

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Discovery of a Novel Mevalonate Pathway and its Potential to Produce Biofuels

**Permalink**

<https://escholarship.org/uc/item/81s1f174>

**Author**

Vinokur, Jeffrey Vinokur

**Publication Date**

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Discovery of a Novel Mevalonate Pathway and its  
Potential to Produce Biofuels

A dissertation submitted in the partial satisfaction of the  
requirements for the degree Doctor of Philosophy  
in Biochemistry, Molecular and Structural Biology

by

Jeffrey Michael Vinokur

2017

© Copyright by  
Jeffrey Michael Vinokur  
2017

## ABSTRACT OF THE DISSERTATION

Discovery of a Novel Mevalonate Pathway and its Potential to Produce Biofuels

by

Jeffrey Michael Vinokur

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2017

Professor James U. Bowie, Chair

The mevalonate pathway is present in eukaryotes, archaea, and some bacteria, where it produces the building blocks used to make cholesterol, vitamin A, natural rubber, and over 25,000 other biomolecules collectively called isoprenoids. Here we report the discovery of a novel mevalonate pathway in archaea. We describe the identification and characterization of three new enzymes, two new metabolites, a novel structure, and a new mechanism. We also show that the new pathway is unique to extreme acidophiles that grow below pH 2. We attempt to re-design three mevalonate pathway enzymes (mevalonate 3-kinase, mevalonate 5-phosphate decarboxylase, and mevalonate 5-pyrophosphate decarboxylase) to produce isoprenol through the decarboxylation of mevalonate. Isoprenol is a 5-carbon alcohol that has a higher energy density than ethanol. Both rational design and random mutagenesis strategies are described. Finally, we report a new method for the soluble expression of recombinant proteins in *E. coli* through the co-expression of archaeal thermosomes and prefoldins.

The dissertation of Jeffrey Michael Vinokur is approved.

Todd O. Yeates

Robert P. Gunsalus

James U. Bowie, Committee Chair

University of California, Los Angeles

2017

*This work is dedicated to my grandparents  
Yefim Susunikashvili and David Vinokur.*

## TABLE OF CONTENTS

Abstract of the Dissertation	ii
Table of Contents	v
List of Figures and Tables	vi
Vita	ix

### **Chapter 1: Evidence of a Novel Mevalonate Pathway.**

1.1	Abstract	1
1.2	Introduction	1
1.3	Results	3
1.4	Discussion	8
1.5	Methods	11
1.6	References	28

### **Chapter 2: Adaptation to Life in Acid Through a Novel Mevalonate Pathway**

2.1	Abstract	31
2.2	Introduction	32
2.3	Results	34
2.4	Discussion	39
2.5	Methods	42
2.6	References	58

### **Chapter 3: Structural Analysis of Mevalonate 3-Kinase**

3.1	Abstract	63
3.2	Introduction	64
3.3	Results & Discussion	66
3.4	Conclusions	70
3.5	Methods	71
3.6	References	88

### **Chapter 4: Mutating Mevalonate Pathway Enzymes to Produce Isoprenol**

4.1	Abstract	93
4.2	Introduction	93
4.3	Results & Discussion	96
4.4	Conclusions	105
4.5	Methods	106
4.6	References	136

### **Chapter 5: A Novel Protein Expression Technology Using Thermosomes**

5.1	Abstract	137
5.2	Introduction	138
5.3	Results & Discussion	140
5.4	Future Directions	143
5.5	Methods	143
5.5	References	161

## LIST OF FIGURES AND TABLES

### Chapter 1: Evidence of a Novel Mevalonate Pathway.

1.1	Mevalonate pathways	16
1.2	Enzyme Specificity	17
1.3	Electrospray ionization mass spectrum after Ta1305 activity	18
1.4	$^{13}\text{C}$ NMR of completed enzymatic reactions	19
1.5	Full $^{13}\text{C}$ NMR spectrum of the no enzyme control from figure 1.3A	20
1.6	Full $^{13}\text{C}$ NMR spectrum of mevalonate-5-kinase activity from figure 1.3B.	21
1.7	Full $^{13}\text{C}$ NMR spectrum of mevalonate-3-kinase activity from figure 1.3C	22
1.8	Electrospray ionization mass spectrum after Ta0762 and Ta1305 activity	23
1.9	Full $^{13}\text{C}$ NMR spectrum of activity of both enzymes in tandem from figure 1.3D	24
1.10	Biochemical characterization of the novel enzymes	25
1.11	Michaelis-Menton plot for mevalonate-3-kinase	26
1.12	Clustal Omega sequence alignment	27

### Chapter 2: Adaptation to Life in Acid Through a Novel Mevalonate Pathway

2.1	Mevalonate pathways	49
2.2	The reaction scheme	50
2.3	Ta0893 possesses key residues for decarboxylation	50
2.4	Purification and identification of MBD	51
2.5	Activity of Ta0893 homologs	52
2.6	Phylogeny and sequence alignment	53
2.7	Purification of <i>P. torridus</i> MBD and substrate specificity	54
2.8	Temperature dependence of <i>P. torridus</i> MBD activity	55
2.9	Optimum pH of <i>P. torridus</i> MBD	55
2.10	Time course of IP production by <i>P. torridus</i> MBD	56
2.11	pH dependence of activity of <i>R. castenholzii</i> MMD	57
T2.1	Table: Proteins identified in gel fragments with MBD activity	57

### Chapter 3: Structural Analysis of Mevalonate 3-Kinase

3.1	Enzymatic Reactions	76
3.2	Overall Structure of M3K	76
3.3	SDS-PAGE gel showing dimer and monomer forms	77
3.4	Comparison of the thermal stability of the monomer and dimer forms	78
3.5	Substrate Binding to M3K	79
3.6	Sequence Alignment of M3K and MDDs	79
3.7	Binding Site Comparison of MDD and M3K	80
3.8	Catalysis of Phosphate Transfer	81
3.9	Michaelis-Menten plot for wildtype mevalonate-3-kinase	82
3.10	Michaelis-Menten plot for a T275A mutant of mevalonate-3-kinase	82



3.11	Michaelis-Menten plot for a L18A mutant of mevalonate-3-kinase.	83
3.12	Michaelis-Menten plot for a S105A mutant of mevalonate-3-kinase	83
3.13	Michaelis-Menten plot for a R185K mutant of mevalonate-3-kinase.	84
3.14	Residues Implicated in Decarboxylation	85
3.15	Decarboxylase homology models	86
T3.1	Kinetic parameters of M3K variants	86
T3.2	X-ray crystallography statistics (molecular replacement).	87

#### Chapter 4: Mutating Mevalonate Pathway Enzymes to Produce Isoprenol

4.1	Mevalonate pathways.	113
4.2	Reaction scheme for mevalonate pathway enzymes	114
4.3	Catalytic residues of mevalonate pyrophosphate decarboxylase	115
4.4	Adding decarboxylation residues to mevalonate 3-kinase	116
4.5	Orientation of T275D and L19K in the M3K double mutant	116
4.6	Adding binding residues to mevalonate pyrophosphate decarboxylase	117
4.7	SDS-PAGE Analysis of MDC mutants	118
4.8	Binding residues of MMD overlaid with M3K	119
4.9	Alternative variants of MMD	120
4.10	Michaelis-Menten plot for the activity of hydroxyethylthiazole kinase	121
4.11	Farnesol kinase and IP Kinase allow <i>E. coli</i> to utilize isoprenol	122
4.12	Schematic of selection plasmid	122
4.13	Growth of <i>E. coli</i> strain containing plasmid pSELECT-03	123
4.14	Directed evolution of selection strain	123
4.15	The final selection strain	124
4.16	Rescue with plasmid pMBI	125
4.17	Selection strain schematic	126
4.18	Titration of IPTG into the selection strain containing wild type decarboxylases	127
4.19	MMD expression at various IPTG concentrations	128
4.20	Selection strategy	129
4.21	Proposed mechanism that requires a 5-phosphate moiety	130
T4.1	Table: Enzyme variants generated to convert mevalonate to isoprenol	131
T4.2	Table: Decarboxylase variants	132

#### Chapter 5: A Novel Protein Co-Expression Technology Using Thermosomes

5.1	Review of the GroEL/ES mechanism	149
5.2	Structural comparison of GroEL/ES and the thermosome	150
5.3	SDS-PAGE gel of truncated Ta0893 co-expressed with GroEL/ES	151
5.4	Map of Takara parent plasmids used in this work	152
5.5	SDS-PAGE analysis of the chaperone expression from the 12 plasmids	153
5.6	Activity of Ta0893 co-expressed with each of the 12 thermosome plasmids	154
5.7	SDS-PAGE of Ta0893 co-expressed with each of the 12 thermosome plasmids	155
5.8	Ta0893 homologs co-expressed with plasmid 10	156
5.9	Differential expression of <i>Acidiplasma</i> MBD with plasmid 10	157
T5.1	Table: List of plasmids constructed for this work	158
T5.2	Table: Test set of 24 eukaryotic proteins	159
T5.3	Table: Test set of 24 archaeal proteins	160

## ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Jim Bowie, who is hereby nominated as the best advisor on the planet Earth. Jim made this research possible by giving me the freedom to pursue an accidental discovery in my rotation, which led to the uncovering of new enzymes, new metabolites, new structures, a new pathway, and a new expression technology. I am grateful for him allowing me to simultaneously pursue this research while undertaking intense science outreach efforts. Any success I have going forward is thanks to his incredible support, guidance, and encouragement.

I would also like to thank my family and friends for being there for me through this roller coaster ride. Oleg, Laza, and Mitch, you guys are the best. I have also been lucky to have incredible co-workers at UCLA that helped advance this research: Tyler Korman and Zheng Cao helped analyze data and conduct NMR experiments for our ACS Biochemistry publication adapted here as chapter 1 (DOI: 10.1021/bi500566q). Duilio Cascio, Michael Sawaya, and Michael Collazo helped with crystallography for our Protein Science publication adapted here as chapter 2 (DOI:10.1002/pro.2607). Matt Cummins conducted experiments and Tyler Korman analyzed data for our Scientific Reports publication adapted here as chapter 3 (DOI:10.1038/srep39737). Chapters 4 and 5 represent ongoing work and involves the researchers already mentioned as well as Mark Arbing. This work was supported by the DOE (DE-FC02-02ER63421), the NIH Chemistry Biology Interface Training Program (NIGMS 5T32GM008496), and the NSF Graduate Fellowship (DGE-1144087). Lastly, none of this would not be possible without the legions of researchers that came before me and it means little without the future researchers that will build upon this work.

## VITA

### DEGREES

- 2008 Bachelors of Science, Biochemistry  
University of Wisconsin – Madison
- 2013 Masters of Science, Biochemistry  
University of California – Los Angeles

### EXPERIENCE

- 2006-2007 Summer Researcher  
Waksman Institute  
Rutgers University
- 2007-2008 Research Assistant  
Biochemistry Lab of John Golbeck  
Pennsylvania State University
- 2008-2010 Undergraduate Researcher  
Great Lakes Bioenergy Research Center  
University of Wisconsin – Madison
- 2010-2012 Undergraduate Researcher  
Chemistry Lab of Sam Gellman  
University of Wisconsin – Madison
- 2012-2017 Graduate Student  
Biochemistry Lab of Jim Bowie  
University of California, Los Angeles

### AWARDS

- 2010 Wiscontreprenuer Scholarship
- 2012 UCLA University Fellowship
- 2012-2014 Dean's Scholar Award
- 2013-2014 UCLA-NIH Chemistry Biology-Interface Trainee
- 2014-2017 NSF Graduate Research Fellowship
- 2017 Audree Fowler Award in Protein Science

## NOTABLE PRESENTATIONS

2010	Wisconsin Institutes for Discovery
2011	Liberty Science Center
2011	St. Louis Science Center
2011	Maryland Science Center
2011	World Science Festival
2012	USA Science & Engineering Festival
2012	Smithsonian Institute
2013	Nebraska Science Festival
2014	Singapore Science Festival
2014	Aspen Science Festival
2016	Caltech
2017	Genentech

## PUBLICATIONS

**Vinokur, JM**, Cummins, MC, Korman, TP & Bowie, JU. (2016) An Adaptation To Life In Acid Through A Novel Mevalonate Pathway. *Scientific Reports* 6, 39737. 10.1038/srep39737

**Vinokur JM**, Korman TP, Sawaya MR, Collazo M, Cascio D, & Bowie JU. (2015). Structural analysis of mevalonate-3-kinase provides insight into the mechanisms of isoprenoid pathway decarboxylases. *Protein Science*, 24(2), 212–20. doi: 10.1002/pro.2607

**Vinokur JM**, Korman TP, Cao Z, Bowie JU. (2014). Evidence of a novel mevalonate pathway in archaea. *Biochemistry*, 53(25), 4161–8. doi:10.1021/bi500566q

Korman, TP, Sahachartsiri B, Li D, **Vinokur JM**, Eisenberg D, Bowie JU. (2014). A synthetic biochemistry system for the in vitro production of isoprene from glycolysis intermediates. *Protein Science*, 23(5), 576–85. doi:10.1002/pro.2436

Haft RJF, Keating DH, Schwaegler T, Schwalbach MS, **Vinokur J**, Tremaine M, Peters JM , et al. (2014). Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. *PNAS*, 111(25), E2576–85. doi:10.1073/pnas.1401853111

## CHAPTER 1

### EVIDENCE OF A NOVEL MEVALONATE PATHWAY

#### 1.1 ABSTRACT

Isoprenoids are a remarkably diverse class of more than 25,000 biomolecules that include familiar compounds such as cholesterol, chlorophyll, vitamin A, ubiquinone, and natural rubber. The two essential building blocks of all isoprenoids, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), are ubiquitous in the three domains of life. In most eukaryotes and archaea, IPP and DMAPP are generated through the mevalonate pathway. We have identified two novel enzymes, mevalonate-3-kinase and mevalonate-3-phosphate-5-kinase from *Thermoplasma acidophilum*, which act sequentially in a putative alternate mevalonate pathway. We propose that mevalonate-3,5-bisphosphate is acted upon by a yet unidentified ATP-independent decarboxylase, yielding isopentenyl phosphate, which is subsequently phosphorylated by the known isopentenyl phosphate kinase from *T. acidophilum* to generate the universal isoprenoid precursor, isopentenyl pyrophosphate (IPP)

#### 1.2 INTRODUCTION

Cholesterol, chlorophylls, hemes, ubiquinones, natural rubbers, and archaeal membrane lipids are just a few examples of more than 25,000 biomolecules that make up the diverse class of organic molecules called isoprenoids (1-3). Isoprenoids are found in all three domains of life and are involved in essential processes such as electron transport, post-translational modification, regulation of membrane fluidity, and cytoskeleton assembly (4). All isoprenoids are composed of two or more isoprene building blocks (5-atom branched hydrocarbons) derived

from isopentenyl pyrophosphate (IPP) and its isomer, dimethylallyl pyrophosphate (DMAPP). Plants and most eubacteria generate IPP and DMAPP from pyruvate and glyceraldehyde-3-phosphate via the deoxy-xylulose-5-phosphate pathway (DXP pathway) (5). Eukaryotes and archaea use a separate pathway called the mevalonate pathway, which relies only on acetyl-CoA as the sole carbon source (6). The mevalonate pathway in archaea is especially important because IPP and DMAPP are used to make branched lipids which are connected to glycerol through ether linkages to form membrane lipids (7). These branched lipids and ether linkages are thought to promote membrane stability at high temperatures (8, 9).

The canonical mevalonate pathway of eukaryotes can be conceptually separated into 2 parts, which we refer to as the upper and lower stages (Fig. 1.1). In the upper stage, three acetyl-CoAs are condensed and reduced by NADPH to yield mevalonate. In the lower stage, mevalonate is sequentially phosphorylated to make mevalonate pyrophosphate, then decarboxylated to yield IPP, which can be converted to DMAPP by an isomerase (10). Recent phylogenetic analyses of archaeal genomes identified strong homologs for the first three enzymes in the pathway leading up to mevalonate. However, in most cases, no homologs for phosphomevalonate kinase or mevalonate pyrophosphate decarboxylase are found (11). Furthermore, some archaea such as *Thermoplasma*, *Flavobacteria*, and *Gramella* have no detectable mevalonate kinase (MVK) homologs (11).

While it has been hypothesized for over a decade that archaea possess an alternative mevalonate pathway based on computational genomics (12), a complete pathway was not elucidated until last year in the organisms *Roseiflexus castenholzii* (13) and *Haloferax volcanii* (14). As shown by the blue arrows in Fig. 1.1, instead of two phosphorylations followed by a

decarboxylation as seen in eukaryotes, some archaea phosphorylate once, decarboxylate with mevalonate-5-phosphate decarboxylase, then phosphorylate again with isopentenyl phosphate kinase. The end product is the same, IPP, but the enzymatic reactions are distinct.

We were interested in obtaining thermophilic enzymes for *in vitro* reconstruction of the mevalonate pathway (15). As part of this effort, we targeted enzymes from an archeon, *Thermoplasma acidophilum*, which is naturally found in coal refuse piles in the northeast United States, where it grows optimally at 59°C (16). One of the target enzymes from *T. acidophilum*, encoded by the *ta1305* gene, was annotated as a mevalonate pyrophosphate decarboxylase in Genbank (we refer to genes in italics (e.g. *ta1305*), and their protein products capitalized without italics (e.g. Ta305)) (17). Upon characterization of the expressed enzyme, however, we discovered that it acts on mevalonate to generate mevalonate-3-phosphate, a previously unknown activity (see Fig. 1.1). We were then able to identify a second novel kinase in *T. acidophilum* encoded by the *ta0762* gene, that phosphorylates mevalonate-3-phosphate to make mevalonate-3,5-bisphosphate, another previously unknown metabolite. We propose that these two new enzyme activities are part of a novel mevalonate pathway utilized by *T. acidophilum* (Fig 1.1).

### 1.3 RESULTS

**Ta1305 is active on mevalonate.** Three genes from *Thermoplasma acidophilum* annotated as mevalonate pyrophosphate decarboxylases in Genbank (17), *ta1305*, *ta0893*, and *ta0461* were cloned and expressed in *E. coli*. Known mevalonate pyrophosphate decarboxylases hydrolyze ATP to ADP as part of their mechanism (21) and can therefore be conveniently studied using standard kinase assays (19). No activity was detected in a kinase

assay using the expected substrate, 1 mM (*RS*)-Mevalonate pyrophosphate. We therefore re-tested the three enzymes with mevalonate and mevalonate-5-phosphate. To our surprise, Ta1305 showed significant ATP consumption in the presence of 1 mM (*RS*)-Mevalonate (Fig. 1.2). No activity was detected with any other substrate or enzyme combination.

**Ta1305 is not a mevalonate-5-kinase.** The only known enzyme to act on mevalonate is mevalonate-5-kinase, which phosphorylates the 5-OH group. We therefore tested if Ta1305 was a mevalonate-5-kinase using a sequential assay. Mevalonate and ATP were incubated with Ta1305, followed by addition of phosphomevalonate kinase (PMVK), which consumes ATP in the presence of mevalonate-5-phosphate. No activity was detected upon the addition of PMVK. When we replaced Ta1305 with an authentic mevalonate-5-kinase from *S. cerevisiae*, however, robust activity was observed. Thus, Ta1305 is not a mevalonate-5-kinase.

**Ta1305 shows no decarboxylase activity as suggested by homology.** The fact that Ta1305 shares 18% identity with mevalonate pyrophosphate decarboxylase from *S. cerevisiae* (22), yet acts on mevalonate suggests that it might directly decarboxylate mevalonate. The expected product of mevalonate decarboxylation would be 3-methylbut-3-en-1-ol (isoprenol) instead of isopentenyl pyrophosphate. To test this hypothesis, we allowed the reaction to go to completion from 1.0 mM mevalonate and analyzed the mixture for any isoprenol production by extracting with hexane followed by gas chromatography. No isoprenol was detected, even after allowing the reaction to incubate for 48 hours at 37°C to promote spontaneous decarboxylation. Positive controls made by spiking the enzymatic reaction with authentic isoprenol indicated that we would have detected isoprenol production as low as 0.01 mM.



**Ta1305 is a kinase.** While the Ta1305 enzyme consumed ATP in the presence of mevalonate, the fate of the phosphate remained unclear. To confirm that this enzyme does not decarboxylate mevalonate and release free phosphate, we assayed the reaction for free phosphate after completion of the reaction using 1 mM ATP and 1 mM (*RS*)-Mevalonate (20). We found no detectable free phosphate, suggesting that the substrate is phosphorylated. The same experiment utilizing authentic mevalonate pyrophosphate decarboxylase from *Saccharomyces cerevisiae* as a positive control and its substrate (*RS*)-Mevalonate pyrophosphate yielded 0.57 mM free phosphate as expected for the racemic substrate mixture.

Phosphorylation of mevalonate was further verified by ESI mass spectrometry. High resolution negative ion electrospray ionization mass spectra were collected on the reaction products. We observed a mass of 227.0313 m/z which is within 0.00008 m/z of the mass expected for a phosphorylated mevalonate [ $C_6H_{13}O_7P - H$ ] (Fig 1.3). These results suggest that Ta1305 phosphorylates mevalonate in an ATP dependent manner.

**Ta1305 generates mevalonate-3-phosphate.** The results so far indicate that Ta1305 is a mevalonate kinase, but there are three potential sites of phosphorylation on mevalonate: the 3-OH group, 5-OH group, and the carboxylate. To identify the site of phosphorylation, we used  $^{13}C$  NMR. While mevalonate alone produces 6 single peaks in  $^{13}C$  NMR, coupling of the carbons to a phosphate will generate doublets for any carbon within three bonds of a  $^{31}P$  atom (23). The three possible phosphorylated products of Ta1305 would each generate a unique set of doublets which allows for positive identification of the phosphorylated species.

To prepare samples for natural abundance NMR analysis, we wanted to keep the ATP concentrations low to simplify the NMR spectra. We therefore re-generated ATP *in situ* from PEP using pyruvate kinase. In this manner, ATP is kept at a very low concentration and recycled, leaving only contributions from the much simpler spectrum of PEP. A spectrum of the no enzyme control sample (i.e. 10 mM (*R*)-Mevalonate) showed 6 singlet peaks as expected (Fig 1.4A) and the carbons were assigned with the guidance of computationally predicted spectra from the Human Metabolome Database (24) and ChemNMR Pro 13.0 (25). Another control sample employing yeast mevalonate-5-kinase showed splitting at carbons 4 and 5 as expected for mevalonate-5-phosphate (Fig 1.4B). The product of Ta1305 activity, however, showed splitting at carbons 2, 3, 4, and 6 which is consistent with phosphorylation at the 3-OH position (Fig 1.4C). Thus, we can annotate Ta1305 as mevalonate-3-kinase. (Full <sup>13</sup>C NMR spectra are shown as figures 1.5-1.7).

**Identification of an enzyme that acts on mevalonate-3-phosphate.** To identify the next step in a possible new mevalonate pathway, we screened enzymes that might further phosphorylate the newly identified metabolite, mevalonate-3-phosphate. Seven genes annotated in Genbank as kinases from *T. acidophilum*, were cloned, and the proteins expressed and purified. We chose two putative kinase genes found near *ta1305* in the genome (*ta1304* and *ta1307*), three genes which had homology to the mevalonate kinase family (*ta0344*, *ta0436*, *ta0546*), and an additional two genes which were annotated as small molecule kinases (*ta0762*, *ta0364*), but to our knowledge, had not been characterized.

We screened the putative kinases for their ability to phosphorylate mevalonate-3-phosphate using a sequential kinase assay. We first used Ta1305 to generate mevalonate-3-

phosphate from mevalonate *in situ* (monitored by ADP production). After the reaction had gone to completion, we added the second test enzyme. Of the seven enzymes tested, only Ta0762 showed activity with mevalonate-3-phosphate (Fig 1.2). No activity was detected with mevalonate, mevalonate-5-phosphate, or mevalonate pyrophosphate. Ta0762 had been computationally annotated as 2-phosphoglycerate kinase though no activity was detected with 2-phosphoglycerate (Fig 1.2). This result suggests Ta0762 is miss-annotated and acts in a pathway directly after mevalonate-3-kinase. To determine if Ta0762 is a kinase, we conducted a free phosphate assay with 2.0 mM ATP and 1.0 mM mevalonate in a sequential enzymatic reaction. After monitoring to completion on the microplate reader, we detected only 0.02 mM free phosphate, suggesting mevalonate-3-phosphate is also phosphorylated. To confirm the production of a doubly phosphorylated product, we again employed ESI mass spectrometry. The reaction product had a mass of 307.0016  $m/z$  which is within 0.0031  $m/z$  of the mass expected for mevalonate diphosphate  $[C_6H_{14}O_{10}P_2 - H]^-$  (Fig 1.8).

**Ta0762 generates mevalonate-3,5-bisphosphate.**  $^{13}C$  NMR was employed to determine the position of phosphorylation by Ta0762 using the same procedure used to identify mevalonate-3-phosphate. Splitting was observed at 5 carbons (2, 3, 4, 5, and 6), which is consistent with phosphates attached to the 3-OH and the 5-OH positions simultaneously (Fig 3D, full spectrum in Fig 1.9). In addition, carbon 4 showed a doublet of doublets pattern consistent with contributions from  $^{31}P$  atoms bonded in both the 3-OH and 5-OH positions simultaneously. The observed spectrum is consistent with the structure of mevalonate-3,5-bisphosphate.

### **Biochemical characterization of mevalonate-3-kinase and mevalonate-3-**

**phosphate-5-kinase.** To characterize the kinetic parameters of the new enzymes, we first determined the optimal pH and temperature for each enzyme. As shown in Fig. 1.10, mevalonate-3-kinase performs optimally at pH 8.5 and 55°C. Mevalonate-3-phosphate-5-kinase performs optimally at 60°C, but does not demonstrate significant pH dependence in the range of pH 6.5 – 9.0. Both enzymes are quite stable, retaining more than 95% of their activity after incubation at 60°C for 1 hour. Under optimal conditions with (*R*)-Mevalonate as a substrate, mevalonate-3-kinase was found to have a  $K_m$  of  $97 \pm 6 \mu\text{M}$  and  $k_{\text{cat}}$  of  $5.0 \pm 0.1 \text{ s}^{-1}$  (Fig 1.11) which is comparable to mevalonate-5-kinase from the archeon *Methanosarcina mazei* ( $K_m$  of  $68 \pm 4 \mu\text{M}$ ,  $k_{\text{cat}}$  of  $4.3 \pm 0.2 \text{ s}^{-1}$ ) (26). Mevalonate-3-phosphate-5-kinase is significantly faster with a  $k_{\text{cat}}$  of approximately  $9.0 \text{ s}^{-1}$ . We were unable to accurately determine the  $K_m$  for mevalonate-3-phosphate-5-kinase, however, since the substrate is not commercially available and the detection limit of our assay is 30  $\mu\text{M}$  substrate. Nevertheless, the enzyme was still at  $V_{\text{max}}$  when assayed with 60  $\mu\text{M}$  mevalonate-3-phosphate indicating that the  $K_m$  is well below 60  $\mu\text{M}$ . For both enzymes, ATP was confirmed to be at saturation since for both enzymes, increasing ATP concentration from 1 mM to 5 mM saw no change in the observed rates.

## **1.4 DISCUSSION**

We propose the two enzymes described here constitute part of a novel mevalonate pathway in *T. acidophilum* that splits at mevalonate and rejoins the known archaeal mevalonate pathway at isopentenyl phosphate (Fig 1.1, red arrows). In *T. acidophilum* there is strong evidence for the presence of acetoacetyl-CoA thiolase, HMG-CoA synthase, HMG-CoA reductase, isopentenyl phosphate kinase, and isopentenyl pyrophosphate isomerase based on homology (10) and experiment (27, 28). However, no homologs have been identified for mevalonate-5-kinase or phosphomevalonate kinase in this organism. Most surprising is the

absence of mevalonate-5-kinase because it is the only enzyme previously known to act on mevalonate. We propose that *T. acidophilum* contains a novel pathway where mevalonate-3-kinase phosphorylates mevalonate at the 3-OH position, followed by mevalonate-3-phosphate-5-kinase, which phosphorylates the 5-OH position, and finally an unidentified decarboxylase converts mevalonate-3,5-bisphosphate to isopentenyl phosphate (Figure 1.1, red arrows).

The proposed decarboxylase enzyme would do the same chemical transformation as mevalonate pyrophosphate decarboxylase but in an ATP independent manner since the substrate is already phosphorylated in the correct position for decarboxylation. We tested apparent mevalonate pyrophosphate decarboxylase homologs Ta0461 and Ta0893 for activity on mevalonate-3,5-bisphosphate, but none was detected in our hands. Both enzymes expressed to inclusion bodies and after many attempts we were unable to refold these proteins in high yield, so it remains possible that one of them is the missing decarboxylase. We also considered the possibility of alternative pathway in which IP Kinase acts on mevalonate 3,5 bisphosphate directly. When we incubated mevalonate-3,5-bisphosphate with IP Kinase, however, we saw no release of phosphate (decarboxylation) or ATP consumption (kinase activity).

Observation of stable tertiary phosphorylated mevalonate is striking because the mechanism of mevalonate pyrophosphate decarboxylase is thought to proceed through a transient tertiary phosphorylated intermediate (30, 31). In a previous mechanistic study of mevalonate pyrophosphate decarboxylase, a heavy oxygen atom was incorporated into mevalonate pyrophosphate at the 3-OH position, then after ATP dependent conversion to IPP, the heavy oxygen was detected in free phosphate (30). This suggests a simple mechanism in

which the  $\gamma$ -phosphate from ATP is transferred to the 3-OH position of mevalonate pyrophosphate, activating the substrate for decarboxylation. Current literature predicts this tertiary phosphorylated molecule is inherently unstable and falls apart with concomitant decarboxylation (32, 33). Our observation of mevalonate-3-phosphate and mevalonate-3,5-bisphosphate as stable metabolites at 55°C suggests that the decarboxylation step requires enzyme catalysis, however.

To shed light on the reason why mevalonate-3-kinase does not function as a decarboxylase, the active site residues of five bacterial mevalonate pyrophosphate decarboxylases (MDCs) were aligned with mevalonate-3-kinase from *T. acidophilum*. We focused on 9 active site residues shown to interact with mevalonate pyrophosphate in a MDC crystal structure by Barta *et al.* (33). Only 4 out of 9 highly conserved residues are preserved in mevalonate-3-kinase (Figure 1.12). It is likely that these five non-conserved residues (L19, L20, I23, D190, T276) contribute to the loss of decarboxylase activity.

Due to the apparently low reliability of predicting function through sequence homology, we cannot tell which organisms have a true copy of mevalonate-3-kinase, and which have the classical mevalonate pyrophosphate decarboxylase. Sequence homology obtained through a protein BLAST search (34) revealed five homologs of mevalonate-3-kinase with >30% identity. All five are annotated as “mevalonate pyrophosphate decarboxylase” or “hypothetical protein” and belong to the order Thermoplasmatales: *Thermoplasma volcanium* (67%), *Thermoplasmatales archeon* (54%), *Ferroplasma sp. Type II* (45%), *Ferroplasma acidarmanus* (42%), and *Picrophilus torridus* (39%). For comparison, the sequence identity of mevalonate-3-kinase to classical yeast MDC is 19%.

Taken together, we have identified two novel enzymes, mevalonate-3-kinase and mevalonate-3-phosphate-5-kinase, which act sequentially in a putative alternate mevalonate pathway in *Thermoplasma acidophilum*. Full confirmation of this pathway requires identification of a missing mevalonate-3,5-bisphosphate decarboxylase. Our findings raise important questions about the mechanism of mevalonate pyrophosphate decarboxylases because tertiary phosphorylated mevalonate species clearly exist as stable metabolites.

## 1.5 METHODS

**Materials.** Miller LB media (BD Difco) was used for growth of bacterial strains. *E. coli* BL21 Gold (DE3) (Agilent) was used as the host strain for both cloning and expression of recombinant proteins. Plasmid pET28a(+) was purchased from Novagen. HotStart Taq Mastermix (Denville) was used for gene amplification. *Phusion* DNA polymerase (Finnzymes), Taq DNA ligase (MCLab), and T5 Exonuclease (Epicenter) were purchased separately and used to make the assembly master mix (AMM) used for cloning. Ni-NTA resin and miniprep reagents were purchased from Qiagen. Primers were synthesized by ValueGene. All other chemicals were purchased from Sigma Aldrich unless otherwise noted.

**Cloning.** Genes were amplified by the polymerase chain reaction (PCR) from *Thermoplasma acidophilum* genomic DNA (ATCC 25905D). A modified Gibson method was used to assemble all constructs (18). Primers included 15-20 base-pairs complementary to the ends of the target gene, plus 15-20 base-pairs complementary to the NdeI and XhoI insertion sites of pET28a(+), which allowed for the addition of an N-terminal 6xHis tag. 10 ng of pET28a(+) digested with NdeI and XhoI was mixed with 30 ng of PCR product, and 7  $\mu$ L of assembly mix (0.1 M TRIS-

HCl, pH 7.5, 0.2 mM dNTPs, 1 mM NAD<sup>+</sup>, 5% PEG 8K, 10 mM DTT, 10 mM MgCl<sub>2</sub>, 0.00375 U/μL T5 Exonuclease, 0.012 μL Phusion DNA polymerase, 4 U/μL TaqDNA ligase). After incubation at 50°C for 2 hours, 5 μL was then used to transform BL21 Gold (DE3) and transformants were selected by plating on LB-agar containing 50 μg/mL kanamycin.

**Expression and Purification.** All *E. coli* strains were grown at 37 °C in LB-media with 50 μg/mL kanamycin. 1 L of LB-media was inoculated with 5 mL of overnight starter culture. Protein expression was induced during log phase (OD<sub>600</sub> 0.5-0.8) with 0.5 mM IPTG. After 20 hours, cells were pelleted, resuspended in 12 mL of buffer A (50 mM Tris HCl [pH 7.5], 100 mM NaCl), lysed by sonication, and cell debris was removed by centrifugation at 30,000xg for 20 min. The lysate was mixed with 3 mL of a Ni-NTA slurry and incubated at 4°C with gentle mixing. After 1 hour, the lysate mixture was applied to a column and the Ni-NTA beads were washed 3 times with 25 mL buffer A containing 10 mM imidazole. Protein was then eluted with 4 mL buffer A containing 250 mM imidazole. For kinetic characterization, enzymes were further purified on an AKTA FPLC using a Superdex S200 10/300 GL gel filtration column run with buffer A, at 0.5 mL/min.

**Enzyme Assays.** Enzymes were assayed for the ability to hydrolyze ATP to ADP by coupling the reactions to pyruvate kinase (PK) and lactate dehydrogenase (LDH) from rabbit muscle (19). Assays were done in duplicate using a 200 μL volume at 55°C and contained 1 mM ATP, 10 mM KCl, 5 mM MgCl<sub>2</sub>, 5 mM 2-mercaptoethanol, 1 mM PEP, 0.5 mM NADH, 1.4 μg enzyme, 1.0 uL LDH/PK mix (Sigma), and 50 mM bis-tris propane [pH 8.5]. The pH of all buffers was adjusted at 25°C. All reagents were incubated for 10 minutes at the desired temperature and then substrate was added to 1 mM final concentration. Absorbance at 340 nm was recorded



over 20 minutes on a SpectraMax 5M microplate reader. Assays for activity with mevalonate-3-phosphate were done in a sequential manner using 0.5 mM mevalonate, 3 mM PEP, and 0.75 mM NADH, respectively. The second enzyme was added after complete conversion of mevalonate to mevalonate-3-phosphate as monitored on the microplate reader. All assay conditions included a control by adding 1 mM ADP directly to ensure PK and LDH activities were not rate limiting.

**Free Phosphate Assays.** Phosphate assays were carried out in a 96 well plate using reagents from a glycosyltransferase activity kit (R&D Systems) (20). 2  $\mu$ L of an enzyme reaction was combined with 48  $\mu$ L ddH<sub>2</sub>O, followed by 30  $\mu$ L malachite green reagent A, 100  $\mu$ L ddH<sub>2</sub>O, and 30  $\mu$ L malachite green reagent B. After a 20 minute incubation, the absorbance at 620 nm was measured on a Spectrmax M5 microplate reader alongside five standards of 0.0 to 1.0 mM KH<sub>2</sub>PO<sub>4</sub> prepared identically. Mevalonate pyrophosphate decarboxylase from *Saccharomyces cerevisiae*, prepared as described previously (15), served as a positive control for release of free phosphate through the decarboxylation of (*RS*)-Mevalonate pyrophosphate.

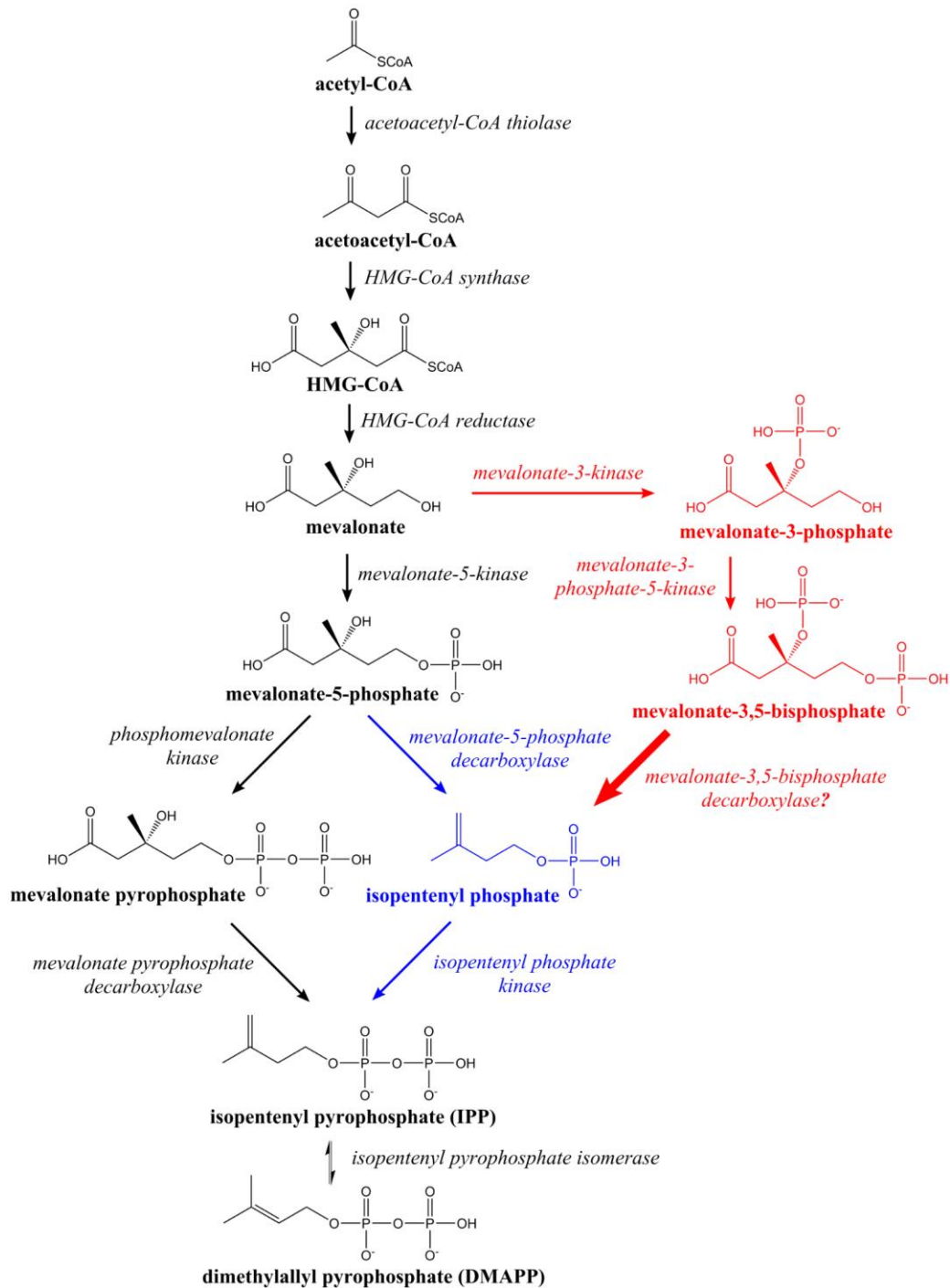
**Gas Chromatography.** Isoprenol was extracted from 200  $\mu$ L of an enzymatic reaction with 600  $\mu$ L hexanes. 2  $\mu$ L of the hexane layer was injected into the GC-FID (HP5890II) equipped with an HP-INNOWAX column (0.320 mm x 30 m, Agilent). The carrier gas was helium with a flow rate of 5 mL/min. The oven temperature was kept at 50°C for 2 min, then raised to 100°C at 10°C/min, then ramped to 250°C at 25°C/min, and finally held at 250°C for 2 min. The inlet and detector temperatures were kept at 250 and 330°C respectively. Isoprenol concentration was determined by comparison to authentic standards.

**Product Identification by NMR.** Enzymatic reactions for NMR were done in a 250  $\mu$ L volume using 1.4  $\mu$ g of enzyme, 20 mM (*R*)-Mevalonate, 20 mM PEP, 1 mM ATP, 1 mM  $MgCl_2$ , and 10  $\mu$ L LDH/PK enzyme mix (Sigma) which was diluted and re-concentrated to remove glycerol. The pH was adjusted to 7.5 at 25°C using 1.0 M KOH with the 20 mM PEP serving as a phosphate buffer in addition to being a substrate for pyruvate kinase. The reactions were incubated for 6 hours at 42°C. The completed 250  $\mu$ L reaction was diluted to 500  $\mu$ L with 99.9%  $D_2O$  from Cambridge Isotope Laboratories and aliquoted into an NMR tube. All spectra were acquired at ambient temperature on a 500 MHz Bruker AV500 spectrometer equipped with a cryoprobe. Data was processed using Topspin 3.1 software. For the spectra of mevalonate-3,5-bisphosphate, PEP concentration in the enzymatic reaction was doubled to 40 mM to allow for full conversion of 20 mM (*R*)-Mevalonate to mevalonate-3,5-bisphosphate.

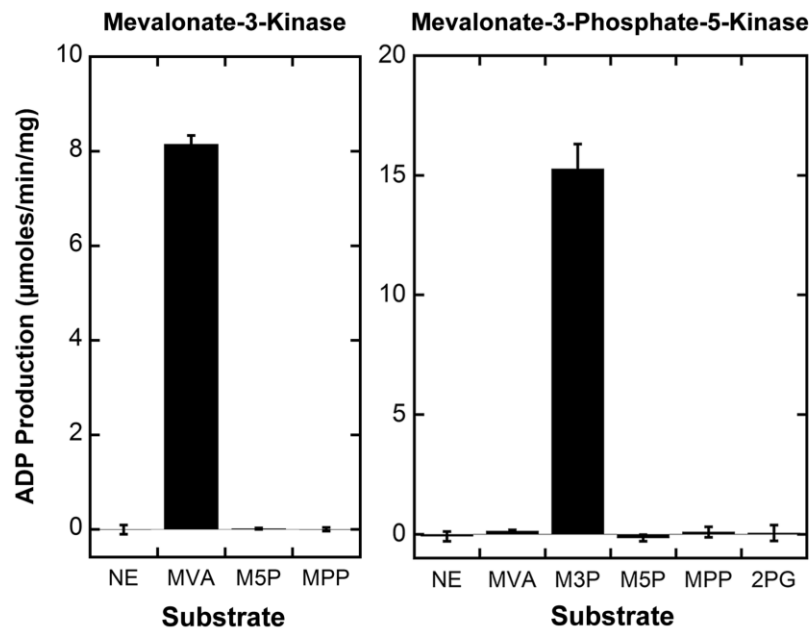
**Product Identification by ESI Mass Spectrometry.** Negative ion electrospray mass spectrometry data was collected with a Waters LCT Premier XE time of flight instrument controlled by MassLynx 4.1 software. Samples from the NMR tubes were transferred to GC vials and injected into the multi-mode ionization source with a Waters Acquity UPLC. The flow injection solvent was 50/50 MeOH/MeCN (LCMS Grade, VWR Scientific) and water blanks were run between all samples. The lock mass standard for accurate mass determination was leucine enkephalin (Sigma L9133).

**Biochemical characterization.** The optimal pH for the enzymes was determined by kinase assays in 0.5 pH unit increments ranging from pH 6.5 to 9.5 using 50 mM bis-tris propane. The optimal temperature of the enzymatic assay was determined in 5°C increments ranging from 25°C to 60°C. We were unable to obtain direct kinetic measurements above 60°C since it is the

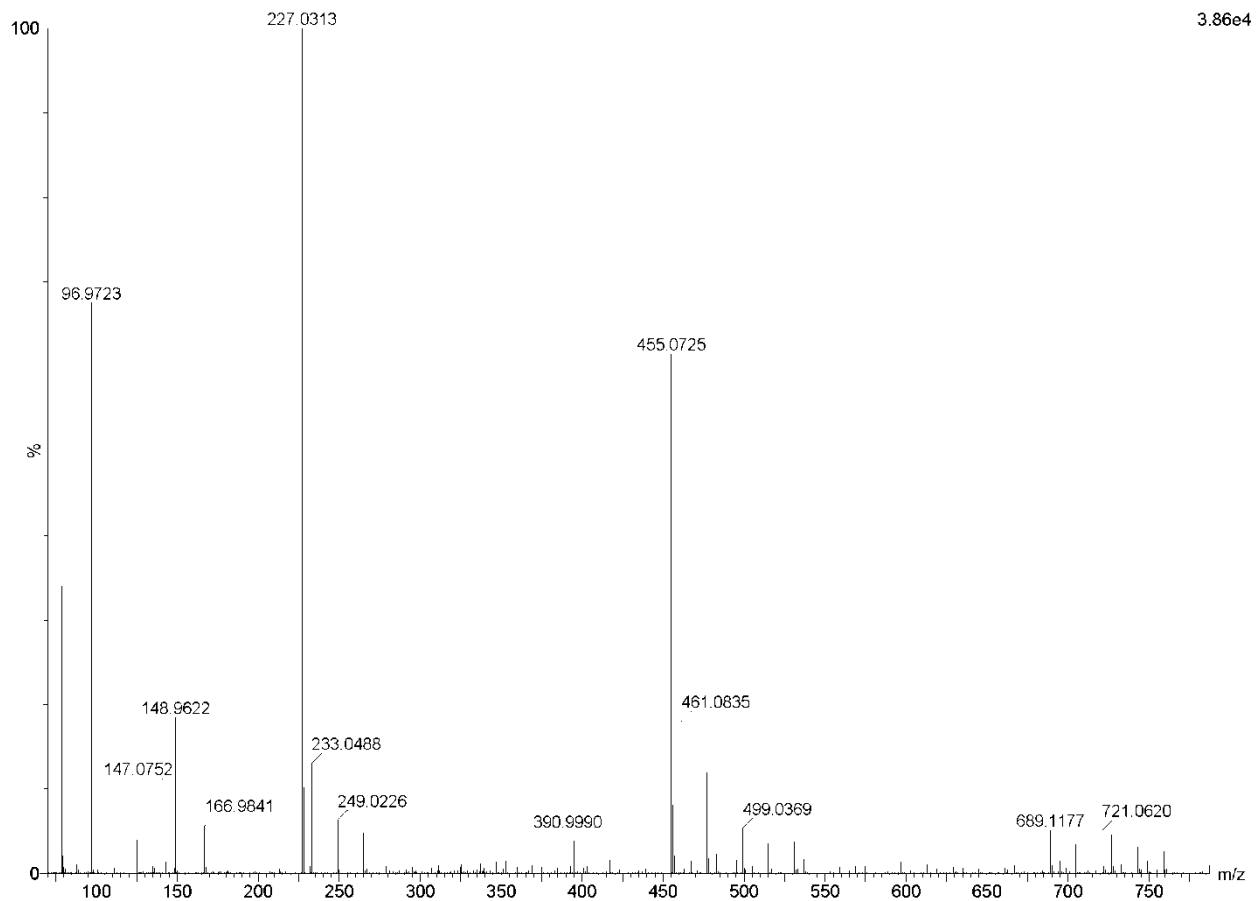
maximum temperature of the SpectraMax M5 microplate reader. To complement this data, temperature stability was assayed by incubating 30  $\mu$ L of 0.170 mg/mL Ta1305 or 0.067 mg/mL Ta0762 for 1 hour at 30°C to 90°C in 4-6°C increments using a thermocycler gradient (Eppendorf ProS PCR machine). Kinetic measurements for Ta1305 were carried out at pH 8.5 and 55°C with 1.4  $\mu$ g enzyme over a range of 0.03 to 4.00 mM (*R*)-Mevalonate. Ta0762 kinetic measurements were performed at pH 8.0 and 60°C with 0.5  $\mu$ g enzyme over the range of 0.03 to 1.00 mM mevalonate-3-phosphate (produced enzymatically).  $k_{cat}$  values were not corrected for temperature since the temperature of each experiment is indicated.



**FIGURE 1-1.** Mevalonate pathways. Eukaryotes use the left side of the fork after mevalonate-5-phosphate (black), and archaea use the right side (blue). The two enzymes identified in this study suggest an alternative pathway branching from mevalonate (red). A predicted decarboxylase (bold arrow) connects mevalonate-3,5-bisphosphate to isopentenyl phosphate.

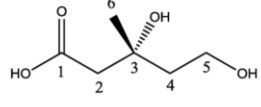
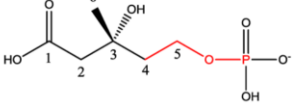
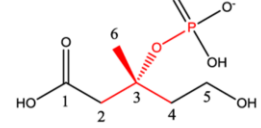
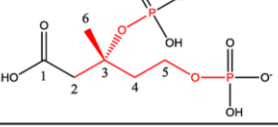


**FIGURE 1-2.** Enzyme specificity. The enzymes were incubated with 1.0 mM of the following substrates: mevalonate (MVA), mevalonate-5-phosphate (M5P), mevalonate-5-pyrophosphate (MPP), mevalonate-3-phosphate (M3P), and 2-phosphoglycerate (2PG). A no enzyme control (NE) was also performed. 1.0 mM ATP was included and ADP production was monitored at 55°C, pH 8.5 for Ta1305 and 60°C, pH 8.0 for Ta0762.

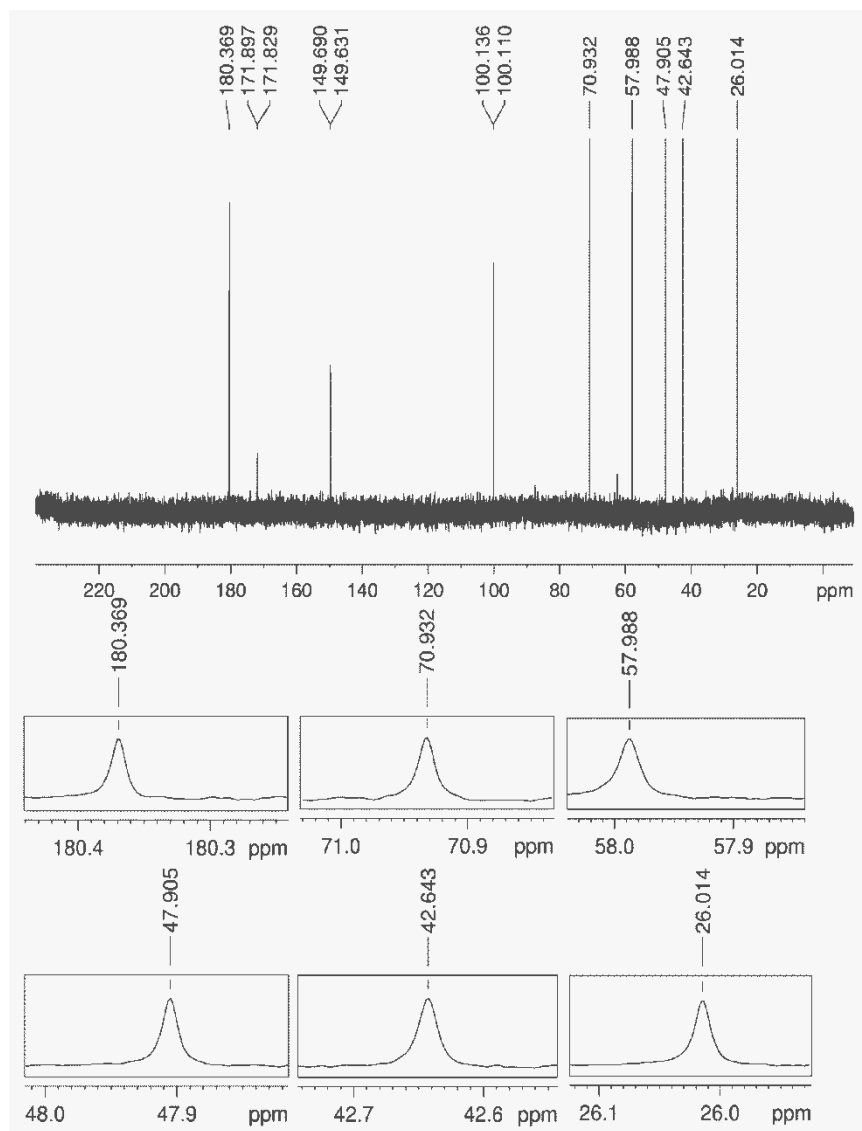


**Figure 1-3.** *Electrospray ionization mass spectrum after Ta1305 activity on (R)-Mevalonate.*

Spectra were collected with a Waters LCT Premier XE time of flight instrument. A sample directly from the NMR experiment was transferred to a GC vial and injected into the multi-mode ionization source with a Waters Acquity UPLC. We observed a mass of 227.0313 m/z which is within 0.00008 m/z of the mass expected for a phosphorylated mevalonate [ $C_6H_{13}O_7P - H$ ].

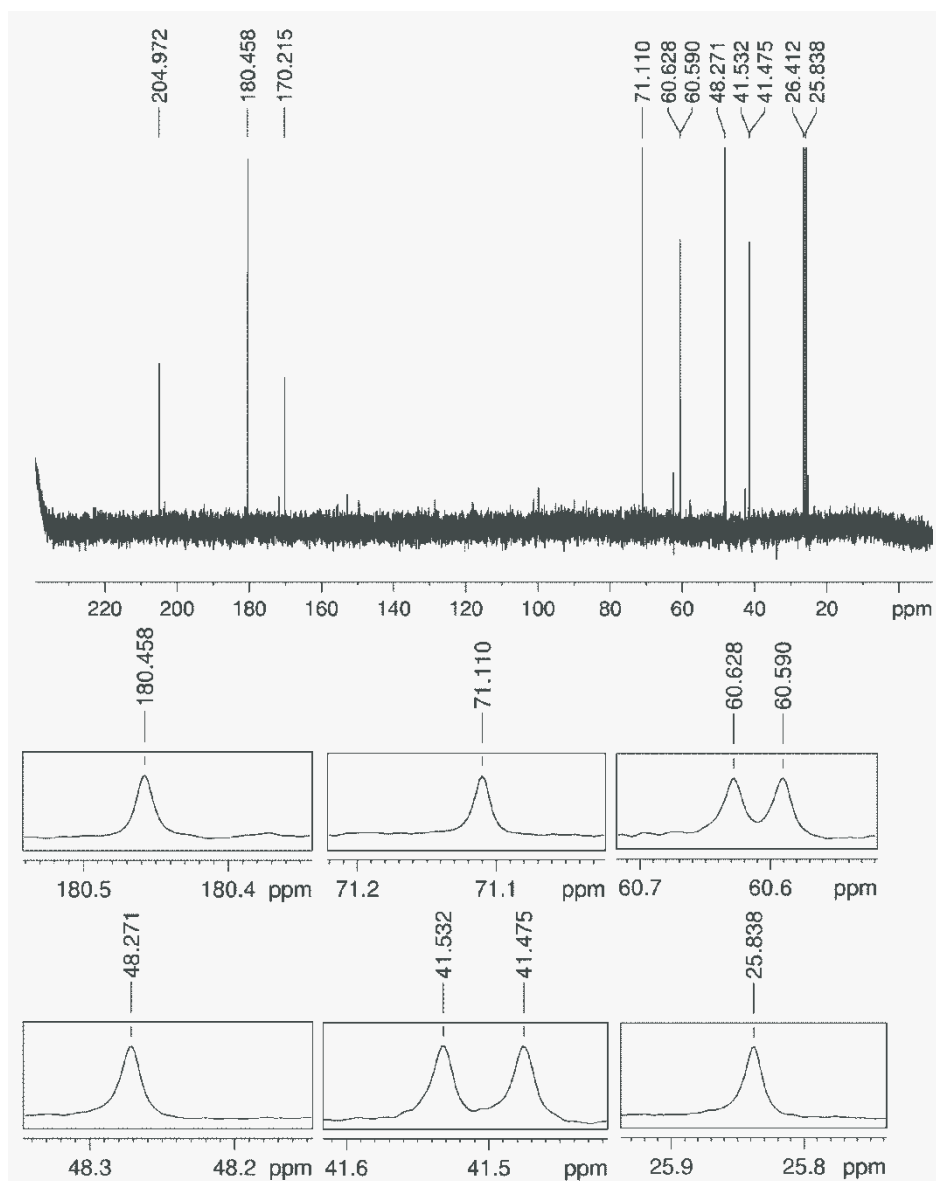
<sup>13</sup> C NMR Chemical Shifts ( $\delta$ ) and Coupling Constants (J)			
mevalonate (no enzyme control)	Carbon	$\delta$ Shift	$J_{C-P}$ (Hz)
<b>A.</b> 	C <sub>1</sub>	180.37	0
	C <sub>2</sub>	47.91	0
	C <sub>3</sub>	70.93	0
	C <sub>4</sub>	42.64	0
	C <sub>5</sub>	57.99	0
	C <sub>6</sub>	26.01	0
mevalonate-5-phosphate	Carbon	$\delta$ Shift	$J_{C-P}$ (Hz)
<b>B.</b> 	C <sub>1</sub>	180.46	0
	C <sub>2</sub>	48.27	0
	C <sub>3</sub>	71.11	0
	C <sub>4</sub>	41.50	7.1
	C <sub>5</sub>	60.61	4.8
	C <sub>6</sub>	25.8	0
mevalonate-3-phosphate	Carbon	$\delta$ Shift	$J_{C-P}$ (Hz)
<b>C.</b> 	C <sub>1</sub>	180.20	0
	C <sub>2</sub>	49.89	3.4
	C <sub>3</sub>	76.72	8.1
	C <sub>4</sub>	41.67	4.1
	C <sub>5</sub>	58.36	0
	C <sub>6</sub>	25.50	2.8
mevalonate-3,5-bisphosphate	Carbon	$\delta$ Shift	$J_{C-P}$ (Hz)
<b>D.</b> 	C <sub>1</sub>	180.25	0
	C <sub>2</sub>	49.74	3.8
	C <sub>3</sub>	76.73	7.1
	C <sub>4</sub>	41.26	7.1, 4.4 <sup>a</sup>
	C <sub>5</sub>	60.99	4.6
	C <sub>6</sub>	25.19	2.4

**FIGURE 1-4.** <sup>13</sup>C NMR of completed enzymatic reactions (10 mM products) in 50% D<sub>2</sub>O. The chemical shifts of the 6 mevalonate carbons are listed along with the J coupling from <sup>31</sup>P (drawn in red). (A) No enzyme (B) yeast mevalonate-5-kinase (C) Ta1305 activity (D) Ta1305 and Ta0762 activity. Reactions were monitored to >96% completion before acquiring spectra. A doublet of doublets was seen at  $J_{C_4-P_1} = 7.1$  and  $J_{C_4-P_2} = 4.4$ . All other splittings are simple doublets. Full spectra are shown in Fig S2-S4, S6.

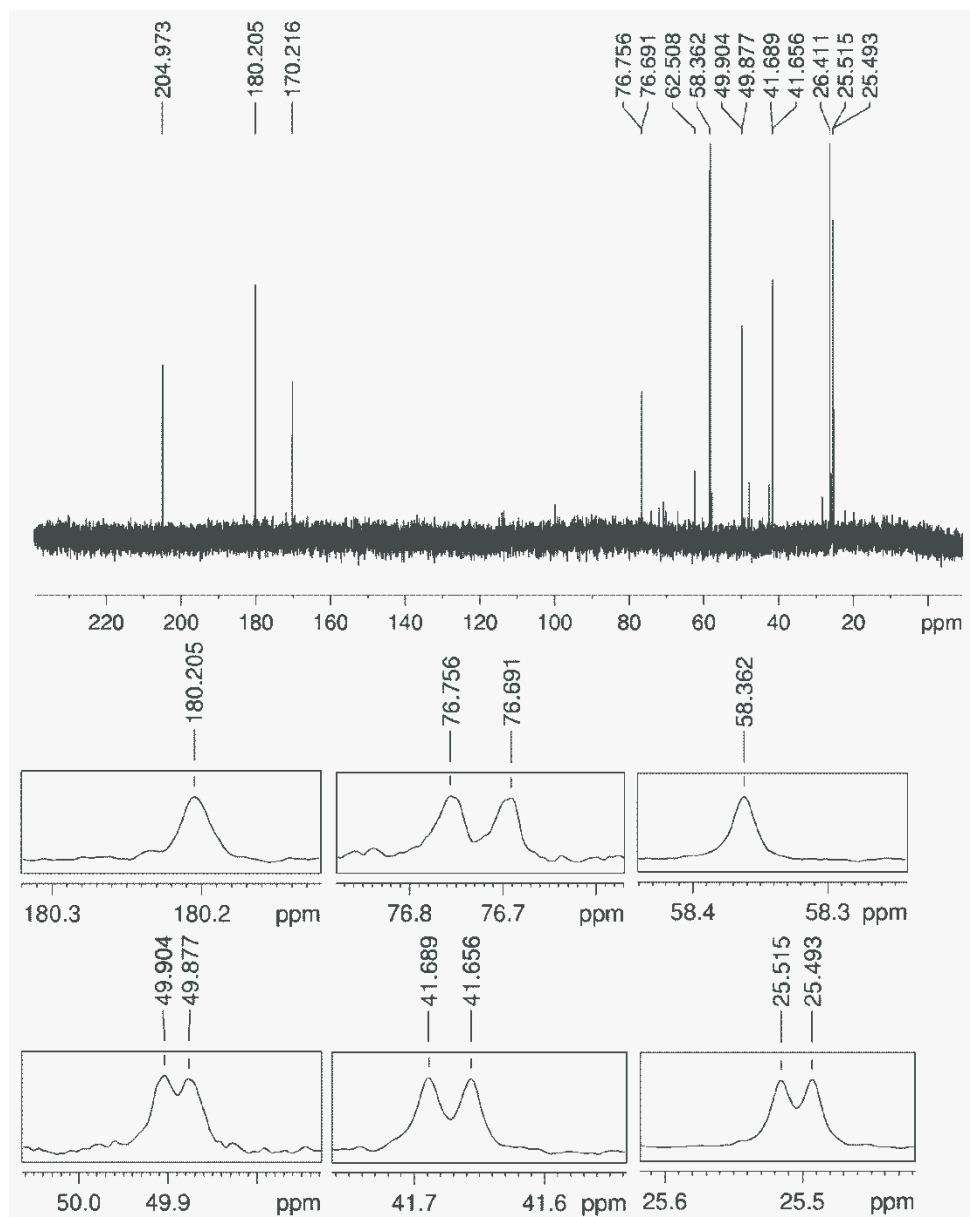


**Figure 1-5.** Full  $^{13}\text{C}$  NMR spectrum of the no enzyme control from figure 3A. Expanded views of the six single peaks corresponding to the six carbons of unreacted mevalonate are shown in the inserts. The additional three peaks between 100 ppm and 175 ppm are from phosphoenolpyruvate that was included in the reaction to regenerate ATP and simplify the spectrum. Spectra were acquired at ambient temperature on a 500 MHz Bruker AV500 spectrometer equipped with a cryoprobe.

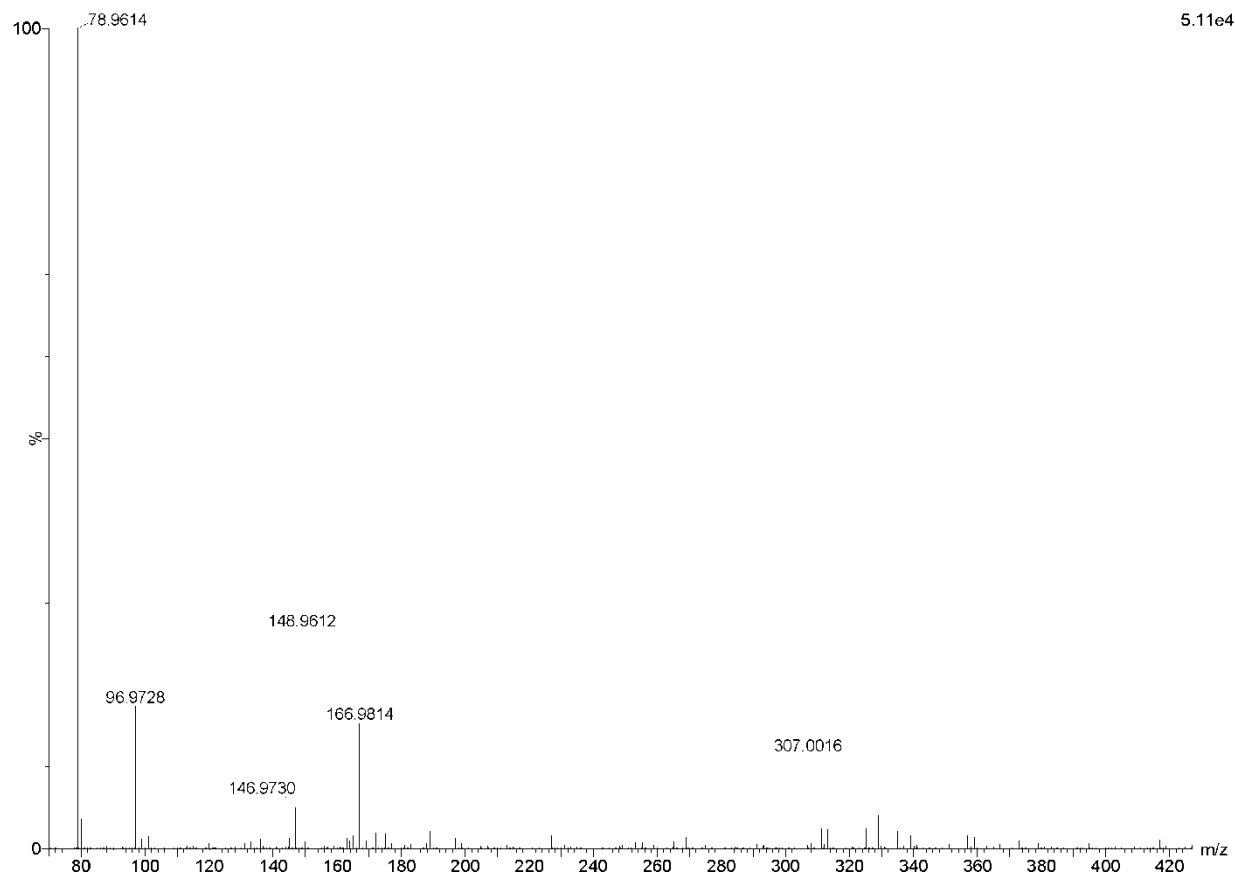




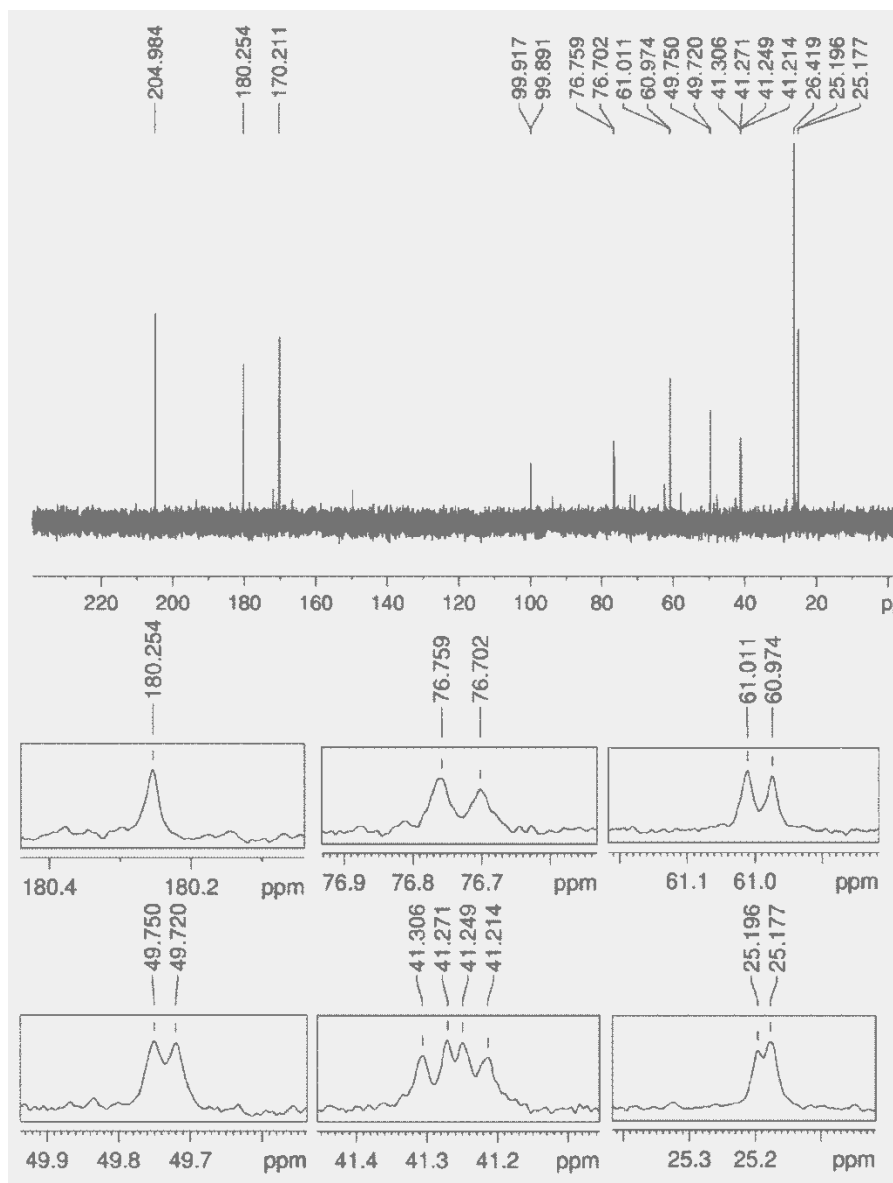
**Figure 1-6.** Full  $^{13}\text{C}$  NMR spectrum of yeast mevalonate-5-kinase activity on (*R*)-Mevalonate from figure 3B. Six peaks corresponding to the six carbons of mevalonate-5-phosphate are shown in the inserts. The additional three peaks at 204.972, 170.215, and 26.412 ppm are from 10 mM pyruvate, which is made during the regeneration of ATP. Spectra were acquired at ambient temperature on a 500 MHz Bruker AV500 spectrometer equipped with a cryoprobe.



**Figure 1-7.** Full  $^{13}\text{C}$  NMR spectrum of mevalonate-3-kinase (Ta1305) activity on (*R*)-Mevalonate from figure 3C. Six chemical shifts corresponding to the six carbons of mevalonate-3-phosphate are shown in the inserts. The additional three peaks at 204.973, 170.216, and 26.411 ppm are from 10 mM pyruvate, which is made during the regeneration of ATP. Spectra were acquired at ambient temperature on a 500 MHz Bruker AV500 spectrometer equipped with a cryoprobe.

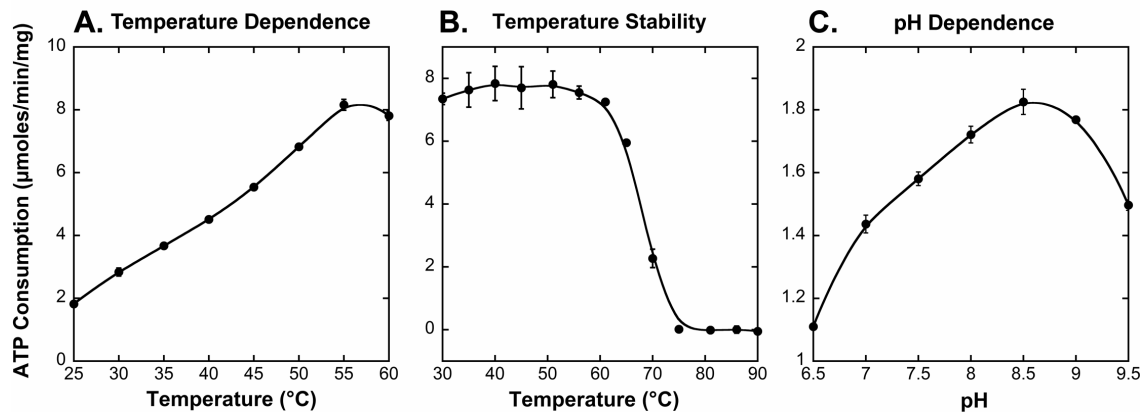


**Figure 1-8.** Negative electrospray ionization mass spectrum after Ta0762 and Ta1305 activity on (*R*)-Mevalonate. The spectrum was collected with a Waters LCT Premier XE time of flight instrument. A sample directly from the NMR experiment was transferred to a GC vial and injected into the multi-mode ionization source with a Waters Acquity UPLC. We observed a mass of 307.0016 *m/z* which is within 0.0031 *m/z* of the mass expected for mevalonate diphosphate [C<sub>6</sub>H<sub>14</sub>O<sub>10</sub>P<sub>2</sub> - H]<sup>-</sup>

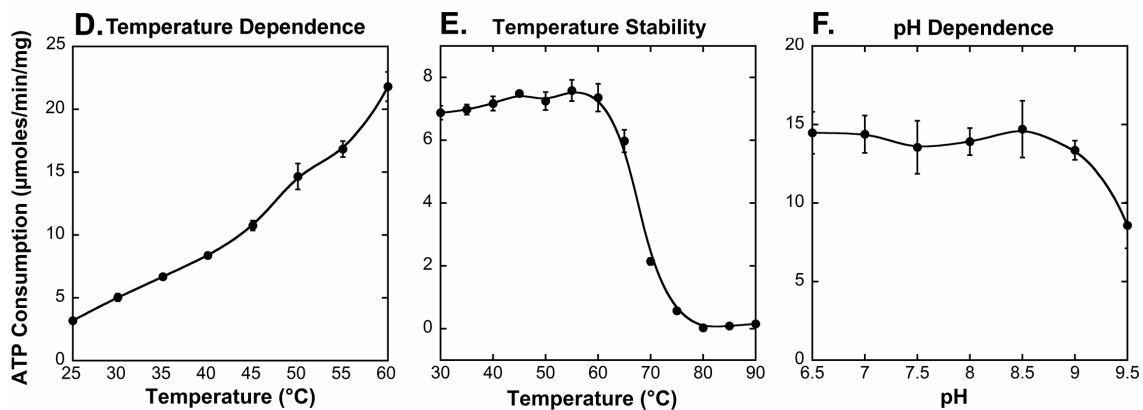


**Figure 1-9.** Full  $^{13}\text{C}$  NMR spectrum of mevalonate-3-kinase (Ta1305) and mevalonate-3-phosphate-5-kinase (Ta0762) activity on (*R*)-Mevalonate from figure 3D. Expanded views of the six peaks corresponding to the six carbons of mevalonate-3,5-bisphosphate are shown in the inserts. The additional three peaks at 204.984, 170.211, and 26.419 ppm are from 10 mM pyruvate, and the doublet at 99.904 is from unreacted PEP. Spectra were acquired at ambient temperature on a 500 MHz Bruker AV500 spectrometer equipped with a cryoprobe.

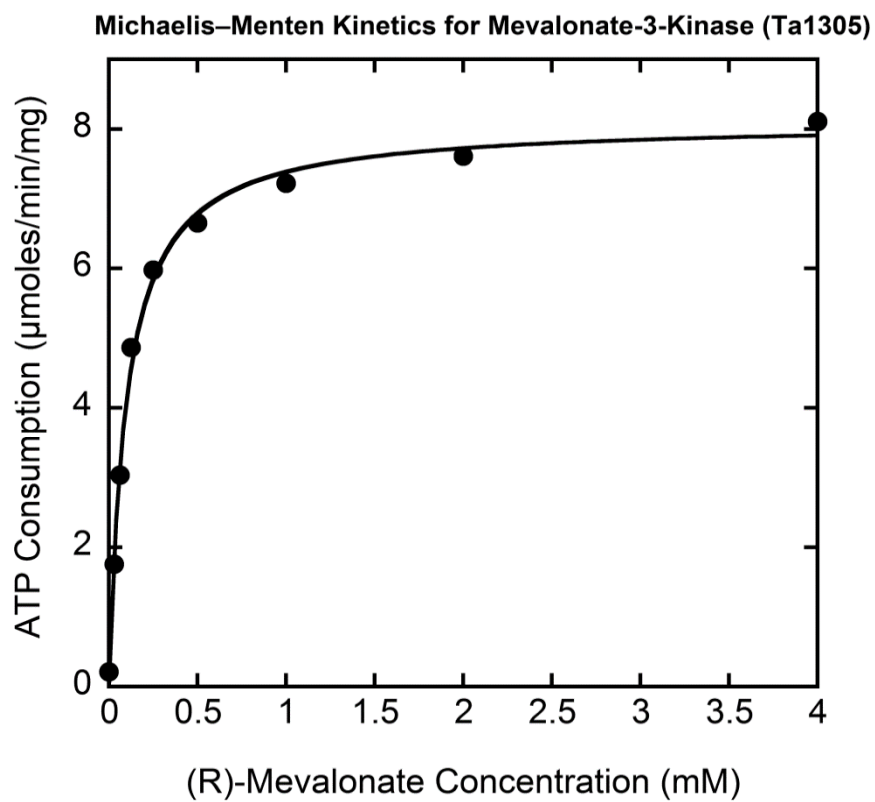
## Mevalonate-3-Kinase



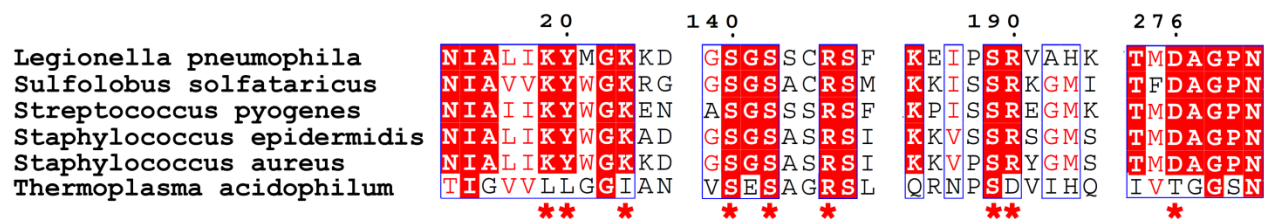
## Mevalonate-3-Phosphate-5-Kinase



**FIGURE 1-10.** Biochemical characterization of mevalonate-3-kinase (A-C) and mevalonate-3-phosphate-5-kinase (D-F). (A) Activity at pH 8.5 over a temperature range of 25-60°C in 5°C increments. (B) Residual activity after incubating at the indicated temperature for 1 hour, then assaying at 55°C and pH 8.5. (C) Activity at 25°C over a pH range of 6.5 – 9.5 in increments of 0.5. (D) Activity at pH 8.0 over a temperature range of 25-60°C in 5°C increments. (E) Residual activity after incubating at the indicated temperature for 1 hour, then assaying at 60°C and pH 8.0. (F) Activity at 60°C over a pH range of 6.5 – 9.5 in increments of 0.5. Curve fits in A-F were generated by a Stineman function (34) and included simply to highlight data trends.



**Figure 1-11.** Michaelis-Menton plot for mevalonate-3-kinase (Ta1305). The enzyme was found to have a  $K_m$  of  $97 \pm 6 \mu\text{M}$  and  $k_{\text{cat}}$  of  $5.0 \pm 0.1 \text{ s}^{-1}$  with respect to (*R*)-Mevalonate.



**Figure 1-12.** Clustal Omega sequence alignment of five bacterial mevalonate pyrophosphate decarboxylases (MDCs) with mevalonate-3-kinase from *T. acidophilum*. Red stars indicate active site residue that interact with mevalonate pyrophosphate in a MDC crystal structure by Barta et al. (33). Red highlighting indicated that at least 5 of 6 amino acids are identical. Only 4 out of 9 active site residues are conserved in mevalonate-3-kinase, suggesting that L19, L12, I23, D190, or T276 may be responsible for the loss of decarboxylase activity.

## 1.6 REFERENCES

1. Reiling, K. K., Yoshikuni, Y., Martin, V. J. J., Newman, J., Bohlmann, J., and Keasling, J. D. (2004) Mono and diterpene production in *Escherichia coli*. *Biotechnol. Bioen.* (87), 200–212.
2. Holstein, S. A., & Hohl, R. J. (2004). Isoprenoids: Remarkable Diversity of Form and Function. *Lipids.* (39), 293–309.
3. Summons, R. E., Jahnke, L. L., Hope, J. M., & Logan, G. a. (1999). 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature*, (6744), 554–557.
4. Goldstein J.L., Brown S.B. (1990) Regulation of the mevalonate pathway. *Nature.* (343), 425-430.
5. Dewick, P.M. (1997) The biosynthesis of C5–C25 terpenoid compounds. *Nat. Prod. Rep.* (14), 111–144.
6. Katsuki H, Bloch K. (1967) Studies on the biosynthesis of ergosterol in yeast. Formation of methylated intermediates. *J. Biol. Chem.* (242), 222–227.
7. Boucher, Y. (2007) Lipids: biosynthesis, function, and evolution, in *Archaea, Molecular and Cellular Biology* (Cavicchioli, R., Ed.) pp 341-353, ASM Press, Washington, DC.
8. Koga, Y., Morii, H. (2005). Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects. *Biosci., Biotechnol., Biochem.* (11), 2019–2034.
9. Faisal, K., Hospital, S. (1990). Physicochemical characterization of tetraether lipids from *Thermoplasma acidophilum*. V. Evidence for the existence of a metastable state in lipids with acyclic hydrocarbon chains. *BBA.* (1024), 54–60.
10. Boucher, Y., Kamekura, M., and Doolittle, W. F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* (52), 515–527.
11. Lombard, J., & Moreira, D. (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Bio. Evol.* (1), 87–99.
12. Smit, A., and Mushegian, A. (2000) Biosynthesis of isoprenoids via mevalonate in archaea: the lost pathway. *Genome Res.* (101), 1468– 1484.
13. Dellas, N., Thomas, S. T., Manning, G., & Noel, J. P. (2013). Discovery of a metabolic alternative to the classical mevalonate pathway. *eLife*, (1), e00672. doi:10.7554/eLife.00672



14. Vannice, J. C., Skaff, D. A., Keightley, A., Addo, J., Wyckoff, G. J., & Miziorko, H. M. (2013). Identification in *Haloferax volcanii* of Phosphomevalonate Decarboxylase and Isopentenyl Phosphate Kinase as Catalysts of the Terminal Enzymatic Reactions in an Archaeal Alternate Mevalonate Pathway. *J. Bacteriol.* (196), 1055-1063.
15. Korman, T.P., Sahachartsiri, B., Li, D., Vinokur, J.M., Eisenberg, D., Bowie, J.U. (2013) A Synthetic Biochemistry System for the in vitro Production of Isoprene from Glycolysis Intermediates. *Prot. Sci.* Just Accepted. doi: 10.1002/pro.2436
16. Ruepp, a, Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., et al. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature*, (6803), 508–513.
17. Benson, D. a, Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2014). GenBank. *Nucleic Acids Res.* 42(Database issue), D32–37.
18. Gibson D.G., Young L., Chuang R-Y., Venter J.C., Hutchison C.A., Smith H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Meth.* (6), 343–345.
19. Technikova-Dobrova, Z., Sardanelli, a M., & Papa, S. (1991). Spectrophotometric determination of functional characteristics of protein kinases with coupled enzymatic assay. *FEBS letters.* (292), 69–72.
20. Wu Z.L., Ethen C.M., Prather B., Machacek M., Jiang W. (2010) Universal phosphatase coupled glycosyltransferase assay. *Glycobiology.* (21), 727–733.
21. Jabalquinto, A. M., Alvear, M., & Emil, E. C. (1988). Physiological aspects and Mechanism of Action of Mevalonate 5-Diphosphate Decarboxylase. *Comp Biochem. Physiol.* (90), 671–677.
22. Krepiy, D., and Miziorko, H. M. (2004) *Protein Sci.* (13), 1875–1881.
23. Davisson, V. J., Woodside, A. B., Neal, T. R., Stremmer, K. E., Muehlbacher, M., & Poulter, C. D. (1986). Phosphorylation of Isoprenoid Alcohols. *J. Org. Chem.* (51), 4768–4779.
24. Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., et al. (2009). HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 37(Database issue), D603–610.
25. Ultra, ChenBioDraw. 11.0, Chemical Structure Drawing Standard, Cambridge Soft Corporation, 100 Cambridge Park Drive, Cambridge, MA 02140, USA.

26. Primak, Y. A., Du, M., Miller, M. C., Wells, D. H., Nielsen, A. T., Weyler, W., & Beck, Z. Q. (2011). Characterization of a feedback-resistant mevalonate kinase from the archaeon *Methanosarcina mazei*. *Applied and environmental microbiology*. (21), 7772–7778.
27. Grochowski, L. L., Xu, H. M., and White, R. H. (2006) *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J. Bacteriol.* (188), 3192–3198.
28. Chen, M., & Poulter, C. D. (2010). Characterization of thermophilic archaeal isopentenyl phosphate kinases. *Biochemistry*. (1), 207–217.
29. Schomburg I., Chang A., Placzek S., Söhngen C., Rother M., Lang M., Munaretto C., Ulas S., Stelzer M., Grote A., Scheer M. & Schomburg D. (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* (41), D764-D772
30. Lindberg M., Yuan C., de Waard A. and Bloch K. (1962) On the formation of Isopentenyl Pyrophosphate. *Biochemistry*. (1), 182-188.
31. Voynova, N. E., Fu, Z., Battaile, K. P., Herdendorf, T. J., Kim, J.-J. P., & Miziorko, H. M. (2008). Human mevalonate diphosphate decarboxylase: characterization, investigation of the mevalonate diphosphate binding site, and crystal structure. *Arch. Biochem. Biophys.* (480), 58–67.
32. Byres, E., Alphey, M. S., Smith, T. K., & Hunter, W. N. (2007). Crystal structures of *Trypanosoma brucei* and *Staphylococcus aureus* mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity. *J. Mol. Biol.* (2), 540–553.
33. Barta, M. L., Skaff, D. A., McWhorter, W. J., Herdendorf, T. J., Miziorko, H. M., & Geisbrecht, B. V. (2011). Crystal structures of *Staphylococcus epidermidis* mevalonate diphosphate decarboxylase bound to inhibitory analogs reveal new insight into substrate binding and catalysis. *J. Biol. Chem.* (286), 23900–23910.
34. Stineman, R. W. (1980) A Consistently Well Behaved Method of Interpolation. *Creative Computing*. (6), 54-57.

## CHAPTER 2

### ADAPTATION TO LIFE IN ACID THROUGH A NOVEL MEVALONATE PATHWAY

#### 2.1 ABSTRACT

Extreme acidophiles are capable of growth at pH values near zero. Sustaining life in acidic environments requires extensive adaptations of membranes, proton pumps, and DNA repair mechanisms. Here we describe an adaptation of a core biochemical pathway, the mevalonate pathway, in extreme acidophiles. Two previously known mevalonate pathways involve ATP dependent decarboxylation of either mevalonate 5-phosphate or mevalonate 5-pyrophosphate, in which a single enzyme carries out two essential steps: (1) phosphorylation of the mevalonate moiety at the 3-OH position and (2) subsequent decarboxylation. We now demonstrate that in extreme acidophiles, decarboxylation is carried out by two separate enzymes acting in tandem: a previously identified enzyme generates mevalonate 3,5-bisphosphate and a new decarboxylase we describe here, mevalonate 3,5-bisphosphate decarboxylase, produces isopentenyl phosphate. Why use two enzymes in acidophiles when one enzyme provides both functionalities in all other organisms examined to date? We find that at low pH, the dual function enzyme, mevalonate 5-phosphate decarboxylase is unable to carry out the first phosphorylation step, yet retains its ability to perform decarboxylation. We therefore propose that extreme acidophiles had to replace the dual-purpose enzyme with two specialized enzymes to efficiently produce isoprenoids in extremely acidic environments.

## 2.2 INTRODUCTION

Extremophiles are organisms capable of surviving in the harshest conditions on earth such as temperatures exceeding 120°C in hydrothermal vents, salinity exceeding 5M NaCl in evaporating lakes, and acidity below pH 0 in acid mine drainage <sup>1-3</sup>. The vast majority of extremophiles belong to the archaeal domain of life, having adapted to conditions prevalent on a primordial earth <sup>4</sup>. Growth in extremely acidic conditions is especially challenging as the organism must maintain a 100,000 fold H<sup>+</sup> gradient across its membrane while allowing for the import and export of metabolites and other molecules <sup>5</sup>.

The lowest pH to support life so far recorded is pH -0.06 (1.2M sulfuric acid) by *Picrophilus torridus*, a member of the archaeal order thermoplasmatales <sup>6</sup>. This unique order contains only 11 characterized organisms, all of which are extreme acidophiles capable of growth at pH 0.5 and below <sup>6-9</sup>. Thermoplasmatales have among the smallest genomes of any free living organism (<2 Mb), reversed membrane potentials, and all but *P. torridus* lack a cell wall entirely <sup>5</sup>. The first line of defense against acidity in thermoplasmatales is a highly impermeable lipid monolayer made of C40 tetra-ether lipids <sup>10</sup>. The C40 alkyl chains are made entirely from tandem repeats of 5-carbon isoprene units, and are connected to polar head groups through ether linkages <sup>10</sup>. Isoprenoid based lipids pack more tightly, making archaeal membranes less permeable to small molecules and the ether linkages impart acid stability <sup>11</sup>.

The 5-carbon precursor for all isoprenoids, isopentenyl pyrophosphate (IPP), is generated by the mevalonate pathway in eukaryotes, archaea, and some bacteria <sup>12</sup>. All known mevalonate pathways first produce (R)-mevalonate through the condensation of three acetyl-CoA molecules followed by a reduction step to yield mevalonate (Fig. 2.1). In eukaryotes,

mevalonate is then phosphorylated twice at the 5-OH position to generate mevalonate 5-pyrophosphate, then decarboxylated to yield IPP (Fig. 2.1, black pathway) <sup>13</sup>. In most archaea however, mevalonate is phosphorylated once to make mevalonate 5-phosphate (M5P), then decarboxylated to isopentenyl phosphate (IP), and finally phosphorylated again to generate IPP (Fig. 2.1, blue pathway) <sup>14,15</sup>. In both the eukaryotic and classical archaeal pathways, decarboxylation is ATP-dependent and proceeds in two sequential steps via a single enzyme: (1) phosphorylation at the 3-OH position of the mevalonate moiety using ATP and (2) decarboxylation (Fig. 2.2) <sup>16</sup>. Thus, the decarboxylases are dual function enzymes. The phosphorylation step adds a phosphate group to the 3-OH position, which acts as a good leaving group and primes the molecule for decarboxylation<sup>16</sup>.

We and another group recently discovered two new enzymes in *Thermoplasma acidophilum*, mevalonate 3-kinase (EC 2.7.1.185) and mevalonate 3-phosphate 5-kinase (EC 2.7.1.186), whose sequential catalysis produces mevalonate 3,5-bisphosphate from mevalonate suggesting that there may be another route to produce isoprenoids besides the canonical archaeal pathway (Fig. 2.1, red pathway) <sup>17,18</sup>. However, the putative mevalonate 3,5-bisphosphate decarboxylase (MBD) remained unidentified to complete this alternative archaeal pathway. Here we report the identification of MBD from *T. acidophilum*, which produces IP through the ATP independent decarboxylation of mevalonate 3,5-bisphosphate (Fig. 2.2, reaction in brackets).

The new pathway is particularly odd, because the 3-OH phosphorylation and subsequent decarboxylation are carried out by two distinct enzymes even though both enzymes are structurally homologous to dual-function decarboxylases. Thus, it appears that the two

enzymes evolved from a dual function decarboxylase, but became specialized. Why? We show that the new pathway is only present in extreme acidophiles, suggesting that low pH may require divergent activities. Indeed we find that the dual function mevalonate 5-phosphate decarboxylase (MMD) from *Roseiflexus castenholzii* is unable to carry out the kinase step at low pH, but retains decarboxylase activity. Thus, it is possible that a separate enzyme became specialized to handle the kinase function at low pH. This adaptation could have evolved through a duplication or horizontal transfer event, followed by specialization to support life in extremely acidic environments.

## 2.3 RESULTS

**Sequence homology suggests that Ta0893 is a decarboxylase.** To search for mevalonate 3,5-bisphosphate decarboxylase (MBD) in *Thermoplasma acidophilum*, we identified proteins homologous to decarboxylases. Two proteins from the archeon *Thermoplasma acidophilum*, Ta1305 and Ta0893, were computationally annotated in the year 2000 as “mevalonate pyrophosphate decarboxylase”<sup>9</sup>. As reported recently, however, Ta1305 is actually mevalonate 3-kinase (EC 2.7.1.185)<sup>17</sup>. A structural homology model of Ta0893 made with PHYRE2 suggested significant similarity to mevalonate 5-pyrophosphate decarboxylase from yeast (PDB: 1FI4, confidence score of 100%) and contained the invariant Asp/Lys/Arg catalytic triad necessary for decarboxylation in the correct positions (Fig. 2.3)<sup>19–21</sup>. Unfortunately, Ta0893 consistently formed inclusion bodies when expressed in *E. coli* and no activity could be detected in extracts. As a result we set out to find the putative missing decarboxylase via direct purification from *T. acidophilum*.

**Identification of MBD activity in *T. acidophilum* lysate.** Following growth of *Thermoplasma acidophilum*, we were able to detect MBD activity in crude lysate using mevalonate 3,5-bisphosphate as a substrate. To initially fractionate the lysate, we separated the *T. acidophilum* lysate by anion exchange chromatography using a HiTrap Q HP column. Ten fractions were collected and assayed for MBD activity<sup>18</sup>. As shown in Fig. 2.4A, fraction 7 showed the highest MBD activity, so we further separated fraction 7 in two side-by-side lanes on a native polyacrylamide gel. Coomassie blue staining (detection limit: ~5 ng) of one lane revealed 25 distinct bands. The adjacent gel lane (unstained) was then cut into 19 fragments (Fig. 2.4B), pulverized and assayed for MBD activity. As shown in Fig. 2.4C, MBD activity was detected in gel fragments 13-16. Each fragment was independently analyzed by NanoLC/MS/MS for protein identification. The results revealed that Ta0893 and 6 other proteins were present in all 4 gel fragments (Table 2.1), suggesting that one of them might be the missing decarboxylase.

**Ta0893 shows decarboxylase activity when truncated or co-expressed with chaperones.** Identification of Ta0893 through mass spectrometry combined with clear homology motivated a larger effort to produce active, soluble Ta0893 in *E. coli*. Nevertheless, all attempts to refold Ta0893 from inclusion bodies failed after varying many parameters such as pH, temperature, salts, and substrates, and screening using the Quickfold kit from Athena Enzyme Systems<sup>22</sup>. We were finally able to detect MBD activity in Ni-NTA eluates under two expression conditions in *E. coli*: (1) truncating Ta0893 by removing 30 AAs from the N-terminus and (2) co-expression of the full-length Ta0893 protein with *E. coli* chaperones, GroES/GroEL (Fig. 2.4D)<sup>23</sup>. We chose to truncate 30 AAs from the N-terminus of Ta0893 because this region was not present in the Ta0893 homolog from *Thermoplasma volcanium*, suggesting the first 30 AAs were not required for function. Size exclusion chromatography showed that the apparent

molecular weight of MBD activity is  $87\pm 4$  kDa in both recombinant Ta0893 co-expressed with chaperones as well as *T. acidophilum* lysate (Fig. 2.4E). The predicted molecular weight of a Ta0893 monomer is 46.3 kDa, suggesting that Ta0893 is a native homodimer. The finding that MBD activity in *E. coli* is dependent on Ta0893 expression and correlates with the same molecular weight as activity in *T. acidophilum* lysate, strongly suggests that Ta0893 is the missing MBD.

**Ta0893 produces isopentenyl phosphate.** To confirm that recombinant Ta0893 produces isopentenyl phosphate (IP), we coupled MBD activity to IP kinase from *T. acidophilum* which specifically phosphorylates IP<sup>24</sup>. Robust ATP consumption was detected when IP kinase was added to a reaction containing Ta0893 pre-incubated with (R)-mevalonate 3,5-bisphosphate (Fig. 2.4F).

**Other extreme acidophiles also display MBD activity.** While we confirmed that Ta0893 is a *bona fide* MBD, the lack of robust Ta0893 expression led us to seek close homologs better suited for recombinant expression to characterize the biochemical properties of MBD. A BLASTp search for homologs of Ta0893 reveals only 8 organisms containing an MBD homolog with greater than 30% sequence identity. Indeed, these 8 organisms represent all the sequenced thermoplasmatales. We cloned each MBD homolog (*Thermoplasma volcanium*, *Picrophilus torridus*, *Ferroplasma acidarmanus*, *Ferroplasma sp. Type II*, and *Acidiplasma sp.*



*MBA-1*)<sup>1,2</sup> into *E. coli* expression vectors. Using standard expression conditions, we observed activity in Ni-NTA eluates for the *T. volcanium* and *P. torridus* MBD homologs (Fig. 2.5). MBD homologs from *Ferroplasma* and *Acidiplasma* formed inclusion bodies under all tested expression conditions.

**The specialized mevalonate pathway is unique to thermoplasmatales.** A BLASTp search using MBD from *T. acidophilum* (Ta0893) shows strong sequence identity for all 8 sequenced thermoplasmatales. This includes *Acidiplasma sp. MBA-1* (68%), *Acidiplasma aeolicum* (68%), *Acidiplasma cupricumulans* (68%), *Thermoplasma volcanium* (63%), *Ferroplasma sp. Type II* (50%), *Ferroplasma acidarmanus* (48%), and *Picrophilus torridus* (40%) (Fig. 2.6A). The next strongest alignments (29%) after thermoplasmatales are proteins from eubacteria of the phylogenetic family chloroflexaceae. Indeed, the chloroflexaceae *Roseiflexus castenholzii* is known to contain a *bona fide* MMD of the canonical archaeal mevalonate pathway<sup>14</sup>. Sequence alignment of *T. acidophilum* MBD against classical mevalonate pathway decarboxylases shows retention of the invariant Asp/Lys/Arg catalytic triad required for decarboxylation, however MBDs are unique in that they are missing both nearly invariant ATP binding residues (Fig. 2.6B)<sup>19,25–27</sup>. Indeed, unlike all other mevalonate pathway decarboxylases, MBD does not require ATP. MBD's complement, mevalonate 3-kinase (Ta1305) has equally strong homologs (39-67% ID) in all 8 thermoplasmatales yet has no

---

<sup>1</sup> The MBD homolog from *Acidiplasma sp. MBA-1* has the identical amino acid sequence as *A. aeolicum* and is 99.7% identical (1AA substitution) to *A. cupricumulans*. We cloned the MBA-1 homolog to represent all 3 species of the *Acidiplasma* genus.

<sup>2</sup> NCBI accession numbers: *T. acidophilum* (WP\_010901303.1), *T. volcanium* (WP\_010916684.1), *P. torridus* (WP\_010917040.1), *F. acidarmanus* (WP\_009887850.1), *F. sp. Type II* (EQB73519.1), and *A. sp. MBA-1* (WP\_048100791.1).

detectable homologs in any other organism, suggesting that this pathway is unique to thermoplasmatales.

**Biochemical characterization of MBD from *Picrophilus torridus*.** Since the *P. torridus* MBD showed robust expression in *E. coli*, we chose to purify this enzyme for further characterization (Fig. 2.7A) and compare it to the classical mevalonate 5-phosphate decarboxylase (MMD) from *Roseiflexus castenholzii*<sup>14</sup>. *Roseiflexus castenholzii* is a member of the chloroflexaceae family whose MMDs are most closely related to MBDs of *Thermoplasmatales* and are therefore the closest known evolutionary precursors. Purified *P. torridus* MBD showed clear specificity for mevalonate 3,5-bisphosphate and no detectable decarboxylase activity on mevalonate, mevalonate 3-phosphate, mevalonate 5-phosphate or mevalonate 5-pyrophosphate (Fig. 2.7B, black bar). *P. torridus* MBD worked optimally at 70°C and a pH around 5 where it had a  $k_{cat}$  of approximately  $7.0 \text{ s}^{-1}$  with respect to (R)-mevalonate 3,5-bisphosphate (Supplemental Fig. 2.8-2.10). This  $k_{cat}$  value is comparable to *R. castenholzii* MMD ( $1.7 \pm 0.1 \text{ s}^{-1}$ )<sup>14</sup>. Interestingly, *R. castenholzii* MMD was also active on mevalonate 3,5-bisphosphate when supplied with ADP as a co-factor (Fig. 2.7B, blue bars). We were unable to accurately determine the  $K_m$  for MBD since it was still at  $V_{max}$  when assayed at the detection limit of our GC-FID assay (30  $\mu\text{M}$ ).

**MMD loses kinase function at low pH.** We characterized the previously un-reported activity of MMD on its reaction intermediate (mevalonate 3,5-bisphosphate) and compared it to its native substrate. Interestingly, at low pH, MMD completely loses its ability to convert mevalonate 5-phosphate to IP (its native reaction), but continues to decarboxylate mevalonate 3,5-bisphosphate, effectively becoming a MBD at low pH because it cannot perform the first

kinase step (Fig. 2.11). The fact that decarboxylase activity remains intact suggests that low pH does not cause global inactivation or unfolding of the enzyme. These results suggest that low pH requires specialization to provide a kinase function. The adaptation could come in the form of adjustments to the normal MMD or the evolution of a new kinase.

## 2.4 DISCUSSION

The novel decarboxylase reported here demonstrates a unique mevalonate pathway in *T. acidophilum*. While the classical archaeal pathway phosphorylates mevalonate at the 5-OH position to yield mevalonate 5-phosphate, and then uses MMD to produce IP in an ATP dependent reaction, the *T. acidophilum* pathway produces the same end product, but uses a completely different set of enzymes and metabolites<sup>17</sup>. The pathway in *T. acidophilum* phosphorylates mevalonate at the 3-OH position and the 5-OH position sequentially by two distinct enzymes to yield mevalonate 3,5-bisphosphate. This is followed by the action of MBD which carries out ATP independent decarboxylation to produce IP. Both archaeal pathways use IP kinase to produce IPP (Fig. 2.1, blue and red routes)<sup>24</sup>.

Localization of this unique pathway to the most acid tolerant organisms on earth suggests that the pathway may confer an evolutionary advantage in extremely acidic environments. *P. torridus* has an internal pH of 4.6 while *T. acidophilum* and *F. acidarmanus* have an internal pH of 5.5 and 5.6 respectively<sup>28-30</sup>. We observe that at pH 4.3, the conversion of mevalonate 5-phosphate by MMD stops completely, while the decarboxylation activity of mevalonate 3,5-bisphosphate by MMD remains intact. We propose that the common ancestor of thermoplasmatales also had a low internal pH and the classical decarboxylation reaction was inefficient, which would have applied evolutionary pressure to adapt. One possible way to adapt

would be to find a way to make both the kinase and decarboxylation steps more effective at low pH. The fact that such a seemingly simple adaptation did not occur, suggests that it may be difficult to accomplish. Interestingly, a different evolutionary pathway was chosen by thermoplasmatales and two separate enzymes developed. We propose a model in which a horizontal transfer or a gene duplication event placed two MMD enzymes into the common ancestor and over time these two enzymes became specialized<sup>9,31</sup>. One lost its decarboxylase function to become a mevalonate 3-kinase, and the second lost its kinase function to become a dedicated decarboxylase (MBD). These two specialized enzymes were more efficient at low pH than a single dual-function enzyme. One possible explanation for the kinase function of *R. castenholzii* MMD stopping while decarboxylation continues at low pH could be the acidity perturbing the protonation state of an active site residue that deprotonates the 3-OH group of mevalonate and triggers nucleophilic attack on the  $\gamma$ -phosphate of ATP.

One peculiar observation is that the two specialized functions do not act in tandem, but are one step removed, with mevalonate 3-phosphate 5-kinase acting between the 3-OH phosphorylation of mevalonate and the decarboxylation by MBD. The reason for this is unclear.

Decarboxylation in the mevalonate pathway has been extensively studied in the eukaryotic homolog, mevalonate 5-pyrophosphate decarboxylase, which converts mevalonate 5-pyrophosphate to IPP in an ATP-dependent manner<sup>16,21,32-35</sup>. The mechanism was previously thought to occur via a single step in which the phosphorylation of the 3-OH position leads to an unstable and transient intermediate that rapidly decomposes into IPP, CO<sub>2</sub> and PO<sub>4</sub><sup>26,27</sup>. Our previous identification of mevalonate 3,5-bisphosphate as a stable metabolite suggested that decarboxylation is not spontaneous and must be carried out by enzyme catalysis<sup>17,36</sup>. The

identification of MBD provides further support for the two-step model since *T. acidophilum* carries out the reactions through two entirely separate enzymes. Furthermore, we show that MMD from *Roseiflexus castenholzii* can accept its intermediate, mevalonate 3,5-bisphosphate, and carry out the second step of its dual function independently of the first step. These results strongly indicate that decarboxylation is catalyzed, not spontaneous, and that the ATP dependent decarboxylases operate via a two-step mechanism.

One unexpected finding of our study is that MMD from *Roseiflexus castenholzii* requires ADP as a co-factor to decarboxylate mevalonate 3,5-bisphosphate. We propose that ADP binding is necessary to form the active enzyme-substrate complex. Structural analysis of mevalonate 5-pyrophosphate decarboxylase from *Staphylococcus epidermidis* showed that ATP binding triggers two flexible loops to close over the substrates in the active site pocket<sup>32</sup>. We suggest that ADP is necessary as a co-factor to mimic the presence of the native substrates (mevalonate 5-phosphate and ATP), which would trigger loop closure and form the active complex. After decarboxylation, the loops would presumably transition to an open state and release IP, ADP, CO<sub>2</sub> and PO<sub>4</sub>. Sequence alignment with mevalonate pyrophosphate decarboxylases shows that MBD does not contain two nearly invariant amino acids that bind the adenine ring of ATP (Fig. 2.6B)<sup>37</sup>. Presumably, the specialization of MBD to become a dedicated decarboxylase involved the loss of its ADP requirement.

In summary, we have identified and characterized mevalonate 3,5-bisphosphate decarboxylase, a novel enzyme which produces isopentenyl phosphate and completes the unique mevalonate pathway of extreme acidophiles comprising the archaeal order thermoplasmatales. We also report that the two steps of the mevalonate 5-phosphate

decarboxylase mechanism can be separated as demonstrated by the robust activity of MMD directly on its intermediate, mevalonate 3,5-bisphosphate. Indeed, at pH values representative of the *P. torridus* cytoplasm (~4.6), MMD loses its kinase function completely and becomes a decarboxylase only. We propose that thermoplasmatales adapted their mevalonate pathway by replacing MMD with two specialized enzymes in order to produce isoprenoids in extremely acidic environments.

## 2.5 METHODS

**Materials.** *E. coli* BL21(DE3) Gold (Agilent) was grown in Miller LB media (Fisher) for both cloning and expression of recombinant proteins. Plasmid pET28a(+) was purchased from Novagen and pBB541 encoding GroEL/GroES was obtained from Addgene<sup>23</sup>. Ni-NTA resin was purchased from Qiagen. Native gels were purchased from Expedeon. All other chemicals were purchased from Sigma-Aldrich unless otherwise noted.

***Thermoplasma acidophilum* growth.** *Thermoplasma acidophilum* was obtained as a live culture from NITE Biological Resource Center (Tokyo, Japan). The organism was grown in NBRC medium 280 at 60°C to OD<sub>600</sub> = 0.5. Cells were harvested by centrifugation at 5000xg. 2g of wet cell pellet (4L culture) were resuspended in 20 mL of 50 mM sodium phosphate buffer [pH 6.5]. The cells were chilled on ice for 30 min and then lysed by sonication. A cell-free lysate was obtained after centrifugation at 30,000xg.

**Anion exchange chromatography.** Initial separation of the *T. acidophilum* lysate was adapted from the procedure of another research group<sup>18</sup>. In brief, 10 mL of *T. acidophilum* lysate was

applied to a 1 mL HiTrap Q HP anion exchange column (strong quaternary ammonium anion exchanger) pre-equilibrated with 50 mM sodium phosphate buffer [pH 6.5], connected to an AKTA FPLC system. The column was washed with 20 mL of 50 mM sodium phosphate buffer [pH 6.5] followed by a gradient from 0 to 1 M NaCl over 20 min and 2 mL fractions were collected. A flow rate of 1 mL/min was used for all steps.

**Native gels.** 20  $\mu$ L of the most active fraction from anion exchange chromatography was loaded into two side-by-side wells of a 20% native polyacrylamide gel (Expedeon). The gel was developed by running towards the anode at 40V for 50 min followed by 150V for 16 hrs in a cold room (4°C). One lane was stained with Expedeon InstantBlue protein stain. The unstained lane was cut into 19 fragments using the stained lane as a reference. Each gel fragment was pulverized by sonication in 500  $\mu$ L of 50 mM sodium phosphate buffer [pH 6.5] containing 500 mM NaCl. 50  $\mu$ L of the resulting slurry was used for the GC-FID assay below.

**GC-FID Decarboxylase Assay.** A 50  $\mu$ L sample of chromatography fraction was added to a 150  $\mu$ L reaction mixture consisting of 50 mM sodium phosphate buffer [pH 6.5], 500 mM NaCl, 10 mM (R)-mevalonate, 20 mM ATP, 5 mM MgCl<sub>2</sub>, 10  $\mu$ g mevalonate 3-kinase (Ta1305), and 10  $\mu$ g mevalonate-3-phosphate-5-kinase (Ta0762). The reaction was incubated for 24 hours at 60°C. Any generated isopentenyl phosphate was then hydrolyzed into isoprenol and free phosphate by adding 100  $\mu$ L of 1 M bis-tris propane [pH 9.0], followed by 30 U of alkaline phosphatase from bovine intestinal mucosa. After incubation at 37°C for 2 hours, the reaction was extracted with 200  $\mu$ L hexanes. 5  $\mu$ L of the hexanes layer was injected into a HP5890 Series II Gas Chromatograph (flame ionization detector) connected to a HP-INNOWAX column (0.320 mm x 30 m, Agilent). The carrier gas was helium with a flow rate of 5 mL/min. Initial oven

temperature was set to 70°C for 2 min, followed by a ramp of 20°C/min for 1 min, and finally a ramp at 50°C/min to a final temperature of 200°C, which was held for 1 min. The inlet was kept at 250°C and the detector at 330°C. Isoprenol eluted at 2.98 min and the sample concentration was determined by comparison to a standard. All GC-FID samples were prepared in duplicate.

**Mass spectrometry.** The four native gel bands from the stained lane corresponding to the highest activities were excised and submitted to ProtTech Inc. for independent analysis via NanoLC/MS/MS. The following steps were carried out by Protech: Peptides were digested in-gel using sequencing grade modified trypsin (Promega) in 100 mM ammonium bicarbonate [pH 8.5] buffer. DTT and iodoacetamide were added for reduction and alkylation of cysteine residues. The digested peptides were extracted with acetonitrile, dried using a Thermo SpeedVac, then redissolved in 2% acetonitrile, 97.5% water, and 0.5% formic acid. Peptides were separated using a high pressure liquid chromatography system (HPLC) fitted with a reversed-phase C18 column (75 µM ID x 8 cm). Samples eluted from the HPLC column were directly ionized by electrospray ionization and analyzed by an ion trap mass spectrometer (LCQ DECA XP Plus, Thermo). MS/MS spectra were acquired via low energy collision induced dissociation. The collected mass spectrometric data were searched against the NCBI protein database using ProtTech's ProtQuest software. Peptides were reported with a mass range of 550 to 1800 Da and a signal to noise ratio greater than or equal to 5.

**General Cloning.** All genes were codon optimized and synthesized by IDT. Each gene was inserted between the NdeI and XhoI sites of the pET28a(+) vector, which allowed for the addition of an N-terminal 6xHis tag. Genes were synthesized with an extra 25 base-pairs complementary to the pET28a(+) vector at the NdeI and XhoI sites. A standard Gibson method



was used to assemble all constructs by mixing 30 ng of synthesized DNA, with 10 ng of pET28a(+) digested with NdeI and XhoI and 7.5  $\mu$ L of Gibson assembly mix<sup>38</sup>. After incubation at 50°C for 2 hrs, 5  $\mu$ L was used to transform *E. coli* BL21(DE3) Gold and transformants were selected on LB-agar plates containing 50  $\mu$ g/mL kanamycin.

**Expression and Purification.** Protein expression and purification was carried out as described previously<sup>17</sup>. In brief, 1 L of LB media was inoculated with 5 mL of *E. coli* overnight starter culture with 50  $\mu$ g/mL kanamycin and/or 100  $\mu$ g/mL spectinomycin as needed. The cells were grown to an OD<sub>600</sub> of 0.5-1.0 and induced at 37°C with 1.0 mM IPTG. After 18 hours, cells were pelleted, resuspended in 5 mL of buffer A (50 mM bis-tris propane [pH 7.5] buffer, 100 mM NaCl, 10 mM imidazole), lysed by sonication, and cell debris removed by centrifugation at 30,000xg for 20 min. The lysate was mixed with 3 mL of a Ni-NTA slurry and incubated for 15 min at 4°C with gentle mixing. The lysate mixture was packed into a column and the Ni-NTA beads were washed 3 times with 20 mL of buffer A. Protein was eluted with 4 mL of buffer A containing 250 mM imidazole. Co-expression of GroEL/GroES was achieved by the addition of plasmid pBB541 (addgene) to the expression strain<sup>23</sup>. For biochemical characterization, *P. torridus* MBD was further purified by heating the eluate at 60°C for 2 hrs to precipitate native *E. coli* proteins followed by one passage through a HiTrap Q HP anion exchange column to remove negatively charged proteins.

**Coupled Enzyme Assay.** The MBD product was confirmed to be IP via coupling to IP kinase. A 100  $\mu$ L reaction containing 25 mM bis-tris propane buffer [pH 6.5], 5 mM (R)-mevalonate, 10 mM ATP, 5 mM KCl, 5 mM MgCl<sub>2</sub>, 5  $\mu$ g mevalonate 3-kinase (Ta1305), 5  $\mu$ g mevalonate 3-phosphate 5-kinase (Ta0893), and 50  $\mu$ L Ta0893 Ni-NTA eluate (0.5 mg/mL total protein) was

incubated for 24 hours at 60°C. To measure ATP consumption, we added 1  $\mu$ L coupling enzyme mix (lactate dehydrogenase and pyruvate kinase mix from rabbit muscle, Sigma), 15 mM PEP, and titrated in NADH via 0.5 mM increments to bring the OD<sub>340</sub> to 1.0. We then recorded the OD<sub>340</sub> over 10 min on a SpectraMax M5 microplate reader. 5  $\mu$ g IP kinase from *T. acidophilum* (Ta0103) was added at the 5 min mark. A negative control replaced Ta0893 eluate with water.

**Biochemical characterization.** The optimal pH for *P. torridus* MBD was determined in 0.5 pH unit increments. Mevalonate 3,5-bisphosphate was enzymatically generated by mixing 10 mM (R)-mevalonate with 20 mM ATP, 5 mM MgCl<sub>2</sub>, 100 mM NaCl, 10  $\mu$ g mevalonate 3-kinase (Ta1305), and 10  $\mu$ g mevalonate-3-phosphate-5-kinase (Ta0762). After 1 hour at 60°C, 150  $\mu$ L of this mixture was combined with 50  $\mu$ L of 1M sodium acetate (for pH 3.5-5.5) or bis-tris propane buffer (for pH 6.5-8.0). pH adjustments were made using HCl and NaOH and confirmed using a micro pH electrode (Hanna Inst. 1083B). The reaction was initiated by the addition of 0.1  $\mu$ g *P. torridus* MBD and incubated at 60°C for 1 hour followed by transferring the vials into boiling water for 1 min to stop the reaction. Analysis was carried out via GC-FID as described above. The optimal temperature was determined in the same manner with pH held constant at 5.5. The data points for optimum pH and temperature were confirmed to be initial rates by quenching a control reaction (pH 5.5, 60°C) at 30, 60, 90, and 120 min as shown in Supplemental Fig. 2.10. Substrate specificity was tested by replacing (R)-mevalonate with commercially available (R)-mevalonate 5-phosphate, or (R)-mevalonate 5-pyrophosphate. Activity on mevalonate-3-phosphate was tested by omitting Ta0762. Kinetic characterization of *P. torridus* MBD were carried out at pH 5.5 and 70°C using the GC-FID assay with 3 ng enzyme over a range of 0.03 to 5 mM (R)-mevalonate 3,5-bisphosphate. 50  $\mu$ L aliquots of this reaction were quenched at 0, 20, 40, and 60 minutes by boiling for 1 min. The enzyme was at  $V_{max}$  for all

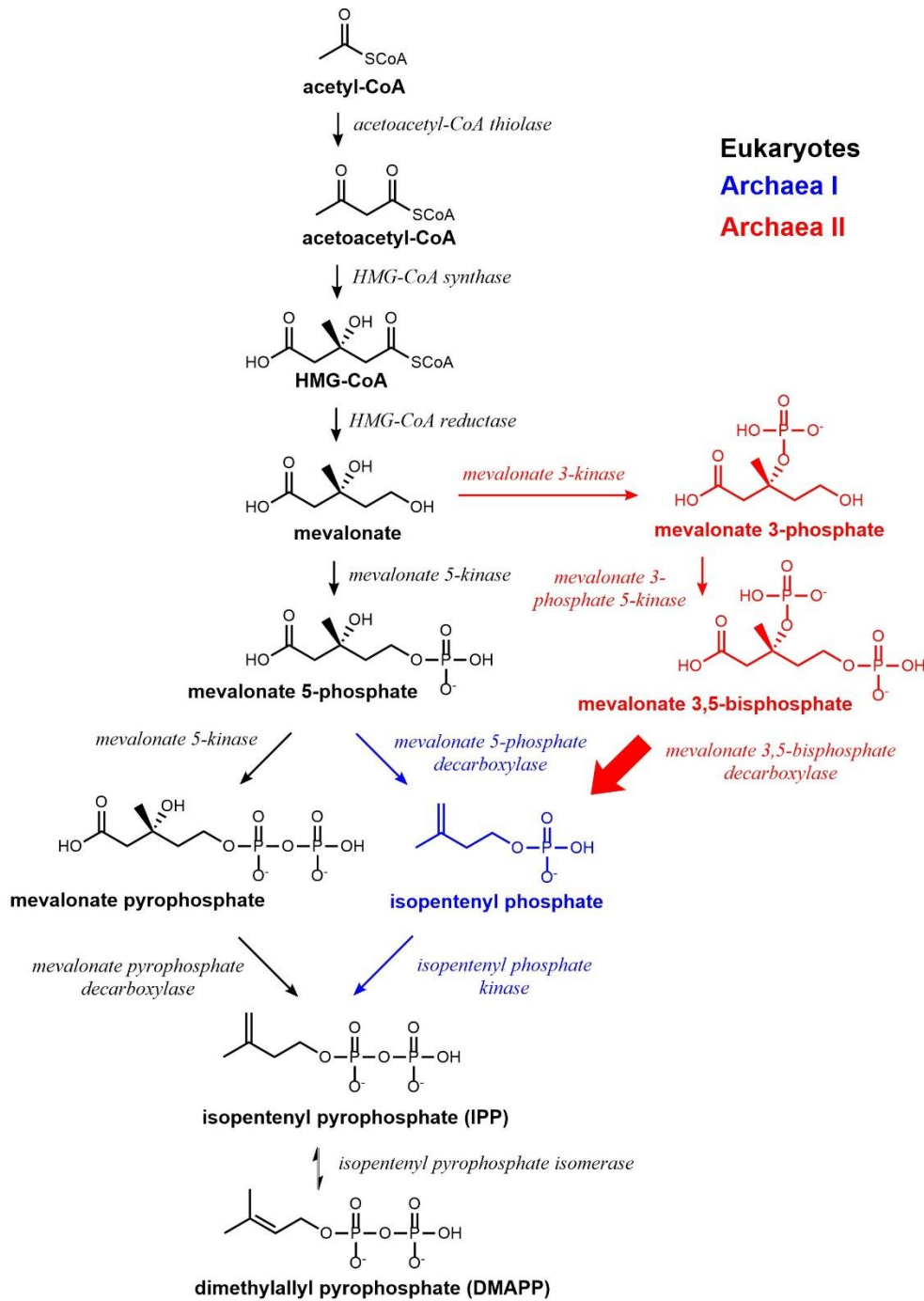
substrate concentrations from 5 mM to the detection limit (30  $\mu$ M). All characterization experiments were carried out in duplicate.

**Size exclusion chromatography.** To estimate the size of the protein complex that causes MBD activity, we mixed 100  $\mu$ L of *T. acidophilum* cell-free lysate or 100  $\mu$ L recombinant Ta0893 with 100  $\mu$ L of Biorad gel filtration standard and injected the mixture directly onto a Superdex S200 10/300 GL gel filtration column. The column was equilibrated in 50 mM sodium phosphate buffer [pH 6.5] and 500 mM NaCl at a flow rate of 0.5 mL/min. 0.5 mL fractions were collected and subjected to our GC-FID decarboxylase assay. Fitting the activity profile to a standard curve made from the internal standards allowed us to estimate the size of the MBD protein complex.

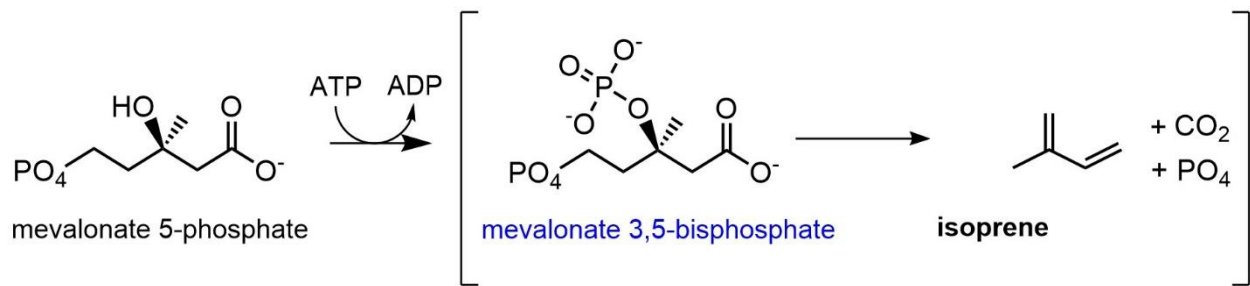
**Sequence Alignment.** Ta0893 was subjected to a standard BLASTP search<sup>39</sup>. The resulting output was aligned using clustal omega with default parameters<sup>37</sup>. Mevalonate pyrophosphate decarboxylases from the PDB were also included in the alignment. The output was rendered with Esript 3.0 and edited with Macromedia Fireworks 2004 to highlight regions of interest<sup>40</sup>.

**Analysis of MMD activity at low pH.** *Roseliflexus castenholzii* mevalonate 5-phosphate decarboxylase was assayed over a pH range of 4.3-9.1 with enzymatically generated substrates. Mevalonate 5-phosphate was produced by mixing 10 mM (R)-mevalonate with 20 mM ATP, 10 mM MgCl<sub>2</sub>, 100 mM NaCl and 20  $\mu$ g mevalonate 5-kinase (M5K) from *M. mazei*<sup>41</sup>. Mevalonate 3,5-bisphosphate was generated in an identical manner except that M5K was replaced with 10  $\mu$ g of mevalonate 3-kinase (Ta1305) and 10  $\mu$ g mevalonate 3-phosphate 5-kinase (Ta0762). After production of the substrates for 4 hr at 37°C and 60°C respectively, we mixed 150  $\mu$ L of each substrate mixture with 50  $\mu$ L of 1M acetate buffer (pH 4.3-5.6) or bis-tris

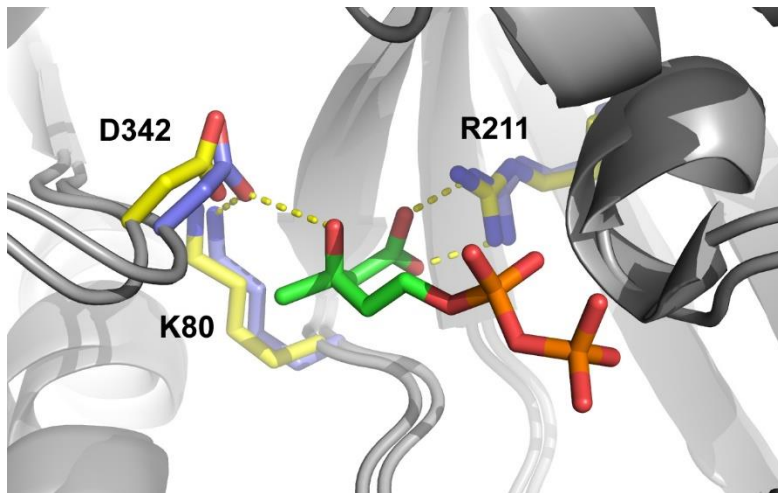
propane buffer (6.5-9.1). A micro pH electrode (Hanna Inst. 1083B) was used to confirm the final pH. 0.5 µg of *R. castenholzii* MMD was then added to each reaction. After incubation at 50°C for 1 hour, we stopped the reaction by immersing the vials in boiling water for 1 min and analyzed it via the GC-FID protocol above. This experiment was carried out in duplicate.



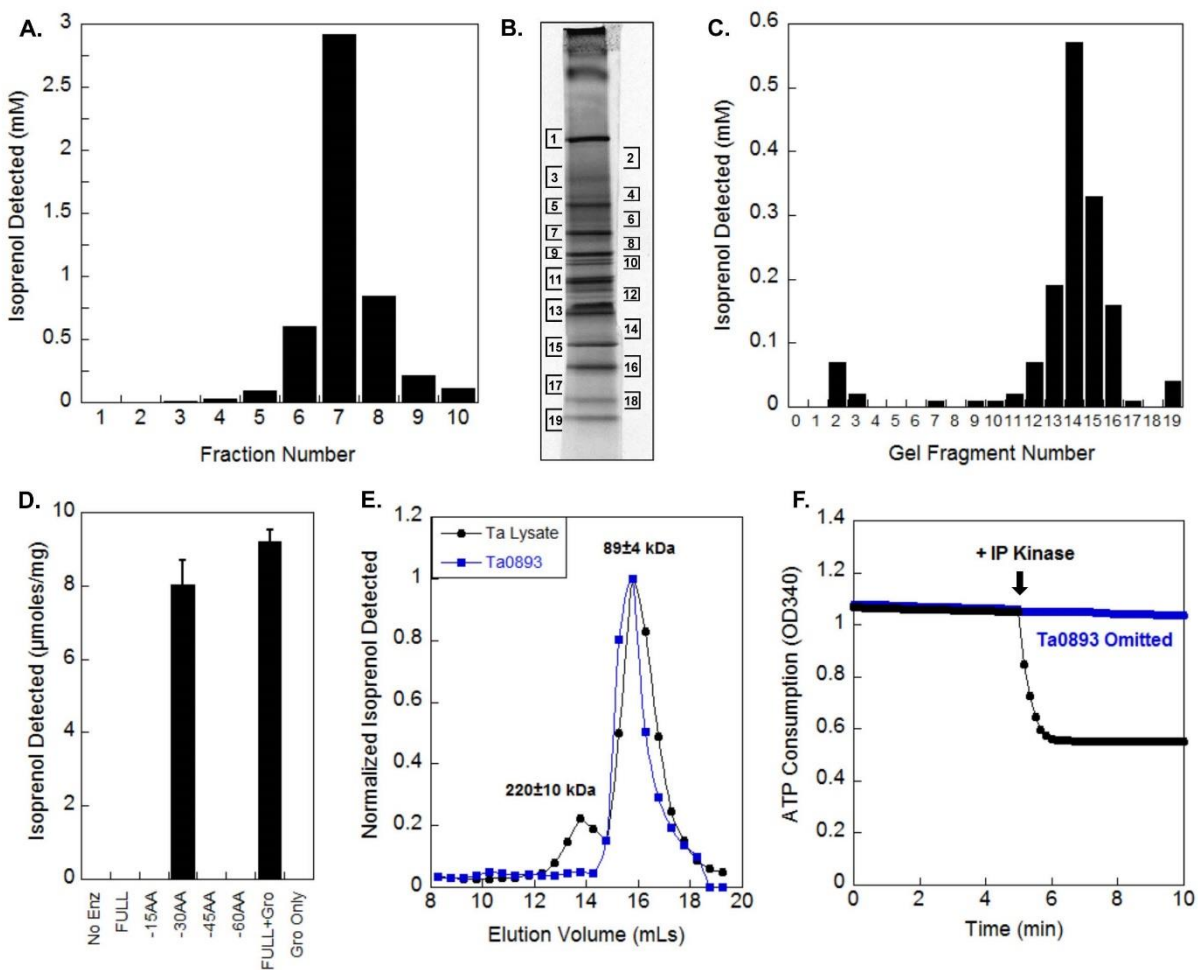
**FIGURE 2.1. Mevalonate pathways.** Eukaryotes use the pathway shown in black while most archaea use the pathway shown in blue. We have identified mevalonate 3,5-bisphosphate decarboxylase (bold arrow) which confirms a third route present in extreme acidophiles (red).



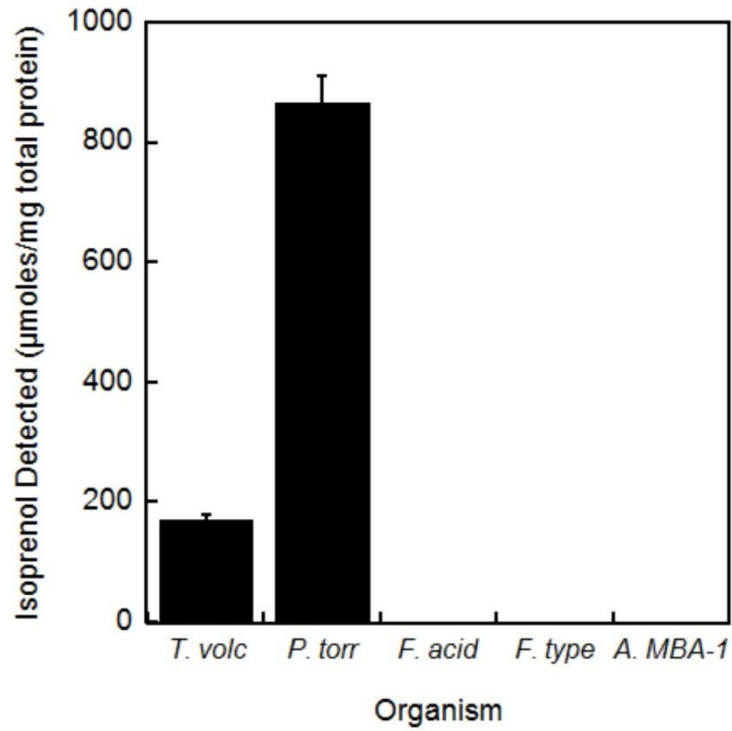
**FIGURE 2.2. The reaction scheme.** Mevalonate 5-phosphate decarboxylase of the classical archaeal pathway employs a two-step mechanism: (1) phosphorylation using ATP and (2) decarboxylation. The enzyme reported here, mevalonate 3,5-bisphosphate decarboxylase, carries out only the second step shown in brackets.



**FIGURE 2.3. Ta0893 possesses key residues for decarboxylation.** A PHYRE model of Ta0893 is overlaid with mevalonate pyrophosphate decarboxylase from *S. epidermidis* containing bound mevalonate 5-pyrophosphate (PDB: 4DU7). Numbered Ta0893 residues are highlighted in yellow and mevalonate pyrophosphate decarboxylase residues are shown in blue.

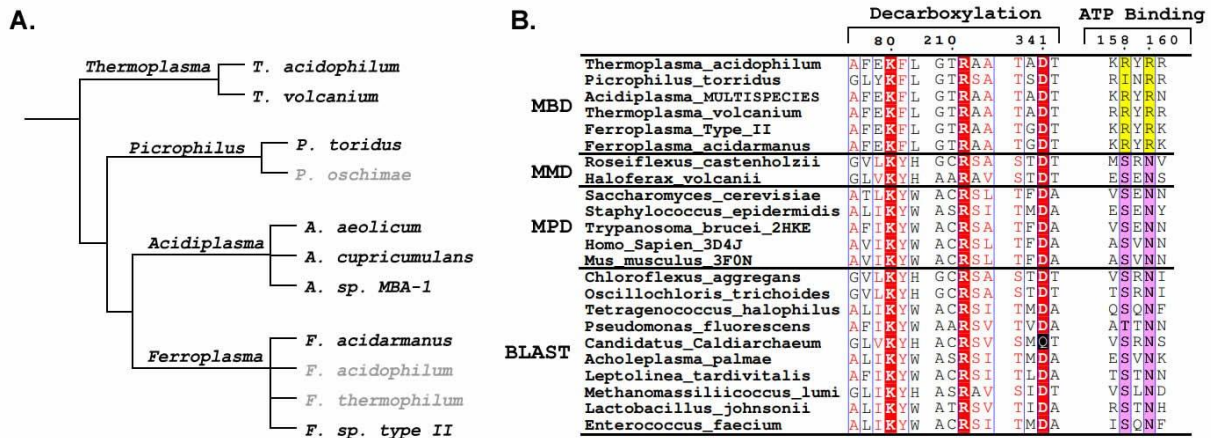


**FIGURE 2.4. Purification and identification of MBD.** (A.) MBD activity in anion exchange fractions of a *T. acidophilum* lysate. (B.) Native gel separation of fraction 7 from anion exchange. (C.) The native gel was cut into 19 fragments, pulverized, and tested for MBD activity. (D.) MBD activity in crude extracts with various N-terminal truncations of Ta0893 expressed in *E. coli* or when the full length protein was co-expressed with GroEL/ES (Gro). (E.) *T. acidophilum* lysate (black) and recombinant Ta0893 co-expressed with GroEL/ES (blue) were separated on a Superdex S200 gel filtration column. The MBD activity peaked at the same volume in both samples, consistent with the identification of Ta0893 as the native MBD. (F.) Addition of IP kinase to the product of Ta0893 causes rapid consumption of ATP (black) suggesting that Ta0893 produces IP. A negative control omitted Ta0893 (blue).

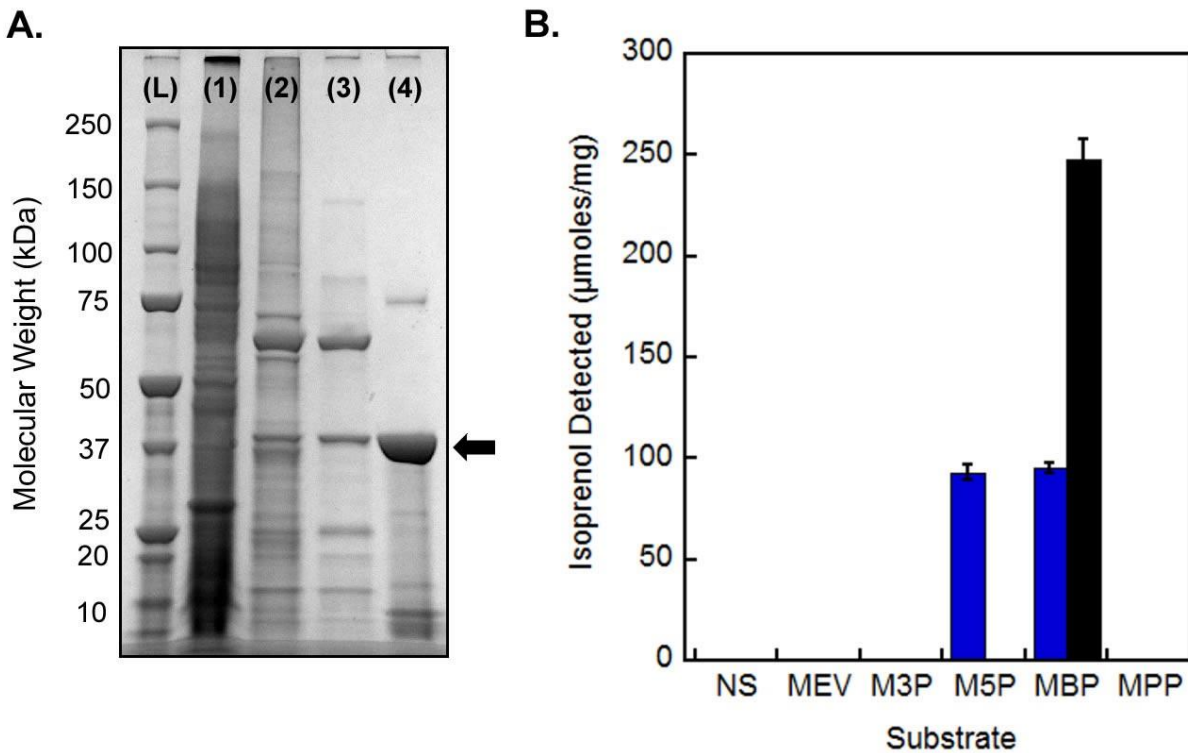


**FIGURE 2.5. Activity of Ta0893 homologs.** Five Ta0893 homologs were his-tag purified and incubated with mevalonate 3,5-bisphosphate to detect MBD activity. Homologs from *Ferroplasma* and *Acidiplasma* formed inclusion bodies under all expression conditions.

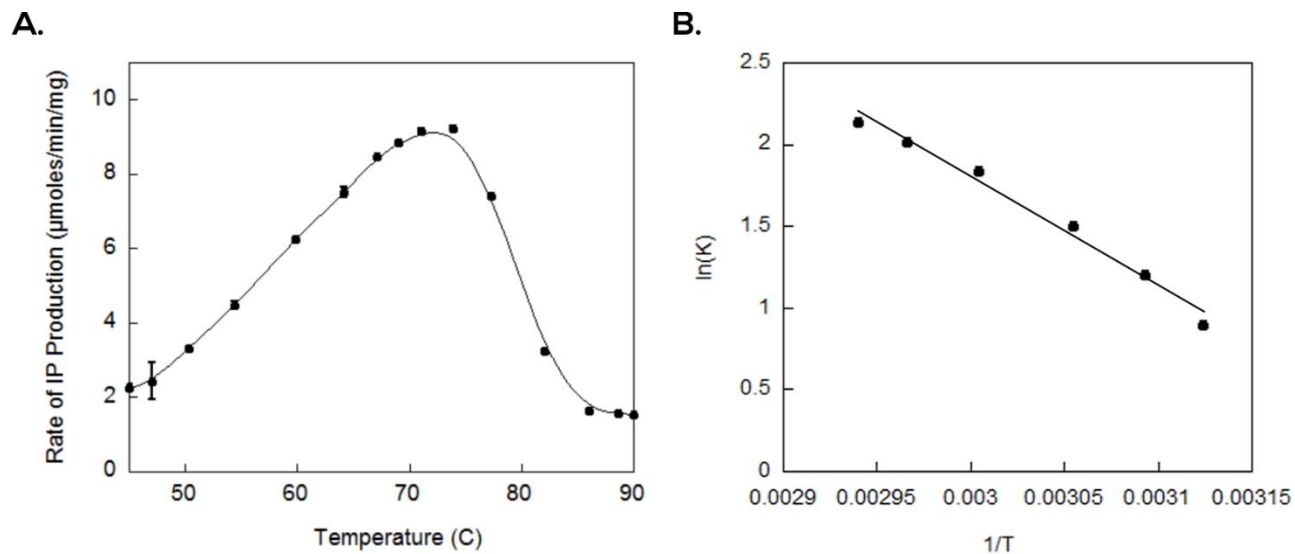




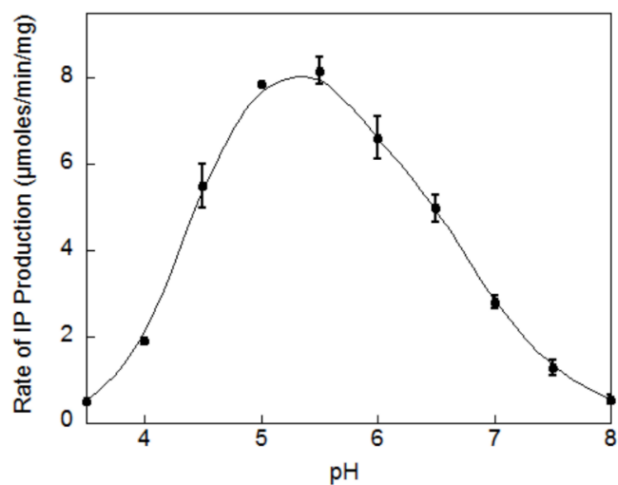
**FIGURE 2.6. Phylogeny and Sequence Alignment.** (A.) A phylogenetic tree of thermoplasmatales based on 16S rRNA. Organisms with no available DNA sequences are shown in grey. Environmental samples were excluded for clarity. (B.) Mevalonate 3,5-bisphosphate decarboxylase homologs representing all 8 sequenced thermoplasmatales were aligned with mevalonate 5-phosphate decarboxylases (MMD), mevalonate 5-pyrophosphate decarboxylases (MPD), and the top Ta0893 BLASTp hits after thermoplasmatales. MBDs retain the Asp/Lys/Arg catalytic triad required for decarboxylation (red), but are missing both ATP binding residues (purple).



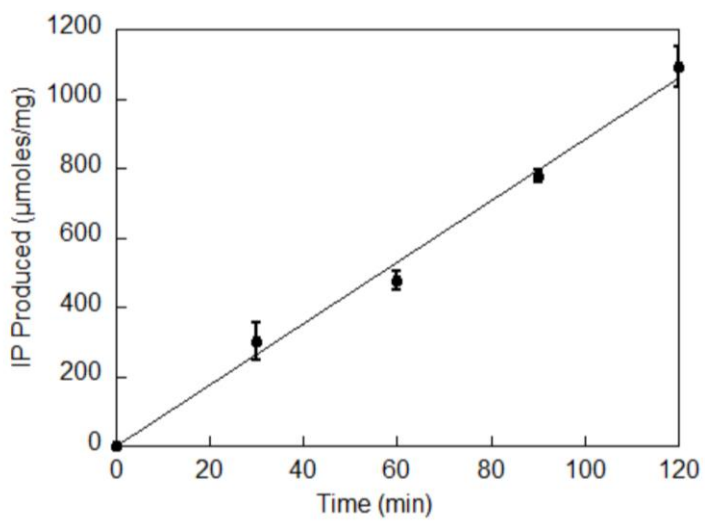
**FIGURE 2.7. Purification of *P. torridus* MBD and Substrate Specificity.** (A.) An SDS-PAGE gel stained with Coomassie blue. Lanes: (L) ladder, (1) crude *E. coli* lysate, (2) Ni-NTA eluate, (3) supernatant after heating for 2 hrs at 60°C, (4) flow through from a Q HP anion exchange column. Yield was 1 mg purified *P. torridus* MBD from 32 L of culture. (B.) *P. torridus* MBD (black) and *R. castenholzii* mevalonate 5-phosphate decarboxylase (blue) were assayed via our GC-FID decarboxylase assay after incubation with the following substrates for 1 hour: (MEV) mevalonate, (M3P) mevalonate 3-phosphate, (M5P) mevalonate 5-phosphate, (MBP) mevalonate 3,5-bisphosphate, (MPP) mevalonate 5-pyrophosphate and (NS) no substrate.



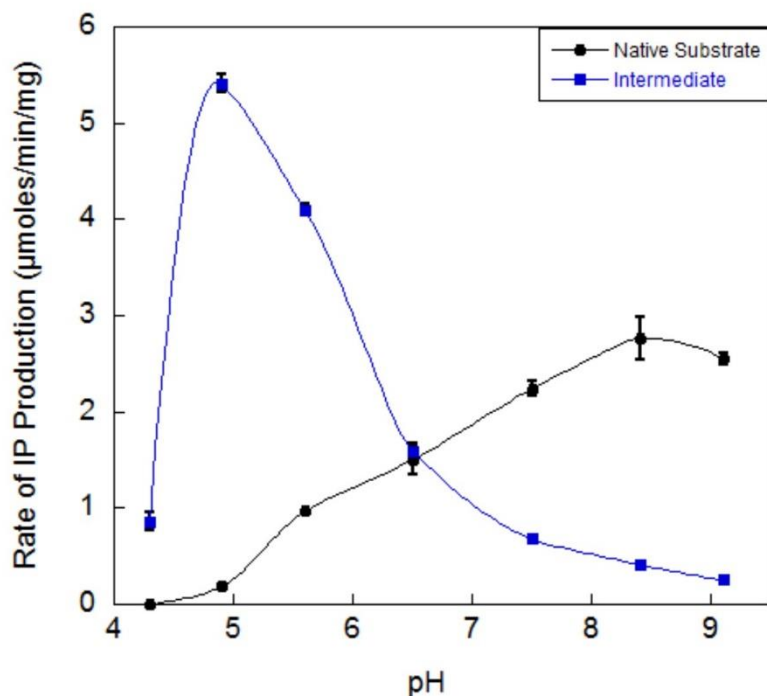
**Figure 2.8. Temperature dependence of *P. torridus* MBD activity.** (A.) *P. torridus* MBD was assayed from 45-90°C with the pH held constant at 5.5. (B.) An Arrhenius plot was generated from the data points from 47-67°C by plotting ln(K) vs. 1/T.



**Figure 2.9. Optimum pH of *P. torridus* MBD.** *P. torridus* MBD was assayed from pH 3.5 – 8 in 0.5 unit increments at 60°C.



**Figure 2.10. Time course of IP production by *P. torridus* MBD.** *P. torridus* MBD was incubated with mevalonate 3,5-bisphosphate at 60°C and pH 5.5. The reaction was stopped at 0, 30, 60, 90, and 120 min. The IP produced by MBD was quantified via our GC-FID assay.



**FIGURE 2.11.** pH dependence of activity of *R. castenholzii* MMD with native substrate (mevalonate 5-phosphate) or intermediate (mevalonate 3,5-bisphosphate). Mevalonate 5-phosphate decarboxylase was assayed at 50 °C over a pH range of 4.3-9.1 with 10 mM mevalonate 5-phosphate (black) or 10 mM mevalonate 3,5-bisphosphate (blue).

Hits	Prot. Mass (kDa)	Peptides	Protein Annotation	NCBI Accession	Rel. Abund.
1	37.7	107	Deoxyhypusine Synthase (Ta0356)	WP_010900784	83.5%
2	44.4	9	S-adenosylmethionine Synth. (Ta0059)	WP_010900487	8.4%
3	46.4	4	Hypothetical Protein (Ta0893)	WP_010901303	0.4%
4	51.1	4	Glutamine Synthetase (Ta1498)	WP_010901897	0.6%
5	55.0	1	Hypothetical Protein (Ta0203)	WP_010900630	0.2%
6	45.0	1	Hypothetical Protein (Ta0204)	WP_010900631	0.1%
7	35.5	1	DNA Repair Protein RadA (Ta1104)	WP_010901514	0.0%

**TABLE 2.1.** Proteins identified in gel fragments with MBD activity. NanoLC/MS/MS identified seven proteins which were present in the entire region of MBD activity (gel fragments 13-16, Fig. 4B). The relative abundance is reported for fragment 14.

## 2.6 REFERENCES

1. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 1092–1101 (2001).
2. Kashefi, K. & Lovley, D. R. Extending the Upper Temperature Limit for Life. *Science* **301**, 934–934 (2003).
3. Fendrihan, S. *et al.* Extremely halophilic archaea and the issue of long-term microbial survival. *Rev. Environ. Sci. Biotechnol. Online* **5**, 203–218 (2006).
4. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1007–1022 (2006).
5. Baker-Austin, C. & Dopson, M. Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* **15**, 165–171 (2007).
6. Schleper, C., Piihler, G., Kuhlmoorgen, B. & Zillig, W. Life at extremely low pH. *Nature* **375**, 741–742 (1995).
7. Golyshina, O. V. *et al.* *Acidiplasma aeolicum* gen. nov., sp. nov., a euryarchaeon of the family Ferroplasmaceae isolated from a hydrothermal pool, and transfer of *Ferroplasma cupricumulans* to *Acidiplasma cupricumulans* comb. nov. *Int. J. Syst. Evol. Microbiol.* **59**, 2815–2823 (2009).
8. Edwards, K. J., Bond, P. L., Gihring, T. M. & Banfield, J. F. An Archaeal Iron-Oxidizing Extreme Acidophile Important in Acid Mine Drainage. *Science* **287**, 1796–1799 (2000).
9. Ruepp, A. *et al.* The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513 (2000).
10. Stern, J., Freisleben, H. J., Janku, S. & Ring, K. Black lipid membranes of tetraether lipids from *Thermoplasma acidophilum*. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* **1128**, 227–236 (1992).

11. Jacquemet, A., Barbeau, J., Lemiègre, L. & Benvegna, T. Archaeal tetraether bipolar lipids: Structures, functions and applications. *Biochimie* **91**, 711–717 (2009).
12. Holstein, S. A. & Hohl, R. J. Isoprenoids: remarkable diversity of form and function. *Lipids* **39**, 293–309 (2004).
13. Lombard, J. & Moreira, D. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol.* **28**, 87–99 (2011).
14. Dellas, N., Thomas, S. T., Manning, G. & Noel, J. P. Discovery of a metabolic alternative to the classical mevalonate pathway. *eLife* **2**, e00672 (2013).
15. VanNice, J. C. *et al.* Identification in *Haloferax volcanii* of Phosphomevalonate Decarboxylase and Isopentenyl Phosphate Kinase as Catalysts of the Terminal Enzyme Reactions in an Archaeal Alternate Mevalonate Pathway. *J. Bacteriol.* **196**, 1055–1063 (2014).
16. Jabalquinto, A. M., Alvear, M. & Cardemil, E. Physiological aspects and mechanism of action of mevalonate 5-diphosphate decarboxylase. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **90**, 671–677 (1988).
17. Vinokur, J. M., Korman, T. P., Cao, Z. & Bowie, J. U. Evidence of a novel mevalonate pathway in archaea. *Biochemistry* **53**, 4161–4168 (2014).
18. Azami, Y. *et al.* (R)-Mevalonate 3-Phosphate Is an Intermediate of the Mevalonate Pathway in *Thermoplasma acidophilum*. *J. Biol. Chem.* **289**, 15957–15967 (2014).
19. Bonanno, J. B. *et al.* Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis. *Proc. Natl. Acad. Sci.* **98**, 12896–12901 (2001).
20. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

21. Krepkii, D. & Miziorko, H. M. Identification of active site residues in mevalonate diphosphate decarboxylase: Implications for a family of phosphotransferases. *Protein Sci.* **13**, 1875–1881 (2004).
22. Marston, F. A. & Hartley, D. L. Solubilization of protein aggregates. *Methods Enzymol.* **182**, 264–276 (1990).
23. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* **7**, 32 (2007).
24. Chen, M. & Poulter, C. D. Characterization of Thermophilic Archaeal Isopentenyl Phosphate Kinases. *Biochemistry* **49**, 207–217 (2010).
25. Voynova, N. E. *et al.* Human mevalonate diphosphate decarboxylase: characterization, investigation of the mevalonate diphosphate binding site, and crystal structure. *Arch. Biochem. Biophys.* **480**, 58–67 (2008).
26. Byres, E., Alphey, M. S., Smith, T. K. & Hunter, W. N. Crystal structures of *Trypanosoma brucei* and *Staphylococcus aureus* mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity. *J. Mol. Biol.* **371**, 540–553 (2007).
27. Barta, M. L. *et al.* Crystal structures of *Staphylococcus epidermidis* mevalonate diphosphate decarboxylase bound to inhibitory analogs reveal new insight into substrate binding and catalysis. *J. Biol. Chem.* **286**, 23900–23910 (2011).
28. Macalady, J. L. *et al.* Tetraether-linked membrane monolayers in *Ferroplasma* spp: a key to survival in acid. *Extremophiles* **8**, 411–419 (2004).



29. van de Vossenberg, null, Driessen, A. J., Zillig, W. & Konings, W. N. Bioenergetics and cytoplasmic membrane stability of the extremely acidophilic, thermophilic archaeon *Picrophilus oshimae*. *Extrem. Life Extreme Cond.* **2**, 67–74 (1998).
30. Searcy, D. G. *Thermoplasma acidophilum*: intracellular pH and potassium concentration. *Biochim. Biophys. Acta* **451**, 278–286 (1976).
31. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742 (2001).
32. Barta, M. L., McWhorter, W. J., Miziorko, H. M. & Geisbrecht, B. V. Structural basis for nucleotide binding and reaction catalysis in mevalonate diphosphate decarboxylase. *Biochemistry* **51**, 5611–5621 (2012).
33. Lindberg, M., Yuan, C., Dewaard, A. & Bloch, K. On the mechanism of formation of isopentenylpyrophosphate. *Biochemistry* **1**, 182–188 (1962).
34. Iyengar, R., Cardemil, E. & Frey, P. A. Mevalonate-5-diphosphate decarboxylase: stereochemical course of ATP-dependent phosphorylation of mevalonate 5-diphosphate. *Biochemistry* **25**, 4693–4698 (1986).
35. Qiu, Y., Gao, J., Guo, F., Qiao, Y. & Li, D. Mutation and inhibition studies of mevalonate 5-diphosphate decarboxylase. *Bioorg. Med. Chem. Lett.* **17**, 6164–6168 (2007).
36. Vinokur, J. M. *et al.* Structural analysis of mevalonate-3-kinase provides insight into the mechanisms of isoprenoid pathway decarboxylases. *Protein Sci.* **24**, 212–220 (2015).
37. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
38. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

39. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

40. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

41. Primak, Y. A. *et al.* Characterization of a feedback-resistant mevalonate kinase from the archaeon *Methanosarcina mazei*. *Appl. Environ. Microbiol.* **77**, 7772–7778 (2011).

## CHAPTER 3

### STRUCTURAL ANALYSIS OF MEVALONATE 3-KINASE

#### 3.1 ABSTRACT

In animals, cholesterol is made from 5-carbon building blocks produced by the mevalonate pathway. Drugs that inhibit the mevalonate pathway such as atorvastatin (Lipitor) have led to successful treatments for high cholesterol in humans. Another potential target for the inhibition of cholesterol synthesis is mevalonate diphosphate decarboxylase (MDD), which catalyzes the phosphorylation of (R)-mevalonate diphosphate, followed by decarboxylation to yield isopentenyl pyrophosphate. We recently discovered an MDD homolog, mevalonate-3-kinase (M3K) from *Thermoplasma acidophilum*, which catalyzes the identical phosphorylation of (R)-mevalonate, but without concomitant decarboxylation. Thus, M3K catalyzes half the reaction of the decarboxylase, allowing us to separate features of the active site that are required for decarboxylation from features required for phosphorylation. Here we determine the crystal structure of M3K in the *apo* form, and with bound substrates, and compare it to MDD structures. Structural and mutagenic analysis reveals modifications that allow M3K to bind mevalonate rather than mevalonate diphosphate. Comparison to homologous MDD structures show that both enzymes employ analogous Arg or Lys residues to catalyze phosphate transfer. However, an invariant active site Asp/Lys pair of MDD previously thought to play a role in phosphorylation is missing in M3K with no functional replacement. Thus, we suggest that the invariant Asp/Lys pair in MDD may be critical for decarboxylation rather than phosphorylation.

## 3.2 INTRODUCTION

Cholesterol, chlorophyll, and vitamin A are just three examples of over 25,000 biomolecules that make up a diverse class of chemicals called isoprenoids.<sup>1-2</sup> In eukaryotes and some bacteria, isoprenoids are synthesized via the mevalonate pathway, which yields the universal isoprenoid precursor isopentenyl pyrophosphate (IPP).<sup>3</sup> Enzymes of the mevalonate pathway have been heavily investigated for the development of cholesterol lowering drugs.<sup>4-7</sup> Indeed one of the best-selling drugs of all time, atorvastatin (Lipitor), works by inhibiting HMG-CoA reductase, thereby lowering the production of mevalonate.<sup>8</sup> Inhibitors of other mevalonate pathway enzymes such as mevalonate diphosphate decarboxylase (MDD) could be useful for lowering cholesterol<sup>9-11</sup> and may also be useful as antimicrobial agents since many pathogens require the mevalonate pathway.<sup>12,13,14</sup>

MDD converts (R)-mevalonate 5-diphosphate (MVAPP) and ATP to isopentenyl pyrophosphate (IPP), ADP, PO<sub>4</sub>, and CO<sub>2</sub>.<sup>15</sup> As shown in Fig. 3.1A, the mechanism can be conceptually divided into two stages, phosphorylation and decarboxylation. The phosphorylation stage involves transfer of the  $\gamma$ -phosphate of ATP to the 3-OH position of MVAPP, generating the intermediate, 3-phospho-mevalonate 5-diphosphate (3P-MVAPP) (Fig 3.1A).<sup>16-18</sup> This intermediate was originally thought to spontaneously decompose to IPP, CO<sub>2</sub> and PO<sub>4</sub> due to the instability of the sterically hindered tertiary phosphate,<sup>14,19</sup> but we recently reported the analogous tertiary phosphorylated intermediates, mevalonate 3-phosphate and mevalonate 3,5-bisphosphate are both stable.<sup>20</sup> Moreover, a fluorinated version of 3P-MVAPP was reported to be stable after isolation from a quenched MDD reaction.<sup>21</sup> It is therefore likely that decarboxylation is not spontaneous so that MDD must also catalyze the decarboxylation step.

We were recently presented with a unique opportunity to probe the mechanism of MDD, by our discovery, along with another group, of a novel mevalonate pathway enzyme, mevalonate-3-kinase (M3K,) in the archeon *Thermoplasma acidophilum*.<sup>20, 22</sup> M3K catalyzes the ATP dependent phosphorylation of (R)-mevalonate to produce (R)-mevalonate 3-phosphate, but without concomitant decarboxylation (Fig. 3.1B).<sup>20</sup> Instead, mevalonate 3-phosphate is released as a stable metabolite. A second enzyme phosphorylates the 5-OH position to yield mevalonate 3,5-bisphosphate, which has been proposed to ultimately be decarboxylated and phosphorylated to the universal isoprenoid building block IPP, although the decarboxylase has not yet been identified.<sup>20, 22</sup> While M3K does not decarboxylate, it does phosphorylate mevalonate in the same position as the homologous decarboxylases, and shares 17–21% sequence identity with MDDs from *H. sapiens*, *S. epidermidis*, and *S. cerevisiae*. Thus, M3K can be considered a defective decarboxylase that catalyzes exactly half the overall reaction of MDD.

To provide insight into the mechanism of MDD, we solved the crystal structure of M3K in the *apo* form and with bound substrates and product, allowing us to perform structural comparisons to MDD, for which nineteen structures have been reported (*H. sapiens*, *M. musculus*, *T. brucei*, *S. cerevisiae*, *L. pneumophila*, *S. aureus*, *S. pyogenes*, *S. epidermidis*).<sup>14, 19, 23-25</sup> Analysis of the M3K active site identified two residues involved in catalyzing phosphate transfer, Arg185 and Ser105, and also identified how M3K binds mevalonate, while excluding mevalonate 5-phosphate and mevalonate 5-diphosphate. Comparison to MDD leads us to propose a model in which both enzymes use similar residues (Arg185/Ser105 in M3K and Lys188 in MDD) to catalyze phosphorylation and we suggest that the invariant Asp/Lys pair unique to MDD plays a critical role in the decarboxylation step rather than phosphorylation.

### 3.3 RESULTS AND DISCUSSION

**Overall Fold.** M3K crystalized in space group C2 with two protein molecules in the asymmetric unit (Fig. 3.2A) that are linked by a disulfide bond between Cys 43 from each monomer. The enzyme purified primarily as a dimer, and became monomeric in the presence of 1 mM  $\beta$ -mercaptoethanol (Fig. 3.3). The monomeric and dimeric forms had equal activity under optimal conditions (pH 8.5 and 55°C). Some archaea are known to possess disulfide bonded proteins in the cytoplasm,<sup>26-28</sup> and the M3K dimer is slightly more stable (Fig. 3.4), suggesting that the natural form of the protein could be a dimer.

We compared our M3K structures to the structure of MDD from *S. epidermidis* (PDB: 4DU7), since the *S. epidermidis* structure contained a bound mevalonate diphosphate (MVAPP), marking the active site.<sup>25</sup> When the M3K *apo* structure is overlaid with the MDD structure from *S. epidermidis*, they align with an overall backbone RMSD of 2.40 Å, despite sharing only 19% sequence identity (Fig. 3.2B). For comparison, the backbone atoms of *S. epidermidis* MDD compared to those of the MDD enzyme from *S. cerevisiae* have an overall RMSD of 2.15 Å. The high structural similarity places M3K into the GHMP kinase family (Galactokinase, Homoserine kinase, Mevalonate kinase, and Phosphomevalonate kinase).<sup>29</sup> All members of this family share a highly conserved overall fold that contains a centrally located cleft with the active site located at the base of the cleft (Fig. 3.2C).<sup>30</sup>

**The Mevalonate Binding Pocket.** To obtain insight into substrate binding and catalysis, we determined a structure of M3K bound to (R)-mevalonate, and a structure of M3K bound to both (R)-mevalonate 3-phosphate and ADP. Mevalonate was found in both proteins of the asymmetric unit (Fig. 3.5A), however mevalonate 3-phosphate and ADP were only observed

in chain A (Fig. 3.5B). Chain B contained mevalonate 3-phosphate in the same position as in chain A, but the ADP was replaced with a density modeled as a sulfate ion.

The structures revealed no major conformational shifts when binding to substrate or products. Global backbone RMSD for all pairwise comparisons of the M3K structures (6 molecules) did not exceed 0.38 Å. The pair with the largest deviation was chain A of the substrate complex and chain B of the product complex. When we restricted analysis to the active site of this pair (residues within 5 Å of bound mevalonate 3-phosphate), we found an RMSD of only 0.31 Å (116 atoms). This high degree of structural conservation is also seen in the active sites of all known MDD structures.<sup>14</sup> These results suggest that M3K does not require major conformational shifts for catalysis.

Mevalonate and mevalonate 3-phosphate bind in the center of a deep cleft located in the middle of the enzyme (Fig. 3.2C). This is the same location that MVAPP binds in MDD.<sup>25</sup> The orientation of mevalonate in M3K is rotated by 31° relative to MVAPP (Fig. 3.5C). This is due to a 1.8 Å shift in the positioning of Arg144 (Fig. 3.6), which binds the carboxylate moiety of mevalonate in both enzymes. Mevalonate is not translated with respect to MVAPP, however, as the C3 carbons of both species are only 0.7 Å apart.

MDD binds the pyrophosphate moiety of MVAPP primarily through four invariant serine residues<sup>25</sup> (Fig. 3.7A). In M3K, the pyrophosphate does not fit due to the protrusion of Glu140 into the middle of the active site. Glu140 replaces an invariant glycine at the analogous position in MDD (Fig 3.6) and provides a polar contact directly to the 5-OH tail of mevalonate. Mevalonate's position in the center of the active site cavity is stabilized by a web of polar

contacts made by His192, Arg28, and Asp189 (Fig. 3.7B). Additionally, an invariant serine (Ser141) in both M3K and MDD can provide an additional hydrogen bond to the 5-OH tail of mevalonate. The steric hindrance introduced by Glu140 explains why mevalonate 5-phosphate and mevalonate 5-diphosphate are not phosphorylated by M3K.<sup>20</sup>

**Residues important for phosphate transfer.** Both Arg185 and Ser105 in M3K make polar contacts to the transferred phosphate and the  $\beta$ -phosphate of ADP (Fig. 3.8A). Arg185 is oriented towards the broken P-O bond, 3.4 Å from the  $\beta$ -phosphate of ADP and 3.3 Å from the transferred phosphate, which is close enough to provide important stabilizing interactions in the transition state.

Arg185 of M3K aligns with an invariant lysine (Lys188) in MDD (Fig. 3.6), both of which are located in a loop region. Surprisingly, this loop is disordered in all wild type MDD structures from *S. epidermidis*, however a D283A mutant of *S. epidermidis*<sup>25</sup> captured this flexible loop in a secondary conformation, where Lys188 of MDD spatially overlays with Arg185 from M3K (Fig. 3.8B). Lys188 is critical for catalysis as a K188A mutant in MDD abolishes all activity.<sup>31</sup> We prepared a R185A mutant in M3K, which also resulted in no detectable activity. To demonstrate an analogous relationship between R185 of M3K and K188 of MDD, we prepared a R185K mutant of M3K, which was active (Table 3.1). Thus, it is likely that the invariant Lys188 in MDD plays the same role as Arg185 in M3K, directly stabilizing a phospho-transfer transition state. All Michaelis-Menten plots corresponding to the mutants in Table 3.1 are shown in figures 3.9-3.13.

Ser105 of M3K is also oriented towards the broken P-O bond. Ser105 is located in a loop region and spatially aligns between conserved Leu and Ala residues in MDD (Fig. 3.6).



Ser105 is 2.4 Å from the transferred phosphate and 3.4 Å from the β-phosphate of ADP. A S105A mutant resulted in a 4.6 fold decrease in  $k_{cat}$ , with only a marginal increase in  $K_m$  (Table 3.1), consistent with a modest role in stabilizing the transition state in phosphate transfer. The transition state is likely further stabilized by hydrogen bonds between the phosphates and amide groups from the protein backbone (Fig. 3.8A).

**Residues involved in decarboxylation.** All MDDs have three invariant residues within the active site (Arg144, Asp283, and Lys17 in *S. epidermidis* MDD).<sup>25</sup> Arg144 positions the substrate in the active site (Fig. 3.5C), and the Asp283/Lys17 pair has been proposed to directly deprotonate the 3-OH of MVAPP, triggering phosphorylation.<sup>15-17</sup> The catalytic role for the invariant Asp/Lys pair of MDD is supported by its active site position and previous mechanistic studies, which showed that a D283A mutant yields a 10<sup>5</sup> fold reduction in activity,<sup>32</sup> and a K17A mutant yields no detectable activity.<sup>31</sup>

In M3K, Arg144 is conserved for positioning the mevalonate moiety, but the invariant Asp/Lys pair in MDD is replaced by Leu18 and Thr275 in M3K, respectively (Fig.3.14). To test the role of Leu18 and Thr275, we characterized mutants L18A and T275A. Both L18A and T275A mutants had no significant effect on  $k_{cat}$  while slightly increasing  $K_m$ . Thus, these residues are not involved in kinase activity (Table 3.1).

Some members of the GHMP kinase family, such as homoserine kinase (PDB: 1H72) are known to facilitate phosphate transfer without the presence of an active site Asp/Lys pair,<sup>33</sup> yet others such as mevalonate 5-kinase (PDB: 2HUF) and phosphomevalonate kinase (PDB: 3GON) have an active site Asp/Lys pair.<sup>34, 35</sup> These observations have led to two proposed

mechanisms of phosphate transfer, one involving simple transition state stabilization, and a second involving Asp/Lys mediated deprotonation, followed by nucleophilic attack on the  $\gamma$ -phosphate of ATP.<sup>14</sup> The current data raises a third possibility.

We propose M3K and MDD both phosphorylate by stabilizing the phospho-transfer transition state (M3K via Arg185/Ser105, MDD via Lys188), leaving the invariant Asp/Lys pair, unique to MDD, to play a role in the decarboxylation step. Testing this model has proven difficult since mutating D283 in MDD to Ala, Glu, Thr, Val or Asn results in no detectable kinase activity. This was explained by a crystal structure of a D283A mutant from Barta *et al.*,<sup>25</sup> in which the mutation caused catastrophic changes in MVAPP binding orientation. Barta *et al.* suggest D283 may play a critical role in positioning the substrate. We can rule out the pyrophosphate moiety as the differentiating factor needed for decarboxylation as a recently characterized enzyme, mevalonate monophosphate decarboxylase from *Haloferax volcanii*<sup>36</sup> and *Roseiflexus castenholzii*<sup>37</sup> were shown to catalyze the identical reaction as MDD, but with mevalonate 5-phosphate. A PHYRE (Protein Homology/Analogy Recognition Engine)<sup>38</sup> model of both mevalonate monophosphate decarboxylases showed retention of the Asp/Lys pair in the same position as MDD, consistent with a role in decarboxylation (Figure 3.15).

### 3.4 CONCLUSIONS

The crystal structure of mevalonate-3-kinase provides new insight into the mechanism of mevalonate diphosphate decarboxylase. Despite sharing nearly identical overall folds, important active site differences can be identified. Glu140 in the center of the M3K active site is responsible for binding mevalonate while excluding mevalonate 5-diphosphate, Arg185/Ser105 catalyze phosphate transfer, and an invariant Asp/Lys pair previously thought to be responsible

for phosphorylation in MDD, is missing in M3K and replaced by non-essential Thr275/Leu18. These findings lead us to propose a model in which M3K and MDD both phosphorylate by stabilizing a phospho-transfer transition state (M3K via Arg185/Ser105, MDD via Lys188), suggesting the invariant Asp/Lys pair unique to MDD may be critical for the decarboxylation step rather than phosphorylation.

### 3.5 METHODS

**M3K expression, purification and assays.** M3K expression, purification and enzyme assays were all conducted as reported previously.<sup>20</sup> In summary, the *ta1305* gene (encoding mevalonate-3-kinase) was cloned into a pET28 plasmid, providing an N-terminal His tag. The plasmid was transformed into *E. coli* BL21 Gold (DE3), cells were grown in LB-media supplemented with 50 mg/L kanamycin, and protein production was induced with 0.5 mM IPTG for 20 hours at 37 °C. M3K was purified by Ni-NTA affinity chromatography followed by elution with 250 mM imidazole. Further purification was accomplished by gel filtration on a Superdex S200 column in 50 mM Tris-HCl pH 7.5 and 100 mM NaCl.

**Crystal Growth Conditions.** A stock of M3K was prepared by concentrating the protein in 7 mM Tris-HCl pH 7.5 and 45 mM NaCl to 7.0 mg/ml using a 30 kDa cutoff Amicon Ultra-15 Centrifugal Filter (Millipore) and stored at 4°C. Some precipitation of the stock was observed upon cooling, but it clarified upon incubation at room temperature for 5 min. Crystallization conditions were screened using the hanging drop method<sup>39</sup> in 96 well plates. For each condition, three 210 nL hanging drops with varying protein stock to reservoir ratios (2:1, 1:1, 1:2) were prepared using a TTP LabTech Mosquito nanoliter-pipetting robot in the UCLA Macromolecular Crystallization Facility. Large single crystals grew after 2 days in ProComplex

condition 68, (Quiagen Cat No. 135468A) which consists of 0.1 M sodium acetate pH 5.0 and 1.0 M ammonium sulfate (2:1 protein:reservoir ratio). Crystals were cryo-protected by a quick soak in a solution consisting of 65% reservoir, 35% (v/v) glycerol, then flash frozen in a cryogenic nitrogen stream and maintained at 100 K for data collection.

**Binding of Substrate and Products.** Substrate complex crystals were obtained by soaking substrates into the *apo* crystals that were prepared as described above. For the mevalonate soaks, a single crystal was removed from the hanging drop and placed directly into 8  $\mu$ L of 65 mM (R)-mevalonate for 2 hours (no mother liquor). During the soak, the crystal morphology remained intact without cracking or dissolving. The crystal was removed from the mevalonate solution and quickly cryo-protected with 20% glycerol, 80% 65 mM (R)-mevalonate solution, and placed into a cryogenic nitrogen stream for data collection. To bind products, a mixture of 25 mM (R)-mevalonate 3-phosphate and 25 mM ADP was prepared enzymatically in a 1 mL reaction consisting of 25 mM (R)-mevalonate, 25 mM ATP, 1 mM  $MgCl_2$  and 36  $\mu$ g mevalonate-3-kinase. The reaction was incubated at 42°C for 1 hr, then cooled to room temperature before attempting crystal soaks. 2  $\mu$ L of the product solution was mixed with 8  $\mu$ L of mother liquor so the final concentration of products was approximately 5 mM. A crystal soaked in this 5 mM solution for 40 min was rapidly transferred into a solution of 35% glycerol, 65% mother liquor for 5 seconds and placed into the cryogenic nitrogen stream for data collection.

**X-Ray Data Collection.** X-ray diffraction data for the *apo* (2.1 Å) and substrate (2.0 Å) crystals were collected with a Rigaku FR-E rotating anode X-ray source, using  $CuK\alpha$  radiation ( $\lambda = 1.5418$  Å) and an R-AXIS HTC imaging plate detector. Data for crystals soaked with products were first collected in-house, and then flash frozen and shipped to the Advanced Photon Source

(Argonne National Laboratory) for data collection on APS-NECAT beamline 24-ID-C with a DECTRIS-PILATUS 6 M detector. A crystal soaked with products diffracted to 2.3 Å on the APS beamline. Reduction and scaling of data were performed using XDS/XSCALE.<sup>40</sup> All data sets were consistent with space group C2 (Table 3.2).

**Structure Determination and Refinement.** The *apo* structure was determined using the automated molecular replacement pipeline, MrBUMP.<sup>41</sup> The path that led to a successful solution employed PHASER<sup>42</sup> for molecular replacement and a search model prepared by CHAINSAW<sup>43</sup> from PDB entry 1FI4, mevalonate 5-diphosphate decarboxylase.<sup>24</sup> Trimming of the 1FI4 coordinate set by CHAINSAW was important for success as the full coordinate set did not lead to a correct molecular replacement solution. The sequence identity between TA1305 and 1FI4 is under 20%. The *apo* model was initially refined using Refmac5<sup>44</sup> with no non-crystallographic symmetry restraints. The residual  $F_o - F_c$  map showed clear positive difference density for side chain atoms that were not included in the model, indicating the correctness of the solution. The automatic chain tracing program, Buccaneer,<sup>45</sup> built in missing parts of the model. This was followed by automated building and refinement with the program ARP/wARP.<sup>46</sup> After each refinement step, the models were visually inspected in COOT,<sup>47</sup> and modified using guidance from both  $2F_o - F_c$  and  $F_o - F_c$  difference maps. Structures for substrate and product bound data sets were solved by refinement against the *apo* structures using phenix.refine.<sup>48</sup> Data collection and refinement statistics are reported in Table 3.2. The models were validated with PROCHECK,<sup>49</sup> ERRAT,<sup>50</sup> and VERIFY3D.<sup>51</sup> The coordinates of the final models and structure factors have been deposited in the Protein Data Bank with PDB code 4RKP (*apo*), 4RKS (substrate), and 4RKZ (products). The structures were illustrated using Pymol<sup>52</sup> and compared to *S. epidermidis* MDD (PDB: 4DU7 & 4DPW) with MDD residues numbered as per their PDB files.

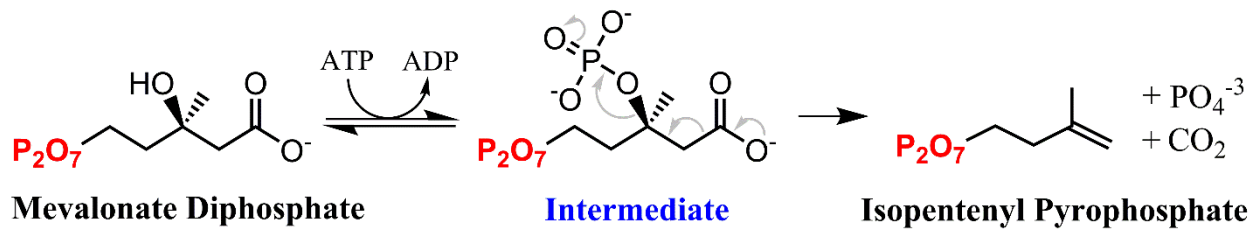
**Mutagenesis.** 100 ng of pET28a plasmid containing the *ta1305* gene was subjected to 32 cycles of the PCR protocol outlined in the manual of PfuUltra II Fusion HS DNA Polymerase (Agilent Cat No. 600670). Mutagenic primers were designed using PrimerX.<sup>53</sup> 50  $\mu$ L of the completed PCR reaction was treated with 1  $\mu$ L DpnI (New England Biolabs) for 1 hour at 37°C, and then 3  $\mu$ L was used to transform 100  $\mu$ L of *E. coli* BL21 Gold (DE3) cells. Transformants were selected on LB-agar containing 50  $\mu$ g/mL kanamycin. Mutations were verified by sequencing the plasmid from the T7 promoter.

**Kinetic Values for M3K Variants.** Kinetic measurements for all mevalonate-3-kinase variants were carried out in duplicate on a SpectraMax M5 microplate reader. Continuous monitoring of M3K activity under optional conditions (pH 8.5 and 55°C) was achieved through a coupled kinase assay as previously described in detail.<sup>20</sup> Initial rates of ATP consumption were determined over a range of (R)-mevalonate concentration, while holding ATP at a saturating concentration of 5 mM. These values were plotted with KaleidaGraph, version 4.0 (Synergy Software, Reading, PA, USA) and fit to a Michaelis-Menten curve to yield  $K_m$  and  $k_{cat}$  values with respect to (R)-mevalonate.

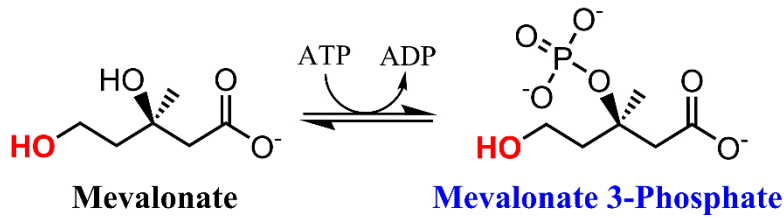
**Sequence Alignment.** A pdb structure for M3K (*apo*) and 8 known MDD structures from the Protein Data Bank (*H. sapiens*, *M. musculus*, *T. brucei*, *S. cerevicie*, *L. pneumophila*, *S. aureus*, *S. epidermidis*, *S. pyogenes*) were submitted to PROMALS3D, a structure alignment server.<sup>54</sup> The output sequences were edited with Macromedia Fireworks MX 2004<sup>55</sup> to improve loop alignment and to fill in structurally disordered residues. The final alignment was rendered with Esript 3.0.<sup>56</sup>

**Comparison of Monomer and Dimer Activity.** Monomer was prepared by incubating enzyme with 1 mM  $\beta$ -mercaptoethanol for 1 hour at 37°C, and dimer was prepared by incubating the enzyme with 100  $\mu$ M copper acetate complexed with 100  $\mu$ M 1,10-phenanthroline for 1 hr at 37°C. Heat stability was tested by incubating 0.16 mg/mL of both enzyme forms at 40-81°C for 60 min in increments of 1-3°C. After heating, residual activity was tested as described previously in an assay buffer without reducing agent.<sup>20</sup>

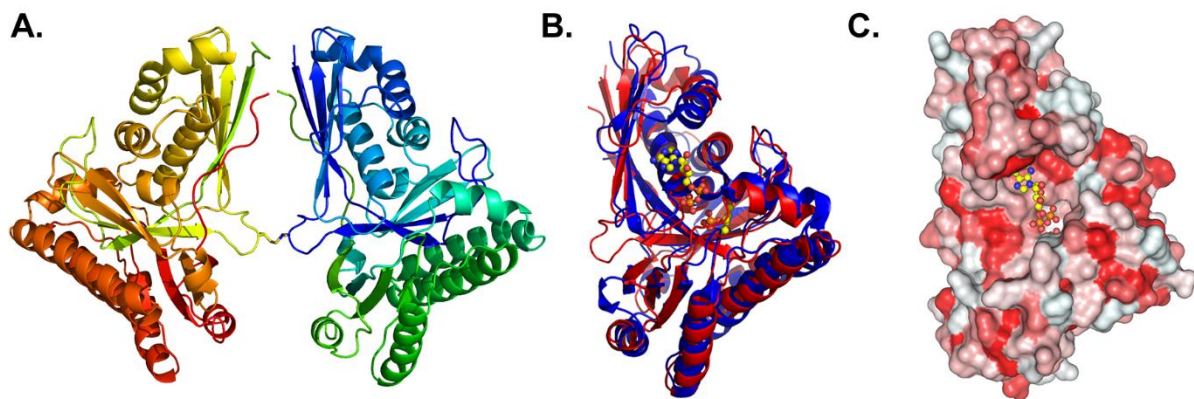
*A. Mevalonate Diphosphate Decarboxylase*



*B. Mevalonate-3-Kinase*

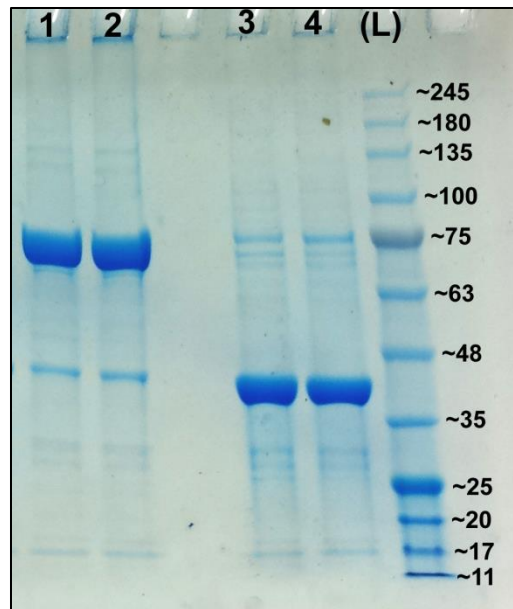


**Figure 3.1. Enzymatic Reactions.** (A) ATP-dependent phosphorylation catalyzed by mevalonate-3-kinase. (B) Identical phosphorylation catalyzed by mevalonate diphosphate decarboxylase, but with concomitant decarboxylation to yield isopentenyl pyrophosphate.

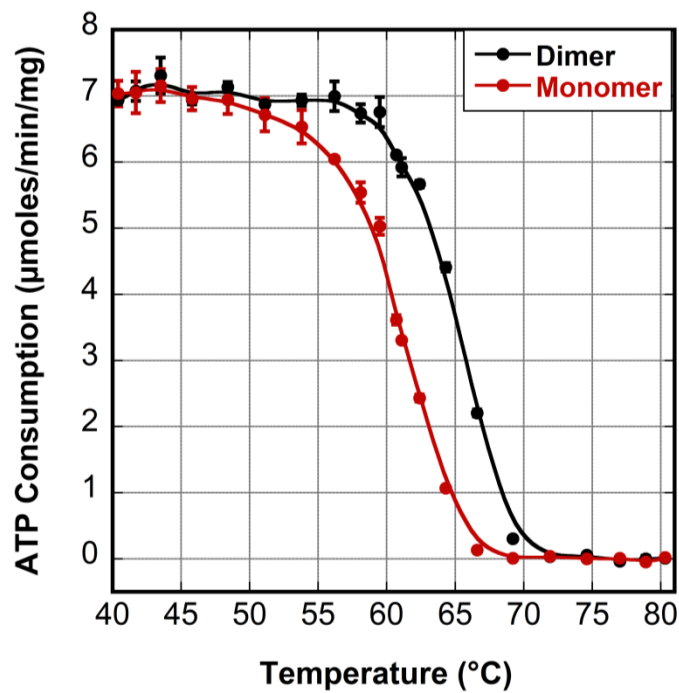


**Figure 3.2. Overall Structure of M3K.** (A) The asymmetric unit of M3K highlighting the disulfide bond between the two proteins. (B) Overlay of M3K (red) with *S. epidermidis* MDD (blue) shows a nearly identical overall fold. (C) M3K surface rendering using a hydrophobicity scale<sup>57</sup> shows the deep active site cleft. Red indicates hydrophobic residues and white indicates hydrophilic residues.

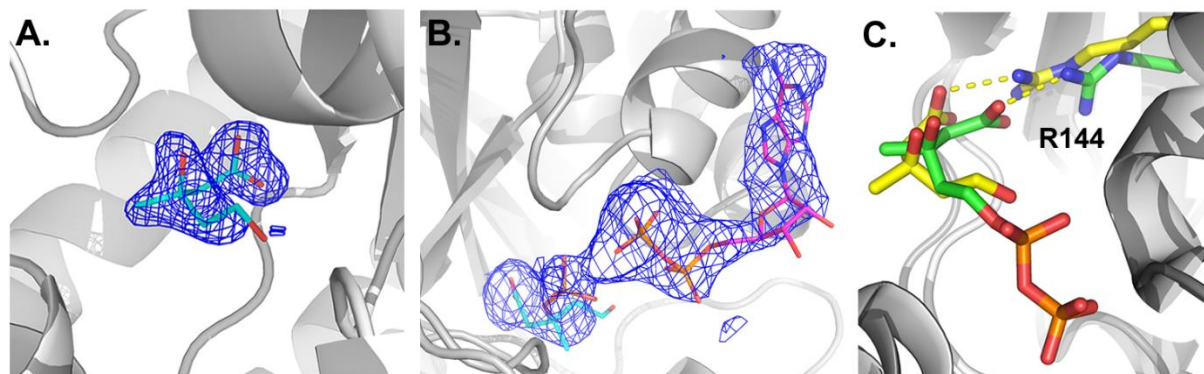




**Figure 3.3.** SDS-PAGE gel showing dimer form (Lanes 1,2) and monomer form (Lanes 3,4) of *mevalonate-3-kinase*. The enzyme was incubated for 1 hour at 37°C in the following conditions: (Lane 1) 1 mM copper-1,10-phenanthroline, (Lane 2) 100  $\mu$ M copper-1,10-phenanthroline, (Lane 3) 10 mM  $\beta$ -mercaptoethanol, (Lane 4) 1 mM  $\beta$ -mercaptoethanol. The gel was a 4-20% SDS-PAGE gel (GenScript, Piscataway, NJ, USA) run at 150V with Tris-MOPS buffer. Lane (L) is an AccuRuler RGB Pre-stained Protein Ladder (Biopioneer, San Diego, CA, USA).



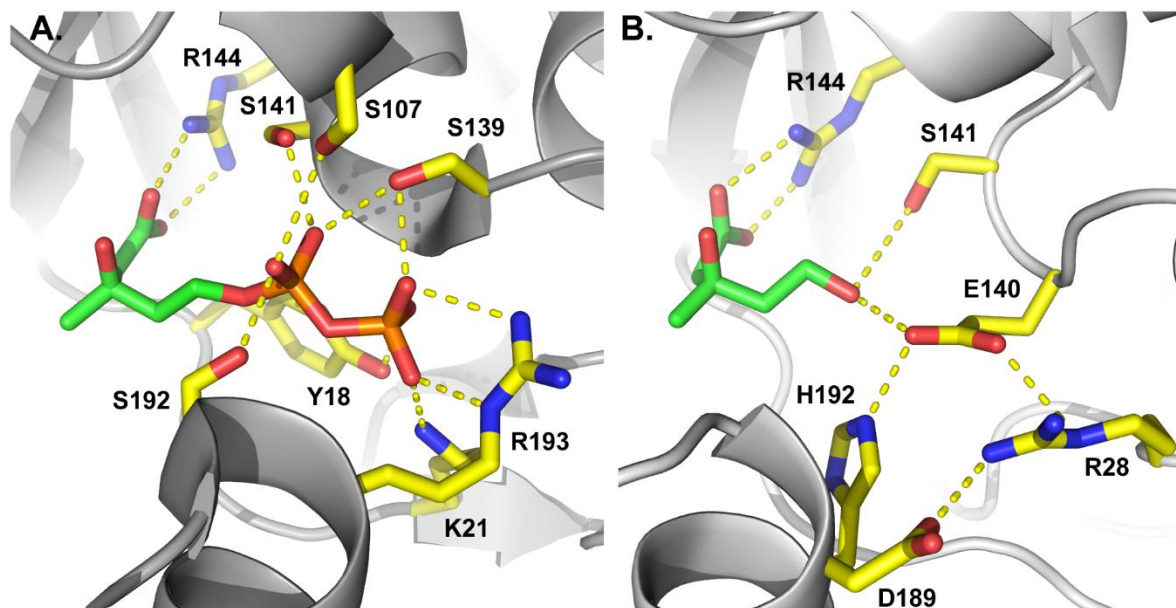
**Figure 3.4.** Comparison of the thermal stability of the monomer and dimer forms of mevalonate-3-kinase. Monomer was prepared by incubating enzyme with 1 mM  $\beta$ -mercaptoethanol for 1 hour at 37°C, and dimer was prepared by incubating the enzyme with 100  $\mu$ M copper acetate complexed with 100  $\mu$ M 1,10-phenanthroline for 1 hr at 37°C. Heat stability was tested by incubating 0.16 mg/mL of both enzyme forms at 40-81°C for 60 min in increments of 1-3°C using a PCR machine gradient. After heating, residual activity was tested as described previously in an assay buffer without reducing agent (see paper, reference 20).



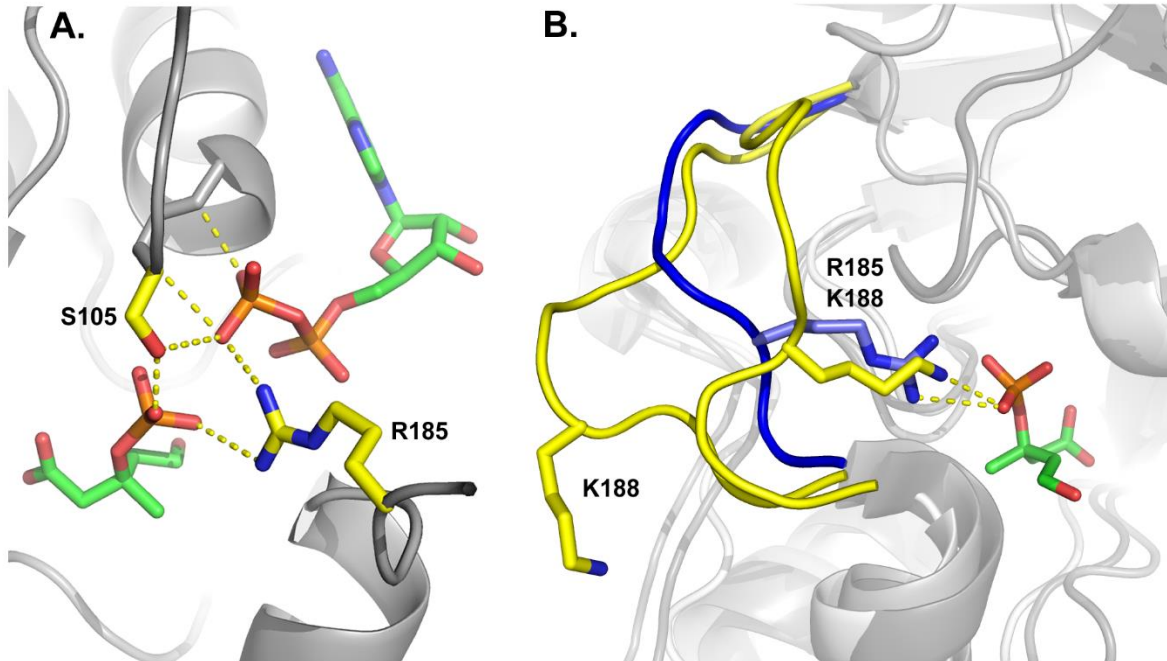
**Figure 3.5. Substrate Binding to M3K.** (A) Electron density for mevalonate. The blue mesh is a simulated annealing  $F_o-F_c$  omit map contoured at  $2.5 \text{ \AA}$ . This omit map was obtained by subjecting the refined protein coordinates without ligands to a round of simulated annealing refinement starting at 10,000 K. (B) A simulated annealing  $F_o-F_c$  omit map for mevalonate 3-phosphate and ADP, generated in the same manner as mevalonate and contoured at  $1.9 \text{ \AA}$ . (C) M3K with bound mevalonate (yellow) is overlaid against MDD with bound mevalonate diphosphate (green) showing the conserved binding location.

	18	28	105	140	144	185	192	275
<i>T. acidophilum</i>	VV <b>L</b> GGI	T <b>R</b> TP	. . . L <b>S</b> GSSD	V <b>S</b> <b>S</b> SA <b>G</b> RS	AF .DYQ <b>R</b> .NP <b>S</b> DV <b>I</b> H	YIV <b>T</b> GG <b>S</b>		
<i>M. musculus</i>	VIKYW <b>G</b> K	LIL <b>P</b>	TAAGL <b>A</b> SSA	G <b>S</b> G <b>S</b> A <b>C</b> RS	VV <b>S</b> AD <b>K</b> KQT <b>G</b> ST <b>V</b> GM	Y <b>T</b> F <b>D</b> AG <b>P</b>		
<i>H. sapien</i>	VIKYW <b>G</b> K	LVL <b>P</b>	TAAGL <b>A</b> SSA	G <b>S</b> G <b>S</b> A <b>C</b> RS	VV <b>S</b> AE <b>K</b> KLT <b>G</b> ST <b>V</b> GM	Y <b>T</b> F <b>D</b> AG <b>P</b>		
<i>S. cerevisiae</i>	TLKYW <b>G</b> K	LNL <b>P</b>	TAAGL <b>A</b> SSA	G <b>S</b> G <b>S</b> A <b>C</b> RS	VV <b>S</b> DI <b>K</b> KD <b>V</b> ST <b>Q</b> GM	Y <b>T</b> F <b>D</b> AG <b>P</b>		
<i>T. brucei</i>	FIKYW <b>G</b> K	LIL <b>P</b>	TAAG <b>M</b> ASSA	G <b>S</b> G <b>S</b> A <b>C</b> RS	VL <b>K</b> GA <b>Q</b> KD <b>V</b> ST <b>K</b> GM	Y <b>T</b> F <b>D</b> AG <b>A</b>		
<i>S. pyogenes</i>	IIKYW <b>G</b> K	KMI <b>P</b>	TAAGL <b>S</b> SSA	A <b>S</b> G <b>S</b> S <b>S</b> RS	VL <b>N</b> AA <b>K</b> K <b>P</b> IS <b>S</b> REG <b>M</b>	F <b>T</b> M <b>D</b> AG <b>P</b>		
<i>S. aureus</i>	LIKYW <b>G</b> K	LII <b>P</b>	TAAGL <b>A</b> SSA	G <b>S</b> G <b>S</b> A <b>S</b> RS	VI <b>N</b> Q <b>H</b> S <b>K</b> K <b>V</b> SR <b>Y</b> GM	F <b>T</b> M <b>D</b> AG <b>P</b>		
<i>S. epidermidis</i>	LIKYW <b>G</b> K	YII <b>P</b>	TAAGL <b>A</b> SSA	G <b>S</b> G <b>S</b> A <b>S</b> RS	VI <b>N</b> N <b>Q</b> S <b>K</b> K <b>V</b> SR <b>S</b> GM	F <b>T</b> M <b>D</b> AG <b>P</b>		
<i>L. pneumophila</i>	LIKY <b>M</b> G <b>K</b>	SN <b>L</b> P	H <b>S</b> S <b>G</b> L <b>A</b> SSA	G <b>S</b> G <b>S</b> S <b>C</b> RS	VI <b>S</b> S <b>Q</b> E <b>K</b> E <b>I</b> PS <b>R</b> V <b>A</b> H	V <b>T</b> M <b>D</b> AG <b>P</b>		

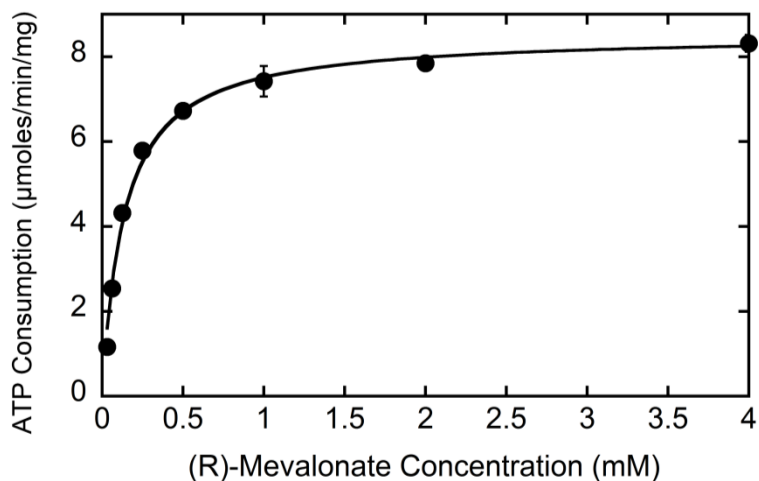
**Figure 3.6. Sequence Alignment of M3K and MDDs.** A structure-based amino acid sequence alignment of M3K from *T. acidophilum* (Ta1305) aligned against the sequences of eight known MDD structures. The numbering corresponds to M3K and residues discussed in this paper are highlighted in green.



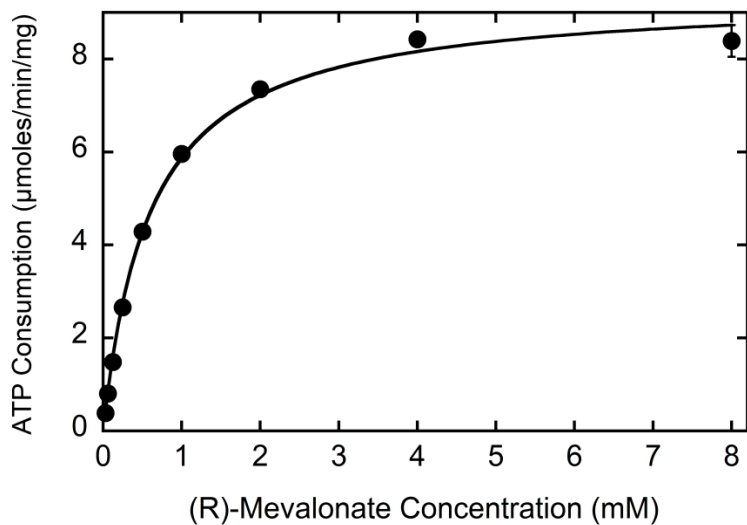
**Figure 3.7. Binding Site Comparison of MDD and M3K.** (A) *S. epidermidis* MDD active site highlighting residues that make polar contacts to the pyrophosphate. (B) M3K active site showing Glu140 and the web of polar contracts that engage the 5-OH tail of mevalonate.



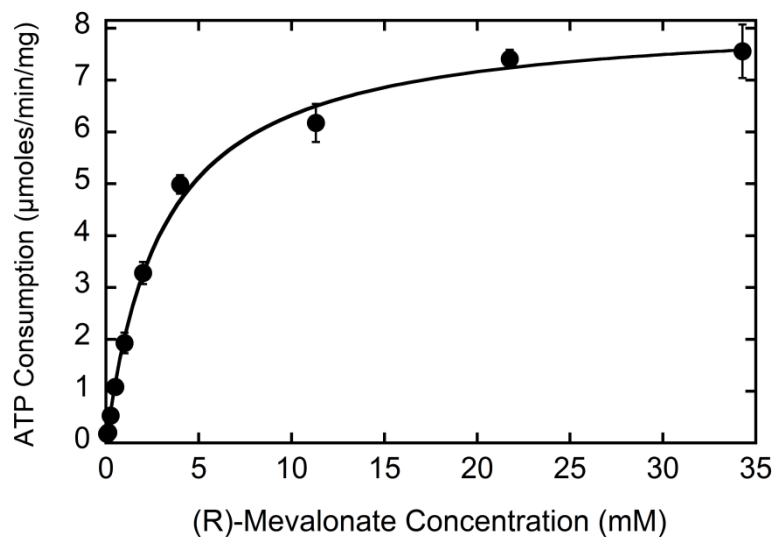
**Figure 3.8. Catalysis of Phosphate Transfer.** (A) Polar contacts of Arg185 and Ser105 to mevalonate 3-phosphate and ADP. (B) The two conformations of the flexible loop of MDD. The outward conformation (yellow) is from *L. pneumophila*, and the inward conformation (also yellow) is a D283A mutant from *S. epidermidis*. Lys188 of MDD (yellow) overlays with Arg185 of M3K (blue) when the loop is oriented inward.



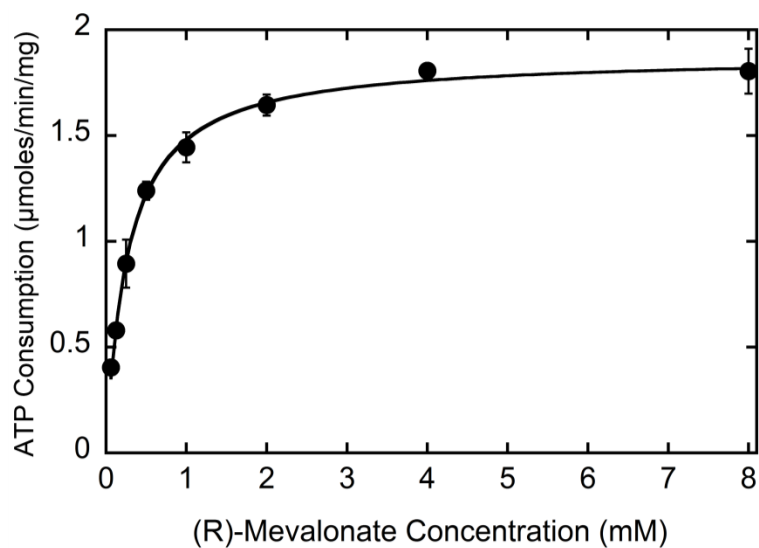
**Figure 3.9.** Michaelis-Menten plot for wildtype mevalonate-3-kinase (Ta1305). The enzyme was found to have a  $K_m$  of  $130 \pm 10 \mu\text{M}$  and  $k_{\text{cat}}$  of  $5.3 \pm 0.1 \text{ s}^{-1}$  with respect to (*R*)-Mevalonate.



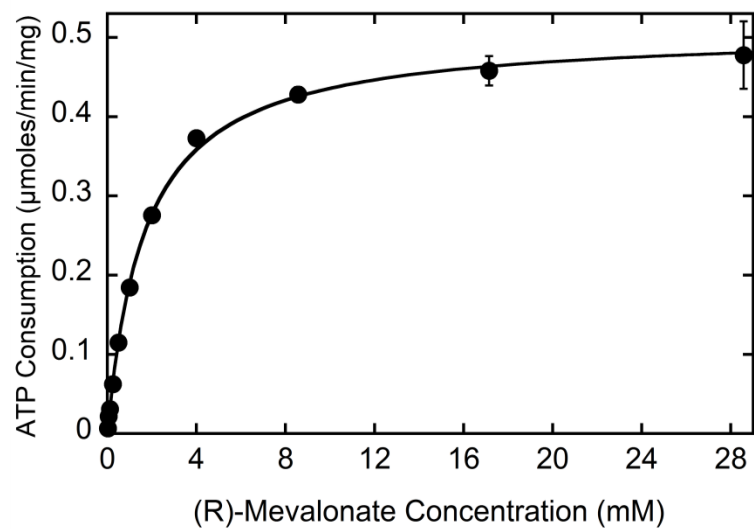
**Figure 3.10.** Michaelis-Menten plot for a T275A mutant of mevalonate-3-kinase. The mutant was found to have a  $K_m$  of  $590 \pm 40 \mu\text{M}$  and  $k_{\text{cat}}$  of  $5.8 \pm 0.1 \text{ s}^{-1}$  with respect to (*R*)-Mevalonate.



**Figure 3.11.** Michaelis-Menten plot for a L18A mutant of mevalonate-3-kinase. The mutant was found to have a  $K_m$  of  $3100 \pm 200 \mu\text{M}$  and  $k_{\text{cat}}$  of  $5.1 \pm 0.1 \text{ s}^{-1}$  with respect to (*R*)-Mevalonate.

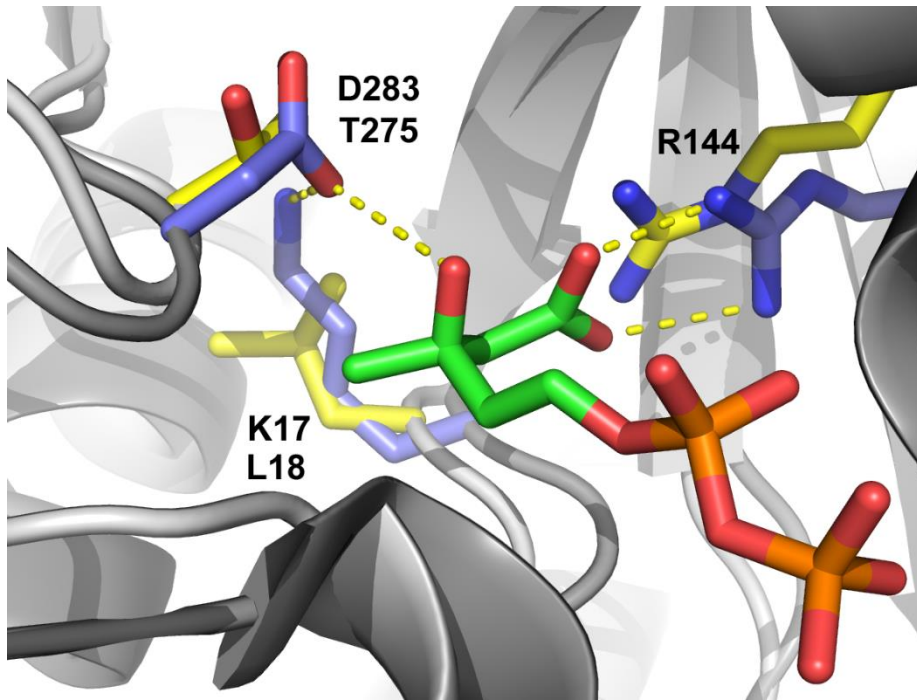


**Figure 3.12.** Michaelis-Menten plot for a S105A mutant of mevalonate-3-kinase. The mutant was found to have a  $K_m$  of  $270 \pm 10 \mu\text{M}$  and  $k_{\text{cat}}$  of  $1.16 \pm 0.01 \text{ s}^{-1}$  with respect to (*R*)-Mevalonate.

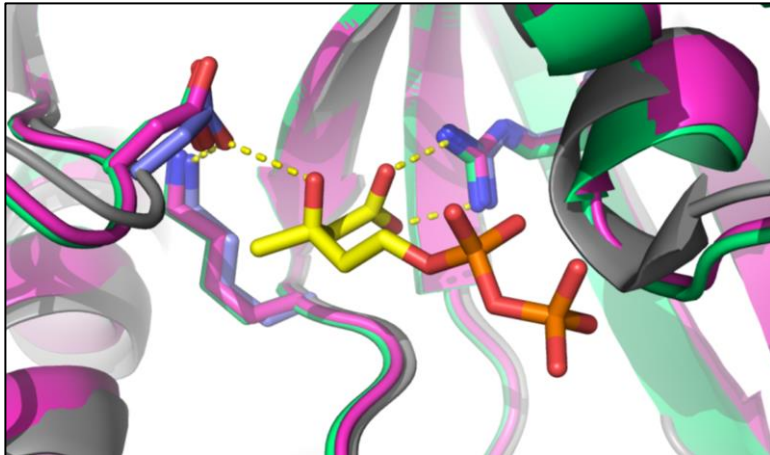


**Figure 3.13.** Michaelis-Menten plot for a R185K mutant of mevalonate-3-kinase. The mutant was found to have a  $K_m$  of  $1680 \pm 60 \mu\text{M}$  and  $k_{\text{cat}}$  of  $0.31 \pm 0.01 \text{ s}^{-1}$  with respect to (*R*)-Mevalonate.





**Figure 3.14. Residues Implicated in Decarboxylation.** The invariant MDD residues of *S. epidermidis* are shown in blue (Arg144, Lys17 and Asp283). In M3K, Lys17 of MDD is replaced by Leu18 (yellow) and Asp283 is replaced by Thr275 (also yellow).



**Figure 3.15.** *Decarboxylase homology models.* PHYRE models of mevalonate monophosphate decarboxylases from *Roseiflexus castenholzii* (Green) and *Haloferax volcanii* (Purple) are overlaid with a known structure of mevalonate diphosphate decarboxylate (Grey, PDB: 4DU7). Retention of the Asp/Lys pair in all known mevalonate pathway decarboxylases and absence in mevalonate-3-kinase, suggests the Asp/Lys pair is not necessary for phosphorylation.

**Table 3.1. Kinetic parameters of M3K variants**

Mutation	$K_m$ ( $\mu\text{M}$ ) <sup>a</sup>	$k_{\text{cat}}$ ( $\text{s}^{-1}$ ) <sup>a</sup>	$k_{\text{cat}}/K_m \times 10^3$ (%) <sup>b</sup>
wildtype	130 ± 10	5.3 ± 0.1	39 ± 9 (100%)
T275A	590 ± 40	5.8 ± 0.1	10 ± 3 (25.6%)
L18A	3100 ± 200	5.1 ± 0.1	1.7 ± 0.5 (4.4%)
S105A	270 ± 10	1.16 ± 0.01	4 ± 1 (10.3%)
R185K	1680 ± 60	0.31 ± 0.01	0.19 ± 0.05 (0.5%)
R185A	0.0 <sup>c</sup>	0.0 <sup>c</sup>	---

<sup>a</sup>In the presence of 5 mM ATP, which was found to be sufficient to saturate all the enzyme variants. <sup>b</sup>Percentage values of  $k_{\text{cat}}/K_m$  are relative to wildtype M3K. <sup>c</sup>Rate not detectable. The minimum detectable rate of the coupled assay is 0.2% relative to wildtype activity.

**Table 3.2.** X-ray Crystallography Statistics (Molecular Replacement).

Mevalonate-3-Kinase	<i>apo</i>	mevalonate	mevalonate-3P / ADP
PDB Accession Code	4RKP	4RKS	4RKZ
<b>Data Collection</b>			
Beamline	UCLA	UCLA	APS-24-ID-C
Wavelength (Å)	1.5418	1.5418	0.9789
Space group	C2	C2	C2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	150.5, 62.5, 103.2	148.0, 61.2, 104.0	146.4, 61.1, 103.6
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 123.72, 90	90, 123.22, 90	90, 122.95, 90
Resolution (Å)	2.1	2.0	2.3
$R_{sym}$ (%)	10.9 (74.3)	7.4 (58.1)	7.4 (68.8)
$I / \sigma I$	16.3 (3.1)	20.0 (3.6)	15.0 (3.2)
CC (1/2)	99.8 (84.0)	99.9 (90.8)	99.8 (91.6)
Completeness	99.5 (99.7)	97.7 (97.7)	99.2 (99.4)
Redundancy	7.4 (7.3)	7.4 (7.3)	6.7 (6.8)
Wilson B-factor (Å <sup>2</sup> )	27.7	27.3	53.0
<b>Refinement</b>			
Resolution (Å)	2.1	2.0	2.3
No. reflections	46796	51657	34194
$R_{work} / R_{free}$	0.182/0.218	0.195/0.242	0.202/0.249
<b>No. atoms</b>			
Protein	4875	4950	4747
Non-protein	227	232	114
Avg. B-factors (Å <sup>2</sup> )	30.5	32.7	57.0
<b>R.M.S.D.</b>			
Bond lengths (Å)	0.008	0.007	0.003
Bond angles (°)	1.013	0.970	0.672
<b>Ramachandran</b>			
Favored (%)	97.8	98.0	97.3
Allowed (%)	2.2	1.7	2.7
Outliers (%)	0.0	0.3	0.0

\*One crystal for each structure was used for collection and refinement.

\*Highest-resolution shell is shown in parentheses.

\* $R_{sym} = \sum ||-<I>| / \sum <I>$ , where  $I$  is the observed intensity and  $<I>$  is the average intensity from observed symmetry-related reflections. CC(1/2) = correlation coefficient between two halves of the data.  $R_{work} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$ , where  $F_{obs}$  and  $F_{calc}$  are the observed and calculated structure factor amplitudes, respectively.  $R_{free}$  is calculated from 10% of reflections not included in the refinement.

### 3.6 REFERENCES

1. Holstein SA, Hohl RJ (2004) Isoprenoids: Remarkable Diversity of Form and Function. *Lipids* 39:293–309.
2. Goldstein JL, Brown SB (1990) Regulation of the mevalonate pathway. *Nature* 343:425-430.
3. Katsuki H, Bloch K (1967) Studies on the biosynthesis of ergosterol in yeast. Formation of methylated intermediates. *J Biol Chem.* 242:222–227.
4. Buhaescu I, Izzedine H (2007) Mevalonate pathway: a review of clinical and therapeutical implications. *Clinical biochemistry.* 40: 575–584.
5. Istvan ES (2002) Structural mechanism for statin inhibition of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Am Heart J* 144:S27-S32.
6. Qiu Y, Li D (2006) Bifunctional inhibitors of mevalonate kinase and mevalonate 5-diphosphate decarboxylase. *Organic letters* 8:1013–1016.
7. Brown MS, Goldstein JL. (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232:34-47.
8. Sever PS, Dahlöf B, Poulter NR, Wedel H, Beevers G, Caulfield M, Collins R, et al. (2003) Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial: a multicentre randomised controlled trial. *The Lancet* 361:1149–1158.
9. Gonzalez-Pacanowska D, Marco C, Garcia-Martinez J, Garcia-Peregrin E (1985) Role of mevalonate-5-pyrophosphate decarboxylase in the regulation of chick intestinal cholesterologenesis. *Biochim Biophys Acta* 833:449-455.
10. Qiu Y, Li D (2006) Inhibition of mevalonate 5-diphosphate decarboxylase by fluorinated substrate analogs. *Biochimica et biophysica acta* 1760:1080–1087.
11. Vlattas I, Dellureficio J, Ku E, Bohacek R, Zhang X (1996) Inhibition of mevalonate 5-pyrophosphate decarboxylase by a proline-containing transition state analog. *Bioorg Med Chem Lett* 6:2091–2096.
12. Heuston S, Begley M, Gahan CGM, Hill C (2012) Isoprenoid biosynthesis in bacterial pathogens. *Microbiology* 158:1389–1401.

13. Wilding EI, Brown JR, Bryant AP, Chalker AF, Holmes DJ, Ingraham KA, Iordanescu S, So CY, Rosenberg M, Gwynn MN (2000) Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol* 182:4319–4327.
14. Barta ML, Skaff DA, McWhorter WJ, Herdendorf TJ, Miziorko HM, Geisbrecht BV (2011) Crystal structures of *Staphylococcus epidermidis* mevalonate diphosphate decarboxylase bound to inhibitory analogs reveal new insight into substrate binding and catalysis. *J Biol Chem* 286:23900–23910.
15. Jabalquinto AM, Alvear M, Emil EC (1988) Mini-Review: Physiological Aspects and Mechanism of action of Mevalonate 5-Diphosphate Decarboxylase. *Comp Biochem Physiol* 90:671–677.
16. Lindberg M, Yuan C, de Waard A, Bloch K (1962) On the formation of Isopentenyl Pyrophosphate. *Biochemistry* 1:182-188.
17. Iyengar R, Cardemil E, Frey PA (1986). Mevalonate-5 diphosphate Decarboxylase: Stereochemical Course of ATP-Dependent Phosphorylation of mevalonate 5-diphosphate. *Biochemistry* 25:4693-4698.
18. Skilleter DN, Kekwick RG (1971) The enzymes forming isopentenyl pyrophosphate from 5-phosphomevalonate (mevalonate 5-phosphate) in the latex of *Hevea brasiliensis*. *The Biochemical Journal* 124:407–417.
19. Byres E, Alphey MS, Smith TK, Hunter WN (2007) Crystal structures of *Trypanosoma brucei* and *Staphylococcus aureus* mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity. *J Mol Biol* 371:540–553.
20. Vinokur JM, Korman TP, Cao Z, Bowie JU (2014) Evidence of a novel mevalonate pathway in archaea. *Biochemistry* 53:4161–4168.
21. Dhe-Paganon S, Magraph J, Abeles RH (1994) Mechanism of Mevalonate Pyrophosphate Decarboxylase: Evidence for a Carbocationic Transition State. *Biochemistry* 33:13355–13362.
22. Azami Y, Hattori A, Nishimura H, Kawaide H, Yoshimura T, Hemmi H (2014) (R)-mevalonate 3-phosphate is an intermediate of the mevalonate pathway in *Thermoplasma acidophilum*. *J Biol Chem* 289:15957–15967.
23. Voynova NE, Fu Z, Battaile KP, Herdendorf TJ, Kim JP, Miziorko HM (2008) Human mevalonate diphosphate decarboxylase: characterization, investigation of the mevalonate diphosphate binding site, and crystal structure. *Arch Biochem Biophys* 480:58–67.

24. Bonanno JB, Edo C, Eswar N, Pieper U, Romanowski MJ, Ilyin V, Gerchman SE, et al. (2001) Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis. *Proc Natl Acad Sci* 98:12896–12901.
25. Barta, ML, McWhorter WJ, Miziorko HM, Geisbrecht BV (2012) Structural basis for nucleotide binding and reaction catalysis in mevalonate diphosphate decarboxylase. *Biochemistry* 51:5611–5621.
26. Jorda J, Yeates TO (2011) Widespread disulfide bonding in proteins from thermophilic archaea. *Archaea* 2011:409156.
27. Beeby M, O'Connor BD, Ryttersgaard C, Boutz DR, Perry LJ, Yeates TO (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol* 3:e309.
28. Mallick P, Boutz DR, Eisenberg D, Yeates TO (2002) Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proc Natl Acad Sci* 99:9679-9684.
29. Cheek S, Zhang H, Grishin NV (2002) Sequence and Structure Classification of Kinases. *J Mol Biol* 320:855–881.
30. Andreassi JL, Leyh TS (2004) Molecular Functions of Conserved Aspects of the GHMP Kinase Family. *Biochemistry*. 43:14594–14601.
31. Qiu Y, Gao J, Guo F, Qiao Y, Li D (2007) Mutation and inhibition studies of mevalonate 5-diphosphate decarboxylase. *Bioorg Med Chem Lett* 17:6164–6168.
32. Krepkij D, Miziorko HM (2004) Identification of active site residues in mevalonate diphosphate decarboxylase : Implications for a family of phosphotransferases. *Protein Sci* 13:1875–1881.
33. Zhou T, Daugherty M, Grishin NV, Osterman AL, Zhang H (2000) Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily. *Structure* 8:1247–1257.
34. Potter D, Miziorko HM (1997) Identification of Catalytic Residues in Human Mevalonate Kinase. *J Biol Chem* 272:25449–25454.
35. Andreassi JL, Vetting MW, Bilder PW, Roderick SL, Leyh TS (2010) Structure of the ternary complex of phosphomevalonate kinase: the enzyme and its family. *Biochemistry* 48: 6461–6468.
- 36 Vannice JC, Skaff DA, Keightley A, Addo J, Wyckoff GJ, Miziorko HM (2013) Identification in *Haloferax volcanii* of Phosphomevalonate Decarboxylase and Isopentenyl Phosphate Kinase as Catalysts of the Terminal Enzymatic Reactions in an Archaeal Alternate Mevalonate Pathway. *J Bacteriol* 196:1055-1063

37. Dellas N, Thomas ST, Manning G, Noel JP (2013) Discovery of a metabolic alternative to the classical mevalonate pathway. *eLife* 2:e00672.
38. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nature Protoc* 4:363–371.
39. McPherson A. *Preparation and Analysis of Protein Crystals* (Wiley, New York, 1982).
40. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125-132.
41. Keegan RM, Winn MD (2007) Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Cryst D* 63:447–457.
42. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 1:658-674.
43. Stein, ND (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Cryst* 41:641-643.
44. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.
45. Cowtan K (2006) Buccaneer software for automated model building. *Acta Cryst. D* 62:1002-1011.
46. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for x-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171-1179.
47. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501.
48. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Cryst D* 68:352-367.
49. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283-291.
50. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511-1519.
51. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356:83-85.

52. The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC ([www.pymol.org](http://www.pymol.org)).
53. Automated design of mutagenic primers for site-directed mutagenesis, PrimerX, 2014 Version, Carlo Lapid ([www.bioinformatics.org/primerx/](http://www.bioinformatics.org/primerx/)).
54. Pei J, Kim BH, Grishin VN (2008) PROMALS3D: a tool for multiple sequence and structure alignment. Nucl Acids Res 6:2295-2300.
55. Macromedia Fireworks MX 2004, Version 7.0, Adobe Systems Inc. ([www.adobe.com](http://www.adobe.com))
56. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucl Acids Res 13:3320-3323.
57. Eisenberg D, Schwarz E, Komarony M, Wall R (1984) Amino acid scale: Normalized consensus hydrophobicity scale. J Mol Biol 179:125-142.



## CHAPTER 4

### Mutating Mevalonate Pathway Enzymes to Produce Isoprenol

#### 4.1 ABSTRACT

In this study we attempt to rationally redesign three mevalonate pathway enzymes (mevalonate 3-kinase, mevalonate 5-pyrophosphate decarboxylase, and mevalonate 5-monophosphate decarboxylase) in order to create a new enzyme we call “mevalonate decarboxylase.” In addition, we developed a robust *E. coli* selection platform to allow random mutagenesis to give rise to mevalonate decarboxylase. Unfortunately, neither rational design nor selection yielded the new enzyme. We believe this is because the enzyme mechanism requires the presence of at least one phosphate attached to the mevalonate moiety in order to trigger a conformational change allowing the active enzyme-substrate complex to form.

#### 4.2 INTRODUCTION

In the last 10 years, significant progress has been made in the microbial production of fuels such as isopropanol, pharmaceuticals such as artemisinin, and commodity chemicals such as acrylic acid.<sup>1</sup> This progress has been dependent on advances in synthetic biology, which focuses on isolating genes from organisms that synthesize the compound of interest, introducing them into microorganisms optimized for industrial biochemical production, and then optimizing expression for maximum production of the compound.<sup>2</sup> Perhaps the best known example is the production of ethanol (grain alcohol) by yeast or *E. coli* which consume glucose and convert it to ethanol through metabolic processes. Decades of modern

biochemistry research has made ethanol production more cost effective by optimizing the fermentation pathway, increasing ethanol tolerance, and reducing competing pathways that consume carbon.<sup>3</sup>

Here we attempt to produce isoprenol, a 5-carbon alcohol that has several desirable properties for use as a biofuel: a very high octane rating, an energy density greater than ethanol, and is less hygroscopic than ethanol.<sup>4,5</sup> These properties make it an ideal drop-in biofuel since it can be used directly in conventional combustion engines and it can be derived from renewable resources. Today the majority of isoprenol is produced through the condensation of formaldehyde and isobutene (a petroleum product). While isoprenol is not directly formed in nature, a phosphorylated form is produced through the native mevalonate pathway of eukaryotes.<sup>6,7</sup> In brief, the mevalonate pathway begins with the condensation of three acetyl-CoA molecules and one reduction step to yield mevalonate. Mevalonate is then phosphorylated twice at the 5-OH position to yield mevalonate 5-pyrophosphate. In the final step, mevalonate 5-pyrophosphate is converted to isopentenyl pyrophosphate (IPP) through the action of mevalonate pyrophosphate decarboxylase (MDC) (Figure 4.1).

Previous attempts to produce isoprenol from glucose have relied on the mevalonate pathway to produce isopentenyl pyrophosphate (IPP) and then coupling this pathway with a promiscuous phosphatase (NudB) to hydrolyze IPP into isoprenol and pyrophosphate.<sup>8-10</sup> At the time of this writing, the highest titer obtained using this method was 2.23 g/L isoprenol.<sup>8</sup> This yield is suspected to be limited by the *in vivo* accumulation of IPP, which is toxic to *E. coli* cells at high concentration.

Our recent structural analysis of mevalonate 3-kinase (M3K) and comparison to mevalonate pyrophosphate decarboxylase (MDC) suggested that it may be possible to produce isoprenol directly from mevalonate (Fig. 4.2).<sup>11</sup> This would avoid toxic accumulation of IPP, consume less ATP, use fewer enzymes, and would ideally increase the final titer of isoprenol.

Mevalonate 3-kinase from *Thermoplasma acidophilum* converts (R)-mevalonate into (R)-mevalonate 3-phosphate in an ATP dependent manner (Fig 4.2A).<sup>12</sup> This reaction is unusual because it represents exactly half the reaction of its eukaryotic homolog: mevalonate pyrophosphate decarboxylase (MDC). MDC attaches a phosphate to the 3-OH position of mevalonate 5-pyrophosphate, but then concomitant decarboxylation follows to produce IPP, CO<sub>2</sub>, and PO<sub>4</sub> (Fig 4.2B).<sup>13</sup> It may be possible to transfer the MDC decarboxylation residues into mevalonate 3-kinase in order to create mevalonate decarboxylase, or in the inverse approach, it may be possible to move M3K binding residues into MDC, in order to achieve the same goal: an enzyme capable of directly converting mevalonate to isoprenol.

Partway through this work, another mevalonate pathway enzyme was discovered: mevalonate 5-monophosphate decarboxylase (MMD) from the green non-sulfur bacteria, *Roseiflexus castenholzii*.<sup>14</sup> MMD converts (R)-mevalonate 5-phosphate and ATP into isopentenyl phosphate, ADP, and PO<sub>4</sub> (Fig 4.2C). While no crystal structure of MMD was available at the time of this research, a PHYRE homology model was generated with a 100.0% confidence interval allowing us to attempt to modify the active site of this MMD so that it can accept mevalonate as a substrate.<sup>15</sup> In all we made 42 mutants of three mevalonate pathway enzymes (M3K, MDC, and MMD) in an attempt to rationally design a mevalonate decarboxylase. However, this effort was ultimately unsuccessful.

We also attempted to select for mevalonate decarboxylase by randomly mutating two known mevalonate pathway decarboxylases (MDC and MMD) and then transforming them into an *E. coli* strain requiring isoprenol for growth. Our selection platform contains a deletion in the *dxs* gene (1-deoxy-D-xylulose-5-phosphate synthase) which blocks native production of IPP through the *E. coli* DXP pathway.<sup>16</sup> We then created a plasmid that produces mevalonate *in vivo* and can convert isoprenol into IPP (an essential molecule). We screened two 10<sup>6</sup> libraries of MDC and MMD mutants but were unable to isolate mevalonate decarboxylase. We suspect our selection was unsuccessful due to the enzyme's inability to form the active enzyme-substrate complex without at least one phosphate attached to 5-OH tail of mevalonate. We propose that polar contacts to the 5-phosphate initiates conformational rearrangements of loop regions surrounding the active site. Movement of these loops sequester the ligand within the binding pocket and forms the enzyme-substrate complex.

### 4.3 RESULTS AND DISCUSSION

**Re-designing mevalonate 3-kinase to add decarboxylation function.** In our previous study we superimposed a structure of mevalonate 3-kinase (M3K) from *Thermoplasma acidophilum* (PDB: 4RKP) with mevalonate pyrophosphate decarboxylase (MDC) from *Staphylococcus epidermidis* (PDB: 4DU7).<sup>11</sup> This showed that they have nearly identical overall fold (RMSD of 2.4Å) and their substrates bind in analogous locations in each enzyme (Fig 3.2, 3.5). All known MDDs characterized to date contain an invariant catalytic triad of arginine, aspartate, and lysine in the active site (Fig 4.3).<sup>17</sup> The arginine positions the mevalonate moiety in the active site and we previously proposed that the Asp/Lys residues were involved in decarboxylation as they are absent from the active site of M3K where they are substituted with threonine and isoleucine, respectively (Fig 3.14).<sup>11</sup> Mutating the threonine or leucine to alanine in M3K had no effect on  $k_{cat}$  indicating they played no role in catalysis. We attempted to

introduce the missing asp and lys residues into the M3K enzyme active site to allow conversion of mevalonate to isoprenol (Fig 4.4). Our first variant has two mutations: L18K and T275D. We also created two variants with only one of the mutations (L18K or T275D). All three enzymes expressed and were soluble, however they did not produce any detectable isoprenol after 24 hours of incubation with 10 mM mevalonate and subsequent analysis by gas chromatography as described in the methods.

We suspected that the Asp and Lys residues might not be correctly oriented in the active site so we followed the protocol from our previous paper to crystalize the L18K/T275D double mutant. A crystal of this mutant protein grew in 0.1 M sodium acetate (pH 4.6), 0.2 M ammonium sulfate, and 30% (w/v) PEG 4000. Data was collected in house and refined to 2.1Å. Analysis of the mutant's active site clearly showed that the sidechains of the asp/lys acid-base pair were not in the expected orientation (Fig 4.5A). The residues pointed away from each other instead of making a polar contact. To address this we made a second generation of mutants in which we mutated three residues surrounding the Asp/Lys pair to more closely mimic the environment of the Asp/Lys residues in MDC (Fig. 4.5B). The three additional mutations were added individually to the variant already containing L18K/T275D to make triple mutants. The added mutations were: G26S, I273T, Y281K (Table 4.1). All triple mutants expressed and were soluble, however they did not produce any detectable isoprenol after 24 hours of incubation with 10 mM mevalonate and subsequent gas chromatography analysis.

It is likely that simply substituting the threonine and leucine of M3K for aspartate and lysine was not sufficient since they must be oriented by a web of interactions with the side chains around it, and in turn, those residues are likely anchored by yet another layer. We

decided that the strategy most likely to yield “mevalonate decarboxylase” would be the inverse approach: instead of attempting to give mevalonate 3-kinase a new function, we would mutate the existing decarboxylase to bind mevalonate.

### **Re-designing mevalonate pyrophosphate decarboxylase to bind mevalonate.**

Since yeast mevalonate pyrophosphate decarboxylase (MDC) already does the required decarboxylation chemistry, we decided to mutate residues in the binding pocket to allow it to accept mevalonate as a substrate. Analysis of the binding pocket of *Staphylococcus epidermidis* MDC shows there are many polar contacts between protein sidechains and the large pyrophosphate group (Fig 3.7).<sup>18,19</sup> In M3K, the binding pocket is significantly different. The key difference is Glu140 protrudes into the middle of the binding pocket and makes a polar contact to the 5-OH tail of mevalonate. Furthermore, Glu140 is held in place by a network of polar contacts with residues H192, S141, D189, and R28.<sup>11</sup> Our strategy was to mutate the pyrophosphate binding residues in MDC to the analogous mevalonate binding residues from M3K (Fig 4.6). A total of 22 MDC variants were made (Table 4.1B). The first generation of mutants contained seven designs. One variant in which all six residues that are responsible for binding mevalonate were introduced into MDC, and six additional variants in which only one of the six mutations was made (Y19L, K22I, N28R, G154E, T209D, or M212H) (Table 4.1). None of the single point mutants showed any activity towards mevalonate when tested through an ATP consumption assay (observed as the absorbance at 340 nm via coupling to pyruvate kinase and lactate dehydrogenase) and the variant which combined all six mutations together was not soluble when expressed in *E. coli*. A second generation of designs was created to revert each one of the 6 mutations to its wild type form. Only reversion of the G154E mutation resulted in MDC becoming soluble (Fig 4.7). This is problematic because G154E is the most important mevalonate binding residue that makes a direct polar contact to the 5-OH tail of

mevalonate. Without this mutation, the other mutations do not serve a purpose. Therefore, a third generation of variants was created in which we started with the most critical mutation, G154E, and built in residues one-by-one to support it (Table 4.1B). These third generation designs contain G154E with an additional 1 to 4 mutations to support it in the correct orientation. Unfortunately, none of the 22 MDC variants produced any detectable isoprenol after 24 hours of incubation at 37°C with 10 mM mevalonate as determined by subsequent gas chromatography analysis. All variants were also tested for decarboxylation activity on mevalonate 3-phosphate by including mevalonate 3-kinase in the assay reaction mixture as described in the methods. We also attempted this reaction using 50 mM phosphate buffer and another variation in which we supplemented the reaction with 10 mM sodium pyrophosphate. No isoprenol was detected in any of these conditions.

#### **Re-designing mevalonate 5-monophosphate decarboxylase to bind mevalonate.**

During the time we were mutating MDC to generate isoprenol, a new enzyme was described in the literature: mevalonate monophosphate decarboxylase (MMD) from *Roseiflexus castenholzii* (18). This enzyme converts mevalonate 5-monophosphate and ATP to isopentenyl phosphate, ADP, and  $\text{PO}_4$  (Fig 4.2C). It essentially carries out the exact same reaction as MDC, but it uses a substrate that has only one phosphate attached to the 5-OH tail of mevalonate instead of two. MMD is potentially a better mutagenesis target because its substrate is more analogous to mevalonate. No crystal structure was available for MMD at the time of this work so we planned our mutagenesis designs based on a homology model generated by the PHYRE2 homology modelling web server.<sup>15</sup> The model generated by PHYRE2 gave a confidence score of 100% with the Asp/Lys/Arg catalytic triad from MDC conserved in the same positions in MMD. Structural superposition of MMD with M3K shows the binding sites are remarkably similar (Fig 4.8). In fact, the binding site of M3K could be replicated completely with just two MMD

mutations: G202E and Y74L. G202E is needed to make a polar contact to the 5-OH of mevalonate; remarkably all other residues which position the glutamate residue in M3K are already present in MMD with the exception of a tyrosine (Y74) below the substrate which needs to be mutated to leucine to completely match the M3K binding site (Fig 4.8). We generated 14 designs of MMD (Table 4.1C). The first design attempted to replicate the binding site of M3K via a double mutant: G202E and Y74L. We produced several additional variations of this double mutant by varying two parameters: (1) Y74 was changed to either phenylalanine, leucine, or alanine to reduce the bulk of Y74, (2) G202 was changed to glutamate or aspartate (Table 4.1C). All proteins were soluble and well expressed but were inactive. One final attempt was made to allow mevalonate-binding by making a polar contact to the 5-OH from an alternative position. In these second generation variants, threonine 255 was mutated to glutamate or aspartate, and alternatively tyrosine 74 was mutated to glutamate or aspartate (Fig 4.9). As previously, none of the 14 enzyme designs produced any detectable isoprenol after 24 hours of incubation at 37°C with 10 mM mevalonate or mevalonate 3-phosphate, as tested with both bis-tris propane and phosphate buffer systems.

**Design of a selection platform to create mevalonate decarboxylase.** Due to the lack of success of rational design, we decided to create a selection platform that would allow us to quickly analyze millions of mutants instead of a few designed mutants that were tested one by one. Such a selection platform would require *E. coli* to produce mevalonate and consume isoprenol. *E. coli* does not natively produce mevalonate but this is a simple fix since there are existing plasmids which allow *E. coli* to produce mevalonate.<sup>20</sup> The challenge is that *E. coli* does not naturally utilize isoprenol and there exists no known enzyme that natively phosphorylates it. Searching the literature suggested two possible enzymes might have promiscuous kinase activity on isoprenol. The first, farnesol kinase, a membrane protein from *Arabidopsis thaliana*,



is known to phosphorylate farnesol.<sup>21</sup> Farnesol resembles three isoprenol molecules linked together end to end. A second candidate is hydroxyethylthiazole kinase, an *E. coli* enzyme identified as having promiscuous activity on isoprenol in a patent which described the conversion of isoprenol to isoprene.<sup>22</sup> *In vitro* testing of hydroxyethylthiazole kinase shows it phosphorylates isoprenol with a  $K_M$  of 8 mM and a  $k_{cat}$  of  $1.5 \text{ s}^{-1}$  (Fig. 4.10). Once isoprenol is phosphorylated to isopentenyl phosphate (IP), we can use IP kinase from *Thermoplasma acidophilum* to generate the essential molecule, IPP.<sup>23</sup>

*E. coli* have an absolute requirement for IPP because it is used to generate all of its isoprenoids.<sup>24</sup> Blocking the *E. coli* DXP pathway, which generates IPP, causes cell death. Indeed, this can be demonstrated by using the antibiotic fosmidomycin (fos), which inhibits the enzyme DXP reductoisomerase (dxr) and causes *E. coli* to stop growing at 50  $\mu\text{M}$  fosmidomycin concentration.<sup>25</sup> To test the ability of *E. coli* to utilize isoprenol, we transformed *E. coli* BL21(DE3) with two plasmids. One containing farnesol kinase or hydroxyethylthiazole kinase and a second plasmid containing IP kinase. As shown in figure 4.11, *E. coli* without any plasmid did not grow on an LB-agar plate containing fosmidomycin and isoprenol because it could not utilize the exogenous isoprenol. The *E. coli* strain expressing farnesol kinase and IP kinase grew because it could convert the exogenous isoprenol into IPP. While this was a strong proof of concept, the resulting strain was not yet ready for selection because: (1) it grew slow, (2) contained two plasmids, (3) required 40 mM isoprenol, and (4) did not produce mevalonate *in vivo*. We next optimized these parameters to create a more robust selection strain.

First, we constructed a single plasmid, which could simultaneously produce mevalonate and utilize isoprenol. This was done by modifying an existing plasmid, pJBEI-6409, created by

researchers at the Joint Bioenergy Institute.<sup>26</sup> pJBEI-6409 already expressed three optimized enzymes necessary to produce mevalonate (Fig. 4.12, grey arrows). Since this plasmid was originally intended to generate limonene, we replaced all the downstream genes with farnesol kinase and IP kinase, with the exception of IDI (Isopentenyl-diphosphate delta isomerase). IDI was left intact to balance the intracellular ratio of IPP and its isomer DMAPP (Fig 4.12, black arrow).<sup>27</sup>

This one plasmid should now allow *E. coli* to produce mevalonate via the first three genes (*atoB*, *HMGS*, and *HMGR*), and convert isoprenol to IPP through the last three genes (IP kinase, farnesol kinase, and IDI). We named this new plasmid pSELECT-03. As shown in Figure 4.13, *E. coli* BL21(DE3) containing pSELECT-03 can grow on a plate containing fosmidomycin and 40 mM isoprenol, but it cannot grow on fosmidomycin without isoprenol.

*E. coli* are known to develop resistance to fosmidomycin quite easily because it is imported into the cells primarily through the glycerol 3-phosphate pump.<sup>25</sup> To prevent this issue, we deleted the DXP synthase gene (*dxs*) in our selection strain *E. coli* BL21(DE3). This deletion permanently blocks the first step of the *E. coli* DXP pathway, which produces IPP.<sup>28</sup> Next we used directed evolution to decrease the selection strain's requirement for isoprenol from 40 mM to 200  $\mu$ M. The strain was grown in LB media supplemented with 40 mM isoprenol and subsequently inoculated in LB containing lower concentrations of isoprenol for multiple days (Fig 4.14). Using this method, we were able to evolve a selection strain that required only 200  $\mu$ M isoprenol and formed large colonies in 24 hours (Fig 4.15).

To confirm the cells were producing mevalonate internally at a level that would allow for its rescue, we transformed plasmid pMBI into the selection strain. Plasmid pMBI contains the known lower mevalonate pathway from eukaryotes (mevalonate 5-kinase, phosphomevalonate kinase, mevalonate pyrophosphate decarboxylase) allowing conversion of mevalonate to IPP.<sup>20</sup> When the selection strain was transformed with plasmid pMBI it was able to grow on LB plates without exogenous isoprenol, even in the nascence of IPTG (Fig 4.16). This test confirms that mevalonate is being produced *in vivo* at a level that allows for plasmid-based rescue, even when no IPTG is added.

The final selection strain we developed, called final-2, represents a robust dead/alive selection, contains all genes on a single plasmid, produces mevalonate internally, contains IDI to balance IPP/DMAPP ratio, grows in 24 hours, has an undetectable spontaneous reversion rate due to the *dxs* deletion, and requires only 200  $\mu$ M isoprenol to grow (Fig 4.17).

We next conducted a control experiment to test the ability of the selection strain to be rescued with wildtype MMD or MDC. The selection strain (final-2) was transformed with pET28a containing MMD from *R. castenholzii* or MDC from *S. cerevisiae*. The addition of the un-mutated decarboxylase did not rescue growth (the expected result); however IPTG is needed to over-express the decarboxylase. When growing the selection strain on a high concentration of IPTG, we saw a rise in spontaneous mutants after 72 hours (Fig 4.18). This issue could be avoided by conducting the screening tests at 10  $\mu$ M IPTG at which point the rate of spontaneous reversion mutants was less than 1 in 43,000 at 72 hours. The SDS-PAGE gel shown in figure 4.19 shows that significant MMD is expressed during induction with 10  $\mu$ M IPTG.

**Screening mutant libraries to create mevalonate decarboxylase.** We created two mutant libraries with  $10^6$  diversity: a library of *R. castenholzii* MMD variants and a library of *S. cerevisiae* MDC variants. The mutagenesis library was constructed by Synbio Technologies and their protocol is described in the methods section. For MMD, there was an average of 1.5 amino acid mutations per protein with 73% of clones having between 1-5 amino acid mutations per protein. For MDC, there was an average of 1.7 amino acid mutations per protein with 85% of clones having between 1-5 amino acid mutations per protein. In both libraries, distribution of the mutations appeared randomly across the 400 amino acids.

**Characterization of hits from the selection experiments.** Two selection experiments were carried out in parallel on LB agar plates containing kanamycin, chloramphenicol, and 10  $\mu$ M IPTG (Fig 4.20). Over 600,000 mutants were subjected to our selection for each library (1.2 million in total). For MMD, there were 27 colonies that gained the ability to grow without isoprenol and 1 for MDC. We picked these 28 colonies, mini-prepped the plasmids and sequenced the DNA. Only nine out of 28 plasmids contained mutations in the decarboxylase gene. We then transformed each of the 9 plasmids coding for mutant decarboxylases into fresh final-2 competent cells. Four out of nine plasmids gave the freshly transformed final-2 cells the ability to grow without isoprenol. Sequencing upstream of the decarboxylase gene in these four plasmids revealed that each of them also contained a mutation in either the lac operator or the LacI protein (Table 4.2). It is likely that the upstream mutations cause the decarboxylase to be constitutively overexpressed, which in turn allows the cells to grow without isoprenol (essentially mimicking the effect of high IPTG in the control experiment (figure 4.18)). Next the four mutant decarboxylases were expressed in *E. coli* BL21(DE3) and affinity purified using their his-tags. None of the four mutant MMDs showed the ability to produce isoprenol, after 24 hours of incubation at 55°C, from mevalonate or mevalonate 3-phosphate. Finally, we tested the ability of

these four mutant decarboxylases as well as *wildtype* MMD and MDC to produce isoprenol when the reaction was supplemented with 1, 10, or 100 mM sodium phosphite in order to take the place of the missing phosphate in the active site. No isoprenol production was detected in any of these conditions.

**Analysis of the selection outcome.** While the selection did not result in the generation of “mevalonate decarboxylase” we believe the selection platform is not to blame. The false positive rate of the selection was 1 out of 43,000 colonies, or 28 colonies out of a total of 1,200,000 subjected to the selection. Even if only one colony out of 1,200,000 grew that contained a true mevalonate decarboxylase, it would be detected under these experimental conditions. Assuming the worst case scenario that only 1 specific AA change is able to create mevalonate decarboxylase, the probability of this mutation occurring at random would be 1 out of 8,000 (400 AA in protein sequence x 19 AA substitutions). At this rate, we would expect to see approx. 75 hits per library. Our hit rate was zero. It is likely that a small number of mutations (1-2) is not enough to enable these decarboxylases to gain the ability to act on mevalonate. The possible reasons are discussed in the conclusion.

#### **4.4 CONCLUSION**

Our inability to mutate M3K, MDD, or MMD to mevalonate decarboxylase suggests that one of our core assumptions was incorrect. We assumed that the phosphate attached to 5-OH tail of mevalonate was not essential to the enzyme mechanism. A review of the literature and our own research provides some support for this theory.

A mutagenesis study of *S. episodemidis* MDC showed that the function of residue S192 is catalytic (Fig 4.21A, black arrow).<sup>18</sup> This was surprising because it sits at the far edge of the binding pocket (distant from the known decarboxylation residues) and makes a polar contact to the pyrophosphate. S192 is positioned at the base of a large loop, which has been observed in both open and closed conformations (Fig. 4.21B, yellow chains).<sup>18,19,29</sup> In our most recent publication, we show that MMD is not active on its intermediate, mevalonate 3,5-bisphosphate, in the absence of ADP, suggesting that the molecule is needed in order to close the loops and form the active enzyme-substrate complex.<sup>30</sup> We propose that closing of this loop (Fig 4.41B, yellow chain) is dependent on invariant S192 making a polar contact to the phosphate group attached to mevalonate. Without the polar contact to phosphate, this loop does not close and catalysis does not occur. This model explains why S192A is reported as being involved in catalysis; S192 likely functions to detect bound substrate and triggers closing of the loop. Re-designing MMD and MDC in order to allow them to accept mevalonate therefore requires the adaptation of two functions simultaneously: (1) mevalonate binding; and (2) creation of an alternative way to open and close the loop without S192 contacting the phosphate. Asking the enzyme to solve both of these issues simultaneously is likely too tall of an order.

## 4.5 METHODS

**Materials.** Miller LB media (BD Difco) was used for growth of bacterial strains. *E. coli* BL21 Gold (DE3) (Agilent) was used as the host strain for both cloning and expression of recombinant proteins. Plasmid pET28a(+) was purchased from Novagen. HotStart Taq Mastermix (Denville) was used for gene amplification. *Phusion* DNA polymerase (Finnzymes), Taq DNA ligase (MCLab), and T5 Exonuclease (Epicenter) were purchased separately and used to make the assembly master mix (AMM) for cloning. Ni-NTA resin and miniprep reagents were purchased

from Qiagen. Primers were synthesized by IDT. All other chemicals were purchased from Sigma Aldrich unless otherwise noted.

**Mutagenesis.** 100 ng of plasmid containing the gene of interest was subjected to 32 cycles of the PCR protocol outlined in the manual of PfuUltra II Fusion HS DNA Polymerase (Agilent Cat. No. 600670). 50  $\mu$ L of the completed PCR reaction was treated with 1  $\mu$ L DpnI (New England Biolabs) for 2 hours at 37°C, and then 3  $\mu$ L was used to transform 100  $\mu$ L of *E. coli* BL21 Gold (DE3) cells. Transformants were selected on LB-agar containing 50  $\mu$ g/mL kanamycin. Mutations were verified by sequencing the plasmid using universal primers found upstream and downstream of the gene.

**Expression and Purification.** All *E. coli* strains were grown at 37°C in LB-media with 50  $\mu$ g/mL kanamycin. 1 L of LB-media was inoculated with 5 mL of overnight starter culture. Protein expression was induced during log phase ( $OD_{600}$  0.5-0.8) with 0.5 mM IPTG. After 20 hours, cells were pelleted, resuspended in 12 mL of buffer A (50 mM Tris HCl [pH 7.5], 100 mM NaCl), lysed by sonication, and cell debris was removed by centrifugation at 30,000xg for 20 min. The lysate was mixed with 3 mL of a Ni-NTA slurry and incubated at 4°C with gentle mixing. After 1 hour, the lysate mixture was applied to a column and the Ni-NTA beads were washed 3 times with 25 mL buffer A containing 10 mM imidazole. Protein was then eluted with 4 mL buffer A containing 250 mM imidazole.

**Isoprenol Analysis by GC-FID Assay.** A 50  $\mu$ L sample of IMAC eluate was added to a 150  $\mu$ L reaction mixture consisting of 50 mM bis-tris propane buffer [pH 8.5], 10 mM (R)-mevalonate, 20 mM ATP, 10 mM KCl, and 5 mM MgCl<sub>2</sub>. The reaction was incubated for 24 hours at 37°C for

MDC variants and 24 hours at 55°C for MMD and M3K variants. The reaction was extracted with 200  $\mu$ L hexanes. Five  $\mu$ L of the hexanes layer was injected into a HP5890 Series II Gas Chromatograph (flame ionization detector) connected to a HP-INNOWAX column (0.320 mm x 30 m, Agilent). The carrier gas was helium with a flow rate of 5 mL/min. Initial oven temperature was set to 70°C for 2 min, followed by a ramp of 20°C/min for 1 min, and finally a ramp at 50°C/min to a final temperature of 200°C, which was held for 1 min. The inlet was kept at 250°C and the detector at 330°C. Isoprenol eluted at 2.98 min and the sample concentration was determined by comparison to a standard. In order to test activity of the enzyme variants on mevalonate 3-phosphate, 10  $\mu$ g of mevalonate 3-kinase (Ta1305) was added to the reaction mixture and incubated at 55°C for 1 hour prior to adding the IMAC eluate.

**M3K Crystallization.** A stock of mutant M3K in 50 mM Tris-HCl [pH 7.5] and 100 mM NaCl was concentrated to 10.5 mg/ml using a 30 kDa cutoff Amicon Ultra-15 Centrifugal Filter (Millipore). A broad screen was setup using a 288 drop experiment. Crystals appeared directly in the screening tray after a few days at room temperature in a condition containing 0.1 M sodium acetate [pH 4.6], 0.2 M ammonium sulfate, and 30% (w/v) PEG 4000 as the reservoir solution. Crystals were cryo-protected by a quick soak in a solution consisting of 65% reservoir solution, 35% (v/v) glycerol, then flash frozen in a cryogenic nitrogen stream and maintained at 100 K for data collection.

**X-Ray Data Collection and Refinement.** X-ray diffraction data for the mutant protein was collected to 2.1 Å in-house on a Rigaku FR-E rotating anode X-ray source, using CuK $\alpha$  radiation ( $\lambda = 1.5418$  Å) and an R-AXIS HTC imaging plate detector. Reduction and scaling of data were



performed using XDS/XSCALE.<sup>31</sup> The structure was solved by molecular replacement using the *apo* structure of mevalonate 3-kinase (PDB: 4RKP) and refined using phenix.refine.<sup>32</sup>

**Construction of pSELECT-03.** The selection plasmid was constructed using pJBEI-6409 as a parent (30). This plasmid, as well as pMBI, were obtained from the plasmid repository at DNASU (Arizona State University) and originated at the Joint Bioenergy Institute. Farnesol kinase from *Arabidopsis thaliana* (Accession: NP\_200664.1) was codon optimized and synthesized by IDT. IP kinase (Accession: WP\_010900530.1) was amplified from *Thermoplasma acidophilum* genomic DNA. The construction was accomplished in two phases: (1) deletion of the third operon of pJBEI-6409 and (2) replacement of the second (except IDI) with IP kinase and farnesol kinase. The sequence of pJBEI-6409 from the beginning of the LacUV5 promoter until the end of HMGR was left intact. All modifications were made using the Gibson assembly method.<sup>33</sup> The fully annotated plasmid sequence of pSELECT-03 is available as a .dna file in the supplement.

**Dilution Series for *E. coli* Survival on Fosmidomycin.** We grew the *E. coli* constructs in LB liquid culture overnight and standardized the culture to an OD<sub>600</sub> of 0.1. A dilution series was made using a 96 well plate. 90  $\mu$ L of LB was aliquoted into each well and 10  $\mu$ L of the 0.1 OD culture was added to the first well. After mixing, 10  $\mu$ L of this dilution was added to the next well. This 10 fold dilution was repeated 6 times. A multi-channel pipette was used to spot 5  $\mu$ L of each dilution on an LB agar plate containing 50  $\mu$ M fosmidomycin and 40 mM isoprenol.

**Deletion of DXP Synthase via Lambda Red Recombineering.** Chemically competent *E. coli* BL21(DE3) was cotransformed pSELECT-03 and the pSIM-6 recombineering plasmid.<sup>34</sup> A

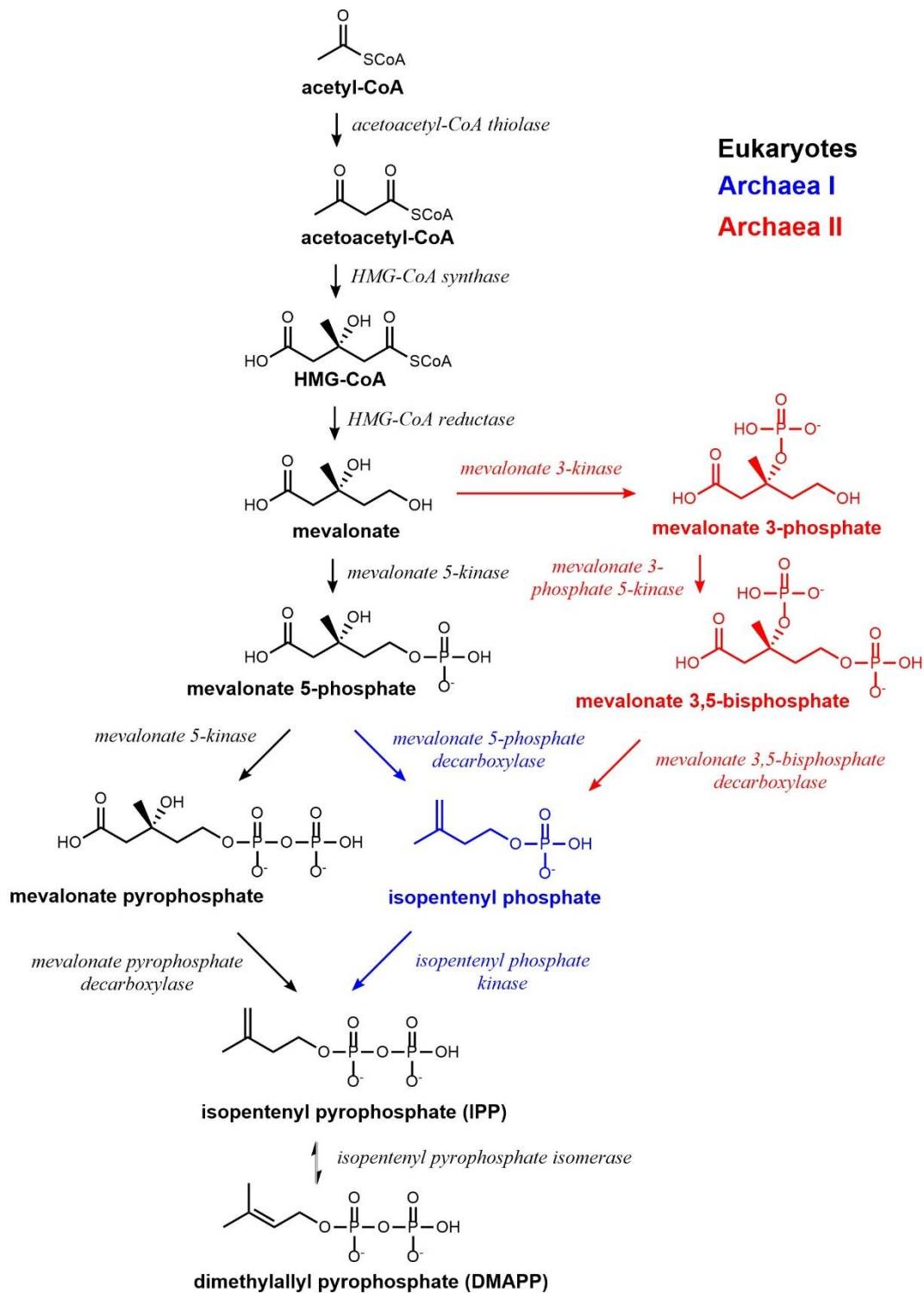
colony was picked and grown at 30°C to an OD600 of 0.5. The culture was transferred to 37°C for 30 min and then 42°C for 30 min to induce protein expression from pSIM-6. The heat-shocked cells were then made electrocompetent and 1 µg of PCR product was transformed. The PCR product was made by amplifying plasmid pKD4 with primers that incorporated the flanking the *dxs* gene of *E. coli*. The outgrowth phase of the electroporation used TB media supplemented with 2% glycerol (v/v) and 40 mM isoprenol. The resulting culture was plated on LB-agar with kanamycin and 40 mM isoprenol at 37°C. To remove the kanamycin cassette, the strain was made electrocompetent and transformed with plasmid pCP20.<sup>35</sup> Outgrowth used TB media with 2% glycerol and 40 mM isoprenol. This electroporation reaction was plated on ampicillin plates supplemented with 40 mM isoprenol and grown at 30°C. The next day colonies were re-streaked on identical plates and grown at 42°C to remove the temperature sensitive pCP20 plasmid. The phenotype of the resulting strain was confirmed by streaking on LB plates with no drug (no growth), isoprenol (growth), isoprenol + ampicillin (no growth), and isoprenol + kanamycin (no growth).

**Competent Cell Preparation.** The evolved selection strain was prepared fresh for each experiment. We found that storing competent cells of final-2 at -80°C severely reduced the transformation efficiency from approx.  $1 \times 10^5$  (fresh) to  $1 \times 10^2$  (frozen). A 5 mL LB culture containing 34 µg/mL chloramphenicol and 200 µM isoprenol was inoculated with a glycerol stock of final-2 and grown overnight (12 hours) in a test tube rotator at 37°C. The next day, 2 mLs of the culture was diluted into 50 mL of fresh LB with 34 µg/mL chloramphenicol and 200 µM isoprenol and grown at 37°C with shaking for approximately 4 hours. When the OD600 reached 0.5, the culture was put on ice for 15 minutes. The culture was spun down for 10 min at 3,000g in a pre-chilled centrifuge and the supernatant was removed. The pellet was gently re-dissolved in 40 mLs of ice cold 0.1 M CaCl<sub>2</sub>. This was pelleted again at 3,000g for 10 min in a



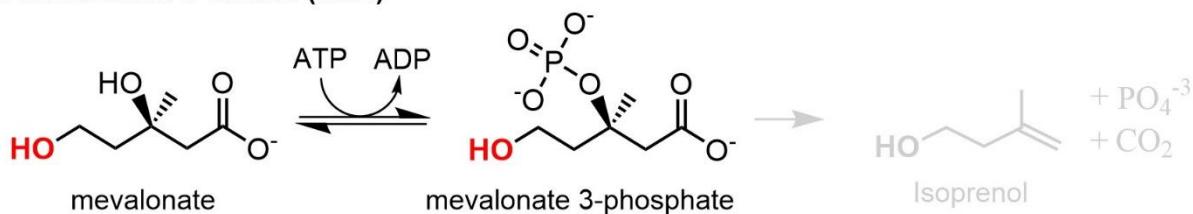
the supernatant (SOC with 1 mM isoprenol) was completely removed. The pellet was re-suspended in 200uL of LB and the entire amount was plated onto an LB plate containing 34 µg/mL chloramphenicol, 10 mM IPTG, and 50 µg/mL kanamycin.

**Selection.** 50 tubes each containing 50 µL of competent final-2 cells were mixed with 1 µL of the mutated plasmid library stock prepared by Synbio tech (70 ng/µL). Each tube was individually transformed using the method above. An additional 3 transformations were included in parallel which used un-mutated plasmid stock to determine the background mutation rate. A control plate received a 1/10 dilution of cells and contained isoprenol and kanamycin which selected only for the addition of the plasmid and allowed us to calculate how many mutants were screened in the experiment. Each plate screened 4000 mutants. The plates were incubated at 37°C for 5 days and checked daily. Colonies were picked, mini-prepped, sequenced, and re-transformed to determine if the ability to grow without isoprenol supplementation travels with the plasmid.

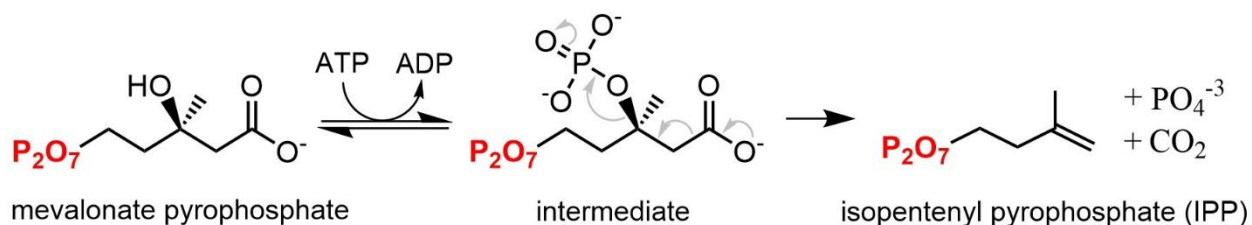


**FIGURE 4.1.** Mevalonate pathways. Eukaryotes use the pathway shown in black. Archaea use the same pathway up until mevalonate 5-phosphate then follow the blue pathway. Extreme acidophiles of the archaeal order thermoplasmatales use the red pathway after mevalonate.

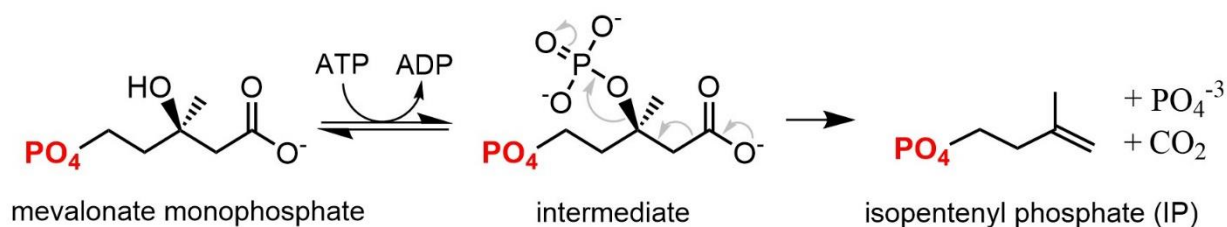
### A. Mevalonate-3-Kinase (M3K)



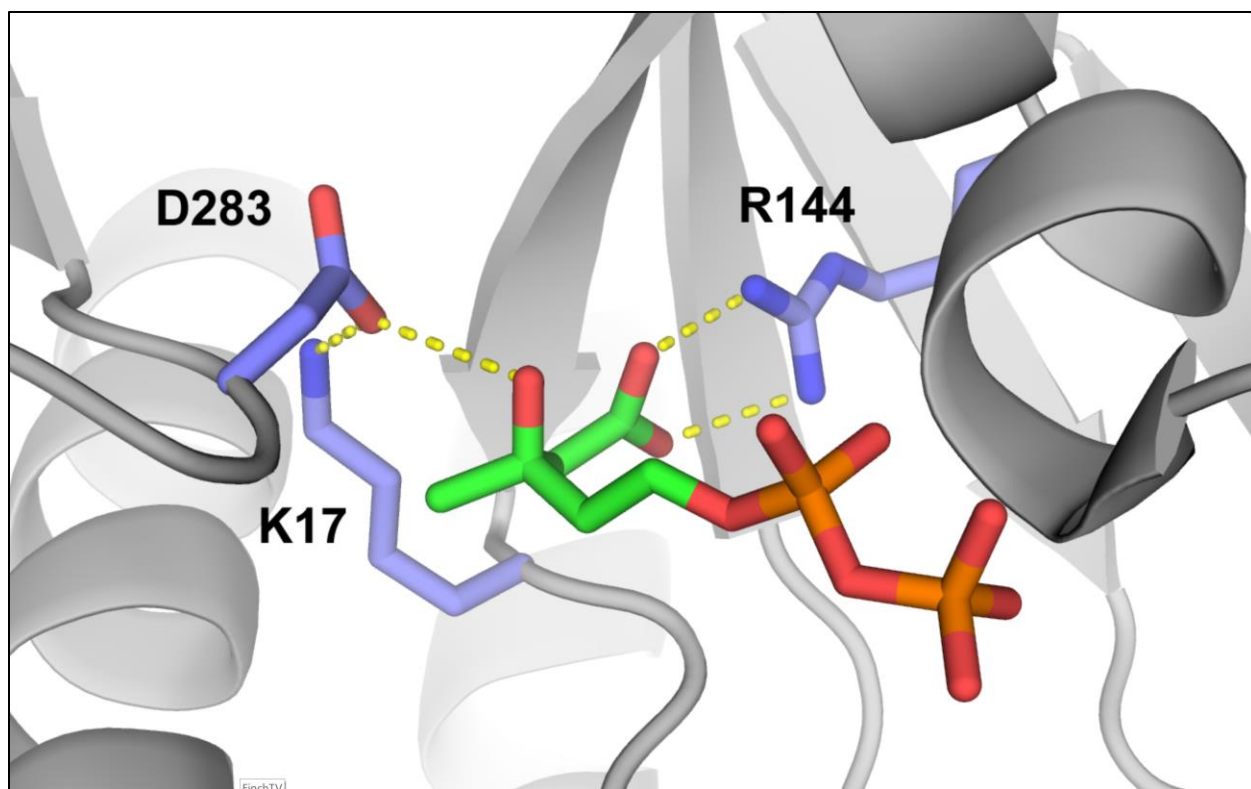
### B. Mevalonate Pyrophosphate Decarboxylase (MDC)



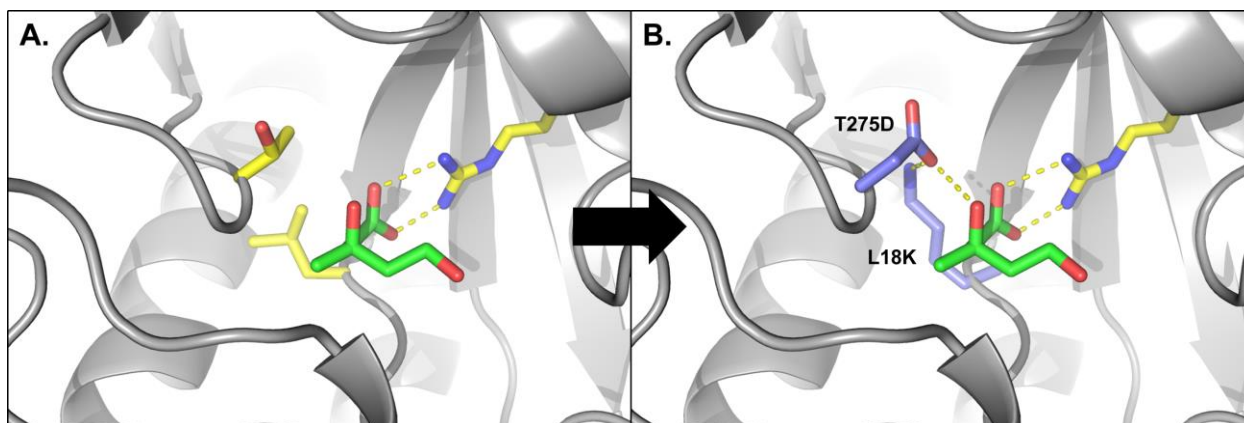
### C. Mevalonate Monophosphate Decarboxylase (MMD)



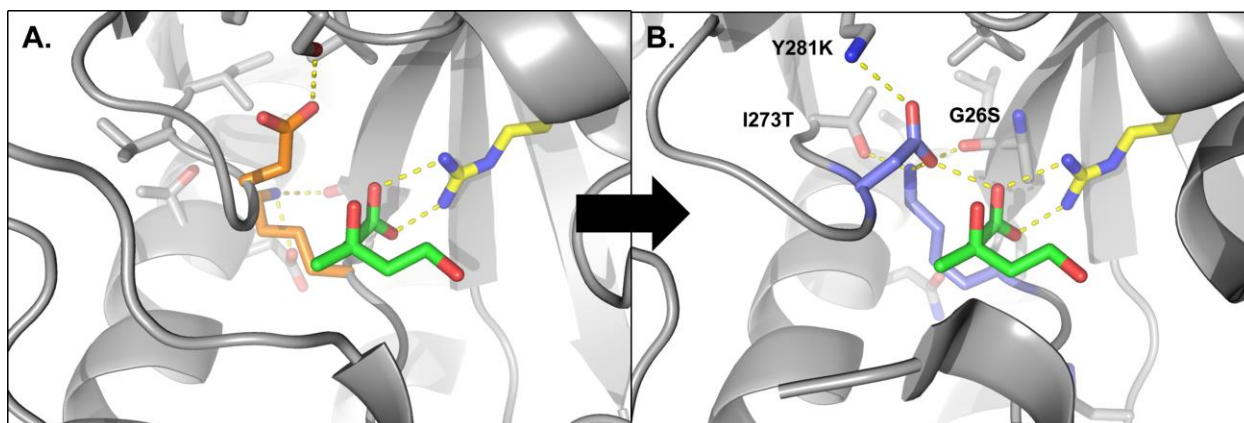
**FIGURE 4.2.** Reaction scheme for mevalonate pathway enzymes. (A.) The reaction scheme for mevalonate 3-kinase. (B.) The reaction scheme for mevalonate pyrophosphate decarboxylase (C.) The reaction scheme for mevalonate monophosphate decarboxylase. The moiety connected to carbon 5 of mevalonate is highlighted in red. M3K does not decarboxylate mevalonate 3-phosphate into isoprenol (shown in grey). Such is the goal of this work.



**FIGURE 4.3** Catalytic residues of *S. epidermidis* mevalonate pyrophosphate decarboxylase (PDB: 4DU7). All MDCs characterized to date contain the aspartate, lysine, and arginine in the orientations shown here.

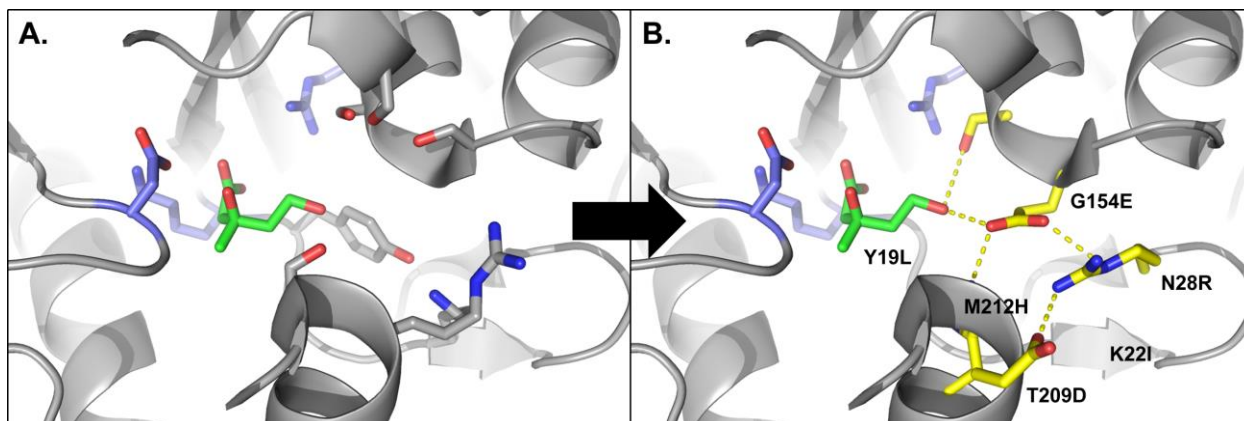


**FIGURE 4.4** Adding decarboxylation residues to mevalonate 3-kinase. (A.) The active site of *wildtype* M3K (PBD: 4RKP) (B.) The two mutations shown (T275D and L18K) were modelled with PyMOL and introduced into M3K by site-directed mutagenesis in order to introduce decarboxylase activity.

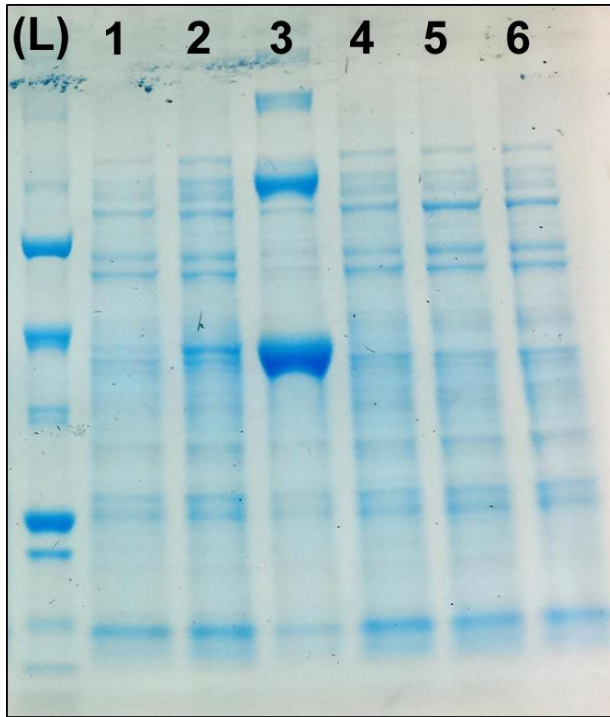


**FIGURE 4.5** Orientation of T275D and L18K in the M3K double mutant. (A.) The actual orientation of the aspartate and lysine in the T275D/L18K mutant of M3K as seen in our crystal structure. (B.) The active site of *S. epidermidis* MDC is shown with the residues surrounding the Asp/Lys pair in grey. The proposed mutations to correctly orient the Asp/Lys sidechains in M3K are shown in stick representation and labelled in black.

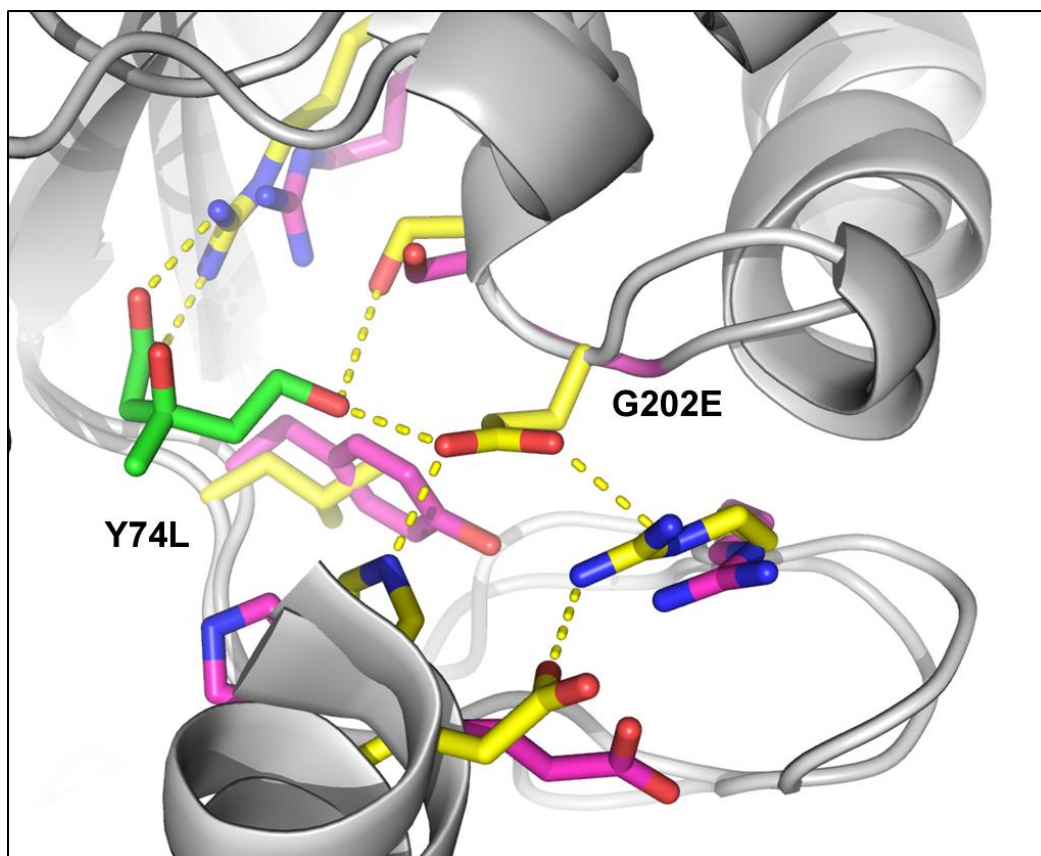




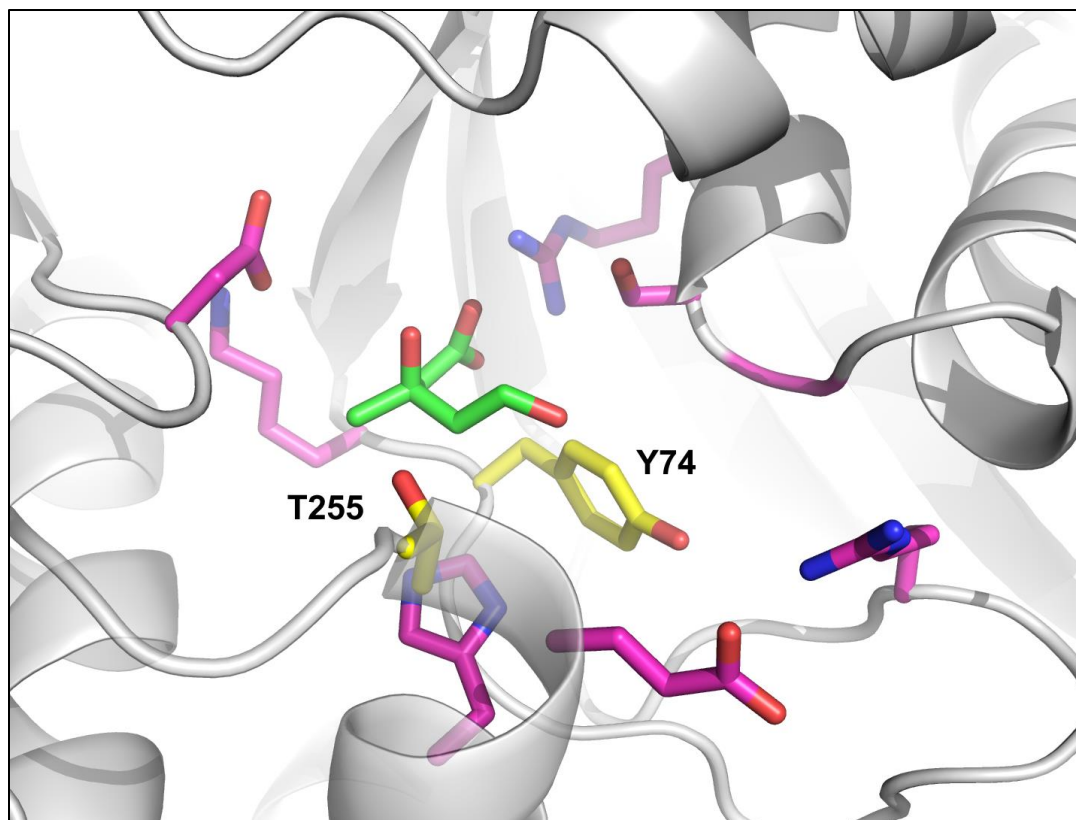
**FIGURE 4.6** Adding residues that bind mevalonate to mevalonate pyrophosphate decarboxylase (MDC). (A.) The active site of MDC from *S. epidermidis* (PDB: 4DU7) with mevalonate modeled into the active site. (B.) The *S. epidermidis* MDC active site with residues from M3K modeled in the active site (yellow).



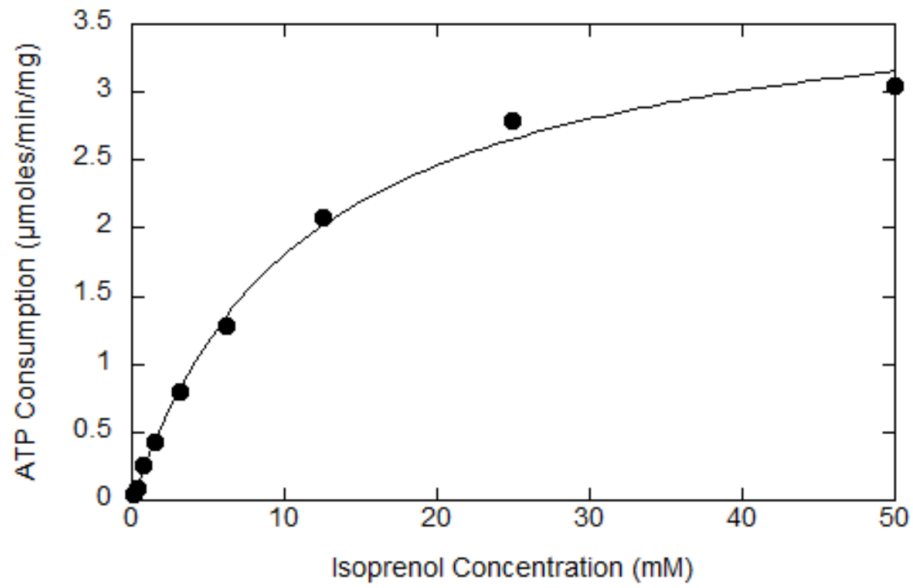
**FIGURE 4.7** SDS-PAGE analysis of MDC mutants. The MDC variant containing all six mutations for binding mevalonate (Y19L, K22I, G154E, N28R, T209D, M212H) was insoluble. We made single mutations reverting each of the introduced mutations to wildtype residues. (L) is protein ladder and the numbers 1-6 correspond to reverting the Y19L, K22I, G154E, N28R, T209D, and M212H mutations, respectively. Only when G154E (lane 3) is reverted to glycine does the enzyme become soluble.



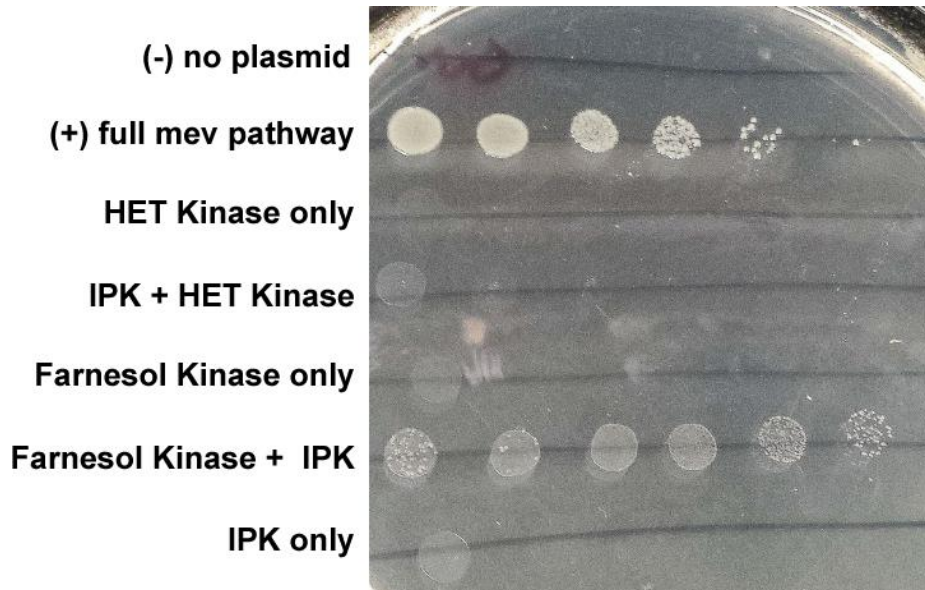
**FIGURE 4.8** Structural superposition of MMD with M3K with substrate binding residues shown in stick representation. Substrate binding residues of M3K are shown in yellow and the residues of MMD are shown in magenta. The substrate binding site of M3K can be mimicked in MMD with only two mutations: G202D and Y74L (PDBs: 4RKP and MMD PHYRE model).



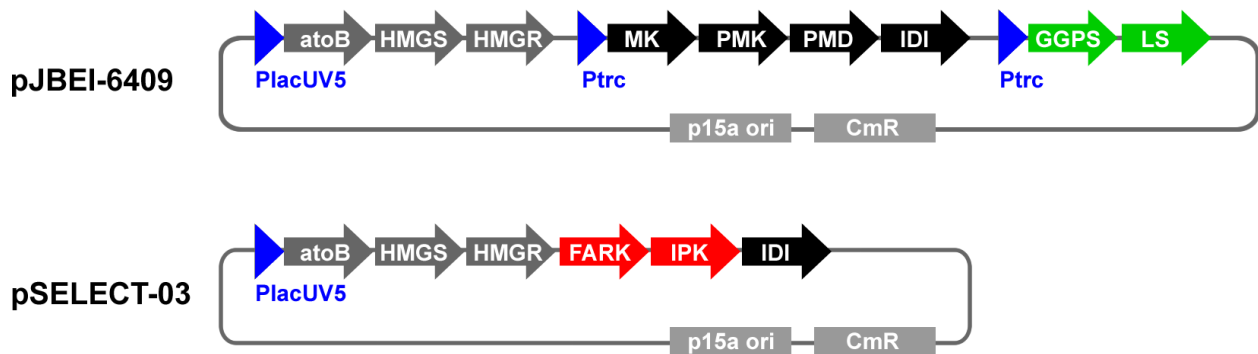
**FIGURE 4.9** Alternative variants of *Roseiflexus castenholzii* MMD. Y74 was mutated to Ala, Leu or Phe, and T255 was mutated to Glu, Asp, or Ser. None of these mutants were active on mevalonate. (MMD PHYRE model shown).



**FIGURE 4.10.** Michaelis-Menten plot for the activity of *E. coli* hydroxyethylthiazole kinase on isoprenol as measured by ATP consumption in a coupled assay.

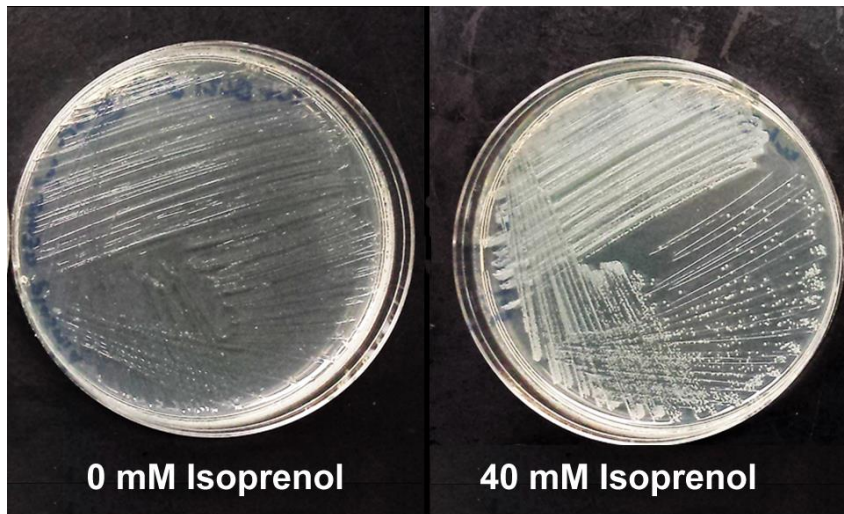


**FIGURE 4.11.** Farnesol kinase and IP Kinase allow *E. coli* to utilize isoprenol. Seven *E. coli* constructs were plated horizontally in a dilution series on an LB-agar plate containing 50  $\mu$ M fosmidomycin, 40 mM isoprenol, and 50  $\mu$ M IPTG. When IP kinase and farnesol kinase were combined, the strain was able to grow in the presence of the fosmidomycin antibiotic due to its utilization of the exogenous isoprenol.

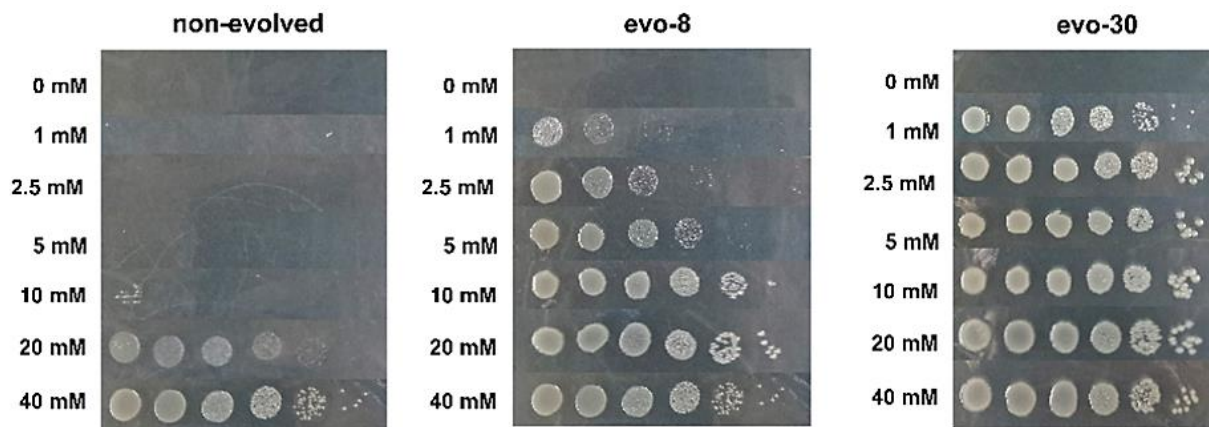


**FIGURE 4.12** Schematic of selection plasmid. Plasmid pSELECT-03 (bottom) was constructed by first deleting the third operon from plasmid pJBEI-6409 (green), and then replacing most of the second operon (black), with farnesol kinase and IP kinase (red).

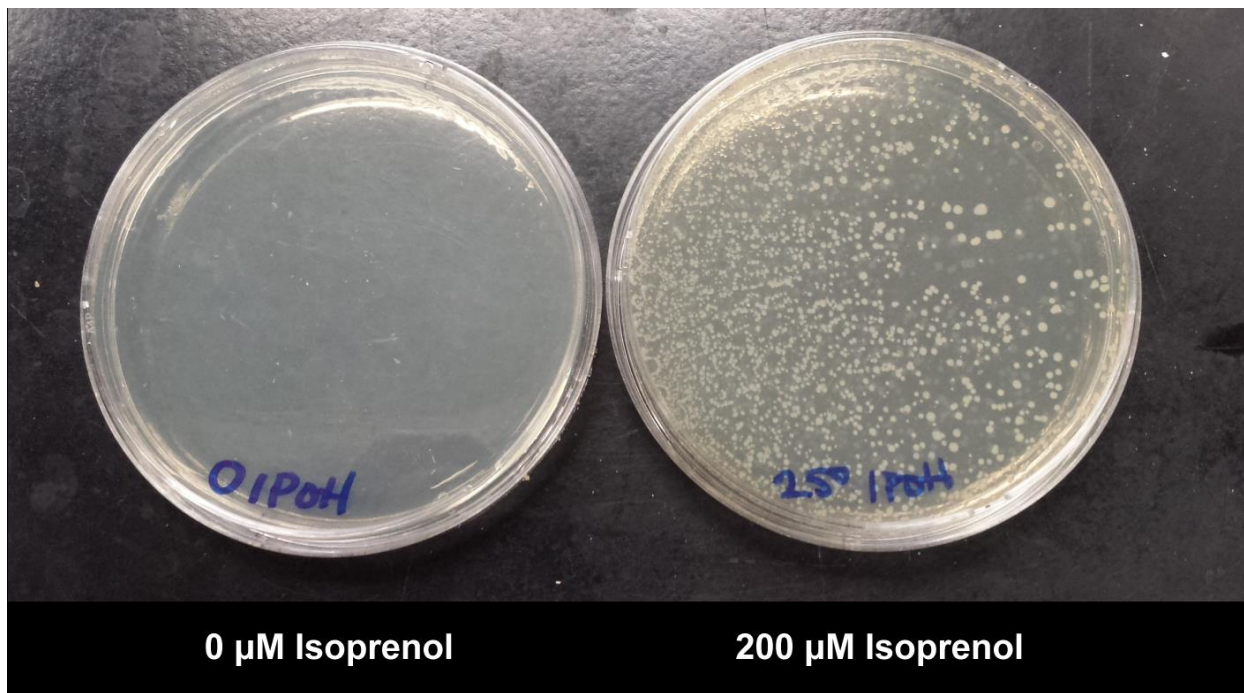




**FIGURE 4.13.** Growth of *E. coli* strain containing plasmid pSELECT-03. This strain did not grow on LB-fosmidomycin plates with no isoprenol (left side), however robust growth was observed in 24 hours when the plate was supplemented with 40 mM isoprenol (right side).

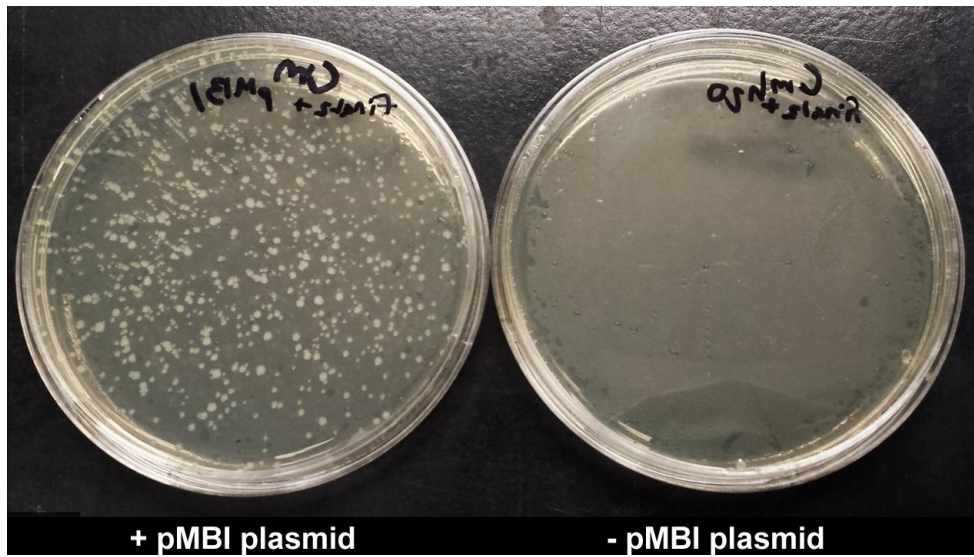


**FIGURE 4.14.** Directed evolution of selection strain. *E. coli* BL21(DE3)  $\Delta dxs$  pSELECT-03 was subcultured in decreasing concentrations of isoprenol. Serial dilutions are shown in horizontal rows for each strain grown at various concentrations of isoprenol. The unevolved strain could not grow below 20 mM isoprenol, however after 30 rounds sub-culturing it could grow at 1 mM isoprenol.

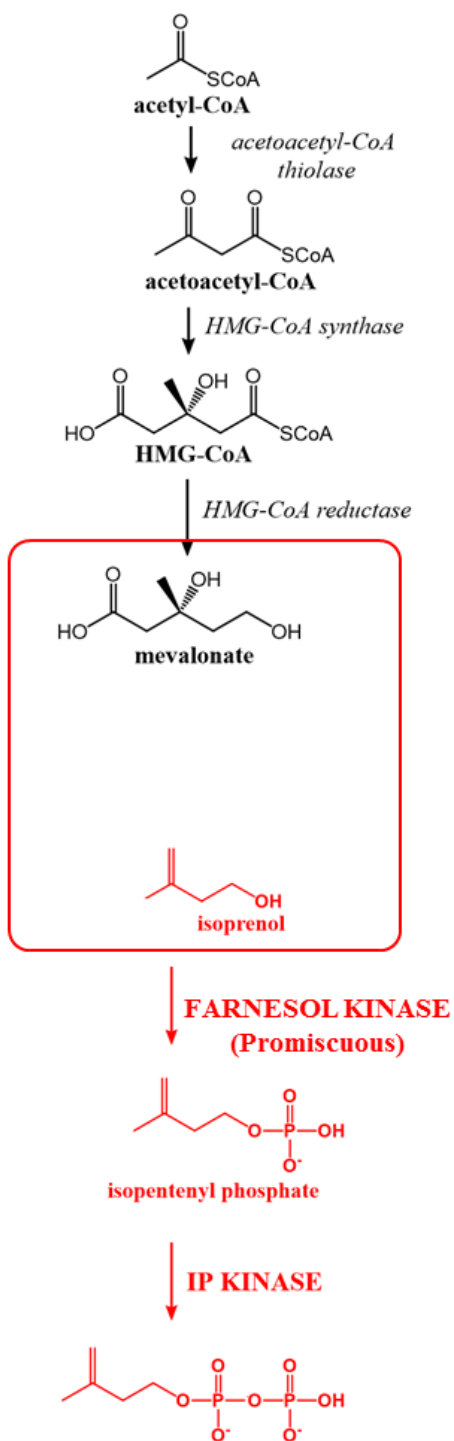


**FIGURE 4.15.** The final selection strain. The selection strain final-2 is *E. coli* BL21(DE3)  $\Delta\text{dxs}$  pSELECT-03 evolved to require only 200  $\mu\text{M}$  exogenous isoprenol to grow to large colonies in 24. In this image final-2 is shown growing on plates containing 250  $\mu\text{M}$  isoprenol.

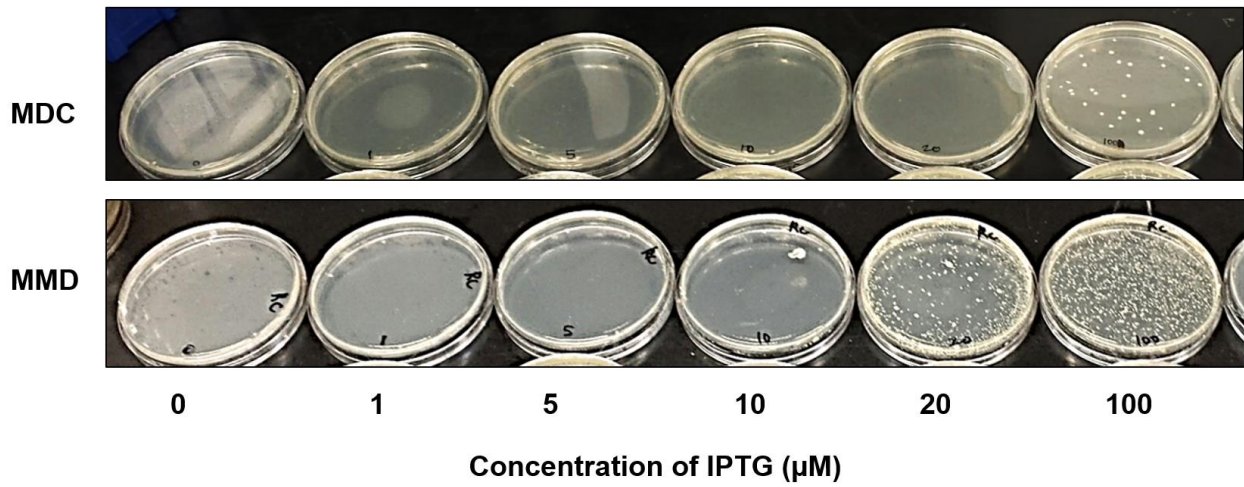




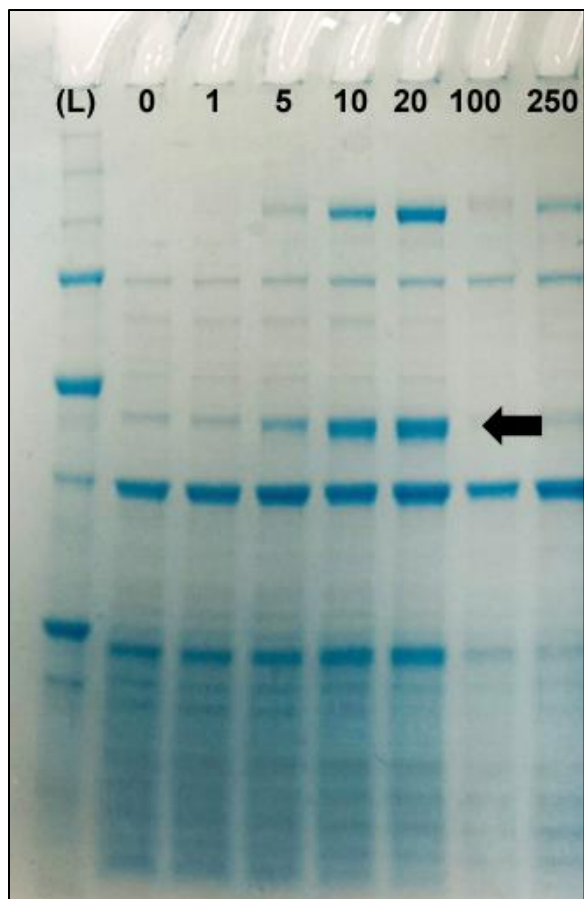
**FIGURE 4.16.** Rescue with plasmid pMBI. When the selection strain (final-2) is transformed with plasmid pMBI (which converts mevalonate to IPP) it gains the ability to grow on LB without isoprenol (left side). When final-2 is transformed with H<sub>2</sub>O instead of pMBI, it cannot grow (right side). There was no IPTG used in this experiment.



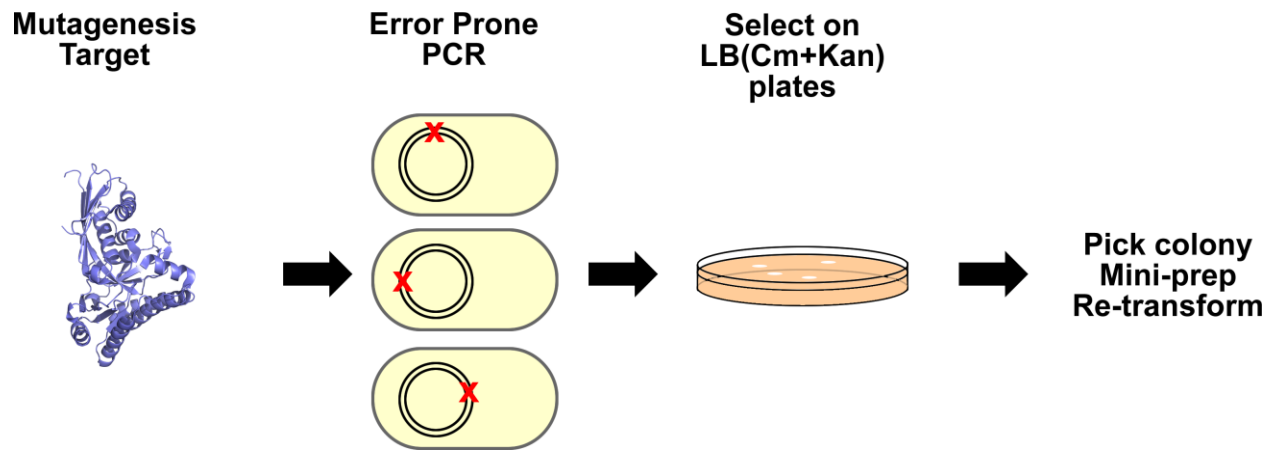
**FIGURE 4.17.** Selection strain schematic. The selection strain (final-2) contains a deletion in the *dxs* gene which blocks endogenous IPP production. The pSELECT-03 plasmid contains the artificial pathway shown here. It produces mevalonate (black) and utilize isoprenol (red). The creation of mevalonate decarboxylase would complete the step shown in the red square.



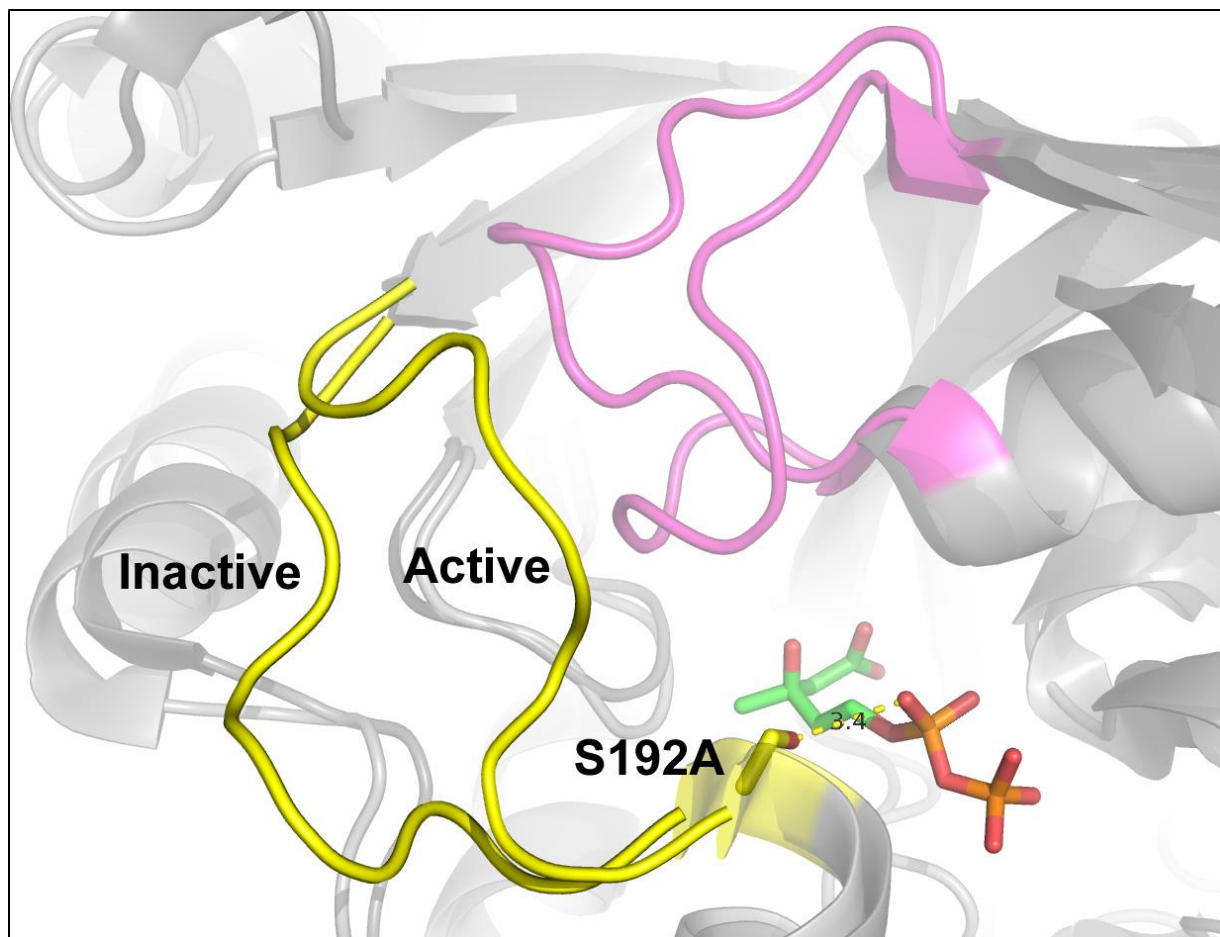
**FIGURE 4.18.** Titration of IPTG into the selection strain (final-2) containing wild type decarboxylases. When final-2 is transformed with wildtype decarboxylase and plated on LB plates with chloramphenicol, there is no detectable growth. If IPTG is added to the plate media, rescue with the native MMD occurs above 20 μM IPTG and rescue with native MDC occurs at 100 μM IPTG. To avoid this issue the selection was conducted at 10 μM IPTG.



**FIGURE 4.19.** MMD expression at various IPTG concentrations. In this expression test, final-2 was grown in LB liquid culture with chloramphenicol, kanamycin, 200 μM isoprenol, and IPTG (concentration indicated across the top). Significant levels of MMD (black arrow) are produced at 10 μM IPTG. The cultures were sonicated, heated at 60°C for 1 hour, and spun down prior to running gel to remove background of *E. coli* proteins (MMD is stable at 60°C).



**FIGURE 4.20.** Selection strategy. MMD and MDC were genetically mutated using error prone PCR. The resulting plasmid library was transformed into the final-2 selection strain and plated on LB-agar plates with chloramphenicol and kanamycin. Any colonies that grew were picked, mini-prepped, sequenced, and isolated plasmids were re-transformed into fresh final-2 cells.



**FIGURE 4.21.** Proposed mechanism that requires a 5-phosphate moiety. The active site of *S. epidermidis* MDC (PDB: 4DU7) is overlaid with *Legionella pneumophila* MDC (PDB: 3LTO). Serine 192 is shown making a polar contact to the pyrophosphate group of mevalonate pyrophosphate. Two loops bordering the active site are highlighted in yellow and purple. The purple loop is known to clamp down when ATP is bound, and the active confirmation of the yellow loop is known to face inwards.<sup>11,19</sup> We believe a polar contact via S192 to the 5-pyrophosphate moiety is required for closure of the yellow loop and activation of the enzyme.

ID	Mutations	Generation	Soluble?	Activity
<b>A. Mevalonate 3-Kinase Variants (M3K)</b>				
1	<i>wildtype</i>	-	Y	N.D.
2	L18K	1	Y	N.D.
3	T275D	1	Y	N.D.
4	L18K, T275D	1	Y	N.D.
5	L18K, T275D, G26S	1	Y	N.D.
6	L18K, T275D, I273T	1	Y	N.D.
7	L18K, T275D, Y281K	1	Y	N.D.
<b>B. Mevalonate Pyrophosphate Decarboxylase Variants (MDC)</b>				
8	<i>wildtype</i>	-	Y	N.D.
9	M212H	1	Y	N.D.
10	Y19L	1	Y	N.D.
11	G154E	1	Y	N.D.
12	N28R	1	Y	N.D.
13	K22I	1	Y	N.D.
14	T209D	1	Y	N.D.
15	Y19L, K22I, N28R, G154E, T209D, M212H	1	N	N.D.
16	Y19L, K22I, N28R, G154E, T209D	2	N	N.D.
17	Y19L, K22I, N28R, G154E, M212H	2	N	N.D.
18	Y19L, K22I, N28R, T209D, M212H	2	Y	N.D.
19	Y19L, K22I, G154E, T209D, M212H	2	N	N.D.
20	Y19L, N28R, G154E, T209D, M212H	2	N	N.D.
21	K22I, N28R, G154E, T209D, M212H	2	N	N.D.
22	G154E, Y19L	3	N	N.D.
23	G154, Y19A	3	Y	N.D.
24	G154E, N28R	3	Y	N.D.
25	G154E, Y19A, N28R	3	N	N.D.
26	G154E, Y19L, N28R	3	N	N.D.
27	G154E, M212H	3	Y	N.D.
28	G154E, Y19A, M212H	3	Y	N.D.
29	G154E, Y19L, M212H	3	N	N.D.
30	G154E, Y19L, K22I, N28R,	3	N	N.D.
<b>C. Mevalonate Monophosphate Decarboxylase Variants (MMD)</b>				
31	<i>wildtype</i>	-	Y	N.D.
32	Y74L, G202E	1	Y	N.D.
33	Y74L, G202D	1	Y	N.D.
34	Y74A, G202E	1	Y	N.D.
35	Y74A, G202D	1	Y	N.D.
36	G202D	1	Y	N.D.
37	G202E	1	Y	N.D.
38	Y74L	1	Y	N.D.
39	Y74A	1	Y	N.D.
40	Y74F	1	Y	N.D.
41	T255S	2	Y	N.D.
42	T255D	2	Y	N.D.
43	T255E	2	Y	N.D.
44	Y74D	2	Y	N.D.
45	Y74E	2	Y	N.D.

**Table 4.1.** Enzyme variants generated to convert mevalonate to isoprenol. (A.) variants of mevalonate 3-kinase (M3K). (B.) variants of mevalonate pyrophosphate decarboxylase (MDC) (C.) variants of mevalonate monophosphate decarboxylase (MMD). Protein solubility is marked as either yes (Y) or no (N). Activity on mevalonate or mevalonate 3-phosphate was not detected (N.D.) for any sample.

Mevalonate Monophosphate Decarboxylase Selection Hits				
Hit Number	Upstream Mutations	MMD Mutations	Soluble?	Activity
1	LacI S97L	A36T, A401E	Y	N.D.
2	Lac Operator G5011A	R39C	Y	N.D.
3	LacI R294C	A142V, H223Y	Y	N.D.
4	Lac Operator G5011A	R39C	Y	N.D.

**Table 4.2.** Decarboxylase variants. The selection experiments yielded 4 plasmids which conferred the ability to grow without isoprenol. Each contained both a mutation upstream of the decarboxylase and a mutation in the decarboxylase. All four were tested *in vitro* and did not show any activity on mevalonate or mevalonate 3-phosphate.



## 4.5 REFERENCES

1. Kung, Y., Runguphan, W. & Keasling, J. D. From Fields to Fuels: Recent Advances in the Microbial Production of Biofuels. *ACS Synth. Biol.* **1**, 498–513 (2012).
2. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379 (2010).
3. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production. *Science* **314**, 1565–1568 (2006).
4. Connor, M. R., Cann, A. F. & Liao, J. C. 3-Methyl-1-butanol production in *Escherichia coli*: random mutagenesis and two-phase fermentation. *Appl. Microbiol. Biotechnol.* **86**, 1155–1164 (2010).
5. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
6. Goldstein, J. L. & Brown, M. S. Regulation of the Mevalonate Pathway. *Nat. Lond.* **343**, 425–30 (1990).
7. Miziorko, H. M. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* **505**, 131–143 (2011).
8. Zheng, Y. *et al.* Metabolic engineering of *Escherichia coli* for high-specificity production of isoprenol and prenol as next generation of biofuels. *Biotechnol. Biofuels* **6**, 57 (2013).
9. George, K. W. *et al.* Metabolic engineering for the high-yield production of isoprenoid-based C5 alcohols in *E. coli*. *Sci. Rep.* **5**, 11128 (2015).

10. Chou, H. H. & Keasling, J. D. Synthetic Pathway for Production of Five-Carbon Alcohols from Isopentenyl Diphosphate. *Appl. Environ. Microbiol.* **78**, 7849–7855 (2012).
11. Vinokur, J. M. *et al.* Structural analysis of mevalonate-3-kinase provides insight into the mechanisms of isoprenoid pathway decarboxylases. *Protein Sci.* **24**, 212–220 (2015).
12. Vinokur, J. M., Korman, T. P., Cao, Z. & Bowie, J. U. Evidence of a novel mevalonate pathway in archaea. *Biochemistry* **53**, 4161–4168 (2014).
13. Lindberg, M., Yuan, C., Dewaard, A. & Bloch, K. On the mechanism of formation of isopentenylpyrophosphate. *Biochemistry* **1**, 182–188 (1962).
14. Dellas, N., Thomas, S. T., Manning, G. & Noel, J. P. Discovery of a metabolic alternative to the classical mevalonate pathway. *eLife* **2**, e00672 (2013).
15. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
16. Campos, N. *et al.* Escherichia coli engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *Biochem. J.* **353**, 59–67 (2001).
17. Krepiy, D. & Miziorko, H. M. Identification of active site residues in mevalonate diphosphate decarboxylase: Implications for a family of phosphotransferases. *Protein Sci.* **13**, 1875–1881 (2004).
18. Barta, M. L. *et al.* Crystal structures of Staphylococcus epidermidis mevalonate diphosphate decarboxylase bound to inhibitory analogs reveal new insight into substrate binding and catalysis. *J. Biol. Chem.* **286**, 23900–23910 (2011).

19. Barta, M. L., McWhorter, W. J., Miziorko, H. M. & Geisbrecht, B. V. Structural basis for nucleotide binding and reaction catalysis in mevalonate diphosphate decarboxylase. *Biochemistry* **51**, 5611–5621 (2012).
20. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802 (2003).
21. Fitzpatrick, A. H., Bhandari, J. & Crowell, D. N. Farnesol kinase is involved in farnesol metabolism, ABA signaling and flower development in *Arabidopsis*. *Plant J. Cell Mol. Biol.* **66**, 1078–1088 (2011).
22. Marlière, P. & Anissimova, M. Process for the enzymatic preparation of isoprene from isoprenol. (2014).
23. Chen, M. & Poulter, C. D. Characterization of Thermophilic Archaeal Isopentenyl Phosphate Kinases. *Biochemistry* **49**, 207–217 (2010).
24. Kuzuyama, T. & Seto, H. Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **88**, 41–52 (2012).
25. Hemmerlin, A., Tritsch, D., Hammann, P., Rohmer, M. & Bach, T. J. Profiling of defense responses in *Escherichia coli* treated with fosmidomycin. *Biochimie* **99**, 54–62 (2014).
26. Alonso-Gutierrez, J. *et al.* Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. *Metab. Eng.* **19**, 33–41 (2013).
27. Agranoff, B. W., Eggerer, H., Henning, U. & Lynen, F. Biosynthesis of Terpenes VII. Isopentenyl Pyrophosphate Isomerase. *J. Biol. Chem.* **235**, 326–332 (1960).
28. Kirby, J. *et al.* Enhancing terpene yield from sugars via novel routes to 1-deoxy-d-xylulose 5-phosphate. *Appl. Environ. Microbiol.* AEM.02920-14 (2014). doi:10.1128/AEM.02920-14

29. Byres, E., Alphey, M. S., Smith, T. K. & Hunter, W. N. Crystal structures of *Trypanosoma brucei* and *Staphylococcus aureus* mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity. *J. Mol. Biol.* **371**, 540–553 (2007).
30. Vinokur, J. M., Cummins, M. C., Korman, T. P. & Bowie, J. U. An Adaptation To Life In Acid Through A Novel Mevalonate Pathway. *Sci. Rep.* **6**, 39737 (2016).
31. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
32. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
33. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
34. Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).
35. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

## CHAPTER 5

### A Novel Protein Co-Expression Technology Using Thermosomes and Prefoldins

#### 5.1 ABSTRACT

Production of recombinant proteins in *E. coli* is one of the cornerstones of modern synthetic biology. The ability to quickly and cheaply produce milligram quantities of soluble protein is often essential for crystallographic and biochemical studies. However, it is a common occurrence for recombinant proteins to suffer from low or no soluble expression due to misfolding or aggregation. Here we demonstrate for the first time a method of solubilizing recombinant proteins using an archaeal chaperone system composed of thermosomes and prefoldins. First we show that an archaeal enzyme, mevalonate 3,5-bisphosphate decarboxylase (MBD) is completely inactive when expressed on its own, however co-expression with thermosomes and prefoldins produces active MBD. The level of activity is ~30 times higher than co-expression with the GroEL/GroES chaperone system. We expanded the analysis beyond our model protein by testing MBD homologs. Soluble expression of four out of five homologs were improved with the thermosome co-expression system. Collaborators are now evaluating our system on a test set of 48 poorly expressing human and archaeal proteins to determine if it can be used as a general strategy to quickly and efficiently produce soluble recombinant proteins in *E. coli*.

## 5.2 INTRODUCTION

Expressing soluble protein in a model organism like *E. coli* has become the most common method for producing large quantities of pure protein for scientific research.<sup>1,2</sup> While the majority of recombinant proteins can be produced in soluble form using *E. coli*, a significant proportion of these targets misfold, aggregate, or form insoluble inclusion bodies.<sup>3,4</sup> There are dozens of methods that can help researchers express recombinant proteins more efficiently to increase yield and stability. Traditional methods include varying expression time, temperature, and amount of inducing agent added to the cultures (typically IPTG, isopropyl  $\beta$ -D-1-thiogalactopyranoside).<sup>2</sup> For more difficult targets, the protein of interest can be fused to a solubilization tag like maltose binding protein, an export signal can be appended to the sequence to send the protein to the periplasm, the target protein can be expressed in an engineered *E. coli* strain optimized for production of toxic proteins, or the recombinant protein can be co-expressed with molecular chaperones.<sup>5-10</sup>

All known organisms contain molecular chaperones, proteins that assist in protein folding.<sup>11</sup> Chaperones are often called Hsp proteins (heat shock proteins) since they are over-expressed when the organism is grown at high temperature. The most well studied chaperone system is the GroEL/ES system of *E. coli*.<sup>12,13</sup> The GroEL/ES complex is a large cage-like structure made up of two stacked rings of GroEL proteins (57 kDa each). Unfolded polypeptide binds to the hydrophobic residues on the outer rim of the cage, then ATP binding induces conformational change, which is followed by the binding of a lid on the GroEL ring (Fig. 5.1A-B). The lid is composed of a 6-8 subunits of GroES proteins. Upon attachment of the lid, the unfolded protein enters the cage of the GroEL/ES complex. Inside the cage is where the physical constraints and unique environment promotes correct protein folding. After the protein has folded correctly, ADP is released, the GroES lid detaches, and the correctly folded protein is

released (Fig. 5.1D-E).<sup>12,13</sup> Numerous studies have shown that over-expressing GroEL/ES improves the yield of soluble recombinant protein and in some cases solubilizes proteins that cannot otherwise fold properly.<sup>8-10,14,15</sup>

GroEL/ES is an example of a group I chaperonin and is found in bacteria. Archaea and eukaryotes use a similar system based on group II chaperonins.<sup>16</sup> While both systems use a cage to trap unfolded proteins and promote proper folding inside a large complex, the properties and mechanism of group I and group II chaperonins are quite different (Fig 5.2). The eukaryotic analog of GroEL/ES, called the TCP-1 ring complex (TRiC), uses a cage composed of 8 unique proteins and is required to fold actin and tubulin.<sup>17</sup> In archaea, the GroEL/ES analog is called a thermosome and the cage is made up of six proteins, most often three  $\alpha$  and three  $\beta$  subunits.<sup>18</sup> The main difference between the chaperonin systems is that group I uses a separate protein as a lid (GroES), while group II have a built in lid (Fig 5.2A-B). Binding of the lid in group I chaperonins causes expansion of the cage, while closing of the built-in lid in group II chaperonins causes constriction of the cage. The chaperonins also contain different surface properties in their cages. The charged residues in the GroES cage are largely negatively charged, while the cavities in eukaryotic and archaeal chaperonins are largely positively charged (Fig. 5.2C-D). Additionally, the cavity of the thermosome is significantly more hydrophilic than that of its bacterial and eukaryotic counterparts (Fig 5.2E-F).<sup>19</sup>

One final difference is that group II chaperonins use prefoldins to deliver unfolded proteins to the cage, while bacteria do not contain a prefoldin homolog.<sup>20</sup> Prefoldins are small hexameric proteins with long  $\alpha$ -helices dangling from a core structural element thus resembling tentacles.<sup>21</sup> It functions as a transfer protein, binding unfolded proteins like actin and tubulin and

delivering them to the thermosome or TCP-1 ring complex. It does not use chemical energy and only binds unfolded protein.<sup>20</sup>

In this study, we show that co-expressing archaeal thermosomes and prefoldins in conjunction with the recombinant protein of interest improves the yield of soluble recombinant protein. We first establish this on a single model protein, mevalonate 3,5-bisphosphate decarboxylase (MBD), from *Thermoplasma acidophilum*.<sup>22</sup> We show that *T. acidophilum* MBD becomes active when co-expressed with thermosomes and prefoldins from *Ferroplasma acidarmanus*. We then expand the study to show that the method improves soluble protein expression in 4 out of 5 MBD homologs. Through a collaboration, we are now evaluating the ability of our system to be used with other recombinant targets by co-expressing thermosomes and prefoldins with a test set of 48 poorly expressing human and archaeal proteins.

### 5.3 RESULTS AND DISCUSSION

**Recombinant mevalonate 3,5-bisphosphate decarboxylase co-purifies with GroEL and GroES.** In our previous publication, we were unable to express *T. acidophilum* mevalonate 3,5-bisphosphate decarboxylase (MBD) in the soluble fraction of *E. coli* lysate using standard molecular biology expression conditions.<sup>2</sup> We tried numerous techniques to produce soluble protein including varying induction parameters, *in vitro* expression, refolding kits, periplasmic export, and various solubility tags. We only saw MBD on an SDS-PAGE gel when a truncated version was co-expressed with GroEL/ES (Fig 5.3). The band for truncated MBD was seen at the expected molecular weight of ~47 kDa on an SDS-PAGE gel, however, there were two larger bands present in the affinity purified fraction at around 57 kDa and 10 kDa, which correspond to GroEL and GroES respectively.



This was particularly striking since neither GroEL nor GroES were his-tagged. We hypothesized that misfolded MBD was binding GroEL at the rim of the cage, so when MBD was purified by its N-terminal his-tag, the GroEL protein was co-purified (For example see Fig 5.1C). Since GroEL forms a double ring with 14 identical subunits, we should expect about 14 times the amount of GroEL on the gel. Additionally, we saw a band for GroES (the lid protein) which can be explained by the fact that GroEL forms two back-to-back cages, and while one side is bound to MBD, the other side could be associated with GroES. Our data suggested that GroEL/ES interacts with truncated MBD, but GroEL/ES could not properly fold the target protein.

**MBD is produced in soluble form when co-expressed with thermosomes and prefoldins.** Since MBD is an archaeal protein, we hypothesized that co-expressing it with the archaeal equivalent of GroEL/ES could increase the yield of active protein. To test this hypothesis, we modified two commercially available chaperone plasmids (pGro7 and pG-KJE8) which express bacterial GroEL/ES and/or DnaKJE, to instead express archaeal thermosomes and prefoldins (Fig 5.4).<sup>15</sup> We simply swapped out the chaperone genes present in pGro7 and pG-KJE8 with genes encoding the  $\alpha$  and  $\beta$  subunits of thermosomes and prefoldins from either *Thermoplasma acidophilum* or *Ferroplasma acidarmanus*. A total of 12 plasmid variations were constructed that express thermosomes only, prefoldins only, both together, and archaeal chaperones expressed in conjunction with DnaKJE.<sup>15</sup> The 12 plasmids are listed in table 5.1 and figure 5.5 shows the expression of chaperones from the constructed plasmids to confirm they function properly. All proteins expressed in soluble form except for prefoldin from *Thermoplasma acidophilum*.

Next, his-tagged MBD was co-expressed with each of the 12 plasmids, affinity purified, and the IMAC eluates were subjected to our GC-FID assay to test for decarboxylase activity. As shown in figure 5.6, co-expressing MBD with thermosomes and prefoldins from *Ferroplasma acidarmanus* (plasmid 10) yielded ~30X the decarboxylase activity as the GroEL/ES system, however a band for MBD was not visible by SDS-PAGE (Fig. 5.7). We believed this is because the amount of soluble MBD produced was insufficient to be visualized on a gel.

**Co-expression with thermosomes and prefoldins improve soluble protein yield in 4 out of 5 MBD homologs.** Next we tested if plasmid 10 can improve soluble expression of MBD homologs from 5 different organisms. The organisms were as follows (sequence identity compared to the *Thermoplasma acidophilum* MBD is shown in parenthesis): *Acidiplasma sp. MBA-1* (68%), *Thermoplasma volcanium* (63%), *Ferroplasma sp. Type II* (50%), *Ferroplasma acidarmanus* (48%), and *Picrophilus torridus* (40%). By co-expressing these MBD homologs with plasmid 10, we were able to improve active protein yield in 4 out of 5 homologs (Fig 5.8).

Three MBD homologs (*Ferroplasma acidarmanus*, *Ferroplasma sp. Type II*, and *Acidiplasma sp. MBA-1*) were completely inactive when expressed without chaperone, however strong activity was detected in the respective IMAC eluates when co-expressed with plasmid 10 (Fig 5.8). Most significantly, we were able to clearly see bands appear on an SDS-PAGE gel for the purified MBD homologs from *Acidiplasma sp. MBA-1*, *Thermoplasma volcanium*, and *Ferroplasma sp. Type II* (Fig 5.8, black arrows). The only protein which was not improved by plasmid 10 was the *P. torridus* homolog. The reason for this is unclear.

**Soluble expression of MBD from *Acidiplasma MBA-1* is greatly improved when thermosomes and prefoldins are expressed together.** We determined if the improved expression of recombinant protein in the presence of plasmid 10 was due to the action of the thermosomes, the prefoldins, or both expressed together. We co-expressed MBD from *Acidiplasma MBA-1* with plasmid 10 under different induction conditions which induced thermosomes only (tetracycline), prefoldin only (arabinose) or both together (arabinose and tetracycline). As show in figure 5.10, a band corresponding to MBD was visible when thermosomes or prefoldins were expressed alone, however induction of both chaperones together produced a much larger band. This data suggests that the ability for plasmid 10 to promote production of soluble protein in the case of *Acidiplasma sp. MBA-1* MBD is due largely to the thermosomes and prefoldins working together (Fig 5.9).

#### 5.4 FUTURE DIRECTIONS

Collaborators in the UCLA Protein Expression Technology Center (PETC) will evaluate the potential of this system to be used as a general strategy for producing soluble recombinant proteins. PETC will test our system on a subset of 48 archaeal and eukaryotic proteins, which were previously scored by the Northeast Structural Genomics Consortium (NESG) as either insoluble or very poorly soluble (Table 5.2 and 5.3).<sup>23</sup> If a significant number of proteins from the test set become soluble, our novel method could be widely implemented by researchers and industry allowing the study of previously unobtainable proteins.

#### 5.5 METHODS

**General Materials.** Miller LB media (Gentrox) was used for growth of bacterial strains. *E. coli* BL21(DE3) (Lucigen) was used as the host strain for both cloning and expression of

recombinant proteins. Plasmid pET28a(+) was purchased from Novagen. *Phusion* DNA polymerase (Finnzymes), Taq DNA ligase (MCLab), and T5 Exonuclease (Epicenter) were purchased separately and used to make the assembly master mix (AMM) used for cloning. Talon metal affinity resin was purchased from Clontech. Primers were synthesized by IDT. All other chemicals were purchased from Sigma Aldrich unless otherwise noted.

**Existing plasmids.** The plasmid expressing truncated *T. acidophilum* MBD was obtained from the DNASU plasmid repository at Arizona State University (Catalog No. TaCD00532176). Plasmid pBB541 was obtained from addgene (Catalog No: 27394). Plasmids pGro7 and pG-KJE8 were purchased from Takara Bioscience.

**Construction of plasmids expressing MBDs.** All MBD homologs were codon optimized and synthesized by IDT. The protein sequences corresponding to the following NCBI accession numbers were used for each homolog: *T. acidophilum* (WP\_010901303.1), *T. volcanium* (WP\_010916684.1), *P. torridus* (WP\_010917040.1), *F. acidarmanus* (WP\_009887850.1), *F. sp. Type II* (EQB73519.1), and *A. sp. MBA-1* (WP\_048100791.1). The DNA sequence of each homolog was inserted between the NdeI and XhoI sites of the pET28a(+) vector, which allowed for the addition of an N-terminal 6xHis tag. This was accomplished using Gibson assembly.<sup>24</sup> In brief, the gene to be inserted was synthesized with an extra 25 base-pairs complementary to the pET28a(+) vector at the NdeI and XhoI sites. 30 ng of synthesized DNA was mixed with 10 ng of pET28a(+) digested with NdeI and XhoI and 10  $\mu$ L of 1X Gibson assembly mix. After incubation at 50°C for 2 hrs, 5  $\mu$ L was used to transform *E. coli* BL21(DE3) and transformants were selected on LB-agar plates containing 50  $\mu$ g/mL kanamycin.

**Construction of plasmids 1-12.** Plasmids expressing thermosomes and prefoldins were constructed using Gibson assembly by modifying plasmids pGro7 and pG-KJE8 (23). Only the chaperone genes within the *araB* and *tet* operons of pGro7 and pG-KJE8 were changed. The NCBI accession numbers for the prefoldins and thermosomes used in this study are as follows: *T. acidophilum* prefoldin  $\alpha$  subunit (WP\_010901487), *T. acidophilum* prefoldin  $\beta$  subunit (WP\_010901545), *T. acidophilum* thermosome  $\alpha$  subunit (CAC12109), *T. acidophilum* thermosome  $\beta$  subunit (WP\_010901684), *F. acidarmanus* prefoldin  $\alpha$  subunit (WP\_009887419), *F. acidarmanus* prefoldin  $\beta$  subunit (WP\_009886466), *F. acidarmanus* thermosome  $\alpha$  subunit (WP\_019841453), *F. acidarmanus* thermosome  $\beta$  subunit (WP\_009886573). Codon optimized DNA sequences corresponding to these chaperones were synthesized by IDT. They were inserted into pGro7 and pG-KJE8 in place of the existing chaperones between the ATG start and TAA stop codons using Gibson assembly.<sup>24</sup> All plasmid sequences used in this study are available as fully annotated .dna files in the supplemental material.

**Expression and Purification.** All *E. coli* strains expressing MBD homologs were grown at 37°C in LB-media with 50  $\mu\text{g}/\text{mL}$  kanamycin and 34  $\mu\text{g}/\text{mL}$  chloramphenicol when appropriate. 100 mL of LB-media was inoculated with 5 mL of overnight starter culture. Chaperone expression was induced during late log phase ( $\text{OD}_{600} \sim 0.7$ ) by the addition of 0.2% w/v L-arabinose and/or 10 ng/mL tetracycline. After 1 hour of incubation, expression of MBD homologs was induced by the addition of 0.5 mM IPTG. After 16 hours, cells were pelleted, resuspended in 1 mL of buffer A (50 mM bis-tris propane [pH 7.0], 200 mM NaCl), lysed by sonication, and cell debris was removed by centrifugation at 5,000xg for 15 min. The lysate was mixed with 200  $\mu\text{L}$  of a Talon affinity resin slurry and incubated at 4°C with gentle mixing. After 15 min, the lysate mixture was pelleted and washed 3 times with 1 mL buffer A containing 5 mM imidazole. The talon resin was again pelleted, but this time resuspended in 200  $\mu\text{L}$  buffer A

containing 300 mM imidazole, vortexed, and then pelleted. The supernatant could be used directly in the GC-FID assay below or analyzed by SDS-PAGE.

**GC-FID Decarboxylase Assay.** A 40  $\mu$ L sample of IMAC eluate was added to a 200  $\mu$ L reaction mixture consisting of 10 mM (R)-mevalonate, 20 mM ATP, 10 mM  $MgCl_2$ , 100 mM NaCl, 5  $\mu$ g mevalonate 3-kinase (Ta1305), and 15  $\mu$ g mevalonate-3-phosphate-5-kinase (Ta0762). The reaction was incubated for 24 hours at 55°C for testing MBD homologs from *P. torridus*, *T. volcanium*, and *T. acidophilum*. The reaction was incubated at 37°C for all other homologs. Any generated isopentenyl phosphate was then hydrolyzed into isoprenol and free phosphate by adding 100  $\mu$ L of 1 M bis-tris propane [pH 8.75], followed by 30 U of alkaline phosphatase from bovine intestinal mucosa (NEB). After incubation at 37°C for 2 hours, the reaction was extracted with 200  $\mu$ L n-hexanes. 5  $\mu$ L of the n-hexane layer was injected into a HP5890 Series II Gas Chromatograph (flame ionization detector) connected to a HP-INNOWAX column (0.320 mm x 30 m, Agilent). The carrier gas was helium with a flow rate of 5 mL/min. The oven temperature was kept at 50°C for 2 min, then raised to 100°C at 10°C/min, then ramped to 250°C at 25°C/min, and finally held at 250°C for 2 min. The inlet and detector temperatures were kept at 250 and 330°C respectively. Isoprenol eluted at 6.44 min and the sample concentration was determined by comparison to an authentic standard.

**SDS-PAGE.** 10  $\mu$ L of IMAC eluate was diluted in buffer A containing 300 mM imidazole to yield a protein concentration of 1 mg/mL. 10  $\mu$ L of this diluted Ni-NTA eluate was mixed with 5  $\mu$ L of 4X SDS-PAGE loading dye containing 10 mM  $\beta$ -mercaptoethanol, vortexed, and heated to 96°C for 3 minutes. 10  $\mu$ L of this solution was loaded onto a 4-20% PAGE gel from Genscript. The gel was run in MOPS buffer at room temperature at 120V for 60 min and stained overnight with

Instant Blue Stain (Expedeon). To reduce background when trying to detect low levels of MBD, the protein solution was heated to 60°C for 1 hour then spun down prior to mixing with loading dye in order to precipitate *E. coli* proteins.

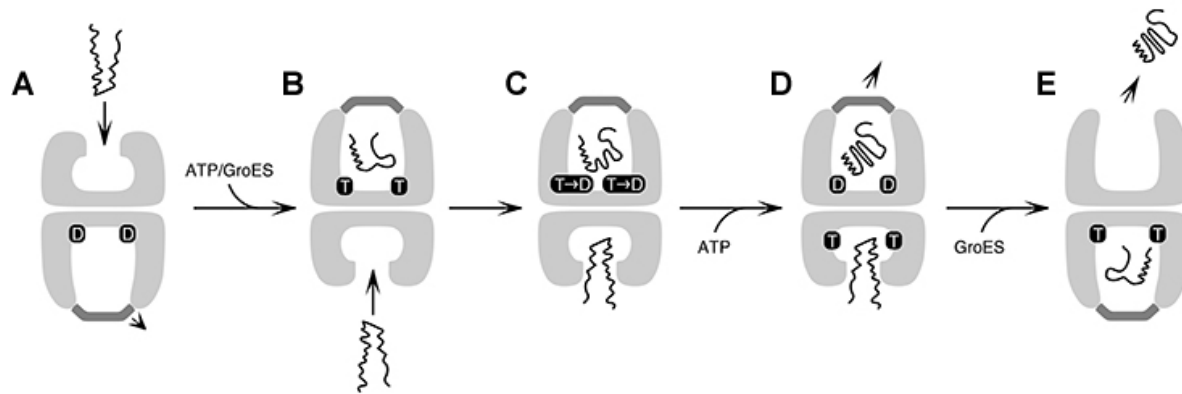
**Induction test of plasmids 1-12.** *E. coli* BL21(DE3) harboring each of the 12 plasmids was inoculated in 5 mL of LB with 34 µg/mL chloramphenicol. When the OD600 reached approximately 0.7, inducer was added as follows: 0.2% w/v L-arabinose and/or 10 ng/mL tetracycline. After induction, cultures were incubated for 20 hours at 37°C in a rotator spinning at 60 rpm. 1 mL of the culture was spun down at 10,000g. The supernatant was removed and the pellet was resuspended in 200 µL buffer A (50 mM bis-tris propane [pH 7.0], 200 mM NaCl). The resulting solution was sonicated on ice using a sonic dismembrator 550 (Fisher Scientific) equipped with a microtip. The sample was spun down at 10,000g and the supernatant was diluted 10 fold before analyzed by SDS-PAGE. *E. coli* containing no chaperone plasmid served as the control.

**Differential expression using plasmid 10.** A glycerol stock of *E. coli* BL21(DE3) harboring plasmid 10 and the recombinant protein of interest was inoculated in 20 mL of LB with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin. The following day 5 mL of this overnight culture was inoculated into each of four 1 L baffled flasks containing LB with the same antibiotics. When the OD600 reached approximately 0.7 we induced the chaperones: the first flask received 0.2% w/v L-arabinose, the second received 10 ng/mL of tetracycline, the third received 0.2% w/v L-arabinose and 10 ng/mL of tetracycline, and the fourth flask received no inducer. After 1 hour at 37°C, IPTG was added to a final concentration of 0.5 mM. Cultures were incubated for 20 hours at 37°C in a rotator spinning at 190 rpm. Cultures were spun down and the cell pellet was

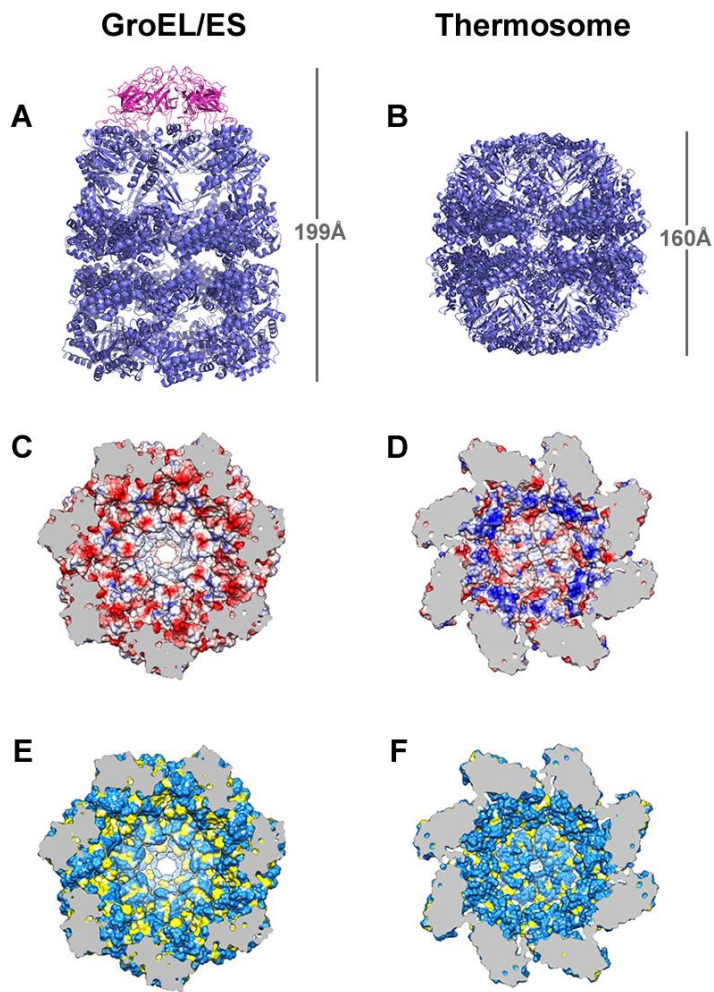
resuspended in 5 mLs of buffer A (50 mM bis-tris propane [pH 7.0], 200 mM NaCl), and processed identically as described in the expression and purification section above.

**Selection of 48 test proteins.** We selected a test set of 24 eukaryotic and 24 archaeal proteins from an initial list of 7,354 protein constructs with known expression characteristics. NESG has previously scored these 7,354 protein constructs for expression and solubility on a qualitative scale of 0-5.<sup>23</sup> The expression score was based on the size of the band observed via SDS-PAGE analysis of the total cell lysate and solubility score was based on the size of the band observed via SDS-PAGE analysis of the soluble fraction. We filtered the list to include only eukaryotic and archaeal proteins that express well (expression score: 4 or 5) but are poorly soluble (solubility score: 0 or 1). These constraints brought the total number of proteins to 236 (majority of the list was bacterial proteins). From this subset, we eliminated all proteins identified as hypothetical, proteins suspected to bind an unusual co-factor (Ex. Fe-S clusters), insoluble partial sequences, and subunits of hetero-oligomeric complexes. This was done via manual inspection of each NCBI entry. This step brought the total to 71 eligible proteins. From this subset, we eliminated suspected membrane proteins and proteins containing export signal peptides by submitting each protein sequence to Protter and manually inspected each graphical output.<sup>25</sup> 5 proteins were eliminated using Protter. Next we submitted each protein sequence to PrDOS, the Protein DisOrder prediction System, which predicts natively disordered regions in proteins.<sup>26</sup> This step eliminated 10 proteins, which brought the total eligible candidates to 56. Finally, eight were eliminated at random to yield the 48 protein test set. The plasmids corresponding to these 48 proteins were obtained from DNASU.

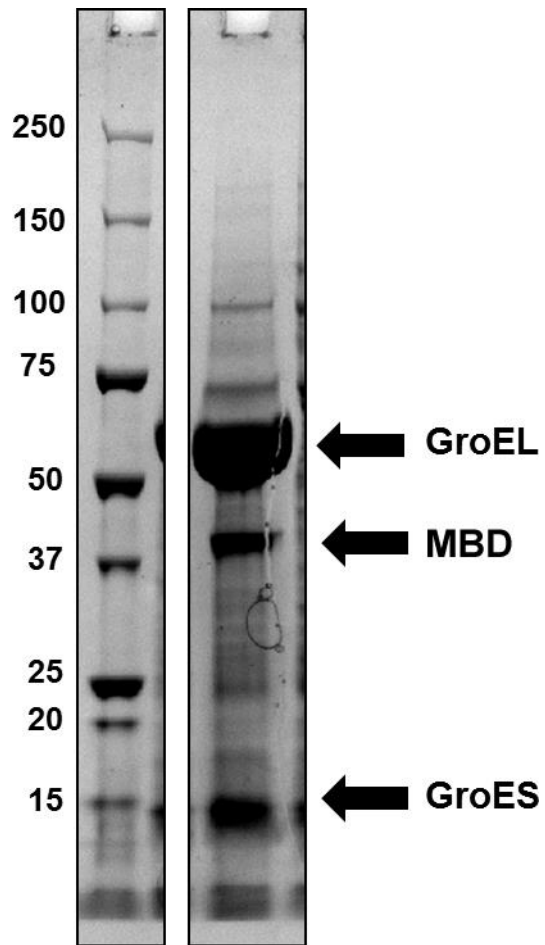




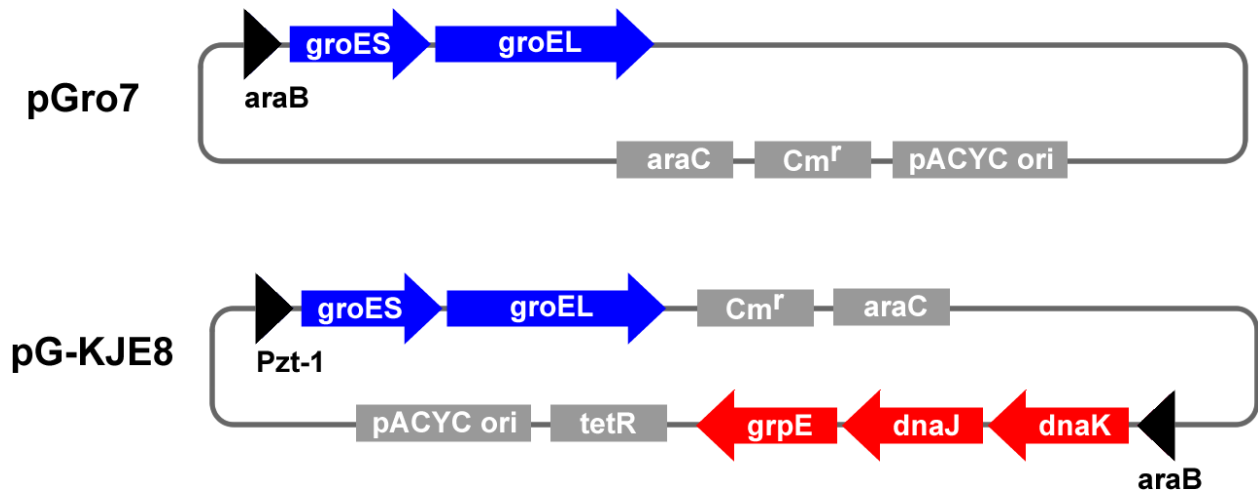
**FIGURE 5.1** Review of the GroEL/ES mechanism. (A.) GroEL proteins form two back-to-back cages and the rim of the cage binds unfolded protein. (B.) Upon binding of ATP and the GroES cap, the unfolded protein enters the cage. (C.) ATP is hydrolyzed to ADP, which reduces the affinity of the GroES cap to the GroEL cage. (D.) The GroES cap dissociates when the opposite side binds ATP. (E.) The folded protein is released and the cycle repeats. A version of this figure originally appeared in *Current Opinion in Structural Biology* and it has been reprinted with permission from Elsevier (License No. 4112191036680).<sup>27</sup>



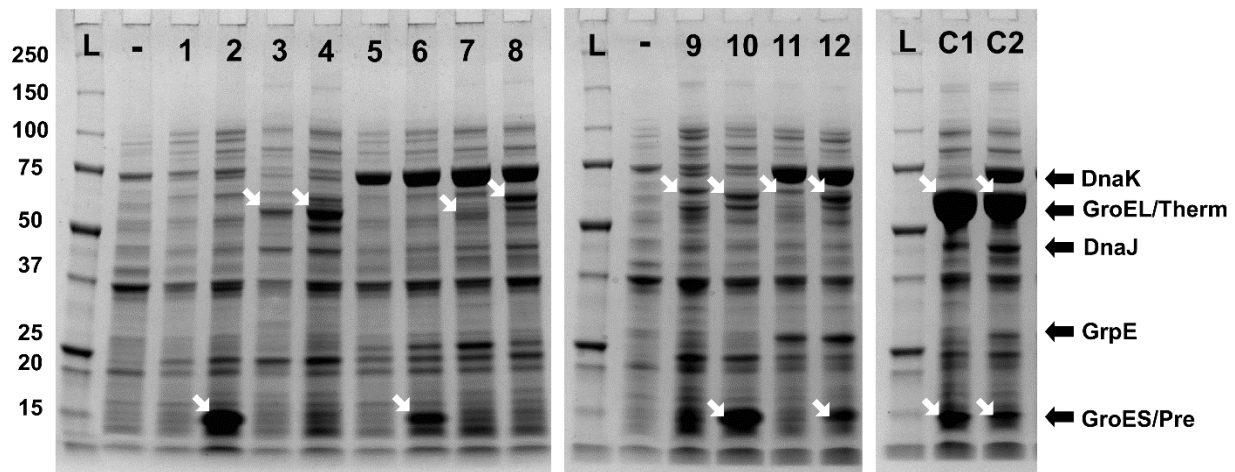
**FIGURE 5.2** Structural comparison of GroEL/ES and the thermosome. (A.) Structure of GroEL/ES with the GroES cap shown in purple. (B.) Structure of the thermosome from *Thermoplasma acidophilum*. (C, D.) Charged residues lining the inside of the two chaperonins are highlighted where red is negatively charged and blue is positively charged. (E., F.) Hydrophobic residues lining the cages of the two chaperonins are shown in yellow. Panels C-F are reprinted here as per PNAS's permissions policy. © 2009 National Academies of Science.<sup>19</sup>



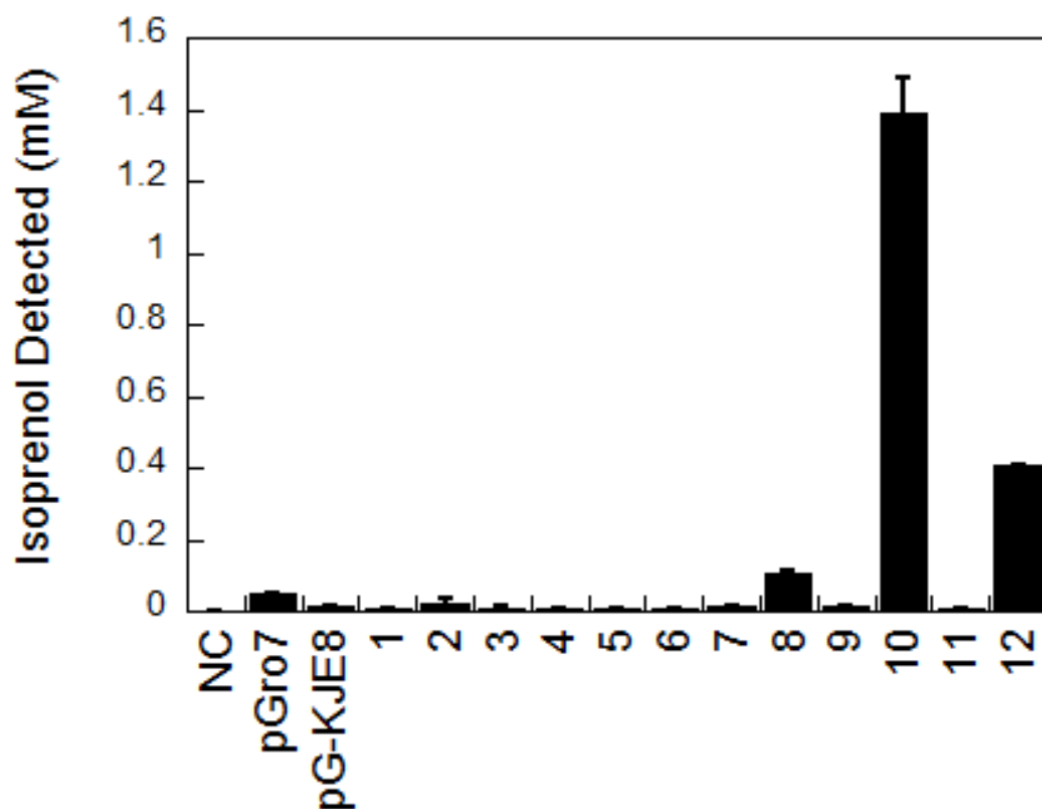
**FIGURE 5.3.** SDS-PAGE gel of truncated *T. acidophilum* MBD (Ta0893) co-expressed with GroEL/ES. His-tagged MBD with a 30 AA truncation on the N-terminus was co-expressed with GroEL/ES (plasmid pBB541). The recombinant protein was purified using standard affinity chromatography procedure and then heated to 60°C for 1 hour to remove *E. coli* proteins (MBD is thermostable). This IMAC eluate was run on an SDS-PAGE gel. The left side shows a standard protein ladder in kilodaltons (kDa).



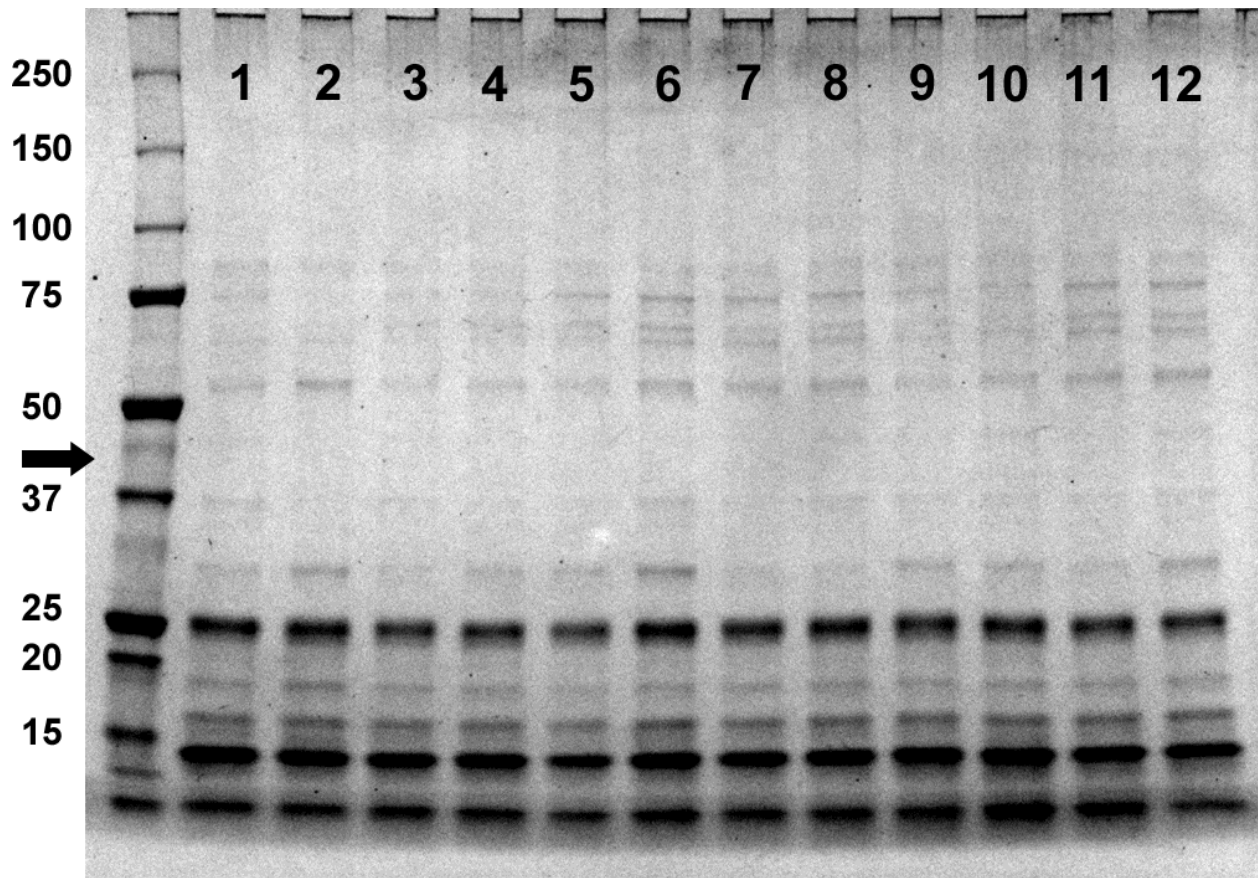
**FIGURE 5.4.** Map of Takara parent plasmids used in this work. The plasmids were modified by replacing the chaperone genes shown blue and red, with thermosomes and prefoldins. Table 5.1 lists the various derivatives constructed from these parent plasmids. Genes are not shown to scale. Construction of plasmids is detailed in the methods.



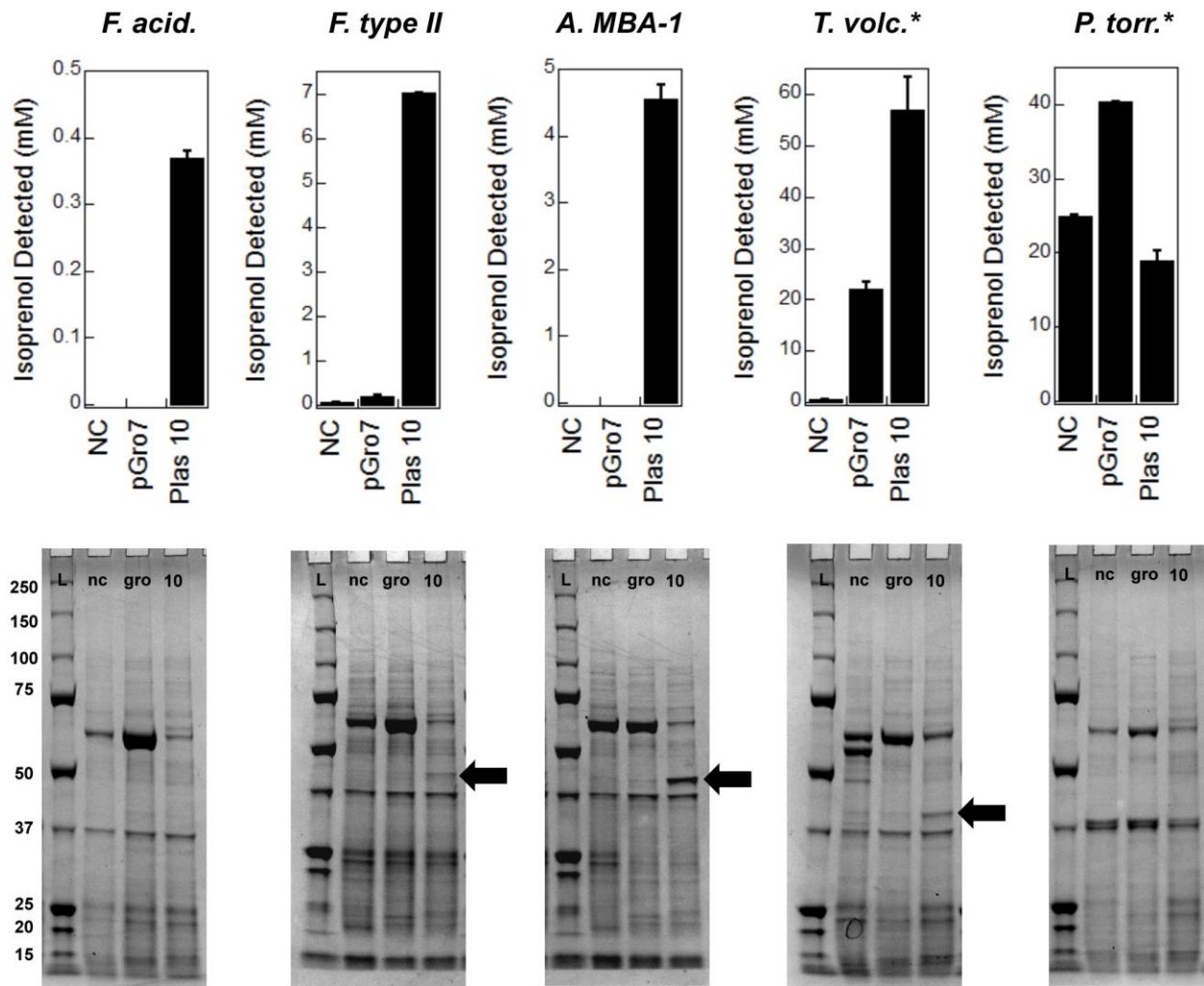
**Figure 5.5.** SDS-PAGE analysis of the chaperone expression from the 12 plasmids constructed in this work. Plasmids 1-4 (pGro7 derivatives) were induced with 0.2% w/v L-arabinose and plasmids 5-12 (pG-KJE8 derivatives) were induced with 0.2% w/v L-arabinose and 10 ng/mL tetracycline. The soluble fraction was run on the SDS-PAGE gel shown above. (-) is *E. coli* without any plasmid. (C1) is pGro7 parent plasmid and (C2) is pG-KJE8 parent plasmid. Bands were visible for all inducible proteins except prefoldin from *T. acidophilum*. The left side shows a standard protein ladder in kilodaltons (kDa).



**FIGURE 5.6.** Activity of MBD co-expressed with each of the 12 thermosome plasmids. Full length *T. acidophilum* MBD was co-expressed with the 12 chaperone plasmids constructed in this study and the two parent plasmids: pGro7 (GroEL, GroES) and pG-KJE8 (GroEL, GroES, DnaK, DnaJ, and GrpE). NC is *E. coli* with no chaperone plasmid.

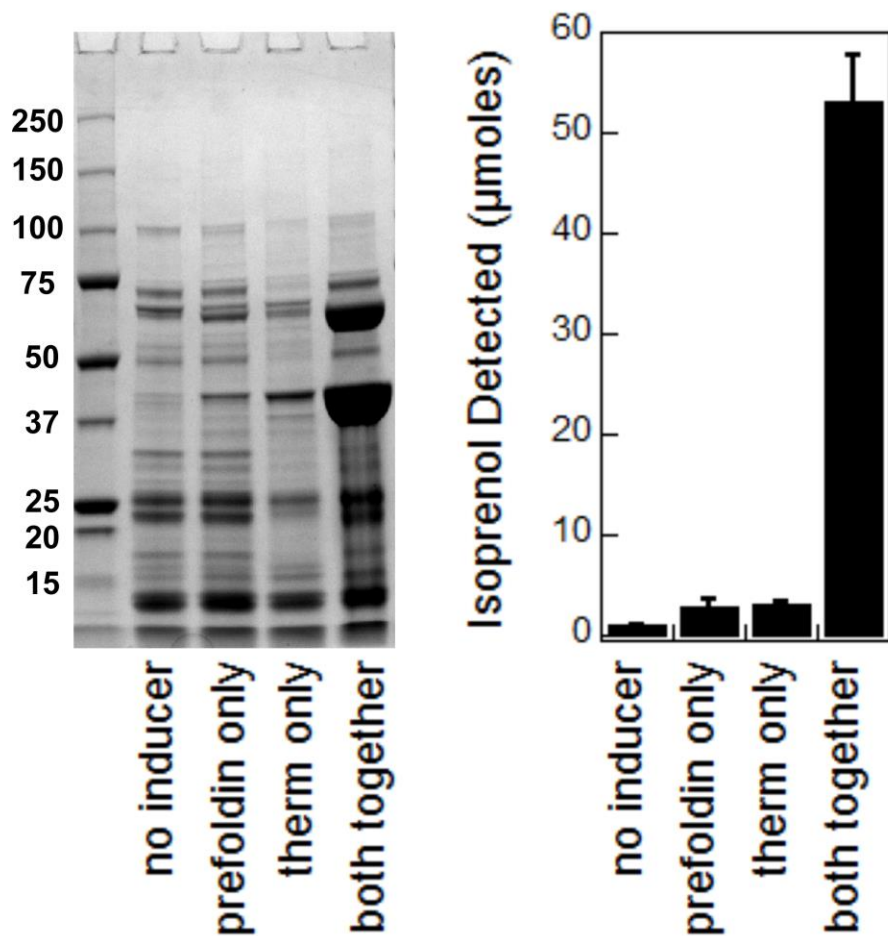


**FIGURE 5.7.** SDS-PAGE of *T. acidophilum* MBD co-expressed with each of the 12 thermosome plasmids. The IMAC eluates from each of the co-expressions was heated at 60°C for 1 hour to reduce background (MBD is thermostable), and then run on a SDS-PAGE gel. No bands were observed near the expected MBD mass of 48 kDa.



**FIGURE 5.8.** MBD homologs co-expressed with plasmid 10. The top row shows the activity of MBD homologs from five organisms co-expressed with either no chaperone (NC), GroEL/ES (gro), or plasmid 10 (Plas 10). Co-expression with plasmid 10 improved yield of active protein in four out of five MBD homologs. The bottom row shows the corresponding SDS-PAGE gel of the IMAC eluates used in the activity assay in the top panels. A band corresponding to MBD (~43 kDa) was visible when *F. type II*, *Acidiplasma MBA-1* and *T. volcanium* MBD were co-expressed with plasmid 10 (black arrows).





**FIGURE 5.9.** Differential expression of *Acidiplasma* MBD with plasmid 10. An *E. coli* strain containing plasmid 10 and a plasmid expressing *Acidiplasma* MBD was grown in various conditions: (1) no inducer, (2) with arabinose to induce prefoldin expression (3) with tetracycline to induce thermosome expression, and (4) with both inducers together to express thermosomes and prefoldins. A large band for *Acidiplasma* MBD was visualized on an SDS-PAGE gel when thermosomes and prefoldins were expressed together. The corresponding activity assay shows that expression of both chaperones together produces the largest increase in active MBD.

Plasmid	Parent	Arabinose Inducible	Tetracycline Inducible
<b>Each Element Independently</b>			
1	pGro7	TA prefoldin	-
2	pGro7	FR prefoldin	-
3	pGro7	TA thermosome	-
4	pGro7	FR thermosome	-
<b>Each Element With DnaK, DnaJ, GrpE</b>			
5	pG-KJE8	TA prefoldin	DnaKJE
6	pG-KJE8	FR prefoldin	DnaKJE
7	pG-KJE8	TA thermosome	DnaKJE
8	pG-KJE8	FR thermosome	DnaKJE
<b>Thermosomes and Prefoldins Together</b>			
9	pG-KJE8	TA prefoldin	TA thermosome
10	pG-KJE8	FR prefoldin	FR thermosome
<b>Thermosomes and Prefoldins with DnaK, DnaJ, and GrpE)</b>			
11	pG-KJE8	TA prefoldin + TA thermosome	DnaKJE
12	pG-KJE8	FR prefoldin + FR thermosome	DnaKJE

**TABLE 5.1.** List of plasmids constructed in this work. TA is short for *Thermoplasma acidophilum* and FR is short for *Ferroplasma acidarmanus*. All prefoldins and thermosomes used in this study have two subunits ( $\alpha$  and  $\beta$ ) and we expressed in tandem. Expression of the chaperones is induced with 0.2% w/v L-arabinose and/or 10 ng/mL tetracycline. Full annotated sequences for all 12 plasmids is included in the supplement as .dna files.

ID	Protein Annotation	Organism	Accession	Size	E/S
1	actin-like protein 6A	<i>Homo sapiens</i>	NP_004292	49	5/0
2	alpha-crystallin B chain	<i>Homo sapiens</i>	NP_001876	19	4/0
3	eukaryotic translation initiation factor 6	<i>Homo sapiens</i>	NP_002203	28	4/0
4	tumor susceptibility gene 101	<i>Homo sapiens</i>	EAW68381	22	4/0
5	heat shock protein beta-2	<i>Homo sapiens</i>	NP_001532	20	5/1
6	MAPK/MAK/MRK overlapping kinase	<i>Homo sapiens</i>	NP_055041	36	5/0
7	ectoderm-neural cortex protein 1	<i>Homo sapiens</i>	NP_001243505	15	5/1
8	dual specificity mitogen-activated protein kinase	<i>Homo sapiens</i>	NP_659731	37	4/0
9	microspherule protein 1	<i>Homo sapiens</i>	NP_006328	40	5/0
10	cyclin-dependent kinase 4	<i>Homo sapiens</i>	AAP36759	35	5/1
11	calpain-5	<i>Homo sapiens</i>	NP_004046	70	4/0
12	Cullin-2	<i>Homo sapiens</i>	ELK38740	13	5/0
13	nitric oxide synthase 2	<i>Homo sapiens</i>	XP_009430818	23	5/0
14	HECT E3 ubiquitin ligase	<i>Homo sapiens</i>	AAR00320	48	5/0
15	dedicator of cytokinesis protein 9	<i>Homo sapiens</i>	NP_001123521	18	4/0
16	cyclin B3	<i>Homo sapiens</i>	EAW89920	18	5/0
17	nephronophthisis 1	<i>Homo sapiens</i>	EAW50635	20	5/0
18	toll-interacting protein	<i>Homo sapiens</i>	NP_061882	32	5/1
19	sedoheptulokinase	<i>Homo sapiens</i>	AAQ02580	53	5/0
20	procaspase-8	<i>Homo sapiens</i>	AAL87632	20	5/1
21	down syndrome candidate reg. 1	<i>Homo sapiens</i>	AAB84371	20	5/1
22	copine-4	<i>Homo sapiens</i>	NP_001276041	63	5/0
23	DET1 homolog isoform 1	<i>Homo sapiens</i>	NP_060466	65	5/0
24	2-oxoglutarate and Fe-dependent oxygenase domain-containing protein 2	<i>Homo sapiens</i>	NP_078899	31	5/1

**TABLE 5.2.** Test set of 24 eukaryotic proteins. The first number of the E/S score is the previously reported expression from 0-5 and the number after the slash is the previously reported solubility from 0-5. Size is given in kilodaltons (kDa).

ID	Protein Annotation	Organism	Accession	Size	E/S
1	menaquinone biosynthesis decarboxylase	<i>Archaeoglobus fulgidus</i>	WP_010877720	56	5/0
2	4-hydroxybutyrate CoA-transferase	<i>Archaeoglobus fulgidus</i>	WP_010878641	55	4/1
3	hydantoin utilization protein A	<i>Archaeoglobus fulgidus</i>	WP_010878406	55	5/0
4	DNA-binding protein	<i>Thermoplasma volcanium</i>	WP_010917235	16	4/0
5	ATP-binding protein	<i>Archaeoglobus fulgidus</i>	WP_010877985	30	5/0
6	FAD/NAD(P)-binding oxidoreductase	<i>Haloarcula marismortui</i>	WP_004962086	63	5/0
7	oleate hydratase	<i>Haloarcula marismortui</i>	WP_011223911	61	5/0
8	ATPase	<i>Methanosarcina mazei</i>	WP_048045837	55	4/0
9	transposase	<i>Haloarcula marismortui</i>	AAV46477	30	5/0
10	thioredoxin	<i>Thermoplasma volcanium</i>	WP_010916989	53	4/1
11	Geranylgeranyl reductase	<i>Methanosarcina mazei</i>	AAM31545	46	4/0
12	ArsR family transcriptional regulator	<i>Methanothermobacter thermautotrophicus</i>	WP_010875835	55	4/1
13	glycosyl transferase	<i>Methanosarcina mazei</i>	WP_011033090	37	5/0
14	Biotin synthase	<i>Methanosarcina mazei</i>	AAM32190	38	5/0
15	3',5'-cyclic-nucleotide phosphodiesterase	<i>Methanosarcina mazei</i>	WP_080503007	27	5/0
16	N-acyl homoserine lactonase family protein	<i>Archaeoglobus fulgidus</i>	WP_010879006	30	5/0
17	cobyric acid synthase	<i>Methanococcus maripaludis</i>	WP_011171159	28	4/0
18	amidohydrolase	<i>Methanosarcina acetivorans</i>	WP_011022897	59	5/0
19	flavodoxin	<i>Methanosarcina acetivorans</i>	WP_011022152	23	5/0
20	Xaa-Pro dipeptidase	<i>Methanosarcina acetivorans</i>	WP_011020271	39	5/0
21	DEXX-box ATPase	<i>Pyrococcus furiosus</i>	WP_011011419	54	4/0
22	HTR-like protein	<i>Haloarcula marismortui</i>	WP_007188402	32	4/0
23	mannosyltransferase	<i>Methanothermobacter thermautotrophicus</i>	WP_010875970	26	5/0
24	anaerobic ribonucleoside-triphosphate reductase activating protein	<i>Methanothermobacter thermautotrophicus</i>	AAB84793	28	5/0

**TABLE 5.3.** Test set of 24 archaeal proteins. The first number of the E/S score is the previously reported expression from 0-5 and the number after the slash is the previously reported solubility from 0-5. Size is given in kilodaltons (kDa).

## 5.6 REFERENCES

1. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411–421 (1999).
2. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Factories* **4**, 1 (2005).
3. Baneyx, F. & Mujacic, M. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **22**, 1399–1408 (2004).
4. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, (2014).
5. di Guana, C., Lib, P., Riggsa, P. D. & Inouyeb, H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67**, 21–30 (1988).
6. Choi, J. H. & Lee, S. Y. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **64**, 625–635 (2004).
7. Fenton, W. A. & Horwich, A. L. GroEL-Mediated protein folding. *Protein Sci.* **6**, 743–760 (1997).
8. Gupta, P., Aggarwal, N., Batra, P., Mishra, S. & Chaudhuri, T. K. Co-expression of chaperonin GroEL/GroES enhances in vivo folding of yeast mitochondrial aconitase and alters the growth characteristics of *Escherichia coli*. *Int. J. Biochem. Cell Biol.* **38**, 1975–1985 (2006).
9. Caspers, P., Stieger, M. & Burn, P. Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*. *Cell. Mol. Biol. Noisy--Gd. Fr.* **40**, 635–644 (1994).

10. Simpanya, M. F., Leverenz, V. R. & Giblin, F. J. Expression and Purification of his-tagged Recombinant Mouse  $\zeta$ -Crystallin. *Protein Expr. Purif.* **69**, 147–152 (2010).
11. Ellis, R. J. & Vies, S. M. van der. Molecular Chaperones. *Annu. Rev. Biochem.* **60**, 321–347 (1991).
12. Keskin, O., Bahar, I., Flatow, D., Covell, D. G. & Jernigan, R. L. Molecular Mechanisms of Chaperonin GroEL–GroES Function. *Biochemistry (Mosc.)* **41**, 491–501 (2002).
13. Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–(ADP)<sub>7</sub> chaperonin complex. *Nature* **388**, 741–750 (1997).
14. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* **7**, 32 (2007).
15. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**, 1694–1699 (1998).
16. Gutsche, I., Essen, L.-O. & Baumeister, W. Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293**, 295–312 (1999).
17. Leroux, M. R. & Hartl, F. U. Protein folding: Versatility of the cytosolic chaperonin TRiC/CCT. *Curr. Biol.* **10**, R260–R264 (2000).
18. Waldmann, T. *et al.* The Thermosome of *Thermoplasma acidophilum* and Its Relationship to the Eukaryotic Chaperonin TRiC. *Eur. J. Biochem.* **227**, 848–856 (1995).
19. Cong, Y. *et al.* 4.0-Å resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc. Natl. Acad. Sci.* **107**, 4967–4972 (2010).

20. Vainberg, I. E. *et al.* Prefoldin, a Chaperone that Delivers Unfolded Proteins to Cytosolic Chaperonin. *Cell* **93**, 863–873 (1998).
21. Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U. & Moarefi, I. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* **103**, 621–632 (2000).
22. Vinokur, J. M., Cummins, M. C., Korman, T. P. & Bowie, J. U. An Adaptation To Life In Acid Through A Novel Mevalonate Pathway. *Sci. Rep.* **6**, 39737 (2016).
23. Boël, G. *et al.* Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* **529**, 358–363 (2016).
24. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
25. Omasits, U., Ahrens, C. H., Müller, S. & Wollscheid, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**, 884–886 (2014).
26. Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460-464 (2007).
27. Grantcharova, V., Alm, E. J., Baker, D. & Horwich, A. L. Mechanisms of protein folding. *Curr. Opin. Struct. Biol.* **11**, 70–82 (2001).