Kinetic characterization, crystallization, and photosynthetic expression of (+)-4R-limonene synthase from *C. sinensis*

A Master’s Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Biochemistry
Daniel Oprian, Advisor

In Partial Fulfillment of the Requirements for

Master’s Degree

by

Sonya Entova

May, 2013
ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to the many people without whose help this thesis would not have been possible. Above all, I want to thank Dr. Daniel Oprian for the privilege of working in his lab. I thank him for his patience and insight as a mentor, as well as for challenging me to think critically and creatively, plan accordingly, and persevere, no matter how daunting the task at hand. It is only by virtue of his unwavering guidance and encouragement that I now feel prepared to continue my journey as a scientist.

In addition, I would like to thank all the members of the Oprian lab, both past and present, and my fellow classmates in the biochemistry department. Learning from them has been a true honor, and I would not be the scientist, or person, that I am today without their advice, support, and companionship.

I would like to thank my parents from the very bottom of my heart for always demanding only the best from me, and for teaching me to do the same. I could never have made it here were it not for their constant love and encouragement. And finally, I would like to thank my close friends and family, whose endless support has made this thesis a reality. To list all their names here would be impossible, but I trust that they recognize who they are, and how much their support has meant to me.
ABSTRACT

Kinetic Characterization, Crystallization, and Photosynthetic Expression of (+)-4R-Limonene Synthase from C. sinensis

A Thesis Presented to the Department of Biochemistry
Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Sonya Entova

The terpenoid family represents one of the most abundant and diverse classes of organic molecules found in nature. Because limonene synthase catalyses the simplest of all monoterpenic (C\textsubscript{10}) cyclization reactions, it is often used as a model for monoterpenic biosynthesis. The study described here was done with the enzyme (+)-(4R)-limonene synthase, which was cloned previously from the flavedo of Citrus sinensis. In this thesis, we discuss the kinetic characterization of this enzyme, in which we observe kinetic properties very different from those previously found for the same enzyme cloned from Citrus limon. We also describe a crystal structure of this enzyme in apoprotein form, and the soaking of such crystals in a substrate-containing solution, where we observe time-dependent changes in electron density for three coordinating Mn\textsuperscript{2+} atoms in the enzyme active site. We hope that these results constitute key steps in identifying the structural components of the enzyme mechanism. Finally, we relate the initial stages of an effort to produce (+)-limonene, a potential biofuel, in the cyanobacterium Synechocystis by engineering our (+)-(4R)-limonene synthase gene into the Synechocystis genome.
# Table of Contents

**Abstract**
iv

**Table of contents**
v

**List of Tables**
vi

**List of Figures**
vii

**Introduction**
1
   *Terpenoids*  
   *Monoterpenes*  
   *Limonene Synthase*  
   *Synechocystis*

**Chapter 1: Limonene Synthase Expression and Purification Optimization**
11
   *Materials and Methods*  
   *Results*

**Chapter 2: Kinetic Characterization**
16
   *Materials and Methods*  
   *Results*

**Chapter 3: Structure Determination by Crystallization**
22
   *Materials and Methods*  
   *Results*

**Chapter 4: Limonene Production in Synechocystis**
29
   *Materials and Methods*  
   *Results*

**Discussion**
41
   *Kinetic Characterization*  
   *Crystal Structure Determination*  
   *Limonene Production in Synechocystis*  
   *Future Directions*

**Appendix: Supplementary Figures**
51

**References**
53
LIST OF TABLES

1 Enzyme activity of protein expressed at 20°C and 37°C 18
2 Data collection and refinement statistics 24
3 Primers used in plasmid construction 31
LIST OF FIGURES

1 IPP and DMAPP production (simplified) 2
2 Monoterpene, sesquiterpene, and diterpene synthesis 4
3 Limonene enantiomers 6
4 Limonene synthase elution profile 13
5 Optimization of expression and purification 14
6 Product analysis by GC-MS 17
7 Limonene synthase kinetic analysis 20
8 Apoprotein crystal structure 25
9 Manganese atoms bound to the active site 26
10 Positioning of the N-terminal tail and the active site loops 28
11 Generating the KSusds plasmid 32
12 Insertion of GPPS’ and LS into KSusds 34
13 Gibson (isothermal) assembly 35
14 Restriction digest of KSusds 36
15 The pUC19a vector 38
16 The pBluescript vectors 39
17 Relative plasmid and fragment lengths for isothermal assembly 40
A1 BG11 mineral media composition 51
A2 Upstream and downstream psbA2 flanking sequences 52
A3 GPPS silent mutation 52
INTRODUCTION

Terpenoids

The terpenoid (or isoprenoid) family represents one of the most abundant and diverse families of organic compounds found in nature, comprising over 55,000 members, many of which play important biological roles in plant defense, attraction of pollinators, and plant signaling. Many terpenoids have been implicated in a variety of industrial and agricultural roles as flavorings, perfumes, solvents, pesticides, and the like. More recently, a variety of pharmaceutical applications have also been discovered for terpenoids, including as anti-inflammatory, [1] anti-tumor and anti-metastatic [2] agents. The extensive commercial importance of terpenoids, together with the relative difficulties often associated with extracting large amounts of them from their native sources, has spurred an effort to apply methods of genetic and metabolic engineering to producing these compounds recombinantly in a variety of alternate hosts [3, 4].

Despite the broad diversity of the terpenoid family, all terpenes are composed of multiples of one common C_5 precursor, isoprene, linked together in a “head-to-tail” manner. The first step in the synthesis of any terpene is the production of two isomeric activated forms of isoprene: isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Because these two isomeric subunits are products of both the mevalonic acid pathway and the mevalonic acid-independent pathway [5], they are amply available in virtually all living species (Figure 1).
Fig. 1 IPP and DMAPP production (simplified) – Isoprene, the subunit from which all terpenoids are synthesized, is found in two isomeric activated forms: isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (A). IPP and DMAPP are products of both the mevalonic acid pathway (B) and the mevalonic acid-independent pathway (C).

One IPP unit and one DMAPP unit can be joined together to form the C\textsubscript{10} terpene precursor, geranyl pyrophosphate (GPP). The 1-4 addition of an additional IPP unit to GPP gives the C\textsubscript{15} terpene precursor, farnesyl pyrophosphate (FPP), and the subsequent addition of a third IPP addition to FPP gives the C\textsubscript{20} terpene precursor, geranylgeranyl pyrophosphate (GGPP) [6] (Figure 2). The cyclization of GPP and subsequent redox modification yields the wide variety of C\textsubscript{10} terpenes known as monoterpenes. Similarly, the cyclization and modification of FPP yields a variety of C\textsubscript{15} terpenes, which are known as sesquiterpenes, and the cyclization and modification of GGPP yields a variety of C\textsubscript{20} terpenes, called diterpenes. The focus of this study will be on the enzyme (+)-4R-limonene synthase, which produces the monoterpane, limonene.
**Monoterpenes**

The conversion of GPP into monoterpenes occurs through a series of carbocation intermediates, beginning with the removal of the pyrophosphate group to form the linalyl cation. Non-cyclic monoterpenes such as linalool can be formed from this cation, but the formation of cyclic monoterpenes requires a subsequent isomerization of the linalyl cation via a (+)-3S-linalyl pyrophosphate intermediate, generating a conformation that sterically favors cyclization to the \(\alpha\)-terpinyl cation [5]. The \(\alpha\)-terpinyl intermediate can then be transformed into various cyclic monoterpenes, the simplest of which is limonene. These basic monoterpene skeletons can then be modified further via a series of oxidation reactions to yield the great diversity of cyclic monoterpenes observed in nature (Figure 2).

A number of monoterpane synthases from various plant sources have been studied and characterized in either their native or recombinant form [6-9]. As a rule, these enzymes have a molecular mass of 50- to 100-kDa in their native form, and are found as either monomers or homodimers. They are generally soluble, but are often targeted to the plant’s plastid membrane when expressed \textit{in vivo}. Several studies have confirmed the presence a highly conserved RR motif in monoterpane synthases located approximately 50-70 residues from the N-terminus. This arginine pair is believed to stabilize the negatively-charged diphosphate that is hydrolyzed from the GPP substrate [10]. It is
Monoterpene, sesquiterpene, and diterpene synthesis – The three terpenoid precursors geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), and geranylgeranyl pyrophosphate (GGPP) are generated by the addition of IPP and DMAPP subunits. GPP then undergoes a cyclization reaction to form the monoterpene limonene, which can be modified to form other monoterpenes. FPP undergoes a similar process to generate various sesquiterpenes, and GGPP undergoes a similar process to generate various diterpenes. Adapted from Dewick [5].

believed that the sequence upstream of this motif corresponds to an N-terminal transit domain which targets the monoterpene to its final destination within the plastid, and it has been shown that enzymes truncated at a position just prior to this conserved motif take on their mature, native conformation and are fully functional when expressed
recombinantly [6, 7, 10]. In addition, all monoterpenes require a divalent metal cation (generally Mg$^{2+}$ or Mn$^{2+}$) for activity. A conserved DDxxD motif in virtually all monoterpenes active sites is understood to coordinate the metal ions necessary for catalysis [6-8].

One particularly curious property of monoterpenes synthases is their ability to be highly specific in some instances, exclusively producing one enantiomer of a given product, and highly promiscuous in others, capable producing several different products from one common GPP substrate [8, 11]. A phylogenetic analysis of the amino acid sequences of several members of the monoterpenes synthase family has found that, aside from the highly conserved DDxxD and arginine motifs, sequence similarity among these enzymes is driven by speciation rather than by function – that is, enzymes of different function from closely related species generally show more sequence relatedness than do enzymes of similar function from distantly related species [6]. One structural study has suggested that this is a result of a terpene synthase evolution mechanism by which plants use one basic, versatile scaffold for several enzymes, which can be modified slightly to give rise to the wide variety of product-specific terpene synthases necessary for the plant’s survival [9].

**Limonene Synthase**

Because limonene is the simplest of the monoterpenes undergoing a cyclization reaction from the linear substrate GPP, the enzyme limonene synthase is often taken as a model for the study of the subset of monoterpenes synthases that catalyze cyclization. Limonene
has a single chiral center located at C₄ and is thus found in two different enantiomeric forms: (+)-4R-limonene, commonly found in citrus fruit [12], and (-)-4S-limonene, found in a wide variety of plants, including mint, pine, sage, and caraway [5] (Figure 3).

While some plants, such as orange and mint, are known to contain a single limonene enantiomer almost exclusively, others have been found to contain both enantiomers in various ratios [7]. The production of either enantiomer is thought to be dictated by the orientation in which the GPP substrate initially enters the active site, indicating that more promiscuous enzymes must be able to bind both right- and the left-handed conformations of GPP, while more specific enzymes must select for a single conformation [13, 14].

Much work has already been done to characterize limonene synthase from a wide variety of sources. Studies have shown that both (+)- and (-)-limonene synthase share the N-terminal transit peptide and the RR and DDxxD motifs conserved in other monoterpenes synthases [7, 10, 11]. The kinetics of limonene synthase from a variety of native and recombinant sources including (-)-limonene synthase from Mentha x piperita (peppermint) and Mentha spicata (spearmint) [15] and (+)-limonene synthase from C. limon (lemon) have also been reported [11]. It has been found that both (+)- and (-)-limonene synthase prefer the Mn²⁺ cation over the Mg²⁺ cation in vitro [14, 16].
A crystal structure of (-)-4S-limonene from *Mentha spicata* liganded to a substrate analog and Mn$^{2+}$ has been previously determined. This cocrystallization technique yielded crystals of the $I_4$ space group that diffracted to 2.7-Å resolution [14]. The structure of (-)-limonene synthase was found to be similar to those previously determined for (+)-boranyl pyrophosphate synthase from *Salvia* [17] and 5-epi-aristolochene synthase from tobacco [18], in that it is composed of two helical domains – the C-terminal domain, which contains the active site, and an N-terminal domain, which has been observed in several other terpenoid synthases but whose function remains undetermined [19]. Like previous structures, (-)-limonene synthase was found to have an N-terminal tail that folds over the C-terminal active site, forming a cap to shield the carbocation intermediates in the active site from solvent during catalysis. Significant rms deviations for all Cα in the (-)-limonene synthase structure indicated that the enzyme is generally flexible, particularly around the active site [14]. This agrees well with a reaction mechanism in which the enzyme must switch between open and closed conformations upon substrate binding and product release. Three Mn$^{2+}$ metal ions, coordinated by the DDxxD motif, are easily observable in this structure.

This thesis will focus on a recombinant (+)-4R-limonene synthase, which has previously been cloned from *Citrus sinensis*, the navel orange [16]. The kinetics of (+)-4R-limonene synthase from *Citrus limon* (lemon) have been previously characterized, showing a $K_m$ of 0.7µM and an activity profile consistent with substrate inhibition [11]. Initial activity assays of our (+)-limonene synthase from *C. sinensis* indicated that the $K_m$
of this enzyme is significantly higher than that of its counterpart from *C. limon* [16]. In this study we present a more complete analysis of the kinetics of (+)-limonene synthase from *C. sinensis* that confirms these early findings. In fact, the $K_m$ determined for our enzyme is 130µM, more than 100-fold greater than the published $K_m$ for enzyme from *C. limon*, and shows no indication of substrate inhibition.

As of the time of this publication, no structures have been published of (+)-4R-limonene synthase. The optimization of expression and purification of this enzyme from *Escherichia coli* has allowed us to obtain enzyme of high enough purity to crystallize the (+)-limonene apoprotein and perform initial soaking experiments with Mn$^{2+}$ and GPP substrate, in an attempt to observe bound reaction intermediates in the enzyme active site. These results, when compared with the known structure for (-)-4S-limonene synthase described above, represent a significant step towards pinpointing the particular structural elements involved in the enzyme’s reaction mechanism.

**Synechocystis**

*Synechocystis* sp. PCC 6803 is a photoautotrophic freshwater cyanobacterium. The physiology and genetics of this organism have been well-documented [20]. It is known that the *Synechocystis* genome consists of a single, 3.6 Mbp chromosome in addition to several plasmids of various sizes [21], and that it is able to spontaneously take up foreign DNA into its own genome by double homologous recombination with very high efficiency [22, 23], making it a popular candidate for the recombinant expression of various proteins.
*Synechocystis* utilizes photosynthetic oxygen-generating machinery very similar to that of plants. It has been shown that the genes that code for photosystem II in *Synechocystis*, *psbA2* and *psbA3*, are redundant, meaning that *Synechocystis* can continue to function normally in the absence of either one of these two genes [24]. One study has been able to successfully replace *psbA2* with the isoprene synthase gene from *Pueraria montana* (kudzu) [25], allowing isoprene synthase to be expressed under the *psbA2* promoter. This resulted in isoprene being produced photosynthetically by the *Synechocystis* [26]. This finding is significant because, as a combustible hydrocarbon, isoprene produced this way is a model for the photosynthetic production of related terpene biofuels. In fact, the manipulation of various microorganisms to produce bulk amounts of biofuels is currently an active field of study [27, 28]. While this method has not been used yet to produce more than one recombinant protein at a time in *Synechocystis*, a different study, which was able to heterologously express several different [NiFe] hydrogenases polycistronically in the related cyanobacterium *Synechococcus elongatus* [29], suggests that polycistronic expression in *Synechocystis* may be possible as well, allowing for the photosynthetic production of more complex hydrocarbons.

In this study we complete the initial stages necessary to engineer a pathway in *Synechocystis* which would produce (+)-limonene photosynthetically. Because limonene contains more energy per molecule and has a much higher boiling point than does isoprene (176 °C vs. 34 °C), it has more tangible applications as a biofuel. However, the
heterologous production of limonene is also more complex than that of isoprene, as it must occur in two steps: the production of GPP first from IPP and DMAPP subunits, and the cyclization of GPP into (+)-limonene second. Accordingly, this pathway requires two enzymes to be functionally expressed in *Synechocystis*: GPP pyrophosphate synthase for the production of GPP, and (+)-limonene synthase for the cyclization into the final (+)-limonene product. In this thesis, we describe the first steps taken to insert the GPP synthase and (+)-limonene synthase genes into the *Synechocystis* genome, following a protocol modeled closely after that of Lindberg et al. for the production of isoprene in the same organism [26].
CHAPTER 1:

LIMONENE SYNTHASE EXPRESSION AND PURIFICATION OPTIMIZATION

MATERIALS AND METHODS

Expression

The gene used to express (+)-4R-limonene synthase has been isolated in the lab previously from *C. sinensis* (orange), truncated to remove the plastidial targeting sequence, and tagged with an N-terminal 6×His tag. The truncated and tagged gene had been cloned into the pET28a vector (Novagen) and expressed in *E. coli* strain BL21-CodonPlus-RIL (Agilent Technologies), which is abundant in additional copies of tRNAs found rarely in *E. coli*, and thereby facilitates the expression of arginine-, isoleucine- or leucine rich proteins such as (+)-limonene synthase [16]. The pET28a vector confers kanamycin resistance, while the CodonPlus-RIL strain confers chloramphenicol resistance; thus, selection was performed by plating on 25 µg/mL kanamycin/34 µg/mL chloramphenicol LB plates. Starter cultures were grown overnight in LB supplemented with 25 µg/mL kanamycin and 50 µg/mL chloramphenicol, at 37°C and shaking at 220 rpm, and used to inoculate 1L cultures of unsupplemented LB the next morning.
Cultures were grown at 37°C to an OD$_{600}$ of 0.6, then induced with 1 mM IPTG and incubated at 37°C for 2 hours to allow for expression of the protein. For expression at 20°C, cultures were grown at 37°C to an OD$_{600}$ of 0.6, then allowed to equilibrate at 20°C for 30 minutes before induction with 1 mM IPTG, and finally incubated at 20°C for approximately 20 hours. To test for protein stability over the course of the 20-hour expression, culture samples were taken at induction and at points 3 hours, 8 hours, and 20 hours post-induction. Following expression at either temperature, cells were harvested by centrifugation at 4°C and stored in pellet form at -80°C.

**Purification**

Frozen cell pellets were thawed and resuspended in 50 mL breaking buffer (50mM Tris, 100 mM NaCl, 20 mM imidazole, 10 µg/ml lysozyme, 10 µg/ml DNase I, pH 7.5) supplemented with a protease inhibitor cocktail (La Roche) per the manufacturer’s instructions. The cell suspension were sonicated with five 20-second pulses at approximately 45 watts separated by 40-second rest periods, and then centrifuged at 16,000g at 4°C for 45 minutes. Supernatant was then filtered and loaded onto a 5-mL Ni-Sepharose column (GE). After washing with wash buffer (20mM Tris, 100mM NaCl, 40mM imidazole, pH 7.5), limonene synthase was eluted using a linear 40-500mM imidazole gradient (also in 20mM Tris, 100mM NaCl, pH 7.5) over 12 column volumes and stored short-term at 4°C.
RESULTS

Purification of the His-tagged (+)-limonene synthase using a Ni-Sepharose column (GE Technologies) gave a clean elution profile (Figure 4). Protein was eluted using an imidazole gradient, with the bulk of the protein eluting between 60 and 140 mM imidazole.

![Limonene synthase elution profile](image)

**Fig. 4 Limonene synthase elution profile** – Purification on a Ni-Sepharose column yielded a clean elution profile with a single peak. Typically, limonene synthase eluted at imidazole concentrations between 60 and 140 mM.

Previously, purification had been performed using the breaking and wash buffers listed in the absence of imidazole. It was hypothesized that adding low concentrations of imidazole to both of these buffers would increase the purity of the final product by discouraging nonspecific binding of untagged protein to the Ni-Sepharose column. In fact, it was found that the addition of 20 mM imidazole to the breaking buffer and 40 mM imidazole to the wash buffer led to a significantly purer elution of limonene synthase. The purity of the limonene synthase product was assessed by Coomassie-staining of
purified fractions on SDS-PAGE gels, which showed that the addition of imidazole lead to a reduction of contaminants in these fractions (see Figure 5, A and B).

**Figure 5 – Optimization of expression and purification**– Protein bands corresponding to limonene synthase in eluted fractions are shown outlined in rectangles. Purification with the addition of imidazole to the breaking and wash buffers (B) yielded a significantly purer product than purification using imidazole-free buffers (A), as indicated by the presence of fewer contaminating bands on Coomassie-stained SDS-PAGE gels. Expression at 20°C (D) instead of at 37°C (B) corresponded to an almost 10-fold increase in yield. Limonene synthase could be expressed at 20°C for up to 20 hours without any noticeable product degradation (C).

We noted that limonene synthase degrades if expressed at 37°C for longer than about 2 hours. To test whether a lower temperature could lead to a more stable, active enzyme, expression was tested at 20°C. At this temperature limonene synthase could be
expressed for as long as 20 hours without any detectable degradation (Figure 5, C).

Furthermore, purification of protein expressed at 20°C yielded significantly more enzyme than at 37°C. This was observed qualitatively by Coomassie-staining (Figure 5, B and D) as well as quantitatively – typical yield for a 1 L culture expressed at 37°C was 2-3 mg, while a typical yield for the same volume of culture expressed at 20°C was 15-25 mg, corresponding to an almost 10-fold increase in yield.
CHAPTER 2:

KINETIC CHARACTERIZATION

MATERIALS AND METHODS

Kinetic Assays

The activity of purified (+)-4R-limonene synthase was measured according to the single-vial assay technique described by O’Maille et al. [30]. Reactions were run in 1-mL volumes containing ~0.1 µM enzyme in reaction buffer (15mM MOPS [pH 7.0], 2 mM DTT, 10 mM MgCl₂) supplemented with GPP substrate (Echelon Biosciences) at concentrations varying from 25-600 µM. Immediately following mixing, the reactions were overlaid with 1-mL ethyl acetate and run at room temperature in septum-capped glass vials. Incubation time was varied between 0-40 minutes with several time points taken for each substrate concentration, to allow for determination of product formation velocity. Reactions were then terminated by vigorous vortexing for 30-45 seconds, thereby also extracting the limonene product into the ethyl acetate. The organic layer was then extracted and analyzed for product formation by GC-MS, as described below. Reaction velocity data were plotted and analyzed using Microsoft EXCEL.
Quantification by GC-MS

Ethyl acetate samples were analyzed for the presence of limonene using a GC unit (Agilent Technologies 7890A, column model HP-5HS) coupled to a mass selection detector (Agilent Technologies 5975C). The GC was run at a He flow rate of 3 mL/min. An automated injection of 5 μL was used, with a split ratio of 50:1 and an injector temperature of 220°C. Initial oven temperature was 50°C, and temperature was increased at a rate of 13°C/min to a temperature of 240°C. A solvent delay of 2.70 minutes was included in the run time. A peak corresponding to limonene was consistently observed at 6 minutes and 22 seconds post-injection (Figure 6, A).

**Fig. 6 Product analysis by GC-MS** – During separation by gas chromatography, a peak corresponding to limonene was routinely observed at 6 minutes and 22 seconds (A). The product was successfully identified as limonene by mass spectrometry (B).
The MSD was operated at 70 keV, with a source temperature of 230°C and a quad temperature of 150°C. Product was successfully identified as limonene by mass spectrometry (Figure 6, B). Following data collection, product peaks were analyzed using the program TCMSD Data Analysis (version E.20.00.493, Agilent Technologies).

RESULTS

Activity of Enzyme Expressed at 20°C vs. 37°C

In addition to being significantly more stable during expression, enzyme expressed at 20°C was also found to be more active than enzyme expressed at 37°C. When the same assay was performed with both 20°C and 37°C enzyme in the presence of 200 µM GPP, it was found that a reaction with the 20°C enzyme at a concentration of 0.07 µM had a velocity roughly comparable to a reaction with 37°C enzyme at a concentration of 0.26 µM, giving rise to $k_{cat}$ values of 0.13 s$^{-1}$ and 0.04 s$^{-1}$ for enzyme expressed at 20°C and 37°C, respectively (Table 1). These results indicate that limonene synthase is roughly 3 times more active when expressed at 20°C than when expressed at 37°C.

<table>
<thead>
<tr>
<th>Expression Temp.</th>
<th>Enzyme Conc.</th>
<th>Product t = 10 min</th>
<th>Reaction Velocity</th>
<th>Approx. $V_{max}$</th>
<th>Approx. $k_{cat}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20°C</td>
<td>0.07 µM</td>
<td>4.09 µM</td>
<td>0.38 µM/min</td>
<td>0.53 µM/min</td>
<td>0.13 s$^{-1}$</td>
</tr>
<tr>
<td>37°C</td>
<td>0.26 µM</td>
<td>4.35 µM</td>
<td>0.49 µM/min</td>
<td>0.64 µM/min</td>
<td>0.04 s$^{-1}$</td>
</tr>
</tbody>
</table>

Table 1 Enzyme activity of protein expressed at 20°C and 37°C – In a side-by-side comparison, it was found that enzyme expressed at 20°C is roughly 3 times more active than enzyme expressed at 37°C.
The decreased protein degradation during 20°C expression described in Chapter 1, and the increased stability and activity of limonene synthase expressed at 20°C, described here, indicates that the protein is better able to fold into its proper conformation at this temperature than at 37°C. This is not surprising – at 20°C *E. coli* function is slowed, allowing the protein to be expressed and processed more slowly, and at this lower temperature the folded protein itself also experiences less thermal agitation. Furthermore, because the physiological temperature for the navel orange (from which this limonene synthase is cloned) is expected to be around 20°C, enzyme folded at this temperature is likely to resemble its native confirmation more closely than enzyme folded at higher temperatures. Thus, all further kinetic analysis was done solely with enzyme expressed at 20°C.

**Kinetic Characterization of Limonene Synthase**

Limonene synthase activity was quantified using the method described in Materials and Methods. The rates of limonene production at various GPP concentrations, normalized to enzyme concentration, are shown in Figure 7A. The dependence of the reaction velocity on GPP concentration appeared to agree with the classical kinetic relationship between reaction rate and substrate concentration as described by the Michaelis-Menten equation,

\[
\frac{v}{K_m + [S]} = \frac{k_{cat}}{K_m} [E] [S]
\]

where \(v\) is the reaction velocity, \(k_{cat}\) is the turn-over number for the enzyme, and \(K_m\) is the substrate concentration at which the reaction velocity is at half of its maximum possible
value. A plot of the observed reaction rates, fitted with a Michaelis-Menten curve, is shown Figure 7C.

<table>
<thead>
<tr>
<th>[GPP]</th>
<th>v/[E] sec⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 µM</td>
<td>0.013</td>
</tr>
<tr>
<td>50 µM</td>
<td>0.029</td>
</tr>
<tr>
<td>75 µM</td>
<td>0.040</td>
</tr>
<tr>
<td>100 µM</td>
<td>0.047</td>
</tr>
<tr>
<td>150 µM</td>
<td>0.064</td>
</tr>
<tr>
<td>200 µM</td>
<td>0.075</td>
</tr>
<tr>
<td>300 µM</td>
<td>0.073</td>
</tr>
<tr>
<td>600 µM</td>
<td>0.079</td>
</tr>
</tbody>
</table>

FIG. 7 Limonene synthase kinetic analysis – The relationship between the observed rates of product formation and initial substrate concentrations (A) were plotted and fitted to the Michaelis-Menten equation (C) and the Lineweaver-Burke plot (B). Analysis of these two plots indicates that the $k_{cat}$ for the reaction is 0.11 s⁻¹, and the $K_m$ is 130 µM.
The double-reciprocal Lineweaver-Burke relationship, according to the equation

\[
\frac{1}{v} = \frac{K_m}{k_{cat}[E]} \cdot \frac{1}{[S]} + \frac{1}{k_{cat}[E]}
\]

was also plotted and used to determine the values of \(K_m/k_{cat}\), and \(1/k_{cat}\). A plot of the observed reaction rates, fitted with a Lineweaver-Burke line, is shown in Figure 7B. Analyses of both the Michaelis-Menten and Lineweaver-Burke plots indicate that the \(k_{cat}\) for our recombinant (+)-limonene synthase is 0.11 s\(^{-1}\), and the \(K_m\) is 130 μM.
CHAPTER 3:

STRUCTURE DETERMINATION BY CRYSSTALLIZATION

MATERIALS AND METHODS

Crystallization and Soaking

All crystallization and soaking experiments were done with Dr. P. Ramasamy. Enzyme purified as described previously (Chapter 1) was used to grow crystals. Crystal trays were set up using the Phoenix crystallization robot (Art Robbins Instruments), with a stock enzyme solution at 15 mg/ml and crystallization buffer (12-16% PEG-8000, 100 mM Tris-HCl pH 7.5-9.0, 200-350 mM sodium tartrate) in a 1:1 ratio. Crystals were grown at 20°C and typically appeared after 10-15 days.

For soaking, crystals still in their original crystallization solutions were supplemented to 10 mM MnCl$_2$ and 15% glycerol (final concentrations), and incubated at room temperature for 1 hour. Crystals were then transferred to a new solution containing 500 µM geranyl pyrophosphate substrate, 10 mM MnCl$_2$ and 15% glycerol, and incubated for variable lengths of time (0-10 min, with 30 second intervals). In this study,
we analyze the structures of crystals soaked in geranyl pyrophosphate substrate solution for 0, 1, and 4 minutes, which will be referred to as $t = 0$, $t = 1$, and $t = 4$, respectively.

**Data Collection and Refinement**

After soaking in substrate solution, crystals were flash frozen in liquid nitrogen. Data were collected at beam line 8.2.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA) using ADSC Q315R CCD detectors (Area Detector Systems Corporation) at a temperature of 100 K. Data were integrated using MOSFLM [31] and scaled using SCALA [32] from the CCP4 software suite v6.3 [33, 34].

Structures were solved by molecular replacement with the help of PHASER, [35] using a previously determined crystal structure of the enzyme apoprotein as a model (P. Ramasamy, not yet published). Refinement was performed using the function phenix.refine [36] in PHENIX software suite v1.8, [37] and model building was done in COOT v0.7 [38].

The statistics from data collection and the early stages of refinement are shown in Table 2. However, all three structures require further refinement prior to submission to the PDB.
Table 2

<table>
<thead>
<tr>
<th>Soaking Time</th>
<th>0 min</th>
<th>1 min</th>
<th>4 min</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space Group</td>
<td>P4,2,2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>68-3.45</td>
<td>67-3.2</td>
<td>72-2.9</td>
</tr>
<tr>
<td>Highest res. range (Å)</td>
<td>3.64-3.45</td>
<td>3.37-3.2</td>
<td>3.08-2.9</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td>a = b = 86.06, c = 222.63</td>
<td>a = b = 85.5, c = 215.91</td>
<td>a = b = 86.19, c = 216.57</td>
</tr>
<tr>
<td></td>
<td>α = β = γ = 90°</td>
<td>α = β = γ = 90°</td>
<td>α = β = γ = 90°</td>
</tr>
<tr>
<td>Total Reflections*</td>
<td>149,990 (22,261)</td>
<td>141,341 (19,404)</td>
<td>235,862 (40,348)</td>
</tr>
<tr>
<td>Unique Reflections*</td>
<td>21,030 (3,034)</td>
<td>24,044 (3,492)</td>
<td>18,945 (2,987)</td>
</tr>
<tr>
<td>Completeness %*</td>
<td>98.9 (98.6)</td>
<td>94.2 (94.2)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Rmerge *</td>
<td>0.17 (0.47)</td>
<td>0.08 (0.46)</td>
<td>0.113 (0.460)</td>
</tr>
<tr>
<td>I/σI*</td>
<td>6.9 (3.1)</td>
<td>14 (3.5)</td>
<td>13.2 (4.6)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.1 (7.3)</td>
<td>5.9 (5.6)</td>
<td>12.4 (13.4)</td>
</tr>
</tbody>
</table>

**Refinement Statistics**

| Resolution Range (Å) | 67-3.45 | 67-3.2 | 67-2.9 |
| # reflections used | 13,077 | 14,148 | 20,943 |
| Rwork | 0.24 | 0.20 | 0.21 |
| Rfree | 0.30 | 0.25 | 0.26 |
| Mn²⁺ atoms | 1 | 2 | 3 |
| R_{bad}, bond length (Å) | 0.011 | 0.009 | 0.009 |
| R_{bad}, bond angle (°) | 1.3 | 1.2 | 1.2 |

**Table 2 Data collection and refinement statistics** – Here we show the statistics for the data collection and initial refinement of crystals soaked in geranyl pyrophosphate substrate for 0, 1, and 4 minutes, respectively.

*Values in parentheses are for highest resolution bin.

**RESULTS**

All three crystals were found to be of the P4,2,2 space group and had very similar unit cell dimensions (a = b ≈ 86, c ≈ 220, α = β = γ = 90°, Table 2). Our (+)-4R-limonene synthase has an overall fold very similar to that of (-)-4S-limonene synthase described previously by Hyatt et al. [14] Two distinct domains can be observed in the structure: a C-terminal domain, containing the active site, and an N-terminal domain, both of which resemble those published previously for (-)-4S-limonene synthase (Figure 8). However,
unlike Hyatt et al., we were not able to model the N-terminal tail in our structure, indicating that this region is too disordered to generate a significant electron density.

**Fig. 8 Apoprotein crystal structure** – In our structure, we observe the same two domains previously described for (-)-limonene synthase [14]. The C-terminal domain is shown in green, with the three bound manganese atoms (shown in purple) observed after soaking in geranyl pyrophosphate solution, coordinated by four aspartates: D348, D352, D493, and D494. The N-terminal domain, which is missing the N-terminal tail, is shown in blue.

Initial refinement of the $t = 0$ structure showed a strong $3\sigma |F_o-F_c|$ positive density near D348, the first of the aspartates of the DDxxD motif, prompting us to manually model a manganese at this location (Figure 9, A). Initial refinement of the $t = 1$ structure also showed this density, and further refinement using this model also generated a second
positive density 3.7 Å away, near D352, the third aspartate of the DDxxD motif, leading us to add a second manganese (Figure 9, B). Initial refinement of the t = 4 structure similarly showed the presence of two manganese cations in the vicinity of D348 and D352, as well as additional positive density indicating a third manganese coordinated by D493 and D494 (Figure 9, C).

**FIG. 9 Manganese atoms bound to the active site** – Only one manganese atom was observed for the t = 0 crystal (no substrate soaking) (A). For a crystal soaked in geranyl pyrophosphate solution for 1 minute, two manganese atoms were observed, coordinated by D348 and D352 of the DxxD motif (B). For a crystal soaked for 4 minutes, a third, additional manganese was also observed, coordinated by D493 and D494 (C). All electron densities are shown at 1σ.
No density was observed that would indicate the presence of geranyl pyrophosphate in the active site of any of these three structures, or in any of the crystals analyzed, including those soaked in substrate solution for up to 10 minutes (data not shown). Furthermore, the N-terminal tail of all three structures was found be highly disordered, indicating that the tail was not its “closed”, active site-capping conformation. A comparison of our structure to that published by Hyatt et al. further revealed a very significant distinction: the previously published substrate-bound structure contains two loops which serve to cover the substrate in the active site, one from above and one from below, in addition to the N-terminal tail, which covers the active site from the front. In our structure, the loop region from 574-584, which moves to cap the substrate-bound active site from below, was also observed to be in a highly disordered state (Figure 10, B). In addition, the short, helical loop from 497-504, which caps the substrate-bound active site from above, was observed in our structure to be significantly shifted from the active site (Figure 10, C).

Additional work done by P. Ramasamy also revealed some unaccounted-for $3\sigma$ $|F_o-F_c|$ positive density in the active site of some soaked crystals, which we suspect may correspond to a pyrophosphate coordinated by the three Mn$^{2+}$ atoms. This density, which appears in certain crystals soaked in substrate for more than 1 minute but less than 4 minutes, is not strong enough to warrant the addition of an entire pyrophosphate to our model, but does nonetheless raise important questions concerning the occupancy of geranyl pyrophosphate in the active sites of soaked crystals (P. Ramasamy, data not yet published).
Fig. 10 Positioning of the N-terminal tail and the active site loops – A comparison of the substrate-bound (+)-limonene synthase structure previously published by Hyatt et al. (gold/red) and our t = 4 apoprotein structure (light blue/dark blue). Manganese atoms are shown in purple. The N-terminal tail of our structure was found to be too highly disordered to model accurately (A). The 574-584 loop, which caps the active site from below in the substrate-bound structure, was similarly quite disordered (B). In addition, the small helical 497-504 loop, which caps the active site from above in the substrate-bound structure, was found to be significantly removed from the active site in our structure (C).
CHAPTER 4:
LIMONENE PRODUCTION IN SYNECHOCYSTIS

MATERIALS AND METHODS

Synechocystis Culture

*Synechocystis* sp. PCC 6803 starter culture was obtained from Professor Mary Allen of Wellesley College (Wellesley, MA). Liquid cultures were grown in BG11 mineral media (see Appendix, Figure A1) in spinner flasks at room temperature with 10 μmol photons/m²/s provided by a 20W cool white fluorescent bulb. For plating, BG11 media was supplemented with 1% agar.

Synechocystis DNA Extraction

To extract DNA, cells were first macerated in TE buffer, then pelleted by centrifugation for 2 min at 13,000 rpm at room temperature and resuspended in 500 μL TE buffer. 50 μl of 50 mg/ml lysozyme was added to the suspension. After a 15 minute incubation at room temperature, the suspension was brought to 1% SDS, incubated an additional 15 minutes at 70°C, and cooled to room temperature. One volume of 1:1 phenol:chloroform was added, and the mix centrifuged for 10 minutes at room temperature, after which the supernatant was extracted twice using 100% chloroform. The supernatant was supplemented with 0.1 volume 3M sodium acetate and 2 volumes 100% ethanol and kept
at -20˚C for 2 hours. The precipitant was then collected by centrifugation for 10 min at 4˚C at 10,000 rpm, washed with 500 µl of 100% ethanol, centrifuged again, and allowed to air dry. The DNA pellet was resuspended in 100 µl TE and stored at -20˚C.

**Plasmid Construction**

*E. coli* strain DH5α (New England Biolabs) was used for the following cloning and plasmid propagations. The method for constructing a plasmid for the replacement of the *Synechocystis* gene *psbA2* with the geranyl pyrophosphate synthase (GPPS) and limonene synthase (LS) genes by double homologous recombination was modeled after that of Lindberg et al. [26]. The 0.5 kb regions of the *Synechocystis* genome immediately upstream and downstream of the *psbA2* gene (see Appendix, Figure A2) were amplified by PCR using the following primers, designed by Lindberg et al.: A2us_Eco_F and A2us_NdeI_Bam_R for amplification of the upstream region, and A2ds_Bam_F and A2ds_SacI_R for amplification of the downstream region (see Table 3 for primer sequences). The upstream fragment was cloned into the *Eco*RI and *Bam*HI sites of the pBluescript KS+ vector (Stratagene) first; this fragment also included a *Nde*I site slightly upstream of *Bam*HI site to allow for later insertion of the GPPS and LS genes in-frame with the translation start site for the *psbA2* gene, allowing us to make use of all the initiation factors normally used by *Synechocystis* to translate the gene at this location.
Table 3

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2us_Eco_F</td>
</tr>
<tr>
<td>2</td>
<td>A2us_NdeI_Bam_R</td>
</tr>
<tr>
<td>3</td>
<td>A2ds_Bam_F</td>
</tr>
<tr>
<td>4</td>
<td>A2ds_SacI_R</td>
</tr>
<tr>
<td>5</td>
<td>GPPS_BamHImut_for</td>
</tr>
<tr>
<td>6</td>
<td>GPPS_BamHImut_rev</td>
</tr>
<tr>
<td>7</td>
<td>GPPS_NdeI_for</td>
</tr>
<tr>
<td>8</td>
<td>GPPS_1D4_NarI_rev</td>
</tr>
<tr>
<td>9</td>
<td>LS_NarI_C8_for</td>
</tr>
<tr>
<td>10</td>
<td>LS_BamHI_rev</td>
</tr>
<tr>
<td>11</td>
<td>IsoA_BamHI_F</td>
</tr>
<tr>
<td>12</td>
<td>IsoA_NdeI_R</td>
</tr>
<tr>
<td>13</td>
<td>IsoA_GPPS_F</td>
</tr>
<tr>
<td>14</td>
<td>IsoA_LS_R</td>
</tr>
</tbody>
</table>

**TABLE 3 Primers used in plasmid construction** – Primers 1-4 were used for the PCR amplification of the fragments of the upstream and downstream of the *Synechocystis psbA* gene. Primers 5 and 6 were used of the mutagenesis of the GPPS gene. Primers 7-10 were used for the amplification of the LS and GPPS’ genes. Primers 11-14 were used for cloning by Gibson (isothermal) assembly.

To perform the cloning, the KS+ vector and the upstream PCR-amplified fragment were digested with *Eco*RI and *Bam*HI overnight at 37°C. Both digested pieces were then gel purified using the QIAquick Gel Extraction Kit (Qiagen), and combined in a ligation reaction with T4 DNA ligase overnight at 4°C. The resultant ligation mixture was used to transform DH5α cells. The downstream fragment was then cloned into the *Bam*HI and *Sac*I sites of the resulting vector following the same protocol as above but with a step-wise digest, digesting first with *Bam*HI and then with *Sac*I, to avoid star activity. In this case the sample was purified between digests using a QIAquick PCR...
Purification Kit (Qiagen) between digests, following the manufacturer’s protocol, to remove residual *BamHI* enzyme and buffer. After gel purification, ligation, and transformation, the final vector was confirmed by sequencing, and given the name KSusds (Figure 11).

**Fig. 11 Generating the KSusds plasmid** – The 0.5 kb sequences directly up- and downstream of the *psbA2* gene were PCR-amplified from *Synechocystis* DNA. The three restriction sites indicated in the figure were then used to subclone the two fragments into the pBluescript KS+ vector to give the plasmid KSusds. An *NdeI* restriction site in the upstream fragment allows insertion of genes at this site that are in frame with the *psbA2* start site.

The gene for GPPS was synthesized in the pUC57 vector (Genewiz) using the known sequence for GPPS from *Abies grandis* (GenBank accession number AF513111) [39]. Before the gene could be cloned into the KSusds vector, it was necessary to remove the *BamHI* cut site located at position 496 of GPPS (see Appendix, Figure A3). This was done with a silent point mutation, using the primers GPPS_BamHImut_for and GPPS_BamHImut_rev to change the cut site, 5’-GGG GAT CCA-3’, to 5’-GGA GAT
CCA-3’ (see Table 3 for primer sequences). The substitution was confirmed by sequencing, and the gene designated GGPS’.

Next, primers were designed to allow insertion of the GPPS’ and LS genes, in tandem, into the Ndel and BamHI restriction sites of KSusds by a three-part ligation, as well as to add a carboxy-terminal 1D4 antibody tag to GPPS’ and an amino-terminal C8 antibody tag to LS. A unique tag was used for each gene so that the two enzymes could be differentially identified by immunoblotting and purified after expression in Synechocystis. Primers GPPS_NdeI_for and GPPS_1D4_NarI_rev were used to PCR amplify and tag the GPPS’ gene, and primers LS_NarI_C8_for and LS_BamHI_rev were used to PCR amplify and tag the LS gene (see Table 3 for primer sequences). The NarI restriction site found in the 1D4 tag was used to connect the two genes by extending the 5’ end of the LS_NarI_C8_for primer to include not only C8 tag sequence, but also the end of the 1D4 tag, starting with the NarI restriction site. The LS_NarI_C8_for primer also contained the stop codon for the 1D4 tag and the start codon for the C8 tag, which were made to overlap using the 5-nucleotide sequence 5’-TAATG-3’, in the hope that this would encourage polycistronic expression of the two tagged genes (Figure 12). The KSusds plasmid and GPPS’ and LS PCR-amplified fragments were digested with the appropriate restriction enzymes overnight at 37°C, purified, and ligated together with T4 DNA ligase overnight at 4°C before transformation. As a control, the same three-part ligation was later performed using the GPPS’ and LS genes and the pUC19a vector, which also contains Ndel and BamHI restriction sites placed approximately 230 basepairs apart.
The GPPS’ and LS genes were PCR amplified out of their respective vectors using primers that added a C-terminal 1D4 antibody tag to GPPS’ and a N-terminal C8 antibody tag to LS. The stop codon TAA of the 1D4 tag and the start codon ATG of the C8 tags were made to overlap in order to encourage polycistronic expression in *Synechocystis* (A). A NarI restriction site found in 1D4 was supposed to allow assembly of the desired plasmid by a three-part ligation (B).

To transfer the upstream and downstream fragments to the pBluescript SK+ vector (Stratagene), the cloning procedure described previously was repeated, using restriction enzymes *Eco*RI and *Sac*I to generate the new vector, SKusds, which was confirmed by sequencing. We then attempted to insert the GPPS’-LS tandem gene fragment previously assembled in pUC19a into the Ndel and BamHI sites in SKusds. As a control, a 250 basepair-long *Ndel*-BamHI fragment excised from the pUC19a vector was also cloned into these same two sites.

When cloning into SKusds by restriction digest proved ineffective, we tried GPPS’ and LS insertion using an alternate approach recently described by Gibson et al. for plasmid assembly in a single, isothermal reaction (Figure 13) [40]. Primers were designed
to allow PCR amplification of the SKusds vector from BamHI to NdeI and of the GPPS'-LS tandem insertion from pUC19a in such a way that the sequences of the two fragments would overlap by ~30 basepairs at each end. Primers IsoA_BamHI_F and IsoA_NdeI_R were used to amplify the SKusds fragment, and IsoA_GPPS_F and IsoA_LS_R were used to amplify GPPS’ and LS (see Table 3 for primer sequences). PCR amplification was performed using 1-2 ng of template DNA and 31 thermocycles. PCR product was then digested with DpnI to remove residual template and fragments were purified by gel extraction. Cloning was performed in a single 20µl reaction volume containing 100 ng of vector backbone DNA and 60 ng of GPPS’-LS insert in buffer (5% PEG-8000, 100mM Tris-Cl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 0.2 mM each of the four dNTPs, 1 mM NAD) supplemented with T5 exonuclease, Phusion polymerase, and Taq ligase as described by Gibson et al. [40] The reaction was incubated at 50°C for 1 hour before transformation into DH5α cells.

**Fig. 13 Gibson (isothermal) assembly** – In the plasmid assembly method described by Gibson et al., overlapping DNA fragments are incubated with together T5 exonuclease, Phusion polymerase, and Taq ligase in a single reaction at 50°C for up to 1 hour. Initially, the T5 exonuclease chews away at the 5’ ends of both fragments, but as the reaction progresses this enzyme denatures. The complimentary ends of the two fragments can then anneal, while the thermophylic polymerase and ligase complete the assembled plasmid. Figure courtesy of Gibson et al. [40]
RESULTS

Both the purification of *Synechocystis* DNA and the PCR amplification of the 0.5 kb fragments directly upstream and downstream of the *psbA2* gene were successful. The upstream fragment was inserted first into the *Eco*RI and *Bam*HI restriction sites of the pBluescript KS+ vector, and the downstream fragment was inserted second into the *Bam*HI and *Sac*I restriction sites of the resulting vector. The plasmid containing both flanking sequences was given the name “KSusds” and confirmed by both restriction digest (Figure 14) and sequencing. Mutagenesis of the GPPS gene to remove the *Bam*HI restriction site at position 496 by a silent mutation was also confirmed by sequencing, and was similarly successful.

![Fig. 14 Restriction digest of KSusds](image)

Lane 1 – KSusds plasmid, no digest
Lane 2 – *Eco*RI/*Bam*HI digest
Lane 3 – *Sac*I digest
Lane 4 – *Bam*HI/*Sac*I digest
Lane 5 – *Eco*RI/*Sac*I digest

Fig. 14 Restriction digest of KSusds – The completed KSusds plasmid was confirmed by restriction digest. Cutting with *Eco*RI/*Bam*HI (2) or *Bam*HI/*Sac*I (4) resulted in fragment slightly smaller than cutting with *Sac*I alone (3), and a very faint band corresponding to one of the 0.5 kb sequences could be observed. Cutting with *Eco*RI/*Sac*I (5) gave a still smaller fragment and a distinct second band corresponding to 1 kb, the combined length of the upstream and downstream sequences.

However, the next step of plasmid construction, a three-part ligation of KSusds, GPPS’ and LS (Figure 12) could not be achieved using the standard subcloning method.
described above. Despite several repeated attempts, this ligation routinely resulted in a much less efficient transformation into DH5α cells than a background (negative control) transformation of just the restriction digested KSusds vector, and analysis by restriction digest of any successfully transformed clones consistently showed that the GPPS’ and LS were not present in the produced plasmid. This remained the case even if double-digestion was done step-wise, guaranteeing that the restriction enzyme pairs did not interfere with each others’ activity. To help facilitate a proper ligation, we tried an additional digest of the KSusds fragment with CIP (calf intestinal alkaline phosphatase). Because CIP cleaves the 5’ phosphate from the ends of linear DNA, such a digest would prohibit the ends of the KSusds fragment from coming together, making a three part ligation with GPPS’ and LS more favorable. Unfortunately, even with this addition, the three-part ligation did not seem to be effective.

As a control, we tried performing the same three-part ligation using a different vector, pUC19a, which also has a BamHI restriction site at position 417 in its multiple cloning site (MCS), and an NdeI restriction site at position 183, upstream of its MCS (Figure 15). The GPPS’ and LS genes assembled perfectly this vector, and the GPPS’-LS construct was confirmed by sequencing. However, even if both GPPS’ and LS were taken from pUC19a as a single fragment (i.e., by digestion with NdeI and BamHI), they could not be subcloned into the corresponding restriction sites in KSusds. Furthermore, because the pUC19a does not contain the EcoRI and SacI sites necessary for the insertion of the upstream and downstream flanking sequences, it could not easily be used as a permanent replacement for the KS+ vector.
Because the GPPS’ and LS genes could successfully be assembled in an alternate vector, we hypothesized that the KS+ vector itself might be preventing the insertion of the genes into KSusds. For example, a leaky promoter at the start of the lacZ operon could be allowing the GPPS’ and LS synthase genes to be expressed spontaneously, generating a product (either the geranyl pyrophosphate intermediate or the limonene itself) that was toxic to the E. coli cells, thus selecting against the plasmid we were trying to construct. To test this possibility, we transferred the upstream and downstream flanking sequences to a related vector, pBluescript SK+, to give the plasmid SKusds, and attempted to insert the GPPS’-LS fragment from pUC19a once again. The SK+ vector is identical to the pBluescript KS+ vector with the exception of MCS, which contains all the same restriction sites but in the reverse order (Figure 16). Thus, if all the necessary DNA fragments are assembled in the SK+ vector, they can be in the correct order relative to each other, but backwards relative to the lacZ operon, so that even a leaky promoter cannot have an adverse effect on cells carrying the plasmid.

**FIG. 15 The pUC19a vector** – As a control, the GPPS’ and LS genes were assembled in the pUC19a vector, which also has an NdeI restriction site (183) and a BamHI restriction site in the multiple cloning site (417). The GPPS’ and LS fragments assembled perfectly in the pUC19a vector. (Figure courtesy of Thermo Scientific)
Fig. 16 -- The pBluescript vectors – The pBluescript KS+ and SK+ vectors share identical features with the exception of the multiple cloning site, which contains all the same restriction sites in both vectors, but in one order (KpnI-SacI) in KS+, and in the reverse order (SacI-KpnI) in SK+. When assembled in SK+ instead of KS+, the GPPS’ and LS genes should not be able to be expressed, even in the case of a leaky lacZ promoter. (Figure courtesy of Agilent Technologies)

However, repeating the cloning using the new SKusds plasmid still did not result in a correctly assembled plasmid. To test the functionality of the NdeI and BamHI insertion sites in the SKusds plasmid, we also repeated the cloning using a control fragment: the ~230 bp sequence between the NdeI and BamHI restriction sites in the pUC19a vector (Figure 15). Surprisingly, we found that, unlike the GPPS’-LS insertion, the NdeI-BamHI fragment from pUC19a subcloned easily into the appropriate sites in the SKusds plasmid, and could be confirmed by sequencing. Thus, it would seem that both the GPPS’-LS and SKusds are able to undergo subcloning using the standard molecular biological methods under certain conditions. Nonetheless, when used together, they do not assemble as expected.

As an alternate approach to subcloning, we also attempted to complete our construct using the protocol for isothermal assembly recently determined by Gibson et al. [40]. Using this method, PCR-amplified DNA fragments with 10-100 bp sequence overlap at either end can be assembled into a single plasmid during a single 50°C reaction
in which a heat-sensitive 5’ exonuclease degrades the 5’ ends of both sequences, allowing the complementary ends to anneal, while thermophilic polymerase and ligase complete the construct (Figure 13). In our work, it so happened that the PCR-amplified segment of the SKusds vector was virtually the same length as the vector itself. It was also found that the pUC19a vector carrying the GPPS’ and LS genes, in its super-coiled form migrated the same distance on an agarose gel as the linear GPPS’-LS fragment (Figure 17).

![Image](image.png)

**Fig. 17 Relative plasmid and fragment lengths for isothermal assembly** – In our isothermal assembly reaction, the length of the SKusds plasmid and the BamHI-Ndel fragment are virtually the same (lanes 1 & 2). In addition, the super-coiled form of the pUC19a plasmid with the GPPS’-LS insert migrates the same distance on an agarose gel as does the linear GPPS’-LS fragment itself (lanes 3 & 4). Thus, separating these two fragments from the template DNA used in the cloning could not be done by simple gel purification, necessitating the additional of a DpnI digest post-PCR, to digest any template plasmid contaminant.

Thus, these fragments could not be separated from the template DNA in each PCR reaction simply by gel purification. To minimize contamination of both DNA fragments by the template plasmids, we added an additional post-PCR digest with DpnI, which digests methylated DNA, to the protocol published by Gibson et al.

However, despite this extra precaution taken to reduce plasmid contamination, our isothermal construction did not yield any correctly-assembled plasmids. Instead, sequencing of several clones transformed with the isothermally assembly reaction product repeatedly showed that the only plasmid being propagated was, surprisingly, the original SKusds plasmid. Thus, this method has so far been similarly unsuccessful in constructing plasmid capable of transforming *Synechocystis* with the GPPS’-LS tandem gene pair.
DISCUSSION

Kinetic Characterization

Our analysis of (+)-4R-limonene synthase cloned from *C. sinensis* (orange) found that this enzyme has a $K_m$ of 130µM. This is consistent with the preliminary kinetic analysis of the same clone performed in the lab previously [16]. A previous characterization done by Lücker et al. [11] had found that the $K_m$ of (+)-limonene synthase cloned from *C. limon* (lemon) was only 0.70 µM, less than 1/100 of the value we determined for our clone. The value published by Lücker et al. appears to agree with additional reports which have determined $K_m$ values in the low micromolar range for a wide variety of other monoterpene synthases [8, 11].

It is not clear why our (+)-limonene synthase appears to behave so differently from the one cloned by Lücker et al. from *C. limon*. However, it is important to note that the study performed by Lücker et al. also observed evidence of substrate inhibition kinetics, with activity for (+)-limonene synthase and two additional monoterpene cyclases decreasing at substrate concentrations greater than 10 µM, including substrate concentrations as high as 100- and 180 µM. Thus, the reported $K_m$ of 0.70 µM for (+)-limonene synthase from *C. limon* was subsequently determined by ignoring all substrate inhibition effects (i.e., excluding data from reactions run at geranyl pyrophosphate concentrations greater than 10 µM). Contrary to the findings of Lücker et al., the data from the reactions observed in our study do not show any indication of substrate inhibition, even at substrate concentrations as high as 600 µM. Rather, the reaction rate
continues to increase with substrate concentration as the reaction approaches its maximum velocity. Furthermore, the maximum reaction velocity per enzyme molecule reported by Lücker et al., even when extrapolating to ignore substrate inhibition, is several orders of magnitude slower than the maximum velocity observed per molecule of our enzyme. Thus, we suspect that there is some element of the assay protocol other than the enzyme mechanism itself that is causing Lücker et al. to report evidence of a substrate inhibition behavior which is not actually characteristic of the enzyme, thereby lowering the observed $K_m$ significantly from its true value.

No $k_{cat}$ value for (+)-limonene synthase has yet been published. However, a previous study using (-)-limonene synthase extracted from Mentha × piperita (peppermint) and Mentha spicata (spearmint) determined a $k_{cat}$ of 0.3 s$^{-1}$ for native (-)-limonene synthase [15]. This value is comparable to the observed $k_{cat}$ of 0.11 s$^{-1}$ for our recombinant (+)-limonene synthase, especially given that previous studies have noted that monoterpen synthase activity can be slightly compromised by recombinant, rather than native, expression [8].

**Crystal Structure Determination**

The fold of our (+)-limonene synthase apoprotein matches that described previously by Hyatt et al. [14] for (-)-limonene synthase cocrystallized with substrate analog, indicating that our enzyme is most likely crystallizing in its native, or at least in its active, conformation, despite the absence of a ligand. In addition, the positions of all three
manganese ions observed in the t = 4 structure agree well with the manganese observed coordinating the substrate analog in the (-)-limonene structure.

In completing the crystal soaking experiments described in Chapter 3, we had hoped to trap reaction intermediates in the enzyme active site, allowing us to better understand the mechanism by which geranyl pyrophosphate substrate is converted to (+)-limonene. Previous trials had suggested that soaking with substrate and Mn\(^{2+}\) results in the appearance of pyrophosphate in the active site (P. Ramasamy, data not yet published), a phenomenon which was similarly observed in this set of soaking experiments. However, the lack of electron density corresponding to a geranyl pyrophosphate molecule in the active site suggests that substrate is not properly bound in any of the crystals observed. This supposition is further supported by the observation that the N-terminal tail is highly disordered all three structures analyzed – had substrate bound, we would expect to see an ordered, “closed” conformation of the tail as it moved to cap the active site in preparation for hydrolysis of the pyrophosphate group. In addition, the two loop regions which cap the active site from above (497-504) and below (574-584) in the substrate-bound structure by Hyatt et al. are found in our structure to be either similarly disordered, as in the case of the 574-584 loop, or else significantly shifted away from the active site, as in the case of the 407-504 helical loop. These key differences support the absence of properly bound substrate in the active site of our substrate-soaked crystals.

It is important to note that all three crystals were soaked in MnCl\(_2\) solution for a full hour before the introduction of geranyl pyrophosphate for substrate soaking. In light
of this fact, it is curious to observe that the number of bound manganese atoms increases from one bound manganese with no substrate soaking, to two bound after one minute of substrate soaking, to all three bound with four or more minutes of substrate soaking, despite the fact that the presence of the substrate itself in the active site is not observed in any of these three structures. One possible explanation for this phenomenon is that geranyl pyrophosphate is recruiting Mn$^{2+}$ ions to their appropriate positions in the active site, as if to initiate substrate binding, but is sterically hindered from actually entering the active site while the enzyme is in its rigid, crystallized formation. In a crystal formation, the active site of our soaked crystals may be constrained to a single particular conformation, preventing it from accommodating the introduction of a molecule as bulky as geranyl pyrophosphate, while still allowing it to take up the Mn$^{2+}$ ions that would be necessary for substrate binding. However, this explanation would not explain the suspected appearance of pyrophosphate in the active site.

Alternatively, it could be that the substrate is, in fact, able to enter the active site, but that inability of the N-terminal tail and the two active site loops to move to their capping positions while crystallized leaves the substrate in the active site too flexible to generate a significant electron density, giving the impression that it has not bound. Similarly, it is possible that geranyl pyrophosphate is able to enter the active site and even to undergo the hydrolysis of the pyrophosphate, but that the inability of the N-terminal tail and the two active site loops to shield the active site from the surrounding solvent results in the reaction of the remaining cation with water to form an alcohol. Both species can then diffuse from the active site, explaining the time-dependent appearance of
the suspected pyrophosphate density between 1 minute and 4 minutes (P. Ramasamy, data not yet published). Finally, it is possible that geranyl pyrophosphate enters the active site but is simply converted to (+)-limonene too quickly to be observed bound to the active site, after which the active site reverts back to its “open” conformation. In all three of these cases, the three Mn$^{2+}$ observed in the t = 4 would be required to coordinate the hydrolyzed pyrophosphate.

Regardless of the precise nature of the substrate in the crystal active site, our results show that (+)-limonene synthase is able to crystallize in its active conformation in the absence of Mn$^{2+}$, and that soaking with geranyl pyrophosphate can subsequently place all three Mn$^{2+}$ ions into their correct positions in the active site of a folded enzyme. These observations further support a reaction mechanism by which Mn$^{2+}$ is not required for correct enzyme folding, but is recruited to the active site by geranyl pyrophosphate, and must be correctly positioned in the active site prior to substrate binding and hydrolysis.

**Limonene Production in Synechocystis**

The difficulties we experienced in constructing a plasmid with which to transform the geranyl pyrophosphate synthase and (+)-limonene synthase genes into the *Synechocystis* genome remain unresolved. It has been shown that the two genes can be combined into a single fragment flanked by *NdeI* and *BamHI* sites in an alternate plasmid, pUC19a, indicating that the nucleotide sequences themselves are not toxic to the *E. coli* cells. It has similarly been shown that a ~230 basepair control fragment excised from the pUC19a
vector can be successfully subcloned into the correct Ndel and BamHI sites in the SKusds plasmid, suggesting that these restriction sites are accessible and fully functional. These results indicate that the restriction sites themselves are not the cause of the problems experienced with subcloning the GPPS’ and LS genes into the SKusds vector. The similar failure of the Gibson isothermal assembly method, which does not rely on the use of restriction enzymes, to generate the correct construct further indicates that the restriction sites in question are not the source of the problem.

It has recently come to our attention that, in addition to a translation start site, the region directly upstream of the psbA2 gene in *Synechocystis* contains a promoter region for the psbA2 gene. In *Synechocystis*, this promoter is, not surprisingly given the gene’s function, light-inducible, causing the cell to produce higher levels of photosystem II in high-light conditions. The sequence is of *E. coli* consensus type, and it has been shown that when found in *E. coli*, this promoter is constitutively active [41]. In light of this finding, it now seems very likely that our GPPS’-LS fragment, when inserted properly after the upstream fragment, is being constitutively transcribed and expressed by the *E. coli* cell, producing a build-up of a toxic compound that quickly kills the cell. This would explain why we are clearly able to subclone the upstream fragment and the GPPS’-LS fragment independent of one another, but are never able to see them correctly assembled.

It has been shown that *E. coli* can tolerate limonene up to a concentration 0.25% (v/v) [27], or 1.5 mM. It is not clear at this point whether our *E. coli*, when transformed with the GPPS’-LS insertion correctly placed behind the psbA2 promoter, could produce
limonene at a sufficiently high concentration so as to cause *E. coli* cell death. A previous study by Carter et al. [28] reports the successful production of small amounts of geranyl pyrophosphate and (-)-limonene as intermediates in the construction of (-)-carvone and (-)-carveol pathways in *E. coli*, but notes that this comes at the cost of a high metabolic burden to the cells. Thus, it remains likely that either the production of limonene, or the diversion of its precursors from key metabolic pathways, in our subcloning experiments is significantly hindering the survival of cells transformed with the desired plasmid, effectively selecting *against* our goal construct.

**Further Directions**

Further studies of (+)-4R-limonene synthase are needed to continue all three initiatives described in Chapters 2-4 of this thesis. The significant disparity observed between the kinetics of our (+)-limonene synthase from *C. sinensis* and the kinetics of the same enzyme from *C. limon* described by Lücker et al. [11] has yet to be resolved. It seems improbable that such a large inconsistency could simply be the result of speciation, given that the two enzymes demonstrate 95% sequence identity [16]. The (+)-limonene synthase gene from *C. limon* has previously been cloned, but never expressed or kinetically characterized, in our lab by S. N. Olsen [16]. It would be necessary to perform a Michaelis-Menton analysis of this *C. limon* clone in order to more fully understand the true nature of the (+)-limonene synthase reaction kinetics. Such an analysis would further allow us to determine whether the difference observed between the behavior of our enzyme from *C. sinensis* and the one from *C. limon* studied by Lücker et al. is rooted in
the enzymes’ dissimilar species of origin, or in some overlooked detail of the assay methodologies used in their respective characterizations.

Similarly, much work remains to be done to further elucidate the reaction mechanism of (+)-limonene synthase by crystal structure determination. The three structures analyzed in this study require additional refinement prior to submission to the PDB. However, these three existing structures could only be determined to a resolution of ~3.0Å; crystallization conditions need to be further optimized in order to produce more stable crystals capable of diffracting to a higher resolution and allowing us to observe more clearly the changes occurring in the active site. In addition, our study shows that soaking in substrate solution, even for as long as 10 minutes, does not yield a crystal structure which includes a properly bound geranyl pyrophosphate molecule, although the presence of a pyrophosphate at certain time points in the soaking is suspected. The exact reason for this observation is not yet understood, but it is possible that the constraints placed on the enzyme by crystal formation are inhibiting proper substrate binding or reaction. Thus, in order to observe the enzyme reaction mechanism further, it may be necessary to employ cocrystallization methods with a substrate analog, like the 2-flourogerananyl pyrophosphate used by Hyatt et al. [14], rather than soaking with substrate. The addition of such analogs for cocrystallization may also serve to stabilize the enzyme structure further, allowing the structure to be determined at higher resolutions.
Finally, the construction of a plasmid with which to transform *Synechocystis* with the geranyl pyrophosphate synthase and (+)-limonene synthase genes remains to be completed. In light of new findings that the upstream fragment contains a promoter that is constitutively active in *E. coli*, it seems likely the GPPS’ and LS genes, when inserted at their proper location in the SKusds plasmid, are simply toxic to the cell. This new hypothesis will now need to be tested by adding deactivating mutations to both genes, and repeating the subcloning into the appropriate sites in SKusds. If the deactivated genes can be successfully used to generate our desired construct, this will be strong evidence that the subcloning technique is sound, but that the product of the active GPPS’ and LS genes is too toxic for the plasmid to be carried in *E. coli*. If we find this to be the case, further work will include putting the two genes under an inducible promoter, instead of the constitutively active one found in the upstream fragment, allowing us to suppress gene expression until the entire construct has been moved into *Synechocystis*.

Once the insertion of the geranyl pyrophosphate synthase and (+)-limonene synthase genes has been completed, the only remaining step in the plasmid construction will be the addition of an antibiotic-resistance gene after the GPPS’-LS gene fragment, which will allow for selection of correctly transformed *Synechocystis* clones. In our original experimental design, we had planned to use the kanamycin-resistance gene, *KanR*, amplified from the pET28a vector, for this purpose. The region directly upstream of *KanR* gene includes a Shine-Delgarno sequence that we hope will facilitate the polycistronic expression of this gene immediately following the expression of (+)-limonene synthase. Upon completion of the plasmid, *Synechocystis* can successfully be
transformed according to previously determined protocols [22, 23], allowing us to produce (+)-limonene photosynthetically. A method for the extraction and analysis of this potential biofuel from the cell culture can then be optimized.
**APPENDIX: SUPPLEMENTARY FIGURES**

**BG11 Media Composition**

<table>
<thead>
<tr>
<th>Volume (mL)</th>
<th>Concentration (g/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>150.0 NaNO₃</td>
</tr>
<tr>
<td>10</td>
<td>7.8 K₂HPO₄</td>
</tr>
<tr>
<td>10</td>
<td>2.0 Na₂CO₃</td>
</tr>
<tr>
<td>10</td>
<td>5.8 Na₂SiO₃·9H₂O</td>
</tr>
<tr>
<td>10</td>
<td>7.5 MgSO₄·7H₂O</td>
</tr>
<tr>
<td>10</td>
<td>3.5 CaCl₂·2H₂O</td>
</tr>
<tr>
<td>10</td>
<td>0.1 diNa EDTA</td>
</tr>
<tr>
<td>10</td>
<td>0.6 each Fe citrate-citric acid</td>
</tr>
<tr>
<td>1</td>
<td>Trace Elements:</td>
</tr>
<tr>
<td></td>
<td>H₃BO₃</td>
</tr>
<tr>
<td></td>
<td>MnCl₂·4H₂O</td>
</tr>
<tr>
<td></td>
<td>ZnSO₄·7H₂O</td>
</tr>
<tr>
<td></td>
<td>NaMoO₄·2H₂O</td>
</tr>
<tr>
<td></td>
<td>CuSO₄·5H₂O</td>
</tr>
<tr>
<td></td>
<td>Co(NO₃)₂·6H₂O</td>
</tr>
</tbody>
</table>

Add salts to distilled water in order given. Bring to 1 liter with dH₂O.

**FIG. A1 BG11 mineral media composition** – The media described above was used for *Synechocystis* liquid culture and for the production of BG11 agar plates. Protocol courtesy of Professor Mary Allen, Wellesley College, Wellesley, MA.
**Synechocystis psbA2 Flanking Sequences**

5’ CTSCCAAGTATGCAGGCTCTTGCAGATCGCAATGGTGAAGAAGCTTCTCAAGGTCGAGTATGATTGCTGCTGGGTTAAA

**FIG. A2** Upstream and downstream psbA2 flanking sequences – A portion of the Synechocystis genome is shown, indicating the psbA2 gene (in italics) and the 0.5 kb fragments directly up- and downstream (in blue). These fragments were amplified by PCR and inserted into the pBluescript KS+ vector to make the plasmid KSusds.

**GPPS Mutagenesis**

5’ TTTGATTTCAAGGAGTACTTGCATTCCAAAGCAATATCAGTGAATGAGG CACTGGAGAGGGCTGTCCCACTTCGCTATCCTGAAAAA

**FIG. A3** GPPS silent mutation – The GPPS originally synthesized (Genewiz) was found to contain a BamHI restriction site, 5’-GGATCC-3’ (underlined). The silent mutation GGG → GGA was introduced, changing the former restriction site to 5’-AGATCC-3’. GPPS gene, GenBank Accession #AF513111 [39].

52
REFERENCES


