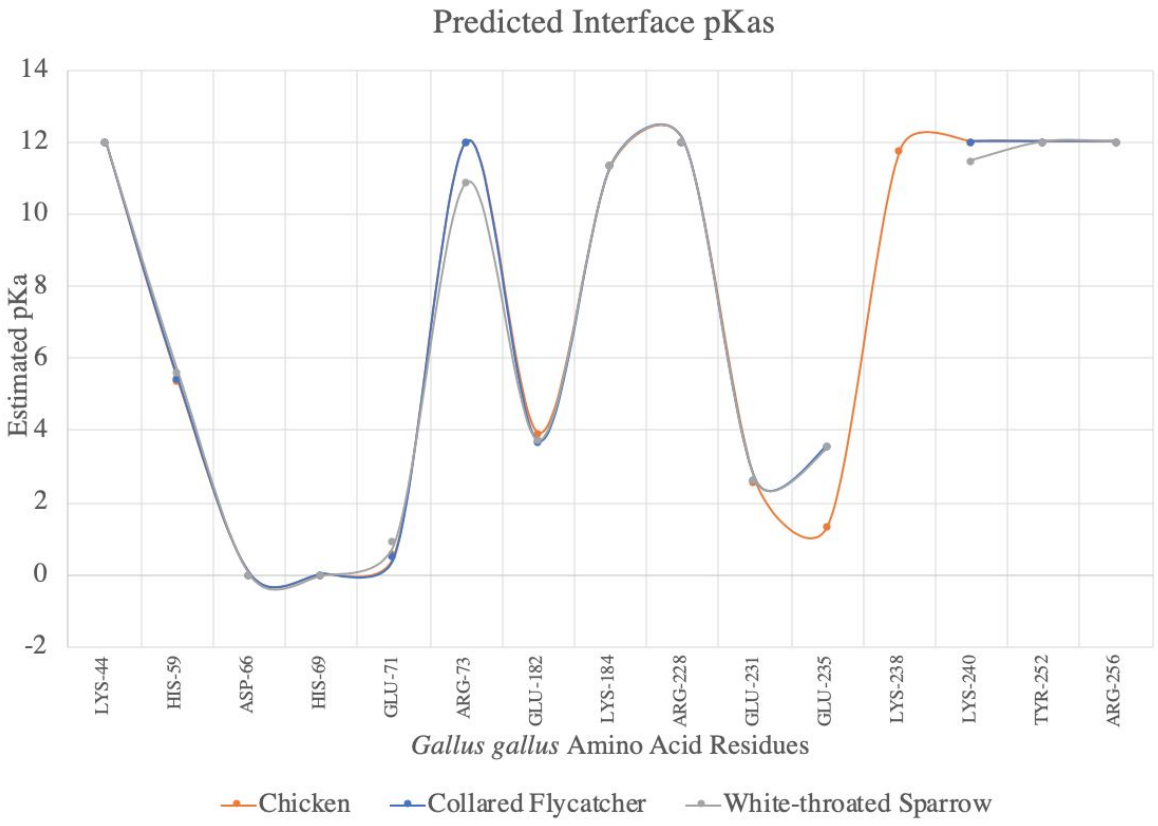


Figure 28: **Comparison of pKa Values at Interface Regions.** Using the predicted interface regions from the alignment to human MDH2, the pKas of relevant residues (labeled using *G. gallus*) of the three species were plotted. This revealed that there is some variation in pKas at Arg-73, Glu-235, and Lys-240. Additionally, the *G. gallus* has one additional residue in the dimer interface region at Lys-238.

Chapter 5: Conclusions

The creation of an expression vector for *G. gallus* mitochondrial MDH was successful. The *G. gallus* MDH2 gene was correctly inserted into the pET28(a)+ expression vector backbone. This construct creates the opportunity to study avian MDH2 in various capacities, whether that be post-translational modifications, migration, or something else. Before any studies can be conducted, however, the protein induction conductions still need to be optimized. When compared to the optimized human cytoplasmic MDH induction, the protein production is minimal. Further testing of the induction conditions would be the first step, examining the effects of lowering the temperature of growth and induction before performing any other tests. By lowering the temperature, the stability of the protein may be improved through slower protein synthesis. Once these tests are conducted, if protein expression is still low, then codon optimization is a clear next step. Codon optimization of the expression vector has been done electronically and it will require the replacement of several codons. Future research would focus on selecting the most important codons to change, perform site-directed mutagenesis to alter these codons, and then perform induction tests. If induction is successful, then *in vitro* assays could be conducted with the wild type MDH2. The K_{cat}, V_{max}, and K_m values of wild-type MDH2 can be determined. Beyond wild-type analysis, mutagenesis could be utilized to change an amino acid from the *G. gallus* sequence to one found in the other two migratory species. Comparing *in vitro* assay results could provide insight into metabolism and differences that exist as a result of varying physiological needs.

The sequence alignments between the three species revealed that the enzymes may be regulated through phosphorylation or have varied binding abilities as a result of sequence variations outside the substrate binding, active, and dimer interface sites. The conservation of the

key catalytic and binding sites likely means that all three forms of the enzyme will have similar binding abilities to substrates and in dimerization. The use of docking software to test substrate binding may provide a further understanding of how the enzymes function.

The slight variations in the 3-D structures can largely be accounted for by the lack of modeling on the beginning and end of protein sequences. This is the result of the proteins used to generate the models. The alignment between the target protein and template proteins most frequently include only the middle portions of the sequences. Around the first twenty and the final ten amino acid residues in the species we are modeling do not align with the template structures. The region between is well modeled because there is a solved structure to base the prediction on. Since there is no template to generate a predicted model, the beginning and end of the protein cannot be modeled accurately. The majority of the secondary structures are well conserved otherwise and this is further supported by the removal of outliers when calculating RMSD.

The differences in pKa data further support the conclusion that the enzyme is regulated outside the active sites and the dimer interface. Arg-109 and Lys-202 in the active sites and Glu-71, Arg-73, and Lys-240 in the dimer interface have less than one pH unit of difference between them. These pKa differences are unlikely to play a role unless there is a specific microenvironment present. Glu-235 and Lys-238 are the two residues with the most significant (>1) pKa difference between them. Lys-238 is glutamine in the migratory species, *Z. albicollis* and *F. albicollis*. Lys-238 does not appear to have much effect due to its location pointing away from the dimer interface (see Figure 24) Further analysis could be conducted by varying pH conditions of the enzyme assay and by testing the effect of mutation, in particular at Lys- 238 and Glu-235, on calculated pKas and activity.

Future opportunities for research should focus on optimizing induction with the *G. gallus* expression vector to perform mutagenesis to test the effects of the amino acid differences and phosphorylation between the nonmigratory chicken and the two migratory birds studied here. Lys-238 is a particular residue of interest given its location with the dimer interface and its role in affecting the pKa of Arg-109. The isoleucine residues and the glycine/alanine residues are also potential targets for mutagenesis, as they have an effect on Arg-109 as well. Mutating the expression vector residues to match the migratory species could provide insight into whether or not these changes affect the enzyme's activity or substrate binding ability.

In addition, phosphomimetics of the serine, threonine, and tyrosine residues both inside and outside the key sites could provide further information about the regulation of the enzyme and if phosphorylation or another possible post-translational modification plays a role. The data generated by these experiments could illuminate a path to a deeper understanding of how MDH is regulated in avian metabolism. The regulation could provide an explanation for the activity differences seen between migratory and non-migratory species. Beyond this, the physiological and metabolic differences between avian species could be further explored.

References

Addgene: Gibson Assembly Protocol. https://www.addgene.org/protocols/gibson-assembly/. Accessed November 29, 2019.

Ait-El-Mkadem S, Dayem-Quere M, Gusic M, et al. Mutations in MDH2, Encoding a Krebs Cycle Enzyme, Cause Early-Onset Severe Encephalopathy. Am J Hum Genet. 2017;100(1):151–159. doi:10.1016/j.ajhg.2016.11.014

Arndt C., Koristka S., Bartsch H., Bachmann M. (2012) Native Polyacrylamide Gels. In: Kurien B., Scofield R. (eds) Protein Electrophoresis. Methods in Molecular Biology (Methods and Protocols), vol 869. Humana Press, Totowa, NJ

Banerjee, S., & Chaturvedi, C. M. (2016). Migratory preparation associated alterations in pectoralis muscle biochemistry and proteome in Palearctic-Indian Emberizid migratory Finch, Red-headed Bunting, Emberiza bruniceps. Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics, 17, 9–25. https://doi.org/10.1016/j.cbd.2015.11.001

Birktoft JJ., Rhodes G., Banaszak LJ. Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5-.ANG. resolution. Biochemistry. 1989;28(14):6065-6081. doi:10.1021/bi00440a051

Boardman PE, Sanz-Ezquerro J, Overton IM, et al. A Comprehensive Collection of Chicken cDNAs. *Curr Biol*. 2002;12(22):1965-1969. doi:10.1016/S0960-9822(02)01296-4

Bordogna, A., Pandini, A. and Bonati, L. (2011), Predicting the accuracy of protein–ligand docking on homology models. J. Comput. Chem., 32: 81-98. doi:10.1002/jcc.21601

Brule CE, Grayhack EJ. Synonymous Codons: Choose Wisely for Expression. Trends Genet. 2017;33(4):283-297. doi:10.1016/j.tig.2017.02.001

Cox, B., Chit, M.M., Weaver, T., Gietl, C., Bailey, J., Bell, E. and Banaszak, L. (2005), Organelle

and translocatable forms of glyoxysomal malate dehydrogenase. The FEBS Journal, 272: 643-654. doi:10.1111/j.1742-4658.2004.04475.x

Dawson, N. J., Ivy, C. M., Alza, L., Cheek, R., York, J. M., Chua, B., … Scott, G. R. (2016). Mitochondrial physiology in the skeletal and cardiac muscles is altered in torrent ducks, Merganetta armata, from high altitudes in the Andes. The Journal of Experimental Biology, 219(Pt 23), 3719–3728. https://doi.org/10.1242/jeb.14271

Ganapathy G, Howard JT, Ward JM, et al. High-coverage sequencing and annotated assemblies of the budgerigar genome. Gigascience. 2014;3(1):11. doi:10.1186/2047-217X-3-11

Gold Biotechnology. Protocol IPTG Induction and Extraction of Proteins. St. Louis, Missouri; 2018.

Gordon, J., Myers, J., Folta, T., et al. "H++: a server for estimating pKas and adding missing hydrogens to macromolecules." Nucleic acids research vol. 33,Web Server issue (2005): W368-71. doi:10.1093/nar/gki464

Goward CR, Nicholls DJ. Malate dehydrogenase: A model for structure, evolution, and catalysis. Protein Sci. 1994;3(10):1883-1888. doi:10.1002/pro.5560031027

Huang Y., Li Y., Burt DW., et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat Genet. 2013;45(7):776–783. doi:10.1038/ng.2657

Hudson LN., Isaac NJ., Reuman DC. The relationship between body mass and field metabolic rate among individual birds and mammals. J Anim Ecol. 2013;82(5):1009–1020. doi:10.1111/1365-2656.12086

Kakizawa, R., Kuroda, N., & Utida, S. (1982). Evolution of Mitochondrial Malate Dehydrogenase in Birds (II). 山階鳥類研究所研究報告, *14*(2–3), 131–136. https://doi.org/10.3312/jyio1952.14.131

Kelley LA., Mezulis S., Yates CM., Wass MN., Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845–858. doi:10.1038/nprot.2015.053

Kufareva I, Abagyan R. Methods of protein structure comparison. Methods Mol Biol. 2012;857:231–257. doi:10.1007/978-1-61779-588-6_10

Kuroda, N., Kakizawa, R., Hori, H., Osaka, Y., Usuda, N., & Utida, S. (1982). Evolution of Mitochondrial Malate Dehydrogenase in Birds. *Journal of the Yamashina Institute for Ornithology*, *14*(1), 1–15. https://doi.org/10.3312/jyio1952.14.1

Lehninger, AL., Cox, MM., Nelson, DL. Lehninger Principles Of Biochemistry. New York : W.H. Freeman, 2017. Print.

Minárik, P., et al. "Malate Dehydrogenases-Structure and Function." Gen. Physiol. Biophys, vol. 21, 2002, http://www.gpb.sav.sk/2002_03_257.pdf.

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. XP_014128222.1, Zonotrichia albicollis malate dehydrogenase, mitochondrial, protein; [cited 2020 03 17]. Available from: https://www.ncbi.nlm.nih.gov/protein/XP_014128222.1?report=genpept

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. XP_005056570.1, Ficedula albicollis malate dehydrogenase, mitochondrial, protein; [cited 2020 03 17]. Available

from: https://www.ncbi.nlm.nih.gov/protein/XP_005056570.1?report=genpept

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. XP_415765.2 Gallus gallus malate dehydrogenase, mitochondrial, protein; [cited 2020 03 17]. Available from: https://www.ncbi.nlm.nih.gov/protein/XP_415765.2?report=genpept

Protein Expression - E.Coli - Optimisation of Expression Levels - EMBL. https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/optimisation_expression_levels/. Accessed 11 Apr. 2020.

The PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC.

The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D506–D515, https://doi.org/10.1093/nar/gky1049

Roderick SL, Banaszak LJ. The Three-Dimensional Structure of Porcine Heart Mitochondrial Malate Dehydrogenase at 3.0-A Resolution*. Vol 261.; 1986.

Shapiro MD, Kronenberg Z, Li C, et al. Genomic diversity and evolution of the head crest in the rock pigeon. Science (80-). 2013;339(6123):1063-1067. doi:10.1126/science.1230422

Taylor AB, Smith BS, Kitada S, et al. Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. Structure. 2001;9(7):615-625. doi:10.1016/S0969-2126(01)00621-9

Ugochukwu, E., et al. “Crystal Structure of Human Malate Dehydrogenase Type 2.” TO BE PUBLISHED, doi:10.2210/PDB2DFD/PDB.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., III,

Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S. and Richardson, D.C. (2018), MolProbity: More and better reference data for improved all‐atom structure validation. Protein Science, 27: 293-315. doi:10.1002/pro.3330