HARDESTY, GRANT ALAN, M.S. The Characterization of MmgE from *Bacillus subtilis*. (2012)
Directed by Dr. Jason J. Reddick. 56 pp.

Bacillus subtilis is a rod shaped gram-positive bacterium that has been extensively studied for decades. The scientific interest lies, in large part, to its use as a model organism for the study of sporulation, which is a simple form of cell differentiation. It is also a very important organism in industry and everyday life as it is used to produce antibiotics, and other important compounds. *B. subtilis* is also used as a probiotic and as an additive to organic fertilizers and plant treatments because of its antifungal properties.

The regulation of the initiation and continuation of sporulation has been well studied, but the metabolic changes the cell must go through are still in question and the subject of current study. One of the metabolic changes that happens in the cell is the activation of the *mmg* operon by σ^E in the mother cell during sporulation. This operon is composed of six genes that encode a putative fatty acid degradation pathway and the methylcitric acid cycle. This cycle takes propionyl-CoA (a fatty acid degradation metabolite) and condenses it with oxaloacetate into 2-methylcitrate, and through the rest of the cycle produces pyruvate and succinate that can enter the citric acid cycle.

One gene in the middle of the putative methylcitric acid cycle is *mmgE*. This gene has not previously been studied. Its sequence, both genetic and proteomic, are a very close match to 2-methylcitrate dehydratases found in other organisms like *Escherichia coli* and *Salmonella typhimurium.*

The *mmgE* gene was cloned into an overexpression strain of *E. coli* and the protein was isolated for study through Ni-NTA chromatography. Through UV/Vis

spectrometry, high performance liquid chromatography (HPLC), and mass spectrometry the activity of this enzyme, on 2-methylcitrate, was studied. It was found that this enzyme does perform 2-methylcitrate dehydratase activity.

THE CHARACTERIZATION OF MMGE FROM *BACILLUS SUBTILIS*

by

Grant Alan Hardesty

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2012

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair_____

Committee Members_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

BACKGROUND

I. A Sporulation

The overall study of *Bacillus subtilis* mainly focuses on the mechanism of sporulation. (1) This is just one physiological process of the bacterium but it is complex and important to understand. (1) Just how complex is fairly well understood as more than a hundred genes (at least partially), and many other factors that control transcription and translation, are involved in this entire process. (1) Sporulation is an energetically intensive process that creates a thermally, and chemically resistant compartment containing genetic material and necessary proteins. (2) This compartment is called a spore and can survive everything short of incineration. (2) The initiator of sporulation is a transcription factor encoded by Spo0A. (2)

Spo0A is the ultimate initiator of sporulation. It exists in an inactivated state until phosphorylated. There exists a system called the phosphorelay to transfer a phosphate onto Spo0A, activating it. This involves three histidine kinases which eventually pass the phosphate on to Spo0F which then passes it onto Spo0B, and finally to Spo0A. This relay makes for easy regulation at multiple points as each step in the path can be inhibited (3). For example a regulatory protein, RapA, inhibits Spo0F through dephosphorylation. Since this is an energetically sensitive time for a *B. subtilis* cell the phosphorelay allows

for strong regulation without the energy expense of a kinase cascade. A kinase phosphorylation cascade would require the use of an ATP at each phosphorylation so the amount of energy that would go into initiating phosphorylation would be much higher. The phosphorelay only requires a single ATP for each run through the whole pathway.

Once activated as Spo0A~P this protein acts as a transcription factor that is responsible for the appearance of σ^F (1). The way that the phosphorylated form of Spo0A controls the entry into sporulation is by the ratio of Spo0A~P:Spo0A. Higher and higher concentrations of Spo0A~P cause activation of more and more processes as the series of sporulation events is linked to the concentration of the active form of Spo0A (4) (3). The concentration of Spo0A increases once activated due to a positive feedback loop that is controlled by Spo0A itself as it acts as its own transcription factor.

Under favorable growth conditions the bacterial cell divides in two after doubling in size. This is the normal life cycle of the cell and known as stage zero of sporulation. Under any of the aforementioned unfavorable conditions the cells leaves the static stage zero and enters stage one. Stage one is characterized by the chromosomes lining up along the long axis of the cell as the septum begins to form at one pole directed by a ring of FtsZ, an analog of tubulin, that is a major component of eukaryotic cell division (4). FtsZ along with SpoIIE direct the division of the cell to one of the poles. Without SpoIIE activation the cell would divide along the cellular midpoint between the poles, just like a stage zero cell (4).

Stage II of sporulation starts once the chromosome has lined up from pole to pole and the septum forms (4). This internal peptidoglycan wall segregates the early prespore

from the mother cell. With the way the chromosome lines up in the cell, approximately thirty percent of it is in the prespore with the rest having to be transported across by SpoIIIE (4). This DNA translocase forms a pore in the septum and allows the rest of the necessary DNA to be transported into the prespore (4). The next thing to happen after division is to have simultaneous gene transcription follow different programs in the mother cell and prespore, since these two components are to become vastly different. There are two sigma factors that are responsible for the difference in transcription between the mother cell and prespore: σ^E in the mother cell, and σ^F in the prespore. The prespore σ^F is regulated by two other factors: SpoIIAB and SpoIIAA. SpoIIAB is an anti-sigma factor as it binds to and inactivates σ^F. The SpoIIAB- σ^F complex exists during the normal life cycle of *B. subtilis* but since they are complexed it is inactive.

Once the cell has divided into the prespore and mother cell in stage II SpoIIAA is transcribed in the prespore (4). It interacts with the SpoIIAB- σ^F complex and causes σ^F to be displaced and active. After σ^F has been activated in the prespore spoIIR is transcribed and moves to the mother cell to activate σ^E (1). This sigma factor controls transcription of genes in the mother cell (1). Before activation σ^E exists in a pre-σ^E form (1) (5). This inactive form of σ^E is signaled to begin transcription, pre-division, by Spo0A~P, so its production is linked to sporulation initiation and is not present at any other time (1) (6). This sigma factor is a member of the *spoIIG* operon and Spo0A~P also induces transcription of SpoIIGA which is the enzyme that activates σ^E by cleaving a twenty-seven residue amino acid sequence from the N-terminus of the pre-σ^E (1) (6)

After transcription differentiation has occurred the cell moves into the stage III of sporulation, which is engulfment. The mother cell produces proteins that slowly degrade the peptidoglycan septum. This removes the separation between the prespore, which remains in its subterminal location, and the mother cell. This is described as a "phagocytic" process (1). This is due to the fact that the mother cell engulfs the prespore. At this point the mother cell also demonstrates some control over the gene transcription in the prespore. The SpoIIIG gene begins to be transcribed an hour after engulfment and σ^E is necessary for this to occur (1). The signaling mechanism is not known beyond the link to σ^E. This gene is necessary for the completion of engulfment since it encodes σ^G ((4). This sigma factor controls entry into the rest of sporulation (4). At this point the prespore is far from complete as it is lacking the tough outer coatings characteristic of a mature spore.

Two tough coatings are formed and cover the prespore in the next two stages of sporulation. The interior of the prespore changes drastically as well. This internal change is formed by a multitude of small proteins, produced in the prespore, that cover the DNA. During this process dipicolinic acid, produced in the mother cell, and large amounts of metal ions are absorbed by the prespore (4). This all goes towards hardening the prespore to protect the DNA against any harsh (4). While the prespore is changing internally, the cortex, a peptidoglycan cell wall, is formed on the outside. This is quickly followed by the spore coat which is a protein-based covering that forms on the exterior of the cortex. Once these two coverings are formed and the core completed the spore is ready and the mother cell lyses.

I. B The *mmg* Operon

There are many genes in the *B. subtilis* genome that encode for the use of various sources of carbon as nutrients for cellular needs (7). Of these there seems to be a distinct need for branched short-chained acids for producing the fatty acids needed for various cellular demands such as membrane production (7). *B. subtilis* can make and use a wide variety of these molecules and this is where the *mmg* operon becomes important in sporulation. The metabolic processes for the vegetative state and early sporulation are known and well-studied, such as glycolysis and the Krebs cycle (5). Once sporulation begins it causes distinct metabolic changes in the cell as it prepares the new spore (5). As stated previously gene transcription becomes differentiated between the prespore and the mother cell, and this is true for metabolism. In the later stages of sporulation the prespore starts producing glucose dehydrogenase as the mother cell begins glycogen production (5). Intermediate stage metabolism isn't as well understood and is also when the *mmg* operon becomes active.

As stated previously σ^E controls transcription of several genes in the mother cell during sporulation (5). One such set of genes are the "mother cell metabolic genes" that are known as the *mmg* operon. This operon consists of six genes: *mmgA, mmgB, mmgC, mmgD, mmgE*, and *yqiQ* (8). The first five genes in this operon were described by Bryan et al in 1996 after systematically screening for σ^E dependent open reading frames. *MmgA-E* were found through this screening and later on Kunst et al found the sixth member of the operon through genome sequencing and gene annotation (7) (5). This last member was named yqiQ (7).

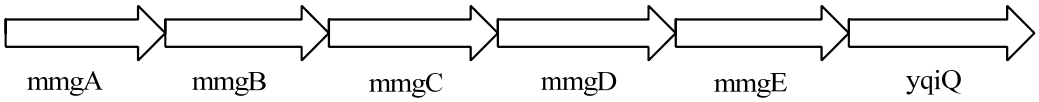


Figure 1. The *mmg* operon. (8)

The first three enzymes of the operon, mmgA-C, are part of fatty acid degradation. *MmgA* was found to encode an acetoacetyl-CoA thiolase by Reddick and Williams (8). Work in the Reddick group also showed that *mmgB* encodes a 3-hydroxybutryl-CoA dehydrogenase (5) (9) and *mmgC* encodes an acyl-CoA dehydrogenase (5) (10) (11). MmgD is the third citrate synthase to be found in *B. subtilis*, and the Reddick group has shown that it has 2-methylcitrate synthase activity (12) (5) (11). The next gene in the operon, *mmgE*, is similar by sequence to 2-methylcitrate dehydratase. YqiQ is likely a 2-methylcitrate lyase (7) (13) (11).

I. C Methylcitric Acid Cycle

Escherichia coli has previously been shown to have a methylcitric acid cycle (14). This cycle becomes fully active under glucose deficiency, and in the presence of propionate (14). Short chain fatty acids can be used as a carbon source by having propionyl-CoA be converted to citric acid cycle substrates (pyruvate and succinate). The ability for an organism to do this not only allows for continued survival on another carbon source, but propionyl-CoA has been shown to be an inhibitor if it is allowed to build up within the cell (15) Having this capability also allows for gluconeogenesis if the cell is starved for glucose. (15)

The genes that make it possible for *E. coli* to metabolize propionate are the common methylcitrate enzymes encoded by the *prp* operon which has the genes prpBCDE (16) (17). These genes encode the ubiquitously required genes for the methylcitrate cycle: methylcitrate lyase, methylcitrate synthase, methylcitrate dehydratase and propionyl-CoA synthase (16) (17). The *prp* genes do not work in this order though. Firstly, propionate needs to be activated by PrpE which attaches coenzyme A to propionate forming the active compound, propionyl-CoA, which is now ready to enter the main body of the methylcitric acid cycle (16) (18). The next step is for propionyl-CoA to condense with oxaloacetate (a citric acid cycle intermediate) to form 2-methylcitrate. This reaction is catalyzed by PrpC, also known as methylcitrate synthase (the reaction is shown in Figure 3) (16) (19) (20). The 2-methylcitrate that is formed by PrpC is then acted upon by methylcitrate dehydratase. This enzyme removes water to create, in *E. coli*, 2-methyl-cis-aconitate (16) (21). Now at this point another enzyme outside of the operon is used to create the substrate for methylisocitrate lyase (16) (21). Aconitase, which is a citric acid cycle enzyme, is necessary to add water, into the aconitate, to form 2-methylisocitrate (16) (18). Methylcitrate lyase is the enzyme at the very end of the methylcitric acid cycle. The products (pyruvate and succinate) are intermediates of the citric acid cycle and can be used to reform oxaloacetate for another round of propionate catabolism (16) (18).

It was shown by Horswill et al, that *Salmonella typhimurium* needs the *prp* operon to grow on propionate (19). The operons in both *E. coli* and *S. typhimurium* encode the same genes so they were given the same designation for propionate metabolism (16) (19)

(17). The *S. typhimurium prp* operon does have one more open reading frame: *prpR*. This gene encodes a sigma-54 family transcription activator, which is part of the promoter that controls gene activation (22). This organism has the same basic methylcitric acid cycle as *E. coli* (19) (16). It has the *prpBCDE* genes that encode: methylisocitrate lyase, methylcitrate synthase, methylcitrase dehydratase, and propionyl-CoA synthase respectively (19). These genes perform the same tasks as stated above and metabolize propionate into pyruvate and succinate. Both *S. typhimurium* and *E. coli* are gram-negative examples of bacteria species that have 2-methylcitric cycles so it is something that is common through bacterial species as a whole, especially since *B.subtilis* is a gram-positive species that has a putative methycitrate cycle.

The entire *mmg* operon in *B. subtilis* is likely involved with branched fatty acid degradation, as *mmgABC* are presumably responsible for breakdown of long-chain, and branching fatty acids as the indication is that about 85% of *B. subtilis* fatty acids are branched (23) (14). All three of these genes are involved in odd-iso, even iso, and anteiso branched fatty acid degradation (14). Odd-iso fatty acid metabolism leads straight into the citric acid cycle since it produces acetyl-CoA directly, but anteiso and even-iso produce propionly-CoA which, with oxaloacetate, is the primary substrate for the methylcitrate cycle (14). These fatty acids are good sources of energy and very abundant in all of *B. subtilis*'s natural habitats (23). Not only can fatty acids be pulled from the environment for metabolism but they are present in a multitude of forms in the membranes of the *B. subtilis* cell. This means that propionate can be funneled into the methylcitrate cycle from catabolizing its own lipids. With how the *mmg* operon is

activated in the mother cell during sporulation it makes sense that it would break down cellular lipids to pay the large energy expense likely required for sporulation (5).

The breakdown of odd chain fatty acids will lead to propionate production (17). This means that *mmgABC* likely supply of propionate, which goes straight into the main part of the methylcitric acid cycle that is encoded by *mmgDE* and *yqiQ* (14). *MmgD* is a gene that transcribes into a protein that is homologous to the *prpC* genes mentioned earlier as it is a citrate synthase, and actually the third one in *B. subtilis* (6) (16) (19). This was a key factor in hypothesizing that *B. subtilis* has a methyl citric acid cycle since protein structure dictates function, having similar structure gives a great possibility of similar capabilities. (14)

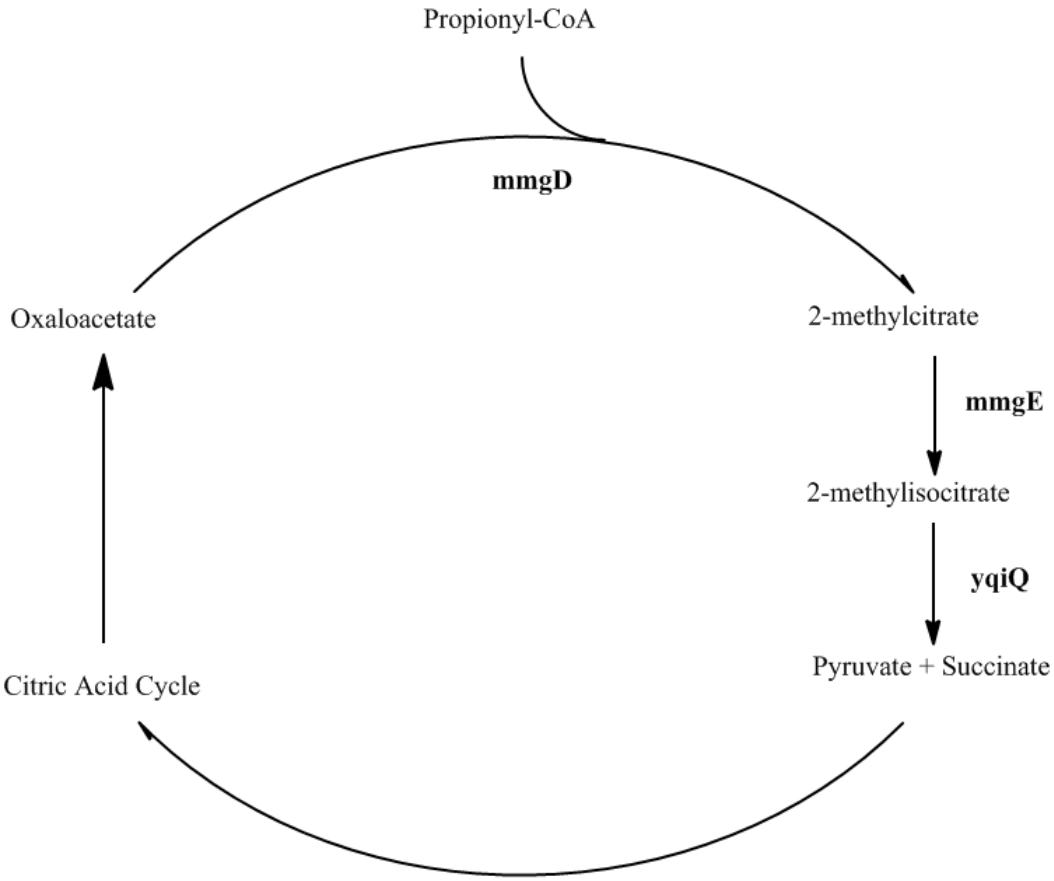


Figure 2. The methylcitric acid cycle. All relevant enzymes, from *B. subtilis*, are in bold**.** (14)

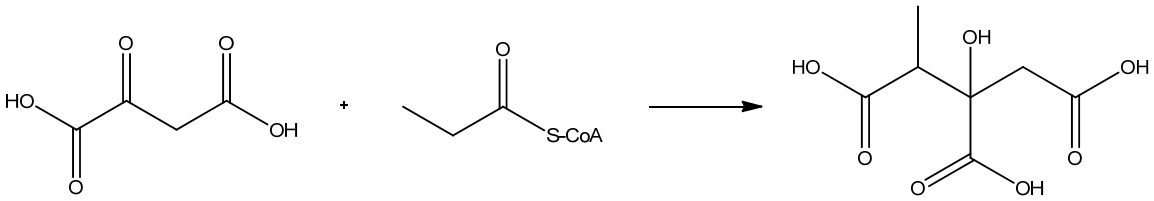
I. D Introduction to MmgE

MmgE is a previously uncharacterized enzyme encoded by one gene of the *mmg* operon of *Bacillus subtilis*. The gene is 1416 base pairs long and transcribes to a 471 amino acid protein that is 52,737.6 Da in mass (11)

Salmonella typhmurium is one of the organisms that has been identified to have a methylcitric acid cycle and as such has a 2-methylcitrate dehydratase (19). In this organism the operon that encodes this set of genes is the *prp*BCDE operon. The cycle was characterized by mutations in the respective genes and analysis was done by NMR to see what intermediates built up in solution (19). When *prpD* was mutated to be non-functional, 2-methylisocitrate did not appear and concentrations of 2-methylcitrate built up showing that this enzyme was likely a 2-methylcitrate dehydratase that catalyzes the tautomerization of 2-methylcitrate into 2-methlyisocitrate (19).

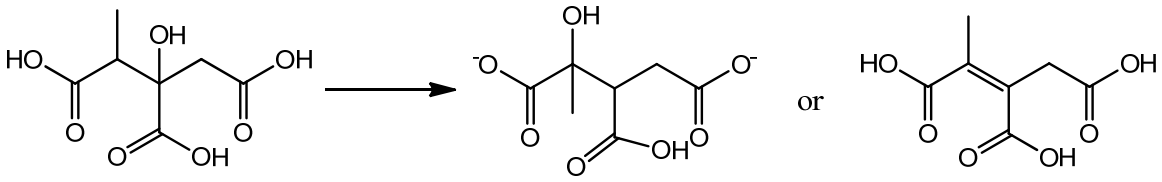
Another organism that has been shown to have a methylcitric acid cycle is *E. coli* (16) (18). The prpD protein in *E. coli* was determined to dehydrate 2-methylcitrate into 2-methyl-cis-aconitate requiring the activity of the citric acid cycle aconitase to complete the conversion to 2-methylisocitrate, which the isocitrate lyase can cleave into pyruvate and succinate (16). Brock et al also discovered that the PrpD protein was similar to MmgE through sequence analysis.

Work done with *Salmonella enterica* showed that the prpD enzyme in that organism is in fact 2-methylcitrate dehydratase (21). The enzyme converted 2-methylcitrate into 2-methyl-cis-aconitate and could not go any further. This means that in some organisms there is the need for another enzyme to convert the methyl aconitate to 2-methylisocitrate. (21)



Oxaloacetate Propionyl-CoA 2-Methylcitrate

Figure 3. The reaction catalyzed by MmgD to form 2-methylcitrate, the substrate for MmgE (24)**.**



2-Methylcitrate 2-methylisocitrate 2-methyl aconitate

Figure 4. The expected reaction catalyzed by MmgE.

I. E ClustalW Alignment

```
CLUSTAL 2.1 multiple sequence alignment


prpD            MSAHISNVRPDFDREIVDIVDYVMNYEITSKVAYDTAHYCLLDTLGCGLEALEYPACKKL 60
e.              MSAQINNIRPEFDREIVDIVDYVMNYEISSKVAYDTAHYCLLDTLGCGLEALEYPACKKL 60
B.              --------MPKTDRVIEEITDYVLEKEITSAEAYTTAGHVLLDTLGCGILALRYPECTKL 52
                         *. ** * :*.***:: **:*  ** ** : ********: **.** *.**

prpD            LGPIVPGTVVPNGARVPGTQFQLDPVQAAFNISAMIRWLDFNDTWLAAEWGHPSDNLGGI 120
e.              LGPIVPGTVVPNGVRVPGTQFQLDPVQAAFNIGAMIRWLDFNDTWLAAEWGHPSDNLGGI 120
B.              LGPIVPGTTVPNGSKVPGTSYVLDPVRAAFNIGCMIRWLDYNDTWLAAEWGHPSDNLGGI 112
                ********.**** :****.: ****:*****..******:******************

prpD            LATADWLSRNAVAAGKAPLTMKQVLSGMIKAHEIQGCIALENAFNRVGLDHVLLVKVAST 180
e.              LATADWLSRNAVASGKAPLTMKQVLTAMIKAHEIQGCIALENSFNRVGLDHVLLVKVAST 180
B.              LAAADYVSRVRLSEGKEPLTVRDVLEMMIKAHEIQGVLALENSLNRVGLDHVLFVKVATT 172
                **:**::**  :: ** ***:::**  ********* :****::*********:****:*

prpD            AVVAEMLGLTRDEILNAVSLAWVDGQSLRTYRHAPNTGTRKSWAAGDATSRAVRLALMAK 240
e.              AVVAEMLGLTREEILNAVSLAWVDGQSLRTYRHAPNTGTRKSWAAGDATSRAVRLALMAK 240
B.              AVAAKLLGGGREEIKNALSNAWIDNAALRTYRHSPNTGSRKSWPAGDATSRGVHLALMSL 232
                **.*::**  *:** **:* **:*. :******:****:****.*******.*:****:

prpD            TGEMGYPSALTAKTWGFYDVSFKGETFRFQRPYGSYVMENVLFKISFPAEFHSQTAVEAA 300
e.              TGEMGYPSALTAPVWGFYDVSFKGESFRFQRPYGSYVMENVLFKISFPAEFHSQTAVEAA 300
B.              KGEMGYPTALSAPGWGFQDVLFNKKEIKLARPLDAYVMENVLFKVSYPAEFHAQTAAESA 292
                .******:**:*  *** ** *: : ::: ** .:*********:*:*****:***.*:*

prpD            MTLYEQMQAAGKTAADIEKVTIRTHEACLRIIDKKGPLNNPADRDHCIQYMVAVPLLFGR 360
e.              MTLYEQMQAAGKTAADIEKVTIRTHEACIRIIDKKGPLNNPADRDHCIQYMVAIPLLFGR 360
B.              VILHPQVKNR---IDEIDRVVIRTHESAIRIIDKKGPLHNPADRDHCLQYITAIGLLFGD 349
                : *: *::       :*::*.*****:.:*********:********:**:.*: ****

prpD            LTAADYEDEVAQDKRIDALREKIVCYEDPAFTADYHDPEKRAIGNAITVEFTDGSRFGEV 420
e.              LTAADYEDNVAQDKRIDALREKINCFEDPAFTADYHDPEKRAIANAITLEFTDGTRFEEV 420
B.              ITAQHYEAETANDPRIDKLRDKMEVTENKTYTEDYLKPDKRSISNAVQVHFKDGTSTEMV 409
                :** .** :.*:* *** **:*:   *: ::* ** .*:**:*.**: :.*.**:    *

prpD            VVEYPIGHARRRADGIPKLIEKFKINLARQFLTRQQQRILDVSLDRARLEQMPVNEYLDL 480
e.              VVEYPIGHARRRQDGIPKLVDKFKINLARQFPTRQQQRILEVSLDRARLEQMPVNEYLDL 480
B.              ECEFPLGHRFRREEAVPKLLEKFSDNLKTHFPDKQHKHIYERCTSYETLQTMRVNEFVDM 469
                  *:*:**  ** :.:***::**. **  :*  :*:::* : . .   *: * ***::*:

prpD            YII 483
e.              YVI 483
B.              FCM 472
                : :
```

Figure 5. ClustalW alignment of prpD from *Salmonella typhimurium, E.coli* strain K-12, and *Bacillus subtilis* strain 168 (24) (25) (26) (27)**.**

The ClustalW alignment shown in Figure 15 is an alignment of the amino acid sequences of the three organisms listed. The reason for doing this is to show the similarity between the sequences of related proteins. The program works by taking

sequences given and aligning them so that the highest percentage of amino acids match between sequences. ClustalW alignments are not necessarily perfect amino acid matches as the software will give a higher score (higher means better match) to similar amino acids (i.e. acidic, basic, etc…) as a similar amino acid may also allow for similar function. An * on the alignment means that that residue is exactly conserved between the amino acid sequences used in the alignment. When a very similar residue is conserved between the sequences a colon is used and a slightly similar amino acid will have a period below the alignment at the residue position. The score given does not equate to a percentage of similarity. It is a score that is derived from the Gonnet PAM250 matrix, which determines how similarity is taken into account and not just exact matches. Less than perfect matches lower the score but still show similarity as they should allow the enzyme to have similar function (25) .

The *S. typhimurium* to *E. coli* alignment had a score of 93.0; *S. typhimurium* to *B. subilis* had a score of 59.0; *E. coli* to *B. subilis* had a score of 61.0 (25). The higher the score the better the match and all of these proteins have high sequence identification with each other. This means that they are all similar in structure which means they likely have similar functionality. As long as active site geometry and composition has been conserved they should still work the same. *S. typhimurium* has a 2-methylcitrate dehydratase that is very highly matched to the homologous enzyme in *E. coli,* but when looking at the ClustalW alignment there may be a ":" or "." which indicate less than perfect matches but those symbols are for all of the sequences in the alignment. In this case, the two sequences in question may have a perfect match but since the third is only

similar it shows as a lower level of similarity on the alignment. For a true sense of the level of similarity between different sequences one must look at the score and compare them relative to the others.

The scores between *B. subtilis* and both *S. typhimurium* and *E. coli* show that there is a high level of sequence conservation between the 2-methylcitrate dehydrates of these two organisms and MmgE of *B.subtilis*. Both scores are very similar, which is to be expected since the proteins from *S. typhimurium* and *E. coli* are so very similar to each other. Since both *S. typhimurium* and *E. coli* are Gram-negative a greater level of similarity would be expected since they are more similar to each other than to the Gram-positive *B.subtilis*. The fact that MmgE has such a high degree of similarity to PrpD from the other two organisms is compelling evidence for it being a 2-methylcitrate synthase.

CHAPTER II

EXPERIMENTAL

II. A Project Overview and Goals

Currently there are complete genome sequences for several organisms, and new ones being completed frequently. The genome of *B. subtilis* has been complete for some time now. This allows for amino acid sequences, and probable protein designations, to be assigned, but there is an incomplete understanding of metabolic pathways in this or any other organism. This is a problem because, without a complete understanding of an organism as a whole it cannot be effectively used as a model organism.

A *long term goal* of the Reddick research group is to characterize the *mmg* operon for a better understanding of *Bacillus subtilis*. The *overall objective* of this thesis is to characterize *mmgE*, which is one of the genes in the *mmg* operon. The *central hypothesis* is that *mmgE* encodes 2-methylcitrate dehydratase, which would play a critical role in a probable methylcitrate cycle. The *rationale* that underlies the research in this thesis is that it will provide knowledge of homologous proteins that have the functionality that is being proposed for *mmgE*. Many other similar organisms have methylcitrate cycles, which gives credence to the hypothesis that *B. subtilis* may have one as well. Characterization of this metabolic pathway allows for new investigations of the organism's metabolism since it will be understood at a more complete level. This is important because *B. subtilis* is a model bacterium used in multitudes of research.

Complete understanding of its own inner workings is necessary to fully this bacterium as a model and biotechnologically useful organism. This comprehension cannot be achieved through reading sequence data alone as it can be incorrect as to the function of specific genes. We have cloned the *mmgE* gene into the BL21 STAR over expression strain of *E. coli* (confirmed by sequencing), as well as purified the His-tagged protein through Ni-NTA chromatography. UV/Vis spectrophotometry as well as HPLC were used to demonstrate enzymatic activity, and direct infusion/LC-MS for identification purposes. Progress towards the long-term goal of the Reddick laboratory and the accomplishment of this Master's project's overall objective will be realized through the following two specific objectives:

1. To produce the *B. subtilis mmgE* protein. In this specific objective we utilized an overall approach involving cloning of the gene into an over expression strain of *E. coli,* BL21 STAR. The gene was cloned into a pET-200 vector that gave antibiotic selectability (kanamycin), and a N-terminus poly-histidine tag to allow for affinity chromatography purification of the protein. Successful cloning was confirmed through DNA sequencing and protein purification was checked through SDS-PAGE, and MALDI analysis after trypsin digestion, which confirmed protein identity.

2. To determine the biochemical function of the *mmgE* protein in *B. subtilis*. The working hypothesis that we tested is that *mmgE* encodes a 2-methylcitrate dehydratase. This is an important enzyme in the methyl citrate cycle, which is part of branched fatty

acid degradation and propionate metabolism. We tested this hypothesis with *in vitro* activity assays. These assays used substrate turned over by the previously characterized protein *mmgD* (a citrate/2-methylcitrate synthase), as the 2-methylcitrate substrate for *mmgE* as well as recently commercially available 2-methylcitrate purchased from Sigma Aldrich. Activity assays were performed using a UV/Vis spectrophotometer, as well as HPLC, to analyze activity and MS direct infusion to identify HPLC peaks.

II. B Expected Significance

Right now, the full genome of *B. subtilis* is known, which is the start of having a full understanding of the organism. Understanding *mmgE,* and by extension the *mmg* operon, is important to the *B. subtilis* field. Characterizing the full operon proves or disproves proposed protein designations and shows what metabolic pathways are present within the organism. Knowing these pathways allows for better use of *B. subtilis* as a model organism as well as add to the knowledgebase of metabolic biochemistry and bioinformatics. Better understanding of the biochemistry of one organism can help with others since new methods and comparisons can be made. Gene annotation is important in this endeavor, but annotation tools are not infallible and errors have been found before. Experimentally characterizing each gene is necessary to be completely certain of identities, functions, and to elucidate metabolic pathways

II. C Cloning of *mmgE*

The *mmgE* gene was first isolated through PCR using Phusion polymerase and genomic DNA isolated from *B. subtilis* strain 168. The primers used were: 5'-TTACATGCAGAACATGTCTACGAATTCATTTAC-3', and 5'-CACCATGCCGAAAACGGATCGTG-3', with the following thermocycler conditions: 30 s at 98 °C; 30 cycles of 98 °C (10 s), 55 °C (30 s), and 72 °C (60 s); after the 30 cycles the temperature was held at 72 °C for 10 min. After the reaction was complete the sample was held at 4 °C until the next steps.

The PCR product was purified from the reaction using the PCR Purification Kit from Qiagen, according to the manufacturer's instructions. The resulting sample was analyzed by 1% agarose gel electrophoresis. The correctly-sized PCR fragment (*mmgE* = 1416 bp) was immediately cloned into pET-200 using topoisomerase-based cloning, using the TOPO-cloning kit from Invitrogen, according to the manufacturer's instructions. All transformations and propagation of plasmids were done using TOP10 chemically competent *E. coli* (also from Invitrogen).

The PCR design was such that the *mmgE* gene was cloned in-frame with an N-terminal His_6-Tag encoded by the pET-200 plasmid. The plasmid was sequenced by dye-terminated sequencing (SeqWright, Inc.), and the length of the *mmgE* gene necessitated that this was done by two rounds of sequencing, first using universal primers to sequence roughly 600 bp of each end of the gene, followed by new primers designed from these sequences to yield complete coverage of the middle region of the gene. These two

experiments together confirmed that the *mmgE* thus cloned into pET-200 matched the expected sequence reported in the genome project.

II. D Overexpression of the MmgE Protein

The vector, containing *mmgE*, was purified from 5 mL cultures of the TOP10 cells using the Plasmid Miniprep Kit from Qiagen, according to the manufacturer's instructions. This isolated DNA was verified by PCR using the cloning primers, with analysis by agarose gel electrophoresis, and used to transform BL21 STAR *E. coli* using the kit from Invitrogen and their instructions. The transformations were plated on LB-agar plates containing 30µg/mL kanamycin and those with growth were stored at 4°C. Colonies were selected and after being restreaked, they were grown up into individual 5mL cultures and the plasmid was purified as before. Each of these purified plasmids was analyzed by PCR and agarose gel to confirm the presence of *mmgE*. The colony that gave the most intense PCR band was selected and used to make subsequent plates and a 10% glycerol cell stock that was stored at -80°C. Large one liter flasks of LB were used to grow up enough bacteria for protein-purification. The LB (10g tryptone, 10g NaCl, 5g yeast extract, 1 liter nanopure water pH 7.5) was infected with 1 mL of a starter culture after adding kanamycin (30µg/mL final concentration) then grown until an OD595 of 0.6 was achieved. Then, IPTG (1mM final concentration) was added to induce the culture and it was allowed to shake overnight at 37°C. The next day the culture was centrifuged at 7,480g for 30 minutes at 4°C. These pellets were stored at -80°C, or were lysed immediately (in 20mL of 1x binding buffer) by sonicating for six

minutes in thirty second bursts, while on ice. This cell slurry was then spun at 16,000g, for 30 minutes, at 4°C.

II. E Purification of MmgE

A 2 mL Ni-NTA column was equilibrated with the following: 6mL of 1x strip buffer (0.5M NaCl, 20mM Tris-HCl Ph 7.9), 6mL of DI H2O, 10mL 1x charge buffer (50mM nickel sulfate), then 6mL of 1x binding buffer (0.5M NaCl, 20mM Tris-HCl pH 7.9, 5mM imidazole). The column can be stored at 4°C with excess binding buffer. The bacteria pellet was resuspended in 10mL of 1x binding buffer and sonicated for six minutes (thirty second bursts), while on ice. Then, after spinning down (16000g), the supernatant was added to the column through a 0.45µm syringe filter. This solution was allowed to flow through the column under gravity to load the protein onto the column. After loading, the column was washed with: 20mL of 1x binding buffer, then 12 mL of 1x wash buffer (0.5M NaCl, 40mM Tris-HCl pH 7.9, 60mM imidazole), and lastly 12mL of 1x elute buffer (1M imidazole, 0.25M NaCl, 10mM Tris-HCl pH 7.9). Once elute buffer was added 1mL fractions were taken. As fractions were collected they were tested for protein with a quick qualitative assay. To 1mL of Bradford reagent was added 33µL of one of the fraction. If the Bradford reagent turned blue it was indicative of the presence of protein so that fraction was saved. Fraction collection was stopped once there was no more protein coming off the column. Fractions collected were analyzed through SDS-PAGE to check for a correct sized protein of 57kDa. Protein containing fractions were dialyzed, on the bench top, in four liters of 27mM Tris buffer at pH

7.5. After dialysis the protein solution was aliquoted into microcentrifuge tubes (with 10% glycerol) and stored at -80°C for use in assays. MmgE modification: A 1M imidazole concentration in the elute buffer caused a precipitation problem which was solved by decreasing it to 200mM imidazole. MmgD modification: Binding, wash, and elute buffers all have 10% glycerol, but was compatible with 1M imidazole in the elute buffer.

II. F Agarose Gel Protocol

To 30mL of 1x TAE, 300mg of agarose was added and heated by microwave until completely melted and the solution was homogeneous. Then, 10µL of ethidium bromide was added to the agarose solution, and it was poured into a casting apparatus. Once cooled and set, the gel was placed in the running tray, and samples were added to the wells. Next, 1x TAE was added to the tray until the gel was completely covered, then the cover was attached and the gel was run at 120V and 400mA until the dye front was about 1cm from the bottom edge. The gel was then imaged using a gel camera and trans UV light.

Samples were prepared by adding 3µL of 6x loading dye to 15µL of the PCR product. The ladder sample was prepared by mixing 1µL of 2-log ladder to 1µL of 6x loading dye, and 4µL of microbiology grade water.

II. G SDS-PAGE Protocol

The resolving gel was made by mixing the following reagents in a 50mL Falcon tube in order: DI H_20: 3400µL, Bis-Acrylamide/Acrylamide: 4000µL, 1.5M Tris Buffer: 2500µL, Ammonium Persulfate (APS) at 100mg/900µL DI H_2O: 120µL20% SDS: 50µL, TEMED: 15µL. Samples were prepared by mixing 25µL of the protein sample with 75µL of SDS sample buffer. These samples were heated for five minutes at 95°C. The samples were then centrifuged at 15,700g for two minutes. Once the gel was polymerized it was set in the running tub and the inner chamber was filled to the top with 1x run buffer. The outer chamber was filled to the proper (2 or 4 gel) line, then the lid was placed and the gel was run at 120V and 80mA until the dye front was about a centimeter from the bottom. The gel was then transfered to a tip box and covered with staining solution, from 1 liter stock (100mL glacial acetic acid, 500mL DI H_2O, 400mL methanol, and 1g Coomassie Brilliant Blue R250). This was then shaken for half an hour before being poured off. Next, destain (same as stain minus the Coomassie Brilliant Blue R250) solution was added and allowed to shake for half an hour and the destain was repeated until the bands were clearly visible. SDS-PAGE gels were imaged in the same manner as agarose gels.

II. H MmgD Purification and Analysis (28)

MmgD was isolated and purified using the previously stated Ni-NTA column protocol, using previously made cell stocks stored at -80°C (from Rejwi Acherya). After purification, and subsequent confirmation SDS-PAGE, an activity assay was performed (as per instructions in the thesis written by Rejwi Acharya) to confirm MmgD was isolated and make sure the protein was active. The assay required: 300µM propionyl-CoA, 5,5'-dithiobis-(2-nitrobenzoate) or DTNB 0.1mM, 20µL of MmgD (any concentration), and 500µM oxaloacetate in 1M Tris buffer (28). Once the MmgD protein was determined to be active the same reaction, minus DTNB, was performed to create the substrate for MmgE.

II. I MmgE Activity Assay With MmgD Product

MmgD was allowed to turnover product for one hour then the reaction was stopped by heating to 95°C for five minutes. The reaction tube was then cooled and centrifuged for two minutes at 15,700g to remove the denatured protein. Next, the solution was transferred to a cuvette and placed in the spectrophotometer, which was set at 240nm (16). The MmgD reaction solution was used to blank the spectrophotometer, and then a 100µL aliquot of MmgE was added to the cuvette. The reaction was inverted a few times to mix then immediately placed in the spectrophotometer and data was collected for an hour.

II. J MmgE Activity Assay with Commercial 2-Methylcitrate

Recently 2-methylcitrate has become commercially available, as a trisodium salt of (2RS,3RS)-2-methylcitrate mixture of diastereomers, and new reactions were performed using this purchased substrate. The 2-methylcitrate was stored at -20°C in 1mL aliquots at a concentration of 20mM. Reactions were set up using a 1mL total volume. The substrate was at a 1mM concentration and 100µL of isolated MmgE protein (any concentration) was added to the reaction buffer (20mM Tris pH 7.5). The reaction was analyzed for one hour at 240 and 220nm in a UV/Vis spectrophotometer.

Timed assays were also performed for HPLC analysis. The reaction components were the same as for the UV/Vis spectrophotometer though the protein amount was varied in some reactions to 40µL, and 60µL. These reactions were allowed to run on the bench top in 5mL volumes. Every fifteen minutes a 1mL sample was taken and quenched with 100mM potassium phosphate buffer (monobasic) at pH 2.9. The samples were then centrifuged at 15,700g for five minutes before being carefully transferred to HPLC vials and injected.

One other assay was done with the commercial 2-methylcitrate. A 1mL solution was prepared using the same concentrations as the UV/Vis assay with commercial substrate and allowed to run for an hour with a broad spectrum scan (200nm-400nm) taken every ten minutes.

II. K HPLC Analysis

The first tries at HPLC used two pumps (one with nanopure H2O with 0.1% TFA, and the other with methanol and 0.1% TFA). A C18 column (Waters 3.9mmx150mm 4 micron) was attached to the HPLC. Several 250µM solutions were prepared and sent through the HPLC to test for separation. These samples were run with a mobile phase of 95% water and 5% methanol with a flow rate of 1mL/min and the detector set at 212nm. Both MmgD and MmgE reactions were analyzed by the HPLC as well.

A new column (Phenomenex Synergi 4u Hydro-RP 80A 250x4.60mm 4 micron) was used for the following analyses. The samples were run through the column at a flow rate of 0.700 ml/min. The mobile phase consisted of 20mM potassium phosphate buffer, pH 2.9 and methanol. A gradient was used that started out 0% methanol which increased up to 15% during ten minutes, and held at that concentration for another forty minutes then the concentration of methanol decreased to zero over the next ten minutes.

II. L Mass Spectrometry

Mass spectrometry was performed using two methods on the same triple quad device. The first was direct infusion. This was performed by pumping a solution directly into the ESI source at a rate of 0.5µL/minute. The lines and syringe were cleaned with an excess of acetonitrile solvent between samples. Data was collected as 200 scans each taken every half second. This data was then averaged to give relative abundance over the whole set of scans.

The second method was LC-MS. This involved using an UPLC in line with the ESI source on the triple quad mass spectrometer. The method used on the UPLC was a fifteen minute run with an organic gradient. The gradient started out with 100% water (with 0.19% formic acid) and at three minutes was 90% water/10% methanol. This gradient reverted back to starting conditions at twelve minutes.

The enzymatic reaction that was analyzed by this method was the same as analyzed by HPLC. but with ammonium bicarbonate buffer (20mM ammonium bicarbonate pH 7.5, MmgE, 1mM 2-methylcitrate). The reaction was allowed to run for an excess of two hours to be sure the reaction had run to completion. No quenching was used. Control experiments were lacking either MmgE or substrate

II. M CitB Expression

The *B. subtilis* AWS198 mutant strain expressing His-tagged CitB directly from the chromosome, was obtained from the Sonensheim group as a glycerol stock and stored at -80°C (29). This stock was used to streak LB agar plates with 34µg/mL chloramphenicol. These plates were grown overnight at 37°C. Individual colonies off of these plates were selected and used to infect 5mL LB starter cultures with chloramphenicol and allowed to grow overnight with the cap loose since *B. subtilis* is an obligate aerobe. Starter cultures were used to infect 1L Difco Sporulation Media (DSM), at pH 7.6, broth and allowed to grow to an OD_{600} of approximately 1.0 and then spun down and stored in the -80°C freezer as with other bacteria pellets (29). The DSM was made by adding 8g of bacto nutrient broth (Difco) to 1L of DI water that contains 10%

(w/v) KCl and 1.2%(w/v) $MgSO_4 \cdot 7H_2O$ (29). Just prior to use 1mL of each was added: 1M $Ca(NO_3)_2$, 0.01M $MnCl_2$, and 0.001M $FeSO_4$ (29). The liter of media was then infected with a milliliter of a starter culture and allowed to shake at 37°C until the culture reached an OD_{600} of about 1.0. At this point that culture was split into two large centrifuge bottles, balanced, and then spun at 7,480g for 15 minutes at 4°C to pellet the bacteria. These pellets were then washed twice with cold tris-citrate buffer (8g citric acid, 20.1g tris, 500mL water) then stored at -80°C until used (29).

II. N CitB Purification

The his-tagged protein (CitB aconitase) was purified through Ni-NTA chromatography. The Ni-NTA column was equilibrated with: 6mL of 1x strip buffer, 6mL of DI water, 10mL of 1x charge buffer, and 6mL of binding buffer then stored at 4°C until use. The pellet was resuspended in 10mL of a special buffer (100mL: 1.5g KCl, 0.79g Tris-HCl, 10mL glycerol, 10mL nonidet P-40, 0.007g EDTA, 0.017g PMSF, 0.007g dithiothreitol) and lysed with eight minutes of sonication (30 second bursts on ice). The slurry from sonication was then centrifuged at 15,700g for twenty minutes. The supernatant was then added through a 0.45micron syringe filter to the Ni-NTA column and allowed to flow under gravity. Then 20mL of binding buffer(0.5M NaCl, 20mM Tris-HCl pH 7.9, 5mM imidazole), 12 mL of wash buffer (0.5M NaCl, 40mM Tris-HCl pH 7.9, 60mM imidazole), and 12 mL of elute buffer (1M imidazole, 0.25M NaCl, 10mM Tris-HCl pH 7.9) were run through the column. The elute buffer was collected in individual 1mL fractions. These fractions were tested for protein content by

mixing 33µL of the fraction with 1mL of Bradford reagent. If the Bradford reagent turns blue then protein was present in the fraction. This was a qualitative assay as an amount was not determined this way. Samples of each fraction were kept before dialysis (4L 25mM Tris-HCl, pH 7.5 left on bench top overnight) and stored in -20°C.

II. O Concentration of Protein Samples

Multiple (5) one liter cultures were done to get enough protein solution (10mL) to concentrate with a VivaSpin concentrator. Once the isolation procedure had been performed as before and enough protein solution isolated it was all placed in a VivaSpin column (5000 MWCO). This column was placed in a refrigerated centrifuge and spun at 3000g until concentrated to 1mL.

II. P CitB Activity Assay

Before the assay was started the protein was first allowed to incubate for an hour with an activation buffer (100mM Tris-HCl pH 7.4, 1mM $Fe(NH_4)_2(SO_4)_2$, 50 mM L Cysteine pH 7.4) (30). The reaction contained 90mM Tris-HCl pH 8.0, 20mM DL-Isocitric acid trisodium and 50µL of the activated protein fraction (29). The reaction was monitored in an UV/Vis spectrophotometer at 240nm and blanked with just buffer and enzyme (29). The reaction was started by addition of the substrate (isocitric acid). The reaction was monitored for an hour.

CHAPTER III

RESULTS AND DISCUSSION

III. A Cloning of *mmgE*

After PCR, *mmgE* was cloned by topo isomerase methodology into an over expression vector pET200, which was transformed into *E. coli*, BL21 STAR from Invitrogen. Transformants from the transformation were tested for the presence of *mmgE* by PCR and one transformation was selected for use. This was due to there being a stronger single band at the correct size (1416 Kb) shown in Figure 6. This transformant was used to make a glycerol stock and used for all subsequent experiments. For complete confirmation, DNA sequencing was performed, which conclusively confirmed that *mmgE* was successfully cloned into the pET200 vector.

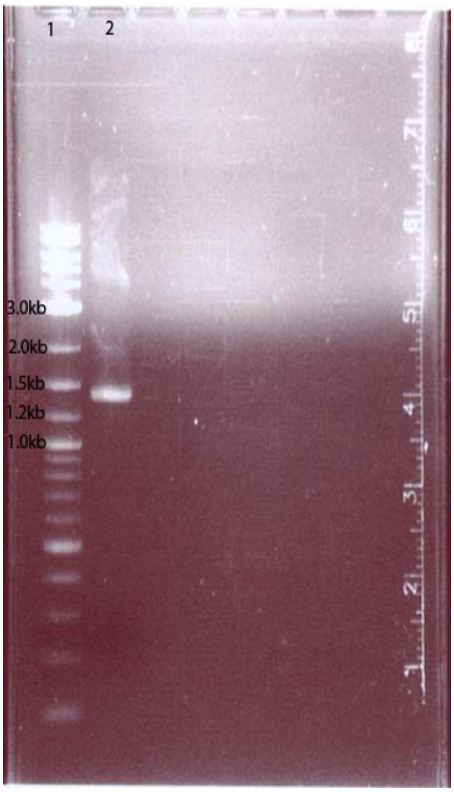


Figure 6. Agarose gel confirming the presence of *mmgE*. Lane 1 was 2-log ladder and lane 2 contained the PCR sample.

III. B Purification of MmgE

By cloning *mmgE* into this vector a His_6 tag was added to the N-terminus. This allowed for Ni-NTA affinity chromatography for isolation and purification of the protein encoded by *mmgE*. The one problem encountered while trying to purify MmgE was that the 1M imidazole concentration in the elute buffer caused a salting out effect that resulted in all the protein in solution to precipitate. This was solved by reducing the imidazole concentration, in that buffer, to 200mM. SDS-PAGE was performed to check for a good purification and that a protein of the correct mass was eluted.

The protein that eluted from the column was used in all UV/Vis spectrophotometry and HPLC work. Before this was done, an in-gel trypsin digest and MALDI-MS confirmed that the purified protein matched the sequence of MmgE from *B. subtilis.*

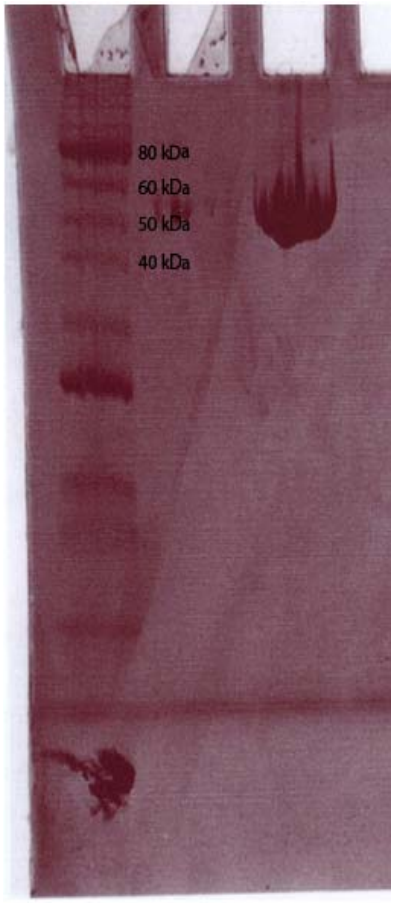


Figure 7. SDS-PAGE of MmgE

III. C MmgD

To produce the 2-methylcitrate needed as the substrate for MmgE, MmgD was purified from a lab strain previously made and stored at -80°C. The protein had a similar sensitivity to precipitating out of solution as MmgE but this was solved by adding 10% glycerol to all buffers when performing nickel affinity chromatography, as previously described. Purification was confirmed using SDS-PAGE (see figure 3). Using an UV/Vis spectrophotometer the activity assay, involving DTNB (5.5'-dithiobis-(2-nitrobenzoate)), for MmgD was performed which showed that MmgD was highly active. MmgD was subsequently used, without the DTNB, to produce 2-methylcitrate, the substrate for MmgE.

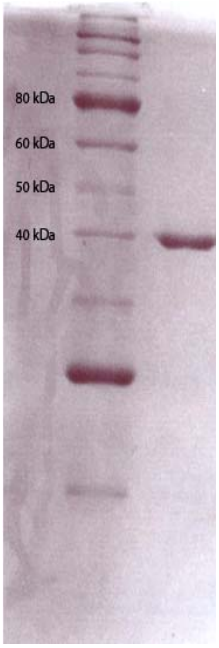


Figure 8. SDS-PAGE of MmgD

Each time it was used to produce substrate for MmgE, MmgD was allowed to turn over product for an hour, or more, and the change in absorbance in the activity assay for MmgD showed that it was performing a reaction. UV/Vis spectroscopy did not provide definitive answers as to the activity of MmgE. There was a slight slope but never enough to equivocally state that MmgE was catalyzing a reaction. Control experiments were done without enzyme and without substrate. Neither produced the same slight slope as the full reaction. Later UV/Vis spectroscopy was revisited with commercial 2-methylcitrate. Activity was observed by an increase in absorbance at 240nm and it was theorized that the CoA-SH that was a product of the MmgD reaction was interfering with observing the MmgE reaction.

III. D HPLC

Due to the trouble of getting an answer through using the UV/Vis spectrophotometer, HPLC was used to check for enzymatic activity. The first column used had a C18 stationary phase and the problem was that this column does not have the properties to separate the polar compounds that were being analyzed. The stationary phase of this column consists of a carpet of eighteen-carbon chains which would interact the most with non-polar compounds. This is problematic for getting retention and separation of the acids that are both the substrates and products of the enzymes encoded by the *mmg* operon. 2-methylcitrate is quite polar with the three carboxylic acid groups in its structure. We tried to get around this by adding trifluoro acetic acid to change the protonation state of the acid groups, as well as varying amounts of methanol in the

mobile phase. None of these variations in protocol changed retention or separation so that column was abandoned.

III. E Phenomenex Synergi Column

The previously mentioned Synergi column from Phenomenex was put in place of the C18 column and the mobile phase was changed to 20mM potassium phosphate buffer, pH 2.9, and a low concentration methanol gradient was added to help get compounds to elute off of the column since the retention was so strong it was taking over an hour to see peaks. As a test, citrate (Figure 9), aconitate (Figure 11), and isocitric acid (Figure 10) were easily separated. Moreover methylcitrate was retained on the column as well, so it was decided to proceed with HPLC. Running a 1mM sample of 2-methylcitrate (Figure 12) through the column showed that the addition of the single extra methyl group caused the compound to stay on the column several minutes longer. Because of this citrate and methylcitrate would elute off of the column at different times.

Isocitrate was the quickest compound to elute off of the column at 4.88 minutes, referring to Figure 9. Citrate eluted at 9.43 minutes and the aconitate came off after that at 11.54 minutes. The 2-methylcitrate eluted off of the column with two peaks at 13.99 minutes and 11.52 minutes (Figure 12). This means that if an HPLC chromatogram of a full MmgE enzymatic reaction has a peak elute after 2-methylcitrate, it should correspond to methyl-cis/trans-aconitate. If a new peak elutes before the 2-methylcitrate, it would likely correspond to 2-methylisocitrate.

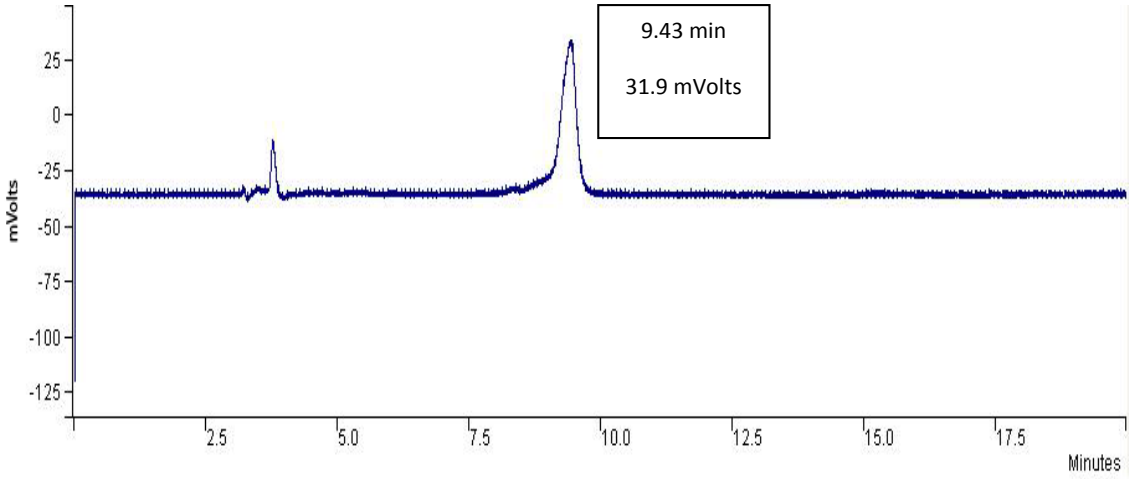


Figure 9. HPLC chromatogram of 1mM citrate (isocratic 100% phosphate buffer).

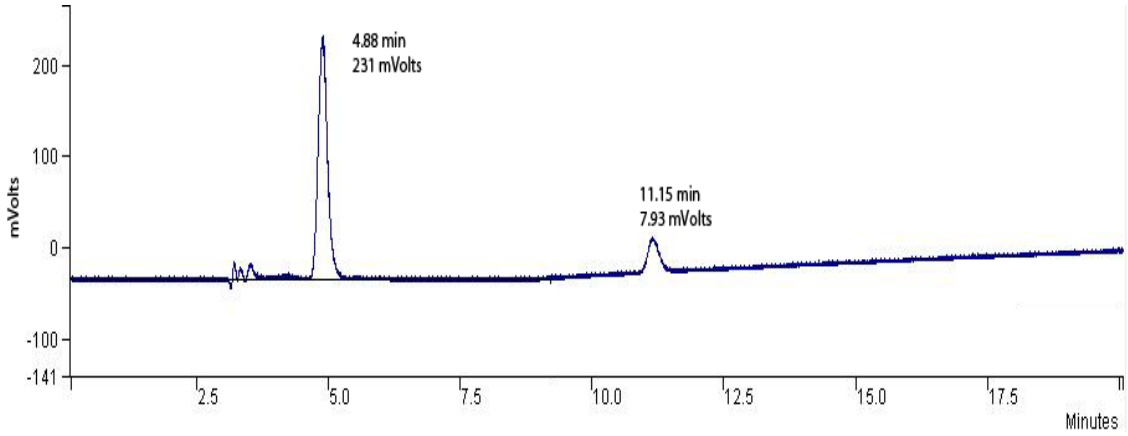


Figure 10. 1mM isocitrate in phosphate buffer pH 2.9 (isocratic 100% phosphate buffer).

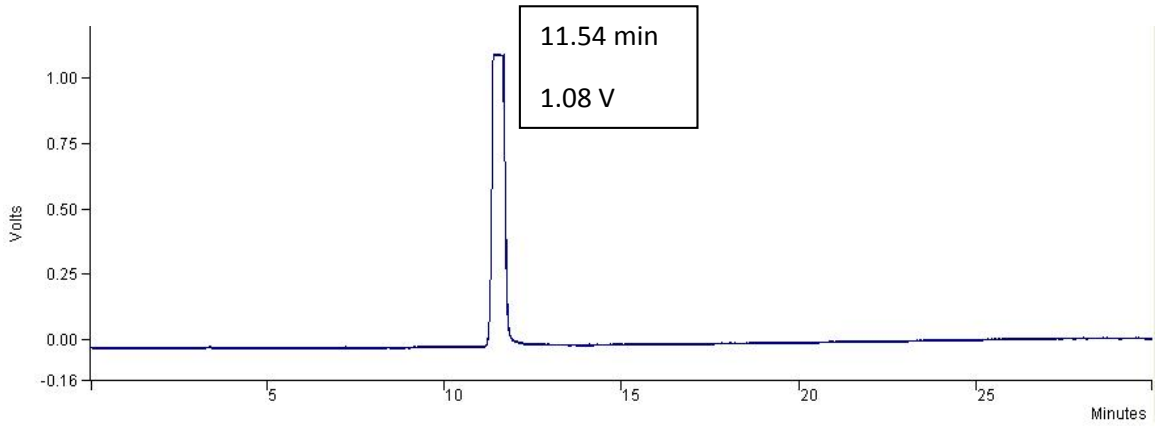


Figure 11. 1mM Trans-aconitate in phosphate buffer pH 2.9 (isocratic 100% phosphate buffer).

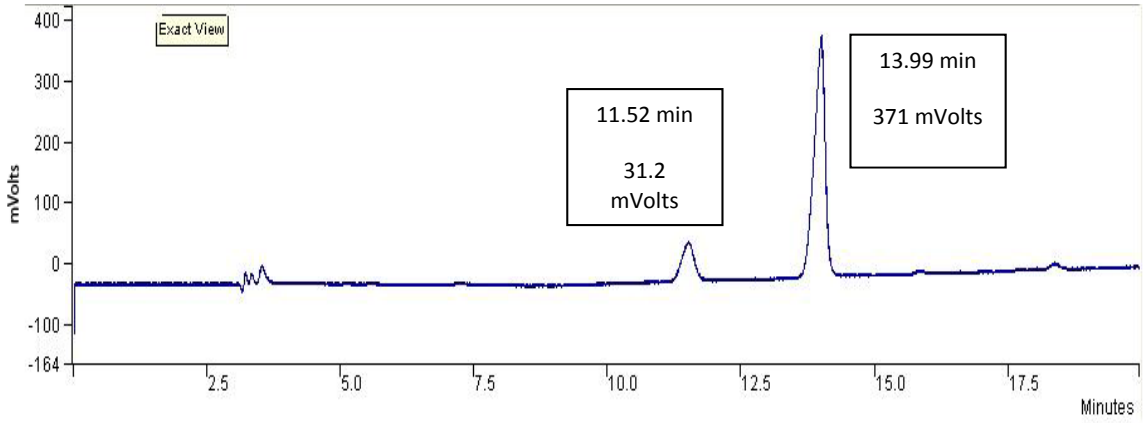


Figure 12. HPLC chromatogram of 1mM 2-methylcitrate (isocratic 100% phosphate buffer).

The use of the potassium phosphate buffer, added to the sample for a final concentration of 100mM, and to bring the pH to 2.9, was found to sharpen the peaks as they were broad when injecting directly with Tris buffer or acetonitrile. This made the purpose of the phosphate buffer quench twofold: first it was to stop the enzyme reaction

so that assays could be analyzed at different stages of completion, and second it improved the peak shapes and resolution.

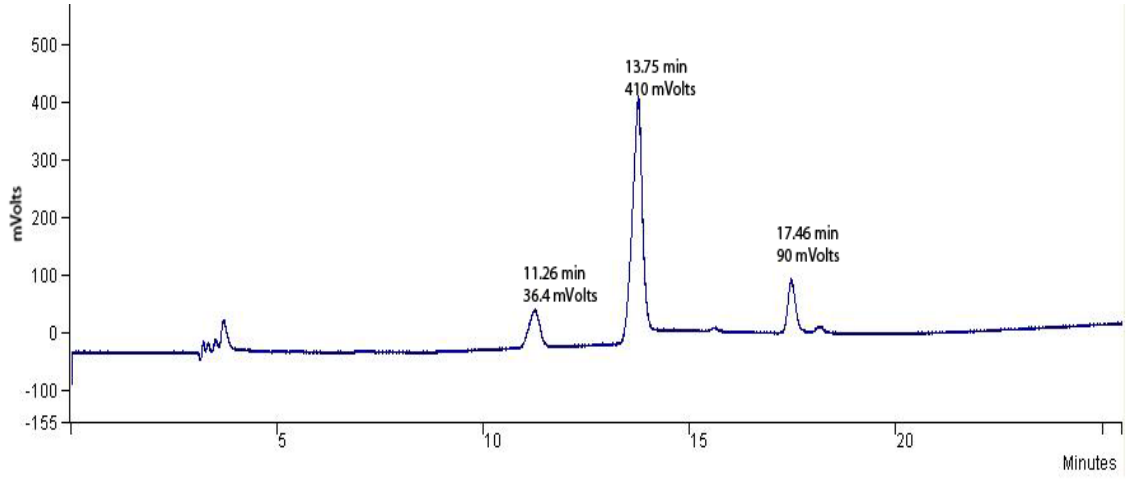


Figure 13. MmgE enzyme reaction at 15 minutes. Quenched with 100mM potassium phosphate buffer pH 2.9. (15% MeOH gradient in phosphate buffer).

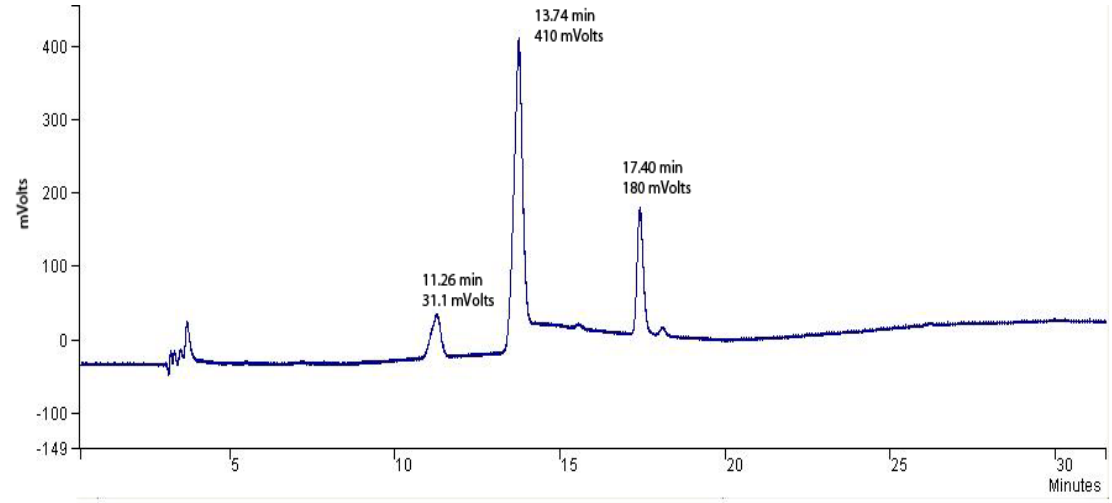


Figure 14. MmgE enzyme reaction at 30 minutes. Quenched with 100mM potassium phosphate buffer pH 2.9 (15% MeOH gradient in phosphate buffer).

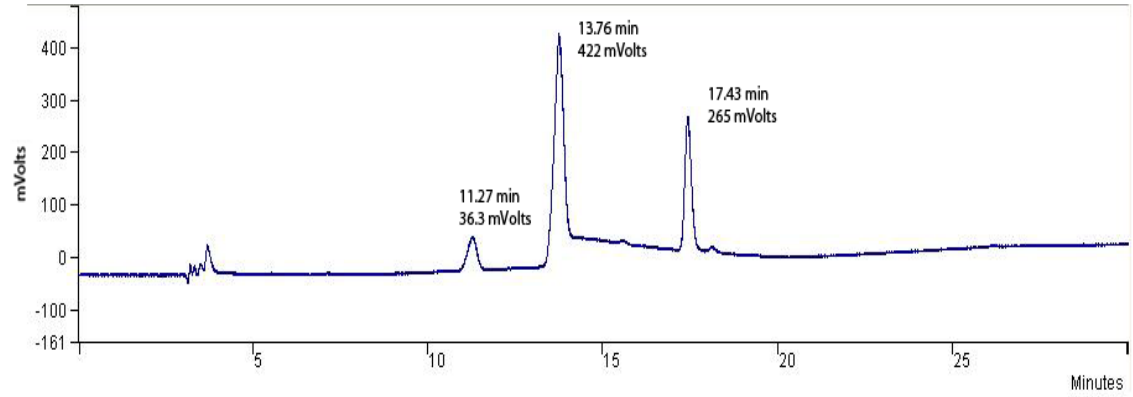


Figure 15. MmgE enzyme reaction at 45 minutes. Quenched with 100mM potassium phosphate buffer pH 2.9 (15% MeOH gradient in phosphate buffer).

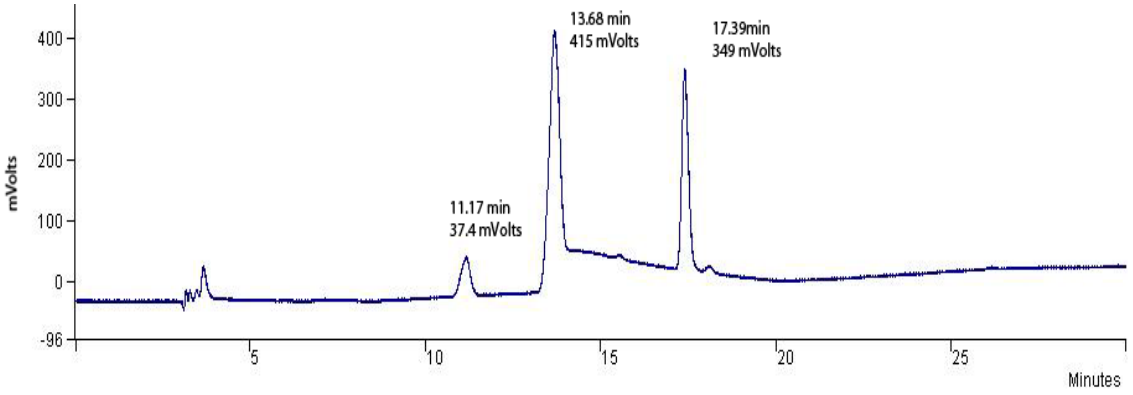


Figure 16. MmgE enzyme reaction at 60 minutes. Quenched with 100mM potassium phosphate buffer pH 2.9 (15% MeOH gradient in phosphate buffer).

Figures 13-16 show a timed reaction of MmgE and 1mM 2-methylcitrate. Every fifteen minutes a sample was taken and quenched with 100mM phosphate buffer, pH 2.9 then centrifuged for five minutes at 15,700g and then injected into the HPLC. The timed assay shows that the MmgE protein is active and using 2-methylcitrate as its substrate. The 2-methylcitrate peak does not decrease in amplitude, in an inverse relationship with the product because the commercially available compound is a racemic mixture of 85% purity. This is important because the enzyme can only use one form of it so 75% of the

2-methylcitrate remains unreacted no matter how long active enzyme is present. The enzyme can only work with one of the four diastereomers present in the commercial 2-methylcitrate. Every fifteen minutes the product peak, at approximately 17.40 minutes, increases in size (essentially doubling) showing that 2-methylcitrate is being converted into product.

The non-methylated compounds can be used as a template for the methylated compounds. Isocitrate eluted first, before citrate, and both aconitate isomers (cis and trans) eluted after citrate. Taking that into account, it is likely just from the HPLC data that the product is 2-methyl-cis/trans-aconitate as the product peak elutes after 2-methylcitrate. The isomeric form of the aconitate cannot be derived just from this data since there was no appreciable separation of the non-methylated aconitate isomers.

In addition to the HPLC data, the reaction was viewed again on the UV/Vis spectrophotometer as another way to see the reaction in progress. This time the commercial 2-methylcitrate was used as substrate instead of relying on MmgD to synthesize it, so that concentration could be controlled with greater confidence.

III. F Direct Infusion Mass Spectrometry

The first mass spectrometry, in this project, done was direct infusion mass spectrometry on a triple quad mass spectrometer. Doing this analysis allowed us to quickly see an average of the masses ionized by the ESI source. Each data set was of 200 scans that were taken every half second. The data for 2-methylcitrate, Figure 18, shows the various masses that 2-methylcitrate exists in: 229, 251, 273, and 295m/z. All of these

masses are missing in Figure 19, which is the sample lacking 2-methylcitrate confirming the identity, of the previously mentioned peaks in Figure 18, as 2-methylcitrate. Each of these masses are 22m/z apart because of a different number of sodium cations associating with the negatively charged oxygens on the compound. This figure shows which masses to look for in subsequent reactions to see if the substrate, or product, is present.

The full reaction is shown in Figure 20, and one thing to note is that there are small peaks corresponding to the masses for 2-methylcitrate as shown in Figure 17. Some of the substrate remains in solution, and this is expected, because the compound was purchased as a mixture of four diastereomers. The enzyme can only complex with one form of the substrate so the others should remain unreacted. A possible mass for the product is the 189m/z peak visible in Figure 20. This corresponds for the expected mass of methylaconitate. The difference between 2-methylcitrate and the product is the loss of water which is 18m/z This can be compared to the LC-MS results, specifically Figure 21where the 206m/z peak is visible. The difference in mass is likely due to the association of an ammonium (18m/z) with the product. The mass of 2-methylcitrate is 206 mass units, but that peak is not present in Figure 17. This is because positive ions are associating with negatively charged oxygen. Since the 206m/z peak is missing in the chromatograms, of samples having only 2-methylcitrate, we can see that 2-methylcitrate normally complexes with at least one ion, so both 189m/z and 206m/z are likely candidates for being methylaconitate.

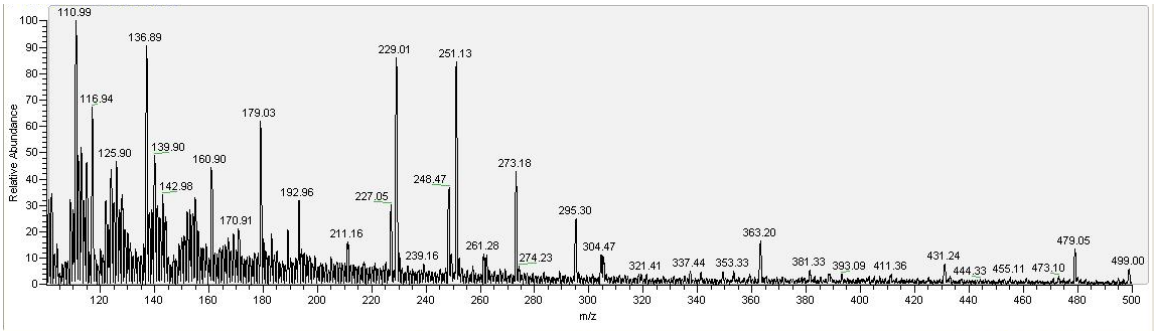


Figure 17. Direct infusion lacking MmgE. Sample was in 20mM ammonium bicarbonate buffer.

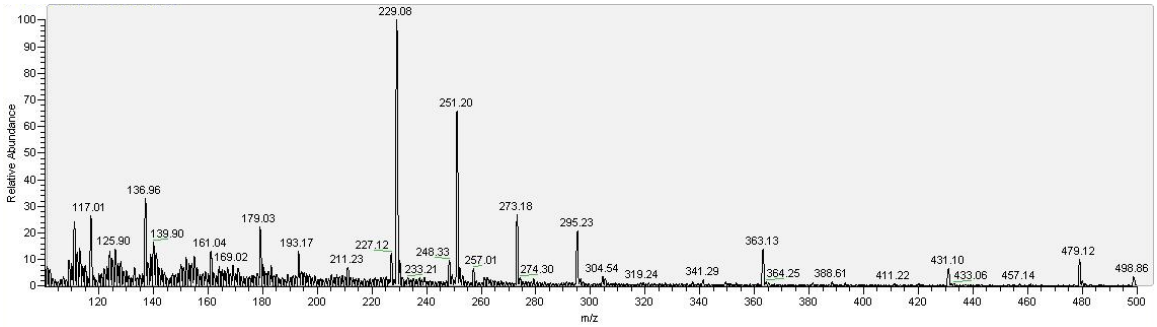


Figure 18. Direct infusion of 2-methylcitrate. Sample was in nanopure water

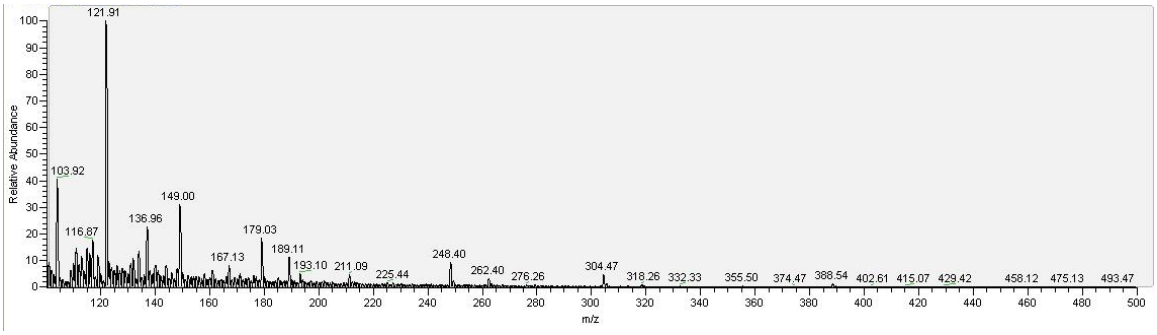


Figure 19. Direct infusion lacking 2-methylcitrate.

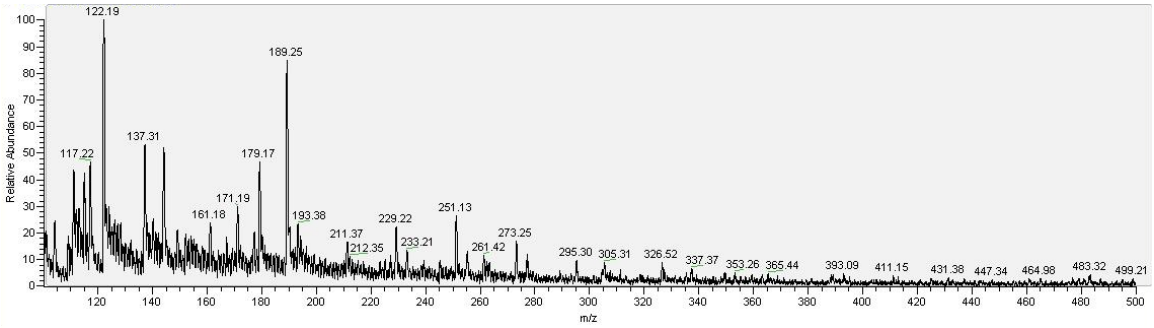


Figure 20. Direct infusion of the full reaction.

III. G LC-MS

The previously mentioned HPLC experiments elucidated activity, by showing a new peak appear in a new location. A probable identification, of the product compound, was made relative to the HPLC chromatograms of 2-methyl-cis/trans-aconitate. This was due to comparing elute order and peak characteristics between the methylated and non-methylated forms of the compounds. For identification purposes, LC-MS (UPLC in line with a triplequad mass spectrometer) and direct infusion MS (triplequad mass spectrometer) were performed. These experiments helped to confirm the identity of compounds by giving exact masses of compounds that are in the solutions.

Full reactions (MmgE protein, 20mM ammonium bicarbonate buffer pH 7.5, and 1mM 2-methylcitrate) were allowed to run at a room temperature and for an excess of two hours to allow for maximum completion. This was to ensure the maximum amount of product was available for mass spectrometry. The controls were the same as the full reaction with each missing a single component for comparison with the complete reaction. One control was lacking 2-methylcitrate and the other was lacking enzyme.

The controls were compared to the full enzyme reaction to allow common masses to be eliminated so that the product of the reaction is obvious.

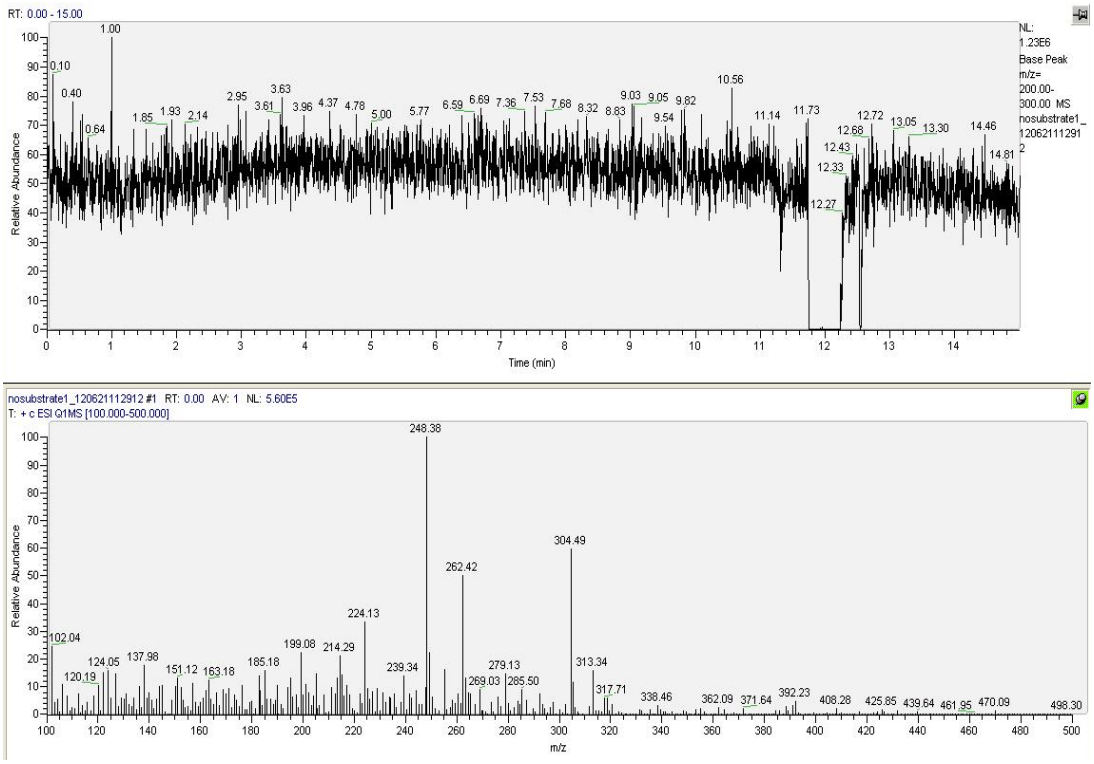


Figure 21. Control experiment lacking 2-methylcitrate. MS spectrum of mass range 100-500 m/z.

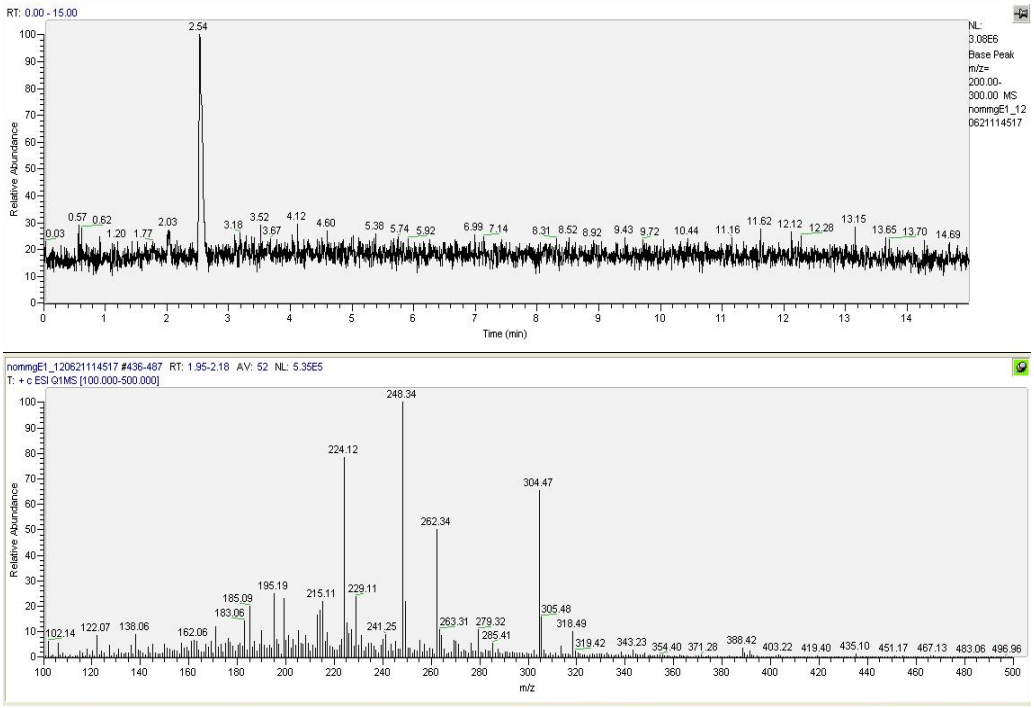


Figure 22. Control experiment lacking MmgE. MS spectrum of peak centered at 2.54 min.

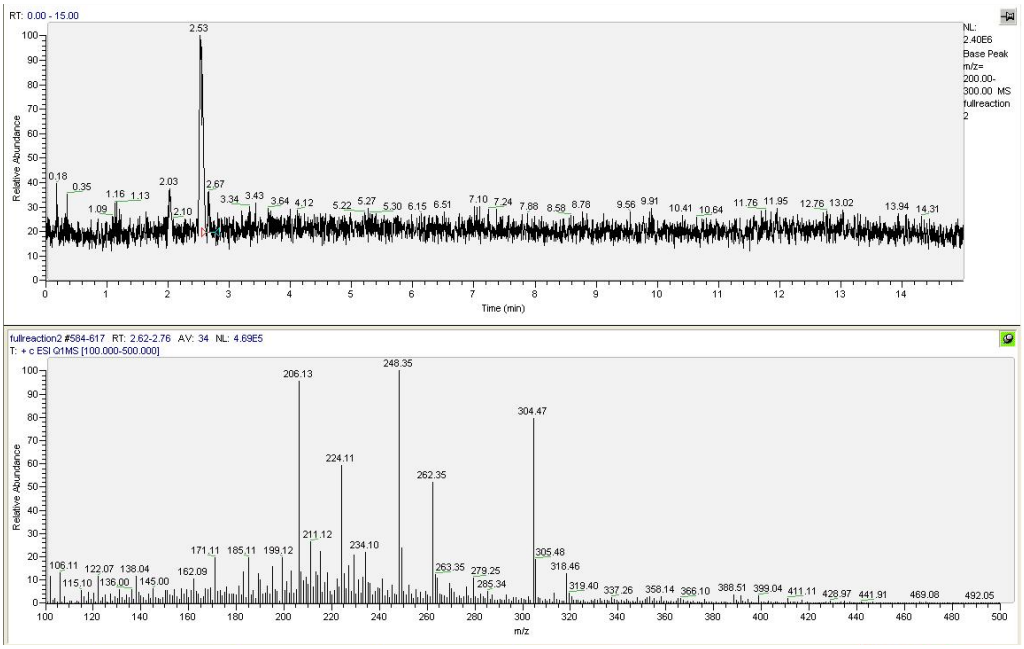


Figure 23. LC-MS analysis of MmgE reaction. MS spectrum of masses centered on peaks at 2.53 and 2.67min

Figure 20 shows a series of masses (229, 251, 273, and 295m/z) that correspond to the substrate, 2-methylcitrate. These are all due to different protonation states and the fact that it is a trisodium salt. Each of those masses are approximately twenty-two m/z apart with corresponds to the loss of a proton and the gain of a single sodium atom. The peak at 248m/z was in all LC-MS spectra which likely means it is due to solvent and/or contamination and there is a peak right next to it, slightly higher m/z that it is masking. This higher mass peak that is being masked would likely correspond to one of the 2-methylcitrate expected masses. Figure 22 shows very similar masses to Figure 20, both of which contain 2-methylcitrate, the only difference being that the sample in Figure 24 was not dissolved in ammonium bicarbonate buffer, but in nanopure water. The slight difference in m/z between the two figures is likely due to protonation since the water that made up the bulk of the mobile phase, for UPLC, was slightly acidified with formic acid. The acidification was necessary for creating positive ions through protonation as analysis was performed in positive mode.

The full enzymatic reaction, which is shown in Figure 21, a single mass stands out. At 206m/z a high relative abundance peak shows up which corresponds to the mass expected for methylaconitate if it has complexed with an ammonium. The formula mass of 2-methylcitrate in its pure form is 206 g/mol, but this mass was not present in any analysis done lacking MmgE. The lack of this peak when 2-methylcitrate is not in the presence of enzyme shows that it is not present in this form in solution. It always associates with an ion, and positively charged ammonium likely accounts for the change in mass. This means that MmgE catalyzes one step of the reaction that takes 2-

methylcitrate to 2-methylisocitrate. If MmgE had produced methylisocitrate a new mass would not be visible since the difference between substrate and product, in that case, is the location of a hydroxyl group. This means that another enzyme would be necessary for the second step of the reaction to produce 2-methylisocitrate. Aconitase, encoded by *citB,* in *B. subtilis,* is the likely candidate for this activity.

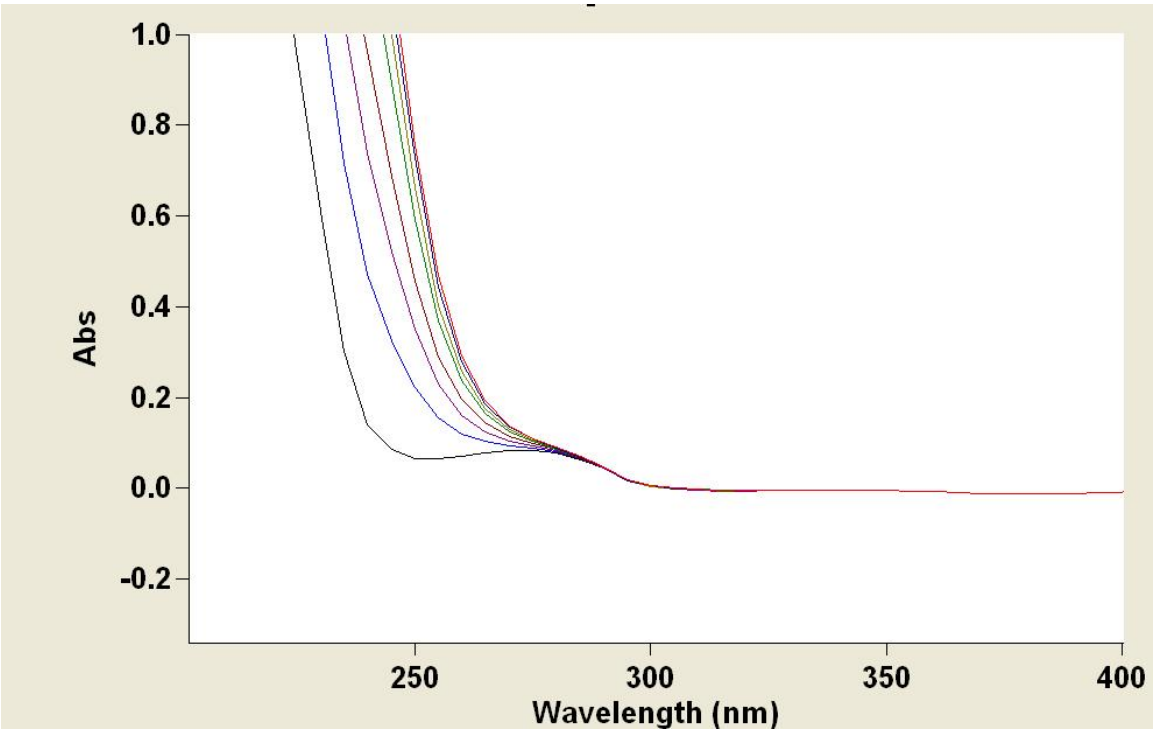


Figure 24. A broad spectrum (200-400nm) scan of the MmgE reaction. Scans taken every ten minutes.

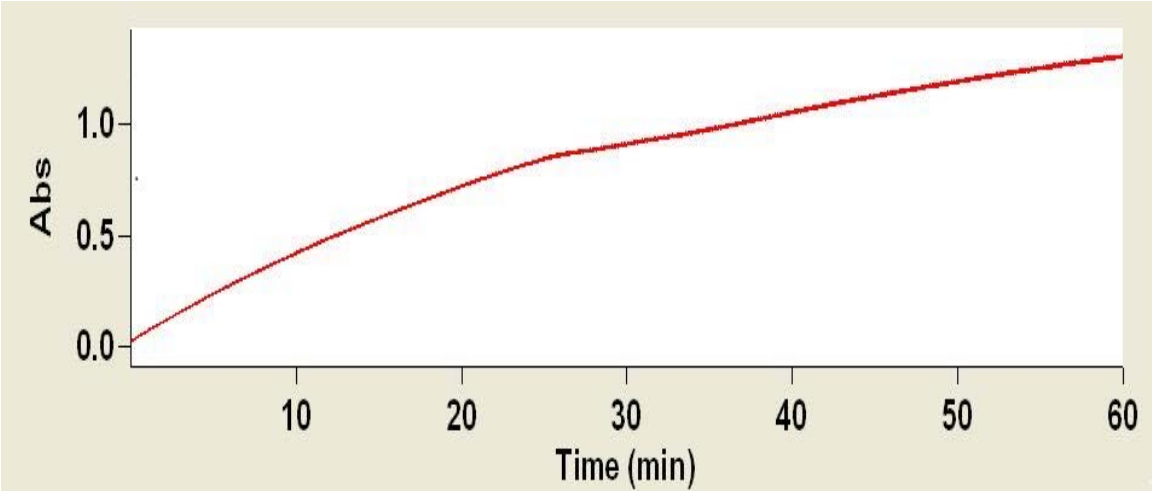


Figure 25. UV/Vis time-course of the mmgE enzymatic reaction at 240nm.

To complement the time-course experiments done with HPLC UV/Vis spectrometry was used to give another way to visualize the reaction. The time-course, at 240nm, in Figure 26 shows good activity over an hour. Previous experiments using the UV/Vis spectrophotometer may have had lower levels of activity due to MmgD only generating low levels of 2-methylcitrate. Also, there may have been interference from HS-CoA being released into solution from the condensation of propionyl-CoA and oxaloacetate to 2-methylcitrate. Coenzyme-A absorbs at this wavelength, which could interfere. This assay confirms that the shift in peaks seen on the HPLC was in fact enzymatic activity due to the MmgE protein interacting with its substrate, 2-methylcitrate. Since the product of MmgE was shown to likely by either the cis or trans isomer of methyl-aconitate the results are consistent with the dehydrating role of MmgE. This removal of water forms the methyl-aconitate but rehydrating activity would have to be present to go one more step to 2-methylisocitrate. There must be another enzyme that performs this rehydrating step. The aconitase enzyme encoded by *citB* is a good

candidate for this activity as aconitase serves the same role in both *E. coli* and *Salmonella*.

III. H CitB

The goal of the citB project was to check for methylcitrate cycle activity of *B. subtilis* aconitase. There were many difficulties with this project with the biggest being protein isolation and purification. It was possible to get protein from the nickel affinity chromatography but it was always in very low concentration, as it was difficult to get a good band on a SDS-PAGE. Partially because of the low concentration, getting activity was not possible. The attempt to remedy this was to purify protein out of several pellets of bacteria and then concentrate them using vivaspin columns, which is shown in Figure 26. This did not provide an increase in activity as shown in Figure 31.

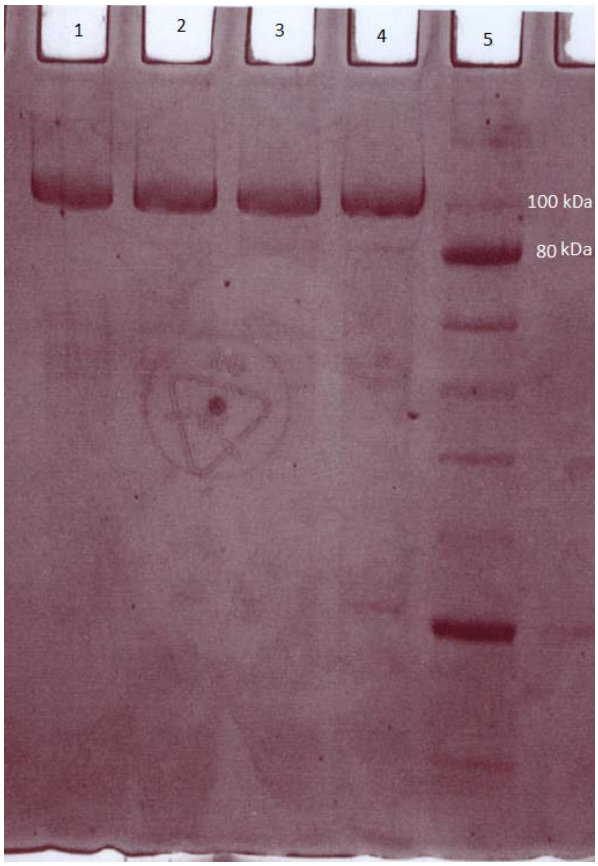


Figure 26. SDS-PAGE of CitB Protein. Lane 5 has Protein ladder and lanes 1-4 have highly concentrated citB protein samples.

The next attempt to increase activity came with using activation buffers, described by the Sonenshein group and a Sigma Aldrich protocol (30) (29). They all had reducing conditions and iron in solution; because the aconitase enzyme has an iron-sulfur cluster that is oxygen sensitive. What this means is that simply by exposing the protein to the atmosphere (which is done upon lysing) oxygen will interact with the cluster disrupting activity. Reducing the cluster and reconstituting with fresh iron can restore activity (29) (30). This was shown through use of an activation buffer on commercial aconitase that had great activity. This was tried on the purified citB protein

with no improvement. This could have been due to the low concentration problems or that the cluster was in some way different causing the activation buffer to be ineffective.

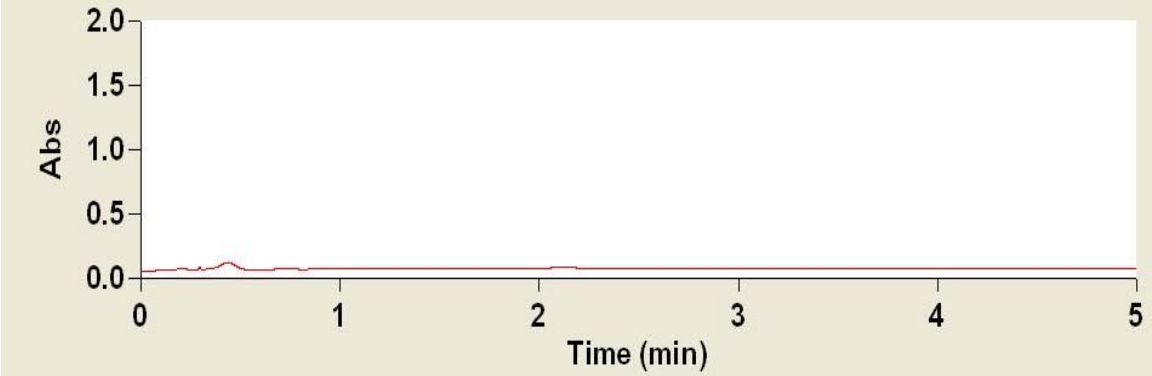


Figure 27. Aconitase assay with activated CitB.

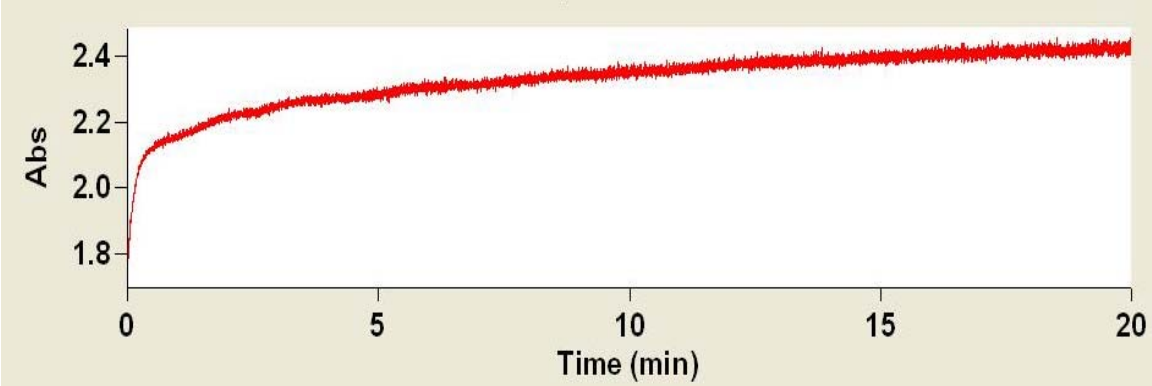


Figure 28. Activity assay of activated CitB and isocitrate. The initial jump in absorbance was from the addition of protein.

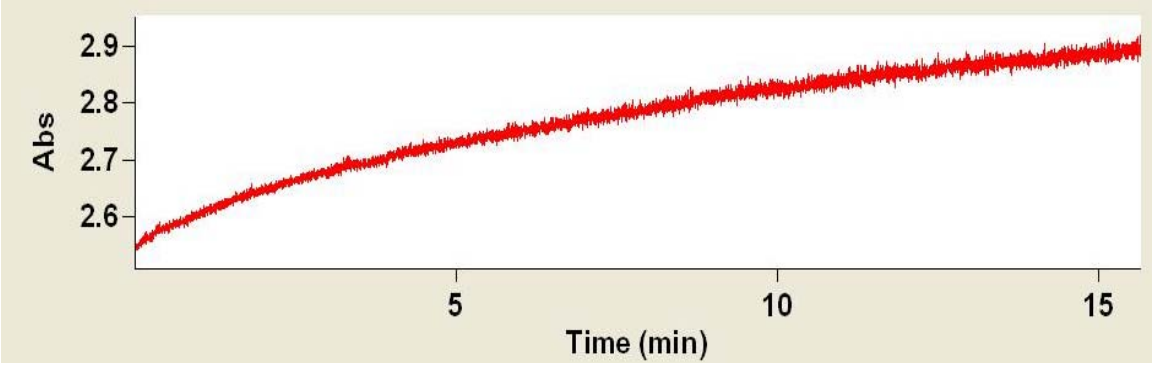


Figure 29. Control experiment without CitB

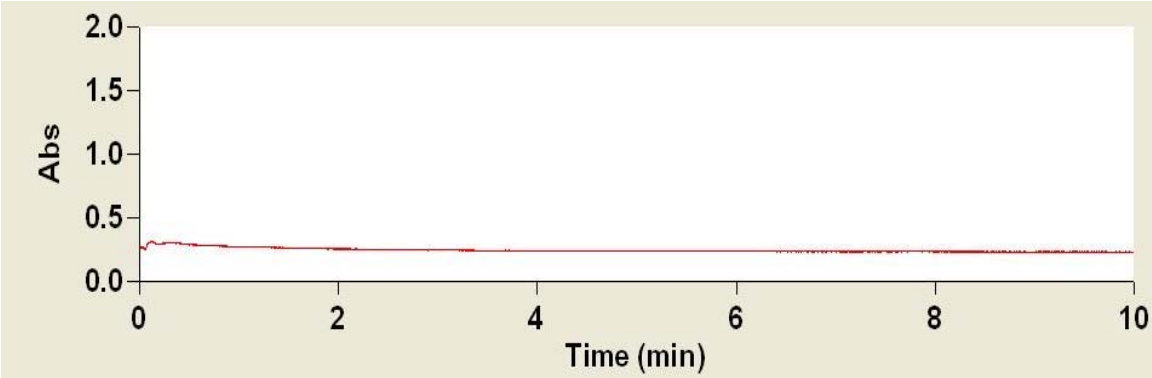


Figure 30. Activity assay with unactivated CitB.

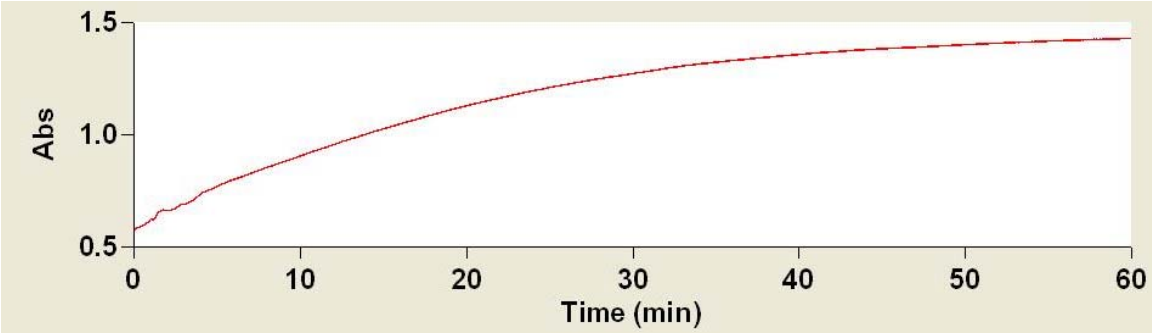


Figure 31. Activity assay of aconitase purchased from Sigma Aldrich.

In figures 28-31 various activity assays of citB are shown. In figures 27 and 30 unactivated enzyme was used with isocitrate and monitored at 240nm. Neither of these assays showed any activity as there was no slope to the line. Figure 28 shows a plot that at first glance looks like great activity, especially with the initial nearly vertical jump in absorbance. This initial increase in absorbance was from the addition of CitB to the reaction mixture. Any increase in absorbance beyond that was likely due to the rest of the reaction mixture as Figure 30 shows that even without any protein there was a steady increase in absorbance, which likely means there was a side reaction in the solution. This would likely be due to the activation buffer that was used in this experiment, and the heavily reducing conditions it provides. It was not possible to find activity with these

assays due to great inconsistency with the results. One trial would look like Figure 29, with a steady increase in absorbance, then the next (which was exactly the same) would look like Figure 28 with no change in absorbance. Either that and/or the control experiments would show that the increase in absorbance was due to something other than enzymatic activity. Figure 31 shows an activity assay using aconitase purchased from Sigma Alrdrich, and the activation buffer from Sigma was used. Aconitase should produce repeatable results like in Figure 31, but it did not.

III. I Conclusion

In the course of this work the *mmgE* gene was successfully cloned and overexressed in *E.coli,* which was confirmed by dye-terminating sequencing. The protein encoded by the *mmgE* gene was also successfully isolated and purified using Ni-NTA chromatography, which was confirmed by both SDS-PAGE and enzyme activity assays. The activity seen shows that this protein is 2-methylcitrate dehydratase. This was shown by the activity in the HPLC which showed that its substrate is in fact 2-methylcitrate and that its product is a methyl-aconitate. LC-MS was utilized to confirm the product's identity as methyl-cis/trans-aconitate. This means that the enzyme has dehydratase but not hydratase activity needed to take the pathway to methylisocitrate.

REFERENCES

1. *Molecular Genetics of Sporulation in Bacillus subtilis.* **Straiger, Patrick and Losick, Richard.** 1996, Annu. Rev. Genetics, pp. 297-341.

2. *Krebs Cycle Function is Required for Activation of the Spo0A Transcription Factor in Bacillus subtilis.* **Ireton, Keith, et al., et al.** 1995, Proc. Natl. Acad. Sci. USA, pp. 2845-2849.

3. *Genetic Networks Controlling the Initiation of Sporulation and the Development of Genetic Competence in Bacillus subtilis.* **Grossman, Alan D.** 1995, Annu. Rev. Genetics, pp. 477-508.

4. *Regulation of Endospore Formation in Bacillus subtilis.* **Errington, Jeff.** 2003, Nature Reviews Microiology, pp. 117-126.

5. *A σE-Dependant Operon Subject To Catabolite Repression During Sporulation in Bacillus subtilis.* **Bryan, Edward M., Beall, Bernard W. and Morgan Jr., Charles P.** 1996, Journal of Bacteriology, pp. 4778-4786.

6. *Sporulation of Bacillus subtilis.* **Piggot, Patrick J. and W., Hilbert David.** 2004, Current Opinion in Microbioligy, pp. 579-586.

7. *The Complete Genome Sequence of the Gram-Positive Bacterium Bacillus subtilis.* **Kunst, F., et al., et al.** 1997, Nature, pp. 249-256.

8. *The mmgA Gene from Bacillus subtilis Encodes a Degradative Acetoacetyl-CoA Thiolase.* **Reddick, Jason J. and Williams, Jayme K.** 2008, Biotechnol Lett, pp. 1045-1050.

9. *Abstracts, 63rd Southeast Regional Meeting of the American Chemical Society.* **Vegunta, Yogesh and Reddick, Jason J.** Richmond : s.n., 2001. SERM-487.

10. *Abstracts of Papers, 238th ACS National Meeting.* **Smith, Jeffrey and Reddick, Jason J.** Washington DC : s.n., 2009. BIOL-175.

11. SubtiList World-Wide Web Server. [Online] Institut Pasteur, 2004. [Cited: June 7, 2012.] http://genolist.pasteur.fr/SubtiList/.

12. *Abstracts of Papers, 238th ACS National Meeting.* **Acharya, Rejwi and Reddick, Jason J.** Washington DC : s.n., August 16-20 2009. Biol-158.

13. *Abstracts, Joint 66th Southwest and 62nd Southeast Regional Meeting of the American Chemical Society.* **Booth, William T., et al., et al.** New Orleans : s.n., 2010. SESW-57.

14. *Genome-Wide mRNA Profiling in Glucose Starved Bacillus Subtilis Cells.* **Koburger, Torsten, et al., et al.** 2005, Mol Gen Genomics, pp. 1-12.

15. *On the Mechanism of Action of the Antifungal Agent Propionate, Propionyl-CoA Inhibits Glucose etabolis in Apsergillus nidulans.* **Buckel, Wolfgang and Brock, Mathias.** 3227-3241, s.l. : Eur. J. Biochem, 2004, Vol. 271.

16. *Oxidation of Propionate to Pyruvate in Escherichia coli.* **Brock, Matthias, et al., et al.** 2002, Eur. J. Biochem., pp. 6184-6194.

17. *Role of the Methylcitrate Cycle in Propionate Metabolism and Detoxification in Mycobacterium smegmatis.* **Upton, AM and JD, McKinney.** 3973-82, s.l. : Microbiology, 2007, Vol. 153.

18. *Propionate Oxidation in Escherichia coli Evidence for Operation of a Methycitrate Cycle in Bacteria.* **Textor, Susanne, et al., et al.** 1997, Arch Microbiol, pp. 428-436.

19. *Salmonella typhimurium LT2 Catabolizes Propionate via the 2-Methylcitric Acid Cycle.* **Horswill, Alexander R. and Escalante-Semerena, Jorge C.** 1999, Journal of Bacteriology, pp. 5615-5623.

20. *Propionate Catabolism in Salmonella typhimurium LT2: Two Divergently Transcribed Units Comprise the prp Locus at 8.5 Centisomes, prpR Encodes a Member of the Sigma-54 Family of Activators, and the prpBCDE Genes Constitute and Operon.* **Horswill, Alexander R. and Escalante-Semerena, Jorge C.** 3, s.l. : Journal of Bacteriology, 1997, Vol. 179. 928-940.

21. *In Vitro Conversion of Propionate to Pyruvate by Salmonella enterica Enzymes: 2-Methylcitrate Dehydratase (PrpD) and Aconitase Enzymes Catalyze the Conversion of 2-Methylcitrate to 2-Methylisocitrate.* **Horswill, Alexander R. and Escalante-Semerena, Jorge C.** 2001, Biochemistry, pp. 4703-4713.

22. *Crystal Structure of Salmonella typhimurium 2-Methylcitrate lyase (PrpB) and Its Complex with Pyruvate and Mg2+.* **Simanshu, Dhirendra K. and Satheshkumar, P. S.** 1, s.l. : Ciocheical and Biophysical Research Communications, 2003, Vol. 311. 193-201.

23. *Fatty Acids of the Genus Bacillus: an Example of Branched-Chain Preference.* **Kaneda, Toshi.** s.l. : Bacteriological Reviews, p391-418, 1977, Bacteriological Reviews, pp. 391-418.

24. **NCBI.** PrpD protein [Salmonella enterica subsp. enterica serovar Typhi]. *The Protein database.* [Online] [Cited: June 12, 2012.] http://www.ncbi.nlm.nih.gov/protein/CAD08825.1. GenBank: CAD08825.1.

25. **EMBL-EBI.** ClustalW2 - Multiple Sequence Alignment. [Online] European Bioinformatics Institute , 2012. [Cited: June 12, 2012.] ClustalW2 - Multiple Sequence Alignment. http://www.ebi.ac.uk/Tools/msa/clustalw2/.

26. **NCBI.** 2-methylcitrate dehydratase [Escherichia coli str. K-12 substr. W3110]. *The Protein database.* [Online] [Cited: June 12, 2012.] http://www.ncbi.nlm.nih.gov/protein/YP_488628.1. YP_488628.1.

27. Gene name search with the pattern 'mmge'. *SubtiList Web Server.* [Online] 2004. [Cited: June 12, 2012.] http://genolist.pasteur.fr/SubtiList/genome.cgi.

28. *Overexpression, Purification, and Characterization of MmgD from Bacillus subtilis Strain 168.* **Acharya, Rejwi.** http://libres.uncg.edu/ir/uncg/f/Acharya_uncg_0154M_10227.pdf, s.l. : UNCG Masters Thesis, 2009.

29. *Bacillus subtilis Aconitase Is Required for Efficient Late-Sporulation Gene Expression.* **Serio, Alisa W., Pechter, Kieran B. and Soninshein, Abraham L.** 2006, Journal of Bacteriology, pp. 6396-6405.

30. **Sigma Aldrich.** Enzymatic Assay of Aconitase. *Sigma Aldrich.* [Online] [Cited: June 1, 2012.] http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/General_information/2/aconitase.Par.0001.File.tmp/aconitase.pdf.

31. *Studies of Propionate Toxicity in Salonella enterica Identify 2-Methylcitrate as a Potent Inhibitor of Cell Growth.* **Horswill, Alexander R., Dudding, Andrea R. and Escalante-Semerena, Jorge C.** 19094-19101, s.l. : The Journal of Biological Chemistry, 2001, Vol. 267.