IN VITRO MULTIENZYME PATHWAY ASSEMBLY FOR ISOPRENOIDS AND ISOPRENOID PRECURSORS PRODUCTION

CHEN XIXIAN

(B.ENG. NATIONAL UNIVERSITY OF SINGAPORE)

THE THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN CHEMICAL AND PHARMACEUTICAL ENGINEERING (CPE)

SINGAPORE-MIT ALLIANCE

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Chen Xixian
Date
4th Dec 2014

ACKNOWLEGEMENTS

The biggest reward of my PhD study was not the dissertation alone but the changes I saw in myself. There are many awakening moments that change my philosophy of life entirely. All these were not possible without the dedications and constant challenges from my main supervisor, Professor Too Heng Phon. I am deeply indebted to him for his inspiration and tolerance to my mistakes. The countless psychiatric sessions Professor Too spent with me were certainly life changing and from which I learnt the attitude of a wise man who would put in 110% effort to make the world a better place. “Do something that changes the world” is one of the many motivational sentences that we heard from him. I was once skeptical about that and have transformed into deeply believing in it. This would never occur to me had I did not take the PhD course under the great mentorship of Professor Too.

I am also blessed to have Professor Gregory Stephanopoulos (MIT, ChemEng) as my co-supervisor. Being a great leader in the field of Metabolic Engineering, Professor Stephanopoulos never failed to give insightful suggestions and warm encouragement. The attachment in Professor Stephanopoulos’ lab was an eye-opening experience which I witnessed open discussion and collaborations.

I would like to offer my special thanks to other faculty members in the Singapore-MIT Alliance, Chemical and Pharmaceutical Engineering (CPE) program, especially Professor Li Zhi, Professor Saif Khan and Professor Raj Rajagopalan, who have given me valuable comments during program meetings.

It was my privilege to have worked with so many dynamic and friendly lab members. My heartfelt gratitude to Dr Zhou Kang, who would selflessly offer his warm encouragement and insightful opinions to assist the progress of my thesis study. My deep appreciation to Dr Zou Ruiyang, whose innovative ideas and thought-provoking discussions significantly helped shape the studies in the thesis. The collaboration with Dr Zhang Congqiang has been an enjoyable experience that I acquired knowledge of statistical experiment design. My gratitude to Dr Wong Long Hui and Dr Seow Kok Hui who have been great companions in my PhD years and the post-lunch soul-searching sessions would certainly be dearly missed. I am also grateful for the moral and intellectual support rendered by Dr Wan Guoqiang, Dr Zhou Lihan, Jeremy Lim, Dr Sarah Ho, Chin Meiyi, Christine Chan, Justin Tan, Seow Vui Yin and Sha Lan Jie.

This thesis is also dedicated to my parents and grandparents who keep the faith in me and give me endless love and moral support throughout the years of my PhD studies. Without which, I would not have gone as far as I am today. Lastly, I would like to acknowledge the wonderful Singapore-MIT Alliance programs, for which I have found my life partner, Cheng He, whose intelligence and passion never failed to inspire me to be a better person.

TABLE OF CONTENTS

SUMMARY

This thesis is focused on the *in vitro* reconstitution and optimization of the multienzymatic biosynthetic pathways to produce isoprenoids and isoprenoid precursors.

A significant challenge in a multi-enzymatic reaction is the need to simultaneously optimize the various steps involved to obtain high-yield of a product. In this study, statistical experimental design was employed to test the hypothesis that an optimal multienzymatic composition can be identified rapidly for high-yield *in vitro* biosynthesis. We demonstrated the synthesis of amorpha-4,11-diene (AD), a key precursor to artemisinin, from mevalonic acid (MVA) by assembling seven enzymatic steps in one-pot. Guided by Taguchi method, the AD yield was significantly improved from 5% to 20%, when the multienzymatic concentrations were optimized. Meanwhile, an inhibitory step, farnesyl pyrophosphate synthase (IspA), was identified where its product precipitated when accumulated to a sufficiently high concentration. To mitigate this limitation, the subsequent enzymatic reaction, amorphadiene synthase (Ads), was found to be a critical step whereby increasing the enzymatic activity resulted in a remarkable improvement of AD yield to approximately 100%.

Next, mechanistic investigation of the interplay among the enzymes and metabolites was carried out to unravel the regulatory behavior of *in vitro* reconstituted AD synthetic pathway. With the aid of Lin-log approximation, a *hitherto* unrecognized inhibition of ATP on Ads activity was identified. Further structural analysis indicated that the polyphosphate moiety elicited the inhibitory

effect. Hence, another novel product inhibitor, pyrophosphate, was identified that potently inhibited the Ads activity. Therefore, an ATP-recycling enzyme (pyruvate kinase) and pyrophosphate-hydrolysis enzyme (pyrophosphatase) were included in the reaction to minimize the inhibitor concentrations. As a result, the kinetics was significantly enhanced by more than 3 fold.

Recycling the pathway enzymes is cost-effective, and able to enhance the specific AD yield. Enzyme immobilization is a desirable strategy often exploited in industrial bioprocesses. Therefore, the multi-biocatalysts were co-immobilized onto immobilized nickel resins via engineered histidine-tags. Based on the regulatory topology of the AD synthetic pathway, a rationally designed bi-modular system was implemented, which successfully improved the AD yield from 40% to ~100%. Furthermore, the multienzymes can be effectively reused for 7 cycles of reaction. Taken together, approximately 2.2g/L of AD was produced within 4 days, which was greater than 6-fold enhancement of AD specific yield as compared to the free enzymatic system.

Furthermore, the oxidation of AD to downstream artemisinic acid was explored with yeast whole-cell biotransformations. A hybrid *in vivo* and *in vitro* platform was demonstrated to produce the cytotoxic compounds, dihydroartemisinic acid and artemisinic acid and achieved ~80% conversion.

Finally, the non-mevalonate pathway was reconstituted and co-immobilized *in vitro* to produce 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MEC). The first committed step, 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) suffered from interfacial inactivation. By omitting Dxs and co-

immobilizing the other pathway enzymes, ~50% of substrate was converted to MEC within 10 minutes.

The findings in the thesis highlighted the advantages of cell free biosynthesis which are flexible, easily controlled and manipulated, and transcending the cellular barrier. The thesis study significantly contributed to optimizing and balancing the multienzymatic system *in vitro*, by first rapidly identifying the optimum enzymatic ratio via statistical experimental design, subsequently identifying the bottleneck step and its regulations, and lastly devise specific strategies to minimizer inhibitors' concentrations and de-bottleneck the multienzyme pathway. The strategies demonstrated are applicable to other *in vitro* multienzyme biosynthesis system beyond the scope of study. This bottom-up approach is re-emerging as a powerful, complementary method to cellular based biosynthesis.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Amorpha-4,11-diene
ADP	Adenosine-5'-diphosphate
ADS	Amorpha-4,11-diene synthase
ATP	Adenosine 5'-triphosphate
CDP-ME	4-diphosphocytidyl-2C-methyl D-erythritol
CDP-MEP	4-diphosphocytidyl-2C-methyl D-erythritol 2-Phosphate
CO2	Carbon dioxide
DMAPP	Dimethylallyl diphosphate
DMAPP	Dimethylallyl pyrophosphate
DXP	1-deoxy-D-xylulose 5-phosphate
dxr	1-deoxy-D-xylulose 5-phosphate reductase
dxs	1-deoxy-D-xylulose 5-phosphate synthase
E.coli	Escherichia coli
ERG12	Mevalonate kinase
Erg19	Diphosphomevalonate decarboxylase
ERG8	Phosphomevalonate kinase
FPP	Farnesyl pyrophosphate
FPP	Farnesyl pyrophosphate
GAP	Glyceraldehyde 3-phosphate
GC-MS	Gas chromatography mass spectrometry
GPP	Geranyl pyrophosphate
GRAS	Generally regarded as safe
gTME	Global transcription machinery engineering
HMBPP	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
idi	Isopentenyl diphosphate isomerase
IPP	Isopentenyl diphosphate
IPP	Isopentenyl pyrophosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ispA	Farnesyl pyrophosphate synthase

ispD	4-diphosphocytidyl-2C-methyl-D-erythritol synthase
ispE	4-diphosphocytidyl-2-C-methylerythritol kinase
ispF	2C-methyl-D-erythritol 2.4-cyclodiphosphate synthase
ispG	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate sythase
ispH	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase
LC-MS	Liquid chromatography mass spectrometry
MEC	2C-methyl-D-erythritol 2.4-cyclodiphosphate
MEP	2C-methyl-D-erythritol 4-phosphate
MVA	Mevalonic acid
NADPH	Nicotinamide adenine dinucleotide phosphate
PCR	Polymerization chain reaction
Pi	Phosphate
PMVA	Phosphomevalonic acid
Ppi	Pyrophosphate
PPMVA	Diphosphomevalonic acid
PYR	Pyruvate
TPP	Thiamine diphosphate

Chapter 1 Introduction

1.1 Motivation

1.1.1 In vivo metabolic engineering and its challenges

Isoprenoids are a diverse class of natural products that have many important biological functions. There are at least 55,000 different isoprenoids identified to date, but only a limited number of these has been shown to have biological functions [1]. This is because many of these compounds accumulate in trace amounts in their natural host, making it difficult to extract at sufficient amounts for further purifications and applications [2]. The quantum leap in the development of recombinant DNA technology has offered an opportunity to produce these valuable natural products in surrogate producers such as the robust microorganism *Escherichia coli* (*E. coli*) with high titers [3, 4]. Substantial experience in manipulating the multi-component biosynthetic pathway inside the microbes has been achieved in the last few decades. For example, Ajikumar and co-workers demonstrated multivariate-modular approach to balance the non-mevalonate pathway in vivo and successfully produce 1 g/L taxiadiene, an important precursor to the anti-cancer drug, Taxol [4]. Despite the recent successes in designing novel genetic circuits and in re-directing intracellular carbon flux for high yield production of metabolites, major challenge remains. The daunting complexity of cells, the unpredictable interference between native and synthetic parts, and the fact that cells adapt and evolve are some of the significant barriers yet to be resolved [5].

1.1.2 In vivo metabolic engineering and its challenges

One way to overcome these challenges in using cells as bioreactor is the exploitation of cell-free synthesis. *In vitro* cell-free synthesis plays a critical role and is re-emerging as a powerful platform to optimize and understand the biosynthetic pathways without the need to be concerned with cell viability [6, 7]. Therefore, it renders the freedom and flexibility to manipulate and adjust its components and abiotic environment at levels that may be intolerable to living organisms. Moreover, *in vitro* multienzymes synthesis ensures the regio- and stereo- selectivity of the products, and is able to direct metabolism to produce desired compounds at mild conditions [7, 8]. One successful application of *in vitro* multienzyme biosynthesis was the production of hydrogen (H_2) from the pentose phosphate pathway intermediates [9]. The central metabolism intermediates that would otherwise be efficiently utilized by microorganisms to produce biomass was directed to synthesize the user-defined product. As a result, the yield surpassed the theoretical yield by anaerobic fermentation from 4 H_2 per glucose (*in vivo* biosynthesis) to 12 H_2 per glucose (*in vitro* biosynthesis). Another well-cited application of *in vitro* multienzyme system was the production of the bacteriostatic agent, enterocin [6]. The cytotoxic compound was synthesized by assembling 12-enzymatic steps, and achieved 25% conversion. However, to successfully assemble a network of multiple biocatalysts to synthesize the products at the optimal productivity requires fundamental understanding of the system and extensive optimization. For example, even when each step of the enzymatic reaction can achieve a yield of 90% conversion, the

overall yield would be equal or less than 25% for a pathway consisting of 12 enzymatic steps. To balance the enzymatic fluxes in the multienzyme system is paramount to minimize the accumulation of inhibitory intermediates and byproducts, improve the life span of enzymes and enhance the productivity of the system. To our best knowledge, very few studies on systematic optimization of multienzymatic reaction was demonstrated in the literature before the project was initiated. Therefore, this thesis study aims to demonstrate strategies to systematically optimize and balance the isoprenoid, amorpha-4,11-diene (AD), biosynthetic pathway *in vitro*. Novel industrial optimization tools and system-level modelling approach were explored to improve the AD titer to near theoretical yield. The insights gained and strategies employed will be valuable and applicable to other *in vitro* multienzymes biosynthesis beyond the scope of the study.

1.2 Thesis Objectives

The main objective of the thesis was to assemble and optimize multienzymes reactions *in vitro* so as to produce isoprenoids and isoprenoid precursors. The important mevalonate pathway was used as an exemplary pathway to be reconstituted and optimized to produce amopha-4,11-diene (AD), an important precursor to the antimalarial drug artemisinin. Mathematical modelling and predictions were extensively utilized to aid in addressing three important research questions: (1) is there an optimal enzymatic compositions present in this multienzymatic biosynthesis system? (2) Are there any regulations

present among the metabolites and the enzymes? (3) How to improve the kinetics and specific yield of the system after understanding its fundamental behaviours? They are elaborated specifically as follows.

(1) We aim to identify the optimum enzymatic ratio of the seven enzymatic steps, in order to balance the metabolic flux, and understand how each enzymatic activity would affect the final yield of AD. Moreover, we hope to identify the key rate-limiting step(s) that its concentrations need to be fine-tuned. In the field of *cell-free* multienzymatic biosynthesis, the concept of optimizing and balancing multienzymatic activities were often under represented as compared to designing unnatural biochemical pathways and producing the final product. The conventional method to adjust the enzymatic concentrations was by trial-and-error method, which was iterative, time-consuming and sometimes failed to capture the true optimum condition. Therefore, the purpose of the first study (Chapter 3) was to demonstrate the important concept of balancing multienzymatic flux *in vitro* by combinatorial approach and rapidly identify the optimum enzymatic levels.

(2) Following the pathway balancing *in vitro*, we was able to achieve ~100% conversion of AD from 5mM mevalonic acid within 12 hours of incubation. However, theoretical calculations based on the turnover rate of the key rate-limiting step, amorphadiene synthase (Ads), revealed that to convert 5mM mevalonic acid required only 3-4 hours.

Therefore, we hypothesized that the critical step Ads was inhibited in the course of reaction. To unravel the limiting factor, we aim to study the regulatory network topology present in the multienzyme system. Single enzyme analysis is advantageous to unravel the molecular interactions between the inhibitor and the enzyme. However, this approach entails too many factors and experimental runs and might not accurately depict the multienzymatic system behavior. Therefore, we pursued a system approach to rapidly quantify an important parameter in metabolic control analysis: elasticity coefficient, which reflects the biological interactions among the metabolites and the pathway enzymes, to some extent. The insights we learnt would guide us for further optimization.

(3) Based on in-depth understanding of the multienzyme pathway system, we aim to develop specific strategies to shorten the reaction cycle and improve the specific AD yield (mg AD per mg of enzyme). Immobilization strategies were also explored in order to save the costly purification process and recycle the multienzymatic system. As a result, it would render the system industrially favourable.

(4) Extending from the AD synthesized pathway, we further aim to explore semi-*in vitro* multienzymatic biosynthesis, which potentially could be scaled-up and used for industrial applications. Since in vitro multienzymatic reaction is flexible and transcends cellular barrier, we seek to mix different in vitro reaction formats together in single vessel

to further convert AD to downstream oxidized artemisinin precursors. The membrane-bound cytochrome enzyme, CYP71AV1, was perturbing the upstream AD production *in vivo*. Therefore, a novel and integrated *in vivo* and *in vitro* hybrid reaction system was proposed and aimed to surpass the conversion yield (~60%) achieved in the literature.

(5) The non-mevalonate pathway is not only valuable for the production of the key building blocks of isoprenoids but also intensively studied for novel drug discovery. Therefore, we aim to isolate the pathway enzymes *in vitro*, reconstitute them in a single vessel, and develop a recyclable platform to study and screen for novel antibiotic drugs in a multienzymatic fashion.

1.3 Thesis Organization

This thesis consists of eight Chapters. Chapter 1 provided a brief introduction and the objectives of the thesis, while Chapter 2 reviewed the literature related to the thesis topics. In Chapter 3, statistical experimental design methodology was employed to optimize the enzymatic ratio of the pathway. It successfully predicted an inhibitory enzymatic step, farnesyl pyrophosphate synthase (ispA) by which its product farnesyl pyrophosphate (FPP) would precipitate in the reaction. Moreover, amorphadiene synthase (Ads) was shown to be a critical enzymatic step in order to achieve almost 100% conversion of substrate to the product. Chapter 4 described the comprehensive investigation of

the components in the *in vitro* multienzyme system to identify interactions among the metabolites and the enzymes. Lin-log approximation was used and predicted a novel inhibitory effect of ATP on amorphadiene synthase (Ads). It is due to the polyphosphate moiety of ATP that resembles the by-product of Ads, pyrophosphate, which was found to be a novel, potent inhibitor of Ads. An ATP recycling system and pyrophosphatase were used to maintain a low concentration of ATP and the removal of pyrophosphate, respectively. The strategies successfully improved the rate of AD production to near theoretical productivity. The same strategy was used to devise a bi-modular co-immobilized multienzyme system that could effectively reuse the enzymes for 7 cycles of reactions, (Chapter 5). Chapter 6 extended the study on the pathway so as to convert amorphadiene to artemisinic acid (AA) and/or dihydroartemisinic acid (DHAA). A whole cell yeast biocatalysis was also explored, which by mixing different formats of in vitro system (whole-cell, cell lysates), approximately 80% of AD was converted to the downstream products (AA and DHAA). Chapter 7 reported the assembly of the non-mevalonate pathway on heterologous surface to produce an important intermediary metabolite 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MEC). Chapter 8 summarized the important findings of the thesis and provided some recommendations for future work.

Chapter 2 Literature Review

2.1 Metabolic engineering of isoprenoids

Natural products represent a rich source of therapeutic agents to defend against human disease. Despite the advancement in synthetic combinatorial chemistry, natural products and their mimics are still the dominant forms used in drug development and disease treatment. Of 155 small molecules used for cancer treatment, 47% of them were derived from natural products [10]. Their remarkable chemical diversity has motived intense scientific interest in mining natural products and identifying their novel functions [1]. Among them, terpenoids is the largest and most diverse class of natural products that possess many important biological functions [11]. There are more than 55,000 terpenoids identified so far, and many of them have functions that are yet to be explored [1]. One barrier that curbed the wide utility of terpenoids is their extreme scarcity in their natural producing host; some of them were only present in parts per million (ppm) level. Therefore, to obtain sufficient amount of a specific kind of terpenoids will require massive harvesting and extensive extraction. This space- and time- consuming process inevitably limit the supply of high-therapeutic-value terpenoids for downstream analysis, and is definitely not sustainable to supply global demands. *De novo* chemical synthesis of many of these terpenoids are not industrially viable due simply to the complex chirality of the chemical structures and chemical synthesis often result in undesirable side product contaminants. Thus, cells as surrogate producers that redirect the metabolite flux towards terpenoids production would provide an innovative means to circumvent these

challenges. Fortunately, despite the structural diversity of terpenoids, they share common building blocks- the 5-carbon molecules, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). The two precursors are condensed to form various lengths of hydrocarbon scaffold. According to the number of carbons in the skeletal structure, typically in units of five carbons, the terpenoids can be classified into hemi- (5 carbon), mono- (10 carbon), sesqui- (15 carbon), di- (20 carbon), and triterpenes (30 carbon) [12, 13]. Over the past decade, metabolic engineering and synthetic biology, using microbial hosts to enhance IPP / DMAPP production, have significantly improved the yield and numerous useful tools to better control the microbial production were developed [14]. Ajikumar *et al.* employed a multivariate-modular approach to metabolic-pathway engineering and successfully increased the titer of taxadiene—precursor to taxol—to 1 g/L in *E. coli* strain [4]. Moreover, Paddon *et al*. combined the genetic engineering and fermentation strategy to create a high-producing yeast strain and achieved 25 g/L artemisinic acid production [3]. These emerging technologies using microbial fermentation (Figure 2.1) is replacing conventional, less environmentally friendly chemical processes, and are becoming industrially viable [15].

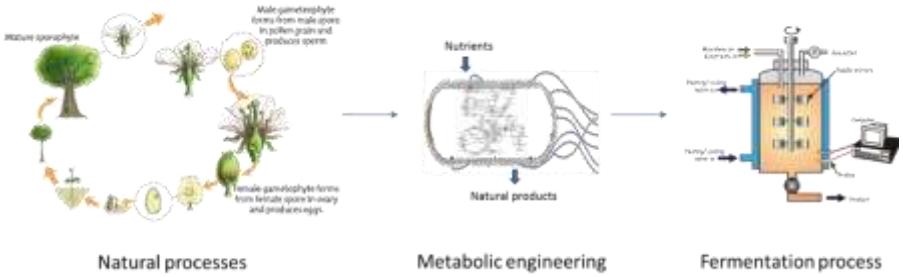

Figure 2.1. From natural processes to fermentation based production of natural products.

In the field of metabolic engineering, pathway balancing is paramount to ensure high yield, since it is well accepted that metabolic pathways are not limited by a single rate-limiting step and that optimized pathways require the balanced expression of several enzymes [16]. Accumulation of pathway intermediates may result in compromised productivity and lead to growth retardation [13]. Therefore, numerous innovative combinatorial and rational design strategies were devised to control and balance the pathway flux, such as using global transcriptional machinery engineering (gTME) [17] and co-localization of pathway enzymes [18]. However, biology is complex, and our ability to rationally redesign biosynthetic pathways is often limited by our understandings of the interplay among biomolecular networks [19]. Moreover, the enzymatic levels are hard to control when overexpressed in heterologous hosts as these proteins may be insoluble when overexpressed and attempt to predict the solubility of the proteins is still a significant challenge [20]. Due to the complexity with *in vivo* based production, *in vitro* multienzyme synthesis (Figure 2.2) has emerged as powerful alternatives to cellular based metabolic engineering [21]. Recently, Cheng *et al.* had successfully reconstituted the polyketide pathway *in vitro*, achieving an approximately 25% overall yield, which *in vivo* production had limited success [6, 21]. Walsh and co-workers have produced terrequinone A, an antitumor fungal agent, in three enzymatic steps, and revealed a novel function of TbiE, that was not identified *in vivo* [21, 22]. Therefore, *in vitro* multi-enzyme synthesis not only provides an alternative route for natural products synthesis, but also provides

additional insights into identifying regulatory pathways that may assist in both *in vivo* and *in vitro* biosynthesis.

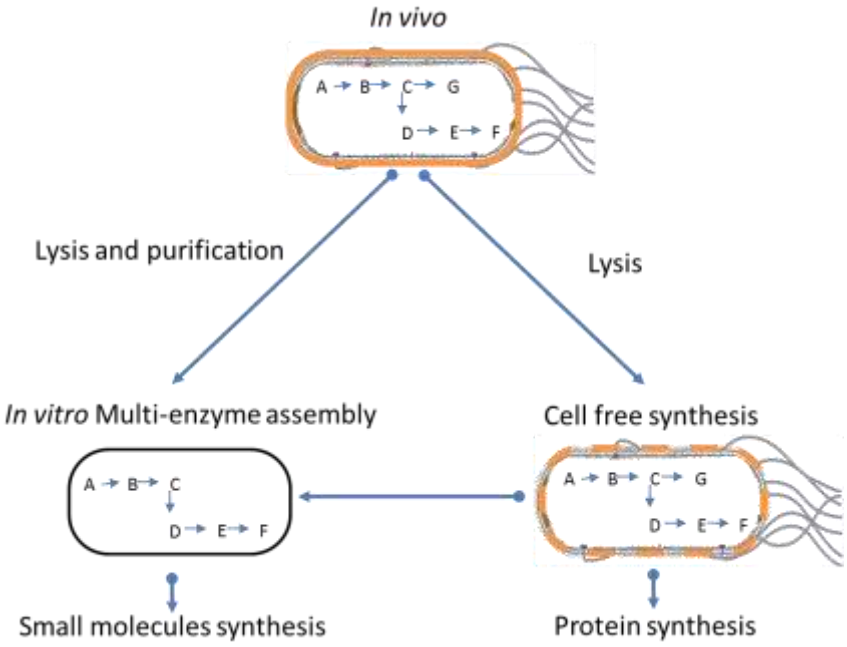

Figure 2.2. Different platforms for biosynthesis [23].
Majority of metabolic engineering and synthetic biology projects are performed in vivo. In vitro systems are emerging as a complementary technology. In vitro systems can be further subdivided into cell free synthesis and multi-enzyme assembly. The former was obtained by lysis of cells and mainly used for biomolecules production, whereas multienzyme assembly is obtained with lysis and purification, and mainly used for small molecules production. (Reproduced with permission from Elsevier.)

2.2 Cell-free and multienzyme biosynthesis in vitro

The power of *in vitro* biotransformation was first appreciated over a hundred years ago. In 1897, Eduard Buchner used yeast extract to convert sugar to ethanol and carbon dioxide for which he won the Nobel Prize (1907 Chemistry) [23]. Since then, cell-free synthesis has gained exciting advancement both in academic research and commercial applications. Indeed, successful *in vitro* biotransformation have been achieved for several decades, due largely to rapid developments of genomic tools and high-throughput engineering techniques [24,

25]. Majority of *in vitro* biocatalytic transformation has focused on single transformations. One classic example is the polymerase chain reaction, developed by Nobel Prize winner Kary Mullis in 1983, that involved only one heat-stable DNA polymerase from *Thermus aquaticus*. It has become an indispensable technology in molecular biology and have revolutionized modern biological research [26]. Another noteworthy application of single-enzyme is their ability to modify the backbones of natural products [13, 27, 28]. Britstol-Myers Squibb Pharmaceutical Research Institute has developed an enzymatic process to add an acetyl group to C-10 position hydroxyl group of 10-deacetylbaccatin III, and convert it to baccatin III (Figure 2.3), an important precursor to taxol, with 51% yield [29]. Due to the specific action of enzyme, protection of other free hydroxyl groups is unnecessary (Figure 2.3). With protein engineering and directed evolution, enzymes are now able to accept a broader spectrum of substrates, thus expanding its capability in serving as novel catalysis to diversify the structures of natural products further [30, 31].

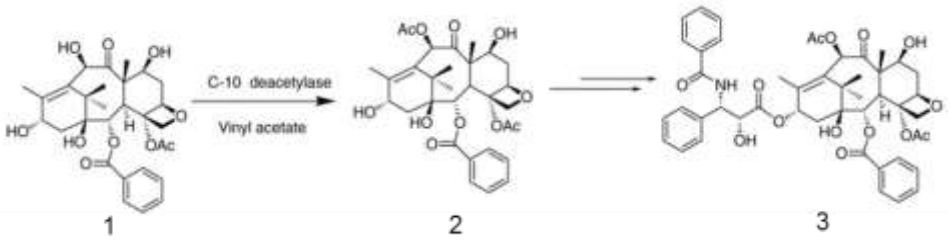

Figure 2.3. Schematic representation of conversion of 10-deacetylbaccatin III (1) to baccatin III (1) by C-10 deacetylase, which is an important precursor to taxol (3) [29].
(Reproduced with permission from Elsevier.)

On the other hand, a vast number of cascade enzyme reactions has been developed in the past decades, ranging from the simple combination of oxidoreductase with a suitable cofactor regeneration system to highly complex multienzymatic synthetic systems [32]. Importantly, many enzymes are fairly compatible with each other within certain ranges of operating conditions, as they are able to co-exist inside living organisms [32]. For example, an alternative route with alcohol dehydrogenase and NAD(P)H cofactor recycling enzyme has been used to enrich alcohol enantiomers instead of lipase-catalyzed method [32]. Monti and co-workers elegantly applied the co-factor specificities rule of certain oxidoreductase, and combined 5 enzymatic oxidation and reduction reactions in a single vessel for the synthesis of 12-ketoursodeoxycholic acid from cholic acid (Figure 2.4); this was not possible by conventional chemistry synthesis [33].

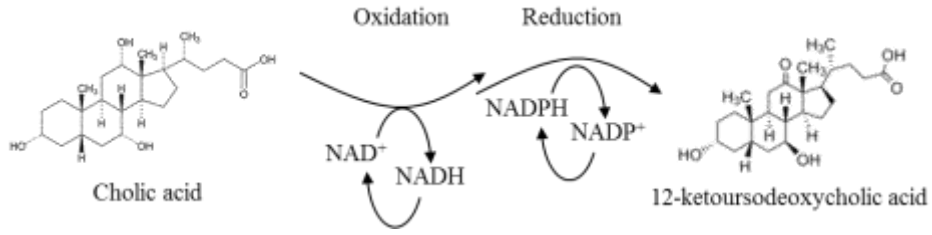

Figure 2.4. Multienzyme synthesis of 12-ketoursodeoxycholic acid from cholic acid using sequential oxidation and reduction [33].
(Permission to o reproduce the Wiley Materials)

This thesis is focused mainly on multienzymatic cascade reactions. A common workflow for multienzyme reaction is shown in Figure 2.5 that comprises of the overexpression of each enzyme in a microbial host cell, purifying the overexpressed enzymes from host cells, and reassembling them in a reaction vessel for biocatalysis.

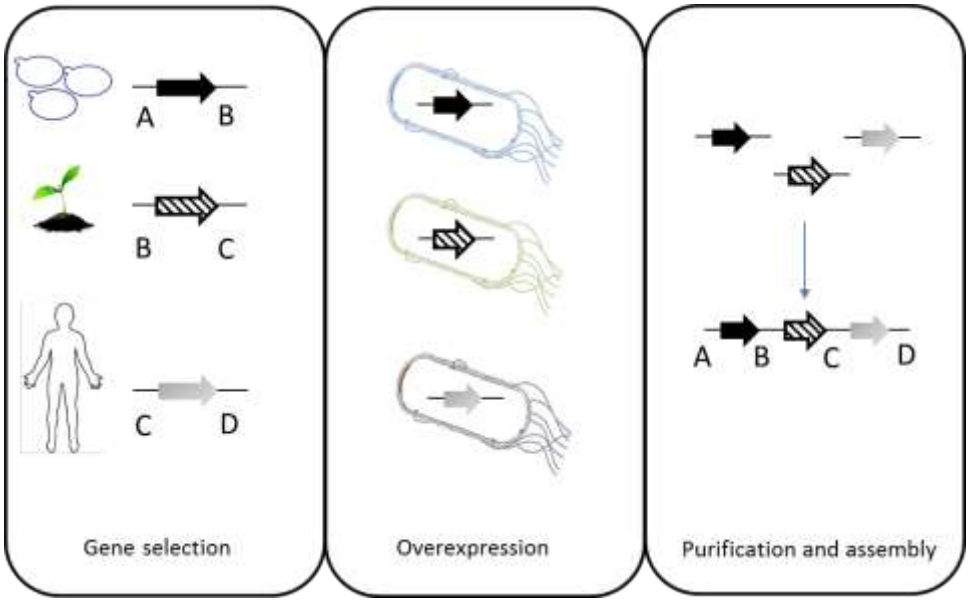

Figure 2.5. The common workflow of multienzyme reaction.

2.2.1 Advantages of in vitro *multienzyme synthesis*

As part of their inherent property, enzymatic catalysis is regio- and stereo-specific, and operate under mild reaction conditions. Many natural products contain several chiral centers, hence, resulting in low yield when synthesized through traditional chemical synthetic process [34]. For example, the total chemical synthesis of taxol resulted in a yield of ~0.4% [4]. Therefore, biocatalysts are increasingly favoured by industries to synthesize and modify complex target molecules [35-37]. In addition, the use of multienzymes in one-pot fashion is particular appealing since it saves time from the laborious process of isolating the intermediates and hence effectively preventing the loss of yield during purification [32]. Sometimes, the auxiliary reaction, such as co-factor

regeneration, would further enhance the yield by providing thermodynamic drive towards product formation and/or irreversibly removing by-products [38].

In contrast to *in vivo* production, *in vitro* multienzyme based technology successfully bypassed cellular barriers [23]. Thus, resources could be entirely channeled towards synthesis of user-defined products, and the substrates could be added freely without a need to consider toxicity. The main advantage of *in vitro* multienzymatic systems as compared to engineered cellular systems is their much reduced complexity and therefore ease of controlling the reactions [39]. The process could be easily optimized by varying enzyme and substrate concentrations, addition of co-substrates and solvents, and by varying pH and temperature. Due to the simplified network, mathematical modeling of *in vitro* multienzyme reaction is made easier; the recent modeling methods will be reviewed in section 2.3.2. Lastly, for *in vitro* multienzyme reaction, the purity of the final products is much higher as compared to cellular systems where competing side reactions and cell metabolites may result in a mixture of contaminating compounds [39]. A comparison between *in vivo* and *in vitro* multienzyme reactions is summarized in Table 2.1.

Table 2.1. Comparison of multienzyme in vivo and in vitro process [40]. (Reprint with Permission)

Characteristics	*In vivo* process	*In vitro* process
Cell/Biocatalysis Constrains		
Substrate inhibition	Possible	Possible
Product inhibition	Possible	Possible
Catalytic stability	Low	High (if immobilized)
Cost production	Low	High
Reaction Constrains		
Reaction reproducibility	Variable	Reproducible
By products	Possible	Unlikely
Operating conditions	High dependence	High dependence
Process Modeling		
Process understanding	Mechanism not fully understood	Possible
Reaction structure	Complex metabolic networks	Simpler reactions
Mathematical model interpretation	Difficult	Possible
Potential Process Controllability		
Regulatory control (T, pH, DO)	Possible	Possible
Supervisory control (concentrations)	Difficult (for intermediate products)	Possible
Process Monitoring		
Online measurements	Possible	Possible
Intermediate products	Unlikely	Possible
Downstream Processes		
Product recovery	Possible	Possible
Recycling (cell / biocatalysis)	Possible	Possible (immobilized)
Others		
Green / renewable process	Yes	Yes
Current research activity	High	Low

2.2.2 Applications of in vitro multi-enzyme pathway assembly

There are three major applications in the field of *in vitro* multi-enzyme reaction: directed synthesis of user-defined products (Section 2.2.2.1), screening for inhibitors (section 2.2.2.2In vitro multienzyme pathway assembly for drug screening), and biochemical analysis of pathway (section2.2.2.3).

2.2.2.1 Directed synthesis of user-defined products

In vitro multienzymatic synthesis is versatile enough to stop at any step along the enzymatic pathway. Pathway intermediates are important to study the biochemical properties of the downstream enzymes. Most of the time, isotope-labelled intermediates served as a good tracker to identify downstream pathways. As compared to cell based biosynthesis, multienzyme synthesis would be more efficient to obtain such intermediates at high quantity. Schuhr *et al*. assembled the first five 1-deoxy-D-xylulose 5 phosphate (DXP) pathway enzymes together with co-factor regeneration systems in one pot (Figure 2.6A), and produced a variety of ^{13}C- or ^{14}C- labeled 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MEC) with a 80% conversion yield from pyruvate [41]. They emphasized the importance of multienzyme reaction that enabled them to introduce isotope labels at specific positions with the same procedure as well as the high efficacy of the process that allow them to scale down the reaction so that the products were not diluted. The isotope-labeled MEC is important to elucidate the mechanism of downstream ispG and ispH enzymes, which have been recently identified as bottleneck steps in the DXP pathway [42, 43].

In addition, *in vitro* enzyme synthesis is highly tractable, allowing the assembly of enzymes from different metabolic pathways and creation of novel pathways, as shown in Figure 2.6B. Zhang *et al*. has demonstrated an unnatural synthetic enzymatic pathway to produce hydrogen from economical starting material [9]. They assembled 13 enzymes, consisting of enzymes from pentose phosphate pathway, and successfully produced hydrogen at a much higher theoretical yield than would be achieved *in vivo* (Figure 2.6B). The entropy gained when hydrogen was produced in the gas phase resulted in a negative thermodynamic drive and hence making the unnatural process spontaneous. This would be challenging for *in vivo* production since the substrates used such as glucose-6-phosphate would be easily diluted inside a living cell. There are many more examples demonstrating the capability of *in vitro* multienzyme synthesis which surpasses nature [6, 44-47].

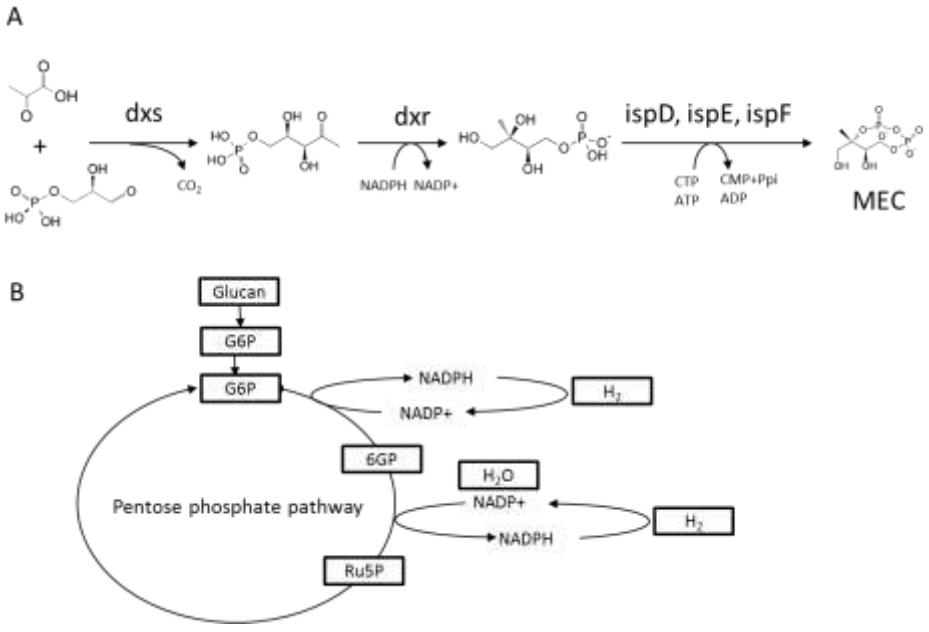

Figure 2.6. Examples of in vitro multienzyme biosynthesis.
(A) synthesis of 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MEC) by the 1-deoxy-D-xylulose 5 phosphate (DXP) pathway. (B) The unnatural synthetic pathway to produce hydrogen

2.2.2.2 In vitro multienzyme pathway assembly for drug screening

Microorganism have unique and essential pathways that are absent in eukaryotic cells. Many of them serve as valuable targets to develop anti-bacterial and anti-infective agents. Moreover, due to increasing cases of drug resistance in various pathogenic strains, drugs targeting multiple enzymatic sites would be preferred to elicit more potent anti-bacterial effect. Therefore, *in vitro* multienzymes assembly provides a focused platform to screen for drugs inhibiting multiple pathway enzymes simultaneously. Potential drug leads can be added at will without concerning its transport across the cellular membrane. Peptidoglycan, for example, is a unique structure in bacteria and is not found in mammalian system. Its biosynthetic pathway serve as attractive drug targets. However, the lack of pathway intermediates severely hampered the screening process [48]. Therefore, El Zoeiby and co-workers reconstructed the peptidoglycan biosynthetic pathway of 6 Mur enzymes, for the purpose of developing inhibitors to the pathway [49]. The shikimate pathway is an important pathway found in plants, fungi and microorganisms, but not in animals, for the synthesis of ubiquinone and folate [50]. Deletion of the pathway is lethal to the host cell, so its constituent enzymes are valuable drug targets. By means of *in vitro* assembling the purified pathway enzymes, a promising high-throughput assay was developed to screen for inhibitors to any of the enzymes simultaneously [51].

2.2.2.3 Understanding of biochemical properties of the pathway

In a much simplified setting, *in vitro* multienzyme synthesis is critical to understand the fundamental biochemical properties of the metabolic pathway, such as the steady-state kinetic behavior as well as the interplay among enzymes and metabolites. Yu and colleagues have reconstituted the *E. coli* fatty acid synthase (FAS) using 10 purified protein components, and did a detailed kinetic analysis of the *in vitro* system [52]. By doing so, they were able to deduce an optimal molar ratio of the 10 protein components based on the influence of individual FAS subunit concentrations on FAS activity. One interesting observation made by them was the unusual dependence of FAS activity on some but not all of the subunits. Another example is the elucidation of polyketide pathway. Sun *et al.* has been able to identify the biosynthetic pathway to produce tetronate RK-682, which is a potent inhibitor of protein phosphatases and of HIV-1 proteinase [53]. Their work found an unusual enzyme RkD which played a central role that is responsible for the synthesis of RK-682.

2.2.3 Challenges of in vitro *synthesis*

Despite the many advantages, there are also a number of limitations of *in vitro* systems. A major limitation for *in vitro* reaction is the need for enzyme catalysts that must be obtained in sufficient quantities for refolding and reconstitution, functional and stable under *in vitro* reaction conditions [39]. To tackle the problem, different ways of enzyme synthesis has been demonstrated; one promising technology is *in vitro* translation system where 32 purified

enzymes from cellular translation machinery are known to be dedicated entirely to producing target enzyme at a rate of 160 µg/ml/h [54]. Innovative reaction formats have been devised and implemented to enhance the performance of biocatalysts. This is reviewed in section 2.4. Moreover, co-factor dependent enzyme reactions involving dehydrogenases, oxidases, kinases, and phosphatases require an enzymatic co-factor regeneration system that continuously replenishes expensive co-factors by catalyzing the opposite reaction using an inexpensive substrate [39]. Majority of the enzymatic steps are reversible in nature, which would limit the conversion yield of a multienzyme reaction. With co-factor recycling system, the reaction could be further driven to completion [38, 55].

2.3 Mathematical tools aided in vitro multienzyme process optimization

To engineer a biological system, the application of mathematical tools is invaluable in the design process. One of the biggest fine chemicals producer, LONZA, routinely applies modeling to avoid formation of inhibiting by-products of biotransformation and also as tools for industrial bioprocess integration [56]. Therefore, mathematical simulations are increasingly applied industrially to devise the most cost-effective process and maintain control over the process.

Two kinds of mathematical tools are important here for *in vitro* multienzyme process optimization: (1) statistical experimental design methodology; (2) kinetic and dynamic process modeling. They have been widely used independently as well as in combinations to identify the optimum process conditions in order to achieve the user-defined objectives [57-59].

2.3.1 Statistical experimental design methodology for process optimization

The conventional method of process optimization involves the study of one variable at a time, which is not only cost and labour intensive but also ignores the interactions between the confounding factors [60]. Statistical modeling and experimental design methods overcome the limitations and have been extensively applied for standard industrial process optimization. It is a "black-box" modeling strategy that views experiment as simply connecting inputs (factors) and output (responses), and predicts the best relationship between the inputs and outputs [61]. Generally, in the design of a statistically based experiment, it involves several steps: (1) selection of responses; (2) identification of the confounding factors; and (3) choice of the different levels or treatments [62]. Usually, partial factorial design is carried out to reduce the experimental efforts.

One important design methodology was formulated by Dr Genichi Taguchi, and is now known as, "Taguchi Methods" [62]. They were first introduced as a means to design robust products and processes, and is now widely used in many interdisciplinary areas including biotechnology. Taguchi methods use orthogonal arrays to minimize and randomize the experimental runs, and ensure all the levels of a factor appear at equal number of times [62]. Moreover, the methods have flexible structures that can be applied for two-, three- and mixed- level fractional factorial designs, and accurately acquire highly reliable technical information [62]. For example, Taguchi method has been successfully applied to fermentation processes that produce recombinant proteins as well as metabolites [63-66]. Sirisansaneeyak and co-workers have used an L_{18} orthogonal

array, one factor at two levels and seven factors at three levels, to increase the lactic acid production by 7 fold, and identified the main factor contributing to the yield increment was the yeast extract levels [66].

The strategy adopted by Taguchi method is vastly different from classical design of experiments (DOE); it involves the empirical minimization of an expected loss function over an uncontrollable "noise space" to determine the best design of the controllable factors / variables [67]. The uncontrollable noise space could include all the external conditions (such as human operator skill levels) that deviate the outcome from the target values [67]. Target values are specified by the experimentalists, and are fixed in the following three types: (1) the closer to some nominal value the better; (2) the bigger the better; (3) the smaller the better. Normally, to improve the production yield, the second target, 'the bigger the better", is applied. The statistical model would then be able to determine the main contributing factor(s), and generalize how individual controllable factor would affect the yield by average effect analysis. Analysis of variance (ANOVA) will be conducted to access the significance of main controllable factors on the yield. Prediction of optimal combinations of the factors would be calculated to guide further designs and validations.

Recently, Taguchi method has been compared with response surface methodology (RSM), and artificial neural network (ANN). Aggarwal and co-workers have noted that RSM requires much more experimental efforts than Taguchi methods, while they predicted nearly similar results [68]. However, in certain applications, Taguchi methods would lead to non-optimal solutions, since

the central objective was to reduce variations. Therefore, Taguchi method and RSM were often used in sequence to reduce the experimental efforts and identify the optimum condition. Teng and co-workers [69] adopted the strategy to optimize the culture condition for whole-cell lipase production process. They used L_{18} Taguchi method to narrow down the 8 controllable factors into 4 main factors, and further optimized the concentrations of the four key factors by RSM. The process effectively reduced the experimental efforts from 80 to 35 experimental runs, and improved the yield by 2.2 fold [69].

2.3.2 Kinetic and dynamic modeling for network description

Understanding the underlying mechanisms of a multienzyme network would definitely offer distinctive advantages for better process design and optimization. Mathematical modeling is the art of describing the essentials of a system in mathematical terms [70]. It digitized a complex biological system into a simplified analogue that is easier to analyze, interrogate, predict, manipulate and optimize than the biological system itself [71]. Despite the long history of modeling, describing *in vivo* system in models has achieved limited success due to the complex multilevel nature of the system. Instead, *in vitro* multienzyme system is much more simplified with known components, and hence, a mathematical model can be composed. A spectrum of models is available ranging from mechanism driven to data intensive models. Despite the multitude of modeling methods, they typically encompass nine phases to construct such models: (1) data selection; (2) collection of information on network structure and

regulation; (3) specification of assumptions and simplifications; (4) selection of a mathematical modeling framework; (5) estimation of parameter values; (6) model diagnostics; (7) model validations; (8) model refinements; and (9) model application [71].

2.3.2.1 Mechanism-based modeling

One very important mechanism in biochemical analysis is the Michaelis-Menten Rate Law (MML) (Equation 2.1). which assumes that the enzyme and substrates interact transiently, and hence a steady state would be achieved [71]. Therefore, defined rate equations could be constructed if the kinetics parameters such as K_m and V_{max} are known. Moreover, the equation is important to study the allosteric effects of enzymes by determining the change in K_m and V_{max} when an inhibitor is present. Ideally, an *in vitro* multienzyme system could be represented by a set of ordinary differential rate equations, governed by MML, to simulate the dynamic behavior of the system. However, there are certain limitations that make the modeling method less practical. Firstly, many of the kinetics parameters (K_m and V_{max}) have to be determined experimentally. Although they are "constants" for individual enzyme, they vary across species and different microenvironment [72]. This variation is important to generate diversity in populations and behavior adaptations. Secondly, the validity of the assumption may be challenged when substrate concentration become very low, especially when the networks involve protein-protein interactions [73] and multienzymatic complex formation. Lastly, the MML equations become very complex and difficult to analyze when there are

multiple substrates and mechanisms involved; many enzymes would interact with more than one substrate, and their mechanisms are not fully unraveled [74].

$$v = \frac{V_{max} S}{K_m + S}$$

Equation 2.1. The Michaelis Menten Equation.
The symbols are as follows: v is the rate of enzymatic reaction. S: the substrate concentration. V_{max} is the maximum enzymatic activity. K_m: is the Michaelis Menten constant.

In view of the shortcomings of MML, theoretical biologists have made effort to modify the equation. Equation 2.2 described the extended version of MML, which would be able to capture reversible one-substrate-to-one-product reaction mechanism. Moreover, Hill equation [75] attempts to describe the cooperative and allosteric interactions between substrate and enzyme. This stems from the observation that *in vivo* substrate concentration did not vary drastically but could elicit drastic change in enzymatic activity [76].

$$v = \frac{V_f\left(\frac{C_S}{K_{m,S}}\right) - V_r\left(\frac{C_P}{K_{m,P}}\right)}{1 + \left(\frac{C_S}{K_{m,S}}\right) + \left(\frac{C_P}{K_{m,P}}\right)}$$

Equation 2.2. The extended Michaelis Menten equation.
The symbols are as follows: v is the rate of enzymatic reaction. V_f and V_r are the forward and reverse rate of reaction respectively. C_s and C_p are the concentration of the substrate and product respectively. $K_{m,S}$ and $K_{m,P}$ are the Michaelis Menten constant for substrate and product respectively.

To capture the multi-substrate / product characteristics of enzymes, more sophisticated rate equations have been developed, such as the Monod-Wyman-Changeux (MWC) model [77]. However, mechanism driven model requires *a priori* knowledge of the biochemical pathway structure. Therefore, quite limited applications have been demonstrated for complex *in vitro* multienzymatic reaction

system. Ku and co-workers have attempted to model a C_{30} carotenoid pathway comprising farnesyl pyrophosphate synthase (ispA) and 4,4'-diapophytoene (DAP) synthase (crtM) [78]. Detailed kinetics for individual enzyme has been carried out under optimized reaction conditions. They were then combined with mass balance equations to model the two-step *in vitro* reaction. With the aid of modeling, Ku *et al*. was able to identify the product DAP is unstable in aqueous medium. Hence, they introduced a novel two-phase reaction system, and achieved a quantitative conversion from isopentenyl pyrophosphate (IPP) to DAP.

2.3.2.2 Canonical modeling: lin-log approximation

An alternative approach to modeling dynamic behavior of multienzyme systems is canonical modeling method; in particular the lin-log approximation approach proposed by Visser and Heijnen which have gained importance and attention [79]. It is an extension of Metabolic Control Analysis (MCA), but allows larger perturbations of enzyme and metabolite concentrations around the steady state [79]. Hence, it is technically feasible to collect the experimental data. Moreover, Lin-log approximation is able to capture the nonlinear dynamics of the enzymatic reactions while keeping the model relatively simple [71]; the reaction rate (Equation 2.3) is a nonlinear function of metabolite concentrations and is proportional to enzyme levels [80]. The detailed mathematical descriptions will be presented in chapter 4. The objective of lin-log approximation was to estimate the elasticity, an important parameter in MCA that describes the effect of a metabolite on enzymatic activity around the pre-determined steady state.

Therefore, elasticity would shed light on the allosteric interactions among the metabolites and enzymes; negative elasticity would normally represent an inhibitory effect, while positive elasticity correspond to a stimulatory effect [79, 80]. Wu *et al*. demonstrated the use of lin-log modeling method to understand a linear reconstituted pathway for pyruvate production from glycerate-3-phosphate (part of glycolysis pathway) , and were able to estimate elasticities that were consistent with biological knowledge [80]. Costa *et al*. did a comprehensive comparison among the available models by using *E. coli* central metabolism as a benchmark and concluded that Lin-log approximation would be the most suitable approach to model complex large-scale networks [81]. Therefore, lin-log approximation is a powerful method in structure identification and multienzyme network topology elucidation; it provides an initial screen of any possible regulations present in an *in vitro* multienzyme system, and hence narrowing down the search space.

$$\frac{v_i}{v_i^o} = \frac{e_i}{e_i^o}\left(1 + \sum_m \varepsilon^0_{i,x_m} ln\frac{x_m}{x_m^o}\right), \qquad \varepsilon^0_{i,x_m} = \frac{x^o}{v_i^o}\frac{dv_i^o}{dx^o}$$

Equation 2.3. The lin-log approximation equation.
The symbols are as follows: v and v^o are reaction flux and steady state flux respectively. e and e^o represent the enzyme concentrations and steady state enzyme concentrations respectively. ε^0 is the elasticity around the steady state which represent the relative change in the reaction rate due to a relative change in metabolite level [79]; the mathematical representation is given, too. x is the metabolite concentrations. i and m represent the number of reactions and the number of metabolites in the multienzyme reaction system respectively.

The drawback of lin-log modeling is that substantial error would be resulted if the substrate concentrations are very close to 0, due to the inherent property of logarithm operations that log(x) approaches negative infinity when x value approaches 0 [82]. Hybrid models, combining both MML rate equations

and canonical modeling method, might be a better strategy for modeling in biological systems [83].

2.4 Formats used in in vitro multienzyme reaction

Another exciting aspect of engineering *in vitro* multienzyme reaction is the creative reaction formats developed to further enhance the performance of the system. The interdisciplinary applications of material research and synthetic biology have overcome some limitations of *in vitro* multienzyme system (section 2.2.3Challenges of *in vitro* synthesis) and significantly improved the biotransformation efficiencies [39]. For example, the industrially important enzyme lipases, when immobilized on hydrophobic solid support, showed interfacial activation [84].

2.4.1 Co-immobilization of purified multienzyme system

Enzyme concentration is much more diluted *in vitro* as compared to the densely packed environment *in vivo*. Diffusion of substrates into the catalytic sites is a significant limitation to reaction kinetics. Increasing enzyme / substrate concentration effectively increased the proximity of the catalytic sites and hence enhanced the productivity of the multienzyme reaction. Nature has its ways of bringing the enzymes closer by creating multienzyme complexes and synthetic biology has exploited such strategies recently. There are many ways of bringing enzymes into close proximity such as chemical-crosslinking, genetic fusion and

co-immobilization. Here three particularly interesting strategies would be reviewed.

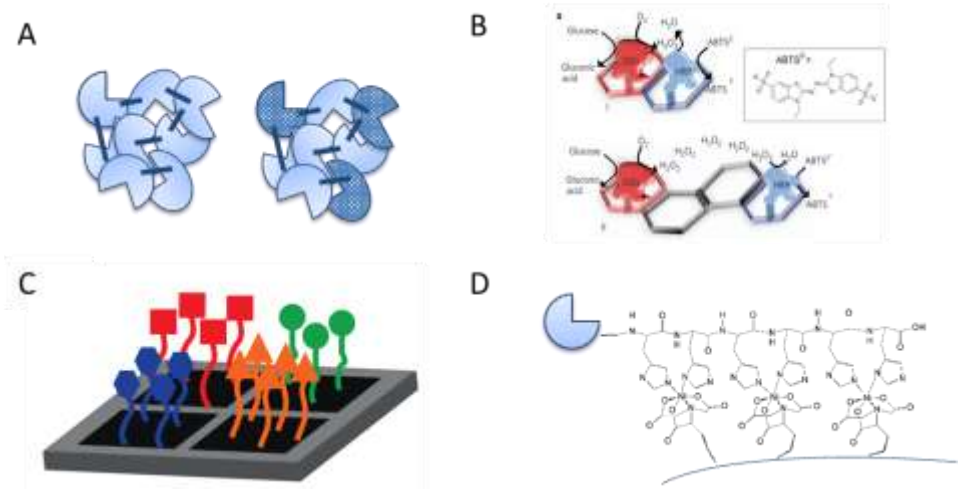

Figure 2.7. Co-immobilization techniques.
(A) cross-linked enzyme aggregates (CLEA), (B) DNA-directed co-immobilization [85]. (C) mRNA display [23]. (Reproduced with permission from Elsevier.) (D) His-tag and Ni-NTA interactions. (Reprinted by permission from Macmillan Publishers Ltd)

2.4.1.1 Cross-linked enzyme aggregates (CLEA)

Inspired by the idea of cross-linked enzyme crystals (CLEC), initially developed for X-ray diffraction study, cross-linked enzyme aggregates (CLEA) is a facile way of immobilizing enzymes without any physical support. The method does not require laborious protein crystallization, which required enzymes to be highly purified [86]. The enzymes can be easily precipitated out of solutions by simple addition of salts, a concentration step that is commonly applied [86]. Then the aggregates is treated with cross-linkers such glutaraldehyde to increase its mechanical stability (Figure 2.7A). Additives and modifiers can then be co-precipitated with the enzymes to further enhance the biocatalytic efficiency.

Several industrially important enzymes, such as lipases have been immobilized by CLEA and has effectively enhanced its activity in organic solvent by 12 fold [87].

For cascade enzymatic reactions, combi-CLEA is used, namely multiple enzymes were co-precipitated together to form the insoluble but functional agglomerate. Mateo *et al*. [88] demonstrated the utility combi-CLEA method on a bienzymatic system to produce enantiomerically pure (S)-mandelic acid. They reported that the product of combi-CLEA was 98%, which was higher than that of individual CLEA particle (94% ee). They attributed to the substrate channeling effect of combi-CLEA that prevented diffusion of intermediate into aqueous phase and hence possible racemization [88]. The limitations of the method are that the enzymatic function may be affected after precipitation and the inter-enzymatic distance is difficult to control.

2.4.1.2 DNA-directed immobilization (DDI)

To have a spatial arrangement of multienzymes, DNA-directed co-immobilization is another increasingly important method for *in vitro* multienzyme system. The strand specificities of DNA primers are able to self-assemble into scaffolds with nanometer-scale dimensions, termed “DNA origami” [89]. Wilner *et al.* reported a DNA-scaffold activated enzyme cascade reaction, comprising of glucose oxidase (GOx) and horseradish peroxidase (HRP) (Figure 2.7B) [85]. Without the DNA scaffold, there would not be any reactions observed. Since the dimensions of a DNA strand are defined, they were able to control the inter-enzymatic distance and hence control the product formation rate.

Inspired by the idea of RNA display (Figure 2.7C), Jung *et al*. devised a functional protein chip using hybridization between mRNA and DNA to capture five enzymes to produce trehalose [90]. They found a linear concentration range that the amount of enzymes captured correlated with its catalytic efficiency. Therefore, by using the novel chip, the five-enzyme pathway for trehalose production *in vitro* could be effectively optimized.

A significant technical limitation is to link DNA with the enzymes. Although the sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (Sulfo-SMCC) chemistry is efficient in cross-linking, the ratio of DNA amount per protein requires careful titration. Moreover, the orientation of the protein on the chip is difficult to control, which might affect the enzymatic activity. There have been many modifications to both DNA and proteins as suggested in literature to overcome these issues. This include the use of aptamer [91-93], fusion proteins [94, 95] and chemical modifications to protein [96]. One interesting strategy was to modify DNA strand with Ni-NTA, a metal chelating group that would non-covalently interact with proteins engineered with polyhistidine-tag [97]. This strategy would be reviewed in the following section 2.4.1.3.

2.4.1.3 Immobilized metal affinity chromatography (IMAC): His-tag and Ni-NTA

IMAC was invented a decade ago [98] and is extensively used in protein purification and immobilization (Figure 2.7D). The advantage of this method is that the orientation of the immobilized protein can be controlled, since the his-tag

can be engineered at either C-terminal or N-terminal of the protein of interest. Correct orientation of the enzymes was shown to be critical to enzymatic activities on solid support [99, 100]. Moreover, IMAC essentially combines the purification and immobilization into one single step, which is a cost-effective way to immobilize enzymes [101, 102]. By this method, Shao *et al*. co-immobilized a five-enzyme biosynthetic pathway for uridine 5'-diphosphate *N*-acetylglucosamine production, and achieved gram-scale productivity [103]. Liu and co-workers [104] demonstrated the direct immobilization of seven enzymes onto Ni-NTA agarose beads from cell extracts for UDP-galactose production. They were able to alter the spatial arrangement of the enzymes by either immobilize one enzyme on one bead or co-immobilize seven enzymes onto one bead. They showed that the co-immobilized multienzyme system was able to achieve higher product titer than the free enzymes after incubating for 48h.

A drawback of the method is the weak, non-covalent interaction between his-tag and Ni-NTA, which is approximately in submicromolar to micromolar range [105]. Therefore, high surface density of Ni-NTA was used to modify the solid support [106].

*2.4.2 Semi-*in vitro *synthesis and whole-cell Biocatalysis*

To scale-up the *in vitro* multienzyme synthesis, other semi-*in vitro* method was explored (Figure 2.2), such as whole-cell Biocatalysis [107], permeabilized whole-cell Biocatalysis [108] , and cell-free synthesis [109, 110]. These methods omit the enzyme purification step and enable the scale-up of *in vitro* system. The

development in the technology has increased dramatically (Figure 2.8), with production in the industrial scale (>100L).

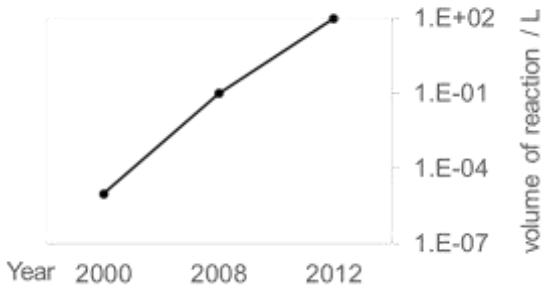

Figure 2.8. Milestones in the scalability of batch *E. coli* extract cell-free protein synthesis reactions [23]. (Reproduced with permission from Elsevier.)

Recently, Zawada *et al.* reported successful scale-up of cell-free cytokine production from microscale to 100L scale and achieved 700mg/L by using *E. coli* cell extracts [111]. In their study, they showed almost linear extrapolation when the reaction size was scaled-up. Their method has been applied for industrial antibody and vaccine production. This study successfully dispelled the myth that *in vitro* system is not scalable, and showed a great promise for the technology to be industrialized.

2.5 Amorpha-4,11-diene and Artemisinic acid synthesis pathway

Artemisinin is the most effective drug ingredient to combat the deadly disease malaria [112]. Recently, its potential applications in cancer treatment are promising and attract international attentions [113]. It is naturally derived from the leaves of the sweet wormwood, *Artemisia annua*. However, the yield was extremely low, ranging from 0.03-0.3%. Moreover, the supply of artemisinin

fluctuates owing to the weather and overall harvest [114]. Therefore, metabolic engineering of microorganism to produce the precursor, artemisinic acid, from glucose was developed and achieved approximately 25g/L yield in 7 days [3]. To our best knowledge, the production of dihyroartemisinic acid (DHAA) has not been demonstrated *in vivo* or *in vitro*. Therefore, we would like to explore the multienzyme biosynthesis technology to optimize the production of AD and DHAA *in vitro*.

DHAA is shown to be the immediate precursor to artemisinin, which can be easily converted by sun light [115]. It was biosynthesized through a series of oxidation reactions from the key precursor amorphadiene (AD), as shown in Figure 2.10. AD, which is the first key cyclization product from the ubiquitous precursor farnesyl pyrophosphate, represents a key branching point from the upstream pathways. As stated in section 2.1, all isoprenoids shared the common building blocks: isopentenyl pyrophosphate (IPP) and Dimethylallyl pyrophosphate (DMAPP), which both are derived from either the mevalonate (MVA) pathway or the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway. Recently much attention has been given to the DXP pathway, since the enzymes are promising targets for anti-infective and anti-bacterial agents [55, 116]. Moreover, the pathway has higher theoretical carbon yield with balanced co-factors, which are advantageous for industrial IPP / DMAPP based isoprenoids production [117]. Nevertheless, MVA pathway is also very important, and it is involved in many research areas dealing with hypercholesterolemia and metabolic bone diseases [118].

2.5.1 The mevalonate (HMG) pathway

The mevalonate (HMG) pathway for the biosynthesis of isoprenoids from acetate represents the initial steps in a series of enzymatic reactions for production of polyisoprenoids (e.g. dolichol) and sterols (e.g. lanosterol, ergosterol, cholesterol) in fungi, plant cytoplasm, animals, most other eukaryotes, archaea and some eubacteria [119, 120]. The study in this thesis was focused on the production of amorphadiene, a sesquiterpene from mevalonic acid shown in Figure 2.9.

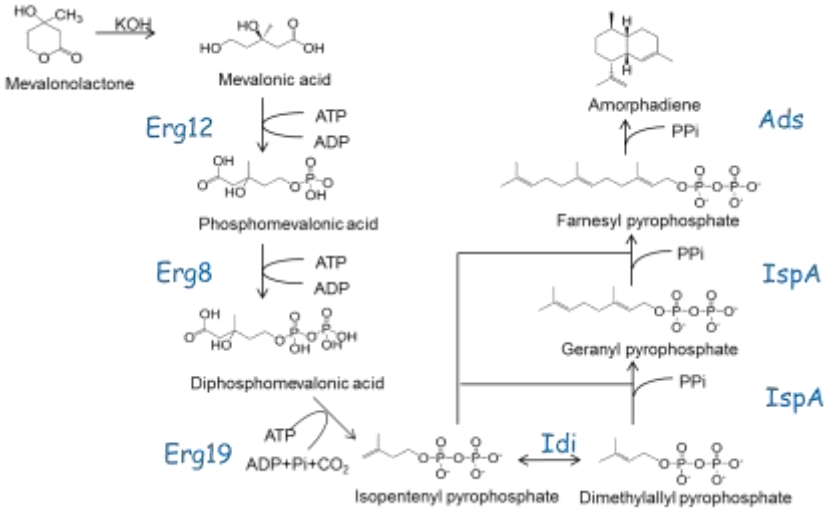

Figure 2.9. The schematic representation of the amorphadiene synthesis pathway.
The enzymes are as follows. Erg12: mevalonate kinase. Erg8: Phosphomevalonate kinase. Erg19: Mevalonate diphosphate decarboxylase. Idi: Isopentenyl diphosphate isomerase. IspA: Farnesyl diphosphate synthase. Ads: amorphadiene synthase.

Mevalonate kinase (Erg12, EC 2.7.1.36) is the first enzyme that transfers γ-phosphate group of ATP to mevalonic acid and form mevalonate 5-phosphate. The enzyme from different genomic background has been amplified and studied extensively. Despite the fact that they catalyze the same reaction, their allosteric property differs significantly. Erg12 from both human and *S. cerevisiae* was

demonstrated to be potently inhibited by the downstream diphosphate compounds, such as farnesyl pyrophosphate (FPP) [121, 122]. On the contrary, Primak *et al.* [123] recently reported a feedback resistant mevalonate kinase from archaeal organism *Methanosarcina mazei*. Structural insight and basic biochemical analysis revealed that FPP would compete with the ATP binding site [124]; there is a region of several hydrophobic amino acid side chains positioned near the ATP binding site, which would then interact with the polyisoprenoid chain of FPP. Changes in orientation and solvent exposure of the region resulted in different feedback inhibition pattern of Erg12 [124].

Phosphomevalonate kinase (Erg8, EC 2.7.4.2) is the subsequent enzyme that catalyzes the reversible reaction of mevalonate 5-phosphate and ATP to form mevalonate-5-diphosphate and ADP. Distinctive biochemical and structural property exist with Erg8 isolated from higher eukaryotes or from fungi and bacteria [125]. Bloch *et al*. discovered and isolated Erg8 from yeast extracts [126], and found the enzymatic activity decreased when ATP concentration was increased above 10mM. Although the enzyme contains an ATP binding site, it is not regulated by downstream pyrophosphate compounds. Biochemical analysis of Erg8 from *Streptococcus pneumonia* suggested a random sequential reaction mechanism [127].

Mevalonate diphosphate decarboxylase (Erg19, EC 4.1.1.33) is the last enzyme in the mevalonate pathway that catalyzes the ATP dependent decarboxylation of mevalonate 5-diphosphate. The enzymatic reaction is essential for yeast viability [128]. It has been suggested that Erg19 is regulated since sterol

production decreased when Erg19 was overexpressed [128]. Therefore, metabolite analogs such as 6-fluoro-MVAPP was developed as an potent inhibitor to block sterol biosynthesis [129].

Isopentenyl diphosphate isomerase (Idi, EC 5.3.3.2) catalyze a crucial activation step in the isoprenoid biosynthesis pathway [130]. Although this enzyme may seem not to be important in the DXP pathway, it is a critical enzyme for the mevalonate pathway to produce the electrophile dimethylallyl pyrophosphate (DMAPP). It has been reported that Idi plays an important role in control of the isoprenoid flux [131]. Structure of *E. coli* Idi was recently unraveled and revealed the unusual mechanism involving protonation of an unactivated carbon-carbon double bond, a reaction rarely encountered in nature [130]. Several inhibitors to Idi have been developed in the literature, such as the epoxide of isopentenyl pyrophosphate which is also a stimulator of γδ T cells [132].

Farnesyl diphosphate synthase (IspA, EC 2.5.1.10) represents one important enzyme for condensation of IPP and DMAPP into C10 and C15 isoprenoid backbone [133]. The enzyme was found to be remarkable in differentiating the isomers IPP and DMAPP and to initiate the chain elongation with very high stereoselectivity [133]. Single site mutation of IspA can catalyze the formation of GPP and GGPP [134]. Detailed kinetics analysis of FPP reaction with IPP and GPP revealed an ordered bi-bi reaction mechanism [135]. Recently, this enzyme has been a target to treat bone diseases and cancer treatment, because

nitrogen-containing bisphosphates are effective blockers of IspA, and are also inhibitors of osteoclastic bone resorption [136, 137].

2.5.2 Terpene synthase: Amorphadiene synthase

Terpene synthase is a class of important enzymes that produce some of the diverse isoprenoid structures. The crystal structure of taxadiene synthase was reported recently, and revealed the structural resemblance among the terpene synthases [138]. Yet, terpene synthases showed remarkable product selectivity and specificity despite using the same substrate. The availability of amorpa-4,11-diene synthase (Ads, EC 4.2.3.24) gene has opened up the possibility of genetically engineering microbes to produce amorphadiene (AD) [139, 140]. Ads acts on a critical branch point that cyclizes FPP into AD. The structure of Ads is not yet available, but reaction mechanism involving a shift in the diphosphate group has been proposed based on NMR studies [141]. Similar to other tepene synthase, Ads exhibited extremely low catalytic efficiency [142], which would probably be due to the slow release of hydrophobic product into aqueous medium [143]. Therefore high-throughput assay has been developed to screen for terpene synthase with improved catalytic efficiency [144]. Since accumulation of FPP is toxic to host cells, Zhao *et al.* developed a screening method based on cell viability to screen for variants of terpene synthase [145].

2.5.3 Cytochromes P450: CYP71AV1

Cytochrome p450 enzyme catalyzes a critical biotransformation of terpenoids, which enabled downstream functionalization of the terpenoids. The thermodynamically unfavorable addition of oxygen atom to hydrocarbon bond has always been an interest to organic chemists. The cytochrome p450 (CYP71AV1, EC 1.14.13.158), responsible for amorphadiene oxidation, has been isolated recently, and shown to catalyze three consecutive oxidation of amorphadiene to form artemisinic acid (AA) (Figure 2.10) [146]. The membrane bound enzyme is NADPH dependent and requires a redox partner, cytochrome p450 reductase (CPR), which was also isolated from *Artemisia annua*. Komori *et al*. has reported a key cysteine residue in CYP71AV1 that without which the stereoselectivity of the enzyme would be compromised [147]. CYP71AV1 has been overexpressed together with the amorphadiene synthesis pathway in different host cells, such as tobacco [148] , *E. coli* [149] and Yeast [3], and high production of artemisinic acid is successfully obtained in yeast. It was observed that overexpression of CYP71AV1 would affect upstream AD production [3]. Therefore, a novel semi-biosynthetic route was suggested; instead of using CYP71AV1, the substrate promiscuous $p450_{BM3}$ was used to add an epoxide group on AD and subsequently converted to AA (Figure 2.10) [150]. Another biosynthetic route was also proposed, comprising three other enzymes, alcohol dehydrogenase (ADH1) [3], aldehyde dehydrogenase (ALDH1) [151], and double-bond reductase (Dbr2) [152]. Dihydroartemisinic acid would be produced with the combined action of the three enzymes and CYP71AV1 (Figure 2.10).

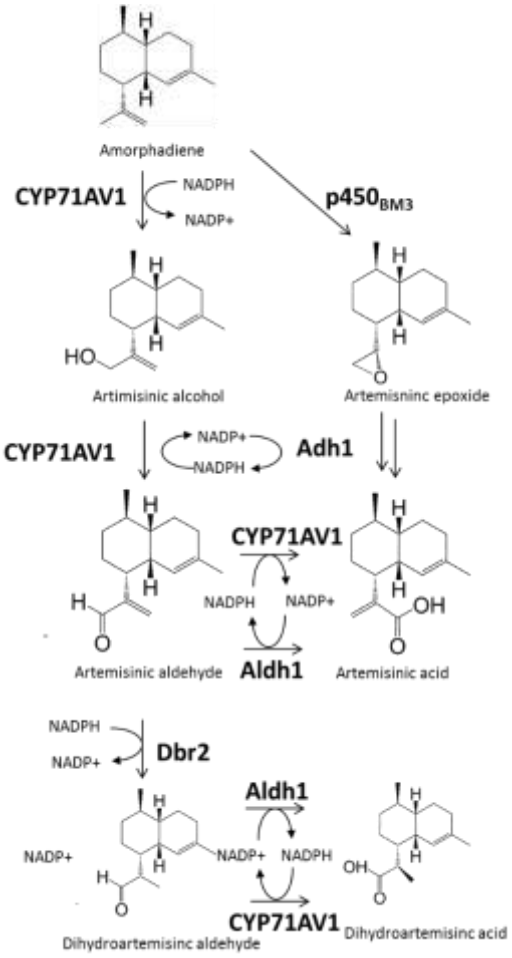

Figure 2.10. Schematic representation of downstream pathways that convert armorphadiene to artemisinic acid and/or dihydroartemisininc acid.

The enzymes are as follows: CYP71AV1: amorphadiene oxidase. Adh1: alcohol dehydrogenase. Aldh1: aldehyde dehydrogenase. Dbr2: double bond reductase.

Chapter 3 Statistical experimental design guided optimization of a one-pot biphasic multienzyme total synthesis of amorpha-4,11-diene

3.1 Introduction

Malaria is a contagious disease that has claimed millions of life annually, and continues to infect more than 0.5% global population, especially in less developed nations [153]. Artemisinin is the key ingredient of the most potent treatment to Malaria [112]. It belongs to the diverse class of isoprenoids that are derived from the building blocks: IPP and DMAPP. Traditional supply of artemisinin solely depended on extraction from the leaves of the sweet wormwood plant *Artemisia annua* [154]. Since the growth of crops is slow and seasonal, this method inevitably results in supply fluctuation of artemisinin [155]. Efforts in metabolic engineering and synthetic biology have made some promises to even out the supply cycle by engineering fast growing microbes to produce artemisinin and its precursor, artemisinic acid [146]. Recently, a multistep semi-synthesis of artemisin has been reported where yeast cell was used to produce precursors for further chemical conversions [3]. Invariably, the complex cellular environment renders any optimization process a challenge to control and maximize productivity [39].

In vitro multienzyme pathway assembly is a useful approach complementing *in vivo* metabolic engineering [156]. As mentioned in Chapter 2, Cheng et al. demonstrated the feasibility of producing polyketide by enzymatic total synthesis [6]. They assembled 12 pathway enzymes from different production hosts and were able to achieve an overall yield of 25% from simple

raw material. This bottom-up method successfully bypasses cellular barriers and allows a higher degree of freedom for pathway manipulation. At the same time, it still ensures regioselectivity and enantioselectivity of the product. Moreover, enzymatic reactions involve fewer chemicals that can simplify purification and reduce the cost of downstream process. *In vitro* multienzyme biosynthesis has been touted as a promising technology that may replace many chemical synthesis processes due to its high efficiency [8].

A significant challenge in a multi-enzymatic reaction is the need to optimize the various steps involved simultaneously so as to obtain high-yield of a product. Here we demonstrated the synthesis of amorpha-4,11-diene (AD), a key precursor to artemisinin, from mevalonic acid (MVA) by assembling seven enzymatic steps in one-pot with two-phase reaction condition (Figure 2.9). Subsequently, we set to optimize the pathway productivity by means of Taguchi orthogonal array design in an attempt to balance the enzymatic levels under pre-determined reaction conditions. The information gained led us to identify an inhibitory step of farnesyl pyrophosphate synthase (IspA), the critical factor Ads and significantly improved the AD yield. Therefore, this technique developed in the present study demonstrated a complementary way of producing valuable drug precursors with the ability to identify the limiting steps and balance the enzymatic flux in an efficient manner.

3.2 Results

3.2.1 Enzymatic purification and characterization

Individual enzyme was overexpressed in *E. coli* strains. Sodium dodecylsufate polyacrylamide gel electrophoresis (SDS-PAGE) results showed that the enzymes were expressed at high levels. However, the yield of purified individual enzymes obtained by immobilized metal affinity chromatography differed significantly (Table 3.1). This was mainly due to the differences in the solubility of the enzymes (Figure 3.1A) [20]. In particular, there was almost no detectable soluble fraction of Ads (Figure 3.1B). This led us to extensively optimize the strains, growth conditions and enzyme extraction methods for Ads. Repeated freeze-thaw method [157] was found to be effective in isolating Ads (1.6 mg/L) from cells with high purity as compared to detergent based lysis method. An initial attempt was made by mixing equal mass of the six enzymes in one pot with an overlay of dodecane phase where amorpha-4,11-diene was found to be produced in trace amounts (results not shown).

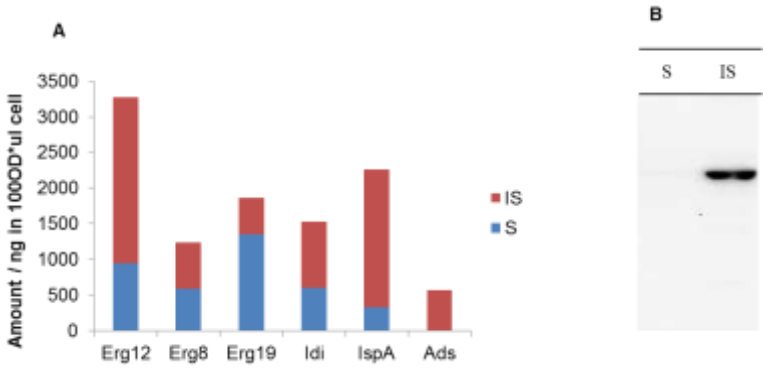

Figure 3.1 Solubility study of the pathway enzymes.
A: solubility of the pathway enzymes measured based on the method described previously. The amount is estimated and normalized by the cell number, which is equivalent to 0.1ml of cell OD600=1. [20]. B: western blot of the supernatant and pellet of the bacterial strains overexpressing ADS, by anti-his_6 primary antibody.

In order to better understand and optimize the *in vitro* system, steady-state kinetics of each enzyme was initially measured. The results were summarized in Table 3.1. The enzymatic concentrations were determined to ensure the measurement of the initial rate of reaction were linear in the first 15 mins. From the results, Ads displayed a significantly lower turnover number, a 2 orders of magnitude lower than the other five enzymes. It seemed to be an intrinsic property of terpene synthases, which was proposed to be limited by the release of the product [142, 143, 158]. Therefore, Ads was identified to be the bottleneck step in the multienzyme synthesis reaction.

Table 3.1. Purification and characterization of each pathway enzyme.

Enzyme (Synonyms)	EC #	MW**	K_m/μM	K_{cat} /s^{-1}	Enzyme Yield* /mg/L
Mevalonate kinase (Erg12)	2.7.1.36	49524	460±153 (MVA)	5.5±1.6	2-8
Phosphomevalonate Kinase (Erg8)	2.7.4.2	51520	780±280 (PMVA)	22.0±7.0	1.5-2.5
Diphosphomevalonate Decarboxylase (Erg19)	4.1.1.33	45181	190±52 (PPMVA)	2.8±0.5	15-60
Isopentyl pyrophosphate isomerase (Idi)	5.3.3.2	21331			6.5-28
Farnesyl pyrophosphate synthase (IspA)	2.5.1.92	32982	200±92 (IPP)	1.5±0.6	0.75-2
Amorpha-4,11-diene synthase (Ads)	4.2.3.24	64624	43.7±10	0.05±0.01	0.3-1.6

The bracket contains the specific substrate that the K_m is measured for.
*The enzyme yield is defined as the final amount of enzymes obtained after purification from a liter bacterial culture. The results have been repeated 3 times.

**The molecular weight (MW) of the enzyme was calculated based on its amino acid sequence.

3.2.2 Tuning enzymatic levels by Taguchi orthogonal array design

Table 3.2. Taguchi L16 (4^5) orthogonal arraray design and results.

	Levels*						
Runs	A	B	C	D	E	AD Yield	Specific AD Yield**
1	2	3	4	1	2	11%	34
2	2	4	3	2	1	12%	62
3	1	2	2	2	2	9%	63
4	1	4	4	4	4	18%	16
5	1	1	1	1	1	5%	49
6	1	3	3	3	3	10%	29
7	2	1	2	3	4	8%	15
8	4	2	3	1	4	14%	24
9	3	3	1	2	4	13%	27
10	4	3	2	4	1	24%	33
11	2	2	1	4	3	16%	23
12	3	4	2	1	3	9%	39
13	3	2	4	3	1	10%	22
14	4	4	1	3	2	26%	72
15	3	1	3	4	2	13%	19
16	4	1	4	2	3	16%	32
Predicted Max AD yield 1 (P1)	4	4	4	4	2	15%	17
Predicted Max AD yield 2 (P2)	4	4	4	4	1	2%	2
Experimentally Max AD yield	4	4	1	3	2	20%	56

Analysis of variance	
Model p-value	0.0046
R^2	0.94
Adj-R^2	0.85

* Refer to Table 3.3 for the actual enzyme concentrations corresponding to the coded levels. A: mevalonate kinase (Erg12), B: phosphomevalonate kinase (Erg8), C: diphosphomevalonate decarboxylase (Erg19), D: isopentenyl pyrophosphate isomerase (Idi), E: farnesyl pyrophosphate synthase (IspA).

** Specific AD yield is a dimensionless parameter measuring mg/L product produced by mg/L of enzyme. (refer section 3.5.6 for the detailed description).

To balance the enzymatic flux and to analyze the contribution of the other five enzymes to the final yield of AD, a combinatorial approach was carried out assisted with Taguchi orthogonal array design [67, 159]. The reaction conditions were fixed at pH 7.4 and 30 ^{o}C, with a constant Ads concentration of 100 mg/L (1.5 μM). The results were summarized in Table 3.2, and the corresponding enzyme concentrations were shown in Table 3.3. Remarkably, we observed divergent AD yield with varying amounts of enzymes, where the best enzymatic ratio (run14) produced 5 fold more of AD as compared to the lowest ratio (run 5) in which the enzymes had equal activities.

Table 3.3. Actual enzyme concentrations corresponding to the coded levels in Taguchi orthogonal array design.

	Coded levels			
Variables (mg/L)	1	2	3	4
A: Erg12	0.8	4	20	80
B: Erg8	0.2	1	5	20
C: Erg19	1.5	7.5	37.5	150
D: Idi	1.8	9	45	180
E: IspA	1.8	9	45	180

To further examine the influence of each enzyme on AD yield, the average effects analysis was determined. Figure 3.2A and B showed the average values of each level of the five enzymes on the AD yield. The five enzymes could be classified into two main groups: A (Erg12), B (Erg8) and D (Idi) positively enhanced AD yield when their activities were increased (Figure 3.2A). However, varying the activities of Erg19, and IspA did not have appreciable effect over AD yield (Figure 3.2B). Moreover, the half normal plot (Figure 3.2C) clearly indicated that Erg12, Erg8 and Idi were three main factors that had stronger

influence to maximize AD yield. Therefore, among the 16 runs, higher AD yield was obtained from combinations where all the three main enzymes were at higher activities (run 14 and run 12). The model predicted the maximum AD yield would be achieved when the first four enzymes were at their highest activities (Table 3.2). Attempt to validate this finding showed no significant improvement over AD yield (results not shown), suggesting that the activity ratio of Erg12:Erg8:Erg19:Idi:IspA:Ads of 100:100:1:25:5:1 (run 14) was near the local optimum. Further improvement of AD yield may require a change in the pre-determined reaction conditions. More importantly, this combination of enzymatic levels resulted in the highest specific AD yield among all the experimental runs designed by Taguchi method. Therefore, this enzymatic ratio was chosen as the reference condition to further optimize the multienzymatic synthesis system.

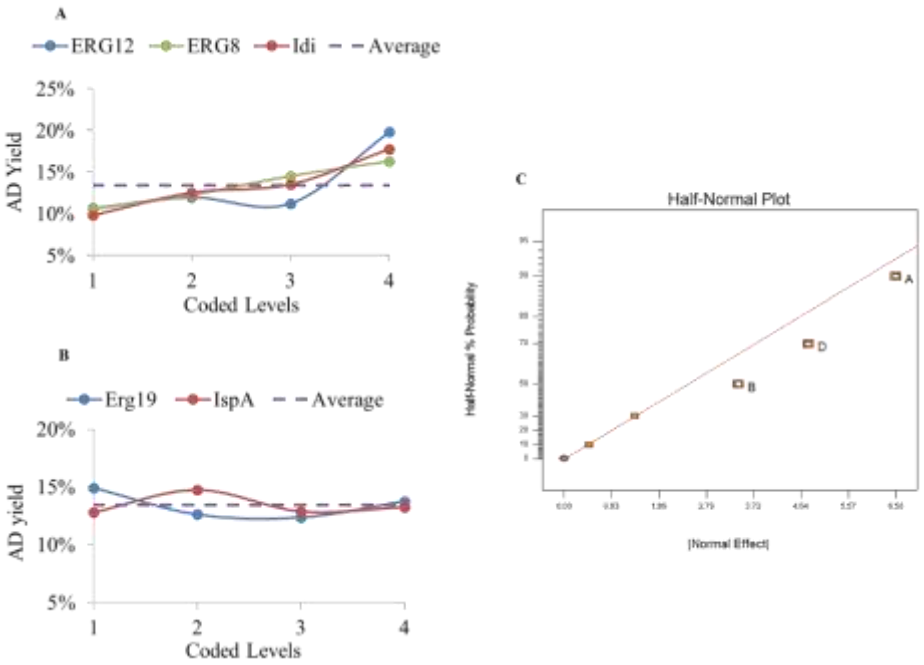

Figure 3.2. The Taguchi orthogonal array design results.
To examine the influence of each enzyme on amopha-4,11-diene (AD) yield, the average effects analysis was determined. The five enzymes can be classified into two main groups. A: Average values of each level of factors Erg12, Erg8 and Idi on AD yield. The group of enzymes has a

positive correlation with AD yield. B: Average values of each level of factors Erg19 and IspA on AD yield. The group of enzymes has little or no effect on AD yield. C: The half-normal plot indicates the significant factors on AD yield. Factor A, B, D represent Erg12, Erg8 and Idi respectively. The abbreviations are as follows. Erg12: mevalonate kinase, Erg8: phosphomevalonate kinase, Erg19: diphosphomevalonate decarboxylase, Idi: isopentenyl pyrophosphate isomerase, IspA: farnesyl pyrophosphate synthase.

3.2.3 Optimize IspA and Ads levels to enhance AD yield

One notable conclusion drawn from the Taguchi model prediction was that, to maximize AD yield, the activities of the first four enzymes were required to be maximized while retaining the activity of the fifth enzyme IspA, at moderate levels (Table 3.2). This alluded to the possibility of an inhibitory effect of IspA enzyme, since intuitively the yield should increase with increasing enzyme concentration. This led us to conduct a set of separate experiments where IspA concentrations were optimized, while retaining the other four enzymes at their reference levels.

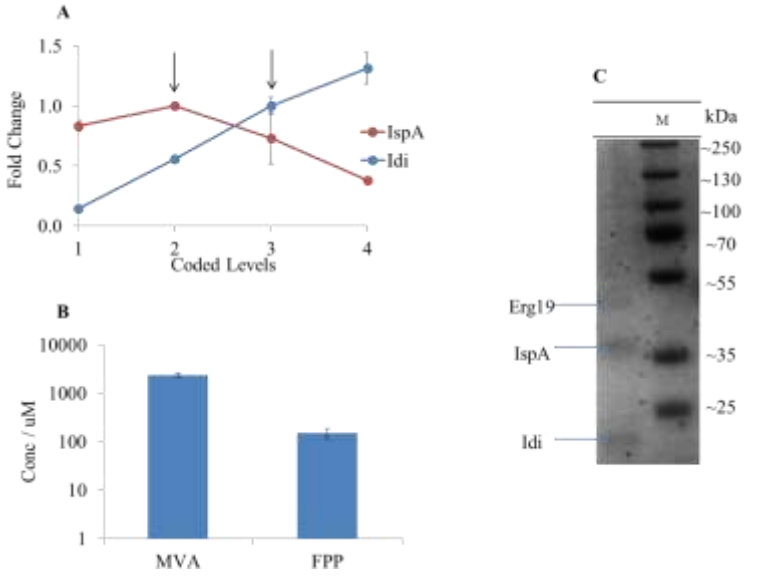

Figure 3.3. Inhibitory effect of IspA and analysis of the precipitates.
A set of separate experiments was conducted to validate the inhibitory effect of IspA. This was attributed to the precipitation of FPP. A: Fold change in amorpha-4,11-diene (AD) yield when increasing IspA and Idi concentrations while keeping other enzymes at reference level. Fold

change in AD yield was calculated by normalizing against AD yield obtained by reference enzyme levels, as indicated by the arrows. Presented data were average of triplicates and standard errors were drawn on the plot. B: UPLC-(TOF)MS analysis of the intermediates in the precipitates. Presented data were average of triplicates and standard errors were drawn on the plot. C: SDS-PAGE analysis of enzymes in the precipitates. The molecular weight of the each band present in the protein marker is indicated. The abbreviations are as follows. Erg19: diphosphomevalonate decarboxylase, Idi: isopentenyl pyrophosphate isomerase, IspA: farnesyl pyrophosphate synthase, MVA: mevalonic acid, FPP: farnesyl pyrophosphate.

Figure 3.3A showed the fold change in AD yield, with respect to that obtained by the reference enzymatic levels, when either Idi or IspA concentrations were varied. Idi concentrations were optimized as a control as it displayed a positive correlation with AD yield (Figure 3.3A). In contrast, a remarkable inhibitory effect was found when IspA activity was increased above its reference level (Figure 3.3A). A critical lead at this point of the study was that precipitates were formed in the reaction when IspA activity was increased. This interesting observation led us to hypothesize that the inhibitory effect of IspA was correlated with the precipitation. Interestingly, LCMS analysis revealed that the precipitates contained FPP and MVA which are the product of IspA, and the raw material respectively (Figure 3.3B). SDS-PAGE indicated that enzymes were co-precipitated, and therefore, exacerbating the overall productivity of the multienzyme reaction (Figure 3.3C).

Table 3.4. Stepwise reaction to identify the cause of precipitation.

Reaction	Precipitation	Estimated concentration of FPP
Erg12-Erg8-Erg19	No	-
Erg12-Erg8-Erg19-Idi-IspA	Yes	0.4mM
Erg12-Erg8-Erg19-Idi-IspA-Ads	Yes	0.4mM

To further identify which factor was the main reason that induced precipitation, the multienzyme reaction was analyzed stepwise. Precipitates were only visible when FPP was produced (Table 3.4). Therefore, it was hypothesized that the negative relationship of the increased activity of IspA and the overall productivity may be due to the accumulation of FPP in the context of the system examined. In order to test this hypothesis and further improve the AD yield of the system, response surface methodology was carried out to optimize the activities of IspA and Ads, so as to minimize the accumulation of FPP. The concentration ranges of IspA and Ads were chosen to be above and inclusive of the reference levels (Table 3.5). The experimental data obtained based on the design was fitted to a linear mathematical model (Equation 3.1).

$$AD = -0.096 - 0.18 \times IspA + 3.68 \times Ads$$

Equation 3.1. The linear mathematical relationship generated by Response Surface Methodology between amorphadiene yield and the factors, IspA and Ads concentration.

The R^2 and adjusted R^2 values were 0.93 and 0.91 respectively (Table 3.5), which indicated that the model was suitable to represent adequately the real relationships between the factors used. Interpretations from the model coefficients suggested a marked agreement with previous observation that IspA level was negatively correlated with AD yield. Moreover, the model implied a positive correlation between Ads level and AD yield, and Ads had a more pronounced effect on AD production. Thus, to validate the model, Ads activity was increased twice when compared to the reference enzymatic ratio. Unexpectedly, the conversion from MVA to AD was ~100%, which was ~5 fold improvement of

AD yield as compared to the reference condition (Figure 3.4). This further confirmed that Ads was the rate limiting step in the multienzyme synthesis reaction for AD production.

Table 3.5. Coded level combinations for a five-level, two factor response surface methodology with central composite design.

	Levels		
Run	A	B	AD (mg/L)
1	1	5	15.9
2	3	3	8.1
3	3	3	10.4
4	5	5	15.6
5	1	1	3.6
6	0.17	3	11.9
7	5	1	3.8
8	3	0.17	0.3
9	5.8	3	9.9
10	3	3	10.7
11	3	5.8	24.9
12	3	3	9.5

Actual enzyme concentrations corresponding to the coded levels

	Coded level variables	
Variables (mg/L)	1	5
A: ispA	36	180
B: ADS	36	180
Alpha (Rotatable)	1.41	
Analysis of variance		
Model p-value	<0.0001	
R^2	0.93	
Adj-R^2	0.91	

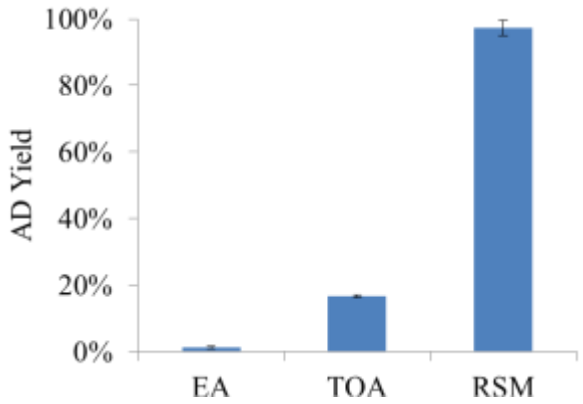

Figure 3.4 Summary of optimization of amorphadiene production.
EA: equal activities of the enzyme, which their concentrations in terms of Taguchi coded levels are Erg12(1), Erg8(1), tMvd(1), Idi(1), IspA(1) . TOA: optimized enzymatic activities by Taguchi orthogonal array method, which their concentrations in terms of Taguchi coded levels are Erg12(4), Erg8(4), tMvd(1), Idi(3), IspA(2). This combination of enzyme concentrations was used as the reference condition. RSM: response surface design suggested increasing Ads activity. The other five enzymes were kept at reference level.

3.2.4 Enhancement Ads specific activity by buffer optimization

Next, an attempt was made to enhance AD specific yield by examining the contribution of ions in the buffer. Potassium ion has recently been shown to significantly improve the activity of a terpene synthase by interacting with the H1-α loop of the enzyme [158]. Interestingly, a similar structure was found in Ads (Figure 3.5).

Figure 3.5 Sequence alignment of H-α1 loop [158] and Ads.

Thus, to test the effectiveness of monovalent ion, we supplemented the Ads reaction buffer with potassium ions. Figure 3.6A and B showed the change in Ads specific activity and the fold change in AD yield with respect to the reference condition respectively. As predicted, Ads specific activity was found to be

enhanced approximately twice with 100 mM potassium ion (Figure 3.6A). More interestingly, the overall AD yield by the reference enzymatic ratio was significantly improved about three times in the presence of 100 mM potassium (Figure 3.6B). Other monovalent ions, sodium chloride and ammonium chloride, were also titrated into the multienzyme reaction. A similar trend was observed that, with either 100 mM sodium ion or 50 mM ammonium ion, there was a marked three-fold improvement of AD yield (Figure 3.6B).

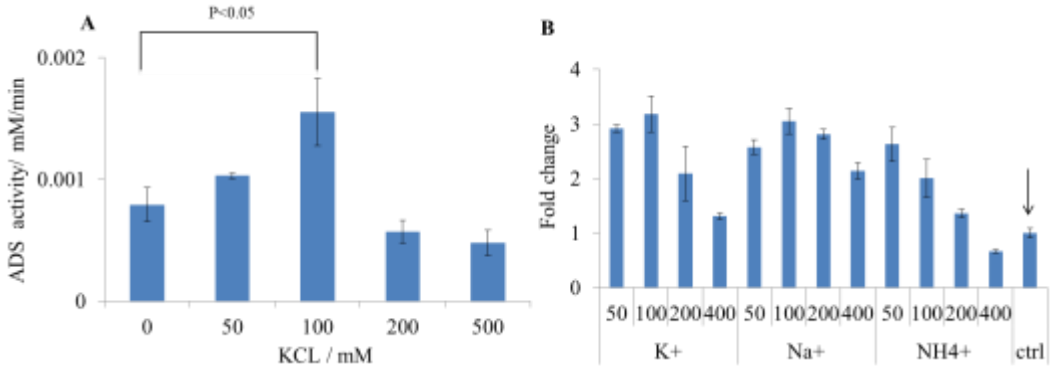

Figure 3.6. Effects of monovalent ions.
Monovalent ions were used to increase the specific activity of amorpha-4,11-diene synthase (Ads) and hence the specific amorpha-4,11-diene (AD) yield of the multienzyme synthesis reaction. A: Titration of potassium chloride concentrations, and their effects on Ads specific activity. Presented data were average of triplicates and standard errors were drawn on the plots. Student's t-Test with paired two samples for means was used to calculate the p-value in the statistical analysis. B: Titration of different monovalent ions concentrations and their effects on AD yield by reference enzymatic levels. Fold change in AD yield was calculated by normalizing against AD yield obtained by reaction without addition of monovalent ions, as indicated by the arrow. Presented data were average of triplicates and standard errors were drawn on the plots.

To further explore the effect of the buffer used, we varied pH from 6 to 9.1 and magnesium concentrations from 5 mM to 20 mM for AD production with reference enzymatic levels. Figure 3.7 showed the fold change in AD yield, with respect to the reference condition, when either pH or Mg^{2+} concentrations was varied. Remarkably, AD yield increased 3 times when the pH increased from 7.4

to 8.2, and there was no amorpha-4,11-diene detected when the pH was reduced to 6. By keeping pH at 7.4, the optimum Mg^{2+} concentration was found to be 15 mM, which resulted in a moderate 1.8 fold improvement of AD yield. However, no synergistic effect was observed at pH 8.2 and 15 mM Mg^{2+}. The optimum condition found was at pH 8.2 with 10 mM Mg^{2+}, which significantly enhanced specific AD yield three times (Figure 3.7).

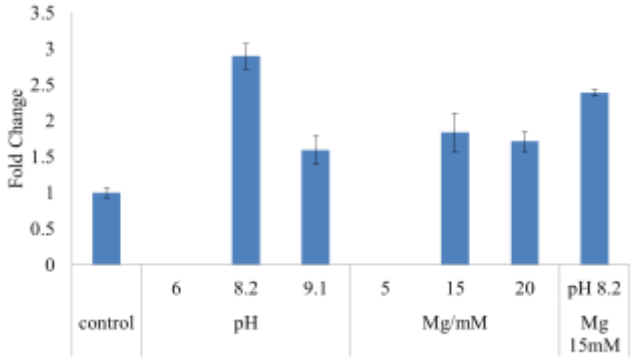

Figure 3.7. Optimization of buffer pH and magnesium concentration.
Varying buffer pH and magnesium concentrations was found to be helpful to enhance the specific amorpha-4,11-diene (AD) yield. Fold change in AD yield was calculated by normalizing against AD yield obtained with buffer pH 7.4 and 10 mM Mg2+, as indicated by the arrow. Presented data were average of triplicates and standard errors were drawn on the plots.

3.3 Discussion

In metabolic engineering and synthetic biology, an essential process is the *in vivo* balancing of metabolic pathway flux to achieve optimal productivity. Early successes of controlling pathway flux *in vivo* have been mainly achieved by tuning the promoter strength, ribosomal binding site sequences and plasmid copy numbers [160]. Controlling enzymatic concentration and activities *in vivo* is a significant challenge where limitations in global cellular and biochemical mechanisms including the solubility of overexpressed enzymes are not easily predictable [20]. Hence, cell-free enzymatic reaction would be an enabling technology that offers an alternative to these challenges. Recently, multienzyme synthesis for therapeutic products has been successfully demonstrated [23, 41, 161] In this study, the mevalonate pathway together with the downstream plant enzyme amorpha-4,11-diene synthase was individually purified and reconstituted in a single vessel and the overall biochemical reaction achieved an almost complete theoretical yield (~100%, 340 mg/L) of amorpha-4,11-diene production. This surpassed the carbon yield of *in vivo* biosynthesis which was approximately 15-20% [162].

One prerequisite for the system is the availability of functional and purified enzymes. Most of the recombinant enzymes produced in this study were insoluble and thus pose a challenge to obtaining sufficient amounts for *in vitro* reconstitution. When Ads was overexpressed in *E. coli*, most of the enzyme was found in inclusion bodies [163]. This is possibly due to the nature of the enzyme which more than 30% of the amino acids of Ads containing hydrophobic

structures, thus rendering the enzyme less soluble as compared to other pathway enzymes. Purification of terpene synthases is challenging, that some active enzymes were only recovered from inclusion bodies by *in vitro* refolding [164-166]. For Ads, using commercially available detergent to lyse the cells, we were unable to recover any functional enzymes. Repeated freeze-thaw method was found to be an effective method to improve the recovery of functional and soluble Ads for *in vitro* reaction.

Although the mevalonate pathway has been extensively optimized in engineered microbes, the *in vivo* mechanistic interactions of the combinations of pathway enzymes affecting amorpha-4,11-diene production are currently unknown. One advantage of the *in vitro* multienzyme system is the ability to precisely control and modulate the enzymatic activities in the same reaction condition. This allows the identification of interacting components or factors which can then guide further optimization. Conventionally, trial-and-error method was used to optimize the pathway enzyme concentrations. For 6 enzymes, each varied at 4 concentrations while keeping the rest at a constant level, there is a minimum of 24 reactions to be carried out. This is time-consuming and iterative, and sometimes fails to capture the optimum condition. Full factorial analysis requires too much experimental effort to be realistically carried out (5^4=625 reactions). Therefore, statistical experimental design methodology, in particular, Taguchi orthogonal array design was used in this study to rapidly and efficiently identify the local optimal ratio of enzymatic activities. Attempts to further manipulate the ratio did not result in significant improvement in the yield of AD.

This approach was rather helpful in identifying the negative effect of IspA which was due mainly to the accumulation and precipitation of FPP as well as the enzymes. Attempts were made to understand the involvement of MVA in the formation of precipitates. There was no obvious precipitates formation when MVA but not FPP was present in the reaction, suggesting a stochastic process that possibly resulted from the charge interaction between high concentration of MVA and Mg^{2+}. Similarly, no precipitation was found when enzymes were present without FPP. Hence, it was likely that FPP accumulated to sufficiently high levels may then precipitate, an observation consistent with a previous report [167]. FPP contains a hydrophobic 15-carbon moiety and is involved in post-translational lipid modification by protein farnesyltransferases [96, 168]. It is not unexpected that Mg^{2+} counterions may shield the negative charges of the pyrophosphate moiety resulting in the precipitation of FPP along with the enzymes. By increasing Ads activity, the flux of FPP towards AD production would have increased, and hence minimizing the accumulation of FPP. A benefit from such an increase in metabolite flux is the remarkable increase in the conversion of mevalonic acid (~100%). It will be interesting to examine the effect of co-immobilizing IspA and Ads to improve the channeling of substrate by proximity effect in the future.

By optimizing the buffer conditions, the specific AD yield was further enhanced, demonstrating the flexibility of the *in vitro* system conditions often intolerable to cells. Monovalent ions were found to be effective in improving both the specific activity of Ads and the specific AD yield by 2 and 3 fold,

respectively. Monovalent ions are well known to be required for many enzymatic activities [169]. Whether the monovalent ions may act as an allosteric activator to Ads by binding to specific secondary structures, e.g. H1-α loop of the enzyme [158], remains to be verified.

pH and magnesium ions play a critical role in moderating the enzymatic activities. Individual enzyme displayed vast different kinetic property under different pH and magnesium concentrations. Literature suggested that purified Ads displayed a higher catalytic efficiency when pH was increased in the presence of Mg^{2+} [142]. Therefore, the increased productivity of the multienzymatic pathway with increased buffer pH would likely be due to the enhanced specific activity of Ads in a more alkaline medium and the hypothesis remained to be verified.

3.4 Conclusion

In summary, amorpha-4,11-diene was successfully produced and achieved quantitative conversion by a multienzyme, biphasic system. We have demonstrated the utility of Taguchi method to efficiently identify the local optimum enzymatic ratio and identified the inhibitory effect of IspA, resulting in the accumulation of FPP. Further optimization of IspA and Ads activities by response surface methodology identified that Ads was a critical factor. By increasing the Ads activity, almost 100% conversion from raw materials to AD has been achieved. Further buffer optimization of monovalent ions, pH and magnesium ions, was able to enhance the specific AD yield 2-3 fold significantly.

The work-flow demonstrated here will be valuable to produce other isoprenoids in an efficient manner.

3.5 Materials and Methods

3.5.1 Bacteria strains and plasmids

Bacteria strains and plasmids used in this study were summarized in Table 3.6. The pET-11a (Stratagene, CA) was modified by replacing the T7 promoter with tac promoter to facilitate the transfer of the plasmids among different strains. A 5' SacI site and a 3' XhoI site were introduced downstream from the 6xHis open reading frame. The mevalonate pathway enzymes, namely mevalonate kinase (Erg12), phosphomevalonate kinase (Erg8) and pyrophosphomevalonate decarboxylase (Erg19), were amplified from *S. cerevisiae* genomic DNA with forward and reverse primers that contain corresponding SacI and XhoI sites. The PCR products were ligated into the modified pET-11a vector (Stratagene, CA) and transformed into competent *E. Coli* strain DH10B. Isopentenyl pyrophosphate isomerase (Idi) and IspA were from our previous study [20]. Ads gene was codon optimized and synthesized by Genescript with sequences encoding C-terminal 6xHis-tag, and subsequently cloned into a modified pBAD-B vector (Invitrogen, CA) using 5' SacI site and 3' XhoI site. The primers used for amplification of the genes were listed in Table 3.7. All the plasmids were transformed and harboured from *E. coli* XL10-gold (Stratagene, CA) and then transformed to respectively strains for enzyme overexpression (Table 3.6).

Table 3.6. Bacterial strains and plasmids used in this chapter.

Name	Description	Reference
E. coli BL21-Gold (DE3)	F^- ompT hsdS (r_B^- m_B^-) dcm^+ Tet^r gal λ(DE3) endA Hte	Stratagene
E. Coli DH10B	araD139 Δ(ara-leu)7697 fhuA lacX74 galK (Φ80 Δ(lacZ)M15) mcrA galU recA1 endA1 nupG rpsL Δ(mrr-hsdRMS-mcrBC)	NEB
E. Coli XL10-Gold	Tetr D(mcrA)183 D(mcrCB-hsdSMRmrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F9 proAB lacIqZDM15 Tn10 (Tetr) Tn5 (Kanr) Amy]	Stratagene
pTrc-His_6-Erg12	Plasmid for overexpression of Erg12 in *E. Coli* DH10B	This study
pTrc-His_6-Erg8	Plasmid for overexpression of Erg8 in *E. Coli* DH10B	This study
pTrc-His_6-Erg19	Plasmid for overexpression of Erg19 in *E. Coli* DH10B	This study
pET-His_6-Idi	Plasmid for overexpression of Idi in *E. Coli* BL21-Gold (DE3)	[20]
pET-His_6-IspA	Plasmid for overexpression of IspA in *E. Coli* BL21-Gold (DE3)	[20]
pBAD-Ads-His_6	Plasmid for overexpression of Ads in E. Coli DH10B	This study

Table 3.7. List of primers used for cloning the genes.

Primer Name	Sequnce
SacI-Sc.ERG12_F	GCGAGCTCTCATTACCGTTCTTAACTTCTGC
Sc.ERG12-XhoI_R	GCCTCGAGTTATGAAGTCCATGGTAAATTCG
SacI-Sc.ERG8_F	GCGAGCTCTCAGAGTTGAGAGCCTTCAGT
Sc.ERG8-XhoI_R	GCCTCGAGTTATTTATCAAGATAAGTTTCCGGA
SacI-Sc.Erg19_F	GCGAGCTCACCGTTTACACAGCATCCG
Sc.Erg19-XhoI_R	GCCTCGAGTTATTCCTTTGGTAGACCAGTCT
SacI-Ec_idi_F	GCTTAGAGCTCCAAACGGAACACGTCA
Ec_idi-XhoI_R	GTAACCTCGAGTTATTTAAGCTGGGTAAATGC
SacI-Ec_ispA_F	GCTTAGAGCTCGACTTTCCGCAGCAACT
Ec_ispA-XhoI_R	GTAACCTCGAGTTATTTATTACGCTGGATGA
SacI-Ads_F	GCGGAGCTCTCTCTGACTGAGGAAAAACCA
Ads-His6-XhoI_R	CGCCTCGAGTCAGTGATGGTGATGATGATG

Underlined text: restriction enzyme site

3.5.2 Expression and purification of Erg12, Erg8, Erg19, Idi and IspA.

Newly transformed colonies were picked from the agar plate, inoculated into 2xPY medium (20 g/L Peptone, 10 g/L Yeast extract, and 10 g/L NaCl, pH 7) containing 100 mg/L ampicillin and grew till stationary phase overnight at 37 ^{o}C

in an incubator-shaker (Shin Saeng Shaking Incubator, Finetech, Korea). The culture was then further transferred into fresh 2xPY medium (1% inoculation) with ampicillin for another 2.5 h at 37 ^{o}C, till optical density A_{600} reached 0.6-1.0. The enzyme expression was induced with 0.1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Temperature was reduced to 20 ^{o}C after induction for higher solubility of the enzymes [20]. The culture was grown for another 48 h and harvested by centrifugation. The cell pellets were stored at -20 ^{o}C till further use.

To purify the enzymes, the frozen cell pellets were resuspended in B-PERII reagent (Pierce, IL), according to the manufacturer's instruction, and vortexed at room temperature for 30 mins to completely lyse the cells. The soluble proteins were contained in the supernatant, which was diluted 15 times in NPI10 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) and incubated with 200 mg Ni-NTA resin (USB, Affymetrix, CA) at 4 ^{o}C for 2 h. The resin was washed with NPI10 buffer after discarding the binding supernatant, and the enzymes were eluted and collected by 400 μl NPI400 (50 mM NaH_2PO_4, 300 mM NaCl, 400 mM imidazole, pH 8). The enzymes were further concentrated by 3K Amicon ultra-0.5 ml centrifugal filter unit (Millipore, MA), and the concentrations were measured by Micro BCA protein assay kit (Thermo scientific, MA). The purified enzymes were further confirmed by sodium dodecyl sulfate-12% polyacrylamide gel electrophoresis (Bio-Rad, CA).

3.5.3 Expression and purification of Ads.

Bacteria culture was grown in 2xPY medium at 20 ^{o}C till stationary phase after Ads expression was induced with 10 mM L-arabinose. The cells were harvested by high speed centrifugation and resuspended in phosphate saline buffer (PBS). To purify Ads, cells were lysed by three rapid freeze-thaw cycles by -80 ^{o}C freezer and 37 ^{o}C incubator. The released enzyme was separated from cell debris by centrifuging at 3000 *g* for 15 min, and purified by Ni-NTA resin as described above.

3.5.4 Enzyme kinetics

The pathway enzyme kinetics was determined individually by initial rate measurements. In brief, the substrates and cofactors were added to 100 mM Tris/HCl reaction buffer (pH 7.4), and the reaction was initiated by adding pre-determined enzyme amount to ensure less than 10% substrate was consumed in 15 mins at 30 ^{o}C. The substrates concentrations were varied in equal steps in reciprocal space from 0.1 to 1 mM. The reaction was terminated by adding equal volume of 1% ammonium hydroxide and diluted 10 times into cold methanol. After high speed centrifugation, the supernatant was subject to UPLC-(TOF)MS analysis. Double-reciprocal plots of each enzymatic activity were constructed for the determination of K_m and K_{cat} values for the respective substrates.

3.5.5 Multienzyme reaction

The multienzyme reaction was carried out in a buffer (25 µl) consisted of Tris/HCl (100 mM, pH 7.4), $MgCl_2$ (10 mM), (±)mevalonic acid (10 mM), ATP (15 mM) and the purified enzymes. The reaction was performed at 30 ^{o}C with an overlay of dodecane phase that contained trans-caryophyllene (50 mg/L) as an internal standard. At the end of the reaction, the dodecane phase was diluted 10 times in ethyl acetate and subject to GCMS analysis. The (±)mevalonic acid was prepared by complete alkaline hydrolysis of 2 M (±)mevalonolactone (Sigma, MO) with equal volume of 2 M potassium hydroxide at 37 ^{o}C for 1.5 h, and neutralized by adding 1 M hydrochloric acid to pH 7 [170].

3.5.6 Experimental design

Taguchi orthogonal array design and response surface methodology with central composite design were calculated using Design Expert® V8 Software (Stat-Ease, Inc). Taguchi L_{16} (4^5) orthogonal array was constructed which can accommodate five control factors corresponding to the five pathway enzymes, each varied at four levels of concentrations (Table 3.3). The four enzymatic levels were normalized against Ads activity (AA), ranging from 1xAA, 5xAA, 25xAA and 100xAA to achieve sufficient coverage. The lowest level was equalized enzymatic activity, whereas the highest level was comparable enzymatic concentrations. The level of Idi was varied according to IspA. 16 randomized experimental runs were conducted to maximize AD yield (Equation 3.2). The specific AD yield (Equation 3.2) was another indicator of the pathway

productivity but was not considered in the design experiment. The two dimensionless readouts were calculated as follows:

$$\text{AD yield} = \frac{\text{Actual AD yield (mg/L)}}{\text{Theoretical yield based on 5mM raw material (mg/L)}}$$

$$\text{Specific AD yield} = \frac{\text{Actual AD yield (mM)}}{\text{Total enzymatic concentrations (mM)}}$$

Equation 3.2. The equations to determine the AD and specific AD yield.

Optimizing IspA and Ads activities was carried out by response surface methodology with central composite design, which involved the investigation of two factors (concentrations of IspA and Ads), each varied at five levels and four centre points for replication. The AD production at 6 h was taken as the response, before the reaction reached completion and any visible precipitations were formed. The experimental data obtained were fitted based on the most suitable model suggested by the software.

3.5.7 UPLC-(TOF)MS analysis of mevalonate pathway intermediates

The analysis was done based on the method developed previously with slight modification [42]. In brief, 5 µl samples were injected into a UPLC C18 column (Waters CSH C18 1.7 µm, 2.1 mm x 50 mm) connected to UPLC (Waters ACQUITY UPLC)-(TOF)MS (Bruker micrOTOF II, MA). Elution was carried out with a step change from 100% aqueous solution containing 15 mM acetic acid and 10 mM tributylamine (0.5 min) to 10% aqueous solution with 90% methanol for another 3.5 min. Electrospray ionization was used and mass spectrometry was

operated to scan 50–800 m/z in negative mode with 2500 V end plate voltage and 3200 V capillary voltage. Nebulizer gas was provided in 2 bar, dry gas flow rate was 9 ml/min, and dry gas temperature was 200 ^{o}C. At the assay condition, all the intermediates were detected in the form $[M-H]^-$. Retention time was subsequently determined for each intermediate with respective synthetic standards and the set m/z extraction range. The peak area was calculated and subsequently used to compute the intermediates concentrations with the software provided by the manufacturer. The calibration curves were constructed with synthetic standards prepared under similar reaction conditions without enzymes. Linearity of the assays were determined individually with coefficients of determinants (R^2) greater than 0.90.

3.5.8 GCMS analysis of amorpha-4,11-diene

The analysis was carried out based on the modified method developed by Martin et al. by scanning three ions; the m/z values are 117, 189 and 204 [170, 171]. 1 µl sample was injected into HP-5 column (Agilent Technologies 7890A gas chromatograph-mass spectrometry, Agilent, CA) with a linear temperature increase of 50 ^{o}C/min from 80 ^{o}C to 300 ^{o}C and hold at 300 ^{o}C for another minute. The peak area was calculated and subsequently used to compute the amorpha-4,11-diene concentrations with the software provided by the manufacturer. Amorpha-4,11-diene concentrations were determined relative to the internal standard trans-caryophyllene of known concentration.

Chapter 4 Unraveling the regulatory behaviour of in vitro reconstituted amorpha-4,11-diene synthesis pathway by Lin-log approximation

4.1 Introduction

In Chapter 3, we have reconstituted and optimized the amorpha-4,11-diene (AD) synthesis pathway comprising of mevalonate kinase (Erg12), phosphomevalonate kinase (Erg8), diphosphomevalonate decarboxylase (Erg19), isopentenyl pyrophosphate isomerase (Idi), farnesyl pyrophosphate synthase (IspA) and amorpha-4,11-diene synthase (Ads), as shown in Figure 2.9, for AD production. However, the regulatory behavior of the pathway was still unclear, especially the interactions of the metabolites with the key enzyme Ads and this is important for further *in vitro* pathway engineering.

The MVA pathway is mainly present in eukaryotes, archaea and some eubacteria and provides the cell with a variety of compounds which are involved in multiple cellular processes [118]. Enzymes stem from different species origin have unpredictable allosteric metabolic regulatory properties [120]. As mentioned in section 2.5.1, a farnesyl pyrophosphate (FPP) resistant mevalonate kinase was isolated from *Methanosarcina mazei* [123], while its counterpart from *Saccharomyces cerevisiae* and *Streptococcus pneumonia* was highly sensitive towards FPP [121] and mevalonate-5-diphosphate (PPMVA) [172, 173] respectively. Therefore, the intricacy of metabolic networks and its revolutionary diversity renders an intuitive analysis of a biochemical pathway behavior a difficult task.

Mathematical modeling approaches aim to reduce the complexity and provide insights for rational engineering, making them an important tool for metabolic engineering and synthetic biology [174]. Two main modelling strategies have been described in section 2.3. The conventional kinetics modeling governed by Michaelis Menten law is usually difficult to obtain satisfactory kinetic parameters and mathematical equations in the context of complex enzymatic reactions [175]. Moreover, it requires *a priori* network structure information and extensive parameter tuning to have a more accurate mathematical representation of a biochemical pathway [176, 177]. Stoichiometric modeling that based on mass balance is a powerful tool to estimate the pathway behavior, although it is still heavily dependent on the prior knowledge of regulatory pattern of the network [175]. Data-driven modeling such as neural network can be applied when mechanistic information is missing. However, time- and cost-intensive experiments are required to obtain sufficient data [178, 179]. Recently developed canonical modelling method such as Lin-log approximation is more promising to predict the regulatory patterns of the biochemical pathway based on steady state fluxes. Due to its defined model structure and basis on thermodynamic principle of enzyme catalyzed reaction, Lin-log approximation requires less number of parameters to be estimated without compromised accuracy [81, 180] and technically allows larger perturbations to predict the regulatory behavior (up to 20 fold difference) [79]. Therefore, it is highly suitable for structure identification for biochemical systems.

In this Chapter, the Lin-log approximation method was to unravel the regulatory behavior of *in vitro* reconstituted AD synthesis pathway from mevalonic acid (Figure 4.1), with the aim of further boosting up the productivity of the pathway. The use of reconstituted pathway is advantageous, (section 2.2), since it is highly reproducible and open to manipulation by changing concentrations of its constituents [181]. The known regulations for the pathway are depicted in Figure 4.1with black dashed line. With the strategy used, we successfully identified a novel interaction between ATP and Ads (Figure 4.1). Subsequently, upon the addition of ATP recycling system, the space- and time-yield of the pathway was further enhanced by more than 3 fold. To our best knowledge, this is the first investigation into the regulation of the MVA and AD synthesis pathway at a multienzymatic system level and the strategy developed here is applicable to other *in vitro* assembled pathways beyond the scope of this study.

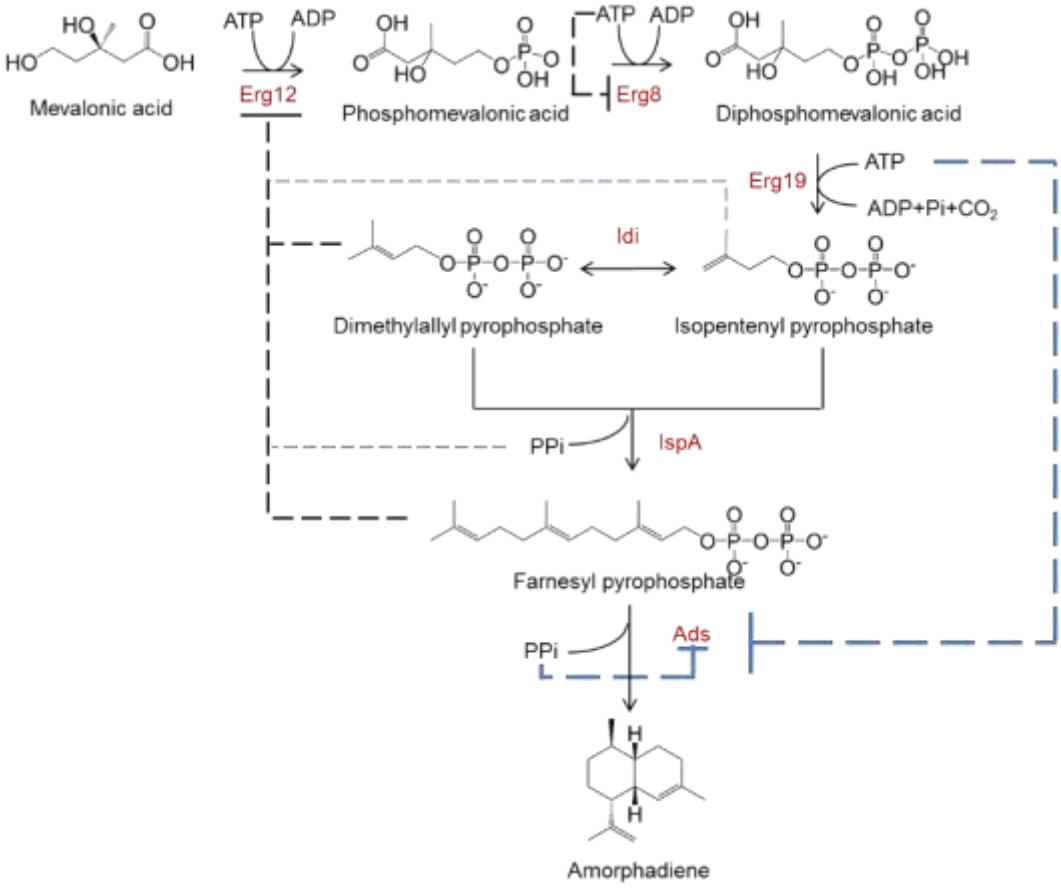

Figure 4.1. The schematic representation of the amorphadiene synthesis pathway with regulations shown.

The black dash line showed the known feedback regulation for the mevalonate pathway enzymes in the literature. The light blue dashed line is the additional regulation that has been found in the study for the mevalonate pathway enzymes. The dark blue dashed line showed the novel regulations been identified in the study; ATP and pyrophosphate are inhibitory to Ads activity. The abbreviations are as follows. Erg12: mevalonate kinase. Erg8: Phosphomevalonate kinase. Erg19: Mevalonate diphosphate decarboxylase. Idi: Isopentenyl diphosphate isomerase. IspA: Farnesyl diphosphate synthase. Ads: amorphadiene synthase. Ppi: Pyrophosphate.

4.2 Results

4.2.1 Elasticity estimation using the Lin-Log approach

One important pre-requisite of Lin-log approximation is to establish a steady state as the reference state. It is technically difficult to achieve a real steady state for *in vitro* reconstituted system; a quasi- steady state is considered as a reasonable approximation [181]. Therefore, enzymatic concentration was determined experimentally to ensure minimum fluctuation of the levels of the reaction intermediates. The time course of the reaction intermediates has been measured and shown in Figure 4.2. They were maintained at approximately the same level from 30 mins to 60 mins, and the AD production was almost linear during that period of time. Therefore, 30min was taken as the reference time point to determine the change in steady-state flux and metabolite concentrations.

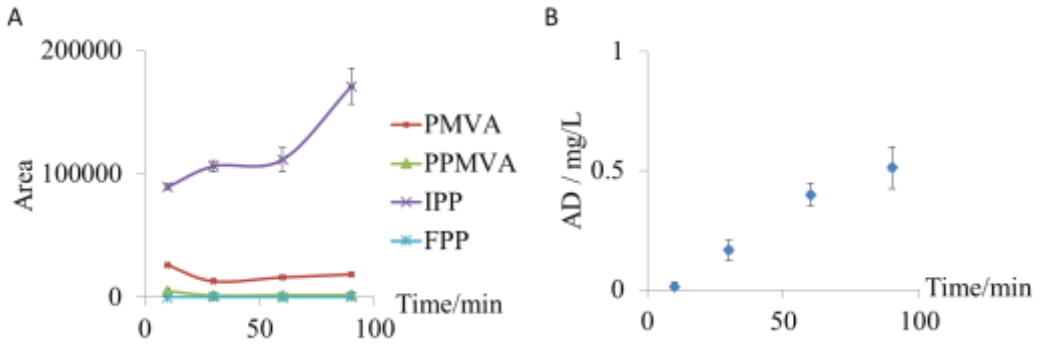

Figure 4.2. Time course of (A) intermediate concentrations, and (B) amorphadiene (AD) production.
From 30 to 60 min, the metabolite concentrations fluctuate the least, and AD production is linear. Therefore, 30min time point was taken to assay the change in the metabolite concentration change for Lin-Log approximation. All the data point was done in triplicates, and the standard deviation is shown.

Table 4.1. Experimental data obtained by perturbation experiments.

	Parameters							Metabolites (μM)							Cofactor (A.U.)	Flux (μM/h)
		Erg12	Erg8	Erg19	Idi	IspA	Ads	MVA	PMVA	PPMVA	IPP	DMAPP	GPP	FPP	ATP	
1	Erg12	0.10	0.20	0.50	0.15	0.15	0.40	2425	356	36	1085	189	11	76	3817	3.1
2		0.20	0.20	0.50	0.15	0.15	0.40	1806	474	73	891	217	13	93	1840	3.3
3	Erg8	0.15	0.16	0.50	0.15	0.15	0.40	2100	467	21	718	337	9	70	2209	2.2
4		0.15	0.24	0.50	0.15	0.15	0.40	2105	375	75	874	340	9	78	1644	2.3
5	Erg19	0.15	0.20	0.40	0.15	0.15	0.40	2125	401	24	881	374	12	63	1610	2.0
6		0.15	0.20	0.50	0.15	0.15	0.40	2131	426	31	702	317	7	58	1378	1.8
7	Idi	0.15	0.20	0.60	0.10	0.15	0.40	2250	429	48	950	266	9	43	1515	1.8
8		0.15	0.20	0.50	0.20	0.15	0.40	2305	382	44	769	342	13	82	1708	3.1
9	IspA	0.15	0.20	0.50	0.15	0.10	0.40	2279	430	43	788	384	6	64	1514	3.2
10		0.15	0.20	0.50	0.15	0.20	0.40	2058	422	31	741	311	7	81	1461	3.1
11	Ads	0.15	0.20	0.50	0.15	0.15	0.50	2489	402	63	685	419	9	80	2181	3.2

Subsequently, individual enzyme concentrations were perturbed around the pre-determined concentrations to obtain 11 independent measurements shown in Table 4.1. The variation in each metabolite concentrations across the 11 experiments was almost all within 2-3 folds to ensure higher accuracy of Lin-log approximation and the total sum of the measured concentrations from each experiment was within 10% variation from the initially added substrate concentration as shown in Figure 4.3A [80]. Multiple linear regression was performed under the 'maximum connectivity' assumption to obtain 11 sets of elasticities (equation Equation 4.6) [181]. Each had one out of the 11 experiments as a reference state. The optimal reference state was evaluated again based on the criterion of adjusted coefficient of determinant (R_a^2 , Equation 4.9) [80]. Figure 4.3B showed that run 9 has the highest R_a^2 value (0.958) among the 11 regressions, which indicated that the approximation was sufficient to reflect the true relationship between the metabolite concentrations and the steady state flux. Thus, the corresponding set of elasticities, with run 9 as reference state, and their 95% confidence intervals are given in Table 4.2, and was used for subsequent analysis.

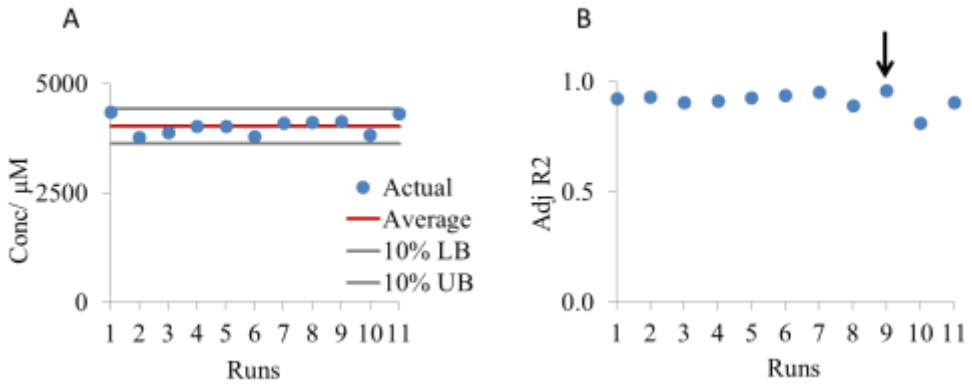

Figure 4.3. Analysis of the perturbation experiments.

(A). The sum of metabolite concentrations for each perturbation. The average value is approximately 4000μM, which is the same as the initial amount of substrate added. The 10% variation boundaries are indicated on the graph. All the perturbation experiments have total measured concentration varied with the 10% boundaries of the average value. This showed that the measurement was accurate to a large extent. (B). Adjusted coefficient of determination for 11 different reference states. "Maximum connectivity" assumption was used. The best reference state, run 9, is indicated by the black arrow.

We calculated flux and metabolite concentrations based on the mass balance equations derived with Metabolic Control Analysis (MCA) coefficients (Equation 4.7 and Equation 4.8), and plotted them against the actual measured values in the parity plot (Figure 4.4).

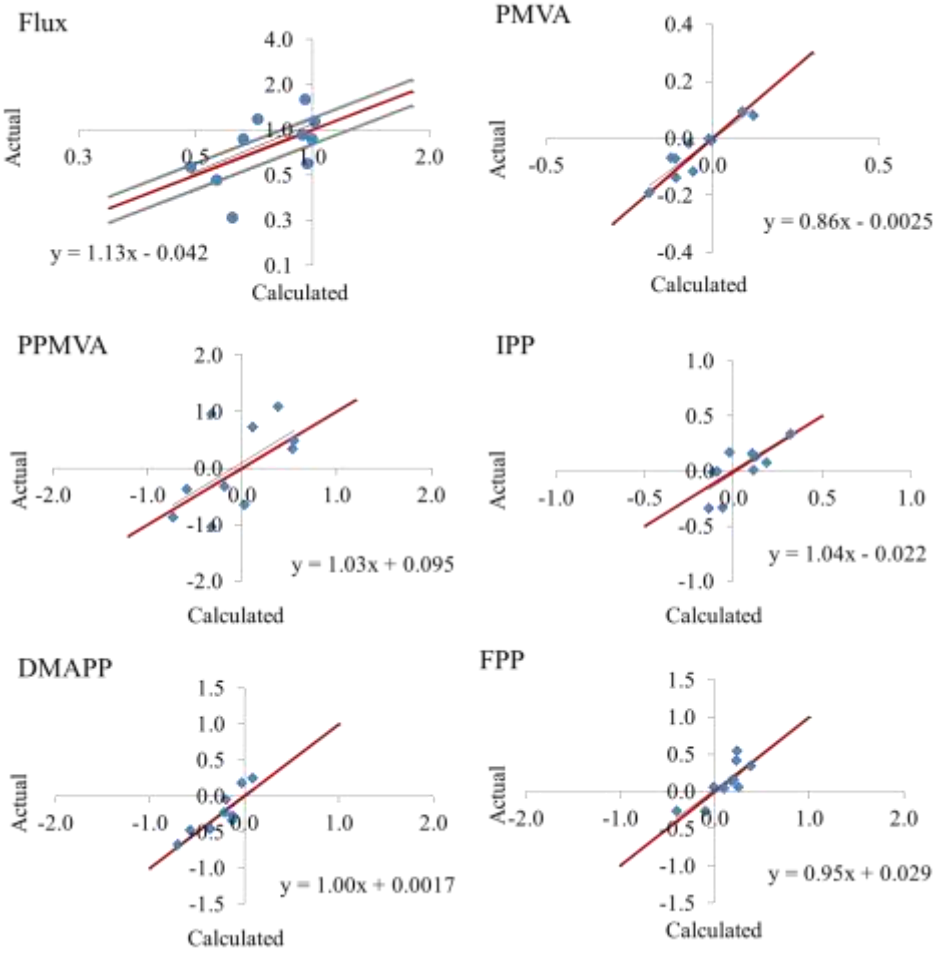

Figure 4.4. Parity plot of the calculated flux and metabolite concentrations with the actual measurements.

All the values are normalized with the reference state (run 9) measurements. The equations of trend lines are given.

Evidently, Lin-log approximation was able to accurately predict the actual measurements to a satisfactory level; the coefficients of all the trend lines between the calculated and actual measurements, including flux and metabolite concentrations, were close to 1. Majority of the calculated flux and metabolite concentrations were within 20% variation of the actual data. Hence, we used the lin-log approximation results to further guide our analysis.

Elasticities would infer to how a metabolite influence a reaction rate around the reference state. Usually, positive elasticity means a stimulatory effect, and negative elasticity means inhibitory effects [80]. Based on the 95% confidence interval, we were able to exclude half of the elasticity values in the Table 4.2, which its sign is hard to determine. Relatively larger confidence intervals occurred when the enzymatic reaction was directly involved with IPP and DMAPP, such as IspA and Idi reaction. Since the two compounds are isomers, which are hardly differentiated by chromatographic measurement used, an additional acid treatment step was introduced in order to remove acid labile DMAPP. Therefore, further measurement error was introduced and exacerbated the model accuracy. Moreover, the substrates, products and cofactors of the corresponding enzymes, highlighted in blue in Table 4.2, were excluded from further analysis. Finally, among the remaining elasticities, those with absolute values close to 1 were highlighted in red; they likely suggest interactions with biological significance [80]. The negative elasticity of DMAPP on Erg12 suggested an inhibitory effect. This was in agreement with literature [123],

lending confidence to the results. It further suggested stimulatory effect of IPP on both Erg12 and Erg8. Moreover, interestingly, a strong inhibitory effect of ATP on Ads was also suggested. These new relationships required further experimental validation in order to confirm whether it was a true biological interaction or purely minimization of the least square error [80].

Table 4.2 . Estimated elasticities at the optimal reference states.
The 95% confidence interval value is indicated in the bracket. The substrate, product and co-factor of the corresponding enzyme is colored in blue. The elasticities that might have biological significance are highlighted in red. The p-values have been calculated which all were less than 0.005.

Run 9	Enzymes					
Metabolite	Erg12	Erg8	Erg19	Idi	IspA	Ads
MVA	3.7 ± (0.67)	3.8 ± (0.12)	3.2 ± (3.61)	2.5 ± (5.30)	3.8 ± (7.93)	3.4 ± (2.15)
PMVA	1.1 ± (0.45)	2.8 ± (0.08)	2.0 ± (2.46)	2.0 ± (3.61)	3.1 ± (5.40)	2.1 ± (1.47)
PPMVA	-0.2 ± (0.07)	-0.1 ± (0.01)	-0.1 ± (0.370	0.1 ± (0.54)	0.0 ± (0.80)	-0.1 ± (0.22)
IPP	1.2 ± (0.30)**	1.1 ± (0.05)**	1.4 ± (1.61)	1.0 ± (2.36)	2.1 ± (3.53)	1.5 ± (0.96)
DMAPP	-0.5 ± (0.18)**	-0.3 ± (0.03)	-0.1 ± (0.95)	-0.3 ± (1.40)	0.2 ± (2.09)	-0.4 ± (0.57)
GPP	-0.3 ± (0.09)	-0.1 ± (0.02)	-0.1 ± (0.46)	-0.2 ± (0.68)	-0.4 ± (1.02)	-0.2 ± (0.28)
FPP	1.5 ± (0.20)	1.6 ± (0.04)	1.7 ± (1.11)	0.9 ± (1.63)	1.5 ± (2.43)	1.7 ± (0.66)
ATP	-0.4 ± (0.19)	-0.6 ± (0.04)	-0.7 ± (1.04)	-0.4 ± (1.53)	-0.6 ± (2.28)	-0.8 ± (0.62)**

4.2.2 Inhibition of Ads by ATP

To validate the results, we supplemented various metabolites: MVA (5 mM), IPP (1 mM), DMAPP (1 mM) and ATP (15mM) into the Ads reaction assay. A remarkable 10-fold reduction in Ads specific activity was only observed with ATP but not other metabolites (Figure 4.5A). Furthermore, to rule out the

possibility that the marked inhibitory effect of Ads was not a result of non-specific charge interaction between ATP and Mg^{2+}, an essential co-factor for Ads activity [142], a range of ATP concentration, from 0 mM to 15 mM was added into Ads reaction. More severe inhibition was observed when ATP concentration was above 10 mM. Nevertheless, with an addition of 0.2 mM ATP into Ads reaction, a concentration far below the Mg^{2+} concentration, a modest yet statistically significant ($p<0.05$) reduction of Ads specific activity was observed, as shown in Figure 4.5B. The apparent K_i value, estimated from Morrison Equation, was 6.4mM, which was much lower than the amount of Mg^{2+} supplemented in Ads reaction.

To further confirm the difference in the Ads activity was real, the unreacted substrate (FPP) concentration was also assayed by UPLC-(TOF)MS method (section 4.5.4). Figure 4.5B clearly depicted a gradual increase in the unreacted FPP concentration when the amount of ATP added into the reaction was increased. This increase could not be due to the matrix effect or ion suppression of ATP when assayed with UPLC-(TOF)MS, since ATP concentration was higher when FPP concentration was high, too. Therefore, it strongly suggested that Ads was inhibited by ATP in a specific manner.

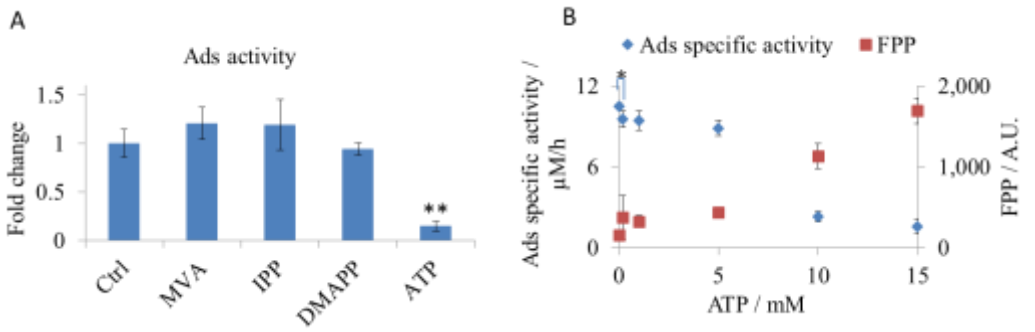

Figure 4.5. Validation of lin-log approximation for Ads reaction.
(A). Fold change in the specific Ads activity when different metabolites were added separately into the Ads reaction. It was obtained by dividing the control ads specific activity (Ctrl), where no additives were added. (B). Ads specific activity when a range of concentration of ATP from 0mM to 15mM was added. The amount of unreacted FPP were measured and its integrated area of LCMS measurement was used (please refer to section 4.5.4). All the measurement were done in triplicates, and their standard deviations were shown. Student t-test with paired two samples for means were performed to calculate the p-value in the statistical analysis.

4.2.3 Inhibition of Ads by Pyrophosphate

To further investigate which of the structure moiety of ATP was responsible for the inhibition of Ads, we used another three structurally similar compounds. ADP contains the adenosine and ribose moiety, and only differs from ATP with one phosphate group. When similar concentration of ADP was supplemented into Ads reaction, only 20% reduction in Ads activity was resulted (Figure 4.6A). Therefore, it was likely that the inhibition was attributed to the effect of the polyphosphate moiety. Therefore, to test the hypothesis, CTP and pyrophosphate (Ppi) were used, since both have the polyphosphate structure. Indeed, by assaying Ads activity with CTP or Ppi separately, a severe ~10-fold inhibition was seen (Figure 4.6A). This result strongly suggested that the polyphosphate tail moiety of ATP was responsible for the inhibition of Ads. Interestingly, pyrophosphate is the by-product of Ads. Hence, the result further

suggested a novel product inhibition of Ads. To confirm the results, we titrated the Ppi concentration from 0mM to 15mM, and assayed the rate of amorphadiene produced. An evident ~30% reduction in Ads specific activity was observed when 0.2 mM Ppi was added (Figure 4.6B). Its apparent K_i was 0.6mM, which implying Ppi as a more potent inhibitor of Ads activity. The unreacted FPP concentration accumulated as pyrophosphate concentration increased; it was another piece of evidence to support the idea that product inhibition of Ads was by Ppi (Figure 4.6B). Therefore, pyrophosphate is a more potent inhibitor of Ads than ATP, and it was a specific feedback regulation of Ads instead of a random depletion of available Mg^{2+}.

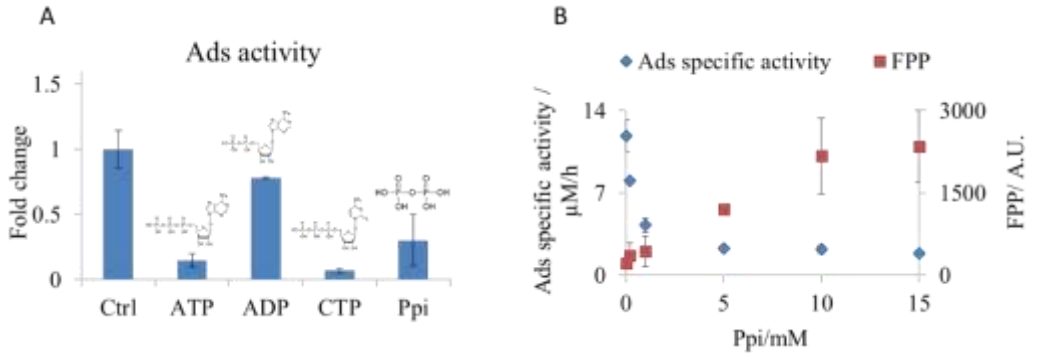

Figure 4.6. Ads activity assay with inhibitors.
(A). Fold change in Ads specific activity when it was screened with ATP (15 mM), ADP (15 mM), CTP (15 mM), Pyrophosphate (Ppi, 15 mM). It was calculated by dividing the control Ads activity, where no additives were added. (B). Ads specific activity when a range of concentration of Ppi from 0mM to 15mM was added. The amount of unreacted FPP were measured and its integrated area of LCMS measurement was used (please refer to section 4.5.4 for LCMS measurement). All the measurement were done in triplicates, and their standard deviations were shown.

4.2.4 Inhibition of the mevalonate pathway enzyme by pyrophosphate

Due to the structural similarity between ATP and pyrophosphate, we further examined whether pyrophosphate would have an inhibitory effect on the

mevalonate pathway enzymes, since they all of them use ATP as an essential co-factor. Figure 4.7 clearly showed that all the three enzymes, Erg12, Erg8 and Erg19, were inhibited with 10 mM Ppi. Interestingly, Erg12 and Erg8 were more sensitive towards the inhibition of pyrophosphate, that a remarkable 10 fold decrease in activity is seen. In contrast, Erg19 was less inhibited by the same concentration of pyrophosphate; a 2-fold decrease in activity was observed. Therefore, these observations demonstrated differential regulatory mechanisms by pyrophosphate on the ATP-dependent mevalonate pathway enzymes.

Lin-log approximation predicted the negative elasticities of DMAPP on Erg12 reaction. This was in agreement with previous report in the literature [123]. Moreover, it predicted a stimulatory effect of IPP on Erg12 and Erg8. In the separate experiment, we observed a more than 5-fold reduction in Erg12 activity when either 500 μM IPP or DMAPP was added into the reaction mixture (Figure 4.7). To our best knowledge, this was the first study showing an inhibitory effect of IPP on *S. cerevisiae* Erg12, although this was also seen for human Erg12 [122]. More strikingly, FPP was the most potent inhibitor of Erg12 with ~100 fold reduction in activity when only 15 μM of FPP was supplemented into the Erg12 reaction (Figure 4.7) [121]. Furthermore, no significant enhancement of Erg8 activity was observed with IPP, an observation that disagreed with Lin-log approximation. This was probably a false positive result predicted from the assumption of "maximum connectivity". There were little changes in Erg8 and Erg19 activities when downstream diphosphate compounds were added into the reaction medium respectively. A slight inhibition of FPP on Erg8 was observed,

but to a much lesser extent as compared to FPP inhibition on Erg12 (Figure 4.7). In contrast, FPP enhanced Erg19 activity mildly and statistically significant (Figure 4.7). The results further suggested different regulation mechanisms of the three enzymes although all of them use ATP as phosphate donor.

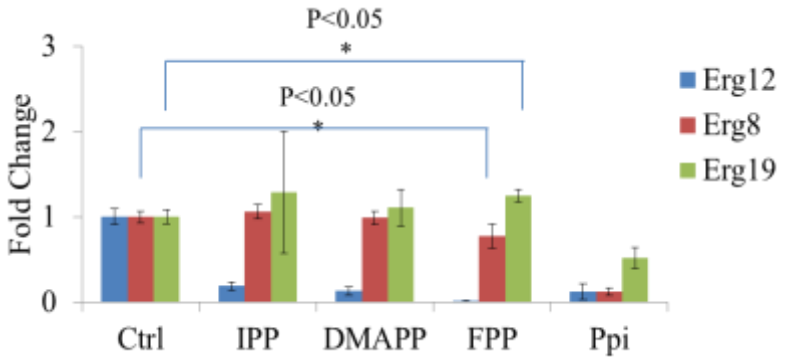

Figure 4.7. Mevalonate pathway enzyme activities in the presence of downstream metabolites.
Fold change in Erg12, Erg8 and Erg19 activities when downstream diphosphate products, IPP (500μM), DMAPP (500 μM), FPP (150 μM), and Ppi (10 mM), were added into their reaction medium respectively. It was calculated by dividing its respective control activity, where no additives were added into the reaction medium. All the measurements were done in triplicates, and their standard deviations were shown. Student t-Test with paired two sample for mean is used to calculate the p-value in statistical analysis.

4.2.5 Enhance AD production with ATP recycling and Pyrophosphatase

In Chapter 3, we have identified Ads as the critical factor that would improve the overall productivity of the multienzyme pathway when its activity was increased. By discovering the novel regulations of Ads by ATP and pyrophosphate, we further examined if the productivity of the multienzyme pathway might increase if both ATP and pyrophosphate concentrations could be kept low. Therefore, two additional enzymes were introduced in the system, as shown in Figure 4.8.

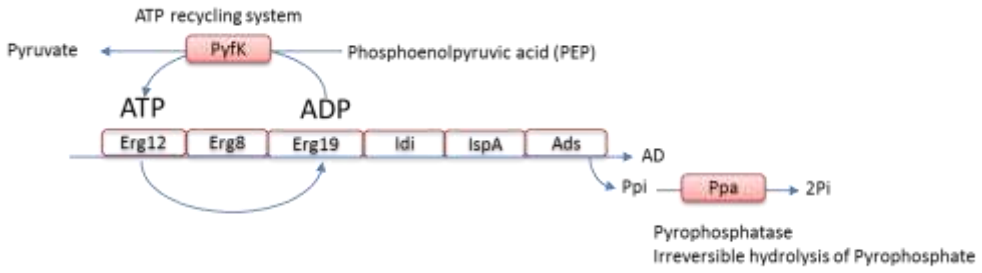

Figure 4.8. Schematic representation of PyfK and Ppa introduced in addition to the amorphadiene (AD) synthesis pathway.
PyfK is pyruvate kinase which can use PEP as phosphate donor and recycle ADP back into ATP; Ppa is pyrophosphatase that can irreversibly hydrolyze pyrophosphate (Ppi) into inorganic phosphate (Pi).

Firstly, pyruvate kinase (PyfK) was the commonly used enzyme that uses phosphoenolpyruvic acid (PEP) as the phosphate donor that recycle ADP back to ATP [182]. With ATP recycling system, catalytic amount of ATP would be needed to ensure that all the substrate would be converted. As a proof of concept, we first tested the hypothesis with 5mM ATP, which was lower than the apparent K_i value. In the future, more robust ATP recycling system could be devised to further reduce the initial amount of ATP. The second enzyme was inorganic pyrophosphatase (Ppa), which could hydrolyze pyrophosphate to inorganic phosphate [55, 183]. Therefore, the toxic by-product could be removed. The enzymes were cloned from *E. coli* strain Mg1655 genomic DNA and purified by immobilized metal affinity chromatography (IMAC) to near homogeneity. Their enzymatic activities have been verified separately (results not shown).

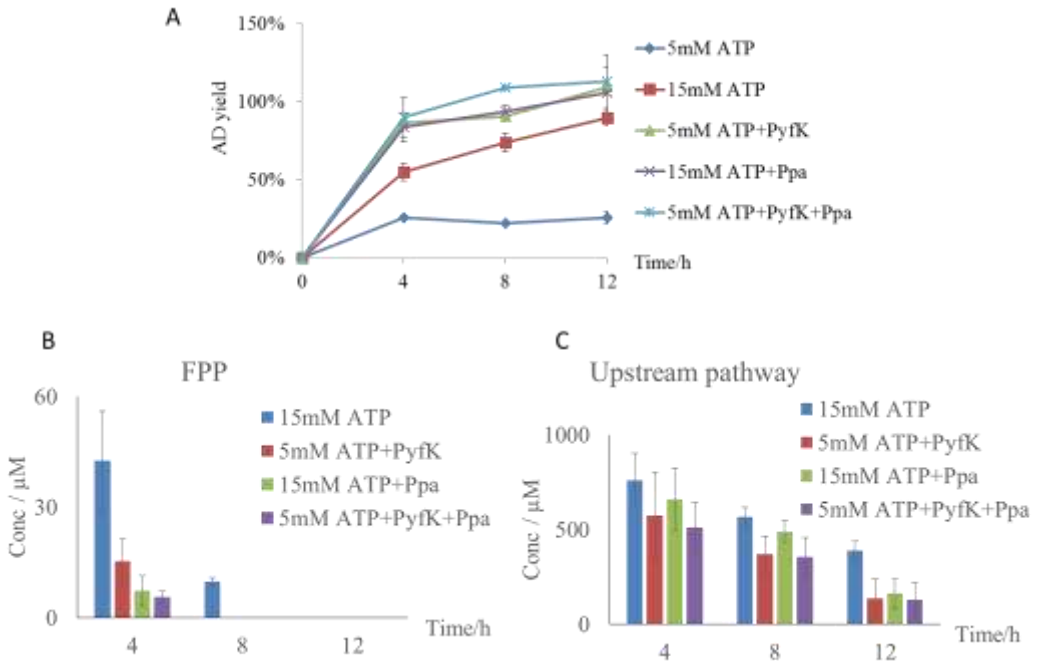

Figure 4.9. Increasing the *in vitro* multienzyme production of AD by pyruvate kinase and pyrophosphatase.
(A). Time course of *in vitro* multienzyme production of AD. PyfK: pyruvate kinase. Ppa: pyrophosphatase. (B). FPP concentration at different time point. (C). The total sum of upstream metabolite concentration at each time point. It was calculated by adding the concentrations of mevalonic acid (MVA), phosphomevalonic acid (PMVA), diphosphomevalonic acid (PPMVA), Dimethylallyl pyrophosphate (DMAPP) and Isopentenyl pyrophosphate (IPP). All the measurements were done in triplicates and the standard deviation was shown.

Figure 4.9A showed the time course of AD production under 5 different conditions. Evidently, more than 90% of MVA was gradually converted to amorphadiene (AD) after 12 h incubation, when the reaction contains 15mM ATP. When ATP concentration was lowered to 5mM, it was not sufficient to convert all the mevalonic acid (MVA) to AD. Upon the addition of PyfK which recycles ADP back to ATP, there was a remarkable 3-fold kinetic enhancement of the multienzyme synthesis system that similar level of conversion (~90%) achieved in 4 hours. Unexpectedly, when Ppa which hydrolyzes pyrophosphate was mixed with the multienzyme reaction containing 15 mM ATP, a 3-fold

kinetic enhancement of the production was also observed. This observation suggested that Ppa was able to rescue the system from ATP inhibition. Lastly, we introduced both PykF and Ppa into the multienzyme synthesis reaction. An apparent, albeit mild, improvement in the overall yield of multienzyme synthesis was observed as compared to that with either one of the enzymes added.

To gain a better understanding, concentrations of intermediate metabolite were assayed at each time point and the four conditions which resulted >90% AD yield were analyzed in Figure 4.9. As expected, there was an accumulation of FPP (~45µM), substrate of Ads, at 4h and 8h when only 15mM of ATP was present as compared to other three conditions that PyfK and/or Ppa were added (Figure 4.9B). Moreover, the total amount of upstream metabolites (MVA+PMVA+PPMVA+IPP+DMAPP) left unreacted at 4h and 8h were at similar levels for all four conditions (Figure 4.9C), indicating that the higher concentration of FPP observed was not due to the faster reaction of the upstream pathway. This further confirmed that Ads activity was suppressed when high concentration of ATP and pyrophosphate were present. Moreover, it further suggested that the ATP recycle system was successfully, that the ATP-dependent upstream pathway was not limited despite the lower amount of ATP added. In contrast, accumulation of upstream metabolites was observed when only 5mM ATP was added into the reaction (results not shown). The interesting observation of Ppa being able to reverse the ATP inhibition might be explained by alleviating the inhibition effect of pyrophosphate on the upstream pathway enzyme so that they consumed ATP in a more efficient manner. This was supported by the

observation, that with Ppa, there was a slight decrease in upstream metabolites concentration as compared to that when Ppa was not added (Figure 4.9C). However, this hypothesis remained to be tested. Lastly, the synergistic effects of PykF and Ppa were reflected in further reductions in the FPP concentrations, especially at 4h (Figure 4.9B).

4.2.6 Enzyme stability in a multienzyme reaction pot

The experimental design used for estimating elasticities could be used to identify the least stable in the multienzyme pathway. This allowed us to access the enzymatic stability from a system perspective, since the presence of multiple enzymes may have a crowding effect. By pre-incubating the multienzymes at 30^{o}C for a known period of time, the least stable enzyme should have its substrate accumulated. Evidently, Figure 4.10A shown that phosphomevalonate (PMVA), the substrate of Erg8, was accumulated when the pre-incubation time was increased. This suggested that Erg8 could be the least stable or rate-limiting enzyme among the 6 enzymes assembled. The decrease in the subsequent intermediate levels could be a result of deactivation of Erg8, as less substrates to the downstream enzyme were produced. In order to test the stability of Erg8 in the reaction condition, we conducted a separate experiment by pre-incubating the enzyme alone under the same reaction conditions and found its half-life to be ~50 min (Figure 4.10B). When 20% glycerol was added, its activity was maintained for more than 1h (results not shown). Thus, this enzyme could be the next target to engineer so as to enhance its stability under *in vitro* reaction conditions.

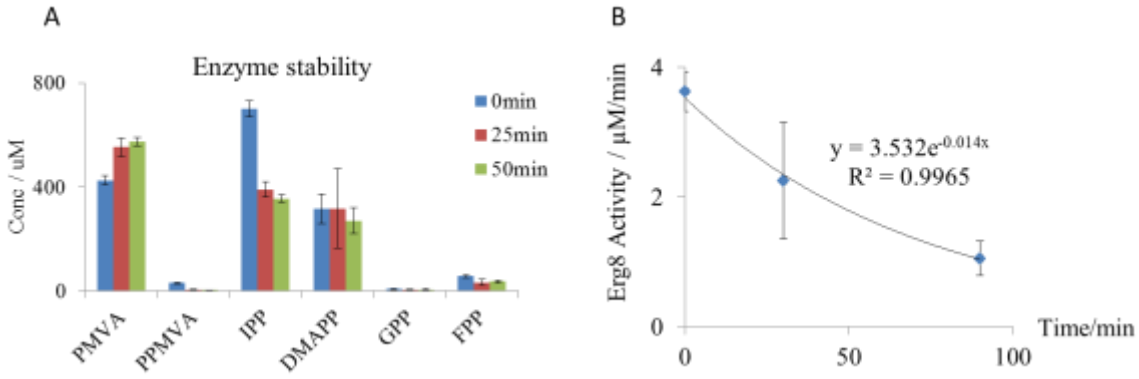

Figure 4.10. Analysis of the pathway enzyme stability.
(A). The change in the intermediate metabolite concentrations when the multienzyme was incubated at 30°C for 0min, 25min and 50min. (B). The activity decay of Erg8 after incubating at 30°C for a period of time. The half-life of the enzyme calculated based on the fitted exponential curve is 50min. All the measurements were done in triplicates and the standard deviations were shown.

4.3 Discussion

To better engineer and optimize a biochemical pathway, both *in vivo* and *in vitro*, understanding the network structure and the interplay among individual pathway components is paramount. However, by conventional trial-and-error approach would be time-consuming and ineffective to reach an optimal solution. The beauty of mathematical modelling is to reduce the experimental efforts and efficiently identify the likely optimal range. An ideal mathematical model for a biochemical pathway would be able to capture the mechanistic information of the multienzyme system, and allow the biologists to analyse and interpret the biological meaning of the parameters estimated. To obtain such model requires undue experimental and computation efforts, since many examples have shown that even kinetics information was abound for individual enzyme but would not result in a satisfactory mathematical representation of the pathway, and

furthermore some information may be contradictory to each other [184]. Recent developed Lin-log formalism reduces the model complexity and permits identifying the biochemical structure around a reference state. It is an extension of metabolite flux analysis method that allows large perturbations around a reference state [79], and was sufficiently capable to unravel the regulatory behaviour of an *in vitro* system [184]. Therefore, we harnessed the benefit of the modelling technology which enabled the identification of a novel inhibitory effect of ATP on Ads.

The inhibition of ATP was more apparent when its concentration was very high. Therefore, when multienzyme system was initiated, high ATP concentration would inhibit Ads activity, resulting in little or no amorphadiene conversion at the start, even though there was already accumulation of FPP at higher concentration (results not shown). This may be detrimental to the whole multienzyme system, since we reported in the previous chapter that accumulation of FPP would lead to precipitation and prevent efficient catalysis of the pathway enzymes. There are three hypotheses that could account for the inhibition of ATP on Ads. Firstly, the phosphate of ATP is complexed with Mg^{2+}, and hence depleting the availability of divalent ions for Ads catalysis. Secondly, there may be a structural similarity between FPP and ATP, since FPP was identified to competitively occupy the ATP binding site of several enzymes including mevalonate kinase (Erg12) [120]. Moreover, there was a glycine rich region identified near the catalytic site of Ads. The glycine cage has been reported to bind the nucleotide portion of ATP in the literature [185]. This might provide a docking site for ATP and prevent the entry

of FPP. Lastly, ATP entails a pyrophosphate moiety, which is a by-product of Ads. From our results, it was likely that the polyphosphate moiety of ATP was responsible for the inhibition effect, since ADP containing the same nucleotide group did not elicit significant inhibition. Hence, a novel feedback regulation of Ads by its product pyrophosphate was identified. Ads belongs to the important class of enzyme, terpene synthase, that is responsible for the diversity of terpenoids [12]. Hence it might be interesting to examine if the inhibitory effect was a universal phenomenon to terpene synthases, in particular, sesquiterpene synthases.

In addition to the novel finding of an inhibitory effect of ATP and pyrophosphate on Ads, regulation of Erg12 by downstream metabolites including IPP, DMAPP, GPP and FPP were also observed [123]. To our best knowledge, we were the first to observe the novel inhibition of *S.cerevisae* Erg12 by IPP, similar to that reported for the human form of Erg12 [122]. We attempted to examine the mechanism of IPP inhibition on Erg12, and found both IPP and DMAPP were uncompetitive inhibitors, determined by lineweaver burk plot, to the substrate mevalonic acid (MVA); they were possibly able to compete with ATP binding pocket and reduce the enzyme efficacy. *S. cerevisea* Erg8 activity was reported to be inhibited by high concentration of ATP [126]. This was effectively reflected by the lin-log approximation. However no inhibition of downstream metabolites, albeit minute inhibition of FPP, was observed for Erg8 catalyzed reaction. The false positive results drawn from Lin-log approximation could be due to the assumption made by “maximum connectivity”. Further

optimization of Lin-log approximation could be carried out with a priori information to strengthen the predictability of the model.

We further examined the inhibition of pyrophosphate (Ppi) on the ATP-dependent mevalonate pathway enzymes. All the three enzymes were affected by Ppi but the extent of inhibitioin varied among the three enzymes. This interesting observations reiterated the diversity of enzymes, and the uniqueness of ATP binding pocket that each enzyme entails that makes the prediction of inhibition less obvious.

By unraveling the regulation of the amorphadiene synthesis pathway, we were able to rationally engineer the multienzyme pathway with two additional enzymes: PyfK that recycles ATP and Ppa that hydrolyze pyrophosphate. This further enhanced the kinetics of the multienzyme system by 3 fold. We have done a theoretical calculation that if Ads is the rate limiting step, which means the production of AD is solely dependent on the activity of Ads, to catalyze all 5mM mevalonic acid to AD, it would only require 3-4h. (Refer to the previous chapter for Ads activity, which is $0.05s^{-1}$, and approximately 3μM of Ads is added). Hence, our results demonstrated that, with PyfK and Ppa, the multienzyme system can reach ~90% conversion at 4h, and hence is kinetically optimal. Further optimization may require engineering Ads to alter its catalytic efficiency.

Another application of the assembled pathway has enabled us to identify the least stable enzyme along the pathway. The stability of enzymes may change when its environment changes. When multiple enzymes are present, some enzymes may lose its activity faster due to aggregation, or a stabilizing effect may

be resulted due to crowding effect. Therefore, by assembling the pathway together and measure its steady state metabolite profile change may provide a more accurate depiction of how enzyme stability may be affected in a multienzyme condition. With that, we have identified the second enzyme, Erg8, as likely to be the least stable enzyme, and should be the target of further optimization.

4.4 Conclusion

In summary, we have identified a novel inhibition of ATP on Ads, and it was due to the presence of polyphosphate moiety in ATP, that is structurally similar to Ads byproduct, pyrophosphate. It was further confirmed that pyrophosphate is a potent product inhibitor of Ads. Moreover, high concentration of pyrophosphate would inhibit the ATP-dependent mevalonate pathway enzymes, although with different extent of inhibition. After unravelling the regulation of the amorphadiene synthesis pathway, we introduced the ATP recycling system and pyrophosphatase into the multienzyme system, and was able to enhance the time yield of the system by 3-fold. According to our theoretical calculation, the system was kinetically optimized. The study emphasized the importance of understanding the pathway regulations and the advantages of *in vitro* multienzyme system. With which, we were able to rationally design targeted engineering strategies to overcome the challenge, and enhance the productivity. We believe the method employed here may be useful for other *in vitro* reconstituted system beyond the scope of this study.

4.5 Materials and Methods

4.5.1 Bacteria strains and plasmids

The bacteria strains and plasmids for the amorphadiene synthesis pathway enzymes were the same as Chapter 3. The Bacteria strains and additional plasmids used in this chapter were summarized in Table 4.3. The cloning of the two additional plasmids were as described [186]. Briefly, Pyruvate kiase (PyfK, EC 2.7.1.40) and inorganic pyrophosphatase (Ppa, EC 3.6.1.1) were amplified from *E. coli* MG1655 strain genomic DNA. The vector backbone was amplified from pBAD-B plasmid (Invitrogen, CA). The primers used for amplification of the genes were listed in Table 4.4. The PCR products were treated by 0.3% Iodine at 70^{o}C for 5min to chemically cleave the phosphorothioate bond. Then ethanol precipitation step was carried out to purify and concentrate the treated DNA. The insert and vector were mixed at approximately 1:1 molar ratio, and then annealed with heating at 80^{o}C for 5min and gradually cooling down to 20^{o}C with a ramp rate of 0.1^{o}C/s. Then 1µl of mixture with final DNA concentration of ~100ng/ µl was transformed into *E. coli* XL10-gold (Stratagene, CA) to self ligate the insert and vector together. The colonies formed were examined by colony PCR with forward primer targeting the vector and the reverse primer targeting the gene of interest. The colony harbouring the correct plasmid was then grown till optical density reached 2 and then was harvested to purify the plasmids. The plasmids were then transformed to respectively strains for enzyme overexpression (Table 4.3).

Table 4.3. The bacteria strain and plasmids used in Chapter 4.

Name	Description	Reference
E. coli BL21-Gold (DE3)	F^- ompT hsdS (r_B^- m_B^-) dcm^+ Tet^r gal λ(DE3) endA Hte	Stratagene
E. Coli DH10B	araD139 Δ(ara-leu)7697 fhuA lacX74 galK (Φ80 Δ(lacZ)M15) mcrA galU recA1 endA1 nupG rpsL Δ(mrr-hsdRMS-mcrBC)	NEB
E. Coli XL10-Gold	Tetr D(mcrA)183 D(mcrCB-hsdSMRmrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F9 proAB lacIqZDM15 Tn10 (Tetr) Tn5 (Kanr) Amy	Stratagene
pBAD-His-PyfK	Plasmid for overexpression of PyfK in *E. coli* DH10B	This study
pBAD-His-Ppa	Plasmid for overexpression of Ppa in *E. coli* DH10B	This study

Table 4.4. The CLIVA primers used in Chapter 4. * indicated the phosphorothioate modification.

Primer Name	Sequnce
I-pBADT(-)_F	TGGCGGATGAGAGAAG*ATT
I-pBADhis(-)_Rn	GTGATGGTGATGGTGA*TGA
I-PykF(His)_f	TCACCAT*CACCATCAC*ATGAAAAAGACCAAAATTGTTT
I-PykF(BADt)_r	CTTCTCTC*ATCCGCCA*TTACAGGACGTGAACAGATGC
I-Ppa(His)_f	TCACCAT*CACCATCAC*ATGAGCTTACTCAACGTCCC
I-Ppa(BADt)_r	CTTCTCTC*ATCCGCCA*TTATTTATTCTTTGCGCGC

4.5.2 Expression and purification of PyfK and Ppa.

Purification of the amorphadiene synthesis pathway enzymes were the same as described in Chapter 3. Expression and purification of the two additional enzymes were described as follows. Newly transformed colonies were picked from the agar plate, inoculated into 2xPY medium (20 g/L Peptone, 10 g/L Yeast extract, and 10 g/L NaCl, pH 7) containing 100 mg/L ampicillin and grew till

stationary phase overnight at 37 ^{o}C in an incubator-shaker (Shin Saeng Shaking Incubator, Finetech, Korea). The culture was then further transferred into fresh 2xPY medium (1% inoculation) with ampicillin for another 2.5 h at 37 ^{o}C, till optical density A_{600} reached 0.6-1.0. The enzyme expression was induced with 10 mM L-arabinose. Temperature was reduced to 20 ^{o}C after induction for higher solubility of the enzymes [20]. The culture was grown for another 48 h and harvested by centrifugation. The cell pellets were stored at -20 ^{o}C till further use.

To purify the enzymes, the frozen cell pellets were resuspended in B-PERII reagent (Pierce, IL), according to the manufacturer's instruction, and vortexed at room temperature for 30 mins to completely lyse the cells. The soluble proteins were contained in the supernatant, which was diluted 15 times in NPI10 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) and incubated with 200 mg Ni-NTA resin (USB, Affymetrix, CA) at 4 ^{o}C for 2 h. The resin was washed with NPI10 buffer after discarding the binding supernatant, and the enzymes were eluted and collected by 400 µl NPI400 (50 mM NaH_2PO_4, 300 mM NaCl, 400 mM imidazole, pH 8). The enzymes were further concentrated by 3K Amicon ultra-0.5 ml centrifugal filter unit (Millipore, MA), and the concentrations were measured by Micro BCA protein assay kit (Thermo scientific, MA). The purified enzymes were further confirmed by sodium dodecyl sulfate-12% polyacrylamide gel electrophoresis (Bio-Rad, CA).

4.5.3 Multienzyme reaction

The multienzyme reaction was carried out in a buffer (25 µl) consisted of Tris/HCl (100 mM, pH 7.4), $MgCl_2$ (10 mM), (±)mevalonic acid (10 mM), ATP and the purified enzymes. If not indicated otherwise, the amount of amorphadiene pathway enzymes were added according to the optimum enzyme levels found in Chapter 3. PyfK and Ppa were both added in excess at the final concentration of 50mg/L. The reaction was performed at 30 ^{o}C with an overlay of dodecane phase that contained known concentration of trans-caryophyllene as an internal standard. At the end of the reaction, the dodecane phase was diluted 10 times in ethyl acetate and subject to GCMS analysis. Equal volume of ice cold methanol was added to stop the reaction. Half of the stopped reaction mixture was further diluted with MeOH and 0.1% ammonium hydroxide, and injected into UPLC column for measurements. Another half of the stopped reaction mixture was used to determine the concentrations of IPP and DMAPP. In order to differentiate the isomer IPP and DMAPP, an acid treatment step was carried out to degrade the acid labile DMAPP [187]. In brief, equal volume of 2M HCl was mixed with stopped reaction mixture, and then incubated at 37^{o}C for 10min. Then the reaction was stopped with 1% acetic acid. It was further diluted in methanol before UPLC-(TOF)MS analysis. To determine the enzyme stability, the enzymes were first mixed together according to the concentration of the optimal steady state determined for Lin-log approximation (run 9). Then they were incubated at 28^{o}C for 0min, 25min and 50min respectively, before the mixture of substrate and co-factors were added to initiate the reaction. The (±)mevalonic acid was prepared

by complete alkaline hydrolysis of 2 M (±)mevalonolactone (Sigma, MO) with equal volume of 2 M potassium hydroxide at 37 ^{o}C for 1.5 h, and neutralized by adding 1 M hydrochloric acid to pH 7 [170].

4.5.4 UPLC-(TOF)MS analysis of mevalonate pathway intermediates

The analysis was done based on the method developed previously with slight modification [42]. In brief, 5 µl samples were injected into a UPLC C18 column (Waters CSH C18 1.7 µm, 2.1 mm x 50 mm) connected to UPLC (Waters ACQUITY UPLC)-(TOF)MS (Bruker micrOTOF II, MA). Elution was carried out with a step change from 100% aqueous solution containing 15 mM acetic acid and 10 mM tributylamine (0.5 min) to 10% aqueous solution with 90% methanol for another 3.5 min. Electrospray ionization was used and mass spectrometry was operated to scan 50–800 m/z in negative mode with 2500 V end plate voltage and 3200 V capillary voltage. Nebulizer gas was provided in 2 bar, dry gas flow rate was 9 ml/min, and dry gas temperature was 200 ^{o}C. At the assay condition, all the intermediates were detected in the form $[M-H]^-$. Retention time was subsequently determined for each intermediate with respective synthetic standards and the set m/z extraction range. The peak area was calculated and subsequently used to compute the intermediates concentrations with the software provided by the manufacturer. The calibration curves were constructed with synthetic standards prepared under similar reaction conditions without enzymes. Linearity of the assays were determined individually with coefficients of determinants (R^2) greater than 0.90.

4.5.5 GCMS analysis of Fanesyl pyrophosphate (FPP) and amorpha-4,11-diene (AD)

To measure the FPP concentration, a hydrolysis reaction was carried with acid phosphatase (Sigma, MO) according to the manufacturer's instruction. In short, FPP was incubated with acid phosphatase at 37^{o}C for 10min, and was stopped by extracting the product farnesol (FOH) with 3 volumes of ethyl acetate. The standard curve of FPP was constructed with synthetic standards that known concentration of synthetic FPP was treat with acid phosphatase and then subjected into GCMS analysis. AD analysis was carried out based on the modified method developed by Martin et al. by scanning three ions; the m/z values are 117, 189 and 204 [170, 171]. For both FPP and AD, 1 µl sample was injected into HP-5 column (Agilent Technologies 7890A gas chromatograph-mass spectrometry, Agilent, CA) with a linear temperature increase of 50 ^{o}C/min from 80 ^{o}C to 300 ^{o}C and hold at 300 ^{o}C for another minute. The peak area was calculated and subsequently used to compute the FPP and AD concentrations with the software provided by the manufacturer. FPP concentration was determined based on the external standard curved constructed. Amorpha-4,11-diene concentrations were determined relative to the internal standard trans-caryophyllene of known concentration.

4.5.6 Lin-log modelling

The lin-log modelling was done based on the previous study demonstrated by Wu et al [80]. It was modified from metabolic control analysis (MCA). The important MCA parameters and relations for linear pathway were summarized in Table 4.5 and Table 4.6.

Table 4.5. Definition of MCA parameters [188].

Elasticity with respect to x_m	$\varepsilon_{i,m}^{x0} = \frac{x_m{}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial x_m}\right)^o$
Elasticity with respect to e_j	$\varepsilon_{i,j}^{e0} = \frac{e_j{}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial e_j}\right)^o$
Elasticity with respect to c_r	$\varepsilon_{i,r}^{c0} = \frac{c_r{}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial c_r}\right)^o$
Flux control coefficient	$C_{i,j}^{v0} = \frac{e_j{}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial e_j}\right)^o$
Metabolite control coefficient	$C_{m,j}^{x0} = \frac{e_j{}^o}{x_k^o} \cdot \left(\frac{\partial x_k}{\partial e_j}\right)^o$
Flux response coefficient	$R_{i,r}^{v0} = \frac{c_r{}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial c_r}\right)^o$
Metabolite response coefficient	$R_{m,r}^{x0} = \frac{c_r{}^o}{x_k^o} \cdot \left(\frac{\partial x_k}{\partial c_r}\right)^o$

Table 4.6. MCA relations for a linear pathway. [80]
All the equation below are for matrix operations. J^o is the steady state flux. S is the stoichiometric matrix for the network of enzymatic reaction. E^{x0} and E^{c0} are the elasticity matrix for dependent metabolite and independent metabolite respectively. *i* is a unit vector.

Metabolite control coefficient	$C^{x0} = -(S[J^o]E^{x0})^{-1} \cdot S[J^o]$
Flux control coefficient	$C^{J0} = E^{x0}C^{x0} + i$
Metabolite response coefficient	$R^{x0} = C^{x0}E^{c0}$
Flux response coefficient	$R^{J0} = C^{J0}E^{c0}$

An important parameter in MCA is elasticity (Equation 4.1), which reflects the kinetic property of individual enzyme towards a dependent metabolite concentration (x_i) or an independent metabolite concentration (c_i).

$$\varepsilon_{i,m}^{x0} = \frac{{x_m}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial x_m}\right)^o \qquad \varepsilon_{i,r}^{c0} = \frac{{c_r}^o}{v_i^o} \cdot \left(\frac{\partial v_i}{\partial c_r}\right)^o$$

Equation 4.1

In the Lin-log approximation, the nonlinear relationship between reaction rate and the reaction components, namely the enzymes and the metabolites were shown in Equation 4.2, where the reaction rate v_i was assumed to be proportional to the enzyme level e_i and a linear sum of logarithms of m dependent (x) and r independent (c) metabolite concentrations. $p_{i,m}$ and $q_{i,r}$ are estimated coefficients for mth dependent metabolite and rth independt metabolite respectively for ith enzymatic reaction.

$$v_i = e_i(a_i + \sum p_{i,m} ln x_m + \sum q_{i,r} ln c_r)$$

Equation 4.2

The rationale behind the logarithmic concentration terms has its origin in the concept that the rate of a chemical reaction is proportional to the reaction

affinity when the reaction is close to equilibrium [189]. The affinity term is linearly proportional to the logarithmic term of the reactant concentration.

A reference state can be introduced in Equation 4.2, as shown in Equation 4.3 that the reaction rate term (v_i^o), enzymes (e_i^o), dependent (x_m^o) and independent (c_r^o) metabolite concentrations were all modified with the "o" symble to show they are measured at steady state.

$$v_i^o = e_i^o(a_i + \sum p_{i,m} ln x_m^0 + \sum q_{i,r} ln c_r^o)$$

Equation 4.3

Therefore, by the definition of elasticity (Equation 4.1) and the reaction rate (Equation 4.3), it can be expressed in terms of enzymes and the coefficients at steady state (Equation 4.4).

$$\varepsilon_{i,m}^{x0} = p_{i,m} \frac{e_i^o}{v_i^o} \qquad \varepsilon_{i,r}^{c0} = q_{i,r} \frac{e_i^o}{v_i^o}$$

Equation 4.4

Combining Equation 4.2, Equation 4.3, and Equation 4.4, the coefficients a_i, p_i and q_i can be eliminated and simplified into the following Equation 4.5.

$$\frac{v_i}{v_i^o} = \frac{e_i}{e_i^o} (1 + \sum \varepsilon_{i,m}^{x0} ln \frac{x_m}{x_m^o} + \sum \varepsilon_{i,r}^{c0} ln \frac{c_r}{c_r^o})$$

Equation 4.5

When a multienzymatic reaction system is considered, Equation 4.5 can be written in a matrix form.

$$\frac{\mathbf{v}}{\mathbf{v^0}} = \left[\frac{\mathbf{e}}{\mathbf{e^0}}\right] \cdot (\mathbf{i} + \mathbf{E^{x0}} \cdot \ln(\frac{\mathbf{x}}{\mathbf{x^0}}) + \mathbf{E^{c0}} \cdot \ln(\frac{\mathbf{c}}{\mathbf{c^0}}))$$

Equation 4.6

Therefore, technically, elasticities ($\mathbf{E^{xo}}$, $\mathbf{E^{co}}$) can be estimated by linear regression when measurements were obtained with perturbations of enzymatic levels ($\mathbf{e/e^0}$), metabolite concentrations ($\mathbf{x/x^0}$, $\mathbf{c/c^0}$) and reaction rate ($\mathbf{v/v^0}$) around a steady state. After obtaining the elasticities, the control coefficients listed in Table 4.5 can be calculated based on the relationship given in Table 4.6. With the control coefficients, Lin-log approximation could give an analytical solution of the steady state mass balances of the metabolic network; the steady state flux (Equation 4.7) and dependent metabolite concentrations (Equation 4.8) could then be re-calculated by enzymatic levels and independent metabolite concentrations.

$$C^{J0} \cdot \left(\frac{v}{v^o} \cdot \frac{e^0}{e}\right) = i + R^{J0} \cdot \ln\left(\frac{c}{c^0}\right)$$

Equation 4.7

$$ln\left(\frac{x}{x^0}\right) = -C^{xo} \cdot \left(\frac{v}{v^0} \cdot \frac{e^0}{e}\right) + R^{x0} \cdot ln\left(\frac{c}{c^0}\right)$$

Equation 4.8

To determine the optimal steady state, the criteria of adjusted coefficient of determinant, aR^2, is used, which is calculated as follows, where SSE is the sum of square error and SST is the sum of square total, N is the number of observations and P is the number of parameters to be estimated.

$$aR^2 = 1 - \frac{SSE/(N-P)}{SST/N}$$

Equation 4.9

Chapter 5 Co-immobilization of Multienzymes for amorpha-4,11-diene synthesis

5.1 Introduction

In Chapter 4, the amorphadiene synthesis pathway was further optimized to achieve near theoretical productivity. However, free enzymes usually suffer from low enzymatic stability and require cost- and time- consuming purification processes. Immobilization of the enzymes is an attractive biorenewable approach offering advantages in enhanced stability, volume specific biocatalyst loading, recyclability, simplified downstream processing and is often the key to optimizing the operational performance of an enzyme in industrial processes [86, 190]. Many parameters contribute to the success of immobilized enzyme technology and continual efforts to optimize attachment chemistries of the enzyme onto compatible surfaces combined with modern analytical tools to improve activity and stability has gained recent attention [191]. As a rule of thumb, the enzyme cost in commercial applications should not exceed more than a few percent of the total production costs [86]. For example, 600kg of 6-aminopenicillanic acid (6-APA) is produced for every kg of immobilized penicillin G amidase, and the specific yield is 11,000kg/kg enzyme for food production [192]. Therefore operational stability of the biocatalysts is critical to ensure the economic viability of the process. Immobilization is able to enhance the enzymatic stability [193, 194]; aggregation due to random Brownian motion of the enzymes was reduced. Moreover, the enzymes are easily separated from the products and reused continuously for subsequent cycles of catalysis [195]. Therefore, the specific yield is enhanced. In addition, immobilization would be able to concentrate the

otherwise diluted enzymes [196]. Hence kinetic efficiency might be enhanced as a consequence of substrate channeling and close proximity of the catalytic sites [197, 198].

A myriad of exciting immobilization strategies have been developed, due to the synergistic growth of material research and biotechnology. As reviewed in chapter 2, cross-linked enzyme aggregates (CLEA), DNA-directed immobilization (DDI) are promising technologies to co-immobilize multienzymes in an elegant way. However, the methods lacked the control over the orientation of the enzymes, which has been shown to be critical to enzymatic activities at interface [99, 100, 195]. Therefore, adsorption via site-directed interaction between the insoluble support and the enzymes is attractive, and it is simple, cheap and causes no or little damage to enzyme [199]. Among them, immobilized metal affinity chromatography (IMAC) technology is a facile approach of immobilizing enzymes. Enzymes can be easily engineered with polyhistidine-tag, and one-step protein purification and immobilization is achieved to avoid expensive protein purification [200, 201]. Moreover, enzymes immobilized via IMAC technology has minimal impact on enzymatic conformation and reflects solution phase enzymatic activities [99, 202].

In this chapter, the pathway enzymes for amorpha-4,11-diene biosynthesis were co-immobilized on Ni-NTA modified solid surface (Figure 5.1). Initial attempt to recycle the co-immobilized AD synthesis pathway did not yield satisfactory productivity and recyclability. This was likely attributed to the presence of high concentration of ATP and pyrophosphate, as identified in chapter

4. Therefore, when pyruvate kinase and inorganic pyrophosphatase were added to remove the inhibitors, both the rate of production and overall AD yield were improved by 3 and 2.5 fold respectively. Further guided by the regulatory topology of the multienzyme pathway, a rationally designed, novel bi-modular system was implemented, which a further 25% improvement in AD yield was observed and a near quantitative conversion of the substrate after 12h was achieved. The multienzymes can be effectively reused for more than 7 cycles of reaction. Taken together, approximately 2.2g/L of amorphadiene could be produced within 4 days, which is a greater than 6-fold enhancement of AD specific yield as compared to the free enzymatic system.

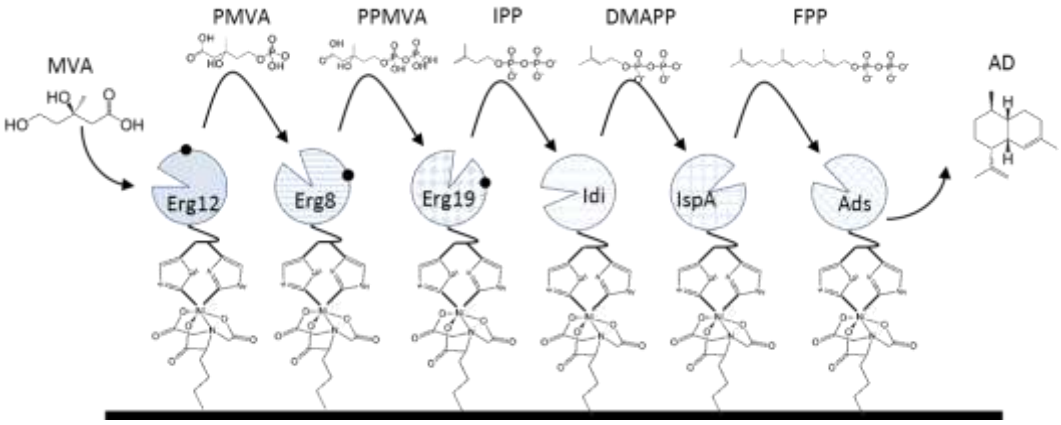

Figure 5.1. Schematic representation of the co-immobilized multienzymes.
The enzymes were arranged in tandem, and the intermediates were shown. The black circle in the graph represent the co-factor ATP. The abbreviations are as follows. Erg12: mevalonate kinase, Erg8: phosphomevalonate kinase, Erg19: diphosphomevalonate decarboxylase, Idi: isopentenyl pyrophosphate isomerase, IspA: farnesyl pyrophosphate synthase.Ads: amorpha-4,11-diene synthase. MVA: mevalonic acid. PMVA: phosphomevalonic acid. PPMVA: diphosphomevalonic acid. IPP: Isopentenyl pyrophosphate. DMAPP: Dimethylallyl pyrophosphate. FPP: Farnesyl pyrophosphate. AD: Amorpha4,11-diene.

5.2 Results

5.2.1 Immobilize enzyme on Ni-NTA functionalized beads

The overexpressed enzymes were engineered with hexahistidine tag (his_6-tag). Therefore, they can be conveniently purified and immobilized by immobilized metal affinity chromatography (IMCA) [98], which the surface of solid resin was functionalized with nickel-chelated nitrilotriacetic acid (Ni-NTA). Since the immobilization process is reversible, a facile dilution method was used to decrease concentration of imidazole and immobilize purified enzymes. For convenience, we used enhance green fluorescence protein (eGFP; Figure 5.2B) as a model protein to monitor the efficiency of the immobilization method. From Figure 5.2A, almost all the eGFP proteins were adsorbed onto the solid resin, when 5µg of protein was mixed with 1mg of resin (an optimal protein to resin ratio recommended by manufacturer). By measuring the residual fluorescence in the supernatant, the immobilization yield was determined to be greater than 90%. We have increased the protein to solid resin ratio twice, the immobilization yield was still above 90%, suggesting a high protein immobilization capability of the resin. For the subsequent experiments, the optimal enzyme to resin ratio (5ug:1mg) was used so that the surface was not saturated.

The dilution method was subsequently applied to immobilize the enzymes involved in the biosynthesis of amorphadiene (AD synthesis pathway). Mevalonate kinase (Erg12) is the first enzyme in the pathway, and is used as an exemplary enzyme to determine the immobilization yield and activity yield (section 5.5.2). From Figure 5.2C, more than 90% of Erg12 was immobilized

after mixing with the resin for 10 min at room temperature and the activity yield had reached a plateau at 10 min. There was a decrease in the immobilized Erg12 activity, approximately 30% activity remained after immobilization, possibly due to adverse interfacial phenomenon that has previously been reported [203]. Nevertheless, the enzyme was efficiently immobilized onto the Ni-NTA grafted solid support, and 10min mixing time will be used subsequently to co-immobilize the multiple enzymes.

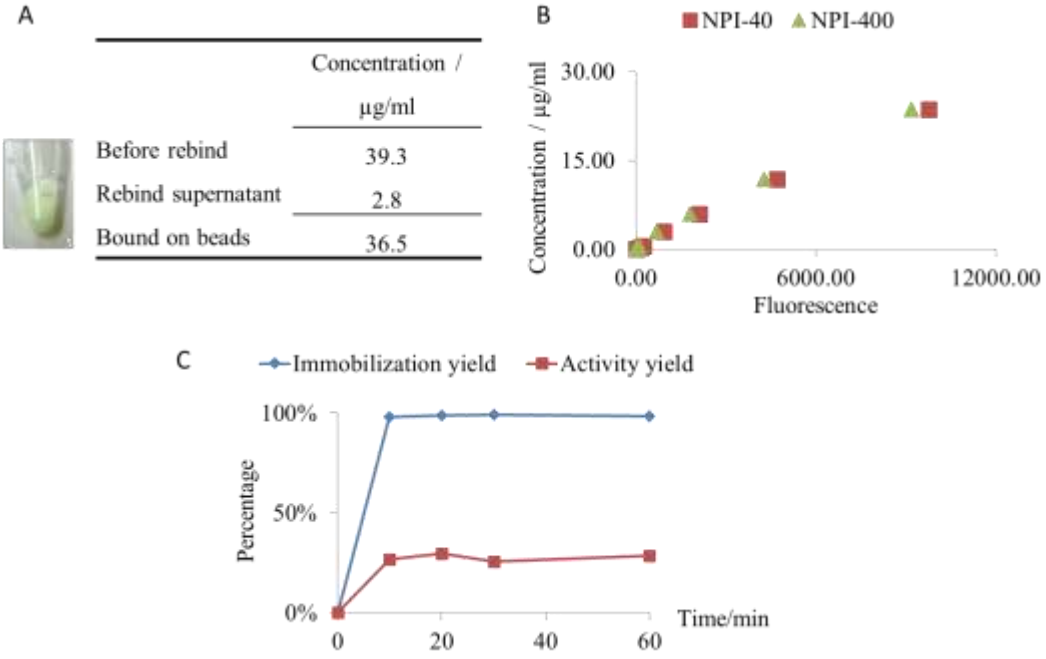

Figure 5.2. Immobilizing his_6-tag enzyme onto Ni-NTA functionalized solid resin. (A). Demonstration of immobilizing eGFP onto Ni-NTA resin by dilution method. The concentrations were determined by its fluorescence intensity. (B). The standard curve of eGFP concentration against its fluorescence intensity (Ex=490nm, Em=500nm). NPI-40 and NPI-400 were two buffers that contained sodium chloride (N), phosphate (P) and imidazole (I). 40 and 400 refer to 40mM and 400mM imidazole concentration. The standard curve showed a good linearity between the fluorescence and concentration of the eGFP protein. It was used to calculate the concentration of eGFP in the solution. (C). The immobilization and activity yield of Erg12. For definition of the two terms, please refer to materials and methods (Section 5.5.2).

5.2.2 Production of Amorpha-4,11-diene

The AD synthesis pathway enzymes were assembled and co-immobilized according to the enzymatic ratio identified previously (Chapter 3) for amorpha-4,11-diene (AD) production (Figure 5.3, system I). SDS PAGE analysis showed that majority of the enzymes were co-immobilized onto the solid support with greater than 90% yield, (Figure 5.4A, lane 3-4).

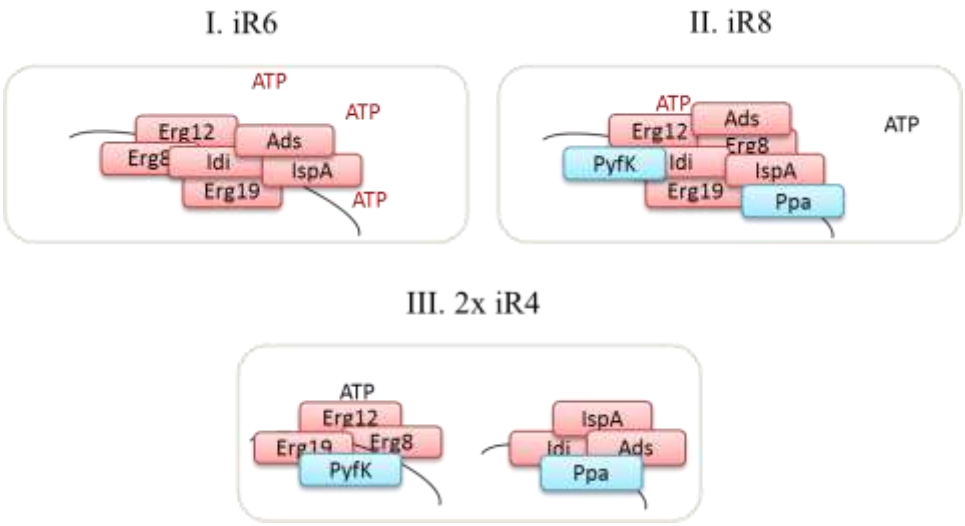

Figure 5.3. Schematic representation of the co-immobilized systems.
System I (iR6): randomly co-immobilize the 6 pathway enzymes on Ni-NTA grafted solid support. System II (iR8): randomly co-immobilize the 6 pathway enzymes together with pyruvate kinase (PyfK) and inorganic pyrophosphatase (Ppa) onto Ni-NTA grafted solid support. System III (2x iR4): bi-modular system that the 8 enzymes were divided into two groups: the ATP dependent mevalonate pathway enzymes were co-immobilized with PyfK that can recycle ATP; the downstream enzymes was co-immobilize with Ppa that can hydrolize pyrophosphate.

Amorphadiene was successfully produced with the co-immobilized system, and could reach approximately 40% AD yield after 12 h incubation (Figure 5.4B). The recyclability of system I was very poor. Less than 10% AD yield was retained after 5 cycles of reaction (Figure 5.4C). LCMS analysis indicated that FPP was accumulated at 12h in the first cycle of reaction (Figure

5.4B), suggesting Ads reaction was inhibited due to the presence of high concentration of ATP and pyrophosphate in the reaction mixture, as discovered in Chapter 4. Therefore, the two additional enzymes, pyruvate kinase (PyfK) and inorganic pyrophosphatase (Ppa), were introduced to co-immobilize with the pathway enzymes in a random fashion (Figure 5.3, system II). Remarkably, AD yield was enhanced to 80% after 12h incubation, a 2 fold enhancement as compared to system I (Figure 5.4B). This was probably attributed to the enhanced Ads activity, as FPP concentration was significantly reduced which is more than 6 fold lower than system I. Surprisingly, the productivity of system II was maintained at greater than 60% after 7 cycles of reactions (Figure 5.4C). Judging from the trend, the enzymes can be reused for more cycles of reactions. This is probably correlated to the minimized amount of FPP left unreacted in the reaction, which will induce precipitation (Chapter 3) and deactivate the enzymes.

The improved system productivity spurred us to optimize the spatial arrangement of the multiple enzymes. A novel bi-modular reaction system was proposed, as shown in Figure 5.3, System III. Since the mevalonate pathway enzymes are ATP dependent, they are co-localized with PyfK which can recycle ATP. The downstream enzymes produce pyrophosphate (Ppi) are co-localized with Ppa that can hydrolyze Ppi. We hypothesize that by segregating Ads from PyfK, the local ATP concentration would be lowered around Ads by substrate channeling effect, since the ATP regenerated by PyfK can be effectively consumed by the mevalonate pathway enzymes before it is diffused to the external reaction medium. Similarly, by bringing Ads in close proximity with

Ppa, Ppi would be hydrolyzed immediately. As a result, an enhanced productivity would be expected. The concept was successfully illustrated as an improved AD yield was observed for system III that approximately 100% conversion was achieved in the first cycle of reaction. This was a 2.5 and 1.25 fold improvement as compared to system I and II respectively (Figure 5.4B). Moreover, System III was kinetically enhanced as AD was produced in a higher rate than the other two systems (~ 50% enhancement as compared to system II at 4 hour, Figure 5.4B). The enhanced rate of reaction was probably due to the further reduction in FPP concentration, as negligible amount of FPP was detected by LCMS in system III. However, to our disappointment, AD yield was gradually decreased to ~67% after 7th cycle of reaction without any appreciable enhancement as compared to system II (Figure 5.4C).

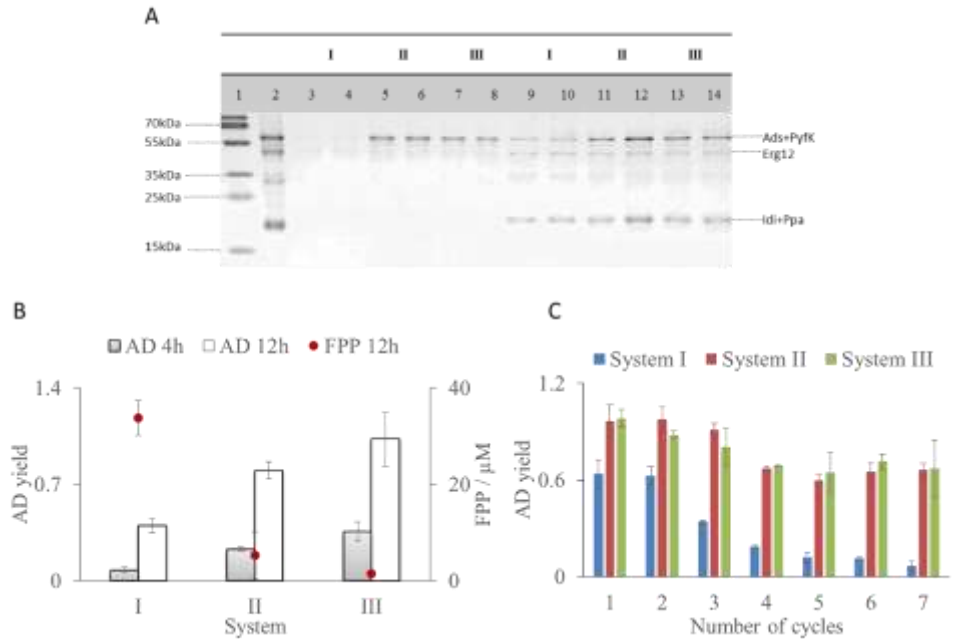

Figure 5.4. Production of amorpha-4,11-diene by co-immobilized AD synthesis pathway.

(A). SDS page analysis of the enzymes. Lane 2 is the total enzyme before immobilization. Lane 3-8: the residual enzymes in supernatant after immobilizing. Lane 9-14: the enzymes that were

still immobilized on the solid support after 7th cycle of reaction. Its concentration was diluted with equal volume of NPI400, which is half of that loaded in lane 2. (B). The AD yield produced by the three systems at 4h and 12h. The FPP concentration was measured by LCMS at 12h. (C). The AD yield at each cycle of 12h reaction. All the measurements were done in triplicates, and the standard errors were shown.

In order to evaluate the decrease in the AD yield after repeated use of the multienzymes, LCMS (Figure 5.5) analysis were conducted to access the accumulated metabolites at the end of 7^{th} cycle of reaction. Surprisingly, apart from FPP, IPP/DMAPP was accumulated to a high level for all three systems. This was probably due to the leakage of farnesyl pyrophosphate synthase (IspA) from the Ni-NTA resin, as reflected by SDS PAGE that IspA was hardly detected (Figure 5.4A, Lane 9-13). Moreover, the mevalonate pathway enzymes were less efficient for both system II and III, as the substrate MVA and PMVA were found to be accumulated. This could be due to the inactivation of PyfK that was not able to continuously supply ATP to the system. The hypothesis remained to be tested.

Nevertheless, the immobilized multienzyme system was successfully recycled. Approximately 2.2g/L AD was produced in 4 days. This translated to approximately 80% overall AD yield and a significant more than 6 fold enhancement in AD specific yield as compared to free enzymatic system.

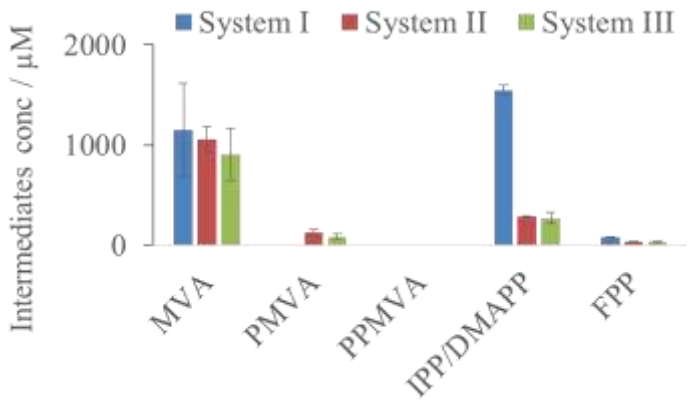

Figure 5.5. UPLC-(TOF)MS analysis of the reaction intermediates in the 7th cycle of reaction.
All the measurements were done in triplicates, and the standard errors were shown.

5.3 Discussion

Co-immobilization of the enzymes is an attractive technology to develop the *in vitro* multienzymatic biosynthetic platform for industrial application. In addition to the advantages associated with the single enzyme immobilization, multienzyme co-immobilization can achieve spatial arrangement of the multi-biocatalysts by artificially varying the inter-enzymatic distance. A number of reports have described the benefits of substrate channeling where the intermediates were efficiently shuffled towards product when the in-tandem enzymes were brought in close proximity [18, 196, 198]. More importantly, the toxic intermediates can be removed effectively to prevent them from inactivating the pathway enzymes [18]. We applied the concept of co-immobilization to the AD synthesis pathway, and designed a bi-modular system to boost the productivity from 40% to 100%. The *in situ* removal of the byproduct pyrophosphate and recycling of ATP were found to improve the recyclability of

the system. Since there were less inhibitors accumulated along the pathway, the activity of the enzymes should be preserved longer, and hence the specific yield could be further enhanced. In the future, more innovative and stable ATP recycling system can be used. One interesting idea is to use a coupled pyruvate oxidase and acetyl kinase system to directly use pyruvate and phosphate as the starting materials to regenerate ATP via the formation of high energy compound acetyl phosphate [204]. In such a way, the phosphate formed after hydrolyzing pyrophosphate by inorganic pyrophosphatase can be used as energy source to replenish ATP in the system. Therefore, the *in vitro* system should achieve self-sustainability. Furthermore, DNA-directed immobilization technique can be explored to further program and address the enzymes rationally to enhance the substrate channeling effect.

One limitation of the co-immobilized system when using IMAC is the steric hindrance of the liquid and solid interface. The high interfacial energy might inactive the enzymes and hampers the diffusion of the substrates to the reaction site. Therefore, the turnover rate of the system may be reduced, as seen for immobilized Erg12. Therefore, modification of the surface with a spacer, such as polyethylene glycol (PEG), might provide a means to overcome such challenges [203].

5.4 Conclusion

In summary, a recyclable multienzymatic platform was established for AD production. The productivity of the system was systematically improved from 40% to ~100%, with the knowledge that high concentrations of ATP and Ppi inhibited Ads activity. The platform could be effectively reused for more than 7 cycles of reactions with AD yield maintained at more than 60%. Thus, the specific yield of the system was significantly improved for more than 6 fold as compared to the free enzymatic system.

5.5 Materials and Methods

5.5.1 Expression and purification of the enzymes

The bacteria strains and plasmids for the amorphadiene synthesis pathway enzymes were the same as Chapter 3 and Chapter 4. The purification methods were the same as described in the previous chapters. The enhanced green fluorescence protein (eGFP) was from the previous study [20]. Its concentration was determined with Micro BCA protein assay kit (Thermo scientific, MA). The standard curve of eGFP was constructed between its florescence (λ_{ex}=490, λ_{em}=510nm) and known concentration. Different buffer was used to dilute the eGFP enzymes, to exclude any background fluorescence.

5.5.2 Immobilize His_6-tag enzymes by dilution

The purified proteins were contained in NPI400 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 400 mM imidazole, pH 8) which was diluted 10 times with phosphate saline buffer. Then the enzyme was immobilized onto Ni-NTA grafted resin (USB, Affymetrix, CA) at a ratio of 5µg enzyme per mg of resin, and vortex at room temperature for 10min. Then the supernatant was carefully remove after centrifuging the mixture. Portion of the supernatant was subject to dodecyl sulfate-15% polyacrylamide gel electrophoresis (Bio-Rad, CA), to measure the residual amount of the protein. To determine the amount of enzymes immobilized, the resin was resuspended in NPI400 buffer, and both the resin and enzymes were loaded into SDS page gel. When immobilizing mevalonate kinase (Erg12), the yield was determined by measuring its kinetic activity before it was immobilized (A_{tp}), the activity on the solid support (A_{im}) and the residual activity in the supernatant (A_{re}). Then the immobilization yield and activity yield was determined by Equation 5.1 and Equation 5.2 respectively [103].

$$Immobilization\ yield\ (\%) = \frac{A_{tp} - A_{re}}{A_{tp}} \times 100\%$$

Equation 5.1

$$Activity\ yield\ (\%) = \frac{A_{im}}{A_{tp} - A_{re}} \times 100\%$$

Equation 5.2

5.5.3 Co-immobilized Multienzyme reaction

There were three systems of co-immobilized multienzymes (Figure 5.3). For system II and III, the enzymes were co-immobilized by directly mixing the enzymes with the solid resin in the ratio as stated above. For system III, the two groups of enzymes were mixed with corresponding amount of resin according to the ratio separately, and then combined the co-immobilized enzymes into one-pot. Subseqeuntly, the multienzyme reaction was initiated upon the addition of substrates (10 mM), (±)mevalonic acid contained in a buffer (50 µl) consisted of Tris/HCl (100 mM, pH 7.4), $MgCl_2$, ATP. If not indicated otherwise, the amount of amorphadiene pathway enzymes were added according to the optimum enzyme levels found in Chapter 3. PyfK and Ppa were both added in excess at the final concentration of 50mg/L. The reaction was performed at 30 ^{o}C with an overlay of dodecane phase that contained known concentration of trans-caryophyllene as an internal standard. At the end of the reaction, the dodecane phase was diluted 10 times in ethyl acetate and subject to GCMS analysis. At each time point, a volume of reaction mixture was drawn and equal volume of 1% NH_4OH and 8 volumes of ice cold methanol were added to stop the reaction and subjected to UPLC-(TOF)MS column for measurements. At the end of each cycle of reaction, the immobilized enzymes were separated from the aqueous solution by centrifuging at 10,000 g for 5min, and the supernatant was carefully removed. Then the solid support was washed with 50µl of NPI10 with 20% glycerol, and the wash buffer was removed by repeating the centrifuging step. Then the resin was resuspended in NPI10 for the next cycle of reaction. The (±)mevalonic acid was prepared by

complete alkaline hydrolysis of 2 M (±)mevalonolactone (Sigma, MO) with equal volume of 2 M potassium hydroxide at 37 ^{o}C for 1.5 h, and neutralized by adding 1 M hydrochloric acid to pH 7 [170].

5.5.4 UPLC-(TOF)MS analysis of mevalonate pathway intermediates

The analysis was done based on the method developed previously with slight modification [42]. In brief, 5 µl samples were injected into a UPLC C18 column (Waters CSH C18 1.7 µm, 2.1 mm x 50 mm) connected to UPLC (Waters ACQUITY UPLC)-(TOF)MS (Bruker micrOTOF II, MA). Elution was carried out with a step change from 100% aqueous solution containing 15 mM acetic acid and 10 mM tributylamine (0.5 min) to 10% aqueous solution with 90% methanol for another 3.5 min. Electrospray ionization was used and mass spectrometry was operated to scan 50–800 m/z in negative mode with 2500 V end plate voltage and 3200 V capillary voltage. Nebulizer gas was provided in 2 bar, dry gas flow rate was 9 ml/min, and dry gas temperature was 200 ^{o}C. At the assay condition, all the intermediates were detected in the form $[M-H]^{-}$. Retention time was subsequently determined for each intermediate with respective synthetic standards and the set m/z extraction range. The peak area was calculated and subsequently used to compute the intermediates concentrations with the software provided by the manufacturer. The calibration curves were constructed with synthetic standards prepared under similar reaction conditions without enzymes. Linearity of the assays were determined individually with coefficients of determinants (R^2) greater than 0.90.

5.5.5 GCMS analysis of amorpha-4,11-diene (AD)

AD analysis was carried out based on the modified method developed by Martin et al. by scanning three ions; the m/z values are 117, 189 and 204 [170, 171]. 1 µl sample was injected into HP-5 column (Agilent Technologies 7890A gas chromatograph-mass spectrometry, Agilent, CA) with a linear temperature increase of 50 ^{o}C/min from 80 ^{o}C to 300 ^{o}C and hold at 300 ^{o}C for another minute. The peak area was calculated and subsequently used to compute AD concentrations with the software provided by the manufacturer. Amorpha-4,11-diene concentrations were determined relative to the internal standard trans-caryophyllene of known concentration.

Chapter 6 *In vitro* biosynthesis of Artemisinic acid and Dihydroartemisinc acid by cytochrome p450 system

6.1 Introduction

The conversion of amorphadiene to dihydroartemisinic acid which can then be chemically synthesized to artemisinin is an important step in producing artemisinin [149]. The first step of this conversion is the oxidization of amorphadiene to artemisinic alcohol and then to artemisinic aldehyde and subsequently to artemisinic acid. This bioconversion is by the cytochrome p450 system from *Artemisia annua* (Figure 6.1A) which consists of cytochrome p450 enzyme (CYP450) and its electron supply partner cytochrome P450 reductase (CPR) [146]. Despite the successful production of 25g/L of artemininc acid in a lab-scale reactor [3], the yield is still less than the highest amorpha-4,11-diene (AD) titer reported, which was approximately 40g/L [162]. The co-overexpression of CYP450 with AD synthesis genes in cells is known to inhibit the upstream AD production, a phenomenon that was observed similarly in our lab too.

To overcome the limitation of cellular production of artemisinin precursors, a hybrid *in vivo* and *in vitro* platform for artemisnic acid production is demonstrated in this chapter. Amorphadiene and CYP450 were separately produced in *E. coli* and yeast cells, respectively. Subsequently whole cell biocatalysis was carried out to convert AD to artemisnic acid. In this way, 80% of amorphadiene was effectively converted to artemisnic acid. Therefore, the new platform developed is a powerful tool for cytotoxic product production and definitely have wide application in biotechnology and biosynthesis.

6.2 Results

6.2.1 Genetic optimization of cytochrome p450

As a membrane bounded enzyme, CYP450 was targeted to the inner membrane of *E.coli* by the first 29 amino acids at N terminal of the protein with a membrane binding sequence (MBS). It has previously been shown that the enzyme might not be functionally optimal inside bacteria with this naive MBS. So we replace the sequence partially (first 6, 15, 22 or 29 amino acids) with a short leading sequence from *Bovine* CYP450, which has been shown to help with the expression of other CYP450 in *E.coli* [205] (Figure 6.1B). In addition to the native CPR from *Artemisia annua* (CPRaa), another CPR from *Taxus baccata* which has been successfully introduced to *E.coli* for the oxidization of another isoprenoid - taxadiene was also tested by fusing to the CYP450 (Figure 6.1B).

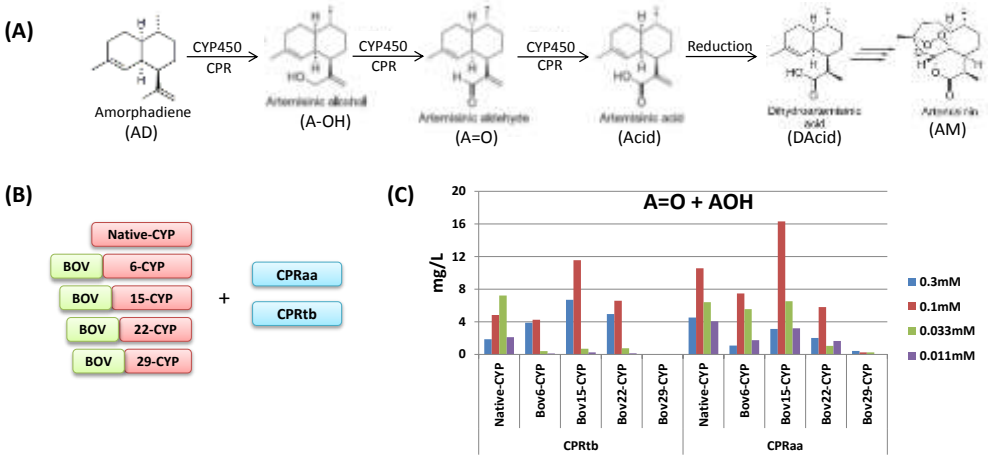

Figure 6.1. protein engineering of cytochrome p450 system.
(A), the reaction catalyzed by the cytochrome p450 system (CYP450) (B), scheme of protein engineering for cytochrome p450 system. Abbreviations: *Artemisia annua* cytochrome p450 enzyme (Native-CYP), short membrane binding sequence from *Bovine* cytochrome p450 enzyme (BOV), engineered *Artemisia annua* cytochrome p450 enzyme with the first 6, 15, 22 or 29 amino acid removed (6-CYP, 15-CYP, 22-CYp, 29-CYP), *Artemisia annua* cytochrome P450 reductase (CPRaa), *Taxus baccata* cytochrome P450 reductase (CPRtb). (C), the production of artemisinic alcohol and artemisinic aldehyde by engineered cytochrome p450 systems. TM2-dxs, TM3-iDF, TM1-ADS, MG1655 strain producing amorphadiene with DXP pathway was used as the parent strain to harbor the cytochrome p450 systems.

MG1655 strain overexpressing the DXP pathway was used as the parental strain to test the performance of the different engineered CYP450 strains. In all tested strains, the major product of CYP450 system was artemisinic alcohol (AOH), and only very tiny amount of artemisinic acid (AA) could be detected. The overall conversion of amorphadiene (Figure 6.1C) showed that BOV15-CYP-CYPaa (CYP15) was the best strain and was used for further analysis. However, the maximum yield was only about 5% of the amorphadiene yield (300mg/L for DXP pathway).It appeared that some major bottlenecks still exist for the function of this plant CYP450 system inside *E.coli*. Other parameters such as growth condition and co-factor supply system (NADPH-FMN-FAD) could be further explored to optimize the system.

6.2.2 Overexpressing CYP71AV1 in E. coli *strains for whole cell biocatalysis*

As CYP450 is membrane bounded, the overexpression of this enzyme may affect growth of *E.coli* as well as the production of amorphadiene. Furthermore, it was very difficult to adjust and optimize the reaction conditions for high yield *in vivo*. To overcome those difficulties, we explored the *in vitro* bioconversion of amorphadiene using CYP450. These were produced separately and then mixed together in a cell free system (Figure 6.2A).

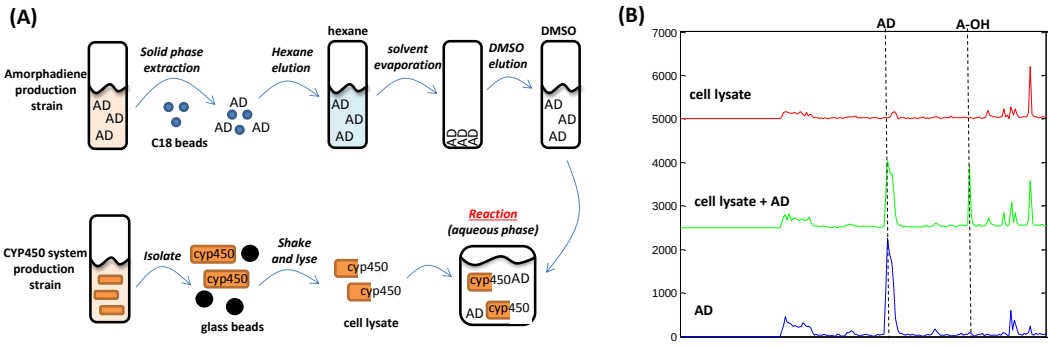

Figure 6.2.*in vitro* biosynthesis of cytochrome p450 system.
(A), illustration of *in vitro* biosynthesis process consist of amorphadiene extraction and CYP450 system extraction. (B), demonstration of in vitro biosynthesis. Samples were collected at 2h after reaction (28^{o}C, pH7).

Amorpha4-11-diene (AD) was firstly purified and extracted by solid phase with C18 beads and then eluted using hexane. The hexane solvent was then removed by evaporation and the AD was dissolved in DMSO which was miscible with the reaction in aqueous phase. Approximately 50% of the AD produced *in vivo* could be recovered by this method. By supplementing AD into the whole cell (final O.D. 10-20) catalyzed reaction mixture at 28^{o}C, pH7, conversion of AD to AOH was successful (Figure 6.2B).

To further optimize the *in vitro* reaction, several *E. coli* strains were screened to select the best strain for functional expression of CYP15 by monitoring the bioconversion of exogenously added AD (Figure 6.2A). We observed severe growth inhibition for M15 strain when CYP15 was induced, which suggested cytotoxicity of the gene product (Figure 6.3A). Among the strains screened, BL21 strain displayed the highest yield and cell viability, although only ~7% conversion from AD to AOH was observed (Figure 6.3A) in two hours. Prolonging the reaction time to 12 hours did not further enhance the

yield significantly, which showed that the reaction was stalled completely after the conversion reached certain level. We have verified that this loss of activity was not due to the inactivation of the CYP15 enzyme, as shown in Figure 6.3C that similar level of conversion was reached even though the whole cell biocatalysts was pre-incubated at 28^{o}C for two hours. Therefore, this suggested that enzyme inactivation was unlikely. Subsequently, we supplemented the reaction with essential co-factors, such as iron(III) chloride, FMN/FAD and a NADPH recycling system with in-house produced glucose dehydrogenase (GDH) from *Bacillus Megaterium*. However, little improvement of final yield was observed. This observation was similar to *tobacco*, a plant cell. The conversion of AD to AA was incomplete, even though CYP71AV1 was successfully overexpressed [148]. Hence, the intracellular environment may be important to ensure activity of CYP15.

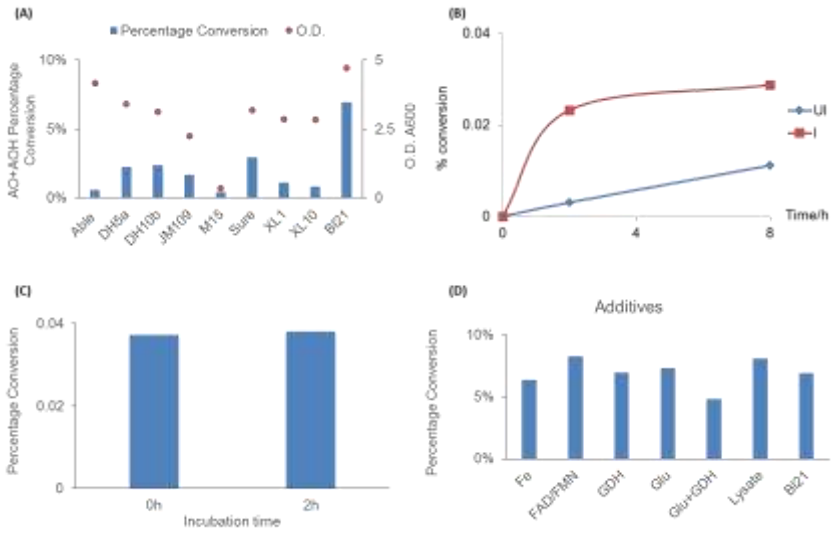

Figure 6.3 *in vitro* whole-cell CYP15 enzyme reaction.
(A), Different bacteria strains were screened for the best host for CYP15 expression and activity. (B), The time course of BL21 whole cell CYP15 biotransformation. The abbreviations are as follows. UI is uninduced cells. I is induced cells. (C), investigation of heat inactivation of whole cell CYP15 biocatalysis. (D). Additives screening to enhance the conversion yield of BL21 whole cell CYP15 biotransformation.

6.2.3 ***Overexpressing CYP71AV1 in* S.cerevisae *W303 strain for whole cell biocatalysis***

Baker yeast is a better host to express plant P450 enzymes, and they are widely used for industrial biotechnology [206]. Therefore, we overexpressed the CYP15 enzyme in the wild type W303 strain. Although hardly any visible protein band could be seen on SDS page gel (Figure 6.4A), production of AA was observed when mixing the whole cell (final O.D. ~20) with AD at 28^{o}C, pH7 for two hours, as shown in Figure 6.4B. It clearly indicated that only the cell harbouring the plasmid p416-TEF-CYP15 was able to catalyze three consecutive oxidation of AD into AA.

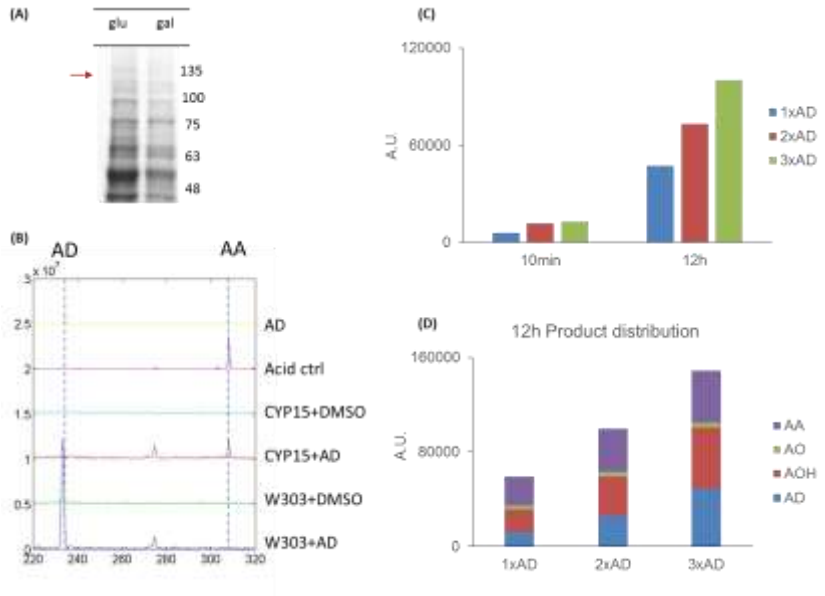

Figure 6.4 *in vitro* production of artemisinic acid (AA) by yeast W303 whole cell CYP15.

(A), SDS PAGE gel of yeast expression CYP15. The red-arrow indicate the expected sized of CYP15, which could not be seen on the gel. Glu: glucose medium. Gal: galactose medium. (B), Demonstration of the production of acid. (C), the total amount of oxidized products produced when the substrate concentration was increased linearly from 1x AD (~400mg/L), 2x AD (800mg/L) to 3xAD (1.2g/L). The integrated area (A.U.) from GCMS analysis was reported. (D), the product distribution profile for three different initial substrate concentrations. The abbreviations are as follow. AD: amorphadiene. AOH: artemisinic alcohol. AO: artemisinic aldehyde. AA: artemisinic acid.

When the substrate AD concentration increased linearly from 400mg/L to 1.2g/L, more products accumulated after 12h of incubation in a linear fashion (Figure 6.4C), where about 40-50% of the final oxidized product was AA (Figure 6.4D). The product distribution was fairly similar despite the initial concentration of substrate; they mainly consisted of AOH and AA. By detailed analysis of the reaction kinetics, the substrate concentration was saturated at 2xAD which was ~800mg/L (~4mM). A two fold decrease in total percentage of conversion at 10min was observed when the initial AD concentration was at 3x AD (~1.2g/L AD / ~6mM AD) (Figure 6.5A). However, the final conversion yield at 12h was approximately the same (~70%) for all three different initial substrate concentrations (Figure 6.5B). One plausible explanation was that the enzymatic catalysis sites were saturated with AD, and thus depriving subsequent oxidation of alcohol to aldehyde and acid. This implied that CYP15 would release each intermediate oxidation product from its catalysis pocket and allowing them to diffuse to the next reaction site for further oxidation. It was further supported by the observation that AOH and AO were both detected in the product pool, too. Hence, the rationale way to improve the conversion further was to increase the available biocatalyst sites. One way was to enhance the turnover rate of CYP71AV1. Literature suggested that glucose may affect biocatalysis efficiency when resting *E. coli* cells were used [107], since glucose would affect the co-factor availability. Therefore, by supplementing 10mM glucose into the reaction medium, a slight enhancement, ~50% increment in kinetics, was observed for all three different initial AD feeding (Figure 6.5A). However, it did not change the

substrate saturation concentration; the percentage of converted product at 10min still decreased at 3x AD concentration. Neither did glucose change the final product distribution profile. The final AA yield was about 40-50% for all the three initial AD concentration. Therefore, we increase the biocatalyst yield in order to increase the catalytic efficiency.

Another noteworthy observation was that CYP15 activity was lost after the yeast cells were lysed by shaking with glass beads (Figure 6.5B); less than 10% conversion was achieved when the cell lysate was used. This might be due to the disruption of the membrane fraction of the enzyme, or could be due to the change of environment leading to the loss of activity.

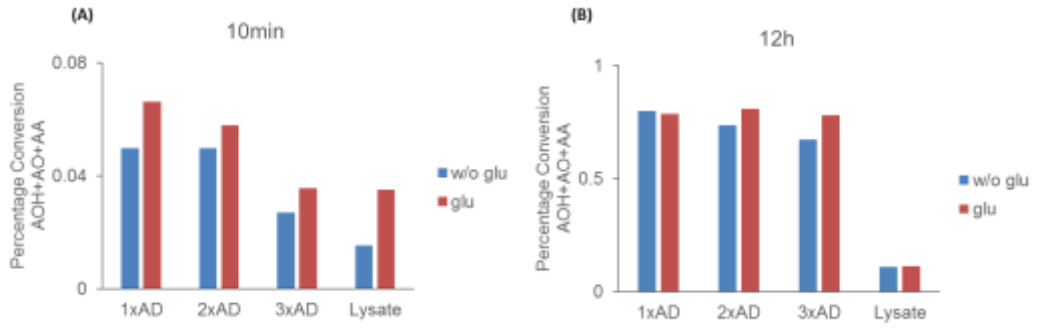

Figure 6.5 analysis of kinetics of CYP15 biotransformation.
Percentage of AD being converted to downstream oxidized products, including AOH, AO and AA, after incubating at 28^{o}C for(A) 10mins and (B) 12 hours. Both reaction conditions with (glu) and without glucose (w/o glu) were shown. The abbrevations are as follows: 1x AD: ~400mg/L amorphadiene. 2x AD: ~800mg/L amorphadiene. 3xAD: ~1.2g/L amorphadiene. Lysate: cells were lysed with glass beads.

The yeast plasmid p416-TEF used is a low copy number (1-2 copies) centromeric expression vector with a constitutive strong promoter. Therefore, a high copy number 2µ-based plasmid pYES260-Gal1 was used instead. The additional benefit associated with the latter vector was the promoter is tightly regulated by Gal1, and hardly any leaky expression was observed when the cells

were grown in glucose medium. Thus, it would avoid any toxicity effect associated with the gene product during cell growth and control gene transcription when cells attain certain biomass. When the yeast cells were transformed with the new plasmid, the growth profile did not alter significantly (Figure 6.6A). Therefore, this high copy plasmids did not exert undue metabolic burden to the host cell. More than 50% improvements in AA yield was achieved (from 50% to 80% overall conversion), with W303 strain harbouring pYES-Gal1-CYP15 plasmid as compared to that with p416-TEF-CYP15 (Figure 6.6B).

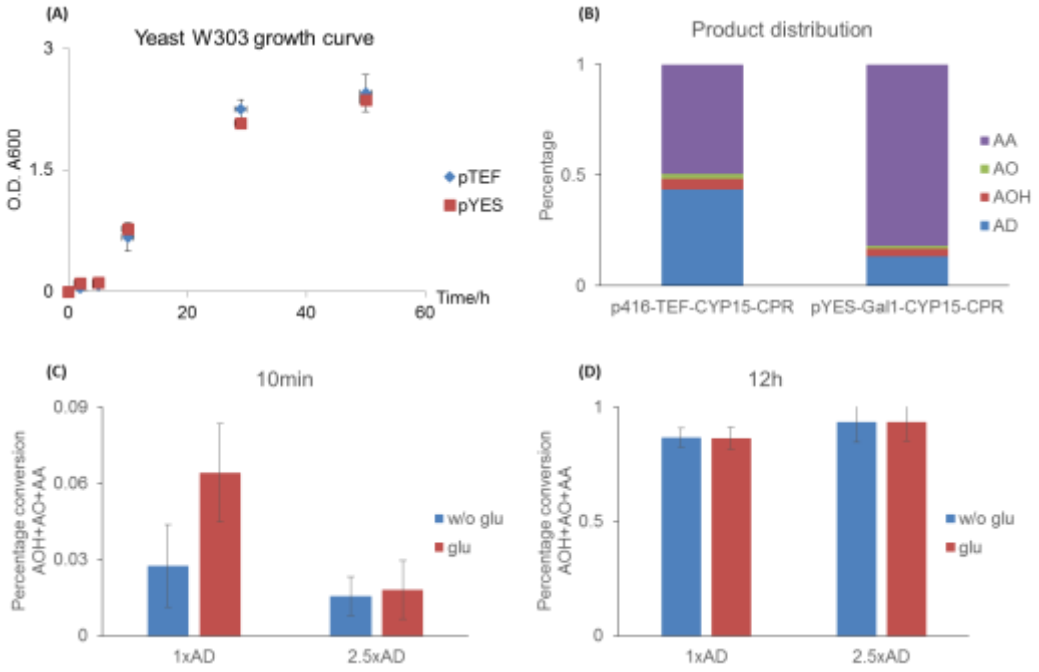

Figure 6.6 Improving the CYP15 enzymatic yield by use of high copy-number plasmid pYES-Gal1.

(A), the growth curve of the W303 strain harbouring both p416-TEF-CYP15 (pTEF) and pYES-Gal1-CYP15 (pYES) plasmids. (B), comparison of product distribution between two vector systems. The abbreviations are as follows. AD: amorphadiene. AOH: artemisinic alcohol. AO: artemisinic aldehyde. AA: artemisinic acid. percentage of AD being converted to downstream oxidized products, including AOH, AO and AA, after incubating at 28^{o}C for(C) 10mins and (D) 12 hours. Both reaction conditions with (glu) and without glucose (w/o glu) were shown. The abbreviations are as follows. 1x AD: ~400mg/L amorphadiene. 2.5xAD: ~1g/L amorphadiene.

The increase in the conversion yield was likely to be attributed to the enhanced enzyme expression. However, the percentage conversion at 10min was still low when 1g/L AD was added (2.5x AD) (Figure 6.6C). Glucose was not effective to improve the rate of conversion, although an enhancement of conversion rate was observed at low initial AD concentration (400mg/L, 1xAD). Therefore, it was likely that specific competitive inhibition by substrate accounted for the lagged conversion efficiency, since the same binding pocket was used for all three steps of oxidation. However, despite the reduced kinetics at the start of the reaction, ~80% conversion could be achieved after 12h incubation (Figure 6.6D). Approximately 800mg/L artemisinc acid could be produced, calculated based on using the AD standards.

6.2.4 ***Overexpressing CYP71AV1 in* S.cerevisae *BY4741 strains for whole cell biocatalysis***

In view of the low cell density of W303 cells (final O.D was about 2), we explored another yeast strain BY4741. Moreover, the uptake of NADPH by W303 strain was limited, as supplementing NADPH directly into the reaction medium did not enhance the productivity of the whole cell biocatalysis (results not shown). This indicated that the lysed cells were not useful. Therefore, permeabilized whole cell biocatalysis might be an alternative solution, so as to allow the ease of uptake of reaction substrates and additives across the membranes. Instead of enzyme-dependent or organic-solvent-treatment permeabilization, we set to explore yeast cells with thin cell walls. CWP1 and

CWP2 encode the two major mannoproteins of the outer cell wall; deletion of the two genes caused increased yeast sensitivities to genotoxic chemicals [207]. Ergosterol, as the predominant sterol component in the cell membrane, can affect membrane permeability. Deletion of Erg6 gene was not lethal but could alter membrane permeability [207]. Thus, BY4741 cells with each of these three genes deleted (denoted as dCWP1, dCWP2, dERG6) were obtained from Euroscarf (Frankfurt, Germany) to overexpress CYP15 and used for subsequent studies.

BY4741 strains grew much faster and to a higher cell density than W303 strain in glucose medium, and single gene deletion did not affect cell growth (Figure 6.7A). All four phenotypes were explored for the catalytic conversion of AD (~500mg/L) to AA. All of them were able to achieve more than 70% conversion within 12 hours. BY4741 cells harbouring pYES-Gal1 plasmid was more productive than the strain carrying p416-TEF, where AA was almost the sole product in the oxidized products (Figure 6.7B). Among the four phenotypes, dCWP2 strain displayed the highest conversion yield, albeit there was very slight enhancements. This suggested that transporting AD across yeast cell wall may not be a bottleneck. We further compared the kinetics of dCWP2 strain and BY4741 strain with and without additives in the medium. Rationally, with thinner cell wall, additives might be easier to diffuse through the membrane and more effectively speed up the catalytic reaction. As reasoned, when glucose was supplemented in the reaction medium, dCWP2 strain exhibited a more significant change (~3 fold enhancement) in rate of conversion as compared to the wild type BY4741 (~2 fold) at 1h (Figure 6.7C). Moreover, when initial AD concentration

was increased from 500mg/L (1x AD) to 1g/L (2x AD), the percentage of oxidized products in dCWP2 strain at 1h increased from 12% to 20%, respectively (Figure 6.7C). In contrast, 10% of oxidized products were produced with BY4741 strain in both initial AD concentrations. Therefore, dCWP2 strain displayed a more competitive advantage than wild type BY4741, and was used for further downstream reaction medium optimization.

Buffer optimization was carried out, by screening a panel of important additives. NADPH is the external co-factor that drives the chain event of electron transfer. Supplementing glucose into the reaction medium resulted in a positive enhancement in the yield and was likely to be due to the enhanced amount of NADPH. Other additives were also explored to enhance the final yield of acid. The electron shuttle enzyme CPRaa used FMN/FAD system to transfer the electron to the iron-sulfur center of P450 enzyme. Depletion of the co-factors may result in the loss of enzymatic activities. Thus, more FMN/FAD was supplemented in the reaction medium. Moreover, reactive peroxide will be formed in the course of the reaction, which is detrimental to the enzymatic activity. L-proline [208] and L-ascorbic acid are singlet oxygen species scavengers . By supplementing these additives into the reaction medium, further 50% enhancement in the final acid yield was observed (Figure 6.7D). A combinatorial optimization of additive concentrations could be explored to examine any synergistic effect so as to further boost up the CYP15 enzyme efficiency in future studies.

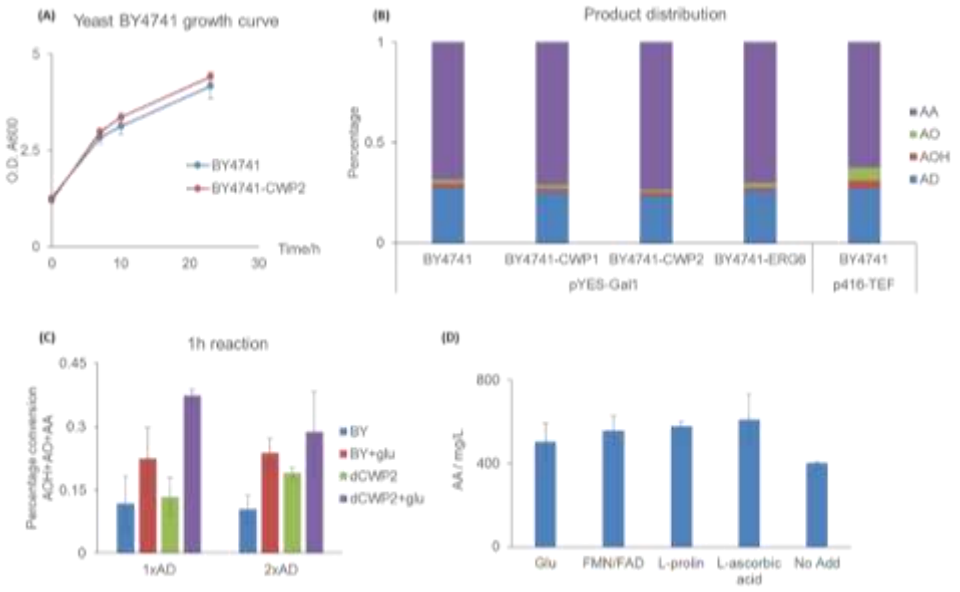

Figure 6.7 *in vitro* production of artemisinic acid (AA) by yeast BY4741 whole cell CYP15.

(A), growth curve of BY4741 strain and BY4741 strain with CWP2 gene deleted. (B), comparison of product distribution between pYES-Gal1 and p416-TEF vector systems. The abbreviations are as follows. AD: amorphadiene. AOH: artemisinic alcohol. AO: artemisinic aldehyde. AA: artemisinic acid. (C), Comparison of the CYP15 conversion efficiency between BY4741 (BY) strain and BY4741 strain with CWP2 gene deleted (dCWP2). Percentage of AD being converted to downstream oxidized products, including AOH, AO and AA, after incubating at 28^{o}C for 1 hour. Both reaction conditions with and without glucose were shown. The abbreviations for the x-axis label is as follows. 1x AD: ~500mg/L amorphadiene. 2xAD: ~1g/L amorphadiene. (D), screening of additives to enhance the final AA yield when dCWP2 whole cell CYP15 was used.

6.2.5 Production of dihydroartemisinc acid (DHAA)

Research showed that the precursor of Artemisinin is DHAA instead of AA [152]. Therefore, it is more desirable to produce DHAA enzymatically. In the process of optimization, we observed that ~70% of oxidized products were outside the cell, regardless of the phenotype (Figure 6.8A). Hence, it served as an avenue for us to use the auxiliary pathway externally that further converted AOH to dihydroartemisinc acid (DHAA) through three consecutive enzymatic steps: alcohol dehydrogenase (Adh1), double bond reductase (Dbr2) and aldehyde dehydrogenase (Aldh1). The advantage is that the three auxiliary enzymes could

be individually overexpressed in separate hosts to its maximum yield. As shown in Figure 6.8B, high amount of enzymes were obtained when the genes were overexpressed with 0.1mM IPTG. Therefore, by mixing the cell lysates with the dCWP2 cells expressing CYP15, a tiny amount of DHAA, ~20% of AA, was successfully produced (Figure 6.8C), and the characteristic MS peak was identical to that preciously reported [151]. There is a need to fine tuning the enzyme amount, as literature suggest that Dbr2 is likely the bottleneck step for high production of artemisinin in *artemisia annua* [209].

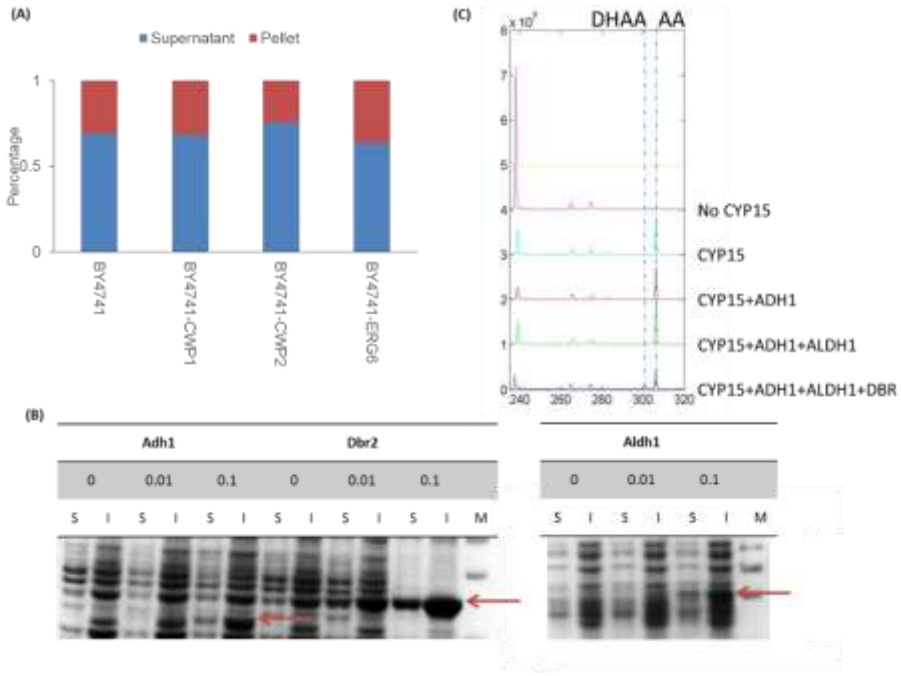

Figure 6.8 auxiliary reaction for dihydroartemisinc acid (DHAA) production.
(A), the distribution of acid in the pellet and reaction supernatant when different BY4741 strains were used for whole cell catalysis. (B), SDS page analysis of Mg1655 (DE3) lysate overexpressing Adh1, Dbr2 and Aldh1 respectively. The inducer IPTG concentration was indicated in the table, which is 0mM, 0.01mM and 0.1mM. The abbreviations are as follows. Adh1: alcohol dehydrogenase. Dbr2: double-bond reductase. Aldh1: aldehyde dehydrogenase. S: soluble fraction of cell lysate. I: insoluble fraction of cell lysate. (C). demonstration of the production of DHAA.

6.2.6 Exploring different reaction formats: hybrid in vivo-in vitro *and total* in vitro *synthesis*

Purified AD (~1g/L) was fed into the growing cells at different growth stages of BY4741 and dCWP2 strains. Interestingly, the addition of AD to cells induced to overexpress CYP15 gene resulted in severe growth retardation in wild type BY4741 but not dCWP2 (Figure 6.9A and B). This could be due to the toxicity of AA produced [210], where dCWP2, cells with thinner cell wall, were able to get rid of this toxic products faster. Although this is an attractive hypothesis to be tested, there was no significant enhancement in the conversion yield between the two phenotypes (Figure 6.9C). The toxic effect was less obvious when the substrate was introduced at mid-exponential growth phase, suggesting other mechanism may be involved in the growth retardation observed when AD was fed at early exponential phase.

Lastly, total *in vitro* synthesis was carried out, which the upstream total enzymatic AD synthetic pathway was also added, (Chapter 3), into the reaction mixture. Total *in vitro* synthesis was able to further convert more than 50% AD to oxidized products (Figure 6.9D). More AOH was present in the oxidized products, suggesting an the existence of yet to be characterized bottleneck(s).

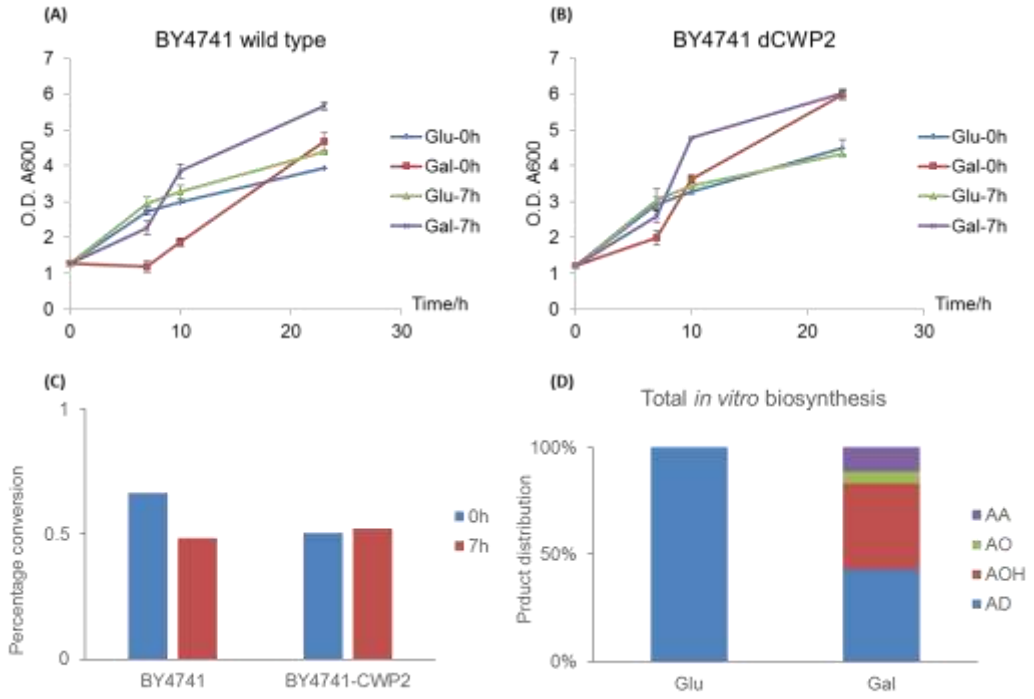

Figure 6.9 Different reaction format for AA production.

Growth curve of (A) BY4741 and (B) BY4741 with CWP2 gene deleted. Purified AD (final concentration ~1g/L) was fed into the living cells at different growth stage: 0h (early exponential phase), and 7h(mid-exponential phase). 2% Galactose (Gal) medium was used to induce the expression of CYP15, whereas no expression of CYP15 when the cells were grown in 2% glucose medium (Glu). (C), percentage of AD being converted to downstream oxidized products, including AOH, AO and AA, after the cells were grown in 28^{o}C for 10hours. (D), The product distribution of total *in vitro* synthesis method using 5mM mevalonic acid at 28^{o}C for 12h. The dCWP2 cells grown in Glu condition was used as control to quantify the amount of AD produced by multienzyme synthesis, whereas the dCWP2 cells grown in Gal condition was used to further catalyze the conversion from AD to AA. The abbreviations are as follows. AD: amorphadiene. AOH: artemisinic alcohol. AO: artemisinic aldehyde. AA: artemisinic acid.

6.3 Discussion

Selective oxidation and reduction of the undecorated hydrocarbon backbone of amorphadiene is a difficult task [150]. It is metabolized by the membrane bound enzyme cytochrome p450 monooxygenase (CYP71AV1), which is a class of important heme containing enzyme that enabled functionalization of AD to bioactive Artemisinin [146]. The enzyme was discovered recently and much is still unknown; for example, the enzyme has not been kinetically characterized. Our initial attempt to overexpress and purify the membrane bound enzyme from *E. coli* was not successful. No apparent protein band was shown on the SDS page gel, after the host *E. coli* cells were cultured at 20^{o}C for 24h (results not shown). It was likely to be due to the cytotoxicity of the gene product which may have resulted in the loss of plasmid. Moreover, CYP71AV1 in *E. coli* was unable to catalyze the three consecutive oxidation reactions and displayed extremely low activity. In light of this, an engineered substrate promiscuous $p450_{BM}$, an enzyme from *Bacillus Megaterium* that display high kinetic efficiency in *E. coli*, was used to convert amorphadiene to artemisinic epoxide (section 2.5.3) [150]. In contrast, *S. cerevisiae* was a better host to overexpress cytochrome p450 [211]; yeast whole cell biocatalysis has been applied for the production of many important fine chemicals [206]. CYP71AV1 showed a very high catalytic efficiency when overexpressed in yeast and artemisinic acid was effectively produced. This could be due to the difference in the intracellular redox environment between *E. coli* and *S. cerevisiae* [212]. The conversion rate of CYP71AV1 was reduced at high initial concentration of AD,

possibly due to the saturation of catalytic sites. Therefore, controlled pulse addition of AD could be explored to overcome the challenge. Moreover, the enzymatic yield could be increased when high copy number plasmid was used to overexpress the enzyme. 80% conversion yield was achieved from 1 g/L AD, which corresponded to 800mg/L artemisinic acid produced in a small scale. The titer was comparable to the highest yield reported to date and that was produced in a shaking flask [3]. The productivity was expected to be further improved by co-expressing the auxiliary enzymes, Adh1 and Aldh1.

6.4 Conclusion

In summary, a hybrid *in vivo* and *in vitro* system was established to produce dihydroartemisinic acid and/or artemisinic acid, by which the production of amorpha-4,11-diene (AD) would not be affected by the downstream overexpression of cytotoxic P450 enzyme. The conversion of AD to downstream oxidized products was systematic optimized through enhanced protein expression and choice of more robust yeast strain. So far 80% of AD can be effectively converted to downstream acid, which will be further optimized by introducing auxiliary enzymes.

6.5 Materials and Methods

6.5.1 Bacteria strains and plasmids

The microbial strains and plasmids used in this study were summarized in Table 6.1. The pET-11a (Stratagene, CA) was modified by replacing the ampicillin with kanamycin maker. Adh1, Aldh1 and Dbr2 genes were codon optimized and synthesized by Genescript, and they were cloned into the modified pETK-11a plasmid by CLIVA method as described in Chapter 4 [186]. To modify the N-terminal of CYP15, different PCR primers were designed to target at different site of the open-reading frame. Then CLIVA was used to ligate the truncated gene into pCL plasmid. The CYP71AV1 with the first 15 amino acid being truncated displayed the highest enzymatic activity. Therefore, the corresponding gene was amplified by the primers (Table 6.2), and ligated into shuttle vector p416-TEF, and pYES-Gal1 respectively. Both plasmids carried the Ura3 auxotrophic marker. The plasmids were transformed into and amplified in *E. coli* DH10B strain. Then the purified plasmid was transformed into the yeast strains with lithium acetate mediated method developed by Gietz *et al.* [213]. The colonies were formed after incubating at 30^{o}C for 3-4 days. Then colony PCR was carried out to confirm the presence of the plasmid. The correct colony was then grown in selection medium lacking uracil till stationary phase for whole cell biocatalysis.

Table 6.1. The bacteria strains and plasmids used in Chapter 6.

Strain	Genotype
Mg1655(DE3)	F^-, $lambda^-$, rph-1, λ(DE3)
W303	MATa/MATα leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 [phi^+]
dCWP1	BY4741; Mata; his3D1; leu2D0; met15D0; ura3D0; YKL096w::kanMX4
dCWP2	BY4741; Mata; his3D1; leu2D0; met15D0; ura3D0; YKL096w-a::kanMX4
dErg6	BY4741; Mata; his3D1; leu2D0; met15D0; ura3D0; YML008c::kanMX4
BY4741	MATa; his3Δ 1; leu2Δ 0; met15Δ 0; ura3Δ 0
p416-TEF-CYP15	Low copy plasmid used to overexpress CYP15 in yeast
pYES-Gal1-CYP15	High copy plasmid used to overexpress CYP15 in yeast
pETK-Adh1	Plasmid used to overexpress Adh1 in Mg1655 (DE3)
pETK-Aldh1	Plasmid used to overexpress Adh1 in Mg1655 (DE3)
pETK-Dbr2	Plasmid used to overexpress Dbr2 in Mg1655 (DE3)
pCL-CYP15	Plasmid used to overexpress CYP15 in *E. coli* strains

Table 6.2. The primers used in Chapter 6.

Primer Name	Sequence
pYES-TT-f	CGAGAATCTTTATTTTCAGGG
pYES-Pro-r	CTTAATATTCCCTATAGTGAGTCGT
CYPaa-CPRaa-RBSf	CGATTAAATAAGGAGGAATAACAT
CYPaa-CPRaa-r	TCTGTTTTATCAGACCGCTTCT
p416-TEF-f	TCATGTAATTAGTTATGTCACGCTT
p416-TEF-r	AAACTTAGATTAGATTGCTATGCTTT

6.5.2 Yeast growth and protein expression

10ml culture of was grown in synthetic medium lacking uracil (20g/L Glucose, 6.7g/L drop-out base (USbiological, MA), 2g/L drop-out mix without uracil (USbiological, MA)), till stationary phase at 28^{o}C, 300rpm in an incubator-shaker (Shin Saeng Shaking Incubator, Finetech, Korea). The yeast growth was monitored by the optical density at A_{600}. For cells harbouring p41-TEF plasmid, the culture was directly harvested for downstream biotransformation, since TEF is a constitutive promoter. For cells harbouring pYES-Gal1 plasmid, CYP15 expression was induced with one step change of glucose synthetic medium to 2% galactose synthetic medim (20g/L Galactose, 6.7g/L drop-out base (USbiological, MA), 2g/L drop-out mix without uracil (USbiological, MA)). In brief, the cells were grown in glucose medium till OD_{600} reach approximatel 2. Subsequently, the cells were harvested by a brief centrifugation at 3000 g for 5min. The supernatant was carefully decanted and the culture was diluted with fresh galactose medium to OD_{600} = 0.5. Then, the culture was allowed to grow at 28^{o}C for 6h till OD_{600} is about 2 and used for subsequent biocatalysis.

6.5.3 Amorpha-4,11-diene purification

The microbial production of amorphadiene (AD) via mevalonate pathway has been optimized by promoter engineering and medium optimization (unpublished work). Approximately 2-4g/L of AD could be produced by *E. coli* Mg1655 strain carrying the mevalonate pathway genes and Ads gene in a 100ml culture growing in 2xPY medium supplemented with 10g/L glucose in 3 days.

The production was initiated with 0.01mM IPTG when the cell has grown till OD_{600}=0.6, and 2g C18 beads were added simultaneously to capture the AD produced. At the end of production, the beads were allowed to sediment, and the bacteria culture was removed by carefully discarding the medium containing the cells. Then the beads were resuspend in deionized water, and packed into a chromatography column (Bio-Rad, CA). The beads were washed with 3 bed volumes of deionized water twice to remove any bacterial contamination. Then the AD was eluted with equal bed volume of hexane (~1ml) three times. It was then subjected to GCMS analysis to determine the extraction efficiency. Usually, more than 50% capture AD could be eluted with first hexane elution. The hexane was further evaporated by vacuum concentrator, and AD was mixed with DMSO which would be miscible for whole cell Biocatalysis.

6.5.4 Yeast whole cell Biocatalysis and product extraction

Approximately 1ml, OD_{600} 2, yeast cells were collected and mixed with AD in a 100µl reaction containing 100mM Tris-HCL, 10mM $MgCl_2$ and additives as indicated in the text. The reaction was carried out in a 2ml GC vial, sealed with screwed cap to prevent evaporation of AD. The AD evaporation was minimized when it was dissolved in DMSO and mixed with yeast cells, as similar level of AD was detected before and after incubation at 28^{o}C for 12h in the control reaction mixture. 20µl of the reaction mixture was drawn at designated time point. It was acidified with 10% 2M HCl, and 100µl of ethyl acetate containing 200mg/L octylbenzoic acid was added to extract amorphadiene and its oxidized

products. 5mm glass beads were added to lyse the yeast cells and facilitate the extraction process. After vortexing at room temperature for 10min, the organic solvent layer was separated from the aqueous phase by centrifugation at 15,000 *g* for 10min. Then organic solvent was derivatized with TMS-diazomethane (Sigma, MO) and 10% methanol and subjected to GCMS measurement.

6.5.5 GCMS analysis of AD, AOH, AO and AA

1 µl sample was injected into HP-5 column (Agilent Technologies 7890A gas chromatograph-mass spectrometry, Agilent, CA). The gas chromatography oven temperature was programed with an initial fast linear temperature increase of 80 ^{o}C/min from 80 ^{o}C to 200 ^{o}C, and then a gradual linear temperature increase of 20^{o}C/min from 200^{o}C to 300^{o}C and hold at 300 ^{o}C for another minute. Scan mode was used to detect the compounds. With that temperature profile, amorphadiene (AD), artemisinic alcohol (AOH), artemisinic aldehyde (AO) and artemisinic acid (AA) were eluted at 3.86min, 5.04min, 4.70min and 5.14min respectively. The peak area was calculated with the software provided by the manufacturer. AD concentrations were pre-determined relative to the internal standard trans-caryophyllene of known concentration. Then the conversion yield was calculated based on the peak areas of the compounds as Equation 6.1, and the yield of AA was determined based on the total amount of AD added at the start of the reaction.

$$\%\ conversion = \frac{Area_{AOH} + Area_{AO} + Area_{AA}}{Area_{AD} + Area_{AOH} + Area_{AO} + Area_{AA}} \times 100\%$$

Equation 6.1

Chapter 7 Multi-biocatalytic synthesis of 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MEC) via the non-mevalonate pathway

7.1 Introduction

The mevalonate-independent pathway or 1-deoxy-D-xylulose-5-phosphate (DXP) pathway (Figure 7.1) demonstrates the complexity and diversity of natural metabolic networks where two distinctive pathways are involved in synthesizing the same 5 carbon molecule, isopentenyl pyrophosphate (IPP). Both pathways co-exist in plants. Interestingly, the DXP pathway was found to be the sole pathway in unicellular green algae [214]. Similarly, many pathogenic bacteria can only synthesize essential metabolic products via the DXP pathway, including the quinone co-factors for respiration [215]. Therefore, the enzymes in this pathway can serve as multiple, valuable drug targets to screen and develop anti-bacterial agents [116, 216]. This concept has been validated with fosmidomycin, a well-known antibiotic, which is a potent inhibitor of 2-C-methylerythritol-4-phosphate reductase (Dxr), the second enzyme along the DXP pathway [217, 218].

The DXP pathway is superior stoichiometrically as compared to the MVA pathway in synthesizing isoprenoids [219]. It produces less by-products and the co-factors are better balanced [219]. Recent efforts in metabolic engineering to enhance the flux through the DXP pathway have identified new bottlenecks, which are the last two enzymatic steps catalyzed by (E)-4-hydroxy-3-methyl-but-2-enyl phyrophosphate synthase (IspG) and (E)-4-hydroxy-3-methyl-but-2-enyl phyrophosphate reductase (IspH) [42, 220, 221]. *In vitro* analysis showed that these two oxygen-sensitive enzymes shared many similar features, such as they both contain [4Fe-4S] prosthetic group [222-224]. In our lab, we have established

a robust and reliable liquid chromatography and mass spectrometry coupled assay to detect the DXP pathway intermediates *in vivo* and discovered that 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEC), the substrate of IspG, accumulated intracellularly and effluxed out of *E. coli* [42]. This provided the insight that endogenous IspG and IspH were kinetically inefficient in shuffling MEC towards isoprenoids production downstream. More intriguingly, the production of amorpha-4,11-diene, a sesquiterpene, was suppressed when IspG and IspH were overexpressed [186]. Therefore, understanding the mechanisms and limitations of these two enzymes would be valuable to overcome the bottlenecks and further boost up the productivity of the DXP pathway. In addition, a better understanding would facilitate the design of novel antimicrobial drug leads against IspG and IspH. In order to investigate the biochemical properties of the two enzymes, there is a need to produce large quantity of MEC, the natural substrate of IspG [225]. Both chemical synthesis [225] and free enzymatic synthesis methods [41] have been reported previously. However, the former method resulted in exceptionally low yield (<10% from the starting material) and the latter method appeared to be useful for small scale production and the enzymes used were non-recyclable thereby increasing the cost of production.

In this study, a recyclable multienzymatic platform to produce MEC *in vitro* by co-immobilized DXP pathway enzymes was explored. Initially, the production of MEC (~50% yield within 1hr) was found to be in a free multienzymatic system, consistent with previous reports [41]. An attempt was then made to produce MEC by co-immobilized DXP pathway enzymes and this

was found to perform poorly. Further investigation showed that immobilized Dxs was inactivated when immobilized. By co-immobilizing the DXP pathway starting from Dxr, high yield of MEC (~50% of DXP was converted to MEC) was observed within 10 minutes. The accumulation of DXP to a comparable level as MEC in both free and co-immobilized enzymatic formats spurred us to investigate possible regulations of Dxr. We found that the enzyme was highly reversible and sensitive to NADP+ concentration. By coupling Dxr with glucose dehydrogenase (GDH) that regenerates NADPH from glucose, almost all the DXP was eventually converted to MEP. Further optimization is still on-going to optimize the multienzymes system on the solid platform.

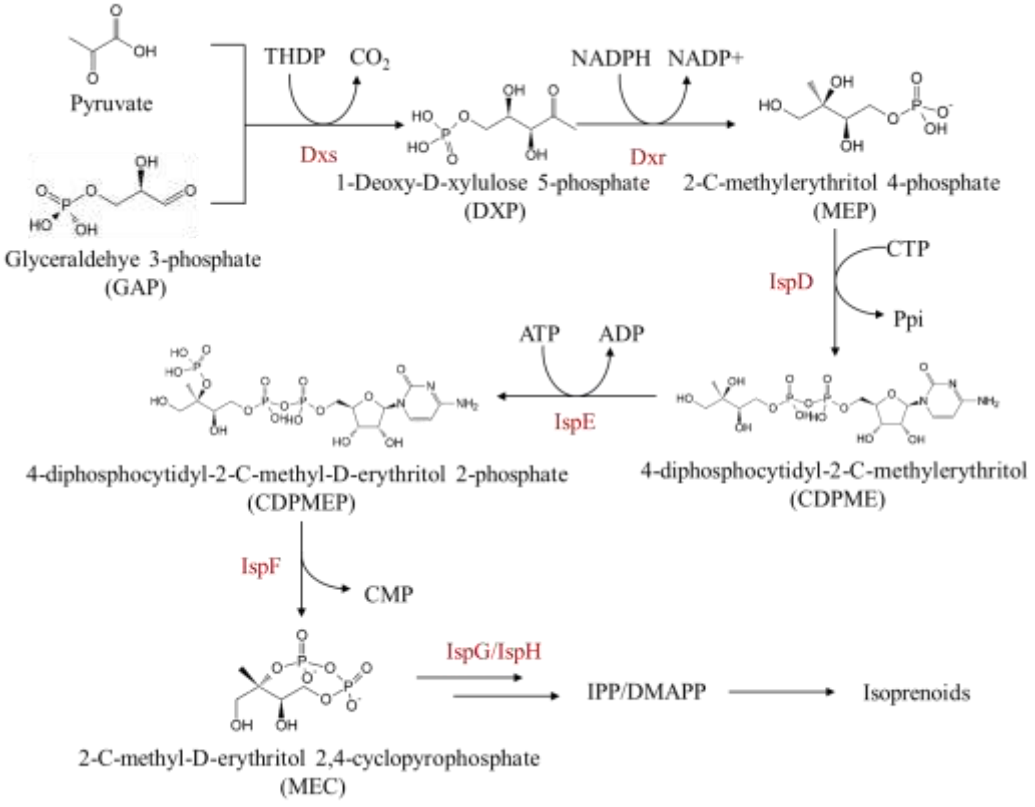

Figure 7.1. Schematic representation of the DXP pathway.

7.2 Results

7.2.1 Enzymatic purification and quantification

The genes encoding each enzyme were amplified from *E. coli* genomic DNA. They were subcloned into either pET-11a or pBAD-B vector with engineered N-terminal hexahistidine tags (His_6-tag). Solubility of the enzymes was examined to determine the optimal strain to overexpress the enzymes (Table 7.3) [20]. All the pathway enzymes were purified by immobilized metal affinity chromatography (IMCA) to near homogeneity, as validated by SDS-PAGE (results not shown). Initial velocity measurements were carried out to determine the steady-state kinetics of the enzymes (Table 7.1). As the substrates for 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) were not available, the kinetics parameters were extracted from literature [226, 227]. From the enzyme kinetics analyses, Dxs displayed the lowest catalytic efficiency among all the pathway enzymes, implying a potential kinetic bottleneck in the pathway. An initial attempt was made by mixing equal concentration of the five enzymes in one pot at 37^oC for 2 hours. MEC was successfully produced with an overall conversion of 50% from the starting materials (pyruvate and glyceraldehyde-3-phosphate). This means that on average, each step of enzymatic reaction was able to achieve about 90% conversion.

Table 7.1. Purification and characterization of individual pathway enzymes from bacterial culture.

Enzyme (synonym)	EC #	MW*	K_m/ μM	K_{cat}/ s^{-1}	Yield#/ mg/L
1-Deoxy-D-xylulose 5-phosphate synthase (Dxs)	2.2.1.7	67616.94	280 (Pyruvate) 467 (GAP)	1.3	8
2-C-methylerythritol-4-phosphate reductase (Dxr)	1.1.1.267	43387.98	198.4 (DXP) [228]	17.6	33
4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD)	2.7.7.60	25737.26	216.9 (MEP)	47.9	67
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE)	2.7.1.148	30925.37	*150 (CDP-ME) 420 (ATP)*	*16.8*	4
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF)	4.6.1.12	16897.54	*339*	*1.01*	33

* MW is calculated based on the primary amino acid sequence.
# Yield refers to the final protein yield after purification
The italic values were adapted from literature. IspE kinetics was assayed at 30°C [226]. IspF kinetics was assayed at 20°C [227].

7.2.2 Production of MEC by co-immobilized DXP pathway enzymes

The initial success with free multienzymatic synthesis prompted us to attempt to conduct the synthesis with co-immobilized multienzymes. Since all the DXP pathway enzymes have a his_6-tag, they can be anchored onto Ni-NTA functionalized surface. In Chapter 5, the established immobilization protocol was used here to immobilize the expressed enzymes directly from cell lysates. After

the cells were disrupted by glass beads, the lysate was prepared with approximately equal concentration of individual enzyme. The mixture was then diluted in phosphate saline buffer (PBS) and incubated with Ni-NTA modified silicon beads at 4^{o}C for 2 h. The unbounded protein was then washed away with PBS. In order to compare the performance of the immobilized multienzymes system with the free enzyme system, the beads with bound enzymes were divided into two equal parts. To one part, NPI400 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 400 mM imidazole, pH 8) was added to elute the enzymes from the beads, and to the other part, equal volume of NPI10 (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) was added. From Figure 7.2, almost all the enzymes were released from the beads (lane 3) in NPI400 buffer, whereas the enzymes were captured on beads in NPI10 buffer (lane 8). The amount of enzymes were comparable between the two systems.

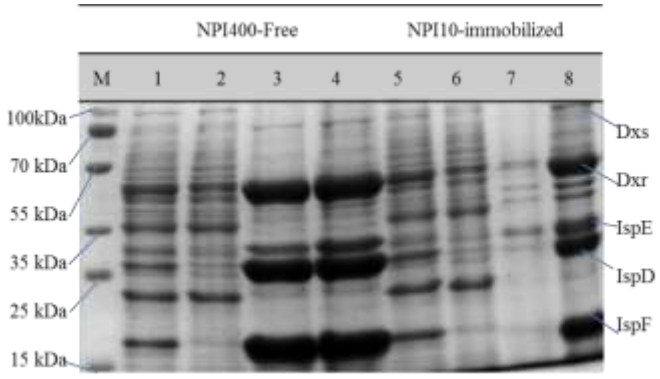

Figure 7.2 SDS PAGE of dxp pathway enzymes in free enzymatic and immobilized enzymatic formats.

Lane 1 and 5: cell lysate mixture before incubating with Ni-NTA beads. Lane 2 and 6, supernatant after incubating with Ni-NTA beads. Lane 3 and 7, enzymes in eluent after either NPI400 or NPI10 was added. Lane 4 and 8, enzymes in eluent and on beads after either NPI400 or NPI10 was added.

Synthesis of MEC was conducted at 37°C and pH 7.4 (100mM HEPES buffer) in a mixture (25µl) containing pyruvate (5mM), DL-GAP (10mM), ThDP (1mM), NADPH (5mM), CTP (5mM), ATP (5mM), $MgCl_2$ (10mM). The reaction was initiated by the addition of the pathway enzymes, either in free or immobilized enzymatic format. A time course study revealed that the synthesis of MEC saturated within 1 hour (Figure 7.3). Approximately 50%, namely 2.5mM of MEC was produced with free enzymes. However, less than 10% of MEC was produced with the immobilized enzymes. The extremely low yield prompted a further investigation into possible bottleneck(s) in the co-immobilized enzymatic system.

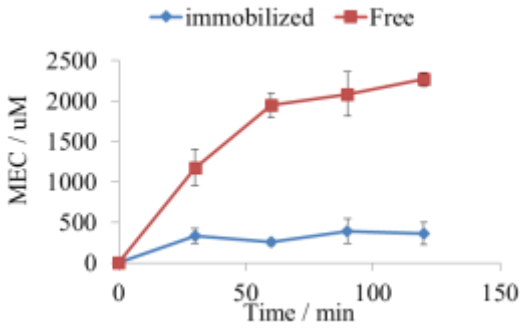

Figure 7.3. Time course of in vitro synthesis of MEC by free or co-immobilized DXP pathway enzymes.
All the measurements were done in triplicates with the standard errors drawn on the plots.

7.2.3 Immobilized Dxs activity was reduced

One hypothesis was that Dxs activity was reduced after immobilization, since it was the first committed step and possibly subjected to interfacial diffusion limitations [41, 191, 229]. To test the hypothesis, Dxs was immobilized onto the Ni-NTA beads, and was separated into free and immobilized format as described

above. Its reaction was conducted at the same reaction condition in the mixture of pyruvate (5mM), DL-GAP (10mM), ThDP (1mM) and $MgCl_2$ (10mM). The time course study clearly demonstrated that immobilized Dxs was approximately 10 times less efficient in catalysis as compared to free Dxs (Figure 7.4).

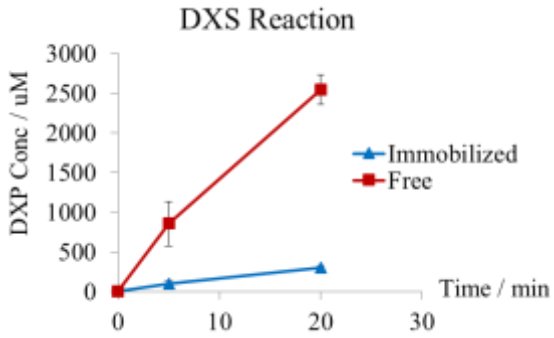

Figure 7.4. Time course of Dxs reaction in either free or immobilized form.
All the measurement were done in triplicates with the standard errors drawn on the plot.

This observation led us to investigate if it was due to mass transport limitation [191] or interfacial enzyme inactivation [230]. To decouple the two possible effects, the kinetic behavior of Dxs prior to and following immobilization was examined. The apparent Michaelis Menten parameters would provide insight into the plausible barrier on the reduced Dxs efficiency. Generally, the apparent K_m would increase if diffusion was a limitation, and V_{max} would decrease if the enzyme was inactivated. From Table 7.2, the maximum reaction rate, V_{max}, for immobilized Dxs was significantly lower than free Dxs (~10 fold lower), whereas K_m was slightly decreased after immobilization (~2 fold lower). Therefore, it strongly suggested that immobilized Dxs was less active at the liquid-solid interface.

Table 7.2. Comparison of the apparent kinetics of Dxs both in free and immobilized format.

	Free	Immobilized
K_m / µM	466.9	313.9
V_{max} / µM/min	275.1	24.2

Interestingly, no appreciable activity difference between immobilized and free IspD was observed (Figure 7.5A), suggesting that the reduced immobilized Dxs efficiency was likely an enzyme-dependent inactivation rather than a general mass transport phenomenon. Furthermore, when assembling the DXP pathway from Dxr (the second pathway enzyme) instead of Dxs, the production of MEC proceeded with similar rates between co-immobilized and free enzymes (Figure 7.5B). This observation provided further evidence that immobilized Dxs activity was severely impaired serving as the major bottleneck for the use of co-immobilized DXP pathway enzymes. Future work might involve bi-modular production of MEC, whereby DXP would be produced with soluble Dxs first, and then converted to MEC with downstream co-immobilized DXP pathway enzymes.

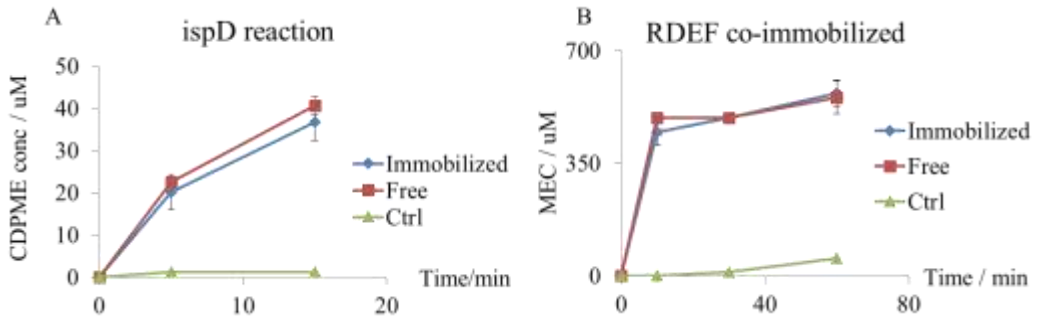

Figure 7.5. Time course of A. IspD reaction both in free and immobilized forms; B. assembled Dxr, IspD, IspE and ispF reaction both in free and co-immobilized forms.
All the measurements were done in triplicates and the standard errors were drawn on the plot.

7.2.4 DXP was accumulated in the multienzymes synthesis reaction

Another intriguing observation was that DXP was accumulated to a comparable level as MEC in the multienzymes reaction medium (Figure 7.6A). Moreover, Dxr was unable to convert the accumulated DXP in the multienzymes reaction mixture when additional purified Dxr (4µg) and NADPH (1mM) were supplemented (Figure 7.6B). The initial hypothesis was that DXP enantiomers were produced by substrate promiscuous Dxs [231]. However the hypothesis was rejected because about 50% chemically synthesized, enantiopure DXP was also left unreacted when it was used as the starting material. Hence, this alluded to the possible regulation of Dxr activity.

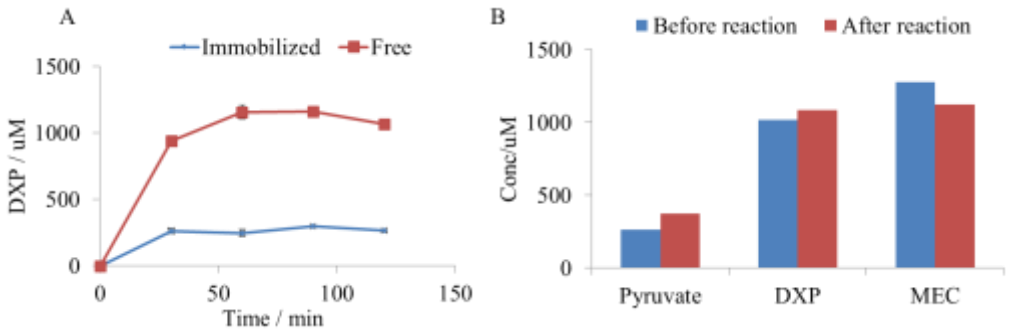

Figure 7.6 A. Time course of accumulation of DXP in the multienzyme reaction both in free and immobilized forms.
All the measurements were done in triplicates and the standard errors were drawn on the plot. B. The metabolite levels before and after adding fresh Dxr enzymes. It showed no significant difference in DXP level, suggesting that Dxr activity was inhibited by the mixture of metabolites.

To test the hypothesis, the products and by-products of the multienzymes pathway were supplied to the Dxr reaction separately. By monitoring the Dxr activity photometrically at 340nm, we noticed a slight drop in the initial rate of reaction when NADP+ was added (Figure 7.7A). NADP+ was the by-product of

Dxr, suggesting a possible reversible reaction mechanism. To verify the observation, glucose dehydrogenase (GDH), cloned from *Bacillus Megaterium* genomic DNA, was added into Dxr reaction to recycle NADPH. In the presence of GDH, catalytic amount of NADPH was sufficient to achieve almost 100% conversion of DXP to MEP, as shown in Figure 7.7B. However, without GDH, majority of DXP was left unreacted even when sufficient amount of NADPH was added into the reaction medium. Therefore, the accumulation of DXP in the multienzymes synthesis was due to the reversible nature of the Dxr reaction [228]. GDH is essential to recycle NADPH and drives the reaction to completion. Work is ongoing to further optimize the multienzymes reaction with GDH to enhance the MEC yield.

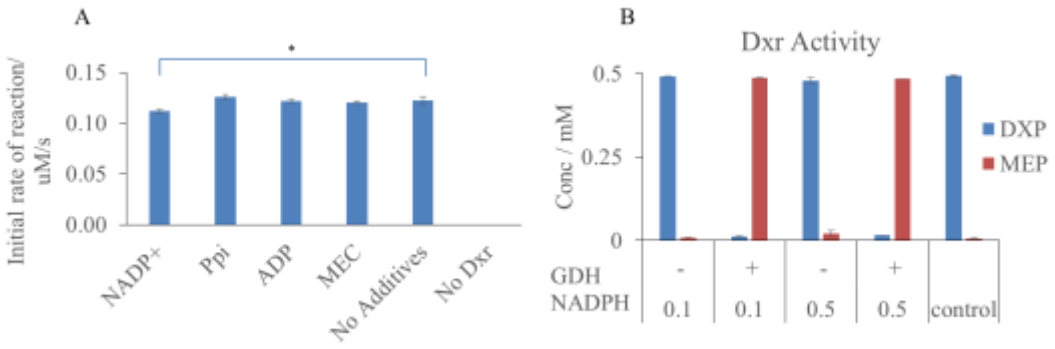

Figure 7.7. The Dxr reaction with downstream metabolites.
A. different downstream products and by-products supplemented into the reaction medium. The activity was monitored by absorbance at 340nm; B. coupled with or without gluose dehydrogenease and in low (0.1mM) or high (0.5mM) NADPH. All the measurements were done in triplicates and the standard errors were drawn on the plot.

7.3 Discussion

The DXP pathway has drawn tremendous interest recently, since the pathway enzymes are uniquely essential constituents of many pathogenic microorganisms, and thus are excellent targets for developing antimicrobial

agents [116, 227, 229]. MEC has been synthesized from glucose using purified enzymes [41, 55]. In this study, the aim was to explore the use of a co-immobilized multienzymes platform for both MEC biosynthesis and for future antimicrobial drug screening.

The initial hypothesis was to increase the enzymatic efficiency via co-immobilization and substrate channeling. Since the *in vitro* enzymes are diluted and dispersed, a stark contrast to nature, co-localization is a way to bring enzymes into close proximity and hence increases local substrate concentration [85, 232]. However, this proposal was not successfully realized as the first commited enzyme, Dxs, was severely inactivated at the liquid-solid surface. This was probably due to the incorrect orientation of the Dxs on the solid support. Since the his_6-tag was engineered at N-terminal of Dxs, where the active site is located, anchoring the enzyme onto the solid support might render the active site inaccessible to the reactants. Attempts have been made to re-engineer Dxs with C-terminal his_6-tag. However, little or no activity was detected when the C-terminal of Dxs was modified. An alternative solution would be to modify the solid support with inert spacer, such as polyethylene glycol [203]. Dxs was discovered more than a decade ago, and has been extensively characterized [233, 234]. It has a promiscuous substrate binding pocket that is able to accept a range of substrates [235, 236]. Moreover, detailed mechanistic insights have revealed that GAP and pyruvate exhibit substrate inhibition [237], and IPP is a product inhibitor of Dxs [238]. These studies suggest that Dxs is highly susceptible to inactivation and is a known bottleneck of the DXP pathway [131, 239]. More recently, an alternative

enzymatic route has been discovered in *E. coli* and Arabidopsis that phosphorylates 1-deoxy-D-xylulose to DXP [240, 241]. Moreover, in the field, there is some disagreement as to the identity of the first enzymatic reaction of the DXP pathway, as some suggested Dxr as the first committed step instead of Dxs [229, 238, 239, 242].

The accumulation of DXP in the multienzymes reaction spurred us to investigate whether Dxr was regulated by downstream metabolites. Similar to this observation is that we have observed DXP also accumulates *in vivo* by using the LCMS method developed [42]. DXP accumulation may also due to Dxr activity as it is highly reversible and the equilibrium is biased in favour of the substrates [228]. With the aid of glucose dehydrogenase, the NADP+ level can be kept at minimal level and shifting the equilibrium towards MEP formation.

Dxr is the second pathway enzyme, and fosmidomycin is a specific inhibitor of the enzyme. The drug has been demonstrated to cure rodent malaria caused by *Plasmodium vinckei* [243]. Interestingly, a new family of enzyme has been discovered in *Brucella abortus* that confers Dxr like (Drl) enzymatic activity[244]. This was likely the reason that certain bacteria have developed drug resistance through changes in biochemical pathway by re-routing and evolving new enzymatic functions. Therefore, a combination of antibacterial drugs that inhibit multiple enzymes simultaneously would limit the chance of developing drug resistance. However, to screen a library of compounds usually consists of hundreds of thousands of chemicals, the cost of preparing the target proteins would be high. Therefore, enzyme recycling and the use of miniaturized

devise, such as immobilized microfluidic enzyme reactor (IMER) [245], would have a great impact to reduce the cost incurred. We believe the technology developed here would have industrial importance and could be applied as a recyclable platform for drug screening.

7.4 Conclusion

In summary, we are in the process of developing a recyclable platform that co-immobilize the DXP pathway enzymes for MEC biosynthesis and antimicrobial drug screening. Since Dxs was found to be severely inactivated at the liquid-solid interface, we aimed to develop a two stage reaction whereby Dxs reaction would be performed in solution, and Dxr, ispD, ispE and ispF would be co-localized on Ni-NTA functionalized surface. With that approximately 50% yield was achieved within 1 hour. The yield can be further optimized with auxiliary reaction such as glucose dehydrogenase to drive the reversible Dxr reaction to completion.

7.5 Materials and Methods

7.5.1 Bacteria strains and plasmids

Bacteria strains and plasmids used in this study were summarized in Table 7.3. The DXP pathway enzymes, namely 1-Deoxy-D-xylulose 5-phosphate synthase (Dxs), 2-C-methylerythritol-4-phosphate reductase (Dxr), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE), and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) were amplified from *E. coli* genomic DNA with

his_6-tag. A 5' SacI site and a 3' XhoI site were introduced downstream from the 6xHis open reading frame. The PCR products were ligated into the modified pBAD-B (Invitrogen, CA) and pET-11a vector (Stratagene, CA) and transformed into competent *E. Coli* strain DH10B [20]. All the plasmids were transformed and harboured from *E. coli* DH10B and then transformed to respectively strains for enzyme overexpression (Table 7.3).

Table 7.3. The bacteria strains and plasmids used in Chapter 7.

Name	Description	Reference
E. coli BL21-Gold (DE3)	F^- ompT hsdS (r_B^- m_B^-) dcm^+ Tet^r gal λ(DE3) endA Hte	Stratagene
E. Coli DH10B	araD139 Δ(ara-leu)7697 fhuA lacX74 galK (Φ80 Δ(lacZ)M15) mcrA galU recA1 endA1 nupG rpsL Δ(mrr-hsdRMS-mcrBC)	NEB
E. Coli XL10-Gold	Tetr D(mcrA)183 D(mcrCB-hsdSMRmrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F9 proAB lacIqZDM15 Tn10 (Tetr) Tn5 (Kanr) Amy	Stratagene
pBAD-His-Dxs	Plasmid for overexpression of Dxs in *E. coli* XL10	
pET-His-Dxr	Plasmid for overexpression of Dxr in *E. coli* BL21	
pET-His-IspD	Plasmid for overexpression of IspD in *E. coli* BL21	[20]
pET-His-IspE	Plasmid for overexpression of IspE in *E. coli* BL21	
pET-His-IspF	Plasmid for overexpression of IspF in *E. coli* BL21	

7.5.2 Enzyme expression and purification

Newly transformed colonies were picked from the agar plate, inoculated into 2xPY medium (20 g/L Peptone, 10 g/L Yeast extract, and 10 g/L NaCl, pH 7) containing 100 mg/L ampicillin and grew till stationary phase overnight at 37 ^{o}C in an incubator-shaker (Shin Saeng Shaking Incubator, Finetech, Korea). The culture was then further transferred into fresh 2xPY medium (1% inoculation) with ampicillin for another 2.5 h at 37 ^{o}C, till optical density A_{600} reached 0.6-1.0. The enzyme expression was induced with 10mM L-arabinose for Dxs and 0.1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for the other four pathway enzymes. Temperature was reduced to 20 ^{o}C after induction for higher solubility of the enzymes [20]. The culture was grown for another 48 h and harvested by centrifugation. The cell pellets were stored at -20 ^{o}C till further use.

To purify the enzymes, the frozen cell pellets were resuspended in B-PERII reagent (Pierce, IL), according to the manufacturer's instruction, and vortexed at room temperature for 30 mins to completely lyse the cells. The soluble proteins were contained in the supernatant, which was diluted 15 times in NPI10 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) and incubated with 200 mg Ni-NTA resin (USB, Affymetrix, CA) at 4 ^{o}C for 2 h. The resin was washed with NPI10 buffer after discarding the binding supernatant, and the enzymes were eluted and collected by 400 μl NPI400 (50 mM NaH_2PO_4, 300 mM NaCl, 400 mM imidazole, pH 8). The enzymes were further concentrated by 3K Amicon ultra-0.5 ml centrifugal filter unit (Millipore, MA), and the concentrations were measured by Micro BCA protein assay kit (Thermo

scientific, MA). The purified enzymes were further confirmed by sodium dodecyl sulfate-12% polyacrylamide gel electrophoresis (Bio-Rad, CA).

7.5.3 Enzyme kinetics

The pathway enzyme kinetics was determined individually by initial rate measurements. In brief, the substrates and cofactors were added to 100 mM HEPES reaction buffer (pH 7.4), and the reaction was initiated by adding pre-determined enzyme amount to ensure less than 10% substrate was consumed in 15 mins at 37 ^{o}C. The substrates concentrations were varied in equal steps in reciprocal space from 0.1 to 1 mM. For Dxr, the enzymatic activity was monitored photometrically at 340nm. For Dxs and IspD, the reaction was terminated by adding equal volume of 1% ammonium hydroxide and diluted 10 times into cold methanol. After high speed centrifugation, the supernatant was subject to UPLC-(TOF)MS analysis. Double-reciprocal plots of each enzymatic activity were constructed for the determination of K_m and K_{cat} values for the respective substrates.

7.5.4 Multienzyme reaction

The multienzyme reaction, starting from Dxs, was carried out in a buffer (25 µl) consisted of HEPES (100 mM, pH 7.4), $MgCl_2$ (10 mM), (±)Glyceraldehyde 3-phosphate (10 mM), pyruvate (5mM), Thiamine pyrophosphate (ThPP, 0.5mM), NADPH (5mM), CTP (5mM), ATP (15 mM) and the purified enzymes. The reaction was performed at 37^{o}C. At the end of the reaction, the reaction was terminated by adding equal volume of 1% ammonium

hydroxide and diluted 10 times into cold methanol. The multienzymes reaction starting from Dxr was conducted at the same reaction condition in a buffer (25µl) consisted of HEPES (100 mM, pH 7.4), $MgCl_2$ (10 mM), 1-Deoxy-D-xylulose 5-phosphate (DXP, 1mM), NADPH (1mM), CTP (1mM), ATP (1 mM) and the purified enzymes.

To co-immobilize the pathway enzymes, the cells were disrupted by glass beads, and the cell lysate mixture was prepared with approximately equal enzyme concentration of individual enzyme. The mixture was then diluted in phosphate saline buffer (PBS) and incubated with Ni-NTA grafted silicon beads at 4^oC for 2 h. The unbounded protein was washed away with PBS. Subsequently, the beads with bound enzymes were divided into two equal parts. To one part, NPI400 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 400 mM imidazole, pH 8) was added to elute the enzymes from the beads, and to the other part, equal volume of NPI10 (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) was added. Then predetermined amount of enzymes were added into the reaction mixture to initiate the reaction.

7.5.5 UPLC-(TOF)MS analysis of DXP pathway intermediates

The analysis was done based on the method developed previously with slight modification [42]. In brief, 5 µl samples were injected into a UPLC C18 column (Waters CSH C18 1.7 µm, 2.1 mm x 50 mm) connected to UPLC (Waters ACQUITY UPLC)-(TOF)MS (Bruker micrOTOF II, MA). Elution was carried out with a step change from 100% aqueous solution containing 15 mM acetic acid and 10 mM tributylamine (0.5 min) to 10% aqueous solution with 90% methanol

for another 3.5 min. Electrospray ionization was used and mass spectrometry was operated to scan 50–800 m/z in negative mode with 2500 V end plate voltage and 3200 V capillary voltage. Nebulizer gas was provided in 2 bar, dry gas flow rate was 9 ml/min, and dry gas temperature was 220 ^{o}C. At the assay condition, all the intermediates were detected in the form $[M-H]^-$. Retention time was subsequently determined for each intermediate with respective synthetic standards and the set m/z extraction range. The peak area was calculated and subsequently used to compute the intermediates concentrations with the software provided by the manufacturer. The calibration curves were constructed with synthetic standards prepared under similar reaction conditions without enzymes. Linearity of the assays were determined individually with coefficients of determinants (R^2) greater than 0.90.

Chapter 8 Conclusion and recommendation of future works

8.1 General Conclusion

Cell free synthesis is a re-emerging powerful technology for biomolecules and small molecules production. In this thesis, production of isoprenoids and its precursors via cell-free synthesis has been demonstrated by assembling both the mevalonate and non-mevalonate pathways. A summary of main findings in the thesis is shown in Figure 8.1. A substantial portion of the thesis was dedicated to the optimization of amorpha-4,11-diene (AD) production using the mevalonate pathway. Through which, it repeatedly demonstrated that the key limiting step was amorpha-4,11-diene synthase (Ads). Firstly, its enzymatic turnover rate was more than 2 orders of magnitude lower than the other pathway enzymes. Secondly, its substrate farnesyl pyrophosphate precipitated in the presence of magnesium, an essential ion for the enzymatic reactions (chapter 3). Lastly, Ads activities were inhibited by high concentration of ATP and its by-product pyrophosphate (Ppi) (chapter 4 and 5). Equipped with thorough understanding of Ads behaviour, we devised specific strategies to improve Ads activities through buffer optimization and removal of inhibitors (ATP regeneration and pyrophosphatase to hydrolyse Ppi). Consequently, the production of AD was improved from less than 5% conversion in 12h to approximately 100% conversion in 4h, which approached the theoretical rate of reaction based on Ads activity. This was a remarkable enhancement of time and space yield. We further demonstrated the production of AD on solid platform with the intention to recycle the pathway enzymes. By designing a novel bi-modular reaction system, the

productivity of the solid platform was able to maintain after 8 cycles and approximately 2.2g/L AD could be produced in a small scale.

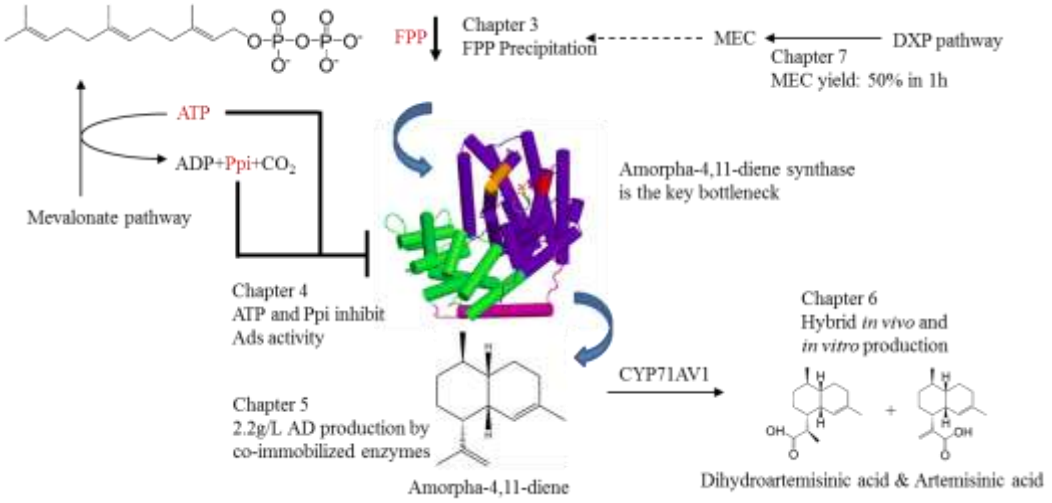

Figure 8.1 Schematic diagram summarizing the main findings in the thesis.
The work mainly focused on establishing and optimizing the mevalonate pathway *in vitro* to enhance amorpha-4,11-diene production. The optimization strategies applied are believed to have broad utility.

8.2 Future studies

8.2.1 Crystal structure of amorpha-4,11-diene synthase and rational protein engineering

As demonstrated in the thesis, Amorpha-4,11-diene synthase (Ads) is the main rate limiting step. Increasing its activity has shown to improve both *in vitro* and *in vivo* AD yield. Despite the improvement of its specific activity with buffer optimization, Ads still displayed remarkably low enzymatic efficiency. One plausible hypothesis was that the hydrophobic product was released slowly into aqueous medium. Doping the medium with amphiphilic substances such as detergent and ionic liquid was not found to be helpful. On the contrary, they prevented FPP from accessing the Ads catalytic site, since FPP is amphiphilic and would bind to detergent-like molecules. Therefore, rational protein engineering would be more desirable approach to improve Ads activity at molecular level. The structurally optimized Ads will be applicable to both *in vitro* and *in vivo* AD synthesis. To this end, structural information of Ads will provide useful information in determining the key amino acids that interact with both the substrate and product. Subsequently, *in silico* tools will be explored to aid re-modeling of the catalytic site of Ads, which will then be verified experimentally.

8.2.2 Increase Ads enzyme yield by **in vitro** *re-folding*

As shown in Chapter 3, Ads is entirely insoluble when overexpressed in *E. coli*. Although repeated freeze-thaw is able to release some soluble Ads into the lysis medium, majority of Ads enzymes were lost as insoluble aggregates. To

recover Ads from the inclusion bodies, *in vitro* protein re-folding is recommended. Initial attempt has been carried out, and found that commonly used chaotropic agent such as 8M urea was unable to recover Ads from the inclusion body. Instead, alkaline solution (pH=12.5) with 2M urea was used to partially dissolve Ads. Lowering urea concentration is favourable, since it will not totally disrupt the partially folded polypeptide chains which resemble native secondary structure [246]. Ionic liquid can be explored in the future to treat the inclusion body to enhance Ads polypeptide chain solubility. Subsequently, on column refolding will be carried out to in situ remove extraneous proteins and refold Ads. Combinations of natural and artificial chaperons as well as additives (Table 8.1) will be optimized to drive the equilibrium toward the correctly folded enzymes. The refolding efficiencies can be determined and compared by measuring the total activity recovery (Equation 8.1).

$$Total\ ativity\ recovery(\%) = \frac{m_r \times A_r}{m_t \times A_n} \times 100\%$$

Equation 8.1.Refolding efficiency is determined by the percentage of total activity recovered after refolding.
The abbreviations are as follows. m_r is the mass of refolded soluble protein. A_r is the specific activity of refolded protein. m_t is the total protein recovered from inclusion body. A_n is the specific activity of the native protein by standard purification.

Table 8.1 Lis of commonly used additives for protein refolding.

Additives	Concentrations used	Effect
Glycerol, sucrose	100mM to 2.5M	Stabilizer
L-Arginine	100mM to 2.5M	Stabilizer
Mg2+	75mM	Salt
GSSG/GSH	0.01mM to 10mM	Redox reagents

Cetyltrimethylammonium bromide (CTAB) + β-cyclodetrin (β-CD)	2.4mM (CTAB) 5mM (β-CD)	Artificial chaperone [247]
ClpB, DnaK, DnaJ, GrpE	-	Chaperones [248]

8.2.3 Scale-up cell free synthesis

The application of cell free synthesis has always been questioned for its capability to scale up, due to the cost incurred to purify enzymes and the use of expensive co-factors. The recent breakthrough demonstrated by Zawada et al. [249] quelled the myth. They have successfully scaled up cell-free synthesis to 100L by using cell lysates and produced 700mg/L human granulocyte macrophage colony-stimulating factor. This was the first ever industrially relevant cell-free protein synthesis reaction. Therefore, in the future, semi-*in vitro* synthesis with cell lysates containing overexpressed enzymes will be used to scale-up the multienzymes biosynthesis. It allows a higher degree of freedom to manipulate the production pathway enzymes than in a living cells, since cell viability is no longer a limiting factor. The balancing of pathway components can be better controlled without high cost incurred. Preliminary data suggested that AD can be synthesized through semi-*in vitro* method. The yield can be further improved when Ads activity is optimized. Accurate quantification and manipulation of enzyme amount will be realized via advanced mass spectroscopy technology. A combination of chaperons can be co-expressed with the pathway enzymes to enhance their stability. To avoid dilution of the pathway enzymes, they can be co-localized by facile means such as IMAC to increase the local

enzyme concentrations. Specific biochemical pathways can be activated to regenerate co-factors. Moreover, since the product formed could be in situ removed by the second organic layer (e.g., dodecane), continuous feeding of substrates to the cell lysates mixture is feasible. Thus, the cell lysate mixture can be effectively recycled.

8.2.4 DNA-directed assembly of multienzymes

DNA directed assembly technology is advancing by leaps and bounds. With the drastic improvement of DNA synthesis technology, using DNA as a novel biomaterial has become a reality. Due to the defined dimension of oligonucleotide and specific base pairing, it is possible to generate almost any geometries and structures. However, the real application of DNA directed assembly technology is still limited. Despite the ample applications of DNA microarray, DNA directed assembly of cascade enzyme reactions is still in its infancy. Therefore, there is a great potential to adapt our *in vitro* system onto the platform. We have shown that by dividing the pathway enzymes into two modules would enhance the productivity of the system. Tuning the inter-enzyme distance at submicron level would result in much higher precision and is expected to further enhance the specific yield of the system.

Bibliography

1. Ajikumar, P.K., et al., *Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms.* Mol Pharm, 2008. **5**(2): p. 167-90.
2. Kim, S.W., et al., *Identification of an alternative translation initiation site for the Pantoea ananatis lycopene cyclase (crtY) gene in E. coli and its evolutionary conservation.* Protein Expr Purif, 2008. **58**(1): p. 23-31.
3. Paddon, C.J., et al., *High-level semi-synthetic production of the potent antimalarial artemisinin.* Nature, 2013.
4. Ajikumar, P.K., et al., *Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-4.
5. Kwok, R., *Five hard truths for synthetic biology.* Nature, 2010. **463**(7279): p. 288-90.
6. Cheng, Q., et al., *Enzymatic total synthesis of enterocin polyketides.* Nat Chem Biol, 2007. **3**(9): p. 557-8.
7. Harris, D.C. and M.C. Jewett, *Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry.* Curr Opin Biotechnol, 2012. **23**(5): p. 672-8.
8. Santacoloma, P.A., et al., *Multienzyme-Catalyzed Processes: Next-Generation Biocatalysis.* Org Process Res Dev, 2011. **15**(1): p. 203-212.
9. Zhang, Y.H., et al., *High-yield hydrogen production from starch and water by a synthetic enzymatic pathway.* PLoS One, 2007. **2**(5): p. e456.
10. Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs over the 30 years from 1981 to 2010.* J Nat Prod, 2012. **75**(3): p. 311-35.
11. Gershenzon, J. and N. Dudareva, *The function of terpene natural products in the natural world.* Nat Chem Biol, 2007. **3**(7): p. 408-14.
12. Withers, S.T. and J.D. Keasling, *Biosynthesis and engineering of isoprenoid small molecules.* Appl Microbiol Biotechnol, 2007. **73**(5): p. 980-90.
13. Klein-Marcuschamer, D., P.K. Ajikumar, and G. Stephanopoulos, *Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene.* Trends Biotechnol, 2007. **25**(9): p. 417-24.
14. Barkovich, R. and J.C. Liao, *Metabolic engineering of isoprenoids.* Metab Eng, 2001. **3**(1): p. 27-39.
15. Lee, S.Y., D. Mattanovich, and A. Villaverde, *Systems metabolic engineering, industrial biotechnology and microbial cell factories.* Microb Cell Fact, 2012. **11**: p. 156.
16. Santos, C.N. and G. Stephanopoulos, *Combinatorial engineering of microbes for optimizing cellular phenotype.* Curr Opin Chem Biol, 2008. **12**(2): p. 168-76.
17. Alper, H. and G. Stephanopoulos, *Global transcription machinery engineering: a new approach for improving cellular phenotype.* Metab Eng, 2007. **9**(3): p. 258-67.

18. Dueber, J.E., et al., *Synthetic protein scaffolds provide modular control over metabolic flux.* Nat Biotechnol, 2009. **27**(8): p. 753-9.
19. Wilkinson, B. and J. Micklefield, *Mining and engineering natural-product biosynthetic pathways.* Nat Chem Biol, 2007. **3**(7): p. 379-86.
20. Zhou, K., et al., *Enhancing solubility of deoxyxylulose phosphate pathway enzymes for microbial isoprenoid production.* Microb Cell Fact, 2012. **11**: p. 148.
21. Fecik, R.A., *Natural product biosynthesis moves in vitro.* Nat Chem Biol, 2007. **3**(9): p. 531-2.
22. Balibar, C.J., A.R. Howard-Jones, and C.T. Walsh, *Terrequinone A biosynthesis through L-tryptophan oxidation, dimerization and bisprenylation.* Nat Chem Biol, 2007. **3**(9): p. 584-92.
23. Hodgman, C.E. and M.C. Jewett, *Cell-free synthetic biology: thinking outside the cell.* Metab Eng, 2012. **14**(3): p. 261-9.
24. Schmid, A., et al., *Industrial biocatalysis today and tomorrow.* Nature, 2001. **409**(6817): p. 258-268.
25. Zaks, A., *Industrial biocatalysis.* Current Opinion in Chemical Biology, 2001. **5**(2): p. 130-136.
26. Bartlett, J.M. and D. Stirling, *A short history of the polymerase chain reaction.* Methods Mol Biol, 2003. **226**: p. 3-6.
27. Srinivasan, A., et al., *Bacterial P450-catalyzed polyketide hydroxylation on a microfluidic platform.* Biotechnol Bioeng, 2004. **88**(4): p. 528-35.
28. Kwon, S.J., et al., *High-throughput, microarray-based synthesis of natural product analogues via in vitro metabolic pathway construction.* ACS Chem Biol, 2007. **2**(6): p. 419-25.
29. Patel, R.N., A. Banerjee, and V.V. Nanduri, *Enzymatic acetylation of 10-deacetylbaccatin III to baccatin III by C-10 deacetylase from Nocardioides luteus SC 13913.* Enzyme Microb Technol, 2000. **27**(6): p. 371-375.
30. Schoemaker, H.E., D. Mink, and M.G. Wubbolts, *Dispelling the myths - Biocatalysis in industrial synthesis.* Science, 2003. **299**(5613): p. 1694-1697.
31. Kwon, S.J., et al., *Expanding nature's small molecule diversity via in vitro biosynthetic pathway engineering.* Current Opinion in Chemical Biology, 2012. **16**(1-2): p. 186-195.
32. Ricca, E., B. Brucher, and J.H. Schrittwieser, *Multi-Enzymatic Cascade Reactions: Overview and Perspectives.* Advanced Synthesis & Catalysis, 2011. **353**(13): p. 2239-2262.
33. Monti, D., et al., *One-Pot Multienzymatic Synthesis of 12-Ketoursodeoxycholic Acid: Subtle Cofactor Specificities Rule the Reaction Equilibria of Five Biocatalysts Working in a Row.* Advanced Synthesis & Catalysis, 2009. **351**(9): p. 1303-1311.
34. Croteau, R., et al., *Taxol biosynthesis and molecular genetics.* Phytochem Rev, 2006. **5**(1): p. 75-97.

35. Panke, S., M. Held, and M. Wubbolts, *Trends and innovations in industrial biocatalysis for the production of fine chemicals.* Current Opinion in Biotechnology, 2004. **15**(4): p. 272-279.
36. Muthana, M.M., et al., *Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from Bifidobacterium longum (BLUSP).* Chem Commun (Camb), 2012. **48**(21): p. 2728-30.
37. Kuberan, B., et al., *Enzymatic synthesis of antithrombin III-binding heparan sulfate pentasaccharide.* Nat Biotechnol, 2003. **21**(11): p. 1343-6.
38. van der Donk, W.A. and H. Zhao, *Recent developments in pyridine nucleotide regeneration.* Curr Opin Biotechnol, 2003. **14**(4): p. 421-6.
39. Lopez-Gallego, F. and C. Schmidt-Dannert, *Multi-enzymatic synthesis.* Curr Opin Chem Biol, 2010. **14**(2): p. 174-183.
40. Santacoloma, P.A., et al., *Multienzyme-Catalyzed Processes: Next-Generation Biocatalysis.* Organic Process Research & Development, 2011. **15**(1): p. 203-212.
41. Schuhr, C.A., et al., *Studies on the non-mevalonate pathway - Preparation and properties of isotope-labeled 2C-methyl-D-erythritol 2,4-cyclodiphosphate.* Eur J Org Chem, 2001(17): p. 3221-3226.
42. Zhou, K., et al., *Metabolite profiling identified methylerythritol cyclodiphosphate efflux as a limiting step in microbial isoprenoid production.* PLoS One, 2012. **7**(11): p. e47513.
43. Xiao, Y., et al., *IspG-catalyzed positional isotopic exchange in methylerythritol cyclodiphosphate of the deoxyxylulose phosphate pathway: mechanistic implications.* Chembiochem, 2011. **12**(4): p. 527-30.
44. Muthana, M.M., et al., *Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from Bifidobacterium longum (BLUSP).* Chemical Communications, 2012. **48**(21): p. 2728-2730.
45. Roessner, C.A., et al., *Genetically engineered synthesis of precorrin-6x and the complete corrinoid, hydrogenobyrinic acid, an advanced precursor of vitamin B12.* Chem Biol, 1994. **1**(2): p. 119-24.
46. Gaitatzis, N., B. Kunze, and R. Muller, *In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11136-41.
47. Wang, Y., et al., *Biohydrogenation from biomass sugar mediated by in vitro synthetic enzymatic pathways.* Chem Biol, 2011. **18**(3): p. 372-80.
48. Bugg, T.D. and C.T. Walsh, *Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance.* Nat Prod Rep, 1992. **9**(3): p. 199-215.
49. El Zoeiby, A., et al., *In vitro reconstruction of the biosynthetic pathway of peptidoglycan cytoplasmic precursor in Pseudomonas aeruginosa.* FEMS Microbiol Lett, 2001. **201**(2): p. 229-35.

50. Maeda, H. and N. Dudareva, *The shikimate pathway and aromatic amino Acid biosynthesis in plants.* Annu Rev Plant Biol, 2012. **63**: p. 73-105.
51. Noble, M., et al., *The kinetic model of the shikimate pathway as a tool to optimize enzyme assays for high-throughput screening.* Biotechnol Bioeng, 2006. **95**(4): p. 560-73.
52. Yu, X.Y., et al., *In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(46): p. 18643-18648.
53. Sun, Y., et al., *In vitro reconstruction of tetronate RK-682 biosynthesis.* Nat Chem Biol, 2010. **6**(2): p. 99-101.
54. Shimizu, Y., et al., *Cell-free translation reconstituted with purified components.* Nat Biotechnol, 2001. **19**(8): p. 751-5.
55. Illarionova, V., et al., *Nonmevalonate terpene biosynthesis enzymes as antiinfective drug targets: substrate synthesis and high-throughput screening methods.* J Org Chem, 2006. **71**(23): p. 8824-34.
56. Vasic-Racki, D., U. Kragl, and A. Liese, *Benefits of enzyme kinetics modelling.* Chemical and Biochemical Engineering Quarterly, 2003. **17**(1): p. 7-18.
57. Maiwald, T., et al., *Dynamic pathway modeling: feasibility analysis and optimal experimental design.* Ann N Y Acad Sci, 2007. **1115**: p. 212-20.
58. Valdramidis, V.P., et al., *Development of predictive modelling approaches for surface temperature and associated microbiological inactivation during hot dry air decontamination.* Int J Food Microbiol, 2005. **100**(1-3): p. 261-74.
59. Swinnen, I.A., et al., *Quantifying microbial lag phenomena due to a sudden rise in temperature: a systematic macroscopic study.* Int J Food Microbiol, 2005. **100**(1-3): p. 85-96.
60. Singh, S.K., et al., *Comparative one-factor-at-a-time, response surface (statistical) and bench-scale bioreactor level optimization of thermoalkaline protease production from a psychrotrophic Pseudomonas putida SKG-1 isolate.* Microb Cell Fact, 2011. **10**: p. 114.
61. Gaia Franceschini, S.M., *Model-based design of experiments for parameter precision: State of the art.* Chem Eng Sci, 2008.
62. Rao, R.S., et al., *The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal.* Biotechnol J, 2008. **3**(4): p. 510-23.
63. Prakasham, R.S., et al., *Enhancement of acid amylase production by an isolated Aspergillus awamori.* J Appl Microbiol, 2007. **102**(1): p. 204-11.
64. Prakasham, R.S., et al., *Alkaline protease production by an isolated Bacillus circulans under solid-state fermentation using agroindustrial waste: process parameters optimization.* Biotechnol Prog, 2005. **21**(5): p. 1380-8.
65. Venkata Mohan, S. and M. Venkateswar Reddy, *Optimization of critical factors to enhance polyhydroxyalkanoates (PHA) synthesis by mixed*

culture using Taguchi design of experimental methodology. Bioresour Technol, 2013. **128**: p. 409-16.
66. Sirisansaneeyakul, S., et al., *Optimization of lactic acid production by immobilized Lactococcus lactis IO-1.* J Ind Microbiol Biotechnol, 2007. **34**(5): p. 381-91.
67. Pignatiello, J.J., *An Overview of the Strategy and Tactics of Taguchi.* Iie Transactions, 1988. **20**(3): p. 247-254.
68. Aggarwal, A., et al., *Optimizing power consumption for CNC turned parts using response surface methodology and Taguchi's technique—A comparative analysis.* Journal of Materials Processing Technology, 2008. **200**(1–3): p. 373-384.
69. Teng, Y. and Y. Xu, *Culture condition improvement for whole-cell lipase production in submerged fermentation by Rhizopus chinensis using statistical method.* Bioresource Technology, 2008. **99**(9): p. 3900-3907.
70. Giersch, C., *Mathematical modelling of metabolism.* Current Opinion in Plant Biology, 2000. **3**(3): p. 249-253.
71. Chou, I.C. and E.O. Voit, *Recent developments in parameter estimation and structure identification of biochemical and genomic systems.* Math Biosci, 2009. **219**(2): p. 57-83.
72. Morohashi, M., et al., *Robustness as a Measure of Plausibility in Models of Biochemical Networks.* Journal of Theoretical Biology, 2002. **216**(1): p. 19-30.
73. Ciliberto, A., F. Capuani, and J.J. Tyson, *Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation.* PLoS Comput Biol, 2007. **3**(3): p. e45.
74. Savageau, M.A., *Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions.* Journal of Theoretical Biology, 1969. **25**(3): p. 365-369.
75. Hill, A.V., *The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves.* J physiol, 1910. **40**(4): p. iv-vii.
76. Segel, I.H., *Enzyme kinetics*. 1993: Wiley New York.
77. Monod, J., J. Wyman, and J.-P. Changeux, *On the nature of allosteric transitions: A plausible model.* Journal of Molecular Biology, 1965. **12**(1): p. 88-118.
78. Ku, B., et al., *Preparation, characterization, and optimization of an in vitro C30 carotenoid pathway.* Appl Environ Microbiol, 2005. **71**(11): p. 6578-83.
79. Visser, D. and J.J. Heijnen, *Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics.* Metab Eng, 2003. **5**(3): p. 164-76.
80. Wu, L., et al., *A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics.* Eur J Biochem, 2004. **271**(16): p. 3348-59.
81. Costa, R.S., et al., *Hybrid dynamic modeling of Escherichia coli central metabolic network combining Michaelis-Menten and approximate kinetic equations.* Biosystems, 2010. **100**(2): p. 150-157.

82. del Rosario, R.C.H., E. Mendoza, and E.O. Voit, *Challenges in lin-log modelling of glycolysis in Lactococcus lactis.* Iet Systems Biology, 2008. **2**(3): p. 136-U30.
83. Bulik, S., et al., *Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws – a promising method for speeding up the kinetic modelling of complex metabolic networks.* FEBS Journal, 2009. **276**(2): p. 410-424.
84. Hanefeld, U., L. Gardossi, and E. Magner, *Understanding enzyme immobilisation.* Chemical Society Reviews, 2009. **38**(2): p. 453-468.
85. Wilner, O.I., et al., *Enzyme cascades activated on topologically programmed DNA scaffolds.* Nat Nanotechnol, 2009. **4**(4): p. 249-54.
86. Sheldon, R.A., *Enzyme Immobilization: The Quest for Optimum Performance.* Adv Synth Catal 2007. **349**: p. 1289-1307.
87. López-Serrano, P., et al., *Cross-linked enzyme aggregates with enhanced activity: application to lipases.* Biotechnology Letters, 2002. **24**(16): p. 1379-1383.
88. Mateo, C., et al., *Synthesis of enantiomerically pure (S)-mandelic acid using an oxynitrilase–nitrilase bienzymatic cascade: a nitrilase surprisingly shows nitrile hydratase activity.* Tetrahedron: Asymmetry, 2006. **17**(3): p. 320-323.
89. Castro, C.E., et al., *A primer to scaffolded DNA origami.* Nature Methods, 2011. **8**(3): p. 221-229.
90. Jung, G.Y. and G. Stephanopoulos, *A functional protein chip for pathway optimization and in vitro metabolic engineering.* Science, 2004. **304**(5669): p. 428-31.
91. Liu, Y., et al., *Aptamer-directed self-assembly of protein arrays on a DNA nanostructure.* Angew Chem Int Ed Engl, 2005. **44**(28): p. 4333-8.
92. Tsuji, S., et al., *RNA aptamer binding to polyhistidine-tag.* Biochem Biophys Res Commun, 2009. **386**(1): p. 227-31.
93. Chhabra, R., et al., *Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures.* J Am Chem Soc, 2007. **129**(34): p. 10304-5.
94. Jongsma, M.A. and R.H. Litjens, *Self-assembling protein arrays on DNA chips by auto-labeling fusion proteins with a single DNA address.* Proteomics, 2006. **6**(9): p. 2650-5.
95. Kampmeier, F., et al., *Site-specific, covalent labeling of recombinant antibody fragments via fusion to an engineered version of 6-O-alkylguanine DNA alkyltransferase.* Bioconjug Chem, 2009. **20**(5): p. 1010-5.
96. Duckworth, B.P., et al., *A universal method for the preparation of covalent protein-DNA conjugates for use in creating protein nanostructures.* Angew Chem Int Ed Engl, 2007. **46**(46): p. 8819-22.
97. Russell P. Goodman, C.M.E., Jonathan Malo, Wei M. Ho, Mireya L. McKee, Achillefs N. Kapanidis, Andrew J. Turberfield, *A Facile Method for Reversibly Linking a Recombinant Protein to DNA.* Chembiochem, 2009. **10**: p. 1551-1557.

98. David C. Wyllie, J.C.-T.T., *Immobilized metal affinity chromatography*. 1999, Schering Corporation, Kenilworth, N,J: United states.
99. Cha, T., A. Guo, and X.Y. Zhu, *Enzymatic activity on a chip: the critical role of protein orientation.* Proteomics, 2005. **5**(2): p. 416-9.
100. Sawasaki, T., et al., *Arabidopsis HY5 protein functions as a DNA-binding tag for purification and functional immobilization of proteins on agarose/DNA microplate.* FEBS Lett, 2008. **582**(2): p. 221-8.
101. Cassimjee, K.E., et al., *One-step enzyme extraction and immobilization for biocatalysis applications.* Biotechnology Journal, 2011. **6**(4): p. 463-469.
102. Wang, L.A., et al., *Specific and reversible immobilization of NADH oxidase on functionalized carbon nanotubes.* Journal of Biotechnology, 2010. **150**(1): p. 57-63.
103. Shao, J., et al., *Biocatalytic synthesis of uridine 5 '-diphosphate N-acetylglucosamine by multiple enzymes co-immobilized on agarose beads.* Chemical Communications, 2002(21): p. 2586-2587.
104. Liu, Z., et al., *Combined biosynthetic pathway for de novo production of UDP-galactose: catalysis with multiple enzymes immobilized on agarose beads.* Chembiochem, 2002. **3**(4): p. 348-55.
105. Knecht, S., et al., *Oligohis-tags: mechanisms of binding to Ni2+-NTA surfaces.* J Mol Recognit, 2009. **22**(4): p. 270-9.
106. You, C., et al., *Affinity capturing for targeting proteins into micro and nanostructures.* Analytical and Bioanalytical Chemistry, 2009. **393**(6-7): p. 1563-1570.
107. Olaofe, O.A., et al., *The influence of microbial physiology on biocatalyst activity and efficiency in the terminal hydroxylation of n-octane using Escherichia coli expressing the alkane hydroxylase, CYP153A6.* Microb Cell Fact, 2013. **12**(1): p. 8.
108. Zhang, W., et al., *Bioreduction with efficient recycling of NADPH by coupled permeabilized microorganisms.* Appl Environ Microbiol, 2009. **75**(3): p. 687-94.
109. Goerke, A.R., et al., *Cell-free metabolic engineering promotes high-level production of bioactive Gaussia princeps luciferase.* Metab Eng, 2008. **10**(3-4): p. 187-200.
110. Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnol Bioeng, 2004. **86**(1): p. 19-26.
111. Zawada, J.F., et al., *Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines.* Biotechnol Bioeng, 2011. **108**(7): p. 1570-8.
112. Eastman, R.T. and D.A. Fidock, *Artemisinin-based combination therapies: a vital tool in efforts to eliminate malaria.* Nat Rev Microbiol, 2009. **7**(12): p. 864-74.
113. Huang, C., et al., *Artemisinin rewires the protein interaction network in cancer cells: network analysis, pathway identification, and target prediction.* Molecular Biosystems, 2013. **9**(12): p. 3091-3100.

114. Paddon, C.J. and J.D. Keasling, *Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development.* Nature Reviews Microbiology, 2014. **12**(5): p. 355-367.
115. Brown, G.D., *The Biosynthesis of Artemisinin (Qinghaosu) and the Phytochemistry of Artemisia annua L. (Qinghao).* Molecules, 2010. **15**(11): p. 7603-7698.
116. Tidten-Luksch, N., et al., *IspE inhibitors identified by a combination of in silico and in vitro high-throughput screening.* PLoS One, 2012. **7**(4): p. e35792.
117. Handke, P., S.A. Lynch, and R.T. Gill, *Application and engineering of fatty acid biosynthesis in Escherichia coli for advanced fuels and chemicals.* Metab Eng, 2011. **13**(1): p. 28-37.
118. Henneman, L., et al., *Inhibition of the isoprenoid biosynthesis pathway; detection of intermediates by UPLC-MS/MS.* Biochim Biophys Acta, 2011. **1811**(4): p. 227-33.
119. Bloch, K., *The biological synthesis of cholesterol.* Science, 1965. **150**(3692): p. 19-28.
120. Miziorko, H.M., *Enzymes of the mevalonate pathway of isoprenoid biosynthesis.* Arch Biochem Biophys, 2011. **505**(2): p. 131-43.
121. Gray, J.C. and R.G. Kekwick, *The inhibition of plant mevalonate kinase preparations by prenyl pyrophosphates.* Biochim Biophys Acta, 1972. **279**(2): p. 290-6.
122. Hinson, D.D., et al., *Post-translational regulation of mevalonate kinase by intermediates of the cholesterol and nonsterol isoprene biosynthetic pathways.* J Lipid Res, 1997. **38**(11): p. 2216-23.
123. Primak, Y.A., et al., *Characterization of a feedback-resistant mevalonate kinase from the archaeon Methanosarcina mazei.* Appl Environ Microbiol, 2011. **77**(21): p. 7772-8.
124. Fu, Z., et al., *Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism.* Biochemistry, 2008. **47**(12): p. 3715-24.
125. Smit, A. and A. Mushegian, *Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway.* Genome Res, 2000. **10**(10): p. 1468-84.
126. Bloch, K., et al., *Mevalonic acid pyrophosphate and isopentenylpyrophosphate.* J Biol Chem, 1959. **234**: p. 2595-604.
127. Pilloff, D., et al., *The kinetic mechanism of phosphomevalonate kinase.* J Biol Chem, 2003. **278**(7): p. 4510-5.
128. Berges, T., D. Guyonnet, and F. Karst, *The Saccharomyces cerevisiae mevalonate diphosphate decarboxylase is essential for viability, and a single Leu-to-Pro mutation in a conserved sequence leads to thermosensitivity.* J Bacteriol, 1997. **179**(15): p. 4664-70.
129. Reardon, J.E. and R.H. Abeles, *Inhibition of cholesterol biosynthesis by fluorinated mevalonate analogues.* Biochemistry, 1987. **26**(15): p. 4717-22.

130. Durbecq, V., et al., *Crystal structure of isopentenyl diphosphate:dimethylallyl diphosphate isomerase.* EMBO J, 2001. **20**(7): p. 1530-7.
131. Yuan, L.Z., et al., *Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in E. coli.* Metab Eng, 2006. **8**(1): p. 79-90.
132. Wouters, J., et al., *Structure and Mechanism of Action of Isopentenylpyrophosphate-Dimethylallylpyrophosphate Isomerase.* Journal of the American Chemical Society, 2003. **125**(11): p. 3198-3199.
133. Thulasiram, H.V. and C.D. Poulter, *Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity.* J Am Chem Soc, 2006. **128**(49): p. 15819-23.
134. Reiling, K.K., et al., *Mono and diterpene production in Escherichia coli.* Biotechnol Bioeng, 2004. **87**(2): p. 200-12.
135. Ku, B.S., et al., *Preparation, characterization, and optimization of an in vitro C-30 carotenoid pathway.* Applied and Environmental Microbiology, 2005. **71**(11): p. 6578-6583.
136. Mitrofan, L.M., J. Pelkonen, and J. Monkkonen, *The level of ATP analog and isopentenyl pyrophosphate correlates with zoledronic acid-induced apoptosis in cancer cells in vitro.* Bone, 2009. **45**(6): p. 1153-60.
137. van Beek, E., et al., *Farnesyl pyrophosphate synthase is the molecular target of nitrogen-containing bisphosphonates.* Biochem Biophys Res Commun, 1999. **264**(1): p. 108-11.
138. Koksal, M., et al., *Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis.* Nature, 2011. **469**(7328): p. 116-20.
139. Wallaart, T.E., et al., *Amorpha-4,11-diene synthase: cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin.* Planta, 2001. **212**(3): p. 460-5.
140. Mercke, P., et al., *Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in Artemisia annua L.* Arch Biochem Biophys, 2000. **381**(2): p. 173-80.
141. Picaud, S., et al., *Amorpha-4,11-diene synthase: Mechanism and stereochemistry of the enzymatic cyclization of farnesyl diphosphate.* Archives of Biochemistry and Biophysics, 2006. **448**(1-2): p. 150-155.
142. Picaud, S., et al., *Expression, purification, and characterization of recombinant amorpha-4,11-diene synthase from Artemisia annua L.* Arch Biochem Biophys, 2005. **436**(2): p. 215-26.
143. Cane, D.E., et al., *Pre-steady-state kinetic analysis of the trichodiene synthase reaction pathway.* Biochemistry, 1997. **36**(27): p. 8332-9.
144. Lauchli, R., et al., *High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase.* Angew Chem Int Ed Engl, 2013. **52**(21): p. 5571-4.
145. Lishan Zhao, L.X., Patrick Westfall, Andrew Main, *Methods of developing terpene synthase variants.* Patent Application Publication US 20120196315A1, 2012.

146. Ro, D.K., et al., *Production of the antimalarial drug precursor artemisinic acid in engineered yeast.* Nature, 2006. **440**(7086): p. 940-3.
147. Komori, A., et al., *Comparative functional analysis of CYP71AV1 natural variants reveals an important residue for the successive oxidation of amorpha-4,11-diene.* FEBS Lett, 2012.
148. Zhang, Y., et al., *The production of artemisinin precursors in tobacco.* Plant Biotechnol J, 2011. **9**(4): p. 445-54.
149. Chang, M.C.Y., et al., *Engineering Escherichia coli for production of functionalized terpenoids using plant P450s.* Nature Chemical Biology, 2007. **3**(5): p. 274-277.
150. Dietrich, J.A., et al., *A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3).* ACS Chem Biol, 2009. **4**(4): p. 261-7.
151. Teoh, K.H., et al., *Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in Artemisia annua.* Botany-Botanique, 2009. **87**(6): p. 635-642.
152. Zhang, Y., et al., *The molecular cloning of artemisinic aldehyde Delta 11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in Artemisia annua.* Journal of Biological Chemistry, 2008. **283**(31): p. 21501-21508.
153. Grayson, M., *Malaria.* Nature, 2012. **484**(7395): p. S13.
154. Klayman, D.L., *Qinghaosu (artemisinin): an antimalarial drug from China.* Science, 1985. **228**(4703): p. 1049-55.
155. Van Noorden, R., *Demand for malaria drug soars.* Nature, 2010. **466**(7307): p. 672-673.
156. Roessner, C.A. and A.I. Scott, *Achieving natural product synthesis and diversity via catalytic networking ex vivo.* Chem Biol, 1996. **3**(5): p. 325-30.
157. Johnson, B.H. and M.H. Hecht, *Recombinant Proteins Can Be Isolated from Escherichia Coli Cells by Repeated Cycles of Freezing and Thawing.* Nat Biotechnol, 1994. **12**(13): p. 1357-1360.
158. Green, S., et al., *Defining the potassium binding region in an apple terpene synthase.* J Biol Chem, 2009. **284**(13): p. 8661-9.
159. Rao, R.S., et al., *Xylitol production by Candida sp.: parameter optimization using Taguchi approach.* Process Biochem, 2004. **39**(8): p. 951-956.
160. Anthony, J.R., et al., *Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene.* Metab Eng, 2009. **11**(1): p. 13-9.
161. Monti, D., et al., *One-Pot Multienzymatic Synthesis of 12-Ketoursodeoxycholic Acid: Subtle Cofactor Specificities Rule the Reaction Equilibria of Five Biocatalysts Working in a Row.* Adv Synth Catal, 2009. **351**(9): p. 1303-1311.

162. Westfall, P.J., et al., *Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.* Proc Natl Acad Sci U S A, 2012. **109**(3): p. E111-8.
163. Picaud, S., M.E. Olsson, and P.E. Brodelius, *Improved conditions for production of recombinant plant sesquiterpene synthases in Escherichia coli.* Protein Expr Purif, 2007. **51**(1): p. 71-9.
164. Hill, A.M., et al., *High level expression of Ricinus communis casbene synthase in Escherichia coli and characterization of the recombinant enzyme.* Arch Biochem Biophys, 1996. **336**(2): p. 283-9.
165. Lee, S. and C.D. Poulter, *Cloning, solubilization, and characterization of squalene synthase from Thermosynechococcus elongatus BP-1.* J Bacteriol, 2008. **190**(11): p. 3808-16.
166. Huang, K.X., et al., *Overproduction, in Escherichia coli, of soluble taxadiene synthase, a key enzyme in the Taxol biosynthetic pathway.* Protein Expr Purif, 1998. **13**(1): p. 90-6.
167. Christensen, D.J. and C.D. Poulter, *Enzymatic synthesis of isotopically labeled isoprenoid diphosphates.* Bioorg Med Chem, 1994. **2**(7): p. 631-7.
168. Rose, M.W., et al., *Evaluation of geranylazide and farnesylazide diphosphate for incorporation of prenylazides into a CAAX box-containing peptide using protein farnesyltransferase.* J Pept Res, 2005. **65**(6): p. 529-37.
169. Page, M.J. and E. Di Cera, *Role of Na+ and K+ in enzyme function.* Physiol Rev, 2006. **86**(4): p. 1049-92.
170. Martin, V.J., et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.* Nat Biotechnol, 2003. **21**(7): p. 796-802.
171. Martin, V.J.J., Y. Yoshikuni, and J.D. Keasling, *The in vivo synthesis of plant sesquiterpenes by Escherichia coli.* Biotechnol Bioeng, 2001. **75**(5): p. 497-503.
172. Lefurgy, S.T., et al., *Probing ligand-binding pockets of the mevalonate pathway enzymes from Streptococcus pneumoniae.* J Biol Chem, 2010. **285**(27): p. 20654-63.
173. Andreassi, J.L., 2nd, K. Dabovic, and T.S. Leyh, *Streptococcus pneumoniae isoprenoid biosynthesis is downregulated by diphosphomevalonate: an antimicrobial target.* Biochemistry, 2004. **43**(51): p. 16461-6.
174. Patil, K.R., M. Akesson, and J. Nielsen, *Use of genome-scale microbial models for metabolic engineering.* Curr Opin Biotechnol, 2004. **15**(1): p. 64-9.
175. Gombert, A.K. and J. Nielsen, *Mathematical modelling of metabolism.* Current Opinion in Biotechnology, 2000. **11**(2): p. 180-186.
176. Ishii, N., et al., *Dynamic simulation of an in vitro multi-enzyme system.* Febs Letters, 2007. **581**(3): p. 413-420.
177. Hold, C. and S. Panke, *Towards the engineering of in vitro systems.* J R Soc Interface, 2009. **6 Suppl 4**: p. S507-21.

178. van Can, H.J., et al., *An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances: Part II.* Biotechnol Bioeng, 1999. **62**(6): p. 666-680.
179. van Can, H.J., et al., *An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances.* Biotechnol Bioeng, 1997. **54**(6): p. 549-66.
180. Tusek, A. and Z. Kurtanjek, *Lin-log Model of E-coli Central Metabolism.* Acta Chimica Slovenica, 2010. **57**(1): p. 52-59.
181. Giersch, C., *Determining elasticities from multiple measurements of flux rates and metabolite concentrations. Application of the multiple modulation method to a reconstituted pathway.* Eur J Biochem, 1995. **227**(1-2): p. 194-201.
182. Kim, D.M. and J.R. Swartz, *Prolonging cell-free protein synthesis with a novel ATP regeneration system.* Biotechnology and Bioengineering, 1999. **66**(3): p. 180-188.
183. Baykov, A.A., et al., *Functional characterization of Escherichia coli inorganic pyrophosphatase in zwitterionic buffers.* Eur J Biochem, 1999. **260**(2): p. 308-17.
184. Hold, C. and S. Panke, *Towards the engineering of in vitro systems.* Journal of the Royal Society Interface, 2009. **6**.
185. Brakoulias, A. and R.M. Jackson, *Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching.* Proteins, 2004. **56**(2): p. 250-60.
186. Zou, R., et al., *Combinatorial Engineering of 1-Deoxy-D-Xylulose 5-Phosphate Pathway Using Cross-Lapping In Vitro Assembly (CLIVA) Method.* PLoS One, 2013. **8**(11): p. e79557.
187. Hahn, F.M., A.P. Hurlburt, and C.D. Poulter, *Escherichia coli open reading frame 696 is idi, a nonessential gene encoding isopentenyl diphosphate isomerase.* J Bacteriol, 1999. **181**(15): p. 4499-504.
188. Visser, D. and J.J. Heijnen, *The mathematics of metabolic control analysis revisited.* Metab Eng, 2002. **4**(2): p. 114-23.
189. Onsager, L., *Reciprocal Relations in Irreversible Processes. I.* Physical Review, 1931. **37**(4): p. 405-426.
190. Bayne, L., R.V. Ulijn, and P.J. Halling, *Effect of pore size on the performance of immobilised enzymes.* Chem Soc Rev, 2013. **42**(23): p. 9000-10.
191. Liese, A. and L. Hilterhaus, *Evaluation of immobilized enzymes for industrial applications.* Chem Soc Rev, 2013. **42**(15): p. 6236-49.
192. Tischer, W. and V. Kasche, *Immobilized enzymes: crystals or carriers?* Trends Biotechnol, 1999. **17**(8): p. 326-35.
193. Wang, W., D.I. Wang, and Z. Li, *Facile fabrication of recyclable and active nanobiocatalyst: purification and immobilization of enzyme in one pot with Ni-NTA functionalized magnetic nanoparticle.* Chem Commun (Camb), 2011. **47**(28): p. 8115-7.

194. Rodrigues, R.C., et al., *Modifying enzyme activity and selectivity by immobilization.* Chem Soc Rev, 2013. **42**(15): p. 6290-307.
195. Dean Brady, J.J., *Advances in enzyme immobilization.* Biotechnol Lett, 2009. **31**: p. 1639-1650.
196. Zhang, Y.H.P., *Substrate channeling and enzyme complexes for biotechnological applications.* Biotechnology Advances, 2011. **29**(6): p. 715-725.
197. H. Olin Spivey, J.O., *Substrate Channeling.* Methods, 1999. **19**: p. 306-321.
198. You, C. and Y.H. Zhang, *Self-assembly of synthetic metabolons through synthetic protein scaffolds: one-step purification, co-immobilization, and substrate channeling.* ACS Synth Biol, 2013. **2**(2): p. 102-10.
199. Zhang, Y.H.P., et al., *Toward low-cost biomanufacturing through in vitro synthetic biology: bottom-up design.* Journal of Materials Chemistry, 2011. **21**(47): p. 18877-18886.
200. Harris, J.M., K.L. Epting, and R.M. Kelly, *N-terminal fusion of a hyperthermophilic chitin-binding domain to xylose isomerase from Thermotoga neapolitana enhances kinetics and thermostability of both free and immobilized enzymes.* Biotechnol Prog, 2010. **26**(4): p. 993-1000.
201. Wang, L., et al., *Specific and reversible immobilization of NADH oxidase on functionalized carbon nanotubes.* J Biotechnol, 2010. **150**(1): p. 57-63.
202. Ley, C., et al., *Immobilization of histidine-tagged proteins on electrodes.* Colloids Surf B Biointerfaces, 2011. **88**(2): p. 539-51.
203. Kang, E., et al., *Specific adsorption of histidine-tagged proteins on silica surfaces modified with Ni2+/NTA-derivatized poly(ethylene glycol).* Langmuir, 2007. **23**(11): p. 6281-8.
204. Kim, D.M. and J.R. Swartz, *Prolonging cell-free protein synthesis with a novel ATP regeneration system.* Biotechnol Bioeng, 1999. **66**(3): p. 180-8.
205. Haudenschild, C., et al., *Functional expression of regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha spp.) in Escherichia coli and saccharomyces cerevisiae.* Arch Biochem Biophys, 2000. **379**(1): p. 127-36.
206. Pscheidt, B. and A. Glieder, *Yeast cell factories for fine chemical and API production.* Microb Cell Fact, 2008. **7**: p. 25.
207. Zhang, M., et al., *Deletion of yeast CWP genes enhances cell permeability to genotoxic agents.* Toxicol Sci, 2008. **103**(1): p. 68-76.
208. Kaul, S., S.S. Sharma, and I.K. Mehta, *Free radical scavenging potential of L-proline: evidence from in vitro assays.* Amino Acids, 2008. **34**(2): p. 315-20.
209. Zhang, Y., et al., *The molecular cloning of artemisinic aldehyde Delta11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in Artemisia annua.* J Biol Chem, 2008. **283**(31): p. 21501-8.
210. Mira, N.P., M.C. Teixeira, and I. Sa-Correia, *Adaptive response and tolerance to weak acids in Saccharomyces cerevisiae: a genome-wide view.* OMICS, 2010. **14**(5): p. 525-40.

211. Pompon, D., et al., *Yeast expression of animal and plant P450s in optimized redox environments.* Methods Enzymol, 1996. **272**: p. 51-64.
212. Delic, M., D. Mattanovich, and B. Gasser, *Monitoring intracellular redox conditions in the endoplasmic reticulum of living yeasts.* FEMS Microbiol Lett, 2010. **306**(1): p. 61-6.
213. Gietz, R.D. and R.H. Schiestl, *Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method.* Nat Protoc, 2007. **2**(1): p. 1-4.
214. Disch, A., et al., *Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium Synechocystis PCC 6714.* Biochem J, 1998. **333 (Pt 2)**: p. 381-8.
215. Rohdich, F., A. Bacher, and W. Eisenreich, *Isoprenoid biosynthetic pathways as anti-infective drug targets.* Biochem Soc Trans, 2005. **33**(Pt 4): p. 785-91.
216. Matsue, Y., et al., *The herbicide ketoclomazone inhibits 1-deoxy-D-xylulose 5-phosphate synthase in the 2-C-methyl-D-erythritol 4-phosphate pathway and shows antibacterial activity against Haemophilus influenzae.* Journal of Antibiotics, 2010. **63**(10): p. 583-588.
217. Kuzuyama, T., et al., *Fosmidomycin, a specific inhibitor of 1-deoxy-d-xylulose 5-phosphate reductoisomerase in the nonmevalonate pathway for terpenoid biosynthesis.* Tetrahedron Letters, 1998. **39**(43): p. 7913-7916.
218. Zinglé, C., et al., *Isoprenoid Biosynthesis via the Methylerythritol Phosphate Pathway: Structural Variations around Phosphonate Anchor and Spacer of Fosmidomycin, a Potent Inhibitor of Deoxyxylulose Phosphate Reductoisomerase.* The Journal of Organic Chemistry, 2010. **75**(10): p. 3203-3207.
219. Dugar, D. and G. Stephanopoulos, *Relative potential of biosynthetic pathways for biofuels and bio-based products.* Nature biotechnology, 2011. **29**(12): p. 1074-1078.
220. Partow, S., et al., *Reconstruction and evaluation of the synthetic bacterial MEP pathway in Saccharomyces cerevisiae.* PLoS One, 2012. **7**(12): p. e52498.
221. Carlsen, S., et al., *Heterologous expression and characterization of bacterial 2-C-methyl-D-erythritol-4-phosphate pathway in Saccharomyces cerevisiae.* Appl Microbiol Biotechnol, 2013. **97**(13): p. 5753-69.
222. Zepeck, F., et al., *Biosynthesis of isoprenoids. purification and properties of IspG protein from Escherichia coli.* J Org Chem, 2005. **70**(23): p. 9168-74.
223. Grawert, T., et al., *Structure of active IspH enzyme from Escherichia coli provides mechanistic insights into substrate reduction.* Angew Chem Int Ed Engl, 2009. **48**(31): p. 5756-9.
224. Lee, M., et al., *Biosynthesis of Isoprenoids: Crystal Structure of the [4Fe-4S] Cluster Protein IspG.* Journal of Molecular Biology, 2010. **404**(4): p. 600-610.

225. Narayanasamy, P. and D.C. Crick, *Enantiomeric Synthesis of 2-C-Methyl-D-Erythritol 2, 4-Cyclodiphosphate.* Heterocycles, 2008. **76**(1): p. 243-247.
226. Bernal, C., et al., *A spectrophotometric assay for the determination of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase activity.* Anal Biochem, 2005. **340**(2): p. 245-51.
227. Bitok, J.K. and C.F. Meyers, *2C-Methyl-d-erythritol 4-phosphate enhances and sustains cyclodiphosphate synthase IspF activity.* ACS Chem Biol, 2012. **7**(10): p. 1702-10.
228. Proteau, P.J., *1-Deoxy-D-xylulose 5-phosphate reductoisomerase: an overview* Bioorganic Chemistry 2004 **32**: p. 483-493.
229. Brammer, L.A., et al., *1-Deoxy-D-xylulose 5-Phosphate Synthase Catalyzes a Novel Random Sequential Mechanism.* Journal of Biological Chemistry, 2011. **286**(42): p. 36522-36531.
230. Ardao, I., et al., *One step purification–immobilization of fuculose-1-phosphate aldolase, a class II DHAP dependent aldolase, by using metal-chelate supports.* Enzyme and Microbial Technology, 2006. **39**(1): p. 22-27.
231. Brammer, L.A. and C.F. Meyers, *Revealing Substrate Promiscuity of 1-Deoxy-D-xylulose 5-Phosphate Synthase.* Organic Letters, 2009. **11**(20): p. 4748-4751.
232. Schoffelen, S. and J. van Hest, *Chemical approaches for the construction of multi-enzyme reaction systems.* Current opinion in structural biology, 2013. **23**(4): p. 613-621.
233. Sprenger, G.A., et al., *Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 12857-62.
234. Lois, L.M., et al., *Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis.* Proc Natl Acad Sci U S A, 1998. **95**(5): p. 2105-10.
235. Leighanne A. Brammer, C.F.M., *Revealing substrate promiscuity of 1-deoxy-D-xylulose 5-phosphate synthase* Org Lett, 2009. **11**: p. 4748-4751.
236. Schürmann, M., M. Schürmann, and G.A. Sprenger, *Fructose 6-phosphate aldolase and 1-deoxy-d-xylulose 5-phosphate synthase from Escherichia coli as tools in enzymatic synthesis of 1-deoxysugars.* Journal of Molecular Catalysis B: Enzymatic, 2002. **19–20**(0): p. 247-252.
237. Brammer, L.A., et al., *1-Deoxy-D-Xylulose 5-phosphate synthase: A novel random sequential mechanism in thiamine Diphosphate-dependent enzymology.* J Biol Chem, 2011.
238. Banerjee, A., et al., *Feedback inhibition of deoxy-D-xylulose-5-phosphate synthase regulates the methylerythritol 4-phosphate pathway.* J Biol Chem, 2013. **288**(23): p. 16926-36.

239. Xiang, S., et al., *Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis.* J Biol Chem, 2007. **282**(4): p. 2676-82.
240. Hemmerlin, A., et al., *A cytosolic Arabidopsis D-xylulose kinase catalyzes the phosphorylation of 1-deoxy-D-xylulose into a precursor of the plastidial isoprenoid pathway.* Plant Physiol, 2006. **142**(2): p. 441-57.
241. Wungsintaweekul, J., et al., *Phosphorylation of 1-deoxy-D-xylulose by D-xylulokinase of Escherichia coli.* Eur J Biochem, 2001. **268**(2): p. 310-6.
242. Dhiman, R.K., et al., *1-Deoxy-D-xylulose 5-phosphate reductoisomerase (IspC) from Mycobacterium tuberculosis: towards understanding mycobacterial resistance to fosmidomycin.* J Bacteriol, 2005. **187**(24): p. 8395-402.
243. Yajima, S., et al., *Structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in a quaternary complex with a magnesium ion, NADPH and the antimalarial drug fosmidomycin.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2007. **63**(Pt 6): p. 466-70.
244. Sangari, F.J., et al., *A new family of enzymes catalyzing the first committed step of the methylerythritol 4-phosphate (MEP) pathway for isoprenoid biosynthesis in bacteria.* Proceedings of the National Academy of Sciences, 2010. **107**.
245. Neuzi, P., et al., *Revisiting lab-on-a-chip technology for drug discovery.* Nat Rev Drug Discov, 2012. **11**(8): p. 620-32.
246. Singh, S.M. and A.K. Panda, *Solubilization and refolding of bacterial inclusion body proteins.* J Biosci Bioeng, 2005. **99**(4): p. 303-10.
247. Dong, X.Y., L.J. Chen, and Y. Sun, *Refolding and purification of histidine-tagged protein by artificial chaperone-assisted metal affinity chromatography.* J Chromatogr A, 2009. **1216**(27): p. 5207-13.
248. Goloubinoff, P., et al., *Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13732-7.
249. Zawada, J.F., et al., *Microscale to Manufacturing Scale-up of Cell-Free Cytokine Production-A New Approach for Shortening Protein Production Development Timelines.* Biotechnology and Bioengineering, 2011. **108**(7): p. 1570-1578.

Appendices

List of Publications

1. **Chen X**, Zhang C, Zou R, Zhou K, Stephanopoulos G, Too HP. (2013) Statistical Experimental Design Guided Optimization of a One-Pot Biphasic Multienzyme Total Synthesis of Amorpha-4,11-diene. PLoS ONE 8: e79650. (Chapter 3)

2. Zhang C, **Chen X**, Zou R, Zhou K, Stephanopoulos G, Too HP. (2013) Combining Genotype Improvement and Statistical Media Optimization for Isoprenoid Production in E. coli. PLoS One 8: e75164.

3. **Chen X**, Zhang C, Zou R, Stephanopoulos G, Too HP. Unravel the regulatory behavior of the in vitro reconstituted amorpha-4,11-diene synthesis pathway by Lin-log approximation. Manuscript in preparation (Chatper 4 and Chapter 5)

4. **Chen X**, Zou R, Zhang C, Stephanopoulos G, Too HP. Hybrid in vivo and in vitro production of artemisinic acid. Manuscript in preparation. (Chapter 6)

5. Zhang C, **Chen X**, Zou R, Zhou K, Stephanopoulos G, Too HP. (2014) Experimental design aided systematic pathway optimization of glucose uptake and deoxyxylulose phosphate pathway to enhance the production of amorphadiene. Biotechnology & Bioengineering. (manuscript submitted)

List of invention disclosures

1. In *vitro* synthetic multi-biocatalytic system for the total synthesis of isoprenoids and isoprenoid precursors. US Provisional Application No. 61/871,940. Inventor: Heng Phon TOO, Xixian CHEN, Congqiang ZHANG, Ruiyang ZOU