A Structural Study of Fumarase C from *E. coli*

Adam Oken[1], Douglas Juers[1] and Dr. Paul A. Sims[2]

1. Program of Biochemistry, Biophysics, and Molecular Biology, Whitman College 2. Department of Chemistry and Biochemistry, University of Oklahoma

Abstract:

Fumarase C (fumC) catalyzes the interconversion of fumarate to L-malate as part of the metabolic Krebs's cycle. Previous studies of fumC from *Eschericha coli* have noted two dicarboyxlate binding sites known as the A site and the B site. Structures have been published for the apo enzyme, the enzyme with inhibitors bound in the active site, and for two mutagenesis experiments. These studies have shown the A site to be the catalytic site within the enzyme and the B site able to bind substrate. Not much is known about the B site, however it is thought to be an activator site from kinetic experiments. Here, we have attempted to continue the structural analyzes of fumarase C in order to elucidate the function of the B site as well as bind the substrate, L-malate, in the catalytic A site in order to confirm the key active site residues and two catalytic bases in the proposed mechanism.

Introduction:

The citric acid cycle, otherwise known as the Krebs cycle, is one of the many significant metabolic pathways in humans and other organisms. The enzyme fumarase, within the citric acid cycle, has been linked to a progressive metabolic encephalopathy. Specifically, an inborn error of the gene, a missense mutation at position 955 causing a G to C transversion, results in an inactivation of the protein due to a single amino acid substitution. In result, the inactive protein causes psychomotor retardation, microcephaly, and abnormal posture with hypotonia contrasting with hypertonia of limbs in humans[8]. Understanding fumerase could lead to cures or interventions for this human disease in the near future. The goal of this research is to build upon current research and characterize the structure of fumarase's

active site so that we can begin to understand the kinetic attributes and mechanistic details of this important protein.

The citric acid cycle is a vital process in all aerobic organisms that releases stored energy through the oxidation of acetyl-CoA, its substrate. The cycle generally receives the acetyl-CoA from the breakdown of glucose, but the substrate can be obtained via other carbohydrates, proteins, and fats as well. In figure 1, the complete citric acid cycle is shown as well as the pyruvate dehydrogenase reaction[9]. The citric acid cycle is an important biological process because it generates guanosine triphosphate (GTP), flavin adenine dinucleotide ($FADH_2$) and nicotinamide adenine dinucleotide (NADH). These three molecules are used in the creation of adenosine triphosphate (ATP), the universal energy source for the cell. In one completion of the citric acid cycle, the starting products acetyl CoA and oxaloacetate are reduced and then regenerated as the final products. As the ten reactions in the Krebs cycle progress, two molecules of CO_2, one molecule of GTP, 3 molecules of NADH, and one molecule of $FADH_2$ are generated. After being produced, the NADH and $FADH_2$ are fed into the oxidative phosphorylation pathway otherwise known as the electron transport pathway and eventually made into ATP. Approximately 10 ATP molecules are produced during each round of the citric acid cycle (2.5 for each NADH and 1.5 for each $FADH_2$). This process occurs in the matrix of the mitochondria for eukaryotes and in the cytosol for prokaryotes. In this study we are interested in one of the eight enzymes used in the citric acid pathway.

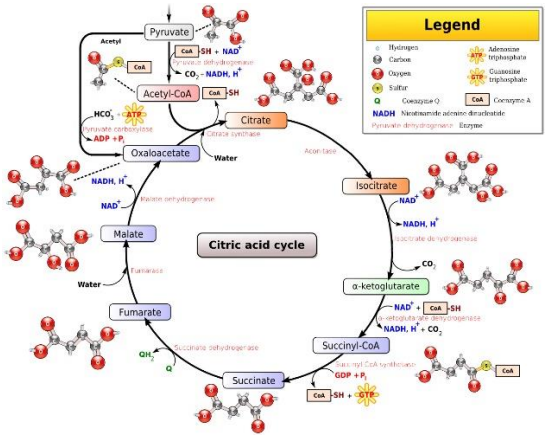
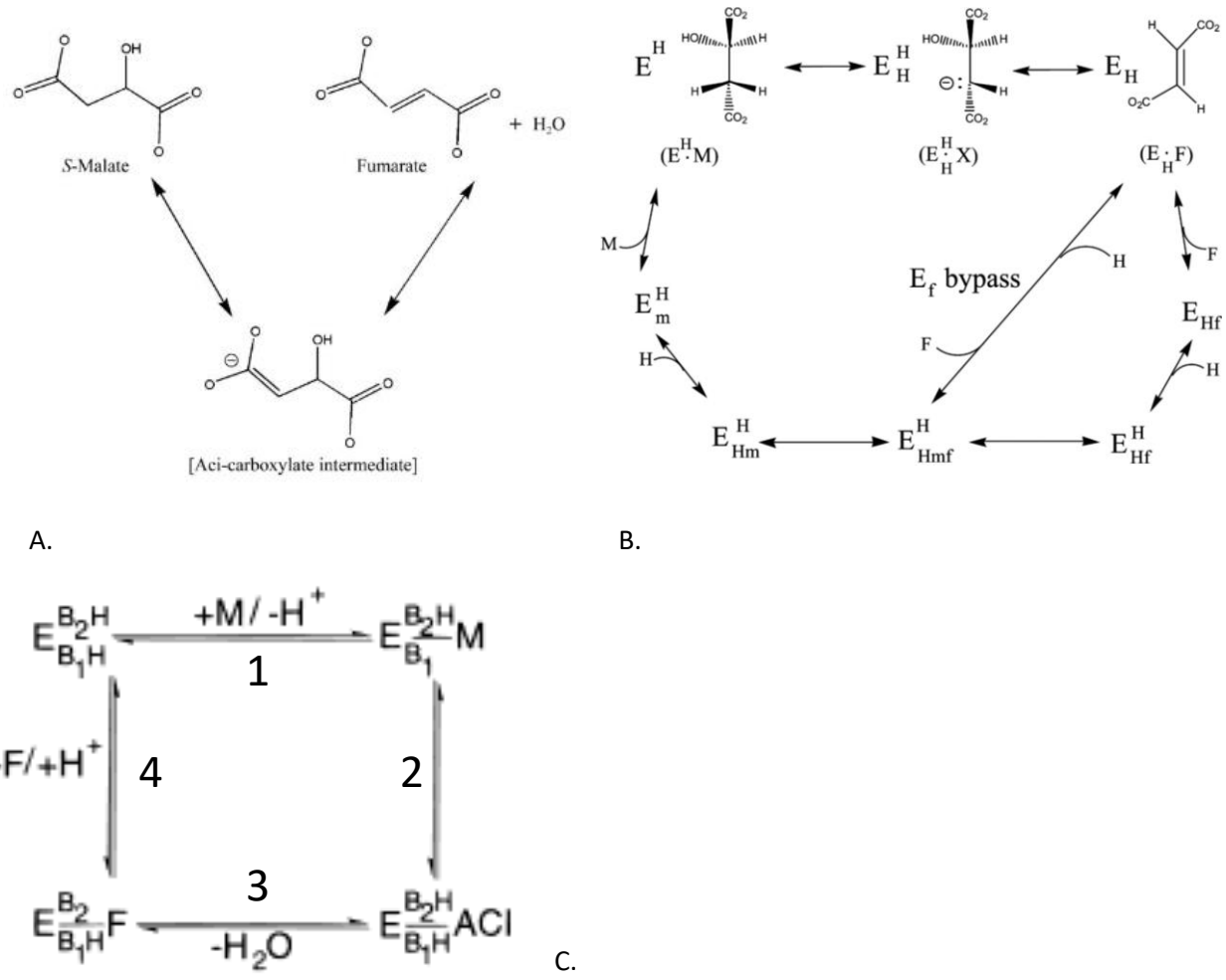


Figure 1: The Citric Acid Cycle. This figure show the entire Citric Acid Cycle with all the enzymes, substrates and products as well as the reaction of pyruvate to acetyl CoA.

In the citric acid cycle, fumarase catalyzes the reversible hydration/dehydration reaction between fumarate and malate. It is predicted that the mechanism is an acid-base catalyzed elimination with an aci-carboxylate intermediate (figure 2) which requires two basic amino acids in the enzyme[10]. In figure 2.B the two residues are labeled A and B. In figure 2.C they are labeled B_1 and B_2. One of the residues, B_1, is thought to be a –COOH group because it has a pK which is not temperature dependent. Since the other residue, B_2, has a ΔH of ionization similar to an imidazole ring, it is believed to be a histidine. The active site includes an aspartate and a histidine, but these residues have not yet been proven to be the two catalytic bases. Continuing to the mechanism, to create fumarate, the enzyme initially loses a proton from B_1 into solution as malate binds (step 1, figure 2.C). Next, abstraction of the malate C3 proton by base B_1 forms the unstable carbanion (step 2, fig. 2.C). The resulting carbanion is stabilized as the *aci*-carboxylate (or enediolate) intermediate[11]. Rearrangement of the *aci*-carboxylate intermediate along with proton donation from the second catalytic base (B_2H) leads to the removal of the hydroxyl group from the C2 position in the form of H_2O (step 3, fig. 2.C). Finally, a proton is added back from surrounding water to the enzyme and fumarate is released (step 4, fig. 2.D.). Figure 2.B shows the mechanism for the catalysis in the direction of formation of fumarate. The K_{eq} for the reaction is 4.4 at 25^o C and pH 7.3, in the direction of malate formation[6]. Additionally, it has been shown that the reaction rate of malate to fumarase is increased more than expected with a high substrate (malate) concentration[6]. At low substrate concentration the enzyme exhibits Michaelis-Menten kinetics while at substrate concentrations 5 times K_m, fumarase exhibits substrate activation. This activation via high substrate concentration will be further discussed in the structural studies of fumarase.



A. B. C.

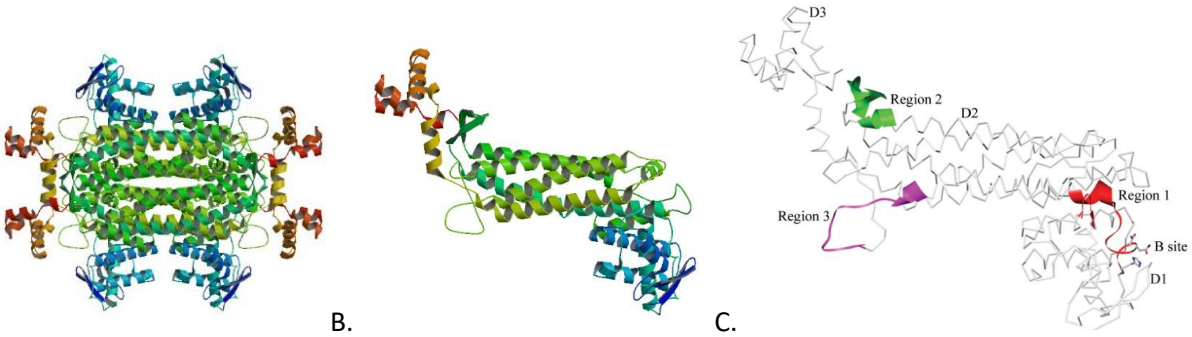
Figure 2: Reaction scheme and mechanism of Fumarase enzyme. Figure 2.A shows the scheme of the reversible reaction from malate to fumarate and water catalyzed by fumarase[1]. Importantly, it also shows the aci-carboxylate intermediate that is formed in the reaction. Figure 2.B shows the arrow pushing reactions for the catalysis in the direction of formation of malate. [12]. Figure 2.C is another view of the mechanism where ACI is the aci-carboxylate intermediate, M is malate, F is fumarate, and E is the enzyme with basic residues B1/B2[10]. This view of the mechanism clarifies the protonation states of the two basic residues within fumarase as the reaction occurs.

There are two classes of fumarase enzymes, class I and class II, both of which are found in *E. coli*. Class I fumarases are heat liable, iron dependent, 4 Fe- 4 S cluster containing dimeric proteins of 120 kDa[7]. Two examples of class I fumarases are fumarase A and fumarase B. In contrast, class II fumarases

are homotetrameric, thermostable, and iron-independent proteins with a molecular mass of around 200 kDa[11]. Additionally, there are numerous orthologs of fumarase from different organisms. Each member of the class II fumarase family can be identified via a GSSxMxKxnxxPxE specific sequence located between Gly317 and Glu331 (as numbered from the fumarase enzyme from *E. coli*) as well as increased homology at two other regions[10]. In this structural study, the specific class II fumarase protein to be analyzed is Fumarase C, otherwise known as fumC from *E. coli*. Each subunit of the fumC enzyme is 467 amino acids long.

Previous research has been conducted on fumarase C from *E. coli* to characterize the structure of the protein[1,3,8, and 10], determining the active sites as well as some important residues within the enzyme. In 2004, the structure of apo fumarase C was solved by Todd Weaver[1]. This is one of four key structural studies that have provided information about fumarase. As shown in the ribbon diagrams below, the quaternary structure of fumarase is a tetramer in its active form (fig 3.A). The monomer has three different domains: Domain 1, Domain 2, and Domain 3 (Fig 3.B). Domain 2 comprises a five-helix bundle that forms a 20-superhelix structure upon tetramer formation. The other two domains, Domain 1 (right end) and Domain 3 (left end), are outshoots from the Domain 2 core. Additionally, there are three highly conserved regions within the superfamily of fumarases that have high sequence identity with other fumarase proteins (region 1, region 2, and region3) (Fig 3.C). Region 1 in Domain 1 consists of the residues Thr96-Thr100 and His 129-Thr146 which are colored red in figure 3.C. Region 2 in Domain 3 consists of residues Gly185-Glu200 which is colored green in figure 3.C. Region 3 in Domain 2 consists of residues Gly317-Glu331 which is colored magenta in figure 3.C. Parts of these conserved regions make up the two active sites of fumarase. The two dicarboxylate binding sites are located 12 angstroms apart in the protein. The A site or active site is formed by residues from regions 1-3 with each region coming from a different subunit of the tetrameric protein. The A site also contains a water molecule within the pocket. The second site or B site is formed exclusively from residues in one subunit of the tetramer, Arg

126-Asn135. The A and B sites in the protein have different effects on the enzyme that are relevant to its activity.



A. B. C.

Figure 3: Previous structural studies of Fumarase C. Part A shows the monomeric form of fumarase with the three domains colored slightly differently. In red/orange is domain 3, in green is the 5 helixes of domain 2, and in blue is domain 1. Part B shows the tetramer form of fumarase with the 20-superhelix structure in the center. Part C shows the three regions that are highly conserved between the superfamily of fumarases in different colors. Red for region 1, green for region 2, and magenta for region 3. Additionally, the three domains of the fumarase monomer and the B site are labeled in the diagram. Note, the A site isn't labeled because it is made up of residues from three different monomers of the tetrameric protein.

The question regarding which of the two active sites was the catalytic site was ansered in 1996 with the second of four key structural studies of fumarase[3]. It is believed that the A site is the catalytic site because no catalytically active monomeric form of the protein has ever been described. Since the A site is a collection of regions from three different monomers, it supports the aforementioned conclusion. To test this hypothesis, the activity of the two sites was determined via mutagenesis, replacing a histidine in each active site to an asparagine[3]. Specifically, the mutation at the B site was H129N and at the active site (A site) was H188N. Starting with the B site, the results from the first experiment showed the H129N mutant had little or no effect on the catalytic activity of the enzyme. Additionally, the H129N mutant also revealed that dicarboxylic acids (like malate) no longer bound to

the B site. For the other protein mutant, the H188N mutation showed a large decrease in catalytic activity. Also the water molecule in the A site was moved approximately 0.70 angstroms when compared to the wild-type. Thus, since a mutation in that A site directly affects the catalytic activity of the enzyme, it confirms that the active catalytic site is indeed the A site. The function of the B site is still not fully known, but it is thought to be an activator site that increases catalytic function when occupied by malate. This is because it has been shown that with increased concentrations of malate present, there is an increase in the rate of reaction. Thus, malate binding in the B site could affect the activity of the A site through intramolecular interactions or slight conformational shifts.

Previous structural studies of co-crystals between fumarase and other molecules indicate important residues in the A site and B site. These molecules are inhibitors and include citrate, pyromellitic acid, and 3-trimethylsilylsuccinic acid. These experiments are the third important structural study in providing key information about fumarase. Due to the nearly identical characterized structures of the three inhibitors, we will solely focus on citrate. More generally, by co-crystalizing a protein with its substrate or an inhibitor, the active site as well as the important catalytic residues and non-covalent interactions in the active site can be determined.

In the previously solved structure of fumarase C co-crystalized with citrate, a strong competitive inhibitor, the key residues in the active site and B site were elucidated[10]. The key residues in the catalytic A site interact with the two carboxylates of citrate. From figure 4.A, it is evident that residues K324, T187, and N326 interact with the first carboxylate, and N326 and T100 interact with the second. Additionally, residue H188 is positioned between the two carboxylate groups so it can interact with both. There are two other notable aspects of the residues in the A site. The first notable aspect is E331c forms an ion pair with H188d which resembles a charge-relay system (the c and d notation indicate which monomer of the protein the residues are located in). A charge relay system is a series of amino acids that are next to each other and activate a nucleophile to perform covalent catalysis. In this case,

the glutamate would hydrogen bond to the histidine, making the imidazole nitrogen a strong base which can activate a nucleophile. As evident in figure 4.A, there isn't a specific residue that is clearly the activated nucleophile, but there is an alcohol on T187d that could be activated. The activated basic hydroxyl group could correspond to the first base (B_1) in Figure 2C since it could extract a proton from the substrate. The second notable aspect is the binding of residues S140b and N141b. This is important because this region leads into the B site generated by residues 126-132. This may lead to interactions between the A and B sites. Looking at the B site of fumarase, the residues interact with the carboxylate groups of malate or β-(trimthylsilyl)-maleate (TMSM) in the same manner that the carboxylate groups interact with that of citrate in the A site as seen in figure 4.A. The important residues are N131, D132, R126, and H129 with the first two interacting with one carboxylate and the last two interacting with the other carboxylate. There are two residues of importance in the B site, residues His129 and Asn131. These two residues have different conformations in the presence or absence of a small molecule in the B site. With the presence of a small molecule in the B site (malate) His129 rotates out of the B site and contacts Asp133 and Arg 126. When in the free state, His129 rotates into the B site preventing a small molecule from binding. The two different confirmations of the B site residues can be seen in figure 4.B where the free and occupied structures are overlaid upon each other. Thus, the important catalytic residues of the A and B sites have been determined by binding competitive inhibitors to the active site to view the interactions.

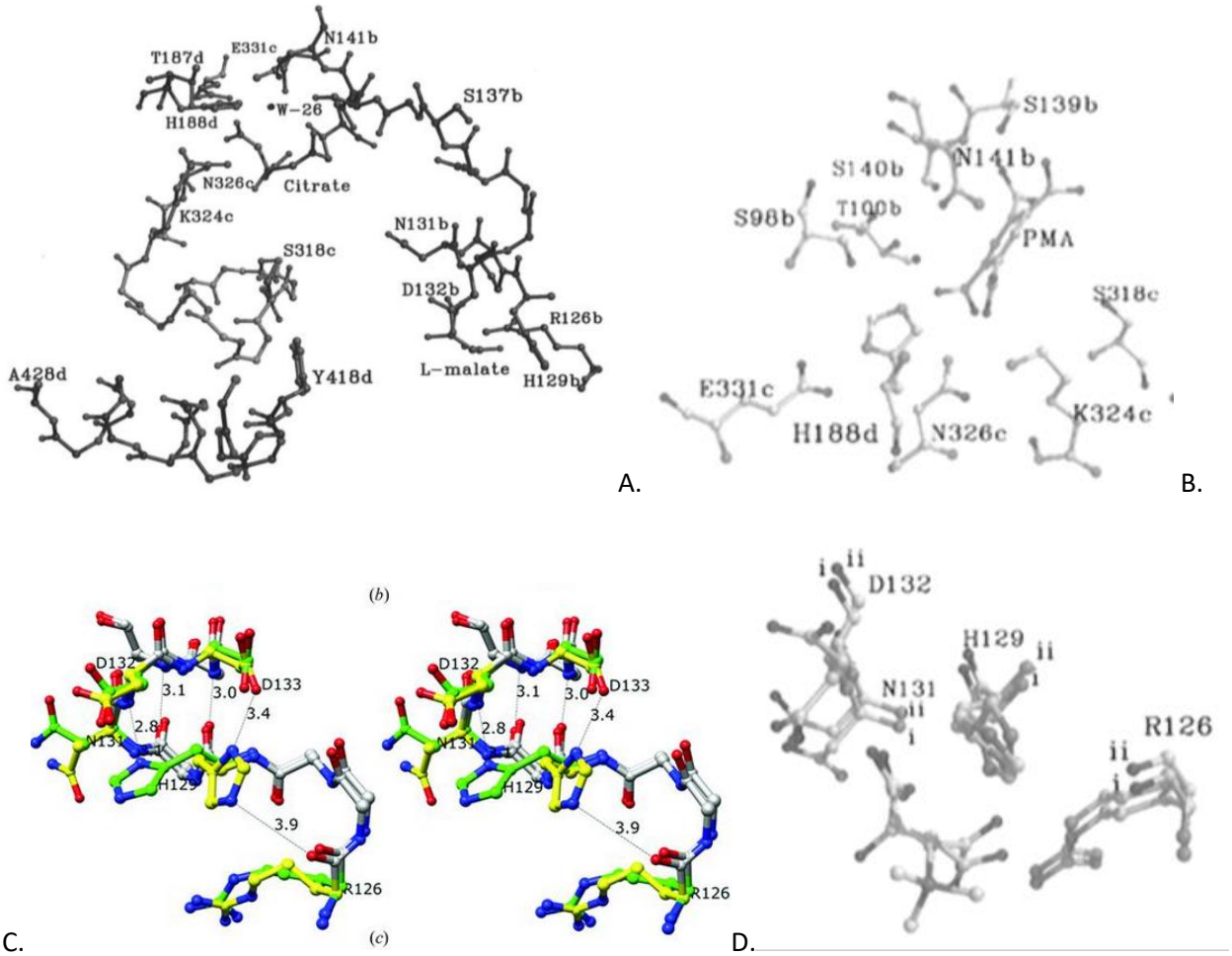
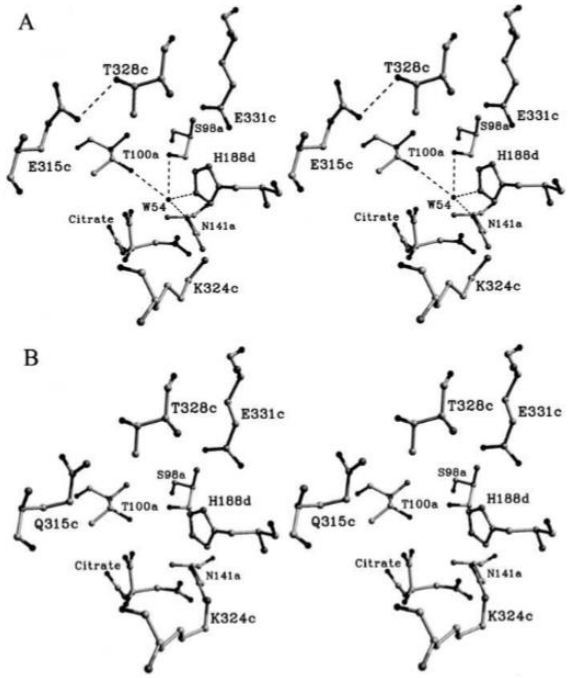


Figure 4: Fumarase C with citrate (A) bound in the active site with malate modeled into the B site. Part A shows the ball and chain structure of the A site with the inhibitor citrate bound and the B site with malate modeled in. Part B shows the stereo drawing that illustrate the A site of the enzyme with PMA, a competitive inhibitor, in the active site. Part C shows the two different conformations between the free and occupied B sites. Note the large rotation of H129 between the two overlaid structures. The free FumC was crystallized from a solution of 50 m*M* MOPS pH 7.5, 100 m*M* $LiSO_4$ and 12%(*w*/*v*) PEG 4000. In contrast, the occupied crystals were grown from a solution of 300 mM citrate, pH 6.0, and 14% PEG 4000. Part D shows the stereo drawling illustrating the native and occupied B site of the enzyme with β-(trimethylsilyl) maleate. Note the movement of the side chain of D132 in this image.

The last of the four important structural studies of fumarase is another mutagenesis experiment where researchers mutated glutamate 315 to a glutamine[8]. This study was performed because two human sisters with progressive metabolic encephalopathy were linked to a missense mutation in their fumarase gene. Specifically, the mutation E315Q in *E. coli* is equivalent to the same missense mutation in the human fumarase they had. The location of the mutation of residue 315 is very close to the A site of the enzyme which could affect the activity of the enzyme. Upon kinetics testing, there was essentially no change in the K_m value[8]. However, the K_{cat} decreased approximately 10 fold in both directions of catalysis. Both rate values can be seen in Figure 5B. Additionally, the structural study revealed two important differences between the mutated and wild-type enzymes which can be seen in Figure 5A. The first is the absence of a hydrogen bond between E315 and T328. The second difference is the unexplainable lack of the conserved and highly coordinated water molecule in the A site. These two specific differences in the mutant's active site resulted in the dramatic decrease in the rate of catalysis. Thus, this study shows the importance of the hydrogen bond between E315 and T328 as well as the more significant, highly conserved active site water molecule.



A.

Enzyme	Native	E315Q
S-malate → fumarate		
V_{max} (μmol substrate/min/mg enzyme)	178.6	16.11
K_m (mM)	0.857	0.885
k_{cat} (s^{-1})	595.2	55.32
k_{cat}/K_m (M^{-1} s^{-1})	6.95E5	6.25E4
Fumarate → *S*-malate		
V_{max} (μmol substrate/min/mg enzyme)	344.8	32.2
K_m (mM)	0.207	0.248
k_{cat} (s^{-1})	1149	107.1
k_{cat}/K_m (M^{-1} s^{-1})	5.56E6	4.32E5

B.

Figure 5: Stereoview Comparisons between wild-type and E315Q fumarase. Part A shows the two structural differences between the wild-type and E315Q mutant. For reference, the structure A is wild-type and B is the E315Q mutant. Note the hydrogen bond between E315 and T328 is gone as well as the black dot indicating the conserved active site water molecule is gone from structure A to B. Part B show the kinetics between the wild-type and E315Q mutant. Note the K_{cat} is around a 10 fold decrease between the two enzymes.

An important aspect of the fumarase enzyme is the conserved and highly coordinated water molecule in the A site. The water molecule, W-26 in the native and occupied as seen in Figure 4, is trapped in the catalytic A site and is coordinated with four different residues within the enzyme; Ser98, Thr100, Asn141, and His188. As shown in the last structural study, without the water molecule the K_{cat} decreased by approximately 10 fold. This greatly reduces the catalytic acitivty of fumarse. Aditionally, the studies with the competitive inhibitor citrate bound and the free strucutre of fumarase both had the water molecule in the A site. This suggests that the conserved water molecule in the A site is important to the activity of the enzyme. However, further research is needed to determine the specific affect the water has on the reversible reaction catayzed by fumarase.

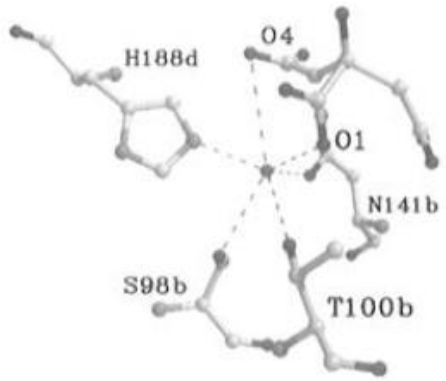


Figure 6: The coordination of the active site water molecule. This conserved water molecule in the active site is coordinated to four different residues for a total of six non-covalent interactions. The water

interacts with His188, Ser98, Thr100, and Asn141. Note His188 interacts with the competitive inhibitor citrate in the A site as well as the active site water molecule.

In this paper, we have completed another structural study of fumarase C from *E. coli* in order to elucidate the active sites and mechanism of the enzyme further. There are currently no published structures with L-malate bound in the A site of the native enzyme. Additionally, the only structure with malate in the B site has a competitive inhibitor in the A site. The goal of our experiment was to bind the substrate, L-malate, in the catalytic A site as well as the B site of fumarase in order to determine the specific interactions between the protein and its substrate. This structure will confirm the active residues that were discussed in the structural studies with competitive inhibitors and the two basic residues that are required for the proposed mechanism. Additionally, this experiment will hopefully give more insight into the conserved and highly coordinated water molecule in the A site. By determining the waters role in the catalytic activities of the enzyme, this study could explain the dramatic effects corresponding to the water's absence. As shown in the last structural study, this water plays a crucial role in the rate of catalysis. Without it, there are serious clinical implications of an inefficient fumarase enzyme which causes genetic disease, specifically a progressive psychomotor retardation in humans. Finally, since there is an increase concentration of malate, the B site of fumarase will also likely be coordinated with malate. This would provide more information about the activation of the reaction caused by the presence of high concentrations of malate. Altogether, by further characterizing the structure of fumarase's active site we will be able to explore the clinical affects, kinetic attributes, and mechanistic details of this important protein.

Methods:

Plasmid Cloning by PCR

The *E. coli* K-12 template DNA was obtained from the Blattner Lab, University of Wisconsin-Madison. Polymerase Chain Reaction, otherwise known as PCR, was then used with unique primers to

add a C terminus His tag and amplify the *fumC* gene. The primers to add the C-terminal His tag were as follows: forward primer: T7Cter_Ec_fumC_fw: 5' GAA GGA GAT ATA CAT ATG AAT ACA GTA CGC AGC GAA AAA G 3' and reverse primer: T7Cter_Ec_fumC_rv: 5'GTG ATG GTG GTG ATG ATG ACG CCC GGC TTT CAT ACT GCC 3'. The master solution for nine individual 25 µL samples contained: 45 µL of reaction buffer, 1.125 µL of each DNTP at a concentration of 10 µM, 11.25 µL of 10 µM forward primer, 11.25 µL of 10 µM reverse primer, 2.25 µL of polymerase, 103.5 µL water, 2.25 µL template DNA, and 45 µL of GC enhancer. The solution was distributed into nine PCR tubes and the amplification program run with a temperature gradient of 50-65 ^{o}C set for the annealing phase. Specifically, the PCR program had seven steps that amplified the DNA: 1) 98^{o}C for 30 seconds, 2) 98^{o}C for 10 seconds, 3) gradient 50^{o}C-65^{o}C for 30 seconds, 4) 72^{o}C for 45 seconds, 5) repeat steps 2-4 20 times, 6) 72^{o}C for 5 minutes, 7) hold at 4^{o}C forever.

2 µL of the aforementioned amplified DNA sequence was transformed into 100 µL of competent 10-G *E.coli* cells. The cells were placed on ice for 30 minutes, shocked in a 42^{o}C water bath for 45 seconds, and then immediately placed on ice for 2 minutes. The cells were brought to a total volume of 0.35 mL with the addition of Super Optimal Broth with catabolite repression, otherwise known as SOC broth (2% w/v tryptone, 0.5% w/v Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and $_{dd}H_2O$), and placed in an incubation shaker at 37^{o}C and 200 rpm for one hour. Next, 50 µL of the cells were spread on agarose plates containing kanamycin and left to grow overnight in a 37^{o}C incubator. Colony PCR was used to detect bacteria that had taken up the vector with the *fumC* C terminus His tag gene. Seven colonies from the above incubated plate were chosen and amplified with PCR. From the previous gradient PCR reaction, an annealing temperature of 62 ^{o}C was selected. The PCR program was as follows: 1) 94^{o}C for 2 minutes, 2) 94^{o}C for 30 seconds, 3) gradient 62^{o}C for 30 seconds, 4) 72^{o}C for 2.5 minutes, 5) repeat steps 2-4 24 times, 6) 72^{o}C for 5 minutes, 7) hold at 4^{o}C forever. From the results of

the colony PCR reactions, colony seven from the agar plate was swabbed into two different vials and grown up in a 5 mL culture of LB broth.

Purification

A miniprep (mostly homemade, but the spin columns are from Epoch Life Science's and the alkaline protease is from Pro-mega) was used to purify the plasmid DNA that contained the *fumC* C terminus His tag gene. Two minipreps were done side by side, one for each of the two growths. From the 5 mL culture above, 1.5 mL was extracted, spun down at 16300 x G for one minute, and the supernatant removed. This step was repeated 3 times, so in total 4.5 mL of the culture was used. The pellet was then suspended in 250 µL of resuspension solution and vortexed so the cells were dislodged. Next 250 µL of a cell lysis solution were added and the mix inverted four times. Protease (10 µL)was added, inverted four times to mix, and incubated for five minutes at room temperature. 350 µL of a neutralization solution were added and inverted four times to mix. This mixture was spun down at max x G for 10 minutes. After the spin, the supernatant was inserted into a spin column and spun at max x G for one minute. 750 µL of wash solution were added and spun down at max for one minute. Another 250 µL of wash solution was added and spun down at max for one minute. The spin column was then transferred to a new sterile 1.5 mL tube, 100 µL of nuclease free water was added, and then spun down at max for one minute. This procedure resulted in the purified *fumC* C terminus His tag gene. Another PCR reaction was used to amplify the two purified plasmid DNA samples. A restriction enzyme digest for each of the two samples of pure amplified plasmid DNA was performed to ensure the gene of interest was cloned into the plasmid in the correct orientation. Two restriction enzymes, EcoNI and HindIII from New England Biolabs, were used to digest the two purified DNA plasmid.

The purified and confirmed DNA plasmid containing the *FumC* C terminal His tag was transformed into BL21 cells for overexpression of the protein. The procedure was almost the same as described above with the transformation of the DNA into the competent 10-G cells. The only difference

is that 150 µL of cell were plated on agarose and kanamycin. After the transformation, instead of growing a 5 mL culture, one culture/colony from the plate was added to 50 mL of LB, 19 g/L lactose and 0.03 g/L kanamycin. This was placed in an incubation shaker at 37°C and 200 rpm for about seven hours. After the incubation time, the starter culture was split into two 1 L volumes of LB, lactose, and kanamycin which was left overnight in an incubation shaker at 37°C and 200 rpm. After the cultures grew overnight, they were centrifuged at 10,000 rpm for 20 minutes. The supernatant was discarded and the pellet containing the cells was flash frozen in liquid nitrogen and placed in a -20 °C freezer until use. After harvesting the cells, the cells thawed at room temperature and were lysed using a combination of lysozyme (0.05 g/10 g cell paste) and sonication. For sonication, the cells were disrupted for 3 x 3 minute pulses while in an ice bath. The cellular debris was then removed via centrifugation at 10,000 rpm for 20 minutes and the protein lysate was collected for purification.

The recombinant protein was purified using Ni-NTA affinity chromatography. A gravity fed Ni-NTA affinity chromatography column (resin from https://www.mclab.com/) was equilibrated with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole adjusted to pH 8). The protein lysate was then loaded and allowed to flow through. The previous step was repeated two more times and then wash buffer was then loaded and allowed to flow through the column three times. Next, a gradient of 10 increasing concentrations of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole adjusted to pH 8) were added one at a time to elute the protein from the column. The gradient was as follows: 5 mL wash 0 mL elution, 4.5 mL wash 0.5 mL elution, 4 mL wash 1 mL elution and so on until 0 mL wash 5 mL elution. Upon review of the corresponding SDS gel, there was still significant amounts of protein left in the load flow through, so the loading and elution process was repeated two additional times using the load flow through from the previous run. After all purification processes were finished, the samples with pure protein were collected, concentrated, and buffer exchanged in an amicon device. To concentrate the protein solution, the solution was reduced down to approximately 10 mL. Next, to exchange the

elution buffer, 20 mL of the new buffer (50 mM KCl and 20 mM HEPES at pH 7.5) was added and allowed to concentrate down to 5 mL. This step was repeated one more time to ensure the elution buffer was completely removed. For these processes the amicon was run at a pressure of 50 PSI by injecting nitrogen gas into the device. The concentration and activity of the protein solution was determined via spectroscopy by measuring the absorbance at a wavelength of 280 nm. The protein concentration calculation used this absorbance along with the extinction coefficient determined from the number of tyrosine's and tryptophans in fumarase. The extinction coefficient to determine the protein concentration was 33835 cm^{-1} M^{-1}. The above process was repeated for the second sample that was generated after the plasmid purification. The first round was labeled as round one and the second round was labeled as round 2. The only difference between the two rounds was the buffer the protein was exchanged into. The round one buffer was 50 mM KCl and 20 mM HEPES at pH 7.5 while the round two had a higher salt concentration at 100mM KCl.

Crystallization

Upon obtaining purified protein, many crystal trays with varying conditions were set up and observed to detect the crystalized protein. The initial crystal screens used the conditions described in previous research of fumC. The initial conditions used PEG 4000, $LiSO_4$, DTT, and MOPS. Screens were set up to vary different PEGS (4000 vs 3350), concentration of the PEG, protein concentration, MOPS concentration, and DTT concentrations as well as micro seeding the drops. After resulting in poor quality and small crystals, new conditions were identified. After obtaining three 72 well screens for different conditions, there were better quality crystals in some drops. These crystallization conditions involve a salt, a PEG, and Bis-Tris Propane (BTP). Upon obtaining these new conditions, crystallization screens were made to vary many different aspects. These included: different PEGS (4000 vs 3350 vs 400 vs 8000), percent composition of the PEG (7%-20%), protein concentration (5-50 mg/mL), salt composition (potassium vs sodium salts as well as different halogens or anions: Cl-, Br-, I-, and CN^-), salt

concentrations, pH, temperature, malate concentrations, BTP concentrations, micro seeding drops and drop volume. All these conditions were varied in different screens and repeated with other variations. Additionally, some crystals were placed in new drop with an increased concentration of malate (75mM and 0.5 M) to soak the substrate into the crystal.

The crystal of fumC which was used for data collection was produced via the hanging-drop method out of a solution of xxxxx. The crystals were of the C222 habit with cell dimensions of a = 62.04 Å, b = 121.14, c = 128.24 Å, and α = β = γ = 90°. The crystals dimensions were relatively small at 200 μm x 80 μm x 80 μm.

Table 1: Data Collection and Processing Statistics

Data set	fumC pre-refinement	fumC post-refinement
Space group	I222	
A (Å)	62.04	
B (Å)	121.14	
C (Å)	128.24	
Resolution (lower to higher limit) (Å)	13.9-2.2	
Unique reflections	24497 (3437)	
Redundancy	3.5 (2.8)	
% Completeness	98.1 (95.8)	
I/σ(I)	5.7 (0.7)	
CC(1/2)	0.982 (0.197)	
R_{meas}	0.277 (2.11)	
R_{free}	0.2456	0.2450
R_{work}	0.2466	0.1847
R. m. s bond length (Å)	0.011	0.014
R. m. s bond angles (°)	1.454	1.325
R. m. s. dihedrals (°)		20.533

Average mosaicity	0.67	
Mean B-factor		49.64????
Ramachandran favored (%)		94.5

Data collection and Processing

All relevant data sets were collected at room temperature using a xxxxx dector and xxxxxx radiation…..

One crystal was used for the data set and the program CrysalisPro was used for data collection as well as CCP4 was used for scaling and merging the data. Details of the data sets are contained in Table 1, with the resolution being 2.2 Å.

Refinement

Three rounds of refinement were conducted using the program COOT utilizing the coordinates generated by CCP4. The values for R_{free} and R_{work} stayed under 0.26 and 0.21 respectively for the refinements. In each round of refinement, water molecules were added and after refinement, confirmed with the difference maps.

Results:

The PCR, expression, and crystallization results discussed below will lead up to the structure of fumarase with its substrate in the active site. Starting with the gene for fumC, a 5x C terminal Histidine tag was added to the sequence using specially designed primers with PCR. To determine the best annealing temperature for the amplification of the DNA, the PCR was run with an 18^{o}C temperature gradient, from 50-68 ^{o}C. Figure 7 shows the PCR results run on a 1% agarose gel which indicate that 62^{o}C was the best annealing temperature. The length of the gene is 2408 base pairs without the 6 C-terminus His tags. This is because the intensity of the band in lane 6 was the strongest, signifying the most DNA was present in that PCR sample.

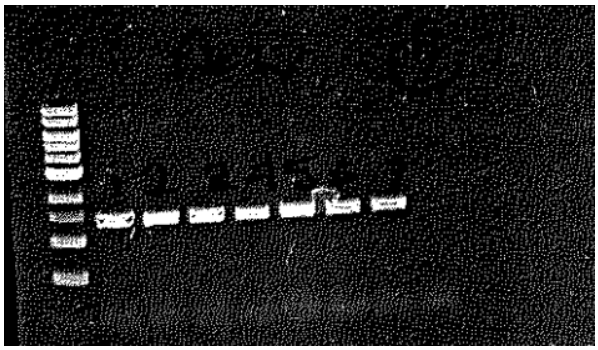


Figure 7: 1% agarose gel showing PCR with a temperature gradient of fumC with primers to add the C terminal His tag. The lanes left to right are: 1 kb ladder, 50-68^{o}C incremented every 3^{o}C. The bands are at approximately 1.3 kb. From this gel, the anneal temperature of 62^{o}C was selected because it was the strongest, most intense band.

Using the PCR sample from lane 6, the DNA was transformed and colony PCR was conducted on seven of the colonies which formed on the consequential plate. The colony PCR amplified the DNA for each colony in order to determine whether the colony contained the plasmid with or without the gene. In figure 8 there are two distinct sizes at around 2.4 Kb and 1.5 Kb corresponding to the plasmid with the fumC gene inserted and just the plasmid respectively. Furthermore, colonies one, three, and five appear to not have the inserted fumC gene. Colonies two, four, six, and seven all appear to have the plasmid with the inserted fumC gene. Additionally, based on the intensities of the bands on the gel, colony seven appears to have the most DNA present in the PCR sample.

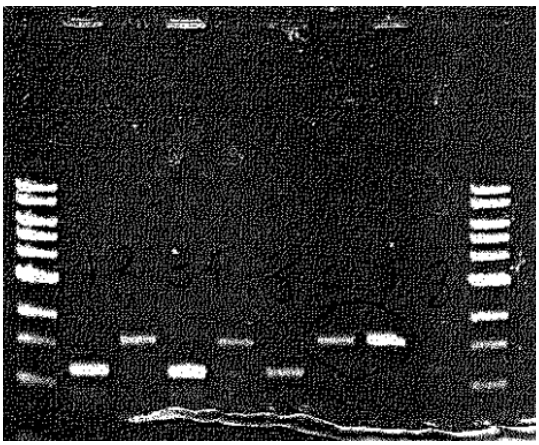


Figure 8: 1% Agarose gel showing the colony PCR results of fumC C-terminal His tag plasmid DNA. From left to right the gel shows: 1 kb ladder, colony 1 up to colony 7, and then another lane of 1kb ladder. There are two clear DNA lengths present in the gel at approximately 2.4 Kb and 1.5 Kb. The DNA at the 2.4 Kb level indicate the fumC gene was taken up by the plasmid.

The two cultures grown up from colony seven were purified with a miniprep and restriction digests were performed to confirm the fumC gene was transformed into the plasmid correctly. The previous gel indicated DNA of roughly the correct length was inserted into the plasmid, the restriction digests confirmed the inserted DNA was in fact the fumC gene. HindIII and Eco-NI were the two restriction enzymes used, each with different cut sites within the plasmid. Table 1 indicates the predicted cut sites and lengths of the fragmented DNA for the plasmid. Figure 9 shows the experimental results of the restriction digests. The restriction digest for the first miniprep of colony seven shows bands at around 2.1 Kb and 1.6 Kb for Eco-NI and bands at around 2.2 Kb and 1.3 Kb for HindIII. These bands approximately agree with the predicted results for the respective enzymes. For the second miniprep of colony seven, there are bands at around 2.2 Kb and 1.6 Kb for Eco-NI and bands at around 2.2 Kb and 1.2 Kb for HindIII. These bands also approximately agree with the predicted results for the respective enzymes.

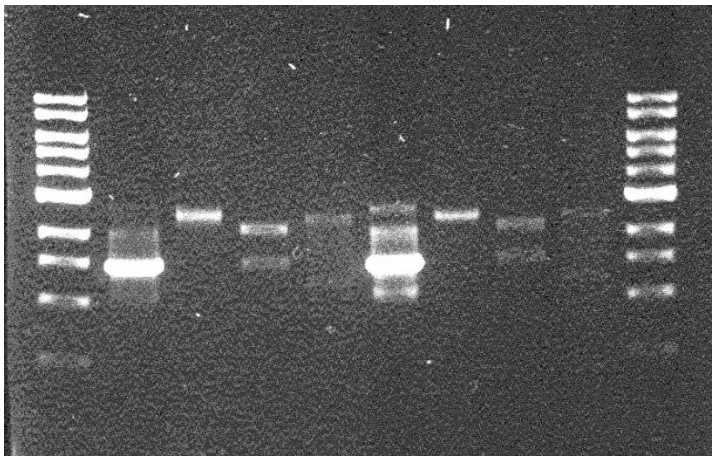


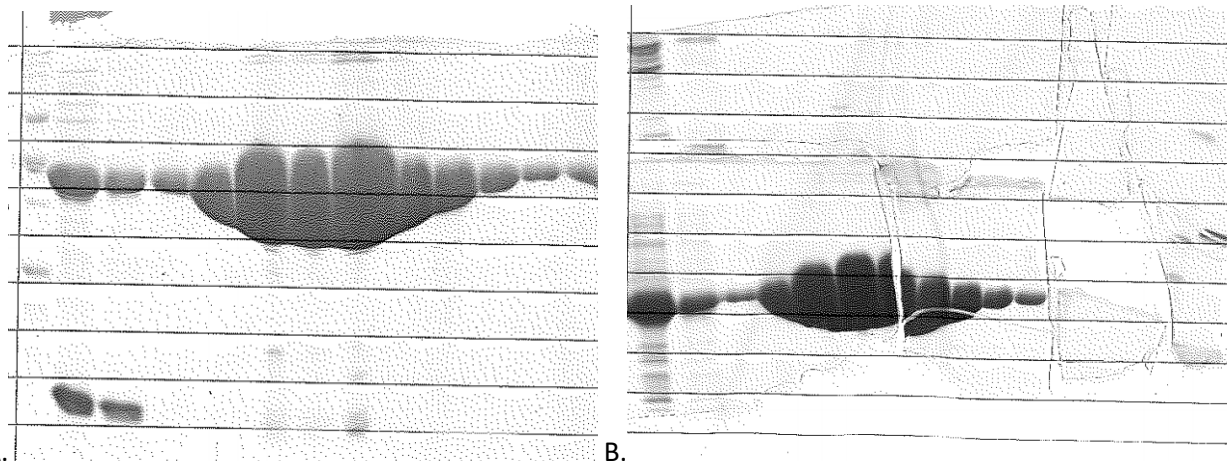
Figure 9: Agarose gel of restriction enzyme digests from two colonies of the plated miniprep purifications. The two restriction enzymes used were EcoNI and HindIII. On the gel the lanes left to right are: 1kb ladder, PCR colony 7 first prep, uncut plasmid colony 7 first prep, EcoNI digest for colony 7 first prep, HindIII digest for colony 7 first prep, PCR for colony 7 second prep, uncut plasmid for colony 7 second prep, EcoNI digest for colony 7 second prep, HindIII digest for colony 7 second prep, and the 1 kb ladder again. Each digestion for both colonies matched the predicted sizes of cut DNA.

Table 2: Restriction enzyme cuts sites for fumC inserted into plasmid.

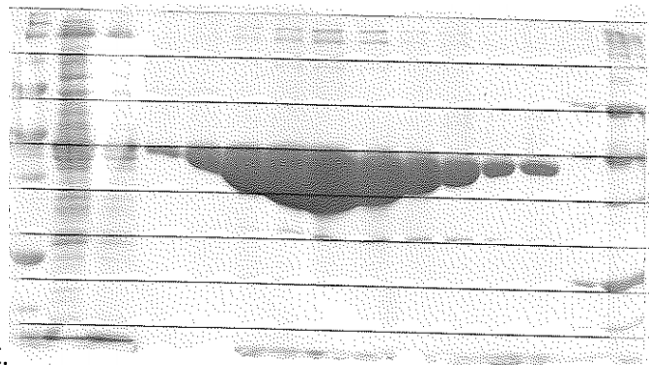
#	Ends	Coordinates	Length (bp)
1	HindIII-HindIII	2893-1675	2416
2	HindIII-HindIII	1676-2892	1217
1	EcoNI-EcoNI	966-3101	2136
2	EcoNI-EcoNI	3102-965	1497

After introducing the DNA plasmid to *E. coli* for overexpression of the protein, the cells were lysed and the protein was purified using a Ni-NTA affinity column. Figure 10 shows the corresponding 12% SDS-phage gels for each of the three rounds of purification. In part A of Figure 10, the gel from the first round of purification is shown. There is a large quantity of fumarase in lanes 2-7 as well as the load

flow through. This means the beads of the column were overloaded and fumarase was flowing through the column. As such, two subsequent purification runs were performed using the load flow through from the run previous. As seen in Figure 10.B the second round and 10.C the third round, the band corresponding to fumarase in the load flow through became increasing smaller after each run. The amount of protein in each elution was still large, but decreased with each round of purification. Additionally, there is not much other protein contamination in the lanes with fumarase, indicating the purification were working efficiently. From the three purifications, the mostly pure elution's were collected, buffer exchanged and concentrated. Specifically, we pooled fractions 2-6 from the first column run, fractions 2-7 from the second column run, and 2-9 form the third column run. The concentration of the protein was determined via spectroscopy. An absorbance of 0.239 was read at the wavelength of 280 nm. The protein concentration calculation used this absorbance along with the number of tyrosine's and tryptophans in fumarase. The final protein concentration for fumarase was found to be 44 mg/mL.



A. B.

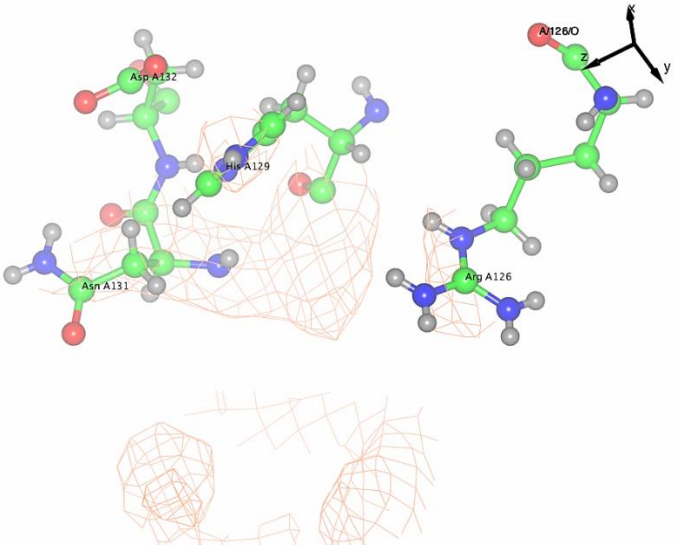


C.

Figure 10: 12% SDS gels of three purifications of fumC with a C terminus Hist tag. Part A shows the first purfication run of fumC C-terminal His tag. From left to right the lanes are: ladder, load flow through, wash flow through and elution fractions 1-10 with increasing concentrations of elution buffer. Note the large amount of protein still in the load flow through lane. Part B shows the second round of purification, running the load flow through the eluted column again. From left to right the lanes are: load flow through, wash flow through and elution fractions 1-10 with increasing concentrations of elution buffer. The ladder lane was lost when the gel ripped. Part C shows the third round of purification of fumC with the load flow throw of the second round of purification. From left to right the lanes are: ladder, load flow through, wash flow through, elution fractions 1-10 with increasing concentrations of elution buffer, and another lane of the ladder.

Upon crystalization, data colletion, and data processing, an electron density map was produced. Using the electron density map, an omit map between the residues and the electon map was created to

view the two active sites . This was done to determine whether there was any electron density in the active site that could correspond to L-malate which was soaked into the crystal when it was grown at a 100 mM concentration. Starting with the B site in Figure 11A, residues Arg126, His129, Asn131, and Asp 132 as well as the electron density from the omit map are shown. Note the presence of electron density surrouding the residues, while the active site in the center of the residues does not have any electron density. Refering back to Figure 4 in the introduction, the first carbolylate of malate is coordinted to Asn131 and Asp132 while the second carboxylate is coordinated to Arg126 and His129. We see no such density in any of these locations in Figure 11A. Additionally, Figure 11B shows the published and deposited PDB structure of apo fumarase superimposed on the structure determined from this study. Notice that His129 is in the same conformation as well as all the other residues in the image.



A.

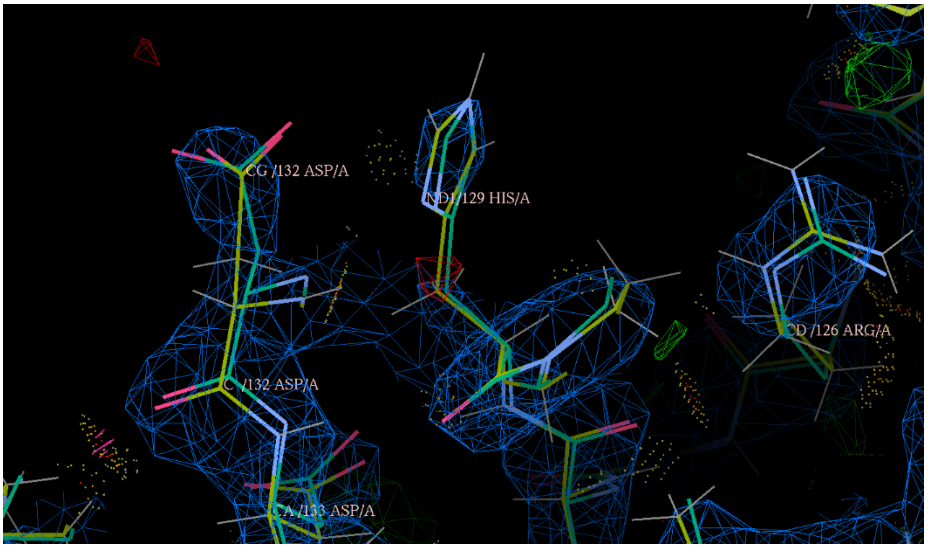
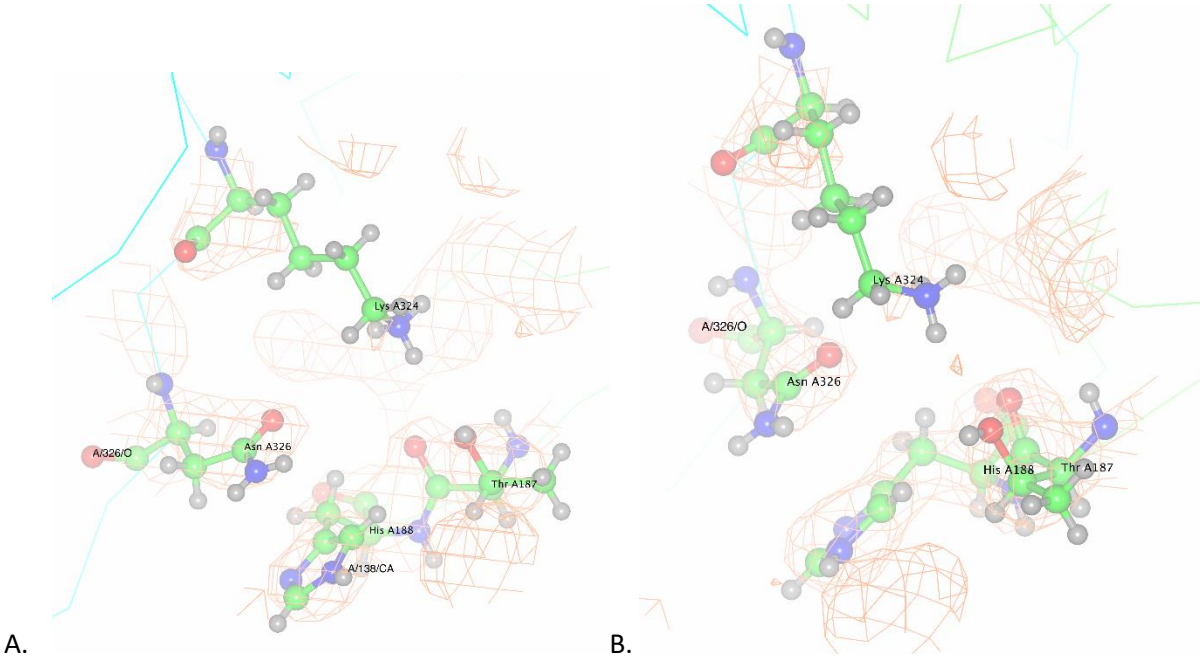
B. 

Figure 11: The omit map of the B site and superimposed image of the free fumarase deposited in the PDB. Part A shows the omit map of the B site. It contains the known active site residues, Arg126, His129, Asn131, and Asp132 as well as the electron density of the determined structure. Note the lack of density between the four key residues in the B site. In previous studies, the malate found in the B site would be centered between the four key residues. Part B show the structure determined in this experiment superimposed onto the PDB file of free fumarase. The two overlaying structures are nearly identical.

Moving to the A site of fumarase, Figure 12 has two different views of the omit map with the corresponding active site residues. The active site cavity is formed by the residues Thr187, His 188, Kys324, and Asn326. In the omit map below, we see no clear electron density in the active site cavity.



A. B.

Figure 12: Two different views of the omit map of the A site within fumC. The A site catalytic cavity is formed by the residues Thr187, His 188, Kys324, and Asn326. When looking at parts A and B of the figure, there is no electron density in the center of the cavity that could correspond to a small molecule in the A site.

Discussion:

Obtaining a concentrated stock of the fumC protein was the initial step of this research. Starting from the gene, a His-tag was added to the C-terminus of the DNA sequence, amplified, and transformed into 10-G *E.coli* cells. The transformations were confirmed successful after colony PCR was performed because the length of DNA (2.4 kb) matched the predicted length of the gene plus the plasmid. Additionally, restriction digests confirmed the transformation because each digest matched the expected DNA lengths for the corresponding enzyme. The DNA was then transformed into BL21 cell for overexpression in *E. coli*. After affinity chromatography, the pure protein was collected and

concentrated down to 15 mL. The final concentration was determined to be 44 mg/mL via UV absorption spectroscopy.

Producing crystals that diffracted to below 3 angstroms was a difficult task for fumC. There were two different issues that when compounded, proved to be problematic in the production of the crystals; size and malate concentration. Starting with the first obstacle, the crystallization conditions that were tested produced an average crystal that was less than 50 µm x 50 µm x 100 µm in size. At this size, adequate diffraction was unattainable for the quality of the crystals. Thus, attempts at optimizing the crystallization conditions to generate larger crystals was performed. Two sets of conditions were optimized using a variety of substitutions and techniques ranging from switching salts to changing the pH of the buffers and micro-seeding the droplets to droplet volume. There were two completely different crystallization conditions tested. The first conditions were found via three 72 well plates at the beginning of the process which established conditions that generated crystals using bis-tris propane, PEG 3350, and a salt. These conditions produced two morphologies of crystals, well-formed rectangular prisms and long, but narrow needles. When optimizing this set of conditions, trends became apparent regarding the significant variables. The important trends were: lower concentrations of PEG 3350 (9-14%), lower salt concentrations using KCl (XXXX mM), micro-seeding helped with crystal size, higher protein concentration (30 mg/mL or more), room temp growth, 300 µL well solution with 4.5 µL drops, and higher pH (8.4 for the BTP). The second set of conditions came from the published papers for the protein, and in particular the conditions from the apo paper were used; PEG 4000, MOPS, and $LiSO_4$. The other published conditions for fumarase C contained competitive inhibitors that would out compete malate from the active sites. These conditions produced crystals with a rectangular prism morphology. The important trends for these crystallization conditions were similar to the other set. Higher protein concentration, lower PEG, and lower $LiSO_4$ concentrations produced larger crystals. After optimizing the both conditions, the crystals from the conditions generated with a salt were approximately twice as

large, with dimensions attaining approximately 200 µm x 100 µm x 100 µm. With these crystals, pre-tests as well as full datasets were collected that showed diffraction up to approximately 2.2 angstroms. At this level of diffraction, it was possible to determine the electron density of the two active site in order to ascertain if a small molecule was present in either active site.

The next challenge in generating crystals for this experiment was the malate concentration. The experiment was attempting to co-crystalize malate in one or both of the active sites in fumarase C. In order to accomplish this, different approaches were attempted in order to soak malate into the crystals. To begin, the K_m of malate is 0.025 mM and that of fumarate is 0.067 mM. The values for K_m determined the minimum concentration needed for malate to have high affinity for the active site. Thus, the first method integrated a concentration of 100 mM L-malate directly into the crystallization droplet when the crystal tray was created. At this concentration of substrate, malate should be bound to at least one of the active sites. For the conditions with a salt, this technique did not inhibit the growth of the crystals. Additionally, the optimized crystals under this method produced quality diffraction up to 2.2 angstroms. On the other hand, this approach was not amenable to the conditions that used MOPS as no or very small crystals would grow. The next approach was to soak malate into the crystal by transferring the already grown crystal to a new droplet containing a set concentration of malate. These concentrations varied from 5 mM to 100mM. This technique did not attain any results for either of the crystallization conditions because the crystals would disintegrate after 12 hours. Additionally, in tests before the 12 hours, the crystals would not diffract to anything better than 6 angstroms in pre-test experiments. Thus, this approached prevented any reasonable data collection and viewing of the active sites. The third approach was to combine the last two techniques to grow the crystals with a malate concentration and then soak them in another drop solution that contained malate. This approach did not work for either of the crystallization conditions as the crystals disintegrated or did not diffract below 3 angstroms. Altogether, the best approach and the one used to obtain the data found in this paper, was to grow

crystals with a malate concentration in the initial drop rather than attempting to soak in malate after crystallization.

The two carboxylate binding sites have already been determined by previous studies and the key residues are known. The catalytic A site is formed by residues Thr187, His 188, Kys324, and Asn326. When looking at Figure 12, which shows the omit map between the four key residues and the corresponding electron density for the area, there is density around each amino acid, yet no electron density in the active site cavity. As a reference for where electron density should be located, the case of citrate bound in the A site can be considered. The residues K324, T187, and N326 interact with the first carboxylate, and N326 and T100 interact with the second. Thus, if a small molecule was present, there should be density at and between those residues which does not appear to be the case in the omit map. This would indicate that there is no small molecule present in the A site because there appears to be no electron density that could correspond to a small molecule.

Moving to the other active site of the protein, the B site is constructed from the residues Arg126, His129, Asn131, and Asp132. Looking at the omit map of the B site in Figure 11A, the electron density is mostly around the active site residues. There appears to be no electron density in the active site cavity which would suggest that there is no small molecule present in the B site. Additionally, recall from the introduction that residue His129 had two different conformations corresponding to the bound and free structure. With the presence of a small molecule in the B site (malate) His129 rotates out of the B site and contacts Asp133 and Arg 126. When in the free state, His129 rotates into the B site preventing a small molecule from binding. In Figure 11B, the structure from this experiment is superimposed upon the published structure of the free or apo enzyme. The two structures are nearly identical and specifically, His129 is in the same conformation between the two structures. This strongly indicates that there is no small molecule present in the B site because of the lack of electron density and the conformation of His129.

The structure determined in this study appears to be free fumarase C from *E. coli* with newly identified crystallization conditions. An additional measurement that helps confirm apo fumarase is the unit-cell dimensions for this structure. This structure was I222 with a = 62.04 Å, b = 121.14, c = 128.24 Å which can be compared to the known apo structure with dimensions of a = 121.6 Å, b = 128.0, c = 62.1 Å. Now established this study has generated the structure of apo fumarase C, a general comparison with the published structure can be performed. The approximately 2.2 angstrom resolution from the structure is very comparable to the 2.19 angstrom structure that is deposited in the PDB (1YFE). Some differences between the structures is the R_{free} and R_{work} values are slightly higher for this experiment at 2.45 and 1.85 versus 2.10 and 2.03 respectively for both experiments. The signal to noise ratio, $I/\sigma(I)$, from the published experiment is quite high at 13 and best at 1.4 compared to this study at 5.7 with a best of 0.7. The other metrics for comparison are very close between the two studies. Looking at the comparisons of the two structures, it appears that they are very similar in resolution and had similar issues with attaining strong diffraction.

Returning to the kinetics, crystallization conditions and structure of fumarase, possible reasoning for why there wasn't malate present in either of the active sights can be hypothesized. Looking at the published crystallization conditions for the citrate bound (300 mM citrate, pH 6.0, and 14% PEG 4000) and the apo structure (50 m*M* MOPS pH 7.5, 100 m*M* $LiSO_4$ and 12% PEG 4000), one of the main differences is the pH of the conditions. With the isoelectric point of fumarase C at 6.51, one of the conditions is below it, making the protein have a net positive charge, and the other above it, making the protein have a net negative charge. When the pH was altered in the crystallization screens of this study, a switch from pH 8.4 to 6, the crystal sizes decreased greatly. This led to the conditions staying at a pH of 8.4, while a pH below 6.51 could have increased the enzymes affinity for malate in either of the active sites. The next possible reason for malate not being in the A site would be that it is being converted to fumarate. It was shown that at high concentrations of malate, the catalysis happens at an

excited rate. It is possible that the equilibrium favored fumarate. However, to complicate this possible reason, the known K_m of malate and that of fumarate are 0.025 mM and 0.067 mM respectively. At the 100 mM concentration of malate in the drop solution, even if it was converted to fumarate, one of the two small molecules would have a high affinity for the enzyme. This would suggest that there is a possible steric effect blocking the active sites. Altogether, the reason as to why there was no small molecule, either malate or fumarate, in either of the active sites is unknown.

Conclusion:

In this study, an attempt to co-crystalize the protein fumarase C from *E. coli* with its substrate, L-malate, was performed. Upon data collection and refinement, the crystals that were produced correspond to the previously deposited structure of apo fumarase C. Thus, a new set of crystallization conditions were found for comparable resolution to the deposited PDB file (1YFE). An important difference between the two experiments is that the newly found crystallization condition incorporates a 100 mM concentration of L-malate into the drop solution. Further research could be performed to co-crystallize the enzyme with its substrate in order to further elucidate the catalytic residues in the A site and determine the mechanism of the proposed activator B site. Additionally, the role of the conserved water molecule in the A site could be determined and its importance to the mechanism established.

Citations:

1. Weaver, Todd. "Structure Of Free Fumarase C From Escherichia Coil." *Acta Crystallographica: Section D (Wiley-Blackwell)* 61.10 (2005): 1395-1401. *Academic Search Premier*. Web. 5 Jan. 2017.
2. “fumC - Fumarate Hydratase Class II - Escherichia Coli (Strain K12) - fumC Gene & Protein.” Accessed January 5, 2017. http://www.uniprot.org/uniprot/P05042.

3. Weaver, T., M. Lees, and L. Banaszak. “Mutations of Fumarase That Distinguish between the Active Site and a Nearby Dicarboxylic Acid Binding Site.” *Protein Science : A Publication of the Protein Society* 6, no. 4 (April 1997): 834–42.
4. **"Purification and characterization of two types of fumarase from Escherichia coli."** Ueda Y., Yumoto N., Tokushige M., Fukui K., Ohya-Nishiguchi H. J. Biochem. 109:728-733(1991) [PubMed] [Europe PMC] [Abstract]
5. **"Purification and characterization of two types of fumarase from Escherichia coli."** Ueda Y., Yumoto N., Tokushige M., Fukui K., Ohya-Nishiguchi H. J. Biochem. 109:728-733(1991) [PubMed] [Europe PMC] [Abstract]
6. Hill, R. L., & Teipel, J. W. (1971) Enzymes 5, 539 - 571.
7. Guest, J. R., Miles, J. S., Roberts, R. E., & Woods, S. A. (1985) J. Gen. Microbiol. 131, 2971 - 2984.
8. Estévez, Marcel, Jeremy Skarda, Josh Spencer, Leonard Banaszak, and Todd M. Weaver. “X-Ray Crystallographic and Kinetic Correlation of a Clinically Observed Human Fumarase Mutation.” *Protein Science* 11, no. 6 (June 1, 2002): 1552–57. doi:10.1110/ps.0201502.
9. Accessed February 11, 2017. https://upload.wikimedia.org/wikipedia/commons/0/0b/Citric_acid_cycle_with_aconitate_2.svg.
10. Weaver, Todd, and Leonard Banaszak. “Crystallographic Studies of the Catalytic and a Second Site in Fumarase C from Escherichia Coli,.” *Biochemistry* 35, no. 44 (January 1, 1996): 13955–65. doi:10.1021/bi9614702.
11. Puthan Veetil, Vinod, Guntur Fibriansah, Hans Raj, Andy-Mark W. H. Thunnissen, and Gerrit J. Poelarends. “Aspartase/Fumarase Superfamily: A Common Catalytic Strategy Involving General Base-Catalyzed Formation of a Highly Stabilized Aci-Carboxylate Intermediate.” *Biochemistry* 51, no. 21 (May 29, 2012): 4237–43. doi:10.1021/bi300430j.

12. “Fumarase.” *Wikipedia*, July 23, 2016.

https://en.wikipedia.org/w/index.php?title=Fumarase&oldid=731146896.